



Asian Journal of Mathematics & Statistics

ISSN 1994-5418

Selection of a NLR-Model by Re-Sampling Technique

S.A. El-Shehawy

Department of Mathematics, College of Science, Qassim University,
P.O. Box 6644, Buriedah 51452, Kingdom of Saudi Arabia

Abstract: The aim of this study is to discuss a method for the selection of parametric models of the moisture retention characteristic MRC for natural soils. The author presents computational tools in S-Plus for the adequate selection of these parametric models as non-linear regression models (NLR-models) and this method of selection based on cross-validation and bootstrap (re-sampling technique). Early, these current computational tools were developed in some works about the selection of the nonlinear regression models. The current results will be compared with the obtained results which are based on the simulation studies and some calculations of measures of nonlinearity in NLR models. Finally, the results indicate to the best selected model which is identical with the selected model by the simulation study and the calculation of the measures of nonlinearity.

Key words: Nonlinear regression models, cross-validation, model selection, retention function, profile, measure of nonlinearity

INTRODUCTION

Mathematical modeling of water flow in unsaturated natural soils leads to partial differential equation-Richard's equation (Kutilek and Nielsen, 1994). To solve this equation one needs the description of the relationships between soil moisture content, hydraulic conductivity and capillary pressure. The moisture retention characteristic (the relation between water content Θ and pressure head h) can be measured for natural soils. A variety of parametric models have been proposed to describe MRC. All these models are NLR-models, which are able to describe the S-shaped retention curve.

A well known MRC model was studied and moreover a now widely used class of functions for parametrizing measured soil water characteristics

$$\Theta(h) = \Theta_r + \frac{(\Theta_s - \Theta_r)}{(1 + (\alpha h)^n)^m} \quad (1)$$

is presented, where α , n and m are positive parameters defining the MRC's shape (Van Genuchten, 1980; Vereecken *et al.*, 1989). If we consider the saturated water content Θ_s and the residual water content Θ_r as parameters too, we will have a MRC model with five parameters, which have to be estimated from observed soil water retention data and this model is denoted by (VG1).

In King (1965), the following model with five real parameters Θ_s , h_0 , b , ε and γ is suggested

$$\Theta(h) = \Theta_s \times \left[\frac{\cosh((h/h_0)^b + \varepsilon) - \gamma}{\cosh((h/h_0)^b + \varepsilon) + \gamma} \right] \quad (2)$$

where, b having negative values and this model denoted by (K1).

Table 1: Some variants of van Genuchten model (VG1) and King model (K1).

Model	(VG2)	(VG3)	(VG4)	(VG5)	(K2)
Case	$m = 1-1/n$	$m = 1-2/n$	$m = 1$	$\Theta, = 0, m = 1$	$\varepsilon = 0$

In El-Shehawy (2008), some special cases of the models (VG1) and (K1) with fewer parameters are studied. Table 1 indicates to these special cases (VG2), (VG3), (VG4) and (VG5) of the model (VG1) and also the model (K2) as a special case of the model (K1).

The model selection in El-Shehawy (2008) is executed by using the simulation study, which based on hundred sets of simulated data and the measures of non-linearity (curvature, bias and skewness).

The results of analyzing experimental data using a parametric model may heavily depend on the chosen model. It is unknown whether a certain regression model describes the unknown true regression function sufficiently well.

In this study, we propose computational tools in S-Plus for the adequate selection of NLR-models based on cross-validation and bootstrap.

The aim of the proposed technique in this study is to select the best parametric NL-model, to compare the results of the used method of the selection with the results of other selection methods, to indicate the caused parameter of this nonlinearity and finally to avoid this defect.

DESCRIPTION OF RE-SAMPLING TECHNIQUE

Let us consider a general NLR-model

$$y_{ij} = f(x_i, \beta) + \varepsilon_{ij}, \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, n_i, \quad \sum_{m=1}^k n_m = N \quad (3)$$

where, y_{ij} is the values of the response variable Y at fixed values x_i of exploratory variables (nonrandom design points). Here, the real valued function $f(\cdot, \beta)$ is expectation function and it is known up to a P -dimensional vector of parameters $\beta = (\beta_1, \beta_2, \dots, \beta_p) \in \mathbb{R}^P$ and this function is twice continuously differentiable in β . Moreover, the random perturbations ε_{ij} (the measurement errors) are uncorrelated random variables with zero mean and unknown variance σ^2 , which do not depend on x_i . Given the data

$$(x_i, y_{ij}), \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, n_i, \quad \sum_{m=1}^k n_m = N \quad (4)$$

the estimation of the regression function f reduces to estimate its parameters.

If the object of the analysis is primarily the estimation of the regression function, that is, of the values of the regression function itself over a region \mathfrak{R} of interest, then the weighted cross-validation criterion will be a convenient criterion characterizing the performance of the model $f(x, \beta)$ (Bunke *et al.*, 1998, 1999):

$$CV = \sum_{i=r}^S c_i \sum_{j=1}^{n_i} |y_{ij} - f(x_i, \hat{\beta}^{(i)})|^2 \quad (5)$$

Here, $\hat{\beta}^{(i)}$ denotes the least squares estimator (LSE), which is calculated from the $(n-1)$ observations left after deleting the observation (x_i, y_{ij}) , with $x_i \in \mathfrak{R}$. Thereby we assume that the (ordered) values $x_r \leq \dots \leq x_s$ of the independent variable are in the region.

If $\mathfrak{R} = [a, b]$ is the interval of the interest with $a \leq x_r \leq \dots \leq x_s \leq b$, then one will use weights

$$c_i = \frac{d_i}{2n_i(b-a)}; \quad 1 \leq r < i < s \leq k, \quad (6)$$

with and $d_i = x_{r+i} - x_{r-1}$; $d_r = 2(x_{r+i} - x_r)$, $d_s = 2(x_s - x_{s-1})$ and $1 \leq r < i < s \leq k$.

If one would not like to specify the interval $\mathfrak{X} = [a, b]$, then $a = x_1$ and $b = x_2$ should be the standard values. The use of a criterion like (5) is justified by the fact, that it estimates the weighted sum of the actual squared prediction errors. The cross-validation criteria (5) will be adequate, if prediction or estimation of the regression function (here the MRC) is a primary objective of the data analysis.

Following Bunke *et al.* (1998, 1999), the model selection could be done by using the cross-validation criteria (5) in three steps:

- List alternative models $f_1(\cdot, \beta_1), \dots, f_M(\cdot, \beta_M)$ corresponding to the theoretical or practical experience in the field of application.
- Select among these models f_m , $m \in \{1, 2, \dots, M\}$ a model $f_{\hat{m}}$ with the smallest value of the chosen cross-validation criteria.
- The data analysis is then done with the estimated regression function $f_{\hat{m}}(x, \hat{\beta}_{\hat{m}})$.

As an alternative to the Ordinary Least Squares Estimator (OLSE) in view of possibly heteroscedastic variances $\sigma_i^2 = \sigma^2(x_i)$ a (non-homogeneous) variance model could be chosen and a corresponding Weighted Least Squares Estimator (WLSE) should be used. Furthermore, it is possible to assess with a procedure in S-Plus the accuracy of the parameter estimations using re-sampling methods.

With respect to the goodness of fit, one can use the criteria $RSS = S(\hat{\beta}_m)/N$, where

$$S(\hat{\beta}_m) = \min_{\beta_m} \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - f_m(x_{ij}, \beta_m))^2, m \in \{1, \dots, M\}. \quad (7)$$

ACCURACY AND PROFILE

Estimates (for instance parameter estimates) will be only of value, if they are complemented by an assessment of their accuracy. Therefore, statistical packages offer the possibility to calculate characteristics like the standard deviation or confidence intervals. These calculations are often based on linear models (linearization) or asymptotic theories (Bates and Watts, 1988; Seber and Wild, 2005). In case of NLR-models combined with typical small data sets these calculated results may differ significantly from the true values. Especially strong nonlinear parameters may have confidence intervals which will be essentially differing from the calculated intervals using a linearization of the NL-model. To overcome these difficulties it is possible to use re-sampling methods -bootstrap- or corrected versions of the confidence intervals based on nonlinear theories.

Accuracy of Parameter Estimates

The bootstrap is a re-sampling method that tries to mimic the (in general unknown) distribution of the observations (measurement error) by generating artificial independent bootstrap samples from the fitted model. The estimates of the model parameters are repeatedly calculated for these bootstrap samples. The accuracy of the original estimates is inferred from the empirical distribution of the bootstrap estimates. In Bunke *et al.* (1998, 1999), a moment oriented bootstrap procedure, which is based on a non-linear transformation (Box-Cox-transformation) to get nearly normally distributed residuals and on the estimation of variance (second moment) of this distribution, is used. The bootstrap can be applied when a regression and variance model were selected. Then it is possible to generate bootstrap samples based on realizations of corresponding normally distributed pseudo random numbers and to compute confidence intervals and some other measures of accuracy (variance, bias and mean absolute relative error) based on this simulation.

Profiles of Parameter Estimates

Another way to get exact confidence intervals for a single component of the parameter vector is the use of so called profile t functions and plots (Bates and Watts, 1988; Seber and Wild, 2005). Plots of the profile t function $\tau(\beta_r)$ provide exact likelihood confidence intervals of the parameters and in addition, reveal how nonlinear the estimation situation for the corresponding parameter is.

The profile t function represents the profile of the residual sum of square function as a function of a single component of the parameter vector in a neighborhood of the estimated parameter vector.

In case of a linear regression model (LR-model) the squares root of this residual sum of square function-the profile t function- is a linear function $\delta(\beta_r)$ of the corresponding component β_r of the parameter vector β . Moreover, the plot of this function against the component would be a straight line.

For a nonlinear component of the parameter vector in a NLR-model the plot of the profile t function will be curved and the amount of curvature gives some information about the nonlinearity of the model and of the component of the parameter vector. Using these profiles it is possible to construct exact confidence intervals for the components of the parameter vector of the model under consideration.

RESULTS

In this section we present the results of the model selection with re-sampling technique and assessment of accuracy of parameter estimates.

Model Selection with Re-Sampling

Now, we consider the model selection based on the weighted cross-validation criterion (5). Using this criterion we will look for a retention model, which minimizes the weighted sum of the actual squared prediction errors. That means, we look for a model, which is stably able to predict values of the retention function based on the given data situation.

We consider the first simulated data set SIMUL_1 of sand (El-Shehawy, 2008), which is given in Table 2.

Since the simulation uses a homogenous error variance, we consider only all combinations of retention models with a homogenous variance for the measurement errors.

Table 3 contains the ranking of the studied models with respect to the RSS criterion and the CV criterion. These results are obtained by using the S-PLUS procedure in Appendix A.

We remark that the model (VG5) is not flexible enough and therefore convergence problems during the calculation of the CV value occur.

An application of the cross-validation approach to our problem would lead to the selection of the model (VG4) with four parameters.

Table 2: The first simulated MRC-data for sand

h	1.00	1.00	3.16	3.16	10.00	10.00	31.62	31.62	100.00	100.00
SIMUL_1	0.431	0.376	0.391	0.382	0.232	0.193	0.071	0.068	0.046	0.040
h	199.53	199.53	630.96	630.96	2512.59	2513.30	15848.92	15848.92		
SIMUL_1	0.029	0.072	0.053	0.040	0.039	0.034	0.071	0.036		

Table 3: Ranking of retention models with respect to CV and RSS for the first simulated data

Model	RSS $\times 10^4$	Ranking	CV $\times 10^4$	Ranking
(VG4)	2.497	5	4.701	1
(VG2)	2.458	3	4.820	2
(K2)	2.491	4	4.858	3
(VG1)	2.458	2	4.928	4
(VG3)	2.500	6	4.946	5
(K1)	2.455	1	5.544	6
(VG5)	12.747	7	34.884	7

Referring to El-Shehawey (2008), if we are interested in the model (VG4) with four parameters, we will try to improve the problematic qualities by the suggested re-parameterization (RVG4) of the model (VG4) through replacing the parameter n by $1/\tilde{n}$.

Since the values of RSS and CV are independent of parameterization, then the model (RVG4) is identical with respect to these criteria with the model (VG4). The models (VG1) and (K1) with five parameters are very flexible and are having optimal values with respect to RSS.

These results are neither depending on the used simulated data set of a soil nor on the simulated type of soil. As a rule, we get optimal values with respect to the cross-validation criterion for the model (VG4) and its re-parametrization (RVG4). Moreover, the models (VG1) and (K1) are having the best fit with respect to the values of RSS.

If we consider data sets with a possibly inhomogeneous variance structure, we will obtain similar ranking of retention models in combinations with an inhomogeneous variance model. Referring to Bunke *et al.* (1999), variance modeling can improve the parameter estimation for a given model using the weighted least squares in place of the ordinary least squares.

Assessment of Accuracy of Parameter Estimates

The assessment of accuracy is a necessary part of the parameter estimation for a given NLR-model. We will consider accuracy statistics for some parameters of the model (VG4) and its re-parameterization (RVG4).

Table 4 gives such statistics for the parameters α , n and \tilde{n} of the model (VG4) and (RVG4), respectively combined with the simulated data SIMUL_1 of sand. These statistics are obtained by

Table 4: Accuracy statistics for α , n and \tilde{n} of the models (VG4) and (RVG4) for SIMUL_1

Parameter	Estimated value	CI = [lb; ub]	CI ^p = [lb; ub]	CI* = [lb*, ub*]
α	0.106	[0.090; 0.122]	[0.099; 0.138]	[0.092; 0.126]
n	2.410	[1.328; 3.492]	[1.469; 2.947]	[1.681; +∞)
\tilde{n}	0.415	[0.229; 0.601]	[0.330; 0.650]	(-∞; 0.599]

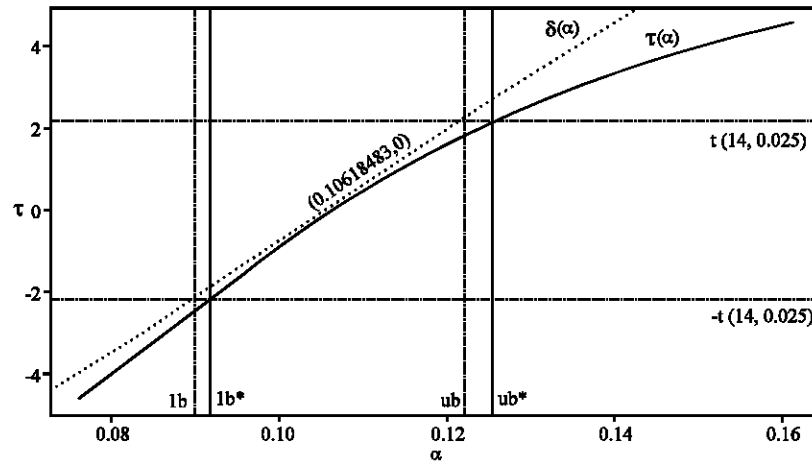


Fig. 1: Profile t function $\tau(\alpha)$ and linear profile $\delta(\alpha)$ for the parameter α in the model (VG4) with respect to the first simulated set of sand

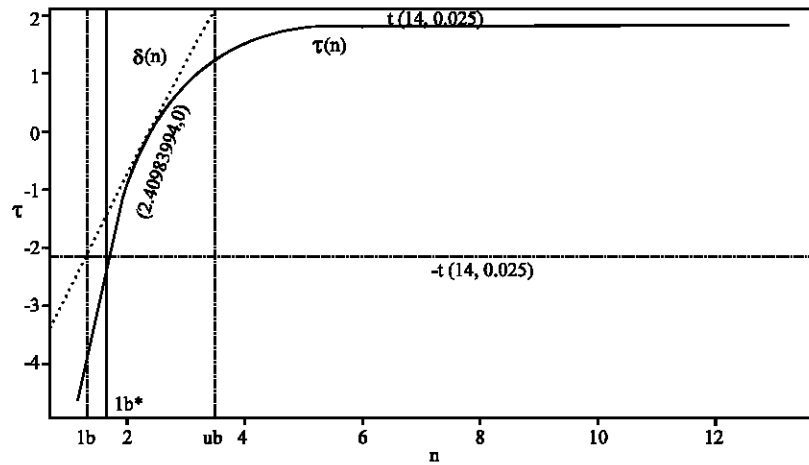


Fig. 2: Profile t function $\tau(n)$ and linear profile $\delta(n)$ for the parameter n in the model (VG4) with respect to the first simulated set of sand

using the S-PLUS procedure in Appendix B. This table contains the estimated values of the corresponding parameter of the model and the used data set and several 95% confidence intervals CI, CI^b and CI^* .

We start with classical asymptotic results, which are based on a linearization of a NL-model. The confidence interval CI with lower bound lb and upper bound ub shows these results. The calculations of these confidence intervals are based on the theory of LR-models.

In case of a parameter, which has near linear behavior, this linear approximation of a NLR-model provides adequate inferential results.

Unfortunately, in a nonlinear analysis and parameters with strong nonlinear behavior combined with typical small data sets this approximation will be woefully inadequate. For example, the parameter α shows weak nonlinear behavior, the parameter n has stronger properties of nonlinearity with respect to its estimation. Referring to El-Shehawey (2008), the corresponding parameter \tilde{n} of the re-parameterization (RVG4) has improved properties.

In the second attempt to improve the accuracy statistics based on a bootstrap simulation we calculated a 95% confidence interval CI^b of the estimated parameter. The calculation of this confidence interval is based on bootstrap sample of hundred data sets generated with NLRCHOICE (Bunke *et al.*, 1998, 1999). These accuracy statistics are inferred from the empirical distribution of the estimated parameters for this bootstrap sample.

The third possibility is the calculation of an exact likelihood interval CI^* with lower bound lb^* and upper bound ub^* . That means, if we are interested in this exact interval for the nonlinear parameter β_r , we will use the profile t function $\tau(\beta_r)$ to correct the asymptotic confidence interval CI.

In Fig. 1 and 2, which are obtained by using the S-PLUS procedure in Appendix B, the curves (solid) show the profiles $\tau(\alpha)$ and $\alpha(n)$, the dotted lines are the linear approximations $\delta(\alpha)$ and $\delta(n)$ and the lines (dashed) describe the construction of 95% exact confidence intervals for the parameters α and n , respectively.

In case of the near linear parameter we get only a slight modification CI^* of the classical asymptotic confidence interval CI, but in case of the parameter n with strong nonlinear estimation behavior the exact confidence interval CI^* is completely different from CI. This exact confidence interval CI^* is unbounded above. That means, for typical data sets frequently very large values of n will be estimated.

CONCLUSIONS

Model selection of NLR-models depends on the main aim of the analysis. A retention model has to be able to describe the typical S-shaped retention curve.

If one is interested in the best fitting model, very flexible models with more parameters should be used. As a criterion ordinary or in case of heteroscedasticity weighted least squares can be used. In case of a retention curve the models (VG1) and (K1) with five parameters are optimal with respect to the goodness of fit.

If one is primary interested in the prediction of future values of the depending variable -here water content-a cross-validation criterion like will be used. This criterion describes the stiffness of the model that means the ability to predict values of the depending variable based on the given data. Optimal with respect to the cross-validation criterion is the model (VG4) with $m = 1$.

The results of model selection are identical with the corresponding results which were obtained by the simulation studies or the calculation of measures of nonlinearity.

Using profile t functions, we note that:

- The selected model (VG4) has a strong nonlinear parameter n and also it has a possible appropriate re-parameterization (RVG4) with respect to this parameter.
- It is possible to correct classical confidence intervals of estimated parameters, which having clear nonlinear behavior.

APPENDIXES

A. S-PLUS Procedure for Model Selection

```
> Sandsimul118<-list(x=c(1.00,1.00,3.16,3.16,10.00,10.00,31.62,31.62,100.00,100.00,199.53,199.53,630.96,630.96,2512.59,
2513.30,15848.92,15848.92),y=c(0.4310391,0.3756147,0.391471,0.3820379,0.2320795,0.1925222,0.0705075,
0.0679685,0.0461746,0.0399038,0.0288667,0.0715104,0.0534139,0.0397278,0.0388411,0.0337290,0.0713605,0.0355392),
xname="Pressure head",yname="Water content")
> Sandsimul118list<-structure(.Data=list(
list(fkt=(yr~p1+(p2-p1)/(1+(p3*xr)
** p4)** p5) [[3]] , par = structure (. Data = c ( 0 . 045430811889044 , 0 . 404116316275956 ,
0 . 129247124828146 , 2 . 86958567927214 , 0 . 678016268058138 ) , Names = c ("p1" , "p2" , "p3" , "p4" , "p5" ) ) , p = 5 , ccode = 1 ,
lambda = 1 , case = "A" , inv = (xr~(1/p3)*(((p2-p1)/yr-p1))**((1/p5)-1)**(1/p4)) [[3]] , name = "VG1" ) ,
list ( f k t = ( y r ~ p 1 + ( p 2 - 1 + ( p 3 * x r ) ** p 4 ) * ( 1 - 1 / p 4 ) ) [ [ 3 ] ] , p a r = s t r u c t u r e ( . D a t a = c (
0.045367561011416,0.403791312302805,0.13093415030614,2.91956331160232) , Names = c ("p1" , "p2" , "p3" , "p4" ) ) , p = 4 ,
ccode = 1 , lambda = 1 , case = "A" , inv = (xr~(1/p3)*(((p2-p1)/(yr-p1))**((p4/(p4-1))-p1)**(1/p4)) [[3]] , name = "VG2" ) ,
list ( f k t = ( y r ~ p 1 + ( p 2 - p 1 ) / ( 1 + ( p 3 * x r ) ** p 4 ) * ( 1 - 2 / p 4 ) ) [ [ 3 ] ] , p a r = s t r u c t u r e ( . D a t a =
c(0.0448921002615711,0.399732245739772,0.146138035177302,3.72718111706133) , Names = c ("p1" , "p2" , "p3" , "p4" ) ) ,
p = 4 , ccode = 1 , lambda = 1 , case = "A" , inv = (xr~(1/p3)*(((p2-p1)/(yr-p1))**p4/(p4-2))-1)**(1/p4) [[3]] , name = "VG3" ) ,
list ( f k t = ( y r ~ p 1 + ( p 2 - p 1 ) / ( 1 + ( p 3 * x r ) ** p 4 ) ) [ [ 3 ] ] , p a r = s t r u c t u r e ( . D a t a = c ( 0 . 0463234428638109 , 0 . 407598503941853 ,
0 . 10618483313943 , 2 . 40983994858276 ) ,
Names = c ("p1" , "p2" , "p3" , "p4" ) ) , p = 4 , ccode = 1 , lambda = 1 , case = "A" , inv = (xr~(1/p3)* (((p2-p1)/(yr-p1))-1)**(1/p4)) [[3]] , name = "VG4" ) ,
list ( f k t = ( y r ~ p 1 + ( p 2 - p 1 ) / ( 1 + ( p 3 * x r ) ** ( 1 / p 4 ) ) ) [ [ 3 ] ] , p a r =
structure (.Data=c(0.0463234427833878,0.407598504025092,0.106184833124216,0.414965318943241) , Names = c ("p1" , "p2" ,
"p3" , "p4" ) ) , p = 4 , ccode = 1 , lambda = 1 ,
case = "A" , inv = (xr~(1/p3)*(((p2-p1)/(yr-p1))-1)**p4) [[3]] , name = "RVG4" ) ,
list ( f k t = ( y r ~ p 1 / ( 1 + ( p 2 * x r ) ** p 3 ) ) [ [ 3 ] ] , p a r = s t r u c t u r e ( . D a t a = c ( 0 . 444272556041495 , 0 . 0989142375227744 ,
1 . 23110558339949 ) , Names = c ("p1" , "p2" , "p3" ) ) , p = 3 , ccode = 1 , lambda = 1 , case = "A" , inv = (xr~(1/p2)*((p1/yr)-1)**(1/p3))
[[3]] , name = "VG5" ) ,
list ( f k t = ( y r ~ p 1 * ( ( c o s h ( ( x r / p 2 ) ** p 3 ) + p 4 ) - p 5 ) / ( ( c o s h ( ( x r / p 2 ) ** p 3 ) + p 4 ) + p 5 ) ) [ [ 3 ] ] ,
par = structure (.Data=c(0.0458700029292177,0.0818102381174718,0.842307290815704,-0.170583024339882,-
0.796739670954662) , Names = c ("p1" , "p2" , "p3" , "p4" , "p5" ) ) , p = 5 , ccode = 1 , lambda = 1 , case = "A" , inv = (xr~p2*((acosh(p5*((p1+yr)/(p1-
yr))))-p4*(1/p3)) [[3]] , name = "K1" ) ,
list ( f k t = ( y r ~ p 1 * ( ( c o s h ( ( x r / p 2 ) ** p 3 ) - p 4 ) / ( ( c o s h ( ( x r / p 2 ) ** p 3 ) + p 4 ) ) ) [ [ 3 ] ] ,
par = structure (.Data=c(0.0458700029292177,0.0818102381174718,0.842307290815704,-0.796739670954662) , Names = c ("p1" ,
```



```
"p2","p3","p4"),p=4,ccode=1,lambd=1,case="A",inv=(xr-p2*((acosh(p4*((p1+yr)/(p1-yr))))*(1/p3)))[[3]],name="K2"),class
="fktlist")
> Sandsimul118list.parinit<-parinit(Sandsimul118,Sandsimul118list,graph=T)
> Sandsimul118.vpred<-selectmv(Sandsimul118,Sandsimul118list,graph=T)
> Sandsimul118.vpredb<-sensitive(Sandsimul118.vpred[[1]],nb=100)
> Sandsimul118.vpredb<-confpar(Sandsimul118.vpredb,nb=100,coverage=.95)
> Sandsimul118.vpredb<-dblbootstrap(Sandsimul118.vpredb,nb2=2)
> Sandsimul118.tpred<-selectmt(Sandsimul118,Sandsimul118list,lambd=c(2,1,.5,0,-.5,-1,-2),graph=T)
> Sandsimul118.tpredb<-sensitive(Sandsimul118.tpred[[1]],nb=100)
> Sandsimul118.tpredb<-confpar(Sandsimul118.tpredb,nb=100,coverage=.95)
> Sandsimul118.tpredb<-dblbootstrap(Sandsimul118.tpredb,nb2=2)
```

B. S-PLUS Procedure for Calculating the Exact Confidence Intervals and the Profile of the Parameters

```
> parameters(simbsp1)<-list(p1=0.0463234427459947,p2=0.407598504097564,p3=0.106184833134302,
p4=2.40983993648298)
> D1MVG4<-deriv(y~ p1+(p2-p1)/(1+(p3*x)*p4),c("p1","p2","p3","p4"),function(p1,p2,p3,p4,x)NULL)
> modD1MVG4<-nlS(SIMUL.1~D1MVG4(p1,p2,p3,p4,H),data=simbsp1,trace=T)
0.00449479 : 0.0463234 0.407599 0.106185 2.40984
> modD1MVG4.prof<-profile(modD1MVG4)
> profile(modD1MVG4)
> conf.int<-function(profile.obj,variable.name,confidence.level=0.95)
+ (if(is.na(match(variable.name, names(profile.obj))))
+ stop(paste("Variable", variable.name, "not in the model")))
+ resid.df<-attr(profile.obj, "summary")["df"][[2]]
+ tstat<-qt(1-(1-confidence.level)/2, resid.df)
+ prof<-profile.obj[[variable.name]]
+ approx(prof,"tau", prof[, "par.vals"])
+ [,variable.name],c(-tstat, tstat))[[2]]
> conf.int(modD1MVG4.prof, "p1", conf=.95)
[1] 0.03398892 0.05833241
> conf.int(modD1MVG4.prof, "p2", conf=.95)
[1] 0.3827173 0.4373069
> conf.int(modD1MVG4.prof, "p3", conf=.95)
[1] 0.09200045 0.12564798
> conf.int(modD1MVG4.prof, "p4", conf=.95)
[1] 1.681325 NA
> ubp1<-0.0463234+(0.00563829*qt(0.975,14))
> lbp1<-0.0463234-(0.00563829*qt(0.975,14))
> ubp2<-0.407599+(0.0127062*qt(0.975,14))
> lbp2<-0.407599-(0.0127062*qt(0.975,14))
> ubp3<-0.106185+(0.00749918*qt(0.975,14))
> lbp3<-0.106185-(0.00749918*qt(0.975,14))
> ubp4<-2.40983994+(0.504646*qt(0.975,14))
> lbp4<-2.40983994-(0.504646*qt(0.975,14))
> (0.504646*(-2.144866879))+2.40983994
[1] 1.327441
> (0.504646*2.144866879)+2.40983994
[1] 3.492238
> ubp1
[1] 0.05841633
> lbp1
[1] 0.03423047
> ubp2
[1] 0.4348511
> lbp2
[1] 0.3803469
> ubp3
[1] 0.1222691
> lbp3
[1] 0.09010086
> ubp4
[1] 3.492198
> lbp4
```

```
[1] 1.327482
> modD1MVG4.prof<-profile(modD1MVG4,4)
> plot(modD1MVG4.prof,4)
> abline(-2.40983994/0.504646,1/0.504646,lty=3)
> abline(v=c(1.327441,3.492238),lty=3)
> abline(v=1.681325,lty=1)
> abline(h=c(-2.144866879,2.144866879),lty=3)
WARNING: Line out of bounds: x1 = 0.677647, y1 = 2.14487, x2 = 13.754, y2 = 2.14487
> text(1.9,1.5,"delta(p4)",adj=0)
> text(6,1.7,"tau(p4)",adj=0)
> text(1.1,-4.7,"lb",adj=0)
> text(3.5,-4.7,"ub",adj=0)
> text(2.44,0,"(2.40983994,0)",adj=0)
> text(9,-2.5,"t(14,0.025)",adj=0.5)
> text(9,2,"t(14,0.025)",adj=0.5)
> text(1.8,-4.7,"lb**",adj=0)
> title("Figure 4.1 : Profile t function tau(n) and linear profile delta (n) for the parameter n in (VG4)")
> (0.00749918*(-2.144866879))+0.10618483
[1] 0.09010009
> (0.00749918*2.144866879)+0.10618483
[1] 0.1222696
> modD1MVG4.prof<-profile(modD1MVG4,3)
> plot(profile(modD1MVG4,3))
> abline(-0.10618483/0.00749918,1/0.00749918,lty=3)
> abline(v=c(0.09200045,0.12564798),lty=1)
> abline(v=c(0.09010009,0.1222696),lty=3)
> abline(h=c(-2.144866879,2.144866879),lty=3)
> text(0.088,-4.7,"lb",adj=0)
> text(0.093,-4.7,"lb**",adj=0)
> text(0.119,-4.7,"ub",adj=0)
> text(0.126,-4.7,"ub**",adj=0)
> text(0.13,4.5,"delta(p3)",adj=0)
> text(0.147,4.5,"tau(p3)",adj=0)
> text(0.15,-2.5,"t(14,0.025)",adj=0)
> text(0.15,2.5,"t(14,0.025)",adj=0)
> text(0.1075,0,"(0.10618483,0)",adj=0)
> title("Figure 4.1 : Profile t function tau (alpha) and linear profile delta(alpha) for the parameter alpha and the model (VG4) w.r.t the first simulated set of sand")
> modD1MVG4.prof<-profile(modD1MVG4)
> profile(modD1MVG4)
> par(mfrow=c(2,2))
> plot(modD1MVG4.prof)
> mtext("Figure 3: Profile for the parameters in the model (VG4)", side=3, outer=T, line=-0.5)
```

REFERENCES

- Bates, D.M. and D.G. Watts, 1988. Nonlinear Regression Analysis and its Applications. John Wiley and Sons Inc., New York.
- Bunke, O., B. Dorge and J. Pohlzehl, 1998. Splus tools for model selection in nonlinear regression. *Comp. Stat.*, 13: 257-281.
- Bunke, O., B. Dorge and J. Pohlzehl, 1999. Model Selection, transformations and variance estimation in nonlinear regression. *Statistics*, 33: 197-240.
- El-Shehawy, S.A., 2008. On the selection of models in nonlinear regression. *AJMS.*, 1(1):1-13.
- King, L.G., 1965. Description of soil characteristics for partially saturated flow. *Soil Sci. Soc. Am. Proc.*, 29: 359-362.
- Kutilek, M. and D.R. Nielsen, 1994. Soil hydrology. *Catena Cremling-Destedt*.
- Seber, G.A.F. and C.J. Wild, 2005. Nonlinear Regression. *Jonh Wiley and Sons*, New York.
- Van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.*, 44: 892-898.
- Vereecken, H., J. Feyen, Maes and P. Darius, 1989. Estimating the soil moisture retention characteristic form texture, bulk density and carbon content. *Soil Sci.*, 148: 389-403.