Pontifícia Universidade Católica
do Rio de Janeiro

Andréa Regina Nunes de Carvalho

Tactical capacity planning in an ETO production setting using optimization models: A real-world industrial context

TESE DE DOUTORADO

Thesis presented to the Programa de Pós-Graduação em Engenharia de Produção of the Departamento de Engenharia Industrial, PUC-Rio as partial fulfillment of the requirements for the degree of Doutor em Engenharia de Produção

Advisor: Prof. Luiz Felipe R. R. Scavarda do Carmo
Co-Advisor: Prof. Fabricio Carlos Pinheiro de Oliveira

Rio de Janeiro
September 2015

PONTIFÍCIA UNIVERSIDADE CATÓLICA
DO RIO DE JANEIRO

Andréa Regina Nunes de Carvalho

Tactical capacity planning in an ETO production setting using optimization models: A real-world industrial context

Thesis presented to the Programa de Pós-Graduação em Engenharia de Produção of the Departamento de Engenharia Industrial do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor.

Prof. Luiz Felipe R. R. Scavarda do Carmo
Advisor
Departamento de Engenharia Industrial - PUC-Rio

Prof. Fabricio Carlos Pinheiro de Oliveira
Co-Advisor
Departamento de Engenharia Industrial - PUC-Rio

Prof. Eduardo Galvão Moura Jardim
Universidade Federal do Rio de Janeiro - UFRJ

Profa. Laura Silvia Bahiense da Silva Leite
Universidade Federal do Rio de Janeiro - UFRJ

Prof. Leonardo Junqueira Lustosa
Independent consultant

Prof. Reinaldo Morabito Neto
Universidade Federal de São Carlos - UFSCar

Prof. Silvio Hamacher
Departamento de Engenharia Industrial - PUC-Rio

Prof. José Eugenio Leal
Coordinator of the Centro Técnico Científico da PUC-Rio

Rio de Janeiro, September 25th, 2015

Andréa Regina Nunes de Carvalho

Andréa Regina Nunes de Carvalho obtained her undergraduate degree in Industrial Engineering at Universidade Federal do Rio de Janeiro and holds an M.Sc. degree in Industrial Engineering from Pontifícia Universidade Católica do Rio de Janeiro. She has recently published in the "International Journal of Production Research" and the "International Journal of Production Economics". Currently, she is a researcher at Instituto Nacional de Tecnologia (INT), where she has been a leader in computer simulator projects aimed to support production planning in several industrial organizations. She has worked on the specification, development, training and implementation of finite capacity simulators in large companies such as Moto Honda, Siemens, Philips, Sony, CSN, Xerox, CCE, Furukawa, FICAP, Petrobras, Fundição Tupy, Abnc, AGFA, among others. Additionally, she teaches in the area of production planning and logistics at MBA level at Fundação Getúlio Vargas (FGV), Sociedade Educacional de Santa Catarina (SOCIESC) and Centro de Ensino Empresarial (CEEM).

Bibliographic data

Carvalho, Andréa Regina Nunes de

Tactical capacity planning in an ETO production setting using optimization models: a real-world industrial context / Andréa Regina Nunes de Carvalho ; advisor: Luiz Felipe Roris Rodriguez Scavarda do Carmo ; co-advisor: Fabricio Oliveira. – 2015.
117 f. : il. (color.) ; 30 cm

Tese (doutorado)–Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Industrial, 2015.
Inclui bibliografia

1. Engenharia Industrial – Teses. 2. Engenharia sob encomenda. 3. Planejamento agregado de produção. 4. Sistema de apoio à tomada de decisão. 5. Programação matemática. 6. Otimização robusta. I. Carmo, Luiz Felipe Roris Rodriguez Scavarda do. II. Oliveira, Fabricio. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Industrial. IV. Título.

CDD: 658.5

Acknowledgements

These four and a half years at PUC-Rio have meant a lot to me and I am so thankful to everybody who helped make this a memorable experience. I consider myself a very lucky person to have had Prof. Luiz Felipe Scavarda and Prof. Fabricio Oliveira as my advisor and co-advisor, whose combined expertise were essential for the development of this thesis, and to have had the opportunity to address a real-life problem with the support of an industrial management team, who opened the doors and gave me all the assistance I needed throughout this research.

Many people contributed to this thesis and I would like to thank some of them in particular. In the first place, I would like to express my sincere gratitude to my thesis advisor, Prof. Luiz Felipe Scavarda, who has provided me with invaluable advice and encouragement and who has been a continuous source of support and guidance throughout the process of developing this thesis. It has been a great honor to work with him on this research and I hope we will be able to continue this in the coming years. I would also like to thank my co-advisor, Prof. Fabricio Oliveira, whose idea of using robust optimization techniques enhanced greatly our research study. Fabricio encouraged me to push further the boundaries of what I thought I could do using optimization models. Every meeting was a fruitful experience that yielded relevant insights. It has been most pleasant to work with him.

This research could have never been done without the support of the industrial management team of the studied company. They shared their planning problems and highlighted the shortcomings within their planning process. They helped me to define the main assumptions for the developed models and validated this work in many circumstances. They were very enthusiastic with the possibility of enhancing their planning method through optimization models. They are truly co-authors in this research.

I am indebted to Prof. Silvio Hamacher, who encouraged me from the beginning to pursue a doctoral degree at PUC and who introduced me to mathematical optimization. I am very grateful for the opportunity to have had classes with Silvio on practical applications of mathematical models. This formed a starting point for this research. Many thanks are due to Prof. Leonardo Lustosa for his interest in my work. This thesis benefited greatly from his comments in other phases of this research process. I thank Prof. Eduardo Jardim, who has been a role model to me ever since I met him and whose enthusiasm and energy are truly exceptional. Although we met less during the last years, his advices have always inspired and guided my professional life ever since my undergraduate years. I am also very grateful to Prof. Marcio Thomé, who kindly helped me at the beginning of the research process and with whom I had some instructive discussions. I also express my gratitude to Prof. Reinaldo Morabito and Prof. Laura Bahiense for participating in the thesis committee.

I greatly acknowledge the support provided by the Instituto Nacional de Tecnologia (INT) and I thank all my DEAP colleagues from this institute for offering a pleasant working environment. I hope to be able to make my contribution to the work of this department in the coming years. In special I am very grateful to Manoel Saisse and Euclydes da Cunha, with whom I had many enlightening discussions about this thesis' research problem, and to my supervisor Valeria Said for her constant support. A special thanks goes to Ricardo Costa and my colleagues from Trilha. Although we have been apart along the last few years, I am truly grateful for all the industrial projects experiences we had together and recognize that they were essential for the development of this research. I am also very greatful to Denis Pinha (Ph.D. student) for the fruitful exchange of ideas, on our related research problems, along the last few years. I also wish to acknowledge the support of Claudia, Isabel, Fernanda, Gilvan and Eduardo in the Industrial Engineering Department Office.

Finally, a very special thanks goes to my family and friends. In particular I thank my beloved husband, Haroldo, and my boys, Bady and Elias. I would not have been where I am now without their unconditional love and support. I owe them my sense of purpose. I am so grateful to my dear parents, Francisco and Leila, who have always supported me. They have been a tremendous source of inspiration to me in many circumstances of my life. A special thanks goes to my sister Débora and her husband Wilfredo, although living so far from us, are always so present and have helped me gather relevant references for this thesis. I am also grateful to my sisters (Clara and Silvia), my brother and his wife (Marcos and Poly), my nieces and nephew (Julia, Paula and Vicente), and aunt Dyla and to my husband's family (Ione, Lucia, Iéia, Paula, Crica, André, Valquiria, Thiago and André Filho) for their intense support always. I wish to thank my best friend, Isabella, for the many occasions she went to meet me at PUC, and my good friends, Catia, Neli and Sergio, who are always encouraging me. And I cannot forget Michelly for all the assistance she gave me during all these years.

This thesis is dedicated to the memory of my grandmother Maria Delfina, who left us before I could finish this work. She inspired me and encouraged me in all of my pursuits. I am sure she would have been proud now.

To my beloved parents, Francisco and Leila

Abstract

Carvalho, Andréa Regina Nunes; Carmo, Luiz Felipe R. R. Scavarda do (Advisor), Oliveira, Fabricio Carlos Pinheiro de (Co-Advisor). **Tactical capacity planning in an ETO production setting using optimization models: A real-world industrial context.** Rio de Janeiro, 2015. 117p. D.Sc. Thesis – Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Many engineering-to-order (ETO) organizations are multi-project capacity-driven production systems in which capacity planning is of major importance in the order acceptance phase. The academic literature, in this area, presents a research–practice gap with a lack of studies on the application of decision support tools to address capacity planning problems in real-world ETO settings. Within this context, the goal of this thesis is to develop a tactical capacity planning solution to support the order acceptance phase of a real-world multi-project organization that produces customised equipments on the basis of ETO policy. This research study lays in the development of mixed integer linear programming models and their practical application to solve production planning problems in the studied organization. As for the theoretical contributions of this thesis, first a deterministic model is presented in which modelling issues that are either not entirely explored in other studies or that have to be adapted to the specificities of the studied setting are taken into account. Moreover, a robust optimization model extends the former model by considering uncertainties of the planning problem. The models were fed with real-world data and solved in order to check whether they actually reflect the planning problem. Furthermore, alternative scenarios were also generated to assist the management board in the order acceptance phase. As for practical implications, for the company´s manufacturing planning team, the proposed solution enhanced the decision-making process regarding tactical capacity planning, addressing different shortcomings of the company´s current planning method. Empirical results suggest that with a slight increase in cost (0.02%) a part component should be processed in-house instead of being outsourced and that with a 0.8% increase in cost (which includes hiring 21% more personnel) the probability of violating the production plans decreases from 90% to 15%, representing a much more stable (protected against uncertainty) situation. From an academic perspective, this research adds empirical evidence to enrich the existing literature, as it not only presents a real case application, but also highlights issues that must be considered and managed in a real-world context in order to develop and implement appropriate techniques to cope with the aforementioned planning problem.

Keywords

Engineer-to-order; aggregate production planning; decision support system; mathematical programming; robust optimization.

Resumo

Carvalho, Andréa Regina Nunes; Carmo, Luiz Felipe R. R. Scavarda do (Orientador), Oliveira, Fabricio Carlos Pinheiro de (Co-Orientador). **Planejamento tático da capacidade na produção ETO usando modelos de otimização: o contexto de um problema real na indústria.** Rio de Janeiro, 2015. 117p. D.Sc. Tese de Doutorado– Departamento de Engenharia Industrial, Pontifícia Universidade Católica do Rio de Janeiro.

Muitas organizações de produção por projeto (i.e., também conhecidas pela sigla inglesa ETO, engineering-to-order) são sistemas de produção multi-projeto em que o planejamento da capacidade, na fase de negociação de novos pedidos, é de suma importância. A literatura acadêmica, nesta área, apresenta uma lacuna entre teoria e prática em função da falta de estudos sobre a aplicação de ferramentas de apoio à tomada de decisão para resolver problemas de planejamento de capacidade em ambientes reais de produção ETO. Dentro deste contexto, o objetivo deste trabalho é desenvolver uma solução para o planejamento tático da capacidade produtiva, apoiando essa fase de negociação, numa organização multi-projeto fabricante de equipamentos especiais sob encomenda. Este estudo envolve o desenvolvimento de modelos de programação linear inteira mista e sua aplicação para resolver problemas de planejamento da produção na organização estudada. Quanto às contribuições teóricas desta tese, é apresentado um modelo determinístico em que são consideradas questões de modelagem não totalmente exploradas em outros estudos ou que tem de ser adaptadas às especificidades do contexto estudado, como a representação da capacidade extra, de processos com múltiplos estágios e a relação de precedência entre as atividades. Além disso, um modelo de otimização robusta, baseado na abordagem proposta por Bertsimas e Sim (2004), estende esse modelo determinístico, considerando incertezas relativas aos tempos de processamento das atividades. Os modelos foram alimentados com dados do mundo real e executados para fins de validação de sua utilidade para resolver o problema de planejamento em questão. Cenários alternativos também foram gerados para apoiar a tomada de decisão dos gestores dessa empresa na fase de negociação de novos pedidos. Com relação às implicações práticas, para a equipe de planejamento da empresa, a solução proposta aprimora o processo de tomada de decisão no que tange o planejamento tático da capacidade produtiva. A solução, além de resolver algumas deficiências do método de planejamento atual da empresa, fornece informações mais detalhadas sobre o problema, permite a intervenção do gestor na construção dos planos de capacidade e incorpora dados relativos à variabilidade nos tempos de processamento permitindo assim uma postura pró-ativa mediante as incertezas. Resultados empíricos mostram que, com um aumento relativamente pequeno no custo (0.02%), um componente deveria ser preferencialmente produzido na própria empresa (ao invés de ser subcontratado). Além disso, com um aumento de 0.8% no custo (o que inclui a contratação de 21% a mais de mão-de-obra direta), a probabilidade de violação dos planos de produção é reduzida de 90% para 15%, representando um plano mais estável e protegido contra incertezas. Do ponto de vista acadêmico, esta pesquisa acrescenta evidências empíricas para enriquecer a literatura existente, uma vez que não só apresenta um caso real, mas também destaca questões que devem ser consideradas e gerenciadas em um contexto do mundo real para que se possa desenvolver e implementar técnicas adequadas para lidar com o problema de planejamento estudado.

Palavras-chave

Engenharia sob encomenda; planejamento agregado de produção; sistema de apoio à tomada de decisão; programação matemática; otimização robusta.

Summary

List of Tables

List of Figures

Glossary

Activity: an aggregation of operations or processing steps within the same type of production process.

Backlogging: the accumulation of work or customer orders

Bill of materials: the list of materials, including raw materials and intermediate items, needed to manufacture an end product.

Cadence: the flow or rhythm of events (for instance, the production process occurs according to a cadence or rhythm)

Committed workload: the ongoing projects that have been confirmed and are associated to fixed customer orders

Component: a part of a project and comprises a set of interconnected activities

Deadline: due date (e.g., a project's due date)

Incoming orders: demands (i.e., customer projects) that are arriving

Intensity of an activity: proportion of an activity's processing time

Milestones: reference points indicating the completion of phases within a project’s execution

Manufacture-to-order: refer to make-to-order and engineer-to-order production strategies

Processing stage: an activity

Processing time: the duration of an activity or the time needed to complete an activity

Product's routing: the description of a product's processing stages

Release date: earliest starting time

Work centre: the resources (i.e., machines and tools) within a type of production process

Workload: an amount of work (expressed in hours in this thesis)

1 Introduction

Engineering-to-order (ETO) has become increasingly important in production systems, particularly for delivering customised products (Gosling and Naim, 2009; Radke and Tseng, 2012; Grabenstetter and Usher, 2013, 2014), a demand which is growing nowadays (Powell et al., 2014). However, highly customised ETO environments have received much less attention from researchers than high volume, standardised, make-to-stock (MTS) environments (Gosling and Naim, 2009; Yang, 2013; Grabenstetter and Usher, 2014; Willner et al., 2014).

In the ETO context, the production flow is entirely driven by actual customer orders with the decoupling point located at the design stage (Gosling and Naim, 2009; Grabenstetter and Usher, 2014; Powell et al., 2014). ETO supply involves physical stages (e.g., component manufacturing, assembly and installation) and non-physical stages (e.g., tendering, engineering, design and process planning) (Bertrand and Muntslag, 1993). ETO processes are highly knowledge intensive and are often built on tacit knowledge, as product structures are subject to constant changes in terms of design and full automation of production processes is often not feasible due to the customer specific requirements (Willner et al., 2014). The ETO context is associated with chaotic production in high-complexity/ high- uncertainty situations (Little et al., 2000; Gosling and Naim, 2009; Yang, 2013), where the ability to address instability in demand and to respond to demand modifications over time is crucial (Hicks and Braiden, 2000; Hicks et al., 2000; Little et al., 2000; Hans et al., 2007; Zorzini et al., 2008).

Due to complex production processes and long lead times, production often starts before the overall project design has been completed (Monostori et al., 2010). Missing information and engineering revisions caused by this overlapping of both non-physical and physical stages are major sources of uncertainty that complicate the management of ETO manufacturing (Hicks and Braiden, 2000; Hicks et al., 2001; Hans et al., 2007). In spite of this, project accept/reject decisions must be made and due dates must be set. ETO customers require reliable due dates as part of the service mix offered, so being able to quote accurate and reliable due dates is a major competitive advantage in this context (Hans et al., 2007; Grabenstetter and Usher, 2014, Mourtzis et al., 2014). Moreover it is a common practice that organizations accept as many projects as they can possibly acquire and promise delivery dates as early as possible. This is done without sufficiently assessing the impact of incoming projects on the resource capacity, leading to the overload of resources and affecting delivery performance and the profitability of the production system (Hans et al., 2007).

Many ETO organizations are multi-project production systems driven by capacity, meaning that their operations are constrained by various scarce resources. Capacity planning is an important aspect in this context and refers to the problem of matching demand for the availability of resources in the medium term or tactical level (Gademann and Schutten, 2005). It refers to an aggregate production planning (APP) that typically encompasses a time horizon from 3 to 18 months and focuses on determining the optimum production, workforce, and inventory levels for each period of the planning horizon for a given set of production resources and constraints (Ramezanian et al., 2012; Jamalnia and Feili, 2013; Diaz-Madroñero et al., 2014; Wang and Yeh, 2014).

Contrary to the operational planning level, the tactical planning level is characterised by high degree of capacity flexibility (e.g., overtime and subcontracting) and requires methods that can exploit this flexibility by supporting the planner in assessing trade-offs between the delivery performance and the expected costs of using nonregular capacity (Hans et al., 2007). Furthermore, many authors recognize that a capacity planning system that supports decision makers to quickly analyse the impact of potential orders on capacity plans is of major importance during the order acceptance phase to determine reliable due dates and price quotations (Giebels, 2000; Gademann and Schutten, 2005; Hans et al., 2007; Sawik, 2009; Montreuil et al., 2013; Nobibon et al., 2015). These decisions are crucial in the ETO context and are related to setting important milestones for each project and for bid preparation (Bertrand and Muntslag, 1993; Gademann and Schutten, 2005; Hans et al., 2007; Zorzini et al., 2008;

Grabenstetter and Usher, 2013; Montreuil et al., 2013). Additionally, other authors (Bertsimas and Thiele, 2006, Aghezzaf et al., 2010) preconize that a tactical production planning model which does not integrate the variability of critical parameters in the planning process results often in worthless plans or at the best in plans which must be revised frequently.

Project-oriented planning approaches (e.g., resource constrained project scheduling - RCPS) and rough-cut capacity planning (RCCP) methods serve as planning tools at aggregate levels in MTO (make-to-order) or ETO systems (Gademann and Schutten, 2005; Hans et al., 2007; Tolio and Urgo, 2007; Alfieri et al., 2011, 2012; Radke et al., 2013) and have been frequently addressed in the scientific literature (Neumann and Zimmermann, 2000; Markus et al., 2003; Tolio and Urgo, 2007; Monostori et al, 2010; Alfieri et al., 2012; Rieck et al., 2012; Artigues et al., 2015). Moreover, mathematical models have traditionally been used to solve planning problems at the tactical level for the manufacture–to-order (i.e., make- or engineer-to-order) context (Hans et al., 2007; Zorzini et al., 2008), where as some consider uncertainties in manufacturing systems (Wullink et al. 2004; Ebben et al. 2005; Genin et al. 2008; Aghezzaf et al., 2010; Lusa and Pastor, 2011; Zhen, 2012; Alem and Morabito, 2012; Aouam and Brahimi, 2013; Rahmani et al., 2013; Munhoz and Morabito, 2014).

Despite this recognized importance of tactical capacity planning for ETO environments, there is a gap in the APP literature between theory and practice (Ramezanian et al., 2012; Jamalnia and Feili, 2013; Lingitz et al., 2013; Liu et al., 2013; Diaz-Madroñero et al., 2014). In a recent literature review, comprising 250 optimization models for tactical production planning, Diaz-Madroñero et al. (2014) outline that, although some proposals have been validated in real environments, very few are reported to be implemented and incorporated into the planning systems of the companies considered. The authors conclude that there is a gap between academic research and industry and that production-planning models should be solved with highly customizable and easy-to-use tools that integrate into firm's current information systems in order to bridge this gap. Ramezanian et al. (2012) affirm that many contributions in this field address the development of solution algorithms, whereas the proposed models rarely consider the industrial environment realistically. Additionally, studies by Buxey (2003, 2005), Corti et al. (2006) and Sharda and Akiya (2012) have shown that approaches to APP and correlated topics associated with tactical capacity planning are often not put into practice because they are too complicated or do not make the actual problem description applicable. This is reinforced in the ETO context with the need for techniques to assist management (Little et al., 2000; Gosling and Naim, 2009; Yang, 2013; Grabenstetter and Usher, 2014; Pero and Rossi, 2014; Powell et al., 2014).

Within this context, the goal of this thesis is to develop a tactical capacity planning solution based on optimization models to support the order acceptance phase of a real-world multi-project organization that produces customised equipments on the basis of ETO policy. This research study centers on the development of mixed integer linear programming models and their practical application to solve production planning problems in the studied organization. As for the theoretical contributions, a deterministic model is presented in which modelling issues that are either not entirely explored in other studies or that have to be adapted to the specificities of the studied setting are taken into account. Moreover, a robust optimization model extends the deterministic model by considering uncertainties of the planning problem. The application of the two models in a real-world ETO production setting contributes to reducing the research-practice gap by providing academic investigators with information on relevant issues that must be considered and managed in a real-world context in order to develop and implement appropriate techniques to cope with the aforementioned tactical planning problem.

This thesis is organized in 6 chapters. The introduction is presented in this first one. Chapter 2 offers a review of tactical capacity planning models. Chapter 3 describes the planning problem in the real-world ETO setting studied and the research method adopted. Chapter 4 presents the deterministic mathematical programming model and its application in a real-world ETO production setting, whereas Chapter 5 extends this model to deal with uncertainty by presenting the robust optimization model and its application. The conclusions and suggestions for future research are given in Chapter 6.

2 Theoretical background

This chapter provides a literature review on the basis of the ETO concept and its main issues, on tactical capacity planning in ETO and on existing mathematical models to solve planning problems at the tactical level for the manufacture-to-order context.

2.1. A brief overview in ETO

In their literature review, Gosling and Naim (2009) identified a lack of clarity concerning the appropriate terminology used to describe ETO supply chain type. Nevertheless, the authors conclude that the production flow in this context is driven by actual customer orders with the decoupling point (i.e., order penetration point - OPP) located at the design stage, as seen in Figure 1. Other expressions associated to this context are project manufacturing and multi-project organizations (Gademann and Schutten, 2005; Hans et al., 2007; Yang, 2013) where different projects are executed together competing for the same production resources (Herroelen and Leus, 2004; Chtourou and Haouari, 2008; Van de Vonder et al., 2008; Deblaere et al., 2011; Alfieri et al., 2011, 2012; Artigues et al., 2013).

Product delivery strategy	*Design*	*Fabrication & procurement*	*Final assembly*	*Shipment*
Make-to-stock (MTS)				OPP
Assemble-to-order (ATO)			OPP	
Make-to-order (MTO)		OPP		
Engineer-to-order (ETO)	OPP			

Fig. 1: Product delivery strategies (adapted from Olhager, 2003)

ETO products have deep and complex structures, with many assembly levels, needing coordination with component supply (Hicks and Braiden, 2000; Cameron and Braiden, 2004; Alfieri et al., 2011; Gosling et al., 2014; Grabenstetter and Usher, 2014). These products are super-value goods that are highly customised, produced in low volumes (often one-of-a-kind) and have long engineering trajectory, with many disruptions and adaptations due to specification changes demanded by the customer (Hans et al., 2007; Pandit and Zhu, 2007; Alfieri et al., 2011; Powell et al. 2014). The production processes are typically non-repetitive yet labor intensive, often demanding highly skilled labor (Powell et al. 2014). In this context, customers change their requirements over the time of product fabrication and, thus, the ability to respond to these modifications is a prerequisite of success (Little et al., 2000; Cameron and Braiden, 2004; Zorzini et al., 2008; Montreuil et al., 2013) and the main order winning characteristic in this context is fitness for purpose (Little et al. 2000). Hicks and Braiden (2000), Cameron and Braiden (2004), Grabenstetter and Usher (2014) and Willner et al. (2014) also highlight that price, reduced lead times and delivery performance are important aspects of customer service as most contracts include financial penalties for late delivery.

In the customer enquiry stage, considered the most critical stage in the ETO context, a customer provides an invitation to tender for a particular product to prospective suppliers, requiring the determination of a price and due date (Aslan et al., 2012). These authors highlight that these decisions require: the estimation of lead times; archiving and retrieval of product data; assessment of available design/production skills and facilities; estimation of costs/profit margins; and effective coordination and communication between all departments involved in the activities listed above. In this stage, there are often a number of phases of negotiation with suppliers that aim to match overall project cost and lead-time with anticipated customer and market requirements (Hicks et al., 2000). Zorzini et al. (2008) define this as a multi-stage decision process, involving complex trade-offs and so requiring an inter-disciplinary teamwork and preparing attractive and reliable bids is only possible with considerable expenditure in terms of time and other resources.

In the order acceptance process of ETO companies, production planning and control may be complex because capacity planning must take into account potential incoming orders which do not have the bill of material structures fully available during this early planning stage (Aslan et al., 2012; Liu et al., 2013). Due to complex production processes and long lead times, production often starts before the overall project design has been completed (Monostori et al., 2010). For planning purposes, much of the data relating to lead times are based on estimates, as historic data is often sparse and unreliable due to operational difficulties in collecting data on shop floor (Hicks and Braiden, 2000). Some of the most common problems associated with ETO planning processes are difficulties in estimating lead-times and delivery dates and conflicts between projects and manufacture schedules (Pandit and Zhu, 2007). Hicks et al. (2000) stress that delivery dates in tenders are based on lead-time estimates which are usually produced without information on available capacity, as it is common for there to be several “floating” quotations awaiting responses from potential customers. Management has to rely on a rough estimation of the impact of an incoming order on resource utilization and eventually has to adjust capacity, as micro process planning (i.e., detailed technological planning of production activities that result in manufacturing instructions - Giebels, 2000) has not been performed yet (Zijm, 2000; Grabenstetter and Usher, 2014). This issue represents a major source of uncertainty that complicates the management of ETO manufacturing (Hicks and Braiden, 2000; Hicks et al., 2001; Hans et al., 2007). Additionally, as the pattern of demand (i.e., in terms of the level and mix of work) fluctuates significantly over time, ETO companies cannot accurately forecast demand, neither order materials nor produce in advance (Powell et al., 2014). In this context, it is difficult to balance production due to the dynamic nature of constraints (Hicks and Braiden, 2000).

Nevertheless, project accept/reject decisions must be made and due dates must be set. ETO customers require reliable due dates as part of the service mix offered, so being able to quote tight and reliable due dates is a major competitive advantage in this context (Hans et al., 2007; Grabenstetter and Usher, 2014, Mourtzis et al., 2014). In practice, order acceptance and capacity planning decisions are often functionally separated, according to Ebben et al. (2005) and Huang et al. (2011), since the sales department is responsible for order acceptance and the production department takes care of production and capacity planning. According to a survey on ETO firms, Hicks and Braiden (2000) affirm that some companies had formalized methods of rough-cut capacity planning based upon spreadsheets but none had formalized systems for capacity requirements planning or finite loading. In consequence, production plans were often not consistent with capacity constraints. For instance, Monostori et al. (2010) affirm that project planners typically try to sequence activities as early as possible relying on the conventional wisdom that it can never be wrong to get work done early. To acquire projects, companies tend to promise a delivery date that is as early as possible (Hans et al., 2007). Moreover it is a common practice that organizations accept as many projects as possible and promise delivery dates as early as possible. This is done without sufficiently assessing the impact of incoming projects on the resource capacity, leading to the overload of resources and affecting delivery performance and the profitability of the production system (Little et al., 2000; Hans et al., 2007).

2.2. Tactical planning in ETO

To deal with the planning complexity in ETO organizations, the planning process needs to be broken down into more manageable parts using a model for hierarchical planning and control based on the three managerial decision levels (strategic, tactical, and operational) (Hans et al., 2007). To adequately perform multi-project planning, projects must be considered simultaneously at all planning levels, while taking into account that those different levels have different objectives, constraints, degrees of aggregation, and capacity flexibility. In a positioning framework, Giebels (2000) details these three levels into three categories: technological, company management, and production planning (see Figure 2). The author highlights that there are many different planning functions to be performed on the higher levels of production planning that are denominated capacity planning and distinguishes long-term capacity planning (to decide on investments in capacity expansion on the basis of market expectations), capacity planning in the order acceptance phase, and the resource loading, thus to enable feasible lower-level schedules (already in the operational level).

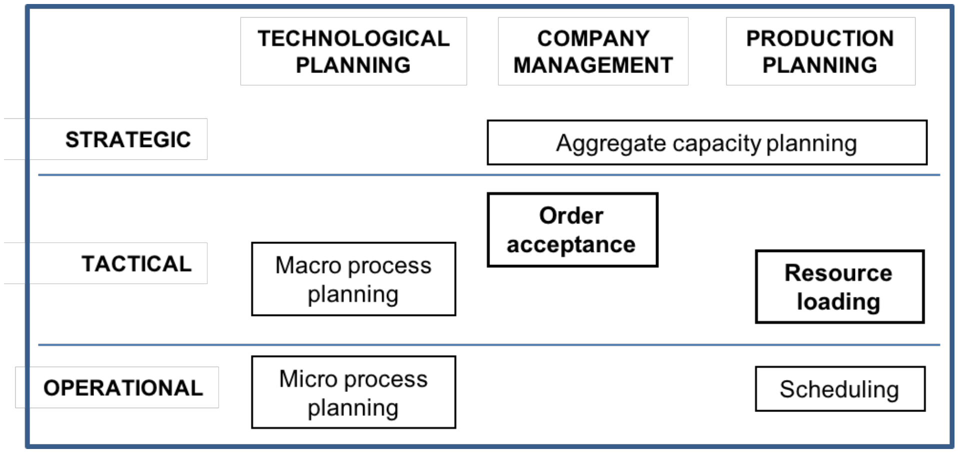


Fig. 2: Positioning framework (from Giebels et al., 2000)

In this planning hierarchy, Bushuev (2014) affirms that tactical planning is a middle-level activity connecting strategic planning and operations control and that the basic problem to be solved is the allocation of resources (i.e., capacity, workforce availability, storage) over a medium-range planning horizon. Within the tactical level, the order acceptance decision involves analysing the consequences of accepting these new orders as well as determining their delivery dates and prices (Little et al., 2000; Giebels, 2000; Ebben et al., 2005). This decision underlines the difficulties encountered in managing trade-offs resulting from the conflicting objectives of sales/marketing and production (Ebben et al., 2005; Zorzini et al., 2008; Huang et al., 2011; Aslan et al., 2012).

At the negotiation and order acceptance stage, adequate capacity planning methods that assess the consequences of decisions for the production system are crucial (Giebels, 2000; Gademann and Schutten, 2005; Hans et al., 2007; Sawik, 2009; Montreuil et al., 2013). When only rough information is available, important outputs (e.g., internal and external due dates, milestones and required capacity levels) of such methods (e.g., RCCP) serve as the basis for accepting or not an incoming project, acquiring additional resources if necessary, ordering raw materials and final fixing of due dates (Hans et al., 2007). Additionally, tactical capacity planning methods ideally should use capacity flexibility (e.g. by working in nonregular time or by subcontracting) to support the planner in making a trade-off between the expected delivery performance and the expected costs of exploiting this flexibility (Hans et al., 2007). Ishii et al. (2014) state that the objective in the order acceptance process is to maximize profits with production capacity limitations. Mestry et al. (2011) and Wang et al. (2013) emphasize that integrating order acceptance and capacity planning provides tremendous opportunities to maximize the operational profits of MTO operations.

Moreover, in multi-project contexts, such as ETO, projects are subject to considerable uncertainty, which may lead to numerous schedule disruptions (Herroelen and Leus, 2004; Corti et al., 2006; Tolio and Urgo, 2007; Chtourou and Haouari, 2008; Van de Vonder et al., 2008; Deblaere et al., 2011; Alfieri et al., 2012; Artigues et al., 2013). According to these authors, this uncertainty derives from several causes (e.g., revisions in project scope, due date changes, variability in processing times, unavailability of resources) and has a significant impact on the stability and the performance of the production system by affecting the due dates achievement, the efficient resource allocation and the usage of nonregular working force (Tolio and Urgo, 2007).

In reactive planning approaches, uncertainties are not considered and the baseline schedules must be revised when unexpected events occur (Herroelen and Leus, 2004; Chtourou and Haouari, 2008; Deblaere et al., 2011; Alfieri et al., 2012). According to Khakdaman et al. (2015), demand, process and supply can be considered the three main types of uncertainties which will make any medium term plan obsolete, thus forcing a re-planning cycle. In practice, usually manual intervention on the production plan is typically required to deal with overloaded resources and violated deadlines. According to Monostori et al. (2010) experience, medium-term production plans are readjusted frequently and less than 50% of the original plan is actually executed.

On the other hand, under uncertainty conditions, many authors (Herroelen and Leus, 2004; Tolio and Urgo, 2007; Chtourou and Haouari, 2008; Van de Vonder et al., 2008; Deblaere et al., 2011; Alfieri et al., 2012; Artigues et al., 2013; Radke et al., 2013) support that the baseline schedule must be robust and therefore should incorporate a certain degree of variability anticipation. According to these authors, the robustness concept may refer to the solution robustness or stability (i.e., the insensitivity of planned activity start times to schedule disruptions) or to the solution quality robustness (i.e., the insensitivity in terms of the objective function). Lagemann and Meier (2014) highlight that robustness includes both stability (i.e., resilience) and flexibility (i.e., ability to adapt to unforeseen events) and therefore a robust plan should contain a stable basic plan and one or several back up plans that need to work together.

Proactive planning approaches, in this sense, try to incorporate information about uncertainty in the baseline schedule so that it can be protected, as well as possible, against future disruptions, i.e., aiming at stability (Herroelen and Leus, 2004; Chtourou and Haouari, 2008; Deblaere et al., 2011; Alfieri et al., 2012). Policella et al. (2004) defines that a plan is robust when it can absorb disruptions (external events) without loss of consistency while keeping the pace of execution, whereas Khakdaman et al. (2015) interpret that a robust plan is one that remains valid for a longer time and is insensitive to the effects of uncertainties.

2.3. Tactical capacity planning models

Mathematical models have traditionally been proposed to solve planning problems at the tactical level for the manufacture-to-order (i.e., make- or engineer-to-order) context (Hans et al., 2007; Zorzini et al., 2008). As this research is related to the tactical-level capacity planning problem coupled with production decisions in the tendering phase, this section highlights papers referring to capacity planning (i.e., determining ideal levels of workforce, production, subcontracting) over a medium-range planning horizon to satisfy demand requirements, order acceptance/rejection decisions, and related issues (e.g., due date setting) that attempt to exploit interaction with capacity planning.

Table 1 summarizes these papers, classifying them according to four groups of attributes that are discussed afterwards. The first group refers to the production planning functions addressed (i.e., workload analysis, due date setting, capacity adjustments, allocation of orders and scheduling). The second group is related to the solution approach and comprises the model type (i.e., whether deterministic, stochastic, robust), the solution method (i.e., whether exact or heuristic) and the goal or performance criteria pursued (i.e., expressed in the objective function). The third group refers to the modelling issues that are pertinent to this research: nonregular capacity, processing stages, precedence relations among activities and the uncertainties considered. Finally, the last group refers to the application approach and comprises the context of application (e.g., ETO, MTO) and the empirical nature of the research.

In a preliminary analysis of this table, one may already conclude that there are few research papers explicitly addressing the ETO context and even fewer that may be classified as real problem-solving papers. Additionally, most of the reviewed researches refer to deterministic models and therefore do not consider uncertainties. A detailed description on the four groups of attributes displayed on Table 1 is given next along with the main conclusions regarding this review on mathematical models.

Table 1: Summary of relevant literature on tactical capacity planning models

Reference	Production planning functions					Solution approach			Modelling issues				Application approach	
	Workload analysis	Due date setting	Capacity adjustments	Allocation of orders	Scheduling	Model type	Method	Goal	Nonregular capacity[1]	Processing stages	Precedence relations[2]	Uncertainties	Context	Nature[3]
ezzaf et al. (2010)			X	X		robust stochastic	exact	min. cost	ov	multiple	typical	demand	not detailed	B
n and Morabito (2012)			X	X		robust	exact	min. cost	ov	single	-	demand, costs	MTO	B
ri et al. (2011)				X		deterministic	exact	min. makespan	-	multiple	feeding	-	ETO	A
ri et al. (2012)				X	X	stochastic	exact	min. makespan	-	multiple	typical	resource need	ETO-MTO	A
am and Brahimi (2013)	X					robust	exact	min. cost	-	one aggregate	-	demand	not detailed	C
estin et al. (2007)			X	X	X	deterministic	heuristic	resource level	-	multiple	GPRs	-	MTO	C
ominas et al. (2012)			X	X		deterministic	exact	max. profit	ov, hir, wta	bottleneck	-	-	not detailed	C
et al. (2006)		X	X	X	X	deterministic	heuristic	min. cost	ov, out	multiple	typical	-	MTO	C
en et al. (2005)	X		X	X	X	stochastic	heuristic	max. utilization	ov, out, hir	multiple	typical	processing times	MTO	C
eman and Schutten (2005)			X	X		deterministic	heuristic	min. cost	not detailed	multiple	typical	-	Multi-project	C
n et al. (2008)			X	X		robust	exact	min. cost	ov, hir	one aggregate		demand	MTS-MTO	B
s (2001)			X	X		deterministic	heuristic	min. cost	ov, out, hir	multiple	typical	-	MTO	C
ntari et al. (2011)	X	X	X	X	hierarchical	deterministic	exact	min. cost	ov, out	multiple	typical	-	MTS-MTO	C
sman (2000)	X		X	X	X	deterministic	heuristic	min. overtime, WIP, meet dd	ov, sub	multiple	typical	-	MTO	C
Lusa and Pastor (2011)			X			stochastic	exact	min. cost	wta, ov	multiple	typical	demand	not detailed	C
Markus et al. (2003)			X	X	hierarchical	deterministic	exact	min. cost	sub	multiple	feeding	-	ETO-MTO	B
Mestry et al. (2011)	X		X	X	X	deterministic	heuristic	max. profit	ov	multiple	typical	-	MTO	C
Monostori et al. (2010)				X	hierarchical	deterministic	exact	min. overuse	-	multiple	feeding	-	ETO	B
Munhoz and Morabito (2014)				X		robust	exact	min. cost	-	one aggregate	-	fruit acidity	MTS-MTO	A
Neumann and Zimmermann, (2000)			X		X	deterministic	heuristic	resource level, max. net present value	ov	multiple	GPRs	-	not detailed	C
Nobibon et al. (2015)	X		X	X		deterministic	exact	max. revenue	hir	one aggregate	-	-	ETO-MTO	C
Rahmani et al. (2013)			X	X		robust stochastic	exact	min. cost	ov, out, hir	2-stage	typical	demand, costs	not detailed	B
Rieck et al. (2012)			X	X	X	deterministic	exact	resource level, min. overload	not detailed	multiple	GPRs	-	not detailed	C
Rom and Slotnick (2009)	X				X	deterministic	heuristic	max. profit	-	bottleneck		-	MTO	C
Sawik (2009)		X		X	X	deterministic	exact	min. delay	-	multiple	typical	-	MTO	B
Schwindt and Paetz (2015)				X	X	deterministic	exact	min. makespan	-	multiple	GFPRs	-	not detailed	C
Slotnick and Morton (2007)	X				X	deterministic	heuristic	max. profit	-	bottleneck	-	-	MTO	C
Tolio and Urgo (2007)			X	X	X	stochastic	exact	min. outsource cost	out	multiple	typical	resource need	ETO	B
Tunali et al. (2011)		X		X	hierarchical	deterministic	exact	min. cost	sub, inv	multiple	typical	-	MTO	B
Wang et al. (2013)	X				X	deterministic	heuristic	max. revenue	-	2-stage	typical	-	MTO	C
Wullink et al. (2004)	X		X	X		stochastic	heuristic	min. cost	out	multiple	typical	work content	MTO	C
Zhen (2012)			X			stochastic	exact	min. cost	out	one aggregate	-	demand	not detailed	C
Zhong et al.(2014)	X				X	deterministic	heuristic	min. makespan	-	bottleneck	-	-	not detailed	C

1. Nonregular capacity : hir = hiring; wta = working time accounts; ov = overtime; out = outsource; subcontract = sub; inv = inventory
2. Precedence relations: typical = finish-to-start; GPRs = generalized precedence relationships; GFPRs= generalized feeding precedence relationships
3. Nature (empirical nature): A = real problem-solving papers, B = hypothetical problem-solving papers, C = methodological papers

2.3.1. Production planning functions

Production planning concerns the time-based allocation of orders to resources. Capacity planning comprehends the rough production planning activities (Giebels, 2000), which can also be found under the aggregate production planning context. According to Ramezanian et al. (2012), aggregate production planning is medium-term capacity planning often from 3 to 18 months ahead. It is normally concerned with the lowest-cost method of production planning to meet customer‘s requirements and to satisfy fluctuating demand over the planning horizon.

For the purpose of this research and in accordance to these definitions and to Giebels (2000) framework (see Figure 2), the tactical capacity planning problem comprehends capacity planning in the order acceptance and resource-loading phases. In the order acceptance phase, two functions are defined (1) the analysis of the consequences for workload by accepting specific orders and (2) the analysis or determination of due dates for individual orders. As for resource loading phase, concerned with enabling feasible lower level schedules, two other functions are defined: (1) capacity adjustments (i.e., recognition of capacity problems and the planning for additional capacity) and (2) the allocation of orders to time periods (i.e, orders are disaggregated in jobs, sufficiently detailed in macro process planning, that can be assigned to machine groups or work cells).

Although this research refers to tactical planning level, many researches present models that make use of scheduling methods to support the tactical decision. Other papers present a hierarchical structure in which the scheduling problem is solved after the planning problem has been addressed. However, in an ETO system, which is the focus of this study, detailed engineering and process planning, required for scheduling purposes, is not possible before order acceptance and, in practice, management has to rely on rough estimation for capacity planning (Zijm, 2000; Aslan et al., 2012).

In order to classify the literature according to these aforementioned production planning functions, five categories are defined: workload analysis, due date setting, capacity adjustments, allocation of orders and scheduling. As the functions are closely related, some papers may address more than one. As seen in Table 1, most of the papers concern the resource loading phase which involves adjusting capacity and allocating orders to time periods. Moreover, some papers concern the order acceptance phase (i.e., a higher level decision process) when demand data is roughly specified.

2.3.2. Solution approach

In general, optimization methods focus on optimality of a solution and the problem is considered solved when the optimum is reached. In this process, alternative solutions with almost equivalent values for the objective function are usually discarded. These solutions might constitute an improvement to other criteria than the initial objective, such as robustness, an aspect that is valuable when uncertainties are present (Hans et al, 2007). As a matter of fact, there is a research track that has addressed attention towards increasing the robustness of the planning methods by incorporating uncertainties to them. For instance, different mathematical programming approaches have been used to formulate uncertainty in manufacturing systems (Rahmani et al., 2013), such as stochastic, robust and robust stochastic optimization models.

In Table 1, model type "stochastic" refers to approaches that require full knowledge of the distributions (i.e., probabilistic information) of the uncertain data (Mulvey et al., 1995; Bredstrom et al., 2013), which can be found in many research studies (Wullink et al. 2004; Ebben et al., 2005; Lusa and Pastor, 2011; Zhen, 2012). In a research study, Alfieri et al. (2012) highlight that the explicit consideration of uncertainty, through a stochastic planning approach, can lead to a significant advantage in terms of plan effectiveness with respect to the traditional mean value approach. According to Juan et al. (2014), since uncertainty is present in most real-world processes, considering random processing times, drawn from non-negative, asymmetrical and skewed rightwards probability distributions (such as lognormal, weibull and gamma), represents a more realistic scenario than simply considering deterministic times.

On the other hand, as full knowledge of probabilistic information is rarely available in practice, robust optimization approaches (i.e., classified as model type "robust" in Table 1) have received a lot of attention as the uncertainty of the parameters is modeled as lower and upper bounds with no need for exact distributions (Bredstrom et al., 2013, Gorissen et al., 2015). For instance, Soyster (1973) proposes a linear optimization model to construct excessively conservative solutions that are feasible for all data in a given uncertainty set without specifying these distributions. To address Soyster's overconservatism and also to retain the advantages of his linear programming framework, Bertsimas and Sim (2004) propose a robust optimization approach to address data uncertainty that allows the degree of conservatism of the solution to be controlled (i.e., protection is provided for the case where only a pre-specified number of the input coefficients changes from its base value). These authors' approach was adopted in research studies conducted by Alem and Morabito (2012) and Munhoz and Morabito (2014).

Moreover, there is also the model type "robust stochastic" which assumes that the probabilistic distributions are known but differ from the stochastic models because of the possibility to control the level of conservatism. This can be found in the studies done by Aghezzaf et al. (2010) and Rahmani et al. (2013).

Despite the development of methods that consider uncertainty, most mathematical programming models found in literature assume that the input data (parameters) are precisely known and are set equal to some nominal values (i.e., the deterministic models) (Aouam and Brahimi, 2013). In fact most of the reviewed models in Table 1 are classified as model type "deterministic", although manufacturing environments are in great part characterized by uncertainty (Tolio and Urgo, 2007).

The reviewed papers are also classified in terms of the solution method and goal. In this sense, all of the revised models concern optimization methods that comprise either exact methods (i.e., methods that guarantee an optimum solution of the problem) or heuristic methods (i.e., methods that attempt to yield a good, but not necessarily optimum solution). Additionally, the performance criteria pursued (i.e., expressed in the objective function) in each revised model is also shown in Table 1. Most of the objectives refer to economic parameters (or are related to them) that have to be minimized or maximized (e.g., minimize cost, maximize revenue or profit, minimize work-in-process and overtime work, minimize the overuse of resources, maximize utilization rate) while some are time-based objectives (e.g., minimize makespan, minimize delays, meet due date).

2.3.3. Modelling issues

The use of nonregular capacity is an important characteristic of the tactical planning level because it provides capacity flexibility to meet demand. Ideally, this flexibility should support the planner in making a trade-off between the expected delivery performance and the expected costs (Hans et al., 2007). In general, the basic capacity flexibility options are overtime, hiring personnel, temporary labour, subcontracting, and increasing inventory levels (Mincsovics and Dellaert, 2009; Jamalnia and Feili, 2013; Lingitz et al., 2013).

In manufacture-to-order production (such as ETO companies) and for service providers, the use of inventories to fill eventual gaps when adapting capacity to demand is commonly not an available option as it is for MTS manufacturers (Ballestin et al., 2007; Lusa and Pastor, 2011). However, capacity flexibility may be efficiently achieved by adapting labour capacity to adjust the workforce and working time (Sillekens et al., 2011). The use of overtime could be a means to create this flexibility, particularly when the use of contingent labour is not an alternative to temporarily increase capacity, especially in production settings that require skilled workers (Alp and Tan, 2008).

Table 1 presents many models within the MTO context that consider nonregular capacity, such as working overtime, subcontracting, outsourcing and personnel hiring. However, research studies that address the ETO or the multi-project context do not completely explore any of these types of working time flexibility.

The aggregation of processing stages is usually necessary when representing resources and operations to achieve manageable problems (Monostori et al., 2010; Alfieri et al., 2011). Working on an aggregate level is common in production planning for two reasons: (1) the detailed production process is sometimes unknown during the planning phase and (2) the dimension of the planning problem, especially with respect to the planning horizon, can be reduced in situations where the number of detailed processing stages is prohibitively large (Alfieri et al., 2011).

All aggregate production planning techniques face the problem of a trade-off between the model's accuracy in capturing the relevant features of the production-planning environment and the resulting model complexity (Nam and

Logendran, 1992). The way that aggregate activities are built is crucial to the feasibility and quality of the production plans (Monostori et al., 2010). For instance, many researchers have tried to simplify the planning problem by considering only the bottleneck machines, focusing on a single or a two-stage system (Mestry et al., 2011).

In most of the reviewed research studies shown in Table 1, aggregate activities that correspond to whole production phases are modelled in order to represent multiple processing stages. Some models require even more aggregate data by representing products as one stage (i.e., one aggregate stage), whereas others focus on the bottleneck stage.

The representation of the production flow in the tactical planning level refers not only to the definition of the aggregate activities but also to the precedence relationships among these activities. The most common, easy-to-model temporal relationships between activities are the classical finish-to-start (i.e., given a pair of activities (i,j), it prescribes that activity j can start only after a finishing time of a predecessor activity i) (Markus et al., 2003; Bianco and Caramia, 2011).

Moreover, the generalized precedence relationships (GPRs) (i.e., comprising the following types: start-to-start, start-to-finish, finish-to-start and finish-to-finish) permit minimum and maximum time lags between pairs of activities but also assume that the effort associated with a given activity is constant overtime (Schwindt and Trautmann, 2000; Neumann and Schwindt, 2002; Bianco and Caramia, 2011, 2012). Quintanilla et al. (2012) extend the classical concept of time GPRs to the concept of work GPRs between pairs of tasks that consider work percentages for both tasks involved. Nevertheless, in a production environment, activities may not only overlap but also vary in intensity over time (e.g., starting with low intensity and gradually increasing it.) In these situations, the traditional finish-to-start or the GPRs cannot completely represent these temporal constraints among activities (Alfieri et al., 2011; Bianco and Caramia, 2011).

An alternative to cope with this issue is the use of variable intensity activities (i.e., activities that vary in intensity over time until they are completed) and feeding precedence constraints (i.e., constraints that allow some overlap in the execution of the connected activities and capture the flow of material between them (Kis, 2005)). Kis (2005) defines a single type of feeding precedence relationship to constrain an activity to start only after a certain percentage of its predecessor activities have been completed. Alfieri et al. (2011) extend this work by developing three other types of feeding precedence relationships in an ETO context. Moreover, Schwindt and Paetz (2015) introduce the concept of generalized feeding precedence relationships (GFPRs), which addresses continuous pre-emptive resource constrained project scheduling problems (i.e., activities may be interrupted for several reasons). In this case, given a pair of activities (i,j), the GFPRs require that a specific final portion of j can be started in a pre-determined time interval after the earliest point at which a given portion of activity i has been executed. This formulation represents all the former cases mentioned (i.e., finish-to-start, GPRs and feeding).

Most of the reviewed research studies shown in Table 1 refer to models that make use of precedence constraints between activities and, within these models, most make use of the typical precedence relationship (i.e., finish-to-start). Moreover, the reviewed papers that adopt a variable intensity activities and feeding precedence relationships refer to ETO or project-oriented contexts. This seems justifiable in these contexts because the aggregate activities usually refer to overlapping production stages with long processing times.

The occurrence of uncertain events can have a significant impact on the stability and performance of the production system. Many real-world planning problems involve noisy, incomplete or inaccurate data, which constitute uncertainties in terms of demand, revenues, costs, production rate and capacity (Hans, 2001; Mulvey et al., 1995; Alem and Morabito, 2012; Aouam and Brahimi, 2013). Within the revised models in Table 1, demand is the most common uncertainty parameter considered (Genin et al., 2008; Aghezzaf et al., 2010; Lusa and Pastor, 2011; Alem and Morabito, 2012; Zhen, 2012; Aouam and Brahimi, 2013; Rahmani et al., 2013). Costs (Alem and Morabito, 2012; Rahmani et al., 2013), work content (Wullink et al., 2004), resource need (Corti et al., 2006; Tolio and Urgo, 2007; Alfieri et al., 2012), processing times (Ebben et.al, 2005) are contemplated in other papers. One particular research study (i.e., Munhoz and Morabito, 2014) refers to a citrus industry, where the uncertain parameter "fruit acidity" is modelled since it disturbs production.

Production planning is a sequential decision process which addresses a planning horizon subdivided in time periods in a way that it could be modelled as a multi-stage process (Tolio and Urgo, 2007). In this sense, to simplify the mathematical formulation of whole multi-stage problems, it could be reasonable to avoid considering uncertain events which are far away in the future (i.e., in distant periods) as their relevance in terms of the short-term decisions is normally low (Tolio and Urgo, 2007).

2.3.4.Application approach

In conclusion, the final group of attributes is related to the application approach (i.e., the context and the empirical nature). In this particular review, the majority of the research studies refer to the MTO context. In some papers, the context of application is not described. Only seven of the 33 papers concern the ETO context or a multi-project environment explicitly. Nevertheless, Little et al. (2000), Gosling and Naim (2009), Yang (2013), Grabenstetter and Usher (2014) and Powell et al. (2014) emphasize the demand for solution approaches to support management in the ETO context.

As for the empirical nature of the research studies, the reviewed papers are classified into three groups, adapted from Jahangirian et al. (2010) and defined as follows: (1) real problem-solving papers (i.e., the model has been applied to a real problem with real data) – “Class A”; (2) hypothetical problem-solving papers (i.e., the model has been applied for the purpose of solving a real-world problem, but using artificial data rather than real data) – “Class B” and (3) methodological papers (i.e., research is conducted to enhance the methodological approach regardless of any specific application area) – “Class C”.

The summary presented in Table 1 reveals that most of the papers are methodological research studies and are based on illustrative examples not linked to real-world contexts. Some authors have conducted hypothetical problem-solving research (Markus et al., 2003; Genin et al., 2008; Sawik, 2009; Aghezzaf et al., 2010; Monostori et al., 2010; Tolio and Urgo, 2007; Tunali et al., 2011; Alem and Morabito, 2012; Rahmani et al., 2013), and only three papers applied the model to a real industrial environment (Alfieri et al., 2011, 2012; Munhoz and Morabito, 2014). This lack of real-world cases corroborates the academic literature (e.g., Ramezanian et al., 2012) and highlights the recognized gap in the APP literature between theory and practice (Buxey, 2003, 2005; Corti et al., 2006; Ramezanian et al., 2012; Sharda and Akiya, 2012; Jamalnia and Feili, 2013; Lingitz et al., 2013; Liu et al., 2013; Diaz-Madroñero et al., 2014). It is here where this thesis intends to make one of its main contributions by developing and applying two tactical capacity planning models in a real-world ETO production setting, thus offering the literature new perspectives in modeling the problem and reducing the current research-practice gap existent in the academic literature

2.3.5. Summary of the literature review

This review has covered literature relating to the use of mathematical models that support decision makers to solve middle-level managerial capacity planning problems in manufacture–to-order (i.e., make- or engineer-to-order) context. Some issues are highlighted from this review:

- Only seven out of the 33 papers reviewed explicitly address the ETO or multi-project environment. This is in accordance with Gosling and Naim (2009) and Yang (2013) that highlighted that ETO contexts have received much less attention from the researchers when compared to standardised make-to-stock contexts.
- Several authors (Giebels, 2000; Gademann and Schutten, 2005; Hans et al., 2007; Sawik, 2009; Montreuil et al., 2013) emphasize that, at the negotiation and order acceptance stage, adequate capacity planning methods that assess the consequences of decisions for the production system are crucial. Within the papers that address the ETO or multi-project context, only one concerns the order acceptance phase (i.e., the workload analysis and due date setting). Most of the analysed papers focus on the resource-loading phase (i.e., the capacity adjustments and allocation of orders).
- In ETO contexts, projects are subject to considerable uncertainty (Herroelen and Leus, 2004; Corti et al., 2006; Tolio and Urgo, 2007; Chtourou and Haouari, 2008; Van de Vonder et al., 2008; Deblaere et al., 2011; Alfieri et al., 2012; Artigues et al., 2013). Under these

conditions, these authors support that a planning model should aim at solution robustness or stability by incorporating a certain degree of anticipation of variability (uncertainty). On the other hand, the majority of the reviewed models are classified as model type "deterministic" and, within the 13 models that consider uncertainties, only two of them are ETO models.

- Tactical capacity planning methods ideally should use capacity flexibility (nonregular capacity) to support the planner in making a trade-off between the expected delivery performance and the expected costs of exploiting this flexibility (Hans et al., 2007). However, in terms of the use of nonregular capacity options, only a few consider changes in workforce level, whereas changes in production capacity are generally modelled through the use of overtime or subcontracting. Moreover, the research studies that address the ETO or the multi-project context seem not to completely explore this kind of working time flexibility.
- To represent the production flow in an aggregate form and achieve a manageable planning problem, the aggregation of processing stages and the representation of the precedence relationship among these stages may be necessary. Under this topic, most of the reviewed research studies refer to multiple aggregate activities and make use of the finish-to-start precedence constraints. Moreover, the models that adopt variable intensity activities and feeding precedence relationships refer to ETO or project-oriented contexts. Nevertheless, there is always a trade-off between the model complexity and its accuracy in capturing the relevant issues of the production planning environment (Nam and Logendran, 1992) and the way that aggregate activities are built is crucial to the feasibility and quality of the production plans (Monostori et al., 2010).
- There is a gap in the APP literature between theory and practice (Buxey, 2003, 2005; Corti et al., 2006; Ramezanian et al., 2012; Sharda and Akiya, 2012; Jamalnia and Feili, 2013; Lingitz et al., 2013). In fact, most of the papers reviewed are methodological

researches and are based on illustrative examples not linked to real-world contexts.

In correspondence to the aforementioned issues, this thesis intends to contribute to the academic literature by proposing two tactical capacity planning models (one deterministic and one robust), for a real-world ETO setting, that supports the order acceptance phase, explores capacity flexibility, adequately represents the production flow and incorporates uncertainties in the generated plans aiming at a solution robustness or stability.

3 Industrial problem and research method

This chapter is organized in two sections. Section 3.1 characterizes the tactical capacity planning problem in the real-world ETO production setting studied by describing this company´s planning process and its major shortcomings. Furthermore, Section 3.2 details the research method adopted to produce means for reducing these shortcomings.

3.1. The industrial problem

The real-world production setting considered in this study is a medium-sized project-driven organization that produces a wide range of customized equipments, such as high-pressure boilers and sophisticated reactors, based on ETO policy. The production system is organized into five work centres that correspond to the major manufacturing processes (cutting, stamping, machining, assembly and welding), which require a skilled, dedicated and expensive workforce. The products manufactured are of high value and have complex structures (many components and production stages) with long lead times (five to 18 months).

This company´s main customers are large enterprises in the oil & gas supply chain that have high bargaining power. These customers demand strict conformity with product specification and due date compliance. In general the ordered product will be part of a large project in the sphere of action of the customer which means that delays produce major disruptions. Failure to meet due dates may ultimately disqualify the service provider and this is a major concern for the company.

Figure 3 presents the planning phases and the major functions of this company´s planning process. At first, customers' tenders are analyzed to determine order acceptance or rejection. In the workload analysis, the company's industrial manufacturing director and his management team (i.e., the manufacturing planning team) compare the customers' demands with the available shop floor production capacity. The potential workload associated with not-yet-confirmed orders is taken into consideration here, and different scenarios are created considering subjectively each project's chance of acceptance as the negotiation stage of each project evolves. The main goals of this analysis are balancing the demand with the available capacity and providing information for an eventual bid preparation. The order processing phase begins when the project is accepted and confirmed. During this phase, the project's activities are detailed and scheduled for each part component. A resource loading analysis is also performed to enable feasible lower-level schedules for production scheduling, which is essentially operational.

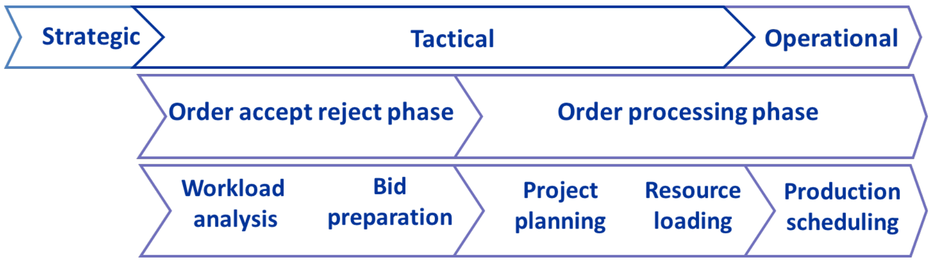


Fig. 3: Planning process in the studied company

In the workload analysis process, which is the focus of this study, the manufacturing planning team is faced with questions such as the following: is it possible to accept a particular set of projects given the burden of already committed orders? When should the production activities be executed? Can the delivery dates imposed by the customers be met? Would overtime hours be needed or would it be necessary to hire more personnel? Which part components could be subcontracted? How much would it cost to manufacture this set of projects?

In other words, the manufacturing planning team must determine if a new order should/could be accepted based on the available capacity, decide on the usage of nonregular capacity and estimate production costs. Attempting to answer these questions, this team manually creates an aggregate production plan to define what demands will be processed and in which time periods. Since information is not accurate nor detailed in this planning stage, as product designs have not yet been conceived, management makes use of rough data, based on historical information on former projects. This is done to estimate processing times in order to define milestones and due dates for these potential incoming projects.

As this is a multi-project planning problem (i.e., projects compete for the same production resources), this plan is developed by assessing simultaneously the demand and the available capacity in order to accommodate overloads through capacity adjustments. In this planning process, demand is classified into categories: (i) committed workload (i.e., the ongoing projects that have been confirmed, detailed and released to the shop floor), (ii) committed new projects (i.e., projects which have been confirmed but have not been detailed enough by the design engineers neither released to production) and (iii) proposals (i.e., projects, in the order accept/reject phase, which do not have specified yet a release date or a deadline with exactness). The committed new projects and the proposals are sensitive to uncertainty referring to the estimated processing times (i.e. since design phase is not concluded). For the committed new projects, this issue is even more critical since their deadlines have already been set. Moreover this is particularly crucial in the short-term when the internal production capacity is assumed fixed, as the process of hiring and training personnel often takes from two to three months.

Due to demand instability that is typical of the ETO context, there are time periods with high levels of workload and other periods where capacity is underutilized. To deal with a possible lack of production capacity, the manufacturing planning team may adopt different strategies to cope with this problem. As the planning focus is the tactical level, high investments in infrastructure, machinery or even process technology are hardly possible. In this planning stage, meticulous adjustments of the internal capacity level are done by authorizing overtime or hiring more operators. Nevertheless, the maximum load capacity limit must be observed. In other words, it may be useless to contract workforce beyond what the equipments and facilities can take.

As to the workforce adjustments, it is not reasonable to adjust the number of employees according to demand oscillations by constantly hiring or dismissing personnel. Investments in training are frequently done but should not be wasted. Moreover, the workforce in this production setting is usually not cross-trained (i.e., the workers execute specific tasks only). Depending on the personnel qualification, it may be difficult to hire certain professionals (e.g., welders, boilermakers). In this sense, the manufacturing planning team tries to maintain some stability in the workforce by minimizing layoffs or hiring personnel.

Subcontracting of product components is also a strategy adopted by this company. However, not all activities of a project can or should be subcontracted. There may not exist a provider with the necessary know-how, or it can be a strategic business decision not to subcontract a specific part of the project. When planning, this limitation is taken into account.

This company has a formalized rough-cut capacity planning method based upon spreadsheets to carry out the workload analysis. Yet, there are major shortcomings that need to be addressed in this company's current planning method to properly support decision making in the tactical planning level:

- Although there are many alternative production plans (i.e., by combining sets of customer orders and the use of nonregular capacity options) to be evaluated, only a few are generated and assessed due to the time-consuming scenario-building process manually performed.
- A cost analysis is not carried out thoroughly while comparing the alternative plans. In this sense the company manufacturing team faces difficulties to decide on the different strategies while planning. Nevertheless, the cost assessment is performed afterwards, but only for the chosen strategy.
- In this aggregate production plan, capacity is viewed as a whole (i.e., capacity is not disaggregated among the five work centres). The multiple processing stages and the precedence relations among the production activities are not taken into consideration. This planning approach, on the one hand, reduces and simplifies the planning problem but, on the other hand, provides rough information that makes it difficult to define a plan for a specific workforce when adjusting internal capacity. This is especially crucial because the workers are highly qualified and not cross-trained.

In essence, these issues need to be addressed when developing an appropriate solution to cope with the aforementioned planning problem. The following section describes the research method adopted to develop and implement a tactical capacity planning solution to support decision making in this industrial setting.

3.2. Research method

The studied problem is characterized by an exploratory approach in which actions and results interact and are mingled in practice. In this sense, the action research method seemed the appropriate choice to organise and direct the investigation. It investigates more than actions: it is participatory, it occurs simultaneously with the action, and it is a sequence of events and approaches used to solve problems (Coughlan and Coghlan, 2002). The objective here is to explore, in a broad sense, the process of modelling and implementing a tactical capacity planning solution in an ETO context to uncover problems.

This research began with a literature review on the production planning issues within the ETO context. The studied company was chosen because it is an ETO with a complex decision-making process. Its manufacturing planning team participated throughout the entire research to improve their planning methods. Within this action research, the researcher made four technical visits to the industrial site over an 18-month period. Between these visits, the researcher gradually developed the proposed planning model and interacted with the company manufacturing planning team to gather additional information and validate the proposed solution model. This procedure is in accordance with the action research developed by Carvalho et al. (2014).

In the first technical visit, the researcher validated assumptions gathered from literature relative to the ETO context and the manufacturing planning team presented, in a comprehensive manner, their planning methods, from the tactical to the operational level. Some shortcomings were highlighted in this presentation, specially referring to the tactical planning level activities, processes and methods. Therefore, in between the first and the second visit, a literature review on tactical planning issues was conducted to gain new insights on how to contribute to improve the company's methods.

The objective of the second visit was to define the problem scope within tactical planning that would be addressed in the action research. In the studied company, tactical planning comprises a set of linked functions in the planning process (i.e., workload analysis, bid preparation, project planning and resource loading) that use distinct tools with specific objectives. A semi-structured interview to identify the core problem to be studied was conducted and was based on the following questions: what decisions are made in the tactical planning level? Are they related? What are the problem sizes (i.e., how aggregated are the data)? What tools are used? What are the major difficulties encountered? What are the major uncertainties? What are the capacity flexibility options considered?

From this second visit, it became clear that the workload analysis (i.e., the planning function that attempts to generate a production plan by levelling demand to the available capacity) in the order acceptance phase was the problem to be addressed. The current workload analysis presented relevant shortcomings that had to be overcome, as its outcomes influence the accept/reject decision as well as the following steps (e.g., bid preparation, project planning and resource loading). Furthermore, the industrial manufacturing director emphasized that the proposed solution should represent the planning problem in a coherent aggregation level when modelling the planning entities (e.g., demand, activities, resources, time). His observations and suggestions were essential to simplify the problem and to cope with the lack of detailed information in this planning phase. By the end of this technical visit, a first version of the conceptual model and the required input data were outlined.

Between the second and the third visits, a literature review was conducted on mathematical models used for tactical planning problems to gain insights on how to represent the studied planning problem. Concurrently, the researcher prepared presentations for the conference calls conducted with the company's manufacturing planning team. The objective was to organize the discussions on the model`s assumptions and constraints. Data were collected through these conference calls and through email exchanges in order to fill the database with real-world data. In parallel, the first version of the proposed mathematical model was developed using Aimms 3.13.

The objectives of the third visit were to present the first version of the proposed model using real-world data and to identify required adjustments in the model. During this visit, it became clear that the feeding precedence modelling adapted from Monostori et al. (2010) to represent the activities flow was not appropriate. A thorough discussion was carried out to identify the needed inputs for the model in order to represent the relationship among the interconnected activities.

Historical data related to former projects' "production curves" (i.e., the accumulated workload informed along the projects' executions) were gathered and analysed together with the company team. At the end, a typical "production curve" was identified to represent the relative pace among the activities within the same project.

Between the third and the fourth visits, the researcher focused on the development of the cadence constraints to properly represent the typical "production curve". For this reason, conference calls were held in a highly iterative process to validate these constraints. Additionally, in this period the researcher gathered additional and updated data to create a real-world database with the customers' demands that the manufacturing planning team wished to simulate in the workload analysis.

During the fourth visit, the manufacturing planning team evaluated the improved version of the proposed model running the real-world database. The main validations were whether the model adequately represented the planning problem and whether this model addressed the shortcomings highlighted in the company´s current planning method. Alternative scenarios were also generated to assist the management board in the order acceptance phase. A validation session was conducted to obtain feedback and new insights from the company's team with respect to the results of this action research.

One of the aspects discussed in the fourth visit refers to the fact that the proposed model is deterministic while the context under study is characterized by many sources of variability, especially the ones related to the work content of the incoming projects and therefore the uncertainties related to the estimated processing times. In this sense, incorporating information about uncertainty in the generated production plans seemed to be the next step in this action research.

To address this issue, after the fourth visit, a robust optimization model was developed, based on Bertsimas and Sim (2004)'s approach, to enhance the former deterministic model by including process uncertainties. This model was tested using a real-world database and a set of scenarios was generated for the manufacturing planning team to assess the adherence of the model to the studied problem.

Finally, in the fifth technical visit, a final validation session was held where the planning team provided feedback relative to the application of the robust model to the real-world planning problem by analyzing the set of scenarios generated. Moreover, within this session, several issues were discussed relative to the implementation of the robust model and the expected impacts of this planning solution.

4 Deterministic approach

This chapter presents the deterministic mathematical programming model that addresses the tactical capacity planning problem studied. This model comprises modelling issues that are either not entirely explored in other research studies or that have to be adapted to the specificities of this particular problem. Furthermore, as this research aims to reduce the research-practice gap in tactical capacity planning, the findings from the application of a preliminary version of the model to a real-world problem, formerly published in Carvalho et al. (2015), are also presented. In this sense, this chapter is organized in two sections (i) a description of the proposed deterministic optimization model and (ii) a presentation of its application.

4.1. The proposed model

This section begins by presenting the main characteristics and assumptions of the proposed deterministic model. In the sequel, its mathematical formulation is detailed and a discussion is presented comparing the model with what has been found in literature concerning the modelling issues highlighted in Table 1.

4.1.1. Introduction to the model

The proposed solution is a tactical capacity planning Mixed-Integer Linear Programming (MILP) model that supports the order acceptance phase by optimally balancing demand with the available capacity and providing information for an eventual bid preparation. It is a cost minimization formulation of an ETO production system that provides an optimal production plan, subject to several problem-related constraints.

Although, in the real world, production planning problems occur in systems that operate indefinitely, decisions must be based on limited information about the future for practical reasons. Additionally, forecasts for remote future tend to be of limited use. In this sense, the studied tactical capacity planning problem is a multi-period problem with a planning horizon that is subdivided into a finite number of equal sized time periods. In terms of application, a rolling horizon planning approach (in which the generated plan is recalculated on a regular basis, or when significant events occur) should be adopted in a way that the planner solves the finite horizon problem and implements the solution only for the more imminent periods. In the following period, the planner solves a new problem with the same length of the planning horizon, yet considering updated input data. A relevant issue concerning this topic refers to the dimension of the planning horizon as it reflects how far into the future forecasts must be made to make optimal first period decisions. The planning horizon should not be excessively long to require unreliable data relative to very distant projects. Nevertheless, its size should be long enough to avoid myopic solutions especially since demand in all subsequent time periods after this horizon is zero.

As this is a multi-project planning problem, demand corresponds to the already committed workload and to new projects that compete for the same resources. In practice, in a rolling horizon approach, the committed workload refers to an existing plan of confirmed orders that is updated and extended with the new projects. The committed workload is assigned to the time periods originally planned. Additionally, each new project has its own negotiated release date (earliest starting time) and the customer's deadline. Because this is a time-driven planning approach, one of the model's assumptions is that backlogging is not allowed and, therefore, deadlines must be met.

Each project consists of a set of components that have specific time windows defined within the project's release date and deadline. A component comprises a set of activities and each activity can be regarded as an aggregation of operations or processing steps within the same type of production process. Furthermore, each work centre comprehends the production resources (i.e., machines and tools) within a type of production process. Therefore, activities may be allocated to work centres in an aggregate level of production planning that is coherent to the tactical planning level. In this sense, the proposed model admits, even in this aggregate level, the representation of a production flow with multiple processing stages.

An activity usually takes several months to be completed. Furthermore, its execution progress mode varies along the time periods. For instance, in this specific problem, an activity typically starts in a low-intensity level mode that increases to a maximum level and then it decreases until it is finally completed. The fraction of an activity done in a time period t is called the intensity of the activity in time period t. Hence, the intensity of an activity is at most one, and the sum of intensities of an activity sums up to one. Usually, there is a limitation on the maximum intensity that may be completed in a single time period and a minimum intensity to guarantee that a project is executed with no interruptions. To represent this issue, the model makes use of the variable intensity activities formulation, which means that each activity may vary its intensity over time until it is completely done.

Activities are interconnected as they represent a sequence of multiple processing stages. Furthermore, activities overlap in time successively, particularly in the production of complex components. Additionally, each activity has its own pace but depends on the evolution of the others that are interconnected. In this sense, the company's planning team refers to milestones, which represent reference points that indicate the completion of phases within a project's execution. The cadence (i.e., the rhythm) of the production flow as a whole is determined by these milestones. Table 2 presents historical data from the company's former projects that specify, for each milestone, the accumulated intensity of five interconnected activities (cutting, stamping, machining, assembly and welding) of a given component. For example, if the workload processed relative to the cutting stage ranges from 55% to 60% (i.e., between milestones 9 and 10), this means that 42 to 47% of stamping, 18 to 20% of machining, 23 to 25% of assembly and 19 to 21% of welding must be concluded. Furthermore, considering this overlapping behaviour according to this cadenced production flow, it is not always possible to establish a precedence relation among many of these activities. In other words, the proposed model admits that the precedence relationship between two activities is not fixed, as one may precede in some periods of time and succeed in other periods (see, for example, in Table 2, that assembly precedes machining until milestone 13 and afterwards it succeeds machining).

Table 2: Cadence data – Accumulated intensities (Data provided by the company)

Milestone	Cutting	Stamping	Machining	Assembly	Welding
0	0.010	0.003	0.002	0.004	0.003
1	0.060	0.018	0.009	0.026	0.019
2	0.220	0.066	0.032	0.097	0.070
3	0.242	0.090	0.050	0.105	0.078
4	0.289	0.140	0.070	0.124	0.096
5	0.346	0.200	0.095	0.148	0.118
6	0.394	0.250	0.115	0.167	0.136
7	0.451	0.310	0.140	0.191	0.157
8	0.498	0.360	0.160	0.210	0.175
9	0.555	0.420	0.184	0.234	0.197
10	0.603	0.470	0.205	0.253	0.215
11	0.650	0.520	0.225	0.273	0.233
12	0.691	0.549	0.299	0.327	0.283
13	0.729	0.577	0.369	0.371	0.327
14	0.773	0.609	0.449	0.421	0.377
15	0.812	0.637	0.519	0.465	0.421
16	0.830	0.648	0.552	0.484	0.438
17	0.850	0.674	0.602	0.521	0.472
18	0.890	0.720	0.692	0.589	0.533
19	0.910	0.745	0.742	0.626	0.567
20	0.930	0.771	0.792	0.664	0.601
21	0.960	0.844	0.875	0.778	0.740
22	0.980	0.935	0.951	0.929	0.916
23	1.000	1.000	1.000	1.000	1.000

In terms of internal capacity, the availability of working hours per period per work centre is proportional to the number of employees allocated in each work centre. Average workforce productivity is considered in order to define the processing times for the activities. Furthermore, the proposed model admits the use of capacity flexibility by considering nonregular capacity alternatives. In this sense, the availability of working hours may be modified by the use of an overtime working shift or by hiring or firing personnel. In order to minimize capacity changes though, the model considers a minimum employment period that restricts dismissing personnel before this period of time and associates a capacity change cost proportional to the number of employees hired or fired.

To resemble reality, the model assumes that, at the beginning of the planning horizon (which corresponds to the fixed capacity periods, a short-term period perspective), changing capacity levels by hiring and firing personnel is not permitted. This assumption seems reasonable as this type of capacity flexibility takes time to be implemented due to the scarcity of specialized workforce available and the necessary training time of the new employees.

Besides the internal capacity, the model also considers external resources through subcontracting, which may incur in higher production costs. These external resources do not have a capacity limitation. One of the model's assumptions is that a component is either all subcontracted (all of its activities are made externally) or all made internally. Furthermore, not all components may be subcontracted for several reasons and this restriction is input information in the model.

As this is a cost minimization optimization model, several cost parameters are considered to calculate the overall production costs. As there might be a significant variation in terms of costs among the production processes, each work centre has its specific hourly production processing cost (relative to the effective use of the production resources, involving energy and lubricants consumptions costs) and overtime cost as well as an average salary (which effectively represents the differences between the lower and higher salary values in a specific work centre). On the other hand, for simplification purposes, a fixed capacity change cost is considered to estimate the cost of hiring and firing employees, regardless of the affected employees' backgrounds (e.g., salary, benefits, rights, employment contract, working shift)

The mathematical formulation of the proposed model is detailed in the sequel.

4.1.2.Mathematical formulation

The sets, parameters and decision variables are presented in Tables 3, 4 and 5, respectively.

Table 3: Sets

a	activities
b	activity types
g	milestones
i	part components of a project
t,l	time periods
w	work centres

Table 4: Parameters

AP_{ia}	1 if activity a belongs to part i, 0 otherwise
AT_{ba}	1 if activity a is classified as activity type b, 0 otherwise
C_{bg}	accumulated intensity of activity type b in milestone g
CAP_w	maximum number of working hours per period at work centre w
CC	capacity change cost relative to hiring or firing one employee
CO_w	average salary per hour of an employee working an overtime hour at work centre w
CR_w	production processing hourly cost at work centre w
CS_w	average salary of an employee working at work centre w
DL_a	deadline of activity a
FC	number of fixed capacity periods
M	“big M”, a sufficiently large number
ME	minimum employment period
MN_a	minimum intensity of activity a in any time period
MX_a	maximum intensity of activity a in any time period
NE_w	number of employees initially allocated at work centre w
NP	number of periods in the planning horizon
OHE_w	number of overtime hours per employee per period at work centre w
PS_i	price of the subcontracted part i
Q_{aw}	processing time of activity a at work centre w
RD_a	release date of activity a
RHE_w	number of regular working hours per employee per period at work centre w
WH_{wt}	number of hours relative to the committed workload allocated to work centre w in period t
XS_i	1 if part i may be subcontracted, 0 otherwise

Table 5: Decision variables

c_a	1 if activity a is processed in-house, 0 otherwise
d_i	1 if part i is processed in-house, 0 otherwise
e_{at}	1 if activity a has already started in period t or in an earlier period; 0 otherwise
ea_{wt}	number of employees allocated at work centre w in period t
ef_{ltw}	number of employees hired in period l and fired in period t at work centre w
eh_{ltw}	number of employees hired in period l and still working in time period t at work centre w
eo_{wt}	number of employees working overtime hours at work centre w in period t
er_{wt}	number of employees working regular hours at work centre w in period t
f_{at}	1 if activity a is entirely processed by period t; 0 otherwise
n_{igt}	1 if part i has completed at least milestone percentual g in period t; 0 otherwise
o_{at}	number of overtime hours processing activity a in period t
r_{at}	number of regular working hours processing activity a in period t
s_{at}	number of subcontracted hours processing activity a in period t
w_{at}	1 if activity a is processed in period t; 0 otherwise
wr_{wt}	number of regular hours relative to the committed workload allocated to work centre w in period t
wo_{wt}	number of overtime hours relative to the committed workload allocated to work centre w in period t
x_{at}	intensity of activity a in period t
z_{at}	accumulated intensity of activity a until period t

The objective function minimizes the overall variable production cost involving production processing (i) and overtime costs (ii) associated with the incoming orders, the production (iii) and overtime costs (iv) of the committed workload, capacity change cost (v), personnel payroll (vi) and subcontracting costs (vii).

Minimize

$$\sum_{a,w,t}[CR_w\,(r_{at}+o_{at})+CO_w o_{at}]+\sum_{w,t}[CR_w\,(wr_{wt}+wo_{wt})+CO_w wo_{wt}+CC(eh_{ttw}+\sum_{l}^{t} ef_{ltw})+CS_w\,ea_{wt}]+\sum_{i}PS_i\,(1-\,d_i)$$

(i) (ii) (iii) (iv) (v) (vi) (vii)

This function is subject to several constraints that are detailed next.

Release date and deadline

The demand is composed by projects that have time windows defined by the release date and the customers' deadlines. All projects consist of components that comprise a set of activities. As an activity may take months to be executed, the model uses variable intensity activities. The processed fraction of an activity (x_{at}) in a time period *t* is called the intensity of the activity in *t*. This intensity is at most one, and the sum of intensities of an activity sums up to one. Constraint 1 ensures that an activity is entirely processed in its time window.

$$\sum_{t=RD_a}^{DL_a} x_{at}=1 \qquad \forall a \tag{1}$$

Non-interruption flow

Constraints 2 to 4 are the non-interruption constraints that guarantee that once started, an activity has to be processed in all subsequent periods until it is finished. When activity *a* starts at time *t*, the binary variable e_{at} takes value 1. When activity *a* is finished at time *t*, the binary variable f_{at} assumes 1. While activity *a* is being processed, the binary variable w_{at} becomes 1.

$$e_{at}\geq\sum_{l=1}^{t} x_{al} \qquad \forall a, RD_a\leq t\leq DL_a \tag{2}$$

$$f_{at}\leq\sum_{l=1}^{t} x_{al} \qquad \forall a, RD_a\leq t\leq DL_a \tag{3}$$

$$w_{at}=e_{at}-f_{at} \qquad \forall a,t \tag{4}$$

Maximum and minimum intensities

Constraints 5 and 6 ensure, respectively, that the intensity of an activity can never be less than the minimum intensity (MN_a) to guarantee no interruptions or more than the maximum intensity (MX_a) permitted in a single time period while being processed.

$$x_{at} \geq MN_a w_{at} \qquad \forall a, RD_a \leq t \leq DL_a \tag{5}$$

$$x_{at} \leq MX_a w_{at} \qquad \forall a, RD_a \leq t \leq DL_a \tag{6}$$

Cadence constraints

In this problem, the overlapping of production activities occurs throughout nearly the entire project's duration. Furthermore, each activity has its own cadence and depends on the evolution of the others. On the other hand, it is not always possible to establish a precedence relation among them, as some precede in some periods and succeed in others.

To represent this complex temporal relation, Constraints 7 and 8 guarantee that all activities of a component are performed at a corresponding pace. The variable z_{at} is the accumulated intensity of activity *a* in time period *t*. The parameter C_{bg} represents the accumulated intensity an activity type *b* should reach in a particular milestone *g*. The binary variable n_{igt} coordinates the paces of all activities that belong to a part component *i* to reach its respective intensity C_{bg} by time period *t*. In other words, n_{igt} only reaches 1 when all z_{at} that refer to part component *i* have reached their respective C_{bg} intensity by period *t*.

$$z_{at} \leq C_{bg} + Mn_{igt} \qquad \forall a, \forall g, RD_a \leq t \leq DL_a,\ AP_{ia} = 1,\ AT_{ba} = 1 \tag{7}$$

$$z_{at} \geq C_{bg} - M\,(1 - n_{igt}) \qquad \forall a, \forall g, RD_a \leq t \leq DL_a,\ AP_{ia} = 1,\ AT_{ba} = 1 \tag{8}$$

Regular hours, overtime and subcontracting

Each activity refers to a manufacturing stage and has to be processed on a specified work centre using regular and/or nonregular hours. The resource requirement of an activity for a given time period will be proportional to the intensity of the activity in that period (i.e., if the total resource requirement of activity *a* is Q_{aw} for work centre *w*, and the intensity of the activity *a* is x_{at} in period *t*, then it requires an amount of $Q_{aw}\, x_{at}$ from work centre *w* in period *t)*. Constraint 9 ensures balance within an activity's processing time and its fractions in terms of regular, overtime and subcontracting hours.

$$Q_{aw}\, x_{at} = r_{at} + o_{at} + s_{at} \qquad \forall a, w, t \mid Q_{aw} > 0 \tag{9}$$

One of the assumptions is that an activity is either entirely subcontracted (i.e., when the binary variable c_a takes value 0) or completely performed internally (i.e., when c_a equals one), as ensured by Constraint 10. Additionally, Constraint 11 guarantees that all the activities of a part component are either all subcontracted or all made internally by establishing a balance between the number of activities subcontracted of a given component and the number of activities that compose this component.

$$\sum_t s_{at} = Q_{aw}\,(1 - c_a) \quad \forall\, a, w, t \mid Q_{aw} > 0 \tag{10}$$

$$\sum_{a/P_{ia}>0} c_a = \Big(\sum_a AP_{ia}\Big)\, d_i \qquad \forall\, i \tag{11}$$

Furthermore, not all components may be subcontracted, and the binary parameter XS_i specifies this condition. If subcontracting is allowed, XS_i equals 1 and the binary variable d_i may take value zero (if component *i* is subcontracted) or value one (otherwise).

$$d_i + XS_i \geq 1 \qquad \forall\, i \tag{12}$$

Maximum number of working hours

Constraint 13 guarantees that the sum of the available regular hours plus the available overtime hours is limited by the maximum number of working hours per period at work centre *w*. When referring to subcontracting, the model assumes that the external resources do not have a capacity limitation.

$$\mathrm{RHE_w er_{wt}} + \mathrm{OHE_w eo_{wt}} \leq \mathrm{CAP_w} \qquad \forall\, \mathrm{w, t} \tag{13}$$

During the fixed capacity periods, the sum of all activities processed in a given work centre minus what is subcontracted must be equal to or smaller than the internal capacity minus the already committed workload. This is ensured by Constraint 14 (i.e., the workload constraint). This issue differs from the previous model published in Carvalho et al. (2015) which did not consider the fixed capacity periods.

$$\sum_{a} \mathrm{Q_{aw}}\, \mathrm{x_{at}} - \sum_{a} \mathrm{s_{at}} \leq (\mathrm{RHE_w} + \mathrm{OHE_w})\, \mathrm{NE_w} - \mathrm{WH_{wt}} \qquad \forall\, \mathrm{w, t} \leq \mathrm{FC} \tag{14}$$

Availability of employees

The availability of the internal capacity in both regular and overtime working shifts is proportional to the number of employees allocated in each work centre. Constraints 15 to 18 establish that the number of employees working in regular ($\mathrm{er_{wt}}$) and overtime hours ($\mathrm{eo_{wt}}$) cannot be greater than the number of employees available and allocated at a given work centre *w* in period *t* ($\mathrm{NE_w}$for the fixed capacity periods, otherwise $\mathrm{ea_{wt}}$).

$$\mathrm{er_{wt}} \leq \mathrm{ea_{wt}} \qquad \forall\, \mathrm{w, t} \mid \mathrm{t} > \mathrm{FC} \tag{15}$$

$$\mathrm{er_{wt}} \leq \mathrm{NE_w} \qquad \forall\, \mathrm{w, t} \mid \mathrm{t} \leq \mathrm{FC} \tag{16}$$

$$\mathrm{eo_{wt}} \leq \mathrm{ea_{wt}} \qquad \forall\, \mathrm{w, t} \mid \mathrm{t} > FC \tag{17}$$

$$\mathrm{eo_{wt}} \leq \mathrm{NE_w} \qquad \forall\, \mathrm{w, t} \mid \mathrm{t} \leq \mathrm{FC} \tag{18}$$

As the number of employees may change due to hiring and firing, the variable $\mathrm{ea_{wt}}$ is defined by the sum of the number of employees hired in earlier time periods and still working in period *t*, as stated by Constraint 19. Constraint 20 keeps $\mathrm{eh_{ltw}}$ updated because it is the number of employees originally hired in period *l* minus the sum of the number of employees that were hired in *l* and fired sometime between *l* and *t*. Furthermore, Constraints 21 and 22 guarantee that the number of employees hired in period *l* for work centre *w* will either decrease or be maintained along the future periods.

$$\mathrm{ea_{wt}} = \sum_{\mathrm{l}}^{\mathrm{t}} \mathrm{eh_{ltw}} \qquad \forall \mathrm{w,t} |\ \mathrm{t}\ > FC \tag{19}$$

$$\mathrm{eh_{ltw}} = \mathrm{eh_{llw}} - \sum_{\mathrm{m}}^{\mathrm{t}} \mathrm{ef_{lmw}} \qquad \forall\ \mathrm{w,l,t} |\ \mathrm{l} \leq \mathrm{t} \tag{20}$$

$$\mathrm{eh_{l\ t+1\ w}} \leq \mathrm{eh_{ltw}} \qquad \forall\ \mathrm{w,l,}\ \ \mathrm{t} < NP \tag{21}$$

$$\mathrm{eh_{l\ t+1\ w}} \leq \mathrm{NE_w} \qquad \forall\ \mathrm{w,l,t} \mid \mathrm{l} = \mathrm{FC}+1, \mathrm{t} = \mathrm{FC}+1 \tag{22}$$

Minimum employment period

To avoid employment instability, the model considers a minimum employment period (ME) that restricts dismissing personnel before this time interval. To ensure this, Constraint 23 defines that the number of employees hired in period *l* and still working in period *t* for work centre *w* must be the same as the number of employees originally hired in period *l* when the interval between *t* and *l* is inferior to the ME. Constraint 24 refers to an interval between *t* and *l* superior to the ME and determines that the number of employees hired in period *l* and still working in period *t* must be equal or inferior to the number of employees originally hired in period *l*.

$$\mathrm{eh_{ltw}} = \mathrm{eh_{llw}} \qquad \forall\ \mathrm{w,l,t} \mid\ \mathrm{l} \leq\ \mathrm{t} \leq \mathrm{l} + \mathrm{ME} \tag{23}$$

$$\mathrm{eh_{ltw}} \leq \mathrm{eh_{llw}} \qquad \forall\ \mathrm{w,l,t} \mid\ \mathrm{t} > l + ME \tag{24}$$

Committed workload

The committed workload associated with the fixed orders is also represented in the proposed model. The parameter $\mathrm{WH_{wt}}$ represents the number of hours relative to the committed workload allocated to work centre *w* in period *t.* On the other hand, there are two decision variables $\mathrm{wr_{wt}}$ and $\mathrm{wo_{wt}}$ that define the number of regular and overtime hours relative to the committed workload allocated to work centre *w* in time period *t*. Constraint 25 establishes the relation of the parameter with the two variables. In other words, this constraint distributes the committed workload allocated in a given time period, at a given work centre, and in its fractions in terms of regular and overtime hour use.

$$wr_{wt} + wo_{wt} = WH_{wt} \qquad \forall\, w, t \tag{25}$$

Constraint 26 ensures the balance between the incoming demand and the committed workload demand attended by regular time hours and the available regular time hours, considering the number of employees working in regular hours. Constraint 27 is equivalent to Constraint 26 but refers to overtime hours.

$$\sum_{a} r_{at} + wr_{wt} = RHE_{w} er_{wt} \qquad \forall\, w, t \tag{26}$$

$$\sum_{a} o_{at} + wo_{wt} = OHE_{w} eo_{wt} \qquad \forall\, w, t \tag{27}$$

Variable domains

$c_a \in \{0,1\} \quad \forall\, a$

$d_i \in \{0,1\} \quad \forall\, i$

$e_{at}, f_{at}, w_{at} \in \{0,1\} \quad \forall\, a, t$

$n_{igt} \in \{0,1\} \quad \forall\, i, g, t$

$ea_{wt} \in Z_{+} \quad \forall\, w, t$

$ef_{ltw}, eh_{ltw} \in Z_{+} \quad \forall\, l, t, w$

All other variables are nonnegative.

Figures 4a-4f summarize the aforementioned constraints of the proposed model in a framework. In these figures, the vertical axis represents the working hours, the horizontal axis refers to the time periods within the planning horizon, and the numbers presented refer to the constraint equations. In Figure 4a, the maximum number of working hours (Constraints 13 and 14) is the upper bound for internal capacity, which is available in terms of regular working hours (Constraint 9) and overtime working hours (also Constraint 9) that are proportional to the availability of employees (Constraints 15 to 22). Additionally, subcontracting (Constraints 9 to 12) working hours may be an alternative for acquiring additional capacity. Figure 4b adds to this framework the already committed workload (Constraints 25 to 27) which must be subtracted from the available capacity to accommodate the incoming orders. Figure 4c includes the incoming orders that are subject to a release date (Constraint 1) and a customer's deadline (also Constraint 1). These orders are subject to a production flow characterized by a non-interruption flow (Constraints 2 and 4), maximum and minimum intensities (Constraints 5 and 6) of the production activities and cadence constraints (Constraints 7 and 8) as displayed in Figure 4d. Figure 4e includes the minimum employment period (Constraints 23 and 24) policy which restricts the capacity changes in the model within the time periods. And finally, Figure 4f adds the fixed capacity periods where hiring and firing personnel is not permitted.

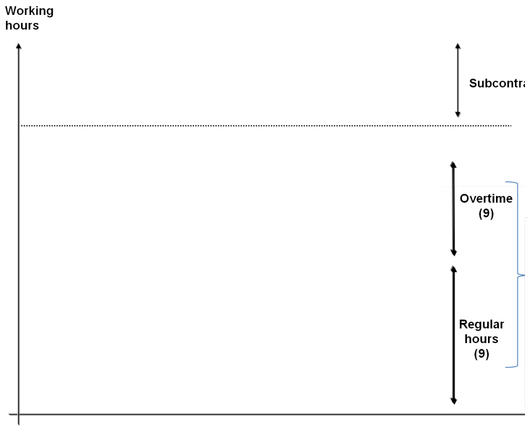


Fig.4a: Constraints 9-22

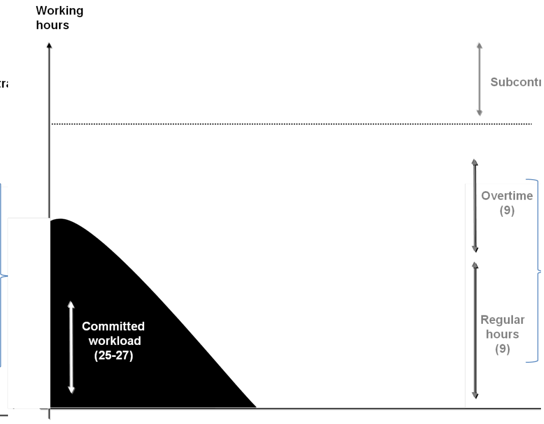


Fig.4b: Fig.4a + Constraints 25-27

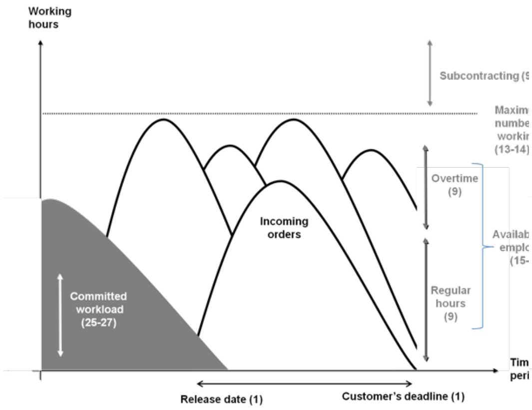


Fig.4c: Fig.4b + Constraint 1

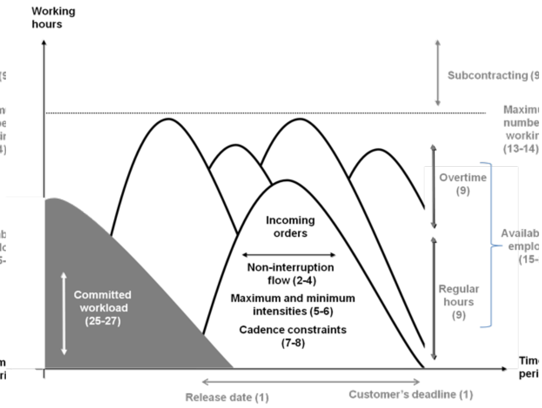


Fig.4d: Fig.4c + Constraints 2-8

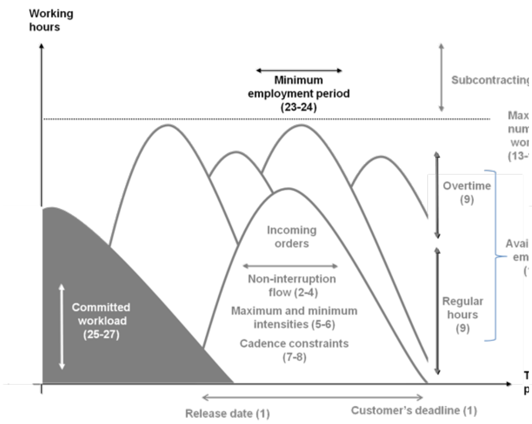


Fig.4e: Fig.4d + Constraints 23-24

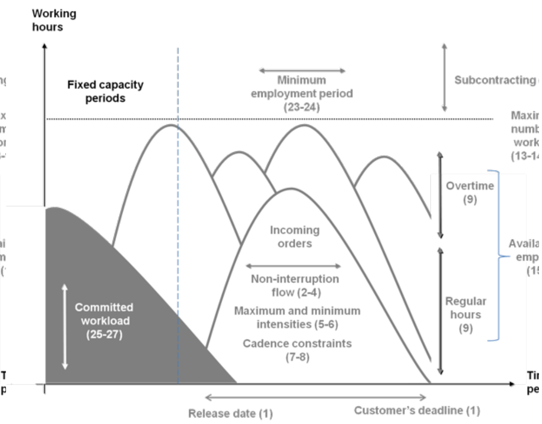


Fig.4f: Fig.4e + Fixed capacity period

Fig. 4: Proposed model's constraints

4.1.3.Discussions

As presented, the proposed solution is a tactical capacity planning MILP model that supports the order acceptance phase by optimally balancing demand with the available capacity and providing information for an eventual bid preparation. It is a cost minimization formulation of an ETO production system that provides an optimal production plan subject to several problem-related constraints. It includes capacity flexibility by considering nonregular capacity alternatives (i.e., overtime, subcontracting and hiring personnel). The model admits the representation of the production flow with multiple processing stages. For the precedence relationship among activities, the model permits an overlapping behaviour considering the cadence of the production flow. The main decision variables considered are the overall variable production costs and the intensities of the production activities throughout the planning horizon.

The literature offers models using exact solution methods to minimize costs in tactical planning problems (see Table 1). Likewise, the proposed MILP model can be solved using an exact solution method that minimizes the overall production costs. Although there are similarities between the reviewed models offered in the literature and the proposed model, this research extends the literature in the tactical planning area for the ETO context, a production environment that has received much less attention from researchers. In fact, only a few of the articles reviewed in Section 2 concern the ETO context, with the majority referring to the MTO context.

The proposed model considers nonregular capacity by explicitly representing overtime and subcontracting and adjusting the internal capacity workforce by hiring and firing personnel. According to Table 1, the research studies that address the ETO or the multi-project context do not entirely explore these types of working time flexibility. However, these nonregular capacity alternatives are consistent with the workforce flexibility options used by the company's manufacturing team to adapt capacity to demand because, in the studied setting, inventories cannot be used to fill gaps between demand and capacity. Moreover, the use of contingent labour to temporarily increase the permanent capacity is not a viable option for the company because its production requires particularly skilled workers who require a significant amount of investment and time in training, an issue highlighted by Alp and Tan (2008). In addition, and in contrast to the reviewed models that address capacity changes by hiring new employees (Hans, 2001; Ebben et al., 2005), the proposed model allows such changes but does so according to a minimum employment period policy to guarantee some stability.

The proposed model also admits multiple processing stages when representing the production process. Although this issue is not a novelty in the literature (see Table 1), it enhances the decision making process for the studied company, which depends on an aggregate production plan that considers capacity as a whole for simplification purposes. Thus, the model offers the managers information on the workforce capacity demand because each stage may correspond to a type of production process that requires specific capabilities. This is particularly crucial in this environment because the workforce is usually not cross-trained, making it difficult to be relocated among the processes.

Moreover, the proposed solution addresses the third modelling issue (i.e., precedence relationships among activities) which comprises three characteristics of the studied production flow: (i) the execution progress mode of each activity varies along the time periods, (ii) the activities overlap in time successively and (iii) the precedence relationship between two activities is not fixed, because one may precede in some periods of time and succeed in other periods. This behaviour was modelled by the use of the cadence constraints that assume the variable intensity formulation (i.e., that enables the variable execution progress mode). Moreover, these constraints admit the overlapping of activities and do not presume a fixed precedence relation between them. They consider that each activity has its own cadence but is conditioned by the progress of the other related activities at each milestone. Apparently, none of the revised models presented in Table 1 exhaustively describe this behaviour. The typical precedence models do not consider the overlapping characteristic. The GPR models consider overlapping but assume a fixed execution progress mode. And finally, the feeding precedence models (see feeding or GFPR in Table 1) consider fixed precedence relation. In this sense, the cadence constraints were developed to overcome these difficulties and, for the company, they enabled the generation of consistent and realistic production plans.

As for the application context, the proposed model addresses a specific but real-world tactical capacity planning problem in an ETO setting. It is designed for situations characterized by a lack of detailed product information, especially when production planning precedes the definition of the bill-of-materials and the product's routings. This aspect does not restrict the use of the model in other types of production systems, such as MTO. However, when detailed information is available during the order acceptance phase, an optimization model that supports details seems to be more appropriate in order to generate more accurate production plans.

4.2.Application

This section presents the results of computational experiments realized to validate the model's behavior and efficiency and analyses the main findings obtained with the application of the capacity planning model proposed in Section 4.1 to solve the real-world ETO tactical planning problem described in Chapter 3. The mathematical model was implemented in Aimms 3.13 and solved using CPLEX 12.1 with its standard configurations. All tests were performed on an Intel® Core™ i5 CPU 1.70G Hz with 6 GBRAM. The model was fed with real-world data and solved in order to check whether it actually reflects the planning problem. Furthermore, alternative scenarios were also generated to assist the management board in the order acceptance phase, and to reinforce the model validation in a real-world context, thereby addressing the research-practice gap in this research area.

4.2.1.Computational experiments

To validate the behavior of the proposed model and to test its efficiency, a set of experiments was realized using a dataset composed of three subsets, each containing instances with five resources and 30, 100 or 250 activities. Additionally, for each instance, different values for the activities time windows (i.e., release dates and deadlines) were considered in order to simulate three sizes of planning horizons (i.e.,12, 15 and 18 months). In this sense, "cumulative" instances, where many activities are scheduled in parallel and "disjunctive" instances, when activities are scattered along the planning horizon, were evaluated. Table 6 shows the CPU times in seconds for the instances tested as well as the number of constraints and variables involved. The optimality tolerance gap considered was 0.5%.

Table 6: Experimental analysis

		30 activities	100 activities	250 activities
12 months	CPU time	0.26 s	142 s	334 s
	Variables	4441 (3120 int)	13853 (9532 int)	31279 (21258 int)
	Constraints	14412	54632	131104
15 months	CPU time	0.64 s	84 s	271 s
	Variables	5371 (3945 int)	14783 (10357 int)	32209 (22083 int)
	Constraints	15552	55772	132244
18 months	CPU time	1.22 s	13 s	273 s
	Variables	6481 (4950 int)	15893 (11362 int)	33319 (23088 int)
	Constraints	16917	57137	133609

4.2.2.Inputs

The input data displayed in Table 7 refer to a particular circumstance that the company manufacturing team was willing to assess. For the sake of confidentiality, data masking was applied to protect original data, but in a way that the data remained useful for the purposes of the model application.

Table 7: Input data

Planning horizon	18 months
Fixed capacity period	3 months
Work centres	5
Running projects	25 (details in Table 9)
Activities per project	5-30
Total activities	250
Range of intensity of the activities	1 to 50% per period (depending on the work centre)
Cadence data	See Table 2
Limit of working hours	4,900 to 58450 per period (depending on the work centre)
Limit of regular hours per employee	150 per period
Limit of overtime hours per employee	25 per period
Initially allocated employees	284 (total for all work centres)
Average salary per month	4,950 monetary units
Overtime hourly cost	50 monetary units
Production process hourly cost	11 monetary units
Capacity change cost (hiring or firing)	5,100 monetary units per employee
Minimum employment period	6 months
Committed workload	89,750 hours (details in Table 10)

Table 8 presents the product structure (i.e., the final products and their main part components), the subcontracting options (i.e., the possibility of subcontracting a given part or not and the cost of subcontracting) and the routing information (i.e., the estimated processing times in each work centre) for a medium-sized boiler and for three types of reactors, which correspond to the incoming demand of this particular planning circumstance. Although these orders are extremely customized projects when analysed in detail, at this planning level similarities among them allow for classifying them into product types. Historical data were used in this case to estimate processing times.

Table 9: Product information

Product structure		Subcontracting		Processing times (h)				
Product type	Parts	Possible	Cost ($1000)	Cutting	Stamping	Machining	Assembly	Welding
Medium boiler	Body			1197		2317	36589	33735
	B2	Yes	4850		1657		1944	1090
	D	Yes	1940	312		157	1048	959
	F	Yes	(105 -222)	312	29	157	1048	959
	H	Yes	360	28	29	40	164	237
	I2	Yes	3222		2		2257	2114
	J	Yes	404	67		40	140	19
	K2	Yes	1565				2257	2114
Reactor X	Body			202	120	522	1362	1451
Reactor Y	Body			196	116	50	1498	1988
Reactor W	Body			129	109	50	775	1052

Table 9 presents data relative to the 25 expected incoming customer orders (i.e., the committed new projects and the proposals in the order acceptance phase) for the current planning horizon, including release dates and deadlines for each project. Table 10 presents the workload already committed for this company for the upcoming 18 months.

Table 10: Expected incoming customer orders

Project	Product type	Release date	Deadline	Category
PJ01	Medium boiler	1	9	Committed
PJ02	Reactor X	2	7	Committed
PJ03	Reactor Y	2	7	Committed
PJ04	Reactor W	2	7	Committed
PJ05	Medium boiler	3	11	Committed
PJ06	Medium boiler	6	14	Proposal
PJ07	Medium boiler	9	17	Proposal
PJ08	Medium boiler	5	12	Committed
PJ09	Reactor Y	2	7	Committed
PJ10	Reactor W	2	7	Committed
PJ11	Reactor X	5	10	Committed
PJ12	Reactor Y	5	10	Committed
PJ13	Reactor W	5	10	Committed
PJ14	Reactor X	5	10	Committed
PJ15	Reactor Y	5	10	Committed
PJ16	Reactor W	8	13	Proposal
PJ17	Reactor X	8	13	Proposal
PJ18	Reactor Y	8	13	Proposal
PJ19	Reactor W	8	13	Proposal
PJ20	Reactor X	8	13	Proposal
PJ21	Reactor Y	13	18	Proposal
PJ22	Reactor W	13	18	Proposal
PJ23	Reactor X	13	18	Proposal
PJ24	Reactor Z	13	18	Proposal
PJ25	Reactor Y	13	18	Proposal

Table 11: Committed workload

Work centres	Time periods																	
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cutting	2000	1500	500	250	200	110	40											
Stamping	5800	2350	1700	1200	500													
Machining	2100	1000	500	450	200	80	20											
Assembly	8200	4800	3300	2300	2100	850	200											
Welding	15400	9600	6100	3050	2200	1500	300											

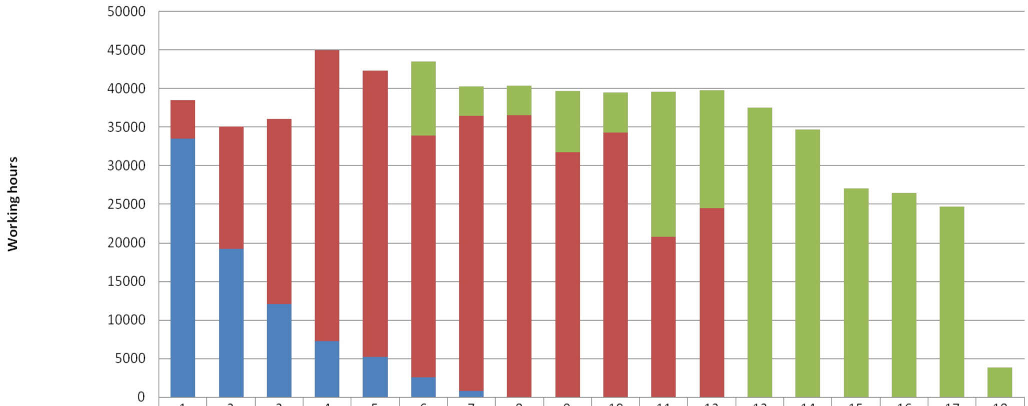
4.2.3.Results

The problem instance relative to the real world circumstance described in Subsection 4.2.2 (See Tables 7, 8, 9 and 10) involved 32,119 variables (17,928 integer variables) and 128,881 constraints. Table 11 summarizes the main outputs considering a stop criterion of 0.3% relative to the optimality tolerance gap (i.e., the solver stops when the gap is smaller than 0.3%).

Table 12: Outputs

Iterations	1,679,940
Solution time	1060 s
Gap	0.28%
Best solution (total cost)	31,634,451

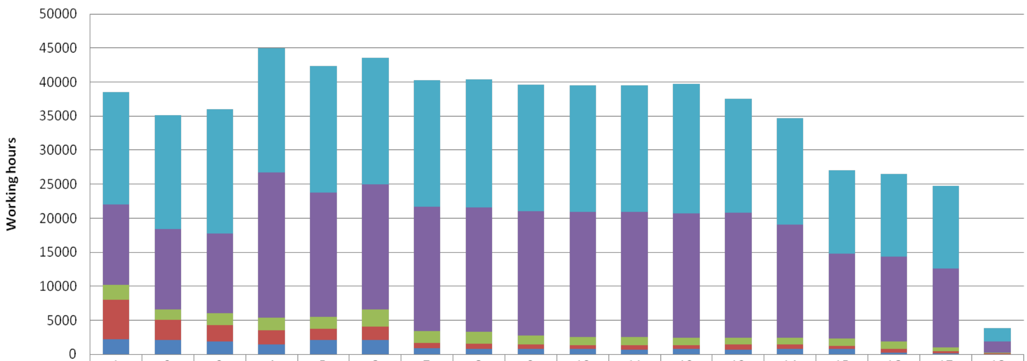
The optimal plan obtained suggests a workload distribution for each demand category (i.e., committed workload, committed new projects and the proposals), such as that shown in Figure 5, where the horizontal axis corresponds to the time periods. One may note that the committed workload refers to the first seven periods, decreasing according to the completion of these projects. Category “new projects” refers to 13 projects that represent almost 53% of the expected demand, whereas the proposals correspond to another 12 projects representing almost 35% of what is planned within the following 18 months.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Proposal						9593	3848	3891	7917	5232	18740	15271	37575	34691	27055	26454	24752	3867
New project	5027	15889	23926	37735	37194	31401	35718	36531	31760	34266	20814	24507						
Committed Workload	33500	19250	12100	7250	5200	2540	760											

Fig. 5: Workload distribution in the optimal production plan

Figure 6 refers to the same optimal plan but exhibits this workload distribution per work centre revealing that most of the planned workload refers to the assembly and the welding work centres. This disaggregate view per work centre is not possible in the current planning method applied by the company, which limits the manufacturing team's ability to fully understand the new demand implications for capacity planning.

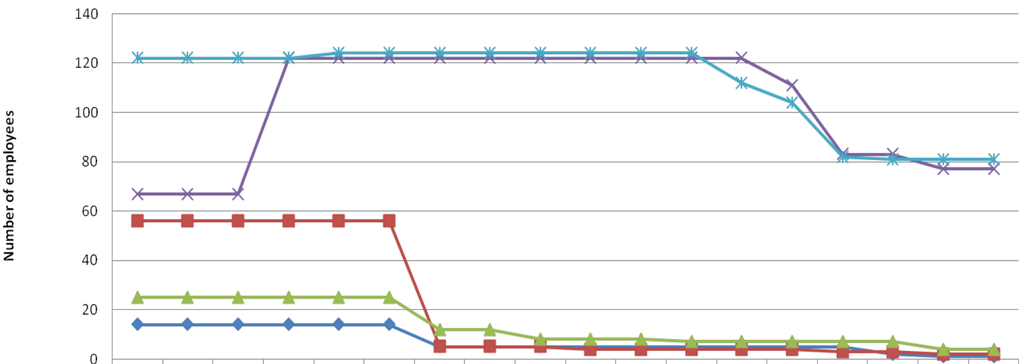


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Welding	16566	16789	18301	18300	18599	18600	18599	18796	18655	18599	18659	19027	16801	15600	12301	12150	12151	1983
Assembly	11725	11725	11726	21350	18300	18371	18301	18301	18302	18302	18301	18302	18302	16650	12452	12450	11550	1626
Machining	2171	1554	1665	1805	1737	2473	1801	1799	1218	1244	1219	1090	1049	1038	1050	1104	600	119
Stamping	5833	2972	2460	2039	1656	1990	750	752	751	601	643	610	700	651	502	450	301	107
Cutting	2232	2099	1874	1491	2102	2100	875	774	751	752	732	749	723	752	750	300	150	32

Fig. 6: Workload distribution per work centre in the optimal production plan

The implementation of this optimal production plan requires capacity changes. Figure 7 shows the changes proposed in terms of the number of employees for each work centre. One may note that in the fixed capacity periods (i.e., in the three initial periods), hiring and firing is not permitted, so the number of employees remains constant along these periods. Additionally, there is also the minimum employment period of six months, which guarantees that employees should not be fired before this period.

According to the results, in period four, 55 assemblers would have to be hired to cope with the incoming demand. This type of disaggregated information is impossible to obtain with precision using the company's current method. Additionally, it should be noted that by the end of the planning horizon, almost 50% of the personnel should be dismissed according to the plan. This only happens because in the model there is no demand after the last time period, as this is a finite planning horizon. In the real world, however, new demand is expected in future periods so these dismissals may not be necessary. In practice, the plan should only be implemented for the more imminent periods within the planning horizon.

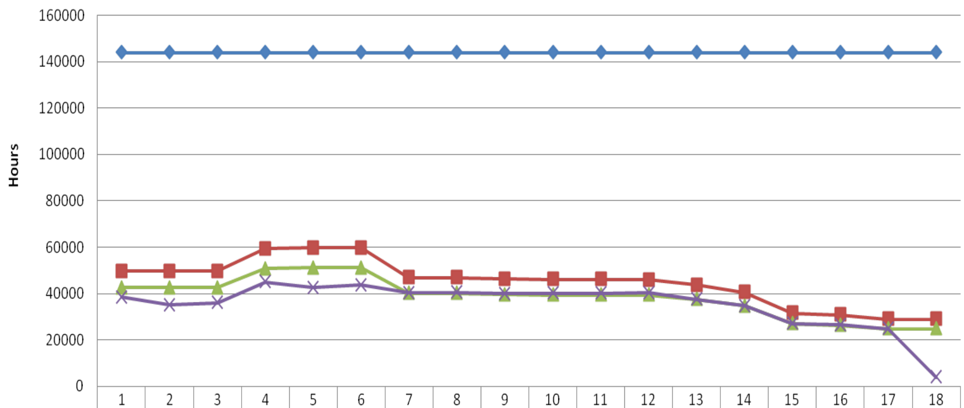


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Cutting	14	14	14	14	14	14	5	5	5	5	5	5	5	5	5	2	1	1
Stamping	56	56	56	56	56	56	5	5	5	4	4	4	4	4	3	3	2	2
Machining	25	25	25	25	25	25	12	12	8	8	8	7	7	7	7	7	4	4
Assembly	67	67	67	122	122	122	122	122	122	122	122	122	122	111	83	83	77	77
Welding	122	122	122	122	124	124	124	124	124	124	124	124	112	104	82	81	81	81

Fig. 7: Employees allocated per work centre in the optimal production plan

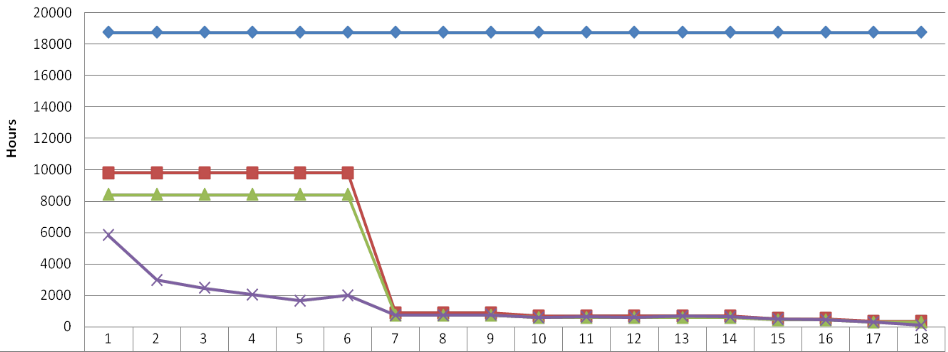
Figure 8 compares the total current demand (i.e., committed workload, committed new projects and proposals) with the total internal capacity, considering the capacity flexibility options inherent to this planning problem. The maximum capacity is the maximum load capacity or physical facility capacity, which represents the potential usage of all infrastructure, machinery and process technology available. This corresponds to the limit of how much can be invested in terms of hiring personnel in the tactical level. Above this border, investments that are out of the scope of this planning level would be necessary. From the results shown in Figure 8, it can be concluded that production resources (e.g., machinery, process technology) are under-utilized when considering the physical facility capacity. In fact, this company has committed to an investment in a higher level of capacity to facilitate later expansion. At the time, there were capital expenditure efficiencies gained by constructing a larger infrastructure and acquiring additional machinery. Clearly there is some risk involved in committing the necessary capital expenditure before demand is certain (which may result in the under-utilization of capacity). Such a lead demand strategy is frequently employed in growing markets and ensures that the operation is likely to be able to meet demand.

In this particular problem, although there is a large physical facility capacity, it is only equipped to a part of its potential. In other words, the number of allocated employees defines the effective capacity in each time period. In this sense, in Figure 8, the regular capacity refers to the regular working hours given the number of employees available at each time period. Overtime capacity refers to this regular capacity plus the overtime hours available according to the employees allocated. Viewed as a whole, the effective capacity seems to meet demand along the planning horizon. Conversely, when analysed for a specific work centre such as stamping (see Figure 9), there is a significant mismatch between the effective capacity and demand during the initial six time periods. These periods of idleness result from the model's assumption of the initial fixed capacity periods and the minimum employment period constraints.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Maximum capacity	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846	143846
With overtime	49700	49700	49700	59325	59675	59675	46900	46900	46200	46025	46025	45850	43750	40425	31500	30800	28875	28875
Regular Hours	42600	42600	42600	50850	51150	51150	40200	40200	39600	39450	39450	39300	37500	34650	27000	26400	24750	24750
Demand	38527	35140	36023	44982	42645	43712	40353	40485	39943	40108	40139	40281	37574	34689	27051	26453	24749	3865

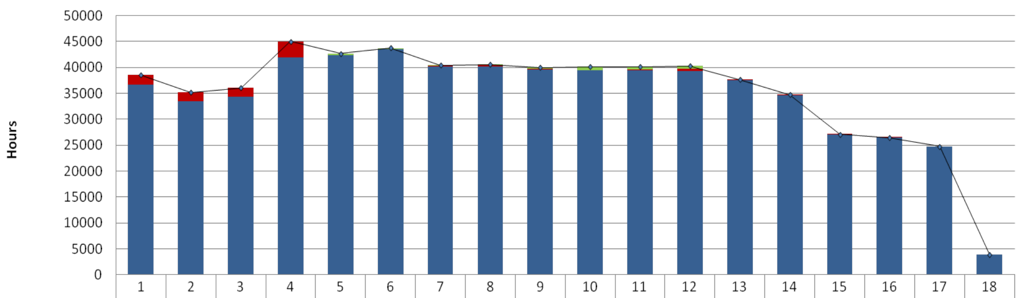
Fig. 8: Current demand and the internal capacity options



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Maximum capacity	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725	18725
With overtime	9800	9800	9800	9800	9800	9800	875	875	875	700	700	700	700	700	525	525	350	350
Regular Hours	8400	8400	8400	8400	8400	8400	750	750	750	600	600	600	600	600	450	450	300	300
Demand	5833	2971	2460	2038	1656	1994	750	752	756	604	646	615	700	650	501	450	300	107

Fig. 9: Current demand and the internal capacity options for the stamping work centre

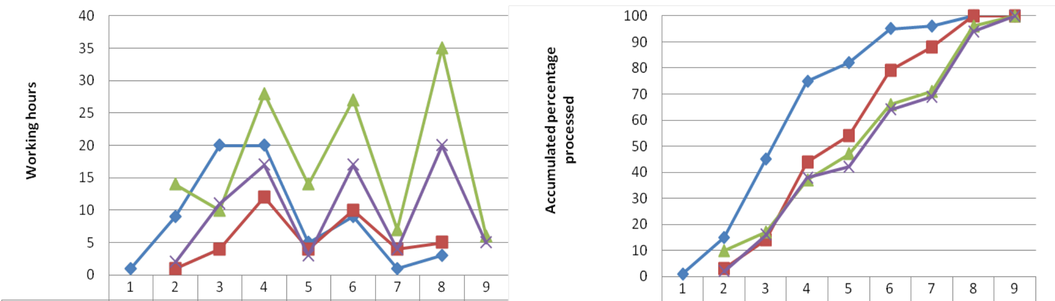
Even though there seems to be sufficient internal capacity (i.e., regular and overtime hours) to meet demand in each time period, as presented in Figure 8, the optimal plan suggests the subcontracting of a part component. In this sense, Figure 10 shows how demand is effectively met, either by using regular, overtime or subcontracting hours.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Subcontracting					254	181	28	66	270	615	588	504						
Overtime	1807	1675	1675	3050		70	125	220	73	43	119	478	100	50	51	53		
Regular	36720	33465	34348	41932	42391	43461	40200	40200	39600	39450	39432	39300	37474	34639	27000	26400	24749	3865
Demand	38527	35140	36023	44982	42645	43712	40353	40485	39943	40108	40139	40281	37574	34689	27051	26453	24749	3865

Fig. 10: Current demand and the capacity options effectively used

To properly represent the synchronism of the production flow, this optimal plan is subject to the set of constraints that define the cadence of the interconnected activities. In this sense, for illustrative purposes, Figure 11 presents the cadenced and overlapping behaviour of four interconnected activities of a specific component along the time periods. The graph on the left presents the workload distribution for each activity of this component along the planning horizon. On the right, the accumulated percentages of these activities are shown. One can notice that the individual curves on the left graph tend to reproduce, during most of the time, a converging behaviour (either increasing or decreasing), which is precisely the cadenced behaviour. Nevertheless, activities have their own specific pace. For instance, cutting usually starts the production process and it supplies materials to the other processing stages. By the end of the 4th period, this activity is 75% complete, whereas only 44% of machining, 37% of assembly and 38% of welding has been processed. According to Table 2, milestone 13 is met in the 4th period.



	1	2	3	4	5	6	7	8	9
Cutting	1	9	20	20	5	9	1	3	
Machining		1	4	12	4	10	4	5	
Assembly		14	10	28	14	27	7	35	6
Welding		2	11	17	3	17	4	20	5

	1	2	3	4	5	6	7	8	9
Cutting	1	15	45	75	82	95	96	100	100
Machining		3	14	44	54	79	88	100	100
Assembly		10	17	37	47	66	71	96	100
Welding		2	16	38	42	64	69	94	100

Fig. 11: Workload distribution and accumulated percentage processed for a component along the time periods

Regarding the overall variable production cost that is minimized by this optimization model, Figure 12 shows its fractions in terms of production processing, overtime, capacity change, personnel payroll and subcontracting. As seen, personnel payroll represents a considerable portion of the total cost (72%).

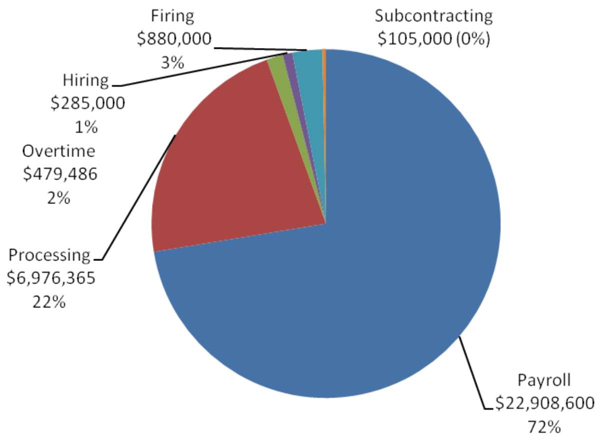
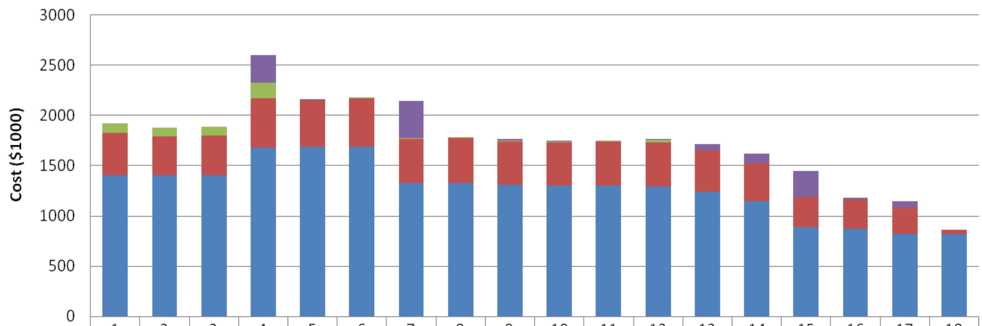


Fig. 12: Costs according to the optimal production plan

Additionally, Figure 13 details some of these costs per time period. Capacity change costs occur in specific periods of time and include hiring (i.e., specifically in periods 4 and 5) and firing employees. The overtime costs occur mostly at the beginning of the planning horizon when hiring personnel is not permitted. The peak in demand and the consequent need to hire more employees in period 4 turns it into a critical period in terms of costs.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Capacity change				275	10		365		20	5		5	60	95	255	20	55	
Overtime	90	84	84	153		4	6	11	4	2	6	24	5	3	3	3		
Processing	424	387	396	495	466	479	444	445	436	434	435	438	413	382	298	291	272	43
Payroll	1406	1406	1406	1678	1688	1688	1327	1327	1307	1302	1302	1297	1238	1143	891	871	817	817

Fig. 13: Costs according to the optimal production plan per time period

4.2.4. What-if scenarios

Although the model generated an optimal production plan, additional three scenarios were simulated in this particular circumstance to support the decision making process, as required by the manufacturing director being them: (i) producing a specific part component in house, (ii) accepting a new incoming order and (iii) postponing production for resource levelling. The analysis of the model solutions in these scenarios was made together with the company´s manufacturing team. Table 12 summarizes the costs of the original plan and three generated scenarios. The details and discussion of each scenario are presented next.

Table 13: Comparison of original plan and simulated scenarios

	Original plan	Scenario 1 (component in house)	Scenario 2 (accept new order)	Scenario 3 (postpone a deadline)
Production processing cost	6,976,365	7,003,920	7,673,314	7,673,314
Overtime cost	479,486	448,982	399,885	464,660
Subcontracting cost	105,000	0	222,000	222,000
Personnel payroll	22,908,600	23,002,650	24,943,050	24,878,700
Hiring cost	285,000	295,000	460,000	390,000
Firing cost	880,000	890,000	1,055,000	985,000
Total cost	31,634,451	31,640,551	34,753,249	34,613,673
Optimality tolerance gap	0.28%	0.31%	0.29%	0.33%

Scenario 1: Producing a specific part component in house

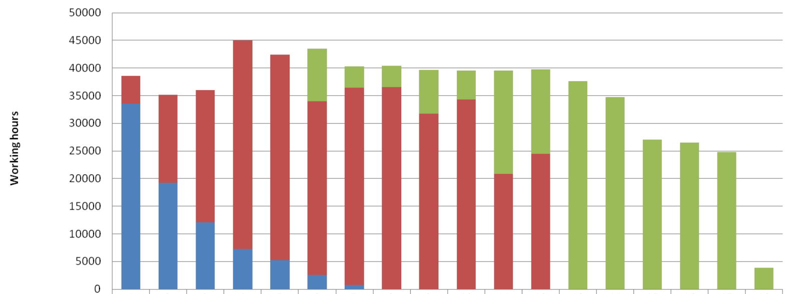
After a detailed analysis of the optimal production plan, which recommends that one of the components should be subcontracted, the company managers decided to assess how much was being saved by subcontracting this particular item. This assessment was performed by prohibiting subcontracting and solving the model again. After this assessment was performed, the result showed an expected increase in the production processing cost and slight increments in the payroll and capacity change costs. However, for the surprise of the manufacturing team, the overall production cost increased only 0.02% (see Table 12), which was considered irrelevant by the managers.

This simulation aided the company in deciding whether to subcontract, although this determination is not made solely on the basis of economic considerations. Acquisition or loss of core competencies and risk mitigation are also involved. In this specific scenario, the company opted not to subcontract this part component due to the low savings associated with this action and the risk associated with using a subcontractor. This information would be impossible to obtain through the company's current planning method.

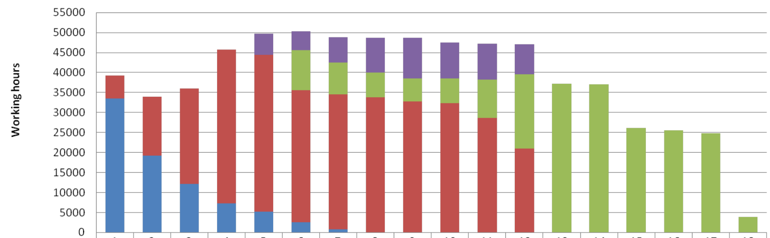
Scenario 2: Accepting a new incoming order

During this planning process, a new incoming order, in the tendering phase, was also being analysed to determine order acceptance or rejection. Maintaining the decision to produce the aforementioned component in house (scenario 1), the company managers decided to use the proposed model to generate a second scenario to estimate the additional costs for accepting this new demand. They also wanted to know if the customer´s due date could be met and if it would be necessary to hire more employees because this demand represented a large project.

Figure 14 shows the workload distribution in the original plan and in scenario 2. This new proposal was planned for the time window ranging from the 5th to the 12th period. Its inclusion in the production plan resulted in the adjustment of the workload distribution for other incoming demands. As expected and shown in Table 12, scenario 2 reflects increases in the production processing costs (the new project refers to additional 60,855 hours), in the personnel payroll and the capacity change costs (7 more assemblers, 9 more welders and 19 employees for the machining work centre should be contracted over the next 5 months). Furthermore, this information enables the recognition of actions that other management areas within the company (e.g., the sales, financial and human resources departments) will need to handle, which is impossible to predict within a short period of time using the current planning method.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Proposal						9593	3848	3891	7917	5232	18740	15271	37575	34691	27055	26454	24752	3867
New project	5027	15889	23926	37735	37194	31401	35718	36531	31760	34266	20814	24507						
Committed Workload	33500	19250	12100	7250	5200	2540	760											



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
New proposal					5206	4765	6457	8759	10246	8924	8911	7587						
Proposal						10026	7955	6170	5651	6260	9663	18533	37222	37089	26083	25586	24786	3862
New project	5704	14674	23884	38542	39277	33043	33728	33769	32806	32289	28576	20984						
Committed Workload	33500	19250	12100	7250	5200	2540	760											

Fig. 14: Workload distribution for the original plan and scenario 2

Scenario 3: Postponing for resource levelling

Maintaining a levelled production plan is one of the company´s goals. Therefore, after a meticulous revision of the optimal production plan in scenario 2 (which includes the new incoming order), the managers found that a demand peak in a specific time period represented the need to hire a considerable number of employees for the machining work centre. In this situation, the managers decided to simulate the postponement of the new proposal in order to level demand. This was accomplished by delaying the due date of this new project and solving the model once more. Figure 15 features the intentional delay in the given project. Figure 16 highlights (see the arrows) the considerable decrease in the number of employees needed from the 5th to the 14th period. This decrease results in a smaller overall cost, as seen in Table 12, by comparing scenario 2 with scenario 3. This is useful information to be presented at the company´s board of directors meeting, so that the commercial/sales director can investigate the possibility of negotiating a new due date with the customer for that specific proposal, even by using yield management techniques.

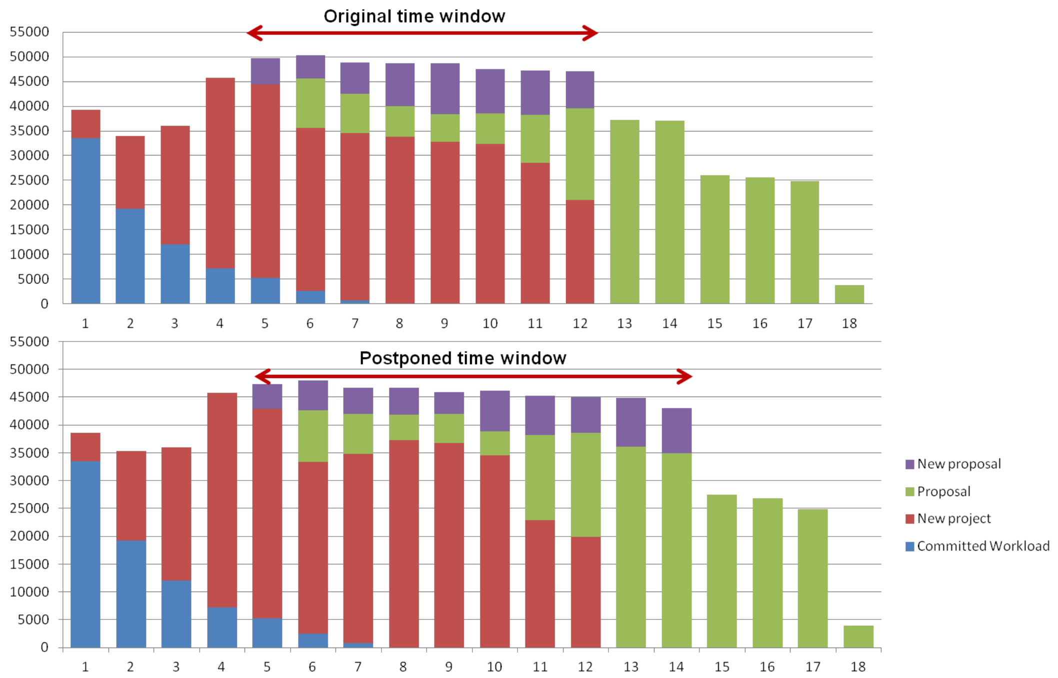


Fig. 15: Workload distribution for scenario 2 and scenario 3

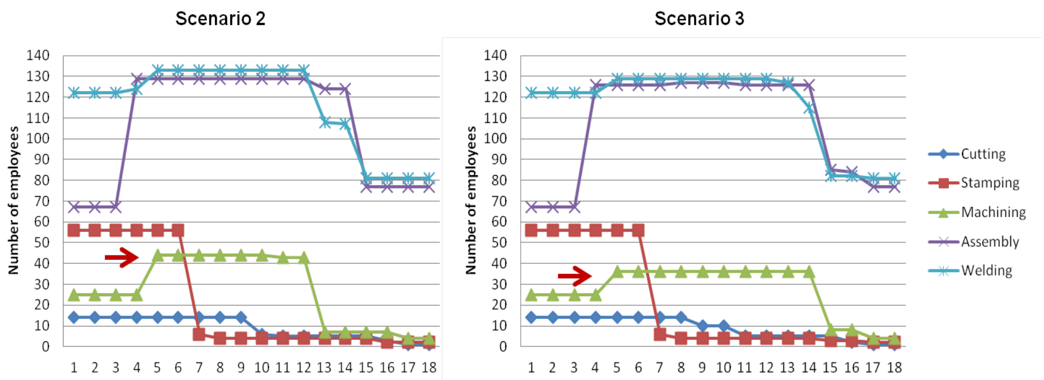


Fig. 16: Employees allocated per work centre according to scenario 2 and scenario 3

4.2.5.Discussion

Applying the model to a real ETO company interested in improving its capacity planning methods in the order acceptance phase was valuable for validating the model itself and for comparison with the company's current planning method. A quantitative comparison analysis, however, was not carried out as the plans generated by the model refer to a longer planning horizon which consider a more detailed level of information than the plans generated by the company's current planning method. Despite this, according to the company´s planning team, the model application proved to be useful to provide information for an eventual bid preparation. Moreover, it assisted this team to balance demand with the available capacity, decide on the usage of nonregular capacity, determine if a new order should/could be accepted and estimate the overall variable production costs. Furthermore, the model includes information that is sufficiently aggregate to address some of the uncertainty present and sufficiently detailed to represent this planning problem. In other words, the model addresses the workload analysis problem in a level of information consistent with the tactical planning phase. For instance, the nonregular capacity options represented in the model are consistent with the issues that the company manufacturing team addresses to cope with the planning problem.

Additionally, different shortcomings in the company's current planning method are addressed by the proposed model and its application. Because it refers to an optimization model, it automatically generates and assesses many alternative solutions in order to find the best outcome. This addresses the fact that only a few alternative plans were evaluated previously by the manufacturing planning team due to the time-consuming process of generating and assessing each plan using the company's current planning method. Another shortcoming is that the cost analysis is not carried out in the current method. Because the proposed model is a cost minimization model that systematically calculates the production costs related to each portfolio, the model is useful for assessing the trade-offs when comparing, for example, internal processing with subcontracting costs or hiring employees and overtime costs. Finally, the difficulty faced in identifying the need for a specific workforce when adjusting internal capacity in the current method is addressed because the proposed model considers capacity in a disaggregate form, making it possible to quantify the number of workers needed in each work centre in every time period of the planning horizon.

The manufacturing planning team reported a good level of satisfaction with the quality of the results, and the following items were referenced during the last validation session:

- Although the proposed model could not be used alone to decide on order acceptance (i.e., the decision to accept or reject a specific customer query depends on a series of criteria that are considered by different departments within the company), it contributes to this process by improving the workload analysis.

- The model represents a diagnostic and decision support tool because it not only helps the managers identify potential gaps between capacity and demand but also assists them by demonstrating how to balance the provision of capacity with demand and how to level demand. Because the model plans demand according to the informed time windows of each project, the managers are able to perform a feasibility check of release dates and due dates. Thus, the model allows the manager to better observe capacity usage and determine milestones for the production of the part components.

- The model also allows the company manufacturing planning team to intervene in the construction of the production plan with managerial decisions (i.e., changing specific input data) in order to generate a concurrent new analysis. Through a sensitivity analysis, it is possible to test different production plans evaluating 'what-if' scenarios and policies in an iterative and interactive planning process comparing the overall variable production costs, as presented in Section 4.2.4. These managerial decisions refer to issues related to demand, supply, capacity, employees and the production flow as detailed in Table 13.

Table 14: Managerial decisions

Type	Actions
Demand	Select the expected incoming orders (proposals) to simulate the plan Adjust deadlines (to determine a deadline or check on the feasibility of a mer's deadline) Modify the original already committed workload plan
Supply	Adjust release dates (material availability) based on information from suppliers and project engineers Allow (or not) the subcontracting of specific part components
Capacity	Alter the maximum number of working hours for specific work centres Authorize overtime for specific work centres
Employees	Change the employment period policy (extending or reducing the minimum employment period) Modify (expand or reduce) the fixed capacity periods
Production flow	Adjust the cadence parameters Regulate the minimum and maximum intensity of the activities

- The model enhances the decision-making process by providing more detailed and precise information about the planning problem that was not available through the company's current method (e.g., the number of employees needed in each time period for each work centre, the number of employees hired and fired in each time period relative to each work centre, the number of overtime hours needed in each time period for each work centre and which part components should be subcontracted).
- This information permits a timely assessment of possible shortcomings and the identification of the primary actions to be taken in order to accept incoming orders. These actions may need to be discussed and negotiated with other management areas (e.g., the industrial manufacturing director may need to contact the human resources department when anticipating a future need to hire more assemblers, may need to inform the financial department of possible overtime hours costs and may need to check the feasibility of outsourcing some part component with the subcontractors).

5 Robust optimization approach

This chapter presents the robust optimization model that extends the former deterministic model addressing the same tactical capacity planning problem. The motivation for this extension derives from the fact that the real-world setting studied is characterized by many sources of uncertainty, which are typical to the ETO context. In fact, multi-project contexts, such as the studied setting, are subject to considerable uncertainty and their planning processes should incorporate the variability of critical parameters, as highlighted by several authors (Herroelen and Leus, 2004; Tolio and Urgo, 2007; Chtourou and Haouari, 2008; Van de Vonder et al., 2008; Deblaere et al., 2011; Alfieri et al., 2012; Artigues et al., 2013; Radke et al., 2013). In this sense, this chapter is organized in two sections: (i) a description of the proposed robust optimization model and (ii) a presentation of the application of the proposed robust model to solve the real world tactical planning problem under study.

5.1. The proposed robust model

This section begins by presenting the main characteristics of the proposed robust optimization approach. In the sequel, the robust model is detailed in its mathematical form and a process to measure the robustness of the generated solutions by calculating probability bounds of constraint violation is described. The section ends with a discussion comparing the proposed robust model with what has been found in literature in terms of the modelling issues highlighted in Table 1, more specifically referring to the solution method and the uncertainties modelled.

5.1.1. Introduction to the robust model

The ETO context is subject to several sources of variability. As this research refers to a tactical planning level, only the uncertainties that affect the medium term were analysed. In this sense, cost uncertainty was discarded as it seems to be a critical parameter in the long term, according to the company's viewpoint. Demand uncertainty was also not considered as the plans are generated for a given and fixed number of incoming orders. On the other hand, the committed new projects and the proposals are sensitive to variability in terms of the production activities processing time, especially since the design phase is not concluded. Therefore, there is a significant level of uncertainty as this process information becomes gradually available. For instance, Figure 17 compares the accumulated percentage of the completed production activities according to the original plan conceived six months before production started, and the actual production of a specific boiler during a 16 period horizon. One may notice that there is a delay between the two curves as the actual production falls behind what was originally planned. When this particular plan was conceived, the processing times, which are subject to many independent arbitrary variables (e.g., changes in component designs, difficulties in handling a new type of material, personnel inefficiency), were underestimated. As highlighted by Khakdaman et al. (2015), process-related uncertainty is one of the main types of uncertainties that make medium term plans obsolete, thus requiring adjustments in the original generated plans.

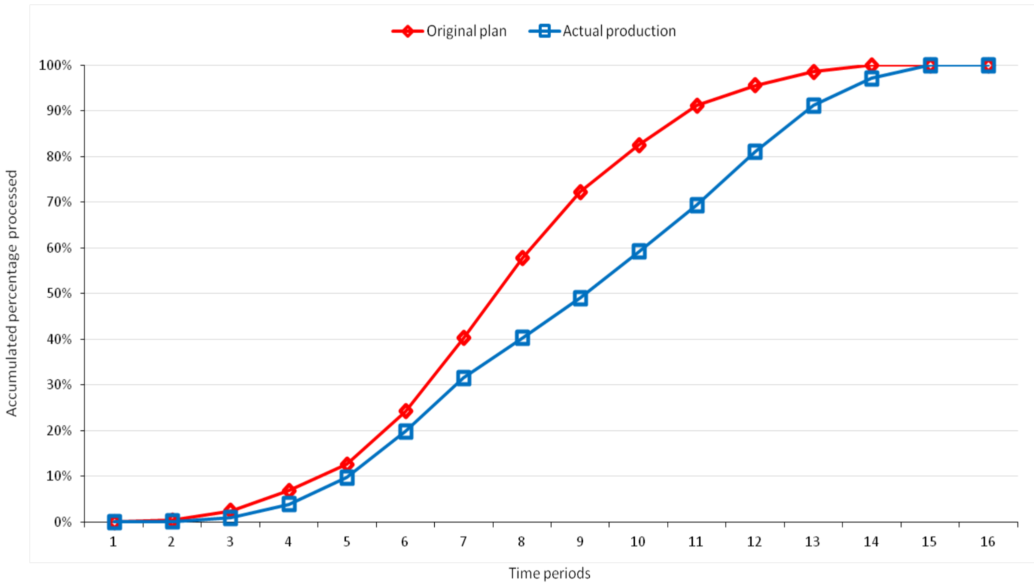


Fig. 17: Comparison between the original plan and the actual production in terms of accumulated percentage processed

To address this problem, a proactive planning approach has been adopted to achieve solution robustness or stability. More specifically, in this study, robustness is achieved whenever the generated production plan (i.e., the original plan) is capable to absorb the activities processing time deviations avoiding readjustments on the original plan, and therefore minimizing the need for a constant replanning cycle. In this sense, a robust optimization model is employed to guarantee that the generated plan remains stable regardless of the variability resulting from this type of uncertainty. The motivation for adopting a robust approach over the so-called traditional two-stage stochastic programming approach (Alem and Morabito, 2012; Munhoz and Morabito, 2014) is the lack of an explicit probabilistic description of the uncertain input data modelled (i.e., the uncertainty of the parameters is modeled as lower and upper bounds without any need for exact distributions).

The proposed robust model is based on Bertsimas and Sim (2004)´s robust approach, which allows the degree of conservatism of the solution to be controlled in terms of probabilistic bounds of constraint violation. More specifically, the plan is protected for the case where only a pre-specified number (i.e., parameter Γ) of deviations occurs. The parameter Γ is introduced in order to adjust the model robustness against the conservatism of the solution. It is also known as the *budget of uncertainty*, which reflects the decision-maker's attitude towards uncertainty. As this budget increases, the model is more protected against processing time variation (i.e., more processing time deviation is incorporated by the model). As Γ is an integer value in this problem, it is interpreted as the maximum number of uncertain parameters that can deviate from their nominal values.

One advantage of the approach proposed by Bertsimas and Sim (2004) is the possibility to quantify explicitly the relationship between the level of conservativeness of the solution and the probability of constraint violation. In this particular problem, when assessing this probability, one is not precisely evaluating the robustness associated to the viability of the generated plans (i.e., as this is not the critical issue since the nonregular capacity options represent a high level of flexibility to fit demand into the available capacity). This probability corresponds to the stability of the generated plans. Therefore, the aforementioned parameter Γ controls the trade-off between this probability and its effect on the objective function. In other words, for a stable plan (i.e., a conservative solution, with a low probability of constraint violation), one should adopt a higher value for Γ, which may represent an extra cost also known as the *price of robustness*. On the other hand, when the protection required is not too high, one can reduce this price by adjusting Γ to a lower value.

In the proposed robust model, uncertainty is only coupled with the first part of the planning horizon (i.e., the fixed capacity periods). In this sense, the activities processing time deviations are considered when calculating the available capacity to process demand only for these fixed capacity periods. This assumption seems appropriate as more accurate decisions must be made in these imminent periods, especially due to the fact that there is less flexibility in these periods, as capacity changes in terms of hiring personnel are not allowed. In real manufacturing environments, although the uncertainty relative to distant events in the future is high, its the relevance is normally low on the short-term decisions (Tolio and Urgo, 2007). Furthermore, as this planning solution considers a rolling horizon planning approach, the uncertainty relative to the distant periods may be addressed when the new generated plans are recalculated especially when relevant events occur.

5.1.2. Mathematical formulation

The proposed robust optimization model is an extension of the former deterministic model (1)-(27), presented in Chapter 4. Table 14 presents the additional parameters and variables included in this version.

Table 15: Additional parameters and variables

Γ_{wt}	Parameter to adjust the model robustness
QD_{aw}	Parameter representing the deviation in the processing time of activity *a* at work centre *w*
k_{at}	Auxiliary variable representing a scaled deviation
Π_{wt}	Robustness variable
p_{at}	Auxiliary robustness variable

The processing time of activity *a* at work centre *w* is given by the parameter Q_{aw}. The parameter QD_{aw} represents the maximum possible deviation of the activity processing time from its mean value, Q_{aw}. In the robust model, each entry Q_{aw} is represented as a symmetric and bounded random variable $\tilde{Q}_{aw}$ with unknown probability distribution and with values in the interval $[Q_{aw}- QD_{aw}, Q_{aw}+ QD_{aw}]$. The subset *K* represents the set of coefficients Q_{aw}, $a \in K$, which are subject to uncertainty. Moreover, the auxiliary variable k_{at} is the scaled deviation of $\tilde{Q}_{aw}$ from its nominal value and is defined by $k_{at}= (\tilde{Q}_{aw} - Q_{aw})/QD_{aw}$ belonging to [-1,1].

The parameter Γ_{wt}, introduced in order to adjust the model robustness against the conservatism of the solution, represents the maximum number of the uncertain parameters that can deviate from their nominal values. Γ_{wt} may take values in the interval [0,|*K*|].

To build the robust counterpart of the deterministic model (1)-(27), it is necessary to modify the formulation of the workload constraint (i.e., Constraint 14) for the fixed capacity periods, in order to consider uncertainty in the parameters Q_{aw}. In other words, the sum of the activities processing times and their deviation must be equal to or smaller than the internal capacity hours plus subcontracting hours minus the already committed workload hours. This is presented in constraint (28).

$$\sum_{a} Q_{aw}\, x_{at} + \mathrm{Max}_k\left\{ \sum_a QD_{aw}\, x_{at} k_{at} \middle| \sum_a k_{at} \le \Gamma_{wt}\,;\ k_{at} \in [0,1] \right\} - \sum_{a} s_{at}$$

$$\le (RHE_w + OHE_w)\, NE_w - WH_{wt} \quad , \forall w, t \le FC \qquad (28)$$

Applying the robust optimization technique developed by Bertsimas and Sim (2004), an auxiliary problem is formulated (29-31). Its objective is to maximize the sum of all deviations over the set of all admissible realizations of the uncertain parameters.

$$\mathrm{Max}_k \sum_{a} QD_{aw}\, x_{at} k_{at} \qquad (29)$$

Subject to

$$\sum_a k_{at} \le \Gamma_{wt} \quad \forall w, t \le FC, \forall a\ | Q_{aw} > 0 \qquad (30)$$

$$k_{at} \le 1 \qquad \forall a, t \qquad (31)$$

If Γ_{wt} =0, the k_{at} for all *a* are forced to 0, so that parameters $\tilde{Q}_{aw}$ are equal to their mean value Q_{aw} and there is no protection against uncertainty. On the other hand, when Γ_{wt} = *K*, the k_{at} for all *a* are forced to 1 (in this particular problem) and constraint (30) is completely protected against uncertainty, which yields a very conservative solution. For values in between 0 and *K*, the decision-maker can make a trade-off between the protection level of the constraint and the degree of conservatism of the solution.

Following the same rationale of Bertsimas and Sim (2004), the dual of model (29)-(31) is stated as follows:

$$\mathrm{Min}_{p,\pi} \Gamma_{wt}\pi_{wt} + \sum_{a} p_{at} \tag{32}$$

Subject to

$$\pi_{wt} + p_{at} \geq QD_{aw}x_{at} \quad \forall\, w, t \leq FC \tag{33}$$

$$\pi_{wt} \geq 0 \quad \forall\, w, t \leq FC \tag{34}$$

$$p_{at} \geq 0 \quad t \geq FC, \forall\, a\, |Q_{aw} > 0 \tag{35}$$

This dual problem has two dual variables (π_{wt}, p_{at}) that are associated to constraints (30) and (31), respectively. By strong duality, as model (29)-(31) is feasible and bounded for all $\Gamma_{wt} \in [0,|K|]$, then the dual problem (32)-(35) is also feasible and their objective function values coincide.

Substituting the model (32)-(35) in Constraint (28), the following robust linear optimization model is obtained. The original problem is now rewritten in its final form.

Minimize

$$\sum_{a,w,t}[CR_w(r_{at} + o_{at}) + CO_w\, o_{at}] + \sum_{w,t}[CR_w\, (wr_{wt} + wo_{wt}) + CO_w wo_{wt} + CC\,(eh_{ttw} + \sum_{l}^{t} ef_{ltw}) + CS_w\, ea_{wt}]$$
$$+ \sum_{i} PS_i\, (1 - d_i)$$

Subject to: (1)-(13), (15)-(27)

$$\sum_{a} Q_{aw}\, x_{at} + \Gamma_{wt}\, \pi_{wt} + \sum_{a} p_{at} - \sum_{a} s_{at} \leq (RHE_{wt} + OHE_{wt})\, NE_{wt} - WH_{wt} \quad \forall\, w, t \leq FC \tag{36}$$

$$\pi_{wt} + p_{at} \geq QD_{aw}x_{at} \quad \forall\, w, t \leq FC \tag{37}$$

$\pi_{wt} \geq 0 \quad \forall\, w, t \leq FC$ (38)

$p_{at} \geq 0 \quad t \leq FC, \forall\, a\,|Q_{aw} > 0$ (39)

Finally, the proposed robust optimization model is composed by the objective function and Constraints 1 to 13 and 15 to 27 (presented in Chapter 4), plus Constraints (36) to (39) to consider uncertainty in the production activities processing time. For the sake of completeness, the entire robust optimization model is described in Appendix 1. This model minimizes the overall variable production costs and guarantees that if up to Γ coefficients change their values within the permitted interval (i.e., $[Q_{aw} - QD_{aw},\ Q_{aw} + QD_{aw}]$), then the solution of the robust optimization model will remain stable. In other words, the solution of this model is a robust solution.

5.1.3. Probability bounds for constraint violation

To select an appropriate value for the Γ parameter, the probability bounds for the workload constraint violation have to be estimated. Therefore, as each Γ represents a distinct production plan, each of these plans is tested in order to evaluate the probability of violation of this specific constraint. Although Bertsimas and Sim (2004) provide theoretical bounds for constraint violation, their approach refers to a particular condition where there is a single constraint to be violated. This characteristic differs from the studied problem as multiple constraints are being assessed (i.e., the workload constraint is indexed for *w* and *t*). Moreover these authors' approach assumes symmetrical probability distributions, an assumption which might be limiting in applications that model processing times where distributions are often known to be asymmetric (Juan et al., 2014). To overcome these difficulties, the probability bounds were calculated through Monte Carlo simulation.

As simulation requires knowledge of the probability distribution on the uncertainty set and this knowledge is unclear, random values for the processing time deviations ($\mathbf{QD_{aw}}$) were drawn from two different distributions. In this sense, a normal distribution was first applied. Additionally, to explore the effects of using a nonsymmetrical distribution, random values were generated considering a lognormal distribution, which is characterized by being non-negative, asymmetrical and skewed rightwards. The two continuous distributions were tested with both the mean and the standard deviation equal to the deterministic processing time deviation.

Within the simulation process, the assessed production plan is executed based on the random values drawn from these distributions. More specifically, the adjusted workload Constraint (40) is checked, considering that there is now an extra term referring to the processing time deviation. In this analysis, variables x_{at} and s_{at} assume the values from the original assessed plan, whereas Q_{aw}, RHE_w, OHE_w, NE_w and WH_{wt} are all parameters.

$$\sum_a Q_{aw}\, x_{at} + \sum_a \mathbf{QD_{aw}\, x_{at}} - \sum_a s_{at} \leq (RHE_w + OHE_w)\, NE_w - WH_{wt} \tag{40}$$

This process is repeated for thousands of times and the results of all iterations are aggregated in order to calculate the percentage of violation occurrences. In this sense, for each assessed production plan (which refers to a specific Γ parameter), it is possible to estimate the probability of constraint violation, that is, the probability of the plan to absorb these deviations within the fixed capacity periods, without amplifying effects to the following periods. This measures the solution robustness of the production plans that are subject to deviations in the processing times. The Monte Carlo simulation process described is summarized in the flowchart displayed in Figure 18.

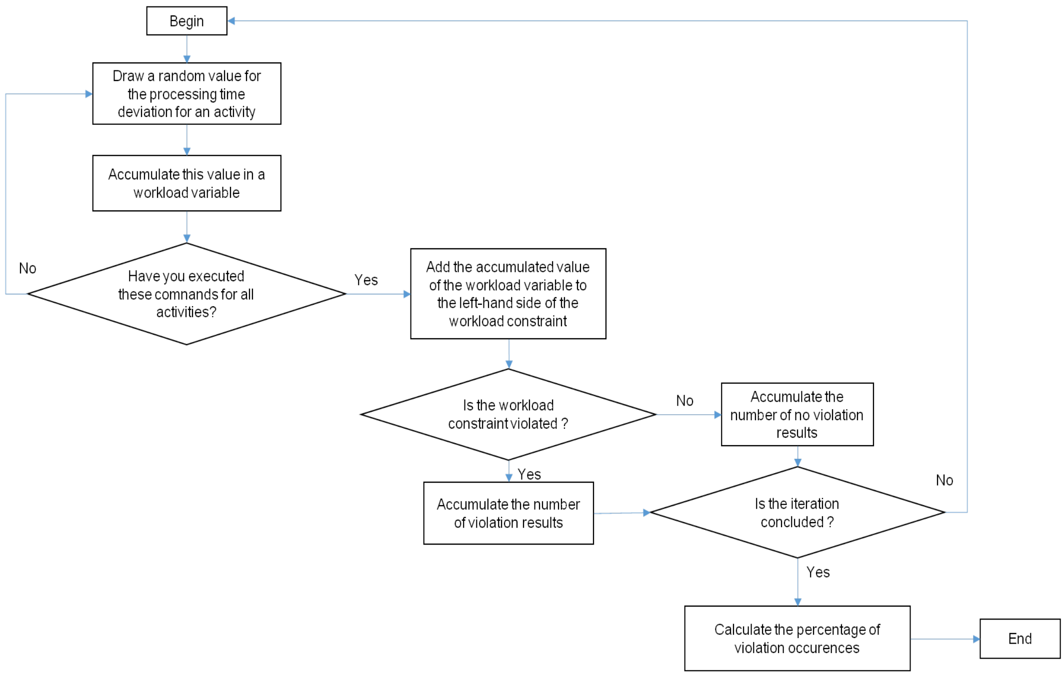


Fig. 18: Monte Carlo simulation process

5.1.4. Discussions

The robust optimization model presented in this chapter is an extension of the former deterministic model presented in Chapter 4. The proposed solution now can be described as a robust tactical capacity planning MILP model that supports the order acceptance phase of an ETO production system. It is a cost minimization model that not only considers capacity flexibility, multiple processing stages and the overlapping cadence of the production flow, but also uncertainties relative to the production process. Additionally, the proposed robust approach permits that the production plans generated by this model are assessed in terms of their robustness through the calculation of the probability bounds of constraint violation.

As seen in Table 1, the literature offers several cost minimization models that consider uncertainties in tactical planning problems, similarly to the proposed robust model. Despite these similarities, this research extends the literature in the tactical planning area for the ETO context. Considering the 33 revised models, only 2 refer to the ETO context addressing some type of uncertainty.

In terms of the model type classification, displayed in Table 1, most of the optimization models presented are deterministic ones. These models generally focus on the optimality of a solution and, therefore, discard alternative solutions with almost equivalent values for the objective function. These solutions, however, might represent an improvement to the robustness of the plans, an aspect that is valuable when uncertainties are present (Hans et al. 2007). Furthermore, the real-world setting studied is characterized by many sources of uncertainty that are typical to the ETO context. In this sense, adopting a proactive approach by incorporating a certain degree of anticipation of variability in the generated plans seems to be the appropriated way to handle uncertainty in the studied planning problem. Through this approach, suboptimal but robust solutions may be considered in order to minimize disruptions in the planning process.

Among the nondeterministic models presented in Table 1, half of them are stochastic, an approach that requires full knowledge of the distributions of uncertain data. The only two ETO models that consider uncertainty are stochastic ones (i.e., Alfieri et al., 2012; Tolio and Urgo, 2007). The drawback in adopting this approach to address the studied planning problem is that in the real-world ETO setting under study, the probability distributions of the activities processing times are not known. In this sense, robust optimization seems to be the suitable approach for developing the proposed solution. As aforementioned, according to this approach, the uncertainty of the parameters is modeled as lower and upper bounds without any need for exact distributions.

As for the types of uncertainties modelled, the proposed robust solution addresses the processing time variability of the production activities. This refers to a type of process uncertainty, which is highlighted by Khakdaman et al. (2015), as one that results in the obsolescence of tactical plans. As seen in Table 1, most models consider uncertainties related to demand. In this particular problem, the model assumes a fixed set of incoming orders and proposals to generate a production plan. In this sense, demand uncertainty is out of the scope of the proposed approach.

5.2. Application

Aiming to reduce the research-practice gap in the tactical planning area, this section presents the findings relative to the application of the proposed robust model to solve the studied problem. In this sense, a set of scenarios evaluates the behaviour of the model for different levels of protection against the workload constraint violation by varying the robust parameter Γ (i.e., each scenario refers to a specific level of protection resulting in a specific capacity plan.). The section ends with a discussion on the contributions of the practical application of the robust model as a tool to enhance and support the decision making process in the studied setting.

5.2.1. Inputs

The input data refer to the same particular circumstance that the company manufacturing team was willing to assess when validating the proposed deterministic model (See Tables 7, 8, 9 and 10 in Section 4.2.2). As presented in the former section, the robust model considers the variability relative to the processing times of the production activities. In this sense, historical data on former projects were used to estimate the processing times and the maximum expected deviations in the studied setting. For instance, for this presentation, these deviations refer to approximately 50% of the processing time, which was considered an appropriate value by the manufacturing planning team.

5.2.2. Results

The results obtained from the robust optimization approach by changing parameter Γ generated different production plans as displayed in Figure 19. It can be noted that under uncertainty, as Γ increases, the model tends to postpone more and more workload from the fixed capacity periods to future periods. In a sense, a capacity buffer is created by explicitly planning “idle” time on a work centre during the fixed capacity periods. That is, time is reserved in case uncertainties occur and the capacity buffer is therefore dimensioned to address this “extra” demand.

In particular, for $\Gamma = 0$, the robust optimization approach corresponds to the deterministic model (1)–(17) presented in Chapter 4. On the other extreme situation ($\Gamma = 85$), the plan suggests the maintenance of a maximum capacity buffer as it assumes that 85 activities (from the 250 activities, 85 may be processed within the fixed capacity periods considering the time windows of their projects) will be penalized with the maximum processing time deviation. This approach coincides with the one presented in Soyster (1973). Comparing the two situations, the former represents a levelled and smoother distribution of workload while the latter suggests a considerable increase of workload from the initial fixed capacity periods to the following periods. The intermediate values of Γ characterize the Bertsimas and Sim (2004)'s approach.

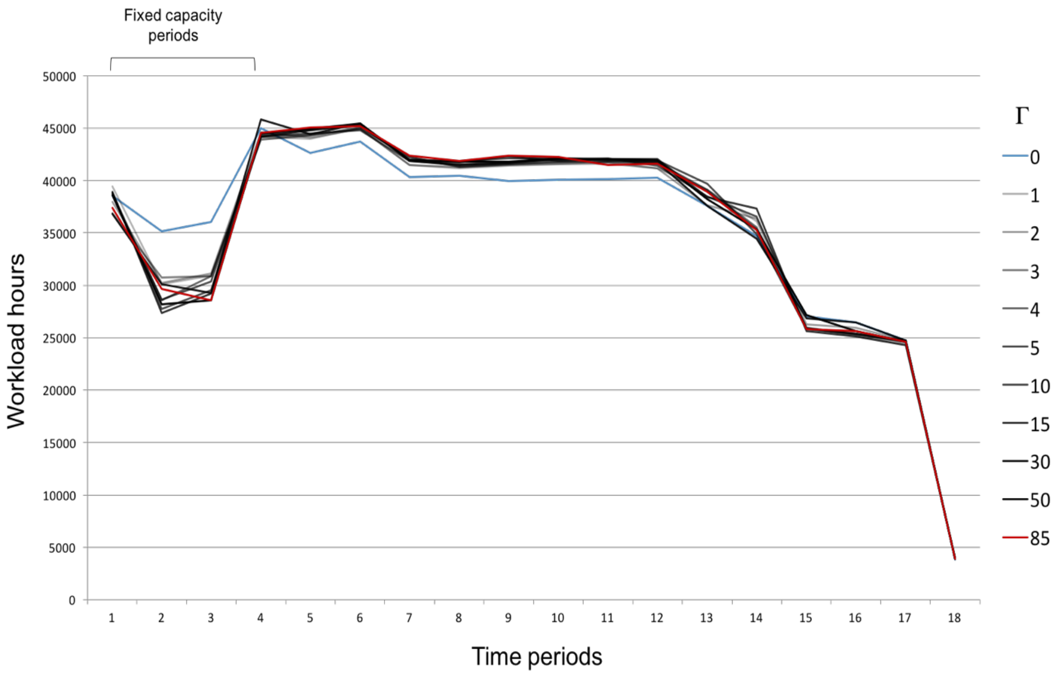


Fig. 19: Workload distributions (for different values of Γ) along the planning horizons

Making a parallel with the knapsack problem, during the fixed capacity planning periods, each work centre's capacity is equivalent to the fixed-size knapsack. The production workload corresponds to the items that are chosen to fill up the knapsack. When considering uncertainty, less workload is allocated to the work centre in a given period as the processing time deviation is also allocated to the work centre. This deviation does not appear explicitly in the production plan, as it is an "idle" time, but it represents the capacity buffer that emerged to protect the plan against uncertainty. For instance, in the ETO or MTO contexts it may be more economical to employ a capacity buffer, rather than to build an inventory buffer to cope with uncertainty. In fact, in these contexts, it may not be feasible to have an inventory buffer due to all of the possible combinations of products or to the lack of information on future demand. In this sense, creating a capacity buffer seems to be the suitable measure to provide flexibility.

Furthermore, the results obtained from the robust optimization approach by changing parameter Γ permits the evaluation of trade-offs between robustness and the total expected solution cost. Table 15 presents, for different values of Γ, the approximate probabilities of the workload constraint violation, the optimal value of the total cost (objective function) and the percentage increase in the objective function. The optimality tolerance gap considered in these experiments is 0.3%. The probability bounds were calculated according to the aforementioned procedure (in Subsection 5.1.2), where random values for the processing time deviations were drawn from a normal and a lognormal probability distributions.

Table 16: Results of the robust solutions for different values of Γ and the corresponding probability bounds of constraint violation for the normal and lognormal distributions, optimal value of the total cost and the percentage increase in the objective function

Γ	Probability bound of workload constraint violation		Optimal value of total cost ($1000)	Increase in the objective function (%)
	Normal	Lognormal		
0	0.906	0.858	31634	-
1	0.453	0.441	31742	0.34
2	0.338	0.349	31765	0.41
3	0.316	0.325	31782	0.47
4	0.300	0.320	31802	0.53
5	0.285	0.309	31816	0.57
10	0.235	0.274	31872	0.75
15	0.160	0.221	31895	0.82
30	0.150	0.197	31896	0.83
50	0.150	0.192	31900	0.84
85	0.150	0.195	31897	0.83

The deterministic solution corresponds to $\Gamma = 0$ when the minimum cost value is not increased and when there is a high probability of constraint violation for both probability distributions. Increasing the protection (i.e., increasing the value of Γ, considering that more parameters are under data uncertainty), the probability of constraint violation decreases, while the minimum cost value increases, representing the *price of robustness*. For the maximum protection case (Γ=85), there is much less chance of constraint violation for the analysed probability distributions. This corresponds to Soyster's (1973) approximation of the worst-case scenario where all uncertain parameter assumes its most adverse value.

Between the deterministic and the maximum protection solutions, there are the intermediate solutions, which represent little variation among the studied distributions. For instance, the normal distribution, which is a light-tailed distribution, quickly decreases for Γ>3. On the other hand, the long right-hand tail, which generates the assymmetry of the lognormal distribution, decreases in a slower pace for the probability bounds for Γ >3.

Figure 20 is a graphical representation of the data in Table 15. It displays the optimal value increase of total cost (%) and the probability bound of the workload constraint violation (%) as a function of Γ. As expected, more conservative Γ's present a lower probability of constraint violation and a higher cost. One can notice that this probability significantly decreases for Γ =15 for both analysed probability distributions. In this particular problem, where the processing time deviation is relatively high (i.e., 50% of the processing time) and as both distributions are supported on infinite intervals, there may be cases where the random values drawn from them represent constraint violation, even for high values of Γ. As seen, the minimum values reached for the probability of constraint violation is 15% for the normal distribution and 19,2% for the lognormal distribution. Figure 21 enlarges a portion of the same graph to reveal the differences between the two distributions.

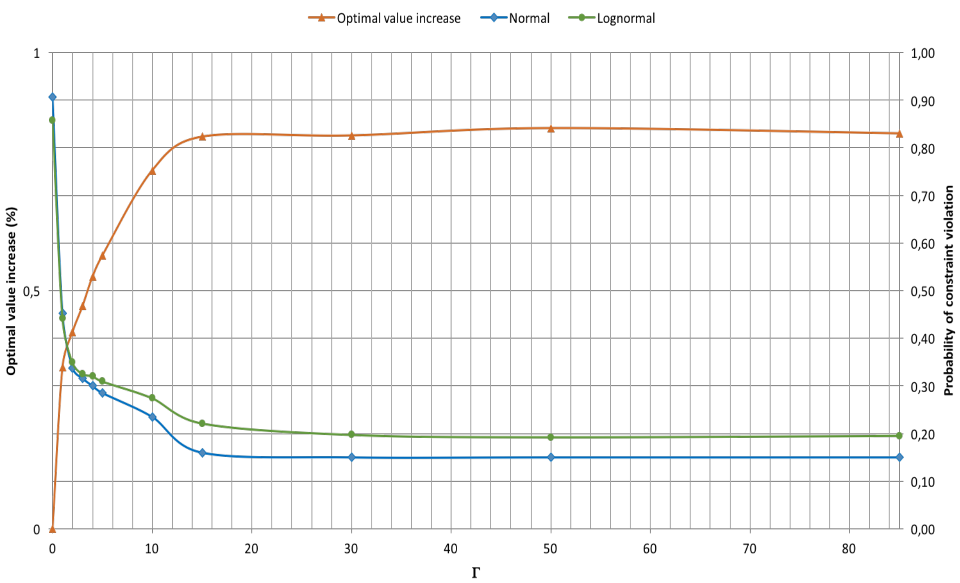


Fig. 20: Optimal value increase and probability bound of constraint violation for a normal and a lognormal distributions as a function of Γ

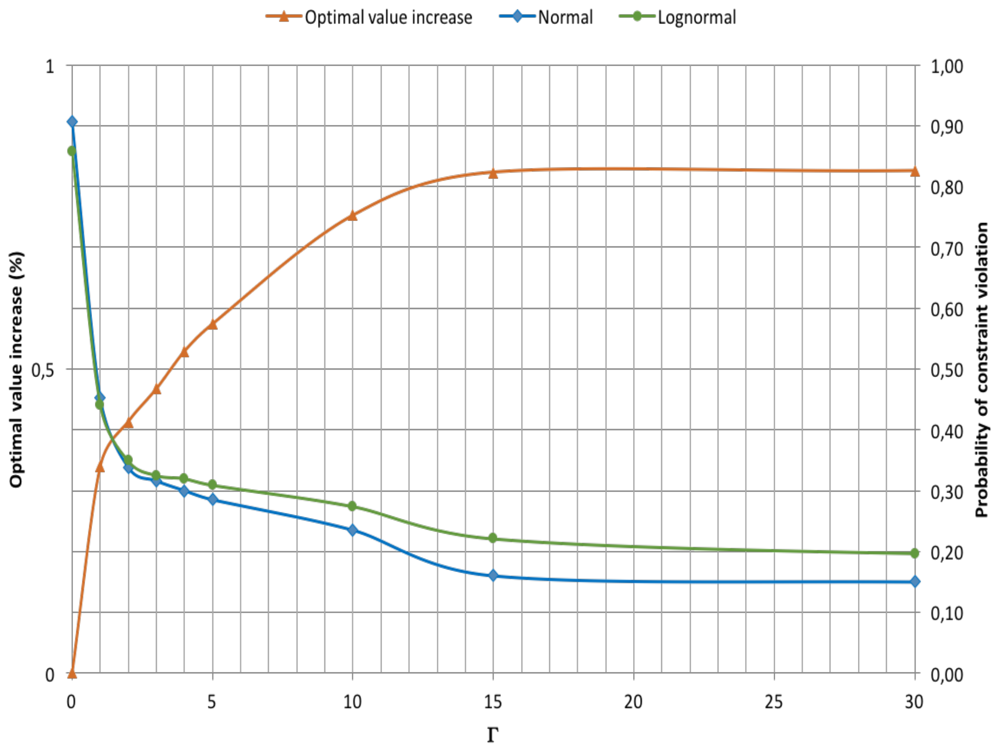
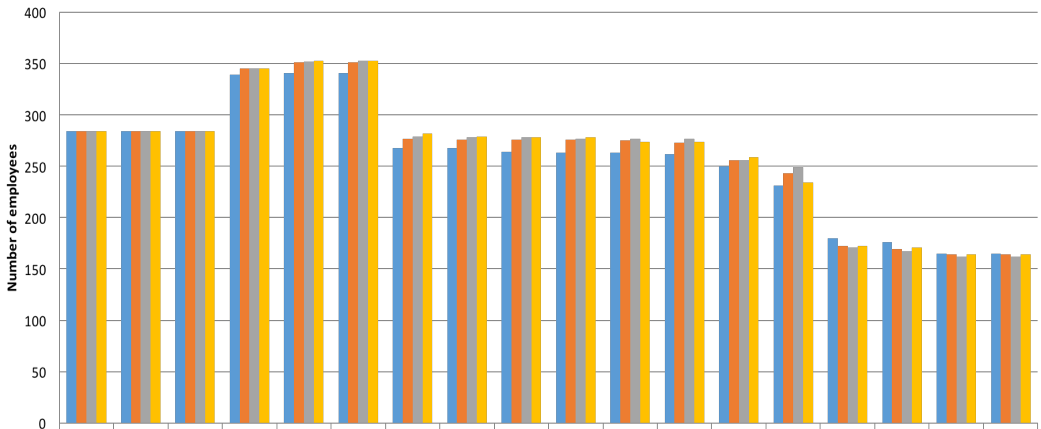


Fig. 21: **Optimal value increase and probability bound of constraint violation for a normal and a lognormal distributions as a function of Γ (zoom of a portion of Figure 20)**

In practice, postponing production results in an increasing need of production capacity in the future periods that leads to capacity changes. This is shown in Figure 22, which presents the number of employees allocated along the planning horizon for four different production plans. Depending on the decision-maker's attitude toward uncertainty, he/she can adjust the level of conservatism through the budget of uncertainty and accordingly decide on the number of employees to contract. These four plans correspond to four different values of Γ (i.e., zero, 5, 15 and 85). The first one (Γ=0) refers to the deterministic solution, which suggests that, after the fixed capacity periods (i.e. more specifically referring to the next three periods), 57 more employees would be needed. The second one (Γ=5) corresponds to a plan that is robust, with a nearly 30% of probability of constraint violation (See Table 15), but with an increase of 0.57% on the optimal total cost. This plan suggests that 67 more employees would be needed (i.e., 10 more when compared to the previous plan). The third and fourth plans, which represent more conservative solutions with nearly 15% of chance of constraint violation, results in an increase of 0.83% on the optimal cost and suggests that 69 more employees would be needed. Since contracting and training personnel takes time, this decision might have to be taken in the short-term.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
0	284	284	284	339	341	341	268	268	264	263	263	262	250	231	180	176	165	165
5	284	284	284	345	351	351	277	276	276	276	275	273	256	243	172	169	164	164
15	284	284	284	345	352	353	279	278	278	277	277	277	256	249	171	167	162	162
85	284	284	284	345	353	353	282	279	278	278	274	274	259	234	172	171	164	164

Fig. 22: Employees contracted (for Γ = 0, 5, 15 and 85) along the planning horizon

At last, in terms of results, given that the robust models for each value of Γ are all linear programming formulations, the computational time to solve them were acceptable and consumed approximately the same solution time of the deterministic model (around 1,000 seconds) in the standard computer used to run the models.

5.2.3. Discussions

This research aims to contribute to reduce the research-practice gap by providing information on relevant issues that must be considered and addressed in a real-world ETO context in order to develop and implement a tactical capacity planning solution. One of the relevant issues modelled refers to process uncertainty, which is characteristic of the ETO context and was highlighted as relevant by the studied company's manufacturing planning team. In this sense the robust optimization model proposed in this chapter extends the former deterministic model addressing the same tactical capacity planning problem by incorporating uncertainty relative to the activities processing times.

In a final validation session, held within the fifth technical visit, the planning team provided feedback relative to the application of the robust model to the real-world planning problem. To assess the adherence of the proposed solution to the studied problem, a set of scenarios, using the former real-world database, was generated considering process uncertainties. By analysing these scenarios, the company's planning team recognized the potentiality of the proposed solution to support decision makers to cope with uncertainty. Within this validation session, several issues, detailed in the sequel, were discussed relative to the robust solution's characteristics and usability.

Acquiring the required input data

In terms of the required data, the planning team concluded that the input data relative to uncertainty could be obtained in an accessible, uncomplicated manner. According to the team, it is reasonable or even instinctive to deduce bounds for the variability of uncertain parameters, based on tacit knowledge and on historical data of previous projects, whereas, determining exact values for the uncertain parameters and specifying their probability distribution would be much more complex for the studied company.

Differentiating uncertainties

As the projects have different characteristics, as ones may be more innovative (i.e., projects that differ considerably from former projects) than others, it is essential, for the company team, to attribute different values of uncertainty among the incoming projects. Moreover, the labor-intensive activities (such as assembly and welding) are more subject to variability than operations that mainly depend on the use of machines (e.g., machining). In this sense the model is prepared to cope with these specificities.

Representing the manufacturing planning team's attitude

The planning team approved the fact that the proposed robust model permits the adjustment of the decision maker's attitude towards uncertainty. They believe that the manufacturing planning team's perspective may be represented by the model in different planning situations. For instance, in cases regarding more innovative projects, projects subject to tight deadlines (i.e., with high financial penalties for late delivery) or projects with critical materials that are imported (i.e. which have a long supply leadtime), they believe they may adopt a more conservative attitude than in cases where projects seem to be similar to previous ones in terms of design characteristics and materials used.

Evaluating trade-offs

After analyzing simultaneously the set of scenarios generated by the robust model, the company's planning team understood that the proposed solution might be useful to assess trade-offs between robustness and the total expected solution cost. In this sense, the model may assist the decision makers in choosing the suitable level of conservatism for the capacity plan by evaluating the model's behavior, in terms of the optimal cost increase, for different levels of protection against the workload constraint violation.

The planning team also gained new insights on how to apply the proposed solution in the studied setting. For instance, the planning team agreed that uncertainty should be considered within the more imminent planning periods (the beginning of the planning horizon) since the impact of distant uncertain events are less relevant in terms of the short-term decisions made. Moreover, since uncertainty is only considered at the beginning of the planning horizon, the variability may also represent a postponement in the starting time of specific components associated to delays in the delivery of supply and project designs to the shop floor and delays in the auditing process imposed by the customers during the production process. In a sense, due to the assumptions considered in the robust model, parameter variability may not be exclusively related to process uncertainty as external (to production) uncertainties referring to supply and customer uncertainties may also be regarded.

Additionally, the planning team realized that the robust plans are for capacity planning and not necessarily for production planning. The robust plans specify the capacity level for non-fixed capacity periods considering a capacity buffer within the fixed capacity periods. In this sense, production is postponed so that the plan may absorb the uncertainties that may occur within the fixed capacity periods. On the other hand, this postponement of production does not reflect what should be effectively planned at the operational level. For instance, when considering uncertainty, less workload should be allocated to the fixed capacity periods, which does not necessarily correspond to the orders that should be dispatched to the shop floor according to the lower-level production schedules. Likewise, this is in accordance to how the company's planning team addresses situations with innovative projects that represent high levels of uncertainty.

In terms of the expected impacts of adopting this planning solution as a substitute for the current planning method, the planning team claims that the main change is to move from a limited (few alternatives, time-consuming method) ''what-if'' analysis of scenarios, without guarantee of optimization (no cost assessment), and no detailed view of capacity to an enhanced (automatically generated by the solver) ''what-if'' analysis of optimized (cost assessment) robust (considering uncertainties) scenarios, with a detailed view of capacity to support decision making.

Particularly, among all the improved issues, the incorporation of uncertainty was considered the most innovative aspect of the proposed solution since it refers to an issue that is difficult to quantify, but that, according to the team, needs to be considered during the planning process in a proactive manner. More specifically, the team understands that the proposed solution should minimize the effort needed in the replanning process, as it generates robust (stable) plans and provides information that permits the identification of actions to be taken in the short-term to address critical situations that may involve renegotiations with the customers, suppliers, the design department and other internal departments.

6 Conclusion and further developments

This chapter summarizes the main contributions of this action research and discusses directions for future work.

6.1.Conclusion

This action research addresses the existing gap between theory and practice with respect to the development of decision support tools for tactical capacity planning functions, especially for ETO organizations. Within this context, the goal of this thesis was to present a tactical capacity planning solution to support the order acceptance phase of a medium-sized multi-project ETO organization. This research study laid in the development and application of two MILP models for this studied setting: (i) a deterministic model, in which modelling issues that are either not entirely explored in other studies or that have to be adapted to the specificities of the studied setting are taken into account, and (2) a robust optimization model which extends the former model by considering uncertainties of the planning problem. Furthermore, the intention of this research was to contribute to the literature by presenting relevant issues that must be considered to properly address this particular but real-world planning problem. This action research provided evidence from a single firm, which limits the extent to which the findings can be generalized. Nevertheless, the ETO context itself offers research attractiveness to the problem because it has received much less attention from researchers than MTS settings, for instance. Additionally, the findings of this study yield knowledge and lessons that are of interest to academics and practitioners, adding empirical matter for enriching the literature in this research area, thereby narrowing the gap between theory and practice.

From a methodological stance, the precise definition of the problem scope (i.e., the choice of focusing on the workload analysis, a specific but relevant planning function) was fundamental to properly address a real-world planning problem. Furthermore, the prototyping approach adopted in this action research, involving the company's manufacturing team in conjunction with the researchers throughout the project, was essential to develop and to validate the model and to assure its adherence to the environment's complex reality. Additionally, the research method comprised a continuous literature review on the topics directly related to the studied problem that helped to gain insights that contributed to the development of the proposed models. Thus, modelling this planning problem was an incremental learning process for the researcher and the company's manufacturing planning team.

Referring specifically to the attributes highlighted in Table 1, this thesis contributes to the academic literature by narrowing the gaps relative to the use of mathematical models that support decision makers to solve middle-level capacity planning problems in ETO contexts. The following items summarize the main findings relative to these attributes:

- The proposed models were developed addressing an ETO organization, a type of context which has received much less attention from researchers when compared to standardized MTS contexts.
- These models support the workload analysis function in the order acceptance phase. This is an important characteristic as, in this phase, adequate capacity planning methods that assess the consequences of decisions for the production system are crucial and there is a lack of research studies relative to models addressing this phase in ETO planning problems.
- The proposed models consider nonregular capacity by explicitly representing overtime and subcontracting and adjusting the internal capacity workforce by hiring and firing personnel. The implementation of nonregular capacity, an issue not fully explored in the ETO setting, was considered a relevant point for the solution's suitability.
- The proposed models admit multiple processing stages when representing the production flow. Although this modelling issue is not a novelty in the literature, from a practitioner's perspective, it enhances the decision making process for the studied company when compared with the company's current planning method (which

aggregates all stages in one), by admitting the representation of a production flow in more details. This helps in providing practitioners a decision support tool, a gap identified in the literature.

- The models were enhanced with the inclusion of the cadence constraints, which refer to the representation of the relationships among interconnected activities. These relationships comprise three characteristics: (i) the execution progress mode of each activity varies along the time periods, (ii) the activities overlap in time successively and (iii) the precedence relationship between two activities is not fixed, because one may precede in some periods of time and succeed in other periods. This modelling issue concerns a finding apparently not yet reported in literature.
- As the real-world studied setting is characterized by many sources of uncertainty, incorporating a certain degree of anticipation of variability in the generated plans seems the appropriate way to cope with uncertainty. Therefore, the proposed robust model refers to a proactive planning approach by considering the processing time variability of the production activities. This particular issue contributes to literature as there are few research papers, within the revised ones, addressing some type of uncertainty in the ETO context.
- The model type (i.e., whether deterministic, robust, or stochastic) was one the issues discussed. In particular, the drawback in adopting a stochastic approach to address the studied planning problem is that the probability distributions of the uncertain parameters are not known. In this sense, the robust optimization approach (i.e., where the uncertainty of the parameters is modeled as lower and upper bounds without any need for exact distributions) seems to be more adequate for developing the proposed solution. In fact, to the best of our knowledge, this is the first robust optimization model for tactical capacity planning that explicitly addresses the ETO context.

- Finally, the empirical nature (i.e., working with a real-world planning problem) of this action research contributed to identify the issues that were relevant to address this particular problem and to provide a more pragmatic view of what can be obtained from mathematical programming approaches to solve real world tactical capacity planning problems in the ETO context.

From the company's manufacturing planning team standpoint, the models not only address different shortcomings of the company's current planning method but enhance the decision-making process on order acceptance. Although this process depends on additional information relative to other departments within the company (i.e., the decision to accept or reject a specific customer query depends on a series of criteria that are considered by different staffs, besides the industrial manufacturing division), the models contribute to it by improving the workload analysis function. For instance, these models represent diagnostic and decision support tools as they help identify the potential gaps between capacity and demand and demonstrate how to adapt capacity to demand and how to optimally balance demand. The models allow the planning team to intervene in the construction of the production plans with managerial decisions that change input data relative to demand, supply, employees and the production flow in order to generate concurrent new analysis. Furthermore, the more detailed and precise information about the planning problem, provided by the proposed models, permits a timely assessment of possible shortcomings and the identification of the primary actions to be taken in order to accept incoming orders.

Referring particularly to the robust model, the company's planning team concluded that its input data can be obtained in an accessible manner and appreciated that different levels of uncertainty could be attributed among the projects and activities. The team approved the fact that the model permits the adjustment of the attitude towards uncertainty and understood that the model is useful to assess trade-offs robustness and total expected cost. As for the expected impacts, the team believes that the robust solution not only represents a proactive planning approach, but that it should minimize the effort needed in the replanning process, as it generates robust plans.

Moreover, relevant insights were also discussed along this action research. For instance, when considering uncertainty, a capacity buffer is employed within the robust plans generated as less workload is allocated within the initial time periods (i.e., the fixed capacity periods). In fact, in ETO contexts it may be more economical to employ this type of buffer rather than maintaining high levels of inventory to cope with uncertainty. Furthermore, when assessing the robust plans, two continuous probability distribution functions, the normal and the lognormal, were used to estimate the probability bounds of constraint violation, through Monte Carlo simulation. The results of these experiments reinforce the resilience of Bertsimas and Sim (2004)'s approach. More specifically, the plans remain stable regardless of the probabilistic description of variability as the distributions tested converge and indicate the same level of protection (Γ value) to significantly reduce the chances of constraint violation.

Referring to the outcomes obtained within the application of the proposed models, empirical results suggest, for instance, that with a slight increase in cost (0.02%) a part component should be processed in-house instead of being outsourced and that with a 0.8% increase in cost (which includes hiring 21% more personnel) the probability of violating the production plans decreases from 90% to 15%, representing a much more stable (protected against uncertainty) situation. Furthermore, the computational time to solve the tactical capacity planning problem under study, by both of the proposed models, refers to 1,000 seconds in a standard computer, which was considered acceptable by the company's manufacturing planning team.

6.2. Further developments

Because comparable empirical studies in the literature are still rare, the present study contributes to the still scant body of empirical knowledge on tactical capacity planning in ETO production settings. Future studies may want to make a more systematic and focused investigation to test and expand on our findings to fill the theory and practice gap in this area. For instance, this research could be extended in several ways.

An important question remains as to what extent the proposed solution is applicable in other contexts. It would be interesting to investigate whether it fits other ETO production settings or even the MTO context, particularly in situations that lack information in the order acceptance phase. In this sense, it would be necessary to check if the modelling issues considered in this research adequately represent these contexts or if new relevant aspects need to be analysed and modelled.

In addition, more efficient MILP formulations could be developed, an issue that was not one of the priorities in this action research. The problem instance analysed comprised approximately 30.000 variables (18.000 integer variables) and almost 130.000 constraints and resulted in a solution time of 1000 seconds. If the problems increase in size and complexity, it may be crucial to seek more efficient formulations.

Another aspect which deserves further investigation refers to the choice of the objective goal of the proposed models (which minimize the overall production cost). In the setting under study, the company's manufacturing planning team has little information on the projects' prices which made it difficult to adopt an objective goal that would maximize profit. On the other hand, minimizing costs without precisely knowing which resources are the production system's bottlenecks may result in profit losses. In this sense, to address this shortcoming, the proposed solution could be extended by adopting the concepts of throughput accounting in the Theory of Constraints.

The development of a user-friendly interface to transform these models into a computational planning system for the company´s manufacturing planning team seems to be another logical next step in further research. Within this direction, one relevant issue refers to the availability, reliability and integrity of information as well as the good use of information technology. For instance, in the studied setting, information is stored in different locations and in different formats and may not be easily available. To ensure the acquisition of all relevant data from these locations, it may be necessary to define communication protocols between information systems.

And finally, an additional aspect refers to the implementation method of this computational planning system as an effective tactical planning tool for the studied setting. This is an interdisciplinary matter as technological, organizational and human issues should be considered simultaneously within this implementation process.

References

Aghezzaf, E., Sitompul, C., Najid, N., 2010. Models for robust tactical planning in multi-stage production systems with uncertain demands. Computers and Operations Research 37, 880-889.

Alem, D., Morabito, R., 2012. Production planning in furniture settings via robust optimization. Computers and Operations Research 39, 139-150.

Alfieri, A., Tolio, T., Urgo, M., 2011. A project scheduling approach to production planning with feeding precedence relations. International Journal of Production Research, 49(4), 995-1020.

Alfieri, A., Tolio, T., Urgo, M., 2012. A two-stage stochastic programming project scheduling approach to production planning. The International Journal of Advanced Manufacturing Technology, 62(1-4), 279-290.

Alp, O., Tan, T., 2008. Tactical capacity management under capacity flexibility. IIE Transactions, 40(3), 221-237.

Aouam, T., Brahimi, N., 2013. Integrated production planning and order acceptance under uncertainty: A robust optimization approach. European Journal of Operational Research, 228, 504-515.

Artigues, C., Kone, O., Lopez, P., Mongea, M., 2015. Mixed-integer linear programming formulations. In: Schwindt, C., Zimmermann, J. (Eds.), Handbook on Project Management and Scheduling. Springer International Publishing, Switzerland, Vol. 1, 17–41.

Artigues, C., Leus, R., Nobibon, F., 2013. Robust optimization for resource-constrained project scheduling with uncertain activity durations. Flexible Services and Manufacturing Journal, 25, 175–205.

Aslan, B., Stevenson, M., Hendry, L., 2012. Enterprise Resource Planning systems: An assessment of applicability to Make-To-Order companies. Computers in Industry, 63, 692–705.

Ballestin, F., Schwindt, C., Zimmermann, J., 2007. Resource Leveling in Make-to-Order production: Modeling and Heuristic Solution Method. International Journal of Operations Research, 4(1), 1-13.

Bertrand, J.W.M., Muntslag, D.R., 1993. Production control in engineer-to-order firms. International Journal of Production Economics, (30-31), 3-22.

Bertsimas, D., Sim, M., 2004. The price of robustness. Operations Research, 52(1), 35–53.

Bertsimas, D., Thiele, A., 2006. Robust and Data-Driven Optimization: Modern Decision-Making Under Uncertainty. Tutorial.

Bianco, L., Caramia, M., 2011. Minimizing the completion time of a project under resource constraints and feeding precedence relations: a Lagrangian relaxation based lower bound. Technical report DII 03-09, University of Rome Tor Vergata, 4OR, forthcoming.

Bianco, L., Caramia, M., 2012. An exact algorithm to minimize the makespan in project scheduling with scarce resources and generalized precedence relations. European Journal of Operational Research, 219, 73-85.

Bredstrom, D., Flisberg, P., Ronnqvist, M., 2013. A new method for robustness in rolling horizon planning. International Journal of Production Economics, 143, 41–52.

Bushuev, M., 2014. Convex optimization for aggregate production planning. International Journal of Production Research, 52(4), 1050-1058.

Buxey, G., 2003. Strategy not tactics drives aggregate planning. International Journal of Production Economics, 85, 331–346.

Buxey, G., 2005. Aggregate planning for seasonal demand: reconciling theory with practice. International Journal of Operations & Production Management, 25(11), 1083-1100.

Cameron, N.S., Braiden, P.M., 2004. Using business process re-engineering for the development of production efficiency in companies making engineered to order products. International Journal of Production Economics, 89, 261–273.

Carvalho, A.N., Oliveira, F., Scavarda, L.F., 2015. Tactical capacity planning in a real-world ETO industry case: An action research. International Journal of Production Economics, 167, 187–203.

Chtourou, H., Haouari, M., 2008. A two-stage-priority-rule-based algorithm for robust resource-constrained project scheduling. Computers & Industrial Engineering, 55, 183–194.

Corominas, A., Lusa, A., Olivella, J., 2012. A detailed workforce planning model including non-linear dependence of capacity on the size of the staff and cash management. European Journal of Operational Research, 216, 445–458.

Corti, D., Pozzetti, A., Zorzini M., 2006. A capacity-driven approach to establish reliable due dates in a MTO environment. International Journal of Production Economics, 104, 536–554.

Coughlan, P., Coghlan, D., 2002. Action research for operations management. International Journal of Operations and Production Management, 22(2), 220–240.

Deblaere, F., Demeulemeester, E., Herroelen, W., 2011. Proactive policies for the stochastic resource-constrained project scheduling problem. European Journal of Operational Research, 214, 308–316.

Díaz-Madroñero, M., Mula, J., Peidro, D., 2014. A review of discrete-time optimization models for tactical production planning. International Journal of Production Research, 52(17), 5171-5205.

Ebben, M.J.R., Hans, E.W., Weghuis, F.M.O., 2005. Workload based order acceptance in job shop environments. OR Spectrum, 27, 107-122.

Gademann, N., Schutten, M., 2005. Linear-programming-based heuristics for project capacity planning. IIE Transactions, 37, 153–165.

Genin, P., Lamouri, S., Thomas, A., 2008. Multi-facilities tactical planning robustness with experimental design. Production Planning & Control: The Management of Operations, 19(2), 171-182.

Giebels, M., 2000. EtoPlan a Concept for Concurrent Manufacturing Planning and Control - Building holarchies for manufacture-to-order environments. PhD thesis, University of Twente, Enschede.

Gorissen, B.L., Yanikoglu, I., den Hertog, D., 2015. A practical guide to robust optimization. Omega, 53, 124–137.

Gosling, J., Naim, M., 2009. Engineer-to-order supply chain management: A literature review and research agenda. International Journal of Production Economics, 122, 741-754.

Grabenstetter, D.H., Usher, J.M., 2013. Determining job complexity in an engineer-to-order environment for due date estimation using a proposed framework. International Journal of Production Research, 51(19), 5728-5740.

Grabenstetter, D., Usher, J., 2014. Developing due dates in an engineer to order engineering environment. International Journal of Production Research, 52(21), 6349-6361.

Hans, E.W., 2001. Resource loading by branch-and-price techniques. PhD thesis, University of Twente, Enschede.

Hans, E.W., Herroelen, W., Leus, R.,Wullink, G., 2007. A hierarchical approach to multi-project planning under uncertainty. Omega, 35, 563–577.

Herroelen, W., Leus, R., 2004. Robust and reactive project scheduling: a review and classification of procedures. International Journal of Production Research, 42(8), 1599-1620.

Hicks, C., Braiden, P. M., 2000. Computer-aided production management issues in the engineer-to-order production of complex capital goods explored using a simulation approach. International Journal of Production Research, 38(18), 4783- 4810.

Hicks, C., McGovern, T., Earl, C.F., 2000. Supply chain management: A strategic issue in engineer-to-order manufacturing. International Journal of Production Economics, 65, 179-190.

Hicks, C., McGovern, T., Earl, C.F., 2001. A Typology of UK Engineer-to-Order Companies. International Journal of Logistics: Research and Applications, 4(1), 43-56.

Huang, S., Lu, M., Wan, G., 2011. Integrated order selection and production scheduling under MTO strategy. International Journal of Production Research, 49(13), 4085-4101.

Ishii, N., Takano, Y., Muraki, M., 2014. An order acceptance strategy under limited engineering man-hours for cost estimation in Engineering–Procurement– Construction projects. International Journal of Project Management 32, 519-528.

Jahangirian, M., Eldabi, T., Naseer, A., Stergioulas, L., and Young, T., 2010. Simulation in manufacturing and business: a review. European Journal of Operational Research, 203, 1-13.

Jamalnia, A., Feili, A., 2013. A simulation testing and analysis of aggregate production planning strategies. Production Planning & Control: The Management of Operations, 24(6), 423-448.

Juan, A., Barrios, B., Vallada, E., Riera, D., Jorba, J., 2014. A simheuristic algorithm for solving the permutation flow shop problem with stochastic processing times. Simulation Modelling Practice and Theory, 46, 101–117.

Kalantari, M., Rabbani, M., Ebadian, M., 2011. A decision support system for order acceptance/rejection in hybrid MTS/MTO production systems. Applied Mathematical Modelling 35, 1363–1377.

Khakdaman, M., Wong, K., Zohoori, B., Tiwari, M., Merkert, R., 2015. Tactical production planning in a hybrid Make-to-Stock–Make-to-Order environment under supply, process and demand uncertainties: a robust optimisation model. International Journal of Production Research, 53(5), 1358-1386.

Kingsman, B.G., 2000. Modelling input-output workload control for dynamic capacity planning in production planning systems. International Journal of Production Economics, 68, 73-93.

Kis, T., 2005. A branch-and-cut algorithm for scheduling of projects with variable-intensity activities. Mathematical Programming, 103(3), 515-539.

Lagemann, H., Meier, H., 2014. Robust capacity planning for the delivery of Industrial Product-Service Systems. Procedia CIRP 19, 99 – 104.

Lingitz, L., Morawetz, C., Gigloo, D.T., Minner, S., Sihn, W., 2013. Modelling of flexibility costs in a decision support system for midterm capacity planning. Procedia CIRP, 7, 539-544.

Little, D., Rollins, R., Peck, M., Porter, J.K., 2000. Integrated planning and scheduling in the engineer-to-order sector. International Journal of Computer Integrated Manufacturing, 13(6), 545-554.

Liu, J., Lin, Z., Chen, Q., Mao, N., Chen, X., 2013. A decision support to assign mould due date at customer enquiry stage in computer-integrated manufacturing (CIM) environments. International Journal of Computer Integrated Manufacturing, 26(6), 571-582.

Lusa, A., Pastor, R., 2011. Planning working time accounts under demand uncertainty. Computers & Operations Research 38, 517–524.

Markus, A., Vancza, J., Kis, T., Kovacs, A., 2003. Project Scheduling Approach to Production Planning. Annals of the CIRP, 52(1), 359-362.

Mestry, S., Damodaran, P., Chen, C., 2011. A branch and price solution approach for order acceptance and capacity planning in make-to-order operations. European Journal of Operational Research, 211, 480–495.

Mincsovics, G., Dellaert, N., 2009. Workload-dependent capacity control in production-to-order systems. IIE Transactions, 41(10), 853-865.

Monostori, L., Erdos, G., Kadar, B., Kis, T., Kovacs, A., Pfeiffer, A., Vancza, J., 2010. Digital enterprise solution for integrated production planning and control. Computers in Industry, 61, 112–126.

Montreuil, B., Labarthe, O., Cloutier, C., 2013. Modelling client profiles for order promising and delivery. Simulation Modelling Practice and Theory, 35, 1-25.

Mourtzis, D., Doukas, M., Fragou, K., Efthymiou, K., Matzorou, V., 2014. Knowledge-based estimation of manufacturing lead time for complex engineered-to-order products. In Proceedings of the 47th CIRP Conference on Manufacturing Systems, Procedia CIRP 17, 499 – 504.

Mulvey, J., Vanderbei, R., Zenios, S., 1995. Robust Optimization of Large-Scale Systems. Operations Research, 43(2), 264-281.

Munhoz, J. R., Morabito, R., 2014. Optimization approaches to support decision making in the production planning of a citrus company: A Brazilian case study. Computers and Electronics in Agriculture 107, 45-57.

Nam, S.J., Logendran, R., 1992. Aggregate production planning – a survey of models and methodologies. European Journal of Operational Research, 61, 255-272.

Neumann, K., Schwindt, C., 2002. Project scheduling with inventory constraints. Mathematical Methods of Operations Research, 56, 513‑533.

Neumann, K., Zimmermann, J., 2000. Procedures for Resource Leveling and Net Present Value Problems in Project Scheduling with General Temporal and Resource Constraints. European Journal of Operational Research, 127, 425-443.

Nobibon, F., Leus, R., Nip, K., Wang, Z., 2015. Resource loading with time windows. European Journal of Operational Research, 244, 404–416.

Olhager, J., 2003. Strategic positioning of the order penetration point. International Journal of Production Economics, 85(3), 319-329.

Pandit, A., Zhu, Y., 2007. An ontology-based approach to support decision-making for the design of ETO (Engineer-To-Order) products. Automation in Construction, 16, 759–770.

Pero, M., Rossi, T., 2014. RFID technology for increasing visibility in ETO supply chains: a case study. Production Planning & Control: The Management of Operations, 25(11), 892-901.

Policella, N., Smith, S., Cesta, A., Oddi, A., 2004. Generating Robust Schedules through Temporal Flexibility. ICAPS-04 Proceedings.

Powell, D., Strandhagen, J., Tommelein, I., Ballard, G., Rossi, M., 2014. A New Set of Principles for Pursuing the Lean Ideal in Engineer-to-Order Manufacturers. In Proceedings of the 47th CIRP Conference on Manufacturing Systems, Procedia CIRP 17, 571 – 576.

Quintanilla, S., Pérez, A., Lino, P., Valls, V., 2012. Time and work generalised precedence relationships in project scheduling with pre-emption: an application to the management of Service Centres. European Journal of Operational Research, 219, 59-72.

Radke, A.M., Tseng, M.M., 2012. A risk management-based approach for inventory planning of engineering-to-order production. CIRP Annals - Manufacturing Technology, 61, 387–390.

Radke, A.M., Tolio, T., Tseng, M.M., Urgo, M., 2013. A risk management-based evaluation of inventory allocations for make-to-order production. CIRP Annals, Manufacturing Technology, 62(1), 459-462.

Rahmani, D., Ramezanian, R., Fattahi, P., Heydari, M., 2013. A robust optimization model for multi-product two-stage capacitated production planning under uncertainty. Applied Mathematical Modelling, 37, 8957-8971.

Ramezanian, R., Rahmani, D., Barzinpour, F., 2012. An aggregate production planning model for two phase production systems: Solving with genetic algorithm and tabu search. Expert Systems with Applications, 39, 1256–1263.

Rieck, J., Zimmermann, J., Gather, T., 2012. Mixed-Integer Linear Programming for Resource Leveling Problems. European Journal of Operational Research, 221, 27-37.

Rom, W.O., Slotnick, S.A., 2009. Order acceptance using genetic algorithms. Computers & Operations Research, 36, 1758-1767.

Sawik, T., 2009. Multi-objective due-date setting in a make-to-order environment. International Journal of Production Research, 47(22), 6205-6231.

Schwindt, C., Paetz, T., 2015. Continuous preemption problems. In: Schwindt, C., Zimmermann, J. (Eds.), Handbook on Project Management and Scheduling. Springer International Publishing, Switzerland, Vol. 1, 251–295.

Schwindt, C., Trautmann, N., 2000. Batch scheduling in process industries: an application of resource–constrained project scheduling. OR Spektrum, 22, 501–524.

Sharda, B., Akiya, N., 2012. Selecting make-to-stock and postponement policies for different products in a chemical plant: A case study using discrete event simulation. International Journal of Production Economics, 136(1), 161-171.

Sillekens, T., Koberstein, A., Suhl, L., 2011. Aggregate production planning in the automotive industry with special consideration of workforce flexibility. International Journal of Production Research, 49(17), 5055-5078.

Slotnick, S.A., Morton, T.E., 2007. Order acceptance with weighted tardiness. Computers & Operations Research, 34, 3029-3042.

Soyster, A. L., 1973. Convex programming with set-inclusive constraints and applications to inexact linear programming. Operations Research 21, 1154–1157.

Tolio, T., Urgo, M., 2007. A Rolling Horizon Approach to Plan Outsourcing in Manufacturing-to-Order Environments Affected by Uncertainty. CIRP Annals, Manufacturing Technology 01.

Tunali, S., Ozfirat, P.M., Ay, G., 2011. Setting order promising times in a supply chain network using hybrid simulation-analytical approach: An industrial case study. Simulation Modelling Practice and Theory, 19, 1967–1982.

Van de Vonder, S., Demeulemeester, E., Herroelen, W., 2008. Proactive heuristic procedures for robust project scheduling: An experimental analysis. European Journal of Operational Research 189, 723–733.

Wang, S., Yeh, M., 2014. A modified particle swarm optimization for aggregate production planning. Expert Systems with Applications, 41, 3069–3077.

Wang, X., Xie, X., Cheng, T.C.E., 2013. Order acceptance and scheduling in a two-machine flowshop. International Journal of Production Economics, 141, 366-376.

Willner, O., Powell, D., Duchi, A., Schönsleben, P., 2014.Globally Distributed Engineering Processes: Making the Distinction between Engineer-to-order and Make-to-order. In Proceedings of the 47th CIRP Conference on Manufacturing Systems, Procedia CIRP 17, 663 – 668.

Wullink, G., Gademann, A.J.R.M., Hans, E.W., van Harten, A., 2004. Scenario-based approach for flexible resource loading under uncertainty. International Journal of Production Research, 42(24), 5079–98.

Yang, L., 2013. Key practices, manufacturing capability and attainment of manufacturing goals: The perspective of project/engineer-to-order manufacturing. International Journal of Project Management, 31, 109–125.

Zhen, L., 2012. Analytical study on multi-product production planning with outsourcing. Computers & Operations Research, 39, 2100–2110.

Zhong, X., Ou, J., Wang, G., 2014. Order acceptance and scheduling with machine availability constraints. European Journal of Operational Research, 232, 435-441.

Zijm, W.H.M., 2000. Towards intelligent manufacturing planning and control systems. OR Spectrum, 22, 313-345.

Zorzini, M., Corti, D., Pozzetti, A., 2008. Due date (DD) quotation and capacity planning in make-to order companies: Results from an empirical analysis. International Journal of Production Economics, 112, 919-933.

Appendix 1

The mathematical formulation of the robust optimization model is detailed in the sequel. The constraint equations are referenced by the same numbers used in the main text.

Minimize

$$\sum_{a,w,t}[CR_w(r_{at} + o_{at}) + CO_w\, o_{at}] + \sum_{w,t}[CR_w\ (wr_{wt} + wo_{wt}) + CO_w wo_{wt} + CC\,(eh_{ttw} + \sum_{l}^{t} ef_{ltw}) + CS_w\, ea_{wt}] + \sum_i PS_i\ (1 - d_i)$$

$$\sum_{t=RD_a}^{DL_a} x_{at} = 1, \quad \forall\, a \tag{1}$$

$$e_{at} \geq \sum_{l=1}^{t} x_{al} \qquad \forall a, RD_a \leq t \leq DL_a \tag{2}$$

$$f_{at} \leq \sum_{l=1}^{t} x_{al} \qquad \forall a, RD_a \leq t \leq DL_a \tag{3}$$

$$w_{at} = e_{at} - f_{at} \qquad \forall\, a, t \tag{4}$$

$$x_{at} \geq MN_a\, w_{at} \qquad \forall a, RD_a \leq t \leq DL_a \tag{5}$$

$$x_{at} \leq MX_a\, w_{at} \qquad \forall a, RD_a \leq t \leq DL_a \tag{6}$$

$$z_{at} \leq C_{bg} + M\, n_{igt} \qquad \forall a, \forall g, RD_a \leq t \leq DL_a,\ AP_{ia} = 1,\ AT_{ba} = 1 \tag{7}$$

$$z_{at} \geq C_{bg} - M\,(1 - n_{igt}) \qquad \forall a, \forall g, RD_a \leq t \leq DL_a,\ AP_{ia} = 1,\ AT_{ba} = 1 \tag{8}$$

$$Q_{aw}\, x_{at} = r_{at} + o_{at} + s_{at} \qquad \forall\, a, w, t \mid Q_{aw} > 0 \tag{9}$$

$$\sum_t s_{at} = Q_{aw}\,(1 - c_a) \quad \forall\, a, w, t \mid Q_{aw} > 0 \tag{10}$$

$$\sum_{a/P_{ia}>0} c_a = (\sum_a AP_{ia})\, d_i \qquad \forall\, i \tag{11}$$

$$d_i + XS_i \geq 1 \qquad \forall\, i \tag{12}$$

$$RHE_w er_{wt} + OHE_w eo_{wt} \leq CAP_w \qquad \forall\, w, t \tag{13}$$

$$er_{wt} \leq ea_{wt} \qquad \forall\, w, t \mid t > FC \tag{15}$$

$$er_{wt} \leq NE_w \qquad \forall\, w, t \mid t \leq FC \tag{16}$$

$$eo_{wt} \leq ea_{wt} \qquad \forall\, w, t \mid t > FC \tag{17}$$

$$eo_{wt} \leq NE_w \qquad \forall\, w, t \mid t \leq FC \tag{18}$$

$$ea_{wt} = \sum_{l}^{t} eh_{ltw} \qquad \forall w, t \mid t > FC \tag{19}$$

$$eh_{ltw} = eh_{llw} - \sum_{m}^{t} ef_{lmw} \qquad \forall\, w, l, t \mid l \leq t \tag{20}$$

$$eh_{l\,t+1\,w} \leq eh_{ltw} \qquad \forall\, w, l, \ t < NP \tag{21}$$

$$eh_{l\,t+1\,w} \leq NE_w \qquad \forall\, w, l, t \mid l = FC + 1, t = FC + 1 \tag{22}$$

$$eh_{ltw} = eh_{llw} \qquad \forall\, w, l, t \mid l \leq t \leq l + ME \tag{23}$$

$$eh_{ltw} \leq eh_{llw} \qquad \forall\, w, l, t \mid t > l + ME \tag{24}$$

$$wr_{wt} + wo_{wt} = WH_{wt} \qquad \forall\, w, t \tag{25}$$

$$\sum_{a} r_{at} + wr_{wt} = RHE_w er_{wt} \qquad \forall\, w, t \tag{26}$$

$$\sum_{a} o_{at} + wo_{wt} = OHE_w eo_{wt} \qquad \forall\, w, t \tag{27}$$

$$\sum_{a} Q_{aw}\, x_{at} + \Gamma_{wt}\, \pi_{wt} + \sum_{a} p_{at} - \sum_{a} s_{at} \leq (RHE_{wt} + OHE_{wt})\, NE_{wt} - WH_{wt} \qquad \forall\, w, t \leq FC \tag{36}$$

$$\pi_{wt} + p_{at} \geq QD_{aw} x_{at} \qquad \forall w, t \leq FC \tag{37}$$

$$\pi_{wt} \geq 0 \qquad \forall w, t \leq FC \tag{38}$$

$$p_{at} \geq 0 \qquad t \leq FC,\ \forall\, a \mid Q_{aw} > 0 \tag{39}$$

$c_a \in \{0,1\} \quad \forall\, a$

$d_i \in \{0,1\} \quad \forall\, i$

$e_{at}, f_{at}, w_{at} \in \{0,1\} \quad \forall\, a, t$

$n_{igt} \in \{0,1\} \quad \forall\, i, g, t$

$ea_{wt} \in Z_+ \quad \forall\, w, t$

$ef_{ltw}, eh_{ltw} \in Z_+ \quad \forall\, l, t, w$

All other variables are nonnegative.