Supplement of :

Reproducibility of the Wet Part of the Soil Water Retention Curve : A European Interlaboratory Comparison

Benjamin Guillaume[1], Hanane Aroui Boukbida[2], Gerben Bakker[3], Andrzej Bieganowski[4], Yves Brostaux[1], Wim Cornelis[5], Wolfgang Durner[6], Christian Hartmann[2], Bo V. Iversen[7], Mathieu Javaux[8], Joachim Ingwersen[13], Krzysztof Lamorski[4], Axel Lamparter[9], András Makó[10], Ana María Mingot Soriano[11], Ingmar Messing[11], Attila Nemes[12], Alexandre Pomes-Bordedebat[1], Martine van der Ploeg[3], Tobias Weber Karl David[13], Lutz Weihermüller[14], Joost Wellens[1], and Aurore Degré[1]

[1]Uliège - Gembloux Agro-Bio Tech, TERRA Teaching and Research Centre, Passage des Déportés 2, 5030 Gembloux, Belgium
[2]Instrumentation, Moyens Analytiques, observatoire en Géophysique et Océanographie (UAR IMAGO), Institut de Recherche pour le Développement (IRD)
[3]Wageningen University and Research, Netherlands
[4]Institute of Agrophysics, Polish Academy of Sciences, Poland
[5]Ghent University, Belgium
[6]Technische Universität Braunschweig, Germany
[7]Department of Agroecology, Aarhus University, Denmark
[8]UCLouvain, Belgium
[9]Federal Institute for Geosciences and Natural Resources, Germany
[10]Department of Soil Physics and Water Management, Institute for Soil Sciences, Centre for Agricultural Research, Herman Ottó street 15, 1022 Budapest, Hungary
[11]Swedish University of Agricultural Sciences, Sweden
[12]Norwegian Institute of Bioeconomy Research, Norway
[13]Institute of Soil Science and Land Evaluation, University of Hohenheim, Germany
[14]Agrosphere Institute IBG-3, Forschungszentrum Jülich GmbH, Germany

Correspondence: Benjamin Guillaume (benjamin.guillaume@uliege.be)

1   The reference sample

A reference sample fulfilling specific requirements was needed for this research. The water release of this reference sample needs to be clearly measurable in the range of water potential studied. Preferably, the water content should be linearly proportional to pF. Ideally, the amount of water release between each 1 pF increase should be equal to or greater than 0.1 ($cm^3.cm^{-3}$). The release of water over time should be comparable to that of a real sample. The sample must be reproducible. The reference sample should also be robust enough to remain stable during the time of the study.
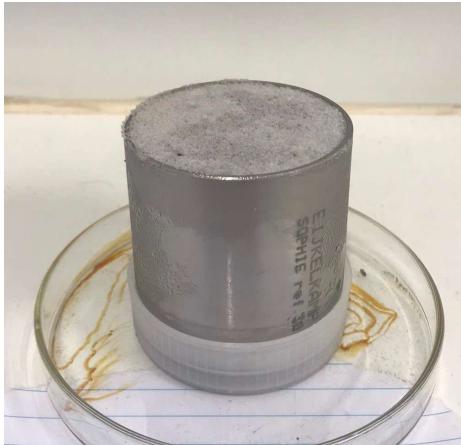
Figure S1. Example of a reference sample used during the interlaboratory comparison.

2 Pre-processing of data

The gravimetric water content (wc) in g.g^{-1} associated with each of the retention points is calculated from mass measurements using the equation S1.

$$wc = \frac{m - m_{\text{dried}} - (m_{\text{entrapped.hpa}} * m_{\text{cloth}})}{m_{dried} - m_{\text{ring}} - m_{\text{cloth}} - m_{\text{rubber}}} \tag{S1}$$

m is the mass (g) of the entire sample (with the ring, the cloth and the rubber) at a given suction. m_{dried} is the dry mass (g) of the entire sample (with the ring, the cloth and the rubber). m_{ring}, m_{cloth} and m_{rubber} are respectively the masses (g) of the steel ring, the dry cloth and the rubber. A correction to account for the water entrapped in cloth depending on the suction is also made. The mass of water entrapped ($m_{entrapped.hpa}$) in the cloth is (in g of water per g of cloth) *0.285* at 10 hPa, *0.172* at 50 hPa, *0.133* at 100 hPa and *0* at 300 hpa.

3 Parameters estimation

Each parameter from the models (described in the main paper), both on the intercept and on the slope, is estimated using Bayesian statistics. In the Bayes theorem, the posterior probability distribution which is calculated by the product of the likelihood with the prior normalized by the probability of the data (eq. S2).

$$\underbrace{P(parameter \mid data)}_{\text{Posterior}} = \frac{\overbrace{P(data \mid parameter)}^{\text{Likelihood}} * \overbrace{P(parameter)}^{\text{Prior}}}{\underbrace{P(data)}_{\text{probability of the data}}} \tag{S2}$$

The likelihood, P(data | parameter), is a function calculated from the data and proportional to the probability that the observed data would be produced by a given parameter value. It is conceptually the same function as in the frequentist approach from

which the maximum is used to estimate the "most probable" parameter value.

The prior distribution allows the researcher to state an appriori belief about the value of the parameter based on its expertise or on previous similar studies. In our case, such study is conduced for the first time and our prior will be mainly non-informative. However, it is possible to restrict parameter values based on the fact that the SWRC is physically bounded to avoid values that are considered impossible or out of range. In our case the priors specification is :

- $(\beta_0, \beta_1) \sim$ uniform[-0.2, 1]

- $(u_{0j}, u_{1j}) \sim$ normal$(0, \sigma_u)$

- $(v_{0k}, v_{1k}) \sim$ normal$(0, \sigma_v)$

- $(w_0, w_1) \sim$ uniform[-0.5, 0.5]

- $\sigma_u \sim$ uniform[0, 0.5]

- $\sigma_v \sim$ uniform[0, 0.5]

- $\sigma_\epsilon \sim$ uniform[0, 0.5]

The last term in the equation from Bayes is the denominator, P(data) or the probability of the data. The calculation of this term involves the integration of the numerator (likelihood * prior) over each model parameter dimension (i.e. $\beta_0, \beta_1, u_{0j}, u_{1j}, v_{0k}, v_{1k}$ and σ_ϵ). This implies a multiple integral that cannot be solved by traditional numerical methods. An alternative approach to avoid calculation of integrals in Bayes equation is to use stochastic simulation to draw samples directly from the posterior distribution thanks to a Markov chain Monte Carlo (MCMC) algorithm. The later algorithm is implemented in C++ through a package called "RStan" (Carpenter et al., 2017).

4 Model sensitivity : Understanding the influence of the Priors

The use of uniform priors tends to lead to high estimates of variance components, especially when the number of groups is low (Gelman, 2006). Hence, a sensitivity analysis of the priors implemented in the model (Eq. 3 in the main paper) is also carried out by rerunning the analysis with different forms of non-informative priors. The sensitivity analysis is limited to the variance components (σ_u, σ_v & σ_ϵ) which are the parameters most influenced by their prior distribution (van Erp et al., 2018). An inverse-gamma(ϵ, ϵ) distribution with ϵ set to 0.1, 0.01 and 0.001 and a half-Cauchy(0, σ) distribution with σ set to 1, 0.1 and 0.01 are used as noninformative or weakly informative priors for this sensitivity analysis (Gelman, 2006). The sensitivity analysis shows that the results are fairly robust to the uninformative priors, except for the inverse-gamma(0.1,0.1) priors (Fig. S2). In this case, the parameter estimates are less certain, with wider posterior distributions. The parameter values of the latent variances parameters (σ_{u0}, σ_{u1}, σ_{v0} and σ_{v1}) are also higher whereas it is robust for the residual variance (σ_ϵ). In this context, the inverse gamma prior is actually informative because the values of σ_{u0}, σ_{u1}, σ_{v0} and σ_{v1} are low (Gelman, 2006). Indeed, the mode of an inverse gamma(0.1, 0.1) distribution is 0.0909 (and 0.0091 for an inverse gamma(0.01, 0.01)). The prior adds
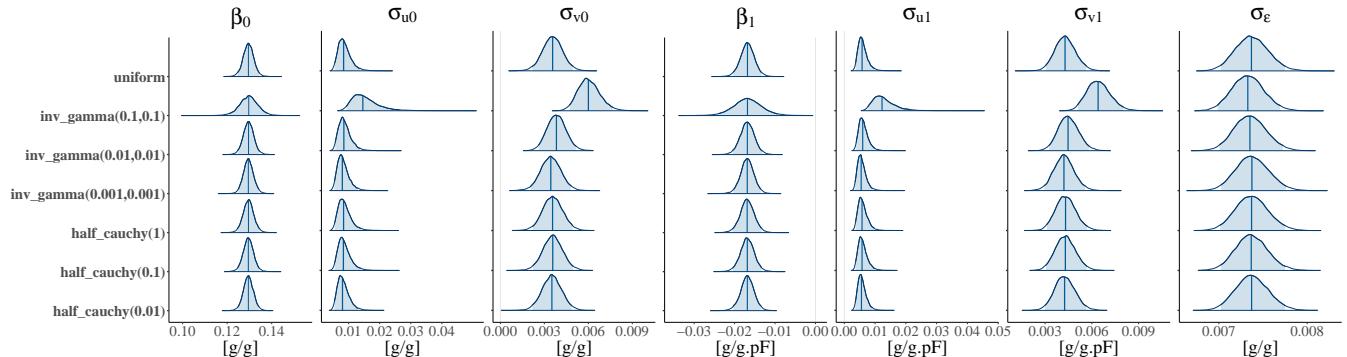
Figure S2. Densities of the posterior probability distributions of parameter estimates of the model (Eq. 3 of the main paper) for the original and alternative priors.

weight towards high values and this has the effect of shifting the posterior distribution of σ_{u0}, σ_{u1}, σ_{v0} and σ_{v1} to the right. Nothing of the sort is observed with half Cauchy priors as the mode is centred in 0.

5 Model validation : How well does the model predict data?

Graphical "prior predictive checks" and "posterior predictive checks" are used to qualitatively assess priors plausibility and model fit, respectively (Gabry et al., 2019). In both cases, modeled data set (y_{rep}) are generated from the posterior predictive distribution and compared to observed data set (y). The distinction between the prior and posterior predictive checks is that the posterior predictive distribution is obtained in the absence of input data for the prior predictive check. In other words, the model is fitted with priors as the only knowledge input for "prior predictive checks" while knowledge from the data is included in "posterior predictive checks". The prior predictive checks ensure that the prior specification encompasses the range of observed data. The posterior predictive checks ensure that the features of the modelled data reflect the observed data. Prior predictive check showed that the prior specification well encompasses the range of observed data (results not shown). Posterior predictive check (Fig. S3) shows how the model (Eq. 3 of the main paper) captures the observed data. Globally, the model represents the observed data correctly. One may recognize that the model fails to represent the bimodal features of the observed data densities. This is particularly true for laboratory number 2. These systematic fitting errors are explained by the non-linearity of some observed SWRCs that are modelled by a linear model.

6 Inference

The presence (or the absence) of an effect is determined by the Bayes factor. The Bayes factor provides a quantification of the predictive performance of a hypothesis (H_1) compared to a null hypothesis (H_0). Bayes factors range from 0 to ∞. The closer the Bayes factor is to 1, the more equally the two hypotheses predict the data. The Bayes factor is the ratio of the marginal likelihoods of both models (Eq. S3), the marginal likelihood being the probability of the observed data (D) given the hypothesis
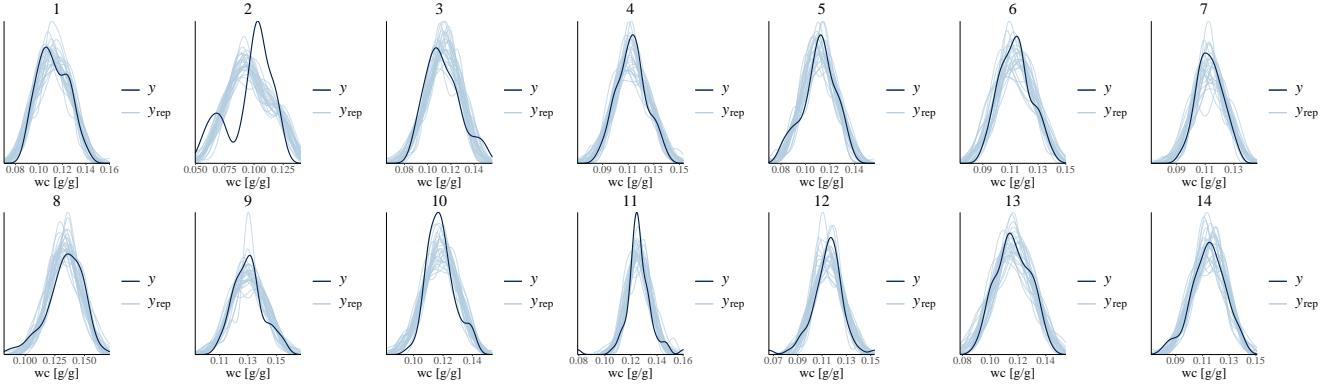
Figure S3. Kernel density estimates for 50 modeled data sets y_{rep} (blue curves) compared to Kernel density estimate of the observed data set y (dark curve) for each of the 14 participating laboratories.

model (H) (van Doorn et al., 2021). This marginal likelihood is estimated from samples of the posterior distribution via Bridge Sampling (see Gronau et al. (2017) for details).

$$\underbrace{\frac{p\left(H_1 \mid data\right)}{p\left(H_0 \mid data\right)}}_{\text{Posterior odds}} = \underbrace{\frac{p\left(data \mid H_1\right)}{p\left(data \mid H_0\right)}}_{\text{Bayes factor 10}} \times \underbrace{\frac{p\left(H_1\right)}{p\left(H_0\right)}}_{\text{Prior odds}} \tag{S3}$$

When an effect is assumed to be present, the magnitude of the effect is assessed by estimating the value of the Bayesian parameter from the posterior distribution with the 95% credible interval. The 95% Bayesian credible interval means that there is a 95% probability that the true value of the parameter lies within the interval given evidences supported by the prior and the data.

Acknowledgements. The rings for reference samples were provided by Royal Eijkelkamp.

References

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., and Riddell, A.: Stan: A probabilistic programming language, Journal of statistical software, 76, 1–32, https://doi.org/10.18637/jss.v076.i01, 2017.

Gabry, J., Simpson, D., Vehtari, A., Betancourt, M., and Gelman, A.: Visualization in Bayesian workflow, Journal of the Royal Statistical Society: Series A (Statistics in Society), 182, 389–402, https://doi.org/10.1111/rssa.12378, 2019.

Gelman, A.: Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper), Bayesian Analysis, 1, 515–534, https://doi.org/10.1214/06-BA117A, publisher: International Society for Bayesian Analysis, 2006.

Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., Leslie, D. S., Forster, J. J., Wagenmakers, E.-J., and Steingroever, H.: A Tutorial on Bridge Sampling, https://doi.org/10.48550/arXiv.1703.05984, 2017.

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., Etz, A., Evans, N. J., Gronau, Q. F., Haaf, J. M., Hinne, M., Kucharský, S., Ly, A., Marsman, M., Matzke, D., Gupta, A. R. K. N., Sarafoglou, A., Stefan, A., Voelkel, J. G., and Wagenmakers, E.-J.: The JASP guidelines for conducting and reporting a Bayesian analysis, Psychonomic Bulletin & Review, 28, 813–826, https://doi.org/10.3758/s13423-020-01798-5, 2021.

van Erp, S., Mulder, J., and Oberski, D. L.: Prior sensitivity analysis in default Bayesian structural equation modeling, Psychological Methods, 23, 363–388, https://doi.org/10.1037/met0000162, 2018.