HydroMe: A Package for estimating soil hydraulic parameters

Omuto, C.T.

(thineomuto@yahoo.com)

January, 2007

Abstract

This document presents a simple tutorial on the estimation of the soil hydraulic parameters from experimental data using HydroMe package for R. HydroMe contains common models in literature for estimating water infiltration and retention characteristics in soil science. The package assumes that a longitudinal data for soil water infiltration or retention exists and there is need to use this data in estimating the hydraulic parameters contained in certain mathematical models for the infiltration or water retention. HydroMe can be freely downloaded from http://cran.r-project.org/, copied into R library, and then loaded into R workspace as a package. Like other packages in R, using HydroMe requires working knowledge in R or S language for the purpose of interpretation of its outputs and further manipulation to improve the accuracy of the estimated hydraulic parameters.

1. Introduction

HydroMe estimates soil hydraulic parameters using experimental data for infiltration or water retention. Soil hydraulic parameters are the coefficients contained in the mathematical models describing infiltration and water retention characteristics (Kutilek and Nielsen, 1994[1]). Some of the common mathematical models for water infiltration and retention in soil science are shown in Table 1 below. These models may be fitted to experimental data obtained at one location, several locations, or in several groups of locations in a study area.

---

[1] Kutilek, M., Nielsen, D. 1994. Soil Hydrology. CATENA-VERLAG, Reiskirchen, 364pp.

Table 1. Common infiltration and water retention models in soil science

Model	**Mathematical expression**	**Parameters**
	Infiltration models	
Green and Ampt	$i(t) = K_s + [A / I(t)]$	K_s, A
Horton	$i(t) = f_c + (f_o - f_c)\exp(-kt)$	f_c, f_o, k
Philip	$i(t) = 0.5S / \sqrt{t} + f_c$	f_c, S
	Water retention models	
Gardner	$\begin{cases} \theta(h) = \theta_s & (ah \le 1) \\ \theta(h) = \theta_r + \dfrac{(\theta_s - \theta_r)}{[1 + \alpha h^n]} & (ah > 1) \end{cases}$	θ_r, θ_s, α, n
Brooks and Corey	$\begin{cases} \theta(h) = \theta_s & (\alpha h \le 1) \\ \theta(h) = \theta_r + (\theta_s - \theta_r)[\alpha h]^{-\lambda} & (\alpha h > 1) \end{cases}$	θ_r, θ_s, α, λ
Campbell	$\begin{cases} \theta(h) = \theta_s & (\alpha h \le 1) \\ \theta_s(\alpha h)^{-b} & (\alpha h > 1) \end{cases}$	θ_s, α, b
van Genuchten	$\begin{cases} \theta(h) = \theta_s & (ah \le 1) \\ \theta(h) = \theta_r + \dfrac{(\theta_s - \theta_r)}{[1 + (\alpha h)^{\alpha}]^m} & (ah > 1) \end{cases}$	θ_r, θ_s, α, n, m (m = 1- 1/n)
Groenevelt and Grant	$\begin{cases} \theta(h) = \theta_s & (ah \le 1) \\ \theta(h) = \theta_{s1} e^{-\alpha_1 h} + \theta_{s2} e^{-\alpha_2 h} & (ah > 1) \end{cases}$	θ_{s1}, θ_{s2}, α_1, α_2

HydroMe contains the following functions for estimating hydraulic parameters:

- *SSphilip* – for parameters in the Philip's model
- *SShort* – for parameters in the Horton's model
- *SSGampt* – for parameters in the Green and Ampt's model
- *SSvan* – for parameters in the van Genuchten's model
- *SSgard* – for parameters in the Gardner's model
- *SSbrook* – for parameters in the Brooks and Corey's model
- *SScamp* – for parameters in the Campbell's model
- *SSMgg* – for parameters in the modified Groenevelt and Grant's model

These functions are invoked in R using the command: *function* (*input*, *parameters*). In this command, *parameters* refer to the parameters in the models as indicated in Table 1 while *function* is the name of HydroMe function. In using this command to estimate the hydraulic parameters in the preferred model, the general mathematical assumption is given by;

$$y_i = f(x_i, \varphi) + \varepsilon_i \quad (1)$$

where y_i are either the measured infiltration rates $i(t)$ or moisture contents of water retention $\theta(h)$, x_i are the corresponding independent variables (either time t in infiltration or h in water retention), and ε_i are the associated error terms due to the estimation y_i with the HydroMe functions. ε_i are generally assumed to have normal distribution and constant variance across y_i. In the command for HydroMe functions, x_i form the *input*. The commands calling HydroMe functions are implemented in R using standard functions in NLME library e.g. nls, nlsList, and nlme (Faraway, 2006[2]). Like any other executable line in R, the outputs of HydroMe functions are stored in R *objects* (e.g. model name) that can be referred to later. Suppose experimental data have been collected according to any of the strategies shown in Table 2.

Table 2. Two possible strategies for the collection of experimental data

One sample		Several samples	
x	**y**	**x**	**y**
x_1	y_1	x_{11}	y_{11}
x_2	y_2	x_{12}	y_{12}
x_3	y_3	.	.
.	.	x_{ij}	y_{ij}
.	.	x_{21}	y_{21}
.	.	.	.
x_i	y_i	x_{ij}	y_{ij}

[2] Faraway, J.J., 2006. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models. CRC Press. New York. 300pp.

Then, a full line command for Equation (1) to estimate (say) the van Genuchten parameters may look like this;

```
>model1 <-nls(y~SSvan(x, Thr, Ths, alp, scal), data)
>model2 <-nlsList(y~SSvan(x, Thr, Ths, alp, scal)|Sample, data)
```

Model 1 above may be for a case of one sample or where there are many samples but are considered as coming from one group and model 2 may be for cases where experimental data are grouped into one or more groups. Once the models have been developed, they can be visualized, summarized, or improved in subsequent command lines using R functions such as *summary, coef, intervals,* etc. For example, in the above two models further lines may be added as follows:

```
>summary(model1) #   to display the summary of the estimated hydraulic parameters
>coef(model2) #   to display the magnitudes of the estimated hydraulic parameters
```

2. *Example – Philip's infiltration parameters*

The data used for this example was taken from a study on the effects of physical degradation on hydraulic properties of soil (Omuto *et al.,* 2006[3]). It consists of infiltration rates (in cm/minute) and time (in minutes) taken from 540 infiltrometer measurements. The data were collected at various points in the field using single-ring infiltrometers with internal diameter of 30-cm. The numbers of records are not uniform for the single-ring measurements and the time intervals between infiltration rates are not equal for all the measurements. The data also contains classes of evidence of soil physical degradation at the sampled locations. These classes are coded E0 (for no evidence of degradation), E1 (for evidence of moderate degradation), and E2 (for evidence of severe degradation). Only 30 locations (labelled as *PlotNo*) out of the 540 locations have been used for this

[3] Omuto, C.T., Minasny, B., McBratney, A.B., Biamah, E.K., 2006. Nonlinear mixed effect modelling for improved estimation of water retention and infiltration parameters. Journal of Hydrology, 330, 748-758.

illustration. This data is stored in HydroMe under the name *Infilt* with the following labels for the columns: *PlotNo, Rate, Time, and Erosion*. The *PlotNo* column contains the grouping factor for locations while *Erosion* column contains the soil physical degradation classes. After attaching the package [using `> library (HydroMe)`], the data can be loading and viewed as shown below.

```
>data (Infilt)
Grouped Data: Rate ~ Time | PlotNo
Sample        PlotNo        Erosion       Time   Rate
 1            11P3          E1             1.0   3.05
 2            11P3          E1             2.0   3.00
 3            11P3          E1             3.0   2.20
…      ……            ….           ….. ………
```

This data can be plotted according to soil physical degradation classes as shown in figure 1 using the following command,

```
>plot(Infilt, outer=~Erosion, xlab="Time (Minutes)",
+ ylab="Rates (cm/min)", layout=c(3,1))# figure 1
```

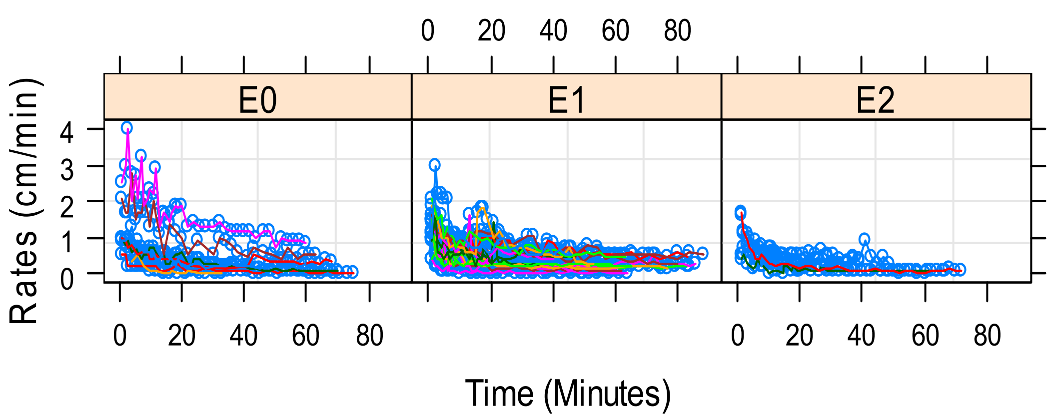

Figure 1. Infiltration characteristics for different soil physical conditions

Although the data is grouped, the following illustrations show model development in HydroMe using *nls, nlsList*, and using mixed-effects' *nlme*.

(a) Model to estimate aggregate parameters for all the 30 points

```
>Infilt.nls <-nls (Rate~SSphilip (Time, fc, S), data=Infilt)

>summary (Infilt.nls)

Formula: Rate ~ SSphilip(Time, fc, S)

Parameters:
      Value        Std. Error   t value
fc    0.174028     0.0226277    7.69091
 S    2.497910     0.1408310    17.73690

Residual standard error: 0.407671 on 1103 degrees of freedom
```

(b) Model to estimate parameter for individual locations

```
>Infilt.nlis <-nlsList (Rate~SSphilip(Time, fc, S)|PlotNo, data=Infilt)

>summary (Infilt.nlis)
Call:
  Model: Rate ~ SSphilip(Time, fc, S) | PlotNo
  Data: Infilt

Coefficients:
   fc
numeric matrix: 30 rows, 3 columns.
PlotNo             Value        Std. Error        t value
101P3              0.511118     0.08080115        -8.306254
111P3              0.115334     0.10142364        -21.296078
131P3              0.513582      0.09870617        -6.750809
..    ….           …..          …..

   S
numeric matrix: 30 rows, 3 columns.

 PlotNo            Value        Std. Error  t value
101P3              2.827428     0.5461741   5.176790
111P3              4.885854     0.5977587   8.173622
131P3              4.357512     0.5868492   7.425266
…..          ……           …….          ……

Residual standard error: 0.3183535 on 1045 degrees of freedom
```

(c) Model to estimate mixed-effects parameters

```
>Infilt.nlme <- nlme (Rate~SSphilip(Time, fc, S)|PlotNo,
 + fixed=list(fc~1, S~1), start=fixef(Infilt.nlis), data = Infilt)

>summary(Infilt.nlme)

Nonlinear mixed-effects model fit by maximum likelihood
```

```
  Model: Rate ~ SSphilip(Time, fc, S)
  Data: Infilt
     AIC          BIC               logLik
  837.33        867.3756             -412.665

Random effects:
 Formula: list(fc ~ 1 , S ~ 1 )
 Level: PlotNo
 Structure: General positive-definite
                StdDev        Corr
      fc    0.9553943       fc
      S    1.8130176    -0.827
Residual 0.3198183

Fixed effects: list(fc ~ 1 , S ~ 1 )
       Value      Std.Error   DF    t-value           p-value
fc   -1.683296   0.1756448       1074     -9.58352    <.0001
S    4.515988   0.3520824        1074    12.82651     <.0001

Standardized Within-Group Residuals:
       Min         Q1        Med        Q3      Max
 -3.585651 -0.6117053 0.02976425 0.5729575 3.882367

Number of Observations: 1105
Number of Groups: 30
```

The parameters for these models can be called using *coef.* For example, the parameter estimates for individual locations in the mixed-effects model is as follows,

```
>coef (Infilt.nlme)

PlotNo     fc               S
31P3       0.056437      5.631674
41P3       0.993500      3.162201
51P3       0.294500      2.384537
61P3       0.265903      3.844253

…. ………              …….
> plot (Infilt.nlme) # produces figure 2.
```

The relative fit information for all these models can also be plotted using *plot*. Such fits may be used to assess the validity of the assumptions (Omuto et al., 2006) or the goodness of fit (Faraway, 2006). Figure 2 shows a plot of standard residuals of the nlme model versus fitted values. This plot was obtained by typing the command, *plot* (*model name*).

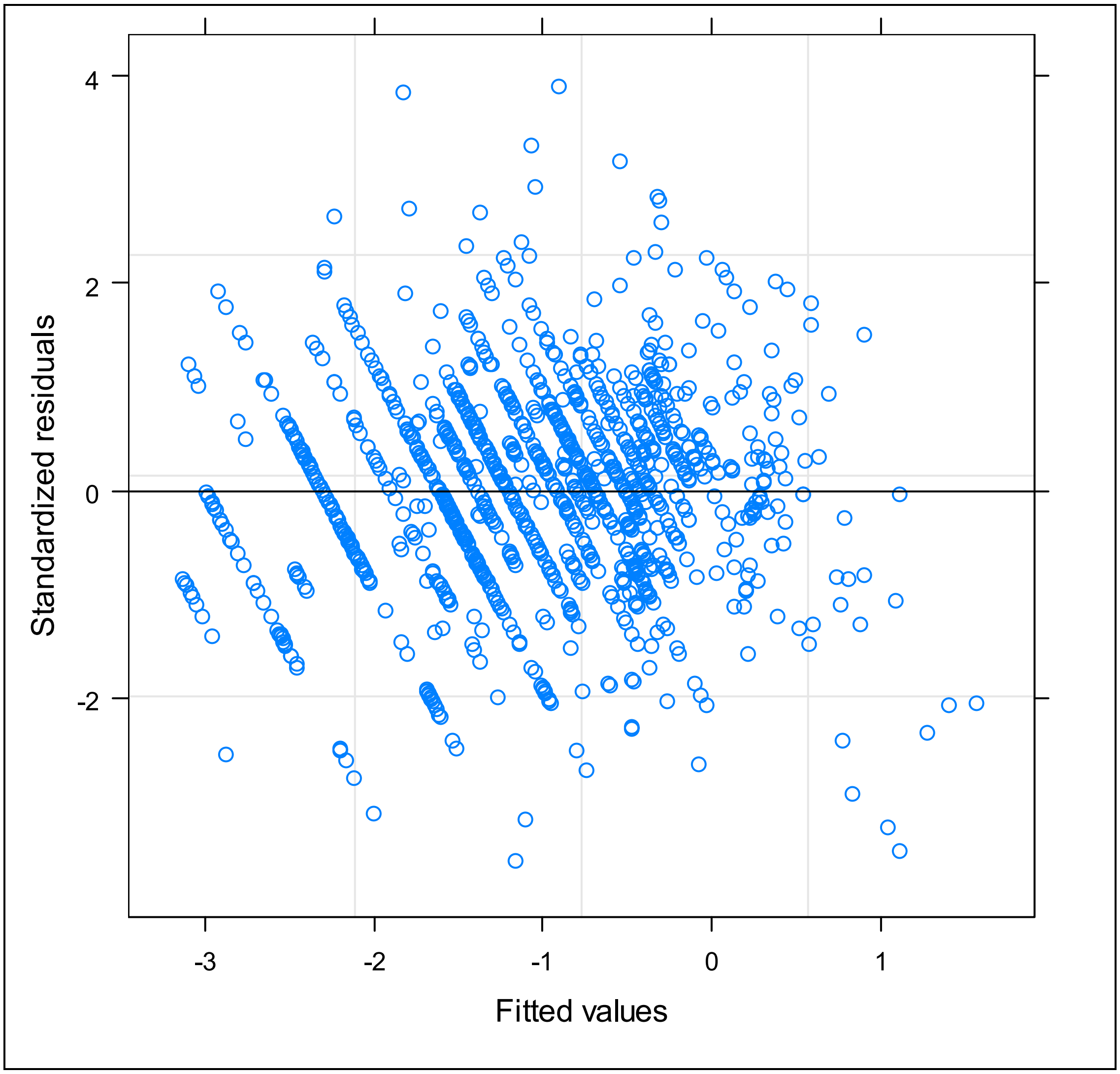

Figure 2. Plots of standardized residuals to assess relative fits for the mixed-effect model

3. *Example-van Genuchten's water retention parameters*

The data used in this example was taken from ISRIC world database of soil water retention characteristics at different locations in many countries of the world (ISRIC, 2004). Only a subset of this data was considered in this example. This set of data consists of 40 samples each with eight levels of suction potential from zero to 150 m and corresponding entry for volumetric moisture content (in $cm^3\ cm^{-3}$). The data contains six columns labelled *Sample, Country, BD* (for bulk density in g cm^{-3}), *x, lnx,* and *y*. The *Sample* column is the grouping variable representing different locations within the countries.

(a) Aggregate parameters for the 40 samples

```
>data(isric)
> isric.nls <-nls(y~SSvan(lnx, Thr, Ths, alp, scal), data=isric))
> summary(isric.nls)

Formula: y ~ SSvan(lnx, Thr, Ths, alp, scal)

Parameters:
     Estimate Std. Error t value Pr(>|t|)
Thr   0.19293    0.04861   3.969 8.95e-05 ***
Ths   0.53555    0.01862  28.763  < 2e-16 ***
alp  -1.41698    0.08629 -16.421  < 2e-16 ***
scal  3.53291    0.96115   3.676 0.000279 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Residual standard error: 0.1262 on 316 degrees of freedom
```

In this model, *Thr* = θ_r, *Ths* = (θ_s - θ_r), *alp* = lnα, *scal* = *n* in the van Genuchten function. The model estimates the average parameters for the whole dataset.

(b) Model to estimate parameters for individual locations

However, in order to estimate water retention parameters for each sampling location, the column containing the number of location is first made the grouping variable. In this data the column is labelled *Sample*. It is declared the grouping variable and the modelling procedure follows accordingly.

```
>isric.nlis <-nlsList(y~SSvan(x,Thr, Ths, alp, scal)|Sample,
 + data=isric)
> isric.nlis
Call:
  Model: y ~ SSvan(x, Thr, Ths, alp, scal)| Sample
   Data: isric
Coefficients:
                  Thr        Ths        alp      scal
Benin3       0.18497203 0.3455683 -1.3104208  4.429375
```

```
Italy4        0.20077604 0.6297568 -1.3776578  5.583299
…..                ……          ……          ……     …..
Nicaragua40   0.20467524 0.5902969 -0.9616822  1.618516
Poland10      0.02648218 0.4011335 -1.4838871  2.932118
……                …….         …..          ….     ….
Togo14        0.07954954 0.3675257 -1.4135902  2.631077
Togo15        0.06362981 0.3700125 -1.2682658  2.303943
…..                …..          ….          …..          …
Degrees of freedom: 144 total; 72 residual
Residual standard error: 0.01397491
```

It is important to note that the model did not converge for some samples (e.g sample number 708 in Netherlands). Such samples may further be checked for any error in the data or any other possible reason for non-convergence. The estimates for the model can be assessed for variability using Figure 3.

```
> plot(intervals(isric.nlis), layout =c(4,1)) # figure 3
```

The vertical axis in Figure 3 that indicates the *Points* column in the *isric* data has been trimmed for the sake of display. The labels under this column are supposed to be shown in the figure. Plots such as illustrated in Figure 3 may be useful in understanding the parameters explaining variability in soil hydrology (Sharma *et al.,* 1980[4]). The models from individual fits may also be used to store the information about the function as well as provide initial estimates for mixed-effects models directly.

(c) <u>Model to estimate mixed-effects parameters</u>

```
>data(Wret)
>Wret.nlis <- nlsList(y~SSvan(x, Thr, Ths, alp, scal), data=Wret)
>Wret.nlme <-nlme(Wret.lis)
>summary (Wret.nlme)
Nonlinear mixed-effects model fit by maximum likelihood
  Model: y ~ SSvan(x, Thr, Ths, alp, scal)
 Data: Wret
        AIC       BIC   logLik
  -866.3474 -814.6464 448.1737
Random effects:
```

[4] Sharma, M.L., Gander, G.A., Hunt, C.G., 1980. Spatial variability and its infiltration in a watershed. Journal of Hydrology 45, 101-122.

```
 Formula: list(Thr ~ 1 , Ths ~ 1 , alp ~ 1 , scal ~ 1 )
 Level: Sample
 Structure: General positive-definite
                  StdDev                  Corr
     Thr   0.097015108     Thr    Ths    alp
     Ths   0.092662217     0.670
     alp   0.009954618     0.926  0.349
    scal   0.212542017    -0.563 -0.289 -0.648
Residual 0.018102168

Fixed effects: list(Thr ~ 1 , Ths ~ 1 , alp ~ 1 , scal ~ 1 )
           Value              Std.Error   DF    t-value     p-value
 Thr   0.1912442  0.01846819  200   10.35533    <.0001
 Ths   0.4999374  0.01762219  200   28.36976    <.0001
 alp   0.0266919  0.00306025  200   8.72212     <.0001
scal   0.8820888  0.04826739  200   18.27505    <.0001

Standardized Within-Group Residuals:
       Min         Q1          Med         Q3      Max
 -2.566365 -0.4872212 -0.06378597 0.4842619 2.894475

Number of Observations: 240, Number of Groups: 30
```

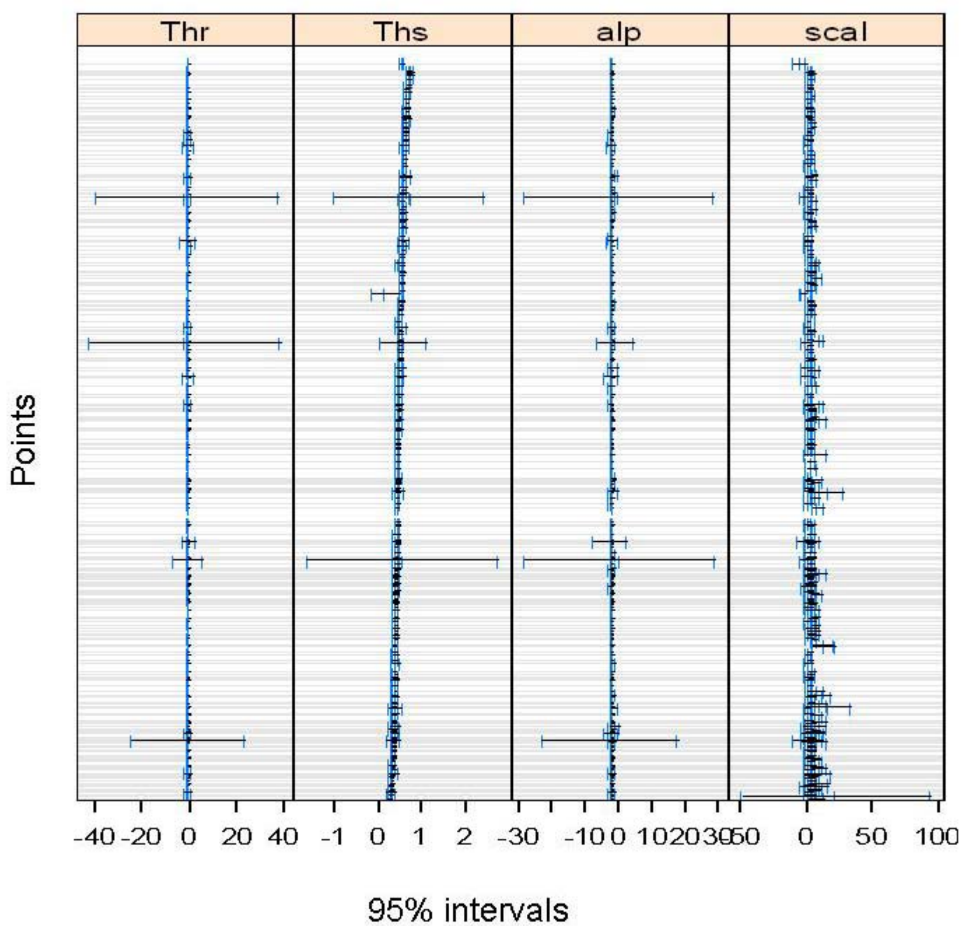

Figure 3: Approximate 95% interval for estimated parameters in the *isric.nlis* model

The fixed-effects in the mixed-effects model are usually comparable with those of the aggregate model. However, the estimated within-group standard errors for the mixed-effects may smaller than those of the aggregate model. This is due to the between-group variability that is accounted for by the mixed-effects model. This is one of the improvements that mixed-effects models introduce. They can also still be further improved by incorporating covariates. For example, bulk density values in the *Wret* data may be added to the above model.

```
> Wret.nlme1 <-update(Wret.nlme, fixed=list(Thr~1,Ths~BULK, scal~1,
 + alp~1), random=list(Thr~1, Ths~1, scal~1),
 + start=c(fixef(Wret.nlme)[1], fixef(Wret.nlme)[2],
+ fixef(Wret.nlme)[3],0,fixef(Wret.nlme)[4]))

>summary (Wret.nlme1)
…….
Structure: General positive-definite
                              StdDev  Corr
Thr.(Intercept)   0.083848671 Thr(I)    Ths
            Ths   0.074871200 -0.497
           scal   0.04343792  -0.363  0.502
       Residual   0.01072998

Fixed effects: list(Thr ~ BULK, Ths ~ 1, scal ~ 1, alp ~ 1)
                  Value       Std.Error   DF       t-value   p-value
Thr.(Intercept)     0.10100  0.009935   199     0.232506    0.8164
       Thr.BULK     0.25513  0.003342   199      1.857641   0.0647
            Ths     0.50000  0.008366   199      0.382355   0.7026
           scal     0.02506  0.002943   199      0.326823   0.7441
            alp     1.25884  0.039068   199      2.522411   0.0124
……
```

A plot of the fitted model and measured values (Figure 4) may also be called to demonstrate the goodness of the fit.

```
> plot(Wret$y~fitted(Wret.nlme1)) # figure 4
```

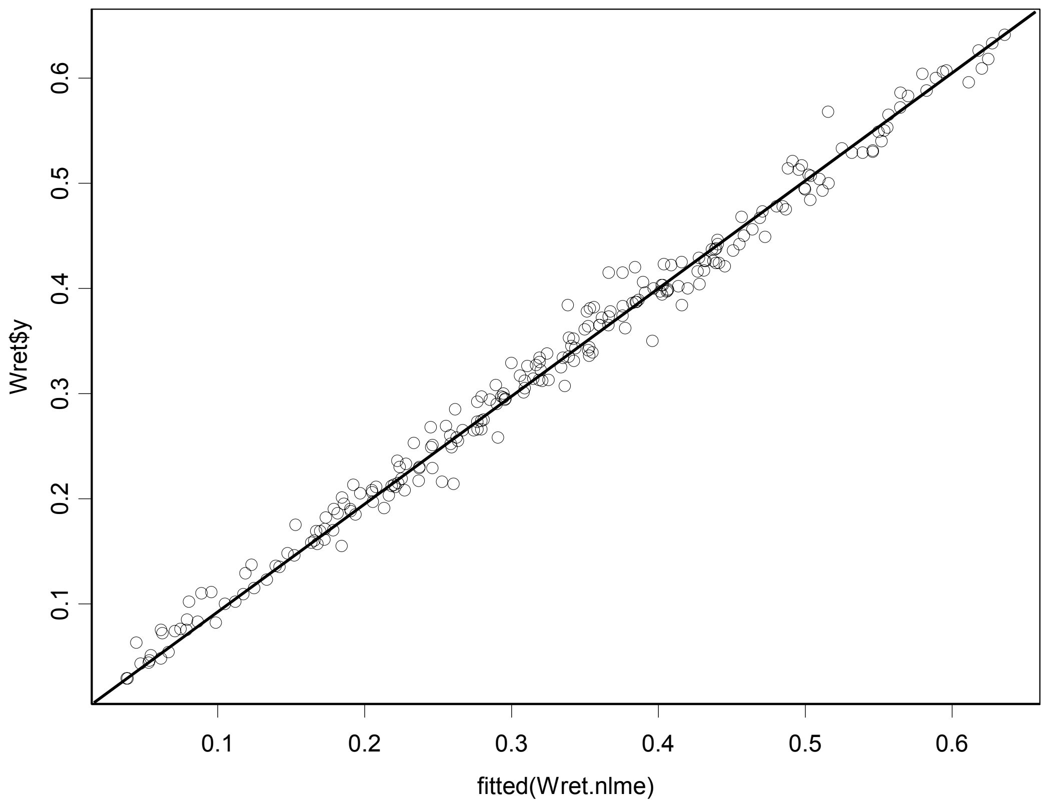

Figure 4: Measured versus within-group fitted values corresponding to *Wret.nlme*1 model

3. Discussions and conclusion

HydroMe presents an efficient and fairly accurate approach for estimating soil hydraulic parameters from experimental data. By computing the parameters of group(s) of experimental data in one singly command, the package beats other statistical packages that need both data entry and parameter estimation to be repeated over the entire numbers of experimental groups in a study. Not only does this eliminate the drudgery and potential errors in data handling but also expedite the estimation process. *HydroMe* also presents a suite of alternatives for improving the accuracy in the estimated parameter. For example, the package incorporates the capabilities of using individual-fits approach (that loops the parameter estimation process for an experimental group) to recognize grouping effects needed to improve the case of aggregating all locations into one (Omuto et al., 2006). It also incorporates mixed-effects approach where there are many opportunities to provide for parameter correlation and also to include other covariates into the estimation process (Faraway, 2006). Mixed-effects opportunities implemented in *HydroMe* provide a good platform to address such issues.

At present, HydroMe does not include models for estimating unsaturated hydraulic conductivity. It is envisaged that this task would be undertaken soon and appropriate models added to the package. The package does also not include all the functionality that may be desirable for specific users. Users are encouraged to adjust the package accordingly.

Acknowledgement

Some of the water retention data used in this package were obtained from ISRIC world database. International Foundation of Sciences (IFS) through the project number C-3953/1 supported the development of this package. I also acknowledge the support from ICRAF through Dr. Markus Walsh, Mr. Richard Coe, and Dr. Keith Shepherd. Valuable comments from Dr. Budiman Minasny of ACPA, University of Sydney and Dr. W. Venebles of the R Development Core Team were also received during the development of *HydroMe*.