

Does vadose zone flow forecasting depend on the type of calibration data?

Wöhling, Th.¹

¹ *Lincoln Ventures Ltd, Lincoln Environmental Research Division, Ruakura Research Centre,
Hamilton, New Zealand*
Email: woehling@lvtlham.lincoln.ac.nz

Unsaturated subsurface water flow is often described by a flow model which is calibrated on either observed soil water content or tensiometric pressure head measurements. For a given model structure the calibration on one data type may lead to significant errors in predictions of the other data type. These errors are difficult to quantify since simultaneous measurements of pressure head and water content are generally not available. Independent vadose zone data of both types were recorded at an intensively investigated experimental field site in the Lake Taupo catchment, New Zealand. A numerical flow model was set up and calibrated (i) using tensiometric pressure head observations, (ii) using soil water content (TDR) observations, and (iii) using both tensiometric and TDR data. The global multi-method search algorithm AMALGAM was used to estimate five soil hydraulic parameters in five model layers, totaling 25 optimized parameters. In the cases (i) and (ii), a single aggregated objective function was defined to fit measurements from four different depths in the vadose zone profile. The third model calibration was placed in a multi-objective context to include the two different data types simultaneously. The trade-off pattern between the fit to the water content and pressure head observations was investigated. Parameter sets from the three calibrations were then used for predicting pressure heads and water content in the vadose zone for independent data, not previously used in the calibration process. The results suggest that predictions of tensiometric pressure head and volumetric water content significantly depend on the type of data used for model calibration. Large differences in the model predictions occur when calibrating to one data type and predicting the other. This demonstrates the need to inform the model about the required prediction data type in the calibration process. This is a prerequisite to make reliable forecasts of vadose zone water flow and to determine realistic uncertainty bounds in vadose zone flow modeling.

Keywords: *Inverse Model calibration, Vadose Zone Modeling, Calibration Data, Global Optimization*

1. INTRODUCTION

The recent development of powerful automatic model calibration algorithms apparently allows the modeler to accurately fit models with increasing complexity and parameter numbers to environmental data (e.g. Vrugt et al. 2008). In the process, errors in both the observation data and the model structure are lumped in “effective” parameter values. However, when the parameter range of measurable parameters can not be constrained by prior information, then these parameters may lose their physical meaning in the calibration. Nevertheless, the calibrated model may be used successfully for predictions if the data type of the predictions was included in the calibration and if the uncertainty of the parameter values is adequately addressed. On the contrary the predictions can be significantly biased when calibrating to one data type and predicting another. This is well demonstrated in this paper for the application of the unsaturated-saturated flow model HYDRUS-1D to a volcanic vadose zone, which is calibrated using

1. tensiometric pressure head observations at four depths
2. volumetric water contents measured at the same depths, and
3. both the pressure head and water content data sets.

This study focuses on the data type used for model calibration and the implications when using the model to predict a data type which was not included in the calibration. The estimation of the parameter uncertainty is addressed in a separate study and is therefore not included in this paper.

2. MATERIALS AND METHODS

2.1. Field Data

The field data originates from the Spydia experimental site in the northern Lake Taupo catchment, New Zealand. Simultaneous measurements of tensiometric pressure heads and volumetric water content were conducted at five different depths through the vadose zone. The vadose zone materials at Spydia encompass a young volcanic soil (0 - 1.6 m depth), unwelded Taupo Ignimbrite (TI, 1.6 - 4.2 m), two older buried Palaeosols layers (PS, 4.2 to 5.8 m depth) and unwelded Oruanui Ignimbrite (OI) below.

Tensiometer (type UMS T4e, Germany, accuracy 0.05 m) and TDR probes were installed at the 0.4, 1.0, 2.6, 4.2, and 5.1 m depths (and at three locations at each depth). The 3-rod TDR probes (0.21 m long) were manufactured in-house and calibrated in the laboratory using calibration cells of in-situ vadose zone material from the installation depths (Stenger et al., 2007). Tensiometer and TDR probes are installed in pairs at each site having a horizontal separation distance of about 0.15 m. The probes are installed horizontally from a cylindrical access caisson (7 m height, 2.3 m diameter) and measurements are recorded by a National Instruments FieldPoint cFP2010 controller at 15 min intervals.

Daily values of potential evaporation were calculated by the Penman-Monteith equation (Allen et al., 1998) using data from the nearby Waihora meteorological station (500 m distance). Precipitation was recorded event based on site using a 0.2 mm bucket gauge and upscaled to hourly values for use in our calculations. A detailed description of the Spydia experimental data and the setup of the experiment can be found in Wöhling et al. (2008) and Barkle et al. (2009) and is therefore not repeated here.

2.2. Vadose Zone Model

The HYDRUS-1D model (Šimůnek et al., 2005) is used to simulate vertical water flow in the Spydia vadose zone. HYDRUS-1D utilizes the Galerkin finite element method based on the mass conservative iterative scheme proposed by Celia et al. (1990). The model solves the one-dimensional Richards equation:

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left(K \left(\frac{\partial h}{\partial z} + 1 \right) \right) - S \quad (1)$$

where θ is the volumetric water content [$L^3 L^{-3}$], t represents time [T], z is the vertical coordinate (positive upward) [L], h denotes the pressure head [L], K is the unsaturated hydraulic conductivity function [LT^{-1}], and S is a sink term representing processes such as plant water uptake [$L^3 L^{-3} T^{-1}$]. The model to describe the soil hydraulic properties is the Mualem-van Genuchten (1980) model:

$$\theta(h) = \begin{cases} \theta_r + \frac{1}{[1 + |\alpha h|^n]^m} & h < h_s \\ 1 & h \geq h_s \end{cases} \quad (2)$$

$$K(S_e) = K_s S_e^l \left[1 - (1 - S_e^{1/m})^m \right]^2 \quad (3)$$

where $S_e = (\theta - \theta_r) / (\theta_s - \theta_r)$ is the effective water content, θ_r and θ_s denote the residual and saturated water content [$L^3 L^{-3}$], respectively, α [L^{-1}] and n [-] are parameters that define the shape of the water retention function, K_s represents the saturated hydraulic conductivity [LT^{-1}], l is the pore-connectivity parameter. Further, it is assumed that $m = 1 - 1/n$, $n > 1$, and $h_s = -0.02$ m. The initial and boundary conditions used to solve Eq. (1) are:

$$\begin{aligned} h(z, t) &= h_i(z) & \text{at } t = 0 \\ h(z, t) &= h_L(t) & \text{at } z = L \end{aligned} \quad (4)$$

and

$$\begin{aligned} -K \left(\frac{\partial h}{\partial z} + 1 \right) &= q_0(t) - \frac{\partial h}{\partial t} & \text{at } z = 0, \text{ for } h_A \leq h \leq h_s \\ h(0, t) &= h_A & \text{for } h < h_A \\ h(0, t) &= h_s & \text{for } h > h_s \end{aligned} \quad (5)$$

where $h_i(z)$ is the initial pressure head derived from linear interpolation of observed tensions at the 0.4, 1.0, 2.6, 4.2 and 5.1 m depths, $h_L(t)$ represents the prescribed (observed) pressure head at the bottom boundary $L = -5.1$ m (depth of the model is 5.1 m), $q_0(t)$ is net infiltration rate (i.e. precipitation minus evaporation) and h_A and h_s denote minimum and maximum pressure head allowed at the soil surface.

The vadose zone stratigraphy (c.f. Wöhling et al., 2008) is represented by five layers in the HYDRUS-1D model. The first three layers represent the Ap (0-0.1 m), B (0.1 – 0.7 m), and C (0.7 – 2.2) horizons of the modern soil, respectively, whereas the fourth and fifth layers represent the Taupo Ignimbrite (2.2 – 4.2 m) and the two Palaeosol layers (4.2 – 5.1 m), respectively. The simulations are set up for a 198 day calibration period (starting April 11, 2006) and a 67 day evaluation period (21 June, 2008 to 27 August, 2008). The initial conditions are described by Eq. (4). The initial pressure heads for the calibration/evaluation periods were -0.41/-0.87, -1.35/-1.21, -1.18/-1.12, -0.85/-1.00, and -0.44/-0.96 m at the 0.4, 1.0, 2.6, 4.2, and 5.1 m depths, respectively.

Eq. (5) describes the atmospheric boundary condition at the soil-air interface (Šimůnek et al., 1996). Infiltration-excess overland flow is neglected and the limit $h_A = -200$ m is used. The plant water uptake, S in Eq. (1), is simulated with the Feddes model (1978) using HYDRUS-1D default parameters for grass and a rooting depth of 0.35 m. For more details regarding the model setup, refer to Wöhling et al. (2008).

2.3. Model Calibration

The soil hydraulic functions as described by Eqs. (2) and (3) require a number of different parameters to be specified for each of the model layers. These parameters are estimated by an inverse modeling procedure that aims to find the best attainable fit between model predictions and corresponding observations of tensiometric pressure head and/or volumetric water content. Three formulations of the objective function are utilized in this study:

$$\min F_1(u) \quad (6)$$

$$\min F_2(u) \quad (7)$$

$$\min F(u) = \begin{bmatrix} F_1(u) \\ F_2(u) \end{bmatrix} \quad (8)$$

where $F_1(u)$ and $F_2(u)$ are the sum of root-mean square errors (RMSE, e.g. Hall, 2001) of the fit between the simulated and observed pressure heads and volumetric water contents at the 0.4, 1.0, 2.6, and 4.1 m depths, respectively, and u is a vector of 25 model parameters to be optimized. Assuming $\theta_r = 0$, five parameters are estimated for each vadose zone layer: θ_s , α , n , K_s , and l .

The model is calibrated using (a) pressure head data (Eq. 6), (b) volumetric water content data (Eq. 7), and (c) both pressure head and water content data simultaneously (Eq. 8). The three different calibrations are subsequently referred to as C1, C2 and C3 runs. Note that C1 and C2 are single-objective optimizations, whereas C3 is posed in a multi-objective framework. To solve the optimization problems, the global search algorithm AMALGAM is utilized, which was previously found to be the most efficient algorithm out of three multi-objective search algorithms tested in a previous study (Wöhling et al., 2008). AMALGAM combines two

concepts, simultaneous multi-method search, and self-adaptive offspring creation, to ensure a reliable and computationally efficient solution to multi-objective optimization problems. The sole algorithmic parameter to be defined by the user is the population size which was set to $s = 100$. To create the initial sample to be iteratively improved with AMALGAM, uniform sampling of the prior defined parameter space is employed. This space is defined by lower and upper parameter bounds: $\theta_s = 0.3 - 0.7$ [$\text{m}^3 \text{m}^{-3}$]; $\alpha = 1 - 20$ [m^{-1}]; $n = 1.1 - 9.0$; $K_s = 1\text{E}^{-7} - 1\text{E}^{-3}$ [m s^{-1}]; $l = -3 - 3$. The optimization runs C1 and C2 were terminated after 50,000 HYDRUS-1D model evaluations, whereas the C3 run was terminated after 100,000 evaluations. A detailed description of AMALGAM has been presented in Vrugt and Robinson (2007) and Wöhling et al. (2008). The single-objective optimizations C1 and C2 result in a single best fitting solution. To account for solutions with similar performance but in possibly different locations of the parameter space, additional parameter sets are selected that are within 10 percent of the lowest aggregated RMSE value.

The multi-objective optimization C3 results in a set of Pareto optimal solutions that represent trade-offs between the two different objectives having the property that moving from one solution to another results in the improvement of one objective while causing deterioration in others (e.g. Gupta et al., 1998). The Pareto efficient solutions are selected for the analysis of C3. The solution of C1 and C2 represent Pareto extremes of the Pareto front derived in C3 (Wöhling et al., 2008).

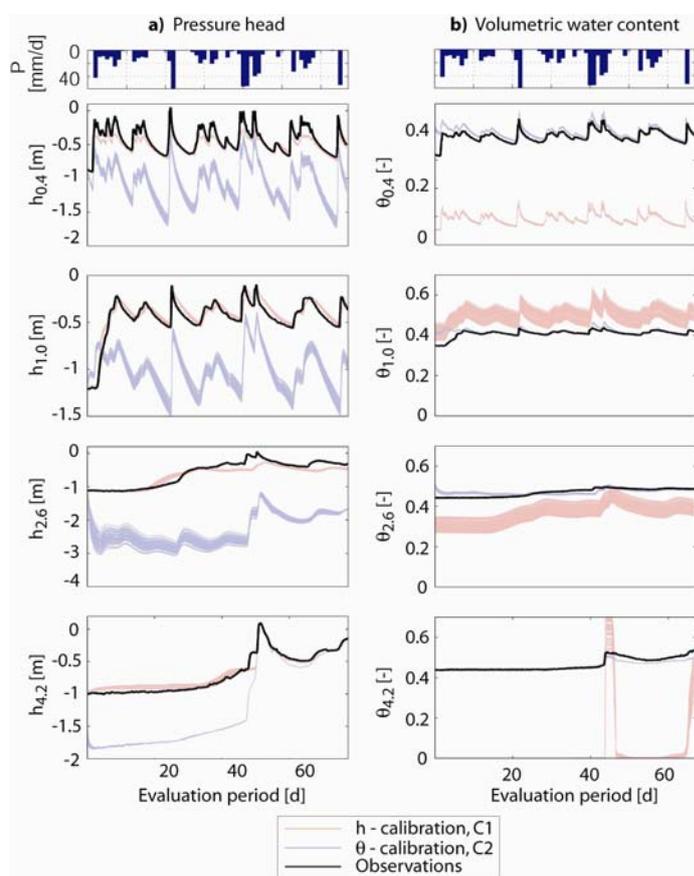


Figure 1. Predictions of (a) pressure head and (b) corresponding volumetric water content at the 0.4, 1.0, 2.6, and 4.2 m depths for the evaluation period using best fitting parameter sets of HYDRUS-1D calibrated on pressure head data (pink) and volumetric water content data (purple).

0.11/0.07/0.16/0.08 for the individual depths. However the dynamic pressure head is generally well described by the model. In contrast, the corresponding predictions of volumetric water content exhibit large discrepancies compared to the observations as shown in Figure 1b. The RMSE of fit to the θ -observations is on average 24.2% for the individual depths and C_e attains large negative values indicating strong bias. The predictions cover a wide range of θ -values at the 1.0 and 2.6 m depths whereas pressure head predictions are generally confined to a narrow range.

Two other criteria are used to measure the fit between observed and simulated tensiometric data of the selected Pareto solutions: the coefficient of determination R^2 and the coefficient of efficiency C_e by Nash-Sutcliffe (ASCE, 1993). C_e may resolve negative numbers if the mean square error exceeds the variance of the observations (Hall, 2001). Model predictions are considered satisfactory if both the values of R^2 and C_e assume values close to unity.

3. RESULTS

3.1. Single-objective model calibration

HYDRUS-1D was calibrated in run C1 to 198 days of pressure head observations at four different depths. The model is well calibrated by the best fitting solution exhibiting the smallest RMSE value of 0.24 m. This was confirmed by individual RMSE - values of 0.07/0.06/0.07/0.04 m at the 0.4/1.0/2.6/ and 4.2 m depths, respectively. The R^2 and C_e criteria attained values greater than 0.87. The pink lines in Figure 1a show predictions of pressure head at the four depths during the 67 day evaluation period using independent data and the parameter sets within 10% of the RMSE of the best C1 solution. The model to measurement misfit is slightly larger during evaluation with RMSE values of

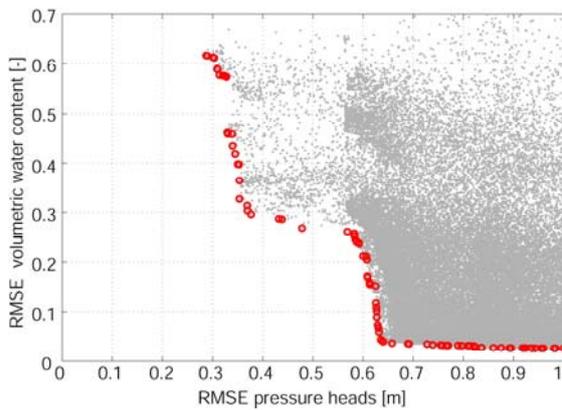


Figure 2. Pareto efficient solutions (red) of the 2D objective space for optimization run C3.

In the optimization run C2, HYDRUS-1D was calibrated to water content data of the calibration period. The model was well calibrated by the best solution ($RMSE = 0.022 \text{ m}^3 \text{ m}^{-3}$, or 2.2 %) which is confirmed by RMSE values of 0.6/0.9/0.5/0.2 % water content and relative large R^2 and C_e values. The purple lines in Figure 1b show that the predictions of volumetric water content during the evaluation period match closely the observations. Again, the parameter sets within 10% of the best aggregated RMSE value are also shown. The RMSE values for the best fitting parameter set are smaller than 2% at the individual depths for the evaluation period. In contrast, large discrepancies were obtained for the corresponding pressure head forecasts as confirmed by RMSE values greater than 0.84 m at the individual depths (Figure 1a).

These results show clearly that a good model fit to a particular data type in the calibration process does not necessarily result in the same good fit in the predictions of another data type. Moreover, large errors in the water contents forecasts can be expected when calibrating flow models on pressure heads and vice versa. If the objective of the modeling study is the simultaneous description of both pressure head and water content (e.g. an accurate estimation of water retention), then the model must be informed about both data types during calibration as demonstrated in the following section.

3.2. Multi-objective model calibration

Both water content and pressure head data was included in the calibration in the multi-objective optimization run C3. Each grey dot in Figure 2 represents one of the 120,000 model evaluations and the red circles indicate the Pareto efficient solutions. A large trade-off exists between the RMSEs of pressure heads (objective 1) and water content (objective 2) with two distinct steps along the Pareto front. The Pareto extremes, i.e. the smallest objective function values, were 0.287 m and $0.022 \text{ m}^3 \text{ m}^{-3}$ for objective 1 and 2, respectively, and correspond well to the single-objective solutions. The sampling density in the vicinity of the objective 1 Pareto extreme is much less as compared to other objective space regions which suggest preferred sampling in the other spaces. However, the expected change in the shape of the Pareto front is believed to be marginal compared to the cost of additional computations.

It can be deduced from the shape of the Pareto front that no parameter set exists which satisfies both objectives equally well. A normalization of the objective function values could have shown this even more clearly but this is not pursued in this study. A total of 138 Pareto solutions were identified and subsequently used to simulate pressure head and volumetric water content during the evaluation pe-

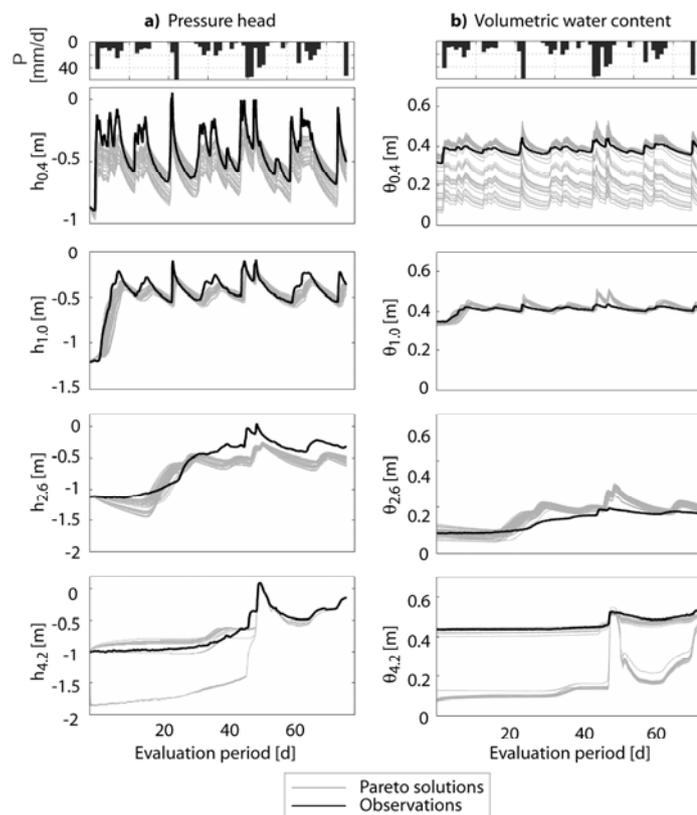


Figure 3. Predictions of (a) pressure head and (b) corresponding volumetric water content at the 0.4, 1.0, 2.6, and 4.2 m depths for the evaluation period using Pareto efficient parameter sets of HYDRUS-1D calibrated simultaneously on pressure head and volumetric water content data.

riod. The corresponding time series are shown in Figure 3. The predictions encompass the observations much better than the predictions of the single data type calibration runs C1 and C2. However, the range of predictions is large for the water contents at the 0.4 m depth (Figure 3b). One explanation for this result is inconsistency in the data. Experimental work has shown that signals of tensiometric pressure and water content changes travel through porous media at different speed. Furthermore, tensiometer and TDR measurements have different zones of influence which depend on the transient system state. Experimental errors also may originate from the calibration of the TDR probes, which was conducted in the laboratory since in-situ calibration was not feasible. Model structural errors are another explanation for the discrepancies. The assumption of uniform flow may be inaccurate for water transport in the active root zone where root channels and biological activity leads to preferential flow paths. These processes in turn can also lead to local variability in the measured state variables, which are then lumped to average conditions in the model.

The model simulations at the 4.2 m depth form two bands, one of which encompasses the observations well and one which significantly underestimates pressure heads or water content (Figure 3). The simulations underestimating pressure heads correspond to parameter sets which are located close to the Pareto extreme for C2 and, vice versa, the simulations underestimating water content correspond to parameter sets which are located close to the Pareto extreme for C1. But the trade-off between water content and pressure head data is relatively small at the 1.0 m and, to a lesser degree, at the 2.6 m depths. These results show that the inclusion of both pressure head and water content data in the calibration significantly improves the overall fit to both data types, although discrepancies remain.

The cumulative recharge flux at the 4.2 m depth during the evaluation period was calculated using the C3 Pareto parameter sets to $0.378 \text{ m}^3 \pm 7\% \text{ CV}$. The relatively small variation is a positive outcome which is, however, partly caused by the definition of the model's lower boundary condition.

3.3. Parameter estimates

The parameter uncertainty of the optimization runs is not formally addressed in this study. The simulated pressure head and water content bands shown in Fig. 1 merely represent results attained with parameter combinations which all calibrate the model relatively well. It is interesting, however, to investigate whether the parameter estimates are unique for the different optimization problems under consideration. Figure 4 shows normalized histograms for the “best fit” parameter combinations of the C1 and C2 runs and the 25 model parameters as well as histograms for the Pareto efficient parameters of run C3. The red and blue dots indicate the parameter value of the overall best (minimum RMSE) solutions of C1 and C2, respectively. In many cases, these values are different for C1 and C2. In some cases they are at opposite ends of the parameter range (θ_2 , α_1 , $K_{s,4}$), but in other cases they attain similar values (θ_4 , n_3 , n_4 , α_4 , $K_{s,1}$, l_1 , l_2). Most parameters exhibit a narrow range for C1 and C2, respectively, which indicates that the solutions are confined to the same area in the parameter space. Parameters exhibiting a broader range of values are typically also less sensitive to the simulation results. The green step function in the panels of Fig. 4 is the normalized histogram of the Pareto values (run C3). The parameter ranges are relatively narrow in most cases, but can differ from the parameter ranges of the single-objective runs. It is interesting to note that the best C1 and C2 solutions are not always included in the range of C3 parameter values.

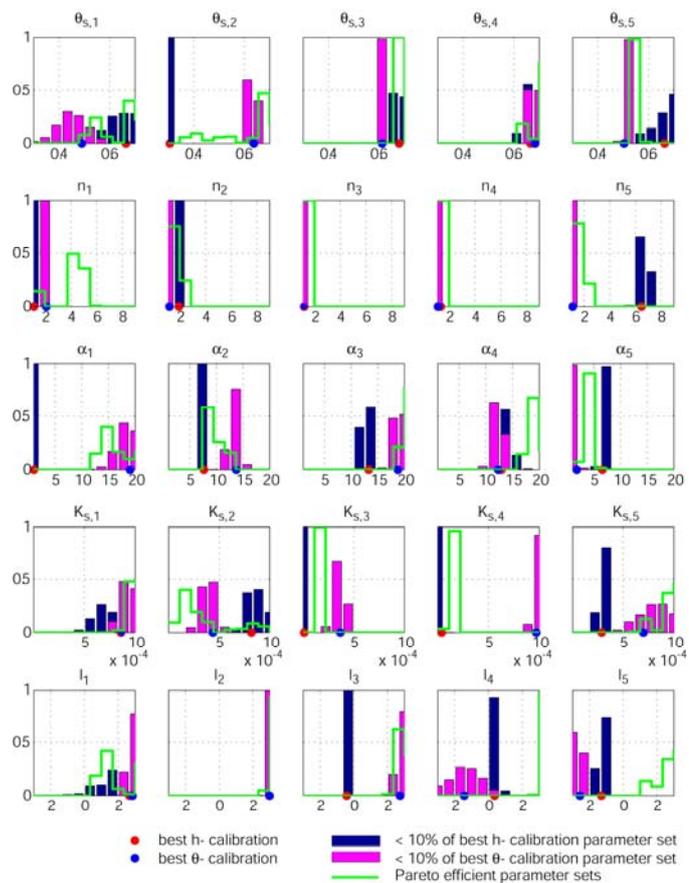


Figure 4. Histogram of best fitting parameters for calibration on pressure head data (blue), on volumetric water content data (magenta) and on both data sets (green).

4. SUMMARY AND CONCLUSIONS

Tensiometric pressure head and water content data were measured at the Spydia experimental site near Lake Taupo, New Zealand. One data type at a time was used to calibrate a vertical vadose zone model of the site using HYDRUS-1D, the global parameter estimation method AMALGAM, and an aggregated RMSE objective function. The simulations with the calibrated model match well to the observations of the calibration type during the evaluation period. When HYDRUS-1D was calibrated using pressure head data, the simulations agree well with independent pressure head measurements at the different depths in the vadose zone profile. A similar good fit is obtained for independent volumetric water content measurements, when the model was calibrated to water content data. However, in both cases a large model discrepancy existed for the data type which was not included in the calibration. This indicates that measurement uncertainties and possibly model structural deficiencies are lumped into the optimized, “efficient” parameter sets. Large errors can be expected when only one data type is calibrated and the other data type is evaluated. This corresponds to only one variable of the soil water retention function being conditioned on actual data. In essence, the results presented in the paper demonstrate that the data type of the predictions should be included in the model calibration for results to be more accurate.

The use of both data type in the calibration exercise was demonstrated by posing the calibration in a multi-objective context. It should be noted that expert knowledge is required for weighting the data types in the optimization scheme accordingly to the accuracy of the measurements. In the presented case study, the calibration on both data types resulted in a much better simultaneous fit to the observed independent pressure head and volumetric data, respectively. Furthermore, the analysis of the trade-off between the fit to the different data types is useful to identify possible discrepancies in the modeling approach and/or the data. The discrepancies are also evident in the different “efficient” parameter values when using different data types in the calibration process. However, parameter uncertainty estimation is not considered here, work on this topic is ongoing and will be subject of another paper.

Based on the findings of this study it is suggested that HYDRUS-1D simulations would not necessarily provide accurate estimates of water fluxes in the vadose zone (for example groundwater recharge) even if both water content and pressure heads were used for model calibration. Water flux data from automated tension plate lysimeters installed at the Spydia site will become available to test this hypothesis.

REFERENCES

- Allen, R.G., L.S. Pereira, D. Raes, and M. Smith (1998), Crop evapotranspiration. *FAO Irrig. Drain. Pap.* 56. FAO, Rome.
- Barkle, G. F., Th. Wöhling, R. Stenger, J. Mertens, B. Moorhead, A. Wall, and J. Claque (2009) Measuring water and contaminants fluxes throughout the vadose zone using Automated Equilibrium Tension Lysimeters (AETLs). *Vadose Zone Journal* (submitted).
- Celia, M.A., E.T. Bouloutas, and R.L. Zarba (1990), A general mass-conservative numerical solution for the unsaturated flow equation. *Water Resour. Res.*, 26(7):1483-1496.
- Feddes, R.A., P. Kowalik, and H. Zaradny (1978), Simulation of field water use and crop yield, *PUDOC*, Wageningen, Netherlands. ISBN 90-220-0676-X.
- Gupta, H.V., S. Sorooshian, and P.O. Yapo (1998), Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, *Water Resources Research*, 34(4):751-764.
- Hall, J.M. (2001), How well does your model fit the data? *Journal of Hydroinformatics*, 3(1):49-55.
- van Genuchten, M.Th. (1980), A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Science Society of America Journal*, 44(5):892-898.
- Šimůnek, J., M.Th. van Genuchten, and M. Šejna (2005), The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media. Version 3.0, Dep. of Env. Sci., University of California Riverside, Riverside, CA, 92521, USA.
- Stenger, R., Th. Wöhling, G.F. Barkle, and A. Wall (2007), Relationship between dielectric permittivity and water content for vadose zone materials of volcanic origin. *Australian J. of Soil Res.*, 45, 299-309.
- Vrugt, J.A., and B.A. Robinson (2007), Improved evolutionary optimization from genetically adaptive multi-method search, *In Proc. of the National Academy of Sci. of the U. S. A. (PNAS)*, 104, pp 708-711.
- Vrugt, J.A., P.H. Stauffer, Th. Wöhling, B.A. Robinson, and V.V. Vesselinov, (2008), Inverse modeling of subsurface flow and transport properties: A review with new developments. *Vadose Zone J*, 7, 843-64.
- Wöhling, Th., J.A. Vrugt, and G.F. Barkle, (2008) Comparison of three multiobjective algorithms for inverse modeling of vadose zone hydraulic properties. *Soil Science Society of America Journal*, 72, 305-319.