PERGAMON

Computers & Geosciences 26 (2000) 319–327

COMPUTERS & GEOSCIENCES

RETMCL: Incorporating maximum-likelihood estimation principles in the RETC soil hydraulic parameter estimation code ☆

K.J. Hollenbeck [a, b], J. Šimunek [c,*], M.Th van Genuchten [c]

[a] *Department of Hydrodynamics and Water Resources, Technical University of Denmark, Lyngby, Denmark*
[b] *Danish Hydraulic Institute, Agern Allé 5, DK-2970, Hørsholm, Denmark*
[c] *US Salinity Laboratory, USDA-ARS, Riverside, CA, USA*

Received 3 May 1999; received in revised form 13 August 1999; accepted 13 August 1999

Abstract

RETC is a public domain computer code for estimating parameters of the water retention curve and hydraulic conductivity functions of unsaturated soils. RETC was developed at the US Salinity Laboratory and is now used world-wide with thousands of copies distributed. Evaluation of the final estimation results in the code has been improved to yield a new version, RETCML, based on the maximum-likelihood theory for the special case of weighted least-squares estimators. This paper first explains the theory of maximum-likelihood and introduces the principles of model adequacy and parameter uncertainty on a formal basis. Next, this paper presents a user guide for the code. RETCML is also free and has been programmed to be almost fully compatible with the original RETC input files, thus making it easy to re-analyze previous data. The output of RETCML includes a thorough evaluation of estimation results. © 2000 Elsevier Science Ltd. All rights reserved.

Keywords: Maximum-likelihood; Estimation; Soil hydraulic properties

1. Introduction

The RETC (RETention Curve) code is a widely used computer program developed at the US Salinity Laboratory for estimating parameters of the retention curve and hydraulic conductivity functions of unsaturated soils (van Genuchten et al., 1991). While the retention curve (often also called the soil water characteristic curve) characterizes the energy status of the soil water, the unsaturated hydraulic conductivity function describes the ability of the porous medium to conduct water. The RETC program uses the parametric models of Brooks and Corey (1964) and van Genuchten (1980a) to represent the soil water retention curve, and the theoretical pore-size distribution models of Mualem (1976) and Burdine (1953) to either predict the unsaturated hydraulic conductivity function from observed soil water retention data or to use those data in the fitting procedure also. The van Genuchten retention function (van Genuchten, 1980a) has been very popular in the fields of soil physics and vadose zone

☆ Code available at http://www.iamg.org/CGEditor/index.htm and http://www.ussl.ars.usda.gov/models/models.html.

* Corresponding author. Tel.: +1-909-369-4865.

E-mail addresses: kjh@dhi.dk (K.J. Hollenbeck), jsimunek@ussr.ars.usda.gov (J. Šimunek).

PII: S0098-3004(99)00138-7

hydrology. The suggested S-shape function and its combination with the pore distribution model of Mualem (1976) is used not only for the description of the soil water characteristic curve (retention curve) but also for functions describing oil retention and transport properties.

The retention and hydraulic conductivity functions constitute properties needed for simulating flow and solute transport in the unsaturated zone. Computer models using analytical descriptions of soil hydraulic properties are now routinely used in both research and management to predict the movement of water and chemicals in the vadose zone between the soil surface and the groundwater table. Interest in the vadose zone has dramatically increased in recent years because of growing evidence that the quality of the subsurface environment is being adversely affected by industrial, municipal, and agricultural activities (van Genuchten et al., 1991).

The public domain RETC code, and its earlier version SOHYP (SOil HYdraulic Properties) (van Genuchten, 1978), have been freely distributed by several agencies, including the US Salinity Laboratory (USSL) of the US Department of Agriculture, Agriculture Research Service (USDA, ARS), the US Environmental Protection Agency (EPA), the International Groundwater Modeling Center (IGWMC), and others. Hundreds of copies of RETC have been distributed by USSL and IGWMC, thousands by EPA, while uncounted numbers of copies have been downloaded from internet sites of these and other agencies. Because of the wide distribution of RETC (and its descendant/successor programs), the implementation of the estimation problem in RETC, including the statistical analysis of estimated parameters, was carried over to numerous other programs, including CFITM (van Genuchten, 1980b), CFITIM (van Genuchten, 1981), ONESTEP (Kool et al., 1985), SFIT (Kool et al., 1987), MULTISTEP (van Dam et al., 1994), 3DADE (Leij and Bradford, 1994), CXTFIT (Toride et al., 1995), HYDRUS-1D (Šimunek et al., 1998) and HYDRUS-2D (Šimunek et al., 1996), and probably many others unknown to the authors. While these programs all correctly implement the estimation problem, the statistical analysis concerning the final estimated parameters is limited because of its imperfect implementation.

We here present an improved version of RETC, called RETCML (RETention Curve — Maximum-Likelihood), for which the uncertainty analysis is based on maximum-likelihood theory for the special case of weighted least-squares estimators. In RETCML, two theorems for evaluating the estimation results are exploited. The first theorem allows an evaluation of whether or not a selected parametric model is "adequate". Adequacy here signifies that the model predictions come so close to the data that any remaining discrepancy is no larger than measurement error in the underlying experiment. Given adequacy, the second theorem provides a basis for evaluating uncertainty in the estimated parameters as a function of measurement uncertainty. To benefit from RETCML's features, the user thus has to specify the measurement uncertainty in the input file. Users of all other above-mentioned programs may also want to implement the suggested changes in their codes in order to be able to more thoroughly evaluate the model adequacy and confidence intervals of optimized parameters in their particular applications.

2. About estimation

Parameter estimation refers to the problem where a variable y is observed as a function of a control variable x, and one has a corresponding model with parameters grouped in a vector $\mathbf{p}$. The optimal parameters are then generally (also in RETC and RETCML) found by minimizing a sum-of-squares (SSQ) objective function:

$$\mathrm{SSQ}(\mathbf{p}) = \sum_{i=1}^{N}\left[\frac{1}{\sigma_{y_i}}\left(y_{i,\,\mathrm{meas}}(x_i) - y_{i,\,\mathrm{model}}(x_i;\mathbf{p})\right)\right]^2 \qquad (1)$$

where the subscripts "meas" and "model" refer to measured and model predicted values of y, and N is the number of measurements. A "model" here means a particular parametric expression for the data, e.g., the van Genuchten retention expression

$$\psi(\theta;\mathbf{p}) = \frac{\left[\left(\dfrac{\theta-\theta_r}{\theta_s-\theta_r}\right)^{[n/(1-n)]} - 1\right]^{1/n}}{\alpha} \qquad (2)$$

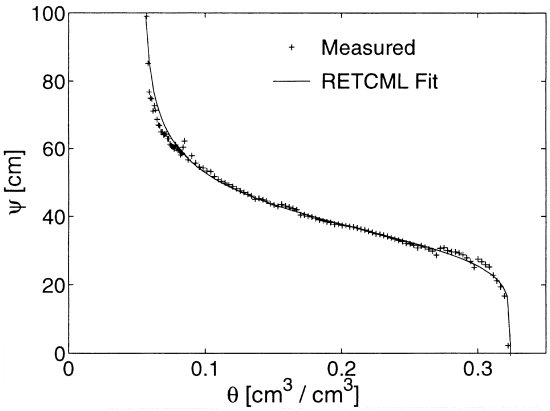

Fig. 1. Some example retention data and corresponding best-fit van Genuchten model.

with parameters $\mathbf{p}=(\theta_r, \theta_s, n, \alpha)$ or any subset thereof if some of these are known by other means. An example of retention data and the corresponding best-fit van Genuchten model is shown in Fig. 1. As this paper focuses on the estimation problem, the reader is referred to the literature for additional examples of hydraulic functions and properties (van Genuchten, 1980a).

A key feature of the above SSQ estimator are the $i = 1, \ldots, N$ weights $\sigma_{y_i}^{-1}$. As the notation implies, the weights must be set to reflect measurement error information about the data:

$$\frac{1}{\sigma_{y_i}} = \left.\frac{1}{\text{measurement error standard deviation}}\right|_i. \tag{3}$$

It is intuitive that information about the quality of the data should determine the quality of the parameter estimates, and that this information hence should be contained in the SSQ estimator. This intuitive principle has a formal mathematical representation. With weights set as given by Eq. (3), the weighted least squares (WLS) estimator Eq. (1) is in fact a maximum-likelihood (ML) estimator. The equivalence between the WLS and the more general ML estimator strictly holds only when all normalized residuals $r_i \equiv (y_i, \text{meas}(x_i) - y_i, \text{model}(x_i; \mathbf{p}))/\sigma_{y_i}$ are (i) independent and (ii) normally distributed. Condition (ii) is not really restrictive because the estimator is robust as long as the error distributions are symmetric and free of a bias and extreme outliers (Press et al., 1992, Chap. 15.6). While condition (i) is more difficult to meet in practice, the case of residuals with unknown correlation is in general difficult to treat, even by more advanced theories. Since this paper is not intended to discuss all aspects of ML estimation, the interested reader is referred to the literature (e.g., Bard, 1974; Beck and Arnold, 1977). Most importantly for practical applications, the error introduced by ignoring some residual correlation is much less severe than the error introduced by not setting the weights properly.

Once the weights are set as in Eq. (3), two theorems for evaluating estimation results can be exploited. The first theorem allows an evaluation of whether or not a selected parametric model is adequate. For adequate models, the second theorem provides a basis for evaluating uncertainty in the estimated parameters. The two theorems and their implementation in RETCML will be presented in the following. In addition, the differences between RETC and RETCML will be pointed out.

2.1. Model adequacy

An adequate model is one that explains the data to such a degree that it is plausible that remaining discrepancies are due only to measurement error noise. In other words for the best fit, the absolute values of the deviations $d_i \equiv (y_i, \text{meas}(x_i) - y_i, \text{model}(x_i; \mathbf{p}))$ should be about as large as the measurement error standard deviations σ_{y_i}, yielding normalized residuals each with $|r_i| \approx 1$, thus $r_i^2 \approx 1$. Furthermore, the SSQ estimator is a sum of N squared residuals r_i^2, so if these are all independent, it can be expected that at best, SSQ $\approx N$. In addition, it is known that a sum of squared random variables that each follow a normal distribution has a chi-square distribution.

The preceding intuitive argument may elucidate the principle behind the formal statistical test for model adequacy. This test is based on the minimum SSQ value, $\text{SSQ}_{\min}$, attained for the optimal parameters $\mathbf{p}_{\text{opt}}$, which is distributed as (Press et al., 1992, Chap. 15.1):

$$\text{SSQ}_{\min} \sim \chi^2_{N-P}. \tag{4}$$

The number of degrees-of-freedom is $N-P$ and not N, because there is a loss of one degree-of-freedom for each of the P parameters.

The expected value (the mean) of a χ^2 function is its number of degrees-of-freedom (Abramowitz and Stegun, 1965 , Eq. 26.4.33):

$$E[\chi^2_{N-P}] = N - P. \tag{5}$$

Hence, from Eq. (4), it can be expected that $\text{SSQ}_{\min} = N-P$. As any estimation problem is best defined when the number of data significantly exceeds that of the number of parameters, a good rule of thumb is that for an adequate model, $\text{SSQ}_{\min}$ should be about the same as the number of data — just as was expected from the introductory intuitive argument.

If the model is inadequate, however (e.g., the situ-

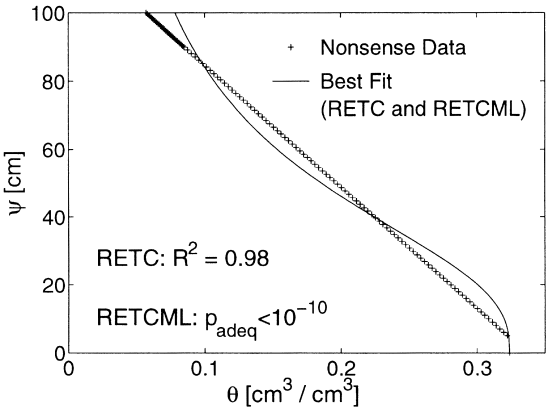

Fig. 2. "Fitting" S-shaped van Genuchten retention curve to synthetic data representing straight line. R^2 value is high even for this inadequate model, whereas p_{adeq} correctly reveals misfit.

ation shown in Fig. 2), SSQ_{min} will be significantly larger than $N-P$. This is so because the model-data deviations d_i will (by absolute value) on average be larger than the measurement errors σ_{y_i}; hence, the squared residuals will on average be larger than 1.

For any observed value of SSQ_{min}, the probability of model adequacy is:

$$p_{adeq} = 1 - Q(SSQ_{min}, N - P) \tag{6}$$

where $Q(\cdot)$ is the χ^2 cumulative density function.

The probability of model adequacy, p_{adeq}, in practical applications is relatively sensitive to outliers in the data. Even a single outlier, giving one large squared residual $r^2_{outlier}$, can result in a high SSQ_{min}. Hence, a rare relatively low p_{adeq} is not necessarily reason for concern.

More often, the reason for a low p_{adeq} is the assignment of too high weights $\sigma_{y_i}^{-1}$, as the SSQ grows with the square of the weights. Too high weights $\sigma_{y_i}^{-1}$ result if the user's specification for the measurement errors σ_{y_i} is too optimistic (too small). Conversely, if these errors are specified too pessimistically (too large), SSQ_{min} will be very small, resulting in an often unrealistic probability of model adequacy essentially equal to 1.

Also the original RETC allows for specification of the weights for the data. On p. 35 in the manual (van Genuchten et al., 1991), the recommendation of setting the weights as inverses of observation errors is given, but its importance is not particularly emphasized. Accordingly, the probability of model adequacy is not computed in RETC, but only the correlation coefficient R^2 between model predictions and data. However, the R^2 value does not follow any distribution, so one cannot evaluate quantitatively *how* good a particular value is. In general, the R^2 values found with RETC are all very close to 1, even for the clearly inadequate model.

The advantage of examining model adequacy rather than a correlation coefficient is illustrated for a "nonsense" example (Fig. 2). The synthetic data used for estimating the van Genuchten retention parameters were set to describe a straight line, which is not a retention curve one can observe in nature. Nevertheless, the R^2 value output by RETC is 0.98, close to the perfect value of 1. In contrast, the probability of model adequacy, p_{adeq}, as computed in RETCML, unequivocally reveals the poor fit. In general, "good" fits are correctly indicated by both R^2 and p_{adeq} (for the fit in Fig. 1, $R^2 = 0.99710$ and $p_{adeq} > 0.99999$). Apparently, the statistic R^2 thus has little power for discriminating between good and poor fits. (Note that measurement error standard deviations were taken to be as large as the crosshairs on the data in Fig. 2 when calculating p_{adeq}.)

2.2. Parameter confidence regions

A confidence region for parameter estimates is defined via a maximum allowable change in the SSQ from its minimum value. The relevant theorem has been implemented in RETCML, but only incorrectly so in the original RETC.

From theory (Press et al., 1992, Chap. 15.6), it is known that the increment $\Delta SSQ(\mathbf{p}) = SSQ(\mathbf{p}) - SSQ_{min}$ follows a chi-square distribution with P degrees-of-freedom:

$$\Delta SSQ(\mathbf{p}) \sim \chi^2_P. \tag{7}$$

Thus, the appropriate value for ΔSSQ_{conf} for a desired level of confidence p_{conf} is:

$$\Delta SSQ_{conf} = Q^{-1}(p_{conf}, P) \tag{8}$$

where $Q^{-1}(p_{conf}, P)$ is the inverse χ^2 cumulative density function.

In practice it would be difficult to compute the SSQ for "all" parameter vectors $\mathbf{p}$ and determine for each one whether $\Delta SSQ(\mathbf{p})$ is either larger or smaller than the allowed ΔSSQ_{conf} value. (Although contouring the SSQ is often a method for analyzing how well-defined estimation problems are (e.g., Toorman et al., 1992).) Fortunately, an approximate method exists based on the Cramer–Rao theorem (Schweppe, 1973, Chap. 12). This theorem gives *a lower bound* for the parameter covariance matrix $\mathbf{\Sigma_p}$ as:

$$\mathbf{\Sigma_p} \geq \mathbf{H}^{-1} \tag{9}$$

where $\mathbf{H}$ is the Hessian matrix. $\mathbf{H}$ can be computed from the sensitivity of the model to its parameters, i.e., from the partial derivatives of the objective function (Press et al., 1992, Chap. 15.5), as

$$\mathbf{H}_{kl} = \sum_{i=1}^{N} \frac{1}{\sigma^2_{y,\,i}} \left[\frac{\partial y_i(\mathbf{p})}{\partial p_k} \, \frac{\partial y_i(\mathbf{p})}{\partial p_l} \right] \quad k, l = 1, \ldots, P. \tag{10}$$

With the parameter covariance matrix known, a geometric description of the confidence region is given through:

$$(\mathbf{p} - \mathbf{p}_{opt})' \mathbf{\Sigma}_{\mathbf{p}}^{-1} (\mathbf{p} - \mathbf{p}_{opt}) = \Delta SSQ_{conf}. \tag{11}$$

Eq. (11) describes an ellipsoid in the parameter space, and all parameter vectors $\mathbf{p}$ within the ellipsoid are part of the confidence region.

Eqs. (10) and (9) suggest that smaller weights imply a larger (less desirable) confidence region. This makes sense because smaller weights express larger measurement errors. Now recall that smaller weights also imply a higher (more desirable) probability of model adequacy. Hence, the magnitudes of the weights

impact the two measures of the quality of the estimation result in opposite directions. Inadequate models can have attractively small parameter confidence regions. The user should not be tempted to set the weights too large to achieve a small confidence region, but at the expense of an unrealistically small probability of model adequacy.

A multivariate confidence region in P dimensions is difficult to visualize, but pseudo-univariate "parameter standard errors" can be taken from the diagonal of the parameter covariance matrix as:

$$\sigma_{p,\,i} = \sqrt{\mathbf{\Sigma}_{\mathbf{p},\,ii}} \quad i = 1,\,\ldots,\,P. \tag{12}$$

A conservative way to find the actual confidence interval for a parameter is to find the projections of the confidence ellipsoid on the parameter axes. The half-length of the symmetric confidence interval for the i-th parameter is (Press et al., 1992, Eq. 15.6.4):

$$c_{p,\,i} = \sqrt{Q^{-1}(p_{\mathrm{conf}}, P)}\,\sigma_{p,\,i}. \tag{13}$$

An example of a confidence ellipse and the projected confidence intervals computed using Eq. (13) is shown in Fig. 3. This schematic example is for two generic parameters here called "p_1" and "p_2".

It is important to be aware of the approximate nature of the Cramer–Rao theorem when choosing a confidence level for the parameters. Only for linear models is the confidence region an ellipsoid, with its size proportional to the norm of the covariance matrix. The uncertainty region for nonlinear models is often odd-shaped, and larger than the Cramer–Rao ellipsoid. The discrepancy grows the higher the confidence level, i.e., the larger the region. For example, while none of the van Genuchten parameters can be negative, for a sufficiently large desired confidence level (a sufficiently large multiplier $\sqrt{Q^{-1}(p_{\mathrm{conf}}, P)}$), the Cramer–Rao ellipsoid will grow into the negative parts of parameter space. Hence, it is not advisable to "stretch" the theorem by computing confidence limits at too high a level. With this in mind, RETCML is coded to output 90% rather than 95% (RETC) confidence limits. Exact and Cramer–Rao confidence regions for the soil hydraulic data are shown in Hollenbeck and Jensen (1998, Fig. 4) . The approximation can be judged admissible.

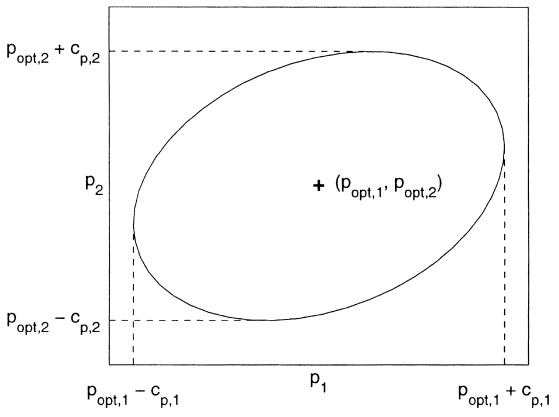

Fig. 3. Schematic representation of two-dimensional confidence ellipse with projections delineating confidence intervals for parameters "p_1" and "p_2" around optimal (estimated) parameters.

In RETC, univariate parameter confidence half-lengths are apparently computed incorrectly. The formula used in the code is:

$$c_{p,\,i} = t^{-1}(p_{\mathrm{conf}}, N-P)\sqrt{\frac{\mathrm{SSQ}_{\min}}{N-P}}\,\sigma_{p,\,i} \quad i = 1,\,\ldots,\,P \tag{14}$$

where $t^{-1}(p_{\mathrm{conf}},\,N-P)$ is the inverse cumulative density of the t distribution. With this formulation, the weight factors almost cancel out, because as seen from Eqs. (9), (10), and (12), the squared weights are contained in the parameter standard error $\sigma_{p,i}$, while from Eq. (1), their reciprocals are contained in $\sqrt{\mathrm{SSQ}_{\min}}$. The net result is that different multiples of weights result in the same computed standard errors. Considering the definitions more closely, the weights enter as multipliers for the deviations, which as a product are squared in $\mathrm{SSQ}_{\min}$, while in the Hessian, the weights themselves are squared. Hence, the two sums defining $\mathrm{SSQ}_{\min}$ and $\mathbf{H}_{kl}^{-1}$ are not exactly reciprocals of each other when not all weights are scaled by the same factor. The use of different weights within a data set (i.e., not a different multiple for all) will impact the estimates to a limited extent. For example, in RETC one can assign larger weights for "wet" retention observations and smaller ones for some of the "dry" data, thus placing more emphasis on the "wet" data as compared to the "dry" data. The overall magnitude of the estimation error, however, will be determined only by $N-P$. Recalling the discussion of model adequacy, the estimation errors as computed from RETC are always about equal to that of an adequate model, irrespective of measurement uncertainty [cf Eq. (5)].

RETC makes use of t statistics in two instances. Comparing RETC's Eq. (14) with RETCML's Eq. (13), it can be seen that the scale factor for the projections of the parameter confidence interval differs. RETC's t_{N-P}^{-1} statistic is only correct for a univariate estimation problem, but not for the generally encountered multivariate case. Furthermore, RETC's confidence analysis output contains the t value for testing the hypothesis H: $p_i = 0$. It is not obvious why one should test whether a particular parameter is equal to zero (some, such as van Genuchten's n, are restricted

to be greater than zero or even one, depending upon the selected soil-hydraulic parametric model). Hence, the t values have not been carried over from RETC's to RETCML's output.

As a closing remark for this Section, it should be mentioned that maximum-likelihood theorems for evaluating estimation results are strictly valid only if the number of data clearly exceeds the number of parameters to be estimated. This restriction does not constrain RETCML relative to RETC, because as seen in Eq. (14), also RETC makes reference to a maximum-likelihood theorem (the Cramer–Rao theorem). A "reasonably" large number of data is required for solving any estimation problem in a meaningful way.

2.3. Assigning proper weights

Although Eq. (3) is a formally correct and complete guideline on how to set weight factors in RETCML, the user might encounter situations where more detailed directions can be helpful. These are given in this Section.

2.3.1. Errors in both x and y or in x only

According to Eq. (3), the SSQ estimator Eq. (1), allows for specifications of errors in y only, when y is a function of x. In practice, however, data may sometimes be available in the form $x(y)$, with errors known for x. For example, in RETC and RETCML, the retention curve is fitted as $\theta(\psi)$, not $\psi(\theta)$. A related situation often encountered in practice is that instrument errors affect not only the measured variable y, but also the independent variable x.

To still use RETC and RETCML in the case of errors in both y and x, or only in x, we suggest the following approach. One can compute a lumped total error in y as:

$$\sigma_{y_{i,\,\mathrm{total}}} = \sqrt{\sigma_{y_i}^2 + \left(\left.\frac{\partial y}{\partial x}\right|_i \sigma_{x_i}\right)^2}. \tag{15}$$

The first term, σ_{y_i}, is the error in y itself, and the second represents the change/error in y caused by a random change/error in x. For the unusual case of error-free y data, the second term under the square root guarantees a non-zero lumped error (otherwise the weight would be "$1/0=\infty$"). The total error information $\sigma_{y_{i,\,\mathrm{total}}}$ then replaces σ_{y_i} in the objective function Eq. (1).

Note that the scheme expressed in Eq. (15) is not part of RETCML. The user has to compute the proper weights separately and then make them part of the input file for RETCML. For application of Eq. (15), it is useful to assume a locally linearized functional dependence of y on x, and to take finite-differences between consecutive data as approximations of the derivatives.

2.3.2. Log-transformed data

RETC's option to log-transform conductivity and diffusivity data has been retained in RETCML. The internal transformation of the input weights performed in RETC has, however, not been maintained. Rather, the weights entering the SSQ are the input weights, because the general philosophy with RETCML is that the user should assess his/her data quality and have full control over the estimation process. The only reason for using the log-transform feature should be to better fulfill the normality assumption for the measurement errors (i.e., one should only log-transform if the log-transformed error is closer to being normally distributed than the raw error). A wide spread in the magnitudes of data in itself is no reason to log-transform. There is also nothing wrong with specifying weights of widely different magnitudes.

2.3.3. Error magnitude unknown

Often in practical applications, the measurement error magnitude is unknown, and not even an educated guess can be made. RETCML can be used to perform ordinary-least-squares (OLS) estimation for this case, essentially emulating RETC. Parameter uncertainty is computed under the assumption of having an adequate model [by setting $\mathrm{SSQ}_{\min}=N-P$ internally, cf Eq. (5)]. The output value of $\mathrm{SSQ}_{\min}$ is, however, that for unit weights. This value can be useful for comparing different parametric models for the same data in a relative sense (e.g., the lower $\mathrm{SSQ}_{\min}$, the more appropriate the model). The probability of model adequacy loses its meaning once $\mathrm{SSQ}_{\min}$ has been scaled according to the quality of the fit.

2.3.4. Mixed estimator

RETC allows for estimation from retention and unsaturated conductivity data simultaneously. In essence, this is performed with a mixed sum-of-squares estimator which uses a relative weight W_1.

$$\mathrm{SSQ}(\mathbf{p}) = W_1\,\mathrm{SSQ}_k(\mathbf{p}) + \mathrm{SSQ}_\theta(\mathbf{p}) \tag{16}$$

where $\mathrm{SSQ}_k(\mathbf{p})$ is a conductivity (or diffusivity) objective function and $\mathrm{SSQ}_\theta(\mathbf{p})$ analogously a retention objective function, both defined as in Eq. (1).

In the preceding sections it was emphasized that each datum should be given its own weight, reflecting its reliability (the inverse of measurement error). Hence, a factor W_1 is formally not only superfluous, but can distort estimation results. Imagine the very realistic case where each sub-SSQ in the estimator Eq. (16) would lead to different optimal parameter vectors $\mathbf{p}$. By manipulating W_1, one can shift the final estimate

from one that is optimal for the conductivity data ($W_1 \rightarrow \infty$) to one that is optimal for the retention data ($W_1 = 0$). The pragmatic suggestion in RETC's manual (p. 36) to set W_1 within limits of 0.1 and 1 makes the problem of arbitrary results less extreme, but it still exists. No theory is available for setting W_1, and as shown here, the factor is not even needed.

Notwithstanding the above discussion, RETCML's input also includes the factor W_1. The main reason for retaining W_1 was to achieve compatibility with the old input file format. W_1 is generally ignored in the computation, with a single exception. When maximum-likelihood estimation is impossible anyway, i.e., when OLS estimation has been invoked, W_1 can be used according to Eq. (16). In this fashion, the conductivity/diffusivity data as a group can be weighed relative to retention data as a group. Still, the user should be aware of possibly distorted results, depending on how each group determines the overall estimate. Therefore, when using mixed estimators, it is recommended to check the output file for how each group contributed to the optimal SSQ (see "weighted sum of squares of observed versus fitted values"). A $W_1 \neq 1$ in WLS analyses or when estimating from retention data only results in a warning note in the output. For a more in-depth discussion of the mixed-estimator problem, see Hollenbeck and Jensen (1998).

3. Using RETCML

3.1. Input

RETCML is compatible with RETC in terms of the input file format (`.in`) and the control file format and name (still `retc.ctl`). Hence the user can easily re-analyze existing data. Likewise, the RETC manual (van Genuchten et al., 1991) is still generally valid for RETCML. However, two cases exist where the input is interpreted differently.

One difference concerns the way in which the OLS estimation is carried out (see Section 2.3.3 for conditions where OLS is appropriate and how to interpret its results). In RETC, an OLS analysis could be invoked by setting all weights to 1. This feature cannot be carried over to RETCML because unit weights could also, just by chance, reflect the true measurement error. Hence, if an OLS analysis is desired with RETCML, at least one weight has to be set to 0. A zero weight is interpreted as missing, and because it does not make sense to conduct a WLS estimation if even only a single weight is missing, one or more zero weight(s) will invoke an OLS analysis. In this context, the user ought to be careful to not accidentally introduce zero weights that can result from application of Eq. (15), i.e., when the gradient between a pair of data points $y(x)$ is infinite and $\sigma_y = 0$. In this situation, the user should remove one point of the pair (one line) from the input file. A data point with zero weight should not affect the estimation results anyway, because the apparent infinite measurement error at this data point implies that it lacks any reliability.

The second difference in input interpretation between RETC and RETCML concerns the weight W_1 (position 16 in `retc.ctl, weight of k/d data with respect to water content data`). As discussed in Section 2.3.4, this weight has no meaning within the context of maximum-likelihood estimation where each data point already has an appropriate weight. Hence, unless OLS for two types of data is conducted, RETCML overrides any input, sets $W_1 = 1$, the neutral value, and prints a warning message. Note that W_1 in RETCML has not exactly the same effect as in RETC, where in addition it is manipulated internally relative to data magnitude. This latter internal manipulation is not transparent to the user and has therefore been abandoned.

3.2. Output

RETCML preserves RETC's output of data and model predictions. The output for the estimation results is, however, different. The individual items are given below in the sequence they appear. Items in parentheses are not always part of the output, but depend on the particular estimation result.

1. **Banner:** Estimation results are announced with a new banner printout. Unlike for RETC, the printout comes before the correlation matrix, which in RETCML is interpreted as an estimation result, too.
2. (**Warning**) if at least one weight is missing (i.e., specified as zero) and an OLS rather than a WLS estimation has been performed (under the assumption of an adequate model; scaling the residuals as discussed in Section 2.3.3).
3. (**Warning**) if there are few data (cut-off set to $N-P < 20$), in which case the theorems used for interpreting estimation uncertainty do not strictly hold (cf Section 2.2).
4. (**Note**) if a relative weight for retention versus conductivity/diffusivity SSQ's $W_1 \neq 1$ was used as input, unless OLS is conducted for two data types. In all other cases, this factor is no longer supported as discussed in Sections 2.3.4 and 3.1.
5. **Correlation matrix:** Same as in RETC. This matrix can help the user to identify ill-defined estimation problems (for which the off-diagonal elements are close to ± 1).
6. **Probability of model adequacy:** The value p_{adeq} as

defined in Eq. (6). This quantity takes the place of RETC's R^2 value for reasons explained in Section 2.1. Note that if OLS is conducted, the residuals are scaled to give an adequate model [cf Eq. (5)]; hence, the now meaningless p_{adeq} is not output.

7. (**Warning**) if p_{adeq} is very small (cut-off set to 10^{-4}). Confidence limits on parameters make little sense for a very likely inadequate model, even though for the sake of completeness, they still are presented. The possible error leading to a very small p_{adeq}, i.e., too large (optimistically specified) weights, is pointed out.
8. (**Warning**) if p_{adeq} is unrealistically large [cut-off set to $SSQ_{min}/(N-P) < 0.01$; cf Eq. (6) and Section 2.1]. Computation of confidence limits in this case becomes numerically unstable [the Hessian has a very high condition number, so its inverse is ill-defined, cf Eq. (9)]. Confidence limits likely become incorrectly small. The possible error leading to a very large p_{adeq}, i.e., too small (pessimistically specified) weights, is pointed out.
9. **Parameter estimates and confidence:** Similar to RETC's output, also including the pseudo-univariate parameter standard errors "std.err." as defined in Eq. (12). Confidence limits are, however, obtained as the optimal parameter $\pm$ the revised (ellipsoid projection) half-lengths as defined in Eq. (13). The chosen confidence limit is 90% as opposed to RETC's 95% for reasons discussed in Section 2.2. The t-values are no longer given as output for reasons also discussed in that Section.
10. **Observed and fitted data:** This part of the output is nearly identical to that of RETC, with the exception of the normalized residuals r_i (labeled ⟨variable-name⟩RES) replacing the deviations d_i (labeled ⟨variablename⟩DEV) when WLS estimation has been performed. The normalized residuals r_i should be close to ± 1 on average (cf Section 2.1). Larger (in absolute terms) values of the residuals may indicate outliers in the data. For OLS, normalized residuals are undefined for lack of a measurement error weight; hence, the deviations d_i are given as output, just as in RETC. A note to that extent is printed out.

4. Practical information

The source code of RETCML is available under http://www.iamg.org/CGEditor/index.htm. The installation file of the Windows version is available under http://www.ussl.ars.usda.gov/models/models.html.

RETCML has been verified by comparing estimation results with those obtained with MATLAB code written for the same purpose, but using MATLAB's own optimization routine `constr`. Only small numerical errors due to different machine precisions were observed; hence, RETCML should be mathematically correct. All features implemented in RETCML (the output listed in Section 3.2) have also been tested. Note that the original RETC can crash with floating point errors when the estimation problem is very ill-defined (e.g., when one tries to estimate the residual saturation but only supplies the near-saturated part of the retention curve). This problem still exists in RETCML, because the optimization algorithm has not been changed. To report any other errors possibly still in the code, please send an email to Jirka Šimunek, simunek@ussl.ars.usda.gov.

4.1. Programming details

RETCML was developed based on the DOS version of RETC. It has been compiled with Microsoft Developer Studio, Fortran Power Station 4.0, on a Windows NT 4.0 station. In the source code, the modifications made are in small letters, whereas RETC is written in ALL CAPS. All modifications are also preceded by comment lines to make the code easier to understand. Some changes in the existing RETC code were necessary for compilation. These are: the standard input unit has been set to 5 instead of RETC's 0, an argument to `dmax1` has been made a double rather than a single precision number (line 132 now), and the terminal (screen) input unit for a character is * instead of 5 (line 60 now).

The original optimization algorithm was not changed. It was considered to introduce analytical objective function gradient evaluation, but comparison with parallel MATLAB code revealed only very small improvements. The required programming effort to introduce analytical gradients for all hydraulic parameter schemes in RETC seemed unwarranted for the rather small gain in accuracy.

Several new (with respect to RETC) functions/subroutines are included at the end of the RETCML source code. These implement the incomplete gamma function needed for computing the χ^2 statistic, which in turn is required for computing the probability of model adequacy [see Eq. (6)]. The functions are called `gammq` and `gammln`, the subroutines `gser` and `gcf`. All these were taken from Press et al. (1992, Chap. 6.2). The function $Q^{-1}(\cdot)$ [see Eq. (8)] is not coded, but rather contained in the form of a lookup table for 1–6 degrees-of-freedom at 90% confidence (lines 35–37 of the code).

References

Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. Dover, New York (1046 pp).

Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press (341 pp).

Beck, J.V., Arnold, K.J., 1977. Parameter Estimation in Engineering and Science. Wiley, New York (501 pp).

Brooks, R.H., Corey, A.T., 1964. Hydraulic properties of porous media. Hydrology Paper 3. Colorado State University, Fort Collins, Colorado (27 pp).

Burdine, N.T., 1953. Relative permeability calculations from pore-size distribution data, Petroleum Transactions. American Institute of Mineral Engineering 198, 71–77.

Hollenbeck, K.J., Jensen, K.H., 1998. Maximum-likelihood estimation of unsaturated hydraulic parameters. Journal of Hydrology 210, 192–205.

Kool, J.B., Parker, J.C., van Genuchten, M.Th, 1985. ONESTEP: a nonlinear parameter estimation program for evaluating soil hydraulic properties from one-step outflow experiments. Bulletin 85-3, Virginia Agricultural Experimental Station, Blacksburg, VA (43 pp).

Kool, J.B., Parker, J.C., van Genuchten, M.Th, 1987. Parameter estimation for unsaturated flow and transport models — a review. Journal of Hydrology 91, 255–293.

Leij, F.J., Bradford, S.A., 1994. 3DADE: a computer program for evaluating three-dimensional equilibrium solute transport in porous media. Research Report 134, US Salinity Laboratory, US Department of Agriculture, ARS, Riverside, CA (81 pp).

Mualem, Y., 1976. A new model for predicting the hydraulic conductivity of unsaturated porous media. Water Resources Research 12 (3), 513–522.

Press, W.S., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 1992. Numerical Recipes in C, 2nd ed. Cambridge University Press, Cambridge (994 pp).

Schweppe, F.C., 1973. Uncertain Dynamic Systems. Prentice-Hall, Englewood Cliffs, NJ (563 pp).

Šimunek, J., Šejna, M., van Genuchten, M.Th, 1996. The HYDRUS-2D software package for simulating water flow and solute transport in two-dimensional variably saturated media, version 1.0, IGWMC-TPS-56. International Ground Water Modeling Center, Colorado School of Mines, Golden, CO (167 pp).

Šimunek, J., Šejna, M., van Genuchten, M.Th, 1998. The HYDRUS-1D software package for simulating water flow and solute transport in one-dimensional variably saturated media, version 2.0, IGWMC-TPS-70. International Ground Water Modeling Center, Colorado School of Mines, Golden, CO (178 pp).

Toorman, A.F., Wierenga, P.J., Hills, R.G., 1992. Parameter estimation of hydraulic properties from one-step outflow data. Water Resources Research 28 (11), 3021–3028.

Toride, N., Leij, F.J., van Genuchten, M.Th, 1995. The CXTFIT code for estimating transport parameters from laboratory or field tracer experiments, version 2.0. Research Report 137, US Salinity Laboratory, US Department of Agriculture, ARS, Riverside, CA (121 pp).

van Dam, J.C., Stricker, J.N.M., Droogers, P., 1994. Inverse method to determine soil hydraulic functions from multi-step outflow experiment. Soil Science Society of America Journal 58, 647–652.

van Genuchten, M.Th, 1978. Calculating the unsaturated hydraulic conductivity with a new closed-form analytical model. Technical Report 78-WR-08. Water Resources Program, Department of Civil Engineering, Princeton University, Princeton, NJ (63 pp).

van Genuchten, M.Th, 1980a. A closed form equation for predicting the hydraulic conductivity of unsaturated soils. Soil Science Society of America Journal 44, 892–898.

van Genuchten, M.Th, 1980b. Determining transport parameters from solute displacement experiments. Research Report 118, US Salinity Laboratory, US Department of Agriculture, ARS, Riverside, CA (37 pp).

van Genuchten, M.Th, 1981. Non-equilibrium transport parameters from miscible displacement experiments. Research Report 119, US Salinity Laboratory, US Department of Agriculture, ARS, Riverside, CA (88 pp).

van Genuchten, M.Th, Leij, F.J., Yates, S.R., 1991. The RETC code for quantifying hydraulic functions of unsaturated soils. Technical Report IAG-DW 12933934, US Salinity Laboratory, US Department of Agriculture, Agricultural Research Service, Riverside, CA (83 pp).