Remote Sensing of Environment 196 (2017) 224–237

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

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

M.H. Afshar, M.T. Yilmaz *

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

ARTICLE INFO

*Article history:*
Received 1 May 2016
Received in revised form 21 April 2017
Accepted 14 May 2017
Available online xxxx

*Keywords:*
Soil moisture
Rescaling
Linear
Nonlinear
Remote sensing

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.

© 2017 Published by Elsevier Inc.

## 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 and 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 and 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;

* Corresponding author.
*E-mail address:* tuyilmaz@metu.edu.tr (M.T. Yilmaz).

http://dx.doi.org/10.1016/j.rse.2017.05.017
0034-4257/© 2017 Published by Elsevier Inc.

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 and 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 and 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 and 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 and Zhan, 2007; Crow, 2007), TCA (Yilmaz and Crow, 2013), quadratic polynomials (Zwieback et al., 2016), copula (Leroux et al., 2014), and Wavelets (Su and 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-Fernandez et al., 2015) and SVM methods (Cortes and 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 and 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.

Download English Version:

https://daneshyari.com/en/article/5754970

Download Persian Version:

https://daneshyari.com/article/5754970

Daneshyari.com