Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Variability of Geotechnical Parameters and Comparison of Machine Learning Algorithms for Predicting the CBR of Lateritic Gravels for road construction in Tropical Regions

Carole Pierre Ongobassomben
carolepierre744@gmail.com
Université de Yaoundé I
Sylvie Désirée Noa Tang
Université de Yaoundé I
Aloys Thierry Ndzié Mvindi
Université de Bertoua
Christophe Enock Embom
Université de Yaoundé I
Michèle Stella Ngono Mbenti Emvana
Université de Yaoundé I
Arnaud Ngo'o Ze
Université de Yaoundé I
Vincent Laurent Onana
Université de Yaoundé I

Research Article

Keywords: Geotechnical variables, Load-bearing capacity, Statistics, Regression analysis, Central Cameroon

Posted Date: November 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11485401/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Variability of Geotechnical Parameters and Comparison of Machine Learning Algorithms for Predicting the CBR of Lateritic Gravels for road construction in Tropical Regions

Abstract

This study aimed to determine the geotechnical characteristics and variability of lateritic gravels in tropical environments, and then to compare Machine Learning Algorithms for predicting the California Bearing Ratio (CBR). A total of 159 samples were collected along the Esse–Soa road section in Central Cameroon. The specific gravity, particle size distribution (PSD), plasticity, derived and compaction parameters, and CBR were determined. The studied materials consist of the <2 mm fraction (14.6–99.4%), skeleton (0.6–85.4%), mortar (12.3–84.6%), and fines (7.3–64%). They are based on the Maximum Dry Density (1.62 – 2.16 g/cm^3) and CBR (12 – 94%) values, suitable for use as subgrade and subbase layers in pavements for all traffic types, and as base layers for light traffics. Data were analyzed using descriptive statistics, resulting in scattered data, characterized by low to high coefficients of variation (4 – 43%). Simple linear regressions (SLR) and principal component analysis were performed to predict the CBR from the other variables. No significant correlation was observed. Multiple linear and ridge regressions were performed following the SLR. They also turned out to be limited. Random forest and support vector machine regressions were also implemented. These methods, based on the same evaluation criteria as the previous learning techniques, produced successively satisfactory models, characterized by low to very low Mean Squared Error, Root Mean Squared Error, and Mean Absolute Error values, and high to very high R^2 values (0.85 – 0.94). PSD and derived parameters were the most suitable for predicting the CBR in the studied materials.

Keywords: Geotechnical variables; Load-bearing capacity; Statistics; Regression analysis; Central Cameroon

1. Introduction

Lateritic gravels are rich soils in iron and aluminum oxides. These soils are formed by the intense weathering of the underlying rock in tropical and subtropical regions [1]. They are characterized by their reddish color, good drainage properties, and relatively high strength. These properties make them valuable materials in road construction [2, 3]. They are used for the reasons mentioned above, often as base or subbase materials, but also because of their availability, low cost, and ability to withstand heavy loads. The CBR is the main geotechnical parameter used to guide the selection and quality control of lateritic gravels in road construction [4, 5]. The values of this parameter enable, for example, the determination of the thickness and composition of pavement layers, thereby ensuring adequate structural support and a certain degree of durability [6]. CBR values can also be used to estimate performances and predict the behavior of pavement and subgrade layers under the effect of repetitive traffic loads. Moreover, an accurate

assessment of the CBR has significant economic and environmental implications. This assessment encourages the use of locally available materials, thereby reducing transportation costs and the environmental impact [7]. Estimating the performances and predicting the behaviors of base and subbase layers using CBR currently poses challenges in tropical and subtropical regions [8]. These challenges are huge, given the significant variation in lateritic soils from one region to another and the numerous contributing factors. These factors include the nature of the bedrock, climate, vegetation, topography, the hydromorphy of the environment, and time [4, 9–14]. Related challenges require careful analysis of the areas, the establishment of relationships between variables and factors, and the development of region-specific models. However, the geotechnical properties of the used lateritic gravels considerably vary, even at the local scale [4, 12–15]. These variations are mainly due to the inherent heterogeneity of lateritic soils [16], which are formed through intense processes of differential weathering [17]. The local variability of lateritic gravels thus makes it difficult to assess and predict the CBR at the regional scale. Accurate methods for local estimation and prediction of CBR are therefore necessary in order to consider extrapolations to regional scales. The traditional techniques used up to now, which generally involve laboratory testing and classical regressions, have numerous limitations. For example, the standard CBR test is very expensive [7], labor-intensive, and time-consuming. Conducting this test requires meticulous experimental procedures in the laboratory. The limitations mentioned above have an impact on the ability to make quick decisions when carrying out road construction projects. The traditional correlation models for predicting CBR from other geotechnical properties of lateritic gravels often result in only moderate or even weak regression outcomes. These models exhibit limited predictive capabilities and practical limitations [4]. Models based on these regressions often assume linear relationships between soil properties and CBR values. These assumptions can lead to an oversimplification of the complex nature and behavior of soil, which, on the contrary, involve nonlinear interactions and spatial variability. The use of linear relationships may therefore be limited in contexts involving heterogeneous soils [18]. Furthermore, traditional approaches tend to inappropriately include or exclude new computational approaches. The latter, it must be admitted, could do a better job of accounting for the heterogeneity and complex behavior of soils. Previous works have shown, for example, that advanced models such as cluster-based data processing, genetic programming, and artificial neural networks outperform traditional techniques. Overfitting here is assessed based on the accuracy and robustness of the CBR's predictions [19–21]. Multiple linear regression, ridge regression, random forest, and support vector machine models, like the earlier mentioned computational techniques, are increasingly being used to estimate and predict soil CBR values [8, 22]. These other algorithms are useful because of their ability to account for the complex relationships between soil characteristics and CBR. Machine learning algorithms, however, have been very little used to estimate and predict the behavior of lateritic materials in the Central African subregion. The objective of this study is, therefore, to determine the geotechnical characteristics and variability of lateritic gravels developed on metamorphic bedrock in tropical environments and then to compare machine learning algorithms for predicting CBR.

2. Geographical and geological contexts

The study area is located in the Centre Region of Cameroon (Fig. 1a), in the Mefou et Afamba Division, specifically in the Esse and Soa Subdivisions. This area is geographically bounded by longitudes 011°33'40"E and 011°54'10"E and by latitudes 03°52'35"N and 04°11'05"N (Fig. 1b). The climate in this area is equatorial, with four distinct seasons: two rainy seasons and two dry seasons. The climate is hot and humid, with annual rainfall estimated at around 1,500 mm. This climate covers the entire southern part of the country [23]. In terms of vegetation, the study area encompasses the bimodal rainfall rainforest zone. This phytogeographic domain consists mainly of semi-deciduous to semi-evergreen and ulmaceae plants, which are well adapted to high humidity and temperatures [24]. The study area is morphologically located in the northern part of the Southern Cameroon Plateau. This morphological unit covers the largest area in Cameroon and remains one of the defining features of Central Africa. It has an average elevation of 750 meters in the vicinity of Yaoundé and rugged terrain with elevations ranging from 475 to 1,075 meters within the study area. This terrain consists of hills and valleys, formed by the differential erosion of various geological formations. The variations mentioned above suggest that the Southern Cameroon Plateau may have been affected by a large-scale undulating movement [23, 25]. The hydrographic network there is shallow. This network exhibits a typical dendritic pattern, whose presence and distribution can be influenced by geological structures. Residual landforms in the Southern Cameroon Plateau are relatively rare and provide evidence of very ancient land surfaces. These landforms generally correspond to cores of hard rocks [23] that have been spared from weathering [25]. Migmatitic gneisses, grenatiferous gneisses with two-mica, and quartzites (Fig. 1b) [26] are the geological formations overlain by the lateritic gravels under study. These formations are part of the Yaoundé Group [25, 27], which is situated within the broader geological context of Central Cameroon and even the Central African subregion. Central Cameroon is a rocky complex ranging from Precambrian formations to more recent ones. These formations are mostly granito-gneissic. They reflect high-grade metamorphic events and are associated with the Pan-African orogeny [27]. The products of their weathering have a significant impact on land use and urban development models.

3. Materials and Experimental Methods

3.1. Collection of samples

The studied lateritic gravels were collected along the 49 km stretch of road connecting the Esse and Soa localities. This road section runs through terrains with elevations ranging from 600 to 700 m. A total of 159 material samples were collected from 19 sampling sites (Fig. 2). The latter are spatially distributed according to the locations of the kilometer points shown in Figure 1. The sampled lateritic gravels were described and then sent to the National Civil Engineering Laboratory of Cameroon (LABOGENIE) for geotechnical testing.

3.2. Geotechnical testing

The particle size analysis of the studied materials was performed by dry sieving and sedimentometry, in accordance with standards NF P 94–56 [28] and NF P 94–057 [29], respectively. The NF P 94–054 standard [30] was used to determine the specific gravity (γ_s) of the solid particles. The liquid limit (LL) and plastic limit (Lp) were determined using the cup and roller methods, respectively, in accordance with standard NF P 94–051(31). The plasticity index (PI) was calculated from LL and Lp using Equation 1.

$$PI = LL - Lp \quad (1)$$

The grading modulus (Gm), the plasticity modulus (Pm), the plasticity product (Pp), and the swelling potential (ε_s) are parameters derived from the grain size distribution and plasticity of materials. These four parameters were determined using Equations 2, 3, and 4 [32] and Equation 5 [33].

$$Gm = [300 - (\% \leq 2\,mm + \% < 425\,\mu m + \% < 75\,\mu m)]/100 \quad (2)$$

$$Pm = PI \times percentage\ of\ particles\ passing\ through\ a\ 425\,\mu m\ sieve \quad (3)$$

$$Pp = PI \times percentage\ of\ particles\ passing\ through\ a\ 75\,\mu m\ sieve \quad (4)$$

$$\varepsilon_s = 1 \times 10exp - 5 \times PIexp2.24 \quad (5)$$

The NF P 94–093 standard [34] was used to determine the compaction parameters. These are the optimum moisture content (OMC) and the maximum dry density (MDD). Material test specimens were prepared and then immersed in water for four days. These test specimens were used to determine the California Bearing Ratio (CBR) through punching. The test was conducted in accordance with standard NF P 94–078 [35] using a universal laboratory testing machine equipped with a pressure gauge.

3.3. Spatial distribution

Variograms illustrating the spatial distributions of the studied materials based on the CBR were generated using ArcGIS version 10.5, following the Topo to Raster method and elevation point type. This is a widely used interpolation technique for creating digital elevation models in hydrogeology. It is based on the ANUDEM program developed by Michael Hutchinson [36]. This technique was implemented in this study with a maximum of 20 iterations, a discretization error factor of order 1, a standard vertical error of 0, and tolerance levels of 2.5 and 100.

3.4. Statistical analysis

The data subjected to statistical analysis are the results of laboratory tests. A population of 159 representative individuals of the studied lateritic gravels and 14 variables were used. The following variables stand out: γ_s, < 2 mm, > 2 mm (skeleton), < 425 µm (mortar), < 75 µm (fines), LL, PI, Gm, Pm, Pp, ε_s, OMC, MDD, and CBR.

3.4.1. Descriptive statistics

Descriptive statistics were performed using XLSTAT software, version 2014.5.03. It involved determining statistical variables such as the minimum (Min), the maximum (Max), the simple arithmetic mean ($\bar{X}$), the standard deviation (τ), and the coefficient of variation (CV). ($\bar{X}$), τ, and CV were determined using Equations 6, 7, and 8, respectively.

$$\bar{X} = (1/n) \sum_{i=1}^{n} x_i \qquad (6)$$

$$\tau = \sqrt{\sigma exp2} \qquad (7)$$

$$Cv = (\tau/\bar{X}) \times 100 \qquad (8)$$

Where n = total number of individuals (159), x_i = the value of the variable for an individual, and

$$\sigma exp2 = (1/n-1) \times \sum_{i=1}^{n} (x_i - \bar{X}) exp2 \qquad (9)$$

3.4.2. Correlation tests

Correlation tests were also performed using XLSTAT software. These tests were conducted on geotechnical variables and their modalities, using Pearson's method. They enabled the identification of the nature and levels of correlations between pairs of variables. The summary of the results is presented in the form of a correlation matrix. The principle used to interpret these correlations is such that, for a given correlation coefficient (R), $0 \leq R < 0.5$ indicates that the variables are weakly correlated. These variables are averagely correlated if $0.5 \leq R < 0.8$. They are highly correlated if $R > 0.8$.

3.4.3. Principal Component Analysis

Principal component analysis (PCA) also helps to assess the relationships between variables. Like the correlation tests, it was performed using Pearson's method. The advantage of ACP is that it summarizes information without truncating it. From a set with n initial variables p, PCA helps to construct subsets by identifying new variables. These are referred to as "principal components." They are usually linearly independent and provide the best explanation for all the observations. Based on n observations of the p initial variables, a descriptive summary in the form of pie charts was produced. The rules for interpreting the positions of the parameters in these charts are such that a variable is considered well-represented if it is close to the circumference of the circle and poorly represented if it is close to the center of the circle. Two well-represented variables and close to each other are positively correlated. Two well-represented and diametrically opposed variables are negatively correlated. Two well-defined variables are considered independent if they can be deduced from one another by rotating them 90 degrees.

3.5. Machine learning algorithms

This section presents the machine learning algorithms used to predict the CBR values of the studied lateritic gravels. The Scikit-Learn library for the Python 3.6 programming language was used to train and validate the machine learning algorithms. The hyperparameters for these algorithms were optimized using the GridSearch CV functionality. These hyperparameters are, in fact, tools that control the robustness and properties of a machine learning model [37]. Four algorithms – namely multiple linear regression, ridge regression, random forest, and support vector machines – were implemented in this study. These algorithms were run on aggregated data. The datasets were as follows: particle size variables (< 2 mm, > 2 mm, < 425 µm, and < 75 µm), plasticity data (LL and PI), derived parameters (Gm, Pm, Pp, and ε_s), and compaction parameters (OMC and MDD). The variable γ_s has been isolated. 30% of the data was used to evaluate the models' performance, and the remaining 70% was used to train the models.

3.5.1. Multiple linear regressions

Multiple linear regression (MLR) is a statistical technique used to model a relationship between a continuous dependent variable and two or more independent variables [38]. It is generally implemented when key assumptions – such as homoscedasticity, normality of the residuals, and the absence of multicollinearity – are reasonably satisfied. MLRs were used in this study because of their ability to fit a linear equation to the observed data, their simplicity, their interpretability, and their computational efficiency.

3.5.2. Ridge regression

Ridge regression, or L2 regularization, is a type of regression that uses linear models [39]. This statistical technique helps to overcome issues of multicollinearity, even when there are a large number of explanatory or predictive variables. It constructs a linear function in a high-dimensional feature space, induces nonlinear regressions [40], and, in the process, reduces the risk of overfitting on the training data. The reduction in risk was achieved by lowering the regression coefficients.

3.5.3. Random forest regression

Random forest regression is a bagging algorithm widely used to predict continuous outcomes. During training, this algorithm randomly constructs a family of models [41]. It generates multiple decision trees using a strategy known as bootstrapping and introduces a random element that takes into account a subset of features at each split. Three hyperparameters – namely, the number of features used to construct each tree, the number of trees, and the minimum number of tree nodes [42] – were used to optimize the algorithm. The final prediction was obtained by averaging the results of all the individual trees. This averaging helped improve the model's robustness, while also reducing the risk of overfitting.

3.5.4. Support vector machines

Support Vector Machines (SVMs) are supervised learning algorithms widely recognized for their strong foundations in statistical learning theory. These techniques are also known for their ability to effectively solve nonlinear regression problems. They

were originally defined for the classification of a qualitative or binary variable. In contrast to traditional techniques, SVMs formulate relationships as convex optimization problems. These problems aim to find a function that approximates the data within a certain tolerance. SVMs are based on finding the optimal margin hyperplane [43], which is defined by maximizing the distance between the closest points to the plane. One of the advantages of this technique is the robustness of the model.

3.5.5. Validation of models

The mean squared error, root mean squared error, mean absolute error, and coefficient of determination were used to evaluate the models' performance.

Mean squared error

The mean squared error (MSE) is a widely used performance metric for evaluating regression models in machine learning. It is defined as the average of the squares of the differences between the predicted values and the real values in a sample of size n (Equation 10). The MSE unit is the square of the unit of the target variable [44].

$$MSE = (1/n)\sum(y_{true} - y_{pred})\, exp2 \qquad (10)$$

Root mean square error

The root mean square error (RMSE) is, like the previous one, a widely used metric for evaluating the performance of regression models in machine learning, where the goal is to predict continuous outcomes. This parameter is defined as the square root of the average of the squared differences between the predicted values and the real values (Equation 11). It expresses the standard deviation of the residuals and has the same unit as the predicted variable [45].

$$RMSE = \sqrt{(1/n)\sum(y_{true} - y_{pred})exp2} \qquad (11)$$

Mean absolute error

The mean absolute error (MAE) is also a commonly used intuitive metric for evaluating the performance of regression models in machine learning. This parameter is defined as the average of the absolute differences between the predicted values and the true values across *n* data points (Equation 12). It quantifies the average range of deviations between predicted and actual values, thereby providing a direct measure of the accuracy of the predictions [46].

$$MAE = (1/n)\sum|y_{true} - y_{pred}| \qquad (12)$$

Coefficient of determination

192 The coefficient of determination (R^2) is one of the most commonly used metrics for evaluating the performance of regression
193 models. It was used in this study to assess the explanatory power of the models. This power is defined as the capacity to quantify the
194 proportion of the variance in the dependent variable that can be predicted based on the independent variables [47, 48] R^2 was
195 determined using Equation 13.

196 $$Rexp2 = 1 - [\sum(y_{true} - y_{pred})exp2 / \sum(y_{true} - \bar{y}_{pred})exp2] \qquad (13)$$

197 **4. Results and Discussion**

198 **4.1. Macroscopic Petrography**

199 The studied materials exhibit reddish-brown to reddish color profiles and visible thicknesses generally ranging from 1.5 to
200 5.5 m. Most of these profiles consist of a multi-level altered set. From the base to the top, iron-rich nodular layers are often clearly
201 distinguishable from the soft clay layers. The nodular layers vary in thickness from 1.2 to 4.3 m. The nodules range in size from
202 millimeters (20–35%) to several centimeters (25–40%) and have angular to subangular and subrounded to rounded shapes (Fig. 3).
203 They are packed in clayey-loamy matrices (15–25%). The sizes of these nodules gradually decrease from the bottom to the top of
204 the profiles in certain sampling sites. These nodules are found in other deposit sites, organized into coarse and fine nodular layers,
205 separated by sharp and undulating boundaries. They are associated with quartz fragments ranging in size from a few millimeters to
206 several centimeters, with angular to subangular shapes, and accounting for 5 to 15 percent of the total. The variations in size, shape,
207 and proportions of the quartz fragments and iron nodules suggest the effects of intense differential dissolution and weathering
208 processes on the underlying rock [17]. These effects are thought to have been conditioned by the heterogeneity of the rock, by the
209 topography, and also by the hydromorphic nature of the environment [9, 11, 12].

210 **4.2. Physical and mechanical properties**

211 Table 1 shows only the average values of the parameters measured at each sampling site. Details are provided in the
212 supplementary data.

213 **4.2.1. Specific gravity**

214 The γ_s values of the studied lateritic gravels range from 2.40 to 3.00 g/cm^3, with an average of 2.73 g/cm^3 (Table 2). 93% of
215 these materials have γ_s values of ≥ 2.6 g/cm^3 (see supplementary data). The γ_s values of the studied lateritic gravels are therefore well
216 within the range of γs values (2.6–3.2 g/cm^3) typically found in equatorial regions [32]. γ_s is a key geotechnical parameter in road
217 construction. The values of this parameter are generally closely related to the mineralogical composition of the particles that make
218 up the material [13,18]. These values influence soil behavior, including its compaction characteristics, strength, and bearing capacity.
219 The higher the γ_s values, the more significant the aforementioned characteristics are [18,49]. The average γ_s values for lateritic gravels

at each quarry site (Table 1) are all greater than 2.58 g/cm^3. These values are closer to 2.85 g/cm^3, a threshold above which lateritic gravels should be considered to perform well [50].

4.2.2. Granularity

The < 2 mm fraction, the skeleton (> 2 mm), the mortar (< 425 µm), and the fine fraction (< 75 µm) were used to characterize the studied materials. These fractions are the most commonly considered when assessing the grain size of lateritic gravels [10]. The content of the < 2 mm fraction ranges from 14.6% to 99.4%. Those of the skeleton range from 0.6% to 85.4%. The mortar and the fine fraction have contents ranging from 12.3% to 84.6% and from 7.3% to 64%, respectively. The average contents of these grain size classes (Table 2) are similar to those of lateritic gravels observed in eastern Cameroon [51]. The particle size distribution of lateritic gravel affects its behavior and suitability for road construction. Materials with high fines content and correspondingly high plasticity indices generally have low bearing capacity. Conversely, those with high concentrations of coarse particles – which are generally much less plastic – very often exhibit high bearing capacities and are well suited for use in subbase and base layers in road construction [13].

4.2.3. Plasticity

Liquidity limit (LL) values range from 39% to 68%, with an average of 55%. The plasticity index (PI) averages 25%, with values ranging from 16% to 36% (Table 2). The average values for these two parameters are higher than those for lateritic gravels in dry or humid savannas [12]. They are also higher than those of soils used for road construction in Yobe State, Nigeria [18]. However, these values are lower than those obtained in the lateritic soils of eastern Cameroon [51]. Plotting the LL and PI values of the studied materials on the Casagrande plasticity diagram (Fig. 4) shows that the mortars (< 425 µm) that make up these materials are moderately clayey to clayey. These mortars are characterized by medium to high plasticity. Their distribution along both sides of the "U" line attests to the high quality of the measured data, which, based on these findings alone, would suggest a significant reduction in performance in building construction. In fact, the practical guide to pavement design for tropical countries recommends PI values in the range of 20–30% and PI < 15% for materials to be used in the construction of the subbase and base layers of high-traffic pavements [52]. The higher the PI values, the greater the compressibility and swelling potential of the soils, thereby reducing their shear strength [53]. The studied materials, considered at the level of the quarries, have average values of $20 \leq PI \leq 31\%$ (Table 1). Based on the conditions mentioned earlier, these materials are not recommended for use as base layers but are primarily suitable as subbase layers.

4.2.4. Derived parameters

The values of the grading modulus (Gm) range from 0.54 to 2.63, with an average of 2.13. Those of the plasticity modulus (Pm) range from 274 to 2639. The average value of this parameter is 732. The plasticity product (Pp) and the swelling potential (ε_s) have values ranging from 173 to 2058 ($\bar{X} = 600$) and from 0.002 to 0.034 ($\bar{X} = 0.016$), respectively (Table 2). Each of these parameters

plays an important role in evaluating the performance and mechanical behavior of pavement layers [10]. Gm, for example, is fundamentally used to characterize the elastic response of materials subjected to repeated traffic loads. 98% of the studied materials (see supplementary data) have Gm values greater than 1.5. This last value represents the minimum threshold below which, for low-traffic roads, lateritic gravels cannot be used as a base layer [50]. Pm and Pp highlight the impact of the plasticity of fine particles on the overall performance of materials. This impact often affects the swelling and compressibility of materials, their strength, and, consequently, the durability of the structures built with them. The use of lateritic gravels in road construction depends on the values of these parameters. The lower these values are, the more they lead to improved performance. For example, the construction of pavement subbase layers for all types of traffic in tropical regions requires Pm and Pp values of less than 500 and 600, respectively [32]. Based on the Pm values, very few (≈ 20%) of the studied materials are suitable for use, but based on the Pp values, a moderate proportion (≈ 53%) is suitable (see supplementary data). ε_s is closely related to the fine fraction of soils, specifically the clay fraction, but also to the presence of certain mineral phases. An increase in the fine-particle content increases the surface roughness of the material. Surface roughness, in turn, increases the affinity for water, which not only leads to an increased risk of swelling but also to an increased risk of creep deformation under prolonged loads [54]. The ε_s values in the studied materials are low. The average value for this parameter is higher than that of lateritic soils in humid and dry savannas [12]. This is also true for lateritic gravels developed on mica schist [13].

4.2.5. Compaction Characteristics

The optimum moisture content (OMC) values for the studied materials range from 7.7 to 22.2%. Those for maximum dry density (MDD) range from 1.62 to 2.16 g/cm^3. The average values for these two parameters are 12.2% and 2.00 g/cm^3, respectively (Table 2). OMC and MDD are the main parameters that determine the optimal compaction of a material. Their controls help prevent problems such as a loss of stiffness and stability in pavement layers [55]. The average OMC value in the studied lateritic gravels is significantly lower than that of lateritic soils (OMC = 28%) in certain tropical regions [49]. It is, however, enclosed by values of wet (12.7%) and dry savanna (11.8%) in Central Cameroon [12]. This value, like the average values at each quarry (Table 1), is greater than 8%, the threshold above which materials are likely to perform less effectively [56]. The CEBTP [52], however, extends this limit to 15%, thereby allowing the use of these materials at all of the studied quarries. In contrast to OMC, the average MDD value for the studied materials is significantly higher than those of the lateritic soils studied by Wahab et al. [49]. This value also falls within the range of the densities of materials found in humid savanna (1.95 g/cm^3) and dry savanna (2.06 g/cm^3) zones [12]. The Practical Guide to Pavement Design for Tropical Countries specifies values of MDD ≥ 2.00 g/cm^3 and MDD ≥ 1.80 g/cm^3 for materials to be used in base and subbase layers of pavements, respectively [52]. These values are, in fact, essential for ensuring their load-bearing capacity and durability [18]. With the exception of four samples from the quarries associated with kilometric points

(KP) 11+000, 20+350, 23+200, and 39+700 (see supplementary data), the materials at all sites are suitable for use as a subbase layer, while approximately 47% are suitable for use as a base layer (Table 1). Compaction characteristics, like derived parameters, are closely related to plasticity. Materials with high PI values are generally characterized by high OMC values. The latter have a direct impact on MDD and, subsequently, on bearing capacity, thereby influencing construction practices [4, 18, 57].

The values of the California Bearing Ratio (CBR) at 95% of the MDD of the studied materials range from 12% to 94% ($\bar{X}$ = 34%). The average value of this parameter is 26% lower than that of lateritic gravels in eastern Cameroon [51], 2% and 19% lower, respectively, than those of lateritic gravels in humid and dry savannas [12], and 13% lower than that of lateritic gravels in Nigeria [50]. The CBR plays a crucial role in road construction, particularly in assessing the suitability of materials. The conditions $25 \leq$ CBR $\leq 35\%$ and $60 \leq$ CBR $\leq 80\%$ must be met for materials to be used in the subbase and base layers of pavements, respectively [52]. 74% of the collected material samples and 79% of the quarries have CBR values $\geq 25\%$ and are therefore suitable for use as a subbase layer. Only 6% of these samples, compared with 0% of the quarries, are suitable for use as a base layer (see supplementary data; Table 1). The iso-CBR variogram, however, shows that approximately 50% of the surface covered by the study area contains materials suitable for use as a subbase layer, compared to 0% for the base layer. This surface is mainly located in the central and eastern parts of the study area (Fig. 5). The volumes of lateritic gravels have a significant impact on their uses. Their availability, which is often assessed based on the size of the quarry, has an impact on construction costs [7]. 84% of the quarries (Table 1) for the studied laterite gravels meet the criteria mentioned above and are suitable for use as a subbase layer. None of these sites, however, are classified as materials for base layers unless adjustments are made. All of these sites, nevertheless, meet the criterion $15 \leq$ CBR $\leq 30\%$ (Fig. 6) established by the Directorate of General Studies and Standardization of the Ministry of Public Works in Cameroon and are specifically suitable for use as a subbase layers in low-traffic roadways (T1). 68% of these sites (Table 1), compared with 58% as shown by the iso-CBR variogram ($30 \leq$ CBR $\leq 60\%$), are suitable for use as a subbase layer in pavements for medium-traffic conditions (T2/T3), as well as for use as a base layer in pavements for T1 traffic [58].

4.3. Variability

4.3.1. Descriptive statistics

The simple statistical analysis of variables for the studied materials shows that the measured parameter values fluctuate around the values of $\bar{X}$. These values slightly differ from those of $\bar{X}$ for some of the parameters but considerably for others (see supplementary data; Table 2). The resulting ranges of values ($0.03 \leq$ Max – Min ≤ 2365) suggest both low and high levels of data disparities. The standard deviation (τ) measures in fact the average distance between each data point and $\bar{X}$. Lower values of τ would indicate a tight clustering of data around $\bar{X}$.

These clustering would then suggest consistencies or low dispersion in the data. Higher values of τ, on the other hand, would result in high dispersion and, consequently, high data variability [59]. An examination of the τ values (0.09–318) shows that the data for the observed variables are even more dispersed than each other. These dispersions are all the more pronounced because the relevant data do not follow the empirical rules associated with normal distributions ($\bar{X} \pm 1\tau \neq 68\%$ and $\bar{X} \pm 2\tau \neq 95\%$) (Table 2). γ_s and the parameters for compaction, plasticity, granularity, and bearing capacity reflect successive levels of data dispersion relative to the overall data [59]. The previous observations are supported by those related to the coefficient of variation (CV). The values for this statistical parameter range from 4 to 39 (Table 2). It is a widely used variable [13] that expresses the extent of variability relative to the mean of a dataset. It provides a standardized measure of dispersion, facilitating comparisons between datasets with different units. Smaller CV values, like those for τ, indicate low variability, and vice versa. The classification by Wilding and Dress [60] was nevertheless used to compare the variability of geotechnical parameters in the studied lateritic gravels. This classification specifies that for CV values < 15, the observed parameter slightly varies. For $15 \leq CV \leq 35$, the observed parameter moderately varies. For CV > 35, the observed parameter highly varies. The CV values are recorded in Table 2. Based on the above criteria, these values show that γ_s, MDD, LL, and Gm slightly vary; PI, OMC, the fraction < 2 mm, and the skeleton moderately vary; and that the mortar, ε_s, the fine fraction, the CBR, Pm, and Pp highly vary. The low variation in γ_s could be related to the mineralogical composition of the solid particles. The latter would highlight the nature and low variability of the substratum from which the studied lateritic gravels are derived [3, 13]. The variability in PI confirms the positions occupied by the mortars of the studied materials in the Casagrande plasticity diagram.

4.3.2. Correlations

The correlation matrix of the variables (Fig. 7) shows 91 levels of correlation, of which 55% are positive and 45% are negative. ≈ 58% of these correlations are low, ≈ 15% are average and ≈ 26% are high. The γs–MDD correlation shows that the slight variation in specific gravity (Table 2) appears to have had a moderate effect on the maximum dry density values [18, 49]. The less than 2 mm fraction behaves in a manner that is antagonistic to the skeleton. These two variables are positively and negatively correlated, respectively, with mortar, the fine fraction, Pm, and Pp, and inversely correlated with Gm. The medium to high plasticity of the studied materials appears to be more closely related to factors other than grain size. No significant correlation was observed between the CBR and the other variables. This means that no single variable, taken individually, would have influenced the high variability of the CBR. The latter would then result from the combined effects of several factors [12], including variations in plasticity, particle size, shape, and proportions, as well as derived and compaction parameters [4, 13, 18, 57].

4.3.3. Principal Components

The percentage of cumulative variance (Fig. 8; Table 3) presented by the F1 and F2 components is not high enough to allow for a reliable analysis. The F3 component (84.87%) was therefore added to reduce the possibilities of misinterpreting the graphs. The results related to the contributions of the variables to the construction of the F1, F2, and F3 components show that 11 of the 14 variables used (γ_s, < 2 mm, > 2 mm, < 425 µm, < 75 µm, LL, PI, Gm, Pm, Pp, and ε_s) are well expressed. OMC, MDD, and CBR are not ($Cos^2 < 0.5$) and therefore cannot be the subject of valid interpretations. These results are obtained using the method of squared cosines (Table 4). The resulting correlation circles F1–F2 (Fig. 9a), F1–F3 (Fig. 9b), and F2–F3 (Fig. 9c) all confirm the observations made earlier in the correlation section. These include the possible influence of γ_s on MDD, as well as regressions of the skeleton and Gm relative to < 2 mm, < 425 µm, < 75 µm, Pm, and Pp. The circles further demonstrate that the moderately to highly plastic properties of the studied materials, combined with their swelling potentials, are independent of their particle sizes. Furthermore, the CBR cannot be derived from any single variable on its individual merits.

4.4. Machine Learning Models

The metric results of the machine learning algorithms are recorded in Table 5. The mean squared error (MSE) values for all combined learning techniques range from 7.9 to 162.5. MSE provides numerical values that help to compare the performance of different learning algorithms. Low values for this metric parameter mean, in terms of mean squared error, that the predictions are, on average, closer to the actual values. These low values thus suggest greater predictive accuracy [46]. The MSE values in this study are moderate to very high. These values are higher than the MSE values obtained for lateritic gravels derived from granites in eastern Cameroon [48]. The very high MSE values obtained in this study are mostly associated with multiple linear regressions and ridge regressions. These two machine learning techniques, it should be noted, are forms of extensions of simple linear regressions (SLR), which themselves very often produce moderate to weak correlations. The very high MSE values associated with multiple linear regressions and ridge regressions could therefore, in line with SLR, indicate low predictive capacities [4] of the CBR. The high MSE values recorded during the execution of random forest and support vector machine regressions could, however, have an entirely other meaning. These two other learning algorithms, it should also be remembered, are optimizing techniques [42, 43] capable of solving linear regression problems, whose final results are characterized by the robustness of the models. MSE itself is reportedly very sensitive to outliers. In fact, squaring the error terms may have allowed this metric to give greater weight to large deviations than to minor ones [44], thereby influencing the quantification of the overall performance of the models implemented in this study.

The root mean square error (RMSE) and the mean absolute error (MAE) were used to limit potential overestimates resulting from the relative influence of large deviations compared to the minor ones. These two metrics facilitate an intuitive assessment of the performance of machine learning models [46, 61]. The RMSE values in this study range from 2.8 to 12.7 (Table 5). Those of MAE range from 0.7 to 10.2. These values are moderate in MLR and ridge regression. They are also moderate in random forest and

SVM regressions, specifically for the plasticity and compaction variables. A low RMSE and MAE, like the MSE, indicate that the predictions are reasonably close to the actual values, thus reflecting better model performances [45]. The SVMs implemented on particle size variables and derived parameters (RMSE = 2.8 and MAE = 0.7) offer, based on the aforementioned evaluation criteria, the best performance in terms of CBR predictions in this study. These observations are confirmed by the values of the coefficient of determination ($R^2 = 0.94$). This other metric indicates the extent to which a model's predicted values would match the observed data. An R^2 value close to 1 (Fig. 10; Fig. 11; Table 5) or even equal to 1 would indicate a perfect prediction, in which the model accounts for all the variability, and would reflect a better fit. In contrast, an R^2 very close to 0 or even equal to 0 would mean that the model performs no better than simply predicting the mean of the data [47]. The random forest regressions practiced on the same granulometric and derived variables (RMSE $\leq$ 4.8, MAE = 3.8, and $R^2 \geq 0.85$) perform secondarily to those of SVMs.

5. Conclusion

Lateritic gravels are materials of great importance in road construction. The CBR is the main geotechnical parameter that guides decisions regarding their applications. The assessment of performances and the prediction of the behaviors of pavement layers using this parameter are challenges for some time in tropical and subtropical regions. These challenges require establishing relationships between variables, which considerably fluctuate. The lateritic quarries located on either side of the road section connecting the localities of Esse and Soa were the subject of this exercise. These two localities are located in a tropical region, specifically in Central Cameroon. The results of geotechnical tests conducted on collected lateritic gravels show that these materials can be used as subbase and base layers for low-traffic roadways. They are, however, moderately recommended for use as a subbase layer in roadways with medium traffic volumes. The properties and performances of these materials considerably vary, although they are developed in climatic, topographic, phytogeographic, and geological contexts that vary little. These variations reflect the generally complex nature of tropical soils. The machine learning algorithms, such as multiple linear regressions and ridge regressions, which are used to build relationships and predict the CBR performances from data on other geotechnical variables, have – like SLR and PCA – proved to be limited. The resulting models are characterized by relatively high MSE, RMSE, and MAE values and very low R^2 values. The random forest and support vector machine regressions, in contrast, based on the same evaluation criteria as the previous machine learning algorithms, were able to generate successively satisfactory models. These models are characterized by low to very low MSE, RMSE, and MAE values, and high to very high R^2 values. Particle size variables and derived parameters were found to be the most suitable parameters for predicting the CBR in the studied materials.

References

[1] Horbe, A. M. C., & Da Costa, M. L. (2005). Lateritic crusts and related soils in eastern Brazilian Amazonia. *Geoderma*, 126(3–4), 225–239. https://doi.org/10.1016/j.geoderma.2004.09.011

[2] Biswal, D. R., Sahoo, U. C., & Dash, S. R. (2016). Characterization of granular lateritic soils as pavement material. *Transportation Geotechnics*, 6, 108–122. https://doi.org/10.1016/j.trgeo.2015.10.005

[3] Onana, V. L., Ngo'o Ze, A., Medjo Eko, R., Ntouala, R. F. D., Nanga Bineli, M. T., Ngono Owoudou, B., & Ekodeck, G. E. (2017). Geological identification, geotechnical and mechanical characterization of charnockite-derived lateritic gravels from Southern Cameroon for road construction purposes. *Transportation Geotechnics*, 10, 35–46. https://doi.org/10.1016/j.trgeo.2016.12.001

[4] Katte, V. Y., Mfoyet, S. M., Manefouet, B., Wouatong, A. S. L., & Bezeng, L. A. (2019). Correlation of California Bearing Ratio (CBR) Value with Soil Properties of Road Subgrade Soil. *Geotechnical and Geological Engineering*, 37, 217–234. https://doi.org/10.1007/s10706-018-0604-x

[5] Tang, K., Zeng, F., Shi, L., Zhu, L., Chen, Z., & Zhang, F. (2023). Mechanical Behavior of Hydrated-Lime–Liquid-Stabilizer-Treated Granular Lateritic Soils. *Sustainability*, 15(6), 5601. https://doi.org/10.3390/su15065601

[6] El-Din Fawzy, H., Mohamdy Abdallah, A., Basha, A., & Elgamal, E. (2025). Review of Pavement Performance Indicators: Measurements, Applications, and Limitations. *IOP Conference Series: Earth and Environmental Science*, 1530, 012036. DOI 10.1088/1755-1315/1530/1/012036

[7] Leśniak, A., & Zima, K. (2018). Cost Calculation of Construction Projects Including Sustainability Factors Using the Case Based Reasoning (CBR) Method. *Sustainability*, 10(5), 1608. https://doi.org/10.3390/su10051608

[8] Marathe, S., & Sheshadri A. (2026). Machine-Learning-Based Prediction of California Bearing Ratio (CBR) for Lateritic–Lithomargic Pavement Subgrades through the Major Index Properties. *Indian Geotechnical Journal*, 56(6), 2779–804. https://doi.org/10.1007/s40098-025-01297-1

[9] Ekodeck, G. E. (1984). L'altération des roches métamorphiques du Sud Cameroun et ses aspects géotechniques. Thèse de Doctorat d'Etat des Sciences Naturelles, IRIGM. Université Scientifique et Médicinal de Grenoble I, France, 375. http://pascal-francis.inist.fr/vibad/index.php?action=getRecordDetail&idt=8557972.

[10] Bagarre, E. (1990). Utilisation des graveleux latéritiques en technique routière. ISTED, 143. https://trid.trb.org/View/1018608

[11] Beyala, V. K. K., Ekodeck, G. E., & Konfor, N. I. (2001). Evolution géotechnique des formations latéritiques dans l'écosystème périforestier du Sud-Est Cameroun : le site de Kandara. *African Journal of Science and Technology*, 2 (1), 19–32. https://doi.org/10.4314/ajst.v2i1.44631

[12] Mvindi, A. T. N., Onana, V. L., Ngo'o Ze, A., Ohandja, H. N., & Ekodeck, G. E. (2017). Influence of hydromorphic conditions in the variability of geotechnical parameters of gneiss-derived lateritic gravels in a savannah tropical humid area (Centre Cameroon), for road construction purposes. *Transportation Geotechnics*, 12, 70–84. https://doi.org/10.1016/j.trgeo.2017.08.003

[13] Ngo'o Ze, A., Onana, V. L., Ndzié Mvindi, A. T., Nyassa Ohandja, H., Medjo Eko, R., & Ekodeck, G. E. (2019). Variability of geotechnical parameters of lateritic gravels overlying contrasted metamorphic rocks in a tropical humid area (Cameroon): implications for road construction. *Bulletin of Engineering Geology and the Environment*, 78(8), 5531–5549. https://doi.org/10.1007/s10064-019-01488-0

[14] Mveng, M. D. M. N., Zame, P. Z., Bayamack, J. F. N. Bisso, D., & Atangana, J. Q. Y. (2024). Spatial variability of geotechnical parameters of lateritic gravels in an Olama–Ngomedzap road section in a humid tropical area (Cameroon). *Discover Applied Sciences*, 6(7) 355. https://doi.org/10.1007/s42452-024-06012-y

[15] Bayamack, J. F. N., Onana, V. L., Mvindi, A. T. N., Ngo'o Ze, A., Ohandja, H. N., & Medjo Eko, R. (2019). Assessment of the determination of Californian Bearing Ratio of laterites with contrasted geotechnical properties from simple physical parameters. *Transportation Geotechnics*, 19, 84–95. https://doi.org/10.1016/j.trgeo.2019.02.001

[16] Popescu, R., Deodatis, G., & Nobahar, A. (2005). Effects of random heterogeneity of soil properties on bearing capacity. *Probabilistic Engineering Mechanics*, 20(4), 324–341. https://doi.org/10.1016/j.probengmech.2005.06.003

[17] Krinitzsky, E. L., Patrick, D. M., & Townsend, F.C. (1976). Geology and geotechnical properties of laterite gravel. United States Army Engineer Waterways Experiment Station, Technical Report S-76-5, Vicksburg. https://books.google.com/books?hl=fr&lr=&id=W9PC8r2mVnEC&oi=fnd&pg=PA7&dq=Geology+and+geotechnical+properties+of+laterite+gravel&ots=faQ_9eLgxB&sig=TUxnftef8ZarXE2DCZvTh3UNpt0

[18] Umar, I. H., Salisu, I. M., Lin, H., & Hassan, J. I. (2024). Soil characterization, CBR modeling, and spatial variability analysis for road subgrade: a case study of Danchuwa–Jajere Road, Yobe State, Nigeria. *Engineering Research Express*, 6(3), 035119. DOI 10.1088/2631-8695/ad78a5

[19] Kurnaz, T. F., & Kaya, Y. (2019). Prediction of the California bearing ratio (CBR) of compacted soils by using GMDH-type neural network. *The European Physical Journal Plus*, 134(7), 326. https://doi.org/10.1140/epjp/i2019-12692-0.

[20] Tenpe, A. R., & Patel, A. (2020). Application of genetic expression programming and artificial neural network for prediction of CBR. *Road Materials and Pavement Design*, 21(5), 1183–1200. https://doi.org/10.1080/14680629.2018.1544924

[21] Trong, D. K., Pham, B. T., Jalal, F. E., Iqbal, M., Roussis, P. C., Mamou, A., Ferentinou, M., Vu, D. Q., Duc Dam, N., Tran, Q. A., & Asteris, P. G. (2021). On Random Subspace Optimization-Based Hybrid Computing Models Predicting the California Bearing Ratio of Soils. *Materials*, 14(21), 6516. https://doi.org/10.3390/ma14216516

[22] Kassa, S. M., & Wubineh, B. Z. (2023). Use of Machine Learning to Predict California Bearing Ratio of Soils. *Advances in Civil Engineering*, 2023, 1–11. https://doi.org/10.1155/2023/8198648

[23] Olivry, J. C. (1986). Fleuves et rivières du Cameroun. Collection « Monographies Hydrologiques ORSTOM » N° 9, Paris, 592.

[24] Letouzey, R. (1985). Notice de la carte phytogéographique du Cameroun au 1 : 500 000. Institut de la Carte Internationale de la Végétation, Toulouse, France, et Institut de la Recherche agronomique (Herbier National), Yaoundé, Cameroun, 5.

[25] Onana, J. B., Njom, B., Owono, F. M., Yannah, M., Ango, T. A., & Ondoa, J. M. (2023). Relative active tectonic in the metamorphic rocks of the Yaounde group: insights from geomorphic indices and topographic analysis. *Journal of Mountain Science*, 20(10), 2946-2974. https://doi.org/10.1007/s11629-023-7971-3.

[26] Dumort, J. F. (1968). Notice explicative sur les feuilles de Yaoundé Est et Douala Est, carte géologique de reconnaissance au 1/500 000. Direction Mines et Géologique, Yaoundé, Cameroun, 69.

[27] Mvondo, H., Owona, S., Mvondo Ondoa, J., & Essono, J. (2007). Tectonic evolution of the Yaoundé segment of the Neoproterozoic Central African Orogenic Belt in southern Cameroon. *Canadian Journal of Earth Sciences*, 44(4), 433–444. https://doi.org/10.1139/e06-107

[28] AFNOR (1996). Sols: reconnaissance et essais. Analyse granulométrique. Méthode par tamisage à sec après lavage. Association Française de Normalisation, Paris.

[29] AFNOR (1992). Sols: reconnaissance et essais. Analyse granulométrique. Méthode par sédimentation. Association Française de Normalisation, Paris.

[30] AFNOR (1991). Sols: reconnaissance et essais. Détermination de la masse volumique des particules solides des sols – méthode du pycnomètre à eau. Association Française de Normalisation, Paris.

[31] AFNOR (1993). Sols: reconnaissance et essais. Détermination des limites d'Atterberg. Limite de liquidité à la coupelle. Limite de plasticité au rouleau. Association Française de Normalisation, Paris.

[32] Charman, J. H. (1988). Laterite in road pavements. London Construction Industry Research and Information Association. Special Publication, 47, London: CIRIA, 71. https://trid.trb.org/View/708714

[33] Millogo, Y., Karfa, T., Raguilnaba, O., Kalsibiri, K., Blanchart, P., & Thomassin, J. H. (2008). Geotechnical, mechanical, chemical and mineralogical characterization of lateritic gravels of Sapouy (Burkina Faso) used in road construction. *Construction and Building Materials*, 22(2), 70–76. https://doi.org/10.1016/j.conbuildmat.2006.07.014

[34] AFNOR (1999). Sols: Reconnaissance et essais. Détermination des références de compactage d'un matériau. Essai Proctor normal – Essai Proctor modifié. Association Française de Normalisation, Paris.

[35] AFNOR (1997). Sols: reconnaissance et essais. Indice CBR après immersion. Indice CBR immédiat. Indice Portant Immédiat – Mesure sur échantillon compacté dans le moule CBR. Association Française de Normalisation, Paris.

[36] Hutchinson, M. F., Xu, T., & Stein, J. A. (2011). Recent Progress in the ANUDEM Elevation Gridding Procedure. *Geomorphometry*, 2011(2), 19–22.

[37] Kumar, R., Dwivedi, S. B., & Gaur, S. (2021). A comparative study of machine learning and Fuzzy AHP technique to groundwater potential mapping in the data-scarce region. *Computers & Geosciences*, 155, 104855. https://doi.org/10.1016/j.cageo.2021.104855

[38] Uyanık, G. K., & Güler, N. (2013). A study on multiple linear regression analysis. *Procedia-Social and Behavioral Sciences*, 106, 234-240. https://doi.org/10.1016/j.sbspro.2013.12.027

[39] McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, 1(1), 93–100. https://doi.org/10.1002/wics.14

[40] Saunders, C., Gammerman, A., & Vovk, V. (1998). Ridge regression learning algorithm in dual variables.

[41] Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[42] Razavi-Termeh, S. V., Sadeghi-Niaraki, A., & Choi, S. M. (2019). Groundwater potential mapping using an integrated ensemble of three bivariate statistical models with random forest and logistic model tree models. *Water*, 11(8), 1596. https://doi.org/10.3390/w11081596

[43] Prasad, P., Loveson, V. J., Kotha, M., & Yadav, R. (2020). Application of machine learning techniques in groundwater potential mapping along the west coast of India. *GIScience & Remote Sensing*, 57(6), 735–752. https://doi.org/10.1080/15481603.2020.1794104

[44] Hassanat, A. B., Alqaralleh, M. K., Tarawneh, A. S., Almohammadi, K., Alamri, M., Alzahrani, A., Altarawneh, G. A., & Alhalaseh, R. (2024). A Novel Outlier-Robust Accuracy Measure for Machine Learning Regression Using a Non-Convex Distance Metric. *Mathematics*, 12(22), 3623. https://doi.org/10.3390/math12223623

[45] Jain, S., Bhowmik, A., & Lee, J. (2025). Machine learning approaches for predicting and validating mechanical properties of Mg rare earth alloys for light weight applications. *Science and Technology of Advanced Materials*, 26(1), 2449811. https://doi.org/10.1080/14686996.2025.2449811

[46] Jonayet, M., Das, A. C., Mozumder, M. S. A., Hasan, M. A., Bhuiyan, M., Islam, M. R., Hossain, M. N., Akter, S., & Alam, M. I. (2024). Machine learning approaches for demand forecasting: the impact of customer satisfaction on prediction accuracy. *The American Journal of Engineering and Technology*, 6(10), 42–53. https://doi.org/10.37547/tajet/volume06issue10-06

[47] Plevris, V., Solorzano, G., Bakas, N., & Ben Seghier, M. (2022). Investigation of performance metrics in regression analysis and machine learning-based prediction models. https://doi.org/10.23967/eccomas.2022.155

[48] Mvindi, W. P. M., Mvindi, A. T. N., Ohandja, H. N., Tchanfa, G. N., Mebouinz, D. L., Tang, S. D. N., & Onana, V. L. (2026). Assessment of CBR of granite-derived lateritic gravels using machine-learning models with derived parameters. *Arabian Journal of Geosciences*, 19(5), 77. https://doi.org/10.1007/s12517-026-12454-y

[49] Wahab, N. A., Rashid, A. S. A., Roshan, M. J., Rizal, N. H. A., Yunus, N. Z. M., Hezmi, M. A., & Tadza, M. Y. M. (2021). Effects of Cement on the Compaction Properties of Lateritic Soil. *IOP Conference Series: Materials Science and Engineering*, 1153(1), 012015. https://doi.org/10.1088/1757-899x/1153/1/012015

[50] Nwaiwu, C. M. O., Alkali, I. B. K., & Ahmed, U. A. (2006). Properties of ironstone lateritic gravels in relation to gravel road pavement construction. *Geotechnical & Geological Engineering*, 24(2), 283-298. https://doi.org/10.1007/s10706-004-6374-7

[51] Nzabakurikiza, A., Onana, V. L., Ngo'o Ze, A., Mvindi, A. T. N., & Ekodeck, G.E. (2016). Geological, geotechnical, and mechanical characterization of lateritic gravels from Eastern Cameroon for road construction purposes. *Bulletin of Engineering Geology and the Environment*, 76(4), 1549–1562. https://doi.org/10.1007/s10064-016-0979-y

[52] CEBTP (1984). Guide pratique de dimensionnement des chaussées pour les pays tropicaux. Centre d'expertise du bâtiment et des travaux publics, Saint-Rémy-lès-Chevreuse, 155.

[53] Olofinyo, O. O., Olabode, O. F., & Fatoyinbo, I.O. (2019). Engineering properties of residual soils in part of Southwestern Nigeria: implication for road foundation. *SN Applied Sciences*, 1(5), 507. https://doi.org/10.1007/s42452-019-0515-3

[54] Courard, L., Rondeux, M., Zhao, Z., & Michel, F. (2020). Use of Recycled Fine Aggregates from C & DW for Unbound Road Sub-Base. *Materials*, 13(13), 2994. https://doi.org/10.3390/ma13132994

[55] Tophel, A., Nguyen, T. M., Walker, J. P., & Kodikara, J. (2025). Soil Moisture Prediction in Pavement Layers Using LSTM Neural Networks. *Geotechnical and Geological Engineering*, 43(6), 318. https://doi.org/10.1007/s10706-025-03293-x

[56] Gidigasu, M. D. (1974). Degree of weathering in the identification of laterite materials for engineering purposes: a review. *Engineering Geology*, 8(3), 213–266. https://doi.org/10.1016/0013-7952(74)90001-5

[57] Sivrikaya, O. (2007). Models of compacted fine-grained soils used as mineral liner for solid waste. *Environmental Geology*, 53(7), 1585–1595. https://doi.org/10.1007/s00254-007-1142-7

[58] DEGN (1987). Recommandation pour l'utilisation en corps de chaussées de graveleux latéritiques naturels. Recommandation 30.004-R. République du Cameroun, Ministère de l'équipement.

[59] Rakrak, M. (2025). Exploring Variability in Data: The Role of Range, Variance, and Standard Deviation. *International Journal of Multidisciplinary Research and Analysis*, 08(03), 1327–1331. https://doi.org/10.47191/ijmra/v8-i03-47

[60] Wilding, L.P., & Dress, L.R. (1983). Spatial variability and pedology. In: Wilding, L.P., Smeck, N., & Hall, G. F. (eds) Pedogenesis and soil taxonomy. Wageningen, the Netherlands, 83–116.

[61] Charde, M. M., Najan, T. P., Cepova, L., Jadhav, A. D., Rash-Inkard, N. S., & Samal, S. P. (2025). Predictive Modelling of Surface Roughness in Grinding Operations Using Machine Learning Techniques. *Manufacturing Technology*, 25(1), 14–23. https://doi.org/10.21062/mft.2025.006

539 **Table 1** Average values of geotechnical parameters of the studied lateritic gravels in each quarry

Quarry	Number of samples (N = 159)	γs (g/cm^3)	< 2 mm (%)	> 2 mm (%)	< 425 µm (%)	< 75 µm (%)	LL (%)	PI (%)	Gm	Pm	Pp	εs	OMC (%)	MDD (g/cm^3)	CBR (%)
1+500	12	2.76	38.3	61.7	28.8	25.2	47	20	2.08	582	510	0.009	12.4	2.01	46
3+600	6	2.71	36.1	63.9	29.1	23.5	57	27	2.11	785	637	0.018	12.9	1.96	44
9+300	10	2.78	35.0	65.0	30.4	25.7	56	25	2.09	757	642	0.015	13.0	1.99	37
10+000	6	2.70	31.8	68.2	27.8	23.7	51	21	2.17	587	505	0.011	12.3	1.99	31
11+000	10	2.71	41.7	58.3	33.5	29.0	55	24	1.96	814	708	0.014	12.6	1.97	30
12+500	12	2.75	34.2	65.8	26.2	21.8	54	25	2.18	661	556	0.014	10.9	2.06	42
14+800	12	2.74	31.1	68.9	26.4	21.8	53	23	2.21	602	502	0.013	12.4	1.98	33
15+900	7	2.75	34.5	65.5	28.6	23.1	60	29	2.14	813	659	0.020	11.2	2.03	33
18+400	5	2.78	42.9	57.1	34.1	26.4	57	27	1.97	924	719	0.017	10.3	2.10	23
20+350	6	2.79	47.8	52.2	37.8	29.1	62	31	1.85	1169	905	0.023	14.0	1.96	37
23+200	10	2.61	43.7	56.3	34.8	26.2	60	30	1.95	1035	785	0.022	11.7	1.95	31
25+025	12	2.73	30.1	69.9	23.0	17.8	57	26	2.29	610	474	0.017	11.4	2.02	29
26+250	7	2.75	34.8	65.2	26.8	19.8	56	25	2.19	698	522	0.016	11.6	2.04	22
27+700	5	2.81	32.4	67.6	26.4	21.2	60	29	2.20	785	631	0.022	12.8	2.02	31
28+625	6	2.79	34.0	66.0	28.3	22.6	56	28	2.15	782	627	0.019	11.6	2.05	28
31+150	15	2.64	31.3	68.7	24.9	19.6	53	24	2.24	594	471	0.013	11.9	1.96	40
34+475	6	2.77	23.9	76.2	21.7	18.2	54	23	2.36	503	422	0.013	13.7	1.97	30
38+800	7	2.83	27.6	72.4	25.5	23.4	62	30	2.23	776	713	0.023	13.6	2.00	24
39+700	5	2.69	47.2	52.8	40.7	35.4	50	21	1.77	888	783	0.010	13.5	1.93	20

540

Table 2 Values of statistical variables of the studied lateritic gravels

Statistical variables	γs (g/cm^3)	< 2 mm (%)	> 2 mm (%)	< 425 µm (%)	< 75 µm (%)	LL (%)	PI (%)	Gm	Pm	Pp	εs	OMC (%)	MDD (g/cm^3)	CBR (%)
Min	2.40	14.6	0.6	12.3	7.3	39	16	0.54	274	173	0.002	7.7	1.62	12
$\bar{X}$	2.73	35.3	64.7	28.6	23.4	55	25	2.13	732	600	0.016	12.2	2.00	34
Max	3.00	99.4	85.4	84.6	64.0	68	36	2.63	2639	2058	0.034	22.2	2.16	94
τ	0.11	12.2	12.2	10.2	8.7	5	4	0.31	318	259	0.006	2.0	0.09	13
$\bar{X}$±1τ	77	74	74	77	76	67	64	76	79	77	71	81	77	13
$\bar{X}$±2τ	100	98	98	98	97	97	96	98	98	98	99	96	98	96
Cv	4	35	19	36	37	10	16	14	43	43	36	17	5	39

Min = Minimum value. $\bar{X}$= Average. Max = Maximum value. τ = Standard deviation. Cv = Coefficient of variation

Table 3 Eigenvalue, variability, and cumulative percentages of component variability

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13
Eigenvalue	7.68	2.92	1.28	0.94	0.91	0.12	0.05	0.04	0.02	0.02	0.01	0.00	0.00
Variability (%)	54.84	20.87	9.17	6.73	6.52	0.84	0.37	0.29	0.17	0.13	0.07	0.00	0.00
Cumulative %	54.84	75.70	84.87	91.61	98.12	98.96	99.33	99.63	99.80	99.93	100.00	100.00	100.00

545 **Table 4** Squared cosines of the measured variables

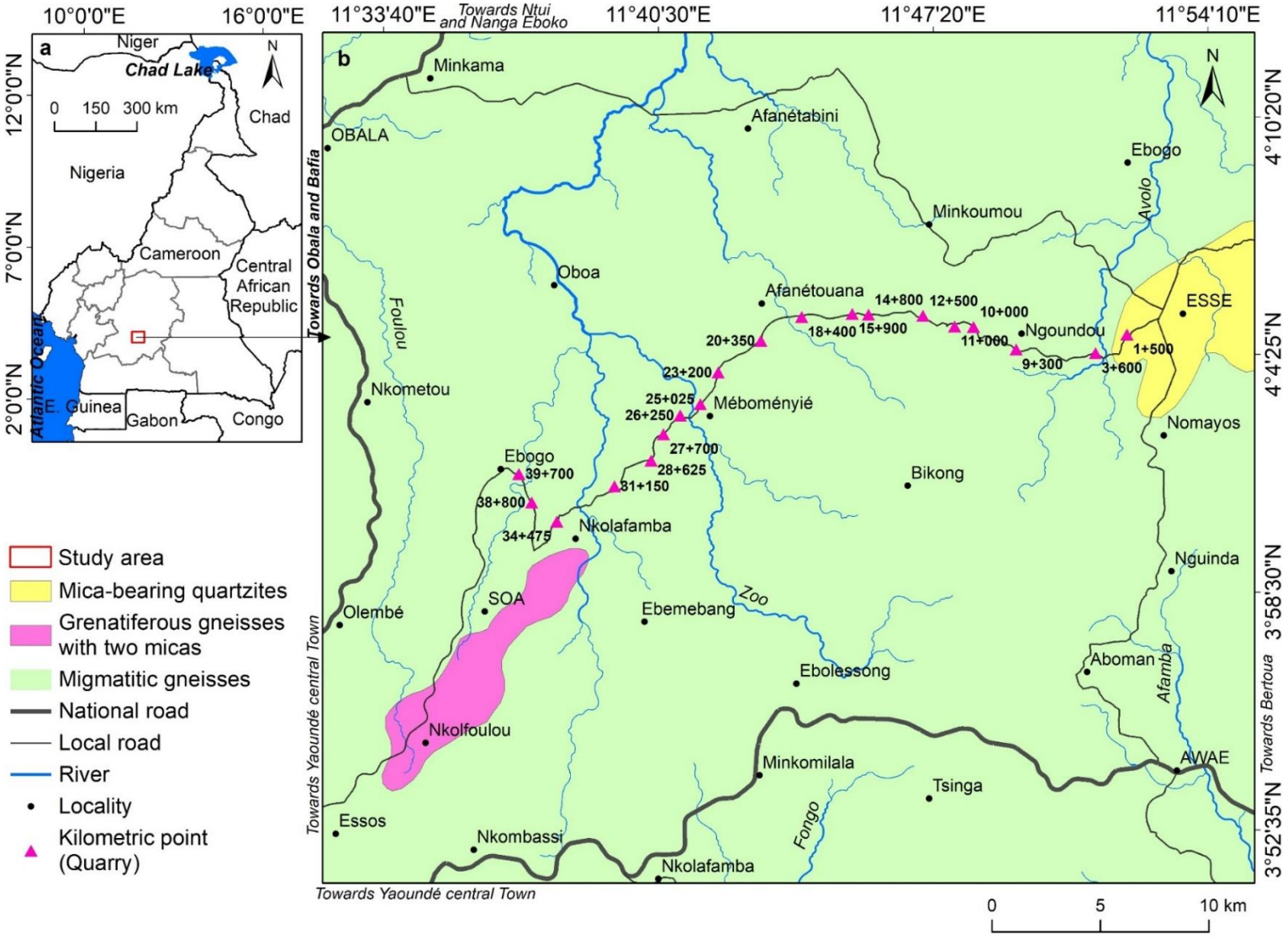
	F1	F2	F3
γs	0.046	0.029	**0.727**
< 2 mm	**0.875**	0.049	0.015
> 2 mm	**0.875**	0.049	0.015
< 425 µm	**0.926**	0.042	0.014
< 75 µm	**0.882**	0.052	0.017
LL	0.121	**0.840**	0.007
PI	0.121	**0.862**	0.002
Gm	**0.923**	0.049	0.016
Pm	**0.937**	0.030	0.008
Pp	**0.940**	0.025	0.011
εs	0.122	**0.844**	0.003
OMC	0.402	0.013	0.006
MDD	0.423	0.037	0.393
CBR	0.085	0.001	0.049

546 *Values in bold correspond for each variable to the factor for which the squared cosine is the largest*

547 **Table 5** Metric results of the machine learning tests conducted in this study

	Particle size distribution parameters				Plasticity parameters				Derived parameters				Compaction parameters			
Model	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2	MSE	RMSE	MAE	R^2
Multiple Linear Regressions	143.8	11.9	9.7	0.10	153.2	12.1	10.2	0.03	141.6	11.9	9.8	0.09	146.1	12.1	10.0	0.07
Ridge Regressions	141	11.8	9.6	0.09	149.9	12.2	10.1	0.05	142.4	11.8	9.9	0.09	146.3	12.1	10.02	0.07
Random Forest Regressions	21.4	4.6	3.8	0.87	87.5	9.8	7.2	0.44	23.2	4.8	3.8	0.85	25.3	5.1	3.9	0.84
Support Vector Regressions	7.9	2.8	0.7	0.94	162.5	12.7	9.9	0.03	8.1	2.8	0.7	0.94	146.8	12.03	10.1	0.08

548 MSE: Mean Squared Error; RMSE: Root Mean Square Error; MAE; Mean Absolute Error; R^2: Coefficient of Determination



549

550 **Fig. 1** Location. (a) The study area in Central Africa. (b) The study area showing the distribution of quarries above the bedrock



551

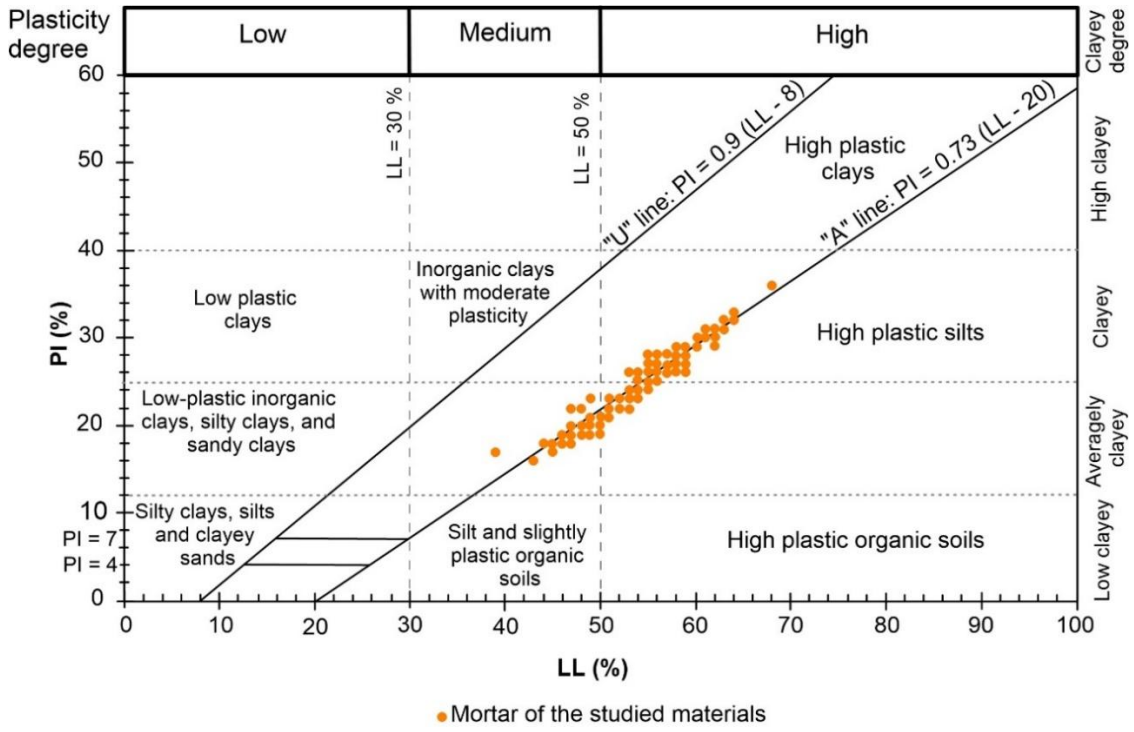
552 **Fig. 2** One of the sites where lateritic gravels were extracted. (a) Panoramic view. (b) Associated profile



553

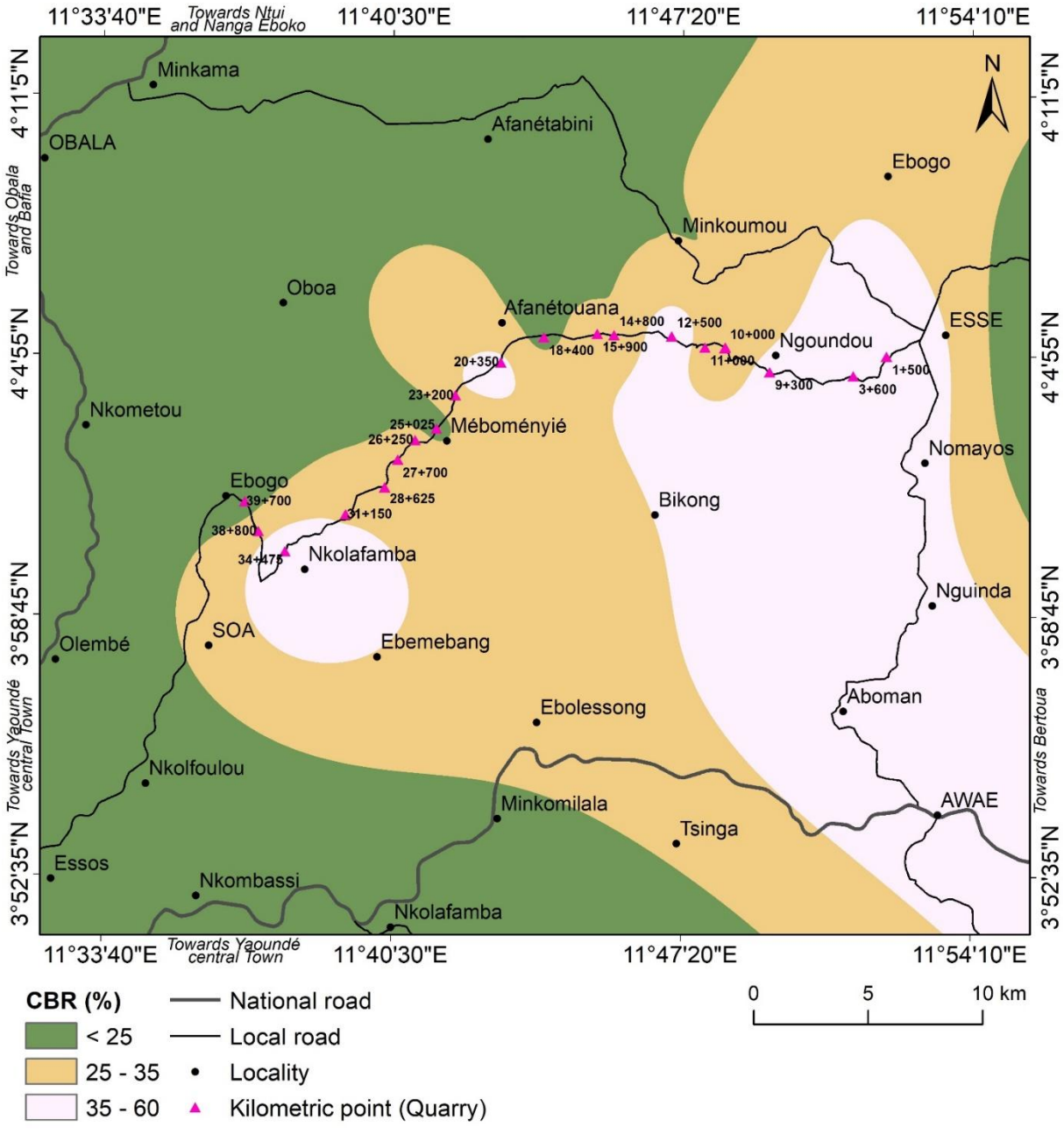
554 **Fig. 3** Representative particle sizes and shapes in the studied lateritic gravels. a) KP1+500. b) KP10+000. c)

555 KP11+000. d) KP18+400. e) KP20+350. f) KP23+200. g) KP27+700. h) KP38+800



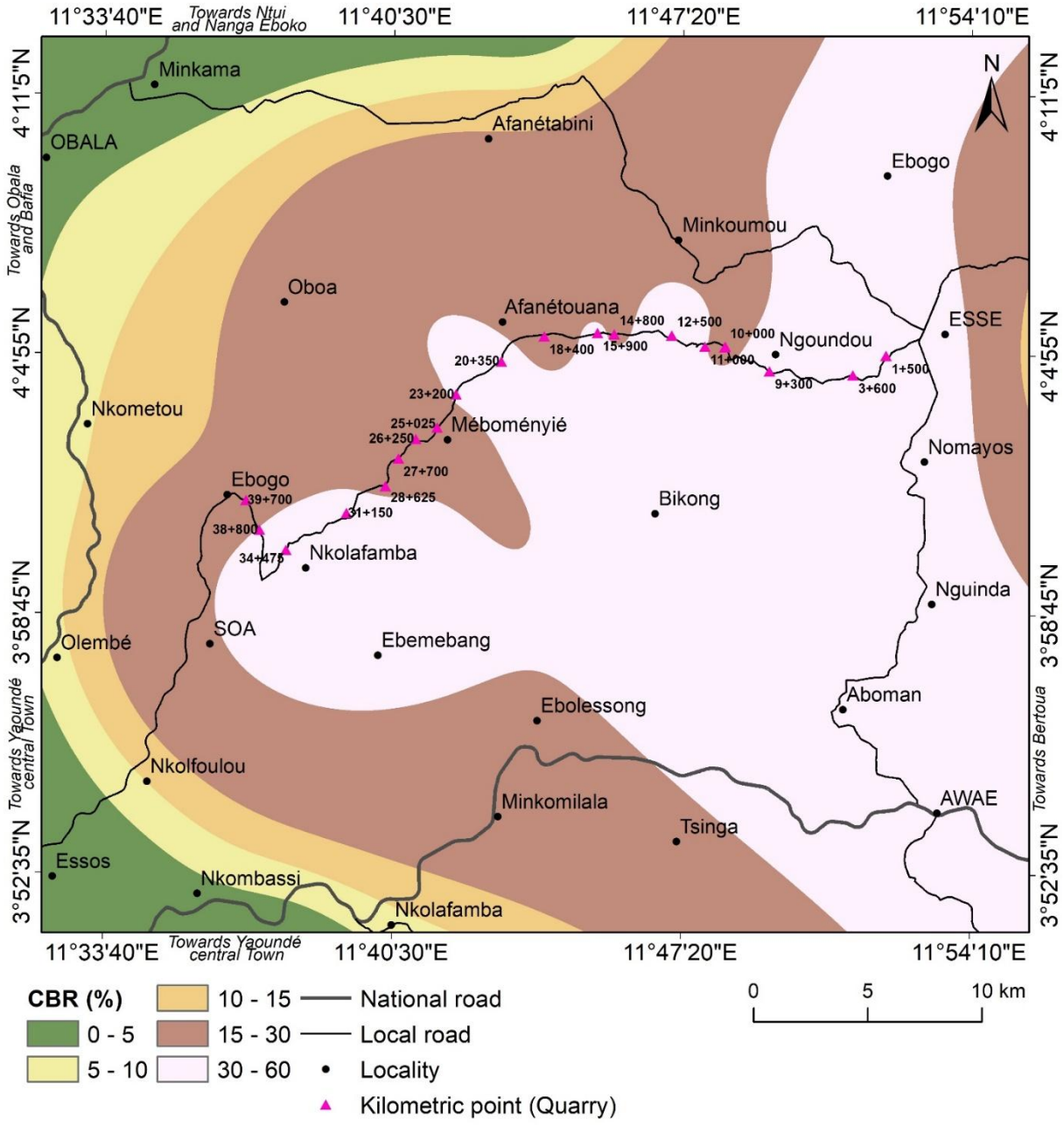
556

557 **Fig. 4** Locations of the studied lateritic gravel mortars in the Casagrande plasticity chart



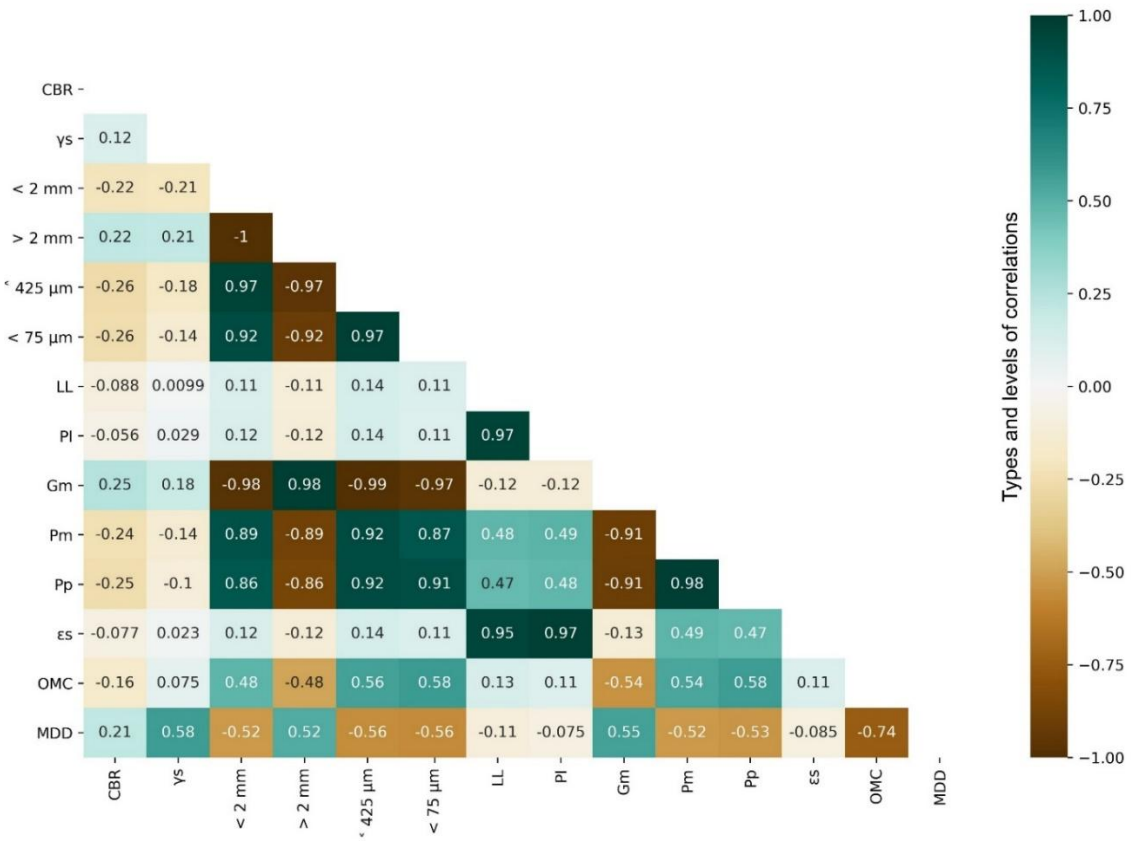
558

559 **Fig. 5** Iso-CBR map of the study area [52]



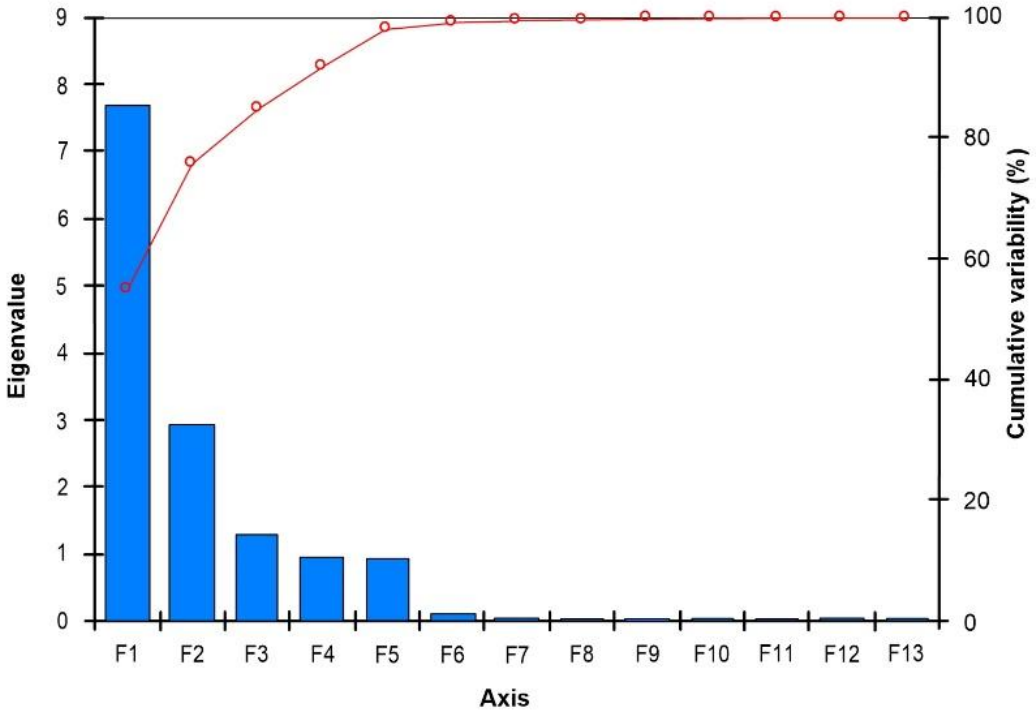
560

561 **Fig. 6** Iso-CBR map of the study area [58]



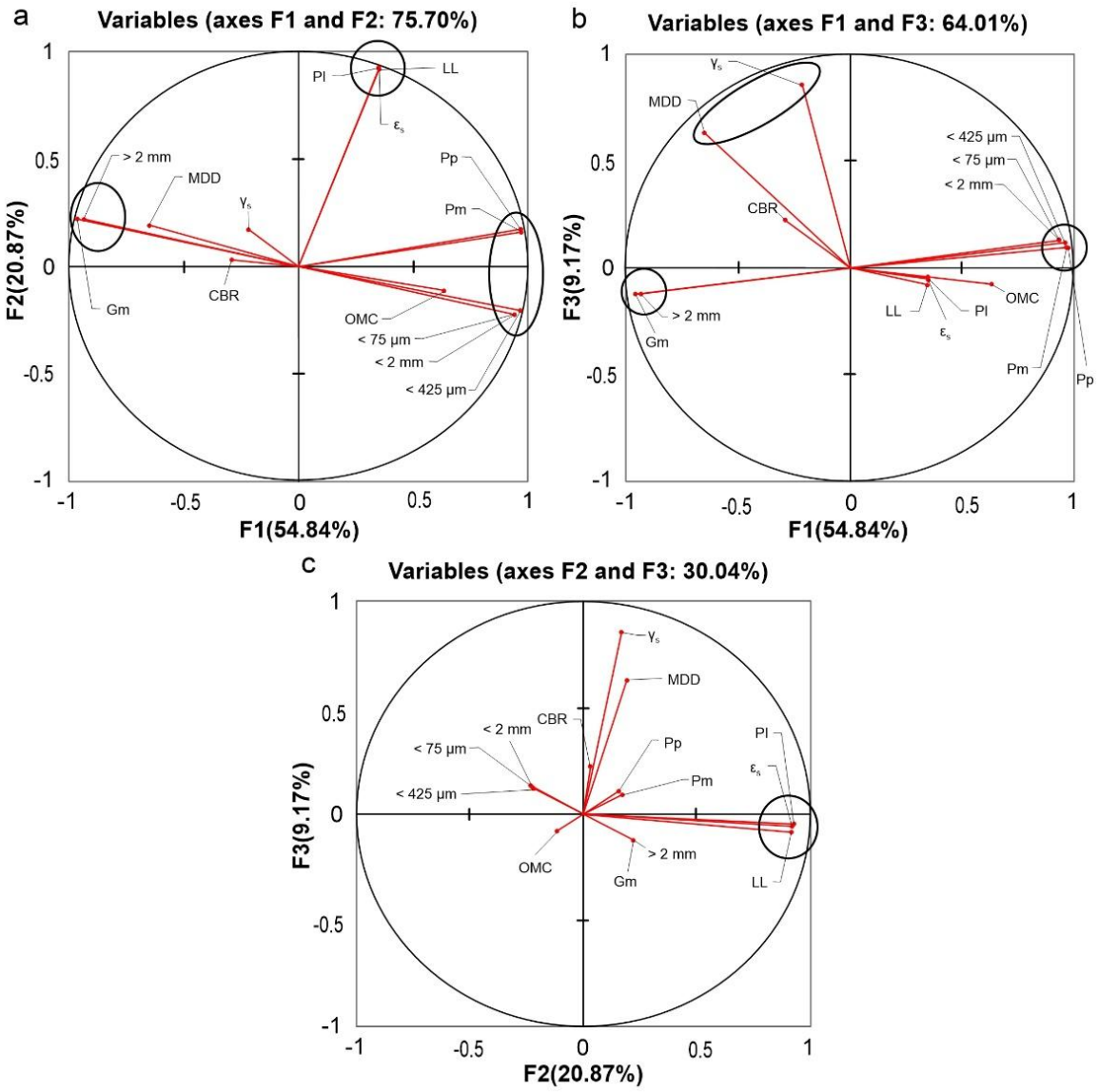
562

563 **Fig. 7** Heatmap showing correlations between geotechnical variables of the studied lateritic gravels



564

565 **Fig. 8** Scree plot of axes and related eigenvalues and cumulated variances



566

567 **Fig. 9** PCA data of geotechnical variables in the studied lateritic gravels

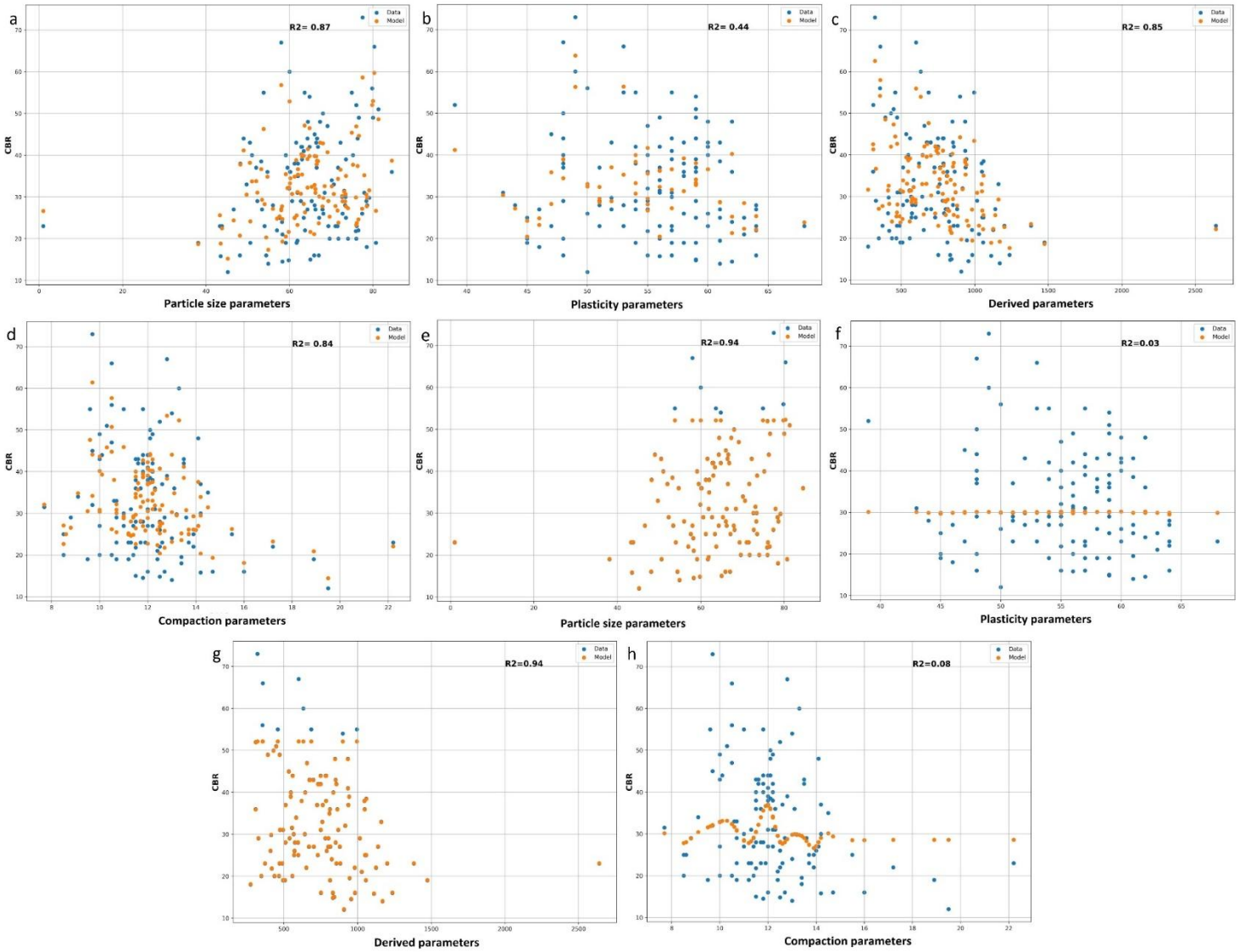
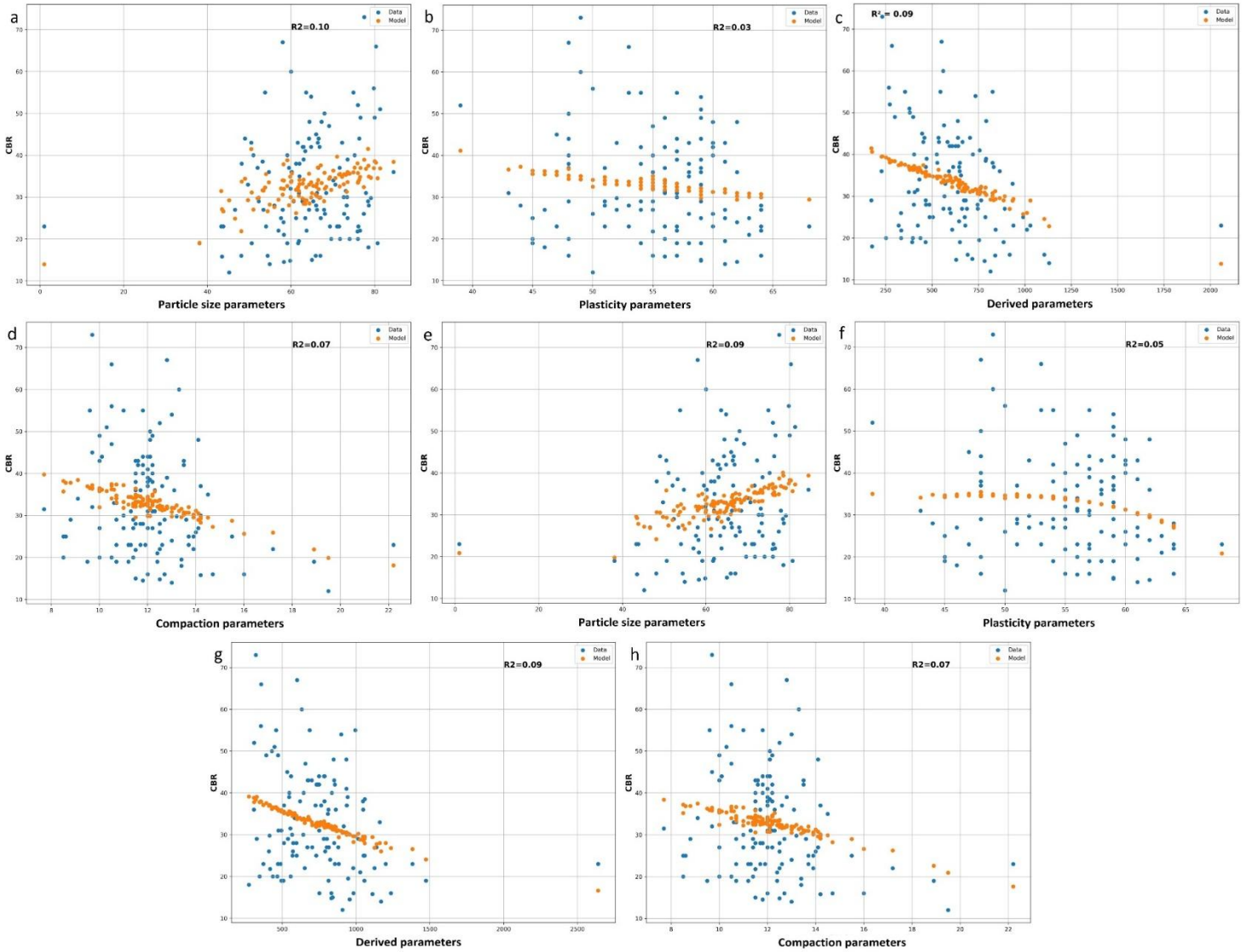


Fig. 10 Relationships between the CBR and the geotechnical groups of variables. a – d) Multiple Linear Regressions. e – h) Ridge regressions



570

571 **Fig. 11** Relationships between the CBR and the geotechnical groups of variables. a – d) Random Forest Regressions. e – h) SVM Regressions

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementarydata.docx