Delay Prediction in Supply Chains: A Hybrid Graph-Based and Machine Learning Approach

Mattia Volpato – Matricola 866316

Relatore: Prof. Michele Ciavotta **Correlatore**: Ing. Bruno Puzzolante

September 2025

1 Introduction

The manufacturing industry faces growing exposure to disruptions driven by global pandemics, geopolitical tensions, and climate change. In such an uncertain environment, building resilience has become a strategic priority.

The **M4ESTRO** project [8], funded by Horizon Europe, aims to develop a Manufacturing as a Service (MaaS) platform to support predictive and adaptive responses to disruptions. Its Supply Resilience App, developed by a collaboration of partner companies [7], integrates diverse data sources to monitor supply chain health and recommend corrective actions. Unlike traditional optimization methods that rely only on historical data, M4ESTRO adopts a proactive framework based on **disruption indicators** to detect anomalies, assess their impact, and enable timely mitigation. Key challenges include selecting relevant metrics, integrating heterogeneous data, and ensuring robust detection.

This thesis addresses one of the project's tasks, focused on developing **Logistic Chain Disruption Indicators (LCDI)** for the detection of internal disruptions. The work focuses specifically on designing and implementing a subset of LCDI aimed at estimating the time from order placement to delivery from supplier to manufacturer, incorporating supplier holidays together with traffic and weather conditions affecting shipment.

We start by formalizing a *graph-based* representation of the supply chain and proceed to define two types of indicators: *historical*, capturing operational times in the order fulfillment process, and *realtime*, dynamically describing shipment status by leveraging the graph structure, historical indicators and external sources. Limited integration with pilot partners restricted the availability of operational data; consequently, we focused on independently defining a formal design and a preliminary implementation of the indicators, which can further be refined as more data become available.

2 Training Data

The dataset, provided by FAE Technology [9], comprised some thousands orders, with the main attributes listed in Table 1. Only around one hundred orders included valid tracking links, which were standardized via an external service to reconstruct complete sequences of shipping events from suppliers to the manufacturer across the various carriers used. The dataset was further enriched with additional sources, supplementing shipment records with traffic and weather data and integrating national holidays into the dispatch processes.

Column	**Type**	**Description**
id	Integer	Order identifier
distributorId	Integer	Supplier identifier
createdAt	Timestamp	Creation timestamp
confirmedDeliveryDate	Timestamp	Delivery timestamp
trackingLink	String	Shipment tracking URL

Table 1: Schema of FAE Technology database.

3 Logistic Chain Modeling

Shipments are modeled as paths in a directed acyclic *supply chain graph*, with supplier sites as sources, carrier facilities as hubs, and the manufacturer as a sink (Figure 1). An arc flow function, satisfying the balance condition [1], is introduced to count shipments, ensuring that all shipments originating from suppliers reach the manufacturer. Routing is modeled probabilistically with discrete-time Markov chains [3], where transitions reflect the **empirical probabilities** of historical routing decisions and path probabilities are computed as products of transitions along maximal paths. Carrier-specific behaviors are incorporated by conditioning flows and transitions on the carrier, allowing the estimation of distinct routing distributions.

The graph is built incrementally from tracking data: new vertices and arcs are added as shipments are recorded, flows are updated, and transition probabilities are recalculated. Although only a single cycle is observed in the dataset, a cycle detection and resolution procedure is applied to ensure the graph remains acyclic, preserving its structural integrity.

All feasible paths from a given vertex to the manufacturer are extracted using a modified depth-first search, yielding a valid probability distribution that serves as input for the subsequent indicators. Although the number of potential combinations is theoretically exponential, the variability of paths is limited in practice by the observed shipment data.

The graph thus provides a probabilistic representation of the logistic network that underlies the predictive indicators developed in this work.

4 Historical Indicators

We characterize operational times in the order fulfillment process using historical data, distinguishing two main components: the interval from order acceptance to dispatch (*dispatch time*) and the interval from dispatch to arrival at the destination (*shipment time*). The overall duration, encompassing both components, is referred to as the *delivery time*.

Dispatch times are modeled at the level of individual supplier sites, while shipment times are modeled for each combination of supplier site and carrier. Each component is

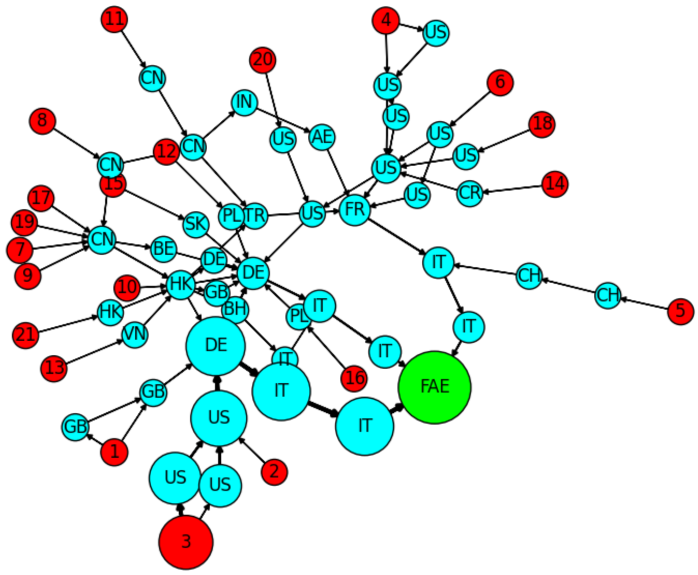

Figure 1: Visualization of the Supply Chain Graph, showing suppliers in red, intermediate facilities in cyan, and the manufacturer in green.

summarized in two complementary ways: a *pointwise* estimate, representing the expected time derived from historical records, and an *interval-based* estimate, which provides a confidence interval reflecting the variability observed in the data.

To ensure a consistent and robust statistical representation, and following established approaches [2], [4], [5], both dispatch and shipment times are modeled with *gamma distributions*, which are well suited for positive, skewed time data. Distribution parameters are estimated from historical records via maximum likelihood estimation (MLE), and the resulting fit is validated to confirm that it accurately reflects the observed central tendency and variability.

The resulting historical indicators provide a statistically grounded basis for planning, yielding both deterministic estimates and uncertainty-aware intervals.

5 Realtime Indicators

Realtime indicators provide updated estimates of shipment progress by combining historical data with current observations. Unlike historical indicators, which rely solely on past distributions, realtime indicators integrate the evolving status of orders, including facility-level events, route conditions, and external factors such as traffic and weather.

The methodology is organized hierarchically. At the facility level, indicators capture processing times based on historical data. At the route level, travel times are derived from historical records and dynamically adjusted with machine learning techniques to reflect current traffic and weather conditions. Path-level indicators then estimate the remaining shipment duration along a *single path*, by aggregating the contributions of the route and facility times registered. Specifically, each path estimate combines two components: the remaining shipment time from the corresponding historical indicator (*Transit Time*, TT) and the estimated remaining time along that path (*Path Time*, PT). These components are combined using a convex combination, $\alpha \cdot \mathrm{TT} + (1-\alpha) \cdot \mathrm{PT}$, where different implementations of the weighting factor α emphasize the historical component early in the shipment, when uncertainty is high, and gradually shift toward the path component as the shipment progresses. Finally, at the system level, path-level estimates are aggregated across all possible paths from a facility to the manufacturer, generalizing the PT as the weighted average of the path times, with weights corresponding to the empirical probabilities of the respective paths. This produces expected delivery times and, when interval-based methods are used, uncertainty bounds that reflect both historical variability and the current shipment state.

Together, these realtime indicators offer a unified framework for adaptive, probabilistically grounded shipment predictions that complement the historical indicators.

6 Evaluation Procedure

Predictive intervals were evaluated using three complementary metrics: *sharpness* (interval width), *coverage* (proportion of observations inside the interval), and *interval score* [6], which balances informativeness and reliability. Metrics were computed at two granularities: **hourly**, using raw model predictions later converted into days for reporting, and **daily**, using predictions discretized into business days and evaluated directly.

Evaluation covered both the **dispatch** stage (where dispatch time must also be estimated) and the **shipment** stage, on both the **training set** and a limited **test set** constructed from a subset of orders not used in any part of model tuning. Multiple model variants were assessed to compare different approaches for computing the weighting factor in the path-level real-time indicator convex combination described above, as well as the machine learning implementation for route time estimation, which incorporates traffic and weather information. With the current limited dataset, the machine learning approach provided only marginal and uncertain improvements, whereas the most effective weighting factor implementation was found to follow a polynomial form. We report only the results for the most robust model, which uses the polynomial weighting factor and relies solely on average values for route time estimation, excluding traffic and weather information.

7 Results

Table 2 reports the results for the most robust model, evaluated at a 90% confidence level. Estimates for the dispatch stage are inherently less precise due to variability in dispatch times, whereas shipment-stage intervals are narrower and more reliable. Considering only the shipment stage, hourly intervals consistently under-cover the observations, while daily intervals provide coverage close to the nominal confidence, achieving a reasonable balance between precision and calibration.

Set	**Stage**	**Sharpness (days)**		**Coverage (%)**		**Interval Score (days)**	
		Hourly	**Daily**	**Hourly**	**Daily**	**Hourly**	**Daily**
Training	Dispatch	2.15	2.20	0.28	0.56	4.22	4.26
	Shipment	1.03	1.18	0.52	0.84	1.45	1.63
Test	Dispatch	2.87	3.00	0.30	0.50	7.81	7.44
	Shipment	1.30	1.55	0.63	0.86	1.99	2.20

Table 2: Evaluation metrics across training and test sets, divided by operational stage (Dispatch and Shipment) and comparing hourly and daily granularities.

We performed a preliminary sensitivity analysis to assess estimation quality at different significance levels. Figure 2 shows observed coverage versus nominal confidence: daily estimates track the ideal trend reasonably well, showing over-reliability up to significance levels of 85% and falling slightly below the ideal trend beyond that point. Hourly estimates consistently fall below the ideal, indicating that, with the current dataset, reliable fine-grained predictions are not yet achievable.

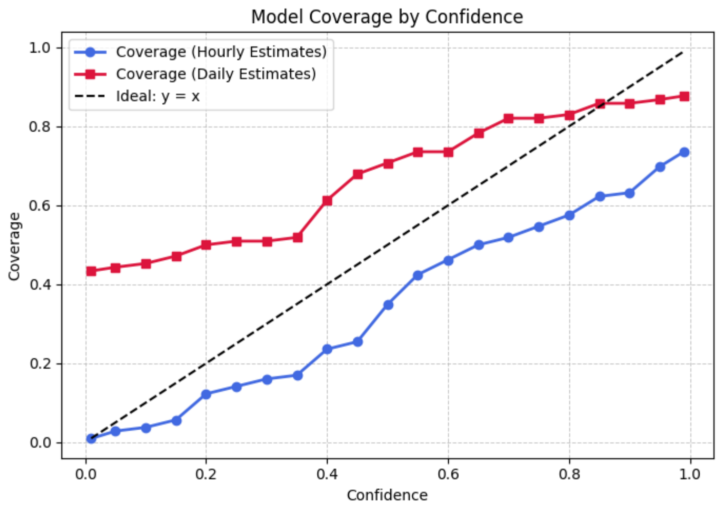

Figure 2: Observed coverage versus nominal confidence.

Figure 3 illustrates the evolution of interval estimates for a representative order. Intervals narrow as the order progresses through shipment stages, reflecting decreasing uncertainty, while the model adapts to real-time deviations, demonstrating responsiveness to unexpected delays.

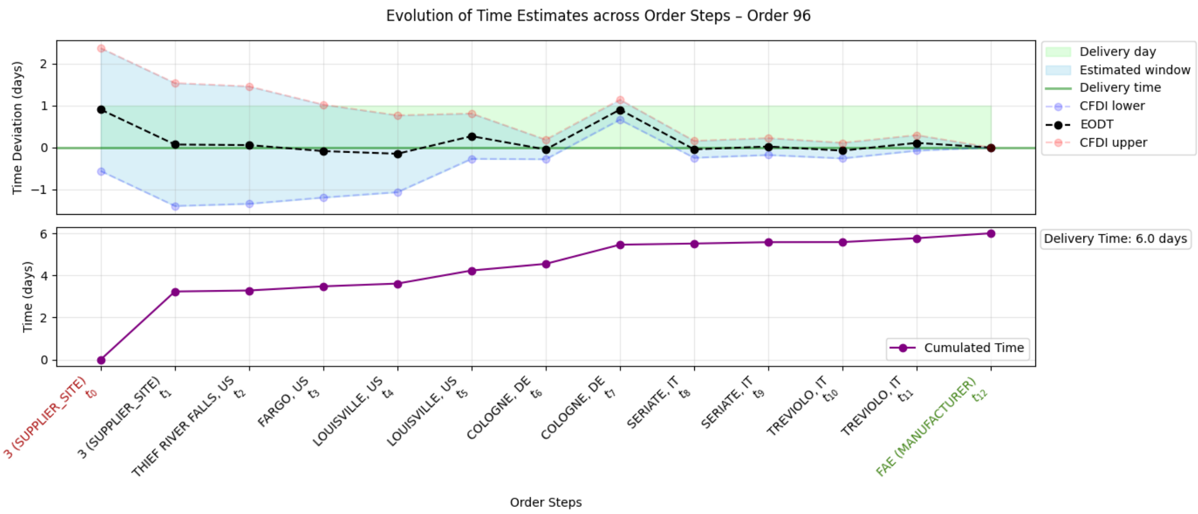

Figure 3: Evolution of interval estimates for a representative order.

Overall, the results suggest that the proposed indicators–particularly at daily granularity–already provide practically useful predictive intervals, achieving a balanced trade-off between sharpness and reliability suitable for operational decision-making. Future work should focus on additional evaluations and potential refinements of the indicator implementations as more data become available.

References

[1] L. R. Ford and D. R. Fulkerson, "Maximal flow through a network," *Canadian Journal of Mathematics*, vol. 8, no. 3, pp. 399–404, 1956.

[2] R. D. Snyder, "Inventory control with the gamma probability distribution," *European Journal of Operational Research*, vol. 17, no. 3, pp. 367–374, 1984. DOI: `10.1016/0377-2217(84)90133-4`.

[3] J. R. Norris, *Markov Chains*. Cambridge University Press, 1998.

[4] K. L. Namit and J. Chen, "Solutions to the (q, r) inventory model for gamma lead-time demand," *International Journal of Physical Distribution & Logistics Management*, vol. 29, no. 6, pp. 373–387, 1999. DOI: `10.1108/09600039910264713`.

[5] A. V. Hill, J. M. Hays, and E. Naveh, "A model for optimal delivery time guarantees," *Journal of service research*, vol. 2, no. 3, pp. 254–264, 2000.

[6] T. Gneiting and A. E. Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[7] M. Consortium. "M4estro partners overview." (2025), [Online]. Available: `https://www.m4estro-project.eu/partners/` (visited on 09/01/2025).

[8] M. Consortium. "M4estro project overview." (2025), [Online]. Available: `https://www.m4estro-project.eu/` (visited on 09/01/2025).

[9] F. Technology. "Fae technology website." (2025), [Online]. Available: `https://fae.technology/` (visited on 09/01/2025).