environmental sciences proceedings

MDPI

Proceedings paper

Application of LSTM in the Analysis of Soil Moisture Time Series Obtained from GNSS-IR†

Patricia Danghian and Asghar Rastbood *

Faculty of Civil Engineering, University of Tabriz, Tabriz, IRAN; p.danghian1400@ms.tabrizu.ac.ir
* Correspondence: arastbood@tabrizu.ac.ir

† Presented at the Chair of the 5th International Electronic Conference on Remote Sensing, German Research Centre for Geosciences, Potsdam, Germany, 7-21 November 2023.

Abstract: Global Navigation Satellite System-Interferometric Reflectometry (GNSS-IR) is a technique for monitoring soil moisture content. In this study, data from station P038 in New Mexico, from 2017 to 2020 are used. After generating signal-to-noise ratio (SNR) data, initial reflection height is estimated, phase is obtained for each satellite, vegetation cover effects are mitigated and removed finally the result is converted to volumetric water content (VWC). LSTM neural network model is used to predict time series of VWC. Model is trained using 80% of observations. By updating network with observed values instead of predicted values, RMSE decreased from 0.09 to 0.04.

Keywords: GNSS-IR; LSTM; SNR; soil moisture; time series

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date

1. Introduction

Estimating VWC of soil can be useful for choosing the kind of plant to be cultivated in that region. Measuring soil moisture has lots of importance. It is a key factor to find out which vegetation can be supported by soil of that region, according to needs the optimum moisture content of the soil for plant growth can be discovered. In Fig. 1(a) SNR curve produced by reflections of a bare soil is displayed. In Fig. 1(b) soil is covered by snow and that causes impressive changes in SNR curve in comparison with bare soil. In Fig. 1(c) there is a layer of vegetation on the soil. This also makes obvious changes in SNR curve. Figure 1(d) shows SNR curve produced by reflections of a wet soil. For predicting time series of VWC we have used LSTM neural network. The LSTM neural network can maintain its content over a long period of time and essentially remember previous information. Gates are also vectors with values between zero and one that determine how old information should be progressed and new information should be added. Generally, one means to pass and zero means to discard information. The input gate specifies which parts of the input data and to what extent they should be added to the memory content. The forget gate determines which parts of the memory content should be removed. The output gate also determines which part of the hidden state content should contain the memory content [1].

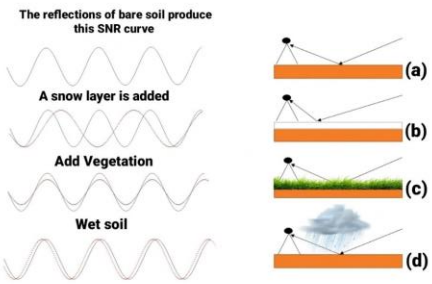


Figure 1. changes occurred in SNR curve in (**a**) bare soil, (**b**) snowy soil, (**c**) vegetated soil, (**d**) wet soil.

In this article, station P038 is selected for our research. The reason for choosing this station is uninterrupted reflection observations of this site during 4 years from 2017 to 2020. In first step the SNR data is generated from Rinex files for these 4 years, then the algorithm suggested by Larson for estimation of soil moisture changes is used [2]. Also for modeling and predicting time series of soil moisture, for the first time LSTM is used and the data is evaluated based on its results. The results show that the soil moisture content has been increased during these years.

2. Soil Moisture Estimation Algorithm Using GNSS

This algorithm is used to generate soil moisture estimation for all stations in PBO networks. This algorithm has 4 main steps. At first useful satellite tracks are selected, in second step a prior reflection height is estimated for each track then SNR metrics are approximate and finally vegetation effects are quantified and removed [3] [4].

2.1. Selection of Useful Satellite Tracks

Satellite tracks should have steady reflections between satellite elevation angles of 5°-25° or 5°-30°. In choosing analyzing area we have to pay attention to buildings and manmade surfaces like roads in order to not abstract the tracks.

2.2. A Prior Reflection Height Estimation for Each Track

The prior reflection height is not exactly known; it should be estimated from SNR data. To characterize the SNR interferogram we can use Equation (4).

$$SNR_{direct=V\,cos(\varphi)} \tag{1}$$

$$SNR_{Reflected}=BVcos(\varphi+\Delta\varphi) \tag{2}$$

$$\Delta\varphi=2\pi L/\lambda;\ L=2Hsin(e);\ A=BV \tag{3}$$

Using Equations (1), (2) and (3):

$$SNR_{Reflected}(e)=Acos(\frac{4\pi H}{\lambda}sin(e)+\varphi) \tag{4}$$

H0 is the a priori reflector height. E is the elevation angle of the satellite, A is an amplitude term, $\boldsymbol{\lambda}$ is the GPS wavelength, and ϕ is a phase shift. Phase and amplitude are calculated using least-squares estimation.

2.3. Vegetation Effect Quantification

An increase in soil moisture causes a decrease in SNR amplitude. If moisture content increases by 0.4 $(\boldsymbol{cm})^{\mathbf{3}}$ $(\boldsymbol{cm})^{-\mathbf{3}}, \boldsymbol{A}_{\mathbf{norm}}$ will decrease from 1.00 to 0.78. Since the difference between residual soil and saturated one for standard soil is about 0.4 $(\boldsymbol{cm})^{\mathbf{3}}$ $(\boldsymbol{cm})^{-\mathbf{3}}$, there shouldn't be a decrease in $\boldsymbol{A}_{\mathbf{norm}}$ beyond 0.78, unless in presence of vegetation growth, so if a satellite track's $\boldsymbol{A}_{\mathbf{norm}}$ time series stays below 0.78, it shows the effect of something other than soil moisture variations, like changes in vegetation water content. In other hand a track that has a time series above 0.78 shows small vegetation effects, in this case it is not easy to understand weather decreases in $\boldsymbol{A}_{\mathbf{norm}}$ are from vegetation or soil moisture changes. If there is a significant vegetation effect in our site, we have to use an algorithm to remove the obstructive effects [5] [6].

3. LSTM Neural Network

Long Short-Term Memory (LSTM) is a type of recurrent neural network architecture that was introduced by Hochreiter and Schmidhuber in 1997, and was later improved by

Jürgen Schmidhuber and his colleagues (2000) as well as Greff and his colleagues (2017). LSTM can be seen as an updated version of RNN. The network takes the output of the previous block ($\mathbf{h_{t-1}}$) and the input vector ($\mathbf{x_t}$) and produces the output of the current block ($\mathbf{h_t}$). The order of input data ($\mathbf{x_t}$) is important for this network.
The relationships of the RNN network are defined in Equation (5), Equation (6) and Equation (7).

$$\boldsymbol{h_t = f_w(h_{t-1}, x_t)} \tag{5}$$

$$\boldsymbol{h_t = tanh(w_{hh}h_{t-1}, w_{xh}x_t)} \tag{6}$$

$$\boldsymbol{y_t = w_{hy}h_{t-1}} \tag{7}$$

Where $\mathbf{w}$ stands for weights and $\mathbf{y_t}$ for network output. Figure 2 shows the block diagram of the RNN network, in which $\mathbf{W}$ shows weights and $\mathbf{L}$ displays the memories. The hyperbolic tangent function ($\mathbf{tanh}$) is the main part of the blocks. A part of the outputs ($\mathbf{y_i}$) is sent to the memories $\mathbf{L_i}$.

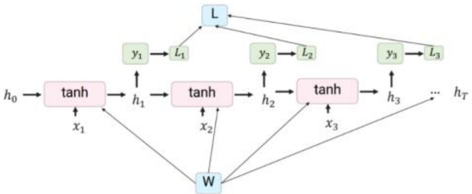


Figure 2. The block diagram of the RNN network shows that the weight matrix $\mathbf{W}$ affects both the input vectors $\mathbf{x}$ and the hidden states $\mathbf{h}$ [7].

4. Case Study

Our discussion draws an example from PBO station called P038. The site shown in Fig. 3(a) is located in Portales, New Mexico at 34.14726°W and -103.40734°S. The station and its monument both has been installed since 2015 and it has been working till today. The data from 2017 till 2020 will be analyzed in our study.

(a) **(b)**

Figure 3. (**a**) Photograph of a GPS site, called P038 in Portales, New Mexico. The antenna phase center at this site is approximately 2 meters above the soil surface. The station has been installed since 2015. (**b**) Google Earth image for PBO site p038 and reflection zones for 2-meter reflector height and satellites with elevation angles between 5-15 degrees.

Before we start our study, we have to calculate the GNSS reflection zones to make sure that we can sense the surface we want to measure. For this we need to check the position of our site and get the gaps in sensing zones and if needed we can have an azimuth or elevation angle mask. As it is displayed in Fig. 3 (b) there is a gap in reflection zone for our site. These are the areas where GPS signals are weak or unavailable, this can happen for several reasons such as obstructions, atmospheric conditions or receiver limitations. The gap below shows the area that satellites have not been able to cover, as there isn't any data recorded in that azimuths there is no need to use azimuth mask. According to Fig. 3(b) all chosen elevation angles and azimuths are suitable. After analyzing the reflection zones of our site and choosing the best elevation angle and azimuths we will generate SNR data from RINEX files, SNR data are converted from dB- Hz to a linear scale (volts/volts) for each rising and setting satellite track. We are going to use multipath signals to generate SNR data [8]. We can take a quick look at our data to quickly test various options (elevation angles, frequencies, azimuths, and quality control parameters) of our site so we can choose the most suitable ones.

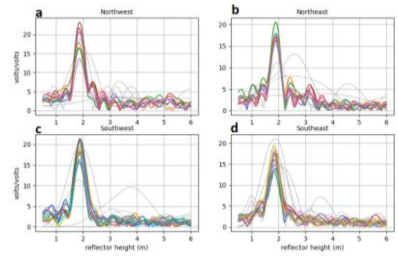


Figure 4. LSP periodogram for SNR data extracted from L2C data captured by P038 GNSS station in four geographic quadrants, (**a**)northwest, (**b**) northeast, (**c**) southwest, (**d**) southeast.

Figure 4 shows all L2C data received by P038 site on 100th day of 2019. Pay attention that the x-axis does not go beyond 6 meters. In addition, note that results on the x-axis begin at 0.5 meters so this region is not allowed since we are not able to resolve very small reflector heights with this method. By using these periodograms we can get whether there is a planar reflector below our antenna or not. As the peaks in the periodograms bunch up around 2 meters, it means that at this site the antenna phase center is ~ 2 meters above the ground. Each color stands for a different satellite. The data that are plotted in grey shows, we have a failed reflection. The quadrants are Northwest, Northeast and so on.

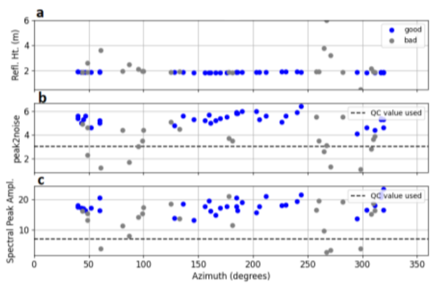


Figure 5. A summary of various quality control metrics, (**a**) reflector height, (**b**) peak to noise ratio, (**c**) spectral peak amplitude.

According to Fig. 5(a) blue dots show the successful Reflector Height (RH) retrievals and grey ones stand for unsuccessful retrievals. The dashed lines show what QC metrics quick Look was using. For our station as shown in Fig. 5(a) we see that the retrieved reflector heights are almost steady at azimuths between 130 and 250 degrees. From Fig. 5(b) we can get that a peak2noise QC metric of 3 is acceptable. As expressed in Fig. 5(c), the amplitudes are generally larger than 11 so we can accept 8 as a minimum value. Then we need to estimate prior reflector heights for our site. Soil moisture algorithm uses GPS satellites. In our case it is effective to use L2C data because it results higher quality than other GPS signals. In station P038 the antenna is 2 meters above the ground. Elevation angle is limited between 5° to 30° because data below elevation angles of 5° may be obstructed by trees or high buildings but the elevation angle that here are used are mostly affected by multipath [2]. In second step, firstly the best satellite tracks are identified, the default is all of rising and setting L2C satellite arcs. After that we are going to estimate the phase for each satellite track on each day. As displayed in Fig. 6, the phase results will be plotted for four geographic quadrates.

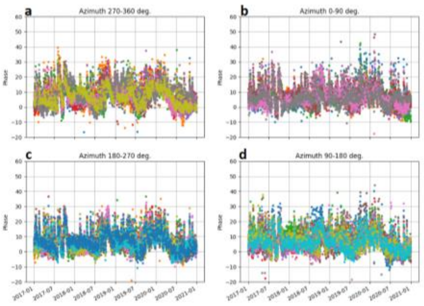


Figure 6. Phase results of all satellites for each day for four geographic quadrates, (**a**)northwest, (**b**) northeast, (**c**) southwest, (**d**) southeast.

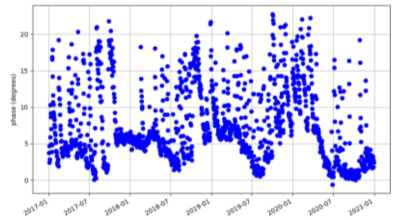


Figure 7. Daily L2C phase results.

Then the phase resulted in previous stage, displayed in Fig. 7, will convert to volumetric water content. Vegetation have significant impact on the phase resulted and it is necessary to remove the effect to achieve accurate soil moisture estimate so before changing units from phase (degrees) to VWC, we will model and remove the vegetation effects by using spectral altitude. In Fig. 8(a) we can see the phase results with and without vegetation correction. Finally, by using soil texture profiles for our site we can level the data that is expressed in Fig. 8(b). It is important to know that nonsense soil moisture values like negative ones are not allowed here.

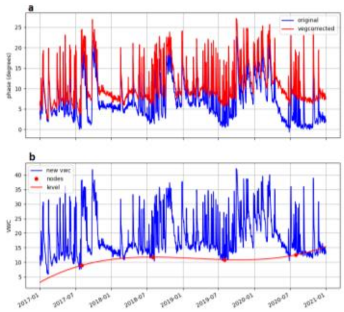


Figure 8. (**a**) Phase results with and without vegetation correction. (**b**) Volumetric water content and it's levelling.

As shown in Fig. 8(b), level of volumetric water content was 8.88 in 2017, these increases to 11.74 in 2018. Then there is a bit decrease through next year. The level reaches to 10.88 in 2019. In 2020 the level rises up to 12.49. In Fig. 9 the final result of our work in expressed. Horizontal axis stands for date and shows our work duration that relates to years 2017 till end of 2020 and vertical axis stands for volumetric soil moisture, on each day of this time period.

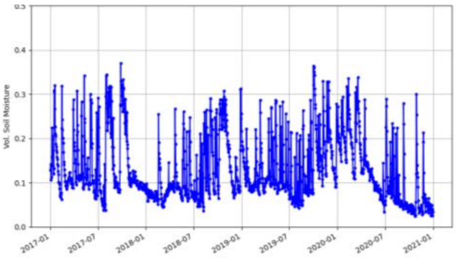
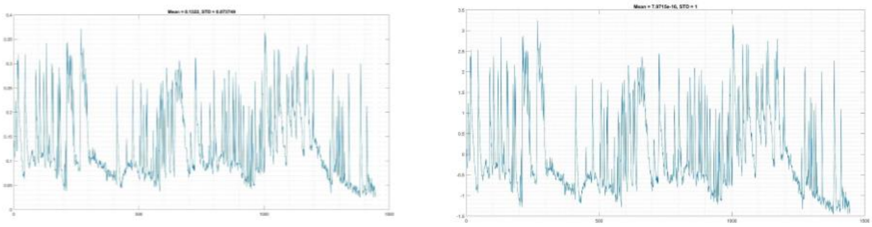


Figure 9. Final Volumetric Water Content (VWC) for P038

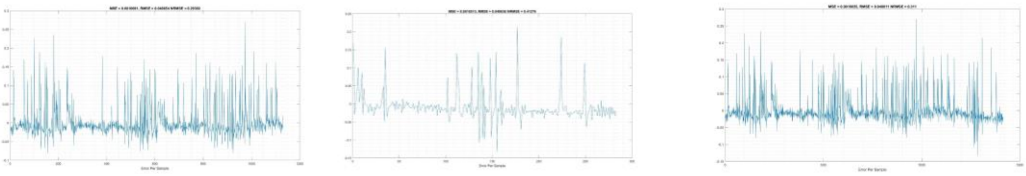
Final step of this paper is to accurately predict the soil moisture multiple days in advance. This solution will help farmers prepare their irrigation schedules more efficiently. For soil moisture time series prediction, it is assumed that there is not much change in the behaviour of the time series at different times. Assuming that the time series is related to a limited past, the future of the time series is predicted using LSTM neural networks with tuned hyper parameters using Bayesian optimization. Model is trained using 80% of observations and tested using 20% left data. We will update network with observed values instead of predicted values, so RMSE decreased from 0.09 to 0.04. Figure 10 shows normalized data versus original data with mean and standard deviation values.



(**a**) (**b**)

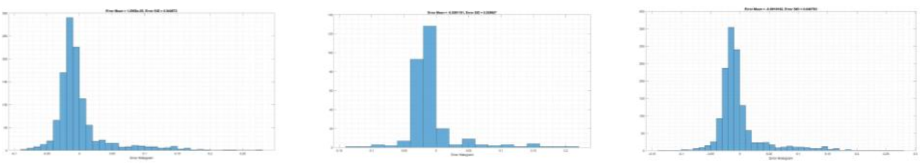
Figure 10. (**a**) Original versus (**b**) Normalized data.

Figures 11 and 12 show the error evaluation and histogram for train, test and all data respectively. The error should be around zero. This is almost the case here.



(**a**) (**b**) (**c**)

Figure 11. Error evaluation for (**a**) train, (**b**) test and (**c**) all data.



(**a**) (**b**) (**c**)

Figure 12. Error histogram for (**a**) train, (**b**) test and (**c**) all data .

6. Conclusion

In this article a multi stage process is presented for estimating near-surface soil moisture around a PBO station called P038, since 2017 till 2020. We used an algorithm to model and remove vegetation effect in estimated soil moisture time series. The primary step for our research was to generate SNR data from RINEX files. During several stage we achieve to phase for each satellite on each day. Then vegetation effect is mitigated and the result will convert to volumetric water content. According to the results the volumetric water content level has reached to 12.49 from 8.88 since 2017 till 2020. Finally soil moisture prediction is done for multiple days in advance. It will be useful for farmers to prepare irrigation schedules more efficiently. In order to get the best results, the hyper parameters tuning of LSTM is done using Bayesian optimization algorithm. The accuracy of the network is checked by different metrics on train, test and all data. In this study Model was trained using 80% of observations and then we by using the remaining 20% of observations the model was tested. We updated network with observed values instead of predicted values and decreased the RMSE from 0.09 to 0.04. Also, R^2 and Rank correlation values for test data were decreased from 0.0299 to -0.0007 so we reached to more accurate prediction. The novelty in this article was the use of LSTM for the first time to model and predict the time series of soil moisture for future days.

Author Contributions: Conceptualization, A.R. and P.D.; resources, A.R.; data curation, A.R. and P.D.; writing—original draft preparation, P.D. and A.R.; writing—review and editing, A.R.; supervision; A.R.; All authors have read and agreed to the published version of the manuscript.

Funding: This research receives no external funding, it's a part of master's thesis called "An investigation in the environmental applications of GNSS systems with interferometric reflectometry.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Barrett, B.W., E. Dwyer, and P. Whelan, *Soil moisture retrieval from active spaceborne microwave observations: An evaluation of current techniques.* Remote Sensing, 2009. **1**(3): p. 210-242.
2. Larson, K.M., et al., *Using GPS multipath to measure soil moisture fluctuations: initial results.* GPS solutions, 2008. **12**(3): p. 173-177.
3. Mao, K., J. Wang, and M. Zhang, *The study of soil moisture retrieval from GNSS_R signals based on AIEM model and experiment data.* High Tech Lett, 2009. **3**: p. 295-301.
4. Masters, D., *Surface remote sensing applications of GNSS bistatic radar: Soil moisture and aircraft altimetry*. 2004: University of Colorado at Boulder.
5. Rodriguez-Alvarez, N., et al., *Vegetation water content estimation using GNSS measurements.* IEEE Geoscience and Remote Sensing Letters, 2011. **9**(2): p. 282-286.
6. Chew, C.C., et al., *Vegetation sensing using GPS-interferometric reflectometry: Theoretical effects of canopy parameters on signal-to-noise ratio data.* IEEE Transactions on Geoscience and Remote Sensing, 2014. **53**(5): p. 2755-2764.
7. Lindemann, B., et al., *A survey on anomaly detection for technical systems using LSTM networks.* Computers in Industry, 2021. **131**: p. 103498.
8. Larson, K.M., et al., *GPS multipath and its relation to near-surface soil moisture content.* IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2009. **3**(1): p. 91-99.