Proc. of the 9th International Symposium on Geotechnical Safety and Risk (ISGSR)
Edited by Zhongqiang Liu, Jian Dai and Kate Robinson
Published by Research Publishing, Singapore. ©2025 The Author(s), CC-BY 4.0 licence.
doi: 10.3850/GRF-25280825_isgsr-189-P329-cd

ADVANCING SOIL-WATER CHARACTERISTIC CURVE ESTIMATION USING BAYESIAN ANALYSIS

Sophia K. Andanje[1], David F. Barba-Galdamez[2], Jungmin. Lee[3], Brian J. Reich[4], Brina M. Montoya[5]

[1,2,3,5] *Department of Civil, Construction and Environmental Engineering, North Carolina State University, 915 Partners Way, Raleigh, NC, 27695-7908, US*

[4]*Department of Statistics, North Carolina State University, Campus Box 8203, Raleigh, NC, 27695-7908, US*

Email: skandanj@ncsu.edu

The soil-water characteristic curve (SWCC) is crucial in understanding soil behavior under unsaturated conditions, with applications in slope stability, foundation design, and environmental engineering. Traditional methods for estimating the SWCC, such as direct measurements, mathematical modeling and grain size distribution (GSD)-based approaches, are often time-consuming, require extensive testing, and are limited by their inability to quantify uncertainties and address correlations among key parameters. This study presents a hierarchical Bayesian framework for SWCC estimation that leverages prior distributions of SWCC parameters derived from soil properties, including Atterberg limits, void ratio, and soil texture, from 170 soil samples. By accounting for parameter uncertainty and correlation, and heteroskedastic variance, the model provides probabilistic SWCC predictions with credible intervals for sandy and clayey soils. The approach is particularly suited for geotechnical applications, as it employs covariates commonly used in geotechnical practice and allows for reliable predictions with limited experimental data. These features make the proposed model a robust alternative to traditional models, enhancing the understanding of soil suction and moisture retention under varying conditions.

Keywords: SWCC, Bayesian, Uncertainty, Prediction Accuracy, Unsaturated Soils.

1. Introduction

The soil-water characteristic curve (SWCC), a fundamental concept in unsaturated soil mechanics, represents the relationship between soil suction and water content. It plays a critical role in predicting a range of unsaturated soil properties such as shear strength, permeability, and volume change behavior (Lu and Likos, 2004; Fredlund et al., 2012). Understanding and estimating the SWCC is vital for geotechnical applications, including slope stability analysis, foundation design, and environmental engineering under unsaturated conditions. As climate change continues to alter water table levels and precipitation patterns, precise SWCC estimation becomes important in assessing soil behavior under changing environmental conditions (Vanapalli et al., 1996).

Traditional methods for determining the SWCC depend on experimental techniques, such as direct measurement using pressure plate extractors, and empirical models based on grain size distribution (Fredlund et al., 2012; Zhai and Rahardjo, 2013). Although these approaches are effective, they are often time-consuming, labor-intensive, and/or statistically limited due to their inherent variability in soil properties (Zhou et al., 2014). In recent years, there has been increasing interest in developing statistical and probabilistic methods to estimate the SWCC from measurable soil properties, such as soil texture (Aubertin et al., 2003; Juang et al., 2019). Several attempts have also been made in the SWCC uncertainty quantification through Bayesian estimation (Wang et al., 2018; Zhang et al., 2022). However, these techniques focus on covariates more related to agronomy and fail to consider collinearity among covariates. In addition, quantification of uncertainty on regression parameters associated with SWCC estimations is insufficient, leaving a gap in both the precision and reliability of current methods.

This study addresses this gap by introducing a Bayesian approach that advances the SWCC estimation, leveraging prior information about soil properties obtained from rapid, cost-effective tests, such as fines content (f), liquid limit (LL), plasticity index (PI), and void ratio (e), and updating these estimates with limited experimental data. The hierarchical Bayesian model used in this analysis accounts for both the inherent variability of soil parameters, uncertainty and correlation in the regression parameters. The model considers a non-constant variance (i.e. heteroskedasticity) improving the understanding of residuals behavior, thereby providing a more robust framework for SWCC prediction. The proposed model will help improve the accuracy and efficiency of SWCC estimation, for applications in civil and geotechnical engineering where resources for extensive experimental testing are limited.

2. Methods

This study employs a hierarchical Bayesian model to estimate the SWCC through a probabilistic framework. A soil database was compiled from laboratory and field experiments reported in the literature, encompassing

diverse soil types. To address variability in the methods and models used to develop SWCCs, their suitability as prior information was evaluated using a two-step validation process implemented with the *soilhypft* R package, following Gupta et al. (2022). The van Genuchten-Mualem (VGM) shape parameters were constrained during this process and a visual inspection of the SWCCs was performed to verify fitness for the VGM model of the experimental data, resulting in a validated dataset of 170 soil samples. These samples, containing information on soil texture, Atterberg limits, void ratios, and SWCC parameters, served as the basis for the study.

2.1.*SWCC Model Development*

The VGM model, which is widely used in geotechnical engineering describes the relationship between degree of saturation *(S)* and soil matric suction(ψ)using parametersα and n, which are specific to the soil sample(van Genutchen-Mualem, 1980). To estimate the VGM model parameters (α and n), a hierarchical Bayesian model is developed to account for the nested data structure, and associated uncertaintiesusing three distinct layers:

Data layer:

$$S_{ij}\left|\ln\alpha_j,\ln n_j,\sigma^2\sim N\left(g_{ij},\sigma^2 g_{ij}\left(1-g_{ij}\right)\right)\right. \quad (1)$$

$$g_{ij}=g\left(\psi_{ij}\left|\alpha_j,n_j\right.\right)=\frac{1}{\left[1+\left(\alpha_j\psi_{ij}\right)^{n_j}\right]^{1-\frac{1}{n_j}}} \quad (2)$$

Process layer:

$$\left\{\ln\alpha_j,\ln n_j\right\}^T\left|\beta_{jk},\boldsymbol{\Sigma}\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})\right. \quad (3)$$

$$\mu_1=\beta_{j1}+\beta_{j2}\,ln\left(f_j+1\right)+\beta_{j3}ln\left(LL_j+1\right)+\beta_{j4}\,ln\left(PI_j+1\right)+\beta_{j5}ln\left(e_j\right) \quad (4)$$

$$\mu_2=\beta_{j6}+\beta_{j7}\,ln\left(f_j+1\right)+\beta_{j8}ln\left(LL_j+1\right)+\beta_{j9}\,ln\left(PI_j+1\right)+\beta_{j10}ln\left(e_j\right) \quad (5)$$

Prior layer:

$$\beta_{jk}\sim N(0,1000)\;for\;k=1,2,\ldots,10 \quad (6)$$

$$\Sigma\sim invWishart\;(2.1,\mathbf{I_2}/2.1) \quad (7)$$

$$\sigma^2\sim InvGamma\;(2.225,0.021) \quad (8)$$

The data layer level models the degree of saturation (S_{ij}) as a stochastic variable following a heteroskedastic normal distribution (Eq. 1), with the VGM-predicted degree of saturation $g_{ij,}$ dependent on matric suction (ψ_{ij}) and the soil-specific parameters α_j and n_j(Eq. 2). The heteroskedastic variance term, $\sigma^2 g_{ij}$ $(1-g_{ij})$, accounts for variability in residuals, with the variance scaling based on the predicted saturation g_{ij}. This formulation allows the model to account for the inherent variability observed in soil-water retention data. The process layer (Eq. 3) models the parameters $\alpha_j$and n_j as dependent on soil-specific covariates through a multi-linear regression frameworkexpressedin Eq. 4 and Eq. 5. These covariates include fines content (f), liquid limit (LL), plasticity index (PI), and void ratio (e). The regression models allow α_j and n_j to vary with soil properties, enabling the model to adapt predictions across diverse soil types. The regression coefficients (β_{jk}) capture the influence of covariates on each parameter, while the residual term ($\sum$) captures unmodeled variability. The prior layer then introduces hyperpriors to regularize model parameters and account for additional sources of uncertainty as follows. First, weak informative priors are assigned to the regression coefficients (Eq. 6)to allow flexibility in the parameter estimation process while discouraging extreme values. Then, the regression coefficient covariance matrix is modelled using an inverse-Wishart distribution (Eq. 7), to ensure realistic correlations among the regression coefficients across multiple covariates. Finally, the variance components in the data layer are governed by an informative inverse-gamma prior (Eq. 8) derived using the residuals obtained from the 170 soil samples, and the inverse-gamma parameters were estimated using the method of moments (MOM) to balance prior information and model flexibility.

3.*Model Fitting and Validation*

To assess model performance, the SWCC data was fit to the VGM model through three stages: predicted versus observed saturation, residual analysis, and posterior error variance estimation. As shown in Figure 2a, the predicted VGM model for soil sample S003 (clayey-silty soil) aligns well with the observed data, capturing key SWCC behavior. Slight deviations between predicted and observed saturation suggest localized variability, which is evaluated through residual diagnostics (Figure 2b). The residuals *(μ)* are small and randomly

distributed, confirming the model's accuracy without systematic bias. This pattern is consistent with the heteroskedastic variance assumption in the model. The posterior estimation of error variance (σ^2)(Figure 2c) is concentrated at small values, indicating that the model effectively explains data variability while maintaining minimal residual error. The alignment of the posterior with the prior demonstrates a balance between prior information and data-driven updates, ensuring that uncertainty is captured without overfitting.

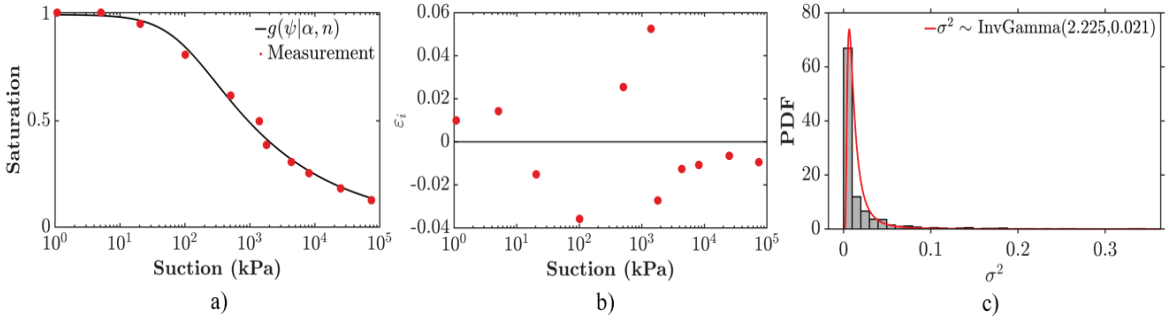


Fig. 2. Model Fitting. a) Example of fit with R package for soil sample S003, b) Example of residuals vs matric suction for soil sample S003, c) Distribution of squared errors for all soil samples used in the model.

To evaluate the interdependence of the VGM parameters and the accuracy of their predictions, the correlation between $ln\alpha$ and $lnn$was analyzed. A weak negative mean correlation between ln α and ln n (Fig. 3a) aligns with expectations of compensatory behavior in SWCC models. The narrow posterior spread indicates that the correlation is well-constrained by the data. To evaluate the estimations obtained with the multilinear regression, we conducted a 10-fold cross-validation scheme on the data. Cross-validation results (Fig. 3b and 3c) show strong agreement between predicted and observed values, clustering around the 1:1 line, demonstrating that the multilinear regression reliably estimates ln α and lnn for different soils. The reported correlation coefficients ($\acute{A}$) are similar to those obtained by Zhang et al. (2022).

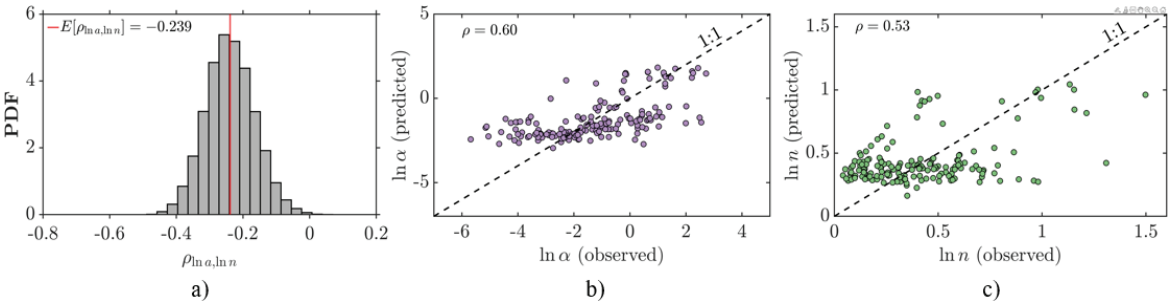


Fig 3. a) Correlation distribution between $ln\alpha$ and lnn. b) Cross-correlation results for $ln(\pm)$. c) Cross-correlation results $ln(n)$.

4. Bayesian Model Application

The accuracy of the hierarchical Bayesian model was validated using two independent soil samples: S112 (sandy soil) and S208 (clayey soil), which were not included in model development.The properties of S112 were$f$= 0.6, LL = 18, PI=0, e=0.84, while S208 had$f$= 0.75, LL = 32, PI=5 and e=2.01. The SWCC was estimated using Eq. 1 based on measured data points for suction and degree of saturation. Figures 4a and 4c show the 95% credible interval (CI) for predictions with two measured data points. For both soils, all experimental points fall within the 95% CI, demonstrating the model's accuracy. However, the larger CI width for the clayey soil indicates higher uncertainty in predicting degree of saturation at a given suction. Increasing the measured data points to four (Figures 4b and 4d), significantly narrows the CI, reducing uncertainty.

A comparison with Zhang's model (Zhang et al., 2022) highlights the advantages of the proposed model. For sample S208, Zhang's model exhibits substantial uncertainty, as most experimental points fall outside the 95% CI (Figures 5a and 5b). This can be attributed to the model's reliance on covariates that are collinear. In contrast, our hierarchical Bayesian model uses geotechnical covariates such as fines content, liquid limit, plasticity index, and void ratio, while accounting for inherent variability, regression parameter uncertainty, correlation, and heteroskedastic variance. These features make the proposed model more robust and aligned with geotechnical engineering needs.

5. Conclusion

The hierarchical Bayesian model effectively predicts the SWCC, outperforming traditional models by addressing variability in soil properties, and parameter uncertainty while utilizing covariates relevant to geotechnical engineering.

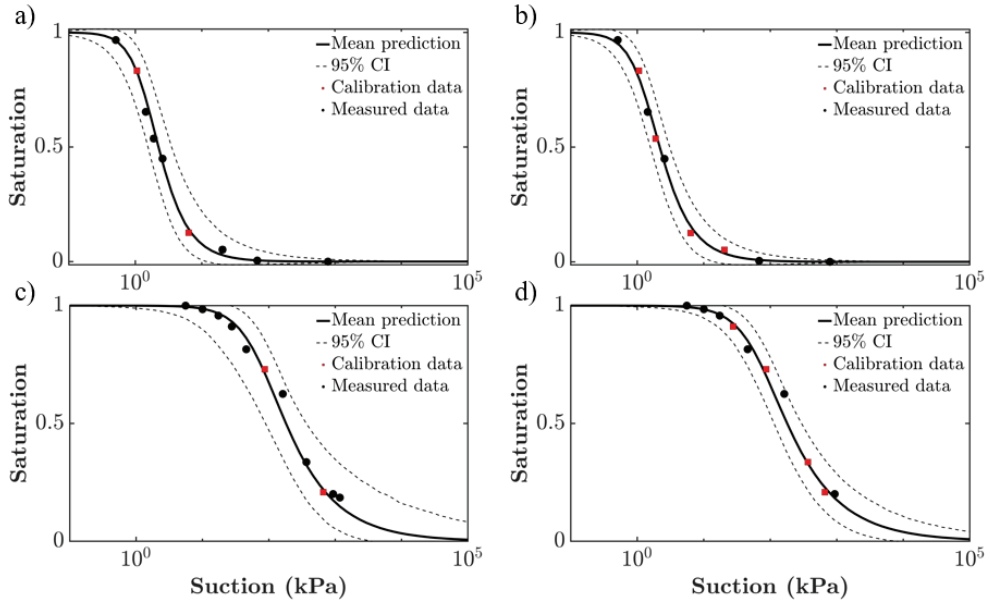


Fig 4. Predicted SWCC using the proposed model: a) S112 (Sandy soil, 2 points), b) S112 (Sandy soil, 4 points), c) S208 (Clayey soil, 2 points), d) S208 (Clayey soil, 4 points).

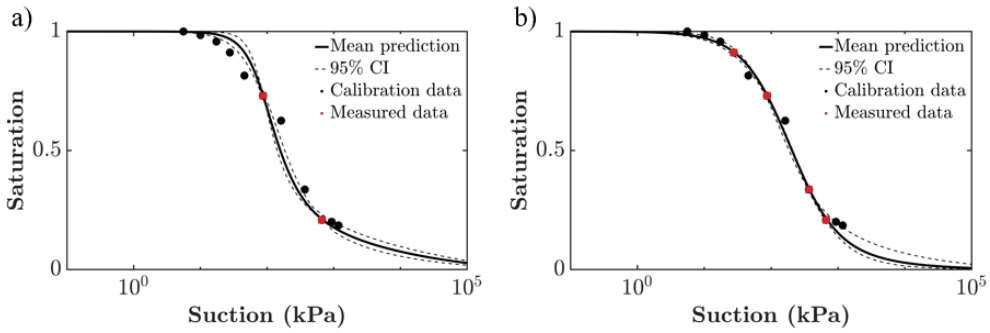


Fig 5. Predicted SWCC using Zhang's model a) S208 (Clayey soil, 2 points), b) S208 (Clayey soil, 4 points).

Results from the Bayesian model show accurate predictions within the 95% CI, with reduced uncertainty as data points increase. By incorporating heteroskedastic variance and robust prior information, the framework delivers reliable SWCC estimates, particularly for clayey soils. This approach improves the accuracy and efficiency of SWCC predictions, providing practical value in geotechnical applications with limited experimental data.

References

Aubertin, M., Mbonimpa, M., Bussière, B., & Chapuis, R. P. (2014). A model to predict the water retention curve from basic geotechnical properties. *Canadian Geotechnical Journal*, 40(6), 1104.

Fredlund, D. G., Rahardjo, H., and Fredlund, M. D. (2012). *Unsaturated Soil Mechanics in Engineering Practice*. John Wiley & Sons, Inc.,Hoboken, New Jersey.

Gupta, S., Papritz, A., Lehmann, P., Hengl, T., Bonetti, S., and Or, D. (2022). Global soil hydraulic properties dataset based on legacy site observations and robust parameterization. *Scientific Data, 9(1).*

Juang, C. H., Zhang, J., Shen, M., & Hu, J. (2018). Probabilistic methods for unified treatment of geotechnical and geological uncertainties in a geotechnical analysis. *Engineering Geology*, 249, 148.

Lu, N., and Likos, W. J. (2004). *Unsaturated Soil Mechanics*. John Wiley & Sons, Inc., Hoboken, New Jersey.

Vanapalli, S.K., Fredlund, D.G., Pufahl, D.E., and Clifton, A.W. (1996). Model for the prediction of shear strength with respect to soil suction. *Canadian Geotechnical Journal, 33(3),* 379–392.

van Genuchten, M.T. (1980). A closed form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J. 44*, 892-898.

Wang, L., Cao, Z., Li, D., Phoon, K., and Au, S. (2017). Determination of site-specific soil-water characteristic curve from a limited number of test data – A bayesian perspective. *Geoscience Frontiers, 9(6),* 1665.

Zhang, J., Yang, S., Zhang, L. L., and Zhou, M. L. (2022). Bayesian estimation of soil-water characteristic curves. *Canadian Geotechnical Journal, 59(4),* 569.

Zhou, W., Yuen, K., and Tan, F. (2014). Estimation of soil–water characteristic curve and relative permeability for granular soils with different initial dry densities. *Engineering Geology, 179*, 1.