Geoderma 195–196 (2013) 79–86

Contents lists available at SciVerse ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

# Soil hydraulic information for river basin studies in semi-arid regions

J.H.M. Wösten [a,*], S.J.E. Verzandvoort [a], J.G.B. Leenaars [b], T. Hoogland [a], J.G. Wesseling [a]

[a] Alterra — Wageningen University and Research Centre, P.O. Box 47, 6700 AA Wageningen, The Netherlands
[b] ISRIC — World Soil Information, P.O. Box 353, 6700 AJ Wageningen, The Netherlands

ARTICLE INFO

*Article history:*
Received 30 May 2012
Received in revised form 23 November 2012
Accepted 28 November 2012
Available online 28 December 2012

*Keywords:*
Pedotransfer function
Van Genuchten parameters
SOTER map unit
Cluster analysis
Hydrological response unit
Soil mapping

ABSTRACT

Water retention and hydraulic conductivity characteristics of the soil are indispensable for hydrological catchment modelling and for quantifying water limited agricultural production. However, these characteristics are often not available for regions and data scarcity for tropical zones is even bigger than for temperate zones. Use of pedotransfer functions which translate soil survey data into soil hydraulic characteristics is an interesting alternative in such cases. In this study, existing pedotransfer functions are identified and their performance is tested for the Limpopo river basin in Africa where distribution of limited water resources is a major challenge. The well performing pedotransfer function developed by Hodnett and Tomasella (2002) was used to translate the map units of the soil and terrain (SOTER) database for southern Africa into hydrological response units. Ten functional soil characteristics were calculated and clustering resulted in a reduction of the 713 SOTER map units for the Limpopo river basin to 14 hydrological response units. The resulting hydrological response unit map provides the required spatial information on soil physical input data, both water retention and hydraulic conductivity, for hydrological modelling of the river basin as well as for assessment of agricultural production. The developed procedure is an attractive approach for other, similar data scarce environments.



## 1. Introduction

Sophisticated models have been developed to describe movement of water and dissolved compounds through the soil. They have become indispensable tools for studies on for example crop growth, erosion, catchment hydrology and effects of climate change. The models are intended to quantify and integrate the most important physical, chemical, hydrological and biological processes in the unsaturated zone with the aim to underpin sustainable natural resource management. Use of models for research and management requires realistic input parameters governing retention and transport of water and chemicals in soils (Pachepsky and Rawls, 2004). At the same time these data are usually fragmented, do not cover the entire area under study, show different degrees of detail, have varying reliability, and are held in different institutes scattered throughout the world (Wösten et al., 2001a, 2001b). Often these models are used for larger river basins as well as for longer periods of time. As a consequence, spatial and temporal variability in hydraulic characteristics have a significant effect on model results. At the same time, these characteristics are notorious for the difficulties and high labour costs involved in measuring them. Thus there is a need to resort to estimating hydraulic characteristics from other more readily available soil data. Compared with other regions, data availability especially for tropical soils is limited (Minasny and Hartemink, 2011). To overcome the problem, existing pedotransfer functions are identified and tested. Next pedotransfer functions can be combined with existing, spatial information on soil and terrain components as recorded in the 'SOTER-based soil parameter estimates for Southern Africa' (version 1.0) (Batjes, 2004). This is a harmonised set of soil parameter estimates for Southern Africa, that was derived from the 1:2 M scale Soil and Terrain Database for Southern Africa (SOTERSAF ver. 1.0) and ISRIC-WISE soil profile database.[1] The database contains interpolated soil profile data for soil and terrain (SOTER) map units for intervals of 0.20 m in the soil profile up to 1.20 m depth.

The focus area of this project is the Limpopo river basin with an area of about 412,000 km$^2$ situated in Mozambique and upstream parts in Zambia, Botswana and South Africa. The main tributary of the Limpopo river is the Elephants river. Land use in the basin is mainly grassland, savannah and shrub land (68%), cropland covers about 26% of which only 1% is irrigated. Wetlands cover 3% and the remaining area is forest and urban. Agriculture in the Limpopo river basin is typically extensive with low input levels utilising mainly available natural resources. Notwithstanding the extensive use, irrigation for agriculture accounts for more than 50% of the total water demand in the Limpopo river basin (LBPTC, 2010). Rainfall in the basin is seasonal and unreliable. In dry years, the upper parts of the river conduct water for only 40 days or less. The basin is arid in the upper parts situated in the Kalahari desert, but becomes less arid

* Corresponding author. Tel.: +31 317 486478; fax: +31 317 419000.
*E-mail address:* henk.wosten@wur.nl (J.H.M. Wösten).

[1] http://www.isric.org/data/soter-based-soil-parameter-estimates-southern-africa-ver-10.

0016-7061/$ – see front matter 
http://dx.doi.org/10.1016/j.geoderma.2012.11.021

further downstream. The middle part drains the Waterberg massif, a region with semi-deciduous forest and low human population density. The lower part is fertile and heavily populated. Catastrophic floods after rainy seasons like in February 2000 are an occasional problem in the lower part. In a study by the National Directorate of Water of Mozambique (NDW, 1989) it was concluded that water abstractions from the river in upstream countries have increased significantly. In order to retain water in the Limpopo river basin 138 major dams have been constructed (Limpoporak, 2012) of which 13 large dams have a storage capacity exceeding 100 $Mm^3$; one in Mozambique; eight in South Africa; three in Zimbabwe and one in Botswana. At present the Limpopo river has been developed nearly to its full potential in Zimbabwe and the remaining runoff makes very little contribution to the flow in the river. Both Botswana and South Africa are separately planning new water developments including construction of new storage dams. For Mozambique information on planned future developments is largely lacking.

Hydrological modelling of the Limpopo river basin is required to optimise water allocation to food production, human and animal consumption, energy generation and nature preservation. To be able to run a hydrological model for the basin, spatial information is required on the soil hydraulic composition of the unsaturated zone. Within the overall objective to generate spatial soil hydraulic information as input data for modelling of catchment management and food production in the Limpopo river basin, specific objectives of the study are to:

- Identify which pedotransfer functions are available for tropical regions.
- Assess the performance of a selected pedotransfer function by comparing measured and predicted soil moisture contents for the Limpopo river basin as selected study area.
- Use the selected pedotransfer function to group existing SOTER soil mapping units for the area into hydrological response units which provide the required information for hydrological modelling.
- Prepare a map of the spatial distribution of the hydrological response units within the Limpopo river basin.

## 2. Methods

### *2.1. Selection and validation of a pedotransfer function*

Due to incompleteness in availability of directly measured soil hydraulic conductivity and water retention characteristics in existing soil databases for African soils, pedotransfer functions (PTF) need to be selected to make predictions for the soil hydraulic conductivity and water retention characteristics at different values of the soil moisture potential. Several pedotransfer functions were developed and tested for tropical soils (e.g. Hodnett and Tomasella, 2002; Minasny and Hartemink, 2011; Van den Berg et al., 1997). Because hydrological modelling of the Limpopo river basin requires a continuous function covering the complete range from saturation to wilting point, the PTF developed by Minasny and Hartemink (2011) could not be used as it predicts soil moisture contents for fixed soil moisture potentials of −10, −33 and −1500 kPa only. The PTF developed by Van den Berg et al. (1997) focuses primarily on Ferralsols and related soils which is reflected by the fact that the Al and Fe contents are used as predictor variables in their continuous PTF. Because the Limpopo river basin consists of a wide range of different soils, the Van den Berg et al. (1997) PTF is considered too restrictive for this study. Consequently, the continuous PTF of Hodnett and Tomasella (2002) is the most appropriate PTF for this study and its performance is tested by comparing predicted and observed soil moisture contents. This PTF predicts parameters of the Van Genuchten equation for the soil water retention curve (Van Genuchten, 1980):

$$\theta(h) = \theta_r + (\theta_s - \theta_r)/\left[1 + (\alpha \cdot h)^n\right]^m \tag{1}$$

where Θ(h) is the volumetric soil moisture content ($cm^3/cm^3$) at soil moisture potential h (cm), $\theta_r$ is the residual and $\theta_s$ is the saturated soil moisture content, α, m and n are parameters, with α in $cm^{-1}$, and n and m are dimensionless, m is derived from n using:

$$m = 1 - \frac{1}{n}. \tag{2}$$

The continuous PTF by Hodnett and Tomasella (2002) was developed using 771 horizons from tropical soils in the International Geosphere–Biosphere Programme Data and Information System (IGBP-DIS) soil database using multiple linear regression techniques. A second-order polynomial (Eq. (3)) is used to predict the parameters α, n, $\theta_s$ and $\theta_r$ of the Van Genuchten equation with the regression coefficients in Table 1. An empty cell in Table 1 means that the corresponding predictor variable is not used.

$$\begin{aligned} X_i = {} & a_{i,1} + a_{i,2}Sa + a_{i,3}Si + a_{i,4}Cl + a_{i,5}OC + a_{i,6}BD + a_{i,7}CEC + a_{i,8}pH \\ & + a_{i,9}Sa * Si + \cdots + a_{i,j}Si * Cl + a_{i,j+1}Sa^2 + \cdots + a_{i,n}Cl^2 \end{aligned} \tag{3}$$

where $X_i$ (i = 1 to 4) are the predicted values of ln(α), ln(n), $\theta_s$ and $\theta_r$, respectively. Sa, Si and Cl are the percentages of sand, silt and clay, OC is the percentage of organic carbon, CEC is the cation exchange capacity (cmol/kg), BD is the bulk density ($kg/dm^3$), and $a_{i,j}$ (i = 1 to 4, j = 1…n) are the coefficients derived by multiple linear regression (Hodnett and Tomasella, 2002).

The K(h) function (Eq. (4)) requires additional values for the saturated hydraulic conductivity ($K_s$) and for the shape parameter *l* (Van Genuchten, 1980).

$$K(h) = K_s \frac{\left(\left(1 + |\alpha.h|^n\right)^{1-1/n} - |\alpha.h|^{n-1}\right)^2}{\left(1 + |\alpha.h|^n\right)^{(1-1/n)(l+2)}}. \tag{4}$$

Values for $K_s$ were found in the Africa Soil Profiles Database 1.0 (Leenaars, 2012) for only 131 soil horizons. However, these values were inconsistent with the soil horizon's texture, with high average values of 52 cm/d for clay, 134 cm/d for clay loam, and 53 cm/d for loamy sand. As a result, values for $K_s$ were not accepted under the quality control criteria of the Africa Soil Profiles Database 1.0. Other data sources for $K_s$ covering the range of soil types occurring in the Limpopo river basin were not found. Also no values were found for the shape parameter *l* for African soils. Due to this lack of consistency as well as data scarcity, values for $K_s$ and *l* are based on values for the dominant soil texture documented for Europe (Wösten et al., 1999) and for The Netherlands (Wösten et al., 2001a, 2001b) and they are reported in Table 2.

The PTF is validated on a set of 92 soil horizons from the Africa Soil Profiles Database 1.0 (Leenaars, 2012), situated within the boundaries

**Table 1**
Regression coefficients in the PTF of Hodnett and Tomasella (2002).

| Predictor variable | ln(α) ($100\ kPa^{-1}$) | ln(*n*)∗100 (−) | $\theta_s$ ($100\ cm^3/cm^3$) | $\Theta_r$ ($100\ cm^3/cm^3$) |
|---|---|---|---|---|
| $a_{i,1}$ | −2.294 | 62.986 | 81.799 | 22.733 |
| Sand (%) | | | | −0.164 |
| Silt (%) | −3.526 | | | |
| Clay (%) | | −0.833 | 0.099 | |
| Organic carbon (%) | 2.440 | −0.529 | | |
| Bulk density ($kg/dm^3$) | | | −31.420 | |
| CEC (cmol/kg) | −0.076 | | 0.018 | 0.235 |
| pH | −11.331 | 0.593 | 0.451 | −0.831 |
| $Silt^2$ | 0.019 | | | |
| $Clay^2$ | | 0.0070 | | 0.0018 |
| Sand∗silt | | −0.014 | | |
| Sand∗clay | | | −0.0005 | 0.0026 |

**Table 2**
Values for $K_s$ and $l$ for the main FAO soil texture classes.

| Soil texture class | $K_s$ (cm/d) | $l$ (−) |
|---|---|---|
| Coarse | 50 | 1 |
| Fine | 4 | −2 |
| Medium | 10 | 0.5 |
| Very fine | 3 | −4 |
| Organic | 5 | 0 |

of the Limpopo river basin. This database contains measured values of soil moisture content for specific values of the soil moisture potential. Selected horizons had at least 5 data points between saturation and wilting point. The criterion of multiple data points on the measured water retention curve is used to guarantee that the PTF would perform well throughout the range of soil moisture potentials which is most important for agricultural production. Based on the available data, the following soil moisture potentials are selected: pF 0.0, 2.0, 2.5, 3.7 and 4.2. All 92 soil horizons had predictor soil properties (Table 3) within the range for which the PTF of Hodnett and Tomasella (2002) was developed and tested. This criterion is a prerequisite for the application of any PTF (Minasny and Hartemink, 2011).

The performance of the PTF is tested by using an integrated measure over the full range of water retention points for each soil horizon. The mean error (ME) (Eq. (5)) and root mean square error (RMSE) (Eq. (6)) are used to assess the performance of the PTF. ME expresses the systematic error (or bias) of the predicted value. RMSE expresses the combined effect of systematic and random error.

$$\mathrm{ME} = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \tag{6}$$

where n = number of paired values for measured ($x_i$) and predicted ($y_i$) moisture content. Mean error (ME) values close to zero indicate that measured and predicted moisture contents do not differ systematically from each other or, equivalently, that there is no consistent bias. Values that differ from zero indicate the presence of systematic deviation or bias. The deviation of ME from zero is tested with a one-sample *T*-test (two-tailed, with a confidence interval of 95%) using the complete dataset of measured and predicted moisture contents without distinction according to soil moisture potentials. Predicted moisture contents (y) are plotted versus measured moisture contents (x) and are best presented in a graph which includes the 1:1 line. The root mean square error (RMSE) is a measure for the scatter of the data points around the 1:1 line. Low RMSE values indicate little scatter, and high RMSE values indicate large scatter. Low RMSE values also imply low ME values.

### *2.2. Application of PTF to the Limpopo river basin*

In order to obtain soil hydraulic characteristics for the Limpopo river basin, the PTF is applied to the soil profile predictor variables (Table 4), available for intervals of 0.20 m to 1.20 m depth from the 'SOTER-based soil parameter estimates for Southern Africa' (2006, version 1.0) (Batjes, 2004). These predictor variables also fall within the range for which the PTF of Hodnett and Tomasella (2002) was developed and tested. The SOTER database contains 713 interpolated soil profile data for soil and terrain (SOTER) map units in the river basin.

Apart from soil and parent material characteristics, terrain characteristics were also used to delineate the SOTER map units such as distinctive, often repetitive, patterns of landform, lithology, surface form and slope as shown in Fig. 1. Consequently, the distinction in soil hydraulic characteristics between map units of the Limpopo river basin, as derived with the PTF, is partly based on these terrain characteristics as well.

### *2.3. Clustering of SOTER map units into hydrological response units*

The Hodnett and Tomasella PTF is used to assign soil hydraulic information to each of the 713 SOTER map units. However, it is argued that each map unit does not necessarily behave differently in a soil hydraulic sense and therefore the map unit could be clustered into larger units for more efficient hydrological modelling of the Limpopo river basin. To cluster SOTER map units into hydrological response units ten functional soil characteristics (Table 5) are calculated.

To prevent strongly correlated characteristics from dominating the clustering of map units, the similarity between characteristics is examined from correlations, and from principal component analysis. Since soil hydraulic characteristics are strongly correlated with soil texture as is also indicated in Tables 1 and 2, the 713 SOTER soil map units are grouped into three texture classes prior to clustering. This a priori grouping is based on the dominant soil texture class in the top 80 cm of the soil (FAO, 1990). The procedure is analogous to the functional hydraulic clustering of Dutch soils developed by Wösten et al. (2012) and resulted in the following three groups: coarse, fine and very fine (Table 6).

After grouping the 713 SOTER map units within the three main soil textural classes, cluster analysis was based on the calculated functional soil characteristics, using the K-means cluster algorithm first reported by MacQueen (1967). The K-means algorithm is a method of cluster analysis which aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean. In this case the algorithm starts with (arbitrarily chosen) centroids of the number of desired clusters. Next all objects are assigned to the cluster for which the distance is smallest. Once all objects have been assigned to a cluster, the centroid is computed again. Allocation of all objects based on the new centroids is done and the process is repeated iteratively until no more changes occur. After obtaining a set of k centroids, k clusters are constructed by assigning each object to the nearest centroid. The aim is to find k clusters which minimise

**Table 3**
Soil properties of the 92 horizons used for validation of the continuous PTF of Hodnett and Tomasella (2002).
Source: Africa Soil Profiles Database 1.0 (Leenaars, 2012).

| | Total sand content (w%) | Silt content (w%) | Clay content (w%) | Bulk density (kg/dm$^3$) | pH $H_2O$ (−) | CEC soil (cmol/kg) | Organic Carbon (%) |
|---|---|---|---|---|---|---|---|
| Min | 4 | 4 | 1 | 0.9 | 4 | 2 | 0.0 |
| Max | 94 | 67 | 83 | 1.7 | 8.7 | 45 | 3.8 |
| Stdev | 27 | 14 | 23 | 0.2 | 0.9 | 9.0 | 0.8 |

**Table 4**
Soil properties of the 713 SOTER map units in the Limpopo river basin used as predictor variables in the continuous PTF of Hodnett and Tomasella (2002).

| | Total sand content (w%) | Silt content (w%) | Clay content (w%) | Bulk density (kg/dm$^3$) | pH $H_2O$ (−) | CEC soil (cmol/kg) | Organic carbon (%) |
|---|---|---|---|---|---|---|---|
| Min | 1 | 0 | 0 | 0.97 | 4.1 | 0.1 | 0.04 |
| Max | 99 | 57 | 81 | 1.86 | 10.8 | 71.7 | 2.8 |
| Stdev | 17 | 9 | 13 | 0.11 | 1.0 | 9.3 | 0.3 |

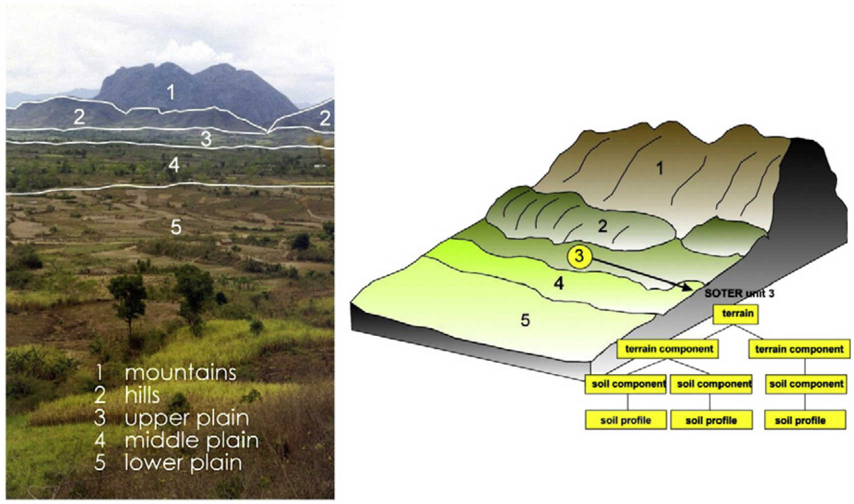


**Fig. 1.** Delineation of SOTER units.
Source: ISRIC — World Soil Information.

the sum of the dissimilarities of the objects within each cluster (Hartigan and Wong, 1979). The optimum number of clusters in each group is assessed based on the following three criteria:

1. Maximum ratio of between cluster variance and within cluster variance or Calinsky criterion C (Calinsky and Harabasz, 1974; Vendramin et al., 2009):

$$C = \frac{B(k)/(k-1)}{W(k)/(n-k)} \tag{7}$$

   where $B(k)$ is the sum of squares between clusters, $W(k)$ is the sum of squares within a cluster, $n$ is the number of observations and $k$ is number of clusters.
2. Maximum difference between cluster centres of the normalised functional soil characteristics (visual inspection of bar plots in Fig. 3b).
3. Minimal areal coverage of clustered map units. If clusters cover only small areas they could be merged with larger clusters.

These criteria are assessed consecutively; first the number of clusters with the largest Calinsky ratio are selected (see Fig. 3a). Here the largest Calinsky ratio of nearly 140 is found for the division in 6 groups. The centre values for the normalised functional soil characteristics of these 6 groups (Fig. 3b) are visually inspected. If cluster centres are very similar, there is little reason to distinguish the clusters, and these clusters could be merged, resulting in fewer groups. The greater the difference between cluster centres, the better the clustering reflects the variation in soil hydraulic characteristics of the area considered. Based on the result of the first two steps, a generalised map of grouped soil hydraulic units, also called hydrological response units, is created from the map of SOTER units. The areal extent and spatial configuration of the grouped soil units are visually inspected from the map. In order to obtain soil hydraulic characteristics for the clustered units a 'typical' soil profile needs to be allocated to each unit. These typical soil profiles are selected from the SOTER database according to the following principles:

1. The summed distances between the normalised functional soil characteristics for the soil profile and the cluster centre are minimal.
2. If there is more than one typical soil profile with a minimum distance to the cluster centre, the profile with the largest areal coverage in the cluster is chosen.
3. The typical soil profile represents a SOTER unit that is realistically situated among other SOTER map units in the cluster (see Fig. 1 for a fictitious SOTER unit 3).

**Table 5**
Calculated functional soil characteristics used for clustering of SOTER map units into hydrologic response map units.

| Functional soil characteristic | Description |
|---|---|
| C | Hydraulic resistance (day) for vertical, saturated groundwater flow |
| kD | Transmissivity ($cm^2$/d) for horizontal water flow |
| EAWr | Easy available water storage ($100 \times cm^3/cm^3$) in the root zone between pF 2.0 and pF 2.6 |
| DAWr | Difficult available water storage ($100 \times cm^3/cm^3$) in the root zone between pF 2.6 and pF 4.2 |
| EAWs | Easy available water storage ($100 \times cm^3/cm^3$) in the subsoil zone between pF 2.0 and pF 2.6 |
| DAWs | Difficult available water storage ($100 \times cm^3/cm^3$) in the subsoil zone between pF 2.6 and pF 4.2 |
| ZCrit1 and ZCrit2 | Maximal depth of the groundwater level (in cm) at which a flux of 1 mm/d and 2 mm/d respectively can reach the lower boundary of the root zone which is at pF 4.2 |
| VCrit1 and VCrit2 | Moisture deficit (mm) at a flux of 1 mm/d and 2 mm/d respectively needed to saturate the soil horizons between the calculated groundwater level and the lower boundary of the root zone |

## 3. Results

### 3.1. Validation of continuous PTF

The continuous PTF proposed by Hodnett and Tomasella (2002) is validated using the Africa Soil Profiles Database 1.0 by first calculating the parameters in the Van Genuchten equation: $\theta_r$, $\theta_s$, $\alpha$ and m using Eqs. (1) and (2), and the regression coefficients in Table 1. Next, soil

**Table 6**
Grouping of SOTER map units for the Limpopo river basin in FAO texture classes.

| Generalised texture | Description | Nr of SOTER map units | Area ($km^2$) | Area (% of total) |
|---|---|---|---|---|
| Coarse | Clay < 18% and sand > 65% | 358 | 414490 | 69.0% |
| Fine | 35% < clay < 60% | 346 | 176932 | 29.5% |
| Very fine | 60% < clay | 9 | 8961 | 1.5% |

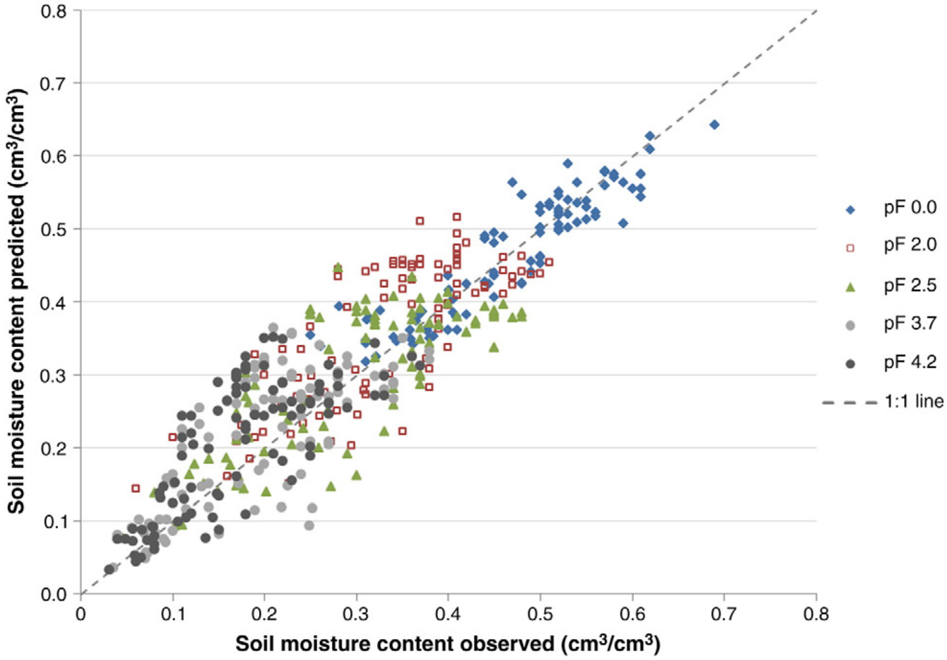


**Fig. 2.** Predicted versus observed moisture contents of 92 soil profile horizons in the Limpopo river basin. Measured moisture content was taken from the African Soil Profiles Database 1.0 (Leenaars, 2012) and moisture content was predicted with the continuous PTF of Hodnett and Tomasella (2002).

moisture contents are predicted for soil moisture potentials of pF 0.0, 2.0, 2.5, 3.7 and 4.2 using Eq. (1). Results are displayed in Fig. 2.

The continuous PTF of Hodnett and Tomasella (2002) yields a good prediction of measured soil moisture contents for the full range of soil moisture potentials (Fig. 2). The mean error (ME), root mean squared error (RMSE) and coefficient of determination ($R^2$) of the predicted versus measured soil moisture contents are 0.017 ($cm^3/cm^3$), 0.064 ($cm^3/cm^3$) and 0.81 respectively. The one-sample *T*-test on the mean error shows a significant deviation from zero of the mean error. However, this deviation is small in terms of soil moisture content (0.017 $cm^3/cm^3$) and irrelevant for most applications of PTFs in river basin studies. Consequently, the PTF is considered to give good predictions of the soil moisture content without systematic bias. Predicted values for residual soil moisture contents ($\theta_r$) and also for saturated soil moisture contents ($\theta_s$) are relatively high (for an example see Table 8). Similar results were obtained for other soil profiles in the Limpopo river basin and this is consistent with reports for tropical soils by Van den Berg et al. (1997), Hodnett and Tomasella (2002) and Minasny and Hartemink (2011). The latter contribute this to relatively high Al and Fe contents of clay minerals in tropical soils as compared to contents in soils in temperate regions.

### 3.2. Clustering of SOTER map units into soil hydraulic units

Table 7 shows that the calculated functional soil characteristics EAWs, ZCrit2 and VCrit2 are strongly correlated to the other characteristics. Therefore, these three functional soil characteristics are excluded and clustering is based on the functional soil characteristics C, kD, EAWr, DAWr, DAWs, ZCrit1, and VCrit1.

As an example, results of the clustering of SOTER map units into soil hydraulic units for the fine textured soil group are shown in Fig. 3.

The colours in the top part (Fig. 3a) indicate the group that each of the 346 fine textured SOTER map units is assigned to. Fig. 3a shows on the horizontal axis the objects that are grouped and on the vertical axis the number of groups distinguished. The Calinsky criterion shows a peak for the optimal number of 6 soil hydraulic groups. Thus based on resemblance of calculated functional soil characteristics, the 346 fine textured SOTER map units can be clustered into 6 soil hydraulic units, or hydrological response units. The bar lengths in the 6 bar plots in the bottom part (Fig. 3b) indicate the normalised values of calculated functional soil characteristics for the centres of each of the 6 groups. For the 358 coarse textured soils, 6 clusters appear to be optimal as well. For the very fine textured soils, the

**Table 7**
Correlation matrix of the ten calculated functional soil characteristics.

| | C | kD | EAWr | DAWr | EAWs | DAWs | ZCrit1 | VCrit1 | ZCrit2 | VCrit2 |
|---|---|---|---|---|---|---|---|---|---|---|
| C | 1.00 | −0.68 | −0.47 | 0.05 | −0.49 | −0.16 | −0.07 | −0.45 | −0.26 | −0.54 |
| kD | −0.68 | 1.00 | 0.23 | −0.51 | 0.30 | −0.30 | 0.05 | 0.69 | 0.19 | 0.74 |
| EAWr | −0.47 | 0.23 | 1.00 | 0.48 | 0.92 | 0.66 | −0.04 | 0.59 | 0.05 | 0.66 |
| DAWr | 0.05 | −0.51 | 0.48 | 1.00 | 0.46 | 0.92 | 0.16 | −0.23 | 0.16 | −0.23 |
| EAWs | −0.49 | 0.30 | 0.92 | 0.46 | 1.00 | 0.72 | 0.03 | 0.59 | 0.14 | 0.62 |
| DAWs | −0.16 | −0.30 | 0.66 | 0.92 | 0.72 | 1.00 | 0.15 | 0.00 | 0.20 | 0.00 |
| ZCrit1 | −0.07 | 0.05 | −0.04 | 0.16 | 0.03 | 0.15 | 1.00 | 0.27 | 0.98 | 0.09 |
| VCrit1 | −0.45 | 0.69 | 0.59 | −0.23 | 0.59 | 0.00 | 0.27 | 1.00 | 0.34 | 0.97 |
| ZCrit2 | −0.26 | 0.19 | 0.05 | 0.16 | 0.14 | 0.20 | 0.98 | 0.34 | 1.00 | 0.19 |
| VCrit2 | −0.54 | 0.74 | 0.66 | −0.23 | 0.62 | 0.00 | 0.09 | 0.97 | 0.19 | 1.00 |

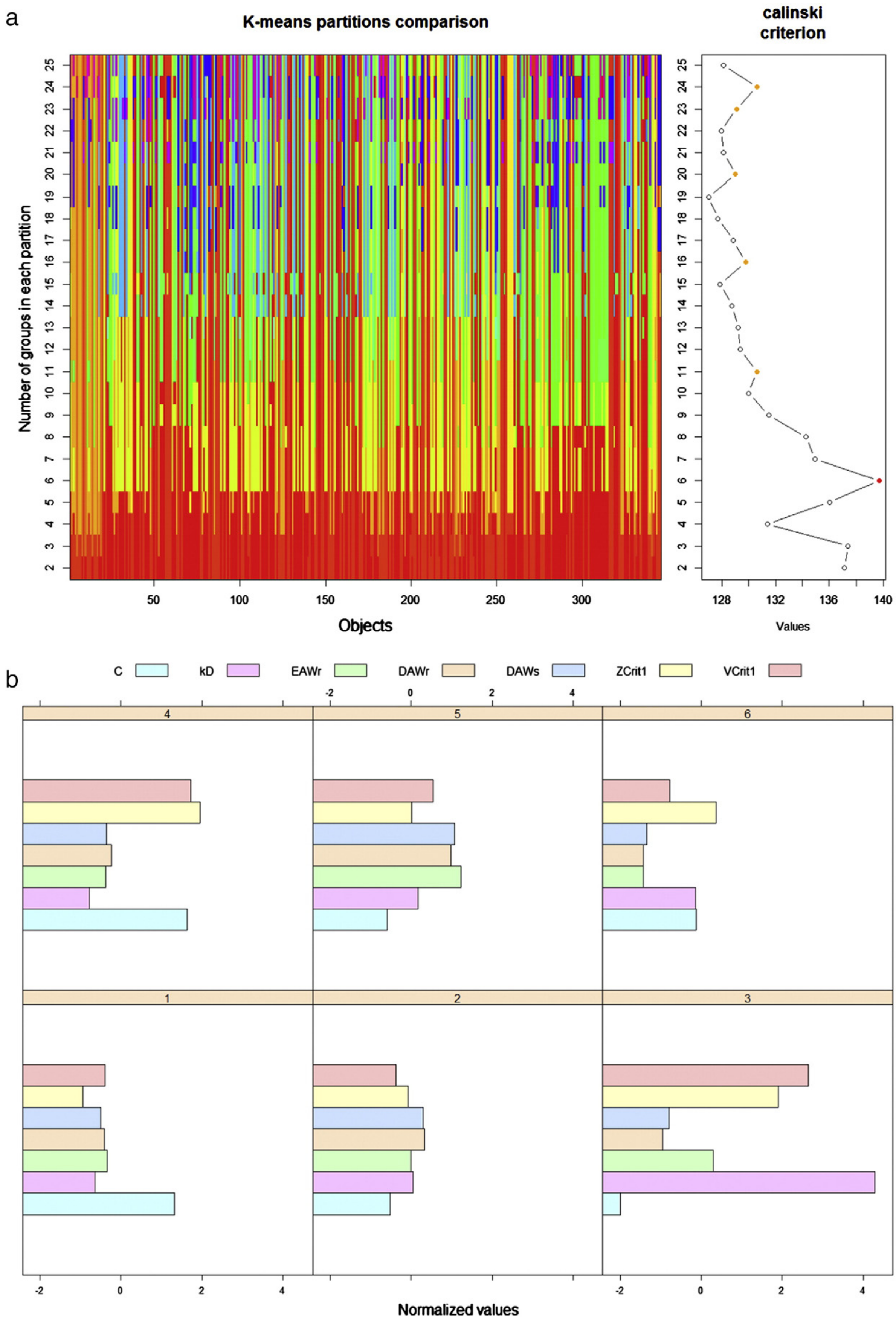


**Fig. 3.** Clustering of 346 SOTER map units (a) into 6 hydrological response map units (b) for the group of fine textured soils.

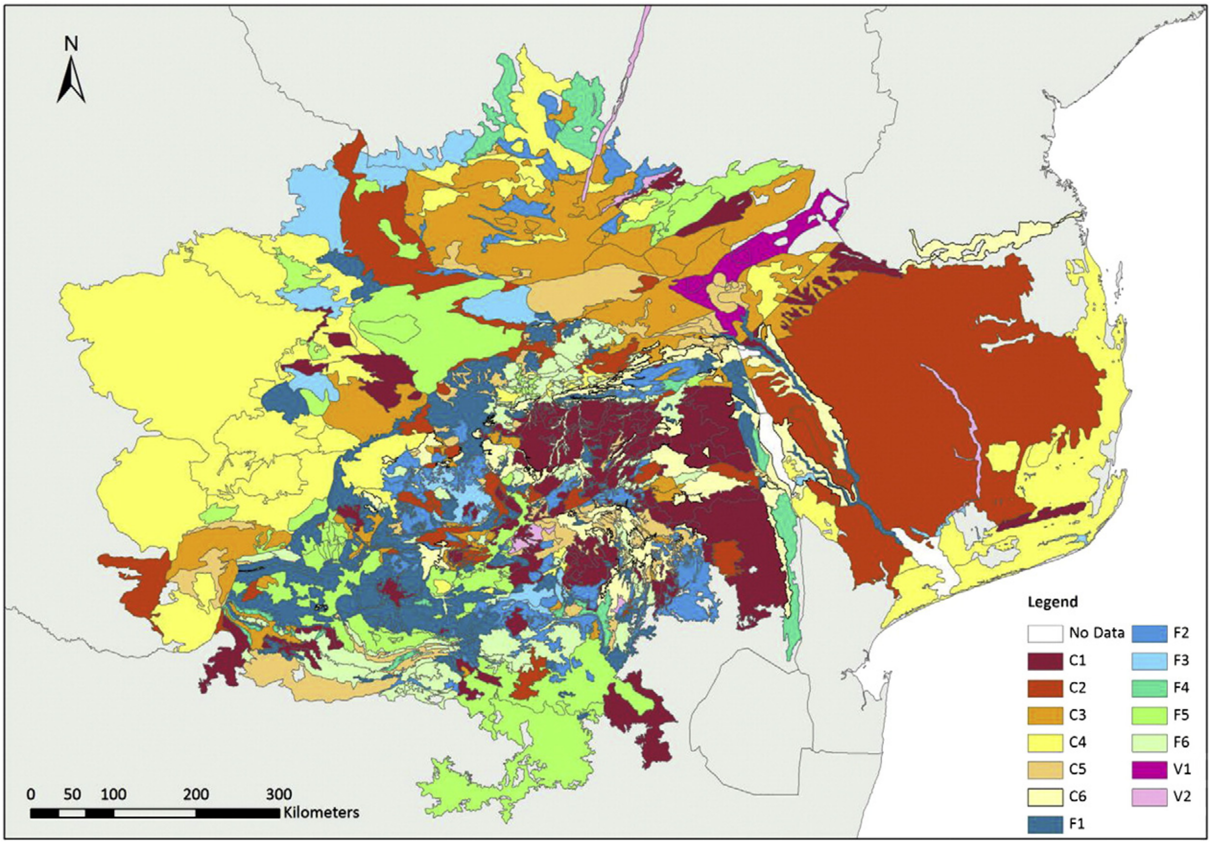


**Fig. 4.** Spatial distribution of hydrological response units in the Limpopo river basin, grouped by SOTER texture classes coarse (C1–C6), fine (F1–F6) and very fine (V1–V2). For the white areas ("null") no soil hydraulic data could be derived within the application boundaries of the PTF.

Calinsky criterion also has its maximum value at 6 clusters, however, since this group contains only 9 SOTER map units, this would result in clusters with only one unit. Therefore, the number of clusters for this group is limited to 2. As a result the 713 SOTER map units for the Limpopo river basin are grouped into 14 hydrological response map units thereby making the hydrological modelling of the river basin much more efficient. With the conversion of SOTER map units in hydrological response map units as key, the map shown in Fig. 4 is produced indicating the soil hydraulic composition of the unsaturated zone of the Limpopo river basin.

Clusters of the coarse textured group cover large areas on the borders and in the central part of the river basin which is in accordance with the soil geography of the area. Consequently, by including the texture classes in the clustering procedure for the hydrological response units, it is assured that the main soil and landscape characteristics are reflected in the clusters.

An example of soil hydraulic characteristics for a soil profile for a cluster in the coarse texture group is given in Table 8.

## 4. Conclusions and discussion

Based on this study, the following conclusions and discussion points are drawn:

- In the case of the Limpopo river basin, clustering of SOTER map units into hydrological response map units results in a reduction of the number of units from 713 to 14. As a result, hydrological modelling becomes more efficient.
- It is attractive to cluster SOTER map units into hydrological response units based on calculated functional soil characteristics as these are the key characteristics that matter in hydrological modelling.
- The developed procedure is an attractive approach for other, similar data scarce environments as the Limpopo river basin and certainly for Sub-Sahara Africa.
- The pedotransfer function developed by Hodnett and Tomasella (2002) performs very well in estimating measured soil moisture contents at various soil moisture potentials for horizons of soil profiles

**Table 8**
Example soil hydraulic data for the soil profile for Cluster C4-4 in the coarse texture group.

| Cluster C4: SOTER map unit ZW95 | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| SOTER map unit name and horizon | Depth of top (cm below surface) | Depth of bottom (cm below surface) | Average $\theta_r$ (100 cm$^3$/cm$^3$) | Average $\theta_s$ (100 cm$^3$/cm$^3$) | Average $K_s$ (cm/d) | Average $\alpha$ (100 kPa$^{-1}$) | Average $l$ (−) | Average n (−) |
| ZW95-D1 | 0 | 20 | 11.1505 | 42.44333 | 30 | 0.309358 | 0.75 | 1.574305 |
| ZW95-D2 | 20 | 40 | 5.641 | 38.0138 | 50 | 0.427504 | 1 | 1.728119 |
| ZW95-D3 | 40 | 60 | 7.3544 | 38.5088 | 50 | 0.456901 | 1 | 1.712827 |
| ZW95-D4 | 60 | 80 | 7.8337 | 38.5627 | 50 | 0.462108 | 1 | 1.701299 |
| ZW95-D5 | 80 | 100 | 7.6764 | 38.51312 | 50 | 0.427009 | 1 | 1.674886 |

situated in the Limpopo river basin. In addition to the water retention curve also the saturated and unsaturated hydraulic conductivity curve can be predicted. This feature makes the PTF approach a versatile instrument in generating soil hydraulic information for modelling.

- In order to test the PTF with respect to hydraulic conductivity it is desirable that in addition to soil moisture contents also saturated and unsaturated hydraulic conductivity are measured for soil horizons in the Limpopo river basin, since the application of the PTF to predict conductivity curves depends on the availability of values for parameters $l$ and $K_s$ in the Van Genuchten equation, for which only few values were found in the literature on tropical soils.
- Combining information on soils and terrains with well performing pedotransfer functions is a powerful tool for generating spatial soil physical input data for hydrological modelling of basins.
- It should be kept in mind that in this project the purpose of clustering was to arrive at a limited number of hydrological response units. In case the purpose would have been clustering into a limited number of soil chemical response units for distinguishing nutrient retention characteristics, organic matter content instead of soil texture could have been the key soil property to look at. However, in both cases basic soil information as recorded in soil survey together with the appropriate pedotransfer function provide the data to arrive at a functional clustering of soil units.
- Due to the inclusion of soil texture information prior to clustering, the spatial distribution of hydrological response units in the Limpopo river basin reflects part of the soil geographical diversity of the basin.
- The spatial distribution of hydrological response units in the Limpopo river basin still needs to be validated by hydrological modelling of the entire basin, e.g. by comparing calculated with measured water discharge. However, not only the soil hydraulic composition of the unsaturated zone affects the hydrological modelling results but also other input parameters such as rainfall and cropping pattern. It is therefore recommended to test the sensitivity of the modelling results for multiple input parameters, including soil hydraulic distribution.

## Acknowledgement

This research was conducted with the support of the Dutch Ministry of Economic Affairs, Agriculture and Innovation in the research cluster More food on smaller foot. Project code KB-11-002-001.

## References

Batjes, N.H., 2004. SOTER-based soil parameter estimates for Southern Africa (ver. 1.0). Report 2004/04 ISRIC — World Soil Information, Wageningen. (http://www.isric.org/isric/webdocs/docs/ISRIC_Report_2004_04.pdf).

Calinsky, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Communications in Statistics 3 (1), 1–27 (2).

Food and Agriculture Organisation (FAO), 1990. Guidelines for Soil Description, 3rd edn. FAO / ISRIC, Rome.

Hartigan, J.A., Wong, M.A., 1979. A K-means clustering algorithm. Applied Statistics 28, 100–108.

Hodnett, M.G., Tomasella, J., 2002. Marked differences between van Genuchten soil moisture-retention parameters for temperate and tropical soils: a new water-retention pedo-transfer function developed for tropical soils. Geoderma 108, 155–180.

Leenaars, J.G.B., 2012. Africa soil profiles database, Version 1.0. A Compilation of Georeferenced and Standardised Legacy Soil Profile Data for Sub-Saharan Africa (with dataset). ISRIC Report 2012/03. Africa Soil Information Service (AfSIS) project and ISRIC — World Soil Information, Wageningen, The Netherlands (148 pp.; 15 fig.; 10 tab.; 25 ref).

Limpopo Basin Permanent Technical Committee (LBPTC), 2010. Joint Limpopo river basin study scoping phase. Final Report. BIGCON Consortium.

Limpoporak, 2012. Website http://www.limpoporak.com/en/default.aspx.

MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. 1. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press, pp. 281–297.

Minasny, B., Hartemink, A.E., 2011. Predicting soil properties in the tropics. Earth-Science Reviews 106, 52–62.

NDW, 1989. Review of the Water Resources of Maputo Province. National Directorate of Water, People's Republic of Mozambique . (Mission report by Euroconsult, Arnhem).

Pachepsky, Ya, Rawls, W.J., 2004. Development of pedotransfer functions in soil hydrology. Developments in Soil Science, 30. Elsevier.

Van den Berg, M., Klant, E., van Reeuwijk, L.P., Sombroek, G., 1997. Pedotransfer functions for the estimation of water retention characteristics of Ferralsols and related soils. Geoderma 79, 161–180.

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

Vendramin, L., Campello, R.J.G.B., Hruschka, E.R., 2009. On the Comparison of Relative Clustering Validity Criteria. Society for Industrial and Applied Mathematics. (http://www.siam.org/proceedings/datamining/2009/dm09_067_vendraminl.pdf).

Wösten, J.H.M., Lilly, A., Nemes, A., Le Bas, C., 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90, 169–185.

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001a. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology 251, 123–150.

Wösten, J.H.M., Veerman, G.J., de Groot, W.J.M., Stolte, J., 2001b. Water retention and hydraulic conductivity characteristics of top and subsoils in The Netherlands: the Staring series. Alterra report, 153 (Wageningen, the Netherlands (in Dutch)).

Wösten, J.H.M., Hoogland, T., Massop, H., de Vries, F., Vroon, H.R.J., Wesseling, J.G., 2012. New soil physical schematisation of The Netherlands. Alterra Report (Wageningen, The Netherlands (in Dutch)).