2024 Fifth International Conference on Intelligent Data Science Technologies and Applications (IDSTA)

Optimizing Facility Location Under Disruptions: A Big Data-Driven Approach

1st Tareq Oshan
College of Engineering
Al Ain University
AbuDhabi, UAE
Tareq.Oshan@aau.ac.ae

2nd Nuha Hamada
College of Engineering
Al Ain University
AbuDhabi, UAE
Nuha.Hamada@aau.ac.ae

3rd Mahmoud Sharkawi
College of Engineering
Al Ain University
AbuDhabi, UAE
mahmoud.sharkawi@aau.ac.ae

***Abstract*—In today's dynamic business environment, companies face the dual challenge of reducing costs and enhancing customer satisfaction. Traditional Facility Location Problem (FLP) models often struggle to process the vast amounts of data generated by modern supply chains, particularly during disruptions. To address this issue, our study uses a mathematical model designed for FLP, specifically considering facility disruptions. Leveraging big data, our model employs linearization and relaxation methods to streamline computations, thereby efficiently identifying optimal warehouse locations and capacities. This approach accounts for storage limitations and effectively manages extensive datasets, showcasing its capability to handle disruptions. By integrating the risk of warehouse failure and utilizing big data, our model provides a robust and resilient framework for supply chain network design, leading to more responsive and reliable supply chains. The model has been validated through a real-world case study involving a Canadian company, using the LINGO software package to derive solutions.**

***Index Terms*—Supply Chain, FLP, Facility disruption, Relaxation method, resilient.**

I. Introduction

In today's dynamic business landscape, companies grapple with the dual challenge of reducing costs and enhancing customer satisfaction. Central to achieving this delicate balance is the design of efficient supply chain networks, where decisions about the number, size, and location of facilities play a pivotal role. Traditionally, facility location models assumed that once facilities were built, they would remain operational indefinitely. However, recent studies recognize a critical reality: constructed facilities may face disruptions at any time.

The advent of big data has transformed the way we approach facility location. As supply chains become increasingly complex and interconnected, the sheer volume of data generated—from real-time demand fluctuations to transportation routes and inventory levels—poses both opportunities and challenges. Big data provides unprecedented insights into customer behavior, market dynamics, and operational efficiency. Yet, harnessing this wealth of information effectively remains a formidable task.

Consider the disruptions that can impact facilities: natural disasters, power outages, labor strikes, and more. These events can disrupt supply chains, rendering traditional facility location models inadequate. For instance, during the 2008 electricity cut-off in China, companies faced production halts due to warehouse unavailability. Similarly, the COVID-19 pandemic disrupted global supply chains, affecting companies like Apple.

To address these challenges, our study introduces a mathematical model specifically similar to the one presented in [1], tailored for facility location in the context of disruptions. We focus on the challenges posed by big data considerations. Besides determining the optimal location and size of needed warehouses, our model assigns two warehouses to each branch: a primary warehouse and a backup warehouse. The primary warehouse serves the branch's demand under normal conditions, while the backup warehouse takes over during disruptions. Importantly, we incorporate a relaxation method to streamline computation. By relaxing certain variables, we facilitate the identification of optimal warehouse locations and sizes in the event of facility failure.

To test our proposed model, we examine a Canadian company operating within a disrupted supply chain context. Analyzing real-world data, we consider factors such as transportation costs, demand patterns, and facility failure probabilities. Leveraging the LINGO software package, we aim to enhance supply chain resilience by strategically locating facilities under disruption risk while accounting for big data challenges.

One method for addressing the facility location problem with large datasets is to utilize mathematical programming techniques. This can include linear programming, integer programming, mixed-integer programming, or non-linear programming to model and solve the problem effectively. These methods allow for the consideration of various constraints and objectives, leading to optimal solutions. For instance, Xu et al. [2] devised a hybrid algorithm merging genetic algorithm, simulated annealing, and tabu search to address the capacitated facility location problem with big data, yielding enhanced performance compared to conventional methods. Such methodologies are effective in managing extensive and intricate problems, delivering high-quality solutions within reasonable timeframes. Zhang et al. [3] developed a parallel ant colony optimization algorithm tailored for large-scale facility location problems, demonstrating impressive efficacy on big datasets.

979-8-3503-5475-1/24/$31.00 ©2024 IEEE

In Liu's work [4], a novel data-driven two-stage sparse distributionally robust risk mixed-integer optimization model was introduced for determining optimal locations of processing plants and distribution centers in uncertain supply chain networks, particularly under worst-case scenarios. This study sheds light on how robust optimization can augment supply chain optimization in uncertain conditions.

A significant advantage of leveraging big data in the facility location problem lies in the ability to make informed decisions based on real-time data. Real-time insights into customer demand, for instance, facilitate identifying optimal facility locations to minimize transportation costs and enhance customer service. Moreover, real-time data on transportation expenses, facility costs, and capacity constraints aid in optimizing the facility location problem, thereby enhancing performance. However, a challenge in addressing the facility location problem with big data is ensuring the scalability of algorithms. As data size increases, algorithms must manage escalating computational complexities while providing precise solutions. This necessitates the utilization of parallel processing, distributed computing, or other high-performance computing techniques. For instance, Liu et al. [5] proposed a hybrid tabu search algorithm for the capacitated facility location problem with big data, implementing parallel computing to achieve superior performance. Furthermore, privacy and security considerations must be addressed when utilizing big data in the facility location problem. Big data may encompass sensitive information like customer data, financial records, and strategic insights, necessitating protection against unauthorized access or disclosure. Thus, implementing appropriate data privacy and security measures is crucial to safeguarding data confidentiality and integrity.

II. Mathematical Model

This section presents a mathematical framework designed to optimize the network configuration of a two-tier supply chain, incorporating warehouses and branches, through the use of Mixed Integer Quadratic Optimization. The primary objective of this model is to determine the most advantageous warehouse locations and sizes while also assigning each branch to both a primary and secondary warehouse to address the issue of facility disruptions.

In this model, branches are identified by the index b, with B representing the set of branches ranging from 1 to m. Warehouse locations are indexed by w, where W encompasses the set of locations ranging from 1 to n. Warehouse sizes are denoted by s, with S covering the set of sizes from 1 to q, each size corresponding to an area of A^s square feet.

Products are categorized into distinct groups, labeled by the index j, and J represents the set of categories ranging from 1 to g. The index j is used when referring to specific product categories.

We define the binary variable x_w^s to represent the construction of a warehouse of size s at location w, with a value of 1 if the warehouse is built and 0 otherwise. Similarly, the binary variable y_{wbj}^{ℓ} indicates whether the demand for product j at branch b is met by warehouse w at level ℓ, where $\ell = 1$ represents the primary warehouse assignment and $\ell = 2$ indicates the secondary warehouse assignment. The vectors x and y comprise all variables x_w^s and y_{wbj}^{ℓ}, respectively.

The cost function comprises three components: fixed costs, operational costs, and transportation costs. The fixed cost is modeled as:

$$C_F(x) = \sum_w (f_w + l_w) \sum_s A^s x_w^s. \tag{1}$$

where f_w represents the cost per square foot, in dollars, over the planning horizon, with A^s indicating the number of square feet for a warehouse of size s constructed at location w; and l_w denotes the cost per square foot, in dollars, over the planning horizon, for industrial land at location w.

To model operational and transportation costs, we introduce the common volume unit K for all product categories. The demand for product j from branch b is $d_{bj}K$. Let ν_j^s be the operational cost of handling one K of product j at a size s warehouse. The dependence of the per unit operational cost on warehouse size allows us to capture cost differences related to economies of size and level of technology.

To account for the risk of warehouse failure in our operational and transportation costs, we must first model this risk. Let $0 < p_w < 1$ represent the probability that warehouse w fails. This probability is location-dependent, as the risk of failure varies by location. For example, a warehouse in a coastal city prone to tropical storms or near a fault line will have a higher probability of failure compared to one situated away from such natural hazards. Similarly, warehouses in politically unstable areas or locations lacking reliable electricity and water supplies will have a higher risk compared to those in more stable and well-supported environments

If we do not account for risk, the primary warehouses will supply their assigned branches and the operational cost is:

$$\sum_{w,s,j} \nu_j^s \sum_b \left(d_{bj} \; x_w^s \; y_{wbj}^1 \right). \tag{2}$$

Considering that the probability of warehouse w remaining operational is $(1 - p_w)$, the anticipated operational cost at the primary warehouses can be represented as:

$$\sum_{w,s,j} \nu_j^s \sum_b \Bigg(d_{bj} \; x_w^s \; y_{wbj}^1 \; (1 - p_w) \Bigg). \tag{3}$$

Considering w as the secondary warehouse, the probability that it will supply its assigned branches is the same as the failure probability of its corresponding primary warehouse, which is:

$$\sum_{w' \neq w} (p_{w'} \; y_{w'bj}^1). \tag{4}$$

Thus, the expected operational cost associated with the secondary warehouses is:

$$\sum_{w,s,j} \nu_j^s \sum_b \left(d_{bj} \; x_w^s \; y_{wbj}^2 \sum_{w' \neq w} (p_{w'} \; y_{w'bj}^1) \right). \tag{5}$$

Putting equations (3) and (5) together gives the expected operational cost function:

$$C_O(x,y) = \sum_{w,s,j} \left[\nu_j^s \sum_b d_{bj} \left(x_w^s \; y_{wbj}^1 \; (1-p_w) + x_w^s \; y_{wbj}^2 \sum_{w' \neq w} (p_{w'} \; y_{w'bj}^1) \right) \right], \tag{6}$$

which is cubic in the binary variables.

The cost of transporting a product j depends on its volume in K units, its physical characteristics, and the locations of the branch and warehouse between which it is shipped. Let τ_{wbj} denote the expense of transporting one unit of product j from warehouse w to branch b. Employing a similar approach as that used for operational cost calculation, the overall expected transportation cost is:

$$C_T(y) = \sum_{w,b,j} \left[d_{bj} \, \tau_{wbj} \left(y_{wbj}^1 \;\; (1-p_w) + y_{wbj}^2 \sum_{w' \neq w} (p_{w'} \;\; y_{w'bj}^1) \right) \right]. \tag{7}$$

The expected cost function is:

$$C(x,y) = C_F(x) + C_O(x,y) + C_T(y). \tag{8}$$

We now proceed with the formulation of constraints. Initially, we ensure:

$$\sum_s x_w^s \leq 1, \; \forall \; w, \tag{9}$$

to enforce a single size selection for each warehouse;

$$\sum_w y_{wbj}^\ell = 1, \; \forall \; b, j, \ell, \tag{10}$$

to guarantee that each branch b is assigned a single primary and a single secondary warehouse for supplying product j;

and

$$y_{wbj}^1 + y_{wbj}^2 \leq 1, \; \forall \; w, b, j, \tag{11}$$

to prevent a warehouse from serving as both the primary and secondary warehouse for product j at branch b.

Moving on to management constraints, we consider a predetermined upper bound, U, on the number of warehouses to be constructed:

$$\sum_{s,w} x_w^s \leq U. \tag{12}$$

To accommodate existing warehouses and those preselected for construction at specific sizes, we impose:

$$x_w^s = 1, \forall \; (w,s) \in B, \tag{13}$$

where $B \subset W \times S$, with W and S representing all indices w and s, respectively. If such warehouses do not exist, then $B = \emptyset$, and constraint (13) is omitted. Additionally, if there is no limit on the number of warehouses to be built, constraint (12) is also omitted.

Subsequently, we verify that the projected demand from designated branches does not exceed the storage capacity V^s allocated for a warehouse of size s:

$$\sum_{b,j} d_{bj} \left(y_{wbj}^1 + y_{wbj}^2 \sum_{w' \neq w} (p_{w'} y_{w'bj}^1) \right) \leq \sum_s V^s x_w^s, \; \forall \; w. \tag{14}$$

In summary, the optimization problem seeks to minimize the total cost while satisfying these constraints:

$$\begin{array}{llcl} \text{M:} \;\; \text{Minimize} & C(x,y) & = & C_F(x) + C_O(x,y) + C_T(y) \\ \quad\;\; \text{Subject to:} & & & \text{(9) - (14)}, \\ & x_w^s & \in & \{0,1\}, \; \forall \; s, w, \\ & y_{wbj}^\ell & \in & \{0,1\}, \; \forall \; w, b, j, \ell. \end{array}$$

The model, denoted as M, represents a binary, cubic optimization challenge. If there exists only one product or category, the subscripts j are omitted. The consistency of model M was demonstrated in [1].

A. Linearization to Model M

Using the standard linearization [6], We initiate the standard linearization process for the products $x_w^s \, y_{wbj}^\ell$. Let's define:

$$z_{wbj}^{\ell s} = x_w^s \, y_{wbj}^\ell, \quad \forall \; s, w, b, j, \ell,$$

and introduce the following $8nmqg$ constraints:

$$z_{wbj}^{\ell s} \leq x_w^s, \tag{15}$$

$$z_{wbj}^{\ell s} \leq y_{wbj}^\ell, \tag{16}$$

$$z_{wbj}^{\ell s} \geq x_w^s + y_{wbj}^\ell - 1, \; \text{and} \tag{17}$$

$$z_{wbj}^{\ell s} \geq 0 \; . \tag{18}$$

These constraints ensure that the continuous variable $z_{wbj}^{\ell s}$ takes binary values and equals 1 only if a warehouse of size s, built at location w, supplies branch b at level ℓ with its demand for product j. The reformulated expected operational cost becomes:

$$C_O(z,y) = \sum_{w,s,j} \left[\nu_j^s \sum_b d_{bj} \left(z_{wbj}^{1s} \; (1-p_w) + z_{wbj}^{2s} \sum_{w' \neq w} (p_{w'} \, y_{w'bj}^1) \right) \right]. \tag{19}$$

As in [1], we proceed to linearize the quadratic terms $z^{2s}_{wbj}, y^1_{w'bj}$ by defining:

$$P^* = \max_w p_w \quad \text{and} \quad P_* = \min_w p_w. \tag{20}$$

Let $z^{2s}_{wbj} \sum_{w' \neq w} (p_{w'} y^1_{w'bj}) = Q^s_{wbj}$, for all w, b, s, j. Then, we have the following constraints:

$$0 \leq Q^s_{wbj} \leq P^* \ z^{2s}_{wbj}, \tag{21}$$

and

$$\sum_{w' \neq w} (p_{w'} \ y^1_{w'bj}) - P^* \ (1 - z^{2s}_{wbj}) \leq Q^s_{wbj} \leq \sum_{w' \neq w} (p_{w'} \ y^1_{w'bj}). \tag{22}$$

Similarly, we define $y^2_{wbj} \sum_{w' \neq w} (p_{w'}, y^1_{w'bj}) = O_{wbj}$, for all w, b, j. The constraints for O_{wbj} are:

$$0 \leq O_{wbj} \leq P^* \ y^2_{wbj}, \tag{23}$$

and

$$\sum_{w' \neq w} (p_{w'}, y^1_{w'bj}) - P^*(1 - y^2_{wbj}) \leq O_{wbj} \leq \sum_{w' \neq w} (p_{w'} \ y^1_{w'bj}). \tag{24}$$

Now, let us define the operational cost function in terms of z and Q:

$$C_O(z, Q) = \sum_{w,s,j} \left[\nu^s_j \sum_b d_{bj} \left(z^{1s}_{wbj} \ (1 - p_w) + Q^s_{wbj} \right) \right]. \tag{25}$$

The transportation cost function in terms of y and O is:

$$C_T(y, O) = \sum_{w,b,j} \left[d_{bj} \tau_{wbj} \left(y^1_{wbj} (1 - p_w) + O_{wbj} \right) \right]. \tag{26}$$

Finally, the capacity constraints are:

$$\sum_{b,j} d_{bj} (y^1_{wbj} + O_{wbj}) \leq \sum_s V^s x^s_w, \quad \forall \ w. \tag{27}$$

Our linearized model, $\boldsymbol{LM}$, is represented as follows:

$$\begin{array}{lll} \boldsymbol{LM}: & \text{Minimize} & C(x, y, z, Q, O) = \\ & & C_F(x) + C_O(z, Q) + C_T(y, O) \\ & \text{Subject to:} & (9)-(13), (15)-(18), \\ & & (21)-(24), (27), \\ & & x^s_w \in \{0, 1\}, \ \forall \ s, w, \\ & & y^\ell_{wbj} \in \{0, 1\}, \ \forall \ w, b, j, \ell. \end{array}$$

Oshan and Caron [1] showed that, similar to M, $\boldsymbol{LM}$ maintains consistency.

B. Relaxation to model $\boldsymbol{LM}$

To relax model $\boldsymbol{LM}$, we replace $y^1_{wbj} \in \{0, 1\}$ with $y^1_{wbj} \geq 0$ to obtain

$$\begin{array}{ll} \boldsymbol{RLM}: \text{Minimize} & C(x, y, z, Q, O) = \\ & C_F(x) + C_O(z, Q) + C_T(y, O) \\ \text{Subject to:} & (9)\text{-}(13), (15)\text{-}(18), \\ & (21)\text{-}(24), (27), \\ & x^s_w \in \{0, 1\}, \ \forall \ s, w, \\ & y^1_{wbj} \geq 0, \ \forall \ w, b, j, \\ & y^2_{wbj} \in \{0, 1\}, \ \forall \ w, b, j. \end{array}$$

Oshan and Caron [1] showed the consistency and validly of Model $\boldsymbol{RLM}$.
Table I presents the relative sizes of our three models.

Model	M	$\boldsymbol{LM}$	$\boldsymbol{RLM}$
Binary Variables	$n(q + 2mg) - \|B\| - \|J\|$	$n(q + 2mg) - \|B\| - \|J\|$	$n(q + mg) - \|B\| - \|J\|$
Continuous Variables	0	$nmg(1 + 3q)$	$nmg(2 + 3q)$
Constraints	$2(n + nm)$ $+nmr + \|B\| + \|J\| + 1$	$mr(2 + 5n + 12nq)$ $+2n + \|B\| + \|J\|$	$mg(2 + 6n + 12nq)$ $+2n + \|B\| + \|J\|$

Table I: Comparison of Problem Size.

III. CANADIAN CASE STUDY

We expand upon the case study presented in [7] by integrating the risk of warehouse failure. The company manages a network comprising two warehouses, catering to 158 branches spread across all Canadian provinces and procuring products from global suppliers. With a diverse inventory of around 19,000 products varying in shapes, sizes, and densities, the company aims to develop a two-tier supply chain model for the forthcoming 15 years.

To streamline analysis, we adopted K to represent 1,000 cubic inches, converting all product demands into K units. This simplification enabled us to concentrate on a single component, eliminating the need for subscript j. Considering a 45-day inventory replenishment cycle at existing warehouses, all costs and demands were computed for this duration.

Upon evaluation, we identified 34 potential warehouse sites based on demographic and geographic factors, offering small ($s = 1$), medium ($s = 2$), and large ($s = 3$) size options. This encompassed the current warehouse locations ($w = 1$ and $w = 2$) operating at their current large scale ($s = 3$), delineated as $B = (1, 3), (2, 3)$.

The amortization of land and fabrication costs was spread over 15 years at a 3% interest rate, with payments scheduled every 45 days. Demand d_b in K units was extrapolated from historical data. Internal estimations were utilized to determine operational costs ν^s, transportation costs τ_{wb}, and storage capacities V^s. Following methodologies outlined in [8], [9], failure probabilities p_w were stochastically generated from a uniform distribution $U \sim [0, 0.05]$.

We solved model $\boldsymbol{LM}$ using LINGO 20.0 x64 and Excel 365 running on an Intel i7 laptop with 16 GB of RAM and a 3.30 GHz processor with four cores. The parameters were stored in an Excel sheet. The model consisted of 64,564 variables, of which 10,844 were binary, highlighting the problem's complexity. After 24 hours, which was the time limit for this study, LINGO was unable to find a feasible solution for model $\boldsymbol{LM}$.

Given LINGO's inability to find a feasible solution for model $\boldsymbol{LM}$, we turned to model $\boldsymbol{RLM}$, where the variable y^1_{wbj} is relaxed to be a non-negative continuous variable. Unlike model $\boldsymbol{LM}$, LINGO found the first feasible solution for model $\boldsymbol{RLM}$ in less than 30 minutes. Table II shows that after 24 hours and approximately 43.5 million iterations, LINGO was unable to find an optimal solution but did find a feasible solution with a toal cost of 4.1 million and an optimality gap of 8.88%. LINGO selected three additional warehouses to be built, in addition to the existing two, comprising two large and one small warehouse.

Table II: Computational Results for Model $\boldsymbol{RLM}$

	$\boldsymbol{RLM}$
Model Class	MILP
Total Variables	64,564
Binary Variables	5,472
Constraints	231,386
Iterations $\times 10^6$	43.570
Time (Hours: Minutes: Seconds)	24:00:00
Best Objective $\times 10^6$	4.1089
Objective Bound $\times 10^6$	3.7478
Selected Warehouses (ID, Size)	(1, 3), (2, 3), (6, 3) (30, 3), and (33, 1)
Status	Feasible

To pursue an optimal solution for model $\boldsymbol{RLM}$, we considered relaxing the y^2_{wbj} variables. However, this could result in assignments between non-built warehouses and branches. To address this, we added the following constraint:

$$y^2_{wbj} \leq \sum_s x^s_w, \quad \forall w \in W,\ b \in B,\ j \in J. \tag{28}$$

We call the model $\boldsymbol{RLM}$ with relaxed y^2_{wbj} and constraint (28) as $\boldsymbol{RLM*}$. Table III shows that model $\boldsymbol{RLM*}$ maintained the same total variable count as model $\boldsymbol{RLM}$, but the number of binary variables decreased to 100, solely related to the variable x^s_w. With this relaxation, LINGO required just over 5.5 hours to obtain an optimal solution for model $\boldsymbol{RLM*}$. LINGO selected three additional warehouses: one large and two medium-sized.

As the solution in Table III provided fractional values for the variable y^ℓ_{wbj}, we decided to enforce the presence of the selected warehouses from model $\boldsymbol{RLM*}$ into model $\boldsymbol{LM}$ and then resolve it. To do so, we updated the set B to:

$$B = \{(1,3),(2,3),(4,2),(28,3),(33,2)\} \tag{29}$$

Table IV outlines the computational outcomes. It shows that LINGO was able to obtain an optimal solution for model $\boldsymbol{LM}$ with constraint (29) in about two hours, with an objective function value of 3.9309×10^6. Notice that the total number of variables decreased from 64,564 to 64,561 as we forced the three warehouses from model $\boldsymbol{RLM}$. The objective function in Table IV is higher than in Table III as expected, since model $\boldsymbol{RLM}$ is a relaxed model allowing fractional assignments that reduce the overall cost.

Table III: Computational Results for Model $\boldsymbol{RLM*}$

	$\boldsymbol{RLM*}$
Model Class	MILP
Total Variables	64,564
Binary Variables	100
Constraints	231,386
Iterations $\times 10^6$	16.730
Time (Hours: Minutes: Seconds)	05:41:25
Best Objective $\times 10^6$	3.8803
Objective Bound $\times 10^6$	3.8803
Selected Warehouses (ID, Size)	(1, 3), (2, 3), (4, 2) (28, 3), and (33, 2)
Status	Optimal

Table IV: Model $\boldsymbol{LM}$ with (29) computational results

	$\boldsymbol{LM}$ with (29)
Model Class	MILP
Total Variables	64,561
Binary Variables	10,841
Constraints	220,642
Iterations $\times 10^6$	13.207
Time (Hours: Minutes: Seconds)	02:01:51
Best Objective $\times 10^6$	3.9309
Objective Bound $\times 10^6$	3.9309
Forced warehouses (ID, size)	(1, 3), (2, 3), (4, 2) (28, 3), and (33, 2)
Status	Optimal

We recommend solving such models in supply chain risk management with large data sets by first relaxing the y variables and then forcing the selected warehouses into model $\boldsymbol{LM}$ before resolving.

IV. Conclusion

Within this investigation, we have presented a cubic, binary optimization framework to strategically determine warehouse placements and branch allocations, incorporating the risk of warehouse failure. Acknowledging the intricacies of the initial model and the extensive array of variables and parameters typical in big data contexts, we introduced linearization and relaxation techniques to facilitate a streamlined and proficient solution approach.

Our approach was rigorously validated through a case study involving a real-world problem faced by a Canadian company. This validation showcased the practical applicability and robustness of our model and solution methodology in addressing the intricate challenges of warehouse location and branch assignment amid potential disruptions.

By incorporating big data considerations into our model, we leveraged extensive datasets to gain deeper insights into supply chain dynamics, demand patterns, and operational risks. This allowed us to create a more realistic and resilient framework for supply chain network design, enhancing the reliability and performance of the network.

The integration of big data not only improved the precision of our optimization model but also facilitated cost reduction and heightened customer satisfaction by enabling more informed and strategic decision-making. Our study underscores the critical role of big data in modern supply chain optimization, providing a pathway to more robust and adaptive supply chain networks.

REFERENCES

[1] T. Oshan and R. J. Caron, "Optimal warehouse location and size under risk of failure," *International Journal of Systems Science: Operations & Logistics*, vol. 10, no. 1, pp. 2208276, 2023, Taylor & Francis.

[2] W. Xu, Y. Hu, W. Luo, L. Wang, R. Wu, "A multi-objective scheduling method for distributed and flexible job shop based on hybrid genetic algorithm and tabu search considering operation outsourcing and carbon emission", Computers & Industrial Engineering, vol.157, 107318, 2021.

[3] X. Zhang, Z.H. Zhan, W. Fang, P. Qian, J. Zhang, "Multipopulation ant colony system with knowledge-based local searches for multiobjective supply chain configuration", IEEE Transactions on Evolutionary Computation, vol. 26(3), 512–526, 2021.

[4] Z. Liu, "Data-driven two-stage sparse distributionally robust risk optimization model for location allocation problems under uncertain environment", Aims Mathematics, vol8(2), 2910–2939, 2022

[5] J. Liu, Y. Wang, G. Sun, T. Pang, "Multisurrogate-assisted ant colony optimization for expensive optimization problems with continuous and categorical variables", IEEE Transactions on Cybernetics, vol. 52(11), 11348–11361, 2021.

[6] F. Glover and E. Woolsey,"Converting the 0-1 polynomial programming problem to a 0-1 linear program" Operations Research, vol. 22, no. 1, pp. 180-182, 1974.

[7] R. J. Caron and T. Oshan, "Optimal warehouse location and size in practice", International Journal of Operational Research, vol. 47, no. 4, pp. 547-557, 2023.

[8] Girish Ch. Dey and Mamata Jenamani, "Optimizing fortification plan of capacitated facilities with maximum distance limits," OPSEARCH, vol. 56, pp. 151-173, 2019.

[9] Q. Li, B. Zeng, and A. Savachkin, "Reliable facility location design under disruptions," *Computers & Operations Research*, vol. 40, no. 4, pp. 901-909, 2013.

V. APPENDIX I: NOTATION

Notation	Description
B, b	B is the set of branch indexed by b.
W, w	W is the set of warehouse location indexed by w.
S, s, q	S is the set of warehouse size indexed by s.
J, j, g	J is the set of product category indexed by j.
B	A subset of $W \times S$ listing warehouses of specific capacities built or to be built.
ℓ	The service level of a warehouse.
x_w^s	A binary variable for warehouse w of capacity s.
y_{wbj}^{ℓ}	A binary variable for the assignment of branch b to warehouse w at level ℓ to satisfy its demand from product j
K	A common volume unit to measure demand.
d_{bj}	The demand, in K, of product j from branch b.
f_w^s	The fabrication cost of warehouse w of size s.
τ_{wbj}	The cost to ship one K of product j from the warehouse w to branch b.
$C_F(x)$	The fabrication and land costs.
$C_O(x, y)$	The operation costs.
$C_T(y)$	The transportation costs.
p_w, p_*, p^*	The probability that warehouse w fails and the min and max failure probabilities.