NHH

INSTITUTT FOR FORETAKSØKONOMI

DEPARTMENT OF BUSINESS AND MANAGEMENT SCIENCE

FOR 16 2017

ISSN: 1500-4066

November 2017

Discussion paper

Creaming - and the depletion of resources: A Bayesian data analysis

BY

Jostein Lillestøl AND **Richard Sinding-Larsen**

Norges Handelshøyskole

NORWEGIAN SCHOOL OF ECONOMICS

NHH

Norges Handelshøyskole

Norwegian School of Economics

NHH
Helleveien 30
NO-5045 Bergen
Norway

Tlf/Tel: +47 55 95 90 00
Faks/Fax: +47 55 95 91 00
nhh.postmottak@nhh.no
www.nhh.no

Creaming -
and the depletion of resources: A Bayesian data analysis

Jostein Lillestøl[1]

Norwegian School of Economics

Richard Sinding-Larsen[2]

Norwegian University of Science and Technology

Nov. 5, 2017 (Revised Feb. 13, 2018)

Abstract

This paper considers sampling in proportion to size from a partly unknown distribution. The applied context is the exploration for undiscovered resources, like oil accumulations in different deposits, where the most promising deposits are likely to be drilled first, based on some geologic size indicators ("creaming"). A Log-normal size model with exponentially decaying creaming factor turns out to have nice analytical features in this context, and fits well available data, as demonstrated in Lillestøl and Sinding-Larsen (2017). This paper is a Bayesian follow-up, which provides posterior parameter densities and predictive densities of future discoveries, in the case of uninformative prior distributions. The theory is applied to the prediction of remaining petroleum accumulations to be found on the mature part of the Norwegian Continental Shelf.

Keywords: Log-normal distribution, sampling proportional to size, resource prediction

[1] Department of Business and Management Science, Norwegian School of Economics, Helleveien 30, N-5045 Bergen, Norway; e-mail: jostein.lillestol@nhh.no

[2] Department of Geoscience and Petroleum, Norwegian University of Science and Technology, Sem Sælands veg 1, N-7491 Trondheim; e-mail: richard.sinding-larsen@ntnu.no

1. Introduction: Context and model

In the search for a valuable resource in a target area, the sizes of consecutive findings often have a declining pattern, consistent with a behavior where the most promising areas are explored first, so-called creaming. Among others, this may be so for the search for petroleum resources, as illustrated in Figure 1, showing the consecutive sizes, measured in million Standard cubic meter oil equivalents (MSm3o.e), of all N=182 development projects on the mature part of the Norwegian Continental Shelf (M-NCS), since the first drilling in 1966 and up to 2014.[3] An issue here is the likelihood of further major discoveries beyond the last one recorded.

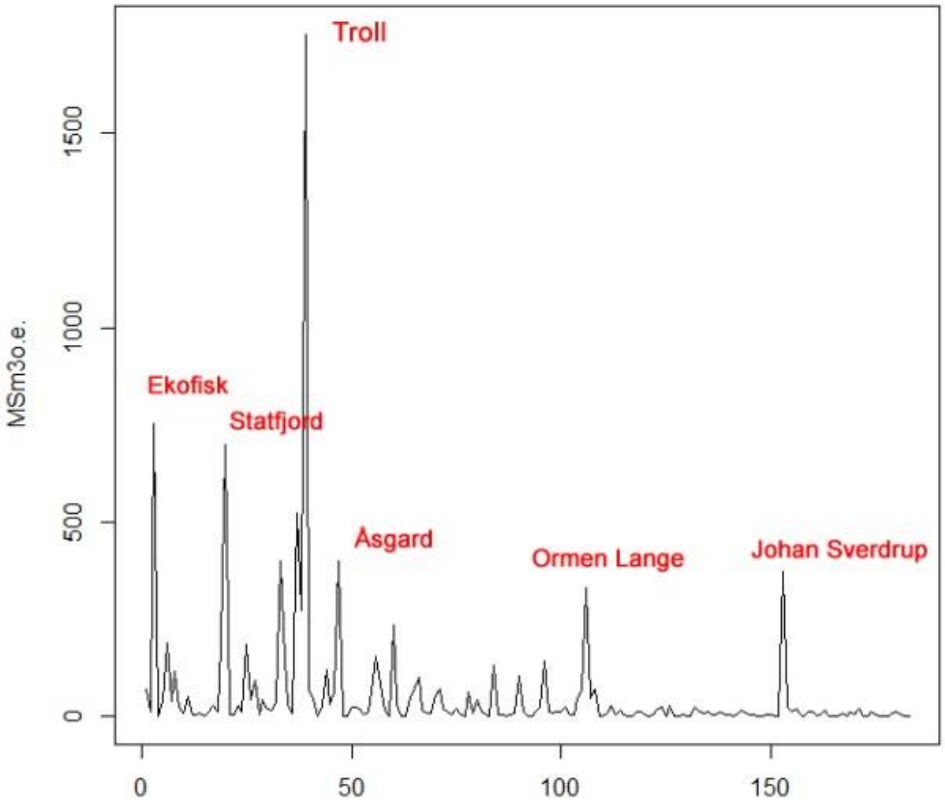

Figure 1. Consecutive sizes of 182 development projects on M-NCS.

[3] By the mature part of the Norwegian Continental Shelf is meant The North Sea and The Norwegian Sea, excluding Jan Mayen and other frontier areas. The source for this data is The Norwegian Petroleum Directorate (NPD), and represents updated values from the petroleum resource account as of 31.12. 2013, prepared as input to the Revised National Budget (RNB 2014).

Size distributions are often considered to be lognormal, preferably supported by the data, but sometimes just because it is analytically convenient. In the current context, the underlying size distribution may be different from the sampled one, since the sampling obviously is not independent identically distributed (i.i.d.) sampling. The conventional way of analyzing this is to assume sampling from a finite population, where the sampling is performed with probabilities proportional to size, and where the population is depleted over time, see Kaufman et. al. (1975). This requires an assumption of the population size N, which is hard to estimate with any precision, and may strongly affect the conclusions. An alternative way is to assume independent sampling, consistent with an infinite population, but with declining size expectations. This provides a basis for easier, but still realistic, probability calculations. Lillestøl & Sinding-Larsen (2017) suggest a model for doing this, which is supported by the data. The model is consistent with lognormal population and lognormal sampling at any consecutive stage in the sampling process. The model is as follows:

Assume that size observations S_t are indexed consecutively by t=1, 2, 3, ... , named time. They appear as coming from a distribution

$$\text{lognormal } (\mu(t),\ \sigma) \text{ where } \mu(t) = \mu + k(t) \cdot \sigma^2$$

and $k(t)$ tending to zero as t tends to infinity. A convenient choice is $k(t) = \alpha \cdot e^{-\beta t}$, thus giving a model with four parameters $(\mu, \sigma, \alpha, \beta)$ to be estimated from the data.

The model is consistent with interpreting μ as the expectation of log-size of the parent distribution and sampling proportional to $x^{k(t)}$ at time t. A decaying creaming function $k(t)$ then reflects that the creaming opportunity gradually diminishes. When t tends to infinity, we get at the limit i.i.d. observations, as if we are sampling from the parent population itself without creaming. Now we cannot expect to make major discoveries, since the parent distribution has a very thin right tail and the opportunity to cream is gone. A key model feature is the constant σ, and how it is linked to both variance and expectation of log-sizes. This is not as weird as seen by first sight, since the formulas for expectation and standard deviation of size itself, having a lognormal$(\mu,\ \sigma)$ distribution, are

$$E(S_t) = e^{\mu+\sigma^2/2} \text{ and } \sigma(S_t) = ES_t\sqrt{e^{\sigma^2} - 1}$$

The coefficient of variation is $\frac{\sigma(S_t)}{E(S_t)} = \sqrt{e^{\sigma^2} - 1}$, and the skewness is $(e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}$, both independent of μ. Moreover, we note that

$$Mode(S_t) = e^{\mu-\sigma^2} < Median(S_t) = e^{\mu} < Mean(S_t) = e^{\mu+\sigma^2/2}$$

For our model, we have to replace μ by $\mu(t) = \mu + k(t) \cdot \sigma^2$, which amounts to adding $k(t) \cdot \sigma^2$ in each of the exponents, and then the dependence on σ^2 will appear in the median as well.

The theoretical basis for the model may be challenged, but the model may be useful anyway, since it is both simple and may be flexible enough to pick up the essential features of the data. Further details on the context and the creaming model is given in Lillestøl and Sinding-Larsen (2017), and are not repeated here. Our approach may be seen as an alternative to the two-stage process model attributed to Kaufman, where the depletion of resources is represented by sapling proportional to size from a finite population, coming from a log-normal superpopulation. Important references to this approach are Barouch and Kaufman (1967), Kaufman et. al. (1975), Andreatta and Kaufman (1986), Lee and Wang (1986), Sinding-Larsen and Xu (2005), Kaufman et.al. (2016). The creaming method and the basic lognormal property was first demonstrated by Meisner and Demirmen (1981). That paper takes a Bayesian approach different from ours. An annotated bibliography of methodology for assessment of undiscovered petroleum resources are given by Charpentier et.al (1995).

2. Bayesian estimation

We will do Bayesian estimation of this model for the M-NCS data above, using a common Markov chain Monte Carlo (MCMC) method with non-informative priors. Common algorithms will provide samples from the joint posterior distribution of $(\mu, \sigma, \alpha, \beta)$: $(\mu_i, \sigma_i, \alpha_i, \beta_i)$ for i=1, 2, 3, …, n. This will be the basis for the construction of corresponding sequences for functions of these parameters, among them the creaming function $k(t)$, the expectation of log-size $\mu(t)$, and the expectation of size $E(S_t)$, according to the formulas given in Section 1. From these sequences, we can estimate their posterior distributions (by smoothed histograms), estimate their posterior expectations (by averaging) and estimate other posterior characteristics, say mode, median or quantiles accordingly. These four sequences will also be the basis for the simulation of predictions beyond t=182, from which predictive distributions and associated probabilities and expectations can be estimated.

We use the MCMC algorithm implemented in OpenBUGS, made accessible directly from R (ref. R Core Team, 2013) by the program R2OpenBUGS (ref. Sturtz et. al., 2005; Yan & Ptates, 2013). BUGS models are expressed in terms of the precision $\tau = 1/\sigma^2$, and the expected log-sizes are then written as

$$\mu(t) = \mu + \alpha \cdot e^{-\beta t} \cdot \frac{1}{\tau}.$$

The four parameters $(\mu, \tau, \alpha, \beta)$ will be assigned independent priors according to no specific prior knowledge (ignorance). BUGS requires proper distributions to represent non-informative cases, which is obtained by priors with large variance, i.e. small precision. As prior for μ is chosen a normal distribution with low precision, and for each of the non-negative parameters (τ, α, β) is taken a gamma distribution with expectation 1 and variance 1000. Exhibit 1 is then our BUGS model, with N being the number of observations (N=182). Note that

we have temporarily rescaled β by a factor of 100, in order to have the parameters of the same order of magnitude at the computing stage.

```
model
{
  for( i in 1 : N ) {
   y[i] ~ dlnorm(m[i],tau)
   m[i] <- mu+alpha*exp(-0.01*beta*i)/tau
   }
  mu ~ dnorm(0.0,1.0E-6)
  tau ~ dgamma(0.001,0.001)
  alpha ~ dgamma(0.001,0.001)
  beta ~ dgamma(0.001,0.001)
  sigma <- 1 / sqrt(tau)
       }
```

Exhibit 1. BUGS model

The BUGS-algorithm imitates sampling from the joint posterior distribution of $(\mu, \tau, \alpha, \beta)$, using Markov chain Monte Carlo methods. This is done iteratively and requires a "burn-in" period, in order to secure stationarity. It turns out that the situation is rather challenging and requires care. This was expected from the structure of the model itself, and clearly exposed by a test run of four chains running in parallel, each with 50 000 iterations, where the first 25 000 iterations were discarded. In practical terms, it is difficult to discriminate between a favorable population with small opportunity for creaming and a less favorable population with larger opportunity for creaming. More on this later.

In Exhibit 2 follows the BUGS call from R and the summary output from R2OpenBUGS for a representative run of 20 000 iterations, from which the 10 000 first were discarded, leaving 10 000 simulated realizations $(\mu_i, \sigma_i, \alpha_i, \beta_i)$, i=1, 2, …, 10 000, from the joint posterior distribution. This is sufficient for our purpose.

```
my.sim <- bugs(data, inits, model.file="BUGSmodel2a.r",
parameters=c("mu", "alpha", "beta", "sigma"), n.chains=1, n.iter=20000)

Inference for Bugs model at "BUGSmodel2a.r",
Current: 1 chains, each with 20000 iterations (first 10000 discarded)
Cumulative: n.sims = 10000 iterations saved
            mean   sd    2.5%     25%     50%     75%   97.5%
mu         -1.66 1.80   -5.75   -2.70   -1.32   -0.34    1.13
alpha       2.16 0.72    1.04    1.63    2.03    2.58    3.92
beta        0.43 0.23    0.16    0.27    0.38    0.52    1.04
sigma       1.61 0.08    1.46    1.55    1.60    1.66    1.77
deviance 1562.46 2.66 1559.00 1561.00 1562.00 1564.00 1569.00

DIC info (using the rule, pD = Dbar-Dhat)
pD = -1.29 and DIC = 1561.00
DIC is an estimate of expected predictive error (lower deviance is better).
```

Exhibit 2. Output from run R2OpenBUGS

The output in Exhibit 2 provides the estimated mean, standard deviation and five useful quantiles of the posterior distribution for each of the four specified parameters, as well as some additional diagnostic information. Hereafter, beta is rescaled by 0.01 to its real value, so that its mean value is 0.01x0.43=0.0043. [4]

The posterior densities of $(\mu, \sigma, \alpha, \beta)$ obtained by smoothing their histograms for the 10 000 simulations are displayed in Figure 2.

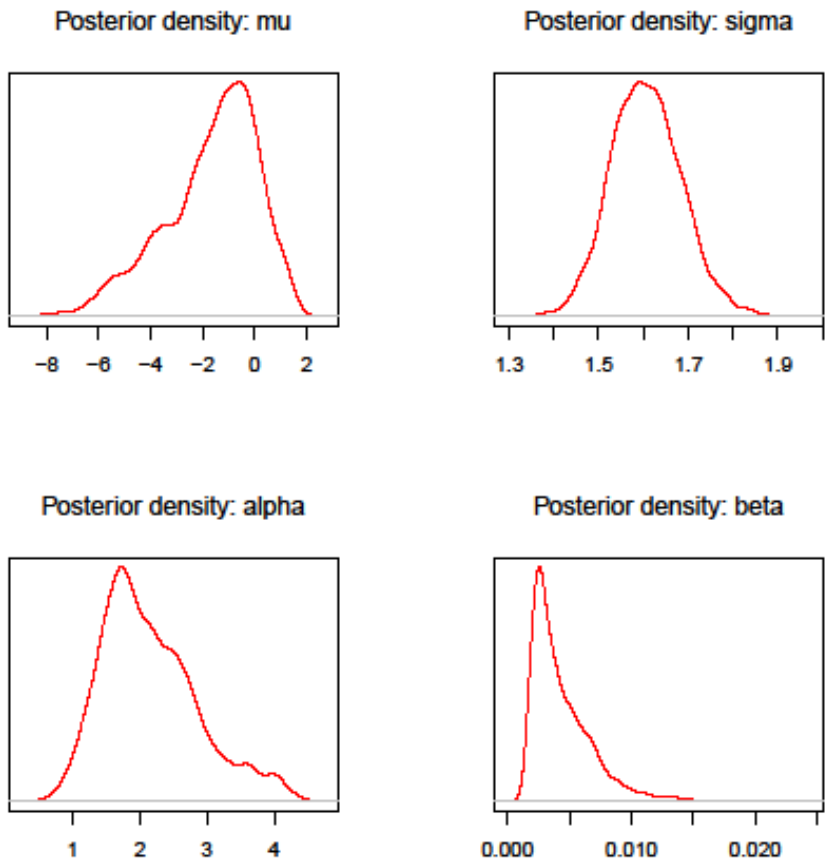

Figure 2. Posterior densities of parameters $(\mu, \sigma, \alpha, \beta)$

The mode of these densities are respectively -0.58, 1.60, 1.73 and 0.0026. The posterior distribution of sigma seems to be nearly symmetric, and is the most stable parameter in repeated computations. The posterior distribution of mu and alpha seem to be moderately skew, with the longest tail to the left and to the right respectively. The tail of these distributions seems to be mirror images of each other, due to statistical dependence between mu and alpha (see later Figure 3), and have humps that indicate polymodality. They both vary

[4] A prior version of this paper contained a numerical error. In fact, a simulated beta-vector was mismatched, i.e. belonged to a different run. This affected some of the insight and conclusions. The necessity to handle a fat-tail problem is now removed, giving similar final predictions to the ones reported earlier.

a lot in repeated computations, which seems disturbing for the purpose of individual parameter estimation. The posterior distribution of beta is also skew with a long upper tail. This parameter governs the rate of decay, and its variability will be important for predictions. It is therefore unfortunate, that this parameter also is very sensitive in the estimation process. However, for the prediction of future discovery sizes, the situation is not that bad, since posterior correlations between the parameters tend to balance individual estimation errors in the log-expectation formula.

In Figure 3 follows six bivariate scatterplots for a sample of 200 observations from the joint posterior distribution. We see that alpha correlates negative with mu (-0.95), beta (-0.70) and sigma (-0.28), and that beta correlates positive with mu (0.78), and the correlations between sigma and mu (0.01), and between sigma and beta (0.04) are minor. The correlations (in parenthesis) are computed from all 10 000 observations, not just the 200 in the graphs.

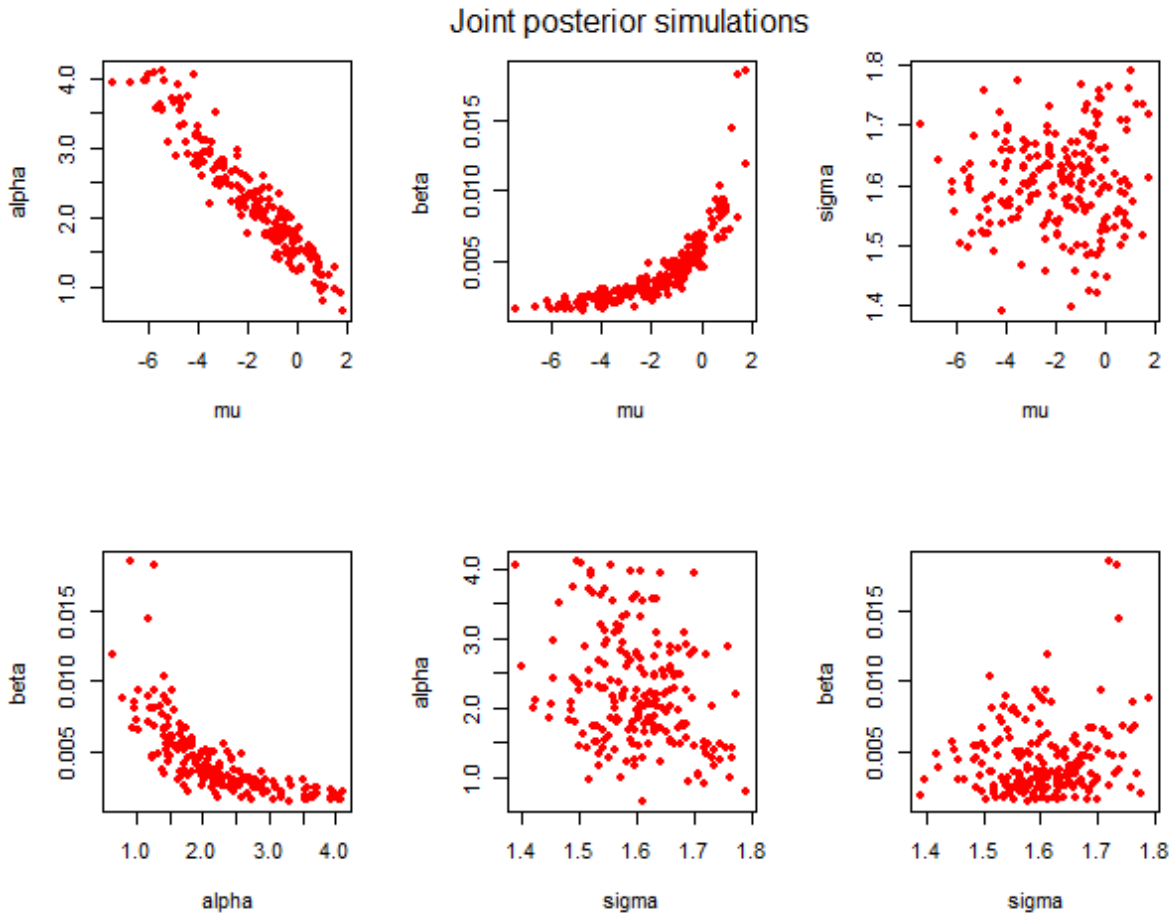

Figure 3. Bivariate scatterplots for a sample of 200 observations from the joint posterior distribution

Looking at the expectation formula, we see that mu and alpha occur linearly, and a deviation from expectation for one of them, is likely compensated by the other, due to the strong negative correlation. For beta we see a more delicate mechanism that involves both mu and

alpha. For instance, a too large beta (fast decay) is likely compensated by a larger mu and smaller alpha, the latter giving less weight to the exponential decay factor.

Remarks. The graphs may seemingly indicate near functional relations between parameters, and thus offer the opportunity for parameter reduction. However, this will destroy the transparency of the model and the opportunity to separate the effects of (dis)favorable population and creaming. The displayed correlation can also throw some light on the estimation process, the difficulties and their relevance for prediction. We have seen that sigma is very stable with posterior mean about 1.60 in repeated runs, while alpha and beta may vary considerably from run to run. However, the variation goes in opposite directions, and may provide the same expectation and similar goodness of fit. The situation is due to a "flat" likelihood-function in one or more directions. For instance, imagine sigma and beta fixed. Then the likelihood is about constant along a linear path in (μ, α)- space. In the non-Bayesian maximum likelihood context, it is like walking on a path at a mountain rim, in search for global maximum, hard to find. However, when the estimation process offers several seemingly different prediction formulas with similar good properties it is, from a mathematical-statistical point of view, immaterial which representative is chosen. From a geoscience point of view it will matter, since some (mu, sigma) combinations may have geologically impermissible consequences, as will be discussed in section 4.

Recall the expressions for mode, median and mean for the size distribution in terms of the parameters $(\mu, \sigma, \alpha, \beta)$ at any instant t, past or future. Consider them as functions of t:

$$Mode(S_t) = e^{\mu(t)-\sigma^2},\ Median(S_t) = e^{\mu(t)},\ Mean(S_t) = e^{\mu(t)+\sigma^2/2}$$
$$\text{with } \mu(t) = \mu + \alpha \cdot e^{-\beta t} \cdot \sigma^2$$

For each t, each of these characteristics have their own posterior distribution, which may also be established from our simulations, by plugging in the simulated $(\mu_i, \sigma_i, \alpha_i, \beta_i)$, i=1, 2, …, 10 000. For each t, a smoothed histogram version of the distributions may be displayed as above. If we want to present the distributions in the time context, we may turn to so-called fan-plots (not shown here). We are here satisfied with just the mean characteristic of these distributions, which is taken as posterior estimates of the mode-, median- and mean-function given above. This is obtained by averaging the 10 000 calculated quantities obtained from each use of one of the formulas above. It should, however, be noted that the arithmetic mean characteristic of these distributions are strongly influenced by the extreme values in the tails.

From the formulas, we see that the ratio between any two of the three measures does not depend on t, and depends only on sigma. In fact, starting with the median, we have approximately

Mean=3.68 x Median Mode= 0.0776 x Median.

This means that the three functions only differ in their scale. Figure 4 displays the function and the scale for mean (i.e. expected) size and median size.

Exhibit 3 shows the numerical values of the estimated mode, median and mean at some selected instants: At the beginning (t=1), for the development project following the last one observed (t=N+1), and then at t=N+10, N+50, N+100 and N+200. The corresponding estimated means (i.e. expectations) of the normal log-size $\mu(t)$ are added as last row of the table. Note that this is about estimation of a parametric function, and not the characteristics of the distribution of new observations at the given instants. They are conceptually different, but coincide by theory for the mean and mean-log, but not for the median and the mode (see later).

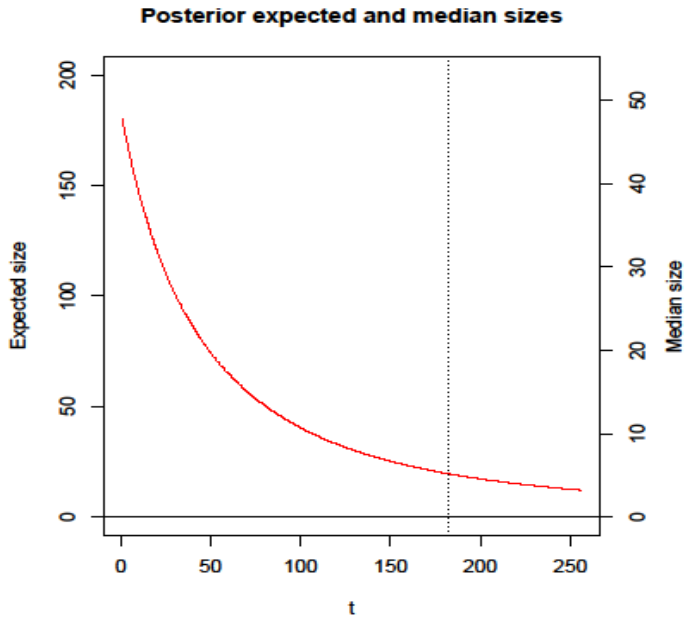

Figure 4. Estimated expectation and median function of t (vertical line at t=N=182)

Parameter	t=1	t=N+1	t=N+10	t=N+50	t=N+100	t=N+200
Mode	3.8	0.28	0.26	0.19	0.14	0.09
Median	49	3.7	3.4	2.5	1.8	1.2
Mean	180	13.4	12.4	9.0	6.5	4.3
Mean-Log	3.85	1.27	1.19	0.85	0.49	- 0.01

Exhibit 3. Posterior estimates of mode, median and mean of the log-normal distribution at instants t=1, N+1, N+10, N+50, N+100, N+200.

Remark. It may be tempting to do this calculation in reverse order, that is, use the parameter estimates of each of the four parameters, obtained by averaging the simulations, and then plug this into the expectation formula. However, this approach will depend on the choice among different parameterizations, and may turn out very different results. For instance, since $(mean(\sigma_i))^2 > 1/\ mean(\tau_i)$, averaging over sigma will give a larger result than averaging over tau. The issue may occur in any Bayesian model estimated by posterior simulations, the log-normal model with its exponential expectation being a vulnerable case. Some ambiguities, like the one above, may be artificially resolved by using the geometric mean instead of the arithmetic mean. [5]

A posterior estimate of the creaming function $k(t) = \alpha \cdot e^{-\beta t}$ may be obtained similarly for each t, by averaging over the 10 000 simulations. The obtained creaming function is given in Figure 5.

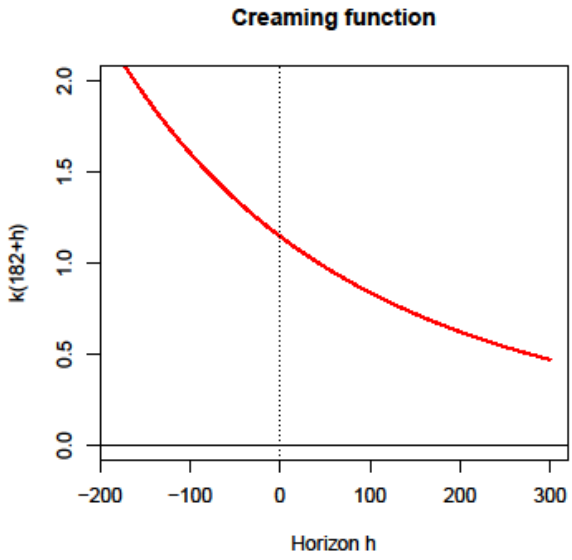

Figure 5. Creaming function estimated by posterior means (vertical line at N=182)

Here the vertical line is the division between the past, for which we have the data ranging back to -182, and the future, with horizons h up to 300. Initially (at h=-182) the expected creaming factor is about two, i.e. sampling proportional to square of size, and then declines close to sampling proportional to size h=0 and down to sampling proportional to square root of size at about h=200. It seems that the creaming opportunity and/or ability is meager beyond a horizon of 300. For the non-Bayesian analysis of this data in Lillestøl and Sinding-Larsen (2017), the horizon H=256 was chosen to give a reasonable match to the whole population in the finite population context. Figure 5 indicates that the horizon might be slightly extended. However, it is not unreasonable, and is kept here for comparison purposes. The choice of

[5] For the non-Bayesian modelling of the problem with parameters fixed, but unknown, and estimated by maximum likelihood, we do not meet the same problem. Instead, we have the issue of bias.

horizon for prediction and decision purposes will be discussed in some detail in a separate section.

It is of some interest to explore in some more detail the dependencies in the posterior solution, in particular how sigma may affect the expectations in the predictive context. Moreover, sigma is a determining parameter for the upper tail of the predictive distributions. In Figure 6, we exhibit bivariate scatterplots of 200 simulated values at t=N+1=183 for each of the creaming factor, log-size expectation and size expectation against its corresponding sigma. These are the same runs as those serving as basis for Figure 3 extended past t=N, which also represent the basis for the predictive evaluations in section 3.

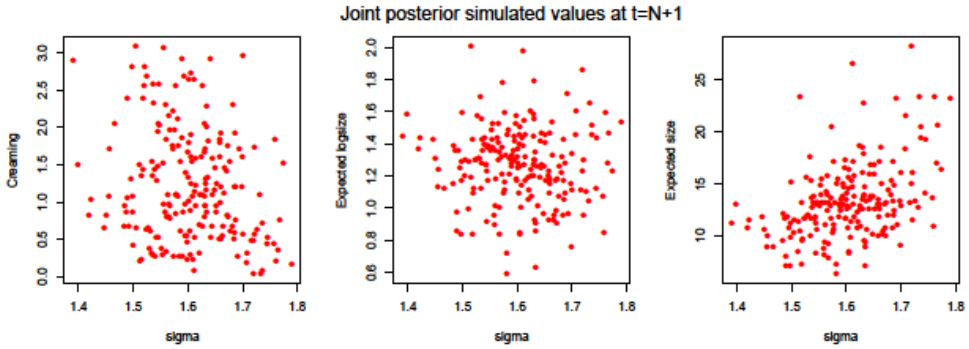

Figure 6. Bivariate scatterplots for posteriors of creaming, log-size expectation and size expectation at t=N+1=183 versus sigma for the sample of 200 observations from the joint posterior distribution.

We see that there is apparently slight negative correlation between sigma and the creaming factor (-0.21), a favorable compensation feature for the additive term of the expectation formula. There is a weak negative correlation between sigma and the log-size expectation (-0.07) and moderate positive correlation between sigma and the size expectation (0.43). The latter is not surprising, since $\sigma^2/2$ adds to the log-size expectation in the exponent (see formula in section 1). However, it may perhaps come as a surprise that the log-size expectation itself does not have a stronger dependence on sigma in the posterior predictive context, since σ^2 appear as a multiplicative factor to the creaming function. The explanation to this is the outweighing effect of alpha shown in Figure 3. As increasing sigma also widens the likely upper range of discovery sizes, there may be an inherent danger over-prediction here. This will be explored further in a later section.

It is of some interest to compare the Bayesian solution with the non-Bayesian in Lillestøl and Sinding-Larsen (2017), based on classic frequentist ideas and maximum likelihood estimates. They are given in Exhibit 4.

Exponential fit: mu, sigma, alpha, beta

Parameter	Estimate	Std.error
mu	-0.298	1.075
sigma	1.592	0.083
alpha	1.690	0.467
beta	0.0055	0.0021

Exhibit 4. Parameter estimates for the non-Bayesian model

The Bayesian posterior means of these parameters were mean-mu=-1.66, mean-sigma=1.61, mean-alpha=2.16, mean-beta=0.0043. We see that these values deviates less than two standard errors from the maximum likelihood values. In fact, sigma is very close, beta within one standard error, alpha at one standard error and mu at about 1.5 standard error away from the values given in Exhibit 4. However, as expected with non-informative priors the modes of the posterior distributions are close to the maximum likelihood estimates.

We also see that the Bayes-estimates are smaller for mu and higher for alpha than the corresponding maximum likelihood estimates. This conforms well to the compensation feature mentioned above. The estimated creaming function turned out lower in the non-Bayesian case, starting at 1.6 at h = -182 and down to 0.6 at h=0. On the other hand, the Bayes estimate of mu is less (i.e. more negative). This means a slightly different emphasis. The Bayesian solution see a less promising population, but a stronger opportunity for creaming. It is therefore not obvious from this, which one of the two solutions will turn out the most optimistic predictions.

Some end remarks on the model and priors: An alternative parameterization is

$$\mu(t) = \mu(1 - \gamma \cdot e^{-\beta t} \cdot \frac{1}{\tau}).$$

A meaningful expression requires mu and gamma to have opposite signs. The most likely value of mu is negative, and therefore gamma positive. In this case, $\gamma e^{-\beta t}/\tau$ is the relative raise in expected log-size at time t due to creaming, which at the start is γ/τ. Relative to the precision, the alternative parameters γ and α play the roles of measure of increased log-size expectations in relative and absolute terms respectively. This model is identical to our previous model by setting $\alpha = \gamma \cdot \mu$. If we use the alternative parameterization, and assign independent priors on $(\mu, \tau, \gamma, \beta)$, this will force μ and α to be negatively correlated, unless the expectation of γ is taken to be zero. Thus, non-informative priors in one of the two parameterization implies informative dependence between two parameters in the other one.

3. Bayesian prediction

From the n=10 000 simulated values of $(\mu, \sigma, \alpha, \beta)$ we may now generate future values according to the model at time N+h, starting at N=182, for different horizons h=1, 2, 3,…, H. This means that

$$\log(S_{N+h}) = \mu + \alpha e^{-\beta(182+h)} \cdot \sigma^2 + \sigma \cdot z_h$$

where $z_h, h = 1,2, \dots, H$ are distributed independent standard normal. We now get 10 000 simulated predictions of S_{N+h} by plug-in of $(\mu_i, \sigma_i, \alpha_i, \beta_i)$ and z_{hi} for i=1, 2, …, 10 000. In fact, we made 10 repeats for each of the 10 000, so that we have 100 000 simulated predictions as basis for computing predictive distributions, by smoothing the histograms. In Figure 7 follows a plot of the predictive distribution of size at t=N+1, i.e. development project 1 after N=182. As expected, the distribution is very peaked at low values, but has a fairly long upper tail.

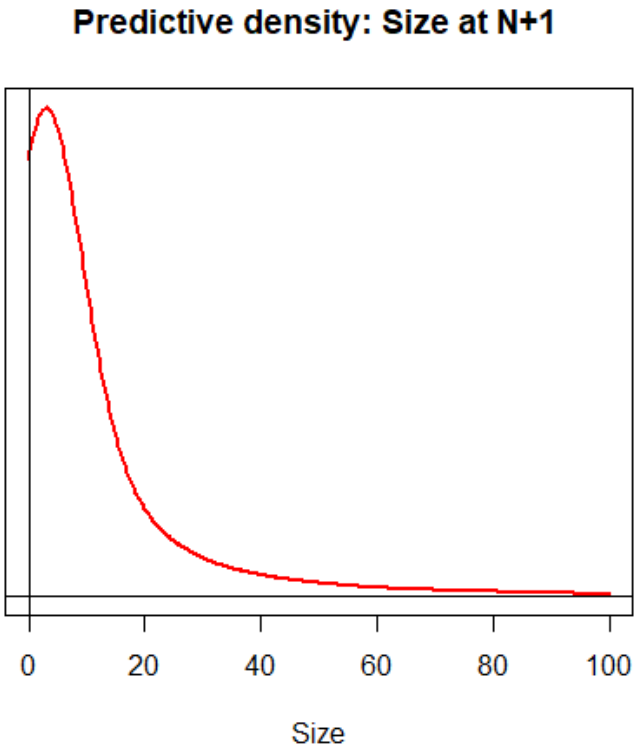

Figure 7. Predictive density of discovery size at time t=N+1 for N=182.

The characteristics median and mean for the predictive distributions of size at instants t=N+1, N+10, N+50, N+100, N+200 and N+256 are obtained by averaging the 100 000 simulated values, and are given in Exhibit 5. We see that the obtained predictive density for t=183 has median=3.6 and mean=13.5, which may be taken as point estimates of the corresponding distribution median and expectation at t=183.

Characteristic	t=N+1	t=N+10	t=N+50	t=N+100	t=N+200	T=N+256
Median	3.6	3.3	2.4	1.6	1.0	0.7
Mean	13.5	12.6	9.2	6.6	4.1	3.4

Exhibit 5. Median and Mean of the predictive distribution of Size at instants t=N+1, N+10, N+50, N+100, N+200 and N+256.

As a check of this result, one may refer to Exhibit 3, which gave posterior estimates of the median- and mean-function expressed by the four parameters $(\mu, \sigma, \alpha, \beta)$. The two exhibits show conceptually different quantities. However, the means should be theoretically identical, but may differ slightly numerically. The median does not possess this property, and the relevant median estimate for the prediction context is the one given here. However, we see that the values in Exhibit 3 and Exhibit 5 are close for both the mean and the median.

The predictive distribution at different horizons may be presented graphically in a so-called fan-chart. Here the quantiles (including the median) appear for different horizons in a time context (not shown here). Instead we present Figure 8, which provides some insight to how the predictive distribution extends the actually observed discoveries beyond t=N=182. Here the curves for mean and median of the predictive distribution of sizes are drawn for t ranging up to N+H=182+256=438.

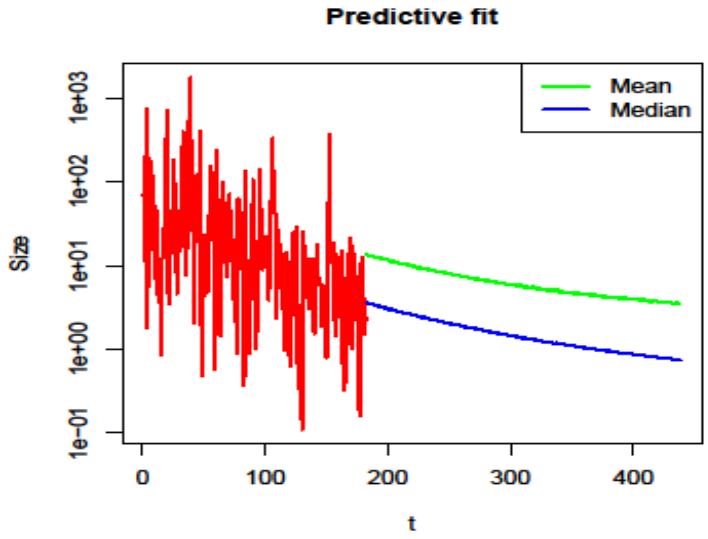

Figure 8. Predictive fit on log-scale: The actual observed sizes (t=1, 2,…, N=182) and the curves for mean and median of the predictive distribution of sizes (t=N+h, for h=1,2,…, 256).

We see that the median curve extends the data nicely, and has a slight concave curvature.[6] The mean curve is seemingly out of bound. However, there may be a reasonable explanation for this. The Bayesian approach, with sigma random, opens up for values on both sides of a

[6] Note also that the gap between these two posterior predictive curves widens with time. This is not contrary to the model itself, which implies a constant ratio between the mean and the median for fixed parameters.

fixed value. This will have different effects on the infinite upside than the downside limited by zero. The obtained predictive density is in theory a mixture of log-normal distributions and will have heavier tail than any of the single log-normal distributions involved. As a mixture, this may be contra-intuitive, but may happen when covariation of the expectation and variance is positive. The mean is strongly affected by heavy tails, and it is no surprise that it will be inflated as seen.

The lesson here is that the useful key characteristic in the current context is not the mean (i.e. expectation), but the median. In terms of prediction error judged by a loss-function, the median is optimal for linear symmetric loss, and the mean optimal for symmetric squared loss. In the current context, it is more reasonable to imagine asymmetric loss. It is more unpleasant to over-estimate the resources than under-estimate them. This favors the median over the mean. In this respect, the mode will also be a viable alternative.

An interesting question is whether it is approximately log-normal for some fixed parameter set, consistent with the non-Bayesian conception of log-normal size distribution for any t. We have performed quantile-quantile plots from the simulated predictions (based on the simulated varying parameters) for t=N+1 and some other t's. They all show points almost in straight line, confirming the approximate log-normality with a sigma larger than the point estimate in the Bayesian model.

We will now study in probability terms what is possible to achieve within a given horizon: short and intermediate horizons like H=10, 25 and 50, and longer horizons like 100, and in particular horizon H=256. The probabilities of at least one development project above given sizes within horizons H=10, 25, 50, 100 are displayed in Figure 9.

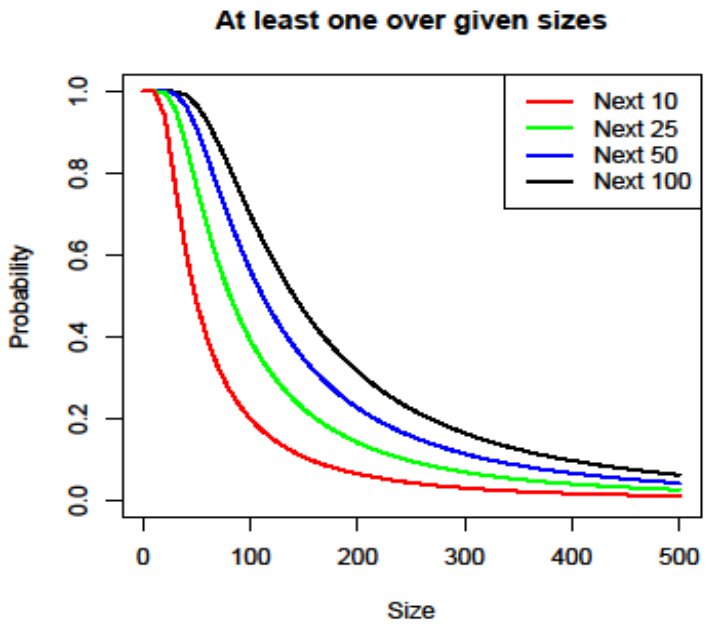

Figure 9. The probability of at least one above given sizes within horizons H=10, 25, 50, 100.

Some actual numbers for some selected size levels, among them that of a giant field (79 Msm3o.e.) and Johan Sverdrup (373.84 Msm3o.e.), for horizons H=10, 25, 50, 100 and 256 are given in Exhibit 6.

Size level	10	20	50	100	200	Giant	Johan Sverdrup
H=10	0.96	0.82	0.44	0.20	0.07	0.24	0.02
H=25	0.99	0.96	0.66	0.35	0.13	0.45	0.04
H=50	0.999	0.99	0.84	0.51	0.21	0.63	0.07
H=100	1.000	0.999	0.92	0.64	0.29	0.76	0.11
H=256	1.000	1.000	0.96	0.74	0.38	0.84	0.15

Exhibit 6. Probabilities of at least one discovery over given size levels 10, 20, 50, 100, 200, Giant (79) and Johan Sverdrup (373.84) within horizons H=10, 25, 50, 100, 256.

These probabilities may be compared with the ones obtained from the non-Bayesian analysis in Lillestøl and Sinding-Larsen (2017). There the probabilities of at least one giant field at horizons H=10, 25, 50 were 0.22, 0.44 and 0.63, respectively. The probabilities of observing another field of size at least that of Johan Sverdrup at horizons H=10, 25, 5, were 0.02, 0.03 and 0.06. We see that the non-Bayesian probabilities match almost perfectly with the corresponding Bayesian ones in Exhibit 6.

The predictive distribution of maximum discovery size within a given horizon may be obtained from our simulations as well. Figure 10 is a fan-plot, showing how some characteristics of the distribution increases with the horizon. It pictures the 50% curve (the median), the 25% and 75% curves (the quartiles Q25 and Q75) and the 10% and 90% curves (our Q10 and Q90). We see how the quartile range [Q25, Q75] and the range [Q10, Q90] increase with the horizon.

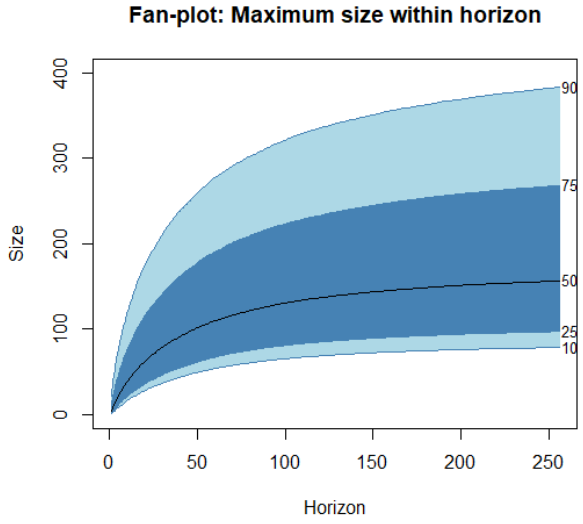

Figure 10. Fan-plot of the maximum discovery size within a given horizon.

We see that the median of the distribution increases rapidly for horizons up to around 50. Then it fades off gradually, and seems to be limited by size 150. It is still some increase beyond 100, but hardly noticeable beyond horizon h=200. The mean will have a similar curve above the median curve, seemingly limited by size 200. These curves are bend down due to the decaying expectations in additional discoveries, over which the maximum is taken. Note that the distribution of maximum size will have a heavy upper tail, and for those the median is more relevant than the mean.

The predictive density of the magnitude of total aggregated resources within horizon H=256 is given in Figure 11. This is done by aggregating the simulations of for h=1, 2, …, 256, and then smooth the histogram of the resulting 100 000 obtained values. This distribution has mean=1665, which may be taken as a point estimate of expected magnitude of total aggregated resources.[7] Quantile characteristics of the distribution are given in Exhibit 7. The 10% quantile Q10 and the 90% quantile Q90 are marked by vertical dotted lines in Figure 11.

Quantile	2.5%	5%	10%	25%	50%	75%	90%	95%	97.5%
Total	635	722	843	1091	1460	1980	2699	3302	3966

Exhibit 7. Quantile characteristics of distribution of total aggregated resources within horizon H=256

[7] Expected total aggregated resources within horizon H may be calculated by the sum of the consecutive expectations. As a check, one could plug-in of point estimates in the expectation formula and add over the horizon. As indicated above, this may not give the correct result.

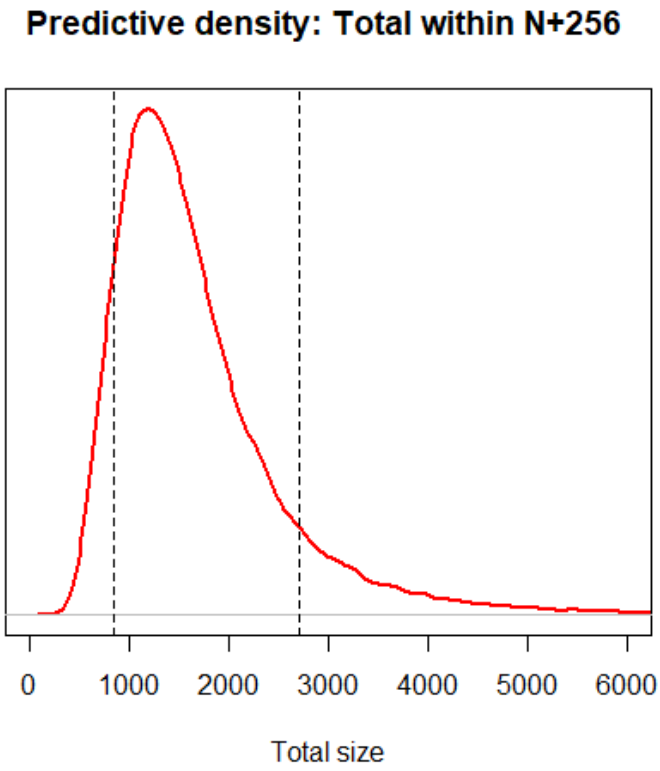

Figure 11. Predictive density of total aggregated resources within horizon H=256

The fan-plot in Figure 12 demonstrates the distribution of total aggregated resources for different horizons. It pictures the median curve and the quantile curves (Q10, Q25, Q75, Q90) for horizons h ranging up to H=256. We see how the quartile range [Q25, Q75] and the range [Q10, Q90] increase with the horizon.

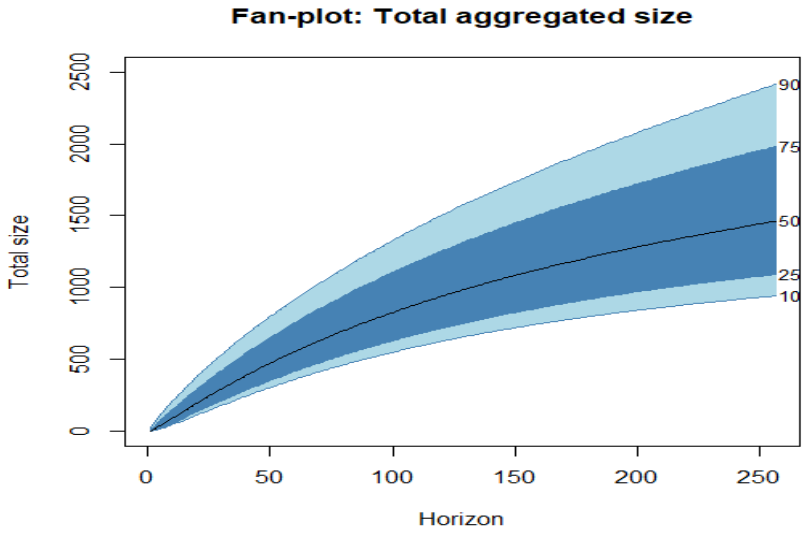

Figure 12. Fan-plot of predictive densities of total aggregated resources within horizons.

Predictive simulations based on our log-normal exponential model using maximum-likelihood estimates in the log-expectation formula are performed in Lillestøl & Sinding-Larsen (2017). In the case of H=256 we obtained mean=1665, median=1603 and quantiles Q10=1129 and Q90=2213. We see that the mean is head on, while the median is smaller in the Bayesian case. This reflects a distribution with long tail towards right. Moreover the quantile ranges are wider. This non-Bayesian calculation reflects the expected decay and the log-normal randomness, taking the estimated parameter values as given. With added parameter uncertainty, we expected this range widened, since the Bayesian approach takes this parameter uncertainty explicitly into account. The results is in line with the typical experience that non-informative priors typically leads to posterior point estimates close to the maximum likelihood solution.

It is also possible to answer question about how large portion of the total, the few large discoveries will account for. In Figure 13 and Figure 14 show this in two ways, for the five largest discoveries within horizons h=5, 6, …, 256. Figure 13 gives the expected aggregated size of the five largest discoveries pictured together with the total aggregated size. The five largest curve bends down rapidly and at horizons 50, 100, 200 and 256, the levels are 506, 689, 840 and 880 respectively. Figure 14 gives the corresponding share of the top five among the total aggregated size. At horizons 50, 100, 200 and 256, the shares are 58%, 45%, 35% and 32% respectively.

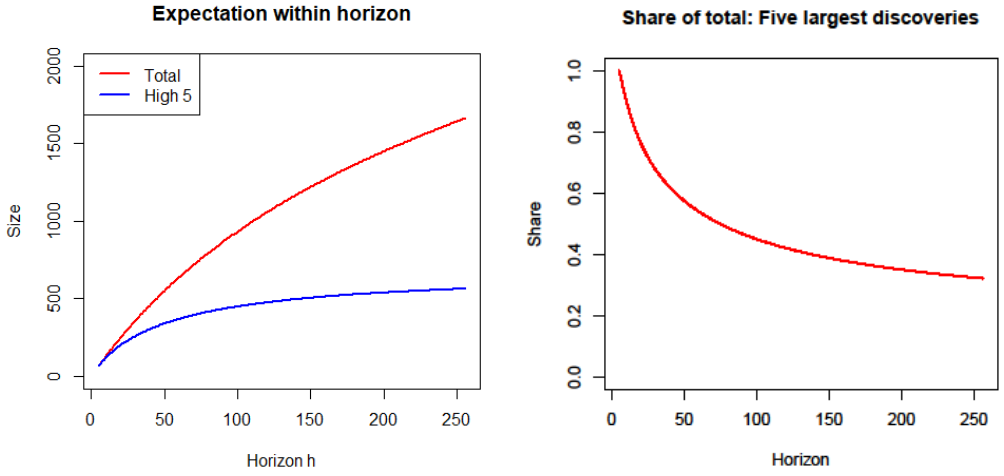

Figure 13. Expected size of five largest discoveries within horizon h=5, 6, …, 256.

Figure 14. Expected share of total for the five largest discoveries within horizon h=5, 6, …, 256.

The Bayesian solution in this section offers a fair comparison with the numbers in the reports of the Norwegian Petroleum Directorate (NPD). The predictive distribution for the totals derived from the NPD-report has mean about 1575, in comparison with our mean=1665.[8] The

[8] Note that the horizon H=256 was selected in Lillestøl and Sinding-Larsen (2017) to match the obtained estimate of expected total aggregated resources with the NPD expectation equal to 1665. Unfortunately, this

quantile characteristics are given in Exhibit 8, together with our corresponding quantile characteristics.

Quantile	0.5%	5%	10%	25%	50%	75%	90%	95%	99.5%
NPD-Total	695	891	993	1192	1461	1815	2272	2645	4396
Bayes-Total	500	722	843	1091	1460	1980	2699	3302	5665

Exhibit 8. Quantile characteristics of distribution of total aggregated resources within horizon H=256

There is an almost perfect match with the median, while our distribution has a slightly higher mean and a longer right tail. Recall that our results are dependent on the choice of horizon H=256. We can interpret a similarity of means and/or medians several ways. If we agree on the expected (or median) remaining resources, then we implicitly agree on the horizon as well. If we agree on the horizon, then we agree on the expected (or median) remaining resources as well. There is, of course the possibility that we disagree on both, and they outweigh each other. Having a range from Q10 to Q90 which is wider than those of NPD, may also be given different interpretations. It may reflect the parameter uncertainty, taken explicitly into account. It may reflect a potential brought forward by the data, not taken into account in the NPD predictions. Or it may reflect specific information available to NPD, not reflected in the data.

The obtained results may give raise to interesting questions to discuss with NPD expertise. Which horizon do they see as a realistic one, connected to their predictive distribution. Is H=256 a realistic one? If so, they may possess some risk reducing information about the future, not inherent in the data. If the horizon imagined by NPD is (considerably) less than 256, then it may either be so that they have not taken into account all information inherent in the data, or that they possess specific information about the future not inherent in the data. One may ask whether all uncertainties are represented.

We stress here that we believe that NPD has lot of information not inherent in the data, and that the results presented here are what the data may indicate without any extra information.

number also included Jan Mayen, a frontier area not covered by our data. With this area excluded, we would have arrived at a horizon somewhat less than 256, and then obtained an even better match with the mean.

4. The choice of horizon

Our prediction of undiscovered resources refers to a discovery horizon beyond 2013 that represents potential future additions of recoverable petroleum resources, either in separate discoveries or as contribution to resource growth. In the past, only 75% of all actual discoveries on the M-NCS have resources where recovery remains likely and of those 40% have their reserves included in existing fields and discoveries. Thus, a horizon of 10 in the prediction scheme may correspond to an exploration effort that discovers 33 potential developments (encompassing about 2.5 years' time span), using past years performance as a guide to the future. A horizon of 100, 200 and 256 therefore encompasses approximately a time span of 25 years, 50 years and 65 years respectively.

The predictive distribution of total aggregated resources is clearly dependent on the choice of horizon, which may be open for discussion. We see this as an opportunity, rather than a drawback. The traditional approach, assuming sampling proportional to size from a finite population requires specification of population size. This is also open for discussion, and may, or may not, be supported by geological expertise. For decision making purposes, some extra considerations on the time scale have to be added, answering questions on when the remains of the population are so small that they are not economic attractive, or the time span goes beyond time when petroleum is phased out anyway. The creaming model approach with decaying creaming factor is very flexible, and turns out predictions for any horizon. For both scientific and decision-making purposes, we have a common framework, and the choice of horizon is a well-defined issue, open for discussion with wide participation. The choices may well come out as follows: H=50 or less (short-term investors), H=100 (reports, to present medium-term economic potential), H=200 or more (scientists, to mimic all petroleum resources there are). The choice of horizon is, in some contexts, determined prior to looking at the data, in other contexts, derived from data itself. In the latter case, the following reasoning may be relevant. Creaming may be seen as the use of geological expertise to go for the most promising first. Thus, a decaying creaming function will represent diminishing ability or opportunity to provide useful information. At some point, the creaming factor is close to zero, meaning that the geological opportunities are close to exhausted. It then makes sense not to let the horizon go beyond this point, in our data an H in the range 200 to 300.

5. New discoveries

It may be of interest to compare our predictions, based on data up to and including year 2013, with the discoveries actually made in the following years. For the year 2014-2016, there were 33 discoveries listed in Exhibit 11 in increasing order.[9]

Discovery	mill Sm3 o.e.	Discovery year
6407/7-9 A	0,2	2016
34/6-3 A (Akkar)	0,3	2014
6407/7-9 S	0,4	2016
6406/2-8 (Imsa Sør)	0,7	2015
35/11-20 B	0,7	2016
35/8-6 A	0,8	2016
34/8-16 S (Tarvos)	0,9	2015
35/11-20 S	1,0	2016
2/4-22 S (Romeo)	1,3	2015
35/11-18 (Syrah)	1,3	2015
15/6-13	1,3	2015
6407/1-7	1,4	2014
6507/10-2 S (Novus)	1,5	2014
16/1-26 S	1,5	2016
6406/6-4 S	1,8	2015
35/11-17 (F-Vest)	2,0	2014
30/11-10 (Krafla Nord)	2,0	2014
35/11-18	2,0	2015
6706/11-2 (Gymir)	2,0	2015
25/5-9 (Trell)	2,1	2014
30/11-9 A (Askja Øst)	2,2	2014
6406/12-3 A (Bue)	2,2	2014
6706/12-3 (Roald Rygg)	2,4	2015
34/10-54 A (Valemon Nord)	2,4	2014
26/10-1 (Zulu Øst)	2,5	2015
34/10-54 S (Valemon Nord)	3,5	2014
6406/12-4 S (Boomerang)	3,5	2015
6707/10-3 S (Ivory)	5,1	2014
6706/12-2 (Snefrid Nord)	6,3	2015
2/4-23 S (Julius)	7,0	2015
25/2-18 S (Langfjellet)	7,8	2016
36/7-4 (Cara)	7,9	2016
6406/12-3 S (Pil)	18,6	2014

Exhibit 11. Discoveries made in the years 2014-2016 (following t=182).

These contain a number of smaller discoveries under evaluation as of December 31st 2016.[10] Some of them may be included in existing fields, others are candidates for becoming standalone developments that preferably can be tied back to a nearby infrastructure. However, many of the smaller discoveries may fail to be developed or considerably delayed. It is also a possibility that two or more medium sized discoveries, not individually worth developing, are combined to a development project. It is therefore not possible, at this stage, to pinpoint how many development projects these 33 discoveries will lead to.

[9] As before, the source for data is the yearly reports from The Norwegian Petroleum Directorate (NPD), prepared as input to the Revised National Budget (RNB2017).

[10] In a press release of 02.11. 2017 ENGIE E&P, the operator of the Cara-licence, has announced an increase in their resource estimate of 36/7-4 (Cara) from 7.9 MSm3o.e. to 11.8 MSm3o.e., the second largest in Exhibit 11.

Based on the performance in the past, referred to in Section 5, the 33 discoveries in Exhibit 11 will correspond to a horizon H=10. Let us say that at the end of year 2013 (at t=182), we were sure that a discovery of size above 10 would lead to a development project. For the horizon H=10, the chance of getting at least one discovery above size 10 was close to one, which in fact materialized. Although we had about 96% chance of getting at least one above 20, it did not materialize as an individual discovery, but may in principle materialize in conjunction with some other discovery. In this respect, our calculations are not inconsistent with the empirical findings.

We may also compare some characteristics of the empirical distribution of the actual sizes with the ones predicted from our model. The empirical median is a relevant and convenient characteristic. In Exhibit 12 we give the expected empirical median of H observations, taken over t=N+1, N+2,…, N+H, for some values of H.

H	5	10	15	20	25	50	100
E(Median)	5.2	4.4	3.9	3.7	3.6	3.1	2.5

Exhibit 12. Expected median size among H development projects after t=182

Note that this is derived directly from the simulations, and is not directly related to the median functions or the median of the predictive distributions at t=N+1, N+2,…, N+H.

The size distribution in Exhibit 11 may be judged under different assumptions. One possibility is to assume that the about 10 largest discoveries will end up as separate development projects. In Exhibit 11, we see 11 discoveries of size 2.4 or larger that may be sufficient for being a candidate for development, either separately or together with neighboring fields, if they exist. The median of the 11 largest numbers in Exhibit 11 is 5.1, while the expected median of size for the 11 next development projects after t=182 is 4.1, slightly less than the 4.3 median for H=10 in Exhibit 12. This is close when taking into account its variability. If we assume that some of the discoveries above 5.1 are matched with smaller ones into projects, this will not affect the median. However, the median will be raised if one of the 11, smaller than 5.1 is matched with one or more discoveries less than 2.4, and this brings the sum of their sizes above 5.1. However, these findings do not discredit in any our model.

6. Extensions

The model may be extended in various directions, in order to pretend being more realistic, allowing for more features in the data. A possibility is to include autocorrelation between successive discoveries, to be estimated from the data as well. With positive correlation, the variance of the total discoveries as a sum will increase. However, inclusion of autocorrelation is not expected to have a noticeable effect on the expected aggregated resources. Moreover,

this will destroy the nice features of our basic model, which allows simple simulations based on independent sampling. Another possibility is based on the observation that the data is indicative of a slight excess of small discoveries. The situation may be modelled as a two-regime-model, with a pure noise process of small discoveries is running concurrent with the log-normal exponential decay process. At each instant, there are fixed but unknown probabilities of coming from the one regime or the other. Both model extensions require additional parameters with attached uncertainties to be expressed by prior distributions. Both model extensions may lead to a more troublesome estimation process. Anyway, they are not expected to give noticeable different results from those above, except that it is likely that the quantile ranges in the predictive distribution of the accumulated sizes of discoveries will be widened.

In this work, we have used non-informative priors. There may exist prior information that legitimate the use of informative priors, in terms of expectations and statements of uncertainty. We may then combine this information with the data. We then face the problem of translating this information into priors for the current model parameters. A more basic question is whether the available information is implicitly based, at least partly, on the M-NCS discovery sequence itself. We may then face the trap of using the same data twice, which is illegitimate.

The possibility to adopt a creaming model with truncated log-normal discovery sizes is under investigation. The creaming features of this model are given in Lillestøl & Sinding-Larsen (2017). This is a model, which may still fit the data, but freed from any heavy tail problems. It is also amenable for full Bayesian solution, which is easily implemented in R, using R2OpenBugs.[11] The choice of truncation point will be somewhat arbitrary. We have tentatively tried a model with truncation point T=2000 and T=2500, with promising results. For the estimation phase, the parameter sigma seem to average about 1.6, as before. The Bayes estimates, i.e. posterior means, of mu and alpha seem to be more negative and positive, respectively, while beta is smaller. This means a less favorable population, but better opportunity for creaming than the non-truncated solution in section 4. The posterior distribution of mu seems to be multimodal, more so with lower truncation point. The quantile characteristics for the predictive distribution of total aggregated resources for Horizon H=256 are given in Exhibit 13, for the case of truncation T=2000, and for comparison, together with those for the unrestricted model and the NPD-predictions in section 4.

Quantile	0.5%	5%	10%	25%	50%	75%	90%	95%	99.5%
Bayes - 2000	547	765	871	1083	1483	1845	2387	2825	4219
Bayes - ∞	500	722	843	1091	1460	1980	2699	3302	5665
NPD-Total	695	891	993	1192	1461	1815	2272	2645	4396

[11] For estimation, this requires a log-transform of the data in the Bugs-model and reference to the truncated normal distribution. For simulation of future discovery sizes, one may use the function *rlnormTrunc,* available from the R-package EnvStat.

Exhibit 13. Quantile characteristics of distribution of total aggregated resources within horizon H=256 for truncated lognormal model (T=2000).

As expected, the truncation at T=2000 narrows the quantile ranges, and the effect is clearly seen at both ends of the distribution. At the low end, the quantiles are shifted a bit upwards, and at the high end downwards. The comparison with the NPD-quantiles reveals a noticeable difference at the high end, where the 95%-quantile of NPD is the lowest, and this is reversed when T=2000, for the more extreme 99.5% quantile. This is a clear revelation of the effect of adopting a truncated model with "low" truncation point.

In order not to be overly optimistic, it may be tempting to adopt a truncated lognormal model. As said, the choice of truncation point is somewhat arbitrary, and the effects of truncation need further investigation. However, it should be noted that non-truncation also contain a truncation decision, by allowing unlimited sizes.

7. References

Andreatta G, Kaufman GW (1986) Estimation of finite population properties when sampling is without replacement and proportional to magnitude, Jour. Amer. Statist. Assoc., vol. 81, no. 395, 657-666.

Barouch E, Kaufman GW (1967) Oil and gas modelled as sampling proportional to random size, Cambridge, Mass: MIT Alfred P. Sloan School of Management.
http://dspace.mit.edu/handle/1721.1/48701

Charpentier DR, Dolton GL, Ulmishek GF (1995) Annotated bibliography of methodology for assessment of undiscovered oil and gas resources, Nonrenewable Resources, vol. 4, no. 2, 154-185.

Halbouty MT (2001) Giant oil and gas fields of the 1990s. An introduction. Presentation at Symposium, Giant Oil and Gas Fields of the Decade 1990-2000, AAPG Convention, Denver, CO, June 5, 2001. (http://www.searchanddiscovery.com/documents/halbouty03/, downloaded Jan 12. 2016)

Kaufman GM, Balcer Y, Kruyt D (1975) A probabilistic model of oil and gas discoveries. In Methods of Estimating the Volume of Oil and Gas reserves (ed. J.D. Dunn). American Assoc. Petroleum Geologists, Studies in Geology no.1, 113-142.

Kaufman GM (1992) Statistical Issues in the Assessment of Undiscovered Oil and Gas Resources, MIT-CEEPR-92-010. WP.
http://dspace.mit.edu/bitstream/handle/1721.1/50204/35719963.pdf?sequence=1

Kaufman GM, Faith R, Schuenemeyer JH (2016) Has the largest field been discovered yet? PETRIMES and GRASP 25 years later, Math Geosci (2016). DOI:10.1007/s11004-016-9652-z

Lee PJ, Wang PCC (1985) Prediction of Oil and Gas Pool sizes when discovery record is available, Math Geol. v.17, No.2, p 95-113.

Lillestøl J, Sinding-Larsen R (2017) Creaming and the likelihood of discovering additional giant petroleum fields, Mathematical Geosciences, 49(1), 67-83, DOI: 10.1007/s11004-016-9657-7.

Meisner J, Demirmen F (1981) The creaming method: A Bayesian procedure to forecast future oil and gas discoveries in mature exploration provinces, Journ. Royal Statist. Soc., ser. A, vol. 144, no.1, 1-31.

R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/

RNB 2014 (2014) Resource account for the Norwegian Continental Shelf as of December 31, 2013. The Norwegian Petroleum Directorate. (Data file: Publiserte-tabeller-RNB2014.xlsx) http://www.npd.no/en/Topics/Resource-accounts-and--analysis/Temaartikler/Resource-accounts/2013/

Sturtz, S, Ligges, U, and Gelman, A (2005). R2WinBUGS: A Package for Running WinBUGS from R. Journal of Statistical Software, 12(3), 1-16.

Xu J, Sinding-Larsen R (2005). How to choose priors for Bayesian estimation of the discovery process model. Natural Resources Research, 14(3), 211-233.

Yan J, Prates M (2013) rbugs: Fusing R and OpenBugs and Beyond.
https://CRAN.R-project.org/package=rbugs

Sources of data:

The Norwegian Petroleum Directorate: The petroleum resources on the Norwegian Continental Shelf

http://www.npd.no/Global/Norsk/2-Tema/Ressursregnskap-og-analyser/Norsk/RNB2014/Publiserte-tabeller-RNB2015.pdf

http://www.npd.no/Global/Norsk/2-Tema/Ressursregnskap-og-analyser/Norsk/RNB2015/Publiserte-tabeller-RNB2016.pdf

http://www.npd.no/Global/Norsk/2-Tema/Ressursregnskap-og-analyser/Norsk/RNB2017/Publiserte-tabeller-RNB-2017-to-desimaler.pdf