As already stressed, uncertainties are endemic in the description of subsurface reservoirs. This first chapter provides the basics to address the estimation of reserves under uncertainty.
We first discuss why considering uncertainty is important. Then, we provide a few useful definitions about reserves and resources. The last section of this chapter recaps different techniques that are used to evaluate reserves and to estimate the uncertainties in these evaluations.
As previously mentioned, characterizing underground geological formations is a tedious task for at least three reasons: geological heterogeneity, lack of data and measurement errors. Therefore, the description of such formations is uncertain.

A first natural question is to ask what uncertainty is. Referring to Wikipedia, uncertainty is “a term used in a number of fields, including physics, statistics, engineering... It applies to predictions of future events, to physical measurements that are already made, or to the unknown. Uncertainty arises in partially observable and/or stochastic environments”. In reservoir characterization and modeling, uncertainty can be understood as the lack of certainty in describing an object, a feature or a process (Bardossy and Fodor, 2003).
Why do we need to quantify uncertainty?

- Questions to address in oil and gas industry
  - How much oil and gas exist?
  - How much can be produced?
  - What are the main sources of uncertainties?
  - How do they impact our estimates?
  - How to account for them?

- Why?
  - To evaluate risk
  - To make good decisions

The decisions to develop oil and gas reservoirs are taken under conditions of uncertainty because of the uncertainty in the description of these reservoirs. Some of the questions to address to help decision-makers are about the volumes of oil and gas available, the volumes that can be produced, the various sources of uncertainties and the way they influence the estimated volumes.

The identification of uncertainty sources and the understanding of the way they impact the volumetric estimates is essential to evaluate risk and make good decisions.
Due to the complexity of reservoir characterization, estimating oil and gas volumes in geological formations is complex.

Yet, a reliable estimate of petroleum reserves is required for making an idea about the world’s energy supply, for assessing the value of individual oil and gas companies and for making it possible to compare the values of these companies.

This emphasizes the need for an international standard with common definitions for quantifying the volumes of oil and gas, and for assessing uncertainties in these estimates.
A few acronyms are often used in the oil and gas industry, of which:

- **HCIIP**: quantities of HydroCarbons (oil + gas) Initially In Place. The term « initially in place » means that we are interested in the existing volumes of hydrocarbons before the commencement of production.

- **PIIP**: quantities of Petroleum Initially In Place.

- **OIIP**: quantities of Oil Initially In Place. It is also abbreviated STOIIP for Stock Tank Oil Initially In Place or STOOIP for Stock Tank Original Oil In Place. The term “Stock tank barrels” refers to the volume of oil at surface pressure and temperature (as opposed to reservoir conditions).

- **GIIP**: quantities of Gas Initially In Place.

- **RF**: recovery factor or fraction of the HCIIP (oil, gas or both) that can be produced.

- **UR**: ultimate recovery or cumulative volume of hydrocarbons (oil, gas or both) produced up to a specific time. When this time is long enough, UR is given by the product HCIIP × RF.
Oil and gas in place differ from oil and gas reserves. The reserves correspond to the hydrocarbon volume in the reservoir that is technically and economically recoverable. The volumes in place are estimated from the available data: the wells drilled in the reservoir and the complementary geophysical and geological surveys. These volume estimates vary with time as more and more information is provided by new wells or seismic surveys.

Volumes in place are defined for a reservoir or a reservoir unit. Three distinct categories of volumes in place are considered:
- Proven oil in place: volumes considered as certain (zones penetrated by wells)
- Probable oil in place: data interpretation shows the interval may be impregnated, without absolute certainty
- Possible oil in place: due to lack of data, high uncertainties exist on fluid contacts, presence of reservoir facies, etc.

Uncertainties in volume estimates depend on the dataset available for the construction of the geological model as well as the knowledge of geoscientists about the reservoir.
The example described above is based on two wells drilled behind a cliff in Yorkshire (UK). The two wells are numbered 10 (left) and 2 (right). The thickness of the interval is about 30m and the well spacing 300m.

The best porous reservoir facies corresponds to fluvial channel infill (in red, orange and yellow). Minor and thinner drains in the middle of the section correspond to crevasse splays (thin isolated layers in orange color).

The objective of the geologist is to correlate the two wells using all available data in order to predict the reservoir architecture.
An example of correlation is shown using these two wells.

In this case, the proposed layering is quite simple and mostly parallel. It is based upon lithostratigraphic considerations solely.

The arrow on the left indicates a level of crevasse splays (overbank deposits), which is generally connected to the main fluvial channel. Referring to the sedimentological section, the geologist can argue that there is a channel somewhere in the field in the stratigraphic interval. However, the location of the channel is unknown as far as there is no high resolution seismic available.
A third well (well 9) is drilled. It does not really bring any new information as it is very similar to well 10.

In addition, two vertical sections are measured on the cliff outcrop (sections 6 and 3). They provide data characterizing the presence and the geometries of channels. This makes it possible to refine the correlation study with new channels in the middle of the interval. Without these two vertical sections, the general correlation would not have changed that much.
In a subsequent step, we account for the information provided by 4 new wells and vertical sections observed along the cliff.

As expected, adding more data to the field database permits to build a much more detailed reservoir model.

There is now a better control of the internal architecture of the reservoir. In such conditions, we improve the accuracy of the estimates of the volumes in place.
Reserves

- Reserves = « quantities of petroleum which are anticipated to be commercially recovered from known accumulations from a given date forward »

- There are also quantities of petroleum potentially recoverable, but not considered as reserves
  - Contingent (discovered but not commercial) resources
  - Prospective (undiscovered) resources

There is general consensus to name reserves or economic reserves the quantities of petroleum, which are anticipated to be commercially recovered from known accumulations from a given date forward (SPE, 1997).

Other quantities of petroleum are potentially recoverable, but not considered as reserves. They encompass the discovered quantities that are not currently considered economically viable: these resources are called contingent (or marginal) resources. Other potentially recoverable petroleum quantities include undiscovered resources: they are often referred as prospective. The term “resources” is used in opposition to reserves.
Reserves

- Depending on the degree of uncertainty, there are three main types of reserves:
  - Proved
  - Unproved
    - Probable
    - Possible

- Reserves are given by HClP × RF

Reserves are split into three main classes: proved, probable and possible. This discrimination provides insights about the uncertainty degree in the estimated petroleum volumes.

A first step consists in placing the reserves into either the proved or unproved classes. The unproved reserves are less certain to be recovered than the proved reserves.

Second, the unproved reserves are sub-classified as probable and possible reserves to denote progressively increasing uncertainty in their recoverability.

As stated above, reserves are calculated as the quantities of hydrocarbons initially in place multiplied by the recovery factor. The volumetric estimations are usually considered as part of the job of geologists as they are based on the analysis of static data: cores, logs and geological maps. On the other hand, the recovery factor is a dynamic variable and as such it is estimated by reservoir engineers. It depends on fluid properties (formation volume factor, viscosity, density, etc.), fluid-rock interactions and production mechanisms. The recovery factor is one of the most important, yet the most difficult variable to estimate.
Only part of the oil or gas in place is ultimately produced. The recovery factor (RF) is defined as the ratio of the volume produced to the volume initially in place.

Typical RFs in the oil and gas industry are reported above depending on the reservoir drive mechanisms.

In the case of oil, production is sub-divided into 3 successive phases.

- During the **primary** recovery stage, reservoir drive comes from a number of natural mechanisms such as the expansion of the natural gas at the top of the reservoir (RF = 30 to 60%), the expansion of the gas initially dissolved in the crude oil (RF = 2 to 30%), or natural water displacing oil towards the well...

- Over the lifetime of the well, fluid pressure falls. At some point, it is not enough to force the oil to the surface. **Secondary** recovery methods are then applied in the form of water or gas injection to increase reservoir pressure. The typical RF for water drive is 2 to 50%.

- **Tertiary** or enhanced oil recovery (EOR) methods increase the mobility of the oil in order to boost production. They rely on the injection of steam, CO₂, surfactants, polymers... Tertiary recovery allows another 5% to 15% of the reservoir's oil to be recovered (Bentley *et al.*, 2009).

At the end, even though it is essential to know the uncertainty in the oil in place, it is even more important to quantify the uncertainty in the volume of oil to be produced.
According to SPE (1997), proved reserves are those quantities of petroleum that can be estimated with reasonable certainty to be commercially recoverable from a given date forward from known reservoirs. In this case, the expression “with reasonable certainty” means “with a high degree of confidence”. Proved reserves can be developed or undeveloped.

Probable reserves are less likely to be recoverable than proved reserves, but more likely than possible reserves. They include reserves in formations that appear to be productive based on well log analysis, but lack core data or definitive tests. They also consist of reserves in formations separated (but less deep) from the proved area by faulting.
Reserves

- Possible reserves: unproved reserves less likely to be recoverable than probable reserves
  - *e.g.*, possible incremental reserves attributed to infill drilling subject to technical uncertainty

- Reserve estimates are revised as additional data are available or as economic conditions change

Possible reserves are those unproved reserves less likely to be recoverable than probable reserves. They can be reserves in formations that appear to be petroleum bearing based on log and core analysis, but may not be productive at commercial rates. They can also correspond to incremental reserves attributed to infill drilling subject to technical uncertainty.

Reserve estimates are generally revised as additional data are available or as economic conditions change.
The words commonly-used to describe uncertainty in traditional definitions of reserves are for instance reasonable certainty, less likely, more likely, probable, proved, possible... These terms are quite unclear when used to quantify levels of confidence. Because of ambiguity associated to these terms, probabilistic definitions that quantify uncertainty are increasingly used in the industry.
Probabilistic definitions

- **Proved** reserves: there should be at least a 90% probability that the quantities actually recovered equal or exceed the estimate $\rightarrow P_{90}$

- Probable reserves: there should be at least a 50% probability that the quantities actually recovered equal or exceed the sum of the estimated **proved** plus **probable** reserves $\rightarrow P_{50}$

- **Possible** reserves: there should be at least a 10% probability that the quantities actually recovered equal or exceed the sum of the estimated **proved** plus **probable** plus **possible** reserves $\rightarrow P_{10}$

Within this framework (Derminen, 2007), proved reserves are associated to a probability of at least 90% that the quantities actually recovered equal or exceed the estimate. The estimate corresponding to the 90% probability is called $P_{90}$.

For probable reserves, there is at least a 50% probability that the quantities actually recovered equal or exceed the estimate called $P_{50}$. This estimate is the sum of estimated proved plus probable reserves. The term 2P is often used for proved plus probable.

Likewise, for possible reserves, there is at least a 10% probability that the quantities actually recovered equal or exceed the estimate called $P_{10}$. This is the sum of estimated proved plus probable plus possible reserves. In this case, we refer to 3P reserves.
The notations introduced in the previous slide are recapped in the figure above (Derminen, 2007). The red curve is the probability density function (pdf) characterizing the reserves. The P90 is the estimate for which there is a 90% chance that the quantities actually recovered are greater. In other words, the integral of the pdf for reserve values greater than P90 is 0.90.

The P50 is the median estimate. There is a 50% chance that the quantities actually recovered are greater. The P10 is the estimate for which there is a 10% chance that the quantities actually recovered are greater.

The P10, P50 and P90 values make it possible to evaluate uncertainty. The narrower the P90 to P10 range, the less the uncertainty.
The probability density function (pdf) is turned into a descending (or reverse) cumulative density function (cdf). In this case, the cdf for the x estimate gives the area under the probability density function from the x estimate to infinity.

The P90 estimate is then associated to 0.90. This value is the area under the pdf from P90 to infinity. It is also the probability of having reserves greater than P90.

Likewise, the P50 is associated to 0.50. This is the median or the halfway point: 0.50 is the probability of having reserves greater than P50.
An intuitive idea is that uncertainty decreases as additional data become available. Thus, we expect uncertainty becomes smaller with increasing field maturity. This idea is however not supported by facts (Derminen, 2007). The figure above shows a typical trend of the uncertainty range (in the pre-production stage) of estimated gas initially in place for a field in the North Sea (Stoessel, 1994). The range of uncertainty is given by the interval between the high and low estimates. It is nearby constant over time. However, in the meantime, the middle estimate evolves significantly. The integration of new data is crucial as it helps refine the estimate. However, it does not necessarily decrease uncertainty. This conclusion seems to hold not only for the pre-production stage, but also during production (Derminen, 2007). A reasonable explanation (Derminen, 2007) is that our ability to properly quantify uncertainty is rather limited due to the endemic complexity of the problem.
Three main families of methods are known to estimate reserves: analogy, volumetric and performance methods.

The analogy methods are the simplest ones. They are based upon analogy with a nearby producing reservoir. In such conditions, the resulting reserve estimates are appropriate provided the analogy is valid.

The volumetric and performance estimation methods are more elaborate. They differ mainly in the types of data used. Volumetric methods refer to the analysis of static data while performance methods use both static and dynamic data. This actually makes the difference between pre- and post-production phases.

Volumetric methods are split into two main groups: deterministic and stochastic methods. In addition, we can stress different levels of complexity among stochastic techniques depending on the refinement of reservoir modeling. The reservoir model can be considered as a unique grid block with a unique, even though random, porosity value, oil saturation value, water oil contact... Conversely, it can be described by a three-dimensional grid populated by random realizations for porosity, oil saturation, etc. The ability to capture the uncertainty due to given geological features depends on the refinement level. The main steps of reservoir modeling will be described in the following chapters, thus making it possible to generate several reservoir models, hence to use advanced stochastic techniques to estimate reserves.
Last, performance methods can be applied when there are enough pressure and production data to allow prediction of future performance. They include decline-trend analysis: the known production data are used to define a trend that is extrapolated until the economic limit is reached. This provides the estimated reserves.

In the case of material balance techniques, the reservoir is considered as a tank. The pressure behavior is analyzed in several steps in response to fluid withdrawal.

Last, reservoir simulation techniques call for fluid flow simulations. Given a reservoir model, fluid displacements are predicted from a fluid flow simulator. This provides the cumulative volume of oil produced. Clearly, this approach leads to reliable results provided the reservoir model matches the observed production history. We will see in chapter 4 that history-matching is an ill-posed problem meaning that there exist many reservoir models matching production data. Again, a stochastic approach is preferred at this stage.
Reserve estimation – Which methods?

- **Depends on the objective**
  - For quick look evaluation (data room, discovery, prospect)
    - Analogy method
    - Deterministic volumetric method
  - For mature projects (reservoir management plan, field with production history)
    - Stochastic volumetric method
    - Performance methods

The choice of method depends on field maturity, data available and reservoir heterogeneity. Analogy and deterministic methods are preferred for quick look evaluation while stochastic volumetric and performance methods are more suited for mature projects.

In the case of volumetric estimations, basic stochastic methods require less computation time than advanced stochastic methods. Likewise, in the case of performance methods, decline-trend analysis is less CPU-time demanding than material balance techniques, which are also less demanding than reservoir simulation methods.

It is worth stressing that volumetric techniques generally involve greater errors and uncertainty than performance methods as they do not account for dynamic data.
As previously pointed out, oil reserves are derived from the product of the volume of oil initially in place (OIIP) by the recovery factor (RF).

The OIIP can be evaluated from mathematical formulas. The one reported here holds for the metric system. In short, OIIP equals rock volume multiplied by porosity $\phi$ multiplied by oil saturation divided by oil formation volume factor $B_o$ at initial reservoir conditions. Porosity $\phi$ is the pore space in the rock, that is the fraction of voids over the total volume. $S_w$ is the water saturation. When there is no gas, oil saturation equals $1-S_w$. When oil is produced, the reservoir temperature and pressure decreases to surface conditions and gas bubbles out of the oil. Thus, the volume of the oil decreases. Oil reserves are calculated for standard surface conditions (15°C and 101.325 kPa) instead of reservoir conditions. The oil formation factor $B_o$ is defined as the ratio of volume at reservoir conditions to volume at standard surface conditions.

RF is a very important, yet very difficult variable to estimate. It is determined by analogy, analytically or fluid flow simulation.

For many reservoirs, the most uncertain parameters are the rock volume and RF.
The formula above also yields the volume of oil initially in place, but referring to the imperial system. Again, the volume is calculated for standard surface conditions (60°F and 14.7 psi).

Similar formulas are applied to calculate the volume of gas initially in place (Garb and Smith, 1987).
For simplicity, we apply a basic stochastic approach to assess uncertainty in reserve estimates. OIIP is a variable that depends on several input variables. The general practice is to use continuous probability density functions (pdfs) to characterize the input variables and to combine these distributions to generate a pdf for reserves.

The input pdfs are combined either analytically (Capen, 1992) or by random sampling (Monte Carlo simulation).
For Monte Carlo simulation, each parameter (drainage area, thickness, porosity, water saturation, oil formation volume factor) is seen as a random variable satisfying some pdf (blue distributions on the top figure). A value is randomly selected for each of the parameter: this is a trial or realization. The resulting reserve estimate is calculated for each set of drawn values: this is the output. The experiment must be repeated hundreds or thousands of times. All the output reserve estimates are grouped into a histogram (red distribution on the top figure) or cumulative histogram.

A large number of iterations is actually required to get stable results.

The histogram can be plotted in terms of frequency (number of outcomes in each category or bin) or density (number of outcomes/total number × width of bin). The histogram approximates the probability density function when expressed against density.

Note that on the example shown above, the input pdfs are either normal or lognormal. These functions are described by two parameters: the mean and the variance. When the pdf is normal, the mean and the variance are those of the normally distributed variable. When the pdf is lognormal, the mean and the variance are those of the associated normally distributed log(variable). More details are provided on this topic hereafter.
Reserves are usually calculated from random variables having continuous probability density functions (pdfs). A pdf is a function that describes the relative likelihood for a random variable to take on a given value.

The available data are grouped to build histograms or sample distributions. Then, probability density functions (or distributions) are used to model these histograms. There exist numerous probability density functions, for example the uniform distribution, the triangular distribution, the normal and the lognormal distributions, the beta distributions... They have different shapes and may be preferred for a given type of data depending on its histogram.

For instance, the drainage area or rock volume being positively skewed with fat right tails, the lognormal distribution is often favored in this case. A "skewed right" distribution is one in which the tail is on the right side while a "skewed left" distribution is one in which the tail is on the left side.

In addition, these distributions may have to be bounded above or below to respect physical limits. For example, saturation is between 0 and 1. Cutoffs must also be accounted for when applied (for instance to porosity) to discriminate between reservoir and non-reservoir rocks.

The choice of the probability density function for a given input variable depends on data and professional judgment.
The uniform distribution, also known as rectangular distribution, is a distribution with constant probability on the interval from \(a\) to \(b\). It is thus defined from 2 parameters. The uniform distribution can be used to characterize a variable when the only information is that it is in the range \([a, b]\).

The triangular distribution is defined from 3 parameters: lower bound \(a\), upper bound \(b\) and mode \(c\) with \(c\) in the range \([a, b]\). The mode is the value at which the probability takes its maximum value. The triangular distribution is flexible as a probability model as it can be symmetric, right- or left-skewed. For instance, the right-skewed shape is of interest for a variable, whereby high probability of low values tapers off to low probability of high values.
The normal distribution, also called Gaussian distribution, is the most commonly used
distribution in many disciplines. It is defined from 2 parameters: mean $m$ and variance $\sigma^2$.
The standard deviation is the square root of the variance: it shows how much dispersion
exists from the mean. The normal distribution is also called bell curve as its plot
resembles a bell.

Due to the symmetric shape of the normal distribution, the mean is also the median and
the mode.

As the normal distribution is non-zero over the entire real line, it has often to be truncated
in practical applications, thus resulting in a truncated normal distribution.
A lognormal distribution is the distribution of a variable whose logarithm is normally distributed. Thus, if the random variable $X$ is lognormally distributed, $Y = \log(X)$ is normal with mean $m$ and variance $\sigma^2$. Logarithm generally refers to natural logarithm. In practice, however, the decimal logarithm is often preferred in petroleum engineering.

Like the normal distribution, the lognormal distribution is completely described by two parameters: mean $\mu$ and variance $\nu^2$. They are the mean and the variance of $\log(X)$. The mean and variance of $X$ are denoted $\mu$ and $\nu^2$, respectively. Note that the mean and the variance of $\log(X)$ are not equal to the mean and the variance of $X$.

A random variable that is log-normally distributed takes only positive real values. Its distribution is skewed to the right meaning that it has a long right tail.
We assume that $X$ is lognormally distributed. Therefore, $\log(X)$ has a normal distribution. The mean and the variance of $\log(X)$ are denoted $m$ and $\sigma^2$, while those of $X$ are $\mu$ and $\nu^2$. The lognormal distribution is defined in terms of the mean and the variance of the associated normal distribution instead of its own mean and variance. The relationship between the mean and variance of $X$ and the mean and variance of $\log(X)$ is shown above.

The plot on the left displays the distribution of $\log(X)$ while the one on the right shows the distribution of $X$. The mean and variance of $\log(X)$ are -1.1 and 0.15, respectively. It results in mean and variance of 0.36 and 0.02 for $X$.

$$
\mu = \exp\left(m + \frac{\sigma^2}{2}\right)
$$
$$
\nu^2 = \exp\left(2m + \sigma^2\right)\left[\exp(\sigma^2) - 1\right]
$$
The beta distribution is defined on the interval [0; 1] and has two free positive parameters, $\alpha$ and $\beta$. It is a power function of the $x$ variable and its reflection $1-x$. The beta function $B(\alpha, \beta)$ (not to be confused with the beta distribution itself) appears as a normalization constant.

The shape of the distribution depends on the two shape parameters $\alpha$ and $\beta$. The above plot shows the beta distribution for various $\alpha$ and $\beta$ values.

This type of distribution can be used for instance to describe the uncertainty in $K_v/K_h$ ratio (ratio of vertical to horizontal permeability).
The lognormal distribution plays a central role in the stochastic estimation of reserves even though distributions other than lognormal describe the individual input variables. One of the main reasons is the Central Limit theorem: any multiplicative process converges towards lognormality.

When input distributions are lognormal, with no truncation, and independent from one another, the resulting product for HCIIP can be obtained analytically.

Otherwise, regardless of the form of the individual distributions, Monte Carlo simulation results tend to a skew distribution that looks lognormal.
Even when using all available data, there is still considerable uncertainty in the description of reservoir properties. This issue is addressed within a stochastic framework. Given this context, we introduced different methods to estimate reserves, of which the stochastic volumetric methods. In the simplest case, the reservoir is considered as a unique grid block with properties randomly assigned on the basis of probability density functions. Monte Carlo simulation can then be run to assess the degree of uncertainties in reserve estimates.

Yet, this approach remains rough. Advanced stochastic volumetric methods and reservoir simulation based performance methods have been developed to refine uncertainty quantification.

The following chapters (2 and 3) investigate the building of realistic three-dimensional reservoir models from static data. Various tools are described that make it possible to generate several likely reservoir models. Instead of using the input distributions directly as explained in the current chapter, all volumetric calculations are based on the generated models. Such an approach provides more flexibility to evaluate how the uncertainty in a broad range of parameters (e.g., top depth, location of a sealed fracture, net-to-gross ratio, etc.) impacts the volumetric estimates. In addition, it makes it possible to analyze the spatial distribution of fluids as well as the associated uncertainties.

A finer geological modeling boils down to an improved description of the spatial distribution of the fluids initially in place. However, it says nothing about recovery, which is the ultimate response of interest. An additional mandatory step consists in building models respecting not only static, but also dynamic data. Then, fluid flow simulations are performed to assess uncertainties in oil and gas production. This topic will be investigated in Chapter 4.