Multi-objective Network Optimization for Roche-Genentech

by

Diego Vesga
M.S. Industrial Engineering, Universidad de Los Andes, 2020

and

Tejveer Singh Oberoi
B.S. Supply Chain & Information Systems, The Pennsylvania State University, 2021

SUBMITTED TO THE PROGRAM IN SUPPLY CHAIN MANAGEMENT
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF APPLIED SCIENCE IN SUPPLY CHAIN MANAGEMENT
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 14, 2025

Signature of Author: ____
Department of Supply Chain Management
May 14, 2025

Signature of Author: ____
Department of Supply Chain Management
May 14, 2025

Certified by: ____
Dr. Jarrod Goentzel
Principal Research Scientist, Center for Transportation and Logistics
Capstone Advisor

Accepted by: ____
Prof. Yossi Sheffi
Director, Center for Transportation and Logistics
Elisha Gray II Professor of Engineering Systems
Professor, Civil and Environmental Engineering

Multi-objective Network Optimization for Roche-Genentech

by

Diego Vesga

and

Tejveer Singh Oberoi

Submitted to the Program in Supply Chain Management
on May 14, 2025 in Partial Fulfillment of the
Requirements for the Degree of Master of Applied Science in Supply Chain Management

Abstract

Roche-Genentech is a global biotech company that operates three distribution centers in the US. Inventory levels for each SKU are set on a country level based on historic demand and epidemiological forecasts. However, there was no formal guidance for determining the product mix and inventory levels for regions within a country. This study introduces two models to support inventory placement decisions. A Multi-Criteria Optimization (MCO) model is developed to advise on which customers to serve from each stocking point, and which SKUs to stock at each location by considering the cost and emissions associated with serving each customer from each distribution center (DC). Additionally, we develop an inbound allocation model that recommends SKU replenishment quantities to maintain target inventory levels at each site given the quantity of inbound imports.

The initial run of our models finds a 11% reduction in US network-wide transportation cost and emissions over the current state customer-DC and product mix allocations. We recommend a revision of company policies that constrain service of the highest value customers and highest revenue segment products exclusively from a single distribution node.

Capstone Advisor: Dr. Jarrod Goentzel, Ph.D.
Title: Principal Research Scientist, Center for Transportation and Logistics

Acknowledgements

Tejveer and Diego extend their sincere thanks to Stacy Malueg, Brett Gravois and Jason Sawicki from Roche-Genentech, for their generous sponsorship and support of this project—welcoming us into their fold, enthusiastically granting access to teams and organizational knowledge, and patiently fielding our many interview and data requests.

We are grateful to our advisor, Dr. Jarrod Goentzel, for his steady guidance, thoughtful feedback, and for helping us connect the dots. We also thank Dr. Milena Janjevic, whose teaching provided the analytical foundation for this project.

We thank our writing coach, Toby Gooley, for her countless recommendations on how to improve the clarity of this piece, and for her calming presence throughout the project.

Table of Contents

List of Figures

List of Tables

1 Scope Definition

1.1 Motivation

Effective inventory management is critical for ensuring the timely availability of potentially life-saving products for patients, while balancing the cost and environmental impact of serving them. The pharmaceutical industry faces additional complexities in effective supply chain management due to materials handling requirements such as temperature controls, high holding costs due to high product costs, and Food and Drug Administration (FDA) regulations around supply chain visibility and site certification.

Given the criticality of drug availability for patients in a complex operating environment, our sponsor company, Roche-Genentech, a leader in developing innovative therapies in areas such as oncology, immunology, and neuroscience, sought to determine the optimal location and level of its finished goods inventory among multiple distribution centers in the US.

Roche-Genentech's US distribution network consists of three distribution centers. The primary site (W1) receives finished product inventory from production facilities in Europe and US and serves approximately 70% of US demand by product value. W2 is co-located with a manufacturing site and handles about 15% of finished goods inventory by value. An additional third-party logistics (3PL) facility (W3) is leveraged to provide surge capacity and redundancy, serving as a buffer against value-at-risk concentration at a single node.

This project develops two network models to balance inventory across the three distribution centers. An MCO model to determine optimal customer and product mix allocation, and an inbound allocation model to determine the split of inbound inventory between the locations to support the recommendations generated by the MCO model. Emphasis was placed on considering the distance, transportation cost, and carbon footprint associated with serving each customer from each of the distribution centers.

1.2 Problem Statement

Roche-Genentech currently determines aggregate inventory levels for commercial finished goods in the United States based on production plans informed by epidemiological forecasts and historical demand patterns. However, a standardized policy for inventory staging across the domestic distribution network was lacking. This gap presented an opportunity to explore inventory regionalization to pursue cost savings, enhance service levels, and reduce the carbon footprint of the network. Additionally, Value-at-Risk (VaR) at each distribution node was recognized by management as an important proxy for network resilience; however, robust scenario analysis capabilities to support this assessment were absent. The primary objective of this study is to define and implement a systematic methodology to inform inventory positioning decisions across the U.S. distribution network. The research questions guiding this project are:

1. How should Roche-Genentech's US Commercial Distribution team formalize an inventory management framework that incorporates analytical models and tools to optimize the balance of finished goods inventory across distribution centers?
2. How can the optimization process measure and weigh the objectives of logistics cost, sustainability, and customer service?
3. What factors contribute to supply resilience, and how can those factors be quantified and incorporated into the inventory management framework?

1.3 Scope: Project Goals and Outcomes

This study introduces a network optimization approach to balance Roche-Genentech's 103 distinct commercial SKUs among three Distribution Centers (DCs) to serve ~2,300 customers. The key contribution of this study is a US Finished Goods Inventory Management Framework consisting of two analytical models:

A MCO model (annual cadence) optimizes customer-DC assignments, SKU mix at each DC, and target inventory levels for each SKU.

An Inbound allocation model (monthly cadence) allocates inbound finished goods shipments across DCs, considering inbound volume and current inventory levels, to maintain the target levels established by the MCO model.

1.4 Plan of Work

The study encompassed the following elements:

1. **Stakeholder Interviews and Data Gathering:** We met with several teams at Roche-Genentech, including Sales & Operations Planning, Transportation Logistics, and Supply Chain Sustainability, to understand the current state of demand and inventory planning, and to identify and request relevant data sources and fields for model parameters.
2. **Site Visits:** We visited two distribution centers, one company-owned, and the other contracted. On these visits, we documented operational considerations such as products per pallet, temperature zones, and capacities by temperature zone.
3. **Current State Network Model:** We mapped all distribution center and customer locations and visualized the current flow of finished goods across the U.S. network. This analysis provided a baseline understanding of the existing customer-to-DC assignments, shipment patterns, and geographic coverage, serving as a foundation for identifying optimization opportunities.
4. **Optimized Network Model:** We developed two models (detailed in Section 3.2) to enhance network performance. The MCO model, run annually, optimizes customer-to-DC assignments, SKU placement, and target inventory levels. The inbound allocation model, run monthly, allocates inbound shipments across DCs to maintain the targets set by the MCO model.
5. **Sensitivity Analysis:** With the models established, we subsequently adjusted constraints to enable what-if analysis capabilities and systematically reported the resulting insights in a Tableau dashboard.

2 State of the Practice

The central focus of our capstone is determining how our sponsor company can optimize its product placement and customer-DC allocation to minimize both costs and emissions while maintaining desired service levels. To address this challenge, we conducted a literature review, where we began by exploring supply chain management strategies in the pharmaceutical industry, followed by an analysis of multi-criteria optimization and inventory balancing approaches, along with a review of emissions estimation methods.

2.1 Pharmaceutical Supply Chains

It is essential to align a firm's supply chain strategy to the overall strategy of the business (Lapide, 2006). The pharmaceutical industry runs on a "blockbuster drug" business model, where much of the revenue comes from breakthrough drugs. The competitive advantage for these drugs tends to last 10-12 years, until patents expire, and lower-priced generics enter the market (Garnier, 2008). As such, pharmaceutical supply chains need to be rapidly scalable to meet the high demand for new innovative treatments and conversely be able to respond to lower demand for products when the competitive advantage is lost.

Pharmaceutical supply chains must consider several specific factors, including the distinction between high-volume commercial products and specialized clinical products, which have stricter security and handling requirements, as mandated by the Food and Drug Administration (FDA). Additionally, these supply chains face time-sensitive challenges related to product expirations, emergency demands, and varying temperature needs for different SKUs. These add another layer of complexity compared to standard supply chain management practices. Furthermore, external elements such as pandemics and shifts in public health needs increase demand variability, making it even more crucial to address forecasting and inventory management. A network model that captures the inherent complexities of pharmaceutical supply chains through constraints, combined with flexible input parameters, can enhance supply chain visibility and facilitate effective scenario planning.

2.2 Multi-Criteria Optimization

Supply Chain Network Design (SCND) problems typically involve mathematical programming to optimize a given objective within specific constraints and decision variables. When making supply chain network decisions, such as facility location, mode selection, or inventory levels, firms often consider multiple objectives, such as cost, service level, and emissions, resulting in a multi-criteria optimization (MCO) problem. MCO helps firms optimize these competing goals simultaneously to improve efficiency and performance across the supply chain.

Two primary methods for solving multi-criteria optimization problems are the weighted-sum method and the epsilon-constraint method. In the weighted-sum method, each objective is assigned a weight reflecting its importance, and all objectives are aggregated into a single objective function by multiplying each by its weight. The decision maker then selects the solution that maximizes the weighted sum. This method is particularly effective when there is a clear hierarchy of objectives, but the challenge lies in selecting appropriate weights, which can be subjective (Mesquita-Cunha et al., 2023). The epsilon method optimizes one primary objective while transforming other objectives into constraints, setting bounds for each and turning the multi-objective problem into a series of single-objective problems. This approach is useful when one objective is significantly more important than the others. It also allows flexibility in meeting secondary objectives (Mesquita-Cunha et al., 2023).

Despite its advantages, MCO faces challenges, including the complexity of modeling, particularly in large supply chains with many variables. High-quality, accurate data is essential for MCO's success, and incomplete or inaccurate data can lead to suboptimal solutions. Additionally, the subjective nature of weighting objectives in the weighted-sum method can cause inconsistencies in decision making, depending on the preferences of the stakeholders involved (Silver et al., 2016).

2.3 Inventory Balancing and Network Models

Effective inventory balancing ensures that stock levels at various locations align with demand patterns, minimizing the risk of both stockouts and excess inventory. This balancing act is vital for managing holding costs, reducing unnecessary transportation, and customer service levels. Additionally, inventory balancing helps prevent the need for costly inter-DC transfers, allowing for smoother product flows across the network.

Silver et al. (2016) emphasize that inventory balancing is a dynamic process that requires constant adjustment to match fluctuating demand. Supply chain models aim to develop heuristics for determining optimal inventory levels at various nodes within a network and deciding where to place stock. In particular, multi-echelon inventory models developed by Graves (1985) address different scenarios, such as fixed reorder intervals, one-for-one replenishment, and continuous review policies. These models advance our understanding of how inventory control policies can be optimized across complex supply chains by considering both inventory levels and the flow of goods between distribution centers.

Typically, companies use heuristic methods or common policies, such as Economic Order Quantity (EOQ), to determine optimal order quantities and inventory levels. EOQ is part of a broader set of inventory policies, which Silver et al (2016) categorize into different families based on various factors like demand patterns, lead-time variability, and the number of echelons in the supply chain. These families include deterministic models, which assume fixed demand and lead times, and stochastic models, which account for uncertainty in demand and supply. Silver et. al also mention how periodic review policies and continuous review policies are employed to manage stock levels, with the choice of policy depending on factors such as the cost of stockouts, the frequency of replenishment, and the criticality of maintaining service levels.

Similar to the scope of this capstone, Ambrosino & Grazia Scutellà (2005) explore the optimization of goods flows within established distribution networks, particularly in multi-stage supply chains. They use mathematical programming models to optimize resource allocation and the movement of goods across a pre-existing network of facilities, with the goal of minimizing total costs, including inventory holding, transportation, and handling costs. Their research highlights the importance of coordinating logistics to maintain responsiveness while controlling costs, which is directly applicable to Roche-Genentech's supply chain optimization efforts.

2.4 Sustainability Measurement

Well-to-Tank (WTT), Tank-to-Wheel (TTW), and Well-to-Wheels (WTW) are commonly used standards for reporting supply chain emissions. WTT accounts for emissions from fuel extraction, processing, and distribution. WTT emissions depend on the energy requirements and carbon intensity of fuel production and the logistics of fuel delivery. TTW captures emissions generated during vehicle operation. TTW emissions are primarily influenced by package weight, distance traveled, and the carbon intensity of the transportation mode. WTW encompasses both stages, providing a full life-cycle view (Gustafsson et al., 2021).

For the purposes of this study, we focus on TTW emissions, as they are more directly impacted by operational logistics decisions, specifically routing and mode choice, and can be more readily modeled. We derive an emissions equation based on weight, distance, and transportation mode intensity.

3 Methodology

After defining the research problem and reviewing existing methodologies, we concluded that the most suitable approach for optimizing inventory balancing across Roche-Genentech's U.S. distribution network is based on multi-criteria optimization (MCO) and multi-echelon inventory models. These models supported balancing inventory levels across the three distribution centers (W1, W2 and W3) and streamlining the flow of finished goods while addressing competing objectives such as cost, service levels, sustainability, and value at risk.

An evaluation of current inventory management practices provided a comprehensive understanding of the supply chain's existing state, including inventory allocation policies, customer assignment strategies, and associated risks. Based on these insights, we identified and weighed relevant optimization criteria, such as cost minimization and environmental impact reduction in collaboration with key Roche-Genentech stakeholders. These weighted criteria formed the foundation of a multi-criteria optimization model designed to optimize inventory distribution across the three facilities.

Finally, the results were shared with a data team at Roche-Genentech to integrate into a decision-support tool that enables Roche-Genentech's U.S. Commercial Distribution team to make informed, data-driven inventory decisions. This tool will provide real-time recommendations based on the optimization model, enhancing operational efficiency and responsiveness under varying scenarios.

3.1 Data Gathering

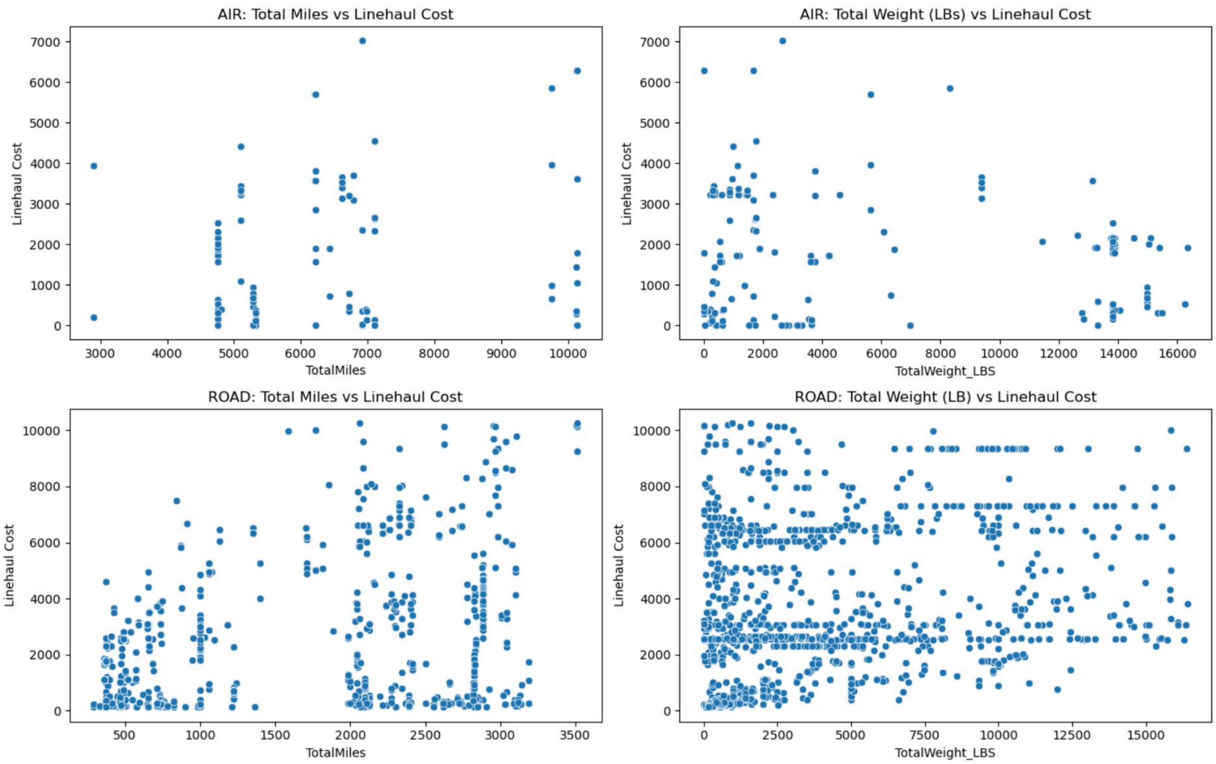
To support the development of our network optimization models, we compiled and processed several key datasets reflecting the current operational landscape. These inputs capture customer demand patterns, transportation costs, emissions intensity, and distribution center (DC) constraints. Each dataset was cleaned and transformed as needed to ensure consistency and relevance to the optimization problem. Below, we summarize the data sources and preparation steps for each input category.

Demand: We obtained historical shipment data detailing customer orders fulfilled from each warehouse over an eight-week period.

Customer Locations: We received a file containing customer addresses, which we aggregated by ZIP code. Using the Google Maps Geocoding API, we obtained the central latitude and longitude coordinates for each customer ZIP code to represent their geographic location. Out of 2300 customer locations in the dataset, 228 were active in the 8-week demand period that was provided. However, inactive customers were still included in the model and were allocated to warehouses based solely on proximity. This represents a limitation, as their actual demand patterns were not captured. This issue could be addressed by using a longer demand period that includes all customers with at least some activity, enabling a demand-driven allocation.

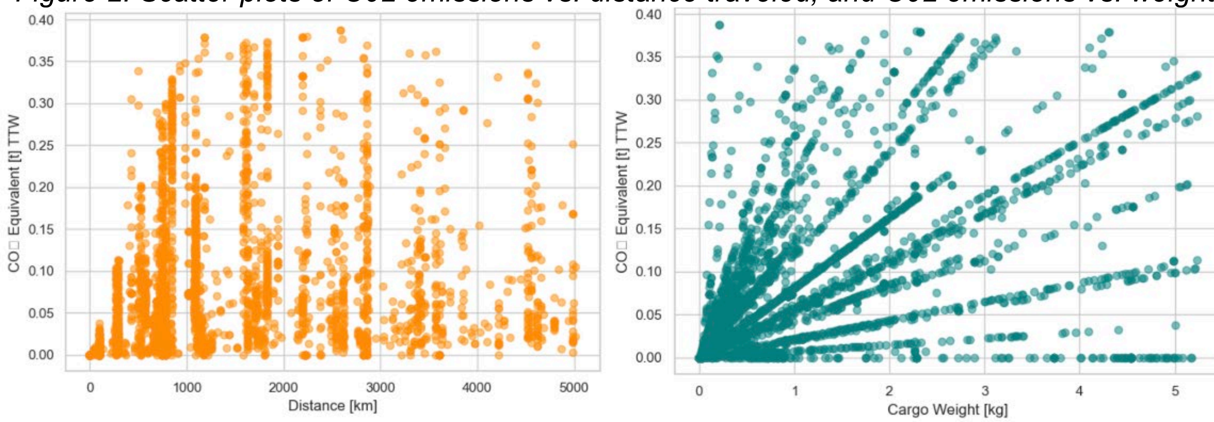
Transportation Cost: A dataset with 4,207 observations spanning from January 2, 2024, to January 8, 2025, was provided. The dataset includes key fields such as total miles traveled, mode of transport (TL, LTL, or Air), total shipment weight (in kilograms), total cost, and postal codes for both the source and destination. We cleaned the dataset by removing null values and removing outliers using the interquartile range (IQR) method. The dataset captures a full year of shipment activity, making it broadly representative across seasonal demand cycles. Figure 1 visually represents the cost and distance/weight relationships in the dataset received.

Figure 1: Scatter plots of line haul cost vs distance, weight by mode



Emissions Estimation: An emissions data extract was provided, containing estimated emissions for 6,545 historic global shipments from a single month—July 2024. The dataset includes key fields such as cargo weight (in tons), distance traveled (in kilometers), and tank-to-wheel (TTW) CO_2 emissions (in tons). These fields formed the basis of our regression analysis, through which we inferred an average emission factor in kilograms of CO_2. While the dataset reflects a one-month snapshot, this is considered appropriate for our analysis, as emissions behavior is not expected to vary significantly by season. The resultant emissions equation is presented in Section 4.1. Figure 2 visually represents the TTW C02 emissions and distance/weight relationships in the dataset received.

Figure 2: Scatter plots of C02 emissions vs. distance traveled; and C02 emissions vs. weight



Product and DC Parameters: We received product data and DC specifications, including masked cost of goods sold (COGS) for each product, storage requirements for each SKU, and pallet capacity of each distribution center (DC) across various temperature zones (e.g., refrigerated, ambient). The use of these parameters is discussed in greater detail in Section 3.2.2.

3.2 Application of Methodology

This study develops two models: A MCO model recommends the annual allocation of customers to distribution centers (DCs), along with the optimal product mix and inventory levels for each DC. An inbound allocation model generates monthly recommendations for the quantity of each product to be shipped between DCs (given upcoming international inbound shipments), ensuring that the recommended inventory levels are maintained.

3.2.1 Baseline Network Model

We first develop an as-is view model of the current state distribution network. We plot distribution centers, customer locations, and arcs based on historic order fulfilled-from decisions for each customer in the dataset. This model provides baseline transportation cost and emission values for the network, against which further optimization models were calibrated and compared.

3.2.2 Multi-Criteria Optimization (MCO) Model

The first model is a Multi-Criteria Optimization (MCO) approach that seeks to minimize the weighted sum of transportation costs and emissions. The weighting is assigned based on the relative importance of each factor and previous cost and emission estimates.

Table 1 presents the sets utilized in the initial model. For modeling purposes, each temperature zone within a DC was treated as a distinct site. For instance, W1_A represents the ambient storage area of the W1 facility, while W1_R denotes its refrigerated section. Each zone was assigned an independent storage capacity constraint, and products were restricted to a single temperature zone according to their specific storage requirements.

Table 1: Model Sets

Symbol	Description
W	Set of warehouses, indexed by w
C	Set of customers, indexed by c
K	Set of product SKUs, indexed by k

The parameters employed in the model are summarized in Table 2. Warehouse capacities, units per pallet, and customer demand were sourced from company-provided datasets, as described in Section 3.1. Distances between warehouses and customers were computed using the us_freeway_geograph library in Python and subsequently incorporated into the transportation cost and emissions formulations. Formulas 1 and 2 show how transportation mode and emissions are calculated for each possible arc, using the coefficients from Section 4.1. The Big M method is used to make sure that SKUs are only assigned to warehouses with matching temperature zones. If the temperature zone of an SKU does not match that of the warehouse, the model multiplies the cost and emissions by a large number (M), making the mismatched option unattractive to the model, thus eliminating that arc in the final solution.

$$c_{i,j,k} = c.intercept_{t.mode[j]} + (weight_k * cost.per.lb_{t.mode[j]})$$
$$+ (d_{i,j} * cost.per.mile_{t.mode[j]}) * \begin{cases} 1 & if\ zone_i = zone_k \\ M & if\ zone_i \neq zone_k \end{cases}, \quad \forall i \in W, \forall j \in C, \forall k \in K \qquad (1)$$

$$e_{i,j,k} = (weight_k * emissions.per.lb_{t.mode[j]})$$
$$+ (d_{i,j} * emissions.per.mile_{t.mode[j]}) * \begin{cases} 1 & if\ zone_i = zone_k \\ M & if\ zone_i \neq zone_k \end{cases}, \quad \forall i \in W, \forall j \in C, \forall k \in K \qquad (2)$$

Given that transportation costs and emissions are mode-dependent, each customer was assigned their most frequently utilized transportation mode based on historical delivery data. To address the model's multi-objective structure, an initial weight of 50% was assigned to both cost and emissions, as no preferred metric was defined at this stage. While further analysis of different weight combinations could provide valuable insights, it would require a more advanced model that treats transportation mode as a decision variable rather than a fixed parameter, an extension that falls outside the scope of this work. Finally, to accurately represent operational constraints, the model imposes service restrictions to a set of customers defined by the company that must be served exclusively from a specific warehouse.

Table 2: Model Parameters

Symbol	**Description**
s_i	Capacity of warehouse $i \in W$ (pallets)
u_k	Number of units per pallet for SKU $k \in K$
γ_{jk}	Demand of customer $j \in C$ for SKU $k \in K$ (units)
$distances_{ij}$	Distance between warehouse $i \in W$ and customer $j \in C$ (miles)
c_{ijk}	Unit cost on arc $(i,j) \in A$ for SKU $k \in K$ (\$/unit)
e_{ijk}	Unit emissions on arc $(i,j) \in A$ for SKU $k \in K$ (kg/unit)
t_j	The transportation mode assigned to customer $j \in C$
$weight_c$	Weigh placed on cost objective
$weight_e$	Weight placed on emissions objective
RC	List of customers restricted to only be served from W1
RW	Warehouses not allowed to serve restricted customers

The primary decision variable in the model is the quantity of each SKU to be served from each warehouse, to each customer, which implicitly determines customer allocation. However, to enforce warehouse capacity constraints, it is necessary to track pallet utilization. To this end, an auxiliary variable, referred to as pallets_used, was introduced. A detailed description of the decision variables, including pallets_used, is provided in Table 3.

Table 3: Model Variables

Symbol	Description
x_{ijk}	Flow of SKU $k \in K$ from warehouse $i \in W$ to customer $j \in C$ using transportation mode $t \in T$ in temperature zone $z \in Z$ (units)
$pallets\_used_{ik}$	Number of pallets used at warehouse $i \in W$ for SKU $k \in K$

We formally define the objective function and constraints of the optimization problem as follows:

Minimize

$$weight_c * \sum_{i \in W} \sum_{j \in C} \sum_{k \in K} c_{ijk} * x_{ijk} + weight_e * \sum_{i \in W} \sum_{j \in C} \sum_{k \in K} e_{ijk} * x_{ijk} \tag{3}$$

Subject to

$$\sum_{i \in W} x_{ijk} \geq \gamma_{jk}, \quad \forall j \in C, \forall k \in K \tag{4}$$

$$\text{pallets\_used}_{i,k} \geq \sum_{k \in C} \frac{x_{ijk}}{u_k}, \quad \forall i \in W, \forall k \in K \tag{5}$$

$$\sum_{k \in K} pallets\_used_k \leq s_i, \quad \forall i \in W \tag{6}$$

$$\sum_{k \in K} x_{ijk} = 0, \quad \forall i \in RC, \forall j \in RW \tag{7}$$

$$x_{ijk} \geq 0, \quad \forall \, i \in W, \forall j \in C, \forall k \in K \tag{8}$$

$$\sum_{j \in C} \sum_{k \in K} x_{ijk} * \ Cost_k \leq \ VaR.Limit_w, \quad \forall w \in W \tag{9}$$

The objective function (3) minimizes the total weighted sum of transportation costs and emissions associated with fulfilling customer demand from each distribution center (DC). Constraint (4) ensures that the demand for each SKU at every customer location is fully satisfied. Constraint (5) computes the number of pallets utilized based on the quantity of each product shipped from warehouses to customers, while Constraint (6) guarantees that the available capacity of each DC, within its respective temperature zone, is not exceeded. Constraint (7) restricts specific warehouses from serving designated customers, in accordance with operational requirements. Finally, Constraint (8) enforces the non-negativity condition on all decision variables.

As an additional consideration, the model could incorporate a value-at-risk constraint (9) to ensure that no warehouse holds inventory exceeding a predefined monetary threshold. This extension would provide further risk mitigation by limiting financial exposure at individual sites. The potential impact of introducing this constraint is examined in the scenario analysis (Section 4.2.1).

We implement this formulation in a Python Jupyter Notebook, using Gurobi Optimizer (version 12.0.1) as the optimization solver. The results of the optimal solution(s) are then exported and visualized in a Tableau workbook for further analysis and presentation.

3.3.3 Inbound Allocation Model

Table 4 presents the sets used in the inbound allocation model. *Inbound nodes* refer to the locations where new inventory is initially received, while *warehousing nodes* are the potential destinations that may store and distribute this inventory.

Table 4. Inbound Model Sets

Symbol	Description
IB	Set of inbound nodes, indexed by *b*
WH	Set of warehousing nodes, indexed by *h*

A key modeling assumption is the consolidation of W1 and W3 into a single virtual warehousing hub, due to their proximity and operational integration. As such, only two destination nodes are considered: WH1 and WH2. Inbound inventory is sourced from two locations: IB1 (representing shipments from Europe) and IB2 (the W2 manufacturing site). Table 5 summarizes these equivalences.

Table 5: Model Equivalences

Model Terminology	Equivalence
IB1	Inventory coming from Europe
IB2	W2 Manufacturing Site
WH1	W1 + W3
WH2	W2

To initiate the model, we calculate each warehouse's excess and shortage using the expressions in Formula 10 and Formula 11, respectively:

$$\text{Excess}_h = \text{Max}(\text{Inventory}_h - \text{Demand}_h\,, 0) \quad \forall h \in W \tag{10}$$

$$\text{Shortage}_h = \text{Max}(\text{Demand}_h - \text{Inventory}_h\,, 0) \quad \forall h \in W \tag{11}$$

Additionally, the model computes an allocation score to guide proportional distribution of remaining inventory, based on each warehouse's relative demand and capacity as shown in Formula 12. This score weights capacity and demand equally but could be adjusted as desired.

$$\text{Score}_{h1} = \frac{\%\text{Demand}_{h1} * \%\text{Capacity}_{h1}}{(\%\text{Demand}_{h1} * \%\text{Capacity}_{h1}) + (\%\text{Demand}_{h2} * \%\text{Capacity}_{h2})} \tag{12}$$

$$\text{Score}_{h2} = 1 - \text{Score}_{h1}$$

The allocation algorithm proceeds through four sequential phases, each applying a rule-based logic to assign inbound inventory to the appropriate warehouse. After each phase, inventory levels, shortages, and excesses are recalculated to inform subsequent decisions. This ensures a dynamic reallocation process that promotes local fulfillment and reduces redundant transfers.

The algorithm is illustrated in Figure 3, which provides a visual representation of the rule logic applied in each of the following phases:

Phase 1. Proximal Demand Fulfillment: Allocate inbound inventory to the closest warehouse where demand exists and cannot be met with current warehouse inventory, minimizing travel distance.

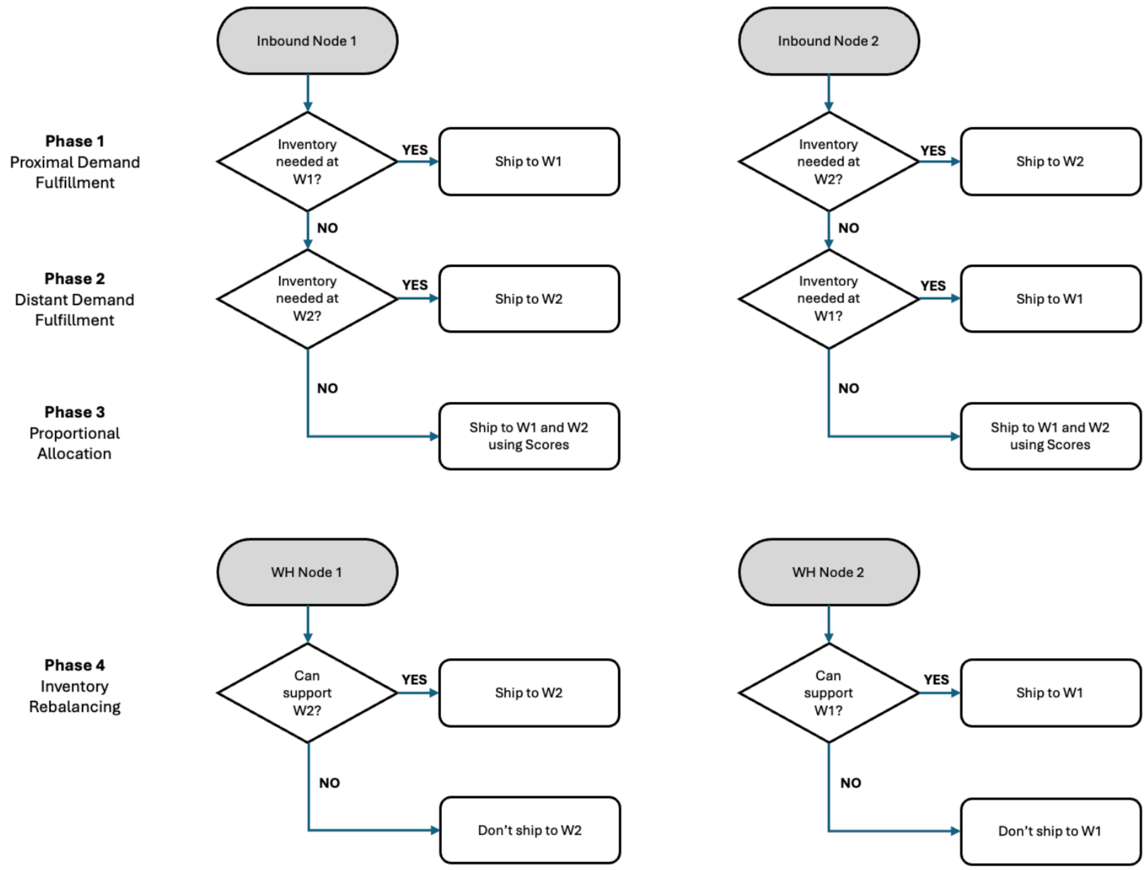
Phase 2. Distant Demand Fulfillment: Assign inventory from inbound nodes to the farther warehouse if demand remains unmet.

Phase 3. Proportional Allocation: Distribute any remaining inventory using the computed allocation scores to reflect relative demand and capacity.

Phase 4. Inventory Rebalancing: Reallocate inventory between warehouses if one facility holds excess stock and can support unmet demand at the other.

Upon completion of all four allocation phases, the model generates a comprehensive shipment plan and updated inventory levels for each warehouse. This ensures that all inbound inventory is efficiently assigned and stored, while minimizing unnecessary inter-warehouse transfers. Following this initial allocation, a secondary rebalancing step is performed between W1 and W3 to further disaggregate their combined inventory. This step enables the model to assign shipments independently to each warehouse based on updated requirements and capacity, thereby refining the network's operational balance.

Figure 3: Inbound Inventory Allocation Model



4 Results & Discussion

We begin by presenting the inferred model inputs related to transportation cost and CO_2 emissions. Next, we outline the results of the MCO model, followed by scenario analyses and findings from the inbound allocation model.

4.1 Emission & Cost Inferences

First, we estimate emissions and transportation costs using the datasets described in Data Gathering (Section 3.1) as inputs to the MCO model. Emissions were inferred using a linear regression model, with tank-to-wheel (TTW) CO_2 emissions (in lbs.) as the dependent variable. The resulting equations are:

$$\text{Mode}_{Air}\ \text{lb}\ CO_2 = 4.0776 * \text{Cargo}\ Weight_{lbs} + 0.8784 * Distance_{mi} \tag{13}$$

$$\text{Mode}_{Road}\ \text{lb}\ CO_2 = 0.0195 * \text{Cargo}\ Weight_{lbs} + 0.1353 * Distance_{mi} \tag{14}$$

Transportation cost was similarly estimated for each mode through a separate linear regression, with linehaul cost as the dependent variable. The resulting equations are:

$$\text{Mode}_{Air}\ \text{Cost} = 755.5454 + 0.2035 * Distance_{mi} \pm 0.0024 * \text{Cargo}\ Weight_{lbs} \tag{15}$$

$$\text{Mode}_{Road}\ \text{Cost} = 111.0641 + 0.0488 * Distance_{mi} + 0.1000 * \text{Cargo}\ Weight_{lbs} \tag{16}$$

The R^2 values (Appendix A) for the transportation cost inferences are modest, which is expected given the inherent non-linearity in the dataset. The regression for air transportation cost yielded a particularly low R^2; However, this is acceptable for modeling purposes, as the distance coefficient for air remains higher than that of road modes, aligning with known cost structures. The purpose of these regressions is not to achieve high predictive accuracy but to generate reasonable cost approximations that can inform the broader optimization framework.

4.2 Multi-Criteria Optimization (MCO) Model

The first model developed in this study, a Multi-Criteria Optimization (MCO) model, offers actionable insights into optimal inventory placement across Roche-Genentech's three U.S. distribution centers. The model determines an optimal customer allocation of 65% from W1, 18% from W2 and 17% from W3, and Out of a total of 103 SKUs, the recommended staging includes 88 at W1, 69 at W2, and 85 at W3, as detailed in Table 6. This new allocation results in approximately a 11% reduction in transportation variable costs compared to the baseline model (Table 7), as it prioritizes serving customers from the closest warehouse, thereby minimizing the distance traveled.

Consistently, transportation costs and emissions follow a similar distribution: approximately 50% of the total cost and environmental impact is attributed to W1, while W2 and W3 each account for about 25%.

Table 6: MCO model results

	COGS Served (MM)	Customers Served	Materials
W1	$22,816.09	1777	88
W2	$3,466.32	517	69
W3	$4,526.51	454	85
	$30,808.92	2748	

Table 7: MCO model vs. baseline model results

	Transportation Cost (MM)	Emissions
Baseline	$52.82	173.11
Optimal (MCO)	$46.90	154.52
Improvement	11.2%	10.7%

The optimization results illustrate a redistribution of customer demand across the network. As shown in

Figure 4, W1 continues to serve as the central distribution hub, fulfilling the majority of customer orders, while W2 and W3 serve as complementary nodes that absorb regional demand and help balance capacity constraints. This allocation strategy leverages geographic proximity while adhering to operational constraints, such as temperature zones and customer-specific warehouse assignments as defined by the business.

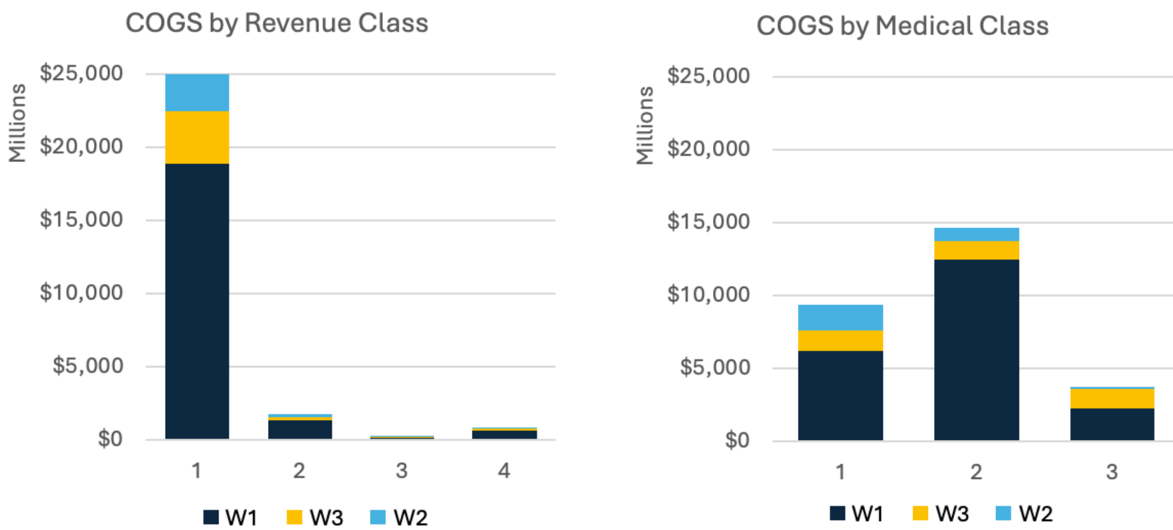
Figure 4: Network Visualization

Legend: Blue - Served from W1, Gray – Served from W2, Yellow – Served from W3

The solution incorporates business rules that restrict service of highest volume customers from W1. As reflected in the revenue-based segmentation (Figure 5), 74% of total demand value is allocated to W1, while W2 and W3 serve 11% and 15% respectively. Products in Revenue Class 1, which represent the majority of revenue, are consistently staged at W1, in accordance with its established operational capacity and service agreements. This is illustrated in Figure 5**,** which shows Cost of Goods Sold (COGS) breakdowns across revenue class and medical class.

Interestingly, the SKU allocation by Medical Class is more evenly distributed under the optimized model. This redistribution reduces the dependency on any single warehouse and supports operational resilience in the event of localized disruptions.

Figure 5: COGS by Revenue and Medical Class



Finally, Appendix B presents the model's recommended demand allocation for each SKU. Notably, no SKU is exclusively staged at a single warehouse. This diversification improves flexibility and mitigates the risk of stockouts in the event of a supply chain disruption. While the scenario analysis (Section 4.2.1) focuses on warehouse-level resilience, the results also raise the opportunity to assess value-at-risk at the SKU level.

4.2.1 Scenario Analysis

To assess the model's responsiveness to changes in key constraints, we conducted a series of scenario analyses. These explored how adjustments to operational assumptions impact customer allocation, warehouse utilization, and overall network feasibility. Results are summarized in Figure 6.

Scenario 1. Relaxing Highest-Value Customer Assignment Rules

In the MCO model, the top customers, defined by order value, are restricted to being served exclusively from W1. In this scenario, we allow the top four customers to be served from any warehouse. The model responds by reallocating one of the high-value customers from W1 to W3, maintaining geographic proximity while improving warehouse capacity balance.

This suggests that the original constraint, while operationally grounded, may limit flexibility. By relaxing the assignment rule for key customers, the network achieves a more efficient allocation of demand across facilities, reducing the load on W1 and enhancing resilience.

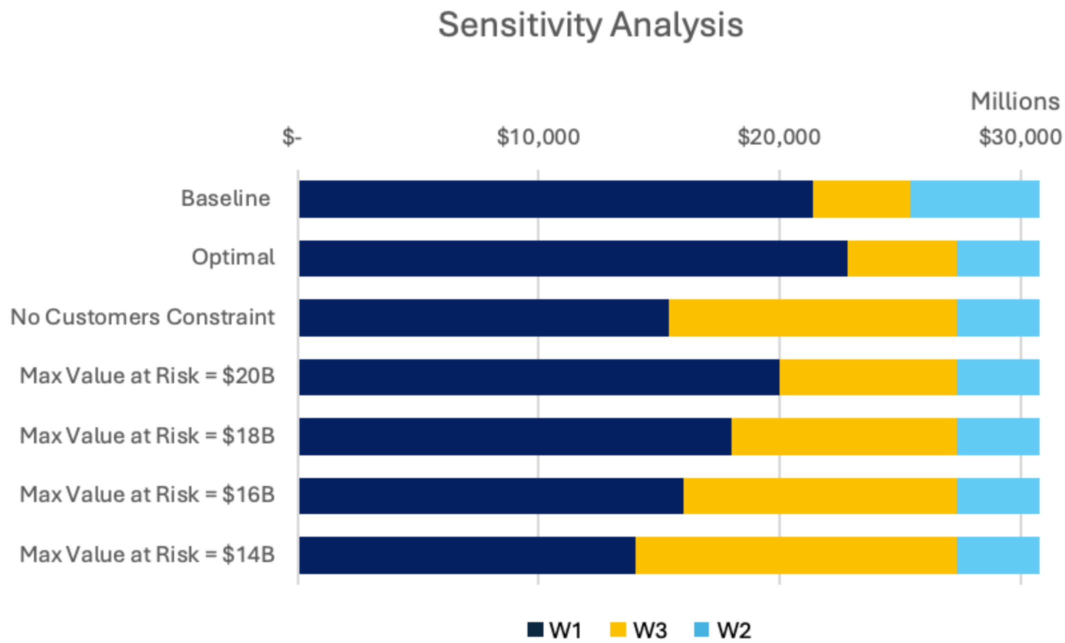
Scenario 2. Value-at-Risk Sensitivity

In this scenario, we analyze the sensitivity of the model to a value-at-risk (VaR) constraint, which limits the total value of demand that can be served from a single warehouse. The constraint is applied progressively, ranging from $20 billion to $14 billion. As shown in Figure 6, decreasing the VaR threshold leads to a redistribution of customers from W1 to W3, mitigating concentration risk.

However, when the value-at-risk (VaR) constraint is set too aggressively (e.g., at $10 billion), the model becomes infeasible. This illustrates a key real-world trade-off: overly strict financial exposure limits on inventory can result in insufficient capacity to meet total network demand. For instance, in our model, the total 8-week demand across all customers is approximately $30.8 billion. If the allowable inventory value across all warehouses is capped at $10 billion, the model cannot fulfill the remaining $0.8 billion in demand beyond what inventory levels can support.

It is important to note that setting a tighter VaR threshold not only reduces financial risk but also limits operational flexibility. With lower inventory levels, the system becomes less resilient to disruptions, and inventory policies must be adjusted accordingly, potentially compromising service levels or responsiveness. The same logic applies when constraints are applied at the SKU level, while this can help prevent overstocking at a specific location, it may also reduce the ability to respond to local fluctuations in demand, leading to reduced flexibility in the distribution network.

Figure 6: Scenario Analysis – COGS Allocation

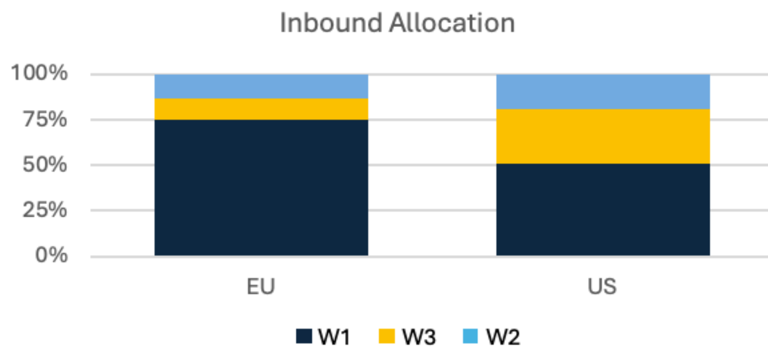


Other scenarios were analyzed, including variations in the weights assigned to transportation cost and emissions in the objective function. As expected, these changes did not affect customer allocation, as they only alter the objective value and not the optimal solution. Even in extreme cases (e.g., assigning 99% weight to transportation cost and 1% to emissions), the model still prioritizes minimizing distance. Adjusting these weights would only have a meaningful impact if transportation mode were included as a decision variable.

4.3 Inbound Allocation Model

The primary objective of the second model is to determine how to distribute inbound inventory monthly, considering current inventory levels at each warehouse, customer allocations, and warehouse capacity constraints. As a result of the rule-based allocation logic, the model achieves a well-balanced distribution of inbound quantities. Specifically, approximately 75% of inventory arriving from Europe is assigned to W1, its geographically closest warehouse, while the remaining 25% is equally split between W3 and W2. Figure 7 illustrates that inventory from the production facility co-located with W2 is largely retained at W2 to avoid unnecessary shipments, with only surplus quantities being redistributed to other facilities. Figure 7 shows the distribution of inbound inventory among the three sites recommended by the inbound allocation model.

Figure 7: Inbound Allocation



Consistent with the results of the MCO model, most of the inventory value remains concentrated at W1, which serves the largest share of total demand and has the highest available capacity. As shown in Table 8 and Figure 8, approximately 67% of inventory by value is allocated to W1. This trend is further reflected in Figure 9, which breaks down inventory cost by revenue and medical class, showing a similar concentration aligned with demand distribution.

Table 8: Inbound allocation model results

	W1	W2	W3
Current Inventory Cost (MM)	$ 19,682	$ 2,737	$ 5,368
Incoming Inventory Cost (MM)	$ 13,656	$ 1,291	$ 7,307

Figure 8: Inbound allocation model results

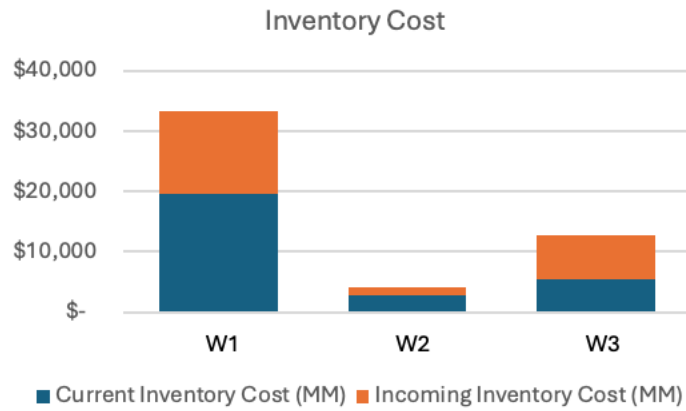
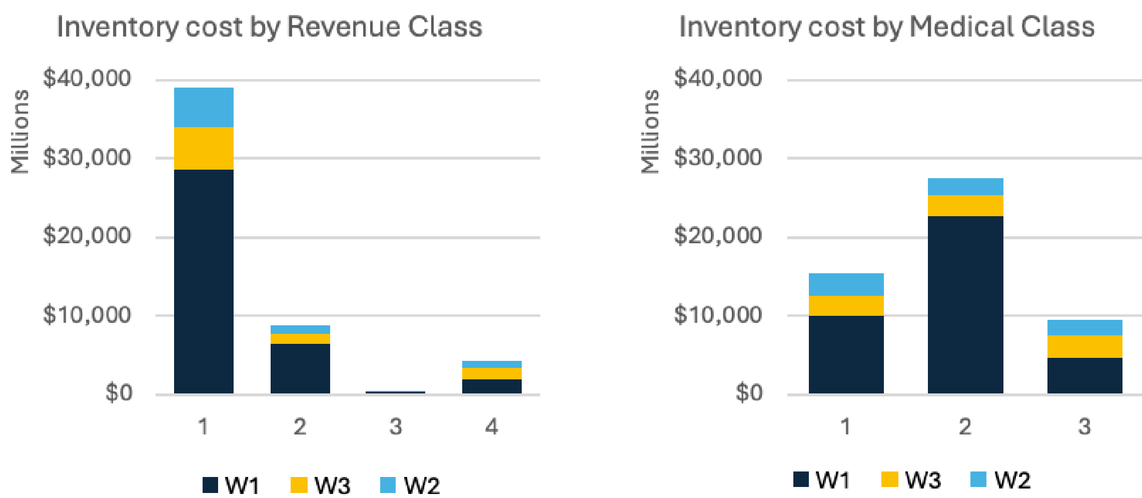
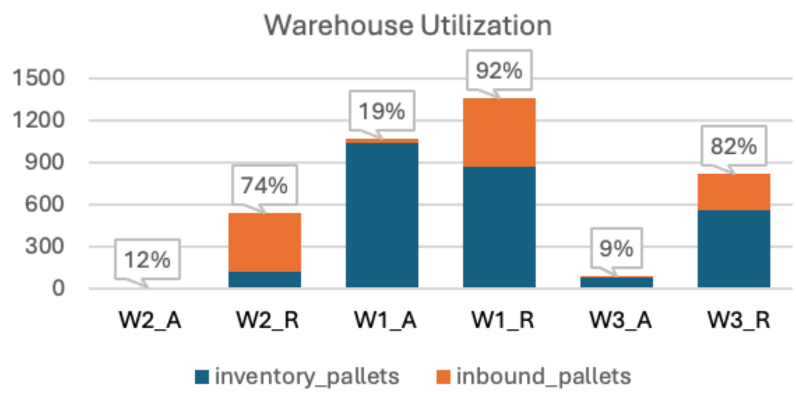


Figure 9: Inventory Cost by Revenue and Medical Class



Finally, warehouse utilization analysis highlights a consistent underutilization of ambient storage zones (A). Although overall inventory levels across the network are sufficient to meet demand, the total volume stored remains relatively low. In contrast, refrigerated zones (R) exhibit significantly higher utilization rates, particularly at W1, where space usage approaches 90%, as depicted in Figure 10 . This elevated usage is largely driven by inbound pallets, suggesting that high-turnover products constitute a significant share of recent inbound shipments.

Figure 10: Warehouse Utilization

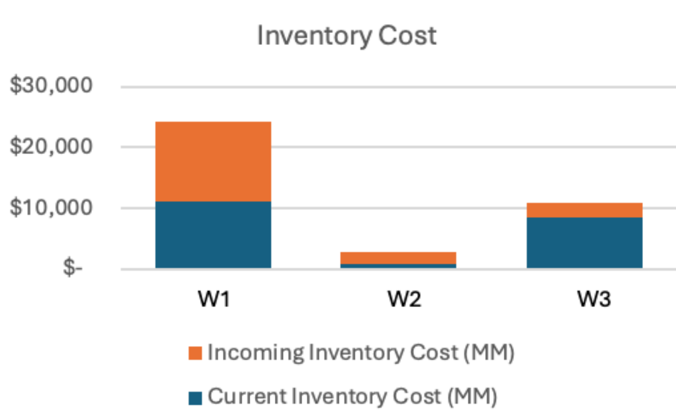


To evaluate the model's behavior over time, we ran a second iteration to simulate two consecutive months. The second iteration builds on the inventory results from the first, while demand remains constant and incoming shipments correspond to the following month. As expected, the results of this second iteration demonstrate the consistency of the inbound allocation model (Section 4.3), assigning incoming inventory in similar proportions as in the first run, as shown in Table 8 and Figure 11.

Table 8: Inbound allocation model results - second run

	W1	W2	W3
Current Inventory Cost (MM)	$ 11,049	$ 824	$ 8,520
Incoming Inventory Cost (MM)	$ 13,136	$ 2,032	$ 2,368

Figure 11: Inbound allocation model results - Second Run



4.4 Limitations

Historic on-time and in-full (OTIF) performance is not directly incorporated into the model, as the focus of this study is not on operational factors such as mode or carrier selection. Rather, the assumption is that improvements in network design, specifically in customer allocation and inventory placement, will indirectly enhance OTIF performance by reducing lead times and improving product availability closer to customers.

Customer-DC allocations were made for the 286 customer locations that registered demand during the 8-week orders extract provided. The MCO model should be re-rerun on a larger demand dataset capturing a longer historical period and broader customer base to make DC allocations for the remaining customer sites.

The proximity of W1 and W3 introduces the possibility that the MCO model's customer allocation decisions are influenced by marginal differences in distance. To address this, additional cost components (e.g., handling fees, service level differentiation) or minimum distance thresholds could be incorporated in future work to better reflect operational realities and prevent arbitrary splits between near-identical sites.

Additionally, Demand is assumed to be deterministic, and as such the models recommend a target cycle stock level for each site which does not consider safety stock.

5 Conclusion

This study addressed the absence of formal guidance for determining regional product mix and inventory levels within Roche-Genentech's U.S. distribution network. While national inventory targets for each SKU were set based on historic demand and epidemiological forecasts, allocation across the three U.S. distribution sites lacked a structured approach.

To support inventory placement decisions for approximately 103 SKUs, we developed two models. The first, a Multi-Criteria Optimization (MCO) model, identifies optimal customer-to-DC assignments and appropriate SKU mix by jointly minimizing transportation costs and emissions associated with serving each customer. An inbound allocation model is developed to recommend replenishment quantities required to maintain target inventory levels given varying inbound shipment volumes.

Initial results show that applying these models leads to a 11% reduction in transportation costs and emissions across the U.S. network relative to the current state. Based on these findings, we recommend revisiting policies that restrict service of high-value customers and high-revenue SKUs to a single distribution center, as more flexible allocation strategies may yield both operational and environmental benefits.

5.1 Management Recommendations

1. **Continue to leverage modeling to inform strategic trade-offs:** This study highlights the value of network modeling in quantifying trade-offs between cost, emissions, and operational flexibility. By simulating alternative inventory placement and customer allocation scenarios, the models enable data-driven decision-making that goes beyond heuristics or historical precedent. The ability to evaluate policy impacts under various demand and capacity conditions provides a structured framework for ongoing network design decisions and long-term strategic planning.

2. **Reevaluate customer-DC assignments**: Reallocating customer assignments across distribution centers based on a joint cost and emissions minimization objective presents an opportunity to reduce transportation costs and carbon footprint by 11%, without requiring additional capital investment. Re-running the MCO model using a historic demand dataset with a longer time horizon than 8 weeks would enable more representative customer-DC allocations by capturing demand from a broader set of customer sites.

3. **Diversify placement of revenue class 1 SKUs:** The optimized model distributes SKUs more evenly across classes and DCs. No SKU is staged at only one DC in the optimal scenario; our model recommends a 74:11:15 percentage split for revenue class 1 SKUs between W1, W2 and W3 respectively. This diversified staging strategy enhances operational flexibility and buffers against the risk of stockouts in the event of localized disruptions.

4. **Re-evaluate policy to serve top-value customers exclusively from W1:** Shifting one of the top 3 customers by order value from W1 to W3 improves inventory balance

across all three sites without affecting transportation cost significantly. This reallocation would help alleviate capacity pressure on W1's highly utilized refrigerated zones and supports more efficient use of underutilized ambient storage at other sites.

5.2 Future Work

For this project, we restricted our scope to an optimized baseline network design (i.e., finding the optimal flow and inventory levels within the existing network). Future work may expand the scope to consider facility location decisions.

Additionally, demand forecasting was excluded from this study. Instead, the distribution network was modeled using an 8-week demand snapshot. Future research could incorporate predicted demand as a dynamic input to enhance model accuracy.

Further developments could also evolve this strategic model into an operational tool to support day-to-day decision-making, such as selecting transportation modes to improve OTIF and overall service levels. In such cases, the relative weighting of cost and emissions in the objective function becomes increasingly important and may need to be dynamically adjusted based on operational priorities.

References

Ambrosino, D., & Grazia Scutellà, M. (2005). Distribution network design: New problems and related models. *European Journal of Operational Research*, *165*(3), 610–624. https://doi.org/10.1016/j.ejor.2003.04.009

Garnier, J.-P. (2008, May). Rebuilding the R&D Engine in Big Pharma. *Harvard Business Review*. https://hbr.org/2008/05/rebuilding-the-rd-engine-in-big-pharma

Graves, S. C. (1985). A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment. *Management Science*, *31*(10), 1247–1256. https://doi.org/10.1287/mnsc.31.10.1247

Gustafsson, M., Svensson, N., Eklund, M., & Fredriksson Möller, B. (2021). Well-to-wheel climate performance of gas and electric vehicles in Europe. *Transportation Research Part D: Transport and Environment*, *97*, 102911. https://doi.org/10.1016/j.trd.2021.102911

Lapide, L. (2006, April). The Essence of Excellence: MIT's SC2020 Project. *Supply Chain Management Review*, *10*(3), 18–24.

Mesquita-Cunha, M., Figueira, J. R., & Barbosa-Póvoa, A. P. (2023). New ϵ−constraint methods for multi-objective integer linear programming: A Pareto front representation approach. *European Journal of Operational Research*, *306*(1), 286–307. https://doi.org/10.1016/j.ejor.2022.07.044

Silver, E. A., Pyke, D. F., & Thomas, D. J. (2016). *Inventory and Production Management in Supply Chains* (0 ed.). CRC Press. https://doi.org/10.1201/9781315374406

Appendix

Appendix A OLS Regression Outputs

Emissions (air mode)

```
[AIR MODE]
Sample size: 891
                            OLS Regression Results
=========================================================================================
Dep. Variable:     co2_equivalent_[lbs]_ttw   R-squared (uncentered):                 0.966
Model:                                  OLS   Adj. R-squared (uncentered):            0.966
Method:                       Least Squares   F-statistic:                        1.252e+04
Date:                      Mon, 12 May 2025   Prob (F-statistic):                      0.00
Time:                              22:55:08   Log-Likelihood:                       -8640.7
No. Observations:                       891   AIC:                                1.729e+04
Df Residuals:                           889   BIC:                                1.729e+04
Df Model:                                 2
Covariance Type:                  nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
cargo_weight_[lbs]     4.0776      0.051     79.561      0.000       3.977       4.178
distances_[mi]         0.8784      0.040     22.131      0.000       0.801       0.956
==============================================================================
Omnibus:                       17.596   Durbin-Watson:                   1.574
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               33.400
Skew:                          -0.006   Prob(JB):                     5.59e-08
Kurtosis:                       3.948   Cond. No.                         3.01
==============================================================================
```

Emissions (road mode)

```
[ROAD MODE]
Sample size: 3818
                            OLS Regression Results
=========================================================================================
Dep. Variable:     co2_equivalent_[lbs]_ttw   R-squared (uncentered):                 0.630
Model:                                  OLS   Adj. R-squared (uncentered):            0.630
Method:                       Least Squares   F-statistic:                            3250.
Date:                      Mon, 12 May 2025   Prob (F-statistic):                      0.00
Time:                              22:55:27   Log-Likelihood:                       -24396.
No. Observations:                      3818   AIC:                                4.880e+04
Df Residuals:                          3816   BIC:                                4.881e+04
Df Model:                                 2
Covariance Type:                  nonrobust
======================================================================================
                         coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------------
cargo_weight_[lbs]     0.0195      0.000     48.411      0.000       0.019       0.020
distances_[mi]         0.1353      0.003     49.885      0.000       0.130       0.141
==============================================================================
Omnibus:                      512.702   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              813.875
Skew:                           0.931   Prob(JB):                    1.86e-177
Kurtosis:                       4.284   Cond. No.                         6.96
==============================================================================
```

Transportation cost (road mode)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     LinehaulBlendedCost   R-squared:                       0.400
Model:                             OLS   Adj. R-squared:                  0.397
Method:                  Least Squares   F-statistic:                     132.9
Date:                 Wed, 14 May 2025   Prob (F-statistic):           5.80e-45
Time:                         13:00:01   Log-Likelihood:                -2387.1
No. Observations:                  402   AIC:                             4780.
Df Residuals:                      399   BIC:                             4792.
Df Model:                            2
Covariance Type:             nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const            111.0641      9.832     11.296      0.000      91.736     130.393
TotalMiles         0.0488      0.005     10.474      0.000       0.040       0.058
TotalWeight_LBS    0.1000      0.007     15.367      0.000       0.087       0.113
==============================================================================
Omnibus:                       46.575   Durbin-Watson:                   1.954
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               64.800
Skew:                           0.803   Prob(JB):                     8.49e-15
Kurtosis:                       4.134   Cond. No.                     3.70e+03
```

Transportation cost (air mode)

```
                            OLS Regression Results
==============================================================================
Dep. Variable:     LinehaulBlendedCost   R-squared:                       0.058
Model:                             OLS   Adj. R-squared:                  0.051
Method:                  Least Squares   F-statistic:                     8.220
Date:                 Wed, 14 May 2025   Prob (F-statistic):           0.000344
Time:                         13:02:25   Log-Likelihood:                -2346.6
No. Observations:                  268   AIC:                             4699.
Df Residuals:                      265   BIC:                             4710.
Df Model:                            2
Covariance Type:             nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
const            755.5454    453.286      1.667      0.097    -136.954    1648.045
TotalMiles         0.2035      0.060      3.374      0.001       0.085       0.322
TotalWeight_LBS   -0.0024      0.018     -0.131      0.896      -0.038       0.034
==============================================================================
Omnibus:                        4.840   Durbin-Watson:                   1.827
Prob(Omnibus):                  0.089   Jarque-Bera (JB):                4.936
Skew:                           0.314   Prob(JB):                       0.0848
Kurtosis:                       2.779   Cond. No.                     4.81e+04
```

Appendix B **COGS Allocation by Material**

	W1	W3	W2	COGS (MM)
P1	57%	38%	5%	$3,451
P2	68%	32%	0%	$21
P3	76%	13%	11%	$1,743
P4	91%	2%	7%	$38
P5	67%	31%	1%	$111
P6	58%	42%	0%	$3
P7	90%	0%	10%	$6
P8	79%	8%	13%	$0
P9	83%	11%	7%	$259
P10	79%	15%	6%	$8
P11	30%	66%	5%	$2
P12	19%	64%	17%	$10
P13	94%	0%	6%	$62
P14	66%	15%	19%	$9,536
P15	94%	1%	4%	$50
P16	94%	0%	6%	$63
P17	73%	26%	1%	$256
P18	88%	2%	9%	$4
P19	80%	15%	5%	$11
P20	87%	8%	6%	$13,886
P21	91%	0%	8%	$23
P22	96%	0%	4%	$143
P23	96%	0%	4%	$41
P24	58%	42%	0%	$57
P25	83%	9%	7%	$202
P26	81%	6%	12%	$11
P27	70%	30%	0%	$3
P28	75%	25%	0%	$0
P29	95%	0%	4%	$103
P30	58%	42%	0%	$134
P31	76%	24%	0%	$38
P32	46%	48%	7%	$178
P33	44%	56%	0%	$8
P34	74%	17%	9%	$341
P35	74%	11%	15%	$6
	76%	14%	10%	$30,809