Hydrocarbon volume prediction performance in the Dutch subsurface and the impact of selection bias

by

Vincent van der Kraan

Utrecht University | Master's thesis for Earth, Life & Climate master (M-profile Geo-Resources)

Supervisors: H.L.J.G. Hoetz MSc. (EBN) and Dr. W.W.W. Beekman (Utrecht University)

March 2020 | Student no. 4275608

Utrecht University

ebn

Abstract

The E&P industry is frequently characterized by disappointing project outcomes. Specifically, the industry fails to deliver what is promised in term of hydrocarbon volumes due to overly optimistic predictions *(NPD Resource Report, 2018)*. Although this prediction bias problem is well-known amongst specialists involved, literature is scarce. Two suggested causes of the prediction bias are evaluation tool bias (e.g. imprecise seismic interpretation) and cognitive bias (e.g. individual motivational bias). In addition, Hoetz *(2016)* proposed the idea of a Selection Bias (SB) in the E&P industry. SB is based on the idea that more attractive prospects are assumed to be more matured; when these overly optimistic projects are drilled, they therefore result in disappointing volume delivery. In the Netherlands, the state-owned company EBN participates in virtually all E&P projects and has been reporting disappointing volumes for years *(EBN Focus, 2019)*. This study aims to address this problem by identifying and quantifying key parameters that contribute to the volume prediction bias.

First, using EBN data of the Dutch subsurface, a statistical look-back analysis is performed to check the quality of subsurface parameters used for volumetric assessments. Past volumetric performance is evaluated for both exploration and development wells in the time frame of 2004-2019 (215 cases). This is then broken down in relevant subsurface parameters. Using statistical tools, the prediction bias for recoverable volume, top reservoir depth, GWC, PHIE, Sw, GRV, NRV, N/G and pressure is quantified. Significant prognosis errors on single well scale are observed as well as a significant bias on portfolio scale. Prediction errors often indicate over-optimism. GRV and Sw are found to be major contributors to the observed volumetric bias.

Secondly, the effects of prediction bias are modelled on portfolio scale using synthetic portfolio modelling. A stochastically generated synthetic drilling portfolio is designed. Prospects are ranked based on attractiveness and drilled on paper. As the industry works with incomplete/noisy data the perception of a prospect often differs from reality. Hence for each prospect a prognosis is synthesized using Monte Carlo simulation. When the synthetic portfolio is drilled on paper the prediction quality is assessed by comparing the generated prognosis and its actual. Findings are that prediction bias can be modelled on portfolio scale based on the concept of SB. A volume bias is unavoidable due to SB as the ranking process prefers large prospects (being truly large or perceived as large). Matured portfolios in which prospects sizes are more clustered, might lead to increased SB. SB is a function of evaluation uncertainty though, so rigorous technical work might reduce SB.

The main outcome of this study is that the look-back analysis of the EBN portfolio shows a volume bias of 42%. This can partly be explained by SB, but other factors contribute. The findings of this study can help prioritize which parameters need more careful attention in reviewing project proposals. Furthermore, including the effect of EB in resource prediction tools might help to improve prognosis quality.

Keywords: E&P industry; prediction bias; Selection Bias; statistical look-back analysis, synthetic portfolio modelling

Content

List of Figures

List of Tables

Abbreviations

Avg.	Average
BRB	Basis Registratie Boringen (EBN database)
BRP	Basis Registratie Prospects (EBN database)
CI.	Confidence Interval
E&P	Exploration and Production
EAGE	European Association for Geoscientists and Engineers
EXP	Expectation volume
GIIP	Gas Initially In Place
GRV	Gross Rock Volume
IPRes	IPResource
MSV	Mean Success Volume
N/G	Net/Gross
NCS	Norwegian Continental Shelf
NRV	Net rock volume
NPD	Norwegian Petroleum Directorate
OCM	Operating Committee Meeting
OGA	Oil and Gas Authority
PHIE	Porosity
PoS	Probability of Success
RF	Recovery Factor
RV	Recoverable Volume
SB	Selection Bias
STD	Standard Deviation
Sw	Water Saturation
TCM	Technical Committee Meeting

1. Introduction

1.1 Problem statement

A solid understanding of the subsurface is of critical importance for successfully drilling wells. This applies especially to the oil and gas industry, where projects are of high complexity and large budgets are at stake. Projects in the Exploration and Production (E&P) industry are however frequently characterized by large degrees of uncertainty. This is related to incomplete data from the subsurface and challenging decision making which is based on a combination of hard data, soft data and interpretations. This uncertainty often results in project outcomes falling short of pre-drilling promises in term of resource volumes. Many studies show systematic underperformance of the E&P industry *(e.g. Milkov, 2017; Rudolph & Goulding, 2017; NPD Resource Report, 2018)*. Specifically, they report a failure of the industry to deliver promised volumes and value. This underperformance is proven to be directly related to overly optimistic evaluations of prospects. The prediction bias is a problem that is well known amongst specialists involved. Despite this awareness and despite remarkable technological advances in computing software and seismic technology, there are few signs of improvement over the last 30+ years *(NPD Resource Report, 2018)*.

For the Netherlands, the Dutch state-owned energy company EBN has been reporting disappointing volume outcomes for years *(EBN Focus, 2019)*. EBN participates as a non-operating partner in essentially all E&P projects in the Netherlands. With its extensive knowledge about drilling operations in the Netherlands, EBN is in a good position to assess the industries performance. Their 2019 annual report shows that over the past decade expected annual total volumes are consistently higher than the actual realized ones. Overall, for the period of 2008 until 2018 total found hydrocarbon volumes from exploration account to only 51% of the expected volumes *(EBN Focus, 2019)*. Internationally, a similar trend is observed. The recoverable resource volumes are reported to be generally 2-4 times less than predicted *(Quirk et al., 2018; and references therein)*.

Although the overprediction problem is well-known amongst specialists involved, no clear consensus on the cause exists. Surprisingly little work compiling exploration and production performance has been published and studies assessing possible explanations for the prediction bias in volumes are scarce. This is not surprising for the oil and gas industry is a highly competitive market with not only substantial amounts of money at stake, but also a strong reputational aspect. Most projects within the industry are therefore confidential and most companies are reluctant to share data, thereby limiting the possibility to perform reliable assessments of the industries performance.

Recently, at the 2018 annual European Association of Geoscientists & Engineers (EAGE) convention, 65 people met for a workshop to discuss the prediction bias. Suggested causes for the disappointing results are, amongst other things, problems with inputs to probabilistic models, general over-optimism, unrealistic Net/Gross-estimates and uncertainties in trap geometry and other rap-related issues *(Quirk et al., 2018)*. Another possible mechanism is the phenomenon of Selection Bias (SB) as formulated by Hoetz *(2016)*. Here is stated that as targets are selected with great care, they are also subject to large uncertainty. As the selection process favours large structures (or large resource volumes) those models that show large structures do have a greater likelihood of being matured as project. This evolves in a tendency of realizing (or drilling) overly optimistic projects which, statistically, result in disappointing project outcomes. Alternatively, if the selection criteria are ignored and drilling the exploration portfolio would take place randomly, no bias would show up.

In the current time of climate change and the energy transition that follows from it, where society has increasingly less confidence in the oil and gas industry, and with the prevailing unstable geopolitical environment, it is critical for the oil and gas industry to be able to deliver what is promised. Moreover, a lot of money is at stake and unfounded policy making based on biased data needs to be avoided. Furthermore, certain new energy systems, e.g. Geothermal and CCS, do also rely on subsurface estimates, and thus a prediction bias could be present also for these systems. Hence a better understanding of the factors that impact prediction uncertainty and prediction bias are relevant beyond the classic petroleum industry.

1.2 Objectives and research questions

The following research questions aim to address the above-mentioned problem:

What are the key parameters that contribute to the observed prediction bias in volumes?

- What is the quality of the prediction of the subsurface parameters being used for volumetric assessments? (E&P drilling projects)
- Can prediction bias effects on portfolio scale be modelled?

1.3 Research approach

To answer the research questions, the research is comprised of two components: (1) a statistical look-back analysis to quantify and decompose prediction bias, and (2) modelling the effect of prediction bias on portfolio scale. The statistical look-back analysis is based on EBN data and is focused on hydrocarbon reservoir volumes and underlying parameters that are used in making volume predictions. As EBN is a Dutch state-owned company only data concerning the Dutch subsurface are available and are used. The first step of this analysis is the design of a database for the pre-drill vs post-drill hydrocarbon volume estimates. Next, to decompose the prediction bias in volumes, similar databases are designed for parameters used in volumetric estimates. For one of these parameters, namely top reservoir depth, a database already exists. This database, set up by Hoetz *(2016)*, is expanded. If the statistics do allow, further breakdowns are carried out e.g. prediction accuracy per operator.

When the bias is adequately decomposed using statistical analysis, the effect of prediction bias is modelled on portfolio scale. A stochastically generated synthetic drilling portfolio is designed. After specifying the portfolio ranking criteria, the synthetic portfolio is drilled *on paper*. By comparing the portfolio prognosed values and the actuals, the prediction quality is assessed. Based on this, key parameters contributing to prediction bias can be identified.

1.4 Thesis outline

In this thesis first a literature study is presented to better understand the volume prediction bias as presented in literature and to summarize some of the suggested causes. As the research consists of two main components, the statistical look-back analysis and the synthetic portfolio modelling, this thesis is set up in a similar manner. First the statistical look-back analysis is presented with its own methods, results and discussion sections. Next the synthetic portfolio modelling is introduced in a comparable structure. In the end of this thesis, the conclusions section will summarize findings from both research components. The thesis is finalized by giving recommendations for future studies.

2. Background

To put the results of this study in perspective, a comprehensive overview of the prediction bias and its suggested causes as presented in relevant literature is necessary. In this chapter, first an overview of previous pre-drill vs post-drill assessments are presented. Then, some suggested causes that might contribute to the observed prediction bias are summarized.

2.1 Volumetric prediction bias in literature

A comparison should be made with other datasets, to later on put the results of the statistical look-back analysis in perspective. Preferentially one of the North Sea to keep the comparison as solid as possible, with similar lithostratigraphy and geological history of the basin, and of roughly same time frame to consider technical advances that are made over time. This is where it becomes visible that little data on this topic is publicly available. Although not extensively searched for, no studies based on data outside the North Sea could be found. For the North Sea the most extensive studies are produced by the Norwegian Petroleum Directorate (NPD). This is the Norwegian government agency responsible for the regulation of the Norwegian petroleum resources on the Norwegian Continental Shelf (NCS). Their 2018 annual report shows a study on all hydrocarbon targets drilled on the NCS in the period 2007 – 2016 (fig. 1). Concerning gas targets, roughly 47 percent of finds are within, 16 percent above and 37 percent below the uncertainty range of predicted estimates. The companies, according to this study, overestimate resource expectations by an average factor of 2.1 *(NPD, 2018)*.

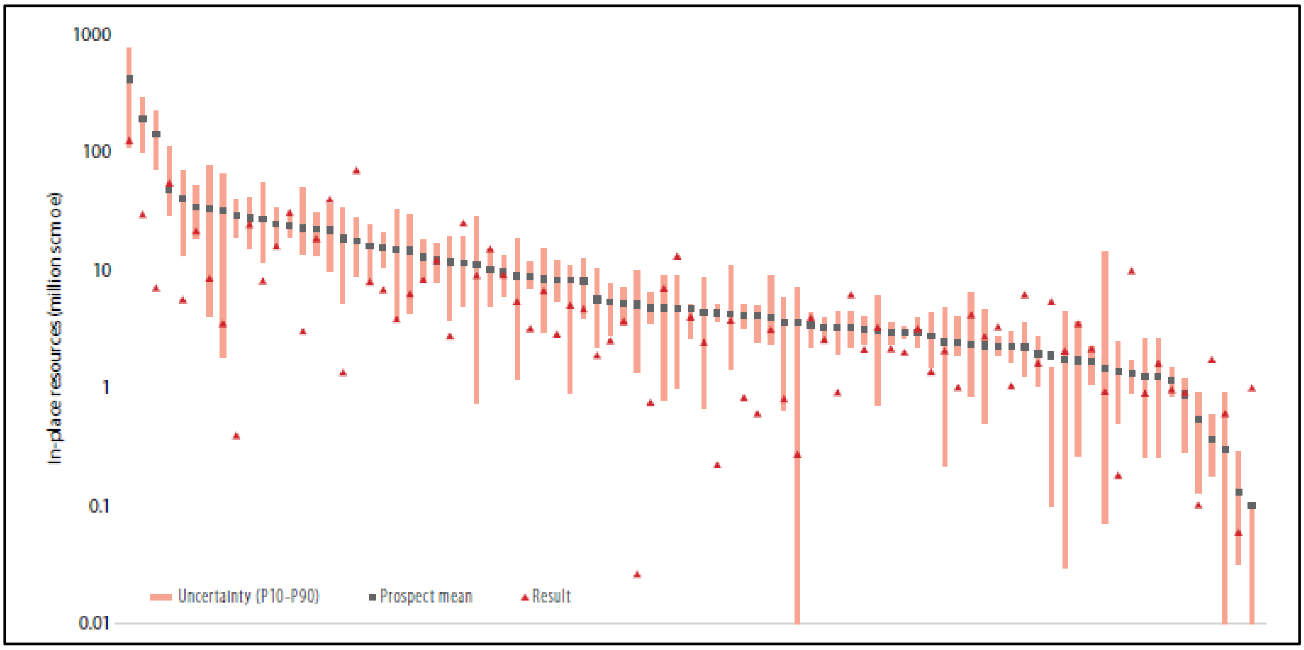

Figure 1.: Company pre-drilling estimates for gas, compared with post-drilling discovery size. Figure is taken from the NDP Resource Report 2018. The red area shows the P10-P90 range. The squares are the expected discovery size pre-drilling, while the triangles represent the estimated discovery size post-drilling. (NPD, 2018)

No comparable study for the British part of the North Sea has been found in the public domain. The only study that is of some interest is an extensive post well-analysis of the Oil and Gas Authority (OGA) on wells in the Moray Firth area. The focus of this study is whether projects were successful and if not, what the reason for failure is *(OGA, 2015; Mathieu, 2016)*. Two other relevant datasets that address the quality of prognosis are Milkov *(2017)* and Rudolph & Goulding *(2017)*.

Milkov *(2017)* presents the results of a dataset comprising of 25 exploration wells drilled by Lundin Petroleum on the NCS during the period 2011 – 2015. Milkov shows that Lundin explorers consistently underestimate the geological Probability of Success (PoS) and significantly overestimate the success case volumes. Lundin's average discovery is approximately 4 times smaller than the average estimated prospect. On portfolio scale just over half of what was promised (Expectation Volumes) was actually found. Milkov makes a strong case about neglect of base rate information related to exploration success and discovery sizes by explorers in their exploration areas. Base rate information meaning historical data on which constraints for new predictions can be based. In principle this aspect of systematic overpromise could be forestalled by EBN for the Dutch situation. As EBN has access to all data and as an investor evaluates all new made predictions, EBN should be able to estimate prognosis corrections if base rate information indicates to.

Rudolph and Goulding *(2017)* present the result of an analysis of Exxon Mobil's conventional wildcat predictions versus results in 44 countries from 1994 to 2015. Interestingly, Rudolph and Goulding on first sight appear to be the only ones (based on the literature used for this study) reporting an underprediction instead of overprediction of volumes. They report that the sum of the pre-drill volume is 27% lower than the post-drill volume. Though closer examination shows that this concerns risked pre-drill volumes. When un-risked pre-drill mean volumes for the successful wells are analysed, the pre-drill volumes are actually 4% greater than the actual post-drill sum. Although still a relatively small bias, this look-back study also shows an overprediction of pre-drill volumes on portfolio scale. Furthermore Brown et al. *(2000)* report that in the period 1987 – 1996, Exxon Mobil discovered worldwide only half the total predicted volume of hydrocarbons.

Many older studies also confirm that explorers are commonly overoptimistic in their predictions of hydrocarbon volumes *(e.g. Rose 1987; Johns et al., 1998; Harper, 2000)*. As these studies are based on data significantly older than presented in this study, no further attention will be given to these analyses. They do show that apparently publicly available look-back studies might use to be more common and that probably the industry appears to have become less willing to give insight in this type of business performance data. Without a doubt further insight would present itself if larger statistical look-back studies would be available. For example, in 2000 a book was published by Ofstad et al. combining the papers presented at the Norwegian Petroleum Society conference *Improving the Exploration Process by Learning from the Past* held in Haugesund in September 1998 *(Ofstad et al., 2000)*. This book touches various aspects of the exploration process with the aim of further developing and improving the process for the future.

Summarizing, few recent studies are available assessing the prediction bias in hydrocarbon volumes, let alone on the underlying parameters. Yet the problem is well known amongst insiders. Although no clear consensus exists on the cause of the bias, some possible contributors are suggested. In the next section, in correspondence with Hoetz et al. *(2020, in review),* these proposed ideas are divided in the following categories: *evaluation tool induced bias* and *cognitive bias*.

2.2 Evaluation tool induced bias

In exploration, as with all geological subsurface work, the "real" geological situation is often unknown. Models are built to reproduce reality as accurate as possible. These models are mainly based on seismic data and surrounding wells. They represent an interpretation based on seismic interpretation and assumptions of what is likely to occur between datapoints *(Lelliot et al., 2009)*. Often soft and hard data are not enough to define the distribution of parameters in the reservoir model. Hence stochastic algorithms are used to provide a measure of uncertainty based on petrophysical parameters and lithofacies. As the uncertainties of each input data used to build the static reservoir model, cannot be expressed in a deterministic realization, a probabilistic model is often the outcome *(Rose, 2006; Binns & Corbett, 2012)*.

As mentioned, the static probabilistic reservoir models, on which the prediction of volumes is based, depend significantly on interpretation of seismic and wells. Whereas estimates for e.g. porosity (PHIE), are often straightforward and based on well data, parameters such as Gross Rock Volume (GRV) and column height are more complex and very specific for each individual prospect. These parameters depend largely on seismic interpretation and can have large uncertainties. Seismic imaging is an imprecise tool even if imaging appears good. An interesting example is given by Quirk and Ruthrauff *(2008):* because of the relatively low vertical resolution in seismic (usually>25m) we often assume that the base of the overlaying seal does equal the top of the reservoir. However, this is not always the case and some waste zone is present. This approach tends to result in overestimating volumes.

Another issue with founding volume predictions on these probabilistic models is the choice of distributions. Quirk and Ruthrauff show in their 2006 paper that three different volumes can be predicted using identical P90 and P10 values in GRV, Net/Gross (N/G) and hydrocarbon column height. Just by doing nothing other than changing from lognormal to stretched beta distributions. Certainly, this affects prediction accuracy. A lot of other issues with the method of predicting hydrocarbon volumes are also presented. Such as that static models do not take into account the three-dimensional aspect of reservoir properties *(Quirk & Ruthrauff, 2008; Binns & Corbett, 2012)*. Another example of bias being introduced by the evaluation method is known from seismic time-depth conversion *(Hoetz, 2016)*. In case the velocity model is too simplistic an important effect might be overlooked: rock velocities tend to increase with increasing depth. A more elaborate velocity parametrization is required to take this into account. Figure 2 illustrates how ignoring this effect can result in systematic overpredicting volumes. So, the conclusion is that prediction errors and prediction bias can be introduced simply by the methodology or evaluation tool by which volumes are prognosed.

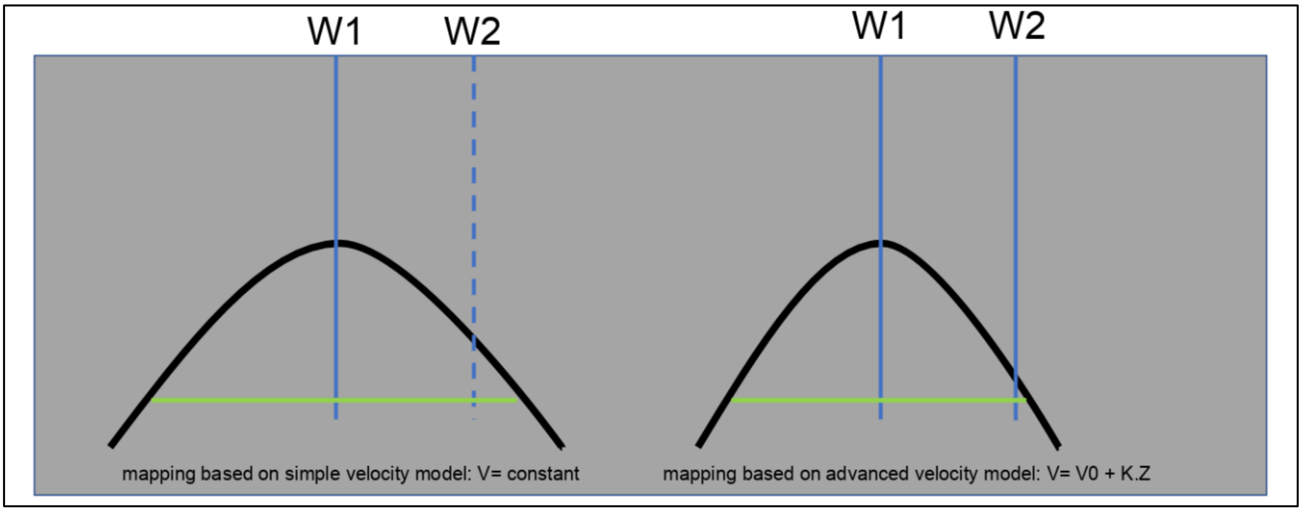

Figure 2.: Example of evaluation tool induced bias resulting in over-optimistic volume predictions from Hoetz (personal correspondence). Early seismic data was often not properly depth converted, e.g. because the velocity model used was too simplistic and ignored the effect of burial compaction. In this example a development well (W2) was planned on a gas bearing anticline that was discovered earlier by W1: drilled at the crest of the structure. Mapping and time-depth conversion of the entire accumulation used a constant velocity calibrated at W1 (fig. left). Subsequently W2 was drilled and found the flank deeper than prognosed. Advances in velocity model building (introducing more sophisticated V0-K velocity functions that honor the effects of burial compaction) shows that the anticline is narrower than mapped initially (fig. right).

In this category we are however limited by the technology being deployed. We can pursue that the tools that we have at our disposal, are used correctly. We should avoid treating the tools as black boxes and recognize potential errors as soon as possible. To do so, we should be aware of improper use of evaluation tools. An example would be to create awareness of proper time-depth conversion methodology.

2.3 Cognitive bias

Given the degree of subjectivity of volume predictions, the influence of cognitive bias is an important consideration. The general definition of a cognitive bias is a predictable and repeatable, unconscious mental error in the processing of information, which can result in illogical judgement and decisions *(Baddeley et al., 2004)*. Causes of bias, specifically on decision making, have been described in general by Kahneman in this best-selling book *Thinking, Fast and Slow (2011)*. Cognitive biases in particular in Earth Sciences are addressed by Baddeley et al. *(2004)*. As the sources and implications of cognitive bias in the exploration process are very widespread *(Baecher, 1988; Baddeley et al., 2004; Binns & Corbett, 2012)*, it would be beyond the scope of this thesis to try and list them all. Instead attention will be focused on three forms of cognitive bias that are perceived to have the largest impact on prediction quality. Namely overconfidence, individual motivational bias and base rate neglect *(Baddeley et al., 2004; Binns & Corbett, 2012; Milkov, 2017)*.

2.3.1. Overconfidence

Overconfidence is a well-established bias in which one's subjective confidence in a judgement is greater than the objective accuracy of the judgement. Kahneman *(2011)* describes a distinction between a swift, intuitive response to a situation and a more thoughtful, analytical approach. This is argued by Binns and Corbett *(2012)* to be particularly applicable to the E&P industry. Rapid, intuitive response to a project proposal based on experience instead of a more slow but considerate response will surely affect prediction performance. According to Myers *(2018)* the potential effects of overconfidence can easily be counteracted by the appropriate use of historical data.

2.3.2. Individual motivational bias

In a sense exploration geoscientist have conflicting roles when generating and reviewing prospects. On one hand, they must accurately evaluate available information and make a prediction as close to reality as possible. Alternatively, they are expected to be creative in generating opportunities and be persuasive in maturing them. This might influence the quality of the predictions. In general terms it is hard to judge which motivational factors might affect prediction quality. Does, for example, the possibility of not getting a prospect drilled impact the assessment of prospect size *(Bond and Carragher, 2018)*? Motivational bias can be under unconscious control but might be conscious too. *(Baddeley et al., 2004)*.

2.3.3. Base rate neglect

Historical base rate information can help set constraints on future predictions. However, people tend to rather focus on specific information at the expense of historic base rate information *(Kahneman and Tversky, 1973; Baddeley et al., 2004)*. This shortcoming in sufficiently weighting a-priory information in reasoning is known as base rate neglect. An example is demonstrated by Milkov *(2017)*: Milkov shows that, based on a dataset of 25 exploration wells from Lundin Petroleum, Lundin explorers disregard information about recent discoveries and instead base volume assessments on individual prospect information. This ultimately contributes to a volume bias.

In the examples of cognitive bias above, the role of the explorationist is described as one of an individual. However, in reality the exploration process is based on group work and how experts collaborate and confer in teams. This generates other, more complex forms of bias associated with group interactions *(Baddeley et al., 2004)*. Also, the mix of biases and to what extend they influence the prediction process varies per company and per individual. Quirk and Ruthrauff *(2008)* state for example: *"Our experience is that volumetric* *assessments of the same pre-dill prospect made by different interpreters and by different companies commonly vary by more than a factor of two."* Quantifying the effect of cognitive bias on prediction performance might therefore be a tough, if not impossible, task. We are not aware of studies assessing the amount of cognitive biases being present in predictions from geoscientists. It might be an area of fruitful and useful future research. Until then the effect of cognitive bias should be restrained as much as possible by consistency in the prediction process. For example, via the use of historical data *(Myers, 2018)* or workflows *(Milkov, 2015)*. After all, subjective judgements are not necessarily problematic as long as they are derived in a consistent manner and can thus be accounted for *(Cox, 1946)*. Another useful advice for mitigating cognitive bias is introducing thorough technical project challenge (e.g. peer reviews) by others who have different motivations.

3. Statistical look-back analysis

To quantify the volume bias and to pinpoint key parameters that contribute to it, a statistical look-back study based on historic data is performed. In this study the quality of the prediction of the subsurface parameters being used in volumetric assessments is analysed. This chapter first describes the method of the statistical analysis. Subsequently the results of this research are presented and discussed.

3.1 Methodology

The statistical look-back analysis is based on EBN data. As stated in the introduction, EBN participates as a non-operating partner in virtually all E&P projects in the Netherlands and therefore has access to all data regarding the projects. The pre-drill prognosed values of reservoir characteristics are generally supplied by the operator as part of the proposal for the project. Once executed, the post-drill measurements (actuals) are also provided. Of the various parameters, the reservoir depth prognosis can easily be checked from the well logs. PHIE, water saturation (Sw), GWC and N/G require some additional petrophysical analysis but can also be fairly well constrained after inspection of the well log. Gas Initially In Place (GIIP) and Recoverable Volumes (RV) are estimates based on well tests and/or updated static models. Analysing prediction quality of specific reservoir parameters allows to investigate which static model input parameters are dominating the observed prediction bias in gas volumes. All the supplied data is stored by EBN in their proprietary well database Basis Registratie Boringen (BRB). To ensure full auditability, all compiled datasets used in the study are based on data from the BRB. Parameters assessed in this study are RV (recoverable volume), GIIP, top reservoir depth, GWC, PHIE, Sw, Net rock volume (NRV), GRV, N/G and pressure. Note that all these parameters are separate entries (data fields) in the BRB. For example, N/G data used in this study is not derived from NRV and GRV data from the database. Rather N/G data is stored separately in the BRB and values are directly taken from the BRB. Results presented in this chapter are based on a download of the BRB from 06-02-2020.

As data being used in evaluations is typically incomplete and imperfect, it is likely that a prediction is estimated higher or lower compared to the “truth”. Hence there is little point in assessing individual predictions on their quality. On portfolio scale though, the cumulative predictions should approach the measured actual. A statistical approach looking at a significantly large set of wells is therefore taken to assess prediction quality and the overall portfolio performance. The focus of this study lays therefore on acquiring results with as much statistical relevance as possible via large sample sizes. Most results that are shown in this report are therefore based on a dataset compiled of both exploration and development wells. Furthermore, datapoints are only from drilling projects with a hydrocarbon objective from the period 2004 until 2019. Datapoints from before 2004 are excluded due to incompleteness. Also, technical failures are removed from the compiled datasets whilst dry holes (regular outcomes of exploration wells) are kept in as data entries. The reason for that is as follows: in the case of a technical failure the prognosis could, due to circumstances not be checked with a reliable actual measurement. Hence this datapoint is considered inconclusive. In the case of dry holes on the other hand, both prognosis and actual are available and hence constitute valid datapoints.

Where required, further details regarding the various datasets are given in the results section where they are presented. When sample size (aka the number of samples in a data set) does allow, further in-depth analyses are performed. For example, top reservoir depth has a large sample (>300), so further detailed analysis such as bias per operator are possible. However, these detailed analyses are not documented in this thesis for confidentiality reasons. As all operators deliver data in a different manner and with a different format, some wells/data entries in the EBN database have more information than others. This reflects in various sample sizes of the different datasets used for the statistical analysis. The fact of the matter is that within the EBN database some entries have complete pre-drill and post-drill data on volumes and all underlying parameters, while others only have incomplete data (e.g. no prognosed porosity).

Data entries in the various datasets are checked by using the source documents e.g. well proposals or post-drilling well summaries. In particular outliers, i.e. very large prognosis errors, have been quality controlled. Also, some random checks, around 10% of the dataset, were conducted to get a feel for the accuracy of the EBN databases. These source documents can be found in the EBN online archive as well as the online NLOG database. NLOG is a database managed by the Geological Survey of the Netherlands and contains all subsurface data that is made publicly available under Dutch mining legislation. Where relevant, data outliers and data discrepancies were reported and the EBN database, and the dataset used in this research, was updated.

All datasets are compiled in Microsoft Excel® and analysed using Excel's built-in tools for statistical analysis. For each parameter the error is calculated by subtracting the actual from the prognosis for each well. The mean of the prediction errors is a measure for the prediction bias for that particular parameter. To quantify the spread of prediction errors the standard deviation (STD) is determined. A relative bias (bias %) is calculated by dividing the bias by the mean of the actual. In addition, as the used datasets only represent a subset of the total population (as stated: the EBN database is not complete) a confidence interval (CI.) is calculated. In this way it is possible to determine whether an observed bias is statistically significant. Example: from the period being reviewed (2004 – 2019), 643 data entries (wells) are present in the EBN database and these represent virtually all hydrocarbon wells drilled in the Netherlands for that period. Of these, only 328 have complete pre-drill and post-drill top reservoir depth data. As the sample of data is thus just over half of the entire population present, a CI. is necessary to estimate within which range bias parameters are for the full population. Several statistical tools were tested for estimating CI values: Tibco Spotfire®, R® and Excel. the latter was found to be most practical and hence used in the further analyses. These duplicate test analyses are not included in this report.

To further substantiate the statistical analysis, paired t-tests are performed on the datasets after Heggland et al. *(2000)*. Heggland et al. have conducted a similar study in which they compared post-drill hydrocarbon volumes and volumetric parameters with their respective pre-drill predictions. This research was based on data from the NCS. A t-test is a statistical tool used to determine if there is a significant difference between the means of two datasets. As there is always the factor of statistical random noise, small differences between prognosed values and their actual measurements can occur without a bias in the ability to prognose. The t-test basically accounts for this statistical noise and checks whether any potential differences between a prognosis and actual can be attributed solely to the noise, or that other factors are involved. As prognosis and actual are inherently related, a paired t-test is used. This type of correlated t-test applies to datasets of matched pairs of similar units. As the actual measurement can come in either higher or lower than prognosis, a so called *two tailed* paired t-test is performed on all datasets. The null hypothesis for the tests is that there would be no difference between the prognosis and the actuals other than random noise. This is following Heggland et al. and is also the custom null hypothesis used in statistical studies. Furthermore, the t-tests were executed with a significance level (alpha) of 0.05. Meaning that a risk of 5% was taken in concluding that a difference exists when there is no actual difference between the two populations. A 5% chance of incorrectly rejecting the null hypothesis was thus deemed acceptable. The significance level of 0.05 was chosen as this is standard in statistics.

The t-test produces two results that are of interest to determine whether to reject the null hypothesis or not: the p-value and the t-statistic. The p-value is the probability of obtaining an effect (in this case a prognosis error) at least as extreme as in the sample data, assuming the null hypothesis is correct. When a p-value is thus less than or equal to the significance level, in this case 0.05, you can reject the null hypothesis. In most cases though, the p-value will be orders of magnitude smaller than the alpha value when a bias is present. In addition to this test of the null hypothesis, the t-test produces a t-statistic. The t-statistic is a ratio of the differences between two datasets. The larger the t-statistic, the more difference there is between datasets and vice versa. For example, a t-statistic of 3 means that the datasets are three times as different from each other. The t-statistic can be compared with the t-critical value, the value that a t-score must exceed for the null hypothesis to be rejected. As a two-tailed t-test is taken, the absolute value of the t-statistic is taken.

Lastly, the EBN database that is used contains more information than just pre-drill and post-drill parameters. Specifically, the post-drill values are the first, often pre-liminary, measurements after the well has been drilled. Often these values are updated over the project duration as more measurements are done over time. We can reasonably assume that new measurements are a closer approximation to reality. To indicate how close the first measurements used in this study are to further updates and thus how stable the calculated bias is with respect to later updates of the actuals. Hence an additional analysis is performed assessing the frequency and magnitude of volumetric updates. Before, during and after the drilling project there is regular communication between the operator and EBN. In spring and fall Technical Committee Meetings (TCM) and Operating Committee Meetings (OCM) are organized. In these meetings the activities of the past year are discussed and evaluated by the operators. EBN is informed about any updated measurements and these updates are then stored in the national hydrocarbon resource database IPResource (IPRes). This database has the purpose of monitoring and prognosing production rates and volumes. To assess the frequency and magnitude of volumetric updates, information from this database is used to calculate the percentual change of volumes per year. Projects with complete GIIP data over the timespan 2009 – 2018 are selected. Updates where the volumes were adjusted with more than 100% percent are regarded outliers, for example due to erroneous measurements, and are not taken into account.

3.2 Results

3.2.1. Volumes

Figure 3 shows the pre-drill Recoverable Volume (RV) and its actual for each project in the compiled RV dataset. Recoverable Volume (RV) = Expectation Volume (EXP = Means Success Volume * Probability of Success) * Recovery Factor (RF). Only wells targeting gas have been selected to avoid unnecessary and complicated volume conversions. Also, the major part of the Dutch E&P projects target gas, whilst only 13 wells in the dataset targeted oil. Of the 215 gas wells, 96 wells have a RV result below the low-case estimate. This category includes 54 dry holes. Another 53 wells have an actual RV between the mid-case and the low case. 34 wells delivered on prognosis. (note that often no new RV values are being calculated when measurement do not differ much from prognosis). 32 wells delivered better than the base case. This means 69% of the wells in this dataset fell short of delivering the mid-case expectation volume. For this dataset only 58% of the prognosed (risked) volumes are found. This implies a 42% volume prediction bias.

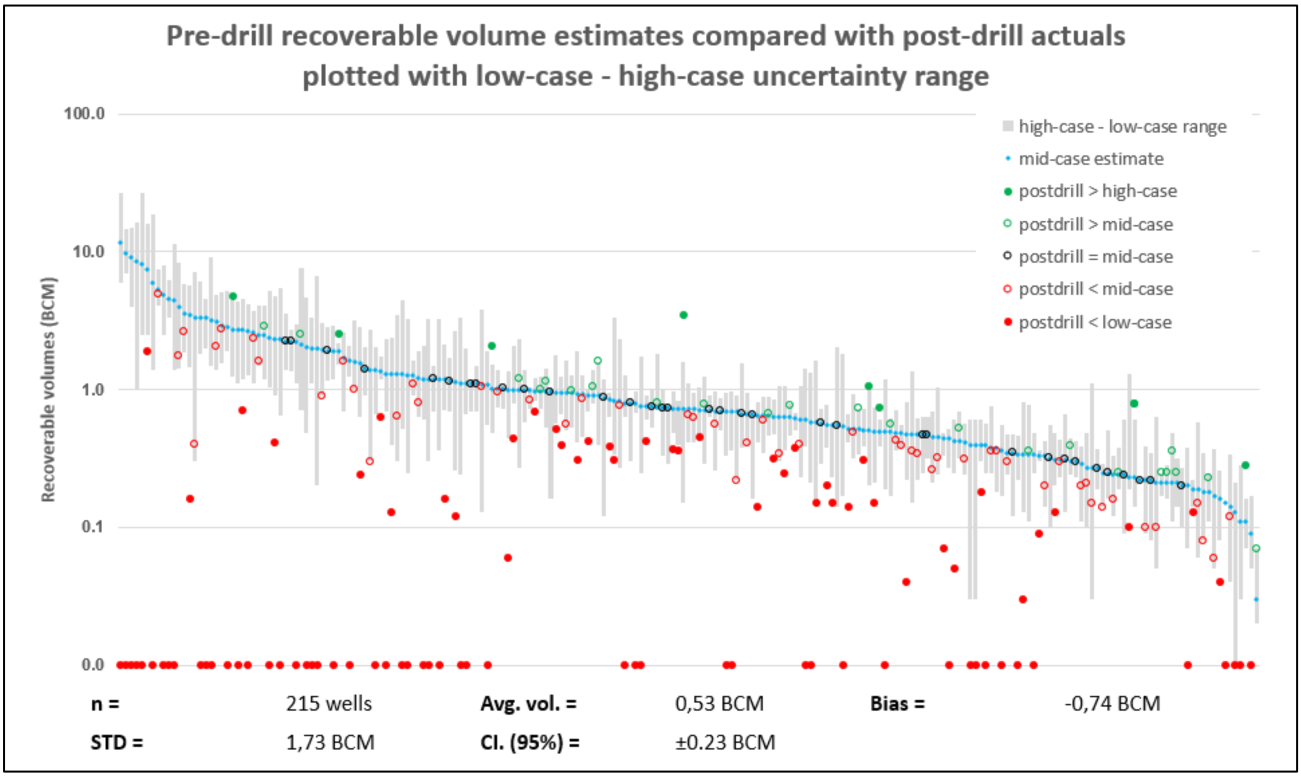

Figure 3.: RV prediction vs actual diagram. For each drilling project a mid-case recoverable volume estimate is plotted (blue rhombus) with a low-case to high-case uncertainty range (grey bar). The spheres represent the actual volume measurement after drilling. Dry holes are plotted at the 0.0 RV axis. Projects are sorted on mid-case estimate size. Note the logarithmic scale.

3.2.2 Depth prognosis

Top reservoir depth is an important factor in determining the hydrocarbon column height, the position of possible spill points and the GRV. Often, in case the actual reservoir depth as encountered in the well, turns out deeper than prognosed, the volumes have to be adjusted downwards. Figure 4 presents the top reservoir depth prognosis errors. 321 Wells within the timeframe of this study are found in the EBN database that contain the pre-drill prognosed top reservoir depth and the post-drill measured actual. For this analysis and all the following ones, both gas target wells and oil target wells are considered, in contrast to the volume dataset. This was done to strive for as much statistical relevance as possible. The depth errors of these wells are plotted with the differentiation per well type: exploration, appraisal and production. The maximum underestimation is -343 meters and the maximum overprediction 225 meters. Based on visual inspection of the graph it would appear that the largest errors are generally referring to exploration wells. Overall 61% of the wells show depth error resulting in an overestimated volume. This bias towards over-optimism is also represented in the fact that the chart is lob sided to the left.

The argument can easily be made that it is unrealistic to expect from geoscientists that they predict top reservoir depth (and GWC for that matter) exactly on point. Depth prognoses are based on seismic and well data containing noise plus assumptions. Unfortunately, no information regarding predrill depth prediction uncertainty is available in the EBN database. Nevertheless, an attempt has been made to put these depth errors into context. A rough rule of thumb says that estimating depth with a depth accuracy of up to 1% percent is typical and reasonable *(Hoetz, personal correspondence)*. In figure 5 the depth errors can be referenced with respect to 1% of the target depth. Also, this figure represents a subset of the depth error dataset. Here only exploration wells have been selected. Often for well reviews different well types are kept separate in order to reduce the risk of comparing apples and oranges. In this graph significant prognosis errors can be observed outside the 1% uncertainty range. Furthermore, the graph is also left skewed, implying bias to overestimating volumes in exploration wells.

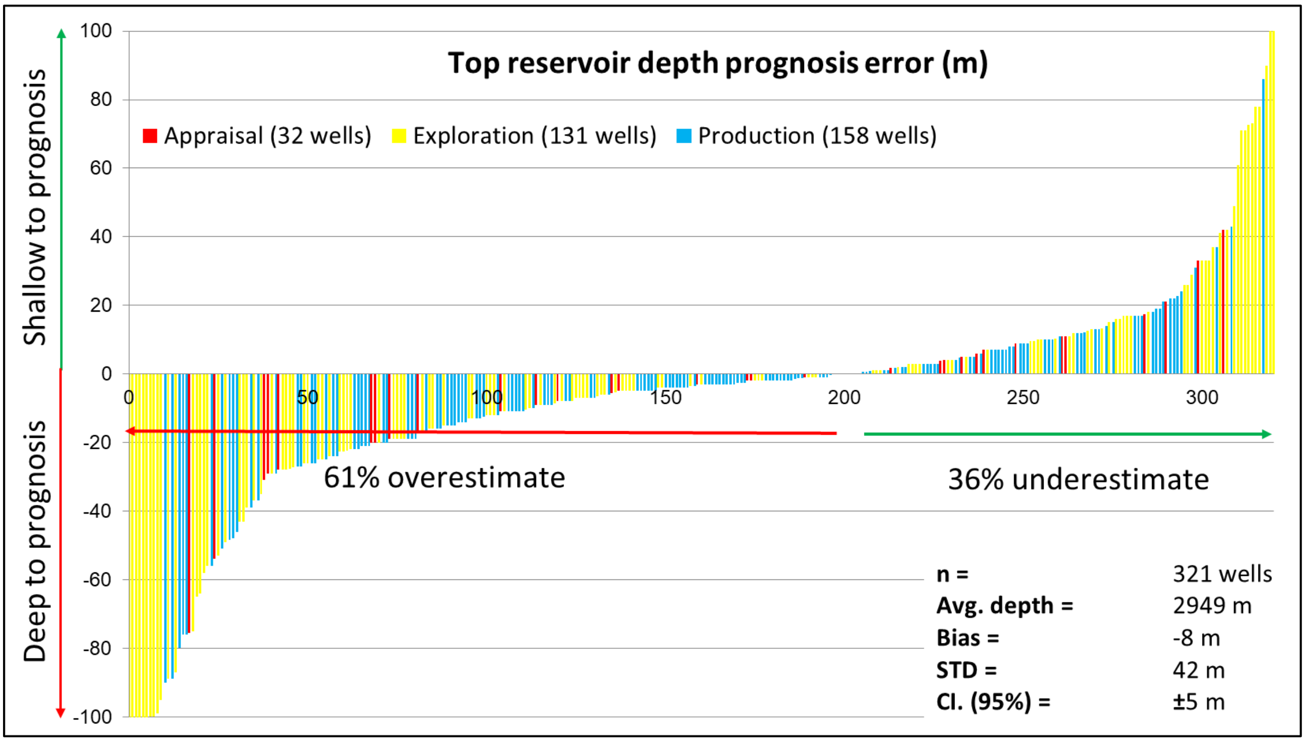

Figure 4.: Top reservoir depth prognosis error (m) plot. Prognosis errors are ranked from deep to prognosis to shallow to prognosis with a distinction per well type. The displayed error is with regard to the mid-case prognosis. The difference in percentage between over- and underestimates indicates the presence of bias in the prognoses. The bias (-8 m) is larger than the CI (+/- 5m) indicating that the bias is statistically significant.

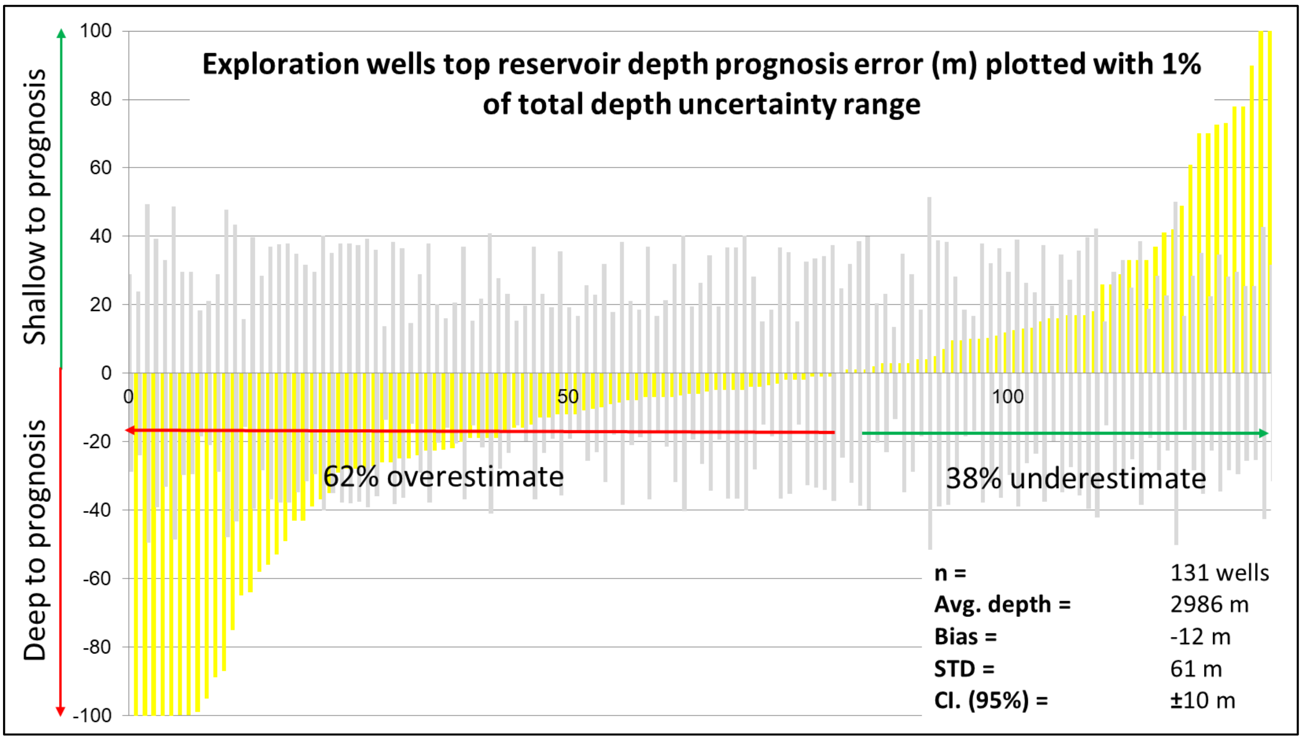

Figure 5.: Top reservoir depth prognosis error (m) plot of only exploration wells. An uncertainty range of 1% of the total depth is added around each project.

3.2.3 Contacts

For assessing the GWC depth prognosis quality, significantly less data is present. Therefore, in this (and subsequent) analysis no differentiation is made between well type. The sample size can still be considered relatively large compared to similar studies from elsewhere. In contrast to the top reservoir depth, where *shallow to prognosis* can be considered a positive surprise as it generally results in a larger hydrocarbon column, a GWC that comes in *shallow to prognosis* would indicate a smaller than predicted hydrocarbon column. In figure 6 the GWC prognosis error data is plotted. Also here a similar tendency to over-optimism and overestimations in volume is observed.

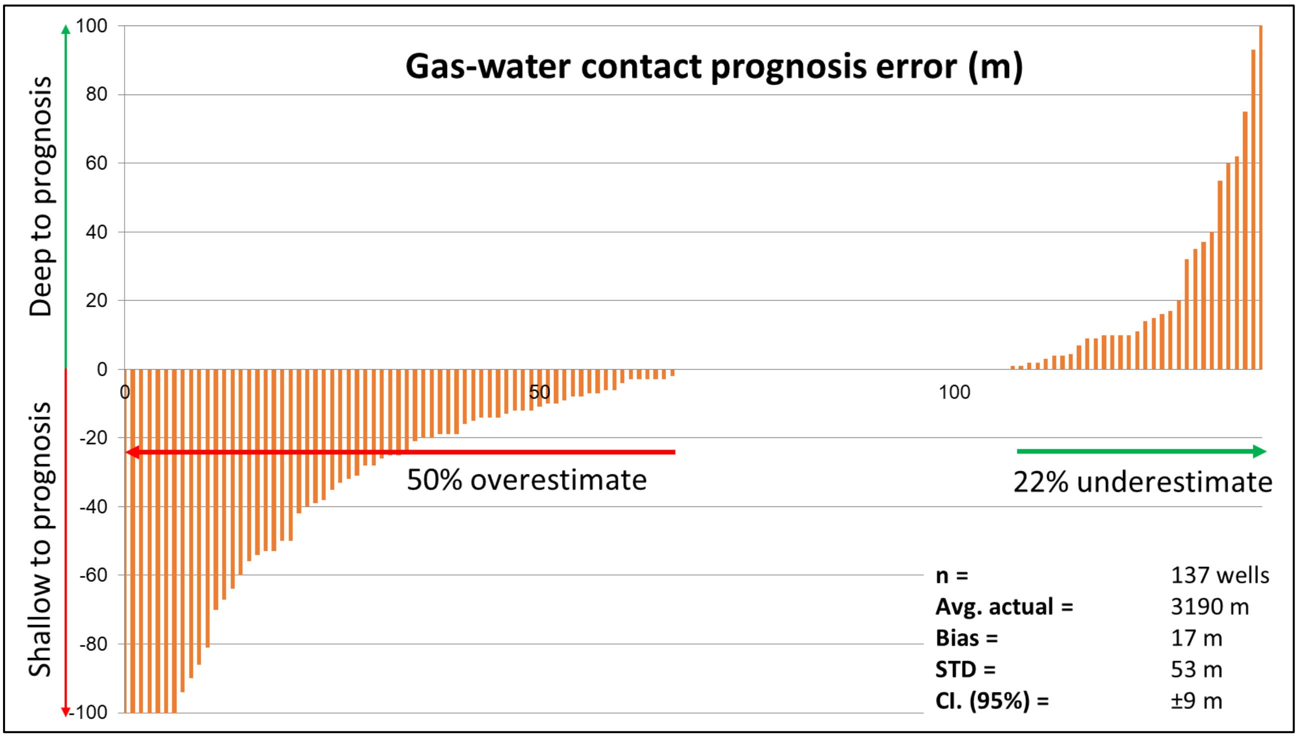

Figure 6.: GWC prognosis error (m) plot. Prognosis errors are ranked from deep to prognosis to shallow to prognosis. The displayed error is with regard to the mid-case prognosis.

Combining the top reservoir depth and the GWC yields the column height. This data cannot be found in the EBN database directly. Figure 7 presents the column height prognosis errors. For this figure wells with both prognosed and actual top reservoir depth and GWC are selected. By simply subtracting the top reservoir depth from the GWC depth, the prognosed and actual column heights are determined. Clearly a tendency towards over-optimism and potential overestimation of volumes observed in the top reservoir depth and GWC depth results, translates to the column height optimism. Based on this parameter 76% of the 135 projects do result in an overestimation of volumes.

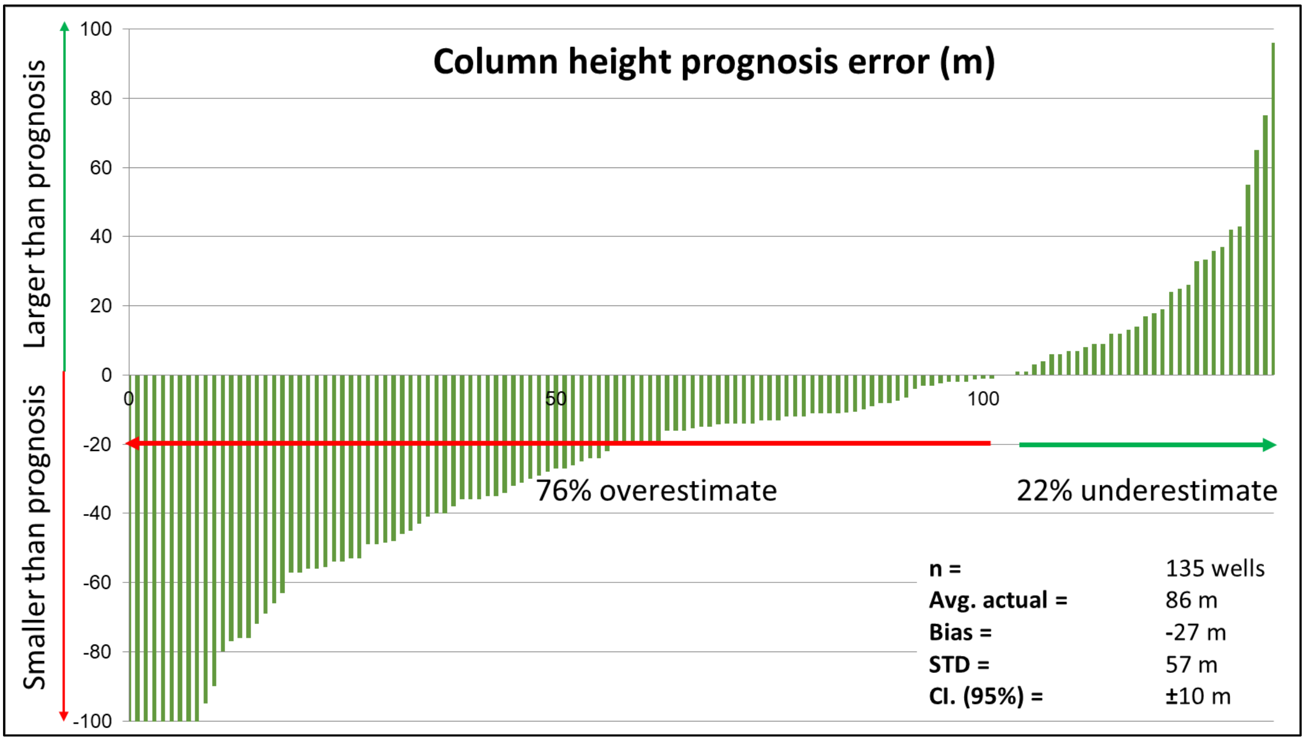

Figure 7.: Column height prognosis error (m) plot. Prognosis errors are ranked from smaller than prognosis to larger than prognosis. The displayed error is with regard to the mid-case prognosis.

3.2.4 Rock properties, pressures and GRV

Figure 8 displays the results of the other tested volumetric parameters. The porosity data indicates an (modest) tendency towards overestimating. Porosity (PHIE) is given in Porosity Units (PU) with 1 PU implying 1% of the rock volume being porosity. High porous rocks can contain more hydrocarbons, hence overestimating porosity means overestimating volumes. With 0.3 on 14 PU (i.e. a relative bias of 2%). The porosity bias is modest and appears, considering the CI. of 0.4 PU, statistically not significant.

In case the actual Sw is larger than the prognosed value, the impact on hydrocarbon volumes is negative. The Sw error graph is clearly lob sided to the left, indicating a tendency to overestimation. The absolute 10% Sw bias translates for a 21% relative Sw bias (i.e. with respect to the actual measured value). (Do note the reversed y-axis for consistency with respect to the impact on volumes).

GRV and NRV directly translate to reservoir size so a larger number means larger volumes. Both show the same pattern of over-optimism (fig 8.C., 8.D.). A strong tendency to parameter overestimation with 68% of the wells overestimated with respect to the mid-case prognosis for GRV and 65% for NRV. The statics for both errors show comparable values for the bias, STD and CI.

Alternatively, the N/G errors appear to be bias free (fig. 8.E.). Over and under-prediction balance each other quite well. Although the bias would indicate a minor tendency to overestimation, the amount is not significant given the CI.

Reservoir pressure is also a parameter in the static models being used for volume prediction. A higher pressure is favourable as it results, via the gas expansion factor, in larger recoverable volumes. In this pressure dataset, the relative bias is small (~3%). Fig 8.F. shows a lob sided graph and it illustrates the fraction of the wells (58%) which overestimated the gas pressure. Table 1 summarizes the relative percentages of overprediction for all tested parameters.

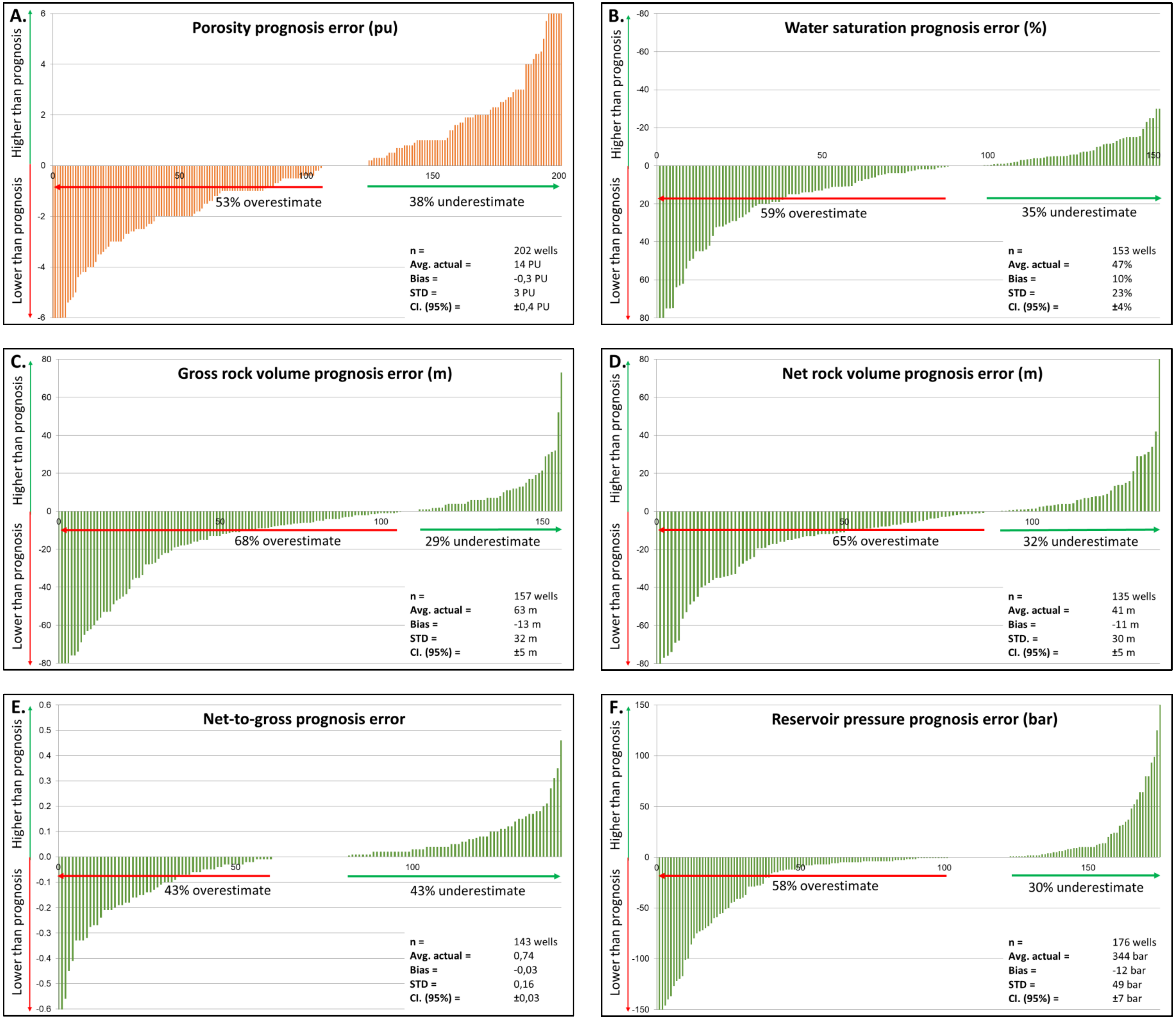

Figure 8.: Prognosis error plots for other subsurface parameters assessed. A.) PHIE (pu), B.) Sw (%), C.) GRV (m), D.) Net (m), E.) N/G, F.) pressure (bar). Prognosis errors are ranked from lower than prognosis to higher than prognosis. The error magnitude is with regard to the mid-case prognosis.

Table 1.: The factor of overprediction in percentages for all tested volumetric parameters. The percentage of overprediction is calculated by dividing the bias by the mean actual. Since top reservoir depth and GWC have such large values for their mean actuals (~3000 m), their respective overprediction percentages are quite small.

Parameter	***Overprediction***
Top reservoir depth	<1%
GWC	<1%
Column height	31%
Sw	21%
Phie	2%
GRV	18%
NRV	26%
N/G	2%
Press	4%

3.2.5. t-test

As aforementioned the look-back study is supported further by the use of a paired t-test. The results of this t-test are summarized in table 2. The null hypothesis of the test is that no systematic differences are present between the actual and the prognosis with a set significance level of 0.05 (i.e. 95% confidence). The null hypothesis is rejected for a parameter when the absolute value of the t-statistic is larger than the t-critical value, and when the p-value is larger than 0.05. Column height is not included as a tested parameter since this was not data directly extracted from the BRB, but rather derived from top reservoir depth and GWC. Furthermore, noteworthy is that all t-critical values are approximately the same value. This is because the t-critical is a constant based on confidence level and sample size. The t-critical is always in the range of 1.965 to 1.984 for samples sizes between 100 and 500 with a confidence level of 95%.

Table 2.: Two-tailed paired t-test results. Null hypothesis is no difference between prognosis and actual with significance level (alpha) of 0.05.

	RV	***Top reservoir depth***	***GWC***	***PHIE***	***Sw***	***NRV***	***GRV***	***N/G***	***Pressure***
Samples	215	321	137	202	153	135	157	143	176
t Stat	6.26	-3.46	3.81	1.23	-5.38	4.23	5.05	2.13	3.22
P-value	2.01E-9	6.02E-4	2.09E-4	0.221	2.74E-7	4.25E-5	1.19E-6	0.0345	0.00151
t Crit.	1.97	1.97	1.98	1.97	1.98	1.98	1.98	1.98	1.97

Except Phie, all parameters fail the t-test meaning that the null-hypothesis of no difference between the population of prognosed values and actual values is rejected. There is thus a large enough difference between the two populations that it cannot be attributed to statistical noise/randomness. As evident from the table the PHIE data do not reject the null hypothesis and thus a potential bias is probably absent. This is in line with the statistical analysis using CI. presented above. The minor difference between the prognosis and the actual might therefore be attributed to statistical randomness/noise. Do note that this t-test is based on a 5%

significance level (which is standard in the industry), a lower significance level most likely would put more weight on the small difference between prognosed values and actuals and thereby perhaps have a different outcome of the t-test. The N/G does just pass the t-test. Although its p-value is smaller than 0.05 and its t-statistic is smaller than the critical t-value, when a significant difference exists between the populations the p-value is usually orders of magnitude smaller than the significance level of 0.05 (as can be seen from the other tested parameters).

3.2.6. Volume updates over project life

Above error statistics are all based on the prognosis and the actual as measured straight after drilling. As reservoir parameters do get updated over project life, an assessment of the magnitude of these updates is interesting too. Figure 9 presents the volume updates over project life. For the timespan 2009 to 2018, 321 projects are selected with complete GIIP values. GIIP is chosen to filter out technical factors and to focus solely on the total gas volume present. The total size of this portfolio selection fluctuates between approximately 4445 and 4525 BCM while no new projects are added or removed. On average GIIP's are yearly updated with between -7 and 4%. Over the whole timeframe individual project GIIP's are on average updated with -3% from the first post-drill actual.

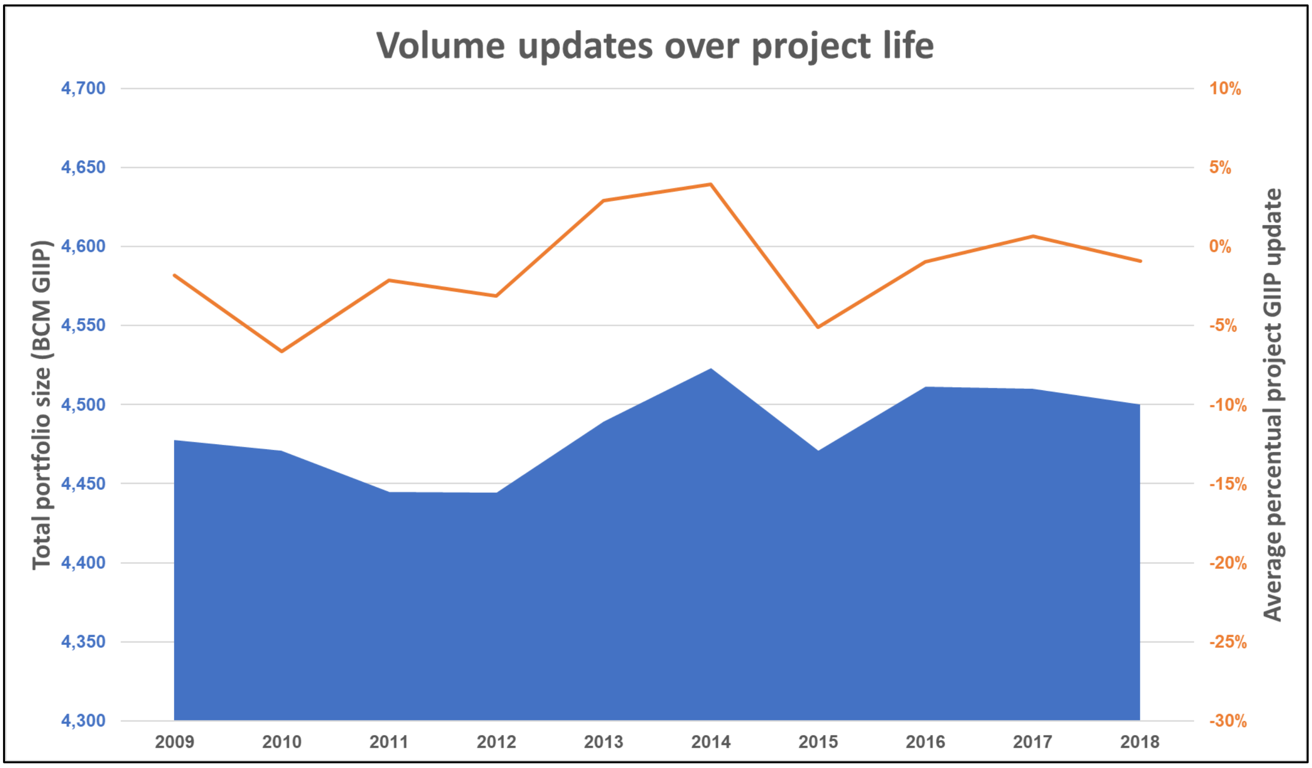

Figure 9.: Volume updates over project life plot. The blue area represents the total size of the selected dataset in BCM GIIP. The orange line displays the average project GIIP update (%). Over time, the size of the portfolio fluctuates although no projects are added or removed. This is due to volumetric updates of individual projects in the portfolio.

3.2.7. Prediction accuracy over time

An interesting question is whether prediction accuracy has improved over time given advances in technology. Figure 10 and 11 show respectively the RV and the top reservoir depth prognosis errors sorted per year. Using a moving average, overall trends in prediction errors are assessed. Based on the size of the datasets for the RV a moving average window of 50 wells was chosen and for the top reservoir depth a moving average window of 100 wells. Although both curves fluctuate a lot due to large spread in prognosis errors, over the time interval selected in this study, RV prediction as well as depth prediction appear to have improved slightly.

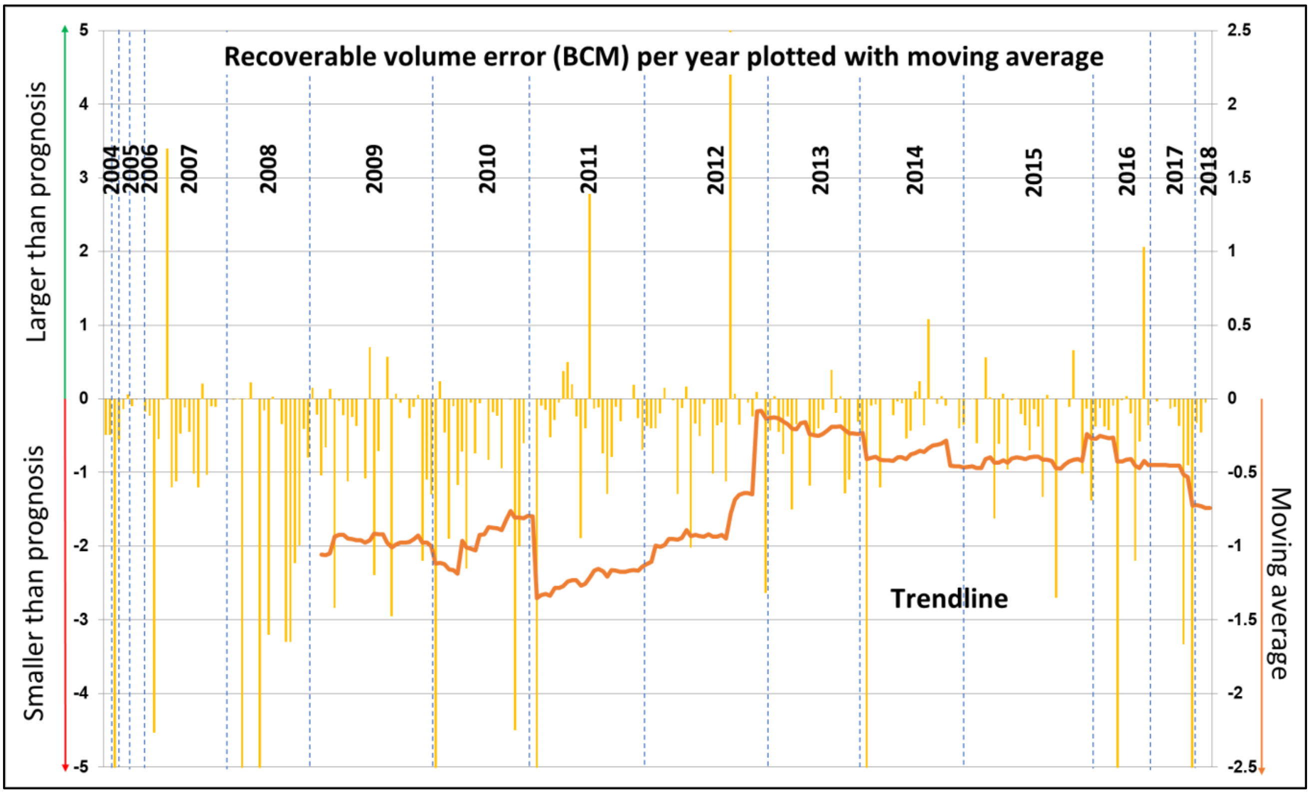

Figure 10.: Recoverable volume prognosis error (m) plot ranked per year with a moving average. The moving average window is 50 wells.

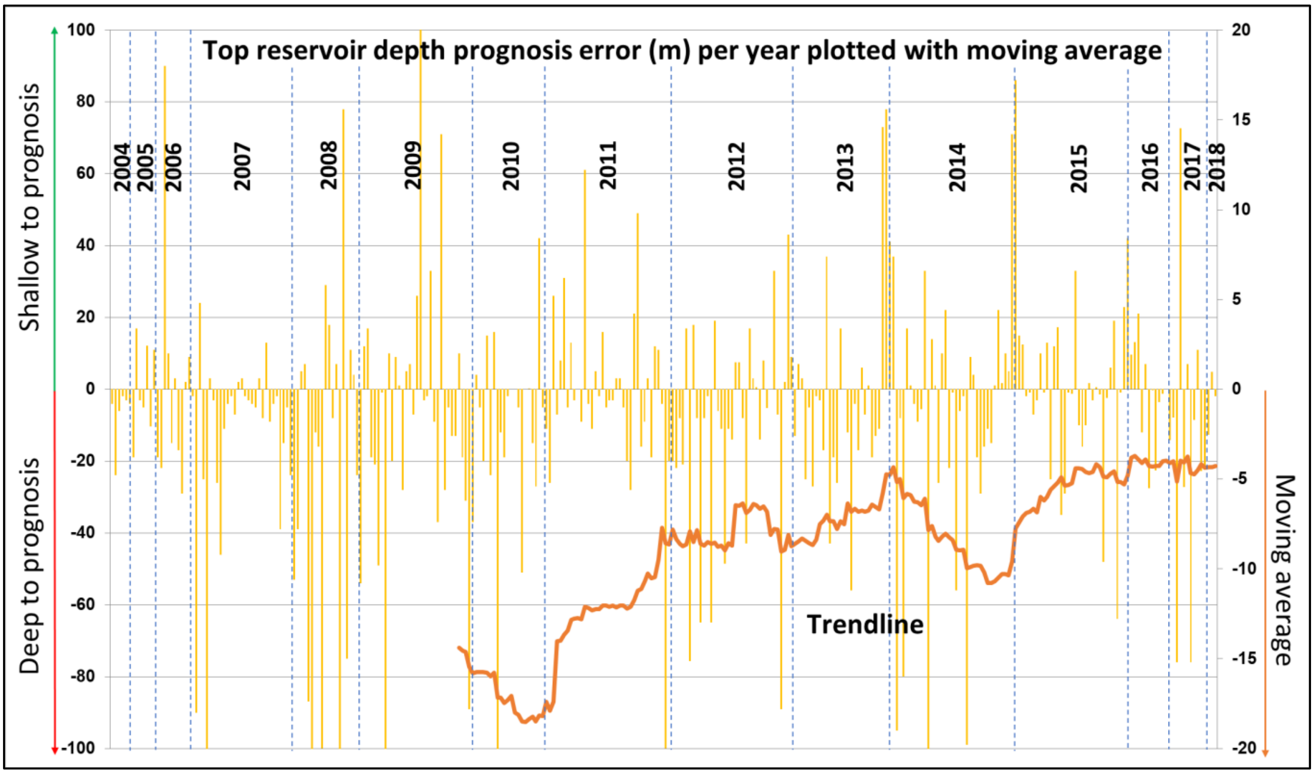

Figure 11.: Top reservoir depth prognosis error (m) plot ranked per year with a cumulative moving average. The moving average window is 100 wells.

3.3 Discussion

3.3.1. Volumes, depth prognosis and contacts

Estimating the actual presence of hydrocarbon volumes and the volume of hydrocarbons present can be considered two fundamentally different things. Also, the dry holes are likely to largely influence the volume error statistics. Hence some analysts exclude dry holes cases from the aggregate expectation volumes for comparison with pre-drill expectations. However, in this comparison all risked expectation volumes are taken into account as these risked volumes are also part of EBN gas production forecasts.

It should also be noted that the terminology used in the RV figure is low-, mid- and high-case instead of the more industry standard P10, P50 and P90. Reason for this is that the operators are not uniform in reporting volume ranges. For example, there are operators that rather use (or used) the P15, P50 and P85 estimates for reporting. Hence for practical purpose EBN uses High Mid Low estimates and these can be equalled to P10, P50 and P90 for most cases. More precise results might be obtained if the EBN database would support better quantified (and standardized!) low and high case estimates. This might be particularly relevant when assessing the prediction performance of individual operators.

Likewise, more insight might be obtained by adding more parameters in the prediction quality database. E.g. differentiating in stratigraphic targets could be helpful in understanding where the key challenges are in subsurface parameter prediction. In the volume plots presented here both appraisal, exploration and development wells with various differencing factors like onshore/offshore and diverse lithostratigraphic targets are included. This occurs because the focus of this study lays on quantifying the prediction bias and looking for contributing factors. More solid results should be obtained by comparing the same type of wells, such as for instance offshore exploration wells with all a post-salt target. A nice example is the EBN based study of Janssen *(2019)* in which she compares 126 exploration wells. Janssen reports that on portfolio scale only 69% of expectation volumes were found, thus concluding a bias of 31%. This differs from the 42% bias observed in this study. However, the difference in bias percentages can partly be explained by the fact that the sizes of the used datasets and thus bias percentages are different, since the bias percentages are relative to the mean actuals of the used datasets. Moreover, the volumes in Janssen's study are based on a project specific EXP GIIP multiplied by a generic Recovery Factor of 75%. Project RV's used in this study have both project specific EXP's and project specific RF's incorporated in the data from the BRB.

As mentioned in the Background (2.1): it would be useful to compare prediction performance from this study with other (foreign) datasets analysed elsewhere. The best study for comparison is the one by the NPD *(NPD, 2018)*. In this study the time frame is roughly the same (2007 – 2016) and prediction quality is checked in a similar manner as presented here (fig. 1). The NPD reports that roughly 47% of finds are within the uncertainty range with an average overestimation factor of 2.1. In the RV analysis presented here this percentage is 36% with an overestimation factor of 2.4. Do note this regards percentage of finds within uncertainty range, in contrast to the bias in total volumes found which is 42% as mentioned earlier. Although there are differences between the numbers, they are of roughly the same magnitude, even more so when considering that the NPD does not take into account dry holes in its calculations and considering the larger timeframe and sample size of this study. If dry holes would not be taken into account in this study, the percentage of finds within uncertainty range would rise to 48%, being close to NPD assessment (47%). The results from the NPD and those of this study are consistent and provides a clear message regarding prediction quality in the North Sea: volumes are generally over-estimated by a factor of two and almost half of the finds fall outside the uncertainty range accompanying the volume prognosis.

Regarding top reservoir depth, it is no surprise that exploration wells do account for the largest depth errors. Appraisal and production wells are typically drilled near existing wells constraining the mapping and depth estimates much better. A similar reasoning can explain why a significant portion of the wells were able to predict the GWC depth fairly well. Production wells target reservoir areas that have been penetrated before and contacts are pinpointed already reasonably well. Also, most of the time the wells are drilled slightly down-dip of the structure. For example, to avoid drilling near to a fault. Obviously, this reflects also in the accuracy of the top reservoir depth prognosis, which shows the importance of adding uncertainty ranges. By drilling the well in a slightly different place than where the prognosis is based on, small errors arise. These add to the bias but shouldn't be taken into account as they do not reflect on the ability to prognose correctly.

Another interesting addition is to investigate how much the different prognosis errors are correlated. For example, the top reservoir depth error with the contact depth error. If both parameters are in error by the same amount, the resulting impact on volumes is null.

3.3.2. Rock properties, pressures and GRV

It is stated in Brown et al. *(2000)* that the GRV is the most important parameter in determining reservoir volumes. This is affirmed by Heggland et al. *(2000)* who show, using linear regression modelling of hydrocarbon pore space volume differences, that variability in the GRV prediction errors is the most important factor when explaining the differences in pre-drill and post-drill volumes. Sw is by far the most important petrophysical contributor according to this study. Interestingly GRV is the only parameter that cannot be directly measured after drill. GRV remains dependent on top reservoir depth and GWC depth and is influenced by a large number of uncertainties. For example, velocity modelling, depth conversion, jump correlation across faults, horizon correlation and geological interpretations. All are largely subjected to interpretation and often poorly understood. Ongoing research to gain further insight in these parameters is therefore necessary.

PHIE seems, based on this study, less important as it shows only a small (relative) bias. PHIE predictions are correct on average and errors that do occur can be attributed to statistical randomness. The possible reason is that PHIE estimates can often be successfully taken from analogue wells testing the same lithostratigraphy and the same depth. Noticing that drilling targets in the Dutch subsurface are dominated by two lithostratigraphic units, the Bunter and the Rotliegend, and most targets are at a depth of around 3 km, it is not surprising that enough data of this is present to accurately predict PHIE.

N/G prediction errors seem to contribute little to the volume prediction bias. Heggland et al. *(2000)* state that there is therefore no need to better understand this parameter. Based on the results in this research, we would nonetheless argue that further insight is desirable as there are significant errors observed in the prediction of the GRV which probably translates via the N/G to the NRV (NRV=N/G*GRV) and ultimately the volume prediction. Considering the triangular relationship between GRV, NRV and N/G, the somewhat similar observed bias in NRV with respect to GRV is to be expected. After all, if one parameter is multiplied by a factor, the related parameters will so too. Also, because the N/G is considered an important factor in defining reservoir quality and in assessing the economics associated with reservoir development better prediction accuracy should be sought after.

As mentioned Sw appears the most important petrophysical factor that contributes to volume prediction errors. Sw is difficult to predict as it is a function based on the Archie equation and it varies between free water-level, oil/gas-water contact and the productive oil/gas-water contact. Also, Sw varies in 3D space whilst in predictions it is represented by a 1D value based on spatial averaging.

3.3.3. Result quality

The sample sizes for all parameters assessed in this study are sufficiently large to make well substantiated statements. However, it does stand out that for some parameters the sample size is considerably larger than for others. E.g. the sample size for the top reservoir depths analysis is 321 wells, while the NRV statistics are based on 135 wells. The question arises for what reason this data is missing from the database. Assuming EBN puts all the data that is provided in its databases, some operators might be considered reluctant in supplying data.

Table 3.: Pairwise t-test results Heggland et al. (2000). Test statistics for null hypothesis: no difference between prognosis and results

Parameter	*t*-statistic	Df	$P > \|t\|$
Hydrocarbon pore volume	3.4282	119	0.0008
Bulk rock volume	5.4012	84	0.0000
Hydrocarbon column height	4.5151	64	0.0000
Reservoir thickness	3.3342	81	0.0013
Net/gross	2.0800	186	0.0389
Porosity	0.3978	157	0.6914
Water saturation	3.5744	81	0.0006

The t-test results are in good confirmation with those obtained by Heggland et al. *(2000)* (table 3). Column height was in this research not taken into account in the t-test as it was derived from the top reservoir depth and the GWC depth, instead of being provided as a value by the operators. In both t-test result sets PHIE passes the null-hypothesis, indicating predictions are generally correct, while N/G barely passes.

All calculated prediction errors and subsequent biases are based on comparing the pre-drill estimate with the first post-drill measurement. Later measurements might provide a closer approximation of reality. Comparing GIIP updates over project duration shows however that, although significant individual updates do occur, on average project updates are minimal, only -3%. This indicates that as the actual volume becomes smaller and the difference between the commonly over-optimistic prognosis increases, the prediction bias enlarges slightly. This small update percentage also shows that the first measurement can be considered quite good.

Based on the cumulative averages of RV prediction errors and top reservoir depth prediction errors, we can be consciously optimistic about improvement of the prediction performance over time. To compare: the NPD reports in their 2018 annual report average accuracy factors over the various license rounds. For the period 1990 – ’97 resources where overestimated by a factor of 2.5. For ’98 – ’07 this value is equally 2.5 after which it lowers slightly to 2.1 between ’07 and ’16. These better top reservoir depth predictions probably result in better RV predictions. This is also showed by Janssen in her 2019 study. The bias percentage of the dataset in her study goes from 31% to 25% when only wells with a depth error smaller than 25 meters are taken into account in the pre-drill and post-drill volume sum.

4. Synthetic portfolio modelling

4.1 Selection Bias in the E&P industry

The results of the statistical look-back analysis do raise questions. What mechanisms can explain the observed prediction bias? Can the prediction bias be modelled on portfolio scale? These questions are addressed in this chapter using synthetic portfolio modelling. This chapter describes the synthetic portfolio modelling in which specifically attention is given to the phenomenon of Selection Bias as presented by Hoetz *(2016)*. Three variations of synthetic portfolios will be presented (fig. 12):

1. Equal Prospect Portfolio (EPP) – a simplified synthetic portfolio demonstrating the concept of SB in the E&P industry
2. Varying Prospect Portfolio (VPP) – an expanded version of EPP which is calibrated to actual data to make it more realistic and intended to quantitatively estimate SB.
3. Varying Clustering Portfolio (VCP) – an expanded version of EPP to assess the effect of portfolio maturation on SB. VCP is modelled with three variations: 3A. low clustering, 3B. medium clustering & 3C. high clustering. The latter represents a highly creamed portfolio (i.e. a portfolio where attractive prospects have been drilled already)

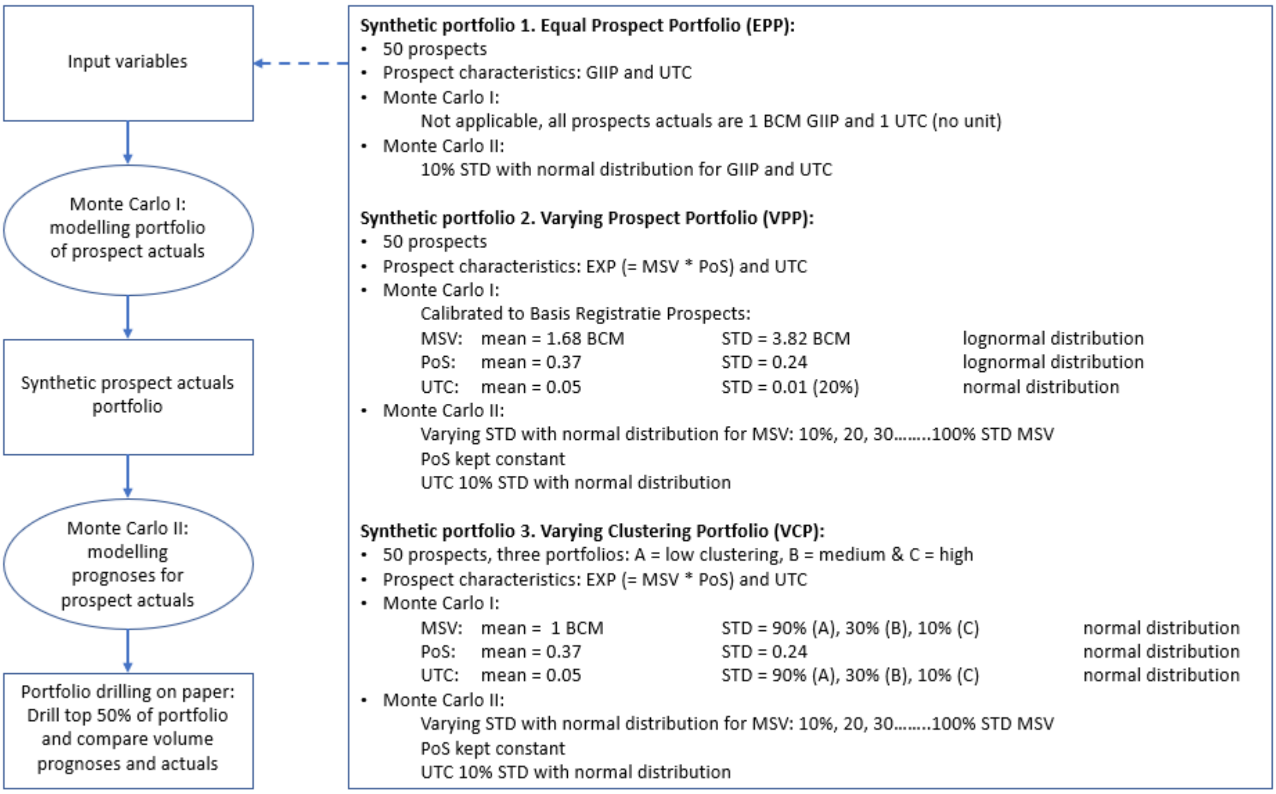

Figure 12: Schematic overview of the synthetic portfolio modelling workflow with input variables for the three synthetic portfolios presented.

4.2 Methodology

4.2.1. Equal & Varying Prospect Portfolios

In the context presented in this study, SB in portfolio ranking implies that selected datapoints (executed drilling projects) are not representative of the entire population (the project portfolio). In the E&P industry there is a clear preference for drilling lucrative (big) prospects or targets. This effect is demonstrated in synthetic gas prospects portfolio 1 (EPP): figure 13 shows a synthetic portfolio comprised of 50 prospects. Every prospect is characterised by just 2 parameters:

1) Volume

   Every drilling project aims to produce gas. The size of the accumulation is expressed as GIIP. We assume here that all prospects have a PoS of 100% and a Recovery Factor of 100%. All 50 prospect volumes in the EPP are quantified: 1 (units are arbitrary but could be, for example, 1 BCM).

2) Costs

   Every drilling project comes with costs. Unit Technical Cost (UTC) is defined here as all (monetary) costs involved divided by the volume of gas for each prospect. UTC's in the EPP are all quantified at 1. (Cost units are arbitrarily but could, for example be 0.10 euro/m^3)

Using the above parametrization any prospect can be represented as a point in the GIIP/UTC diagram. Any prospect portfolio is a collection of points in the GIIP/UTC diagram. Uncertainty in these parameters can be represented as ellipse-shaped clouds around these points (figure 13).

As geoscientists work with missing, incomplete or even misunderstood data, often actual parameters of the prospects are rarely prognosed correctly. To simulate prediction errors, for the 50 EPP prospects (the yellow dot in fig 13), a prognosis (blue dots) is synthesized using Monte Carlo simulation. Set constraints are 10% STD over the GIIP and UTC, with a normal distribution. In other words: 68% of the randomly generated realizations of the estimate are within range 90%-110% of the actual value (Monte Carlo II in fig. 12). Subsequently this synthetic portfolio is sampled (*drilled*) based on ranking by attractiveness, which is here controlled by value. Prospect monetary value, and hence ranking, can be determined by volume and cost. Prospects with (predicted) high volume and (predicted) low cost will be pursued first. The prospects that get drilled first persistently yield negative volume errors. In figure 13 the following example is given: the perceived most attractive prospect is 1.12 BCM. However as stated earlier: all the 50 prospects are actually 1 BCM. Ergo, the first drilled prospect yields an overprediction of 0.12 BCM. If the top 50% of the portfolio is drilled on paper, the prognosed volume is 27 BCM. The actual volume is 25 BCM, so a bias of 2 BCM occurs (the SB in this example amounts to 2/27= 7.4%). As we favour the perceived most attractive prospects, we tend to select the ones that deliver a disappointing volume outcome with respect to the prognosis. Of course, in reality not all prospects are equal size. Hence a more comprehensive model is designed (synthetic portfolio 2).

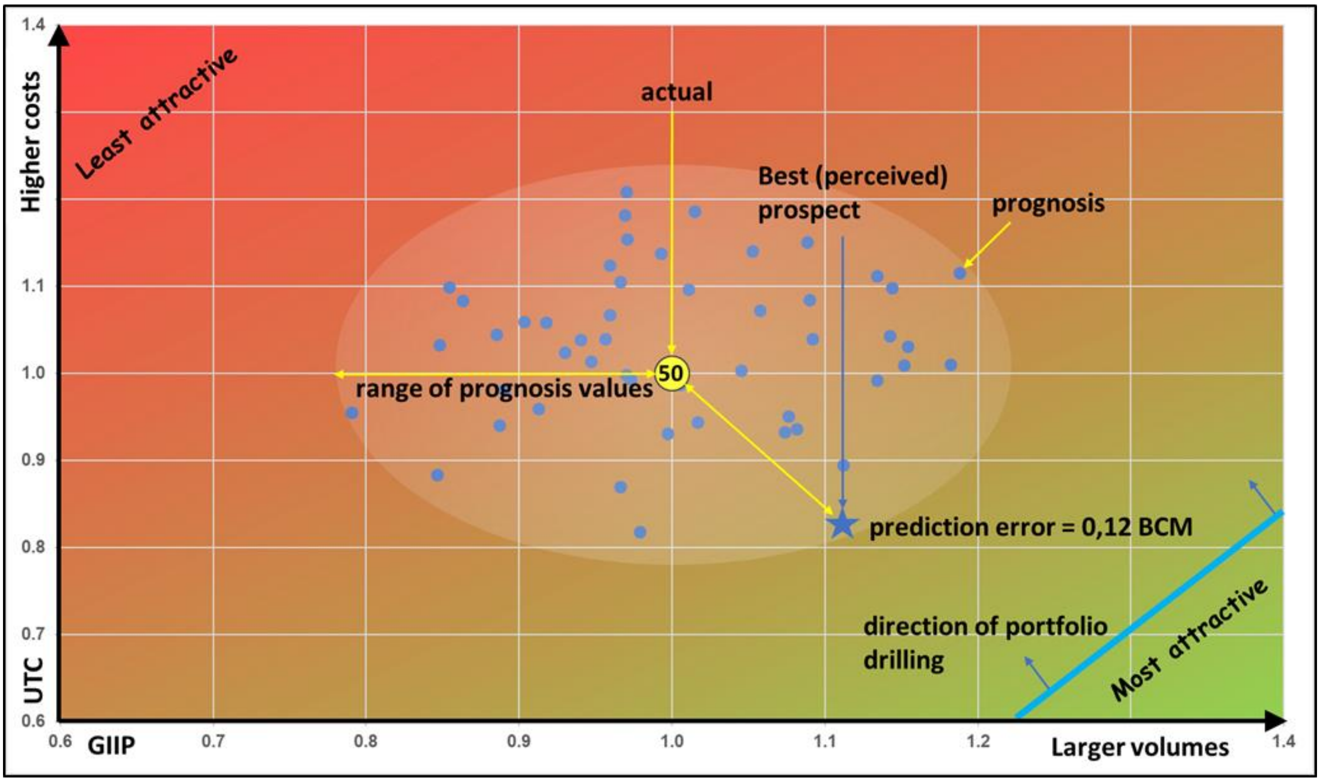

Figure 13.: Synthetic portfolio 1. "Equal Prospect Portfolio" in the GIIP/UTC diagram. Prospect actuals are plotted as yellow dot. Predictions (based on Monte Carlo runs assuming imperfect data) by blue dots. The portfolio is "drilled" from bottom right (attractive) towards top left. The perceived most attractive prospect is indicated by a blue star. Modelling parameters include: 50 prospects of which the top 25 are selected based on perceived attractiveness.

With the aim to more closely resemble reality, the VPP synthetic portfolio has been created. The VPP includes varying prospect sizes with statistical characteristics based on real EBN data. Exploration drilling is stochastically modelled using Monte Carlo simulations. The model is based on Microsoft Excel® in combination with Oracle Crystal Ball®, a spreadsheet-based application for predictive modelling. In general, the model is set up in the following manner:

Comparable to modelling the EPP, the VPP exploration portfolio makes use of major simplifications. A portfolio of 50 prospects is generated (Monte Carlo I in fig. 12) in which each prospect is represented again by two parameters only: 1) Expectation Volume (=EXP = Risked Volume = Probability of Success (PoS) x Mean Success Volume (MSV)) and 2) Unit Technical Cost (UTC). The PoS is the likelihood of a successful outcome, in this case a gas-find. The MSV represents the unrisked mean hydrocarbon volume in case of successful drilling. All above volumes are considered recoverable volumes (i.e. recovery factor is included). The VPP portfolio is based on gas prospects as this is most representative for the Dutch subsurface and to avoid confusing and unnecessary gas volume to oil volume conversions. A portfolio size of 50 prospects is chosen for we feel this number is an adequate representation of the portfolio of a typical operator. Also, on the practical side, stochastic portfolio modelling generates vast amounts of data. This affects processing capabilities and hence processing time significantly. We decided that 50 prospects is an adequate number to demonstrate the potential effects of SB without losing statistical significance or generating to much data which would result excessive model running times.

The parameters PoS and MSV are calibrated to actual EBN data of the Dutch exploration portfolio. For this purpose, the PoS and MSV values from all prospects in the national prospect database Basis Registratie Prospects (BRP) were extracted to calculate the mean and Standard Deviation (STD). Moreover, a histogram (fig. 14) is generated to visually inspect the distribution of the data. A lognormal distribution for both PoS and

MSV is recognized. This is in agreement with relevant literature *(e.g. Quirk & Ruthrauff, 2006)* and hence the Monte Carlo simulation is based on a lognormal distribution (Monte Carlo I in fig. 12). Cut-offs of the distributions in the Monte Carlo simulation are inserted so that a negative MSV is not possible in generating prospects and that the PoS values always are in the range of 0 to 1.

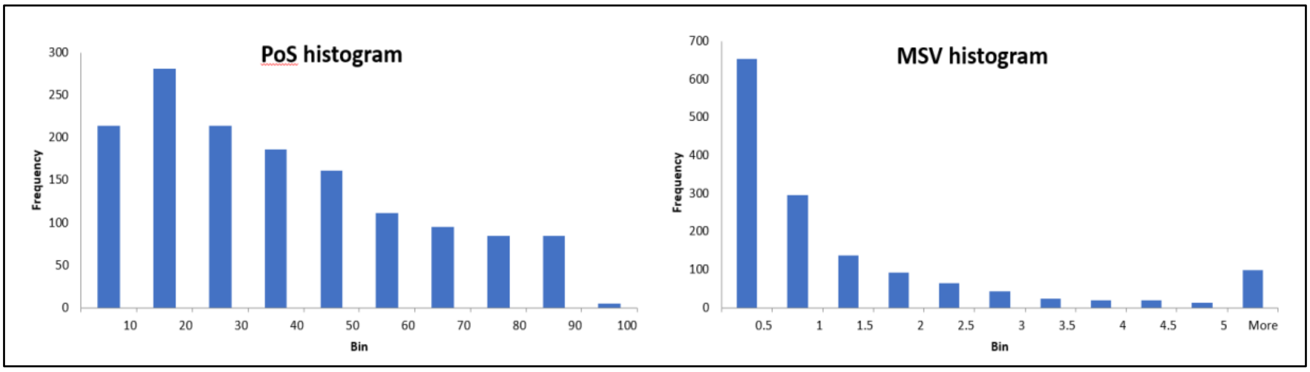

Figure 14.: Prospect portfolio characteristics from the national prospect database. Left: PoS, Mean = 0.37. STD = 0.24. Right: MSV. Mean = 1.68 BCM. STD = 3.82 BCM. These parameters have been used to constrain the simulated portfolios.

For the UTC no easily accessible and useable data is present to constrain the VPP model. For this reason, the characteristics of this parameter are based on the extensive experience with the E&P industry of supervisor G. Hoetz. An UTC of 0.05 +/- 0.01 euro per cubic meter gas is considered reasonable. Assuming a normal distribution this translates in UTC mean of 0.05 €/m3 and a UTC STD of 20%,

With these input parameters, 50 prospects are stochastically generated (Monte Carlo I in fig. 12). These generated values represent the prospect actuals. To model the effect of noisy data, for each prospect actual (the yellow dots), a prognosis (blue dot) is simulated within set constraints. These prognosis values are simulated again using Monte Carlo simulation (Monte Carlo II in fig. 12), only in this instance with a normal distribution instead of the lognormal distribution used for modelling the prospect portfolio. This implies an equal chance of over-prediction as well as under-prediction. The UTC STD and PoS is kept constant while over various runs the MSV STD is adjusted (scaled to the MSV actual of the prospect), thus influencing the range of prognosis values, to simulate varying degrees of uncertainty. When the synthetic portfolio is drilled on paper, the simulated prognosis is compared to the actual on which it is based in order to calculate the prediction error. The sum of all prediction errors gives the volume bias. The relative volume bias is the ratio of the volume bias compared to the predicted volumes.

Of the 50 prospects in the synthetic portfolio, the top 50% is “drilled on paper”. To determine which prospects are the most attractive and thus belonged to the top 50%, ranking criteria are drawn up. Each prospect is given a (monetary) value and thereafter ranked based on the following formula:

*Prospect value = EXP * (gas price – UTC)*

*= (MSV * PoS) * (gas price – UTC)*

The “drilling on paper” is done as shown in figure 15. Prospects with a high value (i.e. high EXP and low UTC) are drilled before prospects with lower value. Only half the portfolio is drilled assuming the remaining part of the portfolio is sub-economic.

An important aspect of the model is the overlapping of the various prognosis ranges, in figure 15 these ranges are the egg-shapes. The “eggs” represent the evaluation uncertainty expressed in percentage STD MSV and percentage STD UTC. To test the impact of the evaluation uncertainty in the VPP, multiple runs are executed with varying degrees of evaluation uncertainty. To do so the prognosis values for the prospects are generated

(Monte Carlo II in fig. 12) with varying STD percentages in MSV (STD MSV ranging from 10% to 100% in steps of 10%). With increasing STD, the range in which a prognosis can occur increases which is represented by increasing *egg* size.

Overlapping of the prognosis ranges (the "eggs") implies that a small prospect could be perceived larger than another actually bigger prospect. An example of this process is given in figure 15.D. Prospect 30 is perceived more attractive than prospect 18 since its prognosis has a larger EXP. In reality however, the actual of prospect 30 has not only a lower EXP but also a higher UTC. Based on perceived attractiveness prospect 30 is drilled on paper before prospect 18. This ranking error results in a negative volume error.

Alternatively to the drilling on paper based on prospect attractiveness, if the same portfolio is drilled randomly no SB can occur. We assume that no, or minor, volume bias would be the result. To test this, the VPP synthetic portfolio is not only drilled on paper based on prospect ranking by attractiveness, but also by randomly selecting prospects.

4.2.2. Model uncertainty

For each run (= step in STD MSV), 100 synthetic portfolios are processed, and averages are taken of the outcomes. This is done to reduce statistical randomness in the Monte Carlo simulation. Further noise reduction can be achieved with larger sample sizes. Some test runs are executed based on 1000 synthetic portfolios to inspect the effect, but minimal improvement was observed while the large sample size did affect the running time of the model significantly. Alternatively, the statistical uncertainty in the model is also effectively assessed by executing the same run multiple times and checking for the variation in the outcomes.

4.2.3. Varying Clustering Portfolio (VCP)

Additionally, the effect of portfolio maturation which results in prospect clustering in the EXP/UTC domain is analysed using the VCP synthetic portfolios. We hypothesize that with high clustering, meaning a low spread of the prospect EXP and UTC (actual) values, the observed bias is larger than when the prospects are far apart/low clustered. To assess this clustering influence three simplified prospect scenarios are generated. One representing low clustering (synthetic portfolio 3A), one medium (3B) and one high clustering (3C).

4.2.4. Effect of gas price on ranking

Lastly, the effect of the gas price on the ranking process is analysed. In this assessment no new synthetic portfolio is designed. It is rather based on the VPP synthetic portfolio but with varying gas price in the ranking formula. Besides the "standard" gas price of synthetic portfolio 2 with a value of 0.15 (€/m^3), model runs are executed with a gas price of 0.1 and 0.5 (€/m^3)

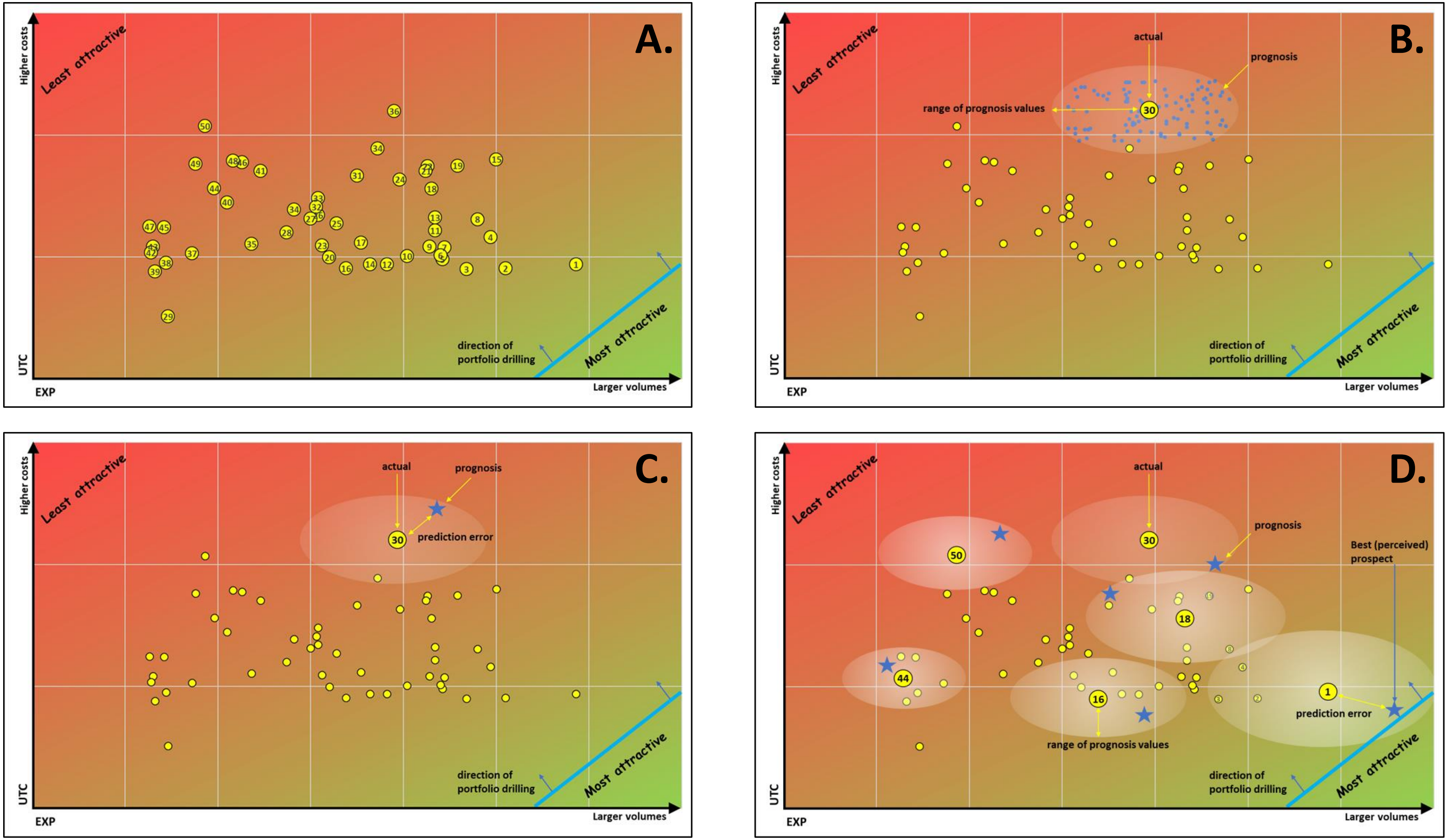

*Figure 15.: Overview of the general build-up of the synthetic portfolio 2. A.) The portfolio contains 50 prospects, each with two characteristics: EXP and UTC. Ranking is based on the formula: Prospect value = EXP * (gas price – UTC), B.) For each prospect a range of prognoses is generated. Each within a range of prognosis values (the "egg"), C.) The blue star represents a stochastically generated prognosis of the actual (yellow circle). The distance between the prognosis and the actual defines the prediction error, D.) Prospects actuals are plotted as yellow dot. Predictions (based on Monte Carlo runs assuming imperfect data) by blue dots (here only indicated for prospect #30). The randomly selected prediction is indicated by a blue star. The portfolio is "drilled" from bottom right (attractive) towards top left. Modelling parameters include: 50 prospects of which the top 25 are selected based on perceived attractiveness.*

4.3 Results

4.3.1. Varying Prospect Portfolio (VPP)

Figure 16 displays the prospect distribution as used in the VPP synthetic portfolio (Monte Carlo I in fig. 12). The prospect spread clearly displays a lognormal distribution: a lot of small prospects, some medium sized ones and few relatively large prospects. This is in good correspondence with the BRP and relevant literature that all report a lognormal spread of prospects *(e.g. Quirk & Ruthrauff, 2006)*.

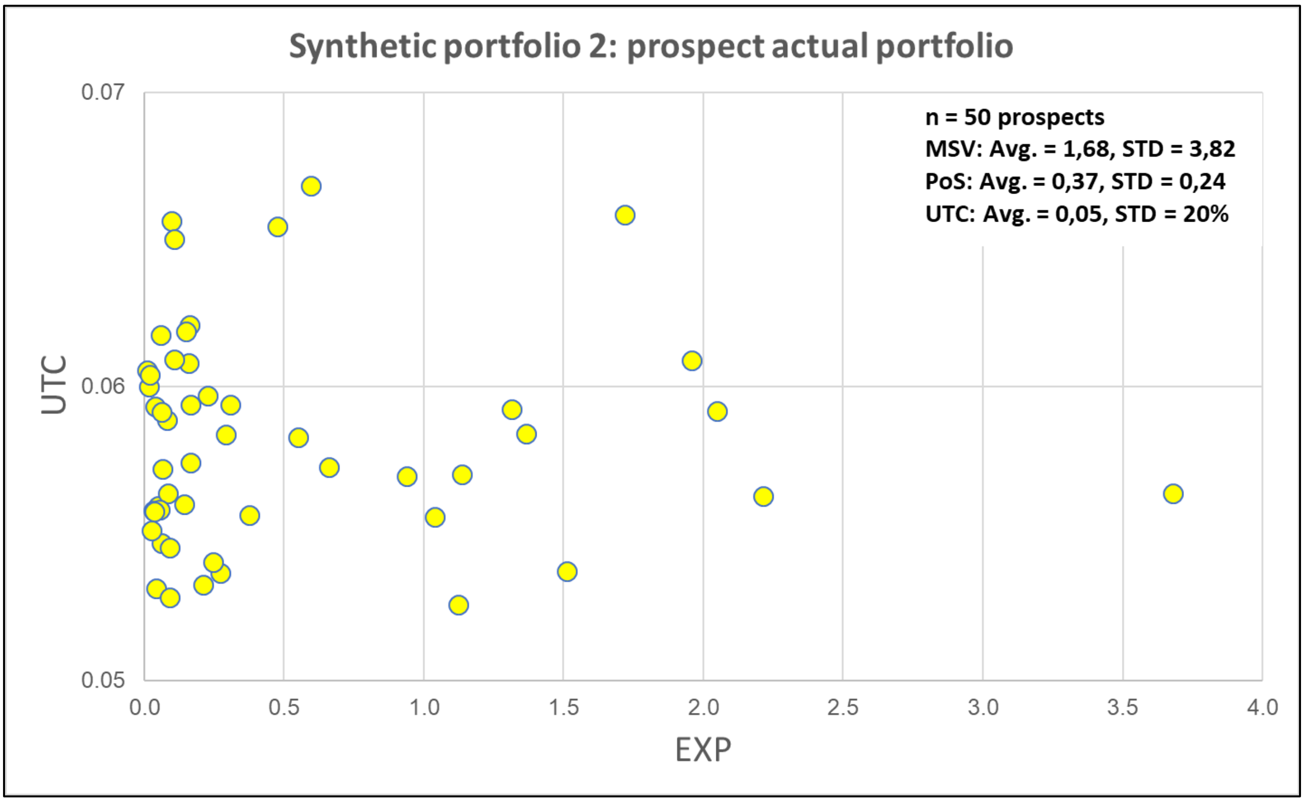

Figure 16.: Synthetic portfolio 2: prospect actual portfolio. Prospect distribution statistics is based on the BRP. The figure represents the result of the "Monte Carlo I" step in figure 11. The cumulative EXP of the top 25 prospects is 61 BCM.

Figure 17 and 18 show the results of the VPP. This portfolio is calibrated using EBN data and tries to attain results as close as possible to reality within the set simplification of the model. With every step of 10% increase in the STD of the MSV the prognosed volume over the portfolio grows. With relatively small prediction uncertainty ranges ("eggs") of up to 30% STD in MSV, the prognosed volume is (very) close to 61 BCM, which is the actual volume of the drilled part (top 50%) of this portfolio. With increasing uncertainty (St. Dev MSV), the prognosed volume increases steadily to around 83 BCM at 100% STD MSV. This is in line with expectation as with increasing prediction uncertainty the prognosis can deviate further from the actual. The actual however remains the same, which we observe in the relatively steady actual drilled volume. This value starts at around 61 BCM with 0% STD in MSV values and ends at around 59 BCM at 100% STD. This (modest) reduction in EXP actuals is related to the fact that uncertainty influences which part of the portfolio is to be drilled. Uncertainty impacts the portfolio ranking and hence some prospects are selected that actually do not belong in the top 50% of best prospects. The increasing difference between the prognosed and actual drilled volume obviously translates to an increasing prediction bias. We see this in the blue line representing the bias percentage. This increases steadily from 0% to 30% bias in predicted volumes. From the graph a clear relationship between prognosis bias (%) and evaluation uncertainty (STD MSV %) can be observed, although this relationship is not linear.

Further comparing the outcome of the model to the situation in the Dutch subsurface yields that the prediction uncertainty range for the Dutch reality corresponds roughly to a 62% or even 74% STD in the MSV's in the model. This is based on the dataset used in figure 3, which shows the recoverable volume mid-cases, low- to high-case uncertainty range and the actual measurement. This information which originates from Dutch operators based on pre-drill assessments, can be used to estimate the evaluation uncertainty. By dividing the uncertainty range by two and then dividing it by the mid-case prognosis we find a value of 62% (i.e. pre-drill uncertainty estimate). The 74% in an empirical uncertainty estimate and results from calculating the absolute difference between prognosis and actual and dividing this by the actual. This means that, on average and assuming normal distributions, the volumetric uncertainty (one STD) is +/- 62% or +/-74% of the base case volumes depending on the way of estimating uncertainty. Based on the model outcome as shown above this level of evaluation uncertainty would generate a volume bias percentage of 13% or 17%.

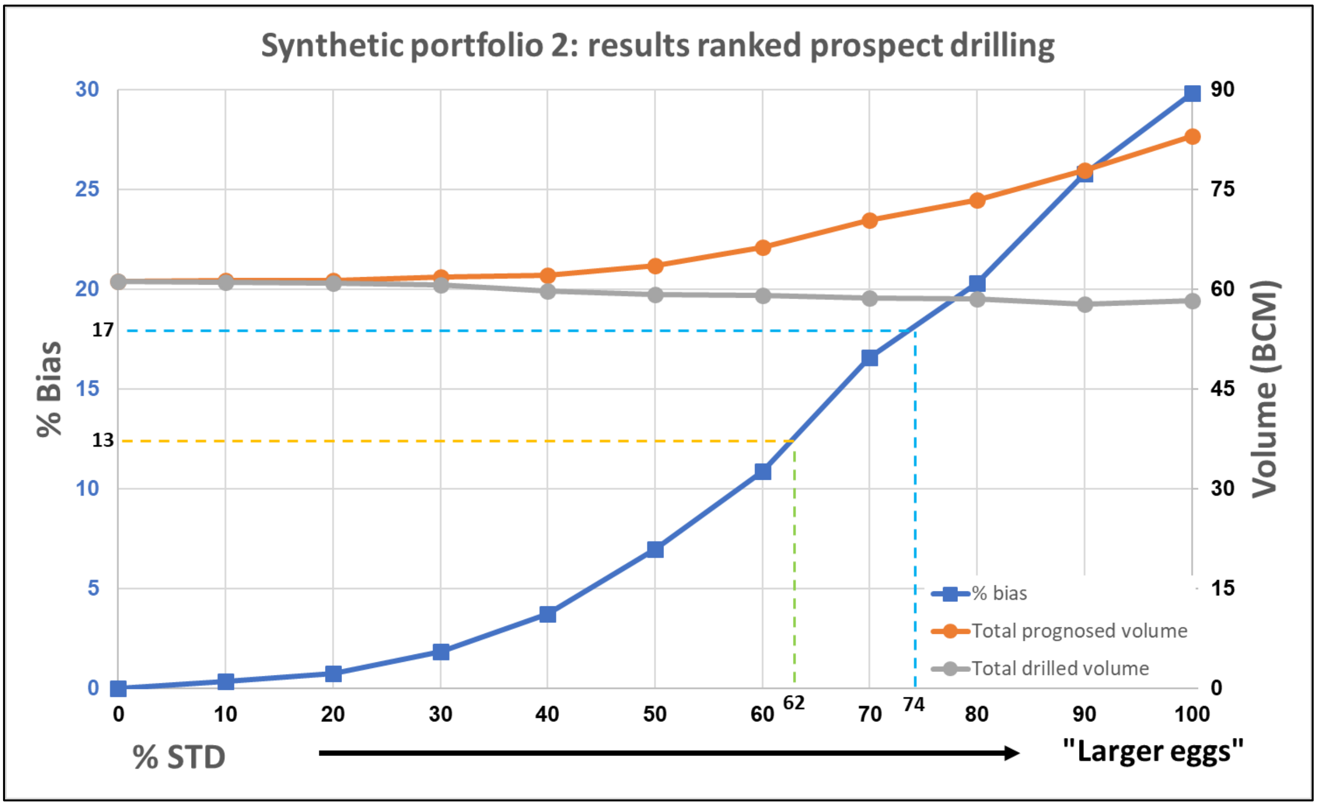

*Figure 17: Synthetic portfolio 2: results of ranked prospect drilling. Drilling on paper is done based on prospect ranking via formula: value = EXP * (gas price – UTC). Assuming an evaluation uncertainty of 62% results in 13% volume bias at portfolio level. An evaluation uncertainty of 74% results in 17% volume bias at portfolio level.*

In figure 18 the results for the VPP portfolio as figure 16 are shown only now if drilling on paper occurs randomly instead of ranking. Immediately striking are the lower prognosed and actual volumes. Both prognosed and actual start at around 35 BCM for no (or low) evaluation uncertainty. Prognosed volumes then rise to approximately 45 BCM and the actual drilled steadily remains to be 35 BCM (with some small changes amongst the varying STD percentages). Bias percentages start at relatively high STD percentages to be significant. Only from 40% STD onwards we see a noteworthy difference between prognosed and actual volumes. At 40% even a small negative bias percentage is observed, so the actual measured volume is larger than the prognosed. Bias percentages become as large as 24% in this scenario.

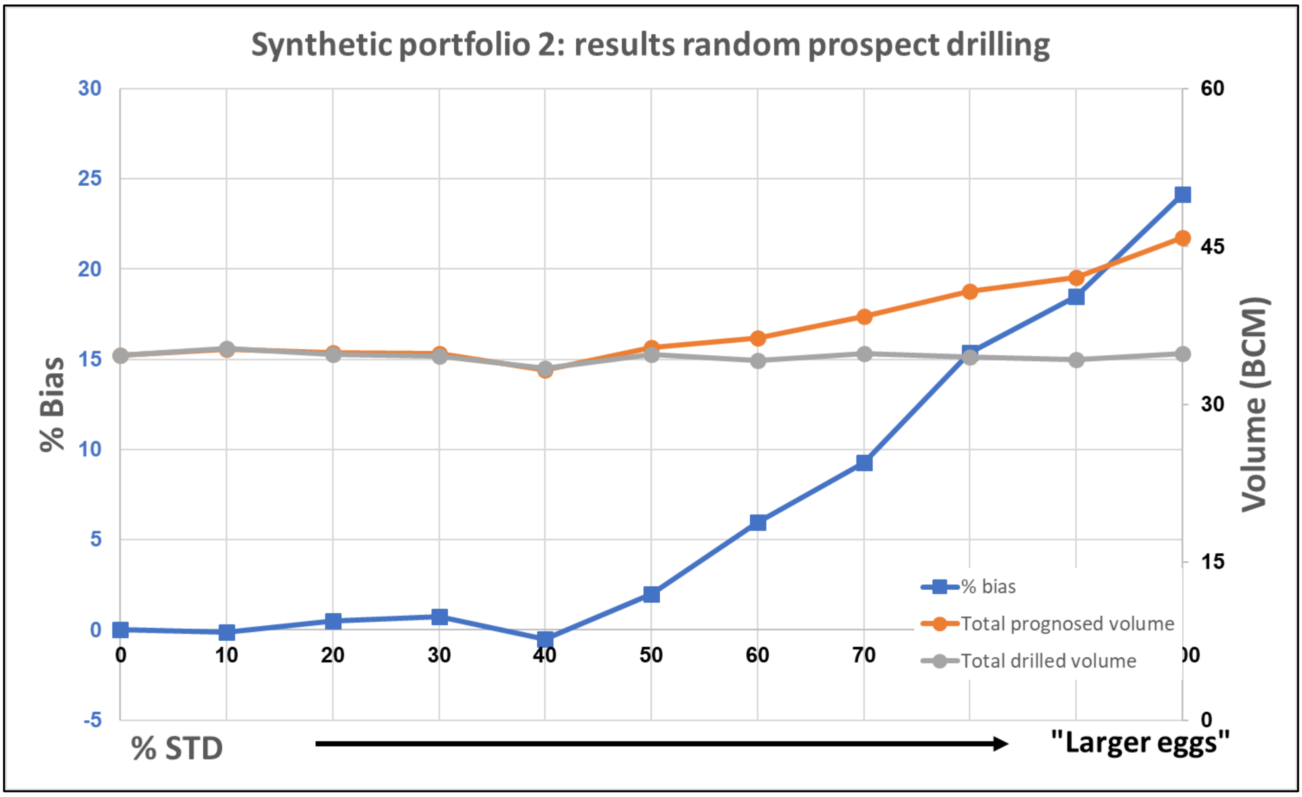

Figure 18.: Synthetic portfolio 2: results of random prospect drilling. Drilling on paper is done based on random prospect selection.

4.3.2. Modelling uncertainty

To assess the uncertainty/statistical noise in the modelling, the same run (thus constant STD%) is executed ten times. This produced a variation of 0.5% in volume bias for prospect drilling based on attractiveness and 1.2% for random drilling. Each value in the results can thus statistically fluctuate by 0.5% or 1.2% depending on the method of drilling on paper. This is a relatively small variation and hence it is concluded that the number of stochastic realizations (100) is sufficient.

4.3.3. Varying Clustering Portfolio (VCP)

Figure 18 displays the results showing the amount of volume bias for the varyingly clustered portfolios (VCPs). Evidently, with increasing clustering of the prospect actuals the bias percentage rises. As these synthetic portfolios are not tuned to real data the absolute bias percentages might be not so meaningful. However, the fact that the bias gets larger with increasing clustering and the overall trend of the graphs are still insightful. Especially the (near) linear appearing trend of the high clustering scenario, with its immediate rise from 0% STD, is noteworthy.

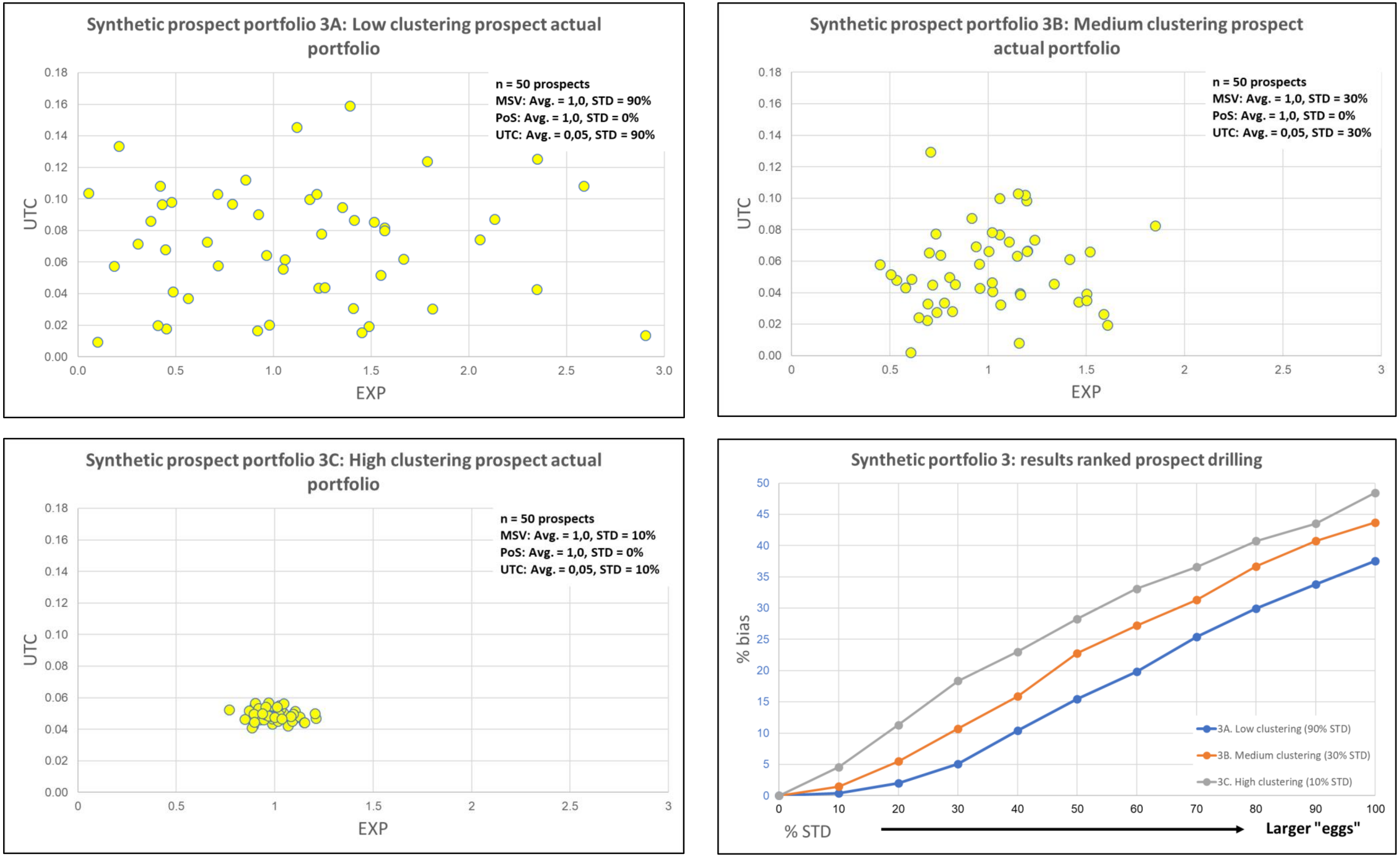

Figure 19.: VCP results. Simplified portfolios are stochastically generated to check the dependency of SB on prospect clustering. A.) Low clustered portfolio, B.) Medium clustered portfolio, C.) Highly clustered portfolio, D.) Bias percentages for these three portfolios as function of evaluation uncertainty (%STD MSV).

4.3.4. Effect of gas price on ranking

As mentioned earlier the prospect value is based on the following formula:

*Prospect value = EXP * (gas price – UTC)*

In this formula the gas price is kept constant, so the formula has three variables and can be plotted with two axes using iso-value lines (fig. 20). In case of varying gas price, displaying this formula in principle requires a multivariate plot with three axes. In this study the synthetic portfolio is projected in two-dimensional plots with on the x-axis the EXP and on the y-axis the UTC and iso-value lines for a particular gas price. The hypothesis is that, with a changing gas price, the direction with which the portfolios are drilled in the plots, and thus the ranking, varies depending on the direction of the iso-value lines. To test the impact of the varying gas price, two additional scenarios are run using the VPP portfolio. Next to the base case gas price of 0.15, a low case of 0.1 (fig. 20) and a high case of 0.5 (fig. 21) have been used (the latter might be unrealistically high given today's market conditions but is intended for illustration only). First the direction of portfolio drilling was visually displayed in the 2D plots. The prospect value formula was plotted in the R® extension Plotly® as a heatmap with contour lines representing the iso-value lines. Figures 20 and 21 show the effect a changing gas price has on prospect ranking. Clearly with increasing gas price the volume factor (EXP) becomes increasingly influential in the ranking process. This effect is more pronounced for small prospects. The contour lines indicating prospect value become steeper when EXP is smaller. In these plots steeper contour lines indicate that the ranking is more influenced by variation in EXP than in UTC.

Effectively the ranking of the prospects indeed varies with gas price. For example, prospect #6 (based on low case gas price) will be drilled on paper as #5 with a gas price of 0.5. When running the model to check which effect the changing gas price has on the volume bias percentages, no clear effect can be deducted. Figure 22 shows the bias percentage for the three gas price scenarios used here: 0.1 (low case), 0.5 (high case) and 0.15 (base case) as reference. It seems that the high gas price of 0.5 results in a lower bias of 29% with respect to the bias percentage of 30% from the 0.1 gas price scenario. This difference is small however and not significant given the statistical noise in these simulations as seen earlier. We conclude that the gas price does indeed have an effect on the ranking of the prospect in the portfolio but the effect on the volume bias is minor.

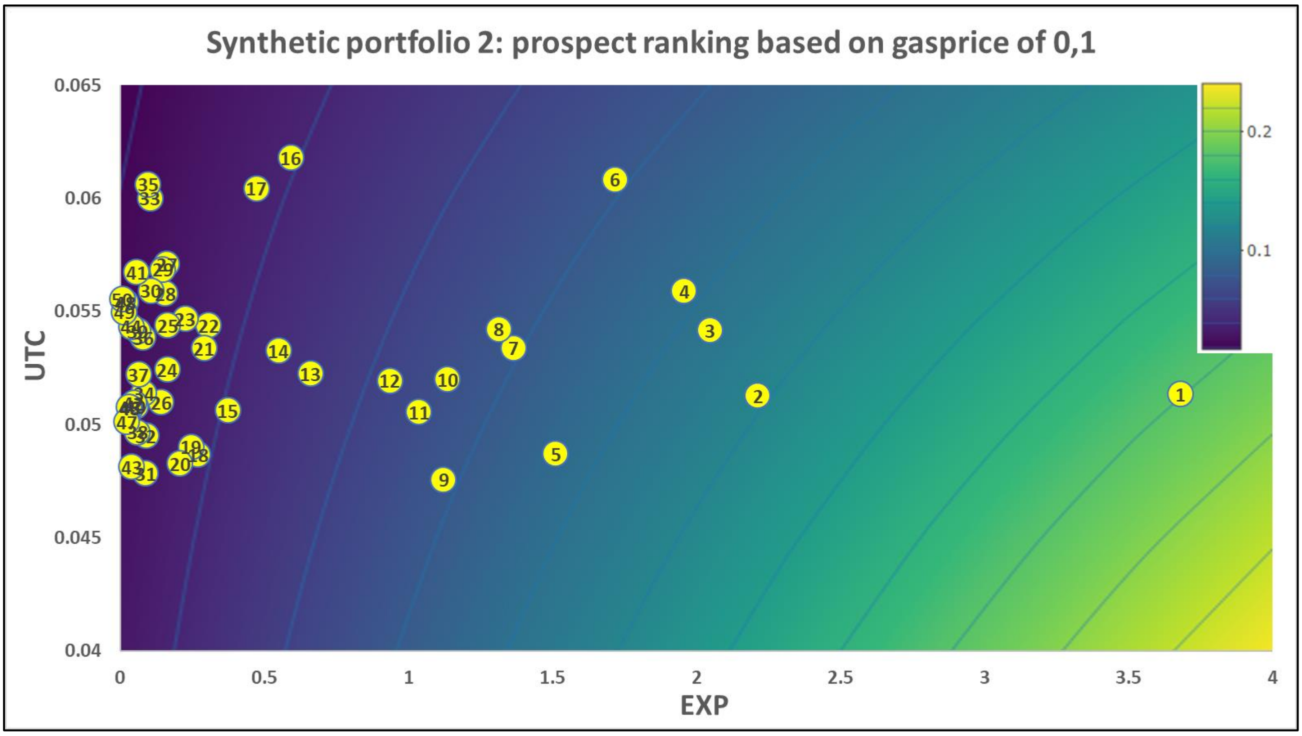

Figure 20.: VPP: prospect ranking based on gas price of 0.1 plot. Contour lines represent iso-prospect values. Prospect Values are in E07 million euros, EXP in BCM and UTC in euro/m3.

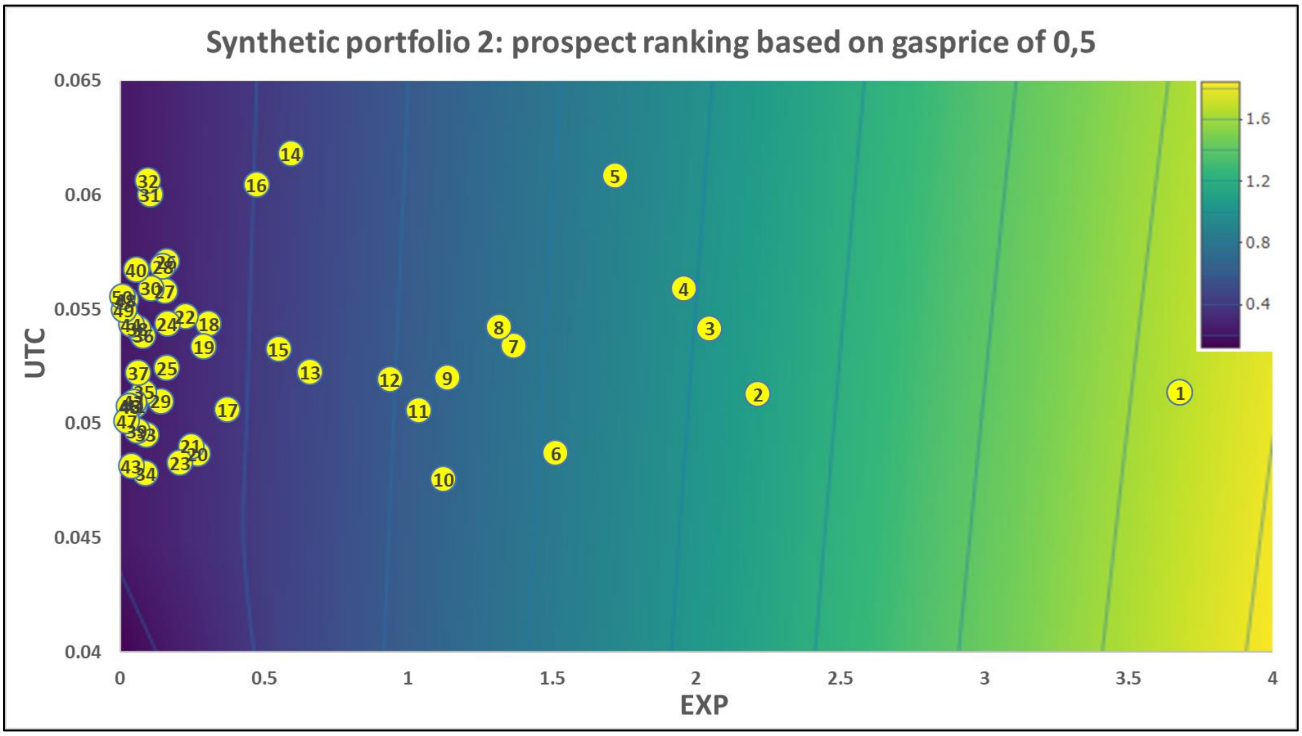

Figure 21.: VPP: prospect ranking based on gas price of 0.5 plot. Contour lines represent iso-prospect values. Prospect Values are in E07 million euros, EXP in BCM and UTC in euro/m3.

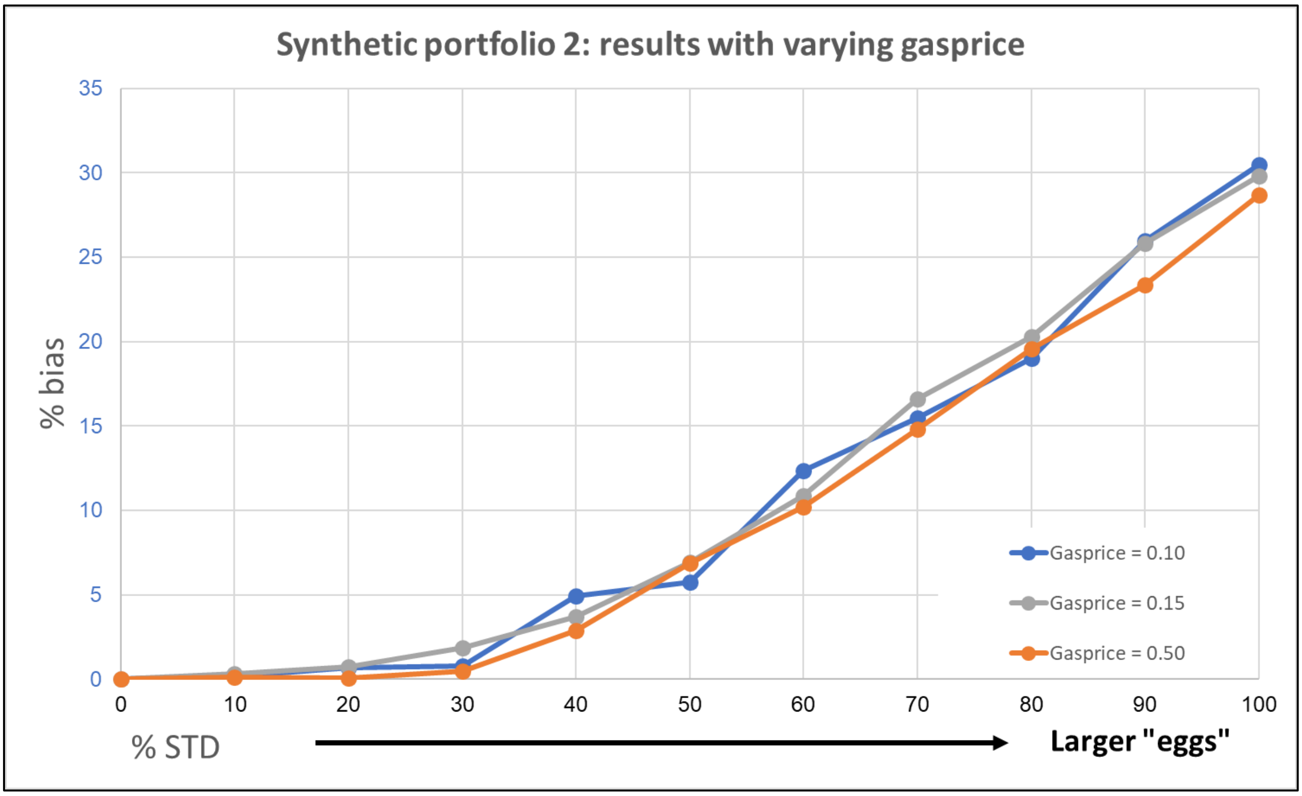

Figure 22.: VPP: portfolio modelling results using. The graph shows the volume bias resulting from SB for varying evaluation uncertainty (% STD) and 3 different gas prices.

4.4 Discussion

4.4.1. Varying Prospect Portfolio (VPP)

As with all models our synthetic portfolio model is an approximation of realty. It is important to keep an eye of the inherent limitations. Modelling results of a portfolio with realistically varying prospect sizes and costs are summarized in figure 17. It shows that the volume bias percentage remains small (< 5%) until the prediction uncertainty (STD in MSV) exceeds 20%. It appears also that as the prediction uncertainty (represented by *eggs* in the EXP - UTC portfolio diagram) become larger, no significant change in cumulative volume drilled from the portfolio occurs. Probably due to the relatively wide range of prospect volumes (in other words: little clustering) the ranking is hardly impacted by the evaluation uncertainty. The largest prospects still get drilled whether they are under- or overpredicted. As the prediction uncertainty range relies on a normal and hence symmetrical distribution (Monte Carlo II in fig. 12), both possibilities are equally plausible. Overprediction in the drilled prospects should thus theoretically occur as often as underprediction. As the results are averaged over 100 synthetic portfolio realizations, the under- and overpredictions therefore probably cancel each other out and no significant bias is observed. Alternatively, it appears that, from 30% STD MSV onwards, the topmost attractive prospects start to have prediction uncertainty ranges large enough so that significant overlapping *of* uncertainty ranges occurs, and the prospect ranking is affected. The prospect portfolio as illustrated in fig 15.D. illustrates how these uncertainty *eggs* can overlap

As the prediction uncertainty ranges start to overlap, it becomes a possibility that prospects that appear more attractive start to rise in the prospect ranking. This however is based on perceived attractiveness. These prospects might actually be of smaller EXP or larger UTC, compared to another prospect that is then downgraded in the ranking due to perceived less attractive characteristics. We see this effect back in the increasing discrepancy in the prognosed and drilled volumes. The rise in prognosed volumes does not come as a surprise as with increasing prediction uncertainty the prognosis can be further away from the actual, thus increasing the prediction error. The fact that the total drilled volume decreases, although slightly, proves that with increasing prediction uncertainty indeed prospects that are actually less attractive indeed might get favoured in the ranking process.

For this test the model is run up to MSV STD of 100% for which a volume bias percentage of some 30% is observed. This model predicts a bias percentage of 13% or 17%, depending on the method of calculation, for the prospect portfolio of the Netherlands. This is based on an average prediction uncertainty of 62% and 74% respectively, expressed as STD in MSV values using the dataset illustrated in figure 3. The method used here to determine this percentage average prediction uncertainty might be optimised further. This could be done by dropping the assumption of normal distributions for EXP estimates. However, this requires more complex simulations and is outside the scope of this project. Another limitation of the current approach is that the current prediction uncertainty is estimated by using the pre-drill information only. By using the actual post-drill parameters, a better estimate of the evaluation uncertainty can be obtained.

The phenomena of SB leads to the hypothesis that random drilling (i.e. no ranking and selection) would not result in volumetric bias. In that case the bias percentage should be zero. Indeed, the bias percentage with random prospect selection is lower, however it is not zero, in particular for larger values of STD EXP (fig 18). This can be explained as follows. Although the prediction ranges for the prospects are based on a normal distribution (Monte Carlo II in fig. 12), in reality, and also in these models, the distributions are not truly normal. Normal distributions are symmetrical, infinite and do include negative values. In practice portfolios are truncated and do not include negative EXP and or UTC values. This introduces skewness of the distributions and this translates into volumetric bias. Although not tested here, it is expected that randomly drilling synthetic portfolios which do include negative volumes would indeed result in bias free prediction errors.

4.4.2. Model uncertainty

The results of the portfolio drilling do contain some noise, no perfect linear relationship is present. The measured uncertainty of the model of 0.5% for most attractive prospect drilling or 1.2% for random prospect drilling does smooth these possible anomalies, though not to the full extend. No clear explanation except for statistical randomness can be given for this. Potentially better results could be obtained by basing the results on averages of 1000 synthetic portfolios. As mentioned, this does however result in excessively long, impractical running times of the model. While developing the modelling code special attention was given to operational performance in order to allow Monte Carlo runs with sufficient sampling. Nevertheless, it appears that statistical noise due to under sampling is present in the results which expresses itself as “jittery” curves in the graphs (e.g. fig 22). Better computational performance might be achievable using alternative code like Python®.

4.4.3. Varying Clustering Portfolio (VCP)

The results for the clustering assessment are in good correspondence with the hypothesis. Indeed, as more clustering of the prospects occurs due to decreasing spread in the prospect sizes, the bias percentage rises. The mechanisms behind this is most likely that with increasing clustering the prediction uncertainty ranges, the “eggs”, will start to overlap significantly more. A STD in MSV of only 10% in the high clustering scenario results in a prediction uncertainty range where most prospects do already overlap. When clustering is low and the ranges do not overlap much, a some of the selected prospects (up to 50%) do show underprediction. This translates into modest bias. In the case of the high clustering scenario nearly all *eggs* do overlap leading to many cases of overpredictions. Increased volumetric bias percentage is a result (fig 19.D.). The mechanism described above does also explain the steady rise in bias percentages of the high clustering scenario. While the medium clustering and low clustering scenario generate a bias percentage of approximately 2 and 1%

respectively, the high clustering scenario shows a much more linear trend and gives a bias percentage of nearly 5%. Possibly the linear characteristics of the bias function are also largely dependent on the clustering.

Prospect clustering can be considered a characteristic of a matured portfolio. As the portfolio become more exhausted, the larger prospects have been developed and the smaller volumes do remain. The modelling presented here indicates that in more matured portfolios the influence of SB increases. This might specifically apply to the Dutch subsurface as this is considered a very mature hydrocarbon basin.

4.4.4. Effect of gas price on ranking

Finally, the gas price assessment shows that the gas price clearly influences the prospect ranking process. Hence one can assume that a variation in the gas price should have an effect on the sum of volumes drilled and thus on the volume bias percentage. However, in the assembly of prospects in the synthetic portfolio used here, this effect does not show up. Indeed, the ranking sequence changes slightly, but whether a relatively large prospects get drilled for example as #5 or as #6 in the drilling sequence doesn't really matter. In the end this prospect is selected and drilled on paper anyway and thus it contributes to the total volume. Surely the ranking sequence is altered but this only affects those prospects that drop out of the "top 50%" selection.

To conclude, the synthetic portfolio drilling shows that the bias observed in the statistical look-back analysis can be explained - to some degree - by the phenomena of SB. The drilling on paper of a synthetic portfolio which is calibrated with real data and with specific assumptions shows that volumes are overestimated by 13% or even 17% solely by the mechanism of Selection Bias. As the volume bias percentage observed in the statistical look-back amounts to 42%, other factors might play a role. Other biases creating factors such as cognitive biases and evaluation tool-induced biases appear to play a role too. The interesting aspect of SB is, that the impact of this mechanism can be modelled with a limited set of assumptions. This has not been done so far in the Netherlands. To advise the Dutch Ministry of Economic Affairs and Climate, EBN uses the resource prediction tool ExploSim to estimate future gas resources *(Lutgert et al., 2005)*. Including the effect of SB might lead to better predictions. A first step in this could be to apply the ranking process as used in ExploSim to the model built in this study.

Overestimating resources due to SB, might be unavoidable as it appears inherent to the method of portfolio ranking prevailing in the E&P industry. Although unavoidable, the effects of SB can be reduced by reducing the project uncertainties. In particular geological parameters tend to be highly uncertain as our subsurface is highly heterogenous and under sampled. It is the task of the geoscientists to think careful about data gathering campaigns and to perform high quality subsurface evaluation work where uncertainties are being understood, quantified and reduced as much as possible.

5. Conclusions

The main research question of this study is to identify and quantify the key parameters that contribute to prediction bias in hydrocarbon volumes. This is addressed by reviewing the quality of the predictions of the subsurface parameters being used for volumetric assessments. Prediction quality is assessed in a statistical manner based on historical portfolio data of the Dutch subsurface. Furthermore, prediction bias effects are modelled by the means of synthetic portfolio modelling. The following conclusions are be drawn:

- Significant prognosis errors are present both in volumes and in underlying geological/petrophysical parameters on single well scale. Prognosis errors often hint towards over-optimism.
- The volume prognosis errors translate to a significant bias on portfolio scale as revealed by the statistical look-back analysis of 215 drilling projects from EBN's database. Volumes are generally over-estimated by almost a factor of two and only half of the finds are within the uncertainty range (high case – low case range as defined pre-drill).
- Multiple underlying geological parameters do contribute to the observed volumetric bias. Gross rock volume and water saturation are major contributors as these are most often predicted too optimistically.
- Prediction bias can be understood and modelled based on the concepts of Selection Bias.
- A bias in hydrocarbon volumes is unavoidable due to Selection Bias. As the ranking process of prospects prefers large prospects (being truly large or perceived large) over-optimistic projects are drilled preferentially. In case of really random sampling in the drilling portfolio no Selection Bias occurs.
- Selection Bias is a function of evaluation uncertainty. Hence well-planned data acquisition and high-quality technical work should reduce Selection Bias.
- Selection Bias is modelled to increase with maturation (creaming) of the portfolio. Matured portfolios are characterized by increased clustering of the prospects in the EXP/UTC diagram.
- Synthetic portfolio modelling shows that 17% bias in recoverable volumes can be expected as a result of Selection Bias for a portfolio comparable to the Dutch prospect database. As a 42% volume bias has been observed in this look-back study, it appears that also other mechanisms do play a role in causing volume error bias. Cognitive bias and Evaluation Tool Induced bias have been suggested to contribute here.

6. Recommendations

Based on the outcomes of the research, some further recommendations are:

- Findings of the look-back analysis can help prioritize which subsurface parameters (and/or operators) need more careful attention in reviewing project proposals.
- Extending the national well database Basis Registratie Boringen with data on uncertainty range of reservoir parameters might lead to more robust prediction quality assessments.
- Conducting high quality technical subsurface work should reduce Selection Bias. Communicating this message to EBN's partners might assist in generating more realistic predictions in project proposals
- The EBN tool *ExploSim* aims to predict future exploration resources for the Dutch state. Incorporating the effect of Selection Bias might help to improve the quality of the *ExploSim* prognoses.
- Geothermal developments might also be subject to Selection Bias, in particular if ranking is driven by uncertain subsurface parameters e.g. permeability. Further study might help to mitigate the effect of Selection Bias on business decisions for geothermal projects.

Acknowledgements

I first and foremost would like to thank my supervisor Guido Hoetz for the unwavering support and feedback. Without him this thesis surely would not have been possible. Gratitude also goes to Martin Ecclestone, who showed great interest and great insight in the project. I enjoyed working on an EAGE abstract[1] with both these gentlemen. Fred Beekman is thanked for acting as supervisor from the university and for providing feedback. Next I would like to thank my fellow interns with whom I shared the experience. I also greatly value the technical help and support from my other EBN colleagues. In this regard especially Pieter Slabbekoorn is thanked for his help with the data for the statistical analysis. Finally, I greatly appreciate the support from friends and family during the process of writing this thesis.

[1]Hoetz, H.G., Ecclestone, M., van der Kraan, V. (2020). Drilling portfolio performance and the role of Survival Bias in volume estimates. *Pending abstract, 82nd EAGE Conference & Exhibition 2020, June 2020.*

References

Baddeley, M. C., Curtis, A., & Wood, R. (2004). An introduction to prior information derived from probabilistic judgements: elicitation of knowledge, cognitive bias and herding. *Geological Society, London, Special Publications*, *239*(1), 15-27.

Baecher, G. B. (1988). Judgemental probability in geotechnical risk assessment. *The Office of the Chief, US Army Corps of Engineers, Tech. Rep*.

Binns, P., & Corbett, P. (2012). Risk and uncertainty from frontier to production–A review. *First Break*, *30*(6), 57-64.

Bond, M., & Carragher, P. D. (2018, June). Addressing the causes of uncalibrated predictions & underperformance in oil & gas ventures. In *80th EAGE Conference & Exhibition 2018 Workshop Programme* (pp. cp-556). European Association of Geoscientists & Engineers.

Brown, M. L., Fosvold, L., Garza, A. J., & Cook, D. M. (2000). A look to the past to avoid old traps in the future. In *Norwegian Petroleum Society Special Publications* (Vol. 9, pp. 9-14). Elsevier.

Cox, R. T. (1946). Probability, frequency and reasonable expectation. *American journal of physics*, *14*(1), 1-13.

EBN B.V. (2019). Focus – Energie in Beweging.

Harper, F. G. (2000). Prediction accuracy in petroleum prospect assessment: A 15 year retrospective in BP. *Improving the exploration process by learning from the past: Norwegian Petroleum Society (NPF) Special Publication*, *9*, 15-21.

Heggland, K. (2000). Volumes before and after exploration drilling: results from the project: Evaluation of Norwegian Wildcat Wells (Article 2). *Improving the Exploration Process by Learning from the Past*, 33.

Hoetz, H.G. 2016 Observations from systematic depth conversion reviews and its impact on the drilling portfolio. *Conference Proceedings, 78th EAGE Conference and Exhibition 2016, May 2016, Volume 2016*, p.1 – 5

Hoetz, H.G., Ecclestone, M., van der Kraan, V. (2020). Drilling portfolio performance and the role of Survival Bias in volume estimates. *Pending abstract, 82nd EAGE Conference & Exhibition 2020, June 2020.*

Janssen, L. (2019). An offshore exploration drilling review. *Dutch Exploration Day 2019 presentation.*

Johns, D. R., Squire, S. G., & Ryan, M. J. (1998). Measuring exploration performance and improving exploration predictions – with examples from Santos' exploration program 1993 – 1996. *The APPEA Journal*, *38*(1), 559-569.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological review*, *80*(4), 237.

Lelliott, M. R., Cave, M. R., & Wealthall, G. P. (2009). A structured approach to the measurement of uncertainty in 3D geological models. *Quarterly Journal of Engineering Geology and Hydrogeology*, *42*(1), 95-105.

Lutgert, J., Mijnlieff, H., & Breunese, J. (2005, January). Predicting gas production from future gas discoveries in the Netherlands: quantity, location, timing, quality. In *Geological Society, London, Petroleum Geology Conference series* (Vol. 6, No. 1, pp. 77-84). Geological Society of London.

Mathieu, C. J. (2018, January). Exploration well failures from the UK North Sea. In *Geological Society, London, Petroleum Geology Conference series* (Vol. 8, No. 1, pp. 267-272). Geological Society of London.

Milkov, A. V. (2015). Risk tables for less biased and more consistent estimation of probability of geological success (PoS) for segments with conventional oil and gas prospective resources. *Earth-Science Reviews*, *150*, 453-476.

Milkov, A. V. Base Rate Neglect: A Common Logical Fallacy of Oil and Gas Explorers?. In *AAPG Annual Convention and Exhibition*.

Myers, K. (2018, June). Analogue play statistics for improved pre-drill risking: North Sea case study. In *80th EAGE Conference & Exhibition 2018 Workshop Programme* (pp. cp-556). European Association of Geoscientists & Engineers.

Norwegian Petroleum Directorate. (2018). Resource report 2018

Ofstad, K., Kittilsen, E. J., & Alexander-Marrack, P. (Eds.). (2000). *Improving the exploration process by learning from the past*. Elsevier.

Oil and Gas Authority UK (2015). Moray Firth–Central North Sea Post Well Analysis. *Oil in Gas Authority, United Kingdom*.

Quirk, D. G., & Ruthrauff, R. (2006). Analysis of reserves discovered in petroleum exploration. *Journal of petroleum Geology*, *29*(2), 125-146.

Quirk, D. G., & Ruthrauff, R. G. (2008). Toward consistency in petroleum exploration: A systematic way of constraining uncertainty in prospect volumetrics. *AAPG bulletin*, *92*(10), 1263-1291.

Quirk, D. G., Archer, S. G., Keith, G., Herrington, P., Ramirez, A. O., & Bjørheim, M. (2018). Can oil and gas exploration deliver on prediction?. *First Break*, *36*(10), 83-88.

Rose, P. R. (1987). Dealing with risk and uncertainty in exploration: How can we improve?. *AAPG bulletin*, *71*(1), 1-16.

Rose, P. R. (2007). Measuring what we think we have found: Advantages of probabilistic over deterministic methods for estimating oil and gas reserves and resources in exploration and production. *AAPG bulletin*, *91*(1), 21-29.

Rudolph, K. W., & Goulding, F. J. (2017). Benchmarking exploration predictions and performance using 20+ yr of drilling results: One company's experience. *AAPG Bulletin*, *101*(2), 161-176.