This study addresses constructing Minimum Spanning Trees (MST) in stochastic weighted distribution networks, where edge costs have inherent uncertainties with known means and variances. Traditional deterministic methods often fail, and existing stochastic approaches are frequently unstable or computationally complex under high uncertainty. A novel variance-based deterministic transformation algorithm is proposed. Its core feature is transforming stochastic edge costs into robust deterministic equivalents by computing an aggregate variance term from the largest (n – 1) edge variances, enabling MST construction via classical algorithms. This method fundamentally enhances stability and ensures feasibility, particularly in high-variance scenarios, improving upon traditional confidence interval-based techniques. The algorithm's efficacy was rigorously validated. Its performance was compared against a probabilistic Q_{ij}-based method under moderate variance, demonstrating consistent and accurate MSTs. It was then applied to a complex 21-edge distribution network with high variance parameters. Results confirm the algorithm's broad applicability, precision, and capability to construct reliable spanning trees under both moderate and substantial uncertainty. The algorithm demonstrates significant computational efficiency (O(r log r)), ensuring practicality and scalability across varying uncertainty levels. Unlike iterative or constraint-heavy models, this algorithm simplifies optimization while preserving uncertainty representation. This makes it well-suited for large-scale networks and real-world systems where cost variability is critical. Future research includes expanding this approach to multi-objective optimization or dynamic networks

Keywords: spanning tree, stochastic graph, variance-based algorithm, deterministic transformation, uncertainty modeling, network optimization

UDC 621.316.925
DOI: 10.15587/1729-4061.2025.329685

DEVELOPMENT OF A VARIANCE-BASED DETERMINISTIC ALGORITHM FOR STOCHASTIC MST IN DISTRIBUTION NETWORKS

Anna Angela Sitinjak
Magister of Science*
Saib Suwilo
Corresponding author
Professor, Doctor of Philosophy, Magister of Science*
E-mail: saib@usu.ac.id
Mardiningsih Mardiningsih
Doctor of Philosophy, Magister of Science*
Sutarman Sutarman
Doctor of Philosophy, Magister of Science*
*Department of Mathematics, Doctoral Program, FMIPA
Universitas Sumatera Utara
Dr. T. Mansur str., 9, Medan, Indonesia, 20222

Received 03.04.2025
Received in revised form 15.05.2025
Accepted 10.06.2025
Published 25.06.2025

How to Cite: *Sitinjak, A. A., Suwilo, S., Mardiningsih, M., Sutarman, S. (2025). Development of a variance-based deterministic algorithm for stochastic MST in distribution networks. Eastern-European Journal of Enterprise Technologies, 3 (4 (135)), 42–51. https://doi.org/10.15587/1729-4061.2025.329685*

1. Introduction

Industrial zones are areas specifically designed for production and manufacturing activities, and at the same time serve as key centers for economic growth and technological advancement [1]. These zones are typically equipped with various supporting facilities such as factories, warehouses, and other infrastructure to accommodate a wide range of industrial operations, including heavy industry, chemical manufacturing, textiles, and high-technology sectors. The infrastructure in these zones must be capable of supporting essential processes such as transportation, raw material distribution, and the transformation of goods into finished products [2]. In many cases, industrial zones are also equipped with strict security systems and structured waste management to maintain environmental stability [3]. In addition, they foster the growth of related sectors such as financial services, marketing, and logistics, while also serving as hubs for research and innovation that accelerate technological development and production efficiency. One of the most vital infrastructure components that ensures the continuity of industrial operations is an effective clean water distribution system [4]. Clean water is not only essential for human consumption but is also used in various industrial processes such as machine cooling, cleaning, and as a raw material in certain industries. Therefore, an efficient water distribution network is critical not only to support productivity but also to maintain environmental sustainability. A limited water supply or uneven distribution can result in serious disruptions to industrial activities and negatively affect the health and well-being of workers [5].

However, designing a clean water distribution system in industrial zones poses a major challenge due to inherent uncertainties – particularly in installation costs. These costs may vary depending on topographical conditions, pipe material types, operational factors, and broader economic dynamics including resource availability and pricing volatility. Specifically, in paper [6] highlighted the use of chance-constrained programming for stochastic network design problems, underscoring the need to manage uncertainty in costs. The study conducted by [7] contributed to methods for stochastic programs with integer recourse, which are relevant when considering discrete decisions in uncertain cost scenarios. Furthermore, in paper [8] explored probabilistic analysis of MST in smart water networks using fuzzy edge weights, demonstrating how different uncertainty representations can impact network design. As such, a comprehensive modeling approach that accounts for these uncertainties is essential. Graph theory offers a suitable framework for modeling such systems, where each water-consuming facility is represented as a vertex, and the water pipelines are modeled as edges weighted by estimated costs. The primary objective is to determine a Minimum Spanning Tree (MST) that connects all vertices while minimizing total cost. MST algorithms have been widely applied in industrial contexts to design cost-efficient networks. Nevertheless, when edge weights are stochastic, typically assumed to follow a normal distribution, conventional deterministic MST algorithms often fail to yield feasible or accurate results. In this context, stochastic cost represents the potential variability in pipeline costs due to external uncertainties such as fluctuations in material prices or weather conditions. Although prior research has introduced several probabilistic spanning tree models, many of them remain limited in their applicability, particularly under high variance. Considering these unresolved challenges, the development of new spanning tree algorithms that remain computationally efficient and reliable under high-variance conditions is necessary. Therefore, research on the development of spanning tree models under high stochastic variance is relevant.

2. Literature review and problem statement

The Minimum Spanning Tree (MST) problem under stochastic conditions has been a significant area of research, particularly in scenarios where edge costs are modeled as random variables. The foundational study by [9] introduced a stochastic spanning tree model, offering a theoretical framework to solve uncertainties in edge costs. Their approach utilized probabilistic methods to identify spanning trees under stochastic settings, but the complexity of the computations posed challenges for larger networks.

Building on this foundation, in paper [10] a learning automata-based heuristic algorithm was proposed, which improved computational efficiency through dynamic adaptation to changing network conditions. While this method reduced computational overhead compared to earlier probabilistic approaches, its applicability to large-scale networks remained limited. Prior research, such as paper [11], introduced chance-constrained programming frameworks for robust decision-making in uncertain power systems. While providing structured approaches to handle stochastic parameters, these works did not directly address optimal distribution network partitioning or the specific uncertainties in load and distributed energy resources tackled by the methodology presented here. Extending the stochastic MST paradigm, study from [12] developed a distributional robust model. This approach reduced the reliance on complete distributional knowledge, demonstrating improved adaptability to uncertain conditions by considering a set of possible distributions rather than a single, precisely known one.

Similarly, in paper [13], the field was advanced with a sample-based stochastic MST method, optimizing performance under limited data availability while maintaining computational efficiency. In paper [14], MST construction was examined under edge and vertex uncertainties, leading to the introduction of the MST-U and V-MST-U models. This research focused on specific graph structures, such as cactus graphs, and analyzed the competitive bounds of their proposed algorithms, providing valuable insights into structural uncertainties. However, it did not specifically address high-variance conditions, which remain a significant challenge. In paper [15], the Random Spanning Trees in Random Environment (RSTRE) model was introduced, investigating the effects of disorder strength on tree configurations. Their findings highlighted how structural variability influences the properties of spanning trees in stochastic settings, contributing to a deeper understanding of tree resilience under uncertainty.

While paper [9] and subsequent studies laid the groundwork for MST under stochastic conditions, challenges related to scalability, computational efficiency, and restrictive distributional assumptions persist. Many existing approaches are either computationally intensive or limited by their reliance on probabilistic and heuristic methods. Specifically, the problem of developing robust and computationally efficient spanning tree algorithms that can reliably handle high-variance stochastic edge costs in large-scale networks remains largely unexplored. This gap exists primarily because high variance introduces significant complexity, making it difficult to guarantee both solution quality and computational tractability. Traditional probabilistic methods often struggle with the wide range of possible outcomes in high-variance scenarios, while heuristic methods may lack the necessary robustness. Furthermore, integrating high-variance considerations with scalability for large networks poses a formidable challenge that has not been comprehensively addressed. These limitations underscore the need for methodologies that prioritize robustness and efficiency without compromising applicability to large-scale networks, particularly under extreme uncertainty.

Solving these gaps, this study proposes a variance-based deterministic transformation algorithm. By converting stochastic edge costs into deterministic equivalents using the square root of the sum of the $(n-1)$ the largest variances $\left(\sqrt{PD}\right)$, the proposed method enhances computational efficiency and reliability. This approach provides a scalable solution for stochastic spanning tree problems and advances the state of the art in this domain.

3. The aim and objectives of the study

The aim of the study is to propose a heuristic algorithm involving a variance-based deterministic transformation that is capable of solving the stochastic Spanning Tree (MST) problem in environments characterized by high uncertainty. This approach is designed to enhance the reliability of distribution network design by converting stochastic edge costs into deterministic values while preserving the representation of uncertainty.

To achieve this aim, the following objectives are accomplished:

– to propose the concept and architecture of a variance-based deterministic algorithm for stochastic MST;

– to validate the proposed variance-based deterministic algorithm by comparing its performance with Ishii's Q_{ij}-based method under moderate variance conditions;

– to apply the proposed algorithm to a stochastic distribution network under high variance conditions and construct the resulting spanning tree.

4. Materials and methods

The object of the study is a stochastic weighted graph, which represents a distribution network operating under uncertainty. In this graph, each edge is modeled as a random variable with a known mean and variance, representing variability in costs or other metrics influenced by stochastic conditions. This study hypothesizes that the proposed algorithm can make a spanning tree on such a graph through a modified method of variance-based algorithm. The methodology assumes that all edge costs follow a normal distribution, simplifying the analysis by excluding correlations or dependencies between edges.

To validate this approach, a case study is conducted on a stochastic graph comprising 21 edges, each defined by its mean and variance. The proposed algorithm is applied by transforming the edge costs, running a new hybrid algorithm, and verifying the validity of the resulting spanning tree. The deterministic cost of the tree is calculated as the sum of the deterministic costs of its edges. This methodology is justified by its ability to integrate uncertainty into a structured deterministic framework. The proposed algorithm addresses limitations in scalability and robustness observed in probabilistic or heuristic methods and ensures computational efficiency in stochastic graph applications.

5. Results of variance-based deterministic algorithm

5. 1. Concept and architecture of the proposed algorithm

The proposed algorithm introduces a variance-based deterministic transformation to solve the uncertainty in edge costs and incorporates this transformation into the process of spanning tree construction. This approach constructs the spanning tree from the transformed deterministic costs. By combining the variance-based transformation with a classical deterministic spanning tree algorithm, the proposed approach enables efficient handling of uncertainty while ensuring compatibility with established computational techniques. For clarity, the relevant abbreviations used in the algorithm's description are provided below:

– $\bar{m}_j$ – mean of edge cost;

– σ_j^2 – variance of edge cost;

– K_a – confidence multiplier from the standard normal distribution table;

– MST – minimum spanning tree;

– $\sqrt{PD}$ – square root of the sum of the $(n-1)$ largest variances.

In this stochastic weight model, each edge has a known mean $(\bar{m})$ and variance (σ^2), representing variability in costs or other metrics influenced by stochastic conditions. In this stochastic weight model, each edge has a known mean, $\bar{m}_j$, and variance, σ_j^2. To convert the stochastic weight MST model into a deterministic form, an auxiliary parameter such as $\sqrt{PD}$ may be added for the proposed algorithm. The value of PD is defined as the sum of $(n-1)$ largest variances, where 'n' is the number of vertices in the graph. As a result, the deterministic cost, c_j, of each edge can be expressed using the following formula

$$c_j=\sqrt{PD}\cdot\bar{m}_j+K_\alpha.\sigma_j^2, \tag{1}$$

where K_α – the inverse of the standard normal cumulative distribution function at confidence level, a. The transformation of edge costs is performed using the square root of the sum of the $(n-1)$ largest variances, and the resulting deterministic costs are then used for choosing edges to form the spanning tree. This technique avoids the instability present in confidence interval-based methods and ensures feasibility even when variance values are high. The algorithm proposed in this study is an extension of the classical Minimum Spanning Tree (MST) algorithm, incorporating an additional stage that converts stochastic edge costs into deterministic ones using this variance-based approach. The pseudocode below outlines the proposed variance-based deterministic algorithm for constructing a Minimum Spanning Tree (MST) from a stochastic graph.

Input: Graph with vertices V and edges E
Mean cost of each edge $j \in E$
Variance of each edge $j \in E$
Number of vertices
Output: Output: T edges
// Set of edges in the MST

Begin
// Step 1: Compute $\sqrt{PD}$
variances ← [σ_j^2 for all j ∈ E]
sort(variances) in descending order
PD ← sum of first (n − 1) elements in variances
sqrt_PD ← sqrt(PD)
// Step 2: Transform stochastic costs to deterministic costs
for each edge j ∈ E do {
deterministic_cost[j] ← $\sqrt{PD}\bar{m}_j + K_\alpha\sigma_j^2$
}
// Step 3: Sort edges by deterministic cost in ascending order
sorted_edges ← sort (E) by deterministic_cost
// Step 4: Choose edges
T ← ∅
DS ← makeDisjointSet(V)
i ← 0
while |T| < (n − 1) and i < |sorted_edges| do {
j ← sorted_edges[i]
(u, v) ← endpoints of edge j
if find(DS, u) ≠ find(DS, v) {
T ← T ∪ {j}
union(DS, u, v)
}
i ← i + 1
}

// Step 5: Compute total of minimum cost

Fig. 1. Architecture of the proposed algorithm based on variance-based

5. 2. Validation against Ishii's Q_{ij}-based method under moderate variance

To validate the proposed algorithm, a case study was conducted using a scenario inspired by Ishii's methodology. The results confirm that under conditions of moderate variance, the proposed algorithm produces spanning trees identical to those generated using Ishii's Q_{ij} metric. This alignment validates the proposed approach in scenarios where variance is not excessively high.

Mathematical model of minimum spanning tree with uncertain cost (m) and assumed to be normally distributed, $N\left(\bar{m}_j,\sigma_j^2\right)$, with the following mathematical model

$$\min \sum_{j=1}^{r} m_j x_j. \tag{2}$$

Subject to:

$$\sum_{j=1}^{r} x_j = n-1, \tag{3}$$

$$\sum_{x_j \in (s)} x_j \le |S|-1, \forall S \subset V(G), |S| \ge 3. \tag{4}$$

Decision variables.

Let x_j be the decision variable associated with edge, j, in the distribution network graph. This binary variable indicates whether a particular edge is included in the resulting spanning tree. The variable takes the value: $x_j \in \{0, 1\}$, where x_j is 1 denotes that edge, j, is selected as part of the minimum spanning tree, and x_j is 0 for case otherwise. So, the equation above changes to Z (through chance constrained programming)

$$Z: \text{ Min} = \sum_{j=1}^{r} \bar{m}_j x_j + 1.0 \left(\sum_{j=1}^{r} \sigma_j^2 x_j \right)^{1/2}, \tag{5}$$

where x_j indicates edge selection in the spanning tree, subject to: $x_j \in \{0, 1\}$, where x_j is 1 denotes that edge-j^{th} is selected as part of the minimum spanning tree, and x_j is 0 otherwise. In 1981, the stochastic spanning tree problem was first formulated by modeling each edge cost as a random variable following a normal distribution [9]. This case is illustrated in Fig. 2, which is presented as a single grouped object to ensure layout integrity.

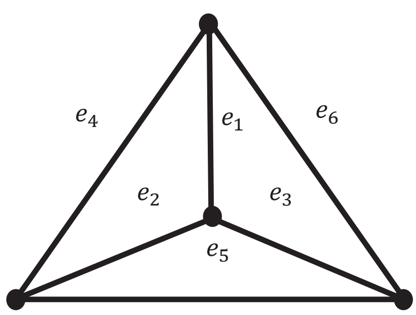

Fig. 2. Graph of edges

The Q_{ij}-based method used in this case follows the formulation presented by [9]. From Fig. 1, each edge connecting a pair of vertices is assumed to have a cost that follows a normal distribution with known parameters mean and variance, $\left(\bar{m}_j, \sigma_j^z\right)$, with a confidence level, α is 0.8413. Edge e_1 follows the distribution $N(16, 0.6)$, while edge e_2 follows $N(49/3, 0.1)$. Edge e_3 is distributed as $N(14, 1)$ and edge e_4 $N(44/3, 0.7)$. Furthermore, edge e_5 follows $N(15, 0.2)$ and edge e_6 follows $N(43/3, 0.2)$, which is approximately $N(14.33, 0.2)$.

From case above, it is known that each cost of $G\left(N\left(\bar{m}_j, \sigma_j^z\right)\right)$ and α is 0.8413. In the standard normal distribution table $N(0, 1)$: K_α is gotten from $F^{-1}(0.8413)$ and equal to 1.0. To compute the case, firstly the pairwise edge-comparison indices Q_{ij} and formulates as

$$Q_{ij} \triangleq \frac{K_\alpha \left(\sigma_i^2 - \sigma_j^2\right)}{\left(\bar{m}_j - \bar{m}_i\right)}, \tag{6}$$

where $\bar{m}_i$ and $\bar{m}_j$ – the mean cost of edges i and j, and σ_i^2 and σ_j^2 their variances. These indices allow identification of edge pairs, which cost distributions differ significantly under the specified confidence level and the resulting Q_{ij} values for all pairs are summarized in Table 1.

Table 1

Q_{ij} comparison indices between edges

$i \backslash j$	1	2	3	4	5	6
1	1	1.5	0.2	0.75	–0.4	1.4
2	0.0*	2	2.7/7	1.8/5	0.3/4	0.05
3	0.0*	0.0*	3	0.9/2	0.8	2.4
4	0.0*	0.0*	0.0*	4	1.5	–1.5/2
5	0.0*	0.0*	0.0*	0.0*	5	0
6	0.0*	0.0*	0.0*	0.0*	0.0*	6

To determine which values of Q_{ij} are acceptable, a threshold interval is computed using the aggregated edge variances. Let n be the number of vertices. The selection requires $(n - 1)$ edges. Define pD as the sum of the $(n - 1)$ smallest variances and PD as the sum of the $(n - 1)$ largest variances

$$pD = \sum\nolimits_{j \in S_{\min}} \sigma_j^2, \; PD = \sum\nolimits_{j \in S_{\max}} \sigma_j^2, \tag{7}$$

where $S_{\min}$ and $S_{\max}$ represent the sets of edges with the smallest and largest variances, respectively. These are converted into threshold parameters using the transformation

$$Lower\_Bound = 2\sqrt{pD}; \; Upper\_Bound = 2\sqrt{PD}. \tag{8}$$

Only Q_{ij} values within this range are retained. The selected edge set T is then constructed from those satisfying

$$Q_{ij} \in \left[2\sqrt{pD}, 2\sqrt{PD}\right]. \tag{9}$$

Based on the valid Q_{ij} values, three edges are selected: e_2, e_5, e_6. These define a spanning tree denoted as x^L is (0, 1, 0, 0, 1, 1). The total cost is evaluated using the following objective function

$$Z^L \triangleq \sum_{j=1}^{r} \bar{m}_j x_j^L + K_\alpha \left(\sum_{j=1}^{r} \left(\sigma_j^2 x_j^L\right)^{1/2} \right). \tag{10}$$

The corresponding values for $\bar{m}_j$ and σ_j of the selected edges are listed in Table 2, and the total cost is computed accordingly.

Table 2

Mean and variance of selected edges

Edge	$\bar{m}_j$	σ_j^2
e_2	49/3	0.1
e_5	15	0.2
e_6	4/3	0.2

Subsequent steps in the Q_{ij}-based method involve iteratively computing intermediate threshold values $\bar{Q}$ by averaging adjacent Q_{ij} values and adjusting the edge cost function accordingly

$$deterministic\_\cos t = \bar{Q}\bar{m}_j + K_\alpha \sigma_j^2. \tag{11}$$

Each deterministic cost is then used to re-evaluate the spanning tree using classical MST algorithm. The result is refined MST configuration, which, in this case, also yields the same optimal tree as in the initial iteration, comprising edges e_3, e_4, e_6. The final solution is (0, 0, 1, 1, 0, 1) as an optimal solution and the associated objective value is

$$Z(Q) \triangleq \sum_{j=1}^{r} m_j x_j^Q + K_\alpha \left(\sum_{j=1}^{r} \sigma_j^2 x_j^Q \right)^{1/2}. \quad (12)$$

The result is a spanning tree with minimum cost total is 44.378. Now, this case will be solved by proposed algorithm.

The proposed hybrid algorithm proceeds as follows:

1. Compute the aggregate variance term

$$\sqrt{PD} = \sqrt{\sum_{j \in S_{\max}} \sigma_j^2}, \quad (13)$$

where $S_{\max}$ consists of the $(n-1)$ largest edge variances.

2. Transform stochastic edge costs into deterministic cost equivalents, c_j as shown in the Table 3.

Table 3

Deterministic costs of edges after transformation

Edge	$\bar{m}_j$	σ_j^2	c_j
e_1	16	0.6	24.92
e_2	49/3	0.1	24.92
e_3	14	1	22.28
e_4	44/3	0.7	22.99
e_5	15	0.2	23.00
e_6	43/3	0.2	21.98

3. Sort all edges based on their deterministic cost, c_j' in ascending order.

4. Select the $(n-1)$ edges with the smallest c_j, ensuring no cycle is formed in the selection, thereby constructing a valid spanning tree.

This process yields a minimum spanning tree that maintains the uncertainty structure of the original stochastic model but resolves it via a deterministic transformation. In the evaluated scenario, the algorithm produces a tree consisting of three edges, selected based solely on their transformed deterministic costs. Fig. 3 illustrates the final MST structure obtained from the proposed algorithm, confirming equivalence with the tree produced by the Q_{ij}-based method.

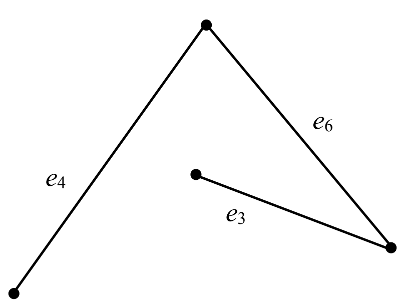

Fig. 3. MST Edges e_3, e_4, and e_6

This outcome confirms that the hybrid algorithm performs consistently under moderate variance and is capable of producing valid and optimal spanning trees, even when traditional probabilistic methods require iterative comparisons. Its use of direct variance aggregation enhances computational efficiency while preserving robustness against uncertainty.

5. 3. Application of the proposed algorithm under high variance

This section assesses the robustness of the proposed method; it was applied to a case characterized by high-cost variance. The following is an example of edge cost data for installing water pipelines. The data include the mean cost and variance for each edge, assumed to follow a normal distribution. These values are summarized in Table 4.

Table 4

Mean and variance of edge costs

Edge	From	To	Mean	Variance
e_1	A	B	16	1
e_2	A	C	18	2
e_3	A	D	14	3
e_4	A	E	20	1
e_5	A	F	22	4
e_6	A	G	17	2
e_7	B	C	19	3
e_8	B	D	15	2
e_9	B	E	18	5
e_{10}	B	F	16	1
e_{11}	B	G	21	3
e_{12}	C	D	23	2
e_{13}	C	E	20	4
e_{14}	C	F	24	3
e_{15}	C	G	25	1
e_{16}	D	E	18	2
e_{17}	D	F	22	4
e_{18}	D	G	20	1
e_{19}	E	F	19	2
e_{20}	E	G	21	3
e_{21}	F	G	16	5

The data in Table 4 provide the essential input for applying the proposed algorithm. Each edge is characterized by a mean cost and a variance, reflecting the uncertainty cost. Based on the data above, the mathematical model is formed using the deterministic transformation of stochastic costs. The model uses the chance-constrained programming, where

$$Z = \sum_{j=1}^{r} \bar{m}_j x_j + K_\alpha \left(\sum_{j=1}^{r} \sigma_j^2 x_j \right)^{1/2}, \quad (14)$$

with $x_j \in \{0,1\}$, where the value of x_j is 1 of edge j is selected [9]. From the standard normal table, K_α is 1.0 because of $F^{-1}(0.8431)$. Therefore, the problem is transformed into a deterministic form as:

$$\min Z = \sum_{j=1}^{r} \bar{m}_j x_j + 1.0 \left(\sum_{j=1}^{r} \sigma_j^2 x_j \right)^{1/2}, \quad (15)$$

$$\sum_{j=1}^{r} x_j = n-1 \text{ (select } n-1 \text{ edges)}, \quad (16)$$

$$\sum_{x_j \in (s)} x_j \le |S| - 1, \ \forall S \subset V(G), \ \|S\| \ge 3 \text{ (no cycle allowed)}, \quad (17)$$

$$x_j \in \{0,1\}. \quad (18)$$

Table 4 illustrates (is an example of) the structure of a water distribution network modeled as a graph, where each vertex represents a factory, and each edge represents a connection between two factories. The costs assigned to each edge reflect the cost associated with installing the pipeline, and are assumed to follow a normal distribution with a known mean and variance. This graph serves as the basis for applying the proposed algorithm to identify the minimum spanning tree under stochastic conditions.

Algorithm for finding the minimum spanning tree using [9] method:

1. Determine the value of Q_{ij}.

For edge of e_1 to e_2

$$Q_{ij} \triangleq \frac{K_\alpha\left(\sigma_i^2-\sigma_j^2\right)}{\left(\bar{m}_j-\bar{m}_i\right)}. \tag{19}$$

The above calculation can be performed on all edges, and the data obtained is Q_{ij} as the Table 5 below.

2. Then select the value of Q_{ij} with $2\sqrt{pD} \leq Q_{ij} \leq 2\sqrt{PD}$ and the result is $5.3 \leq Q_{ij} \leq 10$.

From the Table 5 above, it appears that there is no the value of Q_{ij} that can be selected, so algorithm above cannot be used for cases that have large variance.

When variance levels increase significantly, Ishii's Q_{ij} metric fails to maintain computational efficiency and solution robustness. The proposed heuristic approach addresses this limitation by incorporating variance directly into the deterministic transformation, enabling reliable results even under high-variance conditions. To apply the proposed algorithm, the deterministic cost, c_j, of each edge which K_α is 1.0 is computed using the following formula

$$c_j = \sqrt{PD}\cdot\bar{m}_j + 1.0\cdot\sigma_j^2, \tag{20}$$

where $\sqrt{PD}$ is 5, computer from the sum of 6 largest variance edges. The cost of each edge can be seen in the following Table 6.

After that, choose the smallest cost with the number of selected edges as many as $(n-1)$ edges and no cycles, if the cost values are the same, choose the edge that has the highest variance. Because there are 7 vertices, 6 edges must be selected and there is no cycle, so it is selected, and it can be seen in the Table 7.

Table 5

Q_{ij} values under high variance

Edge	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1.00	–0.50	1.00	0.00	–0.50	–1.00	–0.67	1.00	–2.00	unde-fined	–0.40	–0.14	–0.75	–0.25	0.00	–0.50	–0.50	0.00	–0.33	–0.40	unde-fined
2	0.0	2.00	0.25	0.50	–0.50	0.00	–1.00	0.00	unde-fined	–0.50	–0.33	0.00	–1.00	–0.17	0.14	unde-fined	–0.50	0.50	0.00	–0.33	1.50
3	0.0	0.0	3.00	0.33	–0.125	0.33	0.00	1.00	–0.50	1.00	0.00	0.11	–0.17	0.00	0.18	0.25	–0.13	0.33	0.20	0.00	–1.00
4	0.0	0.0	0.0	4.00	–1.50	0.33	2.00	0.20	2.00	0.00	–2.00	–0.33	unde-fined	–0.50	0.00	0.50	–1.50	unde-fined	1.00	–2.00	1.00
5	0.0	0.0	0.0	0.0	5.00	–0.40	–0.33	–0.29	0.25	–0.50	–1.00	2.00	0.00	0.50	1.00	–0.50	unde-fined	–1.50	–0.67	–1.00	0.17
6	0.0	0.0	0.0	0.0	0.0	6	–0.50	0.00	–3.00	–1.00	–0.25	0.00	–0.67	–0.14	0.13	0.00	–0.40	0.33	0.00	–0.25	3.00
7	0.0	0.0	0.0	0.0	0.0	0.0	7.00	–0.25	2.00	–0.67	0.00	0.25	–1.00	0.00	0.33	–1.00	–0.33	2.00	unde-fined	0.00	0.67
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	8.00	–1.00	1.00	–0.17	0.00	–0.40	–0.11	0.10	0.00	–0.29	0.20	0.00	–0.17	–3.00
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	9.00	–2.00	0.67	0.60	0.50	0.33	0.57	unde-fined	0.25	2.00	3.00	0.67	0.00
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.00	–0.40	–0.14	–0.75	–0.25	0.00	–0.50	–0.50	0.00	–0.33	–0.40	unde-fined
11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	11.00	0.50	1.00	0.00	0.50	–0.33	–1.00	–2.00	–0.50	unde-fined	0.40
12	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	12.00	0.67	–1.00	0.50	0.00	2.00	–0.33	0.00	0.50	0.43
13	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	13.00	0.25	0.60	–1.00	0.00	unde-fined	–2.00	1.00	0.25
14	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	14.00	2.00	–0.17	0.50	–0.50	–0.20	0.00	0.25
15	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	15.00	0.14	1.00	0.00	0.17	0.50	0.44
16	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	16.00	–0.50	0.50	0.00	–0.33	1.50
17	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	17.00	–1.50	–0.67	–1.00	0.17
18	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	18.00	1.00	–2.00	1.00
19	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	19.00	–0.50	1.00
20	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	20.00	0.40
21	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	21.00

Table 6

Deterministic edge costs (transformed)

Edge	c_j
1	81
2	92
3	73
4	101
5	114
6	87
7	98
8	77
9	95
10	81
11	108
12	117
13	104
14	123
15	126
16	92
17	114
18	101
19	97
20	108
21	85

Table 7

Selected edges for final MST

Edge	Deterministic cost (units)	Connected edge
e_3 (AD)	73	A-D
e_8 (BD)	77	A-D-B
e_1 (AB)	81 (not selected due to cycle)	A-D-B
e_{10} (BF)	81	A-D-B-F
e_{21} (FG)	85	A-D-B-F-G
e_6 (AG)	87 (not selected due to cycle)	A-D-B-F-G
e_2 (AC)	92	A-D-B-F-G-C
e_{16} (DE)	92	A-D-B-F-G-C-E

$$TotalCost\_MST = \sum_{j=1}^{r} \bar{m}_j x_j + 1.0\sqrt{\sum_{j=1}^{r} \sigma_j^2 x_j} = 100.88 \text{ unit} \quad (21)$$

These selected edges and the resulting structure are shown in Fig. 4.

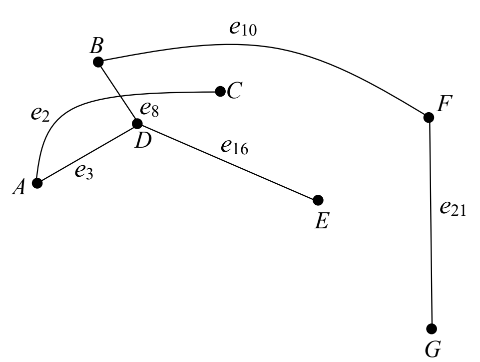

Fig. 4. MST from variance-based algorithm

From the Table 7 above, 6 edges are obtained, namely e_3, e_8, e_{10}, e_{21}, e_2 and e_{16}. The total cost of the spanning tree formed is 100.88 units. The algorithm proposed in this study is an extension of the classical Minimum Spanning Tree (MST) algorithm. Incorporating an additional stage that converts stochastic edge costs into deterministic ones using a variance-based approach. The overall complexity of this algorithm can be analyzed based on the following three main stages: the calculation of $\sqrt{PD}$ from the $(n-1)$ largest variances, which involves sorting the list of all edge variances and selecting the top $(n-1)$ values; the conversion of stochastic costs into deterministic costs; and choosing the selected edges from the smallest until the number of edges is $(n-1)$ and no cycle. From this algorithm, it is resulting in a time complexity of $O(r \log r)$ as the dominant operations remain the sorting of variances and the sorting of edges for MST construction. The additional conversion step introduces only a linear computational cost, which does not affect the algorithm's overall time complexity order.

6. Discussion of results of development of deterministic algorithm

6. 1. Performance and applicability of the variance-based deterministic proposed algorithm

The findings of this study reveal that the proposed variance-based deterministic transformation algorithm provides a robust and computationally efficient solution to the stochastic Spanning Tree problem, particularly in conditions of high uncertainty. This is achieved by transforming stochastic edge costs, originally modeled with known means $(\bar{m})$ and variances (σ^2) under a normal distribution, into deterministic equivalents c_j using an aggregate variance term $\sqrt{PD}$ derived from the square root of the sum of the $(n-1)$ largest variances. This methodological innovation, detailed as the core concept in Section 5. 1, enables the construction of a stable and resilient spanning tree by effectively penalizing edges with higher inherent uncertainty directly within their cost structure. The procedural steps for this transformation and subsequent MST construction are outlined in the pseudocode presented in Fig. 1.

The results obtained providing clear evidence of the algorithm's performance and applicability. Firstly, regarding validation under moderate variance conditions, the proposed algorithm was rigorously validated against Ishii's Q_{ij}-based method, specifically by comparing the outputs in Table 1 and Table 3, the proposed algorithm consistently produced spanning trees identical to those generated using Ishii's Q_{ij} metric under moderate variance conditions. This alignment, visually confirmed by the equivalence of the resulting MST structures illustrated in Fig. 2, 3, demonstrates the proposed method's consistency and validity in scenarios where variance is not excessively high. For example, the final set of selected edges (e_3, e_4, e_6) and the total cost derived from the proposed algorithm's transformed costs matched the optimal tree and objective value obtained by Ishii's iterative approach. This confirms that the proposed algorithm can yield valid and optimal spanning trees while offering a more direct and computationally efficient path compared to traditional probabilistic methods that often require iterative comparisons [9].

Secondly, the robustness and practical utility of the proposed method were assessed by applying it to a more complex case characterized by high cost variance. The comprehensive edge cost data, including both mean and high variance values, are summarized in Table 4. A critical finding was the clear inability of Ishii's Q_{ij} method to yield a valid solution under these challenging conditions, as many Q_{ij} values became undefined due to the large variances. This failure highlights a significant limitation of existing methods when faced with substantial uncertainty. In stark contrast, the proposed algorithm successfully computed stable deterministic edge costs, as shown in Table 4, and subsequently identified a minimum spanning tree comprised of edges e_3, e_8, e_{10}, e_{21}, e_2, and e_{16}. This resulted in a total cost of 100.88 units, as detailed in Table 5 and visually represented in Fig. 3. This result unequivocally demonstrates the algorithm's effectiveness, reliability, and superior applicability under real-world stochastic conditions where high variance is prominent.

The proposed variance-based deterministic algorithm offers significant advantages when compared to alternative existing solutions, addressing key limitations prevalent in the literature.

Classical probabilistic techniques, such as the one proposed by [9] and further explored in paper [16], fundamentally rely on the pairwise Q_{ij} metric to compare edges under a chosen confidence level. However, such methods prove impractical and computationally unstable when variances are large or not uniformly distributed, often rendering many Q_{ij} values undefined and the algorithm itself inapplicable. In direct contrast, the proposed method directly incorporates variance into a deterministic transformation (expressed as $c_j = \sqrt{PD} \cdot \bar{m}_j + K_\alpha \cdot \sigma_j^2$, enabling it to yield reliable results even under high-variance conditions and effectively bypassing the computational failures encountered by Q_{ij}-based approaches. This direct embedding mechanism is a core feature that provides a distinct advantage.

Furthermore, while in paper [17] introduced a chance-constrained programming formulation for MST under Gaussian assumptions, their model typically requires complex constraint handling and does not explicitly incorporate variance directly into the edge cost definition in the simplified and integrated manner of our approach. More recently, [18] reported a distributional robust MST framework, which enhances traditional stochastic models by minimizing reliance on full distributional information. However, the algorithm presented in this work offers a novel and practical reformulation by directly embedding the variance as a penalizing factor into the deterministic transformation of edge costs. This allows it to retain essential uncertainty information without resorting to iterative procedures, complex constraint sets, or exhaustive probabilistic comparisons.

This direct embedding of variance significantly enhances computational efficiency, achieving a favorable time complexity of $O(r \log r)$, where 'r' is the number of edges. This efficiency is maintained because the additional transformation step introduces only a linear computational cost, which does not alter the algorithm's overall asymptotic time complexity. Beyond just runtime, this method enhances reliability and robustness, as edges with higher variance are inherently penalized in their transformed cost structure. This inherent characteristic mirrors practical risk aversion strategies often employed in real-world decision-making under uncertainty, making the solution more resilient.

The solutions obtained in this study directly address and effectively close the problematic part of the persistent challenge of developing computationally efficient and reliable spanning tree algorithms capable of operating effectively under high-variance conditions, especially for large-scale networks. This research successfully fills an existing niche by proposing an algorithm that demonstrates both robustness and efficiency, even in scenarios where traditional probabilistic methods like Ishii's Q_{ij} approach fail due to extreme variance.

This achievement is made possible precisely due to the novel variance-based deterministic transformation. By converting stochastic costs into a manageable deterministic form, the algorithm elegantly integrates uncertainty directly into a well-established and computationally tractable deterministic framework. The demonstrated ability of the algorithm to consistently produce valid Minimum Spanning Trees (MSTs) for cases with both moderate and significantly high variance, coupled with its optimal $O(r \log r)$ computational complexity for large networks, provides compelling evidence that the set aim of developing a reliable and efficient heuristic algorithm for stochastic MST under high uncertainty has been comprehensively achieved. Furthermore, the algorithm's reliance on readily accessible statistical parameters (mean and variance of edge costs) contributes significantly to its practical utility, bridging the gap between abstract theoretical stochastic models and their tangible real-world applications where complete probability distributions might be unavailable or difficult to ascertain.

Despite its strengths, this study has certain limitations that should be acknowledged for practical application and future theoretical studies. The proposed method assumes that all edge costs follow a normal distribution. While common in many applications, this assumption might not universally hold true, and the algorithm's direct applicability or optimality might be affected by highly skewed or non-standard cost distributions, which were not explicitly tested in this work. Furthermore, the current approach simplifies the analysis by excluding explicit correlations or dependencies between edges. In more complex, interdependent networks, ignoring such correlations could potentially influence the truly optimal spanning tree.

A notable disadvantage or inherent trade-off of the variance-based deterministic transformation lies in its tendency to lead to a more conservative solution. By inherently penalizing edges with higher variance, the algorithm prioritizes reliability and stability under uncertainty. However, this could potentially lead to the selection of a path that is not the absolute minimum cost if the actual costs of high-variance edges happen to materialize at the lower end of their probabilistic distribution. This inherent conservatism is a design choice aimed at mitigating risk, but it implies that the solution might not always represent the lowest possible cost in retrospect, only the lowest reliable cost under uncertainty.

Building upon the foundational work presented here, several promising directions exist for future research development, each offering potential to enhance the algorithm's scope and applicability. Firstly, it would be valuable to investigate the adaptability and necessary modifications of the variance-based transformation to accommodate non-normal distributions. This could involve exploring non-parametric methods or incorporating concepts from robust statistics to handle diverse real-world data characteristics, thereby broadening the algorithm's applicability. Secondly, integrating dynamic edge weights to address time-evolving networks represents a crucial next step. This would enhance the algorithm's utility for infrastructure systems where costs and operating conditions fluctuate over time, enabling more adaptive network design. Thirdly, exploring multi-objective optimization frameworks, such as simultaneously optimizing for cost, reliability, and other critical performance metrics, could provide more comprehensive and nuanced solutions for complex network design problems in uncertain environments. Finally, conducting extensive validation with larger-scale, real-world network data from various industrial sectors would further establish the algorithm's practical impact and scalability. These developments would contribute significantly to the creation of more robust, scalable, and versatile optimization frameworks for stochastic network design, building directly upon the fundamental step provided by this research.

7. Conclusion

1. This study successfully proposed the concept and architecture of a novel variance-based deterministic algorithm designed to solve the stochastic Minimum Spanning Tree (MST) problem. The core of this approach lies in its unique transformation mechanism, which converts stochastic edge costs (defined by their mean and variance under a normal distribution) into deterministic equivalents. This conversion is facilitated by incorporating an aggregate variance term $\left(\sqrt{PD}\right)$ and a confidence multiplier (K_a), enabling the use of classical deterministic MST algorithms for problems previously complicated by uncertainty.

2. This study successfully validated the proposed variance-based deterministic transformation approach. Under moderate variance conditions, the algorithm consistently yielded spanning tree configurations equivalent to those obtained by the Q_{ij}-based probabilistic method, thereby demonstrating its accuracy and consistency in scenarios of controlled uncertainty.

3. The algorithm was effectively applied to a stochastic distribution network characterized by high variance. This application successfully constructed a minimum spanning tree with a total deterministic cost of 100.88 units, unequivocally confirming the algorithm's strong applicability and reliability even under conditions of substantial uncertainty.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence:

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Luo, S., Yimamu, N., Li, Y., Wu, H., Irfan, M., Hao, Y. (2022). Digitalization and sustainable development: How could digital economy development improve green innovation in China? Business Strategy and the Environment, 32 (4), 1847–1871. https://doi.org/10.1002/bse.3223
2. Feng, R., Qu, X. (2023). Innovation-Driven Industrial Agglomeration Impact on Green Economic Growth in the Yellow River Basin: An Empirical Analysis. Sustainability, 15 (17), 13264. https://doi.org/10.3390/su151713264
3. Mor, S., Ravindra, K. (2023). Municipal solid waste landfills in lower- and middle-income countries: Environmental impacts, challenges and sustainable management practices. Process Safety and Environmental Protection, 174, 510–530. https://doi.org/10.1016/j.psep.2023.04.014
4. Sinha, S. K., Davis, C., Gardoni, P., Babbar-Sebens, M., Stuhr, M., Huston, D. et al. (2023). Water Sector Infrastructure Systems Resilience A Social-Ecological-Technical System-of-Systems and Whole-Life Approach. Cambridge Prisms: Water, 1–50. https://doi.org/10.1017/wat.2023.3
5. Israilova, E., Voronina, A., Shatila, K. (2023). Impact of water scarcity on socio-economic development. E3S Web of Conferences, 458, 08027. https://doi.org/10.1051/e3sconf/202345808027
6. Koutsokosta, A., Katsavounis, S. (2024). Stochastic transitions of a mixed-integer linear programming model for the construction supply chain: chance-constrained programming and two-stage programming. Operational Research, 24 (3). https://doi.org/10.1007/s12351-024-00856-3
7. Talebi, A., Haeri Boroujeni, S. P., Razi, A. (2024). Integrating random regret minimization-based discrete choice models with mixed integer linear programming for revenue optimization. Iran Journal of Computer Science, 8 (1), 21–35. https://doi.org/10.1007/s42044-024-00193-w
8. Heidari, A., Shishehlou, H., Darbandi, M., Navimipour, N. J., Yalcin, S. (2024). A reliable method for data aggregation on the industrial internet of things using a hybrid optimization algorithm and density correlation degree. Cluster Computing, 27 (6), 7521–7539. https://doi.org/10.1007/s10586-024-04351-4
9. Ishii, H., Shiode, S., Nishida, T., Namasuya, Y. (1981). Stochastic spanning tree problem. Discrete Applied Mathematics, 3 (4), 263–273. https://doi.org/10.1016/0166-218x(81)90004-4
10. Akbari Torkestani, J., Meybodi, M. R. (2010). A learning automata-based heuristic algorithm for solving the minimum spanning tree problem in stochastic graphs. The Journal of Supercomputing, 59 (2), 1035–1054. https://doi.org/10.1007/s11227-010-0484-1

11. Zhao, L., Huang, Y., Dai, Q., Yang, L., Chen, F., Wang, L. et al. (2019). Multistage Active Distribution Network Planning With Restricted Operation Scenario Selection. IEEE Access, 7, 121067–121080. https://doi.org/10.1109/access.2019.2936936
12. Xu, Y., Zhang, L. (2023). Target-based Distributionally Robust Minimum Spanning Tree Problem. arXiv. https://doi.org/10.48550/arXiv.2311.10670
13. Hoeksma, R., Speek, G., Uetz, M. (2024). Stochastic Minimum Spanning Trees with a Single Sample. arXiv. https://doi.org/10.48550/arXiv.2409.16119
14. Mathwieser, C., Çela, E. (2024). Special cases of the minimum spanning tree problem under explorable edge and vertex uncertainty. Networks, 83 (3), 587–604. https://doi.org/10.1002/net.22204
15. Makowiec, L., Salvi, M., Sun, R. (2024). Random spanning trees in random environment. arXiv. https://doi.org/10.48550/arXiv.2410.16830
16. Adhikary, K., Pal, P., Poray, J. (2024). The minimum spanning tree problem on networks with neutrosophic numbers. Neutrosophic Sets and Systems, 63, 258–270. Available at: https://fs.unm.edu/NSS/SpanningTree16.pdf
17. Hu, Z., Sun, W., Zhu, S. (2022). Chance constrained programs with Gaussian mixture models. IISE Transactions, 54 (12), 1117–1130. https://doi.org/10.1080/24725854.2021.2001608
18. Chatterjee, S., Pandurangan, G., Pham, N. D. (2020). Distributed MST. Proceedings of the 21st International Conference on Distributed Computing and Networking, 1–10. https://doi.org/10.1145/3369740.3369778