Published September 12, 2019

Original Research

Core Ideas

- Measurements of field-scale saturated hydraulic conductivity ($<K_s^f>$) were performed.
- Two pedotransfer functions for field-scale $<K_s^f>$ were developed.
- A map of field-scale $<K_s^f>$ at catchment scale was obtained using two approaches.

A Pedotransfer Function for Field-Scale Saturated Hydraulic Conductivity of a Small Watershed

Tommaso Picciafuoco, Renato Morbidelli, Alessia Flammini,* Carla Saltalippi, Corrado Corradini, Peter Strauss, and Günter Blöschl

Classical experimental techniques to determine point values of saturated hydraulic conductivity (K_s) are complex and time consuming; therefore, the development of pedotransfer functions, PTFs, to derive K_s from easily available soil properties is of great importance. However, PTFs have been generally developed at the local scale, while hydrological modeling requires K_s estimates at larger scales. A small Austrian catchment, where detailed soil characteristics were available, was selected to address this issue. Values of field-scale saturated hydraulic conductivity ($<K_s^f>$), observed in a number of catchment areas by double-ring infiltrometers, were used to develop two PTFs, one by multiple linear regression (PTF_{MLR}) and one by ridge regression (PTF_R). Training and validation of the PTFs in the monitored areas indicate that the PTF_R provides better outcomes with smaller average errors. This suggests that the ridge regression is a valid alternative to the classical multiple linear regression technique. Predictions of $<K_s^f>$ by the PTFs in the remaining areas, where infiltration measurements were not performed, were also made to obtain a map of $<K_s^f>$ for the whole catchment. Two alternative approaches were used: Method A—soil properties were first interpolated and then the PTFs applied; Method B—the PTFs were first applied to sites with available soil properties and then interpolated. The maps of $<K_s^f>$ obtained by the PTF_{MLR} are not representative of the $<K_s^f>$ spatial variability. On the other hand, the map generated by the PTF_R with Method A is consistent with catchment morphology and soil characteristics.

Abbreviations: ANN, artificial neural network; FDF, frequency density function; GCV, generalized cross-validation; GMER, geometric mean error ratio; MLR, multiple linear regression; PTF, pedotransfer function; VIF, variance inflation factor.

T. Picciafuoco, R. Morbidelli, A. Flammini, C. Saltalippi, and C. Corradini, Dep. of Civil and Environmental Engineering, Univ. of Perugia, via G. Duranti 93, Perugia, Italy; T. Picciafuoco and G. Blöschl, TU Wien, Centre for Water Resource Systems, Karlsplatz 13, A-1040, Vienna, Austria; G. Blöschl, TU Wien, Institute of Hydraulic Engineering and Water Resources Management, Karlsplatz 13, A-1040 Vienna, Austria; P. Strauss, Federal Agency for Water Management, Institute for Land and Water Management Research, A-3252 Petzenkirchen, Austria. *Corresponding author (alessia.flammini@unipg.it).

Received 13 Feb. 2019.
Accepted 22 July 2019.

Citation: Picciafuoco, T., R. Morbidelli, A. Flammini, C. Saltalippi, C. Corradini, P. Strauss, and G. Blöschl. 2019. A pedotransfer function for field-scale saturated hydraulic conductivity of a small watershed. Vadose Zone J. 18:190018. doi:10.2136/vzj2019.02.0018

Saturated hydraulic conductivity, K_s, is a crucial soil property in most hydrological models. Its accurate estimate is fundamental in the representation of local infiltration into both homogeneous (Philip, 1969; Te Chow et al., 1988; Corradini et al., 1997) and layered soils (Corradini et al., 2000; Morbidelli et al., 2014). The assessment of K_s is also essential at the field scale to properly represent the effect of soil heterogeneity on the infiltration process in distributed rainfall–runoff models. Smith and Goodrich (2000) presented a model that can simulate areal-average infiltration of rainfall in areas exhibiting random variation in K_s through the point infiltration model of Parlange et al. (1982) and the Latin hypercube sampling method. Govindaraju et al. (2001) and Corradini et al. (2011) formulated two models for field-scale infiltration into vertically homogeneous and layered soils, respectively, characterized by a random spatial variability of K_s with a probability density function of the lognormal type. A study area was partitioned into a finite number of regions, each characterized by the areal-average value of K_s, $<K_s>$, and the associated coefficient of variation, $CV(K_s)$. The expected field-scale infiltration rate in each region was expressed as a function of an expected time through the values of $<K_s>$ and $CV(K_s)$. The model by Govindaraju et al. (2001) was then extended to account for the run-on process (Morbidelli et al., 2006). These models for field-scale infiltration have in common the essential requirement of a spatial characterization of K_s, which in the

absence of sampled data can be synthetically approximated by a lognormal random field with values of $<K_s>$ and CV(K_s) fixed in advance. Therefore, at least an experimental characterization of K_s is required to assess the last two quantities.

Measurements of K_s are often time consuming and exhibit a high spatial variability linked with land use or soil type (Papanicolaou et al., 2015; Baiamonte et al., 2017). Values of K_s observed in situ through classical devices are always influenced by local discontinuities in the soil matrix, such as cracks, worm holes, or roots, which often determine preferential flow paths for infiltration (Picciafuoco et al., 2019).

Recently, a global database of infiltration measurements, which is useful to investigate many problems linked with the determination of K_s, was assembled by Rahmati et al. (2018).

An alternative approach to estimate K_s at the point scale relies on the use of pedotransfer functions (PTFs) (Bouma, 1987) that allow to predict hard-to-measure soil properties, such as K_s, from readily available information (e.g., soil texture or bulk density). Vereecken (1989) classified soil hydraulic function related PTFs into two groups: point-based PTFs that directly provide soil hydraulic properties and parametric PTFs that predict parameters involved in different hydrological models. The mathematical complexity of existing PTFs varies from simple tables that provide hydraulic parameters for particular textural classes (Wösten et al., 1995; Tietje and Hennings, 1996), hereafter called class PTFs, to linear-regression-based approaches (Minasny et al., 1999; Pachepsky et al., 2001) and artificial neural network (ANN) models (Schaap et al., 2001; Parasuraman et al., 2006; Sedaghat et al., 2016). Wösten et al. (2001) and more recently Van Looy et al. (2017) provided a detailed review of the developed PTFs. Usually, PTFs are deduced from the relationships between input data available in soil databases (e.g., textural properties or topographic variables) and K_s. Schaap et al. (2001) developed an ANN-based software, ROSETTA, which allows estimation of K_s by application of five hierarchical PTFs generated with a large number of data obtained from three databases (Schaap and Leij, 1998). Wösten et al. (1999) developed a database of European soil properties (HYPRES) and used it to derive PTFs for K_s. Pedotransfer functions for applications on spatial datasets with European coverage were also provided by Tóth et al. (2014). The use of large databases allows identification of a wide range of combinations and relationships among soil properties, in principle making the obtained PTFs applicable in different scenarios. However, the PTFs available in the literature are not always applicable with acceptable accuracy to different regions (Tietje and Tapkenhinrichs, 1993; Kern, 1995; Cornelis et al., 2001; Wagner et al., 2001; Nemes et al., 2003). Therefore, attempts have been made to recalibrate published PTFs by adapting the parameters to soil conditions different from those used in their development (Abdelbaki, 2016).

As mentioned above, to apply models for estimating field-scale infiltration, K_s as a random variable can be represented through $<K_s>$ and CV(K_s) of each catchment region. In principle, the last two quantities could be estimated using PTFs. However, the PTFs already developed allow K_s to be derived only at the local scale, and therefore one may need to use a procedure to upscale the point-based K_s values to the field scale by taking into account the spatial resolution of the available soil properties. Alternatively, different types of PTFs should be formulated to provide a direct estimate of $<K_s>$. In addition, criteria useful to predict CV(K_s) at the field scale by PTF should be defined.

The upscaling would give higher errors in predictions because of the inability to allow for all the elements that produce K_s spatial variability such as local discontinuities in the soil matrix. Therefore, the main objective of this study was to formulate PTFs to deal with a direct estimate of $<K_s>$ at the field scale. Two PTFs able to predict values of $<K_s>$ representative of the different areas belonging to a particular catchment were developed. In this context, the basic element is represented by K_s measurements performed by double-ring infiltrometers at a sufficient number of plots as defined by Picciafuoco et al. (2019). A small Austrian watershed with known detailed soil characteristics was selected. The PTF development relied on both the classical technique of multiple linear regression and the ridge regression approach. The latter approach is sometimes applied in statistical analyses with different issues (Hoerl and Kennard, 1970a, 1970b). The developed PTFs were used to predict $<K_s>$ throughout the catchment, including areas where infiltration measurements were not performed. Finally, maps of $<K_s>$ were derived investigating two alternative spatial interpolation approaches of the variables involved. Additional objectives of this study were understanding the possible benefits of the ridge regression compared with the widely used multiple linear regression and the role of the spatial interpolation on the resulting $<K_s>$ maps.

Materials

Study Area

The study area is the Hydrological Open Air Laboratory (HOAL) catchment, which is located in Petzenkirchen, in the western part of lower Austria. The basin has an elevation that ranges from 268 to 323 m asl, an area of 0.66 km^2, and a mean slope of 8%. The catchment mostly comprises arable lands (87%), and the remaining is covered by forested lands (6%), paved areas (2%), and pasture (5%). The climate can be characterized as humid with a mean annual temperature of 9.5°C and a mean annual precipitation of 823 mm yr^{-1} from 1990 to 2014 (Blöschl et al., 2016).

Winter wheat (*Triticum aestivum* L.), winter barley (*Hordeum vulgare* L.), maize (*Zea mays* L.), and rapeseed (*Brassica napus* L.) are the main cultivated crops in the catchment. Crop rotation is associated with green manure to ensure natural fertilization of the ground. However, N and natural fertilizers, e.g., pig manure, as well as plant protection agents such as plant growth regulators, fungicides, and broad-spectrum insecticides are also applied. The harvest of the winter crops usually occurs in July, while tillage and seedbed preparation are usually scheduled in late August or September.

A soil survey campaign of 300 cores, sampled on the nodes of a 50- by 50-m grid (light-gray squares in Fig. 1), mapped the soil textural composition of the catchment, providing information about organic matter (OM), clay (CL), silt (SI), and sand (SA) contents at multiple depths. According to the USDA soil classification, the topsoil of the catchment is constituted of 75% silty loam, 20% silty clay loam, and 5% silt. Furthermore, a high-resolution digital terrain model was used to derive elevation (EL) and local slope angles (S) across the catchment. Table 1 summarizes the information available in the grid nodes.

Table 1. Information available at the 300 nodes of the catchment grid.

Statistic	Organic matter	Clay	Silt	Sand	Slope	Elevation
	%				°	m asl
Min.	0.5	6.6	48.6	3.4	0.5	257.6
Max.	8.4	39.7	87.6	26.8	27.7	324.6
Mean	2.2	21.8	70.7	7.4	5.9	291.7
SD	0.7	6.4	6.8	2.5	4.0	14.4
CV, %	33.1	29.2	9.7	34.2	68.0	4.9

Training and Testing Datasets

An infiltration measurement campaign was performed between March and September 2017 in the HOAL catchment. A total of 131 point measurements of K_s in 12 different plots (dark-gray circles in Fig. 1) were collected using double-ring infiltrometers (DRIs). The DRI was chosen considering the ease of installation and the low cost of the equipment. Simultaneous measurements to reduce operation time were performed using four DRIs. Picciafuoco et al. (2019) proposed a methodology to identify the minimum number of measurements required for estimating $\langle K_s \rangle$ as geometric mean of K_s in a plot of given dimensions. Specifically, through an uncertainty analysis, they found that the width of the 95% confidence interval of $\langle K_s \rangle$ experienced little reduction with increasing the number of measurements beyond a specific threshold, making further measurements unnecessary. For an area of 80 m^2, this analysis suggested a minimum number of six measurements with an uncertainty interval equal to $\langle K_s \rangle$.

Following Picciafuoco et al. (2019), in the selected plots, subareas of approximately 80 m^2 were considered. In each subarea, six of the observed K_s values out of those available were used to derive the geometric mean at the field scale that is henceforth denoted by $\langle K_s^f \rangle$. Furthermore, the values of OM, CL, SI, SA, S, and EL available for the sampled nodes (Fig. 1) were weighted to estimate their average value in each subarea. These average values along with the $\langle K_s^f \rangle$ data were used as a database for developing and validating the PTF. Table 2 summarizes the basic statistics of this database. The average value of $\langle K_s^f \rangle$ across the subareas is 13.2 mmh^{-1}, and the associated average error with respect to $\langle K_s^f \rangle$ deduced using all the available measurements is 6.6 mm h^{-1}. This error is considered acceptable because it produces deviations within the observed $\langle K_s^f \rangle$ domain.

Because spatially averaged quantities are considered, the variation in SI, CL, OM, and SA is small, with CVs ranging from 1.7 to 10%, and only S and EL have moderate variability across the measurement subareas. Despite the small changes in soil characteristics, $\langle K_s^f \rangle$ exhibits a large variability, with a minimum of 2.7 mm h^{-1}, a maximum of 48.6 mm h^{-1}, and a CV of 71.8%.

Training and testing of the PTF required splitting the database, where two-thirds of the data constituted the training dataset used to develop the PTF, while the remaining data constituted the testing dataset. The split required the presence in the training dataset of at least one observation from each measurement plot and considered the data frequency distribution.

Methods

Development of the Pedotransfer Function for Areal-Average Saturated Hydraulic Conductivity

Due to the absence of information on parameters frequently involved in the definition of local PTFs for K_s such as bulk density

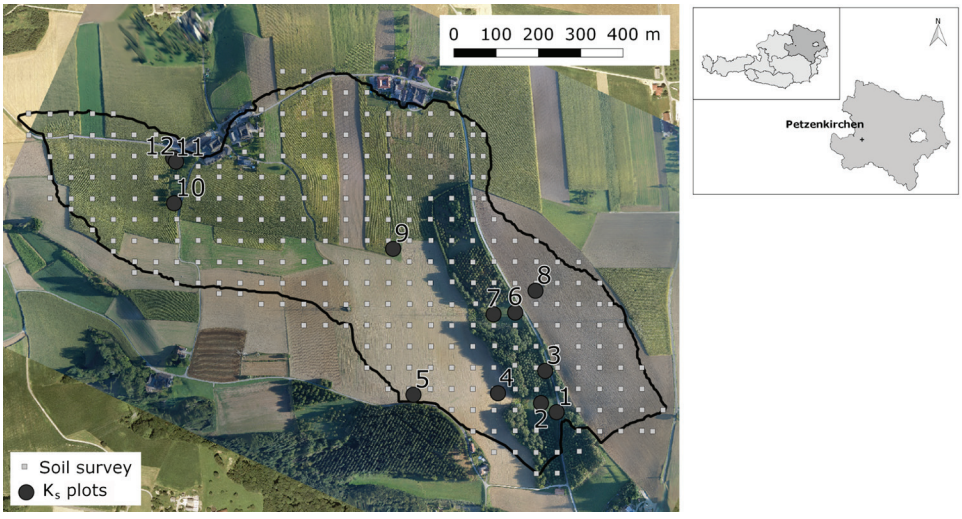

Fig. 1. Plots with available data in the catchment (dark-gray circles) and survey campaign locations of soil texture (light-gray squares). The catchment location in the Lower Austria region is also shown.

Table 2. Basic statistics of the complete database, including the experimental field-scale saturated hydraulic conductivity (<K_s^f>).

Statistic	Organic matter	Clay	Silt	Sand	Slope	Elevation	<K_s^f>
	%				°	m asl	mm h^{-1}
Min.	1.9	18.2	69.3	6.3	3.5	263.6	2.7
Max.	3.2	22.8	73.7	10.2	8.9	314.7	48.6
Mean	2.6	20.3	70.8	8.9	6.1	272.7	13.2
SD	0.3	0.9	1.2	1.0	0.8	14.2	9.5
CV, %	9.6	4.2	1.6	10.9	13.3	5.2	71.8

and total or effective porosity (Ahuja et al., 1984; Jabro, 1992; Spychalski et al., 2007), only the quantities OM, CL, SI, and SA were used. These quantities were incorporated in both linear and quadratic forms, in accordance with many published local PTFs (Brakensiek et al., 1984; Saxton et al., 1986; Wösten et al., 1999; Rahmati, 2017). This choice could introduce a certain degree of collinearity among the independent variables, a condition that was checked as described below.

To develop a PTF to estimate an unknown quantity (dependent variable) using a training dataset composed of both dependent and independent variables, regression techniques are often used.

Consider the standard model for multiple linear regression:

$$\mathbf{y}=\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon} \quad [1]$$

where the matrix of regressors $\mathbf{X}$ has dimensions ($n \times p$) and rank p, the unknown parameter vector $\boldsymbol{\beta}$ is ($p \times 1$), the vector of dependent variables $\mathbf{y}$ is ($n \times 1$), and the error vector $\boldsymbol{\varepsilon}$ is ($n \times 1$).

The objective is to find the $\boldsymbol{\beta}$ that minimizes the sum of squared errors:

$$\min_{\boldsymbol{\beta}} \left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2 \quad [2]$$

where the subscript 2 indicates the L2-norm of the vector. Assuming that $\boldsymbol{\varepsilon}$ is normally distributed, with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2$, and that the errors are uncorrelated, the ordinary least squares estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}=\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \quad [3]$$

The $(\mathbf{X}'\mathbf{X})^{-1}$ matrix always exists if the regressors are linearly independent, that is, if no column of the $\mathbf{X}$ matrix is a linear combination of the other columns. It is also known that the variance–covariance matrix of the regression coefficients is

$$\mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)=\sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \quad [4]$$

If we let $\mathbf{C} = (\mathbf{X}'\mathbf{X})^{-1}$, the variance of the $\hat{\beta}_j$ coefficient is $\sigma^2 C_{jj}$ and the covariance between $\hat{\beta}_i$ and $\hat{\beta}_j$ is $\sigma^2 C_{ij}$. Finally, the vector of fitted values $\hat{\mathbf{y}}$ corresponding to the observed values $\mathbf{y}$ is

$$\hat{\mathbf{y}}=\mathbf{X}\hat{\boldsymbol{\beta}} \quad [5]$$

Inferences such as those illustrated can be easily performed when the regressors are orthogonal, i.e., there is no linear relationship among them. However, when there are near-linear dependencies among the regressors, the inferences can be misleading or erroneous and the model is said to be affected by multicollinearity. In such circumstances, the $\mathbf{C}$ matrix tends to infinity, while the regression coefficients (Eq. [3]) and their variance–covariance matrix (Eq. [4]) become very large and may change erratically in response to small changes in the model or the data. According to Wold et al. (1984), the ridge regression is one of the methods that can be used to accomplish a stabilization of the regression estimates. To stabilize the parameter and to control the general instability associated with the least squares estimates, one can use

$$\hat{\boldsymbol{\beta}}^*=\left[\mathbf{X}'\mathbf{X}+k\mathbf{I}\right]^{-1}\mathbf{X}'\mathbf{y} \quad [6]$$

where $\mathbf{I}$ is the identity matrix ($p \times p$), $\hat{\boldsymbol{\beta}}^*$ is the ridge regression estimator of $\boldsymbol{\beta}$, and $k > 0$ is a positive quantity added to the diagonal of $\mathbf{X}'\mathbf{X}$. The ridge estimator can also be considered as the solution of the least squares problem with penalty $k\|\boldsymbol{\beta}\|_2{}^2$:

$$\min_{\boldsymbol{\beta}} \left\|\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right\|_2^2+k\left\|\boldsymbol{\beta}\right\|_2^2 \quad [7]$$

where the parameter k is the Lagrange multiplier used to resolve the minimization problem. It represents the bias introduced to reduce both the variance associated with the regression coefficients and the magnitude of the coefficients themselves, and as such it is named the shrinkage parameter. To account for possible different units among the regressors, the components of the $\mathbf{X}$ matrix are often standardized by subtracting from each $x_{i,j}$ the corresponding column means (μ_j) and dividing the results by the column standard deviations (SD_j). This is particularly helpful when the selection of k is performed by observing the ridge trace plot because, as explained below, all the coefficients assume the same magnitude.

According to Hoerl and Kennard (1970a, 1970b), the best method for achieving a better estimate of $\hat{\boldsymbol{\beta}}^*$ is using the ridge trace plot to select a single value of k and a unique $\hat{\boldsymbol{\beta}}^*$. The ridge trace is the plot of $\hat{\beta}_j{}^*$ for increasing k, and the selection criterion should consider that (i) coefficients should not have unreasonable absolute values (this is why standardizing $\mathbf{X}$ is important), (ii) coefficients with apparently incorrect signs at $k = 0$ should be changed to have the proper sign, and (iii) coefficients should not significantly change their values beyond the selected k. Furthermore, Golub et al. (1979) suggested the generalized cross-validation (GCV) method as an alternative approach to estimate k. In particular, the optimal k value is the one that minimizes the GCV statistics, which can be calculated as (Liu and Jiang, 2012):

$$\mathrm{GCV}\left(\hat{\boldsymbol{\beta}}^*\right)=\frac{\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}^*\right)'\left(\mathbf{y}-\mathbf{X}\hat{\boldsymbol{\beta}}^*\right)}{\left[n-\mathrm{tr}\left(\mathbf{H}\right)\right]^2} \quad [8]$$

where $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

To establish if a problem of multicollinearity among the independent variables exists, it is possible to derive the variance inflation factors (Marquardt, 1970) defined for the jth regressor as

$$\mathrm{VIF}_j = C_{jj} = \left(1 - R_j^2\right)^{-1} \quad [9]$$

where R_j^2 is the coefficient of determination obtained when x_j is regressed on the remaining $p - 1$ regressors; when x_j is nearly linearly dependent on some subset of the remaining regressors, R_j^2 is near unity and C_{jj} is large. Because the variance of the jth regression coefficient is $\sigma^2 C_{jj}$ (Eq. [4]), we can view VIF_j as the factor by which the variance of $\hat{\beta}_j$ is increased due to near-linear dependencies among the regressors. According to Montgomery et al. (2012), one or more large VIFs indicates multicollinearity, and practical experience suggests that values >10 are an indication that the associated regression coefficients are poorly estimated.

In this study, two PTFs were developed. The first PTF, henceforth designed as $\mathrm{PTF}_{\mathrm{MLR}}$, was derived by performing a multiple linear regression (MLR) analysis on a subset of independent variables whose selection is a step of fundamental importance in the regression procedure. In this context, Puckett et al. (1985), for example, performed a correlation analysis and computed correlation coefficients between pairs of variables; those properties with correlation coefficients $r > 0.7$ were selected for the regression. Ferrer Julià et al. (2004) performed a correlation analysis and chose as independent variables those that presented the best correlation values with the dependent variable. In this study, the independent variables used to derive the $\mathrm{PTF}_{\mathrm{MLR}}$ were those presenting at least a moderate correlation with the dependent variable $\langle K_s^f \rangle$ (Evans, 1996). Once the regressors were selected, after testing of the normal distribution of the errors and an absence of multicollinearity based on the Shapiro–Wilk test (Shapiro and Wilk, 1965) and the VIF, respectively, MLR was performed.

The second PTF, henceforth designated as $\mathrm{PTF}_{\mathrm{R}}$, was derived considering all the examined independent variables as regressors. As a consequence, multicollinearity in the matrix of regressors was detected by the VIF inspection. Multiple linear regression could not be applied and, therefore, the ridge regression technique was chosen to develop the PTF.

Evaluation Criteria

The developed PTFs were compared through an analysis of accuracy in the training and validation phases. The accuracy was assessed by matching the observed $\langle K_s^f \rangle$ and the corresponding predicted data for each dataset. It was quantified through the root mean square error (RMSE):

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left[\log_{10}\left\langle K_{\mathrm{s,m}i}^{\mathrm{f}}\right\rangle - \log_{10}\left\langle K_{\mathrm{s},ei}^{\mathrm{f}}\right\rangle\right]^2} \quad [10]$$

and the geometric mean error ratio (GMER):

$$\mathrm{GMER} = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\ln\left(\varepsilon_i\right)\right) \quad [11]$$

where the subscripts m and e represent measured and estimated data, respectively, n is the number of data, and ε_i is the error ratio, expressed as

$$\varepsilon_i = \frac{\left\langle K_{\mathrm{s},ei}^{\mathrm{f}}\right\rangle}{\left\langle K_{\mathrm{s,m}i}^{\mathrm{f}}\right\rangle} \quad [12]$$

An RMSE of 0 corresponds to a perfect match between observed and estimated values, while the optimal value of GMER is 1. A value of GMER <1 indicates underestimation of the predictive model.

Generation of a Continuous Map

After the development of the PTFs, maps of $\langle K_s^f \rangle$ were generated at a resolution of 25 by 25 m to minimize the effect of the interpolation step and highlight the performance of the PTFs considered. Two alternative methods were adopted (Fig. 2): Method A, with the independent variables available on the 50- by 50-m grid (Fig. 1) first interpolated onto the 25- by 25-m grid and with a PTF then applied at the 25- by 25-m grid; and Method B, with a PTF first applied to estimate $\langle K_s^f \rangle$ in the locations where the independent variables were known and these values then interpolated on the 25- by 25-m grid. The ordinary kriging interpolation approach (Matheron, 1963) was used in each method.

Finally, a qualitative comparison of the predicted $\langle K_s^f \rangle$ maps was performed, together with a comparison with the experimental $\langle K_s^f \rangle$values. These comparisons were performed with the purpose of having a criterion to establish which combination of PTF and Method (A or B) provided better results in terms of $\langle K_s^f \rangle$ spatial variability. A quantitative evaluation of the correspondence between experimental and predicted values was not possible because the spatial scales of the two quantities were different.

Results

Pedotransfer Function Development and Validation

In spite of the limited spatial variability of the examined soil properties, slight trends in the observed $\langle K_s^f \rangle$ with respect to them can be deduced. According to Fig. 3, $\langle K_s^f \rangle$ is positively correlated with SI and EL, while a negative correlation exists with OM, SA, and S. A minor link can be observed between $\langle K_s^f \rangle$ and CL. The presence of a link between the observed $\langle K_s^f \rangle$ values and the respective soil attributes suggests an in-depth analysis to derive an appropriate PTF.

The $\mathrm{PTF}_{\mathrm{MLR}}$ was developed through an analysis of the existing level of correlation between the dependent variable, $\langle K_s^f \rangle$, and the examined independent variables. Figure 4 shows the correlation among the different possible regressors: the color scale ranges from a very strong positive correlation (blue) toward zero correlation (white) down to a very strong negative correlation (red). According to Evans (1996), only OM^2 presents a strong correlation with $\langle K_s^f \rangle$, while moderate correlation can be detected in relation to SI, SI^2, SA, and EL and weak or very weak correlation exists between $\langle K_s^f \rangle$ and OM, CL, CL^2, SA^2, and S. On this basis, only the quantities at least moderately correlated ($r > 0.4$) with the

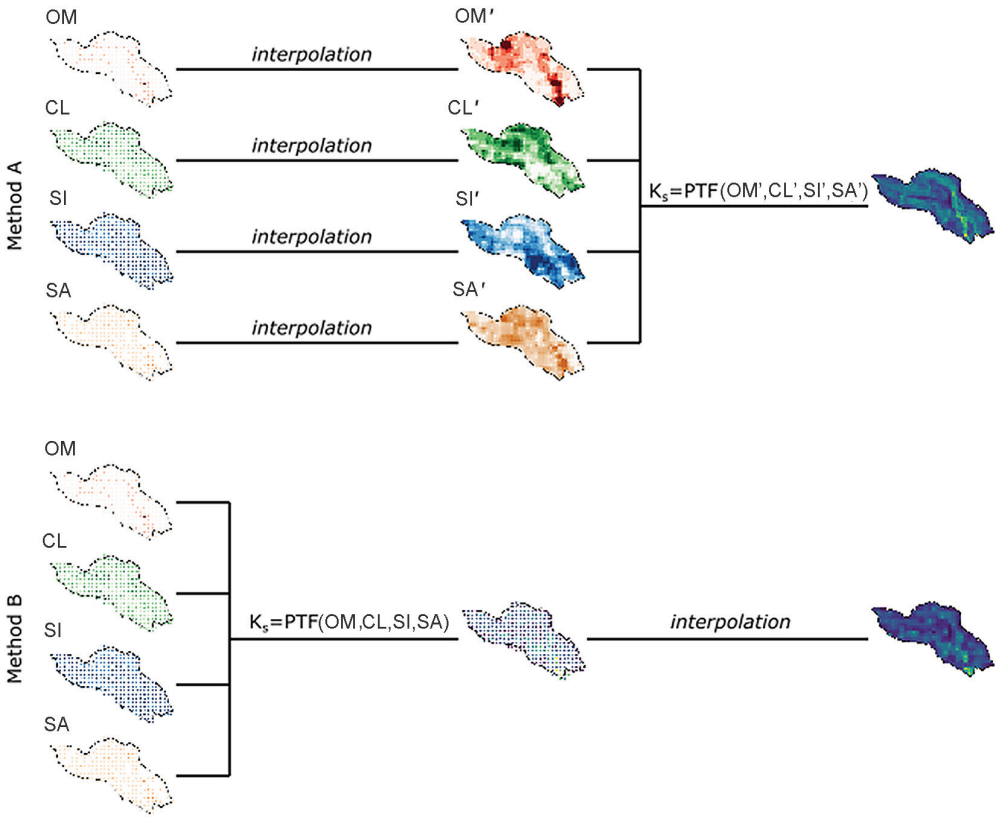

Fig. 2. Qualitative illustration of interpolation types to derive saturated hydraulic conductivity (K_s): Method A with a PTF applied to interpolated independent variables of organic matter (OM), clay content (CL), silt content (SI), and sand content (SA); and Method B with a PTF applied first to the available independent variables and then interpolated at a smaller scale.

dependent variable were chosen as regressors in the development of the PTF_{MLR}.

Estimates of the $\hat{\beta}_j$ regression coefficients (Appendix A) were made by applying Eq. [3] to the training dataset and successively verifying the assumptions of normally distributed errors and absence of multicollinearity in the matrix **X**. Figure 5 shows the magnitude of the VIFs, i.e., the indices that quantify how much the variances associated with the $\hat{\beta}_j$ estimated regression coefficients are affected by near-linear dependencies among the regressors. Because all the VIFs have values of the same order of magnitude and are equal to or less than the empirical limit of 10 (Montgomery et al., 2012), it can be concluded that multicollinearity did not affect the results.

The final step was application of the PTF_{MLR} to both the training and testing datasets and the evaluation of its accuracy. Figure 6a shows the results of the training procedure, with a remarkable accuracy of the PTF_{MLR} for $<K_s^f>$ lower than ~20 mm h^{-1} and a tendency to underestimate the experimental values for $<K_s^f>$ ranging between 20 and 40 mm h^{-1}. The overall performance is quantified by a RMSE of 0.21 and a GMER of 1.09, which support the suitability of the MLR because they approximate the optimal values. Next, Eq. [A1] was applied to the testing dataset. The outcomes of the testing procedure (Fig. 6b) are somewhat similar to those of the training dataset. The experimental $<K_s^f>$ values are adequately predicted up to approximately 20 mm h^{-1}, whereas a clear tendency to underestimate them is shown for $<K_s^f>$ between 20 and 40 mm h^{-1}. In any case, the indices RMSE and GMER for the testing dataset have values (0.24 and 0.98, respectively) comparable to those obtained during the training stage.

The PTF_R was developed considering all the available independent variables. In this case, the MLR technique could not be applied because the conditions of a normal distribution of the errors and an absence of multicollinearity were not satisfied. This was suggested by an analysis of the VIFs associated with the matrix **X** (Fig. 7) that exhibited five (out of 10) values greater than the empirical limit of 10. In addition, because the factors associated with CL, SI, and SA were two orders of magnitude larger than the others, the application of an MLR would have produced coefficients with large variances and thus poor prediction capability.

Therefore, ridge regression was used and both the methods suggested by Hoerl and Kennard (1970a, 1970b) and Golub et al. (1979) were applied to select the shrinkage parameter k. The ridge trace plot (Fig. 8a) was obtained by calculating (Eq. [6]) several $\hat{\boldsymbol{\beta}}^*$ values for multiple trial values of k in the range 0 to 10. Each line represents how a single coefficient $\hat{\beta}_j^*$ changes for different possible shrinkage parameters. Similarly, the GCV–k plot (Fig. 8b) was generated by Eq. [8] using the same trial shrinkage

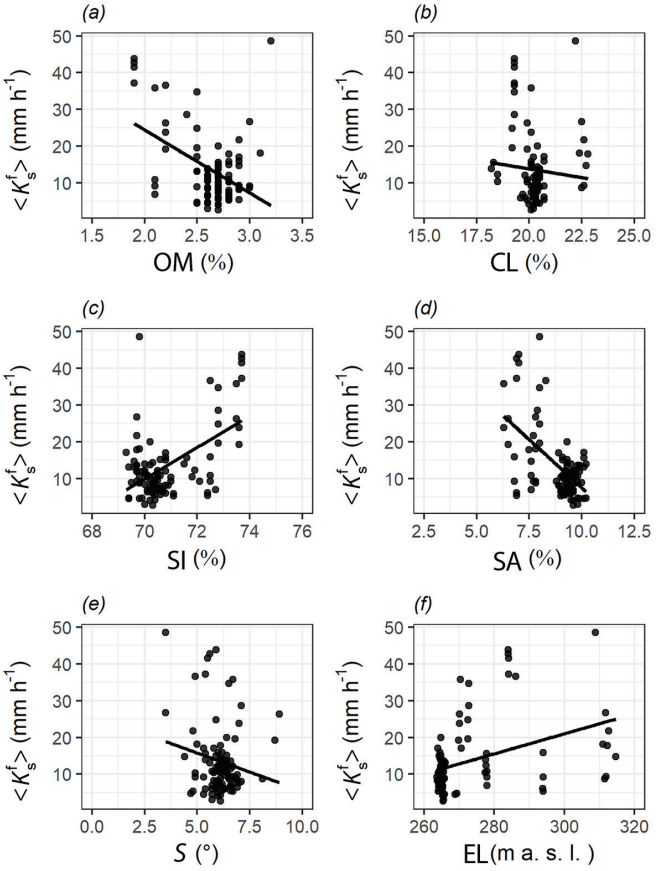

Fig. 3. Field-scale saturated hydraulic conductivity, <K_s^f>, plotted against (a) organic matter content (OM), (b) clay content (CL), (c) silt content (SI), (d) sand content (SA), (e) slope angle (S), and (f) elevation (EL).

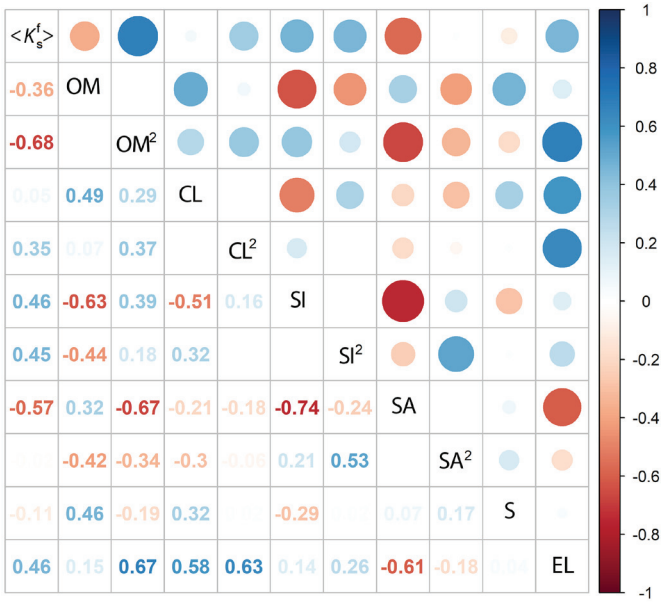

Fig. 4. Correlation matrix among the different available quantities: field-scale saturated hydraulic conductivity (<K_s^f>), organic matter content (OM), square of the organic matter content (OM^2), clay content (CL), square of the clay content (CL^2), silt content (SI), square of the silt content (SI^2), sand content (SA), square of the sand content (SA^2), slope angle (S), and elevation (EL). The size of the circles is proportional to the correlation coefficient r, while the color scale ranges from very strong positive correlation (blue) toward zero correlation (white) down to very strong negative correlation (red).

parameters and thus the same $\hat{\boldsymbol{\beta}}^*$. Finally, the selected k was the one that provided the smallest GCV and satisfied also the conditions for k selection through the ridge trace plot set out by Hoerl and Kennard (1970a, 1970b).

The developed PTF$_R$ (Eq. [A2] in Appendix A) was then applied to the training dataset to evaluate its performance at this stage. Figure 9a shows a fairly good accuracy for experimental <K_s^f> values lower than ~30 mm h^{-1} and a slight tendency to underestimate values >30 mm h^{-1}. The RMSE and GMER in the training step have values of 0.18 and 1.09, respectively. Similarly, Eq. [A2] was applied to the testing dataset (Fig. 9b). Many points are somewhat close to the 1:1 line and indicate a satisfactory relationship between experimental and predicted values. The indices RMSE and GMER of the validation phase have values of 0.18 and 1.01, respectively, which are comparable to those obtained during the training stage.

Comparison of Field-Saturated Hydraulic Conductivity Maps

The developed PTFs present similar performance in terms of RMSE and GMER in both the training and the validation stages, even though the PTF$_R$ produces errors which on average are smaller than those obtained with the PTF$_{MLR}$. However, for a complete assessment of the two applied regression methodologies, a comparison of the maps generated for the whole catchment was also performed.

Both Methods A and B involve experimental variograms fitted by an exponential model (Oliver and Webster, 2014) because this provided the smallest residual squares sum. Figure 10 shows the catchment <K_s^f> map obtained by applying the PTF$_{MLR}$ with Method A to the interpolated independent variables. The majority of the predicted values are on the blue portion of the color scale, i.e., around 20 mm h^{-1}, while a few values are >40 mm h^{-1} or <20 mm h^{-1}. Table 3 summarizes the basic statistics evaluated for the <K_s^f> predicted values: the minimum is 10.7 mm h^{-1} while the maximum is 45.6 mm h^{-1}, and the coefficient of variation is 12.7%, which is rather unusual for K_s even considering averaged values. Figure 11 shows the results obtaining by applying the PTF$_R$ with Method A to the spatially interpolated independent variables. Even though the majority of the predicted <K_s^f> values are in the central portion of the color scale, the map is characterized by a higher spatial variability than that of Fig. 10. This larger spatial variability is also reflected in a CV of 36.8% (Table 3) and a wider variation range of <K_s^f>, characterized by a minimum of 2.7 mm h^{-1} and a maximum of 43.9 mm h^{-1}.

Figure 12 shows the map generated by the PTF$_{MLR}$ with Method B. The minimum predicted value of <K_s^f> (Table 3) is 7.4 mm h^{-1}, which is smaller than that obtained by the same PTF with Method A. Similar evaluations can be made for the maximum

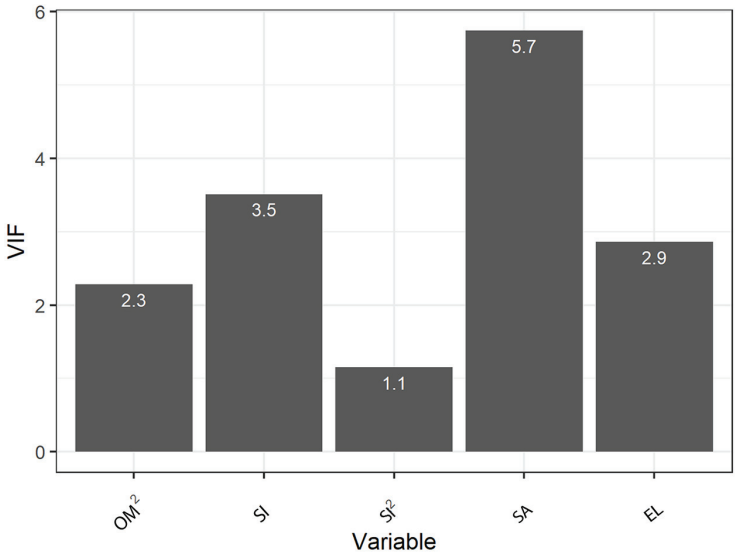

Fig. 5. Variance inflation factors (VIFs) associated with the regressors chosen to develop the pedotransfer function by a multiple linear regression. The independent variables are: square of organic matter content (OM^2), silt content (SI), square of silt content (SI^2), sand content (SA), and elevation (EL).

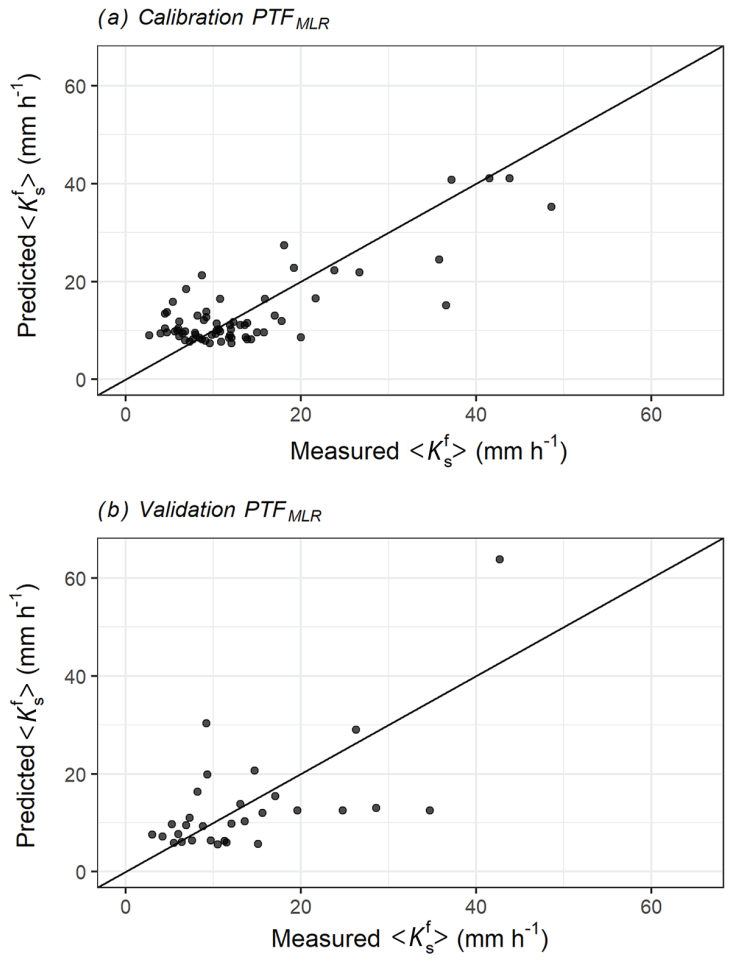

Fig. 6. Experimental vs. predicted field-scale saturated hydraulic conductivity, $<K_s^f>$, values obtained applying the pedotransfer function derived by multiple linear regression (PDF_{MLR}) to the (a) training and (b) testing datasets. The 1:1 line is also shown.

predicted value, which is 43.8 mm h^{-1} and smaller than that obtained by Method A. The map is also characterized by a higher spatial variability than the one shown in Fig. 10, with a less uniform color distribution and a greater CV of 20.5%. Finally, Fig. 13 is the resulting map achieved by the PTF_R with Method B. The spatial distribution of the predicted $<K_s^f>$values appears to be more random and therefore less influenced by the catchment morphology than the map obtained by Method A (Fig. 11). The variation range of $<K_s^f>$ is the widest among all the maps generated, with a minimum value of 1 mm h^{-1}, a maximum of 56.7 mm h^{-1}, and a CV of 42.4%.

Discussion

A moderate correlation was detected between the observed $<K_s^f>$ and soil characteristics. It was expected because of the close relationship between infiltration and soil composition. Texture alone has been reported to be a good predictor of K_s in sandy soils (Jaynes and Tyler, 1984; El-Kadi, 1985), and clay content was the main parameter correlated with K_s by Puckett et al. (1985). However, in the present study, CL exhibited a scarce correlation with $<K_s^f>$ and this made it unsuitable for the regression analysis. On the other hand, OM appeared to be a good predictor of $<K_s^f>$, as was also shown by Rawls et al. (1982, 1983), and Wösten et al. (1999), who successfully used this quantity as a PTF input. A strong relationship between OM and K_s was not always observed. For example, Ferrer Julià et al. (2004) detected almost no correlation between them in their study to construct a K_s map of Spain. Nevertheless, they decided to use OM as input in the developed PTF because of the well-known influence of this soil characteristic on water movement. The negative correlation between $<K_s^f>$ and SA is unusual. However, it could be due to the fact that the measurement locations with low values of SA coincide with agricultural catchment areas. In the study area, land management operations produced a significant increase in the experimental values of $<K_s^f>$ (Picciafuoco et al., 2019) and this accounts for the higher $<K_s^f>$ values associated with low SA.

Soil texture and topographical characteristics were used for training and validating the developed PTFs. Soil texture has often been used as a predictor of K_s, as well as other soil properties such as effective porosity and bulk density (Saxton et al., 1986; Wösten, 1997; Suleiman and Ritchie, 2001; Spychalski et al., 2007). However, these latter quantities were not available in this study. Furthermore, S was not generally used as a predictor despite its well-known influence on infiltration (Morbidelli et al., 2015).

For both the PTF_{MLR} and PTF_R, the training procedure (Fig. 6a and 9a) was found to be accurate to some extent for $<K_s^f>$ values lower than ~30 mm h^{-1}. Due to the lognormal shape of the K_s probability density function, the $<K_s^f>$ observations have mainly values <30 mm h^{-1} and

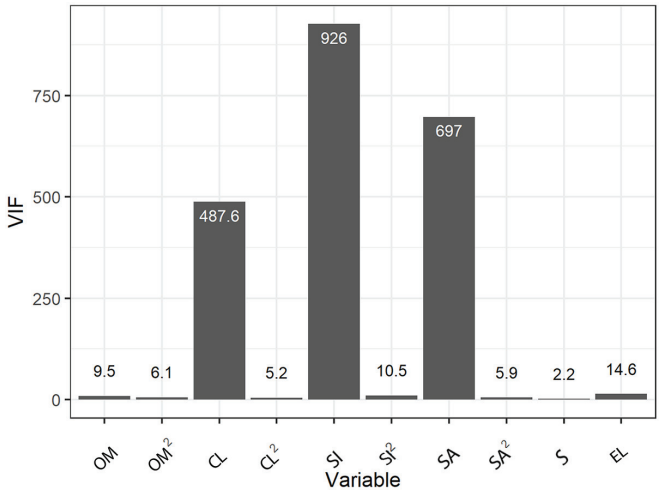

Fig. 7. Variance inflation factors (VIFs) associated with the regressors chosen to develop the pedotransfer function by ridge regression. The independent variables are: organic matter content (OM), square of organic matter content (OM^2), clay content (CL), square of the clay content (CL^2), silt content (SI), square of the silt content (SI^2), sand content (SA), square of the sand content (SA^2), slope angle (S), and elevation (EL).

the regression is more accurate in this range. In contrast, the regression for $<K_s^f>$ >30 mm h^{-1} appears to be less accurate, with observed values generally underestimated with the PTF_R. The performance is slightly better if the PTF_{MLR} is applied. Overall, the average errors calculated in the training stage by the two PTFs are comparable, with RMSEs of 0.21 and 0.18 for the PTF_{MLR} and PTF_R, respectively, which are better than those typically found in the literature for local predictions of K_s. For example, Schaap and Leij (1998) used ANN-based PTFs to predict K_s from different sets of soil properties including texture, bulk density, and retention points at 10 and 33 kPa. In the training stage, they obtained RMSE values ranging from 0.83, if only texture was used, to 0.66 if all the variables were utilized in the prediction. Parasuraman et al. (2006) used ANNs to model K_s for two distinct sites using bulk density and SA, SI, and CL. The RMSEs in the training stage ranged from 0.21 to 0.23 for the first site and from 0.36 to 0.38 for the second site, depending on the input variables utilized and the ensemble algorithm. Twarakavi et al. (2009) derived a set of four PTFs from the same database utilized to develop the software ROSETTA (Schaap et al., 2001) but applying a new methodology called support vector machines (SVM). They obtained K_s estimates for the training dataset with RMSEs between 0.55 and 0.71.

In the validation stage, the PTF_R provided better results than the PTF_{MLR} for $<K_s^f>$ values <10 mm h^{-1} (Fig. 6b and 9b), with the points aligned and close to the 1:1 line. The observed values were better predicted by the PTF_R also for $<K_s^f>$ between 20 and 40 mm h^{-1}, being the points in this interval closer to the bisector. However, both the PTFs poorly reproduced measured $<K_s^f>$ >40 mm h^{-1}, probably due to insufficient data included in the training stage. On average, the PTF_R provided better predictions

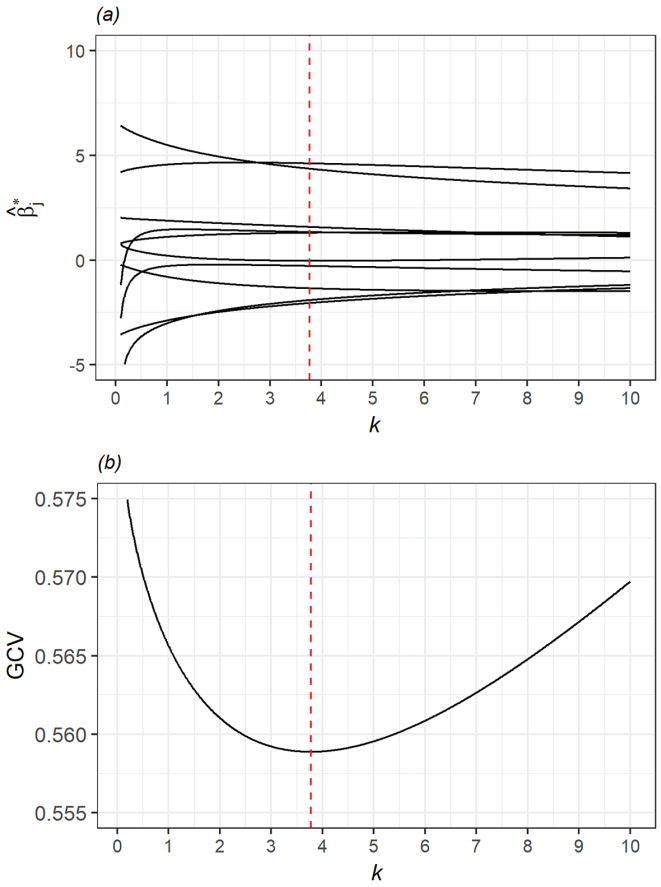

Fig. 8. Selection of shrinkage parameter k: (a) the ridge trace plot represents how each regression coefficient $\hat{\beta}_j^*$ changes for different possible shrinkage parameters; (b) generalized cross-validation (GCV) statistics as a function of k—the minimum GCV value associated with the best shrinkage parameter is indicated by the red dashed lines.

than the PTF_{MLR}, quantified by RMSEs of 0.18 and 0.24, respectively, which are better than those typically found in the literature for local predictions. Schaap and Leij (1998) obtained RMSEs relative to the validation stage ranging from 0.71 to 0.84 for different developed PTFs. Parasuraman et al. (2006) obtained RMSE values that varied from 0.20 to 0.23 for a first site and from 0.48 to 0.50 for a second site when an ANN-based PTF was used for the testing dataset. Twarakavi et al. (2009) achieved RMSE values for the validation dataset between 0.56 and 0.72 for the four SVM-based PTFs mentioned above.

In terms of map generation capability, the two PTFs developed in this study present substantial differences independently of the applied Method (A or B). From a merely qualitative analysis, the maps obtained with the PTF_{MLR} (Fig. 10 and 12) highlight the limited capacity of this function to reproduce the high spatial variability that usually characterizes $<K_s^f>$. The two maps are characterized by an almost homogeneous $<K_s^f>$ distribution. On the other hand, the maps obtained applying the PTF_R (Fig. 11 and 13) appear to better describe the $<K_s^f>$ spatial heterogeneity. These last two maps allow examination into how the two Methods (A and B) influence the final result. As can be easily deduced from Fig.

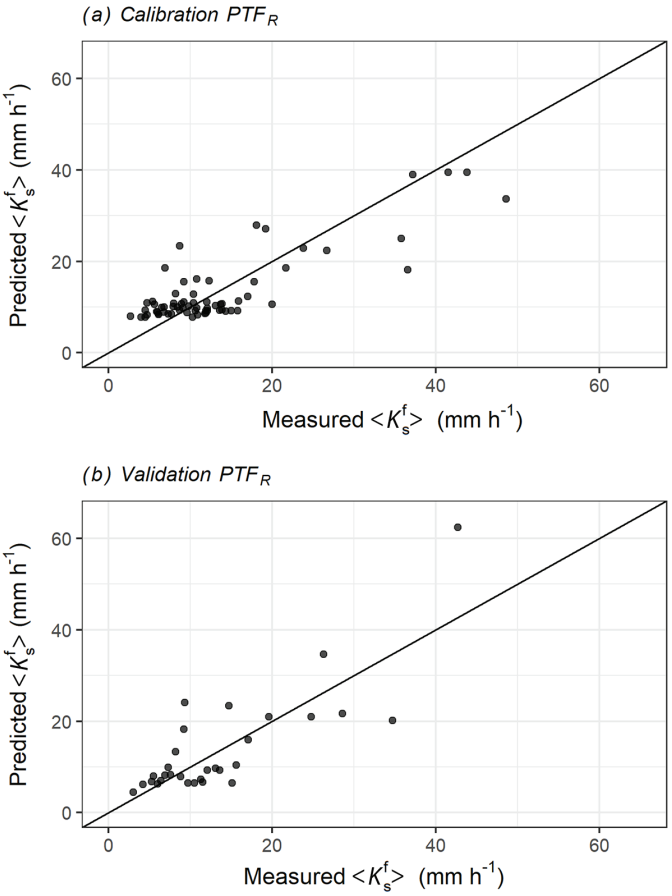

Fig. 9. Experimental vs. predicted field-scale saturated hydraulic conductivity, <K_s^f>, values obtained applying the pedotransfer function derived by the ridge regression (PTF$_R$) to the (a) training and (b) testing datasets. The 1:1 line is also shown.

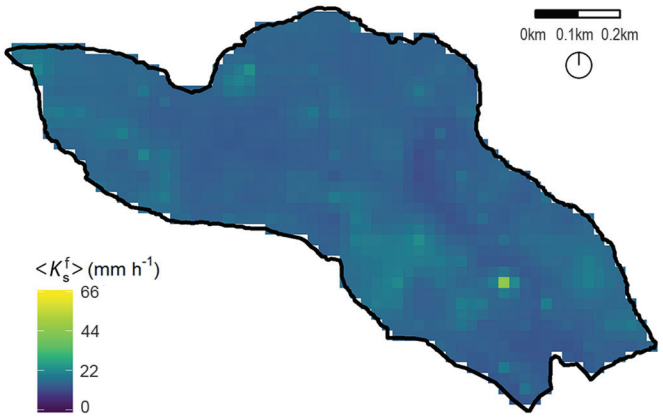

Fig. 10. Map of the field-scale saturated hydraulic conductivity, <K_s^f>, obtained by the pedotransfer function derived by multiple linear regression (PTF$_{MLR}$). Method A, involving spatial interpolation of the independent variables first and then application of the PTF$_{MLR}$, was used.

Table 3. Basic statistics of the field-scale saturated hydraulic conductivity (<K_s^f>) values of the maps obtained by the pedotransfer functions derived with both multiple linear regression technique (PTF$_{MLR}$) and ridge regression technique (PTF$_R$).

	Method A†		Method B‡	
Statistic	PTF$_{MLR}$	PTF$_R$	PTF$_{MLR}$	PTF$_R$
Min., mm h^{-1}	10.7	2.7	7.4	1.0
Max., mm h^{-1}	45.6	43.9	43.8	56.7
Mean, mm h^{-1}	15.0	14.5	14.6	13.9
SD, mm h^{-1}	1.9	5.7	3.0	5.9
CV, %	12.7	36.8	20.5	42.4

† Soil properties were first interpolated and then the PTFs applied.
‡ The PTFs were first applied at sites with available soil properties and then interpolated.

11 and 1, where the stream and forested areas are characterized by a higher variability while the cultivated areas have more uniform values, Method A is affected by the site morphology. This pattern is supported by the previous outcomes of Picciafuoco et al. (2019), who highlighted how, in the study area, the K_s variability, in terms of CV, reduces by about 25% in the agricultural areas compared with the forested land. On the other hand, Method B seems to be influenced by the soil survey sites where the independent variables were measured (light-gray squares in Fig. 1). In Fig. 13, many pixels stand out because their value is significantly different from that of the surrounding ones. They appear to be located on a grid, and their locations coincide with those where the soil surface was sampled to measure OM, CL, SI, and SA. These results are probably related to the smoothing effect that often affects ordinary kriging estimates (Isaaks and Srivastava, 1989), where usually small values are overestimated and large values are underestimated. This effect is uneven in space, being zero at the data locations and increasing with distance from these (see also Journel et al., 2000). The smoothing is less detectable if Method A is applied because it affects the estimates of the textural components, which are characterized by small variability across the catchment, and does not directly influence the predicted <K_s^f>. As a result, the final application of the PTF$_R$, in any case, generates a map linked with the catchment morphology. On the contrary, using the PTF$_R$ with Method B, the spatial interpolation step is the last phase of the map generation process. Therefore, the smoothing effect is more apparent because it directly influences the <K_s^f> estimates.

The qualitative analysis of the maps highlights the inability of the PTF$_{MLR}$ to represent the <K_s^f> spatial variability. However, a more quantitative analysis based on real observations is needed to fully understand which of the two examined methodologies provides the best results. For example, more <K_s^f> observations could be taken in areas of the catchment different from those already monitored and then the experimental dataset could be dived into three groups to be used to train the PTFs, to test the developed PTFs, and to evaluate the generated maps, respectively. Currently, in the absence of additional data, the predicted <K_s^f> values (Table 3) can be compared only with the available experimental <K_s^f> values (Table 2). However, a preliminary remark is necessary: the experimental <K_s^f> values refer to areas of about 80 m^2, while the predicted <K_s^f> values obtained from the application of the two

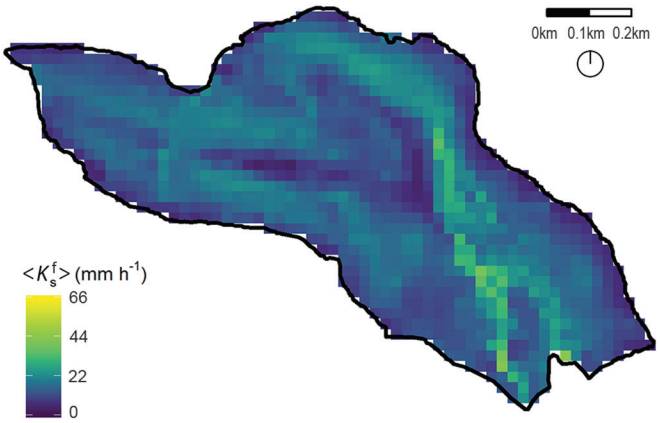

Fig. 11. Map of the field-scale saturated hydraulic conductivity, <K_s^f>, obtained by the pedotransfer function derived by ridge regression (PTF_R). Method A, involving spatial interpolation of the independent variables first and then application of the PTF_R, was used.

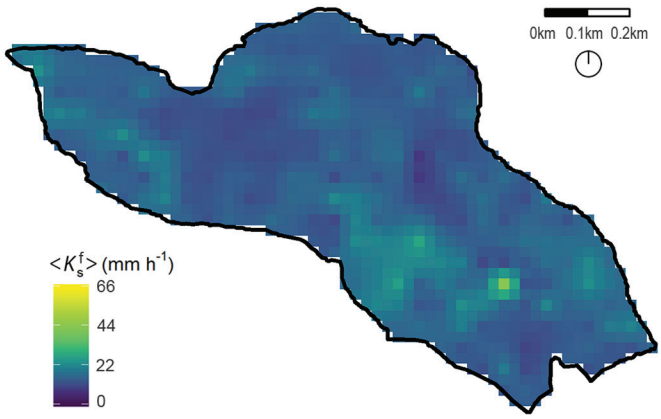

Fig. 12. Map of the field-scale saturated hydraulic conductivity, <K_s^f>, obtained by the pedotransfer function derived by multiple linear regression (PTF_{MLR}). Method B, involving application of the PTF_{MLR} to the independent variables first and then spatial interpolation, was used.

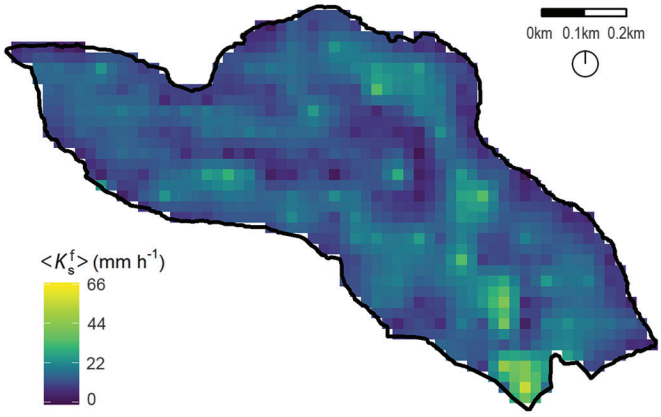

Fig. 13. Map of the field-scale saturated hydraulic conductivity, <K_s^f>, obtained by the pedotransfer function derived by ridge regression (PTF_R). Method B, involving application of the PTF_R to the independent variables first and then spatial interpolation, was used.

methods are representative of much larger areas (about 625 m^2). Therefore, the following comparison has to be considered as a discussion directed to understand the positive aspects of one methodology over the other. In terms of minimum value, the PTF_R provides 2.7 mm h^{-1} (Method A) and 1 mm h^{-1} (Method B), while the PTF_{MLR} gives 10.7 mm h^{-1} (Method A) and 7.4 mm h^{-1} (Method B). In both methods, the PTF_R minima are closer to the experimental one, which was 2.7 mm h^{-1}. However, in terms of maximum value, the PTF_{MLR} provides 45.6 mm h^{-1} (Method A) and 43.8 mm h^{-1} (Method B), which are more similar to the experimental value of 48.6 mm h^{-1} than the maxima obtained with PTF_R, in particular in the case of Method B. As far as the mean value is concerned, both the PTFs provided good estimates regardless of the applied method. The greatest differences between the two PTFs can be seen in the CVs. A reduction of the CV associated with the predicted <K_s^f> values compared with that obtained from the experimental data is expected because of the different scales that characterize these two quantities. Nonetheless, the CVs achieved utilizing the PTF_{MLR} (12.7% for Method A and 20.5% for Method B) are much lower than those obtained by the PTF_R (36.8% for Method A and 42.4% for Method B). Neither of the developed PTFs reproduces the variability of the complete database (CV = 71.8%), but better outcomes are provided by the PTF_R. Despite the higher CVs observed when Method B is applied, Method A provides better results because the derived map is influenced by the catchment morphology, which is what one would expect for a map representing the pattern of soil properties.

The striking element that delineates the PTF_R as the best option is the observation of the different frequency density functions (FDFs) associated with the set of observed <K_s^f> and the sets derived from the maps. In this context, because of the differences in sample size and representative scale between the observed and predicted FDFs, a rigorous quantitative analysis cannot be performed. In any case, Fig. 14 clearly shows that the PTF_{MLR} provides, with both methods, FDFs significantly different from that describing the observed <K_s^f> database.

Our outcomes point out that the PTF_R enables us to derive satisfactory estimates of <K_s^f>, not only in terms of mean but also in terms of variability across the catchment. On the basis of the soil characteristics of our catchment, it is possible to examine the prediction efficiency of PTFs proposed earlier. Among the available class PTFs (Wösten et al., 1995), we have selected the one implemented in the first hierarchical level of the ROSETTA software. Its application would give a catchment with only three values of <K_s^f>, as its area is characterized by a topsoil classified as silty loam (75%), silty clay loam (20%), and silt (5%). The RMSE associated with the application to the testing dataset would have acceptable magnitude, but the spatial structure of <K_s^f> would be completely lost (Fig. 15). The other five PTFs proposed above were also applied to the testing dataset, and their accuracy in terms of RMSE was compared with that characterizing the PTFs developed in this study. As can be seen in Fig. 15, all the PTFs, developed for different catchments and different

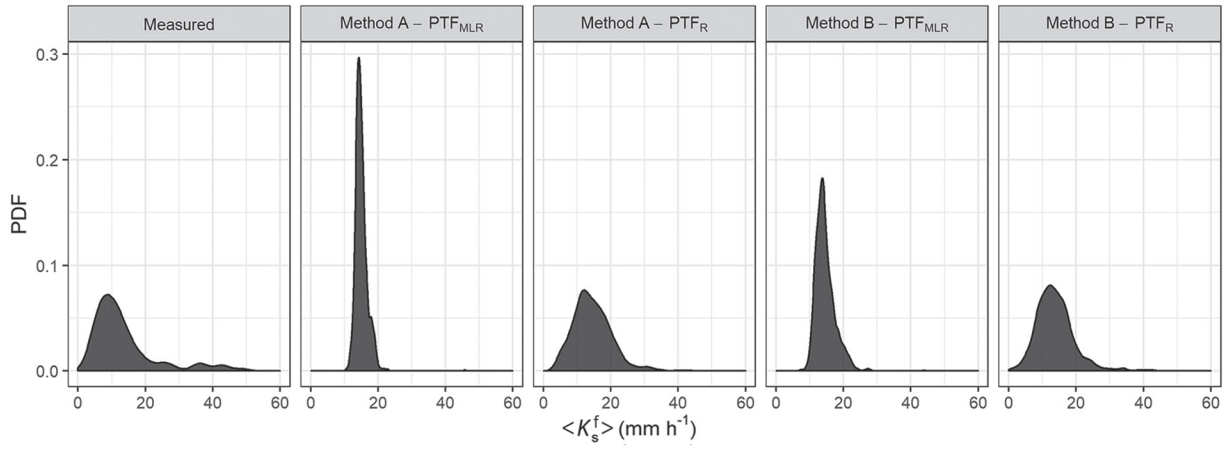

Fig. 14. Frequency density functions (FDFs) characterizing the set of measured field-scale saturated hydraulic conductivity, <K_s^f>, values and the sets derived from the maps. The maps were obtained by the pedotransfer functions derived by multiple linear regression (PTF$_{MLR}$) and ridge regression (PTF$_R$) and using Methods A (spatial interpolation of the independent variables first and then application of the PTF) and B (application of the PTF to the independent variables first and then spatial interpolation).

scales, produce RMSEs higher than those associated with PTF$_{MLR}$ and PTF$_R$. The RMSE values are comparable with those found in other studies (Schaap and Leij, 1998; Twarakavi et al., 2009); however, the predicted values are extremely inaccurate and do not allow for the <K_s^f> spatial variability. These predicted values are almost constant (see Fig. 15) because the input data for our catchment do not vary much.

In any case, these results could be expected because the existing PTFs were generally developed at the local scale, while in our study the PTFs were generated on the basis of a field-scale dataset. This approach allows the existence of random variability sources that are different from soil composition and cannot be quantified.

Conclusions

Two pedotransfer functions developed using the MLR technique and the ridge regression technique were compared in terms of accuracy and capability of reproducing <K_s^f> variability for map generation.

The accuracy at the training stage was found rather similar for the PTF$_{MLR}$ and PTF$_R$. A significant difference in the prediction accuracy of <K_s^f> was obtained in the subareas used for testing the PTFs, with lower RMSE for the PTF$_R$. Therefore, the PTF$_R$ can be considered a valid alternative to the classical MLR technique. This conclusion is also supported by a comparison of the maps generated for the whole area of interest, which were developed by applying the two predictive functions together with a spatial interpolation technique required to obtain an appropriate spatial resolution. Specifically, two alternatives were followed: Method A with the input soil parameters first interpolated, obtaining their continuous maps, and the PTFs then applied to estimate the <K_s^f>

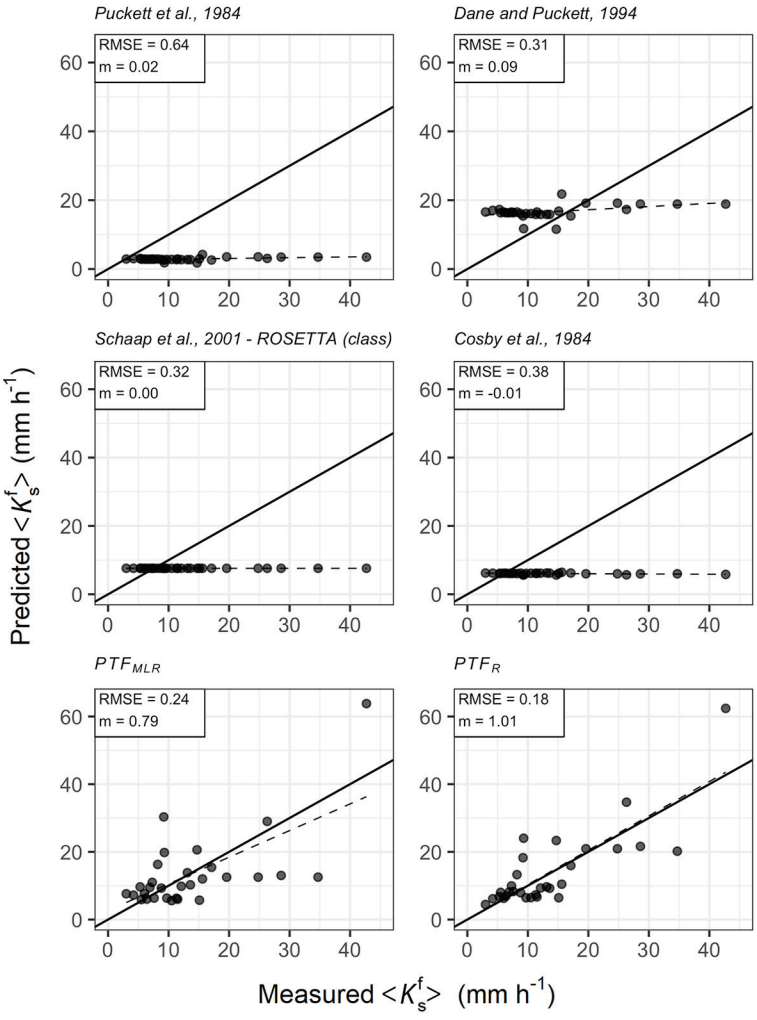

Fig. 15. Measured vs. predicted field-scale saturated hydraulic conductivity, <K_s^f>, values obtained applying literature pedotransfer functions (PTFs) (first and second rows) and the PTFs developed in this study and derived by multiple linear regression (PTF$_{MLR}$) and ridge regression (PTF$_R$) (third row) to the testing dataset; *m* is the slope of the fitting line (dashed line). The 1:1 line is also shown.

maps; and Method B with the PTFs first applied to the input parameters, generating partial maps of $\langle K_s^f \rangle$, and then interpolated to derive continuous maps of $\langle K_s^f \rangle$.

Independently of the two adopted alternatives, the maps obtained with the PTF_{MLR} show an almost uniform distribution of $\langle K_s^f \rangle$, which is not realistic in the study catchment. On the other hand, the maps generated by the PTF_R have a much more variable spatial pattern and a more appropriate CV. The statistical characteristics of the PTF_R–related maps are similar to those of the observed $\langle K_s^f \rangle$ database. In particular, the FDFs associated with the $\langle K_s^f \rangle$ values predicted by the PTF_R using Methods A or B have a shape that is very close to the one characterizing the FDF of the observed $\langle K_s^f \rangle$ database, while those associated with the PTF_{MLR} have a shape significantly different.

Furthermore, the use of the PTF_R with Method A can be considered to provide better results because the derived map is influenced by the catchment morphology, which is what one would expect for a map representing the pattern of a soil property. On the other hand, the map obtained by the PTF_R with Method B appears to be highly influenced by the location of the samples of the independent variables.

Finally, some PTFs proposed earlier were not able to predict $\langle K_s^f \rangle$ for our catchment because (i) the specific texture of the catchment is characterized by low soil variability across the investigated subareas, and (ii) they were generally developed at the local scale, while in our study the PTFs were generated on the basis of a field-scale dataset.

Appendix A

The equation of the PTF obtained using the MLR technique (PTF_{MLR}) is

$$\left\langle K_s^f \right\rangle = -2.24 + 46.2\mathrm{OM}^2 + 36\mathrm{SI} + 32\mathrm{SI}^2 + 23.7\mathrm{SA} + 0.1\mathrm{EL} \quad [A1]$$

The equation of the PTF obtained using the ridge regression technique (PTF_R) is

$$\left\langle \mathbf{K}_s^f \right\rangle = 13.5 + \tilde{\mathbf{X}}\hat{\boldsymbol{\beta}}^{\star} \quad [A2]$$

where $\tilde{\mathbf{X}}$ is the standardized matrix of regressors, $\langle \mathbf{K}_s^f \rangle$ is the vector of the predicted field-scale saturated hydraulic conductivity values, and $\hat{\boldsymbol{\beta}}^{\star}$ is the vector of the ridge coefficients:

$$\hat{\boldsymbol{\beta}}^{\star\mathrm{T}} = \begin{bmatrix} -1.35 & 4.62 & -1.91 & 1.30 & 1.34 & 4.36 \\ & -0.28 & -2.05 & 1.52 & -0.03 \end{bmatrix}$$

It is crucial to note that the matrix $\tilde{\mathbf{X}}$ needs to be generated considering the exact following order of the columns: OM, OM^2, CL, CL^2, SI, SI^2, SA, SA^2, S, and EL expressed in the units described above.

To estimate the $\langle \mathbf{K}_s^f \rangle$ vector with regressors additional to those used in the training stage, the standardized matrix $\tilde{\mathbf{X}}$ is generated according to the following transformation:

$$\tilde{x}_{ij} = \frac{x_{ij} - \mu_j}{\mathrm{SD}_j} \quad [A3]$$

where μ_j and SD_j are the jth elements of the vectors $\boldsymbol{\mu}$ and $\mathbf{SD}$, respectively, (given in mm h^{-1}) by

$$\boldsymbol{\mu} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 5.89 & 273 \end{bmatrix}$$

$$\mathbf{SD} = \begin{bmatrix} 0.12 & 0.12 & 0.12 & 0.12 & 0.12 & 0.12 \\ & 0.12 & 0.12 & 0.78 & 14.32 \end{bmatrix}$$

Acknowledgments

We would like to acknowledge financial support provided by the Austrian Science Funds (FWF) as part of the Vienna Doctoral Programme on Water Resource Systems (DK W1219-N22) and the Italian Ministry of Education, University and Research (PRIN 2015).

References

Abdelbaki, A.M. 2016. Using automatic calibration method for optimizing the performance of pedotransfer functions of saturated hydraulic conductivity. Ain Shams Eng. J. 7:653–662. doi:10.1016/j.asej.2015.05.012

Ahuja, L.R., J.W. Naney, and D.R. Nielsen. 1984. Scaling soil water properties and infiltration modeling. Soil Sci. Soc. Am. J. 48:970–973. doi:10.2136/sssaj1984.03615995004800050003x

Baiamonte, G., V. Bagarello, F. D'Asaro, and V. Palmeri. 2017. Factors influencing point measurement of near-surface saturated soil hydraulic conductivity in a small Sicilian basin. Land Degrad. Dev. 28:970–982. doi:10.1002/ldr.2674

Blöschl, G., A.P. Blaschke, M. Broer, C. Bucher, G. Carr, X. Chen, et al. 2016. The Hydrological Open Air Laboratory (HOAL) in Petzenkirchen: A hypothesis-driven observatory. Hydrol. Earth Syst. Sci. 20:227–255. doi:10.5194/hess-20-227-2016

Bouma, J. 1987. Transfer functions and threshold values: From soil characteristics to land qualities. In: K.J. Beek et al., editors, Quantified land evaluation procedures, Proceedings of the International Workshop, Washington, DC. ITC Publ. 6. Int. Inst. Geoinf. Sci. Earth Obs., Enschede, the Netherlands. p. 106–110.

Brakensiek, D.L., W.J. Rawls, and G. R. Stephenson. 1984. Modifying SCS hydrologic soil groups and curve numbers for rangeland soils. ASAE Pap. PNR-84203. Am. Soc. Agric. Eng., St. Joseph, MI.

Cornelis, W.M., J. Ronsyn, M. Van Meirvenne, and R. Hartmann. 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. Soil Sci. Soc. Am. J. 65:638–648. doi:10.2136/sssaj2001.653638x

Corradini, C., A. Flammini, R. Morbidelli, and R.S. Govindaraju. 2011. A conceptual model for infiltration in two-layered soils with a more permeable upper layer: From local to field scale. J. Hydrol. 410:62–72. doi:10.1016/j.jhydrol.2011.09.005

Corradini, C., F. Melone, and R.E. Smith. 1997. A unified model for infiltration and redistribution during complex rainfall patterns. J. Hydrol. 192:104–124. doi:10.1016/S0022-1694(96)03110-1

Corradini, C., F. Melone, and R.E. Smith. 2000. Modeling local infiltration for a two-layered soil under complex rainfall patterns. J. Hydrol. 237:58–73. doi:10.1016/S0022-1694(00)00298-5

El-Kadi, A.I. 1985. On estimating the hydraulic properties of soil: 2. A new empirical equation for estimating hydraulic conductivity for sands. Adv. Water Resour. 8:148–153. doi:10.1016/0309-1708(85)90055-7

Evans, J.D. 1996. Straightforward statistics for the behavioral sciences. Brooks/Cole Publ., Pacific Grove, CA.

Ferrer Julià, M., T.E. Monreal, A.S. del Corral Jiménez, and E. García Meléndez. 2004. Constructing a saturated hydraulic conductivity map of Spain using pedotransfer functions and spatial prediction. Geoderma 123:257–277. doi:10.1016/j.geoderma.2004.02.011

Golub, G.H., M. Heath, and G. Wahba. 1979. Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics 21:215–223. doi:10.1080/00401706.1979.10489751

Govindaraju, R.S., R. Morbidelli, and C. Corradini. 2001. Areal infiltration modeling over soils with spatially correlated hydraulic conductivities. J. Hydrol. Eng. 6:150–158. doi:10.1061/(ASCE)1084-0699(2001)6:2(150)

Hoerl, A.E., and R.W. Kennard. 1970a. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12:55–67. doi:10.1080/00401706.1970.10488634

Hoerl, A.E., and R.W. Kennard. 1970b. Ridge regression: Applications to nonorthogonal problems. Technometrics 12:69–82. doi:10.1080/00401706.1970.10488635

Isaaks, E.H., and R.M. Srivastava. 1989. An introduction to applied geostatistics. Oxford Univ. Press, New York.

Jabro, J.D. 1992. Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data. Trans. ASAE 35:557–560. doi:10.13031/2013.28633

Jaynes, D.B., and E.J. Tyler. 1984. Using soil physical properties to estimate hydraulic conductivity. Soil Sci. 138:298–305. doi:10.1097/00010694-198410000-00007

Journel, A.G., P.C. Kyriakidis, and S. Mao. 2000. Correcting the smoothing effect of estimators: A spectral postprocessor. Math. Geol. 32:787–813. doi:10.1023/A:1007544406740

Kern, J.S. 1995. Evaluation of soil water retention models based on basic soil physical properties. Soil Sci. Soc. Am. J. 59:1134–1141. doi:10.2136/sssaj1995.03615995005900040027x

Liu, X.-Q., and H.-Y. Jiang. 2012. Optimal generalized ridge estimator under the generalized cross-validation criterion in linear regression. Linear Algebra Appl. 436:1238–1245. doi:10.1016/j.laa.2011.08.032

Marquardt, D.W. 1970. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. Technometrics 12:591–612. doi:10.2307/1267205

Matheron, G. 1963. Principles of geostastistics. Econ. Geol. 58:1246–1266. doi:10.2113/gsecongeo.58.8.1246

Minasny, B., A.B. McBratney, and K.L. Bristow. 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93:225–253. doi:10.1016/S0016-7061(99)00061-0

Montgomery, D.C., E.A. Peck, and G.G. Vining. 2012. Introduction to linear regression analysis. 5th ed. John Wiley & Sons, Hoboken, NJ.

Morbidelli, R., C. Corradini, and R.S. Govindaraju. 2006. A field-scale infiltration model accounting for spatial heterogeneity of rainfall and soil saturated hydraulic conductivity. Hydrol. Processes 20:1465–1481. doi:10.1002/hyp.5943

Morbidelli, R., C. Saltalippi, A. Flammini, M. Cifrodelli, C. Corradini, and R.S. Govindaraju. 2015. Infiltration on sloping surfaces: Laboratory experimental evidence and implications for infiltration modeling. J. Hydrol. 523:79–85. doi:10.1016/j.jhydrol.2015.01.041

Morbidelli, R., C. Saltalippi, A. Flammini, E. Rossi, and C. Corradini. 2014. Soil water content vertical profiles under natural conditions: Matching of experiments and simulations by a conceptual model. Hydrol. Processes 28:4732–4742. doi:10.1002/hyp.9973

Nemes, A., M.G. Schaap, and J.H.M. Wösten. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67:1093–1102. doi:10.2136/sssaj2003.1093

Oliver, M.A., and R. Webster. 2014. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. Catena 113:56–69. doi:10.1016/j.catena.2013.09.006

Pachepsky, Y.A., D.J. Timlin, and W.J. Rawls. 2001. Soil water retention as related to topographic variables. Soil Sci. Soc. Am. J. 65:1787–1795. doi:10.2136/sssaj2001.1787

Papanicolaou, A.T.N., M. Elhakeem, C.G. Wilson, C.L. Burras, L.T. West, H.H. Lin, et al. 2015. Spatial variability of saturated hydraulic conductivity at the hillslope scale: Understanding the role of land management and erosional effect. Geoderma 243–244:58–68. doi:10.1016/j.geoderma.2014.12.010

Parasuraman, K., A. Elshorbagy, and B.C. Si. 2006. Estimating saturated hydraulic conductivity in spatially variable fields using neural network ensembles. Soil Sci. Soc. Am. J. 70:1851–1859. doi:10.2136/sssaj2006.0045

Parlange, J.-Y., I. Lisle, R.D. Braddock, and R.E. Smith. 1982. The three-parameter infiltration equation. Soil Sci. 133:337–341. doi:10.1097/00010694-198206000-00001

Philip, J.R. 1969. Theory of infiltration. Adv. Hydrosci. 5:215–296. doi:10.1016/B978-1-4831-9936-8.50010-6.

Picciafuoco, T., R. Morbidelli, A. Flammini, C. Saltalippi, C. Corradini, P. Strauss, and G. Blöschl. 2019. On the estimation of spatially representative plot scale saturated hydraulic conductivity. J. Hydrol. 570:106–117. doi:10.1016/j.jhydrol.2018.12.044

Puckett, W.E., J.H. Dane, and B.F. Hajek. 1985. Physical and mineralogical data to determine soil hydraulic properties. Soil Sci. Soc. Am. J. 49:831–836. doi:10.2136/sssaj1985.03615995004900040008x

Rahmati, M. 2017. Reliable and accurate point-based prediction of cumulative infiltration using soil readily available characteristics: A comparison between GMDH, ANN, and MLR. J. Hydrol. 551:81–91. doi:10.1016/j.jhydrol.2017.05.046

Rahmati, M., L. Weihermüller, J. Vanderborght, Y.A. Pachepsky, L. Mao, S.H. Sadeghi, et al. 2018. Development and analysis of the Soil Water Infiltration Global Database. Earth Syst. Sci. Data 10:1237–1263. doi:10.5194/essd-10-1237-2018

Rawls, W.J., D.L. Brakensiek, and K.E. Saxton. 1982. Estimation of soil water properties. Trans. ASAE 25:1316–1320. doi:10.13031/2013.33720

Rawls, W.J., D.L. Brakensiek, and B. Soni. 1983. Agricultural management effects on soil water processes: I. Soil water retention and Green and Ampt infiltration parameters. Trans. ASAE 26:1747–1752. doi:10.13031/2013.33837

Saxton, K.E., W.J. Rawls, J.S. Romberger, and R.I. Papendick. 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50:1031–1036. doi:10.2136/sssaj1986.03615995005000040039x

Schaap, M.G., and F.J. Leij. 1998. Database-related accuracy and uncertainty of pedotransfer functions. Soil Sci. 163:765–779. doi:10.1097/00010694-199810000-00001

Schaap, M.G., F.J. Leij, and M.Th. van Genuchten. 2001. ROSETTA: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. J. Hydrol. 251:163–176. doi:10.1016/S0022-1694(01)00466-8

Sedaghat, A., H. Bayat, and A.A. Safari Sinegani. 2016. Estimation of soil saturated hydraulic conductivity by artificial neural networks ensemble in smectitic soils. Eurasian Soil Sci. 49:347–357. doi:10.1134/S106422931603008X

Shapiro, S.S., and M.B. Wilk. 1965. An analysis of variance test for normality (complete samples). Biometrika 52:591–611. doi:10.2307/2333709

Smith, R.E., and D.C. Goodrich. 2000. Model for rainfall excess patterns on randomly heterogeneous areas. J. Hydrol. Eng. 5:355–362. doi:10.1061/(ASCE)1084-0699(2000)5:4(355)

Spychalski, M., C. Kaźmierowski, and Z. Kaczmarek. 2007. Estimation of saturated hydraulic conductivity on the basis of drainage porosity. Electron. J. Pol. Agric. Univ. 10(1):04. http://www.ejpau.media.pl/volume10/issue1/art-04.html

Suleiman, A.A., and J.T. Ritchie. 2001. Estimating saturated hydraulic conductivity from soil porosity. Trans. ASAE 44:235–339. doi:10.13031/2013.4683

Te Chow, V., D.R. Maidment, and L.W. Mays. 1988. Applied hydrology. McGraw-Hill, New York.

Tietje, O., and V. Hennings. 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69:71–84. doi:10.1016/0016-7061(95)00050-X

Tietje, O., and M. Tapkenhinrichs. 1993. Evaluation of pedotransfer functions. Soil Sci. Soc. Am. J. 57:1088–1095. doi:10.2136/sssaj1993.03615995005700040035x

Tóth, B., M. Weynants, A. Nemes, A. Makó, G. Bilas, and G. Tóth. 2014. New generation of hydraulic pedotransfer functions for Europe. Eur. J. Soil Sci. 66:226–238. doi:10.1111/ejss.12192

Twarakavi, N.K.C., J. Šimůnek, and M.G. Schaap. 2009. Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. Soil Sci. Soc. Am. J. 73:1443–1452. doi:10.2136/sssaj2008.0021

Van Looy, K., J. Bouma, M. Herbst, J. Koestel, B. Minasny, U. Mishra, et al. 2017. Pedotransfer functions in earth system science: Challenges and perspectives. Rev. Geophys. 55:1199–1256. doi:10.1002/2017RG000581

Vereecken, H. 1989. Pedotransfer functions for the generation of hydraulic properties for Belgian soils. Ph.D. diss. Faculty of Agric. Sci., K.U. Leuven, Leuven, Belgium.

Wagner, B., V.R. Tarnawski, V. Hennings, U. Müller, G. Wessolek, and R. Plagge. 2001. Evaluation of pedo-transfer functions for unsaturated soil hydraulic conductivity using an independent data set. Geoderma 102:275–297. doi:10.1016/S0016-7061(01)00037-4

Wold, S., A. Ruhe, H. Wold, and W.J. Dunn III. 1984. The collinearity problem in linear regression: The partial least squares (PLS) approach to generalized inverses. SIAM J. Sci. Statist. Comput. 5:735–743. doi:10.1137/0905052

Wösten, J.H.M. 1997. Pedotransfer functions to evaluate soil quality. Dev. Soil Sci. 25:221–245. doi:10.1016/S0166-2481(97)80037-2

Wösten, J.H.M., P.A. Finke, and M.J.W. Jansen. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66:227–237. doi:10.1016/0016-7061(94)00079-P

Wösten, J.H.M., A. Lilly, A. Nemes, and C. Le Bas. 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90:169–185. doi:10.1016/S0016-7061(98)00132-3

Wösten, J.H.M., Y.A. Pachepsky, and W.J. Rawls. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic characteristics. J. Hydrol. 251:123–150. doi:10.1016/S0022-1694(01)00464-4