Using GEE to estimate soil moisture with data augmentation and SVR

Wei Xu (1), Qiming Qin (1), Zhaoxu Zhang (1)

[1] Institute of Remote Sensing and Geographic Information System, Peking University, Beijing, 100871, China

Email: xuwei1995@pku.edu.cn; qmqin@pku.edu.cn; zhaoxuzhang@pku.edu.cn

Abstract: Soil moisture content is an important parameter in hydrological, meteorological, and agricultural applications. At present, soil moisture estimation with Synthetic Aperture Radar (SAR) data shows great potential and promising practice. And some researches use machine learning methods to monitor soil moisture as more and more high-quality remote sensing data are available. However, the data driven feature of these methods brings a new challenge, i.e., every machine learning method needs sufficient in-situ data to be trained. Considering this problem, we proposed a data augmentation method based on alpha approximation method in this paper. Through this method, more training data can be produced, and then these data are provided to train a support vector regression (SVR) machine for soil moisture estimation. It turns out the SVR trained by original data and augmented data shows much better predictive capability than the SVR trained by original data only when using a proper data pretreatment and network, and it also obtains much better results than the original alpha approximation method. By contrast, the root mean squared error (RMSE) of our method decreased to 0.0339 and R2 increased to 0.6491. At last, Google Earth Engine (GEE) is used to collect relevant remote sensing data of Yancheng, Henan province and then soil moisture is estimated with our proposed method efficiently. In conclusion, a new method is proposed in this paper to meet the data driven feature of machine learning methods. And soil moisture of a region can be estimated efficiently by combining our method and the application of GEE.

Keywords: alpha approximation method, GEE, SAR, soil moisture, SVR

1. INTRODUCTION

Soil moisture is an important parameter in hydrological, meteorological and agriculture applications.As a result, it is meaningful to estimate soil moisture accurately and fast. In recent years, many researches have worked on soil moisture inversion with Synthetic Aperture Radar (SAR) data [1]-[3]. And several SAR sensors characterized by short-repeating cycles and high resolutions have been launched, which can provide more open and high-quality data. Therefore, it shows powerful potential and promising practice in estimating soil moisture. However, the result was not very good when the soil was covered by vegetation such as agriculture areas.

Some methods were proposed to improve the estimation, such as polarimetric decomposition method [4] and change detection technique [5]. Aditionally, Balenzano et al. [6] developed alpha approximation method using dense temporal series of Cand L- band SAR data based on change detection technique. This method avoided the impacts of vegetation and roughness of soil surface and resulted in a good estimation. Nevertheless, it needed a lower and an upper bound to solve an underdetermined system of equation while the bounds are usually difficult to obtain and solving the underdetermined system of equations may affect the accuracy of soil moisture retrieval. Some researches about the confirmation of more accurate constraints were conducted [7], [8], but there is little improvement on solving the underdetermined system of equations for its the inherent problem of this method.

This paper avoids finding the constraints or solving the underdetermined system of equation by applying alpha approximation method to augment observed data rather than retrieve soil moisture from these equations directly. Afterwards, a support vector regression (SVR) machine [9] was trained with these extended data. And soil moisture can be estimated with the SVR subsequently. Finally, relevant remote sensing data of Yancheng, Henan province was collected with Google Earth Engine (GEE) and soil moisture of this area was estimated with our proposed method.

2. STUDY AREA AND DATASET

2.1 In-situ data

Two field measurements were conducted. One was from April 6 to 8, 2018 and the other was from May 31 to June 1, 2018. The winter wheat was at jointing stage during the first measurement. While during the second measurement, the winter wheat was at maturity stage, and some has been harvested. In the first measurement, 34 samples were selected in farmland planted with winter wheat. The second measurement comprised 23 samples. The distribution of these samples is illustrated in Fig. 1.

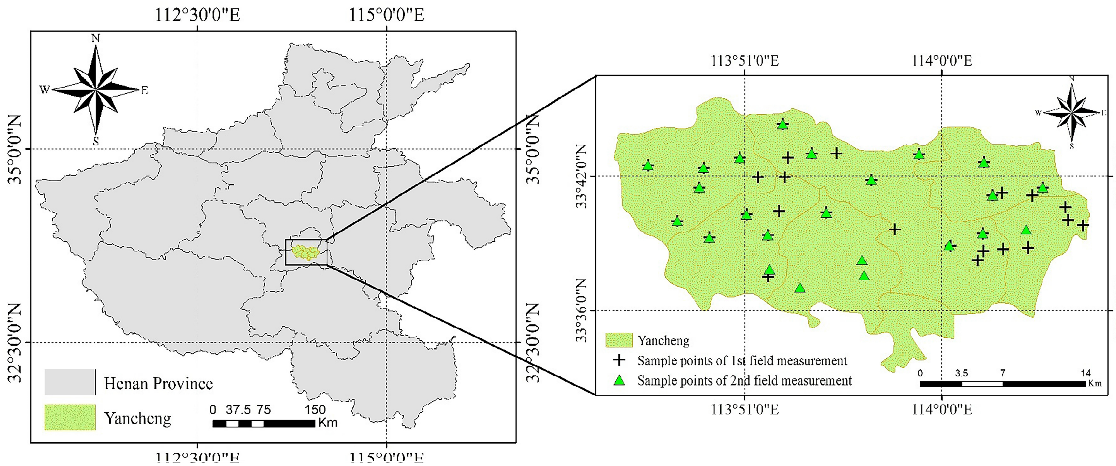

Fig.1. Distribution of selected samples in two field measurements.

2.2 Sentinel-1 data

Four scenes on April 7, 19, May 1, and June 1, 2018 of Sentinel-1 images were used in this study. The images were in Interferometric Wide (IW) swath mode and belonged to Ground Range Detected (GRD) products. Two polarization modes (i.e., VV & VH) were in every scene with 10-m resolution. The original images must be pre-processed including calibration, speckle filtering and terrain correction with Sentinel-1 toolbox of ESAs Sentinel Application Platform (SNAP). Then, the information of the pixels in accordance with field measurement samples was extracted from the images, including local incidence angle and backscatter coefficients of VV and VH. The detailed information of the four Sentinel-1 scenes is illustrated in Table I

TABLE I

DETAILED INFORMATION OF THE FOUR SENTINEL-1 SCENES USED IN THIS PAPER

Data	Scene	Incidence Angle (°)	Polarization
2018.04.07	S1A_IW_GRDH_1SDV_20180407T102821_20180407T102846_021360_024C3C_72B6	42 - 48	VV &VH
2018.04.19	S1A_IW_GRDH_1SDV_20180419T102821_20180419T102846_021535_0251B1_249E	42 - 48	VV &VH
2018.05.01	S1A_IW_GRDH_1SDV_20180501T102822_20180501T102847_021710_025733_460E	42 - 48	VV &VH
2018.06.01	S1A_IW_GRDH_1SDV_20180601T102014_20180601T102039_022162_026591_0867	31 - 37	VV &VH

2.3 Sentinel-2 data

Accordingly, four scenes of Sentinel-2 images were used, too. They were on April 8, 18, 28, and May 28, 2018. These four scenes were 1C products with clouds less than 20%. Atmospheric correction was conducted with the Sen2Cor module of ESAs SNAP, which would turn the 1C products to 2A products. Every image had 12 bands and just Band 4 (Red band) and Band 8 (Near Infrared band) were used here to provide normalized difference vegetation index (NDVI) of every sample as an input variable of SVR. Absolutely, the resolution of these two bands is 10 m. The detailed information of the four Sentinel-2 scenes is illustrated in Table II.

TABLE II

Detailed information of the four Sentinel-2 scenes used in this paper

Data	Scene
2018.04.08	S2B_MSIL1C_20180408T030539_N0206_R075_T49SGT_20180408T060931
2018.04.18	S2B_MSIL1C_20180418T030539_N0206_R075_T49SGT_20180418T055826
2018.04.28	S2B_MSIL1C_20180428T030539_N0206_R075_T49SGT_20180428T064305
2018.05.28	S2B_MSIL1C_20180528T030539_N0206_R075_T49SGT_20180528T064505

3. METHODS

Alpha approximation method is appealing for soil moisture estimation because of its simplicity [7]. More specifically, it avoids the influence of vegetation and surface roughness parameters. This method can be expressed as follows [6]:

$$\frac{\sigma_0^2}{\sigma_0^1} \approx \left| \frac{\alpha_{PP}^2(\varepsilon_s, \theta)}{\alpha_{PP}^1(\varepsilon_s, \theta)} \right|^2 \tag{1}$$

$$|\alpha_{HH}(\varepsilon_s, \theta)| = \left| \frac{\varepsilon_s - 1}{\left(\cos\theta + \sqrt{\varepsilon_s - \sin^2\theta}\right)^2} \right| \tag{2}$$

$$|\alpha_{VV}(\varepsilon_s, \theta)| = \left| \frac{(\varepsilon_s - 1)[\sin^2\theta - \varepsilon_s(1 + \sin^2\theta)]}{\left(\cos\theta + \sqrt{\varepsilon_s - \sin^2\theta}\right)^2} \right| \tag{3}$$

where, σ_0^i represents the backscatter coefficient at time i, ε_s indicates the soil dielectric constant, θ donates the incidence angle and $PP = HH \ or \ VV$ represents polarization.

Moreover, the dielectric constant ε_s is linked to soil moisture m_v directly through a empirical expression of Hallikainen et al. [10] and the coefficients can be settled according to the soil components of experiment area.

In this paper, the first field measurement was conducted from April 6 to 8, 2018, and Sentinel-1 data on April 7, 2018 were available. Once these two types of data are used as a priori information, the soil moisture on April 19, and May 1, 2018 when another two Sentinel-1 scenes are available can be retrieved easily. Sometimes some outliers may emerge in the augmented data, and thus, all these data would be removed to eliminate interference. In this paper, the values of soil moisture content less than $0.05\,\mathrm{cm^3/cm^3}$ or more than $0.45\,\mathrm{cm^3/cm^3}$ are treated as outliers. It is worth mentioning that the augmentation would result in most outliers when using Sentinel-1A and Sentinel-1B data together. Therefore, only Sentinel-1A data with 12-day repetition period were used to augment measured data.

As mentioned above, 34 measured values of soil moisture were obtained in the first field measurement. Firstly, the measured data are divided into two parts, i.e., training (21 points) and testing (13 points). These two datasets should both contain the points distributing in all-region of values. Subsequently, data augmentation is just carried out on training data in order to maintain the purity of testing data. Thus, two different training datasets can be obtained before and after data augmentation while testing dataset stayed the same. Based on the data augmentation, three different methods were used to retrieve soil moisture, i.e, $\mathrm{Method_1}$ (original alpha approximation method), $\mathrm{Method_2}$ (SVR trained by only part of measured data) and $\mathrm{Method_{new}}$ (SVR trained by part of measured data and the augmented data together). And the results were compared and the optimal method can be determined. Afterwards, the optimal method was conducted with the relevant remote sensing data of Yancheng, Henan province, which was collected with GEE. Therefore, the soil moisture map of agriculture area in Yancheng can be obtained.

4. RESULTS AND DISCUSSIONS

The results of Method_1, Method_2, and $\mathrm{Method}_{\mathrm{new}}$ are compared in this section. The scatter diagrams are illustrated in Fig. 2. RMSE and R^2 are listed in Table III.

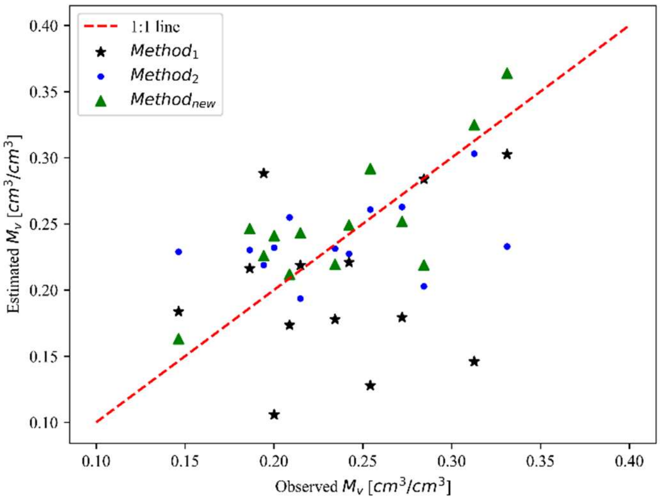

Fig.2. Results of Method_1, Method_2, and $\mathrm{Method}_{\mathrm{new}}$.

TABLE III

RMSE AND R^2 OF METHOD_1, METHOD_2 AND $\mathrm{METHOD}_{\mathrm{NEW}}$

Method	RMSE	R^2
Method_1	0.0775	0.0467
Method_2	0.0478	0.1419
$\mathrm{Method}_{\mathrm{new}}$	0.0339	0.6491

For Method_1, only 6 points from all 13 points were distributed around 1:1 line and the others deviated much from the line and showed irregular distribution. Hence, RMSE of Method_1 is 0.0775 and R^2 is just 0.0467. Method_2 gives a better prediction of soil moisture content. Compared to Method_1, we can find that many points of Method_2 are closer to 1:1 line. However, almost all of the calculated values are around 0.18 to 0.27. The prediction is not good when the observed values are low or high. Comparatively, $\mathrm{Method}_{\mathrm{new}}$, which combines the advantage of alpha approximation method and SVR machine, gives a satisfying prediction in the whole value region. As showed in Fig. 2, the points at observed $\mathrm{M_v} \approx 0.15$ and 0.35 are close to 1:1 line, which means the SVR machine can give a good estimation in low or high values region after augmenting the training data with alpha approximation method. Accordingly, RMSE of $\mathrm{Method}_{\mathrm{new}}$ decreased to 0.0339 and R^2 reached to 0.6491.

And the soil moisture map of agriculture area in Yancheng on April 7, 2018 is shown in Fig. 3.

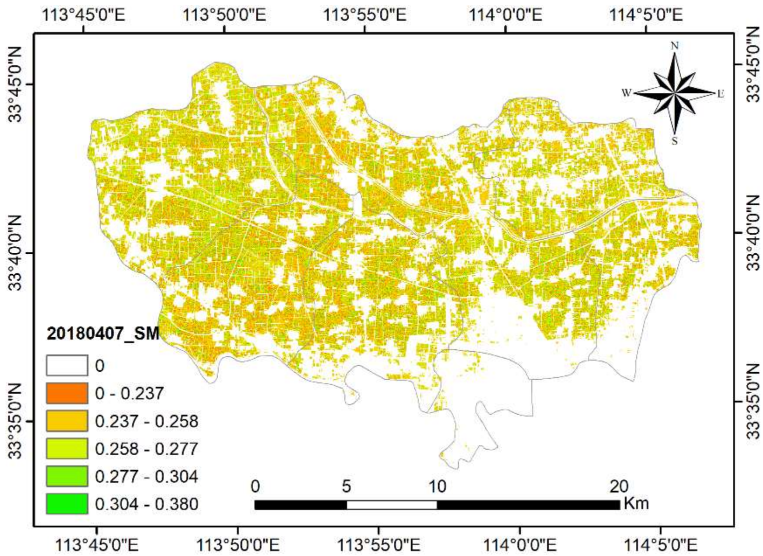

Fig.3. Soil moisture map of agriculture area in Yancheng on April 7, 2018.

5. CONCLUSION

In this paper, alpha approximation method is used to augment measured data for an SVR machines training, and then, the SVR is applied to estimate soil moisture content. After optimization, $\mathrm{Method}_{\mathrm{new}}$ performed much better than $\mathrm{Method}_1$and Method_2. The biggest advantage of this method is that the use of just one-third measured data to obtain the same or very close results with SVR, if three according scenes of SAR images are available. And based on this method, the soil moisture map can be obtained efficiently with GEE.

References

[1] S. Paloscia, S. Pettinato, E. Santi, C. Notarnicola, L. Pasolli, and A. Reppucci, "Soil moisture mapping using Sentinel-1 images: Algorithm and preliminary validation," REMOTE SENSING OF ENVIRONMENT, vol. 134, pp. 234–248, 2013.

[2] Y. H. Kerr, P. Waldteufel, J. P. Wigneron, J. M. Martinuzzi, J. Font, and M. Berger, "Soil moisture retrieval from space: the Soil Moisture and Ocean Salinity (SMOS) mission," IEEE Transactions on Geoscience & Remote Sensing, vol. 39, no. 8, pp. 1729–1735, 2002.

[3] E. G. Njoku, T. J. Jackson, V. Lakshmi, T. K. Chan, and S. V. Nghiem, "Soil moisture retrieval from AMSR-E," IEEE Transactions Geoscience & Remote Sensing, vol. 41, no. 2, pp. 215–229, 2003.

[4] S. R. Cloude and E. Pottier, "A review of target decomposition theorems in radar polarimetry," IEEE transactions on geoscience and remote sensing, vol. 34, no. 2, pp. 498–518, 1996.

[5] E. J. Rignot and J. J. Van Zyl, "Change detection techniques for ERS- 1 SAR data," IEEE Transactions on Geoscience and Remote sensing, vol. 31, no. 4, pp. 896–906, 1993.

[6] A. Balenzano, G. Satalino, V. R. N. Pauwels, and F. Mattia, "Dense Temporal Series of C-

and L-band SAR Data for Soil Moisture Retrieval Over Agricultural Crops," IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, vol. 4, no. 2, pp. 439–450, 2011.

[7] H. Lian, Q. Qin, R. Panciera, M. Tanase, and H. Yang, "An Extension of the Alpha Approximation Method for Soil Moisture Estimation Using Time-Series SAR Data Over Bare Soil Surfaces," IEEE Geoscience & Remote Sensing Letters, vol. PP, no. 99, pp. 1–5, 2017.

[8] J. D. Ouellette, J. T. Johnson, A. Balenzano, F. Mattia, G. Satalino, S. B. Kim, R. S. Dunbar, A. Colliander, M. H. Cosh, and T. G. Caldwell, "A Time-Series Approach to Estimating Soil Moisture From Vegetated Surfaces Using L-Band Radar Backscatter," IEEE Transactions on Geoscience & Remote Sensing, vol. PP, no. 99, pp. 1–8, 2017.

[9] V. Vapnik, The nature of statistical learning theory. Springer science & business media, 2013.

[10] Hallikainen, Martti T., et al. "Microwave dielectric behavior of wet soil-part 1: Empirical models and experimental observations." IEEE Transactions on Geoscience and Remote Sensing 1 (1985): 25-34.