The Thesis Committee for David Petutschnig
Certifies that this is the approved version of the following Thesis:

An analysis of the petroleum industry's inability to deliver on early production forecasts: Shortcomings in probabilistic modeling

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor: ______________________________
Larry W. Lake

Co-supervisor: ______________________________
J. Eric Bickel

An analysis of the petroleum industry's inability to deliver on early production forecasts: Shortcomings in probabilistic modeling

by

David Petutschnig

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Engineering

The University of Texas at Austin

December 2019

Acknowledgements

I am grateful for the support and guidance of my supervisors Professor Larry Lake and Professor Eric Bickel. I further want to thank Professor Reidar Bratvold from the University of Stavanger for his insights and advice. As my time at the University of Texas at Austin comes to an end, I want to express my gratitude to all my friends and colleagues that have made this educational experience truly unique. Lastly, I thank my family for their unconditional support and love throughout the years.

Abstract

An analysis of the petroleum industry's inability to deliver on early production forecasts: Shortcomings in probabilistic modeling

David Petutschnig, M.S.E.

The University of Texas at Austin, 2019

Supervisor: J. Eric Bickel

Previous studies have confirmed that production forecasts in the oil and gas industry are exposed to a variety of biases. This thesis extends those previous findings by investigating the quality of production forecasts for oil fields on the Norwegian Continental Shelf, which were approved between 1995 and 2017. The research focuses on optimism and overconfidence biases.

Both biases are observable in the production forecasts provided by the Norwegian Petroleum Directorate. By comparing annual production data with production forecasts, it is possible to draw conclusions pertaining to the quality of those forecasts. A variety of methods are applied to investigate and illustrate the magnitude of those biases. The findings illustrate that the reason operators do not attain set project goals is because of aforementioned biases rather than unexpected events. The systemic inability to deliver on what was promised is observable through the lack of forecasting quality improvement over time.

Two correction processes are proposed to reduce the encountered biases. A reference class is established to put past outcomes in a distributional setting. Uplift and scaling factors are drawn from the class to adjust the biased production forecasts. The results show a clear improvement in the quality of production forecasts through the use of reference class forecasting. A second process is introduced in which a Bayesian framework is suggested to calculate updated production forecasts. The same reference class is used to provide a prior distribution, which is then updated by the initial forecast (signal) to determine a posterior distribution. The posterior distribution exhibts on average a greater variance and a lower mean than the initial forecast. Therefore, the updated production forecasts are better calibrated and the impact of the biases is reduced.

Limitations arise regarding the availability of additional data, however preliminary results from the analyses are encouraging. Drawing on past experience to debias production forecasts is of paramount importance.

Table of Contents

List of Tables

List of Figures

1. Introduction

OVERVIEW AND MOTIVATION

Oil and gas projects are typically characterized by their capital intensity as well as their longevity. Knowing how much and when hydrocarbons will be produced is important as it provides the basis for any cash flow calculations. Hence, an integral part of the Final Investment Decision (FID) for any oil and gas project is the production forecast. The production forecast should be the result of a mindful assessment of the surface and subsurface conditions - i.e. reservoir properties, geological factors, well completion constraints etc.

Oil and gas companies are usually interested in three values obtained from such a production forecast. The three values are a low estimate, a mean assessment (or a P50) and a high estimate. Uncertainty ranges in the oil and gas industry are typically given by a low value and a high value. In this study the definitions set forth by the Norwegian Petroleum Directorate will be utilized, which specify the uncertainty ranges as follows: The low estimate is represented by the P10 value, and the high estimate by the P90 value. For a continuous distribution, the P10 value marks the point for which 10% of the observations will be lower than the P10. The P90 value marks the point for which 90% of the observations will be lower than the P90. The quality of those production forecasts will be the subject of this study.

The industry is plagued by frequent schedule delays and cost overruns.[1] Observed time and cost results tend to deviate from forecasts systemically in one direction, indicating

[1] (Welsh et al. 2005), (Ernst & Young 2014),

the presence of biases. Those deficiencies are also observable in other uncertainty assessments of oil and gas projects, such as production forecasts.[2]

Various research groups have found substantial shortcomings in the petroleum industry's ability to produce unbiased probabilistic forecasts.[3] It was observed that the forecasts are optimistic in terms of expected production and overconfident regarding the range of possible outcomes. There are many reasons why production forecasts consistently miss their estimates, which we will elaborate on in one of the subsequent chapters. Given the pivotal role of future production values, it is consequential to suggest the need to revise those forecasts and highlight the shortcomings with hopes that the industry will adapt and improve.

OBJECTIVES

The objectives for this research thesis are threefold. Chapter 2 briefly discusses previous research on biases in infrastructure projects. Chapters 3 and 4 introduce the data set and applied methodology used in this thesis. The first objective is to investigate whether the shortcomings in probabilistic production forecasts found by other research groups still persist today. Therefore, chapter 5 analyzes and assesses production forecasts and actual production data from the Norwegian Continental Shelf (NCS) pertaining the quality of the production forecasts. Extending the analysis beyond the base case comparison and having access to low and high production estimates allow for an exhaustive study about the quality of probabilistic forecasts. The second objective is to analyze and illustrate any biases found. Chapter 6 introduces previously suggested processes to reduce encountered biases. The last

[2] (Mohus 2018)
[3] (Welsh et al. 2005), (Nandurdikar and Wallace 2011), (Flyvbjerg et al. 2014)

objective is to propose methods on how to improve probabilistic production forecasting in the oil and gas industry. Chapters 7 and 8 demonstrate two different approaches on how to debias forecasts.

2. Literature review

SHORTCOMINGS IN LARGE INFRASTRUCTURE PROJECTS

Large, capital-intensive projects are prone to cost and time overruns. The examples are abundant, and the incurred value loss can be substantial.[4] A study by Flyvbjerg and COWI highlighted significant deviations between forecasts and outcomes for large infrastructure projects.[5] The study concluded that nine out of ten infrastructure projects have experienced cost overruns to varying extent. Cost escalation was observed in different industry sectors including roads, rail, energy and others.[6] Infamous instances of project disasters are found across industries and countries with no immediate improvement over time.[7] For example: The Sydney Opera House was completed ten years behind schedule. The scaled-down version was over-budget by $95 million, against an original estimate of $7 million.[8] The Channel tunnel, connecting France and the United Kingdom, was over-budget by 80% and forecasted revenues were halved upon completion.[9]

There are many reasons why forecasted project goals were delayed or not attained at all. However, the research in this thesis will focus on biases, rather than computational errors. Flyvbjerg et al. ascribe the causes for those shortcomings to two predominant categories of biases.[10] They differentiate between delusional and deceptive biases, both of which will be discussed in chapter 5.[11]

[4] (Priemus et al. 2008), (Cantarelli et al. 2012)
[5] (Flyvbjerg et al. 2004)
[6] (Buhl et al. 2003)
[7] (Buhl et al. 2002), (Flyvbjerg et al. 2005)
[8] (Flyvbjerg 2014)
[9] (Moore 2010)
[10] (Flyvbjerg 1996), (Buhl et al. 2002)
[11] (Flyvbjerg et al. 2014)

There is strong evidence that those shortcomings are not confined to the public sector. Projects in the private sector are similarly exposed to cost and time overruns.[12]

The upstream oil and gas sector is a project-based industry, where individual projects typically compete against each other for internal funding. Those circumstances make the oil and gas industry an ideal environment for a variety of biases to occur.

BIASES IN THE OIL AND GAS INDUSTRY

The petroleum industry's susceptibility to biases has been the focus of several research groups.[13] A seemingly counterintuitive finding predicates that even experts which specialized knowledge exhibit a susceptibility towards biases.[14] Industry experience alone offers limited help in avoiding these biases. Welsh et al. published a research paper in 2005, in which the potential of reducing biases is highlighted if a person undergoes risk training, which will be elaborated on in chapter 6.[15]

The financial impact of such biases can be substantial. Findings by Welsh et al. in 2007, indicate that biased input parameters provide erroneous uncertainty distributions not only for those input parameters, but also for the resulting outputs.[16] The net present value (NPV) is a commonly used financial metric in the oil and gas industry. Like other key performance indicators, NPV is determined using a variety of inputs, which are uncertain. Welsh et al. demonstrate that the NPV calculated from biased input parameters is lower, compared to the NPV calculated from unbiased input parameters. In one of the illustrated examples the technical reserves were estimated at 360 million barrels of oil, with an

[12] (Ernst & Young 2014)
[13] (Hawkins et al. 2002), (Welsh et al. 2005), (Welsh et al. 2007)
[14] (Welsh et al. 2005)
[15] Ibid.
[16] (Welsh et al. 2007)

expected value of 346 million USD for the project. Figure 1 shows the impact overconfidence levels can have on the project NPV. A probability distribution is said to be overconfident if the range of distributed values is too small. In Figure 1, the technical reserves remain constant at 360 million barrels with varying degrees of overconfidence (the actual reserves are a function of economics and would therefore change).

While real-life examples may exhibit different responses to overconfidence levels, the discussed case demonstrates the possibility of value erosion. In some instances, it could potentially lead to the approval of a project which - if the input parameters were unbiased - would have yielded a negative NPV.

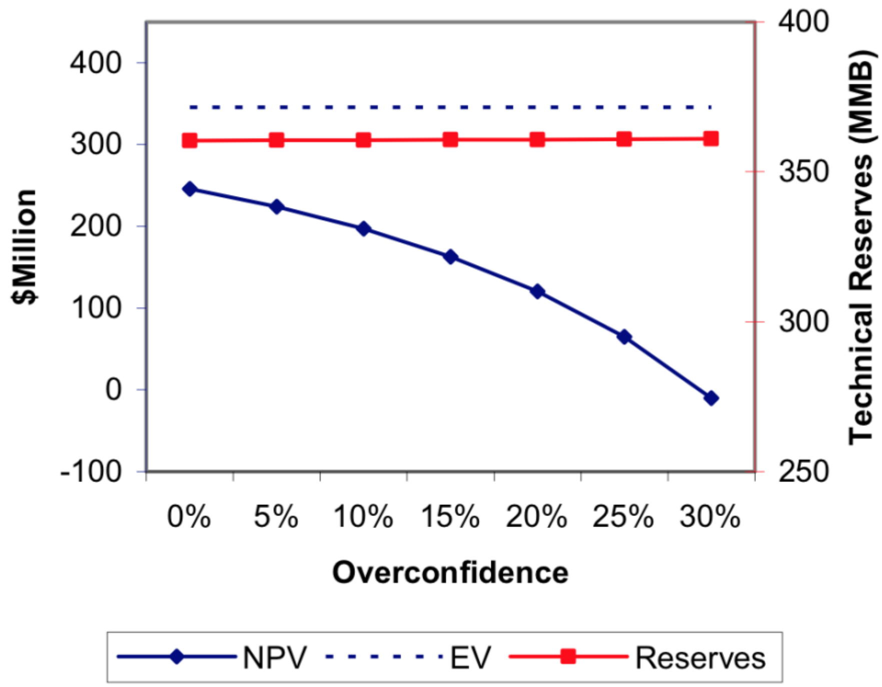


Figure 1. NPV change with varying degree of overconfidence (Welsh et al. 2007)

Welsh et al. focused on the effect of the overconfidence bias, the trust heuristic bias, and the availability bias. However, subsequent research has confirmed that the industry seems to be affected by an even broader range of biases.[17]

The financial ramifications of different biases vary depending on the type of bias encountered. A study by Mohus in 2018 examined 56 oil fields regarding lost value due to cost and time overruns as well as underproduction.[18] The accrued value erosion due to biased production forecasts was estimated at 56.8 billion USD for those fields. While the impact of overconfidence production forecasts appears to be substantial, other biases in the oil and gas industry seem to be less damaging. In 2008, Begg and Bratvold published the results of a study that investigated the errors associated with project and portfolio selection. They concluded that the impact of the selection bias, relative to other biases in the oil and gas industry, seems comparatively small.[19] It is worth mentioning, that the selection bias is potentially more severe if the initial NPV estimates carry more uncertainty, which would be reflective in the uncertainty ranges of the estimates.

[17] (Welsh and Begg 2015), (Mohus, 2018)
[18] (Mohus, 2018)
[19] (Begg and Bratvold 2008)

3. Data

ATTRIBUTES OF PROBABILISTIC FORECASTS

Future production is unknown and, therefore their forecasts are associated with varying degrees of uncertainty. Best practice stipulates that probabilistic forecasts are generated to assess the uncertainties. There are several characteristics of probabilistic production forecasts that are of interest to this study. A set of criteria must be met to consider a probability distribution as well-calibrated. A well-calibrated production forecast is unbiased and consistent with the forecasters' knowledge.

First, the lower (P10) and upper (P90) percentiles must be calculated, so that 80% of the time, the actual production outcomes fall within the range set by the two values. If this is not the case, the confidence interval is either too narrow or too wide, the former being more common in the production forecasting context and is denoted as overconfidence.[20]

The forecasted P50 of a well-calibrated probability distribution specifies the value at which 50% of the actual outcomes exceed the P50 and half of the actual outcomes fall short of the P50. If more than half the actual values are greater or smaller than the P50, the probability distribution is either optimistic or pessimistic. An illustration of a biased forecast can be found in Figure 2. This biased forecast has a higher mean and a narrower distribution.

[20] (Welsh et al. 2005)

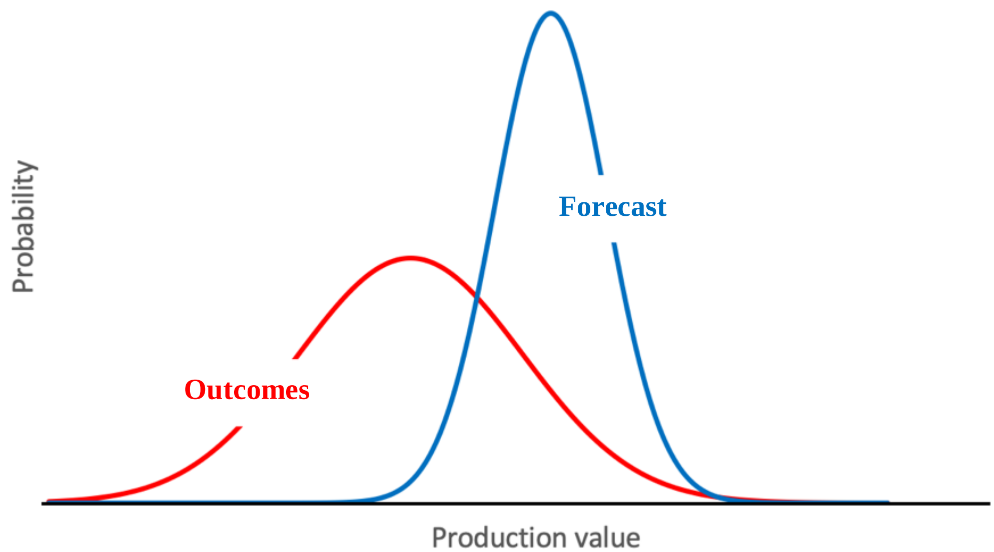


Figure 2. Overconfident and optimistic production forecast vs. actual production outcomes

DATA PRE-TREATMENT

The data set utilized for the analysis in this research project comprises of production forecasts for 56 oil fields, located on the NCS. The production forecasts consist of a P10, a mean and a P90 estimate (rather than a continuous distribution) for every production year for every field. The forecasts were made before the time of the FID. Those fields reached their FID and were approved between 1995 and 2017. The production data for any fields on the NCS is publicly available through the Norwegian Petroleum Directorate's (NPD) website.[21] However, the production forecasts for those fields are not publicly accessible and therefore they were provided exclusively for research purposes under the condition that they are made anonymous.

[21] (Norwegian Petroleum Directorate 2019)

The NPD specifies that the operator must report probabilistic values on all petroleum resources in a way such that the uncertainties "*... shall to the extent possible be designated by P10 - P expected - P90.*"[22] It is plausible to assume that a majority of the production forecasts were generated using a base case approach. The base case is calculated using inputs and assumptions that are consider most likely to occur. The best estimate is defined by the NPD as the "*best estimate of petroleum volumes that are expected to be recovered from a project*" and "*If the best estimate is determined by a stochastic method, the best estimate shall be considered as the expected value*".[23] In some cases P10 and P90 might have been calculated directly from the base case, i.e. using a multiplier, rather than being drawn from a continuous probability distribution. However, for the purpose of this study it is assumed that the reported values were determined from continues probability distributions and thus can be evaluated using attributes of probabilistic forecasts.

The forecasts are part of the required documentation for the Plan of Development, requested by the NPD. Four values are of interest for the subsequent analysis, namely the actual production and the three forecasted values: the P10, mean, and P90. In some instances, the P50 rather than the mean will be used. For most fields, the forecasted mean is close but not equal to the P50. The implications of the inequality will be discussed in Chapter 4. Only some minor adjustments to the data set are necessary, since the data has already been used in a previous research project.[24]

[22] (Norwegian Petroleum Directorate 2018a),(Norwegian Petroleum Directorate 2018b)
[23] (Consortium 2016)
[24] (Mohus, 2018)

START-UP DELAYS

Mohus (2018) found that the average initial production delay between intended start-up and actual start-up for this data set is 202 days.[25] With an estimated mean development time of 2 years and 8 months, the time overrun averages at around 20% for the study period.[26] To avoid confounding the impact of poor production forecasts with the impact of time delays, the first actual production year was set equal to the first forecasted production year. Table 1 shows the estimates (P10, mean and P90) and the actual production for one of the fields on the NCS, using anonymized data.

Estimates			Actual Production
P10	Mean	P90	
0.58	0.59	0.60	0.00
2.26	2.28	2.30	0.00
2.22	2.27	2.29	0.29
1.94	2.22	2.24	2.20
1.63	1.98	2.26	2.85
1.16	1.76	2.08	0.57
0.81	1.52	1.93	1.40
0.71	1.34	1.78	2.31
0.48	1.12	1.62	2.80
0.34	0.77	1.47	2.63
0.26	0.54	1.33	2.49
0.20	0.40	1.21	2.02
0.17	0.32	1.01	1.96
0.13	0.24	0.70	1.92
0.10	0.19	0.52	1.78
0.07	0.16	0.40	1.30

Estimates			Actual Production
P10	Mean	P90	
0.58	0.59	0.60	0.29
2.26	2.28	2.30	2.20
2.22	2.27	2.29	2.85
1.94	2.22	2.24	0.57
1.63	1.98	2.26	1.40
1.16	1.76	2.08	2.31
0.81	1.52	1.93	2.80
0.71	1.34	1.78	2.63
0.48	1.12	1.62	2.49
0.34	0.77	1.47	2.02
0.26	0.54	1.33	1.96
0.20	0.40	1.21	1.92
0.17	0.32	1.01	1.78
0.13	0.24	0.70	1.30
0.10	0.19	0.52	
0.07	0.16	0.40	

Table 1. Correction for start-up delays for one of the fields on the NCS, using anonymized data

[25] (Mohus 2018)
[26] (Haukaas and Mohus 2016)

On the left side the actual production is unchanged, and on the right side the actual production is corrected for start-up delays. It can be assumed that the forecaster considered the first forecasted production year as the first actual production year. In 2017, the 56 fields had a total of 603 forecasted production years. The amended data set (corrected for start-up delays) comprises of 55 fields and 549 production years, which are used for further analysis.

4. Methodology – Probability distribution fitting

LOGNORMAL FITTING APPROACH

A key goal of this study is to assess the quality of production forecasts provided by the operators. It is assumed that the operators have provided production estimates as suggested by the NPD guidelines, i.e., P10/Mean/P90 values for each year. Production forecasts will be investigated for cumulative production years as well as individual production years. Cumulative production forecasts shall be denoted as production forecasts on an *aggregated basis*, which consider production values from previous years. Individual production forecasts shall be denoted as production forecasts on an *individual basis*, which is the forecast for any specific year of interest.

The cumulative mean production forecast (μ_n) of a field for any production year *n* is calculated by simply adding the mean forecasts (μ_i) of the previous years.

$$\mu_n = \sum_{i=1}^{n} \mu_i$$

The resulting cumulative mean forecast can be directly compared to the cumulative production for that field. However, P10 and P90 forecasts for a given aggregation year cannot be added in the same way, as the sum of P10 values differs from the P10 of the sum.[27]

Therefore, obtaining a field's cumulative forecasted P10 or P90 for any aggregation year will require first fitting a distribution to each of that field's yearly forecasts for two out of the three forecasted values, namely P10, mean, and P90. There is no single continuous distribution that will have an exact fit for all three forecasted values, for every

[27] (Kreifeldt and Nah 1995)

given input triplet. While there might be instances where the unmatched, third estimate is equal to the equivalent value drawn from the fitted distribution, this remains the exception. As a consequence, a probability distribution was chosen that honors two out of the three forecasted values. A lognormal distribution was selected since it is bounded at zero on the low end but also allows to model potential production upside through the tail of the distribution on the high end (Figure 3). The two parameters of the lognormal distribution, the mean and the standard deviation, can be matched using any combination of two out of the three forecasted values – i.e. the P10 and mean, the P90 and mean, or the P10 and P90.

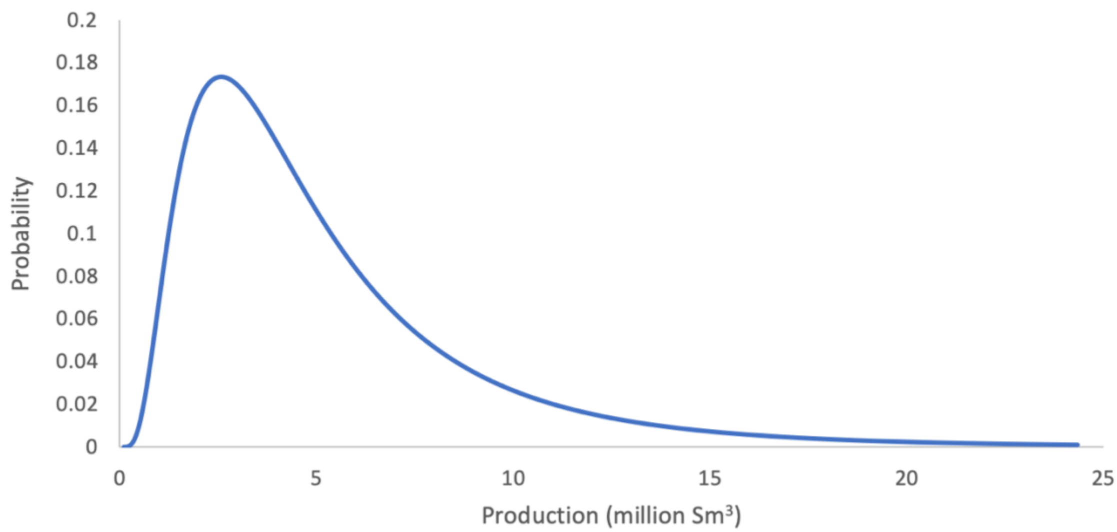


Figure 3. P10/mean lognormal fit for one field on the NCS

Selecting the P10 and mean as input values for the distribution fitting, rather than any of the other two combinations, will allow for more conclusive comparison, as the data is skewed towards the lower value (Figure 3). Using the mean as one of the fitting parameters seems appropriate as most of the time and effort goes into determining the mean forecast. Additionally, operators should pay close attention to downside risk in production forecasts, validating the choice for the P10 as one of the fitting parameters, ensuring an exact fit for the P10 production forecast. The steps of the fitting approach are outlined below.

1. The probability density function for a lognormal distribution is given by

$$f(x) = \frac{1}{x\sigma\sqrt{2\pi}} exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \qquad x > 0, \sigma > 0$$

where the logarithm of the random variable x is normally distributed

$$\ln(x) \sim N(\mu, \sigma^2)$$

with a mean (μ) and variance (σ^2) of

$$\mu = ln\left(\frac{m}{\sqrt{1 + \frac{v}{m^2}}}\right)$$

$$\sigma^2 = \ln\left(1 + \frac{v}{m^2}\right)$$

.

2. The mean (m) of the non-logarithmized sample (x) – i.e. the forecasted mean production, is used to calculate the mean (μ) and variance (σ^2) of the logarithmized sample $\ln(x)$. The two equations are:

$$m = e^{\left(\mu + \frac{\sigma^2}{2}\right)}$$

$$F_x(z) = \phi\left(\frac{\ln(z) - \mu}{\sigma}\right)$$

where z is the percentile value (i.e. -1.282 for the P10 and 1.282 for the P90) for percentile α of the lognormally distributed production forecast. The mean (m) of the production forecast and the cumulative normal distribution function (F_x) are expressed in terms of μ:

$$\mu = \ln(m) - \sigma^2/2$$

and

$$\mu = \ln(z) - \sigma\Phi^{-1}(\alpha)\ .$$

The two equations for μ are combined and the resulting equation can be solved using the quadratic formula. It will yield either zero, one or two solutions for σ. The (real) solution must be chosen, so that σ is positive and the calculated mean matches the given mean using σ:

$$\sigma^2 - 2\Phi^{-1}(\alpha)\sigma + 2[\ln(z) - \ln(m)] = 0$$

$$\sigma = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}$$

With

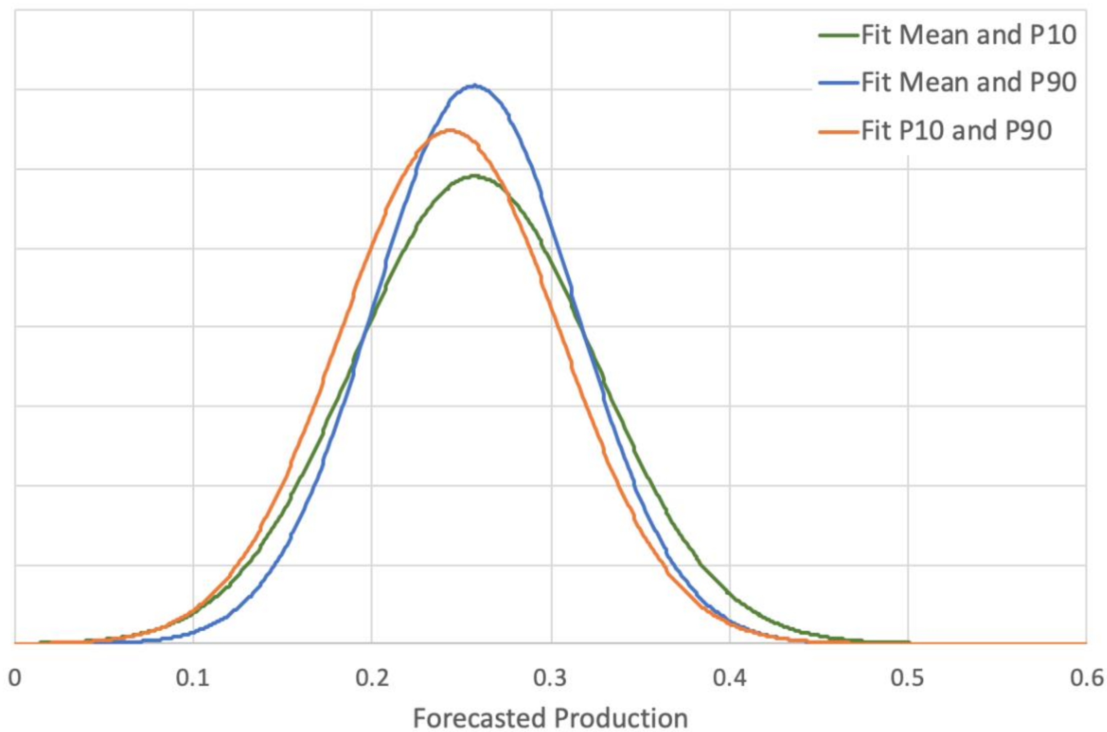
$$A = 1$$

$$B = -2\Phi^{-1}(\alpha)$$

$$C = 2[\ln(z) - \ln(m)]$$

The mean of the normal distribution is calculated using

$$\mu = \ln(m) - \sigma^2/2$$



With the mean and the standard deviation, it is possible to get an exact fit for the mean and the selected percentile and an approximate fit for the remaining percentile. The fitting method can also be used with the mean and the P90 value, as well as the P10 and P90 value, demonstrated in Appendix A-1. A comparison between different fitting methods is in Figure 4.

Figure 4. Lognormal distribution fit for one field in year 0, using P10 and mean, P90 and mean and P10 and P90.

3. Next, the mean and standard deviation of the fitted lognormal distributions now enable the summation of the production years for the individual fields, to obtain cumulative

production forecasts. Given the assumption that the production forecasts are independent, it is possible to calculate the cumulative variance (${\sigma_n}^2$) for aggregation year *n* by adding up the variances of the previous years (${\sigma_i}^2$).

$$\sigma_n^2 = \sum_{i=1}^{n} {\sigma_i}^2$$

To re-calculate the P10 and P90 of the aggregated production forecasts, values were transformed from the input *data space* (lognormal mean and variance) to the *calculation space* (non-logarithmized mean and variance) using the following two equations, which we already used previously to define the normal distribution.

$$\mu = \ln\left(\frac{m}{\sqrt{1 + \frac{v}{m^2}}}\right)$$

σ^2 is the variance of the non-logarithmized input value.

$$\sigma^2 = \ln\left(1 + \frac{v}{m^2}\right)$$

In some cases, the operator failed to provide valid P10 (and/or P90) forecasts. Such an invalid forecast would comprise of a P10 (and/or P90) value equal to the mean forecast, a P10 (and/or P90) value of zero or the absence of a P10 (and/or P90) value. In such instances those fields were omitted from the analysis. The resulting probability distributions of the aggregated production forecasts allow the identification of any statistics, including the P10, mean, P50 and P90.

SKEWNESS OF LOGNORMAL DISTRIBUTIONS

Some of the subsequent analyses use the P50, rather than the mean forecast. For most fields in this study, the mean of the forecasted aggregated production is close to the P50 of the forecasted aggregated production, yet they are still inequal. To validate the use of the P50 instead of the mean, the skewness of the distributions needs to be quantified.

The input parameters used for the lognormal distribution fitting exhibit positive skew, as the ration $(P90 - P50)/(P50 - P10) > 1$. Thus, the lognormal distributions also exhibit positive skew with $mean\ values > P50$. Pearson's second skewness coefficient was determined for each field used in the aggregation calculations. Aggregation year 4 was used as an example to demonstrate the range of skewness coefficients.

For aggregation year 4, the resulting skewness coefficients range from 0.04 and 0.8, with an average of 0.2. A skewness of 0 indicates a perfectly symmetric distribution. A rule of thumb dictates that if skewness is between -1 and -0.5 or between 0.5 and 1, the distribution is moderately skewed. If skewness is between -0.5 and 0.5, the distribution is approximately symmetric (Bulmer 1979).

With increasing number of distributions being added, the skewness reduces in accordance with the Central Limit Theorem. Positive skewness is expected in the production forecasting context as production is bounded on the low end and unbounded on the high end. With an average skewness coefficient of 0.2 for aggregation year 4 (reduced for subsequent aggregation years), evaluating biases based on P50 values - rather than means - appears valid. Despite $P50 \sim \text{mean}$, the analyses should use the mean values when convenient to honor inputs provided by operators.

5. Results and Analyses

No fields from the data set were omitted from any of the analyses unless otherwise indicated. Using the method presented in the previous chapter, a series of methods will be introduced to illustrate the performance of production forecasts.

FORECASTED PRODUCTION PROFILE VS. ACTUAL PRODUCTION PROFILE

Figure 5 shows actual production and mean forecasted production for all 55 fields, and Figure 6 shows the fields' cumulative production by year. Again, in these graphs time delays have been eliminated so that actual production start equals forecasted production start for all of the fields.

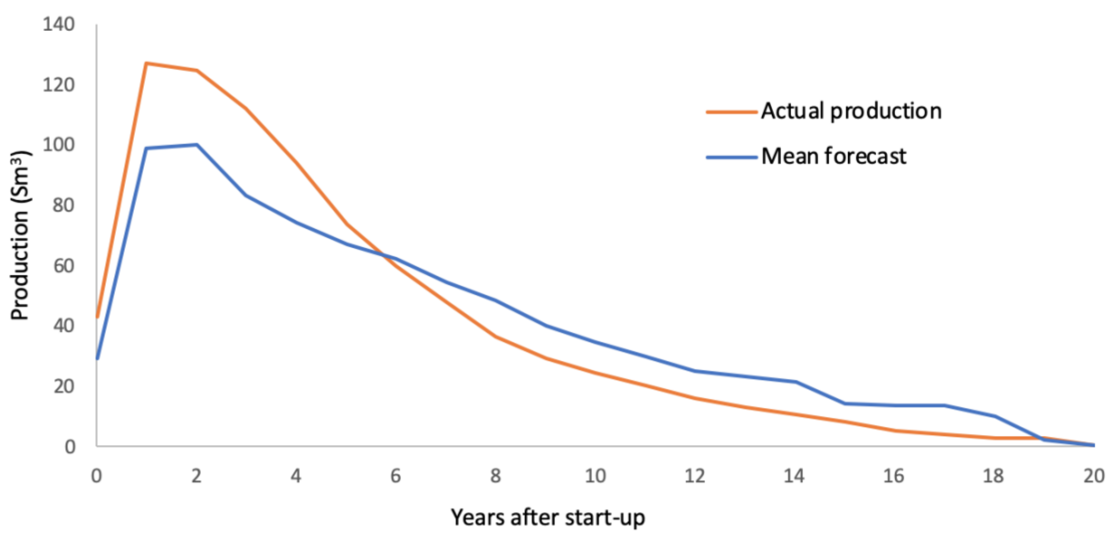


Figure 5. Actual production and mean forecasted production for all fields.

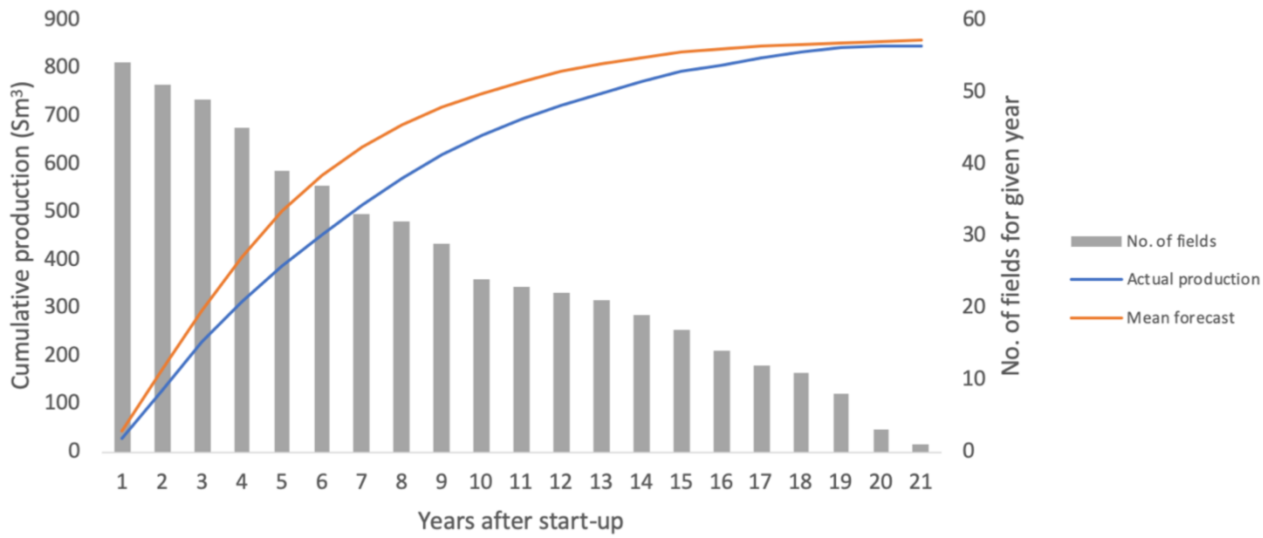


Figure 6. Cumulative actual production and cumulative mean forecasted production for all fields. The grey bars indicate the number of fields available for the analysis in each production year.

Figure 5 shows that for each of the first five years, average actual production from the 55 fields was short of average forecasted-mean production. However, from year six onward, average actual production was greater than average forecasted-mean production. Figure 6 shows that in year 20, cumulative actual production is close to cumulative forecasted-mean production. However, as will be discussed in detail later, most of the fields delivering on the expected ultimate recovery (or exceeding it) required additional investment beyond what was used in the production forecast made at the time of FID. Additional investment types included, but is not limited to: redevelopment projects, additional wells, etc. The grey bars indicate the number of fields available for the analysis for any production year.

Optimism bias

Operators are interested in production forecasts on an individual basis, when compared to actual production, to judge whether production goals were met. Likewise, the comparison between an aggregated production forecast and cumulative production will indicate whether an operator has delivered over the life of a field. Biases that are observable for both, the aggregated and individual production years, will also render the argument that individual production forecast might suffer from one-off unexpected events rather than biases, invalid.

The degree of optimism bias for any aggregated production year *n* can be assessed by using cumulative distributions of the normalized actual production of all fields. Comparing the cumulative distribution with the normalized forecasted mean production for aggregation year *n* will also demonstrate the quantitative impact of optimism. The generation of the cumulative distributions over normal actual production values was preceded by a number of steps.

1. First, the aggregated mean production forecast for year *n* has been normalized so that the mean forecast is 1.0.

2. Next, each field's actual production outcome for aggregation year *n* was normalized by the field's mean forecasts, to allow the distribution of forecasted mean production to be compared with the distribution of actual production.

3. Because the ratio of the forecasted mean production versus actual production varies by field, a Metalog distribution[28] was fit to the normalized actual production data. The Metalog distribution was chosen because it is simple, flexible and provides an easy fit.

Figure 7. CDF for normalized actual production, with the normalized mean forecast indicated. The CDF is for the aggregation year 3, including all fieldsFigure 7 displays the CDF for aggregation year three (chosen randomly) with the red line indicating the normalized mean production forecast. The horizontal axis denotes the aggregated, normalized actual production values for the aggregation year, and the vertical axis denotes

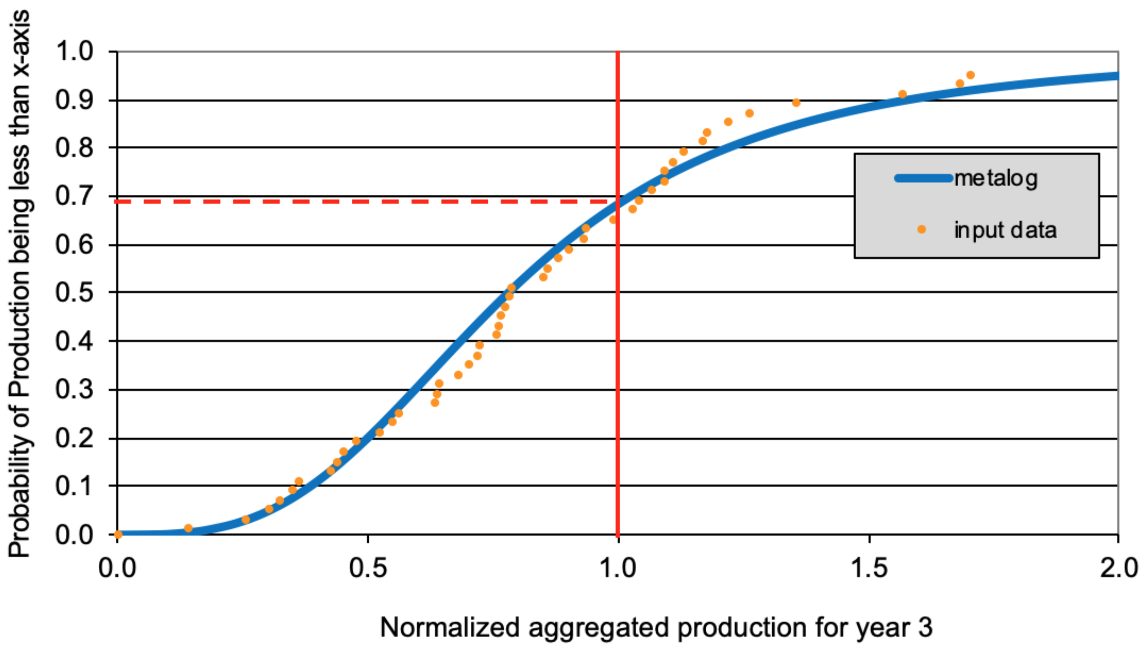


Figure 7. CDF for normalized actual production, with the normalized mean forecast indicated. The CDF is for the aggregation year 3, including all fields

the probability of the actual production being less than the value of the horizontal axis.

[28] (Keelin, 2019)

If the production forecasts were unbiased, the normalized actual production values would be less than the normalized mean production forecast approximately 50% of the time (recall $P50 \sim \text{mean}$ for aggregation years). Figure 7 shows that for aggregated year 3, optimism in the mean forecasts can be observed, as the probability of production being less than the normalized mean forecast is greater than 50%. In fact, for 69% of the fields, the normalized actual production was less than the normalized mean forecast production (x=1.0).

SENSITIVITY ANALYSIS FOR OPTIMISM BIAS

A sensible question to ask at this point is if optimism bias is also observable in other aggregation years. Figure 8 shows the results of a sensitivity analysis of the optimism bias as a function of production year. The orange dots show the percentage of fields whose cumulative actual production did not exceed the field's aggregated forecasted P50, to number of aggregation years n. Notice that for the sensitivity analysis, the P50 (determined from the fitted lognormal distribution) was used instead of the mean. Using the P50 allows for valid comparison between the number of instances where the cumulative production was less than the aggregated P50 and the aggregated P50 value itself. Sensitivity years n range between 1 and 8, because beyond year 8 fields are much more likely to be subject to redevelopment. Those redevelopments would not have been specified in the FID. The gray bars show the number of fields included as a function of n with the scale on the right-hand side of the graph. The number of fields with valid P50 forecasts decreases as n increases, because of data availability and field life. Figure 8 shows that there are no major improvements in the optimism bias throughout the years, as cumulative forecasted P50 systemically overestimate production. The aggregated forecasted P50 values are consistently above the black dotted line – the line indicating where an unbiased, well-calibrated P50 would fall.

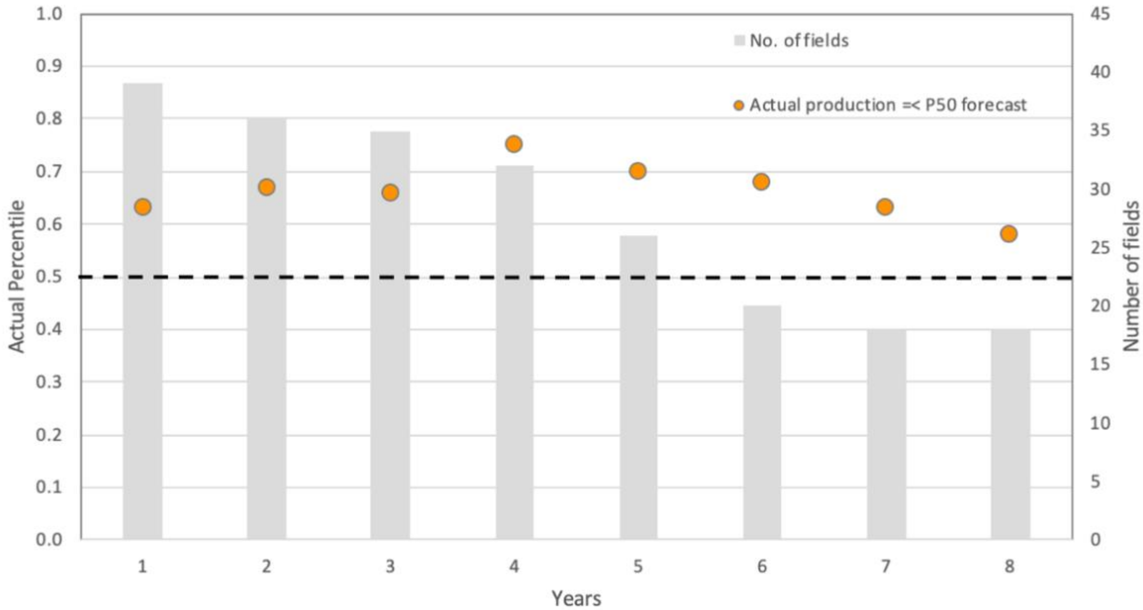


Figure 8. Sensitivity analysis of the optimism bias as a function of production year, on an aggregated basis

OVERCONFIDENCE BIAS

Recall that for a well-calibrated production forecast, the P10 and P90 values should bound the range where the actual production values fall 80% of the time. Using a similar approach as with the optimism bias, the probability of the normalized actual production being less than the normalized P10 forecast is determined (Figure 9). The same CDF plots are used to see whether the normalized P10 forecast for aggregation year 3 (chosen randomly) for all fields is overconfident. For an unbiased normalized P10 forecast the probability of normalized actual production being less than the normalized P10 forecast should, of course, be 10%. If the normalized P10 forecast was unbiased and a red, vertical line was constructed using the normalized P10 forecast on the horizontal axis, the red, vertical line would cross the CDF with a corresponding value on the vertical axis of 0.1.

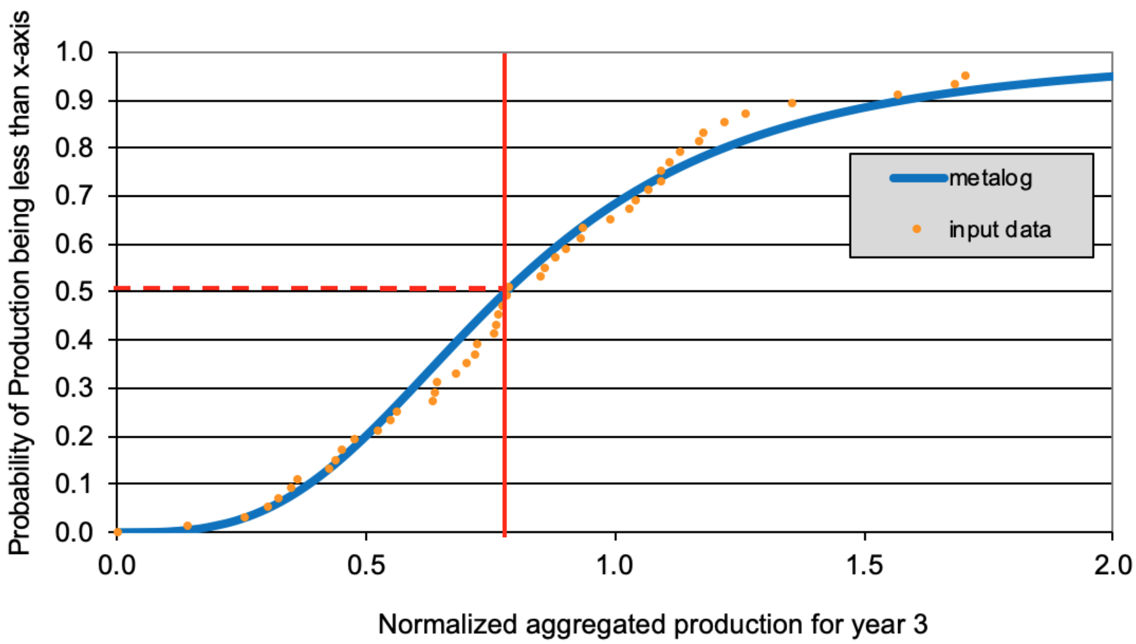


Figure 9. CDF for normalized actual production, with the normalized P10 forecast indicated. The CDF is for the aggregation year 3, including all fields.

The dotted red line in Figure 9 indicates that the normalized P10 value for aggregation year 3 is overconfident. For 51 % of the fields, the normalized actual production was less than the normalized P10 forecast production.

SENSITIVITY ANALYSIS FOR OVERCONFIDENCE BIAS

Similar to the optimism bias, a sensitivity analysis was conducted to investigate whether the overconfidence bias is observable throughout the first eight production years. Figure 10 shows the results of a sensitivity analysis of the overconfidence bias as a function of production year. The orange dots show the percentage of fields whose cumulative actual production did not exceed the field's aggregated forecasted P10, to number of aggregation years n. The gray bars show the number of fields included as a function of n with the scale on the right side of the graph. The number of fields with valid P10 forecasts decreases as n increases, because of data availability and field life. There are no major improvements in the overconfidence bias throughout the years.

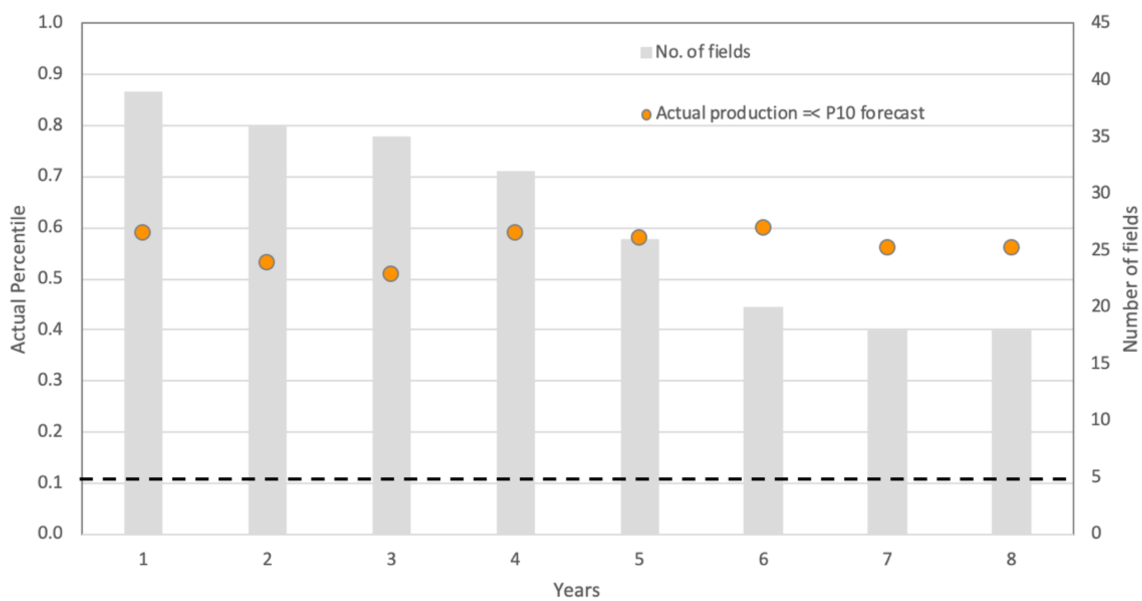


Figure 10. Sensitivity analysis of the overconfidence bias as a function of production year, on an aggregated basis

Overconfidence seems to be even more severe than optimism for this specific data set. It is reasonable to assume that most of the time and effort goes into determining the mean production forecasts, resulting in mean assessments that are less affected by biases as compared to low (and/or) high estimates.

Overconfidence is more apparent in P10 estimates, as P90 estimates might not capture the full extent of overconfident forecasts – i.e. it is not possible to exceed the P100. Therefore, overconfidence and the impact on the 80% interval is better examined using the P10 value.

FIELD SIZE SENSITIVITY ANALYSIS

An argument could be made that field size might play a role in the occurrence of biases and that optimism depends of field size. Some might argue that smaller fields, compared to larger fields, need to be more optimistic to get them approved, as smaller fields might be more marginally economical. If optimism is intentional to get those field approved, the bias becomes a motivational bias.

Figure 11 shows a sensitivity analysis on the field size vs. the occurrence of overconfidence and optimism. There seems to be no observable trend that indicates that field size plays a role in how well calibrated forecasts are.

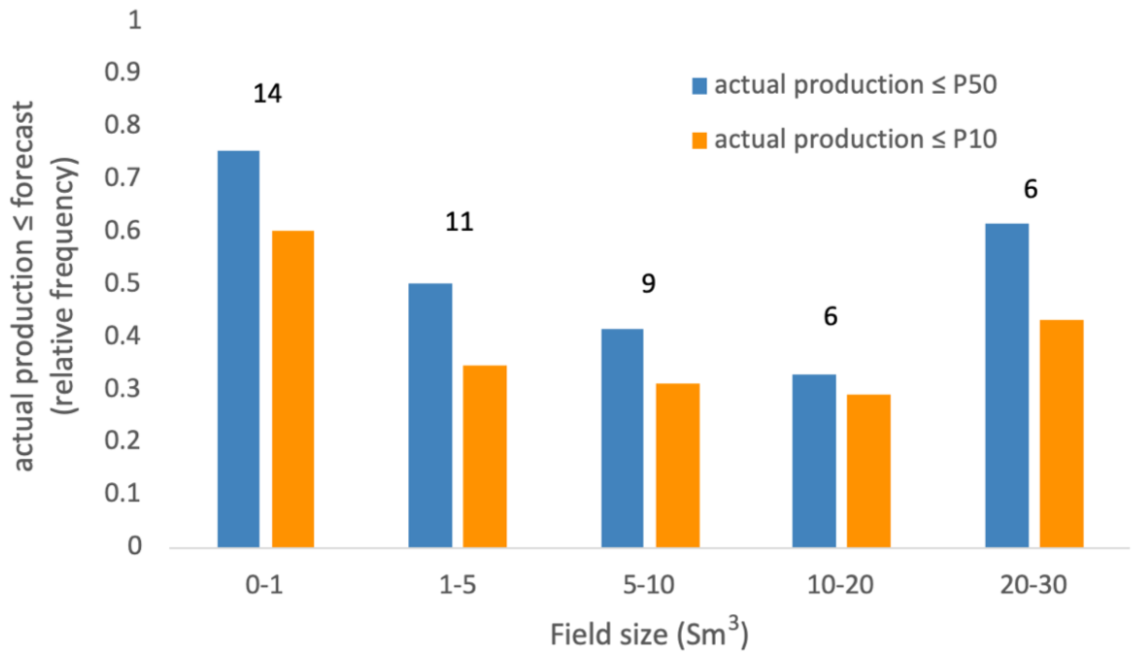


Figure 11. Sensitivity analysis on field size vs. occurrence of overconfidence and optimism. The bars depict the relative frequency with which the forecasted P50 is equal or greater than the actual production (in blue) and the relative frequency with which the forecasted P10 is equal or greater than the actual production (in orange). The number of fields that fall into each field size category varies between 14 and 6 fields.

Limitations of the Analysis

There are several limitations of the analysis that will be briefly discussed in this chapter. First, production forecasts and rates are reported annually, which makes it impossible to determine the month for first oil production, based on the available data. Imagine the following scenario: The forecasted start-up for the first year was intended to be early in that year (e.g. February or March), a delay might push the start-up out a few months but not into a new calendar year. This will create a seemingly overconfident or optimistic first year production forecast. Of course, the opposite scenario is also possible.

Therefore, to reliably judge the quality of the production forecasts in year 1, further data refinement will be necessary.

Secondly, as the production years increase the data availability decreases. It becomes increasingly difficult to draw conclusive findings supported by statistics.

DO FORECASTERS LEARN FROM THEIR MISTAKES?

The study period covers 22 years in which there have been significant technical advances in the industry. A lot of attention and effort has been spent on increasing the sophistication of uncertainty models. But the question remains whether those improvements have led to a decrease in biases in those forecasts. Figure 12 shows the results of a sensitivity analysis on the production performance of all 55 fields over the 22 years span.

The FID year is displayed on the horizontal axis, with each blue dot representing the production excess or shortfall of the mean forecast relative to the actual production, for a field that was approved (FID) in that year. The mean production forecast for the first four years of each field was aggregated and measured against the actual production. For the years 2015, 2016 and 2017, the first three, two and one year(s) were used respectively. A blue dot below the black line (at value zero) indicates production shortfall, and a dot above the line indicates production excess. If there are improvements pertaining the quality of production forecasts, the moving average of all dots would converge to zero over time. The red curve shows a LOESS curve (local polynomial regression) for the entire time span.

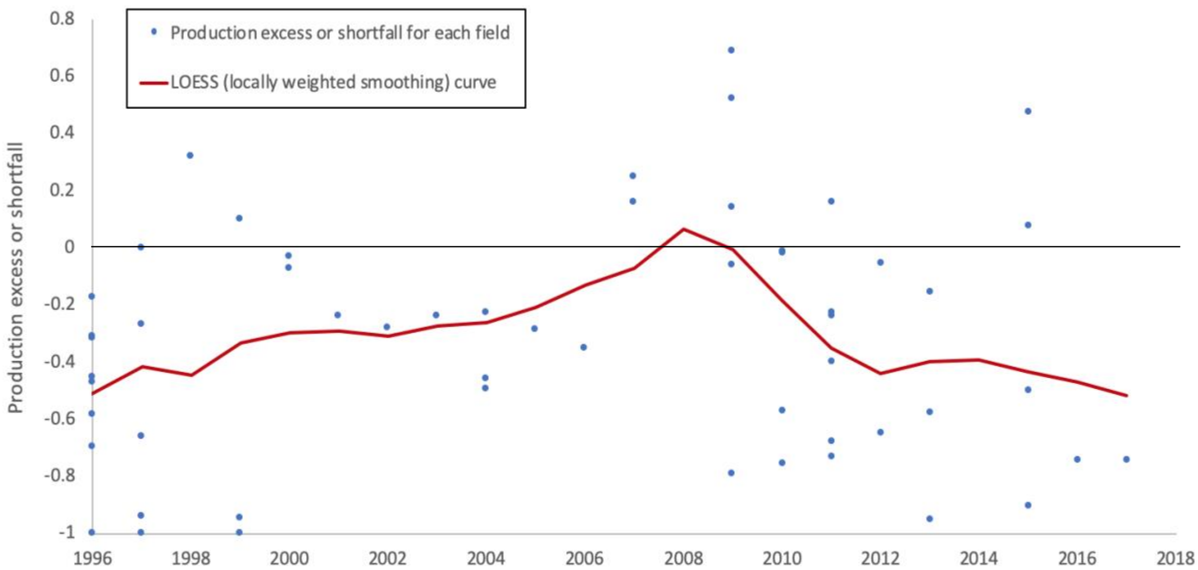


Figure 12. Production excess or shortfall for cumulative first four-year production displayed in the year the FID was made for the field. The red LOESS curve is a moving polynomial regression

From year 1996 to 2008 there is gradual change in the LOESS curve from a low value of -0.51 to a high of 0.07, indicating a general improvement in the quality of production forecasts. After 2008 the trend reverses and in 2017 the local polynomial regression reaches a value of -0.51 again. Therefore, operators do not seem to learn and improve the quality of their production forecasts over time, even with the advent of more sophisticated uncertainty models over the past two and a half decades. The increase in application of more sophisticated models does not in and of itself remove bias or improve forecasting performance.

REASONS FOR OVERCONFIDENCE AND OPTIMISM

Future production is uncertain, and this uncertainty creates an environment in which unforeseen events will occur. However, a distinction must be made between unexpected events and biases in production forecasting. In a major study by Flyvbjerg, the reasons for shortcomings in forecasts were classified as *causes* and *root causes*.[29] Production is associated with a variety of uncertainties, and some of the *causes* that may occur include development time delays, unexpected geological features, reservoir complexity, flow constraints etc. While the causes for not attaining the forecasted production are numerous, one should expect an improvement over time. In a professional setting, the financial ramifications of not delivering on forecasted production would lead to efforts to mitigate future production shortfalls. With no immediate improvement in sight, the explanations for those shortcomings must be found elsewhere. The focus why forecasts systemically fail to deliver must shift to *root causes*. The latter are those factors that persist in the face of statistical analysis. The *root causes* are that forecasters continuously underestimate and, in some cases, even ignore the previously mentioned *causes*. The ignorance or misjudgments of those uncertainties are ultimately manifested in the observed biases. Forecasting errors typically fall in two categories. They can either be unintentional (i.e. delusion) or intentional (i.e. deception).

Delusion

Decision makers and forecasters fall victim to what Kahneman and Tversky (1974) call the planning fallacy.[30] This fallacy will manifest itself in forecasters exhibiting a delusional optimism when assessing uncertainties. The tendency to create production

[29] (Flyvbjerg 2011)
[30] (Tversky and Kahneman 1974)

forecasts based on scenarios of success might be appealing but forecasters will miss the potential for mistakes, even if it is unintentional. Those forecasts will create scenarios in which operators do not deliver and what was promised and likewise the expected financial returns are not attained. Kahneman and Tversky's initial work was later extended by Lovallo and Kahneman (2003) who argue that the biases occur because decision makers often take an inside view when generating those forecasts. Viewing the problem at hand form an inside view will result in forecasters focusing on the unique characteristics of the project at hand. Several studies have shown, that adopting an outside view can reduce the level of delusion in producing uncertainty estimates.[31]

Deception

Another explanation, why forecasts exhibit continuous biases, is deception. Strategic and deliberate misrepresentation of future production is a common occurrence in a project-based industry.[32] Forecasters and decision makers will intentionally overestimate production to increase the chances that their project will receive the necessary approval and the subsequent funding. The deliberate emphasis of advantageous project characteristics and the misrepresentation of potential downside risk will make the project appear superior than it actually is. This fosters an environment in which biases are likely to occur. Financial metrics can also be used to measure the consequences of those biases. Value is destroyed, as capital is not allocated in the most efficient way possible.[33] Such strategic misrepresentations can be countered by enhancing transparency pertaining project forecasts within companies. Providing clarity and aligning incentives, in such a way that

[31] (Flyvbjerg et al. 2014), (Kahneman and Lovallo 2003)
[32] (Flyvbjerg et al. 2014)
[33] (Ernst & Young 2014)

deliberate distortion of project details is discouraged, would be first steps in the right direction.

6. Improvements to production forecasting

ORGANIZATIONAL IMPROVEMENTS

Strong evidence has been presented showing that production forecasts of oil fields on the NCS do indeed suffer from optimism and overconfidence. The focus should now shift towards possible bias mitigation measures. There are several organizational improvements proposed in different research works that have been proven to reduce the biases and their impacts.

Number of expert opinions

Welsh et al. (2007) showed that if an increasing number of expert opinions are considered, overconfidence decreases with a rate that is dependent on the extend of agreement between different experts.[34] They found that the decrease in overconfidence levels, with the increasing number of expert opinions, is non-linear and tends to reduce as the number of experts increases, shown in Figure 13. Thus, more expert opinions only translate into marginal overconfidence reductions after a certain threshold.

[34] (Welsh et al. 2007)

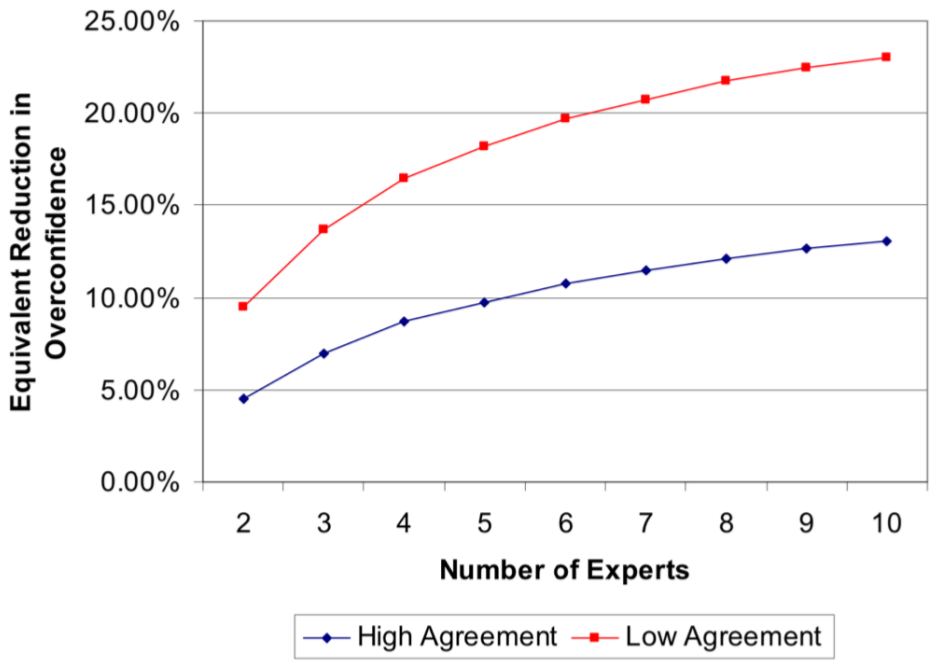


Figure 13. Reduction of overconfidence levels by number of experts and level of agreement

In a corporate setting production forecasts are typically reviewed by more than one person but seldomly do those numbers increase into the double digits. Potentially a greater impact stems from the fact that in most companies, expert opinions - while sometimes reached individually - will be incorporated in such a way that there is generally a consensus reached among those who produce or review the production forecasts. The data available in this study does not indicate the number of expert judgements used to produce the production forecasts.

Risk awareness training

In 2005 Welsh and colleagues published the results of a study in which they asked oil and gas industry professionals a set of questions to highlight common biases.[35] One of the biases that was investigated was overconfidence. The results confirmed that overconfidence commonly occurs in the oil and gas industry. Participants of the study were also asked regarding their industry expertise as well as state their previous experience with risk awareness training and when they completed such training. Findings highlighted that there is no significant reduction in overconfidence with increasing industry experience – which might be most surprising or even worrying, since we commonly rely on industry professional with experience. However, the subject individuals who had undergone recent risk awareness training were performing slightly better than those whose training has been further in the past.

A subsequent study that was published by the same authors confirmed previous findings regarding the reduction of overconfidence with individuals who have had some sort of risk training.[36] This study made the difference even more apparent as a group of students was given two test, one prior to risk awareness training and one immediately after the risk awareness training. The results showed an increase in the range estimates by 20% for the answers given in the post-training test. The importance of frequent risk awareness training was made apparent when the same group of students outperformed a group of industry professionals of whom half of those professionals had some sort of risk awareness training in the past.

The extend of the debiasing effect seems to differ not only regarding the timing of when risk awareness training has been received, but also on the industry and its ability to

[35] (Welsh et al. 2005)
[36] (Welsh et al. 2006)

highlight those shortfalls in a practical setting.[37] Kahneman has found that the effect of debiasing forecasts is mitigated once applied in practice.[38] Therefore, there is a need to actively be aware of those biases, which can be achieved through tracking and updating forecasting performances.

Tracking performance

There is clear evidence that tracking past performances will aid in reducing biases encountered in uncertainty assessments.[39] Fondren et al. published a research study in 2013 in which a database was built to track probability assessments and their outcomes. In one of the examples, shale production forecasts were compared to actual production outcomes to assess the biases and investigate possible corrections. The findings show that using past results can help better calibrate production forecasts. Figure 15 and Figure 14 summarize their findings.

37 (Sellier et al. 2019)
38 (Kahneman 2011)
39 (Capen 1976), (Fondren et al. 2013),

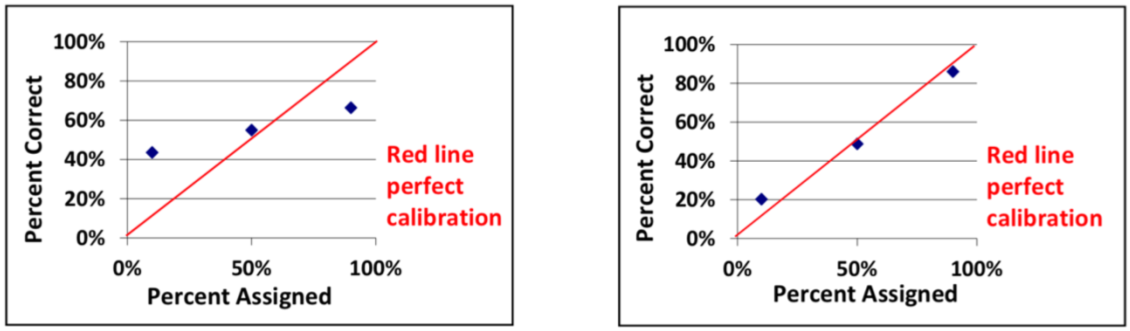


Figure 15. Biased long-term forecast (left) and debiased long-term forecast with 1.5 years of production history. (Fondren et al. 2013)

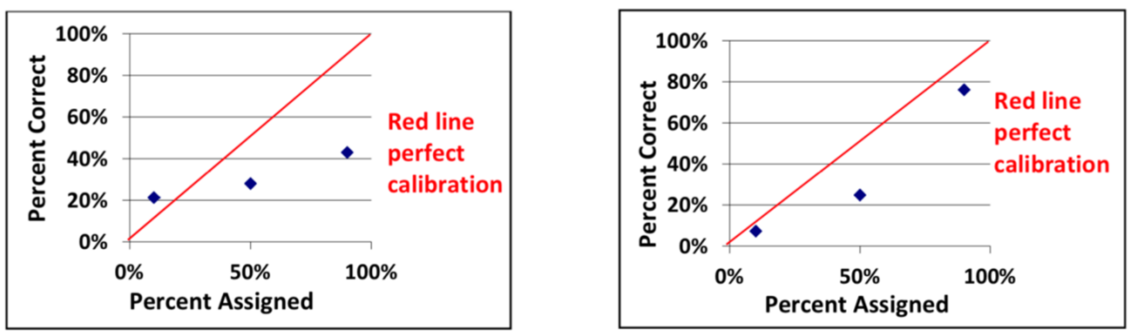


Figure 14. Biased long-term forecast (left) and debiased long-term forecast with 3 years of production history. (Fondren et al. 2013)

The plots indicate the percentiles on the horizontal axis and the percentiles based on the actual outcomes on the vertical axis. For an unbiased forecast the data points would fall on the 45-degree line, i.e. a forecasted Pxx would be an observed Pxx. However, as the authors point out, there are limitations to their study. Given the sparse amount of actual production history, the future actual production was generated using “hind-casting”. Shale formations are characterized by fast build-ups and declines, which limits the number of usable production years. The attention then shifts towards the question whether a few years of actual production history are sufficient to measure the full extent of the impact of those biases?

Tracking performance through past production history is beneficial and easy to implement, yet organizational measures maybe often be cumbersome to introduce, and their adaptability might vary depending on company culture. Therefore, the next two chapters will briefly introduce two methods that can be used to reduce biases in production forecasts, that do not rely on any organizational measures, but rather on statistical methods.

7. Reference Class Forecasting

The previous chapters highlighted the possibility of using past performances and actual outcomes to improve uncertainty assessments in probabilistic forecasting. One such method that has gained increasing popularity in the past decades is reference class forecasting.

Literature

Reference class forecasting (RCF) is based on early work from Daniel Kahneman and Amos Tversky who propose the selection of a reference class "*… for which the distribution of outcomes is known, or can be assessed with reasonable confidence.*"[40] A reference class is a group of comparable, historic projects. To establish an appropriate reference class the number of projects included must be large enough to allow for statistical conclusions to be drawn. However, it must also capture the characteristics of the project in question, such as size, complexity, duration etc., which will in contrast limit the size of the class.

The idea behind reference class forecasting is to provide an outside view on projects with the goal to mitigate some of the biases described earlier. The outside view is achieved by gathering information regarding outcomes and by establishing a distribution of past projects in a statistical setting. This distribution can then be used to improve the specific project at hand. Kahneman and Tversky concluded that decision makers and forecasters that focus on individual estimates, without considering distributional information, will be optimistic in their assessments.[41] Various research studies have proven that if a suitable

[40] (Kahneman and Tversky 1979)
[41] Ibid.

reference class is applied, the accuracy of the investigated forecasts will improve.[42] In one such study, Flyvbjerg et al. investigated the performance of time and cost estimates for roadwork projects in Hong Kong. Projects were divided in different categories depending on what completion stage they were in. Optimistic and overconfident forecasts were identified, and a reference class was used to correct the forecasts. The resulting uplift factors (U) are a function of the probability (p) of a cost or time overrun. For the probabilities between 0 and 1, the maximum overrun was established that was not exceeded in the historic data, where x are the overruns and X is a given value of x. (inf being the infimum; the greatest lower bound for this set). The results are illustrated in Figure 16.

$$U(p) = inf\{x: p \leq Pr(X \leq x)\}$$

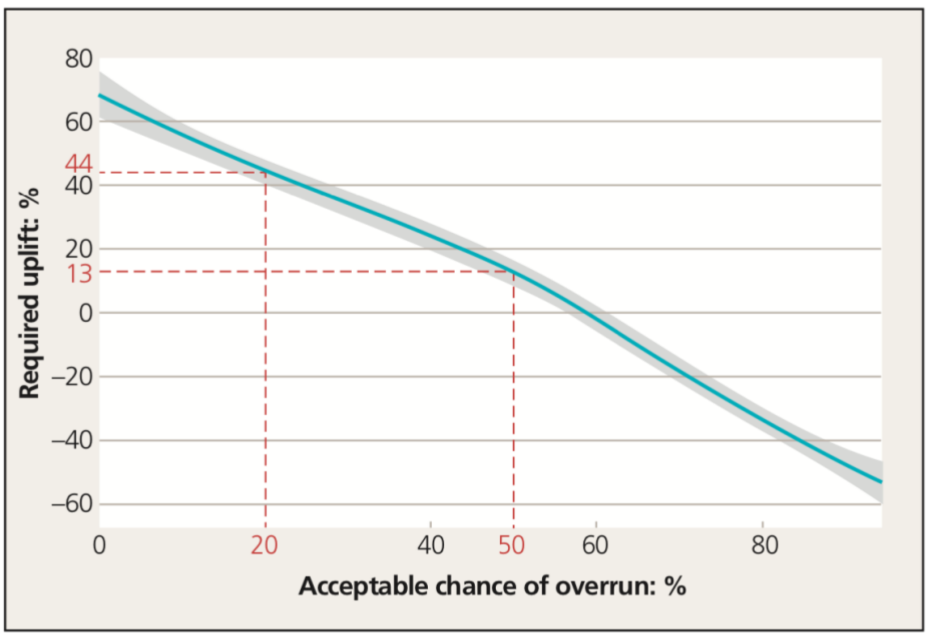


Figure 16. Uplift factors for different acceptable chance of cost overrun for one of the project categories. The grey band indicates the 95% confidence interval. Flyvbjerg et al. 2016

[42] (Kahneman and Tversky 1979), (Gilovich et al. 2002), (Flyvbjerg et al. 2004), (Flyvbjerg et al. 2016)

The uplift factors to adjust the initial, biased forecasts for all projects are summarized in Figure 17. The factors can be used as a multiplication factors for the original, biased forecast. For example: an uplift of 13% will yield in a multiplication factor of 1.13. If the forecasts were pessimistic, the multiplication factor would be smaller than one.

	Level of certainty	Required uplifts		
		Category C	Category B	Category A
Base cost uplift[a]	P50	+13%	+7%	–1%
	P80	+44%	+34%	+14%
Schedule uplift[b]	P50	+26%	+22%	+8%
	P80	+75%	+53%	+26%

[a]Total estimated budget, excluding inflation and excluding bottom-up contingencies

[b]Measured from the date of the estimate to the date of project completions

Figure 17. Uplift multipliers for roadwork project forecasts in HK early stage projects (Category C) and late stage projects (Category A); (Flyvbjerg et al. 2016)

APPLICATION

The application of reference class forecasting is detailed in the following three steps[43]

[43] (Flyvbjerg 2005)

1.) Past projects that are comparable to the project at hand will be identified and included in the reference class. The selection process may vary on the user's preferences and which projects she or he deems comparable.

2.) A cumulative probability distribution for the parameter of interest (usually completion time or cost) is generated using outcomes from past projects that were selected in step 1. One possible approach would include determining the required probability of meeting a production forecast, based on the historic data, and fitting a cumulative distribution to the past outcomes.

3.) Finally, the project at hand is compared with the distribution obtained in step 2. The comparison will yield uplift and/or scaling factors that can be used to amend the probability assessment for the project at hand, as shown in Figure 17. An example will be demonstrated in the subsequent paragraphs.

To date, there has not been an extensive, public study on the application of RCF for production volumes (or reserves) in conventional oil and gas reservoirs. There might be several reasons why this is the case. Oil and gas companies are naturally inclined not to disclose sensitive data, making it difficult to gain access to production forecasts. While the theory behind reference class forecasting has been developed by Kahneman and Tversky in the 1970s and early 1980s, it took almost 30 years to utilize the theory in a practical setting.[44] Flyvbjerg et al. presented the first major study using RCF in 2004.[45]

[44] (Tversky and Kahneman 1974, Kahneman and Tversky 1979)
[45] (Flyvbjerg et al. 2004)

The reference class for this study was established by considering all fields that were used in the investigation for the optimism and overconfidence bias. While the geological features might differ significantly within the Norwegian Continental Shelf, the projects are comparable in their fundamental nature. They are all producing from a well-known geological system, where ample data is available through historic projects. Further, the projects used for the reference class are all producing oil. Any fields not producing oil i.e. NGL, condensate and gas were omitted from this study at the beginning.

Step 2 puts the production outcomes in a distributional setting so that information can be drawn from the probability distribution. One of the advantages of a continuous probability distribution is that we can draw any value of interest and compare it with the forecasted production. Details of step 2 are outlined below.

RCF FOR FIELDS ON THE NCS

Reference classes were established for each of the first eight production years. The first reference class contains first year production outcomes for all fields. The second reference class contains aggregated second year production outcomes for all fields, and so on. No fields were omitted from the study. With increasing production year, the size of the reference classes decreases, as some fields do not have eight years' worth of production. Table 2 shows the number of fields included in each reference class.

Production year	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Number of fields included in RC	52	52	50	48	44	40	35	33

Table 2. Number of fields included in each reference class

Similar to Chapter 5, the actual production outcomes for any aggregation year are normalized by the P50 forecasts (we are using the normalized P50s here, rather the

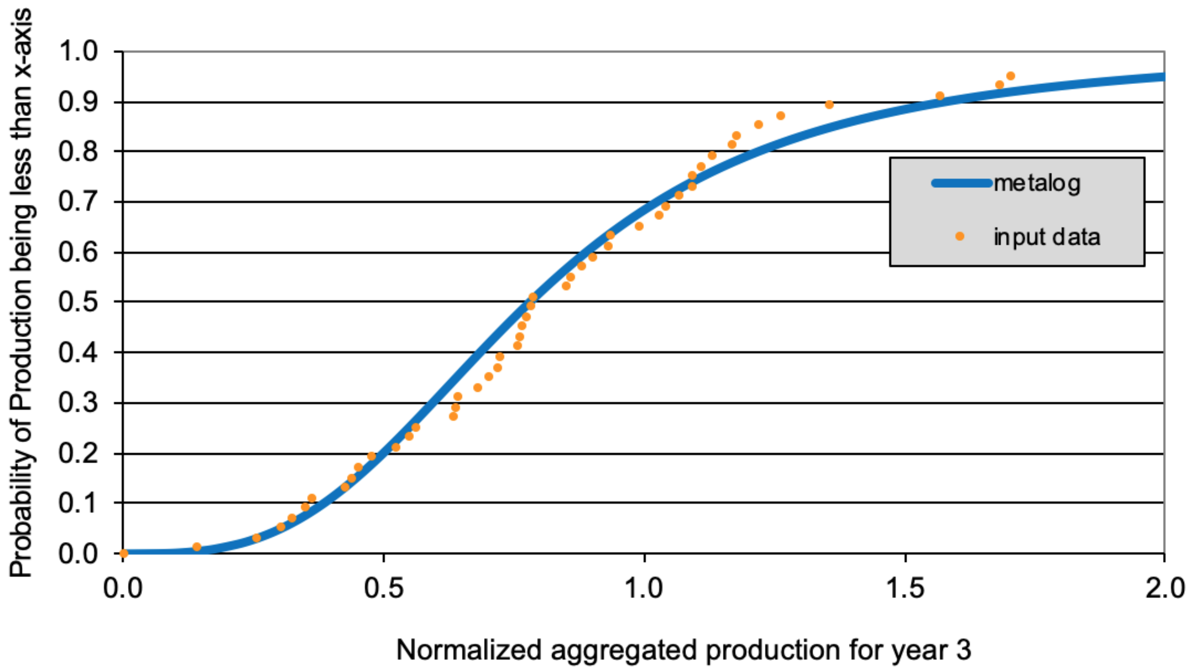


Figure 18. Cumulative distribution of normalized production outcomes for aggregation year 3

normalized means as it is more conclusive in this example). The normalized actual production values are sorted in ascending order by magnitude and analogous to Chapter 5, Keelin's metalog distribution was used to fit a cumulative distribution to the normalized actual production data. Figure 18 shows the cumulative distribution for production year 3.

After the cumulative distribution is fitted, the axes are switched and thus the distribution becomes inverted, shown in Figure 19. This will allow the determination of any multiplication factors.

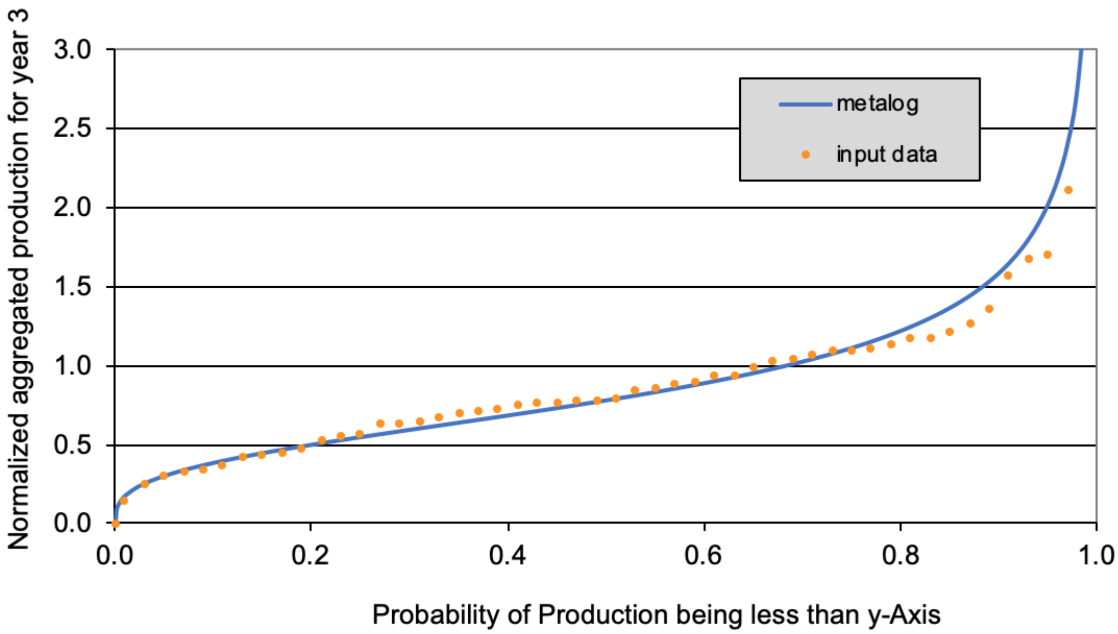


Figure 19. Inverted cumulative distribution for normalized aggregated production in year 3

Recall that the production outcomes were normalized by the P50. Any P50 forecast of any project at hand (for that specific aggregation year) can now be normalized (such that x=1.0) and be compared to the inverted distribution. The horizontal axis shows the acceptable probability of producing less than the forecasted value. The vertical axis in this context shows the RCF multipliers (i.e. multiplication factors, uplift factors). Therefore, for an unbiased P50 forecast the cumulative distribution would return a factor of 1.0 when looking at the 50% value on the horizontal axis. The plot for aggregation year 3 is again depicted in Figure 20 for better visualization.

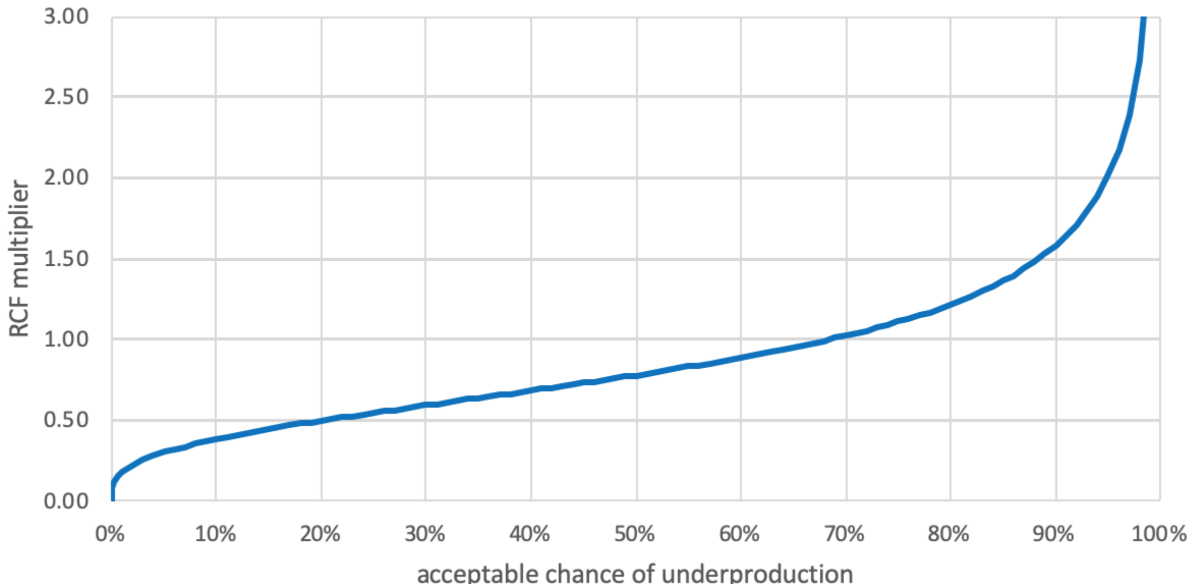


Figure 20. Acceptable chance of not attaining production and according RCF multiplier for aggregation year 3

If the P50 production forecast is optimistic the multiplication factor for the P50 forecast will be less than 1. To account for the overconfidence bias, similar adjustments are expected for the 10th percentile. Figure 20 confirms that the multiplication factor for the biased P10 forecasts for aggregation year 3 are lower than 1. The RCF multipliers for the P10 and the P50 forecast are indicated in Figure 21, highlighted by the orange dotted lines.

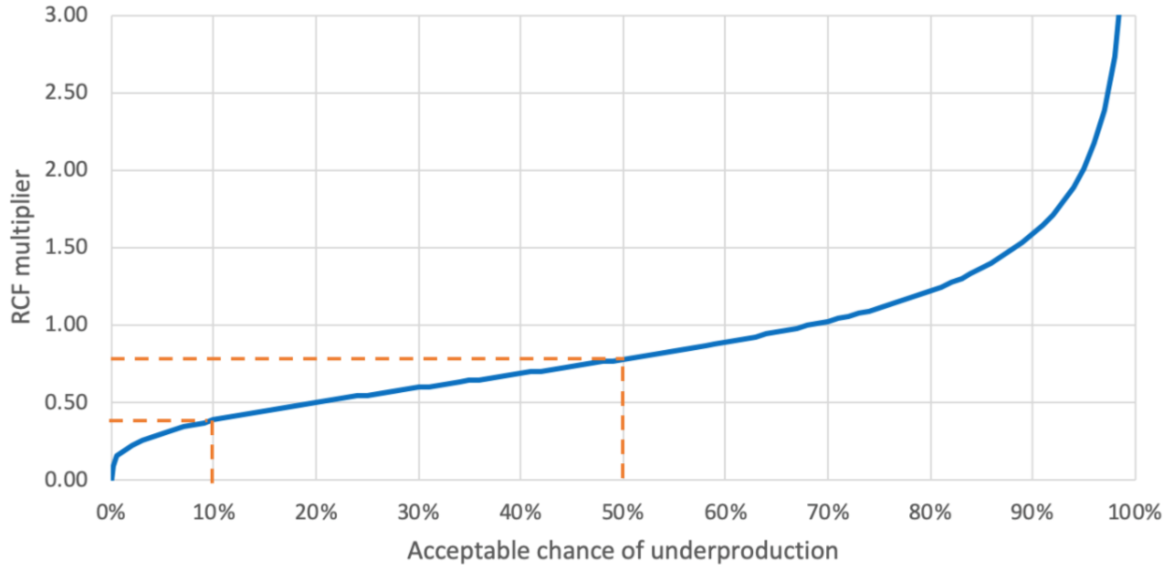


Figure 21. Acceptable chance of not attaining production and respective RCF multiplier for aggregation year 3 with indicated P10 and P50 multipliers

RCF UPLIFTS FOR YEAR 1 TO YEAR 8

Figure 22 and shows the P10 values both the biased and debiased forecasts, on an aggregated basis. The red dots indicate the probability of producing less than the biased P10 forecasts. The green dots show the debiased forecasts.

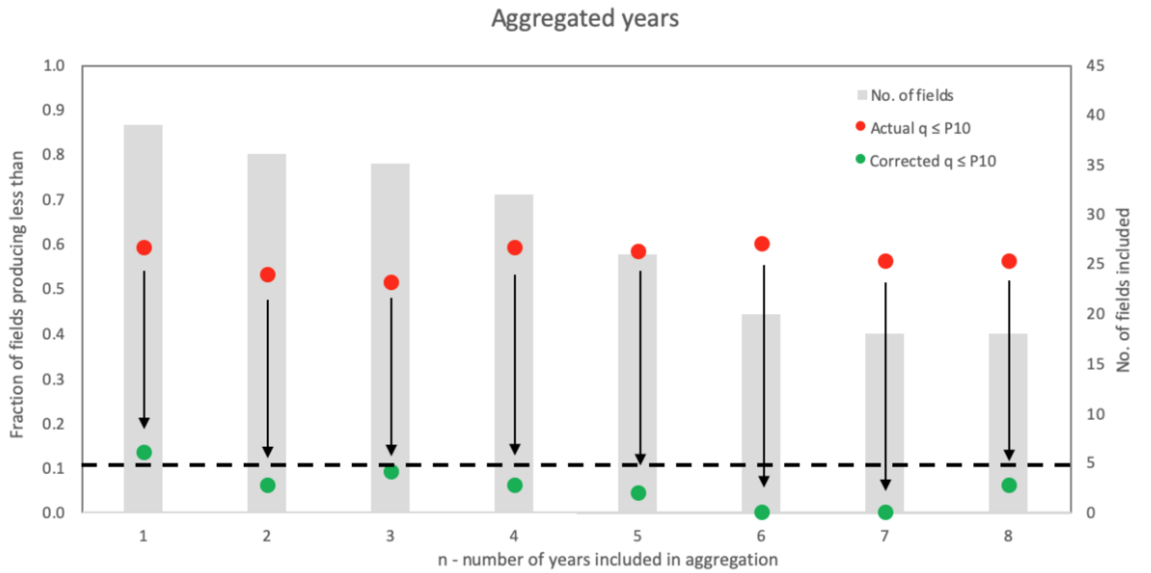


Figure 22. P10 percentiles before and after the application of RCF over the first 8 production years

The debiased forecasts were obtained by applying the aforementioned multipliers to the initial forecasts, and then determining the new probability of producing less than the forecasted production. An improvement is observable for the P10 forecasts, once the multiplication factors are applied. The quality of the forecasts improves, as the actual percentiles move closer to the values that would constitute an unbiased forecast.

Figure 23 shows the P50 for both the biased and debiased forecasts, on an aggregated basis. The red dots indicate the probability of producing less than the biased P50 forecast. The green dots show the debiased forecasts.

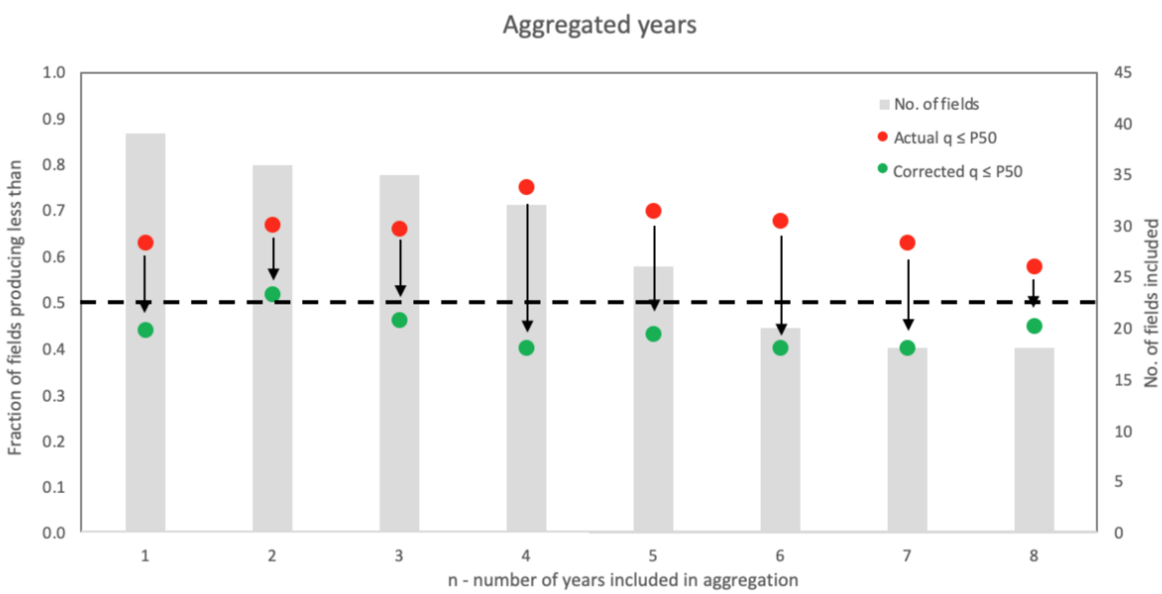


Figure 23. P50 percentiles before and after the application of RCF over the first 8 production years

Again, an improvement is observable for the P50 forecasts, once the multiplication factors are applied. The quality of the forecasts improves, as the actual percentiles move closer to the values that would constitute an unbiased forecast.

The multiplication factors are summarized in Table 3. The table also highlights one of the challenges encountered when using RCF. The number of available fields after year 5 is reduced to the point where the obtained scaling factors for the P10 forecasts result in debiased P10 values close to or equal to zero.

Year	Without applying RCF multipliers			With applying RCF multipliers		
	P10	P50	P90	P10	P50	P90
1	0.59	0.63	0.41	0.13	0.44	0.87
2	0.53	0.67	0.47	0.06	0.52	0.94
3	0.51	0.66	0.49	0.09	0.46	0.91
4	0.59	0.75	0.41	0.06	0.40	0.94
5	0.58	0.70	0.42	0.04	0.43	0.96
6	0.60	0.68	0.40	0	0.40	1
7	0.56	0.63	0.44	0	0.40	1
8	0.56	0.58	0.44	0.06	0.45	0.94

Table 3. RCF multipliers for P10s and P50s for the first eight production years

LIMITATIONS OF RCF

There are several limitations pertaining to reference class forecasting.[46] If the projects span over an extended time period, then the technology applied might differ significantly between individual projects. The uncertainty assessments may also differ substantially between individual fields. In the production forecasting context for fields on the NCS, those arguments do not hold true. It was already established that the increase in sophistications in uncertainty modeling does not in and itself lead to a reduction in the magnitude of biases.

A second argument could be made about the class size. What is an appropriate size for a reference class, so that on the one hand, the projects are comparable but on the other

[46] (Hájek 2007), (Leleur et al. 2015)

hand the comparison is meaningful? It might be challenging to find a measure of comparison to quantify appropriate levels of similarities. In this study, given the limited amount of data, no samples were excluded from the reference class. Therefore, an inferential approach, to calculate the robustness of the reference class must be taken. To test the robustness of the class, a “one-out validation” was chosen. The following three steps were implemented. Aggregation year 3 was chosen to test the robustness:

1.) Chose a random field from the reference class and exclude it from the class.

2.) Recalculate a new reference class with the remaining fields.

3.) Test if the new reference class would achieve the same or better results than the complete reference class, once the multipliers are utilized.

The results of the one-out validation method are partly summarized in Table 4, the entire table can be found in Appendix A-5. Ten random fields have been chosen to show the results of the validity study. The P10 and the P50 corrected with the adjusted reference class shall be denoted as

$$P10_{n-1} \text{ and } P50_{n-1}$$

Where *n* is the class size of the entire reference class. Table 4 shows that the P10 forecasts on average were only slightly impacted by the removal one field from the reference class, indicating that the reference class in terms of determining the P10 scaling factors is robust. The P50 uplift factors in comparison seem to be moderately impacted by the removal of a single field. On average the removal of one field from the reference class resulted in a probability of 57% of producing less than the P50. For comparison, the entire reference class will yield P50 forecasts that on average have a 46% probability of producing less than the forecasts. There is limited variation for the $P10_{n-1}$ and the $P50_{n-1}$ for different $n-1$ reference classes. It can therefore be concluded, that the reference class is robust and large enough in size.

Field number	$P10_{n-1}$	$P50_{n-1}$	avg. $P10_n$	avg. $P50_n$
30	0.104	0.583	0.090	0.460
47	0.104	0.604	0.090	0.460
3	0.104	0.542	0.090	0.460
35	0.104	0.583	0.090	0.460
14	0.083	0.563	0.090	0.460
39	0.104	0.604	0.090	0.460
12	0.083	0.563	0.090	0.460
42	0.104	0.604	0.090	0.460
37	0.104	0.604	0.090	0.460

Table 4. 10 randomly excluded fields with unbiased forecasts for the entire dataset excluding that specific field

If the reference class would be sensible to the exclusion of one field, the reference class could be refined pertaining geological characteristics. Unfortunately, the geological data of each field are not available for this study. Another refinement to RCF in a production forecasting context could be the distinction regarding field size. However, as pointed out previously, for this specific case field size does not seem to significantly impact the magnitude of biases.

8. Rho Signal Information System

Reference class combined frameworks – Literature

The application of RCF to oil and gas production can be considered a brute force method. The convenience regarding calculations and utilization is contrasted by the uncertainty whether the class is representative and thus applicable to the project at hand. The second approach, presented in this chapter, is based on Bayesian probability theory. A Bayesian framework will be used, in which historic data is used as a prior and the distributional forecast is used as a likelihood. Those two distributions will be used to calculate a posterior probability distribution. Previous research has produced a variety of different approaches to build such a model.

Bordley published a study in 2014, in which a Bayesian framework was utilized by combining statistical modeling and past outcomes (reference class).[47] A posterior probability was calculated by updating the reference class data with the statistical model. The study investigated the predicted healthcare cost under a voluntary employee benefit association and showed that the resulting posterior probability had a greater variance and a larger mean than the model-based approach, thus mitigating the biases of the original cost forecasts. While this might work well for cases with access to the model on which the forecasts are based on, it is not applicable in the existing case. The model(s) used for the production forecasts was (were) not made available for this study.

A similar approach to Bordley can be found in a study published by Leleur et al. in 2015, in which RCF is combined with expert judgement pertaining the uncertainty

[47] (Bordley 2014)

ranges.[48] Similarly, Leleur et al. found an improvement in the calibration of forecasts once RCF is applied. The use of expert judgement resulted in scaling factors that are either greater or smaller than 1, thus increasing or decreasing the uncertainty range for the reference class. The authors suggest that if the use of expert judgement yields a scaling factor greater than one, the uncertainty of the reference class is adjusted, taking into account higher uncertainty implied by the expert judgement. If the opposite case is present, and a scaling factor of less than one is obtained from expert judgement, the authors suggest using the uncertainty range specified by the original reference class.

RHO SIGNAL INFORMATION SYSTEM - APPLICATION

Therefore, the attention should be shifted to a model that will integrate the reference class data with the probability distribution over the forecasted production. A Bayesian framework, where the reference class information is used to define the prior and the forecasted production (likelihood function) is used as a signal to calculate the posterior probability distribution parameter, will be specified. The term rho signal information system (RSIS) was first coined in 2008 in a value-of-information focused paper by Eric Bickel.[49] As part of this study, the value was determined for which the value of an information system correlated with a normally distributed signal, using ρ, is equal to $\rho \times 100\%$ the value of perfect information. A similar approach is taking to develop the RSIS for the present data.

[48] (Leleur et al. 2015)
[49] (Bickel 2008)

The assumption is made that the logarithm of the reference class data and the production forecasts are distributed according to a bivariant Gaussian distribution. The approach to creating such a model is detailed in the following steps:

1.) The reference class consists of the actual outcomes for all fields that had valid P10, mean and P90 production forecasts. A maximum likelihood estimation (MLE) was used to find the parameters of the probability distribution that best fits the reference class data. The MLE is used to find the best estimate for the parameter values that will maximize the likelihood function that most closely approximates the observed data. The parameter values of the MLE are called maximum likelihood estimators. In this particular case, the desired probability distribution over the reference class data is a lognormal distribution. The detailed derivation of the maximum likelihood parameters for a lognormal distribution can be found in Appendix-A2. The resulting maximum likelihood estimators (mean and variance) for the lognormal distribution over the reference class data are:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} ln(X_i)}{n}$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(ln(X_i) - \frac{\sum_{i=1}^{n} ln(X_i)}{n} \right)^2}{n}$$

Using the MLEs we were able to fit a lognormal distribution to the actual production outcomes. The lognormal distribution was then transformed into a normal distribution (Figure 24), in order to use properties of bivariant Gaussian distribution, which will become apparent in the next steps.

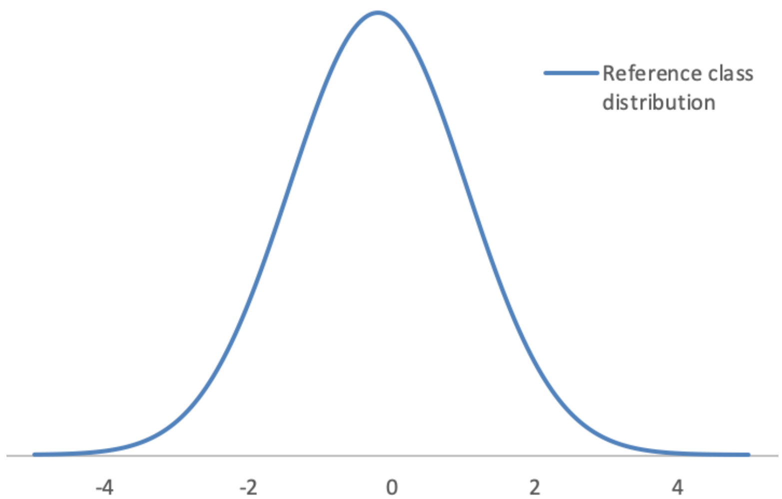


Figure 24. Reference class distribution in normal space

The following equations were used to transform the mean (m) and variance (v) of a lognormal distribution into the mean (μ) and variance (σ^2) of a normal distribution.

$$\mu = ln\left(\frac{m}{\sqrt{1+\frac{v}{m^2}}}\right)$$

$$\sigma^2 = \ln\left(1+\frac{v}{m^2}\right)$$

2.) In step 2, a Pearson correlation coefficient ρ is used, to correlate the mean forecasts with the actual outcomes. For a pair of random variables X and Y, the correlation coefficient is defined as:

$$\rho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

The covariance is defined as follows:

$$cov(X,Y) = \mathrm{E}\big[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y]\big] = \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]$$

With E being the expected value. The rho between the forecasted mean and the actual production for this data set is 0.83.

3.) The fitted lognormal distributions from chapter 4 were converted into normal distributions, using the same equations as in step 1. The forecast will be the likelihood function (i.e. the signal) in the Bayesian framework.

4.) Next, an appropriate value from the production forecast distribution must be chosen to update the prior reference class and thus obtain the posterior probability. As stated previously, it is assumed that most of the time and effort will go into determining the expected production forecast, thus the mean was chosen as a representative value from the probability distribution.

5.) In the final step, the posterior probability must be calculated. The variance of the likelihood function in the RSIS is fixed and known, with the mean becoming the model parameter. Given that the prior and likelihood are Gaussian distributed, we can infer that the posterior will also be Gaussian distributed (bivariate Gaussian distribution - see Figure 25) with the parameters as followed (the detailed derivations can be found in Appendix A-3):

$$\sigma_n^2 = (1 - \rho^2)\sigma_0^2$$

$$\mu_n = \mu_0 + \rho(x_i - \mu)\frac{\sigma_0}{\sigma}$$

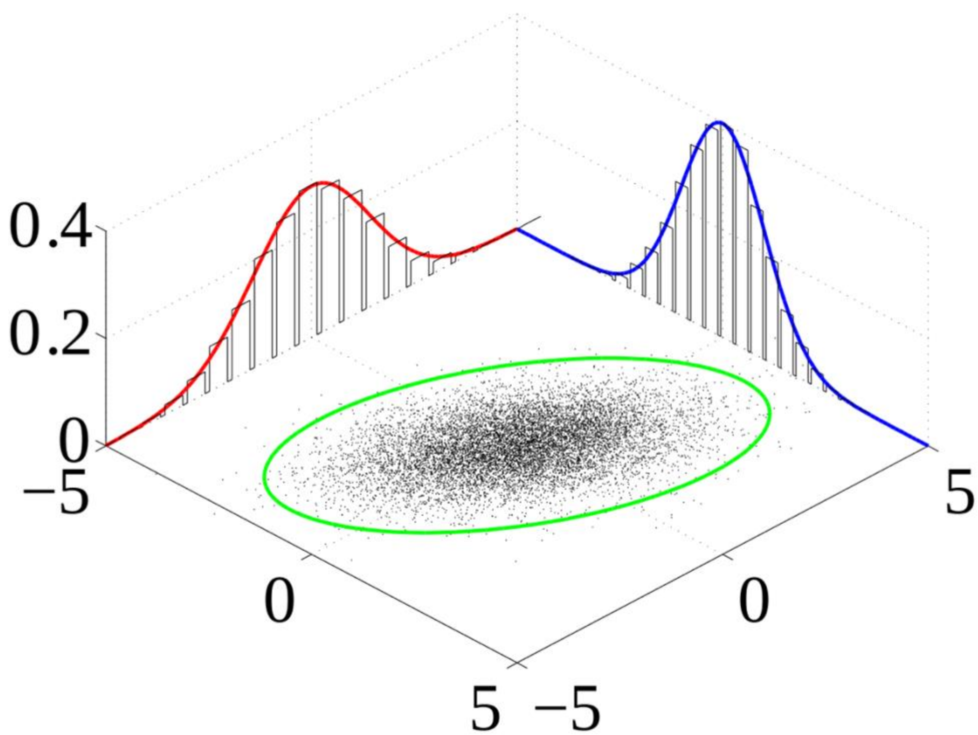


Figure 25. Schematic illustration of a bivariate Gaussian distribution (green).
Matlab code provided by Bscan (2019).

RESULTS OF RSIS

Figure 26 shows the rho signal information system results for a randomly chosen field and year.

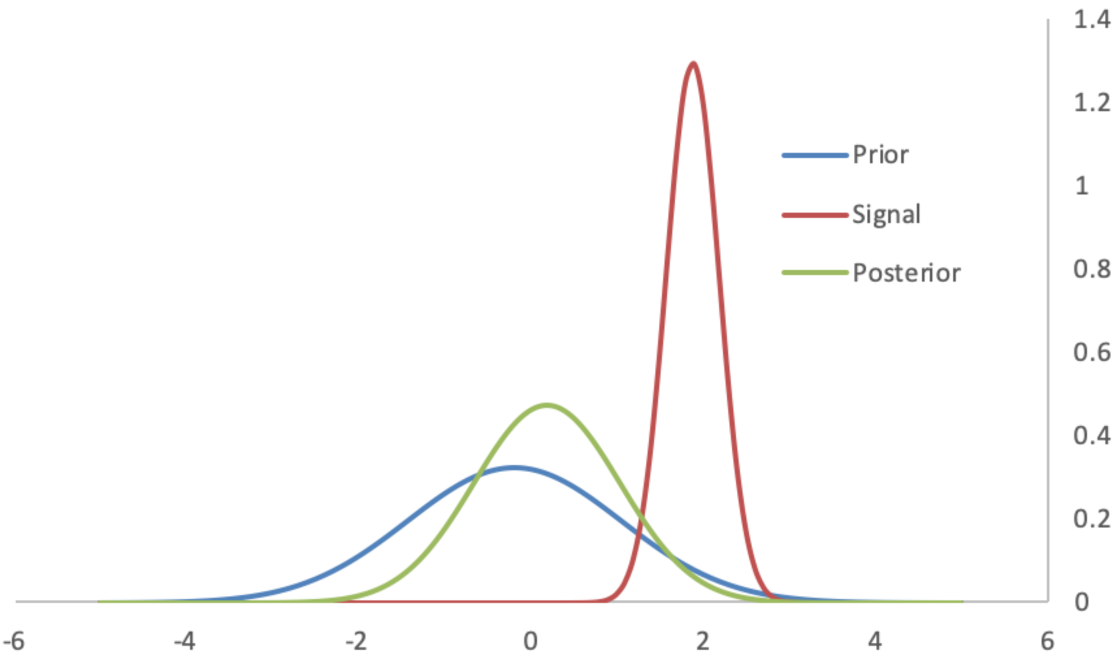


Figure 26. Rho-information system for a random field and a random production year

The prior (RCF) is shown in blue, the signal (production forecast) is depicted in red and the adjusted production forecast (posterior) is shown in green. Recall that all distributions are depicted in normal space. The mean of the posterior is lower than the mean of the likelihood distribution, but higher than the mean of the prior distribution. This is a result that is expected. The RCF would likely underestimate the mean production but the signal and likelihood are used to update the prior beliefs and will result in a mean that is between the historic data of actual production and the production forecast generated for that field and year. The variance of the production forecast is considerably smaller than the variance of the reference class, resulting in a narrower distribution for the likelihood function. The reference class provides a wider distribution, based on historic outcomes which will also increase the variance of the posterior distribution. This correction will mitigate the effects associated with the overconfidence bias. Hence in a Bayesian updated forecast, we see that historic past data can be used to adjust a biased production forecast.

Similar to RCF, a sensitivity analysis pertaining the first eight production years was conducted. Figure 27 depicts the probability of exceeding the production forecasts, before and after application of the rho signal information system, on an individual basis. Production years for the rho signal information system were investigated on an individual basis, rather than an aggregated basis (like in the RCF chapter) because a single distribution representing all production years was used as a prior. Thus, adding up variances as suggested previously to determine aggregated production forecasts would yield results that are not representative.

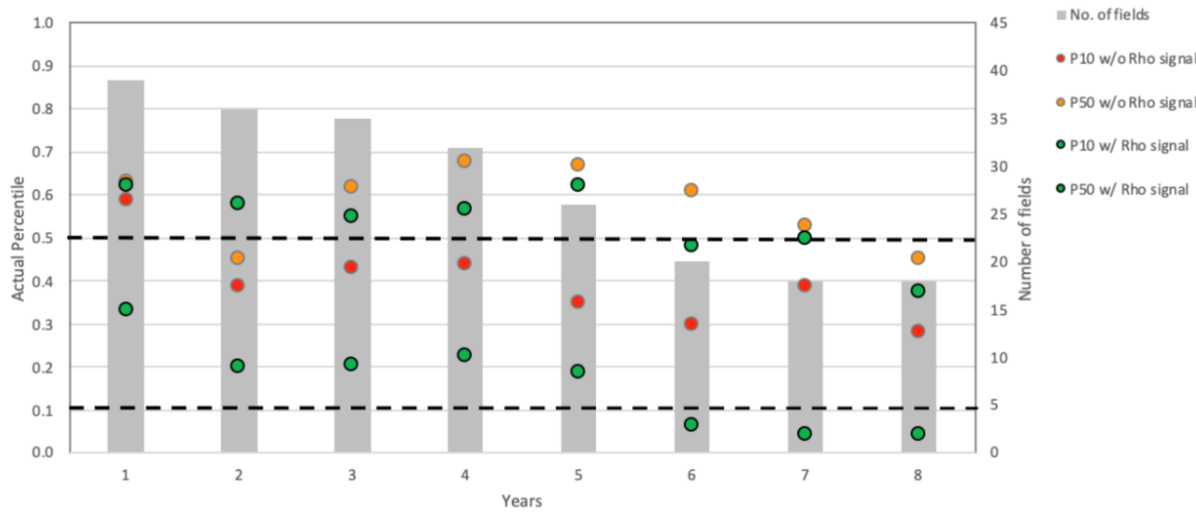


Figure 27. P10 and P50 percentile before and after the application of the rho signal information system for the first 8 production years, on an individual basis

A quick view reveals that the adjusted percentiles are more accurate for the P10 values, but the P50 forecasts only improve marginally. The first two production years deviate more significantly from the actual percentiles, even after the rho signal correction. Given the possible negative implications of overconfidence on NPV, the rho signal information system application is worthwhile.

LIMITATIONS

There are several limitations that arise when the rho signal information system is used. Keeping the variance of the prior fixed will result in a posterior distribution with a constant variance. In a practical setting it is expected that the variance for various fields differ. However, RSIS in this chapter was used to mainly illustrate that a Bayesian framework, where historic data is used to update the production forecast, can be used to adjust biased estimates. If a Bayesian framework is used to update the forecasts in future work, more sophisticated models such as the Markov Chain Monte Carlo approach or the Normal-Gamma conjugate pair would likely yield better results.

Another limitation arises from the constraints faced when choosing an appropriate value from the likelihood distribution. While it is sensible to choose the mean, a different value will yield different posterior distributions. Picking the mean or mode will give disproportional weight to those values. One potential approach to overcome this limitation would be a sequential use of the P10, mean and P90, where these three values are used as three signals to determine the posterior.

9. Discussion on limitation and next steps

Despite having access to almost 600 years' worth of forecasts and actual production data from 55, limitations still arise. First, the data is reported annually, which is especially detrimental for the initial production data. Without more detailed production data (i.e. monthly) it will be impossible to further refine the impact of the start-up delays on the production in the first year. Despite this limitation, the analysis based on the annual data provide valuable information regarding the quality of probabilistic production forecasts for oil fields on the NCS.

As briefly alluded to earlier, the reference class should be refined to get more conclusive results from the reference class. Some of the characteristics, which might provide feasible class selection criteria are geological settings, field size, applied technology etc. Without any additional data for the fields investigated in this study it will be difficult to refine the reference class.

Finally, the application of any model should naturally always be scrutinized in terms of its consistency with data at hand and rationale with regard to constraints. A good starting point is the recommendation made by Leleur regarding expert judgements. Picking the results with a wider uncertainty range, whether it is based on historic data or expert judgement, will yield better results on average. Given strong evidence of overconfident production forecasts, this argument also seems logically sound for this study.

There are several steps that can be taken beyond what was covered in this thesis. The utilization of a database that keeps track of past forecasting performance and allows for easy access to data for any reference class seems imperative. It is assumed that most operators do have such a database, yet what remains uncertain is to what extend operators draw on past experiences to produce production forecasts. With ample evidence of past forecasting studies from different disciplines, that keeping track is beneficial, the creation and appropriate utilization of such a database is strongly advised.

Lastly, the Norwegian Petroleum Directorate is unique in its approach to improve transparency in the industry. More government agencies should advocate public accessibility to data to improve the industries understanding of this problem. Individually, operators may not always be aware that production shortfalls suffer from suffer biases rather than “unexpected events”.

10. Summary and conclusions

Production forecasts for oil fields on the NCS, approved between 1995 and 2017, are biased. The analysis demonstrates clear evidence of optimism and overconfidence. The mean forecasts are, on average higher than the actual production, i.e. they are optimistic. The P10 (and P90) forecasts bound an uncertainty range, in which less than 80% of the actual production falls, i.e. they are overconfident. The causes commonly presented to justify production shortfalls might be numerous, but the real causes are traceable to biases. There are generally two categories of biases that are predominant. Intentional biases (deception) and unintentional biases (delusion) can both lead to value destruction.

The biases vary depending on the production year. Given that the first reinvestment occurred in year 8, most of the analyses focus on the first eight production years. CDF plots are generated to quantify the impact of those biases.

There is a number of mitigation and correction processes that can aid in debiasing the forecasts. Reference class forecasting is a tool that puts past projects in a distributional setting and lets the forecaster determine uplift or scaling factor to the project at hand. It is a convenient and fast method that allows for convenient use. The results show significant improvement in the production forecasts when compared to the biased, uncorrected production forecasts. The rho signal information system is a more sophisticated model, compared to the RCF, with the latter considered as somewhat of a “brute force method”. In the rho signal information system, a Bayesian framework is used to update the prior probability distribution with the production forecasts. This framework can also be considered as a more wholesome approach, since it will honor all inputs provided by the experts to determine the production forecasts.

As shown by various studies, the financial impact of these biases can be substantial. The biases could, for example, lead to a project being approved when it would not have been approved, if the forecasts were unbiased. Moreover, even when the same projects would be approved, the concept choices would likely be different given unbiased forecasts. The reduction of biases in production forecasts is of importance since the NPV can be reduced by the aforementioned biases.

Based on the widespread biases encountered in the industry, it is best to echo what Welsh and Begg have already stated in earlier work: "*In fact, in light of what we know about how bias affects decision making and the economic impacts of this, it could reasonably be claimed that debiasing of* [oil and gas] *industry decisions has greater potential to improve economic outcomes than time and money put into honing technological and modelling processes.*"[50]

[50] (Welsh and Begg 2015)

11. Appendix A

A-1. ALTERNATE LOGNORMAL FITTING METHOD

A log-normal fit can also be obtained using the forecasted P10 and P90 values as followed. Setting up the following equations, with known z_1, z_2 values for α_1,α_2 percentiles, will give

$$\ln(z_i) = \mu + \sigma\Phi^{-1}(\alpha_i).$$

Combining the information from the two percentiles yields

$$\sigma = \frac{\ln(z_2) - \ln(z_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}$$

$$\mu = \frac{\ln(z_1)\Phi^{-1}(\alpha_2) - \ln(z_2)\Phi^{-1}(\alpha_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}.$$

μ can also be calculated directly, once σ is obtained, as μ is a function of σ and can be calculated using either of the original equation presented in Appendix A-1, i.e.,

$$\mu = \ln(m) - \sigma^2/2.$$

A-2. MAXIMUM LIKELIHOOD ESTIMATION[51]

A lognormal distribution was chosen to be fit to the reference class data with the density function of such a distribution being,

$$f(X|\mu,\sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)X}} exp\left[-\frac{(ln(X) - \mu)^2}{2\sigma^2}\right] \qquad X > 0, -\infty < \mu < \infty, \sigma > 0$$

In order to compute the maximum likelihood estimators of the two-parameter lognormal distribution, the likelihood function needs to be established first.

[51] (Ginos 2009)

$$L(\mu, \sigma^2|X) = \prod_{i}^{n}[f(X_i|\mu, \sigma^2)]$$

$$= \prod_{i}^{n}\left(\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right)\frac{1}{X_i}\exp\left[\frac{-(ln(X_i) - \mu)^2}{2\sigma^2}\right]\right)$$

With the log-likelihood function being:

$$\mathcal{L}(\mu, \sigma^2|X) = \ln\prod_{i}^{n}\left(\left(\frac{1}{\sqrt{(2\pi\sigma^2)}}\right)\frac{1}{X_i}\exp\left[\frac{-(ln(X_i) - \mu)^2}{2\sigma^2}\right]\right)$$

$$= -\frac{n}{2}ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}(ln(X_i) - \mu)^2}{2\sigma^2}$$

$$= -\frac{n}{2}ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}[ln(X_i)^2 - 2ln(X_i)\mu + \mu^2]}{2\sigma^2}$$

$$= -\frac{n}{2}ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}ln(X_i)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}2ln(X_i)\mu}{2\sigma^2} - \frac{\sum_{i=1}^{n}\mu^2}{2\sigma^2}$$

$$= -\frac{n}{2}ln(2\pi\sigma^2) - \sum_{i=1}^{n}\ln(X_i) - \frac{\sum_{i=1}^{n}ln(X_i)^2}{2\sigma^2} + \frac{\sum_{i=1}^{n}ln(X_i)\mu}{\sigma^2} - \frac{n\mu^2}{2\sigma^2}$$

In order to find the maximum likelihood estimators, the two parameters $\hat{\mu}$ and $\hat{\sigma}$ need to be determined, such that the equation $\mathcal{L}(\mu, \sigma^2|X)$ is maximized. This is achieved by taking the gradient of $\mathcal{L}$ with respect to μ and σ^2 then setting it equal to 0.

$$\nabla\,\mathcal{L} = \frac{\delta\mathcal{L}}{\delta\mu} = \frac{\sum_{i=1}^{n}ln(X_i)}{\hat{\sigma}^2} - \frac{2n\hat{\mu}}{2\hat{\sigma}^2} = 0$$

$$\rightarrow\ \frac{n\hat{\mu}}{\hat{\sigma}^2} = \frac{\sum_{i=1}^{n} ln(X_i)}{\hat{\sigma}^2}$$

$$\rightarrow\ n\hat{\mu} = \sum_{i=1}^{n} ln(X_i)$$

$$\rightarrow\ \hat{\mu} = \frac{\sum_{i=1}^{n} ln(X_i)}{n}$$

Analogous $\nabla\,\mathcal{L}$ with respect to σ^2 will yield:

$$\nabla\,\mathcal{L} = \frac{\delta\mathcal{L}}{\delta\sigma^2} = -\ \frac{n}{2\hat{\sigma}^2} + \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{2(\hat{\sigma}^2)^2} = 0$$

$$\rightarrow\ \frac{n}{2\hat{\sigma}^2} = \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{2\hat{\sigma}^4}$$

$$\rightarrow\ \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(ln(X_i) - \hat{\mu})^2}{n}$$

$$\rightarrow\ \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}\left(ln(X_i) - \frac{\sum_{i=1}^{n} ln(X_i)}{n}\right)^2}{n}$$

A-3. PROBABILITY DISTRIBUTION (POSTERIOR PROBABILITY) OF A BIVARIANT NORMAL DISTRIBUTION WITH MULTIPLE OBSERVATIONS[52]

The derivation of the posterior probability for a multivariant gaussian is shown in the normal space. For a lognormal distribution the steps are the same, with the only difference that the data is exponentiated.

Assume a normal distributed prior with fixed variance such that,

$$\mu \sim N(\mu_0, \sigma_0^2)$$

With μ_0, σ_0^2 being the prior hyperparameters. Given multiple independent observations X,

$$X_i | \mu \sim N(\mu, \sigma^2)$$

Which yields

$$f(\mu|\underline{x}) \propto f(\mu) f\left(\underline{x_i}|\mu\right)$$

$$= \frac{1}{\sqrt{2\pi\sigma_0^2}} exp\left[\frac{-(\mu-\mu_0)^2}{2\sigma_0^2}\right] \times \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} exp\left[\frac{-(x_i-\mu)^2}{2\sigma_0^2}\right]$$

$$= \frac{1}{(2\pi)^{\frac{(n+1)}{2}} \sqrt{\sigma_0^2 \sigma^{2n}}} exp\left[\frac{-\mu^2 + 2\mu\mu_0 - \mu_0^2}{2\sigma^2} - \sum_{i=1}^{n} \frac{x_i^2 + 2\mu x_i - \mu^2}{2\sigma^2}\right]$$

Since the product of two Gaussian is a Gaussian

[52] (Murphy 2007), (Stony Brook University 2019)

$$\propto exp\left[\frac{-\mu^2(\sigma^2+n\sigma_0^2)+2\mu(\mu_0\sigma^2+\sigma_0^2x_1+\cdots+\sigma_0^2x_n)-(\mu_0^2\sigma^2+\sigma_0^2x_1^2+\cdots+\sigma_0^2x_n^2)}{2\sigma_0^2\sigma^2}\right]$$

$$\propto exp\left[\frac{-\mu^2+2\mu\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}-\left(\frac{\mu_0^2\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\right]$$
$$\times exp\left[-\frac{\mu_0^2\sigma^2+\sum_{i=1}^n\sigma_0^2x_1^2}{2\sigma_0^2\sigma^2}\right]$$

$$\propto exp\left[-\frac{\left(\mu-\frac{\mu_0\sigma^2+\sum_{i=1}^n\sigma_0^2x_i}{\sigma^2+n\sigma_0^2}\right)^2}{2\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}}\right]$$

Setting σ_n^2 and μ_n as (alt. matching coefficients):

$$\sigma_n^2=\frac{\sigma_0^2\sigma^2}{\sigma^2+n\sigma_0^2}=\frac{1}{\sigma^{-2}+n\sigma_0^{-2}}$$

$$\mu_n=\frac{\mu_0\sigma^2+\sum_{i=1}^n x_i\sigma_0^2}{\sigma^2+n\sigma_0^2}=\frac{\mu_0\sigma_0^{-2}+\sum_{i=1}^n x_i\sigma^{-2}}{\sigma_0^{-2}+n\sigma^{-2}}=\sigma_n^2\left(\mu_0\sigma_0^{-2}+\sum_{i=1}^n x_i\sigma^{-2}\right)$$

Rewritten as

$$\sigma_n^2=\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)$$

and

$$\mu_n=\sigma_n^2\left(\frac{\mu_0}{\sigma_0^2}+\frac{n\bar{x}}{\sigma^2}\right)$$

Therefore the posterior distribution is

$$X_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

Recall the pearson coefficient expressed in terms of the covaraince:

$$\rho_{x,y} = \frac{cov(x,y)}{\sigma_x \sigma_y} \xrightarrow{yields} cov(x,y) = \rho_{x,y}\sigma_x\sigma_y$$

Thus, the variance and mean expressed with ρ becomes

$$\sigma_n^2 = (1-\rho^2)\sigma_0^2$$

$$\mu_n = \mu_0 + \rho(x_i - \mu)\frac{\sigma_0}{\sigma}$$

A-5. Validity results for reference class

Table of uplift factors for the "leave-one-out" process to test the robustness of the reference class in chapter 7.

Field	P10	P50
28	0.104	0.563
30	0.104	0.583
32	0.104	0.583
31	0.104	0.583
7	0.083	0.542
33	0.104	0.583
39	0.104	0.604
20	0.083	0.563
6	0.083	0.542
16	0.083	0.563
36	0.104	0.604

43	0.104	0.604
10	0.083	0.563
8	0.083	0.563
0	0.104	0.563
5	0.083	0.542
1	0.104	0.542
47	0.104	0.604
17	0.083	0.563
34	0.104	0.583
44	0.104	0.604
45	0.104	0.604
37	0.104	0.604
22	0.104	0.563
21	0.104	0.563
11	0.083	0.563
12	0.083	0.563
14	0.083	0.563
25	0.104	0.563
19	0.083	0.563
38	0.104	0.604
18	0.083	0.563
27	0.104	0.563
2	0.104	0.542
41	0.104	0.604
48	0.104	0.604
42	0.104	0.604
9	0.083	0.563
23	0.104	0.563
35	0.104	0.583
15	0.083	0.563
40	0.104	0.604
26	0.104	0.563
24	0.104	0.563
29	0.104	0.583
3	0.104	0.542
13	0.083	0.563
46	0.104	0.604
4	0.104	0.542

12.References

Begg, Steve and Bratvold, Reidar. 2008. Systematic Prediction Errors in O&G Project and Portfolio Selection. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA.

Bordley, Robert. 2014. Reference Class Forecasting: Resolving Its Challenge to Statistical Modeling. *The American Statistician* **68** (4): 221-229.

Buhl, Soren, Flyvbjerg, Bent, and Holm, Mette. 2002. Underestimating Costs in Public Works Projects: Error or Lie? *Journal of the American Planning Association* **68** (Summer): 279-295.

Buhl, Soren, Flyvbjerg, Bent, and Holm, Mette. 2003. How common and how large are cost overruns in transport infrastructure projects? *Transport Reviews* **23** (1): 71-88.

Cantarelli, C., Molin, C., van Wee, J. et al. 2012. Different Cost Performance: Different Determinants? The Case of Cost Overruns in Dutch Transport Infrastructure Projects. *Transport Policy* **22** (1): 49-56.

Capen, E. 1976. The Difficulty of Assessing Uncertainty. *SPE Journal of Petroleum Technology* **08** (5579).

Consortium, York Health Economics. 2019. Base Case Analysis [online], 2019).

Ernst & Young, EY. Megaprojects - The New Norm in the Oil and Gas Industry.

Flyvbjerg, Bent. 1996. The Dark Side of Planning: Rationality and "Realrationalität", Center for Urban Policy Research Press.

Flyvbjerg, Bent. 2005. Policy and Planning for Large-Infrastructure Projects: Problems, Causes, Cures.

Flyvbjerg, Bent. 2011. Over Budget, Over Time, Over and Over Again: Managing Major Projects. In, 321-344.

Flyvbjerg, Bent. 2014. What You Should Know about Megaprojects and Why: An Overview. *Project Management Journal* **45** (April-May).

Flyvbjerg, Bent, Garbuio, Massimo, and Lovallo, Dan. 2014. Delusion and Deception in Large Infrastructure Projects: Two Models for Explaining and Preventing Executive Disaster. In, 468-491.

Flyvbjerg, Bent, Glenting, Carsten, and Rønnest, Arne. 2004. *Procedures for Dealing with Optimism Bias in Transportation Planning*. London (Reprint).

Flyvbjerg, Bent, Holm, Mette, and Soren, Buhl. 2005. How (In)accurate Are Demand Forecasts in Public Works Projects? *Journal of the American Planning Association* **71** (Spring): 131-146.

Flyvbjerg, Bent, Hon, Chi-keung, and Fok, Wing Huen. 2016. Reference class forecasting for Hong Kong's major roadworks projects. *Civil Engineering* **169** (CE6): 17-24.

Fondren, Mark, Duane, McVay, and Gonzalez, Raul. 2013. Applying Calibration to Improve Uncertainty Assessment. Presented at the SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, USA.

Gilovich, Thomas, Griffin, Dale, and Kahneman, Daniel. 2002. *Heuristics and biases: The psychology of intuitive judgment.* New York: Cambridge University Press.

Ginos, Brenda. 2009. *Parameter Estimation for the Lognormal Distribution*, Brigham Young University, Utah, USA.

Haukaas, A. and Mohus, E. 2016. *The Petroleum Industry's Inability and Lack of Will to Fulfil Its Promises - an Analysis of Overruns on the Norwegian Continental Shelf*, University of Stavanger, Stavanger.

Hawkins, J., Coppersmith, E., and Cunningham, P. 2002. Improving Stochastic Evaluations Using Objective Data Analysis and Expert Interviewing Techniques. Presented at the ATCE, San Antonio.

Hájek, Alan. 2007. *journal article*. The reference class problem is your problem too. *Synthese* **156** (3): 563-585. https://doi.org/10.1007/s11229-006-9138-5.

Kahneman, Daniel. 2011. *Thinking, fast and slow*. New York, NY, USA: Farrar, Straus and Giroux.

Kahneman, Daniel and Lovallo, Dan. 2003. Delusions of Success: How Optimism Undermines Executives' Decisions. *Harvard Business Review*.

Kahneman, Daniel and Tversky, Amos. 1979. Intuitive Prediction: Biases and Corrective Procedures. *TIMS Studies in Management Science* **12**: 313-327.

Kreifeldt, John and Nah, Keoun. 1995. Adding and Subtracting Percentiles - How bad can it be? *Proc.,* Proceedings of the Human Factors and Ergonomics Society Annual Meeting5.

Leleur, S., Salling, K., Pilkauskiene, I. et al. 2015. Combining Reference Class Forecasting with Overconfidence Theory for Better Risk Assessment of Transportation Infrastructure Investment. *European Journal for Transport and Infrastructure Research* **15** (3): 362-375.

Mohus, Erlend. 2018. *Over Budget, Over Time, and Reduced Revenue, Over and Over Again - An Analysis of the Norwegian Petroleum Industry's Inability to Forecast Production.*, University of Stavanger, Norway.

Moore, Simon. 2019. Strategic PPM - Project and Portfolio Management, https://strategicppm.wordpress.com/2010/11/16/project-failure-channel-tunnel/2019).

Murphy, Kevin. Conjugate Bayesian Analysis of the Gaussian distribution. *The University of British Columbia*.

Nandurdikar, Neeraj Shashikant and Wallace, Luke. 2011. Failure to Produce: An Investigation of Deficiencies in Production Attainment. Presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA. 2011/1/1/. https://doi.org/10.2118/145437-MS.

Norwegian Petroleum Directorate, NPD. 2018a. *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)* (Reprint).

Norwegian Petroleum Directorate, NPD. 2018b. *RNB2019 General Guidelines.* (Reprint).

Norwegian Petroleum Directorate, NPD. 2019. *Reporting to the Revised National Budget 2019 (RNB2019)* (Reprint).

Priemus, H., Flyvbjerg, B., and van Wee, B. 2008. *Decision-Making on Mega-Projects: Cost-Benefit Analysis, Planning and Innovation.* Cheltenham, UK, Edward Elgar Publishing (Reprint).

Sellier, Anne-Laure, Scopelliti, Irene, and Morewedge, Carey. 2019. Debiasing Training Improves Decision Making in the Field. *SAGE Journals* **30** (9): 1371-1379.

Tversky, Amos and Kahneman, Daniel. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* **185** (4157): 1124. http://science.sciencemag.org/content/185/4157/1124.abstract.

University, Stony Brook. Bayesian Inference to the Normal Distribution.
Welsh, Matthew and Begg, Steve. 2015. What have we learnt? Insights form a decade of bias research. *Australian Petroleum Production and Exploration Association Journal*.
Welsh, Matthew, Begg, Steve, and Bratvold, Reidar. 2006. Correcting Common Errors in Probabilistic Evaluations: Efficacy and Debiasing. Presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, USA.
Welsh, Matthew, Begg, Steve, and Bratvold, Reidar. 2007. Modelling the Economic Impact of Cognitive Biases on Oil and Gas Decisions. Presented at the ATCE, Anaheim, California.
Welsh, Matthew, Bratvold, Reidar, and Begg, Steve. 2005. Cognitive Biases in the Petroleum Industry: Impact and Remidation. Presented at the ATCE, Dallas.