Vulnerability and risk associated with geohazards

General Non-Gaussian Probability Models for First-Order Reliability Method (FORM)
A State-of-the-Art Report

ICG Report 2004-2-4
NGI Report 20031091-4 30 September 2004

Partners in ICG

Client: The Research Council of Norway

Contact person:
Contract reference:

For

Project Manager: Suzanne Lacasse

Report prepared by: Kok-Kwang Phoon

Reviewed by: Farrokh Nadim

Work also carried out by:
Summary

Structural reliability theory has a significant impact on the development of modern design codes. Much of its success could be attributed to the advent of the first-order reliability method (FORM). This report presents a state-of-the-art review on modelling dependent non-Gaussian random vectors for FORM. FORM solution in standard space using Cholesky factorisation is recommended to ensure that variables are properly scaled and the correlation matrix is positive definite. The equivalent Gaussian technique widely used in FORM is easier to use when non-Gaussian random variables are coupled using a Gaussian-based fractile correlation approach. For the Gaussian-based product-moment approach, approximate probability transform based on Hermite polynomials is easier to use because a simple power series solution is available to relate the observed correlation with the underlying equivalent Gaussian correlation. If the observed data cannot be reasonably fitted to Gaussian-based product-moment/fractile models, it is necessary to consider more general copula-based models.
Contents

1 INTRODUCTION....................................................................................... 4
2 FIRST-ORDER RELIABILITY METHOD ............................................... 6
   2.1 Theory............................................................................................ 6
   2.2 Optimization in original space....................................................... 8
3 CONSTRUCTION OF UNCORRELATED VECTORS ......................... 10
   3.1 Cholesky and spectral transforms.................................................. 10
   3.2 Optimization in original versus standard space ............................. 12
4 POSITIVE DEFINITENESS .................................................................. 16
   4.1 Physical significance ................................................................. 16
   4.2 Verification using EXCEL ............................................................ 17
5 DEPENDENT NON-GAUSSIAN VECTORS......................................... 19
   5.1 Translation vectors....................................................................... 19
   5.2 Moment-based models............................................................... 21
   5.3 FORM comparison ...................................................................... 23
   5.4 Correlations in FORM ............................................................... 27
   5.5 Available solutions ..................................................................... 29
   5.6 Solution based on Hermite polynomials....................................... 32
   5.7 Solution based on modified four-term Hermite expansion.............. 34
6 FRACTILE CORRELATION ................................................................. 35
7 COPULA THEORY................................................................................ 38
8 CONCLUSIONS.................................................................................... 39
9 REFERENCES........................................................................................ 41
1 INTRODUCTION

Structural reliability theory has a significant impact on the development of modern design codes. Much of its success could be attributed to the advent of the first-order reliability method (FORM) – which provides a practical scheme of computing small probabilities of failure at high dimensional space spanned by the random variables in the problem. The basic theoretical result was given by Hasofer and Lind (1974). Additional elaborations were summarized by Rackwitz (2001). Ang and Tang (1984) presented numerous practical applications of FORM in their well known book on Probability Concepts in Engineering Planning and Design.

In simple terms, FORM essentially involves: (1) transforming a general random vector into a standard Gaussian vector, (2) locating the point of maximum probability density (most likely failure point, design point, or simply β-point) within the failure domain, and (3) estimating the probability of failure as \( p_f \approx \Phi(-\beta) \), in which \( \Phi(.) \) is the standard Gaussian cumulative distribution function. With reference to time-invariant reliability calculation, Rackwitz (2001) observed that: "For 90% of all applications this simple first-order theory fulfils all practical needs. Its numerical accuracy is usually more than sufficient." There is a note of finality in this claim that is probably justifiable for correlated Gaussian random vectors and independent non-Gaussian random vectors.

In principle, FORM is able to handle dependent non-Gaussian random vectors as well. It is well known that a general probability preserving transformation called Rosenblatt transformation exists. This transformation is simply based on the definition of conditional probability, as illustrated below for a general random vector with three components:

\[
F(x_1, x_2, x_3) = \frac{F(x_1, x_2, x_3)}{F(x_1, x_2)} \frac{F(x_1, x_2)}{F(x_1)} \frac{F(x_1)}{F(x_1)} = F(x_3 | x_1, x_2) F(x_2 | x_1) F(x_1)
\]

in which \( F \) is the n-dimensional cumulative distribution function (n is the number of arguments in \( F \)). It is immediately clear that by setting \( \Phi(u_1) = F(x_1) \), \( \Phi(u_2) = F(x_2|x_1) \), and \( \Phi(u_3) = F(x_3|x_1, x_2) \),

\[
\Phi(u_1, u_2, u_3) = \Phi(u_1) \Phi(u_2) \Phi(u_3) = F(x_1, x_2, x_3)
\]

and the requisite probability preserving Rosenblatt transformations are (see Appendix B, Ang and Tang 1984 for details):

\[
\begin{align*}
  u_1 &= \Phi^{-1} F(x_1) \\
  u_2 &= \Phi^{-1} F(x_2|x_1) \\
  u_3 &= \Phi^{-1} F(x_3|x_1, x_2)
\end{align*}
\]
The catch is that closed-form equations for the conditional probabilities in Equation (3) are not available except for the Gaussian case, a class of elliptical distributions (Azzalini and Capitanio 2003; Arslan 2004), and a handful of others (Devroye 1997). It is evident from the statistics literature that simulation of random samples from multivariate distributions is not trivial, much less problems relating to statistical estimation and model identification. In practice, it is further recognized that a general random vector could only be characterized reliably up to the marginal distribution and a measure of dependency such as the popular product-moment or Pearson correlation coefficient. With the exception of the Gaussian case and generalizations such as elliptical (Arslan 2004) and skew elliptical (Azzalini and Capitanio 2003) distributions, a general random vector could not be defined uniquely based on such limited information. An extensive body of literature on simulation of random processes exists (much of which are fairly recent) to attest to the theoretical and computational difficulties involved (e.g., Grigoriu 1984; Grigoriu 1988; Yamazaki and Shinozuka 1988; Deodatis and Micaletti 2001; Phoon et al. 2002; Puig et al. 2002; Sakamoto and Ghanem 2002; Puig and Akian 2003; Field and Grigoriu 2004). STRUREL (1991) implements Equation (3) for structural reliability analysis, but it is not clear how these complicated conditional probabilities could be constructed from limited data.

In view of the above theoretical and practical complexities, it is somewhat surprising that the FORM literature either downplay this problem by appealing to the Rosenblatt transformation (Rackwitz 2001) or assuming that dependent non-Gaussian vectors could be converted to correlated Gaussian vectors using a probability transform alone (Ang and Tang 1984; Low and Tang 2004). It is widely accepted that soil parameters are correlated spatially (Vanmarcke and Fenton 2003) and there is some evidence that correlation between different soil parameters exists as well (Carsel and Parrish 1988; Zhang et al. 2003). However, it is acknowledged that research on the latter aspect is very limited. The same could be said about the marginal distribution of soil parameters. Only second-moment soil statistics are established with some degree of reliability at present (Phoon and Kulhawy 1999a; 1999b). This may partially explain the apparent lack of interest in modelling dependent non-Gaussian random vectors more rigorously within FORM.

Nevertheless, a broader appreciation of dependent non-Gaussian random vectors is of practical interest because: (1) there are sufficient data to establish the non-Gaussianity of most load and resistance variables (e.g., Gumbel distribution for wind speed, shear resistance is always positive), (2) correlation could result from physics (e.g., sliding resistance increases with normal load, undrained shear strength increases with loading rate, etc.), and (3) curve-fitted parameters such as cohesion and friction angle in a linear Mohr-Coulomb failure envelope and SHANSEP parameters for normalized undrained shear strength are correlated by construction. It must be appreciated that validation of FORM using simulation only indicates that the non-Gaussian probability model selected for simulation happens to be compatible with the model pro-
duced implicitly by the above-mentioned assumption in FORM. This report would explain this opportune result theoretically in Section 6. Because dependent non-Gaussian probability models are non-unique, it is actually more important to check whether the selected probability model for FORM or simulation is compatible with observed data.

The objective of this report is to present a state-of-the-art review on modelling dependent non-Gaussian random vectors for FORM. The well known FORM solution for correlated Gaussian random vectors is first presented to provide the necessary technical background for understanding how the general non-Gaussian problem could be approached. Recent advances made in the literature on simulation of non-Gaussian stochastic processes are introduced to highlight their relevance to FORM. The focus of this report is on pragmatic computational issues. In particular, the practical significance of some relatively obscure but important theoretical considerations is illustrated in a concrete way using numerical examples. Discussions are restricted to single limit states containing only one $\beta$-point.

2 FIRST-ORDER RELIABILITY METHOD

2.1 Theory

The general reliability problem consists of a performance function $G(x_1, x_2, \ldots, x_n)$ and a multivariate probability density function $f(x_1, x_2, \ldots, x_n)$. The former is defined to be zero at the limit state, less than zero when the limit state is exceeded ("fail"), and larger than zero otherwise ("safe"). The latter specifies the likelihood of realizing any one particular set of input parameters $(x_1, x_2, \ldots, x_n)$.

The objective of reliability analysis is to calculate the probability of failure, which can be expressed formally as follows:

$$p_f = \int_{G<0} f(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n$$

The domain of integration is illustrated by a shaded region in the left panel of Figure 1a and the probability density function is represented by a 2D surface with elliptical contours for a bivariate Gaussian random vector (right panel). Exact solutions are not available even for this simple case unless the performance function is linear.
Figure 1. (a) General reliability problem for correlated Gaussian random vector and (b) Solution using FORM

The approximate solution obtained from FORM is easier to visualize in a standard space spanned by uncorrelated Gaussian random variables with zero mean and unit standard deviation (Fig. 1b). The probability density function for a standard bivariate Gaussian vector is a 2D surface with circular contours (right panel). If one replaces the actual limit state function ($G = 0$) by an approximate linear limit state function ($G_L = 0$) that passes through the $\beta$-point, that it follows immediately from rotational symmetry of the circular contours that:

$$p_T \approx \Phi(-\beta)$$
The practical result of interest here is that Equation (4) simply reduces to a nonlinear constrained optimisation problem:

$$\beta = \min \sqrt{\mathbf{u}^T \mathbf{u}} \quad \text{for } \{\mathbf{u}: G(\mathbf{u}) \leq 0\}$$

(6)

in which $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$. It is often cited that the $\beta$-point is the "best" linearization point because the probability density is highest at that point. In actuality, the choice of the $\beta$-point requires asymptotic analysis (Breitung 1984). In short, FORM works well only for sufficiently large $\beta$ - the usual rule-of-thumb is $\beta > 1$ (Rackwitz 2001).

### 2.2 Optimization in original space

Low and Tang (1997; 2004) presented a simple and practical computational procedure by exploiting the nonlinear optimization function in EXCEL (Solver add-in) and the equivalency between Equation (6) and the following expression in original variable space:

$$\beta = \min \sqrt{(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})} \quad \text{for } \{\mathbf{x}: G(\mathbf{x}) \leq 0\}$$

(7)

in which $\mathbf{x} = (x_1, x_2, \ldots, x_n)^T$, $\mathbf{m}$ = mean vector of $\mathbf{X}$, and $\mathbf{C}$ = covariance matrix of $\mathbf{X}$. The key advantage is that Equation (7) can be implemented using built-in functions in EXCEL without programming and EXCEL is widely available on PCs.

The procedure can be illustrated using the simple example shown in Figure 2. Under undrained compression loading, the performance function at ultimate limit state can be expressed as (Phoon and Kulhawy 2005):

$$G = Q_{kc} + M_{tc} Q_{tc}^{1.12} - W - F$$

$$Q_{sc} = \alpha (\pi B D s_{u1}) = [0.33 + 0.17 / (s_{u1} / p_a) + \varepsilon](\pi B D s_{u1})$$

$$Q_{tc} = [(6.17 s_{u2} \zeta_{cd} + q) (\pi B^2 / 4)]$$

(8)

in which $\alpha$ = adhesion factor, $s_{u1}$ = average undrained shear strength over shaft depth, $s_{u2}$ = average value within depth $B$ below the tip, $p_a$ = atmospheric pressure (100 kN/m$^2$), $\varepsilon$ = model error for adhesion factor, $M_{tc}$ = model factor for tip resistance, $\zeta_{cd}$ = depth modifier = 1 + 0.33 tan$^{-1}(D / B)$, and $q = \gammaD$ ($\gamma$ = soil total unit weight). The statistics of the random input parameters are summarized in Table 1. In addition, it is assumed that the spatial correlation between $X_1$ and $X_2$ is 0.5. The Gaussian assumption is adopted to simplify the illustration. With the exception of $\varepsilon$, the log-Gaussian model is usually more realistic because the parameters can only take positive values. A more realistic analysis using the log-Gaussian model is given in Section 5.2. The deterministic input parameters are: $B = 1.5$ m, $D = 10$ m, $W = 424$ kN, and $\gamma = 20$ kN/m$^3$. The mean factor of safety for this example is about 3.
The EXCEL solution is summarized in Figure 3. The fields in the Solver menu are intuitive and need no further explanation. More elaborate cell formulas are given in Table 2. Note that multiplication of vectors and matrices in Equation (7) can be implemented directly using array formulas in EXCEL (entered by pressing CTRL-SHIFT-ENTER). The remaining key items to take note of are:

1. check stability of the solution by using different initial trial values in (B16:B20) and
2. use "automatic scaling" under Solver options.

This example is stored in ‘Original space’ in the file "Compress (Gauss).xls".

![Drilled shaft under compression](image)

**Figure 2. Drilled shaft under compression**

<table>
<thead>
<tr>
<th>Statistics of random input parameters</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Random variable</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td>X&lt;sub&gt;1&lt;/sub&gt;</td>
<td>$s_{u1}$</td>
</tr>
<tr>
<td>X&lt;sub&gt;2&lt;/sub&gt;</td>
<td>$s_{u2}$</td>
</tr>
<tr>
<td>X&lt;sub&gt;3&lt;/sub&gt;</td>
<td>$\epsilon$</td>
</tr>
<tr>
<td>X&lt;sub&gt;4&lt;/sub&gt;</td>
<td>$M_{tc}$</td>
</tr>
<tr>
<td>X&lt;sub&gt;5&lt;/sub&gt;</td>
<td>F</td>
</tr>
</tbody>
</table>
### Table 2. Cell formulas for EXCEL solution

<table>
<thead>
<tr>
<th>Cell</th>
<th>Description</th>
<th>Cell formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>I9:M13</td>
<td>C^{-1}</td>
<td>MINVERSE(B9:F13)</td>
</tr>
<tr>
<td>D16:D20</td>
<td>x - m</td>
<td>B16:B20-C3:C7</td>
</tr>
</tbody>
</table>
| F16   | G           | (0.33+0.17*100/$B$16+$B$18)*PI()*SGS3*SGS4*SGS16+SGS19*((6.17*SGS7*SGS17+SGS6)*PI()*(SGS3^2/4)^1.12-0.01-SGS20)
|       |             | =SQRT(MMULT(MMULT(TRANSPOSE(D16:D20),I9:M13),D16:D20)) |
| I16   |             | NORMDIST(-I16) |
| I20   |             | NORMDIST(-I16) |

**Figure 3. EXCEL solution of FORM using Solver Add-in**

### 3 CONSTRUCTION OF UNCORRELATED VECTORS

#### 3.1 Cholesky and spectral transforms

It is crucial to understand that optimization in the original space is a matter of computational expediency and in no way implies that the underlying standard Gaussian space is unnecessary. In fact, it would be shown in the Section 5 that there are no simple means of generating dependent non-Gaussian vectors other
than applying a transform on Gaussian vectors, i.e., existing multivariate non-Gaussian models are derived from the multivariate Gaussian model.

This sub-section discusses the methods for transforming correlated Gaussian vectors into uncorrelated Gaussian vectors and Section 3.2 examines the need to perform these transforms in computation. First, observe that the covariance matrix \( C \) can be factorised as follows:

\[
C = L L^T
\]

(9)

in which \( L \) = lower triangular Cholesky factor. This is always possible because \( C \) is a symmetric positive definite matrix. If \( \mathbf{X} \) is a correlated Gaussian vector, it can be verified that the following transform would produce a standard Gaussian vector \( \mathbf{U} \):

\[
\mathbf{U} = L^{-1} (\mathbf{X} - \mathbf{m})
\]

(10)

by observing that:

\[
\mathbf{m}_U = E\mathbf{U} = L^{-1} E(\mathbf{X} - \mathbf{m}) = 0
\]

(11)

\[
C_U = E\mathbf{U}\mathbf{U}^T = L^{-1} E(\mathbf{X} - \mathbf{m})(\mathbf{X} - \mathbf{m})^T (L^{-1})^T = L^{-T} C (L^{-1})^T = (L^{-1}L)(L^{-1}L)^T = I
\]

(12)

in which \( E(.) = \) expectation operator, \( \mathbf{m}_U \) and \( C_U \) = mean vector and covariance matrix of \( \mathbf{U} \), respectively, \( \mathbf{0} \) = vector containing zeros, and \( \mathbf{I} \) = identity matrix.

For concreteness, consider a bivariate Gaussian vector with covariance given by:

\[
C = \begin{bmatrix}
\sigma_1^2 & \rho \sigma_1 \sigma_2 \\
\rho \sigma_1 \sigma_2 & \sigma_2^2
\end{bmatrix}
\]

(13)

The lower triangular Cholesky factor \( L \) is:

\[
L = \begin{bmatrix}
\sigma_1 & 0 \\
\rho \sigma_1 \sigma_2 & \sigma_2 \sqrt{1 - \rho^2}
\end{bmatrix}
\]

(14)

Explicit analytical expressions for \( L^{-1} \) is available for any \( n \times n \) covariance matrix (Phoon et al. 2005a), but it is more usual to compute \( L \) numerically. It is possible to do this in EXCEL by downloading a free matrix and linear algebra add-in from http://digilander.libero.it/foxes/index.htm. The array formula to do this is MAT_CHOLESKY. This add-in can be easily installed from the EXCEL main menu:
"Tools" > "Add-ins..." > "Browse" for the file "matrix.xla" and install by clicking on it.

If matrix.xla is installed correctly, the item "Matrix and Linear Algebra for Excel v. no." should appear in the "Add-ins available" list.

An alternate transform based on spectral decomposition is available (e.g., Ang and Tang 1984; Haldar and Mahadevan 1999):

\[ U = D^{-1/2} P^T (X - m) \]  \hspace{1cm} (15)

in which \( D \) = diagonal matrix containing eigenvalues in the leading diagonal and \( P \) = matrix whose columns are the corresponding eigenvectors. Again, it can be verified that Equation (15) would produce the desired standard Gaussian vector \( U \):

\[ C_U = E U U^T = D^{-1/2} P^T E (X - m)(X - m)^T P D^{-1/2} \]
\[ = (D^{-1/2} P^T) C (P D^{1/2}) = (D^{-1/2} P^T) (P D P^T) (P D^{-1/2}) = I \]  \hspace{1cm} (16)

The last step relies on the fact that \( P^{-1} = P^T \) because \( P \) is orthogonal and \( C \) is orthogonally diagonalizable, i.e. \( C = P D P^T \), because \( C \) is symmetric. Equation (15) is significantly more expensive than Cholesky factorisation in terms of computational cost. Hence, Equation (10) is preferred.

For the covariance matrix shown in Equation (13), the spectral transform matrix corresponding to \( L \) is:

\[ P D^{1/2} = \begin{bmatrix} \sigma_1 \sqrt{(1+\rho)/2} & \sigma_1 \sqrt{(1-\rho)/2} \\ \sigma_2 \sqrt{(1+\rho)/2} & -\sigma_2 \sqrt{(1-\rho)/2} \end{bmatrix} \]  \hspace{1cm} (17)

For a general \( n \times n \) covariance matrix, \( D \) and \( P \) can be computed using the EXCEL array formulas contained in the matrix and linear algebra add-in (http://digilander.libero.it/foxes/index.htm) MATEIGENVALUE_JACOBI and MATEIGENVECTOR_JACOBI, respectively.

### 3.2 Optimization in original versus standard space

Based on the above theoretical results, it is easy to demonstrate that Equations (6) and (7) are mathematically equivalent:

\[ u^T u = (x - m)^T (L^{-1})^T L^{-1} (x - m) \]
\[ = (x - m)^T (L L^T)^{-1} (x - m) = (x - m)^T C^{-1} (x - m) \]  \hspace{1cm} (18)

\[ u^T u = (x - m)^T P D^{-1/2} D^{-1/2} P^T (x - m) \]
\[ = (x - m)^T P D^T P^T (x - m) = (x - m)^T C (x - m) \]  \hspace{1cm} (19)
However, this does not imply that they are *computationally* equivalent. Low and Tang (2004) argued that optimization in original space using Equation (7) is an advantage because it does not require the additional step of transforming $\chi$ to $u$. However, they failed to recognize that computation of $C^{-1}$ in Equation (7) is *slower* than Cholesky factorisation because it does not exploit the symmetry of $C$. In addition, Cholesky factorisation is extremely stable numerically and is an efficient way to test whether a symmetric matrix is positive definite (details discussed in Section 4). These important numerical issues are well explained in Numerical Recipes (Press *et al.* 1992).

Figure 4 shows that optimization in the standard space only requires some minor adjustments to Figure 3:

1. compute lower triangular Cholesky factor $L$ in (I9:M13) using the array formula `MAT_CHOLESKY(B9:F13)`
2. trial values are iterated in standard space "By Changing Cells: $\$D$16:$\$D$20" in Solver
3. compute values in original space $x = Lu + m$ using the array formula `MMULT(I9:M13, D16:D20) + C3:C7`
4. compute reliability index in cell $\$I$16$ using `SUMSQ(D16:D20)^0.5`

![EXCEL solution of FORM using Cholesky factorisation](image-url)
This example is stored in ‘Standard space (Chol)’ in the file "Compress (Gauss).xls".

Similar adjustments could be made for optimization in standard space using spectral decomposition. Although the procedure is practical from an implementation point of view, it is computationally inefficient because eigenvalues and eigenvectors are expensive to compute in contrast to $C^{-1}$ and $L$. This example is stored in ‘Standard space (eigen)’ in the file "Compress (Gauss).xls".

Table 1 indicates that physical parameters could differ by orders of magnitudes. It is not surprising that the Solver option "Use Automatic Scaling" is necessary. Without invoking this option, the reliability index for Figure 3 would be 3.224, rather than 2.975. The solution $\beta = 2.975$ is verified by COMREL, which can be downloaded from http://www.strurel.de/EPages

Another sign of instability is that the result changes depending on the initial values assigned into cells ($B16:B20$) in Figure 3. This scaling problem in original space is recognized by STRUREL (1991). When optimization is carried out in standard space (Fig. 4), the result remains the same, regardless of the option and initial trial values. This illustrates that mathematical equivalency between Equations (6) and (7) does not say much about the behaviour of numerical schemes. A more direct example is that the intermediate iteration steps shown in cells ($B16:B20$) are completely different for Figures 3 and 4.

Besides re-running Solver using different initial trial values, it may be judicious to check the following identity:

$$u^* = \alpha^* \beta$$

$$\alpha^* = -\frac{[\partial G/\partial u_1 \quad \partial G/\partial u_2 \quad \cdots \quad \partial G/\partial u_n]^T}{\sqrt{\left(\partial G/\partial u_1\right)^2 + \cdots + \left(\partial G/\partial u_n\right)^2}}$$

in which $u^* = \text{solution vector}$ and $\alpha^* = \text{normalized gradient vector}$. Unfortunately, gradient information is not revealed by Solver and some finite difference method is needed to estimate $\alpha^*$. Note that the normalized gradient vector provides a measure of sensitivity, which is arguably more informative than $\beta$ in practical problems.

Figure 5 illustrates a simple forward difference method of computing this gradient information:
Figure 5. Computation of gradient

1. Enter the array formula \( \{=0.001*\text{M_ID()}+\text{D16:D20} \} \) in cells (A24:E28). These cells contain the following matrix \( \mathbf{u} + \Delta \mathbf{u} \):

\[
\begin{bmatrix}
  u_1 + 0.001 & u_1 & u_1 & u_1 \\
  u_2 & u_2 + 0.001 & u_2 & u_2 \\
  u_3 & u_3 & u_3 + 0.001 & u_3 & u_3 \\
  u_4 & u_4 & u_4 & u_4 + 0.001 & u_4 \\
  u_5 & u_5 & u_5 & u_5 & u_5 + 0.001
\end{bmatrix}
\]

2. The corresponding matrix in original space \( \mathbf{x} + \Delta \mathbf{x} = \mathbf{L}(\mathbf{u} + \Delta \mathbf{u}) + \mathbf{m} \). Hence, the array formula in cells (G24:K28) is expressed as \( \{=\text{MMULT}((I9:M13,A24:E28)+C3:C7) \} \).

3. The value of the performance function corresponding to each set of input parameters [stored column-wise in cells (G24:K28)] can be easily computed by copying the same formula from cell A32 to cell E32.

4. The gradient vector \( \nabla G = (\partial G/\partial u_1, \partial G/\partial u_2, ..., \partial G/\partial u_5)^T \) is stored in cells (G32:G36) and computed using the array formula \( \{=\text{TRANSPOSE}(((A32:E32)-$F$16)/0.001) \} \).

5. The magnitude of \( \nabla G \) is stored in cell G37 and is given by the formula \( \text{SUMSQ(G32:G36)}^{0.5} \).

6. The normalized gradient vector \( \mathbf{a}^* \) is stored in cells (H32:H36) and computed using the array formula \( \{=-(G32:G36)/G37 \} \).
7. Finally, Equation (20) is checked in cells (I32:I36) using the array formula 
\{=H32:H36*I16\}. If the correct β-point is found, the results in cells (I32:I36) must agree with the results in cells (D16:D20).

8. Another method is to check that \( \mathbf{u}^* \) and \( \mathbf{g}^* \) are parallel in cell F19 using 
\{=MMULT(TRANSPOSE(D16:D20),H32:H36)/I16\}. The result in this cell should be close to 1.

This example is stored in ‘Standard space (Chol) (2)’ in the file "Compress (Gauss).xls".

Finally, it is usual to estimate the error produced by FORM using importance sampling in the vicinity of the β-point. The formulation for this step is in standard space (STRUREL 1991).

4 POSITIVE DEFINITENESS

4.1 Physical significance

It has been noted above that the covariance matrix must be symmetric and positive definite. The former property follows directly from the definition of the covariance. The latter property is much harder to appreciate and it is natural to wonder if one should worry about this in practice.

A symmetric matrix \( \mathbf{C} \) is defined as positive definite if:

\[
\mathbf{a}^\top \mathbf{C} \mathbf{a} > 0 \quad \text{for all} \quad \mathbf{a} \neq \mathbf{0} \quad \text{(21)}
\]

in which \( \mathbf{a} = (a_1, a_2, \ldots, a_n)^\top \) is a vector of real numbers. The practical significance of Equation (21) is most readily appreciated by defining a random variable \( Y \) as a weighted linear sum of components in a random vector \( \mathbf{X} = (X_1, X_2, \ldots, X_n) \), i.e.:

\[
Y = \sum_{i=1}^{n} a_i X_i \quad \text{(22)}
\]

If \( \mathbf{C} \) is the covariance matrix for \( \mathbf{X} \), then it can be shown that the variance of \( Y \) is:

\[
\sigma_Y^2 = \mathbf{a}^\top \mathbf{C} \mathbf{a} \quad \text{(23)}
\]

Hence, if Equation (21) is not satisfied, some combination of weights could produce a random variable with negative variance! This contradicts the definition of the variance.
4.2 Verification using EXCEL

Despite the above significance, there is no easy way of recognizing whether this property has been violated. Phoon (2004) provided the simple example given below. The $4 \times 4$ matrix on the left is a valid correlation matrix, while the one on the right is not, even though both matrices are symmetric and all entries are bounded between -1 and 1.

\[
\begin{bmatrix}
1 & -0.163 & -0.893 & -0.526 \\
-0.163 & 1 & 0.211 & -0.281 \\
-0.893 & 0.211 & 1 & 0.783 \\
-0.526 & -0.281 & 0.783 & 1
\end{bmatrix}
\begin{bmatrix}
1 & -0.171 & -0.902 & -0.544 \\
-0.171 & 1 & 0.221 & -0.294 \\
-0.902 & 0.221 & 1 & 0.797 \\
-0.544 & -0.294 & 0.797 & 1
\end{bmatrix}
\]

It is not possible to verify positive definiteness by brute force using Equation (21) because one needs to consider all possible vectors $\mathbf{a}$. Low and Tang (2004) proposed verifying $(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m}) > 0$ at every Solver iteration. If this inequality is not satisfied at any iteration, the covariance matrix $\mathbf{C}$ is inadmissible. The advantage is that the comparison could be made at no cost because the term $(\mathbf{x} - \mathbf{m})^T \mathbf{C}^{-1} (\mathbf{x} - \mathbf{m})$ appears in Equation (7). The disadvantage is that one is none the wiser if the inequality is satisfied at every iteration.

It turns out that the Cholesky factor $\mathbf{L}$ or eigenvalue matrix $\mathbf{D}$ computed in the Section 3 for optimization in standard space is also a rigorous test for positive definiteness. It is even possible to isolate the offending entries in the matrix and adjust them to achieve positive definiteness. A systematic EXCEL procedure is outlined in Figure 6 and computational steps summarized in Table 3. The input cells are (A5:D8) and the output cells are (A11:D14). The output cells contain the array formula MAT_CHOLESKY. This example is stored in 'Positive definite' in the file "Compress (Gauss).xls".

The theoretical basis is that Cholesky factorisation is hierarchical, because an $n \times n$ covariance matrix $\mathbf{C}_n$ can always be partitioned as follows:

\[
\mathbf{C}_n = \begin{bmatrix}
\mathbf{C}_{n-1} & \mathbf{c}_{n-1} \\
\mathbf{c}_{n-1}^T & c_{nn}
\end{bmatrix} = \begin{bmatrix}
\mathbf{L}_{n-1} & 0 \\
\mathbf{L}_{n-1}^T & 1
\end{bmatrix} \begin{bmatrix}
\mathbf{I}_{n-1} & 0 \\
0 & 1
\end{bmatrix}
\]

in which $\mathbf{C}_{n-1} = \text{covariance sub-matrix for first (n-1) variables}$, $\mathbf{c}_{n-1}^T = \{c_{n1}, c_{n2}, \ldots, c_{nn-1}\}$, and $c_{nn} = \text{variance of n^{th} variable}$. Direct matrix multiplication yields:

\[
\begin{align*}
\mathbf{C}_{n-1} &= \mathbf{L}_{n-1} \mathbf{L}_{n-1}^T \\
\mathbf{c}_{n-1} &= \mathbf{L}_{n-1} \mathbf{1}_{n-1} \\
\mathbf{c}_{nn} &= \mathbf{1}_{n-1}^T \mathbf{1}_{n-1} + \mathbf{1}_{nn} \mathbf{1}_{nn}
\end{align*}
\]
It is clear from Equation (25a) that $L_{n-1}$ is the Cholesky factor of $C_{n-1}$, i.e. Cholesky factor of $C_n$ contains the Cholesky factor of $C_{n-1}$. The approach in Table 3 follows this inductive structure. If $L_{n-1}$ exists (i.e., $C_{n-1}$ is positive definite), then the offending entries that prevent square root operation to compute $l_{nn}$ in Equation (25c) must lie in the vector $l_{n-1}$, which is a function of the correlations in $C_{n-1}$.

**Table 3** EXCEL procedure to produce a positive definite correlation matrix based on Cholesky factorisation

**Step 1:** Assign an identity matrix to the input cells, i.e. assume all variables are uncorrelated.

**Step 2:** Enter the correlation between variable 2 and 1 = 0.2 in the input cells (shown in a separate area in Fig. 5 for clarity) and check the second diagonal entry in the Cholesky factor (L) if it is positive.

**Step 3:** Enter the correlations between variable 3 and 1 = -0.8 and between variable 3 and 2 = 0.6. The third entry in L is negative. Hence, the matrix is not positive definite.

**Step 3: (re-do)** Adjust the inputs in Step 3 until the third entry in L is positive.

**Step 4:** Expand the correlation matrix to include the next variable.

![EXCEL solution for constructing positive definite correlation matrix](image-url)
5 DEPENDENT NON-GAUSSIAN VECTORS

In practice, there are seldom enough data to characterize marginal distributions reliably, much less bivariate and higher order distributions. Hence, the Rosenblatt transformation is of limited use under such circumstances. It is widely accepted that a practical procedure for modelling dependent non-Gaussian vectors should involve at most the marginal distributions and the covariance matrix. The challenge is to find a multivariate distribution function that is consistent with this second-order information. It should be emphasized that the solution is not unique.

5.1 Translation vectors

The multivariate Gaussian distribution is available and can be defined uniquely by a mean vector and a covariance matrix. Hence, it is not surprising that the following cumulative distribution function (CDF) transform has been proposed:

\[ X_i = \Phi^{-1}F_i(Y_i) \]  

in which \( X_i \) = standard Gaussian random variable with zero mean and unit variance (but \( X_i \) and \( X_j \) may be correlated) and \( Y_i \) = non-Gaussian random variable with cumulative distribution function \( F_i \).

Note that the widely used equivalent Gaussian technique in FORM is simply a first-order Taylor series expansion of Equation (26):

\[ X_i \approx \Phi^{-1}F_i(y^*) + \left( \frac{dX_i}{dY_i} \right)(Y_i - y^*) \]
\[ = Y_i - y^* \left( \frac{dY_i}{dX_i} \frac{dX_i}{dY_i} \Phi^{-1}F_i(y^*) \right) \quad \text{(27)} \]
\[ = \frac{Y_i - m_N^i(y^*)}{s_N^i(y^*)} \]

in which \( y^* \) = linearization point, \( s_N^i = \frac{dY_i}{dX_i} = \frac{\phi[\Phi^{-1}F_i(y^*)]/f_i(y^*)}{\Phi^{-1}F_i(y^*)} \), \( \phi(.) \) = standard Gaussian density function, and \( f_i(.) \) = probability density function of \( Y_i \). The functions \( \Phi^{-1}(.) \) and \( \phi(.) \) are implemented in EXCEL as NORMSINV and NORMDIST(.,0,1,0), respectively.

For concreteness, consider a log-Gaussian random variable \( Y \) with mean = \( m \) and variance = \( s^2 \). By definition, \( \ln(Y) \) is Gaussian distributed with variance = \( \xi^2 = \ln(1+s^2/m^2) \) and mean = \( \lambda = \ln(m) - 0.5 \xi^2 \). The equivalent Gaussian standard deviation \( (s^N) \) and mean \( (m_N) \) are:

\[ s_N^i = y^*\xi \]
\[ m_N^i = y^*(1 - \ln(y^*) + \lambda) \]  

\[ \text{(28)} \]
It is also possible to approximate $Y_i = F_i^{-1}(\Phi(X_i))$ using Hermite polynomials:

$$Y_i = \sum_{k=0}^{\infty} a_i k H_k(X_i)$$  \hspace{1cm} (29)

in which the Hermite polynomials $H_j(.)$ are given by:

$$H_0(X) = 1$$
$$H_1(X) = X$$
$$H_2(X) = X^2 - 1$$
$$H_{k+1}(X) = X H_k(X) - k H_{k-1}(X)$$

and the coefficients $a_i j$ are given by:

$$a_i k = \frac{1}{k!} \int_0^1 F_i^{-1}(t) H_k[\Phi^{-1}(t)] dt$$ \hspace{1cm} (31)

Note that the first coefficient $a_{i0} (k = 0)$ is simply the mean of $Y_i (m_{Yi})$. This can be derived by substituting $H_0 = 1$ into Equation (31) and by a change of variable: $y = F_i^{-1}(t)$, $dt = f_i(y)$ dy, upper (lower) integration limit = maximum (minimum) value of $Y_i = y_{max} (y_{min}) = F_i^{-1}(1) [F_i^{-1}(0)]$:

$$a_{i0} = m_{Yi} = \int_{y_{min}}^{y_{max}} y f_i(y) dy$$ \hspace{1cm} (32)

The Hermite polynomials can be generated efficiently using the 3-term recurrence relation shown in Equation (30), but the coefficients have to be evaluated numerically in most cases. An analytical expression is available if $Y$ is log-Gaussian (Berveiller et al. 2004):

$$a_i k = \frac{\xi_i^k}{k!} \exp \left( \lambda_i + \frac{\xi_i^2}{2} \right) \frac{\xi_i}{k} a_{i,k-1}$$ \hspace{1cm} (33)

Figure 7 shows that a Hermite expansion consisting of 6 terms ($k = 0, 1, \ldots, 5$) is sufficient to approximate a highly skewed log-Gaussian random variable. Note that the lower probability tail has been magnified using log-scale because this region is important for reliability analysis.
5.2 Moment-based models

The observed data may be sufficient to produce reliable moment estimates only (typically up to kurtosis), rather than the full probability density function. Winterstein and Kashef (2000) discussed a suite of moment-based non-Gaussian probability models for such applications. One version is the four-term modified Hermite expansion that only requires the third and fourth moment of the non-Gaussian random variable (Winterstein 1987):

\[ Y_i = m_{Y_i} + s_{Y_i} k_i [X_i + h_{3i} (X_i^2 - 1) + h_{4i} (X_i^3 - 3X_i)] \]  

in which \( m_{Y_i} \) and \( s_{Y_i} \) = mean and standard deviation of \( Y_i \), respectively. The constants \( k_i, h_{3i}, \) and \( h_{4i} \) are functions of the skewness and kurtosis (Table 4).

Analytical expressions for the central moments for a Gaussian random variable are available to any order:

\[ \gamma_k = E[(X_i - m_{X_i})/s_{X_i}]^k = 0 \quad \text{for } k \text{ odd} \]
\[ = (1)(3)(5)(7) \ldots (k-1) \quad \text{for } k \text{ even} \]  

---

Figure 7. Hermite expansions for log-Gaussian random variable with \( \lambda = 0 \) and \( \xi = 1 \) (probability density function shown in bottom right corner)
Table 4. Coefficients of four-term modified Hermite expansion (modified from STRUREL 1991).

<table>
<thead>
<tr>
<th>Coefficient</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skewness: $\gamma_3$</td>
<td>$E[(Y_i - m_{Y_i})/s_{Y_i}]^3$</td>
</tr>
<tr>
<td>Kurtosis: $\gamma_4$</td>
<td>$E[(Y_i - m_{Y_i})/s_{Y_i}]^4$</td>
</tr>
</tbody>
</table>

For $\gamma_4 < 3$:

<table>
<thead>
<tr>
<th>$h_4$</th>
<th>$h_3$</th>
<th>$k_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\left(\frac{\gamma_4 - 3}{24}\right) - 27\left(\frac{\gamma_4 - 3}{24}\right)^2$</td>
<td>$\frac{\gamma_3}{6(1 + 24 h_4)}$</td>
<td>$\frac{1}{\sqrt{1 + 10 h_3^2 + 42 h_4^2}}$</td>
</tr>
</tbody>
</table>

For $\gamma_4 > 3$:

<table>
<thead>
<tr>
<th>$h_4$</th>
<th>$h_3$</th>
<th>$k_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\frac{1.5\gamma_4 - 3.5 - 1}{18}$</td>
<td>$\frac{\gamma_3}{6(1 + 6 h_4)}$</td>
<td>$\frac{1}{\sqrt{1 + 2 h_3^2 + 6 h_4^2}}$</td>
</tr>
</tbody>
</table>

Hence, it can be shown after some tedious algebraic manipulations that the skewness and kurtosis of the random variable defined by Equation (34) are respectively:

$$\gamma_3 = E[(Y_i - m_{Y_i})/s_{Y_i}]^3 = k_i^3(6 h_{3i} + 36 h_{3i} h_{4i} + 8 h_{3i}^3 + 108 h_{3i} h_{4i}^2)$$  \hspace{1cm} (36)

$$\gamma_4 = E[(Y_i - m_{Y_i})/s_{Y_i}]^4 = k_i^4(3 + 48 h_{3i}^2 + 3348 h_{4i}^4 + 24 h_{4i} + 1296 h_{4i}^3 + 60 h_{3i}^2 + 252 h_{4i}^2 + 2232 h_{3i}^2 h_{4i}^2 + 576 h_{3i}^2 h_{4i})$$  \hspace{1cm} (37)

For the log-Gaussian random variable, the 4-term Hermite expansion based on Equation (33) is:

$$Y_i = m_{Y_i} + s_{Y_i} k_i \left[ X_i + \frac{\xi_i}{2} (X_i^2 - 1) + \frac{\xi_i^2}{6} (X_i^3 - 3X_i^1) \right]$$  \hspace{1cm} (38)

in which $m_{Y_i} = \exp(\lambda_i + 0.5 \xi_i^2)$, $s_{Y_i} = m_{Y_i} \sqrt{\exp(\xi_i^2) - 1}$, and $k_i = m_{Y_i} \xi_i / s_{Y_i}$. The skewness and kurtosis for the log-Gaussian case are available analytically:

$$\gamma_3 = E[(Y_i - m_{Y_i})/s_{Y_i}]^3 = [\exp(\xi_i^2) + 2] \sqrt{\exp(\xi_i^2) - 1} - 1$$  \hspace{1cm} (39)

$$\gamma_4 = E[(Y_i - m_{Y_i})/s_{Y_i}]^4 = \exp(4\xi_i^2) + 2\exp(3\xi_i^2) + 3\exp(2\xi_i^2) - 3$$  \hspace{1cm} (40)

Table 5 compares skewness [Eq. (36)] and kurtosis [Eq. (37)] produced by the modified Hermite expansion with the corresponding theoretical solutions given in Equations (39) and (40), respectively. It is clear that this expansion only works well for relatively small $\xi$. The coefficients of variation (COVs) for this range of $\xi$ are sufficient to cover most load variables in structural engineering.
practice, but are not typical for geotechnical variabilities (Phoon and Kulhawy 1999a).

Table 5. Skewness and kurtosis produced by four-term modified Hermite expansion for the log-Gaussian model

<table>
<thead>
<tr>
<th>ξ</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Log-</td>
<td>Skewness</td>
<td>0.3018</td>
<td>0.9495</td>
<td>1.7502</td>
</tr>
<tr>
<td>Gaussian</td>
<td>Kurtosis</td>
<td>3.1623</td>
<td>4.6449</td>
<td>8.8984</td>
</tr>
<tr>
<td>Modified</td>
<td>Skewness</td>
<td>0.3007</td>
<td>0.9326</td>
<td>1.6777</td>
</tr>
<tr>
<td>Hermite</td>
<td>Kurtosis</td>
<td>3.2815</td>
<td>5.7798</td>
<td>12.9616</td>
</tr>
<tr>
<td>Hermite</td>
<td>Skewness</td>
<td>0.3017</td>
<td>0.9451</td>
<td>1.6903</td>
</tr>
<tr>
<td></td>
<td>Kurtosis</td>
<td>3.1616</td>
<td>4.5677</td>
<td>7.9188</td>
</tr>
</tbody>
</table>

More significantly, this expansion can produce negative values with significant probability when ξ is large (Fig. 8). Hence, application in FORM could produce physically non-meaningful results if the non-Gaussian random variables can only take positive values.

Equations (36) and (37) are re-computed using the actual Hermite coefficients shown in Equation (38), i.e., \( h_{3i} = \xi_i/2 \) and \( h_{4i} = \xi_i^2/6 \). Results are shown in the last row of Table 5. These actual Hermite coefficients produce better fit to the skewness and kurtosis when \( \xi_i \) is moderate, say less than 0.5. The modified Hermite coefficients (Table 4) produce a better fit to kurtosis when \( \xi_i = 1 \), but do not perform as well for skewness.

5.3 FORM comparison

The example shown in Figure 2 is re-analysed by assuming that all the random variables in Table 1 with the exception of \( \varepsilon \) are log-Gaussian. Because the equivalent Gaussian standard deviation \( s_i^N \) is a function of the linearization point and changes within each iteration, it is more efficient to re-write Equation (7) as follows:

\[
\beta = \min \sqrt{(y - m)^T S^{-1} R^{-1} S^{-1} (y - m)} \quad \text{for } \{y; G(y) \leq 0\} \tag{41}
\]

in which \( C = (S R S) \), \( R = \) correlation matrix, and \( S = \) diagonal matrix containing \( s_i^N \) in the leading diagonal. In doing so, the correlation matrix is inverted only once before FORM iterations and \( S^{-1}(y - m) \) is simply evaluated as the following vector directly:

\[
S^{-1}(y - m) = \begin{bmatrix}
1/s_i^N & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & 1/s_n^N
\end{bmatrix}
\begin{bmatrix}
y_i - m_i^N \\
\vdots \\
y_n - m_n^N
\end{bmatrix}
= \begin{bmatrix}
(y_i - m_i^N)/s_i^N \\
\vdots \\
(y_n - m_n^N)/s_n^N
\end{bmatrix} \tag{42}
\]
Equation (42) is the version used by Low and Tang (2004) for FORM optimization in EXCEL. Figure 9 shows that the modification to Figure 3 is quite minor:

1. compute $m_i^N$ and $s_i^N$ in cells (L3:L7) and (M3:M7), respectively
2. change array formula in (D16:D20) to Equation (42)
3. change cell formula in I16 to Equation (41), i.e.:
   
   \[ \{\text{SQRT(MMULT(MMULT(TRANSPOSE(D16:D20),I9:M13),} \]
   \[ \text{D16:D20))} \}

This example is stored in ‘Original space (Equiv Gauss)’ in the file "Compress (LogGauss).xls".

**Probability density function**

![Probability Density Function](image(a).png)

**Cumulative distribution function**

![Cumulative Distribution Function](image(b).png)

(a) $\xi = 0.1$

(b) $\xi = 0.3$
It is immediately clear that optimization in standard space is not possible using Equation (41). For this simple example, the reliability index from Solver is 3.193 and 3.621, when automatic scaling is on and off, respectively. Different initial trial values would also produce different results when automatic scaling is off.

Figure 8. Accuracy of modified Hermite expansion for various log-Gaussian distributions
Figure 9. **EXCEL solution of FORM in original space using equivalent Gaussian method**

Figure 10 shows that Hermite expansions can be used in FORM by making the following minor modifications to Figure 4:

1. covariance matrix stored in (B9: F13) is the same as the correlation matrix because \( \mathbf{X} \) contains Gaussian random variables with zero mean and unit variance
2. initialise \( H_0 = 1 \) and \( H_1 = \mathbf{X} \) in cells (B23:B27) and cells (C23:C27), respectively. The rest of the Hermite polynomials are computed using the 3-term recurrence relation given in Equation (30)
3. compute Hermite coefficients in cells (I23:N27) using Equation (33)
4. insert array formula in cells (B16:B20) to compute \( Y_i = \sum a_k H_k(X_i) \):

\[
\{=\text{MMULT}(B24:G24,\text{TRANSPOSE}(I24:N24))\}
\]

This example is stored in ‘Standard space (Chol-Hermite)’ in the file "Compress (LogGauss).xls".

The number of terms required to match the equivalent Gaussian solution is shown in Table 6. COMREL produces the same solution \( \beta = 3.193 \). For all practical purposes, a 4-term expansion is sufficient. Phoon (2003) arrived at the same conclusion based on a laterally loaded pile example with the effective stress friction angle modelled as a beta random variable. Note that a beta distribution is more challenging to approximate because it has two finite (lower and upper) bounds, while the Gaussian random variable is unbounded.
Figure 10. EXCEL solution of FORM in standard space using Hermite polynomials

Table 6. Comparison between Hermite and equivalent Gaussian approximations in FORM

<table>
<thead>
<tr>
<th>Method</th>
<th>$\beta$</th>
<th>$p_f$</th>
<th>No. of Solver iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent Gaussian</td>
<td>3.193</td>
<td>$7.03 \times 10^{-4}$</td>
<td>8</td>
</tr>
<tr>
<td>Hermite: 2-term</td>
<td>3.019</td>
<td>$1.27 \times 10^{-1}$</td>
<td>10</td>
</tr>
<tr>
<td>3-term</td>
<td>3.167</td>
<td>$7.69 \times 10^{-4}$</td>
<td>13</td>
</tr>
<tr>
<td>4-term</td>
<td>3.198</td>
<td>$6.92 \times 10^{-4}$</td>
<td>12</td>
</tr>
<tr>
<td>5-term</td>
<td>3.193</td>
<td>$7.03 \times 10^{-4}$</td>
<td>12</td>
</tr>
<tr>
<td>6-term</td>
<td>3.193</td>
<td>$7.03 \times 10^{-4}$</td>
<td>12</td>
</tr>
</tbody>
</table>

Computationally, the Hermite expansion is less efficient than the equivalent Gaussian expansion, because the Hermite coefficients have to be determined through numerical integration. However, it permits optimization in standard space, where the variables are better scaled. In addition, the CDF transform [Equation (26)] is approximated as an independent step, rather than being coupled to the FORM iterations as shown in Equation (41). More importantly, the Hermite expansion can be used for simulation, such as importance sampling in the vicinity of the $\beta$-point. It is not possible to do this with the equivalent Gaussian expansion because $m_i^N$ and $s_i^N$ are not constants, but are functions of the linearization point.

5.4 Correlations in FORM

There is one glaring difficulty to the translation approach of modelling dependent non-Gaussian random vectors. This difficulty arises because the correlation...
between two non-Gaussian random variables ($\rho_{Y_1Y_2}$) is not the same as the correlation between the equivalent Gaussian random variables ($\rho_{X_1X_2}$) appearing in Equation (26). The exact relation is given by:

\[ m_{Y_1}m_{Y_2} + \rho_{Y_1Y_2}s_{Y_1}s_{Y_2} = \int_{-\infty}^{\infty} \left[ F_1^{-1}(\Phi(x_1))F_2^{-1}(\Phi(x_2)) \right] \phi(x_1, x_2; \rho_{X_1X_2}) \, dx_1 \, dx_2 \quad (43) \]

\[ \phi(x_1, x_2; \rho_{X_1X_2}) = \frac{(1 - \rho_{X_1X_2}^2)^{-1/2}}{2\pi} \exp \left[ -\frac{x_1^2 - 2\rho_{X_1X_2}x_1x_2 + x_2^2}{2(1 - \rho_{X_1X_2}^2)} \right] \quad (44) \]

Note that the observed data produce $\rho_{Y_1Y_2}$ and one needs to solve Equation (43) to obtain $\rho_{X_1X_2}$. The example in Section 5.3 assumes that $\rho_{X_iX_j}$ in cells (B9: F13) are known, which circumvent this important practical consideration.

It is reasonable to question if Equation (43) is only of theoretical interest. In other words, can one safely assume that $\rho_{Y_1Y_2} \approx \rho_{X_1X_2}$ for all practical purposes? The general answer is no. It is most convenient to illustrate this using the log-Gaussian model because Equation (43) can be evaluated analytically:

\[ \rho_{Y_1Y_2} = \frac{\exp(\xi_1\xi_2\rho_{X_1X_2}) - 1}{\sqrt{[\exp(\xi_1^2) - 1][\exp(\xi_2^2) - 1]}} \quad (45) \]

Figure 11 shows some relationships between $\rho_{Y_1Y_2}$ and $\rho_{X_1X_2}$ for various combinations of $\xi_1$ and $\xi_2$. The assumption of $\rho_{Y_1Y_2} \approx \rho_{X_1X_2}$ only works well when $\xi_1$ and $\xi_2$ are small. It can be seen in Figure 8 that the log-Gaussian probability density function becomes increasingly skewed as $\xi$ increases. Hence, it may be concluded (not surprising) that this assumption does not work for strongly non-Gaussian distributions, i.e., those that do not look "bell" shaped. More significantly, the bounds for $\rho_{Y_1Y_2}$ could be less than 1 and/or greater than -1. This is proven theoretically for the general case by Arwade (2004). The practical ramification is that some observed data cannot be modelled using translation vectors.
Figure 11. Some relationships between observed correlation ($\rho_{Y1Y2}$) and correlation of underlying equivalent Gaussian random variables ($\rho_{X1X2}$)

5.5 Available solutions

The above difficulty is well known in the literature on simulation of non-Gaussian processes. Grigoriu (1984; 1988) established a rigorous framework for such dependent translation vectors and proposed solving the integral equation [Equation (43)] or an equivalent differential equation given by Price theorem (Deutsch 1962) explicitly. This direct approach is not easy to apply in practice. Yamazaki and Shinozuka (1988) and Deodatis and Micaletti (2001) presented a simpler numerical solution based on iterative updating of the power spectral density function. This approach exploits the fact that a sufficiently long random process can be represented by a Fourier series. For a short process
or a random vector with few components, this representation must be replaced by the Karhunen-Loeve (K-L) series (Huang et al. 2001). The key difference is that two covariance functions only differ in their power spectral densities when the process is long. The trigonometric basis functions in the expansion do not change. On the other hand, for a short process, the basis functions in the K-L series are eigenfunctions and they are not independent of the covariance function. It suffices to note here that this approach cannot be applied to random vectors.

A more promising iterative approach based on the more general K-L series was proposed by Phoon et al. (2002; 2005a). The basic steps for a zero-mean non-Gaussian random vector $Y$ are summarized as follows:

1. $Y^{(k)} = P D^{1/2} Z^{(k)}$  
2. Initialise the K-L random variables $Z^{(1)} = \tilde{U}$  
3. Compute $Y^{(1)}$ using Equation (46) and determine the empirical CDF for each component, $\tilde{F}_i$  
4. Transform $Y^{(1)}$ to match the target non-Gaussian CDF of each component $F_i$ by $Y^{(2)} = F_i^{-1} \tilde{F}_i [Y^{(1)}]$  
5. Compute the mean of $Y^{(2)} = m^{(2)}$ and update $Z^{(2)} = D^{-1/2} P^T (X^{(2)} - m^{(2)})$  
6. Make $Z^{(2)}$ uncorrelated and repeat the procedure

Note that $\tilde{U}$, $P$, and $D$ were defined in Section 3.1. The above approach was developed for simulation of strongly non-Gaussian processes, but could be applied to random vectors with no modification. Once $Y$ is defined through simulation, $X$ and its correlation structure can be determined directly by Equation (26). Note that this approach is far superior to a brute force simulation approach (always a possibility) in which one guesses $\rho_{X1X2}$, simulates correlated samples of $X$, transforms them to get correlated samples of $Y$, and verifies that the desired $\rho_{Y1Y2}$ is achieved. A $n \times n$ correlation matrix has $n(n-1)/2$ distinct entries and this trial and error brute force approach would be tedious.

The trial and error approach based on simulation is simple to implement in EXCEL. The computational steps are summarized in Table 7.
Table 7. Determination of $\rho_{X1X2}$ from $\rho_{Y1Y2}$ for translation vectors using simulation

<table>
<thead>
<tr>
<th>Step</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Simulate 2 columns of uncorrelated standard Gaussian numbers $U = (U_1$ and $U_2)$ using: Tools &gt; Data Analysis &gt; Random Number Generation 1000 numbers are stored in cells (A2:B1001)</td>
</tr>
<tr>
<td>2</td>
<td>Remove the means (can be significant for small sample size) from the simulated numbers and store them in cells (D2:E1001)</td>
</tr>
<tr>
<td>3</td>
<td>Compute the covariance matrix of these zero-mean numbers ($C_U$) using the function COVAR (can be different from identity for small sample size)</td>
</tr>
<tr>
<td>4</td>
<td>Compute Cholesky factor ($L$) of the Gaussina correlation matrix $R$ in cells (J2:K3) using the array formula {$=\text{MAT}_\text{CHOLESKY}(G2:H3)$}</td>
</tr>
<tr>
<td>5</td>
<td>Compute the inverse Cholesky factor of $C_U$ ($L^{-1}$) in cells (J6:K7) using the array formula {$=\text{MINVERSE}(\text{MAT}_\text{CHOLESKY}(G6:H7))$}</td>
</tr>
<tr>
<td>6</td>
<td>Compute 2 columns of correlated Gaussian numbers with zero mean and unit standard deviation using $X = (U - m)(L \times L^{-1})^T$. The array formula is: {$=\text{MMULT}(D2:E1001,\text{TRANSPOSE(\text{MMULT}(J2:K3,J6:K7)))}$}</td>
</tr>
<tr>
<td>7</td>
<td>Compute 2 columns of correlated log-Gaussian numbers using: $Y_1 = \exp(\lambda_1 + \xi_1 X_1)$ and $Y_2 = \exp(\lambda_2 + \xi_2 X_2)$</td>
</tr>
<tr>
<td>8</td>
<td>Compute the correlation between $Y_1$ and $Y_2$ using CORREL(P2:P1001,Q2:Q1001)</td>
</tr>
<tr>
<td>9</td>
<td>Change the Gaussian correlation value ($\rho_{X1X2}$) in cells (G2:H3) until the desired non-Gaussian correlation ($\rho_{Y1Y2}$) is achieved.</td>
</tr>
</tbody>
</table>

Note: EXCEL worksheet (Fig. 12) found in ‘Correlation simulation’ in file "Compress (LogGauss).xls".
5.6 Solution based on Hermite polynomials

Recent papers (Puig et al. 2002; Sakamoto and Ghanem 2002; Puig and Akian 2003) also focus on the Hermite expansion because it presents a relatively simpler power series solution to Equation (43) (Sakamoto and Ghanem 2002):

$$EY, Y_2 = \sum_{k=0}^{\infty} k! a_k a_{2k} (\rho_X)_{2k}$$

It is not possible to evaluate Equation (47) directly because factorials are notorious for growing very rapidly (e.g., 20! \approx 10^{18}). It is easier to compute the sum to n terms recursively as follows:

$$S_n = n a_n a_{2n} (\rho_X)_{2n}$$

$$S_{n-1} = (n-1)(\rho_X)^2 a_{n-1} a_{2n-1} + S_n$$

$$S_k = k \rho_X a_{2k} (a_k a_{2k} + S_{k+1})$$

$$S_1 = \rho_X a_{21} (a_1 a_{21} + S_2)$$
\[ \sum_{k=0}^{n} k! a_k a_{2k} (\rho_{X1X2})^k = a_{10} a_{20} + S_1 \] (49)

Note that the unknown in Equation (47) is \( \rho_{X1X2} \) and analytical solutions are not available in the general case. However, it is not difficult to solve Equation (47) using numerical methods such as the nonlinear optimization Solver function in EXCEL. Puig et al. (2002) noted that optimization is preferred rather than applying a nonlinear equation solver because it is possible to impose the added constraint of positive definiteness.

For log-Gaussian model, Equation (47) can be expressed analytically as follows:

\[ EY_1 Y_2 = \sum_{k=0}^{\infty} \frac{(\xi_1 \xi_2 \rho_{X1X2})^k}{k!} \exp \left( \lambda_1 + \lambda_2 + \frac{\xi_1^2}{2} + \frac{\xi_2^2}{2} \right) \] (50)

Hence, the correlation between the original log-Gaussian random variables \( \rho_{Y1Y2} \) is given by:

\[ \rho_{Y1Y2} = \left( \sum_{k=0}^{\infty} \frac{(\xi_1 \xi_2 \rho_{X1X2})^k}{k!} \exp \left( \lambda_1 + \lambda_2 + \frac{\xi_1^2}{2} + \frac{\xi_2^2}{2} \right) - m_{Y1} m_{Y2} / \right) / \left( \sqrt{[\exp(\xi_1^2) - 1][\exp(\xi_2^2) - 1]} \right) \] (51)

because \( m_{Y1} = \exp(\lambda_1 + 0.5 \xi_1^2) \) and \( s_{Y1}^2 = \exp(2\lambda_1 + \xi_1^2)[\exp(\xi_1^2) - 1] \). A comparison between Eqs. (45) and (51) shows that they are identical. Figure 13 shows that a 4-term Hermite expansion is sufficient to reproduce the relationship between \( \rho_{Y1Y2} \) and \( \rho_{X1X2} \). It was mentioned in Section 5.3 that a 4-term expansion is also sufficient for the CDF transform. Hence, the Hermite expansion is quite practical. It is surprising that this expansion is not more widely exploited in FORM.

Assuming that the observed correlation matrix is positive definite, there is no guarantee that the correlation matrix of the equivalent Gaussian random variables is positive definite, even if the correlations \( \rho_{X1X2} \) are determined exactly from Equation (43). This is a fundamental theoretical constraint related to the characterization of multivariate dependence based on pair-wise measures of dependence. Hence, performing Cholesky factorisation on the correlation matrix of \( X \) is still necessary.
Figure 13. Relationship between $\rho_{Y1Y2}$ and $\rho_{X1X2}$ for $\xi_1 = 1$ and $\xi_2 = 1$

### 5.7 Solution based on modified four-term Hermite expansion

Although the four-term modified Hermite expansion is quite limited as discussed in Section 5.2, it could produce $\rho_{X1X2}$ from $\rho_{Y1Y2}$ in closed form. First, observe that Equation (49) reduces to the following special case when $n = 3$:

$$
\rho_{Y1Y2} = k_i k_j (\rho_{X1X2} + 2 h_i h_j \rho_{X1X2}^2 + 6 h_i h_j h_k \rho_{X1X2}^3)
$$

(52)

It is also possible to derive Equation (52) exactly from Equation (34) by using multivariate Gaussian moments, which are known in closed-form (Isserlis 1918). The unknown $\rho_{X1X2}$ can be computed as follows (STRUREL 1991, Winterstein 1987):

$$
\rho_{X1X2} = A^{1/3} - \frac{B}{A^{1/3}} - \frac{C}{9}
$$

(53)

The parameters $A$, $B$, an $C$ are given in Table 8. Equation (53) is compared with the theoretical solution [Eq. (45)] in Figure 14. Although it is absurd to talk about correlations larger and smaller than 1 and -1, respectively, Figure 14 illustrates that such inadmissible solutions can be produced by Equation (53). In practice, it is simple to check for this violation. However, it is more difficult to check for violations of theoretical upper and lower bounds. For example, Figure 14b indicates that it is possible to produce a seemingly valid $\rho_{X1X2} > -1$ when $\rho_{Y1Y2} = 0.5$ from Equation (53), but such a solution does not exist theoretically!
Table 8. Formulas for coefficients in Equation (53)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c_k$</td>
<td>$k_i k_j$</td>
</tr>
<tr>
<td>$c_3$</td>
<td>$h_i h_j$</td>
</tr>
<tr>
<td>$c_4$</td>
<td>$h_i k_j$</td>
</tr>
<tr>
<td>$A$</td>
<td>( \frac{27c_3 c_4 c_k + 243c_4^2 \rho_{Y1Y2} - 4c_3 c_k + 9c_4 \sqrt{3\lambda}}{2916c_4^2 c_k} )</td>
</tr>
<tr>
<td>$\tilde{A}$</td>
<td>( \frac{6c_4 c_k^2 - c_3^2 c_k^2 + 54 c_3 c_4 \rho_{Y1Y2} + 243 c_4^2 \rho_{Y1Y2}^2 - 8 c_3 c_k \rho_{Y1Y2}}{162c_4^2} )</td>
</tr>
<tr>
<td>$B$</td>
<td>( \frac{9c_3 - 2c_3^2}{162c_4^2} )</td>
</tr>
<tr>
<td>$C$</td>
<td>( \frac{c_3}{c_4} )</td>
</tr>
</tbody>
</table>

It is not surprising that Equation (53) works well when $\xi_1$ and $\xi_2$ are small. But $\rho_{X1X2} \approx \rho_{Y1Y2}$ in this situation. Hence, application of Equation (53) for geotechnical practice is limited. It is interesting to note that the 4-term Hermite expansion produces much superior agreement (Fig. 13) than the 4-term modified Hermite expansion (Fig 14b) for the log-Gaussian case with $\xi_1 = \xi_2 = 1$.

6 FRACTILE CORRELATION

The difficulty in computing correlations under the nonlinear transform given in Equation (26) is fundamentally caused by the presence of the non-Gaussian CDFs ($F_1$ and $F_2$) in Equation (43). A simple and practical method to circumvent this difficulty is to adopt fractile correlations (sample analog being Spearman rank correlation), rather than the usual product-moment correlations (sample analog being Pearson correlation) (Phoon et al. 2005b).

The fractile is defined as:

\[
V_i = F_i(Y_i)
\]  

(54)

which is uniformly distributed between 0 and 1 with mean $= EV_i = 1/2$ and variance $= E(V_i - EV_i)^2 = 1/12$. By definition, the fractile correlation is given by:

\[
r_{ij} = \frac{E(V_i V_j) - EV_i EV_j}{\sqrt{E(V_i - EV_i)^2 E(V_j - V_j)^2}} = 12 E(V_i V_j) - 3
\]  

(55)

It is evident from definition that the fractile correlation is invariant under monotone transforms, specifically of the type given by Equation (26):

\[
V_i = F_i(Y_i) = F_i[F_i^{-1}(\Phi(X_i))] = \Phi(X_i)
\]  

(56)
Hence, a non-Gaussian random vector generated using Equation (26) will match the fractile correlation function of the underlying Gaussian random vector exactly.

The remaining problem is to determine the relationship between \( r_{ij} \) and the product-moment correlation \( \rho_{XjXj} \) of the underlying Gaussian random vector. Fortunately, an analytical solution for Gaussian random variables does exist (Hotelling and Pabst 1936):
\[ \rho_{X_iX_j} = 2 \sin \left( \frac{\pi}{6} r_{ij} \right) \]  

(57)

The difference between \( \rho_{X_iX_j} \) and \( r_{ij} \) is small with a maximum deviation of only 0.018 as shown in Figure 15. Note that the mapping between \( \rho_{X_iX_j} \) and \( r_{ij} \) extends over the full range \([-1, 1]\). Again, it must be emphasized that Equation (57) only produces a valid correlation matrix if it is positive definite, i.e. the property of positive definiteness is an additional requirement. The counter-example shown in Section 4.2 is based on Equation (57). Although the prescribed fractile correlation matrix \( r_{ij} \) (left matrix) is positive definite, the equivalent Gaussian product-moment correlation matrix \( \rho_{X_1X_2} \) (right matrix) is not. It is of practical significance to note that positive definiteness could be violated by minor adjustments shown in Figure 15 (less than 0.02%).

Variations of this fractile correlation approach have been proposed in the wider literature beyond the domain of structural reliability (e.g., Iman and Conover 1982; Fackler 1991). It is very important to appreciate that the application of a multivariate Gaussian distribution to generate fractiles/ranks [Equation (56)] is non-unique. There are other possibilities and this is briefly discussed under a general framework of copula theory in Section 7. Iman and Conover (1982) observed that the use of the Gaussian distribution will produce "natural elliptical" correlation scatter plots while Fackler (1991) felt that it is a convenient choice among many other possibilities. The practical reason is that correlated samples could be simulated efficiently from uncorrelated samples using the Cholesky factorisation [Eq. (10)] only for elliptical distributions (Gaussian is the most well known).

![Figure 15. Difference between Gaussian product-moment correlation, \( \rho_{X_iX_j} \), and fractile correlation, \( r_{ij} \)](image-url)
Low and Tang (2004) assumed that $\rho_{Y_1Y_2} = \rho_{X_1X_2}$ in their proposed FORM solution for dependent non-Gaussian random vectors. It is clear from Figure 11 that this assumption could be poor for translation vectors. However, simulation results using @RISK (www.palisade.com) agree well with FORM results based on $\rho_{Y_1Y_2} = \rho_{X_1X_2}$. The theoretical reason is that @RISK produces dependent non-Gaussian realizations using the Gaussian-based fractile approach described above and the prescribed fractile correlations ($r_{ij}$) happen to be close to the equivalent Gaussian product-moment correlations ($\rho_{XiXj}$) as shown in Figure 15. It is too optimistic to expect such opportune theoretical results to be present in other possibilities. In summary, the FORM solution proposed by Low and Tang (2004) is only valid if the underlying data follow the Gaussian-based fractile model, rather than the Gaussian-based product-moment model (i.e., translation vectors). For comparison, STRUREL (1991) models dependent non-Gaussian vectors as translation vectors.

7 COPULA THEORY

Copula theory provides a completely general theory of constructing multivariate distributions from their one-dimensional marginals in a theoretically rigorous way (Schweizer 1991). Essentially, a copula function couples marginal distributions together in a nonparametric way so that the resulting multivariate distribution is consistent in the Kolmogorov sense (i.e., all possible lower order distributions are valid probability distribution functions) and the desired degrees of dependence between the components are achieved. Note that there are many possible measures of dependence, although the product-moment and fractile correlations are probably the most widely known.

For differentiable probability distribution functions, it is possible to prove mathematically that a non-Gaussian multivariate probability density function $f(y_1, \ldots, y_n)$ can always be expressed as the product of its marginals $\{f_1(y_1), \ldots, f_n(y_n)\}$ and a copula density function $c$ (Sklar’s theorem):

$$f(y_1, \ldots, y_n) = f_1(y_1) \times \ldots \times f_n(y_n) \ c[F_1(y_1), \ldots, F_n(y_n)] = c(v_1, \ldots, v_n)$$

(58)

in which $c = \frac{\partial^n C}{\partial F_1 \ldots \partial F_n}$ and $C =$ copula distribution function. Note that $c$ is a function of the fractiles $(v_1, \ldots, v_n)$. Hence, it is a multivariate probability density function whose marginals are standard uniforms defined between 0 and 1. The Gaussian-based fractile approach presented in Section 6 is a special case of Equation (58) with $c$ being the multivariate Gaussian copula density function (Phoon 2004). Another possibility is the Frank copula:

$$C_\delta (v_1, v_2) = -\frac{1}{\delta} \ln \left[ 1 + \frac{e^{-\delta v_1} - 1 \left( e^{-\delta v_2} - 1 \right)}{e^{-\delta} - 1} \right]$$

(59)
in which \( \delta \) = scalar controlling the degree of association. Note that the fractile correlation can be computed from any copula function using:

\[
\rho_{12} = 12 \int_0^1 \int_0^1 C(v_1, v_2) \, dv_1 \, dv_2 - 3
\]  

(60)

Hence, the \( \delta \) parameter in Frank copula can be related to the more physically meaningful fractile (rank) correlation. This relationship is shown in Table 9.

The bivariate probability density function \( f(y_1, y_2) \) with \( f_1(y_1) \) and \( f_2(y_2) = \log-Gaussian \) probability density function is different depending on the copula density function used (Fig. 16). Copula theory provides a larger family of dependent non-Gaussian probability models to choose from, which is very useful as existing Gaussian-based product-moment or fractile based models may not fit some observed data. In addition, it is possible to use the multivariate probability density function given by Equation (58) for Rosenblatt transform [Eq. (3)]. Research is on-going to examine if copula-based non-Gaussian random vectors could be applied to FORM efficiently.

8 CONCLUSIONS

This report presented a state-of-the art review on modelling dependent non-Gaussian random vectors for FORM. Because FORM solution requires non-linear optimization, proper scaling of the variables is important. In addition, correlation matrices are quite tricky to specify because they have to be positive definite to be admissible. These two separate issues can be resolved computationally in an efficient way by performing Cholesky factorisation. It must be appreciated that the construction of dependent non-Gaussian probability models using only correlation information is non-unique. The equivalent Gaussian technique widely used in FORM is easier to use when non-Gaussian random variables are coupled using a Gaussian-based fractile correlation approach. For the Gaussian-based product-moment approach (translation vectors), approximate probability transform based on Hermite polynomials is easier to use because a simple power series solution is available to relate the observed correlation with the underlying equivalent Gaussian correlation. If the observed data cannot be reasonably fitted to Gaussian-based product-moment/fractile models, it is necessary to consider more general copula-based models.
Figure 16. Two bivariate log-Gaussian probability density functions with the same fractile correlation = 0.7
Table 9. Relationship between Frank parameter and fractile correlation

<table>
<thead>
<tr>
<th>Frank parameter (δ)</th>
<th>Fractile correlation (r₁₂)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>0.60</td>
<td>0.1</td>
</tr>
<tr>
<td>1.22</td>
<td>0.2</td>
</tr>
<tr>
<td>1.88</td>
<td>0.3</td>
</tr>
<tr>
<td>2.61</td>
<td>0.4</td>
</tr>
<tr>
<td>3.44</td>
<td>0.5</td>
</tr>
<tr>
<td>4.47</td>
<td>0.6</td>
</tr>
<tr>
<td>5.82</td>
<td>0.7</td>
</tr>
<tr>
<td>7.92</td>
<td>0.8</td>
</tr>
<tr>
<td>12.3</td>
<td>0.9</td>
</tr>
<tr>
<td>18.3</td>
<td>0.95</td>
</tr>
</tbody>
</table>

9 REFERENCES


<table>
<thead>
<tr>
<th>Oppdragsgiver/Client</th>
<th>The Research Council of Norway</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kontraktsreferanse/Contract reference</td>
<td></td>
</tr>
<tr>
<td>Prosjektleder/Project Manager</td>
<td>Suzanne Lacasse</td>
</tr>
<tr>
<td>Utarbeidet av/Prepared by</td>
<td>Kok-Kwang Phoon</td>
</tr>
<tr>
<td>Dokument nr/Document No.</td>
<td>ICG Report 2004-2-4</td>
</tr>
<tr>
<td></td>
<td>NGI Report 20031091-4</td>
</tr>
<tr>
<td>Dato/Date</td>
<td>2004-09-30</td>
</tr>
<tr>
<td>Deklarasjon/Distribution</td>
<td>Fri/Unlimited</td>
</tr>
<tr>
<td></td>
<td>Begrenset/Limited</td>
</tr>
<tr>
<td></td>
<td>Ingen/None</td>
</tr>
<tr>
<td>Emneord/Keywords</td>
<td>Land, fylke/Country, County</td>
</tr>
<tr>
<td>Kommune/Municipality</td>
<td>Havområde/Offshore area</td>
</tr>
<tr>
<td>Sted/Location</td>
<td>Feltnavn/Field name</td>
</tr>
<tr>
<td>Kartblad/Map</td>
<td>Sted/Location</td>
</tr>
<tr>
<td>UTM-koordinater/UTM-coordinates</td>
<td>Felt, blokknr./Field, Block No.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Kvalitetssikring i henhold til/Quality assurance according to NS-EN ISO9001</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kontrollert av/Reviewed by</td>
</tr>
<tr>
<td>Helhetsvurdering/General Evaluation</td>
</tr>
<tr>
<td>Dato/Date</td>
</tr>
<tr>
<td>Språk/Style</td>
</tr>
<tr>
<td>Teknisk/Technical</td>
</tr>
<tr>
<td>- Skjønn/Intelligence</td>
</tr>
<tr>
<td>- Total/Extensive</td>
</tr>
<tr>
<td>- Tverrfaglig/Interdisciplinary</td>
</tr>
<tr>
<td>Utforming/Layout</td>
</tr>
<tr>
<td>Slutt/Final</td>
</tr>
<tr>
<td>Kopiering/Copy quality</td>
</tr>
</tbody>
</table>

* Gjennomlesning av hele rapporten og skjønnsmessig vurdering av innhold og presentasjonsform/On the basis of an overall evaluation of the report, its technical content and form of presentation

<table>
<thead>
<tr>
<th>Dokument godkjent for utsendelse/Document approved for release</th>
<th>Dato/Date</th>
<th>Sign.</th>
</tr>
</thead>
</table>

c:/documents and settings/cvepkk/my documents/sabbatical/project 2/icg-2004-2-4.doc
KKP/tha