

Estimating soil-water characteristic curve based on soil type and best-fitting regressions derived from a simplified method using Aburra Valley dataset

Johnatan Ramos-Rivera*, Daniel Parra-Holguín, Yamile Valencia-González, and Oscar Echeverri-Ramírez

National University of Colombia, School of mines, Geotechnical Research Group, Medellín

Abstract. In unsaturated soil mechanics, many attempts have been made to estimate the SWCC based on soil texture and grain-size distribution. This paper proposes a simplified method to estimate the soil-water characteristic curve (SWCC) for both coarse and fine-grained soils using SWCC data and machine learning computer code in the Aburra Valley. Fredlund and Xing parameters has been used to estimate the SWCC correlations. Soil samples collected from field survey were subjected to laboratory testing, SWCCs were estimated using filter paper method. Each SWCC data set from Aburra Valley was fitted with Fredlund and Xing curve using multiple regression analysis, correlations were derived for those four parameters based on predictors derived from machine learning. The proposed method gives a good estimation and low residual errors of the SWCC.

1 Introduction

The Soil-water characteristic curve (SWCC) has been developed as a key tool in agricultural water management, engineering, hydrology, and soil science research. SWCC is considered as a relationship between the water content (or degree of saturation) in soil and water potential also called soil suction (Williams et al., 1983 [1]).

The SWCC has been widely used for the estimation of unsaturated soil properties (i. e shear strength, coefficient of unsaturated permeability, and volumetric water content and its relative volume change (Fredlund and Rahardjo 1993 [2]).

SWCC is generally obtained by laboratory tests. However, measuring these characteristics is time-consuming, labour intensive and therefore expensive. Attempts have been developed to get the SWCC based on soil index properties, such as void ratio, initial gravimetric water content, clay content, silt content, organic matter, density, and salinity. Those approaches based on index properties are highly desirable due to its simplicity and low cost.

Approaches can be divided in (4) main strategies. The first of these strategies is based upon statistical estimation of water contents at selected matric suction values. These water contents, at each suction value, are correlated to soil properties.

The second approach includes those methods that correlate, by regression analysis, soil properties with the fitting parameters of an analytical equation that represents the SWCC. The third approach includes the methods that estimate the SWCC using a physics based conceptual model. It involves physical models based upon the conversion of the PSD (Particle Size Distribution) into a

pore-size distribution. Artificial intelligence methods, such as neural networks, genetic programming, and other machine learning methods (Pachepsky et al. 1996 [3]; Koekkoek and Booltink 1999 [4]; Johari et al. 2006 [5]).

Matric suction in residual soils play an important role on the stability of residual soil slopes against rainfall. Aburra Valley has a tropical rainforest climate with a bimodal annual cycle seasons, uniform temperature of 22°C mean, an average high humidity about 80%, and annual rainfall close 2500 mm/yr. Due to the Climate conditions and the susceptibility to triggered landslides in the Aburra Valley.

In this study, the proposed simplified method has been validated considering experimental data and statistical simulations, using machine learning algorithm to get a simplified a best fitted curve. It concludes that the KNN technique is simple, easy to apply, versatile and accurate capturing the bimodal behaviour. Additionally, the equations have a rational basis which can readily be extended to different soils.

2 Equation for the SWCC

For the past five decades, numerous closed-form and empirical equations have been proposed to describe the SWCC. Leong and Rahardjo 1997 [6] reviewed the popular soil-water characteristic curve equations and showed that they can be derived from the following generic equation:

$$a_1\theta^{b_1} + a_2 \exp(a_3\theta^{b_1}) = a_4\psi^{b_2} + a_5 \exp(a_6\psi^{b_2}) + a_7 \quad (1)$$

* Corresponding author: jramosr@unal.edu.co

where $a_1, a_2, a_3, a_4, a_5, a_6, a_7, b_1,$ and b_2 are constants; ψ is matric suction; and Θ is normalized volumetric water content (i.e., $(\theta_w - \theta_r)/(\theta_s - \theta_r)$), where θ_w is volumetric water content, θ_r is residual volumetric water content, and θ_s is saturated volumetric water content.

In this study Fredlund and Xing 1994 [7] equation has been used regards to its capability for heterogeneous and bimodal soil based on the Van Genuchten and Brooks-Corey models in which the interconnected soil pores and pore-size distribution were introduced. The Fredlund and Xing (1994) equation shall be used in this paper and is given as:

$$\theta_w = c(h)x \left[\frac{\theta_s}{\ln \left[\exp(1) + \left(\frac{h}{a} \right)^b \right]^c} \right] \quad (2)$$

$$C(h) = \left[1 - \ln \frac{\left(1 + \frac{\psi}{\psi_r} \right)}{\ln \left(1 + \frac{10^6}{\psi_r} \right)} \right] \quad (3)$$

where e is the natural number (i.e., 2.718281); $a, n,$ and m are curve-fitting parameters; Ψ is matric suction (kPa); ψ_r is matric suction corresponding to the residual volumetric water content, θ is water content ($\text{cm}^3 \text{cm}^{-3}$), $s,$ and r are the saturated water content and the residual water content ($\text{cm}^3 \text{cm}^{-3}$), respectively.

3 Materials and methods

3.1 Local geology

Colombia is a country in the north of South America, bounded on the north by the Caribbean Sea, the northwest by Panama, the west by the Pacific Ocean, the south by Ecuador and Peru, the southeast by Brazil, and the east by Venezuela. Colombia has the Andes mountain range that extends from south to north of the country. Geodynamical context of Colombia is strongly influenced by the Nazca and Caribbean plates subduct, under the South American within most of the emerged territory of Colombia covers, and North Andean Block, which is colliding with the Panama-Choco Block.

The Aburra Valley, composed of 10 localities, included Medellín, is an elongated depression located at the Central Cordillera of the Colombian Andes, with an approximate area of 1154 km². AV narrows at both north and south and gets wide at the centre in Medellín with 7km approximately, divided by the Aburra river AV serves as basin with a concentrated urbanization that has increasing the complexity of the hydrological process by the urban heat island and the changing vegetation coverage, parameters that will are not further discuss in this work.

AV base stands at 1450m above mean sea level, and it is composed of alluvial material and about 1 km higher the highlands are mainly escarpments. The valley has a wide range of lithologies, ranging from Paleozoic rocks to quaternary deposits, and the structure among those and

the subsurface configuration have been studied by the authors Henao & Monsalve, 2017.

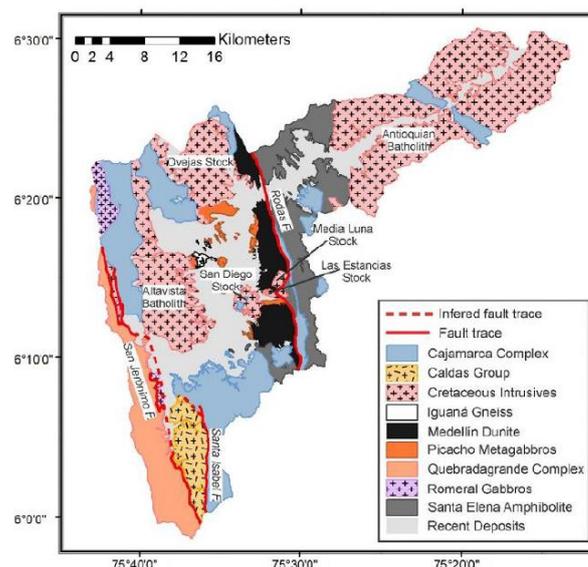


Fig. 1. Local geology of the Aburra Valley. (Henao Casas and Monsalve 2018) [8].

Fig. 1 shows in terms of origin and composition, there are metamorphic rocks, such as schists, amphibolites, migmatites, gneisses; igneous rocks, such as granodiorites, dunites, gabros and basalts; volcanic sedimentary rocks; alluvial deposits and landslide deposits, in addition to manmade fills.

Climate change has shown to change the variability more than the mean values of the atmospheric processes which produces extreme values of different variables. Those extremely weather conditions, specifically on precipitation, triggered various hazard slope instabilities, which results in economic and human losses.

3.2 Database

Soil properties that are known to impact the moisture retention properties of the soil were extracted from the main database of the Geotechnical Research group of National University of Colombia-Medellín. The database involves a range of soils composed of (24) experiments in Aburra Valley. S1 to S17 represents many of the neighbourhoods in the Aburra Valley and it is summarized in Table 1.

Table 1. Soil physical indicators

ID	LL	IP	Sieve #200	USC S	Gs	e
Site 1/M1	37.0	16.0	63.0	CL	2.78	0.96
Site 2 / M1	85.0	39.0	25.0	SM	2.53	1.20
Site 3 / M1	53.0	8.0	92.0	MH	2.79	1.54
Site 4 / S1	85.0	39.0	25.0	SM	2.53	1.20
Site 4 / S2	72.0	22.0	98.0	MH	2.89	1.89
Site 5 / S1	56.0	7.0	95.0	MH	2.83	1.01
Site 5 / S2	54.0	7.0	92.0	MH	2.87	1.72
Site 5 / S3	57.0	8.0	98.0	MH	2.77	2.01
Site 6 / S1	64.0	12.0	95.0	MH	3.06	2.01
Site 7 / S1	57.9	24.5	57.0	MH	2.80	1.10

ID	LL	IP	Sieve #200	USC S	Gs	e
Site 8 / S1	73.0	31.0	88.0	MH	2.89	1.30
Site 9 / S1	62.0	6.0	69.0	MH	2.56	1.16
Site 9 / S2	55.0	7.0	72.0	MH	2.56	1.50
Site 10 / S1	58.0	13.0	97.0	MH	2.66	1.45
Site 10 / S2	36.0	3.0	97.0	ML	2.75	1.07
Site 11 / S1	38.0	2.0	43.0	SM	2.80	1.40
Site 12 / S1	38.0	1.0	50.0	SM	2.75	1.02
Site 13 / S1	54.0	18.0	57.0	MH	2.80	1.00
Site 14 / S1	60.4	15.2	88.0	MH	2.66	1.39
Site 14 / S2	44.2	3.1	70.0	ML	2.63	1.16
Site 15 / S1	82.5	23.7	98.1	MH	2.56	2.04
Site 15 / S1	53.2	11.8	71.9	MH	2.55	1.70
Site 16 / S1	38.8	6.7	68.1	ML	2.57	1.17
Site 17 / S1	46.0	12.0	60.0	ML	2.69	1.22

3.4 Approach to develop a simplified method to estimate the SWCC

The Nelder–Mead method or also called “downhill simplex method” is a numerical direct search method used to find the minimum or maximum of an objective function in a multidimensional space [9]. It consists of multiple steps, which included the math processes of sorting, reflection, expansion, outside contraction, inside contraction, and shrink, which deep explanation can be found on Nelder and Mead, 1995. Although lacking a satisfactory convergence theory, the Nelder-Mead method generally performs well for solving small dimensional real-life problems, which is the case of the Fredlund and Xing equation.

For two variables, a simplex is a triangle, and the method is a pattern search that compares function values at the three vertices of a triangle. The worst vertex, where $f(x, y)$ is largest, is rejected and replaced with a new vertex. A new triangle is formed, and the search is continued. The process generates a sequence of triangles

(which might have different shapes), for which the function values at the vertices get smaller and smaller. The size of the triangles is reduced, and the coordinates of the minimum point are found. The algorithm is stated using the term simplex (a generalized triangle in N dimensions) and will find the minimum of a function of N variables. It is effective and computationally compact.

4 Results

4.1 Statistical Analysis

Before looking for a relationship between the soil index properties and the fitted parameters of the model, the experimental SWCC data are analyzed. There are four (4) different USCS soil classifications in the database (MH, ML, SM, CL). For the CL classification there is one sample, so the analysis is presented with MH, ML and SM. The IP index, percentage pass sieve #200 (P200) and LL parameters show a clear grouping of the ratings (see Fig. 2), with noise characteristic of any measurement process.

The mean of the squared errors, root mean squared error, RMSE, and coefficient of determination, R². For all the criteria, the objective was to minimize estimation errors for the experimental data sets at the population level, which is quantified by RMSE. reports the systematic errors between the measurements and estimated values. RMSE provides the accuracy of the estimation, in terms of standard deviation. The correlation between the measured and estimated SWCC is evaluated by R².

It is worth mentioning the USCS is performed with the soil granulometry and the Atterberg limits, so the quality of the laboratory results is checked. On the other hand, the analysis was useful to find relationships in the properties that are not used in the USCS, the void relationship shows a grouping of ML in low ranges and MH predominantly in high ranges.

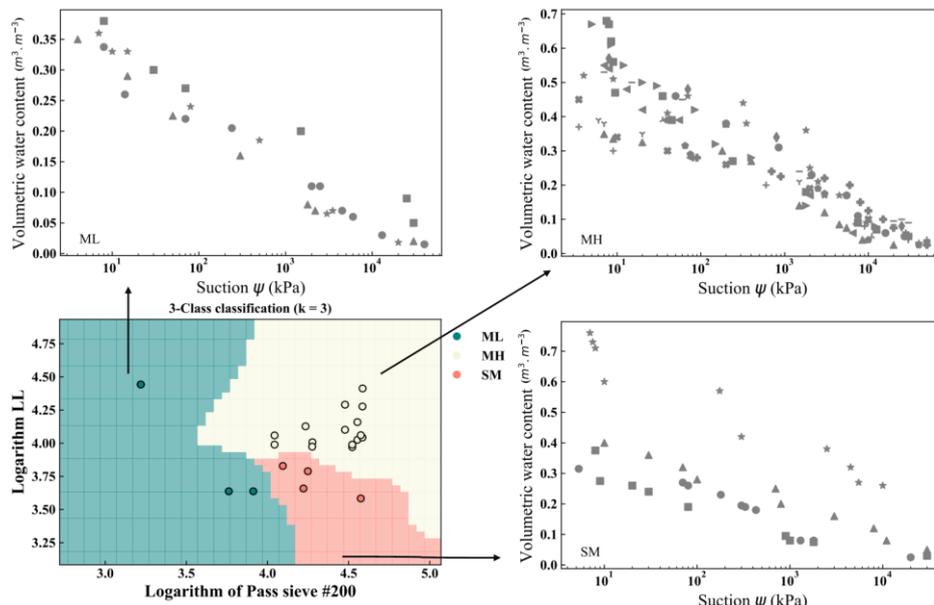


Fig. 2. K-nearest neighbour (KNN) statistical analysis using liquid limit and passing #200

4.2 Soil-Water Characteristic Curve (SWCC)

According to USCS, it is found that the bimodal characteristic of the SWCC is common in the four (4) ML samples, therefore a fit with the Fredlund and Xing equation is not appropriate. The behaviour for MH and SM soils is not clear, which, it is appreciated, have a high variation in the initial water content in the lower suction range, but in the high suction range the curves seem to agree.

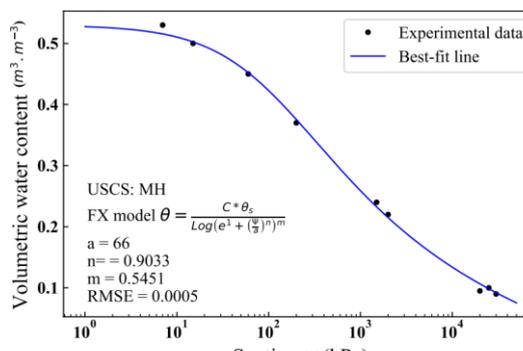
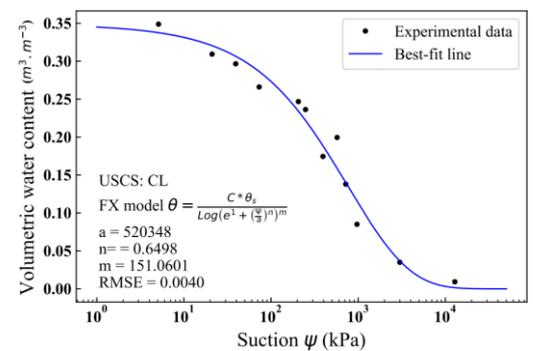
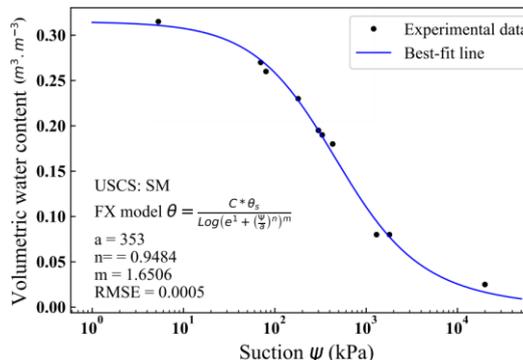
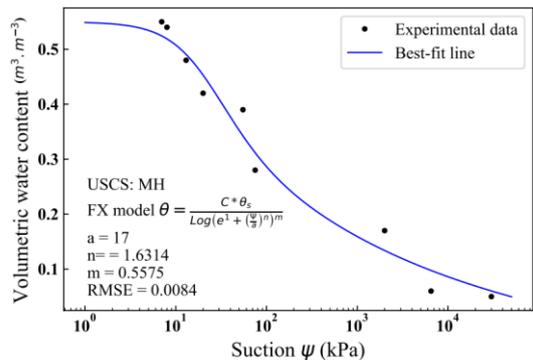
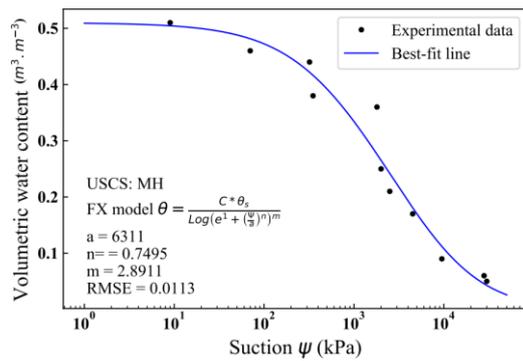
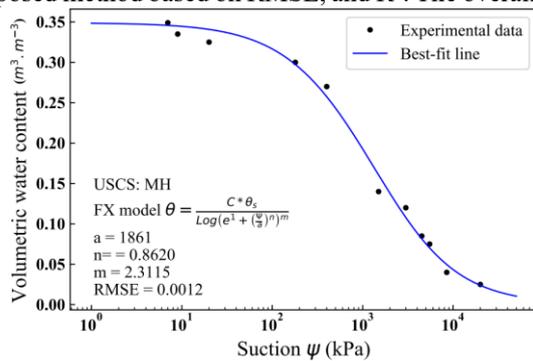
The validity of the SWCC estimation for both fine- and coarse-grained soils using the proposed equation was examined using Aburra Valley dataset. Fig. 3 show the best (8) samples with best fitting and the latest (4) with the lower accuracy.

In this study, error analysis was performed to evaluate the proposed method based on RMSE, and R^2 . The overall

performance of the proposed method with the other one-point methods in terms of RMSE, and R^2 is summarized in Table 3.

The proposed method has the lowest RMSE and the highest R^2 values out of all the one-point methods. This suggests that the proposed method performed better than previous research.

The results showed that the proposed method is viable and performs better combined with Fredlund and Xing equation. From a study of the predicted features, and validations by laboratory tests, it is concluded that the technique is simple, easy to apply, versatile and accurate capturing the bimodal behaviour. Additionally, the equations have a rational basis which can readily be extended to different soils.



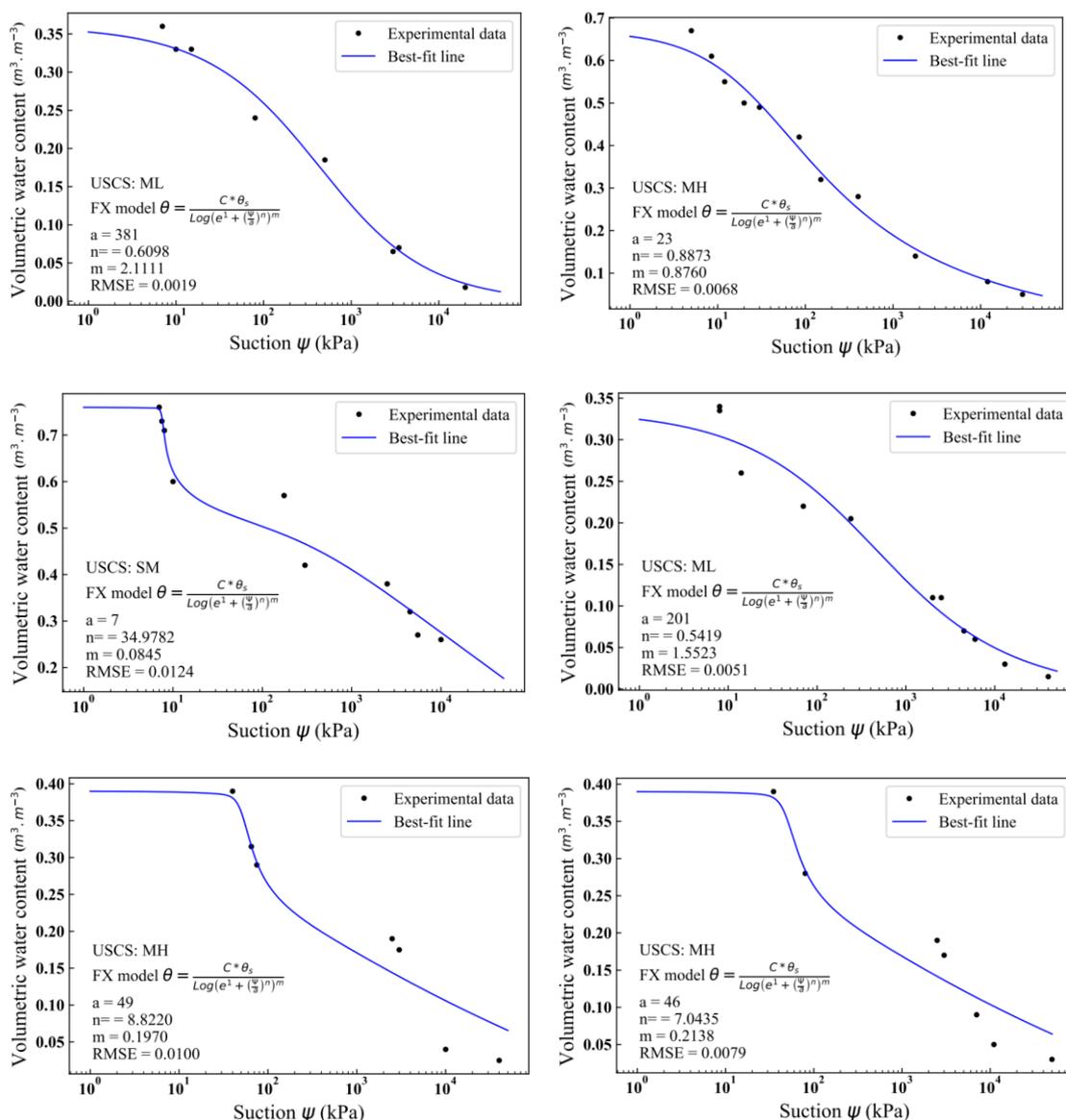


Fig. 3. SWCC regression curves obtained by implementing Nelder-Med fitting

4 Conclusions

Estimating the initial water content is the greatest uncertainty in the adjustment procedure, even more when θ_s was chosen as the highest value of the experimental data, given that the relation $\theta_s = (0.0143 * (wIP 0.75)) + 0.36$, from Zapata. This value underestimates the initial water content, resulting in a setting that is not good at any suction range.

By using the Nelder–Mead downhill simplex method, fewer calculations were required to estimate the optimum SWCC shape. Due to applying the Fredlund and Xing regression, the calculations for the developed system did not become trapped in a local optimum, and the algorithm could then determine the global optimum after a few iterations of the Nelder–Mead downhill simplex method.

Some patterns were found with a limited database and important conclusions about fitting parameters and USCS

were established, however, it had been found by different works that using a large database in conjunction with a knowledge-based system is perhaps the best way to select the optimal SWCC and fitting parameters.

Different strategies like use gradient descent as a derivative technique, make θ_r a learnable parameter, use Houston equations to approximate data, and finally using Nelder Mead Method with Gradient descent, with not significant performance improve. Use Nelder Mead was an appropriate methodology to the fitting process.

References

1. J. Williams, R. E. Prebble, W. T. Williams, & C. T. Hignett, (1982). The influence of texture, structure, and clay mineralogy on the soil moisture characteristic. *Australian Journal of Soil Research*, **21**:1, 15–32.

2. D. G. Fredlund, & H. Rahardjo, (1993). *Soil Mechanics for Unsaturated Soils*.
3. Y. A. Pachepsky, D. Timlin, & G. Varallyay, (1996). Artificial Neural Networks to Estimate Soil Water Retention from Easily Measurable Data. *Soil Science Society of America Journal*. **60**:3, 727–733.
4. Koekkoek, E. J. & Booltink, H. (1999). Neural network models to predict soil water retention. *European Journal of Soil Science*. **50**:3, 489–495.
5. A. Johari, G. Habibagahi, & A. Ghahramani, (2006). Prediction of Soil – Water Characteristic Curve Using. *Journal of Geotechnical and Geo-environmental Engineering*. **132**:1, 661–665.
6. E. C. Leong, & H. Rahardjo, (1997). Review of soil-water characteristic curve equations. *Journal of Geotechnical Engineering*. **123**:12, 1106–1117. [https://doi.org/10.1061/\(ASCE\)1090-0241\(1997\)123:12\(1106\)](https://doi.org/10.1061/(ASCE)1090-0241(1997)123:12(1106))
7. G. Guzmán, C. D. Hoyos (2017). Associating Land Surface Temperature Retrieved from Satellite and Unmanned Aerial Vehicle Data with Urban Cover and Topography in Aburrá Valley. *Conf. Amer. Geophysical Union*.
8. D. Henao & G. Monsalve (2017). Geological inferences about the upper crustal configuration of the Medellín – Aburra Valley (Colombia) using strong motion seismic records, *Journal Geodesy and Geodynamics*. **9**:1, 67-76. <https://doi.org/10.1016/j.geog.2017.06.005>
9. J. A. Nelder & R. Mead (1965). "A simplex method for function minimization". *Computer Journal*. **7**:4: 308–313. doi:10.1093/comjnl/7.4.308