# Estimation of the soil unit weight of mining tailings through the application of machine learning techniques

*Helena Paula* Nierwinski[1#], *Talita* Menegaz[2], *Ricardo José* Pfitscher[1], *Edgar* Odebrecht[3], *Fernando* Schnaid[4] *and Fernando* Mantaras[5]

[1]*Federal University of Santa Catarina, PPGEC, R. Dona Francisca, 8300 – Bloco U Zona Industrial Norte Joinville – SC – Brazil*
[2]*University of São Paulo, IGc, R. do Lago, 562 - Butantã, São Paulo – SP - Brazil*
[3]*University of Santa Catarina State, PPGEC, R. Paulo Malschitzki, 200 - Zona Industrial Norte, Joinville – SC - Brazil*
[4]*Federal University of Rio Grande do Sul, PPGEC, Av. Paulo Gama, 110 - Bairro Farroupilha - Porto Alegre - Rio Grande do Sul*
[5]*Geotechnical Consultant, Geoforma, R. Ten. Antônio João, 2195 - Bom Retiro, Joinville – SC - Brazil*
[#]*Corresponding author: helena.paula@ufsc.br*

## ABSTRACT

There are several correlations in the literature that allow an estimate of the soil unit weight for natural soils, but when dealing with materials whose actual specific gravity of solids is outside the range of natural soils for which the correlations were developed, doubts arise, as occurs in the interpretation of tests on mining tailings. Therefore, the present paper aims to evaluate the application of a previously developed approach supported by machine learning techniques for estimating soil specific weights for mining tailings. This approach was developed considering a more comprehensive range of the specific gravity of solids. So, this work relies on a database with results of CPTu tests carried out in different mining tailings deposits from Brazil to estimate specific weights. The values of the specific weights obtained from the machine learning model were compared with literature data, presenting a suitable fit. The research demonstrates that artificial intelligence can contribute positively to the estimation of reliable design parameters and add security to the development of designs of mining tailings containment structures.

**Keywords:** Soil unit weight; CPTu tests; Multiple linear regression; Artificial neural networks.

## 1. Introduction

An essential geotechnical parameter for the proper interpretation of field tests, understanding the soil behavior, and developing geotechnical designs is the natural unit weight of the soil ($\gamma_t$), also known as Bulk unit weight. The soil natural unit weight is defined as the ratio between the total weight of the soil and the total volume of the soil mass. The Bulk unit weight is distinguished from the dry unit weight by considering the amount of natural moisture in the soil. In designs, for example, the natural unit weight value is necessary for the designer to predict the level of geostatic stresses, which is directly related to the interpretation of the test results and the evaluation of material strength parameters, offering more security for design development. A more accurate methodology for determining the $\gamma_t$ value is characterizing quality undisturbed samples in laboratory tests. In these cases, also some factors influence the reliability of the results. The first, according to Coile (1936) and Stewart (1943), is to guarantee the collection of undisturbed samples, preserving the structure of the material in the field. The others consist of having advanced and up-to-date technology equipment and having trained staff. Investments are necessary to guarantee the development of geotechnical research and a culture of broad geotechnical-geological research; however, in South America, especially in Brazil, these resources are scarce. This deficit makes the accurate measuring of the unit weight hard, forcing most designs to use standard values adopted from the literature for these parameters.

Furthermore, like most mining tailings, cohesionless soils have a particular characteristic that makes it impossible to collect undisturbed samples using traditional methods, requiring technologies that are not widely used in South America, such as freezing techniques.

Field and laboratory tests complement each other; however, field tests have some attractions (Lunne, Robertson and Powell, 1997). The Cone Penetration Test (CPTu) is used to obtain soil parameters, and some correlations allow estimation of the soil unit weight value for general cases. This test is carried out with the support of a set of steel rods and a conical metal tip with electrical sensors installed, which penetrate the soil at a standard constant speed, obtaining the data every 2 cm, for standard tests. CPTu test measures three records concurrently: tip resistance ($q_c$), lateral friction ($f_s$), and generated pore pressure ($u$). Even though the test does not collect samples from the soil, empirical or statistical formulations can be applied to estimate other soil properties based on the measurements, like the natural soil-specific weight (Mayne, 2014). These methods consist of empirical formulations developed from natural soil databases, containing both CPTu records and directly measured values of the soil mechanical parameter under consideration, encompassing a range of specific gravity of solids ($G$) values, usually 2.5 to 2.7 (Mayne, 2007). For instance, organic soils (Lengkeek et al., 2018; Straz and Borowiec, 2020) and mining tailings (Menegaz et al., 2022) have a different range for some properties, which reduces the precision of predictions. Also, considering the extensive applicability of cone tests, encompassing all types of soils, it is counterproductive to develop a specific model for each soil.

Currently, many researchers applied statistical regressions and machine learning to estimate the soil unit weight. Using the characteristics of soils and applying statistical regressions, Mayne (2014) proposes an equation with the variations of $\gamma_t$ as a function of $f_s$, $q_t$, and $m_q$ (cone resistance-depth ratio). The outcoming shows an $R^2$ varying around 0.62, providing an admissible fit. In the same way, Robertson and Cabal (2010) have used dimensionless parameters of resistance ($q_t/\sigma_{atm}$), sleeve friction ($R_f = f_s/\sigma_{atm}$), and the average specific gravity ($G$), proposing two types of equations. The original study did not provide a value of $R^2$ for estimation. However, according to the evaluated soil, Menegaz et al. (2022) show that $R^2$ varies in a range from 0.2 to 0.79. For organic soils, based on laboratory-determined leading parameters using artificial neural networks (ANN), Straz and Borowiec's (2020) paper shows the ANN model can estimate $\gamma_t$ for organic soils with an $R^2$ superior to 0.94. Even though current research presents a good fit to estimate the unit weight in the evaluated scenarios, the accuracy drops when we use a database of heterogeneous soils, as shown by Menegaz et al. (2022).

Recent research from Nierwinski et al. (2023) proposed a practical approach to estimating the soil unit weight using machine learning based on parameters obtained from CPTu tests. The research shows through clustering analysis that the data set does not present any similarities in soil parameters and creates multiple regression models for individual groups. Then, to understand the correlations between parameters and soil unit weight, they performed a statistical analysis, which allowed them to propose an equation based on linear regression. To abstract the hidden relations between CPTu parameters and $\gamma_t$ in multiple soils, they count on an Artificial Neural Network (ANN) using the $q_t$, $f_s$, and $u$ values obtained in CPTu tests and the specific gravity ($G$) from laboratory tests. The $G$ parameter is considered necessary in the analyses because it is an intrinsic characteristic of the soil, capable of distinguishing one soil from another. The $G$ values obtained in the laboratory were correlated with the CPTu data obtained in the interval corresponding to the collection depth of the tested sample. This proposal shows an estimated soil unit weight with an $R^2$ of 0.82. They published an online web application that estimates the soil unit weight to enhance its applicability. In this context, the present study used this application offered by Nierwinski et al. (2023) to test a database encompassing bauxite, iron, zinc, and gold mining tailings and compare the soil unit weight results with the values shown in literature for the same materials.

## 2. State-of-the-art for soil unit weight estimation

Soil unit weight ($\gamma_t$) is the weight of a soil sample per unit volume, expressed in KiloNewton per meter cubic (kN/m³). Several factors can affect this value, for instance, soil type, moisture content, and consolidation. A widespread way to determine the soil unit weight is to collect an undisturbed soil sample and obtain the weight and volume ratio (Coile, 1936; Stewart, 1943). However, it is necessary to guarantee the sample quality for reliable $\gamma_t$ values. For example, as the collection depth increases, it can interfere with the calculated soil unit weight because the soil sample is submitted to a consolidation process.

Aiming for greater practicality and speed in estimating the $\gamma_t$ some authors have developed proposals to estimate the natural specific weight based on the results of field tests, such as the case of CPTu.

Robertson and Cabal (2010), developed approximate contours of unit weight values as a function of dimensionless resistance and lateral friction parameters provided by CPTu tests (Eq. 1). The authors point out that Equation 1 is applied to the vast majority of soils with a specific gravity of solids ($G$) in the range of 2.5 to 2.7. For soils with $G$ outside this range, some interference may occur in the proposed correlation, so they proposed Eq. 2, where the value of $G$ is introduced.

$$\gamma_t/\gamma_w = 0.27\left[\log R_f\right] + 0.36\left[\log(q_t\sigma_{atm})\right] + 1.236 \quad (1)$$

where $q_t$ = corrected tip resistance and $\sigma_{atm}$ = atmospheric pressure (kPa).

$$\gamma_t/\gamma_w = \left[0.27\left[\log R_f\right] + 0.36\left[\log(q_t\sigma_{atm})\right] + 1.236\right] G/2.65 \quad (2)$$

Mayne and Peuchen (2012) proposed a first evaluation relating the unit weight of the materials with the soil plasticity index, verifying a tendency to reduce the specific weight with the plasticity increase. However, as the plasticity index is not obtained using the CPTu test, the authors investigated and identified a relationship between the plasticity index and the ratio between the tip resistance and the depth, denominated $\mathrm{m_q}$ ($\mathrm{m_q} = q_t/z$). Analyses were performed, and two equations were obtained (Eq. 3 and Eq. 4).

$$\gamma_t = \gamma_w + \mathrm{m_q}/8 \quad (3)$$

$$\gamma_t = 0.636\,(q_t)^{0.072}(10 + \mathrm{m_q}/8) \quad (4)$$

Mayne (2014) proposes to use another parameter obtained through CPTu, relying only on the relationship of soil unit weight with the lateral cone friction ($f_s$), a.k.a. sleeve friction, measurements. The two proposals are presented in Eq. 5 and Eq. 6.

$$\gamma_t = 26 - \frac{14}{1 + [0.5 log(f_s + 1)]^2} \quad (5)$$

$$\gamma_t = 12 + 1.5\ln(f_s + 1) \quad (6)$$

Even though such empirical approaches may result in precise estimations for specific scenarios, some soils cannot adopt these formulations in general cases due to their specific characteristics. For example, organic soils (Lengkeek et al., 2018; Straz and Borowiec, 2020) and mining tailings (Menegaz et al., 2022) have non-usual properties, that reduce the accuracy of predictions. In a recent study, Menegaz et al. (2022) evaluated the empirical equations proposed in the literature (Robertson & Cabal, 2010; Mayne & Peuchen, 2013; Mayne, 2014) using a database composed of mining tailings soil parameters. This analysis uses the Pearson correlation coefficient (*r*-value) to evaluate whether the measured and estimated specific weights are related to each other. Results showed a good fit (*r*-value superior to 0.5) for estimating $\gamma_t$ for zinc mining tailings and inaccuracy (*r*-value less than 0.5) for bauxite mining tailings.

However, machine learning was applied by some researches to estimate soil parameters. Fang et. al (2023), used an artificial neural network (ANN) with 1069 shear-wave velocity ($V_s$) measurements to estimate the liquefaction phenomenon potential of soil. Entezari et al. (2022) tested a Random Forest model to estimate the shear-wave velocity ($V_s$), using a dataset with 104,809 CPTu data which resulted in a $R^2$ of 0.58 for the general soils test set. Concerning the soil unit weight estimation, Straz and Borowiec (2020), claim that specific soils, for example, organic soils, demand the design of a specific model. The authors used a Rzeszow Site in Poland, with disturbed and undisturbed samples of structures of organic soils, to perform regressions and train an ANN. With the regression-based estimation of $\gamma_t$ in the function of the organic content ($LOI_T$) and water content ($w$), Eq.7, achieved an $R^2 = 0.94$, the ANN model achieved an $R^2 = 0.98$.

$$\gamma_t = 17.4603 + 0.0407\mathrm{LOI}_T - 0.0307w \quad (7)$$

Even though promising results, this proposal depends on laboratory tests to extract the soils organic and water content, which preclude models' application in scenarios that depend exclusively on in situ test results.

Considering the limitations of regression-based models and the promising results presented in the estimations based on machine learning, Nierwinski et al. (2023) proposed an approach supported by machine learning based on regression model and artificial neural network model that use parameters extracted from CPTu tests taken from a large heterogeneous dataset of soils, including mining tailings. We intend to test the model proposed by the authors in a new mining tailings database and verify the answers found about the specific weight values expected for the material listed in the literature.

## 3. Materials and Methods

This section presents the database used to test the model proposed by Nierwinski et al. (2023) and details of the tested model are presented.

### 3.1. Dataset

This paper uses data obtained in experimental investigation campaigns realized by different companies from Brazil. The dataset contains 294 entries with geotechnical data from 4 distinct tailings, encompassing bauxite, iron, zinc, and gold mining tailing. Those campaigns provide tip resistance ($q_t$), sleeve friction ($f_s$), and soil pore pressure ($u_2$) for all materials present in the database. The soil-specific gravity value was present in 75% of the test reports, and for the 25% remaining, values were used as indicated in the literature. In the following items, we explain the database content.

- Iron - the dataset contains 91 entries of iron mining tailings (30.95%) collected in experimental campaigns realized by a private company in Brazil;
- Bauxite - the bauxite mining tailing data make 159 entries in the dataset (54.08%). Its composition is entirely from investigation campaigns conducted by a private company at sites in the north and northeast of Brazil;
- Zinc - the dataset contains 35 samples of zinc mining tailing (11.9% of the total). Such data is from a field campaign realized by a private company in a dam in Minas Gerais, southeast Brazil;
- Gold - data are from a field trial realized by a private company in Brazil and contain nine entries (3.06%).

### 3.2. Model description

The model used for soil unit weight estimation through CPTu data was proposed by Nierwinski et al. (2023). In this paper, the authors proposed two models to estimate the soil unit weight: linear regression and artificial neural networks (ANNs). To train the models, they rely on a dataset containing 1862 entries with geotechnical data from 10 distinct soil types. The linear regression model uses a logarithmic transformation in two variables, $q_t$ and $u$, and results in a good fit, with an $R^2$ of 0.62 in the trained dataset. The regression model is defined by Equation 8:

$$\gamma_t = -1.1795 + 3.33G + 2.90 \log_{10} q_t + 0.21 \log_{10} u_2 \quad (8)$$

On the other hand, using five hidden layers, the ANN model abstracts the nuances that regression models do not capture in the dataset; as a result, it obtained an R2 of 0.82. The ANN model relies on a resilient backpropagation algorithm with weight backtracking. Also, each neuron processes a logistic function to compute the weights of each input. More implementation details are available on the code published on GitHub ([https://github.com/ricardopfitscher/Sweet](https://github.com/ricardopfitscher/Sweet)), and the model is available for free access at the SWEET—Soil unit WEight Estimator (http://shinyapps.io/) link.

The database described in item 3.1 will be used to estimate the specific weights of mining tailings using regression and ANN models. The specific weight values provided by the model will be analyzed and compared with reference data from the literature.

## 4. Results and discussion

Figures 1 and 2 demonstrate the distribution of specific weight values ($\gamma_t$) provided for the regression and ANN models, respectively.

It can be seen that the distribution of the specific weight values is different for the regression model and ANN model. Table 1 presents the estimated mean and standard deviation values for each evaluated mining tailing. The values estimated using the regression models are bigger than those estimated using the ANN model. The standard deviation is also more significant when using the regression model. This behaviour can be observed in Fig. 1 and Fig. 2.

The mean specific weight values estimated using the regression model are outside the expected range for the mining tailing studied. Obtaining these values possibly comes from the large dispersion observed in the regression model.

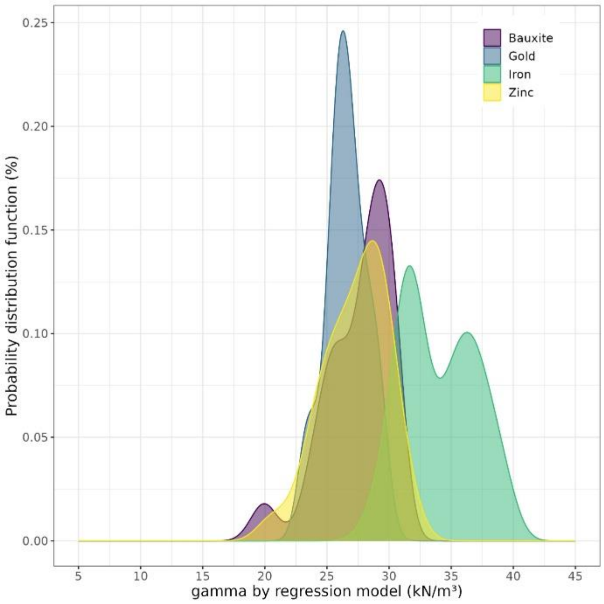


**Figure 1.** Specific weight values (gamma) distribution of mining tailings dataset estimated by regression model.

The ANN model tends to provide less dispersed specific weight values. The ANN model is likely to capture significant physical properties for model development. It can be observed, for example, that the ANN model identifies an increase in soil-specific weight with increasing $q_t$ and $G$. The iron mining tailing was the only material that showed greater dispersion in the estimated values . However, the standard deviation of the ANN model is still smaller than the standard deviation obtained by the regression model.

According to the literature, the values for the studied mining tailings present variation, as shown in Table 2. For the data used in this research, specific weight measurements obtained in the laboratory were not available in a specific manner, making it possible to present a comparison only with the expected range of variation.

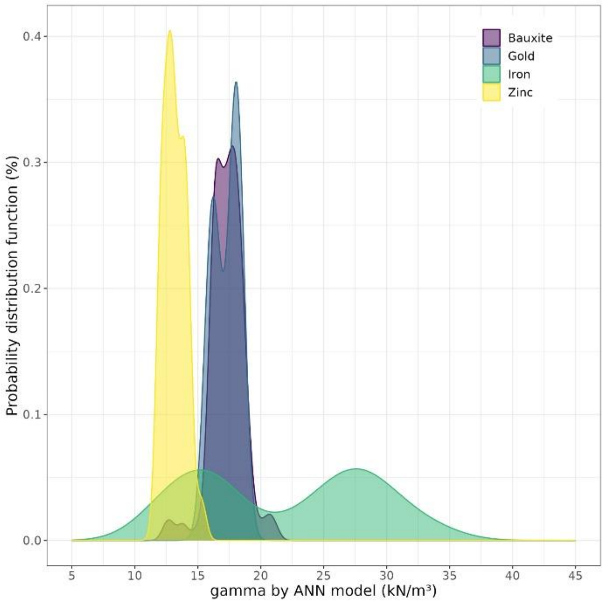


**Figure 2.** Specific weight values (gamma) distribution of mining tailings dataset estimated by ANN model.

**Table 1.** Estimated specific weight values (in kN/m³) statistical parameters

| | **Regression Model** | | **Ann Model** | |
|---|---|---|---|---|
| Mining Tailing | **Mean** | **Standard Deviation** | **Mean** | **Standard Deviation** |
| **Bauxite** | 27.49 | 2.68 | 17.31 | 1.30 |
| **Gold** | 26.55 | 1.59 | 17.24 | 0.95 |
| **Iron** | 34.95 | 9.34 | 21.92 | 6.84 |
| **Zinc** | 27.29 | 2.56 | 13.14 | 0.83 |

**Table 2.** Specific weight values (in kN/m³) from literature

| Mining Tailing | **Specific weight variation (kN/m³)** | **Reference** |
|---|---|---|
| **Bauxite** | 17.3 to 18.4 | Nierwinski *et al.*, 2020 |
| **Gold** | 18.6 to 20.5 | Bedin, 2010 |
| **Iron** | 21.72 to 23.35 | Morgenstern *et al.*, 2016 |
| **Zinc** | 11.27 to 14.92 | Hlenka, 2012 |

Comparing the specific weight values estimated by the models with parameter values provided in the literature, it is observed that the ANN model provided promising estimates, with values very close to those in the literature. A greater variation was observed for iron mining tailing; the model would require calibration for more accurate use for this specific tailing.

## 5. Conclusions

This work sought to test models for estimating the unit weight of soil from a database of mining tailings from Brazil, using regression and ANN models. Through this study, it was possible to define which method best suited the database tested.

The regression model indicated that in most of the materials tested, a value was found above the values found in the literature for the same type of soil. On the other hand, the ANN model showed values very close to those in the literature, making it a good fit for predicting mining tailing specific weight.

A more detailed evaluation and calibration is recommended to use the model for specific weight estimates in iron mining tailings.


## Acknowledgements

The authors would like to acknowledge Brazilian Companys for gently providing the reports to develop our database.

CRediT author statement

Helena Paula Nierwinski: Conceptualization, Methodology, Formal analysis, Writing - Review & Editing, Validation. Ricardo José Pfitscher: Software, Writing - Original Draft. Talita Menegaz: Data Curation, Formal analysis, Writing - Review & Editing. Edgar Odebrecht: Investigation, Resources, Writing - Review & Editing. Fernando Schnaid: Investigation, Resources, Writing – Review & Editing. Fernando Mantaras: Investigation, Resources, Writing – Review & Editing.


Data availability

Part of the data is confidential, but we can provide access upon request by email.

Code availability section

Name of the code/library: SWEET - Soil unit WEight Estimator

Contact: ricardo.pfitscher@ufsc.br

Program language: R

The source codes are available for download at the link under the MIT License: https://github.com/ricardopfitscher/Sweet.

## References


Bedin, J.”Estudo do Comportamento Geomecânico de Resíduos de Mineração.” Tese de doutorado em Engenharia. Programa de Pós-Graduação em Engenharia Civil, UFRGS, Porto Alegre, 2010.

Coile T, Soil samplers, Soil Science 42 (2) (1936) 139–142.

Entezari I, Sharp J, Mayne P, A data-driven approach to predict shear wave velocity from CPTu measurements, in: Cone Penetration Testing 2022, CRC Press, 374–380, 2022.

Hlenka, L. Estudo dos efeitos da velocidade de carregamento na estimativa de parâmetros geotécnicos em resíduos de mineração de zinco. Dissertação de mestrado apresentada ao Programa de Pós-Graduação da UFRGS, Porto Alegre, 2012.

Lengkeek, H. J., De Greef, J., Joosten, S. (ed) 2018. CPT based unit weight estimation extended to soft organic soils and peat. Cone Penetration Testing 2018: 389 – 394.

Lunne, T., Robertson P. K., Powell, J.J.M. (ed) 1997. Cone Penetration Testing. London: Blackie Academic & Professional.

Mayne, P.W. (ed) 2014. Interpretation of geotechnical parameters from seismic piezocone tests. Proceedings, 3rd International Symposium on Cone Penetration Testing: 47-73

Menegaz T., Odebrecht E., Nierwinski H. P., Schnaid F., Soil unit weight prediction from CPTs for soils and mining tailings, in: Cone Penetration Testing 2022, CRC Press, 566–569, 2022.

Nierwinski, H.P., Reus, A., Heidemann, M., Odebrecht, E., Barra, B. S., Hastenpflug, D., & Pérez, Y. A. G. Caracterização Física de Amostras de Rejeito de Mineração de Bauxita e Ouro. Brazilian Applied Science Review, v. 4, n. 6, p. 3530-3544, 2020.

Nierwinski, H. P., Pfitscher, R. J., Barra, B. S., Menegaz, T., & Odebrecht, E.  A practical approach for soil unit weight estimation using artificial neural networks. Journal of South American Earth Sciences, 131, 104648, 2023.

Robertson, P. K., Cabal, K.L. (ed) 2010. Estimating soil unit weight from CPT. 2nd International Symposium on Cone Penetration Testing: 2 – 40.

Straz, G., Borowiec, A. Estimating the Unit Weight of Local Organic Soils from Laboratory Tests Using Artificial Neural Networks, Applied Sciences 10 (7) (2020) 2261.

Stewart, A. A soil tube for obtaining wet clay cores in an undisturbed structural condition, Soil Science 55 (3) (1943) 247–252.