UC Merced

UC Merced Previously Published Works

Title

Machine Learning Based Soil Moisture Retrieval from Unmanned Aircraft System Multispectral Remote Sensing

Permalink

https://escholarship.org/uc/item/0405g35t

ISBN

9781728163741

Authors

Araya, Samuel N
Fryjoff-Hung, Anna
Anderson, Andreas
et al.

Publication Date

2020-10-02

DOI

10.1109/igarss39084.2020.9324117

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

MACHINE LEARNING BASED SOIL MOISTURE RETRIEVAL FROM UNMANNED AIRCRAFT SYSTEM MULTISPECTRAL REMOTE SENSING

Samuel N. Araya[1], Anna Fryjoff-Hung[2], Andreas Anderson[2], Joshua H. Viers[2,3], and Teamrat A. Ghezzehei[2,4]

[1]Earth System Science, Stanford University, Stanford, CA, USA, araya@stanford.edu
[2]Center for Information Technology in the Interest of Society and the Banatao Institute, University of California, Merced, CA, USA
[3]Department of Civil & Environmental Engineering, University of California, Merced, CA, USA
[4]Life and Environmental Science, University of California, Merced, CA, USA

IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9324117

ABSTRACT

We developed machine learning models to retrieve surface soil moisture (0-4 cm) from high resolution multispectral imagery, terrain attributes, and local climate covariates. Using a small unmanned aircraft system (UAS) equipped with a multispectral sensor we captured high resolution imagery in part to create a high-resolution digital elevation model (DEM) as well as quantify relative vegetation photosynthetic status. We tested four different machine learning algorithms. The boosted regression tree algorithm provided the best accuracy model with mean absolute error of 3.8 % volumetric water content. The most important variables for the prediction of soil moisture were precipitation, reflectance in the red wavelengths, potential evapotranspiration, and topographic position indices (TPI). Our results demonstrate that the dynamics of soil water status across heterogeneous terrain may be adequately described and predicted by UAS remote sensing data and machine learning. Our modeling approach and the variable importance and relationships we have assessed in this study should be useful for management and environmental modeling tasks where spatially explicit soil moisture information is important.

Index Terms— Machine learning, boosted regression tree, multispectral, unmanned aerial vehicle, remote sensing, soil moisture, digital elevation model

1. INTRODUCTION

The relatively small quantity of water stored in the upper layers of soil plays a key role in terrestrial biology, biogeochemistry, and atmospheric water and energy fluxes [1], [2]. Remote sensing methods of retrieving soil moisture provide spatially distributed and frequent observations over a large area, which is difficult to achieve using conventional field measurements [3].

Soil moisture signals exist in the visible and near infrared regions of multispectral images, specially when the soil is bare or with moderate canopy cover [4], [5]. The interpretation of soil moisture from multispectral remote sensing is, however, difficult [6]. Reflectance in multispectral bands are primarily the result of soil and vegetation which have a very complex relationship with soil moisture. Furthermore, the temporal dynamics and spatial distribution of soil moisture (and associated changes in vegetation and reflectance) is influenced by soil and topographic variables [7], [8] as well as meteorological variables (precipitation and evapotranspiration in particular). Estimating the complex and interdependent relationship of soil moisture with the observed reflectance and multiple surface and climate variables using conventional statistical or physical-based models is extremely difficult. This makes the problem interpreting soil moisture from such observations a good candidate for machine learning techniques.

Machine learning techniques are able to learn and approximate complex non-linear relationships. These techniques also do not require assumptions on data distribution and they can be used to integrate data from different sources with poorly-defined or unknown probability density functions [9]. Machine learning techniques have often been shown to outperform parametric approaches [9], [10].

Remote sensing from unmanned aircraft systems (UAS) has the potential to address several limitations of traditional remote sensing. The most attractive feature of UAS is their high spatial resolution, frequent or on-demand image acquisition, and low operating costs [11]–[14].

The purpose of this research is to advance soil moisture change measurement, process understanding, and prediction using remote sensing products from UAS and machine learning methods.

The specific goals of this study are to: (1) develop an adaptable method to retrieve information on surface soil moisture from small UAS remote sensing products and machine learning techniques, (2) identify appropriate spatial resolutions of reflectance images and terrain variables for estimating soil moisture, and (3) explore the nature of the relationship between soil moisture and surface properties.

2. METHODS

Using a small, fixed-wing UAS (Finwing Sabre, Finwing Technology) equipped with multispectral camera (Parrot Sequoia Sensor, Paris, France), we collected high-resolution (6.85 cm ground resolution) multispectral images. Images were collected on six different days spread across the 2018

Authorized licensed use limited to: Univ of Calif Merced. Downloaded on April 20,2022 at 17:02:53 UTC from IEEE Xplore. Restrictions apply.

water year to capture the landscape soil moisture dynamics of the study site. Images were mosaicked, orthorectified, and radiometrically calibrated using Pix4D photogrammetry software (Pix4D, Lausanne, Switzerland). Digital elevation model (DEM) data were generated from stereo-images photogrammetrically using Pix4D software and structure-from-motion techniques.

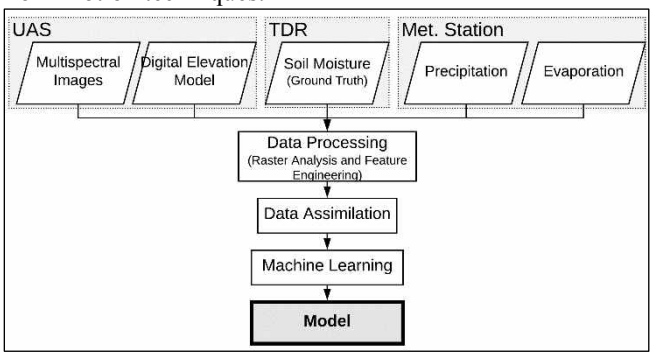

Figure 1 Process flowchart of model development. Three data sources: UAS-borne multispectral imagery, in-situ soil moisture from time-domain reflectometry (TDR) and variables from meteorological station are used to predict soil moisture.

Concurrently with the image acquisition flights, in-situ top 4-cm soil moisture content was measured using a time-domain reflectometry (TDR) probe (FieldScout TDR-300, Spectrum Technologies Inc., IL, USA). Six soil sampling transects were identified and their geographic coordinate precisely measured with real-time kinematic (RTK) positioning survey. Approximately 10 in-situ soil moisture measurement 5 meters apart were then taken across each transect. A total of 406 soil moisture measurements were collected across the six measurement times.

The study site was a small grassland catchment (0.6 km^2 area) located at the Merced Vernal Pools and Grassland Reserve, California. The study site has a Mediterranean climate with hot, dry summers and cool, wet winters with an average annual precipitation of 330 mm. The dominant soils of the area are Redding gravelly loam (Fine, mixed, active, thermic Abruptic Durixeralfs).

The in-situ soil moisture measurements, multispectral reflectance images, terrain variables derived from DEM, rainfall, and potential evapotranspiration (ET) data were then aggregated into a data table and used to train machine learning models to predict the soil moisture. Figure 1 illustrates the model building process.

Several variables were calculated based on the multispectral reflectance, terrain, and meteorological data to be used to train a machine learning model as predictor variables. Several topographic variables derived from DEM, such as slope, curvature, etc., are scale dependent. To identify appropriate scale for the study, we calculated all topographic variables on six different resolution DEM. For this, we first upscaled the DEM from the original resolution of 6.85 cm to 15, 30, 60, 100, 300, and 500 cm cell resolution. We then calculated topographic variables on all the resolutions. The calculation of topographic position index (TPI) does not only depend on DEM resolution but on the definition of inner and out radii of the annulus [15], we thus calculated TPI for multiple different inner and outer radii sizes.

We employed three methods of variable selection: test of linear correlation and linear dependencies among variables, and recursive feature elimination [16], [17]. The final data used for building the models had 46 variables, of which five are meteoric, nine are reflectance variables, and 32 are topographic variables.

3. MACHINE LEARNING

We compared five popular machine learning algorithms: artificial neural network (ANN), support vector regression (SVR), relevance vector regression (RVR), random forest (RF) and boosted regression trees (BRT).

The overall procedure of building the machine learning models is illustrated in Figure 2. The computationally demanding steps of model training and testing were run at a high-performance computing cluster.

The data was split into training and testing sets of approximately 75-25 percent, respectively. The testing set was a hold-out set used only to evaluate final trained models. The split was done by randomly selecting sampling transects for the different dates. To test if there was bias due to the training-testing set split, we generated 30 unique training-testing set splits upon which we trained 30 separate models based on each separate training set. The performance of each model was also assessed on its respective testing set.

The model hyperparameter tuning was done by a 30-fold cross-validation set generated by splitting the training data into 80-20 percent training-validation split by randomly selecting a transect for every day. Optimum model parameters were selected using a comprehensive grid search method[18]. The final performance of models was assessed on the separate hold-out test dataset. The performance of models is measured in terms of mean absolute error (MAE), mean bias error (MBE) and the coefficient of determination (R^2).

The statistical significance of each predictor variable with respect to its effect on the generated model (variable importance) was calculated internally the tree-based models, RF and BRT, and using recursive feature elimination method for the rest of the model algorithms. The relationship between the predictor variables and outputs was analyzed using model-independent accumulated local effects (ALE) plots [19], [20].

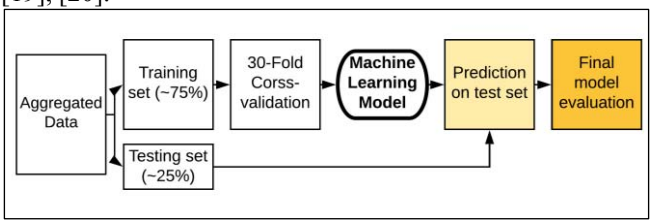

Figure 2 Flowchart showing the model training process.

Authorized licensed use limited to: Univ of Calif Merced. Downloaded on April 20,2022 at 17:02:53 UTC from IEEE Xplore. Restrictions apply.

4. RESULTS

4.1 Model Performance

We tested five machine learning algorithms: ANN, SVR, RVR, RF and BRT. The BRT algorithm models had the best performance with MAE 3.8 followed closely by RF (MAE 3.9). The MAE for RVR, SVR, and ANN models was 4.3, 4.4 and 4.5, respectively. Figure 3 shows the measured water contents from the testing data sets against the prediction by all the BRT models trained with the 30 distinct training-testing sets.

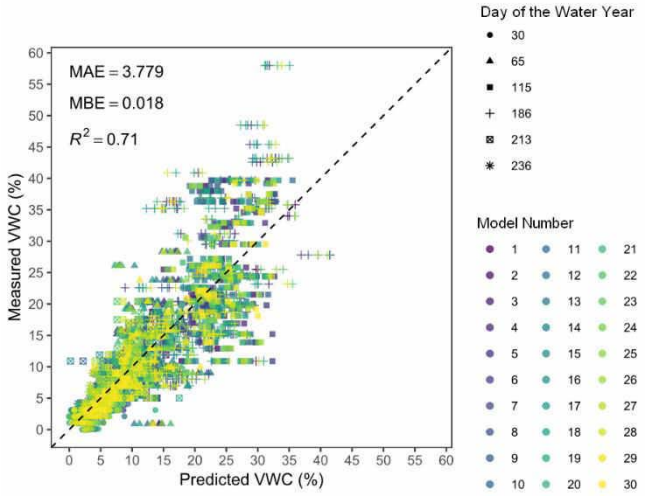

Figure 3 Scatter of the measured versus predicted soil water content of the testing sets around 1:1 line for the best BRT model. MAE, MBE, and R^2 are averaged across the 30 models.

Prediction of soil volumetric water content from the best BRT model for a portion of the study area for January 23, 2018, is shown in Figure 4.

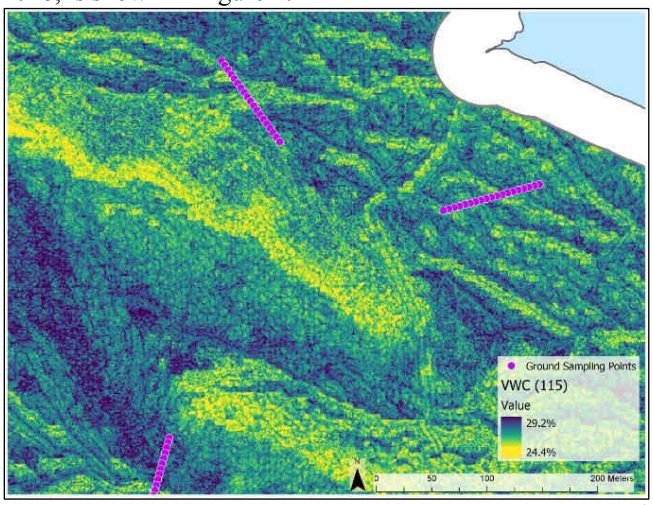

Figure 4 Predicted volumetric soil water content (%) map for 115th day of the water year (January 23, 2018).

4.2 Predictor Variable Importance

The relative importance of predictors from the BRT model grouped by variable type are shown in Figure 5. The meteoric and reflectance variables are the only temporally dynamic variables; topographic variables are not considered time dependent. The four most important variables are precipitation, reflectance in the red band, ET, and TPI. TPI and flow accumulation are the most important of the topographic variables.

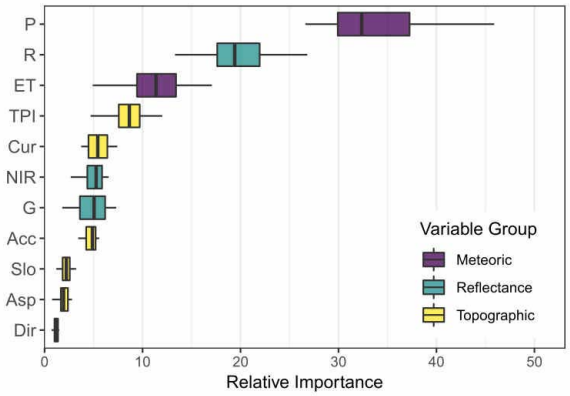

Figure 5 Sum of the relative variable importance distribution of the 30 models grouped by variable type: precipitation (P), red band (R), evapotranspiration (ET), topographic position index (TPI), curvature (Cur), near infrared band (NIR), green mand (G), flow accumulation (Acc), slope (Slo), aspect (Asp), and direction (Dir).

4.3 Effect of Predictor Variables

We used the ALE plots to investigate the effects all the predictor variables. Topography has a strong control on soil moisture distribution at landscape scales [21] and our models indicated the relatively higher importance of some of the topographic variables. We investigated how these topographic variables relate to soil moisture. Figure 6 shows the ALE plots for TPI, curvature, and flow accumulation variables of selected resolutions on soil moisture estimates.

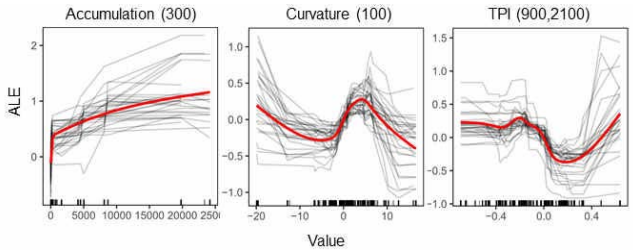

Figure 6 ALE plots for flow accumulation, profile curvature, and TPI variables, numbers in parentheses indicate DEM scale in cm. TPI scale is a combination of the inner-outer diameters. Black curves represent individual effects of the 30 models, and thick, red curves are smoothed trendlines overall individual models. Marks along the x-axis show the distribution of data in the model training set.

Predicted soil moisture generally increased with flow accumulation. Soil moisture initially decreased as surface become less convex and increased as surface curvature transitioned from convex to concave (negative to positive values). TPI is particularly scale-dependent but still showed a similar relationship pattern with soil moisture at all scales. Negative TPI values indicate trends towards valleys, zero values indicate either flat (if slope is shallow) or mid-slope

Authorized licensed use limited to: Univ of Calif Merced. Downloaded on April 20,2022 at 17:02:53 UTC from IEEE Xplore. Restrictions apply.

(if slope is significant) areas, and positive TPI values indicate trends towards ridgetops [15]. Across all scales, there was a negative relationship between TPI and soil moisture at negative values and a positive relationship at positive TPI values; this result indicates valleys and ridge tops were wetter than mid-slope areas.

5. CONCLUSION

This research serves as a proof of concept: surface soil moisture can be interpreted with reasonable accuracy from multispectral UAS remote sensing using machine learning methods. Although the machine learning models used are considered non-spatial models (i.e., they do not consider sampling location information and spatial autocorrelation [22], [23]), the inclusion of spatially dependent variables such as curvature, flow accumulation, and TPI means that the models do account for spatial context and thus should make the predictions more spatially dependent.

As a data mining technique, machine learning model performance and reliability are closely tied to the quantity, and quality of data used in parameterization. Additional ground measurement data should improve future model outcomes.

6. ACKNOWLEDGMENTS

This work was made possible by the support of U.S. Fish and Wildlife Service Agreement #P1740401 as administered by the California Department of Fish and Wildlife; Monique Kolster and the UC Merced Vernal Pools and Grassland Reserve of the Natural Reserve System; and Francesca Cannizzo of the UC Merced Physical and Environmental Planning team. AFH was partially supported by the UC Water Security and Sustainability Research Initiative (UC Office of the President's Multi-Campus Research Programs and Initiatives #MR-15-328473). We gratefully acknowledge computing time on the Multi-Environment Computer for Exploration and Discovery (MERCED) cluster at UC Merced, which was funded by National Science Foundation Grant No. ACI-1429783.

7. REFERENCES

[1] K. E. Trenberth, J. T. Fasullo, and J. Kiehl, "Earth's Global Energy Budget," *Bull. Am. Meteorol. Soc.*, vol. 90, no. 3, pp. 311–323, Mar. 2009.

[2] S. I. Seneviratne *et al.*, "Investigating soil moisture-climate interactions in a changing climate: A review," *Earth-Science Rev.*, vol. 99, no. 3–4, pp. 125–161, 2010.

[3] G. P. Petropoulos, G. Ireland, and B. W. Barrett, "Surface soil moisture retrievals from remote sensing: Current status, products & future trends," *Phys. Chem. Earth*, vol. 83–84, pp. 36–56, 2015.

[4] J. Yuan, X. Wang, C. Yan, S. Wang, X. Ju, and Y. Li, "Soil Moisture Retrieval Model for Remote Sensing Using Reflected Hyperspectral Information," *Remote Sens.*, vol. 11, no. 3, p. 366, Feb. 2019.

[5] D. B. Lobell and G. P. Asner, "Moisture Effects on Soil Reflectance," *Soil Sci. Soc. Am. J.*, vol. 66, no. 3, p. 722, 2002.

[6] B. Zaman, M. McKee, and C. M. U. Neale, "Fusion of remotely sensed data for soil moisture estimation using relevance vector and support vector machines," *Int. J. Remote Sens.*, vol. 33, no. 20, pp. 6516–6552, 2012.

[7] P. Jiang and K. D. Thelen, "Effect of Soil and Topographic Properties on Crop Yield in a North-Central Corn–Soybean Cropping System," *Agron. J.*, vol. 96, no. 1, p. 252, 2004.

[8] C. W. Bobryk *et al.*, "Validating a digital soil map with corn yield data for precision agriculture decision support," *Agron. J.*, vol. 108, no. 3, pp. 957–965, 2016.

[9] I. Ali, F. Greifeneder, J. Stamenkovic, M. Neumann, and C. Notarnicola, "Review of Machine Learning Approaches for Biomass and Soil Moisture Retrievals from Remote Sensing Data," *Remote Sens.*, vol. 7, no. 12, pp. 16398–16421, Dec. 2015.

[10] S. Paloscia, P. Pampaloni, S. Pettinato, and E. Santi, "A Comparison of Algorithms for Retrieving Soil Moisture from ENVISAT/ASAR Images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 10, pp. 3274–3284, 2008.

[11] M. Elarab, "The Application of Unmanned Aerial Vehicle to Precision Agriculture: Chlorophyll, Nitrogen, and Evapotranspiration Estimation," Utah State University, 2016.

[12] J. A. J. Berni, P. J. Zarco-Tejada, L. Suárez, V. González-Dugo, and E. Fereres, "Remote sensing of vegetation from UAV platforms using lightweight multispectral and thermal imaging sensors," *Int. Arch. Photogramm. Remote Sens. Spat. Inform. Sci*, vol. 38, p. 6 pp, 2009.

[13] I. Colomina and P. Molina, "Unmanned aerial systems for photogrammetry and remote sensing: A review," *ISPRS J. Photogramm. Remote Sens.*, vol. 92, pp. 79–97, Jun. 2014.

[14] K. Anderson and K. J. Gaston, "Lightweight unmanned aerial vehicles will revolutionize spatial ecology," *Front. Ecol. Environ.*, vol. 11, no. 3, pp. 138–146, Apr. 2013.

[15] J. Jenness, B. Brost, and P. Beier, "Land Facet Corridor Designer," no. July. 2013.

[16] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Mach. Learn.*, pp. 389–422, 2002.

[17] X. Chen and J. C. Jeong, "Enhanced Recursive Feature Elimination," *Proc. - 6th Int. Conf. Mach. Learn. Appl. ICMLA 2007*, pp. 330–335, 2007.

[18] S. Chan and P. Treleaven, "Continuous Model Selection for Large-Scale Recommender Systems," in *Handbook of Statistics*, vol. 33, Elsevier, 2015, pp. 107–124.

[19] D. W. Apley, "Visualizing the Effects of Predictor Variables in Black Box Supervised Learning Models," 2016.

[20] B. M. Greenwell, "pdp: An R Package for Constructing Partial Dependence Plots," *R J.*, vol. 9, no. 1, 2017.

[21] R. Sørensen, U. Zinko, and J. Seibert, "On the calculation of the topographic wetness index: Evaluation of different methods based on field observations," *Hydrol. Earth Syst. Sci.*, vol. 10, no. 1, pp. 101–112, 2006.

[22] S. Georganos *et al.*, "Geographical Random Forests: A Spatial Extension of the Random Forest Algorithm to Address Spatial Heterogeneity in Remote Sensing and Population Modelling," *Geocarto Int.*, pp. 1–12, Apr. 2019.

[23] T. Hengl, M. Nussbaum, M. N. Wright, G. B. M. Heuvelink, and B. Gräler, "Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables," *PeerJ*, vol. 6, p. e5518, Aug. 2018.

Authorized licensed use limited to: Univ of Calif Merced. Downloaded on April 20,2022 at 17:02:53 UTC from IEEE Xplore. Restrictions apply.