

## Optimization framework for calibration of constitutive models enhanced by neural networks

Rafał F. Obrzud<sup>1,\*</sup>,<sup>†</sup>, Laurent Vulliet<sup>1</sup> and Andrzej Truty<sup>2</sup>

<sup>1</sup>*Soil Mechanics Laboratory, École Polytechnique Fédérale de Lausanne, EPFL, Lausanne, Switzerland*

<sup>2</sup>*Institute of Geotechnics, Cracow University of Technology, PK, Cracow, Poland*

### SUMMARY

A two-level procedure designed for the estimation of constitutive model parameters is presented in this paper. The neural network (NN) approach at the first level is applied to achieve the first approximation of parameters. This technique is used to avoid potential pitfalls related to the conventional gradient-based optimization techniques, considered here as a corrector that improves predicted parameters. The feed-forward NN (FFNN) and the modified Gauss–Newton algorithms are briefly presented. The proposed framework is verified for the elasto-plastic modified Cam Clay model that can be calibrated based on standard triaxial laboratory tests, i.e. the isotropic consolidation test and the drained compression test. Two different formulations of the input data to the NN, enhanced by a dimensional reduction of experimental data using principal component analysis, are presented. The determination of model characteristics is demonstrated, first on numerical pseudo-experiments and then on the experimental data. The efficiency of the proposed approach by means of accuracy and computational effort is also discussed. Copyright © 2008 John Wiley & Sons, Ltd.

Received 6 August 2007; Revised 20 February 2008; Accepted 23 February 2008

**KEY WORDS:** back analysis; parameter identification; elasto-plastic model calibration; neural networks; principal component analysis; optimization

### 1. INTRODUCTION

Numerical simulations have become a standard tool in geotechnical engineering, used in the design and monitoring of existing constructions. The complexity of the boundary value problems (BVPs) to be solved in geotechnical practice is mainly concerned with a highly non-linear behavior of geomaterials. In the past decade many advanced constitutive models for soils were formulated to reproduce stress–strain characteristics exhibited by soils. Apart from hydro-mechanical burdens, soil behavior can also be a result of other forms of loading, e.g. varying temperature, saturation effects, etc. Each model enhancement yields a larger number of material parameters that can be

---

\*Correspondence to: Rafał F. Obrzud, LMS-ICARE, Station 18, EPFL, CH-1015 Lausanne, Switzerland.

<sup>†</sup>E-mail: rafal.obrzud@epfl.ch

obtained by means of laboratory testing and/or *in situ* measurements. With the rising complexity of models and the increasing number of parameters, there seems to be less confidence in assessing material parameters [1]. Some parameters can be estimated directly through closed-form solutions; others can be found only through time-consuming trial-and-error and curve-fitting procedures [2]. Consequently, the difficulties related to the assessment of parameters deter professional engineers using advanced soil models, even if they are recommended for a specific geotechnical task. Hence, there is an understandable demand for efficient automated identification toolboxes for commercial applications, which may encourage practical engineers to use more elaborate models and run numerical computations with a higher level of confidence.

In this paper, a two-level identification strategy of assessing model characteristics for elastoplastic models is investigated. This strategy is based on the feed-forward neural network (FFNN) approach and gradient-based optimization (GBO).

One of the first papers concerning constitutive modeling by means of neural networks (NNs) [3] encouraged researchers to begin exploring artificial intelligence abilities, adopting the NNs to solve various mechanical problems. Ghaboussi and co-workers originally proposed an NN-based framework for constitutive modeling in geomechanics [4, 5]. They introduced a concept of nested adaptive NNs, which considers the nested structure of the material test data, e.g. dimensionality, stress path dependency or drainage conditions. By means of the finite element (FE) method and the autoprogressive training algorithm proposed in [6], they trained NNs with experimental non-uniform triaxial test data, in order to capture and reproduce the non-linear response of the soil without conventional concepts of the theory of plasticity. In addition, further research proved that the NN-constitutive models can be successfully embedded within the FE codes to compute the consistent tangent stiffness matrix [7, 8]. Hashash *et al.* [8] demonstrated that a tangent stiffness matrix can be derived from the NN-based material models, using the explicit formulation represented by network parameters. However, the main drawback of the NN-constitutive models is that it is valid only for a specific material for which a new NN has to be adopted each time. Moreover, a material model loses its 'flexibility', which is inherent in the case of conventional models and which is controlled by parameters explicitly describing concepts of plasticity, such as yield surface, flow rule and hardening law.

As NNs are able to generalize a multi-variable complex relationships, they can be applied to solve typical geotechnical problems such as prediction of slope movements [9], prediction of deflections of diaphragm walls [10] or mapping geophysical measurements of the complex permittivity of a soil–water electrolyte system onto soil properties such as water content and degree of saturation or density [11]. The NN technique has also been applied to determine the limit state surface for reliability analysis [12]. Recently, NNs have been used to solve inverse problems such as the identification of various mechanical properties from structural tests [13, 14] or some geotechnical characteristics from *in situ* experimental data [15]. Shin and Pande [14] proposed an identification of the orthotropic elastic constants by forming a set of equations representing components of the computed tangent non-symmetric stiffness matrix for the NN-based constitutive model with the unknown elements of the conventional orthotropic stiffness matrix. Pichler *et al.* [15] also suggested using genetic algorithms for initial determination of network weights at the beginning of NN training. Such an improved training algorithm may reduce the probability that the gradient-based training algorithm will approach a local minimum during the minimization of the global error for NN approximations. However, this algorithm would unnecessarily complicate the one-time procedure of NN training, as network parameters can be adjusted within a few trial training runs.

Generally, the problem of parameter identification lies in proposing a set of characteristics, which would minimize the difference between the quantities obtained from the stress–strain analysis and experimental measurements. Such an inverse problem can be classically solved using the direct approach, which requires processing the standard stress analysis in an iterative manner [16, 17]. As the non-linear problem may prove to be very complicated to solve analytically, the FE analysis is often used as an engine; hence, the direct problem can be formulated as

$$\mathbf{F}_{\text{ext}}(t) - \mathbf{F}_{\text{int}}(\mathbf{u}, \mathbf{b}, t) = 0 \quad \text{s.t. boundary condition and } \mathbf{b} \in \mathcal{R}_b \quad (1)$$

where  $\mathbf{F}_{\text{int}}(\mathbf{u}, \mathbf{b}, t)$  is the internal force vector that evolves in time  $t$  and depends on model variables  $\mathbf{b}$  adhered to the physical parameter space  $\mathcal{R}_b$ ,  $\mathbf{u}$  is the displacement vector and  $\mathbf{F}_{\text{ext}}(t)$  denotes the vector of external forces. Usually, the back analysis is performed by means of automated procedures involving suitable gradient-based minimization techniques such as conjugate gradient [18], Newton [19, 20] or quasi-Newton [2] algorithms. Performing the regression analysis, the level of disagreement between numerical experimental data can be measured by a certain objective function  $S$  [16, 17, 21–24]. Clearly,  $S$  depends on the vector of model parameters  $b_j$ , where  $j = 1, \dots, \text{NP}$  and the problem of parameter identification is equivalent to minimization of the objective function

$$S(\mathbf{b}) = \frac{1}{t_1 - t_0} \int_{t_0}^{t_1} \|y(t) - y'(\mathbf{b}, t)\| dt \quad \text{where } \mathbf{b} \in \mathcal{R}_b \quad (2)$$

where  $t_1 - t_0$  is the length of the observation time,  $y(t) - y'(\mathbf{b}, t)$  denotes the difference between experimental data and calculated quantities and NP is the number of parameters.

The GBO techniques directly search for the minimum of  $S(\mathbf{b})$  using its value and derivatives with respect to each parameter  $X_{ij} = \partial y'_i / \partial b_j$ . In the case of a non-linear problem, the gradient of  $S(\mathbf{b})$  is computed employing, for instance, the FE analysis. Taking into account an iterative manner of the direct search, this procedure requires a high computational effort. Moreover, the number of FE analyses can be multiplied by the number of structural tests considered in the calibration procedure. The minimization is terminated if a strong local minimum is met. Hence, the final solution strongly depends on the accuracy of the user-specified initial vector of parameters  $\mathbf{b}^0$ . Starting from different initial vectors, the inverse problem may exhibit the non-uniqueness of the solution because the algorithm may yield different local minima. Moreover, in certain cases, the ill-posed initial vector may lead to instability of the FE analysis because of the algorithm search for the solution in the inadmissible space of parameters. Consequently, the function  $S$ , losing its continuity, is no longer differentiable. Thus, the identification procedure should involve time-consuming trial-and-error runs using various starting points in order to find the global minimum.

In order to avoid potential pitfalls related to GBO, there is a need for methods that approximate the final solution, i.e. that find the vicinity of the global minimum. Typically, statistical methods can be used to solve an inverse problem. The mixed formulation of genetic algorithms and NNs [15] or the stochastic approach [25] can be found in the literature. Recently, genetic algorithms have also been adapted to solve the model calibration problems, cf. [26–28]. However, the efficient application of genetic algorithms requires generating a large number of individuals.

Hereafter, the NN approach is proposed to search the entire multi-dimensional space of parameters and to approximate soil model characteristics. Based on experimental data, an approximated solution is generated by NN and then optimized through the gradient-based minimization. This strategy will be verified on the triaxial compression tests for which a numerical specimen will be

considered as a material point and will be subjected to a homogeneous stress state. Numerical results will be used both to generate NN training samples and to perform the regression analysis. The methodology of parameter identification is presented in detail in the following section.

The following sections present components of the proposed identification strategy, i.e. the NN tool with two different approaches of a network input vector definition, also with an application of the principal component analysis (PCA), and, finally, an adopted optimization algorithm. For the sake of transparency of the demonstrated identification strategy, the simplest critical state model was adopted in order to clearly present procedure routines. Thus, the modified Cam Clay (MCC) model is briefly recalled [29]. It is also advantageous that all of the MCC model characteristics have physical meaning so that a reliable comparison between values obtained numerically and by hand can be demonstrated. The performance of the method is presented on benchmark laboratory triaxial tests, i.e. isotropic consolidation (CI) and consolidated isotropic drained (CID) compression. These simple homogeneous and one-phase BVPs were deliberately chosen to illustrate the method to be as transparent as possible and to allow a direct analytical assessment of parameters for the comparative analysis. The determination of the parameters of the model is initially presented for numerical pseudo-experiments and finally for experimental data.

## 2. IDENTIFICATION METHODOLOGY

The FFNN approach is proposed to be the first step of the calibration of elasto-plastic constitutive models. The FFNN algorithm is used as a predictor that instantly maps the experimental data of a given test onto model parameters. Such a prediction constitutes the initial vector of parameters used later on by GBO regarded as the correction step. Such a prediction can be regarded as an inverse solution of the identification problem whereas the optimization involves the direct approach. The strategy of the two-level parameter identification is schematically presented in Figure 1.

The self-organizing ability of NNs is used to prepare the prediction tool. The supervised NN training (see Section 3) is performed only once based on the results from numerical pseudo-experiments. The numerical model that is used to generate numerical measurements should reflect the BVP of the real system. The sufficient number of training patterns allows NN to gain enough knowledge in order to be able to generalize further inputs. Numerical simulations are carried out for random variations of model parameters,  $b_j$ ,  $j = 1, \dots, NP$  and  $\mathbf{b} \in \mathcal{R}_b$ . Vectors of parameters constitute the NN training target sets  $\mathbf{t} = \{\mathbf{b}_k^T\}$ , i.e. the values of parameters expected for the corresponding results of previously computed  $NS^{(t)}$ , the number of pseudo-experiments, where  $k = 1, \dots, NS^{(t)}$ . Correspondingly, the numerical results are collected in the training input set  $\mathbf{p} = \{\mathbf{y}_k^T\}$ , where each vector  $\mathbf{y}_k^T(\mathbf{b}_k)$  contains ND, the number of discrete measurements.

As constitutive models approximate only real soil behavior, the numerical simulations tend to idealize the test response. Hence, the numerical response may never be fitted ideally to the real curve. This suggests that numerical data would lead to the overfitting of NN and, consequently, to an improper prediction. Hence, NN can be correctly trained by means of two approaches of discretization. The first approach assumes that only gradient-delivering discrete points are chosen and provided as the training input. An example will be presented in Section 6.3. The second method consists of introducing noise into the smooth computed equilibrium paths using, for instance, the Gaussian probability density function (pdf) and presenting to the network all observational measurements. In this case, the mean value of the noise is prescribed as equal to zero in each

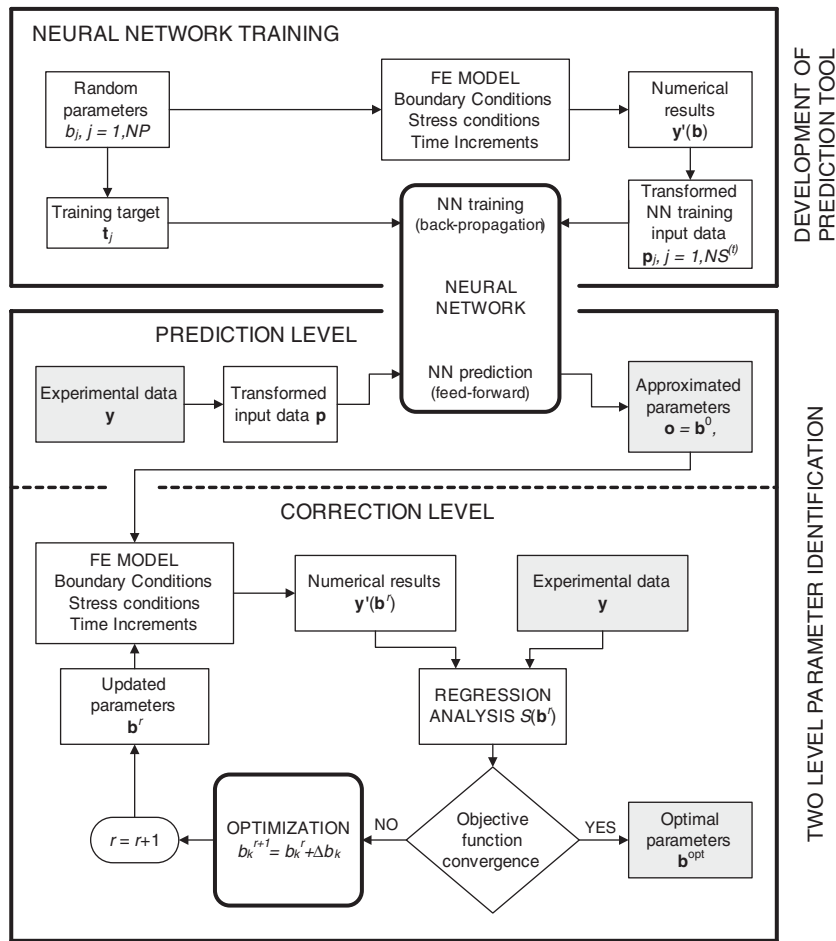


Figure 1. Scheme of the two-level parameter identification with the development of a prediction tool.

equilibrium point whereas the standard deviation can be derived from a variability of the data dispersion observed for experimental curves. The noise simultaneously introduces material and geometrical imperfections regarded, among others, as sources of geotechnical uncertainty [12]. As the whole range of measurements of an experiment is considered, a dimensional reduction in data is desired. This can be easily carried out by applying PCA, which transforms data into a new coordinate system [30].

The choice of the method may depend on the possible level of data fitting, which is influenced by smoothness of the response of a model subject to given boundary conditions.

For a considered constitutive model, NN training can be regarded as the one-time process, only if the broad spans of parameter values are covered, for the generation of training patterns. Obviously, a further development of an NN inverse model is not precluded, and any expected parameter extensions can be reconsidered to retrain the existing NN. Hence, with the single computational effort related to the NN model generation, the efficiently trained NN becomes a robust prediction

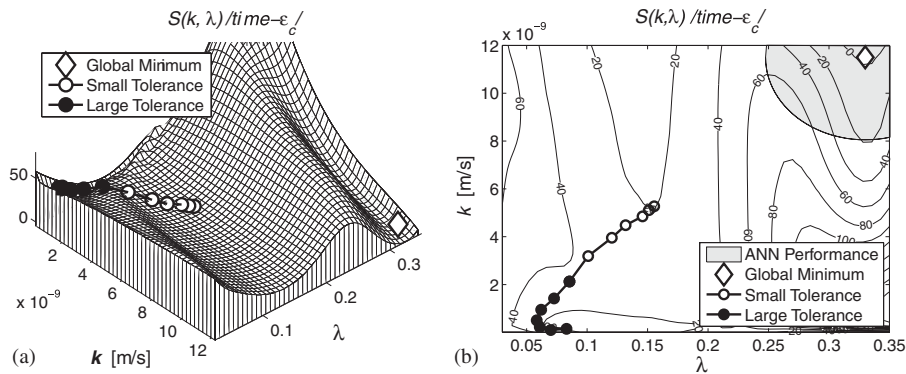


Figure 2. An example of trapped optimization before reaching the global minimum and the possible space of prediction with neural networks: (a) mesh and (b) contour.

tool ready to solve any inverse problem for a variety of soil stress–strain responses for a given BVP. The activated trained NN with experimental data will respond, generating the vector of material properties  $b_j^0$ . Then, the quality of predicted parameters can be easily controlled by performing the regression analysis of experimental data  $\mathbf{y}$  and numerical results  $\mathbf{y}(\mathbf{b}^0)$ . As NN is created to seek the vicinity of the global minimum, the identified parameters may need only slight corrections using the GBO technique. The relatively poorly recognized parameters  $b_k$  can be quickly updated within  $r$ -iterations with the substantially reduced computational effort.

Considering the high complexity of non-linear soil constitutive models and the number of parameters included, the performance of a traditional trial-and-error calibration with a large number of tests is often computationally expensive and does not always lead to satisfactory results. These problems related to GBO can be clearly illustrated using an example of parameter identification for the cavity expansion BVP. Figure 2 presents a highly non-linear character of the error function with respect to consolidation characteristics (permeability coefficient  $k$  and slope of the normal consolidation line (NCL)  $\lambda$ ). The presented error function is associated with the variation of cavity strain in time, for a simulation of the pressuremeter pressure holding test. In order to transparently illustrate the problem, the error function was extracted from the overall error function, including the error function related to the second curve provided by the test, i.e. pore pressure dissipation in time. It is shown that starting from a remote point, the calibration can be finished early if the small tolerance for one of the convergence criteria is set. Despite the raising of the tolerance threshold, the procedure can be terminated if the broad valley in the functional is met. On the other hand, Figure 2(b) shows that the possible region of network prediction is close to the global minimum. The area of a potential prediction can be easily deduced from the NN post-training analysis.

### 3. NN AS THE PREDICTION STEP

The concept of an artificial NN was inspired by the complex architecture of the human brain regarded as a highly non-linear, parallel operating system [31]. Each NN is composed of an assembly of single interconnected processing units that represent neurons of a human brain

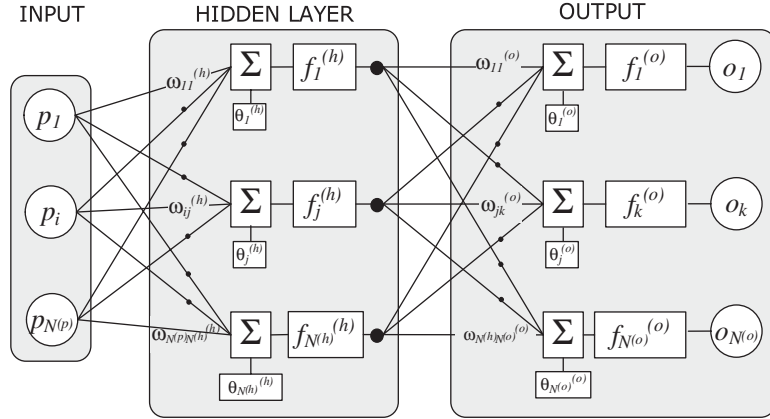


Figure 3. One-hidden-layer neural network topology.

(Figure 3). An elementary neuron consists of synaptic weights,  $\omega_{ij}^{(n)}$ , which store ‘knowledge’ gained by the network and compute the excitation level in the summing junction as

$$\xi_j^{(n)} = \sum_{i=1}^N \omega_{ij}^{(n)} p_i + \theta_j^{(n)} \quad \text{where } j = 1, \dots, M \quad (3)$$

where  $n$  is an identity of the layer,  $N$  is the number of elements in the input vector,  $M$  is the number of neurons in the  $n$  layer and  $\theta_j = \omega_{0j}^{(n)}$  denotes a bias acting as a form of threshold. The output of the  $j$ th processing unit is the unity value transformed by the non-linear differentiable transfer function, e.g.

$$f(\xi) = \frac{2}{1 + e^{-2\xi}} - 1 \quad (4)$$

As  $N$  single units can be interconnected through  $n$  hidden layers, they can create a collective multi-layered network that is able to solve even highly non-linear problems.

### 3.1. Feed-forward procedure

The multi-layered network is designed hierarchically and consists of hidden layers. For the sake of simplicity, the network with a single hidden layer is considered in the following theoretical review. Each layer comprises processing units that are fully interconnected with units of the next layer (Figure 3). The input layer consisting of  $N^{(p)}$  input units activates a hidden layer containing  $N^{(h)}$  neurons. Then the signal is transmitted across outgoing connections to the  $N^{(o)}$ -dimensional output layer. In other words, the feed-forward algorithm maps the dimensionless values of the input vector  $p_i$ ,  $i = 1, \dots, N^{(p)}$  into the output vector  $o_k$ ,  $k = 1, \dots, N^{(o)}$ . The output vector for a network with one hidden layer can be expressed thus

$$o_k = f_k^{(o)} \left( \sum_{j=1}^{N^{(h)}} \omega_{jk}^{(o)} f_j^{(h)} \left( \sum_{i=1}^{N^{(p)}} \omega_{ij}^{(h)} p_i + \theta_j^{(h)} \right) + \theta_k^{(o)} \right) \quad (5)$$

This formula can be regarded as the closed-form solution to a given problem if network weights are correctly adjusted. Such a designed NN with an ability to generalize can reconstruct an unknown complex non-linear function that represents any posed problem.

It may happen that input and output vectors contain data comprising physical values often varying in order of magnitude. In such cases, it may be useful to scale/normalize the input and output vectors into dimensionless values falling into an interval  $(-1, 1)$ .

### 3.2. NN training

Apart from the ability to generalize, NN possesses an important feature of self-organizing in the light of a posed problem. The self-organization requires adjusting network weights by means of performing supervised training. NN training is equivalent to a minimization of discrepancies between output values produced by the untrained NN and target solutions considered as a part of a training set.

In the supervised training  $NS^{(t)}$ , the number of representative patterns comprising input vectors and corresponding targets is used to adjust network weights until NN satisfactorily approximates a function. This is usually performed by one of the GBO techniques. The quality of network performance can be validated by means of  $NS^{(s)}$ , the number of simulation patterns, that has not been presented to the network before. As mentioned above, each training pattern comprises the input vector  $p_i$  within  $N^{(p)}$ , the number of elements ( $i = 1, \dots, N^{(p)}$ ), and the corresponding output vector  $o_k$  of  $N^{(o)}$  elements ( $k = 1, \dots, N^{(o)}$ ). During the first activation of the untrained network, the network weights are chosen randomly, and therefore, discrepancies between each output vector  $o_k$  and target vector  $t_k$  are expected. Hence, the global error function to be minimized,  $F^{(g)}$ , for the total number of training sets,  $NS^{(t)}$ , can be calculated as follows:

$$F^{(g)} = \frac{1}{NS^{(t)}} \sum_{t=1}^{NS^{(t)}} F_t = \frac{1}{NS^{(t)}} \sum_{t=1}^{NS^{(t)}} \frac{1}{N^{(o)}} \sum_{k=1}^{N^{(o)}} (t_k(\mathbf{p}) - o_k(\mathbf{p}))^2 \rightarrow \text{minimize} \quad (6)$$

where  $t$  denotes the index of the training pattern ( $t = 1, \dots, NS^{(t)}$ ).

NN training aims at minimizing the global error by iterative adjustment of the network weights according to the generalized delta rule [32]. The global error is gradually reduced in each network activation, using one of the gradient descent optimization algorithms. Partial derivatives of the global error with respect to the weights,  $\omega_{jk}^{(o)}$  and  $\omega_{ij}^{(h)}$ , determine the gradient of weight increments,  $\Delta\omega_{jk}^{(o)}$  and  $\Delta\omega_{ij}^{(h)}$ , by means of the generalized delta rule. Hence, the learning process is referred to the back-propagation term as the derivative of the global error is back-calculated contrary to the direction of calculations of the activated network. The increments of weights are functions of the partial derivatives of the global error with respect to weights

$$\Delta\omega_{jk}^{(o)} = \Delta\omega_{jk}^{(o)} \left( -\eta \frac{\partial F^g}{\partial \omega_{jk}^{(o)}} \right) \quad \text{and} \quad \Delta\omega_{ij}^{(h)} = \Delta\omega_{ij}^{(h)} \left( -\eta \frac{\partial F^{(g)}}{\partial \omega_{ij}^{(h)}} \right) \quad (7)$$

where  $\eta$  denotes the learning-rate parameter, superscripts  $(o)$  and  $(h)$  concern, respectively, the output and the hidden layers whereas  $i, j, k$  denote the number of elements of a relevant layer presented in Figure 3,  $i = 1, \dots, N^{(p)}$ ,  $j = 1, \dots, N^{(h)}$ ,  $k = 1, \dots, N^{(o)}$ . For details see the literature on the topic (e.g. [31, 32]).



Prior to NN training, the network topology is designed arbitrarily containing  $N^{(h)}$ , the number of neurons in hidden layers. The number of units highly depends on the complexity of a formulated task. This can be influenced by the number of elements in the input and output vectors, as well as by a dimension of a training set. Usually, the number of neurons in hidden layers remains constant for a given training trial and is modified as necessary if the performance of the network is not satisfactory. However, the search of an optimal topology can be automatized using a dynamic update of the architecture during the training process [4, 33]. In our approach, optimal network topologies have been found with a few trial training runs.

### 3.3. Principal component analysis

As mentioned above, training data samples are generated by means of FE analyses. Each analysis is run in the incremental format and produces rows of numerous numerical measurements. Many of these measurements, provided also with artificial noise imitating a natural scatter, can be correlated and, in the context of parameter identification, may be redundant. Hence, it is useful to compute the most meaningful basis to filter out the noise from numerical and further experimental measurements and to extract the most meaningful information from a large data set. PCA is a statistical method that reduces multidimensional data sets to lower dimensions [31]. Reducing a large data set, PCA reveals a simplified hidden structure that can be nested in the test measurements. PCA is an orthogonal linear transformation that transforms the data to a new coordinate system so that the greatest variance of the projected data lies on the first principal axis, the second greatest variance on the second axis, etc. Ignoring the higher-order principal components, the most affecting data in the analysis are retained by lower-order ones.

Supposing that column vectors  $\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_n$  contain scaled observations of  $m = ND$  variables of  $n = NS$  stress–strain analyses, means and deviations along each dimension  $m$  are mapped, respectively, onto 0 and 1 according to the following equation:

$$\hat{\mathbf{x}}_j = \frac{(\mathbf{x}_j - E[\mathbf{x}_j])}{\sqrt{E[\mathbf{x}_j^2] - (E[\mathbf{x}_j])^2}} \quad \text{and} \quad j = 1, \dots, m \quad (8)$$

The empirical correlation matrix ( $m \times m$ ) can be estimated as [34]

$$\mathbf{R} \cong \frac{1}{n} \sum_{i=1}^n \hat{\mathbf{x}}_i \hat{\mathbf{x}}_i^T = \frac{1}{n} \mathbf{X} \mathbf{X}^T \quad (9)$$

where  $\mathbf{X} = [\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_i, \dots, \hat{\mathbf{x}}_n]$ . Then, using the orthogonal similarity of transformation

$$\mathbf{W}^T \mathbf{R} \mathbf{W} = \mathbf{V} \quad (10)$$

the matrix of eigenvectors  $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_j, \dots, \mathbf{w}_m]$  and the diagonal matrix of eigenvalues  $\mathbf{V} = \text{diag}[v_1, \dots, v_j, \dots, v_m]$  can be computed and arranged according to the descending order. The principal component vectors can be thus computed as

$$\mathbf{y}_j = \mathbf{w}_j^T \hat{\mathbf{x}}, \quad j = 1, \dots, l \quad (11)$$

where  $l$  denotes the reduced number of observations  $1 \leq l < m$  under the condition that  $v_l \gg v_{l+1}$ . This operation transforms the  $\mathcal{R}_m$  space into reduced dimension  $\mathcal{R}_l$  retaining those variables that contribute most to the variance.

## 4. OPTIMIZATION AS THE CORRECTION STEP

For the purpose of this research, from a variety of optimization techniques, the modified Gauss–Newton (MGN) algorithm, reported in [23], has been adopted as the corrector for the NN-assigned parameters. This iterative standard regression algorithm is regarded as an effective technique when the weighted least-square objective function is adopted to measure discrepancies between discrete measurements  $\mathbf{y}$  and  $\mathbf{y}'(\mathbf{x}|\mathbf{b})$

$$S(\mathbf{b}) = [\mathbf{y} - \mathbf{y}'(\mathbf{x}|\mathbf{b})]^T \cdot \boldsymbol{\omega} \cdot [\mathbf{y} - \mathbf{y}'(\mathbf{x}|\mathbf{b})] \quad (12)$$

where the diagonal weighting matrix  $\boldsymbol{\omega}$  stores uncertainties of measurements assigning the importance of each variable. The MGN commences the iterative minimization of  $S(\mathbf{b})$  with the initial vector  $\mathbf{b}^0$  assigned by NN. At each iteration  $r$ , the algorithm searches for the objective function gradient  $\mathbf{d}^r$  solving linear equations

$$(\mathbf{C}^T \mathbf{X}_r^T \boldsymbol{\omega} \mathbf{X}_r \mathbf{C} + m_r \mathbf{I}) \mathbf{C}^{-1} \cdot \mathbf{d}^r = \mathbf{C}^T \mathbf{X}_r^T \boldsymbol{\omega} (\mathbf{y} - \mathbf{y}'(\mathbf{b}^r)) \quad (13)$$

where  $\mathbf{X}$  is the sensitivity matrix evaluated directly by FE subroutine using forward finite difference scheme;  $\mathbf{C} = (\mathbf{X}^T \boldsymbol{\omega} \mathbf{X})_{kk}^{-1/2}$  is a diagonal scaling matrix;  $m_r$  is the Marquardt parameter and  $\mathbf{I}$  is the identity matrix. The Levenberg–Marquardt term  $m_r \mathbf{I}$  is introduced to accelerate the convergence performance. As suggested in [23], the Marquardt parameter is prescribed initially as  $m_0 = 0$  and updated in each iteration according to the rule  $m_{r+1} = 1.5m_r + 0.001$ .

At each iteration, the vector of parameters is updated as follows:

$$\mathbf{b}^{r+1} = \rho_r \cdot \mathbf{d}^r + \mathbf{b}^r \quad (14)$$

where the scaling parameter  $\rho_r$  is introduced to control the length of  $\mathbf{d}^r$  so that the maximal absolute value of the fractional parameter value  $\rho_r = (b_j^{r+1} - b_j^r) / |b_j^r|$  changes less than the user-specified tolerance  $d_{\max}$ . The scaling parameter is a scalar so that the search direction  $\mathbf{d}^r$  is preserved.

The iterative search procedure is accomplished if the largest and absolute change of each optimized parameter in iteration  $r$  is less than the user-specified tolerance  $|d_k^r / b_k^r| \leq \varepsilon_{\text{tol}}, \forall k = 1, \dots, \text{NP}$ .

As the solution domain for the elasto-plastic models exhibits usually a highly non-linear character, the accuracy of the gradient-based algorithm also highly depends on the guess of the initial vector of parameters. Starting from a point that is located far away from the global minimum, the search in the solution domain can be terminated early due to the occurrence of the local minimum. At the same time, the efficiency of the process can be clearly reduced as the number of iterations leading to the optimal solution increases. The choice of the initial vector has also a strong influence on the stability of optimization. An ill-defined starting point may drive the search into a completely opposite direction, causing the instability or lack of convergence of the FE simulation. The use of the NN technique limits the admissible parameter space into the region near the global minimum so that all aforementioned difficulties can be avoided.

## 5. ELASTO-PLASTIC CONSTITUTIVE MODEL

The proposed strategy will be applied to the isotropic elasto-plastic MCC model [29]. Despite certain limitations this model is the simplest critical state 1 and can be used for normally consolidated

or lightly overconsolidated clay. Furthermore, it is often the kernel of more complex models such as multi-mechanism models [35]. Although the feasibility of the identification technique is investigated with the MCC model, it can be adopted for much more complex constitutive models. Such a simple model was selected to reliably verify the proposed parameter identification strategy. Although most of the model parameters have a clear physical meaning and can be directly estimated for the considered BVPs, in the case of numerical parameter identification, they can be artificially perceived as ‘non-physical’ properties. This is so because all these parameters may control and affect the stress–strain response of the model. This assumption is introduced only for the demonstrative purpose because it is obvious that the MCC parameter estimates can be easily assessed using the conventional triaxial test curves.

The yield and plastic potential surface is defined in terms of effective stress invariants  $p' = -I_1/3$  and  $q = \sqrt{3J_2}$  and the preconsolidation pressure  $p_{co}$  that defines the overconsolidation ratio  $R_p = p_{co}/p'$ . The equation for the state boundary surface for MCC (cf. Figure 4(b)) is given as

$$F(\sigma', p_{co}) = q^2 + M^2 \cdot p'(p' - p_{co}) = 0 \quad (15)$$

The constitutive model is characterized by five model parameters  $M$ ,  $\lambda$ ,  $\kappa$ ,  $p_{co}$  and  $\nu$ . The  $M$  parameter is the slope of the critical state line (CSL) in the  $p'$ – $q$  plane (Figure 4) and can be expressed as  $M = 6 \sin(\phi'_{tc}) / (3 - \sin(\phi'_{tc}))$ , where  $\phi'_{tc}$  is the effective friction angle determined from the triaxial compression test,  $\nu$  is the Poisson ratio,  $\lambda$  and  $\kappa$  denote, respectively, the slope of NCL and unload–reload compression line in isotropic compression tests (Figure 4(a)). In the model, lines of virgin loading and swelling are approximated linearly in the  $e - \ln p'$  axes, which implies consequently a linear relationship between bulk modulus,  $K$ , and the effective mean stress,  $p'$

$$K = K(p') = \frac{1 + e_0}{\kappa} p' \quad (16)$$

where  $e_0$  denotes the initial voids ratio. In that case,  $G/K = \text{const.}$  and is expressed by the formula

$$\frac{G}{K} = \frac{3}{2} \frac{1 - 2\nu}{1 + \nu} \quad (17)$$

Hence, the current shear modulus is computed through

$$G = \frac{3(1 - 2\nu)(1 + e_0)p'}{2(1 + \nu)\kappa} \quad (18)$$

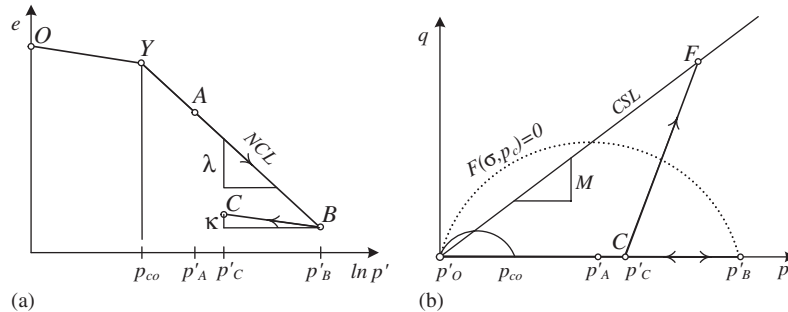


Figure 4. Scheme of test stress paths in (a) isotropic and (b) deviatoric plane.

The total strain increment (rate) is decomposed into elastic,  $\dot{\boldsymbol{\varepsilon}}^e$ , and plastic parts,  $\dot{\boldsymbol{\varepsilon}}^p$

$$\dot{\boldsymbol{\varepsilon}} = \dot{\boldsymbol{\varepsilon}}^e + \dot{\boldsymbol{\varepsilon}}^p \quad (19)$$

where the effective stress rate is computed with the aid of the elastic stiffness matrix  $\mathbf{D}^e$

$$\dot{\boldsymbol{\sigma}}' = \mathbf{D}^e \cdot \dot{\boldsymbol{\varepsilon}}^e \quad (20)$$

$$\mathbf{D}^e = (K - \frac{2}{3}G)\mathbf{I} + G2\mathbf{I}^s \quad (21)$$

The  $\mathbf{I}$  symbol represents the unity tensor ( $I_{ijkl} = \delta_{ik}\delta_{jl}$ ) and  $\mathbf{I}^s$  is the symmetrized unity tensor ( $I_{ijkl}^s = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})/2$ ). Plastic strains occur when the stress state reaches the state boundary surface given by Equation (15) and the flow rule is assumed to be associated

$$\dot{\boldsymbol{\varepsilon}}^p = \dot{\lambda} \frac{\partial F}{\partial \boldsymbol{\sigma}} \quad (22)$$

where  $\dot{\lambda}$  is a plastic multiplier.

The hardening parameter modifies the state boundary surface and its evolution connected to the irreversible (plastic) strains,  $\varepsilon_v^p$ , is given by:

$$\dot{p}_c = -\frac{1+e_0}{\lambda-\kappa} p_c \dot{\varepsilon}_{kk}^p \quad (23)$$

## 6. EXAMPLES OF NUMERICAL IMPLEMENTATION

### 6.1. Selected BVPs

In order to investigate the robustness of the proposed identification scheme on a well-defined BVP, routinely performed triaxial tests, i.e. (i) CI test, (ii) CID compression test, were adopted to compare the quality of analytically and numerically evaluated/predicted characteristics of the MCC model. These two BVPs were also chosen to illustrate the adoption of two approaches for the training input discretization that have been presented in Section 2.

Performing the CI test for the preconsolidated material, consolidation characteristics  $\kappa, \lambda, p_{co}$  can be obtained if the test is run through the elastic and plastic domains. The simulation of CI does not require running FE analyses due to the linear solution in  $\ln p' - e$  space (Figure 4(a)). To obtain a numerical response of the model, each simulation contains five isotropic equilibrium paths (points  $O, Y, A, B, C$ , see Figure 4(a)) corresponding to two measured variables, i.e. the mean pressure at points,  $p'_O, p_{co}, p'_A, p'_B, p'_C$ , and the volumetric strain,  $\varepsilon_O^v, \varepsilon_Y^v, \varepsilon_A^v, \varepsilon_B^v, \varepsilon_C^v$ . Assuming that the slopes of consolidation lines can be evaluated using the gradients of  $A-B$  and  $B-C$  paths, the pressure  $p'_A$  is kept constant for each numerical simulation whereas the ratios  $p'_B/p'_A$  and  $p'_C/p'_B$  are randomly imposed within the interval presented in Table I. Introducing a random variable of  $p_{co}/p'_A$  that imposes an occurrence of plastic straining, various degrees of preconsolidation are taken into account. Furthermore, to allow the identification to be executable regardless of the applied stress increments, a normalization of mean stresses  $p'_i$  with respect to the maximal mean stress  $p'_{max}$  was introduced applying dimensionless variables  $p'_i/p'_{max}$ . A discussion on the choice of space of parameters as well as details concerning NN training will be presented in the following subsections.

Table I. Intervals of variables considered in NN training.

Variable	Interval	
	CI test	CID test
$e_0$		(0.10, 1.20)
$v$	—	(0.25; 0.40)
$M$	—	(0.50; 1.50)
$\lambda$		cf. Figure 5
$\Lambda$		(0.50; 0.95)
$p_{co}/p'_A$	(0.2; 0.8)	—
$p'_B/p'_A$	(1.5; 2.1)	—
$p'_B/p'_C$	(1.1; 16.0)	(1.1; 2.2)

Although the simulation of CI does not require running FE analyses due to the simplicity of the solution, the axisymmetrical simulations of the drained compression test were performed with the use of an FE model. Simulations of the compression under triaxial conditions were carried out using the MCC driver within the Z-Soil v.6 code [36]. In each computation, a strain-controlled compression test was performed on the isotropically preconsolidated material subject to the stress path  $C-F$  (Figure 4b). The degree of preconsolidation was imposed through initial stress conditions  $\sigma_3 = p'_C$  and the random values of  $R_p = p'_B/p'_C$  within the interval depicted in Table I. This interval ensures the occurrence of yielding on the ‘wet’ side of boundary surface. For each simulation, three variables, i.e. axial strain  $\varepsilon_1$ , volumetric strain  $\varepsilon_v$  and deviatoric stress  $q$  were computed in 31 increments until  $\varepsilon_1 = 20\%$ . Small increments were imposed at the beginning of the test and then were gradually extended toward the end of the test in order to capture the elastic gradient regardless of the chosen set of parameters as well as to ensure the accurate convergence of the FE analysis. For practical reasons, the normalization of the measured  $q$  was applied so that the identification could be executed regardless of the magnitude of the initial confining pressure  $p'_C$

$$\mathbf{q}^N = \mathbf{q}/p'_C \quad (24)$$

## 6.2. Space of parameters

The choice of the admissible space of parameters  $\mathcal{R}_b$  plays an important role in NN training. Parameters should be chosen from ranges that are observable in geotechnical practice in order to (i) reduce the parameter domain, (ii) regularize the distribution in this domain and (iii) avoid the non-physical vector of variables. The reduction in the multi-dimensional space of parameters may lead to a significant decrease in the number of generated patterns with a simultaneous improvement of accuracy. This may be of utmost importance if the large-scale FE models are considered in pattern generation.

The intervals adopted in the generation of training patterns are presented in Table I. Typical correlations between  $e_0$  and  $\lambda$  for clayey soils reported in [37–41] can be taken between the upper and lower bounds, cf. Figure 5. Observable relationship between  $\lambda$  and  $\kappa$  represented through the plastic volumetric strain ratio  $\Lambda = 1 - \kappa/\lambda$  may fall into an interval  $\Lambda \in (0.5; 0.9)$  [42]. The intervals for the remaining model characteristics,  $M$  and  $v$ , were set according to commonly observable values for clay.

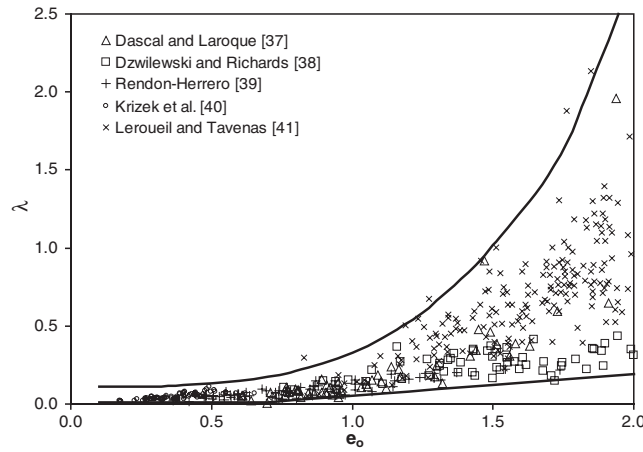


Figure 5. Typical observable relationship between  $e_0$  and  $\lambda$ .

Among other methods reported in [43], the random method of selecting training patterns was adopted. Parameter and stress state distributions within the intervals were assumed to be uniform except for the voids ratio  $e_0$ , for which the exponential pdf was adopted to generate random numbers due to the logarithmic approximation of the upper bound of  $\lambda$  (Figure 5). The exponential distribution is given by

$$P(x|\mu) = \int_0^x \frac{1}{\mu} e^{t/\mu} dt = 1 - e^{-x/\mu} \quad (25)$$

where  $\mu$  is the mean value,  $\sigma$  the standard deviation and  $x$  the random value from uniform distribution. The log-normal distribution of  $e_0$  is to ensure a uniform distribution in the  $\lambda$  interval.

### 6.3. Parameter identification for the CI test

For the purpose of the identification of soil properties from the CI test, a two-hidden-layer FFNN was optimally designed for the 5-component input as NN(5|7-7|3), i.e. seven neurons in each hidden layer for 3-output parameters. The optimal number of neurons was reduced to a minimum to avoid an over-fitting tendency of the network [33]. Apart from the training set containing  $NS^{(t)} = 130$  numerical patterns, a set of 70 pseudo-experiments was prepared additionally to illustrate the efficiency of the approximation. For each numerical experiment  $p'_A, p'_B, p'_C$  and the corresponding  $\varepsilon_v$  were measured. Thus, the dimensionless input vectors corresponding to the measured variables and the outputs regarded as unknown parameters are, respectively, given as

$$\mathbf{p}_t = [\varepsilon_A^v, \varepsilon_B^v, \varepsilon_C^v, p'_A/p'_B, p'_C/p'_B]^T \mapsto \mathbf{NN} \mapsto \mathbf{o}_t = [\kappa, \lambda, p_{co}/p'_B]^T \quad (26)$$

where  $t = 1, \dots, NS^{(t)}$ . Note that the proper value of  $p_{co}$  is obtained by multiplying the dimensionless component by the maximum applied mean pressure  $p'_B$ . To improve the performance of NN training, the input and output vectors were scaled so that their components fall into  $(-1, 1)$  interval.

The quality of the approximation performance was assessed by presenting unknown measurements to the trained network (Figure 6). The linear regression analysis performed after 70 activations

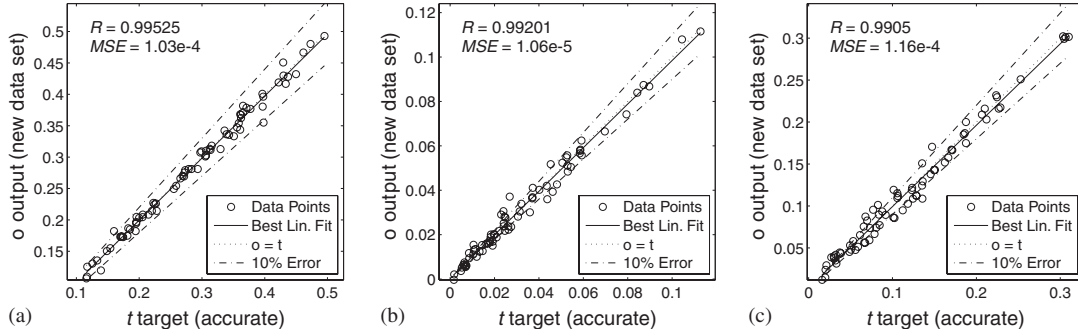


Figure 6. Isotropic consolidation BVP: post-training regression analysis for the data set never presented to the trained network: (a)  $p_{co}/p'_B$  prediction; (b)  $\kappa$  prediction; and (c)  $\lambda$  prediction.

of NN exhibited a satisfactory agreement between accurate and NN-predicted values of  $p_{co}$ ,  $\lambda$ ,  $\kappa$ , with the correlation coefficient  $R$  equal to 0.99 for each parameter and small values of mean-squared error (MSE) (cf. Equation (6)). Moreover, the NN inverse models exhibit the best linear fits that overlap the perfect linear fits  $o=t$ . A visual inspection of post-training analysis shown in Figure 6 confirms the acceptable degree of approximation, generally falling in 10% bounds of the relative error for target values. An example of the parameter identification for real experimental measurements is presented in Section 7.2.

#### 6.4. Parameter identification for the drained compression test

For each numerical simulation, three groups of variables were incrementally computed, i.e. axial strain  $\varepsilon_a$ , deviatoric stress  $q$  and volumetric strain  $\varepsilon_v$ . From among 32 equilibrium paths for  $\varepsilon_a \in \langle -0.2, 0 \rangle$ , the post-failure results were skipped, i.e.  $\varepsilon_a \in \langle 0, \varepsilon_f^a \rangle$  where  $\varepsilon_f^a$  corresponds to  $q_t = q_{max}$ . Then, 31 discrete observational points  $\tilde{\varepsilon}_a$  were randomly chosen with corresponding interpolated variables  $\tilde{q}$ ,  $\tilde{\varepsilon}_v$ . Moreover,  $\tilde{q}$  was normalized into dimensionless variable  $\tilde{q}^N$  (Equation (24)). Moreover, in order to avoid the overfitted solutions and improve generalization, the artificial noise was introduced into ‘smooth’ numerical results for each discrete observation of  $\tilde{q}$ ,  $\tilde{\varepsilon}_v$ . The random noise was introduced using the Gaussian pdf, with the coefficient of variation  $c_v = \sigma/\mu$  equal to 3% for any numerical measurement. This artificially burdens the numerical results with testing uncertainties, i.e. device measurement errors, geometrical imperfections and natural variability of an individual specimen.

Each training pattern consisted of 3-element target vector representing the identified parameters  $M$ ,  $\lambda$ ,  $\kappa$  and the dimensionless input vector that comprised three normalized state variables (cf. Equation (8)), i.e. Poisson’s ratio  $\nu$ , initial voids ratio  $e_0$  and overconsolidation ratio  $R_p$ , and the noisy measurements of three variables

$$\hat{\mathbf{p}}_t = [\hat{\nu}, \hat{e}_0, \hat{R}_p, \hat{\varepsilon}_1^a, \dots, \hat{\varepsilon}_{ND}^a, \hat{\varepsilon}_1^v, \dots, \hat{\varepsilon}_{ND}^v, \hat{q}_1^N, \dots, \hat{q}_{ND}^N]^T \quad (27)$$

where ND=31 discrete measurements. Therefore, the extensive dimension of the 96-component input vectors for NS=175 training patterns was reduced by invoking PCA before NN training. The empirical correlation matrix (Equation (9)) and the transformation matrix were calculated

(Equation (10)). The input vector was thus reduced from 96 to  $NR = 15$  components. The calculated variances of the corresponding 15 dominant principal components for the training set under study were 90.510, 71.287, ..., 4.128 for which the total variability explained by dominant principal components was 82.76%. The reconstruction error  $E_r$  [31] corresponding to the variance of 16th principal component is thus:

$$E_r = E[\|\mathbf{X} - \mathbf{W}^T \mathbf{Y}\|] = 3.262 \quad (28)$$

where the matrix  $\mathbf{X} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_{NS}]$  contains all data with  $m = 96$  scaled observations in each vector, while the matrix  $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{NS}]$  contains  $NS$  dimensionally reduced vectors where each  $NR$ -component vector is the projection of the corresponding data vector from matrix  $\mathbf{X}$  onto the basis vectors contained in the columns of matrix  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_{NR}]$ . Thus, taking into account only 15 principal components the extensive 96 data input is significantly reduced, preserving a small reconstruction error.

Then, a two-hidden-layer NN with the input vectors projected on the principal axes was optimally designed with six neurons in each hidden layer and the training set comprised 175 training sets. The identification can thus be expressed as

$$\hat{\mathbf{p}}_t \mapsto (\text{PCA})\mathbf{p}_t \mapsto \text{NN}(15|6-6|3) \mapsto \mathbf{o}_t = [M, \kappa, \lambda]^T \quad (29)$$

The quality of the network performance was assessed with a set of 75 results of the random pseudo-experiments comprising noisy measurements never presented to the network. The results of the performed regression analysis presented in Figure 7, indicate the fair correctness of the parameters' prediction in terms of the correlation coefficient  $R$  and MSE. It is also indicated that the best linear fits and the perfect models  $o = t$  exhibit conformity. A visual inspection of  $M$  predictions shows an excellent performance of the estimator whereas the predictions of remaining parameters reveal a slight scatter outside of 10% of relative error of target values. This can be explained by the fact that the measurements comprised artificial noise. However, such an accuracy is absolutely sufficient to find a neighborhood of the global solution. An example of parameter identification for real experimental measurements is presented in the following section.

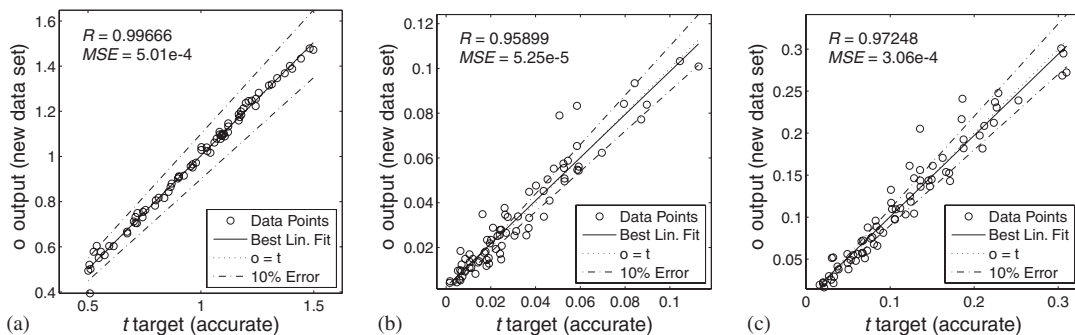


Figure 7. Drained compression BVP: post-training regression analysis for the data set never presented to the trained network: (a)  $M$  prediction; (b)  $\kappa$  prediction; and (c)  $\lambda$  prediction.



## 7. PARAMETER IDENTIFICATION FOR REAL EXPERIMENTAL DATA

The following section demonstrates an application of the identification strategy for real experimental data. Model parameters are conventionally evaluated based on measurements coming from laboratory tests carried out by the authors of this paper, using a triaxial chamber. The compressed specimens were subject to boundary conditions as described in Section 6.1.

### 7.1. Material and test procedure

A series of CI and drained triaxial compression tests was performed on the Boleiy clayey silt samples (liquid limit  $w_L=31.8\%$ ; plastic limit  $w_P=16.9\%$ ; unit weight of skeleton  $\gamma_s=27.1\text{ kN/m}^3$ ; clay fraction 27%). The remolded and homogenized specimens were initially formed in a 80-mm-diameter tube. The 120 mm specimens were first isotropically consolidated with the pressure of  $p_c=100\text{ kPa}$  and then unloaded and carefully trimmed into the standard dimensions of  $38 \times 76\text{ mm}$ . The experiment comprised (i) one triaxial CI test with the following sequence of isotropic loads  $p'=15, 30, 60, 120, 240$  and  $480\text{ kPa}$  and unloading to  $p'=280\text{ kPa}$  and (ii) three triaxial drained shear tests preceded by the CI of specimens to  $p'_B=150, 210$  and  $180\text{ kPa}$  and unloading with the same overconsolidation ratio  $p'_B/p'_C=1.25$ . In the case of CID, the standard shear paths  $dq/dp'=3$  were applied (Figure 4(b)). The initial value of voids ratio for CI test was  $e_0=0.535$ , while for CID tests  $e_0$  at the beginning of shearing paths was, respectively, equal to  $0.478, 0.488$  and  $0.482$ . Based on the measured variation of  $\varepsilon_3$  against  $\varepsilon_1$ , the calculated mean value of the Poisson ratio was equal to  $\nu=0.368$ .

### 7.2. Example of parameter identification for the CI test

An identification of the three variables  $\kappa$ ,  $\lambda$  and  $p_{co}$  was carried out by presenting to the trained network the variations of the volumetric strain  $\varepsilon_v$  of the last three measurements and corresponding mean pressures  $p'$  normalized with respect to the maximal one. An illustrative comparison of the test measurements and the model response using the NN-predicted parameters is shown in Figure 8.

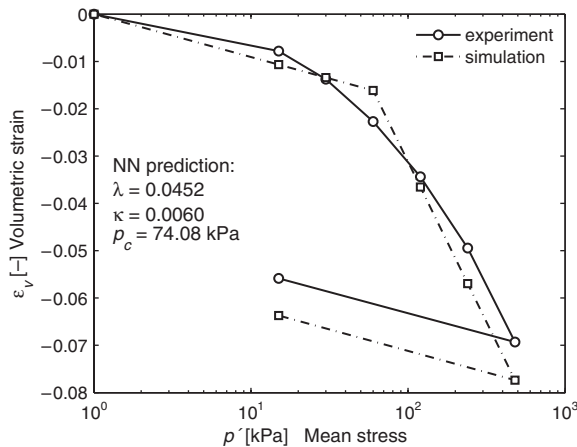


Figure 8. Comparison between experimental data and the model response using the NN-predicted parameters  $\lambda$ ,  $\kappa$ ,  $p_{co}$  for the isotropic consolidation test.

Table II. Identified parameters for isotropic consolidation test.

Variable	Analytical back-calculation value	NN prediction	
		Value	Error (%)
$\kappa$	0.0060	0.0060	0.00
$\lambda$	0.0440	0.0452	0.27
$p_{co}^*$	77.00 kPa	74.08 kPa	3.79

The specimens preconsolidated with real pressure  $p_c = 100$  kPa.

Values of NN-identified parameters are also compared with the values that were back-calculated using the conventional method, i.e. the linear regression of loading curves. The results presented in Table II show a good approximation of the identified parameters with the maximal relative error of 3.79% corresponding to  $p_{co}$ . Note that the identified  $p_{co}$  is smaller than the real consolidation pressure ( $p_{co} = 100$  kPa) due to the earlier mobilization of plastic strains and possibly the trimming effect.

### 7.3. Example of parameter identification for the drained compression test

A numerical identification of three variables  $\kappa$ ,  $\lambda$  and  $M$  was carried out separately for each test by presenting to the trained network variations of three measured variables, i.e. axial strain  $\varepsilon_a = \varepsilon_1$  and corresponding volumetric strain  $\varepsilon_v$  and normalized deviatoric stress  $q^N$  taken within  $\varepsilon_a \in (-0.12, 0)$ . Thirty-one measurements of each variable completed with the values of the Poisson ratio, voids ratio and preconsolidation ratio were collected in the input vector  $\mathbf{p}$ . Then, using the transformation matrix  $\mathbf{W}_i^T$  the normalized input  $\hat{\mathbf{p}}_i^T$  (Equation (27)) was projected onto the subspace spanned by the eigenvectors and presented to the network. The NN-determined parameters, their mean values and relative errors with reference to optimal values are presented in Table IV. The identified parameters of the second test were taken to illustrate the quality of the prediction in comparison with the values obtained through the optimization. This analysis shows the effect of the measured data discretization on the evaluated parameters. Two optimization runs were invoked, i.e. (i) for all 31 discrete points along the whole interval of  $\varepsilon_1$  and (ii) for the chosen points within the elastic and failure domains (the points in between were skipped). Equal importance of  $q$  and  $\varepsilon_v$  measurements was assumed by introducing weights as the inverse of error variance where the errors were adopted as equal to  $e_q = 0.5$  kPa and  $e_\varepsilon = 0.0001$  (cf. [44]). During iterations, only identified parameters were updated. The results presented in Figure 9 show that the fitting of all discrete points leads to the underestimating of  $M$  with reference to the maximal value of  $q/p'$ . The relative error that appeared reaches 4.5%, see Table III. On the other hand, the fitting of arbitrarily chosen fitting zones leads to the value of  $M$  for the maximal  $q/p'$  with a loss of the degree of the fitting in the middle of  $\varepsilon_1$ - $q$  curve. Clearly, the choice of the discretization approach will depend on the assumed criterion imposed by a further geotechnical problem. It is worth noting that the model response with NN gives a very close prediction especially for  $\varepsilon_1$ - $q$  variation (the error of the identified parameter  $M$  is equal to 0.16%). This demonstrates that NN additionally preceded by PCA is able to extract by itself the most important information from the whole family of measurements. It also means that the network weights provide the sensitivity of model parameters. Moreover, NN considers the overall material behavior, and not only its specific state, which can be illustrated by the sensitivity of NN to the dilatant behavior of the soil at the final part of the

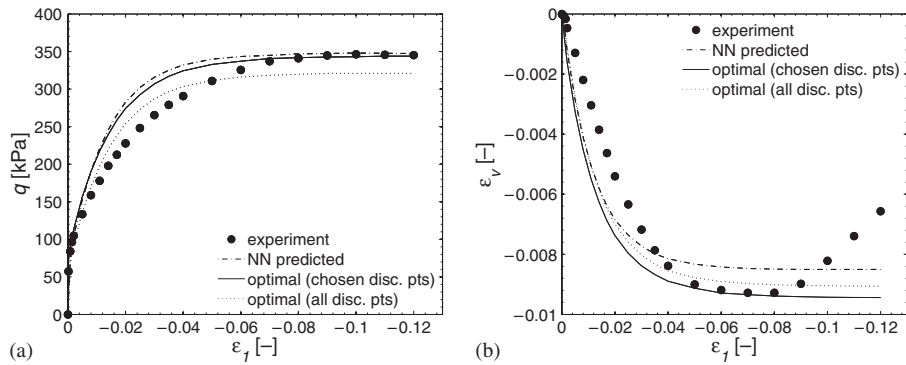


Figure 9. Optimization with the NN-predicted initial vector of parameters for the drained compression test ( $p'_0 = 120 \text{ kPa}$ ). A comparison of fitting for the discrete points of the arbitrarily significant importance and for all the discrete measurements.

Table III. Comparison of the identified parameter  $M$  by means of the analytical back-calculation, NN prediction and optimization with important discrete points and an entire vector of measurements.

Variable	$\max(q/p')$ value	NN prediction		Optimal: chosen discrete measurements		Optimal: all discrete measurements	
		Value	Error (%)	Value	Error (%)	Value	Error (%)
$M$	1.223	1.221	0.16	1.218	0.41	1.168	4.50

$\varepsilon_1$ – $\varepsilon_v$  curve. NN considers the overall variation of  $\varepsilon_v$ , providing an averaged value of  $\lambda$ , which is smaller than the value identified for the maximal volume change (cf. Table IV).

Hereafter, the fitting procedures are carried out merely for chosen discrete points in the elastic and failure regions. The results of the three separate parameter identification runs and one mutual optimization containing data of three tests are presented in Table IV. In the table, the relative errors of the identified parameters for the NN predictions and analytical calculations are calculated with respect to the values obtained through the subsequent GBO runs. A surprisingly large error of the evaluated value of  $\kappa$  can be explained by the high order of the identified parameter and this is not significantly reflected in the model response presented in Figure 10. The NN-recognized parameters are quickly updated with the minimum computational cost within two iterations. Figure 10 reveals the limitations of the MCC model, which makes it impossible to obtain a perfect agreement between experimental and numerical data.

It can be noted that the optimal values of  $\kappa$  and  $\lambda$  obtained from consolidation and compression tests are inconsistent. This is due to prescribing the value of  $R_p = 1.25$  with the measured confining pressures  $p'_B$  and  $p'_C$ . This state parameter could be incorporated into the optimized vector under the condition that additional results of CI are included.

#### 7.4. Efficiency test

The quantitative test of effectiveness of the proposed two-level scheme was performed on one of the compression tests starting from different remote points chosen arbitrarily. As a large change of

Table IV. Comparison of variables evaluated by means of analytical back-calculation, NN prediction and optimization for three drained compression tests: 1,  $p'_0 = 120$  kPa; 2,  $p'_0 = 168$  kPa; 3,  $p'_0 = 144$  kPa, respectively.

Variable	Analytical back-calculation			NN prediction				Optimization	
	Value	Mean	Errors (%)	Value	Mean	Errors (%)		Single run	Three curves
$\kappa$	0.0013		85.7	0.0019		171.4		0.0007	
	0.0013	0.0012	67.5	0.0030	0.0028	25.0	57.4	0.0040	0.0018
	0.0019		18.8	0.0036		125.0		0.0016	
$\lambda$	0.0010		92.8	0.0137		4.9		0.0144	
	0.0011	0.0104	93.5	0.0156	0.0147	6.6	2.0	0.0167	0.0150
	0.0112		22.2	0.0148		2.8		0.0144	
$M$	1.223		0.4	1.221		0.2		1.218	
	1.240	1.227	0.1	1.209	1.203	2.4	1.5	1.239	1.222
	1.217		0.5	1.180		2.6		1.211	

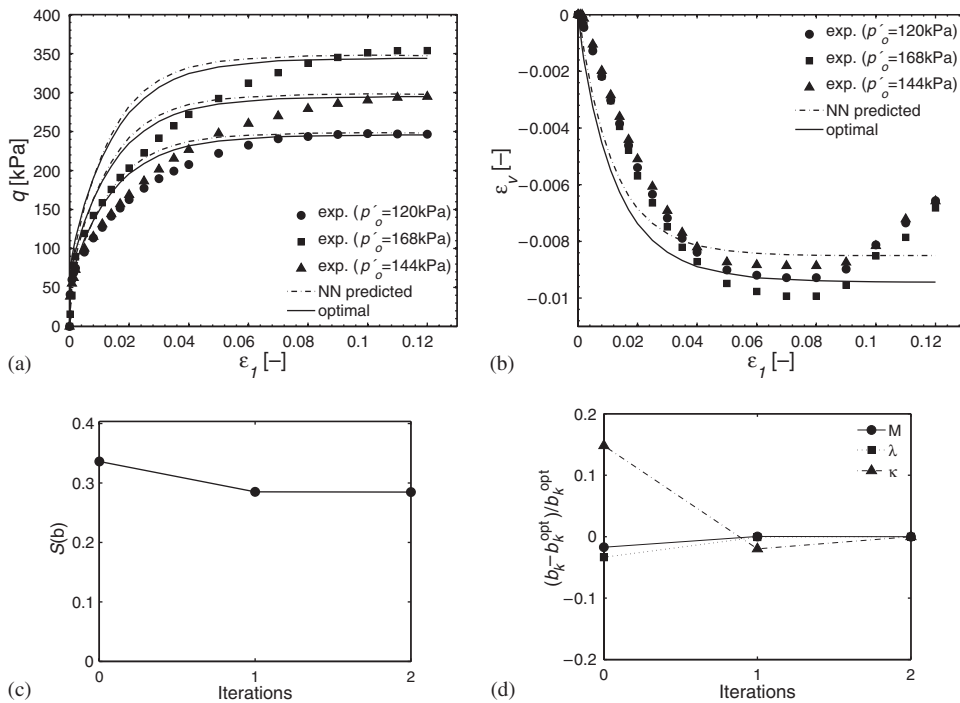


Figure 10. Optimization with the NN-predicted initial vector of parameters: (a), (b) model calibration with experimental data for three drained compression tests; (c) convergence rate of GBO; and (d) convergence rate of the NN-assigned parameters.

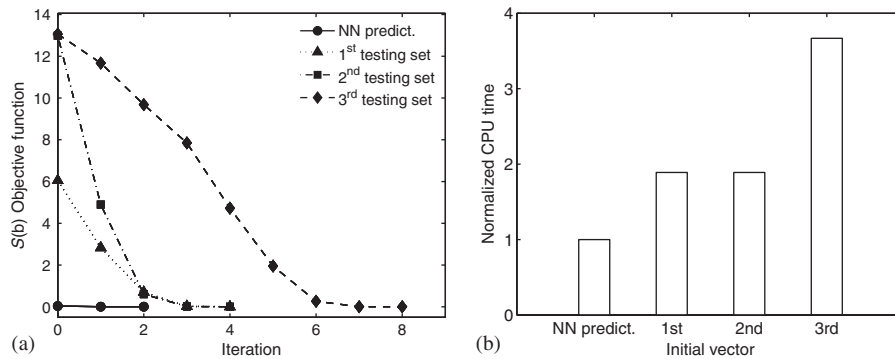


Figure 11. Efficiency test of the optimization preceded by the NN prediction in comparison with three optimization runs with the user-specified initial vector: (a) rate of convergence and (b) computational effort.

parameters may lead to an instability of optimization, convergence was controlled by the scaling parameter prescribed in each run equal to  $\rho_r = 0.5$ . The rate of the convergence was measured for the optimization with the NN-assigned and three testing the user-specified initial vectors. Testing patterns comprised the values of  $M$ ,  $\lambda$ ,  $\kappa$  given, respectively, as  $\mathbf{b}_1 = [0.8, 0.1, 0.01]$ ,  $\mathbf{b}_2 = [1.6, 0.1, 0.01]$  and  $\mathbf{b}_3 = [0.8, 1.0, 0.1]$ . Each minimization converged to the same optimal solution. Note that starting from remote initial points, instabilities of the solution may often occur, for instance, in the case of calibration of hydro-mechanical problems (cf. Figure 2).

Figure 11(a) clearly shows the fast convergence of the optimization with the NN-assigned parameters. Automated minimization is accomplished within two iterations whereas optimizations regarded as trial-and-error examples with user-assigned parameters require at least four iterations. Note that in each iteration, besides the error function computation, the number of FE runs is expanded by the number of optimized parameters if the forward finite difference method is used to evaluate the sensitivity matrix. This number can be multiplied if the model calibration is carried out for several experimental tests simultaneously.<sup>‡</sup> Consequently, the increasing number of iterations results in the increase of computational cost that was augmented for the considered testing sets between 2 and 4 times (Figure 11(b) presents the CPU time that is normalized with respect to the convergence time of the NN-enhanced optimization).

Clearly, the number of iterations may also increase with the enhancement of the constitutive model, resulting in an increase in the number of parameters and experimental curves.

## 8. CONCLUSIONS

This paper presents the efficient generic scheme of the automated calibration procedure of constitutive models. The potential of the two-level parameter identification method combining the standard gradient-based optimization (GBO) technique supported by the FFNN algorithm is presented in detail. The efficiency of the presented method has been proved on the elasto-plastic model where

<sup>‡</sup>Each test requires running an individual model due to different initial stress conditions.

soil specimen is subject to the triaxial boundary conditions. Some general conclusions can thus be drawn

- The application of the neural network (NN) technique to solve an inverse problem provides an efficient and accurate search of the vicinity of a global solution avoiding local minima. The close approximation of parameters preserves the minimal computational expense during the preceding optimization and helps the user to avoid performing time-consuming trial-and-error runs. This is a crucial feature, if the large-scale FE analyses, required in the case of modeling of non-homogeneous stress state BVPs, e.g. field tests are considered during the subsequent direct searching with GBO.
- The choice of an admissible space of parameters plays an important role in NN training leading to the reduction of the multi-dimensional space of parameters and a significant decrease in the number of generated patterns with a simultaneous improvement in the accuracy.
- Incorporating PCA into NN training results in the significant reduction of large-dimensional data with no significant loss of accuracy of evaluated parameters.
- The choice of discrete points during curve fitting has a significant importance. The examples presented in this paper demonstrate that NNs possess the ability to extract meaningful information about the parameters' sensitivity from the training data set.
- A potential application of the NN-based technique for the parameter identification of enhanced constitutive models basically involves delivering relevant data that illustrates the physical phenomena that are considered in the model. Obviously, in the case of a calibration of more complex soil models, the parameters that can be directly estimated with the high degree of confidence can constitute a part of the NN input vector (e.g. the Poisson coefficient or the overconsolidation ratio in the presented example), whereas the characteristics of no physical meaning for a given BVP become the objects of optimization.

Using a well-designed interface, the proposed parameter identification can be adapted to any constitutive model or BVP with only few modifications. Different models or structural tests can be attached as modules to the fully automated pattern generator. The numerical results are used once to properly adjust network weights that are stored with negligible storage cost. As NNs are trained with broad intervals of parameters, they are able to correctly map any experimental set of measurements. Thus, the trained NNs can be part of the automated optimization module incorporated into numerical solution applications such as FE packages. This can be of particular value for multi-parameter models characterized by the complexity of parameter assessment.

Further research is directed at adopting the method for complex BVPs where the efficiency of the existing closed-form solution is restricted due to imposed assumptions. Assembled results of the study of the NN-enhanced parameter optimization for the coupled hydro-mechanical BVP of self-boring pressuremeter tests will be subsequently reported.

#### REFERENCES

1. Graham J. The 2003 R.M. Hardy Lecture: soil parameters for numerical analysis in clay. *Canadian Geotechnical Journal* 2006; **43**:187–209.
2. Anandarajah A, Agarwal D. Computer-aided calibration of soil plasticity model. *International Journal for Numerical and Analytical Methods in Geomechanics* 1991; **15**:835–856.
3. Ghaboussi J, Garrett JH, Wu X. Knowledge-based modeling of material behavior with neural networks. *Journal of Engineering Mechanics* 1991; **117**:132–151.

4. Ghaboussi J, Sidarta DE. New nested adaptive neural networks (NANN) for constitutive modeling. *Computers and Geotechnics* 1998; **22**(1):29–52.
5. Sidarta DE, Ghaboussi J. Constitutive modeling of geomaterials from non-uniform material tests. *Computers and Geotechnics* 1998; **22**(1):53–71.
6. Ghaboussi J, Pecknold DA, Zhang M, Haj-Ali RM. Autoprogressive training of neural network constitutive models. *International Journal for Numerical Methods in Engineering* 1998; **42**:105–126.
7. Shin HS, Pande GN. On self-learning finite element codes based on monitored response of structures. *Computers and Geotechnics* 2000; **27**:161–178.
8. Hashash YM, Jung S, Ghaboussi J. Numerical implementation of a neural network based material model in finite element analysis. *International Journal for Numerical Methods in Engineering* 2004; **59**:989–1005.
9. Mayoraz F, Vulliet L. Neural networks for slope movement prediction. *International Journal of Geomechanics* 2002; **2**(2):153–173.
10. Jan JC, Hung SL, Chi SY, Chern JC. Neural network forecast model in deep excavation. *Journal of Computing in Civil Engineering* 2002; **16**(1):59–65.
11. Ding W, Shang JQ. Neural networks in soil characterization. In *ISC-2 on Geotechnical and Geophysical Site Characterization*, Fonseca AVd, Mayne P (eds). Millpress: Rotterdam, 2004; 889–896.
12. Goh ATC, Kulhawy FH. Reliability assessment of serviceability performance of braced retaining walls using a neural network approach. *International Journal for Numerical and Analytical Methods in Geomechanics* 2005; **29**:627–642.
13. Waszczyszyn Z. Some new results in applications of backpropagation neural networks in structural and civil engineering. In *Advances in Engineering Computational Technology*, Topping BHV (ed.). Civil-Comp Press: Edinburgh, 1998; 173–187.
14. Shin HS, Pande GN. Identification of elastic constants for orthotropic materials from structural test. *Computers and Geotechnics* 2003; **30**:571–577.
15. Pichler B, Lackner R, Mang HA. Back analysis of model parameters in geotechnical engineering by means of soft computing. *International Journal for Numerical Methods in Engineering* 2003; **57**:1943–1978.
16. Sakurai S, Takeuchi K. Back analysis of measured displacements of tunnels. *Rock Mechanics and Rock Engineering* 1983; **16**:173–180.
17. Gioda G, Sakurai S. Back analysis procedure for the interpretation of field measurements in geomechanics. *International Journal for Numerical and Analytical Methods in Geomechanics* 1987; **11**:555–583.
18. Ou CY, Tang YG. Soil parameter determination for deep excavation analysis by optimization. *Journal of Chinese Institute of Engineers* 1994; **17**(5):671–688.
19. Zentar R, Hicher PY, Moulin G. Identification of soil parameters by inverse analysis. *Computers and Geotechnics* 1994; **28**:129–144.
20. Finno RJ, Calvello M. Supported excavations: the observational method and inverse modeling. *Journal of Geotechnical and Geoenvironmental Engineering* (ASCE) 2005; **131**(7):826–836.
21. Caillaud G, Pilvin P. Identification and inverse problems related to material behaviour. In *Inverse Problems in Engineering Mechanics*, Swoboda G (ed.). Balkema: Rotterdam, 1994; 79–86.
22. Ledesma A, Gens A, Alonso EE. Estimation of parameters in geotechnical backanalysis—I. Maximum likelihood approach. *Computers and Geotechnics* 1996; **18**(1):1–27.
23. Poeter EP, Hill MC. Documentation of UCODE, a computer code for universal inverse modeling. *US Geological Survey Water-Resources Investigations Report 98-4080*, 1998.
24. Cividini A, Maier G, Nappi A. Parameter estimation of a static geotechnical model using a Bayes' approach. *International Journal of Rock Mechanics and Mining Sciences and Geomechanics Abstract* 1983; **20**(5):215–226.
25. Cekerevac C, Girardin S, Klubertanz G, Laloui L. Calibration of an elasto-plastic constitutive model by a constrained optimisation procedure. *Computers and Geotechnics* 2006; **33**:432–443.
26. Pal S, Wathugala GW, Kundu S. Calibration of a constitutive model using genetic algorithms. *Computers and Geotechnics* 1996; **19**(4):325–348.
27. Samarajiva P, Macari EJ, Wathugala W. Genetic algorithms for the calibration of constitutive models for soils. *International Journal of Geomechanics* 2005; **5**(3):206–217.
28. Levasseur S, Malecot Y, Boulon M, Flavigny E. Soil parameter identification using a genetic algorithm. *International Journal for Numerical and Analytical Methods in Geomechanics* 2008; **32**(2):189–213.
29. Roscoe KH, Burland JB. On the generalized strain–stress behaviour of 'wet' clay. In *Engineering Plasticity*, Heyman J, Leckie FA (eds). Cambridge University Press: Cambridge, MA, 1968; 535–609.

30. Kaczmarczyk Ł, Waszczyszyn Z. Identification of characteristic length of microstructure for second order continuum multiscale model by Bayesian neural networks. *Computer Assisted Mechanics and Engineering Sciences* 2007; **14**:183–196.
31. Haykin S. *Neural Networks. A Comprehensive Foundation* (2nd edn). Prentice-Hall: Upper Saddle River, NJ, 1999.
32. Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation. In *Parallel Distributed Processing*, Rumelhart DE, McClelland JL (eds). MIT Press: Cambridge, MA, 1986; 287–302.
33. Flood I, Kartam N. Neural networks in civil engineering. I: principles and understanding. *Journal of Computing in Civil Engineering* 1994; **8**(2):131–147.
34. Diamantaras KI, Kung SY. *Principal Component Neural Networks: Theory and Applications*. Wiley: New York, 1996.
35. Hujeux JC. Une loi de comportement pour le chargement cyclique des sols. In *Genie Parasismique*, Davidovici V (ed.). E.N.P.C.: Paris, 1985; 287–302.
36. Z\_Soil.PC 2003. *User Manual: Soil, Rock and Structural Mechanics in Dry or Partially Saturated Media*. ZACE Services Ltd, Software Engineering: Lausanne, 2003.
37. Dascal O, Laroque GS. Caractéristique de compressibilité des argiles du complex Nottaway–Broadback–Rupert (Baie James). *Canadian Geotechnical Journal* 1973; **10**(1):41–63.
38. Dzwilewski PT, Richards AF. Consolidation properties of Wilkinson basin soils. *Journal of Geotechnical Engineering* (ASCE) 1974; **100**(GT10):1175–1179.
39. Rendon-Herrero O. Universal compression index equation. *Journal of Geotechnical Engineering* (ASCE) 1980; **106**(GT11):1179–1200.
40. Krizek RJ, Corotis RB, El-Moursi HH. Probabilistic analysis of predicted and measured settlements. *Canadian Geotechnical Journal* 1977; **14**(1):17–33.
41. Leroueil S, Tavenas F, Le Bihan JP. Propriétés caractéristiques des argiles de l'est du Canada. *Canadian Geotechnical Journal* 1983; **20**(4):681–705.
42. Mayne PW. Cam-Clay predictions of undrained strength. *Journal of Geotechnical Engineering* (ASCE) 1980; **106**(GT11):1219–1242.
43. Chang CC, Chang TYP, Xu YG, To WM. Selection of training samples for model updating using neural networks. *Journal of Sound and Vibration* 2002; **249**(5):867–883.
44. Calvello M, Finno RJ. Calibration of soil models by inverse analysis. In *8th International Symposium on Numerical Models in Geomechanics—NUMOG VIII*, Pande GN, Pietruszczak S (eds). Balkema: Rotterdam, 2002; 107–113.