UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

COLLABORATIVE SYSTEMS IN LOGISTICS AND TRANSPORTATION

TESIS PARA OPTAR AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA

FRANCO FABIÁN BASSO SOTZ

PROFESOR GUÍA:
ANDRÉS WEINTRAUB POHORILLE

MIEMBROS DE LA COMISIÓN:
VLADIMIR MARIANOV KLUGE
FERNANDO ORDOÑEZ PIZARRO
MIKAEL RÖNNQVIST

Este trabajo ha sido parcialmente financiado por CONICYT

SANTIAGO DE CHILE
2018

ABSTRACT
AUTHOR: FRANCO FABIÁN BASSO SOTZ
DATE: 2018
ADVISOR: ANDRÉS WEINTRAUB POHORILLE

COLLABORATIVE SYSTEMS IN LOGISTICS AND TRANSPORTATION

During recent years, horizontal collaboration in logistics has gained attention because of achieved potential benefits such as cost reduction, an increase in fulfillment rates, and a decrease in CO2 emissions owing to reductions in traveled distances. Successful real-world cases, however, are rare since horizontal cooperation in logistics is not usually sustainable. The first paper of this thesis pays attention to this paradox of the lack of cases and discusses 16 identified practical issues that could explain this phenomenon. We propose a taxonomy composed of four categories categorizing the practical issues according to a value chain approach: design, planning and operations, market/business, and behaviors. Furthermore, we propose and discuss some measures to mitigate these problems.

The second paper provides another explanation for the lack of real application of horizontal collaboration in logistics. Until now, the operational research models used to study the horizontal collaboration in transportation have not included competition between firms. Contracts are signed, and both quantities and prices are fixed. Without competition, agreements always save on costs and it is then a matter of allocating costs savings wisely. In the second paper, we consider a coalition formation game but prior to market equilibrium; that is, we propose a collaborative model in which, after the agreements are signed, the different firms and coalitions compete in multiple markets in Cournot fashion. When this happens, the formation of one set of coalitions affects prices and production levels of all other competitors, something that did not occur in the previous literature. A possible partnership among these firms is allowed and studied. We propose multiple models to respond the question of which coalitions will be formed in this setting, including stability constraints and the restriction that the agreement should be cleared by antitrust authorities. Our main finding is that opposed to what has been found in the literature to date, forming coalitions that are beneficial to firms in the agreement and at the same time be susceptible to be cleared by antitrust authorities, is actually quite hard. This could help explain why collaboration has not been observed as much as expected.

Finally, in the third paper, we study a shared-information system in which users of an urban expressway share with a centralized system, data such as position, type of vehicle and speed. We develop accident prediction models for a stretch of the urban expressway Autopista Central in Santiago, Chile, using disaggregate data captured by free-flow toll gates with Automatic Vehicle Identification (AVI) which, besides their low failure rate, have the advantage of providing disaggregated data per type of vehicle. The process includes a random forest procedure to identify the strongest precursors of accidents, and the calibration/estimation of two classification models, namely, Support Vector Machine and Logistic regression. We find that, for this stretch of the highway, vehicle composition does not play a first-order role. Our best model accurately predicts 67.89% of the accidents with a low false positive rate of 20.94%. These results are among the best in the literature even though, and as opposed to previous efforts, (i) we do not use only one partition of the data set for calibration and validation but conduct 300 repetitions of randomly selected partitions; (ii) our models are validated on the original unbalanced data set (where accidents are quite rare events), rather than on artificially balanced data.

RESUMEN DE LA TESIS PARA OPTAR
AL GRADO DE DOCTOR EN SISTEMAS DE INGENIERÍA
POR: FRANCO FABIÁN BASSO SOTZ
FECHA: 2018
PROF. GUÍA: ANDRÉS WEINTRAUB POHORILLE

COLLABORATIVE SYSTEMS IN LOGISTICS AND TRANSPORTATION

Durante los últimos años, la colaboración horizontal en logística ha ido ganando atención en la literatura especializada debido a los beneficios reportados, como la reducción de costos, un aumento en las tasas de cumplimiento y una disminución en las emisiones de CO_2 debido a las reducciones en las distancias recorridas. Sin embargo, los casos exitosos en el mundo real son raros, ya que la cooperación horizontal en logística no suele ser sostenible. El primer *paper* de esta tesis presta atención a esta paradoja discutiendo 16 problemas prácticos los cuales podrían explicar este fenómeno. Proponemos una taxonomía compuesta por cuatro categorías que clasifican los problemas prácticos de acuerdo con un enfoque de cadena de valor, a saber: diseño, planificación y operaciones, mercado/negocio y comportamientos. Además, proponemos y discutimos algunas medidas para mitigar estos problemas prácticos.

El segundo *paper* proporciona otra explicación para la falta de aplicaciones reales de la colaboración horizontal en logística. Hasta ahora, los modelos de investigación operativa utilizados para estudiar la colaboración horizontal en el transporte, no han incluido competencia entre empresas. Los contratos están firmados, y tanto las cantidades como los precios son fijos. Consideramos un juego de formación de coaliciones, pero previo al equilibrio de mercado; es decir, proponemos un modelo de colaboración en el que, una vez firmados los acuerdos, las diferentes firmas y coaliciones compiten en múltiples mercados según un modelo Cournot. Cuando esto sucede, la formación de un conjunto de coaliciones afecta los precios y los niveles de producción de todos los demás competidores, algo que no ocurre en la literatura previa. Se permiten y estudian posibles asociaciones entre estas empresas. Proponemos múltiples modelos para responder la pregunta de qué coaliciones se formarían en este entorno, incluidas restricciones de estabilidad y la restricción de que el acuerdo debe ser aprobado por las autoridades antimonopolio. Nuestro principal hallazgo es que, a diferencia de lo que se ha encontrado en la literatura hasta la fecha, la conformación de coaliciones que son beneficiosas para las empresas, y al mismo tiempo, susceptibles de ser aprobadas por las autoridades antimonopolio, es bastante difícil.

Finalmente, en el tercer *paper*, estudiamos un sistema de información compartida en el que los usuarios de una autopista urbana comparten con un sistema centralizado, datos como la posición, el tipo de vehículo y la velocidad. Desarrollamos modelos de predicción de accidentes para un tramo de la autopista urbana Autopista Central en Santiago, Chile, utilizando datos desagregados capturados por peaje de flujo libre con identificación automática de vehículos (AVI) que, además de su bajo índice de fallas, tienen la ventaja de proporcionar información desagregada por tipo de vehículo. El proceso incluye un procedimiento *Random Forest* para identificar los mejores precursores de accidentes, y la calibración/estimación de dos modelos de clasificación, a saber, *Support Vector Machines* y regresión logística. Encontramos que, para este tramo de la carretera, la composición de los vehículos no juega un papel importante. Nuestro mejor modelo predice el 67,89 % de los accidentes con un índice de falsos positivos del 20,94 %. Estos resultados se encuentran entre los mejores en la literatura, aunque, a diferencia de esfuerzos previos, nuestros modelos están validados con datos reales en los cuales los accidentes son eventos raros.

To all the people that, in one way or another, made this thesis possible.

Acknowledgements

I would like to acknowledge my parents Myriam and Patricio, my lovely wife Francisca, and my *chanchos* friends for their support. I would like to express my deep gratitude to Professor Andrés Weintraub and Mikael Rönnqvist, my research advisors, for their patient guidance, enthusiastic encouragement and useful critiques of this research work. I wish also to acknowledge the help provided by Professor Leonardo Basso, Professor Mauricio Varas and Professor Raúl Pezoa. Finally, I would like to acknowledge Autopista Central for their support with the collection of the data used in the fourth Chapter of this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In this thesis, we will study collaborative systems in logistics and transportation. We understand a collaborative system as a predefined scheme in which multiple agents cooperate in order to obtain individual benefits. The information of each collaborative agent is collected by a centralized system which can be a third-party or a subset of the collaborative agents. This centralized system uses a clearly defined cooperative procedure to define in each period a cooperation plan. The plan interacts with the collaborative agents' through an interface. This scheme is summarized in Figure 1.1.

The objective of the thesis is to study collaborative systems in two different environments, namely, collaborative freight transportation (Chapter 2 and 3) and private urban transportation (Chapter 4). In all cases, we focus on analyzing the underlying reasons for the lack of real-world applications of this kind of systems despite the multiple benefits reported by the literature. Particularly, we include features often left apart in previous research such as practical issues in Chapter 2, competition in Chapter 3, and full data-base for calibration and validation purposes in Chapter 4.

The main hypothesis of this work is that through the application of collaborative systems in logistic and transportation is possible to improve the process of the collaborative agents'.

Methodologically, we use several techniques coming from Operations Research and Analytics such as Optimization, Game Theory, Random Forest, Support Vector Machine and Logistic Regression. All these methods are revised in the respective Chapters.

This thesis is composed of 5 Chapters. It includes an *Introduction* and *Final Comments* Chapters while the rest of the Chapters are published, submitted or draft manuscripts written in a consistent format. This has the advantage that each Chapter can be read on its own as a stand-alone unit. Yet, it is obtained at the expense of some linearity of the thesis. We now briefly explain the contribution of each Chapter.

The second Chapter studies the practical issues arising when horizontal collaboration in logistics is established. We conduct a structured literature review enhanced with expert knowledge. We found 62 articles in which at least one practical issue is discussed. Overall, we find 16 practical issues which are categorized using a proposed taxonomy which follows a supply chain perspective. The four categories proposed are design, planning and operation, business and market, and behavior.

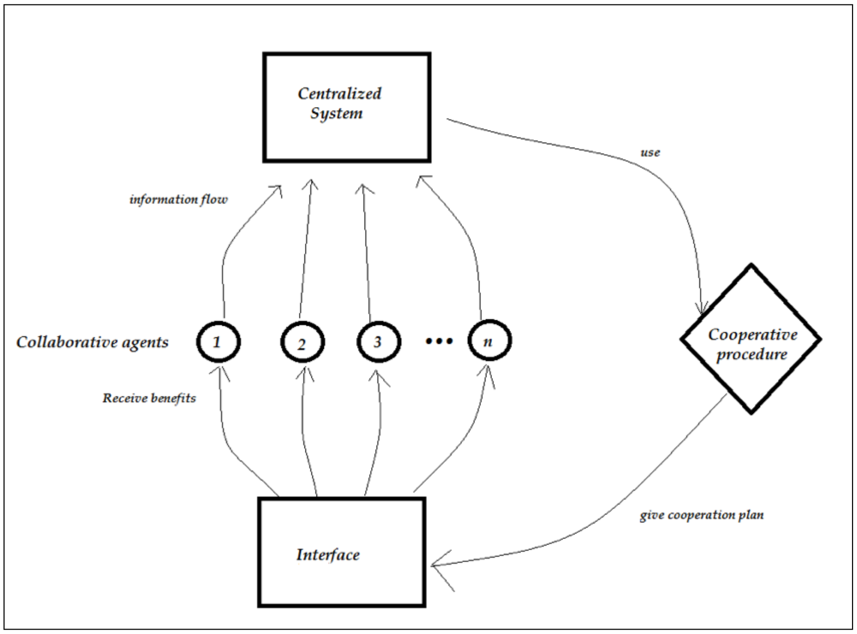


Figure 1.1: Cooperative System.

Our main conclusion is that theoretical models are not usually able to capture all the real-world difficulties which may explain why real applications have last shortly and far from the expected benefits. This Chapter highlights the main complications of such collaboration to encourage researchers to incorporate it into their models.

The third Chapter studies horizontal collaboration in freight transportation among competing firms. Even though, some successful cases have been reported in the forestry and energy industries, they have occurred over limited ranges of time, and far from the breadth that would be consistent with the alleged benefits that it should bring. In this Chapter we attempt to explain this discrepancy considering a coalition formation game prior to market equilibrium; that is, we propose a collaborative model in which, after the agreements are signed, the different firms and coalitions compete in multiple markets in Cournot fashion. When this happens, the formation of one set of coalitions affects prices and production levels of all other competitors, something that did not occur in the previous literature. Possible partnerships among these firms will be allowed and studied. Opposed to previous efforts, we found that collaboration is not always advantageous. Moreover, collaboration among all companies is usually outperformed by a partition of them into smaller sets, which is quite interesting as most related literature has assumed that the grand coalition forms.

The fourth Chapter explores a shared-information system in which users of an urban expressway share with a centralized system, data such as position, type of vehicle and speed. We build predictive models in order to determine situations in which the likelihood of a crash occurrence is higher than usual. These models include a Random Forest approach for determining the most important crash precursors and two classification models, namely, Support Vector Machine and Logistic Regression. Using this information, some preventive measure could be taken in order to decrease accidents rates benefiting the users of the expressway (collaborative agents). Opposed to previous efforts, our models are tested using full, real, online and disaggregated data taken from the Chilean private operated expressway Autopista Central. Our best model accurately predicts 67.89% of the accidents with a false positive rate of 20.94%.

Overall, this thesis seeks to expand the impact of the collaborative systems in logistics and transportation by improving both the success and extension of the models in real-world application.

Chapter 2

A survey on obstacles and difficulties of practical implementation of horizontal collaboration in logistics

Abstract: During recent years, horizontal collaboration in logistics has gained attention because of achieved potential benefits such as cost reduction, an increase in fulfillment rates, and a decrease in CO_2 emissions owing to reductions in traveled distances. Successful real-world cases, however, are rare since horizontal cooperation in logistics is not usually sustainable. This Chapter pays attention to this paradox of the lack of cases and discusses 16 identified practical issues that could explain this phenomenon. We propose a taxonomy composed of four categories categorizing the practical issues according to a value chain approach: design, planning and operations, market/business, and behaviors. Furthermore, we propose and discuss some measures to mitigate these problems.

Keywords: Collaborative logistics, horizontal collaboration, transportation, supply chain.

Difussion: A version of this Chapter has been submitted to *International Transactions in Operational Research* as: Basso, F., Rönnqvist, M., D'Amours, S., and Weintraub, A. (2018). A survey on obstacles and difficulties of practical implementation of horizontal collaboration in logistics.

2.1 Introduction

Horizontal collaboration in logistics has received increasing attention in past years, often driven by the large potential in cost reduction, reduction of uncertainty, and environmental concerns (Verdonck et al., 2013; Du et al., 2016). Many articles deal with new methods in sharing principles and joint planning in literature, though few articles report on successful implementation in practice. In this article, we focus on practical issues in horizontal collaboration. We define horizontal collaboration as cooperation between companies at the same level of a supply chain, e.g. carrier companies. Practical issues are meant to describe real-world problems stakeholders face when they try to implement such collaboration.

The main goal of collaborative logistics is to achieve an improved logistic chain, ensuring that the total fulfillment costs are smaller than the sum of the companies individual costs without collaboration. Examples of horizontal cooperation in logistics include group purchasing, use of a common inventory location to share fixed costs, collaborative transportation, and production lines sharing. Horizontal collaboration in logistics has been studied in the maritime shipping (Sheppard and Seidman, 2001), disaster relief (Schulz and Blecken, 2010; Ergun et al., 2014; Garrido et al., 2015), and airline fields (Oum et al., 2004; Garrette et al., 2009; Weng and Xu, 2014). However, collaborative ground transportation is quite an unexplored area.

In theory, many benefits can be achieved with collaboration, chief among them are cost reduction and increasing fulfillment rate. On the social side, collaboration usually decreases the traveled distance by carriers, which implies fewer emissions. In that way, collaboration encourages green logistics and reduces negative environmental impacts. These benefits have been shown in methodological and case study scientific papers, but few applications have been reported and captured expected benefits. For instance, in Audy et al. (2011), a collaborative transportation agreement is studied for the furniture industry. Even though theoretically important savings could be perceived by collaboration, the negotiation to establish how benefits were shared was impossible to carry out. One of the most important challenges in horizontal collaboration has been to agree on the sharing principles. Recently, Guajardo and Rönnqvist (2016) published an extensive survey in cost allocation methods in transportation collaboration. In Frisk et al. (2010), eight companies analyzed the potential to collaborate with Swedish forest transportation authorities to obtain an expected saving up to 14%. Nowadays, this agreement is no longer in operation. Suzuki and Lu (2017) state that "two real-world cases in which collaborations were attempted based on the idea similar to that of our concept, their results may not be used to assess the cost-saving potential of the concept, because in neither of these two cases, the idea was fully implemented".

This raises a need for a better understanding of the underlying reasons of this phenomenon. We would like to better comprehend why collaboration has so many advantages in theory, but is rarely successful in practice. We have conducted an extensive literature review enhanced with expert knowledge. We have identified what we have called practical issues in logistics collaboration. To the best of our knowledge, no specific papers deal with practical issues in horizontal logistics collaboration.

The literature is rich in vertical cooperation, but the environment is quite different. A recurrent example of vertical collaboration is found in the Collaborative Planning Forecasting and Replenishment (CPFR), where manufacturers and retailers share information and make common forecasting

Figure 2.1: The proposed taxonomy for the practical issues.

to improve demand visibility, thereby improving supply chain efficiency. We can anticipate that some practical issues in vertical collaboration apply also to horizontal collaboration. For a complete review of supply chain collaboration, including some practical issues, we refer the reader to Kanda et al. (2008).

Following a supply chain perspective, we propose a taxonomy for the practical issues with categories: design, planning and operations, business/market, and behaviors (see Figure 2.1). Design practical issues are described as challenges of building efficient, stable, sustainable, and fair collaborations. The planning and operational practical issues are related to the difficulties arising from the implementation of such collaboration. The business/market practical issues comprise the collaborative impact on the firms strategical level and explain how agreements impact the whole market. Finally, we understand the behaviors practical issues as the human relationships challenges.

This classification is organized from a macro to micro perspective with a supply chain point of view. We first focused on practical issues coming from the design process, i.e., issues caused by structuring the collaboration. At this stage, problems are linked to the coalition formation, the benefits sharing policy, and the establishment of a proper coordination mechanism. Cooperation must be implemented with a high-level management standard. Practical issues linked to planning and operations tackle the challenges of implementing the collaboration structure designed previously. At this point, the key concept of information sharing scares managers. Coordination mechanisms are usually based on Decision Support Systems and advanced management techniques. Indeed, the various operational firm cultures should be considered. In third term, we focused on practical issues coming from the difficulties of explaining this novel procedure inside (business) and outside (market) the company. Many fears could be experienced at this juncture. For example, one could be considered as a cartel or lose autonomy. Finally, we have defined the behaviors practical issues by assembling potential problems linked to human relationships. This is crucial in horizontal collaboration, for partners are usually also competitors, and building a trustful partnership is the main challenge to face.

Two main contributions are valued in this article. First, we initially survey practical issues behind implementing horizontal collaboration, providing other researchers and practitioners with an

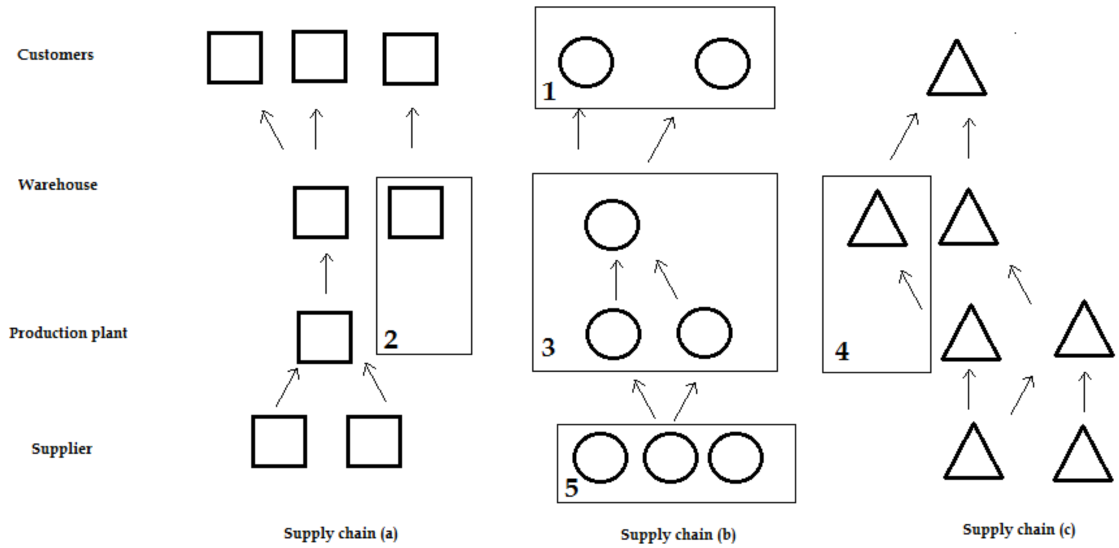


Figure 2.2: Dimensions of the collaboration, revised figure based on Audy et al. (2012b).

overview of examples where methods and processes are developed, but never applied. Second, we propose a categorization to facilitate research into similar issues and search for practical approaches for future use.

In Section 2.2, we briefly describe multiple types of collaboration in logistics. In Section 2.3, we define a taxonomy for practical issues. Finally, in Section 2.4, we conclude with some remarks and further research directions.

2.2 Collaboration in logistics

In this Section, we will analyze multiple types of collaboration. More specifically, we will explain differences between horizontal and vertical collaboration. Horizontal collaboration is defined as cooperation between competitor firms at the same supply chain level. Practical issues could vary dramatically depending on the type of collaboration established. This motivates a clear understanding and description of various logistics collaboration schemes.

To reduce cost, increase market conditions, reduce variability, increase fulfillment, respect operational and environmental constraints, and access new markets, companies establish collaborations with other enterprises. In contrast with vertical collaboration, where cooperation is done in the same supply chain, horizontal collaboration is established by companies at the same level in different supply chains. For example, forestry companies could share a unique terminal to split the fixed costs. Such a terminal often has large investments and requires a set of machines to function. In addition, it can easily handle volumes for several companies. Examples of such terminals abound: truck-train or truck-vessel or truck-train-vessel. Competitors often use this type of collaboration.

Figure 2.2 shows multiple forms of collaboration with graphics. One has a vertical collaboration when business units cooperate within the same supply chain. For example, delimited area 5 con-

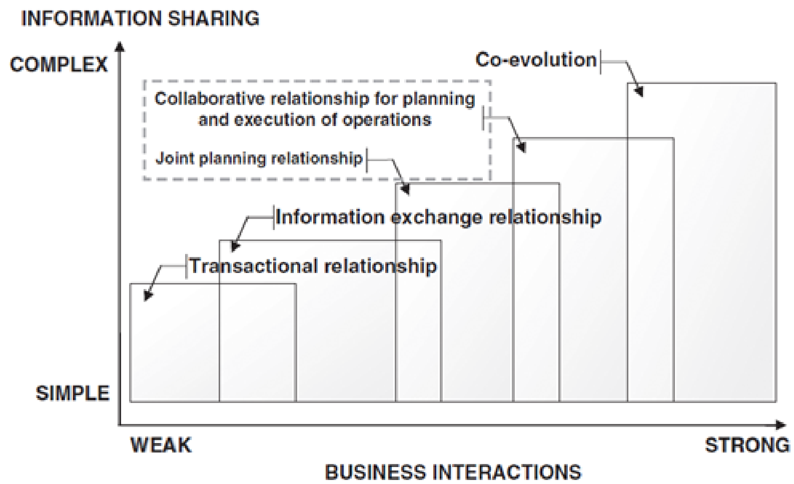


Figure 2.3: Forms of collaborative relationships (Frayret et al., 2003).

siders a horizontal collaboration with three suppliers. Delimited area 3 considers horizontal and vertical collaboration. Finally, we can consider the horizontal collaboration among business units within the delimited area 2 and 4. In this case, we can see that cooperation is not between competitors; collaboration is key for practical issues. Vertical collaboration has been studied extensively: we can mention cooperation to mitigate the bullwhip effect, where information is shared within the supply chain to reduce the variability of the demand. For further information about horizontal and vertical collaboration, we refer readers to Caputo and Mininno (1996), where these concepts are exemplified with the Italian grocery industry.

Collaboration can be achieved at many levels, starting from a simple interchange of information to a common strategic vision. The more complex the collaboration is, the more key practical issues will arise.

In Figure 2.3, the first stage called *transactional relationship* considers all the information exchanges including sometimes a limited amount of operational or tactical data. At this level, the information exchanged is the minimum necessary to accomplish the objective of the collaboration. In the second stage called *information exchange relationship*, partners exchange more important information such as production plans, demand forecasting models etc. In the third stage called *joint planning relationship*, partners share information but also partial planning, ideas, and objectives in order to make some joint decisions. The fourth level is called *collaborative relationship for planning and execution of operations*. This level involves joint implementation of operations and joint contingency management in a spirit of mutual aid. The last stage called *co-evolution* implies cooperation on the strategic level of the firms which implies a long-term relationship. The practical issues are highly depended on the level of cooperation.

Various frameworks have been proposed in collaborative logistics (Simatupang and Sridharan, 2005; Verstrepen et al., 2009; Leitner et al., 2011; Pomponi et al., 2013; Gonzalez-Feliu et al., 2013) usually involving the followings steps: activities planning, computation of the benefits, and the decision of how to distribute them. Some new mechanisms are presented in Audy et al. (2012b) including the presence of a third party which could help avoiding some practical issues.

Decision Support Systems (DSS) are fundamental to plan and execute the collaborative plans. In

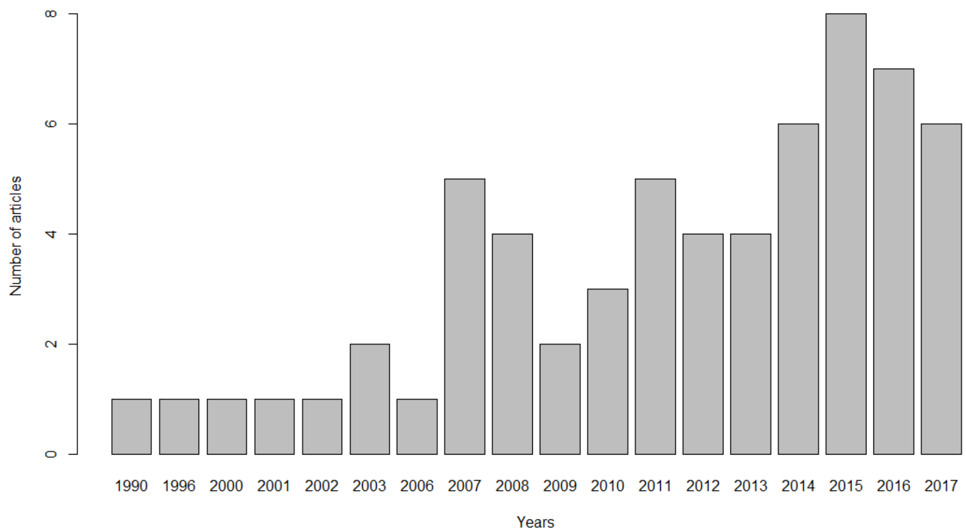


Figure 2.4: Number of articles per year.

this respect, Dahl and Derigs (2011) present a real time DSS for cooperative planning in carriers networks in Belgium. In the case presented by Frisk et al. (2010), a DSS called FlowOpt is used (Forsberg et al., 2005) to identify the potential of collaboration.

2.3 A Taxonomy for the practical issues

The articles reviewed in this survey have been obtained in three steps. In the first, we proceeded with a structured search on the Institute for Scientific Information website (2017). For a topic search, we used the word collaborative with *logistics*, *transportation*, or *inventory*. The second step involved a careful selection of each article. In fact, we aimed to identify articles that discussed practical issues. As we were conducting the review, our list was complemented with other articles we spontaneously noticed in the third step, either because they were frequently cited in the previously selected articles or we tracked references to them. Overall, we have identified 62 published articles published in journals (Table 2.1), especially in the past five years (Figure 2.4).

When companies work together as global economy players, a value chain-based approach is required to reflect these business models (Martinez-Olvera and Shunk, 2006). Consequently, we have classified the practical issues using a value chain-based approach. This classification divides the practical issues into four categories: design, planning and operations, business/market, and behaviors. The first outlines the partnership configuration and the strategies to keep it sustainable. The second is devoted to practical issues from developing and implementing this partnership, particularly the operation and planning process. The third is related to practical issues that threaten the core business of a company and its impact on the whole market. Practical issues belonging to this category are usually caused by a misunderstanding of what horizontal collaboration or cooperation between competitors really means. The fourth is about human behavior and its implications for the partnership. Each practical issue is supported by our expert knowledge or quotation of some of the

Journal	Number of articles
Transportation Research Part E: Logistics and Transportation Review	7
European Journal of Operational Research	5
Supply Chain Management: An International Journal	5
International Journal of Industrial Organization	3
OR Spectrum	3
International Transactions in Operational Research	2
Computers & Industrial Engineering	2
Expert System with Applications	2
International Journal of Production Research	2
Applied Soft Computing	2
Operations Research Letter	1
Transportmetrica A: Transport Science	1
Computers & Operations Research	1
Energy	1
European Journal of Industrial Engineering	1
Decision Support Systems	1
International Journal of Transport Economics	1
Manufacturing & Service Operations Management	1
International Journal of Production Economics	1
Journal of Operational Research Society	1
Strategy: Critical Perspectives on Business and Management	1
Journal of Physical Distribution & Logistics Management	1
Production and Operations Management	1
Production Planning and Control	1
Strategic Management Journal	1
International Journal of Procurement Management	1
Journal of Science and Technology	1
Transportation Science	1
The Review of Economics Studies	1
Research in Logistics and Production	1
Journal of Studies on Manufacturing	1
The Journal of Economic Perspectives	1
International Journal of Logistics Systems and Management	1
Journal of Business Logistics	1
Accounting, organizations and Society	1
Flexible Services and Manufacturing Journal	1
Omega	1
Conference Papers	2
Total	62

Table 2.1: Number of articles per journal.

62 reviewed articles.

2.3.1 Design

The design process is defined as the problem of collaboration stemming from the structure of such a partnership. How is it possible to find an agreement between partners to be engaged in a collaboration scheme? This is one of many fundamental questions in logistics collaboration. This Section we will shed light on some of them.

This topic is usually called ***coalition formation*** in the related literature. For Dahl and Derigs (2011) "on the strategical level the choice of the right set of partners yielding enough consolidation potential as well as mutual trust is a cardinal point". Cruijssen et al. (2007) said to this respect that "according to the respondents the most severe impediments for cooperation are the problems of finding a reliable party that can coordinate the cooperation in such a way that all participants are satisfied". Mathematically, we can identify the coalition formation to a well-known problem called the partitioning problem, where we must split a set into disjoint subsets (Aumann and Dreze, 1974). The collaboration literature usually tackles the coalition formation problem from a cost reduction point of view using OR/Game Theory tools (Lozano et al., 2013; Guajardo and Rönnqvist, 2015; Jouida et al., 2017; Xu et al., 2017). Game theory is used to describe how different partners would act depending on other people actions. Dao et al. (2014) study the Virtual Enterprise (VE), defined as a temporary alliance of enterprises that can share core competencies and resources together to better respond to business opportunities. The consultancy area usually uses this kind of procedure. According to the authors, "the working principle of VE indicates that one of the critical issues to establish a successful VE is to select the right partners, the so-called partner selection problem". Recently, Defryn et al. (2017) propose a multi-objective optimization model to integrate the partner objective in horizontal logistics collaboration. This framework distinguishes between the coalition objective and the player objectives. Additionally, the coalition formation problem does not usually include all the "transactional costs" (Williamson, 1981), estimated to be high, thereby hindering any positive upside of a collaboration.

Some recent research and practical experiences have shown that big coalitions are more likely to fail. Guajardo and Rönnqvist (2015) state that "adding a cost on the coalition cardinality might be useful for reflecting the issues of managing large coalitions. We define this as the ***coalition size*** issue. To mitigate this effect, new models limit the quantity of players per coalition or consider a cost per size in the objective function. Flisberg et al. (2015) states that "clearly, there are many companies involved in each of the integrated instances that make such collaboration more difficult". The right number of partners depends on the industrial context but is rarely higher than two or three, even though some exceptions apply. For example, Frisk et al. (2010) show a real case study of eight forest companies that initially agreed on a collaboration scheme obtaining savings up to 14.2%. Despite the large potential savings, such a collaboration was never implemented. However, three of the companies agreed to test the collaboration and it was successfully implemented for a short period. This experiment worked because a third-party partner was involved, having no interest in any of the companies. This partner represented a research and development organization; when experimenters were searching for a new independent partner, none could be found, putting an end to the three-company collaboration. Some companies worked together following the experiment, though savings were lower. But this collaboration continued because a pairwise collaboration was easier to coordinate (no financial flows are necessary when only volumes were exchanged).

Who will lead the collaboration? This can become a potential source of conflicts. We called this the ***leadership and conflict of interest*** issue. Depending on how the collaboration is structured, a subset of the players (leading companies) performs collaborative planning on behalf of the others. As we will see later, this may cause a conflict of interest, affecting the whole market in some cases. In this Section, we face the leadership issue from a design point of view. To avoid major problems, we need to think about and answer the following questions: How is the coalition led? Should there be a third party? What is the value of being a leading player? How do you trust the leader? How can one avoid the leader taking advantage of their position? Audy et al. (2012a) analyze the second question with a real case study in Sweden using four leadership scenarios. Depending on how agreements are made, the leading companies obtained a 10.6% additional payoff. For Verstrepen et al. (2009), the leading practical issue is the most important: "Finding a reliable party to lead the cooperation and constructing a fair allocation mechanism for the benefits are the impediments that respondents agree with most". This statement is based on a survey presented in the article by Cruijssen et al. (2007). Montoya-Torres et al. (2016) state that "the diverse and (very often) conflicting interests of stakeholders have to be taken into account".

Once the coalition is formed, how are costs and benefits shared? We call this the ***cost allocation*** practical issue. According to Kevin Lynch, CEO for Nistevo Network, "The key to understanding collaborative logistics lies in recognizing how costs are distributed in a logistics network". Academicians have published an significant number of articles on this topic (Özener and Ergun, 2008; Frisk et al., 2010; Audy et al., 2011; Dai and Chen, 2012; Lozano et al., 2013; Sun et al., 2015; Dai and Chen, 2015; Li et al., 2016). For a complete survey on cost allocation in transportation, please see Guajardo and Rönnqvist (2016). Most articles use game theory to propose a solution to this issue using famous methods such as the Shapley Value (Shapley, 1953) and the Nucleolus (Schmeidler, 1969). Moreover, game theory and profit allocation methods have been used in cooperative inventory problems (Triqui Sari and Hennet, 2016; Guajardo and Rönnqvist, 2015). The literature has focused a lot on the problem of sharing the benefits/costs but dividing failure risks when unexpected problems arise has been rarely studied.

The literature usually approaches horizontal collaboration from a static point of view. In other words, it does not matter what will happen when situations go awry, information is missing, or other situations occur. What could possibly go wrong? Some examples include a player leaving the coalition, companies not accomplishing their tasks (e.g., some deliveries are unfinished), data change, or incomplete information. Such examples are defined as ***dynamic*** practical issues. This collaboration is highly unexplored. In Regan and Song (2003), a coalition of mid-sized carriers fulfilling a full truckload pickup and delivery requests is studied. According to Leitner et al. (2011), collaboration "requires a huge effort of coordination and dynamic planning within the network". Hernández et al. (2011) study the deterministic dynamic single carrier collaboration problem fort in the small-to medium-sized less-than-truckload (LTL) industry. Wang and Kopfer (2013) investigate a full truckload pickup and delivery transportation problem with deterministic information and requests to be fulfilled. Wang and Kopfer (2015) study a rolling horizon planning for a dynamic collaborative routing problem with full-truckload pickup. This is an interesting topic for further research.

Another key practical issue to be considered from a structural point of view is the process of ***negotiation*** that must be performed to form the coalition. Firms engaging in alliances incur transaction costs in negotiating, monitoring, and enforcing the contract (Garrette et al., 2009). Bleeke and Ernst (2002) explain how developing the relative negotiation power of the coalition members is key in understanding if an alliance is likely to be successful. These authors consider three factors: partners initial strengths and weaknesses, change of these strengths and weaknesses, potential for competitive conflict. In the same way, Audy et al. (2012a) state that "we should expect that the player with higher negotiation power receives a larger payoff than its weaker counterpart". This idea suggests that the cost allocation must consider the firms negotiation power. In the Frisk et al. (2010) case study, "the results reported rely on the fact that all forest companies agree in advance to accept a cost allocation computed in any of the suggested approaches". In this case, the negotiation was successful. In the transportation planning context, on the one hand, companies with high negotiation power are usually centrally located. On the other hand, peripheral companies have less negotiation power. In the article "Constructive and blocking power in collaborative transportation" by Guajardo et al. (2016), they empirically focus on the negotiation power for companies using game theory models. They concluded that some cost allocation models, such as the modiclus or SM-nucleolus, could help to maintain the coalition sustainable once they consider negotiation power. Wang et al. (2017) propose a profit allocation method to improve negotiation power for logistics network optimization. The proposed approach, in opposition to most previous efforts, regards the willingness of logistics participants to participate.

2.3.2 Planning and operations

Planning and operations practical issues are understood as the challenges that arise from implementing the collaboration. We focus on the operational drawbacks that could appear.

From an operational perspective, it is vital to establish good connections for the ***information flow*** and establish proper ***coordination mechanisms***. In other words, proper and known means of communicating within the coalition need to be provided; they must be secured and trustful. Web-based software has become a promising tool to enhance communication among companies (Kale et al., 2007; Chow et al., 2007; Gonzalez-Feliu et al., 2013; Ilie-Zudor et al., 2015). Ergun et al. (2007) explain in their papers how the three companies "Nistevo, Elogex, and Transplace use the Internet as a common computing platform to give shippers and carriers visibility to hidden costs". Information must be trustful since errors in data could imply higher or lower revenues for the companies. This could be eventually used as a tool to improve archly incomes, meaning that data validation is recommended. Consequently, information compatibility must be a central point. Firms often have many ways to keep information: collaboration involves data homogenization; managing costs should be incorporated into company cost. Planning the collaboration typically requires large amounts of information, and disaggregated data could be necessary for planning calculation. According to Prakash and Deshmukh (2010), "Both the flexibility and collaboration requires high communication between various information systems at one hand and on the other hand the compatibility of their practices". The authors not only focused on information compatibility but also focused on practice compatibility. Stefansson and Russell (2008) define the place where the information and physical goods are exchanged as "interface". They believe that what works today may not work tomorrow. Hence, establishing a fruitful collaborative scheme with the "interface" must be constantly supervised to take fast actions if things are going wrong. For Stefansson and Russell (2008), the bottlenecks tend to be at the interfaces. Recently, Zhang et al. (2017) discussed a collaborative transportation model using an e-commerce platform (e.g., Tmall) for suppliers and retailers to share operational information with multiple logistics service providers (LSPs) forming a collaborative alliance in an e-commerce logistics network.

One of the main issues in implementing collaboration in logistics is the high level required by practitioners since companies must have a qualified human resources team. One idea to partially mitigate this issue is to outsource some technical decisions to a transparent expert or academician. This is called a ***practitioner knowledge*** practical issue. The expert should be respected by all the members and their expertise should be recognized in the industry and academia. For Cruijssen et al. (2010), "transportation is a hands-on and low-tech sector and practical cases have shown that practitioners often regard the problem of constructing a fair gain sharing mechanism as too difficult or academic".

Enterprises could have various accuracy rates, making the collaboration more difficult since companies with the best fulfillment may not be interested in collaborating with those in worse situations, even though a cost reduction could be obtained in theory. Moreover, operational standards must be similar for the collaboration to be successful. This is known as a ***fulfillment and standards*** practical issue. According to Caputo and Mininno (1996), in the collaborative inventory management, "branded industries have to adopt the same standard for bar-coding consumer units, cartons and pallets and they have to respect the chosen standards". To palliate this practical issue, the collaboration contract must stipulate what must be done if a company fails to live up to its commitments. For example, some payments should be made to repair the damages. Determining the amount of those payments is not an easy task, though. In the Wang and Kopfer (2014)article "Collaborative transportation planning of less-than-truckload freight" the authors state that "the calculation of the potential fulfillment costs for all bundles of less-than-truckload pickup and delivery requests with time windows constitutes a very difficult problem, which they do not consider". This practical issue is especially important for horizontal collaboration since all coalition members perform the same activities. In some areas (e.g., retailers), companies place such a value on service levels that it is difficult to envisage any collaboration (Hingley et al., 2011). Further, the partners may have different productivity levels.

Successful cases of horizontal collaboration usually use advanced Enterprise Resource Planning (ERP) or Decision Support System (DSS) for cooperation operations (Buijs and Wortmann, 2014). We call this the ***high-tech*** practical issue. We can mention as an example the transportation collaboration between eight Swedish companies (Frisk et al., 2010) that use the FlowOpt tool to coordinate the exchange of wood between members. For details about this DSS, we refer the reader to Forsberg et al. (2005). This is a practical issue particularly for small- and medium-sized companies that cannot afford expensive computational tools or consultancies. Additionally, these companies benefit most from horizontal collaboration in logistics. Wang and Kopfer (2014) said that "for small and medium-sized freight carriers, horizontal collaboration is considered as a promising support". Unlike large companies that can accomplish requests to a high extent by exploiting the economy of scale, small- and medium-sized carriers are in small surfaces. For Ergun et al. (2014), "Information technology (IT) tools can help facilitate collaboration, but cost and other barriers have limited their use". For example, "the visibility of the transport operation has been greatly enhanced in recent years by the introduction of telematics (..) It was created for three leading grocery manufacturers who collaboratively united to commission the system and is due to go live during the spring of 2006" (Mason et al., 2007). For Hernández and Peeta (2014), collaboration in logistics has grown rapidly in past years because of "the affordability and the increased use of the Internet and information and communication technologies". Montoya-Torres et al. (2016) state that "technological issues are of high relevance to ensure effective and efficient collaboration networks in urban goods transport".

2.3.3 Business/market

We define business practical issues as the challenges faced by the firms at the strategical level. As we mentioned in Section 2.2, collaboration could be established through various dimensions, from a transactional relationship or information exchange to a joint planning or strategic development. On the one hand, closer cooperation will have major implications on the company core business. On the other hand, the market practical issues are related to the interaction between the collaborators and the environment (government, competitors, consumers, etc.)

Collusion is a crucial topic in most countries because of recent cartel discoveries; however, it is not a new procedure. A well-known case study of collusion in first-price auctions is reported by Pesendorfer (2000), where the bidding for school milk contracts in Florida and Texas during the 1980s is studied. Collusion is a key practical topic in horizontal and vertical collaboration. The logistic collaboration must be carried out to in a way to ensure that it is not considered collusion. Achieving this task requires an in-depth study of anti-trust laws in each country. Many authors have studied the collusive implication of integration between two or more firms. Although this integration it is not focused on collaborative logistics, some of the first authors who studied horizontal integration are Shapiro and Willig (1990), who analyzed "the antitrust treatment of collaborative production activities among rival firms". Chen and Ross (2003) studied the particular collaboration in which parent companies produce a critical input that they use to produce final goods. They have in part concluded that "the effect on market efficiency will depend on the extent to which savings in fixed costs compensate for the reduced level of competition between the partners". Basso (2008) studied the effects on capacity and pricing in airports where no collusion regulation was introduced. For Frisk et al. (2010), horizontal cooperation is allowed if it does not interfere with the overall market prices in supply areas. In their case study, eight companies participated in a common transportation planning, but they still competed for buying harvest areas. Some additional precautions could be taken to avoid collusion risk. In the Frisk et al. (2010) case study, an additional practical constraint has been added: the flow balance between any pair of companies should be equal. This implies that there is no need to invoice between companies. Some countries are promoting horizontal collaboration in logistics, implementing legal frameworks to support it. For instance, in Europe, a program called Collaboration Concepts for CO-Modality (CO3) has recently been launched (Vanovermeire and Sörensen, 2014). This framework suggests the use of a trustful third party to coordinate the cooperation. CO3 also offers a legal background describing the entry and exit clauses for collaboration. This program promotes the Shapley Value (Shapley, 1953) to split the benefits among companies. Hezarkhani (2016) study pairwise collusion in bipartite matching games with an application in collaborative logistics.

Depending on the productive area, some useful data to establish cooperation could be considered sensitive information for the companies. These data are referred to as ***sensitive information sharing*** issues. Chow et al. (2007) propose a strategic knowledge-based planning system for the freight forwarding industry in Hong Kong. According to the authors, “most freight forwarders are reluctant to share their secret information such as customer transaction record to others”. Clifton et al. (2008) examine the use of cryptographic techniques to perform collaborative logistics among potential competitors’ carriers without a broker and with a strictly minimum share of information. To illustrate this approach, the authors focused on a problem faced by independent trucking companies with miscellaneous pick-up and delivery tasks. Cruijssen et al. (2010) defined the insinking phenomenon as a pull approach where the service provider initiates the shift of logistics activities selecting companies carefully. The authors think that this procedure is better than outsourcing since it facilitates the attainment of synergy without the difficulties arising from sharing sensitive information between the cooperating companies. According to Leitner et al. (2011), “collaboration entails much more than cooperation, especially in terms of sharing information, risks, knowledge and profit and in the required level of closeness”. Wang and Kopfer (2014) talked about collaboration in less-than-truckload freight. In this case, customer payments and cost structure information are unexposed in the Carrier Transport Provider. Chen (2016) state, “because carriers are generally autonomous units or even competitors, they do not want to reveal their confidential business”. The author proposes a combinatorial clock-proxy exchange to tackle this issue. Recently, Lai et al. (2017) stated that cooperation was “difficult to realize because each party is self-interested who may not share his private information that is necessary for the cooperation”. They proposed an iterative auction scheme that facilitates successful collaboration among carriers.

Some companies do not want to appear related to another one for reasons of ***reputation***, even though the collaboration could be beneficial for them. To avoid this, the collaboration could be done privately. Some companies are suspicious about collaborating with others since they think this kind of procedure could reduce their ***autonomy***. Some papers have confronted this issue. For example, Ghosh and Morita (2012) focused on how collaboration between competitors reduced their product distinctiveness. At this point, it is important to clearly restrict the coalition attributions so companies can decide whether or not to enter the partnership.

2.3.4 Behaviors

In this Subsection, we focus on the behavioral practical issues. In this context, we define the behaviors issues as problems arising from human factors and relations that may impact the collaboration.

The first practical issue that we are going to describe is ***trust***. As we have partially mentioned in other Sections, trust is essential for any collaboration. System, partners, and data are three areas where the trust could be broken. Trust in the system refers to confidence; there is hope that the collaborative partnership will be positive for the company. Here, a deep belief lies in procedures and methodologies of horizontal collaboration in logistics. Trust in the partners refers to the belief that coalition members will act ethically, responsibly, and diligently. Trust in the partners suggests that no stakeholder will take advantage of your company. Shapiro and Willig (1990) explain it like this: “the cooperation intrinsic to a production joint venture may foreclose opportunities that would otherwise arise for one venturer to expand profitably at the expense of another”. Trust in the data refers to players’ confidence in their data and in that of others. Tomkins

(2001) studied the behavioral practical issue when accounting information had to be shared with others. He believed that all business relationships depended on trust to some extent. He says emphatically that “the role that developing more reliable forms of trust and the cost of doing that is rarely considered in the cost-benefit analysis”. Simply put, there is a lack of behavioral analysis when the partnership contributions are studied. Lyons et al. (2006) focused on the preponderance that technologies have over socio-emotional cues. In the Pérez-Bernabeu et al. (2015) case study, the authors demonstrate that important savings in distances and greenhouse gas emissions could be made through collaboration in the vehicle routing problem. They think that “one of the many aspects to analyze in horizontal cooperation is trust between entities”. Meanwhile, Özener and Ergun (2008) emphasize that trust in horizontal collaboration is somehow difficult to build since collaborators are usually competitors. This is a particularly important remark since it is in contrast with the vertical cooperation (partners are not usually direct competitors). This is where trust is simpler to build. Cruijssen et al. (2007) talked about partners’ opportunistic behaviors. This fear of companies could be related to sensitive information that must be shared to perform the cooperation. A trust development evolutionary framework for horizontal collaboration in logistics is proposed by Pomponi et al. (2013), Pomponi et al. (2015), and Montoya-Torres et al. (2016). All authors state that “collaboration in urban logistics requires confidence, trust and information sharing between the actors involved in the process”.

We define ***cultural*** issues as the difficulties stemming from major differences between or within collaborators. This topic is huge in military and civilian environments. Lyons et al. (2006) explain it like this: “the military is facing novel demands in terms of peacekeeping, humanitarian, and disaster relief operations, which require instant and effective logistic collaboration between the local, state, and federal governments; military; and civilian organizations. This collaboration may also traverse cultural and geographic boundaries, which adds another degree of complexity to logistics teams”. Military organizations have a vertical leadership that could collide with civilians ones (more horizontal). Some companies have informal procedures to communicate to others, which could be insurmountable for more organized companies. Ghaderi et al. (2012) explored the potential and impediment for group purchasing collaboration. They stated that “the success factors related to inter-organizational trust, the formality of the group and the uniformity of the group members? are important practical issues to care about.

To perform a good collaboration, ***incentives*** must be offered intelligently. With a classical fulfillment rate scope, managers could be motivated to accomplish only their company task, leaving aside partner requirements. For Hingley et al. (2011), “this resistance may reflect a lack of ability, interest or determination or could simply imply that powerful retail manager gatekeepers block its implementation due to their own self-interest”.

2.3.5 Summary

We have identified and proposed a taxonomy to categorize 16 practical issues that limit the implementation of collaborative logistics (see Table 2.2). Some articles cover several issues. Clearly, problems surface and are described in various application areas and a variety of scientific articles.

Table 2.2: Summary of the practical issues.

Category	Practical Issue	Papers
Design	Coalition formation	Cruijssen et al. (2007) ,Dahl and Derigs (2011) Lozano et al. (2013) , Dao et al. (2014) Guajardo and Rönnqvist (2015),Triqui Sari and Hennet (2016) Jouida et al. (2017) , Xu et al. (2017) Defryn et al. (2017)
	Coalition size	Guajardo and Rönnqvist (2015), Flisberg et al. (2015) Frisk et al. (2010)
	Leadership and conflicts of interest	Cruijssen et al. (2007), Verstrepen et al. (2009) Audy et al. (2012a), Montoya-Torres et al. (2016)
	Cost allocation	Özener and Ergun (2008), Dai and Chen (2012) Audy et al. (2011), Lozano et al. (2013) Sun et al. (2015), Dai and Chen (2015) Guajardo and Rönnqvist (2016) , Li et al. (2016) Triqui Sari and Hennet (2016)
	Dynamic aspects	Regan and Song (2003),Hernández et al. (2011) Leitner et al. (2011), Wang and Kopfer (2013) Wang and Kopfer (2015)
	Negotiation	Bleeke and Ernst (2002), Frisk et al. (2010) Audy et al. (2012a), Guajardo et al. (2016) Wang et al. (2017)
Planning and Operations	Information flow and coordination mechanism	Chow et al. (2007) , Ergun et al. (2007) Kale et al. (2007) ,Stefansson and Russell (2008) Prakash and Deshmukh (2010) Gonzalez-Feliu et al. (2013), Ilie-Zudor et al. (2015), Zhang et al. (2017)
	Practitioner knowledge	Cruijssen et al. (2010)
	Fulfillment and Standards	Hingley et al. (2011), Wang and Kopfer (2014)
	High-tech	Mason et al. (2007) , Frisk et al. (2010) Wang and Kopfer (2014), Hernández and Peeta (2014) Ergun et al. (2014) , Buijs and Wortmann (2014) Montoya-Torres et al. (2016)
Business/Market	Collusion	Shapiro and Willig (1990), Pesendorfer (2000) Chen and Ross (2003), Basso (2008) Frisk et al. (2010), Vanovermeire and Sörensen (2014) Hezarkhani (2016)
	Sensitive information sharing	Clifton et al. (2008), Cruijssen et al. (2010) Leitner et al. (2011), Wang and Kopfer (2014) Chen (2016), Montoya-Torres et al. (2016) Lai et al. (2017)
	Reputation and autonomy	Ghosh and Morita (2012)
Behaviors	Trust	Shapiro and Willig (1990), Tomkins (2001) Lyons et al. (2006), Cruijssen et al. (2007) Özener and Ergun (2008), Pomponi et al. (2013) Pérez-Bernabeu et al. (2015),Pomponi et al. (2015) Montoya-Torres et al. (2016)
	Cultural	Lyons et al. (2006), Ghaderi et al. (2012)
	Incentives	Hingley et al. (2011)

2.4 Concluding remarks

In theory, many benefits could be achieved when implementing horizontal collaboration in logistics. In the last decade, literature in this area has grown rapidly, and many models, methods, and approaches have been suggested and tested in case studies and examples (Palander and Väätäinen, 2005). Real implementations are rare, however. As shown and described, many reasons abound that explain how collaboration is not implemented. Understanding these problems in depth is crucial to alleviate their consequences.

We have proposed a taxonomy for four main categories involving 16 identified issues. These categories are Design, Planning and Operations, Business and Market, and Behavior. They correspond to various parts of the value chain and time horizons. In the articles found, it has been shown that theoretical models are not usually able to capture all the real-world difficulties. This work highlights the main complications of such collaboration to encourage researchers to incorporate it into their models.

From a theoretical point of view, we have found that the cost allocation problem is the most studied practical issue. Researchers have solved it by using Game Theory, OR, and other ad-hoc techniques to split costs fairly. Despite the above techniques, trust and the coordination mechanism are huge obstacles to make the collaboration happen from a real-world point of view. The information must flow in defined and secured channels using sophisticated approaches to react when something unexpected happens. Anti-trust laws are another critical matter, as they directly relate to strict market regulations. Hence, it is important to study anti-trust laws in each country to avoid being taken for a cartel. The agreement must be made in a way that companies do not lose autonomy and maintain their enterprise philosophy.

Chapter 3

Coalition formation in collaborative transportation with competing firms

Abstract: Because of its importance in logistics costs, transportation is usually seen as the primal activity for which horizontal collaboration could have a huge impact. Indeed, some successful cases have been reported in the forestry and energy area, but have occurred over limited ranges of time, and far from the breadth that would be consistent with the alleged benefits that it should bring. We attempt to explain this discrepancy pursuing the following rationale. Until now, the operational research models used to study the horizontal collaboration in transportation have not included competition between firms. Contracts are signed, and both quantities and prices are fixed. Without competition, agreements always save on costs and it is then a matter of allocating costs savings wisely. In this Chapter, we consider a coalition formation game but prior to market equilibrium; that is, we propose a collaborative model in which, after the agreements are signed, the different firms and coalitions compete in multiple markets in Cournot fashion. When this happens, the formation of one set of coalitions affects prices and production levels of all other competitors, something that did not occur in the previous literature. A possible partnership among these firms is allowed and studied. We propose multiple models to respond the question of which coalitions will be formed in this setting, including stability constraints and the restriction that the agreement should be cleared by antitrust authorities. Our main finding is that opposed to what has been found in the literature to date, forming coalitions that are beneficial to firms in the agreement and at the same time be susceptible to be cleared by antitrust authorities, is actually quite hard. This could help explaining why collaboration has not been observed as much as expected.
Keywords: Colaborative transportation, Game Theory, Cournot Model.

Difussion: A version of this paper will be submitted to *Operations Research* as: Basso, F., Basso, L. J. Rönnqvist, M., Weintraub, A, Coalition formation in collaborative transportation with competing firms

3.1 Introduction

Collaborative logistics have received increasing attention in the last years driven mainly by cost, service and environmental concerns. Because of its importance in the total operational costs, transportation has been seen as the primal activity to collaborate. In the literature, three types of collaboration in transportation have been analyzed. The first one corresponds to the vertical collaboration in which stakeholders located in different levels of the supply chain coordinate one or more transportation activities. The rationale for this collaborations is twofold; on one hand, in a vertical structure setting, there is an obvious double marginalization problem which is exacerbated if costs are not minimized, which in this case requires coordination. A second rational are technological constraints. An example of this case is the schedule coordination between the transportation carriers and the grape reception depot in the wine industry where due to oenological constraints, once the grape is harvested, it must be delivered as soon as possible to the grape reception depot in order to keep its quality (Ferrer et al., 2008). The second type of collaborative transportation happens at the horizontal level and may be classified into two types. In the first type, companies that provide transport services in networks that do not overlap much agree to make their networks seamless; agreements range from selling two tickets as just one and taking internal care of luggage to joint pricing. This alliances generate increased revenues for the companies and decreased prices for consumers because horizontal double marginalization is avoided. There is now a large body of literature on the economics and operation of airline international alliances and codesharing agreements (see for example Brueckner (2001) for a seminal article). The prevailing view has been that, as long as the networks do not overlap much, which implies that carriers offer complement rather than substitutes products, these agreements are desirable and have been, therefore usually accepted by antitrust authorities. The second type of horizontal collaboration in transportation, and the one we focus in this paper, is one in which productive firms (i.e., firms that use transport as an input and do not have transport as their main product) agree to share transportation resources in order to decrease costs through minimization of the total traveled distances.

Horizontal transport collaboration between productive firms has been long argued in the literature as desirable and a number of authors have analyzed how cost savings may be apportioned in order to ensure that firms have the incentives to reach the agreement (Frisk et al., 2010; Audy et al., 2012b; Flisberg et al., 2015). Yet, in reality, agreements are rarely seen. In Chapter 2 of this thesis, we offered an explanation on why this may occur; some of the explanations are the lack of mutual trust and proper coordination mechanism. Here we offer another possible explanation: that competition in the output market makes this arrangement either unprofitable for firms or anti-competitive and therefore would raise suspicions of the antitrust authorities or that the calculations needed to compute reasonable coalitions are close to impossible.

Indeed, in all previous papers, the amount of output that firms need to move is fixed: contracts between firms and customers are signed, and the prices agreed upon, implying that the revenues for each firm are constant. Firms, even though producing similar or identical outputs are not competing at the moment of negotiating agreements. Minimizing total transport costs, then, evidently raise profits and the problem that remains is how to allocate cost savings across partners so that everyone in the coalition is better off. Consider for example the model of Frisk et al. (2010) in which they analyzed the benefits of collaboration among N firms based on real-data from Swedish forestry companies. Using cooperative game theory, they analyzed possible partnerships solving multiple

Hitchcock linear problems (Figure 3.1) for all the 2^N possible subsets of the firms, which in Game Theory vocabulary are called coalitions. Using this information, it is possible to compute the characteristic function of the game which provides the cost of each coalition. As it is stated by the authors, the characteristic function satisfies the sub-additivity property which implies that the best option is always the grand coalition (all the firms belong to the same coalition). Then it is a matter of split the cost among the members wisely.

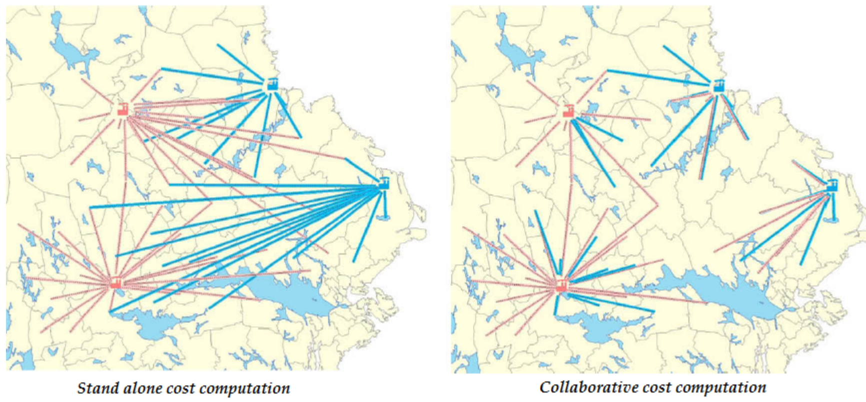


Figure 3.1: Cost computation for the classical transportation problem (source: Frisk et al. (2010)).

Several applications of the model of Frisk et al. (2010) have been reported in the literature. For instance, Audy et al. (2012b) analyzed the benefits of collaborative transportation agreement in the furniture industry, while Flisberg et al. (2015) studied cost reduction in the forest fuel transportation. Yet, all these real-world agreements have last shortly and far from the expected benefits it should bring.

The assumption of fixed outputs could be realistic in the short run if the agreement only covers one period of production. If the agreement, however, would still hold for the next production cycles, then firms would be in a totally different setting, as coalitions would need to decide not only how to move their (collective) output, but also in which markets and at which price to sell. What happens, then, is that the coalition structure affect prices and therefore revenues of all firms, in all coalitions, making that the utility of a coalition is no longer independent of how the rest of the firms organize themselves. This, intuitively and as shown below, destroys subadditivity. Then, which coalition structure ends up being stable, in at least the profitability sense, is far from obvious and, furthermore, how to efficiently calculate the stable coalition structure also is.

To make the point, we resort to the horizontal mergers literature, whose parallel with collaboration agreements is evident (in fact, in this paper, collaboration and merger will be identical). Consider N firms competing in symmetric Cournot fashion, all with identical and constant marginal costs. Salant et al. (1983) show that if $M < N$ firms merge, then that merger agreement is unprofitable for the merging firms unless the merging coalition reaches over 80% market share (see Appendix A for an illustrative example). Thus, the only agreement that would be feasible involves firms reaching very high market shares, close to a monopoly, which, if prices are to be decided, would be immediately challenged by antitrust authorities. If firms have different marginal costs, the overall picture remains unchanged. Yet, importantly, the 80% threshold for a profitable merger paralleled here by collaboration agreements decreases if the merging-collaborating firms achieve a marginal cost that is strictly smaller than the minimum of all marginal costs. This usually receives the name of "true cost synergies" (Farrell and Shapiro, 1990) and is also a condition for the price not to raise.

In our case, the spatial structure of firms, plants, and markets enable true cost synergies through collaborative transportation (Figure 3.1)

In this paper, we introduce Cournot competition in multiple spatially differentiated markets. The competition takes place after the collaboration agreements are signed, thus taking a longer run view. Firms produce a single indistinguishable product with equal unit production cost at multiple capacitated plants located in different points in space, thus making transportation costs to the various markets different. The questions we attempt to answer are: which coalitions may form? Here the stability notion becomes very important, because one may think of a number of conditions that need to be fulfilled. For example, the agreement needs to be profitable for firms in the coalition, but whether this is individually or collectively depends on whether lump-sum transfers are possible. On the other hand, agreements between competitors will raise attention from antitrust authorities and therefore, a condition for stability may be that consumers, or social welfare, are not hurt. This will be the case if the pressure for higher prices caused by decreased competition is smaller than the pressure for smaller prices caused by decreased marginal costs [1]. A second question is how to actually compute a stable coalition structure. Since the value or utility of a coalition is no longer independent on how the rest of the firms are organized, subadditivity properties can no longer be used. We study different avenues to calculate this, discussing efficiency and feasibility of the task. Our main results show that, first, a coalition structure that fulfills the different stability condition exists, showing that true cost synergies play a role, but the structure is with less concentration than when revenues are fixed. We also show that the calculation of such stable coalition structure is not simple, and there are no much efficiency gains in using heuristics or methods such as branch-and-cut.

The rest of the paper is organized as follows. In Section 3.2 we describe the proposed approach; specifically, we define the main assumptions of the model including different notions for stability, describe the methodology to compute the utilities of each coalition for each coalition structure and propose two models to determine the coalitions more likely to form. In Section 3.3, we conduct some numerical experiments through an illustrative example, and analyze a case-study based on real data from the Swedish forestry industry, providing two heuristics for solving strategies. Section 3.4 concludes.

3.2 The proposed approach

3.2.1 Problem definition and assumptions

We will study the problem in which multiple firms, after the collaboration agreements are signed, compete in quantities in multiple markets under a Cournot scheme with a linear demand function. We assume each firm has multiple plants and the transportation cost are different for each pair plant-

[1] Some countries will allow a merger or a collaboration agreement only if consumers are not harmed. This is the case of Australia, the European Union, Indonesia, Mexico, the Netherlands, Norway, Portugal, the United Kingdom and the US. Some other countries have a more flexible stance and allow price increases as long as social welfare is not hurt, that is, that the increase in profits at the industry level is larger than the reduction of surplus for consumers, thus leaving space for a Pareto improving redistribution. This is the case of Canada, South Africa, Singapore, New Zealand, Korea and Canada

market. When collaboration occurs, the companies share their plants and decide their production levels in a centralized way. The markets are independent of each other. Let define K the set of the firms, I_k the set of the plants of the firm $k \in K$, J the set of the markets (Figure 3.2)

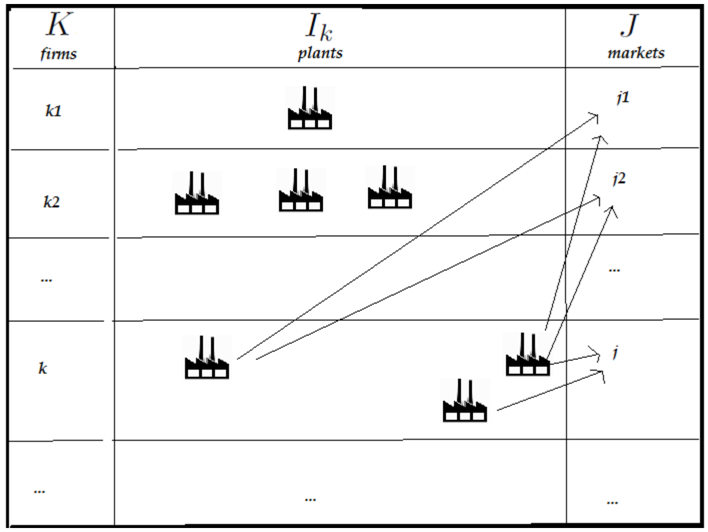


Figure 3.2: Collaborative model under competition.

We will denote $\mathcal{P}$ the set of all the partitions of K and $L = \{1, ..., |\mathcal{P}|\}$ so $\mathcal{P} = (\mathcal{P}_l)_{l \in L}$. Let $\mathcal{P}_l \in \mathcal{P}$ be a given partition of K. Each partition $\mathcal{P}_l$ represents a possible coalitional structure. In this context, each member of $\mathcal{P}_l$ is called a coalition following a game theory vocabulary. The number of partition of a set is given by the Bell numbers which increase very fast. For instance, if $|K| = 10$ then the number of partitions are 115,975 (Bell, 1934; Rota, 1964). Given a coalition $S \in \mathcal{P}_l$, we define the utility $\pi^*_{S,\mathcal{P}_l}$ as the utility perceived by the coalition S if the coalition structure $\mathcal{P}_l$ is attained. As we will see in the next Subsection $\pi^*_{S,\mathcal{P}_l}$ are computed by solving in parallel multiple best answer optimization problems.

3.2.2 Computing the utilities of the coalitions

We will now explicitly define the procedure to compute the utility of each coalition within each partition. To do so, let us consider the following best response optimization model $(O_{S,\mathcal{P}_l})$ $\forall l \in L, S \in \mathcal{P}_l$

Parameters

$a, b \in \mathbb{R}_+$	:	Parameters for demand function.
$c_{kij} \in \mathbb{R}_+$	:	Transportation cost from plant i of the firm k to market j.
$CAP_{ki} \in \mathbb{R}_+$	:	Capacity of the plant i from company k.
$c \in \mathbb{R}_+$	:	Unit production cost.

Variables

$q_{kij} \in \mathbb{R}_+$	:	Amount sent by the firm k from plant i to market j.
$q_{kj} \in \mathbb{R}_+$	:	Amount sent by the firm k to market j.
$Q_j \in \mathbb{R}_+$	:	Amount sent by all the firms to market j.

$$(O_{S,\mathcal{P}_l}) \quad \text{maximize} \quad \pi_{S,\mathcal{P}_l} = \sum_{j\in J}(a - bQ_j)(\sum_{s\in S} q_{sj}) - \sum_{j\in J}\sum_{s\in S}\sum_{i\in I_s}(c + c_{sij})q_{sij}$$

$$q_{sj} = \sum_{i\in I_s} q_{sij} \quad \forall s \in S, j \in J \tag{3.1}$$

$$Q_j = \sum_{k\in K} q_{kj} \quad \forall j \in J \tag{3.2}$$

$$\sum_{j\in J} q_{sij} \leq CAP_{si} \quad \forall s \in S, i \in I_s \tag{3.3}$$

$$q_{sij}, q_{sj}, Q_j \geq 0 \quad \forall s \in S, i \in I_s, j \in J \tag{3.4}$$

The optimization problem seeks to maximize for each coalition the sum of profits of each market $(a-bQ_j)(\sum_{s\in S} q_{sj})$ minus the sum of each market total production-transportation costs $\sum_{s\in S}\sum_{i\in I_s}(c+c_{sij})q_{sij}$. Equations 3.1 and 3.2 define the variables q_{sj} and Q_j. Equation 3.3 imposes a capacity constraint for each supply point. Equation 3.4 establish positivity of the variables. Note that the only structural decision variables are q_{sij}. Then, we can replace the equations 3.1 and 3.2 in the objective function and we obtain:

$$(O_{S,\mathcal{P}_l}) \quad \text{maximize} \quad \pi_{S,\mathcal{P}_l} = \sum_{j\in J}(a - b(\sum_{k\in K}\sum_{i\in I_k} q_{kij}))(\sum_{s\in S}\sum_{i\in I_s} q_{sij}) - \sum_{j\in J}\sum_{s\in S}\sum_{i\in I_s}(c + c_{sij})q_{sij}$$

$$\sum_{j\in J} q_{sij} \leq CAP_{si} \quad \forall s \in S, i \in I_s \tag{3.5}$$

$$q_{sij} \geq 0 \quad \forall s \in S, i \in I_s, j \in J \tag{3.6}$$

Which is a maximization concave quadratic optimization problem with linear constraints. Note that $\pi_{S,\mathcal{P}_l}$ depends on the decision variables $q_{kj} \;\; \forall k \in K \setminus S, j \in J$ through equation 3.2. This strategic interaction is absent in the classical collaborative transportation. That is:

$$\pi_{S,\mathcal{P}_l} = \pi_{S,\mathcal{P}_l}(q_{kij} \;\; \forall k \in K \setminus S, i \in I_k, j \in J) \tag{3.7}$$

The Lagrangian function of the optimization problem is:

$$\mathcal{L}_{S,\mathcal{P}_l}(q,\mu,\lambda) = -\sum_{j\in J}(a-b(\sum_{k\in K}\sum_{i\in I_k}q_{kij}))(\sum_{s\in S}\sum_{i\in I_s}q_{sij}) + \sum_{j\in J}\sum_{s\in S}\sum_{i\in I_s}(c+c_{sij})q_{sij} + \sum_{j\in J}\sum_{s\in S}\sum_{i\in I_s}\mu_{sij}q_{sij} + \sum_{j\in J}\sum_{s\in S}\sum_{i\in I_s}\lambda_{si}(q_{sij}-CAP_{si}) \tag{3.8}$$

$\forall \mathcal{P}_l \in \mathcal{P}, S \in \mathcal{P}_l, s^* \in S, i^* \in I_{s^*}, j^* \in J$ the KKT conditions are:

$$\frac{\partial \mathcal{L}_{S,\mathcal{P}}}{\partial q_{s^*i^*j^*}} = b\sum_{s\in S}\sum_{i\in I_s}q_{sij^*} - (a - b\sum_{k\in K}\sum_{i\in I_k}q_{kij^*}) + c + c_{s^*i^*j^*} + \mu_{s^*i^*j^*} + \lambda_{s^*i^*} = 0 \tag{3.9}$$

$$\sum_{j\in J} q_{s^*i^*j} \leq CAP_{s^*i^*} \tag{3.10}$$

$$\lambda_{s^*i^*}(\sum_{j\in J}q_{s^*i^*j} - CAP_{s^*i^*}) = 0 \tag{3.11}$$

$$\mu_{s^*i^*j^*}q_{s^*i^*j^*} = 0 \tag{3.12}$$

$$\lambda_{s^*i^*}, \mu_{s^*i^*j^*}, q_{s^*i^*j^*} \geq 0 \tag{3.13}$$

As the objective function of each maximization problem $(O_{S,\mathcal{P}_l})$ is concave these KKT constraint are necessary and sufficient to compute the global optimum in terms of the behavior of the non-member of the coalition S. The optimum solution of each $(O_{S,\mathcal{P}_l})$ represents implicitly a best-response function. Solving this fixed point problem we obtain the Cournot-Nash equilibrium conditional on $\mathcal{P}_l$. That is, it is possible to compute the equilibrium values $q^*_{kij} \quad \forall k \in K, i \in I_k, j \in J$. Moreover, by replacing them into the equation 3.7 we can compute numerically the utility values $\pi^*_{S,\mathcal{P}_l} \quad \forall S \in \mathcal{P}_l$.

Theorem 3.1 If $\forall k \in K, i \in I_k \quad CAP_{ki} > |J| \cdot \dfrac{a}{b}$ (no capacity constraint) then:

$\forall l \in L, S \in \mathcal{P}_l, s^* \in S, i^* \in I_{s^*}, j^* \in J$:

(a) if $q_{s^*i^*j^*} > 0$ then:

$$Q^S_{j^*} = \frac{1}{2}\left(\frac{a-c-c_{s^*i^*j^*}}{b} - Q^{-S}_{j^*}\right) \tag{3.14}$$

with $Q^S_{j^*}$ the total amount sent by the coalition $S \in \mathcal{P}_l$ to the market $j^* \in J$ and $Q^{-S}_{j^*}$ the total amount sent by all the firms which are not in S.

(b) Define $c_j(S) = \min\limits_{k'\in S, i'\in I_{k'}} c_{k'i'j} \quad \forall S \subseteq K, j \in J$ as the shortest path from coalition S to market J. If $c_{s^*i^*j^*} > c^*_j(S)$ then $q_{s^*i^*j^*} = 0$

Proof: For $k \in K, j \in J$ since prices are non negatives:

$$\sum_{i \in I_k} q_{kij} \leq \frac{a}{b}$$

adding in j:

$$\sum_{j \in J} \sum_{i \in I_k} q_{kij} \leq |J| \cdot \frac{a}{b} < CAP_{ki}$$

That means

$$\sum_{j \in J} q_{kij} < CAP_{ki} \quad \forall k \in K, i \in I_k$$

Using equation (3.11) in the equilibium we have

$$\forall l \in L, S \in \mathcal{P}_l, s^* \in S, i^* \in I_{s^*} \;\; \lambda_{s^* i^*} = 0$$

(a)
$$\frac{\partial \mathcal{L}_{S,\mathcal{P}}}{\partial q_{s^* i^* j^*}} = b \sum_{s \in S} \sum_{i \in I_s} q_{sij^*} - (a - b \sum_{k \in K} \sum_{i \in I_k} q_{kij^*}) + c + c_{s^* i^* j^*} + \mu_{s^* i^* j^*} = 0 \quad (3.15)$$

$$(3.16)$$

That it is:

$$a - b(Q^S_{j^*} + Q_{j^*}) = c + c_{s^* i^* j^*} + \mu_{s^* i^* j^*} \quad (3.17)$$

$$Q^S_{j^*} + Q_{j^*} = \frac{a - c - c_{s^* i^* j^*} - \mu_{s^* i^* j^*}}{b} \quad (3.18)$$

$$Q^S_{j^*} = \frac{1}{2} \left(\frac{a - c - c_{s^* i^* j^*} - \mu_{s^* i^* j^*}}{b} - Q^{-S}_{j^*} \right) \quad (3.19)$$

By hiphotesis $q_{s^* i^* j^*} > 0$ using equation (3.12) we have $\mu_{s^* i^* j^*} = 0$

$$Q^S_{j^*} = \frac{1}{2} \left(\frac{a - c - c_{s^* i^* j^*}}{b} - Q^{-S}_{j^*} \right) \quad (3.20)$$

(b) By contradiction given $l \in L, S \in \mathcal{P}_l$ suppose that it exists $s^0, s^1 \in S, i^0 \in I_{s^0}, i^1 \in I_{s^1}$ and $j_0 \in J$ such $c_{s^0 i^0 j^0} > c_{s^1 i^1 j^0}$ and $q_{s^0 i^0 j^0} > 0$. This last implies $\mu_{s^0 i^0 j^0} = 0$. By the equation (3.18):

$$Q^S_{j^0} + Q_{j^0} = \frac{a - c - c_{s^0 i^0 j^0}}{b} = \frac{a - c - c_{s^1 i^1 j^0} - \mu_{s^1 i^1 j^0}}{b}$$

$$c_{s^1 i^1 j^0} + \mu_{s^1 i^1 j^0} = c_{s^0 i^0 j^0} > c_{s^1 i^1 j^0}$$

$$\mu_{s^1 i^1 j^0} > 0$$

$$\Rightarrow q_{s^1 i^1 j^0} = 0$$

This imply $(O_{S,\mathcal{P}_l})$ is not optimal $\rightarrow\leftarrow$.

Theorem 3.1 states that each market will be satisfied by the closest plant of the firms composing each coalition. This gives us a simply way to compute the equilibrium when there are no capacity constraints.

3.2.3 Coalition structure models

Once we are able to calculate the utility of a coalition, conditional on the coalition structure, the natural next question is: Which coalitions would form under this environment? This question is intended to be answered in this Section. Clearly, depending on the point of view different answers could be provided. We will build two Mixed Integer Linear Programming (MILP) to maximize the social or the private welfare. First, let recall some classical economic values.

Definition 3.1 We define the Consumer Surplus (CS) for a solution $\mathcal{P}_l \in \mathcal{P}$ as the summation of the consumer surplus of each market:

$$CS_{\mathcal{P}_l} = \sum_{j\in J} \frac{[a - P(Q_j^*)]Q_j^*}{2} \tag{3.21}$$

$$= \frac{b}{2}\sum_{j\in J}(Q_j^*)^2 \tag{3.22}$$

where Q_j^* it the equilibrium quantity sent to the market $j \in J$ by all the firms in the solution $\mathcal{P}_l$ and $P(Q_j^*)$ it is the price. We also define the Private Utility (PU) for a solution $\mathcal{P}_l \in \mathcal{P}$ as the summation of the coalition utilities and the Social Welfare (SW) as the summation of PU and CS :

$$PU_{\mathcal{P}_l} = \sum_{S\in\mathcal{P}_l} \pi^*_{S,\mathcal{P}_l} \tag{3.23}$$

$$SW_{\mathcal{P}_l} = PU_{\mathcal{P}_l} + CS_{\mathcal{P}_l} \tag{3.24}$$

We will use the following decision variable in the coalition structure models:

$x_{\mathcal{P}_l} \in \{0,1\}$: It takes the value 1 if the partition $\mathcal{P}_l$ is formed and 0 if not.

Henceforth, we will consider that $\mathcal{P}_0$ correspond to the non collaborative solution and $\mathcal{P}_{|L|-1}$ to the grand coalition solution. We now define multiple feasible sets to be used in the models. These set correspond to different concepts of stability.

- $\mathcal{F}_0 = \{x \in \{0,1\}^{|L|} : \sum_{l\in L} x_{\mathcal{P}_l} = 1\}$
- $\mathcal{F}_1 = \{x \in \{0,1\}^{|L|} : \sum_{l\in L} x_{\mathcal{P}_l} = 1; \forall l \in L,\ PU_{\mathcal{P}_l}\cdot x_{\mathcal{P}_l} \geq PU_{\mathcal{P}_0}\cdot x_{\mathcal{P}_l}\}$
- $\mathcal{F}_2 = \{x \in \{0,1\}^{|L|} : \sum_{l\in L} x_{\mathcal{P}_l} = 1; \forall l \in L, S \in \mathcal{P}_l, |S| > 1, \bar{S} \subset S, \bar{S} \neq \emptyset,\ \pi^*_{S,\mathcal{P}_l}\cdot x_{\mathcal{P}_l} \geq \left(\pi^*_{S\setminus\bar{S},H(\bar{S},S,\mathcal{P}_l)} + \pi^*_{\bar{S},H(\bar{S},S,\mathcal{P}_l)}\right)\cdot x_{\mathcal{P}_l}\}$
- $\mathcal{F}_3 = \{x \in \{0,1\}^{|L|} : \sum_{l\in L} x_{\mathcal{P}_l} = 1; \forall l \in L, S \in \mathcal{P}_l, \pi^*_{S,\mathcal{P}_l}\cdot x_{\mathcal{P}_l} \geq \left(\sum_{k\in S}\pi^*_{\{k\},\mathcal{P}_0}\right)\cdot x_{\mathcal{P}_l}\}$

where

$$H(\bar{S}, S, \mathcal{P}_l) = \{\bar{S}\} \cup \{S \setminus \bar{S}\} \cup \left(\bigcup_{S' \in \mathcal{P}_l, S' \neq S} \{S'\} \right) \in \mathcal{P}$$

Given a coalition $S \in \mathcal{P}_l$ such $|S| > 1$ and a subcoalition $\bar{S} \subset S, \bar{S} \neq \emptyset$ the partition $H(\bar{S}, S, \mathcal{P}_l)$ is equal to the partition $\mathcal{P}_l$ with the only difference that $S \in \mathcal{P}_l$ is splitted into $\bar{S}$ and $S \setminus \bar{S}$ (Figure 3.3).

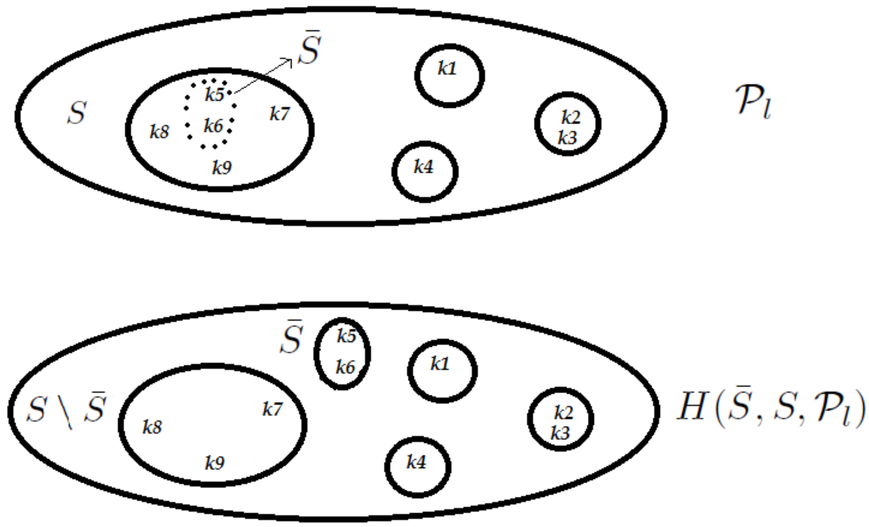


Figure 3.3: An example for the partition $H(\bar{S}, S, \mathcal{P}_l)$.

The feasible set $\mathcal{F}_0$ does not consider any stability constraint. $\mathcal{F}_1$ consider that the partition selected should have a better or equal private utility than the non-collaborative partition so the firms are willing to participate in the collaboration as long as transferable utility is allowed. $\mathcal{F}_2$ imposes a condition to avoid that some members of a coalition has the incentives to split it in two. This idea is similar to the concept of the *core* stability (Aumann, 1961). The feasible set $\mathcal{F}_3$ imposes a stability constraint which could be stated as follows: the coalition structure should be built in a way such each coalition perceive a utility greater or equal than the sum of the stand-alone utility of its member. This idea is similar to the concept of *semicore* stability (Aumann and Hart, 1992) in classical game theory. We define the set $\mathcal{F}$ of the feasible coalition structures which could be equal to $\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2, \mathcal{F}_3$ depending on the decision maker policy.

We define the Private Coalition Structure Problem (PCSP) as follows:

$$\text{maximize} \sum_{l \in L} PU_{\mathcal{P}_l} \cdot x_{\mathcal{P}_l}$$
$$x \in \mathcal{F}$$

Analogously, we define the Social Coalition Structure Problem (SCSP) as follows:

$$\text{maximize} \sum_{l \in L} SW_{\mathcal{P}_l} \cdot x_{\mathcal{P}_l}$$
$$x \in \mathcal{F}$$

The SCPC can be seen as the problem faced up by a central planner which wants to maximize the social welfare subject to stability ($\mathcal{F}_2$ or $\mathcal{F}_3$) and firms willing to participate constraints ($\mathcal{F}_1$). On the other hand, the PCSP is faced up by the firms which want to maximize their private utility subject to stability ($\mathcal{F}_2$ or $\mathcal{F}_3$).

Note that the PCSP do not take into account the consumer welfare. The implementation of such solution could be rejected by anti-trust authorities. For instance, U.S. law, in contrast to that in Canada and other nations, does not explicitly allow 'efficiencies' or cost savings as a defense for merges (Head and Ries, 1997; Ross and Winter, 2005). In our context, we could model the U.S law by adding a *CS constraint* defined as follows

$$x_{\mathcal{P}_l} \cdot CS_{\mathcal{P}_l} \geq x_{\mathcal{P}_l} \cdot CS_{\mathcal{P}_0} \quad \forall l \in L \tag{3.25}$$

This constraint states that the CS of the partition formed should be as good as the CS of the competitive solution. This ensures that the consumers will be better or equal compared to the perfect competition solution. On the other hand, the Canadian law could be modeled including a *SW constraint* defined as follows:

$$x_{\mathcal{P}_l} \cdot SW_{\mathcal{P}_l} \geq x_{\mathcal{P}_l} \cdot SW_{\mathcal{P}_0} \quad \forall l \in L \tag{3.26}$$

This constraint states that the SW of the partition formed should be as good as the SW of non-collaborative solution $\mathcal{P}_0$. Even though it is not possible to a ensure that the consumer will be better with the collaborative solution, it is possible to compensate the consumers in order that both the firms and the consumer are better than in $\mathcal{P}_0$. This is possible due to the cost savings achieved by the firms due to collaboration.

3.3 Numerical experiments

3.3.1 An ilustrative example

In this Section we will study an illustrative example with three firms ($K = \{K1, K2, K3\}$) and two markets ($J = \{J1, J2\}$) and the parameters $a = 30,\ b = 5,\ c = 1$. Each firm has only 1 plant with no capacity constraints. The transportation costs are given in the next figure using block distances:

The set of all the possible partitions of K is composed by 5 elements $\mathcal{P} = \{\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2, \mathcal{P}_3, \mathcal{P}_4\}$ such:

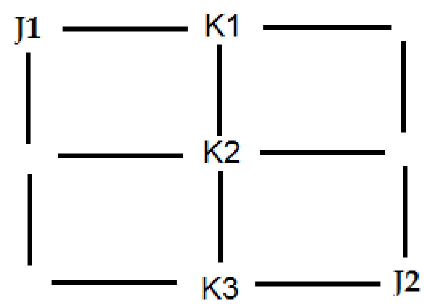


Figure 3.4: Transportation costs.

$$\begin{aligned}
\mathcal{P}_0 &= \{\{K1\}, \{K2\}, \{K3\}\} \\
\mathcal{P}_1 &= \{\{K1, K2\}, \{K3\}\} \\
\mathcal{P}_2 &= \{\{K1\}, \{K2, K3\}\} \\
\mathcal{P}_3 &= \{\{K1, K3\}, \{K2\}\} \\
\mathcal{P}_4 &= \{\{K1, K2, K3\}\}
\end{aligned}$$

Using the equilibrium equation 3.14 we compute the utility for all the coalition within each partition. For sake of explanation and taking into consideration that each firm has one plant we will denote q_{kij} simply q_{kj}.

$\mathcal{P}_0$:

$$\begin{aligned}
2q_{11} + q_{21} + q_{31} &= 5.6 \qquad (s^* = K1, j^* = J1) \\
2q_{12} + q_{22} + q_{32} &= 5.2 \qquad (s^* = K1, j^* = J2) \\
q_{11} + 2q_{21} + q_{31} &= 5.4 \qquad (s^* = K2, j^* = J1) \\
q_{12} + 2q_{22} + q_{32} &= 5.4 \qquad (s^* = K2, j^* = J2) \\
q_{11} + q_{21} + 2q_{31} &= 5.2 \qquad (s^* = K3, j^* = J1) \\
q_{12} + q_{22} + 2q_{32} &= 5.6 \qquad (s^* = K3, j^* = J2)
\end{aligned}$$

The solution of this linear system is shown in the following figure:
It is interesting to remark that the coalitions $\{K1\}$ and $\{K3\}$ obtain more utility than $\{K2\}$ even though the three of them have the same average distance to the markets ($\bar{c}_{kij} = 2$). This occurs because in the market $J1$ the coalition $\{K1\}$ has advantages compared to $\{K2\}$, so $\{K1\}$ can offer more product so $\{K2\}$ is obliged to offer less. The same situation occurs in the market $J2$ with $\{K3\}$ and $\{K2\}$.

$\mathcal{P}_1$:

$$\begin{aligned}
2q_{11} + 2q_{21} + q_{31} &= 5.6 \qquad (s^* = K1, j^* = J1) \\
q_{12} &= 0 \qquad (s^* = K1, j^* = J2) \\
q_{21} &= 0 \qquad (s^* = K2, j^* = J1) \\
2q_{12} + 2q_{22} + q_{32} &= 5.4 \qquad (s^* = K2, j^* = J2) \\
q_{11} + q_{21} + 2q_{31} &= 5.2 \qquad (s^* = K3, j^* = J1) \\
q_{12} + q_{22} + 2q_{32} &= 5.6 \qquad (s^* = K3, j^* = J2)
\end{aligned}$$

The solution of this linear system it is shown in the following figure:

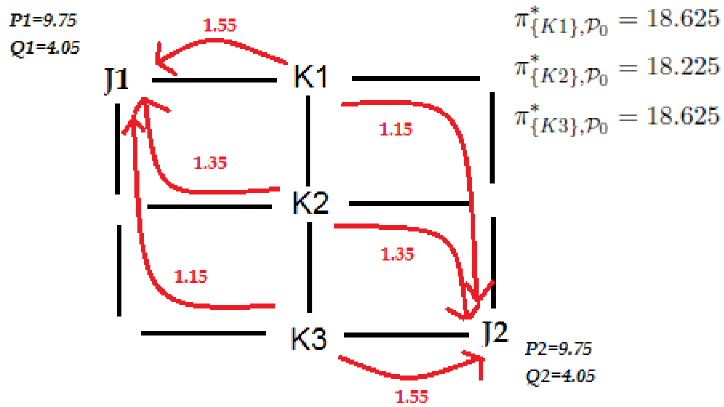


Figure 3.5: Solution for $\mathcal{P}_0$: $PU_{\mathcal{P}_0} = 55.475,\ \pi^*_{\{K1\},\mathcal{P}_0} = 18.625,\ \pi^*_{\{K2\},\mathcal{P}_0} = 18.225,\ \pi^*_{\{K3\},\mathcal{P}_0} = 18.625.$

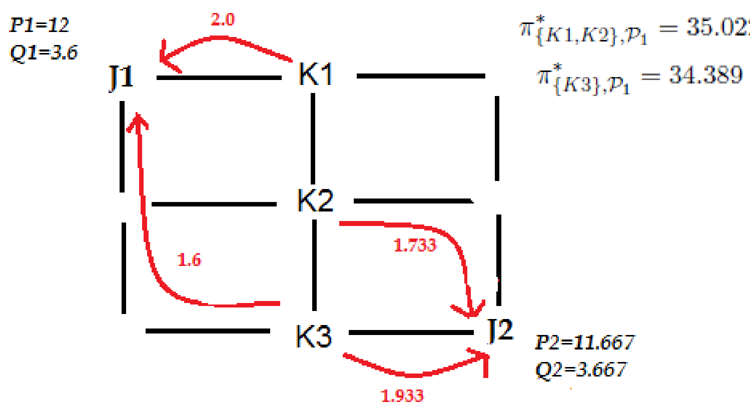


Figure 3.6: Solution for $\mathcal{P}_1$:$PU_{\mathcal{P}_1} = 66.511,\ \pi^*_{\{K1,K2\},\mathcal{P}_0} = 35.022,\ \pi^*_{\{K3\},\mathcal{P}_0} = 31.489.$

$\mathcal{P}_2$:

$$\begin{aligned}
2q_{11} + q_{21} + q_{31} &= 5.6 \qquad (s^* = K1, j^* = J1)\\
2q_{12} + q_{22} + q_{32} &= 5.2 \qquad (s^* = K1, j^* = J2)\\
q_{11} + 2q_{21} + 2q_{31} &= 5.4 \qquad (s^* = K2, j^* = J1)\\
q_{22} &= 0 \qquad (s^* = K2, j^* = J2)\\
q_{31} &= 0 \qquad (s^* = K3, j^* = J1)\\
q_{12} + 2q_{22} + 2q_{32} &= 5.6 \qquad (s^* = K3, j^* = J2)
\end{aligned}$$

The solution of this linear system it is shown in the following figure:

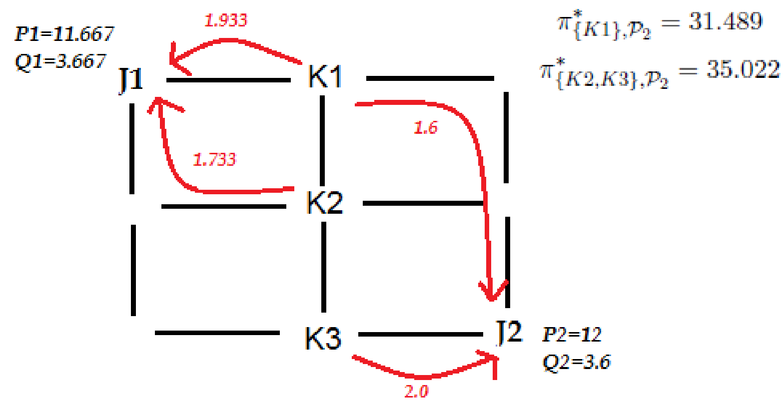


Figure 3.7: Solution for $\mathcal{P}_2$: $PU_{\mathcal{P}_2} = 66.511,\ \pi^*_{\{K1\},\mathcal{P}_2} = 31.489,\ \pi^*_{\{K2,K3\},\mathcal{P}_2} = 35.022.$

$\mathcal{P}_3$:

$$\begin{aligned}
2q_{11} + q_{21} + 2q_{31} &= 5.6 \quad (s^* = K1, j^* = J1)\\
q_{12} &= 0 \quad (s^* = K1, j^* = J2)\\
q_{11} + 2q_{21} + q_{31} &= 5.4 \quad (s^* = K2, j^* = J1)\\
q_{12} + 2q_{22} + q_{32} &= 5.4 \quad (s^* = K2, j^* = J2)\\
q_{31} &= 0 \quad (s^* = K3, j^* = J1)\\
2q_{12} + q_{22} + 2q_{32} &= 5.6 \quad (s^* = K3, j^* = J2)
\end{aligned}$$

The solution of this linear system it is shown in the following figure:

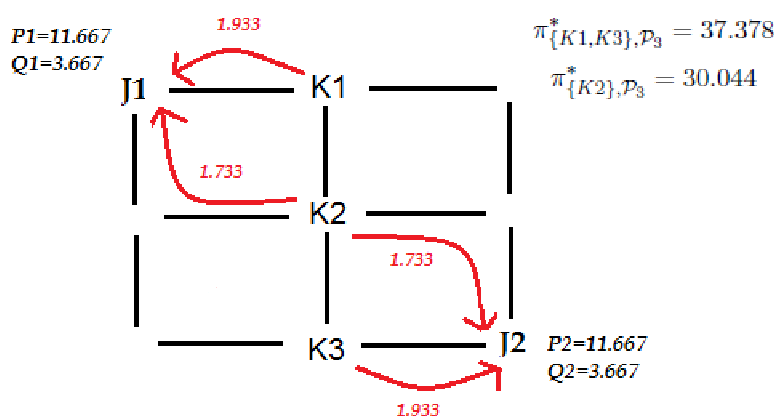


Figure 3.8: Solution for $\mathcal{P}_3$: $PU_{\mathcal{P}_3} = 67.422,\ \pi^*_{\{K1,K3\},\mathcal{P}_3} = 37.378,\ \pi^*_{\{K2\},\mathcal{P}_3} = 30.044.$

$\mathcal{P}_4$:

$$
\begin{aligned}
2q_{11} + 2q_{21} + 2q_{31} = 5.6 \quad & (s^* = K1, j^* = J1) \\
q_{12} = 0 \quad & (s^* = K1, j^* = J2) \\
q_{21} = 0 \quad & (s^* = K2, j^* = J1) \\
q_{22} = 0 \quad & (s^* = K2, j^* = J2) \\
q_{31} = 0 \quad & (s^* = K3, j^* = J1) \\
2q_{12} + 2q_{22} + 2q_{32} = 5.6 \quad & (s^* = K3, j^* = J2)
\end{aligned}
$$

The solution of this linear system it is shown in the following figure:

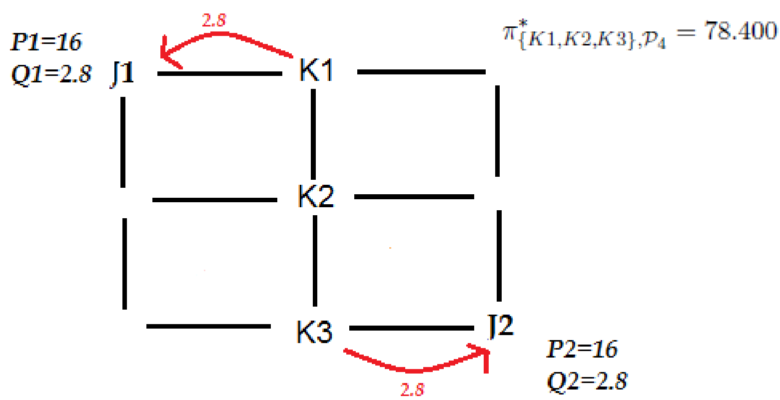


Figure 3.9: Solution for $\mathcal{P}_4$: $PU_{\mathcal{P}_4} = 78.400$, $\pi^*_{\{K1,K2,K3\},\mathcal{P}_3} = 78.400$.

It is clear that if the utilities are properly allocated, any member of the grand coalition $\{K1, K2, K3\}$ would not have the incentive to move to any other the solution.

Let us consider now the same situation described above but with some small changes in the topology as shown in the figure 3.10. The parameter d allow us to study different topologies. If $d = 1$ we retrieve the previous case. The Table 3.3 shows the PU, CS and SW for different values of the parameter d.

For $d = 1$, from the point of view of the firms, the best option is $\mathcal{P}_4$. From the point of view of the consumers, the best option is $\mathcal{P}_0$. From a social point of view, the best solution is also $\mathcal{P}_0$. For $d \in \{2, 3, 4, 5\}$ we observe the SW is maximized for the solution $\mathcal{P}_3$ which it not the grand coalition neither the perfect competition solution. Even though for the consumers the solution $\mathcal{P}_0$ is preferable (non collaborative solution), it is possible for the firms to compensate the consumers. For instance, if $d = 4$ the CS for the perfect competition solution $\mathcal{P}_0$ is 64.800. No regulatory authority would allow a coalitional scheme with a CS less than 64.800. In the solution $\mathcal{P}_3$ the CS is 60.089. In order to keep the solution $\mathcal{P}_3$ socially acceptable the firms should agree a compensation of 4.711 to consumers. The PU for $\mathcal{P}_3$ is 63.289 if we subtract, we obtain 58.578 which it is higher than the PU for the solution $\mathcal{P}_0$ which is 56.000. In conclusion, even taking into account the compensation, the firms would be better.

Table 3.1 shows us that even including multiple stability constraints the solution $\mathcal{P}_3$ is selected by

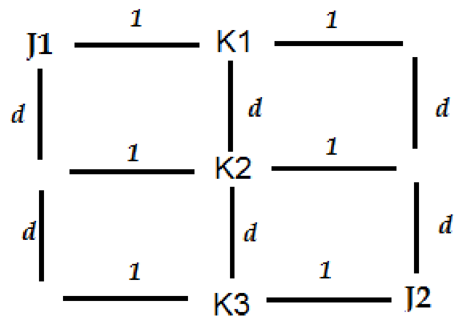


Figure 3.10: Transportation costs.

Constraints	Max SW	Solution
$\mathcal{F}_0$	130.400	$\mathcal{P}_3$
$\mathcal{F}_1$	130.400	$\mathcal{P}_3$
$\mathcal{F}_2$	130.400	$\mathcal{P}_3$
$\mathcal{F}_3$	130.400	$\mathcal{P}_3$

Table 3.1: SCSP models for different stability constraints with $d = 2$.

the models.

3.3.2 A Swedish forestry case study

In this Section, we will test our proposed approach using real-data from the Swedish forestry industry. The main objective of this Section is to analyze whether or not our proposed methodology is computationally solvable for real instances in a reasonable amount of time in order to determine the potential of its application. It is out of the scope of this Chapter to draw specifics conclusions for this case.

Constraints	Max PU	Solution
$\mathcal{F}_0$	78.400	$\mathcal{P}_4$
$\mathcal{F}_1$	78.400	$\mathcal{P}_4$
$\mathcal{F}_1$+SW-C	65.600	$\mathcal{P}_3$
$\mathcal{F}_1$+CS-C	53.900	$\mathcal{P}_0$
$\mathcal{F}_2$	78.400	$\mathcal{P}_4$
$\mathcal{F}_2$+SW-C	65.600	$\mathcal{P}_3$
$\mathcal{F}_2$+CS-C	53.900	$\mathcal{P}_0$
$\mathcal{F}_3$	78.400	$\mathcal{P}_4$
$\mathcal{F}_3$ +SW-C	65.600	$\mathcal{P}_3$
$\mathcal{F}_3$+CS-C	53.900	$\mathcal{P}_0$

Table 3.2: PU maximization models for different stability constraints $d = 2$.

d	Partition	Private Utility	Consumer Surplus	Social Welfare
	$\mathcal{P}_0$	55.475	82.013	137.488
	$\mathcal{P}_1$	66.511	66.011	132.522
1	$\mathcal{P}_2$	66.511	66.011	132.522
	$\mathcal{P}_3$	67.422	67.222	134.644
	$\mathcal{P}_4$	78.400	39.200	117.600
	$\mathcal{P}_0$	53.900	76.050	129.950
	$\mathcal{P}_1$	64.444	62.444	126.889
2	$\mathcal{P}_2$	64.444	62.444	126.889
	$\mathcal{P}_3$	65.600	64.800	130.400
	$\mathcal{P}_4$	78.400	39.200	117.600
	$\mathcal{P}_0$	54.075	70.313	124.388
	$\mathcal{P}_1$	63.489	58.989	122.478
3	$\mathcal{P}_2$	63.489	58.989	122.478
	$\mathcal{P}_3$	64.222	62.422	126.644
	$\mathcal{P}_4$	78.400	39.200	117.600
	$\mathcal{P}_0$	56.000	64.800	120.800
	$\mathcal{P}_1$	63.644	55.644	119.289
4	$\mathcal{P}_2$	63.644	55.644	119.289
	$\mathcal{P}_3$	63.289	60.089	123.378
	$\mathcal{P}_4$	78.400	39.200	117.600
	$\mathcal{P}_0$	59.675	59.123	119.188
	$\mathcal{P}_1$	64.911	52.411	117.322
5	$\mathcal{P}_2$	64.911	52.411	117.322
	$\mathcal{P}_3$	62.800	57.800	120.600
	$\mathcal{P}_4$	78.400	39.200	117.600

Table 3.3: Illustrative example utilities for different values of the parameter d.

Frisk et al. (2010) presented a case in which 8 forestry companies agreed to collaborate in the transportation phase of the wood production process. The companies are private and/or state-owned and they have multiple harvest zones and sawmills in the country. The main transportation task at this stage is to move timber, particularly sawlogs and pulplogs from the harvest zones to the sawmills where the logs are converted into different products. The transportation at this stage is made mainly by truck. We will consider the harvest zones as supply points and the sawmills as markets. Consequently to our approach assumptions, we consider that the supplies and the demands are not fixed.

The companies do not transport the same types of timber. We have decided to choose the timber type which is transported by the maximum number of companies. The selected product is the so-called *tall timber*. This kind of product is transported by 7 of the 8 companies so our case study will be such $|K| = 7$ that implies $|\mathcal{P}| = 877$ possible coalitions structures. The exact location of the harvest zones and sawmills are known so the distances are computed (units: hundreds of kilometers). From an economic point of view, it is not clear we can associate in a straightforward way a specific sawmill with a whole market but rather with multiple sawmills economically connected. It is difficult to argue the number of Swedish sawing markets could exceed 10. To take into account this remark and to analyze the behavior of our proposed approach for different scenarios, we define two parameters I_{max} and J_{max} corresponding to the maximum number of harvest zones and sawmills allowed in the model. The parameters a and b take the values 30 and 5 respectively. Figure 3.4 shows the coalition structure which maximizes the SW subject to $\mathcal{F}_0$.

The maximum number of firms per coalition is 2, particularly the grand coalition never formed. Large coalitions are socially difficult to form. On the other hand, the non-competitive solution $\mathcal{P}_0$

I_{max}	J_{max}	Solution	Social Welfare	Computing time (seconds)
1	1	[0, 1], [2], [3], [4], [5, 6]	78.22	0.48
5	1	[0, 1], [2], [3], [4], [5, 6]	79.83	3.47
10	1	[0, 1], [2], [3], [4], [5, 6]	81.14	8.92
15	1	[0, 1], [2], [3], [4], [5, 6]	81.31	17.21
1	2	[0, 1], [2], [3], [4], [5], [6]	150.63	0.71
5	2	[0, 1], [2], [3], [4, 5], [6]	154.4	6.73
10	2	[0, 1], [2], [3], [4, 5], [6]	156.72	21.73
15	2	[0, 1], [2], [3], [4, 5], [6]	157.47	46.72
1	5	[0], [1], [2], [3], [4], [5], [6]	369.41	2.72
5	5	[0, 4], [1], [2], [3], [5], [6]	376.57	24.82
10	5	[0, 4], [1], [2], [3], [5], [6]	382.23	98.52
15	5	[0, 4], [1], [2], [3], [5], [6]	383.83	220.67
1	10	[0], [1], [2], [3], [4], [5], [6]	741.2	5.06
5	10	[0, 4], [1], [2], [3], [5], [6]	754.29	83.27
10	10	[0, 4], [1], [2], [3], [5], [6]	765.27	318.42
15	10	[0, 4], [1], [2], [3], [5], [6]	769.09	690.21

Table 3.4: Maximum SW for multiple I_{max} and J_{max} with $\mathcal{F} = \mathcal{F}_0$.

Constraints	Max SW	Solution
$\mathcal{F}_0$	225.85	[0, 1], [2], [3], [4, 5], [6]
$\mathcal{F}_1$	225.83	[0, 2], [1], [3], [4, 5], [6]
$\mathcal{F}_2$	225.55	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_3$	225.55	[0], [1], [2], [3], [4], [5], [6]

Table 3.5: SCSP models for different stability constraints.

formed only once. Intermediate solutions are preferred by the model. Fixing the number of markets, the social welfare of the solutions increases when the maximum number of harvest zones I_{max} increases. This is not surprising because if I_{max} increase the unitary transportation costs decrease which implies that a better solution should be found. The computational times are reasonable for the tested instances.

We now turn our attention to the other coalition structures models shown in Section 3.2.3 The parameters a and b are set again as 30 and 5 respectively. We have also fixed $I_{max} = 3$ and $J_{max} = 3$. The solutions were computed by the algorithm shown in Appendix B which was coded in Python 3.5.2.3 and run on a standard computer with an Intel(R) Core(TM) i3-3227U CPU @ 1.90HGz processor.

The results presented in Table 3.5 show that for the constraints set $\mathcal{F}_0$ and $\mathcal{F}_1$ there are two coalitions composed of two firms each. For the constraints set $\mathcal{F}_2$ this two coalition are not allowed because of stability constraints. The sum of the utility of $\{0\}$ plus the utility of $\{2\}$ in the new coalition structure is bigger than the utility of $\{0, 2\}$ in the original coalition structure.

If there is no anti-trust constraints and subject to $\mathcal{F}_1$ stability constraints, the optimal PU is attained in the grand coalition solution (see Table 3.6) which is an expected result according to the classical Cournot competition model. When the constraint SW-C is added to the model the optimal solution is $[0, 3, 5, 6], [1, 4], [2]$ which is an interesting result because: (i) Due to $\mathcal{F}_1$ it is possible to allocate to each firm a utility larger than their stand alone utility in the non-collaborative environment so the companies are willing to participate in the collaboration, (ii) Since the SW of this solution is larger than the non-collaborative solution SW, then it is possible to compensate the consumers

Constraints	Max PU	Solution
$\mathcal{F}_0$	120.02	[0, 1, 2, 3, 4, 5, 6]
$\mathcal{F}_1$	120.02	[0, 1, 2, 3, 4, 5, 6]
$\mathcal{F}_1$+SW-C	88.01	[0, 3, 5, 6], [1, 4], [2]
$\mathcal{F}_1$+CS-C	50.99	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_2$	120.0	[0, 1, 2, 3, 4, 5, 6]
$\mathcal{F}_2$+SW-C	50.99	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_2$+CS-C	50.99	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_3$	50.99	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_3$ +SW-C	50.99	[0], [1], [2], [3], [4], [5], [6]
$\mathcal{F}_3$+CS-C	50.99	[0], [1], [2], [3], [4], [5], [6]

Table 3.6: PUSP models for diferent stability constraints.

in a way that they will be better than in the non-collaborative situation. In fact, the SW of this solution is 219.74 which is 5.81 less than the non-collaborative solution $\mathcal{P}_0$. The PU of the solution $[0, 3, 5, 6], [1, 4], [2]$ is 88.01. The PU minus the compensation is 82.2, thus the non-collaborative solution has a total PU of 50.99, that is, an increase of 61.2% in utility. This implies that if the proper compensations are carried out in a proper way, the collaborative agreement could pass anti-trust regulations, thus, the firms will be much better just as consumers.

3.3.3 Improving computing times: a branch and cut algorithm and an heuristic solution approach for PCSP

As discussed in Section 3.2.2 and explicitly exemplify in Section 3.3.1, computing the utilities of each coalition for each coalition structure requires to solve several linear systems. Even though this could be done fast for small instances, it is not clear whether this is a suitable solving strategy for larger instances. In order to improve on this issue, in this Section, we propose another solving strategy. First, we will announce and prove Theorem 3.2 which gives us a closed formula to compute the coalition equilibrium utilities. Secondly, using this theorem we propose a Branch and Cut (B&C) algorithm and a heuristic solution approach for solving the PCSP.

Theorem 3.2 Given a partition $\mathcal{P}_l$ and a market $j \in J$. If there are no capacity constraints then in the equilibrium we have :

$$Q_j^* = \frac{1}{\sum_{S\in\mathcal{P}_l} \mathbb{1}_{a-c-c_j(S)\geq 0} \cdot \mathbb{1}_{(Q_j^S)^*>0} + 1} \sum_{S\in\mathcal{P}_l} \frac{a-c-c_j(S)}{b} \cdot \mathbb{1}_{a-c-c_j(S)\geq 0} \cdot \mathbb{1}_{(Q_j^S)^*>0}$$

So for each coalition $Z \in \mathcal{P}_l$

$$(Q_j^Z)^* = \left[\frac{a-c-c_j(Z)}{b} - Q_j^*\right]_+ \cdot \mathbb{1}_{a-c-c_j(Z)\geq 0}$$

Proof: By theorem 3.1 we have that $\forall S \in \mathcal{P}_l$:

$$(Q_j^S)^* + Q_j^* = \frac{a - c - c_j(S)}{b} \text{ if } Q^S > 0,\ a - c - c_j(S) \geq 0 \tag{3.27}$$

$$(Q_j^S)^* = 0 \quad \sim \tag{3.28}$$

$$\tag{3.29}$$

Adding up,

$$\Big(\sum_{S \in \mathcal{P}_l} \mathbb{1}_{a-c-c_j(S)\geq 0} \cdot \mathbb{1}_{(Q_j^S)^*>0} + 1\Big) Q_j^* = \sum_{S \in \mathcal{P}_l} \frac{a - c - c_j(S)}{b}$$

$$Q_j^* = \frac{1}{\sum_{S \in \mathcal{P}_l} \mathbb{1}_{a-c-c_j(S)\geq 0} \cdot \mathbb{1}_{(Q_j^S)^*>0} + 1} \sum_{S \in \mathcal{P}_l} \frac{a - c - c_j(S)}{b}$$

Reemplacing,

$$(Q_j^Z)^* = \left[\frac{a - c - c_j(Z)}{b} - Q_j^*\right]_+ \cdot \mathbb{1}_{a-c-c_j(Z)\geq 0}$$

This last theorem allows us to improve the solving strategy because we do not need to solve a linear system to compute the equilibrium quantities and utilities for each partition. For instance, for $|I| = 15, |J| = 10$ and $|K| = 7$ the computing time using this formula is 4.46 seconds while with linear systems approach the computing time is 690.21 seconds, an improvement by two orders of magnitude.

As it is shown in Appendix B, all the experiments shown in previous Sections are solved by enumeration. That is, we compute the equilibrium utilities for all the coalition structures. As discussed in Section 3.2.1, the number of structures grows very fast as the number of firms grows. Accordingly, the logic we pursue by the following two solving strategy, namely B&C and the greedy heuristic, is to avoid to compute all the coalition structures $\mathcal{P}_l$.

Define a vector W as the vector of the coalition structures $\mathcal{P}_l \in \mathcal{P}$ ordered in non-increasing according to its private utility.
Compute the equilibrium utilities $SW_{\mathcal{P}_0}, PU_{\mathcal{P}_0}, CS_{\mathcal{P}_0}$
for $p \in W$ **do**
    if p is feasible **then**
        Remove all the partitions of W which are a child of p
        Compute SW_p, PU_p, CS_p
    end if
end for
Return p such PU_p is feasible and maximized

Algorithm 1: Branch and Cut algorithm for the PCSP.

The B&C algorithm (Algorithm 1) is based on the following idea. We consider a partition $\mathcal{P}_l \in \mathcal{P}$, a coalition $S \in \mathcal{P}_l$ such $|S| > 1$ and a subset $\overline{S} \subset S$ such $\overline{S} \neq \emptyset$ then usually $PU_{H(\overline{S},S,\mathcal{P}_l)} \leq PU_{\mathcal{P}_l}$.

This idea implies that if a feasible solution is found there is no need to compute their children coalition structures. In this context we will say that the partition q is a children of p if there exist $S \in p$ such $|S| > 1$ and a subset $\overline{S} \subset S$ such $\overline{S} \neq \emptyset$ such $H(\overline{S}, S, p) = q$. That is, p is partition formed by the merge of two coalitions of the partition q. Even though our experimental results (Tables 3.7, 3.8 and 3.9) suggest that the B&C based idea is true, we were able to find a counter-example. Consider an instance with $|K| = 3, |I| = 1, |J| = 1, a = 800, b = 200, c = 500$ and the distances to the unique market equals to 100, 0.1, 1 respectively for each plant of the companies. In this case we have $PU_{\{\{0\},\{1,2\}\}} = 94.36$ and $PU_{\{\{0\},\{1\},\{2\}\}} = 99.45$. This instance is quite particular because one of the firm's plant is located quite far from the market compared to the others two.

The greedy heuristic approach (see Algorithm 2) use the fact that generally, the better solutions have fewer coalitions. The procedure computes the equilibrium utilities for the coalition structure with less coalition first. If a coalition structure $p \in \mathcal{P}$ is feasible, the algorithm ends when all the coalitions structures with the same number of coalitions are revised.

Define a vector W as the vector of the coalition structures $\mathcal{P}_l \in \mathcal{P}$ ordered in non-decreasing according the number of coalitions
Compute the equilibrium utilities $SW_{\mathcal{P}_0}, PU_{\mathcal{P}_0}, CS_{\mathcal{P}_0}$
Define $h \leftarrow 1$
Define $s \leftarrow 0$
Define $p^* \leftarrow \mathcal{P}_0$
for $p \in W$ **do**
    if $s = 1$ and the number of coalitions in p is greater than h **then**
        Return p^*
    end if
    if p is feasible **then**
        Compute SW_p, PU_p, CS_p
        $h \leftarrow$ number of coalition of p
        $s \leftarrow 1$
        $p^* \leftarrow p$
    end if
end for

Algorithm 2: Greedy heuristic for the PCSP.

We show now experimental results for the enumeration, B&C and the heuristic solutions. In all the experiments we have fixed the parameters $a = 800, b = 10, c = 100, \mathcal{F} = \mathcal{F}_0 - SW$. The costs were randomly generated in the interval $[1, 300]$ for each instance. In Tables 3.7, 3.8 and 3.9 We show the computational time (seconds) and the number of partition computed (RS) in parenthesis.

In the experiments, the objective function for the enumeration, B&C and the greedy heuristic always coincide. As we can see in Tables 3.7, 3.8 and 3.9, computational times for the heuristic approach are always the lowest. As the number of plants grows the solution tends to have more coalitions (Table 3.7). This implies that the computational times for the three solution strategies tend to be equal because more coalition structures have to be computed in all cases.

$\lvert I \rvert$	Opt. time (OF)	B&C Time (OF)	Heu. time (OF)	Data time
5	0.311 (877)	0.290 (834)	0.286 (715)	0.349
10	0.505 (877)	0.476 (851)	0.386 (715)	0.381
20	0.905 (877)	0.853 (874)	0.839 (855)	0.574
40	1.613 (877)	1.613 (877)	1.613 (877)	0.580
100	3.816 (877)	3.816 (877)	3.816 (877)	0.960

Table 3.7: Experimental results varying the number of plants ($|J| = 3, |K| = 7$).

$\lvert J \rvert$	Opt. time (RS)	B&C Time (RS)	Heu. time (RS)	Data time
5	0.315 (877)	0.203 (534)	0.025 (64)	0.37
10	0.611 (877)	0.580 (809)	0.218 (365)	0.508
20	1.219 (877)	1.194 (850)	0.454 (365)	0.750
40	2.508 (877)	2.445 (870)	1.852 (715)	1.234
100	6.111 (877)	5.868 (873)	4.560 (715)	2.641

Table 3.8: Experimental results varying the number of markets ($|I| = 2, |K| = 7$).

$\lvert K \rvert$	Opt. time (OF)	B&C Time (RS)	Heu. time (RS)	Data time
6	0.046 (203)	0.031 (176)	0.015 (122)	0.031
7	0.234 (877)	0.234 (710)	0.093 (365)	0.375
8	1.125 (4140)	1.078 (3218)	0.328 (1094)	5.672
9	8.409 (21147)	7.892 (13001)	2.078 (256)	176.870
10	- (115975)	- (-)	87.431 (9842)	-
11	- (678570)	- (-)	4915.293 (29525)	-

Table 3.9: Experimental results varying the number of firms ($|I| = 2, |J| = 3$).

3.4 Concluding remarks

Previous efforts in horizontal collaborative transportation have assumed that demands and supplies are fixed. With this assumption the more firms the collaboration has the more the costs are reduced. This implies that the grand coalition always formed and then it is a matter of split the costs wisely. In this Chapter, we have introduced competition to the collaborative transportation classical environment. By changing this, the behavior of each coalition impact on the behavior of the others. Our proposed approach analyzes what happen if after the collaborative contracts are signed, the firms compete in a Cournot fashion. The logic we pursue is to analyze the middle-term collaborative process rather than short-term as it is generally studied in the previous literature.

The utility of each coalition, which in this case depends on the coalition structure of the other firms, it is computed by the Cournot-Nash equilibrium equations. We then proposed several coalition structure models to maximize either the social welfare (SWSP) or the private utility (PUSP) subject to stability and anti-trust regulation constraints. We have conducted numerical experiments through an illustrative example and a case-study based on real data from the Swedish forestry industry. We analyzed several scenarios of potential collaboration under the proposed models. Our main finding is that there exist coalition structures in which the firms are better off than the non-collaborative case and the consumers are susceptible to be compensated in order to be better or equal than the non-collaborative case. As well, and opposed to previous literature, we conclude that forming coalition is quite difficult. Moreover, the number of firms per coalition is usually low. Previous literature have stated that in practice big coalitions are more likely to fail (Flisberg et al., 2015; Guajardo and Rönnqvist, 2015). Our proposed approach could help to explain this discrepancy between the reality and practice.

Further research on this topic should address other competitions schemes such as Bertrand or Stackelberg models. Moreover, economies of scale could be included to encourage coalition formation. This Chapter analyzes the collaboration from a tactical perspective opposed to previous literature which is purely operational. From a strategical perspective, facilities location decisions could be added to the utility computation procedure to take into account a more long-term. Finally, the procedure in which the consumers are compensated in this collaboration scheme remains an open research question.

Appendix A

Let us define a Cournot model in which n firms compete in a unique market. We call q_i the quantity produced bu the firm i and $Q = \sum_{i=1}^{n} q_i$. The demand function in the market is $P(Q) = a - bQ$ where a, b are known parameters.

$$\begin{aligned}
\pi_i(q_i) &= P(Q)q_i - cq_i = (a - bQ)q_i - cq_i = aq_i - bQq_i - cq_i \quad \Rightarrow \\
\frac{\partial \pi_i}{\partial q_i} &= a - 2bq_i - bQ_{-i} - c \quad \text{by symmetry } \forall i \; q_i = q \\
&= a - b(n+1) - c = 0 \quad \Rightarrow \\
q^* &= \frac{a-c}{b(n+1)} \quad \Rightarrow \\
Q^* &= \frac{n(a-c)}{b(n+1)} \quad \Rightarrow \\
P(Q^*) &= \frac{a+nc}{n+1}
\end{aligned}$$

$$\pi_i^* = \left(\frac{a-c}{N+1}\right)^2 \left(\frac{1}{b}\right)$$

Suppose that $m \in \{2, 3, ..., N\}$ firms merges. The merges will be profitable for them iif

$$\begin{aligned}
\left(\frac{a-c}{N-m+1}\right)^2 \left(\frac{1}{b}\right) &> \left(\frac{m(a-c)}{N+1}\right)^2 \left(\frac{1}{b}\right) \\
\frac{1}{(N-m+1)^2} &> \frac{m^2}{(N+1)^2} \\
\frac{1}{(N-m+1)} &> \frac{m}{N+1} \\
\frac{1}{(N-m+1)} - \frac{m}{N+1} &> 0
\end{aligned}$$

The left-hand can be rewrite:

$$\frac{1}{(N-m+1)} - \frac{m}{N+1} = \frac{m^2 - (N+1)(m-1)}{(N+1)(N-m+1)} < \frac{m^2 - (N^2-1)}{(N+1)(N-m+1)} < 0 \;\; \forall m \in \{2, 3, ..., N-1\}$$

Overall, the merge is profitable only for the monopoly case $N = m$.

Appendix B

Compute the equilibrium utilities $\pi^*_{S,\mathcal{P}_l} \ \ \forall l \in L, \forall \mathcal{P}_l \in \mathcal{P}, \forall S \in \mathcal{P}_l$
Compute the social welfare $SW_{\mathcal{P}_l}$ of each coalition structure $\forall \mathcal{P}_l \in \mathcal{P}$
Define a vector W as the vector of the coalition structures $\mathcal{P}_l \in \mathcal{P}$ ordered in non-increasing according to its social welfare.
for $p \in W$ **do**
    Set $x_p = 1$ and $x_{p'} = 0 \ \ \forall p' \in \mathcal{P} : p' \neq p$
    if $\mathcal{F} \neq \emptyset$ **then**
        Return p and SW_p
    end if
end for

Algorithm 3: Exact algorithm solution approach for the SCSP.

Chapter 4

Real-time crash prediction in an urban expressway using disaggregated data

Abstract: We develop accident prediction models for a stretch of the urban expressway Autopista Central in Santiago, Chile, using disaggregate data captured by free-flow toll gates with Automatic Vehicle Identification (AVI) which, besides their low failure rate, have the advantage of providing disaggregated data per type of vehicle. The process includes a random forest procedure to identify the strongest precursors of accidents, and the calibration/estimation of two classification models, namely, Support Vector Machine and Logistic regression. We find that, for this stretch of the highway, vehicle composition does not play a first-order role. Our best model accurately predicts 67.89% of the accidents with a low false positive rate of 20.94%. These results are among the best in the literature even though, and as opposed to previous efforts, (i) we do not use only one partition of the data set for calibration and validation but conduct 300 repetitions of randomly selected partitions; (ii) our models are validated on the original unbalanced data set (where accidents are quite rare events), rather than on artificially balanced data.

Keywords: Real-Time Crash Prediction; Support Vector Machines; Logistic Regression; Automatic Vehicle Identification.

Difussion: A version of this Chapter has been published as: Basso, F., Basso, L. J., Bravo, F., & Pezoa, R. (2018). Real-time crash prediction in an urban expressway using disaggregated data. *Transportation Research Part C: Emerging Technologies*, 86, 202-219.

4.1 Introduction

Car accidents in cities are an important externality caused by traffic. Accidents imply congestion, delays and sometimes fatalities. For example, in Chile, 1,675 persons died in road accidents in 2016, the largest number in the last 8 years, while Rizzi and Ortúzar (2003) calculate that up to USD 1,300,000 are required in safety measures to avoid one death in interurban highways. Thus, understanding under what conditions accidents occur or, in different words, which traffic and external conditions increase the probability of a car accident, may have a sizeable impact. Furthermore, if those conditions were observed on line, then authorities or managers may have the chance to intervene in order to avoid accidents from happening. Nowadays, having traffic data on line is possible because of the new IT technologies which provides quality and bulk data to support monitoring traffic systems (Shi and Abdel-Aty, 2015). The purpose of this research is to study the precursors of car accidents in an urban expressway, using data that is available on-line to the expressway managers, in order to create a real-time accident prediction model which, in the future, may be transformed into a software tool. The on-line data is very rich: every car using this expressway has to have a transponder, so that the expressway can detect and charge them when they cross an Automatic Vehicle Identification (AVI) gate. One specific section of the expressway is studied, looking at data from AVI gates over a period of 18 months. We consider the afternoon rush-time, hence the focus is only on weekdays, using 80% (randomly selected) of the data for calibration purposes, while using the remaining 20% to test the predictive power of our model. Our results are promising: using the best classification model (logistic regression), we are able to predict 67.89% of the accidents (sensitivity), while making only 20.94% of false predictions (false alarm rate). In the binary crash-prediction context, the false alarm rate is defined as the number of misclassified non-accident divided by the total number of observations. The sensitivity is defined as the total number of correct predicted accidents divided by the total number of accidents.

Our approach can be summarized in four steps: (i) The traffic data from AVI gates is aggregated to five minutes averages, and then used to calculate variables that are of interest, such as flows per type of vehicle, speeds, speed change, variance of speeds, density and density change. The data set then will have 0 and 1s, corresponding to no accident or accident respectively. The data set is complemented from other sources that capture external conditions that may affect driving behavior such as, temperature, atmospheric pressure and rain. (ii) We then analyze this data both graphically and statistically, using a random forest procedure, in order to identify what are the variables that appear to be strong precursors of car accidents. (iii) The previous analysis are then used to calibrate two classification models, namely support vector machines (SVM) and logistic regression; for this, the first 80% of the data is used for calibration/training purposes and the remaining 20% for validation. (iv) In order to check for robustness of our models, the following is repeated 300 hundred times: randomly select 80% of the data base, calibrate both SVM and logistic models and then validate using the remaining 20%. This allow us to see dispersion in prediction power as the data changes, thus mimicking what would happen if an online prediction tool was at work, receiving new data continuously. With these results we compare the performances of our models.

There has been previous work on this area -some relevant references are reviewed below- however, there are two main general differences with previous efforts: data and the prediction/performance analysis. Regarding data, in this Chapter we work with data provided by a major tolled urban highway in Santiago, Chile, Autopista Central. This highway spans for 60.5 kms, crossing the metropolitan region from north to south, and connecting with the main interurban highway, Ruta 5. The highway is privately operated, and charge drivers according to the type of vehicle and distance by using AVIs and transponders installed in the vehicles. Since revenues come from AVIs, these devices have a very small failure rate, which enabled the acquisition of a detailed, disaggregated and rich traffic data set, that is, we know exactly at what time and at which speed each vehicle (separated by type) crossed an AVI. This contrast with previous efforts: as far as we know, the majority of the papers in the literature have worked with aggregated data, usually in periods of 30 seconds, without identification of type of vehicle, and using loop detectors which have a sizeable failure rate: according to Ahmed and Abdel-Aty (2012), loop detectors have a failure that ranges between 24% and 29%. Even tough, last years some efforts have been made in order to include AVI data to analyze accident rates (Abdel-Aty et al. (2012); Xu et al. (2013); Yu et al. (2014);Shi et al. (2016)). Disaggregated data differentiated by vehicle type allows us to explore a rather understudied issue: the influence of vehicle composition, and the corresponding speed differences, on the crash likelihood.

The second main difference with the previous literature is how the performance of the resulting models is tested. We improve on this issue on two aspects. First, all the papers reviewed below discarded some of the non-accident observations in order to "balance" the data set and, then, calibrated the model using a fraction of the adjusted data set (typically 70% or 80%) while using the remaining observations for validation. This calibration technique, however, was extended to validation/prediction: to the best our knowledge all previous papers tested the model using the same artificially balanced data, that is, on data that does not show the actual, real pattern of accidents being rare events (Theofilatos et al., 2016). While for the calibration of one of our classification models we do balance the data set (the SVM case), in all cases the performance was tested by attempting to predict accidents using real data, where accidents are indeed very rare events. It is hard to say with certainty how the models calibrated on artificially balanced data would perform on a real-time environment yet, our conjecture is that they necessarily will do worse. Our second improvement is on the robustness of the models. As far as we know, in all papers calibration is made for just one partition of the data which raises the question of robustness: would the parameters of the model be the same if a different partition were used? And would predictive power (also called sensitivity) remain the same? To answer these questions, we created the additional 300 repetitions explained above, in order to calculate 300 values for sensitivity and false positive rates, obtaining then the averages, maximums, minimums and standard deviations. Hence, it is important to keep in mind that, while some papers reviewed below may present performances similar to ours, that performance was achieved -in contrast to our case- in a non-real environment and using just one partition of the data. As we explicitly show, it is quite possible that for that one partition, results end up being much better than for others. The power of the calibrated model was also tested on traffic and crash data that was collected by Autopista Central on a period of time later than the one we had at hand. This test is what comes close to learn what would have been the result should a real-time model been working. The sensitivity was actually better than before: we are able to predict 75.03% of the accidents.

We now briefly review some important references. Golob and Recker (2004) used k-clustering techniques looking at 1000 crashes occurred in 1999 in Southern California, in order to define taxonomies for the flow regimes previous to an accident. Note the emphasis here is on identifying flow regimes that make more likely that an accident will occur, rather than on attaching an actual probability of accident to a particular traffic condition. In the beginning of this project we tried to use k-clustering techniques but its performance was evidently inferior so we did not pursue this more. For a recent review of the effect of flow regimes and climate conditions see Theofilatos and Yannis (2014).

Abdel-Aty et al. (2004) used a logistic model, as we do, but in a matched case-control setting, implying that not all the non-accident data is used, as opposed to what we do. They looked at data from the Interstate 4 in 1999 obtained from the Orlando Police Department and loop detectors installed approximately 0.5 miles apart. This model has a predictive power of 67%. The false alarm rate it not included in the paper.

More recently, SVM has been studied as the classification method for prediction. For instance, Lv et al. (2009) used simulated data obtained from the software TSIS to identify traffic conditions which increase the probability of accidents . Yu and Abdel-Aty (2013) used data from I-70 highway in Colorado to measure the risk of crash-accident in real time using SVM. As opposed to our case, the authors only select some of the observations with no accidents in order to avoid unbalanced data and facilitate calibration. We tackle the issue in a different way when calibrating SVM: a Synthetic Minority Over-sampling Technique (SMOTE) was used, discussed in detail in Section 4.4

Hossain and Muromachi (2012) used random multinomial logit model to identify the most important predictors and applied a Bayesian belief net procedure to predict crashes. The authors used data collected from Shibuya 3 and Shinjuku 4 expressways under the jurisdiction of Tokyo Metropolitan Expressway Company Limited in Japan. They obtained a mean sensitivity (predictive power) of 66% with a 20% false alarm rate. Recently, based on 551 crashes and corresponding speed information collected on expressways in Shanghai, China, Sun and Sun (2015) calibrate a dynamic Bayesian network with time series. They obtained maximum of 76.4% sensitivity and a false alarm rate of 23.7%.

The rest of the article is organized as follows. In Section 4.2 the data is described, explaining how it was processed and providing some descriptive statistics. In Section 4.3 we tackle the variable selection problem, in order to identify what are the strongest precursors of car accidents. In section 4.4 we present and calibrate the Support Vector Machine classification model for the initial partition of data (first 80% for calibration, last 20% for validation) while in Section 4.5 the same is done with the logistic regression model. In Section 4.6 we test for robustness and compare the performance of models, ours and the ones presented in the literature. Section 4.7 concludes.

4.2 Data set and preparation

Autopista Central is an expressway in Santiago, Chile, which is 60.5 kms long and has a north-south orientation (see Figure 4.1). The raw traffic data set they provided us with has traffic information from November 1st 2014 to April 30th 2016. A data point is the time and speed at which a certain vehicle (fully identified by its transponder) passed an AVI gate using any of the available lanes; in other words traffic per lane cannot be distinguished. Vehicles are classified as light (this include SUVs and smaller commercial vehicles), heavy (including trucks and buses) or motorcycle. On the other hand, Autopista Central also provided us with their accident information. This information is recorded manually: when any incident happens, they track it and store their type (accident, broken down car, roadworks, etc), date, time and exact location. In this work we are only interested in the accidents.

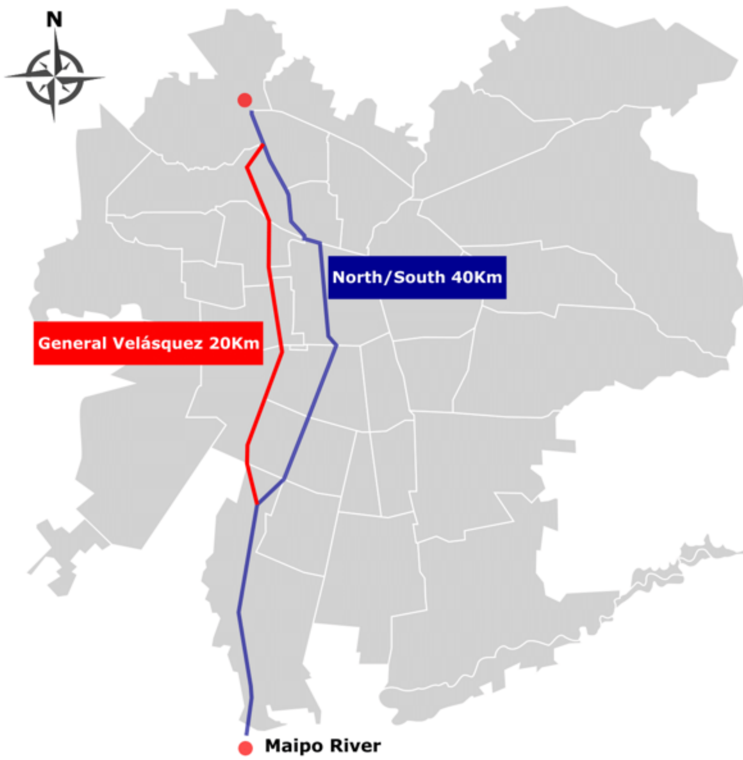


Figure 4.1: Autopista Central, Santiago, Chile.

The highway is divided in sections for managerial reasons. We decided to focus on the section of the expressway that has the largest accident rate per kilometer per unit of time, namely 3.41 accidents per kilometer per month. We studied the north-south direction of this section which spans for 4.7 kms between the Mapocho River and Carlos Valdovinos street, and has six entry ramps and three exit ramps. It has two AVI gates from where traffic information is obtained (see Figure 4.2). We consider the afternoon rush hour, that is, Monday to Friday from 5:30p to 8.30p, which left us with 10,745,766 observations, of which 5,298,683 correspond to the AVI gate AC-09 and 5,447,083 to the AVI gate AC-08. This difference is due to existence of two entry ramps between those AVI gates. By choosing a specific section and period, we think we can avoid the influence of, for example geometry or changes in driving behavior, thus helping us to better predict. This, we think, does not decrease the applicability of the overall approach, since a functioning real-time accident prediction tool may have different models running for different times of the day and different Sections of the road (Kwak and Kho, 2016).

The raw data was used to calculate 17 variables, averaged over periods of five minutes, for each of the two gates, giving us a total of 34 variables. They are, for each type of vehicle: flow, speed, standard deviation of the speed, density (that is, average flow divided in average speed) and density change, simply calculated as the difference with its value in the previous five minutes. A composition variable was also considered, defined as the proportion of each type of vehicle compared with the total flow. The right-hand side variables, for light vehicles and gate, are defined in Table 4.1 below:

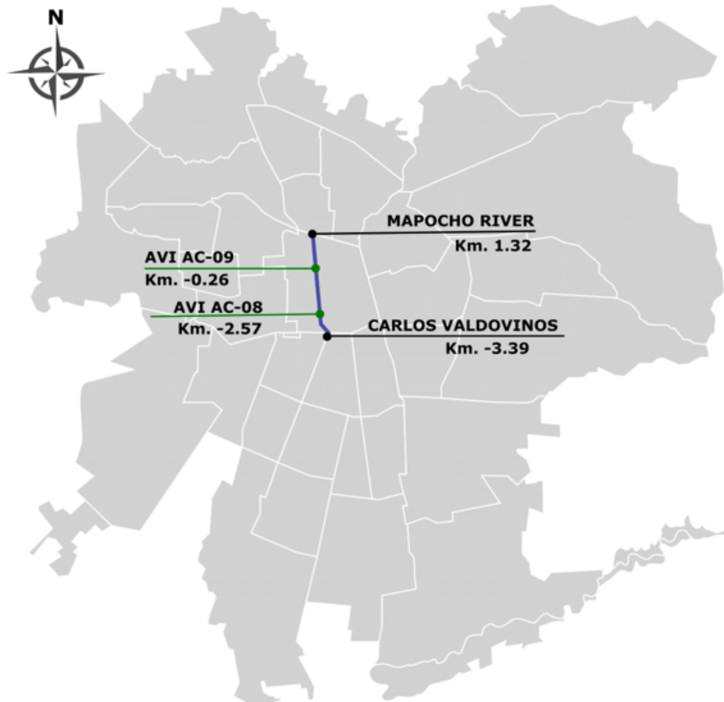


Figure 4.2: Section of the expressway studied.

Variable	Definition
Flow.Light.08	Total flow of light vehicles
Speed.Light.08	Mean speed of light vehicles
StdDev.Speed.Light.08	Standard deviation of the speed of light vehicles
Dens.Light.08	Temporal density of light vehicles, defined as Flow.Light/Speed.Light
Composition.Light.08	Percentage of light vehicles in total flow
Delta.Den.Light.08	Change in Dens.Light compared to the previous five minutes
Delta.Speed.Light.08	Change in Speed.Light compared to the previous five minutes

Table 4.1: Variables used for light vehicles crossing AVI gate 08.

The remaining variables are defined analogously, changing the type of vehicle and/or the AVI gate. The accident data was then used to create the 35^{th} variable, namely, whether there was an accident during the next period of five minutes or not. The variable takes a value of zero if there was no accident and a value of one if there was one. This data set was complemented with -for each period of five minutes - temperature, atmospheric pressure and rainfall. The weather data comes from a station installed by the Department of Geophysics of University of Chile, only 1 kilometer away from the north AVI (AC-09). The final data set has 13,029 observations (5 minutes periods) of which only 39, i.e. 0.30% had an accident, confirming the rare event feature discussed above. Tables 4.2 and 4.3 provide descriptive statistics which enable to show the main features of the traffic conditions of the section of the expressway studied. Figure 4.3 shows the evolution of the average speed of the lights vehicles during the studied period.

From Tables 4.2 and 4.3 is possible to see that the composition of the traffic in the studied section and period is given by an extremely high percentage of light vehicles, which follows from the fact

Table 4.2: Descriptive statistics of AVI gate AC-08.

AVI AC-08					
		Average	**Std. Dev.**	**Minimum**	**Maximum**
Light	Speed [km/h]	76.6	15.8	8.8	102.4
	Flow [veh]	386.6	50.8	118	527
	% Composition	92.9%	1.9%	83.4%	99.1%
	Density [veh/km]	5.5	2.2	1.4	18.9
Heavy	Speed [km/h]	71.2	14.0	7.9	104.5
	Flow [veh]	18.0	6.1	1	43
	% Composition	4.3%	1.3%	0.3%	10.5%
	Density [veh/km]	0.3	0.1	0.0	1.6
Motorcycle	Speed [km/h]	79.7	14.9	18.8	134.1
	Flow [veh]	11.9	5.0	1	39
	% Composition	2.9%	1.1%	0.2%	10.3%
	Density [veh/km]	0.2	0.1	0.0	0.8

Table 4.3: Descriptive Statistics of AVI gate AC-09.

AVI AC-09					
		Average	**Std. Dev.**	**Minimum**	**Maximum**
Light	Speed [km/h]	51.6	17.9	10.8	95.1
	Flow [veh]	379.0	54.3	6	510
	% Composition	93.4%	1.9%	46.2%	99.7%
	Density [veh/km]	8.2	2.7	0.1	17.1
Heavy	Speed [km/h]	50.9	16.0	7.0	104.2
	Flow [veh]	15.2	5.9	1	42
	% Composition	3.7%	1.3%	0.3%	46.2%
	Density [veh/km]	0.3	0.2	0.0	1.0
Motorcycle	Speed [km/h]	58.1	16.0	13.9	117.4
	Flow [veh]	11.8	5.1	1	42
	% Composition	2.9%	1.5%	0.2%	10.6%
	Density [veh/km]	0.2	0.1	0.0	0.9

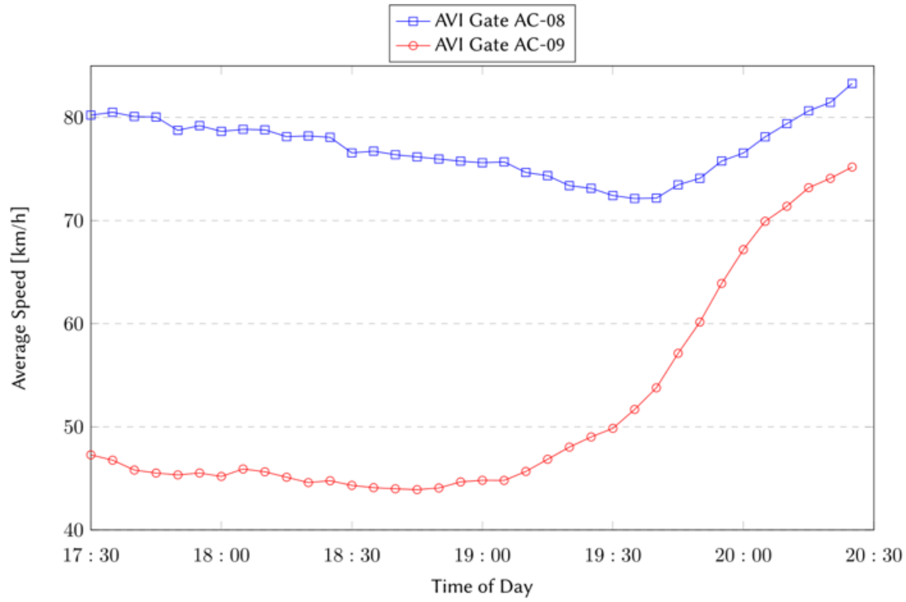


Figure 4.3: Evolutions of average speed of light vehicles in the studied period.

that heavy vehicles have lower fares in the parallel west section of the freeway (General Velásquez; see Figure 4.1). The average participation of heavy vehicles in General Velásquez is around 11% for the same period and the equivalent section, much higher than the 4% showed here.

We also see that the average speed in the AVI gate AC-09 is much lower than the one in the AVI gate AC-08 (for all types of vehicles). This is probably due to the existence of an exit ramp just ahead (less than 200 meters) of the AVI gate AC-09, which connects nicely with Avenida Libertador Bernardo O'Higgins, the main avenue of Santiago. Thus, near this AVI gate at rush time, one of the lanes is nearly blocked by the vehicles leaving the freeway, leaving only two tracks for the rest of the traffic.

Finally, the average flow of light vehicles is higher in the AVI gate AC-08 than in the AVI gate AC-09, which is explained by the same remark made before: there exists two entry ramps between these AVI gates, and this area, located in the center of the city, is a very congested one in the rush time, thus we have many vehicles entering in this section of the freeway that aim to travel to the suburbs.

In Figure 4.4, the distribution of the accidents recorded over the studied period is provided. We can see that the most dangerous time interval is from 7.30p to 8.00p with 11 accidents. That represents 28% of our accident information. We can also note that this sub period coincides with the lowest average speed of light vehicles recorded in AVI gate AC-08, and it has some of the highest average speeds in AVI gate AC-09 (Figure 4.3).

4.3 Variable selection

Having access to a large data set is, undoubtedly, a plus in our goal to predict accidents. But it also brings in the problem of variable selection. Directly including a large number of variables in

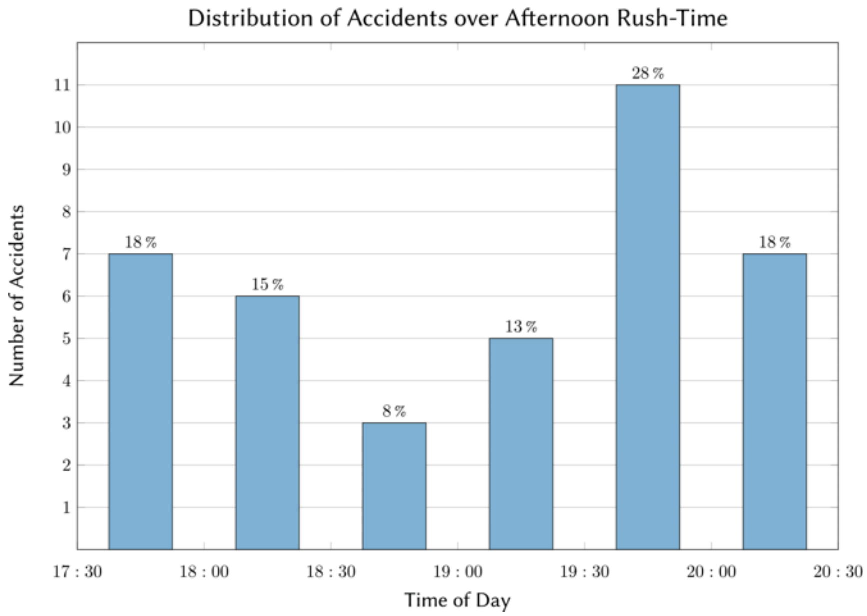


Figure 4.4: Distribution of accidents over afternoon rush time.

a classification or regression model may cause over adjustment of the model (Sawalha and Sayed, 2006) which, in turn, may affect both the interpretation of the interrelation between variables and, more importantly, the use of the model in prediction phase. It thus becomes important to analyze our data in order to identify what are the variables that appear to be strong precursors of car accidents. In order to do this, Pearson correlations, Random Forest techniques (also used by Ahmed and Abdel-Aty (2012)) and graphical analyses are used.

The Pearson correlation ρ_{XY} was computed for each pair of variables, in order to test for linear dependence; see Figure 4.5. We discarded variables with $|\rho_{XY}| > 0.95$ to avoid multicollinearity issues. For instance, this procedure removed Den.Light.08 since it is highly correlated to Speed.Light.08. In the same way, the variables Composition.Light.08 and Composition.Light.09 are discarded because they are correlated to Composition.Heavy.08 + Composition.Bike.09 and Composition.Heavy.08 + Composition.Bike.09, respectively.

We not turn to the Random forest (RF) procedure. RF was used in our study because it is a widely used method to determine variable importance (Lin et al., 2015). In the crash prediction context, this technique was also applied in similar fashion by Abdel-Aty et al. (2008), Ahmed and Abdel-Aty (2012) ,Xu et al. (2013).

RF is a machine learning classification method composed by a collection of decision trees. RF classifies an entry in the class which has been assigned most times by the trees (Breiman, 2001). The construction of each tree of the RF is made through two random processes. First, a random sample with replacement of cases is performed, which serves to grow the tree. Second, a sample is selected among all the variables, which is then used to split the nodes. The unused data is called out-of-bag (OOB) data. The OOB data could be used to determine an unbiased estimation of classification error.

In this Chapter, the RF is used to estimate each variable's importance. The importance of a variable in a decision tree is estimated in its ability to reduce an impurity index of nodes when used as a

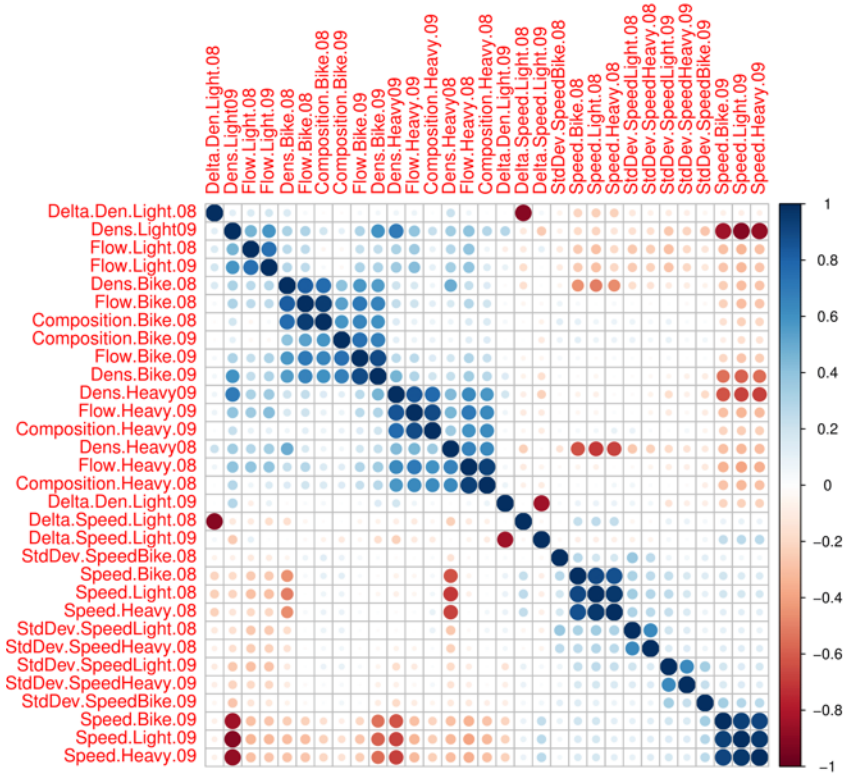


Figure 4.5: Correlation matrix.

split variable. We use the Gini index as a measure of impurity. For a binary tree (i.e. with two classes as is the case in this study: accident/non accident), the Gini impurity index (Breiman et al., 1984) is defined for node as:

$$i(t) = 2p(1/t)p(2/t) \tag{4.1}$$

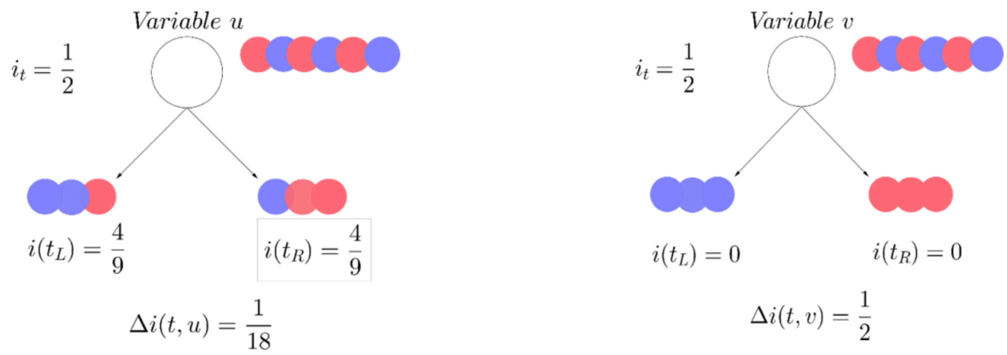
where $p(i/t)$ is the probability of case in class i given node t.

Then, after splitting the tree using the variable u, the decrease in impurity is defined as:

$$\Delta i(t,u) = i(t) - \frac{N_L}{N} i(t_L) - \frac{N_R}{N} i(t_R) \tag{4.2}$$

where N_L and N_R are the number of observations falling into the left and right children of the split, respectively, while $N = N_l + N_R$ is the total number of observations. $i(t_L)$ and $i(t_R)$ are the Gini's impurity index for the left and right children.

Thus, the larger the value of $\Delta i(t,u)$, the more important variable u is. Our RF has 500 trees which include 4 variables randomly chosen. Then, the average decrease in impurity is computed for those 500 trees for each variable. The results are shown in Figure 4.7. We observe that the most important variables are related mainly to the light class, something reasonable given the high proportion of the light class compared to others ones (see Tables 3.3 and 4.3).



(a) Case 1, Variable u. (b) Case 2, Variable v.

Figure 4.6: Example of two splittings, with case 2 (variable V) preferred.

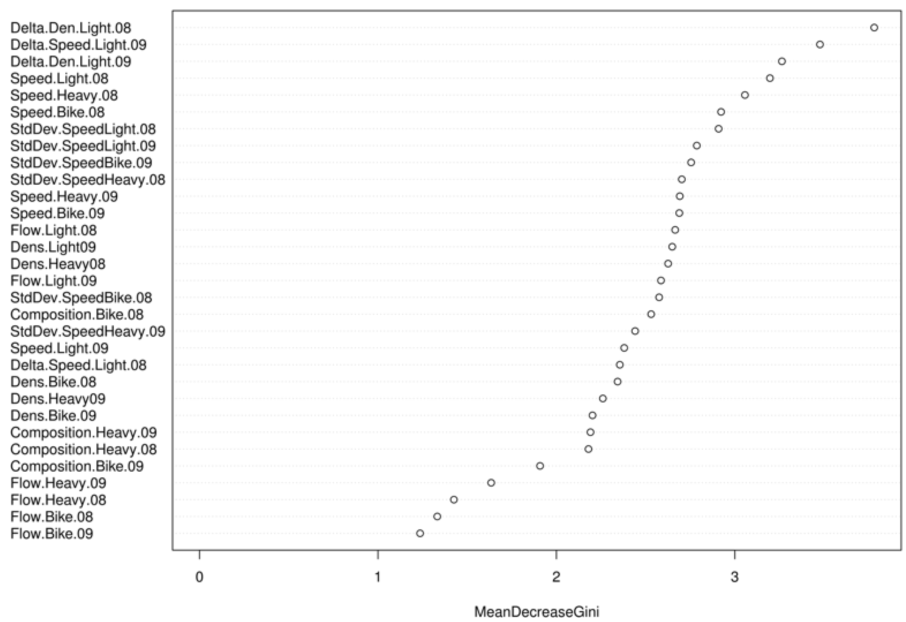


Figure 4.7: Change in Gini impurity index to determine variable importance.

Finally, a graphical analysis was conducted to determine accidents precursors. To do so, we compare the behavior of the mean of each variable around the time of the accident. Our most important findings are (see Figure 4.8 and Figure 4.9) **that the global minimum for the variables Delta.Den.Light.09 and Speed.Light.08 are attained just 5 minutes before the accident**. As these two variables are in the top 5 of Gini's index, we conjectured that these variables would be highly relevant in the classification models and became our starting point when calibrating.

A final point to make is that the RF procedure help to assess the importance of one variable at the time. It does not help, however, to assess whether non-linear combinations of variables help to further separate 0s from 1s or not. Non-linear specifications of the classification models are tested below.

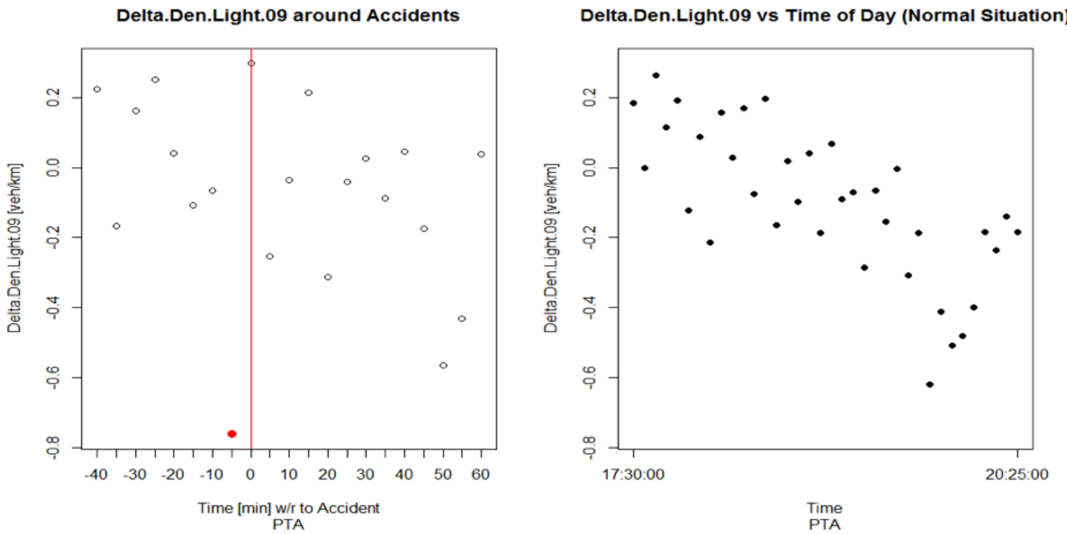


Figure 4.8: Delta.Den.Light.09 behavior prior and after an accident.

4.4 Classification method: Support vector machines

Support Vector Machine (SVM) could be used to solve binary classification problem. We first provide some theoretical background on how SVM works, then use it on our data set. SVM seek to find a separator hyperplane $f(x) = wx + b$ between two classes in order to maximize the distance between the classes and the decision frontier. Given linear separable data $(x_1, y_1), ..., (x_n, y_n)$ $y_i \in \{-1, 1\}, x_i \in \mathbb{R}^i, i = 1...n$, one and only one of the following statement holds:

$$x_i \cdot w + b \geq 1 \qquad for\ \ y_i = 1 \tag{4.3}$$

$$x_i \cdot w + b \leq -1 \qquad for\ \ y_i = -1 \tag{4.4}$$

We can combine the above statements into only one as follows:

$$y_i(x_i \cdot w + b) \geq 1 \qquad \forall i \tag{4.5}$$

It is possible to prove (Cortes and Vapnik, 1995) that the w vector which maximizes the margin must be the solution of the following non-linear optimization problem:

$$\begin{aligned}&\min \tfrac{1}{2}||w||^2\\ &\text{s.t.}\\ &y_i(x_i \cdot w + b) \geq 1 \qquad i = 1, ..., n\end{aligned}$$

If data is not linear separable, it is possible to add slack variables ξ_i to penalize misclassification. Cortes and Vapnik (1995) proposed the following SVM optimization problem:

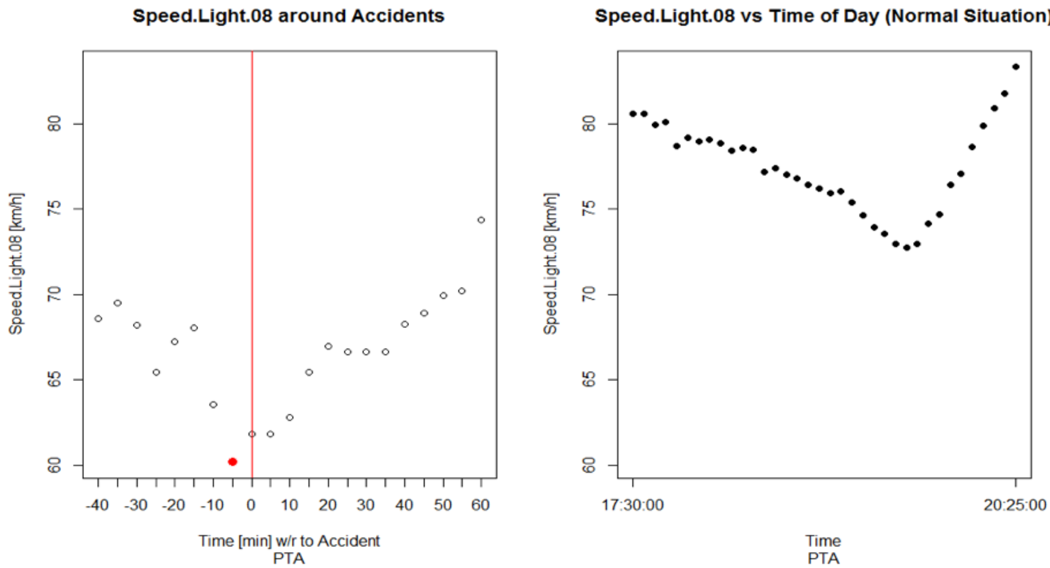


Figure 4.9: Speed.Light.08 behavior prior and after an accident.

$$
\begin{aligned}
&\min \tfrac{1}{2}||w||^2 + C\textstyle\sum_i \xi_i \\
&\text{s.t.} \\
&y_i(x_i \cdot w + b) \geq 1 - \xi_i \qquad i = 1, ..., n \\
&\xi_i \geq 0 \qquad i = 1, ..., n
\end{aligned}
$$

If we introduce the KKT multipliers, the SVM optimization problem can be stated as follows:

$$
\begin{aligned}
&\max \ \textstyle\sum_i a_i - \tfrac{1}{2}\sum_i \sum_j a_i a_j y_i y_j x_i \cdot x_j \\
&\text{s.t.} \\
&\textstyle\sum_i a_i y_i = 0 \\
&0 \leq a_i \leq C \quad \forall i
\end{aligned}
$$

If the decision function is nonlinear, it is possible to map the data to another Euclidean space H through a function Φ. Note that in the dual formulation, the data appears only as product $x_i \cdot x_j$.The mapping to the Euclidean space H could be done by computing the kernel function K which represents the dot product in $H : K(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$: (Friedman et al., 2001). In this Chapter we used the following classical kernels:

- Radial Kernel: $K(x_i, x_j) = exp(-\gamma||x_i - x_j||^2)$
- Polinomial Kernel: $K(x_i, x_j) = (\gamma x_i \cdot x_j + 1)^q$, with $q = 3$
- Sigmoid Kernel: $K(x_i, x_j) = tanh(\gamma x_i \cdot x_j + 1)$

As discussed above, and as can be seen from the actual data, accidents are rare events. This implies that SVM has to calculate the best separating hyperplane with a large number of observations in one class, and a very small number of observations on the other. This has proved to be troublesome for SVM as it ends up providing poor predictions (Akbani et al., 2004). To overcome this problem in the calibration phase, we use the Synthetic Minority Over-sampling Technique (SMOTE). This technique introduced by Chawla et al. (2002) sub-samples the majority class and over-samples the minority class. To do the latter, synthetic examples of the minority class are created. These examples are randomly introduced among the minority class and some of their closest k-neighbors. Some adjustment could be done in order to tune the proportions of both classes. It is very important to point out this artificially more balanced data is only used to calibrate the model; to test the predictive performance the full, original, unbalanced data is used. Previous articles have shown the benefits of using the SMOTE-SVM combined procedure which usually outperforms other classification methods for unbalanced data set (Drosou et al. (2014);Fergani et al. (2016)).

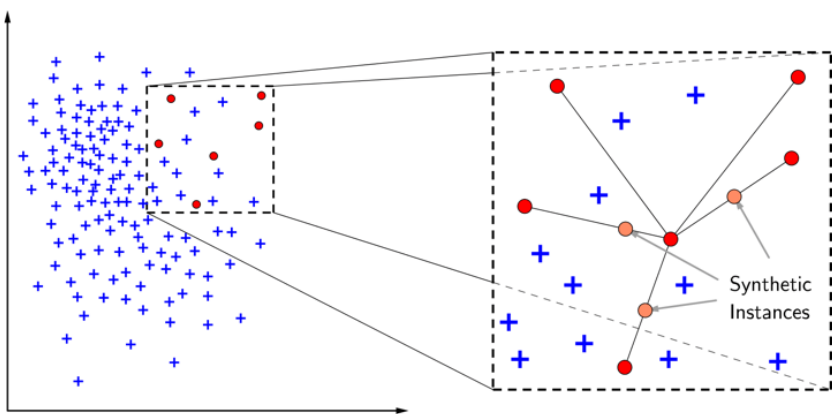


Figure 4.10: Oversampling using SMOTE.

Following to the previous Section, we tried many different specifications using from the most relevant to less relevant variables, according to the RF analysis. As explained, in this section only the first 80% of the data set is used for calibration purposes, leaving the remaining 20% for validation. This ensures that we validate on data that was not used for training or calibrating the model, thus really placing stress on the model capabilities to predict. The 20% corresponds to an average of 7.8 accidents per validation sample. We think this is a number large enough to test the quality of our model. We also tried with 70% vs 30% cross validation ratios, obtaining very similar results.

The calibration data was adjusted through the SMOTE technique, varying some key parameters (the kernel type and the gamma parameter). Regarding variables, we ended up finding that the best specifications had Speed.Light.08 and Delta.Den.Light.09 as main variables. A specification including, additionally, Speed.Light.09 worked well also. In Table 4.4 the best models are shown, using the predictive performance on the full training data (i.e. not adjusted with SMOTE), as a model adjustment metric.

Arguably, the best model is the one that used a radial kernel for the SMOTE procedure. Calibrating without adjusting the data set with SMOTE proved to deliver much worse results. The poor performance of SVM for unbalanced datasets has been, as discussed before, documented in the literature (Akbani et al., 2004). Figure 4.11 shows that the decision frontier is almost a straight line, something that also happens for the sigmoid and polynomial kernels.

In Figure 4.12 we show the SVM decision frontier and the (unbalanced) validation data set: the last 20% of the data we were provided with. It clearly shows why the sensitivity rate was 100%: the frontier perfectly separates the accident events. Yet, as we have discussed before, the sensitivity rate may be "too good": the value may heavily depend on the specific partition. To see this in Figure 4.13 we show the SVM decision frontier and the full data set (100%). What this shows is

Table 4.4: Results of SVM models.

Kernel	Radial	Sigmoid	Polynomial (Grade 3)
gamma	0.001	0.001	1
cost	10	100	1
SMOTE.perc.over	500	500	500
SMOTE.perc.under	100	100	100
Variables	Speed.Light.08 Delta.Den.Light.09	Speed.Light.08 Delta.Den.Light.09	Speed.Light.08 Speed.Light.09 Delta.Den.Light.09
Sensitivity (%)	100	100	87.50
False Positives Rate (%)	20.17	28.56	20.86

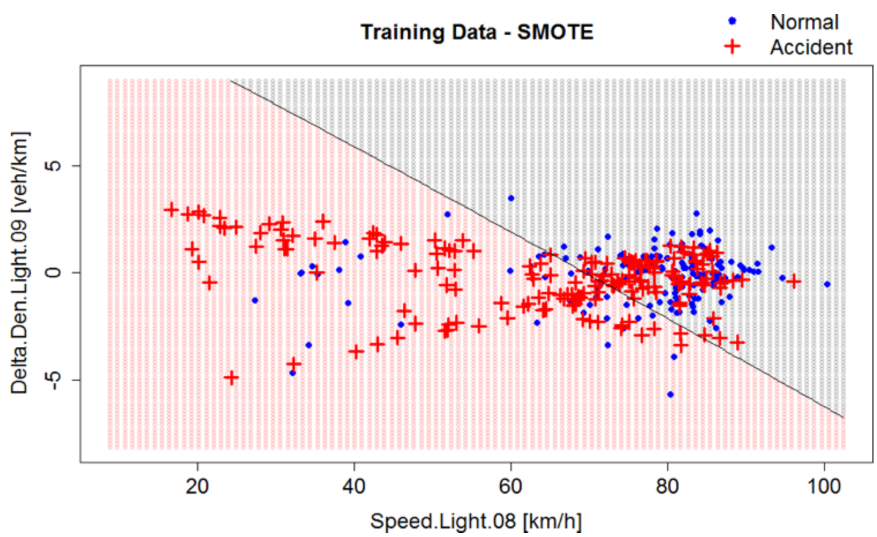


Figure 4.11: Decision frontier for SVM with radial kernel for the training SMOTE data-set.

that 8 accidents are above the frontier and would not be predicted. It just happened that none of those accident took place in the last 3 and half months of the data we received, thus not making it to the validation data set, and enabling a 100% sensitivity. The partition used was a good draw. It is also clear that if a different partition of the data was used the sensitivity would not be as high; in fact, in a really bad draw, the 10 points above the frontier would be all in the validation data set and the result would be very poor. This discussion shows the importance of repeating the calibration/validation process over many random partitions of the data, something we do in Section 4.6, and indicates that sensitivity values coming from calibration/validation on just one partition of the data should be taken with care.

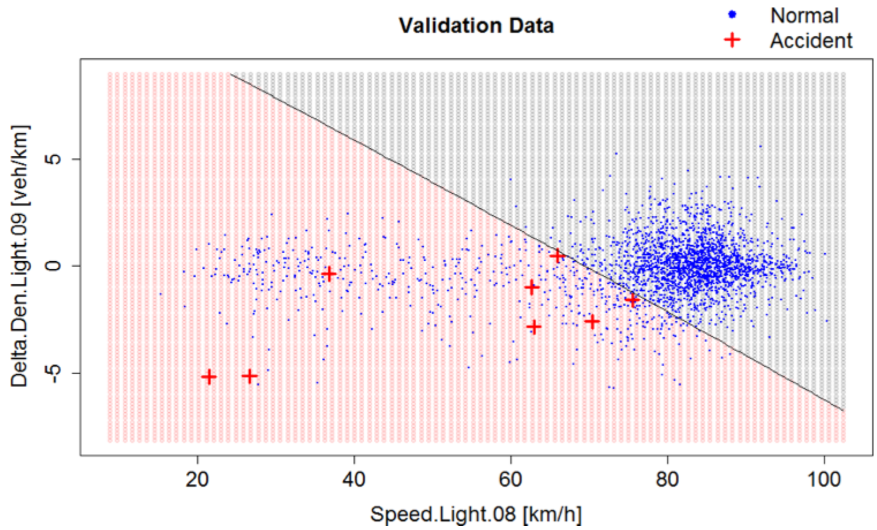


Figure 4.12: Decision frontier for SVM with radial kernel for the full validation data-set.

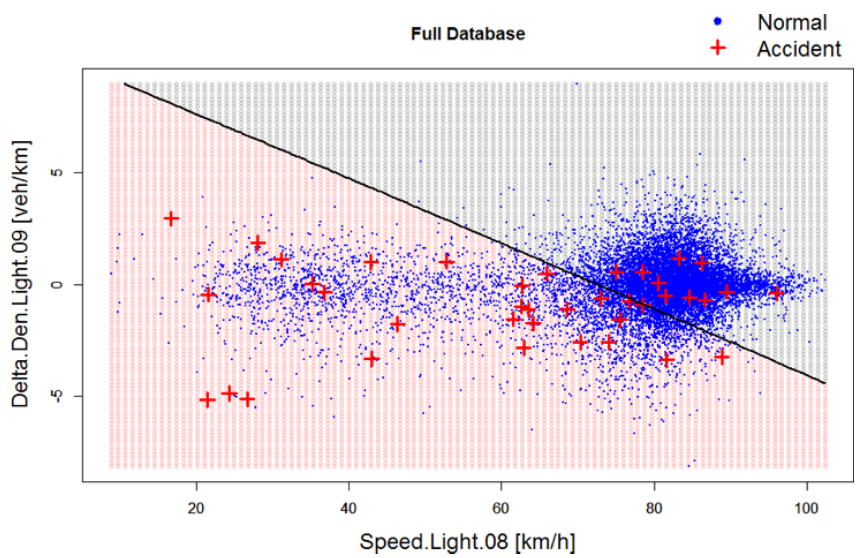


Figure 4.13: Decision frontier for SVM with radial kernel over the full data-set (100%).

4.5 Classification method: Logistic regression

We now explore a second classification method: the logistic regression. The previous Section showed that the SVM classifier frontier was quite similar to a line. Thus, the upside of using a logistic regression model is that one obtains parameters that may be easier to interpret opposed to SVM which is more of a black-box which requires multiple rules extraction (Martens et al., 2007).

The generalized linear models (of which the logistic regression is a particular case) aim to relax the restrictions given by the classical linear model

$$y = \beta_0 + \beta^T x + \varepsilon \tag{4.6}$$

which has to satisfy the Gauss-Markov assumptions in order to have the BLUE (best linear unbiased estimator) property for the ordinary least squares estimators. In particular, the error must be normally distributed with zero mean and has to satisfy the homoscedasticity property: $VAR(\varepsilon_i) = \sigma^2 < \infty$, i.e. constant variance.

As Hastie and Pregibon (1992) remarks, in some situations this is not appropriate. Generalized linear models deal with these problems by introducing a reparametrization to induce linearity and by allowing a non-constant variance (homoscedasticity violation) of the error. Specifically, GLM require:

- Link function, which describes how the mean depends on linear predictors $g(\mu) = \beta_0 + \beta^T x$.
- Variance function that captures how the variance of y depends upon the mean $VAR(y) = \Phi V(\mu)$, with Φ constant.

In our case, we consider $y = 1$ if an accident occurs in the next five minutes, and $y = 0$ in other case, so y$\sim$ Bernoulli (p). As link function between p and the independent variables x, we use the logit link function g given by:

$$g(p) = log(\frac{p}{1-p}) = \beta_0 + \beta^T x \tag{4.7}$$

$$p = \frac{1}{1 + exp(-\beta^t x)} \tag{4.8}$$

The parameters β and β_0 are then estimated by maximum likelihood. As in the SVM case, we tried many different specifications using from the most relevant to less relevant variables, starting with linear specifications. As in SVM, the best results were obtained when the vector x contained Speed.Light.08 and Delta.Den.Light.09. The results of the logistic regression are shown in Table 4.5. The significance tests indicate that the three parameters are different from 0 at 90% confidence level.

Table 4.5: Maximum likelihood parameters for the two variables logistic regression model.

Variable	Estimate	St. deviation	p-value
Intercept	-3.378	0.555	$1.15 \cdot 10^{-9}$
Speed.Light.08	-0.035	0.008	$1.24 \cdot 10^{-5}$
Delta.Den.Light.09	-0.214	0.119	0.073

To use this model in prediction phase we need to define a value p_o such that, when the model delivers a value of $p > p_0$, those traffic conditions are classified as leading to an accident in the next five minutes. In order to define p_0 we calculate the values of p for all observations in the estimation data set and choose p_0 so that the false alarm rate stands at about 20%. This lead to $p_0 = 0.299\%$, a value that may seem low to the reader yet, it leads to satisfactory results in terms of sensitivity, while keeping the false alarm rate at 22% over the training data-set. The low value is explained by the extreme unbalanced characteristic of the data set.

The decision frontier -a straight line indeed- is shown in Figure 4.14. Changing the threshold p_0 moves the decision frontier in parallel fashion. The tradeoff is as follows: If the straight line goes up, both the sensitivity and the false alarms increase. If the straight line goes down, both the sensitivity and the false alarms decrease. We attempted to estimate the Logistic regression model by adjusting the estimation data set using SMOTE, yet the results end up being worse, an interesting fact that shows that balancing data sets, through SMOTE or match-control may not always be the best course of action.

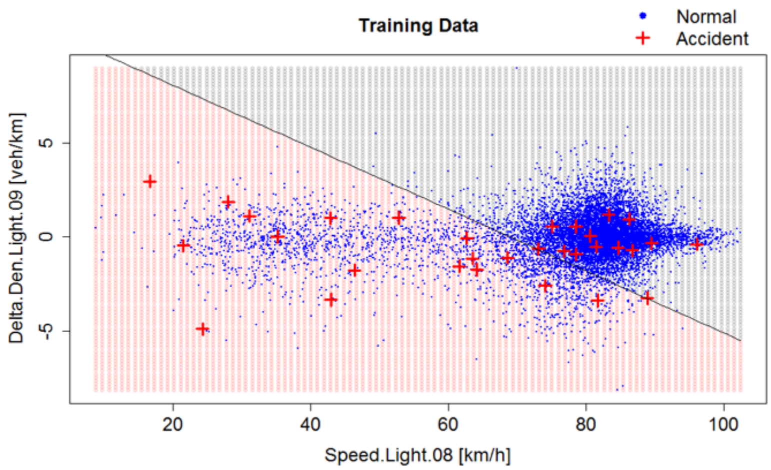


Figure 4.14: Decision frontier for the two variables logistic regression for the training data-set.

We now turn to the validation set, the last 20% of the data. The logistic model delivers, as in the SVM case a sensitivity of 100%, with a false alarm rate of 21.29%. The decision frontier and the validation data set are shown Figure 4.15.

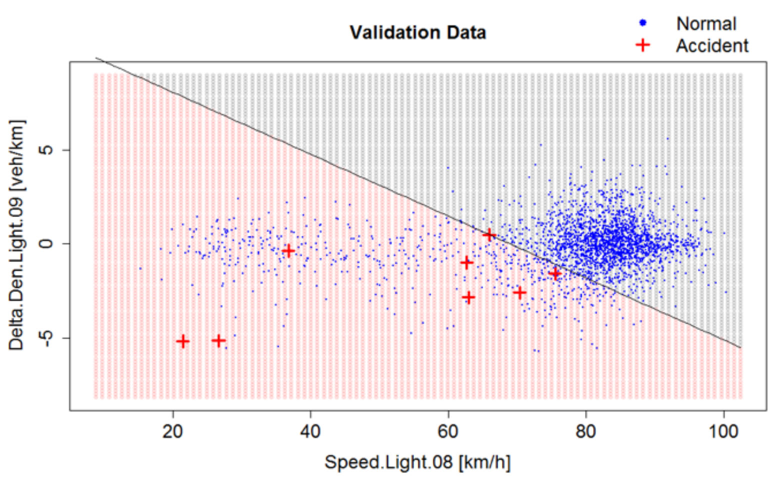


Figure 4.15: Decision frontier for the two variables logistic regression over the validation data-set.

Many observations can be made: first, the decision frontier is similar, yet not identical to the SVM one. Second, the 100% sensitivity is explained by the same reasons as in SVM: the partition that uses the first 14.4 months for estimation and the remaining 3.6 months for validation is, by chance, a very good draw. Third, while SVM and the logistic model are quite similar, the latter has the advantage of providing us with an explicit function for the decision frontier, thus being easier to interpret.

In fact, we can now provide an interpretation of the actual process by which most accidents happen. As Figures 4.8 and 4.9, the parameters of the Logistic model, and Figures 4.14 and 4.15 show, accidents occur when, simultaneously, there is a dramatic, absolutely abnormal drop in density at AVI gate AC09 (upstream) while, downstream (AVI gate AC08), speeds are abnormally low. This means that, upstream, there were atypical congestion conditions that start to ease, leading to vehicles probably speeding more than usual in order to catch up. Those vehicles however, will ran into an abnormally low speed zone downstream, slower than what drivers are used. In a nutshell, the perfect storm occurs when vehicles that were trapped in heavier than usual congestion upstream, race to recover time lost but ran into an unexpected, atypically low speed zone downstream.

The interpretation of the process above hints us that not only the dramatic drop in density at the upstream AVI gate AC09 may be of interest but also the speed. Including that variable linearly, however did not improved the model so we attempted non-linear specifications. After several attempts we obtained better estimation using the processed variables $Speed.Light.08^2$ and $Delta.Den.Light.09 \cdot Speed.Light.09^2$. The estimated parameters for this model are presented in Table 4.6.

Table 4.6: Maximum likelihood parameters for the two variables logistic regression model.

Variable	Estimate	St. deviation	p-value
Intercept	-4.287	0.393	$<2 \cdot 10^{-16}$
Speed.Light.08	$-2.99 \cdot 10^{-4}$	$7.26 \cdot 10^{-5}$	$3.89 \cdot 10^{-5}$
Delta.Den.Light.09 $\cdot$ Speed.Light.09^2	$-5.58 \cdot 10^{-5}$	$2.60 \cdot 10^{-5}$	0.032

From this, it can be seen that both variables (which are a function of three original variables) are significant at the 95% confidence level. The signs of the variables are intuitively correct: high speeds at the downstream gate decrease the likelihood of accidents, and at a more than a linear rate. On the other hand, negative Delta.Den.Light.09, that is, drops in density increase the likelihood of an accident (as found before) but, now, the effect is amplified by the square of the speed. In summary, the situation that causes the highest probability of accidents is: (i) substantially density drops upstream with ensuing high speeds (ii) unusual low speeds at gate AC08.

For this non-linear (in the variables) logistic model, the threshold probability is also set at $p_0 = 0.299\%$, corresponding to a false alarm rate of 21% over the estimation data set. Turning to validation/prediction, the model naturally achieved a sensitivity of 100% but decreased the false alarm rate to 20.17%., as shown in Figure 16. That the non-linear model achieves better results than the linear model can be seen by comparing Figures 14 and 16 below: the linear model had 10 accident events at the wrong side of the decision frontier, the non-linear model only six.

We also tried using a random-parameter logistic regression for both linear and non-linear logistic regression but the performance did not increase. Moreover, the likelihood ratio test suggests that the random-parameters logistic regression is not significantly better than the deterministic logistic regression. The estimated parameters of both models have the same sign and magnitude order. We also tried k-neighbors and CART but they ended up having considerably less predictive power than SVM and logistic regressions.

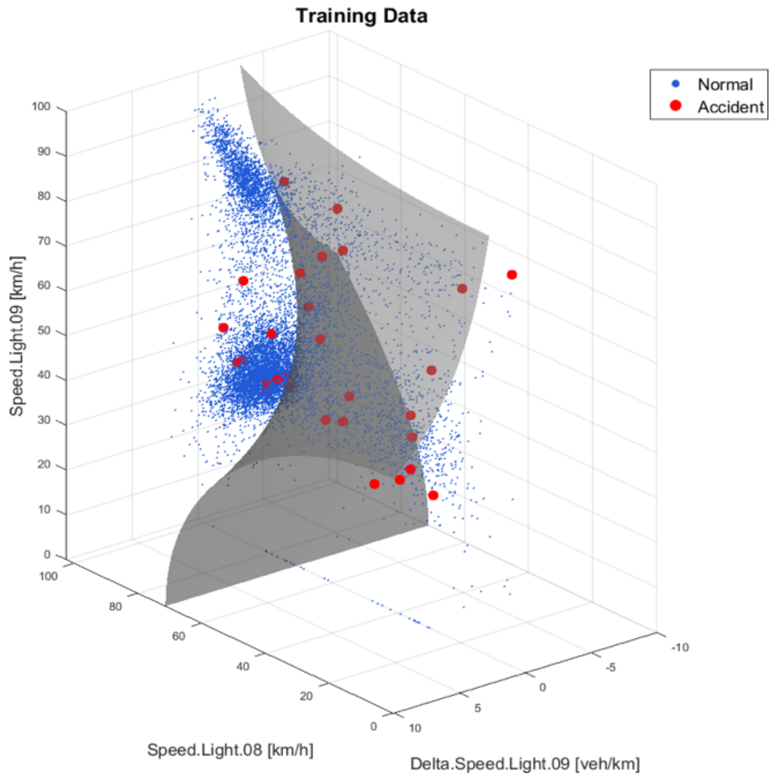


Figure 4.16: Decision frontier for the three variables logistic regression for the training data-set.

4.6 Robustness and model comparisons

We now turn to robustness. As clearly show, the sensitivity of the models may heavily depend on the partition of the data used so, what we did is to repeat 300 hundred times the following: we randomly select 80% of the data base, calibrate both SVM and logistic models and then validate using the remaining 20%. We then calculated 300 values for sensitivity and false positive rates and then calculated the averages, maximum, minimum and standard deviations. The number of repetitions used was based in our experimental results, which showed that the mean of the sensitivity and false alarm rate stabilizes when the number of repetitions was around 200, therefore 300 hundred repetitions were chosen to be on the safe side. Also, is important to remark that with this number of repetitions, the probability that two sub-samples are equal is indistinguishable from zero. The main results are shown in Table 4.7 and Figure 4.17.

Table 4.7 shows that SVM models have high sensitivity percentages, particularly that of a degree 3 polynomial kernel, which reaches a mean prediction of almost 80%, however they provide high false positive rates as they overestimate the zones where accidents should occur, probably due to the base balancing when using SMOTE. Yet if SMOTE is not used for training, SVM does not deliver good sensitivities. On the other hand, logistic regression models show false positive rates near to 20% adjusted on the estimation base; an expected behavior when using cross validation with random selection. In this last category, the non-linear logistic regression model shows the best results, with a mean sensitivity of 67.89% (similar to that obtained through SVM models with radial and sigmoid kernel), and a mean false positive rate of 20.94% (much lower than the same SVM models).

Table 4.7: Prediction power for adjusted models.

Indicators	SVM Radial Kernel	SVM Sigmoid Kernel	SVM Polinomial Kernel	Logistic Regression Linear	Logistic Regression Nonlinear
Mean Sensitivity (%)	69.06	68.50	77.13	62.26	67.89
Maximum Sensitivity (%)	100	100	87.50	100	100
Minimum Sensitivity (%)	53.50	54.64	52.06	45.63	59.17
Mean False Alarm Rate (%)	28.44	27.72	59.78	20.94	20.94
Standard Deviation (%)	1.62	0.88	7.92	0.10	0.07

These results are quite promising when compared to the literature, as shown on Table 4.8 -which draws from Lin et al. (2015). First, because we achieve high sensitivity values with low false alarm rates. More importantly, because our results are averages over 300 randomly selected calibration/estimation data sets and, therefore, we are positive that they are not conditioned by a particular partition of the data set. Also, because the 300 validations were done over non-balanced data sets.

Keeping in mind these differences, note that studies that reach higher sensitivity rates than those we obtain here have a false positive rate much higher than the 20% we achieve. On the other hand, when Ahmed and Abdel-Aty (2012) projected a false alarm rate near to 20%, their sensitivity dropped to about 60% according to the ROC curve of this article. On the other hand, the high sensitivity percentage found by Sun and Sun (2015) has the methodological disadvantage of coming from an artificially balanced base, where for each accident record, only 5 normal situation records where selected, to later use this base for both training and validation purposes; therefore, the percentages shown does not necessarily reflect the predictive power it would actually have in real world situations.

A final robustness check worth of study, is whether the calibrated model would perform well on data collected after the period used for calibration. This would serve two purposes: first, it is what comes closer to learn what would have been the result should a real-time model been at work. Second, it enables a look to how far or close in the future are the models able to predict. Hence, we used the non-linear logistic regression model on traffic and crash data that was collected by Autopista Central on a period of time later than the one we had at hand (June and July, 2016). The results are encouraging: the sensitivity was actually better than the mean, reaching a sensitivity of 75.03%, while keeping the false alarm rate at 22.47%.

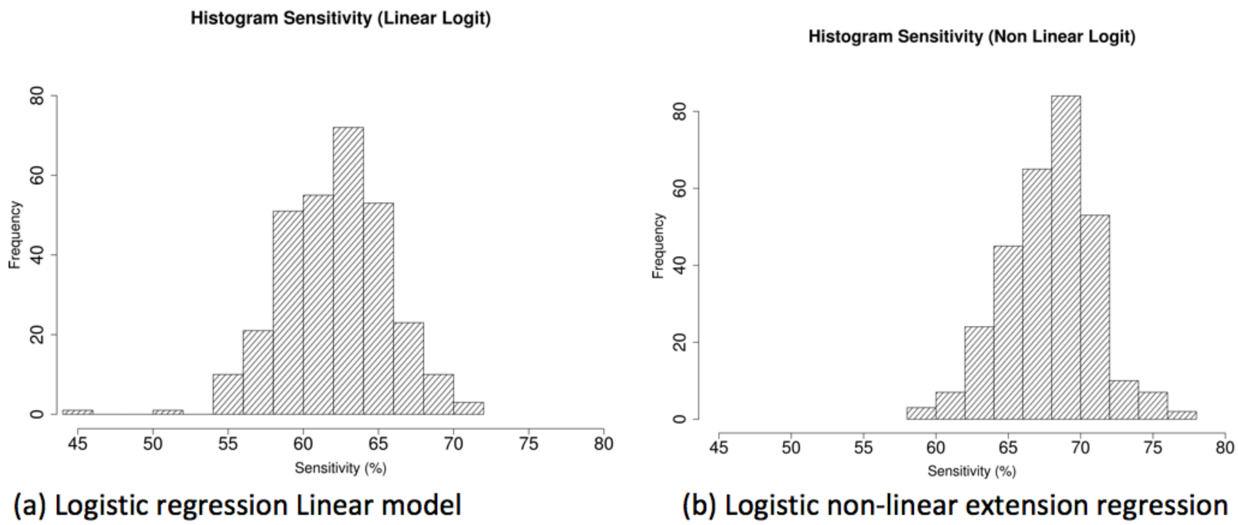


Figure 4.17: Sensitivity histogram (300 repetitions with 5-fold CV) in logistic regression.

4.7 Concluding remarks

Road accidents imply congestion, delays and sometimes loss of human life. That is why in the last two decades researchers have tried to stablish relations between crashes, flow states and environmental variables. Even though loop detectors (electromagnetic dispositive flow data collectors) exist since early 60's, predictive models for crash prediction appeared only in the early 2000's. These models showed that predicting road accidents is possible, but the lack of online data and the high failure rate of loop detectors have truncated the construction of computational tools. Recently, Automatic Vehicle Identification gates have been introduced in some urban expressways, such as Autopista Central in Chile. AVIs have almost no failures (less than 1%) and they are sometimes able to distinguish among vehicles classes such as Cars and SUVs, Buses and Trucks, and Motorcycles. In this study, techniques based on machine learning and logistic regression models to classify and forecast accidents on a stretch of the Autopista Central in Santiago are introduced. To the best of authors' knowledge, this Chapter is the first in making predictions based on disaggregated variables per type of vehicle using AVIs information. This allows isolating the contribution of each class to the increase the probability of accidents. Moreover, this Chapter is also the first to use non-artificially balanced data to validate the predictive models, which is quite important in order to think in a real-time application tool, and the first, as far as we know, to use repetitions to randomly select the calibration/estimation data set, in order to ensure robustness. The procedure described in this Chapter is as follows. First, we defined a stretch in Autopista Central to collect the flow data. The election was done in order to maximize the rate of accident per km per month. Second, we built a Random Forest model to classify the importance of the available variables. We complement it with visual inspection which permitted to identify the main explanatory variables of accident occurrence. Using these, SVM and logistic regression models were adjusted using the first 80% of the available data. Then the models were validated using the last 20%. The best models (radial SVM and non-linear extension logistic regression) predicted the 100% of the accidents with a relatively low false alarm rate of 20% approximately. To prove the robustness of our approach, we made 300 additional repetitions, randomly selecting the calibration/estimation data set and keeping

Table 4.8: Prediction power for previous research.

Authors	Variable selection method	Classification method	Sensitivity (%)	False Alarm Rate (%)
Abdel-Aty et al. (2004)	N/A	Logistic regression	69	N/A
Pande and Abdel-Aty (2006)	Classification tree	Neural Network	57.14	28.83
Abdel-Aty et al. (2008)	Random forest	Neural Network	61	21
Hossain and Muromachi (2012)	Random multinomial logit	Bayesian Network	66	20
Ahmed and Abdel-Aty (2012)	Random forest	Matched case-control method	68	46
Lin et al. (2015)	Frequent Patern tree	Bayesian Network	61.11	38.16
Sun and Sun (2015)	N/A	Dynamic Bayesian network	76.4	23.7

the remaining for validation. We trained and estimated our models in each instance, thus obtaining robust average sensitivity and failure rates.

The main conclusions of this Chapter are:

1. In the studied stretch, the selected modeling variables are related only to vehicles in the ”Car and pickup truck” category, which is directly related to the central location of this stretch, and its intrinsically urban nature. This is reflected on the traffic composition: 93.1% of the vehicles registered in the period studied belong to the ”Car and pickup trucks” category (light). This means that, for this stretch, vehicle composition variables were not relevant, contrary to what one may have conjectured, a finding on itself. Yet, in extensions of the study to more rural stretches, we have find that variables related to the rest of the types of vehicles (particularly regarding the interaction of motor bikes and trucks, and their speed differences) are indeed of first order importance, showing the advantages of using disaggregate data. Moreover, our preliminary work in this stretch, which also has more distant traffic counters, has shown results as positive as the ones described in the Chapter.
2. SVM models reach a high percent of sensitivity, but tend to overestimate the ”accident” prediction zone, prompting high rates of false positives, much higher than the 20% sought a priori. This can be caused by the Synthetic Minority Over-sampling Technique (SMOTE)

we used to balance the data. Yet, without SMOTE, sensitivity itself drops.

3. The non-linear logistic model reaches, at validation, a mean sensitivity of 67.89% with just 20.94% of false positives. This sensitivity is comparable to the best results obtained in contemporary literature although their failure rates are usually higher. The comparison though is not really fair to our model, as we did not use a specific partition of data but used 300 random ones, and we validated on actual data and not artificially balanced data.
4. From this same model (non-linear logistic), the situation where accidents are most likely to occur is identified: In summary, the situation that causes the highest probability of accidents is: (i) substantially density drops upstream with ensuing high speeds (ii) unusual low speeds downstream. This concurs with empirical intuition and experience: a sudden, unusual traffic congestion prompts, once it starts to dissipate, a more aggressive behavior of drivers who try to recover their lost time by speeding yet, a couple of kilometers down they face unusual low speeds, causing braking maneuvers that can lead to crashes.

We believe that our results are promising and that studying the rest of the expressway stretches and in other hours is warranted. We expect that due to differences in traffic geometry, length, and vehicle composition on the different stretches and periods, models will change, both in terms of variables as in terms of specifications. From a methodological point of view, we would like to stress the importance of validating using the original data set and using more than one data partition to ensure robustness.

We conclude by highlighting what we consider is an important avenue of future research: the matter of which preemptive actions could be taken when an accident is predicted. This is indeed crucial, yet escapes the scope of this Chapter. Possibly, providing a reasonable response to this question will require the work of a multidisciplinary team composed by psychologist, occupational safety, health experts and engineers to decide the best measures to prevent the accidents predicted.

Chapter 5

Final comments

In this thesis, we have studied collaborative systems in logistics and transportation. Chapters 2 and 3 are mainly related to horizontal collaboration in freight transportation and logistics, meanwhile, Chapter 4 is related to urban private transportation.

Collaborative systems have shown multiple benefits in transportation, even though real applications of these models are scarce. To tackle this drawback, in this thesis we have incorporated several features usually ignored in collaborative transportation. Particularly, in Chapter 2, we have analyzed the impact of practical issues in the real implementation of collaborative logistics. To so, we determined 16 practical obstacles arising when an implementation of horizontal collaboration occurs. We have learned that even though the majority of the papers tackle the coalition formation and cost allocation problems, real-applications usually fail due to behavior issues such as trust and coordination mechanism. In Chapter 3, we have provided another explanation for the lack of real implementation of horizontal collaborative transportation. The logic we pursued is as follows. Previous efforts have assumed both supply and demand are fixed so the problem was purely operational and no competition occurs. In Chapter 3, we have included competition after the cooperation agreements are signed. We have analyzed the Cournot-Nash equilibriums and proposed several coalition formation models based on private and social welfare subject to stability and antitrust constraints. The main conclusion of our proposed approach is that collaboration is not always advantageous. Moreover, we concluded that forming coalition is quite difficult. In fact, our results reveal that collaboration among all companies is usually outperformed by a partition of them into smaller sets, which is remarkable as most related literature has assumed that the grand coalition forms. Finally, in Chapter 4, we have analyzed a collaborative system in which the users of an expressway cooperate by sharing position and speed information. Using this, we build several predictive models in order to determine when an accident is more likely to happen. Opposed to previous studies, we used the full data-base for both calibration and validation purposes. Thus, our models are tested in a real environment in the path of a real implementable tool. This test is what comes close to learn what would have been the result should a real-time model been working. If a system like this is available for the managers of the expressway, the users of the expressway will benefit from the cooperation by facing fewer accidents rates. Of course, this will happen if and only if the preventive measures are taken.

Overall, this thesis seeks to diminish the gap between the models and the applications in the collaborative systems in logistics and transportation.

Bibliography

Abdel-Aty, M., Pande, A., Das, A., and Knibbe, W. (2008). Assessing safety on dutch freeways with data from infrastructure-based intelligent transportation systems. *Transportation Research Record: Journal of the Transportation Research Board*, (2083):153–161.

Abdel-Aty, M., Uddin, N., Pande, A., Abdalla, F., and Hsia, L. (2004). Predicting freeway crashes from loop detector data by matched case-control logistic regression. *Transportation Research Record: Journal of the Transportation Research Board*, (1897):88–95.

Abdel-Aty, M. A., Hassan, H. M., Ahmed, M., and Al-Ghamdi, A. S. (2012). Real-time prediction of visibility related crashes. *Transportation research part C: emerging technologies*, 24:288–298.

Ahmed, M. M. and Abdel-Aty, M. A. (2012). The viability of using automatic vehicle identification data for real-time crash prediction. *IEEE Transactions on Intelligent Transportation Systems*, 13(2):459–468.

Akbani, R., Kwek, S., and Japkowicz, N. (2004). Applying support vector machines to imbalanced datasets. In *European conference on machine learning*, pages 39–50. Springer.

Audy, J.-F., D'Amours, S., and Rönnqvist, M. (2012a). An empirical study on coalition formation and cost/savings allocation. *International Journal of Production Economics*, 136(1):13–27.

Audy, J.-F., D'Amours, S., and Rousseau, L.-M. (2011). Cost allocation in the establishment of a collaborative transportation agreementan application in the furniture industry. *Journal of the Operational Research Society*, 62(6):960–970.

Audy, J.-F., Lehoux, N., D'Amours, S., and Rönnqvist, M. (2012b). A framework for an efficient implementation of logistics collaborations. *International transactions in operational research*, 19(5):633–657.

Aumann, R. J. (1961). The core of a cooperative game without side payments. *Transactions of the American Mathematical Society*, 98(3):539–552.

Aumann, R. J. and Dreze, J. H. (1974). Cooperative games with coalition structures. *International Journal of Game Theory*, 3:217–237.

Aumann, R. J. and Hart, S. (1992). *Handbook of game theory with economic applications*, volume 2. Elsevier.

Basso, L. J. (2008). Airport deregulation: Effects on pricing and capacity. *International Journal of Industrial Organization*, 26(4):1015–1031.

Bell, E. T. (1934). Exponential polynomials. *Annals of Mathematics*, pages 258–277.

Bleeke, J. and Ernst, D. (2002). Is your strategic alliance really a sale. *Strategy: critical perspectives on business and management*, 4:4.

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

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees, wadsworth international group, belmont, california (1984). *Google Scholar*.

Brueckner, J. K. (2001). The economics of international codesharing: an analysis of airline alliances. *international Journal of industrial organization*, 19(10):1475–1498.

Buijs, P. and Wortmann, J. (2014). Joint operational decision-making in collaborative transportation networks: the role of it. *Supply Chain Management: An International Journal*, 19(2):200–210.

Caputo, M. and Mininno, V. (1996). Internal, vertical and horizontal logistics integration in italian grocery distribution. *International Journal of Physical Distribution & Logistics Management*, 26(9):64–90.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357.

Chen, H. (2016). Combinatorial clock-proxy exchange for carrier collaboration in less than truck load transportation. *Transportation Research Part E: Logistics and Transportation Review*, 91:152–172.

Chen, Z. and Ross, T. W. (2003). Cooperating upstream while competing downstream: a theory of input joint ventures. *International Journal of Industrial Organization*, 21(3):381–397.

Chow, H. K., Choy, K. L., and Lee, W. (2007). A strategic knowledge-based planning system for freight forwarding industry. *Expert Systems with Applications*, 33(4):936–954.

Clifton, C., Iyer, A., Cho, R., Jiang, W., Kantarcıoğlu, M., and Vaidya, J. (2008). An approach to securely identifying beneficial collaboration in decentralized logistics systems. *Manufacturing & Service Operations Management*, 10(1):108–125.

Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3):273–297.

Cruijssen, F., Borm, P., Fleuren, H., and Hamers, H. (2010). Supplier-initiated outsourcing: A methodology to exploit synergy in transportation. *European Journal of Operational Research*, 207(2):763–774.

Cruijssen, F., Cools, M., and Dullaert, W. (2007). Horizontal cooperation in logistics: opportunities and impediments. *Transportation Research Part E: Logistics and Transportation Review*, 43(2):129–142.

Dahl, S. and Derigs, U. (2011). Cooperative planning in express carrier networksan empirical study on the effectiveness of a real-time decision support system. *Decision Support Systems*, 51(3):620–626.

Dai, B. and Chen, H. (2012). Profit allocation mechanisms for carrier collaboration in pickup and delivery service. *Computers & Industrial Engineering*, 62(2):633–643.

Dai, B. and Chen, H. (2015). Proportional egalitarian core solution for profit allocation games with an application to collaborative transportation planning. *European Journal of Industrial Engineering*, 9(1):53–76.

Dao, S. D., Abhary, K., and Marian, R. (2014). Optimisation of partner selection and collaborative transportation scheduling in virtual enterprises using ga. *Expert Systems with Applications*, 41(15):6701–6717.

Defryn, C., Sörensen, K., and Dullaert, W. (2017). Integrating partner objectives in horizontal logistics optimisation models. *Omega*.

Drosou, K., Georgiou, S., Koukouvinos, C., and Stylianou, S. (2014). Support vector machines classification on class imbalanced data: a case study with real medical data. *Journal of Data Science*, 12(4):143–155.

Du, G., Sun, C., and Weng, J. (2016). Liner shipping fleet deployment with sustainable collaborative transportation. *Sustainability*, 8(2):165.

Ergun, Ö., Gui, L., Heier Stamm, J. L., Keskinocak, P., and Swann, J. (2014). Improving humanitarian operations through technology-enabled collaboration. *Production and Operations Management*, 23(6):1002–1014.

Ergun, Ö., Kuyzu, G., and Savelsbergh, M. (2007). Shipper collaboration. *Computers & Operations Research*, 34(6):1551–1560.

Farrell, J. and Shapiro, C. (1990). Horizontal mergers: an equilibrium analysis. *The American Economic Review*, pages 107–126.

Fergani, B. et al. (2016). Comparing hmm, lda, svm and smote-svm algorithms in classifying human activities. In *Proceedings of the Mediterranean Conference on Information & Communication Technologies 2015*, pages 639–644. Springer.

Ferrer, J.-C., Mac Cawley, A., Maturana, S., Toloza, S., and Vera, J. (2008). An optimization approach for scheduling wine grape harvest operations. *International Journal of Production Economics*, 112(2):985–999.

Flisberg, P., Frisk, M., Rönnqvist, M., and Guajardo, M. (2015). Potential savings and cost allocations for forest fuel transportation in sweden: A country-wide study. *Energy*, 85:353–365.

Forsberg, M., Frisk, M., and Rönnqvisty, M. (2005). Flowopt–a decision support tool for strategic and tactical transportation planning in forestry. *International Journal of Forest Engineering*, 16(2):101–114.

Frayret, J., D'Amours, F., and D'Amours, S. (2003). Collaboration et outils collaboratifs pour la pme manufacturière [collaboration and collaborative tools for manufacturing smes]. Technical report, CEFRIO Technical Report, Canada.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York.

Frisk, M., Göthe-Lundgren, M., Jörnsten, K., and Rönnqvist, M. (2010). Cost allocation in collaborative forest transportation. *European Journal of Operational Research*, 205(2):448–458.

Garrette, B., Castañer, X., and Dussauge, P. (2009). Horizontal alliances as an alternative to autonomous production: Product expansion mode choice in the worldwide aircraft industry 1945–2000. *Strategic Management Journal*, 30(8):885–894.

Garrido, R. A., Lamas, P., and Pino, F. J. (2015). A stochastic programming approach for floods emergency logistics. *Transportation research part E: logistics and transportation review*, 75:18–31.

Ghaderi, H., Darestani, S. A., Leman, Z., and Ismail, M. Y. (2012). Horizontal collaboration in logistics: a feasible task for group purchasing. *International Journal of Procurement Management*, 5(1):43–54.

Ghosh, A. and Morita, H. (2012). Competitor collaboration and product distinctiveness. *International Journal of Industrial Organization*, 30(2):137–152.

Golob, T. F. and Recker, W. W. (2004). A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A: Policy and Practice*, 38(1):53–80.

Gonzalez-Feliu, J., Morana, J., Grau, J.-M. S., and Ma, T.-Y. (2013). Design and scenario assessment for collaborative logistics and freight transport systems. *International Journal of Transport Economics*.

Guajardo, M., Jörnsten, K., and Rönnqvist, M. (2016). Constructive and blocking power in collaborative transportation. *OR spectrum*, 38(1):25–50.

Guajardo, M. and Rönnqvist, M. (2015). Operations research models for coalition structure in collaborative logistics. *European Journal of Operational Research*, 240(1):147–159.

Guajardo, M. and Rönnqvist, M. (2016). A review on cost allocation methods in collaborative transportation. *International Transactions in Operational Research*, 23(3):371–392.

Head, K. and Ries, J. (1997). International mergers and welfare under decentralized competition policy. *Canadian Journal of Economics*, pages 1104–1123.

Hernández, S. and Peeta, S. (2014). A carrier collaboration problem for less-than-truckload carriers: characteristics and carrier collaboration model. *Transportmetrica A: Transport Science*, 10(4):327–349.

Hernández, S., Peeta, S., and Kalafatas, G. (2011). A less-than-truckload carrier collaboration planning problem under dynamic capacities. *Transportation Research Part E: Logistics and Transportation Review*, 47(6):933–946.

Hezarkhani, B. (2016). Pairwise mergers in bipartite matching games with an application in collaborative logistics. *Operations Research Letters*, 44(6):818–822.

Hingley, M., Lindgreen, A., Grant, D. B., and Kane, C. (2011). Using fourth-party logistics management to improve horizontal collaboration among grocery retailers. *Supply Chain Management: An International Journal*, 16(5):316–327.

Hossain, M. and Muromachi, Y. (2012). A bayesian network based framework for real-time crash prediction on the basic freeway segments of urban expressways. *Accident Analysis & Prevention*, 45:373–381.

Ilie-Zudor, E., Ekárt, A., Kemeny, Z., Buckingham, C., Welch, P., and Monostori, L. (2015). Advanced predictive-analysis-based decision support for collaborative logistics networks. *Supply Chain Management: An International Journal*, 20(4):369–388.

Jouida, S. B., Krichen, S., and Klibi, W. (2017). Coalition-formation problem for sourcing contract design in supply networks. *European Journal of Operational Research*, 257(2):539–558.

Kale, R., Evers, P. T., and Dresner, M. E. (2007). Analyzing private communities on internet-based collaborative transportation networks. *Transportation Research Part E: Logistics and Transportation Review*, 43(1):21–38.

Kanda, A., Deshmukh, S., et al. (2008). Supply chain coordination: perspectives, empirical studies and research directions. *International journal of production Economics*, 115(2):316–335.

Kwak, H.-C. and Kho, S. (2016). Predicting crash risk and identifying crash precursors on korean expressways using loop detector data. *Accident Analysis & Prevention*, 88:9–19.

Lai, M., Cai, X., and Hu, Q. (2017). An iterative auction for carrier collaboration in truckload pickup and delivery. *Transportation Research Part E: Logistics and Transportation Review*, 107:60–80.

Leitner, R., Meizer, F., Prochazka, M., and Sihn, W. (2011). Structural concepts for horizontal cooperation to increase efficiency in logistics. *CIRP Journal of Manufacturing Science and Technology*, 4(3):332–337.

Li, J., Cai, X., and Zeng, Y. (2016). Cost allocation for less-than-truckload collaboration among perishable product retailers. *OR spectrum*, 38(1):81–117.

Lin, L., Wang, Q., and Sadek, A. W. (2015). A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transportation Research Part C: Emerging Technologies*, 55:444–459.

Lozano, S., Moreno, P., Adenso-Díaz, B., and Algaba, E. (2013). Cooperative game theory approach to allocating benefits of horizontal cooperation. *European Journal of Operational Research*, 229(2):444–452.

Lv, Y., Tang, S., Zhao, H., and Li, S. (2009). Real-time highway accident prediction based on support vector machines. In *Control and Decision Conference, 2009. CCDC'09. Chinese*, pages 4403–4407. IEEE.

Lyons, J., Ritter, J., Thomas, K., Militello, L., and Vincent, P. (2006). Collaborative logistics: Developing a framework to evaluate socio-technical issues in logistic-based networks. In *Collaborative Technologies and Systems, 2006. CTS 2006. International Symposium on*, pages 208–214. IEEE.

Martens, D., Baesens, B., Van Gestel, T., and Vanthienen, J. (2007). Comprehensible credit scoring models using rule extraction from support vector machines. *European journal of operational research*, 183(3):1466–1476.

Martinez-Olvera, C. and Shunk, D. (2006). Comprehensive framework for the development of a supply chain strategy. *International Journal of Production Research*, 44(21):4511–4528.

Mason, R., Lalwani, C., and Boughton, R. (2007). Combining vertical and horizontal collaboration for transport optimisation. *Supply Chain Management: An International Journal*, 12(3):187–199.

Montoya-Torres, J. R., Muñoz-Villamizar, A., and Vega-Mejía, C. A. (2016). On the impact of collaborative strategies for goods delivery in city logistics. *Production Planning & Control*, 27(6):443–455.

Oum, T. H., Park, J.-H., Kim, K., and Yu, C. (2004). The effect of horizontal alliances on firm productivity and profitability: evidence from the global airline industry. *Journal of Business Research*, 57(8):844–853.

Özener, O. Ö. and Ergun, Ö. (2008). Allocating costs in a collaborative transportation procurement network. *Transportation Science*, 42(2):146–165.

Palander, T. and Väätäinen, J. (2005). Impacts of interenterprise collaboration and backhauling on wood procurement in finland. *Scandinavian Journal of Forest Research*, 20(2):177–183.

Pérez-Bernabeu, E., Juan, A. A., Faulin, J., and Barrios, B. B. (2015). Horizontal cooperation in road transportation: a case illustrating savings in distances and greenhouse gas emissions. *International Transactions in Operational Research*, 22(3):585–606.

Pesendorfer, M. (2000). A study of collusion in first-price auctions. *The Review of Economic Studies*, 67(3):381–411.

Pomponi, F., Fratocchi, L., and Rossi Tafuri, S. (2015). Trust development and horizontal collaboration in logistics: a theory based evolutionary framework. *Supply Chain Management: An International Journal*, 20(1):83–97.

Pomponi, F., Fratocchi, L., Tafuri, S. R., and Palumbo, M. (2013). Horizontal collaboration in logistics: a comprehensive framework. *Research in Logistics & Production*, 3.

Prakash, A. and Deshmukh, S. (2010). Horizontal collaboration in flexible supply chains: a simulation study. *Journal of Studies on Manufacturing*, 1(1):54–58.

Regan, A. and Song, J. (2003). An auction based collaborative carrier network. *Transportation Research Part E: Logistics and Transportation Review*, 2.

Rizzi, L. I. and Ortúzar, J. d. D. (2003). Stated preference in the valuation of interurban road safety. *Accident Analysis & Prevention*, 35(1):9–22.

Ross, T. W. and Winter, R. A. (2005). The efficiency defense in merger law: Economic foundations and recent canadian developments. *Antitrust Law Journal*, 72(2):471–503.

Rota, G.-C. (1964). The number of partitions of a set. *The American Mathematical Monthly*, 71(5):498–504.

Salant, S. W., Switzer, S., and Reynolds, R. J. (1983). Losses from horizontal merger: the effects of an exogenous change in industry structure on cournot-nash equilibrium. *The Quarterly Journal of Economics*, 98(2):185–199.

Sawalha, Z. and Sayed, T. (2006). Traffic accident modeling: some statistical issues. *Canadian Journal of Civil Engineering*, 33(9):1115–1124.

Schmeidler, D. (1969). The nucleolus of a characteristic function game. *SIAM Journal on Applied Mathematics*, 17(6):1163–1170.

Schulz, S. F. and Blecken, A. (2010). Horizontal cooperation in disaster relief logistics: benefits and impediments. *International Journal of Physical Distribution & Logistics Management*, 40(8/9):636–656.

Shapiro, C. and Willig, R. D. (1990). On the antitrust treatment of production joint ventures. *The Journal of Economic Perspectives*, 4(3):113–130.

Shapley, L. S. (1953). A value for n-person games. *Contributions to the Theory of Games*, 2(28):307–317.

Sheppard, E. J. and Seidman, D. (2001). Ocean shipping alliances: the wave of the future? *International Journal of Maritime Economics*, 3(4):351–367.

Shi, Q. and Abdel-Aty, M. (2015). Big data applications in real-time traffic operation and safety monitoring and improvement on urban expressways. *Transportation Research Part C: Emerging Technologies*, 58:380–394.

Shi, Q., Abdel-Aty, M., and Yu, R. (2016). Multi-level bayesian safety analysis with unprocessed automatic vehicle identification data for an urban expressway. *Accident Analysis & Prevention*, 88:68–76.

Simatupang, T. M. and Sridharan, R. (2005). An integrative framework for supply chain collaboration. *The International Journal of Logistics Management*, 16(2):257–274.

Stefansson, G. and Russell, D. M. (2008). Supply chain interfaces: defining attributes and attribute values for collaborative logistics management. *Journal of Business Logistics*, 29(1):347–359.

Sun, J. and Sun, J. (2015). A dynamic bayesian network model for real-time crash prediction using traffic speed conditions data. *Transportation Research Part C: Emerging Technologies*, 54:176–186.

Sun, L., Rangarajan, A., Karwan, M. H., and Pinto, J. M. (2015). Transportation cost allocation on a fixed route. *Computers & Industrial Engineering*, 83:61–73.

Suzuki, Y. and Lu, S.-H. (2017). Economies of product diversity in collaborative logistics. *Journal of Business Logistics*.

Theofilatos, A. and Yannis, G. (2014). A review of the effect of traffic and weather characteristics on road safety. *Accident Analysis & Prevention*, 72:244–256.

Theofilatos, A., Yannis, G., Kopelias, P., and Papadimitriou, F. (2016). Predicting road accidents: a rare-events modeling approach. *Transportation Research Procedia*, 14:3399–3405.

Tomkins, C. (2001). Interdependencies, trust and information in relationships, alliances and networks. *Accounting, organizations and society*, 26(2):161–191.

Triqui Sari, L. and Hennet, J.-C. (2016). Cooperative inventory planning in a distribution network. *International Journal of Production Research*, 54(19):5916–5931.

Vanovermeire, C. and Sörensen, K. (2014). Integration of the cost allocation in the optimization of collaborative bundling. *Transportation Research Part E: Logistics and Transportation Review*, 72:125–143.

Verdonck, L., Caris, A., Ramaekers, K., and Janssens, G. K. (2013). Collaborative logistics from the perspective of road transportation companies. *Transport Reviews*, 33(6):700–719.

Verstrepen, S., Cools, M., Cruijssen, F., and Dullaert, W. (2009). A dynamic framework for managing horizontal cooperation in logistics. *International Journal of Logistics Systems and Management*, 5(3-4):228–248.

Wang, X. and Kopfer, H. (2013). Dynamic collaborative transportation planning: A rolling horizon planning approach. In *International Conference on Computational Logistics*, pages 128–142. Springer.

Wang, X. and Kopfer, H. (2014). Collaborative transportation planning of less-than-truckload freight. *OR spectrum*, 36(2):357–380.

Wang, X. and Kopfer, H. (2015). Rolling horizon planning for a dynamic collaborative routing problem with full-truckload pickup and delivery requests. *Flexible Services and Manufacturing Journal*, 27(4):509–533.

Wang, Y., Ma, X., Liu, M., Gong, K., Liu, Y., Xu, M., and Wang, Y. (2017). Cooperation and profit allocation in two-echelon logistics joint distribution network optimization. *Applied Soft Computing*, 56:143–157.

Weng, K. and Xu, Z.-h. (2014). Flow merging and hub route optimization in collaborative transportation. *Journal of Applied Mathematics*, 2014.

Williamson, O. E. (1981). The economics of organization: The transaction cost approach. *American journal of sociology*, 87(3):548–577.

Xu, C., Wang, W., and Liu, P. (2013). A genetic programming model for real-time crash prediction on freeways. *IEEE Transactions on Intelligent Transportation Systems*, 14(2):574–586.

Xu, X.-F., Hao, J., Deng, Y.-R., and Wang, Y. (2017). Design optimization of resource combination for collaborative logistics network under uncertainty. *Applied Soft Computing*, 56:684–691.

Yu, R. and Abdel-Aty, M. (2013). Utilizing support vector machine in real-time crash risk evaluation. *Accident Analysis & Prevention*, 51:252–259.

Yu, R., Abdel-Aty, M. A., Ahmed, M. M., and Wang, X. (2014). Utilizing microscopic traffic and weather data to analyze real-time crash patterns in the context of active traffic management. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):205–213.

Zhang, M., Pratap, S., Huang, G. Q., and Zhao, Z. (2017). Optimal collaborative transportation service trading in b2b e-commerce logistics. *International Journal of Production Research*.