Acta Polytechnica CTU Proceedings **49**:65–70, 2024
https://doi.org/10.14311/APP.2024.49.0065

VALIDATION OF GEOTECHNICAL LABORATORY TESTS DATA OBTAINED FROM VARIOUS SOURCES

Veronika Pavelcová[a,*], Tomáš Janda[b]

[a] *Czech Technical University in Prague, Faculty of Civil Engineering, Department of Geotechnics, Thákurova 7, 166 29 Prague, Czech Republic*

[b] *Czech Technical University in Prague, Faculty of Civil Engineering, Department of Mechanics, Thákurova 7, 166 29 Prague, Czech Republic*

[*] corresponding author: veronika.pavelcova@fsv.cvut.cz

Abstract. Data validation is a critical step in every data analysis. It is a well-known fact that the results of any data analysis are directly dependent on the input values. This paper works with a worldwide database of laboratory tests collected during four year operation of web calibration application ExCalibre and elaborates on the data correctness. ExCalibre application enables automatic calibration of three critical state models based on data from standard laboratory tests. The paper presents the rules and overall methodology for checking the validity of laboratory tests data and summarises the obtained results.

Keywords: Geotechnical laboratory test, data validation, data processing, error, warning, database.

1. Introduction

General observation of usage non-linear geotechnical material models for soils shows that the majority of practical engineers limit themselves to the use of basic non-linear soil models of the Mohr-Coulomb or Drucker-Prager type when solving standard geotechnical tasks. Although these models are sufficient when describing approximate interaction of the structure and the subsoil, some complex tasks require more sophisticated solutions. Critical state models allow for more accurate behaviour prediction, however, they are not as widely used as they could be. The general lack of knowledge of these advanced material models among practical engineers may be one of the reasons.

The second reason is the significantly higher difficulty of determination of advanced non-linear models input parameters in comparison to the basic models. Critical state models require prescription of parameters describing the behaviour of the given material under isotropic compression along with parameters characterising the current state of the soil.

Presented and previous work presented in *Extracting General Knowledge of Model Parameters for Clays out of Numerous Laboratory Tests* [1] is motivated by this non-straight-forward determination of input parameters and works forward to simplify the determination process by using knowledge extracted out of all of the data collected during four year operation of web calibration application ExCalibre [2–4]. This application enables automatic calibration of three advanced constitutive soil models based on detailed data from triaxial and oedometric tests [5]. After finishing all data preparation steps, Bayesian inference methods [6] will be used to set confidence intervals for each material parameter that governs critical state models.

2. ExCalibre

As mentioned before, ExCalibre web application laboratory test database is the main source of data processed in presented research. The application was developed by Faculty of Civil Engineering of CTU in Prague in cooperation with the Faculty of Science of Charles University. It runs from 2018 and performs automatic deterministic calibration of the hypoplastic model for sands [7], hypoplastic model for clays [8, 9] and modified Cam-clay model [10]. A typical workflow of model calibration using ExCalibre is the following:

(1.) Downloading a laboratory protocol template which is prepared in the form of an MS Excel file. The template differs for fine-grained and coarse-grained soils.

(2.) Filling in the results of laboratory tests. A minimum of one isotropically consolidated undrained (CIUP) triaxial test with pore pressure measurement for clay soils or one isotropically consolidated drained (CID) triaxial test for sandy soils must be filled in for successful calibration. In addition, one standard oedometric (OED) test is necessary.

(3.) Uploading the laboratory protocol spreadsheet to the ExCalibre web application. As a result, the application proposes optimal model parameters and graphs comparing the course of measured and simulated laboratory tests.

(4.) Manual modification of individual model parameters by user if required.

The main advantage of all these critical state models is that they distinguishes between material parameters that are constant for all possible states of a particular soil and state variables that evolve during straining. In this sense critical state models are fundamentally

Parameter	Data type	Note
USCS	string	
Specific gravity	float	[g cm^{-3}]
Liquid limit	float	[%]; only fine-grained soils
Plastic limit	float	[%]; only fine-grained soils
Angle of repose	float	[°]; only coarse-grained soils
Soil gradation	2D array	size (2, n); sieve size [mm] – passing [%]

Table 1. Data extracted from IDX AND GRAD worksheet.

Parameter	Data type	Note
Initial void ratio	float	[-]
Oedometric test data record	2D array	dimension (2, n); axial stress [kPa] – axial deformation [-]

Table 2. Data extracted from OED-NAT-n or OED-REC-n worksheet.

different from the Mohr-Coulomb model which employs distinct values of its material parameters for the same soil in different density states.

3. Protocol structure

As mentioned above, laboratory tests uploaded to ExCalibre application are all in the form of pre-prepared template file in MS Excel. This means that all protocols should contain the same type of data and should keep the same structure. Although template differs for fine-grained and coarse-grained soils, the majority of data worksheets remains the same. The only difference is the type of prescribed triaxial test. The content of the individual worksheets of the workbook will be described gradually.

3.1. NOTES worksheet

The introductory worksheet of any laboratory protocol is named NOTES and it contains basic information about sampling: soil type, sampling location, sampling depth, sampling method and possible notes of the laboratory technician. This worksheet is the same for both fine-grained and coarse-grained soils and no data from this worksheet are currently being further processed.

3.2. IDX AND GRAD worksheet

Another type of the laboratory protocol worksheet is named IDX AND GRAD. This worksheet contains soil classification, index characteristics and grain size analysis. This worksheet is the same for both fine-grained and coarse-grained soils. Extracted data to further statistical processing from this worksheet are summarised in Table 1.

3.3. OED-NAT-n worksheet or OED-REC-n worksheet

The laboratory protocol worksheet named OED-NAT-n contains a record of an oedometric test of naturally consolidated soil. The laboratory protocol worksheet named OED-REC-n contains a record of an oedometric test of the reconstituted soil sample. The symbol n in the label of the worksheet indicates the order of the oedometric test, in other words, it represents an increasing integer sequence. This worksheet is the same for both fine-grained and coarse-grained soils. Extracted data to further statistical processing from this worksheet are summarised in Table 2.

3.4. CIUP-NAT-n worksheet or CIUP-REC-n worksheet

The first type of triaxial laboratory protocol worksheet is named CIUP-NAT-n or CIUP-REC-n. CIUP-NAT-n laboratory protocol worksheet contains a record of a consolidated isotropically undrained triaxial test with pore pressure measurement of a naturally consolidated soil. The laboratory protocol worksheet named CIUP-REC-n contains a record of a consolidated isotropically undrained triaxial test with the measurement of the pore pressure of the reconstituted soil sample. The symbol n in the label of the worksheet indicates the order of the CIUP triaxial test, same as in the case of an oedometric test. This worksheet is used only for fine-grained soils. Extracted data to further statistical processing from this worksheet are summarised in Table 3.

3.5. CID-REC-n worksheet

The other, and last described, type of laboratory protocol worksheet is named CID-REC-n. It contains a record of a consolidated isotropically drained triaxial test of the reconstituted soil sample. The template does not assume the possibility of performing CID triaxial test of naturally consolidated test as soil sampling usually disrupts soil integrity. The symbol n in the label of the worksheet again indicates the order of the triaxial test. This worksheet is used only for coarse-grained soils and extracted data to further statistical processing from this worksheet are summarised in Table 4.

Parameter	Data type	Note
Initial void ratio	float	[-]
Initial mean effective stress	float	[kPa]
Triaxial test data record	2D array	dimension (3, n); axial deformation [-] – mean effective stress [kPa] – axial deviatoric stress [kPa]

TABLE 3. Data extracted from CIUP-NAT-n or CIUP-REC-n worksheet.

Parameter	Data type	Note
Initial void ratio	float	[-]
Radial effective stress	float	[kPa]
Triaxial test data record	2D array	dimension (4, n); axial deformation [-] – radial effective stress [kPa] – axial deviatoric stress [kPa] – volumetric deformation [-]

TABLE 4. Data extracted from CID-NAT-n or CID-REC-n worksheet.

4. Data validation

Data validation is a critical step in data preparation for statistical processing as the results of any data analysis are directly dependent on the input values. Even though pre-prepared templates for laboratory protocols uploaded to ExCalibre calibration application help prevent users error inputs, it cannot be assumed that the uploaded data are free of errors. ExCalibre itself does not implement a data validation at this time. Due to this fact and worldwide variability of users it was necessary to validate all data inputs that will be further statistically processed.

A set of rules that should catch all types of error data was created. All rules has been consulted with the head of the geotechnical laboratory at the Faculty of Science of Charles University, the lab technician with many years of practical experience. Created algorithm works on first-catch principle and distinguishes between two levels of error severity. So called *warnings* draw attention to values that are physically possible but not probable to occur and therefore require increased attention. On the other hand, *errors* notify values that should not occur under any circumstances. However, these rules are not strict. They were created based on years of experience not statistical observation. Another important note is the fact that the rules described below are created for soils which physical behaviour can be represented using Modified Cam-Clay, hypoplastic clay or hypoplastic sand model for soils. They do not consider, for example, highly organic soils, which are not suitable for ExCalibre calibration and would not comply with most of the stated rules.

All controls are performed cyclically in a predetermined logical order for all uploaded laboratory protocols. Then it forms two additional dataframe columns, one with a possible caught warning and the other with a caught error. In the algorithm described, the key of mentioned dataframe corresponds to the name of one laboratory protocol. Key values are assigned so that one line of the dictionary represents characteristic data of one laboratory protocol presented in Section 3. Any automated interventions into protocols are avoided. All necessary data corrections or protocol discarding will be done individually and manually with respect to all laboratory protocol data.

4.1. Warning rules

Overall 19 warning rules were created. Individual rules are listed below. The order in which the rules are listed corresponds to the order in which they are implemented in the script.

4.1.1. IDX and GRAD worksheet

(1.) Specific gravity is out of 2.5–2.9 g cm^{-3} interval.

(2.) Liquid limit is greater than 150.0 %.

(3.) Plastic limit is greater than 60.0 %.

(4.) Angle of repose is lower than 30.0° or greater than 40.0°.

(5.) Angle of repose exists but soil consists of more than 10.0 % of fine grains (smaller than 0.063 mm).

(6.) Number of soil gradation inputs is lower than 5.

(7.) The first soil gradation passing is lower than 100 % but greater than 95.0 %.

(8.) Soil gradation sieve for the first non 100.0 % passing is out of 0.001–63.0 mm interval.

4.1.2. OED-NAT-n worksheet or OED-REC-n worksheet

(1.) Number of oedometric test data inputs is lower than 5.

(2.) Axial deformation is negative.

4.1.3. CIUP-NAT-n worksheet or CIUP-REC-n worksheet

(1.) Number of CIUP test data inputs is lower than 50.

(2.) Maximal axial deformation is lower than 1.0 %.

(3.) Initial mean effective stress is out of 20.0–2 000.0 kPa interval.

(4.) Initial axial deviatoric stress is greater than 10.0 kPa.

4.1.4. CID-REC-n worksheet

(1.) Number of CID test data inputs is lower than 50.

(2.) Maximal axial deformation is lower to 1.0 %.

(3.) Initial radial effective stress is out of 20.0–2 000.0 kPa interval.

(4.) Initial deviatoric stress is greater than 10.0 kPa.

(5.) Absolute value of maximal volumetric deformation is greater than 0.5 %.

4.2. Error rules

Overall 24 error rules were created. Individual rules are listed below. The order in which the rules are listed corresponds to the order in which they are implemented in the script.

4.2.1. IDX and GRAD worksheet

(1.) Specific gravity is out of 2.0–3.2 $\mathrm{g\,cm^{-3}}$ interval.

(2.) Plastic limit is greater than or equal to liquid limit.

(3.) Liquid limit is greater than 500.0 %.

(4.) Plastic limit is greater than 80.0 %.

(5.) Angle of repose is lower than 25.0° or greater than 45.0°.

(6.) Angle of repose exists but soil consists of more than 20.0 % of fine grains (smaller than 0.063 mm).

(7.) Number of soil gradation inputs is 0.

(8.) The first soil gradation passing is lower than 95.0 %.

(9.) Soil gradation passing is increasing.

(10.) Any soil gradation passing is negative.

4.2.2. OED-NAT-n worksheet or OED-REC-n worksheet

(1.) Number of oedometric test data inputs is 0.

(2.) Initial void ratio is out of 0.2–3.5 interval.

(3.) Axial stress is out of 1.0–10 000.0 kPa interval.

(4.) Initial axial deformation is non-zero.

4.2.3. CIUP-NAT-n worksheet or CIUP-REC-n worksheet

(1.) Number of CIUP test data inputs is 0.

(2.) Initial axial deformation is non-zero.

(3.) Axial deformation is decreasing (tolerating noise).

(4.) Axial deviatoric stress is negative.

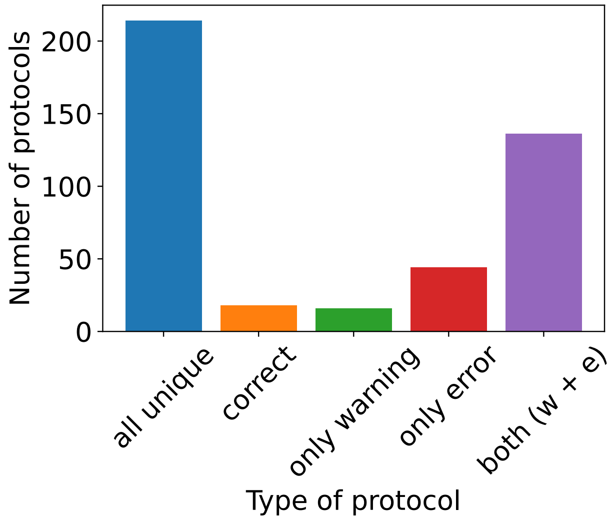


Figure 1. Summary of correct and incorrect laboratory protocols.

4.2.4. CID-REC-n worksheet

(1.) Number of CID test data inputs is 0.

(2.) Initial axial deformation is non-zero.

(3.) Axial deformation is decreasing (tolerating noise).

(4.) Radial effective stress is not constant.

(5.) Axial deviatoric stress is negative (tolerating noise).

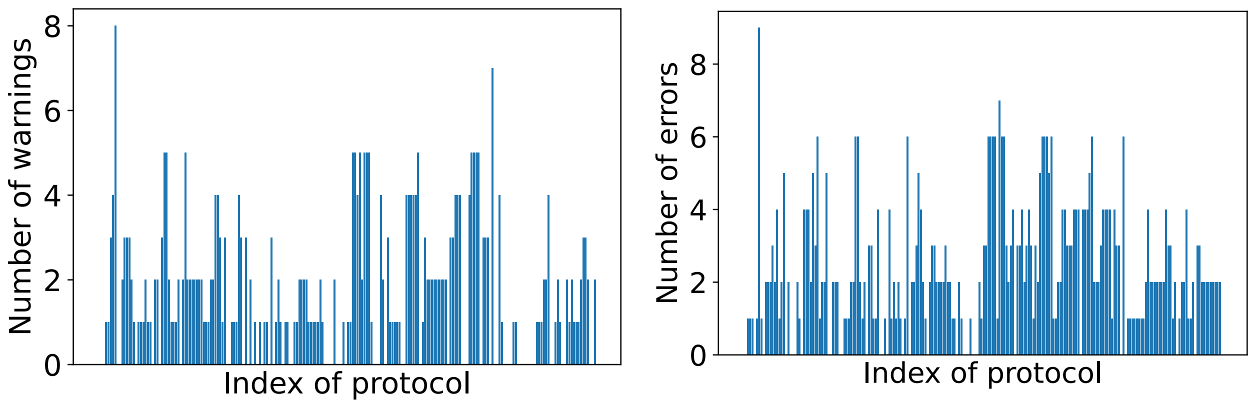
(6.) Initial volumetric deformation is non-zero.

Soil index characteristics and granular analysis depend only on the type of soil. Therefore, it was possible to determine the limit values for individual parameters. On the other hand, the data record of oedometric, and especially triaxial, tests can vary greatly depending on the tested soil and the chosen test procedure. For this reason, it was usually impossible to determine error limits for the individual parameters that govern these tests. Thus, the control rules for these tests were mainly limited to the detection of erroneously entered units of measurement and evident physically impossible results rather than the definition of individual limit values.

5. Results

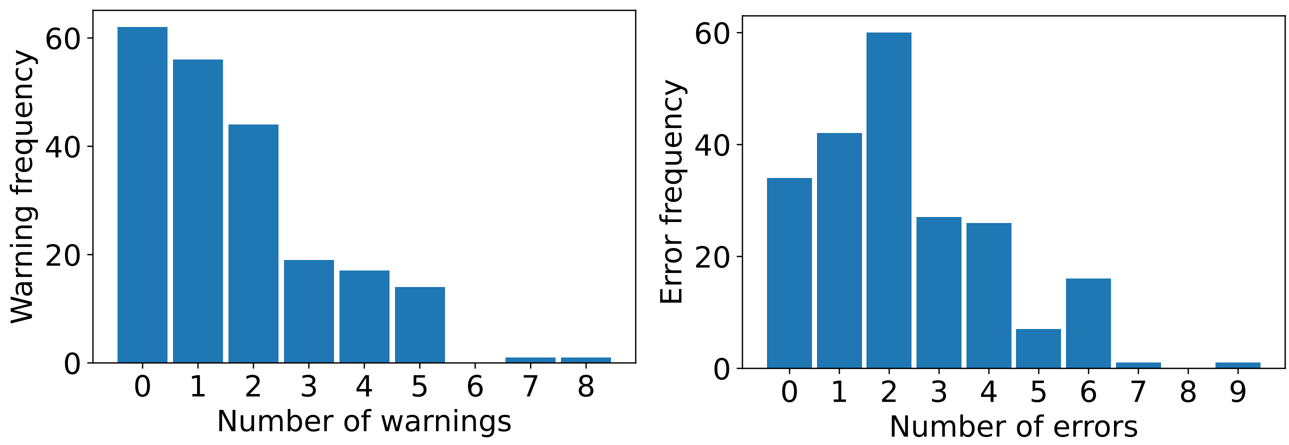
Out of 214 unique laboratory protocols, only 18 seem to be correct as they passed data validation check without any findings. Another 16 protocols have only warnings, no errors and conversely, 44 protocols have only errors and no warnings. Finally, there are 136 protocols left that have both warning's and error's detection. This observation is clearly summarised in Figure 1. However, this figure does not say anything about the amount of errors that each protocol contains.

Figure 2 shows the total detection of invalid data for each protocol, warnings (Figure 2a) and errors (Figure 2b) separately. These counts are related to an



(a). Warnings. (b). Errors.

Figure 2. Number of warnings and errors detected in each laboratory protocol.



(a). Warnings. (b). Errors.

Figure 3. Histograms of warnings and errors detected in laboratory protocols.

index that is unique to each laboratory protocol and with which the protocol is identified. It can be seen that no more than 8 warnings or 9 errors were found in any of the protocols. Majority of protocols reports about 0–2 warnings and 0–4 errors. Histograms in Figure 3 clearly demonstrates this trends.

Table 5 and Table 6 summarise total number of occurrences of individual warnings and errors across all protocols. It can be observed that the most erroneous data were captured in data characterising sandy soils, namely existence of angle of repose for clay soils (control rule 4.1.1.5 and 4.2.1.6) or invalid record of CID triaxial test (control rules 4.1.4 and 4.2.4). Another recurring error are non-zero deformations or unsecured contact of piston and soil sample at the beginning of the deformation tests, most observed during oedometric tests (control rule 4.2.2.3 and 4.2.2.4).

It should be mentioned that the absolute counts of invalid data can be distorted by the repetition of individual problems as part of the control rules are interconnected, meaning that if one data is entered incorrectly and multiple rules are created to check its properties, an error will be displayed for multiple rules, although only one data value is missing or incorrect.

6. Conclusion

The paper introduced the reader to the origin of the researched laboratory protocols, their original use for the calibration of advanced material models in the ExCalibre web application, and the possibility of their future use in statistical analysis. The importance of checking any input data before statistical processing and the effect of invalid data on the results of statistical analysis was pointed out. The content of the geotechnical laboratory test template for ExCalibre application was presented by individual sheets of the Excel workbook in detail. The paper described the methodology chosen by the authors for data validation and specifically states all control rules applied during the control to individual characteristic of laboratory protocol. Finally, the results of the application of the control mechanism to the given data were summarised.

The specific purpose of this work was to give an initial estimation of how the real data collected in the laboratory protocols are realistic and physically possible. A manual control by a natural person is the necessary next step in data control. Attention will be paid only to individual data detected by the control rules listed in Section 4. The correction or removal of

	4.1.1. IDX AND GRAD								**4.1.2. OED**	
No. warning	1.	2.	3.	4.	5.	6.	7.	8.	1.	2.
Frequency	14	1	0	5	56	21	25	4	2	6

	4.1.3. CIUP				**4.1.4. CID**				
No. warning	1.	2.	3.	4.	1.	2.	3.	4.	5.
Frequency	40	11	4	19	28	35	3	44	38

TABLE 5. Frequency of each warning detected in laboratory protocols.

	4.1.1. IDX AND GRAD										**4.1.2. OED**			
No. error	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	1.	2.	3.	4.
Frequency	0	0	1	0	0	50	18	22	5	0	0	0	77	87

	4.1.3. CIUP				**4.1.4. CID**					
No. error	1.	2.	3.	4.	1.	2.	3.	4.	5.	6.
Frequency	0	26	2	12	17	43	17	49	13	55

TABLE 6. Frequency of each error detected in laboratory protocols.

values marked as invalid will always be decided with regard to the overall properties of the tested soil. It is possible that in the course of manual corrections some rules will have to be modified or new rules added. However, the rules presented here are based on years of practical experience in the laboratory and as such they can provide guidance to other researchers dealing with similar issues of validation of standard geotechnical laboratory test records.

This paper, together with previously published *Extracting General Knowledge of Model Parameters for Clays out of Numerous Laboratory Tests* [1], presents in detail one of the options for dealing with the task of preparation of data originating from an MS Excel file for statistical processing using automatic scripts. The further progress of statistical data processing will be presented in the future.

ACKNOWLEDGEMENTS

The support provided by the SGS project No. SGS23/032/OHK1/1T/11 and the GAČR project No. 22-12178S is gratefully acknowledged.

REFERENCES

[1] V. Pavelcová, T. Janda. Extracting general knowledge of model parameters for clays out of numerous laboratory tests. In *Acta Polytechnica CTU Proceedings*, vol. 40, pp. 76–82. 2023. `https://doi.org/10.14311/APP.2023.40.0076`

[2] T. Kadlíček, T. Janda, M. Šejnoha, et al. Automated calibration of advanced soil constitutive models. Part I: hypoplastic sand. *Acta Geotechnica* **17**(8):3421–3438, 2022. `https://doi.org/10.1007/s11440-021-01441-0`

[3] T. Kadlíček, T. Janda, M. Šejnoha, et al. Automated calibration of advanced soil constitutive models. Part II: hypoplastic clay and modified Cam-Clay. *Acta Geotechnica* **17**(8):3439–3462, 2022. `https://doi.org/10.1007/s11440-021-01435-y`

[4] ExCalibre. SoilModels Automatic Calibration. [2023-07-26]. `https://soilmodels.com/excalibre-cs/`

[5] T. Janda, D. Mašín. General method for simulating laboratory tests with constitutive models for geomechanics. *International Journal for Numerical and Analytical Methods in Geomechanics* **41**(2):304–312, 2017. `https://doi.org/10.1002/nag.2558`

[6] J. Kruschke. *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan.* Academic Press, 2014. ISBN 978-0-12-405888-0.

[7] P.-A. von Wolffersdorff. A hypoplastic relation for granular materials with a predefined limit state surface. *Mechanics of Cohesive-frictional Materials* **1**(3):251–271, 1996. `https://doi.org/10.1002/(sici)1099-1484(199607)1:3<251::aid-cfm13>3.0.co;2-3`

[8] D. Mašín. A hypoplastic constitutive model for clays. *International Journal for numerical and analytical methods in Geomechanics* **29**(4):311–336, 2005. `https://doi.org/10.1002/nag.416`

[9] D. Mašín. Clay hypoplasticity with explicitly defined asymptotic states. *Acta Geotechnica* **8**(5):481–496, 2013. `https://doi.org/10.1007/s11440-012-0199-y`

[10] K. H. Roscoe, J. B. Burland. On the generalized stress-strain behaviour of wet clay. In *Engineering plasticity*, pp. 535–609. Cambridge University Press, 1968.