Proceedings RootS25
3rd International Workshop on Soil-Vegetation-Atmosphere Interaction
Boldrin D, Cecconi M, Cotecchia F, Leung A, Pedone G, Perrini P, Romero E, Tagarelli V, Zdravković L (eds)
ISSN: 3107-300X

doi: doi.org/10.53243/RootS2025-22

Hierarchical Bayesian modeling of root distribution for the geotechnical analysis of bioengineered systems

Stefano Collico[1], Marco Uzielli[1]

[1]*Department of Civil and Environmental Engineering, University of Florence, Italy*

ABSTRACT: The geotechnical modeling of rooted soils is challenging due to the strong spatial variability of root systems and to the epistemic uncertainties stemming from the typically limited and sparse available datasets and from the use of simplified analytical models for the estimation of root-induced cohesion. This study introduces a flexible Hierarchical Bayesian Regression model for the modeling of the vertical spatial variability of root area ratio. The proposed approach includes a non-monotonic functional form which ensures improved coherency with empirical data. The model is trained on a preliminary multi-species dataset and subsequently validated on new data, demonstrating the utility of the Bayesian approach in reducing epistemic uncertainty in typical scenarios of limited evidence. The probabilistic outputs of the approach can be used directly in reliability-based design of bioengineered geotechnical systems.

Keywords: Root-induced cohesion; spatial variability; epistemic uncertainty; Hierarchical Bayesian modeling

1 INTRODUCTION

Root reinforcement is increasingly considered in geotechnical engineering to modify the response to gravitational loading, erosion, and other hydroclimatic phenomena due its capability of altering hydro-mechanical parameters and soil state, thereby improving stress path behaviour. The response of root-reinforced soils strongly depends on bio-mass system characteristics, which are plant species-specific, site-specific and climate-dependent, e.g., with respect to temperature (Bischetti et al. 2009). The geotechnical modeling of rooted soils is particularly challenging due to the aleatory spatio-temporal variability of root patterns resulting in a marked heterogeneity of induced reinforcement and has important implications, for instance, in slope stability analysis. The mechanical contribution of roots to soil strength is typically modeled by accounting only for the fiber reinforcement effect (Vergani et al., 2017; Bischetti et al., 2009) and parametrized by an additional cohesion term. However, direct measurements of root-induced cohesion are generally limited, so nonlinear transformation models (i.e., regression) are commonly employed to infer root cohesion indirectly from root diameter measurements. Local (i.e., site and species-specific) transformation models can, in principle, provide estimates of root cohesion. However, when based on limited data such models may suffer from statistical significance and carries a high risk of extrapolation across the site. Alternatively, generic transformation models, may be applied when available. Yet, these models might introduce substantial transformation uncertainty because they attempt accomodate a wide range of soil types and plant species, often resulting in overly conservative lower-bound estimates of parameter of interest. More advanced approaches have been proposed by Xian et al. (2025) on Hierarchical Bayesian Regression Model for predicting root tensile strength from root diameter. However, no prior studies have applied this framework to Root Area Ratio (RAR). This work introduces a Hierarchical Bayesian Regression Model to predict the site- and species-specific spatial variability of the Root Area Ratio (*RAR*), which serves as input to multiple models available in the technical literature for estimating root-induced cohesion. Temporal variation is not addressed here and will be explored in future research. The proposed hierarchical structure transfers information from generic transformation models to local predictions while accounting for site- and plant-specific variability, enabling statistically consistent inference of local root reinforcement parameters even with limited measurements. Finally, a flexible parametric formulation is proposed to capture the aleatory spatial variability of RAR, providing a better fit than the monotonic trend typically used.

2 MODELING OF ROOT-INDUCED COHESION

During soil mass movement, shear stresses develop along the slip surface, and roots intersecting this zone are engaged primarily in tension. In absence of direct measurements, the spatio-temporal root cohesion c_r can then be expressed according to analytical tranformation model such as the one initially proposed by Wu (1976) and Waldron (1977):

$$c_r(\mathrm{t,z}) = k' \sum_{j=1}^{n_d} T_{rj}(t,z) \cdot RAR(t,z) \quad (1)$$

where $T_{rj}(t,z)$ is the root tensile strength for the j-th diameter class, n_d number of diameter classes at z depth and RAR is the Root Area Ratio defined as the ratio between root area and rooted-soil area. The coefficient k' is a correction factor, usually taken as 1.2 (Preti et al., 2010), introduced to account for the fact that roots are not uniformly oriented perpendicularly to the slipping surface.

2.1 *Spatial variability of the Root Area Ratio*

The RAR provides a measure of root density within the soil and depends on site-specific state (i.e., soil moisture), climate characteristics, and land use management (Bischetti et al., 2009). The vertical spatial variability of RAR along depth z at a given temporal instant has typically been modelled using an exponential function (Ng et al., 2019):

$$RAR(\mathrm{z}) = \theta_1 \, z^{-\theta_2} \cdot \varepsilon_{RAR} \quad (2)$$

where θ_1, θ_2 are positive coefficients and ε_{RAR} is the error introduced by the transformation model. The temporal dependency of RAR is not included explicitly for simplicity of notation. The coefficient θ_1 provides the RAR value at ground surface. However, limitations of these monotonic models have been highlighted. Bischetti et al. (2009) observed that, for most species, RAR initially increases and then decreases with increasing depth. To capture non-monotonic behavior, the same authors proposed a flexible transformation model relying on the analytical formulation a Gamma probability density function. However, their formulation forces $RAR(0) = 0$, which conflicts with empirical evidence. Here, a nonlinear transformation model given by the shifted Weibull probability density function equation is proposed:

$$RAR(\mathrm{z}) = A \cdot \frac{k}{\lambda}\left(\frac{z+\delta}{\lambda}\right)^{k-1} e^{\left[-\left(\frac{z+\delta}{\lambda}\right)^{k}\right]} \quad (3)$$

where k and λ are the shape and scale parameters, respectively; δ is a positive depth offset coefficient; and an A is an amplification. The parameters $\{A, k, \lambda, \delta\}$ are inferred from the data within the hierarchical Bayesian regression model introduced in Section 4. This model removes the constraint of null root area at ground surface. The non-zero depth z_p at which RAR achieves its maximum value is obtained by imposing $\partial(RAR)/\partial z$=0, thereby yielding:

$$z_p = \lambda \left({}^{(k-1)}/_{k} \right)^{1/k} - \delta \quad (4)$$

3 DATASET

This study relies on a preliminary novel multi-species dataset which is currently being compiled from an extensive desk study and whose work-in-progress status is given in Geppetti et al. (2025). For each plant species, the dataset allocates information about location, region, mean annual precipitation, soil texture, taxonomic classification, age, diameter of the stem, RAR at different depths, horizontal distance of RAR measurements from the stem. Moreover, it provides average, minimum and maximum values of root tensile strength T_r, and regression's coefficients (a, b) for root diameter (d) - tensile strength (T_r) correlation according to Gray and Sotir (1996) power law model, such that: $T_r = a \cdot d^{-b}$. For this study, a total of 16 distinct plant species are considered, representing a subset of the dataset currently being compiled by Geppetti et al. (2025). The depthwise variability of RAR for all of the species is shown in Figure 1, where each color represents a different plant species. Spline interpolations are applied to measured data values to better highlight the existence of non-monotonic spatial trends.

4 HIERARCHICAL BAYESIAN MODELING

Among Bayesian approaches, Hierarchical Models (aka HBM) have been widely applied within various domains of the geotechnical discipline (e.g., Ching & Phoon, 2021; Collico et al., 2024). In such studies, site-level unknown random variables are quantified from available observed site-level measurements by updating hyperparameters priors that pool information across sites. In applications to soil-root systems, Xian et al. (2025) proposed a HBRM for root tensile strength prediction from root diameter. Compared with conventional machine learning techniques, Hierarchical Bayesian Models provide transparent parameter inference, explicitly quantify uncertainty, and maintain physical interpretability, rather than relying on purely data-driven fitting. Following the same workflow, a Hierarchical Bayesian Regression structure is proposed to model the spatial variability of RAR and allow its inference from limited measurements.

Let m define the number of plant species and let y_{ij} define the j-th observation of RAR for the i-th species (i=1,…,m) with n_i being the number of observations datapoints (j=1,…,n_i). The likelihood function for y_{ij} is given by:

$$p\left(y_{ij} \middle| \Theta_i, \sigma_i^2\right) = \mathcal{N}\left[y_{ij}; g\left(z_{ij} \middle| \beta_i\right), \sigma_i^2\right] \quad (5)$$

in which $g(z_{ij}|\beta_i)$ is the Weibull distribution given in Equation (3), $\beta_i = \{A_i, k_i, \lambda_i, \delta_i\}^T$ is the vector of regression model coefficients, and ε_i~$\mathcal{N}(0,\sigma_i^2)$ is the epistemic transformation error, with σ_i^2 being the species-specific epistemic model variance. The model variance is defined as an a scaled inverse-χ^2 distribution through a two-step procedure involving sequentially: (1) the assignment of an auxiliary inverse-gamma-distributed scale parameter ξ with hyperparameters a_0 and b_0; and (2) the specification of a further hyperparameter ν_0. Note that the positivity constrains imposed on the Weibull coefficients (i.e., $\beta_i > 0$) allow the log-transformation $\Theta_i = \{\ln A_i, \ln k_i, \ln \lambda_i, \ln \delta_i\}^T$. The vector of prior parameters for the i-th species is then modelled as:

$$\Theta_i | B, \Sigma_\Theta \sim \mathcal{N}(B, \Sigma_\Theta) \tag{6}$$

in which $B \sim \mathcal{N}(\mu_0, \Lambda_0)$ is the weakly informative prior for the population-level mean vector and Σ_Θ is the diagonal variance–covariance matrix of the log-coefficients, whose entries are inverse gamma-distributed with hyperparameters n_0 and S_0 since no dependence is assumed among model coefficients. The hyperparameters μ_0, Λ_0, n_0, S_0, a_0, b_0 and ν_0 are common across all species and were selected in agreement with the ones reported in Sosa & Aristizabal (2021) and Ching & Phoon (2021). The posterior distribution is calculated using Bayes‘ theorem as:

$$\begin{aligned} p\left(\{\Theta_i\}_{i=1}^{m}, \{\sigma_i^2\}_{i=1}^{m}, B, \Sigma_\Theta, \xi | y_{ij}\right) \\ \propto \prod_{i=1}^{m} \prod_{j=1}^{n_i} p\left(y_{ij} | \Theta_i, \sigma_i^2\right) \\ \cdot \prod_{j=1}^{n_i} p(\Theta_i | B, \Sigma_\Theta) \\ \cdot \prod_{j=1}^{n_i} p\left(\sigma_i^2 | \xi, \nu_0\right) \cdot p(\xi | a_0, b_0) \\ \cdot p(B | \mu_0, \Lambda_0) \cdot p(\Sigma_\Theta | n_0, S_0) \end{aligned} \tag{7}$$

The nonlinear character of the Weibull model in Equation (3) requires a hybrid MCMC scheme involving: (1) a Gibbs sampler for the conjugate updating of σ_i^2, B, Σ_Θ, and ξ; and (2) slice sampling for the non-conjugate updating of the log-transformed Weibull model parameters Θ_i. The graphical representation of the hierarchical Bayesian structure is reported in Figure 2.

5 APPLICATION

5.1 *Training phase*

In the training phase, the unknown hyperparameters are estimated together with the plant-specific coefficients while explicitly accounting for their associated uncertainties. To ensure statistical robustness, a total of 3000 posterior samples are generated and employed for inference. Example of fitted trends, in terms of mean and 95% credible intervals from the respective posterior distributions, for two different plant species (Picea Abies and Fagus sylvatica) are reported in Figure 3, which illustrate how the adopted parametric form effectively adapts to interspecific differences. From these results, a clear difference is observed with respect to the monotonic exponential formulation of Equation (2). In particular, the proposed model captures the empirically clearly visible existence of peak values of *RAR* which would otherwise be impeded by the monotonic assumption.

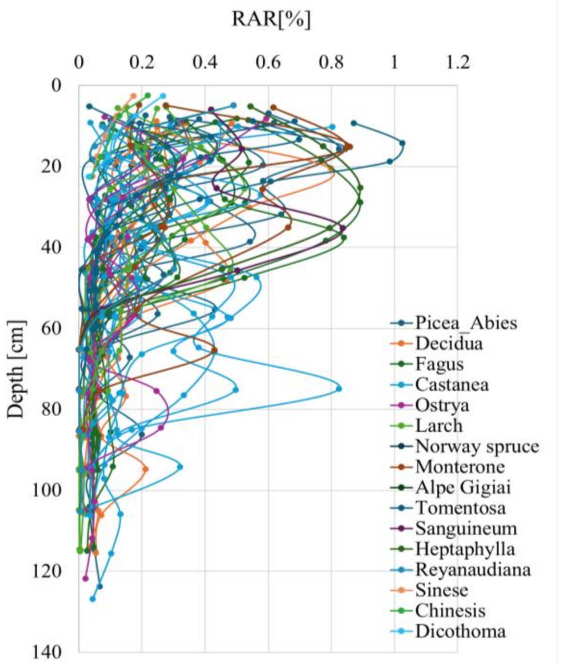

Figure 1. RAR measurements along depth for different plant species with spline interpolations for improved visualization.

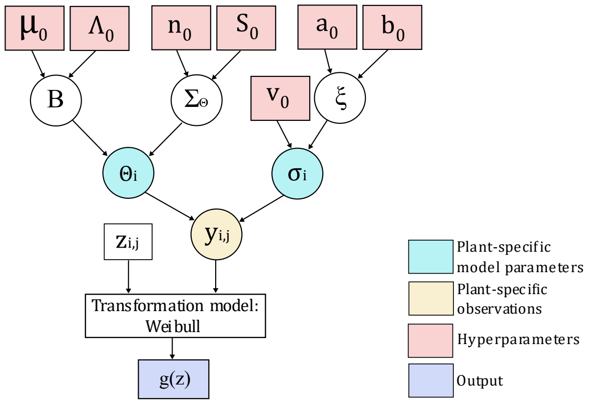

Figure 2. Graph representation of the structure of the Hierarchical Bayesian regression model

5.2 *Validation*

The hierarchical model was validated using the *Heptaphylla* species, for which 10 individual data points were withheld from the training dataset and used exclusively for validation. The spatial prediction of *RAR* is conducted parametrically by reintroducing a variable number n_v of data points. Figure 4 shows the depthwise posterior mean and 95% credible interval for n_v=1, n_v=3, and n_v=10 (i.e., the complete dataset for the species). While for n_v=1 model predictions tend to the trend inferred from the hyperparameters calibrated in the training process (i.e., blue line), increasing the number of measurements improves the adaptation of the prediction to species-specific measurements. The progressive restriction of the credible interaval attests to the decrease in epistemic prediction uncertainty.

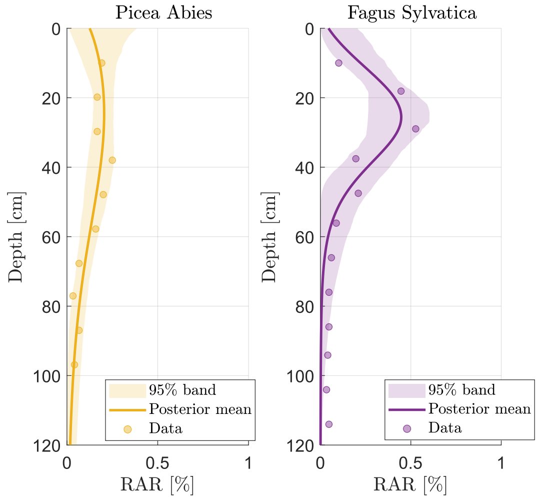

Figure 3. Example of posterior mean and 95% credible intervals for Weibull models of two plant species

6 CONCLUSIONS

This paper illustrates the main modeling steps and seleted results of the application of a Hierarchical Bayesian Regression model to characterize the spatial distribution of the Root Area Ratio. The proposed model directly tackles the challenges posed by the typical scenario of limited datates, plant-specific spatial variability, and the epistemic uncertainty in transformation models. A flexible nonlinear formulation, based on a shifted Weibull distribution, is proposed to allow a more coherent prediciton of the spatial variability of root area ratio to empirical measurements in comparison with widely used monotonic transformation models. The validation of the Bayesian framework demonstrates the reduction of epistemic uncertainty with increasing data numerosity. This improved representation, along with the availability of complete posterior distributions, provide an enhanced basis for estimating root-induced reinforcement in the context of limit-state and reliability-based design of bioengineered geotechnical systems.

7 ACKNOWLEDGEMENTS

This research was partially supported by Horizon Europe project ARAGORN (grant ID 101112723).

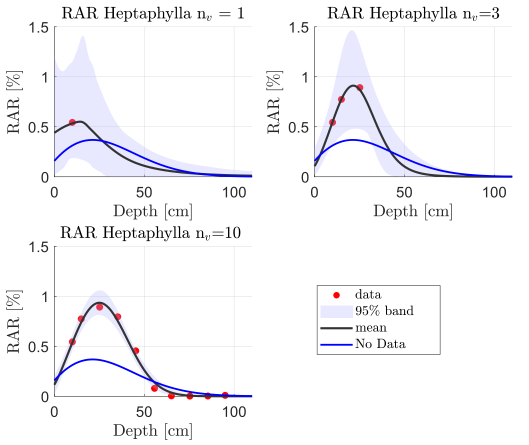

Figure 4. Prediction of RAR variability and uncertainty due to regression coefficients for Heptaphylla species as a function of the number of measurements

8 REFERENCES

Bischetti, G. B., Chiaradia, E. A., Epis, T., Morlotti, E. 2009. Root cohesion of forest species in the Italian Alps, *Plant and soil* **324**(1), 71–89.

Ching, J., Wu, S., Phoon, K. K. 2021. Constructing quasi-site-specific multivariate probability distribution using hierarchical Bayesian model, *Journal of Engineering Mechanics* **147**(10), 04021069.

Collico, S., Spagnoli, G., Romero, E., Fraccica, A. 2024. A Bayesian clustered-multilevel updating for local undrained shear strength prediction of fine-grained soils, *Applied Clay Science* **257**, 107444.

Geppetti A., Dani A. Preti F., Uzielli. M. 2025. A multivariate database for root-induced mechanical reinforcement: status of work and initial insights. *Proceeding of Roots 2025,* Bari, Italy.

Gray, D.H., Leiser, A.T. 1982. *Biotechnical Slope Protection, Erosion Control*, Van Nostrand Reinhold, New York.

Ng, C., Leung, A., Ni, J. 2019. *Plant-soil slope interaction*, CRC Press, Florida, United States.

Preti, F., Dani, A., Laio, F. 2010. Root profile assessment by means of hydrological, pedological and above-ground vegetation information for bio-engineering purposes, *Ecological Engineering* **36**(3), 305–316.

Sosa, J., Aristizabal, J. 2021. A gentle introduction to Bayesian hierarchical linear regression models. arXiv preprint arXiv:2110.10565.

Vergani, C., Giadrossich, F., Buckley, P., Conedera, M., Pividori, M., Salbitano, F., Schwarz, M. 2017. Root reinforcement dynamics of European coppice woodlands and their effect on shallow landslides: A review, *Earth-science reviews* **167**, 88–102.

Waldron, L. J. 1977. The shear stress resistance of root-permeated homogeneous and stratified soil, *Soil Sci. Soc. Am. Pro.* **41**, 843–849.

Wu, T. H. 1976. Investigation on landslides on Prince of Wales Island, Alaska. *Geotech Rpt. No 5*, Dpt. of Civil Eng., Ohio State Univ. Columbus, USA.

Xian, J., Zhu, J., Leung, A. K., Wu, C., Feng, S., Zhang, J. 2025. Using global sensitivity analysis to quantify the uncertainty of root reinforcement in vegetated slope stability, *Plant and Soil* **513**, 1741–1771.