LSTM BASED SOIL MOISTURE PREDICTION

Shivanand Venkanna Sheshappanavar[1], Chilukuri K. Mohan[1] and David G. Chandler[2]

[1] Department of Electrical Eng. and Computer Science
Syracuse University, New York
ssheshap/mohan@syr.edu
[2] Department of Civil and Environmental Eng.
Syracuse University, New York
dgchandl@syr.edu

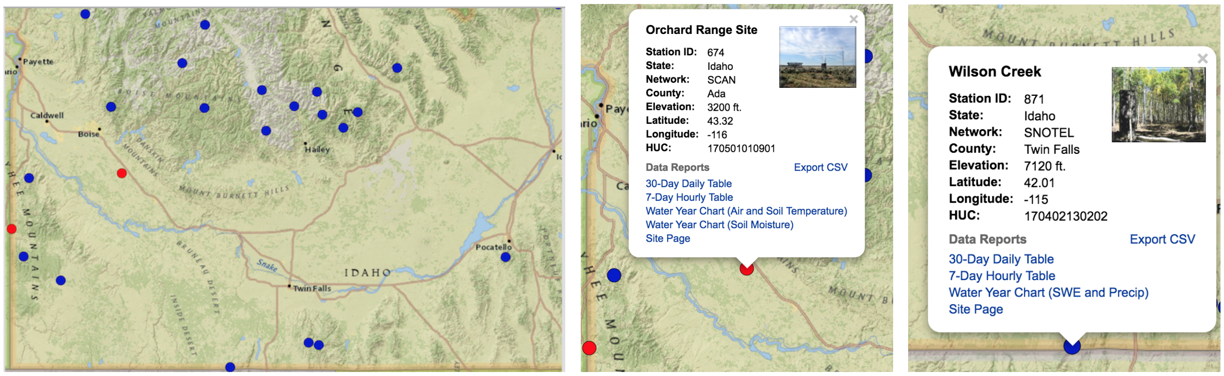
Abstract

Soil moisture content is an important variable that has a considerable impact on agricultural processes and practical weather-related concerns such as flooding and drought. We address the problem of predicting soil moisture by applying recurrent neural networks that use *Long Short-Term Memory (LSTM)* models. The success of our approach is evaluated using a dataset obtained from ground-based sensor infrastructure networks. Feature reduction using a mutual information approach is shown to be more effective than feature extraction using principal component analysis.

1 Introduction

Soil moisture has a colossal impact on several hydrological processes including infiltration, evapotranspiration and subsurface flow. Measurement and prediction of soil moisture help us obtain deeper insights into the localized dynamics of critical ecological processes. Accurate prediction of soil moisture allows the quantification of drought conditions and the prediction of flash floods caused by precipitation run-off. Economic consequences of accurate soil moisture prediction are significant, assisting improvements in crop productivity and agricultural management practices, and permitting precise control over the root zone environment, leads to healthier crops and higher yields. Weather forecasts can be improved, since high soil moisture results in high evaporation, increasing the likelihood of moisture convergence. In addition, monitoring soil moisture provides us with better understanding of how water, energy and carbon are exchanged between land and air.

Neural networks such as multilayer perceptrons have been useful in prediction of stream-flow based on snow accumulation, along with recommends application of Principal Component Analysis (PCA) [1]. More recently, deep networks have been used for prediction, providing greater flexibility in mapping diverse, complex functions. The risk of over-fitting the data (resulting in poor generalization) can be mitigated using regularization techniques or by reducing the feature space dimensionality. Researchers have recently used Deep Belief Networks (DBN) and other techniques for feature learning or extraction [2].



(a) Location of the SNOTEL/SNAP sites[2] (b) SiteID 674 (c) SiteID 871

Figure 1: Two sites selected for this experiment[2]

Recurrent Neural Networks with Long Short-Term Memory (LSTM) rank among the state-of-the-art networks for predicting future values of a time series, with potential application to hydrology [3]. Deep Feed-forward Neural Networks have also been used on datasets obtained using Visible Infrared Imaging Radiometer Suite (VIIRS) from cropland China [4].

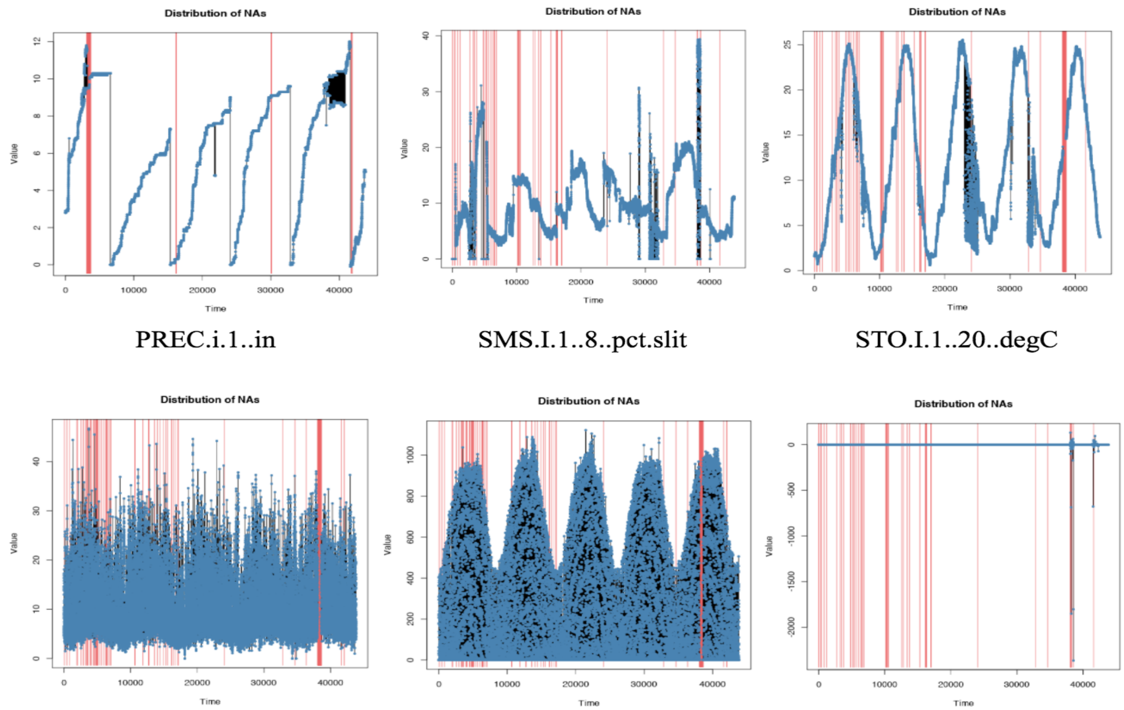
This paper presents the application of LSTM model for soil moisture prediction, using datasets collected from ground using Soil Climate Analysis Network (SCAN) and Snow Telemetry (SNOTEL) networks, unlike prior works which focused on SMAP datasets. We apply this approach to data from different soil climate networks, and evaluate two feature extraction methods on this data. We were able to find the most significant features for soil moisture prediction, and shown the effectiveness of our approach compared to feature extraction methods previously used in hydrology.

2 Data Collection

Two important sources of data used in this paper are from networks SCAN and SNOTEL, managed and operated by National Resources Conservation Services (NRCS) and Natural Water and Climate Center (NWCC) [5]. Figure 1 (a) shows a group of sites (SCAN-red, SNOTEL-blue) in Idaho. We selected 2 sites one from each network, Orchard Range Site (SCAN Station ID 674) and Wilson Creek (SNOTEL Station ID 871) located in Ada and Twin Falls county respectively of Idaho State as in figure 1 (b) and (c). Datasets with predefined features can be downloaded from [6] by specifying start and end dates, selecting hourly options for an annual period. A dataset spanning for 5 years (01/01/2012 to 12/31/2016) has been collected, treated for missing values, divided into training (4 years, 2012-2015) and testing (3 months, Jan-March 2016) sets.

Table 1: List of features with tag names.

All Feature Names (total - 21)
(PREC.I-1, PREC.I-2) Precipitation Accumulation
(TOBS.I-1) Air Temperature Observed
(STO.I-1:-2,8,20) Soil Temperature Observed
(SAL.I-1:-2,8,20) Salinity
(RDC.I-1:-2,8,20) Real Dielectric Constant
(BATT.I-1,BATT.I-2,BATT.I-3) Battery
(WDIRV.H-1) Wind Direction Average
(WSPDX.H-1) Wind Speed Maximum
(WSPDV.H-1) Wind Speed Average
(SRADV.H-1) Solar Radiation Average
(RHUMV.H-1) Relative Humidity Average
(RHUMX.H-1) Relative Humidity Maximum
Soil Moisture Targets (total - 3)
(SMS.I-1:-2,8,20) Soil Moisture Percent



PREC.i.1..in | SMS.I.1..8..pct.slit | STO.I.1..20..degC

WXPDX.H.1..mph | SRADV.H.1..watt. | SAL.I.1..2..gram.

Figure 2: Plot of missing values (red vertical lines) and actual values (blue dots) of few features.

Table 1 lists the features for station ID 674 (other sites may have less or more). For some of the features values are missing at a given site so we have analyzed the data for missingness by plotting the dataset for NA distributions as shown in figure 2 (due to space constraints and for proper visibility including only few original features in figure 2). The red vertical lines represent the missing values. General statistics of missing values shown in figure 3 where first row represents number of data points with no missing values for any of the features. A '1' represents that a value exists for the feature in the given row and a '0' represents missing value. The corresponding number on left most column represents the number of such rows in the dataset. For example only one row is missing all values except for precipitation accumulation 'PREC.I.1..in.' as represented by penultimate row of the table. We removed features 10 percentage or more missing values for example dropped battery 'BATT.I-3' feature resulting in the feature count to 19 as shown in table 3. Mutual information between the

Figure 3: General statistics of missing values

	SMS.I.1..8..pct....silt.	STO.I.1..8..degC.	SAL.I.1..8..gram.	RDC.I.1..8..unit.	SMS.I.1..2..pct....silt.	SAL.I.1..2..gram.	RDC.I.1..2..unit.	TOBS.I.1..degC.	BATT.I.1..volt.	BATT.I.2..volt.	STO.I.1..2..degC.	SMS.I.1..20..pct....silt.	STO.I.1..20..degC.	SAL.I.1..20..gram.	RDC.I.1..20..unit.	WDIRV.H.1..degr.	WSPDX.H.1..mph.	WSPDV.H.1..mph.	SRADV.H.1..watt.	RHUMV.H.1..pct.	RHUMX.H.1..pct.	PREC.I.1..in.	PREC.I.2..in.	
39742	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0
432	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1
2901	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1
110	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	2
60	1	1	1	1	1	1	1	0	0	0	1	1	1	1	1	1	1	1	1	1	1	0	0	5
43	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	1	6
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	7
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	1	0	7
291	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	10
199	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	11
30	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1	1	1	12
2	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	1	1	1	1	1	1	0	1	13
1	1	1	1	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	13
6	1	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	14
2	0	0	0	0	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	1	1	14
4	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	1	18
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	21
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	22
	43	43	43	43	48	48	48	65	65	65	246	540	540	540	540	558	558	558	558	558	558	609	3074	9948

features and the visualization of features as in figure 2 suggested omitting soil moisture percent, two salinity attributes, one real dielectric constant and wind direction average features. We plotted density of the observed data and imputed data as shown in figure 4 to conclude that the observed values neither are identical nor similar to the imputed values (Missing Not at Random).

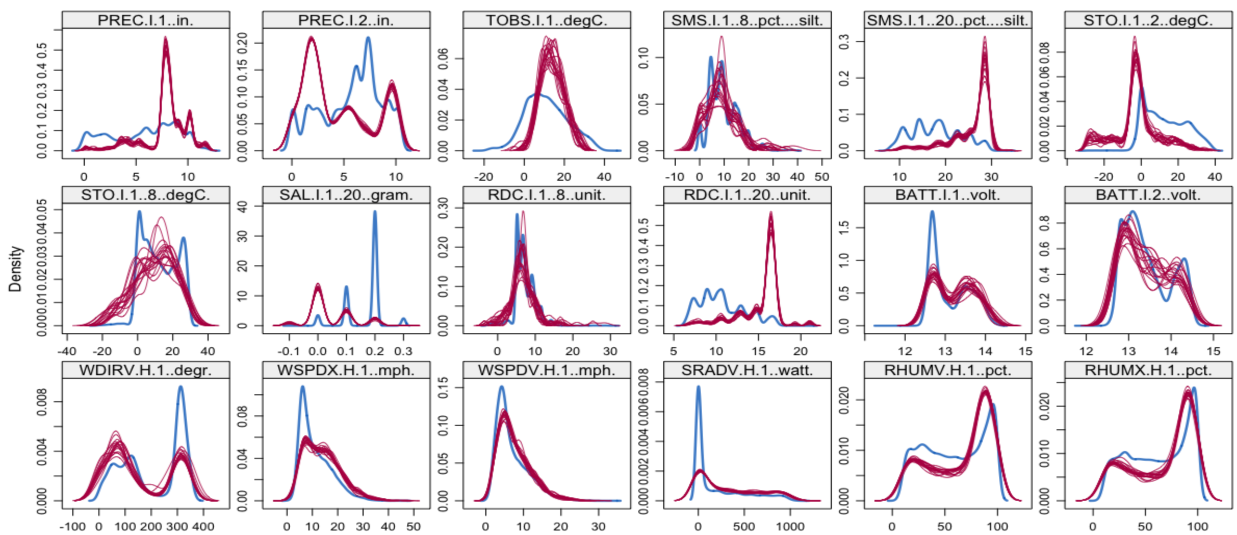


Figure 4: The observed values(blue) are neither identical nor similar to the imputed values(red) which is why the missing of values is concluded as Missing Not at Random (MNAR).

We used two methods for filling in the missing values and reduce feature dimensionality:

- Dataset I: Exponential moving average imputation using R package imputeTS [7] by taking the exponential moving average (with a window of k=24 indicating 24 hours or daily average as a replacement value) for each of the feature and filled the missing values.

- Dataset II: Multiple Imputation by Chained Equations 'MICE', is another R package [8] for filling missing values. MICE fills in data based on each feature and generates internally say a 'P' number of datasets where 'P' is equal to the number of columns (features(15)+targets(3)) in the dataset giving priority to each column. So we used 'with' function to explicitly get hold of the P number of datasets (in case of StationID 674 P=15+3=18) and 'pool' function to combine the datasets (in this case 18) to get a final MICE filled dataset.

- Dataset III: Using PCA on Dataset II we further reduced the features to just 8 from 15, since eight of the original features showed higher mutual information.

In each case, we have three subset datasets labeled A, B and C. All three subsets have only one target soil moisture content feature i.e. either of 2-inch or 8-inch or 20-inch.

- Subset A has all 15 features.

- Subset B has only those features which are relevant to that particular depth for example at depth 8-inch the features set does not include 2-inch or 20-inch soil temperature observed and salinity at 20-inch.

- Subset C is same as subset A but does not have any wind or air temperature related features. In total 3x3 = 9 subsets are considered if all the features and targets are available (or less as in case of StationID 674 with 6 datasets).

Table 2: Types of Datasets generated for Site ID 674 with 8-inch as target soil moisture.

Category	**Dataset II - MICE filled**	**Soil Moisture**	**Datasets generated**
8-inch A	PREC.I-1,TOBS.I-1,STO.I-1:-2,STO.I-1:-8,STO.I-1:-20, SAL.I-1:-20,RDC.I-1:-8,RDC.I-1:-20,BATT.I-1,BATT.I-2, WSPDX.H-1,WSPDV.H-1,SRADV.H-1,RHUMV.H-1,RHUMX.H-1	SMS.I-1:-8	MICE-PCA (8-inch A)
8-inch B	PREC.I-1,TOBS.I-1,STO.I-1:-8,RDC.I-1:-8, BATT.I-1,BATT.I-2,WSPDX.H-1,WSPDV.H-1, SRADV.H-1,RHUMV.H-1,RHUMX.H-1	SMS.I-1:-8	MICE-PCA (8-inch B)
8-inch C	PREC.I-1,STO.I-1:-2,STO.I-1:-8,STO.I-1:-20, SAL.I-1:-20,RDC.I-1:-8,RDC.I-1:-20,BATT.I-1, BATT.I-2,SRADV.H-1,RHUMV.H-1,RHUMX.H-1	SMS.I-1:-8	MICE-PCA (8-inch C)

Table 2 gives a glimpse of feature subsets, the target feature for each of A, B and C category belonging to dataset II generated by MICE filling of missing values. It also lists the datasets which can be generated from these subsets. Similarly keeping 2-inch, 20-inch data as the target value, 6 more subset datasets can be generated.

3 Method

Our method, illustrated in figure 5, includes data collection, data treatment, feature reduction or extraction, then the application of an LSTM network for soil moisture prediction. The setup and the layer wise workings of LSTM has been thoroughly discussed in [3]. We used the model prescribed in KERAS (a python library) for the core implementation of LSTM with epochs 50, batch size 72, 100 hidden units, linear activation as parameters with adam optimizer, root mean squared error as loss function using 48 months of data for training and 3 months of data in testing the model.

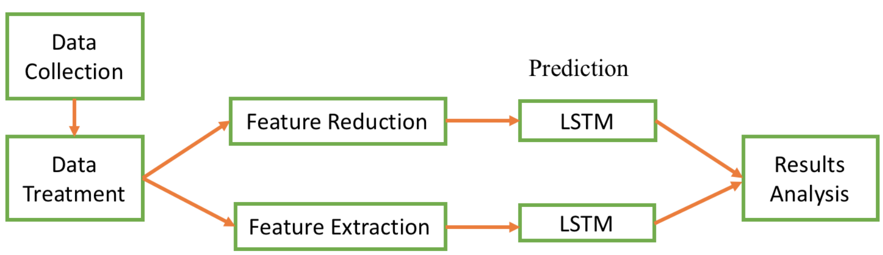


Figure 5: Overview of methodology

3.1 Feature Extraction using Principal Component Analysis (PCA)

Feature extraction is essential for effective summarization of all the features but on a reduced or minimum dimensions. We used a well-established statistical method PCA for features extraction which transforms data into a new space and still hold most of the original information. Useful features that are linear combinations of data attributes are derived by this approach. So our 15 features dataset was finally reduced to just 8 features reducing the dimension of dataset which showed more mutual information than compared to the original dataset.

Table 3: Feature extraction methods used and number of features after each method

Feature extraction	Feature category	Original Features	Features based on percentage of missing values	Features based on visualization and mutual information	Features from PCA
Feature Names	Precipitation Accumulation	1+1	1+0	1+0	Eight significant principal components
	Air Temperature Observed	1	1	1	
	Soil Temperature Observed	1+1+1	1+1+1	1+1+1	
	Salinity	1+1+1	1+1+1	0+0+1	
	Real Dielectric Constant	1+1+1	1+1+1	0+1+1	
	Battery	1+1+1	1+1+0	1+1+0	
	Wind Direction Average	1	1	0	
	Wind Speed Maximum	1	1	1	
	Wind Speed Average	1	1	1	
	Solar Radiation Average	1	1	1	
	Relative Humidity Average	1	1	1	
	Relative Humidity Maximum	1	1	1	
total=N		21	19	15	8

4 Experimental Results

Applying LSTM to predict soil moisture content on 6 and 9 of the data subsets from Station ID 674 and Station ID 871 respectively we arrived at results presented in Table 4 and 5 with results of two datasets (II and III) including all three subsets (A,B and C). Due to space constraints we are only presenting the prediction and RMSE plots of station ID 674 in comparison with a well know machine learning method called Support Vector Machine (SVM) using radial basis function as kernel. All results are obtained with 50 epochs, using adam optimizer. For station 674 using dataset II (MICE filled) the RMSE was lowest among all combinations as seen in table 4.

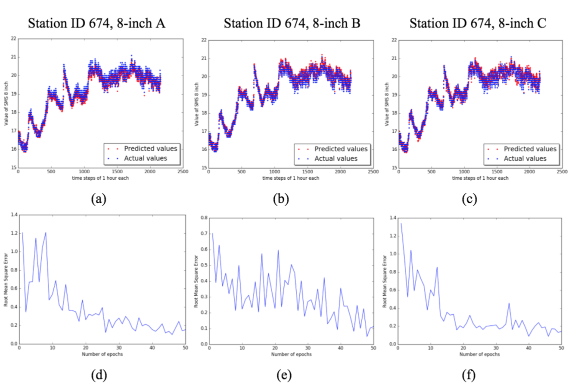


Figure 6: Station ID 674 8-inch Dataset II Subset-Category A, B and C prediction results(a,b,c) with root mean square errors(d,e,f)

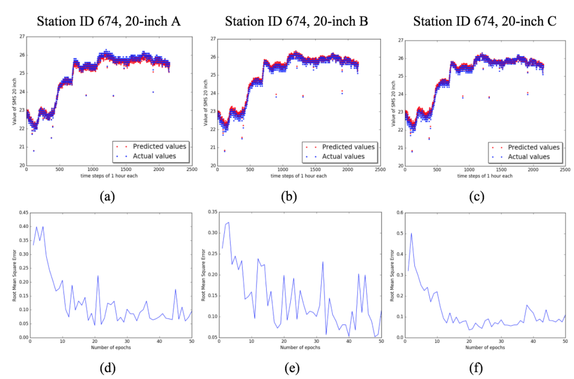


Figure 7: Station ID 674 20-inch Dataset II Subset-Category A, B and C prediction results(a,b,c) with root mean square errors(d,e,f)

Figures 7,8 and 10 show plots of predicted and actual values of soil moisture for station ID 674, plots of RMSE for each of the A, B and C subsets, for each of 8 and 20-inch, for each datasets II and III. Greater variations among values in 8-inch as shown in figure 7 when compared to values in Figure 8 can be attributed to the depth of the sensor device itself i.e. at 8-inch interference could be more compared to 20-inch depth. Figure 9 shows similar plot for station ID 871.

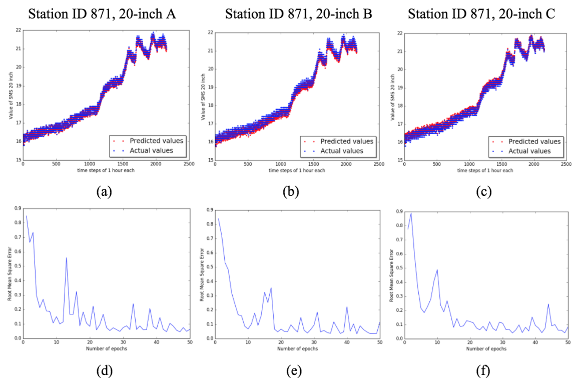


Figure 8: Station ID 871 20-inch Dataset II Subset-Category A, B and C prediction results(a,b,c) with root mean square errors(d,e,f)

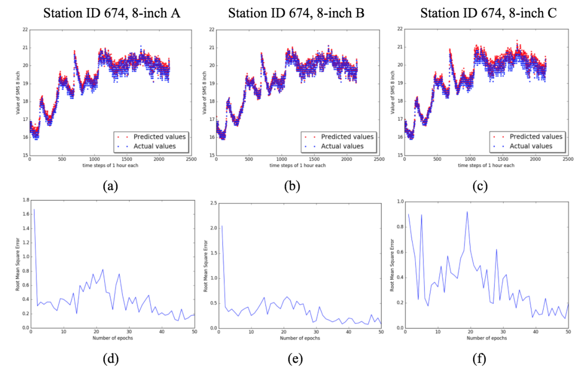


Figure 9: Station ID 674 8-inch Dataset III (MICE+PCA) Subset A, B and C prediction results (a,b,c) with root mean square errors(d,e,f)

Table 4: Station ID 674 RMSE values using LSTM and SVM: minimum, maximum and average values(LSTM) using Dataset II(missing values filled by MICE) and Dataset III(PCA on Dataset II)

	LSTM						SVM	
674	MICE			MICE+PCA			MICE	MICE+PCA
RMSE	min	max	avg	min	max	avg		
8-inch A	0.103	1.209	0.373	0.101	1.671	0.398	6.217	8.201
8-inch B	0.052	**0.326**	0.143	0.079	2.048	**0.340**	6.325	8.205
8-inch C	0.088	1.340	0.345	**0.075**	0.922	0.351	6.222	7.499
20-inch A	0.044	0.401	0.131	0.390	1.450	0.756	11.113	13.895
20-inch B	0.049	0.702	0.301	0.451	**0.83**7	0.546	12.058	17.131
20-inch C	**0.037**	0.502	**0.121**	0.275	0.878	0.393	11.335	15.559

A slight skewness can be seen in figure 7 (c) and more skewness in figure 10 (c) where majority of the predicted values are slightly larger than the actual values which is why the red dots appear above the blue dots. This skewness can be attributed non inclusion of wind related features in the C subset of respective datasets, especially at 8-inch those features must be significant which is why we don't see any skewness for 20-inch plots on C subset datasets (figures 8(c), 9(c)). The greater proximity in actual and predicted values in the plots of figure 8 and 9 can be attributed to lesser disturbances at 20-inch than at 8-inch. Much worse SVM results of Station 871 are excluded from table 5. Also from tables 4 and 5 it is clear that RMSE was higher whenever PCA was used for feature selection.

Table 5: Station ID 871 RMSE values: minimum, maximum and average values using Dataset II(missing values filled by MICE) and Dataset III(PCA feature extraction on Dataset II)

871	MICE			MICE+PCA		
RMSE	min	max	avg	min	max	avg
2-inch A	0.288	1.441	0.625	3.905	8.36	6.83
2-inch B	0.294	2.181	0.836	0.585	2.201	0.818
2-inch C	0.312	1.443	0.627	0.525	2.209	0.850
8-inch A	0.195	1.505	0.614	4.004	6.920	5.65
8-inch B	0.385	1.352	0.714	0.159	1.542	0.667
8-inch C	0.183	2.080	0.600	0.101	1.573	0.592
20-inch A	0.046	0.850	0.165	1.843	3.506	2.715
20-inch B	**0.033**	**0.840**	**0.149**	0.086	**0.979**	**0.235**
20-inch C	0.041	0.891	0.170	**0.052**	1.020	0.287

5 Conclusions

This paper has explored the application of an LSTM based neural network learning model for soil moisture prediction, applied to real (SCAN and SNOTEL) data. In our work the data treatment and feature extraction steps played significant role in obtaining better results. Application of *MICE* for missing values imputation gave best results. We created a benchmark dataset of 6 and 9 subsets for sites 674 and 871 respectively based on sensor depth. PCA based feature extraction with 8 significant features performed well but with some degradation compared to the initial 15 features. Achieved RMSE minimization during both training and testing with an overall minimum RMSEs in range of 0.03 to 0.4 at different depths for majority of the datasets created for both the sites.

In future work, we plan to improve results by using RNNs for filling the missing data as proposed in [9], to make our novel approach an end to end neural network based soil moisture content prediction and forecasting system. Also fine tune LSTM parameters, to achieve negligible RMSE. As Neural networks are sensitive to outliers in the data, we plan to adopt methods in [10] for outlier detection and replacement, and investigate if statistical features can further improve the results.

References

[1] Tarnpradab, Sansiri, et al. ”Neural networks for prediction of streamflow based on snow accumulation.” Computational Intelligence for Engineering Solutions (CIES), 2014 IEEE Symposium on. IEEE, 2014.

[2] Bai, Yun, et al. ”Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models.” Journal of Hydrology532 (2016): 193-206.

[3] Fang, Kuai, et al. ”Prolongation of SMAP to Spatio-temporally Seamless Coverage of Continental US Using a Deep Learning Neural Network.” arXiv preprint arXiv:1707.06611 (2017).

[4] Zhang, Dongying, et al. ”Upscaling of Surface Soil Moisture Using a Deep Learning Model with VIIRS RDR.” ISPRS International Journal of Geo-Information 6.5 (2017): 130.

[5] Schaefer, Garry L., and Ron F. Paetzold. ”SNOTEL (SNOwpack TELemetry) and SCAN (soil climate analysis network).” Proc. Intl. Workshop on Automated Wea. Stations for Appl. in Agr. and Water Resour. Mgmt. 2001.

[6] NRCS website for SCAN and SNOTEL datasets. https://www.wcc.nrcs.usda.gov/scan/

[7] Moritz, Steffen, and Thomas Bartz-Beielstein. ”imputeTS: Time Series Missing Value Imputation.” R package version 0.4 (2015).

[8] Buuren, Stef, and Karin Groothuis-Oudshoorn. ”mice: Multivariate imputation by chained equations in R.” Journal of statistical software 45, no. 3 (2011).

[9] Che, Zhengping, et al. ”Recurrent neural networks for multivariate time series with missing values.” arXiv preprint arXiv:1606.01865 (2016).

[10] Zhao, Zhiruo, Chilukuri K. Mohan, and Kishan G. Mehrotra. ”Adaptive Sampling and Learning for Unsupervised Outlier Detection.” In FLAIRS Conference, pp. 460-466. 2016.