Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Disruption evaluation in end-to-end semiconductor supply chains via interpretable machine learning

Friedrich-Maximilian Jaenichen*, Christina J. Liepold**, Abdelgafar Ismail*,
Maximilian Schiffer***, Hans Ehm*

*Infineon Technologies AG, Am Campeon 1-15, 85579 Neubiberg, Germany (e-mail: Maximilian.Jaenichen@infineon.com).

** School of Management, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

*** School of Management & Munich Data Science Institute, Technical University of Munich, Arcisstr. 21, 80333 Munich, Germany

Abstract: COVID-19 has posed unprecedented challenges to global health and the world economy. Two years into the pandemic, the widespread impact of COVID-19 continues to deepen, impacting different industries such as the automotive industry and its supply chain. This study presents a hybrid approach combining simulation modeling and tree-based supervised machine learning techniques to explore the implications of end-market demand disruptions. Specifically, we apply the concept of born-again tree ensembles, which are powerful and, at the same time, easily interpretable classifiers, to the case of the semiconductor industry. First, we show how to use born-again tree ensembles to explore data generated by a supply chain simulation model. To this end, we demonstrate the influence of varying behavioral and structural parameters and show the impact of their variation on specific key performance indicators, e.g., the inventory level. Finally, we leverage a counterfactual analysis to identify detailed managerial insights for semiconductor companies to mitigate adverse impacts on one echelon or the entire supply chain. Our hybrid approach provides a simulation model enhanced by a tree-based supervised machine learning model that companies can use to determine optimal measures for mitigating the adverse effects of end-market demand disruptions. We close the loop of our analysis by integrating the findings of the counterfactual analysis backward into the simulation model to understand the overall dynamics within the multi-echelon supply chain.

Copyright © 2022 The Authors. This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: Supervised learning, counterfactual analysis, semiconductor supply chain, chip shortage

1. INTRODUCTION

The COVID-19 pandemic is the blueprint for an unprecedented supply chain risk scenario with high uncertainty, ripple-effect like disruption propagation, and concurrent impacts on demand, supply, and logistics (Ivanov et al., 2017; Dolgui et al., 2019; Ivanov, 2020). These disruptions result from various measures imposed by governments worldwide to control the spread of the pandemic (Ivanov, 2020). To guarantee the best recovery scenario, companies rely on trust and collaboration along their supply chain (Domínguez et al., 2014; Jaenichen et al., 2021). Still, end-market disruptions such as the COVID-19 pandemic particularly challenge upstream supply chain members (Fransoo & Udenio, 2020; Sterman, 1989).

The semiconductor industry is particularly susceptible to such disruptions due to its high capital intensity, extended capacity lead times, and long production cycle times (Mönch et al., 2011). Additionally, short life cycles due to fast technological development and highly volatile demand characterize the industry (Aytac & Wu, 2013; Ehm et al., 2011a; Ehm et al., 2011b). Because of these characteristics, semiconductor companies face the bullwhip effect, which itself leads to additional operational disturbances and challenges such as the chip shortage (Fransoo & Udenio, 2020).

According to Chilmon and Tipi (2020), the area of simulation modeling moves toward combining different techniques where artificial intelligence (AI) or machine learning (ML) regularly enhances simulation modeling to improve interpretability. Additionally, recent developments in ML, such as born-again tree ensembles (Vidal & Schiffer, 2020) and the concept of counterfactual analysis (CA) (Wachter et al., 2017; Dandl et al., 2020) facilitate the interplay of simulation and ML, (Chilmon & Tipi, 2020). Against this background, we present a hybrid approach for integrating simulation modelling into a supervised ML framework. Our goal is to evaluate demand disruptions in an end-to-end semiconductor SC focusing on upstream semiconductor companies. We utilize our results to assess how the company can respond to unprecedented demand disruptions.

The remainder of this paper is structured as follows: Section 2 provides a brief overview of related work in supervised ML models. Subsequently, Section 3 presents the proposed hybrid
approach with a simulation model and ML. Section 4 discusses the results of this study. Finally, Section 5 provides concluding remarks and managerial insights.

2. RELATED LITERATURE

Supply chain simulation is a widely studied topic. We refer to Chilmon & Tipi (2020), who predict a trend toward hybrid modeling approaches, i.e., the combination of simulation and ML techniques, for an in-depth literature review. In this context, we will briefly discuss related simulation studies and recent developments in ML in the following.

Fowler et al. (2015) present four aggregation levels for supply chain simulation in semiconductor manufacturing where end-to-end supply chain studies are located at the highest abstraction level and represent, e.g., interactions between companies (Yagi et al., 2014). These have recently been studied in several system dynamics (SD) simulation models: a multi-echelon supply chain of a chemical company (Udenio et al., 2015), a four echelon manufacturing supply chain (Olivares-Aguila & ElMaraghy, 2020), a general supply chain evaluation study under demand disruptions (Fransoo & Udenio, 2020), and an end-market demand disruption model for the semiconductor industry (Jaenichen et al., 2021).

ML techniques can further leverage simulation model outputs (Chilmon & Tipi, 2020). However, sophisticated ML models, such as deep neural networks, may lack comprehensibility (Ventura & Cerquielli, 2019) and interpretability, i.e., understanding of the decision path (Guidotti et al., 2018). However, model interpretability is very important for decision-makers (Rebala et al., 2019). Research has shown that decision trees provide high model interpretability and process diverse data input (Guidotti et al., 2018; Rokach, 2016). However, the greedy nature of decision trees may not lead to good accuracy (Breiman et al., 1984; Rebala et al., 2019). In this context, Ho (1995) suggests the construction of several trees based on subspaces of the data. Breiman (2001) further adds to the concept by introducing the notion of the random forest, which as of today, belongs to the best performing algorithms concerning classification accuracy (cf. Bertsimas & Dunn, 2017), but shows lower comprehensibility due to its design complexity (Guidotti et al., 2018; Rebala et al., 2019; Vidal & Schiffer, 2020). Breiman and Shang (1996) explore the possibility of creating one representation of several decision trees: born-again tree ensembles, which show a much lower test error (Breiman, 1996). However, born-again tree ensembles are generally bigger than classical decision trees (Breiman et al., 1984), and runtime increases substantially due to higher complexity (Breiman & Shang, 1996). Vidal and Schiffer (2020) further investigate these ensembles and introduce a dynamic programming algorithm for building optimal born-again tree ensembles.

Recent developments in explainable AI to further increase interpretability include CA (Verma et al., 2020), which identifies necessary adaptations in the input values to modify the output label (Fernández et al., 2020). The concept of CA was introduced by Wachter et al. (2017) as a mathematical optimization problem and further extended by a multi-objective loss function (Dandl et al., 2020). CA has been applied to random forests in Fernández et al. (2020) and tree ensembles in Lucic et al. (2019). Parmentier and Vidal (2021) explore highly performant optimal counterfactual explanations for tree ensembles.

Combining techniques from the research streams reviewed above, we propose a hybrid combination of SD simulation and supervised ML techniques to leverage results and insights from simulation studies, focusing on the semiconductor industry. By so doing, we identify a clear opportunity to understand how to aid in mitigating the current semiconductor shortage due to strong demand disruptions in the end-market.

3. METHODOLOGY

Our hybrid approach evaluates end-market demand scenarios and mitigates their effect on multi-echelon supply chains. The overall hybrid approach first carries out the simulation based on defined KPIs for data generation before training both a smaller and a more extensive random forest and transforming each into a decision tree counterpart with the same performance, i.e., a born-again tree ensemble. Such a born-again tree ensemble is interpretable by design and subsequently allows to perform a CA to identify mitigation measures for end-market demand disruptions. We first focus on the simulation model and data pre-processing in Subsection 3.1. Subsequently, Subsection 3.2 introduces our tree-based supervised machine learning approach. Finally, we present the CA in Subsection 3.3.

3.1 Simulation Modeling and data pre-processing

The simulation model presented in Jaenichen et al. (2021) consists of four SC echelons: the semiconductor companies, the Tier-2 and Tier-1 suppliers, and the original equipment manufacturer (OEM). The demand is aggregated globally, equaling the global demand for light vehicles. The underlying structure of each echelon bases on the same SD modeling approach (Sterman, 2000, Udenio et al., 2015).

To allow for more detailed analyses, we extend the simulation of Jaenichen et al. (2021) by four key performance indicators (KPIs): demand fulfillment, capacity utilization, inventory coverage, and forecast accuracy. Each of them quantifies a relevant business performance aspect. Each KPI can reach an uncritical (green), semi-critical (yellow), and critical (red) status; we determine thresholds for each status and KPI based on domain knowledge. Our extended model evaluates the supply chain setting and compares different end-market demand scenarios quantitatively.

We generate SC data by running our simulation in a parameter variation experiment for each specified scenario. To this extent, we vary behavioral parameters, structural parameters, and the level of end-market disruption. First, we vary the semiconductor companies’ behavioral parameters, including the inventory adjustment time (IAT) and the forecast adjustment time (FAT). A lower FAT reflects a nervous forecast adjustment, where the received demand signal changes the forecast by the same magnitude. We consider a low, medium, and high FAT scenario. Demand variation is smoothed over an extended period in the high FAT scenario. The hybrid approach considers three inventory adjustment
scenarios: nervous, medium, and smooth. The IAT reflects the speed of adjusting the desired inventory level during a demand drop. Hence, a lower IAT leads to a quicker adjustment in the desired inventory level, while a higher IAT reflects a less nervous adjustment policy. Second, we vary structural parameters, i.e., lead time towards the Tier-1 supplier, production time (based on product complexity), and the semiconductor companies’ desired inventory coverage (DIC) level. Third, the end-market demand disruption is crucial for the overall data generation approach. Each of the scenarios either has a balanced slope of both drop and recovery (symmetric V-shape), a steeper drop (italic V-shape), or a steeper recovery (inverse italic V-shape). The length of the demand drop ranges from 0 weeks, i.e., no disruption, to a maximum of 52 weeks. The magnitude of the demand drop determines the dynamics during each simulation run ranging from 0%, i.e., no disruption in the end-market demand, to the extent of the COVID-19 pandemic end-market demand scenario. Appendix A lists the parameter for the data scenario generation.

For the scope of this study, we aggregate the four KPIs to an overall score with a maximum value of twelve (reaching the value of three for all four KPIs) and a minimum value of four (where each KPI equals one). This aggregation aims to classify the severity of an end-market demand outcome for the semiconductor supply chain and indicates whether there is an immediate need to act upon the anticipated demand scenario. Red scenarios with an overall score lower than eight lead to severe distortion of the end-to-end semiconductor automotive supply chain and require immediate operational attention. Yellow scenarios with a score between eight and ten are less severe but still lead to an undesirable business situation. Scenarios classified with the green status (minimum score of eleven) do not require immediate intervention.

3.2 Tree-based supervised learning approach

We follow the objective of classifying the generated scenarios according to the aggregated score described above. The goal is to indicate whether the company needs to take immediate action to counteract the anticipated demand scenario considering the overall parameter of the respective semiconductor echelon.

At first, we develop a decision tree classifier as a benchmark model and generate first insights on the model performance. We train a decision tree with a depth of three for initial insights into the most critical features’ data structure and decision thresholds. As the ML model in this study is used for a multi-label classification problem, each label from the sample data must be assigned correctly to achieve a high model score (Pedregosa et al., 2011).

Subsequently, we train and test two random forest classifiers with depths of three and eight and perform a hyperparameter optimization to improve their performance. Probst et al. (2019) indicate that random forests require only a minimum of hyperparameter fitting compared to other algorithms, e.g., support vector machines. The ML implementation uses the Scikit learn library version 0.24.1 (Pedregosa et al., 2011) and is run in Python 3.7.1.

We analyze and compare the accuracy and performance of born-again tree ensembles compared to the random forest and decision tree classifier. Finally, born-again tree ensembles are constructed from the two instances of random forests with a depth of three and ten estimators and a depth of eight and 30 estimators. We test the four objectives of born-again tree ensembles based on Vidal and Schiffer (2020) and choose the best performing for the CA.

3.3 Application of counterfactual analysis

We leverage CA to identify actions that improve the overall status of the semiconductor companies. We then test the analysis results, i.e., the counterfactuals, recursively in the simulation model to verify improvements for the supply chain. Before running the supervised learning models, we separate one percent of all data samples for analyses. The born-again tree ensemble constructed from the random forest with estimators at a depth of eight is the basis for the CA. All separated samples are iteratively analyzed concerning changes in the overall class label. We focus on two types of label changes from within the auction space of the semiconductor companies: improvement from red to yellow or green and improvement from yellow to green. In a first step, we assume that the values of all features can be varied as we want to identify the full potential of the CA. Subsequently, we exclude demand scenario and structural parameters; thus, we only consider behavioral parameters and the DIC of the semiconductor echelon. We use the CEMl package by Artelt (2019) that supports tree-based models.

4. RESULTS

The parameter variation experiment, which we performed in AnyLogic 8, runs all combinations of parameters without randomness. Overall, the simulation output has 381,888 samples with eight features, where every single output represents an end-market demand scenario with a specific parameter combination. As the Pearson correlation coefficient and the covariance is low for all eight features, we keep all features. We train and test the benchmark decision tree classifier with default settings from the Sklearn library with a depth of three. We derive the importance of each feature as the normalized total reduction of Gini impurity per feature (Pedregosa et al., 2011) to find the essential features for the classification task. The drop length and magnitude in the end-market demand show feature importance of 0.785 and 0.210, respectively. The IAT shows a value of 0.005.

Figure 1 displays a scatterplot of the two most essential features, the respective overall score, and a clear relationship to better understand the characteristics of the data set. We observe that the values for the features increase with a lower overall sum score. Figure 1 shows mainly red scenarios in its upper right. Thus, we further investigate this set of features to balance the class distribution. Data samples are primarily filtered according to magnitude and drop length since very high values of these inputs predominantly lead to the red state. The filtering results in a much more balanced share between the green, yellow, and red status of 33.4%, 33.6%, and 33.0%, respectively.
Table 1 displays the train and test accuracy and F1-score of the decision tree, random forest with a depth of three, and the pruned born-again tree ensemble. All four objectives for building the born-again tree ensemble lead to the same train decision tree, random forest with a depth of three, and the pruned born-again tree ensemble as the basis for the CA. Of the initially 1,627 separated samples, i.e., one percent of the filtered and balanced data, the CA finds 397 counterfactuals for the decision rules for the classification. We use the pruned born-again tree ensemble to employ this strategy.

### Table 1: Results of supervised learning approaches – 10 estimators

| Classifier | Train-Accuracy | Train-F1 | Test-Accuracy | Test-F1 |
|------------|----------------|----------|---------------|---------|
| (1)        | 0.82605        | 0.82809  | 0.82612       | 0.82807 |
| (2)        | 0.84297        | 0.84111  | 0.84111       | 0.84279 |
| (3)        | 0.84297        | 0.84111  | 0.84111       | 0.84279 |

(1) – decision tree | (2) – random forest with depth of 3 | (3) pruned born-again tree ensemble

The result for the classification with a higher depth in estimators shows a similar pattern (see Table 2). The born-again tree ensemble outperforms the decision tree classifier with a slightly lower classification performance than the random forest with 30 estimators. Only one estimator represents the born-again tree ensemble, providing direct decision rules for the classification. We use the pruned born-again tree ensemble as the basis for the CA. Of the initially 1,627 separated samples, one percent of the filtered and balanced data, the CA finds 397 counterfactuals for the previously determined feature range (c.f. Appendix 1). A total of 185 red samples can be improved to yellow and 210 from yellow to green. The overall status is improved from red directly to green for two samples.

### Table 2: Results of supervised learning approaches – 30 estimators

| Classifier | Train-Accuracy | Train-F1 | Test-Accuracy | Test-F1 |
|------------|----------------|----------|---------------|---------|
| (1)        | 0.95438        | 0.95438  | 0.95261       | 0.95261 |
| (2)        | 0.96278        | 0.96276  | 0.96177       | 0.96173 |
| (3)        | 0.96095        | 0.96094  | 0.95940       | 0.95939 |

(1) – decision tree | (2) – random forest with depth of 8 | (3) pruned born-again tree ensemble

In the following, we present examples of the CA to demonstrate its capabilities in aiding decision-makers. Table 3 shows three different characteristics for the definition of the CA: the initial feature values for cycle time, DIC, FAT, IAT, lead time, drop length, drop magnitude, and type of V-shape, the proposed delta in the respective feature value, and the updated feature values with the improved outcome. The example shows a severe drop of eleven percent for 44 weeks in a reverse italic V-shape, i.e., the value of -0.5 for the last feature value, reflects a shorter recovery than the drop period. As determined by the CA, a ten percent increase in DIC changes the overall status in this setting from red to yellow.

### Table 3: Recommended change in desired inventory coverage

| Initial | Delta | Updated | Class |
|---------|-------|---------|-------|
| 14 10 3 | -1    | 14 11 3 | Red |
| 30 44 11 | -0.5  | 30 44 11 | Yellow |

(1) – Cycle time | (2) – DIC | (3) – FAT | (4) – IAT | (5) – lead time | (6) – drop length | (7) – drop magnitude | (8) – V-shape

Another example is listed in Table 4, where the CA detects the improvement potential of the overall state by adjusting two features simultaneously. In this scenario, a reduced DIC and a smoother forecast behavior, i.e., a higher value for the FAT, improve the overall status from yellow to green. This insight is of particular interest as the inventory level can be decreased relatively quickly should the semiconductor company decide to employ this strategy.

### Table 4: Recommended change in desired inventory coverage and forecast adjustment time

| Initial | Delta | Updated | Class |
|---------|-------|---------|-------|
| 24 14 1 | -3    | 24 16 6 | Yellow |
| 10 30 22 | 0.03 -0.5 | 10 23 22 | Green |

(1) – Cycle time | (2) – DIC | (3) – FAT | (4) – IAT | (5) – lead time | (6) – drop length | (7) – drop magnitude | (8) – V-shape

Overall, the CA identified two situations where the decision-maker may improve the status from red to green. Interestingly, this significant improvement is achieved by only adapting the DIC and the IAT. One of these scenarios (see Table 5) reflects a demand drop of seven percent over 20 weeks in a symmetric V-shaped end-market demand. An increase from eight to 13 weeks of reach combined with a much higher IAT is necessary to achieve the green status.

### Table 5: Recommended change in desired inventory coverage and inventory adjustment time

| Initial | Delta | Updated | Class |
|---------|-------|---------|-------|
| 24 8 7 | 5     | 24 12 12 | Red |
| 10 25 20 | 0.07 0.0 | 10 25 20 | Green |

(1) – Cycle time | (2) – DIC | (3) – FAT | (4) – IAT | (5) – lead time | (6) – drop length | (7) – drop magnitude | (8) – V-shape

We now use the recommended adaptations from Table 5 and feed them back into the SD simulation model. The comparison of the KPIs in Figure 2 compares the simulation output for the overall available capacity (semi), the work-in-process (WIP semi), and the target WIP (Target WIP semi) of the semiconductor manufacturer (a) pre and (b) post adaptation. Based on Figure 2, we find that the two adaptations of the feature values, as listed in Table 5, lead to a better control of the
inventory level during both the demand drop and the market recovery phase.

While the target, i.e., desired WIP, exceeds available capacity after the end-market demand recovery, the updated feature values generate a business environment in which we can significantly reduce this overshoot. This is visualized in the Figure with the overall available capacity, hence, the upper limit of the WIP. Additionally, we observe a significant improvement of order fulfillment in the optimized setting. All orders can now be fulfilled instead of only 94.7% previously. While the initial scene shows a deviation from the desired stock level of more than 15%, the adapted sample does not reach the red status. Finally, improved mean capacity utilization during recovery and better control of the desired WIP level leads to an overall green status for this scenario.

![Figure 2: Simulation output (a) pre and (b) post adaptation](image)

This backward integration of the improved parameter setting identified by CA to the simulation model concludes the result section.

5. CONCLUSION

In conclusion, we were able to apply the methodology presented in Vidal and Schiffer (2020) to a specific application in the semiconductor industry. Extending the insights from the supervised learning analysis, we utilized CA to provide concrete actions for decision-makers when dealing with changes in the end-market demand. To this end, we can provide instructions which parameters need to be adapted to improve the overall supply chain performance. While traditional measures for mitigating supply chain disruptions, e.g., buffering by increased inventory, are vastly popular and intuitive to the practitioner, our analyses revealed settings where such an approach is not beneficial. Accordingly, our CA constitutes a promising extension to the hybrid approach of simulation and ML as it allows the user to increase understanding of the complex system at hand.

Additionally, based on insights from the presented methodology, the semiconductor company can react to anticipated demand patterns and improve the overall status of their supply chain situation from red to yellow or yellow to green. In general, both changes reflect a significant opportunity to improve the supply chain performance for semiconductor companies, which ideally leads to a significant chance of reducing chip shortages along the whole supply chain. The insights from this analysis can also encourage a discussion around inventory management and forecasting with exact values determined through CA. Finally, our study also showed that, in some cases, excess redundant inventory levels are also a risk for the supply chain performance of the semiconductor manufacturer. This indicates that CA provides value on top of supervised learning approaches by providing actual thresholds for relevant supply chain features. The backward integration of the proposed counterfactuals is valuable for decision-makers in practice as it increases the understanding of how scenario modifications impact the supply chain. Future studies may extend the investigation to supply chains of other sectors within the same industry, e.g., consumer-grade electronics, and different notions of disruptions aside from unique outliers, e.g., continuous raw material supply shortages.

REFERENCES

Artelt, A. (2019). CEML-Counterfactuals for Explaining Machine Learning models-A Python toolbox. GitHub repository. URL https://www.github.com/andreArtelt/ceml.

Aytac, B., & Wu, S. D. (2013). Characterization of demand for short life-cycle technology products. *Annals of Operations Research*, 203(1), 255-277.

Breiman, L., Friedman, J.H., Olshen, R.A., & Stone, C.J. (1984). *Classification and regression trees* (1st ed.). Routledge.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2), 123-140.

Breiman, L., & Shang, N. (1996). Born again trees. *University of California, Berkeley, Berkeley, CA, Technical Report*, 1(2), 4.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Bertsimas, D., & Dunn, J. (2017). Optimal classification trees. *Machine Learning*, 106(7), 1039-1082.

Chilmon, B., & Tipi, N. S. (2020). Modelling and simulation considerations for an end-to-end supply chain system. *Computers & Industrial Engineering*, 150, 106870.

Dandl, S., Molnar, C., Binder, M., & Bischl, B. (2020, September). Multi-objective counterfactual explanations. In *International Conference on Parallel Problem Solving from Nature* (pp. 448-469). Springer, Cham.

Ehm, H., Ponsignon, T., & Kaufmann, T. (2011a). The global supply chain is our new fab: Integration and automation challenges. In *2011 IEEE/SEMI Advanced Semiconductor Manufacturing Conference* (pp. 1-6). IEEE.

Ehm, H., Wenke, H., Mönch, L., Ponsignon, T., & Forstner, L. (2011b). Towards a supply chain simulation reference model for the semiconductor industry. In *Proceedings of the 2011 Winter Simulation Conference (WSC)* (pp. 2119-2130). IEEE.

Fernández, R. R., De Diego, I. M., Aceña, V., Fernández-Isabel, A., & Moguerza, J. M. (2020). Random forest explainability using counterfactual sets. *Information Fusion*, 63, 196-207.

Fowler, J.W., Mönch L., Ponsignon, T. (2015). Discrete-event simulation for semiconductor wafer fabrication facilities: A tutorial. *International Journal of Industrial Engineering*, 22(5), 661–682.

Fransoo, J. C., & Udenio, M. (2020). Exiting a COVID-19 Lockdown: The Bumpy Road Ahead for Many Supply Chains. *Available at SSRN 3590153.*
Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR), 51*(5), 1-42.

Ho, T. K. (1995, August). Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition (Vol. 1, pp. 278-282). IEEE.

Ivanov, D., Dolgui, A., Sokolov, B., & Ivanova, M. (2017). Literature review on disruption recovery in the supply chain. *International Journal of Production Research, 55*(20), 6158-6174.

Ivanov, D., Dolgui, A., & Sokolov, B. (Eds.). (2019). *Handbook of ripple effects in the supply chain* (Vol. 276). Springer, New York.

Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review, 136*, 101922.

Jaenichen, F. M., Liepold, C. J., Ismail, A., Martens, C. J., Dörsam, V., & Ehrl, H. (2021, December). Simulating and evaluating supply chain disruptions along an end-to-end semiconductor automotive supply chain. In *2021 Winter Simulation Conference (WSC)* (pp. 1-12). IEEE.

Lucic, A., Oosterhuis, H., Haned, H., & de Rijke, M. (2019). FOCUS: Flexible optimalizing counterfactual explanations for tree ensembles. *arXiv preprint arXiv:1911.12199*.

Mönch, L., Fowler, J. W., Dauzère-Pérès, S., Mason, S. J., & Rose, O. (2011). A survey of problems, solution techniques, and future challenges in scheduling semiconductor manufacturing operations. *Journal of Scheduling, 14*(6), 583-599.

Olivares-Aguila, J., & ElMaraghy, W. (2021). System dynamics modelling for supply chain disruptions. *International Journal of Production Research, 59*(6), 1757-1775.

Parmentier, A., & Vidal, T. (2021, July). Optimal counterfactual explanations in tree ensembles. In *International Conference on Machine Learning* (pp. 8422-8431). PMLR.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine learning in Python. *The Journal of Machine Learning Research, 12*, 2825-2830.

Probst, P., Wright, M. N., & Boulesteix, A. L. (2019). Hyperparameters and tuning strategies for random forest. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery, 9*(3), e1301.

Rebala, G., Ravi, A., & Churiwala, S. (2019). Machine learning definition and basics. In *An Introduction to Machine Learning* (pp. 1-17). Springer, Cham.

Rokach, L. (2016). Decision forest: Twenty years of research. *Information Fusion, 27*, 111-125.

Sterman, J. D. (1989). Misperceptions of feedback in dynamic decision making. *Organizational behavior and human decision processes, 43*(3), 301-335.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. McGraw-Hill.

Udenio, M., Fransoo, J. C., & Peels, R. (2015). Destocking, the bullwhisk effect, and the credit crisis: Empirical modeling of supply chain dynamics. *International Journal of Production Economics, 160*, 34-46.

Ventura, F., & Cerquetteli, T. (2019). What’s in the box? Explaining the black-box model through an evaluation of its interpretable features. *arXiv preprint arXiv:1908.04348*.

Verma, S., Dickerson, J., & Hines, K. (2020). Counterfactual explanations for machine learning: A review. *arXiv preprint arXiv:2010.10596*.

Vidal, T., & Schiffer, M. (2020, November). Born-again tree ensembles. In *International conference on machine learning* (pp. 9743-9753). PMLR.

Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology, 31*, 841.

Yagi, D., Nagasawa, K., Irohara, T., Ehml, H., & Yachi, G. (2014). Semiconductor supply planning by considering transit options to take advantage of pre-productions and order cancellations. *Simulation Modelling Practice and Theory, 41*, 46-58.

### Appendix A. Overview parameter variation

| Feature | Range |
|---------|-------|
| Lead time semiconductor company | Consignment stock: products have a lower communicated lead time  <br>Non-consignment stock: products have a higher lead time |
| Cycle time | Complex product: more complex, requires more production steps  <br>Medium product: less complex, requires a decent amount of production steps  <br>Simple product: less complex product requires fewer production steps |
| Desired inventory coverage | Excess inventory: excess +20%  <br>High inventory: additional +10%  <br>Medium inventory: standard scenario  <br>Low inventory: leaner planning (-10%) |
| Forecast adjustment time | Nervous adjustment: lower value than from calibration  <br>Medium adjustment: equals value from the parameter calibration  <br>Smooth adjustment: higher value than from calibration |
| Inventory adjustment time | Nervous adjustment: lower value than from calibration (-20%)  <br>Medium adjustment: equals value from the parameter calibration  <br>Smooth adjustment: higher value than from calibration (+20%) |
| Demand drop shape | V-shape (symmetric)  <br>Italic V-shape (slower recovery)  <br>Inverse italic V-shape (faster recovery) |
| Demand drop length | Demand drop length: maximum value length of assumed COVID-19 scenario; range from 1-52 weeks |
| Demand drop magnitude | Demand drop magnitude: maximum value from COVID-19 scenario; a step size of 1% |