

Calibrating and comparing reliability analysis procedures for slope stability problems

J. Zhang, W. H. Tang, L.M. Zhang

The Hong Kong University of Science and Technology, Hong Kong SAR, China

Y.R. Zheng

Chongqing Logistic Engineering University, Chongqing, China

ABSTRACT: Reliability analysis, in most cases, can only be implemented in an approximate way. This produces two subsequent questions: (1) how approximate is the approximate procedure is? (2) how the relative validity of different approximate procedures can be compared? In this paper, the practical difficulties for rigorous reliability analysis of Hong Kong slopes are addressed. Two types of approximate reliability analysis procedures for the analysis of Hong Kong slopes have been calibrated and compared using field observation data. It is found that the procedure which does not consider soil suction (P1) is highly biased towards the conservative side, while the procedures considering suction (P2-1 and P2-2) are biased towards the unsafe side. These approximate procedures are calibrated with field observation data to estimate the actual reliability index and the actual failure probability of the system. A comparison among different procedures shows that the calibrated P1 is the best procedure in terms of prediction accuracy. As discussed in this paper, the reliability analysis procedure with a more rigorous soil mechanics background may not always result in better results in terms of prediction accuracy. Whether a procedure with a more rigorous theoretical background could produce a more accurate prediction depends heavily on the quality and quantity of information available.

1 INTRODUCTION

Uncertainties are pervasive in geotechnical engineering. Common sources of uncertainties include: inherent variability of material properties, statistical uncertainty due to limited number of tests, measurement uncertainty, transformation uncertainty, uncertainty associated with the assumed probabilistic distributions and mechanistic model uncertainty. To quantify the effects of uncertainty in geotechnical engineering, reliability methods have been proposed to supplement deterministic approaches. Lacasse and Nadim(1998) gave some case studies where reliability analysis had been applied successfully in various areas of geotechnical engineering.

To fully utilize the power of reliability analysis, it is essential to properly characterize all important sources of uncertainties present in a geotechnical problem and to consider all these uncertainties explicitly in a reliability analysis. In practice, however, the above principle may not be easy to implement either due to limited information or due to limitations in our knowledge. For instance, the model uncertainty associated with a mechanistic model is often not considered appropriately in many reliability analyses because the characteristics of model uncertainty of many geotechnical models have not yet been well quantified. In many situations, in order to implement reliability analysis without heavily increasing the demand for additional information compared with a deterministic analysis, reliability analysis is often carried out in an approximate manner. With this realization, the following questions could be raised from practicing engineers:

- (1) How approximate is the approximate method? Is the approximation on the conservative side? Can the degree of approximation be quantified?

- (2) When several approximate reliability analysis procedures are available, how can one compare the relative plausibility of these approximate procedures?

Clearly, the ultimate test of a geotechnical system is how well it performs in reality. The degree of approximation associated with a reliability analysis procedure may be “back-figured” by systematically comparing the model predictions with the real performances of the geotechnical systems. The relative validity of various procedures can also be measured by comparing the predictions from different procedures with the actual performance data. The objective of this paper is to: (1) address the practical difficulties for rigorously implementing reliability methods in engineering, (2) demonstrate how the degree of approximation can be gauged, and (3) show how different approximate reliability procedures can be ranked utilizing an extensive set of filed performance data for man-made slopes in Hong Kong.

2 CHARACTERISTICS OF SLOPE FAILURES IN HONG KONG

The geology of Hong Kong has been well documented in the literature (e.g. Ruxton, 1960). The main rock types related to slope instability in Hong Kong are granitic and volcanic rocks, which have been severely weathered in situ. In general, the intensity of weathering decreases with depth. The granites can be weathered into soils up to 60m in some places, but is extremely variable both within and between sites. The volcanic rocks are less weathered compared with the granites, with thicknesses of weathered soils up to 20m. The in situ weathered soil which retains the original texture, fabric and structure of the parent rock is called “saprolite” locally. Closest to the ground surface, the grain structure of the soil may have collapsed. Such soil is called “residual soil” locally, and is seldom thicker than a few meters. Saprolite and residual soil are sometimes collectively known as residual materials (Au 1996). Man-made slopes in Hong Kong are mainly found with residual materials. For slopes with residual materials, failures often occur in the saprolite. In addition to the slopes in residual materials, however, there are also some slopes formed with colluvium, which is a material derived from the weathering of any parent rock which has been transported downhill by the agencies of gravity and water (Brand, 1995). Brand (1995) noted that coluvium poses many of the same general characteristics as saprolite, particularly in the context of engineering behavior. For geotechnical engineering purposes, colluvium can be grouped with saprolite. Therefore, in the context of this paper, discussion is mainly focused on saprolite. The completely decomposed granites (CDG) and completely decomposed vocanic rocks (CDV) are two typical saprolite materials in Hong Kong. The regional average cohesion and friction angle of CDG and CDV measured in triaxial compression tests are 3.9kPa, 39.8 degree, and 5.1kPa, 39.9 degree, respectively (Cheung, 2004). Mass permeability of the saprolite is in the order of $10^{-4} - 10^{-6} \text{ ms}^{-1}$ (Brand, 1985).

Due to a lack of flat areas, many man-made slopes have been formed at the lower portions of the hillslopes in many parts of the territory along with the intensive urbanization. Up to now, there are more than 55,000 registered slope features in Hong Kong. Hong Kong slope failures are closely related to rainfall infiltration. In the dry season, the groundwater table in the slope is usually close to the basal rock surface and far away from the slope surface (Lumb, 1975), and hence the slope mass is largely unsaturated. In the wet season, the strength of the soil is reduced by reduction in or loss of suction caused by rainfall infiltration. When the soil strength is decreased to a certain critical value, slope failure occurs. On the average about 200~300 slope failures are observed each year, most of which occur during or soon after periods of heavy rainfall. The association between rainfall and slope failures in Hong Kong has been noticed by a number of researchers which have been summarized by Finlay et al. (1996). Brand (1984) found that an hourly intensity of about 70mm/h seemed to be the threshold above which landslides occur, and that 24 h rainfall of less than 100mm is very unlikely to result in a major landslide incident. Premchitt (1991) noted that landslides are almost certain to occur in Hong Kong whenever the 24h rainfall exceeds 200mm.

3 CURRENT DESIGN PRACTICE

3.1 Challenges for slope stability analysis in Hong Kong

From the viewpoint of analysis and design, saprolite is undoubtedly is one of the “difficult” soils present in geotechnical design practice (Brand 1995). A number of methods have been developed to evaluate slope stability in Hong Kong, including: (1) statistical analysis of the correlation between slope failures and rainfall patterns (e.g., Lumb 1975, Brand 1984, Finlay et al. 1996), (2) terrain evaluation based on geomorphological mapping (e.g., Burnett and Styles 1982), (3) statistical analysis of the correlation between slope failures and slope geomorphology (e.g. Lee et al. 2001), (4) statistical analysis of the correlation between slope failures, slope age and rainfall pattern(e.g., Cheung 2004), and (5) site specific mechanistic slope stability analysis (e.g., GEO 2000). Many of these methods have been discussed in Brand (1985) and Brand (1995). Among these methods, the site specific mechanistic slope stability analysis method is deemed the most reliable method and is widely used for slope design in Hong Kong. Implementation of site specific mechanistic analysis of slope stability, however, is also not easy due to the following reasons:

- (1) The saprolite in Hong Kong is generally very heterogeneous. Material properties may vary widely over short distances which makes the variance of test results within any site of the same order as the variance between different sites (Lumb 1975). The inherent heterogeneity associated with the saprolite makes it difficult to sample and test (Brand 1995). The characterization of mass strength and mass permeability from small samples in this case is difficult and might be unreliable.
- (2) Rainfall infiltration plays a dominant role in slope failures in Hong Kong (Brand, 1995). Predicting pore water pressure distribution during rainfall infiltration is crucial for slope stability analysis. Although many approaches have been suggested in the literature to predict groundwater response based on rainfall infiltration analysis, the application of these methods is dependent on a large number of input parameters which are not usually available. The problem of predicting the suction distribution at the moment of extreme rainfall is further complicated by the spatial variations in flow properties and potential irregularities in the flow region.

3.2 Current design practice

Due to the above difficulties, slope stability analysis in Hong Kong can only be carried out in an approximate way. The guidance for slope stability analysis specified by GEO (1984), which reflects a compromise between practical difficulties and theoretical stringency, is generally followed in engineering practice in Hong Kong. Since the reliability analysis procedures studied in this paper are similar to the slope stability analysis procedure specified in GEO (1984), some key relevant issues such as test methods, pore pressure prediction methods, and slope stability analysis suggested by GEO (1984) are briefly reviewed as follows to set the stage for further discussion.

Test method: Despite the obvious objections to the measurement of shear strengths of saprolite materials by means of laboratory tests, these tests are recommended in GEO (1984) since they seem to be the most satisfactory means of establishing the likely range of shear strength in saprolite (Brand, 1995).

Suction: Although suction may not necessarily be completely eliminated by rainfall infiltration in the wet season, the prediction of long term suction is extremely difficult. Therefore, the contribution of suction to shear strength is not considered in slope stability analysis to avoid risky design.

Groundwater response to rainfall infiltration: The pore pressure distribution in a slope is dependent on both the pattern of rainfall and the hydrogeology of the slope (Brand, 1985). Two methods are suggested in GEO (1984) for the analysis of groundwater response under rainfall infiltration. The first method is the wetting band approach suggested by Lumb (1975), which is based on the limit analysis of rainfall infiltration into the horizontal surface of a porous medium. The thickness of the wetting band (Fig.1) is calculated as follows:

$$h = \frac{k_{sat}t}{n(S_f - S_o)} \quad (1)$$

in which h =thickness of the wetting brand; k_{sat} =saturated permeability of the soil, t= rainfall duration, S_f =final degree of saturation, and S_o =initial degree of saturation. It is then assumed that

when the wetting band descends and meets the groundwater table, the rise in the groundwater level is equal to the thickness of the wetting band (Koo, 1997). A rainfall with a return period of ten years is commonly employed in the wetting band analysis, and the calculated thickness of the wetting band is often about two meters.

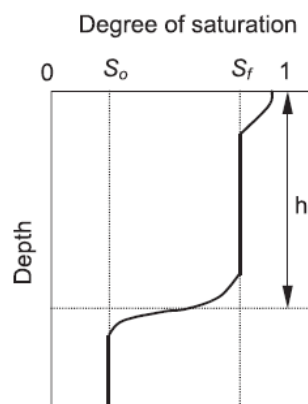


Fig.1 Illustration of the depth of wetting band

Another suggested method for predicting groundwater response with rainfall infiltration is by interpolating the field measured pore pressure data. This method can largely avoid the uncertainties involved in the analytical analysis of pore water pressure. To make this method meaningful, however, pore pressures must be monitored for a sufficiently long period of time, and piezometers must be installed at the appropriate depths at sufficient locations on the slope (Brand, 1985).

Slope stability analysis method: It is noted that slope failures in Hong Kong are often non-circular. In Hong Kong, the method of slices like Janbu (1954) and Morgenstern & Price (1965) which can treat non-circular slip surfaces is suggested to be used for slope stability analysis.

Recommended factor of safety (FOS): Slope design in Hong Kong is currently on a deterministic basis. The recommended FOS against loss of life and economic risk is shown in Table 1 and Table 2 for new and old slopes, respectively. The target FOS for an old slope is generally lower than that of a new slope in consideration to the “proof test” effect.

Table 1 Recommended FOS for new slopes for a ten-year return period rainfall

		Loss of life		
		Negligible	Low	High
Economic loss	Negligible	>1.0	1.2	1.4
	Low	1.2	1.2	1.4
	High	1.4	1.4	1.4

Table 2 Recommended FOS for old slopes for a ten-year return period rainfall

Loss of life	Negligible	Low	High
FOS	>1.0	1.2	1.4

4 RELIABILITY ANALYSIS NEGLECTING THE EFFECTS OF SUCTION

For better landslide management and communication with the public, the value of the risk management approach is increasingly appreciated by local landslide organizations in Hong Kong (Malone, 1997). Estimation of the failure probability of a slope is indispensable for calculating the risk associated with the slope. Among all the available methods, site specific reliability analysis, based on soil mechanics principles, seems to be the most reliable method for estimating the failure probability. Because the site specific mechanistic reliability analysis is based on deterministic slope stability analysis, it has succeeded many practical difficulties present in deterministic slope

stability analysis, as described in the previous section. For example, analysis of the groundwater response to rainfall infiltration in a reliability model is still challenging. Ideally, one can apply the wetting band approach in the reliability model supplemented with statistics on its model uncertainty. In practice, however, the uncertainty caused by the wetting band approach is rarely known. Therefore, the uncertainty caused by the wetting band approach is hard to quantify, and a rigorous site specific reliability analysis sounds appealing but may not be feasible. Similar to deterministic geotechnical analysis, site specific reliability analysis may only be implemented in an approximate manner with the assistance of certain assumptions.

With consideration of both theoretical stringency and practical difficulties, Cheung (2004) proposed two types of procedures for reliability analysis for Hong Kong slopes. The first procedure, which assumes that the soil suction will be fully eliminated during a rainstorm, is first introduced as follows. For ease of presentation, this procedure is called P1 (procedure 1) in this paper. In P1, the general principles for slope stability analysis suggested by the GEO (1984) are largely followed. The shear strength parameters of saturated soils and the groundwater regime are modeled as random variables. The variability of the shear strength parameters of saturated soils is determined by a Bayesian approach that combines regional experience with site-specific information (Zhang et al. 2004). The variability of a site-specific groundwater regime is estimated by a combination of field monitoring records and engineering judgment. The FOS of the slope is calculated using the extended Morgenstern and Price generalized method of slices (Morgenstern and Price, 1983). A spreadsheet method similar to the one developed by Low and Tang (1997) is used to implement the reliability analysis.

One may find the above procedure to be far from perfect. Many important uncertainties that deserve attention are not included in the analysis. These important uncertainties are:

- (1) Due to the inherent heterogeneity in Hong Kong soils, shear strength measured in the lab may differ significantly from the in situ mass strength that governs the slope stability. This uncertainty is not explicitly addressed in the reliability procedure.
- (2) The groundwater variability is judged by interpolating measured data supplemented with judgment. Since in general the period of groundwater monitoring may not be sufficiently, the inferred groundwater variability may differ a lot from the actual variability.
- (3) Suction may not always be fully eliminated during the rainfall infiltration. The uncertainty caused by the assumption of no suction is not addressed in the reliability procedure.
- (4) There are many assumptions associated with the extended Morgenstern-Price method, such as the soil is a Mohr-Coulomb material; the problem is for plain strain; the stress-strain relationship of the soil has no influence on the slope stability; and all slices have the same FOS. The model uncertainty of the extended Morgenstern-Price method is not addressed in the reliability procedure.
- (5) There is always a risk that the critical slip surface may not be detected. The uncertainty caused by this risk is not considered in this procedure.

Although the above imperfections can be easily identified, it seems difficult in practice to appropriately characterize and explicitly incorporate the above uncertainties into the reliability analysis in an economical way. Due to the presence of the above unquantified uncertainties, P1 is only an approximate reliability analysis procedure. Risk planners might be interested in the degree of approximation in P1. Since the ultimate test of a geotechnical system is how well it performs in the real world, a systematic way to identify the degree of approximation for a reliability analysis procedure is to compare the predictions from P1 with the observed performances for a sufficient large number of slopes. This process is called “calibration” in this paper, and the method for calibration is introduced in the following section.

5 CALIBRATING THE RELIABILITY ANALYSIS PROCEDURE

5.1 Methodology

Due to the approximate nature of P1, the calculated reliability index may be considered as notional instead of the actual reliability index of the system. However, the notional reliability index can be calibrated by the field observations to estimate the actual reliability index. Three commonly used relationships are listed in Table 3, where β_n denotes the notional reliability index

and β_{ci} s denotes the calibrated reliability index, which are estimates of the actual reliability index. In this paper, all these three relationships are considered possible to estimate the actual reliability index. The parameters of the three relationships are first determined using the likelihood method as described in the following paragraph, and the relationship that is most supported by the observed data is singled out for subsequent analysis. The other two relationships that are less supported by the observed data are abandoned and will be not discussed further. In such a case, the degree of approximation and modeling uncertainty in P1 may be gauged by the best relationship between β_n and β_c . The task of quantifying modeling uncertainty in a procedure is reduced to determining the parameters in the 3 relationships listed in Table 3 followed by a comparison among three relationships. To begin with, the method used in this paper to determine the parameters in the 3 relationships listed in Table 3 is introduced as follows.

Assume a database of slopes with known performance (failure or non-failure) is available. Also assume in the database there are m non-failure cases and n failure cases. Taking the relationship 1 ($\beta_{ci} = \beta_n + \varepsilon_\beta$) as an example, the parameter to be estimated is ε_β . A weighted likelihood function (Manski and Lerman 1977) of this relationship is:

$$L = w_{stable} \sum_{i=1}^m \ln [\Phi(\beta_{ni} + \varepsilon_\beta)] + w_{failed} \sum_{j=1}^n \ln [1 - \Phi(\beta_{nj} + \varepsilon_\beta)] \quad (2)$$

$$w_{stable} = \frac{1 - Q_p}{1 - Q_s} \text{ and } w_{failed} = \frac{Q_p}{Q_s} \quad (3)$$

where Φ =cumulative probability function of a standard normal variable, β_{ni} =notional reliability index of model i , w_{stable} , w_{failed} =adjustment factors for the unbalanced database, Q_p =proportion of failed slopes in the real world and Q_s = proportion of failed slopes in the database. It should be noted here that the adjustment factors are intended to account for the “choice-based sample bias” in the database, which arises when the portion of the failed slopes in the database (Q_s) is not equal to the portion of failed slopes in the real world (Q_p). After the likelihood function is derived, μ_ε (the best estimate of ε_β) can be determined by maximizing (3), and σ_ε (the standard deviation of ε_β) can be estimated from the second order derivative of the likelihood function at the maximum likelihood point. The parameters of other relationships in Table 3 can be determined in a similar fashion.

Table 3 Candidate calibration models

MODEL NO.	MODEL
1	$\beta_{c1} = \beta_n + \varepsilon_\beta$
2	$\beta_{c2} = N_\beta \beta_n$
3	$\beta_{c3} = N_\beta \beta_n + \varepsilon_\beta$

5.2 Calibration results

To calibrate P1, 145 local slopes (55 failed and 90 not failed), where available information is adequate for a site-specific reliability analysis, are collected and the reliability index of these slopes are calculated (Cheung 2004). In this database, the portion of failed slopes is 0.379 ($Q_s=0.379$). Based on the field record during 1984-2000 from the registered slopes, only 6.9% ($Q_p=0.069$) of slopes have failed, therefore, in the database the effect of failed slopes is overemphasized. The bias in the database can be appreciated by the adjusted weighted likelihood function as described in Eq.(2).

With the above database, the parameters of the three relationships listed in Table 3 are determined. These parameters are then substituted into the three relationships to relate β_n with β_{ci} . Fig.2 plots the best estimated relationships between β_n and β_{ci} . It can be seen that in general β_{ci} is underestimated by β_n . The predictions from the various relationships are seen to be quite different.

To see which relationship is better supported by the observed data, the three relationships are compared by using the Bayesian information criterion (BIC) (Schwarz 1978). The BIC of relationships 1, 2, and 3 are 53.67, 83.91 and 58.24, respectively. Since a smaller BIC indicates a better fit, relationship 1 appears to be the best one to estimate the actual reliability index. Therefore the following discussions are based on the calibrated model 1.

5.3 Characteristics of uncertainty in notional reliability index

Relationship 1 suggests that the uncertainty in the reliability index can be represented by the random variable ε_β . In other words, ε_β can characterize the uncertainties caused by the assumptions made in P1, and it can be considered as a factor indicating the modeling uncertainty associated with P1. ε_β is also found to be asymptotically, normally distributed with a mean of 1.031 and a standard deviation of 0.204. Fig.3 plots β_n , mean and 95% confidence intervals of β_{cl} . In Fig.3, even the lower bound of the 95% confidence interval line of β_{cl} is significantly higher than β_n , so P1 is highly biased towards the conservative side.

5.4 Possible causes of conservatism

One possible assumption responsible for the conservatism might be that suction would be fully eliminated during a severe rainfall and that the long-term stability of a slope cannot rely upon suction. This assumption may be contentious in nature. After a comprehensive review of the field suction measurements at different sites by different researchers in Hong Kong, Shen (1997) noted that suction could exist in the slope during or after a severe rainstorm. Zhang et al. (2004) theoretically showed that long term suction may exist in the soil, especially when the slope is protected by the surface cover. In the absence of surface cover, if the surface soil has relatively lower saturated permeability, it can also act as a surface cover to maintain suction (Zhang et al. 2004). It is also shown that soil suction may even exist in the slope even during and after a prolonged rainfall. Cheung (2004) noted that the presence and magnitude of soil suction are highly site-dependent. The ability to retain suction depends on both the soil type and the degree of slope surface protection. In general, soil suction in CDV soils is higher than that in CDG soils and colluvium. Rigid surface covers such as chunam(soil-cement-lime cover plaster) have larger potential to maintain soil suction than soft cover like vegetated cover. Therefore, it seems rather conservative to universally assume that suction does not contribute to the long-time stability.

Another possible reason is related to the curved strength envelope at low effective stresses in Hong Kong residual soils. Hong Kong slope failures are typically shallow, and the effective stresses at failure are low. When laboratory tests are carried out in the range of high effective stress, the actual strength of the soil at the low stress range is likely to be underestimated by the straightline projection of strengths measured at higher effective stresses (Brand, 1995). Such an underestimation in strength may also introduce conservatism into the reliability analysis.

Besides the above mentioned factors, the assumptions associated with the method of slices, the assumptions on the groundwater regime for stability analysis may also contribute to the conservatism. In addition, the overall conservatism in the above reliability procedure does not mean all the assumptions involved are conservative. For example, the potential inability to detect the most critical slip surface may lead to non-conservative predictions. Therefore the overall conservatism in the above reliability procedure merely implies that the effects of conservative assumptions have overtaken those from the unsafe assumptions.

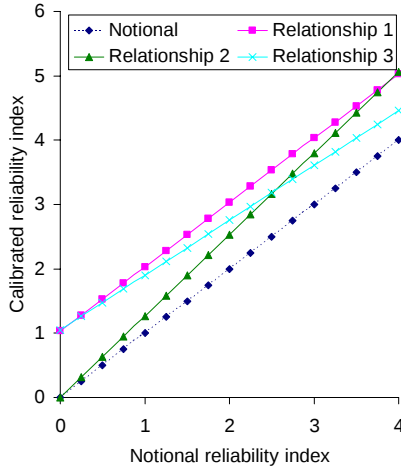


Fig.2 Comparison of best estimated actual reliability index from different calibration models

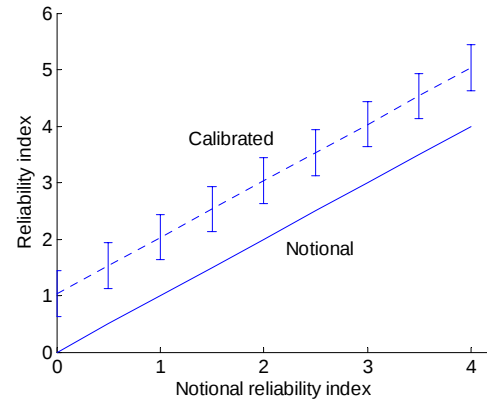


Fig.3 Comparison of notional and calibrated reliability index

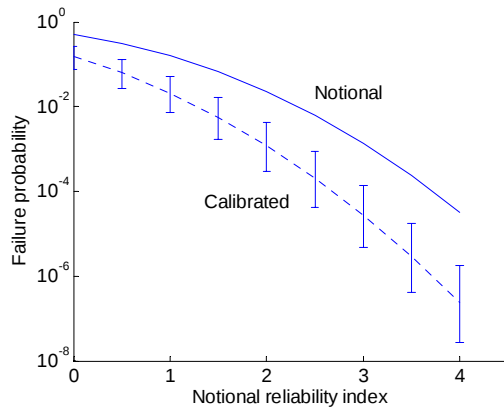


Fig.4 Comparison of notional and calibrated failure probability

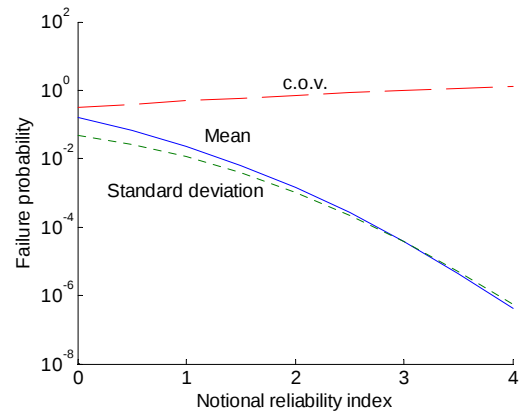


Fig.5 Mean, standard deviation and c.o.v. of the calibrated failure probability

5.5 Uncertainty in notional failure probability

Uncertainty in notional reliability will also propagate to failure probability. Since the actual reliability index is uncertainty, the actual failure probability is also uncertain. Fig.4 plots the notional failure probability with the median and 95% confidence interval of the calibrated failure probability using Monte-Carlo simulation. This figure shows that the calibrated failure probability is significantly lower than the notional failure probability. The amount of uncertainty in notional failure probability varies with calculated reliability index. This phenomenon is more obvious in Fig.5, which shows the mean, standard deviation and coefficient of variance (c.o.v.) of the actual failure probability against the notional reliability index. In Fig.5, the slope of the standard deviation line is negative, indicating the absolute uncertainty in the actual failure probability decreases as the notional reliability index increases. The slope of the c.o.v. line is positive, suggesting the relative uncertainty in the actual reliability index increases as the notional reliability index increases. The above phenomena suggest that the failure probability is more sensitive to modeling uncertainty, and that it is relatively less accurate in calculating small failure probabilities.

6 COMPARISON OF DIFFERENT RELIABILITY ANALYSIS PROCEDURES

6.1 Reliability analysis procedures considering effects of suction

As mentioned above, soil suction plays a very important role in slope stability problems in Hong Kong. It is shown in the above sections that the reliability procedure which does not consider the beneficial effects of suction is highly biased towards the conservative side. One may raise a pertinent question: can we improve our reliability analysis by considering the effect of suction? This problem is not easy since the magnitude of the suction is highly variable, both spatially and temporally, and site-specific suction information is rarely available. Nonetheless, it would be useful to investigate the effects of including soil suction in the shear strength of soils on the prediction of a reliability analysis procedure with assumed statistical parameters for soil suction. The reliability procedure that considers the effect of suction is called P2 in this paper. In P2, the shear strength model used is as follows:

$$\tau = c' + (u_a - u_w) \tan \phi^b + (\sigma_n - u_a) \tan \phi' \quad (5)$$

where c' =effective cohesion, u_a = pore-air pressure, u_w = pore water pressure, σ_n =net stress, ϕ^b =friction angle with respect to suction, ϕ' =effective friction angle.

Local field measurements of soil suction reveal that soil suction is of the order of tens of kPa during rainy seasons, and that CDV soils have a greater ability to maintain suction than CDG soils and colluvium (Cheung, 2004). With this knowledge, the site-specific suction is determined with the following assumptions: mean suction in CDG and colluvium slopes =15 kPa, mean suction in CDV=30 kPa, c.o.v. for soil suction =0.2, and the friction angle with respect to soil suction $\phi^b = \phi'/2$ with variability equal to that of ϕ' .

In the above section, the c.o.v. of the suction is assumed to be 0.2. Since suction is highly variable and a rough judgment may not be reliable, such a small c.o.v. may give doubts that suggested judgmental overconfidence. In such a case, P2 is also assessed assuming that the c.o.v. of suction is equal to 0.5. To distinguish P2 with different assumptions on suction, the analysis with c.o.v. =0.2 and c.o.v.=0.5 on suction are called P2-1 and P2-2 respectively.

6.2 Calibration and comparison of different reliability analysis procedures

To characterize the modeling uncertainty associated with P2-1 and P2-2, the reliability indices of the 145 slopes are first recalculated. For each case of P2-1 and P2-2, the parameters of the three relationships in Table 3 are first determined using the likelihood method, and then the best relationship is selected using BIC as described above. Similar to the situation in P1, it appears that relationship 1 is the best for describing modeling uncertainty in both P2-1 and P2-2. The determined parameters associated with relationship 1 of P1, P2-1 and P2-2 are shown in Table 4 for comparative purpose. It should be noted that μ_ε can be a measure to indicate the average bias associated with each procedure. A positive sign for μ_ε shows that the procedure is biased on the conservative side and the reverse is also true. The absolute value of μ_ε shows the degree of bias of the procedure. Table 4 shows that while the modeling uncertainty in P1 is on the conservative side, the modeling uncertainties in P2-1 and P2-2 are on the unsafe side. Therefore the magnitude of suction might have been overestimated in P2-1 and P2-2. The absolute values of the mean modeling uncertainty factor (μ_ε) in P2-1 and P2-2 are smaller than those in P1, so on average the predictions from P2-1 and P2-2 are less biased than those based on P1.

While μ_ε indicates the average bias associated with each procedure, the c.o.v. of ε_β characterizes the degree of homogeneity of modeling uncertainty of the procedure. Here the term “homogeneity” represents the level of similarity of modeling uncertainty when a procedure is applied to different slopes. Table 4 shows that while P1 is the most biased, its modeling uncertainty is the most homogenous. The modeling uncertainty associated with P2-2 is more homogenous than that for P2-1.

Table 4 Comparison of calibration results from P1, P2-1 and P2-2

	P1	P2-1	P2-2
	1.0313	-0.7249	-0.7699
c.o.v.	0.1981	0.3104	0.2791

Among the three reliability analysis procedures, which one is best in terms of prediction accuracy? Ideally, a preferable procedure is the one with more homogeneous and less biased modeling uncertainty. However, as shown in Table 4, the least biased procedure (P2-1) may not be the one with most homogeneous modeling uncertainty. In such a case, the relative validity of these procedures can be compared in terms of BIC, as described above. To see the effects of calibration, the BIC of three procedures, with and without calibration are all calculated and listed in Fig.5. The BIC of a procedure without calibration is calculated based on β_n , and the BIC of a procedure with calibration is calculated based on the best estimate of β_{cl} .

Table 5 BIC of different procedures

	P1	P2-1	P2-2
Without calibration	81.11	78.20	70.04
With calibration	53.67	73.78	63.69

Several phenomena are noticed. Firstly, for each procedure, the BIC has been reduced significantly after calibration. This implies that the calibrated procedures can provide better prediction than those without calibration. Secondly, before calibration, P2-2 is the best procedure. Note both the bias and the level of homogeneity of modeling uncertainty associated with P2-2 between those of P1 and P2-1. The fact that P2-2 appears to be the best model may suggest that in P2-2 a balance between bias and homogeneity is reached. Thirdly, after calibration, it is interesting to note that the calibrated P1 appears to be the best model among the three calibrated procedures. This is not surprising. Note the calibrated procedures are compared based on the best estimates of β_{cl} , in which the average bias in each procedure has been appreciated. In such a case, the relative validity of the different calibrated procedures is only related to the level of homogeneity of the modeling uncertainty. Since P1 has the least c.o.v. in modeling uncertainty, it is understandable that after calibration P1 is the best procedure. This may indicate that the assumptions on suction made in both P2-1 and P2-2 seem to be inappropriate. A certain amount of inhomogeneous random uncertainties may have been introduced into P2-1 and P2-2. To improve the prediction from the unsaturated procedures, it is highly desirable to obtain more reliable site-specific suction information. It is also noted that P2-1 (the c.o.v. of suction = 0.2) is inferior to P2-2 (the c.o.v. of suction = 0.5). This may suggest that the assumption on suction made in P2-1 is more biased than that made in P2-2.

7 GEOTECHNICAL IMPLICATIONS

In practice, different approximate reliability analysis procedures may be available for the same problem. Because of the different approximations involved in different procedures, the calculated reliability index may be either on the conservative or on the unconservative side. The same notional reliability index from different procedures may not imply the same level of safety. To avoid too risky or over conservative design, it is essential to calibrate the procedure with real observation data.

The above discussion also shows that a reliability procedure with a more theoretically rigorous background may not always produce more reliable predictions. For the P2-1 and P2-2 considered in this paper, the soil suction adopted in the analysis is subjected to a huge amount of uncertainty, which has greatly undermined the prediction accuracy of the unsaturated analysis procedures. However, it is believed that P2 based on the unsaturated model can finally overtake P1 when adequate knowledge on soil suction is available. Based on the above discussion, it seems that whether a procedure based on more rigorous theory can produce a better prediction depends heavily on the information available at hand.

8 CONCLUSIONS

The following conclusion can be drawn from this study:

- (1) Due to limitation in information and knowledge, reliability analysis of Hong Kong slopes can only be implemented approximately. The degree of approximation and the amount of modeling uncertainty of different procedures can be quantified by comparing model predictions with field observations. Such comparison can be implemented systematically using the weighted likelihood method.
- (2) Two types of approximate reliability procedures have been considered in this paper. P1 is the procedure which does not consider soil suction, whereas P2-1 and P2-2 are two procedures considering soil suction. P2-1 and P2-2 differ with each other in the assumptions on suction variability. The c.o.v. for suction adopted in P2-1 and P2-2 are 0.2 and 0.5, respectively.
- (3) The modeling uncertainty of each of these procedures can be characterized by a random variable ε_β . The statistics of ε_β of each procedure can be determined by the weighted likelihood method. While μ_ε reflects the average bias of the modeling uncertainty, the c.o.v. of ε_β shows the level of homogeneity of the modeling uncertainty associated with the procedure.
- (4) As indicated by μ_ε , while P1 (a procedure without considering soil suction) is highly biased towards the conservative side, P2-1 and P2-2 (procedures considering soil suction) are biased towards the unsafe side. TP2-1 and P2-2 are less biased than P1 on the average. The modeling uncertainty in P1 is more homogeneous since it has the least c.o.v. value of ε_β .
- (5) The calculation of failure probability is very sensitive to modeling uncertainty. It's less accurate in calculating smaller failure probabilities.
- (6) BIC can be used to compare the relative validity of different reliability analysis procedures. For cases which have not been calibrated, P2-2 is the best procedure in the sense of prediction accuracy among the three uncalibrated procedures. After being calibrated, P1 can produce the most accurate prediction. No matter whether a case has been calibrated or not being calibrated, P2-1 is always inferior to P2-2. This might suggest that the suction assumption in P2-2 is more realistic.
- (7) When several approximate procedures are available for the same problem, the same notional reliability index from different procedures does not imply the same amount of safety.
- (8) Reliability analysis procedures based on more rigorous soil mechanics models may not always result in better prediction in terms of prediction accuracy. Whether a procedure with a more theoretically rigorous background could produce a more accurate prediction depends heavily on the quality and quantity of information available.

ACKNOWLEDGMENTS

This research was substantially supported by a grant from the NSFC/RGC joint research scheme (Project N_HKUST611/03) and a grant from the RGC (Project HKUST629404E).

REFERENCES

- Brand, E. W. (1984) *Landslides in Southeast Asia: a state-of-the-art report. Proc. 4th Int. Symp. on Landslides, Toronto. 1, p17-59. Bitech, Vancouver.*
- Brand, E.W. (1985) *Predicting the performance of residual soil slopes. Proc. 11th Int. Conf. on Soil Mech. Found. Engng. San. Francisco. p2541–2578. Balkema, Rotterdam.*
- Brand, E.W.(1995) *Keynote Paper: Slope instability in tropical areas. Proc. 6th Int. Symp. on Landslides. Christchurch, New Zealand. vol. 3, p2031-2051. Balkema, Rotterdam.*

Burnett, A.D. and K.A. Styles. (1982) *An approach to urban engineering geological mapping as used in Hong Kong*. Proc. 4th Int. Cong. Int. Assoc. Engng. Geol., New Delhi. p167-176. Oxford & IBH, New Delhi.

Cheung R.W.M. (2004) *Methodology for updating cut slope reliability based on observed performance*, PhD thesis, The Hong Kong Univ. of Science and Technology, Hong Kong.

Finlay, P. J., Fell, R., and Maguire, P. K. (1997) *The relationship between the probability of landslide occurrence and rainfall*. Can. Geotech. J., 34, p811-824.

GEO (1984) *Geotechnical Manual for Slopes*, 2nd Edition (Reprinted, 2000).

Koo, Y.C. (1997) *Some important aspects of design of soil slopes in Hong Kong*, Slope engineering in Hong Kong, edited by K. S. Li, J. N. Kay, and K. K. S. Ho, p21-29. Balkema, Rotterdam.

Janbu, N. (1954) *Application of composite slip surface for stability analysis*. Eur. Conf. on Stability of Earth Slopes. Sweden. vol. 3, p43-49. Stockholm, Sweden.

Lacasse, S., and Nadim, F. (1998) *Risk and reliability in geotechnical engineering*. Proc. 4th int. conf. on Case Histories in Geotechnical Engineering. St. Louis. p1172-1192.

Low, B.K., and Tang, W.H. (1997) *Probabilistic slope analysis using Janbu's generalized procedure of slices*. Comput. Geotech., 21(2), p121-142.

Lumb, P. (1975) *Slope failures in Hong Kong*. Q. J. Engng. Geol., 8, p21-65.

Malone, A. W. (1997) *Risk management and slope safety in Hong Kong*. Slope engineering in Hong Kong, edited by K. S. Li, J. N. Kay, and K. K. S. Ho, p21-29. Balkema, Rotterdam.

Manski C.F. and Lerman S.R. (1977) *The estimation of choice probabilities from choice-based samples*. Econometrica, 45(8), p1977-1988.

Morgenstern, N.R. and Price, V.E. (1965) *The analysis of the stability of general slip surfaces*. Geotechnique, 1965,15(1):p79-93.

Premchitt, J. (1991) *Salient aspects of landslides in Hong Kong*. Proc. 9th Asian Regional Conference on Soil Mechanics and Foundation Engineering, Bangkok, vol. 2, p497-502.

Ruxton, B.P.(1960) *The geology of Hong Kong*. Q.J. Geol. Soc., 115, p233-260.

Shen J.M. (1997) *Soil suction in relation to slope stability: A summary of research carried out in Hong Kong in 1978-1997*, Slope engineering in Hong Kong, edited by K. S. Li, J. N. Kay, and K. K. S. Ho, p21-29. Balkema, Rotterdam.

Zhang, L.L., Fredlund, D.G., Zhang L.M., and Tang W.H. (2004) *Numerical study of soil conditions under which matrix suction can be maintained*. Can. Geotech. J., 41(4):p569-582.

Zhang, L.M., Tang, W.H., Zhang, L.L., and Zheng, J.G. (2004). *Reducing uncertainty of prediction from empirical correlations*. J. Geotech. Geoenviron. Engrg., ASCE, 130(5), p526-534.