Soil Physics Functions in CEPHaStat

by Chawezi Miti and Murray Lark.

1 Introduction

The CEPHaStat collection of functions includes functions to fit soil water retention curves (van Genuchten model) to data on volumetric water content at different tensions. This can be done for a single set of measurements of water content at each tension, for a set of replicated water release curve measurements from which a single model is required, or for a set of observations comprising replicated observations from different subsets of soils which are expected to differ in their water retention properties (e.g. soils from under different treatments, or different soil types). In this latter case it is possible for some parameters to be pooled over the subsets, while others differ. A test of the null hypothesis that the soils all have a common parameter set can be made.

Once a set of parameters is estimated then it is possible to compute a set of soil quality indices, based on the work of Reynolds et al. (2007).

The model fitting to a single set of points on the water release curve uses the general non-linear fitting capability of R in the function `VanGenuchten.fit.single`. For cases where there is between-sample variation to account for the fitting is done in a mixed model setting under which parameter values for individual soil samples are assumed to be a combination of fixed effects (e.g. specifying a mean for a particular soil type) and random effects. This is not straightforward because of the non-linearity of the model that we are fitting. The functions here use an R package called saemix (Comets et al., 2017), which uses what is called the stochastic approximation expectation maximization method. In effect we sample from possible distributions of the parameters, and so we obtain empirical distributions which reflect our uncertainty about their values which can be expressed as standard errors. We shall write up more of the detail of this for publication in due course.

Detail for the functions is provided in the CEPHaStat v3 manual. This document gives some worked examples. The full script for these is provided in a file Soil Physics Examples.R. There are three data files wrcExample.dat, wrcZambia.dat and wrcZimbabwe.dat. The latter two are real data sets extracted from World Soil Information System (WoSIS), described by Batjes et al. (2017) as it stood in 2016 (Batjes et al., 2016).

2 Example 1. Fitting a single model to one set of measurements

.

2.1 Read in and plot the data

.

The data are read in as follows:

```
data.df<-read.table("wrcExample.dat",header=T)
```

The CEPHaStat function `plot.wrc.data` will plot the water release curve data in a data frame, if the names of the variables which contain the tensions (assumed to be in kPa) and volumetric water contents are provided. Optionally, if van Genuchten parameters are provided, the model is also drawn. Here the function is used just to produce a plot of the data (Fig. 1).

```
plot.wrc.data(data.df,hvar="h",thetavar="theta")
```

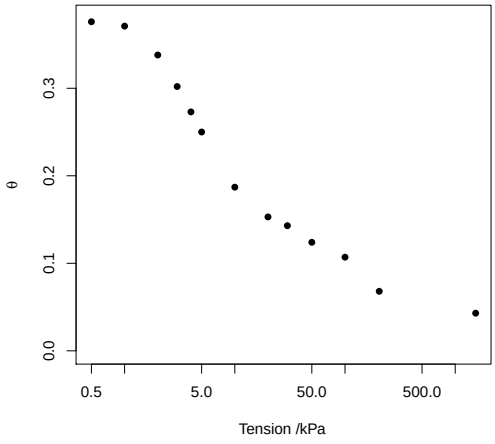

Figure 1: Plot of wrc data

2.2 Fit parameters of the Van Genuchten water release curve

.

The Van Genuchten water release curve has the form below. At a tension of h the volumetric water content is $\theta(h)$ where

$$\theta(h) \;=\; \theta_{\rm r} \;+\; (\theta_{\rm s} - \theta_{\rm r})\,\frac{1}{\{1+(\alpha h)^n\}^m},$$

where $\theta_{\rm s}$ and $\theta_{\rm r}$ are respectively the volumetric water content at saturation and the residual water content, α is related to the reciprocal of air-entry tension, and n is a parameter related to the pore size distribution. Here we do not fit m as a separate parameter but set it to $m = 1 - n^{-1}$.

It is necessary to provide an initial set of parameter values to start the fitting. These are entered in the vector `init.vals` in order `thr,ths,alp,nscal` corresponding to $\theta_{\rm r},\ \theta_{\rm s},\ \alpha,\ n$:

```
init.vals<-c(0.05,0.5,0.05,1.5)
```

The parameters are then estimated as follows:

```
op<-VanGenuchten.fit.single(data.df,init.vals)
```

It is assumed that the tension (kPa) and volumetric water content (proportion) are in data.df as variables `h` and `theta` respectively.

The object `op` contains the estimated parameters. If the plotting function is run again, with these parameters as input, then a plot is produced of the point data with the fitted curve (Fig. 2).

```
plot.wrc.data(data.df,op$Coefficient.estimates,hvar="h",thetavar="theta")
```

2.3 Interpretation of the wrc parameters

.

The wrc parameters might be plugged into a soil water model, or used to compute single quantities such as the water-filled pore space ratio at a specified tension. The `SQ.indices` function has been written to compute a set of soil physical quality

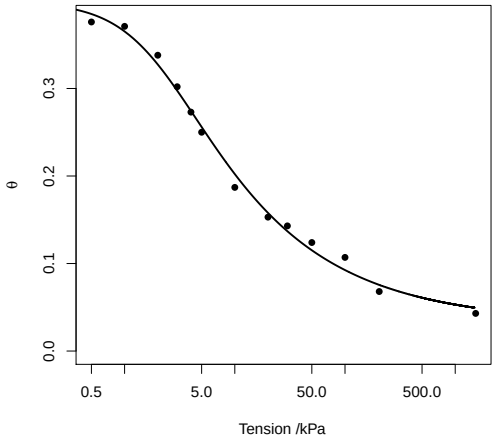

Figure 2: Plot of wrc data with fitted van Genuchten function

indices proposed by Dexter (2004) and Reynolds et al (2007). These indices are listed and explained below.

Dexter's S is the modulus (absolute value) of the gradient of the water release curve at its inflection point (i.e. where the slope stops increasing with increased tension), interpreted in terms of the microstructure of the soil, which is better-defined, with a wider range of pore sizes, when S is large. Dexter (2004) gives an interpretation of values of S which this function reproduces.

S	
$S > 0.035$	Good microstructural quality
$0.02 < S \leq 0.035$	Poor microstructural quality
$S \leq 0.02$	Very poor microstructural quality

Note that S is defined with respect to the water release curve for *gravimetric* water content, unlike in conventional usage. For this reason a bulk density value is required to rescale the volumetric water content computed from the Van Genuchten parameters. If bulk density is not measured directly, then it is computed from the value of θ_S, assuming a particle density of 2.65 g cm^{-3}: – this may underestimate particle density for very ferruginous soils, and overestimate it for soils with a large organic content (Landon, 1984).

Total porosity is equal to θ_S, i.e. the volumetric water content of the saturated soil.

Macroporosity is the difference between total porosity and porosity at a tension when it is assumed that only micropores are filled (matrix porosity). Reynolds *et al.* (2007) suggest three values, by default we use the middle one (4.9 kPa), but this can be changed by setting the optional parameter `t_m` to the selected value. The interpretation is based on Reynolds *et al.* (2007), if macroporosity is $\leq$0.04 (volumetric) then the soil is assumed to be degraded by compaction or consolidation. Otherwise, for medium to fine textured soils, it is regarded as undegraded.

The relative water capacity, RWC, is defined as the ratio of the volumetric water content at field capacity to the total porosity. This is interpreted as optimal for microbial activity in the interval $0.6 < \text{RWC} \leq 0.7$, too dry below the range and too wet above (Reynolds *et al.*, 2007).

Field capacity by default is assumed to be the volumetric water content at a tension of 9.8 kPa, but this can be changed by adjusting the optional argument `t_FC`. Permanent wilting point by default is assumed to be the volumetric water content at a tension of 1471 kPa, but this can be changed by adjusting the optional argument `t_PWP`.

Plant available water capacity, PAW, is the difference between the water content at field capacity and the permanent wilting point. Following Reynolds *et al.* (2007), this is interpreted as follows:

PAW m^3 m^{-3}	
PAW > 0.2	Ideal
$0.15 <$ PAW ≤ 0.2	Good
$0.1 <$ PAW ≤ 1.5	Limited
PAW ≤ 0.1	Poor

The air capacity, AC, of the soil is the difference between the total porosity and the field capacity. Following Reynolds *et al.* (2007) these values are interpreted as follows

AC m^3 m^{-3}	
AC > 0.15	Aeration likely to be adequate for all soils
$0.10 <$ AC ≤ 0.15	Aeration likely to be adequate except for fine-textured soils
AC ≤ 0.1	Crop-damaging aeration deficit likely

a value less than 0.1 (volumetric) is interpreted as likely to lead to crop-damaging aeration deficit, aeration is likely to be adequate if air capacity is between 0.1 and 0.15, except for fine-textured soils, and a value exceeding 0.15 is likely to mean adequate aeration in all soils.
To compute these indices from Van Genuchten parameters in the object `op$` we use the function `SQ.indices` as below

```
index.values<-SQ.indices(op$Coefficient.estimates)
```

The function `print.indices` will extract the output from the object created this way. By default it just prints the index values and their interpretations to the console, but setting `ret=T` will make it create a dataframe of these values.

```
print.indices(index.values,ret=F)
```

The console-output for this first example is shown below

```
Dexter's S              0.052          Good microstructural quality
Total porosity          0.399          ~
Macroporosity           0.14           Undegraded (medium to fine textured)
Relative
 water capacity         0.5            RWC suboptimal for microbial activity
Plant-available
 water capacity         0.15           PAW limited
Air capacity            0.2            AC good for all soils
```

3 Example 2. Fitting a single model to a set of replicate measurements

.

3.1 Read in and plot the data

.

```
data.df<-read.table("wrcZambia.dat",header=T,stringsAsFactors=T)
names(data.df)
```

These data comprise wrc measurements for three acrisols in Zambia. The observations may be plotted as follows

```
plot.wrc.data(data.df,hvar="h",thetavar="theta",main="Three Acrisols from Zambia")
```

3.2 Fitting the models

.

We treat these observations as three replicated observations from a single population. We therefore assume that there is an underlying 'population' water retention curve, and it is those parameters which we want to estimate. This is done with the `saemix` algorithm to fit a non-linear mixed model in which individual replicates have random variation in their Van Genuchten parameters around the mean for their population. In CEPHaStat this is achieved with the function `VanGenuchten.fit.group`. As with fits to values for a single sample, it is assumed that the tension (kPa) and volumetric water content (proportion) are in data.df as variables h and theta respectively. Initial parameter values are presented in init.vals, as before. The final argument of the function indicates a factor which groups observations from a single soil. If you examine the data for this example you will see that the factor levels for the three samples are `Zambia_Acrisol_1, Zambia_Acrisol_2, Zambia_Acrisol_3`.

```
op<-VanGenuchten.fit.group(data.df,init.vals,"Sample")
```

The output from this function includes an extra variable, the standard error for each parameter of the Van Genuchten function, which can be seen as follows

```
print(op)
```

```
$Coefficient.estimates
      Estimate
thr    0.05655
ths    0.45157
alp    0.43878
nscal  1.37400

$Standard.errors
      Standard error
thr          0.03056
ths          0.02265
alp          0.08778
nscal        0.04948
```

Again, the function `plot.wrc.data` can be used to plot the data, with the fitted model (Fig. 3).

```
plot.wrc.data(data.df,op$Coefficient.estimates,hvar="h",thetavar="theta",
main="Three Acrisols from Zambia")
```

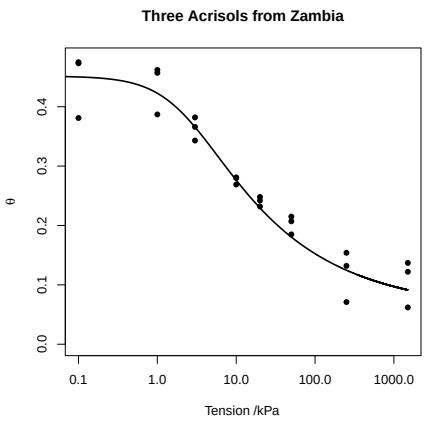

Figure 3: Plot of wrc data for three acrisols from Zambia, with pooled fitted van Genuchten function

the soil quality indices can be computed as before:

```
index.values<-SQ.indices(op$Coefficient.estimates)
print.indices(index.values,ret=F)
```

Dexter's S	0.053	Good microstructural quality
Total porosity	0.45	~
Macroporosity	0.12	Undegraded (medium to fine textured)
Relative water capacity	0.6	RWC suboptimal for microbial activity
Plant-available water capacity	0.19	PAW good
Air capacity	0.17	AC good for all soils

4 Example 3. Fitting a model to data with replicate observations on different soil classes and testing for differences

.

4.1 Read in and plot the data

In this example we have data from two soil types, vertisols and ferralsols, in Zimbabwe. We can read in the data, and plot the points on the wrc curve with the classes, indicated by the `groups` argument of the function, distinguished by colour (Fig. 4; groups are FAO soil classes in this example).

```
data.df<-read.table("wrcZimbabwe.dat",header=T,stringsAsFactors=T)
names(data.df)


plot.wrc.data(data.df,hvar="h",thetavar="theta",
xlab="Tension /kPa",maxth=-1,groups="FAO_class",main="Zimbabwe soils")
```

4.2 Fit the models

We shall now estimate different water release curves for each soil class. There are various options, since distinct models for the classes might be fitted by allowing certain parameters to differ between the classes whilst others are common across the classes. As part of the process of inference we propose a 'null model', the simpler one to be considered. In the most straightforward case this is a model in which all four parameters are common to all soil classes (a single pooled model). We may then evaluate evidence for differences between the classes by fitting, for comparison,

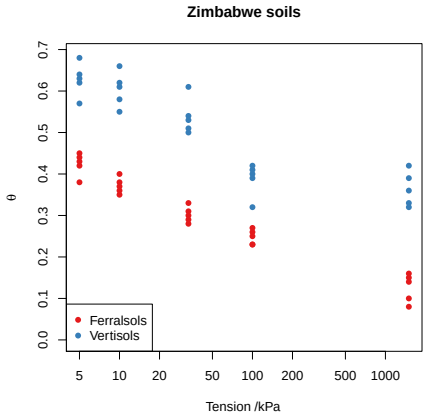

Figure 4: Plot of wrc data for samples of vertisols and ferralsols from Zimbabwe

.

a model in which some or all of the parameters differ among the classes. In this case we allow all parameters to differ in the so-called 'full' model. This is done by setting up two vectors, `par.var.null` and `par.var.null`. Each of these has four entries, corresponding to the van Genuchten parameters in the standard order θ_r, θ_s, α, n. If, for example, the fourth element of `par.var.null` is set to zero then the parameter α is constant over all classes, if the value is 1 then this parameter may differ between the classes. To compare models differing in all parameters (full model) with a model in which all parameters are fixed over classes (null), the vectors are set up as follows.

```
par.var.null<-c(0,0,0,0)  #Which parameters differ between groups in null model?
par.var.full<-c(1,1,1,1)  #Which parameters differ between groups in full model?
```

We then fit the null and full models with the following command

```
op<-VanGenuchten.fit.compare(data.df,init.vals,g.name="Sample",par.var.null,
par.var.full,cov.name="FAO_class")
```

Note that the argument `g.name` identifies the factor which groups together observations on a single soil sample (`Sample` here) and that `cov.name` identifies the factor which defines the different soil classes which are to be compared (`FAO_class` here). We can examine the output by printing the model object:

```
print(op)
```

```
$Coefficient.estimates
      Ferralsols Vertisols
thr      0.04284   0.36279
ths      0.45405   0.61809
alp      0.12535   0.02751
nscal    1.29304   3.31897

$Comparison
      Group.dependent.parameters.null Group.dependent.parameters.full
thr                                 0                                1
ths                                 0                                1
alp                                 0                                1
nscal                               0                                1

$Inference
$Inference$log.likelihood.ratio
[1] 80.95

$Inference$df
[1] 4

$Inference$P.value
[1] 0.000000000000000111
```

The output includes the estimates of the parameters for each class, `$Coefficient.estimates` and the term `$Comparison` is a reminder of which parameters differ between classes in each model. Next the log-likelihood ratio is presented. This statistic is twice the difference between the maximized log-likelihood values for the null and the full model. If the null model is correct then this statistic is asymptotically distributed as chi-squared, with degrees of freedom equal to the difference between the sum of the terms in `par.var.full` and the sum of terms in `par.var.null`, four here. A P-value may therefore be computed to assess the evidence for the full model in comparison with the null. Here that evidence is strong, showing a difference between the two soil classes with respect to their Van Genuchten parameters.
As before the fitted Van Genuchten models may be plotted: (Fig. 5).

```
plot.wrc.data(data.df,op$Coefficient.estimates,groups="FAO_class",
hvar="h",thetavar="theta",main="Zimbabwe soils")
```

The soil quality indices can be computed as before:

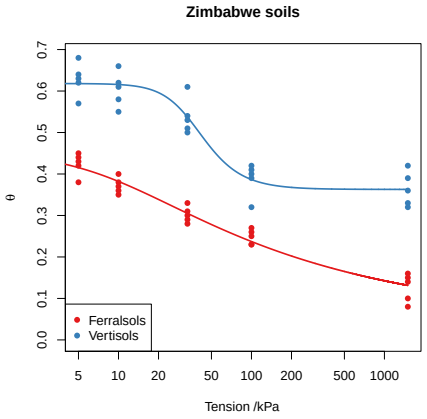

Figure 5: Plot of wrc data for samples of vertisols and ferralsols from Zimbabwe, with fitted van Genuchten functions specific to each class

```
index.values<-SQ.indices(op$Coefficient.estimates)
print.indices(index.values,ret=F)
```

```
Dexter's S          Ferralsols      0.046   Good microstructural quality
Total porosity      Ferralsols      0.45    ~
Macroporosity       Ferralsols      0.04    Macroporosity degraded
Relative
 water capacity     Ferralsols      0.8     RWC exceeds range for optimal microbial
Plant-available                              activity
 water capacity     Ferralsols      0.25    PAW ideal
Air capacity        Ferralsols      0.07    Aeration deficit likely in rooting zone
Dexter's S          Vertisols       0.18    Good microstructural quality
Total porosity      Vertisols       0.62    ~
Macroporosity       Vertisols       0.2     Undegraded (medium to fine textured)
Relative
 water capacity     Vertisols       0.6     RWC suboptimal for microbial activity
Plant-available
 water capacity     Vertisols       0.25    PAW ideal
Air capacity        Vertisols       0.23    AC good for all soils
```

5 References

Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., de Mendes, J. 2016. Standardised soil profile data for the world (WoSIS, July 2016 snapshot),`http://dx.doi.org/10.17027/isric-wdcsoils.20160003`.

Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., de Mendes, J. 2017. WoSIS: Providing standardised soil profile data for the world, Earth System Science Data 9, 1-14, `http://dx.doi.org/10.5194/essd-9-1-2017`

Comets, E., Lavenu, A., Lavielle, M. 2017. Parameter estimation in nonlinear mixed effect models using saemix, an R implementation of the SAEM algorithm. *Journal of Statistical Software*, **80**, 1–41.

Dexter, A.R. 2004. Soil physical quality. Part I. Theory, effects of soil texture, density and organic matter, and effects on root growth. *Geoderma*, **120**, 201–214.

Landon, J.R. (ed), 1984. Booker Tropical Soil Manual (1st Edn) Longman, Harlow. page 97, section 6.5.

Reynolds, W.D., Drury, C.F., Yang, X.M., Fox, C.A., Tan, C.S., Zhang, T.Q. 2007. Land management effects on the near-surface physical quality of a clay loam soil. *Soil & Tillage Research*, **96**, 316–330.