Production Forecasting: Optimistic and Overconfident—Over and Over Again

Reidar B. Bratvold and **Erlend Mohus,** University of Stavanger, and **David Petutschnig** and **Eric Bickel,** University of Texas at Austin

Summary

The oil and gas industry uses production forecasts to make decisions, which can be as mundane as whether to change the choke setting on a well, or as significant as whether to develop a field. These forecasts yield cash flow predictions and value-and-decision metrics such as net present value and internal rate of return.

In this paper, probabilistic production forecasts made at the time of the development final investment decisions (FIDs) are compared with actual production after FIDs, to assess whether the forecasts are optimistic, overconfident, neither, or both.

Although biases in time-and-cost estimates in the exploration and production (E&P) industry are well documented, probabilistic production forecasts have yet to be the focus of a comprehensive, public study. The main obstacle is that production forecasts for E&P development projects are not publicly available, even though they have long been collected by the Norwegian Petroleum Directorate (NPD), a Norwegian government agency. The NPD's guidelines specify that at the time of FID, the operators should report the forecasted annual mean and P10/90 percentiles for the projected life of the field.

We arranged to access the NPD database in order to statistically compare annual production forecasts given at the time of FID for 56 fields in the 1995 to 2017 period, with actual annual production from the same fields. This work constitutes the first public study of the quality of probabilistic production forecasts. The main conclusions are that production forecasts that are being used at the FID for E&P development projects are both optimistic and overconfident, leading to poor decisions.[1]

Introduction

The oil and gas industry spends enormous resources on forecasting future production, costs, and project completion times. The cash flows from these forecasts are used to formulate value estimates, which in turn form the basis for investment decisions regarding new development, redevelopments, drilling and maintenance schedules, etc.

Many oil and gas projects around the world significantly overrun their cost-and-time forecasts (Merrow 2012; Norwegian Petroleum Directorate 2013; Ernst and Young 2014; Taraldsen 2015; Oglend et al. 2016; Smyth 2016). Providers of cost-and-time estimates often justify overruns by referring to unexpected events or managers modifying their numbers for tactical reasons. Furthermore, it is sometimes argued that if people knew in advance the real costs and challenges involved in delivering large projects, nothing would ever get developed—so it is better not to know, because ignorance helps get projects started (Flyvbjerg 2017). Finally, it is claimed that such cost overruns lead to increased ultimate recovery, although this has not been verified. Moreover, an analysis of development projects in Norway showed that increased costs do not correlate with increases in producible reserves (Haukaas and Mohus 2016).

Although we briefly discuss time overruns, our main focus is on production shortfalls. Nandurdikar and Wallace (2011) showed that significant production shortfalls historically have been the rule rather than the exception.

The main contributions of this paper are threefold. First, we extend the work of Nandurdikar and Wallace (2011) by using data from Norwegian fields to investigate whether optimism and overconfidence continue to plague production forecasts at the time of FID. We will demonstrate that the lack of production attainment, relative to forecasts is as bad now as it was in the mid-1990s, with no sign of improvement despite a significant increase in uncertainty modeling sophistication over the past two decades. Second, we introduce and discuss the concepts of delusions (honest mistakes) and deception (strategic manipulation of information), to examine the reasons for these shortfalls. Finally, we briefly discuss how forecasts can be improved by structuring incentives in a way that keeps everyone focused on company-wide goals.

Some Attributes of Probabilistic Forecasts

The purpose of this work is to investigate whether operators on the Norwegian Continental Shelf (NCS) are well-calibrated when they forecast production for new developments. A well-calibrated production forecast is unbiased and consistent with the forecasters' knowledge. A production forecaster is said to be well-calibrated when the set of his or her forecasts satisfies the following:

1. The range of actual production outcomes falls within the range of predicted production outcomes. If too many actual outcomes fall outside the range of predicted possible outcomes, the forecasters are overconfident (Tversky and Kahneman 1974). For the production forecasting context evaluated in this paper, approximately 80% of the actual production outcomes should be within the forecasted P10/P90 range.
2. The average of the forecasted production rates should be close to the average of the actual production rates. If this is not the case, the forecaster is either optimistic or pessimistic.

In general, forecasts tend to be affected by optimism and overconfidence. In an example using synthetic data for produced volume over some time period, these two biases are illustrated in **Fig. 1.**

Consider a forecaster who has forecasted the mean produced volume for enough fields to allow statistically significant inferences about the forecaster's extent of optimism and overconfidence. Let the distribution of these means be represented by the blue distribution in Fig. 1. Now let the distribution of the actual produced volume outcomes be represented by the orange distribution in Fig. 1. The

[1]The conclusions based on the analysis presented in this paper are limited to the set of fields from the NCS. However, other authors have demonstrated the optimism bias in production forecasts from fields around the world (Nandurdikar and Wallace 2011; Nandurdikar and Kirkham 2012).

This paper (SPE 195914) was accepted for presentation at the SPE Annual Technical Conference and Exhibition, Calgary, Alberta, Canada, 30 September–2 October 2019, and revised for publication. Original manuscript received for review 17 December 2019. Revised manuscript received for review 16 February 2020. Paper peer approved 17 February 2020.

mean of the forecast is greater than the mean of the actual outcomes (violating requirement 2), which indicates that the forecaster is optimistic. Furthermore, the range of possible outcomes captured by the forecasts is significantly narrower than the range of actual outcomes (violating requirement 1), which indicates that the forecaster is overconfident.

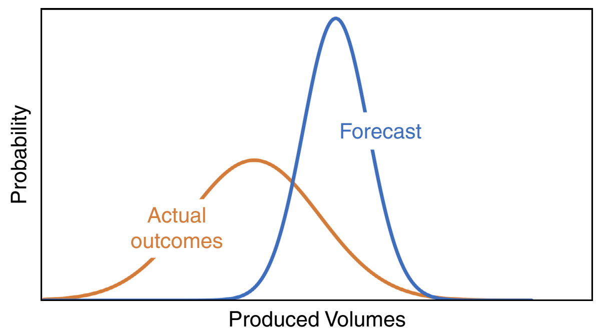

Fig. 1—Example using synthetic data—forecasted and actual outcomes for produced volume.

In a deterministic setting, or if the forecaster has been told to provide only a single produced volume forecast value for each field, we can track actual outcomes vs. forecasts by generating a scatterplot of the actual produced volume vs. the estimated produced volume.

The quality of probabilistic forecasts can be evaluated similarly. **Fig. 2** shows a scatterplot, on the basis of synthetic data, of the P50 forecasts and the associated 80% confidence intervals captured by the P10/P90 forecasts.

The graph shows the forecasted median (P50) values as blue markers and the low and high estimates as ranges ("error" bars) around the P50. The *y*-axis position is determined by actual results. If the forecasts are unbiased,

1. Approximately the same percentage of the P50-markers would be to the left of the 45-degree line, as are to the right of that line.
2. Approximately 80% of the P10/P90 confidence intervals would contain the associated actual value.

The analysis part of this paper will apply the above bias-detection tools to the NCS-based production forecasts.

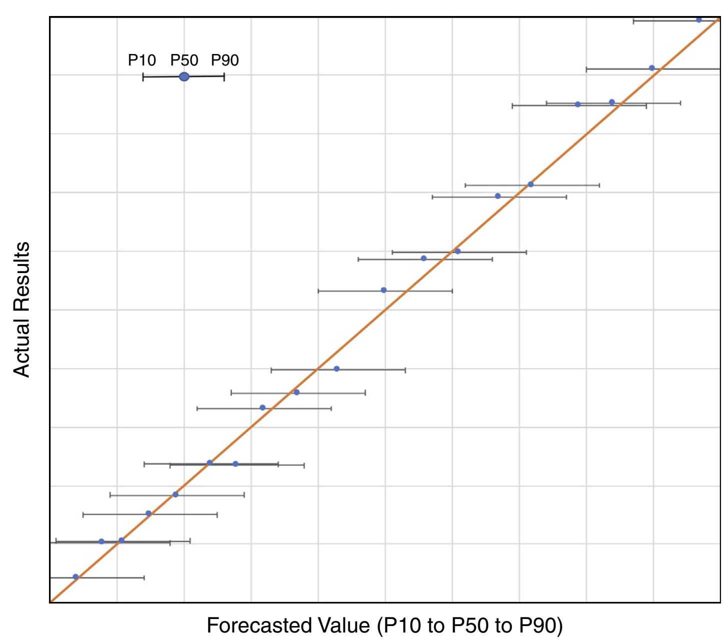

Fig. 2—A scatterplot of synthetic data showing the P50 forecast, the 80% confidence interval ("error" bars) captured by the forecasted P10/P90, and the actual value in brown.

Forecast and Production Data

In this paper, we analyze production data from development projects on the NCS approved from 1995 to 2017. Actual production data at field-level are publicly accessible through the NPD's website, but the production forecasts that the operators provide at the time of FID are not public. However, operators' production forecasts in support of FID are required to be included in the plan for development and operations submitted to the NPD.[2] The relevant guidelines for production reporting are provided in two documents issued by the NPD and the Ministry of Petroleum and Energy (Norwegian Petroleum Directorate 2018a, 2018b).

The forecast data are taken from the operators' annual reporting to the revised national budget (Norwegian Petroleum Directorate 2019a), and the data used as the FID data in this analysis consist of the data set provided by each operator immediately after the FID.

According to the NPD's factpage, (Norwegian Petroleum Directorate 2019b) a total of 85 oil and gas fields were approved for development from 1995 to 2017. Because this work is focused on oil production, forecasts for gas, natural gas liquids, and condensate

[2]For the purpose of this work, the NPD provided access to the forecast data under a confidentiality agreement. Hence, no company or field names will be used.

production were omitted from the analysis. We use production forecasts for oil that were provided at the year of (or closest to) the FID. Some fields were disqualified because of poor or missing data. Furthermore, only fields that had forecasted and actual production data prior to the end of 2017 are included. These criteria resulted in a final data set covering 56 fields.

The production forecasts include year-by-year estimates for the projected lifetime of each field, updated annually. The NPD guidelines make clear that a probabilistic approach should be used for production estimates and specify what the operators should report with respect to volume and production forecasts (Norwegian Petroleum Directorate 2018a). The guidelines state, "The planned production schedule and recovery rate should be described, and uncertainties should be highlighted" (Norwegian Petroleum Directorate 2018a). On page 37, the guidelines further state, "Expected production profiles for oil, gas, condensate, natural gas liquids and water for the entire field and for the individual deposits, and different production facilities, if applicable, must be stated along with the associated uncertainty. [...] All estimates shall be given with uncertainty, i.e., high (preferable P10), low (preferable P90), and base estimate (expected value). [...] These values should always apply when uncertainty is reported" (Norwegian Petroleum Directorate 2018b).[3] The best estimate is defined 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" (Norwegian Petroleum Directorate 2018a).

The early plan for development and operation guidelines did not specify what probabilities the low and high estimate should reflect (Norwegian Petroleum Directorate 2000; Ministry of Petroleum and Energy 2010). Therefore, the estimates of early forecasts might have reflected other probabilities. Because this information is not included in the available data, all estimates are assumed to reflect P10/Mean/P90 values.

Startup Delays

Many of the fields in our data set experienced startup delays. The average estimated development time[4] for a project on the NCS in the 1995 to 2017 period was 2 years and 8 months (Haukaas and Mohus 2016). The delay for these projects averaged 202 days, or approximately 20% of the average estimated time. Seventeen percent of projects started producing early, whereas 69% had initial production later than scheduled. Thirteen fields were more than 6 months delayed, and 17 fields overran their time schedules by more than 20%.

To avoid confounding the impact of poor production forecasts with the impact of time delays, we set the estimated first production year equal to the actual first production year. The elimination of startup delays reduced the data set from 56 fields and 603 forecast years to 55 fields and 549 forecast years.

Fig. 3 shows actual production and mean forecasted production aggregated for all 55 fields, and **Fig. 4** shows the fields' cumulative production by year and the number of fields included in the forecasted and actual cumulative production numbers shown. Again, in these graphs time delays have been eliminated so that actual production start equals forecasted production start for all of the fields.

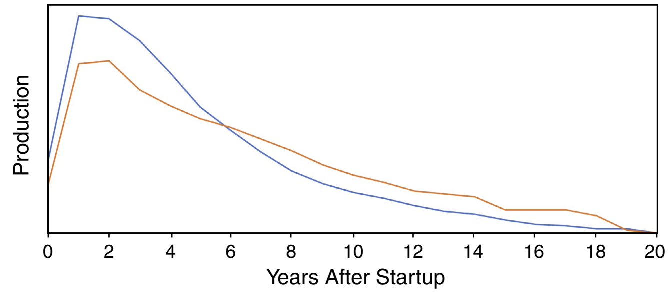

Fig. 3—Actual production and mean forecasted production for all fields.

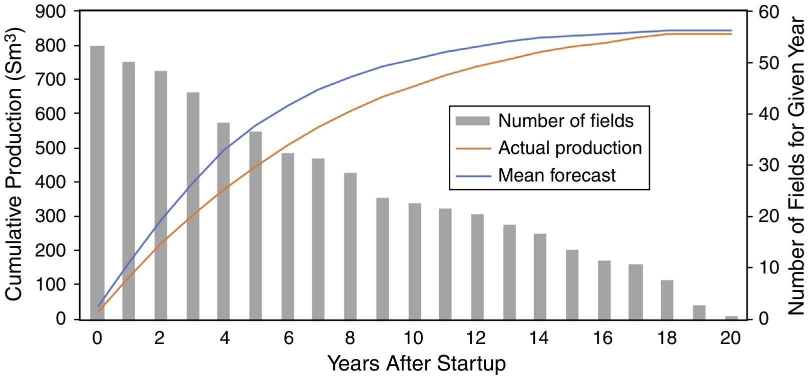

Fig. 4—Cumulative actual production and cumulative mean forecasted production for all fields.

Fig. 3 shows that for each of the first 5 years, average actual production from the 55 fields was short of average forecasted mean production. However, from year six onward, average actual production was higher than average forecasted mean production. Fig. 4 shows that in year 20, cumulative actual production is close to, but still below, cumulative forecasted mean production. However, as will be discussed in detail later, most of the fields delivering on the expected ultimate recovery required investment (e.g., redevelopment projects, additional wells) beyond what was used in the FID forecast. As shown in Fig. 4, the amount of data for later production years is

[3]In the analysis in this work, we use ascending cumulative density functions (CDFs), where P10 indicates a low value and P90 a high value.

[4]Development time is defined as the number of days from FID to actual production start.

too limited to support any statistical conclusions. However, it is likely that without the additional investments, many of the fields would have had lower cumulative production volumes.

In addition to the experienced development time slippages, frequent cost overruns were observed. Lost value due to cost overruns, schedule overruns, and underproduction for field developments on the NCS in the 1995 to 2017 period was estimated at NOK 474 billion, equivalent to USD 56.8 billion in 2017 nominal terms (Mohus 2018).

Analysis of Forecasted Production Profiles vs. Actual Production Profiles

As the base-case methodology is common practice in the oil and gas industry, we assume that the central number often represents the base case, viewed as the most likely outcome and that, in general, the numbers are not based on a consistent and coherent uncertainty quantification. Upon closer inspection of the production forecast data, we assume that forecasters have often used the expected value to populate the mean and P10/P90 values. Nevertheless, we believe that over the period investigated here, there has been an increasing trend in using probabilistic production forecasts with varying degrees of sophistication. This is supported by a plethora of papers discussing production uncertainty modeling (Barker et al. 2001; Hegstad and Henning 2001; Omre and Lødøen 2004; Rivera et al. 2007; Allen 2017). However, a key goal of this study is to assess the quality of production forecasts provided by the operators. Thus, we will assume that the operators have provided production estimates as suggested by the NPD guidelines (i.e., P10/mean/P90 values for each year). This, then, allows us to simply add a given field's central numbers (i.e., the means) for each year to get the cumulative mean forecasted production for that field. This can be directly compared against the cumulative actual production for that field. However, as described in more detail later, it would be incorrect to likewise aggregate the P10/P90 values.

The following discussion will address cumulative production over the first 4 years (years 0 to 3) of field life, which period will be denoted by F4Y. A period of 4 years was chosen because (1) production beyond that period has much less impact on the net present value of cash flows from the project and, consequently, on the development decision; (2) beyond this period, fields are much more likely to be subject to redevelopment, which would not have been specified in the FID; and (3) the results and conclusions drawn from the analysis are not very sensitive to the number of aggregation years in the range of 2 to 7.

Production Aggregation. As pointed out earlier, a given field's cumulative mean forecasted production for the F4Y can be obtained by summing the mean forecasted production for each of those years. However, obtaining a field's cumulative forecasted P10/P90 for the F4Y 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.

Being that no distributions with closed-form expressions would fit these three parameters, we used the lognormal distribution because (1) it has a lower bound of zero, and (2) its two parameters, the mean and standard deviation, can be matched using the mean and P10, the mean and P90, or the P10 and P90. For reasons that will become clear, we generated the lognormal mean and standard deviations using the forecasted mean and P10. Appendix A discusses the details of the fitting algorithm. Refer to Appendix A for an example that shows the different distributions for the three different lognormal fitting methods for one of the fields.

Ten of the 55 fields remaining in the data set were too recently approved for development, to have 4 years of production. For these fields, we included only the actual production years in the aggregation. Furthermore, some of the forecasts did not include the P10/P90 forecasts or had set them equal to the mean forecasts, which prevented the fitting of the lognormal parameters. This left 32 fields used in the aggregation of the forecasted and actual production for the F4Y.

After creating fitted distributions for each year's forecasted production for a given field, the means and variances for each year were added and used to identify a lognormal distribution for the field's cumulative forecasted production over the F4Y. The resulting probability distribution allows the identification of any statistics, including the P10, mean, and P90.

Figures 5 and 6 show the F4Y forecasted production distributions for two arbitrarily chosen fields.

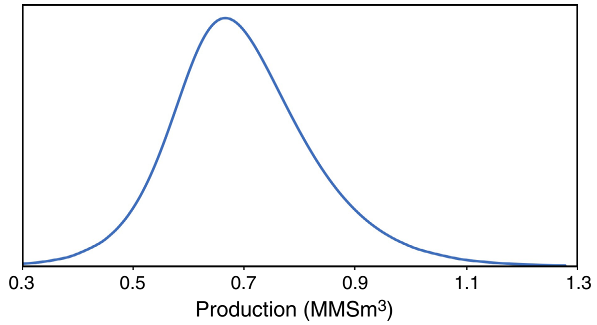

Fig. 5—Forecasted production for the F4Y, Field A.

Analysis. From each field's F4Y forecasted production distribution, we extracted the P10, P50, and P90 to compare estimated production with actual production in a probabilistic sense (i.e., to construct probabilistic scatterplots in the format used in Fig. 2).

The input parameters used for the distribution fitting exhibit positive skew, as the ratio $(\mathrm{P90}-\mathrm{P50})/(\mathrm{P50}-\mathrm{P10})>1$. Thus, the fitted lognormal distributions also exhibit positive skew with mean values > P50. Pearson's second skewness coefficient was determined for each of the 32 fields used in the F4Y calculations. The resulting skewness coefficients range from 0.04 and 0.8, with an average of 0.2.[5] 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 the F4Y, using the mean forecast to represent the P50 in evaluating optimism bias is reasonable.

[5]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).

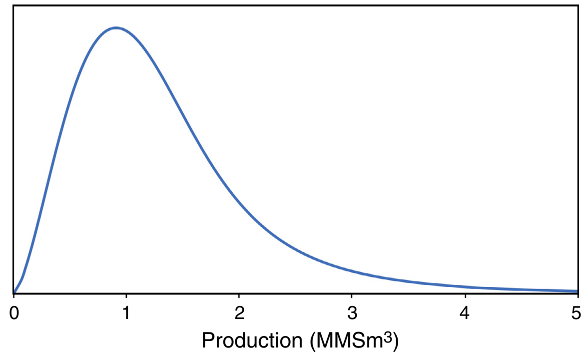

Fig. 6—Forecasted production for the F4Y, Field B.

Overconfidence. **Fig. 7** includes the data for the remaining 32 fields. The horizontal axis shows the forecasted F4Y P10/P50/P90, while the vertical axis shows the corresponding actual four-year production for a given field. More than half of the P50 forecasts are on the right side of the 45-degree line (optimism bias), and fewer than 80% of the range bars touch the solid brown line (overconfidence bias). These biases are evident in the forecasts for fields with more than 10 $MMSm^3$ oil produced in the F4Y. **Fig. 8** provides clearer scatterplots for the fields with less than 10 $MMSm^3$, which show the same forecast bias as for the larger fields.

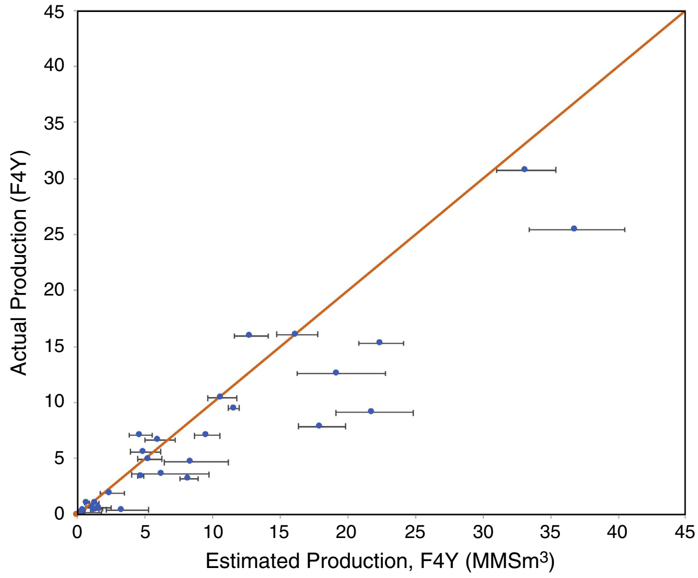

Fig. 7—Scatterplot showing all fields.

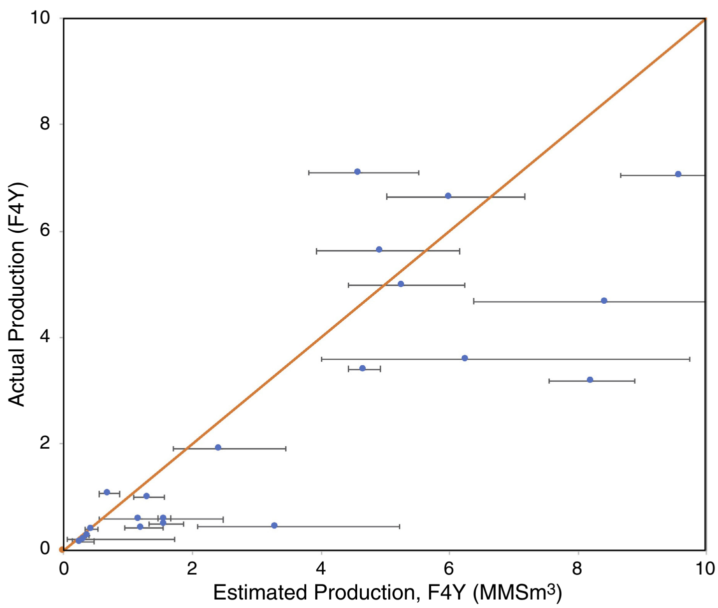

Fig. 8—Scatterplot showing fields with mean forecasts for the F4Y less than 10 $MMSm^3$.

Of the 32 field scatterplots, only 10 (31%) touch the 45-degree line, indicating overconfidence. For 27 of the 32 field scatterplots (84%), the P50 markers fall below the 45-degree line, indicating optimism.

Optimism. The degree of optimism can be assessed by comparing the cumulative distribution of the normalized actual production for the F4Y for all 32 fields, with the normalized forecasted mean and forecasted P10 production for the F4Y for those fields, as shown in **Fig. 9.**

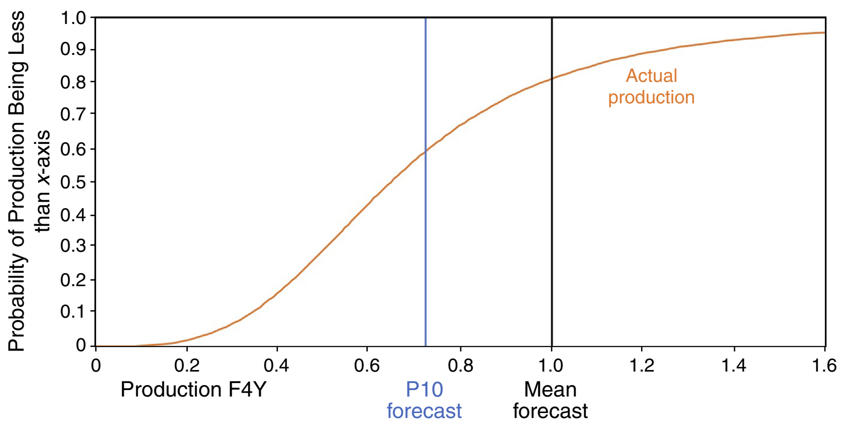

Fig. 9—Cumulative distribution for normalized actual production, normalized forecasted mean, and P10 production for the F4Y for all 32 fields.

The forecasted production has been normalized so that the mean forecast is 1.0. Each field's actual production has also been normalized by the field's mean forecast, to allow the distribution of forecasted mean and P10 production to be compared with the distribution of actual production. The P10 production forecasts aggregated over all fields were obtained using Monte Carlo simulation. Because the ratio of the forecasted mean production vs. actual production varies by field, we fit a metalog distribution (Keelin 2016) to the normalized actual production data, shown in orange in Fig. 9.[6] For 84% of the fields, the normalized actual production was less than the normalized mean forecast production ($x = 1.0$). Similarly, for 59% of the fields, the normalized actual production was less than the normalized P10 forecast production; and for approximately 30% of the fields, the normalized actual production was less than half of the normalized forecasted mean production. The impact of these forecasting biases would have been even greater if we had included the biased forecasts for time to first oil.

Table 1 summarizes the results discussed above and for reference gives the percentiles that would indicate an unbiased forecast. It indicates that only 31% of the fields had actual production that fell within the forecasted 80% range.

	Under P10	Under P50	Under P90	Inside P10 to P90
Unbiased forecasts	10%	50%	90%	80%
Actual forecasts	59%	84%	90%	31%

Table 1—Percentages of fields whose normalized actual production were below the normalized forecasted P10, P50, and P90 for the F4Y.

Sensitivity Analysis on the Number of Aggregation Years. We chose the FnY with $n = 4$ for our analysis. Would a different n lead to different conclusions? **Fig. 10** shows the results of a sensitivity analysis of percentage of fields whose cumulative actual production did not exceed the field's cumulative forecasted P10 or P50, to number of years n.[7] 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 and P10 forecasts for all n years decreases as n increases. Also, both the Actual $q <=$ P50 forecast and the Actual $q <=$ P10 forecast are not very sensitive to the number of years included in the analysis, with the Actual q being the actual production.

Sensitivity Analysis on Field Size. To allow for the possibility that forecasters (or companies) might be more inclined toward optimism when fields are small, being that small fields tend to be more marginal economically than are larger fields, a sensitivity analysis was conducted on field size, shown in **Fig. 11.** The vertical axis represents the fraction of fields for which the actual production is less than or equal to the P50 and P10 forecasts, and the horizontal axis represents field sizes in $MMSm^3$. The number of fields included in the analysis is indicated in black font above the bars and differs between field sizes. Although there might not seem to be enough large fields to make the statistics valid for these fields, there does not seem to be any trend indicating reduced bias with increasing field size.

Causes and Root Causes of Underperformance

Flyvbjerg (2011) distinguished "causes" from "root causes" in explaining cost overruns, benefit shortfalls, and delays in major projects. In the oil and gas industry, the following are often cited as causes of production shortfalls: project complexity, scope changes, technological uncertainty, drilling delays, and unexpected geological features. However, the root cause of underperformance is that forecasters

[6]The metalog distribution was chosen because it is simple, flexible, and easy to fit to data.

[7]Year 8 was the first year of redevelopment across these fields.

and project planners tend to systemically underestimate or even ignore the possibility of complexity, scope changes, drilling delays, etc., during project development and decision-making (Flyvbjerg 2017). Such ignorance or underestimation of uncertainty is what leads to optimism and overconfidence, whereas complexity, scope, technology, etc., are simply specific issues about which planners have been optimistic and overconfident and through which these biases manifest themselves. Scope changes and unexpected events are so common in major projects that the risk of their occurrence should clearly be considered in sound project preparation. But, again, such risks are typically ignored or underestimated, and that is the root cause of underperformance.

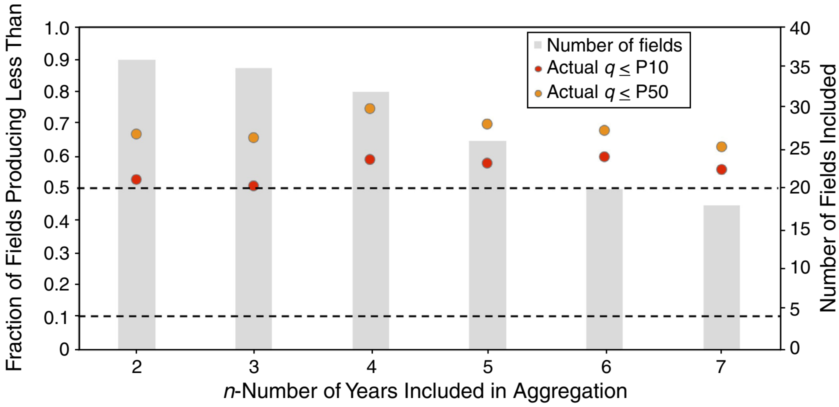

Fig. 10—Sensitivity analysis of how the number of aggregation years affects percentage of fields whose cumulative actual production does not exceed their cumulative forecasted P10 or P50.

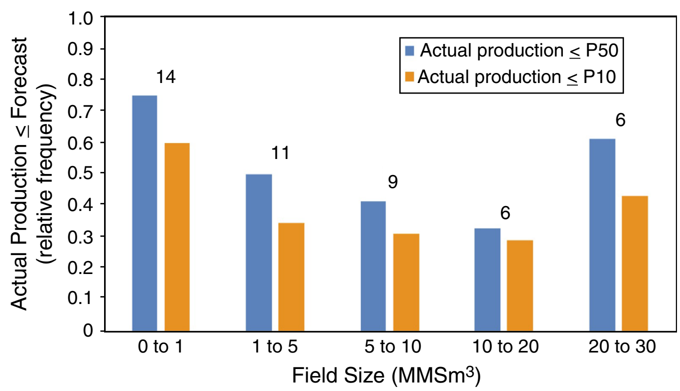

Fig. 11—Sensitivity analysis on field size with regard to optimism bias.

This discussion will focus on root causes of underperformance and not on conventional causes. The underlying reasons for the forecasting errors presented above fall into two categories: (1) delusions or honest mistakes, and (2) deceptions or strategic manipulation of information or processes (Flyvbjerg 2017).

Delusions. Tversky and Kahneman (1974) demonstrated that decision makers and forecasters fall victim to what they called the planning fallacy. The planning fallacy manifests itself when people focus on specific and unique characteristics of the project, which leads forecasters to optimism rather than a rational assessment of possibilities and probabilities. The result is a tendency of overlooking the potential for mistakes and miscalculations, leading to optimistic and overconfident forecasts based on a presumption of success. As a result, the operators pursue initiatives that are unlikely to deliver on production forecasts or expected financial returns. Lovallo and Kahneman (2003) asserted that these biases often are the result of taking an "inside view" in forecasting: forecasters have a strong tendency to consider projects as unique and, thus, focus on the particulars of the case at hand when forecasting. Adopting an "outside view," where information is put in a distributional setting of comparable past projects has been shown to reduce delusions in estimates of costs and schedules (Lovallo and Kahneman 2003; Flyvbjerg et al. 2014).

Deceptions. According to the second explanation—deception—forecasts are flawed by strategic misrepresentation or principal-agent problems (Flyvbjerg et al. 2014). Whereas delusions are not deliberate, deceptions are due to the different preferences and incentives of the forecasters and decision makers. Project planners, forecasters, and decision makers might deliberately and strategically overestimate production (and underestimate costs and time for delivery) to increase the chance that their projects gain approval and funding. They deliberately spin scenarios of success and gloss over the potential for disappointment. This misrepresentation and resulting disappointment can be moderated by enhancing transparency, aligning incentives, and providing decision clarity.

Bad Luck and Unexpected Events

The analysis performed in this work is restricted by there being many project details that were unavailable. However, not having access to the project details ensures that we take an outside view and do not focus on specific or unique characteristics of the project. For every one of the projects with biased forecasts, the forecaster might be able to provide detailed explanations of why the actual production fell short of the forecasted production.

Unexpected events[8] and bad luck, or the unfortunate resolution of one of the major project uncertainties mentioned above, is a reason typically given for a poor outcome. However, the bad-luck argument is refuted by statistical tests. Explanations that account for underperformance in terms of bad luck or unexpected events have survived for decades only because data on project performance have generally been of low quality (i.e., data have been disaggregated and inconsistent) because they came from small samples that did not allow rigorous statistical analyses.

First, if underperformance were truly caused by bad luck and error, we would expect a relatively unbiased distribution of error in performance, with mean error close to zero. However, the empirical data in this study, as well as in other studies on cost overruns and schedule delays, show with high statistical significance that the distribution of error is exceedingly biased and has a mean far from zero.

Second, if bad luck or unexpected events were main explanations of underperformance, we would expect performance to improve, because in a professional setting errors and their sources would be recognized and addressed through the refinement of data, methods, etc., much like in weather forecasting or medical science. As discussed below, substantial resources have in fact been spent over several decades on improving forecasts, including probabilistic forecasting. Nonetheless, this has not led to improved performance in terms of lower cost overruns or production shortfalls. Bad luck or error, therefore, do not appear to explain the data. It is not so-called estimation "errors" or their causes that need explaining. It is the fact that, deliberately or not, in the vast majority of projects, risks of scope changes, schedule delays, high complexity, unexpected geological features, etc., are systemically underestimated, resulting in underestimated costs and overestimated production.

We might agree with proponents of conventional explanations that it is, for example, impossible to predict for the individual project exactly which scope change, complexity, or geological problem will materialize and bring about less than expected production. However, it is possible to predict the risk, on the basis of experience from previous projects, that some such problems will haunt a project and how this will affect production. Clearly, such risks can and should be accounted for in forecasts of production, but typically are not. Moreover, some projects might be prone to what Taleb (2007) calls "black swans" (i.e., extreme events with low probability and high impact), but for the development projects in a well-known geological basin with reasonably well-known technology and concepts, the main reason our forecasts are biased is not black swans or "unknown unknowns," but rather an optimistic and overconfident assessment of key uncertainties. For valid explanations of underperformance, we need to look at explanations in terms of optimism, overconfidence, and strategic misrepresentation.

Do Operators Learn from Their Mistakes?

When one of the authors worked on production forecasting for a major new development on the NCS in the mid-1980s, the word "uncertainty" was not part of the vocabulary. Using then state-of-the-art production models, the forecast for each year in the anticipated production life of 10 to 30 years consisted of a single value, often in double precision. Since then, the oil and gas industry has increasingly invested in and adopted probabilistic forecasting methods. An indicator of this growth is that from the first year to the last year of the time period for which we have production data, 1995 to 2017, the number of papers on the topic of probabilistic production forecasting has grown by more than 600%, as seen in **Fig. 12.**[9]

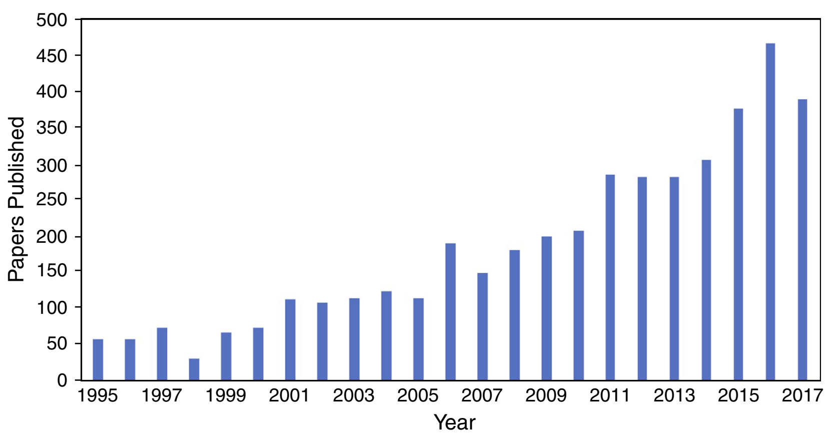

Fig. 12—Growth in the number of papers on probabilistic production forecasting over the 22-years period focused on in this paper.

Production forecasts on the NCS have also increasingly been probabilistic. Equinor (previously Statoil) operates the most fields on the NCS and is a joint venture partner in almost every field on the NCS. The company has been in the forefront in implementing probabilistic approaches, and Statoil-affiliated authors have published 59 papers on the topic since 1995. This quote by Neumann et al. (2012) confirms the company's view of the importance of probabilistic forecasting:

In Statoil there is a requirement to undertake a thorough uncertainty analysis before making an investment decision. The decision to go for the project or not, or the choice between different field development concepts, are made on mean values and not on deterministic reference values. Hence it is important to estimate mean production, mean net present value (NPV), mean cash flow, etc., for the different alternatives and decisions. Understanding the volume uncertainty and estimating the volume uncertainty range are important input for risk mitigation actions such as defining flexibility in the field development plan and finding robust solutions that are optimal over a range of outcomes rather than optimal on the reference case understanding of the field. … a good decision basis should include good estimates of the uncertainty in both the in-place volumes and in the dynamic properties affecting flow.

[8]Although "unexpected event" in colloquial English is understood to mean something not expected, its meaning or interpretation in a probabilistic setting, where the expected value is seldom attained by any single project, is nebulous. It could mean any deviation from the expected value or, as the term usually has a negative connotation, any negative deviation from the expected value. Does it refer to a black swan (Taleb 2007; i.e., an extreme event with low probability and high impact) or to any outcome outside the assigned probability distribution? Or maybe it is a synonym for "I erred in my forecast." The point is that using unexpected events as the reason for why our forecasts are biased, does not foster improvements. Calling something an unexpected event does not provide any guidance on how to improve and avoid such "surprises" in the future.

[9]The number of papers was identified using the following search specification: *probability OR probabilistic OR uncertainty AND production AND forecast.*

A pertinent question then, is whether the increased sophistication in uncertainty quantification has improved production forecasts. **Fig. 13** shows the production excess and shortfall for the F4Y for our study interval.

The horizontal axis shows the FID year, and each blue dot represents the F4Y production excess or shortfall relative to the actual production, for a field whose FID was approved in that year. For years 2015, 2016, and 2017, we used the first 3, 2, and 1 year, respectively. A dot above the 0-line indicates a higher production (excess) than the mean forecast, and a dot below the 0-line indicates a production shortfall. However, if there are improvements in the ability to produce unbiased, probabilistic forecasts over time, the average of all the dots for a given year should approach the 0-line. The red line on the graph shows a locally weighted smoothing (LOESS) curve across the entire time period.[10]

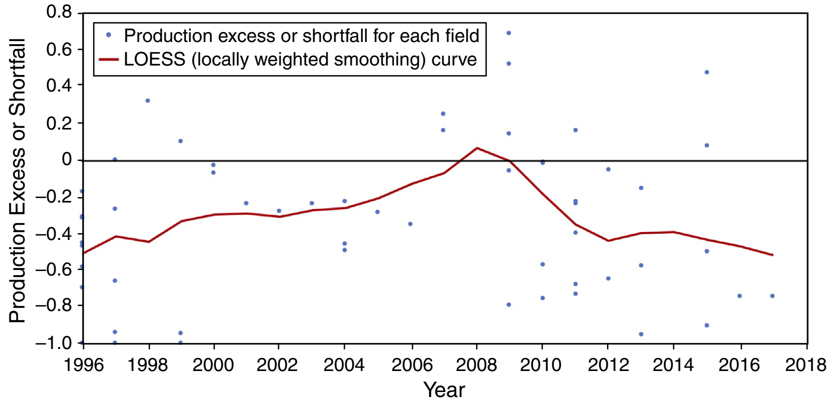

Fig. 13—Production excess or shortfall for the F4Y for each of the 56 fields, shown with the year in which the FID was approved.

The gradual change in the LOESS curve from a low of approximately −0.51 in 1996 to 0.07 in 2008 indicates forecasting improvements, as a result of reduced optimism in the mean. However, from 2008 to 2017 the LOESS curve has a predominantly negative slope, ending up in the same LOESS value of −0.51 as it had for 1996. Thus, the operators do not seem to learn and improve their forecasts over time. This could be because of delusions or deceptions. However, the empirical data show that the increased application of probabilistic forecasting models does not in and of itself remove bias or improve forecasting performance.

Oil Price Correlations

The behavior of the moving average (LOESS curve) in Fig. 13 might be partly correlated with oil price. This can be examined by depicting the data from that figure along with the Brent spot oil price, which is done in **Fig. 14.**

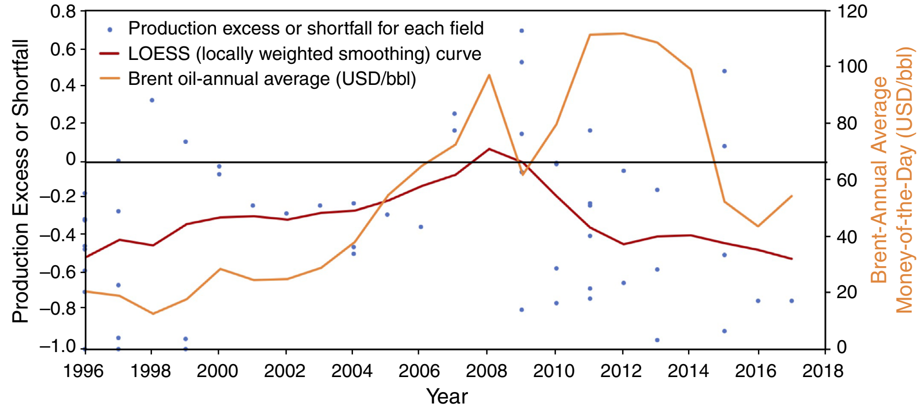

Fig. 14—Brent spot oil price and production excess or shortfall.

There does seem to be a strong positive correlation between the oil price and optimism bias reduction from 1996 to 2008. Then, the correlation become negative for a few years after 2008, before it again turns positive. **Fig. 15** shows the LOESS curve as a function of the annual average Brent spot price.

The upper left display of Fig. 15 shows the year-on-year LOESS vs. oil price, whereas the remaining graphs in Fig. 15 show the same with 1-, 2-, and 3-year delays of the LOESS curve lagging the oil price. The correlation, and explanatory power of the oil price, measured by R^2,[11] increases significantly as these lags are introduced. In other words, the optimism bias in production forecasts increases with lower oil prices, which can be observed from the positive slope in Fig. 15. Because this optimism is hard to attribute to delusion, deception might be at play: low oil prices reduce the profitability of projects, and a higher forecasted production rate is required to get the projects approved.

[10]Local regression, also known as local polynomial regression or moving regression, is a generalization of moving average and polynomial regression (Fox 2008; Merrow 2012). LOESS analysis does not require the specification of a function to fit a model to all of the data in the sample. LOESS is very flexible, making it ideal for modeling complex processes for which no theoretical models exist.

[11]R^2 is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model.

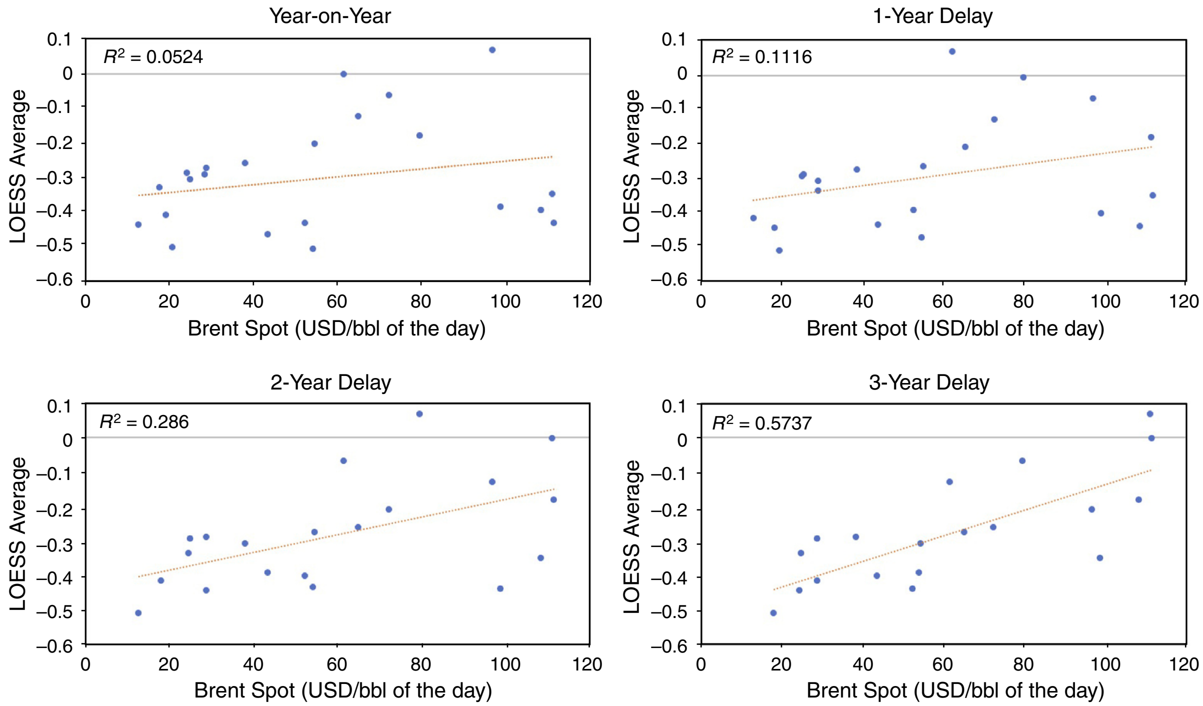

Fig. 15—Scatterplots of LOESS average vs. annual Brent spot price with 0-, 1-, 2-, and 3-years lag for the LOESS curve (production excess or shortfall).

The above discussion was focused on the lack of improvement in optimism bias regarding the mean estimate. Do the empirical data support a similar conclusion with respect to the overconfidence bias? Indeed, they do. **Fig. 16** shows the relative frequency of actual production (F4Y) being within the 80% band for each FID-approval year in our production-data interval.

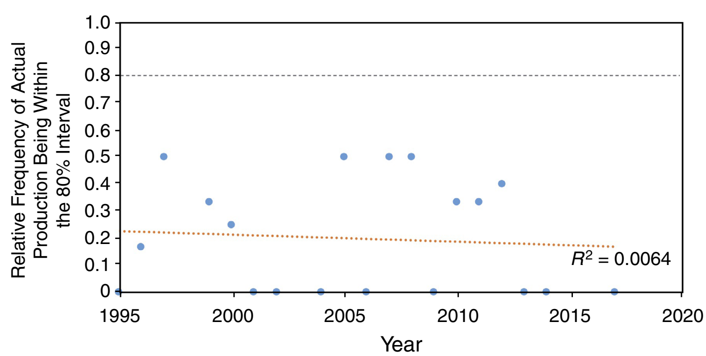

Fig. 16—Percentage of fields whose F4Y actual production fell within the 80% band as a function of time.

All of the blue dots fall below the 80% line, and the regressed line indicates a weak negative trend of the overconfidence bias over time. In other words, on average, the overconfidence in the forecasts over the past 22 years has been persistent, and the forecasts have not improved but, if anything, become slightly more overconfident.

Conclusions

Using empirical data from oil fields on the NCS, we have demonstrated that production forecasts provided by the operators at the time of the FID are biased; namely, optimistic and overconfident. Despite that the NPD requests consistent and unbiased probabilistic forecasts, there is an 84% chance that the actual production for the first 4 years will be less than the mean forecast and a 59% chance it will be less than the P10 forecast. Moreover, the empirical data show that there is only a 31% chance that the actual production will fall within the P10 to P90 range.

There are several reasons for the consistent underperformance in production forecasting. Delusions, or honest mistakes, stem from optimism and overconfidence biases. Deception is often in the form of strategic mismanagement to get one's own project approved. However, there also seems to be a lack of interest, ability, and willingness to improve. This could be due to a lack of awareness of poor forecasting performance or to deceptions. Bad luck does not hold any explanatory power in this context, being that the forecasts are consistently off in one direction (i.e., they are biased). The fact that optimism bias is correlated with oil price also suggests that deception might be at play.

The NCS production shortfall relative to forecasts is as poor now as it was 22 years ago, with no sign of improvement despite the significant increase in operators' uncertainty modeling competence over the past two decades. In contrast, weather forecasting, for example, has experienced significant improvements over the last two decades: 7-day forecasts made today are as accurate as 5-day forecasts were

22 years ago (Haiden et al. 2016). Unfortunately, in the oil and gas industry, the development of probabilistic forecasting systems has not been accompanied by a commensurate effort in developing procedures to assess the performance of these probabilistic forecasts.

We do not believe that the forecasting biases stem from poor thinking and analysis by individuals. Rather they are symptoms of a poor forecasting system or process. Any company that truly values unbiased forecasts should align its incentive and promotion schemes such that forecasters are truly encouraged and rewarded for providing unbiased forecasts. One such method to improve forecasts is using a reference class (Kahneman and Tversky 1979). Reference class forecasting provides an "outside view" to the project at hand through a reference class of comparable past projects. The results of a reference class forecasting application will be presented in a subsequent paper.

The conclusion from this study is not that probabilistic forecasting does not work and should be abandoned. Rather, the argument is that the current approach, with a focus of developing ever-more sophisticated methods for quantifying future production uncertainties, is not sufficient (Bickel and Bratvold 2007). It would be better to use simpler models or to ensure that any use of more complex models involved unbiased input. The first, and absolutely essential, step to improve forecasting ability is to keep score, track forecasting performance, and provide feedback.

Acknowledgment

We thank the NPD for providing access to the forecast data used in this work. The first author thanks Steve Begg of the University of Adelaide for having engaged in stimulating discussions on the issues and challenges addressed in this paper.

References

Allen, D. P. B. 2017. Handling Risk and Uncertainty in Portfolio Production Forecasting. *SPE Econ & Mgmt* **9** (2): 37–42. SPE-185178-PA. https://doi.org/10.2118/185178-PA.

Barker, J. W., Cuypers, M., and Holden, L. 2001. Quantifying Uncertainty in Production Forecasts: Another Look at the PUNQ-S3 Problem. *SPE J.* **6** (4): 433–441. SPE-74707-PA. https://doi.org/10.2118/74707-PA.

Bickel, J. E. and Bratvold, R. B. 2007. Decision-Making in the Oil and Gas Industry—From Blissful Ignorance to Uncertainty-Induced Confusion. Paper presented at the SPE Annual Technical Conference and Exhibition, Anaheim, California, USA, 1 January. SPE-109610-MS. https://doi.org/10.2118/109610-MS.

Bulmer, G. M. 1979. *Principles of Statistics*. New York, New York, USA: Dover.

Ernst and Young. 2014. *Megaprojects—The New Norm in the Oil and Gas Industry*. London, England, UK: EYGM Limited White Paper.

Flyvbjerg, B. 2011. Over Budget, Over Time, Over and Over Again: Managing Major Projects. In *The Oxford Handbook of Project Management*, 321–344. Oxford, England, UK: Oxford University Press.

Flyvbjerg, B. 2017. *The Oxford Handbook of Megaproject Management*. Oxford, England, UK: Oxford University Press.

Flyvbjerg, B., Garbuio, M., and Lovallo, D. 2014. Delusion and Deception in Large Infrastructure Projects: Two Models for Explaining and Preventing Executive Disaster. *Calif Mgmt Rev* **51** (2) : 170–193. https://doi.org/10.2307/41166485.

Fox, J. 2008. *Applied Regression Analysis and Generalized Linear Models*. Thousand Oaks, California, USA: SAGE Publications.

Haiden, T., Janousek, M., Bidlot, J. et al. 2016. Evaluation of ECMWF Forecasts, Including the 2016 Resolution Upgrade. ECMWF Technical Memorandum 792, Berkshire, England (December 2016). https://doi.org/10.21957/vltarm3ql.

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*. Stavanger, Norway: University of Stavanger.

Hegstad, B. K. and Henning, O. 2001. Uncertainty in Production Forecasts Based on Well Observations, Seismic Data, and Production History. *SPE J.* **6** (4): 409–424. SPE-74699-PA. https://doi.org/10.2118/74699-PA.

Kahneman, D. and Tversky, A. 1979. Intuitive Prediction: Biases and Corrective Procedures. *TIMS Stud Mgmt Sci* **12:** 313–327.

Keelin, T. W. 2016. The Metalog Distributions. *Decis Anal* **13:** 243–277. https://doi.org/10.1287/deca.2016.0338.

Lovallo, D. and Kahneman, D. 2003. Delusions of Success: How Optimism Undermines Executives' Decisions. *Harvard Business Review* **81** (7). https://hbr.org/2003/07/delusions-of-success-how-optimism-undermines-executives-decisions (accessed May 2020).

Merrow, E. W. 2012. Oil and Gas Industry Megaprojects: Our Recent Track Record. *Oil and Gas Fac* **1** (2): 38–42. SPE-153695-PA. https://doi.org/10.2118/153695-PA.

Ministry of Petroleum and Energy (MPE). 2010. *Guidelines for PDO and PIO*. Oslo, Norway: Ministry of Petroleum and Energy.

Mohus, E. 2018. *Over Budget, Over Time, and Reduced Revenue, Over and Over Again—An Analysis of the Norwegian Petroleum Industry's Inability To Forecast Production*. Stavanger, Norway: University of Stavanger.

Nandurdikar, N. S. and Kirkham, P. M. 2012. The Economic Folly of Chasing Schedules in Oil Developments and the Unintended Consequences of Such Strategies. Paper presented at the SPE Hydrocarbon, Economics, and Evaluation, Calgary, Alberta, Canada, 24 September. SPE-162878-MS. https://doi.org/10.2118/162878-MS.

Nandurdikar, N. S. and Wallace, L. 2011. Failure To Produce: An Investigation of Deficiencies in Production Attainment. Paper presented at the SPE Annual Technical Conference and Exhibition, Denver, Colorado, USA, 1 January. SPE-145437-MS. https://doi.org/10.2118/145437-MS.

Neumann, K. B., Hegstad, B., Bratli, E. et al. 2012. Uncertainty Study on In-Place Volumes in Statoil. Paper presented at the Ninth International Geostatistics Conference, Oslo, Norway.

Norwegian Petroleum Directorate (NPD). 2000. *PDO/PIO Guidelines*. Stavanger, Norway: Norwegian Petroleum Directorate.

Norwegian Petroleum Directorate (NPD). 2013. *Vurdering Av Gjennomførte Prosjekter På Norsk Sokkel*. Stavanger, Norway: Norwegian Petroleum Directorate.

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)*. Stavanger, Norway: Norwegian Petroleum Directorate.

Norwegian Petroleum Directorate (NPD). 2018b. *RNB2019 General Guidelines*. Stavanger, Norway: Norwegian Petroleum Directorate.

Norwegian Petroleum Directorate (NPD). 2019a. *Reporting to the Revised National Budget 2019 (RNB2019)*. Stavanger, Norway: Norwegian Petroleum Directorate.

Norwegian Petroleum Directorate (NPD). 2019b. *Factpage*. Stavanger, Norway: Norwegian Petroleum Directorate.

Oglend, A., Osmundsen, P., and Sindre, L. 2016. Cost Overruns in Norwegian Oil and Gas Projects: A Long-Tailed Tale. Paper presented at the IAEE Energy Forum, Bergen, Norway.

Omre, H. and Lødøen, O. P. 2004. Improved Production Forecasts and History Matching Using Approximate Fluid-Flow Simulators. *SPE J.* **9** (3): 339–351. SPE-74691-PA. https://doi.org/10.2118/74691-PA.

Rivera, N., Meza, N. S., Kim, J. S. et al. 2007. Static and Dynamic Uncertainty Management for Probabilistic Production Forecast in Chuchupa Field, Colombia. *SPE Res Eval & Eng* **10** (4): 433–439. SPE-100526-PA. https://doi.org/10.2118/100526-PA.

Smyth, J. 2016. *Cost Overruns Near $50bn as Australia's LNG Boom Falters*. Bracken House, London, England, UK: The Financial Times (Reprint).

Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York, New York, USA: Random House Group.
Taraldsen, L. 2015. *Norske Oljeprosjekter Har Sprukket Med over 200 Milliarder På 14 År*. Oslo, Norway: Teknisk Ukeblad.
Tversky, A. and Kahneman, D. 1974. Judgment under Uncertainty: Heuristics and Biases. *Science* **185** (4157): 1124. http://science.sciencemag.org/content/185/4157/1124.abstract.

Appendix A—Fitting a Lognormal Distribution Using the Mean and a Percentile or Two Percentiles

Let μ and σ be parameters (mean and standard deviation, respectively) of a normal distribution. Given the lognormal mean m and the value z for percentile α, we need to find μ and $\sigma > 0$.

Let Φ be the standard normal distribution function. The two pieces of information are

$$m = e^{\mu+\sigma^2/2} \Rightarrow \mu + \sigma^2/2 = \ln(m), \quad \text{(A-1)}$$

$$\ln(z) = \mu + \sigma\Phi^{-1}(\alpha). \quad \text{(A-2)}$$

The above equations can be rewritten and expressed in terms of μ

$$\mu = \ln(m) - \sigma^2/2, \quad \text{(A-3)}$$

$$\mu = \ln(z) - \sigma\Phi^{-1}(\alpha). \quad \text{(A-4)}$$

Subtracting Eq. A-3 from Eq. A-4 and multiplying by 2 gives

$$\sigma^2 - 2\Phi^{-1}(\alpha)\sigma + 2[\ln(z) - \ln(m)] = 0. \quad \text{(A-5)}$$

This can be solved with the quadratic formula

$$\sigma = \frac{-B \pm \sqrt{B^2 - 4AC}}{2A}, \quad \text{(A-6)}$$

where

$$A = 1,$$

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

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

There will be zero, one, or two solutions.

μ is a function of σ and can be calculated using either of the original equations; e.g.,

$$\mu = \ln(m) - \sigma^2/2. \quad \text{(A-7)}$$

A similar approach can also be used to fit two percentiles (e.g., the P10 and P90). Setting up the following equations, with known z_1, z_2 values for α_1, α_2 percentiles, will give

$$\ln(z_i) = \mu + \sigma\Phi^{-1}(\alpha_i). \quad \text{(A-8)}$$

Combining the information from the two percentiles yields

$$\sigma = \frac{\ln(z_2) - \ln(z_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}, \quad \text{(A-9)}$$

$$\mu = \frac{\ln(z_1)\Phi^{-1}(\alpha_2) - \ln(z_2)\Phi^{-1}(\alpha_1)}{\Phi^{-1}(\alpha_2) - \Phi^{-1}(\alpha_1)}. \quad \text{(A-10)}$$

Fig. A-1 compares the use of (1) P10 and P90, (2) the mean and P90, and (3) the mean and P10 for Year 0 for one NCS field.

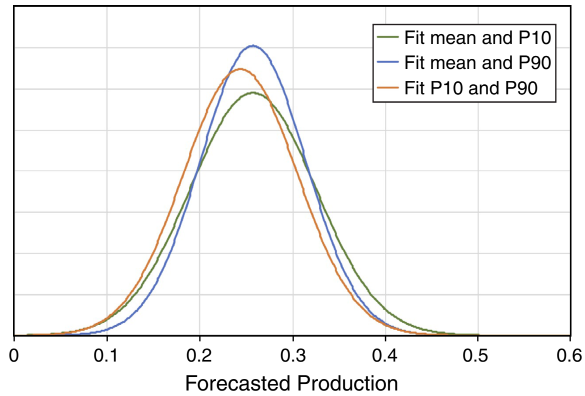

Fig. A-1—Lognormal distribution fit for one field in year 0, using P10 and mean, P90 and mean, and P10 and P90.