The added utility of nonlinear methods compared to linear methods in rescaling soil moisture products

DOI:
https://doi.org/10.1016/j.rse.2017.05.017

Document Version
Accepted author manuscript

Citation for published version (APA):
Afshar, M., & Yilmaz, M. T. (2017). The added utility of nonlinear methods compared to linear methods in rescaling soil moisture products. *Remote Sensing of Environment*, *196*, 224. https://doi.org/10.1016/j.rse.2017.05.017

Published in:
Remote Sensing of Environment

THE ADDED UTILITY OF NONLINEAR METHODS COMPARED TO LINEAR METHODS IN RESCALING SOIL MOISTURE PRODUCTS

M. H. Afshar and M. T. Yilmaz*

Civil Engineering Department, Middle East Technical University, Ankara, Turkey

* Correspondence to: M. T. Yilmaz, tuyilmaz@metu.edu.tr

April 22, 2017

Abstract

In this study, the added utility of nonlinear rescaling methods relative to linear methods in the framework of creating a homogenous soil moisture time series has been explored. The performances of 31 linear and nonlinear rescaling methods are evaluated by rescaling the Land Parameter Retrieval Model (LPRM) soil moisture datasets to station-based watershed average datasets obtained over four United States Department of Agriculture (USDA) Agricultural Research Service (ARS) watersheds. The linear methods include first-order linear regression, multiple linear regression, and multivariate adaptive regression splines (MARS), whereas the nonlinear methods include cumulative distribution function matching (CDF), artificial neural networks (ANN), support vector machines (SVM), Genetic Programming (GEN), and copula methods. MARS, GEN, SVM, ANN, and the copula methods are also implemented to utilize lagged observations to rescale the datasets. The results of a total of 31 different methods show that the nonlinear methods improve the correlation and error statistics of the rescaled product compared

to the linear methods. In general, the method that yielded the best results using training data improved the validation correlations, on average, by 0.063, whereas ELMAN ANN and GEN, using lagged observations methods, yielded correlation improvements of 0.052 and 0.048, respectively. The lagged observations improved the correlations when they were incorporated into rescaling equations in linear and nonlinear fashions, with the nonlinear methods (particularly SVM and GEN but not ANN and copula) benefitting from these lagged observations more than the linear methods. The overall results show that a large majority of the similarities between the LPRM and watershed average datasets are due to linear relations; however, nonlinear relations clearly exist, and the use of nonlinear rescaling methods clearly improves the accuracy of the rescaled product.

Key Words: Soil moisture, rescaling, linear, nonlinear, remote sensing

1. Introduction

Soil moisture is one of the key variables in many geophysical science applications (e.g., those dealing with climate, hydrology, water resources, or agriculture; Lawrence & Hornberger, 2007) owing to its memory (Han et al., 2014) and role in water and energy exchange between land and the atmosphere (Koster et al., 2004). Hence, an accurate estimation of soil moisture is critical for many applications (Dorigo et al., 2012). Different soil moisture time series for the same location and same time period can be retrieved via different platforms (e.g., hydrological models, in situ observations, and remote sensing). It is often desirable to merge these different datasets to obtain more accurate estimates (Anderson et al., 2012; Yilmaz et al., 2012). However, due to the limitations of these platforms (e.g., satellites can monitor only the top few centimeters at relatively coarse resolutions, points in in situ observations have spatial representativeness limitations, and models have different parameterizations (Koster et al., 2009)), these datasets have systematic differences in their horizontal, temporal, and/or vertical supports (Dirmeyer et al., 2004; Koster et al., 2009). As a result, soil moisture values obtained from various platforms often need to be rescaled before they can be meaningfully validated, merged, or used in different applications (Dirmeyer et al., 2004; Reichle & Koster, 2005; Reichle et al., 2008; Yilmaz and Crow, 2013; Yin et al., 2014; Su and Ryu, 2015).

Many different methods are proposed to handle these systematic differences between soil moisture products, where an unscaled original product Y is rescaled to the space of a reference product X. However, the performances of these methods depend on many factors, including sampling errors, the degree to which the rescaling methods' underlying assumptions are met, and the goal of the rescaling efforts. Examples of such goals include minimizing the variability of the difference between the rescaled product (Y^*) and X via a first-order linear regression (REG1), matching the total variability of a dataset Y to an arbitrary reference dataset X (VAR), matching the cumulative distribution function (cdf), and matching only the signal variability of Y to that of X (here, "signal" refers to the true variability of a dataset, where the total variability is composed of true signal variability and noise variability components) using triple collocation analysis (TCA: Hain et al., 2011; Miralles et al., 2011; Parinussa et al., 2011; Scipal et al., 2008; Stoffelen, 1998; Zwieback et al., 2012).

Once the rescaling method is selected for implementation in a specific application, this method can be implemented using different strategies (Yilmaz et al., 2016). For example, a dataset can be rescaled by using a single coefficient for the entire time series by using separate rescaling coefficients for each month or separate coefficients for the anomaly and seasonality components. Such rescaling strategies affect the accuracy statistics of Y^*, even though, by definition, a particular rescaling method is selected to be the optimum method for a particular application (here, the optimum method refers to the method that results in the best statistic of interest, among other methods). To give a more specific example, consider the relative accuracies of X and Y or the differences between the signal-variability-to-noise-variability ratio (Gruber et al., 2016), for X (SNR_X) and Y (SNR_Y). In general, the relative variations of SNR_X and SNR_Y are expected to impact the overall performance of the rescaling methods through the use of various rescaling strategies (Yilmaz et al., 2016) for many applications (e.g., the creation of homogenous time series and data assimilation). For example, if $SNR_X >> SNR_Y$, it is better to rescale Y strongly to X (e.g., by rescaling the seasonality and anomaly components separately using two different rescaling coefficients or rescaling datasets for each month separately using 12 different rescaling coefficients). By contrast, if $SNR_Y > SNR_X$, it is better to weakly rescale Y to X (e.g., by rescaling the entire time series at once and using a single rescaling coefficient). Hence, the performance of any rescaling method (e.g., REG1, VAR, TCA, and CDF) could vary depending on the aggressiveness with which the rescaling strategy is implemented (e.g., weak or strong; Yilmaz et al., 2016).

Both the rescaling method selection (Yilmaz & Crow, 2013) and degree of aggressiveness implemented (Yilmaz et al., 2016) can impact the optimality of the Y^* statistics. Here, the question arises whether the inter-comparisons of rescaling methods make sense, without taking into consideration SNR variations. Yilmaz et al. (2016) investigated the impact of SNR variations using only a particular rescaling method (VAR). Hence, before making comments with high confidence, a sensitivity study that comprehensively investigates the impact of SNR variations on the performances of various rescaling methods is still required. However, in the absence of evidence, it is viable that SNR variations will impact various rescaling methods similarly, though the actual degree of improvement via stronger/weaker rescaling strategies may depend on the particular rescaling method. Accordingly, a universally optimum rescaling method that fits all applications may not exist; the optimality of a rescaling method is largely application specific, particularly if the underlying assumptions inherent to its own methodology are not met. Hence, studies investigating the relative performances of different rescaling methods (both linear and nonlinear) may still contribute to the efforts on the topic of optimal rescaling methods, even without explicitly considering SNR variations.

Satellite-based soil moisture data are often validated using station-based watershed average data (Jackson et al., 2010, 2012), which have considerably higher local nonlinearity, due to the soil moisture dynamics (Crow & Wood, 2002). The spatial support difference between station- and remote sensing-based products (i.e., point vs areal average) is another source that introduces nonlinear relations between different products. In a recent study, Zwieback et al. (2016) introduced nonparametric CDF and used two new parametric methods to extend TCA to investigate the impact of nonlinear relations on the error statistics obtained via TCA. This study particularly stresses the existing quadratic relations (e.g., the saturation of sensitivity of a product with respect to the sensitivity of another product) between the actual signal components of different soil moisture products, which may lead to nonlinear relations. Zwieback et al. (2016) also provided an extensive discussion on the existence of nonlinear relations between soil moisture products. It is, therefore, viable that such existing nonlinear relations between datasets may not be captured using linear methods, and the use of nonlinear methods may be necessary. By contrast, the variety of nonlinear methods used to rescale soil moisture datasets remains very limited, and there is still more room to investigate the performance of such nonlinear methods.

Among the rescaling methods used in soil moisture studies, CDF (Drusch et al., 2005; Reichle & Koster, 2004; Yin et al., 2015; Zwieback et al., 2016) has received particular attention.

Other methods, based on VAR (Crow et al., 2005; Draper et al., 2009; Su et al., 2013), REG1 (Brocca et al., 2013; Crow & Zhan, 2007; Crow, 2007;), TCA (Yilmaz & Crow, 2013), quadratic polynomials (Zwieback et al., 2016), copula (Leroux et al., 2014), and Wavelets (Su & Ryu, 2015) have also been implemented to reduce the systematic differences between soil moisture time series. However, a comprehensive intercomparison of the performances of these methods in a soil moisture rescaling study has not yet been performed.

The above-listed methodologies have been explicitly used in soil moisture rescaling studies, whereas many other methods have not. For example, multiple linear regressions using quadratic equations (REG2) and lagged observations (REGL) have previously been used in a soil moisture TCA framework (*Crow et al.,* 2015; Su et al., 2014; Zwieback et al., 2016), but quadratic equations and lagged observations together (REGL2) have not. Among the many machine learning methodologies, ANN methods (Rochester et al., 1956) have been used to retrieve soil moisture via microwave measurements (Notarnicola et al., 2008; Paloscia et al., 2008; Prigent et al., 2005; Rodriguez et al., 2015) and SVM methods (Cortes & Vapnik, 1995) have been used to predict soil moisture (Gill et al., 2006) in the root zone using data assimilation techniques (Liu et al., 2010). Other methods that can be used to relate the different datasets, such as the nonlinear regression methods GEN (Koza, 1994) and MARS (Friedman, 1991), have not been used in soil moisture-related studies. To our knowledge, none of these methods (REG2, REGL, REGL2, MARS, GEN, SVM, and ANN) have previously been explicitly used to rescale soil moisture datasets.

The soil moisture has a high temporal memory (i.e., autocorrelation), and consecutively retrieved soil moisture observations have high dependence, implying that previously retrieved soil moisture observations could arguably be viewed as a slightly degraded version of the current values. This property is very valuable for satellite-based soil moisture retrievals; lagged soil moisture products could be used as independent observations, given that past observations are quasi-independently obtained from current observations. This dependence has been utilized by many recent studies (Crow et al., 2015; Su et al., 2014; Zwieback et al., 2013), particularly those focusing on soil moisture TCA methods, which require three independent products. Exploiting the same information source, lagged variables are inherently used by some ANN types in building robust relations between the input and output layers. Although many other methods (e.g., multiple linear regression, MARS, GEN, copula, and SVM) could also benefit from such information in the framework of rescaling soil moisture variables, such an effort has not been made to date.

VAR, REG1, TCA, and CDF have unique solutions and are widely implemented in soil moisture rescaling studies. The optimality of linear rescaling methods (VAR, REG1, and TCA) in the context of data assimilation has been investigated both analytically and numerically by Yilmaz and Crow (2014), and some remedies are available for these methods when the underlying assumptions are not met (Crow & Yilmaz, 2014; Su et al., 2014). However, because the implementations of nonlinear rescaling methods remain limited in the context of rescaling soil moisture time series, the performance of these nonlinear methods, which are relative to that of linear methods, remains largely unexplored. Therefore, there is still room to investigate the performances of nonlinear methods relative to those of linear methods to better understand the degree of existing nonlinearity in soil moisture products, even though the degree of existing nonlinearity and degree to which these nonlinear relations can be captured drives the actual difference between the performance of the nonlinear and linear rescaling methodologies.

This study is the first to use a number of methods (REG2, REGL, REGL2, ANN, SVM, GEN, and MARS) and their lagged types to explicitly rescale the soil moisture observations. This study also includes the first comprehensive comparison of the performances of linear methods

(REG1, REG2, REGL, REGL2, VAR, TCA, and MARS) as well as nonlinear methods (CDF, copula, ANN, SVM, and GEN) in rescaling soil moisture datasets. Through these intercomparisons, this study comprehensively analyzes the added utility of lagged observations in a soil moisture rescaling framework. This study is particularly relevant for the efforts to create a homogenous time series in the framework of global soil moisture dataset validation (Leroux et al., 2014) and trend analysis (Dorigo et al., 2012), contributes to the efforts to better understand the optimality of different rescaling methodologies (Yilmaz and Crow, 2013; Yilmaz et al., 2016), and adds to the efforts to identify the degree of the existing nonlinearity in soil moisture products.

2. Linear and Nonlinear Rescaling Methods

2.1. Linear Regression

2.1.1 First-order Linear Regression

Linear rescaling methods have been widely used to rescale soil moisture time series to reduce their inconsistency (Brocca et al., 2013; Crow et al., 2005; Crow & Zhan, 2007). Overall, linear rescaling methods are implemented by considering the most general linear relation between a reference dataset (X) and an original unscaled dataset (Y) in the form of:

$$\mathrm{Y}^{*} = \mu_{\mathrm{X}} + (\mathrm{Y} - \mu_{\mathrm{Y}})c_{\mathrm{Y}}, \quad (1)$$

where Y^{*} is the rescaled version of Y; μ_{X} and μ_{Y} are time averages of X and Y, respectively; and c_{Y} is a scalar rescaling factor (in this study, minimum-maximum fits are not considered). Here, c_{Y} is found using REG1, VAR, and TCA-based linear methods (*Yilmaz and Crow*, 2013):

$$c_{\mathrm{Y}}^{\mathrm{R}} = \rho_{\mathrm{XY}}\, \sigma_{\mathrm{X}}/\sigma_{\mathrm{Y}} \quad (2)$$

$$c_{\mathrm{Y}}^{\mathrm{V}} = \sigma_{\mathrm{X}}/\sigma_{\mathrm{Y}} \quad (3)$$

$$c_{\mathrm{Y}}^{\mathrm{T}} = \Sigma_{\mathrm{xz}}/\Sigma_{\mathrm{yz}}. \quad (4)$$

where Z is a third product that is similar to products X and Y; Σ_{xz} and Σ_{yz} are covariances between X-Z and Y-Z, respectively; c_Y^R, c_Y^V, and c_Y^T are the linear rescaling factors for the REG1-, VAR-, and TCA-based methods, respectively; σ_X and σ_Y are the standard deviations of X and Y, respectively; and ρ_{XY} is the correlation coefficient between X and Y. Accordingly, the rescaled products are estimated as

$$Y_{REG1}^* = \mu_X + (Y - \mu_Y)c_Y^R, \quad (5)$$

$$Y_{VAR}^* = \mu_X + (Y - \mu_Y)c_Y^V, \quad (6)$$

$$Y_{TCA}^* = \mu_X + (Y - \mu_Y)c_Y^T, \quad (7)$$

where Y_{REG1}^*, Y_{VAR}^*, and Y_{TCA}^* are the rescaled products using REG1, VAR, and TCA methods, respectively.

2.1.2. Multiple Linear Regression

Above, the most general linear form (equation 1) is used to represent the relation between soil moisture products. The added utility of quadratic equations (Zwieback et al., 2016) and lagged variables (Su et al., 2014) have been recently investigated in the TCA framework. In this study, three multiple linear regression equations that take advantage of quadratic equations and lagged observations are considered:

$$Y_{REG2}^* = \mu_X + (Y - \mu_Y)c_{Y1} + (Y - \mu_Y)^2 c_{Y2}, \quad (8)$$

$$Y_{REGL_t}^* = \mu_X + (Y_t - \mu_Y)c_{Y3} + (Y_{t-1} - \mu_Y)c_{Y4}, \quad (9)$$

$$Y_{REGL2_t}^* = \mu_X + (Y_t - \mu_Y)c_{Y5} + (Y_{t-1} - \mu_Y)c_{Y6} + (Y_t - \mu_Y)^2 c_{Y7}, \quad (10)$$

where t is the time step; Y_{t-1} is the lagged version of Y_t; Y_{REG2}^*, $Y_{REGL_t}^*$, and $Y_{REGL2_t}^*$ are the rescaled products obtained using second order linear regression, lagged linear regression, and second order/lagged linear regression, respectively. In this study, only higher than second order linear regressions are not used because Zwieback et al. (2016) used second order relations, and our independent analysis also shows that second order relations yield the best results using independent validation data (results not shown). Here, even though the quadratic terms are nonlinear in the explanatory variable, in this study, they are investigated under the linear category as their regression parameters (intercept, slope, and quadratic coefficient) are linear. However, it is stressed that this choice is inconsequential and impacts neither the results nor the conclusions.

2.1.3 Multivariate adaptive regression splines

MARS (*Friedman,* 1991) is an extension of the linear regression method that handles nonlinearities and the dependence between datasets. The MARS algorithm partitions training datasets into splines (i.e., sections) with different slopes, and these splines are later smoothly connected to each other into basis functions (i.e., polynomials). Here, the role of these basic functions is to project Y (unscaled product) to a new variable H by considering a knot value (an inflection point) and hinge functions that are automatically determined by the data (Hastie et al., 2009).

The MARS algorithm consists of two phases of forward and backward stepwise procedures. In the forward stepwise procedure, the model aims to find basis functions that reduce the errors between the rescaled and reference variables the most. However, at the end of the forward phase, the algorithm produces a complex model that gives a poor response for predicting new independent data (Andres et al., 2011). In other words, the developed model in the forward phase will overfit with the training data and therefore require a backward stepwise selection to eliminate ineffective basis functions. The backward phase, in fact, prunes the model to create a more generalized model with better abilities. This phase starts its operations with the most general and simple model (i.e., the mean of the reference dataset) in the forward phase and moves forward by adding basis functions (i.e., polynomial) to the model. The least effective basis functions in the mean square sense are later eliminated, until the change in prediction error is small.

In this study, the training procedure, including the application of forward and backward steps and the locating of knot points, is conducted by using earth package (Milborrow, 2016) in the R environment (a freely available data analysis programming language; R Core Team, 2015). For more details about the MARS and its development procedure see studies of Hastie et al. (2009) and Sharda et al. (2008).

2.2. Nonlinear Rescaling Methods

2.2.1 Artificial Neural Networks

ANNs, which are originally modeled from the existing information processing paradigm of biological neural networks of the human brain (Chen & Billings, 1992), provide methods to establish relations between datasets (e.g., X and Y) through networks of neurons (nodes) in the so-called hidden layers. There are different types of ANNs available in the literature, and they can be classified with respect to their structure (i.e., numbers of layers and the way in which their neurons are connected), training method, and activation function.

The structure of ANNs can be defined depending on the nature of the problem and datasets. Strictly linear systems do not require any hidden layer, while the use of one or two hidden layers is sufficient to solve most (if not all) complex nonlinear problems. However, the optimality of the number of neurons has been an ongoing debate for almost two decades (Huang & Babri, 1998; Kentel, 2009; Murata et al., 1994; Sheela & Deepa, 2013; Xu and Chen, 2008) and is not as clear as the optimality of the hidden layer number.

In this study, four ANN functions [Multi-layer perceptron (MLP; Rosenblatt, 1958), Radial basis function (RBF; Poggio & Girosi, 1990), ELMAN (Elman, 1990), and JORDAN (Jordan, 1997)] with different structures that belong to feed-forward, radial basis function, and recurrent networks, are used to rescale the dataset(s) to the scale of a reference dataset. The optimum number of hidden layers and neurons for each function are separately identified through a grid search within a domain of (1-2) and (1-40) for the number of hidden layers and their neurons, respectively. ANN implementations in this study have been carried out using the RSNNS package, which was written by Bergmeir and Benitez (2012) for the R environment (R Core Team, 2015). The structural properties of the ANN functions (e.g., training method, activation functions) are chosen by following the default values and guideline of the RSNNS package given in Table 1. For more details about the networks used in this study and the differences in their parameters, readers can refer to the user manual of the RSNNS package (Bergmeir & Benitez, 2012).

2.2.2 Genetic Programming

GEN (Koza, 1994; Vladislavleva et al., 2009) is an automatic programming technique that is based on Darwin's theory of population evolution (abandoning poor members of society and creating modified children selectively). GEN uses the Genetic Algorithm (GA) to create tree-structured computer programs as a solution for defined problems (e.g., rescaling unscaled variables to the reference space).

Given the availability of relevant datasets, GEN discovers their relationship through randomly created computer programs that are composed of mathematical functions and arithmetic operators without having *a priori* information about the datasets or their structures. GEN utilizes these functions and picks the best-fitted ones (i.e., refines these functions) in a statistical sense by exchanging information through so-called crossover and mutation operators. Here, the crossover operator combines randomly selected parts of two programs and creates a new program for the new population, while the mutation operator creates a new program by randomly selecting one part of a program and randomly mutating it. This refining process evolves over a series of generations until reaching the termination criteria (e.g., evolving time, maximum generations, error threshold, etc.).

All of the steps of GEN in this study are performed by using the RGP package (Flasch et al., 2014) in the R language programming environment. The preliminary required parameters of GEN (e.g., the causality relationship between unscaled and reference soil moisture products, termination criteria, etc.) are presented in Table 2. The remaining required parameters (e.g., GA operator's probabilities and performing procedure of them) are defined as per their default values following the guidelines of the RGP package (Flasch et al., 2014).

2.2.3 Support Vector Machine

SVM (Vapnik & Chervonenkis., 1974; Vapnik, 1998) is a statistic-based technique for general (nonlinear) classification and regression. The SVM seeks to find the optimal function (as flat as possible) with a margin that contains all points, with an error smaller than ϵ (Hernandez et al., 2009). This flat linear function can be found by using an ϵ-insensitive loss function that penalizes errors greater than ϵ, while the trade-off between flatness and precision is determined by the regularization constant, "C", in an optimization problem as:

$$\min_{\alpha,\alpha^*}\left[\frac{1}{2}(\alpha-\alpha^*)^{T}Q(\alpha-\alpha^*)+\epsilon\sum_{i=1}^{l}(\alpha_i+\alpha_i{}^*)+\sum_{i=1}^{l}y_i(\alpha_i-\alpha_i{}^*)\right], \quad (11)$$

subject to

$$[0 \leq \alpha_i], [\alpha_i^* \leq C], [i = 1, \ldots, l], \left[\sum_{i=1}^{l}(\alpha_i - \alpha_i^*) = 0\right] \quad (12)$$

where (α_i, α_i^*) are Lagrange multipliers, C is the upper bound, Q is a l by l positive semi definite matrix, $Q_{ij} \equiv x_i x_j K(y_i, y_j)$, and $K(y_i, y_j)$ is the kernel function associated with the support vectors of (y_i, y_j). The nonlinear kernel function transforms datasets into a higher dimensional feature space, where the optimized linear function in the new feature space is equal to a nonlinear regression in the original space (Olson & Delen, 2008).

Here, the optimization of ϵ, C, and γ (parameter of kernel function) in the above equations is essential for obtaining the best regression function (Smola & Scholkopf, 2004). Therefore, once the radial basis kernel function is selected, an optimization procedure is implemented for the ϵ, C, and γ hyper parameters based on cross validation (the optimized values are not shown). The domains of the parameters that need to be optimized are 0.01-1, 1-1000, and 0.5-1 for the ϵ, C, and γ parameters, respectively (Hernandez et al., 2009; Meyer et al., 2015). In this study, the above calculations of the regression functions are performed by using the e1071 R package (Meyer et al., 2015) in the R environment. For more details about the SVM and its development procedure, see the studies by Vapnik (1998) and Smola and Scholkopf (2004), and for the e1071 R package, see the study by Meyer et al. (2015).

2.2.4 Cumulative Distribution Function Matching

The CDF (Reichle & Koster, 2004) is among the earliest implemented techniques that aim to reduce the systematic differences between soil moisture datasets by the matching the cdf of the datasets. This method has been widely used in many applications, particularly in studies that focus on data assimilation (Drusch, 2007; Li et al., 2010). CDF aims to match the rankings (i.e., cdf) of a soil moisture dataset to those of a selected reference dataset. The schematic representation of the CDF used in this study is given in Figure 1 (i.e., the path shown by the panels BADE). For more details, please see the study by Reichle and Koster (2004).

2.2.5 Copula

Copula functions are widely used to describe the multivariate dependence between random variables by using their univariate distributions. More specifically, this method enables the estimation of a multivariate cdf of random variables by using copula functions that utilize the univariate cdf of random variables, assuming the marginal probability distributions follow a uniform distribution. The general equation for the estimation of the multivariate distribution in the copula approach is described by Sklar (1959) as follows:

$$\mathrm{C}\left(\mathrm{cdf}_{\mathrm{u}_1}, \mathrm{cdf}_{\mathrm{u}_2}, \ldots, \mathrm{cdf}_{\mathrm{u}_\mathrm{N}}\right) = \Pr(\mathrm{U}_1 \leq \mathrm{u}_1, \mathrm{U}_2 \leq \mathrm{u}_2, \ldots, \mathrm{U}_\mathrm{N} \leq \mathrm{u}_\mathrm{N}) \quad (13)$$

Where C is a unique multivariate copula function that contains all of the dependence information among the datasets through a single parameter (e.g., P or θ). Here, Sklar's theorem implies that for any group of random variables $\mathrm{U}_1, \mathrm{U}_2, \ldots, \mathrm{U}_{\mathrm{N}-1}$, there exists a copula function $\mathrm{C}\left(\mathrm{cdf}_{\mathrm{u}_1}, \mathrm{cdf}_{\mathrm{u}_2}, \ldots, \mathrm{cdf}_{\mathrm{u}_\mathrm{N}}\right)$ that links these variables through an estimation of the multivariate probability distribution of these random variables.

The copula approach explicitly requires a conditional multivariate cdf to find the solution to a rescaling problem, which can be found via the partial derivative of the copula functions in the following form:

$$\mathrm{C}_{\mathrm{U}_\mathrm{N}|\mathrm{U}_1,\mathrm{U}_2,\ldots,\mathrm{U}_{\mathrm{N}-1}} = \frac{\partial \mathrm{C}(\mathrm{cdf}_{\mathrm{u}_1}, \mathrm{cdf}_{\mathrm{u}_2}, \ldots, \mathrm{cdf}_{\mathrm{u}_\mathrm{N}})}{\partial \mathrm{C}(\mathrm{cdf}_{\mathrm{u}_1}, \mathrm{cdf}_{\mathrm{u}_2}, \ldots, \mathrm{cdf}_{\mathrm{u}_{\mathrm{N}-1}})}. \quad (14)$$

Here, the goal is to first estimate $\mathrm{cdf}_{\mathrm{u}_\mathrm{N}}$ and to then retrieve the value of U_N by utilizing the cdf of the observed variables $(\mathrm{U}_1, \mathrm{U}_2, \ldots, \mathrm{U}_{\mathrm{N}-1})$. Here, these observed variables could be selected as observations from different platforms as well as lagged values of the same variable to be predicted. However, the solution of equation 14 requires knowledge of the conditional cdf of the observed variables ($cdf_{U_N|U_1,U_2,\ldots,U_{N-1}}$), which can be found through an iterative procedure (for details on this optimal solution, see the study of Leroux et al., 2014).

The schematic representation of the CDF and copula methods that rescale the variable Y to X is shown in Figure 1. In this example, the conditional cdf of 0.47 gives the optimal copula result (panel C in Figure 1), which has a curved shape compared to the projection line of the cdf (straight line in panel A). The optimal shape and location of this projection line curvature in panel C can be found by optimizing the parameters P, θ, and/or conditional cdf value, whereas the optimality depends on the goal of the application.

The list of copula functions used in this study [five total: NORMAL (Frahm et al., 2003), CLAYTON (Clayton, 1978), GUMBEL (Gumbel, 1960), FRANK (Genest, 1987), and JOE (Joe, 1997)] and their properties are given in Table 3. In this study, all of the steps, including the calculation of the CDFs and the fitting of different copulas, are performed using the R programming language package "Copula", which was written by Hofert et al. (2012). For more information about the mathematical properties of the copula function and families, fitting procedures, and simulation issues, see the studies by Genest and Favre (2004) and Nelsen (2013).

2.2.6 Lagged Types

Soil moisture is a highly autocorrelated variable; accordingly, any given day's soil moisture observations contain valuable information about the next day's actual soil moisture values. This implies that is it is viable to use lagged observations as independent observations (Crow et al., 2015; Su et al., 2014) in addition to non-lagged observations (i.e., two input time series are used to predict a single output time series). Among the rescaling methods used in this study, the performances of the lagged versions of MARS (MARSL), GEN (GENL), SVM (SVML), MLP (MLPL), RBF (RBFL), ELMAN (ELMANL), JORDAN (JORDANL), NORMAL (NORMALL), CLAYTON (CLAYTONL), GUMBEL (GUMBELL), FRANK (FRANKL), and JOE (JOEL) are also evaluated in addition to their non-lagged types.

2.3 Comparison of the Rescaling Methods

In this study, the rescaling methods are compared for their ability to minimize the error variance of Y^* ($\sigma^2_{\epsilon_{Y^*}}$), minimize the error absolute mean bias (AMB), and maximize the ρ between X and Y^* (ρ_{XY^*}). The details of these statistics are given below in chapter 4. Here, ρ_{XY^*} and ρ_{XY} are the same for all linear rescaling methods. Among the linear methods, by definition, REGL2 minimizes the $\sigma^2_{\epsilon_{Y^*}}$ of the training data; hence, REGL2 is preferable over other linear methods (REG1, REG2, REGL, VAR, and TCA) if $\sigma^2_{\epsilon_{Y^*}}$ is the selection criterion when the training and validation datasets are the same. Accordingly, the comparison of linear methods may not be meaningful given that REGL2 yields the minimum $\sigma^2_{\epsilon_{Y^*}}$, whereas all of the methods have an identical ρ_{XY^*} (if REGL3 was used, it would have further reduced the training $\sigma^2_{\epsilon_{Y^*}}$). By contrast, the optimality of REGL2 is not guaranteed when the parameters obtained using the training datasets are applied to independent validation datasets. This implies that the inter-comparison of linear methods for the validation of Y^* is still necessary before confidently making conclusions about their performances.

Linear and nonlinear methods have particular advantages and disadvantages, which impact their optimality for different applications and goals. Among the linear methods, REGL2 minimizes the mean square difference between X and Y^*, VAR matches the total variability components of X and Y, and TCA matches the signal variability components of Y and X so that the error variance of the analysis in data assimilation framework is minimized (Yilmaz & Crow, 2013). Accordingly, the applications that aim to *linearly* create a homogenous dataset for which Y^* is closest to X (i.e., those that seek to minimize mean square errors) may prefer REGL2 (assuming that REGL2 does not severely overfit the datasets). MARS is expected to yield better results than the other linear methods (due to their advantage of the use of splines at different knot points), but this expectation may not be analytically proven because REG2 and REGL2 take advantage of quadratic relations. Given that merging-type studies (e.g., data assimilation) explicitly require the signal variability components of Y^* and X to be the same, TCA is a better candidate for such studies (Yilmaz & Crow, 2013). Among the nonlinear rescaling methods, copula links the CDF_X and CDF_Y multivariate functions instead of matching them, similar to CDF (Figure 1). By contrast, ANN, GEN, and SVM machine-learning methods establish the relationships between datasets and act like a system in which the input-output relations may be too complex to be shown explicitly with equations or perhaps cannot be shown at all. When ANN and GEN are compared, GEN has an advantage: first, the assembly of blocks (i.e., the input variables, target, and mathematical functions) is defined, and then, the optimized structure of the model and its coefficients are determined during the training process. By contrast, in ANNs, the structure of the network is specified first and the coefficients are then obtained during the training process. Conversely, the main drawback of GEN is its high computational cost due to the infinite search space of symbolic expressions.

Overall, the relative performances of methods using independent datasets that are not used in their parameter estimation are not analytically predictable (including the linear methods). Hence, it may not be possible to analytically prove that any particular rescaling method will result in a superior accuracy by using independent validation data. Accordingly, a comparison of the performances of linear and nonlinear methods is still needed to attain a greater understanding of their relative added utility.

Many of the methods discussed here (ANN, GEN, SVM, and copula) have different structures and therefore different complexities. However, currently, these methods can be easily implemented in various applications using data analysis programming languages, such as R, Matlab, IDL, and Python. The available packages or toolboxes in these programming languages train the networks (e.g., optimize the weights of connections among the neurons of layers of the network) such that the considered performance statistics between the reference and predicted values are optimized. These packages require users only to define certain parameters (e.g., the number of hidden layers and neurons and type of functions that ANNs have to implement, such as learning, update, activation, and output functions; Table 2). Despite the fact that these methods have greater computational complexity (i.e., much longer codes running in the background) than other simpler rescaling methods (e.g., linear methods and CDF), these complex methods can be implemented using a couple of lines of codes that run for a very short time, similar to less-complex methods, once the optimized parameter sets are obtained (this optimization phase of these complex methods could require relatively longer computational times). Hence, there is relatively very little difference between the simpler methods (e.g., linear methods) and the more complex methods (e.g., machine learning methods), especially in terms of the computational ease of implementing these rescaling methods, except for the optimization of components.

3. Datasets

The remote sensing-based Land Parameter Retrieval Method (LPRM) soil moisture datasets (Owe et al., 2001, 2008) used in this study utilizes the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) X-band and C-band observations. These datasets are acquired between 2002 and 2009 from the Vrije Universiteit Amsterdam (personal communication with Robert Parinussa, 2013). LPRM uses three parameters (soil moisture, vegetation water content, and soil or canopy temperature) as well as passive-microwave-based, dual-polarized (either 6.925 or 10.65 GHz) observations from AMSR-E for the retrieval of both the surface soil moisture and vegetation water content. The final LPRM soil moisture dataset is gridded to a spatial resolution of 0.25° and has a daily temporal resolution with a revisit time of ~3 day. AMSR-E stopped transmitting data in October 2011 due to antenna problems, and the continuation of LPRM datasets will use observations retrieved from other sensors, such as the Advanced Microwave Scanning Radiometer-2 (*Parinussa et al.,* 2015) and Fengyun-3B (Parinussa et al., 2014). For more details on the LPRM retrieval method, please see the studies by Owe et al. (2001, 2008).

The watershed average in situ soil moisture datasets are obtained for the LPRM local overpass time over the four USDA ARS watersheds: Little River (LR), Little Washita (LW), Walnut Gulch (WG), and Reynolds Creek (RC). These four watersheds contain dense soil moisture sensors (each watershed contains 16 to 29 stations over a 150 to 610 km^2 area, less than a single LPRM pixel area) that make soil moisture measurements at depths from 0 to 5 cm and at intervals of 20 to 60 min over forest, grazing land, semiarid, and mountainous climatic regions. The areas of these watersheds are smaller than one LPRM pixel area. Soil moisture measurements at different stations are averaged to obtain a time series that is representative of each watershed (Jackson et al., 2010). Verification of these watershed average datasets has been performed via comparisons against gravimetric soil moisture observations (Cosh et al., 2006, 2008). These datasets were previously used to validate AMSR-E and Soil Moisture and Ocean Salinity (SMOS) surface soil moisture products (Jackson et al., 2010, 2012). Watershed average datasets are acquired through the International Soil Moisture Network (ISMN; Dorigo et al., 2011). Given that these datasets are available only between June 2002 and July 2009 from the ISMN database, this study is limited between these dates, even though the LPRM dataset is available beyond 2009. Among the available data between these dates, there are 0 soil moisture values for 131, 2, and 52 days, for LW, WG, and RC, respectively; these 0 values are assumed to be missing and are not used in the analyses performed in this study.

Among the linear methods, TCA requires the use of a third product (along with the watershed average and LPRM datasets) to estimate the rescaling coefficient (Stoffelen, 1998). For this purpose, Noah land surface model version 2.7 (Ek et al., 2003) simulations obtained from Global Land Data Assimilation System (GLDAS) simulations (Rodell et al., 2004) are used as the third product in the TCA calculations. NOAH soil moisture simulations representing the top 10 cm from four USDA ARS watersheds are retrieved at a spatial resolution of 0.25° for the LPRM local overpass time. These datasets are obtained from the Goddard Earth Sciences Data and Information System (http://hydro1.sci.gsfc.nasa.gov/dods/). For more information about the dataset, see the study by Rodell et al. (2004).

4. Added Utility of Rescaling Methods

In this study, the LPRM soil moisture values are rescaled to watershed average datasets using linear (VAR, TCA, REG1, REG2, REGL, REGL2 and MARS) and nonlinear (CDF, GEN, SVM, ANN, and copula) methods, where ANN has four types (MLP, RBF, ELMAN, and JORDAN) and copula has five types (NORMAL, CLAYTON, GUMBEL, FRANK, and JOE). Additionally, 12 lagged types are also considered (MARSL, GENL, SVML, MLPL, RBFL, ELMANL, JORDANL, NORMALL, CLAYTONL, GUMBELL, FRANKL, and JOEL). Overall, 31 different methods are considered in this study (7 linear, 12 nonlinear, and 12 lagged methods).

The parameters obtained using training data are later used to rescale the LPRM validation datasets, and the accuracy of the rescaled LPRM datasets ($LPRM^*$) is later assessed using independent watershed average validation datasets and the statistics below:

$$\varepsilon_i = Sta_i - LPRM_i^* \tag{15}$$

$$AMB_i = |\mu_{\varepsilon_i}| \tag{16}$$

$$\sigma_{\varepsilon_i} = \sqrt{\sum\left(\varepsilon_i - \mu_{\varepsilon_i}\right)^2 / (n-1)} \tag{17}$$

$$\rho_i = \frac{\Sigma_{Sta_i LPRM_i^*}}{\sigma_{Sta_i}\sigma_{LPRM_i^*}} \tag{18}$$

where subscript i indicates each watershed (total four), Sta is the station-based watershed average dataset, ε is the error of $LPRM^*$, and μ_ε and σ_ε indicate the temporal mean and standard deviation of the errors, respectively, AMB indicates the error absolute mean bias, n is the number of available observations, and $\sum(\ )$ is the summation operator. The statistics ρ, σ_ε, and AMB are calculated over four watersheds separately.

Only mutually available LPRM and watershed average datasets are used to calculate all of the statistics (equations 16-18) in this study. The datasets are divided into training and validation parts. Some rescaling methods that are explicitly used in the autocorrelation information to rescale, traini and validate datasets cannot be selected via random sampling; accordingly, temporally continuous data are selected for training and validation. To reduce the impact of sampling errors on the results, two separate experiments are implemented: the first experiment uses the first (time-wise) 25% of the data for validation and the remaining 75% for training, whereas the second experiment uses the first 75% for training and the remaining 25% for validation. Later, the statistics (equation 16-18) for these two experiments are averaged, and these averages are presented in this study.

The added utility (U) of the rescaling methods is calculated with respect to the performance of the REG1 method:

$$U_{m,s,l} = M_{m,s,l} - REG1_{s,l}, \qquad (19)$$

where m represents 9 methods (listed below), s represents 4 locations (LR, LW, WG, and RC), l represents 3 statistics (ρ, σ_ε, and AMB is obtained as the average of above defined two experiments), M represents the method of interest, and U is the added utility with respect to REG1. To ensure that U is always positive for the improvements and negative for the degraded results, the bias and standard deviation statistics are multiplied by -1. U is calculated only for the following selected methods: i) REGL2, ii) better performing MARS and MARSL, iii) CDF, iv) better performing GEN and GENL, v) better performing of SVM and SVML, vi) best performing type of copula (including all of the lagged types), vii) best performing type of ANN (including all of the lagged types), viii) the method (among the 31 methods) that gives the best statistical training ("Tr_best"), and ix) the method that gives the best statistics when the validation data are used ("Best"). For example, if MARSL gives the best ρ over LR using training data, then MARSL is selected as the "Tr_best" method for ρ over LR, whereas another method may perform better using the validation data ("Best"). Comparisons of U are performed separately over four watersheds. Similarly, these comparisons are repeated for each performance statistic (ρ, σ_ε, and AMB, 3 total).

5. Results and Discussion

The statistics of the LPRM and watershed average soil moisture datasets are analyzed (Table 4) prior to evaluating the results of the rescaling experiment. On average, there are 1600 days where the LPRM and watershed average data are mutually available between June 2002 and July 2009. Two different experiments are conducted using two different training datasets, and validation dataset are used to check the consistency of the results. On average, 1200 of the available data points are used for training, for both experiments, whereas the remaining (~400) unused data points are left for independent validation. Overall, the statistics (μ, σ, and lag1 autocorrelation) of the datasets (Table 4) are very similar for the training and validation periods for both experiments (statistical significance tests are not performed). Unscaled original LPRM time series have 2-4 times larger μ and σ than the watershed average time series, which can also be seen in the scatterplots of the datasets (Figure 2, upper row). This clearly shows that these datasets should be reconciled in some statistical sense (e.g., Figure 2, middle row) before they can be meaningfully compared or used to create a homogenous and consistent time series. The watershed average time series has 3.4%, 4.5%, 0.1%, and 5.5% missing data (results not shown) for the LR, LW, WG, and RC watersheds, respectively. The time series obtained over LR and RC have more missing data than those obtained over LW and WG, yet the autocorrelation values over RC are statistically significantly higher than the values over LW, WG, and WG (for both the LPRM and watershed average datasets). Higher autocorrelation values, despite more missing data, imply calculation differences between RC, and the remaining 3 watershed average data could be real; they may not be considerably impacted by the missing data, even though the LPRM autocorrelations are, on average, 0.10 lower than the watershed average values (perhaps due to the higher noise component).

The statistics ρ, AMB, and σ_ε (equations 16-18) for the 31 experiments for the training and validation periods are presented in Tables 5-6 and Figures 3-5. Table 5 shows the training and validation results numerically. Figures 3 and 4 show the average values obtained for four watersheds using the data presented in Table 5. Table 6 shows the added utility of the methods (only the best performing types are presented). Figure 5 represents the average values obtained for the four watersheds presented in Table 6. Here, the U values are calculated with respect to the REG1 values (Table 5) using equation 17. In general, a higher ρ is almost always associated with a lower σ_ε for both validation and training datasets (Tables 5-6), implying that these statistics are consistent when representing the accuracy of the analyzed dataset. Overall, the relative performances of these 31 methods are very consistent for the training and validation datasets (i.e., better performing methods using training datasets also performed better when using validation datasets). This consistency can also be seen in the U values (Table 6 and Figure 5). This provides inferences about the relative performances of these rescaling methods when using training datasets, which could provide very meaningful information about independent data scenarios. The consistency between the training and validation results also supports the selection of training and validation periods; these two periods may not have a considerable difference in terms of the relation between the LPRM and watershed average data, as well as in terms of the relative performances of the rescaling methods.

On average, the GENL and ELMANL ANN methods yield a ρ improvement of ~0.05 using independent validation datasets. This improvement is lower (0.02 - 0.04 ρ improvement) for the SVML, MARSL, REGL2, and NORMALL methods (Figure 5 and Table 6). In contrast to its wide use, the CDF method has no added skill (Figure 5); in fact, on average, it yields degraded correlations compared to REG1 when validated using independent data (Table 6). When the method selection is consistent with the training results, these Tr_best methods yield better U values than any method alone, with U values that are similar to the best validation results (“Best”) approximately 75% of the time (Figure 5). These results further support the above discussion that it is better to make a rescaling method selection that is consistent with the training data statistics, when this selection can yield better validation results than the selection of any other method alone.

When the results are averaged over all of the watersheds, all of the nonlinear methods (except for JOE) demonstrated improved correlations compared to the REG1 correlations using the training datasets (Figure 3). When validation datasets are used, MARS, GEN, SVM, all four ANNs, and NORMAL still have superior correlations compared to REG1 (Table 5 and Figure 4). In particular, the improvements over LR, LW, and RC using GENL, SVML, and ELMANL (0.083, 0.090, and 0.135), respectively, are much higher than the improvements over other locations via various methods (Table 6). Compared to the best performing linear method using the validation data (MARSL), on average, the GENL, SVML, ELMAN, ELMANL, JORDAN, and JORDANL nonlinear methods yielded better results (Figure 4). These outcomes stressed the results of the first-order linear regression, which can be improved via higher order or more complex linear methods, and there is still added utility that can be gained via nonlinear methods compared to linear methods. Thus, nonlinear methods have a higher potential to give more accurate results compared to linear methods, and as a result, the existing nonlinear relations cannot be captured through linear methods.

Soil moisture products have high autocorrelation; hence, two of the most recent soil moisture observations have a high linear dependence (Table 4 and Figure 2 bottom row). The use of lagged observations, in addition to the unscaled observations in a first-order linear framework (REGL), improves the statistics compared to REG1. However, the GEN and SVM methods yield, on average, better improvements than linear methods, such as REG1 and MARS (Table 7), particularly over LR and LW (the ρ difference between GENL and GEN over LR is 0.086 and the SVML and SVM difference over LW is 0.083). These results show that the overall nonlinear methods better utilize the lagged observation information (Table 4) and have a higher potential to improve the results compared to the linear methods, even though the degree of improvement varies for different methods. The ANN methods do not have much added skill via lagged observations, perhaps because these methods already utilize the lagged observation information. These results further highlight the higher potential of nonlinear methods in rescaling soil moisture datasets.

When the parameters obtained using the training datasets are implemented over the validation datasets, some skill loss (i.e., artificial skill) is often observed because all of the methods overfit their datasets to some extent. Loosely speaking, an increase of 0.06 or 0.10 in ρ constitutes a statistically significant increase, especially when 1200 or 400 samples are used for training or validation experiments, respectively (e.g., an increase from 0.60 to 0.66 or from 0.60 to 0.70). Accordingly, MARS, SVM, and ELMAN yield significant ρ improvements (with respect to REG1 ρ) over half of the training cases, whereas GEN, FRANK, and JORDAN also yield significant improvements over some locations (Table 5; most of the training improvements are over LW and RC, and only a few are over WG). By contrast, for validation experiments, only ELMAN and JORDAN resulted in significant ρ improvements (both over RC), showing that most of these improvements are artificial skills. Here, the degree to which the methods overfit the datasets is evaluated through the comparisons of ρ for the validation datasets (Figure 4) versus the training datasets (Figure 3), where higher differences indicate a higher degree of artificial skill. These differences show the artificial skill in ρ to be approximately 50% for ANN (ELMANL only have 18% artificial skill); ~65-80% for GEN, MARS, and SVM; and ~ 100% for REGL2, CDF and the copula methods, on average (NORMALL has an artificial skill of only 65%). These results stress the use of independent validation data to avoid artificial skill.

The skills of nonlinear methods are heavily impacted by the number of iterations performed to optimally obtain certain parameters. By contrast, increasing the degree of these iterations eventually results in overtraining and hence overfitting. For example, in this study, the maximum number of iterations for ANN simulations is set at 1000. When this number is increased to 100,000, training correlations can be obtained between the reference and rescaled products (as high as 0.90 for certain cases). However, this gained training skill is quickly lost when the obtained ANN configurations and parameters are utilized on independent validation data. Such dramatic differences are more common for ANN than other methods (GEN, SVM, and copula), whereas the degree of overfitting using other methods does not depend as much on user specifications as ANN (results not shown).

Among the copula methods, CLAYTON, GUMBEL, and JOE have asymmetric tail dependence properties (strong in one tail and weak in the other) and do not perform as well as NORMAL or FRANK, which have symmetric tail dependence for both training and validation experiments (Table 5). Both the copula and CDF methods use CDF_X and CDF_Y to rescale observations. However, it is stressed that the performances of copula methods are very sensitive to the $C_{X|Y}$ values (equation 16), which are selected during training. The optimality of these $C_{X|Y}$ values depends on the objective of the training process (e.g., the minimization of AMB only, the maximization of ρ only, the minimization AMB and σ_ϵ simultaneously, or the minimization of AMB and σ_ϵ, and the maximization of ρ simultaneously). In this study, the penalty function is formed and $C_{X|Y}$ values are obtained in a way that training is penalized for increased AMB and σ_ϵ and decreased ρ. Investigations for the added utility of lagged observations show only Normal

Copula (Elliptical family) utilizing this information, whereas the remaining copula types (Archimedean family) result in degraded rescaled products, especially when lagged observations are also used and validated using independent data (Table 5 and Figure 4). This result is consistent with the study of Afshar et al. (2016), who found the Elliptical family to be better at capturing the dependency among variables than the copula functions of the Archimedean family.

6. Conclusions

In this study, LPRM soil moisture datasets are rescaled to station-based datasets over four USDA ARS watersheds to reduce the systematic differences between datasets. The rescaled datasets are validated by using independent data that are not used in the training part. This study is the first to perform a comprehensive comparison of the performances of various linear (VAR, TCA, REG1, REG2, REGL, REGL2, and MARS) and nonlinear (CDF, GEN, SVM, ANN, and copula) methods (total 31 methods); the first to use the REG2, REGL, REGL2, MARS, GEN, SVM, and ANN methods to explicitly rescale the soil moisture datasets in the framework of soil moisture rescaling; and the first to comprehensively investigate the added utility of lagged observations in the soil moisture rescaling framework.

The relative performances of methods using training and validation datasets are consistent; the rescaling method that results in a more accurate rescaled product using training data also results in a more accurate rescaled product using validation data, and the best performing method using the training datasets yields better results than any other individual method that uses the validation datasets. Although the actual performances of the rescaling methods might change for different datasets, it is viable that a similar consistency would also exist for other datasets that are not used in this study. Such a consistency between the training and validation results gives confidence to the user in their selection of the rescaling method, particularly in the operational implementation of rescaling methods.

A large majority of the related variability between products are due to first-order linear relations. Although multiple linear regression-based rescaling methods slightly improve the rescaled product statistics, the training and the validation statistics consistently show that nonlinear methods resulted in a more accurate rescaled product than linear methods. Overall, GENL and ELMAN improved independent validation dataset correlations the best (on average 0.05), whereas improvements reached as high as 0.14 at individual locations (ELMAN over RC).

Among nonlinear methods, ELMAN exhibits superior performance, particularly when the datasets are highly autocorrelated (over RC), whereas the GEN and SVM methods exhibit superior performance when the lagged observations are also used as predictors (over LR and LW). Although lagged observations improve the rescaled product statistics when datasets are rescaled linearly, nonlinear methods yield better statistics than linear methods. This highlights that lagged observations, which contain valuable information in the soil moisture rescaling framework as in the TCA framework (Crow et al., 2015; Su et al., 2014; Zwieback et al., 2013). Nonlinear methods have higher added utility potential than linear methods in using lagged observations, in addition to their overall higher rescaling potential compared to the linear methods.

The higher rescaling potential and lagged observation utilization potential compared to linear methods clearly show that the soil moisture datasets used in this study have nonlinear relations that cannot be modeled using linear methods. It is also viable that such nonlinear relationships may exist between other soil moisture datasets that are not used in this study. These results imply that the soil moisture inter-comparison studies (Albergel et al., 2012; Brocca et al., 2011; Hain et al., 2011; Mladenova et al., 2014; Parinussa et al., 2014; Wagner et al., 2014) and non-data assimilation type blending studies (Leroux et al., 2014; Liu, et al., 2012, 2014) may benefit from these nonlinear rescaling methods, given the key results in this study. The performance metrics (ρ, σ_ε, and AMB) can be considerably (in some cases statistically significantly) improved via such nonlinear methods, whereas their degree of improvements may be dataset specific.

Recent studies highlight the utility of simple API models compared to more complex models (Crow et al., 2012; Han et al., 2014; Yilmaz et al., 2016), particularly in studies aiming to methodologically improve current techniques (Crow & Yilmaz, 2014; Yilmaz & Crow, 2013). Given that such simple models have better skills in drought studies (Crow et al., 2012), such models can be used to create long and homogenous time series, expanding to historical dates, where precipitation observations are available. To ensure the consistency of the units of the model values with traditional ground observations, this model time series could be rescaled to available ground observations, relying on the consistency found between the training and the validation datasets, where mutually available datasets can be used to retrieve the necessary parameters.

Overall, it is likely that more accurate nonlinearly rescaled products will improve applications that are better related to studies using linearly rescaled products. For example, assimilation experiments require observations to be rescaled into model space before they can be merged. By definition, an assimilation of more accurate observations (e.g., obtained via nonlinearly rescaling methods) in models always results in a more accurate analysis than the assimilation of less accurate observations (unless the underlying assumptions are not met). On the one hand, Yilmaz and Crow (2013) show an assimilation analysis accuracy that depends on the degree to which the signal component of observations should be rescaled to the signal component of the model, rather than the overall product differences that are alleviated directly, as done in this study. Similarly, Su et al. (2014) and Zwieback et al. (2016) show that matching this signal component is also very important for error characterization. Consistently, Yilmaz and Crow (2013) demonstrate TCA matching of the signal components of the datasets and a better rescaling method than REG1 in the assimilation framework. The current study does not involve assimilation experiments and does not compare the actual signal components of the datasets; hence, only the explicit use of nonlinear methods in the assimilation framework (future study) may convey the real added utility via such nonlinear methods in assimilation experiments. On the other hand, an analysis accuracy improvement through the use of more accurate observations is inherent to the definition of assimilation studies. It is our expectation that the marginal gains in the rescaled dataset accuracy (e.g., ~ 0.02 ρ improvements) might not translate into large gains in assimilation analysis errors, whereas statistically significant improvements (e.g., 0.10 – 0.14) might translate into meaningful assimilation analysis improvements. Again, this expectation needs to be validated using a dedicated assimilation experiment.

7. Acknowledgments

The authors would like to thank three anonymous reviewers for their constructive comments. The authors would also like to thank the International Soil Moisture Network for the USDA ARS station-based soil moisture datasets, Vrije Universiteit Amsterdam (Robert Parinussa, personal communication) for the LPRM datasets, and NASA for the GLDAS datasets (downloaded from http://mirador.gsfc.nasa.gov). This research was supported by the EU Marie Curie Seventh Framework Programme FP7-PEOPLE-2013-CIG project number 630110 and The Scientific and Technological Research Council of Turkey (TUBITAK) Grant 3501 project number 114Y676.

8. References

Afshar, M.H., Sorman, A.U., Yilmaz, M.T., 2016. Conditional Copula-Based Spatial-Temporal Drought Characteristics Analysis – Case Study over Turkey, Water-141409

Albergel, C., de Rosnay, P., Gruhier, C., Muñoz-Sabater, J., Hasenauer, S., Isaksen, L., Kerr, Y., Wagner, W., 2012. Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. Remote Sens. Environ. 118, 215–226. doi:10.1016/j.rse.2011.11.017

Anderson, W.B., Zaitchik, B.F., Hain, C.R., Anderson, M.C., Yilmaz, M.T., Mecikalski, J., Schultz, L., 2012. Towards an integrated soil moisture drought monitor for East Africa. Hydrol. Earth Syst. Sci. 16, 2893–2913. doi:10.5194/hess-16-2893-2012

Andrés Suárez, J., Lorca Fernández, P., Cos Juez, F.J. de, Sánchez Lasheras, F., 2011. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). Scopus.

Bergmeir, C.N., Benítez Sánchez, J.M., 2012. Neural Networks in R Using the Stuttgart Neural Network Simulator: RSNNS.

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., Bittelli, M., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. Remote Sens. Environ. 115, 3390–3408. doi:10.1016/j.rse.2011.08.003

Brocca, L., Melone, F., Moramarco, T., Wagner, W., Albergel, C., 2013. Scaling and Filtering Approaches for the Use of Satellite Soil Moisture Observations, in: Remote Sensing of Energy Fluxes and Soil Moisture Content. CRC Press, pp. 411–426. doi:10.1201/b15610-21

CHEN, S., BILLINGS, S.A., 1992. Neural networks for nonlinear dynamic system modelling and identification. Int. J. Control 56, 319–346. doi:10.1080/00207179208934317

CLAYTON, D.G., 1978. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65, 141–151. doi:10.1093/biomet/65.1.141

Cortes, C., Vapnik, V., 1995. Support-Vector Networks. Mach. Learn. 20, 273–297. doi:10.1023/A:1022627411411

Cosh, M.H., Jackson, T.J., Moran, S., Bindlish, R., 2008. Temporal persistence and stability of surface soil moisture in a semi-arid watershed. Remote Sens. Environ. 112, 304–313. doi:10.1016/j.rse.2007.07.001

Cosh, M.H., Jackson, T.J., Starks, P., Heathman, G., 2006. Temporal stability of surface soil moisture in the Little Washita River watershed and its applications in satellite soil moisture product validation. J. Hydrol. 323, 168–177. doi:10.1016/j.jhydrol.2005.08.020

Crow, W.T., Crow, W.T., 2007. A Novel Method for Quantifying Value in Spaceborne Soil Moisture Retrievals. J. Hydrometeorol. 8, 56–67. doi:10.1175/JHM553.1

Crow, W.T., Koster, R.D., Reichle, R.H., Sharif, H.O., 2005. Relevance of time-varying and time-invariant retrieval error sources on the utility of spaceborne soil moisture products. Geophys. Res. Lett. 32, L24405. doi:10.1029/2005GL024889

Crow, W.T., Kumar, S. V., Bolten, J.D., 2012. On the utility of land surface models for agricultural drought monitoring. Hydrol. Earth Syst. Sci. 16, 3451–3460. doi:10.5194/hess-16-3451-2012

Crow, W.T., Su, C.-H., Ryu, D., Yilmaz, M.T., 2015. Optimal averaging of soil moisture predictions from ensemble land surface model simulations. Water Resour. Res. 51, 9273–9289. doi:10.1002/2015WR016944

Crow, W.T., Yilmaz, M.T., 2014. The Auto-Tuned Land Data Assimilation System (ATLAS). Water Resour. Res. 50, 371–385. doi:10.1002/2013WR014550

Crow, W.T., Zhan, X., 2007. Continental-Scale Evaluation of Remotely Sensed Soil Moisture Products. IEEE Geosci. Remote Sens. Lett. 4, 451–455. doi:10.1109/LGRS.2007.896533

De Andrés, J., Lorca, P., de Cos Juez, F.J., Sánchez-Lasheras, F., 2011. Bankruptcy forecasting: A hybrid approach using Fuzzy c-means clustering and Multivariate Adaptive Regression Splines (MARS). Expert Syst. Appl. 38, 1866–1875. doi:10.1016/j.eswa.2010.07.117

Dirmeyer, P.A., Guo, Z., Gao, X., Dirmeyer, P.A., Guo, Z., Gao, X., 2004. Comparison, Validation, and Transferability of Eight Multiyear Global Soil Wetness Products. J. Hydrometeorol. 5, 1011–1033. doi:10.1175/JHM-388.1

Dorigo, W.A., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., Gruber, A., Drusch, M., Mecklenburg, S., van Oevelen, P., Robock, A., Jackson, T., 2011. The International Soil Moisture Network: a data hosting facility for global in situ soil moisture measurements. Hydrol. Earth Syst. Sci. 15, 1675–1698. doi:10.5194/hess-15-1675-2011

Dorigo, W., de Jeu, R., Chung, D., Parinussa, R., Liu, Y., Wagner, W., Fernández-Prieto, D., 2012. Evaluating global trends (1988-2010) in harmonized multi-satellite surface soil moisture. Geophys. Res. Lett. 39, n/a-n/a. doi:10.1029/2012GL052988

Draper, C.S., Walker, J.P., Steinle, P.J., de Jeu, R.A.M., Holmes, T.R.H., 2009. An evaluation of AMSR–E derived soil moisture over Australia. Remote Sens. Environ. 113, 703–710. doi:10.1016/j.rse.2008.11.011

Drusch, M., 2007. Initializing numerical weather prediction models with satellite-derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set. J. Geophys. Res. 112, D03102. doi:10.1029/2006JD007478

Drusch, M., Wood, E.F., Gao, H., 2005. Observation operators for the direct assimilation of TRMM microwave imager retrieved soil moisture. Geophys. Res. Lett. 32, L15403. doi:10.1029/2005GL023623

Ek, M.B., Mitchell, K.E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G., Tarpley, J.D., 2003. Implementation of Noah land surface model advances in the National Centers for Environmental Prediction operational mesoscale Eta model. J. Geophys. Res. 108, 8851. doi:10.1029/2002JD003296

Elman, J.L., 1990. Finding structure in time. Cogn. Sci. 14, 179–211. doi:10.1016/0364-0213(90)90002-E

Flasch, O., Mersmann, O., Bartz-Beielstein, T., Stork, J., Zaefferer, M., 2015. R genetic programming framework. R package version 0.4-1.

Frahm, G., Junker, M., Szimayer, A., 2003. Elliptical copulas: applicability and limitations, Statistics & Probability Letters. doi:10.1016/S0167-7152(03)00092-0

Friedman, J.H., 1991. Multivariate Adaptive Regression Splines. Ann. Stat. 19, 1–67. doi:10.1214/aos/1176347963

Genest, C., 1987. Frank's family of bivariate distributions. Biometrika 74, 549–555. doi:10.1093/biomet/74.3.549

Genest, C., Favre, A.-C., 2007. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. J. Hydrol. Eng. 12, 347–368. doi:10.1061/(ASCE)1084-0699(2007)12:4(347)

Gill, M.K., Asefa, T., Kemblowski, M.W., McKee, M., 2006. SOIL MOISTURE PREDICTION USING SUPPORT VECTOR MACHINES. J. Am. Water Resour. Assoc. 42, 1033–1046. doi:10.1111/j.1752-1688.2006.tb04512.x

Gruber, A., Su, C.-H., Zwieback, S., Crow, W., Dorigo, W., Wagner, W., 2016. Recent advances in (soil moisture) triple collocation analysis. Int. J. Appl. Earth Obs. Geoinf. 45, 200–211. doi:10.1016/j.jag.2015.09.002

Guang-Bin Huang, Babri, H.A., 1998. Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. IEEE Trans. Neural Networks 9, 224–229. doi:10.1109/72.655045

Gumbel, E.J., 1960. Distributions des valeurs extrˆemes en plusieurs dimensions, Publications de l'Institut de Statistique de l'Universit´e de Paris. Distributions des valeurs extrˆemes en plusieurs dimensions, Paris, France.

Hain, C.R., Crow, W.T., Mecikalski, J.R., Anderson, M.C., Holmes, T., 2011. An intercomparison of available soil moisture estimates from thermal infrared and passive microwave remote sensing and land surface modeling. J. Geophys. Res. Atmos. 116, 1–18. doi:10.1029/2011JD015633

Han, E., Crow, W.T., Holmes, T., Bolten, J., 2014. Benchmarking a Soil Moisture Data Assimilation System for Agricultural Drought Monitoring. J. Hydrometeorol. 15, 1117–1134. doi:10.1175/JHM-D-13-0125.1

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning, Springer Series in Statistics. Springer New York, New York, NY. doi:10.1007/978-0-387-84858-7

Hernández, N., Kiralj, R., Ferreira, M.M.C., Talavera, I., 2009. Critical comparative analysis, validation and interpretation of SVM and PLS regression models in a QSAR study on HIV-1 protease inhibitors. Chemom. Intell. Lab. Syst. 98, 65–77. doi:10.1016/j.chemolab.2009.04.012

Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2012. Copula: Multivariate Dependence with Copulas. R package version 0.999-13.

Jackson, T.J., Cosh, M.H., Bindlish, R., Starks, P.J., Bosch, D.D., Seyfried, M., Goodrich, D.C., Moran, M.S., Du, J., 2010. Validation of Advanced Microwave Scanning Radiometer Soil Moisture Products. IEEE Trans. Geosci. Remote Sens. 48, 4256–4272. doi:10.1109/TGRS.2010.2051035

Jackson, T.J., Bindlish, R., Cosh, M.H., Zhao, T., Starks, P.J., Bosch, D.D., Seyfried, M., Moran, M.S., Goodrich, D.C., Kerr, Y.H., Leroux, D., 2012. Validation of Soil Moisture and Ocean Salinity (SMOS) Soil Moisture Over Watershed Networks in the U.S. IEEE Trans. Geosci. Remote Sens. 50, 1530–1543. doi:10.1109/TGRS.2011.2168533

Joe, H., 1997. Multivariate models and multivariate dependence concepts. CRC Press, US.

Jordan, M.I., 1997. Serial Order: A Parallel Distributed Processing Approach. North-Holland/Elsevier Science Publishers, pp. 471–495. doi:10.1016/S0166-4115(97)80111-2

Kentel, E., 2009. Estimation of river flow by artificial neural networks and identification of input vectors susceptible to producing unreliable flow estimates. J. Hydrol. 375, 481–488. doi:10.1016/j.jhydrol.2009.06.051

Koster, R.D., Dirmeyer, P.A., Guo, Z., Bonan, G., Chan, E., Cox, P., Gordon, C.T., Kanae, S., Kowalczyk, E., Lawrence, D., Liu, P., Lu, C.-H., Malyshev, S., McAvaney, B., Mitchell, K., Mocko, D., Oki, T., Oleson, K., Pitman, A., Sud, Y.C., Taylor, C.M., Verseghy, D., Vasic, R., Xue, Y., Yamada, T., 2004. Regions of Strong Coupling Between Soil Moisture and Precipitation. Science (80-.). 305, 1138–1140. doi:10.1126/science.1100217

Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., Puma, M.J., Koster, R.D., Guo, Z., Yang, R., Dirmeyer, P.A., Mitchell, K., Puma, M.J., 2009. On the Nature of Soil Moisture in Land Surface Models. J. Clim. 22, 4322–4335. doi:10.1175/2009JCLI2832.1

Koza, J., 1994. Genetic programming as a means for programming computers by natural selection. Stat. Comput. 4, 87–112. doi:10.1007/BF00175355

Lawrence, J.E., Hornberger, G.M., 2007. Soil moisture variability across climate zones. Geophys. Res. Lett. 34, 1–5. doi:10.1029/2007GL031382

Leroux, D.J., Kerr, Y.H., Wood, E.F., Sahoo, A.K., Bindlish, R., Jackson, T.J., 2014. An Approach to Constructing a Homogeneous Time Series of Soil Moisture Using SMOS. IEEE Trans. Geosci. Remote Sens. 52, 393–405. doi:10.1109/TGRS.2013.2240691

Liu, D., Yu, Z., L, H., 2010. Data assimilation using support vector machines and ensemble Kalman filter for multi-layer soil moisture prediction. Water Sci. Eng. 3, 361–377. doi:10.3882/j.issn.1674-2370.2010.04.001

Liu, Y.Y., Dorigo, W.A., Parinussa, R.M., de Jeu, R.A.M., Wagner, W., McCabe, M.F., Evans, J.P., van Dijk, A.I.J.M., 2012. Trend-preserving blending of passive and active microwave soil moisture retrievals. Remote Sens. Environ. 123, 280–297. doi:10.1016/j.rse.2012.03.014

Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., De Jeu, R.A.M., Wagner, W., van Dijk, A.I.J.M., McCabe, M.F., Evans, J.P., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sci. 15, 425–436. doi:10.5194/hess-15-425-2011

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C., Lin, C., 2015. Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), R package version 1.6-7.

Milborrow, S., n.d. earth: Multivariate Adaptive Regression Splines, R package version 4.4.5.

Miralles, D.G., De Jeu, R.A.M., Gash, J.H., Holmes, T.R.H., Dolman, A.J., 2011. Magnitude and variability of land evaporation and its components at the global scale. Hydrol. Earth Syst. Sci. 15, 967–981. doi:10.5194/hess-15-967-2011

Mladenova, I.E., Jackson, T.J., Njoku, E., Bindlish, R., Chan, S., Cosh, M.H., Holmes, T.R.H., de Jeu, R.A.M., Jones, L., Kimball, J., Paloscia, S., Santi, E., 2014. Remote monitoring of soil moisture using passive microwave-based techniques — Theoretical basis and overview of selected algorithms for AMSR-E. Remote Sens. Environ. 144, 197–213. doi:10.1016/j.rse.2014.01.013

Murata, N., Yoshizawa, S., Amari, S., 1994. Network information criterion-determining the number of hidden units for an artificial neural network model. IEEE Trans. Neural Networks 5, 865–872. doi:10.1109/72.329683

Nelsen, R.B., 1999. An Introduction to Copulas, Lecture Notes in Statistics. Springer New York, New York, NY. doi:10.1007/978-1-4757-3076-0

Notarnicola, C., Angiulli, M., Posa, F., 2008. Soil moisture retrieval from remotely sensed data: Neural network approach versus Bayesian method. IEEE Trans. Geosci. Remote Sens. 46, 547–557. doi:10.1109/TGRS.2007.909951

Olson, D.L., Delen, D., 2008. Advanced data mining techniques. Springer.

Owe, M., de Jeu, R., Holmes, T., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. J. Geophys. Res. 113, F01002. doi:10.1029/2007JF000769

Owe, M., de Jeu, R., Walker, J., 2001. A methodology for surface soil moisture and vegetation optical depth retrieval using the microwave polarization difference index. IEEE Trans. Geosci. Remote Sens. 39, 1643–1654. doi:10.1109/36.942542

Paloscia, S., Pampaloni, P., Pettinato, S., Santi, E., 2008. A Comparison of Algorithms for Retrieving Soil Moisture from ENVISAT/ASAR Images. IEEE Trans. Geosci. Remote Sens. 46, 3274–3284. doi:10.1109/TGRS.2008.920370

Parinussa, R.M., Holmes, T.R.H., Wanders, N., Dorigo, W.A., de Jeu, R.A.M., Parinussa, R.M., Holmes, T.R.H., Wanders, N., Dorigo, W.A., Jeu, R.A.M. de, 2015. A Preliminary Study toward Consistent Soil Moisture from AMSR2. J. Hydrometeorol. 16, 932–947. doi:10.1175/JHM-D-13-0200.1

Parinussa, R.M., Holmes, T.R.H., Yilmaz, M.T., Crow, W.T., 2011. The impact of land surface temperature on soil moisture anomaly detection from passive microwave observations. Hydrol. Earth Syst. Sci. 15, 3135–3151. doi:10.5194/hess-15-3135-2011

Parinussa, R.M., Wang, G., Holmes, T.R.H., Liu, Y.Y., Dolman, A.J., de Jeu, R.A.M., Jiang, T., Zhang, P., Shi, J., 2014. Global surface soil moisture from the Microwave Radiation Imager onboard the Fengyun-3B satellite. Int. J. Remote Sens. 35, 7007–7029. doi:10.1080/01431161.2014.960622

Parinussa, R.M., Yilmaz, M.T., Anderson, M.C., Hain, C.R., de Jeu, R.A.M., 2014. An intercomparison of remotely sensed soil moisture products at various spatial scales over the Iberian Peninsula. Hydrol. Process. 28, 4865–4876. doi:10.1002/hyp.9975

Poggio, T., Girosi, F., 1990. Networks for approximation and learning. Proc. IEEE 78, 1481–1497. doi:10.1109/5.58326

Prigent, C., Aires, F., Rossow, W.B., Robock, A., 2005. Sensitivity of satellite microwave and infrared observations to soil moisture at a global scale: Relationship of satellite observations to in situ soil moisture measurements. J. Geophys. Res. 110, D07110. doi:10.1029/2004JD005087

R Core Team, 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. Austria.

Reichle, R.H., Crow, W.T., Koster, R.D., Sharif, H.O., Mahanama, S.P.P., 2008. Contribution of soil moisture retrievals to land data assimilation products. Geophys. Res. Lett. 35, L01404. doi:10.1029/2007GL031986

Reichle, R.H., Koster, R.D., 2005. Global assimilation of satellite surface soil moisture retrievals into the NASA Catchment land surface model. Geophys. Res. Lett. 32, L02404. doi:10.1029/2004GL021700

Reichle, R.H., Koster, R.D., 2004. Bias reduction in short records of satellite soil moisture. Geophys. Res. Lett. 31, L19501. doi:10.1029/2004GL020938

Rochester, N., Holland, J., Haibt, L., Duda, W., 1956. Tests on a cell assembly theory of the action of the brain, using a large digital computer. IEEE Trans. Inf. Theory 2, 80–93. doi:10.1109/TIT.1956.1056810

Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J.K., Walker, J.P., Lohmann, D., Toll, D., Rodell, M., Houser, P.R., Jambor, U., Gottschalck, J., Mitchell, K., Meng, C.-J., Arsenault, K., Cosgrove, B., Radakovich, J., Bosilovich, M., Entin*, J.K., Walker, J.P., Lohmann, D., Toll, D., 2004. The Global Land Data Assimilation System. Bull. Am. Meteorol. Soc. 85, 381–394. doi:10.1175/BAMS-85-3-381

Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil Moisture Retrieval Using Neural Networks: Application to SMOS. IEEE Trans. Geosci. Remote Sens. 53, 5991–6007. doi:10.1109/TGRS.2015.2430845

Rosenblatt, F., 1958. The perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev. 65, 386–408. doi:10.1037/h0042519

Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., Wagner, W., 2008. A possible solution for the problem of estimating the error structure of global soil moisture data sets. Geophys. Res. Lett. 35, L24403. doi:10.1029/2008GL035599

Sharda, V.N., Prasher, S.O., Patel, R.M., Ojasvi, P.R., Prakash, C., 2008. Performance of Multivariate Adaptive Regression Splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data. Hydrol. Sci. Journal-Journal Des Sci. Hydrol. 53, 1165–1175. doi:10.1623/hysj.53.6.1165

Sheela, K.G., Deepa, S.N., Sheela, K.G., Deepa, S.N., 2013. Review on Methods to Fix Number of Hidden Neurons in Neural Networks. Math. Probl. Eng. 2013, 1–11. doi:10.1155/2013/425740

Sklar, A., 1959. Fonctions de Répartition à n Dimensions et Leurs Marges. Publications de l'Institut de Statistique de L'Université de Paris.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14, 199–222. doi:10.1023/B:STCO.0000035301.49549.88

Stoffelen, A., 1998. Toward the true near-surface wind speed: Error modeling and calibration using triple collocation. J. Geophys. Res. Ocean. 103, 7755–7766. doi:10.1029/97JC03180

Su, C.-H., Ryu, D., 2015. Multi-scale analysis of bias correction of soil moisture. Hydrol. Earth Syst. Sci. 19, 17–31. doi:10.5194/hess-19-17-2015

Su, C.-H., Ryu, D., Crow, W.T., Western, A.W., 2014. Beyond triple collocation: Applications to soil moisture monitoring. J. Geophys. Res. Atmos. 119, 6419–6439. doi:10.1002/2013JD021043

Su, C.-H., Ryu, D., Young, R.I., Western, A.W., Wagner, W., 2013. Inter-comparison of microwave satellite soil moisture retrievals over the Murrumbidgee Basin, southeast Australia. Remote Sens. Environ. 134, 1–11. doi:10.1016/j.rse.2013.02.016

Vapnik, V.N., 1998. Statistical learning theory. Wiley.

Vapnik, V.N., Chervonenkis, A.Y., 1974. Theory of Pattern Recognition [in Russian]. Nauka, USSR.

Vladislavleva, E.J., Smits, G.F., den Hertog, D., 2009. Order of Nonlinearity as a Complexity Measure for Models Generated by Symbolic Regression via Pareto Genetic Programming. IEEE Trans. Evol. Comput. 13, 333–349. doi:10.1109/TEVC.2008.926486

Xu, S., Chen, L., 2008. A novel approach for determining the optimal number of hidden layer neurons for FNN's and its application in data mining.

Yilmaz, M.T., Crow, W.T., Anderson, M.C., Hain, C., 2012. An objective methodology for merging satellite- and model-based soil moisture products. Water Resour. Res. 48, n/a-n/a. doi:10.1029/2011WR011682

Yilmaz, M.T., Crow, W.T., Ryu, D., Yilmaz, M.T., Crow, W.T., Ryu, D., 2016. Impact of Model Relative Accuracy in Framework of Rescaling Observations in Hydrological Data Assimilation Studies. J. Hydrometeorol. 17, 2245–2257. doi:10.1175/JHM-D-15-0206.1

Yilmaz, M.T., Crow, W.T., Yilmaz, M.T., Crow, W.T., 2014. Evaluation of Assumptions in Soil Moisture Triple Collocation Analysis. J. Hydrometeorol. 15, 1293–1302. doi:10.1175/JHM-D-13-0158.1

Yilmaz, M.T., Crow, W.T., Yilmaz, M.T., Crow, W.T., 2013. The Optimality of Potential Rescaling Approaches in Land Data Assimilation. J. Hydrometeorol. 14, 650–660. doi:10.1175/JHM-D-12-052.1

Yin, J., Zhan, X., Zheng, Y., Liu, J., Fang, L., Hain, C.R., Yin, J., Zhan, X., Zheng, Y., Liu, J., Fang, L., Hain, C.R., 2015. Enhancing Model Skill by Assimilating SMOPS Blended Soil Moisture Product into Noah Land Surface Model. J. Hydrometeorol. 16, 917–931. doi:10.1175/JHM-D-14-0070.1

Yin, J., Zhan, X., Zheng, Y., Liu, J., Hain, C.R., Fang, L., 2014. Impact of quality control of satellite soil moisture data on their assimilation into land surface model. Geophys. Res. Lett. 41, 7159–7166. doi:10.1002/2014GL060659

Zwieback, S., Dorigo, W., Wagner, W., 2013. Estimation of the temporal autocorrelation structure by the collocation technique with an emphasis on soil moisture studies. Hydrol. Sci. J. 58, 1729–1747. doi:10.1080/02626667.2013.839876

Zwieback, S., Scipal, K., Dorigo, W., Wagner, W., 2012. Structural and statistical properties of the collocation technique for error characterization. Nonlinear Process. Geophys. 19, 69–80. doi:10.5194/npg-19-69-2012

Zwieback, S., Su, C.-H., Gruber, A., Dorigo, W.A., Wagner, W., Zwieback, S., Su, C.-H., Gruber, A., Dorigo, W.A., Wagner, W., 2016. The Impact of Quadratic Nonlinear Relations between Soil Moisture Products on Uncertainty Estimates from Triple Collocation Analysis and Two Quadratic Extensions. J. Hydrometeorol. 17, 1725–1743. doi:10.1175/JHM-D-15-0213.1

List of Figure Captions

Figure 1: Schematic representations of the CDF and Copula based rescaling methods. The paths in the BADE and BCFE panels represent the CDF and Copula methods, respectively. $C_{X|Y} = 0.47$ is plotted with darker color in panel C to represent the best performing projection line of the Copula.

Figure 2: Scatter plot of the Watershed average and LPRM soil moisture data over four watersheds. Original (unscaled) and rescaled data are given in the upper and middle rows, respectively; lagged unscaled LPRM vs unscaled LPRM are given in the lower row.

Figure 3: Performances of different rescaling methods during the training period were calculated as averages of the statistics given by the equations (16-18). The above values are obtained by averaging the results of two experiments by using different training and validation periods (i.e., the first and the last 75% of the data, respectively) and by averaging the results for four watersheds. Here, the olive green color represents copula, cyan represents ANN, dark green represents the remaining nonlinear methods, orange represents the linear methods that result in a correlation difference, and yellow represents the linear methods with no correlation change.

Figure 4: Performances of different rescaling methods during the validation period. The above values are obtained by averaging the results of two experiments by using different validation periods.

Figure 5: Added utility of the rescaling methods.

Table 1: Parameters of the ANNs used in this study where identity is abbreviated as Id.

<table>
<tr><th rowspan="2">ANN</th><th colspan="5">Function Type</th></tr>
<tr><th>Learning</th><th>Update</th><th>Output</th><th colspan="2">Activation</th></tr>
<tr><td rowspan="4">MLP</td><td rowspan="4">Back-propagation</td><td rowspan="4">Topological order</td><td rowspan="4">Id.</td><td>Input</td><td>Id.</td></tr>
<tr><td>Hidden</td><td>Id.</td></tr>
<tr><td>Context</td><td>---</td></tr>
<tr><td>Output</td><td>Id.</td></tr>
<tr><td rowspan="4">RBF</td><td rowspan="4">Back-propagation</td><td rowspan="4">Topological order</td><td rowspan="4">Id.</td><td>Input</td><td>Id.</td></tr>
<tr><td>Hidden</td><td>Gaussian</td></tr>
<tr><td>Context</td><td>---</td></tr>
<tr><td>Output</td><td>Id.+bias</td></tr>
<tr><td rowspan="4">ELMAN</td><td rowspan="4">Back-propagation</td><td rowspan="4">JE Order</td><td rowspan="4">Id.</td><td>Input</td><td>Id.</td></tr>
<tr><td>Hidden</td><td>Id.</td></tr>
<tr><td>Context</td><td>Id.</td></tr>
<tr><td>Output</td><td>Id.</td></tr>
<tr><td rowspan="4">JORDAN</td><td rowspan="4">Back-propagation</td><td rowspan="4">JE Order</td><td rowspan="4">Id.</td><td>Input</td><td>Id.</td></tr>
<tr><td>Hidden</td><td>Id.</td></tr>
<tr><td>Context</td><td>Id.</td></tr>
<tr><td>Output</td><td>Id.</td></tr>
</table>

Table 2: Defined sets of GEN.

Parameter	Rescaling method	
	GEN	GENL
Causality relationship	$X = f(Y)$	$X = f(Y, Y_{lag})$
Function set	"sin", "cos", "tan", "sqrt", "exp", "log", "+", "-", "*", "/", "^"	
Fitness function	$\frac{(Y^{*} - X)^{2}}{N}$	
Population size	100	
Stop condition	Time (40 minutes)	
where X, Y, Y_{lag}, and Y^{*} are the reference, unscaled, lagged form of unscaled, and rescaled soil moisture products respectively and N is the number of observations.		

Table 3: Copula functions (C_{YX}), parameters (P and θ), and characteristics used in this study. F_X and F_Y indicate CDF_X and CDF_Y, respectively.

Copula	$C_{YX}(F_Y, F_X)$	Tail Dependence	Family
Normal	$\int_{-\infty}^{\emptyset^{-1}(F_Y)} \int_{-\infty}^{\emptyset^{-1}(F_X)} \frac{\exp\left[-\frac{F_Y{}^2 - 2PF_YF_X + F_X{}^2}{2(1-P^2)}\right]}{2\pi(1-P^2)^{1/2}} dF_Y dF_X$	Strong in center	Elliptical
Clayton	$(F_Y{}^{-\theta} + F_X{}^{-\theta} - 1)^{-1/\theta}$	Strong in left tail	Archimidean
Gumbel	$\exp\{[(-\ln F_Y)^{\theta} + (-\ln F_X)^{\theta}]^{\frac{1}{\theta}}\}$	Strong in right tail	Archimidean
Frank	$\frac{-1}{\theta}\ln[1 + \frac{(e^{-\theta F_Y} - 1)(e^{-\theta F_X} - 1)}{e^{-\theta} - 1}]$	Strong in center	Archimidean
Joe	$1 - [(1-F_Y)^{\theta} + (1-F_X)^{\theta} - (1-F_Y)^{\theta}(1-F_X)^{\theta}]^{\frac{1}{\theta}}$	Strong in right tail	Archimidean

Table 4: Statistics of the training and validation datasets for two experiments using the first and the 25% of the data as validation, respectively, and the remaining data as training.

Exp.	Dataset	Loc.	Num. avail. points	Mean		Standard deviation		Lag1 autocorrelation of datasets	
				LPRM	In-situ	LPRM	In-situ	LPRM	In-situ
1	Training (last 75%)	LR	1193	0.311	0.105	0.099	0.046	0.784	0.819
		LW	1154	0.282	0.125	0.104	0.057	0.728	0.863
		WG	1239	0.18	0.046	0.074	0.022	0.801	0.889
		RC	1103	0.227	0.118	0.121	0.075	0.831	0.969
	Validation (first 25%)	LR	396	0.331	0.109	0.098	0.044	0.757	0.805
		LW	383	0.286	0.118	0.099	0.052	0.686	0.751
		WG	411	0.176	0.045	0.083	0.021	0.785	0.849
		RC	366	0.232	0.107	0.104	0.072	0.778	0.974
2	Training (first 75%)	LR	1192	0.316	0.106	0.1	0.046	0.757	0.826
		LW	1153	0.285	0.122	0.109	0.058	0.733	0.841
		WG	1238	0.18	0.044	0.077	0.022	0.789	0.879
		RC	1102	0.234	0.117	0.113	0.077	0.796	0.972
	Validation (last 25%)	LR	397	0.314	0.105	0.095	0.043	0.855	0.784
		LW	384	0.276	0.127	0.077	0.049	0.628	0.828
		WG	412	0.175	0.048	0.076	0.02	0.814	0.881
		RC	367	0.21	0.109	0.129	0.067	0.866	0.963

Table 5: Detailed performance of different rescaling methods during training and validation periods. Best statistics are shown in bold. The best performing method for training (Tr_best) and overall (best) are shown. The ones listed below are obtained by averaging the results of two experiments with different training periods.

Statistic		LOC	ORG	VAR	TCA	REG1	REG2	REGL	REGL2	MARS	MARSL	CDF	GEN	GENL	SVM	SVML	MLP	MLPL	RBF
Training	ρ	LR	0.567	0.567	0.567	0.567	0.580	0.576	0.586	0.600	**0.618**	0.577	0.595	0.608	0.602	0.617	0.579	0.589	0.580
		LW	0.514	0.514	0.514	0.514	0.536	0.531	0.551	0.602	0.630	0.566	0.570	0.635	0.604	**0.674**	0.552	0.569	0.536
		WG	0.696	0.696	0.696	0.696	0.708	0.712	0.733	0.734	0.772	0.721	0.730	0.761	0.730	**0.773**	0.709	0.742	0.708
		RC	0.698	0.698	0.698	0.698	0.709	0.732	0.734	0.727	0.759	0.687	0.727	0.753	0.727	0.765	0.721	0.750	0.709
	σ_ε	LR	0.083	0.042	0.061	0.038	0.037	0.037	0.037	0.036	**0.036**	0.042	0.037	0.036	0.036	0.036	0.037	0.037	0.041
		LW	0.091	0.056	0.071	0.049	0.048	0.048	0.048	0.046	0.044	0.053	0.047	0.044	0.046	**0.042**	0.048	0.047	0.055
		WG	0.062	0.017	0.019	0.016	0.016	0.016	0.015	0.015	**0.014**	0.017	0.015	0.014	0.015	0.014	0.016	0.015	0.019
		RC	0.084	0.059	0.079	0.054	0.053	0.052	0.052	0.052	0.049	0.060	0.052	0.050	0.052	0.049	0.053	0.050	0.059
	AMB	LR	0.208	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.001	0.000	0.031
		LW	0.160	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.006	0.004	0.001	0.000	0.018
		WG	0.135	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.000	0.000	0.006
		RC	0.113	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.003	0.003	0.000	0.001	0.018
Validation	ρ	LR	0.530	0.530	0.530	0.530	0.536	0.530	0.534	0.547	0.551	0.540	0.527	**0.613**	0.539	0.542	0.534	0.532	0.536
		LW	0.495	0.495	0.495	0.495	0.490	0.515	0.506	0.509	0.504	0.504	0.501	0.555	0.503	**0.586**	0.502	0.518	0.492
		WG	0.684	0.684	0.684	0.684	0.680	0.697	0.699	0.686	0.698	0.667	0.680	0.700	0.670	**0.691**	0.682	0.703	0.683
		RC	0.666	0.666	0.666	0.666	0.669	0.704	0.703	0.674	0.710	0.653	0.670	0.699	0.672	0.695	0.676	0.709	0.669
	σ_ε	LR	0.082	0.043	0.061	0.037	0.037	0.037	0.037	0.037	0.037	0.043	0.038	**0.034**	0.037	0.037	0.037	0.037	0.041
		LW	0.077	0.049	0.060	0.043	0.044	0.043	0.043	0.044	0.044	0.054	0.044	0.042	0.045	**0.041**	0.043	0.043	0.048
		WG	0.067	0.018	0.020	0.015	0.015	0.015	0.015	0.015	**0.015**	0.017	0.015	0.015	0.016	0.015	0.015	0.015	0.018
		RC	0.088	0.061	0.084	0.053	0.053	0.050	0.051	0.052	0.050	0.061	0.053	0.052	0.053	0.052	0.052	0.050	0.054
	AMB	LR	0.216	0.003	0.003	0.003	0.003	0.003	0.003	0.002	0.002	0.003	0.002	**0.001**	0.002	0.001	0.002	0.001	0.032
		LW	0.159	0.007	0.007	0.007	0.007	0.007	0.007	0.007	**0.006**	0.006	0.008	0.007	0.007	0.007	0.007	0.008	0.019
		WG	0.129	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.003	0.008
		RC	0.113	0.010	0.010	0.010	0.010	0.010	0.010	0.007	0.006	0.009	0.009	0.009	0.007	0.005	0.010	0.009	0.024

Table 5, continuation.

Statistic		LOC	RBFL	ELM.	ELM.L	JOR.	JOR.L	NOR.	NOR.L	CLA.	CLA.L	GUM.	GUM.L	FRA.	FRA.L	JOE	JOEL	Tr_best	BEST
Training	ρ	LR	0.585	0.595	0.601	0.591	0.597	0.585	0.591	0.581	0.558	0.566	0.546	0.594	0.562	0.517	0.517	**0.618**	**0.618**
		LW	0.558	0.583	0.592	0.556	0.585	0.561	0.570	0.560	0.519	0.550	0.523	0.581	0.531	0.520	0.511	**0.674**	**0.674**
		WG	0.740	0.747	0.748	0.726	0.737	0.722	0.746	0.631	0.655	0.721	0.727	0.725	0.727	0.720	0.727	**0.773**	**0.773**
		RC	0.741	**0.850**	0.844	0.829	0.826	0.708	0.741	0.725	0.759	0.697	0.746	0.709	0.751	0.673	0.734	**0.850**	**0.850**
	σ_ε	LR	0.038	0.037	0.037	0.037	0.037	0.037	0.037	0.037	0.038	0.038	0.039	0.037	0.038	0.039	0.039	**0.036**	**0.036**
		LW	0.048	0.046	0.046	0.048	0.047	0.047	0.047	0.047	0.049	0.048	0.050	0.047	0.049	0.049	0.050	**0.042**	**0.042**
		WG	0.019	0.015	0.015	0.016	0.015	0.015	0.015	0.017	0.017	0.016	0.015	0.015	0.015	0.016	0.015	**0.014**	**0.014**
		RC	0.054	**0.040**	0.041	0.043	0.043	0.054	0.052	0.052	0.050	0.055	0.051	0.055	0.051	0.056	0.052	**0.040**	**0.040**
	AMB	LR	0.024	0.003	0.005	0.006	0.007	0.000	0.000	0.017	0.023	0.001	0.004	0.001	0.004	0.000	0.002	**0.000**	**0.000**
		LW	0.023	0.008	0.007	0.006	0.009	0.000	0.002	0.025	0.029	0.000	0.013	0.000	0.036	0.001	0.005	**0.000**	**0.000**
		WG	0.038	0.002	0.001	0.004	0.003	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	**0.000**	**0.000**
		RC	0.015	0.004	0.001	0.003	0.003	0.001	0.001	0.000	0.001	0.007	0.004	0.001	0.003	0.017	0.013	**0.000**	**0.000**
Validation	ρ	LR	0.532	0.535	0.533	0.535	0.537	0.536	0.534	0.532	0.503	0.524	0.496	0.536	0.505	0.484	0.468	**0.551**	**0.613**
		LW	0.510	0.535	0.543	0.514	0.538	0.502	0.512	0.493	0.445	0.499	0.462	0.511	0.459	0.475	0.458	**0.586**	**0.586**
		WG	0.706	**0.713**	0.711	0.700	0.698	0.678	0.688	0.638	0.608	0.671	0.646	0.684	0.651	0.662	0.635	**0.691**	**0.713**
		RC	0.705	**0.801**	0.785	0.779	0.792	0.666	0.702	0.673	0.696	0.659	0.692	0.661	0.688	0.641	0.682	**0.801**	**0.801**
	σ_ε	LR	0.038	0.038	0.038	0.037	0.037	0.037	0.037	0.037	0.038	0.038	0.039	0.037	0.039	0.039	0.040	**0.037**	**0.034**
		LW	0.043	0.043	0.043	0.043	0.042	0.044	0.044	0.044	0.045	0.044	0.046	0.045	0.047	0.044	0.046	**0.041**	**0.041**
		WG	0.018	0.015	**0.015**	0.015	0.015	0.015	0.015	0.016	0.016	0.016	0.017	0.015	0.016	0.016	0.017	**0.015**	**0.015**
		RC	0.051	**0.043**	0.045	0.044	0.045	0.054	0.052	0.053	0.052	0.055	0.053	0.056	0.054	0.054	0.052	**0.043**	**0.043**
	AMB	LR	0.026	0.004	0.007	0.004	0.006	0.002	0.002	0.016	0.021	0.001	0.006	0.002	0.006	0.001	0.003	**0.003**	**0.001**
		LW	0.024	0.016	0.016	0.007	0.011	0.008	0.007	0.022	0.028	0.009	0.021	0.007	0.035	0.010	0.005	**0.007**	**0.005**
		WG	0.036	0.004	0.004	0.004	0.004	0.003	0.003	0.002	**0.002**	0.003	0.003	0.002	0.002	0.003	0.003	**0.003**	**0.002**
		RC	0.021	**0.005**	0.006	0.007	0.008	0.008	0.007	0.009	0.007	0.012	0.009	0.007	0.008	0.022	0.018	**0.010**	**0.005**

Table 6: Added utility of the selected methods compared to the REG1 validation statistics (Table 5) over four watersheds. Positive values indicate improvements, and negative values indicate degradation.

		ADDED UTILITY OF METHODS AGAINST REG1 STATISTICS								
Stat.	LOC	REGL2	MARSL	CDF	GENL	SVML	ELMAN	NORMALL	Tr_Best	Best
ρ	LR	0.004	0.021	0.010	0.083	0.012	0.005	0.004	0.021	0.083
	LW	0.011	0.008	0.009	0.059	0.090	0.040	0.016	0.090	0.090
	WG	0.016	0.015	-0.017	0.017	0.007	0.030	0.005	0.007	0.030
	RC	0.036	0.044	-0.013	0.033	0.029	0.135	0.036	0.135	0.135
σ_ε	LR	0.000	0.001	-0.005	0.003	0.000	-0.001	0.000	0.001	0.003
	LW	0.000	-0.001	-0.010	0.002	0.002	0.001	-0.001	0.002	0.002
	WG	0.000	0.000	-0.002	0.000	0.000	0.000	0.000	0.000	0.001
	RC	0.002	0.003	-0.008	0.001	0.000	0.010	0.001	0.010	0.010
AMB	LR	0.000	0.001	0.000	0.002	0.001	-0.001	0.001	0.000	0.002
	LW	0.000	0.000	0.000	-0.001	-0.001	-0.009	0.000	0.000	0.002
	WG	0.000	0.000	0.000	0.000	0.000	-0.001	-0.001	0.000	0.000
	RC	0.000	0.003	0.000	0.000	0.004	0.004	0.002	0.000	0.004

Table 7: Added utility of the lagged observations calculated as the statistic of the lagged type of a method minus a non-lagged type (e.g., 0.020 ρ value over LW is obtained as $\rho_{REGL} - \rho_{\mathrm{REG1}}$; and 0.086 ρ value over LR is obtained as $\rho_{GENL} - \rho_{\mathrm{GEN}}$ using values given in Table 5).

		ADDED UTILITY OF LAGGED OBSERVATIONS					
Stat.	LOC	REG1	MARS	GEN	SVM	ELMAN	NORMAL
ρ	LR	0.000	0.004	0.086	0.003	-0.002	-0.002
	LW	0.020	-0.005	0.053	0.083	0.008	0.010
	WG	0.013	0.013	0.020	0.021	-0.002	0.010
	RC	0.038	0.035	0.029	0.023	-0.016	0.036
σ_ε	LR	0.000	0.000	0.003	0.000	0.000	0.000
	LW	0.001	0.000	0.002	0.004	0.000	0.000
	WG	0.000	0.000	0.000	0.001	0.000	0.000
	RC	0.002	0.002	0.002	0.001	-0.002	0.002
AMB	LR	0.000	0.000	0.001	0.000	-0.003	0.000
	LW	0.000	0.000	0.000	0.000	0.000	0.001
	WG	0.000	0.000	0.000	0.000	0.000	0.000
	RC	0.000	0.001	-0.001	0.001	-0.001	0.000

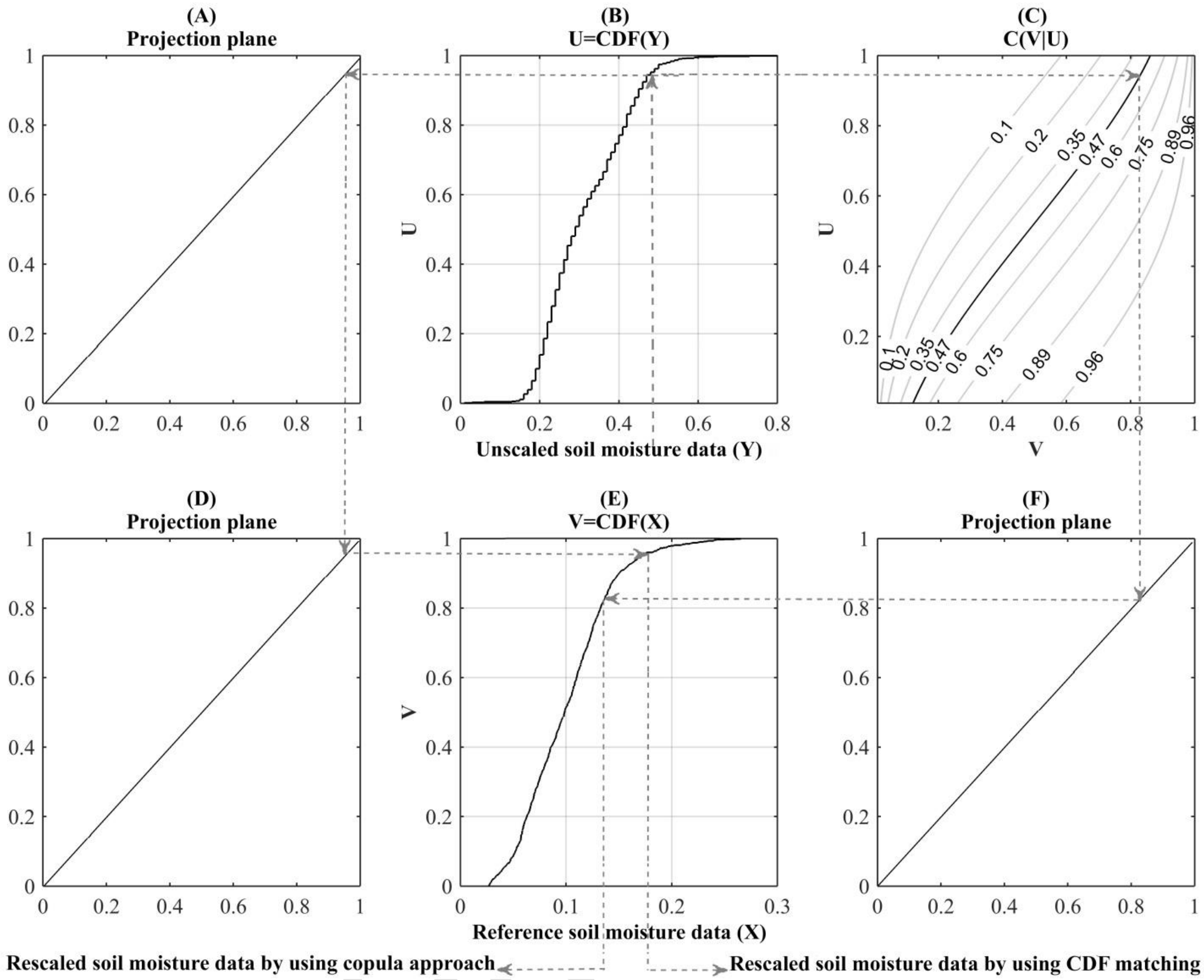

Figure 1: Schematic representations of the CDF and Copula based rescaling methods. The paths in the BADE and BCFE panels represent the CDF and Copula methods, respectively. $C_{X|Y} = 0.47$ is plotted with darker color in panel C to represent the best performing projection line of the Copula.

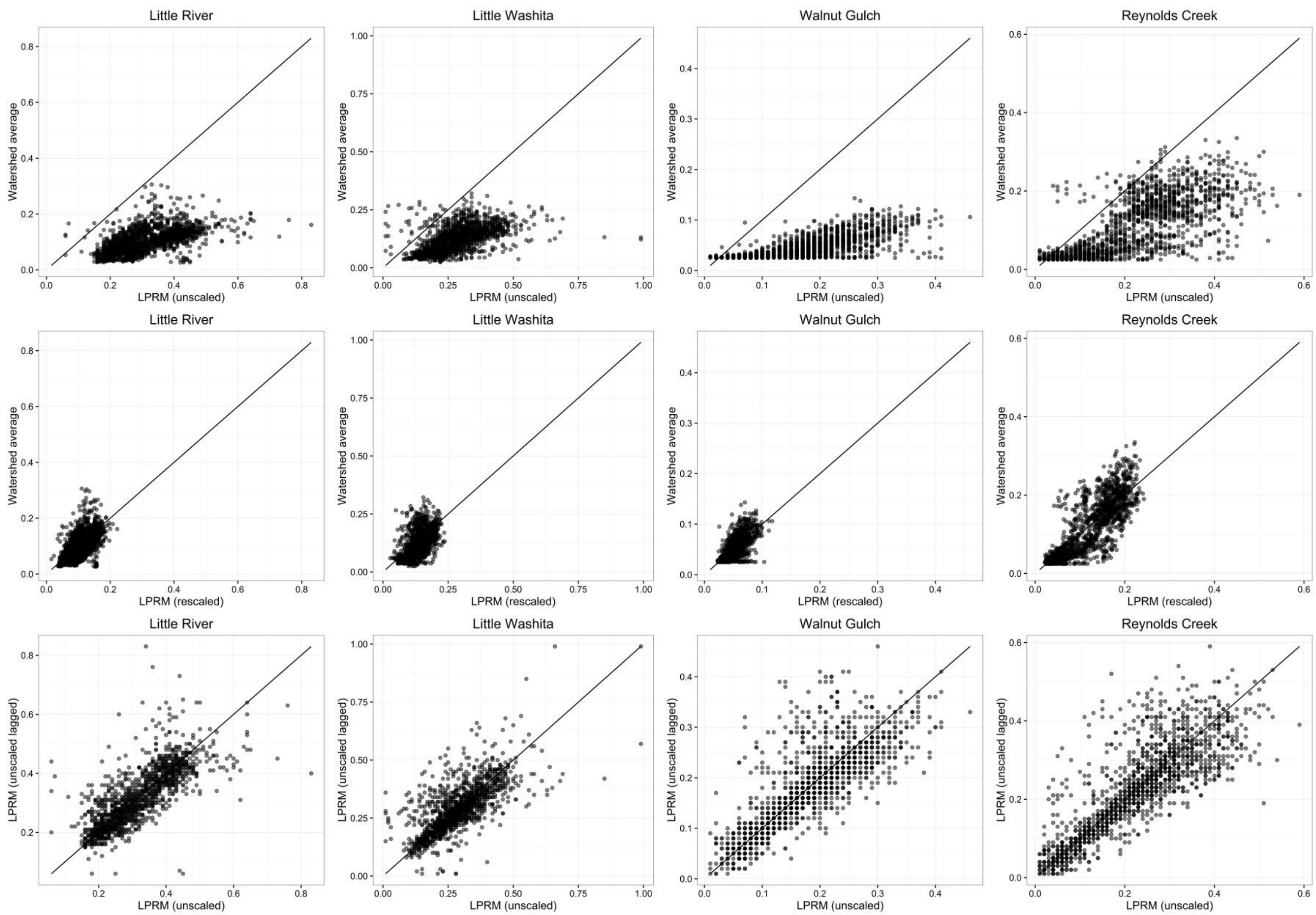

Figure 2: Scatter plot of the Watershed average and LPRM soil moisture data over four watersheds. Original (unscaled) and rescaled data are given in the upper and middle rows, respectively; lagged unscaled LPRM vs unscaled LPRM are given in the lower row.

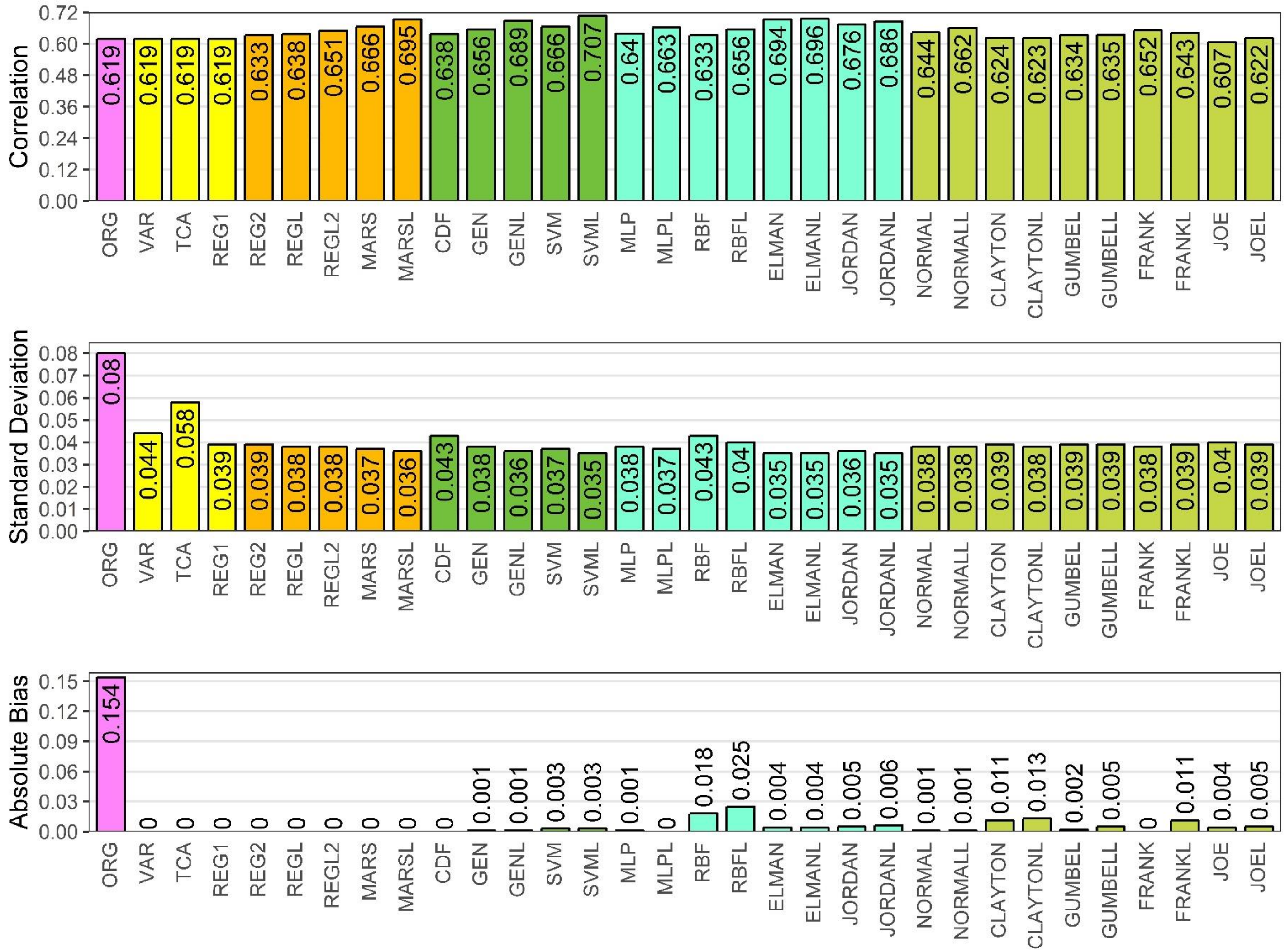

Figure 3: Performances of different rescaling methods during the training period were calculated as averages of the statistics given by the equations (16-18). The above values are obtained by averaging the results of two experiments by using different training and

validation periods (i.e., the first and the last 75% of the data, respectively) and by averaging the results for four watersheds. Here, the olive green color represents copula, cyan represents ANN, dark green represents the remaining nonlinear methods, orange represents the linear methods that result in a correlation difference, and yellow represents the linear methods with no correlation change.

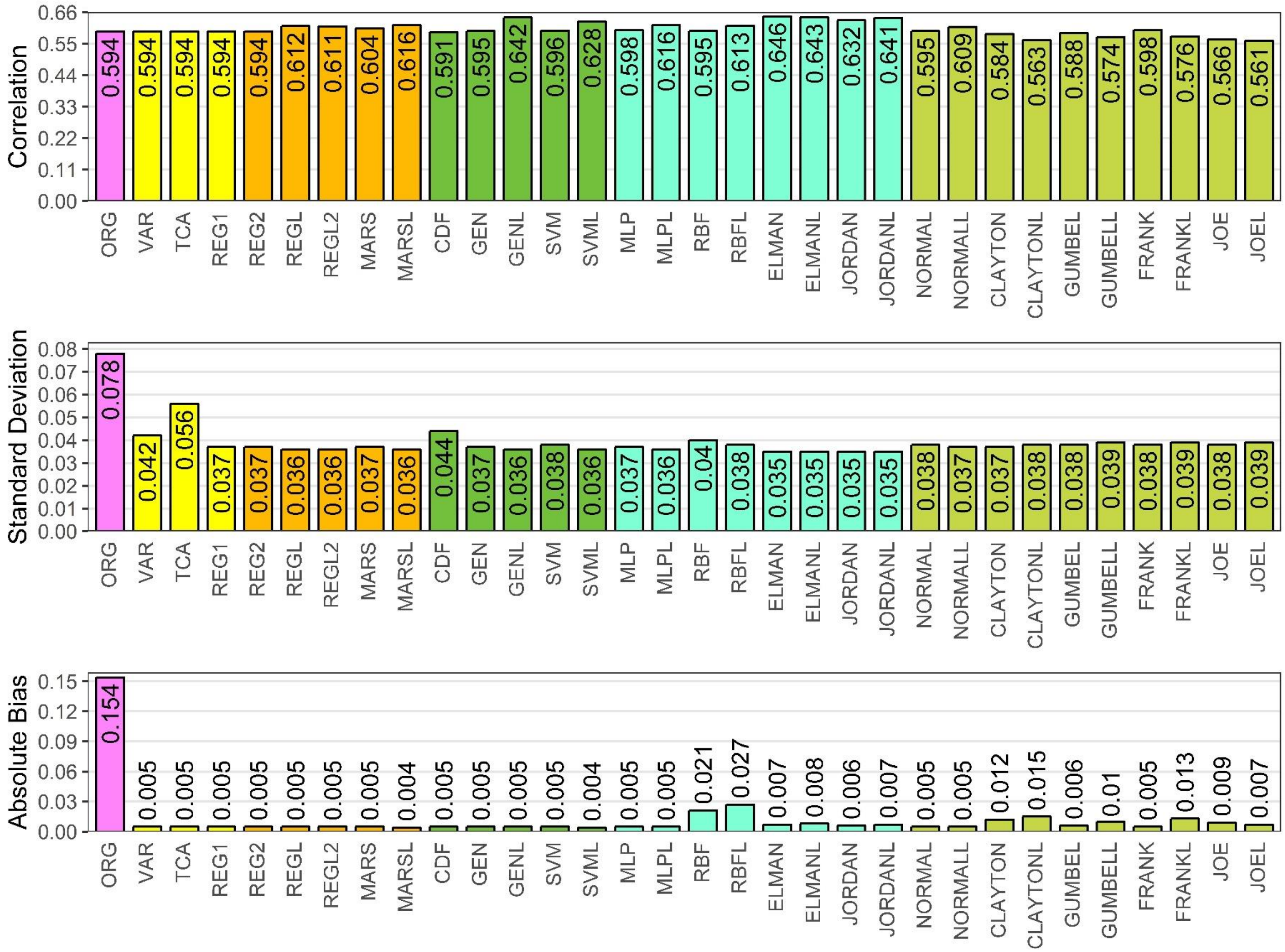

Figure 4: Performances of different rescaling methods during the validation period. The above values are obtained by averaging the results of two experiments by using different validation periods.

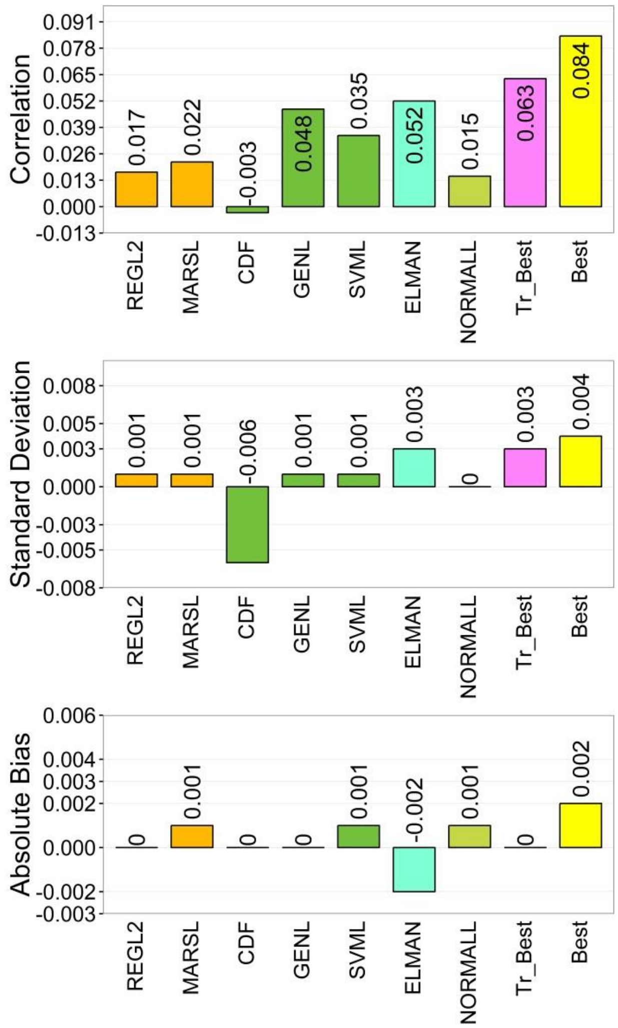

Figure 5: Added utility of the rescaling methods.