Improvement of Demand Fulfillment in Advanced Planning System through Decentralized Decision Support System⋆

ARTICLE INFO

Keywords:
Advanced Planning System
Supply Chain Planning
Decision Support Tool
Mixed-Integer Programming
Customer Allocation

ABSTRACT

In a tight supply situation, industries such as the semiconductor sector face challenges in allocating supply to customers through complex hierarchical supply chain planning systems, which involve the use of Advanced Planning Systems (APS) and human planners in updating allocation plans. This article focuses on designing, developing, and testing a prototype Decision Support System (DSS) based on a mathematical model to assist human planners and improve demand fulfilment systems in APS when demand exceeds supply. By incorporating digitalization and supporting human interventions in planning, our approach enhances these systems.

The bi-objective mathematical model aims to maximize customer service levels while maintaining a maximum amount of stock available for unforeseen planning situations. We developed the proposed mathematical model into a web application decision support tool, called the Regional Customer Allocation Support Tool (ReCAST). The ReCAST prototype was applied to a semiconductor case study, demonstrating its effectiveness in supporting decision-making processes by planners in a real-world context.

1. Introduction

The semiconductor industry is one of the most global and competitive supply chains with unique and complex challenges. These challenges include long production times, high capital costs, a wide range of products, short product lifetimes, global production processes, production yield, and demand uncertainties. To address these challenges, an effective and mature planning system must be implemented within these global value chains.

Such a planning system should be able to make decisions related to global production targets, schedules for each production line, providing customer predictions upon order receipt, demand fulfilment, order shipments, and consideration of limited resources, while also satisfying the company's overall financial strategy. Furthermore, the planning system must account for the interconnections and precedence of these decisions. In this complex environment, a hierarchical planning system is necessary, taking into account dimensions such as production steps, planning horizon, and product types. To achieve this, semiconductor manufacturing employs an Advanced Planning System (APS) within an Enterprise Resource Planning (ERP) system, encapsulated in an information system (Wiers and de Kok, 2017; Stadtler, Kilger and Meyr, 2015; Guzman, Andres and Poler, 2021).

APS improves three competitive factors, namely cost, quality, and time, and provides a basis for advanced optimization techniques to solve complex decision problems (Stadtler et al., 2015). Within APS, product-to-demand allocation is a contemporary challenge for the global supply chain. This Demand Fulfilment (DF) problem often depends on the lead time within a production system, where products are allocated to customers which is called Available-to-Promise (ATP) (Stadtler et al., 2015; Wiers and de Kok, 2017; Pibernik, 2005; Günther and Meyr, 2009). To achieve synchronization between demand and production plans, APS provides an interactive framework to share information on upcoming demands from the customer side and the Order Promising (OP) department of the production system. The planning system uses Advanced Demand Information (ADI) to improve demand forecasts (Thonemann, 2002) in APS master planning (see Figure 1).

Due to complexity within the global supply chain, the allocation problem in the APS is structured hierarchically according to business requirements (Fleischmann and Meyr, 2003). This decomposition is mandatory due to different planning horizons, planning levels (strategic, tactical, operational), and aggregation/disaggregation of products (Stadtler, 2005). In such a hierarchical decision structure, decisions interact. Thus, all decisions need to be synchronized to facilitate optimal allocation decisions. However, the application of APS is not mature enough in terms of integration to achieve optimal decisions (Lin, Hwang and Wang, 2007).

As shown in Figure 1, the APS receives ADI from the customer to generate demand plans for master planning. Master planning by matching the demand with capacity generates ATP for allocation planning of aggregated products

ORCID(s):

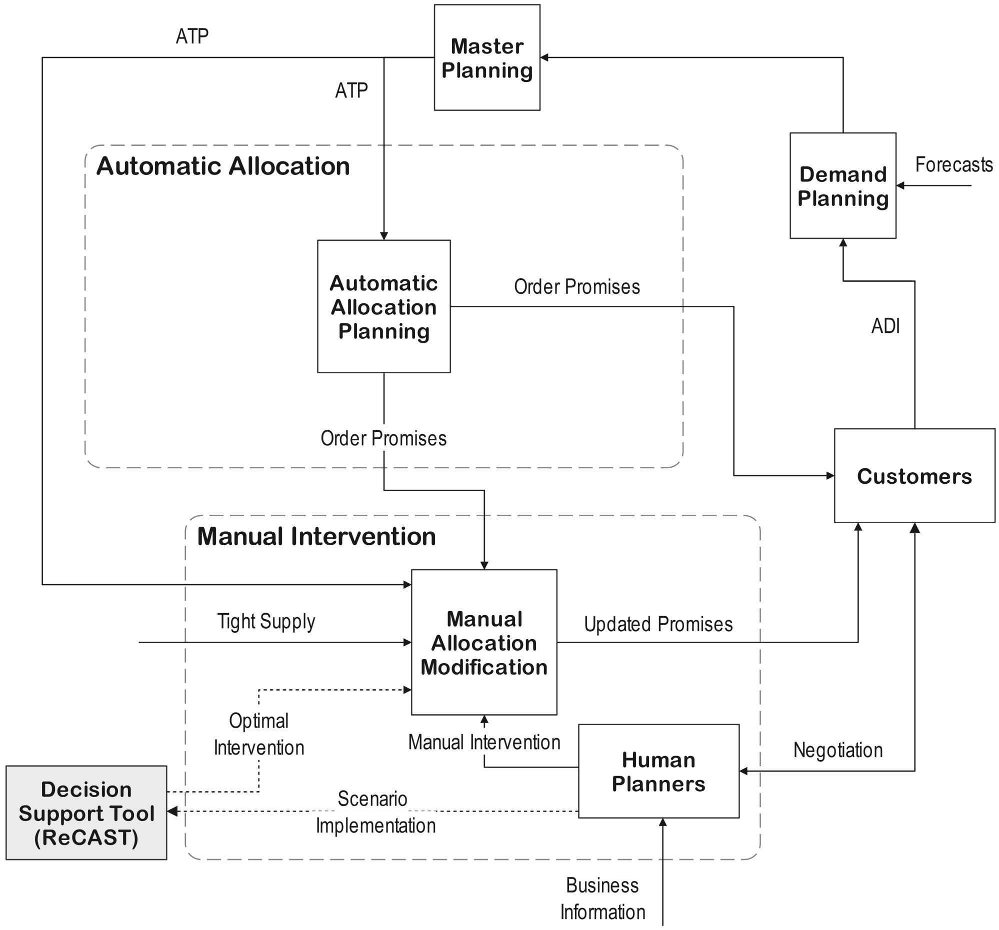


Figure 1: Automatic and Manual Allocation Planning in APS.

between production sites or for stock. Within hierarchical allocation planning, Automatic Allocation Planning (see Figure 1) consumes granulated ATP and sends order promises to end customers. Although ADI and APS support planning, demand and supply uncertainties cause inconsistency within the plans. To deal with uncertainties, APS requires flexibility that must be applied using the intervention of human planners (Moscoso, Fransoo and Fischer, 2008). In practice, existing APS design and application solutions assign expert planners (Wiers and de Kok, 2017) to various planning modules such as capacity planning, production planning, demand forecasting, and order promising (see Manual Intervention in Figure 1).

The role of allocation planners in supply shortages (Tight Supply in Figure 1) is crucial as they need to manually intervene and modify the promises made previously. Planners have a deep knowledge of upcoming uncertainties and business strategies while continuously negotiating with customers to anticipate their expectations (see Human Planners in Figure 1).

To our knowledge, previous studies on allocation planning considered automatic allocation in various forms using mathematical optimization (see Section 2). However, there are gaps in the research assumptions given in the literature. First, they dealt with forecast or ADI uncertainties as entirely unknown or predictable based on the data. They neglect causalities and the chance of decreasing uncertainties by negotiating with customers based on contract terms (Seitz, Akkerman and Grunow, 2022). Second, they consider all uncertainty in product unavailability, while in real cases, changes continuously occur. Third, they ignore the role of planners in plan modifications and updates.

This research investigates the empowerment of allocation planners through a decision support tool powered by mathematical optimization, aiming to fill a research gap in allocation planning and providing support for human planners. As shown in Figure 1, human planners, aware of the business context and customer expectations, can use the proposed DSS to create scenarios and evaluate different optimal allocation plans. The objective of the proposed mathematical model is to maximize customer service levels, directly related to customer satisfaction while maintaining a maximum amount of stock. The results obtained from the mathematical model were validated using real data from the case study company. We developed a decision support tool called the Regional Customer Allocation Support Tool (ReCAST), which was tested as a prototype within the case study company.

Our key contributions are:

- Introduction of a bi-objective mathematical model that maximizes customer service levels while maintaining a maximum amount of stock for unforeseen planning situations, ensuring a balance between customer satisfaction and stock management.

- Development of a prototype Decision Support System (DSS) based on the proposed mathematical model to assist human planners in updating allocation plans when demand exceeds supply, enabling them to make better-informed decisions.

- Design and implementation of the ReCAST, a web application decision support tool that incorporates the proposed mathematical model and facilitates the decision-making process for planners, streamlining their work.

- Application of the ReCAST prototype to a real-world semiconductor case study, demonstrating its effectiveness in supporting decision-making by planners and improving demand fulfillment systems in APS, providing practical evidence of its value.

- Empirical evidence of the benefits of digitalization and supporting human interventions in planning through the use of the developed decision support tool, leading to reduced human errors, decision biases, planning costs, and improved work-life quality for human planners.

- Analysis of the potential for further optimization of planning systems by digitizing and using advanced analytics approaches, leading to shortened response times and improved scalability, offering insights for future development in the field.

- Demonstration of the shifting role of planners from handling calculations and operations to applying insight and business strategies, enabled by the implementation of the proposed decision support tool, illustrating the transformative impact of such tools on the industry.

The remainder of this paper is organized as follows. To familiarize the reader with the use of advanced analytics, operations research, and decision support systems within DF and APS, Section 2 presents a review of the literature in related domains. We discuss the methodology used for this research in Section 3. To provide a description of the case-specific APS and discuss its related demand fulfillment problem, in Section 4, we detail our problem statement. Section 5 discusses the methods and approaches proposed for transferring manual allocation to the developed decision support tool. Details on the development and formulation of combinatorial optimization are given in Subsection 5.1 together with the input data used. The development of the ReCAST DSS is detailed in Subsection 5.2. The results, deployment and tests of the mathematical model and ReCAST are given in Section 6. Finally, in Sections 7 and 8, we discuss the result of the implementation of ReCAST and the feedback collected from the planners within the case study company and conclude the paper.

2. Literature Review

Our research is related to allocation planning, demand fulfilment, supply shortage, hierarchical planning, advanced demand information and support for human planners. In supply chain planning and sequential decision making, a planning system is structured according to decision levels, plan aggregation (hierarchy), planning horizon, and solution approaches (see (Guzman et al., 2021) and (Stadtler et al., 2015) for recent reviews). As described in Section 1, we investigate the standard industrial APS, which hierarchically separates allocation planning using Advanced Demand Information (ADI) (Pibernik, 2005). ADI and industry-4.0-based real-time data sharing offer planning systems to develop finer recursive solutions (Fernandez-Viagas and Framinan, 2022; Thonemann, 2002).

APS provides an upcoming supply in the form of ATP (Framinan and Perez-Gonzalez, 2016) and allocates it through a hierarchical structure to the end customer based on the due date and uncertainties. Hierarchical planning offers the advantages of discrete objectives (such as revenue, time, and service) based on different levels (Vogel and Meyr, 2015). For example, service objectives must be met on the allocation to improve customer satisfaction and optimize the logistic system. In contrast, revenue objectives can be completed at a higher level when aggregated products are allocated to distribution centres.

The main focus of decades of research has been developing mathematical optimization within the various domains of the planning system. However, standard industry solutions still follow simple rule-based algorithms (Guzman et al., 2021) and depend on expert planner manipulation to maintain the flexibility of plans (Wiers and de Kok, 2017; Meyr, 2009). However, we investigate allocation planning solutions under shortage with realistic assumptions for a real case considering industry demands and limitations. The key is not to overwrite all allocation steps with a mathematical model. Instead, what we propose and implement using the insight of human planners by providing them with a DSS based on a mathematical model. Consequently, we review the literature on allocation planning and demand fulfilment in hierarchical APS.

The promising order and the fulfilment of the demand have been investigated according to various structures of the problem. First, the objectives and decision criteria define the goal of allocation planning. Revenue management to increase the profitability of the allocation plan is the main decision criterion discussed in (Quante, Meyr and Fleischmann, 2009; Cano-Belmán and Meyr, 2019; Seitz and Grunow, 2017; Xu and Chen, 2021; Ghomi-Avili, Khosrojerdi and Tavakkoli-Moghaddam, 2019; Papier, 2016; Vogel and Meyr, 2015). Other decision criteria such as customer service level and fairness in delivery (Seitz and Grunow, 2017; Ghomi-Avili et al., 2019) or attempt to control the inventory and delivery system (Topan, Tan, van Houtum and Dekker, 2018; Gayon, Benjaafar and De Véricourt, 2009) are also investigated in the literature.

A second set of researchers investigate the uncertainty of the demand date and how to structure allocation planning. The methods are deterministic demand forecast or ADI, with frequently updated plans Cano-Belmán and Meyr (2019); Seitz and Grunow (2017); Papier (2016) or using imperfect forecast or ADI data with methods to solve uncertainty (Topan et al., 2018; Gayon et al., 2009; Seitz, Grunow and Akkerman, 2020). In this domain, researchers aim to dampen the imperfection of the data by contract management or the precision of the data (Pekgün, Park, Keskinocak and Janakiram, 2019; Seitz et al., 2022).

In Babarogić, Makajić-Nikolić, Lečić-Cvetković and Atanasov (2012), the authors study the allocation of products to customers in a stock-in-stock manufacturing system that is in short supply. During the rolling-horizon planning system, orders of segmented customers are satisfied without accumulation, which means the number of unsatisfied orders does not transfer to the next planning weeks. Additionally, the lowest prior segments met by a limited number and oversupply will be saved in stocks for the following planning weeks. The authors modeled the problem by maximizing the level of customer service, defined as the fraction of customer orders delivered on time. The authors compared the results with the results of rule-based heuristic allocations.

Lečić-Cvetković, Atanasov and Babarogić (2010) dealt with order fulfilment in scarce supply by developing an algorithm to maximize the level of customer service in different groups of customers. The proposed algorithm prioritizes customers of higher importance to the company with total allocations, while only using partial allocations with lower-prioritized customers. Also, they considered back orders in their algorithm. The authors define the level of customer service as the number of satisfied orders and the percentage of promised orders. Although the customer service level was designed to consider long-term business planning, the way customers were classified still followed the revenue management perspective.

Seitz and Grunow (2017) presented a new order promising method to promise orders when products and processes within the supply chain are flexible, customer order lead times are heterogeneous, and demands are uncertain. Their problem was modeled on the basis of a semiconductor manufacturer, where orders are promised online. They considered demand planning and order forecasts as initial steps before promising an order online. These steps help the model cope with changes in production plans resulting from newly arrived orders. In an earlier article Seitz, Grunow and Akkerman (2016) modeled allocation planning in semiconductor manufacturing in which data availability and information sharing were considered at a higher level of granularity (the granularity level defines the level of aggregation/disaggregation within the hierarchical categorization of products or customers). They considered order forecast bias to qualify the data from individual customers. This data exploitation resulted in a better allocation plan, especially for truthful customers.

In a more recent article Cano-Belmán and Meyr (2019) dealt with allocation planning with scarce supply in a multi-

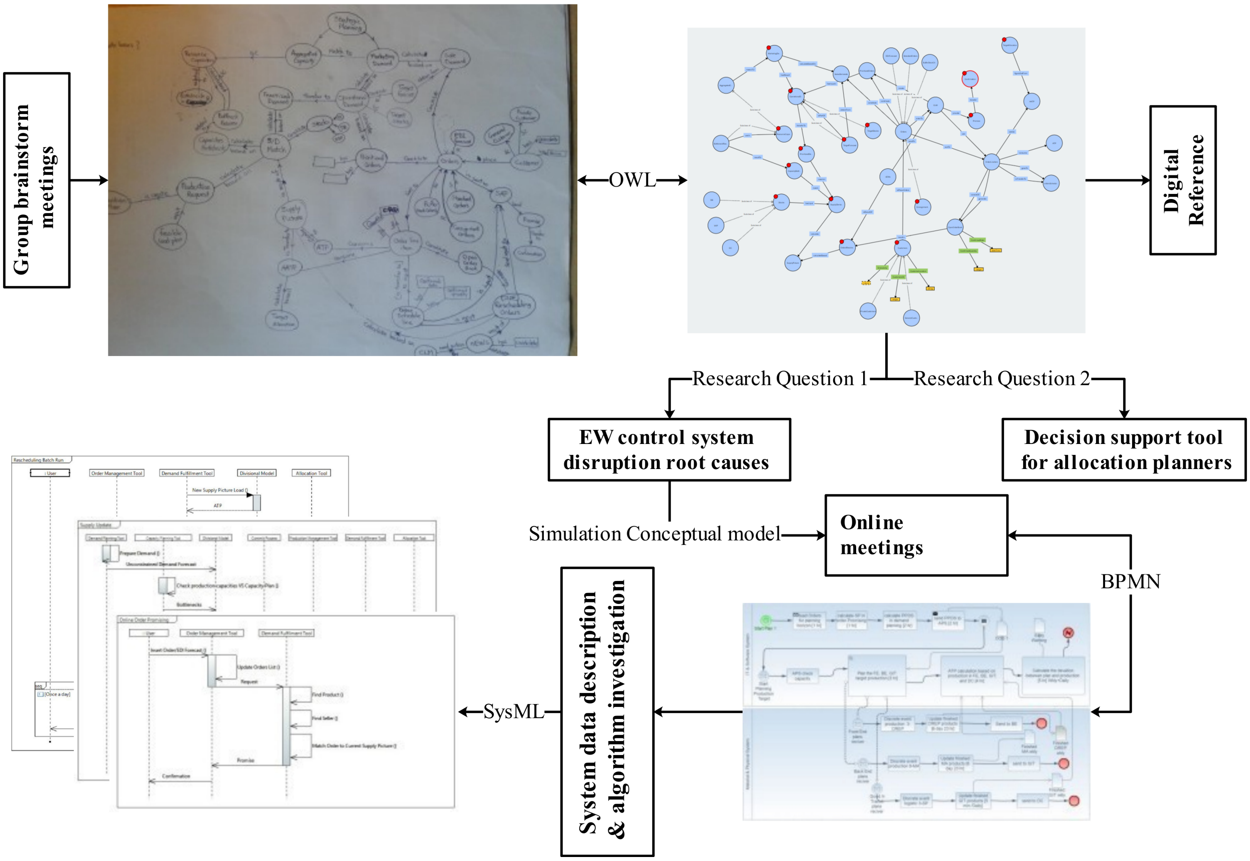


Figure 2: Pathway for implementing and using MBSE approaches for system understanding (Mousavi et al., 2022).

stage customer hierarchy. Central and decentral allocations were evaluated for heterogeneous customers according to their behaviour, location, type of requested products, etc. Other authors added more complexity to encounter uncertainties in order promising such as Grillo, Alemany, Ortiz and Mula (2018) who considered fuzziness and Gössinger and Kalkowski (2015) by considering robustness.

Advanced demand information is not random dynamic imperfect data. Previously, researchers tried to solve the allocation with ADI by applying stochasticity and uncertainty to the data. In actual practice, ADI is not purely arbitrary and some planners take care of customer orders through negotiation, by updating knowledge to end customers about upcoming shortages, supply side obstacles, and increased demands. Therefore, the research conducted in this paper aims to fill a research gap within the literature in the domain of APS and ATP in a hierarchical planning system by considering the role of human planners in the development of solutions to meet demands under shortage. In detail, first, the authors develop research that supports and emphasizes the impact of human decision makers in allocation decisions. Second, it develops a DSS that uses a new combinatorial optimization model for decentralized allocation planning under shortage in the hierarchical planning system. Third, it develops a web-based tool that uses ADI and the prototype tool is tested and evaluated with human planners within the case study company.

3. Methodology

The work described in this article was carried out within the Productive 4.0 project. The overall goal of this project was to digitize manufacturing and SCs in the semiconductor industry. The research reported here was one of the problems addressed in (Mousavi, Heavey, Azzouz, Ehm, Millauer and Knobloch, 2022) where the role of Model-Based System Engineering (MBSE) is highlighted in problem definition and communication with non-experts in quantitative solution methods.

Figure 2 (taken from Mousavi et al. (2022)) describes the pathway of use of MBSE models. In this figure, the hand-drawn diagram was derived and then transferred to the Web Ontology Language (OWL) using Protégé (Protégé, 2020). OWL indicates separable classes (subjects) of the planning system linked by the property of another class or object. These OWL models were found to be useful in communicating our research questions and providing a common language between the stakeholders of the case study company. The output of the OWL models allowed us to define two research questions regarding order management. First, the Early Warning (EW) issue for which we wanted to find the root causes and for which we used simulation (Mousavi, Azzouz, Heavey and Ehm, 2019) and second, an allocation issue with which human planners deal manually, for which we proposed a decision support tool based on operations research. This second problem (see "Decision support tool for allocation planners" in Figure 2) is the focus of this article. To gather further knowledge about the planning system, we use Business Process Model and Notation (BPMN) and Systems Modeling Language (SysML). This information was important for the development of the DSS described in this article, it allows us to better understand the planning system within the case study company.

4. APS and Problem Statement Description

The value chain of the case study company delivers categories of products using a global Supply Chain Network (SCN), which effectively combines the different production sites of the global network. An SC planning system is required to support a global supply chain where flexibility is paramount to enable high capital equipment utilization, which is costly. To manage this SC, a complex sociotechnical system consisting of several subsystems is required, which can be categorized into the following:

- Production or physical system, which are sets of machines and logistics.
- Planning system to maintain plans for the SC system, which differs from the Information Technology (IT) system.
- Expertise system or human planners who are required as "human glue" to manually intervene in SCs, using their cognitive insights.
- An IT system is a set of software modules and databases that interact with humans and integrate with different planning and optimization systems.

In this system, planning is the backbone of the SCN. The strength of an Enterprise Resource Planning (ERP) is well known in the integrated management and transactional backbone of companies (such as human resources, finance, customer service, etc.), but there are planning challenges caused by the complexity and inefficient implementation of ERP, while acknowledging that ERP has useful planning functionality through the use of Material Requirement Planning I (MRP I) and MRP II (Stadtler, 2005). This lack of planning support in ERP belongs to deficiencies of MRP I and MRP II, however, ERP should be seen as an informational backbone of companies. Despite the weaknesses of ERP planning, such systems have been an important enabler for the development of APS, as ERP systems provide the essential data needed by APS. APS are deployed in ERP to fill the planning gap found in ERP (Wiers and de Kok, 2017).

APS is a type of information system with functionality to support planning. APS consists of several software modules, each covering certain tasks in planning and scheduling. APS very often needs to be custom developed specifically based on a company's requirements. APS develops plans based on hierarchical planning using solution approaches such as mathematical optimization, meta-heuristic, and rule-based algorithms. As shown in Figure 3, Stadtler et al. (2015); Stadtler (2005) defines and categorizes the modules and functionalities that are needed to plan and manage SC within the SCP-Matrix. Software vendors customize and deploy some of the APS modules within the ERP system. APS is a set of interactive modules that realize real execution in the supply chain by sending plans or targets (anticipations) and receiving executed actions (i.e., feed-backs). There is no common definition of APS; therefore, we review the definition of APS from a design and implementation point of view.

As defined by Wiers and de Kok (2017), an APS is an interactive planning tool that has a model of the real physical system (implicitly or explicitly) and a set of engines to recalculate the consequences of planning actions immediately. The planning engines run custom-specific algorithms based on the process flows of planning actions to generate plans and schedules. The information required for APS is provided by ERP and stored in random access memory (RAM). Usually, the information is read from different databases according to the design of the process flow of the planning engine. Furthermore, human planners manipulate the plans of an APS using a graphical user interface (GUI).

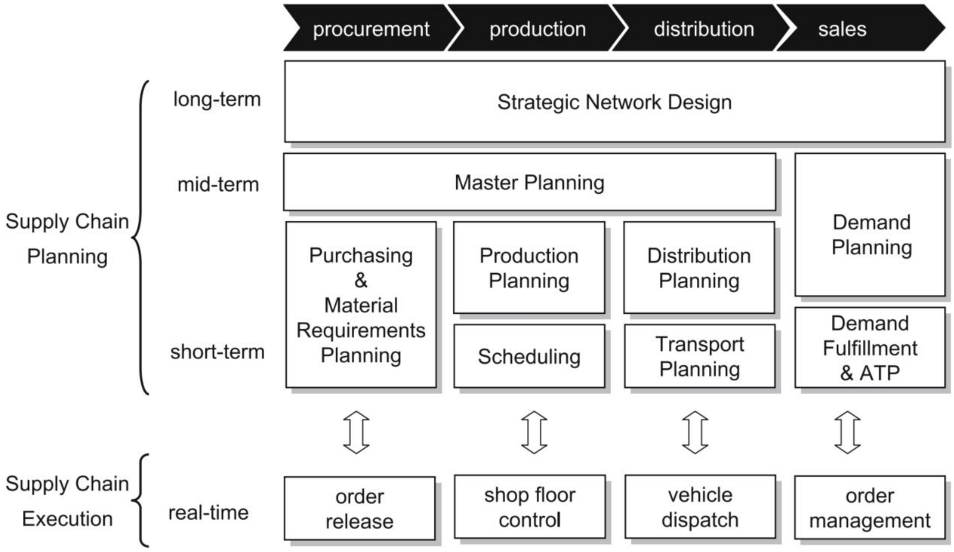


Figure 3: APS module functionalities covering SCP-Matrix obtained from (Stadtler et al., 2015).

The focus of this article is on master planning and DF and ATP functionality within the APS system in the semiconductor case study (see Figure 3). Master Planning aims to match operational demands with capacity bottlenecks to define production planning targets and calculate possible future supplies called ATP. DF in the semiconductor case study aims to provide promises and repromises to customer orders based on ATP and open orders. The main functionalities of APS are handled in the same planning engines with data obtained from integrated databases. Thus, the APS is a centralized decision system with complex process flows which are integrated with databases and the ERP information system.

In the semiconductor case, ATP is generated weekly and is a core aspect of the APS. ATP represents all current and future supplies that are available to fulfill demands. Thus, ATP is the main supply-side input of order management. ATP is consumed by different modules within order management by pegging the ATP to open orders or demands. Pegging does not mean that a segment of ATP is permanently reserved for that demand or order, ATP is rather a method which allows the promising tool within APS to calculate feasible delivery dates for an order (Stadtler et al., 2015).

Although APS consists of several automatic algorithms, it cannot replace human activities. In the case study, the algorithms cover a large number of operations based on formal and simplified planning rules. To perform an end-to-end supply chain planning, human intervention is mandatory. Examples of interventions are the modification of automatic algorithms, handling exceptions, maintenance of plans, management of escalations, modifying APS plans, and other activities. In fact, APS is a central decision support tool for human planners to efficiently manage the planning and scheduling of the entire supply chain. Wiers and de Kok (2017) indicates that human interventions as implicit systems have a large impact on supply chain performance. In this section, we aim to introduce human interventions in order management and allocation within the semiconductor case from its problematic point of view.

In this regard, APS in the case study provides allocation planning for customer orders based on "Available-To-Promise" (ATP). When demand exceeds supply, APS needs the support of experts (human intervention) about the delivery time and amount to be allocated to customers. In the case study planning terminology, it is called an Allocation Situation (AS). This feature of APS maintains the flexibility of planning to make feasible decisions in allocations.

4.1. Problem statement description

Volatility in demand and supply causes instabilities within plans. These instabilities must be mitigated by improving and maintaining APS in a decentralized manner. To deal with instabilities within the DF, a module is used in the Order Management system called Allocation (see Figure 4). The aim of this module is to maintain allocation situations that could not be handled by the Automatic Fulfillment Decision System (AFDS), which requires manual intervention.

The related software is called AM-UI which is a custom-specific software. Therefore, when ATP generates data, it is sent to AM-UI and AFDS. AM-UI has priority over AFDS. Thus, what is consumed from ATP in AM-UI will be denoted as AATP (Allocated-Available-To-Promise) and taken from ATP, with the remaining ATP used in AFDS.

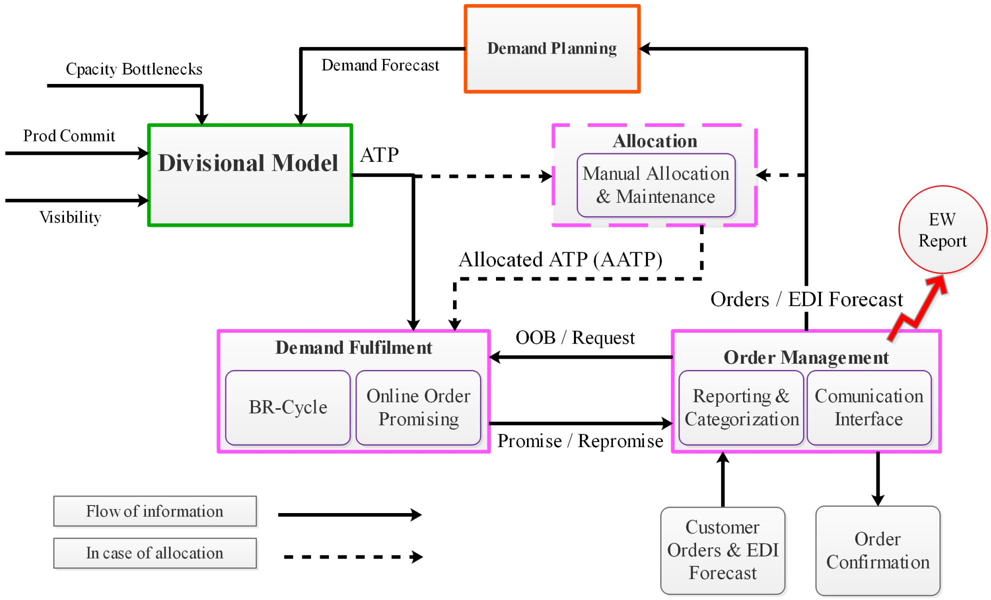


Figure 4: A description of Order Management Modules.

There are two types of allocation within the Allocation module, namely Customer Allocation and Product Allocation, and both are manual allocations. Customer Logistic Management (CLM) is part of the case study order management system. The aim is to develop and manage customer allocations. When demand exceeds supply, the Allocation Manager (ALM) raises a flag, which means that an allocation task is assigned to a Zonal Logistic Planner (ZLP). ZLPs allocate the product in AS to customers. Customer Allocation planning is the area that we focus on in this paper.

Dealing with tight allocations is a daily business of ZLPs. These processes consist of the allocation of one product to many customers for a planning horizon of approximately a year on a weekly basis. For example, if a product needs to be allocated between three customers for forty-five weeks, it means that the ZLP should identify the allocation amount (or at least check). For this manual planning task, they need to check a customer's history of demand, follow the case study's planning strategies according to products on allocation, check the level of ATP for the whole planning horizon, and consider the level of stock and customer orders. In all of these cases, an ZLP should consider customers equally and modify the allocation strategies according to the business situation of the product under allocation. These decision makers require a high level of experience and knowledge.

Based on the investigations, the authors held meetings with planners within the CLM pools in the case study company. A rough estimate (which was not scientifically rigorous) showed that each ZLP deals with customer allocation more than 80 times a month. This manual allocation takes around 30 minutes for an experienced ZLP to be finalized in AM-UI. In the global SC there are 29 ZLPs which handle customer allocation based in the region of the CLM pool. Consequently, more than 1,160 hours of case study human resources are spent on customer allocation per month. It could be implicitly concluded that for each of these allocations, there could be other hidden costs within each manual allocation. For example, costs of teaching and testing of manual allocation for the ZLPs; or cost of intractability or human errors within these processes.

Through meetings with ALMs, ZLPs, business process owners, IT engineers, and SC planning experts, we identified this challenging allocation planning problem. This dependency of APS to planners is necessary, but the way they need to interact manually makes the allocation planning challenging and intractable problems which we aim to solve by automation of allocation planning through a mathematical model and development of a decision support tool. Within this investigation, we developed a mixed-integer programming model to allocate products to customers. The mathematical model developed is used as the core of a decision support tool described in this paper. This decision support tool still needs planners' insights, but it decreases their effort to allocate and improves the quality of the allocation plan. In the next sections, we discuss the developed mathematical model, discuss its performance through validation using real case data, and describe what-if analysis.

5. ReCAST: Description of Mathematical Model and Software Development

Allocation planning in the case study's APS is performed in a hierarchical system in which products or ATP are assigned to the nodes and leaves of the hierarchical tree, where the leaves represent individual customers and the upper nodes represent the aggregation of customers based on different criteria, such as regions, as described by Vogel and Meyr (2015); Cano-Belmán and Meyr (2019). Therefore, the allocation should cover nodes in addition to leaves (customers). In such an environment, the ATP allocated to higher levels (i.e., upper nodes) defines the amount of ATP for each region. The allocated ATP in each node is used to satisfy customers in the region. The ReCAST mathematical model is designed for allocating leaf sellers in CLM pools based on regions.

Allocation of limited resources to demands is the main goal of allocation planning. The allocation of ATP to orders is usually carried out by an automatic algorithm within APS. However, these normal business cycles cannot meet all business situations. The semiconductor case faces situations where demand exceeds available supply. Planners in APS can categorize products for allocation. For example, the allocation situation arises when the performance of the delivery is decreased.

The process of allocation starts with the determination of the general Allocation Situation. Then, according to the type of allocation, the relevant Allocation Logistic Manager (ALM) or ZLPs are assigned to identify an Allocation Situation. After that, the allocation maintenance is performed by the responsible planner according to the defined business logic. These steps require the gathering of data, communication, and decision making by the responsible planner. The result of allocation maintenance is the modification of the quantity or lead times of the orders or the change in targets. The modification needs to be announced to customers.

The responsibilities and roles of the human planner for managing an Allocation Situation are more than what is listed here. However, it is clear that planners are crucial in the management of the supply chain. They are an implicit element within the SC system, where their performance affects its performance. For instance, ZLPs change the amount and time of allocation for specific products by consuming the Gross ATP. In customer allocation, ZLP should allocate the available short supply to the current customer's orders. Human allocation usually has some drawbacks, such as bias decisions between customers, the selection of simplified answers, the inability to select the optimal solution according to a criteria, and the possibility of mistakes. In addition, it is a time-consuming task for ZLPs, which are human capital for the company. Our objective is to provide a decision support tool for allocation planning not only to ease allocation planning for ZLPs, but also to improve the performance of allocations. The decision support tool is based on mathematical optimization and is called the Regional Customer Allocation Support Tool (ReCAST).

5.1. Mathematical formulation

The model formulation was developed to allocate limited supplies to customers in multi-stage allocation planning. The allocation plan will perform this within a time horizon. This model was developed in the context of improving the level of customer service within demand fulfilment. During interviews with experts, business owners, and allocation managers in the order management department, it was revealed that profit is not considered at this level, but products are allocated to improve customer satisfaction. Therefore, profit is considered at a higher level in the planning hierarchy in product allocation. Thus, in an allocation situation executed by ZLPs the aim is to satisfy customers and control stock level. Therefore, the ZLPs consider the following assumptions within their allocation plans:

- Promised date (CMAD) and quantities.
- Orders could be split and satisfied through the allocation horizon.
- Contract clauses for each customer.
- Level of buffer stock that should be kept from ATP weekly.
- Use buffer stock based on business situation.

As discussed, one of the advantages of APS is the availability of data for decision making. AM-UI creates data from an allocation plan which contains ATP and a list of orders for a selected planning horizon, usually a year in weekly time frames. Based on available information and gained knowledge, we proposed a mathematical model. The model aims to develop an allocation plan in the allocation situation. The Mixed Integer Programming (MIP) model aims to increase the level of customer service of allocation plans under the previously presented constraints. Total supply (ATP) and customer demand are deterministic and are known at the beginning of each planning horizon. Given the purpose of the model, these are the parameters and decision variables on which the model is based:

1. Indices
   - i: customers
   - τ: confirmed time
   - t: time
2. Parameters
   - $O_{i,\tau}$: Order is the quantity requested by the customer $i = 1, ..., I$, confirmed at time $\tau = 1, ..., T$. $O_{i,\tau}$ which is known for the whole planning horizon.
   - ATP_t: The Available To Promise (ATP) is known for the whole planning horizon and is equal to ATP_t at each $t = 1, ..., T$. The total demand at time t is usually more than ATP_t.
   - BS_{min_t}: ZLPs specifies the minimum of the cumulative buffer stock, BS_{min_t} which is an input parameter for the model.
   - IBS: Initial Buffer Stock is the level of buffer at the beginning of selected planning horizon.
   - $X_{i,t}$: ZLPs could use buffer stock if customer i follows the lead time for the order at time τ. It is a binary input.
   - RB_t: ZLPs can add more to the buffer stock from ATP_t, which is called Reserved Buffer Stock RB_t. The value of RB_t is also a parameter predefined for each time t and the ZLPs try to reach this level and keep as much as possible from ATP_t, that is not used.
   - λ: The weight defined by the planner for the objective function of allocation(f_1) from ATP.
   - $MaxDelay$; The number of weeks that an order could be delayed.
3. Decision variables
   - $AQ_{i,\tau,t}$: Allocated or promised quantity to customer i in time t refers to an order previously confirmed at time τ. This quantity is consumed from ATP_t.
   - $AS_{i,\tau,t}$: Allocated or promised quantity to customer i in time t refers to an order previously confirmed at time τ. This quantity is consumed from buffer stock. ZLPs are allowed to consume from buffer stock to satisfy the order.
   - AR_t: Allocated quantity from the reserve buffer based on the ZLP goal and the ATP data.
   - BS_t: Buffer stock level for each week.
   - Z_t: Allowance of using from stock for orders in time t. This is a binary variable.

Based on indices, parameters, and decision variables, the mathematical formulation is as follows.

$$\text{Minimize} = -(\lambda \times f_1 + (1 - \lambda) \times f_2) \tag{1}$$

$$f_1 = \sum_i \sum_\tau \left[\sum_{t=\tau} AQ_{i,\tau,t} + \sum_{t\neq\tau} (AQ_{i,\tau,t} \times P) \right] + \sum_i \sum_\tau \left[\sum_t AS_{i,\tau,t} \right] \tag{2}$$

$$f_2 = \sum_t AR_t \tag{3}$$

Subject to:

$$\sum_t AQ_{i,\tau,t} + \sum_t AS_{i,\tau,t} \leq O_{i,\tau} \qquad \forall i \in I, \tau \in T \tag{4}$$

$$BS_t = IBS \qquad for \ \ t = 1 \tag{5}$$

$$BS_t = BS_{t-1} + ATP_{t-1} - \sum_i \sum_\tau AQ_{i,\tau,t-1} - \sum_i \sum_\tau AS_{i,\tau,t-1} \qquad \forall t = 2, 3, ..., T \tag{6}$$

$$\sum_i \sum_\tau AQ_{i,\tau,t} + AR_t \leq ATP_t \qquad \forall t \in T \tag{7}$$

$$\sum_i \sum_\tau AS_{i,\tau,t} \leq BS_t \qquad \forall t \in T \tag{8}$$

$$AR_t \leq RB_t \qquad \forall t \in T \tag{9}$$

$$BS_t \geq BS_{min_t} \qquad \forall t \in T \tag{10}$$

$$\sum_\tau AS_{i,\tau,t} = \sum_t AS_{i,\tau,t} \qquad \forall t \in T, \ \ i \in I, \ \ \tau \in T \tag{11}$$

$$Z_t \leq \left(\frac{\sum_i \sum_\tau AQ_{i,\tau,t}}{ATP_t} \right) + eps \qquad \forall t \in T \tag{12}$$

$$AS_{i,\tau,t} \leq bigM \times X_{i,t} \times Z_t \qquad \forall t \in T, \ i \in I, \ \tau \in T \tag{13}$$

$$P = 1 - \frac{|t - \tau|}{MaxDelay} \tag{14}$$

The objective function in Equation 1 represents a multi-objective optimization problem, aiming to simultaneously maximize two different objectives, f_1 and f_2. Given the conflicting nature of these objectives, the developed decision support tool allows ZLPs to apply various scenarios according to their business goals. These scenarios involve the use of weights (λ) to balance the two functions, enabling ZLPs to manage trade-offs between the competing objectives.

The first objective function, f_1 (as presented in Equation 2), represents the maximization of the level of customer service. It is defined as the sum of the allocated quantities, where the allocated quantity is penalized when $t \neq \tau$ (t represents the current period, and τ represents the previously confirmed time to the customer). The penalty is calculated in Equation 14, which reduces the value of the promised quantity as the promising date moves further away from τ. In essence, the model aims to satisfy demand as close to the previously confirmed time as possible. The second part of this objective function is active only when the customer follows the lead time condition and the Available-to-Promise (ATP_t) quantity cannot cover the requested quantity, resulting in the need to use buffer stock.

The second objective function, f_2 (as presented in Equation 3), models the reserved quantities that the allocation manager aims to maintain. By defining this goal, the allocation manager can effectively control the use of buffer stock and ensure a more efficient allocation process.

The first constraint Equation 4 refers to the total allocation of ATP or stock that must be less than the requested orders. Note that an order could be divided into several parts based on the maximum delay given in Equation 14. The maximum delay or delivery window is a term in the customer's contract within the case study. Equations 5 and 6 aim to control the level of the buffer stock. They refer to the calculation of the buffer stock at time t based on the level of buffer stock at $t-1$ and the amount consumed by ATP and the buffer stock. Furthermore, the total quantity and reserve allocated to ATP and buffer stock at time t should be less than the available promise and buffer stock, which are modeled in Equations 7 and 8. Equation 9 is added to control the level of reserve buffer stock based on the inputs of ZLPs. The buffer stock variable should also be higher than the minimum values defined by the ZLP in Equation 10. Note that this number is different from the Reserved buffer stock that the allocation manager must add to the buffer stock at time $t+1$. Equation 11 describes the allocation from stock that should happen when $t = \tau$. Equation 12 refers to a binary variable Z_t. This variable becomes one when all ATP is consumed by $AQ_{i,\tau,t}$. This binary variable controls the next Equation 13 to force the use from stock when the ATP is finished and there are still available orders. The input binary parameter input by the ZLPs, $X_{i,t}$, is also multiplied in this equation to consider the case study strategies and the goal of the planner.

The model mentioned above gradually evolved through meetings and testing. First, we developed the initial idea in sets of meetings with SC experts and ZLPs. In this step, we were required to understand the decision variables and constraints that ZLP considers implicitly within their cognitive decisions. In addition, what are the case study strategic benchmarks and guidelines required by ZLP to consider which are collected through meetings with ZLPs and stakeholders.

Second, we investigate further the available data. The results of this investigation of the available APS data led to a better understanding of requirements. During a series of meetings, we clarified how ZLPs use these data for their calculation. Note that the decision making between planners may vary based on their understanding of the allocation situation, which makes the understanding of processes challenging.

Third, based on the results of the meetings, available data, and responses to questionnaires, we developed the initial model. In the initial development phase, the MIP model was programmed using the YALMIP toolbox in MATLAB (Löfberg, 2004) with real data from the case study applied to this model. The solver used was GUROBI 9.0 -academic version. The model output was discussed with the allocation business owners and ZLPs, which led to improvements in the mathematical model and a quality improvement in the allocation plans was developed.

During debugging and testing of the initial model in the fourth iteration, we found a missing variable in the mathematical model. Adding this new variable was considered implicitly by ZLPs. This variable is AR_t which changes the shape of the model. This decision variable was examined by ZLPs. After they certify the usefulness of the addition of this new decision variable, we modified the model. This new variable allows us to divide the total allocation to each customer into two decision variables $AQ_{i,\tau,t}$ and $AS_{i,\tau,t}$ which are the quantities allocated from ATP and stock, respectively. Previously, an ZLP considered these two calculations in one output from the model.

Fifth, we tested the new data in the new model. When the quality of the solutions was certified by ZLPs, the capabilities of the model were proved to SC experts. Thus, a mathematical model was developed for use within a decision support tool. Thus, for further improvements to the model, we moved the model to Python. In the next sections, we discuss the developed tool (ReCAST).

5.2. ReCAST development

This subsection will present the details of a prototype web-based decision support tool, developed using the mathematical model presented in the previous section. Planners may not be familiar with mathematical optimization; therefore, our aim was to develop ReCAST as an easy-to-access and user-friendly tool for planners. The motivation of this project is to improve allocation planning by replacing manual calculations with an optimized support tool. Note that this tool will not replace the role of a planner, but will change their role from the need to do manual allocation to optimally allocating based on the objective function. The developed tool is named the Regional Customer Allocation Support Tool (ReCAST).

In the following, the steps and processes for designing and developing ReCAST are detailed, which are:

- **Requirement Analysis:** to define the basic structure of the tool.

- **Design:** this will describe the steps involved in the design of the web-based tool.
- **ReCAST:** the steps involved in the development of this tool.

Results of the mathematical optimization, test, and deployment of ReCAST are presented in the next section.

5.2.1. Requirement Analysis and Design

We conducted a requirements analysis for the ReCAST tool. In the requirements specification, we focus on the analysis of a user type, expected user tasks, user interfaces, user interactions, and system performance. The purpose of the requirement analysis is to define the structure and scope of the project. These specifications were used to build a simple prototype for meeting with stakeholders to validate their requirements. Project managers, ZLPs, developers, and SC experts provided feedback during these meetings, which helped us modify some of the user-interaction operations and business processes based on the previous version of the prototypes we presented. The three main objectives of the ReCAST prototype are:

- The aim is to obtain the optimal customer allocation to be used within AM-UI as benchmarks by ZLPs during allocation planning.
- The optimal solutions generated by ReCAST are based on the ZLP input configurations. These configurations come from the insights that ZLPs have or gain from marketing, product types, business situations, etc. Thus, ReCAST should receive these configurations from the end user.
- ReCAST should have a convenient and easy-to-use design. ZLPs are required to manually import data from a local Excel file, add configuration data, run the mathematical optimization model, select the proper scenario from the generated mathematical solutions, and finally export from ReCAST the allocation plan to AM-UI as a structured Excel file.

The process flow of ReCAST is described in Figure 5. This figure shows that when a Target Allocation flag is raised, the ZLPs are notified that an allocation situation is required and they need to develop a new allocation plan. Allocation plans are manually added using the AM-UI module. When they feel that the allocation situation is complex enough that they need ReCAST, they export the allocation situation to an Excel file. The exported file is then imported to ReCAST. ReCAST reads the data and parses them to a compatible format to be used in the mathematical model. For this, it requires input from ZLPs which are the selection of planning horizon, their strategies by identifying the weights of scenarios, delivery windows, packing unit of product on allocation, minimum buffer quantity, and their goal of keeping ATP as a reserve buffer stock. After entering the inputs, the ReCAST program is run. The model displays scenarios on a new page using visualization tools. These visualizations are designed to facilitate the process of selecting scenarios. After selecting the most relevant scenario, the ZLPs exports the results. Now another parser transfers the results to an Excel file that is compatible with AM-UI. The user then needs to import the ReCAST result into the AM-UI tool.

5.2.2. User groups

In ReCAST, the primary user is an ZLP. In addition, an Allocation Logistic Manager (ALM), a Customer Logistic Manager (CLM), or a Supply Planner may use ReCAST. These users would be defined as high-level users in ReCAST, and this needs to be taken into consideration in the design of the tool. Currently, we have successfully validated the business logic for ZLP. As the software is designed for internal use, there is no requirement for high-level performance because the number of such users is low. Users of ReCAST have also stated that data stored beyond three months could be erased, as the planning horizon moves and historical data will not be required.

There are three states for an ZLP when they use ReCAST (see Figure 6):

- State A: Not logged in state: At Login page, Register page, Forget Password.
- State B: Logged in state: At Index page, Initial Scenario page (Import Data from Excel file), Task Information page. Including All pages at state C.
- State C: Working state: At Config Optimal Model page, ReCAST Analysis Result page, Scenario Modification page.

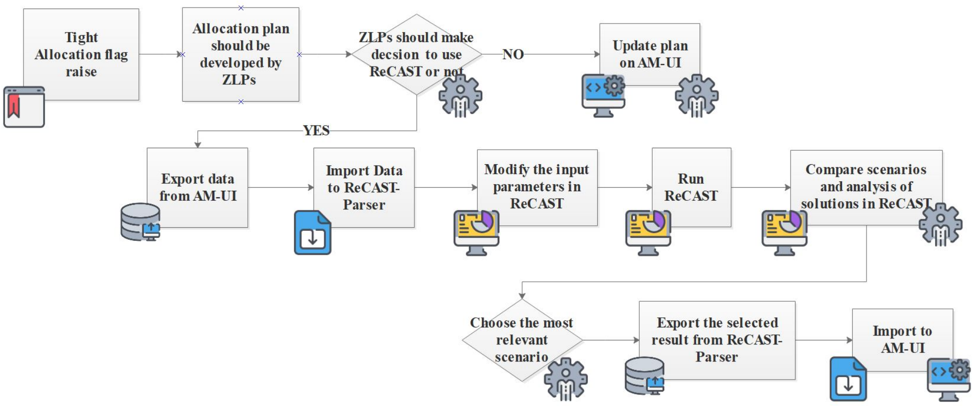


Figure 5: Process flow of using ReCAST in the case study APS.

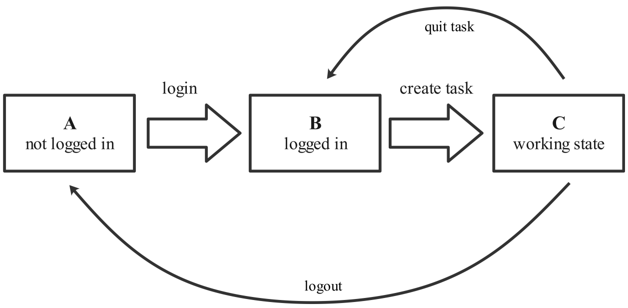


Figure 6: ReCAST user's states and transits.

The transits between states are:

- A to B: From not logged in state to logged in state, by triggering a login event.
- B to C: By creating a task for running ReCAST. A task can be considered as a business logic flow to generate allocation plan scenarios.
- C to B: By quitting the task or finishing the task, the ZLP is currently doing.
- C to A: Logout.

5.2.3. Functional Requirements

This subsection provides information on all known ReCAST functions and services. It is more relevant to specific functional descriptions related to business logic and all user interactions between the system. Use case diagrams are essential to describe functional requirements, based on the actor of the system, showing the interaction of users in the system, and also representing the relationship between each function. Each circle in Figure 7 represents a use case for a system boundary that describes a usage scenario context to capture the requirements for user interaction with the system. In this figure, there are two types of relationships, extended and include. Extend is used when a use case adds steps to another first-class use case. Include is used to extract use case fragments that are duplicated in multiple use cases. The included use case cannot stand alone, and the original use case is not complete without the included one. The main use cases are described as follow:

- **Extract Excel Data to Content (Parser1)**: The use case is used for importing data from an AM-UI Excel file. First, ZLPs upload the file and input parameters manually; then the file with the parameters will be posted to the

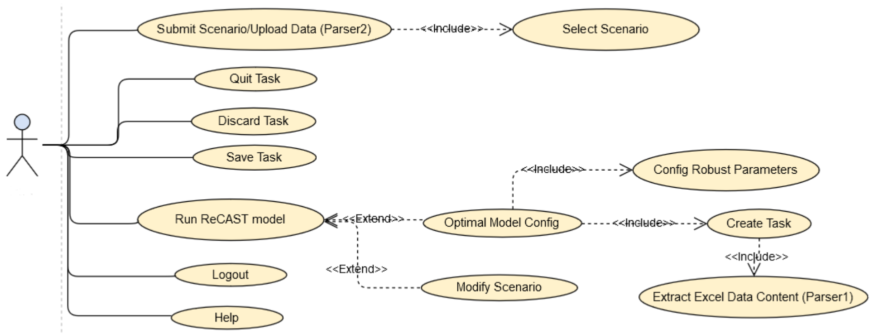


Figure 7: Functional requirement.

server (AWS). Note that the word server here means Amazon Web Service (AWS), which is the server on which the ReCAST was developed. The server then checks the validity of the file and the parameters and extracts all the data from that file using an Excel file parser.

- **Create Task**: When an Excel file is uploaded to the server, users should define a task in ReCAST. The aim of the task definition is to avoid confusion between various allocation plans. Also, assigned task names and descriptions are useful for storing and retrieving data within the database.
- **Submit Scenario/Upload Data (Parser2)**: The use case is used to export selected scenario data to an Excel template file for upload to AM-UI. After showing all the results of the scenario, ZLP will select a scenario and then click 'Export'. The browser will send data to the server, and the server will run Parser2 to generate an Excel file compatible with AM-UI. The server will then send the generated Excel file to the browser to be downloaded by ZLPs. Moreover, the server will write the results and product information to the database for convenient use for the next time.
- **Scenario Selection**: ZLPs select the most appropriate scenario based on the initial input data. In this use case, some visualization features are considered to facilitate scenario selection by ZLPs.
- **Run ReCAST model**: There are two ways for users to run the ReCAST optimization program, from the Configuration page and from the Modification page, but no matter which approach is used, they will always trigger a click event to run the optimization model.
- **Modify Scenario**: On the 'Modification' page, the table of scenario information can be modified by a ZLP. Allocated from ATP (A-ATP) and Allocated from Stock (A-Stock) are editable, but the sum is a fixed number. (There is a relationship between these three parameters: A-ATP + A-Stock = Total-Allocation.) In addition, the user can click the 'Check' browser to post data to the optimizer server to run the ReCAST optimization model. This then returns results to be displayed. If the user does not click 'Check', the data will not be updated. Note that this feature was not developed in the prototype delivered to the company.
- **Optimal Model Config**: Each individual problem requires a modification of the input parameters. In this use case, the ZLPs define the configuration for the mathematical model.

5.2.4. ReCAST: A web based tool

ReCAST aims to be accessible in a browser using the Internet, which makes the application widely available throughout the case study company as a decentralized DSS. Thus, ReCAST is a web-based application. The website is divided into three modules: front-end, back-end, and server.

The front-end is implemented using HTML, CSS, and JavaScript; Website back-end was developed using the Django framework with the Python language; Server connectivity is provided by Amazon Web Service (AWS). The Django built-in server uWSGI was used for development. Most of the front-end user interaction is implemented using the Javascript framework Jquery.js and the style design is implemented using the CSS framework Bootstrap. Figures 8 and 9 represent the website and technical modules used within ReCAST.

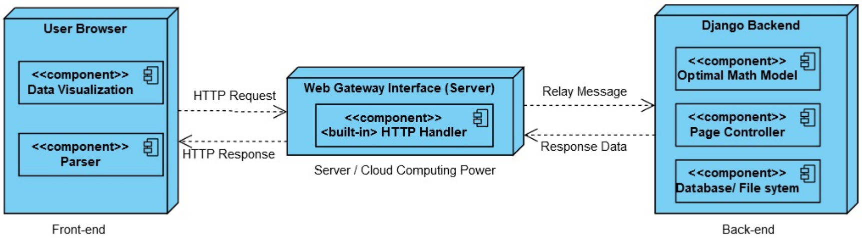


Figure 8: ReCAST system modules.

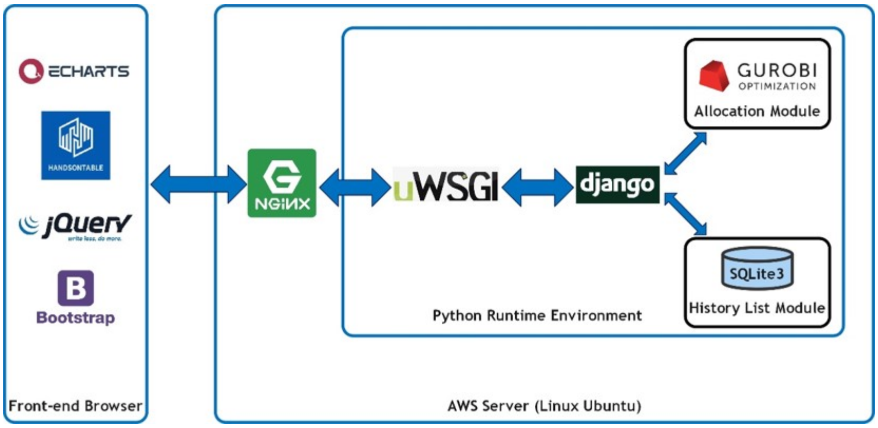


Figure 9: Technical modules.

The data visualization module provides ZLP with visualization of scenarios so that they can be compared. This was implemented using Echarts.js which is developed using JavaScript. The parser is the input and output module of the system that provides functions for processing Excel. This uses two frameworks, HandOnTable.js to implement the display effect in the front end, and the Pandas framework is used to generate the Excel file in the back end. Django's built-in uWSGI can be used as a server to quickly implement access functions by combining with cloud computing platforms, which use the HTTP access request processing function. Optimal Math Model provides the core business algorithm to generate allocation plan scenarios by invoking the Gurobi mathematical API interface. Page Controller defines a webpage jump relationship (route logic) in the Django framework to manipulate which pages should be visible in ReCAST. The database module provides system persistence capability; the user history task records all data, which is provided by Django with its built-in access to SQLite. More details describing the developed ReCAST prototype are presented in Appendix B.

6. ReCAST Implementation

6.1. Mathematical model results

Within the modeling phase of the project, we extracted five real allocation cases. This data set was used for the development of the model, and further testing of new data was performed, which will be discussed in the next section. Here, we use part of this data to examine the mathematical model (see Appendix A Table 2).

The different combinations of weights between f_1 and f_2 are presented in Table 3 and Figure 10. The applied examples contain two customers with 49 weeks of a planning horizon. Note that both functions are similar in units, thus the summation of weighted objective functions would not cause any issue.

Table 3 shows 11 scenarios that start from high to low weights for allocation to customers. When the weight value is one, it shows that the model prefers to allocate to customers first rather than keeping reserve goals. In this table, the total available ATP and the total requested order are also identified. The sum of the allocation from ATP and from stock defines the allocation quantities which will be shipped to customers each week. Finally, the optimal values of each objective are the sum of rows 5 and 6 which gives row 7 (see Table 3). These values are shown in Figure 10. This figure shows the relation between the values of the objective function in different scenarios based on several combinations of weights. When the weight of f_1 is equal to zero, the model prefers to keep all ATP as a reserve buffer stock and nothing is allocated to customers. When the weight of f_1 is less than 50%, the two objective functions become equal. This example was tested based on input parameters that show the effects of the multi-objective on the final solutions.

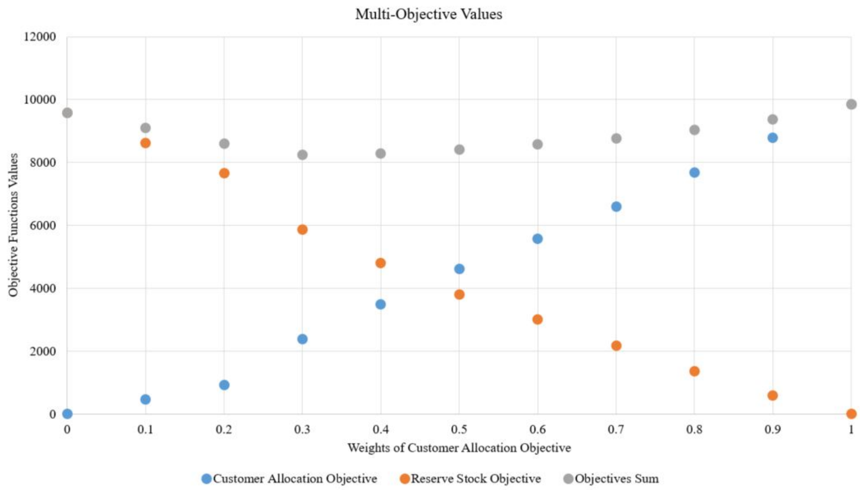


Figure 10: Comparison of multi-objective values between scenarios.

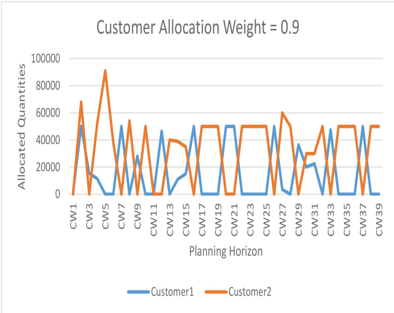
The model was tested in various scenarios that demonstrated the competency of the model and how it depends on the input of the ZLPs. In other words, the model can develop a solution based on a business situation rather than a ZLP's preference. Figure 11 visualizes the allocation plan for four selected scenarios with a selected planning horizon of 39 weeks shown by CW (Current Week). Figure 11a shows the quantities allocated to each customer every week. It shows the allocation plan when the customer weight is 0.9 and the stock weight is 0.1. The quantities allocated to customers one and two at week 17 are equal to 0 and 50000. The comparison between these figures shows the effect of weighted objectives on the quantities allocated to each customer. Comparing Figure 11a with Figure 11c, shows that the model with 0.9 customer allocation weight allocates more quantities to customer two between weeks 1 and 5. However, in Figure 11c a major part of the allocation to customer one occurs in week 6.

6.2. What-If Analysis

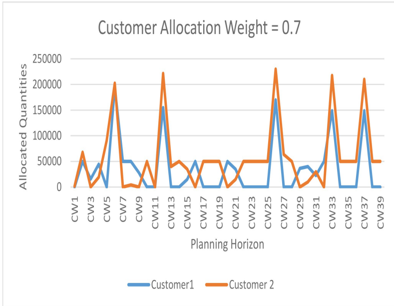
The mathematical model presented solves the allocation problem and provides exact optimal solutions. However, in real cases, it is possible that specific business conditions need to be considered. For example, a customer requests a specific quantity in a critical week for its business. To solve this exception, ZLPs might modify one or two specific weeks. In these situations, ZLPs require a mathematical model to calculate the optimal plans, but might need to modify the optimal solution for a small number of weeks (possibly one or two). For these modifications, ZLPs usually need to use product from stock. To keep the optimal solution and modification flexible, we proposed a heuristic approach for analyzing the effects of using from stock or not for ZLPs.

To provide this analytical tool for ZLPs, we performed a what-if analysis on the model. Since the model is MIP, duality theory could not be applied. Thus, we could not calculate the dual values using solvers. To address this issue, we fixed the binary variable of the optimal solutions to change the model to LP. Then we change the value of the binary variables one by one and calculate the gradient of the objective function. By comparing the value of the binary fixed objective with the global optimal (normal objective), we could see the effects of the change in a plan. The concept of the heuristic algorithm for this what-if analysis is similar to duality since it aims to calculate the cost of change in one of the variables. The pseudocode of the what-if analysis is presented in Algorithm 1. This is similar to calculating the dual variable of binary constraints. The plot of this analysis is presented in Figure 12.

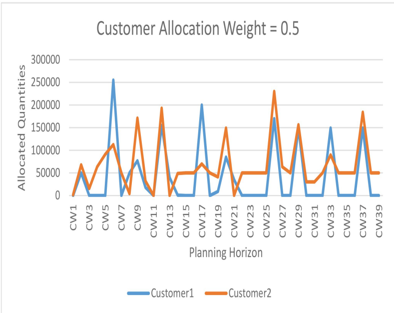
The result of this what-if analysis presents a plot for ZLPs to be used as a benchmark for decision making regarding the modification of using from stock or not. As ZLPs are only allowed to use the product from stock when ATP is consumed, this analysis can support them in understanding if they need to modify the optimal solution of ReCAST.



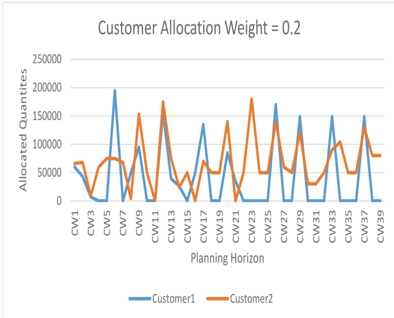
(a) Customer Allocation weight = 0.9.



(b) Customer Allocation weight = 0.7.



(c) Customer Allocation weight = 0.5.



(d) Customer Allocation weight = 0.2.

Figure 11: Comparison of allocation plans based on different scenarios.

This plot supports them in comparing the impact of using or not using from stock in a specific week to select the most suitable weeks for changing the plan of using from stock. As an example, the output of the use of Algorithm 1 which is presented in Figure 12 shows the what-if analysis of using stock for 49 weeks of a planning horizon (49 weeks) in four scenarios, where different customer weights are used. Each point in this figure represents one run by changing Z_t^* in that week from the optimal binary variable to its opposite value (see Algorithm 1).

The y axis of this plot shows the normalization of this difference and the negative values show how much the new model loses in comparison to the optimal model. For instance, the optimal values of the binary variable in week 17 and 18 for Scenario 4 are both zero, which means the model does not use from stock in these weeks. However, if the ZLPs wants to use from stock in this week and now wants to decide which week is the most suitable, then week 17 is recommended, as it has a lower consequence on the optimal value. This means that using from stock in week 17 causes a lower loss in the optimal value of the function.

6.3. ReCAST test and deployment

After the initial development of ReCAST, its performance was tested. Thus, we performed two sets of tests, internal tests, and tests with ZLPs. Within the internal tests, we used five real cases from the case study and tested ReCAST with various combinations of inputs and scenarios. Table 1 [1] shows an example of a test table. This table aims to

[1] SP defines the product type, PU shows the packing unit value of the product, RB_Shift indicate whether the user changes the 'Reserve Buffer goal' or all are the same value, and MB_Shift indicates whether the user changes the 'Minimum Buffer_Stock' or all are the same values.

Algorithm 1 What-If Analysis.

Z_t^* is the optimal solutions of binary variable.
$f_1^* + f_2^*$ is the optimal value of objective functions.
Constraints 12 and 13 are the binary-constraints related to binary variable Z_t
for All scenario weights i **do**
  for All the planning horizon j **do**
    if Z_j^* in secenario i is equal to 0 **then**
      change the value to 1
      set the changed variable as a parameter
    else if Z_j^* in secenario i is equal to 1 **then**
      change the value to 0
      set the changed variable as a parameter
    end if
    Change the binary-constraints by replacing the variable Z_j with identified
    parameter in If statement above.
    Optimize the model with the new binary-constraints.
    Calculate the differences between the new optimal value with new binary
    -constraint and the global optimal $((f_{1,bin-fixed}^* + f_{2,bin-fixed}^*) - (f_1^* + f_2^*))$.
  end for
  Calculate these differences for all scenarios
  Plot the normalized differences for scenarios, in whole planning horizon
end for

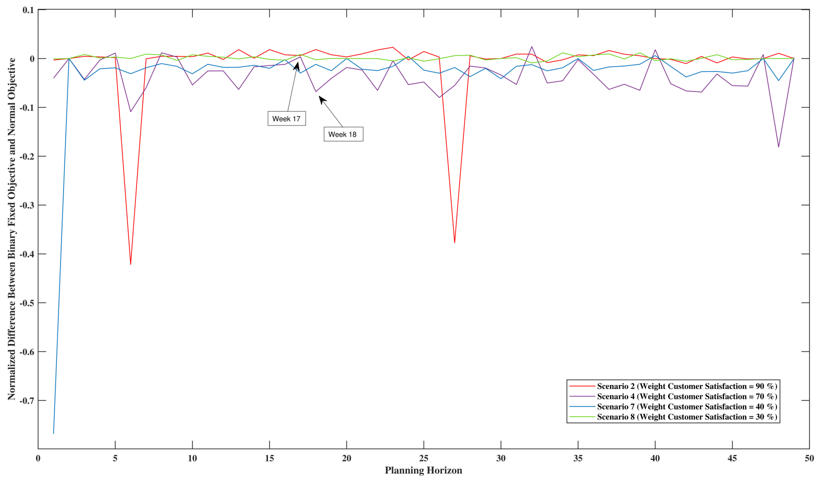


Figure 12: What-If analysis of using from stock or not.

capture input data, add configurations, and identify bugs. In this example, we identify two issues. First, there was an issue with the parser data to select the right current week (CW). The number of weeks in each year is 52 or 53. When the selected CW starts with bigger values and then goes to the next year (for instance, from week 50 in 2019 to week 25 in 2020), the developed module for the parser misses some data. Second, the minimum buffer stock constraint also makes the model infeasible in some cases. This bug was also identified in this example, which led to the reformulation of the mathematical model.

To perform industry tests, we conducted the following steps:

- **Organized and gathered users**: We selected and communicated with planners and SC experts. In the selected

Table 1
Internal testing of ReCAST.

SP #	**Date**	**Test Date**	**Case #**
product SP	181218	200917	01-Mar
Mode	**Consideration**		**Customer #**
Normal	Test the initial stock value		3
PU	**Customer Weights**	**CW**	**Initial Stock Value**
500	0.9	50-25	535500
Max Delay	**Reserve Buffer goal**	**RB_ Shift**	**Minimum Buffer_Stock**
10	200000	No	100000
MB_Shift	**All Use of Stock = 1**	**Bug**	**Rationality of Solutions**
No	Yes	YES	Infeasible

team, there were two ZLPs from Asia, two ZLPs from Europe, one ZLP from the US, one ZLP coordinator and the allocation process owner. This team supported us in examining ReCAST.

- **Introduced ReCAST within two online meetings**: We held a couple of meetings with the selected team. In these meetings, we introduced ReCAST and provided a complete how-to-use tour. In addition, a ReCAST user manual and examples were sent to the ZLPs. The general results of these meetings showed that ReCAST is easily understandable to users. They could grasp the concepts and the definition of parameters. It showed that ReCAST's allocation decision variables were defined in alignment with the planners' cognitive decision processes. However, further suggestions were also collected that could be considered in future steps of the development of the platform.

- **Supported ZLPs during testing**: During testing, ZLPs dealt with issues regarding the definitions of some variables which were resolved by holding online meetings and sending information by email. ZLPs in Asia could not reach the web-application due to network security in the Asia office, which was resolved by using a different network. Consequently, all the testing teams tested ReCAST completely. Note that due to confidentiality agreement, the ReCAST prototype did not store any test examples used by ZLPs to test the prototype.

- **Held feedback meetings to collect ZLP experiences and suggestions**: Two meetings were held to collect feedback. In these meetings, we received their suggestions and discussed the quality of solutions. These are discussed in the next section.

7. Discussion

APS can be improved by applying decentralized mathematical operation research methods. Planning systems are believed to always need human planners and applied mathematics can improve their decision-making. In advanced planning systems, the main functionalities are handled centrally. The central system cannot be fully replaced by mathematical optimization due to the required flexibility. However, the availability of APS data can support decentralized mathematical optimization to improve integrated decisions.

ReCAST reads parameters from specific AM-UI data, and planners' input was based on their understanding of the business situation. In addition, it generates plans based on scenarios defined by planners. In ReCAST, planners must define seven types of input variables. Within the rest of this section, we will discuss the received feed-backs and suggestions.

- The definitions of input by planners should be clarified in a comprehensive and understandable way.

- The logic behind the ReCAST is good, and it would reduce the training required for CLMs/ZLPs as it provides a structured approach to an ill-defined problem area.

- At the current level, ReCAST is a working prototype and needs further development to become a final product. To achieve this goal, ZLP testing of the tool, discussions with IT experts, and recommendations from planning experts are required.

- ReCAST is user-friendly and is easy to use.
- In summary, the ZLPs suggested that a tool to automate the current allocation process would be beneficial. A tool that combines manual and automatic allocation such as ReCAST would considerably improve the as-is process.

In conclusion, the ReCAST prototype was tested by planners and feedback was obtained. This prototype shows how the case study could further optimize their planning system by digitizing and using advanced analytics approaches. Response times could be shortened, and scalability greatly improved. Furthermore, it clarifies how the planner's roles can shift from handling calculations and operations to applying insight and business strategies. Within the testing phase, it has been calculated that ZLPs can develop an allocation plan approximately six times faster by using ReCAST, rather than manually.

8. Conclusion and Future Studies

Operations Research (OR) analytics require to be complemented with a final product that delivers a decision support to the end user within a company. Data analysis and visualization are two key elements required to connect OR analytics to end users in a complex planning system like APS. As demonstrated here, these decentralized decision support tools can integrate well with a centralized planning system (APS), which can lead to more efficient planning, as shown here. This approach is more maintainable. In addition, it changes the role of planners from performing repetitive tasks to more analytical decision making that requires human cognitive capabilities.

In this paper, we proposed a mathematical model for customer allocation in APS. The model is a multi-objective MIP model, which was tested on real data from industry, which showed its competency and capabilities. In the mathematical element of the solution, we presented further investigation regarding possible what-if analysis conducted on the model to better support decision makers.

In addition, we discussed the development steps of a decision support tool called Regional Customer Allocation Support Tool (ReCAST). ReCAST is a web-based application with an OR core that aims to support planners in the calculation of customer allocation plans. The tool was developed as a prototype version and tested with industry users. The results of the test and deployment supported the approach obtained and showed benefits.

In summary, ReCAST empowers planners that require daily allocation planning with advanced mathematical techniques and digitalization. ReCAST facilitates the decision-making process, digitizes customer allocation planning, reduces human errors, avoids human decision biases, increases the quality of allocation plans, reduces planning costs, and improves work-life quality of human planners.

A. Case Study Inputs

In this table, the planning horizon is from $CW1$ to $CW49$ (CWi Current Week i) and there are two customers ($C1$ and $C2$). Reserve buffer stock, which is an important goal of ZLPs to obtain from stock, is denoted by RB_t . We consider this value constant in this example, but ZLPs can modify it based on their understanding of business situations. The Initial Buffer Stock (IBS) in this example is 535500. Another parameter defined by ZLP is the minimum buffer stock, denoted by BS_{min_t}. These values are considered constant in this example based on ZLPs ideas, but in real examples ZLPs can modify them based on their requirements. X C1 and X C2 are representative of the $X_{i,t}$ parameter, which defines whether the customers are allowed to use from stock ($X_{i,t} = 1$) or not ($X_{i,t} = 0$). In this example, all customers are allowed to use stock. Similarly to the previous parameters, ZLPs can modify these values.

As mentioned, the proposed mathematical model is multi-objective. To solve it, we used the weighted sum approach, where the objective functions are aggregated by multiplying them with weights and summing over them. The sum of weights is equal to one. The weights are the input of the planners. In this example, we used 11 combinations of weights:
[[1,0],[0.9,0.1],[0.8,0.2],[0.7,0.3],[0.6,0.4],[0.5,0.5],[0.4,0.6],[0.3,0.7],[0.2,0.8],[0.1,0.9],[0,1]]

Another parameter that needs to be defined by ZLP is the packing unit, which is related to the product type on allocation. In this example, the packing unit is 500. The last input parameter is $MaxDelay$. This parameter identifies how many weeks the model is allowed to postpone and split in an order. The maximum delay is used in the calculation of the penalty function in the mathematical model. The output table discussed above is presented in Table 3.

B. ReCAST Prototype Demonstration

ReCAST is a decision support tool that aims to integrate with the use of APS, providing planning data that the proposed mathematical model can use. Thus, ReCAST is not only an optimization model, but also a web application made up of various technologies to allow easy input of data by ZLPs. ReCAST reads allocation data from the case study's APS, parses data, receives a planner's input, obtains insight and planning strategies from the case study, provides visualization, generates solution scenarios, and finally exports the selected result in the format readable by the case study's APS, through the AM-UI software module. The optimization model and web application can handle any type of allocation situation and input data. In this section, we describe the functionalities and features of the ReCAST prototype.

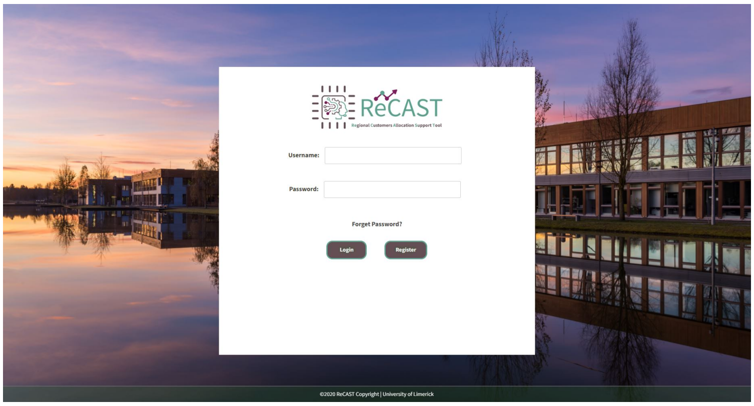


Figure 13: ReCAST login page.

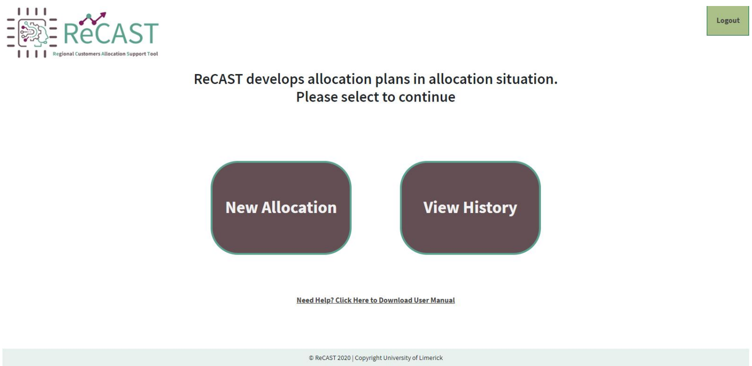


Figure 14: ReCAST index page.

Each planner has a user id and password. They can log into ReCAST from anywhere using an internet connection (see Figure 13). On the first page, they can check the history of their allocation or create a new allocation. The development of a new allocation plan is as follows. When the allocation situation occurs, the ZLPs can download the task as an Excel file. They should upload the Excel file of the allocation situation from their local computer (see Figure 14).

By uploading a file, they can review the content of the allocation situation and define a new task in ReCAST. They choose the planning horizon of the required allocation. For example, in this case, from week number 25 in 2020 to week number 20 in 2021 (see Figure 15). As customer allocation is defined based on product type, planners also need to identify what the packing unit of that product is in the logistic system. Finally, in this page, they create scenarios according to the case study's strategy for the product on allocation. They identify the weights between allocating more products to customers and keeping the level of stock. This feature lets planners define different scenarios to check various strategies and solutions (see Figure 15).

After uploading the file to ReCAST and defining the new task, planners should identify the allocation limitations, goals, and allowances of using a stock product; the screen shown in Figure 16 is shown. This shows that each order could be allocated to different portions of the planning horizon. Max Delay is the maximum number of weeks that the last part of the order could be satisfied. Minimum Buffer stock identifies the lowest level of stock. The planner could simply apply their ideas to the entire planning period or modify week by week (see Buffer Stock Allocation in Figure 16). Reserve buffer stock also shows the goal of keeping product from ATP and reserving it for stock. As can be seen, this value could also be applied globally for all periods in the planning horizon or could be modified weekly. Finally, the planner could restrict the model to allocate some customers in some weeks. By default, all customers are considered similar and all are allowed to use product from stock. However, in some cases, the allowance of using a product from stock needs to be modified by planners (see Figure16).

After a couple of seconds, the MIP programming model described is run and then ReCAST generates and visualizes solutions. Input data and visualization (see Figures 17 and 18, respectively) were developed to simplify understanding to the planner. Users can compare different customers according to scenarios. In the scenario table (see Figure 19), they can see the allocation plans developed by ReCAST based on the defined scenarios.

After comparing the results between scenarios, the planner can select and export the most appropriate allocation plan based on their insights. The exported file is compatible with the destination software module. Thus, users upload the ReCAST output to the advanced planning system. It will automatically update the whole planning system.

The prototype version of ReCAST aims to show the overall functionality of the allocation planning task. Note that the robust model in this figure is an incomplete part of the development of a robust optimization model for the instability of ATP. The robust optimization could be considered as a future extension of the decision support tool.

References

Babarogić, S., Makajić-Nikolić, D., Lečić-Cvetković, D., Atanasov, N., 2012. Multi-period Customer Service Level Maximization under Limited Production Capacity. International Journal of Computers Communications & Control 7, 798–806. doi:10.15837/ijccc.2012.5.1075.

Cano-Belmán, J., Meyr, H., 2019. Deterministic allocation models for multi-period demand fulfillment in multi-stage customer hierarchies. Computers & Operations Research 101, 76–92. doi:10.1016/j.cor.2018.09.002.

Fernandez-Viagas, V., Framinan, J.M., 2022. Exploring the benefits of scheduling with advanced and real-time information integration in industry 4.0: A computational study. Journal of Industrial Information Integration 27, 100281.

Fleischmann, B., Meyr, H., 2003. Planning Hierarchy, Modeling and Advanced Planning Systems, in: Handbooks in Operations Research and Management Science. Elsevier. volume 11 of *Supply Chain Management: Design, Coordination and Operation*, pp. 455–523. URL: http://www.sciencedirect.com/science/article/pii/S0927050703110092, doi:10.1016/S0927-0507(03)11009-2.

Framinan, J.M., Perez-Gonzalez, P., 2016. Available-to-promise systems in the semiconductor industry: A review of contributions and a preliminary experiment, in: 2016 Winter Simulation Conference (WSC), IEEE. pp. 2652–2663.

Gayon, J.P., Benjaafar, S., De Véricourt, F., 2009. Using imperfect advance demand information in production-inventory systems with multiple customer classes. Manufacturing & Service Operations Management 11, 128–143.

Ghomi-Avili, M., Khosrojerdi, A., Tavakkoli-Moghaddam, R., 2019. A multi-objective model for the closed-loop supply chain network design with a price-dependent demand, shortage and disruption. Journal of Intelligent & Fuzzy Systems 36, 5261–5272.

Grillo, H., Alemany, M.M.E., Ortiz, A., Mula, J., 2018. A Fuzzy Order Promising Model With Non-Uniform Finished Goods. International Journal of Fuzzy Systems 20, 187–208. doi:10.1007/s40815-017-0317-y.

Guzman, E., Andres, B., Poler, R., 2021. Models and algorithms for production planning, scheduling and sequencing problems: A holistic framework and a systematic review. Journal of Industrial Information Integration , 100287doi:10.1016/j.jii.2021.100287.

Gössinger, R., Kalkowski, S., 2015. Robust order promising with anticipated customer response. International Journal of Production Economics 170, 529–542. doi:10.1016/j.ijpe.2015.05.026.

Günther, H.O., Meyr, H. (Eds.), 2009. Supply chain planning: quantitative decision support and advanced planning solutions. Springer, Berlin. OCLC: ocn339390372.

Lečić-Cvetković, D., Atanasov, N., Babarogić, S., 2010. An Algorithm for Customer Order Fulfillment in a Make-to-Stock Manufacturing System. International Journal of Computers Communications & Control 5, 783–791. doi:10.15837/ijccc.2010.5.2238.

Lin, C.H., Hwang, S.L., Wang, E.M.Y., 2007. A reappraisal on advanced planning and scheduling systems. Industrial Management & Data Systems doi:10.1108/02635570710822822.

Löfberg, J., 2004. Yalmip : A toolbox for modeling and optimization in matlab, in: In Proceedings of the CACSD Conference, Taipei, Taiwan.

Meyr, H., 2009. Customer segmentation, allocation planning and order promising in make-to-stock production. OR Spectrum 31, 229–256. doi:10.1007/s00291-008-0123-x.

Moscoso, P.G., Fransoo, J.C., Fischer, D., 2008. Human Planners, Planning Structures and the Planning Bullwhip. SSRN Scholarly Paper ID 1499310. Social Science Research Network. Rochester, NY. URL: https://papers.ssrn.com/abstract=1499310.

Mousavi, B.A., Azzouz, R., Heavey, C., Ehm, H., 2019. Simulation-Based Analysis of the Nervousness within Semiconductors Supply Chain Planning: Insight from a Case Study, in: 2019 Winter Simulation Conference (WSC), IEEE, National Harbor, MD, USA. pp. 2396–2407. URL: https://ieeexplore.ieee.org/document/9004936/, doi:10.1109/WSC40007.2019.9004936.

Mousavi, B.A., Heavey, C., Azzouz, R., Ehm, H., Millauer, C., Knobloch, R., 2022. Use of model-based system engineering methodology and tools for disruption analysis of supply chains: A case in semiconductor manufacturing. Journal of Industrial Information Integration 28, 100335.

Papier, F., 2016. Supply allocation under sequential advance demand information. Operations Research 64, 341–361.

Pekgün, P., Park, M., Keskinocak, P., Janakiram, M., 2019. Does forecast-accuracy-based allocation induce customers to share truthful order forecasts? Production and Operations Management 28, 2500–2513.

Pibernik, R., 2005. Advanced available-to-promise: Classification, selected methods and requirements for operations and inventory management. International journal of production economics 93, 239–252.

Protégé, 2020. URL: https://protege.stanford.edu/.

Quante, R., Meyr, H., Fleischmann, M., 2009. Revenue management and demand fulfillment: matching applications, models and software, in: Supply Chain Planning: Quantitative Decision Support and Advanced Planning Solutions. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–32. doi:10.1007/978-3-540-93775-3_3.

Seitz, A., Akkerman, R., Grunow, M., 2022. A contract portfolio perspective on the role of customer order lead times in demand fulfilment processes with supply shortage. International Journal of Production Research 60, 3792–3808.

Seitz, A., Grunow, M., 2017. Increasing accuracy and robustness of order promises. International Journal of Production Research 55, 656–670. doi:10.1080/00207543.2016.1195024.

Seitz, A., Grunow, M., Akkerman, R., 2016. Data Driven Supply Allocation to Individual Customers Considering Forecast Bias. SSRN Electronic Journal doi:10.2139/ssrn.2813835.

Seitz, A., Grunow, M., Akkerman, R., 2020. Data driven supply allocation to individual customers considering forecast bias. International Journal of Production Economics 227, 107683.

Stadtler, H., 2005. Supply chain management and advanced planning-basics, overview and challenges. European journal of operational research 163, 575–588. doi:10.1016/j.ejor.2004.03.001.

Stadtler, H., Kilger, C., Meyr, H., 2015. Supply Chain Management and Advanced Planning: Concepts, Models, Software, and Case Studies. 5th ed., Springer-Verlag, Berlin Heidelberg. doi:10.1007/978-3-642-55309-7.

Thonemann, U.W., 2002. Improving supply-chain performance by sharing advance demand information. European Journal of Operational Research 142, 81–107.

Topan, E., Tan, T., van Houtum, G.J., Dekker, R., 2018. Using imperfect advance demand information in lost-sales inventory systems with the option of returning inventory. IISE Transactions 50, 246–264.

Vogel, S., Meyr, H., 2015. Decentral allocation planning in multi-stage customer hierarchies. European Journal of Operational Research 246, 462–470. doi:10.1016/j.ejor.2015.05.009.

Wiers, V.C., de Kok, A.T.G., 2017. Designing, Selecting, Implementing and Using APS Systems. Springer.

Xu, L., Chen, J., 2021. Real-time order allocation model by considering available-to-promise reserving, occupying and releasing mechanisms. International Journal of Production Research 59, 429–443.

Table 2
Sample data for optimization model.

	Week	ATP	Order C1	Order C2	RB_t	BS_{min_t}	X C1	X C2
1	CW1	101500	60000	68000	100000	400000	1	1
2	CW2	0	50000	68000	100000	400000	1	1
3	CW3	115500	0	68000	100000	400000	1	1
4	CW4	163500	0	0	100000	400000	1	1
5	CW5	191000	0	83000	100000	400000	1	1
6	CW6	138000	195000	183000	100000	400000	1	1
7	CW7	150000	50000	83000	100000	400000	1	1
8	CW8	154000	50000	0	100000	400000	1	1
9	CW9	128000	45000	80000	100000	400000	1	1
10	CW10	150000	0	180000	100000	400000	1	1
11	CW11	93500	0	0	100000	400000	1	1
12	CW12	146500	155000	0	100000	400000	1	1
13	CW13	140000	50000	70000	100000	400000	1	1
14	CW14	150000	15000	70000	100000	400000	1	1
15	CW15	150000	0	0	100000	400000	1	1
16	CW16	150000	135000	0	100000	400000	1	1
17	CW17	150000	50000	70000	100000	400000	1	1
18	CW18	150000	0	80000	100000	400000	1	1
19	CW19	150000	0	80000	100000	400000	1	1
20	CW20	150000	85000	80000	100000	400000	1	1
21	CW21	150000	35000	80000	100000	400000	1	1
22	CW22	150000	0	80000	100000	400000	1	1
23	CW23	150000	0	80000	100000	400000	1	1
24	CW24	150000	0	0	100000	400000	1	1
25	CW25	150000	0	80000	100000	400000	1	1
26	CW26	150000	170000	80000	100000	400000	1	1
27	CW27	163500	0	60000	100000	400000	1	1
28	CW28	150000	0	130000	100000	400000	1	1
29	CW29	136500	149000	40000	100000	400000	1	1
30	CW30	150000	0	30000	100000	400000	1	1
31	CW31	152500	0	30000	100000	400000	1	1
32	CW32	150000	0	70000	100000	400000	1	1
33	CW33	147500	149000	70000	100000	400000	1	1
34	CW34	150000	0	104000	100000	400000	1	1
35	CW35	150000	0	100000	100000	400000	1	1
36	CW36	150000	0	70000	100000	400000	1	1
37	CW37	150000	149000	60000	100000	400000	1	1
38	CW38	150000	0	80000	100000	400000	1	1
39	CW39	150000	0	80000	100000	400000	1	1
40	CW40	150000	0	76000	100000	400000	1	1
41	CW41	163500	0	76000	100000	400000	1	1
42	CW42	136500	149000	76000	100000	400000	1	1
43	CW43	150000	0	76000	100000	400000	1	1
44	CW44	150000	0	76000	100000	400000	1	1
45	CW45	162500	0	0	100000	400000	1	1
46	CW46	137500	149000	0	100000	400000	1	1
47	CW47	159500	0	0	100000	400000	1	1
48	CW48	150000	0	0	100000	400000	1	1
49	CW49	150000	0	0	100000	400000	1	1

Table 3: Comparison of results using different objectives' weight values.

No.	Senarios (f_1 Weights)	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
1	Summation of ATP	7081000	7081000	7081000	7081000	7081000	7081000	7081000	7081000	7081000	7081000	7081000
2	Summation of Orders	4925000	4925000	4925000	4925000	4925000	4925000	4925000	4925000	4925000	4925000	4925000
3	Allocated from ATP (AQ)	4176000	3878500	3554000	3315000	3210500	3171500	3029000	2829000	2287500	2287500	0
4	Allocated from Stock (AS)	749000	1046500	1371000	1610000	1714500	1753500	1749000	1442500	118000	118000	0
5	Customer Allocation Objective	9848	8788.68	7678.8	6594.49	5568.48	4613.1	3496.52	2383.02	926	463	0
6	Reserve Stock Objective	0	582	1354.2	2178	3003.6	3793.5	4792.2	5870.9	7669.6	8628.3	9587
7	Summation of Objectives	9848	9370.68	9033	8772.49	8572.08	8406.6	8288.72	8253.92	8595.6	9091.3	9587

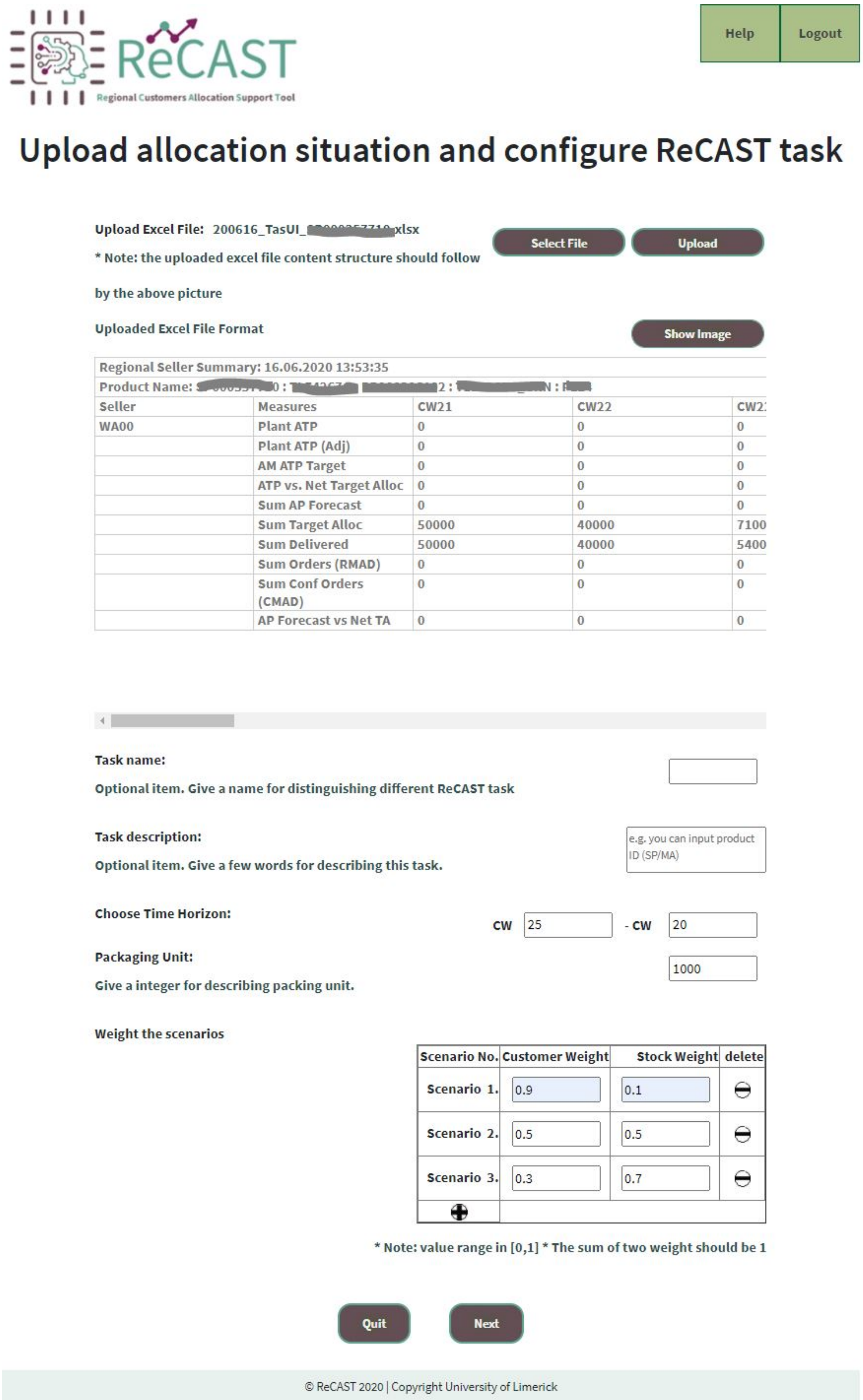


Figure 15: ReCAST new allocation task creation page.

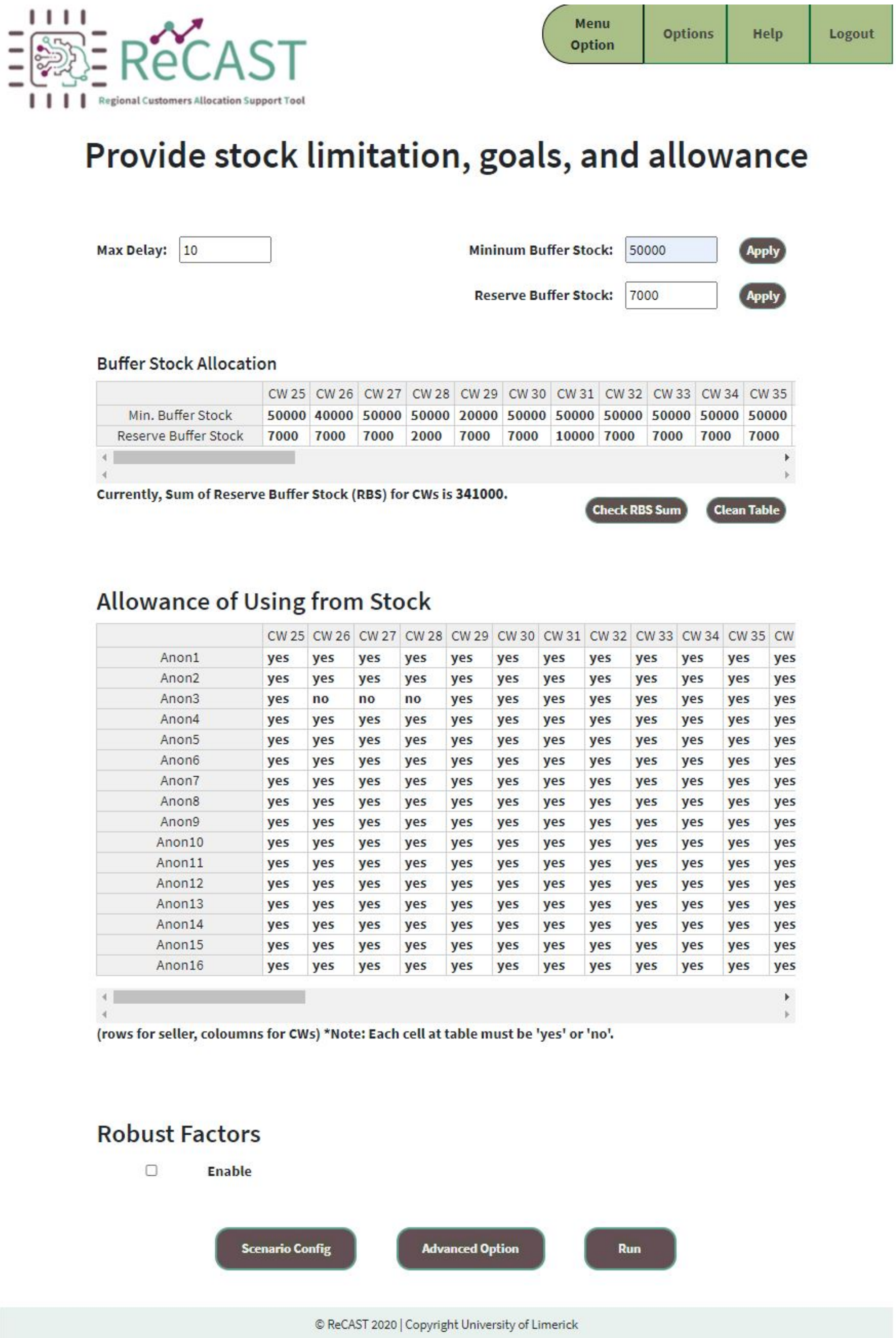
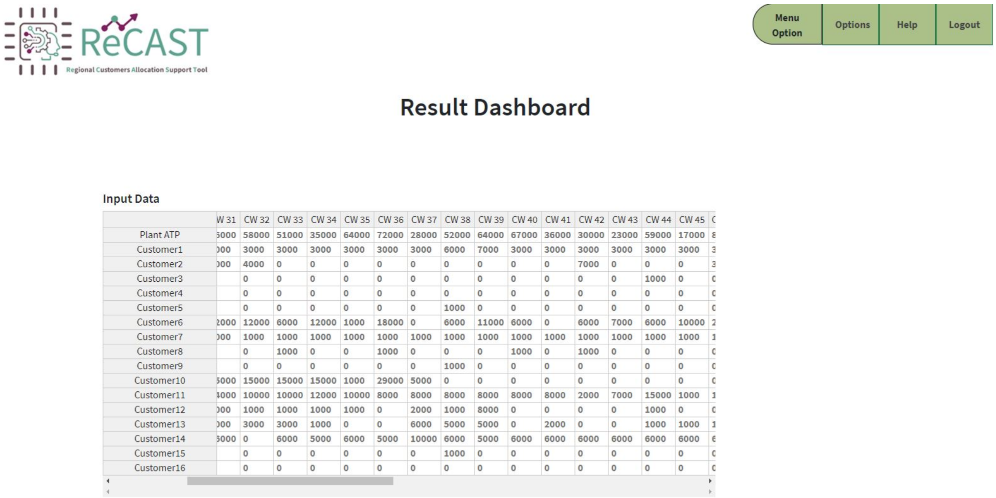
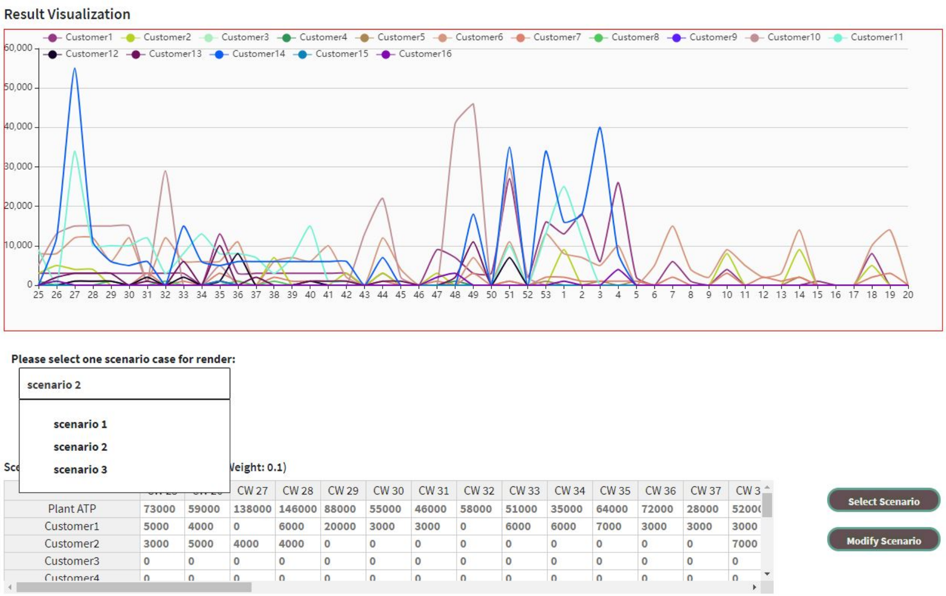


Figure 16: ReCAST business situation inputs and configuration page.



	W 31	CW 32	CW 33	CW 34	CW 35	CW 36	CW 37	CW 38	CW 39	CW 40	CW 41	CW 42	CW 43	CW 44	CW 45	C
Plant ATP	5000	58000	51000	35000	64000	72000	28000	52000	64000	67000	36000	30000	23000	59000	17000	8
Customer1	000	3000	3000	3000	3000	3000	3000	6000	7000	3000	3000	3000	3000	3000	3000	3
Customer2	000	4000	0	0	0	0	0	0	0	0	0	7000	0	0	0	3
Customer3		0	0	0	0	0	0	0	0	0	0	0	0	1000	0	0
Customer4		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer5		0	0	0	0	0	0	1000	0	0	0	0	0	0	0	0
Customer6	2000	12000	6000	12000	1000	18000	0	6000	11000	6000	0	6000	7000	6000	10000	2
Customer7	000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1
Customer8		0	1000	0	0	1000	0	0	0	1000	0	1000	0	0	0	0
Customer9		0	0	0	0	0	0	1000	0	0	0	0	0	0	0	0
Customer10	5000	15000	15000	15000	1000	29000	5000	0	0	0	0	0	0	0	0	0
Customer11	4000	10000	10000	12000	10000	8000	8000	8000	8000	8000	8000	2000	7000	15000	1000	1
Customer12	000	1000	1000	1000	1000	0	2000	1000	8000	0	0	0	0	1000	0	0
Customer13	000	3000	3000	1000	0	0	6000	5000	5000	0	2000	0	0	1000	1000	1
Customer14	5000	0	6000	5000	6000	5000	10000	6000	5000	6000	6000	6000	6000	6000	6000	6
Customer15		0	0	0	0	0	0	1000	0	0	0	0	0	0	0	0
Customer16		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 17: ReCAST result page part 1 of 3.



			CW 27	CW 28	CW 29	CW 30	CW 31	CW 32	CW 33	CW 34	CW 35	CW 36	CW 37	CW 3
Plant ATP	73000	59000	138000	146000	88000	55000	46000	58000	51000	35000	64000	72000	28000	5200
Customer1	5000	4000	0	6000	20000	3000	3000	0	6000	6000	7000	3000	3000	3000
Customer2	3000	5000	4000	4000	0	0	0	0	0	0	0	0	0	7000
Customer3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Customer4	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 18: ReCAST result page part 2 of 3.

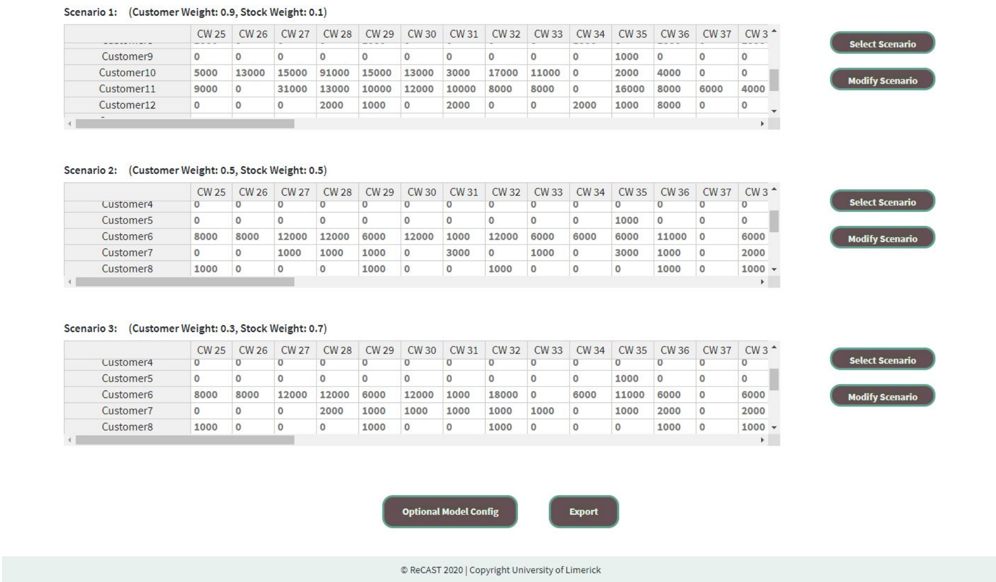


Figure 19: ReCAST result page part 3 of 3.