

1 **Using similarity of soil texture and hydro-climate to enhance soil moisture estimation**

2
3
4 E. J. Coopersmith,¹ B. S. Minsker,¹ and M. Sivapalan^{1,2}

5
6 ¹*Department of Civil & Environmental Engineering, University of Illinois at Urbana-*
7 *Champaign, Urbana, IL 61801, USA*

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

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26 **Abstract**

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

42 **Keywords:** soil moisture, hydro-climatic, edaphic, similarity, prediction, operational models

43

44

45

46

47

48

49

50

51

52 1. Introduction

53 Soil moisture estimates are needed routinely for many practical applications, such as irrigation
54 scheduling and operation of farm machinery. They are typically produced either through remote
55 sensing or sparse networks of *in situ* sensors. Although recent remote sensing studies have
56 confirmed that such measurements approximate *in situ* sensor networks (Jackson et al, 2012),
57 satellite-based sensors provide measurements at a spatial resolution of several kilometers – too
58 large for daily agricultural decision making. On the other hand, *in situ* sensor networks produce
59 values that are difficult to generalize to locations with no proximal sensors. Under these
60 circumstances, dynamic soil moisture evolution models are typically used for soil moisture
61 estimation at the desired location, using information from the nearest available sensors. This
62 method of soil moisture estimation immediately raises the issue regarding the type of model that
63 is most appropriate for such an application. One could think of several different types of models
64 that may be suitable.

65 The first group of soil moisture models considers only the variability of precipitation, as it
66 has been shown that precipitation variability is the primary mechanism for wetting/drying
67 (Entekhabi and Rodriguez-Iturbe, 1994). Many subsequent models employed an “antecedent
68 precipitation index” (API), defining a pre-established temporal window for antecedent rainfall.
69 This index is then used to estimate current levels of soil moisture (Saxton and Lenz, 1967) and
70 has been implemented with recession modeling for soil water in agriculture (Choudhury and
71 Blanchard, 1983) and also in weather prediction (Wetzel and Chang, 1988). Other precipitation-
72 focused approaches utilize stochastic models to estimate the distributions of soil moisture values
73 using an initialization of daily rainfall (Farago, 1985). Both the stochastic and API approaches
74 require some initial condition for soil moisture at the forecast location – requiring either
75 professional judgment or a sensor. While these issues can be addressed using a soil water
76 balance model, this type of model must be recalibrated frequently, which most soil moisture
77 models are not, as its errors are cumulative (Jones, 2004).

78 The second group of models adopts a process-based approach, estimating soil moisture
79 from surface radiation and precipitation (Capehart and Carlson, 1994). These process-based
80 models are typically forced by evapotranspiration demand and precipitation at their upper
81 boundary and, if applicable, by groundwater at their lower boundary. More sophisticated models

82 of this type, such as HYDRUS (Simunek et al, 1998), attempt to improve predictions via detailed
83 knowledge of hydraulic soil parameters, information regarding root structures, soil temperature
84 readings, and detailed atmospheric/meteorological information, which are not widely available,
85 especially for routine applications envisaged here.

86 The third group of models are agriculturally-focused, building model projections outward
87 from existing instrumentation and additional measurements. Gamache et al (2009) developed a
88 soil drying model for which cone penetrometers and soil moisture sensors are required. At most
89 remote sites, these data sources are not currently accessible. Another similar approach employs
90 specific soil type information (theoretically, publicly available data), but ultimately requires
91 proximal sensors to provide the needed soil moisture estimates (Chico-Santamaria, et al, 2009).

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

104 The aim of this paper is to present and test an approach that can help overcome the issues
105 of calibration at ungauged locations associated with the Pan et al. soil moisture estimation model.
106 The proposed solution involves calibrating the Pan (2012) diagnostic soil moisture equation
107 (model) at gauged sites and then extrapolating the calibrated model to ungauged sites by
108 invoking similarity. Similarity here is defined on the basis of hydro-climatic characteristics,
109 using a classification system developed by Coopersmith et al (2012), as well as edaphic (soil)
110 properties. The proposed new scheme maintains the advantage of Pan et al.’s parsimonious soil
111 moisture model in that it does not require specification of initial soil moisture condition, and also

112 there is no need to recalibrate periodically. The model's simplicity also permits implementation
113 of the model in a manner that can easily be refit with new parameters, where necessary. Section
114 2 provides more details on the approach.

115 To calibrate and validate the model, data from the U.S. Department of Agriculture's
116 (USDA) Soil Climate Analysis Network (SCAN) were used. This national array of soil moisture
117 sensors (with co-located precipitation sensors) delivers hourly data at a variety of publically-
118 accessible sites throughout the United States. Fifteen sensor locations with numerous years of
119 high-quality, minimally-interrupted data were selected for further analysis. These sites display
120 considerable hydrologic diversity, which aids in demonstrating that the nationwide application of
121 the proposed soil moisture model using precipitation data represents a feasible goal. With
122 respect to agricultural decision-support, for energy-limited sites, the value of hourly soil moisture
123 estimates is found in the determination of whether or not a field is trafficable – whether heavy
124 equipment will damage fields or become mired. With respect to water-limited sites, the value of
125 soil moisture estimates is found in devising optimal irrigation strategies that utilize limited water
126 resources most efficiently. Of the fifteen SCAN sites examined, the three sites in New Mexico,
127 the site in Colorado, the site in Nebraska, the site in Wyoming, and the two in Iowa are all water-
128 limited (8 in total). The remaining sites (7 in total), located in Pennsylvania (2), Arkansas,
129 Georgia, South Carolina, North Carolina, and Virginia, are all energy-limited. Results of the
130 analysis are given in Section 3, followed by discussion in Section 4 to suggest further
131 improvements and conclusions are presented in Section 5.

132 **2. Methodology**

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

142 addition to hydro-climates. The results of these four steps are then examined to identify which
143 approach to regionalization performs best.

144

145 **Step 1: Calibration Using a Two-Layer Genetic Algorithm**

146 Unlike the original diagnostic soil moisture calibrations, the ultimate objective of this work is to
147 enable agricultural decision support in near real time. To this end, the daily model from Pan
148 (2012) is first modified to yield an hourly model within the same framework. Genetic algorithms
149 are then deployed to calibrate the model, enabling more efficient exploration of the parameter
150 search space than the traditional Monte Carlo search, which was the approach taken by Pan
151 (2012).

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

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

159

$$160 \quad \theta_{est} = \theta_{re} + (\phi_e - \theta_{re})(1 - e^{-c_4\beta}) \quad (1)$$

161

162 Here θ_{est} represents the best estimate of soil moisture during a given hour. θ_{re} denotes residual
163 soil moisture, the minimum quantity of moisture that is present regardless of the length of time
164 without precipitation. ϕ_e , the soil's porosity, signifies the maximum possible soil moisture value,
165 at which point the soil becomes saturated. Finally, c_4 is a parameter related to conductivity and
166 drainage properties, essentially defining the rate at which soil can dry. If c_4 assumes a value of
167 zero, the soil is permanently at its residual soil moisture value, θ_{re} - a soil that dries infinitely
168 rapidly. Conversely, as c_4 becomes large, the soil will permanently assume the value of its

169 porosity, ϕ_e – a soil that dries infinitely slowly. The β term in Equation 1 is calculated in
170 Equation 2 below:

171

$$172 \quad \beta = \sum_{i=2}^{i=n-1} \left[\frac{P_i}{\eta_i} \left(1 - e^{-\frac{\eta_i}{z}} \right) e^{-\sum_{j=1}^{j=i-1} \left(\frac{\eta_j}{z} \right)} \right] + \frac{P_1}{\eta_1} \left(1 - e^{-\frac{\eta_1}{z}} \right) \quad (2)$$

173

174 Here, P_i denotes the quantity of rainfall during hour i (day in the original presentation in Pan et
175 al., 2003). The soil depth at which an estimation occurs is given by z . This convolution
176 summation has a temporal window of n hours for considering past precipitation. For instance,
177 today's soil moisture is strongly influenced by yesterday's rainfall, influenced to a lesser degree
178 by last week's rainfall, and not influenced at all by rainfall from ten years previous. Given the
179 general limitation of our datasets and the fact that shallow-depth soil moisture is most relevant to
180 decision-support, all of our analyses occur with measurements of two inch (~5cm) depth.

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

187 Finally, the estimated soil water loss at hour i , e.g. due to evapotranspiration or deep
188 drainage, is expressed by the term, η_i . As this algorithm does not presume any more detailed
189 knowledge of potential evaporation/drainage behaviors, this “eta series,” representing losses due
190 to evapotranspiration and deep drainage, is modeled as a sinusoid (Pan, 2012) with period 8,760
191 (the number of hours in a year). The eta (η) series is required to calculate the beta (β) series (Eq.
192 2), which is required to use the diagnostic soil moisture equation (Eq. 1). Thus, before any other
193 parameters are chosen, a generalized sinusoidal form of η is estimated as given in Equation 3:

194

195 $\eta = \alpha \sin(i - \delta) + \gamma$ (3)

196

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

204

205

206 **Step 2: Error Correction Using The k-Nearest Neighbors Machine Learning Algorithm**

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

211

212 $\theta_{obs} = \theta_{est} + \varepsilon$ (4)

213

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

215 To correct biases in these errors, the k-Nearest Neighbor algorithm (Fix and Hodges,
216 1951) is employed to predict ε using the characteristics from the training data. More
217 specifically, the data are searched for the most similar matches in terms of time of day, day of
218 year, θ_{est} , $\beta(n)$, and $\beta(M) - \beta(n)$. For example, if the model returns a prediction of $\theta_{est} =$
219 0.35 at 2:00pm during July when rainfall has been heavy recently but drier over a longer period,
220 KNN will search the training set for other estimates near 0.35 made on mid-summer afternoons

221 where a similar recent rainfall pattern has been observed. Next, the algorithm averages the value
222 of the error, ε , associated with those types of conditions, producing an estimated error, ε_{est} .
223 Each validation estimate is then adjusted to be $\theta_{est} + \varepsilon_{est}$. This technique allows consistent
224 model biases, such as underestimating wetter days and overestimating drier days, to be corrected.

225 This error correction model also accounts for diurnal soil moisture variations that were not
226 considered in developing the diagnostic soil equation, which was designed to deliver daily soil
227 moisture estimates. Consider a soil moisture estimate at 4pm, after soil has had a full day of
228 sunlight (theoretically) to dry. As the diagnostic soil moisture equation only considers drainage
229 and evapotranspiration losses on a daily basis, θ_{est} will be larger than θ_{obs} . Yet, because this
230 type of mistake presumably occurred frequently throughout the training data, the algorithm will
231 locate other 4pm estimates, each of which will be biased in the same direction, and our final soil
232 moisture estimates will take this bias into account, improving the results as shown subsequently.

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

$$236 \quad R^2 = 1 - \frac{SSR}{SST} \quad (5)$$

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

239

240 **Step 3: Estimation by Hydro-climatic Similarity**

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

246 This hypothesis was tested at fifteen SCAN sensors for which soil moisture and
247 precipitation data are available hourly for a period of several years. These sensors are located in

248 diverse geographic locations and hydro-climatic classes in Iowa, North Carolina, Pennsylvania,
249 New Mexico, Arkansas, Georgia, Virginia, South Carolina, Nebraska, Colorado, and Wyoming.
250 The data at each of these locations were divided into training/validation sets and parameters were
251 calibrated using training data only. Next, these parameters were employed on the validation sets
252 at the locations for which they were calibrated. The subsequent R^2 values (proportion of
253 variance in soil moisture explained by the machine-learning-enhanced diagnostic soil moisture
254 equation, see Steel and Torrie, 1960, for reference) defined a baseline level of performance for
255 that site.

256 The process of cross-validation is detailed below:

- 257 1. Consider two sites, x and y , chosen from the fifteen available calibrated locations.
- 258 2. Estimate the soil moisture values in the validation dataset of site y , using the parameters
259 calibrated from the training dataset at site x .
- 260 3. Record the difference between the R^2 baseline value at site y (obtained using parameters
261 calibrated at site y) and the performance obtained at site y using parameters calibrated at
262 site x .
- 263 4. Repeat steps 1-3 for all 210 possible (x, y) pairs where $x \neq y$
264 Note: (x, y) and (y, x) are not equivalent. One signifies the performance of parameters
265 calibrated at site x making predictions at site y , the other signifies the performance of
266 parameters calibrated at site y making predictions at site x .

267

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

274

275 **Step 4: Estimation by Hydro-climatic and Edaphic Similarity**

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

280 This information allows sites already deemed hydro-climatically similar to be further sub-
281 divided into sites that are and are not edaphically similar. Analogous to the previous section, we
282 consider pairs of sites, x and y , where parameters are calibrated at site x and validated at site y .
283 In this case, four groups can be defined – the first, where x and y and hydroclimatically similar,
284 the second, where x and y are hydroclimatically similar, but differ edaphically, the third, where
285 x and y are edaphically similar, but differ hydroclimatically, and finally, where x and y are
286 hydro-climatically and edaphically dissimilar.

287

288 **3. Results**

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

294

295 **3.1 Testing the Value of Machine Learning Error Correction for Soil Moisture Prediction** 296 **Using the Diagnostic Soil Moisture Equation**

297 Figure 2 shows the performance of the calibrated parameters for the 15 SCAN sites using only
298 the diagnostic soil moisture equation (Step 1 of the methodology) along with the subsequent
299 improvement in performance following machine learning error correction (Step 2). In each case,
300 the six parameters required for the implementation of the diagnostic soil moisture equation are
301 calibrated using training data from before 2010. Sensors with hourly precipitation and soil

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

302 moisture time series data between 2004 and 2009 (inclusive) provide four to six years of training
303 data (some sites are missing one or two years of data). Only days of the year where snow cover
304 is unlikely are used to train the algorithm (from the 100th to 300th day of the year in all
305 locations, for consistency). Validation data consist of days 100-300 for 2010 and 2011.

306 The results illustrate that in all fifteen test cases, performance within the validation
307 sample is improved by machine learning modeling of residuals from the training set, in some
308 cases, as much as 26.9% of the unexplained variance (site 2091) in soil moisture is corrected
309 from by this technique. The average results (far right column, Fig. 2) illustrate that the
310 diagnostic soil moisture equation explains just 69.2% of the variance in soil moisture ($\rho = 0.83$)
311 before machine learning corrections occur, but explains 77.5% of the variance in soil moisture (ρ
312 = 0.88) thereafter.

313 To explore these findings in more detail, three of the 15 SCAN sites, chosen to represent
314 different hydro-climatic locations – New Mexico (#2015, hydroclimate IAQ/southwestern
315 desert, Loamy Sand), Iowa (#2068, hydroclimate ISCIJ/northern midwest plains, Silty Clay
316 Loam), and Georgia (#2013, hydroclimate LWC/southeastern forest, Sandy Loam) are examined
317 to illustrate how improvements from adding machine learning error models to the diagnostic soil
318 moisture equation differ across sites. These three sites represent three distinct hydro-climatic
319 classes, with significant differences in soil texture, seasonality of precipitation, aridity, timing of
320 maximum precipitation, and timing of maximum runoff. Using error correction models for
321 prediction at these sites increased R²-values by an average of 8.2%, which is similar to the 8.3%
322 improvement in R² averaged across all fifteen sites. Thus, these three locations are
323 representative in terms of both hydro-climatic and edaphic diversity and their responsiveness to
324 machine learning.

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

332 In Figure 3, the diagnostic soil moisture equation is able to trace the general trend of the
333 soil moisture time series ($\rho = 0.860$). However, during the middle of the time series, in which
334 the observed soil moisture values fall below 5%, the benefits of machine learning error
335 correction are most noteworthy. There are other hours scattered throughout the dataset where the
336 green line (ML prediction) follows the blue line (observed values) much more closely than the
337 red line (diagnostic soil moisture equation). The green line ($\rho = 0.917$) not only improves upon
338 the correlation value of Pearson's Rho (the square root of the R^2 value in Eq. 5), but also displays
339 marked improvement for those cases in which the diagnostic soil moisture equation produces
340 significant errors.

341 During the validation period, specifically 2010, wetter conditions were observed than
342 were present during calibration. At this SCAN site, before 2010, the average soil moisture value
343 observed was 28.55%, with only 25% of values exceeding 35% volumetric soil moisture.
344 However, in 2010, the average soil moisture value measured was 33.16% with 45% of values
345 exceeding 35%. The machine learning driven error correction improves the diagnostic soil
346 moisture equation ($\rho = 0.846$) significantly ($\rho = 0.915$), but fails to raise its forecasts to reach
347 some of the wetter conditions experienced in validation. Underestimations of this nature,
348 although detrimental in terms of numerical errors, are not necessarily a problem for decision
349 support of agricultural or construction activities, for example. If a model warns that a site is very
350 wet and in reality, it is even wetter than predicted, the user has still been given adequate warning
351 not to attempt activity at that site. It is important to note that small errors are more significant in
352 terms of decision support (specifically when and where to irrigate) during dry conditions.
353 Generally, the model's errors are smaller, in absolute terms, during drier conditions. This
354 analysis's approach to error correction, as it relies on previous errors to predict future errors, will
355 not address long-term trends within the soil moisture record.

356 In Figure 5, a soil moisture series from Georgia is modeled by the diagnostic soil
357 moisture equation. Even before adding any error correction, the equation performs well ($\rho =$
358 0.936) and the machine learning approach yields a smaller improvement ($\rho = 0.941$). It is worth
359 noting that machine learning does not damage an already excellent performance, offering slight
360 improvements when possible and essentially no correction when training data suggest the model
361 has already performed adequately.

362 Table 1 presents all fifteen sites for which the diagnostic soil moisture equation has been
363 calibrated, including information regarding their hydroclimatic class from Coopersmith et al
364 (2012), their soil textural characteristics, and their performance before and after the KNN bias
365 correction process.

366 **3.2 Bias Correction – More Detailed Results**

367 In addition to generalizing the parameters calibrated in the diagnostic soil moisture
368 equation, the error correction approach allows for systematic biases to be removed by searching
369 training data for similar conditions and then predicting the types of mistakes most likely to occur.
370 Figure 6, by zooming in upon a 30-day period from Figure 2, illustrates how machine learning
371 reduces errors by introducing a diurnal cycle into a model that previously lacked one. The
372 remaining bias is likely explained by a slightly wetter training dataset as compared with the
373 validation data. It is possible that the diurnal cycle at some locations reflects a soil moisture
374 probe's dependency on electromagnetic properties driven by temperature change (apparent
375 permittivity) rather than hydrologic processes (Rosenbaum et al, 2011). However, the model's
376 ability to respond to these nuances would not compromise its performance were these nuances
377 subsequently removed.

378 Any corrective algorithm will, over thousands of validation points, push the estimate
379 away from the observed value in some cases. However, the results from Table 1 demonstrate
380 that its overall performance represents an improvement at all sites, and thereby justifies its use.
381 Regarding the issue of 'measurement artifacts,' whether the diurnal cycle is genuine or an
382 idiosyncratic sensor output, the model is tasked with calibrating itself and correcting biases as
383 defined by the empirically-reported data. Figure 6 illustrates its ability to do so. Were the
384 sensors to no longer report such a diurnal pattern (i.e. it is merely a measurement artifact, and
385 subsequently corrected), the machine learning step would no longer observe those biases, and
386 consequently, no longer introduce such a pattern. The accuracy of the SCAN network is a
387 relevant inquiry, but unfortunately, not within the scope of this paper.

388 By addressing such systematic biases, machine learning enables model performance to
389 improve with each successive growing season as the training dataset expands. For instance,
390 although the fields in Iowa endured flooding during the validation period and subsequently made
391 errors, such errors would eventually populate the training data. The next time such flooding

392 occurs, the model is likely to recognize the occurrence of those same conditions and adjust the
393 diagnostic soil moisture equation's predictions accordingly. In this vein, model performance is
394 likely to improve over time, especially with the models already showing reasonable accuracy
395 using only a few years of training data.

396 Figure 7, 8, and 9 present these results in more detail for each of the three SCAN sites
397 presented in Figures 3, 4, and 5. In each figure, the upper-left image presents the average bias
398 correction (change in % soil moisture) for each hour of the day (0-23). At all three sites, bias
399 corrections display a clear diurnal pattern – that is to say the removal of a diurnal cycle is a
400 substantial role of machine learning under a variety of hydroclimatic and edaphic conditions.
401 The upper-right image of each figure presents the bias correction as a function of the unadjusted
402 soil moisture estimate – essentially, whether there exists a systemic over- or underestimation
403 when values are high or low.

404 The first two sites (Figures 7 and 8) do not present a clear pattern, but Figure 9 displays a
405 trend suggesting that the highest estimates of soil moisture tend to be overestimates and the
406 lowest estimates of soil moisture tend to be underestimates – but these biases are removed via
407 machine learning. The lower-left image presents bias correct as a function of the day of the year
408 (from 100-300, the days of the year when the model is applied). At all three sites, the seasonal
409 cycle does appear in terms of the patterns of bias correction, but the pattern is noisier than the
410 diurnal cycle. The magnitude of the adjustments are largest in the monsoon-affected desert of
411 New Mexico, a bit smaller in the Midwestern plains characterized by less extreme seasonal
412 behavior, and smallest in the Southeast where seasonal variations are low.

413 Finally, the lower-right image relates bias correction to the beta series from the diagnostic soil
414 moisture equation (Pan, 2012), a convolution of a decaying precipitation time series working
415 backwards temporally from the current time. Stated differently, these charts relate bias
416 correction to the amount of antecedent precipitation (with more recent precipitation weighted
417 more heavily). In Figure 7 (Plains, Silty Clay Loam), the model tends to underestimate moisture
418 when large quantities of antecedent rainfall are present, where in Figure 9 (Woods, Sandy
419 Loam), once antecedent precipitation becomes non-trivial, displays the opposite pattern. This is
420 consistent with the finer Midwestern soils' proclivity for ponding/flooding due to larger
421 proportions of clay. In these cases, larger amounts of rain will soak soils from above, and
422 capillary rise might further soak sensors from below, leading to underestimation from the

423 diagnostic soil moisture equation and subsequent machine learning correction. By contrast, with
424 sandier soils, drainage occurs easily, leading to higher rates of loss than the eta series (Pan, 2012)
425 would predict (there is more available water to lose), leading to overestimation with large
426 amounts of antecedent rainfall.

427 **3.3 Cross-Validation Results for Hydro-climatic Similarity: Qualitative Findings and** 428 **Significance Testing**

429 To test the hypothesis that models calibrated in one location can be used in a hydro-climatically
430 similar location, cross-validation was used as described in Step 3 of Section 2. The fifteen SCAN
431 sites yield $15^2 = 225$ possible (x, y) pairs. Fifteen of these 225 pairs occur when $x = y$,
432 establishing the baseline level of performance for a given site (validation performed using the
433 parameters calibrated at that same location). Of the 210 remaining (x, y) pairs, 120 of them
434 consist of paired catchments in which x and y are located in unrelated classes, 60 consist of
435 paired catchments in which x and y are located in a “similar” class (different by a single split
436 within the classification tree), and 30 consist of paired catchments in which x and y fall within
437 the same hydroclimatic class (but x and y do not represent the same catchment). Figure 10
438 presents box plots illustrating the change in R^2 values for these three sets of pairs in a manner
439 analogous to the differences shown in Figure 2. Table 2 presents the quantitative results, again
440 averaging the deterioration of performance in terms of change in R^2 .

441 These findings show that calibrating the model at one location and applying those
442 parameters elsewhere within the same class (green) is preferable to applying those parameters in
443 a similar, but not identical class (yellow) and vastly superior to applying those parameters in an
444 unrelated class (red). The differences between any two clusters (same-class, similar-class,
445 unrelated class) are all significant at the $\alpha = 0.01$ level ($p < .001$ in all cases) as calculated by a
446 two-sample, heteroscedastic t-test (Welch, 1947).

447

448 **3.4 Impact of Soils: Cross-validation Results for Edaphic and Hydro-climatic Similarity**

449 To isolate the impacts of soil types (edaphic similarity) on soil moisture prediction,
450 groups of sensor locations among the 15 SCAN sites that are hydro-climatically similar were

451 analyzed, shown in Figure 11. The soil textural data for each of these fifteen sensors are plotted
452 on a soil texture pyramid diagram in Figure 12. These data were obtained from either Pedon Soil
453 Reports available through the SCAN network (which provide precise percentages of clay, silt,
454 and sand), or, where this information was unavailable, from soil information in the national soil
455 Web database².

456 Of the thirteen sensors from the four hydro-climatic classes with multiple SCAN sensors
457 (light green, blue, dark green, and brown in Figures 11 and 12), 30 (x, y) pairs exist where the
458 model can be calibrated at site x and its parameters applied at site y. Note that (x, y) is not
459 equivalent to (y, x) as the sites for calibration and validation are reversed. Of these 30 pairs, 20
460 pairs are edaphically similar as well. However, 10 of them include a pair of points where the soil
461 types or terrain types are notably misaligned (for example, light green dots in Figure 12 where
462 two of the three sensors are in silty clay loam and the third is in sandy loam– a notably different
463 soil). A similar analysis to the one presented in Figure 10 and Table 2 has been reproduced,
464 comparing the loss in predictive power (R^2) for the 20 pairs with similar hydro-climates and soils
465 against the loss for the 10 pairs in which either the soil texture (Figure 12) or type do not align.
466 The average loss of 1.0% for the 20 very similar pairs is a much smaller decline than the 8.0%
467 average decline observed for the 10 pairs for which soil/terrain information suggests
468 dissimilarity. These results are significant with a p-value of approximately 0.02. Additionally,
469 the upper-most two green dots in Figure 10, where calibrated parameters at one location perform
470 poorly at another of similar hydro-climatic class, fall within these 10 cases.

471 These observations show the importance of soil information, or edaphic similarity. While
472 pairs of calibration/validation locations with similar hydro-climates, but dissimilar soils, show a
473 decline in performance as compared with pairs of locations where both are similar, so too do
474 locations with similar soils, but dissimilar hydro-climates. The shaded circles in Figure 12
475 illustrate groups of sensors that are quite similar in terms of soil textures. However, despite their
476 soil similarities, differences in hydro-climates hinder cross-application, showing a decline in
477 performance of 10.9% for all (x, y) pairs within the shaded regions of Figure 12 for which x and
478 y are not from the same hydro-climatic class.

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

479 As summarized in Figure 13, these results suggest that in cases where both soil type and
480 hydro-climate align, very little performance is lost when parameters are re-applied (1.0%),
481 moderate declines in performance are observed when one of these two factors are aligned (8.0%
482 if hydro-climates align and soil types do not; 10.9% if soil types align, but hydro-climates do
483 not), and large declines in performance appear when neither align (20.5%). Clearly both types of
484 attributes are important and should be considered in future modeling work in which the relative
485 importance of hydroclimates and soil textures can be examined in greater detail.

486

487 **4. Discussion: Future Work to Improve Predictions**

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

494 **4.1 Estimates Enhanced By Topographic Classification**

495 Ultimately, the combination of a hydro-climatic classification system and the diagnostic soil
496 moisture equation demonstrates a generalization of calibrations, facilitating predictions at any
497 location where a viable sensor exists within a similar hydro-climatic class and soil type.
498 However, the lumped, bucket model is not ideally-suited for landscapes with complex
499 topography. Conveniently, the majority of SCAN sites are placed on relatively flat surfaces.
500 Integration of topographic insights is a fertile area for future research. One possible approach to
501 further improving predictive accuracy is to disaggregate the soil moisture estimates as a function
502 of local topography. While SCAN sites used for soil moisture data are generally located on flat
503 surfaces, predictions may be needed at locations located on ridges or in valleys where the soils
504 are likely to be wetter or drier than their surroundings. This requires the notion of regional
505 topological classification. In this manner, the notion of similarity is extended to include hydro-
506 climatology, soil characteristics, and topographic designation (ridge, slope, valley, etc).
507 Preliminary analyses suggest that small-scale topography does play a meaningful role in the

508 wetting/drying process. Future research with more extensive datasets in locations with more
509 complex topological contours could improve soil moisture predictions by enabling the models
510 developed in this work to be adjusted as a function of local topographic classification.

511

512 **4.2 An Enhanced Diagnostic Soil Moisture Equation**

513 The diagnostic soil moisture equation could also be improved in future modeling efforts by
514 considering overland and subsurface flows, specifically in areas characterized by more complex
515 topography. Currently, the model assumes that, in the absence of saturation, all rainfall will
516 ultimately infiltrate, as the porosity parameter serves as an upper bound on soil moisture levels.
517 The diagnostic soil moisture equation was designed originally as a daily model, and it is
518 probably rare that on any given day, a significant fraction of precipitation does not infiltrate.
519 However, at the hourly scale it is quite possible that the water from an intense rainfall event will
520 not make its way into the soil at the location of the sensor. To address this lateral transfer
521 phenomenon, additional parameters can be introduced into the diagnostic soil moisture equation
522 that place an upper bound on the quantity of rainfall that can be infiltrated during any hour (or
523 other interval) of the convolution calculation for any particular soil type. Agricultural decision-
524 support includes trafficability when wet (Coopersmith et al, 2014) and irrigation support when
525 dry. While overland flow is perhaps an unneeded component in water-limited catchments where
526 irrigation schemes represent the most significant soil-moisture-related decision, in wetter
527 catchments, in which trafficability is a real concern, such an addition could improve the model.
528 While this approach would require the fitting of additional parameters, it is likely that predictions
529 would be improved. These additional parameters could also be considered in assessing cross-site
530 edaphic similarity using the methods described above, although they may be highly correlated
531 with existing parameters such as porosity, residual soil moisture, and drainage.

532

533 **4.3 Water Balance Models and Up-Scaling**

534 The diagnostic soil moisture equation used in this paper (Pan et al, 2003; Pan, 2012) was an
535 appropriate choice due to its ability to generate soil moisture estimates without the need for

536 knowledge of antecedent soil moisture conditions. Koster and Mahanama (2012) and Orth et al.
537 (2013) have developed approaches to estimate soil moisture at the watershed scale by leveraging
538 hydroclimatic variability and long-term streamflow measurements in a water-balance model –
539 also without employing previous soil moisture conditions. If the parameters calibrated and then
540 generalized in this work produce point estimates of soil moisture at a diversity of locations,
541 integration with a water balance approach could help with the up-scaling process.

542

543 **5. Conclusions**

544 This work has demonstrated the feasibility of estimating soil moisture at locations where soil
545 moisture sensors are unavailable for calibration, provided they fall within hydro-climatically and
546 edaphically similar areas to gauged locations. By calibrating the diagnostic soil moisture
547 equation via a two-part genetic algorithm, improving its performance via a machine learning
548 algorithm for error correction, then validating that algorithm at the same location in subsequent
549 years, a baseline level of predictive performance is established at fifteen locations. Next, these
550 results are cross-validated – deploying parameters calibrated at a given site at sites of similar and
551 different hydro-climatic classes, demonstrating that parameters can be re-applied elsewhere
552 within the same class, but not without. Finally, by incorporating edaphic information, we
553 observe the strongest cross-validation results when hydro-climatic and edaphic characteristics
554 align. As only 24 hydro-climatic classes describe the entire nation (and only 6 describe a
555 significant majority), it is entirely possible that a couple dozen well-placed soil moisture sensors
556 can enable reasonably accurate soil moisture modeling at any location within the continental
557 United States.

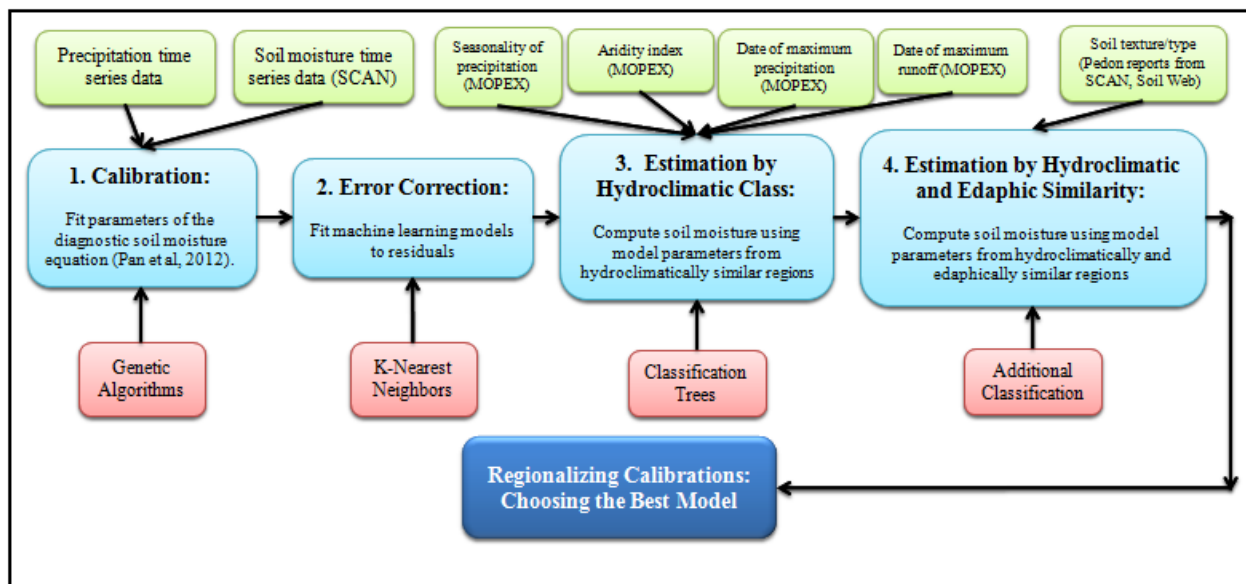
558 It is likely that the types of errors made when parameters are cross-applied between sites
559 of different hydroclimates will differ from the types of errors that appear when the sites differ
560 edaphically. Further research extending beyond model performance into the specific conditions
561 under which models perform less effectively along with the magnitude and bias of those errors
562 would be highly illustrative for future researchers.

563 This analysis can improve agricultural decision-support by offering insight into locations
564 that can benefit from targeted irrigation in drier conditions, or conversely, by minimizing risks of

565 ruts and damaged equipment when fields are no longer trafficable during wetter conditions.
 566 Scaling the results of these models upward can assist with larger-scale assessments of flood risks
 567 or as calibration/validation tools for satellite estimates of soil moisture. Scaling these results
 568 downward can help maximize yields. Given the ubiquity of precipitation data, which are the
 569 only inputs these models require, better understanding of the transferability of modeled
 570 parameters is a step towards far wider availability of soil moisture estimates.

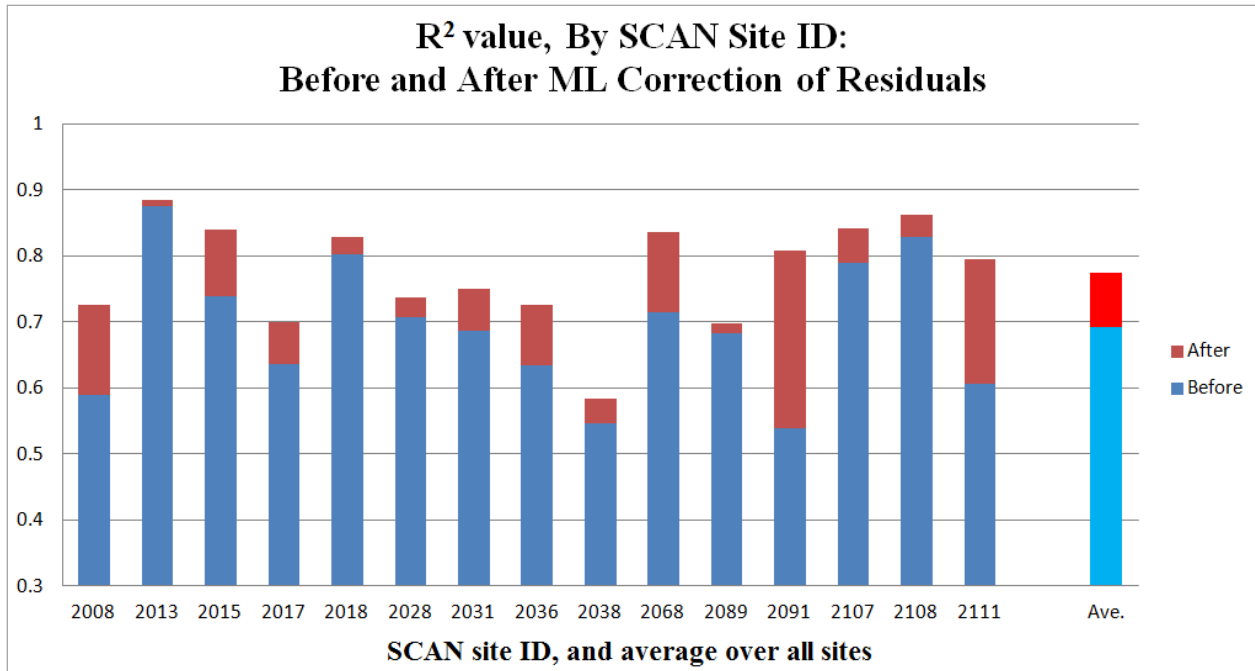
571 Leveraging these findings, the discussion section also presented the results of preliminary
 572 analysis that illustrates how further improvements in soil moisture predictions could be gained
 573 by disaggregating based on local topography. This would enable more accurate predictions at
 574 sites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

578
 579
 580



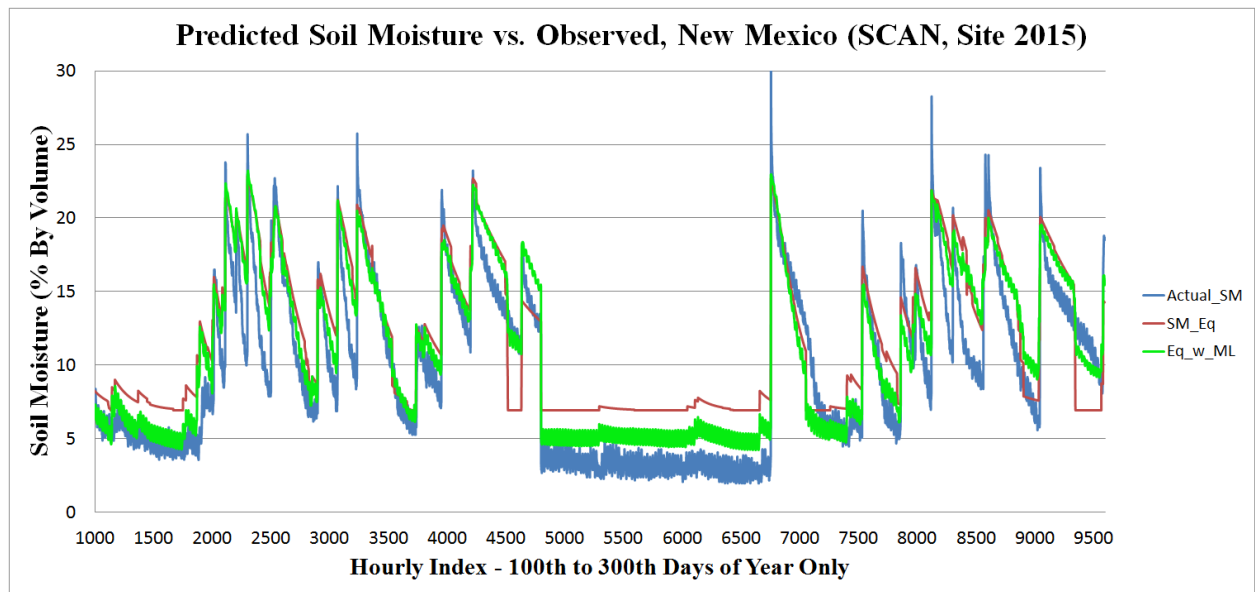
581
 582

Figure 1, Methodological flow chart



583
584

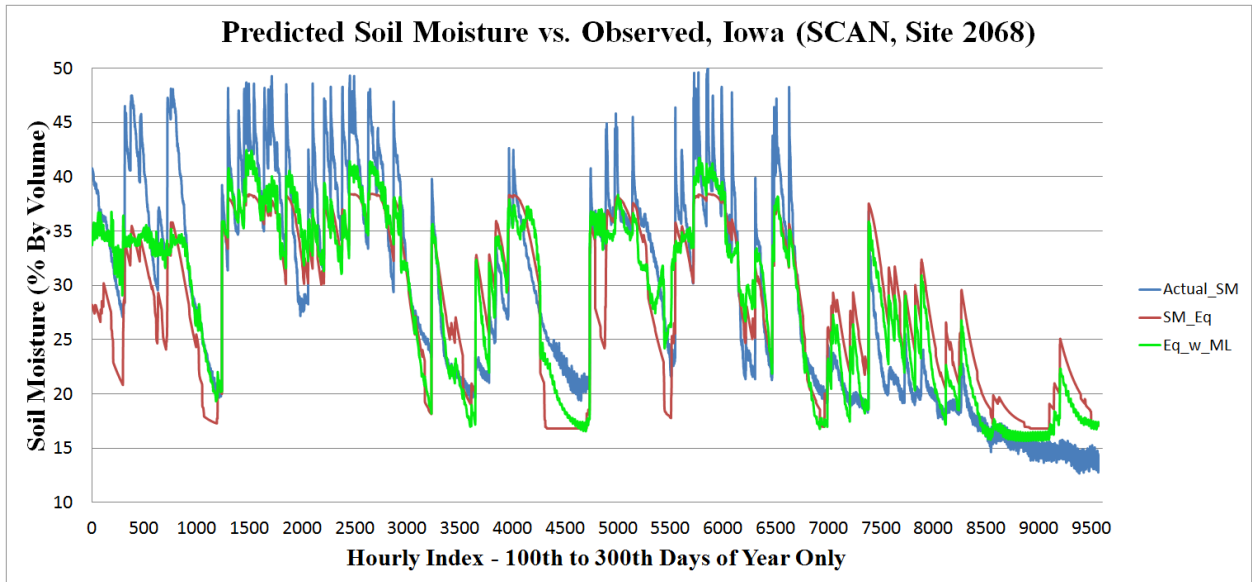
Figure 2, Improvements from machine learning (KNN) models of residuals.



585

**Figure 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).
Hydroclimate: IAQ (Intermediate Seasonality, Arid, Summer Peak Runoff)
Soil Texture: Loamy Sand**

586
587
588
589
590
591
592

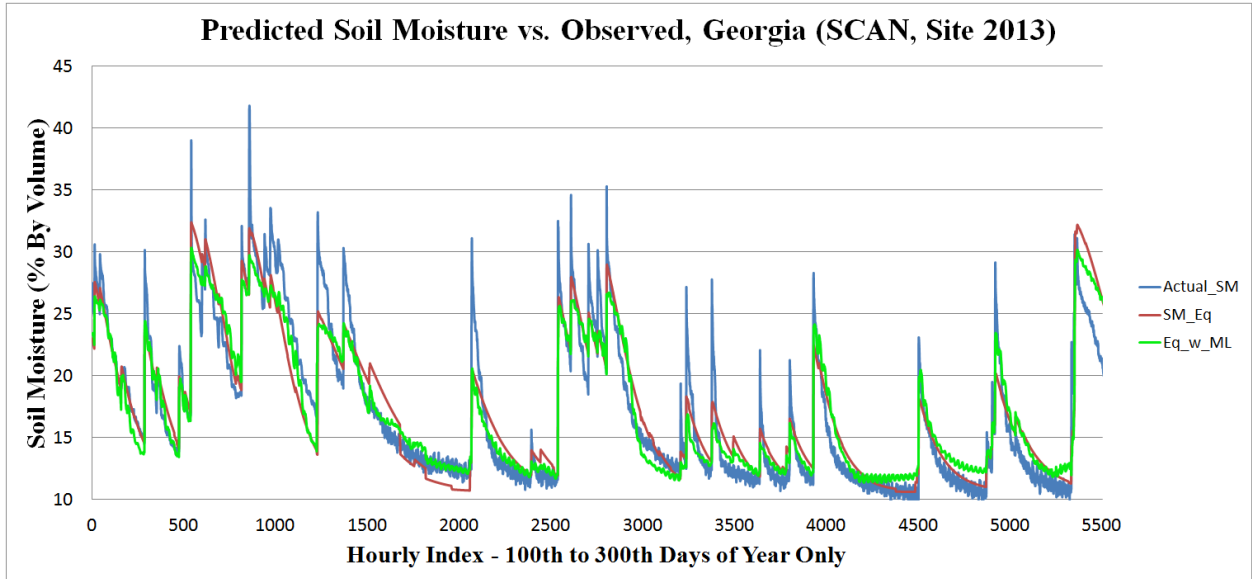


593

594 **Figure 4, SM Time Series, SCAN Site 2068, Iowa (USA), line colors from Fig. 3**
 595 **Hydroclimate: ISCJ (Intermediate Seasonality, Semi-Arid, Winter Peak Runoff, Summer**
 596 **Peak Precipitation)**
 597 **Soil Texture: Silty Clay Loam**

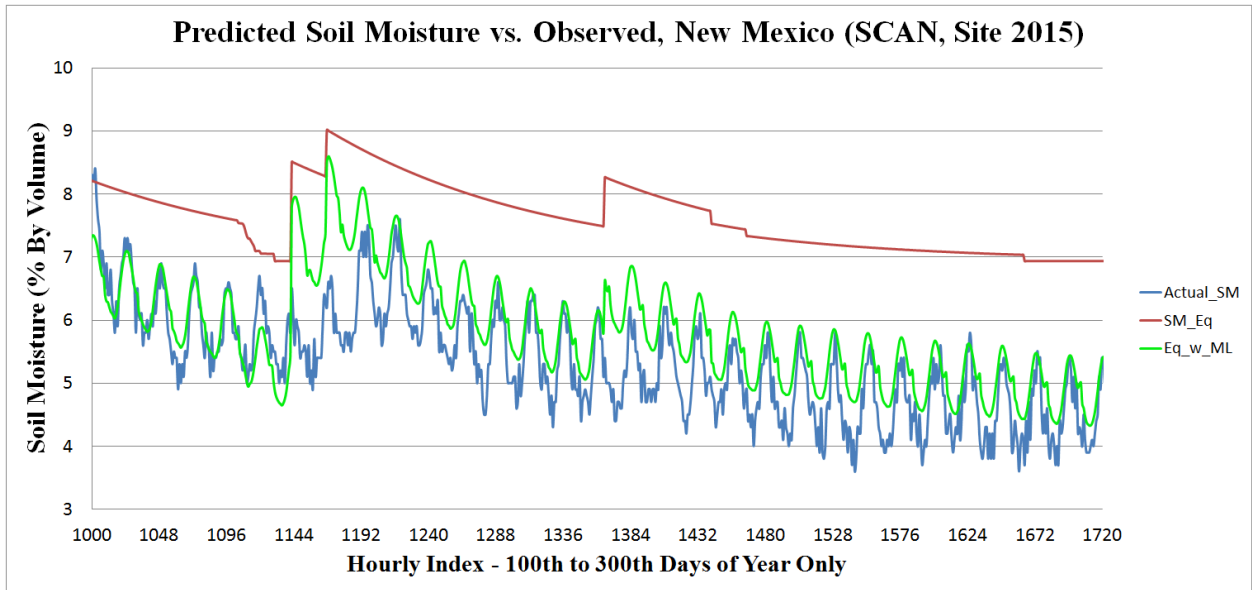
598

599



600

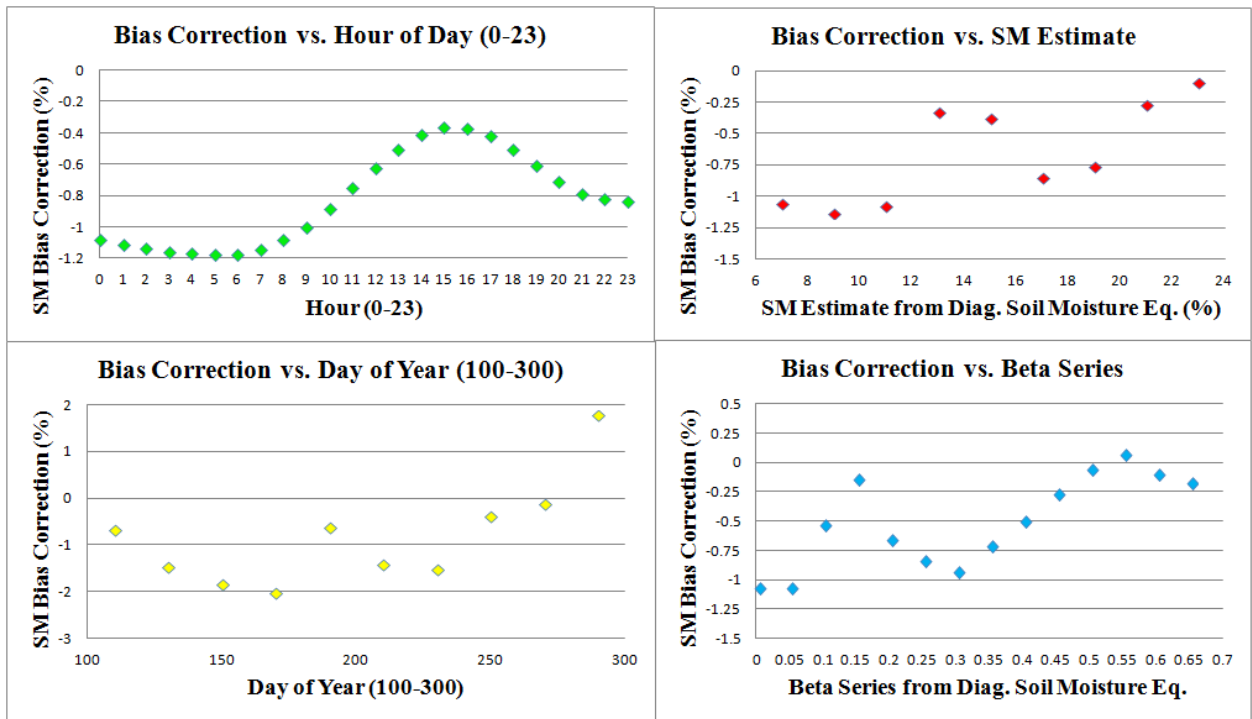
601 **Figure 5, SM Time Series, SCAN Site 2013, Georgia (USA), line colors from Fig. 3**
 602 **Actual Soil Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line),**
 603 **and Diagnostic Soil Moisture Equation with ML Error Correction (Green Line)**
 604 **Hydroclimate: LWC (Low Seasonality, Winter Peak Precipitation, Winter Peak Runoff)**
 605 **Soil Texture: Sandy Loam**



606

607 **Figure 6, Soil Moisture Time Series, SCAN Site 2015, New Mexico (USA), Actual Soil**
 608 **Moisture (Blue Line), Diagnostic Soil Moisture Equation Estimate (Red Line), and**
 609 **Diagnostic Soil Moisture Equation with Machine Learning Error Correction (Green Line)**

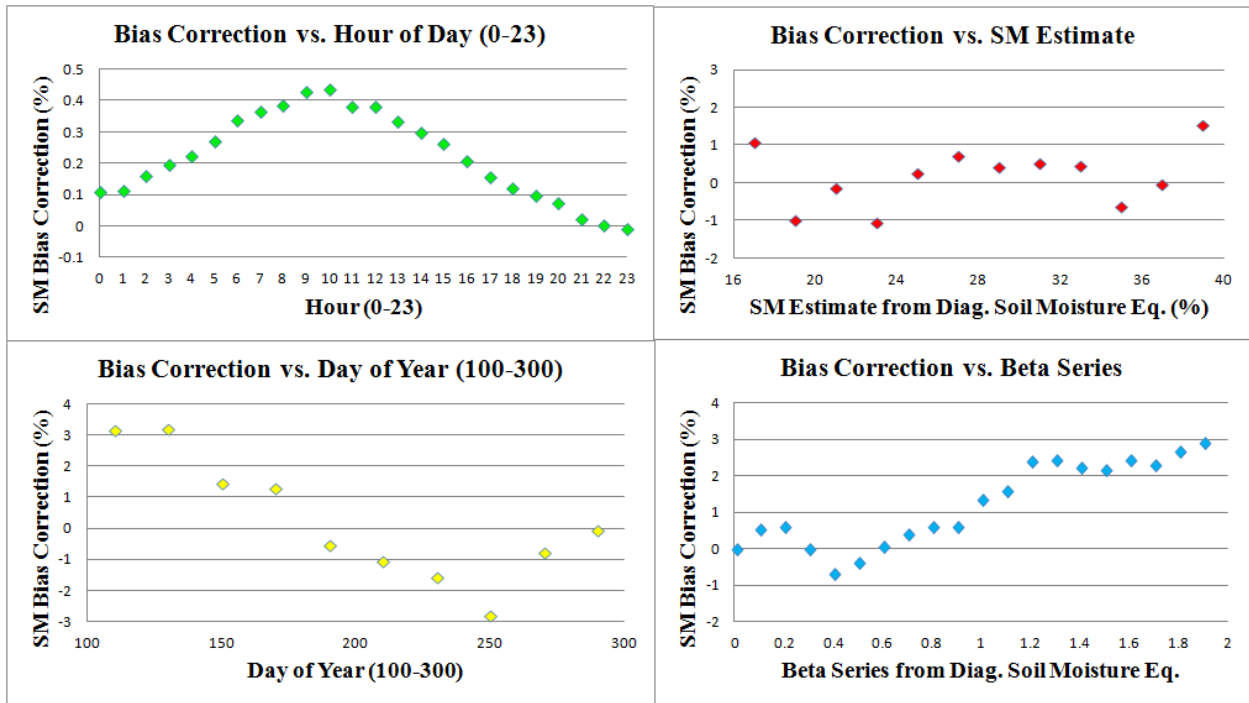
610



611

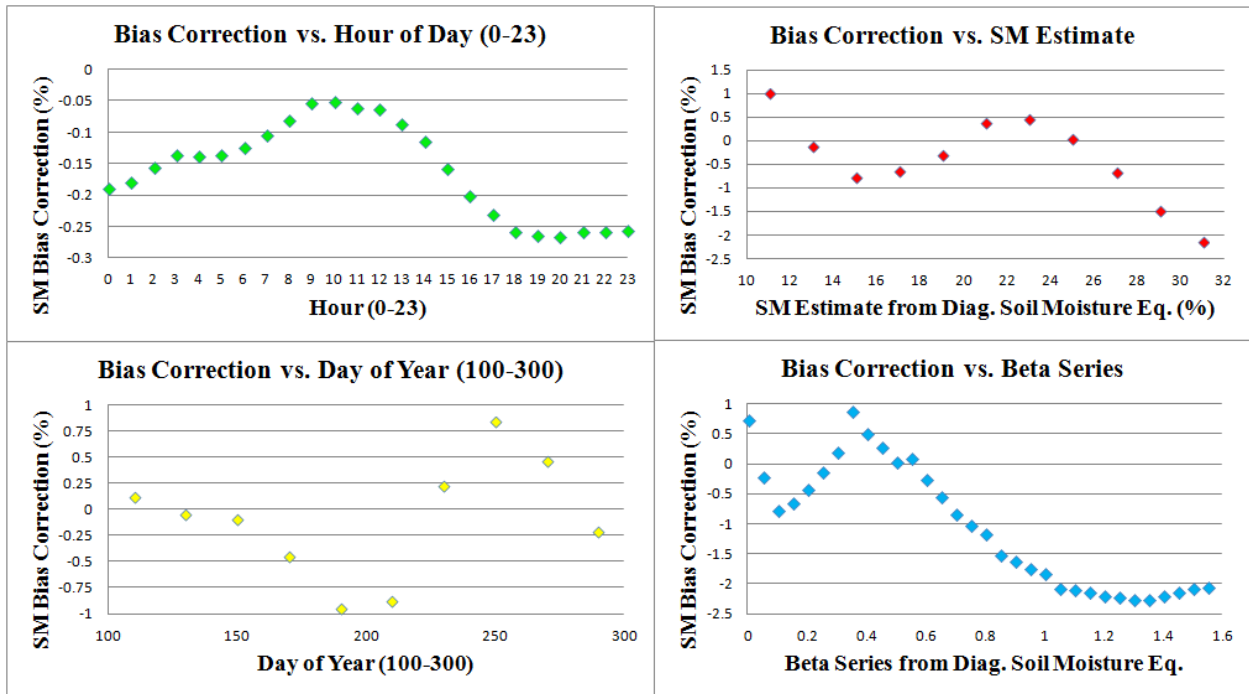
612 **Figure 7, Bias Correction Analysis, SCAN Site 2015 (IAQ, Desert, Loamy Sand)**

613



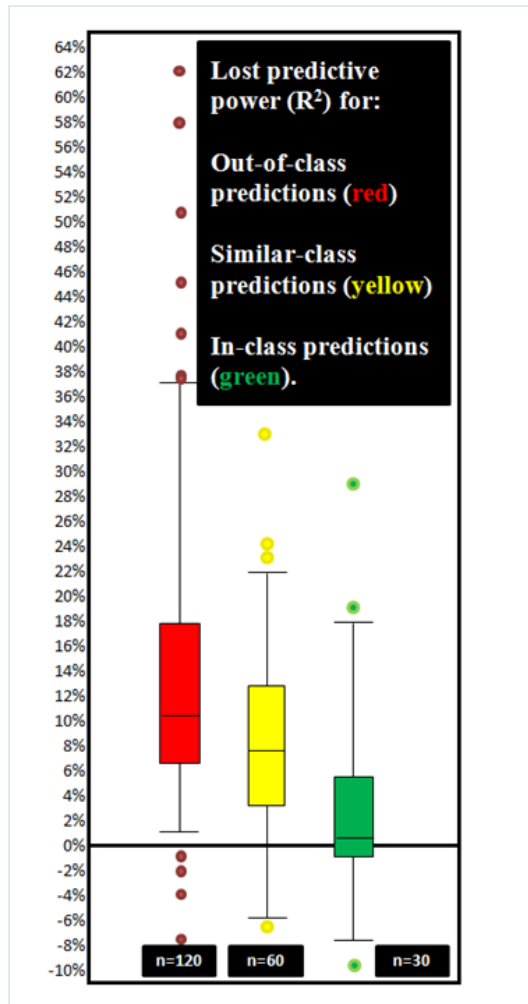
614
615
616

Figure 8, Bias Correction Analysis, SCAN Site 2068 (ISCJ, Plains, Silty Clay Loam)



617
618
619
620
621

Figure 9, Bias Correction Analysis, SCAN Site 2013 (LWC, Woods, Sandy Loam)



622

623

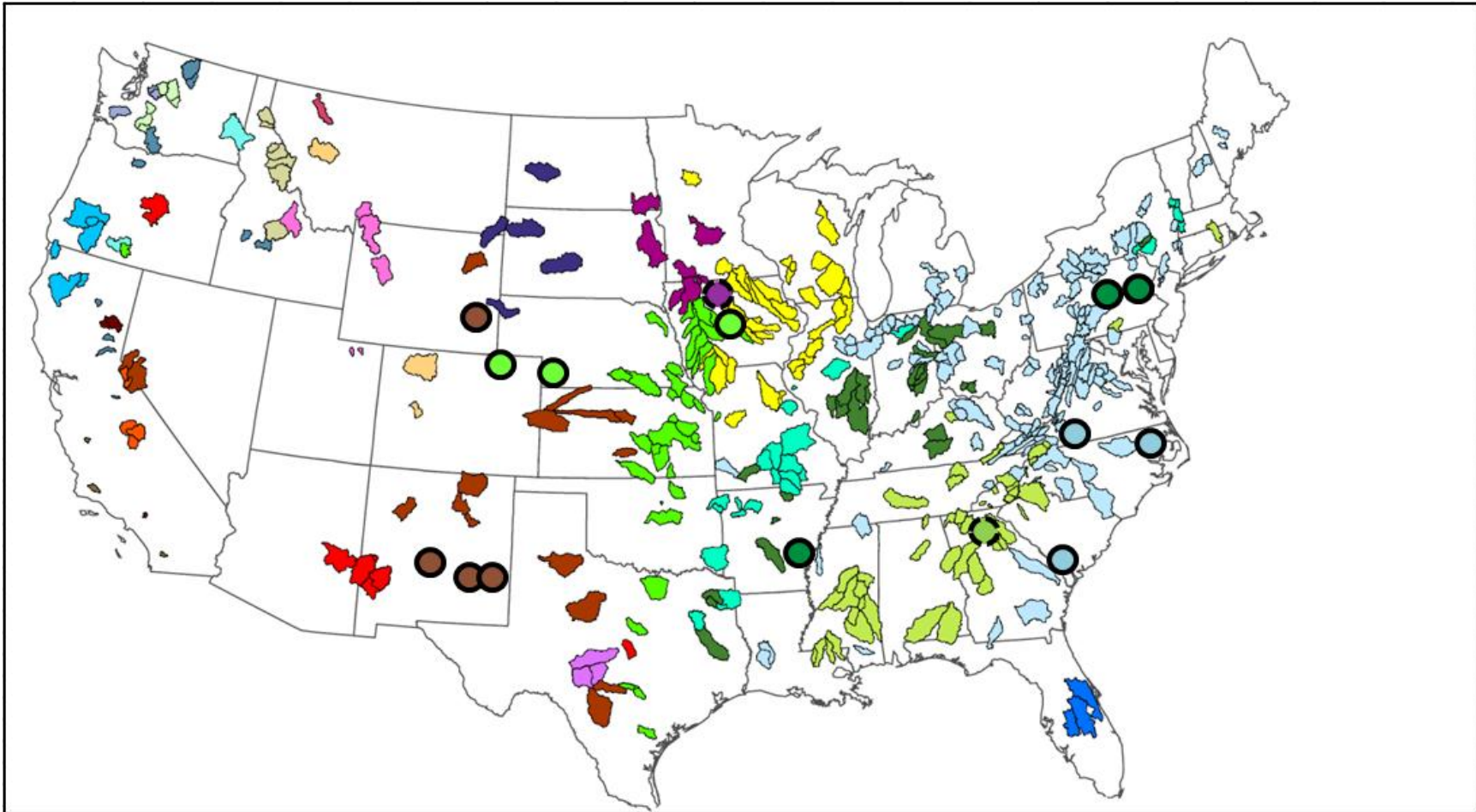
624

625

626

627

Figure 10, 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)



**Figure 11, 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**

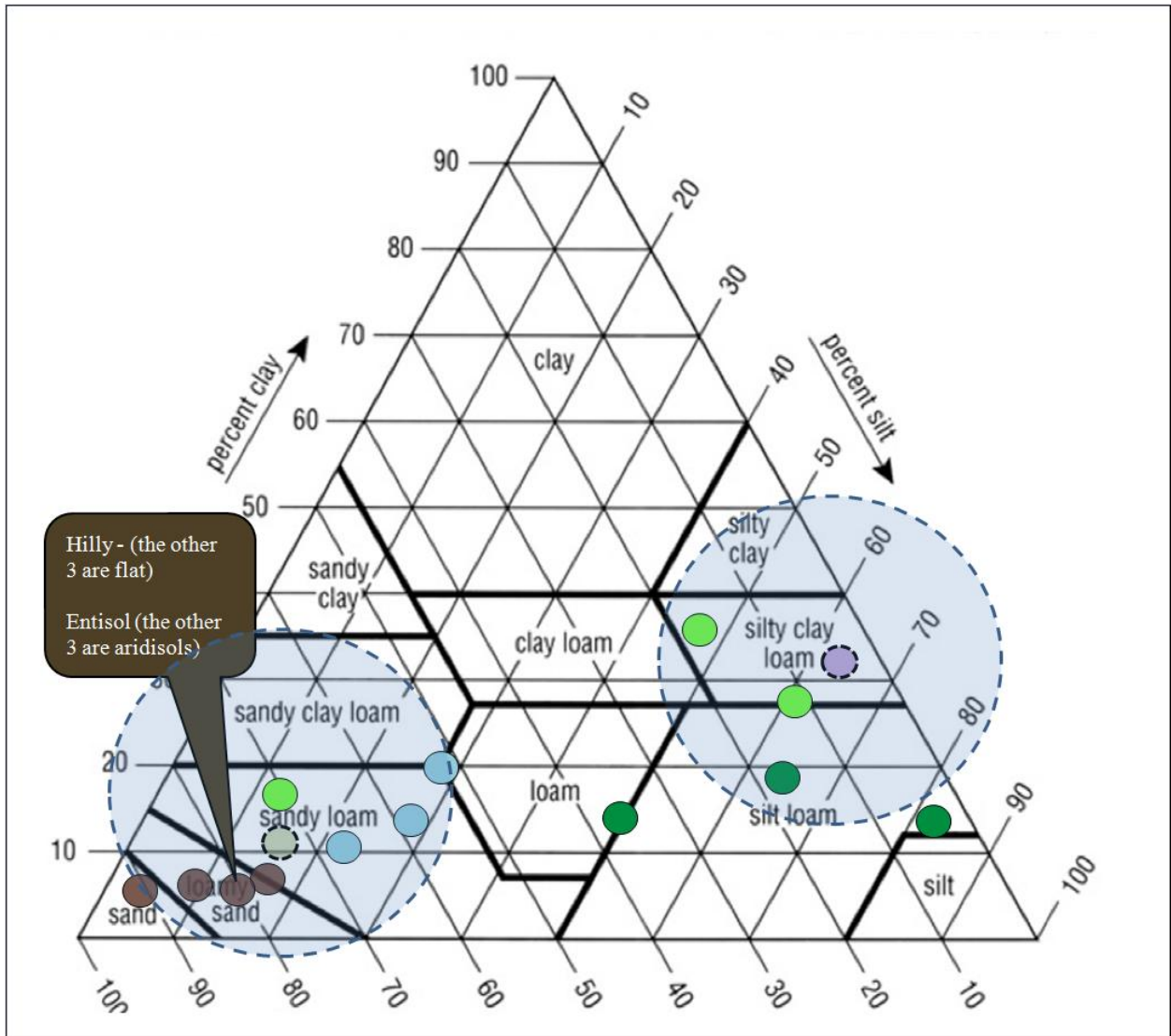


Figure 12, The 15 SCAN sensors, color-coded to match their hydro-climatic class, with similar soil textures shaded.

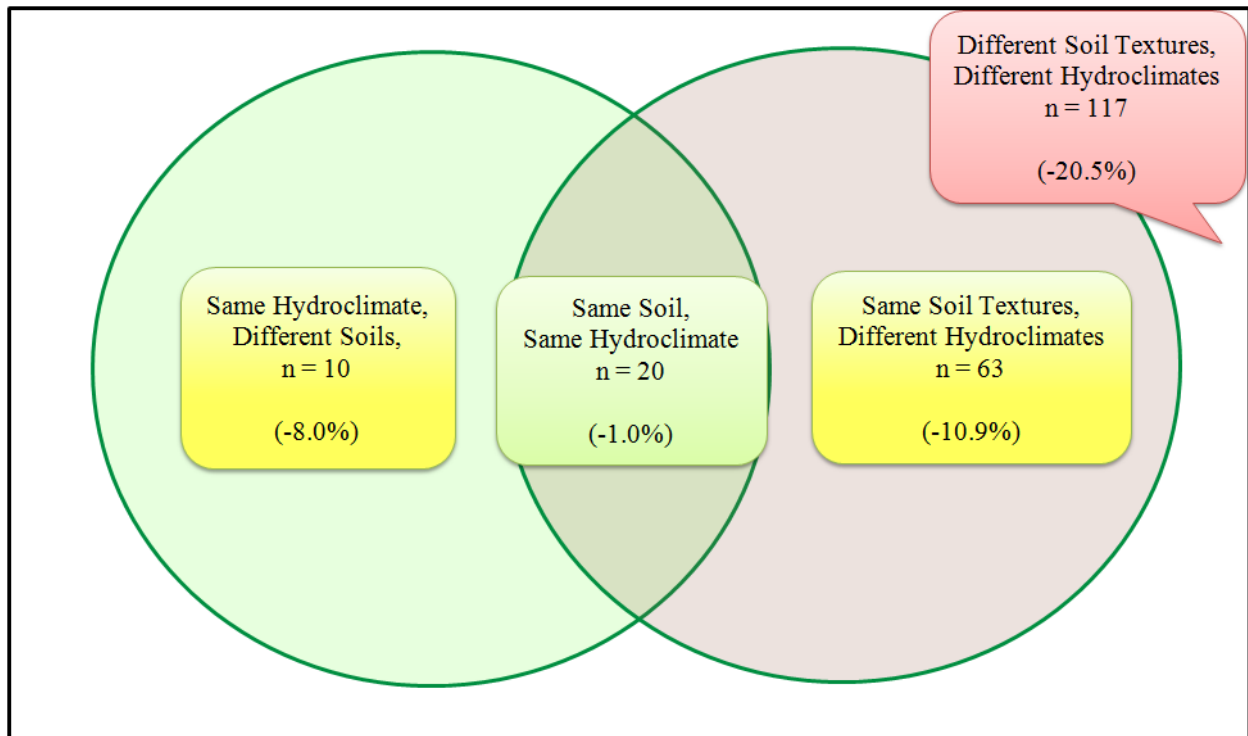


Figure 13 Venn-Diagram of Modeling Errors with Similar and Different Soils and Hydro-climates

SiteID	Hydro-climate	Soil Information	RMSE	RMSE w/ KNN	R ²	R ² w/ KNN
2008	LJ	Sandy Loam	8.38	7.69	0.590	0.726
2013	LWC	Sandy Loam	2.16	2.06	0.876	0.885
2015	IAQ	Loamy Sand	3.29	2.37	0.740	0.841
2017	ISQJ	Sandy Loam	3.62	3.27	0.637	0.701
2018	IAQ	Loamy Sand*	2.23	2.16	0.803	0.828
2028	LPC	Loam	4.89	4.71	0.707	0.738
2031	ISQJ	Silty Clay Loam	5.46	6.00	0.687	0.750
2036	LPC	Silt Loam	4.61	3.95	0.635	0.726
2038	LJ	Sandy Loam	4.81	4.51	0.546	0.584
2068	ISCJ	Silty Clay Loam	5.28	4.03	0.716	0.837
2089	LJ	Sandy Loam	6.7	6.31	0.682	0.697
2091	LPC	Silt	8.12	6.89	0.539	0.808
2107	IAQ	Loamy Sand	1.98	1.85	0.790	0.843
2108	IAQ	Loamy Sand/Sand	1.26	1.12	0.828	0.863
2111	ISQJ	Silty Clay Loam	5.38	5.01	0.607	0.796

*Not similar to other sandy soils, see Figure 9.

Table 1, The Fifteen SCAN Sites: Class & Soil Information and Performance

	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%

Table 2, Cross-Validation Results

References

- Barricelli, N.A. "Numerical testing of evolution theories. Part II. Preliminary tests of performance, symbiogenesis and terrestrial life". *Acta Biotheoretica* (16): 99–126. 1963.
- 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." *Journal of Hydrology*. Vol. 316, Issue 1-4, pp. 129-140, 2006.
- 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." *American Society of Agricultural and Biological Engineers Annual International Meeting 2009*. Volume 4, 2009, Pages 2629-2646.
- Choudhury, B. J. and Blanchard, B.J., "Simulating soil water recession coefficients for agricultural watersheds," *Water Resour. Bull.*, 19, 241–247, 1983.
- Coopersmith E.J., Minsker, B.S., Wenzel, C.E., and Gilmore, B.J., "Machine learning assessments of soil drying for agricultural planning." *Computers and Electronics in Agriculture*. Vol. 104, pp. 93-104, Jun 2014. DOI: 10.1016/j.compag.2014.04.004.
- Coopersmith, E., Yaeger, M., Ye, S., Cheng, L., and Sivapalan, M. "Exploring the physical controls of regional patterns of flow duration curves – Part 3: A catchment classification system based on regime curve indicators." *Hydrol. Earth Syst. Sci.* doi:10.5194/hess-16-1-2012.
- 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(4), 371– 376, 1985.

Fix, E., 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." Geotechnical Special Publication, Issue 192, 2009, Pages 89-94.

Goldberg, D.E. Genetic Algorithms in Search, Optimization, and Machine Learning. Addison-Wesley Professional. 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., Leroux, D., "Validation of soil moisture and ocean salinity (SMOS) soil moisture over watershed networks in the U.S." IEEE Transactions on Geoscience and Remote Sensing, Volume 50, Issue 5, Part 1, 1530-1543. May 2012.

Jones, H. G. (2004). "Irrigation scheduling: advantages and pitfalls of plant-based methods." *Journal of Experimental Botany*, 55, 2427-2436

Koster, R.D. and Mahanama, S.P.P. 2012: "Land Surface Controls on Hydroclimatic Means and Variability." *J. Hydrometeor*, 13, 1604–1620. 2012. doi: <http://dx.doi.org/10.1175/JHM-D-12-050.1>

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. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20110015242_2011016052.pdf (Retrieved 14 May 2013.)

Orth, R.A , Koster, R.D.B, Seneviratne, S.I.A "Inferring soil moisture memory from streamflow observations using a simple water balance model." *J. Hydrometeor*. Vol.14, Iss. 6, 1773-1790, 2013.

Pan, F., "Estimating daily surface soil moisture using a daily diagnostic soil moisture equation. *Journal of Irrigation and Drainage Engineering*, 138(7), 625-631, 2012.

Pan, F., and Peters-Lidard, C. D. (2008). "On the relationship between the mean and variance of soil moisture fields." *Journal of the American Water Resources Association*. 44(1), 235-242.

Pan, F., Peters-Lidard, C. D., and Sale, M. J. (2003). "An analytical method for predicting surface soil moisture from rainfall observations." *Water Resources Research*, 39(11), Art.

Rosenbaum, U., J.A. Huisman, J. Vrba, H. Vereecken and H.R. Bogaen. 2011. Correction of temperature and electrical conductivity effects on dielectric permittivity measurements with ECH2O Sensors. *Vadose Zone J.* 10: 582-593. doi:10.2136/vzj2010.0083. 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.

Sharifat, K., and Kushwaha, R.L., "Sinkage simulation model for vehicles on soft soil." 2000 ASAE Annual International Meeting, Technical Papers: Engineering Solutions for a New Century 1, pp. 2549-2553

Simunek J., Sejna, M., van Genuchten, M. "The HYDRUS-1D software package for simulating water flow and solute transport in two-dimensional variably saturated media. Version 2.0, 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 Resources Research*, 44, W02404, doi:10.1029/2005WR004436.

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 and Tillage Research* 103 (2), pp. 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, 1960, pp. 187, 287

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." *Journal of Hydrology*, 374, pp. 307-317 (2009).