



REGRESSION-BASED PREDICTION OF SOIL PLASTICITY AND CONSISTENCY INDICES IN THE GEOLOGICAL CONTEXT OF THE SIBIU AREA, ROMANIA

Gabriela Mariana DRAGOMIR,^{1,2*}, Mihaela ROCA³, Irina Rozica MIRCEA³,
Bahadır BOZTEPE⁴, Erbil BALIKAYA⁴

¹ Doctoral School of Geology, University of Bucharest, Nicolae Bălcescu No. 1, 010041 Bucharest, gabriela-mariana.dragomir@drd.unibuc.ro

² Prospect Technical Studies, Gara Pasarea Street, Number 3, 077145 Pantelimon, Ilfov, Romania

³ Faculty of Geology and Geophysics, University of Bucharest, Nicolae Bălcescu No. 1, 010041 Bucharest, Romania, mihaela.roca@unibuc.ro

⁴ Makyol Construction Industry Tourism and Trade Co. Inc., Sibiu Branch Office, Romania

DOI: 10.2478/minrv-2026-0058

received – 10.08.2026 / revised – 17.08.2026

accepted – 10.09.2026 / published – 25.09.2026

Abstract: *Atterberg-limit testing is fundamental to geotechnical characterization, although the number of determinations may be limited in routine infrastructure investigations by the available testing programme. This study applies and validates previously established regional regression relationships for predicting the plasticity index (I_p) and consistency index (I_c) of fine-grained soils from the Sibiu area, Romania. A database of 83 samples with complete grain-size fractions, natural water content (w), liquid and plastic limits (w_L , w_P), depth (z), and Atterberg indices was analysed. The plasticity index was predicted from the liquid limit through the linear model $I_p = 1.06627 w_L - 17.4372$, according to [1], whereas the consistency index was predicted using a logarithmically transformed multivariable regression incorporating depth, clay fraction, and natural water content. The I_p model achieved $R^2 = 0.912$ and $RMSE = 3.64\%$. The transformed I_c regression achieved $R^2 = 0.942$. Lower and upper characteristic values for I_p and I_c were derived from the regression predictions using the statistical procedure described in the manuscript. The proposed relationships support preliminary site characterization, interpolation between tested horizons, and quality control of laboratory determinations; they complement rather than replace laboratory testing. Their applicability is limited to geological conditions comparable with the investigated fine-grained Neogene deposits of the Sibiu area.*

Keywords: *Atterberg limits; plasticity index; consistency index; multivariable regression; granulometry; Sibiu;*

1. Introduction

In geotechnical engineering practice, reliable characterization of soil parameters is fundamental to design. Among laboratory-derived index properties, the Atterberg limits - liquid limit (w_L), plastic limit (w_P), plasticity index ($I_p = w_L - w_P$), and consistency index ($I_c = (w_L - w)/I_p$) - are important for classifying fine-grained soils and interpreting their consistency and engineering behaviour [2, 3]. Laboratory testing remains the reference procedure and should follow applicable standards [4, 5, 6].

The practical objective of this study is to evaluate whether missing Atterberg parameters can be estimated within the investigated Sibiu-area database, where fine-grained Miocene Sarmatian marls and silts vary over short depth intervals. In many routine investigations, only a limited number of Atterberg tests are performed. The resulting regression equations are intended as regional engineering tools for preliminary characterization and quality control, not as replacements for laboratory tests or universal empirical relationships.

The regression relationships used in this study were established and evaluated in a previous study by Dragomir et al. [1], in which a continuous predictive modelling approach for geotechnical parameters was proposed as a probabilistic alternative to the discrete estimation of characteristic values. The methodological

* Corresponding author: Gabriela Mariana Dragomir, Doctoral School of Geology, University of Bucharest, Nicolae Bălcescu No. 1, 010041 Bucharest, gabriela-mariana.dragomir@drd.unibuc.ro

contribution of the present research lies in their application to the Sibiu dataset to determine the lower and upper characteristic values using the regression predictions and the statistical procedure described below. The originality of the study does not lie in proposing a new regression algorithm, but in validating transparent and interpretable relationships within a distinct geological setting characterized by heterogeneous Neogene marl and silt deposits, and in assessing their predictive performance from a geological perspective rather than solely through statistical indicators.

Although numerous empirical relationships between Atterberg limits, grain-size distribution and related index properties have been reported, their practical applicability remains constrained by geological variability. Most published predictive relationships are calibrated for specific datasets and assessed principally through statistical performance; less attention is given to the geological conditions that govern validity and transferability. This study therefore treats the equations as regional engineering tools, whose reliability depends on the depositional environment, lithological heterogeneity, granulometric composition and carbonate content for the classification of samples as marly clay.

Accordingly, the objectives are:

- (i) to describe the geotechnical database and the geological context;
- (ii) to present and justify the regression models used to predict I_p and I_c ;
- (iii) to validate these predictions against laboratory values through statistical indicators, agreement plots, frequency distributions, depth profiles and residual analysis;
- (iv) to interpret the results in terms of granulometry, water content and geological heterogeneity;
- (v) to compare the findings with the literature.

Literature Review

The physical basis of the Atterberg limits was established by [2], whose plasticity chart remains the foundation of modern soil classification. Skempton [3] introduced the concept of colloidal activity, defined as the ratio of the plasticity index to the clay fraction, demonstrating that soils with similar clay contents may exhibit markedly different plasticity depending on their dominant clay minerals. Together, these studies show that grain-size distribution alone is insufficient to characterize soil consistency and that mineralogical composition exerts a fundamental control on index properties. These concepts provide the physical basis for investigating predictive relationships among measurable index properties while recognizing that mineralogical variability limits their applicability beyond the calibration domain.

Subsequent studies have related the Atterberg limits to grain-size fractions, natural water content, and mineralogical characteristics. [7] reviewed these empirical relationships and reported that single-variable correlations rarely explain more than approximately 70% of the observed variability, whereas multivariable and nonlinear approaches generally provide higher predictive accuracy. More recently, machine-learning techniques have further improved prediction accuracy by incorporating a larger number of explanatory variables [8]. However, these methods often sacrifice model transparency and physical interpretability. Because the present study evaluates engineering relationships intended to remain transparent, physically interpretable, and readily applicable in routine geotechnical practice, regression-based models were considered more appropriate than black-box machine-learning approaches.

The importance of geological context has been increasingly recognized in recent studies. [9] demonstrated that empirical relationships between geotechnical index properties are inherently formation-dependent and should be interpreted within their geological setting rather than transferred indiscriminately between different depositional environments. Their work emphasized that reliable predictive models require regional calibration that explicitly accounts for lithological and geological variability, providing one of the principal methodological foundations for the present study.

Carbonate-bearing and marly soils introduce an additional level of geological complexity. In such deposits, common throughout the Transylvanian Basin, calcium carbonate occurs as dispersed micrite and cement, thereby modifying soil aggregation, the effective clay fraction, and consequently the measured plasticity. Although carbonate-content data were available for only a subset of the investigated samples, carbonate remains an important geological modifier in marly deposits and is therefore considered when interpreting model applicability and transferability.

A closely related regional study was presented by [1], who combined multivariable regression with the Effective Random Dimension (ERD) methodology to derive characteristic geotechnical values for two Romanian geological domains, reporting coefficients of determination of up to $R^2 = 0.98$ for the liquid limit. Their results demonstrated both the potential of regional calibration and the importance of accounting for geological variability when developing predictive relationships for engineering practice.

Collectively, the available literature demonstrates that empirical prediction of Atterberg-related parameters is feasible, but that predictive performance is strongly influenced by geological conditions. Validation of such relationships for heterogeneous Neogene marl deposits remains limited. The present study addresses this gap by evaluating a previously established regional regression framework for the Sibiu Neogene deposits, defining its domain of applicability, and interpreting predictive performance in relation to the geological setting.

2. Geological Setting

The regional and local geological setting provides the basis for interpreting the statistical results (Figure 1). The proposed regression equations are intended for application within a specific geological domain rather than as universally transferable empirical relationships. Accordingly, characterization of the geological framework is fundamental to defining the limits of their applicability.

As a general principle, the explanatory power of regression models for Atterberg-related properties is influenced by geological homogeneity. Deposits formed under relatively uniform depositional conditions are generally expected to exhibit more consistent predictive relationships than those affected by complex depositional and diagenetic histories.

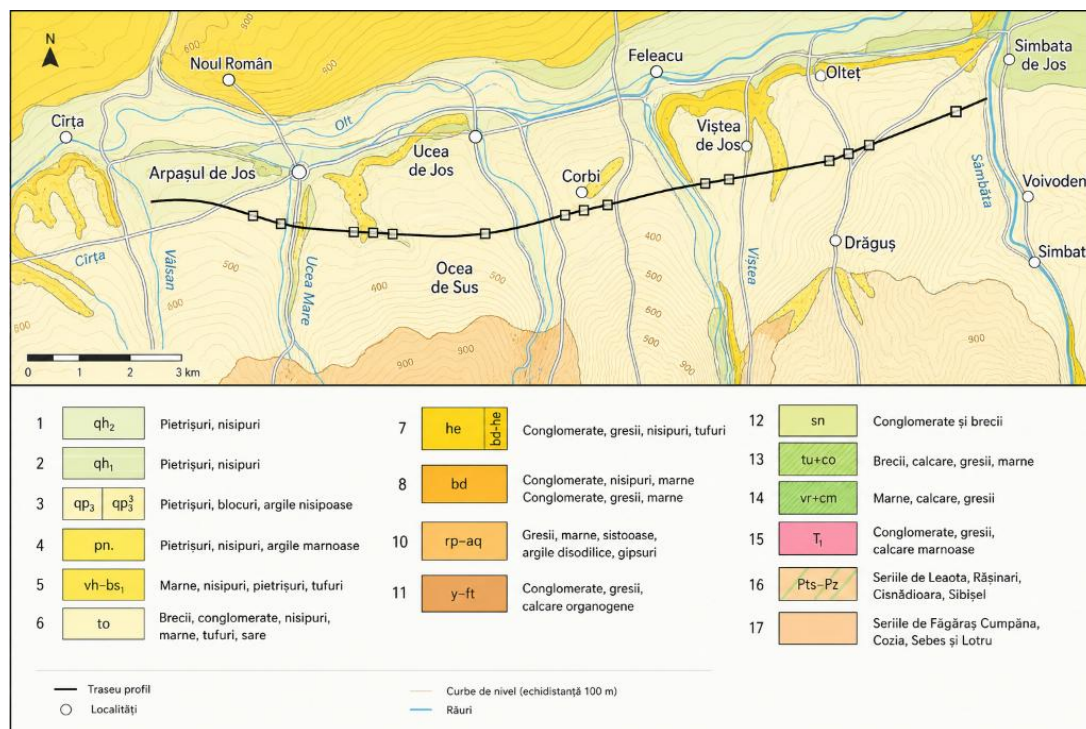


Fig. 1. Geological context of the study area. Schematic geological map of the Sâmbăta–Arpașu sector, redrawn after Codarcea-Dessila, Dimitrescu and Stancu [10]

The Transylvanian Basin is a post-tectonic intramontane sedimentary depression located in central Romania and bounded by the Carpathian fold-and-thrust belt [11, 12]. Following the Late Oligocene–Early Miocene Carpathian collision, the basin accumulated a thick succession of Miocene sediments deposited in predominantly brackish to freshwater environments. In contrast to the strongly deformed Subcarpathian Flysch belt, the Transylvanian Basin is only moderately affected by block faulting and differential subsidence, particularly along the southern margin, where the investigated Sâmbăta–Arpașu sector is located [13].

The stratigraphic sequence encountered in the investigated infrastructure boreholes comprises two principal units. The uppermost interval (approximately 0–5 m) consists of Quaternary proluvial–alluvial deposits associated with the piedmont fans of the Sâmbăta and Arpașu rivers. These deposits are composed predominantly of poorly sorted gravels and sands with a silty-clayey matrix of generally low plasticity. Owing to their predominantly non-plastic character, these materials were excluded from the regression analysis to ensure that model calibration was restricted to cohesive soils for which the Atterberg limits represent physically meaningful and engineering-relevant descriptors.

Below the Quaternary cover, Miocene (Sarmatian–Pannonian) sediments dominate the investigated profile. These deposits consist of a cyclical alternation of grey to bluish-grey clayey marls, silty clays, and

subordinate fine sands deposited within the epicontinental Paratethys Sea [14]. This lithological succession constitutes the geological domain for which the regression models were calibrated and validated.

A defining geotechnical characteristic of these Miocene deposits is the spatially variable occurrence of calcium carbonate (CaCO_3), present as dispersed bioclasts, secondary chemical precipitates, and localized calcareous concretions. Although carbonate-content measurements were available for only a subset of the investigated samples and therefore could not be incorporated into the principal calibration dataset, their geological significance is considered when interpreting the applicability and transferability of the proposed regression models.

CaCO_3 content in the available dataset ranges from near zero to 21.9% by weight, with a strongly right-skewed distribution (mean 2.0%, standard deviation 4.9%). At a constant clay content, carbonate-rich horizons reduce plasticity through two complementary mechanisms: dilution of the active clay fraction by relatively inert carbonate particles and physicochemical flocculation induced by free Ca^{2+} ions, which promote aggregation of clay platelets into coarser pseudo-silt units [15, 16, 17]. These processes modify the engineering behaviour of marl deposits and help explain part of the geological variability that cannot be captured by regression models based solely on granulometric and index-property variables. Accordingly, carbonate content is considered an important geological factor influencing the interpretation, applicability, and transferability of the proposed regression models, even though it was not included as a predictor in the principal calibration.

3. Materials and Methods

3.1 Database and laboratory determinations

The initial experimental database comprised 86 fine-grained soil specimens recovered from continuous borehole sampling campaigns conducted for infrastructure geotechnical investigations in the Sâmbăta–Arpașu sector of the Sibiu area. Continuous sampling was adopted deliberately because selecting only lithologically distinct layers would introduce subjective bias towards visually recognizable materials and would underestimate the natural variability of the geological profile. Instead, continuous sampling captures the full spectrum of geological variability, including thin transitional horizons between marl and silt, thereby providing a representative basis for regression analysis [18]. Sampling depths ranged from approximately 6.0 to 35.0 m below ground level, encompassing both shallow, less consolidated horizons and deeper, more consolidated deposits.

Each database record includes the borehole identifier, sampling depth interval, clay, silt and sand fractions [5], natural water content, (w) [4], liquid and plastic limits, (w_L) and (w_P) [6], the derived indices (I_p) and (I_c), and, where available, calcium carbonate content [19]. After excluding records with incomplete information for the variables required by the regression models, 83 complete samples were retained for calibration and validation. Because calcium carbonate was determined for only a subset of the investigated samples, it was not included in the principal calibration dataset; instead, its influence is discussed separately in the interpretation of model applicability and residual variability.

3.2 Definition of the index properties

The plasticity and consistency indices were computed from the laboratory-determined Atterberg limits and the natural water content using the conventional definitions:

$$I_p = w_L - w_P \quad (1)$$

$$\frac{I_c = (w_L - w)}{I_p} \quad (2)$$

where

- (w_L) is the liquid limit (%),
- (w_P) is the plastic limit (%),
- (w) is the natural water content (%),
- (I_p) is the plasticity index (%), and
- (I_c) is the consistency index (dimensionless).

These conventional definitions were adopted without modification throughout the study. Whereas the plasticity index depends exclusively on the Atterberg limits and therefore reflects the intrinsic plasticity of the soil, the consistency index additionally incorporates the natural water content and characterizes the in-situ consistency state.

3.3 Regression models for the plasticity and consistency indices

The plasticity-index regression relationship adopted from the previous study [1] is:

$$Ip_{pred} = 1.06627wL - 17.4372 \quad (3)$$

The regression coefficient indicates that the plasticity index increases almost proportionally with the liquid limit, while the negative intercept reflects the relatively stable plastic limit characteristic of the investigated deposits. The regression coefficients were independently verified during model implementation.

The consistency index was predicted using a multivariable regression formulated in transformed response space. Rather than modelling (I_c) directly, the regression was calibrated for the logarithm of the ratio between the consistency index and the natural water content, in accordance with [1]:

$$Q = \log_{10} \left(\frac{I_c}{w} \right) \quad (4)$$

$$Q = -0.238095 - 0.270692 \frac{1}{z} + 0.020012 * \log_{10}(clay) - 0.252114 * \sqrt{w} \quad (5)$$

The inverse-depth term represents the progressive effects of consolidation, the logarithmic transformation accounts for the nonlinear influence of clay fraction, and the square-root transformation of natural water content improves model stability across the investigated range of moisture conditions.

Residual adjustment and back-transformation were subsequently applied to obtain the final consistency-index predictions using Equations [7]:

$$I_c = w \cdot 10^Q \quad (6)$$

where

- z is sampling depth (m)
- Clay is clay fraction (%)

The regression relationships in Equations (3), (5), and (6) were adopted from the previous study [1], whereas the Effective Random Dimension (ERD) method used in that study was not applied here. For the Sibiu dataset, lower and upper characteristic values were derived from the regression predictions using the statistical framework of NP 122:2010. The uncertainty term is the regression root mean square error (RMSE), denoted sc in Appendix A.5 of NP 122:2010 and calculated here as the square root of the mean squared residuals, $RMSE = \sqrt{[(1/n)\Sigma e^2]}$. Thus, RMSE is used consistently as the uncertainty estimator in the Methods and Table 3; no separate standard error of regression is introduced. The coefficient $t\alpha = 1.64$ represents the one-sided 95% coefficient used for large samples, consistent with the asymptotic value in Appendix A.5. For I_p , the characteristic bounds are calculated directly in the original response space. For the consistency model, the bounds are first calculated in the transformed response space Q and then back-transformed to I_c using the mean natural water content of the investigated dataset. Because depth is already included as a predictor, the reported values are dataset-level characteristic summaries rather than depth-specific profiles. They are intended as prudent statistical estimates for the investigated geological domain, not as universally applicable design values.

$$X_k = \overline{X_{pred}} + k_n \cdot S_e \quad [8]$$

where $t\alpha = 1.64$ is the one-sided 95% statistical coefficient, X_k is the characteristic value, X_{mean} is the regression-based mean value, and RMSE is the regression root mean square error used as the uncertainty term. The lower and upper bounds correspond to the minus and plus forms, respectively, for the investigated dataset.

The lower and upper bounds are reported as the characteristic lower and upper values, respectively, for the investigated dataset. The bounds are intended as prudent statistical summaries of the investigated geological domain and are not presented as universal design values.

3.4. Model validation

Prediction performance was evaluated by comparison with laboratory measurements of the plasticity and consistency indices. Pearson's correlation coefficient (r), the coefficient of determination (R^2), Spearman's rank correlation coefficient, mean bias, mean absolute error (MAE), and root mean square error (RMSE) were used as quantitative performance indicators. Model behaviour was further evaluated using agreement plots, frequency distributions, depth profiles, and residual analyses. Residuals were defined as $e = \text{predicted} - \text{laboratory}$, and their distribution was assessed using the Shapiro-Wilk normality test. Validation was performed exclusively against laboratory observations; no independent external dataset or cross-validation procedure was used.

3.5. Scope relative to the reference framework

The regression structure follows the previously published methodological framework [1]. The present study applies the established regression equations to the Sibiu dataset and additionally derives characteristic values for I_p and I_c . The Effective Random Dimension (ERD) method is not applied.

4. Results

4.1. Descriptive statistics

The investigated dataset represents a broad range of engineering behaviour within the fine-grained Neogene deposits of the Sibiu area. Laboratory plasticity index (I_p) values range from **10.8%** to **71.8%** (mean **44.5%**, CoV **27.8%**), reflecting the presence of soils spanning from low to high plasticity. In contrast, the consistency index (I_c) is distributed within a relatively narrow interval (**0.596–1.000**; mean **0.919**), indicating that most samples were recovered under comparable stiff to hard consistency conditions despite the observed lithological variability.

These results illustrate the heterogeneous composition of the investigated Neogene succession while confirming that the moisture conditions at the time of sampling remained relatively uniform. The descriptive statistics of both laboratory measurements and regression predictions are summarized in Table 1.

Table 1. Descriptive statistics of laboratory and predicted indices

Variable	n	Mean	SD	CoV (%)	Minimum	Maximum	Units
Laboratory (I_p)	83	44.545	12.376	27.8	10.770	71.800	%
Predicted (I_p)	83	46.298	13.960	—	15.734	79.572	%
Laboratory (I_c)	83	0.919	0.074	8.1	0.596	1.000	—
Predicted (I_c)	83	0.882	0.098	—	0.496	0.980	—

4.2. Prediction performance

The predictive performance of the regression models is summarized in Table 2. For the plasticity index, the linear regression provides a direct prediction of the laboratory I_p , allowing the coefficient of determination to be interpreted directly. The model explains 91.2% of the observed variance and shows strong agreement with laboratory measurements.

The consistency-index procedure differs in structure. Model calibration was performed in the transformed response space defined by $Q = \log_{10}(I_c/w)$, where the multivariable regression achieved a coefficient of determination of **$R^2 = 0.942$** . Following residual adjustment and back-transformation, the final engineering predictions were evaluated against laboratory measurements using the correlation coefficient, mean bias, mean absolute error and root mean square error. These validation statistics describe the performance of the final predicted consistency index and should not be interpreted as the coefficient of determination of the transformed regression.

Overall, the results indicate that the proposed regression procedures reproduce the laboratory observations with high accuracy for the plasticity index and with good agreement for the consistency index, while reflecting the greater variability inherently associated with natural water content.

Table 2. Agreement and residual statistics (residual = predicted - laboratory)

Target	Calibration	r	(R^2)	Spearman	Bias	MAE	RMSE	Target
Plasticity index (I_p)	Direct linear regression	0.977	0.912	0.970	1.753	2.966	3.642	Plasticity index (I_p)
Consistency index (I_c)	Transformed regression	0.985	0.942	—	—	—	—	Consistency index (I_c)

4.3. Plasticity-index prediction

The calibration relationship shown in Figure 3 confirms the strong linear dependence of the plasticity index on the liquid limit. The regression equation was adopted from the previous study [1]:

$$I_p = 1.06627w_L - 17.4372$$

The narrow scatter of the laboratory observations around the fitted regression line reflects the close coupling between these two index properties within the investigated Neogene deposits.

The high coefficient of determination ($R^2 = 0.912$) together with the low prediction error (RMSE = 3.64%) indicates that the liquid limit is a **strong predictor** of the plasticity index in the investigated geological domain. This behaviour is consistent with the physical relationship between the two variables, both of which

primarily reflect the quantity and activity of the fine-grained fraction. The small positive mean bias (1.75%) indicates a slight tendency to overestimate the plasticity index, although the magnitude of this deviation remains small relative to the observed variability of the dataset.

Overall, the results show that the adopted linear regression reproduces the plasticity characteristics of the investigated fine-grained Neogene deposits with good accuracy and can support interpolation, preliminary site characterization, and quality control of laboratory determinations within the evaluated geological domain.

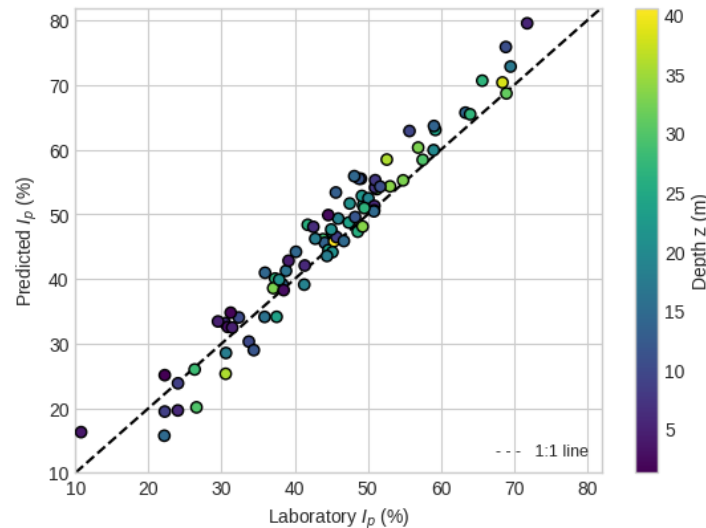


Fig. 2. Comparison between laboratory and predicted plasticity-index (I_p) values. Samples are colored according to sampling depth, and the dashed line represents perfect agreement (1:1)

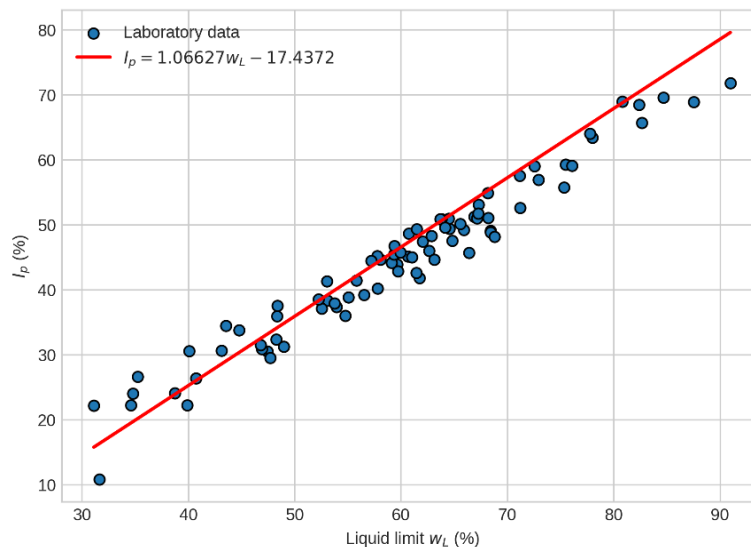


Fig. 3. Calibration of the plasticity-index model showing the relationship between laboratory and the liquid limit (w_L), together with the fitted regression line

4.4 Consistency-index prediction

The consistency-index procedure was calibrated in the transformed response space defined in the previous study [1]. The multivariable regression based on sampling depth, clay fraction, and natural water content achieved $R^2 = 0.942$ for the transformed response.

Following residual adjustment and back-transformation, final consistency-index values were recalculated for each sample using Equation (6) above and compared directly with the laboratory measurements.

The final predictions show good agreement with the observed consistency indices ($r = 0.831$), with a mean absolute error of 0.047, a root mean square error of 0.066, and a small negative mean bias (-0.037). The slight underestimation reflects the conservative residual-adjustment procedure incorporated into the adopted modelling framework rather than a systematic deficiency of the regression itself. The resulting values are used in Figures 4, 6, 8, and 10 and in the characteristic-value calculation.

The greater dispersion relative to the plasticity-index model is physically consistent with the definition of the consistency index. Whereas I_p depends only on the Atterberg limits, I_c additionally incorporates the natural water content and is therefore more sensitive to local variations in moisture conditions and lithological heterogeneity. Nevertheless, the regression reproduces the overall behaviour of the investigated deposits and provides useful engineering estimates for preliminary site characterization, interpolation between tested horizons, and quality control within the calibrated geological domain.

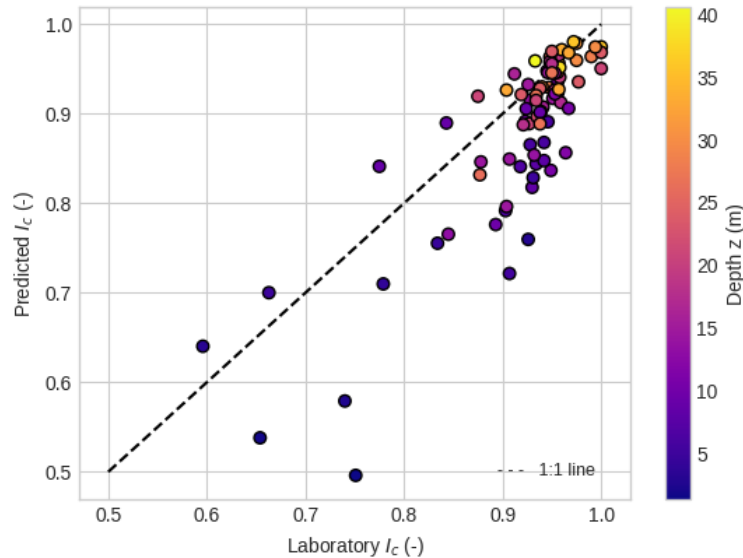


Fig. 4. Comparison between laboratory and predicted consistency-index (I_c) values

4.5 Frequency distributions

The frequency distributions of the laboratory measurements and corresponding regression predictions are compared in Figures 5 and 6 to evaluate whether the proposed models reproduce not only individual observations but also the overall statistical characteristics of the investigated dataset.

For the plasticity index, the predicted and laboratory distributions exhibit a high degree of overlap, both in their central tendency and in their overall spread. This close agreement indicates that the linear regression successfully reproduces the population characteristics of the investigated deposits rather than merely fitting individual observations. The correspondence between the two distributions further supports the robustness of the regional relationship established between the liquid limit and the plasticity index.

The consistency-index distributions also show good overall agreement, although the predicted values display a slightly broader distribution and a modest shift towards lower values. This behaviour is consistent with the small negative mean bias observed during model validation and reflects the greater influence of natural water content on the consistency index. Consequently, the regression captures the overall statistical behaviour of the investigated deposits while preserving the increased variability expected for a moisture-dependent engineering parameter.

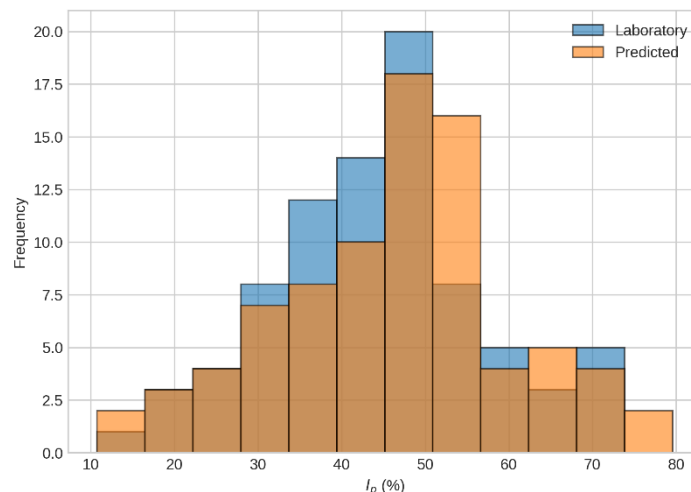


Fig. 5. Frequency distributions of laboratory and predicted plasticity-index (I_p) values

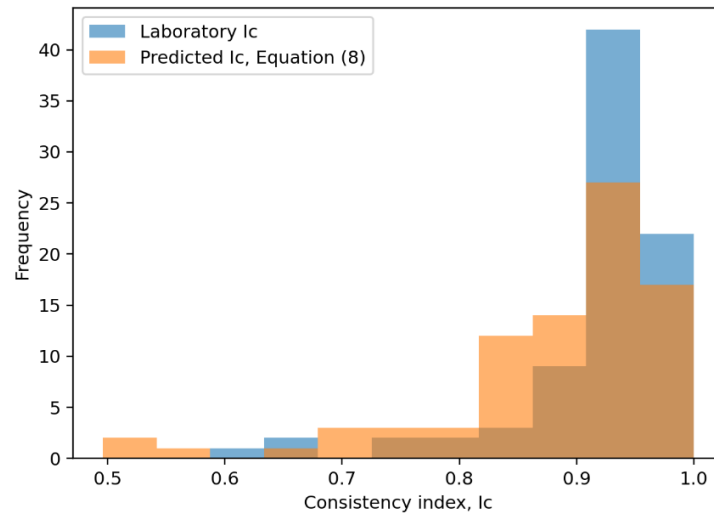


Fig. 6. Frequency distributions of laboratory and recalculated predicted consistency-index (I_c) values, determined using Equation (6)

Overall, the distributional agreement confirms that the proposed regression procedures reproduce the principal statistical characteristics of the laboratory dataset and are suitable for regional interpolation and preliminary engineering assessment within the calibrated geological domain.

4.6 Vertical profiles with depth

Laboratory and predicted values of the plasticity index (I_p) and consistency index (I_c) were compared along the investigated depth interval to evaluate the ability of the regression models to reproduce the vertical geotechnical profile of the Neogene deposits. This comparison provides a practical assessment of the applicability of the proposed equations for reconstructing continuous subsurface profiles from incomplete laboratory datasets.

The predicted plasticity-index profile closely follows the laboratory measurements throughout the investigated section, reproducing both the overall trend and the local variations observed with depth (Figure 7). The gradual increase in plasticity within the deeper intervals is consistent with the progressive transition towards finer-grained, more clay-rich horizons characteristic of the investigated Neogene succession.

The consistency-index profile (Figure 8) reproduces the general vertical behaviour of the deposits, although larger local deviations occur in intervals affected by abrupt lithological changes or variations in natural water content. This behaviour is expected because the consistency index represents a state parameter that depends directly on in-situ moisture conditions in addition to the plasticity characteristics of the soil.

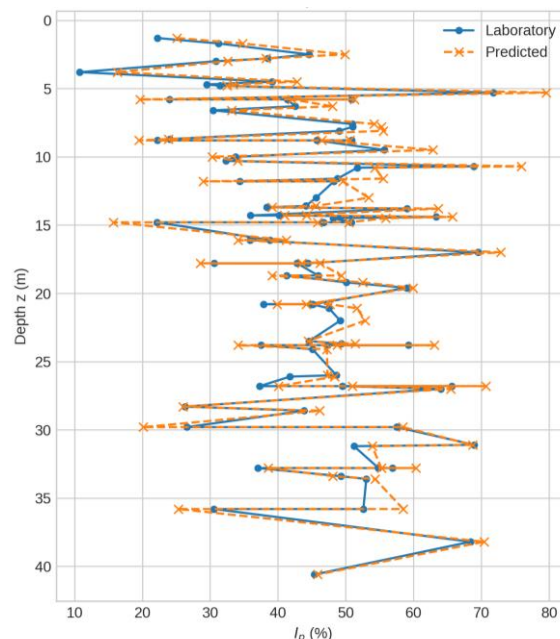


Fig. 7. Comparison between laboratory and predicted plasticity-index (I_p) values along the investigated depth profile

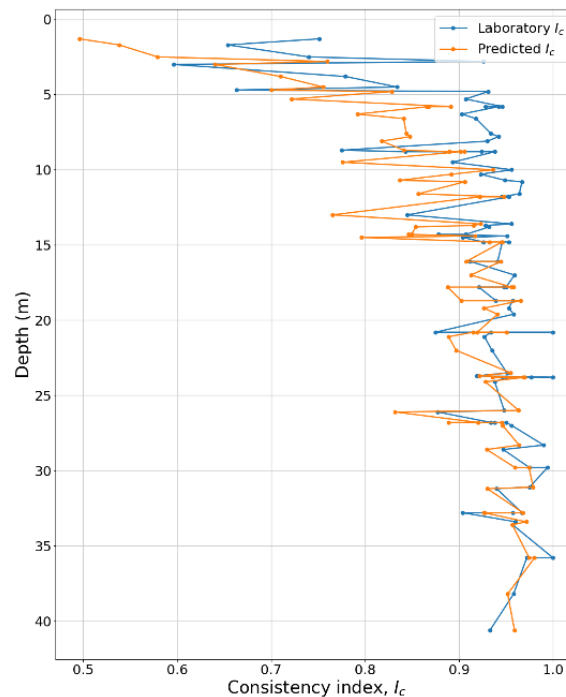


Fig. 8. Comparison between laboratory and recalculated predicted consistency-index (I_c) values along the investigated depth profile, determined using Equation (6)

From an engineering perspective, these results indicate that the proposed regression models can be used to reconstruct continuous geotechnical profiles where laboratory data are incomplete. The plasticity-index model provides reliable interpolation between tested horizons, whereas consistency-index predictions should be interpreted with greater caution in intervals exhibiting marked lithological or moisture variability.

4.7 Residual analysis

Residual analysis was performed to evaluate the adequacy of the regression models and to identify potential systematic errors or heteroscedasticity. Residuals were calculated as predicted minus laboratory values, and their distributions were examined with respect to the fitted values. In addition, the normality of the residuals was assessed using a normal Q–Q plot and the Shapiro–Wilk test.

The residuals of the plasticity-index model are randomly distributed around zero without any evident systematic trend (Figure 9), indicating that the prediction errors remain approximately constant throughout the investigated range. The Shapiro–Wilk test does not reject normality, and the Q–Q plot (Figure 11) shows close agreement between the observed residuals and the theoretical normal distribution. Together, these results support the adequacy of the adopted linear regression model for the analysed dataset.

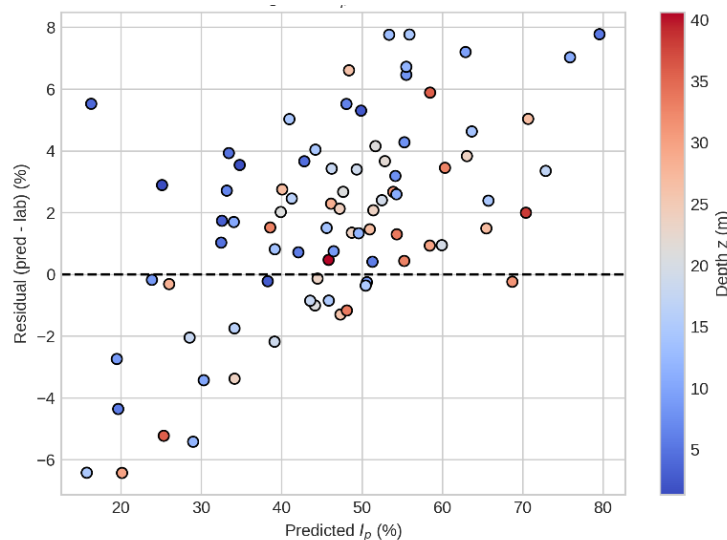


Fig. 9. Residuals of the plasticity-index (I_p) predictions plotted against predicted values

The consistency-index residuals (Figure 10) exhibit greater dispersion and depart from normality (Shapiro–Wilk). This behaviour reflects the additional variability introduced by natural water content and local geological heterogeneity, which cannot be fully represented by the adopted regression model. Although the transformed regression captures the principal consistency trends within the investigated deposits, local deviations remain larger than those observed for the plasticity-index model. Consequently, consistency-index predictions are appropriate for preliminary characterization, interpolation and quality control, but measured laboratory values should remain the preferred basis for analyses requiring high local accuracy.

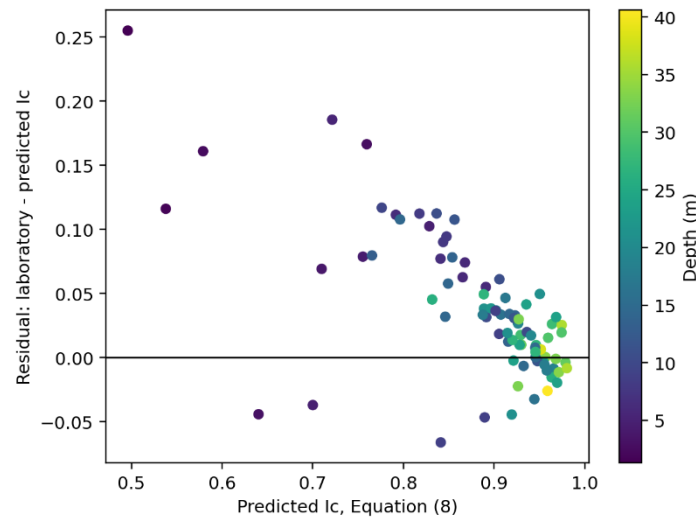


Fig. 10. Residuals of recalculated consistency-index (I_c) predictions determined using Equation (6), plotted against predicted values

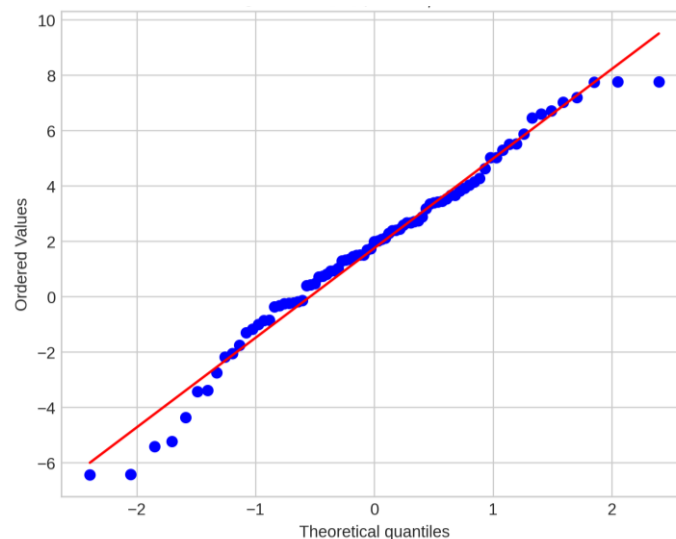


Fig. 11. Normal Q–Q plot of the plasticity-index (I_p) residuals

5. Discussion

The contrasting predictive performance of the two models can be interpreted from the physical meaning of the two indices. The plasticity index (I_p) is controlled primarily by the relationship between the liquid and plastic limits and therefore reflects the intrinsic plastic behaviour of the fine-grained fraction. The strong agreement obtained for I_p is consistent with the close coupling between the liquid limit and plasticity index in the investigated Neogene deposits.

The consistency index (I_c), in contrast, incorporates natural water content in addition to the Atterberg limits. Its larger prediction errors and greater residual dispersion are therefore physically consistent with its sensitivity to local moisture conditions. Variations in water content, together with lithological transitions within the marl and silty-clay succession, provide a plausible explanation for the lower agreement of the I_c predictions.

The variables entering the consistency-index model should be interpreted in relation to the geological setting. Sampling depth is treated as a site-specific proxy for the combined effects of consolidation state, lithological changes and associated variations in moisture conditions; the regression does not establish a direct causal effect of depth. The clay fraction represents an important compositional control because changes in the proportion of fine particles affect the plastic and consistency behaviour of the soil.

Carbonate content provides an additional geological control in the investigated marly deposits. Because carbonate measurements were available only for a subset of samples, its independent statistical contribution could not be quantified in the principal calibration. Nevertheless, carbonate may modify the effective clay fraction and soil aggregation and therefore provides a plausible geological source of residual variability. Its presence should consequently be considered when interpreting deviations from the regression relationships.

The residual patterns provide a further interpretation of model behaviour. The plasticity-index residuals are approximately centred around zero and show no evident systematic trend, supporting the adequacy of the linear relationship within the investigated domain. The consistency-index residuals are more dispersed and depart from normality, which is consistent with the greater sensitivity of I_c to natural water content and local geological heterogeneity.

The characteristic values derived from the regression predictions should be interpreted as statistical values for the investigated geological domain. They summarize the uncertainty associated with the regional predictive relationships and are not intended to represent universal values for all soils or geological formations. Their engineering interpretation therefore remains conditional on the applicability of the underlying regression relationships.

Table 3. Characteristic values derived from the regression predictions using the NP 122:2010 statistical framework

Parameter	Statistical basis	Lower characteristic value	Upper characteristic value
Plasticity index, I_p (%)	$I_{p,pred} \pm 1.64 \cdot RMSE_{Ip}$	40.32512	52.27088
Transformed consistency response, Q	$Q_{pred} \pm 1.64 \cdot RMSE_Q$	-1.35918	-1.24930
Consistency index, I_c (-)	$I_{c,k} = \bar{w} \cdot 10^{Q_k}$	0.79967	1.02990

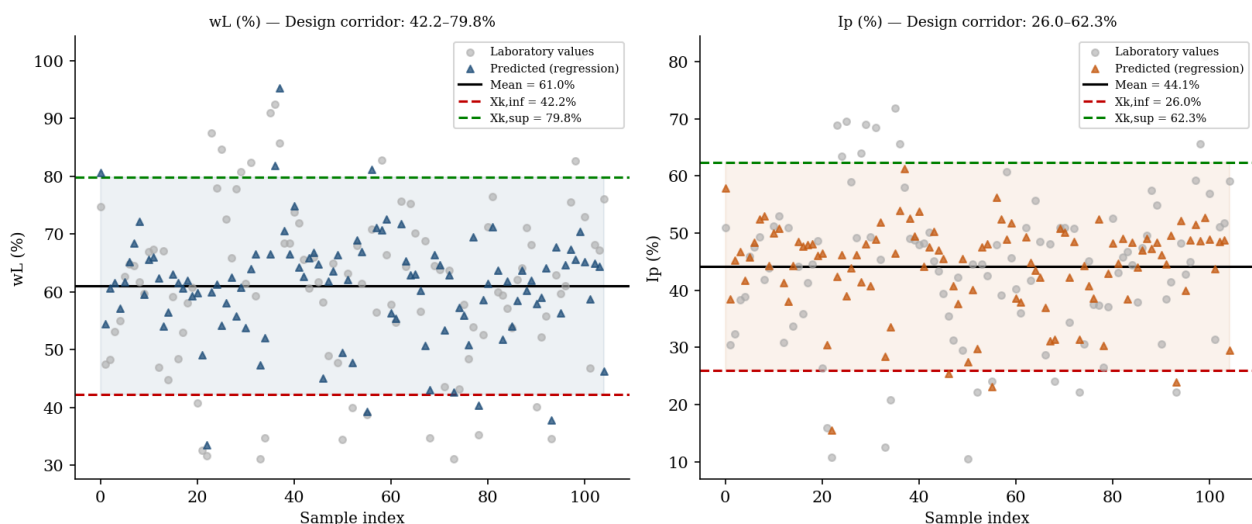


Fig. 12. Eurocode 7/NP122:2010 Characteristic Design Corridors

The findings are consistent with the literature reviewed in the manuscript, which indicates that empirical relationships between Atterberg-related properties are strongly dependent on soil composition and geological setting. In particular, the present results support the view that regression performance should be evaluated within the geological domain for which the relationship is intended, rather than interpreted as evidence of universal transferability.

The present study also differs in scope from the reference framework of Dragomir et al. [1]. The regression relationships were established in [1]; the present work applies and validates those relationships for the Sibiu dataset and additionally derives lower and upper characteristic values. The contribution is therefore regional application, geological interpretation and characteristic-value derivation rather than development of a new regression algorithm.

The characteristic values derived in this study should likewise be understood within the calibration domain and the statistical framework described in Section 3.3. They provide a consistent statistical summary of the regression-based predictions for the investigated dataset, but their use in design outside the investigated geological conditions would require independent verification.

The applicability of the proposed equations is restricted to geological conditions comparable with the fine-grained Neogene deposits represented by the Sibiu calibration dataset. Application to different lithological units, depositional environments, mineralogical compositions, carbonate regimes, moisture conditions or stress histories requires independent validation and, where necessary, recalibration.

6. Conclusions

This study shows that previously established regional regression relationships can provide reliable predictions of the plasticity index (I_p) and consistency index (I_c) when applied to the investigated fine-grained Neogene deposits of the Sibiu area.

The principal findings are as follows:

1. The plasticity index (I_p) was predicted accurately from the liquid limit using the regional relationship adopted from [1], $I_p = 1.06627w_L - 17.4372$. For the Sibiu dataset, the model achieved $R^2 = 0.912$ and RMSE = 3.64%, with no evident systematic trend in the residuals.

2. The consistency-index model achieved $R^2 = 0.942$ in the transformed response space, indicating that depth, clay fraction, and natural water content explain most of the variability of the transformed response. This coefficient of determination applies only to the transformed regression model.

3. The contrasting predictive performance of the two models reflects the different physical controls on the plasticity and consistency indices.

4. The proposed regression equations provide practical support for preliminary site characterization, interpolation between tested horizons, laboratory quality control, and identification of intervals requiring supplementary investigation. They are intended to complement, rather than replace, laboratory testing and engineering judgement.

5. The proposed equations are regionally calibrated and should be applied only to geological conditions comparable with the investigated fine-grained Neogene deposits of the Sibiu area. Application to different lithological units or depositional environments requires independent validation and, where appropriate, recalibration.

6. Lower and upper characteristic values for the plasticity and consistency indices were derived from the regression predictions. These values are specific to the investigated Sibiu geological domain and should not be interpreted as universally applicable design values.

Future research should focus on validating the proposed equations using independent geological formations, incorporating more complete mineralogical and carbonate datasets, and assessing the transferability of the regional regression framework to other fine-grained depositional environments.

References

- [1] **Dragomir G.M., Roca M., Mircea I.R.**, 2026
Continuous predictive modeling of geotechnical parameters: A probabilistic alternative to discrete characteristic value estimation, Geotechnics, 6[3], 59. <https://doi.org/10.3390/geotechnics6030059>
- [2] **Casagrande A.**, 1948
Classification and identification of soils, Transactions of the ASCE, 113, 901-930.
- [3] **Skempton A.W.**, 1953
The colloidal activity of clays, Proceedings of the 3rd International Conference on Soil Mechanics and Foundation Engineering, Zurich, 1, 57-61.
- [4] **EN ISO 17892-1**, 2015
Geotechnical investigation and testing - Laboratory testing of soil - Part 1: Determination of water content
- [5] **EN ISO 17892-4**, 2017
Geotechnical investigation and testing - Laboratory testing of soil - Part 4: Determination of particle size distribution
- [6] **EN ISO 17892-12**, 2018
Geotechnical investigation and testing - Laboratory testing of soil - Part 12: Determination of liquid and plastic limits

- [7] **Spagnoli G., Shimobe S.**, 2020
Statistical analysis of some correlations between compression index and Atterberg limits, Environmental Earth Sciences, 79, 245.
- [8] **Kim M., Senturk M.A., Tan R.K., Ordu E., Ko J.**, 2022
Deep learning approach on prediction of soil consolidation characteristics, Applied Sciences, 12[11], 5654.
- [9] **Bicocchi G., Tofani V., D'Ambrosio M., Tacconi-Stefanelli C., Vannocci P., Casagli N., Lavorini G.**, 2019
and Trevisani, M. Geotechnical and hydrological characterization of hillslope deposits for regional landslide prediction modeling. Bulletin of Engineering Geology and the Environment, 78, 4875-4891. <https://doi.org/10.1007/s10064-018-01449-z>
- [10] **Codarcea-Dessila M., Dimitrescu R., Stancu J.**, 1968
Harta geologică a României, scara 1:200,000, foaia 27-Sibiu (L-35-XIX), Institutul Geologic, Bucharest.
- [11] **Săndulescu M.**, 1984
Geotectonica României, Editura Tehnică, Bucharest, 335 pp.
- [12] **Mutihac V.**, 1990
Structura geologică a teritoriului României, Editura Tehnică, Bucharest, 420 pp.
- [13] **Krészsek C., Filipescu S., Silye L., Matenco L., Doust H.**, 2010
Miocene facies associations and sedimentary evolution of the Southern Transylvanian Basin (Romania): Implications for hydrocarbon exploration, Marine and Petroleum Geology, 27[1], 191–214.
<https://doi.org/10.1016/j.marpetgeo.2009.07.009>
- [14] **Vancea A.**, 1960
Neogenul din Bazinul Transilvaniei, Editura Academiei, Bucharest, 262 pp.
- [15] **Mitchell J.K., Soga K.**, 2005
Fundamentals of Soil Behavior (3rd ed.), John Wiley & Sons.
- [16] **Firincioglu B.S., Bilsel H.**, 2023
Unified Plasticity Potential of Soils, Applied Sciences, 13[13], 7889. <https://doi.org/10.3390/app13137889>
- [17] **Zolfaghari Z., Mosaddeghi M.R., Ayoubi S.**, 2015
ANN-based pedotransfer and soil spatial prediction functions for predicting Atterberg consistency limits and indices from easily available properties at the watershed scale in western Iran, Soil Use and Management, 31[1], 142–154.
<https://doi.org/10.1111/sum.12167>
- [18] **Phoon K.-K., Kulhawy F.H.**, 1999
Characterization of geotechnical variability, Canadian Geotechnical Journal, 36[4], 612-624.
- [19] * * *
STAS 7107/3-74. Teren de fundare. Determinarea conținutului de carbonați. Romanian State Standard, 1974.



This article is an open access article distributed under the Creative Commons BY SA 4.0 license. Authors retain all copyrights and agree to the terms of the above-mentioned CC BY SA 4.0 license.