FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/ Specialization: Master of science/ Industrial Asset Management	The spring semester, 2023 Open
Author: Atefeh Angiz	
Supervisor at UiS: Reidar Brumer Bratvold	
Thesis title: Value loss as a result of poor project planning decisions	
Credits (ECTS): 30	
Keywords: Value loss Petroleum development projects Poor forecasting Petroleum Production Forecasts Development overruns	Pages: 128 Stavanger, 14.06.2023

Abstract

The forecasts of development costs, schedules, and future productions, which are utilized to support investment decisions for projects, have a substantial impact on valuation and decision-making. However, many large projects, including petroleum projects on the Norwegian Continental Shelf (NCS), encounter cost and schedule overruns, as well as production shortfalls. Flyvbjerg [1] argued that exceeding budgeted time and costs is so prevalent in megaprojects that it can be considered a rule.

Although many studies worked on the cost and schedule overruns in various industries, the subject of production underperformance has not been widely covered. This study builds upon the research conducted by Bratvold [2] and Nesvold [3], [4] regarding production forecasts for projects on the NCS. It involves the analysis of production shortfalls along with development time and cost overruns to evaluate the value loss arising from inaccurate forecasts.

The actual cost, time and production values are publicly available on the Norwegian Petroleum Directorate's (NPD) website [5]. Only the forecasted development costs and time schedule were publicly available, while the projected production values remained confidential.

Using the forecasted and actual values of the aforementioned data[1], this study aimed to identify overruns or underperformances, assess the economic value erosion caused by them, and examine the correlation between cumulative underperformances in the first

[1] All the data utilized in this thesis pertains to projects on the NCS, unless specifically mentioned otherwise.

10 years of production and total cost overruns. For this purpose, the Present Value (PV) calculations were performed using two methods suggested by Mohus [6], in addition to Pearson, Spearman, and Support Vector Regression (SVR) analyses, utilizing MS Excel and Python.

After summarizing previous research on overruns and underperformances in various industries, an overview of the development project phases on the NCS and the study's methodology is provided. Based on the analyses conducted in this study, the results can be summarized as follows:

1) Utilizing development cost data from 142 projects with (Plan for Development and Operation) PDO approvals between 2001 and 2022, a total loss of NOK 268 billion was attributed to cost overruns. This indicates a 12.7% overrun of actual costs compared to the initial forecasted costs.
2) Analysing development time data for 76 oil fields with PDO approval from 1990 to 2019, the study found an average development delay of 101 days, resulting in a 12% schedule overrun compared to the initially forecasted development times. The forecasted value loss due to these delays amounted to approximately NOK 39 billion.
3) Examining production data from 67 oil fields with production spanning from 1995 to 2021, it was observed that an average of around 50 million Sm^3 of oil was not delivered during the first 10 years of production, as initially forecasted. This accounts for a 5% underproduction. The forecasted value loss due to production shortfalls, considering only the mean forecasts, amounted to approximately NOK 80 billion.

4) The total value loss resulting from poor forecasts, when considering the first 10 years of production, was calculated to be NOK 387 billion, which is the cumulative sum of the previously mentioned figures.
5) The average correlation coefficients between the cumulative underproduction in the first 10 years and the total cost overruns, calculated using the Pearson, Spearman, and SVR methods, were 0.28, 0.08, and 0.41, respectively. The Pearson coefficient suggests a positive linear correlation, albeit a weak one. The Spearman coefficient indicates a very weak positive linear correlation between the ranks of the data. The SVR coefficient suggests a moderate correlation, which is non-linear, considering that the linear correlation coefficient is considerably smaller.

Moreover, the reasons for such poor forecasts, attributed to human bias, can be categorized as delusion, deception, and bad luck [7]. It is recommended that forecasters be trained about the biases and uncertainties involved in their forecasts, leverage the expertise of superforecasters, utilize historical data, and employ external methods like Reference Class Forecasting (RCF) to improve their forecasting accuracy [8].

To the best of the author's knowledge, based on data available until 2023, there has been no prior exploration of conducting a value loss analysis on the development time, cost, and production data for the first 10 years of production, along with regression analysis. This unique aspect distinguishes the present study from previous works. Considering the substantial value erosion resulting from inadequate forecasts in petroleum projects, it is vital for both the industry and the public to maintain diligent monitoring of project performance.

Acknowledgements

First and foremost, I would like to express my heartfelt gratitude to my supervisor, Professor Reidar Brumer Bratvold, for his invaluable assistance, active involvement, and guidance throughout the entire research and writing process. From our very first meeting, where he expressed that "we are here to learn from each other," I have been privileged to gain immense knowledge and insight from his mentorship.

Additionally, I would like to extend my deepest appreciation to my parents, my brother, and my husband for their unwavering support, constant presence, and encouragement during my master's studies. Their belief in me has been an endless source of motivation.

I am aware that this accomplishment would not have been possible without the unwavering support of these remarkable individuals. Once again, thank you.

Atefeh Angiz

Table of Contents

List of Figures

List of Tables

Chapter1 – Introduction

In this chapter, an introduction to the topic of the thesis, key contributions and thesis questions is provided. Subsequently, an overview of previous research focused on the issue of poor forecasting in development projects is presented. The remainder of this chapter includes the thesis outline and structure, offering a glimpse into the content of the present work.

1.1 Introduction

In recent years, there has been a significant rise in the development of megaprojects, which has sparked considerable academic interest in evaluating the performance of such projects. One of the key objectives of this study is to demonstrate that inadequate forecasts of costs, schedules, and production values of projects at the time of their PDO [9] approval, which manifest as cost overruns, delays, and underperformance in production, can lead to value loss.

In this section, a review of previous works related to overruns in various types of megaprojects is provided. These include infrastructure, wind farm, public housing, transportation, and mining and metals projects. Additionally, examples of underproductions in the domains of food, agriculture, and dams are surveyed. Furthermore, the existing literature on project overruns and production shortfalls in the petroleum sector is discussed. Finally, the outline and structure of the present study is presented, along with its key contributions in addressing some of the gaps found in the existing literature.

The research questions addressed in this study were as follows:

1) To what extent did the actual values deviate from the forecasted development schedule, cost, and production values individually?

2) What was the magnitude of the individual and total value loss resulting from cost and schedule overruns, as well as underproduction in 2022-NOK?

3) Was there any correlation between the cumulative underproduction during the first ten years and the total cost overruns?

1.2 Previous literature

1.2.1 cost and schedule overruns

Project planners have used inefficient methods for estimating the budgeted cost, demand, and scheduled time, which led to inaccurate results and overruns [10]. As stated by Hall and Hall [11] in the book "Great Planning Disasters" historically, megaprojects have experienced frequent cost overruns that have disturbed political sponsors and confused project managers.

Megaprojects can be categorized into four major types: production, consumption, infrastructure, and extraction projects [12]. From an investment viewpoint, they are defined as projects with a budget exceeding $1 billion USD [13]. However, in certain contexts, projects with a budget of $100 million USD can also be considered as megaprojects [14]. Additionally, from an operational standpoint, megaprojects are believed to have long-term and widespread effects on the environment [15].

Following that, Nijkamp and Ubbels analyzed infrastructure projects in the Netherlands and Finland to find the root causes of cost misjudgments. They found that most cost underestimations have occurred due to delays in project delivery [16]. Due to numerous cases of project overruns in megaprojects, cost and schedule overruns in this type of project have been the subject of numerous public and private research studies.

The study conducted by Koch [17] on windfarms found that some projects experienced no cost overruns, while others exceeded their initial budget by up to 65%. Regarding the scheduled development time, the study reported that some projects were completed within 9% of their planned schedule, while others experienced delays of up to twice the planned time. The study found that cost overruns are common in offshore wind projects due to factors like technical challenges, inadequate planning, and regulatory issues.

As another example, Ansar et al. [18] analyzed data from 245 large dams, including 26 major dams constructed over a span of 74 years. They found that the actual costs of these dam projects increased by an average of 96% (almost double) compared to their initial forecasts. Similarly, the study found that the actual development times were delayed by an average of 44% compared to their original forecasted development times. The findings highlighted the need for more accurate cost estimation and better management practices to mitigate cost overruns in hydropower megaprojects.

This trend has also been observed in other sectors, such as transportation and public social housing. In the case of infrastructure projects, Flyvbjerg et al. [10] conducted an analysis of 258 transportation infrastructure projects encompassing different types, regions, and time periods. They discovered that the cost overrun percentages for rail, bridges and tunnels, and roadways projects were 44.7%, 33.8%, and 20.4% respectively. Furthermore, they suggested that the cost forecasts utilized for making decisions regarding project developments were misleading and untrustworthy.

A recent study conducted by Nilsson [19] examined seven large road and railway projects in Sweden, as presented in Table 1, in order to identify the root causes of significant cost overruns in such projects. As can be seen from Table1, the study revealed that the initial forecasted cost for the Southern link road project in Stockholm was 4.0 billion Swedish Krona, whereas its actual final cost amounted to approximately 8.4 billion Swedish Krona. This indicates a cost overrun of 110% for this particular project. The research raises concerns about the accuracy and reliability of cost forecasts used in decision-making, which pose a weakness in the process of conducting a Cost-Benefit Analysis (CBA) for infrastructure projects. Some potential causes identified in the research were inadequate initial cost estimation, underestimation of project complexities, changes in project scope, inaccurate risk assessment, and inadequate project management.

Table1 - The seven Swedish infrastructure projects studied by Nilsson [19]

			Costs			Traffic opening	
		Project	First	Final	Cost overrun[a]	Planned	Actual
Stockholm	Rail	The Third Track …	1.5			1998	
		…that became the City Line	7.5	20.1	168	2011	2017
	Road	The Southern Link	4.0	8.4	110	1997	2004
	Road	The Northern Link	2.1	10.4	395	1996	2014
Malmö	Rail	The City Tunnel	4.8	12.7	165	2000	2010
Gothenburg	Road	Götaleden	1.6	3.5	119	1999	2006
Rural	Rail	The Bothnia Line	9.8	25.2	157	2006	2012
		Total	**29.8**	**80.3**	**173**		
Stockholm	Road	The Bypass. v. 1	5.4			2005	
		The Bypass. v. 2[b]	19.0	*37.7*	*98*	2016	*(2030)*

Note: Cost when the formal time to implement the project and final cost, billion Swedish Krona (SEK) at nominal prices. Planned and actual traffic opening.
[a]((Final cost/estimated cost)−1) × 100.
[b]*Estimated.*

Regarding the public sector social housing projects, Chadee et al. [20] conducted a study on public sector social housing projects in Small Island Developing States of the Caribbean Sea between 2005 and 2010. Their research focused on investigating various public housing construction programs in this region. The study revealed that, on average, the cost overruns in these projects amounted to 75%. The study highlighted challenges such as inadequate planning, limited capacity, bureaucratic inefficiencies, and political influences. These factors created uncertainties in project implementation, leading to delays, cost overruns, and reduced quality in social housing construction.

EY conducted research on 192 global mining and metals megaprojects and found that over 64% of these projects experienced cost overruns, schedule overruns, or both. The study suggests that the reduction of capital productivity, defined by the Australian Bureau of Statistics as the ratio of output to capital input, can be attributed to five key risks [21]:

1. Poor schedule and cost estimation methods;

2. Conflicts between Stakeholders due to insufficient representation of socio-economic value of the projects and superficial relationships;
3. Lack of resilient supply chain;
4. Digital and workforce disruption;
5. Unpredictable external environment.

1.2.2 cost and schedule overruns in the petroleum sector

An early study conducted by the Norwegian government examined the cost performance of oil projects in the North Sea. The study revealed that the development costs of these projects increased by 26%, amounting to approximately 25 billion NOK, from their PDO approval until their last Capital Cost Estimate (CCE). The study encompassed 11 oil field projects that were implemented between 1994 and 1998. This analysis sheds light on the significant cost escalation experienced by these projects during their development phase in the specified time period [22].

The NPD reviewed cost developments for five megaprojects approved between 2006 and 2008, namely Skarv, Yme, Valhall Redevelopment (VRD), Tyrihans, and Gjøa. The findings indicated that, on average, these projects experienced cost overruns of approximately 50%. The majority of these overruns were identified in the early stages of the projects. The study concluded that optimistic forecasts were a result of unrealistic objectives set during project planning [23].

In a study by Merrow the success rates of megaprojects in various sectors were investigated. The findings suggested that non-Oil and Gas development projects, despite their increased size and complexity, had a success rate of approximately 50%. On the other hand, for Oil and Gas projects, the success rate dropped significantly to 22%. The study further revealed that the unsuccessful Oil and Gas projects experienced cost overruns of about 33%, schedule overruns of around 30%, and approximately 64% of these projects faced production shortfalls within the first two years [24].

According to another study by Ernst and Young consulting company between 2015 and 2019, which analysed 500 major projects in the Oil and Gas sector, it was found that 38% of these projects experienced cost overruns, while 60% of them were completed with delays. The research highlighted the importance of adaptability and flexibility in organizations to navigate unpredictable futures [25].

Dahl et al. [26] analyzed 80 Oil development projects in Norway, utilizing approval plans and special permits obtained from the Norwegian Ministry of Oil and Energy between 2000 and 2013. The researchers employed regression analysis to examine the relationship between cost overruns and oil price developments, as well as the number of employees involved. The findings of the study indicated a positive correlation between cost overruns and changes in oil prices, suggesting that fluctuations in oil prices can impact project costs. Additionally, the analysis revealed a positive relationship between cost overruns and the number of employees involved in the projects. These insights contribute to understanding the factors influencing cost overruns in Oil development projects.

In the study by Haukaas and Mohus [27], the focus was on cost overruns in development projects on the NCS. The researchers analyzed cost data from 78 oil fields and reservoir data from 66 oil fields approved over a 23-year period. The key findings of their work are as follows:

1. The total cost overruns for the studied period amounted to 231 billion 2015-NOK, representing approximately a 25% overrun compared to the budget.
2. They observed no relationship between forecasted revenue increases and budget overruns.
3. Megaprojects experienced cost overruns approximately twice as much as small development projects.
4. A comparison of projects awarded to Norwegian yards versus Asian yards revealed that projects commissioned by Asian yards had significantly higher costs and longer development times.
5. The researchers highlighted that despite significant attention being given to cost overruns on the NCS, approximately 85% of the oil fields analyzed still experienced cost overruns. They argued that this finding indicates biased forecasts by operators on the NCS and a lack of improvement in this area.

A subsequent study by Mohus [6], provided further insights into the extent of cost overruns, development delays, and production shortfalls in development projects on the NCS. The study highlighted the potential for improving forecast accuracy through the application of appropriate forecasting methods. The key findings were as follows:

1. Analysis of cost data for 68 oil fields with PDO approval between 1995 and 2017 revealed total cost overruns of approximately 213 billion 2017-NOK, equivalent to a 26% budget overrun.

2. Analysis of development time data for 42 fields developed from 1997 to 2013 showed an average development delay of 202 days, corresponding to a 25% schedule overrun.
3. Analysis of production data for 56 fields with PDO approval between 1997 and 2017 involved comparing actual and forecasted production values. The study calculated the present value of production revenues for the first four years of projects and found a value loss of around 61 billion 2017-NOK due to delays and approximately 200 billion 2017-NOK in lost value due to production shortfalls.
4. In the final part of the study, the Reference Class Forecasting (RCF) method was applied to modify the forecasted production values of the projects. By applying multiplicators to the outcomes, the study observed a reduction in delusional and deceptional biases in the forecasts.

In a cost performance analysis of petroleum development projects on the NCS, cost data from annual national budget reports provided by the Norwegian Ministry of Petroleum and Energy was examined [28]. The study focused on 148 development projects between 2001 and 2022, comparing initial cost estimates with actual costs. When unfinished projects were excluded from the calculations, it was found that 11 out of 47 projects had lower actual costs than the initial estimates, but were outside the -20% uncertainty range. On the other hand, for projects with actual costs higher than the initial forecasts, 35 out of 85 projects were outside the +20% uncertainty range. These findings highlight the significant variability in cost estimates and the challenge of accurately forecasting project costs in the petroleum development sector.

In another section of the same study [28], a graph as Figure 1 was presented showing the ratio of actual cost to initial forecasted cost plotted against the cumulative probability for all the fields analyzed. The graph illustrated that there was a significant likelihood, approximately 60%, of projects having actual costs exceeding their initial forecasts. The study further divided the data into different time intervals, namely "2001 to 2006," "2007 to 2012," "2013 to 2018," and "2019 to 2022." Within these intervals, the probabilities of actual costs exceeding the initial forecasts were approximately 55%, 75%, 50%, and 45%, respectively. These findings highlight the varying degrees of cost overruns and the associated uncertainty in cost forecasts across different time periods.

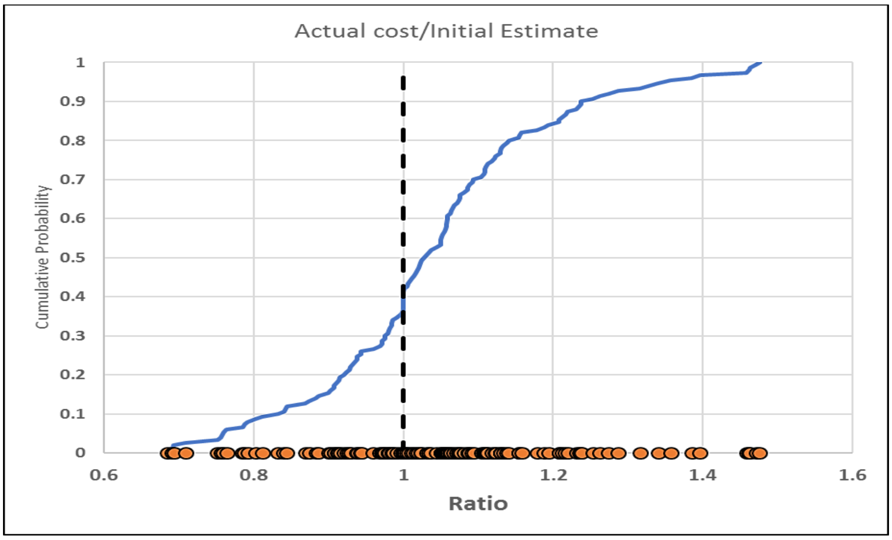

Figure 1 - The CDF plot of (Actual cost/Initial estimat) ratio [28]

According to the analysis conducted in reference [28], it was found that as the duration of projects increased, the deviation of actual costs from initial forecasts also increased. Additionally, when the projects were categorized based on their costs, it was observed that both small projects (less than NOK 3.5 billion) and large projects (more than NOK 7 billion) were more likely to exceed their initial cost estimates. On the other hand, mediumsized projects (between NOK 3.5 and 7 billion) had the potential to have actual costs lower than the initial predictions. These findings suggest that project size and duration play a role in the accuracy of cost forecasts, with larger and longer projects tending to experience more significant deviations from initial forecasts.

1.2.3 production shortfalls

Production shortfall happens when the actual productions fall short to meet the forecasted production values.

In their study, Jayne and Rashid [29], highlighted the impact of inaccurate forecasts in the food industry. They emphasized that inaccurate predictions of crop production and future consumption can lead to negative consequences for food price stability and food security. Through a heuristic example, they illustrated how a mere 13% overestimation of production capacity coupled with an 8% underestimation of consumption can result in a significant 21% shortfall in food supply. This example underscores the importance of precise forecasting in order to mitigate the adverse effects on food availability and ensure the stability of prices and food security.

According to a recent study in the agricultural sector [30], production shortfalls were observed in starchy staples in Ghana. The study focused on the period between 1987 and 2017 and highlighted the production of cassava, plantain, yam, and cocoyam. The findings indicated that these crops consistently fell short of meeting the forecasted production values, with shortfalls of approximately 35% for cassava, 45% for plantain, 46% for yam, and 53% for cocoyam.

A study focusing on electricity generation in the Amazon region of Brazil [31] examined the power capacity of 28 operational dams and found that there was an underproduction of approximately 12GW compared to the projected capacity. To address this issue and meet the electric demands more effectively, the study proposes the utilization of Floating Photovoltaic (FPV) systems. These systems are suggested as a potential solution due to their better environmental impacts and their ability to generate additional electricity to supplement the existing power capacity. Implementing FPV systems could potentially help bridge the gap between projected and actual electricity generation in the Amazon region.

1.2.4 production shortfalls in the petroleum sector

In addition to the common issue of low forecasts of costs and development times in the PDOs for petroleum development projects, another aspect that is often observed is the overestimation of production values. Meaning that the projected production levels outlined in the PDOs tend to be higher than what is actually achieved during the operation of the projects.

Demirmen [32] believed that forecasting producible reserves could pose challenges for the petroleum industry. He discussed various sources of uncertainty along with the procedures involved in such forecasts and recommended ways to improve them. He utilized the data from the NPD to analyze changes in Ultimate Recovery (UR) forecasts for 15 oil fields on the NCS between 1974 and 2004. He stated that, apart from fluctuations, the overall trend indicated a growth in forecasts of expected production values over time. In a subsequent study [33], he examined the trends of forecasted reserves in 38 oil and gas fields developed on the NCS during the same period and demonstrated a 30% growth.

Given that the initial facilities in fields were designed based on development plans, Demirmen [32] , [33] argued that inaccurate forecasts of producible reserves could lead to economic damage. Such poor forecasts may result in the misallocation of capital and underperformance in production, leading to wasted investments. Conversely, overestimations could lead to a lack of additional wells or necessary adjustments to handle the overproduction, causing further economic losses.

Nandurdikar and Wallace [34], conducted a study utilizing data from 147 petroleum projects sourced from the Independent Project Analysis (IPA) over a 16-year period. Their research demonstrated that production deliveries, when compared to the forecasted production volumes at the project sanction stage, declined over time. In 1995, these projects were able to deliver 94% of the promised production values, but by 2011, the actual production only amounted to 75% of the initially forecasted volumes. The researchers asserted that this decline was primarily attributed to poor and overly optimistic production forecasts, which ultimately resulted in production underperformance.

In a 2020 study, Bratvold et al. [2] compared annual data of actual productions with forecasted values (including mean, p10, and p90) using information from the NPD's database. They focused on data from 56 oil fields involved in Exploration and Production (E&P) development projects spanning the period from 1995 to 2017. The researchers found that the forecasts tended to be overly optimistic and overconfident, which led to poor decision-making during the Final Investment Decisions (FID) process. They attributed these shortcomings in production forecasts to factors such as delusion and deception, suggesting that strategic mismanagement or misguided beliefs may have contributed to these inaccuracies.

They used data from 32 out of a sample of 56 oil fields as valid production forecasts, including P10, mean, and P90 forecasts. They demonstrated the biases that emerged when comparing these forecasts to the cumulative actual production during the initial four years of production. On one hand, they revealed that instead of 80%, only 31% of the actual production values fell within the confidence interval between P10 and P90. On the other hand, when considering the P10 forecasts, they illustrated that around 59% of the production values were lower than the forecasted values [2].

Two years later, in 2022, Nesvold and Bratvold [3] corroborated the findings of the previous study regarding biased production forecasts used in the approval of petroleum development projects on the NCS. They examined production data from 71 oil fields that commenced production between 1995 and 2020. The researchers employed and compared three methods for reducing bias, namely the Reference Class Forecasting (RCF), Mean calibration, and Triplet calibration. They specifically focused on oil forecasts beyond the second year of production. Nesvold and Bratvold asserted that production forecasts should be adjusted downwards and that a broader uncertainty range must be incorporated into these forecasts. These findings emphasize the importance of improving the accuracy of production forecasts by employing bias reduction techniques and utilizing more realistic uncertainty ranges in the petroleum industry.

1.3 Thesis outline

Decision-making in projects that involve significant investments, resources, or scope is often complex, highly uncertain, and consequential. This is why the quality of decisions can be influenced by both the decision-makers and those who provide information to them [35].

When considering petroleum projects, as discussed in previous studies, the objective of decision-makers is to maximize project benefits while minimizing risks. However, it has been clearly demonstrated that biased and inaccurate forecasts of project costs, development times, and production volumes have a significant impact on the optimal decision-making process regarding how, when, and whether a project should be developed. These flawed forecasts can have adverse effects on decision-making in multiple ways. Biased cost forecasts can lead to improper allocation of resources, resulting in financial losses or missed opportunities. Inaccurate predictions of development times can cause delays and disrupt project schedules, leading to additional costs and potential revenue losses. Furthermore, erroneous forecasts of production volumes can affect revenue projections and profitability, ultimately resulting in suboptimal investment decisions.

In this study, the focus is on investigating the forecasts provided by operators in the PDO on the NCS. The aim is to identify the disparities between the actual values and the forecasts pertaining to development costs, time, and production values.

Furthermore, the study aims to calculate the loss in value resulting from cost and time overruns, as well as production shortfalls. This assessment allows for a comprehensive understanding of the impact these deviations have on project outcomes.

Lastly, the study delves into the relationship between the cumulative underproduction within the first ten years of the projects and the total cost overruns observed in the same projects. To explore this relationship, three regression methods are employed to analyze and determine any potential correlations.

1.4 Thesis structure

This study consists of six chapters. The first chapter provides a brief overview of previous research on the topic of overruns in development projects, specifically focusing on petroleum projects. It also includes a summary of the thesis outline and structure.

In the second chapter, the Act of 1996, phases of petroleum projects, uncertainties in forecasts of the projects on the NCS are presented. Additionally, the chapter explains the Present Value (PV) method for valuation of cash flows over time and discusses the utilization of Pearson, Spearman, and SVR methods used to establish relationships between variables.

Chapter three is dedicated to describing the databases used in this thesis, as well as the categories of data. The chapter also acknowledges the limitations of the thesis.

Chapter four comprises the analysis of actual and forecasted values of development time, cost, and production data. This analysis enables the calculation of cost and schedule overruns, as well as production shortfalls. Two methods are employed to estimate the total present value loss resulting from these poor and inaccurate forecasts. Furthermore, this chapter investigates the correlation between cumulative production underperformance and total cost overruns during the first ten years of production using Pearson, Spearman, and SVR regression methods.

Based on the analyses conducted in chapter four, chapter five discusses the results and relates them to previous research. It addresses the research questions and offers recommendations for future studies.

Finally, chapter six presents the conclusions drawn from the results obtained in the study.

1.5 Key contributions

As stated by Flyvbjerg [36] due to his wide research in the area of overruns in the megaprojects and as it was mentioned in the section related to the previous works, benefit shortfalls and cost and schedule overruns in megaprojects are more like a well-known rule. Therefore, several private and public research works have focused on the topic of the overruns or shortfalls resulted from biased forecasts in the development projects.

However, it is worth mentioning that to the best knowledge of the author, not many studies have focused on the value loss that such forecasts could bring for development projects, especially in the petroleum industry.

Considering this gap in previous works, this study, using the method for calculating the present value recommended by Mohus [6], compared the actual and forecasted data on the development time, cost and production values for development projects on the NCS.

The data was related to the first ten years of production until 2022 and considered a wider time-frame.

The results of this study explained in detail in chapter four, demonstrated that value loss due to delays was around NOK 39 billion, while this loss for the cost overruns and underperformance in production was almost NOK 268 billion and about NOK 80 billion, respectively.

Another important point to consider, was that although many studies have examined cost overruns and production shortfalls, but when searching for studies investigating a possible relationship between underproduction and cost overruns, there are not many cases out there. Apart from the study by Nesvold and Bratvold [4], that investigated the relationship between biases in production forecasts and field features and another one by Mohus [27], the author did not find any other research working on possible relationships of cost overruns with underproduction. As a result, this study tried to fill this gap by using three regression methods to find the correlation between the cumulative underproduction in the first ten years and the total cost overruns in those years for all the fields. And when using the Pearson, Spearman and Support Vector Regression methods, it turned out that there no linear correlation and weak non-linear relationship between the two aforementioned variables for the development projects in this study.

Chapter 2 – Methodology

This chapter provides a summary of the Act of 1996, outlining its role in governing petroleum activities. It discusses the phases of petroleum projects, addressing the uncertainties associated with forecasting. The chapter also explains the methods and software utilized for analyses of this study. It describes the use of discounting factors and inflation rates incorporated into the PV method for calculating the financial impact of overruns and underproduction. It introduces statistical techniques such as Pearson, Spearman, and SVR to analyze the correlation between cumulative underproductions in the first 10 years and total cost overruns. Overall, this chapter sets the foundation for understanding the legal framework, project phases, forecasting uncertainties, and Present Value and regression methods used in the analysis of chapter 4.

2.1 Act on petroleum activities on the NCS

Norway possesses significant hydrocarbon reserves, primarily located in offshore areas of the NCS, which have played a vital role in the country's economic development. To ensure the profitability of petroleum activities and safeguard the nation's interests, the Norwegian government established the Petroleum Act of 1996. This legislation grants exclusive rights to the Norwegian state for subsea petroleum deposits and resource management. The petroleum deposit aspect recognizes the complexity, uncertainty, and diverse industries involved in these projects. To address these factors, the PDO must comprehensively consider all aspects and challenges associated with the project. The resource management aspect focuses on maximizing reservoir resources and minimizing waste through regular evaluation of production strategies and technical solutions employed by operators. By emphasizing the importance of strategic planning and effective resource management, the Petroleum Act aims to optimize the utilization of Norway's petroleum resources for the benefit of the nation [37].

2.2 Phases of the Petroleum Projects on the NCS

In accordance with the Petroleum Act, all organizations seeking to engage in petroleum activities must follow a standardized process and obtain necessary licenses and approvals from the authorities. Figure 2 illustrates the official process mandated by the Ministry of Petroleum and Energy (MPE) in Norway. The key phases of a petroleum project, including exploration, discovery, and field development, are depicted in the figure. Additionally, the required applications and documents for each phase are outlined. The five main phases, namely concession and production license, producible reserves determination, development, production, and decommissioning, are briefly described below [38]. This process ensures that all necessary steps and considerations are taken into account from the initial exploration phase to the ultimate decommissioning of the field, promoting responsible and efficient management of petroleum resources in Norway.

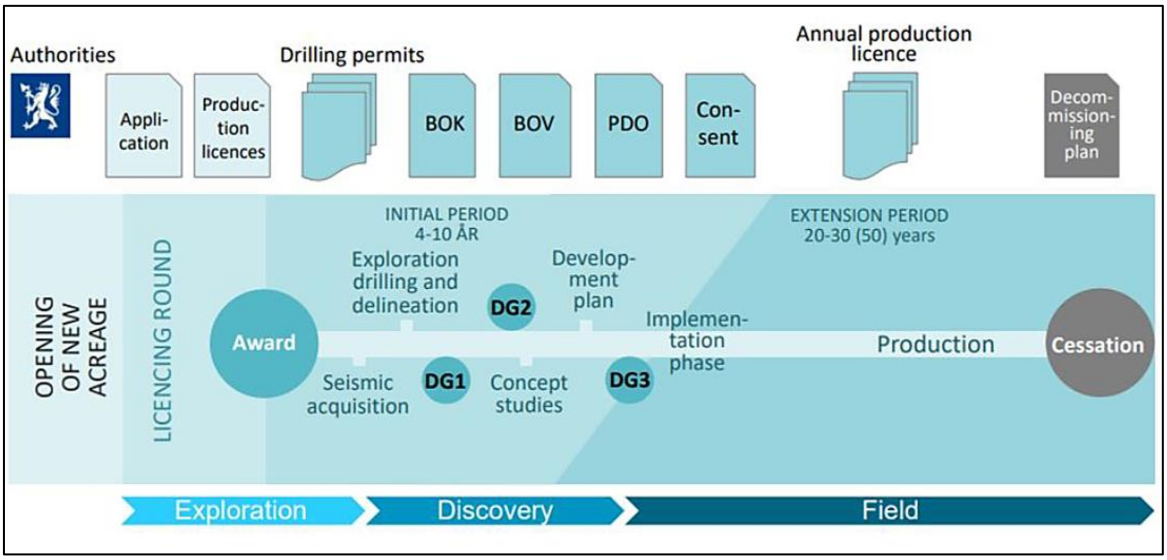

Figure 2 - Phases of petroleum projects on the NCS [39]

Concession and Production License: In the Concession and Production License phase, the areas for petroleum activities on the NCS are opened by the MPE. MPE conducts assessments to evaluate the resources in the area, as well as the economic, social, and environmental impacts of the proposed petroleum activities. Based on these assessments, the Norwegian Parliament makes a decision to either open the area for petroleum activities or not.

Once an area is opened, operators can apply for a production license during the licensing rounds announced by the MPE. The production license grants the operator the rights to explore, drill, and produce petroleum products in the designated area. The specific type of production license granted to each operator is determined by their application and the evaluation process [38].

Producible Reserves Determination:

In the Producible Reserves Determination phase, operators who have been awarded a production license for a period of ten years are required by law to develop a work program. This work program outlines the geological, geophysical, and exploration drilling activities that the operator plans to undertake within specified deadlines, depending on the terms of the license.

The primary objective of the work program is to ensure the thorough exploration of the designated areas. If operators successfully complete their work programs and make any discoveries during the initial ten-year period, they may have the opportunity to extend their production licenses for up to 30 years, subject to the assessment and approval of MPE [38].

Development and Production: During the extension period of the production license, operators are allowed to develop and operate. However, before initiating the planning and implementation of the development projects, they must obtain approval from MPE by submitting a PDO or a plan for installation and operation (PIO) for pipelines and onshore projects. PDOs or PIOs consist of two main parts: the development plan and the impact assessment. The impact assessment part specifically focuses on evaluating the effects of the petroleum projects on the local community, businesses, and environment [38].

The development plan for petroleum projects must encompass comprehensive details regarding hydrocarbon production to ensure profitability. These details include potential installations, constructions, anticipated future production incomes, as well as costs associated with development, operation, and decommissioning. In the guidelines provided by MPE for PDOs and PIOs, the main stages of development planning are outlined as feasibility studies, concept development, primary engineering, submission, and the final decision by the relevant authority [40].

Decommissioning: As part of the licensing requirements, operators with licenses on the Norwegian Continental Shelf are required by MPE to submit a decommissioning plan within two to five years before their license expires. This plan must provide a comprehensive assessment of the impacts associated with the shutdown, including an impact assessment. Additionally, the plan should outline the procedures and details for decommissioning the installations and closing operations, which are typically included in a disposal plan [38].

2.3 Uncertainties in Petroleum Projects Forecasts

Having a comprehensive understanding of the uncertainties associated with reservoir and production forecasts is crucial as it directly influences development decisions. Accurate forecasts are essential for effective planning and resource management. According to the NPD [5], operators on the NCS are required to provide probabilistic forecasts for production values. These forecasts should include p10 (forecast), mean forecast, and p90 (high forecast) to account for the range of possible outcomes and provide a clearer picture of the expected production potential.

The p10 value represents the forecast with a 0.1 probability of being lower than the actual values, while the p90 value represents the forecast with a 0.9 probability of being higher than the actual values. Therefore, if the p10 and p90 forecasts are accurate, approximately 80% of the actual values should fall within this interval.

As projects progress and more information is gathered, the uncertainties associated with the forecasts are gradually reduced. This leads to a narrowing of the interval between the p10 and p90 values, bringing them closer to the mean forecasts. This means that with the accumulation of data and improved understanding of the reservoirs and production processes, the level of uncertainty decreases, resulting in more precise forecasts [6].

2.4 Discounted Cash Flow Valuation

Just like any other project, the profitability of development projects is determined by analyzing their costs and revenues. The costs of these projects can be categorized into development costs, operation costs, and decommissioning costs. In this study, the focus was specifically on development costs and capital investment costs (CAPEX). On the other hand, the revenues considered in this study were generated from the sale of hydrocarbons.

Many public and private decision-making processes take into account the concept that the value of money diminishes over time. This means that there is a discounting effect applied to future values, making a unit of currency worth less in the future compared to its present value. This principle is known as the time value of money. It recognizes that due to factors such as inflation, opportunity costs, and risk, receiving a certain amount of money today is generally considered more valuable than receiving the same amount in the future. Discounting future values allows decision-makers to properly assess the costs, benefits, and profitability of projects by considering the timing and uncertainty associated with cash flows [41].

One way of Discounting Cash Flows (DCF) is calculating the Present Value (PV) of future cash flows or values. Equation 1 is commonly used in PV computation when applying a constant discount rate across all years [42]:

$$PV = \sum_{k=1}^{n} \frac{FV_k}{(1+i)^k} \tag{1}$$

Where:

FV_k = Future Value at year k

i = discounting rate

n = number of periods

In this study, a constant discounting factor or Weighted Average Cost of Capital (WACC) of 0.08 was assumed for all the years.

In addition, When calculating future values with a constant discounting factor for each year, Equation 2 can be used [42]:

$$FV = \sum_{k=1}^{n} PV_k(1+i)^k \tag{2}$$

Where:

FV = Future value

PV_k = Present Value at year k

n = number of periods

If different discounting factors are considered for each year such as the case for annual inflation rates, the factor for converting each value at year k to a future value at year 2022 can be formulated as Equation 3. In this study, same formula is used to convert the cash flows (costs and production revenues) to 2022-NOK values.

Conversion factor for value in year k to a value in year 2022 = $(1+i_k)\ (1+i_{k+1})\ \ldots\ (1+i_{2022})$ (3)

Based on the above formula and the annual inflation rates between 1990 and 2022 [43], the conversion factors used to take values to 2022-NOK are summarized in Table 2.

Table 2 - Conversion factors for taking values at year k to year 2022

		coverting value of year k to value of year k+1	Converting value of year k to value of year 2022
Year (k)	Annual inflation rate	(1+inflation rate%)	(1+%ik)*(1+%i(k+1))*…*(1+%i2022)
1990	4.13	1.0344	2.03
1991	3.44	1.02	1.96
1992	2.33	1.02	1.92
1993	2.29	1.01	1.88
1994	1.38	1.02	1.85
1995	2.46	1.0126	1.81
1996	1.26	1.0257	1.78
1997	2.57	1.0225	1.74
1998	2.25	1.0237	1.70
1999	2.37	1.0309	1.66
2000	3.09	1.0300	1.61
2001	3.00	1.0129	1.57
2002	1.29	1.0249	1.55
2003	2.49	1.0045	1.51
2004	0.45	1.0153	1.50
2005	1.53	1.0233	1.48
2006	2.33	1.0071	1.45
2007	0.71	1.0375	1.43
2008	3.75	1.0220	1.38
2009	2.20	1.0242	1.35
2010	2.42	1.0128	1.32
2011	1.28	1.0070	1.30
2012	0.70	1.0212	1.30
2013	2.12	1.0204	1.27
2014	2.04	1.0217	1.24
2015	2.17	1.0355	1.22
2016	3.55	1.0188	1.18
2017	1.88	1.0276	1.15
2018	2.76	1.0217	1.12
2019	2.17	1.0129	1.10
2020	1.29	1.0348	1.08
2021	3.48	1.0481	1.05
2022	4.81	1.0000	1.00

According to Mohus [6], two methods can be used for the comparison of forecasted and actual values with regard to time. When there is no delay in the actual production start compared with the forecasted start, the two values can be compared in the same years. However, when there is a delay in the actual production start, two methods can be employed to compare the forecasted and actual values based on their respective years. This section provides a brief explanation of these methods. For more detailed information, please refer to [6]:

Method 1 – Shifting to Forecasted Production Start: This method suggests that when comparing actual and forecasted values for delayed fields, the actual production values for the delayed years should be set as zero. This approach takes into account the impact of delays on underproduction.

Method 2 – Shifting to Actual Production Start: The second method proposes that when there are delays in the actual production start, the forecasted production values for the delayed years can be shifted to align with the actual production start for comparison with the actual values. However, this method does not consider the effect of schedule overruns on production underperformance.

Considering the previous clarifications, there are two important subjects worth mentioning:

1) The difference between the results of the two methods demonstrates the value loss due to delays.
2) Even though the sum of actual productions and forecasted productions in the two methods remains the same, due to the time value of money and the discounting

effect, the approximate present value calculated based on each method would be different when shifting the production values in time.

For the purpose of calculating the monetary value lost due to delays and underproduction (or the economic value erosion due to poor forecasts of development time and production values), Equation 4, as suggested by Mohus [6] for the Present Value (PV), was utilized in this study:

$$PV_i = \frac{Production_i * Conversion\ rate * Exchange\ rate_i * Oil\ price_i}{(1+wacc_i)^t} \tag{4}$$

Where:

PV_i = Present Value (in million NOK) of a value at year i

i = Number of years after PDO approval

$Production_i$ = Actual or Forecasted production values in year i in million Sm^3

Conversion rate = 6.29 bbl. for each 1 Sm^3 from [44]

$Exchange\ rate_i$ = For converting USD Dollar to NOK in year i from 1995 to 2021 [45]

$Oil\ price_i$ = Brent Oil spot price in USD Dollar per bbl. (Barrel) from 1995 to 2021 [46]

$WACC_i$ = 8% for the petroleum projects on the NCS

t = PDO approval year = time zero

By utilizing Equation 4, the forecasted and actual production revenues from selling petroleum products were calculated. Subsequently, with a discounting rate set at 8%, they were discounted back to the year of PDO approval, where development decisions were made. To enable a comparison and determination of total forecasted and actual cash flows, the annual inflation rates [43] from 1990 to 2019 were applied to convert the revenues into 2022-NOK values.

2.5 Regression Methods

To investigate the correlation between cumulative underproductions in the first 10 years (y) and total cost overruns (x), regression analysis was employed. Specifically, three regression methods were utilized: Pearson correlation coefficient, Spearman correlation coefficient, and SVR. By employing these regression methods, the study aimed to uncover the nature and strength of the correlation between cumulative underproductions and total cost overruns.

2.5.1 Pearson Regression

In a linear regression model, the relationship between two variables, Y and X, can be represented by "Y = a + bX". The coefficient "a" is the intercept, which represents the value of Y when X is zero. The coefficient "b" is the correlation coefficient or slope, indicating the change in Y for a unit change in X.

The Pearson correlation coefficient measures the linear relationship between two variables and ranges from -1 to +1. A value of -1 indicates a perfect negative linear relationship, meaning that as one variable increases, the other decreases in a perfectly predictable manner. A value of +1 represents a perfect positive linear relationship, where both variables increase or decrease together. A correlation coefficient of 0 suggests no linear relationship between the variable [47].

The Pearson regression method specifically focuses on capturing the linear relationship between variables, considering a constant rate of change. It calculates the correlation coefficient to quantify the strength and direction of the linear relationship between the variables. By analyzing the Pearson correlation coefficient, researchers can gain insights into how closely the variables are associated in a linear manner. Pearson Correlation Coefficient formula involves calculating the covariance between the independent and dependent variables, and dividing it by the product of the standard deviations of the independent and dependent variables as shown in Equation 5 [47]:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-y)^2}} \tag{5}$$

Where:

r_{xy} = Pearson Correlation Coefficient

x_i and y_i = independent and dependent data points

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ = Mean of dependent variables (x_i)

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ = Mean of dependent variables (y_i)

Overall, the Pearson regression analysis has its own benefits and limitations. For instance, it is a simple method that can be easily interpreted and provides a single number (R) that can be used to understand the direction and magnitude of the correlation between two variables. It can be applied to a wide range of continuous datasets, provides a quantitative assessment, and its standardized scale simplifies the comparison of relationship strengths across various variables. On the other hand, it is limited to considering linear relationships with constant rates (slope), and it is greatly affected by outliers, which can distort the calculated value. Additionally, it does not establish causation between the variables [48].

2.5.2 Spearman Regression

The Spearman correlation coefficient measures the strength and direction of the monotonic relationship between two variables. It is based on the ranks of the data rather than the actual values. The Spearman correlation coefficient can range from -1 to +1, where -1 indicates a perfect negative monotonic relationship, +1 indicates a perfect positive monotonic relationship, and 0 indicates no monotonic relationship between the variables. The Spearman correlation coefficient formula involves calculating the squared differences between the ranks of each pair of data points, summing these squared differences, and applying Equation 6 as shown below [49]:

$$I_s = 1 - \frac{6(\sum di^2)}{n(n^2-1)} \tag{6}$$

Where:

N = number of data

d_i = difference between ranking of any pair of data

I_s = Spearman correlation Coefficient

Summarizing the pros and cons of Spearman regression analysis, it is worth mentioning that it is more flexible than Pearson's correlation coefficient. That is because, instead of considering only linear relationships with constant rates, it can capture monotonic (ascending or descending) relationships that are not linear as well. Due to the use of ranks of the data instead of the raw data, it is less influenced by outliers and is less sensitive to the uneven variability of the data as well. It can be used for data that are not normally distributed. On the other side, it has some limitations. For instance, it loses some information by not considering the raw data and instead working with their ranks. For small sample sizes, its results are less accurate than the results obtained using Pearson's correlation coefficient. Additionally, it can only capture monotonic relationships and cannot identify relationships with specific patterns or curvilinear associations [50].

2.5.3 Support Vector Regression

Support Vector Machines (SVM) are a type of supervised machine learning method that can be used for different purposes, such as regression analysis, pattern recognition, and prediction. This method was proposed by Vapnik [51], and combines statistical learning with mathematical optimization. Support Vector Regression (SVR) is an extension of SVM for regression tasks, which considers the closest data points (support vectors) to the regression line or the so-called "hyperplane" in order to reduce the effect of outliers in the analysis, maximize the fit of the data on the regression line, and minimize the errors. SVR can be applied to both linear and non-linear problems.

SVR models allow for a tolerance margin called the epsilon (ε), which determines the width between the data and the hyperplane. Although SVR is capable of handling non-linear relationships effectively and can control the model complexity by adjusting the epsilon and the margin, it is sensitive to the choice of kernel function and hyperparameters. Additionally, SVR models can be complex and provide limited interpretability regarding the data [52].

Kernel functions typically act as a black box in SVRs and encompass various types, including linear, Gaussian Radial Basis Function (RBF), polynomial, or sigmoidal kernels. These kernels are utilized to reduce the dimensionality of the data and effectively handle non-linear relationships [53].

Depending on the type of optimization problem or use case, the steps and formulation of SVR, as well as the libraries used, can vary [54]. When using SVR for regression analysis like what is done in this study using Python, the steps can be listed as follows:

1) Prepare the data by identifying the independent (x) and dependent (y) variables.
2) Ensure that the data is on the same scale to avoid one variable dominating another. The most common approach is to use normalization, which involves scaling both variables between 0 and 1.
3) Determine the formulation of the hyperplane that best fits the data with the desired margin.
4) Plot the data and hyperplane to enhance visual understanding.

To summarize the strengths of the SVR method it can be said that: it is able to capture non-linear relationships between variables, it is less sensitive to outliers by minimizing errors within a certain range, and it aims at minimizing the overall error across the entire dataset. The weaknesses associated with SVR can also be summarized as: being computationally complex, especially when dealing with large datasets, being sensitive to the choice of parameters used in its algorithm, requiring careful selection and tuning and providing limited information about the relationship between variables and is primarily utilized for prediction accuracy rather than interpretability [55].

Chapter 3 – Data

The objective of this chapter is to provide a clear and thorough explanation of how the data used in this study were extracted. Besides, the datasets used for the analyses, description and categories of the of the data in addition limitations and assumptions of the present study are described as well.

3.1 Databases

The databases used for this study can be categorized into three groups. Firstly, the annual reports on the national budget included data on the costs of the projects under development for each year. These data were collected, digitized, and sorted in an unpublished work by Bratvold and Sheikhoushaghi [28].

Secondly, publicly available database from the NPD website [5] was used, which included information on the development time, cost, and actual production values of the petroleum fields on the NCS. Thirdly, a confidential database including data[2] on the forecasts of production provided by the operators on the NCS at the time of PDO approvals.

3.2 Data Categories

The data used in this study for conducting the analyses consisted of four groups: development cost data, development time data, production data, and general data. In the following section, a brief description is provided for each group.

[2] A non-disclosure agreement was signed before the start of the study in order to gain access to and preserve the confidentiality of the available information on production forecasts. Therefore, if the names of any fields or operators are mentioned in any part, they are referring to the public data on NPD.

3.2.1 Development Cost Data

For the purpose of this study, the data related to the development costs included the initial forecasted costs of the projects, the total cost changes at the end of the projects, and the final year of the projects. Subsequently, the total cost changes in the last year were added to the initial forecasts to calculate the actual costs of the projects for use in the analyses. In this part, 142 development projects with PDO approval from 2001 to 2022 were utilized in the analyses.

3.2.2 Development Time Data

Regarding the analysis of development times, the following data points were utilized: the forecasted year of first oil, the actual year of project start, and the year of PDO approval. In this study, the development time was defined as the time between the PDO approval year and the year of production start. The analysis made use of the development time data obtained from the confidential dataset and focused on 76 oil fields with PDO approval years ranging from 1990 to 2019.

3.2.3 Production Data

The actual production data was extracted from the NPD website [56], while the data on forecasted production volumes for each field in each year was obtained from the confidential dataset. In this work, data from 67 oil fields with production years ranging from 1990 to 2021 were utilized in the analysis.

3.2.4 General Data

In the analysis of calculating the economic value loss, several other public datasets were used, which are further explained in the next chapter. These datasets included:

1. The annual Brent oil prices in Europe (in Dollars per Barrel) from 1995 until 2022 [46],
2. The data on annual inflation rates from 1990 to 2022 was used to convert values at different years to 2022-NOK value [43],
3. The historical data on annual exchange rates between 1995 and 2022 for converting US Dollar to NOK extracted form Norges Bank [45],
4. The factor for converting crude oil volume to barrels as follows: 1 Sm^3 crude oil = 6.29 barrels [44],
5. The Weighted Average Cost of Capital (WACC) used for the petroleum projects on the NCS in this study was determined to be 8%. This percentage was derived as an average of the percentages commonly utilized in the petroleum industry in Norway, taking into account the opinions of experienced individuals in the field. In a study by Franc-Dąbrowska, Mądra-Sawicka and Milewska [57], the WACC used by 231 European energy companies in their financial valuations between 2015 and 2019 was examined. According to their findings, the WACC percentage employed in the Oil and Gas E&P industries was approximately 7.2%. A report provided by Vår Energi [58], for the third quarter of 2022 stated a discount rate of 8% to be used in economic evaluations. Another report by OKEA [59], used a WACC of 10% in their post-tax value testing. Additionally, NORECO [60] employed a WACC of 6.14%. Finally, according to a report by Equinor [61], the discount rate for E&P projects on the NCS ranged from 5% to 9%.

3.3 Study Limitations

Based on the datasets used for the analyses in this study and certain aspects of the data, certain limitations were inevitable. Due to incomplete or missing data for certain fields or years, it was decided to exclude them from the study. Specifically, in the data for years 1993, 1998, 2003, and 2016, there was only one available data point either for the actual or forecasted values, making it impossible to compare and analyze overruns or shortfalls. As a result, such data was excluded from the analysis.

Another limitation was observed in the development cost data, which did not include finished projects in 2022. Therefore, the study had to consider projects only until 2021. Additionally, out of a total of 148 projects, data for 6 projects, such as the first forecasted cost, final production year, or total changes in costs, were missing partially or entirely. As a result, those projects had to be excluded from the analysis.

In the forecasted production data, there were instances where the first year of production start preceded the PDO approval. As a result, those fields were excluded from the analysis, assuming that the reporting was done incorrectly.

Furthermore, in this study, only data on oil fields and projects were utilized, considering oil as the most significant petroleum product.

According to Nandurdikar and Wallace [34], for petroleum projects, significant investments are typically received after the first ten years of production, which can alter the nature of these projects from what was initially approved in the PDO. Consequently, this study solely considered production data for the first ten years after the start of production.

In the calculations of production underperformance and value losses, only the mean forecasts of production values were utilized.

There was a lack of monthly production data, preventing the investigation of the monthly effects of delays on underproduction and revenue loss.

Due to the unavailability of monthly production data, development times and dates were only considered in years. This means that projects with delays of less than one year were rounded up to one year.

The accuracy of the study could have been improved if more data on cost, time, and production were available.

Lastly, this study exclusively focused on oil fields located on the Norwegian Continental Shelf. Including projects from other parts of the world could lead to a more comprehensive analysis.

Chapter 4 – Analysis and Results

In this chapter, the analysis performed on the data and the results obtained using Python and MS Excel software are explained. The analyses in this section can be categorized into five groups, which are thoroughly defined in the following subsections:

1) Analysis of the development cost data and the budget overruns related to the projects,

2) Analysis of the development time data and delays in starting the projects,

3) Analysis of the production data to calculate the underperformance in promised production values,

4) Economic analysis of the value loss due to cost and schedule overruns and underproduction separately, as well as in combination. In this part, two methods of adjusting the time zero in calculating an approximation for the PV erosion, as a result of deviations from forecasts, were utilized.

5) Regression analysis using Pearson, Spearman and SVR methods to investigate the correlation between cumulative underproduction in the first 10 years and total cost overruns.

4.1 Analysis of the development cost data

As discussed earlier, numerous research works have explored the topic of cost overruns, and it has become widely acknowledged that overruns in megaprojects are more of a rule, as highlighted by Flyvbjerg [36].

To identify the cost overruns, the initial forecasted costs and the final actual costs, as reported in the last available project reports in the annual national budget [28], were utilized. These costs were then converted to NOK values using conversion factors from the final year of the projects to 2022. By subtracting the forecasted costs from the actual costs, the budgeted cost overruns were calculated. The data was subsequently sorted based on the PDO approval years, and the total amounts of the forecasted development costs and cost overruns were calculated for each year.

In Figure 3, a graph is presented showing the total forecasted costs represented by blue columns and the total cost overruns depicted by orange columns for each PDO approval year. Additionally, the percentage of cost overrun for each PDO year is illustrated on the secondary axis as a grey line.

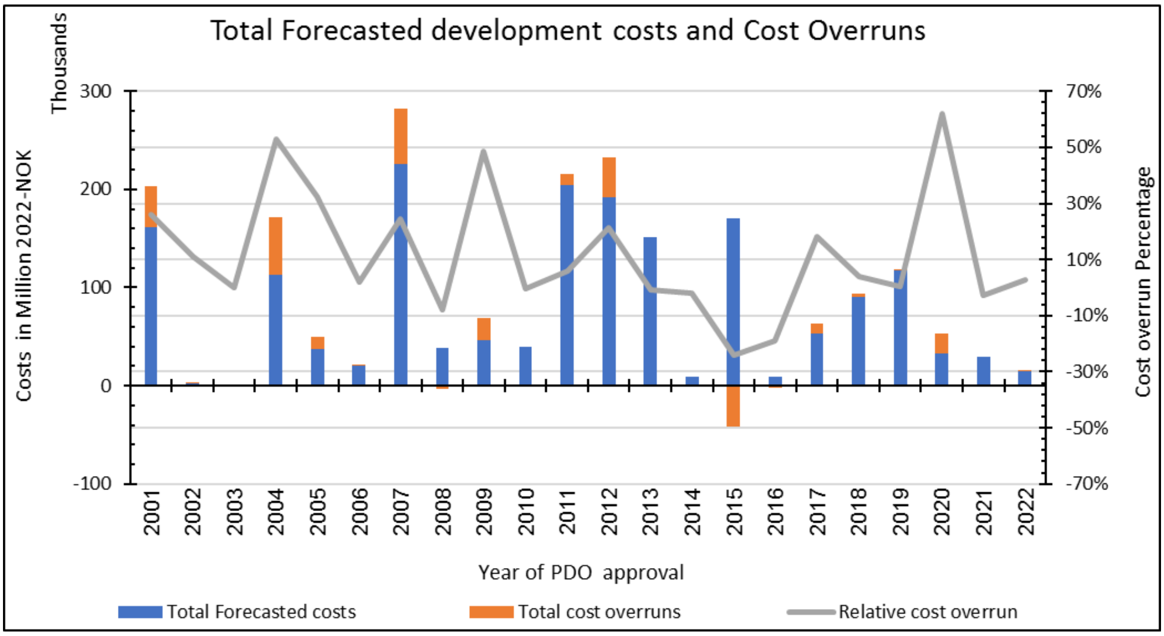

Figure 3 - Total forecasted costs, Total cost overruns and relative cost overruns for oil projects on the NCS in each PDO approval year

As observed in Figure 3, the highest cost overrun percentages were in 2020, 2004, and 2009, reaching 62%, 53%, and 49%, respectively.

In the years 2008, 2013, 2014, 2015, 2016, and 2021, the actual costs were lower than the forecasted costs, indicating cost reductions during those periods.

The highest percentages of cost reductions were observed in 2015, 2016, and 2008, with values of 24%, 19%, and 8%, respectively. Additionally, it is evident from the graph that no PDO was approved in 2003.

From 2001 to 2008, there was a declining trend in cost overruns, with the percentage decreasing from 26% to -8%. However, there were two peaks during this period, occurring in 2004 and 2007, with cost overruns of 53% and 25%, respectively.

Between 2009 and 2015, there was an overall decreasing trend in relative cost overruns, with the percentage decreasing from 49% to -24%. The peak during this period was around 21% in 2012.

From 2016 until 2022, there was a rise in cost overruns, with the percentage increasing from -19% to 3%. However, there were two sharp reductions in 2019 and 2021, with cost overruns of 4% and 65%, respectively.

To analyze the trend of development cost forecasts, the study utilized the Simple Moving Average (SMA)[3] for the percentage of cost overruns in the mentioned projects. The analysis considered both 10-year and 5-year periods. Figure 4 displays the percentages of cost overruns, along with the 10-year SMA (shown in gray) and the 5-year SMA (shown in orange). The data is sorted based on the PDO approval years.

The blue line, representing the percentages of cost overruns, fluctuated throughout the period. It reached its peak of 62% in 2020 and reached its lowest point of -24% in 2015. While the improvement is not very apparent in the 5-year SMA, the 10-year SMA indicates a slight reduction in cost overruns, suggesting a small improvement in the accuracy of development cost forecasts.

[3] The SMA is widely used to identify trends in data. By calculating the average value of a variable over a specific time period, it smooths out short-term fluctuations and highlights the underlying trend.

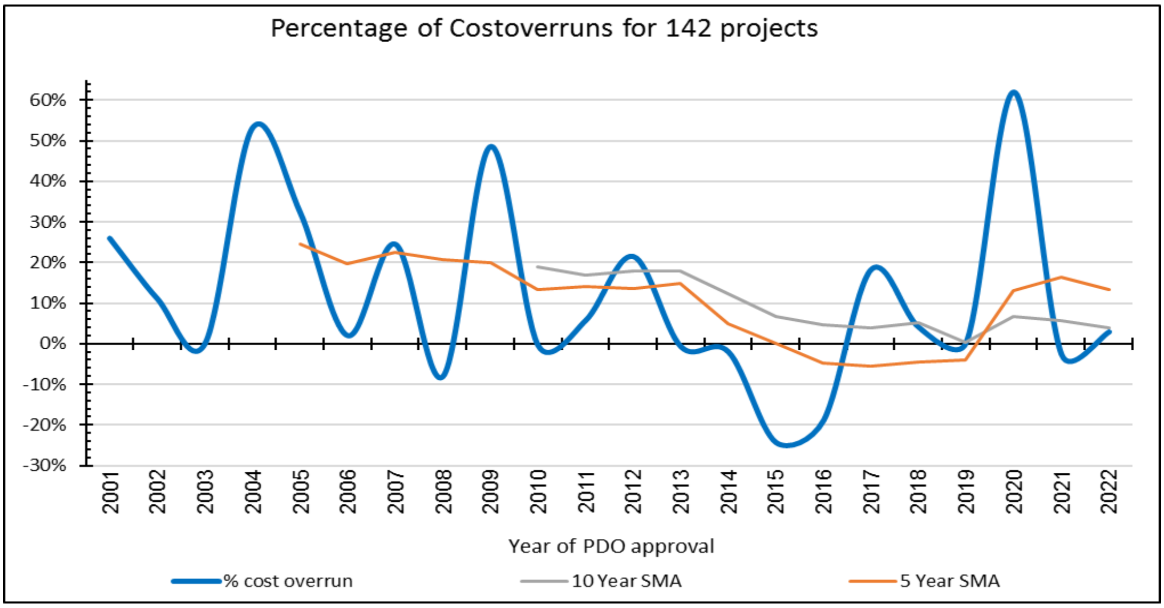

Figure 4 - The trend of cost overrun percentage and the simple moving average lines for 142 projects

Out of the total of 142 projects included in the study, 87 projects (61%) experienced cost overruns, indicating that the actual costs exceeded the initial forecasts. On the other hand, 47 projects (33%) were completed with costs lower than the forecasts, while only 8 projects (6%) were finished with the same forecasted costs.

Between 2001 and 2022, the total actual development costs for the analyzed projects amounted to approximately NOK 2381 billion, while the forecasted development costs were around NOK 2112 billion. This reveals a total cost overrun of approximately NOK 268 billion, corresponding to a 12.7% overrun from the initial budgets.

These findings, along with previous studies on cost overruns in petroleum projects on the NCS, indicate that there is no consistent evidence of projects being delivered at costs lower than the initial forecasts.

4.2 Analysis of the development time data

The calculation of delays in the development projects was based on the comparison between the forecasted production start dates or time intervals provided in the PDOs and the actual production starts. It was assumed that if a project started within the forecasted time frame, it would be considered on schedule. The difference between the forecasted and actual production starts was used to determine the schedule overruns.

It is important to note that the analysis considered the time between the PDO approval year and the production start as the development time, and all the data was measured in years. When the forecasted first oil years were subtracted from the actual production starts for the 76 fields on the NCS with PDO approval between 1990 and 2019, the average development delay was calculated to be approximately 101 days, which corresponds to a 12% schedule overrun on average. Additionally, the average expected development time was computed to be around 2.07 years.

The findings of this study, which indicate an average development delay of approximately 101 days (equivalent to around 3.4 months), are consistent with the information reported in the 2020's report on project execution on the NCS [62]. According to that report, the average schedule overrun for development projects between 2007 and 2018 was stated to be 3.5 months. These aligned results suggest that the analysis in this study is in line with the reported information by NPD.

Figure 5 presents the results of the analysis conducted on the available data regarding forecasted development times, development time delays, and the percentages of development delay for the 76 oil fields included in the study. The data is presented in the form of blue columns representing forecasted development times, orange columns representing development time delays, and a gray line representing the percentages of development delay on a secondary axis. The projects are sorted based on their PDO approval year.

From the analysis, it was found that out of the 76 oil fields, 20 fields (26%) started producing with delays after their forecasted production start, indicating a deviation from the expected schedule. On the other hand, the development of 54 projects (71%) was completed within the forecasted development time as promised. Only 2 fields (3% of the projects) were finished before the forecasted development time, indicating an early completion.

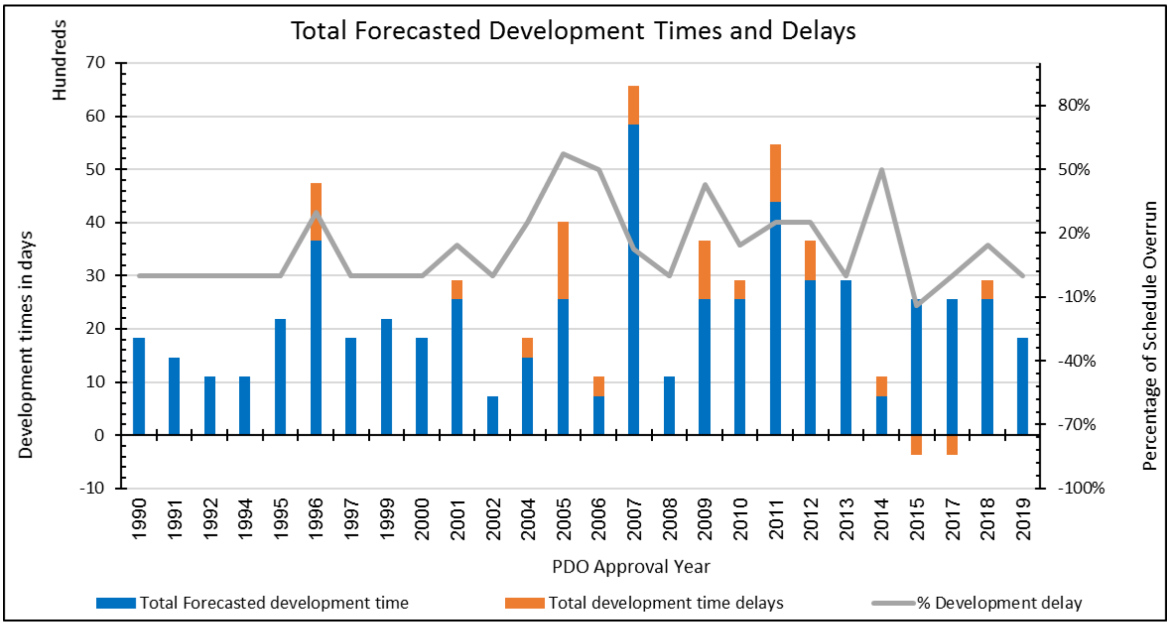

Figure 5 - Forecasted development times, development time delays and percentage of relative delays for the 76 projects between 1990 and 2019 on the NCS

Figure 5 clearly depicts the trends in development delays based on the PDO approval years. The graph highlights that the highest development delays occurred for projects approved in the year 2005, with an average delay of approximately 57% compared to their forecasted production start times. On the other hand, projects approved in 2015 showed an average earlier production start of around 14% compared to their forecasted times. Notably, 2015 and 2017 were the only years where production started earlier than the forecasted times.

It is worth mentioning that there was no available data on development times for the years 1993, 2003, and 2016. Additionally, it is interesting to note that projects approved between 1990 and 1995, those between 1997 and 2000, as well as projects in 2002, 2008, and 2019, experienced no production start delays.

The analysis of schedule overruns in development times reveals interesting trends within specific time intervals. From 1995 to 2005, there was a general increase in schedule overruns, with two notable peaks in 1996 (at 30%) and 2001 (at 14%). This indicates that during this period, projects experienced significant delays compared to their forecasted development times.

However, between 2005 and 2008, there was a sharp reduction of 57% in schedule delays, indicating a period of improved performance in meeting development time targets. This was followed by a sudden increase of 43% in schedule delays in 2009, suggesting a temporary setback.

From 2010 to 2019, there was an overall downward trend in schedule overruns, indicating a general improvement in meeting development time targets. However, there were two sudden increases in schedule delays, observed in 2014 (at 50%) and 2018 (at 14%). These instances highlight temporary deviations from the overall trend of reduction in schedule overruns.

Figure 6 presents the trend of development delays for the 76 oil fields over the study period. The blue line represents the percentage of development delays in days, while the grey and orange lines represent the 10-year and 5-year Simple Moving Averages (SMA) of the delays, respectively.

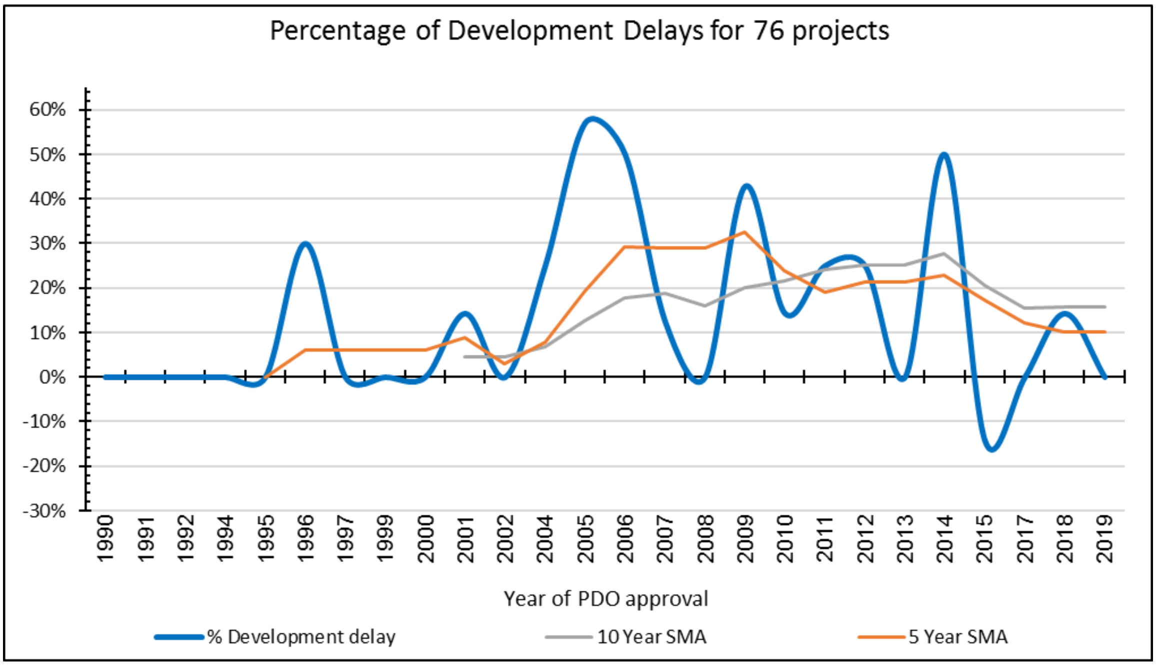

Figure 6 - The trend of schedule overrun percentage and the simple moving average lines for 76 projects

The analysis highlights that the highest percentage of schedule overruns occurred in 2005, reaching 57%. On the other hand, projects approved in 2015 started production on average 14% earlier than their forecasted times, indicating better performance in meeting development time targets.

However, when examining the SMA lines for the 5-year and 10-year periods, no clear signs of improvement in the forecasts of development times are evident. The lines demonstrate fluctuations and no consistent downward or upward trend, suggesting that the industry has struggled to improve their forecasts on production start times over the almost 30-year period analysed.

The results obtained from the analysis of development delays are utilized in the calculation of value loss.

4.3 Analysis of the production data

4.3.1 Forecasted Production Profiles

According to the guidelines provided by the MPE [40], the forecasts of production values included in the operators' PDO must consist of three values: a low forecast, a mean forecast, and a high forecast, representing p10, mean, and p90 values, respectively. However, for the purpose of this study, only the mean values were investigated.

When the mean forecasted production values were sorted based on the number of years after the forecasted production start, clear forecasted production profiles emerged. As argued by Demirmen [32], [33], the forecasts of production values tend to fluctuate throughout the lifecycle of projects, but an average increase of approximately 16% is observed in the forecasts. The mean forecasted production data used in this study also confirm these trends.

Figure 7 clearly depicts the mean forecasted production values for certain fields, represented by blue columns, along with the percentage changes in their mean forecasts, indicated by the orange line on the secondary axis. In general, the majority of fields in the dataset exhibited similar patterns to field A, B, C, or D, where their peak forecasts occurred in the 2nd, 3rd, 4th, or 5th year, respectively. However, the primary objective here was to demonstrate that the fluctuations in the forecasts were evident throughout the lifetime of each field.

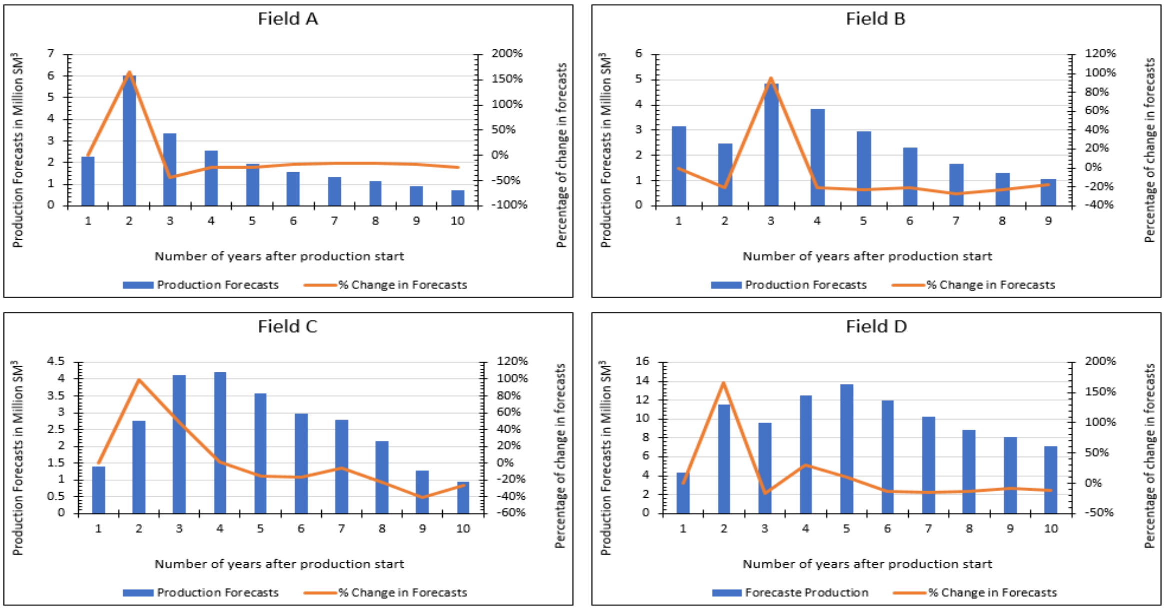

Figure 7 - The fluctuations of the Forecasted Production values for each field

Figure 8 displays the sorted total mean forecasts for the 67 oil fields with production between 1995 to 2021, organized according to the number of years after production start. The blue bars represent the forecasts, while the orange line indicates the proportion of change in these forecasts. The figure clearly shows that there was a substantial increase in the total forecasts during the second year, which resulted in an overall trend indicating a 17% growth in forecasts over the first 10 years.

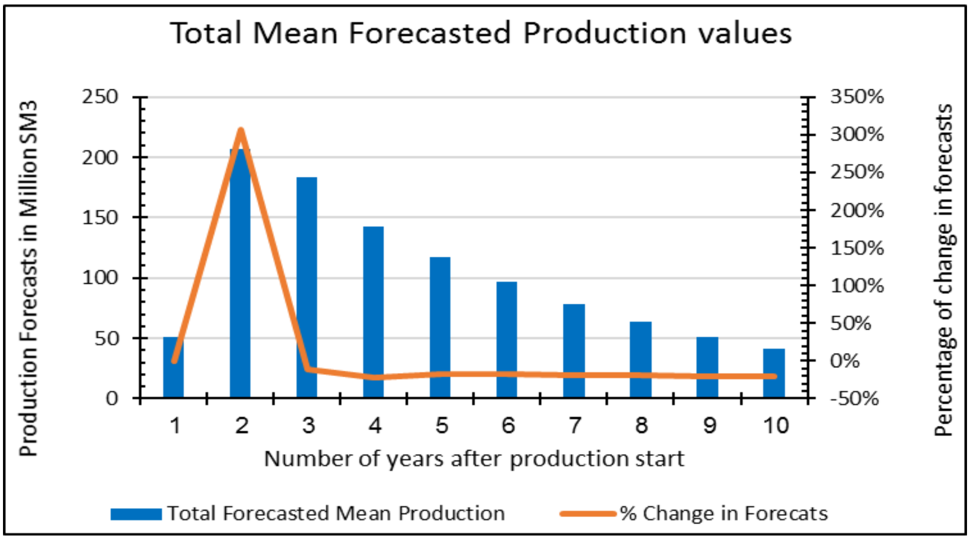

Figure 8 - The fluctuations of the Forecasted Production values for 67 fieldS

4.3.2 Production Forecasts of the 67 Fields

In the context of this study, it is important to note that the forecasts of production values provided by the operators on the NCS are updated annually throughout the entire lifetime of the fields. However, for this specific study, only the forecasts of production values in the first year of PDO approval were taken into consideration. The primary focus of this study was to assess the underperformance in production, and as such, the forecasted values were compared with the actual values for the 67 oil fields with production spanning from 1995 to 2021. This comparison was carried out specifically for the first 10 years of each field.

Figure 9 illustrates the total mean forecasted production values for the 67 oil fields, sorted by their forecasted production years. The forecasted start of production was in 1995, with the lowest total forecasts at that time. The forecasts experienced a consistent increase until 2001, which marked the year with the highest total forecasts. Subsequently, there was an overall decrease in the forecasts up until 2019. However, in the following two years, there was a subsequent increase in the forecasts once again.

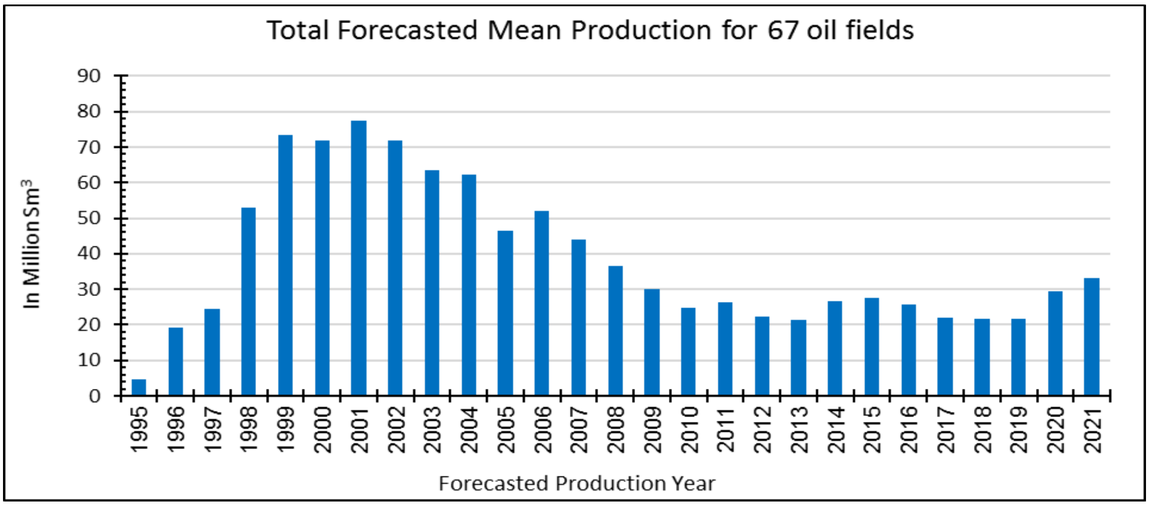

Figure 9 - The annual forecasted mean production values for the 67 fields between 1995 and 2021

4.3.3 Actual Productions of the 67 oil fields

The actual production values for the fields operating on the NCS are publicly available on the NPD's website [56]. However, for the purpose of this study, only the fields with available forecasted values were considered in order to facilitate meaningful comparisons. As a result, a total of 67 fields with production data between 1995 and 2021 were included in the analysis.

Figure 10 depicts the total actual production values for the 67 oil fields considered in this study, sorted by their actual production years. Observing the graph, it can be observed that the highest total production occurred in the year 2001, while the lowest production was recorded in 1995.

Between 1995 and 2001, there was a gradual increase in production values over a period of six years. However, in the subsequent 12-year period until 2013, an overall decreasing trend can be observed, with two peaks in 2004 and 2006.

From 2014 to 2021, a steady growth in production values is evident, punctuated by two drops in 2018 and 2019.

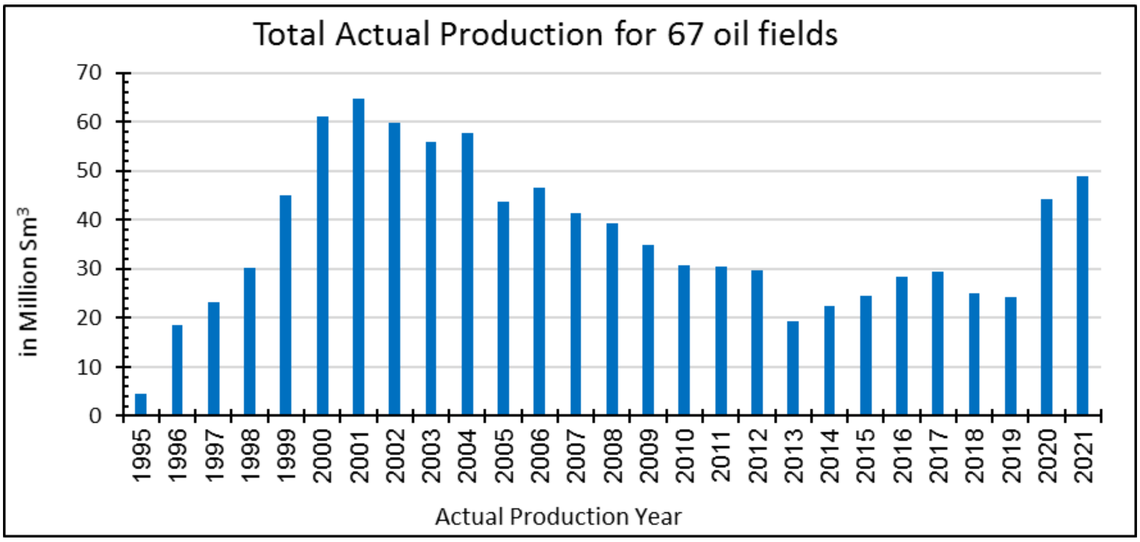

Figure 10 - The annual actual production values for the 67 fields between 1995 and 2021 [56]

4.3.4 Comparing Actual and Forecasted Productions

Previously, the study examined overruns in development time and costs. It was observed that while the forecasts of time did not show any improvement over the study period, the forecasted development costs exhibited a slight improvement. Building upon this analysis, the focus shifted to studying the forecasted production values and actual production values together using Method 2, which does not consider the effect of delays. This approach aimed to determine the trend and magnitude of underproduction.

When the data for the first ten years of production was utilized and the forecasted values of delayed projects were shifted to align with the actual production starts (Method 2), the results were depicted in Figure 11. For example, if a field had a forecasted production start in 2018 but experienced a delay of two years, resulting in the actual production start in 2020, the forecasted production values for 2018 and onwards were shifted to 2020 and onwards. Subsequently, all the forecasted and actual values from 2018 onwards were compared for each year.

Figure 11 presents a visual representation of the actual production values and underproduction for the 67 oil fields with production spanning from 1995 to 2021. In this graph, the actual production values are depicted by blue columns, while the underproduction values are represented by orange columns. The production values are sorted based on the number of years after the production start.

Additionally, the graph includes a gray line on the secondary axis, which represents the percentage of underproduction. This line provides further insight into the extent of underproduction relative to the actual production values.

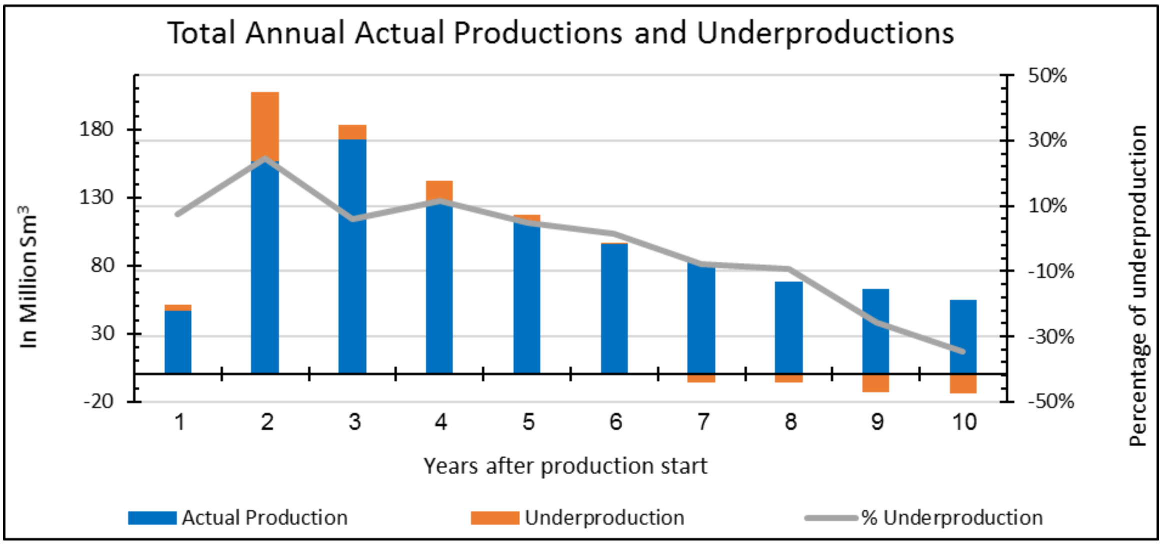

Figure 11 - The annual actual production values and underproductions for the 67 oil fields

As is evident from Figure 11, the highest and lowest yearly underproductions were a result of the forecasts in the 2nd and 10th year of production, amounting to 24% and -35% respectively. Following a 17% increase in underproduction from the 1st to the 2nd year, an overall declining trend persisted until the 10th year. However, a slight 6% increase in underproduction is noticeable between the 3rd and 4th years. Notably, after reaching an underproduction level of nearly zero (1%) in the 6th year, no underproduction occurred between the 7th and 10th year following production initiation. In fact, the annual actual productions consistently exceeded the forecasted quantities as the years progressed.

For a more comprehensive understanding, Figure 12 presents the cumulative actual production and underproduction values. Given that Figure 11 illustrated a general decrease in yearly underproduction as fields mature, Figure 12 highlights the impact of significant underproduction experienced in the second year. Consequently, even after a decade of production, there remained unmet promised production volumes. This indicates that the substantial underproduction in the early stages of production had a lasting impact that persisted over the years.

Figure 12 clearly demonstrates the progression of cumulative underproduction over time. Starting with 7% underproduction in the 1st year, it significantly increases to 21% in the 2nd year before gradually declining. By the end of 10 years of production, the cumulative underproduction reaches 5%.

By utilizing Method 2, it can be deduced that the total of forecasted mean productions was approximately 1028 million Sm^3, while the total of actual production values amounted to around 978 million Sm^3. Consequently, there was an unfulfilled production amount of approximately 50 million Sm^3 of oil across the 67 fields during the initial 10 years of production.

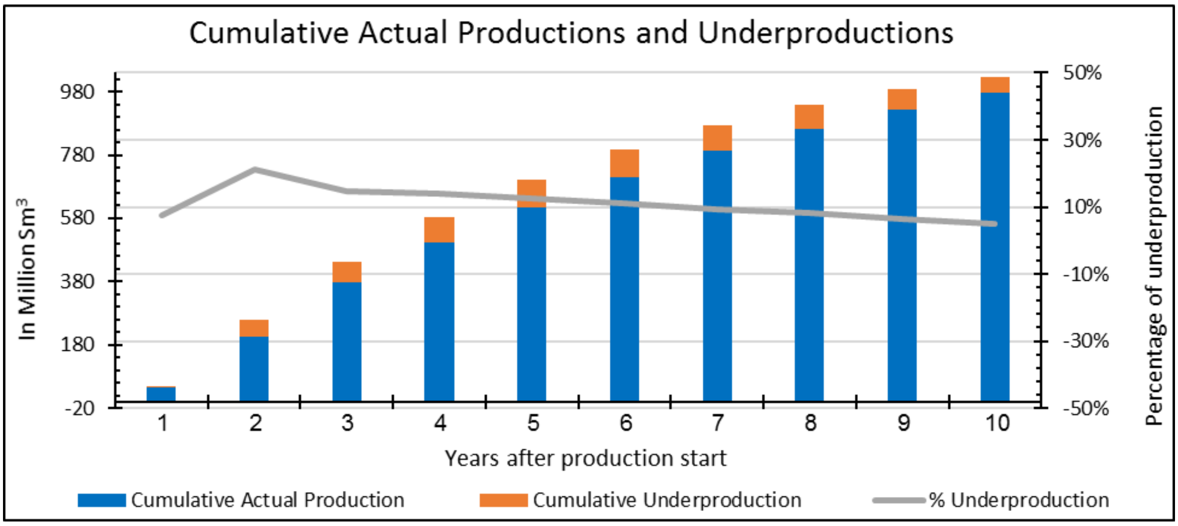

Figure 12 - The cumulative actual production values and underproductions for the 67 oil fields

The trend of underproduction percentages throughout the study period is illustrated in Figure 13. The Simple Moving Average (SMA) was utilized to examine how the production value forecasts changed between 1995 and 2021.

Figure 13 displays the percentage of underproductions, represented by the blue line, along with the 10-year period SMA (gray line) and the 5-year period SMA (orange line). The values are sorted based on production years along the horizontal axis.

The blue line exhibits fluctuations throughout the period, peaking at approximately 36% in 1998 and reaching its lowest points in 2012 and 2021, at nearly -40%. In contrast, the SMA lines for the 5-year and 10-year periods indicate a slight decrease in underproductions, suggesting an improvement in the forecasted production values when considering the initial 10 years of production.

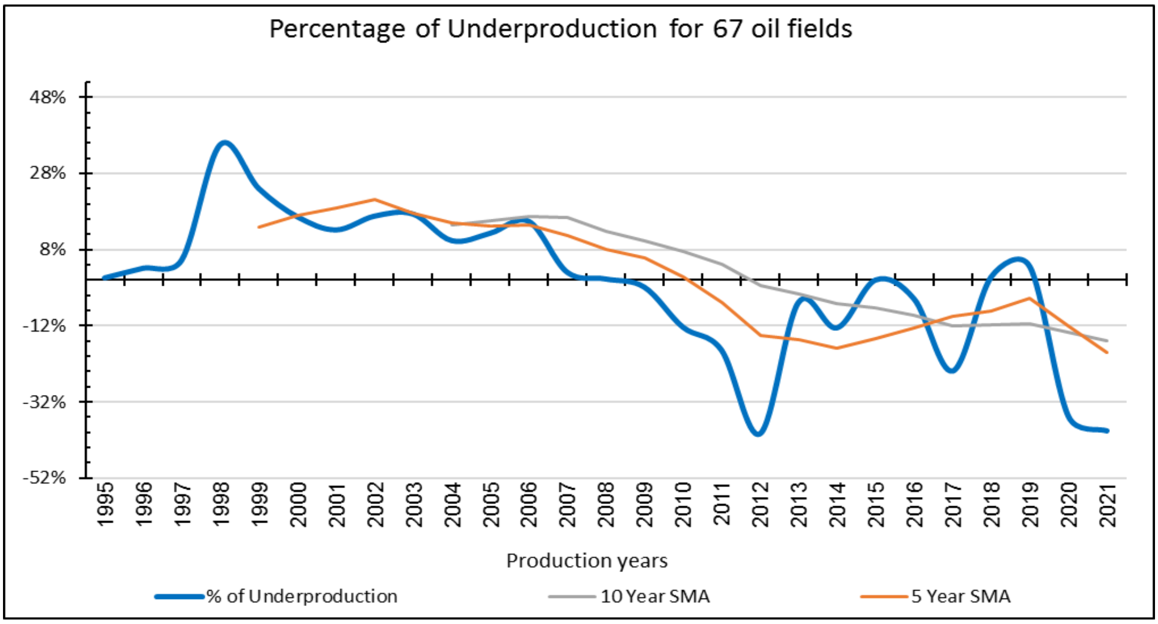

Figure 13 - The trend of underproduction percentage and the simple moving average lines for 67 oil fields

4.4 Analysis of Economic Value Loss

In this section, the economic consequences of underproduction and delays are analysed. The production profiles, comparing the forecasted and actual productions using the two methods described earlier in Chapter 2, are presented and discussed.

4.4.1 Production profiles when Shifting to Forecasted Production Start

When utilizing Method 1 to shift the start of delayed projects to the forecasted years and assuming zero actual production values during the delayed period, the forecasted and actual production values were obtained as shown in Figure 14. It is important to note that only the mean forecasts for the 67 oil fields with production between 1995 and 2021 were taken into account.

Figure 14 illustrates the total annual quantities of the actual and mean forecasted productions, depicted by the blue and orange lines respectively. The data is arranged based on the number of years after the PDO approval.

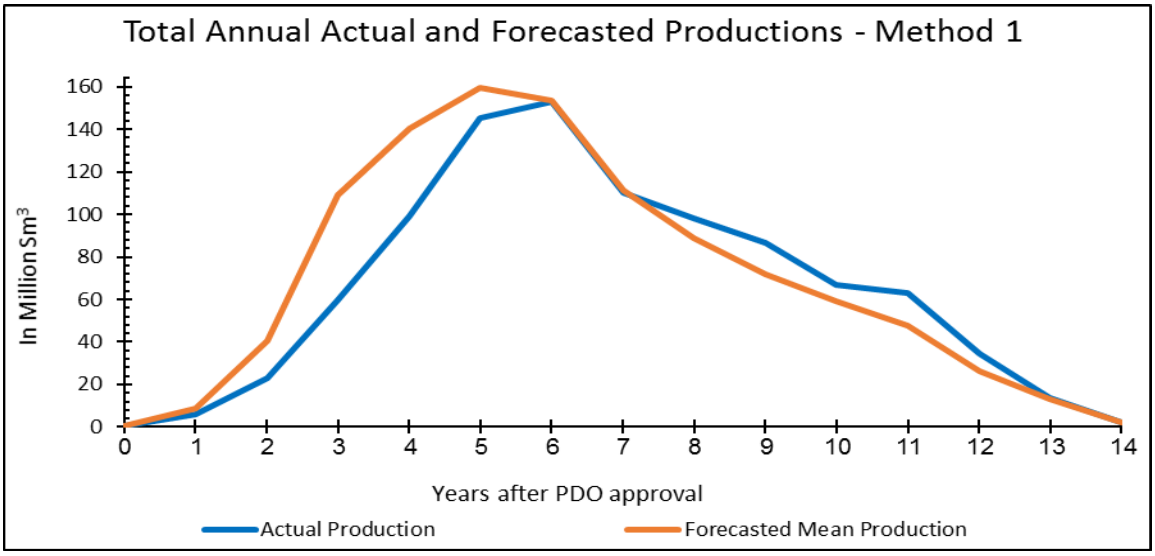

Figure 14 – Annual actual and forecasted production profiles for the 67 oil fields – Method 1

From Figure 14, it is evident that for fields with production occurring within less than 6 years after their PDO approval, both actual and forecasted values exhibited a significant increase. However, throughout the entire period, the annual actual production values remained lower than the mean forecasted production values.

For fields with production spanning between 6 and 7 years after their PDO approval, the actual and forecasted values were very close and both experienced a decline.

Projects with production occurring between 7 and 13 years after their PDO approvals demonstrated a decline in both actual and forecasted productions. However, during this period, it is noticeable that the actual production values were higher than the forecasted values for almost all projects.

Lastly, projects with production ranging from 13 to 14 years had actual production values that closely matched their forecasted values.

Figure 14 highlights that the highest actual and forecasted production values were observed 5 years after PDO approval. Additionally, the maximum deviation between the annual forecasted and actual production values occurred 3 years after PDO approval, reaching approximately 49 million Sm^3.

The cumulative production profiles for the 67 oil fields, utilizing Method 1, are depicted in Figure 15. It is evident that throughout the entire period, the cumulative actual productions remained lower than the cumulative forecasted productions.

The projects that extended their production up to 7 years after PDO approval accounted for the largest cumulative underproduction, totalling 127 million Sm^3.

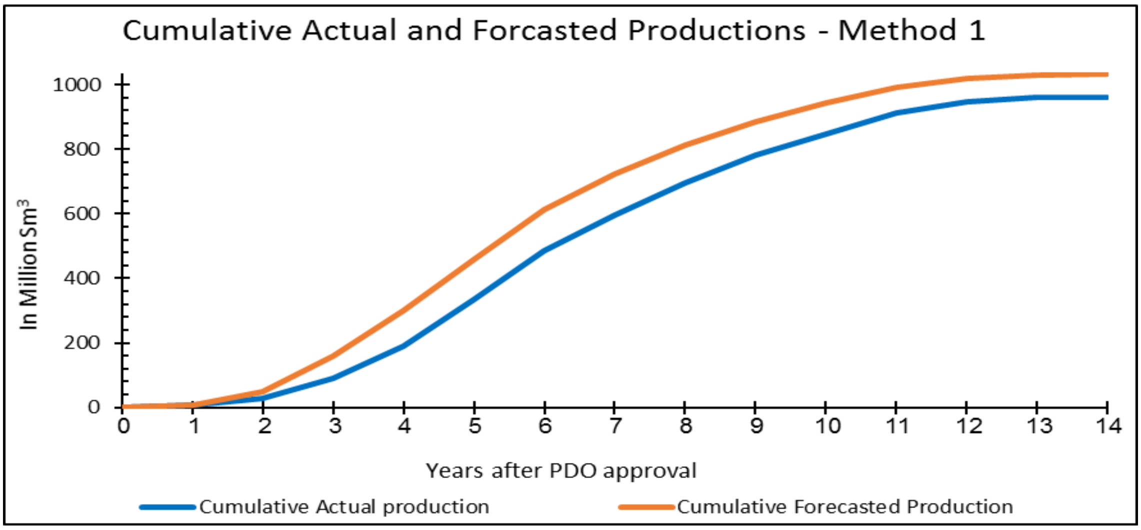

Figure 15 - Cumulative actual and forecasted production profiles for the 67 oil fields – Method 1

4.4.2 Production profiles when Shifting to Actual Production Start

Method 2, as described, involves shifting the forecasted production starts back to the actual production starts for projects with delays, thereby reducing the impact of delays on underproduction. When applying Method 2 to the 67 oil fields and comparing their forecasted and actual production profiles, Figure 16 is generated.

Figure 16 displays the annual actual and forecasted production values, similar to Figure 12, but with the data sorted according to Method 2. It can be observed that between 0 to

7 years after PDO approval, the actual productions were lower than the forecasted productions, although both exhibited an increase. However, from 6 years after PDO approval and then from 8 to 14 years after PDO approval, the actual productions surpassed the forecasted productions.

The highest forecasted production value was observed 5 years after PDO approval, while the highest actual production value occurred 6 years after PDO approval. Both values demonstrated an increase from PDO approval until these respective years, followed by a gradual decline until the 10th year.

Based on the results, when mitigating the impact of delays on underproduction using Method 2, the largest annual underproduction occurred 4 years after PDO approval, amounting to approximately 34 million Sm^3.

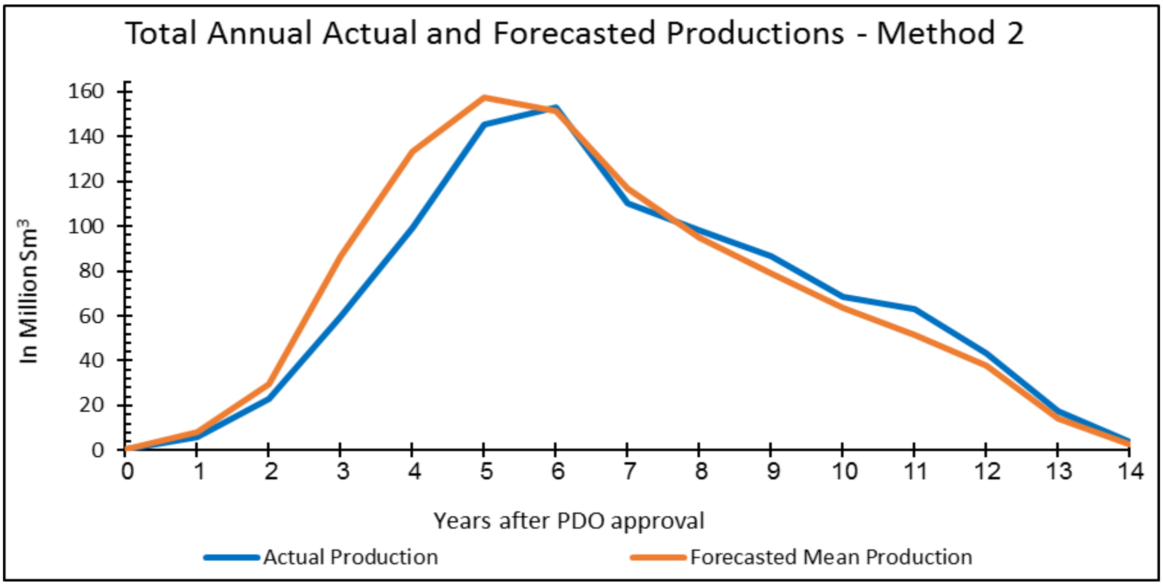

Figure 16 - Annual actual and forecasted production profiles for the 67 oil fields – Method 2

Figure 17 provides an illustration of the cumulative production profiles when utilizing Method 2. This figure displays the cumulative profiles, similar to Figure 15, but sorted according to the logic of Method 2. It is worth noting that the highest cumulative underproduction was associated with projects having production at 7 years after PDO approval, resulting in a cumulative underproduction of approximately 87 million Sm3.

By comparing the area between the cumulative actual and forecasted curves, it becomes evident that Method 2 reduces the effect of delays on underproduction compared to Method 1.

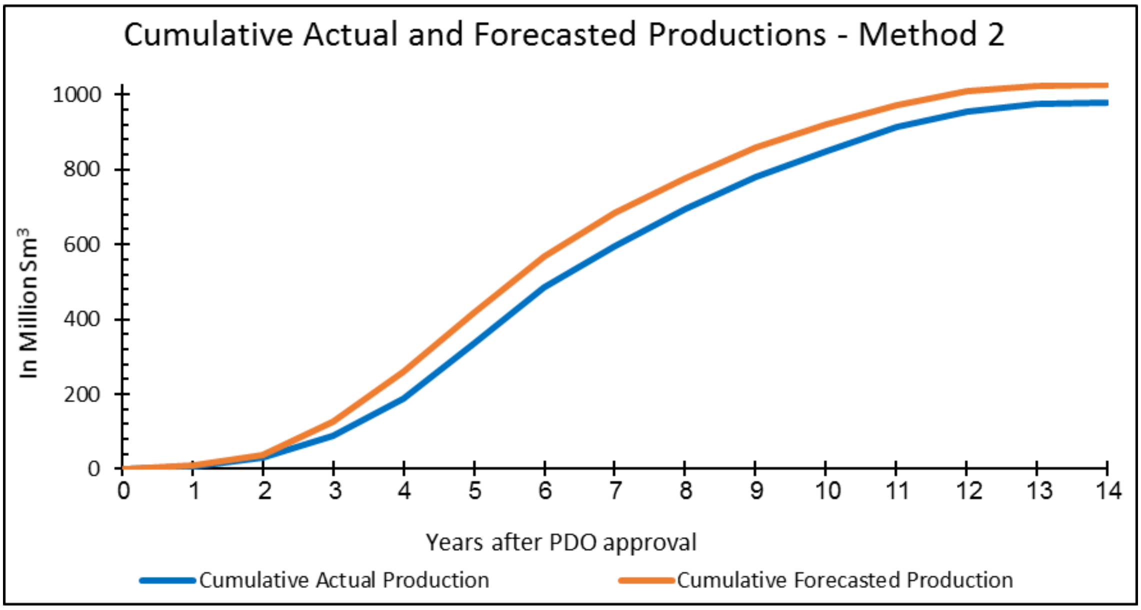

Figure 17 - Cumulative actual and forecasted production profiles for the 67 oil fields – Method 2

4.4.3 Economic value losses

Considering an economic perspective, the production values generate positive cash flows for the projects. Although the cumulative values of the forecasted and actual productions, as demonstrated in Figures 15 and 17, were very close to each other, the value generated from selling productions in later years is reduced.

Using the two methods of sorting the data, the total present values were calculated. Method 1 accounted for underproductions and delays, while Method 2 considered underproductions without the effect of delays. By subtracting the present value of Method 2 from that of Method 1, the value loss due to delays was determined.

Figure 18 visually presents the actual and lost production revenues of the 67 oil fields with production between 1995 and 2021, when using Method 1 for sorting. The blue and orange columns represent the total actual production revenues and revenue losses, respectively, while the grey line indicates the percentage of production revenue loss. The data is sorted based on the year of PDO approval.

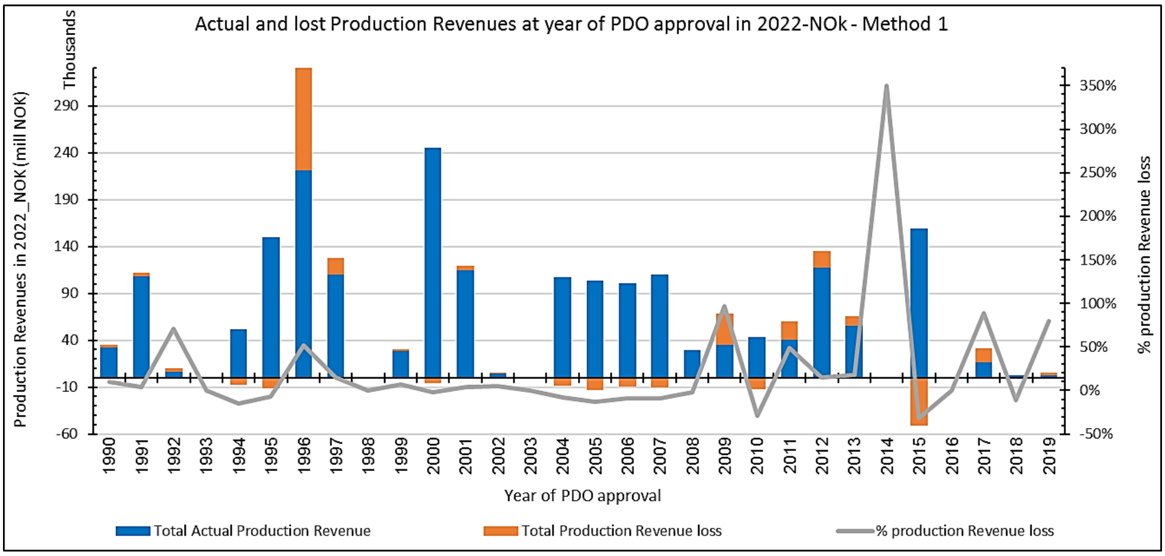

Figure 18 – Total actual and Lost production Revenues in 2022-NOK – Method 1

Figure 18 highlights that the highest amount of production revenue loss occurred in 1996, amounting to NOK 114 billion. Additionally, projects approved in 2014 experienced the largest percentage of revenue losses, where the forecasted cash flows were approximately 3.5 times higher than the actual cash flows.

On the other hand, projects approved in 2015 demonstrated the highest increase in actual production revenues compared to the forecasts. The actual production revenues in 2015 amounted to NOK 51 billion, which was 32% higher than the forecasted production revenues.

The fluctuating trend of production revenues in Figure 18 reflects the monetary values that were lost and earned during these years, providing insights into the variations in cash flows resulting from underproduction and delays.

According to calculations using Method 1, the total present value of actual productions amounted to approximately NOK 1996 billion, while the total present value of forecasted values was nearly NOK 2114 billion. Consequently, by subtracting these present values, the value loss due to underproduction and delays was computed to be approximately NOK 118 billion.

Figure 19 represents the actual and lost production revenues for the same 67 oil fields when Method 2 is employed to sort them. This figure showcases the cash flows when only the effect of underproduction is taken into account. The total actual production revenues and revenue losses are presented in the blue and orange columns, respectively. The percentage of production revenue loss is represented by a grey line, displayed on the secondary axis. The data is sorted based on the year of PDO approval.

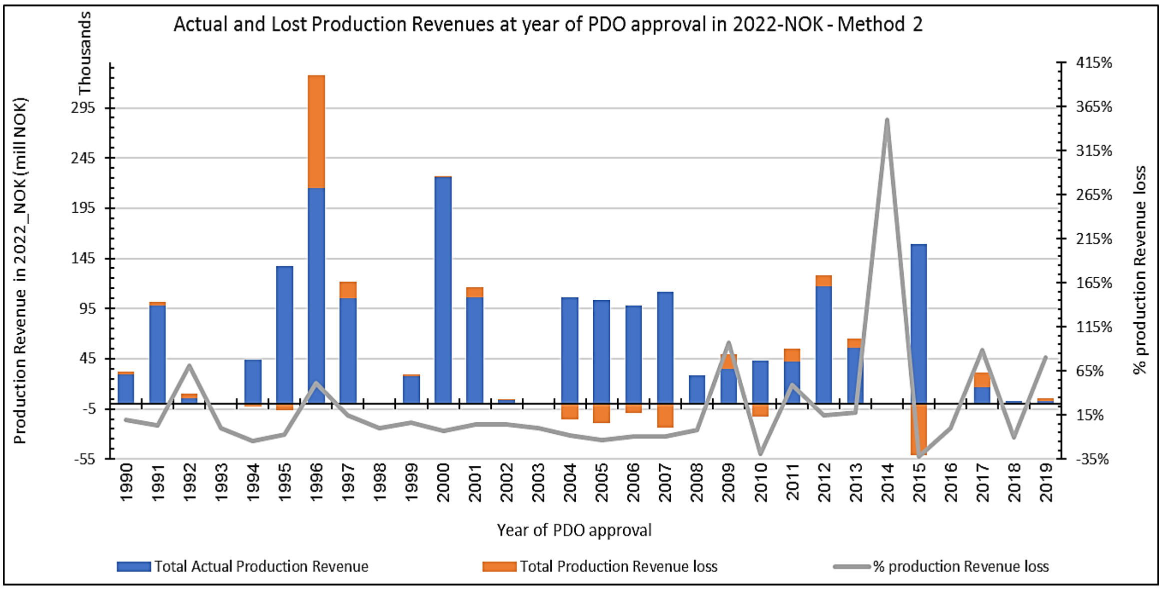

Figure 19 - Total actual and Lost production Revenues in 2022-NOK – Method 2

As depicted in Figure 19, the majority of revenue losses due to underproduction occurred in 2014, where the actual revenues amounted to only about one-third of the forecasted revenues. The highest amount of production revenue losses, totalling NOK 112 billion, was observed in 1996. On the other hand, the year 2015 had the highest percentage of actual revenues exceeding the forecasted revenues, reaching 32%.

Based on the results obtained from the present value calculations using Method 2, the total present value of actual production revenues was approximately NOK 1921 billion, while the total present value of forecasted cash flows was around NOK 2000 billion. By subtracting the present value of forecasted cash flows from the present value of actual production revenues, the value loss attributed solely to underproductions amounted to approximately NOK 80 billion.

Earlier in this study, it was shown that projects on the NCS were delayed by an average of 101 days, which accounted for approximately 12% of the total project duration. By subtracting the PV calculated using Method 2 from the PV calculated using Method 1, the value loss specifically attributed to delays was estimated to be NOK 39 billion.

Furthermore, the total cost overruns for 142 development projects with PDO approvals between 2001 and 2022 were previously computed as NOK 268 billion, representing a deviation of 12.7% over the budgeted costs. Taking into account the PV calculations from both Method 1 and Method 2, in addition to the value lost due to cost overruns, the overall value losses and deviations of actual values from the forecasts can be summarized in Table 3.

Table 3 – Summary of Value losses due to cost overruns, delays and underproduction

Category of Value Loss	In billion NOK	% Deviation of Actual from Forecasted Values
Cost Overruns	268.404	12.7%
Schedule Overruns	38.672	12%
Underproduction	79.574	5%
Total Present Value Loss	386. 65	

As shown in Table 3, the total value loss due to delays, cost overruns, and underproductions resulting from forecasts in PDOs between 1990 to 2022 amounted to approximately NOK 387 billion. This loss was comprised of NOK 268 billion attributed to cost overruns, NOK 38 billion attributed to schedule overruns, and a loss of production revenues of nearly NOK 80 billion. The last column of the table displays the percentage of deviations from the forecasted development costs, schedules, and production values, which were found to be 12.7%, 12%, and 5%, correspondingly.

Figure 20 represents the breakdown of value losses due to delays, cost overruns, and underproduction in red, blue, and orange, respectively, as a proportion of the total value loss. The values are presented in 2022-NOK billion. The figure reveals that 69% of the total value loss was attributed to poor forecasts of development costs, 21% to inaccurate time estimations, and 10% to underproduction. It is notable that the highest value losses were incurred due to cost overruns, while the lowest losses were associated with delays. However, it is important to note that delays can have indirect effects on the other two categories.

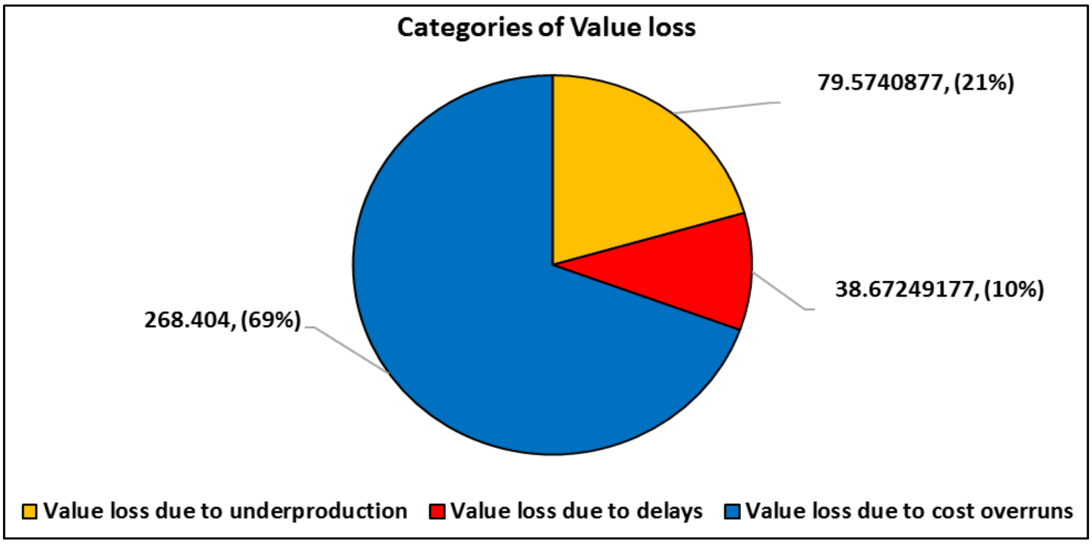

Figure20 – Portion of value losses due to cost overruns, delays and underproduction to the total value loss

4.5 Regression Analysis

In the final part of the analysis on the production data for the 67 oil fields with production between 1995 and 2021, the focus was on examining the correlation between cumulative underproductions in the first 10 years of production and total cost overruns.

However, for 16 fields in the dataset, including BYRDING, FRAM H-NORD, FRØY, GIMLE, GULLFAKS SØR, HEIDRUN, JOTUN, NJORD, NORNE, OSEBERG ØST, SINDRE, STATFJORD

NORD, VARG, VIGDIS, VISUND, and ØSGARD, there was no available information on their cost overruns. Therefore, these fields were excluded from the regression analysis.

As a result, the regression analysis focused on the remaining 51 oil fields. To ensure greater accuracy in the results, both positive and negative values for underproduction and cost overruns were included in the analysis. This means that fields with production equal to or greater than the forecasts, or costs equal to or less than the forecasts, were also included.

In the regression analysis, when there was no production data available for certain years, it was assumed that there was no cumulative production for those years, rather than using the same cumulative production as the previous years.

Considering that investment costs (CAPEX) typically occur earlier than productions in most development projects, the analysis investigated the relationship between cumulative underproductions (y-axis) and total cost overruns (x-axis). Three methods, namely Pearson, Spearman, and SVR, were used for this purpose, and their results are described in the following subsections:

4.5.1 Pearson Regression Analysis

In order to investigate a potential linear relationship with a constant rate between the two variables, the Pearson regression analysis was employed. The results of this analysis, including correlation coefficients (R), R-squared (R^2) values, and scatter plots, are summarized in Table 4. The calculations and visualizations were performed using Excel software.

Table 4 – Summary of correlation coefficients (R) and the R-squared (R^2) values with Pearson Regression Method

	Cumulative Underproduction in X years (Million Sm^3) VS. Total Cost Overruns in Million NOK								
X =	2 yrs.	3 yrs.	4 yrs.	5 yrs.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.
R	0.3375	0.2937	0.4483	0.4411	0.3638	0.3257	0.2516	0.1940	-0.1806
R^2	0.1139	0.0863	0.2010	0.1945	0.1324	0.1061	0.0633	0.0376	0.0326

As observed in Table 4, the highest correlation coefficient (R) and R-squared value were found between cumulative underproduction after 4 years and their corresponding total cost overruns, approximately at 0.45 and 0.2, respectively. On the other hand, the only negative correlation coefficient was associated with the data for fields with 10 years of data, with an R value of -0.18, which also had the lowest absolute value. The corresponding R-squared value for this data was 0.03, indicating a weak relationship.

Figure 21 illustrates the trend of correlation coefficients (Rs) and R-squared values (R^2s) as fields matured. The plot displays the linear correlations obtained using the Pearson method for the data related to the 51 fields with 2 to 10 years of production. It is evident that correlation coefficients decreased from 2 to 3 years, then experienced an increase from 3 to 4 years. Between 4 to 5 years, the coefficients remained relatively stable, but then declined from approximately 0.44 to -0.18. Similarly, R^2, representing the percentage of data that fitted the Pearson regression line, followed the same trend as Rs. It decreased from 11% to 9% for fields with 2 years to 3 years of production, then increased to 20% for fields with 4 years of production. However, for fields with 5 to 10 years of production, the percentage decreased from 19% to only 3%.

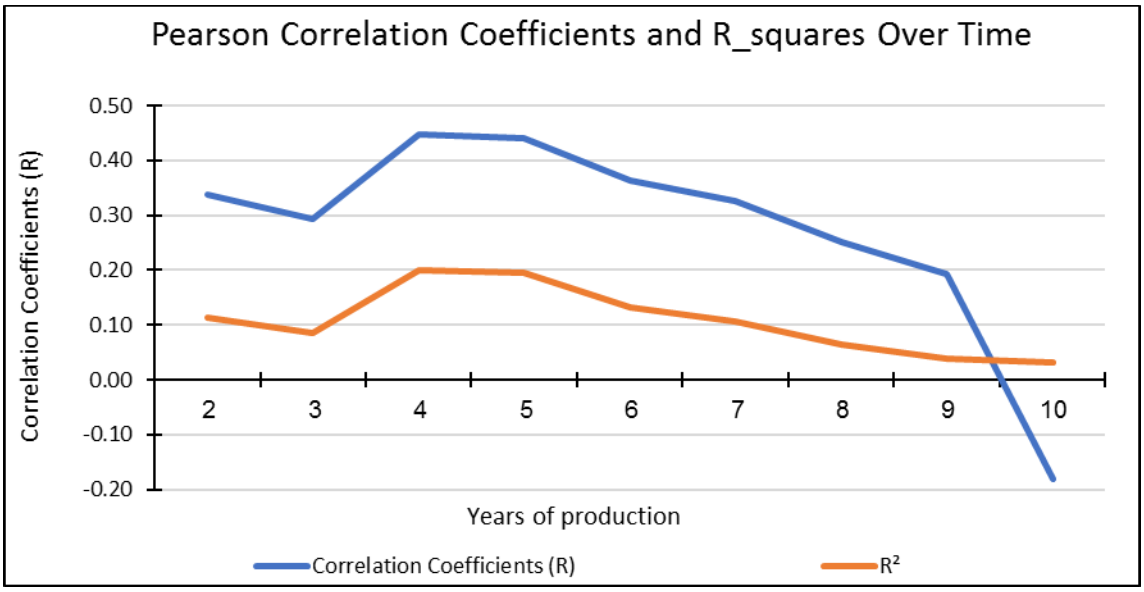

Figure 21 – Trend of Pearson correlation Coefficients and R_squares between the cumulative underproductions from 2 to 10 years and the total cost overruns for the 51 fields

Figures 22 to 30 display scatter plots depicting the relationship between cumulative underproductions in 2 to 10 years (vertical axis) and total cost overruns (horizontal axis). The scatter plots consist of blue dots representing the data points. Additionally, red dotted lines are included to represent the trendline or regression line that best fits the data.

The equation of the best-fit line, which represents the relationship between cumulative underproduction and total cost overruns, is provided in the figures. This equation describes the mathematical relationship between the two variables.

Furthermore, the value of R-squared (R^2) is shown in the figures. R-squared indicates the degree of dependence or correlation between the variations in cumulative underproduction and total cost overruns. A higher R^2 value indicates a stronger relationship between the variables.

Please refer to Figures 22 to 30 for a visual representation of the scatter plots and the corresponding trendlines, as well as the equation of the best-fit line and the R-squared value.

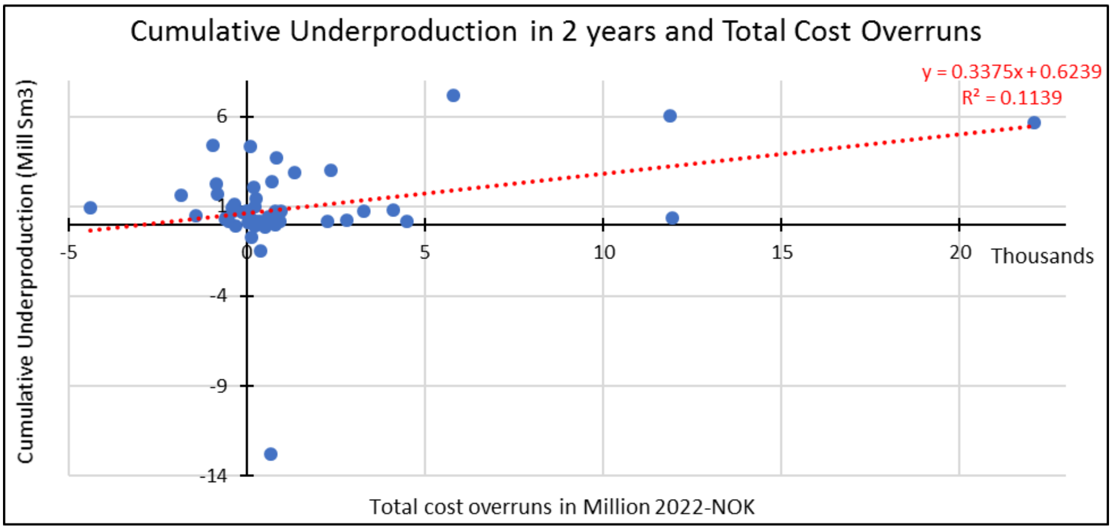

Figure 22 – Pearson scatterplot for cumulative underproduction in 2 years versus the total cost overruns

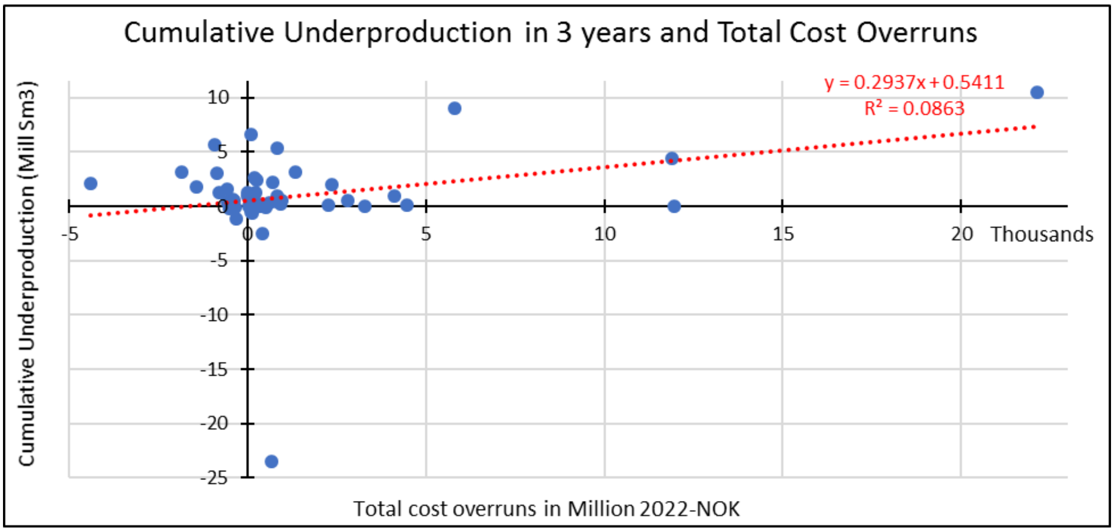

Figure 23 - Pearson scatterplot for cumulative underproduction in 3 years versus the total cost overruns

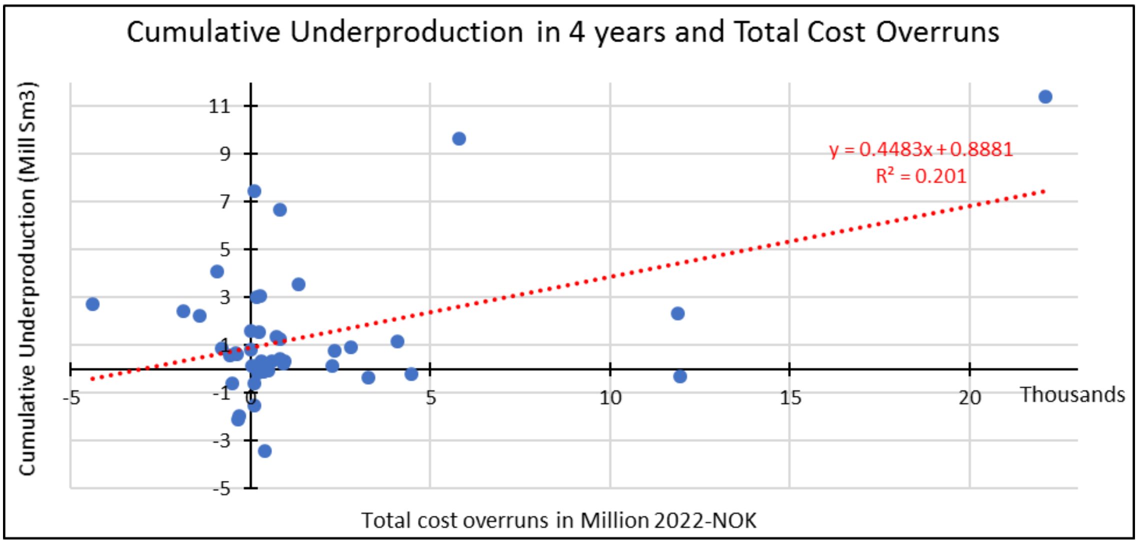

Figure 24 - Pearson scatterplot for cumulative underproduction in 4 years versus the total cost overruns

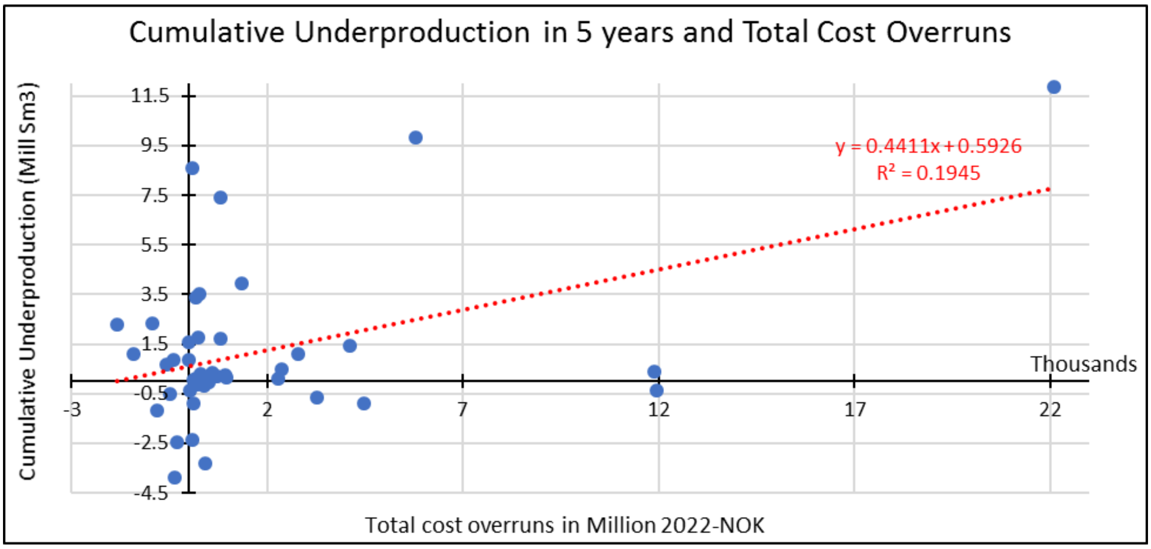

Figure 25 - Pearson scatterplot for cumulative underproduction in 5 years versus the total cost overruns

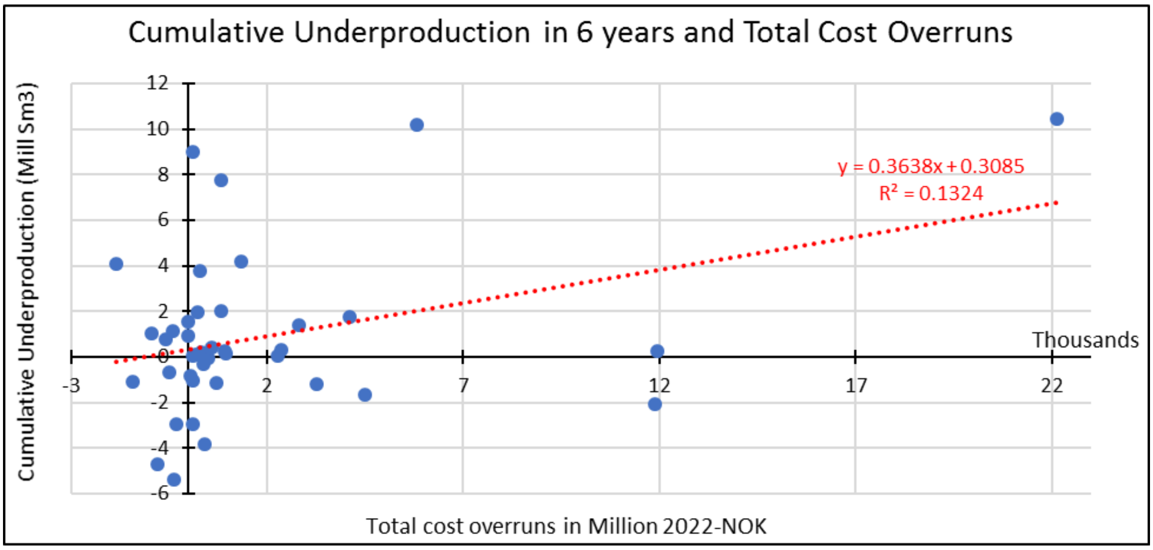

Figure 26 - Pearson scatterplot for cumulative underproduction in 6 years versus the total cost overruns

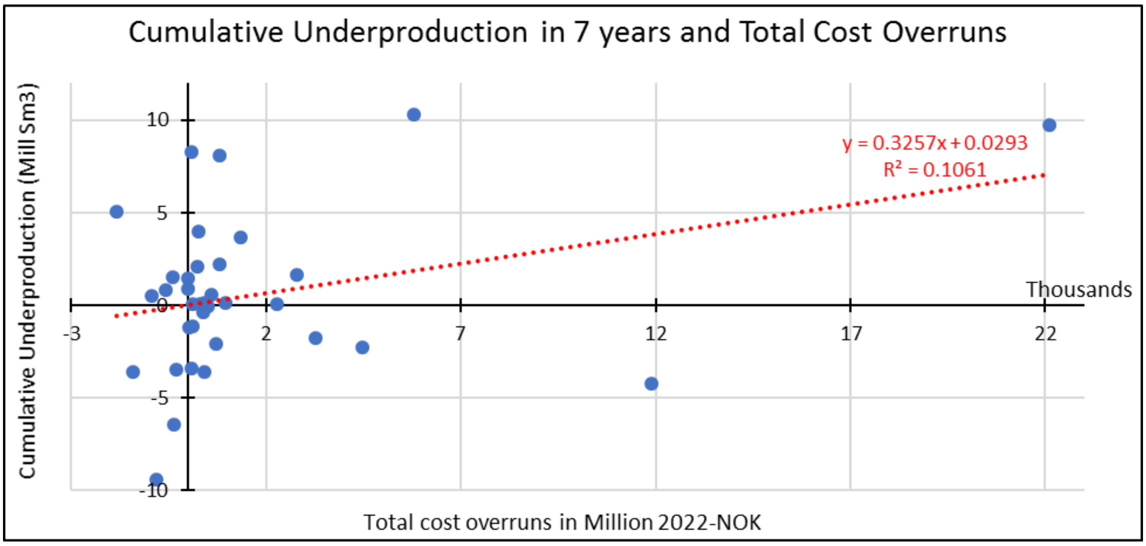

Figure 27 - Pearson scatterplot for cumulative underproduction in 7 years versus the total cost overruns

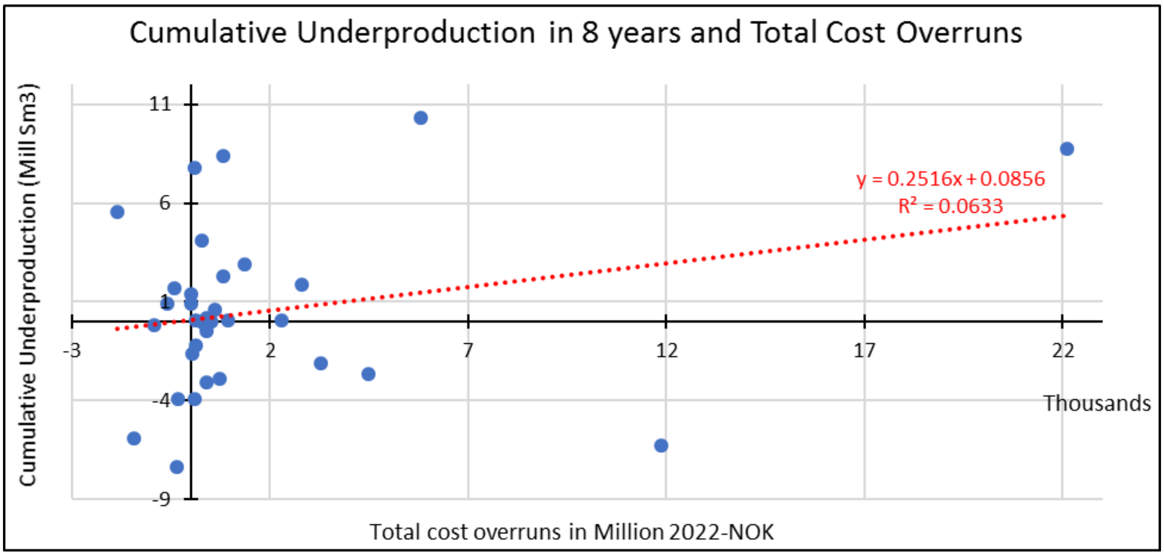

Figure 28 - Pearson scatterplot for cumulative underproduction in 8 years versus the total cost overruns

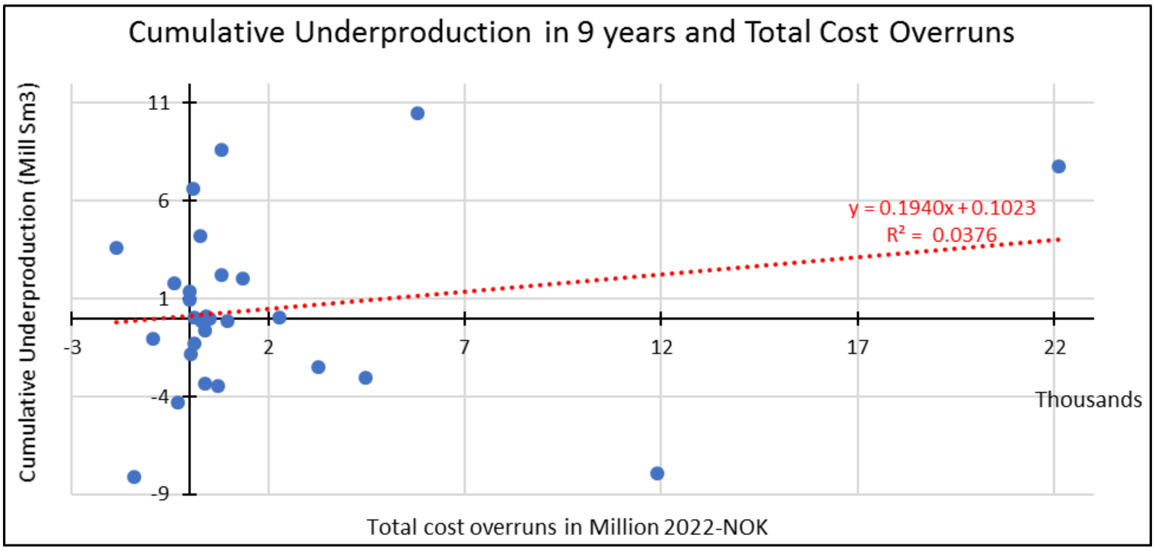

Figure 29 - Pearson scatterplot for cumulative underproduction in 9 years versus the total cost overruns

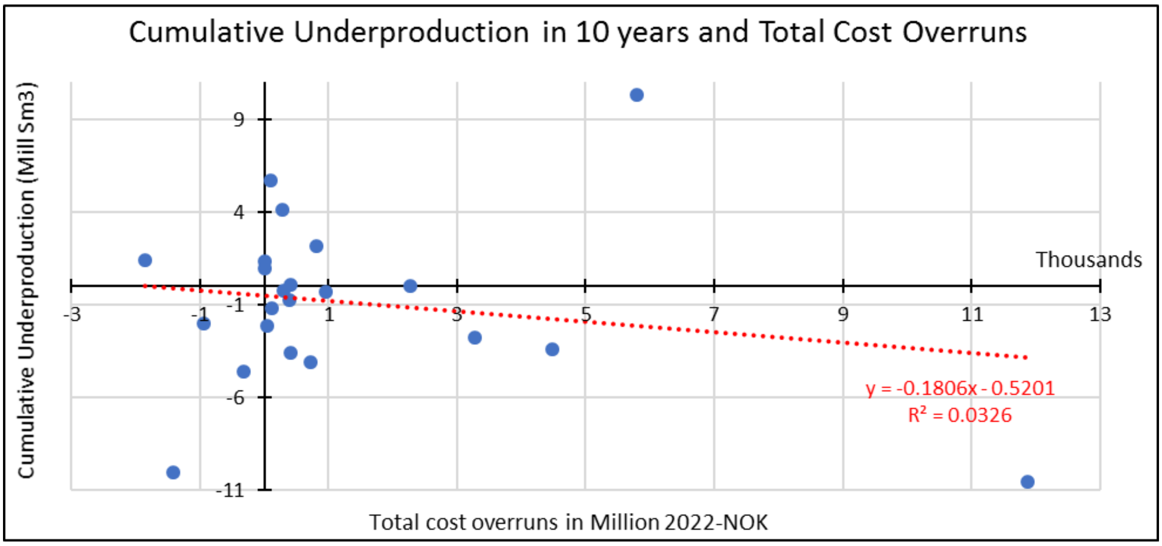

Figure 30 - Pearson scatterplot for cumulative underproduction in 10 years versus the total cost overruns

From Figures 22 to 30, it is evident that the data points are scattered around the regression line. This in addition to the correlation coefficients (R) and R-squared (R2) values indicate a weak linear relationship between the cumulative underproductions in the first 10 years and the total cost overruns using the Pearson method. The dispersion of the data points suggests that there are other factors and variables influencing the relationship between cumulative underproductions and cost overruns, and a simple linear model may not adequately capture the complexity of this relationship.

4.5.2 Spearman Regression Analysis

In order to explore any monotonic relationship between the rankings of the cumulative underproductions and the total cost overruns, the Spearman rank-based regression analysis was employed. This analysis was performed on the same set of 51 oil fields, with the same assumptions and criteria as mentioned in the previous section.

The Spearman rank-based regression analysis was performed using the Excel software and the RANK.AVG function to assign ranks to the data in ascending order. The correlation coefficients (R) between the ranked data were then calculated using the CORREL function. Additionally, the R-squared (R^2) values were computed by squaring the correlation coefficients.

The results of the Spearman regression analysis for the ranked cumulative underproduction and the rank of total cost overruns are summarized in Table 5. It can be observed that the highest correlation coefficient and R-squared value were obtained for the rank data of year 6, with values of approximately 0.18 and 0.03, respectively. Conversely, the ranks of the datasets for the 10 years exhibited a negative correlation coefficient of -0.05, which had the lowest absolute value among all the correlations. The ranks of the 3-year data showed the lowest correlation coefficient and R-squared value, with values of approximately 0.0045 and 0.00002, respectively.

Table 5 - Summary of correlation coefficients (R) and the R-squared (R^2) values with Spearman Regression Method

	Ranked Cumulative Underproduction in X years (Million Sm^3) VS. Ranked Total Cost Overruns in Million NOK								
X =	2 yrs.	3 yrs.	4 yrs.	5 yrs.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.
R	0.0206	0.0045	0.0313	0.1330	0.1811	0.1736	0.1625	0.0944	-0.0520
R^2	0.0004	0.00002	0.0010	0.0177	0.0328	0.0301	0.0264	0.0089	0.0027

Figure 31, depicting the changes in Spearman correlation coefficients and R-squared values as the fields matured, provides insights into the relationship between the rank of data. Initially, as fields progressed from 2 years to 3 years of production, the rank-based correlation coefficient decreased from 0.02 to 0.005. However, as the fields further matured, the correlation coefficient experienced an upward trend, reaching a peak of 0.18 for fields with 6 years of data.

Subsequently, the correlation between the ranks declined significantly, reaching a value of -0.05 for fields with ten years of production. These observations highlight the changing nature of the relationship between the ranks of cumulative underproduction and total cost overruns as the fields progress in their production lifecycle.

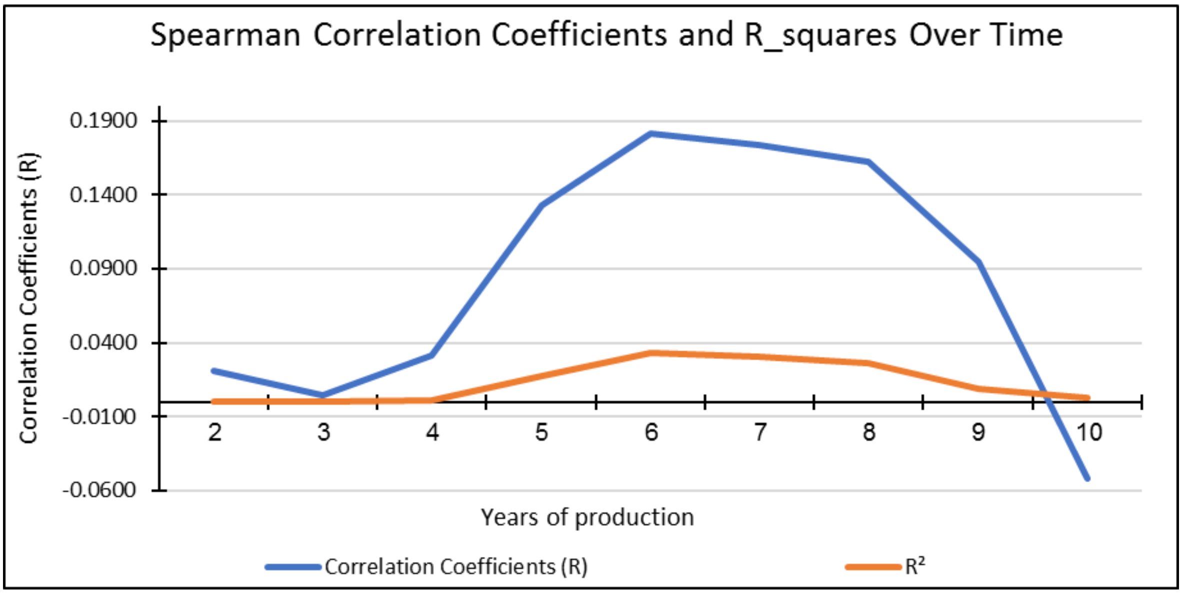

Figure 31 - Trend of Spearman correlation Coefficients and R_squares between the ranked cumulative underproductions from 2 to 10 years and the ranked total cost overruns for the 51 fields

Figures 32 to 40 present the scatterplots for the Spearman regression analysis, showcasing the relationship between the ranks of cumulative underproductions in the first 10 years and the ranks of total cost overruns. In each scatterplot, the ranks of the data are depicted as blue dots, while the regression line, its equation, and the R2 value are shown in red.

Upon observing the scatterplots and considering the correlation coefficients and R2 values, it becomes apparent that there is a very weak linear relationship between the variables. The low R2 values indicate that only a small percentage of the data fits the regression line. The highest R2 value, which reaches 3%, is observed in the ranks of 6 and 7 years. Overall, the dispersion of the ranks of the data suggests a lack of strong correlation between the ranked cumulative underproductions and the ranked total cost overruns in a linear fashion.

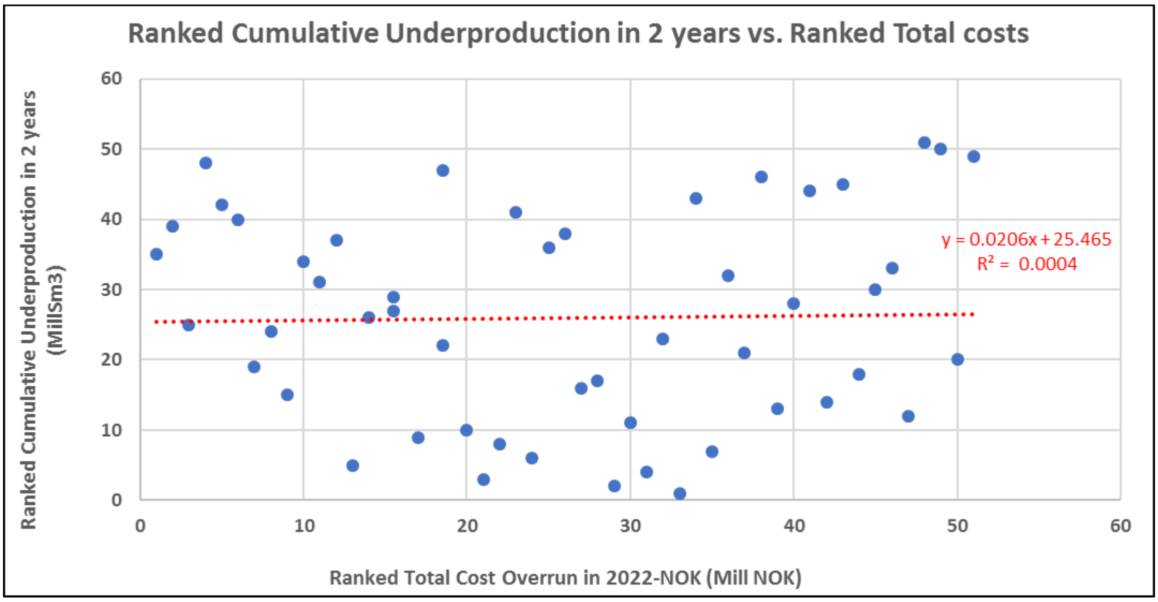

Figure 32 - Spearman scatterplot of the ranked cumulative underproduction in 2 years versus the ranked total cost overruns

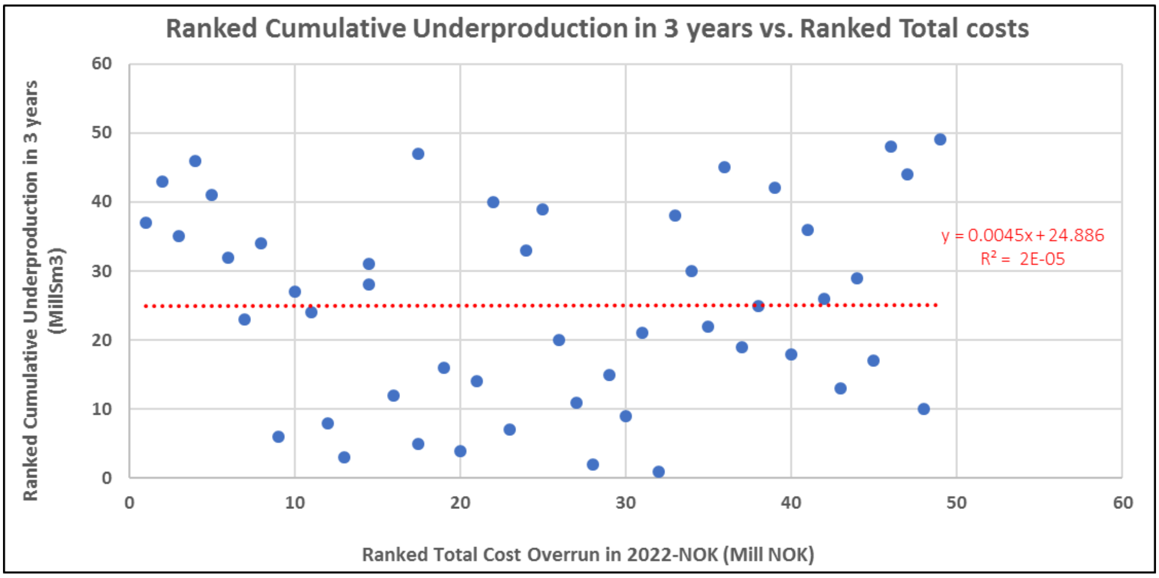

Figure 33 - Spearman scatterplot of the ranked cumulative underproduction in 3 years versus the ranked total cost overruns

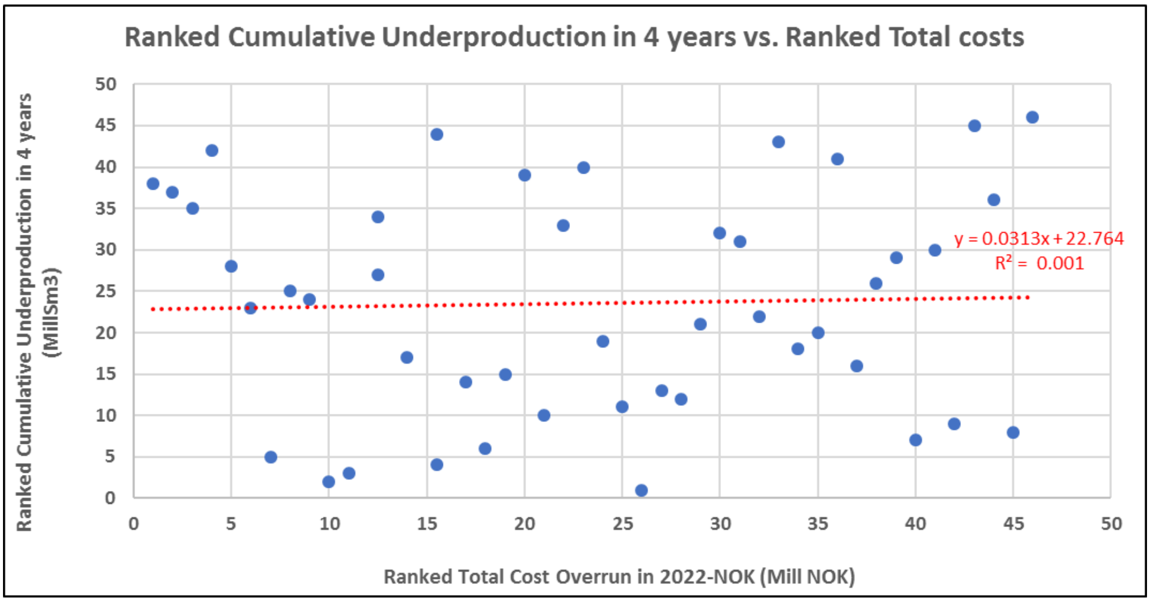

Figure 34 - Spearman scatterplot of the ranked cumulative underproduction in 4 years versus the ranked total cost overruns

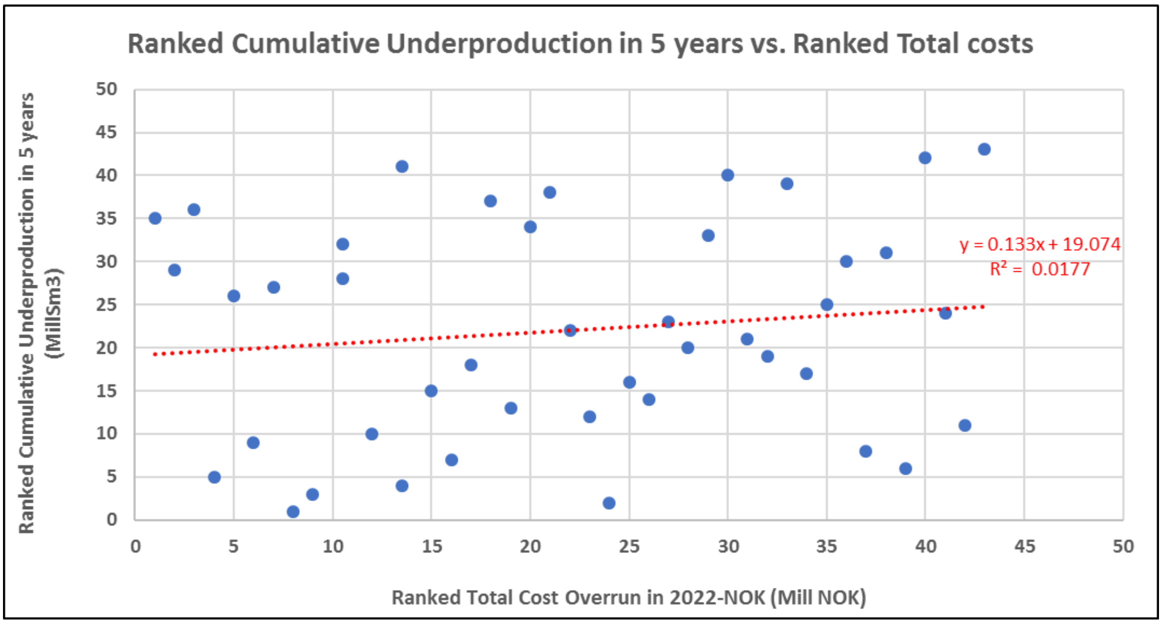

Figure 35 - Spearman scatterplot of the ranked cumulative underproduction in 5 years versus the ranked total cost overruns

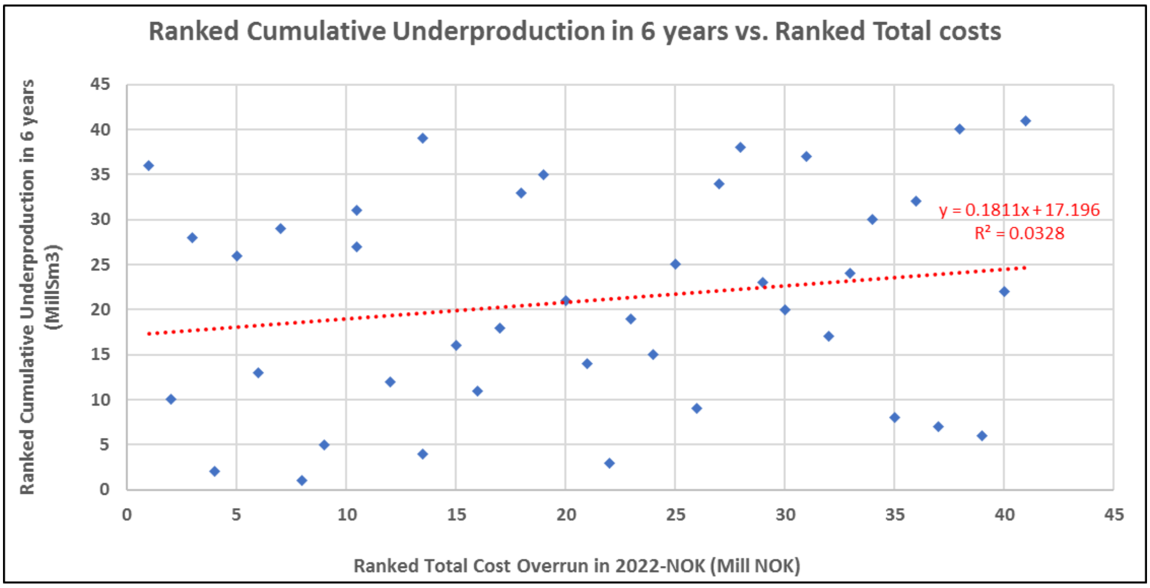

Figure 36 - Spearman scatterplot of the ranked cumulative underproduction in 6 years versus the ranked total cost overruns

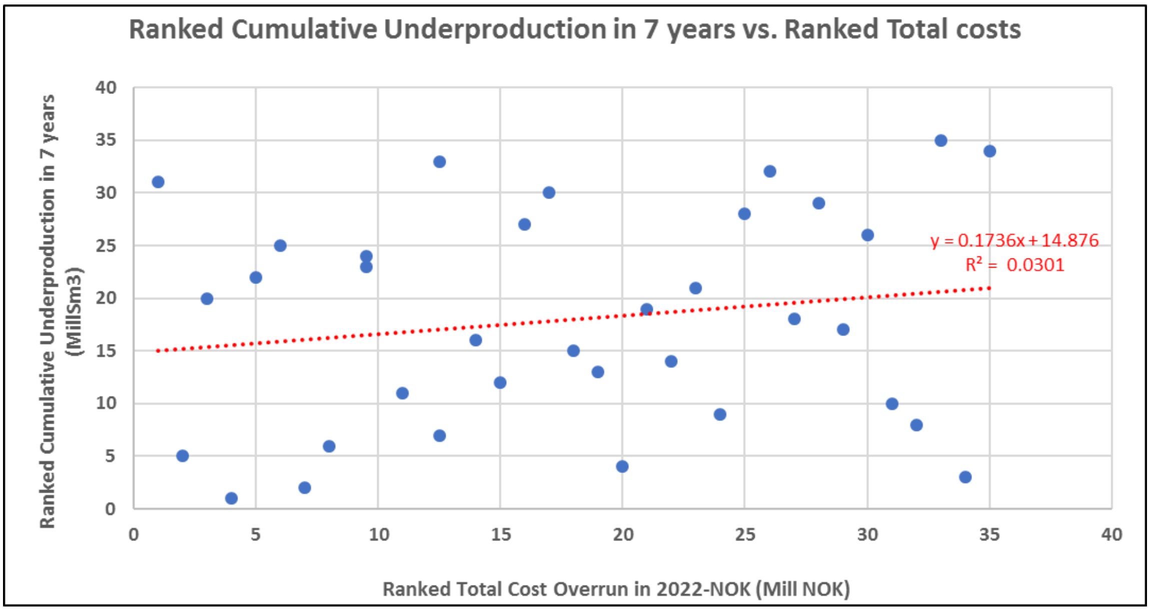

Figure 37 - Spearman scatterplot of the ranked cumulative underproduction in 7 years versus the ranked total cost overruns

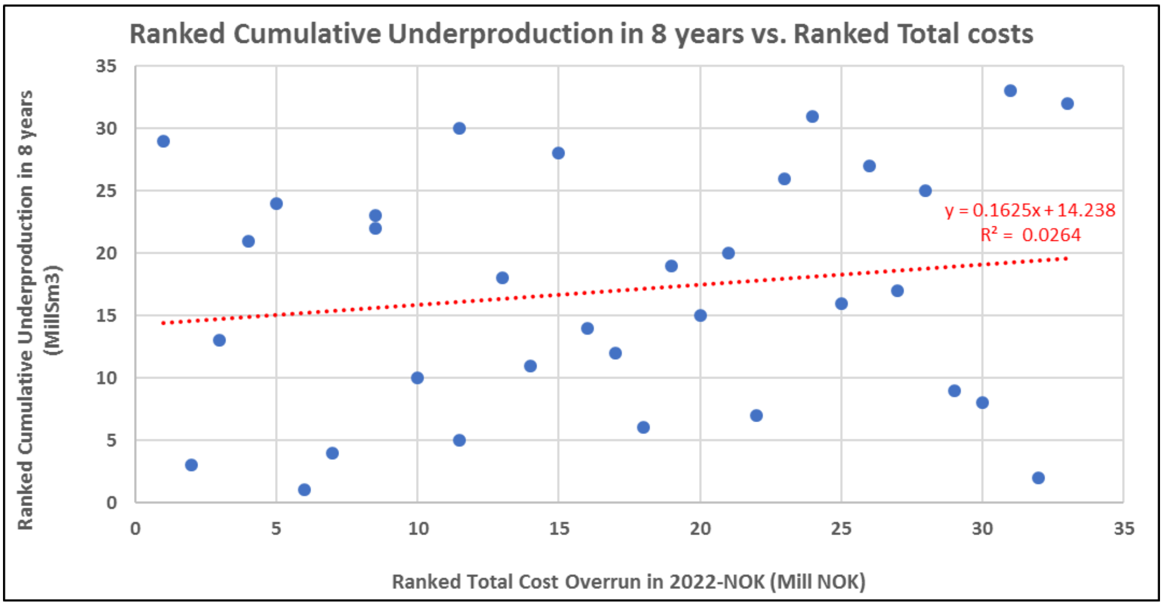

Figure 38 - Spearman scatterplot of the ranked cumulative underproduction in 8 years versus the ranked total cost overruns

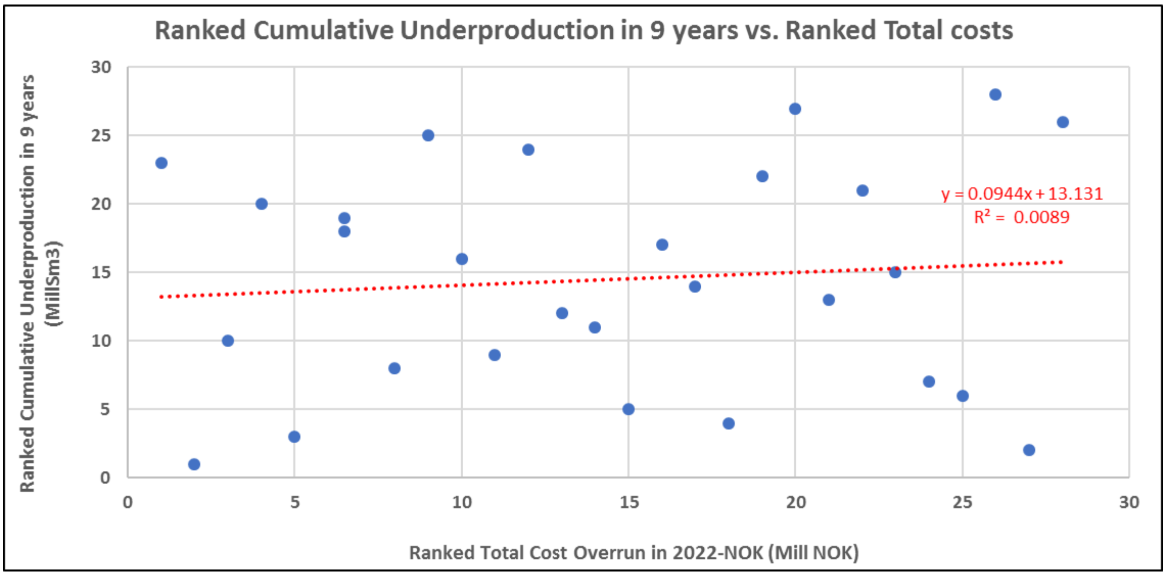

Figure 39 - Spearman scatterplot of the ranked cumulative underproduction in 9 years versus the ranked total cost overruns

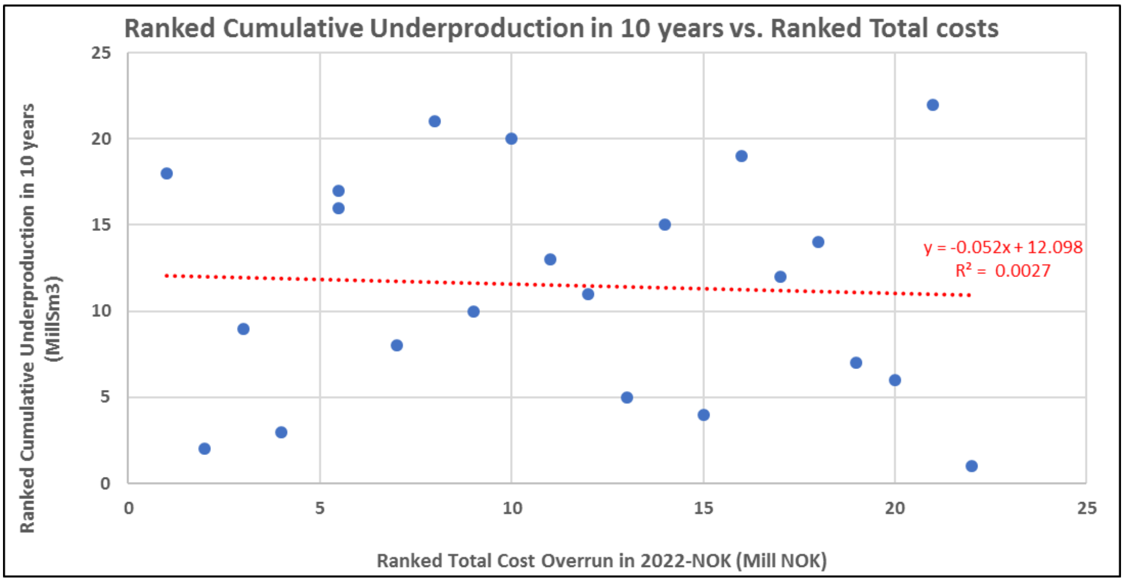

Figure 40 - Spearman scatterplot of the ranked cumulative underproduction in 10 years versus the ranked total cost overruns

To verify the accuracy of the correlation coefficients (R) and R2 values calculated during the Pearson and Spearman regression analyses, Python code was employed. Figure 41 showcases a code snippet utilized for the dataset associated with fields having 2 years of production. The code employs the "pearsonr()" and "spearmanr()" functions from the "scipy.stats" library to perform the calculations.

```
from scipy.stats import spearmanr
from scipy.stats import pearsonr

# calculate pearson's correlation
print("2 Years datasets:")
R2 , p2 = pearsonr(X_2,Y_2)
print("pearsons correlation coefficient is: %.3f" % R2)
print("pearsons p-value is: %.3f" % p2)

# calculate spearman's correlation
coef2, p_value2 = spearmanr(X_2,Y_2)
print('Spearmans correlation coefficient is: %.3f' % coef2)
# interpret the significance
alpha = 0.05
if p_value2 > alpha:
 print('Samples are uncorrelated (fail to reject H0) due to p_value = %.3f' % p_value2)
else:
 print('Samples are correlated (reject H0) due to p_value = %.3f' % p_value2)
```

Figure 41 – Python codes used for Pearson and Spearman regression analysis on 2 years datasets

Furthermore, the results obtained using the aforementioned Python codes are depicted in Figures 42 and 43. As can be seen, the results align with those calculated using Excel.

2 Years datasets:
pearsons correlation coefficient is: **0.337**
Spearmans correlation coefficient is: **0.021**

3 Years datasets:
pearsons correlation coefficient is: **0.294**
Spearmans correlation coefficient is: **0.005**

4 Years datasets:
pearsons correlation coefficient is: **0.448**
Spearmans correlation coefficient is: **0.031**

5 Years datasets:
pearsons correlation coefficient is: **0.441**
Spearmans correlation coefficient is: **0.133**

6 Years datasets:
pearsons correlation coefficient is: **0.364**
Spearmans correlation coefficient is: **0.181**

Figure 42 – Python results for the Pearson and Spearman regression analyses for 2 to 6 years datasets

7 Years datasets:
pearsons correlation coefficient is: **0.326**
Spearmans correlation coefficient is: **0.174**

8 Years datasets:
pearsons correlation coefficient is: **0.252**
Spearmans correlation coefficient is: **0.162**

9 Years datasets:
pearsons correlation coefficient is: **0.194**
Spearmans correlation coefficient is: **0.094**

10 Years datasets:
pearsons correlation coefficient is: **-0.181**
Spearmans correlation coefficient is: **-0.052**

Figure 43 - Python results for the Pearson and Spearman regression analyses for 7 to 10 years datasets

4.5.3 Support Vector Regression (SVR) Analysis

As the Pearson and Spearman analyses revealed a weak linear relationship between the dependent and independent variables, the SVR analysis was utilized to investigate non-linear relationships between cumulative underproductions and total cost overruns. The SVR analysis aims to minimize the impact of outliers on correlation coefficient and best fit curve calculations, thereby increasing the accuracy of the results. Python programming language was employed to implement the SVR analysis and determine the non-linear relationship between the variables.

To evaluate the results, the "r2_score" function from the "sklearn.metrics" library in Python was utilized to compute the correlation coefficients (R) and R^2 values for the respective datasets. The outcomes of these computations are presented in Table 6. The results illustrate that the highest correlation coefficient (R) and R^2 values were obtained for the cumulative underproductions in 2 years and total cost overruns, yielding approximately 0.54 and 0.29, respectively. Conversely, the datasets for fields with 9 years of production exhibited the lowest R and R^2 values, amounting to approximately 0.15 and 0.02, respectively.

Table 6 - Summary of correlation coefficients (R) and the R-squared (R^2) values with SVR Method

	Cumulative Underproduction in X years (Million Sm^3) VS. Total Cost Overruns in Million NOK								
X =	2 yrs.	3 yrs.	4 yrs.	5 yrs.	6 yrs.	7 yrs.	8 yrs.	9 yrs.	10 yrs.
R	0.5433	0.5048	0.5265	0.5426	0.4473	0.4385	0.2061	0.1512	0.3472
R^2	0.2951	0.2548	0.2772	0.2944	0.2001	0.1923	0.0425	0.0229	0.1206

To observe the trend of SVR correlation coefficients (R) and R2 values across the 2 to 10 years of production in the mentioned fields, Figure 44 was generated. The figure showcases the changes in correlation coefficients between underproduction and cost overruns.

From Figure 44, it can be observed that there was a decline in correlation coefficients from 0.54 to 0.50 as fields transitioned from 2 years to 3 years of production. Subsequently, the R values increased to 0.53 for fields with 4 years of production. Following this, there was a gradual reduction in the correlation coefficients from 4 years to 9 years of production, reaching its lowest value at 0.15. However, for fields with 10 years of production, the R value increased to 0.35.

This trend indicates that the correlation between cumulative underproductions and total cost overruns varied throughout the years of production, with fluctuations and a general decline over time, except for a slight increase in the case of 10 years of production.

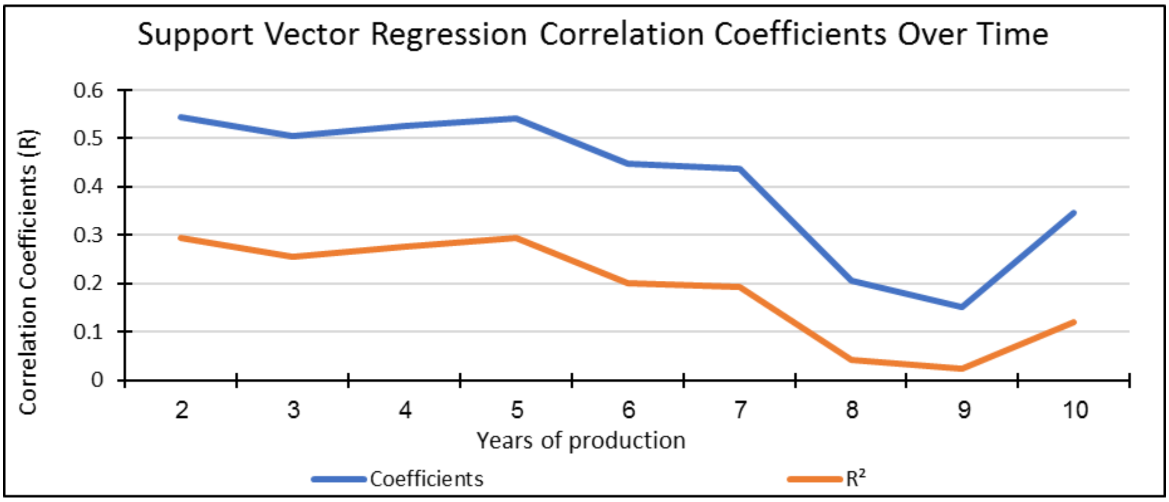

Figure 44 - Trend of SVR correlation Coefficients and R_squares between the cumulative underproductions from 2 to 10 years and the total cost overruns for the 51 fields

After defining the data sets in Python, normalizing them, and utilizing the "SVR" function from the "sklearn.svm" library, scatter plots and SVR curves were generated, as depicted in Figures 45 to 53. Connecting the values of R and R^2 from Table 6 to the corresponding figures, it becomes apparent that larger R^2 values indicate a higher percentage of data being fitted to the SVR curve. The results illustrate a moderate non-linear relationship between the variables. The data sets ranging from 2 to 10 years, along with the regression curves, are represented by blue dots and red curves, respectively.

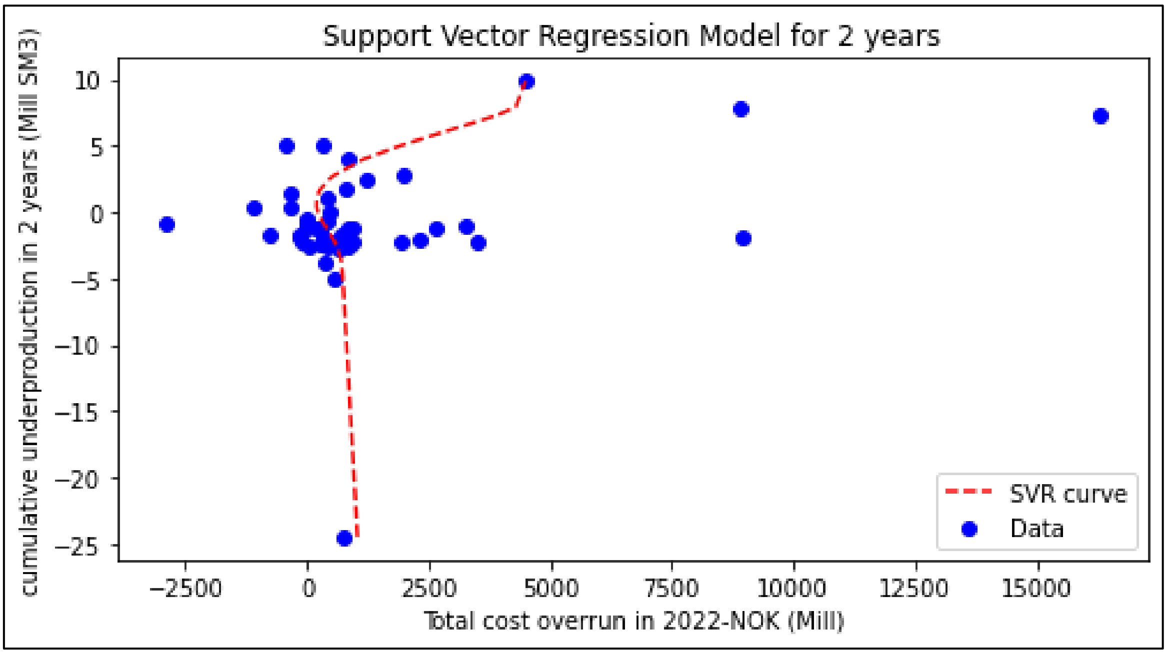

Figure 45 - SVR scatterplot of the cumulative underproduction in 2 years versus the total cost overruns

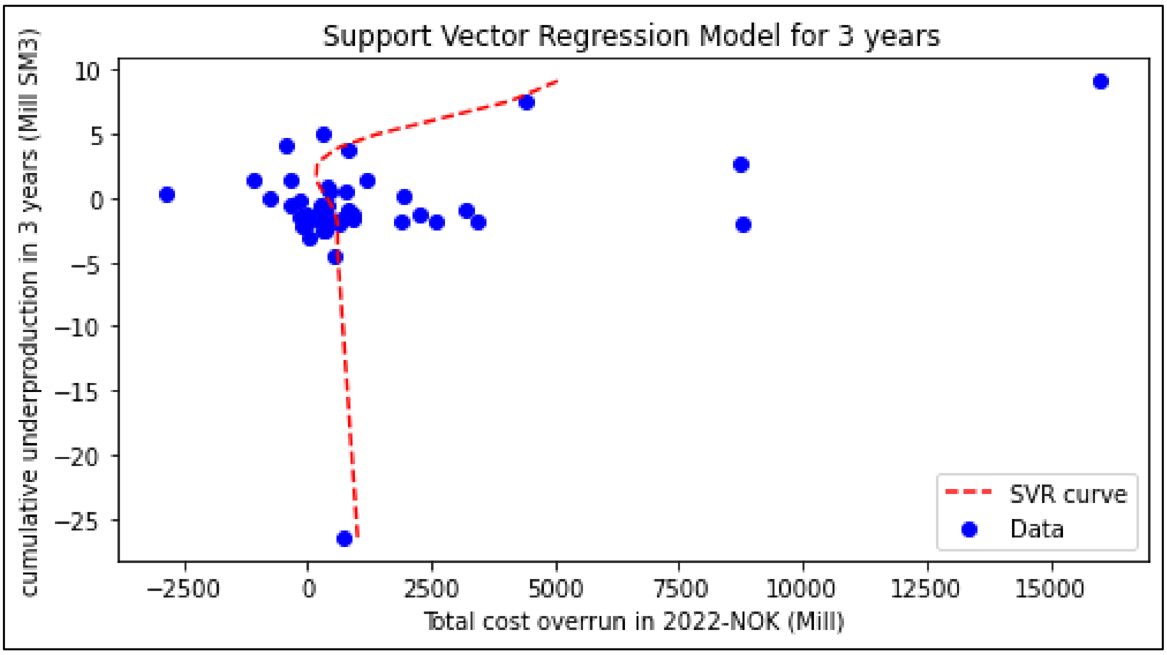

Figure 46 - SVR scatterplot of the cumulative underproduction in 3 years versus the total cost overruns

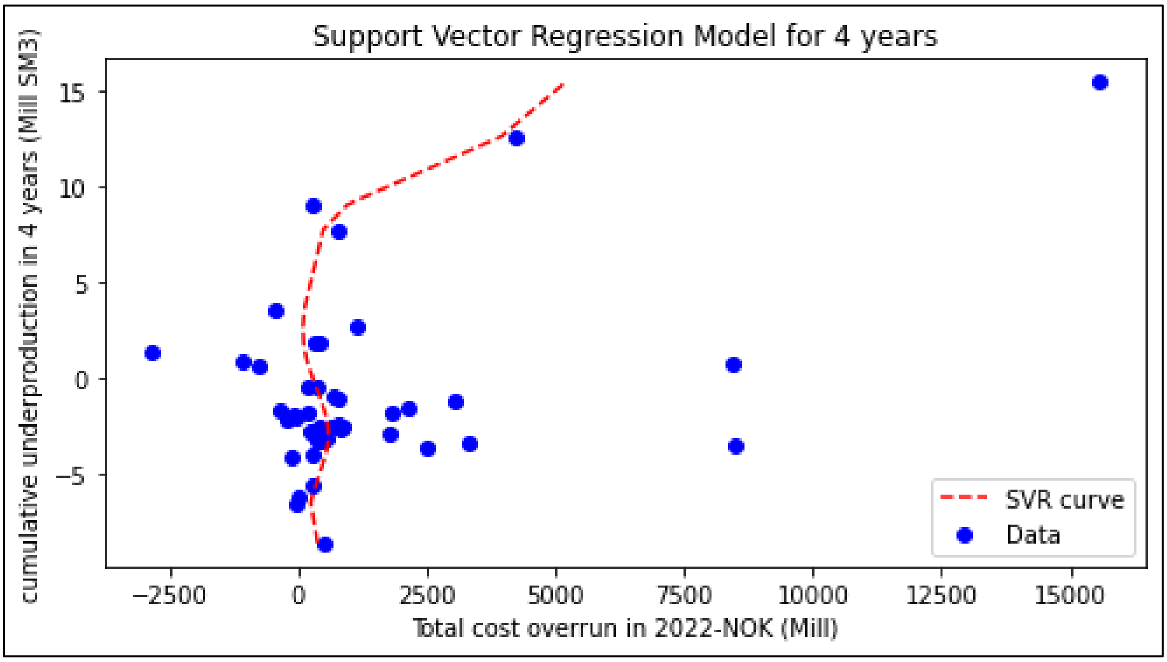

Figure 47 - SVR scatterplot of the cumulative underproduction in 4 years versus the total cost overruns

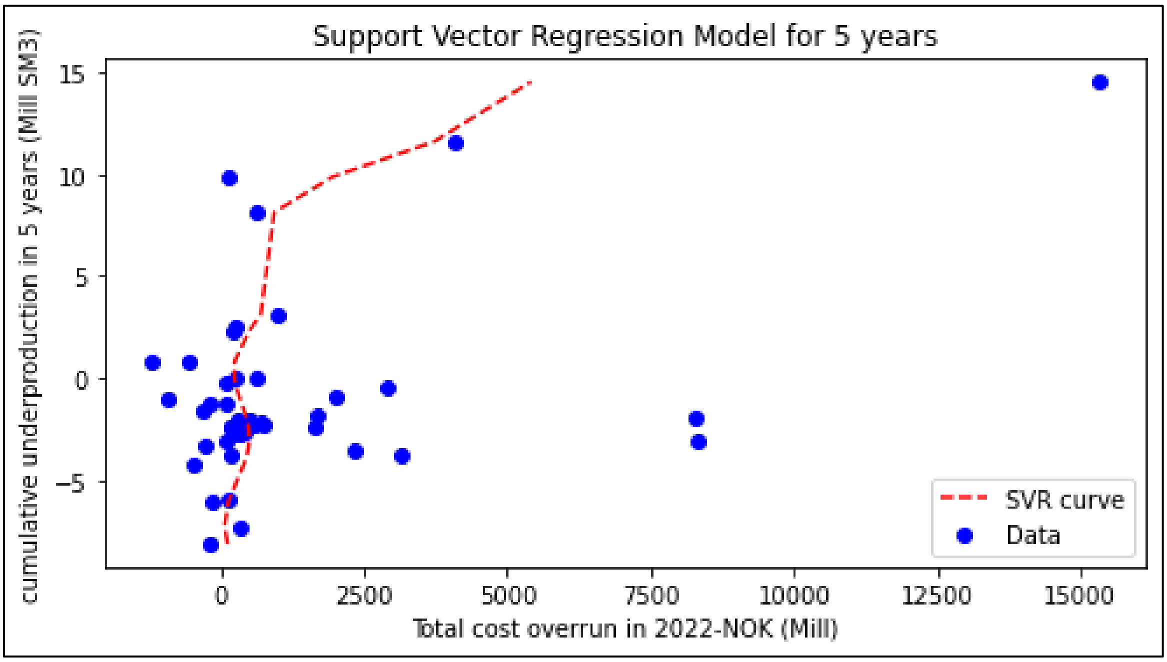

Figure 48 - SVR scatterplot of the cumulative underproduction in 5 years versus the total cost overruns

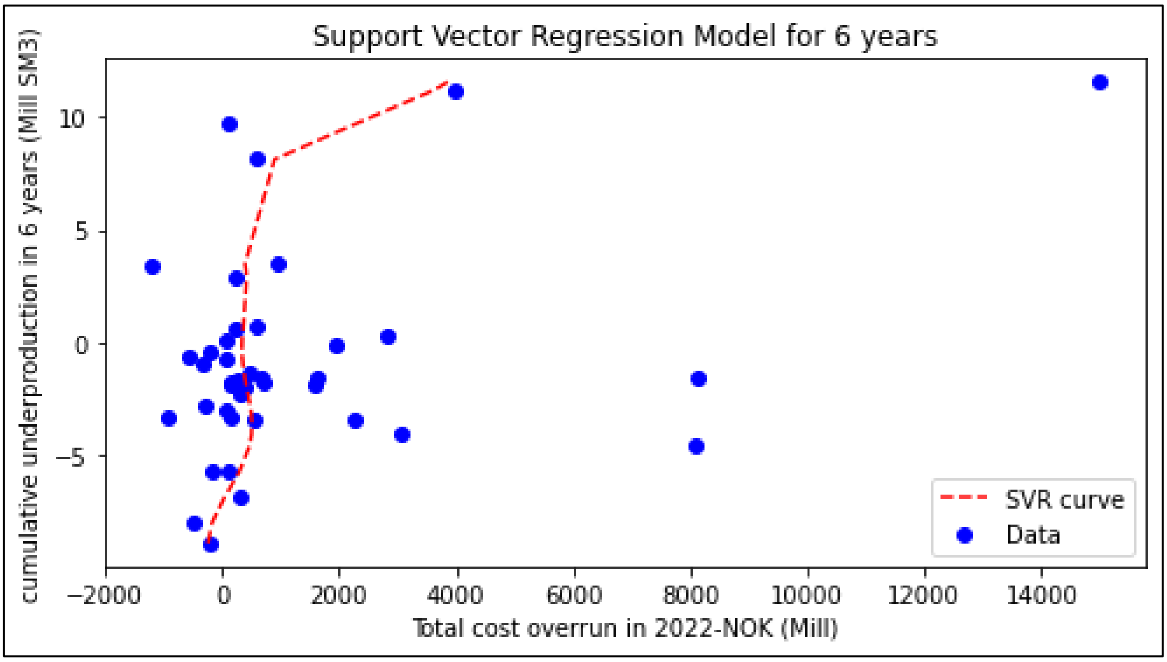

Figure 49 - SVR scatterplot of the cumulative underproduction in 6 years versus the total cost overruns

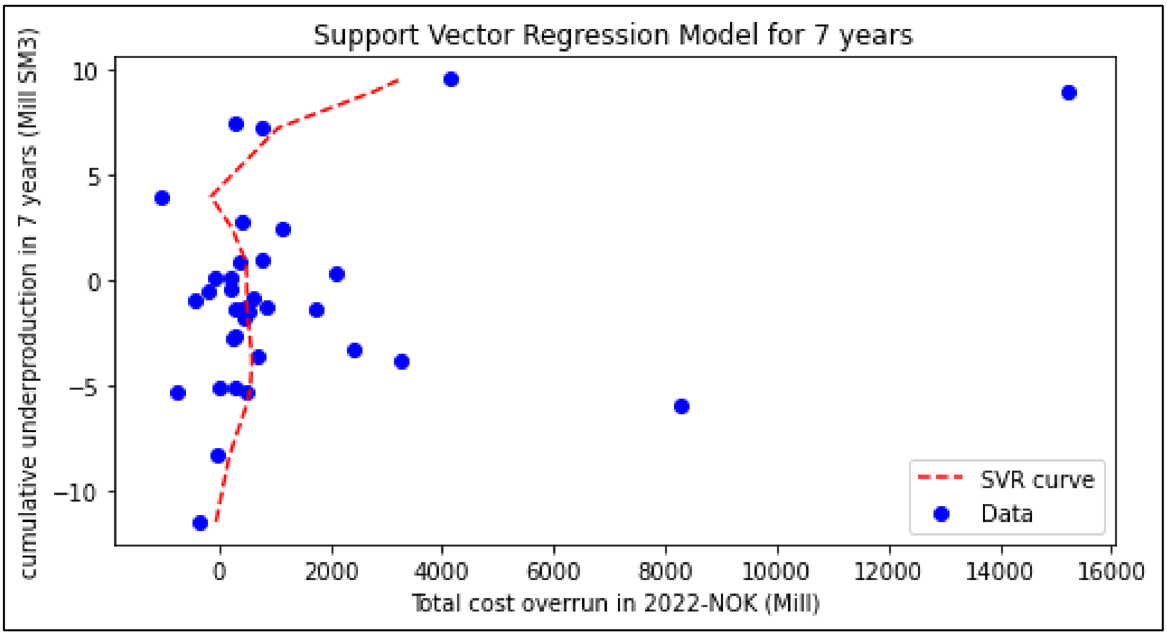

Figure 50 - SVR scatterplot of the cumulative underproduction in 7 years versus the total cost overruns

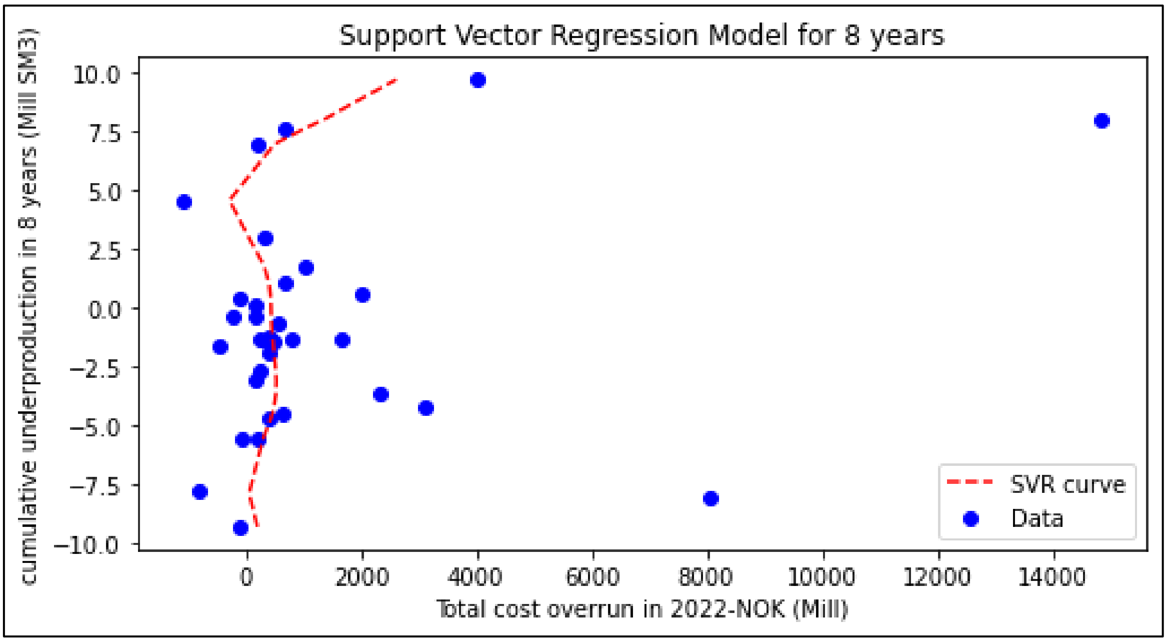

Figure 51 - SVR scatterplot of the cumulative underproduction in 8 years versus the total cost overruns

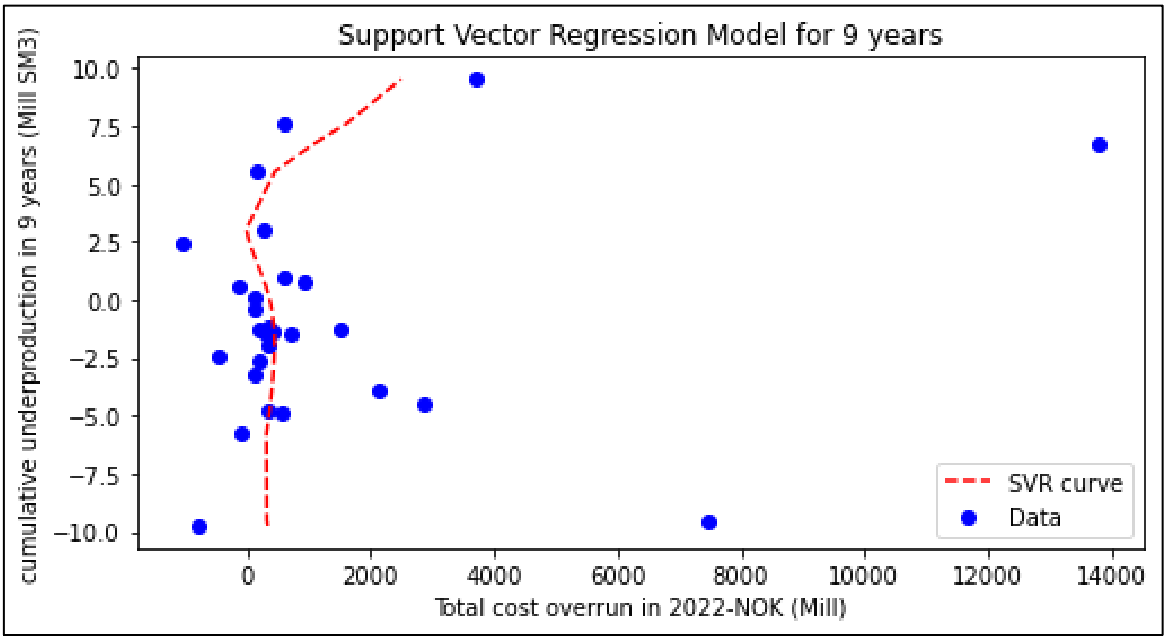

Figure 52 - SVR scatterplot of the cumulative underproduction in 9 years versus the total cost overruns

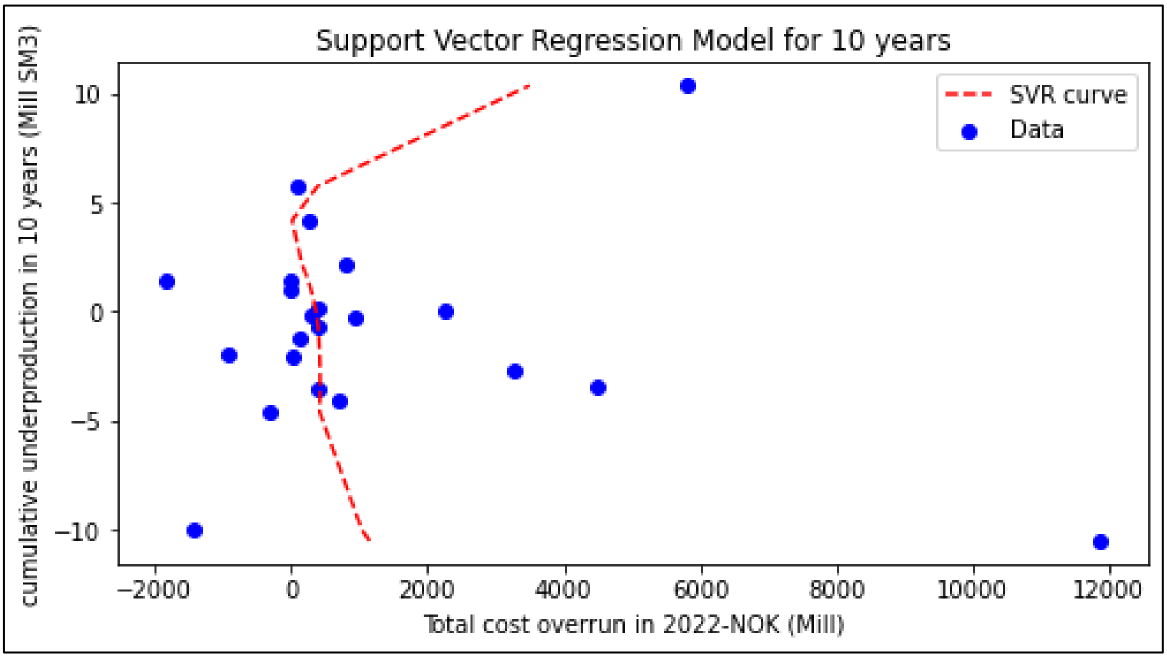

Figure 53 - SVR scatterplot of the cumulative underproduction in 10 years versus the total cost overruns

4.5.4 Comparing the results of regression analysis

To provide a visual comparison between the regression methods, Figures 54 to 62 were created using Python with the "seaborn" and "matplotlib.pyplot" libraries. These figures compare the scatter plots from the SVR and Pearson methods, along with their respective regression lines. Since the Spearman ranks are unitless and not directly comparable with the other two methods, they were not included in the comparison.

Figures 54 to 62 display the SVR curves and Pearson regression lines in gray and red colors, respectively, overlaid on the scatter plots represented by blue dots. From the figures, it is evident that the gray regression curve of the SVR method provides a better fit to the data points compared to the red line associated with the Pearson method as the fields mature.

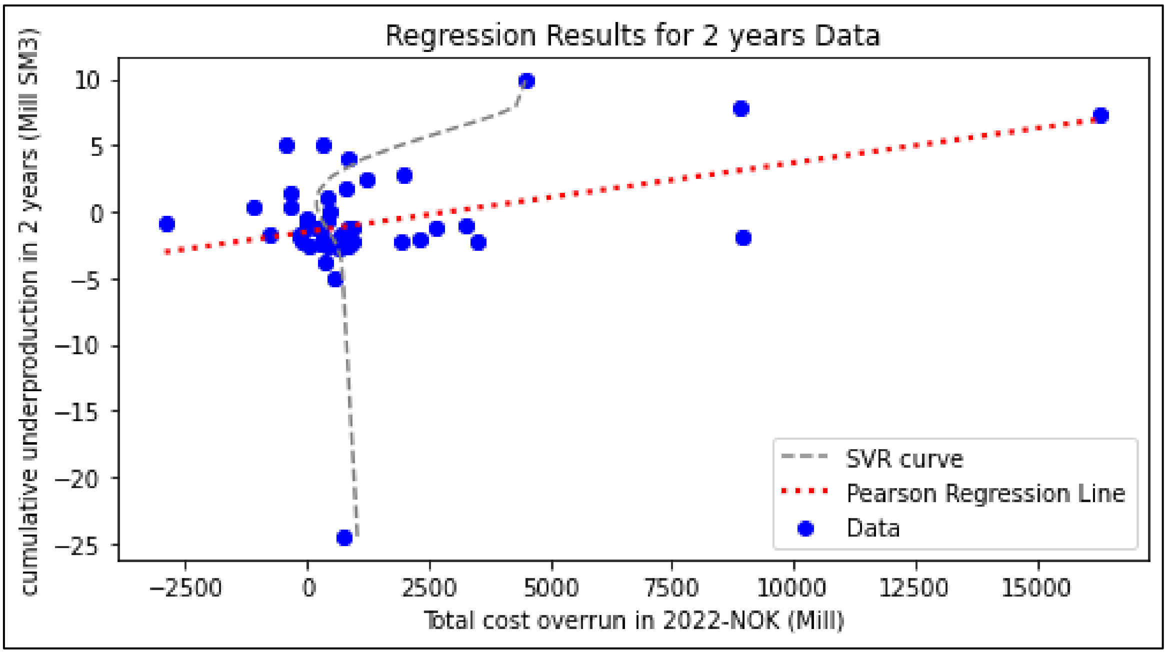

Figure 54 – SVR and Pearson regression lines of the cumulative underproduction in 2 years vs. the total cost overruns

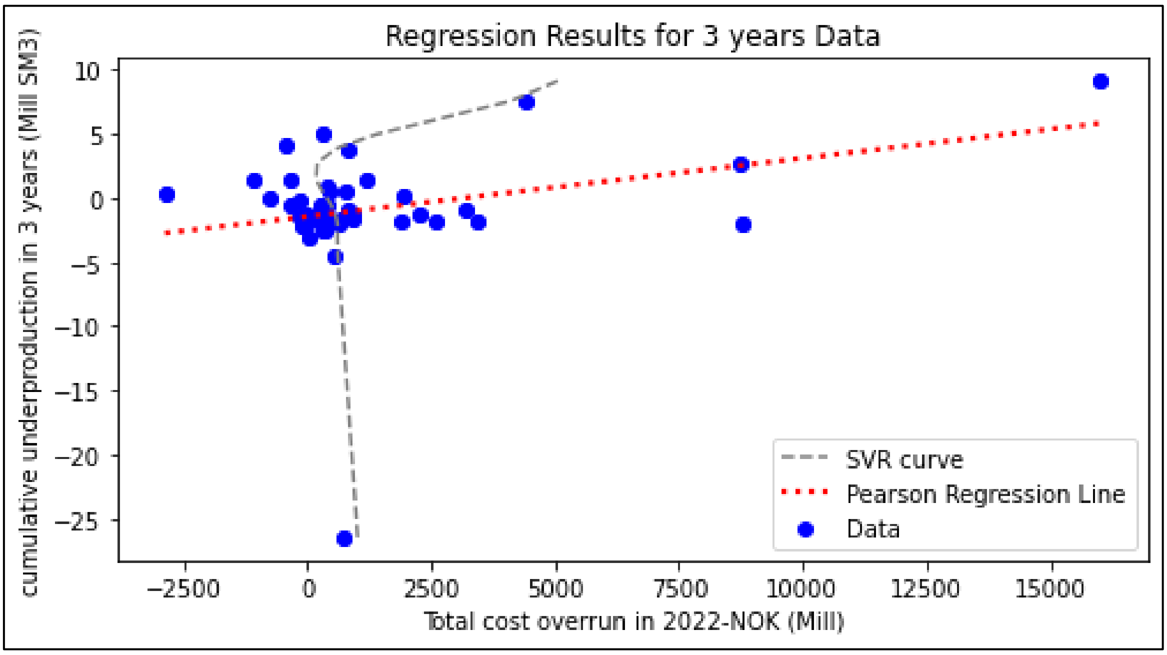

Figure 55 - SVR and Pearson regression lines of the cumulative underproduction in 3 years vs. the total cost overruns

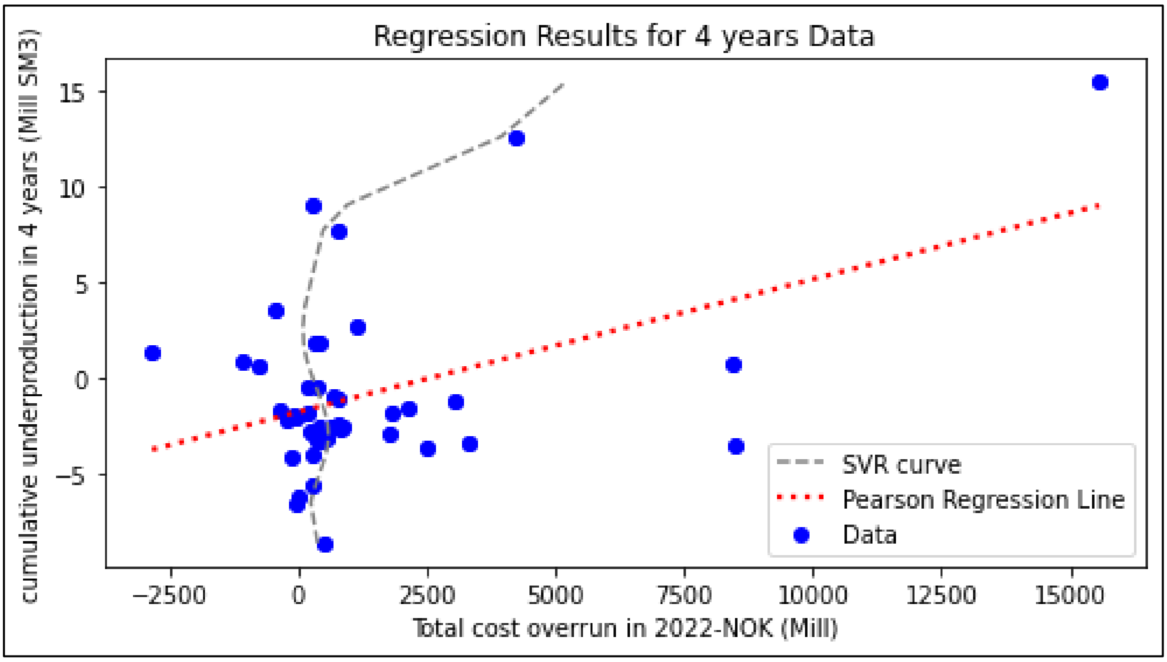

Figure 56 - SVR and Pearson regression lines of the cumulative underproduction in 4 years vs. the total cost overruns

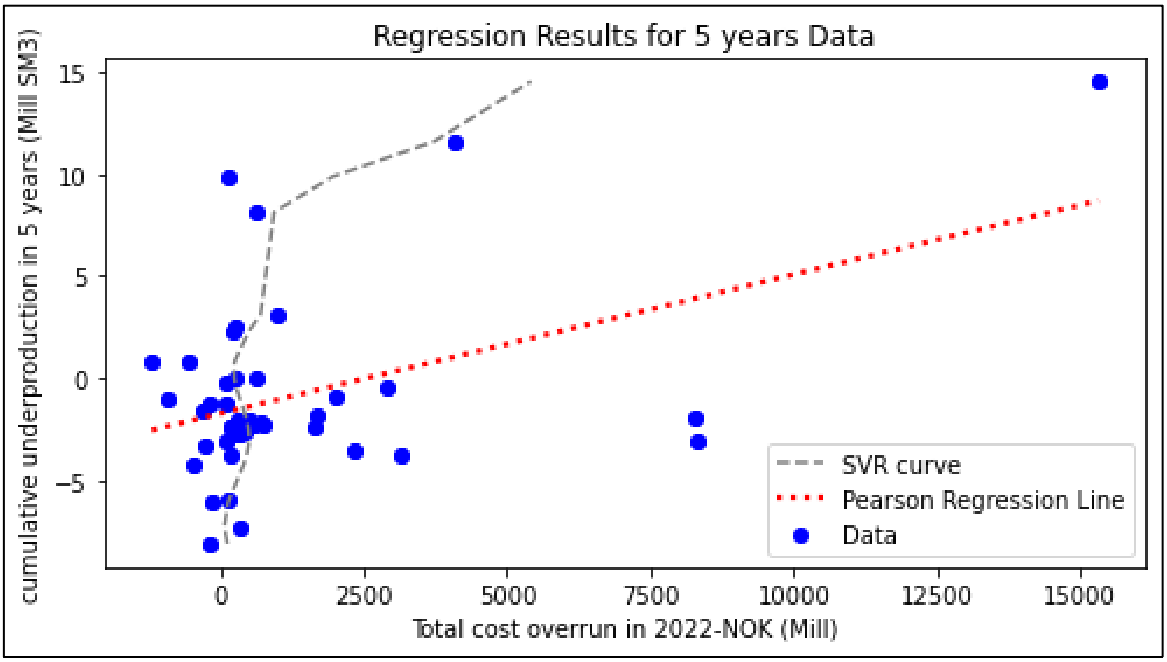

Figure 57 - SVR and Pearson regression lines of the cumulative underproduction in 5 years vs. the total cost overruns

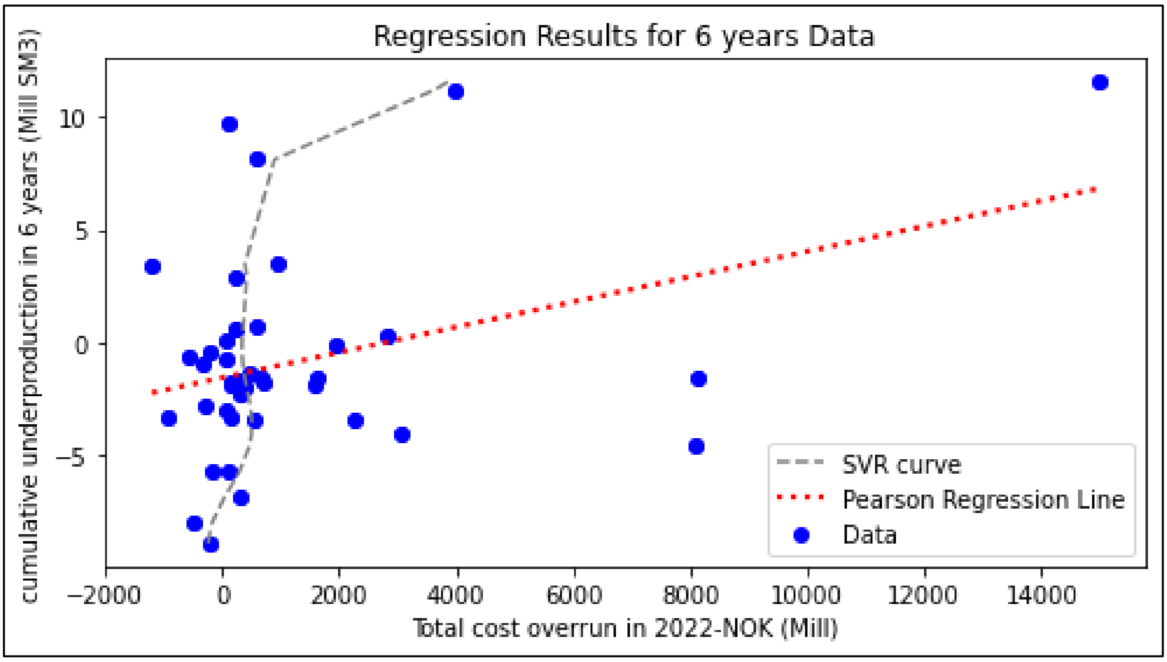

Figure 58 - SVR and Pearson regression lines of the cumulative underproduction in 6 years vs. the total cost overruns

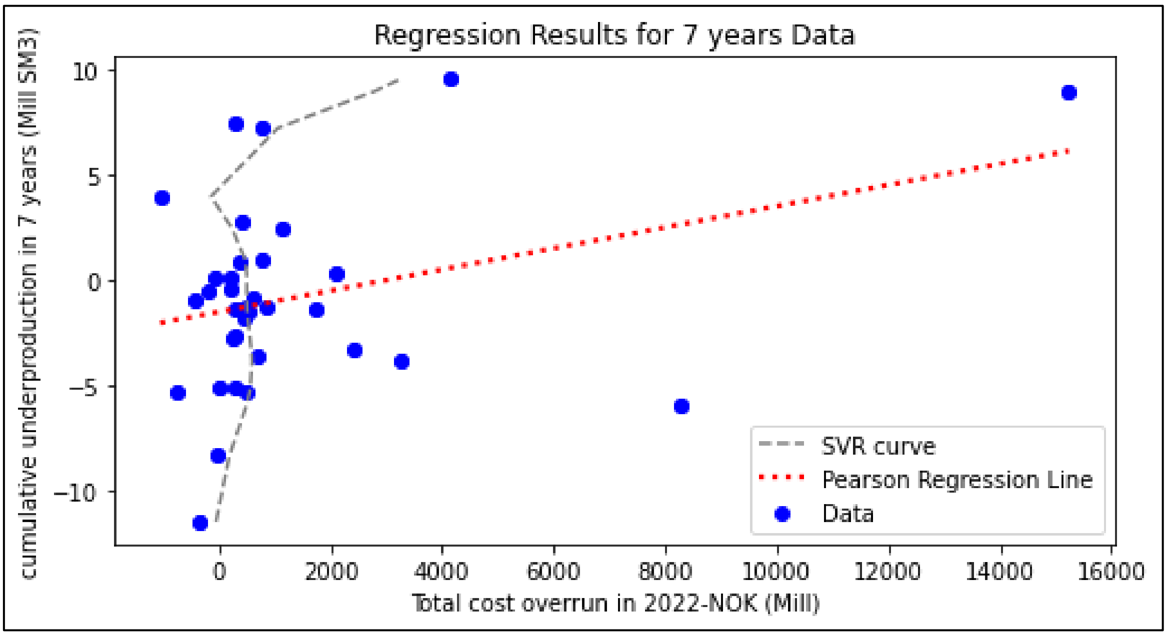

Figure 59 - SVR and Pearson regression lines of the cumulative underproduction in 7 years vs. the total cost overruns

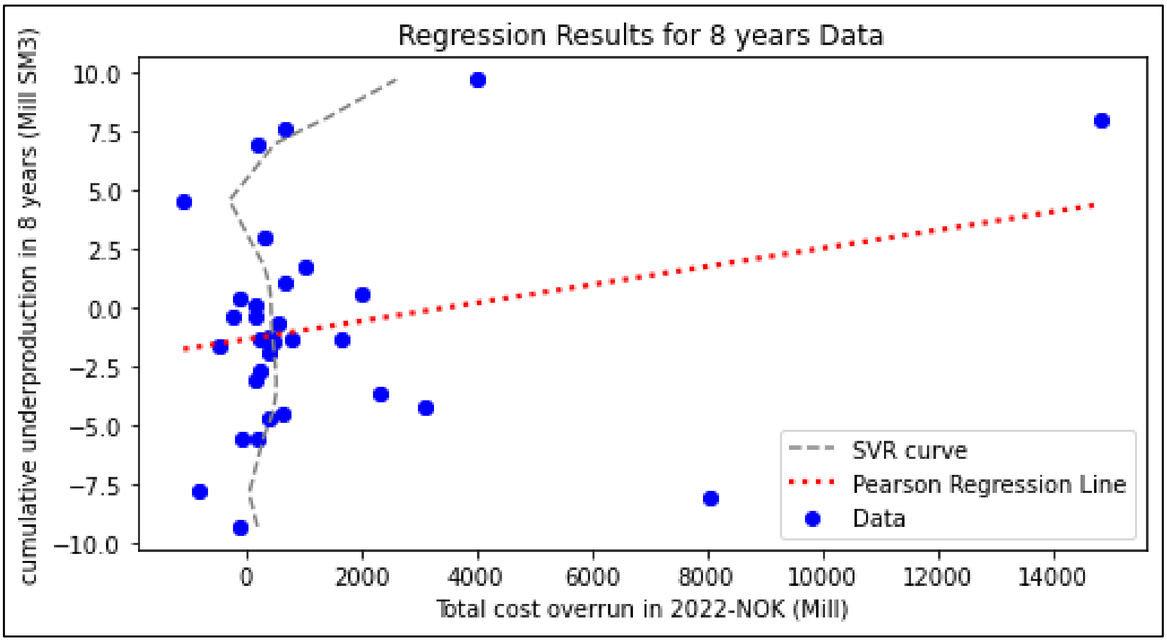

Figure 60 - SVR and Pearson regression lines of the cumulative underproduction in 8 years vs. the total cost overruns

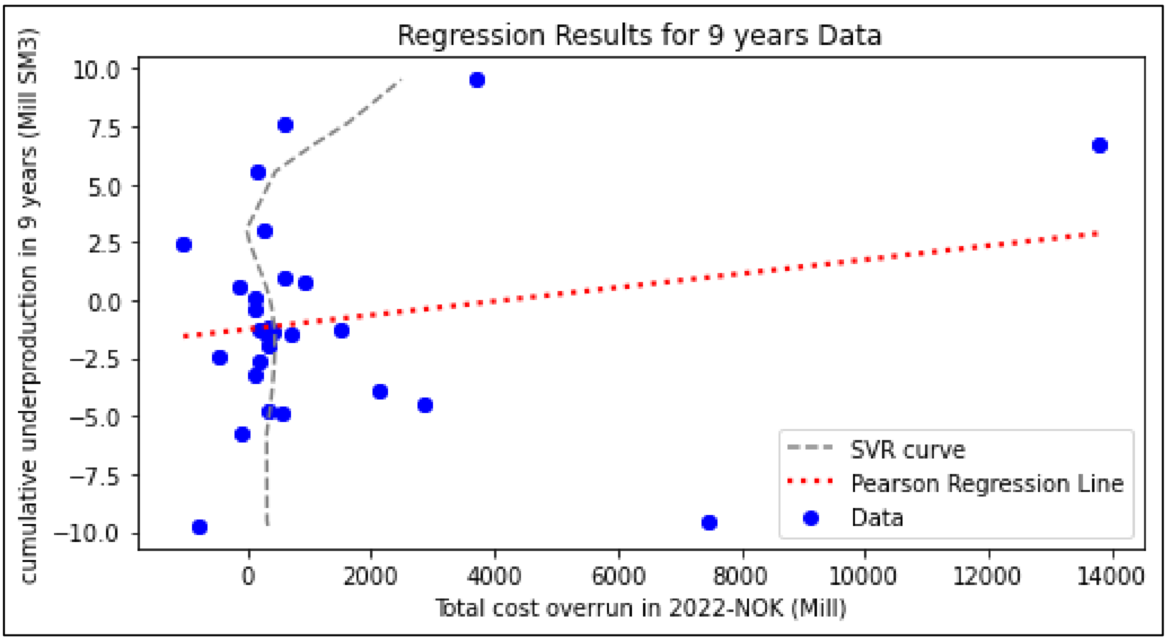

Figure 61 - SVR and Pearson regression lines of the cumulative underproduction in 9 years vs. the total cost overruns

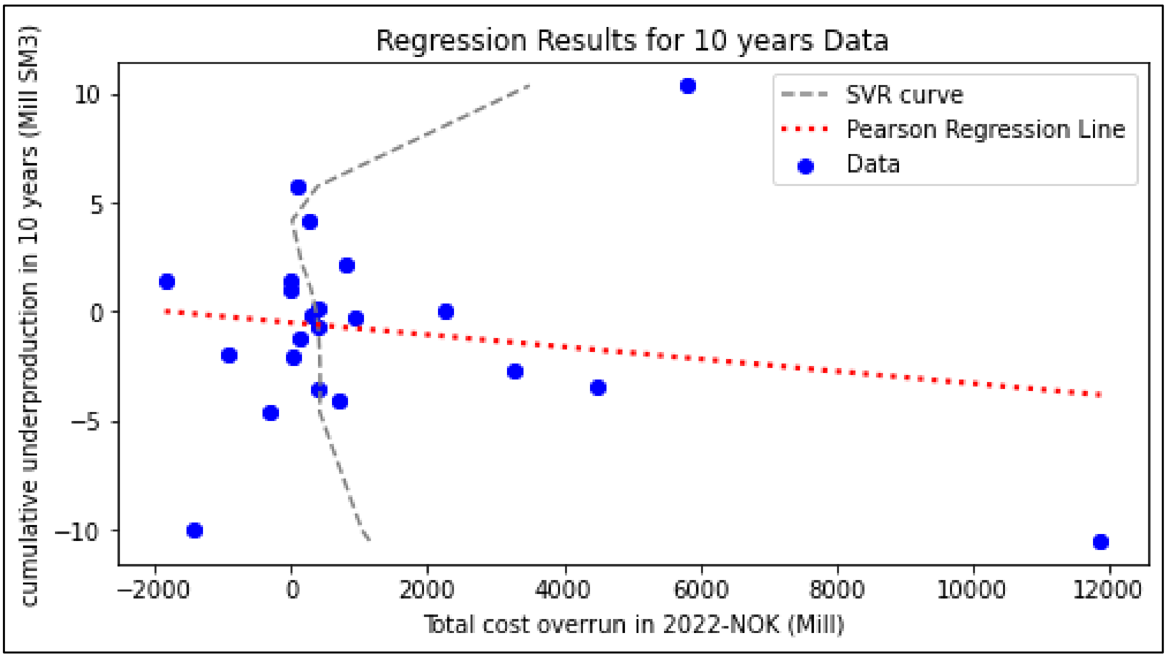

Figure 62 - SVR and Pearson regression lines of the cumulative underproduction in 10 years vs. the total cost overruns

In the following section, Figures 63 to 66 are used to compare the results of correlation coefficients and R2 values from the previous regression analyses.

It is important to note that nonlinear correlation refers to deviations in one variable that are non-linearly correlated to the deviations in the other variable. Similarly, linear correlation means that changes with a constant rate (Pearson) or nonconstant rate (Spearman) in one variable lead to the same rate of changes in the other variable.

Figure 63 provides an overview of the changes in correlation coefficients from the Pearson, Spearman, and SVR regression analyses, represented by blue, orange, and gray lines, respectively. From the figure, it can be observed that for the data sets ranging from 2 to 7 years, SVR initially shows higher non-linear correlation between the variables, followed by the Pearson method. Meanwhile, the Spearman correlation coefficients for these data sets remain relatively low.

Between 8 to 9 years of production, there is a change in the trend of correlation coefficients. Pearson results surpass the SVR results, indicating a stronger linear relationship with a constant rate between the variables during this time period. On the other hand, Spearman results remain at the lower end, suggesting weaker correlation based on ranks.

Finally, for fields with 10 years of production data, the order of the R values is SVR, Spearman, and then Pearson. This indicates a stronger non-linear relationship or a linear relationship with a non-constant rate (monotone) between the variables during this period.

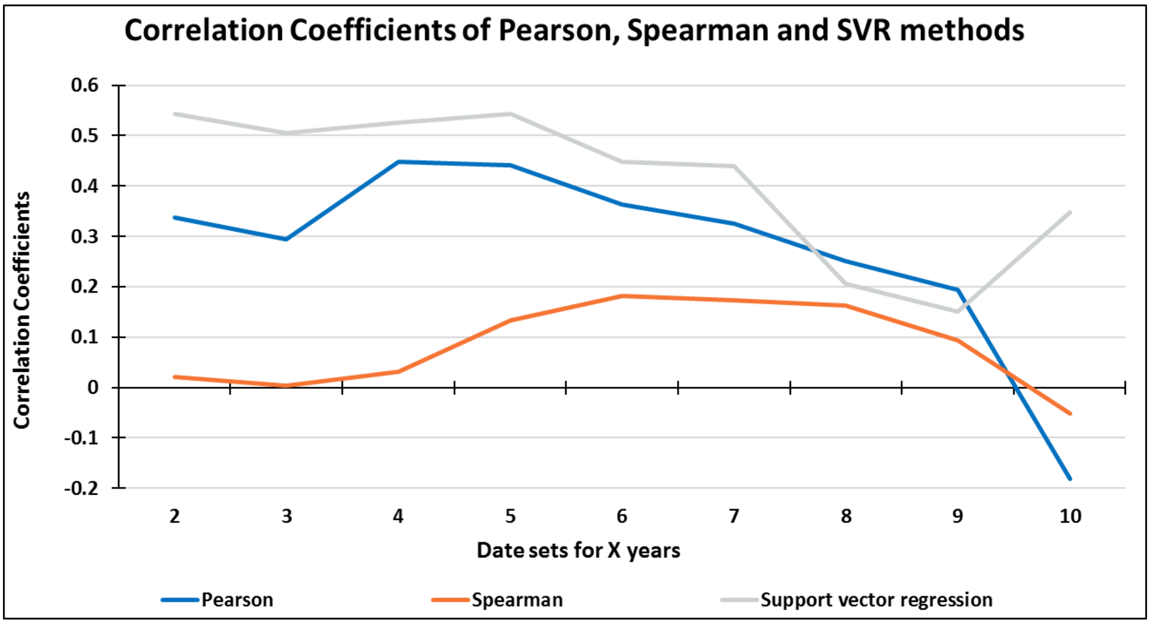

Figure 63 – Pearson, Spearman and SVR correlation coefficients for 51 fields over 10 years

Figure 64 illustrates the average correlation coefficients obtained from the Pearson, Spearman, and SVR methods, represented by the blue, orange, and gray columns, respectively. The figure shows that, on average, the SVR method yielded a correlation coefficient of 0.41 between the cumulative underproductions during the first 10 years and the total cost overruns. In comparison, the Pearson and Spearman methods had average correlation coefficients of 0.28 and 0.08, respectively. This indicates that, on average, the strongest relationship between the variables was non-linear rather than linear. Furthermore, the figure highlights a moderate non-linear relationship, a weak linear relationship between the data, and a very weak relationship between the rank of the data.

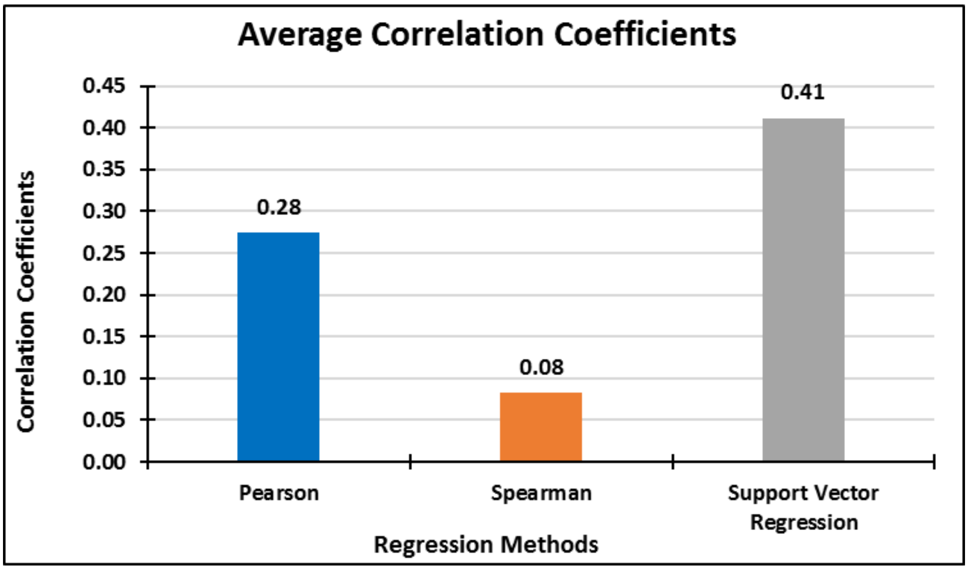

Figure 64 - Pearson, Spearman and SVR average correlation coefficients for 51 fields over 10 years

Figure 65 compares the trend of R2 values obtained from the Pearson, Spearman, and SVR analyses, represented by the blue, orange, and gray lines, respectively. From 2 to 7 years of production, the SVR R2 values remained higher than the Pearson and Spearman R2 values. However, the Spearman R2 values were almost zero for the rank of the data between 2 to 4 years. Both SVR and Pearson R2 values exhibited a similar trend, decreasing from 2 to 3 years of production. Subsequently, an increase in SVR values occurred from 3 to 5 years, while Pearson values increased from 3 to 4 years. Overall, SVR, Pearson, and Spearman results showed a gradual decline in R2 values until 9 years, 10 years, and 10 years, respectively. Notably, between 8 and 9 years of production, the Pearson correlation coefficient was higher than the SVR R value.

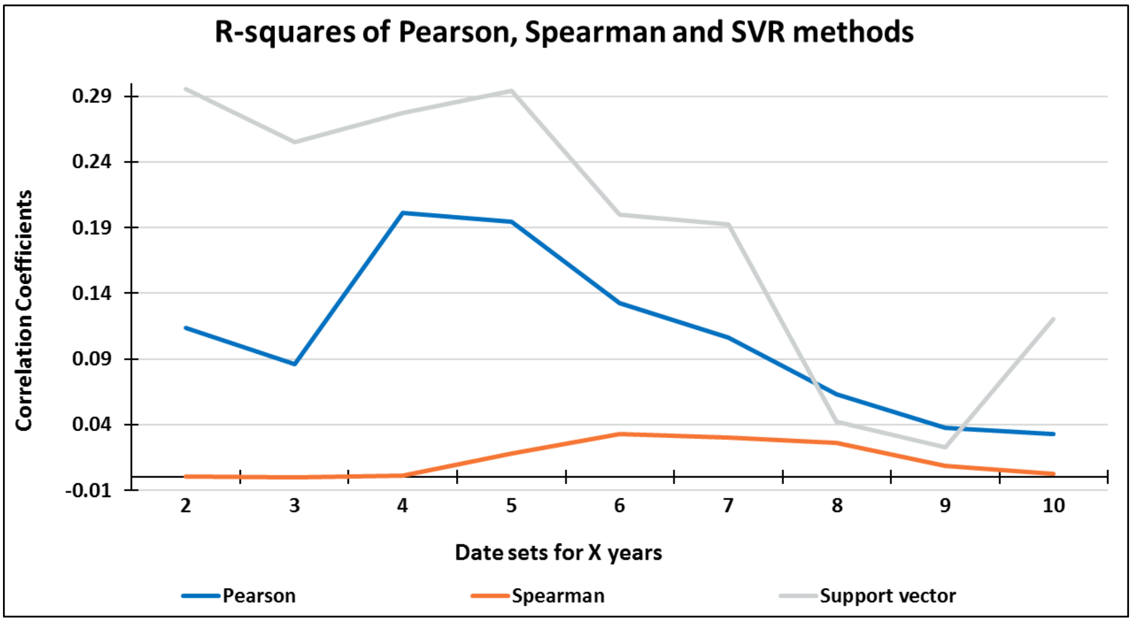

Figure 65 - Pearson, Spearman and SVR R-squares for 51 fields over 10 years

Figure 66 illustrates that, on average and in descending order, the highest percentage of data points fitted on the regression curves was observed in the SVR method, followed by the Pearson method, and finally the Spearman method. Over the 10-year period, approximately 19% of the data points were fitted on the SVR non-linear curve, 11% on the Pearson linear curve with a constant rate, and only 1% of the data ranks were fitted on the linear line with a non-constant rate. This indicates that the SVR method had the highest percentage of data points that closely matched the regression curve, while the Spearman method had the lowest percentage.

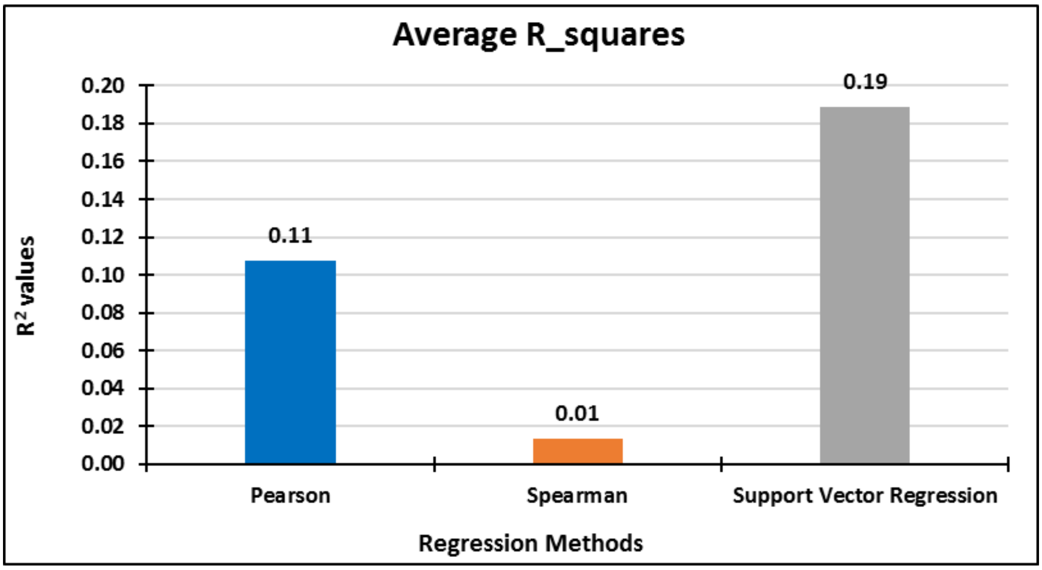

Figure 66 - Pearson, Spearman and SVR average R-squares for 51 fields over 10 years

Based on the results depicted in the previous figures, it is evident that the correlation between the cumulative underproductions in the first 10 years and total cost overruns can be described as a moderate nonlinear relationship. Additionally, there is a weak linear relationship between the variables. Furthermore, the relationship between the data ranks shows a very weak linear correlation.

Chapter 5 – Discussion

5.1 Discussing the results of this study

In this chapter, a brief summary of key findings, answers to the research questions, interpretations of the results, limitations of the findings, and suggestions for further research on the topic of poor forecasts are presented.

Chapter 4 involved an analysis of forecasted and actual development time, cost, and production data for the oil fields on the NCS. The analysis revealed that forecasts of project costs and schedules tended to be underestimated compared to the actual values, while production forecasts were mostly overestimated. Looking at the trends of forecasts during the study period, it was evident that the forecasts experienced many fluctuations. Regarding the improvements in forecasts, it was shown that although the production and development cost forecasts have improved slightly, the forecasted development times showed no signs of improvement.

Furthermore, the present value loss resulting from cost overruns, delays, and underproduction in oil fields on the NCS was calculated separately and in total using two methods for adjusting the production start dates. In the last part of the analysis, the correlation between cumulative underproductions in the first 10 years and total cost overruns for the fields on the NCS was also investigated.

Considering the research questions mentioned at the beginning of the study, the results of the analysis confirmed several key findings:

1) The deviations of the actual values from the forecasted values at the time of PDO approval for the development time, cost, and production data were approximately 101 days (12%), NOK 268 billion (12.7%), and 50 million Sm3 of oil production (5%).
2) In total, NOK 387 billion of value was lost, which includes NOK 268 billion of value loss due to cost overruns, NOK 39 billion of value loss due to delays, and NOK 80 billion of value lost due to underperformance in production.
3) According The regression analyses revealed that the average correlation coefficient from the Pearson, Spearman, and SVR methods was 0.28, 0.08, and 0.41, respectively. This indicates that there is a weak linear relationship and a moderate nonlinear relationship between the cumulative underproduction in the

first 10 years of production and the total cost overruns. Additionally, the analysis showed a very weak linear relationship between the ranks of these variables.

The results of this study were consistent with previous research [27], [6], [2] focusing on forecasts of development time, cost, and production data in petroleum projects. These findings demonstrated that production shortfalls, schedule delays, and cost overruns persist in the industry. While most previous studies examined the first four years of production, this study extended the analysis to encompass the first ten years, revealing that there was still a discrepancy in the delivered production value compared to the promised value. However, there was a slight improvement in production forecasts when considering the longer time frame.

As discussed in Chapter 3, there were several limitations in this study that need to be acknowledged. These limitations included the limited number of datasets due to missing or incomplete information, the focus on oil fields on the NCS, and the consideration of only mean forecasts and delays in years. To conduct a more comprehensive analysis, it is recommended that future studies employ statistical or machine learning methods that incorporate all three forecast scenarios (p10, mean, p90) provided in the PDO approvals. This would provide a more comprehensive understanding of the forecast accuracy and potential deviations in project outcomes.

5.2 Causes of Unreliable Forecasts and their Solutions

In the field of poor forecasts, Kahneman and Tversky [63] have highlighted the tendency for overconfident forecasts, which can be attributed to human judgment biases, such as optimism and neglecting the distributions of the data. Their research suggests that individuals often exhibit a bias towards overly optimistic forecasts, failing to account for the potential range of outcomes and uncertainties associated with the data.

Lovallo and Kahneman [64] have observed a common cognitive bias in project planning, where planners tend to have an optimistic outlook and believe that the project will proceed according to plan. This bias leads to underestimating the potential risks, costs, and duration of the projects. This phenomenon, known as the planning fallacy, can have significant consequences, as it can lead to unrealistic project timelines, budget overruns, and poor resource allocation.

A work by Flyvbjerg, Garbuio, and Lovallo[7] suggests that errors in estimating costs, time, and production values can be categorized into three main groups of deception, delusion, and bad luck provides a framework to understand the underlying factors contributing to these errors.

Deception refers to the intentional presentation of projects in a more favorable light than their actual state. This can occur when there is a tendency to misrepresent or downplay the challenges, risks, or limitations of the project. Deception can stem from strategic motivations, where individuals may have incentives to portray the project in a more positive manner to secure funding, gain support, or meet targets [7].

Delusion, on the other hand, arises from unrealistic judgments and expectations regarding the benefits, losses, and uncertainties associated with the project. It is often a consequence of an "inside view" perspective, where individuals become overly focused on the specific details and circumstances of the project, leading to underestimations of potential challenges or negative outcomes. Delusion can be driven by cognitive biases, such as optimism bias or overconfidence, which can cloud judgment and lead to biased estimations [7].

Lastly, bad luck encompasses external factors and events that are beyond the control of managers and planners. These factors may include unforeseen market fluctuations, natural disasters, regulatory changes, or other unpredictable events that can disrupt the project's progress and outcomes. Bad luck highlights the inherent uncertainties and uncontrollable variables that can impact project performance, despite diligent planning and estimation efforts [7].

The recommendations provided by Welsh, Begg, and Bratvold [65] and the perspectives highlighted by other studies offer valuable insights on improving forecasts in the oil and gas industry.Training decision-makers and enhancing their understanding of biases and uncertainties can be instrumental in improving forecasts. By raising awareness about cognitive biases and common pitfalls in decision-making, decision-makers can become more mindful of their own biases and make more informed and rational judgments. This can help in debiasing forecasts and reducing the influence of overconfidence, optimism bias, and other cognitive biases that may lead to inaccurate forecasts.

The use of superforecasters, as mentioned in some studies [66], [67], involves engaging individuals who have a track record of making accurate predictions. These individuals possess specific skills and approaches that enable them to navigate uncertainties and make better forecasts. Leveraging their expertise and incorporating their insights into forecasting processes can lead to improved understanding and anticipation of future outcomes.

Additionally, Flyvbjerg's [8] suggestion of utilizing historical data and employing the Reference Class Forecasting (RCF) method is another valuable approach. RCF involves analyzing past projects or similar cases to establish a reference class and using that information to generate forecasts for future projects. By gaining an outside view and taking into account the actual performance of similar projects, forecasters can mitigate the impact of individual biases and improve the accuracy of forecasts.

Implementing a combination of these approaches, including training decision-makers, leveraging the expertise of superforecasters, and adopting the RCF method, can contribute to more accurate and reliable forecasts in the oil and gas industry.

5.3 Future Research

Future scientific research can indeed explore various aspects of forecast quality in different industries, such as manufacturing, petroleum, agriculture, and others. While cost overruns have been extensively studied, more attention can be given to production shortfalls as a crucial topic in forecasting.

To enhance the accuracy of forecasts, future studies could consider more granular data, such as monthly production and development schedules or costs. This level of detail can provide a more precise understanding of fluctuations and patterns in project performance. Additionally, expanding the analysis to include data from different regions and types of petroleum projects (e.g., oil, gas, Liquid Natural Gas) would contribute to a more diverse and comprehensive dataset, leading to more robust findings. It would also be valuable to examine the impact of different currencies and trading systems on project forecasts to account for global variations and economic factors.

In terms of regression analysis, alternative regression methods can be explored to identify relationships between schedule and cost overruns and production shortfalls. Utilizing different regression models and techniques can offer additional insights and potentially uncover nuanced relationships that were not captured by the previous analysis. By developing mathematical equations to describe the relationships between overruns, underperformance, and forecasts, it becomes possible to optimize one set of forecasts based on the insights derived from the other set.

By incorporating these suggestions into future research, a deeper understanding of forecast quality and the underlying factors influencing project performance can be gained. The inclusion of more extensive and diverse datasets, adoption of advanced regression techniques, and exploration of different industries and global contexts will contribute to the continuous improvement of forecasting practices across various sectors.

Chapter 6 – Conclusion

The objective of this work was twofold. Firstly, to identify and quantify potential relationships between production shortfalls and cost overruns in oil projects on the Norwegian Continental Shelf. Secondly, to evaluate the value loss resulting from these overruns and production shortfalls.

In the first chapter, an introduction to the topic was presented, along with a review of previous works and an outline of the structure of the thesis. The second chapter summarized the Act of 1996 regarding petroleum activities on the NCS, explained the phases of development projects, addressed forecasting uncertainties, and described the methods used for calculating Present Value (PV) and Correlation Coefficients. Chapter three focused on the databases, data categories, and limitations of the thesis. In chapter four, the analyses and results were elaborated upon. Chapter five provided a discussion of the results, explored the reasons behind poor forecasts, and suggested ways to enhance forecasting accuracy. Finally, this concluding chapter summarizes the findings based on the results obtained throughout the thesis.

The analyses conducted in this study revealed a consistent pattern of underestimation in the forecasts of time and cost for development projects, while the production forecasts tended to be higher than the actual deliveries from the fields. Furthermore, the findings indicated a lack of improvement over time in the accuracy of cost and schedule forecasts. However, a slight improvement was observed in the forecasts of production values when considering the first ten years of production for the oil fields.

In this study, the actual and forecasted values of development time, cost, and production were compared to calculate the value losses and percentages of schedule and cost overruns, as well as production shortfalls. The results revealed an average development delay of 101 days or 12%, cost overruns amounting to NOK 268 billion or 12.7%, and an average underproduction of 50 Million Sm3 or 5% (based on data from 67 oil fields with production between 1995 and 2021 using Method2).

Two methods were employed in this study to calculate the present value losses resulting from inaccurate forecasts. Method 2 focused on underproduction, while Method 1 considered a combination of underproduction and delays. The value loss attributed to delays was determined to be NOK 39 billion, while the value loss associated with underproduction amounted to NOK 80 billion in 2022-NOK. When combined with the value loss caused by cost overruns, the total value loss due to poor forecasts in petroleum projects on the NCS reached NOK 387 billion.

To enhance the accuracy of forecasts in the oil and gas industry, it is important to address the underlying reasons for biased forecasts. These reasons can be categorized into three main groups: delusion, deception, and bad luck. Delusion refers to unrealistic judgments and underestimations of project outcomes caused by an inside view perspective. Deception occurs when projects are presented in a more favorable light than they actually are, often resulting from misleading strategic information. Bad luck encompasses factors that are beyond the control of managers and planners, making them difficult to predict or avoid. To overcome these biases and improve forecasts, a combination of approaches can be implemented: 1) decision-makers should be trained and educated about the biases and uncertainties involved in their decision-making processes. This helps in debiasing their judgments and leads to more informed and accurate forecasts, 2) Leveraging the expertise of superforecasters, who have demonstrated a better understanding of unpredictable future outcomes, can also be beneficial, 3) adopting the Reference Class Forecasting (RCF) method, which incorporates historical data and an outside view perspective, can contribute to improved forecasts. By considering past project performance and comparing it to similar projects, a more realistic forecast of future outcomes can be achieved.

The relationship between the cumulative underproductions during the first 10 years and the total cost overruns was explored using regression methods, namely the Pearson, Spearman, and SVR methods. The average correlation coefficients obtained from the

Pearson, Spearman, and SVR methods were 0.28, 0.08, and 0.41, respectively. The results revealed a weak linear correlation between the variables, a weaker linear relationship between the ranks of the data, and a moderate non-linear correlation between the two. These findings indicate that there is some degree of association between cumulative underproductions and total cost overruns, with the SVR method showing a slightly stronger non-linear correlation compared to the linear correlations observed with the other methods.

Overall, the findings of this study are consistent with previous research on project overruns and production shortfalls, highlighting the need for improved forecasting to mitigate value losses. This study stands out by incorporating a broader range of data, including a larger number of oil fields and longer production periods (10 years), to enhance the understanding of the dynamics within petroleum projects on the NCS. Additionally, by considering data until 2023, this work encompasses the most up-to-date information from oil fields on the NCS.

Finally, it is recommended that production forecasts, similar to development cost and time forecasts, become publicly available. By making production forecasts publicly accessible, it not only encourages operators to enhance their forecasting accuracy but also stimulates further research on this topic.

References

[1] B. Flyvbjerg, "Over budget, over time, over and over again: Managing major projects," 2013.

[2] R. B. Bratvold, E. Mohus, D. Petutschnig, and E. Bickel, "Production Forecasting: Optimistic and Overconfident—Over and Over Again," *SPE Reservoir Evaluation & Engineering,* vol. 23, no. 03, pp. 0799-0810, 2020, doi: 10.2118/195914-pa.

[3] E. Nesvold and R. B. Bratvold, "Debiasing probabilistic oil production forecasts," *Energy,* vol. 258, p. 124744, 2022/11/01/ 2022, doi: https://doi.org/10.1016/j.energy.2022.124744.

[4] E. Nesvold and R. Bratvold, "Field Features Do Not Explain Greenfield Production Forecasting Bias," *SPE Journal,* pp. 1-18, 2022.

[5] NPD. "Facts." Norwegian Petroleum Directorate. https://www.npd.no/en/facts/ (accessed.

[6] E. Mohus, "Over Budget, Over Time, and Reduced Revenue, Over and Over Again – An Analysis of the Norwegian Petroleum Industry's Inability to Forecast Production," Master, Finance and Risk Management, University of Stavanger, Stavanger, Norway, 2018. [Online]. Available: http://hdl.handle.net/11250/2566243

[7] B. Flyvbjerg, M. Garbuio, and D. Lovallo, "Delusion and deception in large infrastructure projects: two models for explaining and preventing executive disaster," *California management review,* vol. 51, no. 2, pp. 170-194, 2009.

[8] B. Flyvbjerg, "From Nobel Prize to project management: Getting risks right," *Proj Manag J,* vol. 37, no. 3, pp. 5-15, 2006.

[9] Norwegian Petroleum. "GLOSSARY." https://www.norskpetroleum.no/en/glossary/ (accessed.

[10] B. Flyvbjerg, M. S. Holm, and S. Buhl, "Underestimating costs in public works projects: Error or lie?," *Journal of the American planning association,* vol. 68, no. 3, pp. 279-295, 2002.

[11] P. Hall and P. G. Hall, *Great planning disasters*. Univ of California Press, 1982.

[12] P. K. Gellert and B. D. Lynch, "Mega-projects as displacements," *International Social Science Journal,* vol. 55, no. 175, pp. 15-25, 2003, doi: https://doi.org/10.1111/1468-2451.5501002.

[13] E. W. Merrow, *Industrial Megaprojects: Concepts, Strategies, and Practices for Success*. John Wiley & Sons, 2011, p. 384.

[14] A. Warrack, "Resource Mega Project Analysis and Decision Making. Victoria, BC: Institute for Research on Public Policy," *Western Resources Program,* 1985.

[15] X. Ren and L. Weinstein, "Urban governance, mega-projects and scalar transformations in China and India," in *Locating Right to the City in the Global South*. New York: Routledge, 2013, pp. 107-126.

[16] P. NIJKAMP and B. UBBELS, "HOW RELIABLE ARE ESTIMATES OF INFRASTRUCTURE COSTS? A COMPARATIVE ANALYSIS," *International Journal of Transport Economics / Rivista Internazionale Di Economia Dei Trasporti,* vol. 26, no. 1, pp. 23-53, 1999. [Online]. Available: http://www.jstor.org/stable/42747355.

[17] C. Koch, "Contested overruns and performance of offshore wind power plants," *Constr Manag Econ,* vol. 30, no. 8, pp. 609-622, 2012.

[18] A. Ansar, B. Flyvbjerg, A. Budzier, and D. Lunn, "Should we build more large dams? The actual costs of hydropower megaproject development," *Energy policy,* vol. 69, pp. 43-56, 2014.

[19] J.-E. Nilsson, "The Weak Spot of Infrastructure BCA: Cost Overruns in Seven Road and Railway Construction Projects," *Journal of Benefit-Cost Analysis,* vol. 13, no. 2, pp. 224-246, 2022.

[20] A. A. Chadee, I. Ray, and X. T. Chadee, "Systemic issues influencing technical certainty in social housing programmes in a small island developing state," *Buildings-Basel,* vol. 11, no. 2, p. 65, 2021.

[21] L. Hudson, G. Ferres, R. Macartney, J. Polglase, and P. O'Flaherty. How better project management can boost mining's capital productivity [Online] Available: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/mining-metals/ey-how-better-project-management-can-boost-minings-capital-productivity-final.pdf?download

[22] M. Emhjellen, K. Emhjellen, and P. Osmundsen, "Cost estimation overruns in the North Sea," *Proj Manag J,* vol. 34, no. 1, pp. 23-29, 2003.

[23] N. P. Directorate, "Evaluation of projects implemented on the Norwegian shelf," *Norwegian Petroleum Directorate,* 2013.

[24] E. W. Merrow, "Oil and gas industry megaprojects: Our recent track record," *Oil and Gas Facilities,* vol. 1, no. 02, pp. 38-42, 2012.

[25] Ernst & Young. *How do you reshape when today's future may not be tomorrow's reality?, Oil and Gas Digital Transformation and the Workforce Survey 2020*, 2020.

[26] R. E. Dahl, S. Lorentzen, A. Oglend, and P. Osmundsen, "Pro-cyclical petroleum investments and cost overruns in Norway," *Energy Policy,* vol. 100, pp. 68-78, 2017/01/01/ 2017, doi: https://doi.org/10.1016/j.enpol.2016.10.004.

[27] A. Haukaas and E. Mohus, "The petroleum industry's inability and lack of will to fulfil its ptomises - an analysis of overruns on the Norwegian Continental Shelf," Bachelor, Insituttet for Petroleumsteknologi, University of Stavanger, Stavanegr, 2016.

[28] R. B. Bratvold and A. Sheikhoushaghi, "How far PDO Estimates are from the Actual Costs," 2022.

[29] T. S. Jayne and S. Rashid, "The Value of Accurate Crop Production Forecasts," 2010.

[30] I. G. K. Ansah, M. Appiah-Twumasi, and F. Tsiboe, "Starchy staples production shortfalls in Ghana: Technical inefficiency effects outweigh technological differences across ecologies," *Plos one,* vol. 18, no. 4, p. e0284780, 2023.

[31] S. Sulaeman, E. Brown, R. Quispe-Abad, and N. Müller, "Floating PV system as an alternative pathway to the amazon dam underproduction," *Renewable and Sustainable Energy Reviews,* vol. 135, p. 110082, 2021.

[32] F. Demirmen, "Reserves Estimation: The Challenge for the Industry," *Journal of Petroleum Technology,* vol. 59, no. 05, pp. 80-89, 2007, doi: 10.2118/103434-jpt.

[33] F. Demirmen, "Reliability and Uncertainty in Reserves: How the industry Fails, and a vision for improvement," in *SPE Hydrocarbon Economics and Evaluation Symposium*, 2005: OnePetro.

[34] N. S. Nandurdikar and L. Wallace, "Failure to Produce: An Investigation of Deficiencies in Production Attainment," 2011.

[35] R. B. Bratvold and S. H. Begg, *Making Good Decisions*. Society of Petroleum Engineers, 2010.

[36] B. Flyvbjerg, "What you Should Know about Megaprojects and Why: An Overview," *Proj Manag J,* vol. 45, no. 2, pp. 6-19, 2014, doi: 10.1002/pmj.21409.

[37] NPD. "

Act 29 November 1996 No. 72 relating to petroleum activities." https://www.npd.no/en/regulations/acts/act-29-november-1996-no2.-72-relating-to-petroleum-activities/#Section-1-1 (accessed.

[38] NPD. "THE PETROLEUM ACT AND THE LICENSING SYSTEM." https://www.norskpetroleum.no/en/framework/the-petroleum-act-and-the-licensing-system/ (accessed.

[39] NPD. "What do you know about the Norwegian Petroleum Directorate - the NPD's duties and roles." https://www.npd.no/globalassets/1-npd/publikasjoner/ressursrapport-2017/pdfer-norsk/tildelingsbrev-til-oljedirektoratet-for-2017.pdf (accessed.

[40] (2022). *Guidelines for plan for development and operation of a petroleum deposit (PDO) and plan for installation and operation of facilities for transport and utilisation of petroleum (PIO)*.

[41] C. Price, *Time, discounting and value*. Blackwell Publishers, 1993.

[42] S. A. Broverman, *Mathematics of investment and credit*. Actex Publications, 2010.

[43] The World Bank Data. "Inflation, consumer prices (annual %) - Norway." https://data.worldbank.org/indicator/FP.CPI.TOTL.ZG?end=2021&locations=NO&start=1990 (accessed.

[44] NPD. "Conversion factors." https://www.npd.no/en/about-us/use-of-content/conversion-table/ (accessed.

[45] N. Bank. "EXCHANGE RATES." Norges Bank. https://data.norges-bank.no/api/data/EXR/B.USD.NOK.SP?format=excel-both&apisrc=nbi&startPeriod=1994-01-01&endPeriod=2022-12-31&locale=en (accessed.

[46] Energy Information Administration (EIA). "PETROLEUM & OTHER LIQUIDS." EIA. https://www.eia.gov/dnav/pet/hist/LeafHandler.ashx?n=PET&s=RBRTE&f=A (accessed.

[47] I. Maryati, T. Sumartini, and D. Sofyan, "Experiences of Pearson formula in analysis regression," in *IOP Conference Series: Materials Science and Engineering*, 2021, vol. 1098, no. 3: IOP Publishing, p. 032088.

[48] J. M. Stanton, "Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors," *Journal of Statistics Education,* vol. 9, no. 3, 2001.

[49] H. Eliufoo, "Project team perception of causes of building cost overruns: A reflection in Tanzania," *Journal of Civil Engineering and Architecture,* vol. 11, no. 1, pp. 149-158, 2017.

[50] J. C. Parcon, *Spearman rank regression*. Western Michigan University, 2003.

[51] V. Vapnik, S. Golowich, and A. Smola, "Support vector method for function approximation, regression estimation and signal processing," *Advances in neural information processing systems,* vol. 9, 1996.

[52] A. Tharwat, "Parameter investigation of support vector machine classifier with kernel functions," *Knowledge and Information Systems,* vol. 61, pp. 1269-1302, 2019.

[53] E. Vinagre, T. Pinto, S. Ramos, Z. Vale, and J. M. Corchado, "Electrical energy consumption forecast using support vector machines," in *2016 27th International Workshop on Database and Expert Systems Applications (DEXA)*, 2016: IEEE, pp. 171-175.

[54] J. Zeng, Z.-H. Tan, T. Matsunaga, and T. Shirai, "Generalization of Parameter Selection of SVM and LS-SVM for Regression," *Machine Learning and Knowledge Extraction,* vol. 1, no. 2, pp. 745-755, 2019.

[55] M. Welling, "Support vector regression," *Department of Computer Science, University of Toronto, Toronto (Kanada),* 2004.

[56] NPD. "FACTPAGES Norwegian Petroleum Directorate." NPD. https://factpages.npd.no/en/field/TableView/Production/Saleable/TotalNcsYear (accessed.

[57] J. Franc-Dąbrowska, M. Mądra-Sawicka, and A. Milewska, "Energy Sector Risk and Cost of Capital Assessment—Companies and Investors Perspective," *Energies,* vol. 14, no. 6, p. 1613, 2021. [Online]. Available: https://www.mdpi.com/1996-1073/14/6/1613.

[58] Vår Energi, "Third quarter 2022 Interim report," 2022. [Online]. Available: https://mb.cision.com/Public/21086/3654171/987f480d256ab9c4.pdf

[59] OKEA, "2022 Q3 Quarterly report," 2022. [Online]. Available: https://www.okea.no/wp-content/uploads/2022/07/q3-2022-quartely-report.pdf

[60] Gurufocus. "Norwegian Energy Co ASA (FRA:CJ1) WACC %." https://www.gurufocus.com/term/wacc/FRA:CJ1/WACC-Percentage/Norwegian-Energy-Co-ASA?&frequency=A (accessed.

[61] Equinor. "2021 Annual report and Form 20-F." https://www.sec.gov/Archives/edgar/data/1140625/000114062522000044/eqnr20211231.pdf (accessed.

[62] NPD, "Project execution on the Norwegian continental shelf," Oslo, Norway, 2020. [Online]. Available: https://www.aph.gov.au/DocumentStore.ashx?id=7c03414b-e040-4a92-8bb3-d6273e52abba

[63] T. Kahneman, "D. kahneman, a. tversky," *Prospect theory: An analysis of decisions under risk,* pp. 263-291, 1979.

[64] D. Lovallo and D. Kahneman, "Delusions of success," *Harvard business review,* vol. 81, no. 7, pp. 56-63, 2003.

[65] M. Welsh, S. Begg, and R. Bratvold, "SPE 102188: Correcting common errors in probabilistic evaluations: efficacy of debiasing," in *Society of Petroleum Engineers 82nd Annual Technical Conference and Exhibition., Dallas, Texas, USA*, 2006.

[66] P. Tetlock and D. Gardner, "Superforecasting: The Art and Science of Prediction, 2015 ed," ed: Broadway Books, New York, NY, 2015.

[67] N. Silver, *The signal and the noise: the art and science of prediction*. Penguin UK, 2012.