B2B Omnichannel Network Design and Inventory Positioning

By: Geoffrey Allen and Shoichi Ishida
Advisor : Dr. Eva Ponce

Topic Areas: Omnichannel, Inventory, Network, B2B

Summary: Our sponsor, a large U.S. foodservice distribution company, aims to minimize the overall cost for the supply chain network, including omnichannel fulfillment options. We built an MINLP model to simultaneously optimize network design and inventory positioning. Our results demonstrate potential cost reductions of 3–9% in transportation, 2–8% in warehouse handling, up to 50% in inventory costs, and nine times greater efficiency than Gurobi.

Prior to MIT, Shoichi Ishida, CSCP, PMP, worked as a managerial consultant in operations and supply chain management at KPMG. He holds a Bachelor of Science in Industrial Engineering from Waseda University and a Postgraduate Certificate in Supply Chain Analytics from University of London.

Prior to MIT, Geoffrey Allen worked in supply chain strategy at Sysco and operations transformations at McKinsey & Co. Geoff holds a Master of Business Administration from the University of Virginia and a Bachelor of Economics and Finance from James Madison University.

KEY INSIGHTS

1. Simultaneously optimizing network design and inventory position yields greater cost reduction than traditional design, especially for capturing safety stock pooling-effect.

2. Omnichannel fulfillment options like cross-docking or direct shipment from suppliers further reduce costs, especially for high-volume products.

3. Tailored algorithms employing sophisticated techniques achieve higher efficiency than commercial solvers for inventory-integrated network design.

Introduction

The B2B foodservice distribution industry in the United States is projected to experience growth of 40% from 2022 to 2025. Although the trend towards omnichannel distribution has been present for B2C sales for over a decade, an increasing number of B2B companies are also adopting an omnichannel supply chain strategy. Recent research shows that B2B sales are now also permanently on a new trajectory towards omnichannel as well.

However, for companies used to a more traditional supply chain network design, implementing an omnichannel distribution network can be complex and potentially costly. Another complicating factor in network design, also related to fulfillment strategy, is inventory allocation. Changes in inventory positioning can significantly influence the ability of the model to leverage inventory pooling, ultimately affecting safety stock costs. Since this pooling effect generally exhibits non-linear behavior, many previous studies have either overlooked it or proposed separate models for network optimization and inventory positioning.

The sponsor for this project, a large B2B foodservice distribution company in the United States, is offering a comprehensive range of food-related products to diverse customers, such as restaurants and schools, through its nationwide supply chain network. It is beginning to offer omnichannel fulfillment to customers via pilots started over the past year.

Inventory positioning and fulfillment strategy at our sponsor company has been managed primarily in a decentralized approach, with the local distribution sites being the main determinant of which items to stock and which customers to serve from the different sites. However, this approach does not produce wholly optimal solutions. Taking this into account, this paper addresses questions including:

- What is the current state of the supply chain network in terms of suppliers, item flows, distribution centers, customer demand, and costs?
- How do customer service and lead time requirements influence the network design?
- What is the least-cost solution to satisfy customer demand?
- What would changing to the new optimized design with new omnichannel fulfillment options be worth in terms of cost of implementation?

Methodology

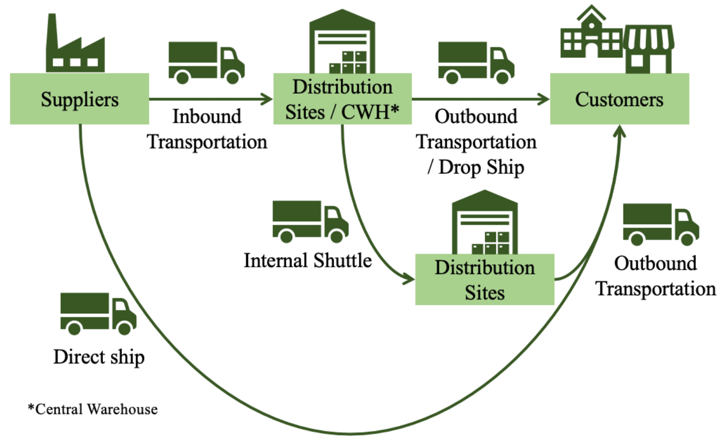
In order to find an optimal structure of supply chain network and inventory allocations with minimized cost, we formulated our problem as an optimization problem. To consider the non-linear safety stock cost, we developed a mixed-integer non-linear programming problem (MINLP) as our optimization model. We subsequently provide details on key assumptions, formulations for our model, and algorithms to solve it.

To determine which products to focus on, we analyzed and segmented our sponsor company's entire stock-keeping unit (SKU) count for the annual period using the ABC-XYZ approach. The product portfolio follows the common pattern of an "L" shape, with a small number of products responsible for a significant portion of demand with low variability, while there is also a long tail of low-volume products with high variability. We also added detail on item storage type, which could be either Dry, Cold (refrigerated), or Frozen. As a result, three products, SKU1, SKU2, and SKU3, were selected.

Our model encompasses all costs incurred throughout the end-to-end supply chain network, including inbound transportation costs (from suppliers to distribution sites), outbound transportation costs (from distribution sites to customers), product cost, warehouse handling cost, and inventory holding costs (cycle stock and safety stock holding costs).

Additionally, the model incorporates multiple omnichannel fulfillment options. The first option is direct ship, where suppliers deliver products directly to customers, bypassing distribution sites. The second option is cross-docking/peer replenishment. Under this option, products are routed to distribution sites via other sites, reducing transportation costs by utilizing more affordable internal shuttles for transportation between sites. However, this option incurs additional costs, such as cross-docking costs, at the second sites. The last option involves the use of central warehouses, which function similarly to regular distribution sites but generally have slightly lower warehouse handling and inventory holding costs. A conceptual image illustrating these fulfillment options is shown in Figure 1.

Figure 1. Possible Supply Chain Network Structure



It is also worth noting that we reformulated the model into a quadratic programming problem so that the problem is solvable with commercial solvers such as Gurobi. In addition, we developed a customized algorithm utilizing outer approximation to expedite the solve time. This acceleration is crucial, as the extensive amount of data and the NP-Hardness of the MINLP pose challenges that could result in excessively long computation times.

Results and Discussion

Our model effectively achieved cost reduction in both total logistics costs and total supply chain costs by optimizing supply chain flows and inventory allocations for each product. The magnitude of the reduction varies across different products, with SKU3 demonstrating the most significant improvement. Actual and enhanced inbound/outbound flows are illustrated in Figure 2.

Figure 2. As-Is vs. To-Be Network (SKU3)

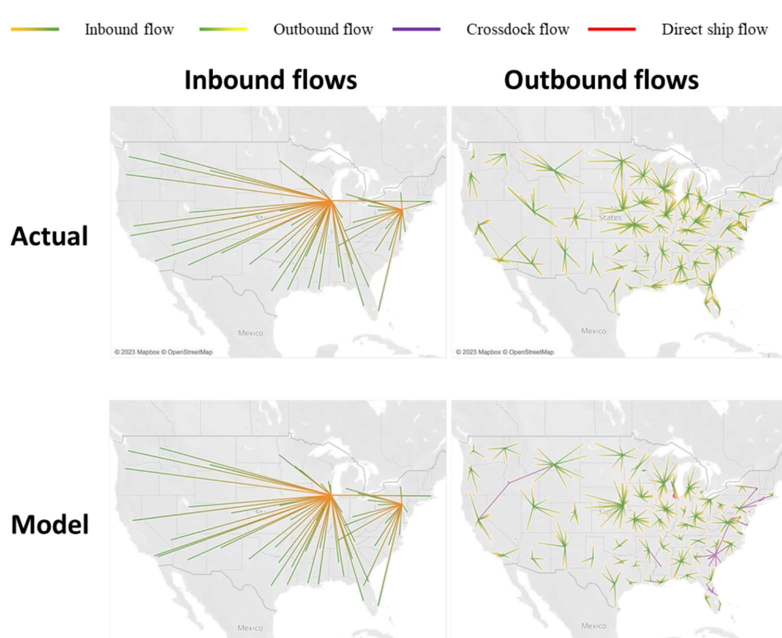


Table 1. Cost Summary (SKU3)

Cost Summary (Thousand USD)	AsIs ($)	Model ($)	Difference to AsIs ($)	Difference (%)
Transportation Cost	15,467	14,380	(1,088)	-7.0
Inbound Transportation Cost	5,871	5,635	(235)	-4.0
Cross-dock Cost	1	379	368	N/A
Direct Ship Cost	-	808	808	N/A
Outbound Transportation Cost	9,595	7,566	(2,029)	-21.1
Warehousing Cost	3,691	3,414	(276)	-7.5
Inventory Cost	234	118	(116)	-49.4
Cycle Stock Cost	61	66	5	8.6
Safety Stock Cost	173	52	(121)	-69.9
Subtotal Logistics Cost	19,392	17,912	(1,480)	-7.6
Product Cost	56,771	56,771	(0)	0.0
Total Supply Chain Cost	76,163	76,683	(1,480)	-1.9

For SKU3, our model successfully reduced the total logistics cost by 7.6% (excluding product cost) and the total supply chain cost by 1.9% (including product cost), primarily driven by significant reductions in outbound transportation cost and safety stock holding cost, as shown in Table 1. This particular item experiences a high concentration of customer demand near the supply point in Chicago, prompting the model to switch to a direct ship approach to minimize costs. The model also incorporates both cross-docking lanes and peer replenishment for this item. The notable improvements may be attributed to its comparatively lower product cost, resulting in a larger relative impact of transportation and inventory costs. Moreover, the inventory targets for this item are set 55% lower than the actual levels, ensuring the target service level is maintained. This implies that the company's current inventory allocations are inefficiently dispersed or not perfectly managed with consideration of supply and demand variability information, unlike our approach, which utilized the standard safety stock equation.

Regarding other products, SKU1 achieved a 7.5% reduction in total logistics cost and a 0.7% reduction in total supply chain cost, while SKU2 experienced reductions of 4.3% and 0.3%, respectively. The variance in cost reduction could be attributed to differences in product cost as well as variations in cost structures (such as the ratio of unit inbound transportation cost to outbound transportation cost, their individual costs, or unit inventory holding cost) and the variability of demand and supply lead times. However, this analysis provides valuable insights for identifying the next target SKU. It suggests that the sponsor company should focus on products that have similar cost structures and demand variability profiles as SKU3 to leverage the benefits of network design.

While the utilization of omnichannel options may vary across products, primarily due to the existing cost structure, it is always worthwhile to consider their implementation. Both SKU1 and SKU2 made partial use of omnichannel fulfillment options, which played a role in reducing the overall total supply chain cost.

Finally, by implementing our customized algorithm, we attained higher efficiency. We compared the average execution time of ten trials for Gurobi's default solve and our tailored outer approximation algorithm. This approach divides the original MINLP problem into MILP and NLP problems, solving them iteratively until convergence. Using the termination criterion (UB-LB)/LB = 0.0001, the average runtime was 5.62 secs, whereas the default solve required 44.6 secs. In essence, we significantly reduced the runtime to approximately 1/9 of the baseline, while maintaining solution quality. This improvement in runtime efficiency is particularly advantageous when scaling up our model to handle larger or more complex scenarios.

Conclusions

Our MINLP model effectively optimized network flows and inventory allocations simultaneously, uncovering opportunities for reduction in the total supply chain cost, particularly in transportation and safety stock holding costs. Furthermore, the implementation of omnichannel fulfillment options proved to be highly effective in achieving further cost savings. Moving forward, the company must verify the feasibility of these proposed modifications from an operational standpoint.

Our model is highly expandable to include multiple items, which would significantly amplify the model's potential impact on the company. In such situations, our tailored algorithm would drastically reduce solve times.