NTNU – Trondheim
Norwegian University of
Science and Technology

Statistical Methods for Calculating the Risk of Collision Between Petroleum Wells

Bjørn Erik Loeng

Master of Science in Physics and Mathematics
Submission date: June 2012
Supervisor: Jo Eidsvik, MATH
Co-supervisor: Erik Nyrnes, Statoil

Norwegian University of Science and Technology
Department of Mathematical Sciences

Abstract

In this thesis we explore several statistical methods for addressing the risk of collision between two petroleum wells. Such a collision is a potentially dangerous but rare event that can occur in situations with directional drilling.

The common practice in the petroleum industry today is to use the two closest points in the two wells in a hypothesis test, in order to make a conclusion on whether we should drill as planned based on the collision risk. We suggest a more accurate version of the hypothesis test, which turns out to be more conservative than the original test.

In order to extend the usual approach of only considering the two closest points in the collision risk calculations, we obtain a joint statistical distribution for the position coordinates of all the survey points in two neighboring wells.

As an alternative measure of the collision risk, we estimate the probability of collision. This is done in two different ways, namely by considering only the two closest points and by considering the whole wells. In the latter case, we use the joint distribution for all the survey points. For some well pair cases, the collision probability is much larger when we consider all the survey points in two wells, than when we only consider the two single closest points.

We estimate the probability values by using Monte Carlo simulation methods. Since a well collision is considered to be a rare event, we introduce two methods in order to increase the accuracy in the situations where the original Monte Carlo method needs an inconveniently large number of samples. These methods give accurate results even when the collision probability is very small.

Sammendrag

I denne masteroppgaven undersøker vi en rekke statistiske metoder som beskriver risikoen for kollisjon mellom to petroleumsbrønner. En slik kollisjon er en potensielt farlig, men sjelden hendelse, som kan oppstå i situasjoner der man benytter seg av retningsboring.

I dag er den vanlige fremgangsmåten i petroleumsindustrien å bruke de to nærmeste punktene i de to brønnene i en hypotesetest for å komme fram til en konklusjon om hvorvidt vi skal bore som vi har planlagt, basert på kollisjonsrisikoen. Vi foreslår en mer nøyaktig versjon av hypotesetesten, som viser seg å være mer konservativ enn den opprinnelige testen.

Som en utvidelse av den vanlige tilnærmingen der man kun tar for seg de to nærmeste punktene i risikoberegningene, finner vi den simultane statistiske fordelingen til posisjonskoordinatene til alle målepunktene i to nabobrønner.

Som et alternativt mål på kollisjonsrisiko, finner vi sannsynligheten for kollisjon. Dette gjøres på to forskjellige måter, nemlig ved å betrakte kun to enkeltpunkter, og ved å betrakte hele brønnene. I det siste tilfellet gjør vi bruk av den simultane fordelingen til alle målepunktene. For noen brønnpar er kollisjonssannsynligheten mye større når vi betrakter alle målepunktene enn når vi kun betrakter de to nærmeste punktene.

Vi estimerer sannsynligheten ved hjelp av Monte Carlo-simuleringsmetoder. Siden en brønnkollisjon anses å være en sjelden hendelse, innfører vi to metoder for å forbedre nøyaktigheten i situasjoner der den opprinnelige Monte Carlo-metoden trenger et upraktisk stort antall simuleringer. Disse metodene gir nøyaktige resultater selv om kollsisjonssannsynligheten er svært liten.

Preface

This thesis marks the end of my Master's degree in Industrial Mathematics at the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. The work has been carried out at the Department of Mathematical Sciences during the spring semester of 2012. The work is made in collaboration with Statoil, which has provided me with the overall problem of the thesis.

I would like to thank my supervisor at NTNU, Associate Professor Jo Eidsvik, for all our inspiring discussions and his helpful advices during the whole semester. I would also like to thank my supervisor at Statoil, Erik Nyrnes, for his useful advices and guidance regarding the industry practice.

Bjørn Erik Loeng
June 25, 2012
Trondheim

Contents

Chapter 1

Introduction

In the petroleum industry, the term *directional drilling* is used for the operation of drilling a petroleum well in directions that are not necessarily vertical. The ability to drill in three dimensions makes it possible to make a complex system of wells far below the surface. For instance, at *Statoil*'s offshore field named *Heidrun*, there is a subsea template that contains 56 slots (Hansen et al., 2011), which leads to a high density of wells. When we drill a new well in such an area, we must be aware of the risk of collision with adjacent wells. In this thesis we will explore various methods for describing and measuring the possibility of such a collision.

One benefit with directional drilling is that a large area can be covered by one stationary rig. Also, it is easier to reach targets that are inaccessible with vertical wells. On the other hand, the task of deciding the position of the drill string at all times is more complex than for vertical wells. In addition, there are a number of error sources that influence the position measurements, complicating the problem.

While it is important to know the position of the well accurately in order to achieve the objective of the well, like hitting a target reservoir, it is also necessary to know as much as possible about the position in order to be able to avoid well collisions.

1.1 Measurement While Drilling

Measurement while drilling (MWD) is a technique for measuring properties of a petroleum well during the drilling operation. The measurements are made at appropriate well length intervals, often at around every 30 meters. This is the regular length of drill pipes (or stands), and an MWD survey is often made during the non-drilling period that arises when a new drill pipe is attached to the drill string. A well point where a survey is made is called a *survey point* or a *survey station*. For an offshore well with nine survey points, the situation is described by Figure 1.1.

Our focus will be on *magnetic measurement while drilling*, because this is the most common kind of MWD in the petroleum industry today. From now on in this

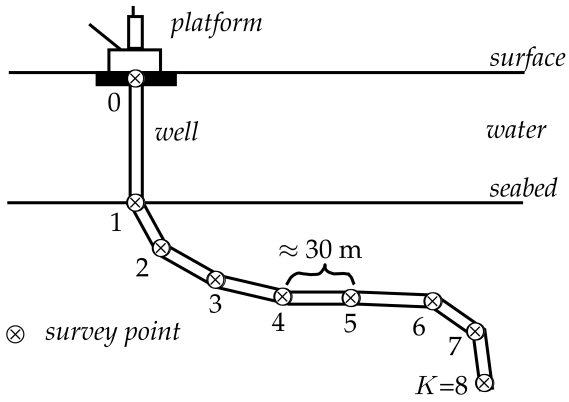

Figure 1.1: A simplified example of an offshore well consisting of $K + 1 = 9$ survey points. The distance between survey points is often about 30 meters.

thesis, *MWD* implicitly means *magnetic MWD*.

The MWD tool, which is a part of the bottom hole assembly of the drill string, measures many different values, for instance the temperature and various mud flow properties. For directional drilling, the most important measurements are the ones from accelerometers and magnetometers. Using these instruments, we can measure the gravity field and the magnetic field at a survey point. We use these values to estimate the *North*, *East* and *Vertical Depth* coordinates, also known as the NEV coordinates, of the survey point.

1.2 Well Collision Avoidance

In the past, all wells were vertical, and the risk of collision between wells was negligible. With directional drilling, the possibility that a well collision occurs must be considered when the well that is being drilled, called the *reference well*, is close to previously drilled wells, called *offset wells*. The risk that is taken should be as small as possible, because such a collision would at best cause major economic damage. If the collision leads to a blow-out, one could in addition potentially face life-threatening and environmental consequences.

If one knew the exact positions of all wells, there should be no risk of collision at all. There are, however, various error sources, which give uncertainty to the NEV position coordinates of the well. This uncertainty can be used to draw conclusions about the risk of collision between two wells.

When it comes to well collision avoidance, the current practice in the petroleum industry is to find the two closest points in the two neighboring wells. One then uses the uncertainty of the positions to make a conclusion in a *hypothesis test* on whether the risk of collision is small enough to proceed the drilling. In this thesis we will explore this hypothesis test, and we will also suggest a new and improved

alternative test.

The task of considering segments of wells in the collision avoidance calculations, rather than only single points, has previously been attempted solved by Tsao et al. (1999), Thorogood et al. (1991) and Brooks (2008). However, they do not take into account the correlation between different areas of the segments or between the two wells. In this thesis, we introduce a method that obtains the joint statistical distribution of all the survey points in two wells, in order to completely describe the position uncertainty.

We also introduce and compare various ways of estimating the *probability* of well collision, as an alternative measure of the risk that is taken if we decide to drill as planned. We obtain this probability in two ways, by first considering only the two closest points and then by considering all the survey points in the wells. In the latter case, we use the previously mentioned joint distribution of the points.

1.3 Overview

In Chapter 2 we introduce the concept of *well positioning*, that is to estimate the position coordinates of the points in a petroleum based on MWD measurements. We also find the statistical distribution of the *NEV* coordinates of the survey points in a petroleum well, by using *error propagation*.

In Chapter 3 the goal is to find various ways of describing the risk of well collision. This involves finding the *probability of well collision* in different (and new) ways. Methods that rely on *Monte Carlo simulations* are computationally challenging, especially in a rare-event situation, which is often the case for a well collision. An important contribution from this thesis is the use of a *cross-entropy method* and the *enhanced Monte Carlo method* in order to reduce the computational efforts in such situations.

In Chapter 4 we introduce a number of well pair test cases in order to explore and compare the methods introduced in Chapter 3. We also gain some initial insight into how the well geometry parameters *crossing angle*, *well direction* and *depth* influence the collision risk for the test cases.

Finally, we summarize the results in Chapter 5.

Chapter 2

Well Positioning and Error Propagation

Our goal in this chapter is to find the *NEV* coordinates of the survey points in a petroleum well, based on the MWD measurements in the points. This task is also known as *well positioning*. In addition, we will obtain the uncertainty of the estimated positions, based on the *error propagation* of various error sources.

We will consider a petroleum well where the MWD measurements are made (or planned) at $K+1$ *survey points* (or *survey stations*) $k \in \{0, \ldots, K\}$. Each survey point belongs to one of the *survey series* $l \in \{1, \ldots, L\}$. Each survey series contains surveys that are made with the same survey tool. The situation is described in Figure 2.1. We assume that the survey series are non-overlapping, although this assumption may not always be true in a real situation.[1]

In Section 2.1 we will define the coordinate systems that are used in the calculations, as well as some useful angles like the inclination I and the azimuth A that describes the direction of the well path in each point. In Section 2.2 we introduce the MWD measurements of the gravity field $\mathbf{G}$ and the magnetic field $\mathbf{B}$, and in Section 2.3 we calculate the angels I and A from $\mathbf{G}$ and $\mathbf{B}$. Finally, using I, A and the measured depth D (from now on called the *DIA* values), we show how to obtain the *NEV* coordinates of the survey points in Section 2.4. Summarized, the procedure is given by

$$D, \mathbf{G}, \mathbf{B} \longrightarrow D, I, A \longrightarrow N, E, V.$$

Simultaneously, we will show how to find the *joint statistical distribution* of the estimated *NEV* positions, including their covariance matrix $\mathbf{\Sigma}_{NEV}$. In Section 2.2 we begin with a number of error sources that are assumed to influence the position measurements, described by the covariance matrix $\mathbf{\Sigma}_{\varepsilon}$. In Section 2.3 we show how the uncertainty propagates to the uncertainty of the *DIA* values,

[1] Even though we have chosen to assume non-overlapping survey series for convenience, there is nothing that indicates that it is impossible to generalize the forthcoming calculations so that they apply to *overlapping* survey series as well.

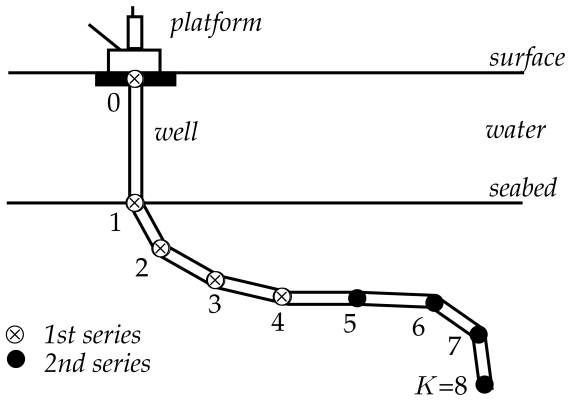

Figure 2.1: A simplified example of an offshore well consisting of $K + 1 = 9$ survey points. There are $L = 2$ survey series in this example. The points 0–4 belong to the first survey series while the points 5–8 belong to the second series.

by approximating the covariance matrix $\mathbf{\Sigma}_{DIA}$, and in Section 2.4 we finally find the statistical distribution of the *NEV* positions, including the covariance matrix $\mathbf{\Sigma}_{NEV}$. In short, the procedure is described by

$$\mathbf{\Sigma}_{\varepsilon} \longrightarrow \mathbf{\Sigma}_{DIA} \longrightarrow \mathbf{\Sigma}_{NEV}.$$

In order to do the transition from one covariance matrix to the next, we use the *error propagation theory* described in Appendix A. In Section 2.5 we generalize the calculations in order to obtain the joint statistical distribution for the survey points in *two or more* wells.

The exploration of well position uncertainty began with the work of Walstrom et al. (1969). Here, all error sources were assumed to be random and independent between different survey stations. Wolff and de Wardt (1981) introduced the concept of systematic (or survey-specific) error sources, while Brooks and Wilson (1996) introduced error propagation theory in order to find the 3×3 covariance matrix for the *NEV* coordinates at a point k. This method was developed further by Williamson (2000), whose normal (Gaussian) error model makes up the basis for this thesis, as it did for Gjerde (2008) and Gjerde et al. (2011).

By introducing matrix notation, much more information than only the marginal 3×3 covariance matrices can be obtained, namely the covariance values *between* different survey points. If we for instance want to simulate the *NEV* positions from the joint statistical distribution of several points in a larger part of the well, rather than considering one single point, it is important to know about the correlation between the different points. Here, we will therefore eventually find the joint $3K \times 3K$ covariance matrix $\mathbf{\Sigma}_{NEV}$ for the *NEV* positions of *all* the survey points in a petroleum well. We assume that the initial position $\mathbf{p}_0$ is known. In the forthcoming

calculations, we will also assume that we are given the measured (or planned) *DIA* values at every survey point in the well.

2.1 Definitions of Coordinate Systems and Angles

In order to describe the position of a point in the well, we will use an *NEV coordinate system*. The NEV coordinates are often given relative to a subsea template or a platform, as seen in Figure 2.2. This coordinate system is spanned by the *N(orth)*, *E(ast)* and *V(ertical depth)*[2] directions. In Figure 2.2 we also see another coordinate system that we will use in the well positioning. With its origin in the MWD tool (drawn as a black square), the *tool coordinate system* is spanned by the X, Y and Z directions.

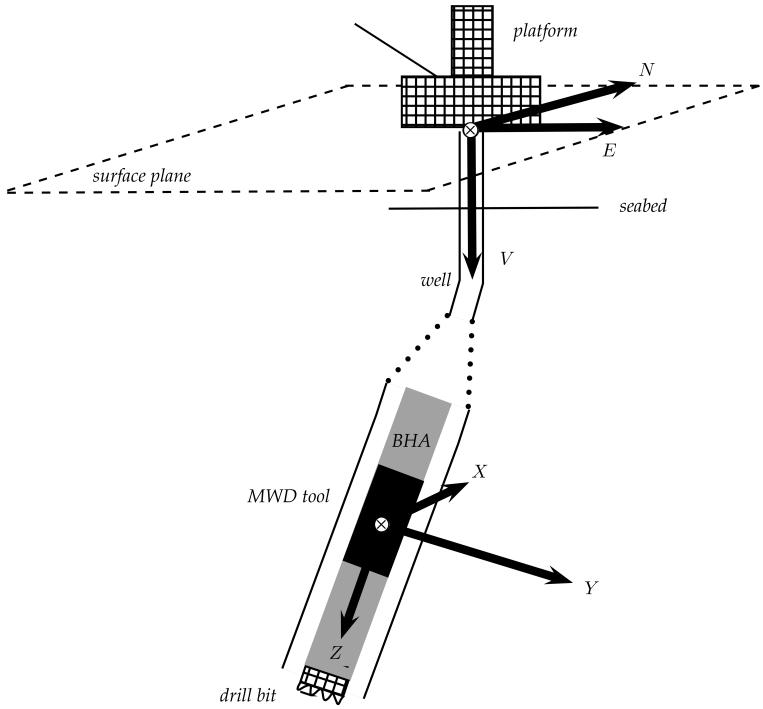

Figure 2.2: The coordinate systems used in this thesis. In an offshore setting, the *NEV* coordinate system often has its origin at the platform, and the *NE* plane is the water surface plane. The tool coordinate system consists of the X, Y and Z directions and has its origin in the MWD tool, which is drawn as a black square. The tool is a part of the *bottom hole assembly (BHA)* of the drill string.

[2]The V direction is also known as *TVD* or *True Vertical Depth*.

The *NEV* coordinate system is actually an extended version of an *universal transverse mercator (UTM)* coordinate system. In a UTM system, the earth's surface is divided into a number of two dimensional zones. By extending to three dimensions within such a zone, we have a Cartesian coordinate system where the N axis is in the direction of the geographical north, while the V axis points in the downward vertical direction. The E axis then corresponds to the direction of the geographical east.

In Figure 2.3 we have depicted the coordinate system after translating the origin to a survey point in the well, in order to see how the *azimuth angle* A and the *inclination angle* I are defined. These angels are used to describe the direction of the well path in the point. Let a direction vector $\mathbf{t}$ start at the well point of interest, and let the vector point in the direction of the well path in that point. The *azimuth* A is defined as the angle from the positive N axis to the projection of the direction vector on the NE plane, while the *inclination* angle I is the angle from the positive V axis to the direction vector.

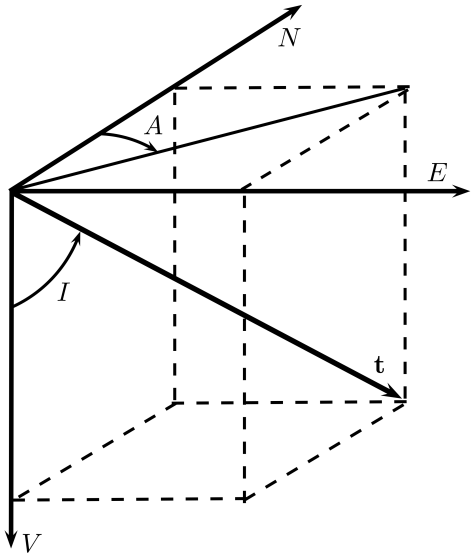

Figure 2.3: The *NEV* coordinate system. The direction vector $\mathbf{t}$ gives the direction of the well path in a well point. After $\mathbf{t}$ is translated to the origin of the coordinate system, one can obtain the azimuth angle A and the inclination angle I.

The *tool coordinate system* is a coordinate system where the origin is in the MWD tool at all times. This is a Cartesian coordinate system as well, where the Z axis points in the direction of the well path (in the drilling direction), while the X and Y axes are fixed (with respect to the tool) in two directions perpendicular on Z, as depicted in Figure 2.4.

The *high-side direction* is defined as the projection of the negative V direction

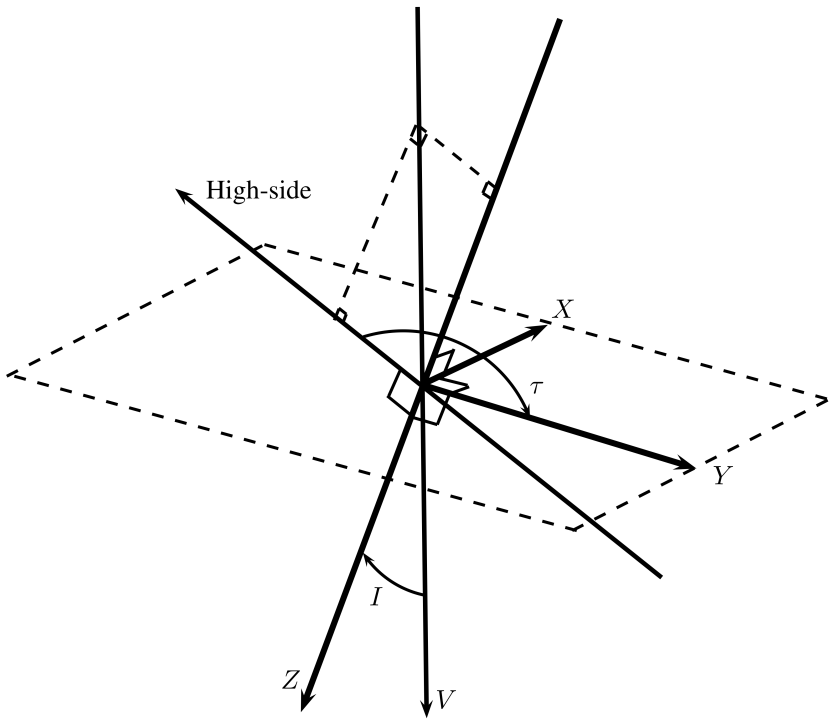

Figure 2.4: The tool coordinate system with the definition of the toolface angle τ and the high-side direction. The inclination angle I is also included.

on the XY plane, as shown in figure 2.4. The *toolface* angle τ is then defined as the angle from the high-side direction vector to the positive Y axis. The toolface angle is useful because the tool rotates during drilling. When we know the toolface angle τ, we also know the directions of the X and Y axes, since these are fixed with respect to the tool. We need to know these directions, since the magnetometer and accelerometer instruments measure values that are decomposed in the X, Y and Z directions. Note that the toolface angle is undefined for wells that are exactly vertical. We will therefore assume that the given well data are such that the wells are never *exactly* vertical, which should be a reasonable assumption in a practical situation.

2.2 Measurements and Error Sources

The accelerometer sensors in the MWD tool measure the gravity field $\mathbf{G}$, decomposed in the three directions of the tool coordinate system, giving G_x, G_y and G_z. Similarly, the magnetometer sensors measure the magnetic field $\mathbf{B}$ in three directions, giving B_x, B_y and B_z. From the decomposed values, the values of the

total fields are given by

$$\begin{aligned} G &= \sqrt{G_x^2 + G_y^2 + G_z^2}, \\ B &= \sqrt{B_x^2 + B_y^2 + B_z^2}. \end{aligned}$$

The direction of the gravity field is in the downward vertical V direction. The direction of the magnetic field is defined by a *magnetic dip angle* θ and a *declination angle* δ, as described in figure 2.5. The dip angle is defined as the angle from the geographical north N to the *magnetic north* N_m, which is a direction in the NE plane. The declination angle is the angle from N_m to the magnetic field vector.

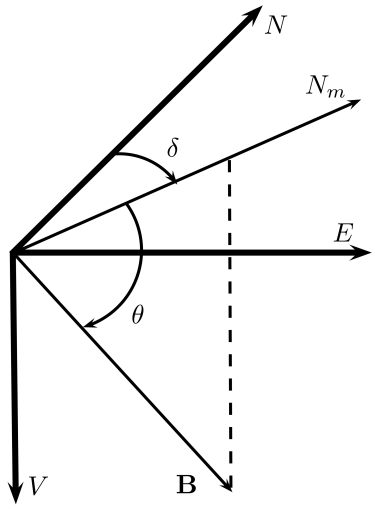

Figure 2.5: The declination angle δ gives the magnetic north N_m, while N_m and the magnetic dip angle θ gives the direction of the magnetic field $\mathbf{B}$.

Another measurement that is made is the *measured depth* D, defined as the distance from the surface to the survey point when one follows the well path. In other words, D is the *along-hole distance* from the surface to the point of interest.

For all these measurements, there are several types of error sources that need to be considered when we eventually want to find the statistical distribution of the NEV positions in the survey points. The 19 MWD errors that makes up the error model that we will use in the forthcoming example cases in this thesis are listed in Table 2.1. The reasoning behind the different error sources is described in Appendix B. We assume that all the errors can be categorized as one of the three *propagation modes* named *random*, *systematic* and *global*. These three types of errors will be explained shortly. We also assume that the 19 different error sources are independent and normally distributed with mean equal to zero.

First, a short note on notation. Consider a d-variate normal distributed vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^t$, where superscript t denotes that a matrix is *transposed*. When

Table 2.1: The error sources $\varepsilon_i, i = 1, \ldots, 19$, used in this thesis. The propagation mode of an error is either random (R), systematic (S) or global (G).

Error number, i	Description	Standard deviation, σ_i	Propagation mode
1–3	G_x, G_y, G_z bias	0.0039 m/s^2	S
4–6	G_x, G_y, G_z scale	0.0005	S
7–9	B_x, B_y, B_z bias	70 nT	S
10–12	B_x, B_y, B_z scale	0.0016	S
13	Depth reference	0.35 m	R
14	Depth scale factor	$6 \cdot 10^{-4}$	S
15	Depth stretch type	$2.5 \cdot 10^{-7}$ m^{-1}	G
16	Declination	0.36°	G
17	Declination B dependent	5000° nT	G
18	Sag	0.08°	S
19	Axial magnetism in string	150 nT	S

$\mathbf{x}$ has mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$, we write

$$\mathbf{x} \sim N_d(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x).$$

Also, let $\mathbf{I}_n$ denote the $n \times n$ sized *identity matrix*.

A *random error* is random between all survey points in the well. In other words, a random error source i gives an error $\varepsilon_{i,k}^R$ at each survey point $k = 0, \ldots, K$, but the error at one point is independent of the errors at all the other points in the well. Their joint distribution is given by

$$\boldsymbol{\varepsilon}_i^R = \left[\varepsilon_{i,0}^R, \varepsilon_{i,1}^R, \ldots, \varepsilon_{i,K}^R\right]^t \sim N_{K+1}\left(\left[0, 0, \ldots, 0\right]^t, \boldsymbol{\Sigma}_{\varepsilon,i}\right), \tag{2.1}$$

with covariance matrix of size $(K+1) \times (K+1)$ given by $\boldsymbol{\Sigma}_{\varepsilon,i} = \sigma_i^2 \mathbf{I}_{K+1}$.

A *systematic error* (or *survey-specific error*) is random between the different survey series $l = 1, \ldots, L$ in the well. Within one series l, the error $\varepsilon_{i,l}^S$ is the same for all points, but the errors for points in different series are independent. These errors are distributed as

$$\boldsymbol{\varepsilon}_i^S = \left[\varepsilon_{i,1}^S, \varepsilon_{i,2}^S, \ldots, \varepsilon_{i,L}^S\right]^t \sim N_L\left(\left[0, 0, \ldots, 0\right]^t, \boldsymbol{\Sigma}_{\varepsilon,i}\right), \tag{2.2}$$

with the $L \times L$ sized covariance matrix $\boldsymbol{\Sigma}_{\varepsilon,i} = \sigma_i^2 \mathbf{I}_L$.

A *global error* is the same for the whole well, or more generally, it is the same for the whole geographical area surrounding the well. Neighboring wells are therefore influenced by the exact same error. The distribution of a global error is simply given by

$$\varepsilon_i^G \sim N_1\left(0, \boldsymbol{\Sigma}_{\varepsilon,i}\right), \tag{2.3}$$

where the covariance "matrix" now equals the scalar value σ_i^2 known as the *variance* of the distribution, that is $\boldsymbol{\Sigma}_{\varepsilon,i} = \sigma_i^2$.

A useful observation is that the random propagation mode and the global propagation mode in fact are special cases of the systematic propagation mode: A systematic error is a random error when there are $K+1$ survey series (equal to the number of survey points), while it is a global error when there is only one survey series. This means that a random error and a global error will propagate to the *DIA* values in the same way as a systematic error. When we consider a random error, we just imagine that there are $K+1$ survey series with one survey point in each. When we consider a global error, we imagine that there is only one survey series containing $K+1$ measurements. Then we can make the calculations for any error as if it is a systematic error.

Since all the 19 errors in Table 2.1 are assumed to be independent, we can combine their distributions given by (2.1) - (2.3) into the joint distribution

$$\left[\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \ldots, \boldsymbol{\varepsilon}_{19}\right]^t \sim N_d\left(\left[0, 0, \ldots, 0\right]^t, \boldsymbol{\Sigma}_\varepsilon\right), \tag{2.4}$$

where the covariance matrix is

$$\boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} \boldsymbol{\Sigma}_{\varepsilon,1} & & & \\ & \boldsymbol{\Sigma}_{\varepsilon,2} & & \\ & & \ddots & \\ & & & \boldsymbol{\Sigma}_{\varepsilon,19} \end{bmatrix}.$$

Here we have disregarded the the superscripts R, S and G on the $\boldsymbol{\varepsilon}_i$'s. The dimension d is given by the sum of the dimensions of the 19 distributions.

2.3 The *DIA* Values and Their Distribution

From the geometry in Figure 2.4 we can find expression for the gravity components that include the inclination I and the toolface τ,

$$G_x = -G \sin I \sin \tau, \tag{2.5}$$
$$G_y = -G \sin I \cos \tau, \tag{2.6}$$
$$G_z = G \cos I. \tag{2.7}$$

By solving equations (2.5) - (2.7) for the inclination angle I, we obtain

$$I = \arctan\left(\frac{\sqrt{G_x^2 + G_y^2}}{G_z}\right). \tag{2.8}$$

We use equations (2.5) - (2.6) to find an expression for the toolface angle,

$$\tau = \arctan\left(\frac{-G_x}{-G_y}\right).$$

Williamson (2000) show that the magnetic components are given by

$$B_x = B\left(\cos\theta\cos I\cos A_m\sin\tau - \sin\theta\sin I\sin\tau + \cos\theta\sin A_m\cos\tau\right) \quad (2.9)$$
$$B_y = B\left(\cos\theta\cos I\cos A_m\cos\tau - \sin\theta\sin I\cos\tau - \cos\theta\sin A_m\sin\tau\right) \quad (2.10)$$
$$B_z = B\left(\cos\theta\cos I\cos A_m + \sin\theta\cos I\right) \quad (2.11)$$

The *magnetic azimuth* A_m is defined as the angle from the *magnetic north* N_m, as seen in figure 2.5, to the vector projection on the NE plane. This is the same definition as for the original azimuth A, except that N is now replaced with N_m. From the expressions (2.9) - (2.11), the magnetic azimuth can be shown (Gjerde, 2008) to be

$$A_m = \arctan\left(\frac{B_x\cos\tau - B_y\sin\tau}{[B_x\sin\tau + B_y\cos\tau]\cos I + B_z\sin I}\right).$$

By combining the declination angle δ and a grid correction (because of meridian convergence) into an estimated correction value δ_{corr}, we have that (Gjerde, 2008)

$$A = \delta_{\text{corr}} + A_m. \quad (2.12)$$

In summary, we need the measured values G_x, G_y, G_z, B_x, B_y and B_z, in addition to given reference values for the angles δ and δ_{corr} as we describe in Appendix C.3, in order to calculate I and A from the expressions (2.8) and (2.12).

Now we will find the approximate statistical distribution of the *DIA* values. Let all the *DIA* values of the $K+1$ survey points be contained in the vector

$$\boldsymbol{\alpha} = \left[D_0, I_0, A_0, \ldots, D_K, I_K, A_K\right]^t. \quad (2.13)$$

By non-linear error propagation theory, as shown in Appendix A, the joint distribution of $\boldsymbol{\alpha}$ is approximated by (A.5), that is

$$\boldsymbol{\alpha} \sim N_{(K+1)\times 3}\left(\left[\mu_{D_0}, \mu_{I_0}, \mu_{A_0}, \ldots, \mu_{D_K}, \mu_{I_K}, \mu_{A_K}\right]^t,\ \boldsymbol{\Sigma}_{DIA}\right), \quad (2.14)$$

with covariance matrix

$$\boldsymbol{\Sigma}_{DIA} = \mathbf{J}_{DIA}\boldsymbol{\Sigma}_{\varepsilon}\mathbf{J}_{DIA}^t,$$

which will be explained in more details shortly.

The mean vector in (2.14) consists of the expected depth values $\mu_{D_k}, k \in \{0, ..., K\}$, and the calculated values μ_{I_k} and μ_{A_k} obtained from (2.8) and (2.12) if we could have used the expected values of the variables on the right-hand sides. We can estimate μ_{D_k} by using measured values, while we can estimate μ_{I_k} and μ_{A_k} by calculating (2.8) and (2.12) using measured values on the right-hand sides.

The covariance matrix $\boldsymbol{\Sigma}_{DIA}$ contains linear combinations of weighted elements of the covariance matrix $\boldsymbol{\Sigma}_{\varepsilon}$, given in (2.4). The *weights* are given by the elements in the Jacobian matrix $\mathbf{J}_{DIA}$. From (A.4), the Jacobian matrix is given by

$$\mathbf{J}_{DIA} = \left[\mathbf{W}_1, \mathbf{W}_2, \ldots, \mathbf{W}_{19}\right]^t, \quad (2.15)$$

where

$$\mathbf{W}_i = \begin{bmatrix} \mathbf{w}_{i,1}^0 & & & & \\ \mathbf{w}_{i,1}^1 & & & & \\ \vdots & & & & \\ \mathbf{w}_{i,1}^{n_1} & & & & \\ & \mathbf{w}_{i,2}^{n_1+1} & & & \\ & \vdots & & & \\ & \mathbf{w}_{i,2}^{n_1+n_2} & & & \\ & & \ddots & & \\ & & & \mathbf{w}_{i,L}^{n_1+\cdots+n_{L-1}+1} & \\ & & & \vdots & \\ & & & \mathbf{w}_{i,L}^{K} \end{bmatrix}, \tag{2.16}$$

for $i \in \{1, \ldots, 19\}$. Here, n_l is the number of survey points in survey series l, when we consider a systematic error source. From the observations in Section 2.2, $n_l = 1$ and $L = K + 1$ when error i is random, while $n_l = K + 1$ and $L = 1$ when it is global. The blank parts of the matrix indicate zeros.

The weighting function $\mathbf{w}_{i,l}^k$ for measurement station number k (which belongs to survey number l) is given by

$$\mathbf{w}_{i,l}^k = \begin{bmatrix} \dfrac{\partial D_k}{\partial \varepsilon_{i,l}} & \dfrac{\partial I_k}{\partial \varepsilon_{i,l}} & \dfrac{\partial A_k}{\partial \varepsilon_{i,l}} \end{bmatrix}^t, \tag{2.17}$$

where all the derivatives are evaluated at the expectation of the variables of the differentiated functions, corresponding to the matrix elements in (A.4). The weighting functions $\mathbf{w}_{i,l}^k, i \in \{1, \ldots, 19\}$, can be found in Appendix C.1.

2.4 The *NEV* Positions and Their Distribution

In this section, we will show how to find the *NEV* position coordinates $\mathbf{p}_k = [N_k, E_k, V_k]^t$ of a survey point $k \in \{1, \ldots, K\}$ in a well, by using a *minimum curvature method*, as derived by Sawaryn and Thorogood (2003). In such a method, two neighboring survey points $k - 1$ and k are assumed to lie on a circular arc in the three dimensional space, as shown in Figure 2.6.

By using the known *NEV* position $\mathbf{p}_0$ of an initial point, which for instance can be given at the water surface, together with the *DIA* values $\boldsymbol{\alpha}$ in (2.13), the *NEV* positions can be calculated recursively by

$$\mathbf{p}_k = \mathbf{p}_{k-1} + \frac{\Delta D f(\kappa)}{2} \begin{bmatrix} \sin I_{k-1} \cos A_{k-1} + \sin I_k \cos A_k \\ \sin I_{k-1} \sin A_{k-1} + \sin I_k \sin A_k \\ \cos I_{k-1} + \cos I_k \end{bmatrix}, \tag{2.18}$$

where

$$\Delta D = D_k - D_{k-1}, \qquad f(\kappa) = \frac{\tan(\kappa/2)}{\kappa/2},$$

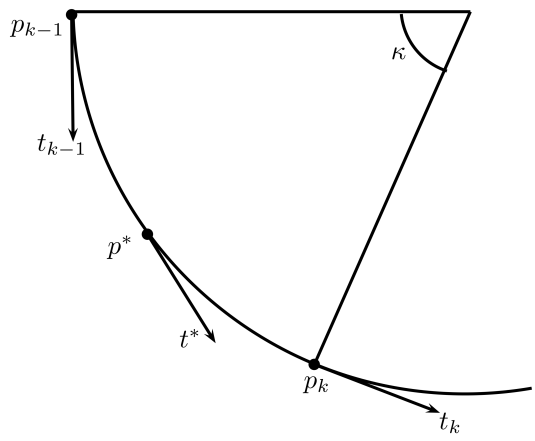

Figure 2.6: The minimum curvature method assumes that the well path between two survey points $k-1$ and k is a circular arc. Here, $\mathbf{p}_j$ is the *NEV* coordinates of point $j \in \{k-1, k\}$, $\mathbf{t}_j$ is the direction vector in point j, and κ is the dogleg angle. In addition, $\mathbf{p}^*$ is the position and $\mathbf{t}^*$ is the direction vector of some point between the survey points.

$$\kappa = 2 \arcsin \left(\sqrt{\sin^2 \left(\frac{I_k - I_{k-1}}{2} \right) + \sin I_{k-1} \sin I_k \sin^2 \left(\frac{A_k - A_{k-1}}{2} \right)} \right). \quad (2.19)$$

The value κ, expressed in (2.19), is known as the *dogleg angle*. This angle is defined as the change in inclination from point $k-1$ to point k, as shown in Figure 2.6.

We will now find the approximated statistical distribution of the estimated *NEV* positions. Let all the *NEV* coordinates of the unknown K survey points be contained in the vector

$$\mathbf{p} = \left[\mathbf{p}_1^t \dots \mathbf{p}_K^t \right]^t = \left[N_1, E_1, V_1, \dots, N_K, E_K, V_K \right]^t.$$

For simplicity, we will use the differences $\Delta \mathbf{p}_k = \mathbf{p}_k - \mathbf{p}_{k-1}, k \in 1, \dots, K$. We will now assume that the dogleg angle κ is small, so that we can use the limit formula

$$\lim_{x \to 0} \frac{\tan x}{x} = 1,$$

implying the approximation $f(\kappa) \approx 1$ in expression (2.18). We then have

$$\Delta \mathbf{p}_k = \frac{D_k - D_{k-1}}{2} \begin{bmatrix} \sin I_k \cos A_k + \sin I_{k-1} \cos A_{k-1} \\ \sin I_k \sin A_k + \sin I_{k-1} \sin A_{k-1} \\ \cos I_k + \cos I_{k-1} \end{bmatrix}. \quad (2.20)$$

From the error propagation theory, the distribution of the position differences is approximated by

$$\left[\Delta \mathbf{p}_1, \Delta \mathbf{p}_2, \dots, \Delta \mathbf{p}_K \right]^t \sim N_{K \times 3} \left(\left[\Delta \boldsymbol{\mu}_{p_1}, \Delta \boldsymbol{\mu}_{p_2}, \dots, \Delta \boldsymbol{\mu}_{p_K} \right]^t, \boldsymbol{\Sigma}_{\Delta p} \right),$$

with covariance matrix

$$\mathbf{\Sigma}_{\Delta p} = \mathbf{J}_{\Delta p}\mathbf{\Sigma}_{DIA}\mathbf{J}_{\Delta p}^{t}.$$

The Jacobian matrix is

$$\mathbf{J}_{\Delta p} = \begin{bmatrix} \mathbf{R}_{1,0} & \mathbf{R}_{1,1} & & & \\ & \mathbf{R}_{2,1} & \mathbf{R}_{2,2} & & \\ & & \ddots & \ddots & \\ & & & \mathbf{R}_{K,K-1} & \mathbf{R}_{K,K} \end{bmatrix}, \tag{2.21}$$

where we have defined the weighting functions

$$\mathbf{R}_{j,k} = \begin{bmatrix} \dfrac{\partial \Delta \mathbf{p}_j}{\partial D_k} & \dfrac{\partial \Delta \mathbf{p}_j}{\partial I_k} & \dfrac{\partial \Delta \mathbf{p}_j}{\partial A_k} \end{bmatrix}.$$

The elements of $\mathbf{R}_{j,k}$ are given explicitly in Appendix C.2.

The relationship between the *NEV* positions $\mathbf{p}_k$ and the differences $\Delta\mathbf{p}_k$ is simply

$$\mathbf{p}_k = \mathbf{p}_0 + \sum_{m=1}^{k} \Delta \mathbf{p}_m. \tag{2.22}$$

We have assumed that the known initial position $\mathbf{p}_0$ has no uncertainty. By using the error propagation theory one last time we find the approximate joint position distribution

$$\left[\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_k\right]^t \sim N_{K\times 3}\left(\left[\boldsymbol{\mu}_{p_1}, \boldsymbol{\mu}_{p_2}, \ldots, \boldsymbol{\mu}_{p_k}\right]^t, \mathbf{\Sigma}_{NEV}\right), \tag{2.23}$$

with covariance matrix

$$\mathbf{\Sigma}_{NEV} = \mathbf{J}_p\mathbf{\Sigma}_{\Delta p}\mathbf{J}_p^t.$$

From (2.22) the Jacobain matrix in (2.23) is

$$\mathbf{J}_p = \begin{bmatrix} \mathbf{I}_3 & & & \\ \mathbf{I}_3 & \mathbf{I}_3 & & \\ \vdots & \vdots & \ddots & \\ \mathbf{I}_3 & \mathbf{I}_3 & \ldots & \mathbf{I}_3 \end{bmatrix}. \tag{2.24}$$

In sum, the total covariance matrix for all the *NEV* postions $\mathbf{p}_k$ in the survey points $k \in \{1, \ldots, K\}$ in a well is given by

$$\mathbf{\Sigma}_{NEV} = \mathbf{J}_p\mathbf{J}_{\Delta p}\mathbf{J}_{DIA}\mathbf{\Sigma}_{\varepsilon}\mathbf{J}_{DIA}^t\mathbf{J}_{\Delta p}^t\mathbf{J}_p^t. \tag{2.25}$$

We can also obtain the position $\mathbf{p}^*$ of a point between two survey points $k-1$ and k, as seen in Figure 2.6, by using the interpolation formula (Sawaryn and Thorogood, 2003) given by

$$\mathbf{p}^* = \mathbf{p}_{k-1} + \frac{\Delta D^* f\left(\kappa \frac{\Delta D^*}{\Delta D}\right)}{2}\left(\mathbf{t}_{k-1} + \mathbf{t}^*\right). \tag{2.26}$$

Here, $\mathbf{t}_{k-1}$ and $\mathbf{t}^*$ are two direction vectors, as shown in Figure 2.6. The vector $\mathbf{t}_{k-1}$ is given by

$$\mathbf{t}_{k-1} = \begin{bmatrix} \sin I_{k-1} \cos A_{k-1} \\ \sin I_{k-1} \sin A_{k-1} \\ \cos I_{k-1} \end{bmatrix},$$

while the the direction vector corresponding to the unknown position is

$$\mathbf{t}^* = \frac{\sin\left(\left[1 - \frac{\Delta D^*}{\Delta D}\right]\right)}{\sin \kappa}\mathbf{t}_{k-1} + \frac{\sin\left(\frac{\Delta D^*}{\Delta D}\kappa\right)}{\sin \kappa}\mathbf{t}_k. \tag{2.27}$$

In addition, ΔD^* is the difference in measured depth between the middle point and survey point $k-1$.

2.5 Several Wells

The method described so far can be generalized in order to obtain a multivariate statistical distribution that includes two or more wells. We will follow the same procedure as for one well. In short, the procedure is described by

$$\mathbf{\Sigma}_{\varepsilon}^{\text{total}} \longrightarrow \mathbf{\Sigma}_{DIA}^{\text{total}} \longrightarrow \mathbf{\Sigma}_{NEV}^{\text{total}},$$

where the covariance matrices in the previous sections are generalized in order to be applied to several wells, as indicated by the superscript *total*.

The covariance matrix associated with the error sources in Table 2.1 for M wells is given by

$$\mathbf{\Sigma}_{\varepsilon}^{\text{total}} = \begin{bmatrix} \mathbf{\Sigma}_{\varepsilon;R,S}^{1} & & & \\ & \ddots & & \\ & & \mathbf{\Sigma}_{\varepsilon;R,S}^{M} & \\ & & & \mathbf{\Sigma}_{\varepsilon;G} \end{bmatrix} \tag{2.28}$$

where $\mathbf{\Sigma}_{\varepsilon;R,S}^{m}$ corresponds to the measurment error covariance matrix for well m as given in (2.4), except that the covariance values for global errors are not included in these matrices. The global error covariances, combined into a diagonal matrix $\mathbf{\Sigma}_{\varepsilon;G}$, are instead placed in the bottom of $\mathbf{\Sigma}_{\varepsilon}^{\text{total}}$. These error sources should only be included once, since the global property implies that the error is the same for all wells.

2.5.1 Distribution of the *DIA* Values for M Wells

We now let all the *DIA* values be contained in $[\boldsymbol{\alpha}^1, \ldots, \boldsymbol{\alpha}^M]^t$, where

$$\boldsymbol{\alpha}^m = \left[D_1^m, I_1^m, A_1^m, \ldots, D_{K_m}^m, I_{K_m}^m, A_{K_m}^m\right]$$

contains the *DIA* values for well m, with $K_m + 1$ equaling the number of survey points in well m. The total distribution is given by

$$\left[\boldsymbol{\alpha}^1, \boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^M\right]^t \sim N_d\left(\left[\boldsymbol{\mu}_{\alpha^1}, \boldsymbol{\mu}_{\alpha^2}, \ldots, \boldsymbol{\mu}_{\alpha^M}\right]^t,\ \boldsymbol{\Sigma}_{DIA}^{\text{total}}\right), \tag{2.29}$$

with covariance matrix

$$\boldsymbol{\Sigma}_{DIA}^{\text{total}} = \mathbf{J}_{DIA}^{\text{total}} \boldsymbol{\Sigma}_{\varepsilon}^{\text{total}} \mathbf{J}_{DIA}^{\text{total},t}. \tag{2.30}$$

The dimension is $d = 3 \times \sum_m (K_m + 1)$, and $\boldsymbol{\mu}_{\alpha^m}$ is the expectation vector for the values in well m. The Jacobian matrix is given by

$$\mathbf{J}_{DIA}^{\text{total}} = \begin{bmatrix} \mathbf{J}_{DIA;R,S}^1 & & & \mathbf{W}_G^1 \\ & \ddots & & \vdots \\ & & \mathbf{J}_{DIA;R,S}^M & \mathbf{W}_G^M \end{bmatrix}.$$

The matrices $\mathbf{J}_{DIA;R,S}^m$ corresponds to the Jacobian matrix for well m as given in (2.15), except that the weighting matrices $\mathbf{W}_i$ that correspond to global errors are removed. These are instead placed at the right side of $\mathbf{J}_{DIA}^{\text{total}}$, in order to correspond with the placement of the covariance values in $\boldsymbol{\Sigma}_{\varepsilon}^{\text{total}}$, as shown in (2.28). If we let n_g equal the number of global error sources, then a weighting matrix $\mathbf{W}_G^m$ in $\mathbf{J}_{DIA}^{\text{total}}$, for $m \in \{1, \ldots, M\}$, is given by

$$\mathbf{W}_G^m = \left[\mathbf{W}_{G,1}^m, \ldots, \mathbf{W}_{G,n_g}^m\right],$$

where the blocks are the weighting matrices obtained from (2.16), which were used when we considered one well and one error source i.

The joint distribution of all the *DIA* values is therefore given by (2.29), with covariance matrix given by (2.30).

2.5.2 Distribution of the *NEV* Positions for M Wells

Following the procedure in the case a single well, we now let all the *NEV* position differences be contained in $[\Delta\mathbf{p}^1, \ldots, \Delta\mathbf{p}^M]^t$, where

$$\Delta\mathbf{p}^m = \left[\Delta\mathbf{p}_1^m, \ldots, \Delta\mathbf{p}_{K_M}^m\right]$$

contains the position difference vectors of well m. The distribution is approximated by

$$\left[\Delta\mathbf{p}^1, \Delta\mathbf{p}^2, \ldots, \Delta\mathbf{p}^M\right]^t \sim N_d\left(\left[\boldsymbol{\mu}_{\Delta p^1}, \boldsymbol{\mu}_{\Delta p^2}, \ldots, \boldsymbol{\mu}_{\Delta p^M}\right]^t,\ \boldsymbol{\Sigma}_{\Delta p}^{\text{total}}\right), \tag{2.31}$$

with covariance matrix

$$\boldsymbol{\Sigma}_{\Delta p}^{\text{total}} = \mathbf{J}_{\Delta p}^{\text{total}} \boldsymbol{\Sigma}_{DIA}^{\text{total}} \mathbf{J}_{\Delta p}^{\text{total},t}.$$

The dimension is $d = 3 \times \sum_m (K_m + 1)$, and the weighting Jacobian matrix is

$$\mathbf{J}_{\Delta p}^{\text{total}} = \begin{bmatrix} \mathbf{J}_{\Delta p}^1 & & & \\ & \mathbf{J}_{\Delta p}^2 & & \\ & & \ddots & \\ & & & \mathbf{J}_{\Delta p}^M \end{bmatrix}$$

where the blocks equals the Jacobian matrices when considering only one well, as shown in (2.21).

Finally, we let all the *NEV* positions for all the M wells be contained in $[\mathbf{p}^1, \ldots, \mathbf{p}^M]^t$, where

$$\mathbf{p}^m = \left[\mathbf{p}_1^m, \ldots, \mathbf{p}_{K_M}^m\right]$$

contains the position vectors of well m. The distribution is approximated by

$$\left[\mathbf{p}^1, \mathbf{p}^2, \ldots, \mathbf{p}^M\right]^t \sim N_d\left(\left[\boldsymbol{\mu}_{p^1}, \boldsymbol{\mu}_{p^2}, \ldots, \boldsymbol{\mu}_{p^M}\right]^t, \ \boldsymbol{\Sigma}_{NEV}^{\text{total}}\right), \tag{2.32}$$

with dimension $d = 3 \times \sum_m (K_m + 1)$ and covariance matrix

$$\boldsymbol{\Sigma}_{NEV}^{\text{total}} = \mathbf{J}_p^{\text{total}} \boldsymbol{\Sigma}_{\Delta p}^{\text{total}} \mathbf{J}_p^{\text{total},t}.$$

The weighting Jacobian matrix is

$$\mathbf{J}_p^{\text{total}} = \begin{bmatrix} \mathbf{J}_p^1 & & & \\ & \mathbf{J}_p^2 & & \\ & & \ddots & \\ & & & \mathbf{J}_p^M \end{bmatrix} \tag{2.33}$$

where the block matrices can be found from (2.24).

We have now obtained that the covariance matrix for the *NEV* positions of all the survey points in M wells can be expressed as

$$\boldsymbol{\Sigma}_{NEV}^{\text{total}} = \mathbf{J}_p^{\text{total}} \mathbf{J}_{\Delta p}^{\text{total}} \mathbf{J}_{DIA}^{\text{total}} \boldsymbol{\Sigma}_{\varepsilon}^{\text{total}} \mathbf{J}_{DIA}^{\text{total},t} \mathbf{J}_{\Delta p}^{\text{total},t} \mathbf{J}_p^{\text{total},t}. \tag{2.34}$$

2.5.3 An Example With Two Wells

In Figure 2.7 we see the well path of two real petroleum wells from the North Sea, along with the *NEV* axes. They are seen to be very close to each other down to about 1000 meters below sea level. For both wells the measurements are made with three different survey tools, which means that each survey point belongs to one of three survey series. The blue reference well has 203 survey stations, while the red offset well has 193 survey stations. We denote the vectors containing the *NEV* coordinates of the two wells by $\mathbf{p}$ and $\mathbf{q}$ respectively.

We can find the covariance matrix $\boldsymbol{\Sigma}_{NEV}^{\text{total}}$ for all the *NEV* coordinates

$$[\mathbf{p}^t, \mathbf{q}^t] = \left[N_1^p, E_1^p, V_1^p, \ldots, N_{203}^p, E_{203}^p, V_{203}^p, N_1^q, E_1^q, V_1^q, \ldots, N_{193}^q, E_{193}^q, V_{193}^q\right]$$

of the two wells by calculating expression (2.34), using given *DIA* values at survey points in the wells in the calculation of the Jacobian matrices.

In order to discuss the *correlation* between the different values, we find the corresponding *correlation matrix* $\mathbf{P}$. This matrix has the same size as the covariance matrix. Its elements are given by

$$P(i,j) = \frac{\Sigma_{NEV}^{\text{total}}(i,j)}{\sqrt{\Sigma_{NEV}^{\text{total}}(i,i)\Sigma_{NEV}^{\text{total}}(j,j)}}.$$

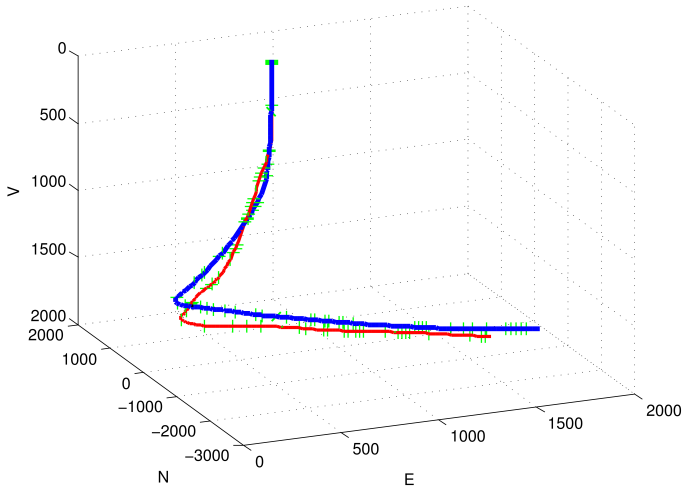

Figure 2.7: Two petroleum wells from the North Sea. All three axes has unit *meter*. The blue reference well (of which the end is the most to the east of the two) has 203 survey stations, while the red offset well has 193 survey stations. The reference well starts out at $\mathbf{p}_0 = [N_0^p, E_0^p, V_0^p]^t = [970.10, 816.50, 0]^t$ going straight down before hitting the seabed at $V_1^p = 349.5$. The offset well starts at $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [976.20, 817.70, 0]$, going straight down to $V_1^q = 336.50$. At the seabed level the wells go in the south-west direction. Later they turn and go in the south-east direction.

The calculated correlation matrix for the two wells is shown in Figure 2.8. Here, the correlation values are represented by colors. By the number of survey points in the wells, the dimension of the matrix is $3 \times (203 + 193) = 1188$, as seen on the horizontal axis. Here we multiplied by 3 because there are three coordinates (N, E and V) at each point. As we have marked on the vertical axis, the 609 first values in each row and column correspond to the NEV values of the first well, while the rest corresponds to the NEV values of the second well. For instance, element $\Sigma_{NEV}^{\text{total}}(2,4)$ shows the correlation between the coordinates E_1^p and N_2^p in the first well, while element $\Sigma_{NEV}^{\text{total}}(611,3)$ gives the correlation between E_1^q in the second well and V_1^p in the first well.

Focusing on the upper left quarter of the matrix, which gives the correlation values between the NEV values in the first well, we observe that there is some kind of change in behavior around row (and column) 130, and then again around row 340. This behavior is because of the systematic error sources in our error model. The correlation values from row 1 to about row 130 corresponds to NEV values calculated from measurements from the first survey series in the first well. The values from about row 130 to about row 240 corresponds to the second survey series, and so on. In general, the correlation values *within* the series are seen to

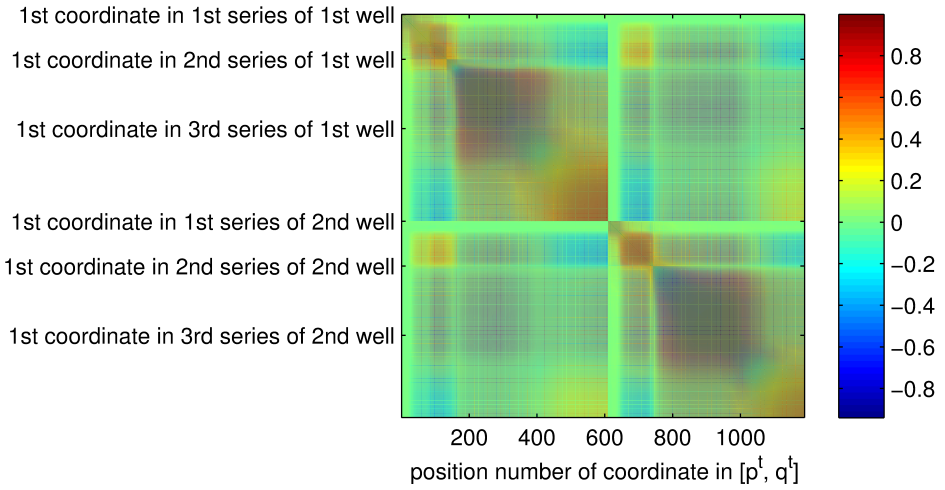

Figure 2.8: Correlation matrix for the *NEV* coordinates of the survey points in the two wells in Figure 2.7.

be higher than the correlation values between points that belongs to two different series. The same observation can be made for the second well, by considering the lower right quarter of the matrix.

Similar observations can be made for the upper right and lower left quarter of the correlation matrix. These parts contain correlation values between pair of points where one point belongs to the first well, and the other point belongs to the second well. These correlation values are different from zero only because of the global errors in the error model, and therefore they are in general smaller than the other correlation values originating from all kind of error sources.

We also include a plot of the variance values of the *NEV* coordinates in both wells in Figure 2.9, equaling the diagonal values of $\mathbf{\Sigma}_{NEV}^{\text{total}}$. In general, the variances tend to increase when the survey number increases. This is as expected, since we are moving further and further away from the known initial positions $\mathbf{p}_0$ and $\mathbf{q}_0$, which should intuitively contribute to an increase in the uncertainty. However, they do not always increase, as seen by the upper plots of the variance of the north coordinates. In this case, the decrease seems to occur for the survey points where the direction in the *NE* plane starts to stabilize, after the turn seen to the left of Figure 2.7. This is confirmed by Figure 2.10, where we plot the variance values against the azimuth values of the survey points. This might suggest that a stable behavior of a well geometry variable like the azimuth may cause a decrease in the uncertainty when we move downwards in a well.

The plots in this example intends to give some confirmation that the method shown in this chapter results in a reasonable covariance matrix for the *NEV* positions of survey points.

In this case we actually obtain a covariance matrix that seems to be close to singular. In order to be able to find the inverse of the matrix in later calculations[3] we follow the strategy of ignoring the most uninformative variables until we are able to find a useable pdf value. This method is thorougly described in Appendix D. In this case, we end up ignoring the uninformative variables that has correlation values above the value of 0.89.

[3]We will need the inverse in Chapter 3 when we perform a *cross-entropy method* (Sec. 3.2.3) which is based on the statistical distributions found in this chapter.

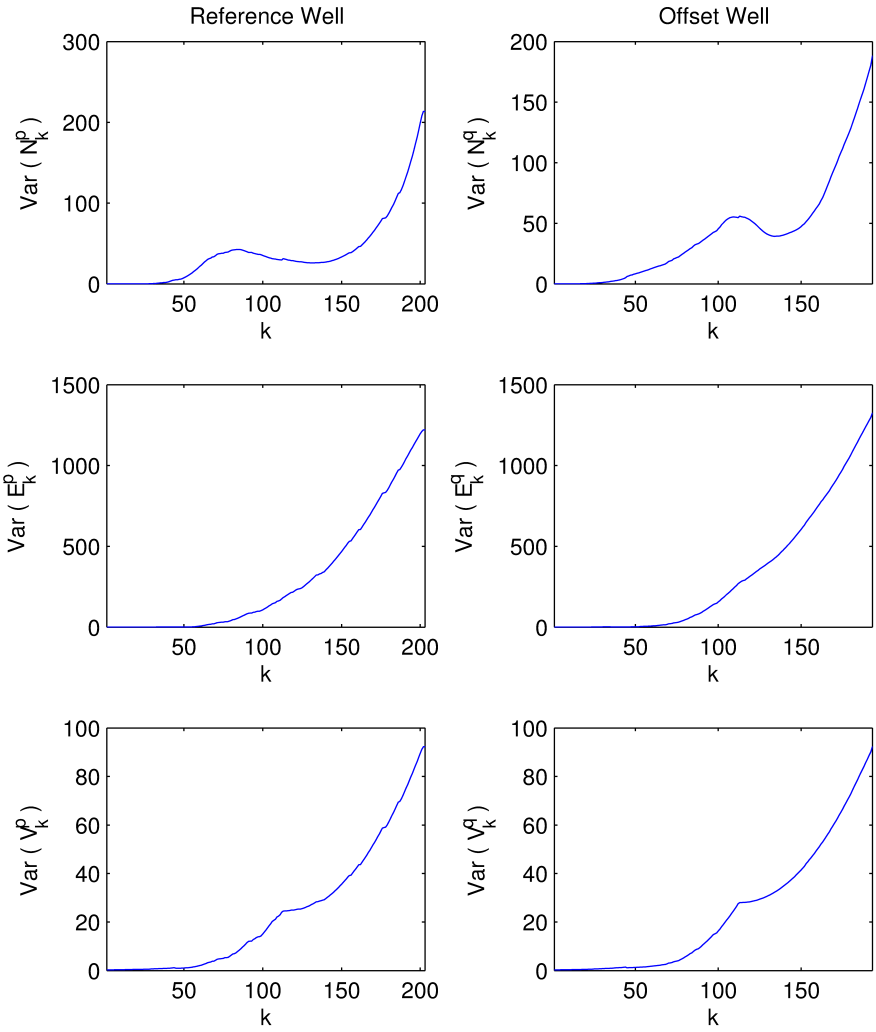

Figure 2.9: Variance values for the *NEV* coordinates of the survey points in the two wells in Figure 2.7.

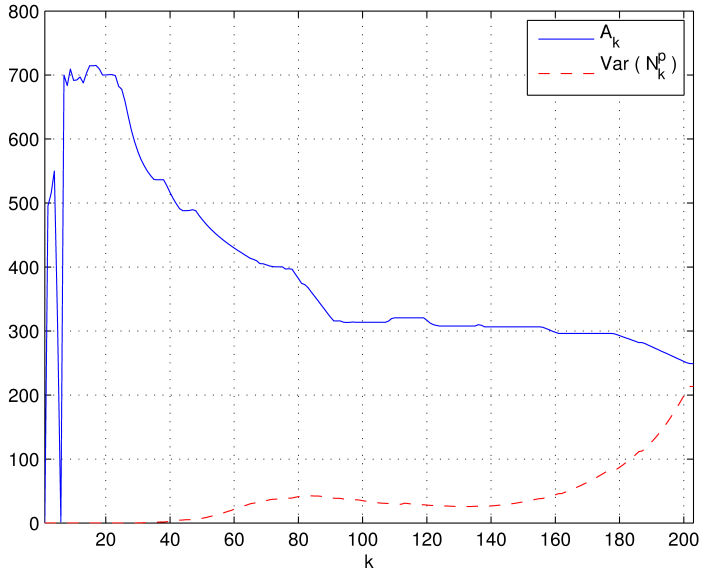

Figure 2.10: Variance values of the north coordinates (red, dashed) compared to the azimuth values (blue) of the survey points in the reference well in Figure 2.7.

Chapter 3

The Risk of Well Collision

When we drill a petroleum well in an area that contains other wells, we want to know the risk that a well collision occurs. If we are not able to avoid such a collision, the consequences could become disastrous. Our goal in this chapter is to describe various ways of addressing the risk of well collision.

Consider the setting where we drill (or plan to drill) a *reference well* close to an *offset well*. We define that a *well collision* happens if

$$r \leq r_1 + r_2, \tag{3.1}$$

where r is the *shortest distance* between the center lines of the two wells, and r_1 and r_2 denote the radii of the reference well and the offset well, respectively. Throughout this thesis, we will use the assumption that $r_1 + r_2 = 0.5 \text{ m} + 0.5 \text{ m} = 1 \text{ m}$, which is reasonable compared to many real situations.

We assume that the shortest distance r is a stochastic value. The value of r depends on the positions of the two wells, of which the *DIA* values and the *NEV* coordinates of the survey points are assumed to behave stochastically according to the statistical distributions (2.29) and (2.32) in Chapter 2. As explained thoroughly in that chapter, we are able to estimate the *NEV* coordinates of the survey points in the two wells by using MWD measurements in addition to some reference values. Let the measured *NEV* coordinates be contained in the position vector

$$[\mathbf{p}^t, \mathbf{q}^t] = \left[N_1^p, E_1^p, V_1^p, \ldots, N_{K_p}^p, E_{K_p}^p, V_{K_p}^p, N_1^q, E_1^q, V_1^q, \ldots, N_{K_q}^q, E_{K_q}^q, V_{K_q}^q\right], \tag{3.2}$$

where K_p and K_q denote the number of survey points in the two wells when we leave out the known initial points $\mathbf{p}_0 = [N_0^p, E_0^p, V_0^p]^t$ and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t$.

In the petroleum industry today, the common practice regarding collision avoidance calculations is to consider only the two closest points in the two wells (ISCWSA, 2011). We denote the *NEV* coordinates of these two points by $\mathbf{p}_c = [N_c^p, E_c^p, V_c^p]^t$ and $\mathbf{q}_c = [N_c^q, E_c^q, V_c^q]^t$. The situation is depicted in Figure 3.1.

The closest points are found by calculating the *NEV* positions of all the survey points in both wells, as expressed by the vector in expression (3.2), possibly in addition to some interpolated points, as described in Section 2.4. We then use

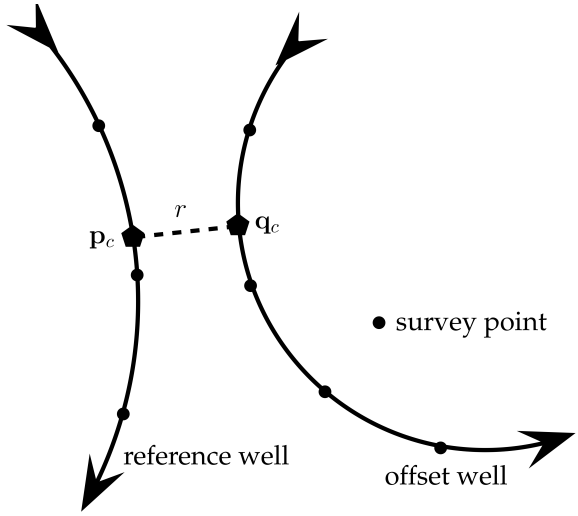

Figure 3.1: A simple sketch of a reference well being drilled close to an offset well. The shortest distance between the wells is denoted by r, while $\mathbf{p}_c = [N_c^p, E_c^p, V_c^p]^t$ and $\mathbf{q}_c = [N_c^q, E_c^q, V_c^q]^t$ are the two closest points.

some method to find the two points in the wells that are closest to each other. Some of these methods are described by ISCWSA (2011) and Gjerde (2008). In the most common method, called *3D closest approach*, we simply estimate the two points that have the shortest Euclidean distance between them. We describe an approximated 3D closest approach method in Appendix E.

One way to draw conclusions regarding the risk of a collision taking place, is to perform a *hypothesis test*. This is the common practice in the industry today. The uncertainty in the measured two closest points is used to perform a test where the null hypothesis states that the expected value μ_r of the shortest distance r leads to a collision, that is

$$H_0 : \qquad \mu_r \leq r_1 + r_2.$$

The hypothesis test will then reveal whether or not we can trust our measurements enough to continue the drilling, based on a chosen significance level. We decide to drill the well only if the null hypothesis is rejected.

We describe two such tests in Section 3.1. In Section 3.1.1 we explore the standard hypothesis test, in which we approximate the distribution of the shortest distance r by a normal distribution, by using the error propagation theory in Appendix A. As described in Section 3.1.2, it turns out that it is also possible to do a more exact test, where the statistical distribution of the *squared distance* r^2 is found *without* the need of an additional normal approximation.

As an alternative to a hypothesis test, we describe how to estimate the *probability of well collision* in Section 3.2. This is the value defined by

$$l = P\left(\text{Well collision}\right) = P\left(r \leq r_1 + r_2\right).$$

An estimate of this value can be used to draw conclusions on whether the well plan should be changed, by comparing it to a predetermined maximum toleration value. In Section 3.2.1 we begin by calculating the probability that the positions of the two closest point are coinciding, in the sense that their distance is within the sum of the radii of the two wells, by following a similar analytical approach as for the more exact hypothesis test. The calculations are based on the statistical distribution of the *NEV* coordinates of the two points. We will also estimate the probability value by performing a *Monte Carlo (MC) simulation method* based on the same distribution.

Up to this point, we have introduced collision risk methods in which we only consider the two closest points. There is a drawback with this: The fact that the two points $\mathbf{p}_c$ and $\mathbf{q}_c$ are expected to be the closest ones, does *not* necessarily mean that there can not be other points (survey points or intermediate points) having uncertainty such that there is a non-zero probability that they too will coincide, which would give an additional contribution to the total collision probability. A method that analyzes the risk of collision should ideally take this into account. We aim to investigate this principle further in Section 3.2.2. There we use a MC simulation method where we simulate the *DIA* values of *all the points* in the two wells, instead of just using the position values of the two closest points. Finding an analytical solution, as we do when we consider only the two closest points, is outside scope of this thesis.

We illustrate the difference between the *two-points method* and the *all-points method* for estimating the collision probability in Figure 3.2. In the two-points MC simulation method we simulate B samples of the NEV coordinate of the two closest points, and calculate the percentage of samples where the simulated points coincide. In the all-points method we simulate B samples of the *DIA* values of all the survey points, and calculate the percentage of simulated well pairs where a collision occurs. The resulting percentages are the estimates of the collision probability for the two methods. The all-points method is obviously more time-consuming than the two-points method, but it should give a more correct result, since we exploit more of the available information about the wells.

An additional important challenge associated with the MC simulation methods for estimating the probability of collision is the fact that a collision is usually considered to be a *rare event*. In Section 3.2.3 we will discuss a *cross-entropy method* in order to make the simulation methods more effective in both the two-points situation and the all-points situation.

Another simulation method for rare events is the *enhanced Monte Carlo method*, which we discuss in Section 3.2.4. We only consider this method in the all-points paradigm in order to have an additional method for estimating the probability value in this situation, since we are not able to obtain any analytical value, which we are in the two-points situation.

In summary, we will look at the following collision risk methods in this chapter.

- Two hypothesis tests when considering the two closest points (Sec. 3.1). We present a normal approximated test (Sec. 3.1.1), which is used by the industry today. We also look at a more exact test (Sec. 3.1.2).

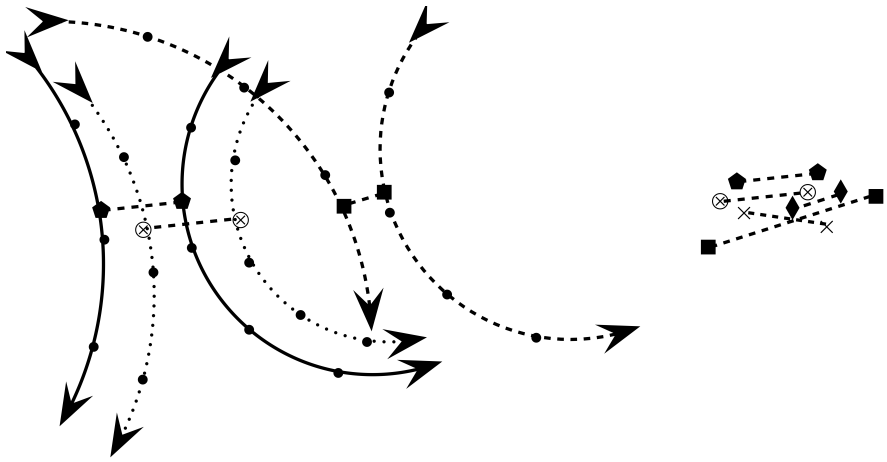

Figure 3.2: An illustration of the difference between the two-points MC simulation method and the all-points MC simulation method that are introduced in this chapter. To the right we have sketched $B = 5$ simulated pairs of points, based solely on the coordinates $\mathbf{p}_c$ and $\mathbf{q}_c$ of the two closest points in Figure 3.1. To the left we have sketched $B = 3$ simulated pairs of wells, based on the *DIA* values of the entire well segments in Figure 3.1. The three well pairs are drawn with different line types.

- An analytical calculation of the probability of collision when considering the two closest points (Sec. 3.2.1).
- Monte Carlo simulation of the probability of collision when considering the two closest points (Sec. 3.2.1).
- Monte Carlo simulation of the probability of collision when considering all the points in the two wells (Sec. 3.2.2).
- Aiming at increasing the accuracy in a rare-event situation, we also look at the cross-entropy method (Sec. 3.2.3) and the enhanced Monte Carlo method (Sec. 3.2.4).

In Chapter 4, we will perform all these collision risk methods on various pairs of petroleum wells, in order to obtain numerical results and discuss the methods more carefully.

3.1 Hypothesis Tests

Today, the industry practice on making conclusions about the possibility of a collision between two wells is to perform a hypothesis test on whether or not the expected positions of the wells implies a collision. The common practice (Williamson, 2000) is to only consider the two closest points in the two wells.

The common practice also includes a normal approximation of the distribution of the shortest distance r. An important disadvantage with this method is that the resulting normal distribution is defined for both positive and negative values, while r should always be non-negative. Gjerde (2008) argues that this approximation is valid when the collision probability is small, which usually is the case. However, in the interesting and important cases where the wells actually are very close, the collision probability might not be that small, which could possibly result in wrong conclusions.

Here, we therefore propose a new hypothesis test that does *not* make the approximation that the shortest distance itself is normal distributed. We first explain the original normal approximated test in Section 3.1.1. In Section 3.1.2, we then show the derivation of a more exact hypothesis test. Later, in Section 4.2.1, we compare the two tests performed on various well cases, in order to try to detect any major differences between them.

In both tests, the null hypothesis is that a collision between the two wells does take place when we consider the expected value of the shortest distance between the wells, that is

$$H_0 : \qquad \mu_r \leq r_1 + r_2,$$

while the alternative hypothesis is defined as

$$H_1 : \qquad \mu_r > r_1 + r_2.$$

The *significance level* α is chosen in order to ensure that

$$P\left(\text{Reject } H_0 \mid H_0 \text{ is true}\right) = \alpha. \tag{3.3}$$

For instance, a typical value to choose is $\alpha = 1/500$. Then we make sure that we reject the null hypothesis only once out of 500 times in the situation where the null hypothesis is in fact true. In other words, the probability that we decide to go on with an existing well plan that is expected to end in a collision is equal to α.

Let $\mathbf{p}_c = [N_c^p, E_c^p, V_c^p]^t$ and $\mathbf{q}_c = [N_c^q, E_c^q, V_c^q]^t$ contain the *NEV* coordinates of the two closest points. The estimated distance between the two points is given by the Euclidean distance

$$r = \sqrt{\left(N_c^p - N_c^q\right)^2 + (E_c^p - E_c^q)^2 + (V_c^p - V_c^q)^2}. \tag{3.4}$$

The vector $[\mathbf{p}_c^t, \mathbf{q}_c^t] = [N_c^p, E_c^p, V_c^p, N_c^q, E_c^q, V_c^q]^t$ is now assumed to be normal distributed as

$$\left[\mathbf{p}_c^t, \mathbf{q}_c^t\right] \sim N_6\left([\boldsymbol{\mu}_{p_c}^t, \boldsymbol{\mu}_{q_c}^t], \boldsymbol{\Sigma}_{p_c,q_c}\right). \tag{3.5}$$

This distribution is motivated by the marginal distribution of the two survey points succeeding $\mathbf{p}_c$ and $\mathbf{q}_c$ in the wells. *That* distribution is in turn obtained by simply ignoring the elements that corresponds to other coordinates in the expectation vector and covariance matrix of the total joint distribution in (2.32) of *all* the *NEV* coordinates $[\mathbf{p}^t, \mathbf{q}^t]$. We estimate the covariance matrix $\boldsymbol{\Sigma}_{p_c,q_c}$ by the marginal covariance matrix of the succeeding points. However, instead of estimating the expectation vector by the coordinates of the succeeding points, we use the measured or planned values of the closest points $[\mathbf{p}_c^t, \mathbf{q}_c^t]$, in order to obtain the distribution in (3.5).

3.1.1 The Approximated Test

As is the common practice in the petroleum industry, the statistical distribution of the shortest distance r can be approximated by using the error propagation theory in Appendix A. Following the error propagation theory, we have that r is *approximately* normal distributed as

$$r \sim N_1\left(\mu_r, \sigma_r^2\right) \tag{3.6}$$

with variance given by

$$\sigma_r^2 = \mathbf{J}_r \mathbf{\Sigma}_{p_c,q_c} \mathbf{J}_r^t. \tag{3.7}$$

When we approximate by using the succeeding survey points, the Jacobian matrix in (3.7) is given by

$$\begin{aligned}
\mathbf{J}_r &= \left[\frac{\partial r}{\partial N_c^p} \quad \frac{\partial r}{\partial E_c^p} \quad \frac{\partial r}{\partial V_c^p} \quad \frac{\partial r}{\partial N_c^q} \quad \frac{\partial r}{\partial E_c^q} \quad \frac{\partial r}{\partial E_c^q}\right] \\
&= \left[\frac{N_c^p - N_c^q}{r} \quad \frac{E_c^p - E_c^q}{r} \quad \frac{V_c^p - V_c^q}{r} \quad \frac{N_c^q - N_c^p}{r} \quad \frac{E_c^q - E_c^p}{r} \quad \frac{V_c^q - V_c^p}{r}\right].
\end{aligned}$$

In the first matrix the differentiated functions are evaluated in the measured or planned values $[\mathbf{p}_c^t, \mathbf{q}_c^t]$.

In a hypothesis test one rejects the null hypothesis when the *p-value* p is lower than the significance level, that is when $p < \alpha$. If $\hat{r}$ is our estimate of the shortest distance based on the measured or planned values $[\mathbf{p}_c^t, \mathbf{q}_c^t]$, then the p-value is defined to be the probability that a new obtained distance r in fact is as extreme as (or more extreme than) $\hat{r}$, given the null hypothesis H_0. In other words,

$$\begin{aligned}
p &= P\left(r \geq \hat{r} \mid \mu_r \leq r_1 + r_2\right) \\
&= P\left(\frac{r - \mu_r}{\sigma_r} \geq \frac{\hat{r} - \mu_r}{\sigma_r} \mid \mu_r = r_1 + r_2\right) \\
&= 1 - \Phi\left(\frac{\hat{r} - (r_1 + r_2)}{\sigma_r}\right),
\end{aligned}$$

where $\Phi(\cdot)$ is the known *cumulative distribution function (cdf)* of the standard normal distribution. In the second equation, we recognize that an equality in the conditional expression will cover the worst case, that is when it is most difficult to detect a collision.

The industry often follows the equivalent decision rule of rejecting H_0 when $\omega > 1$, where ω is known as the *separation factor* given by the expression

$$\omega = \frac{\hat{r} - (r_1 + r_2)}{k_\alpha \sigma_r}, \tag{3.8}$$

where k_α is the $100(1-\alpha)$th percentile of the standard normal distribution. For instance, for the significance level $\alpha = 1/500$ we have that $k_\alpha = 2.878$. An argument for this expression is given in Appendix F.

3.1.2 The More Exact Test

Now, we will show how to find the exact distribution of the squared distance between the two closest points $\mathbf{p}_c = [N_c^p, E_c^p, V_c^p]^t$ and $\mathbf{q}_c = [N_c^q, E_c^q, V_c^q]^t$, that is

$$r^2 = (N_c^p - N_c^q)^2 + (E_c^p - E_c^q)^2 + (V_c^p - V_c^q)^2,$$

and we will use this distribution to create a hypothesis test that is more exact, in the sense that we do not need an additional normal approximation in order to find the distribution of r. However, we still have to do the normal approximations in Chapter 2, which involves the assumption of the *NEV* coordinates $[\mathbf{p}^t, \mathbf{q}^t]$ in expression (3.2) being normal distributed. The statistical distribution of the two closest points $[\boldsymbol{\mu}_{p_c}^t, \boldsymbol{\mu}_{q_c}^t]$ is still assumed to be given by expression (3.5).

Let us start by defining

$$\mathbf{d} = \mathbf{p}_c - \mathbf{q}_c. \tag{3.9}$$

By the error propagation theory in Appendix A, the statistical distribution of $\mathbf{d}$ is a trivariate normal distribution with mean vector $\boldsymbol{\mu}_d = \boldsymbol{\mu}_{p_c} - \boldsymbol{\mu}_{q_c}$, that is

$$\mathbf{d} \sim N_3\left(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d\right).$$

The covariance matrix is given by

$$\boldsymbol{\Sigma}_d = \mathbf{J}_d \boldsymbol{\Sigma}_{p_c,q_c} \mathbf{J}_d^t. \tag{3.10}$$

This transition is actually exact because $\mathbf{d}$ is defined as a linear function of $\mathbf{p}$ and $\mathbf{q}$. The Jacobian matrix is namely

$$\mathbf{J}_d = \begin{bmatrix} \mathbf{I}_3 & -\mathbf{I}_3 \end{bmatrix},$$

where $\mathbf{I}_3$ is the 3×3 identity matrix. The resulting covariance matrix for $\mathbf{d}$ turns out to be

$$\boldsymbol{\Sigma}_d = \boldsymbol{\Sigma}_{p_c} + \boldsymbol{\Sigma}_{q_c} - \boldsymbol{\Gamma}_{p_c,q_c} - \boldsymbol{\Gamma}_{q_c,p_c},$$

where we have defined

$$\boldsymbol{\Sigma}_{p_c,q_c} = \begin{bmatrix} \boldsymbol{\Sigma}_{p_c} & \boldsymbol{\Gamma}_{p_c,q_c} \\ \boldsymbol{\Gamma}_{q_c,p_c} & \boldsymbol{\Sigma}_{q_c} \end{bmatrix}.$$

Now, Sheil and O'Muircheartaigh (1977) show how to find the statistical distribution of

$$r^2 = \mathbf{d}^t \mathbf{d}, \tag{3.11}$$

exploiting the work of Ruben (1962) and Kotz et al. (1967). They use a linear transformation given by

$$\mathbf{b} = (\mathbf{L}\mathbf{V})^{-1} \boldsymbol{\mu}_d.$$

The 3×3 matrix $\mathbf{L}$ is a lower triangular matrix defined by a Cholesky decomposition $\boldsymbol{\Sigma}_d = \mathbf{L}\mathbf{L}^t$. An eigendecomposition of $\boldsymbol{\Sigma}_d$ is given by

$$\boldsymbol{\Sigma}_d = \mathbf{V}\mathbf{D}\mathbf{V}^{-1},$$

where the columns of the 3×3 matrix $\mathbf{V}$ are the eigenvectors of $\mathbf{\Sigma}_d$, and the diagonal matrix $\mathbf{D}$ consists of the corresponding eigenvalues $\alpha_1 \geq \alpha_2 \geq \alpha_3$.

According to Sheil and O'Muircheartaigh (1977), the cdf of $\mathbf{d}^t\mathbf{d}$ is given by

$$P\left(\mathbf{d}^t\mathbf{d} \leq t\right) = \sum_{k=0}^{\infty} c_k F\left(3+2k, t/\beta\right), \tag{3.12}$$

where $F(y,\nu)$ is the cumulative distribution function of a central χ^2 distribution with ν degrees of freedom for some value y. The constant β can be chosen freely; however, it must be contained in $(0, 2\alpha_{\min})$ in order to have a converging series in expression (3.12). Sheil and O'Muircheartaigh (1977) propose $\beta = 0.90625\alpha_{\min}$ for efficiency reasons. The coefficients c_k are given by

$$c_k = \begin{cases} Ae^{-\lambda/2} & , \quad k = 0, \\ \frac{1}{k}\sum_{r=0}^{k-1} g_{k-r} c_r & , \quad k \geq 1, \end{cases}$$

where

$$A = \sum_{j=1}^{3} \sqrt{\frac{\beta}{\alpha_j}}, \qquad \lambda = \sum_{j=1}^{3} b_j^2,$$

$$g_m = \frac{m}{2}\sum_{j=1}^{3} b_j^2 \gamma_j^{m-1} + \frac{1}{2}\sum_{j=1}^{3}\left(1 - m b_j^2\right)\gamma_j^m \qquad (m \geq 1),$$

with

$$\gamma_j = 1 - \frac{\beta}{\alpha_j} \qquad (j = 1, 2, \ldots, n).$$

Since there are an infinite number of terms in the sum in expression (3.12), we must truncate the series after a chosen number of terms N, in order to be able to compute the probability. In this case, Sheil and O'Muircheartaigh (1977) gives an upper limit for the error,

$$\sum_{k=N+1}^{\infty} c_k F\left(3+2k, t/\beta\right) < \left(1 - \sum_{k=0}^{N}\right) F\left(3+2N, t/\beta\right).$$

If we denote our measured realization of $\mathbf{d}$ by $\hat{\mathbf{d}}$, then correspondingly to the normal approximated test, the p-value is defined as the probability that a squared measured distance $\mathbf{d}^t\mathbf{d}$ actually is as extreme as (or more extreme than) $\hat{\mathbf{d}}^t\hat{\mathbf{d}}$, given

the null hypothesis H_0. Then the p-value can be obtained by

$$\begin{aligned}
p &= P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid H_0\right) \\
&= P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d \leq (r_1+r_2)^2\right) \\
&= P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid \boldsymbol{\mu}_d = \underset{\substack{\boldsymbol{\mu}_d \\ \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d \leq (r_1+r_2)^2}}{\operatorname{argmax}} \left\{P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}}\right)\right\}\right) \\
&= P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid \boldsymbol{\mu}_r = \underset{\substack{\boldsymbol{\mu}_d \\ \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d \leq (r_1+r_2)^2}}{\operatorname{argmax}} \left\{1 - P\left(\mathbf{d}^t\mathbf{d} \leq \hat{\mathbf{d}}^t\hat{\mathbf{d}}\right)\right\}\right) \\
&= 1 - P\left(\mathbf{d}^t\mathbf{d} \leq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid \boldsymbol{\mu}_r = \underset{\substack{\boldsymbol{\mu}_d \\ \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d \leq (r_1+r_2)^2}}{\operatorname{argmin}} \left\{P\left(\mathbf{d}^t\mathbf{d} \leq \hat{\mathbf{d}}^t\hat{\mathbf{d}}\right)\right\}\right)
\end{aligned}$$

The maximization in the third equation is made in order to get the most conservative result. The final minimization can be performed by using standard optimization software, which for instance is included in MATLAB. In order to calculate the final p-value, we use expression (3.12).

As a short example, we perform both the exact hypothesis test just described in this section *and* the normal approximated test in Section 3.1.1 on the well pair described in Figure 3.3. These artificial wells are seperated by a distance of 7.5 meters.

The probability distribution functions (pdf's) $f(r)$ under the null hypothesis

$$H_0: \qquad \mu_r \leq r_1 + r_2 = 1 \text{ m},$$

for both tests are plotted in Figure 3.4. The distributions are seen to differ significantly. If $\hat{r}$ denotes the measured shortest distance, then the p-value found as the area to the right of $r = \hat{r}$ below the graphs. Here, the shortest distance is about $\hat{r} = 7.5$ m. Then, the p-value is seen to be smaller when we use the normal approximated distribution in the first tests, than for the exact distribution in the second tests, which would mean that we will less often reject H_0 when performing the exact test. However, we note that the p-values tends to be equal when the distance $\hat{r}$ increases, which fits well with the statement of Gjerde (2008) that the normal approximated test is valid when the collision probability is small.

We will explore both tests more thoroughly in Chapter 4.

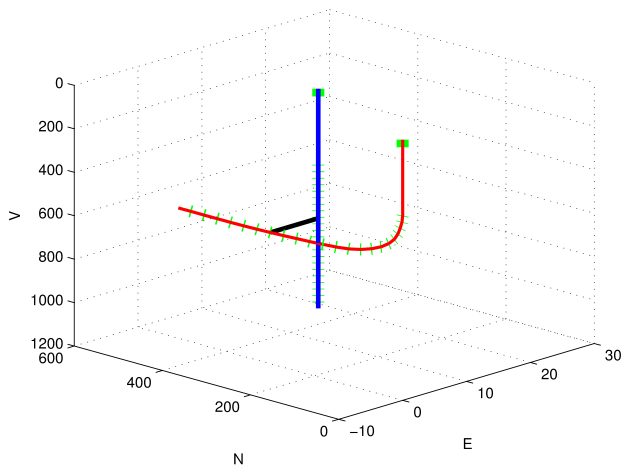

Figure 3.3: Two artificial petroleum wells. The blue reference well is going strictly downwards from the starting point $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [300, 7.5, 0]^t$. The red offset well starts out at $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [0, 0, 0]^t$ going downwards, before the inclination starts to increase. At the end, the well is horizontal, going northwards. The shortest distance between the wells is 7.5 m, marked with a black line.

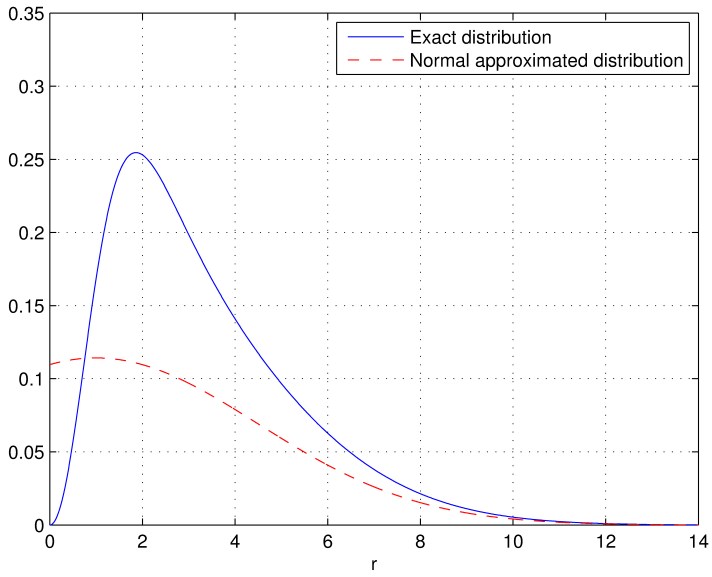

Figure 3.4: The pdf's of the shortest distance r under H_0 for the two hypothesis tests described in this chapter, as performed on the well pair in Figure 3.3.

3.2 The Probability of Collision

We now want to find the *probability* that there will be a collision between two neighboring wells. This is the value defined as

$$l = P\left(\text{Well collision}\right) = P\left(r(\mathbf{x}) \leq r_1 + r_2\right),$$

where $r(\cdot)$ is a function that returns the shortest distance between the two wells based on the stochastic value $\mathbf{x}$ that contains the considered information regarding the positions of the wells. In the two-points methods that we describe in Section 3.2.1, we only consider the *NEV* coordinates of the two closest points, implying $\mathbf{x} = [\mathbf{p}_c^t, \mathbf{q}_c^t]$. In the all-points method that we describe in Section 3.2.2, we consider all the survey points in the wells, meaning $\mathbf{x} = \boldsymbol{\alpha}$, where $\boldsymbol{\alpha}$ is a vector containing the *DIA* values of all the survey points in both wells, that is

$$\boldsymbol{\alpha} = \left[D_0^p, I_0^p, A_0^p, \ldots, D_{D_p}^p, I_{K_p}^p, A_{K_p}^p, D_0^q, I_0^q, A_0^q, \ldots, D_{K_q}^q, I_{K_q}^q, A_{K_q}^q\right], \tag{3.13}$$

where the superscript p indicates the reference well, while q indicates the offset well. As before, $K_p + 1$ and $K_q + 1$ are the total numbers of survey points in each well.

When we consider the two closest points, the probability l_{two} can be calculated in an analytical manner, based on the procedure that we used during the derivation of the more exact hypothesis test in Section 3.1.2. When we consider all the points method, an analytical calculation of the collision probability is outside the scope of this thesis.

In both cases, we will estimate the probability by *crude Monte Carlo (CMC) simulation*. We then need to realize that

$$\begin{aligned} P\left(r(\mathbf{x}) \leq r_1 + r_2\right) &= E_x I\left(r(\mathbf{x}) \leq r_1 + r_2\right) \\ &= \int_{\mathcal{X}} I(r(\mathbf{x}) \leq r_1 + r_2) f\left(\mathbf{x}|\mathbf{u}\right) d\mathbf{x}, \end{aligned}$$

where $f(\mathbf{x}|\mathbf{u})$ is the pdf of $\mathbf{x}$, possibly depending on a parameter $\mathbf{u}$. Also, E_f indicates the expectation over $\mathbf{x}$, $I(\cdot)$ is an indicator function, and $\mathcal{X}$ is some space in which the stochastic vector $\mathbf{x}$ takes values.

A CMC estimate of l is then given by the plug-in estimate

$$\hat{l} = \hat{P}\left(\text{Well collision}\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(r_b \leq r_1 + r_2\right) \tag{3.14}$$

Here we have defined $r_b = r(\mathbf{x}_b)$, where $\mathbf{x}_b$, for $b = 1, \ldots, B$, are B simulated realizations from the distribution of $\mathbf{x}$. [1] The *Strong Law of Large Numbers* assures

[1] In this thesis we only consider the multivariate normal distribution, for which there are known methods for simulation. In order to simulate $\mathbf{x}$ from the distribution $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, a common method consists of simulating the vector $\mathbf{z}$ from the standard normal distribution $N_d(\mathbf{0}, \mathbf{I}_d)$, for instance by using the *Box-Muller algorithm*. Then $\mathbf{x} = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}$, where $\mathbf{A}$ is defined by the Cholesky decomposition $\mathbf{A}\mathbf{A}^t = \boldsymbol{\Sigma}$, is a realization from the distribution of interest.

that the estimate $\hat{l}$ converges to the expected value l when the sample size B is large enough. From the *Central Limit Theorem*, the estimator given in (3.14) tends to be normal distributed when B gets large.

In order to have a measure of the uncertainty in this estimate, we estimate the standard deviation $\sigma_{\hat{l}}$ of $\hat{l}$ (Ripley, 1987) by the *standard error* of $\hat{l}$ given by

$$\hat{\sigma}_{\hat{l}} = \frac{1}{\sqrt{B}} \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(I\left(r_b \leq r_1 + r_2\right) - \hat{l} \right)^2}. \tag{3.15}$$

An alternative error measure is the *relative error* (also known as the *coefficient of variation*) of the estimate. This is defined as

$$RE = \frac{\hat{\sigma}_{\hat{l}}}{\hat{l}} \quad , \tag{3.16}$$

which is simply the standard error scaled according to the size of the estimate $\hat{l}$ itself. The interval

$$\left[\hat{l} - \hat{\sigma}_{\hat{l}}, \hat{l} + \hat{\sigma}_{\hat{l}}\right]$$

is known as the 68 % *confidence interval* of l, which is also a common measure of the error in a Monte Carlo method (Goodman). From expression (3.15) we can conclude that the sample size B should be as large as possible, in order to obtain a narrow confidence interval and a small relative error.

In many cases, the probability of well collision may be very small, say of order $\sim 10^{-6}$. In such cases, we consider a collision to be a *rare event*, while the probability of collision is a *rare-event probability*. In order to obtain an accurate estimate of the probability (which means that it has a narrow confidence interval) by using a Monte Carlo simulation method, we may need a very large number of samples B. Since the CMC simulations are very time-consuming and computer-intensive, especially when we consider all the points in the wells, we might want to use methods that can improve the accuracy of the estimates. In Section 3.2.3 we introduce a cross-entropy method in order to increase the accuracy of the rare-event probability estimate in both the two-points situation and in the all-points situation.

Since we calculate the probability value in an analytical manner only in the two-points situation, we introduce a second rare-event method in the all-points situation, in order to get an even broader foundation for comparison of the methods. This is the enhanced Monte Carlo (EMC) method, which we introduce in Section 3.2.4.

The word *crude* in the term *crude Monte Carlo simulation method* implies that the method does not take advantage of any of the two rare-event methods.

3.2.1 Considering the Two Closest Points

We will now find the collision probability in the same manner as we performed the hypothesis tests in Section 3.1, meaning that we only consider the two closest points in the two wells. That is, we want to find the probability l_{two} that the two

closest points $\mathbf{p}_c = [N_p, E_p, V_p]^t$ and $\mathbf{q}_c = [N_q, E_q, V_q]^t$ are in fact coinciding, in the way that their distance is shorter than $r_1 + r_2$.

In the introduction of the hypothesis tests in Section 3.1, we assumed that the joint statistical distribution of the vector $[\mathbf{p}_c^t, \mathbf{q}_c^t]$ is given by (3.5), that is

$$\left[\mathbf{p}_c^t, \mathbf{q}_c^t\right] \sim N_6\left([\boldsymbol{\mu}_{p_c}^t, \boldsymbol{\mu}_{q_c}^t], \boldsymbol{\Sigma}_{p_c,q_c}\right), \tag{3.17}$$

with expectation vector and covariance matrix as described in that section.

In Section 3.1.2 we described how to calculate the cdf $P\left(\mathbf{d}^t\mathbf{d} < t\right)$, as seen in expression (3.12), where we defined $\mathbf{d} = \mathbf{p}_c - \mathbf{q}_c$. Now we realize that the probability of collision l_{two} actually is equal to that cdf value, when we set $t = (r_1 + r_2)^2$ and approximate the expected values $\boldsymbol{\mu}_{p_c}$ and $\boldsymbol{\mu}_{q_c}$ by the planned or measured values.

That is, the probability value can be obtained analytically by

$$\hat{l}_{\text{two}}^{\text{analytical}} = P\left(\mathbf{d}^t\mathbf{d} < (r_1 + r_2)^2\right),$$

where we use the cdf in expression (3.12).

In order to find a CMC estimate of l_{two}, we use that the shortest distance function is given by

$$r\left(\left[\mathbf{p}_c^t, \mathbf{q}_c^t\right]\right) = \sqrt{(\mathbf{p}_c - \mathbf{q}_c)^t(\mathbf{p}_c - \mathbf{q}_c)}.$$

The CMC simulations $r_1, \ldots, r_B$, are then found by simulating B samples $[\mathbf{p}_c^t, \mathbf{q}_c^t]_b$, $b = 1, \ldots, B$, from the distribution (3.17). Then,

$$r_b = r\left([\mathbf{p}_c^t, \mathbf{q}_c^t]_b\right) = \sqrt{(\mathbf{p}_{c,b} - \mathbf{q}_{c,b})^t(\mathbf{p}_{c,b} - \mathbf{q}_{c,b})}$$

for $b = 1, \ldots, B$. The CMC probability estimate $\hat{l}_{\text{two}}^{\text{CMC}}$ is finally calculated by expression (3.14), that is

$$\hat{l}_{\text{two}}^{\text{CMC}} = \hat{P}_{\text{two}}^{\text{CMC}}\left(\text{Well collision}\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(r_b \leq r_1 + r_2\right).$$

The uncertainty of this estimate can be described by the standard deviation $\hat{\sigma}_{\hat{l}}$ in (3.15) and the relative error RE in (3.16).

3.2.2 Considering All Points

Tsao et al. (1999), Thorogood et al. (1991) and Brooks (2008) all intend to find the probability of collision for *segments* of two wells rather than just *points*, but they all have in common that they do not take into account the correlation between different areas of the segments or between the two wells. In this section we aim to find a method that do take this into account.

We will do this by simulating B realizations of the DIA values

$$\begin{aligned} \boldsymbol{\alpha} &= [\boldsymbol{\alpha}^p, \boldsymbol{\alpha}^q] \\ &= \left[D_0^p, I_0^p, A_0^p, \ldots, D_{D_p}^p, I_{K_p}^p, A_{K_p}^p, D_0^q, I_0^q, A_0^q, \ldots, D_{K_q}^q, I_{K_q}^q, A_{K_q}^q\right] \end{aligned} \tag{3.18}$$

of all the survey points of the two wells from the distribution

$$\left[\boldsymbol{\alpha}^p, \boldsymbol{\alpha}^q\right]^t \sim N_d\left(\left[\boldsymbol{\mu}_{\alpha^p}, \boldsymbol{\mu}_{\alpha^q}\right]^t,\ \boldsymbol{\Sigma}_{DIA}^{\text{total}}\right), \tag{3.19}$$

as given in (2.29) in Section 2.5.1. The dimension is $d = 3(K_p + K_q + 2)$. We estimate the expectation vector by the given (planned or measured) values of $\boldsymbol{\alpha}$ in (3.18), while the covariance matrix $\boldsymbol{\Sigma}_{DIA}^{\text{total}}$ is obtained from expression (2.30).

The shortest distance function $r(\boldsymbol{\alpha})$ will now calculate the *NEV* coordinates from the *DIA* values $\boldsymbol{\alpha}$ by the methods described in Chapter 2. Then a 3D closest approach method, as described in Appendix E, is used to approximate the shortest distance between the wells.

We denote the simulated *DIA* values by $\boldsymbol{\alpha}_b$, for $b = 1, \ldots, B$. Finally, we estimate the probability of collision with the plug-in estimate in expression (3.14), giving the CMC estimate

$$\hat{l}_{\text{all}}^{\text{CMC}} = \hat{P}_{\text{all}}^{\text{CMC}}\left(\text{Well collision}\right) = \frac{1}{B}\sum_{b=1}^{B} I\left(r_b \leq r_1 + r_2\right),$$

where $r_b = r\left(\boldsymbol{\alpha}_b\right)$, for $b = 1, \ldots, B$.

In the same way as in the CMC simulation method for two points, we should now perform an error analysis including the standard deviation $\hat{\sigma}_{\hat{l}}$ in (3.15) and the relative error *RE* in (3.16).

3.2.3 The Cross-Entropy Method

One way to obtain an accurate estimate of a rare-event probability is to use a *cross-entropy (CE) method* as described by de Boer et al. (2002) and Rubinstein and Kroese (2004).

As we described in the introduction of Section 3.2, the probability of collision between two wells is in general defined as

$$\begin{aligned} l &= P(\text{well collision}) = P\left(r(\mathbf{x}) \leq \gamma\right) \\ &= E_f I\left(r(\mathbf{x}) \leq \gamma\right) \\ &= \int_{\mathcal{X}} I(r(\mathbf{x}) \leq \gamma) f\left(\mathbf{x}|\mathbf{u}\right) d\mathbf{x}, \end{aligned}$$

where we have defined the limit value

$$\gamma = r_1 + r_2.$$

The vector $\mathbf{x}$ contains the considered information regarding the positions of the wells, while $r(\mathbf{x})$ is a function that returns the shortest distance between the two wells based on $\mathbf{x}$. We now let $f(\mathbf{x}|\mathbf{u})$ be the pdf of $\mathbf{x}$, which depends on a parameter $\mathbf{u}$.

In the two-points method we consider only the *NEV* coordinates of the two closest points, implying

$$\mathbf{x} = \left[\mathbf{p}_c^t, \mathbf{q}_c^t\right] = \left[N_c^p, E_c^p, V_c^p, N_c^q, E_c^q, V_c^p\right].$$

In this case, $f(\mathbf{x}|\mathbf{u})$ is the pdf of the trivariate normal distribution described by expression (3.17). In the all-points method, we consider the *DIA* values of all the survey points in the wells, implying

$$\mathbf{x} = \boldsymbol{\alpha} = \left[D_0^p, I_0^p, A_0^p, \ldots, D_{D_p}^p, I_{K_p}^p, A_{K_p}^p, D_0^q, I_0^q, A_0^q, \ldots, D_{K_q}^q, I_{K_q}^q, A_{K_q}^q\right].$$

Now, $f(\mathbf{x}|\mathbf{u})$ is the pdf of the multivariate normal distribution described by expression (3.19). In both cases, the parameter $\mathbf{u}$ is the *expectation vector* of $\mathbf{x}$, having dimension d. We assume the covariance matrix $\boldsymbol{\Sigma}$ to be known, giving the pdf

$$f\left(\mathbf{x}|\mathbf{u}, \boldsymbol{\Sigma}\right) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{u})^t \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{u})\right). \tag{3.20}$$

In order to calculate the pdf value, we need the inverse $\boldsymbol{\Sigma}^{-1}$ of the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$, since this is a part of the density (3.20). If the covariance matrix is singular or close to singular because of high correlation, we follow the principle of ignoring the most uninformative variables until we are able to find a useable pdf value. This method is thorougly described in Appendix D.

We previously showed that the CMC simulation method gives the estimate

$$\hat{l} = \frac{1}{B}\sum_{b=1}^{B} I\left(r(\mathbf{x}_b) \leq \gamma\right)$$

when we use B simulated samples $\mathbf{x}_b$, for $b = 1 \ldots, B$, from $f(\mathbf{x}|\mathbf{u})$. In order to obtain a satisfying width of the confidence interval for l, there is need for a very large sample size B, especially when the event $\{r(\mathbf{x}) \leq \gamma\}$ is known to be a rare event. The CE method is a way to overcome this computational challenge.

The procedure is based on *importance sampling*. Instead of sampling from $f(\mathbf{x}|\mathbf{u})$, one samples from another density $g(\mathbf{x})$, which is such that

$$r(\mathbf{x}) \leq \gamma) f(\mathbf{x}|\mathbf{u}) = 0.$$

We then use the fact that l can be represented as

$$l = \int I(r(\mathbf{x}) \leq \gamma) \frac{f(\mathbf{x}|\mathbf{u})}{g(\mathbf{x})} g(\mathbf{x}) d\mathbf{x} = E_g I(r(\mathbf{x}) \leq \gamma) \frac{f(\mathbf{x}|\mathbf{u})}{g(\mathbf{x})}.$$

By simulating B realizations from the density $g(\mathbf{x})$, we can estimate l by the regular plug-in estimate, which is called the *importance sampling estimator* or the *likelihood ratio estimator*, given by

$$\hat{l} = \frac{1}{B}\sum_{b=1}^{B} I(r(\mathbf{x}_b) \leq \gamma) W(\mathbf{x}_b), \tag{3.21}$$

where we have introduced the *likelihood ratio*

$$W(\mathbf{x}) = \frac{f(\mathbf{x}|\mathbf{u})}{g(\mathbf{x})}.$$

Now, the CE method provides a way to find the density $g(\mathbf{x})$ that somehow optimizes the accuracy of the estimator $\hat{l}$ in (3.21). The best choice would be to use

$$g^*(\mathbf{x}) = \frac{I(r(\mathbf{x}) \leq \gamma) f(\mathbf{x}|\mathbf{u})}{l}, \tag{3.22}$$

which would mean that $\hat{l}$ in (3.21) equals l with zero variance. Of course, we can not use this estimator because it depends on l itself. However, we use this observation to find another density $g(\mathbf{x})$ that is useable *and* gives an estimator $\hat{l}$ with a satisfying accuracy. According to de Boer et al. (2002), it is convenient to choose a density g in the same density family as $f(\mathbf{x}|\mathbf{u})$. We therefore set $g(\mathbf{x}) = f(\mathbf{x}|\mathbf{v})$, for some unknown parameter $\mathbf{v}$. We will now write the likelihood ratio as

$$W(\mathbf{x}|\mathbf{u}, \mathbf{v}) = \frac{f(\mathbf{x}|\mathbf{u})}{f(\mathbf{x}|\mathbf{v})}.$$

The *Kullback-Leibler distance* or the *cross-entropy* between two densities $h_1(\mathbf{x})$ and $h_2(\mathbf{x})$ is defined as

$$\mathcal{D}(h_1(\mathbf{x}), h_2(\mathbf{x})) = E_{h_2} \ln \frac{h_1(\mathbf{x})}{h_2(\mathbf{x})} = \int h_1(\mathbf{x}) \ln h_1(\mathbf{x}) d\mathbf{x} - \int h_2(\mathbf{x}) \ln h_2(\mathbf{x}) d\mathbf{x}. \tag{3.23}$$

The cross-entropy method uses this measure of distance to find the density $f(\mathbf{x}|\mathbf{v})$ that is closest to g^* defined in (3.22), by finding the density that minimizes the cross-entropy $\mathcal{D}(f(\mathbf{x}|\mathbf{v}), g^*(\mathbf{x}))$. By minimizing (3.23) with respect to $\mathbf{v}$, they show that for *any* parameter $\mathbf{w}$, the optimal choice of $g(\mathbf{x})$ is the density $f(\mathbf{x}|\mathbf{v}^*)$, where

$$\mathbf{v}^* = \arg\max_{\mathbf{v}} E_w I(r(\mathbf{x}) \leq \gamma) W(\mathbf{x}|\mathbf{u}, \mathbf{w}) \ln f(\mathbf{x}|\mathbf{v}),$$

which can be estimated by simulating $\mathbf{x}_1, \ldots, \mathbf{x}_{B_1}$ from $f(\mathbf{x}|\mathbf{w})$, and then solving

$$\frac{1}{B_1} \sum_{b=1}^{B_1} I(r(\mathbf{x}_b) \leq \gamma) W(\mathbf{x}_b|\mathbf{u}, \mathbf{w}) \nabla_{\mathbf{v}} f(\mathbf{x}_b|\mathbf{v}) = \mathbf{0} \tag{3.24}$$

with respect to $\mathbf{v}$.

Now, equation (3.24) can be solved analytically with respect to $\mathbf{v}$. The solution for an element v_j in $\mathbf{v}$ is (de Boer et al., 2002)

$$v_j = \frac{\sum_{b=1}^{B_1} I(r(\mathbf{x}_b) \leq \gamma) W(\mathbf{x}_b|\mathbf{u}, \mathbf{w}) x_{bj}}{\sum_{b=1}^{B_1} I(r(\mathbf{x}_b) \leq \gamma) W(\mathbf{x}_b|\mathbf{u}, \mathbf{w})}. \tag{3.25}$$

Finally, the procedure of the CE method is as follows.

1. Choose values for the parameter $\mathbf{w}$ and the sample sizes B and B_1.
2. Simulate $\mathbf{x}_1, \ldots, \mathbf{x}_{B_1}$ from $f(\mathbf{x}|\mathbf{w})$.
3. Calculate $\mathbf{v}$ by (3.25).

4. Simulate $\mathbf{x}_1, \ldots, \mathbf{x}_B$ from $f(\mathbf{x}|\mathbf{v})$.

5. Calculate the importance sampling estimate $\hat{l}$ in (3.21) with $g(\mathbf{x}) = f(\mathbf{x}|\mathbf{v})$.

Unfortunately, the result is not useful when the probability is "too small", which may cause a problem when we are dealing with rare-event probabilities. Rubinstein and Kroese (2004) solve this problem by using a *multi-level algorithm*. In the algorithm they use a sequence of *tuning parameters* $\{\hat{\mathbf{v}}_t, t \geq 0\}$ and a sequence of *levels* $\{\hat{\gamma}_t, t \geq 1\}$. In each iteration, an element is estimated in both sequences. They also introduce a *rarity parameter* ρ that should be chosen to be not "too small". Their suggestion is $\rho = 0.1$, which we will use when we perform the method.

The algorithm is summarized as follows.

1. Choose values for the value ρ, the maximum number of iterations $t_{\max}$ and the sample sizes B and B_1.

2. Define $\hat{\mathbf{v}}_0 = \mathbf{u}$ and set the iteration counter $t = 1$.

3. Simulate a B_1-sized sample of realizations $\mathbf{x}_1, \ldots, \mathbf{x}_{B_1}$ from $f(\mathbf{x}|\hat{\mathbf{v}}_{t-1})$. Estimate the sample ρ-quantile $\hat{\gamma}_t$ by the $\lfloor \rho B_1 \rfloor$'th *order statistic* of the sequence $r(\mathbf{x}_1), \ldots, r(\mathbf{x}_{B_1})$.

4. Use the sample $\mathbf{x}_1, \ldots, \mathbf{x}_{B_1}$ found in the previous step to calculate $\hat{\mathbf{v}}_t$, with elements given by (3.25), using $\gamma = \hat{\gamma}_t$ and $\mathbf{w} = \hat{\mathbf{v}}_{t-1}$.

5. If both $\hat{\gamma}_t > \gamma$ *and* $t < t_{\max}$, set $t = t + 1$ and go to step 3. Otherwise, go to step 6.

6. Estimate the collision probability l by the importance sampling estimator in (3.21),
$$\hat{l} = \frac{1}{B} \sum_{b=1}^{B} I(r(\mathbf{x}_b) \leq \gamma) W(\mathbf{x}_b|\mathbf{u}, \hat{\mathbf{v}}_t),$$
using a B-sized sample $\mathbf{x}_1, \ldots, \mathbf{x}_B$ simulated from $f(\mathbf{x}|\hat{\mathbf{v}}_t)$.

In the last step of the algorithm one finds the CE estimate $\hat{l}^{\mathrm{CE}}$, which should preferably give a narrower confidence interval than the CMC estimate while needing less computational effort.

The standard error of the estimate is calculated by

$$\hat{\sigma}_{\hat{l}} = \frac{1}{\sqrt{B}} \sqrt{\frac{1}{B-1} \sum_{b=1}^{B} \left(I\left(r_b \leq r_1 + r_2\right) W(\mathbf{x}_b, \mathbf{u}, \hat{\mathbf{v}}_t) - \hat{l} \right)^2},$$

while the relative error is still defined as

$$RE = \frac{\hat{\sigma}_{\hat{l}}}{\hat{l}} \quad .$$

As an example, we perform the all-points CE method on the artificial well pair described in Figure 3.3. In each iteration of the algorithm just described, a new well pair is estimated, and their *DIA* values are contained in the vector $\hat{\mathbf{v}}_t$. Only two iterations were needed to reach the stopping criteria $\hat{\gamma}_t \leq \gamma = r_1 + r_2 = 1$ m, resulting in a total of three well pairs, including the original pair. In Figure 3.5 we have depicted these well pairs. The blue reference well seem to have about the exact same position in the three pairs. The original well pair is described by $\hat{\mathbf{v}}_0 = \mathbf{u}$, estimated by the planned or measured *DIA* values $\boldsymbol{\alpha}$, while $\hat{\mathbf{v}}_1$ and $\hat{\mathbf{v}}_2$ contains the *DIA* values of the well pairs that have a shorter distance. In step 3 of the algorithm above we obtain the level values $\hat{\gamma}_1 = 3.45$ and $\hat{\gamma}_2 = 0.50$ before we finally estimate the collision probability $\hat{l}_{\text{all}}^{\text{CE}} = 1.668 \cdot 10^{-2}$ in step 6, having a relative error of $RE = 9.14 \cdot 10^{-3}$. In Table 3.1 we give the numerical results obtained when we apply both the CE method and the CMC method on the same well pair. Even though we use a smaller number of samples in total, which decreases the computational run time, the standard error and relative error is seen to be smaller for the CE method than for the CMC method.

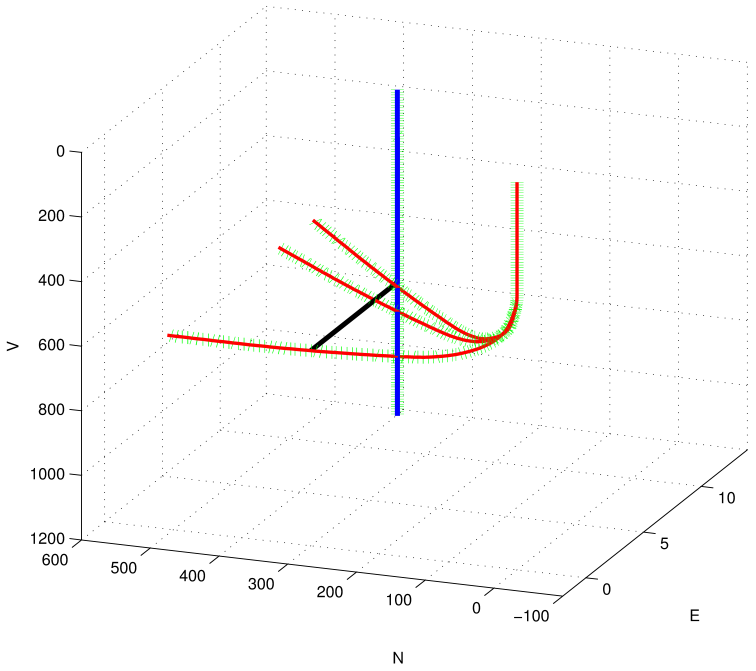

Figure 3.5: The three well pairs found when performing the CE method on the wells in Figure 3.3.

In Chapter 4 we will perform the CE method on various well pair test cases, in order to learn more about this method.

Table 3.1: Numerical results obtained by performing the all-points CMC simulation method and the all-points CE simulation method on the well pair in Figure 3.3. We have used the notation $XeY = X \cdot 10^Y$.

Method	B	B_1	Run time (sec)	$\hat{l}_{\text{all}}$	$\hat{\sigma}_{\hat{l}}$	RE
CMC	1e5	-	1927	1.621e-2	3.99e-4	2.46e-2
CE	5e4	5e3	1122	1.668e-2	1.54e-4	9.14e-3

3.2.4 The Enhanced Monte Carlo Method

Naess et al. (2009) propose a technique called *Enhanced Monte Carlo (EMC)*, which is originally used in estimation of the reliability of various structural problems. This reliability is often expressed as a *failure probability*, which usually is very small. Therefore, one could think that the method would be useful in other rare-event problems, including the problem of estimating the probability of a well collision.

The EMC method is all about increasing the efficiency (that is, reducing the computational running time) in the task of estimating some kind of failure probability p that depends on a *safety margin* M, in the way that $p = P(M \leq M^*)$, for some failure limit M^*. In our case, the safety margin M is the measured shortest distance r between two petroleum wells, while the failure probability p corresponds to the well collision probability l. The failure limit M^* equals the critical distance $\gamma = r_1 + r_2$, that is the sum of the radii of the two wells.

The method exploits the possibility that there are some values of the safety margin other than our actual measured value that would result in a larger failure probability. Calculating the corresponding failure probabilities using these other values can be carried out with improved efficiency compared to the actual problem, since there would be need for a smaller amount of simulations in order to obtain a confidence interval of a similar size. Then, if there exists a relationship between the probability of interest and the probability values that are easier to obtain, we could hope to estimate the wanted probability value based on those other values.

In our case the probability of collision increases when the shortest distance decreases. This means that by shifting one of the wells towards the other, we get a larger collision probability. In order to perform this shifting, we introduce a scaling parameter $\lambda \in [0, 1]$. Given a value λ, the corresponding shortest distance is defined to be

$$r(\lambda) = r - \mu_r (1 - \lambda).$$

Since we estimate the true shortest distance μ_r by a measurement r, we actually have that $r(\lambda) \approx r\lambda$.

Now it is assumed that the probability function $l_\lambda(\lambda)$ satisfies

$$l_\lambda(\lambda) = q \exp\{-a(\lambda - b)^c\}, \qquad \lambda_0 \leq \lambda \leq 1,$$

for some appropriate value λ_0. After taking the logarithm of both sides, we find

the parameters q^*, a^*, b^* and c^* that minimize the mean square error function

$$F(q,a,b,c) = \sum_{j=1}^{M} w_j \left(\log \hat{l}(\lambda_j) - \log q + a(\lambda_j - b)^c \right)^2, \tag{3.26}$$

where the suggested weight factor is given by

$$w_j = \left(\log C^-(\lambda_j) - \log C^+(\lambda_j) \right)^{-2}.$$

Here we have used the limits of the 95% confidence interval for l_λ, as estimated by

$$C^{\pm}(\lambda) = \hat{l}_\lambda(\lambda) \left(1 \pm 1.96\, C_v\left(\hat{l}_\lambda\right) \right) \tag{3.27}$$

with an estimate of the *coefficient of variation*, equaling the value that we earlier defined as the *relative error*, given by

$$C_v\left(\hat{l}_\lambda\right) = RE\left(\hat{l}_\lambda\right) = \frac{\hat{\sigma}_\lambda}{\hat{l}_\lambda(\lambda)} = \sqrt{\frac{1-\hat{l}_\lambda}{B\hat{l}_\lambda}}, \tag{3.28}$$

where B is the number of samples in the MC simulation method for estimating $\hat{l}_\lambda$. For any $\lambda \in [\lambda_0, 1]$, the probability is estimated by

$$\hat{l}_\lambda(\lambda) = q^* \exp\left\{ -a^* (\lambda - b^*)^{c^*} \right\}. \tag{3.29}$$

Then we can estimate the probability of interest l by

$$\hat{l} = \hat{l}_\lambda(1) = q^* \exp\left\{ -a^* (1 - b^*)^{c^*} \right\}. \tag{3.30}$$

As always, we would like to know about the uncertainty of the estimate $\hat{l} = \hat{l}_\lambda(1)$. One way to approximate a confidence interval for $\hat{l}$ is to use the confidence interval in equation (3.27). In order to obtain a value for the coefficient of variation $C_v(\hat{l})$ in (3.28), we use the same value B that we use for estimating the other $\hat{l}_\lambda$ values, even though we never actually estimate $\hat{l}$ using B samples in that way.

We can check if this approximation actually gives a satisfying estimate of the confidence interval, by first estimating the collision probability as exact as possible, which takes a lot of time. Then we do many EMC runs, and count the number of times that the "exact" probability value is inside the respective confidence intervals. The resulting *coverage ratio* gives an indication of the degree to which the confidence intervals actually are meaningful. When we estimate the 95 % confidence intervals of 500 EMC runs, we obtain that the coverage ratio is only about 90 %. This clearly indicates that the width of these confidence intervals are underestimated.

As an example, we once more use the well pair shown in Figure 3.3. When performing the EMC method on these wells, we shift the blue reference well towards the red reference well, as illustrated in Figure 3.6. In each new position we estimate the collision probability using a crude Monte Carlo method.

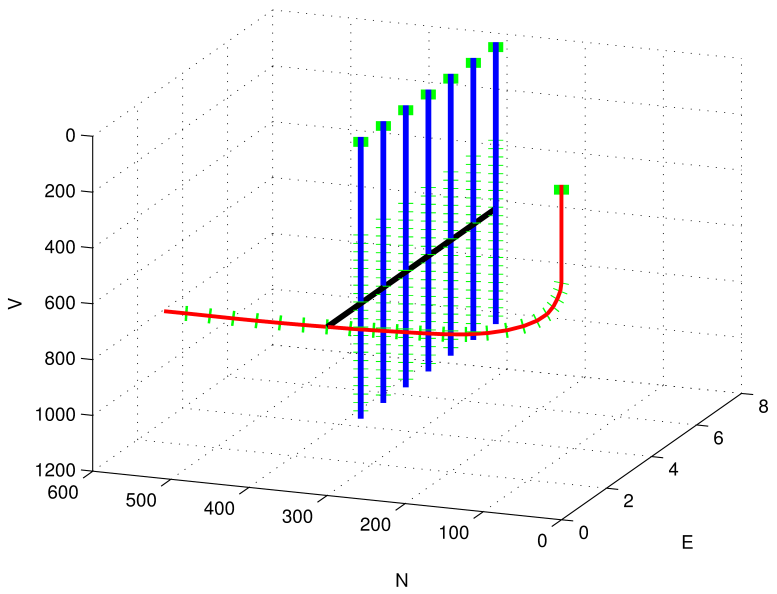

Figure 3.6: The new shifted well pairs used when performing the EMC method on the wells in Figure 3.3.

The probability values are plotted as circles in the two plots of Figure 3.7. These two plots show the same results, except that the ordinal axis is logarithmic in the right-hand plot. We now find the parameters that minimizes the mean square error function in expression (3.26), which gives us the fitted blue regression lines, together with the red (and probably underestimated) 95 % confidence intervals from expression (3.27). The final EMC estimate turns out to be $\hat{l} = \hat{l}_\lambda(1) = 1.690 \cdot 10^{-2}$, having a relative error $RE = 5.39 \cdot 10^{-2}$.

We have plotted the CE estimate $\hat{l} = 1.668 \cdot 10^{-2}$ from the previous section as a star in Figure 3.7. This value seem to be close to the fitted regression line, and it is placed well inside the confidence interval, implying that the EMC method gives a good estimate of $\hat{l}$ in this case.

When we later apply this method on various well pair cases in Chapter 4, we use a larger number of interpolating values in the area $\lambda \in \{0.2, 0.6\}$. These values are more realistic to obtain in a real rare-event situation, because the probability values $\hat{l}_\lambda(\lambda)$ are so small when we are close to $\lambda = 1$. In Chapter 4, we will also compare the EMC results with the numerical results obtained by the CMC method and the CE method.

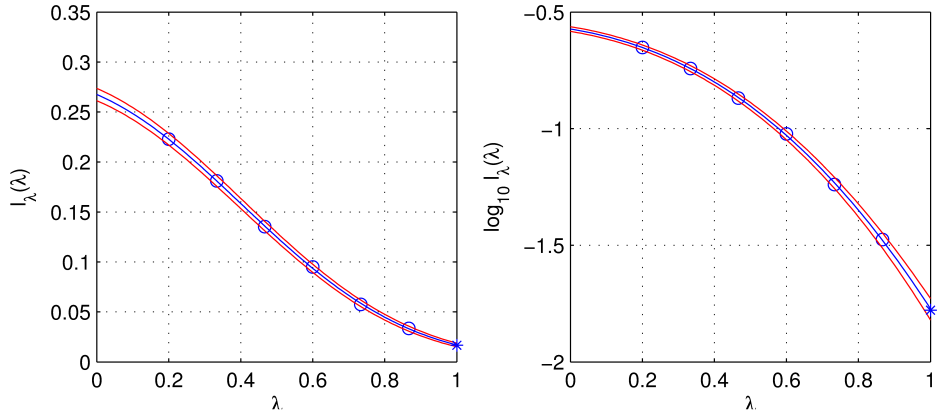

Figure 3.7: The EMC method performed on the well pair in Figure 3.3 results in this interpolation.

Chapter 4

Numerical Results

In this chapter we apply the collision risk methods from Chapter 3 on various well pair cases, in order to compare and learn more about the different methods. In addition, we make some observations regarding the relationship between the geometry of the wells and the risk of collision.

In Section 4.1 we introduce six main test cases, that each consists of a well pair having a distinguishable geometry. Some of the main cases are split into sub-cases, making a total number of 15 test cases.

In the rest of the chapter, there are two main tasks involving numerical results that will be carried out, using the well pair test cases. First, in Section 4.2, we apply the methods from Section 3 on the test cases in order to compare and better describe the various methods. Second, in Section 4.3, we use parts of the numerical results in an attempt to describe the influence of the crossing angle, the well direction in the NE plane and the measured depth on the collision risk.

The *crossing angle*, or *the angle of incidence*, is defined to be the angle between the wells in the area around the two closest points, when we consider the most natural two-dimensional plane in the NEV coordinate system for the specific well pair geometry, for instance the NV plane or the horizontal NE plane. In order to specify this angle uniquely, we must specify the considered plane in addition to the numerical value of the angle. For instance, for the well pair shown in Figure 3.3 in Section 3.1.2, the crossing angle is 90° in the NV plane.

4.1 Well Pair Test Cases

We now begin by describing the six different test cases that we use in the rest of this chapter. Some of the cases consists of sub-cases, resulting in a total number of 15 fundamentally different cases. In addition, we have some shifted versions of some of these cases, meaning it is the same well pair with the same geometry, with the only exception that the shortest distance is adjusted according to our needs.

Most of the cases are made-up artificial well pairs, while Case 6 is a real well pair from the North Sea. We describe and depict the well pairs carefully in Section

4.1.1 through 4.1.6. A whole lot of information will be presented, so we also make a comprehensible summary of the test cases in Section 4.1.8.

4.1.1 Test Case 1: Crossing Angle of 90°

The first test case consists of two sub-cases named 1-A and 1-B. Both pairs consist of an artificial well pair where the crossing angle between the wells is set to be 90° in either the NV plane or the EV plane. The shortest distance between the wells is 7.5 meters in both cases.

Case 1-A is depicted in Figure 4.1. The blue reference well is going strictly downwards, while the red offset well is going downwards at first, before going horizontally *northwards*, passing west of the reference well.

Case 1-B is shown in Figure 4.2, and is very similar to Case 1-A. The only difference is that the red offset well now ends up going horizontally *eastwards*, passing south of the reference well.

4.1.2 Test Case 2: Crossing Angle of 45°

The second case is very similar to the first case, except that while the red offset well ended up having an inclination of zero degrees in the first case, it now ends up having an inclination of only 45°. The crossing angle between the wells is thereby constructed to be equal to 45° in either the NV plane or the EV plane. This case also consists of two sub-cases. Case 2-A and Case 2-B is shown in Figure 4.3 and Figure 4.4 respectively. In the exact same way as for Case 1-A and Case 1-B, the offset well is heading *northwards* in Case 2-A, while it is going *eastwards* in Case 2-B. The shortest distance between the wells is 7.5 meters.

4.1.3 Test Case 3: Parallel Wells

The third case consists of six sub-cases named 3-A, 3-B, 3-C, 3-D, 3-E and 3-F. In all these cases the two wells are parallel, both first going straight down in the vertical direction, and then turning to the horizontal direction. Their distance is constructed to be 3 meters the whole time. In the first three cases, the wells are heading *northwards*, while in the last three cases they are heading *eastwards*.

The cases 3-A, 3-B and 3-C are all shown in Figure 4.5. The two cyan-colored lines divides the wells into three parts. The upper vertical part is Case 3-A, the middle part is Case 3-B, and the ending horizontal part is Case 3-C. The three black lines indicate which two points we chose to consider when we only need the two closest points.

The cases 3-D, 3-E and 3-F are shown in Figure 4.6. They are constructed in the exact same way as the previous three cases, with the only exception that the wells are heading *eastwards* instead of northwards.

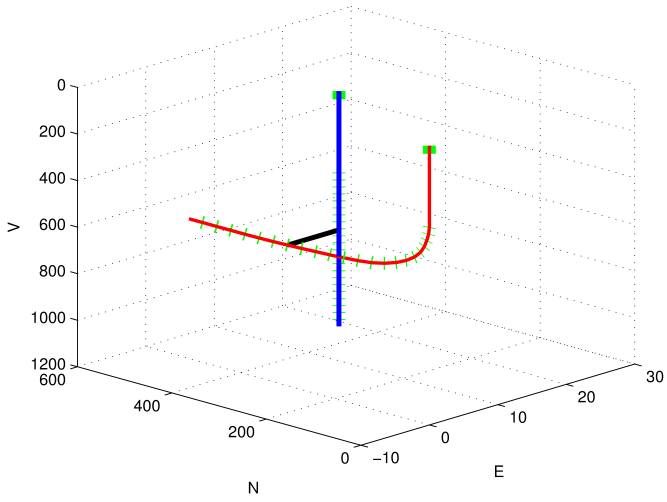

Figure 4.1: The artificial well pair in Case 1-A. This is exactly the same pair of wells that were shown and described in Figure 3.3. The crossing angle is designed to be 90° in the NV plane, while the shortest distance is 7.5 m. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [300, 7.5, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [0, 0, 0]^t$ for the red offset well.

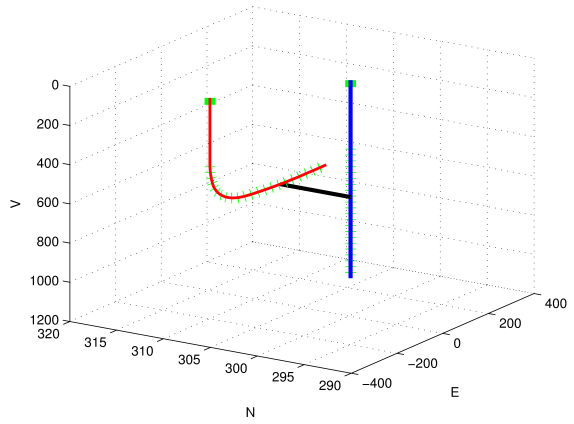

Figure 4.2: The artificial well pair in Case 1-B. The crossing angle is designed to be 90° in the EV plane, while the shortest distance is 7.5 m. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [300, 7.5, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [307.5, -300, 0]^t$ for the red offset well.

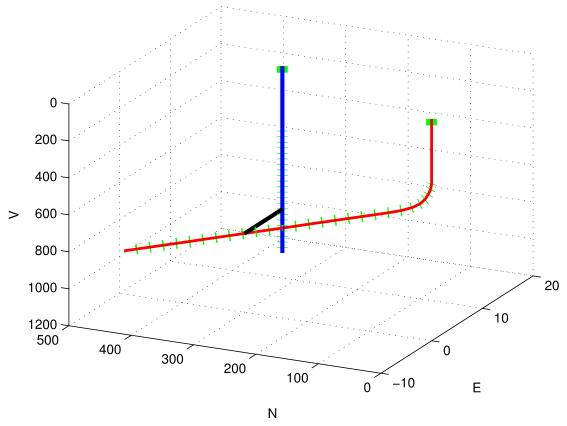

Figure 4.3: The artificial well pair in Case 2-A. This is exactly the same pair of wells that were shown and described in Figure 4.1, except that the red offset well now has an inclination of 45° in the NV plane, resulting in a crossing angle of 45°. The shortest distance is 7.5 m. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [300, 7.5, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [0, 0, 0]^t$ for the red offset well.

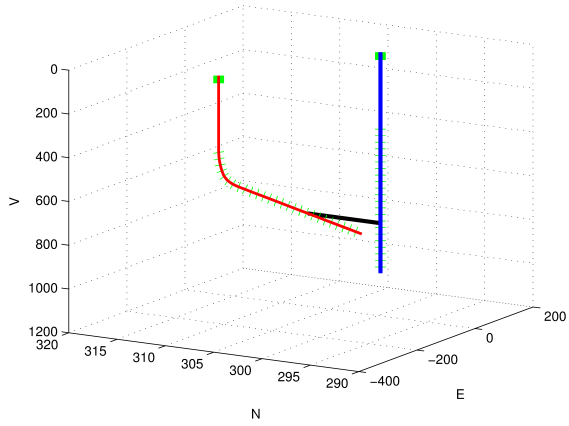

Figure 4.4: The artificial well pair in Case 2-B. This is exactly the same pair of wells that were shown and described in Figure 4.2, except that the red offset well now has an inclination of 45° in the EV plane, resulting in a crossing angle of 45°. The shortest distance is 7.5 m. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [300, 7.5, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [307.5, -300, 0]^t$ for the red offset well.

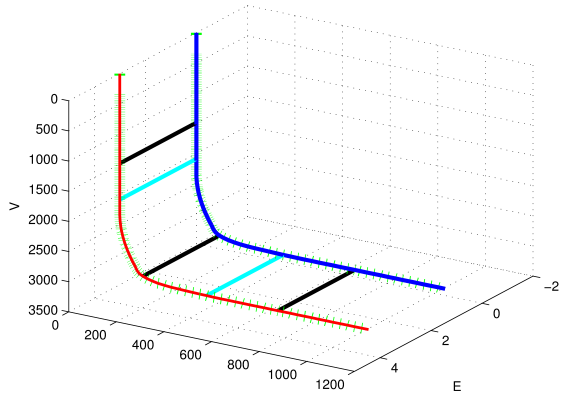

Figure 4.5: The first artificial well pair in Case 3. The wells are designed to be parallel, while their distance is 3 m. The well segments in considered in the three sub-cases 3-A, 3-B and 3-C are limited by two cyan lines (the second and fourth horizontal line from the top). The black lines indicate which points we consider to be the closest ones for each case. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [0, 0, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [0, 3, 0]^t$ for the red offset well.

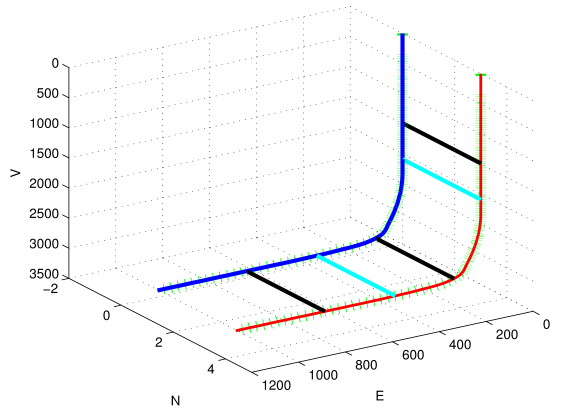

Figure 4.6: The second artificial well pair in Case 3. The wells are designed to be parallel, while their distance is 3 m. The well segments in considered in the three sub-cases 3-D, 3-E and 3-F limited by two cyan lines (the second and fourth horizontal line from the top). The black lines indicate which points we consider to be the closest ones for each case. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [0, 0, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [3, 0, 0]^t$ for the red offset well.

4.1.4 Test Case 4: Horizontally Crossing Wells

In the fourth case, as depicted in Figure 4.7, both wells are going straight down at first. At some chosen deep level, the inclination starts to increase, and they cross with a crossing angle of 90° in the NE plane when both wells are heading in the horizontal direction. As seen in the Figure, this case is parted into three sub-cases, called 4-A, 4-B and 4-C, in order to better investigate the influence of the measured depth of the wells on the risk of well collision. The well pair in Case 4-A crosses at a vertical depth of about 1400 meters. For Case 4-B that vertical depth is about 3400 meters, while it is about 5400 meters for Case 4-C. The shortest distance is about 10 meters in all sub-cases.

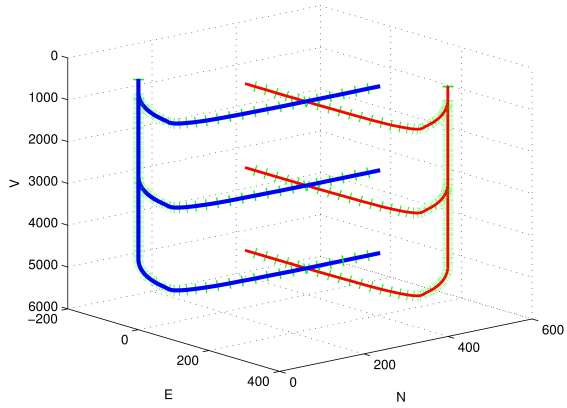

Figure 4.7: The three artificial well pairs in Case 4. In all the sub-cases, the wells are designed to cross when both wells are going in the horizontal direction, while the shortest distance is 10 m. Case 4-A is the upper pair. Case 4-B is the middle pair, while Case 4-C is the deepest pair. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [0, 0, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [400, 400, 10]^t$ for the red offset well.

4.1.5 Test Case 5: Horizontal Crossing Angle of 45°

The fifth case, as depicted in Figure 4.8, is very much the same as Case 4-A, except that the crossing angle is now 45° in the NE plane. Both wells are going straight down at first. Then the inclination starts to increase, and they cross when both wells are heading in the horizontal direction.

4.1.6 Test Case 6: A Real Well Pair

The sixth test case consists of the real well pair that we considered in the *NEV* covariance matrix example in Section 2.5.3. We repeat the well geometry in Figure

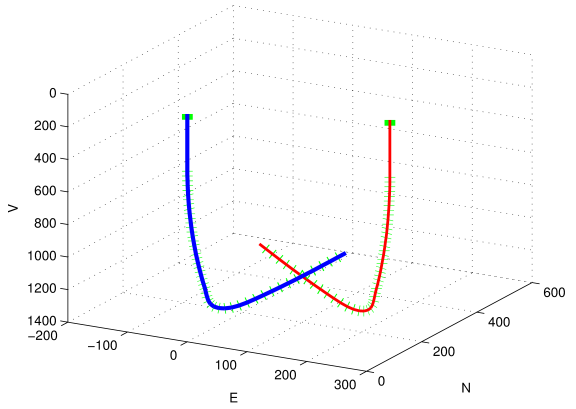

Figure 4.8: The artificial well pair in Case 5. The wells are designed to cross when both wells are going in the horizontal direction, with a crossing angle of 45°. The shortest distance is 10 m. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [0, 0, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [150, 270, 12]^t$ for the red offset well.

4.9. As before, the shortest distance between the wells is 15.59 m.

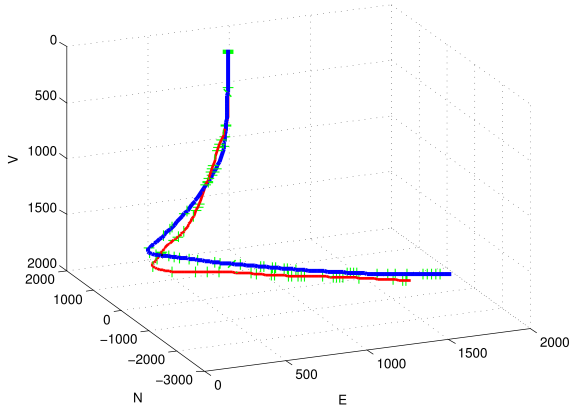

Figure 4.9: The two real petroleum wells in Case 6. This is exactly the same pair of wells that were shown and described in Figure 2.7. The shortest distance between the wells is 15.59 m, which is in the area down to the left where the wells seem to cross. The starting points are $\mathbf{p}_0 = [N_0, E_0, V_0]^t = [970.10, 816.50, 0]^t$ for the blue reference well and $\mathbf{q}_0 = [N_0^q, E_0^q, V_0^q]^t = [976.20, 817.70, 0]^t$ for the red offset well.

4.1.7 Shifted Test Cases

In addition to the test cases described in Section 4.1.1 through Section 4.1.6, we will use some *shifted versions* of some of those cases. By that we mean that we will use some of those well pairs without changing their geometry at all, with the only exception that we shift the starting point of one of the wells in order to obtain a specific value of the shortest distance between the wells.

We have made six such cases, named 1-AX, 1-AY, 2-AX, 2-AY, 6-X and 6-Y. For instance, Case 1-AX will look exactly like Case 1-A, except that the shortest distance between this new well pair is 11.1 meters instead of 7.5 meters. The values of the shortest distances are chosen according to the hypothesis tests that we perform in Section 4.2.1. There we will explain the reasoning behind the chosen distances.

We list the well geometry properties of the shifted test cases in the bottom of Table 4.1 in the next section.

4.1.8 Summary

In Table 4.1 we list the most relevant well geometry properties of the well pairs in all the test cases described in Section 4.1.1 through Section 4.1.7.

For each case in the table, the first two columns contains the number of the well pair test case and its corresponding figure number, as given in the previous sections. The *shortest distance* is the shortest Euclidean distance $\hat{r}$ based on the given (planned or measured) *DIA* values for the two wells. The *crossing angle* is specified along with the considered two-dimensional plane.

The value D is the measured (along-hole) distance of the well in the area around the two closest points. From the definitions of the inclination I and azimuth A in Chapter 2, we know that $I = 0°$ is the vertical direction, while $I = 90°$ is the horizontal direction. Also, $A = 0°$ is northwards, $A = 90°$ is eastwards, $A = 270°$ is westwards, while $A = 315°$ is northwestwards. The azimuth is undefined when a well is vertical, while the crossing angle is undefined when the wells are parallel.

4.2 Testing the Collision Risk Methods

Here, we present all the numerical results obtained by performing the collision risk methods from Chapter 3 on all or some of the test cases that we described thoroughly in Section 4.1.

In Section 4.2.1 we present the numerical results from the two hypothesis tests. In Section 4.2.2 and 4.2.3 we present the estimates of the probability of collision, as obtained by using the two-points methods and the all-points methods, respectively.

During these subsections we also discuss and compare the results, and we make some observations in order to describe the various collision risk methods. The applied test cases cover a variety of well geometry characteristics, which should provide a good basis for making observations about the methods and their results. However, for most of the different test cases, there is surely a large amount of various well parameters that make up the well geometries, making it a challenge

Table 4.1: A comprehensive summary of the well geometry properties of all the test cases described in this chapter. The shifted test cases from Section 4.1.7 are listed in the bottom of the table, below the bar. The value D is the measured (along-hole) distance of the well in the area around the two closest points, while I is the inclination and A is the azimuth. The crossing angle is undefined when the wells are parallel.

Well pair case	Fig.	Short. dist. $\hat{r}$ (m)	**Well geometry around the closest points** Cross. angle	**Reference well** D (m)	I	A	**Offset well** D (m)	I	A
1-A	4.1	7.5	90° NV	620	0°	-	830	90°	0°
1-B	4.2	7.5	90° EV	620	0°	-	830	90°	90°
2-A	4.3	7.5	45° NV	800	0°	-	890	45°	0°
2-B	4.4	7.5	45° EV	800	0°	-	890	45°	90°
3-A	4.5	3.0	-	1490	0°	-	1490	0°	-
3-B	4.5	3.0	-	3290	30°	0°	3290	30°	0°
3-C	4.5	3.0	-	3890	90°	0°	3890	90°	0°
3-D	4.6	3.0	-	1490	0°	-	1490	0°	-
3-E	4.6	3.0	-	3290	20°	90°	3290	20°	90°
3-F	4.6	3.0	-	3890	90°	90°	3890	90°	90°
4-A	4.7	10.0	90° NE	1610	90°	0°	1610	90°	270°
4-B	4.7	10.0	90° NE	3620	90°	0°	3620	90°	270°
4-C	4.7	10.0	90° NE	5600	90°	0°	5600	90°	270°
5	4.8	10.0	45° NE	1610	90°	0°	1580	90°	315°
6	4.9	15.6	40° NE	2450	90°	160°	2400	90°	200°
1-AX	4.1	11.1	90° NV	620	0°	-	830	90°	0°
1-AY	4.2	11.2	90° NV	620	0°	-	830	90°	45°
2-AX	4.3	10.0	45° NV	800	0°	-	890	45°	0°
2-AY	4.3	10.1	45° NV	800	0°	-	890	45°	0°
6-X	4.9	7.4	40° NE	2450	90°	160°	2400	90°	200°
6-Y	4.9	29.5	40° NE	2450	90°	160°	2400	90°	200°

find explanations to the observations on a case-by-case level. Nevertheless, we make some attempts on comparing the overall results from the various methods, in order to detect any apparent differences or similarities between them. We will also make some case-by-case observations in Section 4.3, where we are focusing on the influence of a few chosen well geometry parameters.

In order to apply the various collision risk methods from Chapter 3, we have made a MATLAB implementation that we run on an Intel Core i7 Quad 2.80 GHz processor. When performing the non-analytical methods for estimating the probability of collision, we strive to obtain a relative error RE of order $\sim 10^{-2}$ for each test case.

We present the results using a number of tables, one for each method. All the tables begin with the column *well pair case* that contains the numbers of the applied well pair test cases, where we follow the numbering in Section 4.1.1 through Section 4.1.7. The *shortest distance* is the shortest Euclidean distance $\hat{r}$ based on the given (planned or measured) *DIA* values for the two wells. The *run time* is the computational time given in seconds.

In order to save some space in the tables containing the numerical results, we will from now on use the notation $XeY = X \cdot 10^Y$.

4.2.1 The Hypothesis Tests

In Table 4.2 we present the numerical results obtained by performing the two hypothesis tests in Section 3.1 on the well pair test cases.

The null hypothesis is $H_0 : \mu_r \leq r_1 + r_2 = 1$ m for both the approximated test in Section 3.1.1, as well as for the more exact test in Section 3.1.2. For each case, we give the resulting *p-value* for both tests, and we conclude on whether or not the null hypothesis should be rejected based on the significance value $\alpha = 1/500 = 0.002$. For the approximated test we also include the *separation factor* ω as calculated by expression (3.8).

The first obvious observation to make from Table 4.2 is that the p-value is higher for the more exact test than for the approximated test for all the test cases. In other words, the exact test is *more conservative* than the approximated tests for all the cases. This is further illustrated in Figure 4.10. Here we compare the p-values for the two tests by plotting all the values from Table 4.2 in a $\log_{10}$ plot. The line $\alpha = 0.002$ is plotted in order to easily see when the null hypothesis is to be rejected. We reject H_0 when the p-value is below the significance value α.

For Case 4-A, 5 and 6, the difference between the p-values is seen to be especially extreme, in the way that the approximated test is very far away from rejecting H_0, while the more exact test is very far away from non-rejection. On the other hand, the p-values are not very different in Case 1-A, 1-B, 2-A and 2-B. Also, both the p-values and their difference varies when we consider well pairs that have the same shortest distance between them. In total, this indicates that both the p-values themselves and the difference between them is very much dependent on the well geometry of the well pair of interest.

For the six subcases of Case 3 and the three subcases of Case 4, the pattern seems to be that the p-values from both tests are increasing when the measured

Table 4.2: Numerical results obtained by performing the **hypothesis tests** described in Section 3.1 on the well pair test cases.

Well pair case	Short. dist. $\hat{r}$ (m)	Run time (sec)	**Approximated test** p-value	Sep. fact. ω	Rej. H_0	**More exact test** p-value	Rej. H_0
1-A	7.5	15	3.14e-2	0.647	No	4.10e-2	No
1-B	7.5	16	7.10e-2	0.510	No	0.102	No
2-A	7.5	15	1.85e-2	0.725	No	2.34e-2	No
2-B	7.5	16	5.32e-2	0.561	No	7.41e-2	No
3-A	3.0	16	9.57e-3	0.814	No	0.132	No
3-B	3.0	17	0.121	0.406	No	0.598	No
3-C	3.0	16	0.372	0.113	No	0.872	No
3-D	3.0	16	9.57e-3	0.814	No	0.132	No
3-E	3.0	16	0.141	0.374	No	0.616	No
3-F	3.0	17	0.452	0.065	No	0.926	No
4-A	10.0	17	6.88e-10	2.105	**Yes**	0.175	No
4-B	10.0	17	1.28e-3	1.048	**Yes**	0.228	No
4-C	10.0	18	2.49e-2	0.681	No	0.349	No
5	10.0	17	8.40e-11	2.220	**Yes**	2.48e-2	No
6	15.6	19	3.30e-11	2.269	**Yes**	0.112	No
1-AX	11.1	15	1.99e-3	1.000	**Yes**	2.40e-3	No
1-AY	11.2	16	1.67e-3	1.020	**Yes**	1.99e-3	**Yes**
2-AX	10.0	16	1.99e-3	1.000	**Yes**	2.38e-3	No
2-AY	10.1	16	1.68e-3	1.020	**Yes**	1.99e-3	**Yes**
6-X	7.4	19	1.99e-3	1.000	**Yes**	0.511	No
6-Y	29.5	17	1.06e-13	4.432	**Yes**	1.99e-3	**Yes**

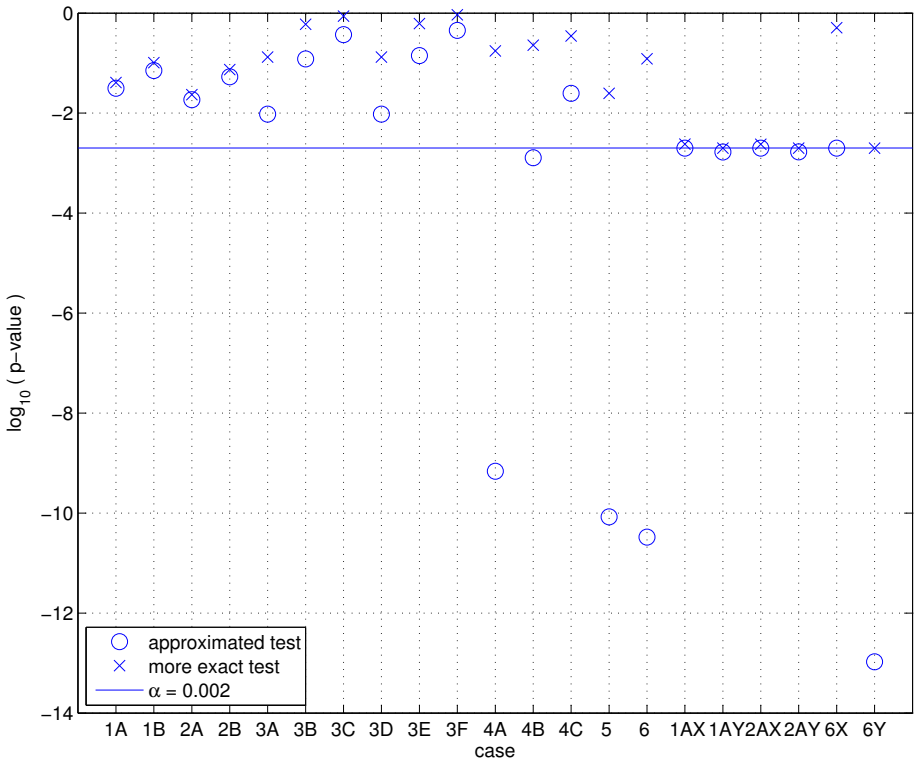

Figure 4.10: Graphical illustration of 10 base logarithm of the p-values for the two different tests, using the values in Table 4.2.

depth of the two closest points is increasing. We will discuss this phenomenon further in Section 4.3.3, where we make more observations about the influence of the measured depth on the collision risk.

Now, we finally explain the reasoning behind the *shifted well pair cases* that we introduced in Section 4.1.7. The hypothesis test results for these cases are shown below the bar in Table 4.2. In the test cases marked with an X, we have tuned the shortest distance of the corresponding original case to be the minimum shortest distance that allows the null hypothesis of the approximated test to be rejected, resulting in the p-value $p = 1.99\text{e-}3$ and the separation factor $\omega = 1.00$. For instance, Case 1-AX is made by shifting one of the wells in Case 1-A away from the second well in the direction of the vector connecting the two closest points. We stop the shifting when we reach the distance for which the null hypothesis in the approximated test is rejected. Then the shortest distance is 11.1 meters instead of the original 7.5 meters. The same argument holds for the test cases marked with a Y, except that we have now tuned the shortest distance according to the more exact test, resulting in a p-value of $p = 1.99\text{e-}3$ in that test.

For Case 1-AX, 2-AX and 6-X, the two tests disagree on whether H_0 should be rejected or not. For all these three shifted test cases, the approximated test is barely rejecting H_0. For Case 6-X, the more exact test is very far from rejection, while for the other two cases, the p-values of the more exact test are much closer to those of the approximated test. The situation is similar for Case 1-AY, 2-AY and 6-Y. For these cases, the more exact test barely rejects H_0. For Case 6-Y, the approximated test is clearly rejecting H_0, while the p-values are much closer for Case 1-AY and 2-AY. In sum, this does not indicate any immediately clear relationship between the rejection areas for the two tests, in the way that when one of the tests is on the border of rejection, having a p-value of $p = 1.99\text{e-}3$, the other test does *not* result in a specific predetermined p-value that is independent of the well geometry.

For the real well pair in Case 6, the probability distribution functions (pdf's) $f(r)$ under H_0 for both tests are plotted in Figure 4.11. Just as for Case 1-A, as seen in Figure 3.4 at the end of Section 3.1.2, the distributions are seen to differ significantly. If $\hat{r}$ denotes the measured shortest distance, then the p-values are found as the area to the right of $r = \hat{r}$ below the graphs. The p-values are seen to be very different at the actual shortest distance $\hat{r} = 15.6$ m for Case 6.

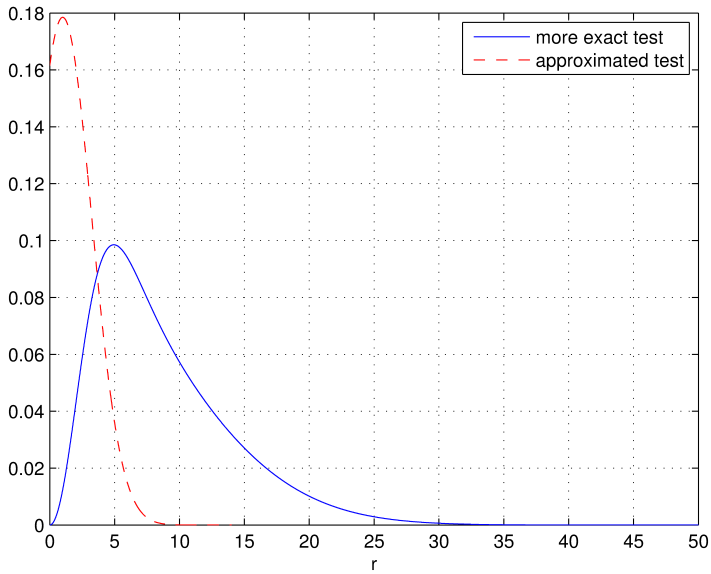

Figure 4.11: The pdf's of the shortest distance r under H_0 for the two hypothesis tests as performed on Case 6. The pdf's are almost exactly the same for the shifted test cases 6-X and 6-Y.

In fact, since the shifted test cases 6-X and 6-Y only differ from Case 6 in their shortest distances, the corresponding pdf's for these cases are found to be virtually indistinguishable from the pdf's for Case 6 shown in Figure 4.11. For these two cases, the p-values are found as the areas below the graphs to the right

of $r = \hat{r} = 7.4$ m and $r = \hat{r} = 29.5$ m, respectively.

For the more exact test, this area is easily seen to be much larger for 6-X than for Case 6, while the area for Case 6 is in turn much larger than for Case 6-Y. For the approximated test, it is not that easy to see these differences directly from Figure 4.11. However, we know that the pdf belongs to a normal distribution, implying that the graph decreases exponentially. In sum, this helps us understand the differences between the p-values for Case 6-X, 6 and 6-Y on the $\log_{10}$ plot in Figure 4.11.

4.2.2 The Collision Probability Considering Two Points

Here we begin by performing the analytical two-points method for calculating the probability of well collision $\hat{l}_{\text{two}}^{\text{analytical}}$, as descibed in Section 3.2.1, on the well pair test cases in Section 4.1. The results are presented in Table 4.3.

For Case 6-Y, the collision probability is in practice equal to zero, but we have chosen to include the obtained value in order to be able to know if the CMC and CE simulation methods actually give a result even when the collision probability is of order $\sim 10^{-41}$.

Now, in the $\log_{10}$ plot in Figure 4.12 we compare the p-values in the hypothesis tests from Table 4.2 with the analytically calculated collision probability values from Table 4.5. From Figure 4.12 we can make the remarkable observation that the collision probability is smaller than both p-values in all the cases. For some cases, there is not a huge difference, while in cases like 3-F, 5, 6 and 6-Y, the difference is seen to be larger. Both the collision probability and its difference to the p-values seems to be dependent on the well geometry of the well pair of interest.

Combining with the observations in the previous section, the pattern is that the collision probability is smaller than the p-value of the approximated hypothesis test, which in turn is smaller than the p-values of the more exact hypothesis test.

The numerical results reveal that there exist well geometries where a risk assessment based on a hypothesis test gives a different result compared to a risk assessment based on the collision probability. For instance, if we demand the collision probability to be below the value of 2e-4 to continue the drilling, then we *would not* accept further drilling for the well pair in Case 1-AY considering the collision probability $\hat{l}_{\text{two}}^{\text{analytical}} = 5.679\text{e-}4$, although we *would* reject H_0 in both hypothesis tests based on the significance level $\alpha = 0.002$. However, for the horizontally crossing well pair in Case 4-C, we would indeed accept further drilling considering the collision probability $\hat{l}_{\text{two}}^{\text{analytical}} = 1.588\text{e-}4$, while we would not reject H_0 in any of the hypothesis tests.

If we had chosen the toleration value for the collision probability to be equal to the significance value α, then the hypothesis tests would be more conservative than the probability requirement for all the test cases, following from the observation that the collision probability is smaller than both p-values for all the cases. However, there is no obvious reason to treat the two collision risk measures in such a similar way, because of the fundamental differences between the two concepts.

From Chapter 3, a p-value is the probability of obtaining the shortest distance $\hat{r}$ given that the expected distance implies a collision, while the collision probability is

Table 4.3: Numerical results obtained by performing the **analytical two-points** method on the well pair test cases for calculating the probability of collision.

Well pair case	Shortest distance $\hat{r}$ (m)	Run Time (sec)	**Collision probability** $\hat{l}_{\text{two}}^{\text{analytical}}$
1-A	7.5	0.08	9.629e-3
1-B	7.5	0.11	1.776e-2
2-A	7.5	0.10	5.908e-3
2-B	7.5	0.11	1.368e-2
3-A	3.0	0.50	1.411e-3
3-B	3.0	0.40	6.878e-3
3-C	3.0	0.41	6.223e-3
3-D	3.0	0.43	1.411e-3
3-E	3.0	0.39	7.795e-3
3-F	3.0	0.67	3.852e-3
4-A	10.0	0.14	1.57e-12
4-B	10.0	0.48	1.370e-5
4-C	10.0	1.16	1.588e-4
5	10.0	0.21	9.90e-13
6	15.6	1.75	1.45e-13
1-AX	11.1	0.08	6.784e-4
1-AY	11.2	0.08	5.679e-4
2-AX	10.0	0.09	6.723e-4
2-AY	10.1	0.10	5.648e-4
6-X	7.4	1.73	1.833e-5
6-Y	29.5	1.63	7.43e-41

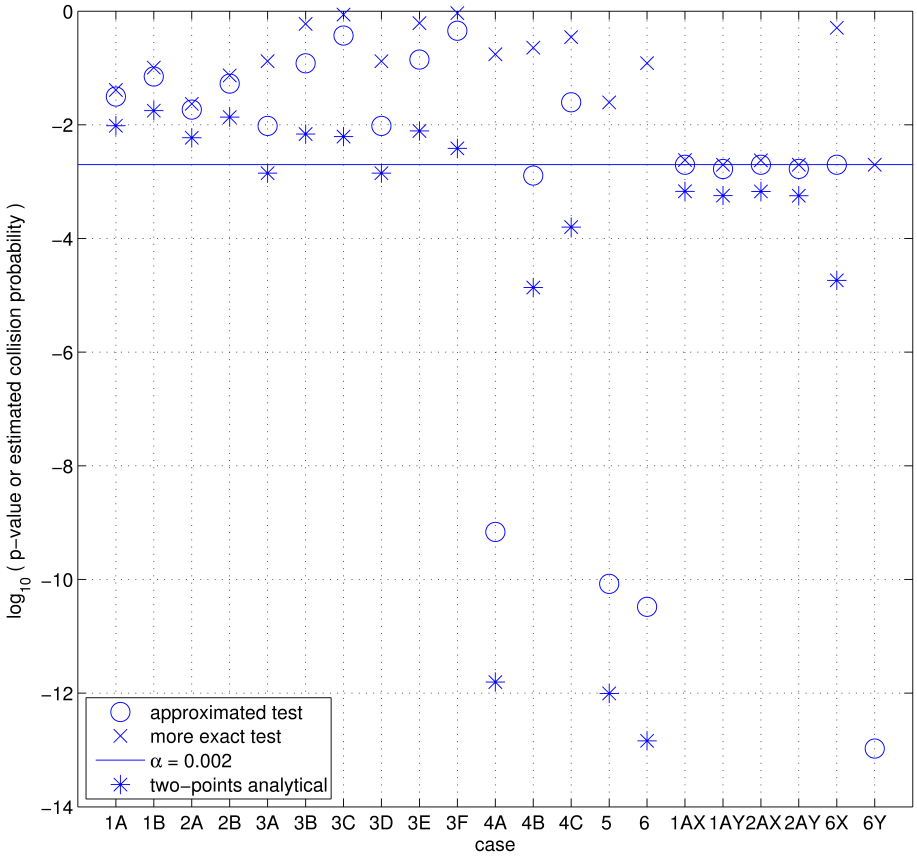

Figure 4.12: Graphical illustration of the 10 base logarithm of the p-values from Table 4.2 together with the calculated collision probability values $\hat{l}_{\text{two}}^{\text{analytical}}$ from Table 4.3. Here we have intentionally not included the probability value $\hat{l}_{\text{two}}^{\text{analytical}} = 7.43\text{e-}41$ for Case 6-Y, which is in practice equal to zero.

the probability of r implying a collision, given that the expected shortest distance is approximated by $\hat{r}$. In Figure 4.13 we plot the pdf's of $r = \sqrt{\mathbf{d}^t\mathbf{d}}$, as first defined in expression (3.11), both when considering the exact hypothesis test and when considering the analytical collision probability, by using expression (3.12) in both situations. The p-value

$$p = P\left(\mathbf{d}^t\mathbf{d} \geq \hat{\mathbf{d}}^t\hat{\mathbf{d}} \mid \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d \leq (r_1 + r_2)^2\right)$$

of the more exact test is found as the area below the blue solid graph to the *right*

of $r = \hat{r} = 15.6$ m, while the collision probability

$$\hat{l}_{\text{two}}^{\text{analytical}} = P\left(\mathbf{d}^t\mathbf{d} < (r_1 + r_2)^2 \mid \boldsymbol{\mu}_d^t\boldsymbol{\mu}_d = \hat{r}^2\right)$$

is found as the area below the red dashed graph to the *left* of $r = r_1 + r_2 = 1$ m.

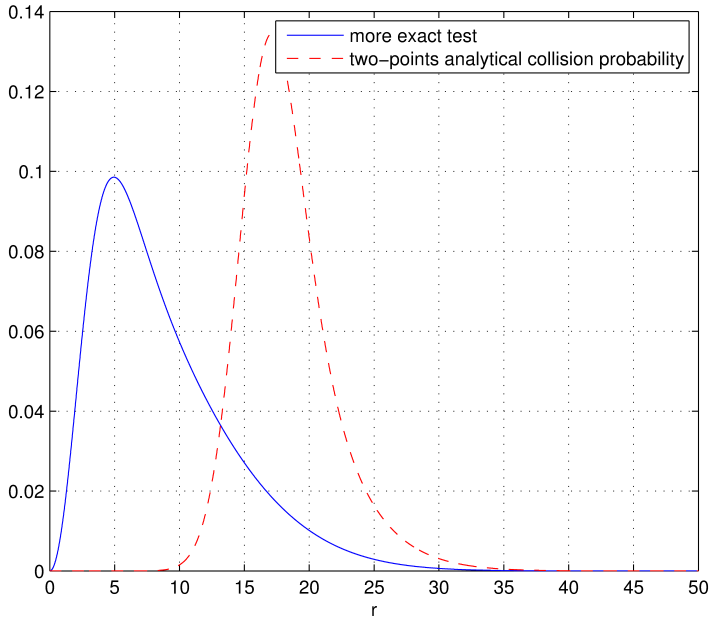

Figure 4.13: The pdf's of the shortest distance r, when considering the more exact hypothesis test, and when considering the analytical two-points collision probability.

All in all, we have to consider two very different distributions, just like when we compared the pdf's of the two hypothesis tests in Figure 4.11. For the test case in Figure 4.13, it is easy to see that $p >> \hat{l}_{\text{two}}^{\text{analytical}}$. However, in the general case, there is need for a deeper study of the properties of the distributions. The fact that we fall short on finding a more rigid explanation of the observed phenomenon about the collision probability tending to be smaller than the p-values, is obviously not a strong argument for the general validity of the observation. On the other hand, the observation is strengthened by the fact that we have used a variety of well geometries in our test cases.

When we consider the shifted well pair cases 1-AX, 1-AY, 2-AX, 2-AY, 6-X and 6-Y in Figure 4.12 we observe that the collision probability $\hat{l}_{\text{two}}^{\text{analytical}}$ is not the same for all the X cases (where H_0 is barely rejected in the approximated test), neither is it the same for all the Y cases (where H_0 is barely rejected in the more exact test). This means that there is no immediately clear relationship between rejection of H_0 and the numerical value of the collision probability, except for the one that the collision probability always seems to be less than both p-values.

In Table 4.6 we now give the numerical results obtained by performing the two-points crude Monte Carlo (CMC) simulation method for estimating the probability of collision, as described in Section 3.2.1, on the well pair test cases. The column B contains the sample size values in the MC simulation method. The remaining columns contain the estimated collision probability $\hat{l}_{\text{all}}^{\text{CMC}}$, its standard error $\hat{\sigma}_{\hat{l}}$, as well as the relative error RE.

In Table 4.5 we give the numerical results for the two-points cross-entropy (CE) simulation method, as described in Section 3.2.3. The table is very similar structured as Table 4.4, except that we now also include the sample size B_1 and the number of iterations needed. When performing the CE method we have used the maximum number of iterations $t_{\max} = 20$ and the rarity parameter $\rho = 0.1$.

It is worth noting that in Case 4-A, 5, 6 and 6-Y, the collision probability is so small that the CMC method is not able to give a result, since it would need a inconveniently large number of samples to gain the needed accuracy. On the other hand, the CE method does indeed give accurate results in these cases. This shows the strength of the CE method versus the CMC method: In a rare-event situation, the CE method is able to give answers when the CMC method is not.

When we compare the results given in Table 4.3, 4.4 and 4.5 we immediately see that the collision probability are very similar for the three methods for all the cases. This supports both the CMC method and the CE in the way that they both seem to give an accurate estimate, at least when the sample size is large enough. This also supports the correctness of the much faster calculation of the analytical calculated probability values $\hat{l}_{\text{two}}^{\text{analytical}}$ in Table 4.3. This in turn supports the correctness of the p-values of the more exact hypothesis test in Table 4.2, since the calculations in both cases are based on the same cdf $P(\mathbf{d}^t\mathbf{d} \leq t)$, as given in expression (3.12).

4.2.3 The Collision Probability Considering All Points

In Table 4.6 we present the numerical results obtained by performing the all-points crude Monte Carlo (CMC) simulation method for estimating the probability of collision, as described in Section 3.2.2, on all the well pair test cases in Section 4.1, except for the shifted cases.

The column B contains the sample size values in the MC simulation method. The remaining columns contain the estimated collision probability $\hat{l}_{\text{all}}^{\text{CMC}}$, its standard error $\hat{\sigma}_{\hat{l}}$, as well as the relative error RE.

In Table 4.7 we give the numerical results for all-points cross-entropy (CE) simulation method, as described in Section 3.2.3. As for the two-points method, we now also include the sample size B_1 and the number of iterations needed. When performing the CE method we have used the maximum number of iterations $t_{\max} = 5$ and the rarity parameter $\rho = 0.1$.

Exactly as for the two-points methods, there are some cases (4-A, 5 and 6) in which the collision probability seems to be so small that we are unable to obtain a result when we apply the CMC method. In these cases, the CE method still seem to give accurate results. This once again indicates that the CE method is better in this kind of rare-event situations.

Table 4.4: Numerical results obtained by applying the **two-points CMC** simulation method for estimating the collision probability on the well pair test cases.

Well pair case	Shortest distance $\hat{r}$ (m)	B	Run time (sec)	**Collision probability & measures of uncertainty** $\hat{l}_{\text{two}}^{\text{CMC}}$	$\hat{\sigma}_{\hat{l}}$	RE
1-A	7.5	1e8	114	9.624e-3	9.76e-6	1.01e-3
1-B	7.5	1e8	115	1.776e-2	1.32e-5	7.44e-4
2-A	7.5	1e6	113	5.915e-3	7.67e-6	1.30e-3
2-B	7.5	1e8	115	1.368e-2	1.16e-5	8.49e-4
3-A	3.0	1e8	116	1.408e-3	3.75e-6	2.66e-3
3-B	3.0	1e8	115	6.878e-3	8.26e-6	1.20e-3
3-C	3.0	1e8	115	6.217e-3	7.86e-6	1.26e-3
3-D	3.0	1e8	115	1.413e-3	3.76e-6	2.66e-3
3-E	3.0	1e8	115	7.789e-3	8.79e-6	1.13e-3
3-F	3.0	1e8	114	3.854e-3	6.20e-6	1.61e-3
4-A	10.0	1e8	116	0	0	-
4-B	10.0	1e8	114	1.376e-5	3.71e-7	2.70e-2
4-C	10.0	1e8	115	1.672e-4	1.29e-6	7.73e-3
5	10.0	2e8	230	0	0	-
6	15.6	1e8	123	0	0	-
1-AX	11.1	1e8	125	6.798e-4	2.61e-6	3.83e-3
1-AY	11.2	1e8	124	5.657e-4	2.38e-6	4.20e-3
2-AX	10.0	1e8	123	6.694e-4	2.59e-6	3.86e-3
2-AY	10.1	1e8	122	5.590e-4	2.36e-6	4.21e-3
6-X	7.4	1e8	123	1.800e-5	4.24e-7	2.36e-2
6-Y	29.5	1e8	125	0	0	-

Table 4.5: Numerical results obtained by performing the **two-points CE** simulation method for estimating the probability of collision on the well pair test cases. For all tests, we have set $t_{\max} = 20$ and $\rho = 0.1$.

Well pair case	Short. dist. $\hat{r}$ (m)	B	B_1	No. of itera-tions	Run time (sec)	**Collision probability & measures of uncertainty** $\hat{l}_{\text{two}}^{\text{CE}}$	$\hat{\sigma}_{\hat{l}}$	RE
1-A	7.5	2e7	2e6	20	94	9.627e-3	7.02e-6	7.29e-4
1-B	7.5	2e7	2e6	20	94	1.775e-2	1.43e-5	8.08e-4
2-A	7.5	2e7	2e6	3	39	5.898e-3	4.22e-6	7.15e-4
2-B	7.5	2e7	2e6	20	95	1.369e-2	1.09e-5	7.99e-4
3-A	3.0	2e7	2e6	3	37	1.410e-3	1.25e-6	8.86e-4
3-B	3.0	2e7	2e6	5	44	6.851e-3	9.69e-6	1.41e-3
3-C	3.0	2e7	2e6	5	45	6.261e-3	1.66e-5	2.66e-3
3-D	3.0	2e7	2e6	3	40	1.413e-3	1.25e-6	8.85e-4
3-E	3.0	2e7	2e6	20	94	7.799e-3	1.12e-5	1.44e-3
3-F	3.0	2e7	2e6	20	94	3.831e-3	1.35e-5	3.54e-3
4-A	10.0	2e7	2e6	5	45	1.59e-12	4.6e-14	2.88e-2
4-B	10.0	2e7	2e6	20	94	1.361e-5	6.23e-8	4.57e-3
4-C	10.0	2e7	2e6	20	95	1.686e-4	9.30e-7	5.52e-3
5	10.0	2e7	2e6	5	47	9.74e-13	1.0e-15	1.03e-2
6	15.6	2e7	2e6	20	94	1.44e-13	1.2e-15	8.62e-3
1-AX	11.1	2e7	2e6	20	92	6.779e-4	5.17e-7	7.63e-4
1-AY	11.2	2e7	2e6	20	91	5.678e-4	4.34e-7	7.65e-4
2-AX	10.0	2e7	2e6	4	41	6.714e-4	5.01e-7	7.47e-4
2-AY	10.1	2e7	2e6	4	44	5.646e-4	4.23e-7	7.49e-4
6-X	7.4	2e7	2e6	5	48	1.822e-5	8.71e-8	4.78e-3
6-Y	29.5	2e7	2e6	20	94	7.70e-41	1.4e-42	1.82e-2

Table 4.6: Numerical results obtained by performing the **all-points CMC** simulation method for estimating the probability of collision on some of the well pair test cases.

Well pair case	Short. dist. $\hat{r}$ (m)	B	Run time (sec)	**Collision probability & measures of uncertainty** $\hat{l}_{\text{all}}^{\text{CMC}}$	$\hat{\sigma}_{\hat{l}}$	RE
1-A	7.5	1e5	1927	1.621e-2	3.99e-4	2.46e-2
1-B	7.5	1e5	1861	3.718e-2	5.98e-4	1.61e-2
2-A	7.5	1e5	1921	1.522e-2	3.87e-4	2.54e-2
2-B	7.5	1e5	1899	3.030e-2	5.42e-4	1.79e-2
3-A	3.0	1e5	1186	4.490e-3	2.11e-4	4.71e-2
3-B	3.0	1e5	1739	0.1118	9.97e-4	8.91e-3
3-C	3.0	1e5	859	3.677e-2	5.95e-4	1.62e-2
3-D	3.0	1e5	1177	4.630e-3	2.15e-4	4.64e-2
3-E	3.0	1e5	1814	0.1393	1.09e-3	7.86e-3
3-F	3.0	1e5	775	2.393e-2	4.83e-4	2.02e-2
4-A	10.0	1e5	1954	0	0	-
4-B	10.0	1e5	4118	1.270e-3	7.96e-5	6.27e-2
4-C	10.0	1e5	2069	1.667e-2	4.05e-4	2.43e-2
5	10.0	1e5	1917	0	0	-
6	15.6	1e5	2152	0	0	-

For the cases where the CMC method does in fact give a result, the probability estimates obtained by the CMC method and the CE method are very similar, which supports both methods.

For both methods we have achieved to obtain a relative error RE of order $\sim 10^{-2}$ for all the cases. The computational running time is in general seen to be much less for the CE method than for the CMC method, meaning that it seems to be more efficient. This observation is further illustrated in Figure 4.14, where we have plotted the relative error RE against the run time for Case 4-B, using various sample sizes B for both methods. Here we can see that the relative error decreases when the run time is increasing for both methods, which is reasonable. We also observe that the RE values seems to stabalize at some value, for which there is a lot of additional run time needed in order to decrease the relative error further. This seems to occur at the sample size $B = 5e4$ for the CMC method, and at the sample size $B = 5e4$ for the CE method.

Table 4.7: Numerical results obtained by performing the **all-points CE** simulation method for estimating the probability of collision l_{all} on some of the well pair test cases. For all tests, we have set $t_{\max} = 5$ and $\rho = 0.1$.

Well pair case	Short. dist. $\hat{r}$ (m)	B	B_1	No. of iterations	Run time (sec)	**Collision probability & measures of uncertainty** $\hat{l}_{\text{all}}^{\text{CE}}$	$\hat{\sigma}_{\hat{l}}$	RE
1-A	7.5	5e4	5e3	2	1122	1.668e-2	1.54e-4	9.14e-3
1-B	7.5	5e4	5e3	2	1242	3.726e-2	3.71e-4	9.95e-3
2-A	7.5	5e4	5e3	2	1278	1.492e-2	1.36e-4	9.12e-3
2-B	7.5	5e4	5e3	2	1152	3.128e-2	3.05e-4	9.74e-3
3-A	3.0	2e5	5e3	5	2921	4.428e-3	1.77e-4	3.99e-2
3-B	3.0	5e4	5e3	1	1167	0.1120	1.64e-3	1.46e-2
3-C	3.0	2e5	5e3	5	2257	3.871e-2	5.04e-4	1.30e-2
3-D	3.0	2e5	5e3	5	2958	4.278e-3	1.73e-4	4.04e-2
3-E	3.0	5e4	5e3	1	1169	0.1378	1.61e-3	1.17e-2
3-F	3.0	5e4	5e3	5	642	2.414e-2	8.17e-4	3.38e-2
4-A	10.0	5e4	5e3	5	1574	3.806e-10	6.277e-12	1.65e-2
4-B	10.0	5e4	5e3	3	1500	1.080e-3	1.50e-5	1.38e-2
4-C	10.0	5e4	5e3	2	1482	1.638e-2	2.89e-4	1.77e-2
5	10.0	5e4	5e3	5	1586	6.267e-11	9.27e-13	1.48e-2
6	15.6	2e5	5e3	5	5475	1.495e-8	6.92e-10	4.63e-2

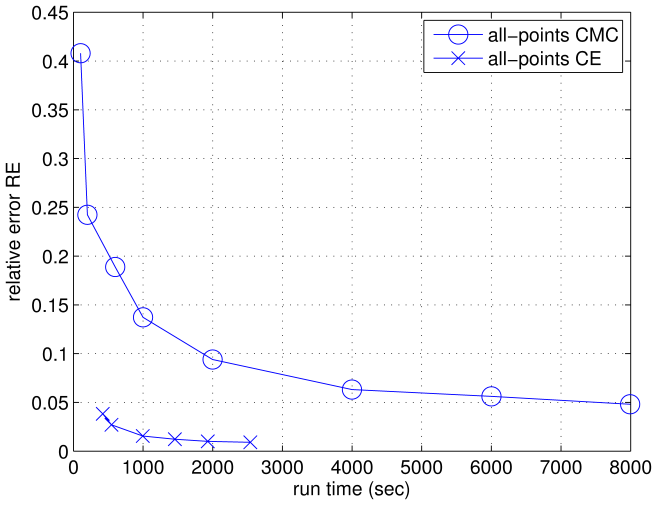

Figure 4.14: The run time versus the relative error RE for a number of different sample sizes B when performing the CMC method and the CE method on Case 4-B. For the CMC method we have used the sample sizes $B \in \{5e3, 1e4, 3e4, 5e4, 1e5, 2e5, 3e5, 4e5\}$. For the CE method we have used the sample sizes $B \in \{5e3, 1e4, 3e4, 5e4, 7e4, 1e5\}$.

We continue by looking at the numerical results from the enhanced Monte Carlo (EMC) simulation method in Table 4.8. Since this method is very time-consuming, we have only applied it on six of the well pair cases. We have especially chosen to use Case 4-A, 5 and 6, since the CMC method gave no results for these cases.

In Table 4.8 the sample size B is the number of samples in the CMC simulation method we use in order to obtain the M probability estimates

$$\hat{l}_\lambda(\lambda_0), \ldots, \hat{l}_\lambda(\lambda_{M-1}).$$

We have chosen to use a number of equally-spaced scaling parameters in the area $\lambda \in \{0.10, 0.82\}$, since for a rare-event, the probability values $\hat{l}_\lambda(\lambda)$ are often too small to be obtained by the CMC method when we are close to $\lambda = 1$.

For each case we have included a regular plot and a $\log_{10}$ plot that shows the resulting interpolation. These are given in Figure 4.15 – 4.20. In the figures, the middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$ obtained by inserting values of $\lambda \in [0, 1]$ in expression (3.29), while the outer red lines indicate the 95% confidence interval from expression (3.27). The M probability estimates

$$\hat{l}_\lambda(\lambda_0), \ldots, \hat{l}_\lambda(\lambda_{M-1})$$

are marked by circles.

For the rare-event cases 4-A, 5 and 6, the relative error RE is extremely large. In all the figures, we have marked the corresponding CE estimate $\hat{l}_{\text{all}}^{\text{CE}}$ with a star. The

Table 4.8: Numerical results obtained by performing the **all-points EMC** simulation method for estimating the probability of collision on some of the well pair test cases.

Well pair case	Short. dist. $\hat{r}$ (m)	B	λ_0	λ_{M-1}	M	Run time (sec)	**Collision probability & measure of uncertainty** $\hat{l}_{\text{all}}$	RE
1-A	7.5	1e4	0.20	0.67	15	2794	1.732e-2	7.53e-2
2-A	7.5	1e4	0.20	0.67	15	2865	1.858e-2	7.27e-2
3-A	7.5	1e4	0.40	0.82	26	3154	4.563e-3	0.148
4-A	10.0	1e4	0.22	0.57	15	2844	8.469e-10	3.44e2
5	10.0	1e4	0.11	0.58	20	3799	4.080e-10	4.95e2
6	15.6	1e4	0.12	0.54	27	5823	9.239e-13	1.04e4

confidence intervals are seen to cover, or at least almost cover, this more accurate estimate in all cases, including the last three. This indicates that the EMC method at least to some degree supports the values obtained by the CE method in the cases where we were unable to get a comparable estimate from the CMC method.

We note that in Figure 4.18 – 4.20, the lower confidence bound is so small that it is essentially equal to zero, which makes it undefined in the $\log_{10}$ plot.

We finish this section by comparing the all-points probability values with the two-points values. As seen in the regular plot in Figure 4.21 and in the $\log_{10}$ plot in Figure 4.22, the probability values $\hat{l}_{\text{all}}^{\text{CE}}$ from the all-points CE method in Table 4.7 are always much larger (often by a factor of order ~ 10) than the corresponding values $\hat{l}_{\text{two}}^{\text{analytical}}$ from the analytical two-points method in Table 4.3. For the all-points values we have used the CE values rather than the CMC values or the EMC values because the relative error is in general seen to be smaller for the CE method. The fact that the probability is larger for the all-points method is exactly as expected from our introduction to Chapter 3.

We observe that the difference between the two methods varies between the cases. For cases like 1-A and 1-B, and for the strictly vertical parallel wells in Case 3-A and 3-D, the difference is not very large. However, for the transition between vertical wells and horizontal wells in Case 3-B and 3-E, and also for the horizontally crossing wells in Case 4-A, 4-B, 4-C and 5, the difference is significant. All in all, the consequence of performing a two-points method instead of an all-points method is seen to be very much depending on the actual well geometry.

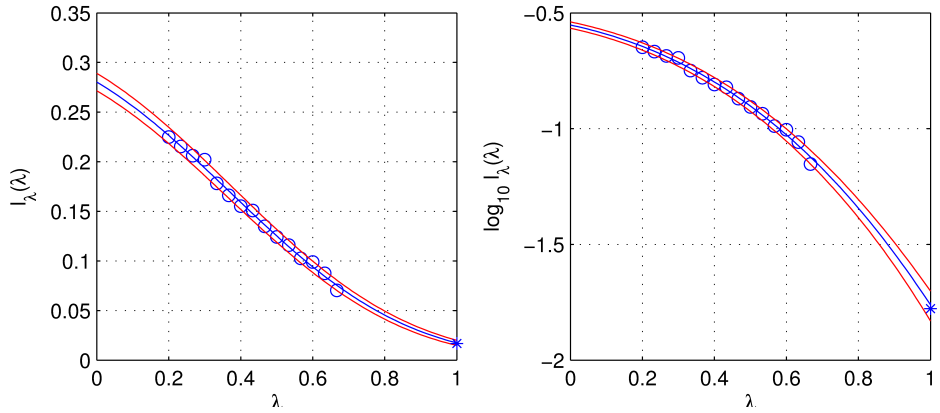

Figure 4.15: The EMC method performed on Case 1-A results in this interpolation. The estimated parameters $q = 0.323$, $a = 0.882$, $b = -0.526$ and $c = 2.838$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) =$ 1.732e-2. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} =$ 1.668e-2, obtained by the CE method, is marked with a star.

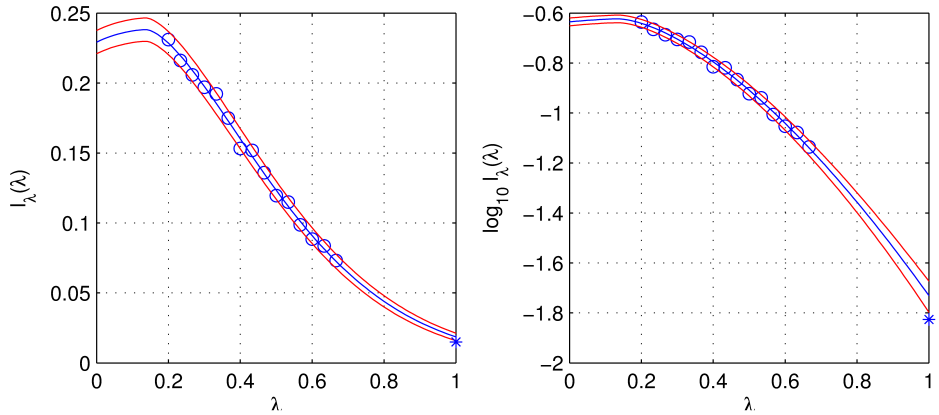

Figure 4.16: The EMC method performed on Case 2-A results in this interpolation. The estimated parameters $q = 0.238$, $a = 3.220$, $b = 0.139$ and $c = 1.561$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) =$ 1.858e-2. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} =$ 1.492e-2, obtained by the CE method, is marked with a star.

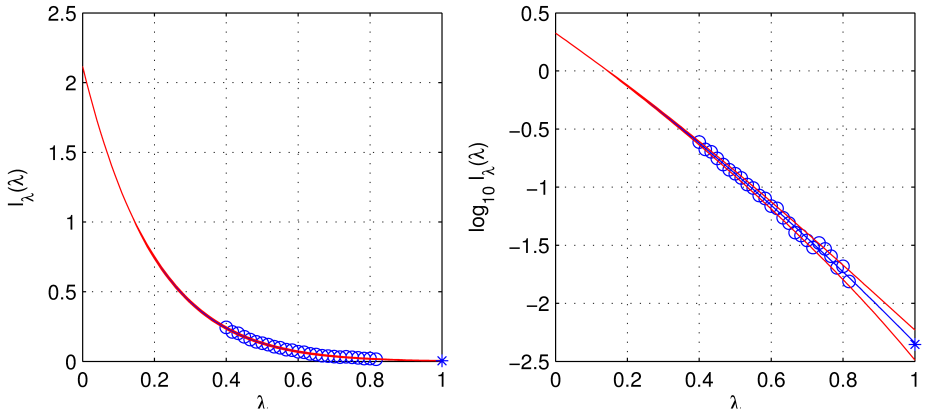

Figure 4.17: The EMC method performed on Case 3-A results in this interpolation. The estimated parameters $q = 2.28\text{e}2$, $a = 1.775$, $b = -1.705$ and $c = 1.816$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) = 4.563\text{e-}3$. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} = 4.428\text{e-}3$, obtained by the CE method, is marked with a star.

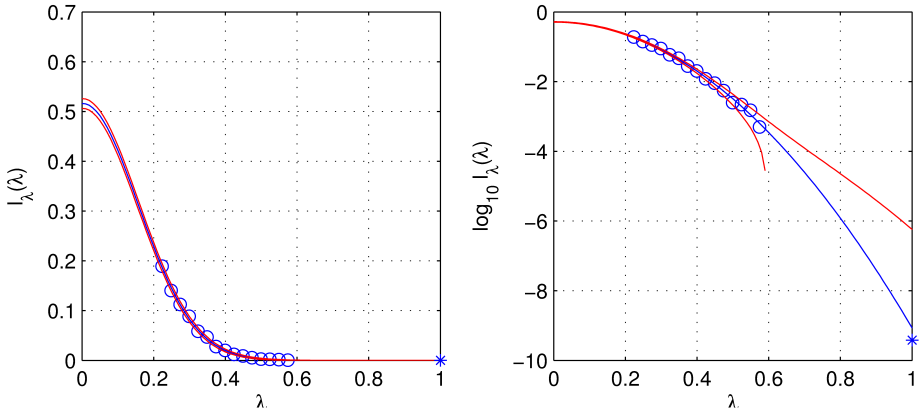

Figure 4.18: The EMC method performed on Case 4-A results in this interpolation. The estimated parameters $q = 0.516$, $a = 20.30$, $b = 1.90\text{e-}3$ and $c = 1.987$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) = 4.469\text{e-}10$. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} = 3.806\text{e-}10$, obtained by the CE method, is marked with a star.

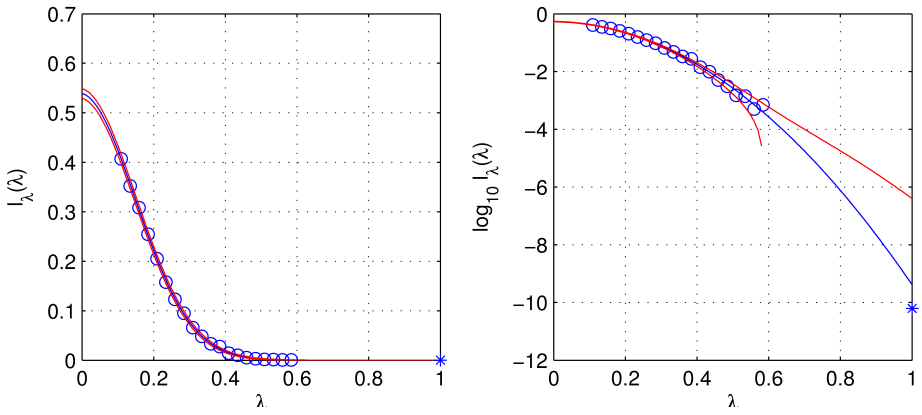

Figure 4.19: The EMC method performed on Case 5 results in this interpolation. The estimated parameters $q = 0.540$, $a = 20.61$, $b = -9.43\text{e-}3$ and $c = 2.010$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) = 4.080\text{e-}10$. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} = 6.267\text{e-}11$, obtained by the CE method, is marked with a star.

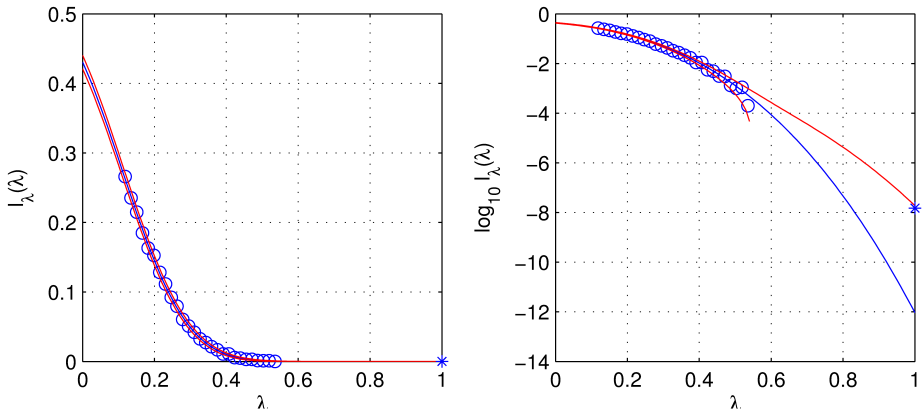

Figure 4.20: The EMC method performed on Case 6 results in this interpolation. The estimated parameters $q = 0.528$, $a = 15.36$, $b = -0.220$ and $c = 2.852$ are inserted in expression (3.30) in order to obtain the estimate $\hat{l}_{\text{all}} = \hat{l}_\lambda(1) = 9.239\text{e-}13$. The middle blue line indicates the estimated probability values $\hat{l}_\lambda(\lambda)$, while the outer red lines indicate the 95% confidence interval. The more accurate estimate $\hat{l}_{\text{all}} = 1.495\text{e-}8$, obtained by the CE method, is marked with a star.

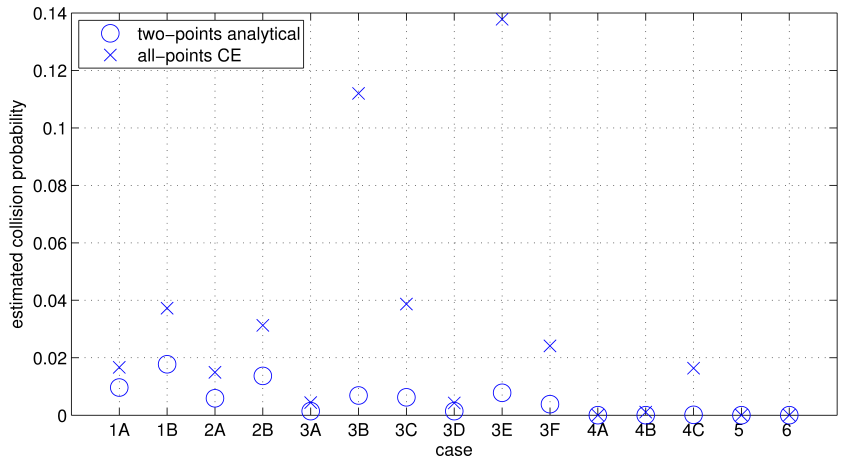

Figure 4.21: Comparison of the two-points probability values $\hat{l}_{\text{two}}^{\text{analytical}}$ and the all-points probability values $\hat{l}_{\text{all}}^{\text{CE}}$.

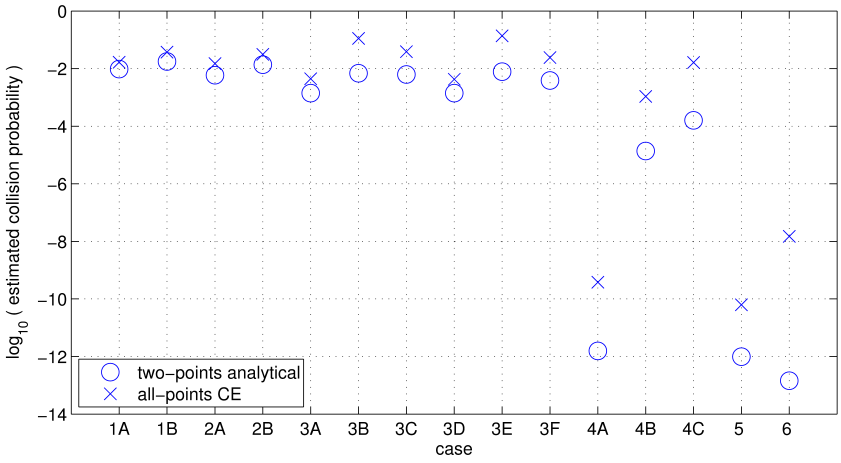

Figure 4.22: Comparison of the 10 base logarithm of the two-points probability values $\hat{l}_{\text{two}}^{\text{analytical}}$ and the all-points probability values $\hat{l}_{\text{all}}^{\text{CE}}$.

4.3 Influence of Well Parameters

We finish this chapter by using the previously obtained numerical results in order to gain some initial insight into how some chosen well geometry parameters might influence the risk of collision.

As we emphasized in the previous section, knowing how a well geometry parameter actually influences the calculations of a collision risk method, is not a trivial task, because of the complexity of these calculations and the variety of variables that makes up the well geometry. Even though we might not be able to make definitive conclusions about the well parameters in general, we can indeed make some comments regarding the results for our particular well pair cases, which could serve as a starting point for future investigations.

Although there is obviously a large amount of additional well parameters that could be investigated, including for instance the inclination angle and the *NEV* coordinates, we have chosen to make observations regarding the following three parameters, by considering the indicated well pair cases.

- The crossing angle (Sec. 4.3.1). We consider the well pairs in Case 1-A, 2-A, 1-B, 2-B, 4-A and 5.
- The well direction in the *NE* plane (Sec. 4.3.2). By this we actually mean the azimuth angle of the wells in the area around the closest points. We consider the well pairs in Case 1-A, 1-B, 2-A, 2-B, 3-B, 3-C, 3-E and 3-F.
- The measured depth in the area around the closest points (Sec. 4.3.3). We consider the well pairs in Case 3-A, 3-B, 3-C, 3-D, 3-E, 3-F, 4-A, 4-B and 4-C.

In the tables presented in these sections, we present the p-values of both hypothesis tests from Table 4.2, the calculated two-points collision probability $\hat{l}_{\text{two}}^{\text{analytical}}$ from Table 4.3 and the estimated all-points collision probability $\hat{l}_{\text{all}}^{\text{CE}}$ from Table 4.7.

In each table we have grouped the well pairs that have about the same geometry, except for the parameter of interest. We will refer to such a collection of well pair cases as a *group*. However, even though we have striven to rule out the influence of other kind of geometry parameters within a group, there might still exist such additional factors. We will comment on this in the forthcoming sections when necessary.

4.3.1 Crossing Angle

The well pairs in Case 1-A, 1-B and 4-A have a crossing angle of 90°, while the well pairs in Case 2-A, 2-B and 5 have a crossing angle of 45°. The collision risk results for these cases are collected in Table 4.9.

In this table we have grouped the well pairs that seem to have the same geometry, except for the crossing angle. This means that in order to make observations regarding the influence of the crossing angle, we should compare Case 1-A against Case 2-A, Case 1-B against Case 2-B and Case 4-A against Case 5.

Table 4.9: The numerical results we use to make observations about the influence of the crossing angle.

Well pair case	Crossing angle	Hypothesis tests: Approximated p-value	Hypothesis tests: More exact p-value	Collision probability: $\hat{l}_{\text{two}}^{\text{analytical}}$	Collision probability: $\hat{l}_{\text{all}}^{\text{CE}}$
1-A	90°NV	3.14e-2	4.10e-2	9.629e-3	1.668e-2
2-A	45°NV	1.85e-2	2.34e-2	5.908e-3	1.492e-2
1-B	90°EV	7.10e-2	0.102	1.776e-2	3.726e-2
2-B	45°EV	5.32e-2	7.41e-2	1.368e-2	3.128e-2
4-A	90°NE	6.88e-10	0.175	1.57e-12	3.806e-10
5	45°NE	8.40e-11	2.48e-2	9.90e-13	6.267e-11

From Table 4.9 we observe that for both hypothesis tests, the p-values are higher in the 90° cases, meaning that the tests are farther away from rejecting H_0 than in the corresponding 45° cases. Also, the collision probability values $\hat{l}_{\text{two}}^{\text{analytical}}$ and $\hat{l}_{\text{all}}^{\text{CE}}$ are higher when the crossing angle is 90°. In sum, all the collision risk methods give the result that there is a higher risk of collision when the crossing angle is 90° than when it is 45° for all these three groups, indicating that the crossing angle is one of the many well geometry parameters that has an influence on the collision risk.

However, in this case there are indeed additional well geometry factors that we have not be able to keep constant within the groups, namely the inclination values and the azimuth values, as given in Table 4.1. Even though it is true that the crossing angle is changed from 90° to 45° within all the three groups, it is also the case that the inclination angle of the offset well is changed from 90° to 45° within the first two groups, while the azimuth angle of the offset well is changed from 270° to 315° within the third group. This might indicate that it is in fact the changes in these angles, and not the crossing angle, that influence the collision risk values. In order to find the correct explanation, more detailed studies that includes more test cases will be needed.

Our suspicion about the influence of additional factors is strengthened by the fact that we could actually expect the collision risk to *increase* when the crossing angle changes from 90° to 45°, since the volume of the coinciding region of two colliding cylinders (wells) is larger when the cylinders are not colliding perpendicularly.

4.3.2 Well Direction in the *NE* Plane

We now consider the direction of the wells in the *NE* plane in the area around the closest points, which is the same as considering the azimuth angles of the two

wells. In Table 4.10 we present the collision risk results for the relevant test cases. Here we have grouped the well pairs that have the same geometry, except for the azimuth angles.

Table 4.10: The numerical results that we use to make observations about the influence of the well direction in the *NE* plane. The azimuth A is given for each well in the area around the closest points. Also, $A = 0^\circ$ is northwards, while $A = 90^\circ$ is eastwards. The azimuth is undefined when a well is vertical.

Well pair	Azimuth A		**Hypothesis tests**		**Collision probability**	
			Approximated	More exact		
case	Ref.	Off.	p-value	p-value	$\hat{l}_{\text{two}}^{\text{analytical}}$	$\hat{l}_{\text{all}}^{\text{CE}}$
1-A	-	0°	3.14e-2	4.10e-2	9.629e-3	1.668e-2
1-B	-	90°	7.10e-2	0.102	1.776e-2	3.726e-2
2-A	-	0°	1.85e-2	2.34e-2	5.908e-3	1.492e-2
2-B	-	90°	5.32e-2	7.41e-2	1.368e-2	3.128e-2
3-A	-	-	9.57e-3	0.132	1.411e-3	4.428e-3
3-D	-	-	9.57e-3	0.132	1.411e-3	4.278e-3
3-B	0°	0°	0.121	0.598	6.878e-3	0.1120
3-E	90°	90°	0.141	0.616	7.795e-3	0.1378
3-C	0°	0°	0.372	0.872	6.223e-3	3.871e-2
3-F	90°	90°	0.452	0.926	3.852e-3	2.414e-2

We observe that for the strictly parallel well segments in Case 3-A and Case 3-D in the third group, all the collision risk values are about the same for both cases. Here, the azimuth angles are undefined in both cases. Hence, their only difference is that the starting positions of the wells in the NE plane are separated in the east direction for Case 3-A and in the north direction for Case 3-D. This difference itself is seen not to have any significant influence on the collision risk for these two cases.

For the other groups, both p-values increases when the azimuth is increased from 0° to 90°, that is when going from the north direction to the east direction. However, for the collision probability values, there is seen to be some inconsistency in the way that the probability increases for all the groups except for the last one, for which the probability decreases.

Summarized, there is seen to be a clear tendency for the hypothesis tests, in the way that their p-values indicates a larger collision risk for the east direction than for the north direction for all the groups. On the other hand, the collision probability estimates show that the collision probability might actually decrease when the azimuth angle increase from 0° to 90°. Nevertheless, there is no doubt

that there is at least some kind of change in the collision probability when the azimuth angle is changed.

4.3.3 Measured Depth

In Table 4.11 we have grouped the well pairs that have the same geometry, except for the measured depth of the wells in the area around the closest points.

Table 4.11: The numerical results that we use make observations about the influence of the measured depth D. The given value is the depth in the area around the closest points. Here the depth is approximately the same for both wells in all the cases.

Well pair case	Depth D (m)	**Hypothesis tests** Approximated p-value	More exact p-value	**Collision probability** $\hat{l}_{\text{two}}^{\text{analytical}}$	$\hat{l}_{\text{all}}^{\text{CE}}$
3-A	1490	9.57e-3	0.132	1.411e-3	4.428e-3
3-B	3290	0.121	0.598	6.878e-3	0.1120
3-C	3890	0.372	0.872	6.223e-3	3.871e-2
3-D	1490	9.57e-3	0.132	1.411e-3	4.278e-3
3-E	3290	0.141	0.616	7.795e-3	0.1378
3-F	3890	0.452	0.926	3.852e-3	2.414e-2
4-A	1610	6.88e-10	0.175	1.57e-12	3.806e-10
4-B	3620	1.28e-3	0.228	1.370e-5	1.080e-3
4-C	5600	2.49e-2	0.349	1.588e-4	1.638e-2

We observe that for all the three groups of well pairs in Table 4.11, the p-values for both hypothesis tests are increasing when the measured depth increases. The same observation applies to the collision probability values in the third group. However, for the first two groups, containing the parallel well pairs, the collision probabilities are the highest at the middle depth level, in the transition between vertical wells and horizontal wells. Although it is difficult to make strong and general conclusions based on these values, the results indicate that the collision probability does *not always* increase when the wells get deeper, at least when it comes to non-straight parallel wells. In this case, it seems that there are other well geometry parameters than the measured depth, like the change of inclination, that provides a larger contribution to the collision probability.

For the p-values, the tendency is much clearer. We observe that for all the three groups in Table 4.11, the p-values for both tests increase with the measured depth. In addition, the p-values for the approximated test seem to increase faster than for the more exact test, which was also illustrated in Figure 4.10 in Section 4.2.1. In order to describe this observation further, we show the pdf's for the two tests

applied on Case 4-A, 4-B and 4-C in Figure 4.23. As before, the p-values are found as the areas under the graphs to the right of $r = \hat{r} = 10.0$ m. Here we easily see that the differences between the p-values for the three cases are relatively larger for the approximated test, as was evident on the $\log_{10}$ plot in Figure 4.10 as well.

This plot also illustrates the tendency that the p-value of the approximated test actually gets closer and closer to the p-value of the more exact test when the measured depth increases, in the way that the width of the pdf graph (which illustrates the variance of the pdf) seems to increase faster for the approximated test. This was also illustrated in Figure 4.10 in Section 4.2.1.

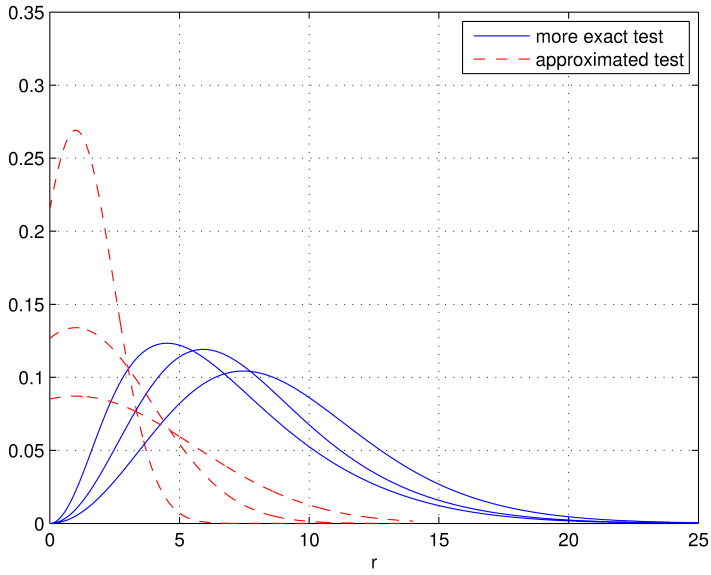

Figure 4.23: The pdf's of the shortest distance r under H_0 for the two hypothesis tests as performed on the three subcases of Case 4. When we consider the rightmost parts of the graphs, the graphs to the left belong to Case 4-A, the middle graphs belong to Case 4-B, and the graphs to the right belong to Case 4-C.

Chapter 5

Closing Remarks

From the results obtained in the previous chapters, especially from the numerical results in Chapter 4, we are able to make some concluding remarks. This includes some topics that might be interesting to investigate in a possible future work on this subject.

5.1 Conclusions

In this thesis we have explained and explored several methods for addressing the risk of collision between petroleum wells.

In Chapter 2 we first explained how to obtain the approximated *NEV* coordinates of points in a petroleum well based on magnetic MWD measurements in addition to some reference values. We also described our new contribution of obtaining the joint statistical distribution of the *NEV* coordinates of all the survey points of two wells. This is used in the methods that estimate the collision probability for the whole wells, rather than only the two closest points.

In Chapter 3 we thoroughly described the various collision risk methods, and in Chapter 4 we applied the methods on a number of well pair test cases in order to obtain numerical results.

The current industry practice regarding collision risk is to perform a hypothesis test based solely on the *NEV* positions of the two closest points. The industry test makes a normal approximation of the statistical distribution of the shortest distance between the points. We have proposed a new and more exact hypothesis test in which there is no need for this approximation. The resulting p-values are found to be different for the two tests. In fact, the more exact test is more conservative than the approximated test for all the test cases, meaning that the null hypothesis is more often rejected than for the approximated test. This implies that there are situations where the approximated test allows further drilling, while the more exact test claims that the probability of obtaining the measured positions, given an expected well collision, is higher than the chosen significance level. Both the p-values themselves and the difference between them depends on the well geometry

of interest.

In addition to the p-values of the hypothesis tests, we have striven to calculate an alternative parameter that gives information about the risk of well collision, namely the probability of such a collision. By deciding a maximum value for this probability, we have a new kind of criterion for whether or not to continue the drilling.

We have estimated the probability of collision in both the two-points situation, where we consider the two closest points, as well as in the all-points situation, where we consider the whole wells.

We have obtained the probability of collision between the two closest points in three ways, using an analytical method, a crude Monte Carlo (CMC) simulation method and a cross-entropy (CE) method. The analytical method is by far the fastest method of these three. All three methods give similar probability estimates, which give support to all the methods. This means that both the CMC method and the CE method tend to give accurate results. However, for well pairs where the probability of collision is very small, the CMC method is not able to calculate an estimate unless the sample size is inconveniently large, while the CE method still give accurate results in such rare-event situations.

When we compare the probability of collision between the two closest points with the results of the hypothesis tests, we observe that the probability of collision is smaller than the p-values of both tests for all our test cases.

Calculating the collision probability analytically when considering the whole wells is outside the scope of this thesis. However, we have estimated this probability in three different ways, using a CMC simulation method, a CE method and the enhanced Monte Carlo (EMC) method. The CE method is more accurate (having a smaller relative error) than the CMC method for the test cases where both methods do give results. The CE method and the EMC method are able to obtain an estimate of the probability for the rare-event test cases where the CMC method fails, exactly as for the two-points methods. The CE method is much more accurate than the EMC method in such a situation, even though the EMC method still gives probability estimates that at least support the CE values to some degree for the cases where the CMC method is useless.

The all-points methods result in larger collision probability estimates than the two-points methods, which is as expected, since these methods take into account that other parts of the well segments around the closest points might collide even when the closest points do not. However, the magnitude of the difference is found to be highly dependent on the well geometry, in the way that for some test cases, the two-points methods and the all-points methods give very similar probability values, while for other cases, the difference is seen to be significant.

At the end of Chapter 4 we made some observations about the influence of some well geometry parameters on the collision risk in the well pair cases, which could serve as a starting point for future investigations. Here we observed that for our test cases, the collision probability does not necessarily increases when the measured depth increases, at least not when it comes to non-straight parallel wells. However, the p-values of the hypothesis tests indeed increased with the measured

depth for the test cases.

We summarize the most important conclusions in the following.

- The approximated hypothesis test is less conservative than the more exact test, which means that it underestimates the collision risk compared to our new and improved test. Implementation of the more exact test is unproblematic, and the method is not time-consuming.
- When it comes to the alternative collision risk measure that is the probability of well collision, the analytical method is much faster than both the CMC method and the CE method in the two-points situation. In the all-points situation, the CE method is superior to the CMC method and the EMC method, both in terms of accuracy and efficiency.
- The magnitude of the difference between the all-points results and the two-points results, in the way that the two-points methods underestimate the collision probability compared to the all-points methods, is found to be depending on the well geometry. For some well geometries the difference is very large, which might suggest a further investigation of this phenomenon. In this thesis we have laid the foundation when it comes to the collision probability.

5.2 Further Work

Finally, there are several aspects of the work in this thesis that could be investigated further. We now list some examples.

- The current industry practice is to perform a hypothesis test based on the two closest points in the wells. In this thesis, we have dealt with the case of calculating the probability of well collision when considering the whole wells instead of only considering the two closest points. Is it possible to perform a hypothesis test that is based on all the points in a similar way?
- More detailed comparisons between the hypothesis tests and the collision probability can be made. Is there a more general relationship to be found?
- We have made some initial observations considering the influence of some well geometry parameters. Is it possible to get a more rigid understanding of the influence of such parameters on the collision risk values?
- What is the effect of the uncertainty of the NEV coordinates of a well on the collision risk? And what is the effect of the correlation values between the positions of two wells?
- What would be the consequences if the MWD error sources given in Chapter 2 are not normal distributed, as we have assumed in this thesis?
- Is it possible to extend the methods in this thesis to apply to other input data, like gyroscopic measurements and other types of MWD directional data?

Bibliography

A. Brooks. A New Look at Wellbore Collision Probability. In *SPE Annual Technical Conference and Exhibition*, 2008.

A. Brooks and H. Wilson. An Improved Method for Computing Wellbore Position Uncertainty and its Application to Collision and Target Intersection Probability Analysis. In *European Petroleum Conference*, 1996.

COMPASS. Landmark of Halliburton, 2011.

P.T. de Boer, D.P. Kroese, S. Mannor, and R.Y. Rubinstein. A Tutorial on the Cross-Entropy Method. *Annals of operations research*, 134, 2002.

T. Gjerde. *A Heavy Tailed Statistical Model Applied in Anti-Collision Calculations for Petroleum Wells.* PhD thesis, MSc thesis, Norwegian University of Science and Technoglogy, 2008.

T. Gjerde, J. Eidsvik, E. Nyrnes, and B.T. Bruun. Positioning and position error of petroleum wells. *Journal of Geodetic Science*, 1(2):158–169, 2011.

J. Goodman. Monte Carlo methods. Chapter 9. URL `http://www.cs.nyu.edu/courses/fall06/G22.2112-001/MonteCarlo.pdf`.

J. Ødegård Hansen et al., editors. *Facts, The Norwegian Petroleum Sector 2011.* Ministry of Petroleum and Energy, 2011.

Collision Avoidance Sub-Committee ISCWSA. Collision Avoidance Calculations - Current Common Practice. Minutes of Meeting, September 2011.

S. Kotz, NL Johnson, and DW Boyd. Series Representations of Distributions of Quadratic Forms in Normal Variables. I. Central Case. *The annals of mathematical statistics*, 38(3):823–837, 1967.

MATLAB. *Version 7.11.0 (R2010b).* The MathWorks Inc., 2010.

A. Naess, BJ Leira, and O. Batsevych. System Reliability Analysis by Enhanced Monte Carlo Simulation. *Structural safety*, 31(5):349–355, 2009.

S. Ripley. *Stochastic Simulation.* Wiley, 1987.

H. Ruben. Probability Content of Regions Under Spherical Normal Distributions, IV: The Distribution of Homogeneous and Non-Homogeneous Quadratic Functions of Normal Variables. *The Annals of Mathematical Statistics*, 33(2):542–570, 1962.

R.Y. Rubinstein and D.P. Kroese. *The Cross-Entropy Method: A Unified Approach to Combinatorial Optimization, Monte-Carlo Simulation, and Machine Learning.* Springer, 2004.

SJ Sawaryn and JL Thorogood. A Compendium of Directional Calculations Based on the Minimum Curvature Method. In *SPE Annual Technical Conference and Exhibition*, 2003.

J. Sheil and I. O'Muircheartaigh. Algorithm as 106: The Distribution of Non-Negative Quadratic Forms in Normal Variables. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 26(1):92–98, 1977.

J.L. Thorogood, T.W. Hogg, and H.S. Williamson. Application of risk analysis methods to subsurface well collisions. *SPE drilling engineering*, 6(4):299–304, 1991.

Y. Tsao et al. Well Collision Avoidance. U.S. patent 5,901,795, May 1999.

JE Walstrom, AA Brown, and RP Harvey. An Analysis of Uncertainty in Directional Surveying. *Journal of Petroleum Technology*, 21(4):515–523, 1969.

H.S. Williamson. Accuracy Prediction for Directional Measurement While Drilling. *SPE Drilling & Completion*, 15(4):221–233, 2000.

CJM Wolff and JP de Wardt. Borehole Position Uncertainty – Analysis of Measuring Methods and Derivation of Systematic Error Model. *Journal of Petroleum Technology*, 33(12):2338–2350, 1981.

Appendix A

Error Propagation Theory

Consider a d-variate normal distributed vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^t$, where superscript t denotes that a matrix is *transposed*. When $\mathbf{x}$ has mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$, we write

$$\mathbf{x} \sim N_d(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x).$$

Now consider a vector $\mathbf{f} = [f_1, f_2, \ldots, f_k]^t$ containing k linear combinations of the variables in $\mathbf{x}$,

$$\mathbf{f}(\mathbf{x}) = \mathbf{a} + \mathbf{B}\mathbf{x}, \tag{A.1}$$

where $\mathbf{a}$ is a vector of constants and $\mathbf{B}$ is a matrix of constants,

$$\mathbf{a} = [a_1, a_2 \ldots a_k]^t, \qquad \mathbf{B} = \begin{bmatrix} B_{1,1} & B_{1,2} & \ldots & B_{1,d} \\ B_{2,1} & B_{2,2} & \ldots & B_{2,d} \\ \vdots & \vdots & \ddots & \vdots \\ B_{k,1} & B_{k,2} & \ldots & B_{k,d} \end{bmatrix}.$$

The uncertainty (or error) of the linear combinations $\mathbf{f}$ will be effected by the uncertainty of $\mathbf{x}$. In other words, the uncertainty of $\mathbf{x}$, described by $\boldsymbol{\Sigma}_x$, propagates to $\mathbf{f}$. By using the properties of the normal distribution, $\mathbf{f}(\mathbf{x})$ can be shown to be distributed as

$$\mathbf{f}(\mathbf{x}) \sim N_k(\mathbf{a} + \mathbf{B}\boldsymbol{\mu}_x, \mathbf{B}\boldsymbol{\Sigma}_x\mathbf{B}^t). \tag{A.2}$$

The distribution of k *nonlinear* functions $\mathbf{f} = [f_1, f_2, \ldots, f_k]^t$ of $\mathbf{x}$ can be approximated by linearizing $\mathbf{f}$ with the use of a first order Taylor series expansion centered at $\boldsymbol{\mu}_x$. In matrix notation, this is given by

$$f(\mathbf{x}) \approx \mathbf{f}(\boldsymbol{\mu}_x) + \mathbf{J}(\mathbf{x} - \boldsymbol{\mu}_x), \tag{A.3}$$

where

$$\mathbf{J} = \begin{bmatrix} \frac{\partial f_1(\boldsymbol{\mu}_x)}{\partial x_1} & \frac{\partial f_1(\boldsymbol{\mu}_x)}{\partial x_2} & \cdots & \frac{\partial f_1(\boldsymbol{\mu}_x)}{\partial x_d} \\ \frac{\partial f_2(\boldsymbol{\mu}_x)}{\partial x_1} & \frac{\partial f_2(\boldsymbol{\mu}_x)}{\partial x_2} & \cdots & \frac{\partial f_2(\boldsymbol{\mu}_x)}{\partial x_d} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_k(\boldsymbol{\mu}_x)}{\partial x_1} & \frac{\partial f_k(\boldsymbol{\mu}_x)}{\partial x_2} & \cdots & \frac{\partial f_k(\boldsymbol{\mu}_x)}{\partial x_d} \end{bmatrix} \tag{A.4}$$

is the *Jacobian matrix* of $\mathbf{f}(\mathbf{x})$. An element $J(i,j)$ of $\mathbf{J}$ is the partial derivative of $f_i(\mathbf{x})$ with respect to x_j, evaluated in $\mathbf{x} = \boldsymbol{\mu}_x$.

The expressions $\mathbf{f}(\boldsymbol{\mu}_x) - \mathbf{J}\boldsymbol{\mu}_x$ and $\mathbf{J}$ in (A.3) corresponds to $\mathbf{a}$ and $\mathbf{B}$ in (A.1), respectively. From (A.2), the distribution of $\mathbf{f}(\mathbf{x})$ is therefore approximately given by

$$\mathbf{f}(\mathbf{x}) \sim N_k\left(\mathbf{f}(\boldsymbol{\mu}_x), \boldsymbol{\Sigma}_f\right), \tag{A.5}$$

with covariance matrix

$$\boldsymbol{\Sigma}_f = \mathbf{J}\boldsymbol{\Sigma}_x\mathbf{J}^t.$$

Appendix B

Description of the Error Sources

We will now describe the 19 error sources for measurment uncertainty as given in Table 2.1. For convenience, we repeat this information in Table B.1.

Table B.1: The error sources $\varepsilon_i, i = 1, \ldots, 19$, used in this thesis. The propagation mode of an error is either random (R), systematic (S) or global (G).

Error number, i	Description	Standard deviation, σ_i	Propagation mode
1–3	G_x, G_y, G_z bias	0.0039 m/s^2	S
4–6	G_x, G_y, G_z scale	0.0005	S
7–9	B_x, B_y, B_z bias	70 nT	S
10–12	B_x, B_y, B_z scale	0.0016	S
13	Depth reference	0.35 m	R
14	Depth scale factor	$6 \cdot 10^{-4}$	S
15	Depth stretch type	$2.5 \cdot 10^{-7}$ m^{-1}	G
16	Declination	0.36°	G
17	Declination B dependent	5000° nT	G
18	Sag	0.08°	S
19	Axial magnetism in string	150 nT	S

The following descriptions are based on the work of Williamson (2000) and Gjerde (2008).

The errors $\varepsilon_1, \varepsilon_2, \varepsilon_3$ are the *gravity bias errors*, implying that they give an additive noise to the the three gravity measurements G_x, G_y and G_z in each survey series. The errors $\varepsilon_4, \varepsilon_5, \varepsilon_6$ are the *gravity scale errors*, which means that they give a contributions proportional to the expected values of the measurements.

As an example, consider the measurable value G_x of the gravity component in the X direction. This value is assumed to be the sum of the unknown expected value μ_{G_x} and the contribution from two measurement errors; a bias error ε_1 and

a scale error ε_4. We have assumed that ε_1 and ε_4 are independent, and

$$\begin{aligned} \varepsilon_1 &\sim N\left(0,\ \sigma_1^2\right), \\ \varepsilon_4 &\sim N\left(0,\ \sigma_4^2\right). \end{aligned}$$

This means that in a well point k, a measured value is distributed as

$$G_x = \mu_{G_x} + \varepsilon_1 + \mu_{G_x}\varepsilon_4 \sim N\left(\mu_{G_x},\ \sigma_1^2 + \mu_{G_x}^2\sigma_4^2\right).$$

The expected value μ_{G_x} can be estimated by a sampled (measured) value of G_x, denoted g_x. Then,

$$G_x = g_x + \varepsilon_1 + g_x\varepsilon_4 \sim N\left(g_x,\ \sigma_1^2 + g_x^2\sigma_4^2\right). \tag{B.1}$$

The errors $\varepsilon_7, \varepsilon_8, \varepsilon_9$ are the *magnetic bias errors*, while the errors $\varepsilon_{10}, \varepsilon_{11}, \varepsilon_{12}$ are the *magnetic scale errors*. These error sources work on the magnetic measurements B_x, B_y and B_z in the exact same way as the first six errors did on the gravity measurements.

The errors $\varepsilon_{13}, \varepsilon_{14}, \varepsilon_{15}$ are related to the measurement D_k of the measured depth in well point k. They are the *depth reference*, *depth scale factor* and *depth stretch type* error sources. For a measurable value D, their contribution is given by

$$D = \mu_D + \cdots + \varepsilon_{13} + \mu_D\varepsilon_{14} + \mu_D\mu_V\varepsilon_{15},$$

where μ_V is the expected vertical depth coordinate of the well point. The dots take the place of other error sources that also influence the measured depth.

The errors ε_{16} and ε_{17} are *declination errors*, meaning they are related to the declination angle δ, as defined in Section 2.2, for instance in Figure 2.5. The error ε_{16} is the *constant declination error*, while the error ε_{17} is the *B-dependent declination error*.[1] These errors are assumed to influence a measured azimuth value as

$$A = \mu_A + \cdots + \varepsilon_{16} + \frac{1}{B\cos\theta}\varepsilon_{17},$$

where the dots take the place of other error sources that also influence a measured azimuth value.

The error ε_{18} is the *sag error*, which is caused by misalignment of the sensors in the MWD tool. When the inclination angle is large, the weight might cause a sag in the tool itself or its sensors. This error is assumed to influence a measured inclination angle as

$$I = \mu_I + \cdots + \sin\mu_I\varepsilon_{18}.$$

The error ε_{19} is the *axial magnetism error in the drill string*, which is caused by other equipment around the MWD tool, which affects the measured magnetic field. This error is assumed to influence a measured azimuth angle as

$$A = \mu_A + \cdots + \frac{\sin I \sin A_m}{B\cos\theta}\varepsilon_{19}.$$

[1] More precicely, the error is dependent on the horizontal projection of the magnetic field, which is denoted by B_H in Williamson (2000).

We should emphasize that these 19 error sources make up an error model that we simply assume to be correct. More error sources could be included, as described by Williamson (2000).

Appendix C

The Weighting Functions

Here, we will find the weighting functions that makes up the Jacobian matrices $\mathbf{J}_{DIA}$ and $\mathbf{J}_{\Delta p}$ in Chapter 2. These matrices are used to find the covariance matrix in the joint distribution of the NEV values, by using the equation (2.25), that is

$$\mathbf{\Sigma}_{NEV} = \mathbf{J}_p \mathbf{J}_{\Delta p} \mathbf{J}_{DIA} \mathbf{\Sigma}_\varepsilon \mathbf{J}_{DIA}^t \mathbf{J}_{\Delta p}^t \mathbf{J}_p^t.$$

C.1 Weighting Functions for the *DIA* Values

The Jacobian matrix $\mathbf{J}_{DIA}$ is given in (2.15). The weighting function $\mathbf{w}_{i,l}^k$ for measurement station number k (which belongs to survey number l) is given by

$$\mathbf{w}_{i,l}^k = \begin{bmatrix} \dfrac{\partial D_k}{\partial \varepsilon_{i,l}} & \dfrac{\partial I_k}{\partial \varepsilon_{i,l}} & \dfrac{\partial A_k}{\partial \varepsilon_{i,l}} \end{bmatrix}^t, \tag{C.1}$$

where all the derivatives are evaluated at the expectation of the variables of the differentiated functions, corresponding to the matrix elements in (A.4).

The *DIA* values depend on some of the error sources in Table 2.1 (equaling Table B.1) only indirectly, that is via some measured value. Therefore, in order to find the elements in $\mathbf{w}_{i,l}^k$, one might have to use the chain rule for derivatives. For instance, for the error sources $i \in \{1, 4\}$, when we use the expression of the measured value G_x in (B.1), we get for the middle element in (C.1) that

$$\frac{\partial I_k}{\partial \varepsilon_{1,l}} = \frac{\partial I_k}{\partial G_{x,k}} \frac{\partial G_{x,k}}{\partial \varepsilon_{1,l}} = \frac{\partial I_k}{\partial G_{x,k}}, \tag{C.2}$$

$$\frac{\partial I_k}{\partial \varepsilon_{4,l}} = \frac{\partial I_k}{\partial G_{x,k}} \frac{\partial G_{x,k}}{\partial \varepsilon_{4,l}} = \frac{\partial I_k}{\partial G_{x,k}} g_{x,k}. \tag{C.3}$$

All the derivatives are again evaluated at the expectation of the variables of the differentiated functions.

Now follow the 19 weighting functions for the *DIA* values, one for each error source in Table B.1.

$$
\begin{aligned}
\mathbf{w}_{1,l}^{k} &= \frac{1}{G}\begin{bmatrix} 0 \\ -\cos I_k \sin\tau_k \\ (\cos I_k \sin A_m \sin\tau_k - \cos A_m \cos\tau_k)\tan\theta + \cot I_k \cos\tau_k \end{bmatrix}, \\
\mathbf{w}_{2,l}^{k} &= \frac{1}{G}\begin{bmatrix} 0 \\ -\cos I_k \sin\tau_k \\ (\cos I_k \sin A_m \cos\tau_k - \cos A_m \sin\tau_k)\tan\theta - \cot I_k \sin\tau_k \end{bmatrix}, \\
\mathbf{w}_{3,l}^{k} &= \frac{1}{G}\begin{bmatrix} 0 \\ -\sin I_k \\ \tan\theta \sin I_k \sin A_{m,k} \end{bmatrix}, \\
\mathbf{w}_{4,l}^{k} &= \begin{bmatrix} 0 \\ \sin I_k \cos I_k \sin^2\tau_k \\ -\left(\tan\theta \sin I_k \left(\cos I_k \sin A_{m,k} \sin\tau_k - \cos A_{m,k}\cos\tau_k\right) + \cos I_k \cos\tau_k\right)\sin\tau_k \end{bmatrix}, \\
\mathbf{w}_{5,l}^{k} &= \begin{bmatrix} 0 \\ \sin I_k \cos I_k \cos^2\tau_k \\ -\left(\tan\theta \sin I_k \left(\cos I_k \sin A_{m,k} \cos\tau_k + \cos A_{m,k}\sin\tau_k\right) - \cos I_k \sin\tau_k\right)\cos\tau_k \end{bmatrix}, \\
\mathbf{w}_{6,l}^{k} &= \begin{bmatrix} 0 \\ -\sin I_k \cos I_k \\ \tan\theta \sin I_k \cos I_k \sin A_{m,k} \end{bmatrix}, \\
\mathbf{w}_{7,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ \frac{1}{B\cos\theta}\left(\cos A_{m,k}\cos\tau_k - \cos I_k \sin A_{m,k}\sin\tau_k\right) \end{bmatrix}, \\
\mathbf{w}_{8,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{B\cos\theta}\left(\cos A_{m,k}\sin\tau_k + \cos I_k \sin A_{m,k}\cos\tau_k\right) \end{bmatrix}, \\
\mathbf{w}_{9,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ -\frac{1}{B\cos\theta}\left(\sin I_k \sin A_{m,k}\right) \end{bmatrix}, \\
\mathbf{w}_{10,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ \left(\cos I_k \cos A_{m,k}\sin\tau_k - \tan\theta \sin I_k \sin\tau_k + \sin A_{m,k}\cos\tau_k\right) \\ \cdot\left(\cos A_{m,k}\cos\tau_k - \cos I_k \sin A_{m,k}\sin\tau_k\right) \end{bmatrix}, \\
\mathbf{w}_{11,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ -\left(\cos I_k \cos A_{m,k}\cos\tau_k - \tan\theta \sin I_k \cos\tau_k - \sin A_{m,k}\sin\tau_k\right) \\ \cdot\left(\cos A_{m,k}\sin\tau_k + \cos I_k \sin A_{m,k}\cos\tau_k\right) \end{bmatrix}, \\
\mathbf{w}_{12,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ -\left(\sin I_k \cos A_{m,k} + \tan\theta \cos I_k\right)\sin I_k \sin A_{m,k} \end{bmatrix},
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{w}_{13,l}^{k} &= \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \\
\mathbf{w}_{14,l}^{k} &= \begin{bmatrix} D_k \\ 0 \\ 0 \end{bmatrix}, \\
\mathbf{w}_{15,l}^{k} &= \begin{bmatrix} D_k \cdot V_k \\ 0 \\ 0 \end{bmatrix}, \\
\mathbf{w}_{16,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \\
\mathbf{w}_{17,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ \frac{1}{B\cos\theta} \end{bmatrix}, \\
\mathbf{w}_{18,l}^{k} &= \begin{bmatrix} 0 \\ \sin I_k \\ 0 \end{bmatrix}, \\
\mathbf{w}_{19,l}^{k} &= \begin{bmatrix} 0 \\ 0 \\ \frac{\sin I_k \sin A_{m,k}}{B\cos\theta} \end{bmatrix}.
\end{aligned}
$$

C.2 Weighting Functions for the *NEV* Positions

The Jacobian matrix $\mathbf{J}_{\Delta p}$ is given in (2.21). Its elements are the weighting functions given by

$$
\mathbf{R}_{j,k} = \begin{bmatrix} \frac{\partial \Delta \mathbf{p}_j}{\partial D_k} & \frac{\partial \Delta \mathbf{p}_j}{\partial I_k} & \frac{\partial \Delta \mathbf{p}_j}{\partial A_k} \end{bmatrix}.
$$

The elements in $\mathbf{R}_{j,k}$ are easily obtained from (2.20). They are given by

$$\begin{aligned}
\frac{\partial \Delta \mathbf{p}_k}{\partial D_k} &= \frac{1}{2}\begin{bmatrix} \sin I_k \cos A_k + \sin I_{k-1} \cos A_{k-1} \\ \sin I_k \sin A_k + \sin I_{k-1} \sin A_{k-1} \\ \cos I_k + \cos I_{k-1} \end{bmatrix}, \\
\frac{\partial \Delta \mathbf{p}_k}{\partial D_{k-1}} &= \frac{1}{2}\begin{bmatrix} -\sin I_k \cos A_k - \sin I_{k-1} \cos A_{k-1} \\ -\sin I_k \sin A_k - \sin I_{k-1} \sin A_{k-1} \\ -\cos I_k - \cos I_{k-1} \end{bmatrix}, \\
\frac{\partial \Delta \mathbf{p}_k}{\partial I_k} &= \frac{D_k - D_{k-1}}{2}\begin{bmatrix} \cos I_k \cos A_k \\ \sin I_k \sin A_k \\ -\sin I_k \end{bmatrix}, \\
\frac{\partial \Delta \mathbf{p}_k}{\partial I_{k-1}} &= \frac{D_k - D_{k-1}}{2}\begin{bmatrix} \cos I_{k-1} \cos A_{k-1} \\ \sin I_{k-1} \sin A_{k-1} \\ -\sin I_{k-1} \end{bmatrix}, \\
\frac{\partial \Delta \mathbf{p}_k}{\partial A_k} &= \frac{D_k - D_{k-1}}{2}\begin{bmatrix} -\sin I_k \sin A_k \\ \sin I_k \cos A_k \\ 0 \end{bmatrix}, \\
\frac{\partial \Delta \mathbf{p}_k}{\partial A_{k-1}} &= \frac{D_k - D_{k-1}}{2}\begin{bmatrix} -\sin I_{k-1} \sin A_{k-1} \\ \sin I_{k-1} \cos A_{k-1} \\ 0 \end{bmatrix},
\end{aligned}$$

where the derivatives are evaluated at the expectation of the variables of the differentiated functions.

C.3 Reference Values

In the calculations of some the previously described weighting functions, there is need for reference values for some of the other values discussed so far. A predicted gravity field from Gjerde (2008) is given by

$$G = \begin{cases} 978030 + 5186 \sin^2 \phi + 0.14 D_V & \text{(offshore well)}, \\ 978030 + 5186 \sin^2 \phi + 0.10 D_V - 0.31 h_0 & \text{(onshore well)}. \end{cases} \tag{C.4}$$

In (C.4), the gravity G is given in units mGal, where one unit equals 10^{-5} m/s^2. Also, ϕ is the latitude position measured in degrees, D_V is the vertical depth measured in meters, and h_0 is the height of the (onshore) installation above the mean sea level.

In the case of the magnetic field, the reference values that are typical for the Norwegian Sea (Gjerde, 2008) are given by

$$B = 50000 \text{ nT}, \tag{C.5}$$
$$\delta = 1, \tag{C.6}$$
$$\theta = 75. \tag{C.7}$$

Some of the weighting functions depend on the toolface angle τ_k, which should ideally be given at every survey point. However, in the case of *planning* a well path, the (future) rotation of the well bore is unknown. In this case, we will simulate the toolface values uniformly from $[0, 2\pi)$, as legitimized by Williamson (2000), even though they also claim that this solution tend to give somewhat optimistic results.

Appendix D

Pdf of Vectors With High Correlation

For some well data, the procedure described in Chapter 2 may result in covariance matrices that are either singular or nearly so.

Consider the d-variate normal distributed vector $\mathbf{x} = [x_1, x_2, \ldots, x_d]^t$, having mean vector $\boldsymbol{\mu}_x$ and covariance matrix $\boldsymbol{\Sigma}_x$, that is

$$\mathbf{x} \sim N_d(\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x).$$

Singularity of $\boldsymbol{\Sigma}_x$ is simply explained by high correlation values between two or more of the variables in the stochastic vector $\mathbf{x}$. In other words, for the corresponding *correlation matrix* $\mathbf{P}$ having elements

$$P(i,j) = \frac{\Sigma_x(i,j)}{\sqrt{\Sigma_x(i,i)\Sigma_x(j,j)}},$$

one or more of the non-diagonal elements $P(i,j),\ i \neq j$, are very close to 1.

The pdf of $\mathbf{x}$ is given by

$$f\left(\mathbf{x}|\boldsymbol{\mu}_x, \boldsymbol{\Sigma}_x\right) = \frac{1}{(2\pi)^{d/2}|\boldsymbol{\Sigma}_x|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu}_x)^t \boldsymbol{\Sigma}_x^{-1}(\mathbf{x}-\boldsymbol{\mu}_x)\right). \tag{D.1}$$

When the covariance matrix $\boldsymbol{\Sigma}_x$ is close to singular, obtaining the pdf of the statistical distribution is a problematic task, since the function involves the inverse $\boldsymbol{\Sigma}_x^{-1}$.

We approach this problem by realizing that a stochastic vector $\mathbf{x}$, that contains highly correlated values, behaves stochastically in the same way as the modified vector $\tilde{\mathbf{x}}$, in which we have ignored the *uninformative* elements. By this we mean that if two or more values are highly correlated, we only consider one of them to be *informative*, while the rest are uninformative. When we actually need the numeric values of the ignored elements, we consider these values to be the equal to the value of the corresponding informative variable.

When we want to calculate the pdf in expression (D.1) of the vector $\mathbf{x}$, we ignore the elements of $\mathbf{x}$ and $\boldsymbol{\mu}_x$, and the rows and columns of $\boldsymbol{\Sigma}_x$, that corresponds to the uninformative elements of $\mathbf{x}$.

Now, we need a limit value for the correlation value, in order to decide when to consider a variable to be uninformative. This limit should be as close to 1 as possible, in order to get an accurate approximation.

In a *singular value decomposition (SVD)* of $\boldsymbol{\Sigma}_x$ we find unitary matrices $\mathbf{U}$ and $\mathbf{V}$, and a diagonal matrix $\mathbf{S}$, so that

$$\boldsymbol{\Sigma}_x = \mathbf{USV}^*,$$

where $\mathbf{V}^*$ is the conjugate transpose of $\mathbf{V}$. The diagonal elements of $\mathbf{S}$ are known as the *singular values* of $\boldsymbol{\Sigma}_x$.

The *condition number cond* of the covariance matrix $\boldsymbol{\Sigma}_x$ is defined as the ratio of the largest singular value of $\boldsymbol{\Sigma}_x$ to the smallest. The *inverse condition number* is defined as $rcond = (cond)^{-1}$.

In MATLAB, the matrix $\boldsymbol{\Sigma}_x$ is considered to be *poorly conditioned* when $rcond < eps$, where *eps* is a measure of the precision of a floating point number. In our version of MATLAB, we have $eps = 2.2204 \cdot 10^{-16}$. The program then gives a warning message about singularity when we try to invert a matrix that fulfills the condition $rcond < eps$. Motivated by this, we choose to ignore the most uninformative elements until we reach $rcond \geq eps$.

Finally, we only consider the remaining informative variables in calculating the pdf in expression (D.1).

Appendix E

An Approximated 3D Closest Approach Method

Well planning programs like COMPASS (2011) finds the positions of the two closest points in two wells and the distance between them by use of a 3D closest approach method. Here, we have developed our own approximated method for finding the closest points, in order to be able to find the shortest distance using MATLAB (2010).

The first approximation that is made in our approximated 3D closest approach method is shown in Figure E.1. To the left we see the true shortest distance $\hat{r}$ between two (three dimensional) wells. As described by the right-hand side of the figure, an approximation $\tilde{r}$ of the shortest distance r is given by the minimum of the distances between a finite number of points in both wells. That is, if we let $\mathbf{p}_i$ and $\mathbf{q}_j$ represent the *NEV* position vectors for the points $i \in \{1, \ldots, H_1\}$ and $j \in \{1, \ldots, H_2\}$, with a total of H_1 and H_2 points in the two wells respectively, then

$$\tilde{r} = \min_{i,j} \|\mathbf{p}_i - \mathbf{q}_j\|,$$

where $\|\cdot\|$ is the Euclidean norm. The position vectors $\mathbf{p}_i$ and $\mathbf{q}_j$ are the positions of survey points, as calculated in expression (2.18), or found by the use of the interpolation formula (2.26) in the minimum curvature method. In order to get a good approximation $\tilde{r}$, we need the *NEV* positions of a lot of intermediate points as well as the survey points.

By now we have found the two closest points when we consider a finite number of survey points and interpolated intermediate points. However, since the two closest points might in fact happen to be *between* a pair of such points, we find a more accurate estimate by approximating straight lines between the points (survey points and intermediate points) in each well. There are known geometrical methods for finding the shortest distance between such line segments, which we will apply in an area around the previously approximated closest points in Figure E.1, in order to find an even better approximation of the positions of the two closest points in

the wells.

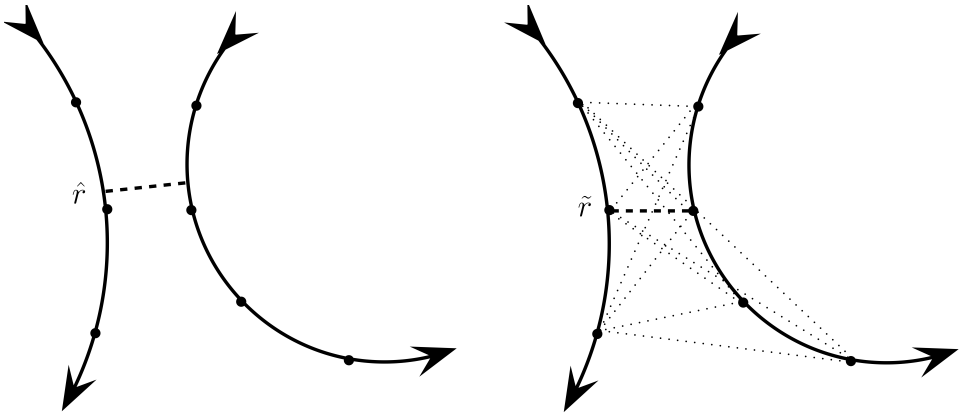

Figure E.1: To the left we have two wells (or parts of wells) that are close to each other. Known positions (survey stations and interpolated positions) are shown as dots, while the shortest distance $\hat{r}$ between the wells is shown as a dashed line. The same two well parts are depicted to the right. Now, each of the dotted lines shows the distance between a pair of known positions, and the approximated overall shortest distance $\tilde{r}$ (dashed line) is given by the minimum of all these distances.

Appendix F

The Separation Factor in the Hypothesis Test

In the hypothesis test in Section 3.1.1, the rejection rule is to reject H_0 if we find that $\hat{r} > r^*$, where $\hat{r}$ is an estimate of r and r^* is an unknown decision limit. We will derive r^* as follows, where we begin with the left-hand side of (3.3).

$$\begin{aligned} P\left(\text{Reject } H_0 \mid H_0 \text{ is true}\right) &= P\left(\hat{r} > r^* \mid r \leq r_1 + r_2\right) \\ &= P\left(\frac{\hat{r} - \mu_{\hat{r}}}{\sigma_{\hat{r}}} > \frac{r^* - \mu_{\hat{r}}}{\sigma_{\hat{r}}} \;\middle|\; r = r_1 + r_2\right) \\ &= P\left(Z > \frac{r^* - (r_1 + r_2)}{\sigma_{\hat{r}}}\right) \\ &= \alpha. \end{aligned}$$

In the first equation we express the words with the defined symbols. In the second equation we subtract both sides of the first inequality by the expectation $\mu_{\hat{r}}$ of $\hat{r}$ and divide by its standard deviation $\sigma_{\hat{r}}$. For the conditioning expression to the right, we recognize that an equality will cover the worst case, that is when it is most difficult to detect a collision. In the third equation we use the assumption that $\hat{r}$ is normal distributed, giving that $Z = (\hat{r} - \mu_{\hat{r}})/\sigma_{\hat{r}}$ is standard normal distributed. We also use that $\mu_{\hat{r}} = r = r_1 + r_2$, which means that there is no longer need for the conditioning expression. In the last line we repeat (3.3). The last two lines implies that

$$\frac{r^* - (r_1 + r_2)}{\sigma_{\hat{r}}} = k_\alpha,$$

where k_α is the $100(1-\alpha)$th percentile of the standard normal distribution. We then obtain

$$r^* = (r_1 + r_2) + k_\alpha \sigma_{\hat{r}}. \tag{F.1}$$

Conclusively, the decision rule corresponding to the significance level α is to reject H_0 (and conclude that there will not be a collision) if $\hat{r} > r^*$, with r^* as in

(F.1). In other words, we chose to reject when

$$\frac{\hat{r}}{r^*} = \frac{\hat{r}}{(r_1 + r_2) + k_\alpha \sigma_{\hat{r}}} > 1.$$

One can easily show that this decision rule is equivalent with the decision rule given in (3.8). In other words,

$$\hat{r}/r^* > 1 \qquad \longleftrightarrow \qquad \omega > 1.$$