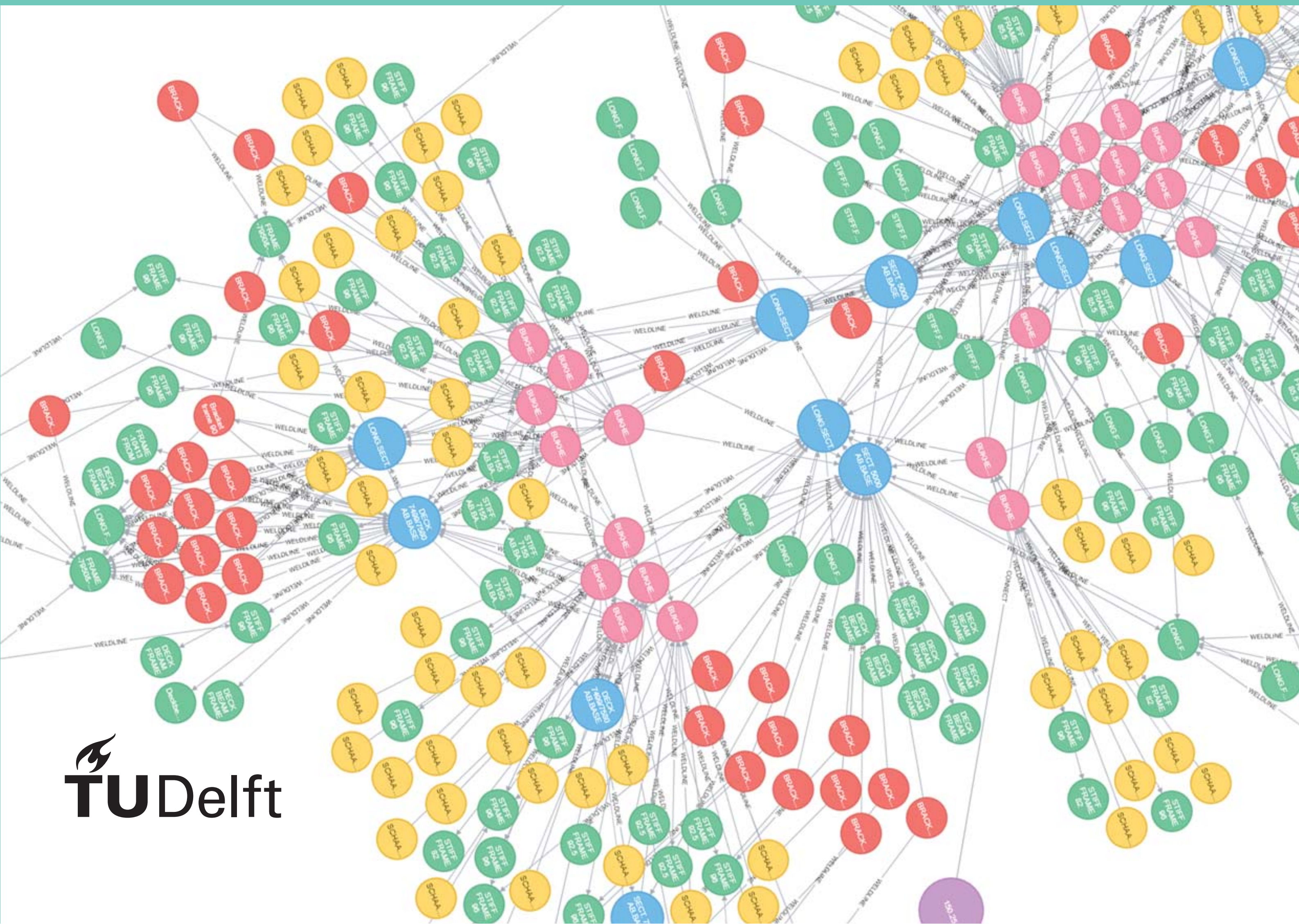
Case Based Reasoning

As a Cost Estimation Method for the Ship Building Industry

S. A. M. T. Visseren

Technische Universiteit Delft
&
Royal IHC

SDPO.17.036.m



TU Delft

Case Based Reasoning

As a Cost Estimation Method for the Ship Building Industry

by

S. A. M. T. Visseren

at Company

Royal IHC

To obtain the degree of Master of Science
in Marine Technology in the specialization Ship Production
at the Delft University of Technology,
to be defended publicly on Wednesday November 29, 2017 at 10:00 AM.

Student number:	4079329	
Project duration:	February 6, 2017 – November 29, 2017	
Thesis committee:	Dr. ir. R. G. Hekkenberg,	TU Delft
	Dr. ir. J. M. G. Coenen,	TU Delft, supervisor
	Ir. Ing. G. Alblas,	TU Delft, supervisor
	Dr. W. W. A. Beelaerts van Blokland,	TU Delft
	C. W. Zevenbergen MScBA	Royal IHC

This thesis is confidential and cannot be made public until 20-11-2022.

TU Delft

“Our differences are the real treasures.”
TEMITOPE IBRAHIM

Abstract

Accurate cost estimations are important for ship production companies such as Royal IHC. The research discussed is based on a feasibility study for a CBR cost estimation method. The current cost estimation method at IHC is a top-down price-to-win analogy cost estimation. This method uses mainly the weight of the ship as an input variable for the estimation. A problem with using only weight is that the actual work performed on a section is left out of the estimation as well as other cost drivers. Another problem occurring is the current estimation method being divided into three phases, with lacking evaluation, optimization and feedback possible. This research discusses two new cost estimation methods. The statistical method is a parametric cost estimation method, that uses equations with the main cost drivers. The graph database method uses the CBR theory as a basis for cost estimations. Problem solving is done by using a solution from an old ship and reusing this solution for a new ship. The graph database stores detailed information about the work performed on the ship and the main cost drivers. The qualitative and quantitative analyses showed that using multiple cost drivers and the work performed on a section creates more accurate cost estimations. With a solution as the graph database method also the possibilities of evaluation and process optimizations are broader.

Preface & Acknowledgments

Before you lies the research on Case Based Reasoning as a cost estimation method for the ship building industry. This research has been conducted in order to complete the master Marine Technology at the faculty Mechanical, Maritime and Materials Engineering of the TU Delft.

During the past 9 months I got the chance to perform this research at Royal IHC. I would like to thank the company because I learned so much during the past 9 months about the company and the ship building industry. I would like to thank T. Roos and C. W. Zevenbergen for being my daily supervisors at IHC and taking the time to help me with my questions or leading me to the right person. Furthermore I would like to thank all the people discussed in chapter 11 they were willing to share their knowledge of the production process and the cost estimation methods at Royal IHC. Also I would like to thank my other colleagues of the Planning and Control department at Royal IHC for their help during the past 9 months. Everybody was willing to share their knowledge, take me along the yard and to help me with my questions.

I would like to thank J. M. G. Coenen for her guidance and assistance during my research. Her feedback was really helpful and I could always call or plan a meeting if I had any questions regarding my research. I would also like to thank R. G. Hekkenberg for being the chair of my thesis committee and W. W. A. Beelearts van Blokland for being part of my thesis committee.

This research is performed as a feasibility study for the cost estimation method designed by G. Alblas for his PhD study. I would like to thank him for the opportunity to conduct this research and for his assistance and his time during my research. It was nice to collaborate with him on this research.

S. A. M. T. Visseren
Delft, November 2017

Summary

Royal IHC is a company specialized in the production of vessels or constructions for the offshore, dredging and mining industries. This research focusses on the production of vessels at the Kinderdijk yard of IHC. The production process starts with the preparation of the section building material. Then the material is welded together into sections and the pre-outfitting and the painting of the sections is performed. When the sections are ready they are assembled into block assemblies or directly assembled into the hull. This can be performed on a slipway or in a dry dock. At IHC they assemble the ships on a slipway. During the hull assembly the slipway outfitting is performed and when the hull assembly is done the hull is painted. Then the ship is launched and moored at a quay where the final outfitting is performed and the ship is delivered to the client.

This research focusses on a feasibility study of a case based reasoning cost estimation method: the graph database method. Two other cost estimation methods will be used in a comparison with the new method: the current IHC method and the statistical method. If the new cost estimation method proves to be feasible and a more accurate cost estimation can be made, this is beneficial for shipyard, Royal IHC.

The current cost estimation methods at Royal IHC are the VC cost estimation, the BVC cost estimation and the hour evaluation. The VC method estimates the total production costs of the vessel and allocates a budget to every department or resource included in the production project. The BVC method are the detailed cost estimations the different departments make for their part of the production process. This research will focus on the BVC cost estimation phase of the ship section building. The hour evaluation is performed at the end of the production process and compares the estimated hours with the final production hours of the ship sections. There are a few problems that occur during the current cost estimation methods. Only one input variable is used, weight, and the work performed on the section is left out of the cost estimation. It is presumed that using more than one cost driver and the work performed on a section will lead to more accurate cost estimations. There is to little integration between the three cost estimation methods because there is no feedback back in to the cost estimation process. There are no evaluation possibilities regarding design or cost optimizations. The feasibility study will compare the two new methods with the current method and base this comparison on the problems occurring with the old method.

The second method is a parametric statistical cost estimation method. This method uses more input variables than the current cost estimation method, but less than the graph database method. The variables that the statistical method uses are the highest correlated cost drivers towards the final production hours. It will therefore be used as a method to see if it is necessary to use more than two high correlated cost drivers or if this method is sufficient for an accurate cost estimation. The statistical method uses a regression analysis to obtain equations that can be used to estimate the total production hours. These equations use cost driver values as variables. The cost drivers are found by a correlation matrix. Variables with high correlations towards the hours are parts and profiles. Weight turned out to have a low correlation.

The third method is based on the CBR theory. This theory solves a problem by reusing a solution from an old case as a solution for the new case. This theory uses a case base with historical data from old cases and tries to find a similar case to the new case in this data base. This can be used in a ship section cost estimation method. By creating a data base with the data of old ship sections a data set for solution seeking can be made. When the cost estimation of a new section is needed, a similar section can be found in the database. The solution of this section needs to be adapted and can then be reused as a solution for the new section. For the third cost estimation method the graph database functions as the case base for the CBR theory. The graphs are the data stores of the ship sections and similar sections can be found inside this graph database. A ship section graph is built up out of nodes and relations. The nodes represent parts or the total ship section and the relations represent the weld lines between the parts. Inside a graph data can be stored both as labels to the nodes or as labels to the weld lines. Not only the cost estimation can be an output of a similarity search, also design problems can be solved with the graph database method.

For the comparison of the methods qualitative and quantitative analyses are performed. The qualitative analyses showed that only using weight as a cost driver cannot estimate the right hours for every ship section set up. Confirming the correlation analysis and the presumption that there are other cost drivers with more effect on the final production hours. The graph database turned out to be the most preferable cost estimation method from the qualitative analysis. This method showed for every structure a difference in the cost estimation and can use the work that is performed on a section. This method can also be used for design suggestions, for instance in the reduction of parts.

The quantitative analyses uses the data available at IHC to estimate the costs of a new section. The IHC BVC hour estimation for the new ship section showed 18% difference with the final production hours of the vessel. For the statistical method the equations were filled in with section data from the new section. This showed an average difference over the equations of 14.6%. The graph database method was tested with a simplified example estimation. The estimation is simplified because the exact database search algorithm and the similarity assessment are not finished. The similarity search was performed to find the most similar section in the dataset. When these were found the similarity assessment adapts the hour of the similar sections. This adaptation is performed in two ways: by using the total difference ratio between the two sections or by looking at the differences between the label values between the sections. The latter can for this research only be performed for the weld length. Both methods show a more accurate cost estimation then the other two methods: the difference ratio results in an average of 9% and the weld length difference results in an average of 11.5% difference with the total production hours.

Both the statistical and the graph database cos estimation methods result in more accurate estimations than the BVC method. The graph database method results in the most accurate cost estimations. This method uses more cost drivers than the other two methods and it shows the work made during the ship section production. Also for the hour evaluation the graph database method is the most preferred method. Because it uses many cost drivers and the production work analyses can be made on these variables. These can be used to optimize the production process or the design of the ship sections.

Case based reasoning is according to the research a feasible cost estimation method for the ship building industry. When the CBR theory is used in the form of a graph database with graphs containing the ship section data the method results in more accurate results. However this research only test graphs that contain much ship section data about cost drivers and about the work performed during production. The statistical method and the current BVC estimation both use less cost drivers and cannot take the work performed into account which results in less accurate cost estimations.

Contents

List of Figures

List of Tables

Abbreviations

AC	Actual Cost
ACWP	Actual Cost of Work Performed
ANOVA	Analysis of Variance
API	Application Programming Interface
BAC	Budget At Completion
BCWP	Budgeted Cost of Work Performed
BCWS	Budgeted Cost of Work Scheduled
BVC	Bedrijfs Voor Calculatie (company pre-calculation phase)
CBR	Case Based Reasoning
CH	Chain weld
COCOMO	Constructive Cost Model
CPE	Central Production Engineering
CPI	Cost Performance Indicator
CSV	Comma Seperated Value
DBMS	DataBase Management System
df	Degrees of Freedom
DMS	Document Management System
EAC	Estimate At Completion
ECAC	Estimated Cost At Completion
EGW	ElectroGas Welding
ERP	Enterprise Resource Planning
ESW	ElectroSlag Welding
ETAC	Estimated Time At Completion
EV	Earned Value
FCST	Forecast of Remaining Work
FP	Full Penetration weld
FSW	Friction Stir Welding
FTG	Estimate To Go
FW	(double) Fillet Weld
GMAW	Gas Metal Arc Welding
GT	Gross Tonnage
HK	Hoekstaal Profiel
HP	Holland Profiel
IACS	International Association of Classification Societies
IHC	Industriële Handels Combinatie
IMO	International Maritime Organization
LL	international convention of Load Lines of Ships
MARPOL	international convention for the Prevention of Pollution at sea
MARS	ERP system for Royal IHC
mh/ton	Man Hours per Ton
MMAW	Manual Metal Arc Welding
MS	Mean Squares
NoSQL	Not Only Structured Query Language
NT	Net Tonnage
O&C	Ontwerp Calculatie

O&M	Offshore and Marine
P&C	Planning and Control
PL	Plate
PNG	Portable Network Graphics
PP	Partial Penetration weld
PS	Planned Spend
SAW	Submerged Arc Welding
SOLAS	international convention for the Safety of Life at Sea
SPI	Schedule Performance Indicator
SQL	Structured Query Language
SS	Sum of Squares
SSW	Solid State Welding
ST	Strip Profiel
STCW	international convention for the Standards of Training, Certification and Watchkeeping for Seafarers
TONNAGE	international convention on Tonnage Measurement
V1	Version 1
V2	Version 2
VC	VoorCalculatie (pre-calculation phase)
VC 1	VoorCalcualtie version 1 (costprice of the vessel)
VC 2	VoorCalcutalie version 2 (budget for the new production project)
VK	Vierkant Profiel
WT	WaterThight

1
Introduction

The research discussed in this report is based on a feasibility study of a new cost estimation method designed by G. Alblas. This new method is based on the Case Based Reasoning theory and will hereafter be called the graph database method. To test the feasibility of this new method for cost estimation the problems occurring with the current method used by IHC will be discussed. These problems lead to the goal the new research has to fulfil. A parametric cost estimation method is designed to see if using multiple cost drivers create a more accurate cost estimation. It also indicated which section building variables are the highest cost drivers. The graph database method will include multiple cost drivers and the work performed on the section during the production. The comparison between the qualitative and a quantitative analyses of the statistical method and the graph database method will show which method results in the most accurate cost estimation. It will also show if it is beneficial for the cost estimation process to use more then the three highest cost drivers and the work performed on a section.

This chapter gives an introduction to the research and the way the research is approached. In section 1.1 the problem occurring with the current cost estimation methods of Royal IHC is introduced, as well as the goal of the research. Furthermore, the main research questions and the sub questions are stated in section 1.3. The approach of the project is illustrated with a flow chart, which is visualized in section 1.4. The section 1.5, shows the report outline.

1.1. Problem Definition and Thesis Objective

The current cost estimation method of Royal IHC is performed in three different phases of the ship production project as shown in figure 1.1. In the beginning of a project the VC cost estimation will determine the production costs of the vessel. The VC estimation will also allocate a budget to every department that is a part of the production project. During the BVC estimation each department will estimate the detailed costs for their tasks in the production project based on the allocated budget from the VC estimation. This project is focussed on the cost estimation of the steel part of the ship section building. These hours are estimated during the BVC phase by the production department. At then end of the project an hour evaluation compares the estimated hours with the realized production hours. The current cost estimation process being divided in three phases leads to a few drawbacks. The phases have little cohesion with each other and use different programs for the estimation of the costs.

Both the VC and the BVC cost estimations use weight as an input unit for their estimation and the output is given in man hours per ton. When using only the weight of a vessel or a ship section, work that is performed during the production is left out of the cost estimation. This may lead to less accurate cost estimations but also the matter in which optimization is possible is limited. An hour evaluation can only provide information in terms of the project progress and the optimization feedback in terms of the weight.

Presumed during the feasibility test is that a more integrated method would lead to more accurate cost estimations and broader evaluation possibilities. Reasoning behind this presumption is that more cohesion between the methods will lead to more feedback between the different estimation phases. Another presumption

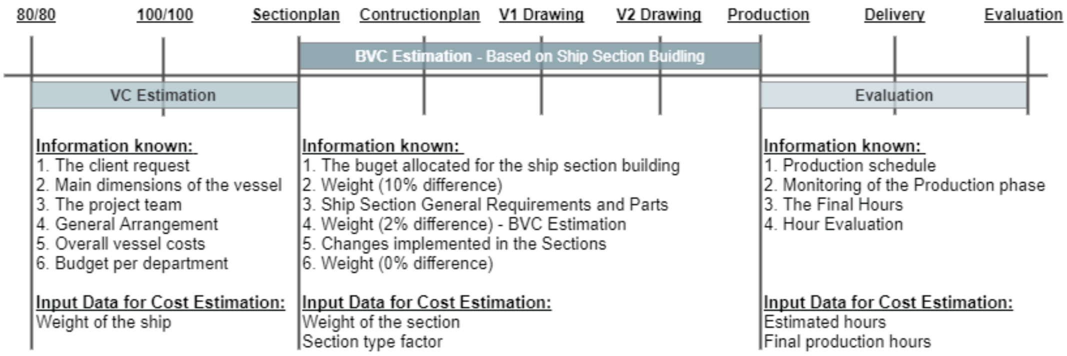


Figure 1.1: Production Project Time Line

is that using multiple cost drivers and including the work that is performed during the production will lead to more accurate cost estimations and more detailed evaluations. If it is possible to integrate more detailed information of the production process, optimizations can be achieved for both the cost estimation process and the production process.

The problems with the current method lead to goals for a new cost estimation method. The first goal is to create more accurate cost estimations with the new method. The goal is presumed to be achieved by using multiple cost drivers and taking the work performed on a section into account. Figure 1.2 shows the preferred time line for the new method, connected to the second goal for the new method: creating a more integrated cost estimation. This goal is presumed to be achieved by limiting the amount of cost estimation phases and offer the possibility of optimizing the production process with the help of the new method. The third goal is linked to the limited of evaluation possibilities. Presumed is that the using more then one cost driver and taking the work performed on a section into account during the cost estimation will lead to broader evaluation possibilities. This may lead to the option of using the evaluation results in new ship building projects.

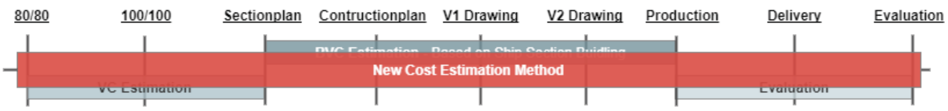


Figure 1.2: Production Project Time Line

1. **Creating a more accurate cost estimation:**
   - Possibility to use multiple cost drivers
   - Observing the work that is performed during the production
2. **Creating more integration between the cost estimation methods:**
   - Optimization possibilities in terms of the production process
   - Limiting the use of different cost estimation methods during the production project
3. **Creating more possibilities for the section evaluation :**
   - Possibility to use the results of the evaluation in future cost estimations
   - Evaluation is possible on more than one cost driver and on the work performed during production

The thesis objective is to perform a feasibility study of a new case based reasoning cost estimation method for the ship building industry introduced as the graph database method. This study will show if the goals for the new cost estimation method are achieved. A third cost estimation method will be introduced as the statistical method. This method lies, in terms of input data used for the cost estimation, in between the current cost estimation method and the graph database method. The results of the statistical cost estimation method will show if it is necessary to use every cost driver and all the work performed as in the graph database method or if less information will also lead to achievement of the discussed goals for a new method.

1.2. Scope of the Research

The research has to be executed in a specific time frame, so a few limitations and assumptions to the research scope are set. These limitations and assumptions will set boundaries for the research. In the conclusions the influence of the assumptions will be discussed.

Focus on the Steel Part of the Section Building Process.

The research is focused on a small part of the ship building process due to time constraints. This research looks into the steel part of the ship section building. Material preparation, outfitting and painting of the ship sections are left out of the analysis. This part of the production process concerns around the 7.5% of the total ship costs.[1] It is however a part of the production process that can benefit from process optimizations because there is much work performed by the yard and its employees. Included in the ship costs are also material and equipment costs which is only a purchase part. The section building part of the production process can thus benefit from design and cost optimizations.

Research Based on the PhD Research of G. Alblas

This research examines the cost estimation as proposed by G. Ablas and verifies the method with a feasibility study for the section building at Royal IHC. The method designed by G. Alblas uses the case based reasoning theory in a cost estimation method. The cases are represented by section graphs in a database, therefore this method will be called the graph database method during this research. The research from G. Alblas will create a graph database that can be used during the whole ship building process and can provide a method to analyse ship sections.

Sections with Identical Design take the same Time to Complete

One of the most important assumptions made in this research is that sections identical in design take the same amount of time to complete. This is not the case in real life. Shipyard conditions and the employees at the shipyard may lead to changes in the production hours. The production process is a process involving human beings as main producers. This means much factors are present that differ per person, such as the welding speed but also breaks and day to day variations in work efforts may lead to changes in production hours performed on a section. The production personnel administrate the hours they spend on the production of a section, which can be a rounded number or a biased number leading to small errors in the hour administration. Then there is a randomness factor including unstable factors in the production process such as the weather, equipment and changes ordered by the client or IHC. Still, even though it is probably not achievable to produce identical sections in exactly the same amount of time, this assumption is necessary to be able to say that the method could provide a valid cost estimation.

Focus on the Royal IHC Production Process & one of the Production Yards

The research will be conducted at Royal IHC so the section building data and cost estimation procedures will be based on their information and documentation. The focus of this project will be based on the Kinderdijk location of Royal IHC. Royal IHC has two main locations, Kinderdijk and Krimpen aan den IJssel. Even though they are both Royal IHC yards they do have some differences. The two locations are for instance different in terms of slipway dimensions, section building space, processes performed to estimate costs.

One Type of Ship and a One Section Type

For the feasibility study the data collection from ship sections used in the analyses must be big enough to perform the comparison of the three cost estimation methods. Therefore is chosen to create a selection in the type of vessel and the type of ship section. The type of ship is a hopper dredger. At the Kinderdijk location they perform the cost calculation of custom build dredgers, this cost estimation will be discussed in the report and used as the reference cost estimation method of Royal IHC. Appendix D shows a section plan of a hopper dredger. In this plan the different section types are named in the table. For the research the focus lies for the VC phase on the 1400 block and for the BVC on the midship buoyancy compartments.

Prior focus on the BVC phase of the cost estimation processes

Due to time constraints for this research the choice is made to start with a feasibility analysis of the new cost estimation method as a replacement for the BVC phase. At this phase detailed information of the ship sections is available at Royal IHC, from new vessels but also from older vessels. This gives a good starting point for an analysis. If the CBR method proofs feasible the method can be tested for the VC phase and as an hour evaluation method for the final phase.

1.3. Research Questions

The main research question of this research is: "Is Case Based Reasoning a feasible cost estimation method for the ship building industry?". This is a question that combines the research of G. Alblas and Royal IHC. This main question can be answered with the level the goals for the new cost estimation are achieved and the following research questions:

- Do the statistical cost estimation method and the graph database method give more accurate cost estimation in comparison to the current IHC cost predictions methods in both the VC and the BVC phase?
- Is it possible to gain more insight, by using the different cost estimation methods, in the cost drivers of the production hours and in aspects that lead to deviations between the cost estimation and the final costs of the ship building?

1.3.1. Subquestions

To answer the research questions discussed in section 1.3 sub-questions are composed. The sub-questions will be answered in the chapter they will be discussed. The sub-questions are:

1. Graph similarity between ship sections is an important part for the comparison of different objects, following sub questions can be asked:
   (a) How can a cost estimation be made with the use of these sections graphs?
   (b) Which information or data from ship sections are important to create graph models and combine them in a graph database?
   (c) How could a ship section be modelled into a Graph?
2. What are the current cost estimation methods that Royal IHC uses?
3. What are the most important aspects to look at during the production of a ship to create a realistic cost estimation?
4. How can the difference in budgeted costs and final ship production costs be explained?
5. How can the CBR method help to create a cost estimation?

1.4. Approach of the Research

The Research is performed for the TU Delft and Royal IHC. During this period the research has to be performed, a report has to be written and a presentation will be given at the end of the project. In figure 1.3 a flowchart of the Research is given.

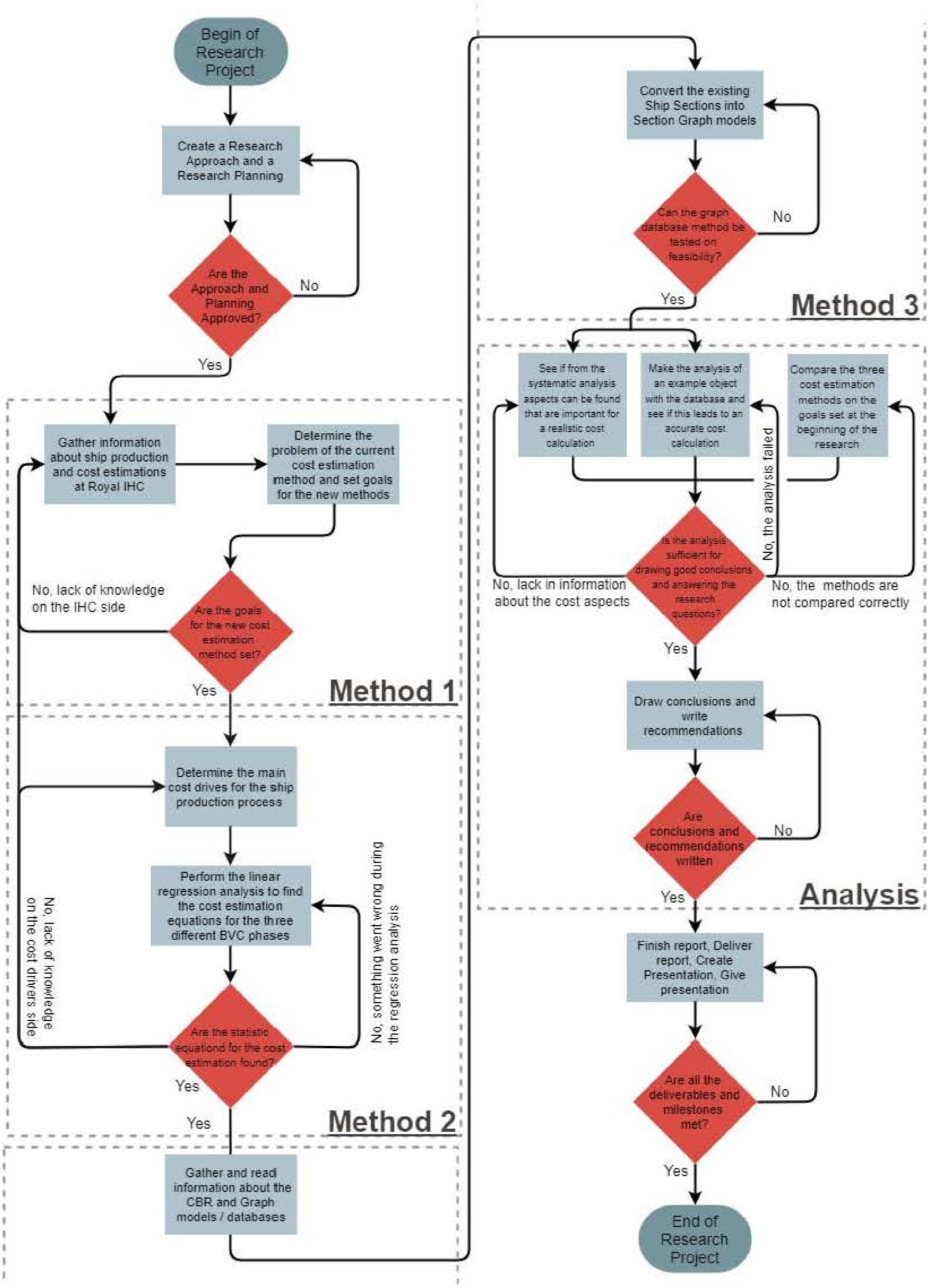


Figure 1.3: Research Flowchart

The flowchart shown in figure 1.3 starts with composing a research approach and planning, these have to be approved before continuing to the next step. For the current cost estimation method the background information is gathered. The needed information is split in two parts. The information needed to understand the ship building process and the current cost estimation methods at IHC. This information is based on literature, documentation at IHC and people references from IHC. The statistical method is described in two steps. The analyses performed for the statistical cost estimation method are a cost driver analysis and a linear regression analysis. The regression analyses will result in equations to estimate the costs of ship sections. For the graph database method first information about the CBR theory is gathered. Then the graph database method is explained. In the analysis part the qualitative and quantitative analyses are performed. The qualitative analysis is performed to see how the three different cost estimation methods, the current method, the statistical method and the graph database method, behave in estimating costs for different structures. Then a quantitative analysis is performed to compare the statistical method and the graph database method with the current BVC estimation. For this analysis an example section is used and the costs are estimated with both the new methods. From these results conclusions can be drawn and the research questions.

1.5. Report Outline

The report outline can be found in figure 1.4. This shows a chapter outline of the report, it also shows which sub-questions will be answered in which part of the report and which method is used to answer these questions.

- **Chapter 1:** Shows the introduction to the research and the report. It includes the problem definition, research questions, the approach of the research and the report outline.
- **Chapter 2:** Shows an introduction into shipbuilding from a literature point of view and the shipbuilding process at Royal IHC.
- **Chapter 3:** Explains the current cost estimation methods performed at Royal IHC and the drawbacks from this method.
- **Chapter 4:** The second cost estimation method will be discussed. The cost driver analysis will show the correlation of the section variables to the realized production hours. The regression analysis will result in equations for the cost estimation.
- **Chapter 5:** Introduces the Case Based Reasoning theory and explains how this could be used in determining cost price in the ship building industry.
- **Chapter 6:** The graph database cost estimation method will be discussed. First is explained why is chosen for a graph database over other database types and which graph database type is chosen. Then the cost estimation process with using section graphs is explained. The labels information used for in the section graphs are listed. In the end the complete purpose of the graph database method for the entire production process is explained.
- **Chapter 7:** The qualitative analysis of the three cost estimation methods. With the use of different ship section structures a comparison between the different cost estimation methods is made.
- **Chapter 8:** The quantitative analysis of the three cost estimation methods. With the use of quantitative data from the sections and the cost estimation methods the costs for a new ship section are estimated and compared for the three cost estimation methods.
- **Chapter 9:** Discuss the conclusions of the research. It will answer the main research questions from the research.
- **Chapter 10:** Explains the future developments at Royal IHC and the implementation plan for the graph database method. The validity of the research and the recommendations for further research are discussed. Then the assumptions and the restrictions of the research are listed as well as some discussion topics.

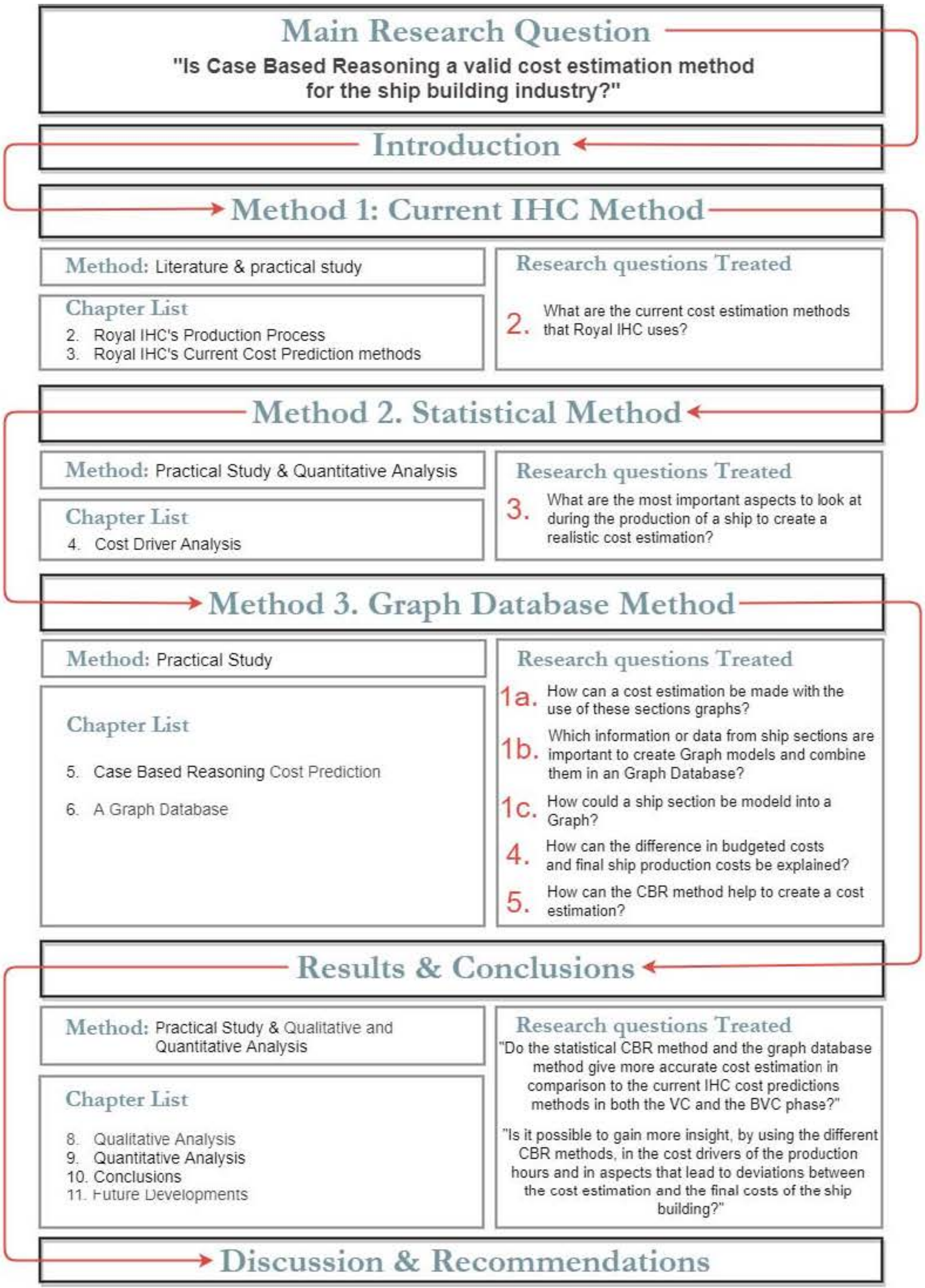


Figure 1.4: Report Outline

2

Royal IHC's Ship Production Process

This chapter will discuss the global ship building process and how this is performed at Royal IHC. Royal IHC is a Dutch company that produces numerous types of vessels. It has two big shipyards in the Netherlands, one in Kinderdijk and one in Krimpen aan den IJssel. At both yards facilities are available for section building and at both yards a covered slipway is present for the hull assembly and the launching of vessels.

In section 2.1 an introduction is given of the general ship building process. Section 2.2 gives an insight in what Royal IHC does in the area of ship building and how their production process differs from the introduced general ship building process.

2.1. Ship Building

How the shipbuilding process is performed depends on different factors, for instance the type of vessel or the availability of the yard. This section is a general introduction of the shipbuilding process taken from literature. An example of the general shipbuilding process is shown in figure 2.1. It starts with a request from a client to the shipyard for a new vessel, the shipyard starts with the basic drawings. When the contract is signed a project planning is made and after that the detailed drawings are made. The production planning is based on these drawing and after the planning has been made and the drawings have been approved the material can be procured. When the material is available the sections can be build and when ship sections are ready they can be assembled into the final ship. After all the blocks have been assembled the (remaining) outfitting is performed and the surface is prepared for the paint job. Throughout the complete production process the vessel and its constituent parts are submitted to quality tests, performed by the classification society and by the client. If these tests have positive outcomes the ship is delivered to the client.

It should be noted that some of the processes shown may also be performed at another stage in the production process. Instead of painting the complete ship at once, the ship sections may be conserved before they are assembled on the slipway or in the dry dock. The same goes for the outfitting job. This can partly been done when the sections have been built, called pre-outfitting, or, after the ship has been launched, at the quay, called quay-outfitting. Some shipyards choose to outsource some of the production steps. The conserving of ship sections for instance can be performed at or by a different company. [27],[32]

For this research the main focus lies on the steel part of the section building performed at Royal IHC, which concerns roughly 7.5% of the total vessel costs.[1] In figure 2.1 the shipbuilding project phases are shown. The dotted lines indicate the cost estimations phases. The circle shows the phase on which the cost calculations for this research project are based. The aim of the research is to validate a new method to estimate the ship section costs more accurately, so it is important to know what activities have to be performed during the ship section production. The activities during these phases have to be examined on what the cost drivers for the section building cost estimation are.

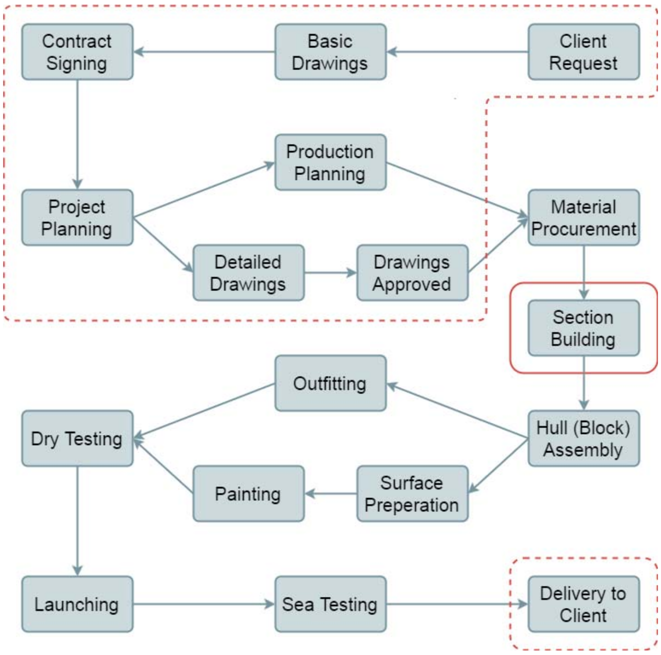


Figure 2.1: An example overview of the shipbuilding process [27], [32] and [35]

2.1.1. Stakeholders

A more accurate cost estimation is beneficial for the shipyard. There are however more parties involved in the shipbuilding process. These stakeholders will be briefly mentioned to show who is involved and might influence the production process.

The order for a new vessel is given by a client, called a shipowner. The ship building process starts with a request to build and design a vessel from a client. When a yard and the client have an agreement, other stakeholders are needed to ensure a good progression of the shipbuilding project. A design team is asked to make the design and engineering drawings. The drawings have to be approved by the class inspection societies and are needed for the production of the vessel. A ship building process in terms of stakeholders is shown in figure 2.2. Some stakeholders are not directly related to a specific part of the shipbuilding project, they are not shown in the figure. The shipyard has to keep them satisfied during the complete project or even when there are no shipbuilding projects performed during a period of time.

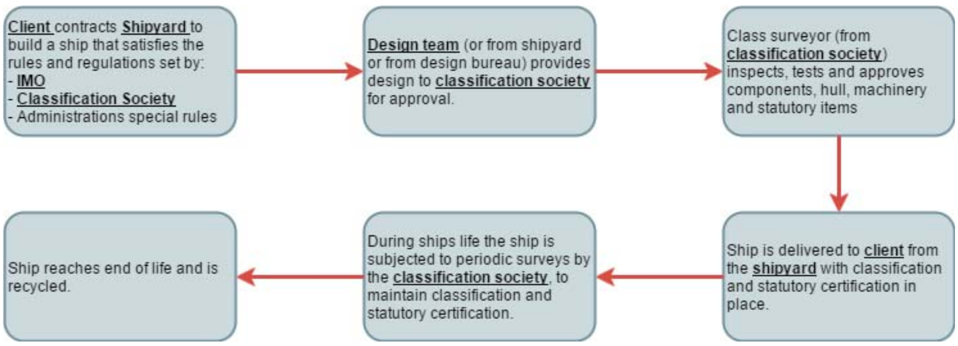


Figure 2.2: An overview of the shipbuilding process in terms of stakeholders. [18]

In figure 2.3 the stakeholders in connection to the shipyard are indicated. The red lines suggest the stakeholders delivering goods to the shipyard and the green lines suggest the shipyard delivering goods to the stakeholder. Appendix B explains the different stakeholders mentioned in figure 2.3 in more detail.

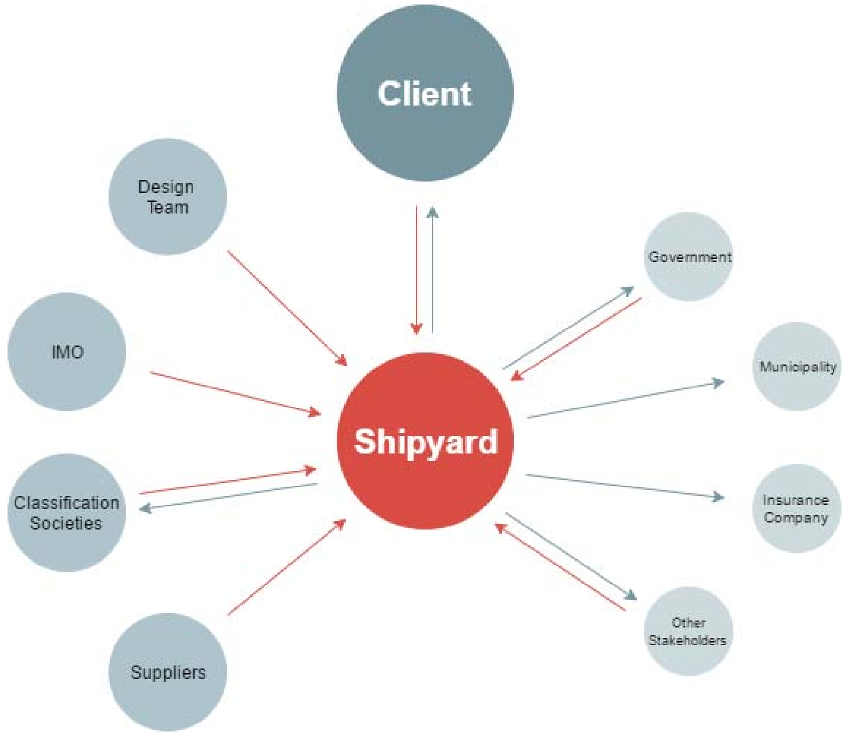


Figure 2.3: Stakeholder overview

The CBR Cost Estimation Method in Relation to the Stakeholders

If the new cost estimation method will prove to provide a more accurate cost price[2] in the total production cycle over the current IHC cost estimation methods, the shipyard will benefit from this. They will have a method to estimate the costs throughout the whole process and a way to evaluate the production process. This may lead to optimizations in the design of the vessel and improvements of the production process. If the new estimation method can provide design suggestions for the new vessel or improve the ship production process, the new method might not only be beneficial for the shipyard but also for the client. The shipyard can suggest new design proposals towards the client or the proposed vessel may be delivered in a shorter amount of time by the optimized production process. For the design team the broader evaluation possibilities are beneficial if they want to see the effects of changing the design of the vessel in the cost price and to make design suggestions using historical data from old vessels.

The other stakeholders that are mentioned might not have a direct benefit from a new method, but they do have influence on the cost price and the shipbuilding process. This can be in a direct way or an indirect way. The stakeholders that have a direct influence on the ship building process and the cost price are the suppliers, IMO and the classification societies. Suppliers have a direct influence on the cost price because they compose the costs of the delivered material. IMO and the classification societies set the rules and regulations the ship has to meet and the yard has to hire a class inspector to perform tests on the new vessel. Stakeholders with an indirect influence on the shipbuilding process are the government, municipality and the insurance companies. The government and municipality set the rules and regulation for the yard itself. This can indirectly influence the building process of an individual shipbuilding project. In terms of cost price they also regulate the employee circumstances and minimal loans which can change the ship costs. The shipyard pays an amount of money to the insurance company, if this fluctuates it can influence the costs of a vessel. The other stakeholders can create an unforeseen influence on the shipbuilding process and costs. The surrounding area can contain other companies or residential areas where changes or problems might occur that can change the production process.

[2] This research only concludes if the method can give more accurate cost estimations for the section building process. Further research into the method of G. Alblas will conclude if the method can also provide an accurate cost estimation for the total ship costs.

2.1.2. Shipyard Production Process

To build a vessel shipyards need equipment and machinery. In this subsection the general equipment needed for the different process steps of ship production will be discussed. This chapter only covers the production phases. Figure 2.4 shows the production process in more detail. The different processes are briefly discussed below. The different steps are situated at the logical place during the shipbuilding process, but if late time changes are made in the production planning some of the steps may be performed in a different order. Furthermore it must be noted that the steps can be performed next to each other or in a different sequences. For instance the section building can be performed next to other steps in the process and sometimes the painting of sections is pushed to the ship painting step because of time issues. So some deviations to this process visualisations must be kept in mind.

Material Preparation Equipment

The needed material can be ordered at supply companies. Some of this material needs extra changes to be able to fit in the ship, see figure 2.4. Steel plates might need bending or post processing before it can be placed in the ship section. The steel is delivered in flat plates with a maximum surface and thickness. This material can then be processed into ship section parts. This can be done by cutting, bending and post processing methods. The shipyard can perform these material preparation processes at the yard itself or the yard can choose to outsource these activities at a different company. The steel plates and profiles are purchased from suppliers.

Section Building

Ships are built in smaller building units, called sections. These sections are assembles into complete ships later in the production process. From the general arrangement of the vessel a section plan is made. An example of a section plan is shown in appendix D. A ship is composed of sections for a couple of reasons the plate material available is not longer than a certain length and if the ship can be divided into smaller parts the project planning can be divided into smaller projects as well. This last aspect provides a clearer overview of the total ship building project. Often sections are again divided into smaller subsections or sub-panels. During this research the production of sub-panels is placed under the section building instead of the material preparation. Welding longitudinal or vertical profiles to plates was performed in the panel street and is still performed in the robot hall. During section building different welding techniques and welding properties can be chosen. More on the welding process can be found in appendix C. [7]

Hull Assembly

During the hull assembly the ship sections are assembled together into the ship hull. Sometimes sections are prior to the hull assembly assembled in blocks, called a block assembly. The assembly of sections is performed using an overhead crane. The hull assembly is performed at the yard, in either a dry dock or on a slipway. The cranes have to be able to reach the stored ship section and the place where the vessel is assembled. [27]

Conserving

Before plates can be used in the production process they are shot-blasted. This is a method that removes rust and mill scale. Then the plates are painted with a primer paint. If sections or the complete assembled ship have to be painted a different procedure is followed. First the surface is cleaned to remove the mill scale, dust, etcetera before painting. If sections are painted before they are assembled a primer paint might be useful to prevent the material from rusting. When the ship is assembled the ship will be fully painted. The paint job for the hull is divided into three parts, below waterline, the waterline and the topside of the hull and superstructure. The inside of the vessel, tanks and living areas is painted with different coatings. [7]

Outfitting

Outfitting is the process where all the parts such as pipes, interior, machinery and electrical wires are placed in the vessel. C. Rose [27] divides the outfitting process of a ship in three stages: pre-outfitting, slipway or

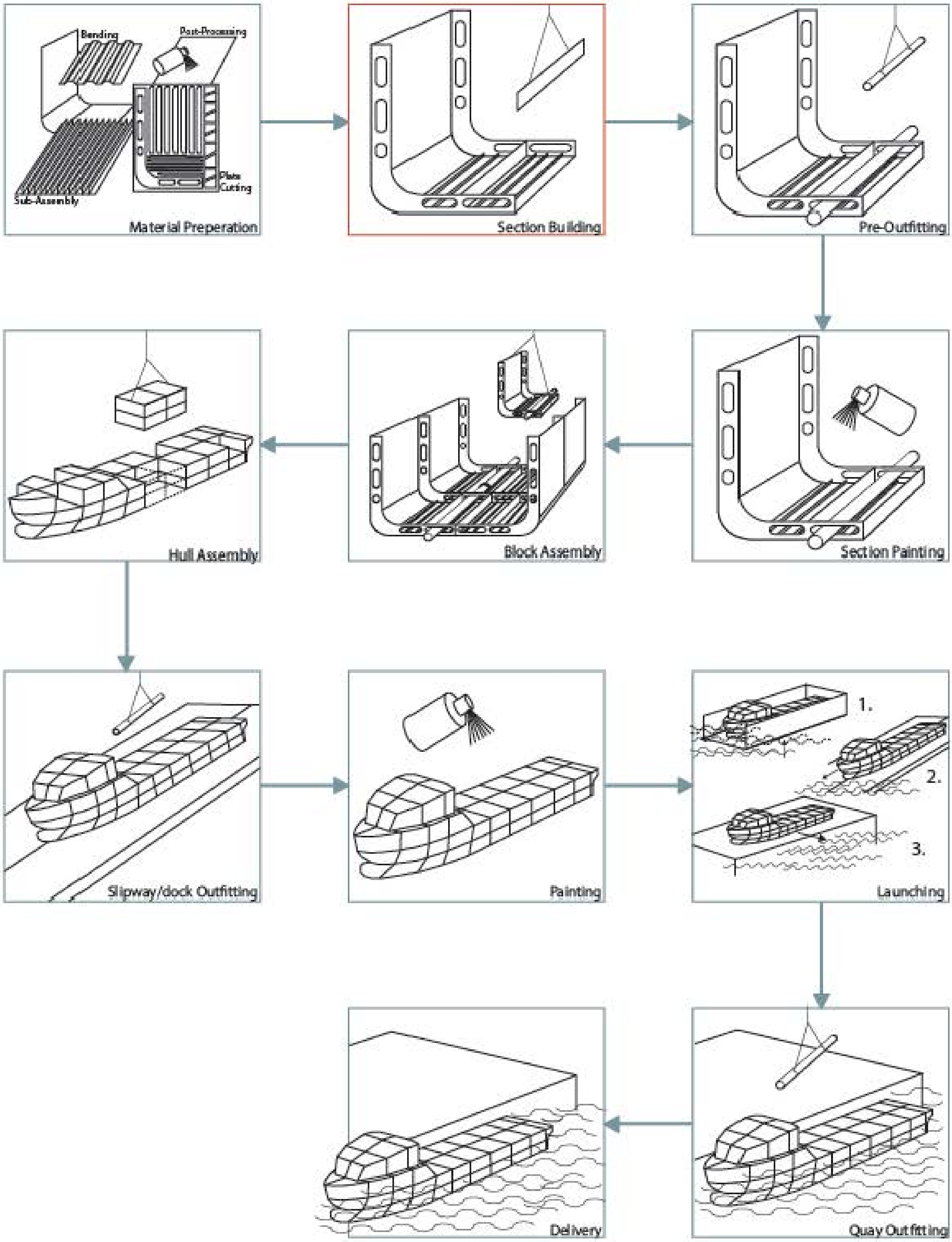


Figure 2.4: Shipbuilding Process

dry dock outfitting and quay outfitting. Pre-outfitting is performed after the section building process. The sections are expanded with the needed components. This can be done during the section building or after the section building. Slipway or dry dock outfitting is performed during the block assembly. This phase needs sufficient planning because machinery can only be placed in the machinery room when the section on top of the machinery room has not yet been assembled, because this machinery has to be placed by cranes. The quay outfitting is performed after the vessel is launched and moored at the quay. During the quay outfitting also the commissioning of the vessel is performed. This means that the vessel is being prepared to set sail so all the machinery is being installed. After this phase the outfitting should be finished and a test run can be made. [27]

Launching

During the launch the vessel is transported from the place where it is built into the water. This can be done by an end launch, a side launch or by using a dry dock. A dock launch is relatively easy, and executed by filling up the dry dock with water and sailing the ship out. An end launch is done on a slipway. The ship is built on a slipway and before the block assembly the slipway is reinforced with keel blocks to secure place for the launching system. If the width of the water into which the ship has to be launched is relatively small the ship is launched using a side launch. This is only done for smaller ships and the ship is launched sideways into the water. [7] In figure 2.4 the three ways of launching a vessel are visualized. The dock launch is shown by 1, the slipway launch is shown by 2 and the sideways launch of the quay is shown by 3.

Testing

The testing is done to check if the vessel meets the requirements and is built according to the rules and regulations. These are set by IMO and by the classification societies. The classification societies send a class inspector to perform the required tests on the vessel. These tests may be the checking of the welding lines or performing stability tests. They can be performed before the ship is launched for the underwater parts or after to perform propulsion tests. [2] Testing is not shown in figure 2.4 because it is performed throughout the whole process.

2.2. Shipbuilding at Royal IHC

"Royal IHC designs, builds and maintains innovative equipment and vessels, working from sea level to ocean floor, for maritime service providers in the offshore, dredging and mining industries. We sell vessels, equipment and services either separately, or in combination. In addition, we are able to supply advanced technology for the international rental market." - Royal IHC [11]

Royal IHC is a company that builds different types of complex vessels. Above the mission of IHC is quoted. If we look at the shipbuilding process as taken from literature that was discussed before, and compare this to the shipbuilding process of Royal IHC they differ at some points. Figure 2.5 shows the differences between the production processes on the different locations. Red circles indicate a difference between the literature and the Royal IHC shipbuilding itself. The blue circles indicate a difference in shipbuilding between the different IHC yards.

2.2.1. Royal IHC Yard: Kinderdijk

At Kinderdijk the process starts with the preparation of the steel plates and profiles. Royal IHC has their own material preparation and pre-processing department, called Metalix. This department is located at Kinderdijk and delivers steel parts or steel kits for Royal IHC yards but also for other yards. A steel kit is a package of steel plates that is preprocessed for production. Metalix can do all sorts of processing for plates and profiles, plate cutting, tapering and bevelling, 2D & 3D modeling and shaping. [21] Metalix cannot provide the subpanels, these are made by a so called panel street. IHC has recently removed the panel street at the Kinderdijk location, so the panels needed for section building will either be hand welded or purchased from other companies.

At Kinderdijk the crane capacity is often not so high that the yard is able to lift big blocks of already

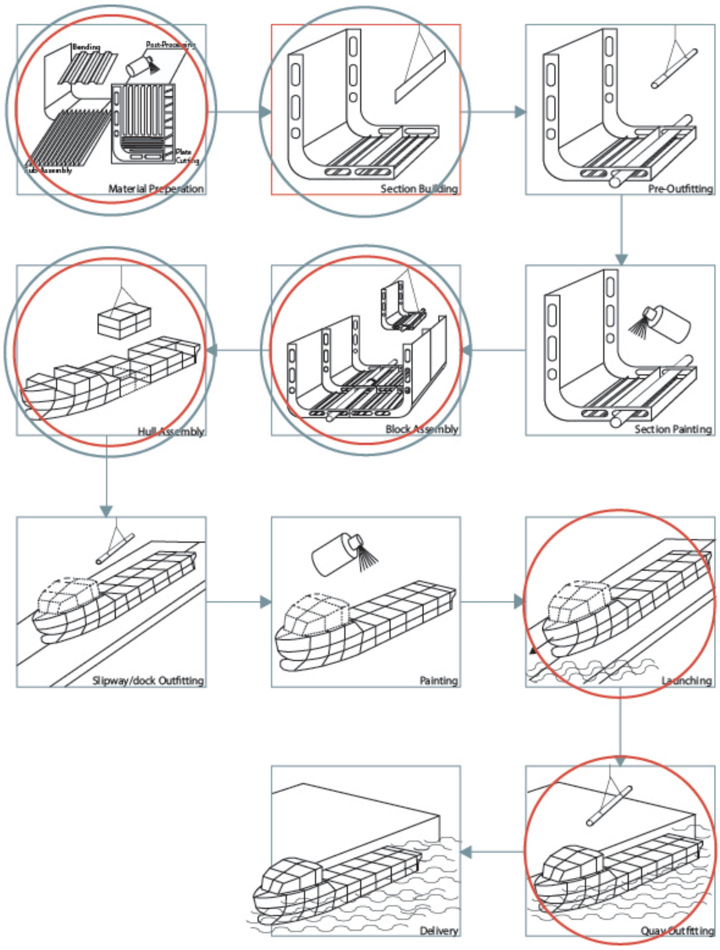


Figure 2.5: IHC Shipbuilding Process

assembled sections. So the sections are assembled at the slipway. The launch of the ship will be done by a slipway launch. The whole ship is assembled on the slipway. At Kinderdijk the slipway often does not have the capacity to launch vessels with the deck house already assembled to the ship. This is due to the fact that the door of the yard is not high enough. If this is the case the deck house is installed during the quay outfitting step.

2.2.2. Outsourcing

When a company orders goods or proceedings from other companies this is called outsourcing. A company may use outsourcing for different reasons, for strategic purposes, when there is a need for a special delivery or when the time on a project is running out. Royal IHC has their own steel preparation unit, Metalix. The steel kits can be ordered there, but if they cannot deliver on time or if their capacity is too low for the project they can be outsourced to different companies.

Currently Royal IHC outsources most of the section building to other companies. In the past they built the sections themselfs, but due to strategic reasons more sections are being built somewhere else. Reasons for outsourcing sections could be that it is cheaper to build the sections at another company, because this company has more capacity or more employees. Also outsourcing has the advantage of not having to invest in the machinery to build the sections. The panel street for instance is not necessary if sections are built elsewhere. This saves the yard money, maintenance and space. The space where the sections used to be built can be used for different purposes such as storage space or as space for assembling smaller ships. Royal IHC still has the capacity to build sections themselves. This is an option that is chosen if the sections are more complicated, generating more building risks or to keep the project running. If there are more places where sections are built the project can run faster.

The data that will be generated or can be used in a database when outsourcing sections is less than if sections are built at Royal IHC. In chapter 10 a closer look is taken on the influence of outsourcing the section building has on a new cost estimation method.

2.2.3. Differences between Royal IHC Yards

As can be seen in figure 2.5 especially in the beginning of the process there are differences between the IHC yards. This is mainly due to the fact that Krimpen aan den IJssel has a larger production floor, slipway and section floor, and the yard has more crane capacity in comparison to Kinderdijk. There are also some differences in the material preparation and in the hull assembly part of the process.

Kinderdijk has the material preparation unit, Metalix, at their own yard so material can be delivered on spot. For Krimpen aan den IJssel the material has to be ordered and afterwards be transported to the yard. This means some planning has to be done for this step in the production chain. The sections at Krimpen aan den IJssel can be bigger than the sections built at Kinderdijk, due to the higher crane capacity and the larger slipway. This makes that the section plans for ships built at Krimpen or at Kinderdijk differ from each other. Also the building process at Krimpen is different from the one at Kinderdijk. Kinderdijk has only room and capacity on the slipway for one big vessel. Capacity limits are set by the space and the crane capacity at this location. At Krimpen aan den IJssel the slipway is bigger and the cranes have more capacity so they can already start building the new vessel when the old vessel is not yet launched. This is called 'tandem building'. The section block assembly can then be moved and positioned on the slipway by cranes after the other ship has been launched.

The differences between sizes and building capacity can be seen in the documentation of the section data. For this research is therefore chosen to only focus on section built at one yard, the Kinderdijk yard.

2.3. Summary

If the new cost estimation method proves to be feasible and a more accurate cost estimation can be performed, this is beneficial for shipyard, Royal IHC. If the method can also provide design insights the method might be beneficial for the client and the design team as well. The rest of the stakeholders might not have a direct benefit from a more accurate cost price calculation. They might have an influence on the shipbuilding process and the cost price of the ship. In a direct way, this goes for the suppliers, IMO and class societies, or in an indirect way, government and insurance companies.

The production process is discussed as background information for the research. The process starts with the preparation of the section building material. Then the material is welded together into sections and the pre-outfitting and the painting of the sections is performed. When the sections are ready they are assembled into block assemblies or directly assembled into the hull. This can be performed on a slipway or in a dry dock. At IHC they assemble the ships on a slipway. During the hull assembly the slipway outfitting is performed and when the hull assembly is done the hull is painted. Then the ship is launched and moored at a quay where the final outfitting is performed and the ship is delivered to the client.

3

Method 1: Royal IHC's Current Cost Estimation Method

This chapter explains the methods Royal IHC uses to estimate the production costs of the new to be built ship in the different phases of cost estimation. These phases are the VC phase when the overall vessel production costs are estimated and when the budget per production task or department involved in the production project is allocated. The BVC or the company pre-calculation phase where each department estimates their budget in detail. The final phase is the hour evaluation phase. In this phase the difference between the estimated and realised costs are analysed for the project evaluation.

The chapter will conclude with the problem definition of the current cost estimation method and the goal for a new cost estimation method. Another important result for the new cost estimation method is the input and output data of the current IHC method. Section 3.1, shows the cost estimation method Royal IHC uses at the Kinderdijk yard. Section 3.2 shows the problem with the current estimation method and the goal for the new method.

3.1. Cost Calculation Phases at Royal IHC Kinderdijk

Royal IHC has two different yard locations, Krimpen aan den IJssel and Kinderdijk, with yard specific cost estimation methods. This research is based on the cost estimation of Kinderdijk. This is a cost estimation method made for dredgers and divided into three phases, the VC, BVC and evaluation phase. This chapter will explain how the current cost estimation methods at Royal IHC - Kinderdijk are performed. For the VC [redacted]

3.1.1. Pre-Calculation Phase

The pre-calculation or the VC phase estimated the total production costs of the vessel at the beginning of the production project. Also the budget per department or task involved in the production project is allocated. For this project the focus lies on the production department, but most of all on the budget allocated to the hull subgroups. All the different subgroups and departments are given in 'het blauwe boekje' or the standard classification from Royal IHC [28]. The hull groups are separated into nine groups, 11 - 19. These hull groups are shown in the section plan in appendix D. The numbers on the section plan correspond with the given list below. The budget after each number only covers the steel part of the section building process, the outfitting or material preparation for these sections is not covered in the budget. It should be noted that these numbers correspond to the Kinderdijk yard and the ship type they belong to are hoppers.

- 11. Aft ship
- 12. Engine room

- 13. Pump room
- 14. Hopper or Midship
- 15. Cylinder compartments
- 16. Fore ship
- 17. Poop Deck
- 18. Deckhouse at aft ship
- 19. Deckhouse at front or amidship

The VC estimation is made by the O&C department at Royal IHC. A VC calculation is divided into two calculations, VC 1 and VC 2. The first calculation is the total production price of the vessel. The VC 2 is the budget per system as explained before. The VC 1 is based on the request the client delivers, a list of requirements, the main dimensions, basic drawings and historical ship data. Especially the steel weight is an important factor in this calculation. The costs for the VC 1 are given in man hours per ton. For the VC 2 more detailed information of the vessel is required. These are based on historical ship data and early drawings and calculations from the design team. The program IHC uses to estimate the VC 1 and the VC 2 cost estimations is explained in appendix F.

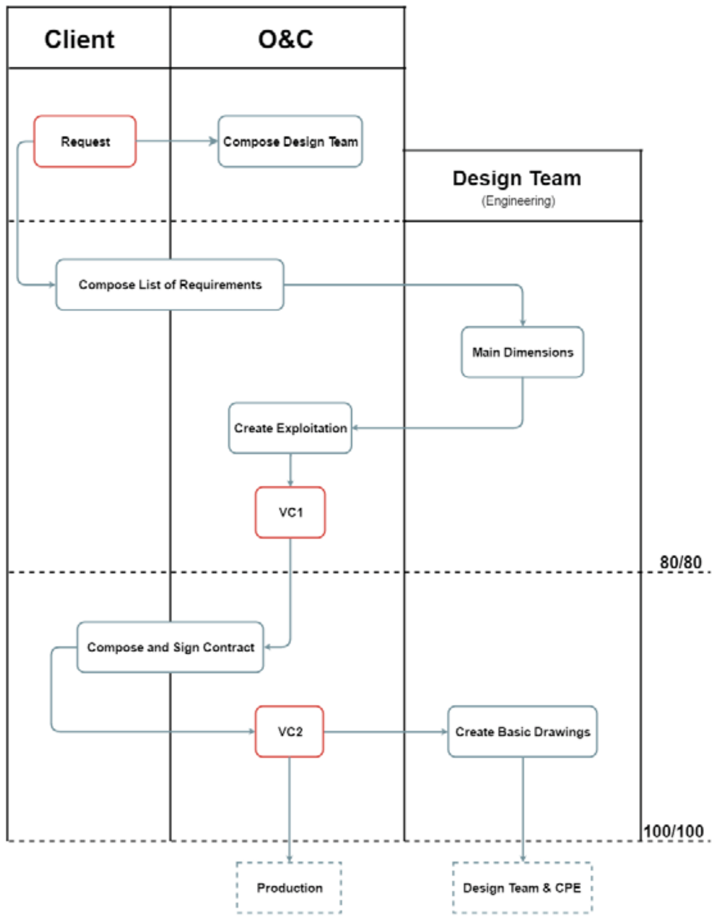


Figure 3.1: VC Calculation steps

The steps that are taken during the VC phase are shown in figure 3.1. First the client has a request for a vessel. Based on the request the O&C department composes a design team for the project and the O&C department start creating a list of requirements for the new to be built vessel. The design team estimates the main dimensions based on this list of requirements. With these main dimensions the O&C department creates an exploitation. After an agreement between the client and Royal IHC the contract is composed and signed by both parties. Where after the O&C department will allocate a budget for the different tasks in the shipbuilding project, planning steps, drawing steps, productions steps etcetera, called the VC 2. For each department different input data is needed. For the section building department this budget is based on the size of the vessel, a ship profile at the point of the building block and dimensions of the different compartments. With this data the weight for the section block is estimated and the costs are given in man hours per ton. This estimation is based on previous similar projects. For the outsourcing tasks the budget is given by the other company. After the VC 2 the design team can start with the basic drawings that will be used in the next phase by the design team and the CPE department.

In figure 3.1 the different project phases are positioned to show which process steps are performed in which phase. The 80/80 phase is the phase where the client has already handed out a request but the contract between client and Royal IHC has not been signed yet. The VC 1 is estimated during this phase. After the contract between client and shipyard has been signed the 100/100 phase starts. During this phase the VC 2 is estimated.

3.1.2. Company Pre-Calculation Phase

For a project manager the complete cost price of the project is important. For a line department such as the production department the total cost price of a project is of less importance, because they have multiple projects running in their department. The BVC combines at a line department the current projects running in the department. The BVC costs are given in man hours. To express the BVC in time is useful to create an adequate schedule for the production floor.

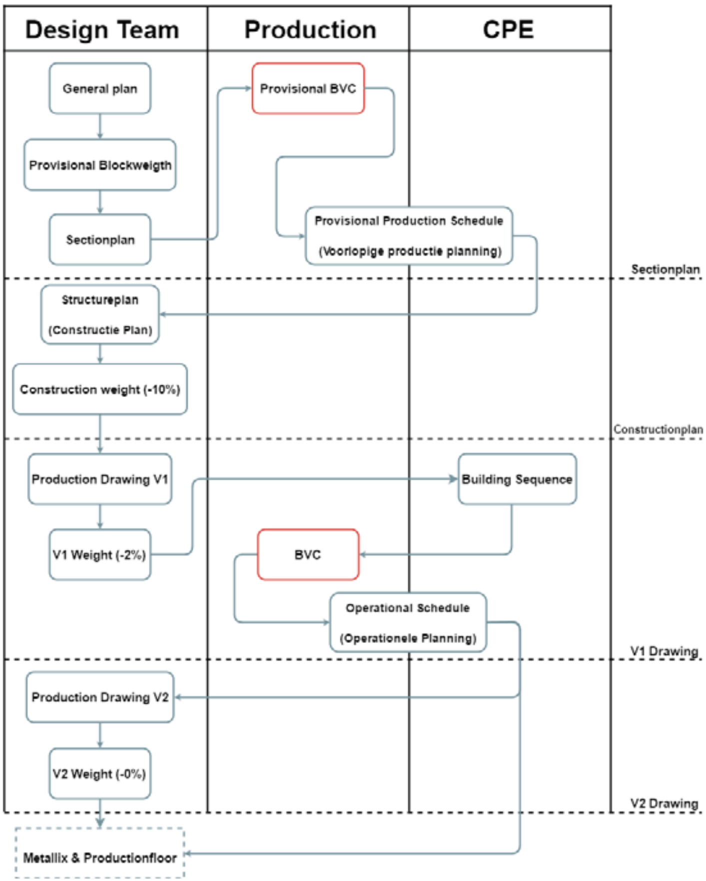


Figure 3.2: BVC Calculation steps

When the VC 2 is handed out to the different departments within the running ship building project, each department can build their own BVC on the budgeted costs for their part of the project. The production department gets an estimated budget delivered by the O&C department in the VC 2. On this budget they base their BVC. The program IHC uses to estimate the costs is given in appendix F. The cost estimation of the BVC phase is based on the weight of the ship sections and a ship section type factor. This factor represents a difficulty factor for the different ship section types. These follow from historical data. The weight is also compared to previous built ship sections of the same type.

Figure 3.2 shows the process to create a BVC for the production department. After the VC 2 has been handed out the design team starts creating a general plan and block weight. Then they create a section plan, on which the production department builds a provisional BVC and a provisional production schedule. This phase is called the section plan phase. An example section plan is given in appendix D. With a section plan the production department can create process steps, needed material and machinery, number of employees and proceedings that are needed for the production process. After this phase the construction plan is made. The design team estimates the construction weight which is accurate to the final steel weight with a maximum deviation of 10%. The construction plan is needed to create this construction weight estimation. This plan is made by the design team on forehand. Based on the construction plan the production drawing V1 is made by the design team, also the V1 weight is estimated which has a maximum deviation of 2% compared to the final weight. Based on this estimation the building sequence is made by the CPE department. This is the sequence in which the sections are build and assembled. With the building sequence the BVC is made by the production department. When the operational planning is made by the production and CPE department, the V1 drawing phase is finished. This operational planning is the schedule for the production floor. The design team creates a V2 production drawing and a V2 weight, which should be equal to the actual weight. These may contain some late changes that have to be revised in the schedule.

3.1.3. Final Ship Building Costs

During the production the budget and planning have to remain within line of the estimated planning and BVC. So keeping track of the schedule made before the start of the project is necessary to meet deadlines and to make sure the project is not making extra costs. At the end of this part the total production costs and production time of the vessel can be summed up creating the final ship production costs.

Figure 3.3 shows the hour evaluation phase at Royal IHC. The production floor gets their tasks for the production of the vessel from the BVC. They start with the ship production process. The keel laying is the start of the ship assembly process. During the production the floor employees have to keep track of their work progress and update this progress in Floor2plan. The hours are administrated by the foreman on the production floor. For the section building the administration can be performed for ironwork and welding hours. The progress of the section building is monitored with the administrated hours by the foremen and the estimated BVC hours. These will be discussed in weekly meetings with the planning department and the foremen. This process will proceed till the section building is finished. After the vessel has been launched the outfitting at the quay is performed and the vessel is being commissioned. During this phase the production and CPE departments keep track of the planning and budget. After the vessel delivery the total hours and money spent on the production of the vessel are summed up and an analysis of these final costs and hours spent is made. The analysis that is performed is an earned value analysis, more information about a EV analysis is given in appendix F.

3.2. Requirements for a new Cost Estimation Method

In order to verify a new cost estimation it is important to know what requirements the new cost estimation should meet and with what kind of input data this new cost estimation should be able to estimate the costs. In chapter 8 these requirements will be compared for the different cost estimation methods, both the current method and new methods.

3.2.1. Input and Output Data Requirements

Table 3.1 shows the estimation output and input data of the cost estimation phases. The input data becomes more detailed from the VC phase to the BVC phase. The new cost estimation method should be able to estimate the steel part of the section building. Currently the weight of the sections is used to estimate the costs. These costs can be based on more accurate data during the BVC phase compared to the VC phase.

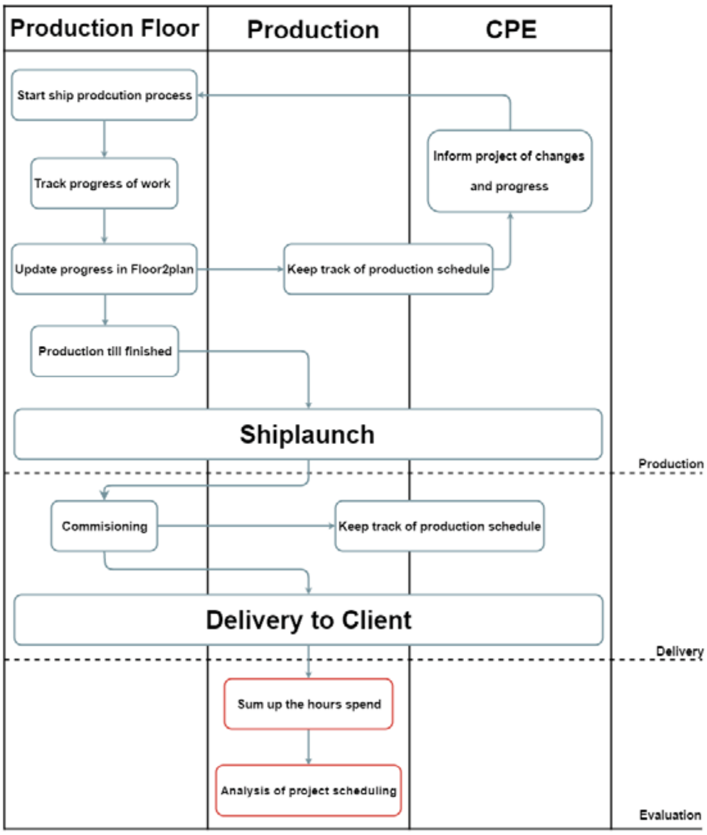


Figure 3.3: Final Cost Calculation steps

For a new cost estimation the level of input data available during the different cost estimation phases will not change. Thus a new cost estimation method should be able to estimate every output of the current phase with the same data available at that moment.

Table 3.1: Differences Between the VC and BVC Cost Estimation Methods

	VC	VC 2 For Section Building	BVC
Estimation of:	Vessel price and overall budget	Budget for the hull groups	Cost price for sections
Output Data	Total vessel cost price		Section cost price in Mh/ton
	Budget per department	Mh/ton budget per hull group	
Input data	Client requirements	Ship profile around hull group	Section Data
	Main dimensions	Main dimensions	Key figures
	Historical ship data	Historical hull group data	Historical section data
			Analysis lists

3.2.2. Evaluation Requirements

When looking at the BVC estimation results compared to the realized costs they do match the realised costs of the complete vessel when the different section hours are combine. When the separate sections are reviewed or the absolute difference between realised and estimated hours are overlooked occur large differences occur. How these changes between the estimated and the realised hours occur cannot be investigated with the current evaluation methods. This might be due to the fact that only weight is used as an input variable for the cost estimations. Presumed is that using more than one cost driver the evaluation possibilities and results become more detailed. Two requirements for a new cost estimation method are that the hour estimation possibilities are broader than the current hour estimations and that it is possible to get more insight in what influences the change between the estimated costs and the realized costs.

3.2.3. Problems Occurring with the Current Cost Estimation Methods

A problem is that both the VC and BVC cost estimation methods use the unit man hours per ton. Different production techniques used and other cost drivers involved in the section building are disregarded in the current cost estimations. Another problem with using only man hours per ton is that the standards used during the VC estimation need updates when the data becomes older and must be monitored if the older ship data can still be used for newer vessels. S. Al-Baz researched this topic (Chapter 11). In his research for the VC phase [3] he discovered that when using man hours per ton as a unit for the VC estimation of the hours these do at first sight not change over time. His first step was to create a graph in which mh/ton are lined out against the building year of the vessel. This gives almost a stable distribution around 37.5 mh/ton. For his second analysis he scaled up the same mh/ton for the size of the vessel. This led to the correlation that the larger the vessel the less mh/ton are needed to build the vessel. With this scale the same graph is made as in the first analysis with a new outcome. The amount of mh/ton needed for building a vessel increases over time. This can be due to different changes in the production process, for instance the increase of curved ships. From the current cost estimation and time monitoring methods these aspects can not easily be found. A new method should provide more insight in the used techniques and parameters to build vessels.

3.2.4. Goals for a New Cost Estimation Method

From the problem definition and the requirements discussed in this chapter goals for a new cost estimation method can be listed.

1. **Creating a more accurate cost estimation:**
    - Possibility to use multiple cost drivers
    - Observing the work that is performed during the production
2. **Creating more integration between the cost estimation methods:**
    - Optimization possibilities in terms of the production process
    - Limiting the use of different cost estimation methods during the production project
3. **Creating more possibilities for the section evaluation:**
    - Possibility to use the results of the evaluation in future cost estimations
    - Evaluation is possible on more than one cost driver and on the work performed during production

The problem definition also leads to the two main research questions discussed in chapter 1. These questions are answered in the conclusion of this report. The sub questions belong to different parts of the research and will be answered in the summary or conclusion of the chapters in which the question is treated.

3.3. Summary

Subquestion 2 - **"What are the current cost estimation methods that Royal IHC?"**

The current cost estimation methods at Royal IHC are the VC cost estimation, the BVC cost estimation and the hour evaluation. The VC method estimates the total production costs of the vessel and allocates a budget to every department or resource included in the production project. The BVC method are the detailed cost estimations the different departments make for their part of the production process. This research focusses on the ship section cost estimation and the belonging BVC estimation. The hour evaluation is performed at the end of the production process and compares the estimated hours with the final production hours of the ship sections. There are a few problems that occur during the current cost estimation methods. Only one input variable is used, weight, and the work performed on the section is left out of the cost estimation. It is presumed that using more than one cost driver and the work performed on a section will lead to more accurate cost estimations. There is to little integration between the three cost estimation methods because there is no feedback back in to the cost estimation process. There are no evaluation possibilities regarding design or cost optimizations. The feasibility study will compare the two new methods with the current method and base this comparison on the problems occurring with the old method.

Due to the time constraint of the research, the first priority is set to analyse if the new cost estimation method works for the BVC phase. During this phase most data of the section building are available. All the detailed information and drawings are available at Royal IHC. This gives the most accurate estimation of the man hours and the costs for hull sections at Royal IHC. Thus it is a good base for an analysis of a new cost method. If the new method proofs feasibility for the cost estimation during the BVC phase, an analysis for the VC, where less information is available, and the final cost phase can be made in further research.

4

Method 2: The Statistical Method

During the section building process different factors determine the total hours of the section. These factors can be properties of the section itself or properties of the work performed on the sections. The can represent for instance the amount of parts or the size or the total weld length made on the section. Not all these properties might have the same effect on the total production hours made on the section. The properties are evaluated in terms of their causality to the production hours of a ship section. Properties that have a high causality towards the hours are called cost drivers.

This chapter introduces a new method for cost estimation for the BVC phase at IHC, a parametric method that uses historical section data to compose equations that can estimate the hours of an section. Ship section properties are ranked on their correlation towards the total hours, in the cost driver analysis. Then the cost driver values will be used in a regression analysis.

4.1. Analysis Explanation

During the BVC phase of ship section production the information available from the sections becomes more detailed. The information in the begin of the BVC phase is less detailed and accurate than at the end of the BVC phase. Therefore it is chosen to create different equations that can be used at different points of the BVC phase. One set of equations for the early BVC cost estimation. One set for the detailed BVC estimation for the total production hours of the section and one for the detailed estimation when the hours are split in welding hours and ironwork hours. For these different BVC phases first the cost drivers analysis is performed and then with the three most correlating variables the regression analysis is performed.

Not every property has the same correlation with the hours of a ship section. The cost driver analysis shows the correlations of the different properties to the hours spend on the sections. In order to identify correlations between section parameters and man hours, a data set is built that contains property values of seven different midship sections from two ship orders. The data used in this analysis is obtained from documented data at IHC. The welding properties are obtained manually. Only the weld length is used in the correlation matrix. More welding properties were obtained, but data from only seven sections turned out to be too little to use in the cost driver analysis.

There are two desired outcomes of the cost driver analysis:

1. The correlation matrix shows which three variables have the highest correlation to the hours spent on the section.
2. The properties used in the different scenarios differ to make a distinction between properties that can be roughly estimated earlier in the design process of the vessel and later when all the properties are known. This will make it possible to create different equations for different stages in the design process.

The regression analysis will result in equations that can be used to estimate the ship section hours. The detailed data from seven sections is shown in table I.5. This table shows data from two different orders and six midship buoyancy compartments and one aft ship buoyancy compartment. There is also data available from 60 sections. The only variable that is not included in the data for 60 section is the weld length because this was a variable that had to be obtained manually. The analyses are performed with the six mid ship buoyancy compartments, the seven buoyancy compartments and the 60 midship buoyancy compartments. The reason behind the six and seven sections is that something can be said about the influence of using different section type. The reason for also using 60 sections is that a bigger dataset creates more significant results. More detailed information about the two different orders and the seven sections can be found in appendix I.

Table 4.1: Primary Properties of the Seven Ship Sections

Sectie ID	Frames	Brackets	Collar plates	Profiles	Parts		Weldlength	
12781421	29	31	10	102	187		1255.10	
12781423	46	60	21	150	279		1577.24	
12781425	18	52	34	114	223		1482.64	
12781225	25	92	20	95	268		1086.96	
12651421	51	83	80	211	530		2389.92	
12651423	52	71	94	158	407		2294.79	
12651425	64	70	98	170	424		2371.91	

4.1.1. Changes in Ship Section Data during the BVC Phase

Currently Royal IHC estimates the costs of the ship sections based on the weight of the ship sections using data from old vessel orders. As discussed in the previous chapters only using weight may not be the best method for section hours. The reason they use weight as a cost driver is because they can estimate the weight of the ship already in the VC phase, and the ship sections early in the BVC phase.

Figure 3.2 in chapter 3 shows that during the BVC phase the weight of the sections changes with a range from 10% to 0%. The real BVC is estimated on a weight difference of 2%. To test these assumptions an analysis is performed to calculate the total percentage that the weight changes from the first estimate to the final weight. Appendix U shows the excel files of the three analyses that are performed of two different orders, 1265 and the 1278. The first sheet shows the weight difference between all different kind of sections. Sheet two shows only midship sections, 14xx numbers, and the final sheet shows the weight difference between midship buoyancy compartments.

Table 4.2: Weight Analysis Results

	Begin Estimation in BVC Phase	V1 Weight Estimation
All Sections	14%	4%
Midship Sections	13%	3%
Buoyancy Compartments	4%	3%

Table 4.2 shows the results leading from the analyses. The first column is the estimated weight in the beginning of the BVC phase. The second column is the V1 weight estimation that is supposed to be 2% off from the real weight of the section. As can be seen the first estimate is maximum 14% off and the V1 weight estimation maximum 4% off.

The parts and weld length are also presumed to be estimated early in the BVC phase. Table 4.3 shows the final weld length and a weld length estimated with only the plates, profiles and blukhead frames included: the early weld length. The difference is maximum 12% so presumed that the weld length difference is in line with the weight difference. Table 4.4 shows the final parts of the section and the estimated parts without brackets and collar plates. The maximum difference is 42% so if there are no amounts of brackets and collar plates known the estimation of the parts will not be accurate. But is the amount of brackets and collar plates can be estimated with 75% accuracy parts can also be a property that can be estimated early in the BVC phase. Where the detailed information of brackets and collar plates will have a difference of 25% the total parts will only have a maximum difference of 11%. Therefore the total amount of parts and weld length will be used as the other two variables for the early estimation BVC phase equations.

Table 4.3: Weld Length Percentage Difference

	Early weld length	Final weld length	Difference
12781421	1199.54	1255.10	4.43%
12781423	1514.19	1577.64	4.02%
12781425	1410.10	1482.64	4.89%
12781225	958.11	1086.96	11.85%
12651421	2158.89	2389.92	9.67%
12651423	2079.66	2294.79	9.37%
12651425	2123.82	2371.91	10.46%

Table 4.4: Part Percentage Difference

	Early part estimation	Final parts	Difference
12781421	146	187	21.93%
12781423	198	279	29.03%
12781425	137	223	38.57%
12781225	156	268	41.79%
12651421	367	530	30.75%
12651423	242	407	40.54%
12651425	256	424	39.62%

4.1.2. Methods Used

The statistical analysis uses a correlation analysis to find the cost drivers and regression analyses to find the equation for the cost estimation. With the use of a scatter plot the type of relation a property has on the hours can be analysed. These methods are further explained in appendix I.

Relation Type Analysis

With scatter plots and trend lines the relation between the cost drivers and the production hours of the ship section can be found. The type of relation can be linear, exponential and logarithmic. With a trend line through the scatter plot points a coefficient of determination, R^2, can be generated. The R^2 shows the percentage of the production hours that is explained with the cost driver used. A higher value means that more of the production hours are explained by the variable. [8] Table 4.5 shows the R^2 values for the different cost drivers compared to the production hours.

Table 4.5: Coefficient for Determination for Different Relation Types per Cost Driver

	Linear	Exponential	Logarithmic	Relation
Parts	0.78	0.76	0.64	**Linear**
Profielen	0.91	0.87	0.80	**Linear**
Frames	0.81	0.76	0.64	**Linear**
Collar Plates	0.73	0.67	0.52	**Linear**
Brackets	0.45	0.51	0.33	**Exponential**
Weigth	0.62	0.64	0.55	**Exponential**
Weld length	0.73	0.72	0.71	**Linear**

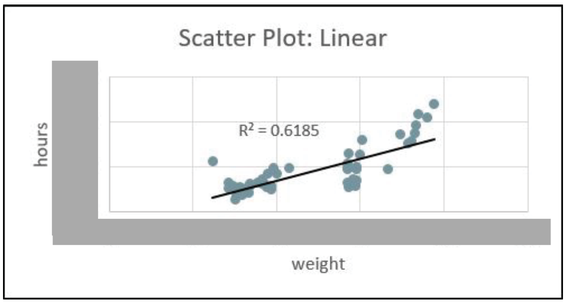
The results of this analysis shows that parts, profiles, frames, collar plates and weld length have a linear relation towards the production hours and brackets and weight an exponential relation.[1] The weld length coefficient for determination is based on seven sections the other variables are based on 60 sections. The statistic analysis tries to fit the trend line to the data used for the analysis. Only trusting the R^2 value might therefore lead to the wrong assumptions. There might be outliers or the variable shows a high correlation but no causality to the dependent variable. [13]

[1]The scatter plots can be found in appendix P till appendix S

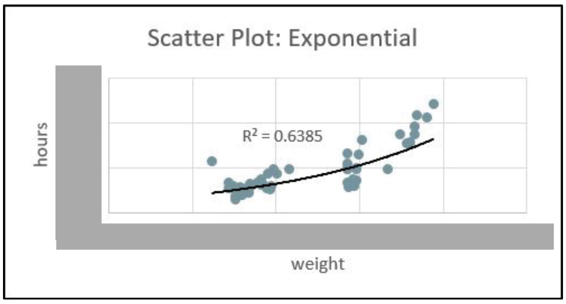
For each variable it is necessary to know what type of relation they have to know what type of regression analysis to use. A few criteria can be found for the type of relation: the R^2 values, the causality the variables have and finally a linear relation is preferred over an exponential equation. Iversen [13] says in his book that from a straight line it is easier to determine which values are deviant then from a curved line. Variables with a linear relation to the dependent variable are used in a linear regression analysis, variables with an exponential relation in a non-linear regression analysis. Also the linear regression analysis is preferred over a non-linear regression analysis, because it is easier to understand and to use then the non-linear regression analysis. Only if the data set cannot be explained by a linear relation the non-linear regression analysis can be used. [9]

When looking at the total amount of parts or the different type of parts it is logical that there is a linear relation towards the production hours. An increase in the amount of parts lead to an increase with the same ratio of the production hours. If an extra part is added, whether this is a profile or a plate, this part is assembled to the section by welding it together. When adding one part, for instance a profiles, to a section does not lead to the hours increasing exponential. Also the scatter plots showed evenly distributed data points for the different parts. That is why for all parts a linear relation is chosen. The weld length is based on the same criteria, the weld length increases the hours with the same ratio and the data distribution is evenly spread. Therefore also the weld length has a linear relation to the production hours.

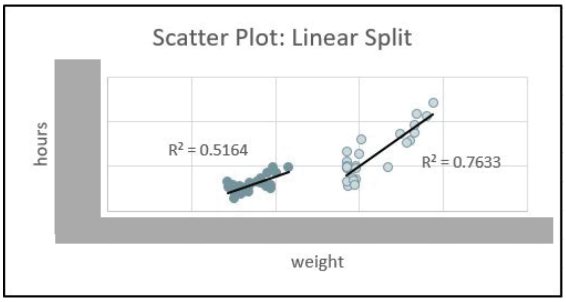
In terms of the R^2 value the weight is exponential. The data distribution is however unevenly spread. Figure 4.1a shows the scatter plot of the weight data with a linear trend line and figure 4.1b shows the scatter plot with an exponential trend line. The data in the scatter plots show that there is a gap between the section weight around 50. This can be because higher sections need more cranes or more people to assemble the section. Figure 4.1c shows the two data clouds separated from each other. For this graph the R^2 values of the linear trend lines are higher than the exponential values. However only the R^2 values of the weight after 50 is higher than the trend line over the complete dataset. Because it is possible to have a linear relation with weight and the hours the linear regression analysis is used for weight as well.



(a) Weight: Linear Trend Line



(b) Weight: Exponential Trend Line



(c) Weight: Linear Trend Line Data Split

Linear Regression Analysis

The regression analysis results in equations to calculate the hours. The linear equations can be built up in the following ways:

- General: $Hours = \beta + x * \alpha$
- With multiple variables: $Hours = \beta + x_1 * \alpha_1 + x_2 * \alpha_2 + x_3 * \alpha_3$

Where: β is the intercept value from the regression analysis, x_n are the property values from the section data, for instance the amount of parts, and α_n are the property coefficients given in the regression analysis.

The statistic method of estimating costs uses these equations. There will be a set of three different type of equations, one that can be used in the beginning of the BVC phase, one in the end of the BVC phase and one that can be used if the welding and ironwork hours have to be estimated separately. The number of variables taken into account for the regression analysis are the three highest correlating variables in regard to the hours. These three variables are analysed combined, in pairs and individually. For the linear regression analysis the P-values are used as indicators to show that the variables and their coefficients are significant.

4.2. Scenario 1: Early BVC Estimation

The early BVC estimation has been carried out to see the correlation between the less detailed section properties, to match the first part of the BVC phase. These properties are the weight, the total parts and the total weld length of the section. First the ranking of the independent variables is analysed and then the linear regression analysis is performed.[2]

Cost Drivers

The cost drivers are ranked by the results of the correlation analysis. As can be seen in table4.6, weight scores the lowest, weld length second and the parts show the highest correlation towards the final production hours. A high correlation does not necessary mean a high causality, but is does show the statistical effect a variable has on the total hours.

Table 4.6: Cost Drivers during the Early BVC Phase

	Part		Weld Length		Weight	
6 sections	1	0.96	2	0.89	3	0.80
7 sections	1	0.96	2	0.86	3	0.78
60 sections	1	0.89			2	0.79

Linear Regression Analysis

The linear regression analysis results in equations for the cost estimation. For the early BVC phase the equations that were used to estimate the costs are listed below. Table 4.7 shows the results for two variables. The current BVC estimations are used as comparison. For 60 sections the results where not accurate because only parts and weight could be used. For six and seven sections using two variables showed the best results with the lowest P-values as discussed below. Appendix P shows the total linear regression analysis for the early BVC phase.

The calculated percentages in table 4.7 show the difference in percentage between the total amount of hours spent on the section and the calculated hours estimated with the data from the regression analysis. The variables that are used in the equations are also shown in the table. It also shows the difference between the hours estimated during the BVC and the total amount of hours spent on the section.[3]

Table 4.7: Results Linear Regression Analysis Early BVC Estimation

		Part & Weight	Part & Weld Length	Weight & Weld Length
6 sections	**Estimated**	6.29%	6.41%	5.37%
	BVC	16.66%		
7 sections	**Estimated**	6.46%	6.40%	6.27%
	BVC	20.65%		

[2]The results belonging to this analysis are given in appendix P

[3]The values in the tables shown in this chapter show the absolute differences to the total realized hours. In general when the separate section hours are summed up it gives a quite accurate estimation of the total hours for all the sections, but to show how much the BVC and statistic estimations are off from the realized hours the absolute values are taken.

1. **Six sections:** Using three variables, parts, weld length and weight, gave the best results. The costs estimated where 5,16% off from the total hours spend on the section. However when looking at the P-values shown in the appendix using these three properties leads to high P-values. Using two properties gives better P-values and hour estimations in the same range as with three variables. The best P-values are generated using the weight and the total weld length as properties for the equation. Using these two properties also generates the best cost estimation from the three regression analyses for two variables.

   $Hours = -85 - x_{weight} * 0.05 + x_{weldlength} * 1.70$

2. **Seven sections:** Again for seven sections using three variables leads to the best cost estimation but high P-values. Same as with the regression analysis of 6 sections using two variables leads to lower P-values and just a slightly increased cost estimation difference. Using weight and weld length also shows for two variables the best P-values.

   $Hours = -15 - x_{weight} * 0.05 + x_{weldlength} * 1.75$

As can be seen the two equations for 6 and 7 sections do not differ that much. The intercept value for 6 sections is lower than the one for 7 sections and for 7 sections the factor that is multiplied by the weld length is a 0.05 higher than the factor for 6 sections. The equations both have subtract the weight from the intercept value, which might lead to wrong estimations if the weight and weld length differ to much from each other. If the weight increases and the weld length decreases in early stages of the BVS phase the equation might result in negative hour values. Therefore a boundary condition is set for the equations.

$$x = \frac{weight}{weldlength}$$

$$20\% < x < 30\%$$

This ratio is based on the current differences between the weld length and the weight. If the values for weight and weld length are not within this ratio the equations cannot be used.

4.3. Scenario 2: Final BVC Estimation

The final BVC estimation has been carried out to see the correlation between the variables available at the end of the BVC phase. These properties are the weight, the total parts and the specific type of parts and the total weld length of the section. During this phase also properties of the welding are known, such as the type and the percentage of watertight welds. First the ranking of the independent variables is analysed and then the linear regression analysis is performed.[4]

Cost Drivers

First the cost driver analysis is performed. The results are shown in table 4.8. The detailed welding properties could not be tested with the cost driver analysis because the data known was to limited to receive accurate results.

Table 4.8: Cost Drivers during the Final BVC Phase

	Part		Weld Length		Weight		Frames		Brackets		Collar Plates		Profiles	
6 sections	2	0.96	4	0.89	5	0.80	7	0.71	3	0.96	6	0.77	1	0.96
7 sections	1	0.96	3	0.86	5	0.77	7	0.73	8	0.62	4	0.78	2	0.93
60 sections	3	0.89			5	0.79	2	0.90	6	0.67	4	0.85	1	0.95

Linear Regression Analysis

The linear regression analysis results in equations for the cost estimation. For the final BVC phase the equations that were used to estimate the costs are listed below. Table 4.9 shows the results for two variables for six, seven

[4] The results belonging to this analysis are given in appendix Q

and 60 sections.[5] Using two variables showed the best results with the lowest P-values as discussed below. Appendix Q shows the total linear regression analysis for the early BVC phase.

Table 4.9: Results Linear Regression Analysis Final BVC Estimation

		Input		
6 sections	Variables	Part & Profile	Part & Bracket	Profile & Bracket
	Estimated	5.37%	5.31%	5.09%
	BVC	16.66%		
7 sections	Variables	Part & Profile	Part & Weld Length	Profile & Weld Length
	Estimated	5.17%	6.40%	9.19%
	BVC	20.65%		
60 sections	Variables	Part & Profile	Part & Frames	Profile & Frames
	Estimated	12.91%	15.95%	12.86%
	BVC	16.21%		

1. **Six sections:** The equations for 3 and 2 variables give results in the same range. The lowest p-values were met when using number of profiles and number of brackets in the same equation. However this still led to relatively high p-values in comparison to the p-values resulting from scenario 1 and the p values from scenario 2 using 60 sections.

   $Hours = -105 + x_{profiles} * 3.80 + x_{brackets} * 7.25$

2. **Seven sections:** The equations for 3 and 2 variables give results in the same range, except for the 2 variable equation using number of profiles and weld length combined. The lowest p-values are met when using the parts and weld length in the same equation. Again this still leads to relatively high p-values.

   $Hours = 220 + x_{parts} * 2.20 - x_{weldlength} * 0.04$

3. **60 sections:** The most accurate results are given by the calculations using 3 or 2 variables, except for the calculation using properties parts and frames combined. The best p-values are met when using profiles and frames combined. So the most accurate equation is the 2 variable equation using both number of profiles and number of frames.

   $Hours = -95 + x_{profiles} * 5.20 + x_{frames} * 4.70$

The equations resulting from these analyses are different from each other and also the properties used in the equations are different. The 6 and 7 sections equations give results in the same range as the scenario 1 equations, only for the scenario 1 equations the P-values are lower. The preferred equations for 6 and 7 sections are thus the ones discussed in scenario 1. For 60 sections the equation has low enough P-values to assume that the regression coefficients are statistically relevant. This equation gives hour estimations that are around 4% lower than the hours estimated during the BVC phase.

4.4. Scenario 3: Final BVC Estimation: Welding and Ironwork Hours Split

The third scenario has been carried out to see the correlation for the different type of hours, the welding hours and the ironwork hours. In this way the hours can be estimated separately from each other. The variables known for this analysis are the same as for scenario 2. First the ranking of the independent variables is analysed and then the linear regression analysis is performed.[6]

Cost Drivers

First the cost driver analysis is performed. The results are shown in table 4.10.

[5] For explanation of the percentages see scenario 1: early BVC phase.
[6] The results belonging to this analysis are given in appendix R

Table 4.10: Cost Drivers during the Final BVC Phase: Iron and Welding Hours split

		Part		Weld Length		Weight		Frames		Brackets		Collar Plates		Profiles	
6 sections	**Iron**	3	0.69	4	0.55	5	0.44	7	0.37	1	0.84	6	0.38	2	0.81
	Weld	1	0.98	3	0.94	5	0.86	7	0.76	4	0.93	6	0.84	2	0.94
7 sections	**Iron**	3	0.66	4	0.42	6	0.32	7	0.31	1	0.72	5	0.33	2	0.66
	Weld	1	0.97	3	0.92	5	0.85	6	0.79	7	0.54	4	0.86	2	0.93
60 sections	**Iron**	3	0.77			5	0.61	2	0.84	6	0.60	4	0.69	1	0.85
	Weld	2	0.77			5	0.61	3	0.88	6	0.67	4	0.88	1	0.95

Linear Regression Analysis

The linear regression analysis results in equations for the cost estimation. When the hours are split into welding hours and iron work hours it could be that the summation of the two estimation results in different estimations than the total hour equations. Table 4.11 shows the results for two variables for six, seven and 60 sections if the equations for iron and welding hours are summed up.[7] Using two variables showed the best results with the lowest P-values as discussed below. Appendix R shows the total linear regression analysis for the early BVC phase.

Scenario 3 uses the same properties as for the total hours. but the hours spend on the sections are split in iron work hours and welding hours.[8] This scenario shows if a more accurate cost estimation can be made by splitting the type of hours. The table shows the variables that are used for the regression analysis. When there are two variables listed behind each other this means that one is used for the equation of the iron hours and the other for the welding hours.

The welding hours are a summation of the welding hours, the panelstreet hours and the panelstreet hours. The panelstreet hours contain both ironwork and welding hours, however most of the work performed in the panelstreet are welding activities. The panelstreet hour administration is not divided in ironwork hours or welding hours so for this research the hours are combined with the total welding hours.

Table 4.11: Results Linear Regression Analysis Final BVC Estimation: Hours Split

		Input		
6 sections	**Variables**	Part & Profile	Part & Bracket/Weld Length	Profile & Bracket/Weld Length
	Estimated	7.34%	6.64%	6.63%
	BVC	16.66%		
7 sections	**Variables**	Part & Profile	Part & Bracket/Weld Length	Profile & Bracket/Weld Length
	Estimated	5.72%	6.19%	7.01%
	BVC	20.65%		
60 sections	**Variables**	Part & Profile	Part & Frames	Profile & Frames
	Estimated	12.91%	15.95%	12.86%
	BVC	16.21%		

1. **Six sections:** The two variables, for ironwork profiles and brackets and for welding profiles and weld length, give the best results in terms of hour estimation. Also in terms of p-values using these variables gave good results.

   $Hours_{ironwork} = 90 + x_{profiles} * 0.35 + x_{brackets} * 2.60$

   $Hours_{welding} = -280 + x_{profiles} * 3.00 + x_{weldlength} * 0.20$

2. **Seven sections:** For seven sections using 3 variables gives the most accurate results. But using 3 variables led to high p-values. The lowest p-values are met when using two variables, profiles and brackets, for ironwork and for welding profiles and weld length the same as for 6 sections. However these cost

[7] For explanation of the percentages see scenario 1: early BVC phase.
[8] The welding hours do not contain the robot hall hours for scenario 3 and 4 which leads to a small difference in BVC percentages.

estimations led to a higher error percentage. This percentage is still more than half of the BVC estimated hour error percentage, so these equations are used.

$Hours_{ironwork} = 80 + x_{profiles} * 0.7 + x_{brackets} * 1.80$

$Hours_{welding} = -145 + x_{profiles} * 3.00 + x_{weldlength} * 0.15$

3. **60 sections:** Both 3 variables and 2 variables give results in the same range. Looking at the p-values using frames and profiles for both ironwork and welding gives the lowest p-values. Also in terms of cost estimations these results are accurate.

   $Hours_{ironwork} = 22 + x_{profiles} * 1.00 + x_{frames} * 2.50$

   $Hours_{welding} = -115 + x_{profiles} * 4.15 + x_{frames} * 2.20$

The equations can be used to estimate the hours for ironwork and welding separately. The equations for welding for 6 and 7 sections are in the same range. The ironwork equations differ a lot. Preferable is to take the 6 or 60 section equations because they are only based on one type of ship section.

4.4.1. Scenario 4: Welding and Ironwork hours split: Weld Factor Example

This scenario has the same goal as scenario three only a different approach has been used. Because the weld length and the position of the welder are known the welding hours can be calculated with the welding factor known at Royal IHC. The difference that remains can be an extra difficulty factor of the section. The variables known for this analysis are the same as for scenario 2. First the ranking of the independent variables is analysed and then the linear regression analysis is performed.[9]

Cost Drivers

First the cost driver analysis is performed. The results are shown in table 4.12. This scenario could only be tested for six and seven sections because there is no welding information known for 60 sections. The difference in the table is the corralation of the variables towards the welding hours minus the weld length times the welding factor.

Table 4.12: Cost Drivers during the Final BVC Phase: Welding Factor

		Part		Weld Length		Weight		Frames		Brackets		Collar Plates		Profiles	
6 sections	Iron	3	0.69	4	0.55	5	0.45	7	0.37	1	0.84	6	0.38	2	0.81
	Difference	2	0.83			4	0.58	6	0.53	3	0.79	5	0.55	1	0.85
7 sections	Iron	3	0.66	4	0.42	6	0.32	7	0.31	1	0.72	5	0.33	2	0.66
	Difference	1	0.80			6	0.43	5	0.45	3	0.68	4	0.49	2	0.70

Linear Regression Analysis

The linear regression analysis results in equations for the cost estimation. Table 4.13 shows the results for two variables for six and seven sections if the equations for iron and welding hours are summed up.[10] Using two variables showed the best results with the lowest P-values as discussed below. Appendix R shows the total linear regression analysis for the early BVC phase.

The idea of this scenario is that Royal IHC already has standards they use to calculate the weld duration in terms of meters per hours. These are based on the welder or weld machine position. PA and PB is underhand welding and has a duration of 4 meters per hour, PF is vertical welding and has a factor of 2 meters per hour. These factors also include the set up time of the welder. Some sections gave good results, mainly sections that are easier to built. The other sections are less straight forward. The variables that have a high correlation with the welding difference are shown in table 4.12. It is seen as a sort of difficulty factor for the welding hours. Because for 60 sections no detailed welding information is known this analysis can only be performed for six ans seven sections.

[9] The results belonging to this analysis are given in appendix S

[10] For explanation of the percentages see scenario 1: early BVC phase.

Table 4.13: Results Linear Regression Analysis Final BVC Estimation: Welding Factor

		Input		
6 sections	**Variables**	Part & Profile	Part & Bracket	Profile & Bracket
	Estimated	5.98%	7.03%	5.95%
	BVC	16.66%		
7 sections	**Variables**	Part & Profile	Part & Bracket	Profile & Bracket
	Estimated	8.59%	6.16%	5.46%
	BVC	20.65%		

1. **Six sections:** The most accurate results were given when using 2 variables, especially using bracket and profiles calculating the difficulty hours for welding and the ironwork hours. The p-values of all the different equations are high so choise is to use the equation giving the best cost estimation.

   $Hours_{ironwork} = 90 + x_{profiles} * 0.35 + x_{brackets} * 2.60$

   $Hours_{welding} = L_{PA,PB}/4 + L_{PF}/2$

   $Hours_{difficultywelding} = -275 + x_{profiles} * 2.80 - x_{brackets} * 0.40$

2. **Seven sections:** The same as explained for 6 sections holds for 7 sections. Only now 3 variables give the best results, but comparing these p-values to the p-values for 2 variables, the 2 variable analyses give lower values. So chosen is to use the 2 variables brackets and profiles.

   $Hours_{ironwork} = 80 + x_{profiles} * 0.7 + x_{brackets} * 1.80$

   $Hours_{welding} = L_{PA,PB}/4 + L_{PF}/2$

   $Hours_{difficultywelding} = -250 + x_{profiles} * 1.40 + x_{brackets} * 2.70$

The equations for the welding difficulty show a lot of difference between 6 or 7 sections. For the difficulty equation data from more sections are needed to say anything about the equations and whether it finally leads to more accurate results using the weld factor and a difficulty factor.

4.5. Section Type Effect

The difference between six or seven section is used to see if a different type of section has an influence on the cost estimation. Currently at Royal IHC they estimate hours by using different factors for each section type. For an actual analysis of multiple types of ship section the data available is to little, therefore only the results of six and seven sections could be used. Looking at the differences between these analyses shows that when using the equations following from the regression analyses, they result in the same range of error percentages. When the BVC estimation increased with 4% adding an aft ship buoyancy compartment the statistic equation results in the same range of error percentage. These equations are however made by using the data of only these 6 or 7 sections. So it is again recommended to perform similar analyses when more data is available to see if it is possible to estimate the costs of different type of sections using one equation. If this is not the case other equations have to be made for each section type.

4.6. Standard Error

To see if the equations from the regression analysis are significant in the current analysis the P-values have been used. Another way to see if a regression analysis provides accurate results is by looking at the standard error. By looking at the equation for the standard error it can be seen it is strongly linked to the number of oberservations toghether with the standard deviation of the measurements.

$$SE = \frac{SD}{\sqrt{n}}$$

This means that a lower standard deviation divided by a large number of observations results in a lower standard error. From each analysis with the same number of sections, the number of observations is constant, it can be seen which combination of variables has the lowest standard deviation, and thus generates results that are close to the expected result. [8]

By looking at the results from the different cost driver analyses, the standard error is the lowest for the regression analyses in which also the P-values are either in the same range as the lowest standard error values or are the lowest. This means that they indicate the same equations as validated with the P-values in this chapter as the most significant equations be used.

4.7. Conclusion

Subquestion 3 - **"What are the most important aspects to look at during the production of a ship to create a realistic cost estimation?"**

The most interesting result of the correlation analysis is that the property that Royal IHC bases most of their current cost estimation on, weight, does not have the highest correlation to the hours spent on the sections. The ranking of the properties per equation phase are shown in table 4.14. This ranking is composed by multiplying the correlation coefficients from the cost driver analyses times the amount of sections used leading to the following ranking of cost drivers.[11] The amount of parts, mainly profiles and brackets, and the total weld length have a higher relation to the total amount hours spend on the sections. A research performed by P. van der Horst [34] on the reduction of parts between order 01265 and 01266 concludes that a reduction on the amount of parts can result in a reduction of production hours and costs. During this research multiple parts were combined in one part which led to a reduction in the amount of parts but an increase of weight. The reduction of parts led to a production cost reduction of €58.000,- and a reduction of €6.000,- for Metalix. The increase of weight resulted in a material price increase of €2.100,-, which is a small increase compared to the reduction in production and Metalix costs. This research supports the conclusion of this statistical analysis that parts have a higher correlation between the production hours spent on a section than the weight of the section has.

Table 4.14: The Ranking of the Properties per Scenario following from the Corralation Matrix

	Early BVC Estimation	**BVC Estimation**	**BVC Estimation: Hours Split**		
	Total	Total	Ironwork	Welding	Welding Difference
1	parts	profiles	profiles	profiles	parts
2	weld length	parts	parts	parts	profiles
3	weight	weld length	brackets	weld length	brackets
4		frames	frames	collar plates	collar plates
5		collar plates	collar plates	weight	weight
6		weight	weld length	frames	frames
7		brackets	weight	brackets	

It should be noted that the properties discussed in this chapter are section properties and not plate or part properties. These properties need to be assessed when more information is available from more ship sections. For the correlation analysis properties such as weld type and weld thickness could not be tested yet. These variables may however have a significant relationship to the total amount of hours spent on the section. For instance the research of P. van der Horst concludes that a higher application of SAW welding during the production process leads to lower costs [34]. These variables are not taken into account for the regression analysis because they did not show a relation high enough to be one of the three variables used for the equations. Table 4.14 shows the ranking for all the properties per scenario.

For the regression analysis the P-values are relatively high but the R^2 are also high, this could have multiple reasons. The sample size is too small. Using 60 sections shows lower P-values than for six or seven sections. Another reason could be that the variables have a high correlation to each other. A good example for this effect can be shown by looking at the analysis where parts, frames and profiles are taken into account. The regression analysis of only profiles and frames have low P-values in comparison to the regression analyses in which parts and profiles or frames are taken into account. Parts and profiles or frames have a high correlation to each other resulting in higher P-values. Recommended would be to perform the regression analysis again

[11] The results belonging to this analysis are given in appendix T

when more data is available for more sections to see if increasing the sample size results in lower P-values giving more significant results. When increasing the sample size does not lower the P-value it would be recommended to use other less correlating properties in the regression equations. For now the equations that show the lowest P-values along with the best cost estimation is taken into account. [8]

Appendix O shows the summary of the regression analysis. The equations resulting from the regression analyses for scenarios 1, 2, 3 and 4 show that using different properties than only weight results in a more accurate cost estimation. This is a conclusion drawn by comparing the BVC calculation IHC made for the sections and the statistic equations that where the result of the regression analysis. The different equations will be verified in chapter 8, when the production hours of a ship section are not included in the statistic analysis.

During the analysis a few equations turn out to be more accurate and result in lower P-values than others. When a cost estimation has to be performed with rough data, in the beginning of the BVC phase, it is preferred to use the 60 sections equations of scenario 2 or 3. These equations have more data to base their estimation on and have lower p-values. The scenario 2 equation can be used to calculate the total production hours. Scenario 3 equations can be used to calculate iron work hours and welding hours separately. During chapter 8 the difference between the equations resulting from 60 sections and 6 sections can be evaluated. For instance if a rough estimation of weld length is also available the difference between outcomes of using equation of 6 sections from scenario 1 and 3 and 60 sections of scenario 2 and 3 can be estimated. The 6 section equation from scenario 1 is preferred over scenario 2 because this led to a lower P-value and the same range of error percentages. Scenario 3 can be used to see if ironwork hours and weld hours need to be calculated separately. If more detailed information is available the same equations can be used to estimate the hours for the new section. By comparing these results to the rough data hour estimations, conclusions can be drawn if the equations also produce accurate cost estimations if the data available is still rough data. Scenario 4 gave accurate results but to use the equation from this scenario, detailed information from the welding position is needed and due to time constraints it is not possible to gather this information from a new section.

5

A Case Based Reasoning Cost Estimation

This chapter introduces the Case Based Reasoning theory and how this theory can be used for estimating ship section production costs. Section 5.1 discusses the different parts of the CBR theory. Section 5.2 discusses how the theory could be implemented as a theory behind the cost estimation method for the ship building industry.

5.1. Case based Reasoning Theory

The case based reasoning theory is based on three aspects that are introduced by the name of the theory: [24]

1. **Case:** a case is a previous task that has already been solved.
2. **Based:** based means that the connection between reasoning is made on cases. Meaning that the reasoning for a new solution is based on old cases.
3. **Reasoning:** reasoning stands for the way a solution is found. In CBR this reasoning is based on previous cases.

The CBR method tries to find solutions for new tasks by looking at already solved tasks. This old case does not necessarily have to be identical to the new case to be able to reuse the solution of the case. [24] This section explains the CBR method and the main aspects of finding a solution with the CBR method in general. Discussed will be what the cases are and how these can be represented. How similarity works and how this can be implemented in the search for a solution. How a solution can be retrieved from the old cases and how evaluation of solutions is possible.

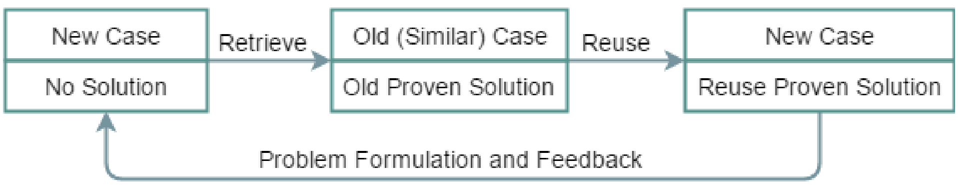


Figure 5.1: Basic Case Based Reasoning Process steps

Figure 5.1 shows the basic CBR process of how to find a solution to a new problem based on an old case. A new problem is encountered and there is no solution yet. By retrieving this old problem from the old experience a new solution can be found. With the use of older cases, that are identical or similar to the new case, a solution can be composed by reusing the old solution. This solution is used for the new case and feedback is made on that solution. This new case is then added to the experience list.

5.1.1. Cases & Case Representation

For the CBR method to work properly experience is needed. One requirement is that there are already solved cases on which the new case solution can be based. Not every experience might be needed for the solution of the new case. There may be three sorts of cases. Cases not useful for the new case because there is a lack of similarity, experience similar or close to the new solution and exact replicas of the new case. Though not every experience might be useful for finding a solution for an old case, every experience is useful for administration and to build up a list of experience. [24]

The total experience can be a list of many different cases that have already been encountered. Cases can be very complex and are therefore divided into parts. Two types of parts can be distinguished, problem parts and solution parts. A problem part is an explanation or a part of the explanation of the problem. A solution part is a part of the solution that was chosen to solve the case. There can be positive cases with solutions that solved the problem and negative cases where the solution could not solve the problem. Both of these cases are useful as a reference for new cases. [24]

The cases are represented most easily by using ’feature-value pairs’. The features are the subject and the value is for instance the number or weight of the subjects. An example: "the bag is blue" the feature is the colour of the bag and the value is blue. The features must be explained for both the problem and the solution. If cases need to be more detailed, other representations are possible.[1] [24]

The representations of cases can also be organised in different ways. According to Richter there are three types: flat, structured and unstructured. Flat structure is meant to organize a small number of cases. It is a way to structure cases in forms of tables. In these tables features and solutions are listed on the left. All the different cases and their values are listed right from the attributes (features and solution). If there is a large number of cases the structured type is a good way of organising them. The structured type organises cases as ’hierarchies’ or ’networks’. Cases can also be written in text form, this is an unstructured way of organising cases, because the attributes and values are hidden in the text. [24]

5.1.2. Retrieval & Similarity

To retrieve a solution for a new problem based on an solved case, it is important to search for cases similar to the new problem. The extent of similarity between two cases is reached when a solution of a previous case can be used for a new problem with limited changes. The extent of similarity varies for every solution search. The most important thing is to find out which case from the experience has the most suitable solution for the problem. The historical cases are organised and stored in a case base. To search for a possible solution in a case base the problem has to be in the same format as the previous cases. Then a comparison can be made between the cases in the case base and the problem. This process is called the ’similarity assessment’. According to Richter [24] if cases are represented by attribute-value pairs the extent of similarity depends on the different attributes and the relative relevance of these attributes. Attributes can have a specific rate to multiply the similarity with. A more relevant attribute has a higher rate then a less relevant attribute. The difference between attribute values can be evaluated in different ways:

- 1 if the values are completely similar
- 0 if the values are not similar in any way
- a number between 0 to 1 as a ratio for similarity between the evaluated attributes

This ratio for similarity is multiplied with the ratio of relevance creating a number. All these numbers are summed up and divided with the sum of the relevance ratios. This summation represents the similarity between the experience case and the problem case. In this way a solution can be found with the CBR theory. [24]

An algorithm can be composed for this calculation. This algorithm is shown by equation 4.1 till 4.6. The goal of this algorithm is to find the optimum solution from the case base. The assumption is made that a case solution is optimal if there is no case with a higher similarity than the optimal case. According to Richter this

[1] For more information about different representation methods see the book of Richter chapter 5.

can be done by a neighbourhood search. With the principle that 'there is no case more useful than the nearest neighbour' - Richter. [24]

Cb: set of experience cases

p: problem case

t: case within the Cb

t$_{sim,i}$**:** the similarity of case i

a$_i$**:** attribute of a case t, where i is the case number

y$_p$**:** attribute of the problem case

T: optimal solution case

q$_j$**:** relevance ratio of attributes, where j is the attribute number

x$_j$**:** the similarity ratio between the case attribute and the corresponding problem attribute

$$t_i \in Cb \tag{5.1}$$

$$a_j = a_1, a_2, \ldots, a_n 1 \tag{5.2}$$

$$x_j = (a_1 \times y_1), (a_2 \times y_2), \ldots, (a_n \times y_n) \tag{5.3}$$

$$sim_1(x_1 \times q_1), sim_2(x_2 \times q_2), \ldots, sim_n(x_n \times q_n) \quad | \quad x_j, q_j \in dom(a_j) \tag{5.4}$$

$$t_{sim,i} = \frac{1}{\sum q_j} \times \sum sim(x_j \times q_j) \quad | \quad 1 \leq j \leq n \tag{5.5}$$

$$t_{sim,i} = T \quad IF \quad t_{sim,i} > \{t_{sim,1}, t_{sim,2}, \ldots, t_{sim,n}\} \tag{5.6}$$

Result: The optimal solution for p from Cb, called T. There is no t with a higher similarity to p then T.

5.1.3. Evaluation of Solutions

Reusing a solution is simple when the case identical to the historical case and if that solution worked in the past. When there is some deviation between the problem case and the experience case the solution has to be evaluated and the new solution has to be adapted to fit the problem case. There are three steps explained by Richter taken in the CBR process: [24]

1. Describe the problem.
2. Perform the similarity assessment and look for the closest neighbour in the case base.
3. Use the retrieved solution for the problem by copying or adapting it.

Using the CBR method can create the risk of using a solution that will not suffice as a solution for the new problem and the risk of having limited similar cases in the case base. The chance of these risks can be minimized by using up-to-date cases and frequently evaluating the case base on relevant cases. Cases can become irrelevant when they become older. For instance if they use different techniques or outdated data. [24]

5.2. Case Based Reasoning Used for Ship Cost Calculation

Royal IHC has performed many previous shipbuilding projects and has built up an experience list in this field. Every vessel they have built can be seen as a project with different tasks or systems. The current cost calculation methods at Royal IHC use historical data for the estimations. However not every cost estimation phase of the current method uses the CBR theory. Other types of cost estimation methods are discussed and examples of CBR cost estimations are given in this section. Also the use of the CBR theory for a section building cost estimation method are discussed.

Cost Estimation Methods

Cost estimation methods can be divided in algorithmic and non-algorithmic methods. Algorithmic cost estimation methods use equations based on historical data and the main cost drivers to calculate the costs. A type of an algorithmic method is the parametric cost estimation method. [17] This research will introduce a parametric cost estimation, the statistical method. Other types of algorithmic methods are commonly used in the software industry and often use an equation with the lines in the software code as a function of the software development costs and effort. An example of this is the COCOMO method. [17]

Non-algorithmic methods can be: [17]

- **Expert Judgement:** the cost estimation is based on the experience of a expert or group of experts. In the past this method was used at IHC and still is if IHC has to perform a quick cost estimation.
- **Analogue Cost Estimation:** a CBR cost estimation method is an example of an analogue cost estimation. This method uses historical data of projects to find the solution for a new project.
- **The Bottom-up or Top-down approach:** top-down estimating starts with the cost estimation of the overall project and then zooms in on the different tasks. This method is often applied in early stages of the project, when the data available is still limited. The bottom-up approach starts with the cost estimation of the smallest tasks of the project, when combining these the overall project cost estimation is found.
- **Parkinson's Law:** the costs depend on the resources that are available during the project duration instead of the work that needs to be performed. This method will therefore determine the costs instead of estimating, it looks at the people that are available over the time of the project.
- **Price-to-Win:** estimates the costs of a project on the maximum price to win the project, in other words the budget of the client. The project is fit into a budget that matches the budget of the client. This might lead to delays.

The current methods of IHC use a mix of cost estimation types. The VC phase is based on a top-down approach combined with the price-to-win method and the analogue cost estimation method. When they need a cost estimation in limited time they use the expert judgement method. The BVC method uses also the top-down approach but for instance already specified on the section building activity. This top-down approach is combined with the analogy method and sometimes expert judgement.

The CBR theory would be characterised as an analogue cost estimation method. The third method that will be introduced in this research, the graph database method, is a method using the CBR theory to estimate the cost of a ship section. One example of the CBR theory as a cost estimation method is given in the research of K. Zima [36]. He performed a research on a CBR cost estimation method for the preliminary phase of a construction project. The type of project is different from a ship section production project, however the CBR method used is the same as discussed in this chapter. The research shows that it is possible to perform cost estimations using the CBR theory. However how accurate the cost estimations are depends on the size of the database and the similarity assessment.

Ship Section Cost Estimation

For cost estimation with the CBR theory it should be known when sections are similar in order to find a similar case and a cost function to reuse the hours. A cost function adapts the old solution for the reuse as a new case solution. In order to search for a similar section the cost drivers of the production process have to be found. A cost driver in the section building process is a property of this process that has a correlation to the realized hours at the end of the production process. The cost drivers with the highest correlation are known from the cost driver analysis.

5.2.1. Cases & Case Representation

A shipbuilding project is a big project that can be divided into different sub-cases with different attributes and values. As explained before the budget of the VC is allocated to the different production departments and tasks. The departments and tasks are called systems and can be seen as the sub-cases for the CBR cost estimation. For this project the ship section systems are discussed. An example of the division of systems is shown in figure 5.2. The different levels can be turned into cases for the CBR estimation. For this project a case represents a ship section. A case has different attributes such as total cost price or total steel weight of the section. Sections are built up by parts, such as plates or profiles. Each part in a ship section has its own attributes, for instance plate thickness or curvature. The red lines in figure 5.2 show the welding connections between the parts. These connections can be seen as different sub-cases with different attributes, type of weld, weld length, etcetera. The section cases can be combined into the hull section block, where all the different systems of a production project can be combined.

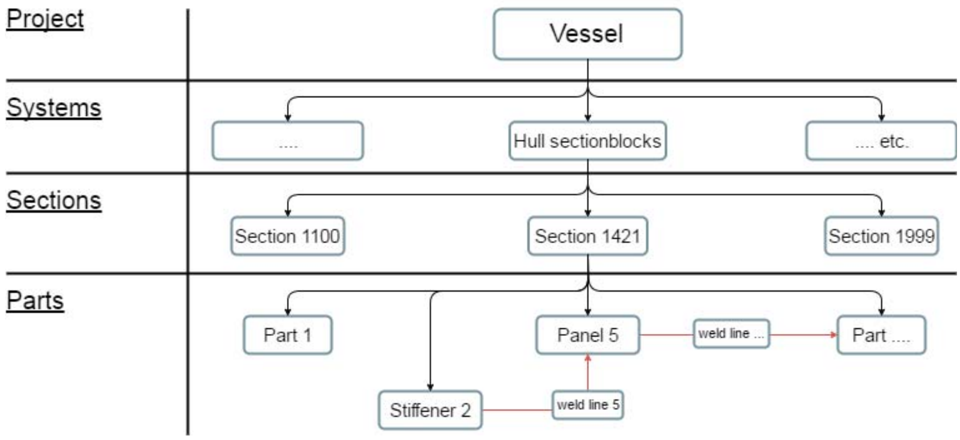


Figure 5.2: Ship Production Project Division

B. Delatte and A. Butler [6] discuss representation forms of cases in the ship building industry. They discuss the sub-case levels in their research which is different from the division shown in figure 5.2. The division and the level of detail of the sub-cases can be different for each problem solving case, but it is important that the division of sub-cases is appropriate for the project. They discuss three types of storages flat, relational and object oriented. The flat type is discussed earlier in this chapter and is not convenient for a case base of ship sections, because the time of analyses increases with the size of the case base. The unstructured type from the earlier discussion organises cases in text form, which is at IHC not the way they store data of ship sections. The structured way is preferred as storage type. Relational and object oriented are examples of structured case base storage types. The research of B. Delatte and A. Butler [6] conclude that the object oriented form of storage is preferred as a case base type for ships because it can store complex cases. A drawback of an object oriented storage type is that the relations between the attributes cannot easily be labelled. [22] Another database type might be preferred for a ship section case base. Chapter 6 discusses different types of storage options for a CBR cost estimation method for ship sections.

5.2.2. Retrieval & Similarity

Ship section can be represented by an attribute-value pair so the similarity assessment can be performed by a comparable method as discussed in section 5.1. For a similarity assessment and solution retrieval from the section case base the different attributes belonging to the different sections have to be listed. These attributes also have to be rated in terms of relevance for the solution.

The relevance ratio will depend on the problem. For this project the focus is on cost estimation. Therefore the relevance ratios will depend on which attribute has a higher correlation to the cost price. By ranking the cost drivers a similar section can be found, the solution can be adapted and used as a solution for the new case. In this research the solution will be a cost price for the new section.

The cost estimation is an iterative process in which more attribute values are generated or refined towards the end of the process. This results in a need to filter between the attributes of the different cases. If the plate thickness is not known yet, it is important to filter out these values from the experience case base because they are not useful at that point in the cost estimation.

Not only the costs of a section can be an output of a CBR estimation. Design principles such as the number of stiffeners can also be estimated by a CBR estimation. The output is the number of stiffeners instead of the costs and the input are the important aspects for the determination of the number of stiffeners. In the database information about costs can be stored but for instance also information about how many profiles or brackets are used, the sizes of these items and which welding attributes are used. These features can be used for the design base for new similar sections. This part of the database is left out of the research but can be explored in later researches.

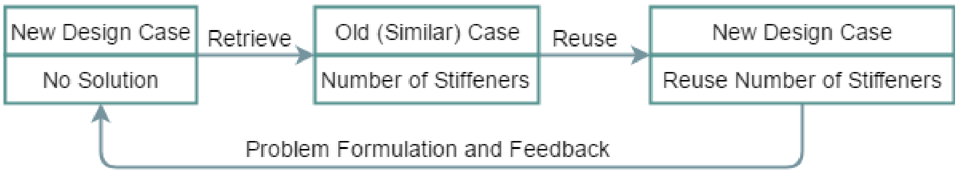


Figure 5.3: Case Based Reasoning process example to determine number of stiffeners for a new ship section

Y. Seo et al. [31] performed a research to create a block assembly planning with the use of the CBR theory. The basis of the research was to look for a more consistent solution method than expert judgement. The planning differed when different experts made the planning for the same block assembly. The research led to the conclusion that it is possible to create a more consistent assembly planning if the case base consists of cases that include their solution for the assembly planning. The CBR steps they performed were the same as discussed in this research. This research showed that also other solution searches are possible with a CBR method if the right data is included.

5.2.3. Evaluation

Copying a solution from previous cases might lead to wrong estimations. It depends on the similarity between the problem and historical sections in the case base. The amount of experience in the case base can also determine the relevance of the found solution. When the case base contains only a few cases the cost price might be based on wrong assumptions. Important for an accurate solution retrieval is: [24]

- Gather as much as possible accurate experience in the case base.
- Check the attribute values on old or expired data.
- Delete or adapt old data or cases from the case base.
- Place finished cases in the case base.
- Evaluate the relevance ratio of the different attributes frequently.
- Before placing a case in the case base evaluate the realised data and the estimated data for deviations in the data.

These actions lower the chance of wrong assumptions in retrieving and adapting a solution from the case base. The more data in the database the higher the validation of the solution is. More cases also show if there is some coherence between the different cost estimation of previous cases. If there is old and expired data in the case base this might lead to wrong assumptions but also to outliers in the similarity assessment. In ship production old data can be old processes or rules that have changed. Relevance ratios might also change in time due to changing rules and processes. So keeping data and estimation standards up to date is important for accurate results.

5.3. Summary

Subquestion 5 - **"How can the CBR method help to create a cost estimation?"**

The CBR theory tries to find a solution for a new case by reusing a solution from an old case. This theory uses a case base with historical data from old cases and tries to find a similar case to the new case in this data base. This can be used in a ship section cost estimation method. By creating a data base with the data of old ship sections a data set for solution seeking can be made. When the cost estimation of a new section is needed, a similar section can be found in the database. The solution of this section needs to be adapted and can then be reused as a solution for the new section.

6

Method 3: The Graph Database Method

The third cost estimation method that will be discussed is the graph database method. It uses the CBR theory to estimate the costs of ship sections. The method bases the costs of a new ship section on the realized production costs of old sections. In the first section an analysis is performed what the best type of data storage is for the ship section data. It concluded that the most preferred type of data storage is a graph database, that is why the third method is called the graph database method.

Section 6.2 explains what a graph model is and which types of graph models exist. It shows examples of graph models both to illustrate a simple graph model and to show how a ship section graph could be visualized. Section 6.3 explains the graph database principle and it shows how the graph database method can be used to estimate the costs of ship sections. This chapter also introduces the program that is used for the graph database, the node and relationship labels and it introduces the complete method of G. Alblas.

6.1. Different Types of Data Storage

In terms of data storage there are different methods. The so called SQL databases and the NoSQL databases. NoSQL databases can be categorized in four different types of databases: key value databases, column family databases, graph databases and document databases. An overview of the different types of databases can be found in figure 6.1. An explanation of the different database types can be found in appendix J. The different SQL and NoSQL methods and the advantages and limitations of SQL or NoSQL methods will be discussed in the appendix. This section draws the conclusion which database type is most preferred for a CBR cost estimation method of ship sections.

The database type should match the principles of CBR, so it should be possible to find a case similar to the new case and reuse the solution. It is also important to keep in mind that the database type should represent ship sections and preferably the work that is performed when building a section. The type of data storage must be able to store information of the section, such as part dimensions, weld length and the production hours. But it should also be possible to create relations from one part to another part of a section. These relations represent the weld lines between the different parts, for instance a profile welded on a plate. The data storage type should also be able to include labels to these connections, thus being able to create a relationship label such as the type of weld and the welding technique. To be able to implement both the CBR theory and to represent a ship section in a database, the different types of data storage are compared with each other by different aspects listed below.

- Being able to create relationships between items and being able to create a label with data added to the relationship.
- Store items of sections and being able to store data within these sections.
- For a proper analysis a normal computer a fast analysing speed should be able to perform the analysis.
- Being able to store big data sets.

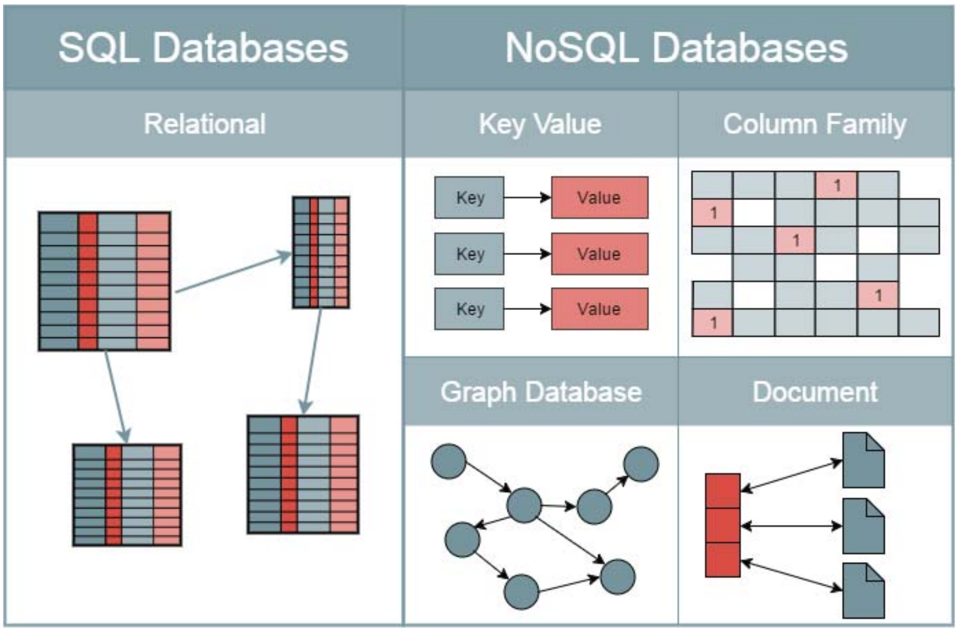


Figure 6.1: Different Types of Databases [25]

- The availability to analyse the database

Table 6.1 shows the analysis of the different aspects that are important for a database. As can be seen the graph database performs well on every aspect, but most importantly this storage type can represent a ship section including the amount of work. The relationships between the parts can have a label and thus weld line attributes can be included in the database.

Table 6.1: Analysis of Different Types of Databases

	SQL	NoSQL			
	Relational	**Key Value**	**Column Family**	**Document**	**Graph**
Store Data	+	++	++	++	++
Create Relationships	-	-	-	-	++
Analyzing Speed	-	++	++	+	+
Size of Data Set	-	++	++	+	+
Availability of Analysis	++	++	++	++	++

6.2. Graph model

As said before the graph model[1] is based on nodes and relationships between those nodes.[25] These nodes can be seen as the individual items. In case of ship section building a node represents a part of the section. The weld lines between these parts can be visualized as lines connecting the nodes. It does not necessarily mean that a node always represents the same part of a ship section, it could be a plate or a bracket. In each section graph there is one node that contains the data of the section itself, such as the total welding hours performed on the section. Different aggregation levels are possible for the graph database.

In this section the different type of graph models, labelled property graph, single-relational graph, hyper graph and triples, will be discussed. A conclusion will be made which one of these types is used for the research.

[1] Note that for the use of the database it is not necessary to know the exact algorithm behind the graph models. This will therefore not be discussed broadly during the report.

6.2.1. Type of Graph Models

In the graph databases book by Robinson [25] one type of graph model is discussed broadly, the labelled property graph. There are other types of graph models that are briefly discussed by the author or in other literature, the single-relational graph, hyper graphs and triples.

The labelled property graph is a type of graph model in which the relationships and nodes have labels. These labels are the attributes in the attribute-value pairs discussed. An example of a labelled property graph is given in figure 6.2. This figure shows two drivers, Alice and Bob, who together own three types of cars. The relationships shows if the drivers are owners or primary owners. Alice owns a Prius and primary owns a Mini and a Range rover and Bob owns a Mini and a Range Rover and primary owns a Prius. This is a simple overview of a labelled property graph where the figure shows that the nodes and relationships are labelled.

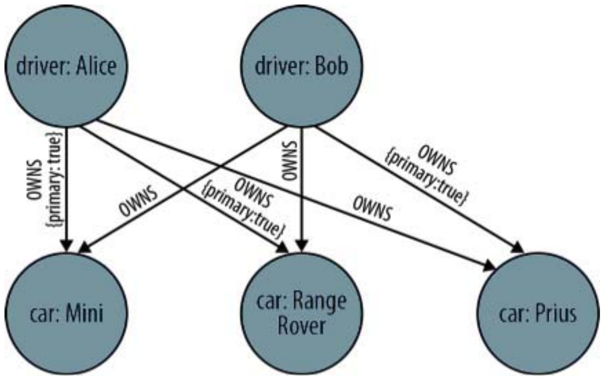


Figure 6.2: Example of a Labelled Property Graph [25]

Labelled property graphs have a few aspects:

- It contains nodes with corresponding relationships
- Nodes describe items
- Nodes can have different labels
- Relationships have a label and are directed
- Relationships have a start and end node
- Relationships can also contain properties

In single relational graphs all the relationships have the same meaning within the graph. A relationship can for instance be friendship between two people. All the relationships within that graph represent friendship. Nodes can represent different things, for instance if a relationship has the label friendship all the nodes will be people, but they can be different types of people. [26] Hyper graphs change from the labelled property graph in a way that one relationship can have more nodes to connect than just a starting node and an ending node. Figure 6.3 shows an example of a hyper graph. It is the same situation as the property graph, two drivers that own three type of cars. For a hyper graph less relationships have to be drawn to create almost the same effect of the labelled property graph. They tend to be more clear than the labelled property graph. However, it is not possible to create different relationships. So if it is necessary to known who is the primary owner of the three cars the hyper graph is preferred above the labelled property graph. [25]

Triples can show relationships between nodes with no clear connection or not coherent at first sight. An example that is given in the book of Robinson: 'Triples can capture relations such as 'Fred likes ice cream' and 'Ginger dances with Fred'.[25] This might not be interesting for small databases because no connection might be found between 'liking ice cream' and 'dancing with people'. However, for big data sets this could lead to more coherence between these two topics. Triples are used to find connections in large data sets. An example of a triple for a ship section could be connecting a welder to a weld line and adding a connection that the welder drinks coffee during his break. At first sight this might not be an interesting relation, but when more data are gathered about what the welders drinks during their breaks and what type of welds they perform interesting analyses could be carried out.

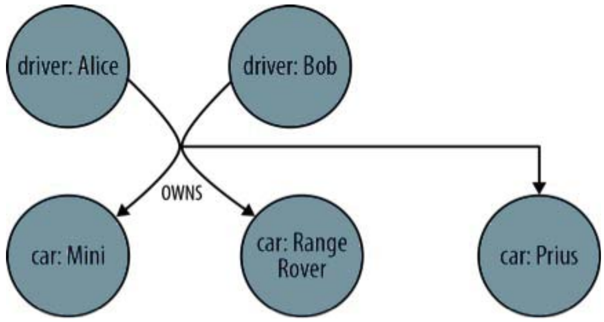


Figure 6.3: Example of a Hyper Graph [25]

For a graph database a few aspects are important, these are more detailed than for the data storage type. Important for the cost estimation is the level of detail in the information and the amount of labels that can be added to the graph. The visualization could be important for the communication and analysing of the graphs. It is also important that there are programs available to work with these types of graphs.

Table 6.2 shows the four types of graph models and how they perform on the important aspects for a graph database. For this study the labelled property graph is deemed to be the most suitable method to use. The single relational type is limited in the information that can be added to the nodes. Therefore the analysis of the database and the detail of the graphs score low in the table. Adding labels and properties to the relationships is not possible with a hyper graph. The hyper graph will therefore not be used in this research. The graphs consist of relations that have a dependent property such as a welding line between two plates, so there are no unspecified relationships between the nodes. Therefore, triples are not used for the graph model. The labelled property database scores well on every important aspect especially on the possibility to label a relationship between nodes.

Table 6.2: Analysis of the Four Graph Model Types

	Single Relational	Labeled Property	Hyper	Triples
Detail of the Graphs	-	++	-	++
Visualization	+	+	++	-
Analysis of the Database	-	+	-	++

6.2.2. Example ship Section Graph

A graph for a double bottom seen in figure 6.4 [5] could look like the graph shown in figure 6.5. In the graph the part between the two solid floors and the side of the ship to the centre girder is shown[2]. The direction of the relations is pointed to the part of the section that is welded on the other node, so if a stiffener is welded on a plate the direction is pointed to the stiffener.

It has to be noted that this section graph is just an example. It is not specified yet if the nodes will be labelled as profile types, such as longitudinal stiffeners or even more detailed inner bottom longitudinal stiffener, or plate types, such as keel plate or floor bracket. It can also be that the nodes will be labelled by two type of labels, plates and profiles and that the properties the node can have contain more detailed information about the profile or plate.

The goal of the ship section graph database is to find sections that are similar to the section that has to be built. This is done to find a similar solution for the new section. To create a graph that represents a ship section and to be able to calculate costs with these graphs a couple of labels and information has to be added to the nodes and relationships. This added information is called a label.[25] For the weld lines the welding length and welding techniques and for plates the type of the plate and the size of the plates may be a label added to the node.

[2] the number of longitudinal stiffeners and brackets differs from figure 6.4, the extra stiffeners and girders will be visualized the same way as the others already visualized

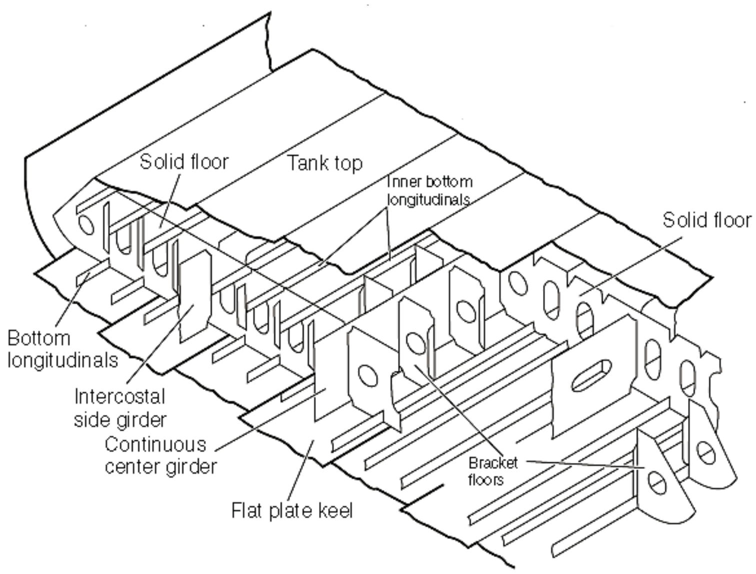


Figure 6.4: Longitudinal stiffened double bottom structure [7]

6.3. Graph Databases

"*A graph database management system is an on-line database management system with Create, Read, Update and Delete (CRUD) methods that expose a graph data model.*" [25]

A graph database has two properties that are important in discussing graph databases: [25]

- **The underlying storage:** If a database uses *native storage*, the database is optimized for storing and analysing graphs. [25]

- **The processing engine:** databases can use *index-free adjacency*, in these databases the nodes of the graph point toward each other. The free-index adjacency can be seen as the index-based search method. This search method is based on creating smaller data sets within a big data set to improve the search speed, in comparison to linear search which looks at all the data of the data set one by one. For instance if a number between 1 and 10 has to be found and the computer can only say higher or lower, it is better to split the search field in half then start with 1,2,... etcetera. [26] The index-free adjacency is seen as a native graph processing. [25]

The program that will be used as a database for the ship graphs is Neo4j. Neo4j combines native storage with native processing. This is preferable for a graph database of ship sections because they are optimized for the storing and analysing of graphs and the processing engine is quicker than the one of a non-native database. Neo4j is an open source graph database using a coding language called cypher. The Neo4j program and options of the program will be discussed in appendix K. The input and output data types will be listed and explained as well as the database options. The examples given are based on the information used in this research. How the information of the ship sections can be computed into graphs will also be explained in this appendix. Figure K.1 shows a section graph composed in the Neo4j database.

To analyse graphs in the database a compute engine is necessary. A graph compute engine can run a created algorithm in the graph database. An analysis can be performed by asking questions such as 'How much is the difference between budgeted costs and realised costs for a section?' or 'What was the total realised welding time of a section?'.[25] However Neo4j was not built with the intention to be used as a graph compute engine, it is still possible to create algorithms to analyse the database. There are other programs available that contain algorithms in which can be used to perform analyses.[4] Further research has discover if there are enough extensions available to use Neo4j for both the storage of graphs and as a graph compute engine.

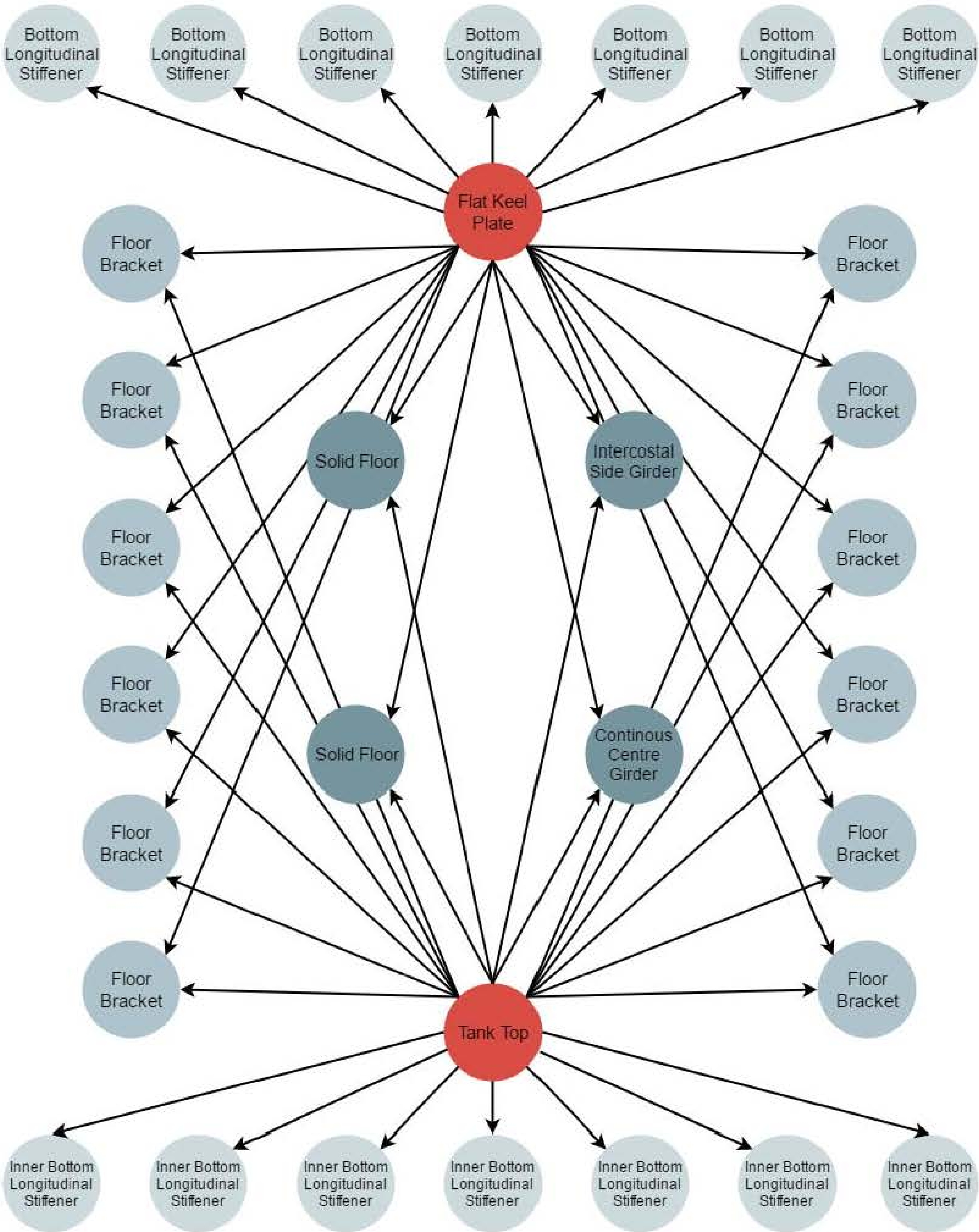


Figure 6.5: Longitudinal stiffened double bottom Graph example

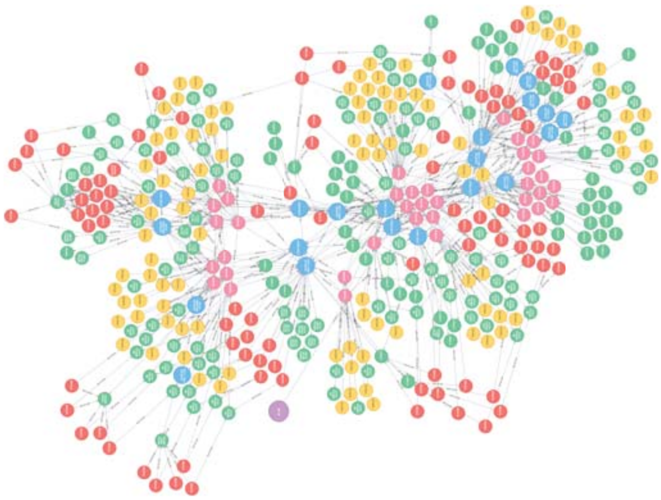


Figure 6.6: Example of a section graph from the graph database

6.3.1. Graph Database as a Case Based Reasoning Method in the Ship Building Industry

The new cost estimation combines the CBR theory and with a graph database to store the data of ship sections. Figure 6.7 shows the cost estimation process of the graph database cost estimation method. The numbers are labelled to the activities and are explained below. There are two points in the process where two options are possible. These numbers are marked with a star.

Table 6.3: Explenation number in figure 6.7

1	Old ship section build at IHC
2	Documentation of the old ship section (Final documentation, preferably includes final costs, final working hours but also budgeted costs and technical data)
3	Create a graph model of the old ship section and label the nodes and relationships properly
4	Create a graph database with numerous data of already built ship sections
5	Order for a new ship comes in, create ship sections from the general arrangement drawing
5*	Analyse the graph database, for instance on difference between hours estimated and hours spent on a ship section
6	Create a Graph Model from the new ship section and create the cost function. The cost function is the code to find the most similar in the database.
7	Run the Graph Model through the graph database and look for similar ship sections
8	Estimate the costs of the new ship section based on the similarities found in the database
9	Build the new ship section
9*	If changes occur in the ship sections anywhere in the process update the ship section, update the Graph Model and estimate costs again
10	Finalize the Section Graph of the new ship section with the realized data, final costs, final working hours etcetera.

The research will be based on the BVC phase, where more detailed information is available than during the VC phase. The historical section graphs contain the most recent section data and the realized production hours. However the section information available when the BVC estimation will be performed might be subjected through some changes before the section is built. A problem that might occur is that the data are not sufficient to create an accurate cost estimation on. Important will be that information about the main cost drivers for ship sections is known and can be added to the graph database.

For further research the problem the validity of the graph database method during the VC phase should be checked. A problem might be that there is not enough information available of the new to build ship sections. During the VC phase the information available is less then during the BVC phase. The VC phase covers the basic information of the vessel that is going to be built.

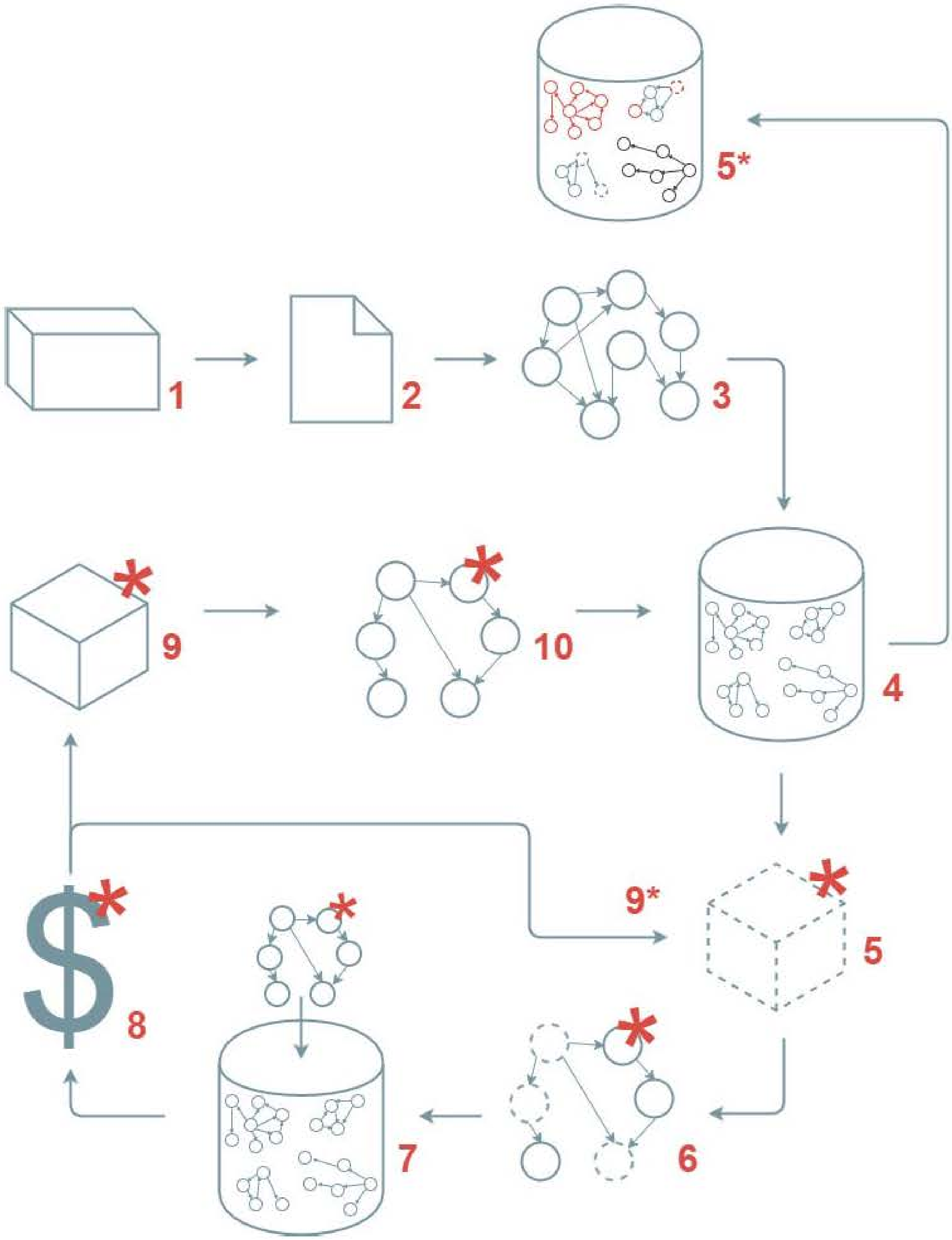


Figure 6.7: Cost Estimation Process based on the Case Based Reasoning Method combined with a Ship Section Graph Database

The process figure number 5* shows another option of the database. With this option it is possible to analyse the ship section database or the hour evaluation directly after the section has been built. This analysis can be used for cost estimation purposes as an hour evaluation. But it can also be for design analysis, for instance if a new welding technique or building material is implemented to see the outcomes of this change or to find similar cases.

6.4. Section Building labels for Graph Database

In this section the labels that are necessary to create a good cost estimation with the graph database and to find relevant similarities between ship sections are discussed. The amount of labels included in a graph is different from the amount of properties that have been used in the statistical method. In a graph more data can be included, however in this research only the cost drivers where a correlation against the hours is made of are used in the test cost estimation.

The labels in a graph database can for instance be the plate dimensions, profile type or the building order of a section. The relationship lines contain the labels that represent the amount of work performed during the production. A relationship label can for instance be the different welding techniques and if it is manually or machine welded. Each ship section contains one node that represents the total ship section, containing labels belonging to the total section. They could be the budgeted and realized costs of the old ship sections or the hours spent by personnel on the section. Figure 6.8 shows an example of a section graph with nodes and relationships with their labels.

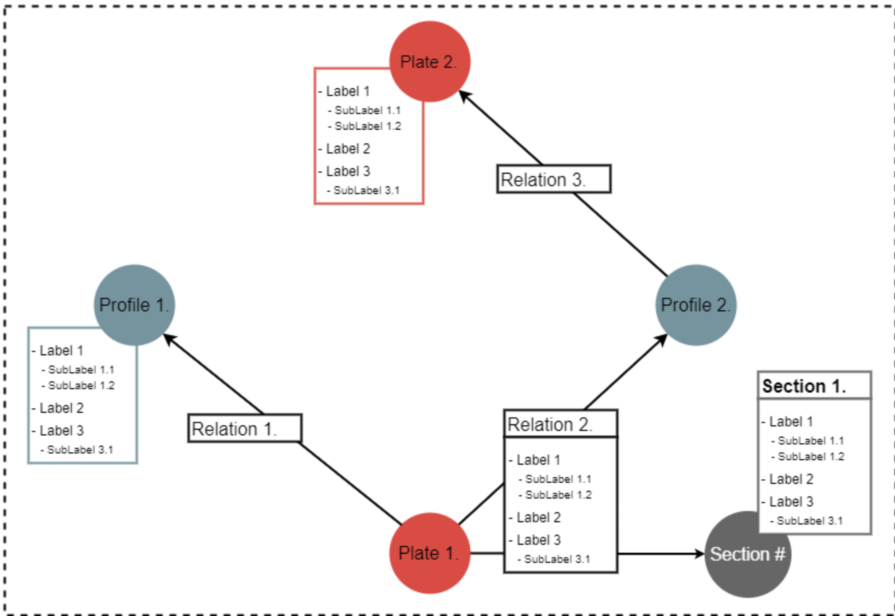


Figure 6.8: An Example of a Section Graph with labels

6.4.1. Information available at Royal IHC

At Royal IHC they have different types of data documentation of the ship sections. These types of documentation are listed below, more information about these data documentation types can be found in appendix L.

1. **Analysis lists**
2. **2D section drawings**
3. **3D model in Nupas**
4. **Royal IHC construction code**

5. **Welding list**
6. **Order administration**

A 3D model of the ship sections is made during the design phase, the program they use these drawings is Nupas. They draw the section drawings with corresponding analysis lists. The 2D drawings of the ship sections contain the geometric information of the section. The data that is available at Royal IHC is used as a basis to create the graph database input lists. Figure 6.9 shows which data is used for which graph database input list.

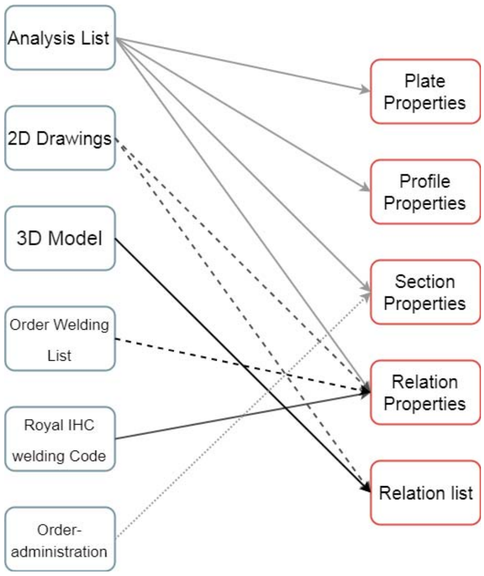


Figure 6.9: The connection between data available at Royal IHC and implementation documents of the Graph database.

The graph database will contain information about different elements of a ship section. The information from plates and profiles is fully extracted from the analysis lists. The section and welding (relation) information is only partly extracted from the analysis lists. Section hours are gathered from the order administration. More information on the welding labels are gathered from the 2D drawings, the Royal IHC welding code and the welding lists and information on the relationship list is extracted from both the 2D drawings and the 3D model.

6.4.2. Node-Labels

The nodes in the graph can be multiple objects, plates, stiffeners, girders and frames. All these objects can have different labels that are important to place in the database, the so called node-labels. The program IHC works with creates two types of data lists one for plates and one for profiles. Under plates all the elements made from only one steel plate are categorized. Plates can be: 1) (Shell-) Plates, 2) Frames, 3) Brackets & 4) Collar plates. For these objects the node labels are listed and explained.

Plates:

Plates are divided into four types, plates used for decks and shells, frame plates, brackets and collar plates. In the analysis lists they are all labelled as plates, PL, and have the same labels. This is why in the database the different plate types will have the same labels, but for visualization and for the analysis of the database the plates are divided. The labels for plates that will be used are:

- **Plate ID:** This is the identification a plate has from IHC and is imported into Neo4j. a typical plate identification number is: 01278142101051. The first five numbers are the order number, 01278. The next six numbers are the section number, 142101 and the last two numbers are the plate number, 51. An additional set of digits could be added if there are sub assemblies in a section.

- **Description:** This is the description of the part. It contains the part type, for instance a bracket or a collar plate, and the location, for instance at base or a distance from the centre line. A description of a shell plate could be 'Long,sect, -7350 from CL' and a description of a bracket Bracket -8550 from CL. For collar plates the plate description also states the type of collar plate.
- **Type:** For plate the type is always PL, plate
- **Width:** The maximum width of the plate.
- **Length:** The maximum length of the plate.
- **Thickness:** The Maximum thickness of the plate.
- **Net Surface:** the total surface of the plate. This is the surface minus the holes and other dimension differences in the plate.
- **Material:** The material of the plate.
- **Weight:** The total weight of the plate. This is the weight minus the holes and other dimension differences in the plate, so the net weight.
- **Shape type:** These are linked to the work-chain numbers, listed in appendix M. These numbers are used to indicate whether a plate is curved or straight. The curvature can be exposed at the end of the plate, such as a flange, or the plate itself is curved, such as the bilge plates.
- **Destination:** The destination given in the analysis lists of Royal IHC show the panel number the plate belongs to. A panel with corresponding stiffeners have the same 'Des' code, an example of a 'Des' code is PL100-1. For the last items, brackets and collar plates that have been assembled when all the panels were assembled, there is a different 'Des' code, VM.
- **Construction order:** This number corresponds to the construction order for the section. Some panels can be built simultaneously, but when assembling these panels they are assembled in the order given in the construction order for that section.

All these elements show the information that is given in the analysis lists of Royal IHC. It shows the basic dimensions of the plates and where in the ship the plate is located. It also shows what type of plate is taken into account and when the plate is used in the construction process of the section. Detailed information about the geometry and the location of the plate in the vessel can be found in the section 2D drawings and in Nupas, the 3D model of the section.

Profiles:

Profiles have nearly the same labels as plates. There are two labels that differ from plate nodes:[3]

- **Profile ID:** The ID for profiles is categorized in the same way as for plates. Profiles often start with number 500 and up.
- **Type:** Profiles can occur in different types in the analysis list, HP, ST, VK and HK.

6.4.3. Relationship-Labels

Relationships in the section graphs are the welding lines between the parts. These welding lines have different labels as well. The weld labels will contain the information about the amount of work that is performed on a section.

- **Item type:** Both the beginning ID and the end ID are labeled as to which type of item they are. in type of nodes there are profiles and plates, but the plates can be categorized in normal plates, brackets, frames and collar plates.
- **Weld line length:** The total length of the weld line that is connecting the two surfaces together. This is the total weld line length. Some weld line types require only one weld seam where others need double weld seams[4], for instance fillet welds. This factor is already calculated in the total weld line.

[3] if the label is the same as for plates, see the plate label explanation
[4] The amount of times a welder or machine has to follow the weld line to fill the entire weld seam is not taken into account.

- **Position and Thickness difference of the seam[5]:** This feature shows the way the weld seam is positioned, face-to-edge or edge-to-edge, and if there is a thickness difference between the welding surfaces. The characters show the type of seam, if there is a thickness difference and the position of the surfaces. These correspondent with the Royal IHC welding codes and are shown in the section drawings. [29] Appendix N shows the first page of the Royal IHC welding code with the explanation of the characters.
  - **N:** Plates in line with equal thickness
  - **B:** Plates in line with different thickness on the top
  - **O:** Plates in line with different thickness at the bottom
  - **E:** Plates in line with different thickness divided over weld line
  - **H:** T-joint
- **Taper[1]:** These characters refer to the welding surfaces and if they are tapered or not. These are also given in the welding code. [29]
  - **N:** Thickness difference not tapered
  - **A:** Thickness difference tapered
  - **B:** Top is tapered
  - **O:** Bottom is tapered
  - **C:** Center: both sided tapered
- **welding method[1]:** This shows if the welds are done by manual welding or automatic welding. [29]
  - **00 - 16:** Are not used
  - **10 - 29:** Manual and mechanical welding
  - **30 - 49:** Automatic welding
  - **40 - 69:** Backup
  - **50 - 99:** All seams
- **Welding seam shape[1]:** the shape of the welding seam shows if the seam of the weld line is tapered or not and in which way.[6] [29]
  - **1:** I-seam
  - **2:** V-seam opening at top
  - **3:** V-seam opening at bottom
  - **4:** Symmetric K- or X-seam
  - **5:** Asymmetric K- or X-seam with the biggest opening on top
  - **6:** Asymmetric K- or X-seam with the biggest opening at the bottom
  - **7-10:** special seams
- **Weld line co-ordination:** This shows the coordination of the weld line. See figure 6.10 for the names of the weld co-ordinations.

  This is the position the welder has to weld and can be categorized as upper-hand, under-hand and vertical welding. Here PA, PB and PC are under-hand welding positions, PF and PG vertical welding positions and PE and PD are upper-hand. An assumption for the welder position is that upper-hand [illegible] welding is preferred as the welder position. The welding co-ordination is derived from the building order for the belonging section and the welding code.

[5]Note that these characters are only used for plate to plate connections where the drawings give a separate notation at the weld seam, for the other plate to plate connections and for profile to plate connections there is another way of characterizing.

[6] The metalix method of applying the taper is left out of the label value.

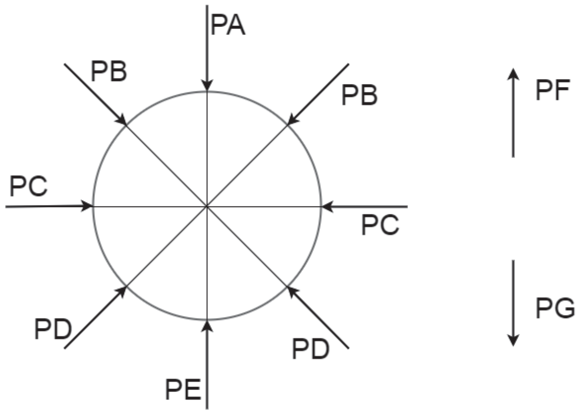


Figure 6.10: The different welding co-ordinations

- **Upper-hand:** Horizontal welding position, with the welding torch upper-hand. As mentioned before this method is avoided as much as possible. This is done by turning the section in a way that as far as possible weld lines can be made under-hand and some of them vertically, but avoiding the upper-hand welding position. That is why this method is left out of consideration in the welding list.
- **Under-hand - Horizontal** Horizontal welding position, with the welding torch above head. The average welding speed is: 4 meter per hour.
- **Under-hand - Vertical**: Vertical welding position, with the welding torch under-hand. The average welding speed is: 2 meter per hour.

- **Edge penetration type:** The way the two surfaces are penetrated by the weld line. This is given in the welding list and differs per order number. [30] Figure 6.11 shows the different weld penetration types. Note that DP welding is the same method as FP welding but the weld line is fuller.
  - **FP:** Full penetration welding.
  - **DP:** Deep penetration welding.
  - **PP:** Partial penetration welding.
  - **FW:** (Double) fillet weld.
  - **CH:** Chain weld.

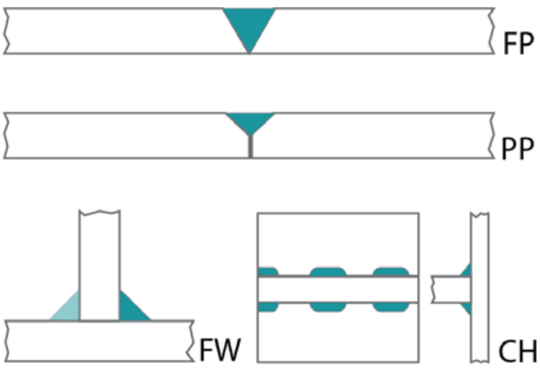


Figure 6.11: The different weld penetrations

- **Weld thickness:** the thickness the weld seam has to cover.
- **Throat thickness:**[7] The thickness of the welding throat. This is given in the corresponding welding list of the section order. [30]

[7] Note that these characters are only used for brackets, profiles and collar plates.

- **Technical Procedure:** These are welding methods discussed in appendix C. As discussed in that chapter they can be SAW, GMAW and MMAW. Looking at these methods SAW is used when a weld can be performed by a SAW machine or a tractor. MMAW and GMAW can be used for roughly the same welds. In the welding code of Royal IHC belonging to the ship orders the GMAW method is taken into account and the MMAW method is left out of consideration. [redacted] the MMAW method is still used at Royal IHC, but mostly outside and by the older welders. New welders use and acquire the GMAW welding method instead of MMAW. For these reasons MMAW is left out of the welding procedures, only GMAW and SAW welding are used as technical welding procedures.
- **Watertight:** If the seams between welding surfaces have to be watertight or not. 1 = watertight and 0 = not watertight.
- **Edge angle:** The angle of two surfaces to each other.
- **Welding factor[2]:** The welding factor is a safety margin for the weld line. For some types of components and in some parts of the ship a higher strength is necessary to obtain a safe weld line. By multiplying the welding factor with the thickness given in the welding list the necessary safe welding thickness is obtained. The welding factor also depends on whether the weld needs to be watertight or not. For welds in watertight tanks the welding factor is higher.

6.4.4. Ship Section Labels

Section labels represent the data that is only available for the total ship section. For the sections estimated hours are available, but also realized production hours. This can be used in an analysis of the graph database. The sections are visualized as nodes in the graph database and connected to the first plate in the analysis list. They are only connected to one plate of the section to shorten the run time of the Neo4j database.

- **Section identification:** The section ID is the order number combined with the section number.
- **Year:** This is the year the section is built at Royal IHC.
- **Type:** the type of the section. For this research the focus lies on buoyancy compartments, but there are more ship section types, such as engine or pump room sections or a hopper coaming.
- **Curvature:** The curvature number shows if the section contains no curvature (0), one degree of curvature for example the bilge plate (1), or if the section contains double shaped plates which can occur in sections in the aft or front sections of the vessel (2).
- **Yard:** as seen in chapter 2 there are differences between the yard locations of Royal IHC. Sections built at Krimpen aan den IJssel can be bigger and heavier than sections built at Kinderdijk due to a bigger crane capacity and a bigger section floor. Also when sections will be outsourced but data is placed in the database it is important to take this into account and place this information into the database.
- **Parts:** The total amount of parts the ship section contains. Parts being plates, frames, brackets, collar plates and profiles. This is not the same value as the amount of weld lines, because parts can be connected to more than one other parts.
- **Weight:** The weight of the total section. As for plates and profiles, this is the net weight of the section.
- **Hours:** The hours of the section are divided in the order administration in VC, BVC and realized hours, but also in welding hours, iron work hours, grinding hours, panel street hours and robot hall hours. The hours for grinding are left out of the analysis in this research. For some orders the grinding hours are made by Royal IHC, for other orders the grinding work is outsourced to a different company. For this reason the grinding hours are left out of the research.

  For the ship sections that are placed in the graph database the hours for welding, iron work, panel street and robot hall, are calculated for both the VC and the BVC and the realized costs are known.

  - **Welding hours:** The hours spent on the welding of the section. These hours also include the set up time of the welder and more welding labels.
  - **Ironwork hours:** These hours consist of the activities that are performed before the welding. These are tasks that belong to the setting and the attachment of steel parts for the section building. This can for instance be the pre-assembling of the panels with tasks such as tack welding.

  - **Panel street hours:** These are the welding and ironwork hours performed on the panels that are assembled in the panel street.
  - **Robot hall hours:** Hours spent by the robot in the robot hall on the assembling of section sub-panels.
- **Total Hours:** The sum of all the hours spent on the section. This can also be calculated for the VC and the BVC estimated hours.
- **Total welding length:** The total welding length that is needed to assemble the section.

6.4.5. Limitations to the Labels Used

The labels that are used show detailed information about the section building and the different production techniques. It does not show external effects for the section building process. Broken machinery or injuries delay the production process and every employee works at a different pace that is probably not constant over time. There is also a randomness factor involved for aspects such as weather influences. The total contribution of these effects are hard to detect with both the current method and the proposed solutions, the current labels do not contain enough information to make an analysis for this contribution. So there will always be some noise when estimating the hours.

Late changes made at the end of the BVC phase are not included in the graph database but are sometimes included in the hour administration of the sections. They are visible in the analysis list but not on the drawings and the 3D drawings. Including them into the relationship lists was not possible for most of these changes, because the location and relations of these parts could not be found. Excluding these changes results in some noise between the label values taken into account and the total production hours of the section.

It should be noted that there is no accessibility factor included in the labels or that the total volume of the weld line is not taken into account for the section labels. An accessibility factor can cope for weld lines that are more time consuming. This is not taken into account because multiple labels are already included that indicate what type of weld is performed. A more detailed cost driver analysis should find out if these labels show enough detail to estimate the hours of the different type of weld lines or if an accessibility factor needs to be added. The total volume of the weld is a label that is preferably included in the welding labels, because it states how many weld material is used and how many layers of weld have to be laid to fill the weld. The calculation of the volume is however a time consuming task. If in the future a less time consuming method is available, it is preferred to make a new label containing the volume of the weld length.

6.5. Possibilities of the Method of G. Alblas

This graph database can be seen as the cost estimation part of the complete method of G. Alblas. His complete method will be able to estimate or analyse elements of the complete design and production process of the vessel. The method is intended as a method that can be used to make cost estimations, design suggestions and section assembly suggestions in both the VC as the BVC phase. Another intended part of the method is that it is both applicable for one of a kind vessels of all different vessel types, but also for vessel types or replicas that have been built before. Figure 6.12 shows the total method of G. Alblas. Table 6.4 explains the numbers shown in the figure. The thick lines in the figure show the method loop, the dotted lines show the database suggestion lines. Some of the possibilities of the total method are based on the section graph database but there are parts that have their own graph structure. For instance the outfitting arrangement can have its own data structure. Every production component could be connected to the vessel in one big database or there could be separate databases that each store their own part of the production cycle.

After the section plan has been designed from the general arrangement of the vessel the detailed design for the vessel and ship sections starts. From that point forward in the design and production phase of the ship the method can be used for suggesting optimizations toward the design and can be used to estimate costs. The possibly, unfinished ship section is turned into a graph and will be run through the database searching for similar sections. There is an extra option that is not directly derived from the method of G. Alblas. It is an extra option using the same input data as used to create a section graph database, the use a simulation program. A program such as the Technomatix Plant Simulation can be used to simulate parts of the section building process. The next step of the method is to search for similar sections in the database. During this similarity search a similar section to the new section can be found and used for design or cost estimation solutions. When the design and engineering of the sections are at their final stage, the production of the sections and the vessel starts. During this process data from the new vessel will be placed in the database. When this is done the data can be used for new projects.

Table 6.4: Explanation of figure 6.12: The complete method of G. Alblas

1	The general arrangement of the new vessel.
2	The section plan of the vessel is drawn from the general arrangement of the vessel.
3	The ship sections are designed and modelled in CAD programs.
4	A model of the possibly, unfinished ship section is generated.
5	The ship section is turned into a section graph.
6	The similarity assessment of the new ship section graph and the old ship section graphs in the database is performed.
6*	There is a possibility to use the information of the section graph in a simulation program.
7	The new section is ready to be built.
8	The building of the section
9	The pre-outfitting part of the section is performed, including activities such as piping.
10	The outfitting is executed.
11	The ship section graph database, updated with the new vessels production information.

6.5.1. Detail Section Cost Estimation

How the method can be used to estimate the detailed costs for ship section is discussed in chapter 6. This method can be used to determine the costs of detailed sections during the BVC phase of the cost estimation. In chapter 7 it will be discussed how this method could be more beneficial than the current cost estimation method of IHC.

6.5.2. Cost Analysis of the Production Process

The method includes many production labels in the database. Analysing these labels using the different values gives more insight in the cost drivers during the production process. Analysis using the database can for instance be done by using the output lists of the database and performing the cost driver analysis as performed in this research. Or when the final hours are known and variate from the estimated hours, the data taken into account can be analysed on variations among the final data. In this way more insight in the cost drivers of the production process can be created.

6.5.3. Overall Detail Cost price

The method can be used to determine the overall ship costs for the to be carried out activities or work. The method is not intended to estimate the steel price. If not only costs of the hull production but also of the overall ship and the other production parts will be included in the database, this can be used to search for similar solutions and their costs. If the method will lead to a more accurate overall cost price must be reviewed by further research.

6.5.4. VC Phase Cost Estimation

If the database will give an accurate cost estimation in the VC phase is not mentioned in this report and must be concluded from new research. But if the complete ship data will be implemented in the database, searching for a similar vessel by filtering out the data that is already known in a early design phase should be possible. The latter will however be hard to do when a one of a kind, new vessel type is being built. New research should therefore not only see if accurate cost estimation can be made in this phase but also how the method performs with new vessel types.

6.5.5. Information Enrichment

By looking at previous ships, analyses can be made for the cost drivers of the ship production process. If new optimizations are adapted in new vessels the outcome of these solutions can be measured with the database.

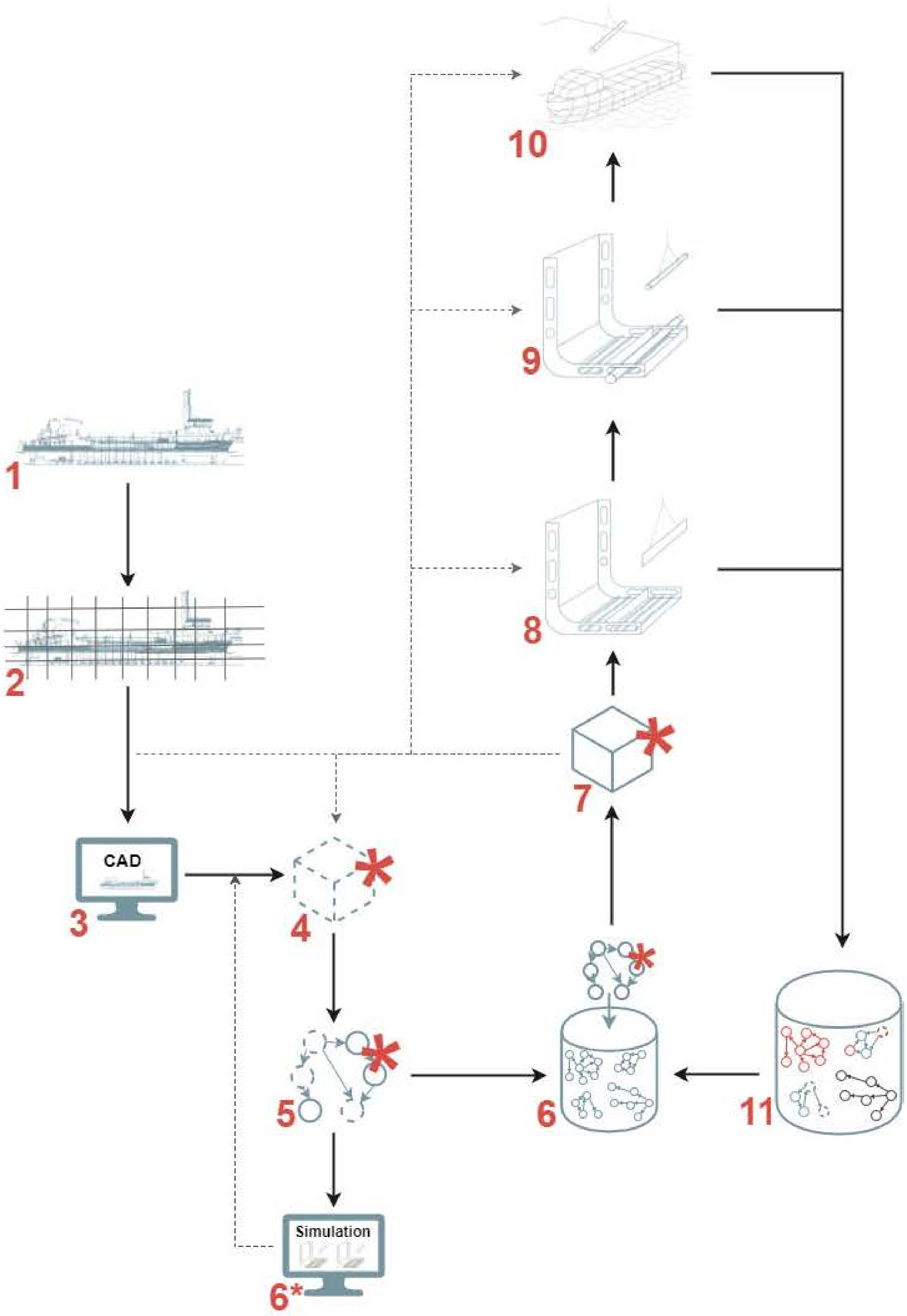


Figure 6.12: The complete method of G. Alblas

Also new solutions can be designed by looking at problems in the production process with the help of the database and by knowing what the cost drivers are. The complete method can thus result in information enrichment to create insights in more production efficient ship designs.

6.5.6. Technique Consequences

The method can be of value to see what effect different production techniques have on the complete production phase of the vessel. For instance if a different type of welding technique is used, the effects of this method can be monitored and estimated. This is not only possible on cost level but also on the production and design level of vessels. A new building material or welding material can for instance be monitored with the database.

6.5.7. Design Proposals

The complete method can also result in design suggestions. For instance if a combination of parts can be reduced to one part as discussed in the research of P. van der Horst [34]. The pattern of these parts can be filtered in the database and the suggestion can be made to alter the section by replacing the different parts by one plate. This can be a standard filter in the database searching for the different part patterns or a filter that can be used per section. Chapter 7 explains an example from the report of P. van der Horst. New design proposals can be monitored with the database in the same way as the new techniques can. Further research should find out if it is possible to create design proposals with the graph database.

6.5.8. Simulation Input

This option is not part of the intended method of G. Alblas but serves as combination of the method with another program that uses the same data input. A simulation program can be used to optimize the production process of the yard. For instance with a Tecnomatix plant simulation different variables can be optimized or monitored. The input data for section building can be extracted from the input data for the graph database. If the simulation program is used to optimize the production of sections the input is the part information from the analysis lists and the relation list combining these parts. The relation list is however used without the detailed welding process information but the welding position and welding process can be taken into account. With this information and the yard production labels, for instance the crane details and employee information, the program can perform a simulation about the speed of the section production. Optimization in terms of yard labels can be tested with a simulation. It is also possible to perform simulations for other parts of the production process, such as piping or the block assembly. Where optimization with the database first has to be tested in real life, the simulation can give some insight in a optimization without testing it in real life. A note has to be made that simulations do not always situate real life, employees differ in work speed and weather influences may not be encountered the same way.

6.5.9. Cost Estimations using Section Plan Variations

The method is not able to give suggestions for the design of a section plan but it can give insights in the cost optimization by alternating the section plan. By testing variations of section plans the most accurate cost price can be found.

6.5.10. Piping & Outfitting Suggestion

The method can also give suggestions on piping and outfitting level. When data from these parts of the production process will be included in a database the method can provide insights in these systems. As the database looks for sections with similar costs, the database can search for vessels with a similar piping arrangement. The costs can be based on this similar vessel but also design suggestions can be made for the piping arrangement. In the same way also other systems connected to shipbuilding can be gathered in databases. Other systems might be the outfitting of the vessel or the mooring arrangement. Further research is necessary to see how the information from these systems will be gathered and structured in a database, how the search algorithm and similarity assessment for these systems will be conducted and whether or not the method can provide accurate cost estimations or design suggestions.

6.6. Summary

Subquestion 1a - **"How can a cost estimation be made with the use of a sections graph database?"**

As discussed in chapter 5, the CBR theory tries to find a similar section to reuse this solution for the new section. The graph database functions as the case base for the CBR theory. The graphs are the data stores of the ship sections and similar sections can be found inside this graph database.

Subquestion 1b - **"Which information or data from ship sections are important to create graph models and combine them in an graph database?"**

The different node and relationship labels are discussed in this chapter. Plate properties are for instance the dimensions of the plate and which building order number it gets. Profile properties are almost the same as the plate properties only the profiles have an extra label called the profile type. This label indicates the type of profile used. Examples of section labels are the total weld length and the hours performed on the section. Relationship labels can be the welding technique used or the position of the welder. The effect every label has on the total production hour has to be investigated in further research.

Subquestion 1c - **"How can a cost estimation be made with the use of these sections graphs?"**

This is answered by the code given in appendix V. This code created graphs in Neo4j with the node and relationship lists. These node lists contain the label data and the ID tag of the node. The relationship list connects the ID tags of the nodes and contains the label data for the relationships.

Subquestion 4 - **"How can the difference in budgeted costs and final ship production costs be explained?"**

This can be due to multiple factors. The graph database method can provide more insight in what led to these differences because more input data from the ship sections is used during the cost estimation. There are still some factors that are not taken into account during the cost estimation. The production task being a human job and the failure of equipment, but also a randomness factor from the weather or other yard conditions. These factors generate noise between the estimated hours and the final production hours.

7

Qualitative Comparison between the Three Cost Estimation Methods

This chapter discuss how the three different cost estimation methods perform on different example structures of a ship section. They are evaluated in a qualitative way. Each example in section 7.1 shows two different ship section parts and how they are represented by a graph. Reasoning is done by looking at differences between the two examples and how the costs would be estimated by the three methods. From this analysis conclusions are drawn in the final section.

7.1. Qualitative Analysis

The results from the cost driver analysis together with the method explanations and input data are used to analyse the section structures in a qualitative way. The main conclusion from the cost driver analysis that is used is that weight is not the main cost driver for the section production hours is used. Also the goals for a new cost estimation method are kept in mind during this analysis. This section gives a few examples to illustrate the way the new cost estimation methods and the current method behave for different ship structures. This part can only be illustrated in a qualitative way because the similarity search algorithm of the graph database method has not yet been finished so only a less detailed cost estimation in a quantitative way can be performed.

7.1.1. Example 1: Simple plate with profiles

Example 1 overlooks two simple plates with three or five profiles. Figure 7.1 shows a plate with three profiles. The graph representation is given on the right side of the figure. Figure 7.2 shows the second set-up, a plate with five profiles. For this example the size of the plates and the profiles indicated stay the same for both the set-ups.

Table 7.1 shows the differences of the main properties from the two set-ups. The weight has increased because of the two profiles added. The number of parts has increased from four parts to six parts between the two set-ups. Also the amount of profiles is increased from three to five. The weld length will also increase because the two extra profiles have to be attached to the plate.

- **Royal IHC current methods:** With the current method the differences in weight will lead to a different cost estimation for the two set-ups. Probably an increase of costs for set-up 2. A main problem that occurs when only the weight is used, is that the increase of the amount of parts or profiles and the weld length is relatively higher than the increase of weight. The plate is a heavy part of the construction, so the profiles have less influence on the weight. When using the conclusion from the cost driver analysis, looking at the other properties may lead to a more accurate cost estimation.
- **Statistic Method:** The statistic method bases the cost estimation on different properties. So if the

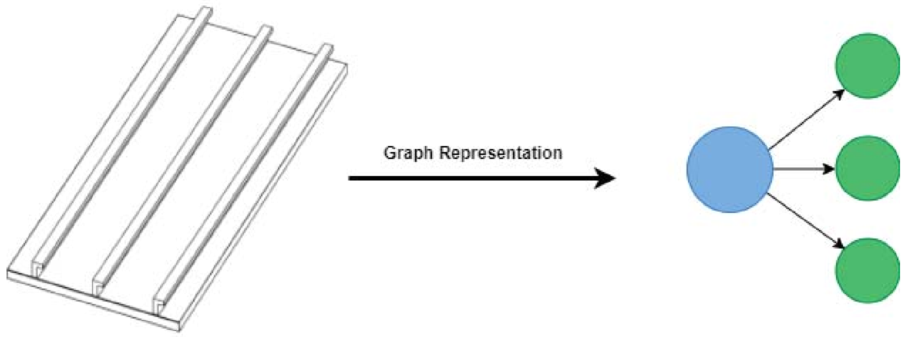


Figure 7.1: Example 1: Set-up 1 Graph Representation

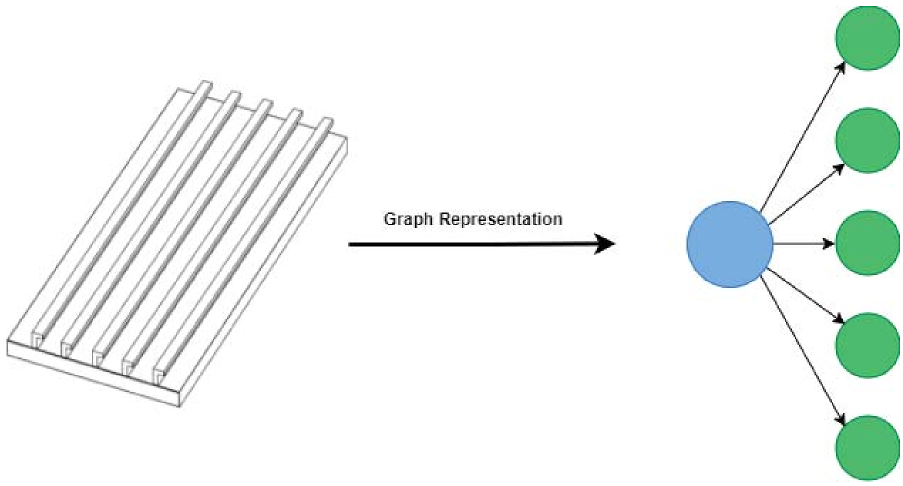


Figure 7.2: Example 1: Set-up 2 Graph Representation

Table 7.1: Example 1. Property Differences

	Setup 1	Setup 2
Weight	Lower	Higher
Parts	4	6
Plates	1	1
Weld Length	Lower	Higher
Profiles	3	5
Frames	-	-
Brackets	-	-
Collar plates	-	-

amount of profiles or parts changes will lead to a different cost estimation. The conclusion from chapter 4 that parts and weld length have a higher correlation to the hours than weight and that the equations give a more accurate cost estimation than the BVC estimation, assumes that this methods will lead to a more accurate cost estimation for the set-ups in this example. Although this method may lead to a more accurate cost estimation than the current IHC method, the limitation of the statistic method is that it is not taking the amount of extra work that is put into the section building into account. Another problem is that not every cost driver is used in the equations, so only if profiles, parts or weld length are variables in the equations the statistical method will give different outcomes.

- **Graph Database:** In the graph database all the information from cost drivers and the work performed on the section building can be stored. A filter can be made on the amount of parts, the weight of the structure and the weld length but also on the welding method, the type of weld that has to be made and the amount of ironwork that is performed in the structure. With this information of a ship section the solution of the most similar section can be adapted and reused. This will have an advantage above the other two methods of being more accurate in terms of cost driver information used. In this example the amount of work that has to be performed for the second set-up is higher than the first set-up. This will be shown in the graph database, but in the other two cost estimation methods the amount of work is neglected.

7.1.2. Example 2: Plates with same amount of profiles

Example two shows two set-ups with five profiles only the first set-up contains one plate and the second set-up two. Figure 7.3 shows the first set-up a plate with five profiles. Figure 7.4 shows the second set-up where two plates create the same surface as the first set-up. Furthermore the size of the profiles and the plate thickness is the same in both set-ups.

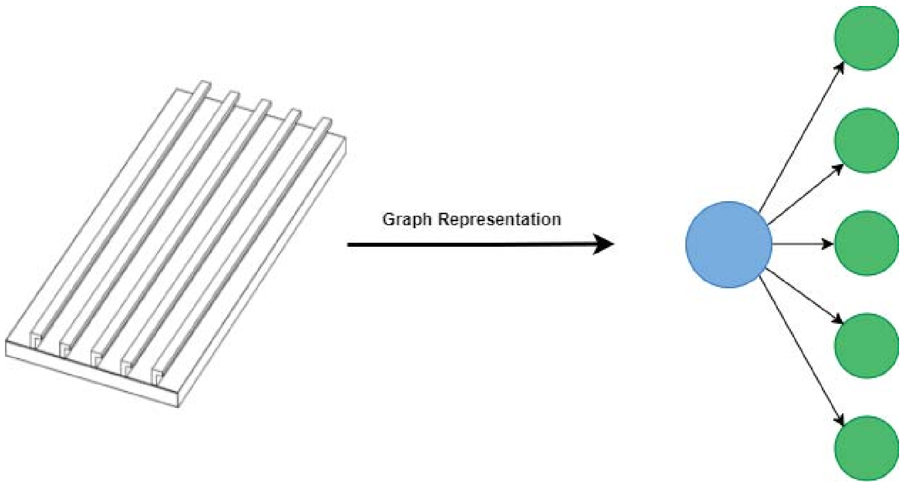


Figure 7.3: Example 2: Setup 1 Graph Representation

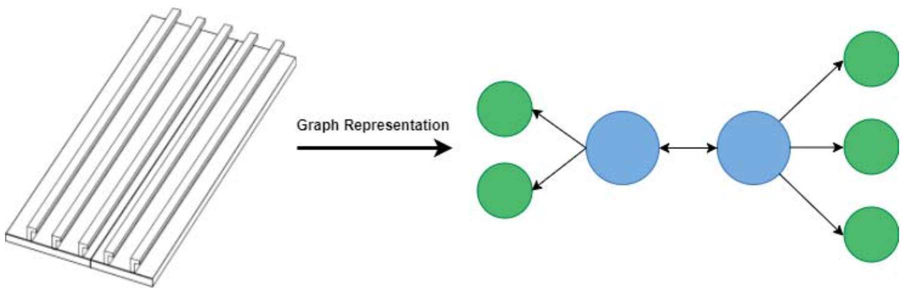


Figure 7.4: Example 2: Setup 2 Graph Representation

Table 7.2 shows the differences between the set-ups. The set-ups have more or less the same weight. The amount of parts is increased with one plate and the weld length will be higher. The main difference between these two set-ups is the amount of work that is delivered for the production of the structures. The second set-up will have more production hours because of the extra weld line.

- **Royal IHC current methods:** the current method will not show a different cost estimation because the weight of the structure stays the same. This shows the limit of the current method used at Royal IHC and the limit of estimating costs based on hours per ton.
- **Statistic Method:** The statistic method will show a different cost price if the cost drivers that change are variables in the equations for the costs. The weld length and the amount of parts are different for the

Table 7.2: Example 2. Property Differences

	Setup 1	Setup 2
Weight	± Same	± Same
Parts	6	7
Plates	1	2
Weld Length	Lower	Higher
Profiles	5	5
Frames	-	-
Brackets	-	-
Collar plates	-	-

set-ups so this will lead to different variable values in the equations. This shows the advantage of using more cost drivers than just weight. Still this method cannot show the impact of a different welding method or the amount of ironwork that is used to build the structure.

- **Graph Database:** Again the graph database can store the same data as used in the other methods but it can also use the welding type and length. Thus the amount of work that is spent on the structure. This will result in a more accurate cost estimation then the other two methods.

7.1.3. Example 3: Two plates combined, 90°or round connection

Example 3 looks at two set-ups that show a 90°corner in a structure. The first set-up, shown in figure 7.5, contains a straight corner with two plates and six profiles. Set-up two, shown in figure 7.6 contains a rounded corner and three plates with again six profiles. The size and plate thickness of the structure stays the same for both set-ups.

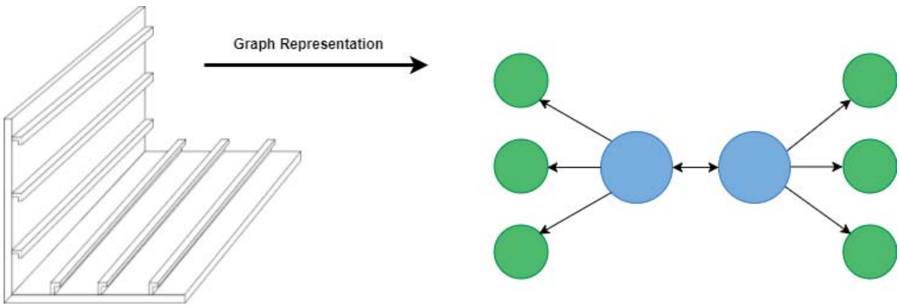


Figure 7.5: Example 3: Setup 1 Graph Representation

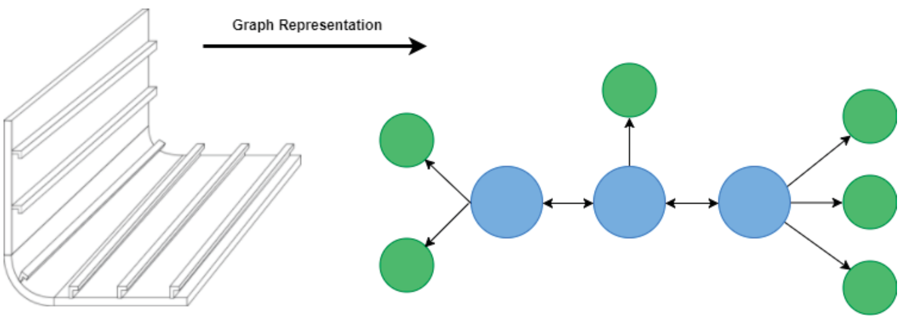


Figure 7.6: Example 3: Setup 2 Graph Representation

Table 7.3 shows the differences in terms of property values. The amount of parts is different because the second set-up contains three plates instead of two in the first set-up and in set-up 2 one of the plates is bended . The weld length is increased, but the rest of the structure is the same.

Table 7.3: Example 3. Property Differences

	Setup 1	Setup 2
Weight	± Same	± Same
Parts	8	9
Plates	2	3
Weld Length	Lower	Higher
Profiles	6	6
Frames	-	-
Brackets	-	-
Collar plates	-	-

- **Royal IHC current methods:** Again this method will not lead to different cost estimations because the weight of the section will not change that much.

- **Statistic Method:** The statistic method will give a change in the cost estimation outcomes. Parts and weld length are used variables in some of the equations and thus when these equations are used the values will change resulting in a different cost estimation. For this example however the amount of extra work may have a higher influence because of the bended plate in the second set-up and the two extra weld lines. To assemble and work with a rounded plate is more time consuming than working with a flat plate. This is not taken into account in the statistic method.

- **Graph Database:** In the database geometric values of the plates are taken into account. Again the graph database contains more detail of the structure but also it contains information about the amount of work spent on a section. Where the current IHC method can only use a section factor, for this example a rounded outer shell. The graph database can go into more detail by looking at every part on its own instead of the complete section.

7.1.4. Example 4: Two plates combined, 90°connection

Example 4 shows two plates with a straight 90°corner. Set-up one, shown in figure 7.7, are two plates combined in the corner, with one plate containing five profiles and brackets partly connecting the both plates. Set-up two, shown in figure 7.8, shows the same size plates but with the profiles and brackets distributed over the two plates. Two profiles on the top plate and three on the lower plate.

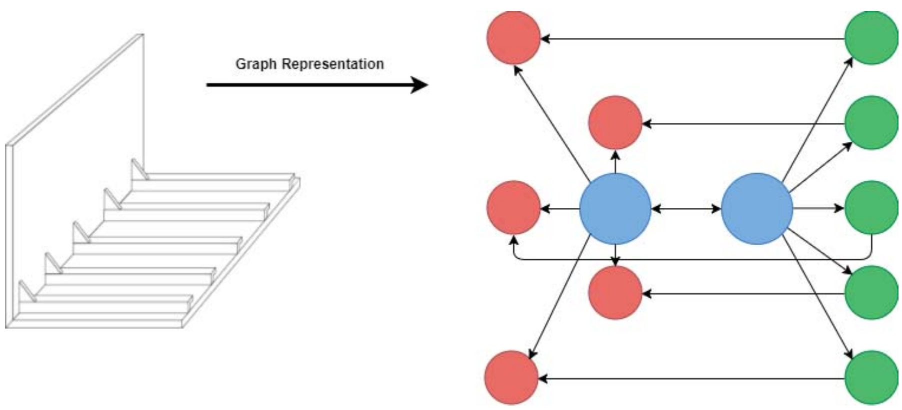


Figure 7.7: Example 4: Setup 1 Graph Representation

Table 7.4 shows the differences between the two set-ups. Again as in the previous two examples there are no weight differences and the plate and profile sizes do not differ between the two set-ups. In this example all the property values are the same, the amount of parts is the same and the weld length is the same. The only thing that shows the difference in the geometrics of these set-ups are the graphs shown in the set-up figures. In this example it shows what value they have over just looking at the cost driver values.

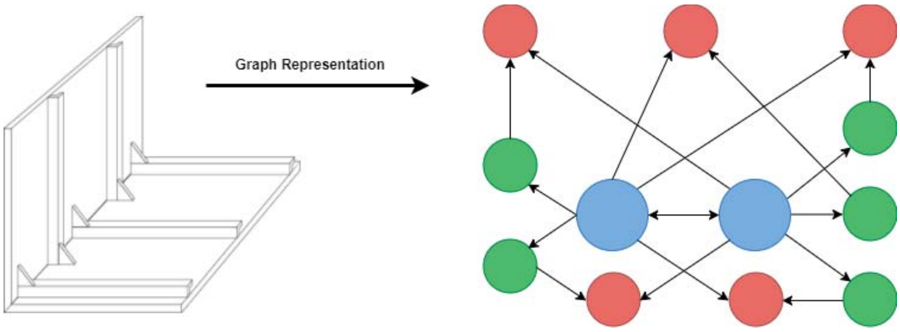


Figure 7.8: Example 4: Setup 2 Graph Representation

Table 7.4: Example 4. Property Differences

	Setup 1	Setup 2
Weight	± Same	± Same
Parts	7	7
Plates	2	2
Weld Length	± Same	± Same
Profiles	5	5
Frames	-	-
Brackets	5	5
Collar plates	-	-

- **Royal IHC current methods:** The current method will not show any differences because the weight of the set-ups stays the same.

- **Statistic Method:** This is the first example in which also the statistic method will not show any differences, because the property values stay the same. This shows the limitations of the statistic method. The equations only use the property values, but not the geometric shape or the relation between the section parts.

- **Graph Database:** The graph database is in this case the method that does deliver a different output from the other methods. Because the method looks at the relations between the different parts and the amount of work that is performed on a structure it will result in a different cost estimation. The graphs shown in the set-up figures show different relations between the parts of the structure and also the label values of the nodes and relations will vary in size. If a closer look is made on the two set-ups the amount of work will vary between them. A plate with five profiles takes more time to produce than a plate with only three profiles but for the second set-up the start-up time of welding profiles to the plates is doubled. This creates higher start-up times for the second set-up. The transport costs are also higher because to avoid above hand welding the crane has to turn the second set-up twice to assemble the brackets to the plates. This is the case for the first set-up as well, but the second set-up is a heavier structure. The plate consists of profiles that are already welded on the plate. Also there might be some trouble reaching the brackets of the second set-up due to the fact that the distribution is different and might be too close to each other to weld comfortably.

7.1.5. Example 5: Collar plate reduction

[redacted] came up with a method of reducing the amount of collar plates in a section. This example illustrates this option for collar plate reduction and discusses how this effects the cost estimation. Set-up one, shown in figure 7.9, shows the set-up with collar plates. The graph shows the collar plates and how they are connected to the rest of the structure with three weld lines creating a watertight connection in the section. The collar plates are connected to the profiles, the standing plate and the bottom plate. The main problem is that the collar plate connections are hard to reach by a welder. The second set-up, shown in figure 7.10, shows the solution that can be used avoid the use of collar plates and reduce the welding time. The connection between the profiles and the standing plate does not necessarily have to be a watertight connection, only the connection between the bottom plate and the standing plate must be watertight. Furthermore in not using collar plates the amount of ironwork in tack welding the collar plates is reduced. A note has to be made that this method ot reducing collar plates can only be used if the connection without collar plates can add enough strength to the construction.

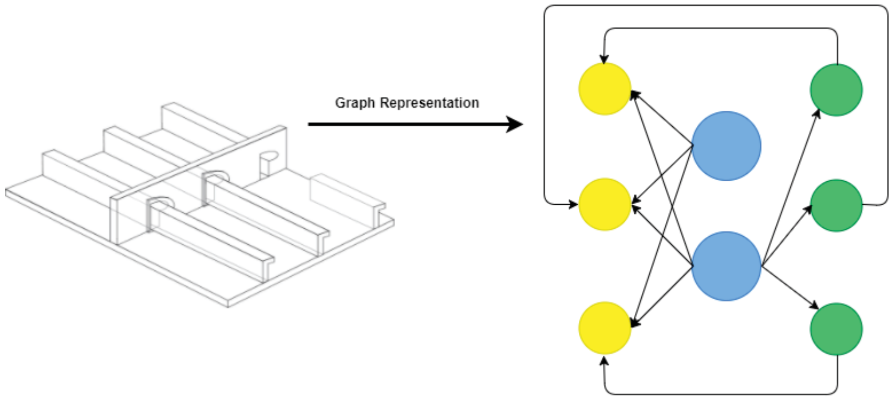


Figure 7.9: Example 5: Setup 1 Graph Representation

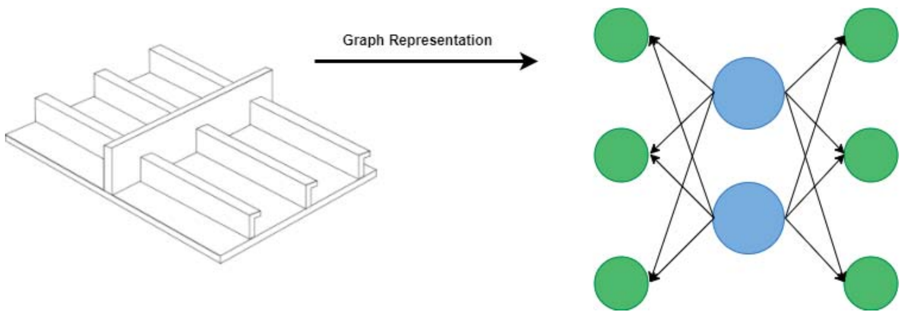


Figure 7.10: Example 5: Setup 2 Graph Representation

Table 7.5 shows the differences between the two set ups. The main difference is in the detailed amount of parts. The collar plates are not seen in set-up two but there are three profiles added. The weight and weld line length stay more or less the same and also the amount of parts stays the same.

Table 7.5: Example 5. Property Differences

	Setup 1	Setup 2
Weight	± Same	± Same
Parts	8	8
Plates	2	2
Weld Length	± Same	± Same
Profiles	3	6
Frames	-	-
Brackets	-	-
Collar plates	3	0

- **Royal IHC current methods:** As holds for the previous examples this method will not give any difference in cost estimation because there is not a significant weight difference.
- **Statistic Method:** The statistic method will show a different cost estimation for both set-ups. To get a good insight in the reduction of costs when avoiding the use of collar plates further research has to be performed. The exact impact of reducing collar plates in a section should be known and taken into

account in creating new equations for this cost estimation method. Chapter 4 did not show a high enough correlation for the collar plates to be taken into account in the equations. The profiles on the other hand are taken into account in the current equations, so the equations will use a higher amount of profiles which leads to a different cost estimation but might not lead to the right cost estimation because of the collar plate reduction but the profile increase.

- **Graph Database:** The graph database method can use the length of watertight welds as a label. This will lead to a reduction of watertight welds for set-up two and thus a reduction in hours. This estimation cannot be performed by the other two methods. The decrease of the amount of collar plates and the increase of profile can also be used in the cost estimation. This results in a more accurate cost estimation than with the other two methods. Another advantage of taking the watertight weld length into account may be that the amount of inspection hours decreases. This is not yet taken into account in the cost estimation method but might be an option for the complete method.

7.1.6. Example 6: Example from the part reduction report [34]

The final example uses a suggestion for part reduction from the research of P. van der Horst [34] in his report on part reduction for the IHC section building. He introduces a method in which the three parts shown in set-up one, figure 7.11, will be reduced to one as shown in set-up two, figure 7.12. The solution combines a profile and two brackets into one plate. This reduces the weld length as well as the setting and attaching time spent on the structure. Only one plate has to be transported instead of three separate parts and the attaching the weld lines between the three different parts is avoided by the solution.

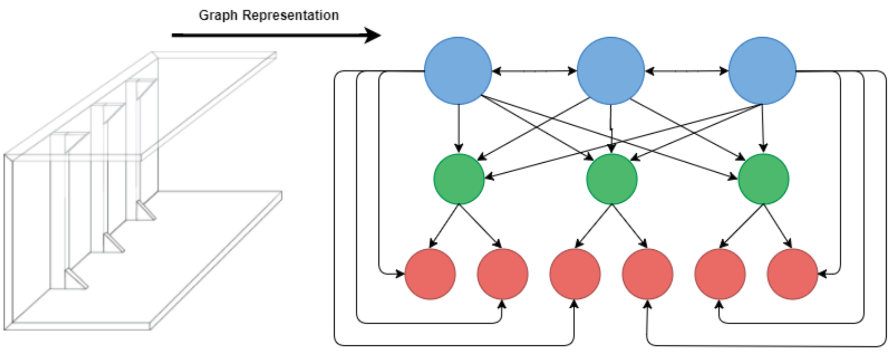


Figure 7.11: Example 6: Setup 1 Graph Representation

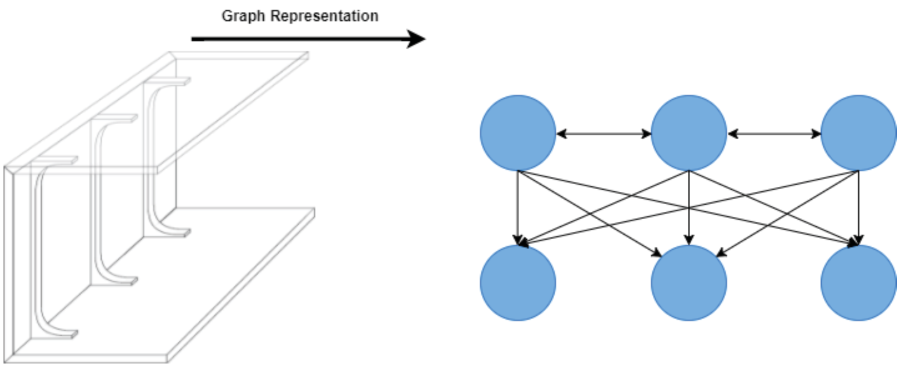


Figure 7.12: Example 6: Setup 2 Graph Representation

Table 7.6 shows the differences between the two set-ups in this example. In terms of weight and plate size there are no differences. But in terms of weld length and the amount of parts there are differences. The amount of parts is halved by implementing plates that give the same strength to the structure as the three similar parts combined.

- **Royal IHC current methods:** The current IHC method will not give any different results because the

Table 7.6: Example 6. Property Differences

	Setup 1	Setup 2
Weight	± Same	± Same
Parts	12	6
Plates	3	6
Weld Length	Higher	Lower
Profiles	3	0
Frames	-	-
Brackets	6	0
Collar plates	-	-

weight of the structure stays equal over the two set-ups. The weight might even increase a bit, according to the conclusion from P. van der Horst [34], leading to a higher cost estimation.

- **Statistic Method:** Using the method in this example will lead to a different cost estimation, because the amount of parts, profiles and brackets decrease in the second set-up. When these variables are used for the equations the estimation will result in a more accurate cost estimation in comparison to the current IHC method.
- **Graph Database:** The graph database will not only take the reduced amount of parts into account, but also the difference in the amount of work. This method can also use the reduction of the work from the second set-up in the cost estimation. This example does also show that this method can also be used for design proposals. An option for the use of a graph database is to create design proposals in the design phase of the ship and the ship sections. If solutions like the one shown in this example can be placed as a pattern into the database then the database can suggest to implement a solution for new section containing similar structures. These recognitions of part constructions can be performed by pattern searches in the database.

7.2. Conclusion

The three methods use different properties and input data to estimate the costs. The qualitative analysis showed with the use of example structures what the limitations and benefits of the methods are. The current cost estimation method at Royal IHC uses the least amount of detail in the input data. It uses the weight of the section to estimate the costs. As was seen in the cost driver analysis, there are higher correlated cost drivers than the weight. In the qualitative analysis was shown that details in terms of parts and the amount of work performed on the sections is left out of the cost estimation if only the weight is used as input data.

The statistic method uses more detail in terms of parts and weld length. This method gave a change in cost estimation between the set-ups for most of the examples. The limitation of this method is that in terms of the work that is performed on the sections it method lacks of input data. Moreover not all the cost drivers are used as variables in the equations from the statistic method. For instance in the collar plate example, example 5, the equations do not contain the amount of collar plates as a variable so the equations might not lead to the desired outcome.

The graph database can contain the most detailed information of the three methods. It can contain all the information from every cost driver and data about the amount of work performed on the sections. From the qualitative analysis follows that using this information may result in the most accurate cost estimations compared to the other two methods. However to be able to use this method, research has to be performed on the impact of changes within sections. Especially for the work performed on a section. The difference in costs for the collar plate example can only be accurately estimated when the effect of a collar plate on the production hours is known. A research that P. van der Horst performed on the reduction of parts is a good example of such a research. This method showed for all the examples that it will lead to a different cost estimation because the graph representations, and thus the data in the database, change per set-up, due to the fact that the database contains information about the relations between the parts in the section.

A graph database can help testing new innovative design or engineering ideas. Whereas the current method tends to be conservative in implementing and testing new principles, the graph database is a more innovative method. For instance the use of a new building material such as aluminium, will not only have an influence on the weight of a section, but also on the production technique and the dimensions of the parts. When implementing a completely new design idea the graph database method will give a more accurate cost estimation and more insight in the design of a new section. The method can search for sections in which a similar principle is implemented instead of only searching for a section with the same weight. But also for previously performed techniques or design principles the database can create a more accurate cost estimation. The example of using fewer parts by combining them into one part shows an example of an optimization of the hours spent on the section. To conclude this chapter a graph database shows the most promising results when analysing the cost estimation of the examples with the three methods. The graph database method can create more accurate cost estimation for both known as unknown cases and it can help by creating cost optimizations in terms of design, engineering and production techniques used.

8
Quantitative Comparison between the Three Cost Estimation Methods

This chapter will compare the results of the three different cost estimation from an example calculation. The methods are: the current BVC method of IHC, the statistical method and the graph database method. The BVC estimations are the estimated hours by Royal IHC. The statistical method estimates the hours with the equations resulting from the regression analysis. For the graph database method the similar section search algorithm and the similarity assessment has not yet been determined, so the cost estimation with this method will be explained with a simplified example. The results of the three methods are set off against the realised production hours.

8.1. Comparison Section: 12781428

The section that will be used for the calculation belongs to order 1278. This section is a midship buoyancy compartment it contains a watertight frame and an inclined tween deck. The section is comparable to the other midship buoyancy compartments. Table 8.1 shows the sections taken into account for the example cost estimation.[1] The table shows the previous sections used for the regression equations calculated in chapter 4, and the new section that will be used as a test section. Section 12781428 has not been used for the cost driver analysis. From the table is can be seen that the new section matches the dataset of sections from the regression analysis, it lies in between the other 1278 sections and the 1265 sections in terms of hours and property values.

The total hours of section 12781428 are composed off:

[redacted]

The section weight changes in the course of the design phase. The current BVC estimation is estimated when the weight can still change from around 2% till 10%. So the new cost estimation should be able to estimate the results within a range of 10%. For the analysis it is assumed that the weld length is also a property containing less detail, as discussed in the regression analysis for scenario 1. The weld length for section 1428 is: 1765.94. This is the total weld length, the weld length for only plates and profiles is also calculated: 1672,82. The difference between these weld lengths is around 100 meter, this is within the range of 10% of the total weld length. From this calculation it is presumed that when the total dimensions of the outer plates, the bulkhead

[1] Section 12781225 is not taken into account for the cost estimation because only 6 or 60 midship buoyancy compartments are used in the equations.

Table 8.1: The sections taken into account in the example cost estimation, including section 12781427

Sectie ID	Frames	Brackets	Collar plates	Profiles	Parts	[redacted]	Weldlength	[redacted]
12781421	29	31	10	102	187	[redacted]	1255.10	[redacted]
12781423	46	60	21	150	279	[redacted]	1577.24	[redacted]
12781425	18	52	34	114	223	[redacted]	1482.64	[redacted]
12781428	**42**	**55**	**49**	**152**	**320**	[redacted]	**1765.94**	[redacted]
12781225	25	92	20	95	268	[redacted]	1086.96	[redacted]
12651421	51	83	80	211	530	[redacted]	2389.92	[redacted]
12651423	52	71	94	158	407	[redacted]	2294.79	[redacted]
12651425	64	70	98	170	424	[redacted]	2371.91	[redacted]

frames and most of the profiles are known the weld length can be estimated early in the BVC process within a 10% error margin from the actual weld length.

8.2. Cost Calculation: Current BVC Method of Royal IHC

From the new section, the 12781428, the BVC hours are known as well as the V1 weight. The BVC estimation for section 12781427 can be seen in table 8.2. The hours of the section were estimated to be much lower than they were realized. Resulting in an estimation error of 18% with the realised costs. The ironwork hours and the welding hours were estimated with an error respectively 18% and 0%. The welding hours are thus estimated without any estimation errors but because of the ironwork hour estimation the total estimation turned out to be too low.

Table 8.2: BVC Hours Compared to the Realized Hours of section 12781428

	[redacted]	Difference
Total	[redacted]	18%
Ironwork	[redacted]	18%
Welding	[redacted]	0%

From the current BVC estimation it is not possible to see directly what caused the estimation error. An analysis about the hour progress over the project can be made, but it is hard to examine what caused the hour estimation to deviate from the realised hours with the current BVC method. The hours are based on the weight and a section difficulty factor, so only these two factors can determine the hour estimation.

8.3. Cost Calculation: Statistic method

The equations that are found with the regression analyses performed are used to test the statistic method in a quantitative way. The method will first be explained by the use of a flowchart. Then an example calculation is performed. The cost estimation of the statistical method is performed see if it is sufficient to use only the two highest cost drivers for a cost estimation. In comparison to the graph database that uses every cost driver and the work performed on the ship section. First of all the real property values are used in the equations and compared to the BVC hours. Later the property values are changed by a maximum of 15% to represent the change in the values during the beginning of the BVC phase and the end of the BVC phase.

8.3.1. Method Description

In chapter 4 the statistic method is introduced. Figure 8.1 shows a flowchart with the suggested strategy how to use the statistic method and when to use which equation. The example cost estimation tries to verify this strategy.

The method starts with looking at the available data at the moment the cost estimation in being performed. If at least the weight and weld length of the new section are known the equation from 6 sections from scenario 1 can be used to calculate the total hours. If this data is not available or is not useful, this method cannot be

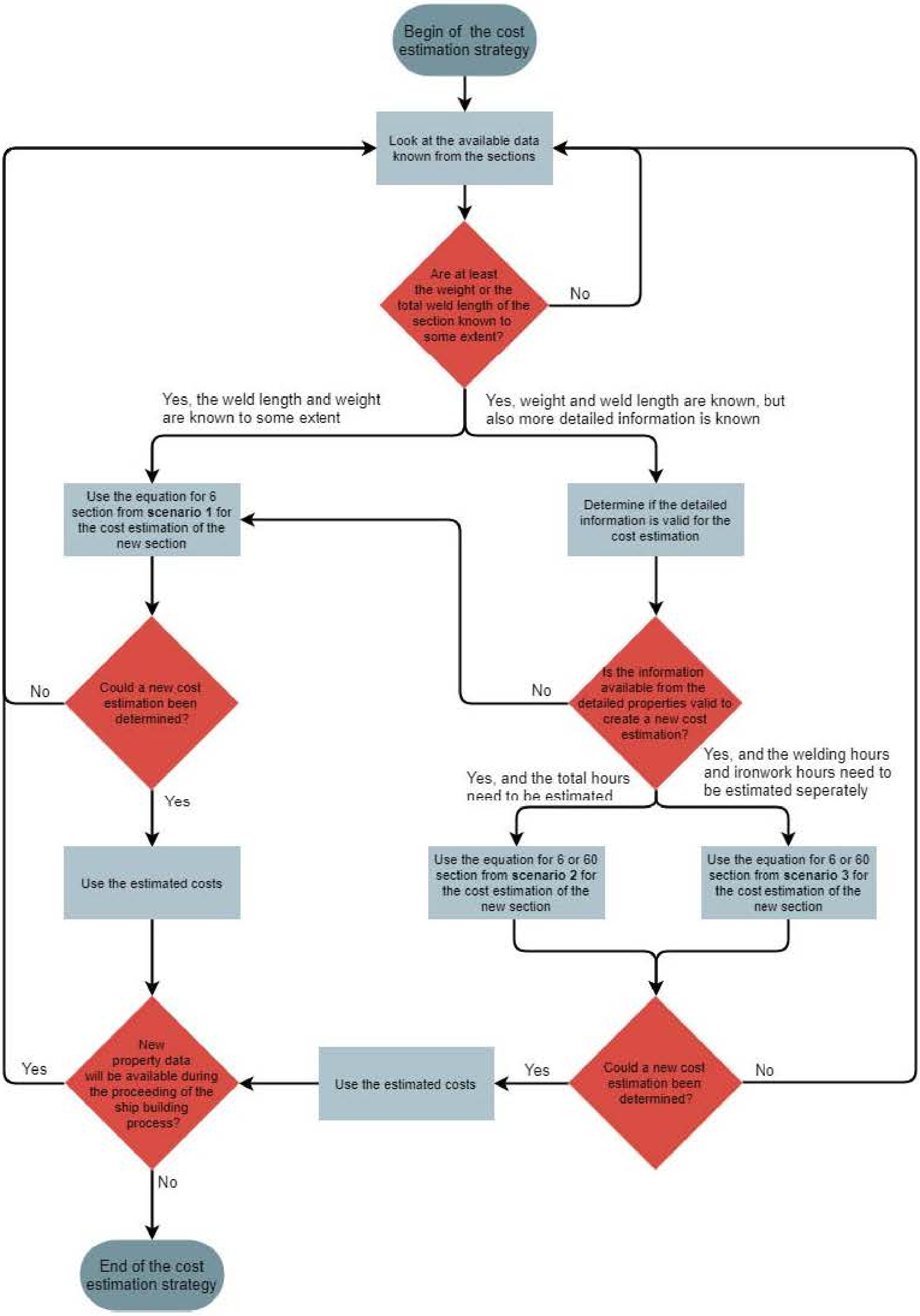


Figure 8.1: Strategic Cost Estimation Method Flowchart

used yet.

If values of more detailed cost drivers such as the amount of profiles and brackets, the other equations can be used to estimate the costs. When the total hours of the section needs to be calculated with more detailed properties the equations for 6 and 60 section from the second scenario are used. If the welding and ironwork hours need to be estimated separately the equations of 6 and 60 sections from scenario 3 are used. When the cost estimations are performed at some point during the BVC phase it could be that changes in the property data occur or that more up to date data is available. The loop of the cost estimation starts from the first decision point in the flowchart.

The equation from the first scenario is used as the first equation because it uses the least detailed properties that are presumed to be estimated quite roughly at the beginning of the BVC cost estimation phase. Also it is assumed that this equation will give a less accurate cost estimation in comparison to the equations from scenario 2. Scenario 2 uses more detailed data from the sections and can be estimated later on in the BVC cost estimation phase. This assumption will be tested in the next subsections.

8.3.2. Example Cost Estimation: Final Property Data

The test will be performed with the different equations chosen from the regression analyses. From the first scenario only the 6 section equation is tested. From the second scenario both the 6 and 60 section equations are used and the ironwork hours and the welding hours are calculated with the third scenario equations for 6 and 60 sections. Below the equations are listed.

- Scenario 1
  - **6 sections:** $Hours = -85 - x_{weight} * 0.05 + x_{weldlength} * 1.70$
- Scenario 2
  - **6 sections:** $Hours = -105 + x_{profiles} * 3.80 + x_{brackets} * 7.25$
  - **60 sections:** $Hours = -95 + x_{profiles} * 5.20 + x_{frames} * 4.70$
- Scenario 3
  - **6 sections - Ironwork hours:** $Hours_{ironwork} = 90 + x_{profiles} * 0.35 + x_{brackets} * 2.60$
  - **6 sections - Welding hours:** $Hours_{welding} = -280 + x_{profiles} * 3.00 + x_{weldlength} * 0.20$
  - **60 sections - Ironwork hours:** $Hours_{ironwork} = 22 + x_{profiles} * 1.00 + x_{frames} * 2.50$
  - **60 sections - Welding hours:** $Hours_{welding} = -115 + x_{profiles} * 4.15 + x_{frames} * 2.20$

In table 8.3 the results of the equations are given. The table also shows the difference between the calculated and the realized hours in percentage and in hours. Looking at the different scenarios, the first scenario gives the least accurate cost estimation. In this equation the properties used are the weight and the total weld length. The total weld length is also used in the equation for the welding hours with 6 sections from the third scenario that gives a accurate estimation. This example estimation shows that using the weight in a cost equation does not give the most accurate cost estimation. The second scenario gives more accurate results compared to the first scenario, especially when the equation for 60 sections is used. Also the third scenario results in a more accurate cost estimation when the 60 section equations are used.

Table 8.3: Results from the Example Cost Estimation

			Percentage	
		[redacted]	6 sections	60 sections
Scenario 1	Total		17%	x
Scenario 2	Total		13%	10%
Scenario 3	Ironwork		14%	15%
	Welding		4%	4%
	Total		18%	11%

8.3.3. Example Cost Estimation: Changing Property Data During the BVC Phase

The results from the weight analysis in chapter 4 shows that the weight estimation differs from 15% to 4% from the real weight during the BVC phase. This subsection will show if the statistic method gives an accurate cost estimation when the property data varies over the BVC phase. Not only the weight data will be changed, also the other properties used in the equations will be changed with the same ratio. Because only the weight changes are documented during the BVC phase at IHC it is assumed that the maximum changes of the other property values are also in the same range as the weight.

As a maximum percentage 15% is taken. Then a middle value of 10% and as lowest variation 2% is used. Also the real property values of the section are rounded as the 0% base. Table 8.4 shows the changed variable data used to calculate the hours. Only the properties that are used in the equations are shown in the table. The tables 8.5, 8.6, 8.7 and 8.8, show the results of the change in property values.

Table 8.4: Alternated Property Data Representing the Value Change during the BVC Phase

		laslengte	profielen	Frames	brackets
real		1765.94	152	42	55
rounded		1750	155	40	55
2% up		1801.259	155.04	42.84	56.1
2% down		1730.621	148.96	41.16	53.9
10% up		1942.534	167.2	46.2	60.5
10% down		1589.346	136.8	37.8	49.5
15% up		2030.831	174.8	48.3	63.25
15% down		1501.049	129.2	35.7	46.75

Table 8.5 shows that there are only small differences between the hour estimations of the real values and the rounded values. Overall when the values are rounded they give a slightly less accurate cost estimation.

Table 8.5: Real values against the Rounded Values

		Difference Real Values		Difference Rounded Values	
		6 sections	60 sections	6 sections	60 sections
Scenario 1	**Total**	17%	x	21%	x
Scenario 2	**Total**	13%	10%	11%	10%
Scenario3	**Ironwork**	14%	15%	14%	15%
	Welding	4%	4%	3%	5%
	Total	18%	11%	18%	10%

Table 8.6 shows the differences between the values that are 15% up of 15% down. The 15% decreased values show higher differences with the realized hours than the real values. The 15% increased values show a more accurate cost estimation. From these results a clear difference between the 6 or 60 section equations cannot be found.

Table 8.6: Property values 15% up against 15% down

		Difference 15% up		Difference 15% down	
		6 sections	60 sections	6 sections	60 sections
Scenario 1	**Total**	4%	x	31%	x
Scenario 2	**Total**	2%	4%	27%	25%
Scenario3	**Ironwork**	11%	11%	17%	19%
	Welding	8%	15%	16%	7%
	Total	3%	4%	33%	26%

Table 8.7 shows the cost estimation with the variable values 10% up and 10% down. The 10% alternation represent the phase around the time the construction weight is estimated. The results of the 10% decreased values are more accurate compared to the estimations made with the 15% values. The 10% increased values are less accurate than the 15% values. Overall the values are becoming more accurate when the alternation is lower.

Table 8.7: Property values 10% up against 10% down

		Differnce 10% up		Difference 10% down	
		6 sections	60 sections	6 sections	60 sections
Scenario 1	Total	8%	x	26%	x
Scenario 2	Total	3%	1%	22%	20%
Scenario3	Ironwork	12%	13%	16%	18%
	Welding	4%	11%	12%	3%
	Total	8%	1%	28%	21%

Table 8.8 shows the difference in cost estimation when the variables are changed 2% up and 2% down with the realized hours. The 2% alternation shows the data values on which normally at IHC the final BVC cost estimations are made. The differences are closest to the cost estimation with the real section property values.

Table 8.8: Property values 2% up against 2% down

		Differnce 2% up		Difference 2% down	
		6 sections	60 sections	6 sections	60 sections
Scenario 1	Total	15%	x	19%	x
Scenario 2	Total	11%	8%	15%	12%
Scenario3	Ironwork	14%	15%	15%	16%
	Welding	2%	5%	6%	3%
	Total	16%	9%	20%	13%

The tables above show that the more the variable values change the less accurate the cost estimations are. Table 8.9 shows the same results. These are the average differences of the total hours. Ironwork hours and welding hours are left out of average value. This table also shows that when the property values are increased the estimations become more accurate. Overall a difference of around 15% with the real values is encountered with the statistical analysis during the complete BVC phase. In which the 15% increased values show the least accurate cost estimation.

Table 8.9: Average Difference between the Different Property Value Alternations

	Real	Rounded	15% up	15% down	10% up	10% down	2% up	2% down
Calculated	14%	14%	3%	29%	4%	24%	12%	16%

It should be noted that the values are all increased or decreased with the corresponding percentage, however in real life these variables do probably not all change with the same ratio. The results of these calculations thus show the maximum and minimum results when looking at the percentages when the property values do all change with the same ratio. The differences with the realized hours could change when the property values alternate with different ranges. For instance the first scenario 6 section equation: $Hours = -85 - x_{weight} * 0.05 + x_{weldlength} * 1.70$, starts with a negative value of which the weight is extracted to the result of this extraction the weld length is added. One can notice that if the weight is increased and the weld length stays the same or decreases. This equation will give less accurate results than when the property values change with the same ratio. Therefore the range introduced in chapter 4 can be used as boundary condition.

8.3.4. Conclusions

The first step in the flowchart in figure 8.1 is to look at the equation of 6 sections from the first scenario. This equation uses the weight and the weld length as variables. It is set as the first equation to use because the weight and weld length can be estimated in an early stage of the BVC cost estimation phase. It was presumed to give the least accurate cost estimation. The results from the example calculation prove this assumption. The equation from the first scenario gives the least accurate cost estimation. Also the equation uses negative variable values which might lead to wrong cost estimation results if the variables change to much. It is preferred to leave this equation out of the strategy for the statistical cost estimation method.

The second scenario equations for 6 and 60 sections use more detailed information than the first scenario equation and estimate the total hours of the section. These equations give more accurate results than the first scenario. Especially the equation for 60 sections gives good results. This is the preferred equation from scenario 2. Scenario 3 also uses more detailed information than the first scenario but it estimates the welding and ironwork hours separately. From this scenario not a clear conclusion can be drawn whether the equations for 6 or 60 sections give the most accurate results. When summing up the results of these equations the 60 sections give a more accurate estimation of the total hours but the 6 section equations result in slightly more accurate ironwork and welding hours. Because this scenario is used to calculate the ironwork and welding hours separately the equation for six section is chosen at the moment. If more research is performed the equation choice might change or the equations can change.

The method shows more accurate results when the values are increased. This means that either the equations from the regression analysis lead to a underestimation or that something went wrong during the production of the section that led to high real production hours. To figure out what effects resulted in the high production hours an analysis has to be performed during the section or more sections have to be estimated with the statistical method. The statistical analysis can give some insight in the hour evaluation at the end. Mainly on the cost drivers used in the equations and on the difference between the estimated hours and the realized production hours.

8.4. Cost Calculation: Graph Database

The general working principle of the graph database method will be tested in this section. Although the graph database algorithm and the similarity assessment have not yet been finished, with a simplified example the method can be tested and conclusions can be drawn on the accuracy of the method. The section starts with the method strategy explained with a flowchart. After the strategy is tested with the same example section used in the other two methods, the 12781428.

8.4.1. Method Description

Figure 8.2 shows the cost estimation strategy when the database is used. The general strategy is known, but as the figure shows there are two black boxes in the strategy of the cost estimation. The first box is the search method for similar sections in the database and the second is the similarity assessment of the chosen sections. The general principle of the black boxes will be explained in this section with an example.

The method starts with modelling the section properties into a section graph in the same way as the other sections in the graph database. The first black box is shown, the section will be run through the database to search for similar sections. If there is no similar section found yet, either because the information is not sufficient to find a similar section or because something went wrong with the search method, the strategy is to go back to the transformation from the new section into a section graph. This is to find the error in the search method or to wait for more data to use. From which point the data is sufficient for the search method has to be examined by doing more research on this topic.

If there are similar sections found in the database the similarity assessment or the cost function will estimate the costs, the second black box in the cost estimation process. The differences between the sections determine the cost estimation. It is not known how much influence a difference in label values has on the total hours of the section. By a simplified example the principle of the similarity assessment will be explained, but to be able to use the assessment on real ship sections more research has to be performed how to assess the differences between sections.

When the similarity assessment is performed it either leads to a cost estimation or something went wrong during the cost estimation process. If the cost estimation is accurate has to be assessed, the requirements of an accurate cost estimation have to be investigated in further research. If no cost estimation can be found the section has to be run through the database again or the similarity assessment has to be performed again. It could be that the data at the beginning of the cost estimation was too limited for an accurate similarity assessment or that the similarity assessment was not performed correctly. When an accurate cost estimation could be received from using the graph database these costs can be used as final estimated costs, when more data becomes available the cost estimation can be performed again.

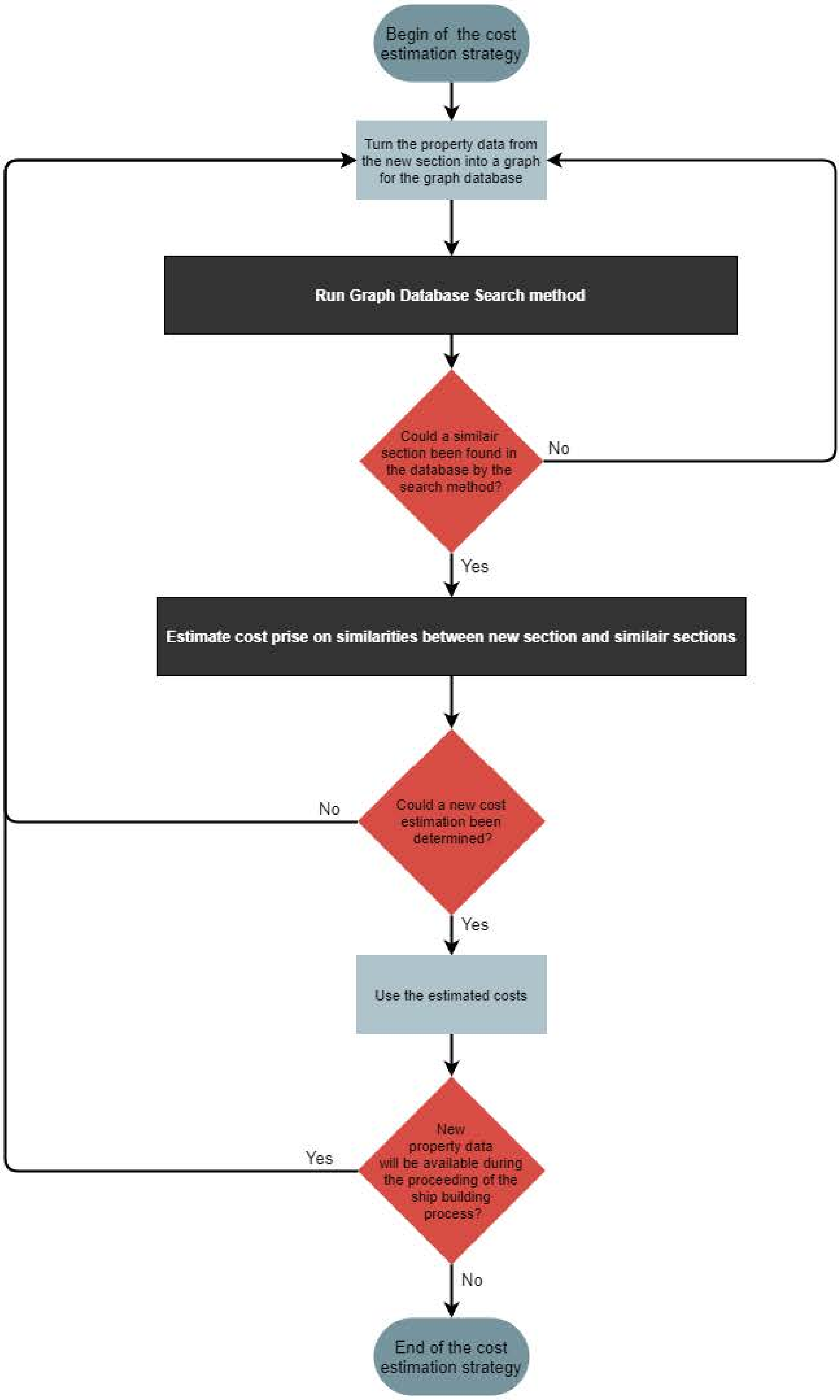


Figure 8.2: Graph Database Cost Estimation Method Flowchart

8.4.2. Example Cost Estimation Explanation

Because there are still two phases of the graph database cost estimation unknown a simplified cost estimation example will show the principles of the method. This example is performed using excel and the labels with their relation to the hours known from the statistic method. First the search for the most similar sections will be performed. Then the similarity assessment will be explained.

Similar Section Search

The similarity assessment is performed with the correlation knowledge gained from the cost driver analysis. It is known which label has the highest relation toward the hours and which label the least. Furthermore the values from these labels can be used to calculate the differences between the values of the new section and the most similar sections can be found.

Appendix W shows the results of the excel search for similar sections. The first table shows the percentage differences between the label values of the new section, 12781428, and the old sections. The second table multiplies these percentages with a correlation ranking number. These ranking numbers are based on how much the label correlates with the total hours. For instance the profile percentages are multiplied by 7 and the frames by 3. Looking at the algorithm discussed in chapter 5, where the ranking number is called a relevance ratio, the equation for this example case will look like this:

$$t_{sim} = 7 * x_{profiles} + 6 * x_{parts} + 5 * x_{brackets} + 4 * x_{weldlength} + 3 * x_{frames} + 2 * x_{collarplates} + 1 * x_{weight}$$

The reason why percentages are used instead of the real values is because the values for the different labels are not in line with each other. If for instance the weight difference would be multiplied, in the same range of ranking numbers, with the profiles difference the weight value would have significantly more influence on the final result than the profiles. The final results of the section similarity search is shown in table 8.10.

Table 8.10: Results of the Example Similarity Search

Section ID	Similarity	[redacted]
12781423	3.30	
12651423	5.73	
12651425	6.21	
12781425	6.89	
12651421	9.34	
12781225	10.39	
12781421	12.07	
12781428	**Total Hours:**	

The table shows that there are three sections more similar to section 12781428 than the others. Sections 12781423, 12651423 and 12651425 rank the closest to the new section. When looking at the total hours these sections are also the most similar to the new section. The other four sections differ more from the new sections. In the similarity assessment these three sections will be used to demonstrate this part of the cost estimation process.

Alternating Section Data

Just as the statistic method, the graph database method is meant to be used during the complete BVC phase. The same analysis used with the final label data is performed with different label data to see if different sections are chosen during the search method when the label data is alternated. The label values are changed with 15%, 10% and 2%. This analysis and the results of the analysis are shown in appendix W.

When the values are changed 2% down or 10% down the same sections in the same order are the result of the search method. When the data is changed 2% up the same sections result from the search method are given but in a different order. The other label variations gave different sections as a result from the similarity search.

The hours of the similar sections are not directly used as the new costs for the new section but they are subjected to the similarity assessment. So using the hours of a section that is the outcome of the similarity search in the beginning of the BVC phase, 15% off, might not be a big issue if the similarity assessment turns their hours into a good cost estimation for the new section.

Simplified Similarity Search Method Limitations

The simplified similarity search method performed in this section has limitations compared to the intended graph database cost estimation method. In the database more labels can be taken into account than used in this simplified search method. The ranking numbers will probably be different from the ones used in this example. Currently the ranking numbers are integers, but they could also be decimals or not used as a number but as a filter. The latter means that if profiles are the highest correlated label to the total hours, the similar sections will first be filtered on profiles and then on the second correlated label.

Similarity Assessment

The similarity assessment uses the outcome of the similar section search algorithm and the differences between these sections and the new section. A similarity assessment, or cost function in CBR terms, approach is not yet known. This part discusses a few options. The similarity assessment can be performed outside the database or inside the database. Many options are still possible. This research uses excel to process the data and perform some similarity assessment tests. If the cost function is made outside the graph database, it can still used to perform the similar section search and for the analysis in the end.

Table 8.11 shows the data and the difference ratio of the three sections most similar to the new section. The simplified similarity assessment performs one cost estimation with the difference ratios and one with the differences between the label values.

Table 8.11: The Data of the Similar Sections Including the Difference Ratio

Section ID	Profiles	Parts	Brackets	Weld Length	Frames	Collar Plates	[redacted]	Difference Ratio
12781427	152	320	55	1765.94	42	49	[redacted]	
12781423	150	279	60	1577.24	46	21	[redacted]	3.30
12651423	158	407	71	2294.79	52	94	[redacted]	5.73
12651425	170	424	70	2371.91	51	80	[redacted]	6.21

Difference in Label Data between the Old and the New Section

The label values between the new section and the old sections show differences. The similarity assessment uses the effect of these differences on the hours to estimate the costs of the new section. With these differences the increase or decrease of the hours can be calculated.

Using this method requires knowledge of the influence the different labels have on the total hours of the section. For instance if the most similar section contains five profiles more than the new section, the hour increase these five profiles will have to be researched. Extra research in the hour increase or decrease the different in labels will have has to be performed. IHC has a factor to calculate the hours needed for welding. This factor will be used to show the principle of the similarity assessment based on the difference of label data. By doing the following analysis, in the real similarity assessment, the correlation of the labels to the total production hours are taken into account. For instance by multiplying the results of the similarity assessment with the ranking order of the correlation matrix.

Table 8.12: Factors for the Difference between Label Data: Weld Length Example

	[redacted]	Percentage Difference
Real Data	[redacted]	9.37%
2% up	[redacted]	8.49%
2% down	[redacted]	11.20%
10% up	[redacted]	8.04%
10% down	[redacted]	18.31%
15% up	[redacted]	6.91%
15% down	[redacted]	18.24%

Table 8.12 shows the results with the weld length factor in the cost estimation. The calculation that multiplies the difference in weld length with the weld factor from IHC, 4 meter per hour[2] These hours are shown under calculated hours in the table. The overall cost estimation using this method is accurate. When the variable values are increased the cost estimation becomes less accurate. When the values are decreased the cost estimations become more accurate. However the differences variate less than with the statistical analysis, so it leads to more stable cost estimations over the BVC phase.

Table 8.13: Different Number of Sections used for the Label Difference Cost Estimation: Weld Length Example

		Percentage Difference		
	[redacted]	**3**	**2**	**1**
Real Data		9.37%	7.89%	10.17%
2% up		8.49%	7.00%	9.28%
2% down		11.20%	8.77%	11.05%
10% up		8.04%	9.78%	6.44%
10% down		18.31%	19.45%	14.60%
15% up		6.91%	4.05%	4.46%
15% down		18.24%	24.70%	29.67%
Average		**11.51%**	**11.66%**	**12.24%**

Table 8.13 shows the results when a variation in the number of similar sections is used to estimate the costs. Three, two or one similar sections are used to estimate the costs by looking at the weld length difference. The Last row shows the average results of the different cost estimations. This row shows that there are not that many differences in using one or more sections. Leading to the conclusion that if the difference between label data from the similar section(s) is turned into an accurate cost corrections for the total hours of the new section it does give as accurate cost estimation. It does however not eliminate the need to search for the most similar sections. Looking for the most similar sections filters the data that are used for the cost estimation, in order to use the most relevant data and to limit the amount of calculations that have to be performed.

The Total Difference between the Old and New Sections

The second similarity assessment approach is the less detailed than using the difference in label data. This method uses the difference factor calculated during the search method. A ranking can be made according to the order of the factors. The final hours of a similar section are multiplied with this ranking number, in this example the total hours of the 12781423 are multiplied by three, the 12651423 by two and the 12651425 by one. The sum of these multiplications is divided by the sum of the ranking numbers, in this case six.

Looking only at the total difference percentages has the advantage that no extra research is needed to figure out what a difference between label data will mean for the section hours. So the need for extra research is minimized and the method can be implemented whenever the first black box, the similar section search algorithm, is finished.

Table 8.14: Difference Factor used in the Similarity Assessment Results

	[redacted]	Percentage Difference
Real Data		4.79%
2% up		4.79%
2% down		9.44%
10% up		1.11%
10% down		16.40%
15% up		6.41%
15% down		19.22%

Table 8.14 shows the results of an example using the difference factor for the similarity assessment. This example uses the three most similar sections because their difference ratios are close to each other and

[2] the assumption is made that the extra weld length is performed with an underhand welding position. Most of the weld length within a section is performed underhand and added or subtracted by the total hours of the similar section.

differ significantly with the other sections. When the final label data is used during the estimation the hour estimation gives a result of 4.79% from the real hours. The lower the variations of the data the more accurate the estimations become.

Table 8.15: Different Number of Sections used for the Percentage Cost Estimation

	[redacted]	Percentage Difference		
		3	**2**	**1**
Real Data		4.79%	6.62%	14.90%
2% up		4.79%	6.62%	14.90%
2% down		9.44%	6.62%	14.90%
10% up		1.11%	0.72%	3.88%
10% down		16.40%	20.54%	14.90%
15% up		6.41%	0.72%	3.88%
15% down		19.22%	26.18%	31.82%
Average		**8.88%**	**9.72%**	**14.17%**

Table 8.15 shows the cost estimation results when the analyses are performed with a variation in the number of sections. The last row of the table shows that if only the most similar sections are taken into account it does not lead to accurate results compared to using more similar sections. If two or more of the most similar sections are used it leads to more accurate results.

8.4.3. Conclusions

The graph database method is tested by an example estimation. The method contains two black boxes, the similar section search algorithm and the similarity assessment. The general principle of these parts are known, but the execution has to be optimized.

The example similarity search shows that when the exact or slightly changed input data from the new section is used the analysis gives the most similar sections as an output of the analysis. When the input data is changed with 10% the output sections are not the most similar sections in terms of final hours and the label values.

Two options for the similarity assessment are discussed in this chapter, based on the differences between label data and one based on the total percentage difference between the new section and the most similar sections. Looking at the results, shown in table 8.13 and 8.15, the option of using a correction for label differences leads to more accurate results than using the total hours. The cost estimation by using a correction for the label difference is less susceptible to data changes and to the use of one or more similar sections.

The downfall of using the differences between the label data is that the time to implement this method is longer than when using only the total difference percentage, because of the research that is needed to state what effect the difference will have on the total hours. More data is needed to significantly investigate the effect of the label value differences. Where in using the total percentage and a ranking of the similar sections requires less additional research.

The example cost estimation uses data that has been gathered during the research and where the correlation to the total costs is known from. These are the amount of detailed parts, weight and the weld length. The real graph database will contain more labels of the sections and if the correlation to the hours are known, these labels can be used in the search method. The similarity assessment is also tested in this example estimation. This example does not state if adding more label data and using an optimized similarity assessment, leads to a more accurate cost estimation.

8.5. Comparison of the Three Methods

This section will compare the three different cost estimation methods with each other. Table 8.16 shows a summary of the percentage difference with the estimated costs and the realized costs. When comparing the BVC estimated hours with the statistical method and the graph database method it shows that the average of the estimations is higher than the results from the other estimations. The statistical method shows a slightly more accurate cost estimation, around 4% improvement compared to the BVC estimation. The graph database method shows even more accurate results. The weld length example shows around 7% improvement and the similarity factor around 10%. When looking at the final production data the graph database method performs even better, for the weld length example 9% improvement and for the similarity factor method 13% improvement.

Table 8.16: Summary of the Three Cost Estimation Methods

	BVC	Statistical	Graph Database	
			Weld Lenght	**Similarity Factor**
Real	18%	14%	9.37%	4.79%
2% up		12%	8.49%	4.79%
2% down		16%	11.20%	9.44%
10% up		4%	8.04%	1.11%
10% down		24%	18.31%	16.40%
15% up		3%	6.91%	6.41%
15% down		29%	18.24%	19.22%
Average	**18%**	**14.6%**	**11.51%**	**8.88%**

Table 8.17 shows a Harris profile of the three cost estimations ranked on the goals and requirements for the new cost estimation method. A Harris profile ranks the different methods on the different goals. In terms of an accurate cost estimation the current BVC estimation method is ranked last. There is no possibility to use multiple cost drivers and to see the work performed during the production process. Further more the summary in table 8.16 showed that the method results in the least accurate cost estimation. The summary shows that using more and other properties than used in the statistical method can create a more accurate cost estimation. The graph database method seems to be the most promising when it comes to the most accurate cost estimation. The weld length example showed improvement in regard to the current method estimation. If the similarity assessment is not only based on weld length but other labels can also be used for the cost estimation, it may lead to even more accurate cost estimations. Estimating the costs with the similarity factor led to the most accurate results. In terms of the possibilities of using more cost drivers and showing the work performed on the production the graph database also ranks the highest, because this is the only method that has the option to use the production work.

In terms of creating more integration between the cost estimation phases, the current method and the statistical method both score low. Neither om them are able to test cost optimizations and the current method uses multiple different programs to estimate the costs for the different phases. The statistical method can estimate the costs for the BVC phase. If the equations perform well with the data known during the VC phase has to be investigated in further research. The graph database performs well on the integration goal. It will be one method that can suggest optimizations regarding the costs and is presumed to estimate the costs for the VC phase more accurate then the other two methods, if this is the case has the be investigated in further research.

The evaluation possibilities of the current BVC method and the statistical methods are fewer than the graph database. For the current method it is not possible to perform an evaluation for future estimations and it is hard to perform an estimation based on the cost drivers used. Only weight can be taken into account for the hour evaluation. For the statistical method it is possible to evaluate the section hours on the properties used in the equations. However the graph database provides more evaluation possibilities, it is possible to provide suggestions for future estimations and to base the evaluation on more than one cost driver.

One thing is added to the goals, the implementation of the method. The current method scores high on this goal because it is already there and only needs to keep the data up to date. The statistical method needs some implementation especially in terms of an improved cost driver analysis and regression analyses including data of more sections. The equations calculated with more sections give the most accurate results. When more data is available the equations can be recalculated. Also if the equations are suitable for one type of section or for more types has to be seen when data from more sections become available. The graph database method needs more work before it can be implemented. However it seems to be the most promising method in terms of the other goals, therefore an implementation plan is made.

Table 8.17: Harris Profile

	BVC				Statistical				Graph Database			
	-2	-1	1	2	-2	-1	1	2	-2	-1	1	2
1. Creating an accurate cost estimation												
1.1 possibility to use multiple costdrivers												
1.2 seeing the work performed on production												
2. Creating more integration												
2.1 Possibility for optimization regarding costs												
2.2 Limiting the use of different programs												
3. Evaluation possibilities												
3.1 Possibility of evaluation for future estimations												
3.2 Evaluation on more than one cost driver												
*** Implementation**												

9

Conclusions

The performed research compares three cost estimation methods during the BVC phase with one another. The first being the current cost estimation method, the second being a statistical method and the third method is the graph database method. The main goal of the research was to perform a feasibility study for the case based reasoning cost estimation method. First the main research question will be answered.

"Do the statistical cost estimation method and the graph database method give more accurate cost estimation in comparison to the current IHC cost predictions methods in both the VC and the BVC phase?"

Quantitative research performed in chapter 8 showed that both the statistical method and the graph database cost estimation method result in more accurate cost estimations than the current method performed by IHC. The test performed compared the cost estimations of the three methods to one another. The BVC hours showed an estimation error of 18% compared to the realized production hours. When the final values of the properties are used to estimate the costs, the statistical method average was around 14% off compared to the realized hours. The graph database method showed even more accurate estimations, an error of 9.5% when using only the weld length differences and 4.5% when using the difference factor.

The BVC hours are estimated with property values that are different from the final property values before production, so for both the statistical and the graph database method the property hours are changed from 0% to 15% to match the BVC property values. The average from the statistical method estimation error varies from 14% with realized hours to 29% when the property values are varied with 15%. When the similarity assessment of the graph database uses the weld length difference for the cost estimation with changed property values, the estimation error remains quite stable across the alternations, having an average of 11.5% and a range from 9.5% to 18%. If instead of the property difference the total percentage difference is used, the estimation errors do increase when the property values are alternated. They increase from 4.5% when the real values are used to 19% when the values are altered with 15%.

Qualitative research of the three methods showed that the property values used have an influence on the cost estimation. The current cost estimation methods of Royal IHC compare the weight of the new case to the weight of old cases and translate this into a cost estimation. The qualitative analysis showed that only looking at weight ignores properties that do have a significant influence on the hours. Examples of ship section panels show that using more properties then only weight will lead to different results from a cost estimation. The statistical method takes the properties with the highest correlations to the hours into account, which show alternations in the cost price for more examples than the current method. When using even more properties and also taking the used techniques into account every example shown gave an alternation in the cost price.

One of the suspected shortcomings, discussed in the introduction, of the current estimation was not taking into account the production techniques from the cost estimation. This research confirms that this is indeed the case. Especially the graph database method solves this problem by including properties describing the welding technique used and more properties that describe the section in more detail. This method led to the most accurate cost estimations during the test.

Another problem mentioned in the introduction is that currently IHC has multiple cost estimation methods, for instance the two different methods for the VC and the BVC phase, leading to less integration over the cost estimation process. During the research the VC phase turned out to be a completely different process compared to the BVC phase, time constraints made it not possible to perform the analyses for both the VC and the BVC phase. So whether the new methods give more accurate cost estimations in the VC phase cannot be concluded from this research. Whether the method can therefore solve the second problem of combining both the VC and the BVC estimation methods into one method cannot be verified. From the BVC example estimations it can however be concluded that the new methods can be implemented in the total BVC phase, so the need for multiple estimation methods based on the level of details known is solved.

"Is it possible to gain more insight, by using the different cost estimation methods, in the cost drivers of the production hours and in aspects that lead to deviations between the cost estimation and the final costs of the ship building?"

Using the current method the insight in the final cost drivers is limited with the weight correlation to the final hours and the effect different type of section have on the final hours. The analysis at the end of the production process is based on the planned hours and the total hours, so there are limited possibilities to get insight into the aspects that lead to the final hours. The statistical analysis uses a correlation analysis and a regression analysis to compose the equations. This gives an insight in the correlations between the final hours and the different production properties. The statistical analysis however does not give any more insights in the end of project analysis than the current methods. There are more properties included in the analysis but the effect of these properties on the final hours are not shown, just an overall correlation between the property and the final hours and an equation to estimate the hours used. For both of these methods the deviations between the estimated hours and the final hours at the end of a project cannot be examined, but from the included properties the correlation of the property values and the final hours are analysed before the project.

The graph database can offer more insight into the effects ship section labels have on the final hours. The correlations between property and the final hours have to be known in order to search for similar sections. This gives an insight into the properties that have to highest correlation between the final hours, however the effect of the properties on the hours are not known yet. For the similarity assessment the effect a property has on the final hours must be known in order to estimate the hours. This method does lead to more insights in what properties are cost drivers in the project. It may also give more insight in what aspects lead to deviations between the cost estimation and the final hours. This can be analysed at the end of the project by using data that is available in the database and changes made in this data during the project. This only counts for the data that is taken into account, so the part characteristics and the production techniques used. There are properties causing noise between the estimated and final hours that are not included in the database, such as human factors or weather influences. The effect of these factors cannot be analysed by the method used at this point.

"Is Case Based Reasoning a feasible cost estimation method for the ship building industry?"

The answer to this question according to the research is that CBR used in a cost estimation is a feasible method. The graph database method uses the CBR theory in the cost estimation, from the analyses this method turns out to be the most accurate. The accuracy of the cost estimation however depends on the amount of data that is used in the cost estimation. The current cost estimation uses an analogy way of estimating the costs and uses historical data. This method is the least accurate cost estimation method. The parametric statistical method shows a more accurate cost estimation using more cost drivers than the current BVC cost estimation. However this method is not as accurate as the graph database method, so the graph database method is the most preferred method of the three discussed methods. The main conclusions of this research is that using more cost drivers than only weight and including the work performed in the production lead to more accurate cost estimations.

10

Future Developments & Discussion

In this chapter future developments for IHC and for the graph database method are discussed. Recommendations are made for further research and the limitations not yet discussed in the report are listed. Furthermore discussion points that arose during the research are described and the type of data that is used during the research and the validity of the research will be discussed.

10.1. Implementation Plan for the Graph Database Method

The implementation of the method is part of the feasibility of the method, if the implementation or creating the database is time consuming it might be hard to start using the method at the company. This section will discuss the implementation of the method.

10.1.1. The Section Cost Estimation Method

Now that the graph database cost estimation turns out to be the most accurate and promising cost estimation method, an implementation plan is made. As discussed in chapter 8 there are still two black boxes in the cost estimation method that need to be optimized.

- **Similar Section Search Algorithm:**
  - Gather more data to fill the graph database.
  - Test the detailed data on correlation towards the production hours. This research only tests the basic label data, because the detailed welding data was too limited to create proper correlations, so this part has to be tested.
  - Create a ranking of the properties to search for similar sections.
  - Create a search algorithm. Whether every property is used and do they have the same weight in the algorithm. Whether the similarity is based on sub-panels or on the complete section. These are examples of questions that need to be answered to be able to create the search algorithm.
  - Investigate when a section is similar. During the example cost estimation in chapter 8 there where more than one similar sections that were close to the new section. To which extent a section is similar has to be researched.
- **Similarity Assessment:**
  - Find out what kind of similarity assessment will be performed. If this is done on the total similarity percentage of the similar sections to the new section or by a different approach. If the similarity percentage is used, the similarity assessment does not need many extra research.

  - Look at the effect of each property towards the hours. For instance if there are five profiles more or less the effect of these five profiles on the hours of the new section has to be determined.
  - Create the similarity assessment method.
  - Determine if this similarity assessment is accurate to estimate the costs of the new section. Before the similarity assessment can be used, it first has to be tested if the assessment estimates the cost of the new section accurate.

- **VC phase:** See if the method is also valid for the VC phase. The cost estimation method is tested for the BVC phase. To see if it works for the VC phase has to be tested.
- **Hour Evaluation:** The method seems promising for the hour evaluation phase from the qualitative analysis. When the method is finished the possibilities for the evaluation phase should be tested.
- **Updates:** implement developments or updates along the way and delete old or incorrect data in the database to be able to keep the cost estimation accurate.

The test performed with the limited data of this research showed accurate results. The gathering of more data has to be performed and the similarity assessment has to be designed to be able to get even more accurate results. The gathering of more data might be a problem for the method. The relation lists were created manually for this research which was a time consuming task. A new CAD program might help to built these relation lists. This program will be implemented at IHC and is called NX.

10.1.2. Implementation of NX

The building of the ship section relation lists turned out to be a time consuming manual task during this research. These lists are however required for the working principle of the graph database method. Implementing the graph database if the sections graphs have to be constructed manually, the data gathering will be a labour intensive part of the method. The current programs used at IHC are not capable of creating these lists in an easier way.

[redacted] explained a new CAD program that will be implemented at IHC. This CAD program can create the relationship lists including the welding properties and the weld length from the 3D drawing of the section. This will reduce the time of creating the lists manually. The welding properties have to be assigned to each weld line, which for most lines can be reduced to standard weld lines. Other than that the program creates a list of the connection between parts. [16]

The properties that are taken into account in the relation lists from NX are:

- **Type:** the type stands for the type of weld technique.
- **ID:** the ID that is in this column is the ID of the relationship line.
- **Weld Code:** this is the same weld code as earlier explained for the graph labels, the two characters and two digits that show which type of weld has been used and which shape the end parts of the plates have.
- **Connected Parts:** indicates which parts have been connected.
- **Joint Type:** the joint type shows what the shape of the connection between the parts is.
- **Length:** is the weld length between the parts.
- **Volume:** this is the total volume of the connection between the parts.
- **Throat Height:** the throat height shows the height of the weld that is laid. This is separated for both sides of the connection.
- **Robot Weld:** shows if a robot weld has been used or not on one side of the connection. This is separated for both sides of the connection.
- **Further Columns:** the connected parts are labelled after the last column by their part ID, the section ID, panel number and subsection number.

Compared to the manually created relationship lists, most of the labels match the properties from NX. Table 10.1 shows where the NX properties differ with the manually made properties in the relation lists.

Table 10.1: Manual Relation List Properties against NX Relation List Properties

Extra Manual	Manual Created Relation List	NX Relation List	Extra NX
Welding Coordination Edge penetration type	Connected part ID	Welded Part ID Section	Weldline ID Volume
Water Tight	Connected part Type	Connected parts	Robot Weld Opposite Side
Edge angle	Weld Length	Length	Throat height Opposite Side
Welding Factor wthickness	Position and Thickness difference of the seam Taper Welding method	Weld code	
	Welding seam shape	Joint Type	
	throat thickness	Throat height [mm]	
	Welding procedure	Weld Type Robot weld	
	Analysis list	SUB PAN Position ID	

Most of the properties are used in both of the lists under the same or a different label name. Both methods do have some extra properties. The NX relation list gives the weld list an ID tag, which could be useful to consider for the manual list in order to find weld lines on ID tag in the graph database instead of looking for both the connected part IDs. Also the volume of the weld line is a label that is useful to take into account because this shows the total amount of welding material needed to weld the structure together. The volume was considered for the manual list, but it was a time consuming calculation to perform for every weld line. With a CAD program the information needed is available so the calculation of the volume is performed by the program. The NX list also takes the other side of the weld line into account for the throat height and if it is performed by a robot weld. For the manual list the properties that are of importance and not taken into account in the NX list are the welding coordination, the edge penetration type and if the weld is water tight or not. These properties tell something about the weld line and the weld techniques that are used. The other properties, the edge angle, welding factor and weld thickness, do come back in some of the other properties or do not say much about the weld techniques used. Weld angle can be found in the welding coordination, the welding factor does not say much about the weld line itself but it is more of a safety factor for the welding procedure and the weld thickness is found in the plate thickness and in the NX list in the volume.

If in the end NX is used as a program that creates the relation list, it is recommended to take at least the welding coordination, edge penetration type and whether the weld is watertight or not into account in the program. If a different method is found to create the welding lists, it is recommended to take all the properties discussed into account from both the NX and the manual lists.

10.2. Royal IHC Developments

A new development at Royal IHC is that the company is going to outsource the section building. This has influences for the need of a new cost estimation method and for the use of this cost estimation method.

Cost Estimation

The need of a new cost estimation will change if sections are not built by the company an accurate cost estimation is not needed for own use but as a validation of the hours proposed by the outsourcing company. An accurate cost estimation can prevent the situation that an outsourcing company proposes a budget that is far off from the yard's own cost estimation. Important is to note that the section building process is based on the yards of IHC, so changes in hours might occur based on changes in the building techniques.

When sections are built at own yards the feedback of the database is done with accurate data and more detailed data than an outsourcing company will give. At the own yard the hours are split in welding and ironwork hours and the total hours are known. Other companies will not be likely to give an overview of what kind of hours they made on one section. More likely is that they give the total hours they spent per section or the overall production hours. In order to keep the database up to date and to make accurate cost estimations it is important to keep some of the knowledge for their own. Having some feedback of sections that are built at the own yard will keep the database up to date which is beneficial for accurate cost estimations.

Complete Method

Although most of the sections are not being built at IHC the complete ship and the sections will be designed and engineered at IHC. The design proposals will still be of benefit, when the sections are built at other companies the welding time does still decrease when a new design can decrease the parts and the weld length. To be able to implement and test new design ideas it is recommended to built some sections at the own yard. This is the case at the moment. Most of the sections will be outsourced but a few sections per order will be built at IHC. These sections can be used to test new design principles that can then be implemented in the database.

For the other parts of the shipbuilding process that will be implemented in the database apply the same principles as for the section building. When the parts are performed by IHC the database can be used as intended and new designs can be tested and implemented in the database. When that part of the system will be outsourced the database can be used for the design of that part of the ship building process and as a validation of the budget given by the outsourcing company. It is however important, to keep knowledge and the option for optimization of the shipbuilding process at IHC, still perform some parts of the building process at the own yards to have the feedback to the database and keep the estimation up to date.

10.2.1. Global Developments

Global developments in the total industry can have influences on the shipbuilding process at Royal IHC. These developments can be tested by the database if the development will be performed at the own yards. If for instance other demands for welding are required by the class societies, these can be tested on a new section and analysed with the older sections in the database.

10.3. Validity of the Research

A source of the gathered information were employees of Royal IHC. They gave information about the current cost estimations performed at IHC, the building process at IHC and their future developments. Also G. Alblas gave information about his cost estimation process. In terms of results, chapter 7 uses a qualitative analysis to receive results. When using qualitative data sources bias may occur. When using interviews both respondent's and researcher's bias may occur. Respondent bias means that the respondent is giving answers that are repetitive, socially accepted or in agreement with what the interviewer wants to hear. Another difficulty that can occur is that employees only tell the positive points of their organisation. Because the interviews during this research were meant to gather information about the programs and the production process used at Royal IHC this type of bias will most likely not occur. The programs were explained like an user manual and the building process like a process description. These interviews were not meant to seek for the problems with IHC's working method. Therefore also researcher bias during these interviews did not occur. There was no interest for the researcher to be gained by any hypothesis confirmed.

The data used in the cost drivers analysis and the example tests is data from previous sections built at IHC. This data is the direct section data, which is the same IHC uses for their production process. The relationships lists have made with information available at IHC. The drawings, the analysis lists and the welding codes where used. Some errors may occur in these lists due to the facts that they were made manually. However these lists are not directly used in this research. They are used for the cost driver analysis, in which the results showed that only the total weld length and the welder position were used. The other properties turned out to have too limited information from seven section to create a good correlation or to use in further research.

A limitation of this research is that there are only seven ship sections used for the detailed information. These seven sections showed some accurate results. But when 60 sections were used in the cost driver analyses they showed more significant results. Looking at more properties and the type of work that is performed in the section building process results in more accurate cost estimations from this research. To verify the quantitative results from this research a new cost driver analysis has to be performed if more section data is available.

10.4. Restrictions and Assumptions

There are some restrictions and assumptions made that are discussed during the report. This section sums up some of these restrictions and assumptions.

- The research focusses solely on the section building of buoyancy compartments for trailing suction hopper dredgers built at royal IHC's yard in Kinderdijk during the BVC phase. For the implementation of the method of G. Alblas it is necessary to perform research on influences of other section and vessel types on both of the yards and for the VC phase.
- This research uses Neo4j as the database type for the graph database because it is an open source program in which different input files can be used. No other database programs have been tested in this research. This might be a consideration for further research.
- The current relationship lines in the graph database are only labelled as weld lines. They can be filtered in the database by the different labels such as filtering on watertight weld lines or which type of weld technique is used. Another option is to split the relationship lines by type of weld line so that filtering is not necessary. This can be done by using different labels such as watertight welds and non-watertight welds or by the different weld technique used. This is already done for the part types, it creates more overview in the different elements in the database.

10.5. Recommendations

- When NX is available and more section data is generated a verification of the results from this research should be performed. Also to see if the NX program solves the problem that the section data manually has to be transformed into a relationship list.
- The hours are currently administrated on the total section by ironwork hours and welding hours. It would be preferable to change the hour administration to the different actions performed or to write hours per sub-panel. This increases the knowledge of what the cost drivers are during the production process because the size of the sample is decreased. Furthermore a good hour administration is important for further research.

10.5.1. Recommendations for the Method of G. Alblas

Some other recommendations concerning the further research of G. Alblas are suggested below.

- An important part of the method verification is to find out if the method gives a more accurate cost estimation for the overall production costs of the vessel. For the BVC phase the results of the example cost estimations were more accurate, but these cost estimations were performed with all the detailed property information. If the method can combine both the VC and the BVC estimations into one estimation method it is important to find out if the method can solve this problem.
- The biggest problem by sourcing out the section building part of the production is the feedback for the database. For up-to-date information in the database and for the most accurate cost estimations it is preferable that some sections are still being built at Royal IHC or that there is open feedback from the outsourcing company about the section building process. The latter will most likely not be done so it is a good idea to perform some sections at IHC to keep knowledge about section building inside IHC and also to be able to perform optimizations on design and cost areas.

10.6. Discussion

The biggest discussion point rising from this research is when a cost estimation is accurate. Is this when a certain error margin is reached or when only noise exists from unforeseen circumstances. Noise can occur from weather influences or working with people. It might be hard to find out when the error percentage between the estimated costs and the final costs is due to these unforeseen circumstances or due to the production method used. During this research the error percentages are compared to the BVC estimations to see if the method results in more accurate cost estimations because these BVC estimations were available. For new to be built sections this approach cannot be performed, because this method is meant to replace the current BVC methods of IHC.

Another problem that arises with these cost estimations is that performing an estimation for a big object such as a section, contains more noise than performing an estimation for smaller objects such as subsections. The cost estimation might therefore be more accurate when it is chosen to do hour administrations per sub-panel. This research cannot conclude what the most accurate cost estimation error for ship sections should be or if it is better to write hours on smaller parts of the section building process. These are still questions to be answered by further research.

Furthermore new developments at shipyards might change the shipbuilding process. The robot hall at Royal IHC, discussed with [redacted], has influences on the building process and therefore the hours spent on the building process. These are developments that can be investigated when IHC chooses to build some sections at their own yard.

Bibliography

[1] Neo4j - the number one platform for connected data.

[2] ABS. Classification services, 2017. URL http://ww2.eagle.org/en/what-we-offer/classification-services.html.

[3] S. Al-Baz. Cost model hull dredgers. Technical report, Royal IHC, November 2016.

[4] Ryan Boyd. Graph compute with neo4j: Built-in algorithms, spark & extensions. *Neo4j*, March 2016.

[5] Transport Canada. *Marine transportation: Chapter 4. Construction*. Government of Canada, January 2010.

[6] Ben Delatte and Alley Butler. An object-oriented model for conceptual ship design supporting case-based design. *Marine Technology*, July 2003.

[7] D. J. Eyres and G. J. Bruce. *Ship Construction*. Butterworth-Heinemann, Oxford, seventh edition, 2012.

[8] A. Field. *Discovering Statistics Using Spss*. SAGE, London, 3th edition, 2009.

[9] Jim Frost.

[10] Jan L. Harrington. *SQL Clearly Explained*. Elsevier, United States of America, third edition, April 2010.

[11] Royal IHC. About ihc: Mission and vision, 2017. URL https://www.royalihc.com/en/about-us/about-royal-ihc/mission-and-vision.

[12] IMO. About imo, 2017. URL http://www.imo.org/en/About/Pages/Default.aspx.

[13] Iver H. Iversen and Kennon A. Lettal. *Experimental Analysis of Behavior, Part 2*. Elsevier, United States of America, first edition, November 1991.

[14] Willem Janse. Presentation: Welding construction steels, March 2017. URL http://compass/uploads/media/Lassen_Young_IHC_20170327.pdf.

[15] J. Kaarsemaker, M. Fiksinski, and M. Rydell. *Offshore & Marine: Handboek Sectiebouw*. Royal IHC, April 2011.

[16] Gerrit Koelewijn.

[17] Sweta Kumari and Shashank Pushkar. Performance analysis of the software cost estimation methods: A review. *International Journal of Advanced Research in Computer Science and Software Engineering*, July 2013.

[18] Marcel LaRoche. *The Role of Classification Societies*. Western Economic Diversification Canada, 2012.

[19] Nisith R. Mandal. *Ship Construction and Welding*. Springer, Singapore, ebook edition, 2017.

[20] H. Maylor. *Project Management*. Financial Times, Prentice Hall - M. U. A., e-book edition, 2011.

[21] *Metalix: Company profile*. Metalix, 2016.

[22] Peter Neubauer. Graph databases: Nosql and neo4j. *InfoQ*, May 2010.

[23] Apostolos Papanikolaou. *Ship Design: Methodologies of Preliminary Design*. Springer, UDordrecht, first edition, 2014.

[24] Michael M. Richter and Rosina O. Weber. *Case-Based Reasoning*. Springer, Berlin Heidelberg, e-book edition, June 2013.

[25] Ian Robinson, Jim Weber, and Emil Eifrem. *Graph Databases: new opportunities for connected data*. O'reilly, United States of America, second edition, June 2015.

[26] Marko A. Rodriguez and Peter Neubauer. The graph transversal pattern. *chapter in Graph Data Management: Techniques and Applications*, April 2010.

[27] C. D. Rose. *Automatic Production Planning for the Construction of Complex Ships*. PhD thesis, Technical University Delft, 2017.

[28] *Standars Classification Royal IHC Shipbuilding*. Royal IHC, July 2016.

[29] *Welding codes*. Royal IHC, February 2017.

[30] *Welding list*. Royal IHC and Vuyk Engeneering Rotterdam, July 2014.

[31] Yoonho seo, Dongmok Sheen, and Taioun Kim. Block assembly planning in shipbuilding using case-based reasoning. *Expert Systems with Applications*, January 2007.

[32] Martin Stopford. *Maritime Economics*. Routledge, New York, third edition, 2009.

[33] D. W. Suijkerbuijk. *Hoe begroot ik een sectie?* Royal IHC, January 2015.

[34] P. van der Horst. Kostenreductie door productievriendelijk construeren, 2014.

[35] April Weber. Ship and boat construction and repair. *ILO*, March 2011.

[36] Krzysztof Zima. The case-based reasoning model of cost estimation at the preliminary stage of a construction project. *Procedia Engineering*, 2015.

Appendix

A

Conference Paper

CASE BASED REASONING

AS A COST ESTIMATION METHOD FOR THE SHIPBUILDING INDUSTRY

Sanne Visseren – TU Delft
Dr. Ir. Jenny Coenen – TU Delft
Ir. Ing. Gerrit Alblas – TU Delft
Wouter Zevenbergen MScBA– Royal IHC

ABSTRACT

Accurate cost estimations are important for ship production companies such as Royal IHC. The current cost estimation method at IHC is a top-down price-to-win analogy cost estimation. This method uses mainly the weight of the ship or ship sections as an input variable for the estimation. A problem with using only weight is that the actual work performed on a section is left out of the estimation as well as other cost drivers. Another problem occurring is the estimation being divided into three phases, with lacking evaluation, optimization and feedback possible. This research discusses two new cost estimation methods. The statistical method is a parametric cost estimation method, that uses equations with the main cost drivers. The graph database method uses the CBR theory as a basis for cost estimations. Problem solving is done by using a solution from an old ship and reusing this solution for a new ship. The graph database stores detailed information about the work performed on the ship and the main cost drivers. The qualitative and quantitative analyses showed that using more cost drivers and the work performed on a section creates more accurate cost estimations. With a solution as the graph database method also the possibilities of evaluation and process optimizations are broader.

Keywords: Cost Estimation, Case Based Reasoning, Shipbuilding, Ship Production, Section Building, Marine Technology

I. INTRODUCTION

"Royal IHC designs, builds and maintains innovative equipment and vessels, working from sea level to ocean floor, for maritime service providers in the offshore, dredging and mining industries. We sell vessels, equipment and services either separately, or in combination. In addition, we are able to supply advanced technology to the international rental market." - Royal IHC [1]

An accurate ship production cost estimation is important for a ship production company such as Royal IHC. If an accurate cost estimation can be made at the beginning of a production project these costs can be monitored during the project. This research will compare three different ship building cost estimation methods.

Cost estimation methods can be categorised by algorithmic methods and non-algorithmic methods. Algorithmic cost estimation methods use equations based on historical data and the main cost drivers to calculate the costs. A parametric cost estimation is an example of an algorithmic cost estimation type. Types of non-algorithmic cost estimations are: [2]

- Expert judgement
- Analogy
- Bottom-up or Top-down approach
- Parkinson's law
- Price-to-win

The current method of Royal IHC uses a top-down price-to-win type or a top-down analogy type of cost estimation. In some cases also experts judgement is used for cost estimations. The current method is divided into three methods. With as main input variable the weight of the ship or ship sections. Problems occurring with the current cost estimation method is that when only looking at the weight as an input variable the amount of work performed is left out of the cost estimation as well as other cost drivers. Presumed is that using more cost drivers and the amount of work will lead to more accurate cost estimations. Another problem with the current method is that the evaluation, process optimization and feedback possibilities are limited.

Two new cost estimation methods will be introduced during this research, the statistical method and the graph database method. The goals for a new cost estimation method are based on the problems occurring with the current estimation method:

1. Creating a more accurate cost estimation

2. Creating more integration between the cost estimation methods

3. Creating more evaluation possibilities

The statistical method is a parametric cost estimation method. The graph database method is an analogy, more specific a case based reasoning, cost estimation method. The goal of this research is to perform a feasibility study for the graph database method. The current method is used to see what the problem is with the current cost estimations and the statistical method is used to see if it is necessary to include every cost driver and the work performed in a cost estimation. A comparison between the three cost estimation methods will result in the answer to the main research question of this research:

"Is Case Based Reasoning a feasible cost estimation method for the ship building industry?"

II. CASE BASED REASONING

The graph database method will be designed on the case based reasoning problem solving theory. The theory can be explained by three aspects:

1. Case: a case is a previous task that is already solved.
2. Based: based means that the connection between reasoning is made on cases. Meaning that the reasoning for a new solution is based on old cases.
3. Reasoning: reasoning stands for the way a solution is found. In CBR this reasoning is based on previous cases.

The CBR theory bases the solution of a new case on the solution of an proven solution from an old similar case retrieved from a historical database. The CBR process can be found in figure 1. The search for similar solutions is performed with a nearest neighbour search algorithm. [3]

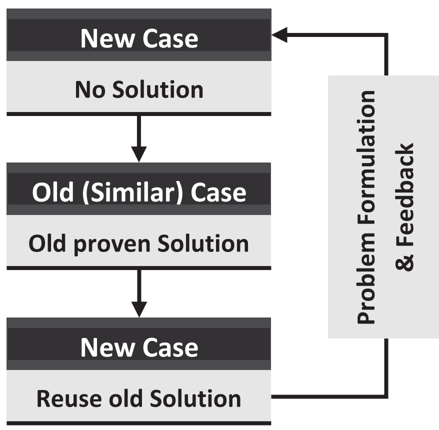


Figure 1. CBR process

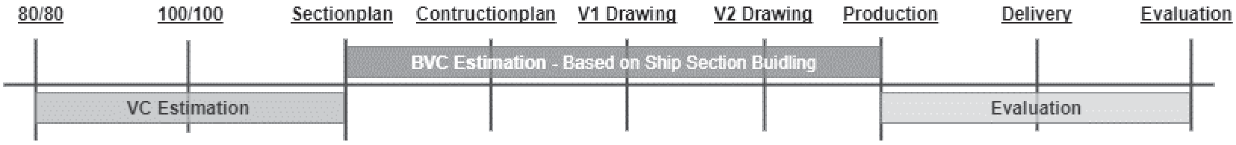


Figure 2.The three phases of the current cost estimation method at IHC

III. CURRENT METHOD

The current cost estimation method at Royal IHC is divided into three cost estimation methods, the VC phase, the BVC phase and the hour evaluation phase. Figure 2 shows a timeline of a ship production process with the different cost estimation phases.

The overall ship production costs are estimated at the beginning of the VC calculation before the contract with the client is signed. This is the 100/100 phase in figure 2. After the cost price of the vessel is approved the VC phase estimates the budget for each department involved in the ship production project. The input variable during the VC phase is the weight of the vessel. The weight of the new vessel is compared with previous ships to find a new cost price.

The second phase of the cost estimation is the BVC phase. The BVC cost estimation uses the allocated VC budget to estimate a more detailed cost estimation per department. The research is focussed on the BVC phase of the section building part of the ship production process. The BVC cost estimation uses the section weight as an input variable together with a section type factor. The weight of the section will be compared to other section and multiplied by the section factor.

When the production of the ship section starts the evaluation phase monitors the production hours. The hours worked on a section are compared with the estimated BVC hours. At the end of the project an hour estimation is made in terms of an EV analysis.

A few problems occur with the current cost estimation method. The first problem is that the current method uses only weight as input variable for the cost estimation. When only weight is used the actual production work performed on a section and other cost drivers are left out of the cost estimation. Presumed is that using multiple cost drivers and the work performed on a section will result in more accurate cost estimations.

A second problem that occurs with the current cost estimation is that there is limited feedback between the different methods. It is not possible to perform process or cost optimizations with the current method. Presumed is that having a more integrated cost estimation method over the three phases it is possible to create optimizations for the cost estimation and production of the ship building project. Another possibility of being able to optimize the production process is to have a more detailed hour and section evaluation. With the current hour evaluation is possible between the estimated hours and the final production hours, but this hour evaluation does not lead to insights in the production process or what caused deviations between the hours. Presumed is that using more cost drivers will increase the insight in the deviations between the hours.

IV. STATISTICAL METHOD

The statistical method is a parametric cost estimation method using equations composed by a statistical analysis of historical section data. With a linear regression analysis equations are composed that can be used to estimate the costs. A cost driver analysis must be performed in order to use the highest correlated ship section variables.

The cost driver analysis is performed by a correlation matrix. This analysis shows which variables have the highest correlation towards the final hours of the section. The resulting ranking of the cost drivers:

1. Profiles
2. Parts
3. Brackets
4. Weld Length
5. Frames
6. Collar Plates
7. Weight

The weight turns out to have the lowest correlation to the final hours. This means that the current method uses a cost driver that has a low correlation to the final hours. Part and profiles have a high correlation towards the hours. P. van der Horst [4] performed a research on part reduction and the influences for the production hours. This research led to the conclusion that a reduction in parts reduces the production hours. A reduction in parts does however not necessarily mean a reduction in weight.

With these cost drivers the linear regression analysis is performed. This analysis results in equations where the costs can be estimated with. The general equation with only one variable is:

$$\boldsymbol{Hour} = \boldsymbol{\beta} + \boldsymbol{x} * \boldsymbol{\alpha} \quad \textbf{(1)}$$

Where β is the standard value of the equation, x is the section variable and α is the standard variable factor.

Multiple variables can be added to the equations:

$$\boldsymbol{Hour} = \boldsymbol{\beta} + \boldsymbol{x_1} * \boldsymbol{\alpha_1} + \boldsymbol{x_2} * \boldsymbol{\alpha_2} + \boldsymbol{x_3} * \boldsymbol{\alpha_3} \textbf{(2)}$$

During the BVC phase the section data becomes more detailed. In the beginning less accurate cost driver data is known then at the end of the BVC phase. Therefore three different set of equations are composed. An early BVC phase set, a final BVC phase set and a set of equations that can estimate the costs for the welding and iron work hours separate. Each set will be composed of an equation for six sections, an equation for seven sections and one for 60 sections. For 60 sections the weld length is not known, but the other cost drivers are and a larger dataset can be used. These equations will be tested during the quantitative analysis.

V. GRAPH DATABASE METHOD

The graph database method uses the CBR theory to estimate the costs. A graph is a structure that stores data from objects. In this case the object is a ship section. A graph is composed of nodes and relations between these nodes. [5] A node in a ship section graph represents a section part, for instance a plate or a profile, or a node can represent the total section. The relations between these nodes are the weld lines between the section parts. Figure 3 shows an example of a section graph.

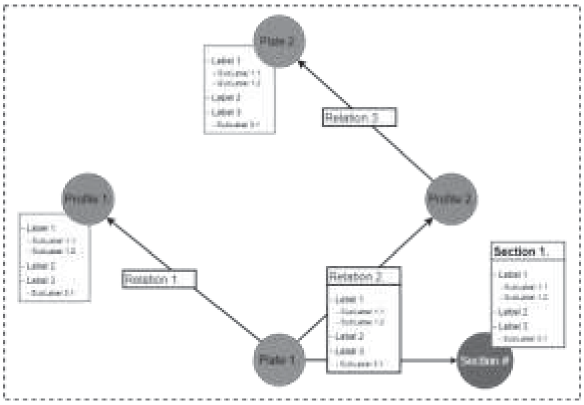


Figure 3. Example of a graph

The nodes and relations in a graph contain labels. [5] The section labels represent the total values of the ship section, for instance the amount of parts or total weld length. Part node labels contain information specified for the parts, for instance the part dimensions or the building order of the part. Relationship labels contain information about the work performed on the section, for instance the welding technique or if the weld is watertight or not.

The graph database method is performed in a loop this the final section data is known. Figure 4 shows the cost estimation process. First a database with old sections is created at number 1. Then the data from a new section is gathered and transformed into a graph, numbers 2 and 3. This data does not have to be the final section data. At number 4 the new section is run through the database to find a similar section, the similarity search. The cost estimation is based on the solution of the most similar section and adapted at number 5 to the cost estimation for the new section, the similarity assessment. When the final section data is used the section graph is places into the database, number 6 and 7. When the data can still change the loop returns to number 2.

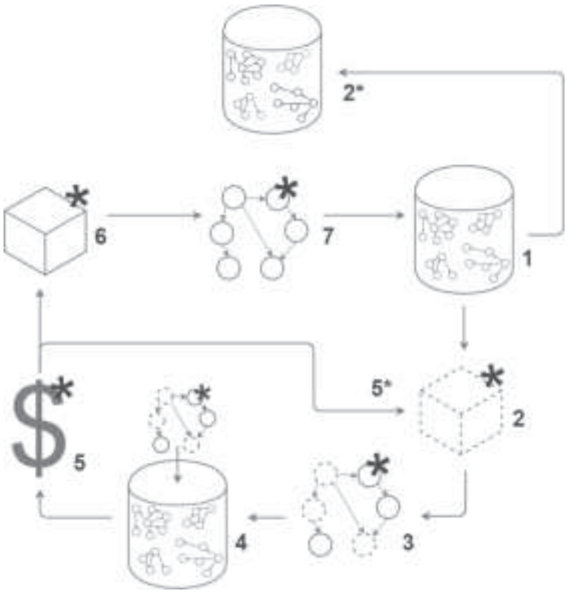


Figure 4. Graph Database Process

Another function of the database is the evaluation possibilities, number 2*. It is possible to find other solution in the database. For instance when a new material is introduced solutions from previous ship sections can be evaluated.

VI. QUALITATIVE ANALYSIS

The similarity search and similarity assessment algorithms are not yet finished so in order to compare the three cost estimation methods a qualitative and a simplified quantitative analysis are performed. The qualitative analysis uses example structures from ship sections and compares the cost estimation methods with each other.

The first example structure is a plate with five profiles or two plates with five profiles. Figure 5 shows a drawing of the structure and the graph representation. The cost estimation differences for this example will be discussed for the three methods.

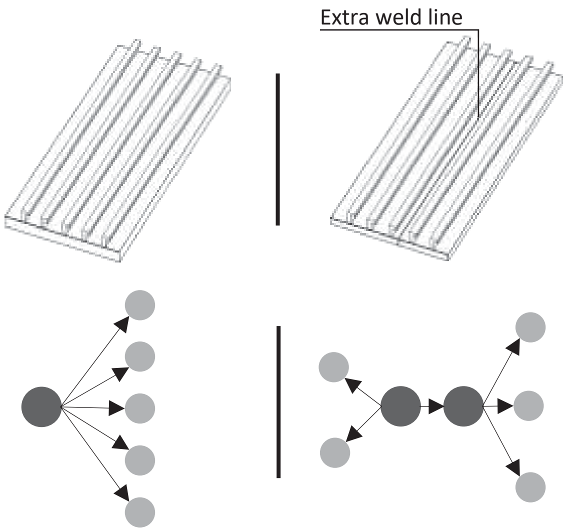


Figure 5. Example structure 1

Royal IHC current methods: The current method will not show a different cost estimation because the weight of the structure stays the same. This shows the limit of the current method used at Royal IHC.

Statistic Method: The statistic method will show a different cost price if the cost drivers that change are variables in the equations for the costs. The weld length and the amount of parts are different for the set-ups so this will lead to different variable values in the equations. This shows the advantage of using more cost drivers then just weight.

Graph Database: The graph database can store the same data as used in the other methods but it can also use the amount of work that is performed on the. This will result in a more accurate cost estimation then the other two methods.

The second example, shown in figure 6, shows an example of the research of P. van der Horst [4]. He designed a part reduction method by changing a structure of two brackets and a profile by one plate. The graph representation for the first set up shows that there are many different weld lines and parts. For the second set up the number of parts and weld lines are decreased.

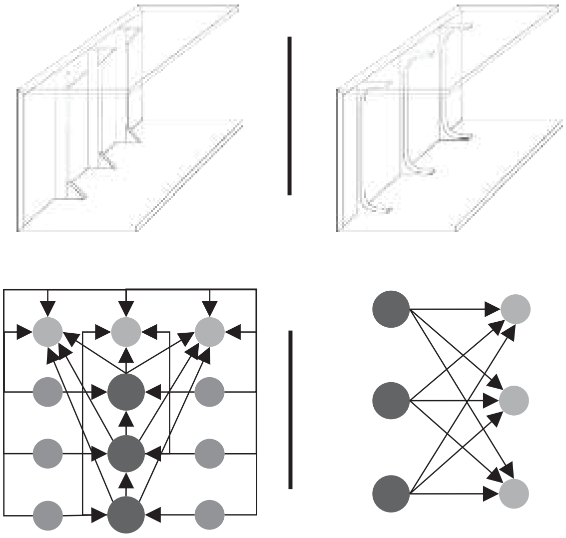


Figure 4. Example structure 2

Royal IHC current methods: The current IHC method will not result in a different cost estimation because the weight of the structure stays equal over the two set ups. The weight might even increase slightly, according to the conclusion from P. van der Horst leading to a higher cost estimation. [4]

Statistic Method: Using this method in this example will lead to a different cost estimation, because the amount of parts, profiles and brackets decrease in the second set-up. When these variables are used for the equations this method will result in a different cost estimation.

Graph Database: The graph database will not only take the reduced amount of parts into account, but also the difference in the amount of work. This example does also show that this method can also be used for design proposals. The database can recognize the pattern of set up one and suggest to change it with a structure similar to set up two.

Conclusion:

The qualitative analysis results in the conclusion that using only weight for the cost estimation might not result in the most accurate cost estimations. The example structures show that when using weight as the only cost driver many work or changes in the section are not taken into account. Also the statistical analysis might not give the most accurate results because this method only works when the cost drivers that change are in the cost estimation equations. The graph database method is tends to be the most accurate cost estimation method. It can take the work performed on the section into account and multiple cost drivers. Furthermore the method can be used to create design optimizations.

VII. QUANTITATIVE ANALYSIS

The three methods are also compared in terms of quantitative data. An example section is used to perform cost estimations with the three methods. The results of the quantitative analysis are shown in table 1.

The BVC cost estimation results are known for the example section. For the statistical method the cost estimation is performed by filling in the equations with the section data from a new section.

The quantitative analysis of the graph database is performed in a simplified way. A similarity search is performed based on the ranking from the cost driver analysis, the seven sections and the data from the new section. The similarity assessment can be performed in multiple ways. For this research two ways are shown: using the difference between label values and by using the total difference ratio between sections. The similarity search is performed by looking at the difference between labels, these differences can be multiplied with the label difference factor to adapt the solution of the similar section to a solution estimation for the new section. This can only be performed for the weld length in this research because for the other cost drivers the difference factors are unknown. The other method uses the similarity factor as a ranking for the most similar sections and uses the solutions with the ranking to estimate the costs for the section. This leads to the most accurate results if two or three similar sections are used.

Table 1. Summary of the Quantitative Analysis

	BVC	Statistical	Graph Database	
			Weld Length	Similarity Ratio
Real	18%	14%	9%	5%
2% Up		12%	8%	5%
2% Down		16%	11%	9%
10% Up		4%	8%	1%
10% Down		24%	18%	16%
15% Up		3%	7%	6%
15% Down		29%	18%	19%
Average	18%	14.60%	11.51%	8.88%

The results in table 1 show that the two new cost estimations result in a more accurate cost estimation. The percentages represent the absolute difference between the final production hours and the estimated hours. The data of the section is changed with 2% till 15% to match the BVC phase. From the start of the BVC phase till the production of the section the data can still change. The average values of these estimation are used as comparison values.

Conclusion:

The statistical method shows a slightly more accurate cost estimation compared the current BVC estimation, from 18% to 14.6%. The graph database however results in the most accurate results. When the difference between the labels are used the cost estimation is reduced from 18% to 11.5% when only the weld length is used. The total similarity ration shows even more accurate results, a reduction from 18% to 8.9%. Presumed is the when more label factors are known the label difference method of similarity assessment can lead to even more accurate results. This method is also the most stable when the data might still change during the BVC phase.

VIII. CONCLUSION

The research leads to the conclusion that the graph database method is a feasible cost estimation method. In terms of the goals set at the beginning of the research the graph database method performs the best.

1. *Creating a more accurate cost estimation:*

   Both the statistical and the graph database method result in more accurate results compared to the BVC estimation. However the graph database results in the most accurate results.

2. *Creating more integration between the cost estimation methods:*

   The graph database is intended to use one method for the complete cost estimation phase. Where the statistical method and the current cost estimation method need more than one program or equation to estimate the costs. The graph database uses only the graph database for the cost estimation.

3. *Creating more evaluation possibilities:*

   The current method only had the option to perform hour evaluations using the final production hours and the estimated hours. The only factor that could be used as insight into the difference between the hours is the weight. The graph database method uses multiple cost drivers and the work performed on a section. So more insight in the factors that led to a difference in hours is available with this method.

So to conclude the research:

"Is Case Based Reasoning a feasible cost estimation method for the ship building industry?"

The answer is yes. However not only the case based reasoning theory is what makes the graph database method so promising. The use of multiple cost drivers and that the work performed on a ship section can be taken into account leads to more accurate cost estimations and broader evaluation possibilities.

IX. ABBREVIATIONS

BVC	-	Bedrijfs Voor Calculatie
CBR	-	Case Based Reasoning
IHC	-	Industriële Handels Combinatie
VC	-	Voor Calculatie

X. REFERENCES

[1] Royal IHC, Royal IHC, 2017. [Online]. Available: https://www.royalihc.com/en/about-us/about-royal-ihc/mission-and-vision.

[2] S. P. Sweta Kumari, "Performance Analysis of the Software Cost Estimation Methods: A Review," *International Journal of Advanced Research in Computer Science and Software Engineering,* 2013.

[3] M. M. Richter and R. O. Weber, Case-Based Reasoning, Berlin Heidelberg: Springer, 2013.

[4] P. v. d. Horst, "Kostenreductie door productievriendelijk construeren," Rotterdam Mainport University, 2014, 2014.

[5] I. Robinson, J. Weber and E. Eifrem, Graph Databases: new opportunities for connected data, United States of America: O'reilly, 2015.