

PRAVDĚPODOBNOSTNÍ ANALÝZA METODOU LATIN HYPERCUBE SAMPLING

PROBABILISTIC ANALYSIS USING LATIN HYPERCUBE SAMPLING METHOD

TOMÁŠ SVOBODA, MATOUŠ HILAR

1 ÚVOD

Mechanické parametry horninového masivu získané v rámci geotechnického průzkumu pro návrh podzemních staveb (případně i jiných konstrukcí) mají často značný rozptyl. Daná skutečnost vyplývá nejen z vlastností geologického prostředí, které zpravidla není homogenní, ale je zapříčiněna také nepřesnostmi prováděných laboratorních či polních zkoušek.

Nejistoty ve vstupních parametrech jsou v deterministických výpočtech v geotechnické praxi zohledněny pomocí bezpečnostních koeficientů. Ty do výpočtů podle Eurokódu 7 (EC7 – ČSN EN 1997-1) vstupují ve formě dílčích součinitelů aplikovaných na materiálové parametry, samotná zatížení, jejich účinky, nebo případně oboje. Při standardním výpočtu mohou být výsledky značně konzervativní, model s redukovanými vstupními parametry se může výrazně odlišovat od skutečného chování. V normě EN 1990:2002 je povoleno použití pravděpodobnostních metod, které zohledňují variabilitu vstupních parametrů a zatěžovacích stavů. Pravděpodobnostní výpočet je formulován jako alternativa, která má být ověřena standardním výpočtem s využitím dílčích součinitelů. Pomocí pravděpodobnostních metod lze určit pravděpodobnost poruchy a index spolehlivosti, jejichž minimální hodnoty pro mezní stav únosnosti a jednotlivé třídy spolehlivosti jsou doporučeny v EC 7. Výsledkem pravděpodobnostního přístupu pak je nejen zda konstrukce vyhoví, či nikoli, ale i stanovení míry rizika porušení souvisejícího s navrženou konstrukcí.

Pravděpodobnostní výpočty se stávají stále dostupnějším nástrojem pro řešení geotechnických úloh. Jejich rozšíření v běžné praxi brání zejména vyšší časové nároky na zpracování výpočtů, nároky na výstupy z geotechnického průzkumu, dále také chybějící implementace pravděpodobnostních metod v běžně používaném programovém vybavení. Následující příspěvek se proto zabývá redukční pravděpodobnostní metodou Latinských hyperkrychlí (Latin Hypercube Sampling), která je alternativou k časově náročné simulační metodě Monte Carlo. Obsahem článku je především popis algoritmu metody, její vývoj, současný stav poznání a aplikace v české geotechnice praxi.

2 NÁHODNÉ PROMĚNNÉ

Výsledky numerického modelování geotechnických úloh jsou velmi citlivé na vstupní parametry (v souvislosti s pravděpodobnostními analýzami lze hovořit o proměnných). Nejistota spojená se stanovením parametrů vede k nejistotě určení výsledku celého systému – řešeného problému. Pro jejich vyjádření je vhodné brát v úvahu jejich charakter náhodných veličin. Náhodná veličina nabývá různých hodnot a je charakterizována rozdělením hustoty pravděpodobnosti. Z praktického hlediska studie dále uvažuje pouze spojité náhodné veličiny, které mohou nabývat všech hodnot z daného intervalu. Náhodné vstupní veličiny jsou u pravděpodobnostních výpočtů reprezentovány sadou deterministických čísel (tzv. realizací, vzorků), která jako celek

1 INTRODUCTION

Mechanical parameters of rock mass obtained for the purpose of designing underground structures (or other structures) within the framework of geotechnical investigation often display significant scatter. This fact follows not only from the properties of the geological environment, which usually is not homogeneous, but it is also caused by inaccuracies in the executed laboratory of field tests.

Uncertainties in input parameters are allowed for in deterministic calculations in the geotechnical practice by means of safety coefficients. They enter the calculations to Eurocode 7 (EC7 – ČSN EN 1997-1) in the form of partial coefficients applied to material parameters, loads themselves, their effects, or both of them. In a standard model calculation the results may be significantly conservative; a model with reduced input parameters may substantially differ from the real behaviour. The EN 1990:2002 standard permits the use of probabilistic methods, which take into consideration the variability of input parameters and loading cases. A probabilistic calculation is formulated as an alternative, which should be verified by a standard calculation using partial coefficients. By using probabilistic methods it is possible to determine the probability of a defect and the index of reliability, the minimum values of which for the ultimate limit state and individual reliability classes are determined in the EC 7. The result of the probabilistic approach is then not only the information whether the structure will or will not satisfy requirements, but also the determination level of the risk of a failure connected with the designed structure.

Probabilistic calculations have been becoming an ever more attainable tool for solving geotechnical problems. Their spreading in the common praxis is prevented first of all by the higher consumption of time for executing the calculations, requirements for outputs from geotechnical investigation and, in addition, due to the absence of the implementation of probabilistic methods in a commonly used software. The following paper therefore deals with the Latin Hypercube Sampling reduction probabilistic method, which is an alternative to the time-intensive Monte Carlo simulation method. The paper content comprises first of all the description of the method algorithm, its development, current state of knowledge and application within the Czech geotechnical practice.

2 RANDOM VARIABLES

Results of numerical modelling of geotechnical problems are very sensitive to input parameters (it is possible in the context of probabilistic analyses to speak about variables). The uncertainty associated with the determination of parameters leads to an uncertainty in the determination of results of the entire system – the problem to be solved. For their formulation it is advisable to take into consideration their character of random quantities. A random quantity assumes various values and is characterised by the probability density distribution. From the practical point of view, the study further takes into consideration only continuous random

tvoří jednu z deterministických úloh, ze kterých se skládá řešený náhodný problém. Vstupní sady parametrů jsou tak použity pro získání odezvy (např. stupeň stability) geomechanického modelu, u níž hledáme statistické a pravděpodobnostní charakteristiky.

Spojité náhodné veličiny X je charakterizována funkcí hustoty pravděpodobnosti $f(x)$ a kumulativní distribuční funkcí $F(x)$ (1), pro něž platí:

$$F(x) = \int_{-\infty}^{+\infty} f(x) dx \quad (1)$$

Polohu hodnot náhodné veličiny X nejlépe vystihuje střední hodnota značená EX (též μ), variabilitu (varianci, rozptýlenost) dat vyjadřuje rozptyl $Var X$, též σ_X^2 (2):

$$EX = \int_{-\infty}^{+\infty} xf(x) dx \quad Var X = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx \quad (2)$$

Odmocnina z rozptylu je nazývána směrodatnou odchylkou ($\sigma_X = \sqrt{\sigma_X^2}$). Dalšími charakteristikami (třetím a čtvrtým momentem) náhodné veličiny vhodnými pro určení asymetrie pravděpodobnosti rozdělení jsou šikmost a špičatost.

2.1 Závislost náhodných proměnných

Důležitou vlastností náhodných proměnných je stupeň jejich vzájemné závislosti. Ta je charakterizována kovariancí $cov(X_i, X_j)$, respektive koeficientem korelace $corr(X_i, X_j)$ (3), který se na rozdíl od kovariance nemění při lineární transformaci proměnných. Korelační koeficient nabývá hodnot od -1 do 1 .

$$cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) \quad corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} \quad (3)$$

kde $i, j = 1, \dots, n$

Při prostorovém popisu mechanických vlastností horninového masivu se předpokládá, že hodnoty dané proměnné ve dvou vzájemně blízkých bodech budou podobné. Naopak ve dvou vzdálených bodech se již výrazně liší a parametry jsou téměř nekorelované. Změnu korelačního koeficientu vzhledem ke vzdálenosti dvou bodů v prostoru popisuje korelační funkce $\rho(X_i, X_j)$. Z uvedených vlastností vyplývá, že s rostoucí vzdáleností bodů hodnota korelačního koeficientu klesá. Závislost jeho poklesu je obvykle popisována exponenciálním tvarem korelační funkce definované polohou bodů x_i, x_j náhodných proměnných X_i, X_j a korelační délkou.

2.2 Pravděpodobnostní rozdělení

Pro statistické zpracování variabilních vstupních parametrů je nejprve nutné tyto parametry popsat pravděpodobnostním rozdělením. Existuje řada metod k tomu určených, například χ^2 (chi kvadrát) test, Kolmogorov–Smirnov test a další, případně je vhodné využít statistických analýz implementovaných do uživatelských programů (např. QC Expert). Pokud je k dispozici málo vstupních dat, či známe pouze interval hodnot, ve kterých se parametry pohybují, je obtížné stanovit odpovídající rozdělení.

Převážná většina studií a prací zabývajících se pravděpodobnostní analýzou geotechnických problémů (např. Hamm a kol. 2006, Flores, 2010) poukazuje na skutečnost, že parametry horninového (zeminného) masivu je vhodné popsat normálním, či log-normálním rozdělením. Výjimkou je použití Weibullova rozdělení, případně exponenciálního rozdělení, a to především z důvodu existence jednoduchého řešení distribuční funkce.

3 METODA LATIN HYPERCUBE SAMPLING

Metoda Latin Hypercube Sampling (LHS) je numerická simulační metoda typu Monte Carlo (MC). Metoda Monte Carlo (obr. 1) je běžně používána pro řešení náhodných problémů, u kterých požadujeme statistické a pravděpodobnostní informace odezvy problému. K jejich přesnému odhadu však metoda

quantities, which can assume all values from a particular interval. Random input quantities are represented in probabilistic calculations by a set of deterministic figures (the so-called realisations, samples), which as a whole forms one of deterministic problems which the random problem being solved consists of. The input sets of parameters are subsequently used for obtaining the response (e.g. the degree of stability) of the geomechanical model at which we seek statistical and probabilistic characteristics.

A continuous random quantity X is characterised by the function of probability density $f(x)$ and the cumulative distribution function $F(x)$ (1), for which it applies:

$$F(x) = \int_{-\infty}^{+\infty} f(x) dx \quad (1)$$

The position of the values of random quantity X is best fitted by the mean value denoted as EX (also μ); data variability (variance, dissemination) is expressed by scatter 'Var' X , also σ_X^2 (2):

$$EX = \int_{-\infty}^{+\infty} xf(x) dx \quad Var X = \int_{-\infty}^{+\infty} (x - EX)^2 f(x) dx \quad (2)$$

The square root of scatter is titled as standard deviation ($\sigma_X = \sqrt{\sigma_X^2}$). Other characteristics (third and fourth moment) of the random quantities suitable for the determination of the probability distribution asymmetry are skewness and kurtosis.

2.1 Dependence of random variables

An important property of random variables is the degree of their interdependence. The interdependence is characterised by covariance $cov(X_i, X_j)$, or the correlation coefficient $corr(X_i, X_j)$ (3), which, in contrast with covariance, does not change when variables are being linearly transformed. The correlation coefficient assumes values from -1 to 1 .

$$cov(X_i, X_j) = E(X_i X_j) - E(X_i)E(X_j) \quad corr(X_i, X_j) = \frac{cov(X_i, X_j)}{\sigma_{X_i} \sigma_{X_j}} \quad (3)$$

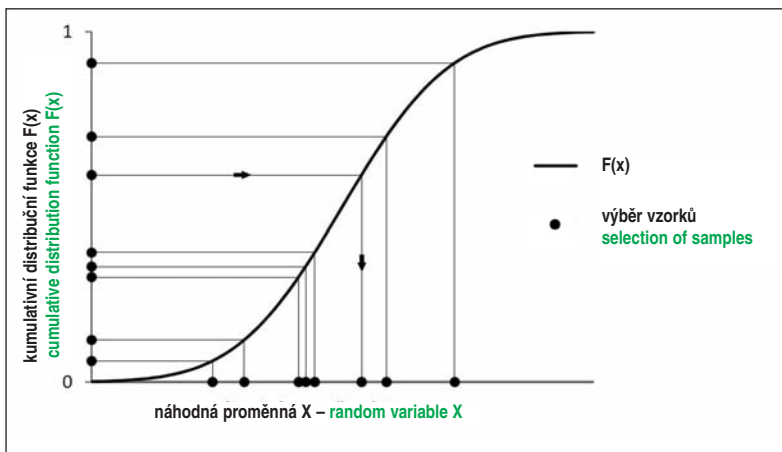
where $i, j = 1, \dots, n$

In the case of a spatial description of mechanical properties of rock mass it is assumed that the values of the particular variable in two points found close to each other will be similar. On the contrary, in two remote points the values already significantly differ and the parameters are nearly uncorrelated. The change in the correlation coefficient with regard to the distance between the two points in space is described by the correlation function $\rho(X_i, X_j)$. It follows from the above-mentioned properties that the correlation coefficient value drops with the growing distance between the points. The dependence of the drop is usually described by means of the exponential shape of the correlation function defined by the positions of points x_i, x_j of random variables X_i, X_j and the correlation length.

2.2 Probability distribution

It is necessary for the statistic processing of variable input parameters to start with the description of the probability distribution. There are several methods available which are designed to this purpose, for example the χ^2 (chi-squared) test, Kolmogorov - Smirnov test and other, or it is advisable to use statistical analyses implemented into user programs (e.g. QC Expert). If the amount of input data is insufficient or we know only the interval of the values within which the parameters vary, it is difficult to determine the relevant distribution.

The overwhelming majority of studies and works dealing with the probabilistic analysis of geotechnical problems (e.g. Hamm et al. 2006, Flores 2010) point out the fact that it is recommendable to describe ground (soil) mass parameters by a normal or log-normal distribution. The use of Weibull distribution or exponential distribution is an exception, first of all because of the existence of a simple solution to the distribution function.



Obr. 1 Princip výběru vzorků metody Monte Carlo
Fig. 1 The principle of the Monte Carlo method sampling

Monte Carlo vyžaduje obvykle velmi mnoho pokusů (běhů programu, simulací), aby bylo dosaženo požadované chyby. Z důvodu snížení počtu simulací a z nich vyplývající značné časové náročnosti byla vyvinuta řada redukčních metod.

První zavedení metody Latinských hyperkrychlí souvisí s řešením a zpracováním nejistot v analýzách bezpečnosti nukleárních elektráren ve Spojených státech amerických. Metoda byla prvně publikována Conoverem a jeho kolegy v roce 1979 a její praktické použití v práci Imana a Conovera (1982). LHS je velmi účinným nástrojem pro provádění statistických analýz, které jsou zaměřeny na stanovení nižších statistických momentů výsledných proměnných. Výrazně snižuje počet simulací při zachování vysoké přesnosti odhadů (počet simulací N je v řádu desítek až prvních stovek). LHS tak konverguje výrazně rychleji ke správnému řešení než metoda Monte Carlo. Výhoda metody plyne ze způsobu výběru realizací, kdy celý rozsah vstupní náhodné proměnné je pokryt rovnoměrně vzhledem k distribuční funkci. Žádná reálná hodnota není předem vyloučena. Současně metoda zachovává zjištěné (odhadnuté) funkce hustoty pravděpodobnosti pro jednotlivé náhodné proměnné a stanovené korelační koeficienty mezi nimi. Aby bylo dosaženo těchto výhod, LHS sestává vysoce závislou sdruženou hustotu pravděpodobnosti vektoru náhodných veličin.

3.1 Výběr vzorků náhodné proměnné

Principem metody je N násobné generování vzorků každé náhodné proměnné. Definiční obor kumulativní distribuční funkce $F(x)$ odpovídající funkci hustoty pravděpodobnosti $f(x)$ dané proměnné je rozdělen na N disjunktivních intervalů (vrstev). Jednotlivé intervaly nabývají shodné pravděpodobnosti $1/N$. Z každé vrstvy je vybrána jedna hodnota, která reprezentuje celý interval a v simulaci se použije právě jednou. Jedná se tedy o stratifikační metodu. Pomocí inverzní transformace distribuční funkce je získána reprezentativní hodnota náhodné proměnné.

Existuje více způsobů výběru vzorků z jednotlivých intervalů na oboru distribuční funkce. Jednou z metod je vygenerování N náhodných čísel n z intervalu $\{0,1\}$ o rovnoměrném rozdělení. Tato čísla jsou následně lineární transformací přiřazena k odpovídajícím intervalům a výše zmíněnou inverzní transformací distribuční funkce stanoveny hodnoty vzorků $x_{i,k}$ (4) proměnné X_i :

$$x_{i,k} = F_i^{-1} \left(\frac{n + (k - 1)}{N} \right) \quad (4)$$

kde k je k -tý vzorek (též vrstva, interval) i -té náhodné proměnné X_i , F_i^{-1} je inverzní distribuční funkce této proměnné, n je náhodně vygenerované číslo z rovnoměrného intervalu $\{0,1\}$, N je počet intervalů a současně počet simulací. Způsob výběru a samotná metoda je často označována jako „LHS – random“.

3 LATIN HYPERCUBE SAMPLING METHOD

The Latin Hypercube Sampling (LHS) method is a numerical simulation method of the Monte Carlo (MC) type. The Monte Carlo method (see Fig. 1) is commonly used for solutions to random problems in which we require statistical and probabilistic information on the problem response. However, for the exact estimation the Monte Carlo method usually requires very many trials (program runs, simulations) to achieve the required error. Several reduction methods have been developed with the objective to reduce the quantity of simulations and the significant time demands following from them.

The first implementation of the Latin Hypercube Sampling method is associated with the solving and processing of uncertainties in analyses of the safety of nuclear power plants in the United States of America. The method was published for the first time by Conover and

his colleagues in 1979 and its practical use was first described in the work by Iman and Conover (1982). The LHS is a very effective tool for executing statistical analyses which are focused on the determination of lower moments of order statistics of resultant variables. It significantly reduces the number of simulations with the high precision of estimations (the number of simulations N is in the order of tens to several hundreds) maintained. In this way the LHS converges to the correct solution significantly faster than the Monte Carlo method. The advantage of this method follows from the method of the selection of realisations, where the entire scope of the input random variable is covered by uniformly with respect to the distribution function. No real value is a priori excluded. The current method preserves the identified (estimated) functions of the density of probability for individual random variables and the determined coefficients of correlation between them. The LHS makes up the highly dependent composite probability of the density of the random values vector.

2.1 Random variable sampling

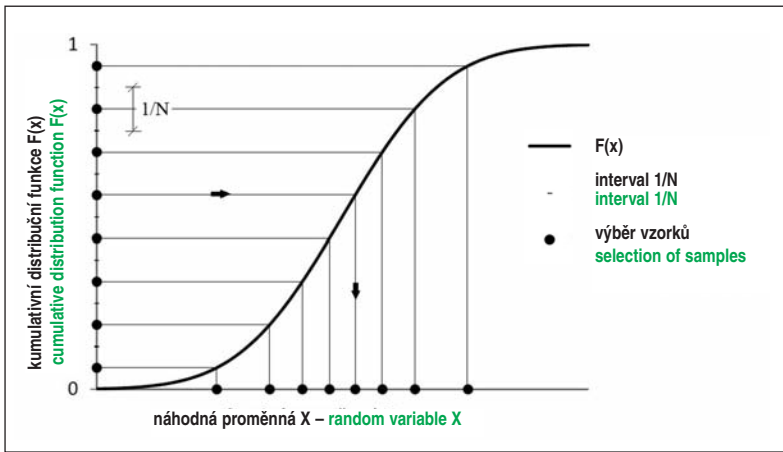
The method principle lies in repeating N times the generation of samples of each random variable. The domain of the cumulative distribution function $F(x)$ corresponding to the probability density function $f(x)$ for the particular variable is divided into N disjunctive intervals (layers). Individual intervals assume identical probabilities of $1/N$. One value is selected from each layer to represent the entire interval and is to be used only once in the simulation. It is therefore a case of a stratification method. The representative value of the random variable is derived by means of the inverse transformation of the distribution function.

There are more methods available for the selection of samples from individual intervals within the domain of the distribution function. One of the methods lies in generating N uniformly distributed random numbers n from the interval $\{0,1\}$. These numbers are subsequently bound to corresponding intervals using linear transformation and the values of samples $x_{i,k}$ (4) of the variable X_i are determined by means of the above-mentioned inverse transformation of the distribution function:

$$x_{i,k} = F_i^{-1} \left(\frac{n + (k - 1)}{N} \right) \quad (4)$$

where k is the k th sample (also layer, interval) of the i th random variable X_i , F_i^{-1} is the inverse distribution function of this variable, n is a randomly generated number from the interval $\{0,1\}$, N is the number of intervals and, at the same time, the number of simulations. The method of the selection and the method itself are often referred to as „LHS – random“.

Another method of generating samples of a random variable which is applied in the majority of studies is selecting a value



Obr. 2 Princip výběru vzorků metody „LHS median“
Fig. 2 The principle of the LHS Median method sampling

Dalším způsobem generování vzorků náhodné proměnné, který je aplikován v převážné většině studií, je výběr hodnoty ze středu intervalu $1/N$ na distribuční funkci, který je spojený opět s přímým použitím inverzní distribuční funkce (5):

$$x_{i,k} = F_i^{-1} \left(\frac{k-0,5}{N} \right) \quad (5)$$

Tato metoda je někdy označována jako „LHS median“. Obr. 2 znázorňuje kumulativní distribuční funkci $F(x)$, která je rozdělena na osm intervalů o stejné pravděpodobnosti $1/N$. Body na vertikální ose představují středy těchto intervalů a následně vyznačují odpovídající hodnoty proměnné X .

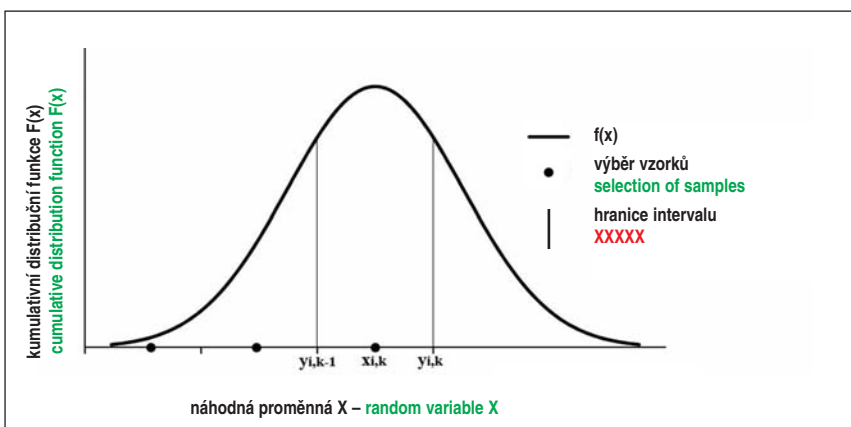
Autoři Huntington a Lyrantzis (1998), Keramat a Kielbasa (1999) poukazují na nevýhody této stratifikační metody. Ty se týkají především krajních vrstev oboru distribuční funkce, které nejvíce ovlivňují hodnotu rozptylu, šikmost a špičatost rozdělení vstupních veličin. Zatímco tato metoda poskytne odpovídající, či velmi blízkou střední hodnotu vzorků té požadované, rozptyl se většinou výrazně liší.

Huntington a Lyrantzis navrhují řešit výběr reprezentativních vzorků $x_{i,k}$ (6) jako střední hodnotu intervalu vymezeného na funkci hustoty pravděpodobnosti dané proměnné X_i . Vzorky se tak nacházejí přesně v těžištích rozdělovacích ploch.

$$x_{i,k} = \frac{1}{N} \int_{y_{i,k-1}}^{y_{i,k}} x f(x) dx \quad (6)$$

Meze integrálu $y_{i,k}$ lze určit ze vztahu (7):

$$y_{i,k} = F_i^{-1} \left(\frac{k}{N} \right) \quad (7)$$



Obr. 3 Princip výběru vzorků metody „LHS mean“
Fig. 3 The principle of the LHS Mean method sampling

from the mid-point of the interval $1/N$ on the distribution function, which is again associated with the direct use of the distribution function (5):

$$x_{i,k} = F_i^{-1} \left(\frac{k-0,5}{N} \right) \quad (5)$$

This method is sometime referred to as the „LHS median“. Fig. 2 depicts the cumulative distribution function $F(x)$, which is divided into eight intervals with equal probability of $1/N$. Points on the vertical axis represent mid-points of these intervals and subsequently mark the corresponding values of the variable X .

Authors Huntington and Lyrantzis (1998), Keramat and Kielbasa (1999) point out disadvantages of this stratification method. These disadvantages are first of all related to extreme layers of the distribution function domain, which most of all affect the variance of input quantities. Whilst this method provides the adequate value or a mean value very close to the required value of samples, the variance is mostly significantly different.

Huntington and Lyrantzis propose that the selection of representative samples $x_{i,k}$ (6) should be solved as a mean value of the interval delineated on the function of the density of probability of a particular variable X_i . In this case the samples are located exactly in the median points of dividing areas.

$$x_{i,k} = \frac{1}{N} \int_{y_{i,k-1}}^{y_{i,k}} x f(x) dx \quad (6)$$

Limits of integration $y_{i,k}$ can be determined from the relationship (7):

$$y_{i,k} = F_i^{-1} \left(\frac{k}{N} \right) \quad (7)$$

The method referred to as the LHS Mean (see Fig. 3) catches the probability density function, where the mean value is determined exactly and the estimation of the variance of the particular variable is significantly closer to the required variance. The improvement of the sampling method is demonstrated by the authors by comparing the statistical accuracies of the two methods for exponentially distributed random variables with various number of samples (simulations). The values of the samples selected using the two methods are nearly identical; they differ only at the distribution boundaries. Fig. 4 demonstrates the difference in the selection of samples using both methods for a normally distributed variable $f(x)$, which is in this particular case characterised by Young's modulus E . The influence of the number of simulations on the dissimilarity of the selection of samples is obvious.

The LHS Mean method converges to a correct solution faster than the LHS Median method, therefore it requires smaller number of simulations. This fact is even confirmed by Vořechovský (2009).

3.2 Form of samples and their arrangement

When the LHS method is applied, one of the two forms of samples arranged in sets for individual simulations is usually taken into consideration. In the first set, after generating representative samples for all random variables taken into consideration, samples for individual simulations are selected in the form of random permutations of whole numbers 1, 2, ... to N (table 1).

The permutation table lines define the number of simulations N and their sequence, whilst the columns define the number of variables X_i taken into consideration when solving the random

Metoda, často nazývaná „LHS mean“ (obr. 3), lépe vystihuje funkci hustoty pravděpodobnosti, kdy střední hodnota je stanovena přesně a odhad rozptylu dané proměnné je značně bližší požadovanému. Vylepšení vzorkovací metody autoři demonstrují na srovnání statistické přesnosti obou metod pro exponenciálně rozdělené náhodné proměnné o různém počtu vzorků (simulací). Hodnoty vzorků vybrané oběma způsoby jsou téměř identické, liší se pouze „okrají“ rozdělení. Obr. 4 demonstruje rozdíl ve způsobu výběru vzorků obou metod pro normálně rozdělenou proměnnou $f(x)$, která je v daném případě charakterizována Youngovým modulem E . Patrný je vliv počtu simulací na odlišnost výběru vzorků.

Metoda „LHS mean“ konverguje ke správnému řešení rychleji než metoda „LHS median“, a tudíž vyžaduje menší počet simulací. Tuto skutečnost potvrzuje i Vořechovský (2009).

3.2 Forma vzorků a jejich uspořádání

Při aplikaci metody LHS se většinou uvažuje jedna ze dvou forem vzorků uspořádaných do sad pro jednotlivé numerické simulace. V první, po vygenerování reprezentativních vzorků pro všechny uvažované náhodné proměnné, jsou vzorky do jednotlivých simulací vybírány ve formě náhodných permutací celých čísel 1, 2, ... až N (viz. tabulka 1).

Řádky permutační tabulky definují počet simulací N a jejich pořadí, sloupce počet proměnných X_i uvažovaných při řešení náhodného problému. Jednotlivé prvky, čísla permutační tabulky určují k -tou vrstvu (interval), ze které je vybrán vzorek $x_{i,k}$ proměnné X_i . Jakmile jsou pořadová čísla nahrazena vzorky, můžeme hovořit o matici X $K \times N$, kde K označuje počet proměnných a N počet simulací.

Druhá forma uvažuje normalizované vzorky uspořádané do matice U $K \times N$.

$$u_{ij} = \frac{x_{i,j} - u_j}{\sigma_j}$$

Tab. 1 Permutační tabulka – počet simulací N /proměnných X_i

Table 1: Permutation table – number of simulations N /variables X_i

	X_1	X_2	X_3	X_4	X_i
1	3	1	4	3	5
2	5	2	2	1	4
3	1	4	3	5	1
4	4	3	5	2	3
N	2	5	1	4	2

problem. Individual elements, numbers in the permutation table, determine the k th layer (interval) which the sample $x_{i,k}$ of the variable X_i is selected from. As soon as the sequence numbers are replaced by samples, we can speak about an X $K \times N$ matrix, where the K denotes the number of variables and N is for the number of simulations.

The second form considers standardised elements arranged in a U $K \times N$ matrix.

$$u_{ij} = \frac{x_{i,j} - u_j}{\sigma_j}$$

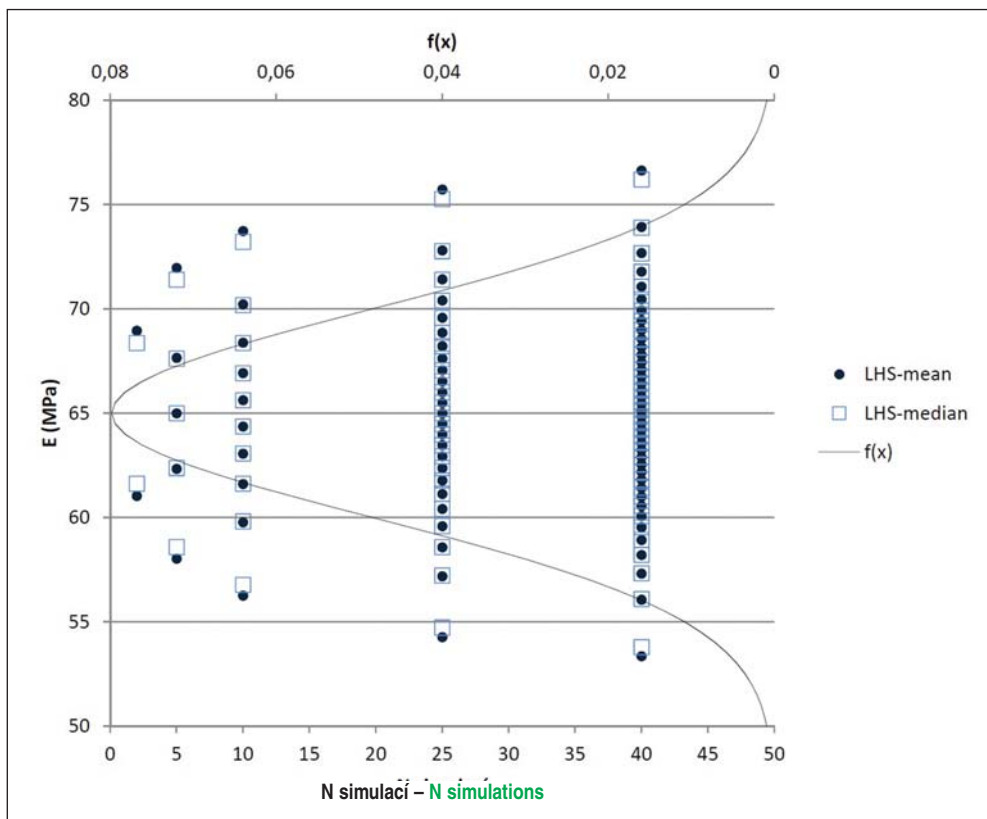
where u_j and σ_j are the mean value and standard deviation of the j th variable respectively. This transformation will ensure that all columns (variables) have a zero mean value and unit variance. Other options are to work directly with the values of samples of individual variables or with random variables from the uniform distribution (0,1) generated in the columns of a $K \times N$ matrix.

However, significant undesired correlations affecting the accuracy and quality of results usually originate among individual variables. The random permutation described above represents the simplest permutation method. The assumption is adopted that the generated vectors of the X matrix are independent, or the dependences which originate are sufficiently small.

3.3. Methods for introducing statistical dependence

It is advisable to verify the measure of statistical dependence of individual vectors (columns of the permutation table) using a Pearson's linear correlation coefficient or a Spearman's rank correlation coefficient. The majority of works (see below) solve the implementation of statistical correlation in the form of exchanging the sequence of samples at individual variables and do not change their values. In this way, the probability distribution of each random quantity is preserved.

There are several methods for selecting samples proposed for the simulation of correlated random vectors available within the framework of Monte Carlo type of simulations, where random variables have arbitrary probability distribution and values of correlation coefficients. This paper is further focused on a brief description and assessment of a group of methods which are based on the optimisation of the sequence of samples



Obr. 4 Porovnání výběru vzorků metodami „LHS median“ a „LHS mean“ se zohledněním vlivu počtu simulací
Fig. 4 Comparison of sampling using the LHS Median and LHS Mean methods with the number of simulations taken into consideration

kde u_j a σ_j jsou střední hodnota a směrodatná odchylka j -té proměnné. Tato transformace zajistí, že všechny sloupce (proměnné) mají nulovou střední hodnotu a jednotkový rozptyl. Dalšími možnostmi je pracovat přímo s hodnotami vzorků jednotlivých proměnných nebo s náhodnými hodnotami z rovnoměrného rozdělení (0,1) vygenerovaných ve sloupcích matice $K \times N$.

Mezi jednotlivými proměnnými však dochází zpravidla ke vzniku značných nežádoucích korelací, které ovlivňují přesnost a kvalitu výsledků. Výše popsaná náhodná permutace představuje nejjednodušší permutační metodu. Je přijat předpoklad, že vygenerované vektory matice X jsou nezávislé, případně vzniklé závislosti dostatečně malé.

3.3 Metody zavedení statistické závislosti

Míru statistické závislosti jednotlivých vektorů (sloupců permutační tabulky) je vhodné ověřit využitím Pearsonova lineárního korelačního koeficientu nebo Spearmanova koeficientu pořadové korelace. Většina prací (viz dále) řeší implementaci statistické korelace formou záměny pořadí vzorků u jednotlivých proměnných a nemění již jejich hodnoty. Tímto způsobem je zachováno pravděpodobnostní rozdělení každé náhodné proměnné.

Existuje řada metod výběru vzorků navržených pro simulaci korelovaných náhodných vektorů v rámci simulací typu Monte Carlo, kdy náhodné proměnné disponují libovolným pravděpodobnostním rozdělením a hodnotami korelačních koeficientů. Článek se dále zaměřuje na stručný popis a hodnocení skupiny metod, která je založena na optimalizaci pořadí vzorků ve sloupcích matice $K \times N$, které byly generovány bez přihlídnutí ke korelacím.

Ortogonalní transformace nezávislých náhodných proměnných v korelované představuje nejrozšířenější způsob zavádění korelace v metodě LHS a vychází z metody publikované autory Scheuerem a Stollerem (1962). Ti navrhli způsob generování korelovaných proměnných s Gaussovým rozdělením, jež je založený na Choleského dekompozici kovarianční matice (viz dále). Iman a Conover (1982) popsali metodu využívající Spearmanův koeficient korelace k popisu statistické závislosti mezi sloupci matice pořadových čísel a Choleského dekompozici korelační matice. Navržená metoda má obecnější charakter ve srovnání s návrhem autorů Scheuera a Stollera, jelikož není omezena na normálně rozdělené náhodné proměnné. Nevýhodou však zůstává předpoklad nekorelovaných vstupních proměnných a přetrvávající významná chyba v korelaci mezi proměnnými vzhledem k požadované korelační matici.

Metoda nazvaná „Updated Latin Hypercube Sampling“ (ULHS) (Florian 1992) vychází z postupu Iman a Conover a byla vyvinuta pro redukci nežádoucích korelací s cílem obdržet nekorelované proměnné. Spearmanův koeficient korelace (8) popisuje závislost mezi sloupci matice náhodně generovaných pořadových čísel \mathbf{R} ($K \times N$):

$$c_{i,j} = 1 - \frac{6 \sum_k (R_{ki} - R_{kj})^2}{N(N-1)(N+1)} \quad (8)$$

kde koeficienty $c_{i,j}$ jsou Spearmanovy koeficienty pořadové korelace mezi proměnnými i a j v intervalu $\{-1,1\}$ a \mathbf{R} jednotlivé prvky matice \mathbf{R} . Korelační matice \mathbf{C} je symetrická a pozitivně definitivní vyjma případů, kdy některé sloupce mají identická pořadí. Choleského dekompozice (9) vyžaduje, aby matice \mathbf{C} byla pozitivně definitivní, což současně znamená, že počet simulací (realizací) musí být větší než počet proměnných, tzn. $N > K$. V případě UHLS má \mathbf{C} tvar jednotkové matice.

$$\mathbf{C} = \mathbf{L}^T \mathbf{L} \quad (9)$$

\mathbf{L} je dolní trojúhelníková matice. Upravená matice pořadí \mathbf{R}^* vychází ze vztahu (10):

in the columns of a $K \times N$ matrix, which were generated without taking correlations into consideration.

The orthogonal transformation of independent random variables into correlated variables represents the most widely spread method for introducing correlation in the LHS method; it is based on a method published by authors Scheuer and Stoller (1962). They proposed a method for generating correlated variables following the Gaussian curve of distribution, which is based on Cholesky decomposition of covariance matrices (see below). Iman and Conover (1982) described a method using Spearman's coefficient of correlation for the description of statistical dependence between columns of a sequential numbers matrix and the Cholesky decomposition of a correlation matrix. The proposed method has a more general character in comparison with the proposal of authors Scheuer and Stoller because of the fact that it is not restricted to normally distributed random variables. Anyway, the assumption of uncorrelated input variables and a persisting significant error in correlation between variables with respect to the required correlation matrix remain to be disadvantage.

The method referred to as the Updated Latin Hypercube Sampling (ULHS) (Florian 1992) starts from the Iman and Conover procedure and was developed for the reduction of undesired correlation with the aim of obtaining uncorrelated variables. Spearman's coefficient of correlation (8) describes the dependence among columns of a matrix of randomly generated sequential numbers \mathbf{R} ($K \times N$):

$$c_{i,j} = 1 - \frac{6 \sum_k (R_{ki} - R_{kj})^2}{N(N-1)(N+1)} \quad (8)$$

where coefficients $c_{i,j}$ are Spearman's coefficients of correlation among variables i and j within the interval $\{-1,1\}$ and R is for individual elements of matrix \mathbf{R} . The correlation matrix \mathbf{C} is symmetrical and positively definite with the exception of cases where some columns have identical serial numbers. The Cholesky decomposition (9) requires the matrix \mathbf{C} to be positively definite, which at the same time means that the number of simulations (realisations) must be greater than the number of variables, i.e. $N > K$. In the case of the UHLS, the \mathbf{C} has the shape of a unit matrix.

$$\mathbf{C} = \mathbf{L}^T \mathbf{L} \quad (9)$$

\mathbf{L} is a lower triangular matrix. The modified matrix of sequence \mathbf{R}^* is derived from the relationship (10):

$$\mathbf{R}^* = \mathbf{R} \mathbf{L}^{-1} \quad (10)$$

The sequential numbers in each column of the matrix of sequence \mathbf{R} are subsequently aligned in a way securing that their sequence is identical with the sequence of numbers in corresponding columns of matrix \mathbf{R}^* . This system of modifying the sequential matrix can be realised iteratively and it theoretically allows significant reduction of correlation (Huntington and Lyrintzis 1998). Thus the matrix \mathbf{R} will contain significantly lower level of undesired correlation in its result. Although, other authors caution that, when used in praxis, the ULHS method tends to converge to sequences which continue to provide erroneous generation among variables. In the case of the simulation of correlated variables, the technique described above encounters certain problems. First of all it is necessary to use the ULHS iterative method to minimise the correlation. Subsequently the matrix \mathbf{T} with required correlation coefficients is subjected to the Cholesky decomposition and a new sequential matrix is obtained using the equation for \mathbf{R}^* . This correlation method can be used exclusively once, without possibility for the improvement by iterations.

Huntington and Lyrintzis (1998) propose an optimisation method called „single-switch“. Instead of using serial numbers it

$$\mathbf{R}^* = \mathbf{R}\mathbf{L}^{-1} \quad (10)$$

Pořadová čísla v každém sloupci pořadové matice pořadí \mathbf{R} jsou následně srovnána tak, aby měla stejná pořadí jako čísla v odpovídajících sloupcích matice \mathbf{R}^* . Tento způsob úpravy matice pořadí může být realizován iterativně a teoreticky dovolí výrazné snížení korelace (Huntington a Lyrantzis, 1998). Matice \mathbf{R} tak ve výsledku bude zahrnovat značně nižší úroveň nežádoucí korelace. Další autoři však upozorňují na skutečnost, že při použití v praxi má metoda ULHS tendenci konvergovat k pořadím, které i nadále poskytují chybnou korelaci mezi proměnnými. V případě simulace korelovaných proměnných má popsaná technika určité problémy. Nejprve je nutné použít iterativně metodu ULHS k minimalizaci korelace. Následně je matice \mathbf{T} s požadovanými korelačními koeficienty podrobena Choleského dekompozici a využitím rovnice pro \mathbf{R}^* získána nová pořadová matice. Tato korelační metoda může být použita výhradně jednou, již zde není možnost zlepšení iteracemi.

Huntington a Lyrantzis (1998) navrhují optimalizační metodu nazvanou „single-switch“. Ta nevyužívá pořadová čísla, ale přímo uspořádává upravené vzorky a pro stanovení korelací využívá Pearsonův korelační koeficient. Matice \mathbf{R} obsahující neseřazené modifikované vzorky (11):

$$\mathbf{s}_{i,j} = \frac{(x_{i,j} - u_j)}{\sigma_j} \quad (11)$$

kde $u_{i,j}$ je střední hodnota j -té proměnné a σ_j je odpovídající směrodatná odchylka. Následující procedura je řešena postupně pro každý sloupec m z uvažovaného (řešeného) $m-1$ počtu sloupců. Vektor \mathbf{T} , obsahující aktuální korelační koeficienty mezi m -tým a každým předchozím sloupcem je definován (12):

$$T_j = \frac{1}{N} \sum_{i=1}^N R_{ij} R_{im} \quad (12)$$

kde $1 \leq j \leq m-1$. Chyba E korelačních koeficientů (nezaměnit s efektivním zatížením E v EC7) je (13):

$$E = \sum_{j=1}^{m-1} (T_j - T'_{jm})^2 \quad (13)$$

kde \mathbf{T}' je matice obsahující požadované korelační koeficienty. Jsou-li uvažovány nekorelované proměnné, tyto koeficienty nabývají nulových hodnot. Principem optimalizační metody je výpočet změny E pro každý pár vzorků m -té proměnné, který nastane při jejich vzájemném prohození. Pár poskytující největší redukci E je prohozen. Tento postup je pro m -tou proměnnou realizován iterativně až do okamžiku, kdy není možné žádné další zlepšení, nebo se korelační koeficienty nacházejí uvnitř definovaných intervalů. Postupně je procedura opakována pro všechny proměnné. Na závěr je k získání matice vzorků \mathbf{S} použita přetříděná matice vzorků \mathbf{R}^* . Autoři ve své studii vyhodnocují nově navrženou metodu na základě porovnání s technikou UHLS. Metoda „single-switch“ poskytuje výrazně vyšší korelační přesnost při nižším počtu proměnných a vzorků ve srovnání s UHLS.

Velmi efektivní a přesnou proceduru pro zavedení požadovaných korelací představuje metoda simulovaného žíhání (např. Morris a Mitchel, 1995, Vořechovský, 2009). Metoda pracuje jak se Spearmanovým koeficientem, tak s klasickým lineárním Pearsonovým a dalšími koeficienty korelace. Pro hodnocení kvality statistických závislostí upřednostňuje normu E , která zohledňuje deviaci všech koeficientů korelace s druhou mocninou (14), před maximálním rozdílem, který není vhodný pro přímou minimalizaci.

directly arranges modified samples and uses Pearson's correlation coefficient for the determination of correlations. The matrix \mathbf{R} containing unaligned modified samples (11):

$$\mathbf{s}_{i,j} = \frac{(x_{i,j} - u_j)}{\sigma_j} \quad (11)$$

where $u_{i,j}$ is the median value of j th variable and σ_j is the corresponding standard deviation. The subsequent procedure is solved step-by-step for each column m from the $m-1$ number of the columns being considered (solved). The vector \mathbf{T} containing current topical coefficients between the m th column and each preceding column is defined (12):

$$T_j = \frac{1}{N} \sum_{i=1}^N R_{ij} R_{im} \quad (12)$$

where $1 \leq j \leq m-1$. The error E of correlation coefficients (not to mistake with effective load E in EC7) is (13):

$$E = \sum_{j=1}^{m-1} (T_j - T'_{jm})^2 \quad (13)$$

where \mathbf{T}' is a matrix containing required correlation coefficients. If uncorrelated variables are taken into consideration, these coefficients assume zero values. The principle of the optimisation method lies in the calculation of the change E for each pair of samples of the m th variable, which originates when they are swapped. The pair providing the greatest reduction E is swapped. This procedure is realised for the m th variable iteratively until no other improvement is possible or the correlation coefficients are found inside the defined intervals. The procedure is repeated step by step for all variables. In conclusion, the resorted matrix of samples \mathbf{R}^* is used for obtaining the matrix of samples \mathbf{S} . In their study, the authors assess the newly proposed method on the basis of a comparison with the UHLS technique. The Single-Switch method provides significantly higher correlation accuracy at a lower number of variables and samples in comparison with the UHLS.

A very effective and accurate procedure for the implementation of required correlations is represented by the Simulated Annealing Method (e.g. Morris and Mitchel 1995, Vořechovský 2009). The method operates with both Spearman's coefficient and classical linear Pearson's coefficient and other correlation coefficients. It prefers the standard E for assessing the quality of statistical dependences, allowing for deviations of all correlation coefficients with a square (14), to the maximum difference, which is not suitable for direct minimisation.

$$E = \sum_{i=1}^{N_V} \sum_{j=1}^{N_V} (A_{i,j} - T_{i,j})^2 \quad (14)$$

where N_V is the number of variables and in brackets there is the difference between topical correlation coefficients $A_{i,j}$ and required $T_{i,j}$ quantities i and j . The method starts from a simple probabilistic method which swaps the serial numbers of a random pair of samples of a randomly selected variable in matrix \mathbf{X} (e.g. standardised samples, serial numbers of samples). In the case of the standard E reduction, the matrix \mathbf{X} was adopted as a new generation, i.e. the initial matrix. But the above-mentioned procedure in the majority of cases ends in a state where no change brings the reduction of the standard and the new matrix cannot be adopted. The method therefore finds the so-called local minimums, without a chance of finding the actual global minimum. The principle of the Simulated Annealing Method is therefore based on the violation of the condition that a new generation of samples may be adopted only in the case of the standard E having been reduced. A new generation originates in each iteration. It is either adopted or not. The simulated annealing is in line in the other case and the

$$E = \sum_{i=1}^{N_V} \sum_{j=1}^{N_V} (A_{i,j} - T_{i,j})^2 \quad (14)$$

Kde N_V je počet proměnných a v závorce rozdíl korelačních koeficientů aktuálních $A_{i,j}$ a požadovaných $T_{i,j}$ veličin i a j . Metoda vychází z jednoduché pravděpodobnostní metody, která zaměňuje pořadí náhodné dvojice vzorků náhodně vybrané proměnné v matici \mathbf{X} (např. normalizované vzorky, pořadová čísla vzorků). V případě, že došlo ke snížení normy E , byla matice \mathbf{X} přijata jako nová generace, tedy výchozí matice. Uvedený postup však ve většině případů končí ve stavu, kdy jakákoli změna nepřináší snížení normy a nová matice nemůže být přijata. Metoda tak nachází tzv. lokální minima bez šance nalezení skutečného globálního minima. Princip metody simulovaného žhání je tedy založen na porušení podmínky, že nová generace vzorků může být přijata pouze v případě, že došlo ke snížení normy E . V každé iteraci dochází k vytvoření nové generace, která je buď přijata, nebo ne. Simulované žhání přichází na řadu v druhém případě a nový vektor s vyšší energetickou konfigurací ($\Delta E > 0$) je přijat s určitou pravděpodobností, která se řídí Boltzmannovo rozdělením (15).

$$P \approx \exp\left(\frac{-\Delta E}{t}\right) \quad (15)$$

Toto rozdělení uvažuje systém, který se nachází v teplotní rovnováze t a má svou energii pravděpodobnostně rozloženou mezi možné energetické stavy ΔE . Nenulová pravděpodobnost v přijetí nové generace vzorků zvyšuje možnost nalezení konfigurace, která se vyhne lokálnímu minimu.

Další vybrané permutační metody budou zmíněny již méně podrobně. Mezi ně náleží i metoda „Symmetric Latin Hypercubes Design (SLHD)“ (Ye a kol., 2000), která vychází ze symetrické matice $N \times K$. SLHD je založen na algoritmu „columnwise-pairwise (CP)“. Ten při hledání optimální symetrické matice v každém kroku dané iterace prohazuje současně pořadí dvou párů vzorků uvnitř sloupce. Na konci iterace je vybrána nejlepší záměna vzhledem ke kritériu minimalizace a použita k úpravě matice. Podoba výsledné matice je výrazně citlivá na počáteční náhodné vygenerování permutační matice, tudíž by celý algoritmus měl být opakován vícekrát s různými počátečními permutacemi. Metodu autoři vyhodnocují na základě porovnání s algoritmem simulovaného žhání a Choleského dekompozicí. Dalším příkladem algoritmu založeného na optimalizaci pořadí prvků pořadové matice, svým řešením blízký Choleského dekompozici, je Gram–Schmidtova pořadová ortogonalizace. Existuje řada dalších metod založených na optimalizaci pořadí vzorků matice prvků, případně pořadové matice, které byly implementovány do metody LHS z důvodu zohlednění korelací mezi proměnnými a nejsou zde podrobněji popsány. Jejich princip je však velmi podobný algoritmům popsaným výše.

Ke spolehlivému stanovení korelovanosti proměnných je nutné mít k dispozici značné množství výsledků z geotechnického průzkumu. Tento předpoklad je velmi obtížné naplnit, jelikož finanční náklady projektu určené na laboratorní experimenty jsou v současné době s ojedinělými výjimkami velmi omezené. Většina prací zaměřených na pravděpodobnostní řešení geotechnických problémů se tak potýká s nedostatkem vhodných statistických dat a je proto nutné hledat možné vztahy v databázích projektů realizovaných v obdobném geologickém prostředí, ačkoli tak dochází k vnášení nepřesností do analýz. I z tohoto důvodu studie, zabývající se vlivem variability vstupních proměnných, často řeší experimentální problém, kde potřebná data jsou dostupná.

new vector with a higher energy configuration ($\Delta E > 0$) is adopted with a certain probability which follows Boltzmann's distribution (15).

$$P \approx \exp\left(\frac{-\Delta E}{t}\right) \quad (15)$$

This distribution takes into consideration a system which is found in temperature balance t and has its energy in terms of probability distributed among possible energy states ΔE . The non-zero probability in the adoption of a new generation of samples increases the chance of finding a configuration which will avoid the local minimum.

Other selected permutation methods will be mentioned in smaller detail. The Symmetric Latin Hypercubes Design (SLHD) method (Ye et al. 2000), starting from a symmetric matrix $N \times K$, belongs among them. The SLHD is based on a Columnwise-Pairwise (CP) algorithm. When an optimum symmetric matrix is being sought, this algorithm concurrently swaps the sequence of two pairs within a column in each step of a particular iteration. The swap which is best with respect to the minimisation criterion is selected at the end of the iteration and is used for the matrix modification. The shape of the resultant matrix is significantly sensitive to the initial random generation of the permutation matrix, therefore the whole algorithm should be repeated several times with different initial permutations. The authors have assessed the method on the basis of a comparison with the simulated annealing algorithm and Cholesky decomposition. Another example of an algorithm based on the optimisation of the sequence of elements of a sequential matrix, which is close to Cholesky decomposition in terms of its solution, is the Bram – Smith sequential orthogonalisation. There are several other methods available based on the optimisation of the sequence of samples of the matrix of elements or sequential matrix, which were implemented into the LHS method with the aim of allowing for correlations among variables and are not described in this paper in more detail. Anyway, their principle is very similar to the algorithms described above.

If the determination of the degree of correlation of variables is to be reliable, it is necessary to have a significant number of the geotechnical investigation results available. Although, it is very difficult to meet this condition because of the fact that financial expenses of a project allocated to laboratory experiments are today, with isolated exceptions, very limited. The majority of works focused on probabilistic solutions to geotechnical problems fight with the lack of suitable statistical data and it is therefore necessary to seek potential relationships in the databases of projects which were realised in a similar geological environment, despite the fact that in this way inaccuracies are inserted into analyses. It is also for this reason that studies dealing with the influence of the variability of input variables frequently solve an experimental problem where there is the required data available.

3.4 Other modifications of the LHS method

In comparison with the Monte Carlo method, the fundamental concept of the LHS method suffers from a significant disadvantage. It does not allow arbitrary adding simulations to the realisations for which the calculations were already completed. This disadvantage follows from the formulation itself and from the method principle, where it is necessary from the very beginning of solving the problem to choose the exact number of simulations. If we randomly added other simulations to the already completed realisations, the loss of consistency, or the uniform distribution within the distribution function domain, would follow. If an insufficient accuracy of the problem response is identified, it is necessary to proceed to solving the problem from the very

3.4 Další úpravy metody LHS

Ve srovnání s metodou Monte Carlo má základní koncept metody LHS výraznou nevýhodu, neumožňuje libovolně přidávat simulace k již vypočteným realizacím. Tato nevýhoda plyne z vlastní formulace a podstaty metody, kdy hned na začátku řešení problému je nutné zvolit přesný počet simulací. Pokud bychom k již vypočteným realizacím nahodile přidali další, došlo by ztrátě konzistence neboli rovnoměrného rozdělení na oboru distribuční funkce. V případě zjištění nedostatečné přesnosti odezvy problému je nutné přistoupit k řešení od samého počátku tentokrát s vyšším počtem simulací. Řešení výše popsané nevýhody nabízí Sallaberry a kol. (2008). Autoři navrhnou algoritmus pro rozšíření výchozí LHS o počtu vzorků m a korelovaných proměnných s odpovídající korelační maticí C na LHS s $2m$ vzorky a korelační maticí blízkou originální. Hierarchical Subset Latin Hypercube Sampling je metoda navržená Vořechovským (2009, 2010) z důvodu řešení možného rozšíření počtu simulací pro korelované náhodné proměnné. Principem metody HSLHS je přidávání simulací v podobě obsazení pouze konkrétních vybraných subintervalů. Přidávané simulace nemají podobu LH – vzorků jako takových, a nelze je tudíž vyhodnotit samostatně. Vždy musejí být zkombinovány se sadou vzorků na nižší úrovni a vyhodnoceny společně již jako přesný LHS celek rovnoměrně rozložený na intervalu (0,1) jako v případě metody LHS – median.

Další úprava LHS a její aplikace, vztažená především k akademické půdě, souvisí se zohledněním prostorové variability vlastností horninového masivu. Metoda je založena na teorii náhodných polí a autokorelace prostorových vlastností geologického prostředí vstupuje do výpočtu ve formě korelační funkce, která je nejčastěji vyjádřena exponenciálním vztahem (např. Markovova funkce).

4 APLIKACE METODY LHS PRO ŘEŠENÍ GEOTECHNICKÝCH ÚLOH V ČR

Využití metody LHS v geotechnické praxi v České republice není příliš časté, převážná většina prací zabývající touto tematikou byla vypracována na akademické půdě.

Statistická analýza numerického modelu tunelu Mrázovka je součástí práce Hilara (2000), kde pomocí LHS byl vyhodnocován vliv pěti základních parametrů Mohr-Coulombova modelu na deformace ostění a povrchu. Dalším příkladem je využití metody LHS pro výpočty tunelu Valík (Hrubešová a kol., 2003), kde je pomocí LHS vyhodnocován vliv 10 vstupních parametrů (např. také součinitel anizotropie, úklon puklin, atd.). Vaněčková (2008) využívá metody LHS pro řešení stability skalního svahu prostoupeného systémem diskontinuit. Geotechnické parametry a další vlastnosti ploch nespojitosti mohou nabývat čtyř různých pravděpodobnostních rozdělení. Součástí práce Paráka (2008) je využití metody LHS pro stanovení vlivu geotechnických parametrů geologických vrstev, parametrů stříkaného betonu a počátečních podmínek na strukturní síly v ostění tunelu kruhového průřezu. Využití LHS v geotechnice se dlouhodobě věnuje doc. J. Pruška (např. Pruška, Šedivý, 2010).

Výše uvedené práce předpokládají lineárně nezávislé vstupní geotechnické parametry, proto v nich nebyla řešena jejich korelace. Dále všechny uvedené práce uvažují typ metody LHS – median (střed intervalu na oboru distribuční funkce). Tato metoda ve srovnání s LHS – mean konverguje k výsledku výrazně pomaleji, pro dosažení uspokojivé přesnosti je doporučován větší počet simulací. Dále je v uvedených pracích zpravidla uvažováno normální pravděpodobnostní rozdělení vstupních parametrů.

Výrazné uplatnění v ČR nachází metoda LHS v oboru železobetonových konstrukcí. Ústav stavební mechaniky Stavební fakulty

beginning, this time with a higher number of simulations. The solution to the above-mentioned disadvantage is offered Sallaberry et al. (2008). The authors propose an algorithm for the expansion of the initial LHS with the number of samples m and correlated variables with the corresponding correlation matrix C to the LHS with $2m$ samples and correlation matrix close to the original matrix. The Hierarchical Subset Latin Hypercube Sampling is a method proposed by Vořechovský (2009, 2010) with the aim of solving possible extension of the number of simulations for correlated random variables. The principle of the HSLHS method lies in adding simulations in the form of occupying only concrete selected sub-intervals. The simulations being added do not have the form of LH-samples as such and it is therefore impossible to assess them independently. They must always be combined with a set of samples at the lower level and be assessed jointly, already as an accurate LHS complex evenly distributed within the interval (0,1), in the same way as in the case of the LHS – Median method.

Another modification of the LHS and its application, relating first of all to university premises, is connected with allowing for the spatial variability of ground mass properties. The method is based on the theory of Random Fields and the autocorrelation of spatial properties of geological environment enters the calculation in the form of a correlation function, which is most frequently expressed by an exponential relationship (e.g. mark correlation function).

4 THE LHS METHOD APPLICATION TO SOLVING GEOTECHNICAL PROBLEMS IN THE CZECH REPUBLIC

The use of the LHS method in geotechnical practise in the Czech Republic is not too frequent; the majority of works dealing with this topic was carried out on university premises.

The statistical analysis of the Mrázovka tunnel numerical model is part of the work of Hilar (2000), where the LHS was used for assessing the influence of five basic parameters of the Mohr-Coulomb model on deformations of the lining and terrain surface. Another example is the use of the LHS method for the Valík tunnel calculations (Hrubešová et al. 2003), where the LHS is applied to the assessment of the influence of 10 input parameters (e.g. the anisotropy coefficient, dipping of discontinuities etc.). Vaněčková (2008) uses the LHS method for solving problems of stability of a rock slope interspersed by a system of discontinuities. Geotechnical parameters and other properties of discontinuity surfaces can assume four different probability distribution forms. Part of the work by Parák (2008) is the application of the LHS method to the determination of the influence of geotechnical parameters of geological strata, parameters of sprayed concrete and initial conditions on structural forces in a circular tunnel lining. Doc. J. Pruška devotes himself to the use of the LHS in geotechnics in the long term (e.g. Pruška, Šedivý 2010).

The above-mentioned works assume linearly independent input geotechnical parameters, which is the reason why they did not deal with their correlation. In addition, all of the above-mentioned works consider the LHS-Median method type (the mid-point of an interval within the domain of the distribution function). This method, in comparison with the LHS-Mean, converges to the result significantly slower; a higher number of simulations is recommended for the purpose of achieving a satisfactory accuracy. Further, normal probability distribution of input parameters is usually assumed in the above-mentioned works.

The LHS method finds significant use in the CR in the field of reinforced concrete structures. The Institute of Building Mechanics of the Faculty of Civil Engineering of the University of Technology in Brno, represented by Prof. Novák and others,

VUT v Brně v zastoupení doc. M. Vořechovského, prof. Nováka a dalších se dlouhodobě zabývá úpravami a vylepšeními metody LHS pro praktické aplikace.

5 SHRNUŤÍ

Metoda Latinských hyperkrychlí (LHS) představuje efektivní pravděpodobnostní metodu typu Monte Carlo pro statistické zpracování vstupních proměnných a odhad statistických momentů odezvy řešeného problému. Největší výhodou je možnost značného snížení počtu simulací oproti standardní metodě Monte Carlo při zachování vysoké přesnosti odhadů. Metoda LHS zachovává pravděpodobnostní rozdělení přiřazené všem simulovaným proměnným a umožňuje zohlednění korelovanosti mezi nimi. Byla vyvinuta a přijata řada modifikací, které zvyšují přesnost metody a současně snižují časovou náročnost simulací. Metoda LHS nachází své uplatnění v řadě oborů, mezi nimiž má své místo i geotechnika včetně podzemního stavitelství. Využití metody LHS pro statické výpočty podzemních staveb může výrazně zpřesnit představu o předpokládaném chování posuzované konstrukce (zejména pak o pravděpodobnosti výskytu extrémních stavů).

Tento příspěvek byl zpracován s podporou grantů TAČR TA01011816 a GAČR P105/12/1705.

**RNDr. TOMÁŠ SVOBODA, Ph.D., svoboda.tomas@3-g.cz,
3G Consulting Engineers, s. r. o.,
DOC. ING. MATOUŠ HILAR, Ph.D., hilar@3-g.cz,
3G Consulting Engineers, s. r. o., a FSv ČVUT**

Recenzoval: doc. Dr. Jan Pruška

deals with problems of modifications and improvements of the LHS method for practical applications in the long term.

5 SUMMARY

The Latin Hypercube Sampling (LHS) method represents an effective probabilistic method of the Monte Carlo type for statistical processing of input variables and estimation of statistical moments of the response of the problem being solved. The greatest advantage lies in the possibility of significant reducing the number of simulations in comparison with the standard Monte Carlo method with the high accuracy of estimations maintained. The LHS method maintains the probability distribution assigned to all variables being simulated and makes allowing for the degree of correlation among them possible. Several modifications improving the method accuracy and at the same time reducing the time intensity of simulations were developed and adopted. The LHS method finds its use in many fields, among which geotechnics including underground construction industry has its position. The LHS use for structural analyses of underground structures may significantly improve the accuracy of the conception of the anticipated behaviour of a structure being assessed (first of all regarding the probability of occurrence of extreme states).

This paper was carried out with the support by grants TAČR TA01011816 and GAČR P105/12/1705.

**RNDr. TOMÁŠ SVOBODA, Ph.D., svoboda.tomas@3-g.cz,
3G Consulting Engineers, s. r. o.,
DOC. ING. MATOUŠ HILAR, Ph.D., hilar@3-g.cz,
3G Consulting Engineers, s. r. o., a FSv ČVUT**

LITERATURA / REFERENCES

- Flores, A. N., Entekhabi, D., Bras, R. L. Reproducibility of soil moisture ensembles when representing soil parameter uncertainty using a Latin Hypercube-based approach with correlation control. *Water Resources Research*, Vol. 46, 2010, 13 pp.
- Florian, A. An efficient sampling scheme. *Updated Latin Hypercube Sampling Probabilistic Engineering Mechanics*, Vol. 7, 1992, pp. 123–130.
- Hamm, N. A. S., Hall, J. W., Anderson, M. G. Variance-based sensitivity analysis of the probability of hydrologically induced slope instability. *Computers & Geosciences*, Vol. 32 2006, pp. 803–817.
- Hilar, M. Numerická analýza tektonicky porušeného horninového masivu s primární výstrojí při aplikaci NRTM. Disertační práce, 2000, ČVUT, Praha.
- Hrubešová, E., Aldorf, A., Ďuriš, L., Vojtasík, K., Svoboda, J. Pravděpodobnostní přístup ke statickému a stabilitnímu řešení ostění tunelu Valík. Sborník konference Podzemní stavby Praha 2003. pp. 131–138.
- Huntington, D. E., Lyrantzis, C. S. Improvements to and limitations of Latin hypercube sampling. *Probabilistic Engineering Mechanics*, Vol. 13, No. 4, 1998, pp. 245–253
- Iman, R. L., Conover, W. J. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics – Simulations and Computing*, Vol. 11, No. 3, 1982, pp. 311–334,
- Keramat, M., Kielbasa, R. Modified Latin Hypercube Sampling Monte Carlo (MLHSMC) Estimation for Average Quality Index. *Analog Integrated Circuits and Signal Processing*, Vol. 19, 1999, pp. 87–98.
- Morris, M., Mitchel, T. Exploratory design for computer experiments. *Journal of Statistical Planning and Inference*, Vol. 43, pp. 381–402.
- Parák, T. Použití pravděpodobnosti a matematické statistiky při výpočtech geotechnických konstrukcí. Disertační práce, 2009, ČVUT, Praha.
- Pruška, J., Šedivý, M. Aplikace metody latinských hyperkrychlí. *Geotechnika*. 2010, roč. 13, č. 3–4, s. 31–34. ISSN 1211-913X.
- Sallaberry, C. J., Helton, J. C., Hora, S. C. Extension of Latin hypercube samples with correlated variables. *Reliability Engineering and System Safety*, Vol. 93, 2008, pp. 1047–1059.
- Scheuer, E. M., Stoller, D. S. On the generation of normal random vectors. *Technometrics*, Vol. 4, pp. 278–281.
- Vaněčková, V. Výpočet stability skalních svahů. Disertační práce, 2008, ČVUT, Praha.
- Vořechovský, M., Novák, D. Correlation control in small-sample Monte Carlo type simulations I. A simulated annealing approach. *Probabilistic Engineering Mechanics*, Vol. 24, Issue 3, 2009, pp. 452–462.
- Vořechovský, M. Extension of sample size in Latin Hypercube Sampling with correlated variables. In Michael Beer, M. Rafi L. Muhanna, Robert L. Mullen, editors, REC 2010, *Proc. of 4th International Workshop on Reliable Engineering Computing*, Singapore, 2010, pp. 353–368.
- Ye, K. Q. Column Orthogonal Latin hypercubes and their application in computer experiments. *Journal of American Statistical Association*, Vol. 93, 1998, pp. 1430–1439.