

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Using hydro-climatic and edaphic similarity to enhance soil moisture prediction

E. J. Coopersmith¹, B. S. Minsker¹, and M. Sivapalan^{1,2}

¹Department of Civil & Environmental Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

²Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

Received: 17 January 2014 – Accepted: 9 February 2014 – Published: 25 February 2014

Correspondence to: E. J. Coopersmith (ecooper2@gmail.de)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Abstract

5 Estimating soil moisture typically involves calibrating models to sparse networks of in situ sensors, which introduces considerable error in locations where sensors are not available. We address this issue by calibrating parameters of a parsimonious soil moisture model, which requires only antecedent precipitation information, at gauged locations and then extrapolating these values to ungauged locations via a hydro-climatic classification system. Fifteen sites within the soil climate analysis network (SCAN) containing multi-year time series data for precipitation and soil moisture are used to calibrate the model. By calibrating at one of these fifteen sites and validating at another, we observe that the best results are obtained where calibration and validation occur within the same hydro-climatic class. Additionally, soil texture data are tested for their importance in improving predictions between calibration and validation sites. Results have the largest errors when calibration/validation pairs differ hydro-climatically and edaphically, improve when one of these two characteristics are aligned, and are strongest when the calibration and validation sites are hydro-climatically and edaphically similar. These findings indicate considerable promise for improving soil moisture estimation in ungauged locations by considering these similarities.

1 Introduction

20 Soil moisture estimates are needed routinely for many practical applications, such as irrigation scheduling and operation of farm machinery. They are typically produced either through remote sensing or sparse networks of in situ sensors. Although recent remote sensing studies have confirmed that such measurements approximate in situ sensor networks (Jackson et al., 2012), satellite-based sensors provide measurements at a spatial resolution of several kilometers – too large for daily agricultural decision making. On the other hand, in situ sensor networks produce values that are difficult to generalize to locations with no proximal sensors. Under these circumstances, dynamic

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

data), but ultimately requires proximal sensors to provide the needed soil moisture estimates (Chico-Santamaria et al., 2009).

Pan et al. (2003, 2012) addressed the shortcomings of the existing modeling approaches reviewed above by developing what they called a “diagnostic soil moisture equation” (i.e., model) in the form of a partial differential equation representing the lumped water balance of a vertical soil column, and representing the soil moisture at any moment in time as a function of the sum of a temporally decaying sequence of observed past rainfall events. The model has the advantage that initial soil moisture conditions are not required (only antecedent precipitation data), nor must the model be recalibrated periodically. However, this approach does require a soil moisture sensor at the relevant location for initial calibration of the model’s parameters. This method has the disadvantage that the presence of soil heterogeneity could necessitate a large number of sensors to account for the spatial variation of soil moisture (Pan and Peters-Lidard, 2008). Furthermore, decision support often requires estimation at locations lacking sensors.

The aim of this paper is to present and test an approach that can help overcome the problems associated with the Pan et al. soil moisture estimation model. The proposed solution involves calibrating the Pan et al. diagnostic soil moisture equation (model) at gauged sites and then extrapolating the calibrated model to ungauged sites by invoking similarity. Similarity here is defined on the basis of hydro-climatic characteristics, using a classification system developed by Coopersmith et al. (2012), as well as edaphic (soil) properties. The proposed new scheme maintains the advantage of Pan et al.’s parsimonious soil moisture model in that it does not require specification of initial soil moisture condition, and also there is no need to recalibrate periodically. The model’s simplicity also permits implementation of the model in a manner that can easily be refit with new parameters, where necessary. Section 2 provides more details on the approach.

To calibrate and validate the model, we use data from the US Department of Agriculture’s (USDA) Soil Climate Analysis Network (SCAN), described in Schaefer et

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

al. (2007). This national array of soil moisture sensors (with co-located precipitation sensors) delivers hourly data at a variety of publically-accessible sites throughout the United States. Fifteen sensor locations with numerous years of high-quality, minimally-interrupted data were selected for further analysis. These sites display considerable hydrologic diversity, which aids in demonstrating that the nationwide application of the proposed soil moisture model using precipitation data represents a feasible goal. Results of the analysis are given in Sect. 3, followed by discussion in Sect. 4 to suggest further improvements and conclusions are presented in Sect. 5.

2 Methodology

The proposed modeling approach involves four steps, summarized in Fig. 1 and described in more detail in the sections below. First, the diagnostic soil moisture model of Pan et al. is calibrated at locations with ample data. Given that the focus of this study is on soil moisture estimation for agriculture, we only consider prediction during the growing season, which is appropriate given that the model does not address snow melt processes. Second, the predictions at these locations are improved using machine learning techniques for error correction. Third, the classification system proposed by Coopersmith et al. (2012) is used to generalize the parameters calibrated at each location, enabling its application at other sites characterized by the same hydro-climatic class. Fourth, sites are examined for edaphic (soil property) similarity in addition to hydro-climates. The results of these four steps are then examined to identify which approach to regionalization performs best.

2.1 Step 1: calibration using a two-layer genetic algorithm

Unlike the original diagnostic soil moisture calibrations, the ultimate objective of this work is to enable agricultural decision support in near real time. To this end, the daily model from Pan et al. (2012) is first modified to yield an hourly model within the same

framework. Genetic algorithms are then deployed to calibrate the model, enabling more efficient exploration of the parameter search space than the traditional Monte Carlo search, which was the approach taken by Pan et al. (2012).

Genetic algorithms (GAs), a subset of evolutionary algorithms, were originally developed by Barricelli (1963) and have become increasingly common in environmental and water resources applications, including the calibration of hydrologic model parameters (e.g., Cheng et al., 2006; Singh and Minsker, 2008; Zhang et al., 2009).

In this work, a simple genetic algorithm uses the operations of selection, crossover, and mutation (for reference, see Goldberg, 1989) to search for parameters that minimize prediction errors from the diagnostic soil moisture equation (Pan et al., 2012):

$$\theta_{\text{est}} = \theta_{\text{re}} + (\phi_e - \theta_{\text{re}})(1 - e^{-c_4\beta}). \quad (1)$$

Here θ_{est} represents the best estimate of soil moisture during a given hour. θ_{re} denotes residual soil moisture, the minimum quantity of moisture that is present regardless of the length of time without precipitation. ϕ_e , the soil's porosity, signifies the maximum possible soil moisture value, at which point the soil becomes saturated and cannot increase its moisture content. Finally, c_4 is a parameter related to conductivity and drainage properties, essentially defining the rate at which soil can dry. If c_4 assumes a value of zero, the soil is permanently at its residual soil moisture value, θ_{re} – a soil that dries infinitely rapidly. Conversely, as c_4 becomes large, the soil will permanently assume the value of its porosity, ϕ_e – a soil that dries infinitely slowly. The β term in Eq. (1) is calculated in Eq. (2) below:

$$\beta = \sum_{i=2}^{i=n-1} \left[\frac{P_i}{\eta_i} \left(1 - e^{-\frac{\eta_i}{z}}\right) e^{-\sum_{j=1}^{j=i-1} \left(\frac{\eta_j}{z}\right)} \right] + \frac{P_1}{\eta_1} \left(1 - e^{-\frac{\eta_1}{z}}\right). \quad (2)$$

Here, P_i denotes the quantity of rainfall during hour i (day in the original presentation in Pan et al., 2003). The soil depth at which an estimation occurs is given by z . This convolution summation has a temporal window of size n for considering past precipitation.

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

| | |
|--------------------------|--------------|
| Title Page | |
| Abstract | Introduction |
| Conclusions | References |
| Tables | Figures |
| ⏪ | ⏩ |
| ◀ | ▶ |
| Back | Close |
| Full Screen / Esc | |
| Printer-friendly Version | |
| Interactive Discussion | |



Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper

To wit, yesterday's rainfall affects today's soil moisture, last week's rainfall is relevant, but less so, and rainfall from ten years ago is probably not relevant.

To choose the appropriate value for n , the value of β is calculated at each hour throughout the dataset – setting n to a very large value (2000 h, denoted by M) initially. Next this “beta series” (where $n = M$) is correlated with a separate beta series, calculated where $n \ll M$. If the correlation coefficient between these two time series approaches unity, then the smaller value of n is selected. Otherwise, n is increased incrementally until the correlation between the $n \ll M$ beta series and the $n = M$ beta series approaches unity.

Finally, the η_i terms signify the estimated potential evaporation/drainage loss at hour i of the calendar year. As this algorithm does not presume any more detailed knowledge of potential evaporation/drainage behaviors, this “eta series” is modeled as a sinusoid (Pan et al., 2012) with period 8760 (the number of hours in a year). The eta (η) series is required to calculate the beta (β) series (Eq. 2), which is required to use the diagnostic soil moisture equation (Eq. 1). Thus, before any other parameters are chosen, a generalized sinusoidal form of η is estimated as given in Eq. (3):

$$\eta = \alpha \sin(i - \delta) + \gamma. \quad (3)$$

Here, α represents the sinusoid's amplitude, γ denotes the vertical shift, and δ signifies the necessary phase shift. These three parameters are fitted via the genetic algorithm such that the correlation between the beta series (using the eta series implied by α , γ , and δ) and the observed soil moisture series (θ_{obs}) is maximized. Once values for the eta series are established, the remaining three parameters of Eq. (1) (θ_{re} , ϕ_{e} , and c_4) are then fitted by a second application of the genetic algorithm, this time minimizing the sum of squared errors between the estimated soil moisture series (θ_{est}) and the observed values (θ_{obs}).

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

2.2 Step 2: error correction using the k-Nearest Neighbors machine learning algorithm

After the parameters of the diagnostic soil moisture equation (Eq. 1) have been calibrated, the hourly precipitation time series is used to generate a soil moisture time series during the growing season months of interest. Discrepancies between the observed soil moisture values (θ_{obs}) and the estimated values (θ_{est}) are computed as shown in Eq. (4):

$$\theta_{\text{obs}} = \theta_{\text{est}} + \varepsilon \quad (4)$$

where ε represents the error associated with any hour's soil moisture estimate.

To correct biases in these errors, the k-Nearest Neighbor algorithm (Fix and Hodges, 1951) is employed to predict ε using the characteristics from the training data. More specifically, the data are searched for the most similar matches in terms of time of day, day of year, θ_{est} , $\beta(n)$, and $\beta(M) - \beta(n)$. For example, if the model returns a prediction of $\theta_{\text{est}} = 0.35$ at 14:00 LT during July when rainfall has been heavy recently but drier over a longer period, KNN will search the training set for other estimates near 0.35 made on mid-summer afternoons where a similar recent rainfall pattern has been observed. Next, the algorithm averages the value of the error, ε , associated with those types of conditions, producing an estimated error, ε_{est} . Each validation estimate is then adjusted to be $\theta_{\text{est}} + \varepsilon_{\text{est}}$. This technique allows consistent model biases, such as underestimating wetter days and overestimating drier days, to be corrected.

This error correction model also accounts for diurnal soil moisture variations that were not considered in developing the diagnostic soil equation, which was designed to deliver daily soil moisture estimates. Consider a soil moisture estimate at 16:00 LT, after soil has had a full day of sunlight (theoretically) to dry. As the diagnostic soil moisture equation only considers drainage and evapotranspiration losses on a daily basis, θ_{est} will be larger than θ_{obs} . Yet, because this type of mistake presumably occurred frequently throughout the training data, the algorithm will locate other 16:00 LT estimates,

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



each of which will be biased in the same direction, and our final soil moisture estimates will take this bias into account, improving the results as shown subsequently.

To assess the performance of the soil moisture models with and without machine learning, an R^2 value as defined in Eq. (5) is used, as this value represents the proportion of variance in soil moisture explained by the developed model.

$$R^2 = 1 - \frac{SSR}{SST} \quad (5)$$

where SSR denotes the sum of squared residuals and the SST term signifies the total sum of squares, i.e., the sample's variance.

2.3 Step 3: estimation by hydro-climatic similarity

This step tests the hypothesis that the classification system by Coopersmith et al. (2012) can be used to generalize the calibrated parameters for the diagnostic soil moisture equation using hydro-climatic similarity. If two locations are assigned the same hydro-climatic classification, then the calibrated parameters from one SCAN sensor within that class will be assumed to perform well at another.

This hypothesis was tested at fifteen SCAN sensors for which soil moisture and precipitation data are available hourly for a period of several years. These sensors are located in diverse geographic locations and hydro-climatic classes in Iowa, North Carolina, Pennsylvania, New Mexico, Arkansas, Georgia, Virginia, South Carolina, Nebraska, Colorado, and Wyoming. The data at each of these locations were divided into training/validation sets and parameters were calibrated using training data only. Next, these parameters were employed on the validation sets at the locations for which they were calibrated. The subsequent R^2 values (proportion of variance in soil moisture explained by the machine-learning-enhanced diagnostic soil moisture equation, see Steel and Torrie, 1960, for reference) defined a baseline level of performance for that site.

The process of cross-validation is detailed below:

1. Consider two sites, x and y , chosen from the fifteen available calibrated locations.

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

2. Estimate the soil moisture values in the validation dataset of site y , using the parameters calibrated from the training dataset at site x .
3. Record the difference between the R^2 baseline value at site y (obtained using parameters calibrated at site y) and the performance obtained at site y using parameters calibrated at site x .
4. Repeat steps 1–3 for all 210 possible (x, y) pairs where $x \neq y$.

Note: (x, y) and (y, x) are not equivalent. One signifies the performance of parameters calibrated at site x making predictions at site y , the other signifies the performance of parameters calibrated at site y making predictions at site x .

At this point, three types of (x, y) pairs emerge. If the hypothesis is correct, then the first type, when x and y fall within the same hydro-climatic class, should display limited losses in predictive power. The second type, when x and y fall within a “similar” hydro-climatic class (two classes differing by a single division of the classification tree developed in Coopersmith et al., 2012) should display greater losses of predictive power. Finally, the third type, when x and y fall in two unrelated classes, should display the largest loss of predictive power.

2.4 Step 4: estimation by hydro-climatic and edaphic similarity

The final step extends the hypothesis proposed in Step 3 by evaluating the impacts of soil texture and type on soil moisture predictive power. The fifteen sites from the SCAN network are examined based upon the soil textural information available from the Pedon soil reports that SCAN provides, as well as data from NRCS’s soil survey database¹.

This information allows sites already deemed hydro-climatically similar to be further sub-divided into sites that are and are not edaphically similar. Analogous to the previous section, we consider pairs of sites, x and y , where parameters are calibrated at

¹<http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx>

site x and validated at site y . In this case, four groups can be defined – the first, where x and y are hydro-climatically and edaphically similar, the second, where x and y are hydro-climatically similar, but differ edaphically, the third, where x and y are edaphically similar, but differ hydro-climatically, and finally, where x and y are hydro-climatically and edaphically dissimilar.

3 Results

This section begins by presenting the results of the machine learning approach used in error correction during the initial calibration step (Sect. 3.1). Next, Sect. 3.2 presents results for the hydro-climatic similarity analysis, illustrating the performance of calibration/validation pairs within the same class and without. Finally, Sect. 3.3 shows how the predictive power improves when both hydro-climatic and edaphic similarity are considered.

3.1 Testing the value of machine learning error correction for soil moisture prediction using the diagnostic soil moisture equation

Figure 2 shows the performance of the calibrated parameters for the 15 SCAN sites using only the diagnostic soil moisture equation (Step 1 of the methodology) along with the subsequent improvement in performance following machine learning error correction (Step 2). In each case, the six parameters required for the implementation of the diagnostic soil moisture equation are calibrated using training data from before 2010. Sensors with hourly precipitation and soil moisture time series data between 2004 and 2009 (inclusive) provide four to six years of training data (some sites are missing one or two years of data). Only days of the year where snow cover is unlikely are used to train the algorithm (from the 100th to 300th day of the year in all locations, for consistency). Validation data consist of days 100–300 for 2010 and 2011.

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

The results illustrate that in all fifteen test cases, performance within the validation sample is improved by machine learning modeling of residuals from the training set, in some cases, as much as 26.9 % of the unexplained variance (site 2091) in soil moisture is corrected from by this technique. The average results (far right column, Fig. 2) illustrate that the diagnostic soil moisture equation explains just 69.2 % of the variance in soil moisture ($\rho = 0.83$) before machine learning corrections occur, but explains 77.5 % of the variance in soil moisture ($\rho = 0.88$) thereafter.

To explore these findings in more detail, three of the 15 SCAN sites, chosen to represent different hydro-climatic locations – New Mexico (#2015, desert), Iowa (#2068, plains), and Georgia (#2013, hills/woods) are examined to illustrate how improvements from adding machine learning error models to the diagnostic soil moisture equation differ across sites. These three sites represent three distinct hydro-climatic classes, with significant differences in seasonality of precipitation, aridity, timing of maximum precipitation, and timing of maximum runoff. Using error correction models for prediction at these sites increased R^2 values by an average of 8.2 %, which is similar to the 8.3 % improvement in R^2 averaged across all fifteen sites. Thus, these three locations are representative in terms of both hydro-climatic diversity and their responsiveness to machine learning.

The base soil moisture model results from applying Step 1 at the three sites are displayed in Figs. 3–5. These predictions are shown with the results produced by deploying the machine learning algorithm (KNN) in Step 2 to remove bias and correct errors. In each image, the blue line represents the observed soil moisture readings, the red line represents the estimates generated by the diagnostic soil moisture equation, and the green line represents those predictions after the machine learning algorithm has removed biases and corrected errors. Soil moisture values (y-axis) are measured in percentage terms (0–100).

In Fig. 3, the diagnostic soil moisture equation is able to trace the general trend of the soil moisture time series ($\rho = 0.860$). However, during the middle of the time series, in which the observed soil moisture values fall below 5 %, the benefits of machine learning

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

error correction are most noteworthy. There are other hours scattered throughout the dataset where the green line (ML prediction) follows the blue line (observed values) much more closely than the red line (diagnostic soil moisture equation). The green line ($\rho = 0.917$) not only improves upon the correlation value of Pearson's Rho (the square root of the R^2 value in Eq. 5), but also displays marked improvement for those cases in which the diagnostic soil moisture equation produces significant errors.

During the validation period, considerable flooding occurred in Iowa (Fig. 4). Flood events of this nature were not experienced during calibration, and the porosity parameter, ϕ_e (Eq. 1) was set at 38.4 %. While this was appropriate for the training data (during which soil moisture did not exceed this level), extreme flooding events caused moisture levels for which the diagnostic soil moisture equation was not properly calibrated. The machine learning driven error correction improves the diagnostic soil moisture equation ($\rho = 0.846$) significantly ($\rho = 0.915$), but without flooding events in the training set, is unable to correct for the errors due to exceedingly wet (flooded) soil. Underestimations due to floods, although detrimental in terms of numerical errors, are not necessarily a problem for decision support of agricultural or construction activities, for example. If a model warns that a site is very wet and in reality, it is even wetter than predicted, the user has still been given adequate warning not to attempt activity at that site.

In Fig. 5, a soil moisture series from Georgia is modeled by the diagnostic soil moisture equation. Even before adding any error correction, the equation performs well ($\rho = 0.936$) and the machine learning approach yields a smaller improvement ($\rho = 0.941$). It is worth noting that machine learning does not damage an already excellent performance, offering slight improvements when possible and essentially no correction when training data suggest the model has already performed adequately.

In addition to generalizing the parameters calibrated in the diagnostic soil moisture equation, the error correction approach allows for systematic biases to be removed by searching training data for similar conditions and then predicting the types of mistakes most likely to occur. Figure 6, by zooming in upon a 30 day period from Fig. 2, illustrates how machine learning reduces errors by introducing a diurnal cycle into a model that

previously lacked one. The remaining bias is likely explained by a slightly wetter training dataset as compared with the validation data.

By addressing such systematic biases, machine learning enables model performance to improve with each successive growing season as the training dataset expands. For instance, although the fields in Iowa endured flooding during the validation period and subsequently made errors, such errors would eventually populate the training data. The next time such flooding occurs, the model is likely to recognize the occurrence of those same conditions and adjust the diagnostic soil moisture equation's predictions accordingly. In this vein, model performance is likely to improve over time, especially with the models already showing reasonable accuracy using only a few years of training data.

3.2 Cross-validation results for hydro-climatic similarity: qualitative findings and significance testing

To test the hypothesis that models calibrated in one location can be used in a hydro-climatically similar location, cross-validation was used as described in Step 3 of Sect. 2. The fifteen SCAN sites yield $15^2 = 225$ possible (x, y) pairs. Fifteen of these 225 pairs occur when $x = y$, establishing the baseline level of performance for a given site (validation performed using the parameters calibrated at that same location). Of the 210 remaining (x, y) pairs, 120 of them consist of paired catchments in which x and y are located in unrelated classes, 60 consist of paired catchments in which x and y are located in a "similar" class (different by a single split within the classification tree), and 30 consist of paired catchments in which x and y fall within the same hydroclimatic class (but x and y do not represent the same catchment). Figure 7 presents box plots illustrating the results of these three sets of pairs and Table 1 presents the quantitative results.

These findings show that calibrating the model at one location and applying those parameters elsewhere within the same class (green) is preferable to applying those parameters in a similar, but not identical class (yellow) and vastly superior to applying

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



those parameters in an unrelated class (red). The differences between any two clusters (same-class, similar-class, unrelated class) are all significant at the $\alpha = 0.01$ level ($p < 0.001$ in all cases) as calculated by a two-sample, heteroscedastic t test (Welch, 1947).

3.3 Impact of soils: cross-validation results for edaphic and hydro-climatic similarity

To isolate the impacts of soil types (edaphic similarity) on soil moisture prediction, groups of sensor locations among the 15 SCAN sites that are hydro-climatically similar were analyzed, shown in Fig. 8. The soil textural data for each of these fifteen sensors are plotted on a soil texture pyramid diagram in Fig. 9. These data were obtained from either Pedon Soil Reports available through the SCAN network (which provide precise percentages of clay, silt, and sand), or, where this information was unavailable, from soil information in the national soil Web database².

Of the thirteen sensors from the four hydro-climatic classes with multiple SCAN sensors (light green, blue, dark green, and brown in Figs. 8 and 9), 30 (x, y) pairs exist where the model can be calibrated at site x and its parameters applied at site y . Note that (x, y) is not equivalent to (y, x) as the sites for calibration and validation are reversed. Of these 30 pairs, 20 pairs are edaphically similar as well. However, 10 of them include a pair of points where the soil types or terrain types are notably misaligned (for example, light green dots in Fig. 10 where two of the three sensors are in silty clay loam and the third is in sandy loam – a notably different soil). A similar analysis to the one presented in Fig. 7 and Table 1 has been reproduced, comparing the loss in predictive power (R^2) for the 20 pairs with similar hydro-climates and soils against the loss for the 10 pairs in which either the soil texture (Fig. 9) or type do not align. The average loss of 1.0% for the 20 very similar pairs is a much smaller decline than the 8.0% average decline observed for the 10 pairs for which soil/terrain information suggests dissimilarity. These results are significant with a p value of approximately 0.02. Additionally,

²<http://websoilsurvey.nrcs.usda.gov/app/WebSoilSurvey.aspx>

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the upper-most two green dots in Fig. 7, where calibrated parameters at one location perform poorly at another of similar hydro-climatic class, fall within these 10 cases.

These observations show the importance of soil information, or edaphic similarity. While pairs of calibration/validation locations with similar hydro-climates, but dissimilar soils, show a decline in performance as compared with pairs of locations where both are similar, so too do locations with similar soils, but dissimilar hydro-climates. The shaded circles in Fig. 10 illustrate groups of sensors that are quite similar in terms of soil textures. However, despite their soil similarities, differences in hydro-climates hinder cross-application, showing a decline in performance of 10.9 % for all (x, y) pairs within the shaded regions of Fig. 10 for which x and y are not from the same hydro-climatic class.

As summarized in Fig. 11, these results suggest that in cases where both soil type and hydro-climate align, very little performance is lost when parameters are re-applied (1.0 %), moderate declines in performance are observed when one of these two factors are aligned (8.0 % if hydro-climates align and soil types do not; 10.9 % if soil types align, but hydro-climates do not), and large declines in performance appear when neither align (20.5 %). Clearly both types of attributes are important and should be considered in future modeling work.

4 Discussion: future work to improve predictions

This section discusses other approaches that could be used in the future to improve and broaden the applicability of the methods developed in this work. First, we will consider micro-topographic effects on soil moisture, as local peaks and valleys can cause soils to dry more or less rapidly. Second, we will discuss a conceptual omission within the diagnostic soil moisture equation – infiltration excess. Finally, we will discuss the role of future satellite data on soil moisture modeling.

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

calculation for any particular soil type. While this approach would require the fitting of additional parameters, it is likely that predictions would be improved. These additional parameters could also be considered in assessing cross-site edaphic similarity using the methods described above, although they may be highly correlated with existing parameters such as porosity, residual soil moisture, and drainage.

4.3 NASA's Soil Moisture Active Passive (SMAP) mission

With NASA satellite data for soil moisture available at the 36, 9, and 3 km scales throughout the United States, and with the SMAP satellite scheduled to launch during 2014 (O'Neill et al., 2011), the models developed in this work will have ample measurements against which to test and improve their results, and can be used to help check the accuracy of satellite measurements. Future research in topographic disaggregation as proposed above, using LiDAR data, could be used to improve satellite soil moisture estimates by accounting for smaller-scale topography.

5 Conclusions

This work has demonstrated the feasibility of estimating soil moisture at locations where soil moisture sensors are unavailable for calibration, provided they fall within hydro-climatically and edaphically similar areas to gauged locations. By calibrating the diagnostic soil moisture equation via a two-part genetic algorithm, improving its performance via a machine learning algorithm for error correction, then validating that algorithm at the same location in subsequent years, a baseline level of predictive performance is established at fifteen locations. Next, these results are cross-validated – deploying parameters calibrated at a given site at sites of similar and different hydro-climatic classes, demonstrating that parameters can be re-applied elsewhere within the same class, but not without. Finally, by incorporating edaphic information, we observe the strongest cross-validation results when hydro-climatic and edaphic characteristics

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

align. As only 24 hydro-climatic classes describe the entire nation (and only 6 describe a significant majority), it is entirely possible that a couple dozen well-placed soil moisture sensors can enable reasonably accurate soil moisture modeling at any location within the continental United States.

Leveraging these findings, the discussion section also presented the results of preliminary analysis that illustrates how further improvements in soil moisture predictions could be gained by disaggregating based on local topography. This would enable more accurate predictions at sites characterized by peaks and valleys that dry faster or slower than the relatively flat locations at which soil moisture algorithms are generally calibrated. Incorporating overland flow into the diagnostic soil moisture equation and integrating satellite data into the approach could also improve predictions in the future.

Acknowledgements. The authors would like to acknowledge Praveen Kumar and Carl Bernacchi of the departments of Civil & Environmental Engineering and Plant Biology respectively at the University of Illinois in Urbana-Champaign for their suggestions, which have led to some of the edaphic analyses presented in this work.

References

- Barricelli, N. A.: Numerical testing of evolution theories, Part II. Preliminary tests of performance, symbiogenesis and terrestrial life, *Acta Biotheor.*, 16, 99–126, 1963.
- Berghuijs, W. R., Sivapalan, M., Savenije, H. H. G., and Woods, R. A.: The seasonal water balance as a window to explore catchment similarity at various time-scales, in preparation, 2014.
- Capehart, W. J. and Carlson, T. N.: Estimating near-surface soil moisture availability using a meteorologically driven soil water profile model, *J. Hydrol.*, 160, 1–20, 1994.
- Cheng, C. T., Zhao, M. Y., Chau, K. W., and Wu, X. Y.: Using genetic algorithm and TOPSIS for Xinanjiang model calibration with a single procedure, *J. Hydrol.*, 316, 129–140, 2006.

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Chico-Santamarta, L., Richards, T., and Godwin, R. J.: A laboratory study into the mobility of travelling irrigators in air dry, field capacity and saturated sandy soils, in: American Society of Agricultural and Biological Engineers Annual International Meeting 2009, Reno, NV, USA, 2629–2646, 2009.

5 Choudhury, B. J. and Blanchard, B. J.: Simulating soil water recession coefficients for agricultural watersheds, *Water Resour. Bull.*, 19, 241–247, 1983.

Coopersmith, E., Yaeger, M., Ye, S., Cheng, L., and Sivapalan, M.: Exploring the physical controls of regional patters of flow duration curves – Part 3: A catchment classification system based on regime curve indicators, *Hydrol. Earth Syst. Sci.*, 16, 1, doi:10.5194/hess-16-1-2012, 2012.

10 Entekhabi, D. and Rodriguez-Iturbe, I.: Analytical framework for the characterization of the space-time variability of soil moisture, *Adv. Water Resour.*, 17, 35–45, 1994.

Farago, T.: Soil moisture content: statistical estimation of its probability distribution, *J. Clim. Appl. Meteorol.*, 24, 371–376, 1985.

15 Fix, E. and Hodges, J. L.: Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties, Technical Report 4, USAF School of Aviation Medicine, Randolph Field, Texas, 1951.

Gamache, R. W., Kianirad, E., and Alshawabkeh, A. N.: An automatic portable near surface soil characterization system, *Geotech. Sp.*, 192, 89–94, 2009.

20 Goldberg, D. E.: *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley Professional, Boston, MA, USA, 1989.

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., and Leroux, D.: Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the US, *IEEE T. Geosci. Remote*, 50, 1530–1543, 2012.

25 Jones, H. G.: Irrigation scheduling: advantages and pitfalls of plant-based methods, *J. Exp. Bot.*, 55, 2427–2436, 2004.

O'Neill, P., Entekhabi, D., Njoku, E., and Kellogg, K.: The NASA Soil Moisture Active Passive (SMAP) mission: overview, NASA, Goddard Space Flight Center, Jet Propulsion Laboratory, available at: http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110015242_2011016052.pdf, last access: December 2013.

30 Pan, F.: Estimating daily surface soil moisture using a daily diagnostic soil moisture equation, *J. Irrig. Drain. E.-ASCE*, 138, 625–631, 2012.

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Pan, F. and Peters-Lidard, C. D.: On the relationship between the mean and variance of soil moisture fields, *J. Am. Water Resour. As.*, 44, 235–242, 2008.

Pan, F., Peters-Lidard, C. D., and Sale, M. J.: An analytical method for predicting surface soil moisture from rainfall observations, *Water Resour. Res.*, 39, 1314, doi:10.1029/2003WR002142, 2003.

Saxton, K. E. and Lenz, A.T.: Antecedent retention indexes predict soil moisture, *J. Hydraul. Div. Proc. Am. Soc. Civ. Eng.*, 93, 223–241, 1967.

Schaefer, G. L., Cosh, M. H., and Jackson, T. J.: The USDA Natural Resources Conservation Service Soil Climate Analysis Network (SCAN), *J. Atmos. Ocean. Tech.*, 24, 2073–2077, 2007.

Sharifat, K. and Kushwaha, R. L.: Sinkage simulation model for vehicles on soft soil, in: 2000 ASAE Annual International Meeting, Technical Papers: Engineering Solutions for a New Century, 1, 9–12 July 2000, Milwaukee, WI, USA, 2549–2553, 2000.

Simunek, J., Sejna, M., and van Genuchten, M.: The HYDRUS-1D Software Package for Simulating Water Flow and Solute Transport in Two-Dimensional Variably Saturated Media, Version 2, IGWMC–TPS–70, International Ground Water Modeling Center, Colorado School of Mines, Golden, CO, 1998.

Singh, A. and Minsker, B. S.: Uncertainty-based multiobjective optimization of groundwater remediation design, *Water Resour. Res.*, 44, W02404, doi:10.1029/2005WR004436, 2008.

Silva, R. B., Lancas, K. P., Miranda, E. E. V., Silva, F. A. M., and Baio, F. H. R.: Estimation and evaluation of dynamic properties as indicators of changes on soil structure in sugarcane fields of Sao Paulo State – Brazil, *Soil Till. Res.*, 103, 265–270. 2009.

Steel, R. G. D. and Torrie, J. H.: Principles and Procedures of Statistics with Special Reference to the Biological Sciences, McGraw Hill, New York, NY, USA, 187 pp., 1960.

Wetzel, P. J. and Chang, J. T.: Evapotranspiration from nonuniform surfaces – A 1st approach for short-term numerical weather prediction, *Mon. Weather Rev.*, 116, 600–621, 1988.

Zhang, X., Srinivasan, R., and Bosch, D.: Calibration and uncertainty analysis of the SWAT model using Genetic Algorithms and Bayesian Model Averaging, *J. Hydrol.*, 374, 307–317, 2009.

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)**Table 1.** Cross-validation results.

| | Unrelated class | Similar class | Same class |
|--------------------|-----------------|---------------|------------|
| Median | −10.5 % | −7.3 % | −0.8 % |
| Mean | −13.7 % | −7.7 % | −3.4 % |
| Standard deviation | 1.0 % | 1.1 % | 1.4 % |

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

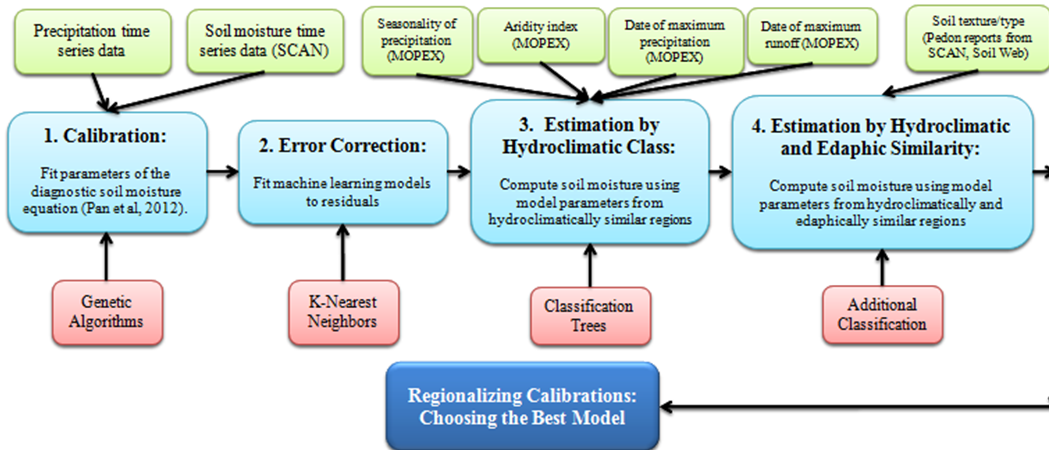


Fig. 1. Methodological flow chart.

| | |
|--------------------------|--------------|
| Title Page | |
| Abstract | Introduction |
| Conclusions | References |
| Tables | Figures |
| ◀ | ▶ |
| ◀ | ▶ |
| Back | Close |
| Full Screen / Esc | |
| Printer-friendly Version | |
| Interactive Discussion | |

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

**R² Value, By SCAN Site ID:
Before and After ML Correction of Residuals**

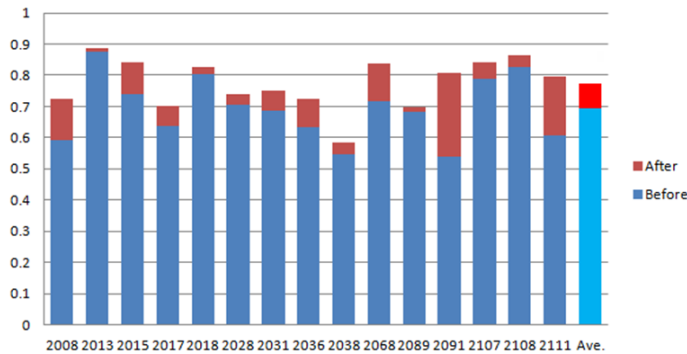


Fig. 2. Improvements from machine learning (KNN) models of residuals.

- Title Page
- Abstract Introduction
- Conclusions References
- Tables Figures
- Navigation: Home, Previous, Next, First, Last
- Back Close
- Full Screen / Esc
- Printer-friendly Version
- Interactive Discussion



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

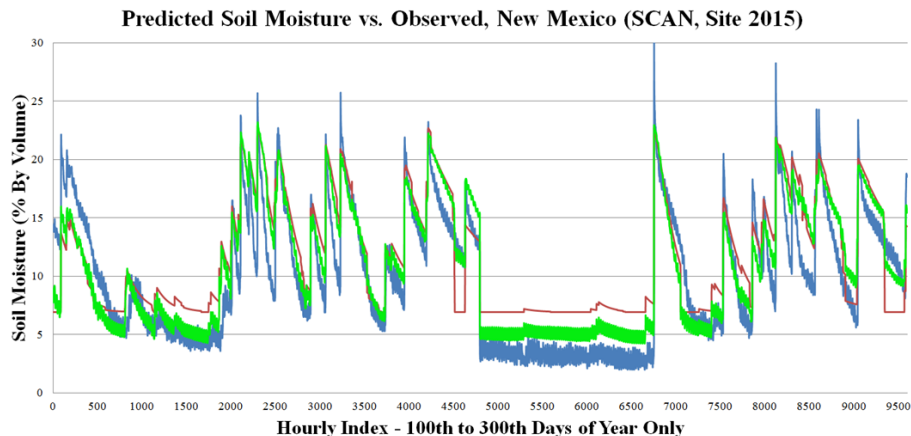


Fig. 3. Soil moisture time series, SCAN site 2015, New Mexico (USA), actual soil moisture (blue line), diagnostic soil moisture equation estimate (red line), and diagnostic soil moisture equation with machine learning error correction (green line).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

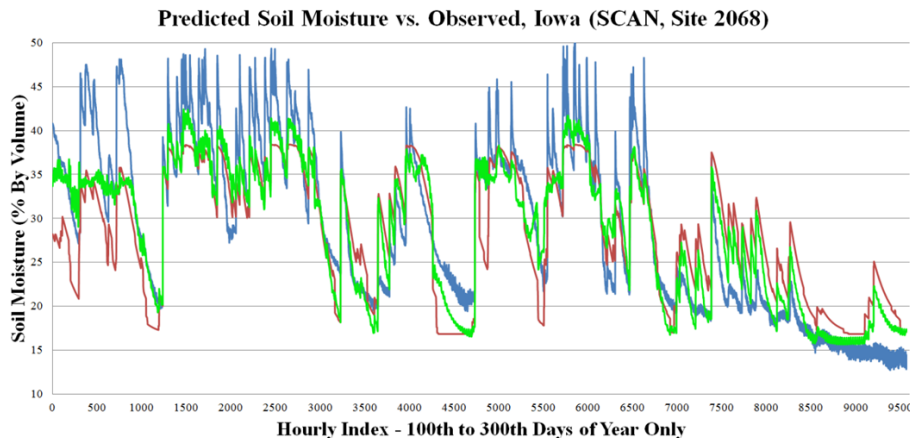


Fig. 4. Soil moisture time series, SCAN site 2068, Iowa (USA), actual soil moisture (blue line), diagnostic soil moisture equation estimate (red line), and diagnostic soil moisture equation with machine learning error correction (green line).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

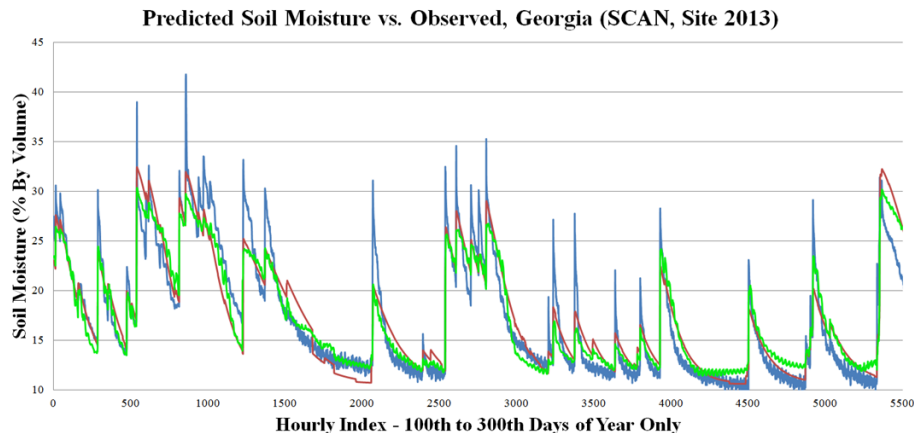


Fig. 5. Soil moisture time series, SCAN site 2013, Georgia (USA), actual soil moisture (blue line), diagnostic soil moisture equation estimate (red line), and diagnostic soil moisture equation with machine learning error correction (green line).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

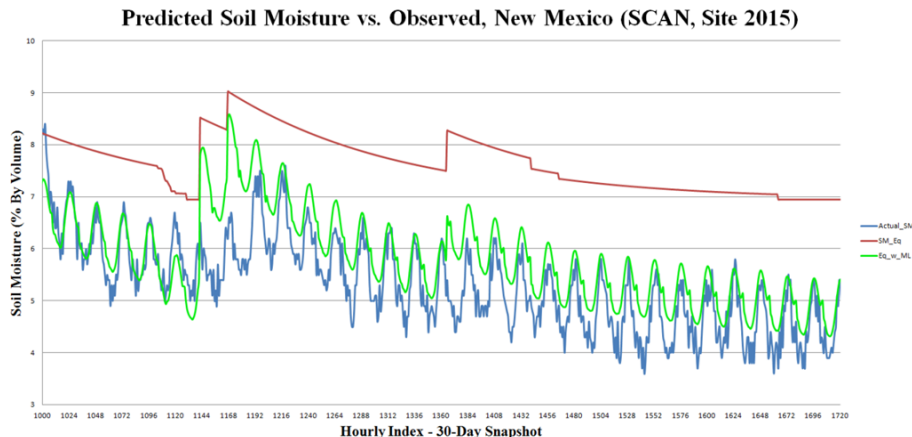


Fig. 6. Soil moisture time series, SCAN site 2015, New Mexico (USA), actual soil moisture (blue line), diagnostic soil moisture equation estimate (red line), and diagnostic soil moisture equation with machine learning error correction (green line).

[Title Page](#)
[Abstract](#)
[Introduction](#)
[Conclusions](#)
[References](#)
[Tables](#)
[Figures](#)
[⏪](#)
[⏩](#)
[◀](#)
[▶](#)
[Back](#)
[Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)

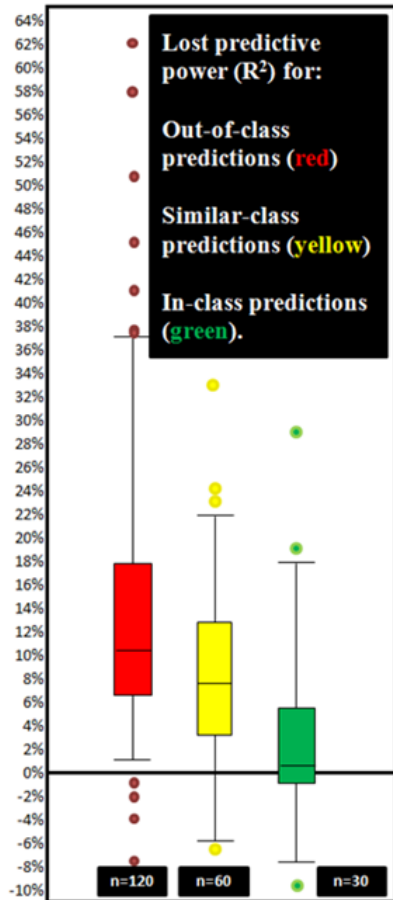


Fig. 7. Loss of predictive power (R^2) (y axis) between baseline predictions (model calibrated in the same watershed) and cross-validation predictions (model calibrated in other watersheds).

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

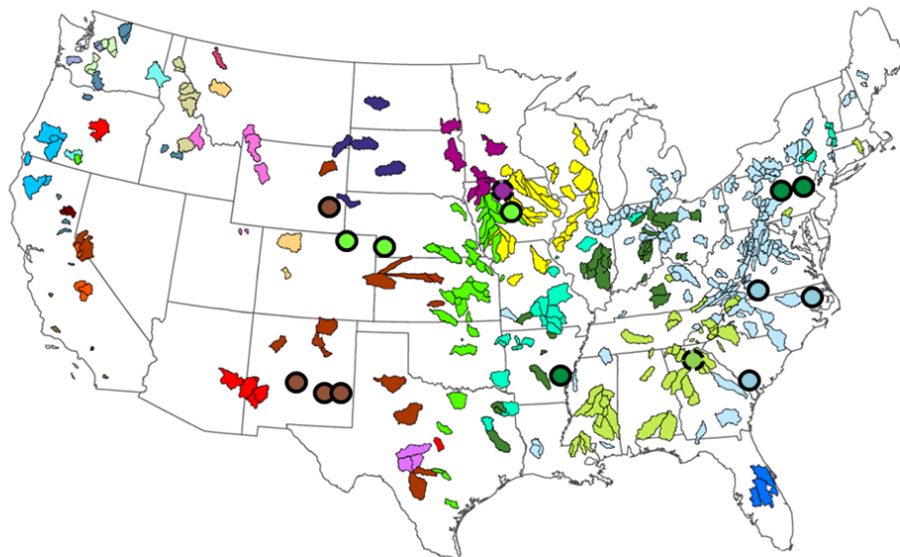


Fig. 8. 428 MOPEX catchments colored by hydro-climatic class (Coopersmith et al., 2012). 15 SCAN sensors (for which the diagnostic soil moisture equation is calibrated) are shown as colored circles. Circle colors correspond to the hydro-climatic class of the point in question. Circles with dotted borders are unique (no other sensor for calibration is available within that class).

HESSD

11, 2321–2353, 2014

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

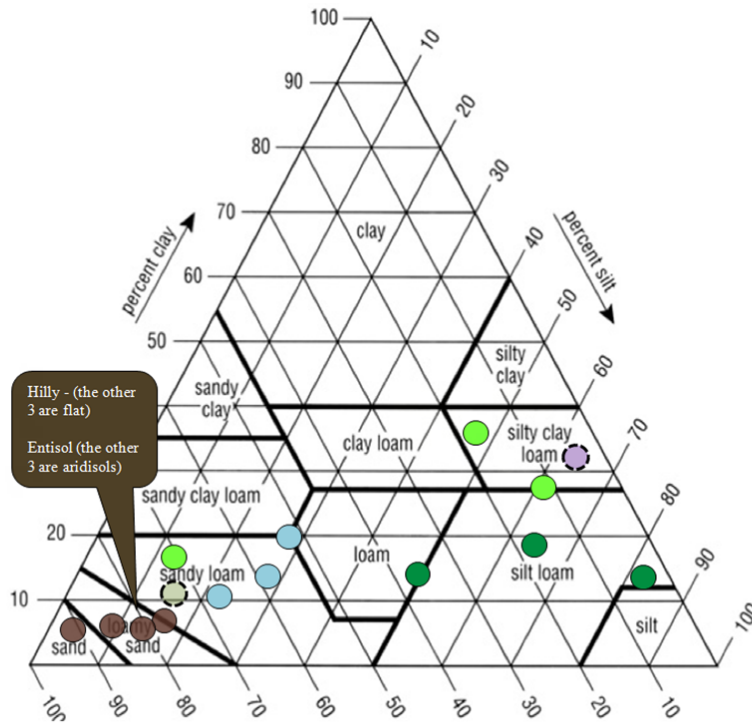


Fig. 9. The 15 SCAN sensors, color-coded to match their hydro-climatic class.

| | |
|--------------------------|--------------|
| Title Page | |
| Abstract | Introduction |
| Conclusions | References |
| Tables | Figures |
| ⏪ | ⏩ |
| ◀ | ▶ |
| Back | Close |
| Full Screen / Esc | |
| Printer-friendly Version | |
| Interactive Discussion | |



Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

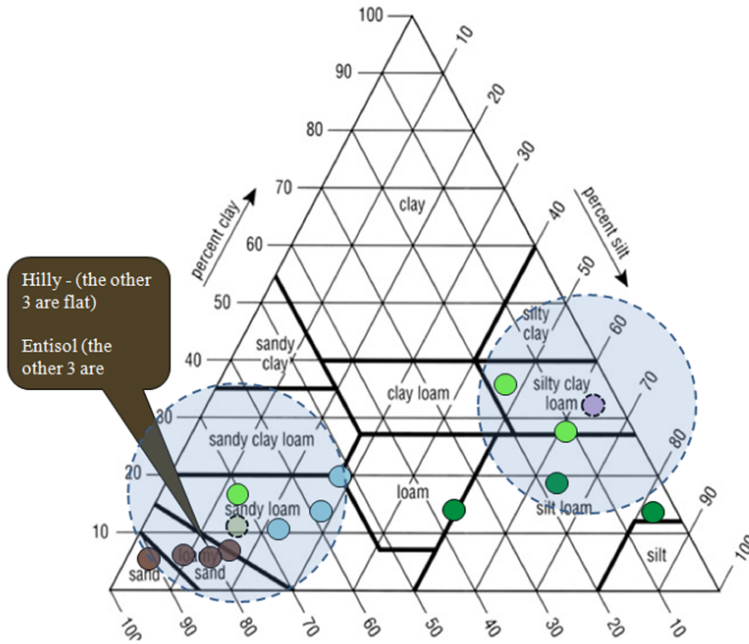


Fig. 10. The 15 SCAN sensors with similar soil textures shaded.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Using hydro-climatic and edaphic similarity to enhance soil moisture

E. J. Coopersmith et al.

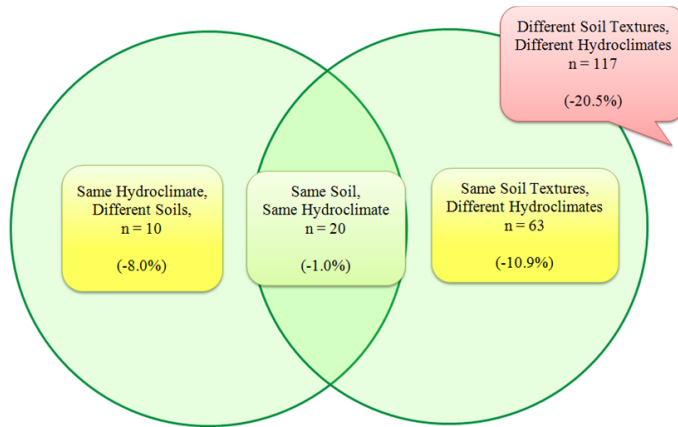


Fig. 11. Venn-diagram of modeling errors with similar and different soils and hydro-climates.

| | |
|--------------------------|--------------|
| Title Page | |
| Abstract | Introduction |
| Conclusions | References |
| Tables | Figures |
| ⏪ | ⏩ |
| ◀ | ▶ |
| Back | Close |
| Full Screen / Esc | |
| Printer-friendly Version | |
| Interactive Discussion | |