26th International Conference on Harbor, Maritime and Multimodal Logistic Modeling & Simulation, 013
21th International Multidisciplinary Modeling & Simulation Multiconference

2724-0339 © 2024 The Authors.
doi: 10.46354/i3m.2024.hms.013

Scenario Simulation to Investigate new Solutions for Port Terminals and Plants based on Intelligent Data Fusion and Supply Chain Modeling

Agostino G. Bruzzone[1,3*], Anna Sciomachen[2], Marco Gotelli[3], Antonio Giovannetti[1], Alessia Giulianetti[2], Farshad Shamlu[1,3]

[1] Simulation Team, via Magliotto 2, 17100, Savona, Italy
[2]Department of Economics and business studies, University of Genoa, Italy
[3] Department of Mechanical, Energy, Management and Transport Engineering, University of Genoa, Italy
*Corresponding author. Email address: agostino.bruzzone@simulationteam.com

This paper explores the use of synthetic data generation and strategic engineering to address complex problems in industrial and logistic sector with an emphasis on port terminals. The research introduces the Generator of Logistics Flow (GOLF), a model leveraging discrete event stochastic simulation and AI to create synthetic data that mirrors real-world container flow scenarios. The study emphasizes the role of strategic engineering, which integrates advanced simulation, AI, and data analytics to provide decision-makers with a robust quantitative foundation for policy implementation and strategy design. GOLF allows users to generate data and analyze scenarios efficiently. This research demonstrates the potential of synthetic data and intelligent systems to address data scarcity, enhance operational efficiency, and foster innovation in supply chain logistics and port terminal management.

Keywords: Modeling and Simulation, Synthetic Data, Supply Chain, Logistics, Optimization, Strategic Engineering

1. Introduction

Nowadays it is common to hear that "Data is the new Gold". Indeed, most advanced business models and industrial processes are organized, managed and optimized using moles of available data. However, for obvious reasons, this approach is limited to the situations when the required data is available.

Generating synthetic data offers a promising solution to the reluctance of organizations to share business and operational data, thereby facilitating the testing and implementation of innovative strategies. The supply chain sector, particularly maritime logistics, exemplifies this challenge. Therefore, this paper presents the development of a logistic flow generator designed to address this issue. One of the most important entities in maritime trade are the port terminals, since their effectiveness have a significant impact on the entire supply chain. Handling millions of containers with limited spaces and time constraints is a complex task and a determining aspect to maintaining competitiveness. Indeed, as highlighted by recent studies (Bruzzone et al., 2019) the efficiency of supply chain operations has a direct impact on global trade flow, economic stability, and the capacity to meet the rising demands of international commerce by facilitating the seamless transfer of goods between sea and land transport systems.

To tackle complexity (Bruzzone et al., 2023) and build a resilient ecosystem against geopolitical and natural factors, supply chains need to leverage modeling, simulation, and digital technologies. A new discipline called strategic engineering has been developed to create and implement these solutions within a structured framework. Strategic engineering combines the development of advanced simulation systems with AI and data analytics to create intelligent systems. These systems provide decision-makers with a quantitative foundation to implement new policies, design innovative strategies, and take real-time action.

In this case, Strategic Engineering approach is applied to develop a stochastic simulator to generate the container flow called GOLF (Generator of Logistics Flow) that leverages Modeling and Simulation (M&S), Data Analytics and AI for statistical data analysis and collection regarding container flows.

This approach allows to take benefits of available statistical data (e.g. container terminal capacities and flows, cargo-related statistics of seaports, demand of goods in different areas etc.), to fuse it and to obtain required inputs for the model. Obviously, while some information could be available directly or calculated in straightforward way, it is not the case for entire scenario. Indeed, to reconstruct all operations performed on a container during its travel from origin to destination, it is necessary to take into account various variables, such as travel time of the ship, time to load and unload it, time for customs controls etc., all with relative stochastic component. Considering this, the authors decided to simulate operations of whole logistic chain based on available information and doing so to produce all required data. At the same time, utilization of Discrete Event Simulation (DES) allows to speed up significantly the process of data generation; indeed, in container flow of even a relatively small geographic area could be at an order of millions of containers per year. The discrete event stochastic simulator generates synthetic data, which is then provided for data analysis and used by decision-makers to evaluate potential scenarios even in cases when the availability of real data characterizing terminal container flows and trades is challenging to obtain, e.g. due to owners' confidentiality concerns and business competitiveness reasons.

To provide a valuable tool for decision-makers, the authors developed an analysis to synthesize the actual properties associated with each container. Each container is characterized in terms of transport modes and specific attributes, with each property being stochastically distributed for comprehensive analysis. Specifically, each container is transported by various modes, including ship, truck, train, or feeder ship, and is characterized by its size, type (e.g., reefer), weight, and whether it contains hazardous materials.

Therefore, developing a generator based on a discrete event stochastic simulator and AI for data fusion helps to overcome the lack of real data, providing a quantitative basis for developing new tools and evaluating future scenarios. The validation process is based on comparing synthetic data with available partial data, expert knowledge, and industry standards to refine and validate the data generation models.

2. State of the art

Generating synthetic data is a crucial advancement in overcoming the data lack problem in supply chain analytics and terminal technological and scientific development. Nevertheless, this issue extends to several sectors and use cases where data scarcity represents also a limitation. For instance, in the healthcare sector (Dahmen et al., 2019) the creation of realistic synthetic behavior-based sensor data is crucial for testing machine learning techniques in healthcare applications. Traditional methods often fall short in terms of complexity and realism. To address these limitations, SynSys, a machine learning-based synthetic data generation method, has been developed. SynSys generates synthetic time series data using hidden Markov models and regression models trained on real datasets. On the other hand, the same philosophy where traditional simulation techniques are frequently limited by the availability of real-time data and the substantial computational resources required to model complex systems has been adopted to predict and optimize interdiction operation in maritime domain (Bruzzone et al., 2017) where these models provide a quantitative basis for evaluating various scenarios and strategies, enhancing decision-makers' ability to respond effectively.

Another important use case is identified related to statistical testing (Soltana et al., 2017). Indeed, it needs knowledge about a system's actual or anticipated usage profile to estimate reliability, often requiring the generation of synthetic test data. This is clearly related to critical aspects that are commonly discussed in several industries when studying predictive models. The synthetic data must mirror the statistical characteristics and logical validity constraints of the actual data the system is going to actually process. For data-intensive systems, a new approach generates synthetic test data that is both statistically representative and logically valid. Initially, a data sample meeting statistical characteristics is created without considering logical constraints. Then, an iterative tweaking process corrects any logical violations while maintaining statistical properties. Such methodologies have significant implications across various sectors where high-value assets are involved. For example, they are utilized in fraud detection on major e-commerce platforms, predictive maintenance of costly machinery, and ensuring the predictive quality of continuous production plants.

At a more fundamental level, synthetic data are also used to review and optimize deep learning algorithms. Advances in rendering pipelines, generative adversarial networks (GANs), and fusion models have significantly improved the generation of synthetic data (de Melo et al., 2022). These synthetic datasets intrinsically don’t have an a priori limit in terms of quantity and they are all pre-labeled. Also, domain adaptation techniques further align synthetic data with real world, paving the way to sophisticated learning methods closely reproducing human cognitive processes. These considerations are obviously investigated in industrial plants sector where data is often confidential, scarce, or heterogeneous. Reliable synthetic datasets to support simulations and data analytics (Libes et al., 2017) that maintains statistical representativeness and logical validity are fundamental to improve efficiency and competitiveness in the market. Therefore, it is clear that developing technologies for synthetic data is an urgent and crucial aspect for research and innovation.

Another example of intelligent systems for supply chain optimization is related to shoe manufacturing (Bruzzone et al., 2022). In this case study, the authors created a decision support system capable to extract and fuse non homogeneous data about relative supply chain, use this information to simulate entire process and to provide decision makers with comparison of AS-IS and TO-BE scenarios, using obtained Key Performance Indicators (KPI).

3. GOLF (Generator of logistic Flow)

Based on analysis of scenario, identification of available data and survey of existing solutions, the authors developed GOLF (Generator of Logistics Flow). As explained previously, the system is composed of modules responsible for data extraction, fusion, stochastic simulation, visualization and validation. At the same time, the solution makes hypothesis where the data is not available directly, e.g. by using population of city and industrial factors to evaluate demand of goods in geographic areas.

3.1. Model Development

The Generator Model calculate the primary and secondary modes of transportation for each container, specifying if a container is transported by truck, train, oceanic ship or feeder ship for each scenario.

The simulator vectorize the different transportation modes in order to compute the number of transport modes for each scenario segmenting the journey into primary and secondary modes of transportation. So far, the container is transported by the four main ports of Genoa Port, Savona, La Spezia and Livorno. Each terminal has different characteristics that are configured via a configuration file, which reflects the specific trade trends of each terminal. The terminal to be simulated and the amount of containers to be shown in the output file are defined in GOLF interface

The simulator estimates travel times for containers, including time spent in the port and travel time for each transportation segment based on the configuration file which gathers information from a configuration file. After the time estimation the simulator assigns transportation carriers to containers. GOLF Simulator is designed with a modular architecture to ensure scalability and maintainability. The System Design includes different modules such as configuration reader, data sheet reader, flow mode generator, distance calculator, journey time calculator, and vectorization functions.

3.2. Equations Implemented

The simulator uses probability distributions to select cities, ports, and transportation modes. This approach mimics real-world variability and randomness in logistics operations. The simulator handles various configurations and scenarios by adjusting input parameters.

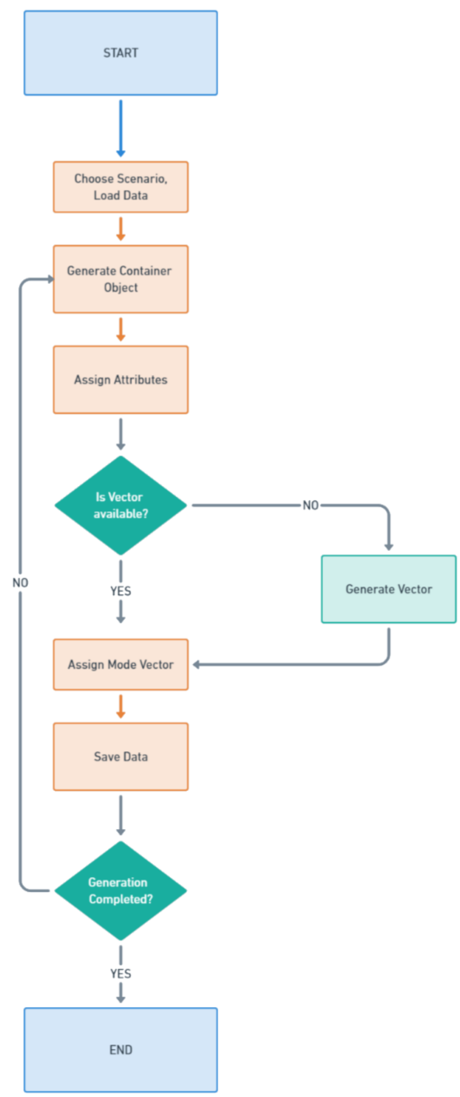

Figure 1. Logistic Scheme GOLF

The distance between two geographical points is calculated using the Haversine formula, which accounts for the Earth's curvature:

$$d = 2r \cdot arcsin(\sqrt{sin^2\left(\frac{\Delta Ø}{2}\right) + \cos(Ø_1) \cdot \cos(Ø_2) \cdot sin^2\left(\frac{\Delta\mu}{2}\right)}$$

Where:

- d is the distance between the points
- r is the Earth’s radius
- $Ø_1, Ø_2$ are the latitudes of two points in radians

- $\Delta\mu$ is the difference between the longintudes

Journey times are calculated based on the selected transportation modes and their respective durations. For each mode, the duration is obtained from the input data sheets. The total journey time includes the time spent in port and the travel time for each segment of the journey, therefore the total journey time is:

$$T_{Total} = T_{First\ Transportation\ mode} + T_{in\ port} + T_{second\ transportation\ mode}$$

Where each time is based on Beta Distribution.

The simulator assigns transportation vectors to containers, grouping them based on their destinations, sources and transport modes.

This step ensures that containers with similar characteristics are processed together, improving the simulation's efficiency.

3.3. GOLF GUI

The model is developed using the Python programming language. To make it easily accessible for users to generate data and analyze new scenarios, we developed a graphical user interface (GUI). This GUI allows users to input the terminal and the number of containers to be generated.

Thus far, GOLF is configured to simulate the main terminals of the Port of Genoa, as well as the ports of Savona, La Spezia, and Livorno. However, users defines other ports in the configuration file and simulate their trade dynamics by specifying the probability of each port trading with other ports and countries. The number of containers to be simulated is specified by the user.

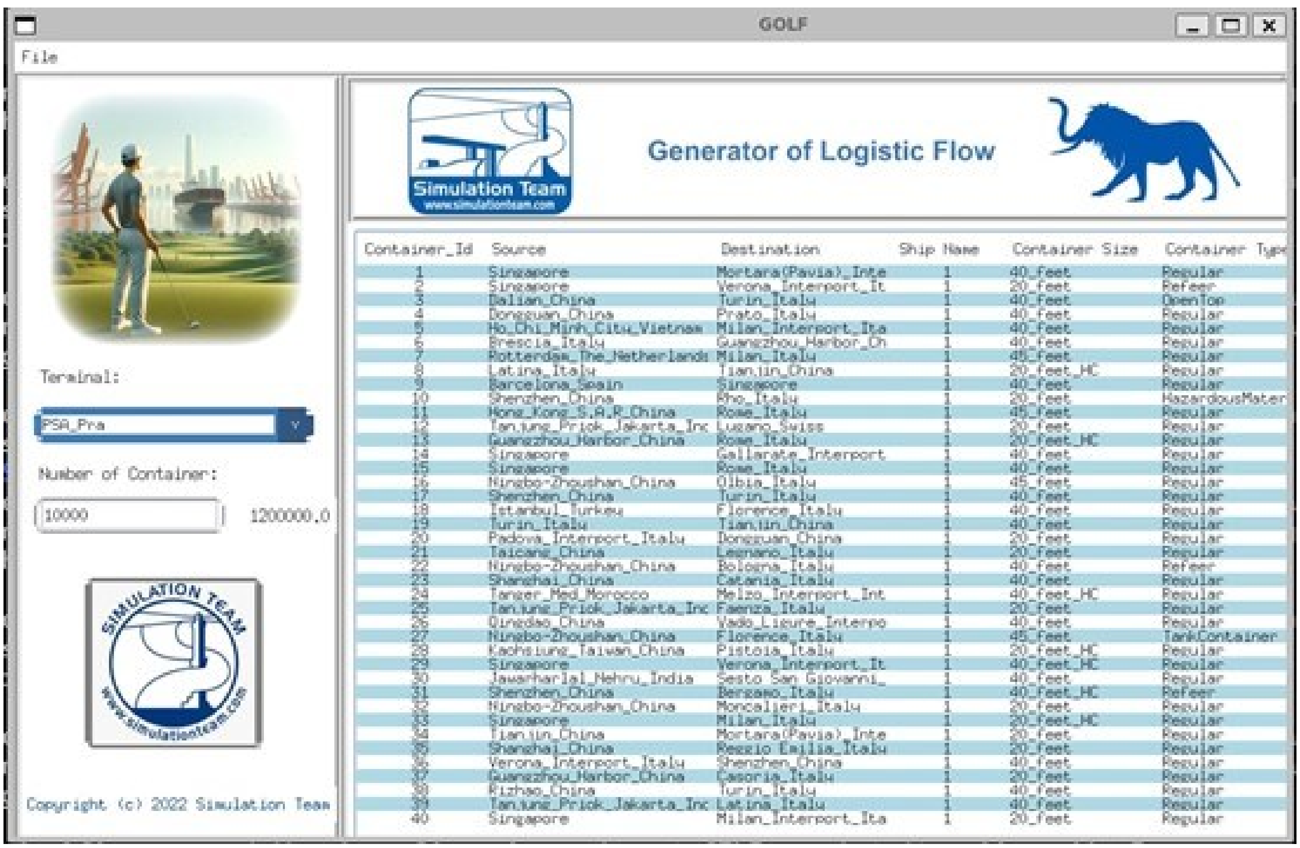

Figure 1 – GOLF GUI

4. Validation of Synthetic Data Generation

One of main criticalities of data generation is related to necessity of ensuring correctness of produced results. Indeed, in order to be applicable for further utilization, the outputs need to be statistically similar to the information obtained from field, e.g. by comparison of delivery times and container flows in real-world with estimations provided by the program.One of methods to validate outputs of stochastic data generator could be by comparison of expected logistic flows with that ones produced by the program. The scenario considered for validation has been set with the generation of 100,000 container in the terminal of Genova PSA Voltri.

Destination	Configuration File	GOLF stochastic algorithm
Shanghai_China	8078	8101
Singapore	6439	6397
Ningbo-Zhoushan_China	5337	5361
Shenzhen_China	4942	4852
Guangzhou_Harbor_China	4153	4230
Qingdao_China	4072	4132
Tianjin_China	3482	3422
Hong_Kong_S.A.R_China	3057	3013
Port_Klang_Malaysia	2357	2393
Busan_South_Korea	2340	2343

Table 1. Number of container for the first ten destinations

In Table 1 are reported the first ten destinations with the associated number of containers generated with the configuration file and after the logistic flow generator has been run. At this point it is possible to compare expected and observed results, which seem to differ only by several percent, while the most important flow has very small error and the biggest relative errors are attributed to relatively small and insignificant sub-flows.

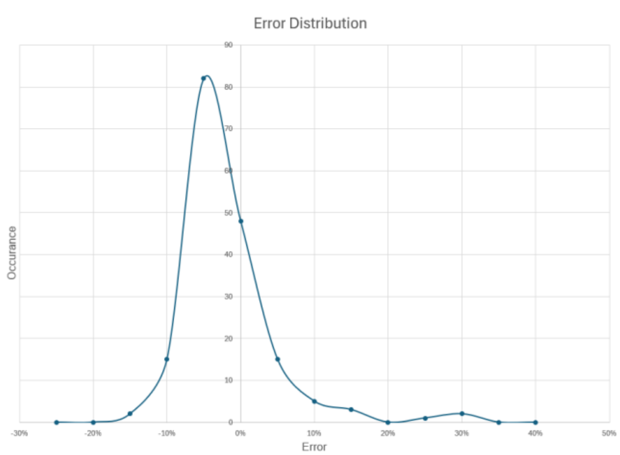

Figure 3. Error and Observation Graph

Considering this, it is possible to employ generated data about logistic flows as input for further modeling and analysis.

5. Conclusions

For many practical applications it is common nowadays to rely on big quantities of available data, however, this approach is not applicable to the cases when only general statistics is available. Considering this, the authors developed a logistic flow data generator, capable to produce statistically accurate information about flows of goods. This approach allows to feed other simulation models in order to perform what-if analysis and optimization of current logistics and transportation networks.

References

Bruzzone, A. G., Gotelli, M., Giovannetti, A., De Paoli, A., Ferrari, R., Pedemonte, M., ... & Frosolini, M. (2023). Strategic Engineering for Decision Making during Urban Crises.

de Melo, C. M., Torralba, A., Guibas, L., DiCarlo, J., Chellappa, R., & Hodgins, J. (2022). Next-generation deep learning based on simulators and synthetic data. Trends in cognitive sciences, 26(2), 174-187.

Dahmen, J., & Cook, D. (2019). SynSys: A synthetic data generation system for healthcare applications. Sensors, 19(5), 1181.

Bruzzone, A. G., Massei, M., Sinelshchikov, K., Fadda, P., Fancello, G., Fabbrini, G., & Gotelli, M. (2019). Extended reality, intelligent agents and simulation to improve efficiency, safety and security in harbors and port plants. In 21st International Conference on Harbor, Maritime and Multimodal Logistics Modeling and Simulation, HMS 2019 (pp. 88-91). CAL-TEK.

Bruzzone, A. G., Di Matteo, R., Maglione, G. L., & Massei, M. (2017, July). Simulation models and artificial neural networks for vessels behavior analysis. In Proceedings of the Summer Simulation Multi-Conference (pp. 1-9).

Bruzzone, A. G., Massei, M. & Frosolini, M. (2022). Redesign of supply chain in fashion industry based on strategic engineering. Procedia Computer Science, 200, 1913-1918.

Soltana, G., Sabetzadeh, M., & Briand, L. C. (2017, October). Synthetic data generation for statistical testing. In 2017 32nd IEEE/ACM International Conference on Automated Software Engineering (ASE) (pp. 872-882). IEEE.

de Melo, C. M., Torralba, A., Guibas, L., DiCarlo, J., Chellappa, R., & Hodgins, J. (2022). Next-generation deep learning based on simulators and synthetic data. Trends in cognitive sciences, 26(2), 174-187.

Libes, D., Lechevalier, D., & Jain, S. (2017, December). Issues in synthetic data generation for advanced manufacturing. In 2017 IEEE International Conference on Big Data (pp. 1746-1754). IEEE.