3.5 ESTIMATION OF SOIL MOISTURE USING AN ARTIFICIAL NEURAL NETWORK

Hongli Jiang* and William R. Cotton
Department of Atmospheric Science, Colorado State University Fort Collins, Colorado

1 INTRODUCTION

Accurate estimation of soil moisture is essential for the successful simulations of boundary layer evolution (Golaz et al. 2001), mesoscale circulations (Fast and McCorkle 1991), and convection (Clark and Arritt 1995; Gallus and Segal 2000). However, since soil moisture observations are insufficient for direct real-time initialization, many alternative methods and techniques are proposed and tested for retrieval of soil moisture. Among them are the direct use of passive microwave remotely sensed data (Vinnikov et al., 1999) or assimilating into the land data assimilation system (LDAS) (Burke et al. 2001), and assimilation of satellite-derived infrared heating rates into a numerical model (Wetzel et al. 1984; McNider et al. 1994; Jones et al. 1998a,b) or combining satellite-derived infrared skin temperature and vegetation index (Gillies and Carlson, 1995).

The low resolution, passive microwave data are not very useful for cloud and meso- scales applications because of its large footprint (from ∼25 km at 37 GHz to 150 km at 6.6 GHz). Data from these satellites are not ideal for measuring soil moisture in situations where there is significant vegetation cover (Owe et al. 1999). The National Oceanic and Atmospheric Administration's (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and Geostationary Operational Environmental Satellite (GOES) Visible-Infrared Spin Scan Radiometer (VISSR) provide extensive spatial and temporal coverage. In addition, AVHRR data can be used to retrieve various geophysical parameters such as IR skin temperature and vegetation index closely related to soil moisture (Leese et al. 2001).

The purpose of this study is to develop a new methodology capable of estimate soil moisture from remotely sensed IR data with sufficient spatial and temporal resolution. The new methodology is an artificial neural network (ANN) based technique, adapted from the ANN model developed for precipitation estimation based on infrared (IR) satellite data (Hsu et al. 1997; Hsu et al. 1999; Sorooshian et al. 2000).

*Corresponding author address, Hongli Jiang, Department of Atmospheric Science, Colorado State University, Fort Collins, CO 80523; e-mail: jiang@atmos.colostate.edu.

Artificial neural network provides a promising alternative to meteorological applications that the characteristics of the processes are difficult to describe using physical equations. Although there is no known physical equations relating soil moisture with other meteorological and surface variables, variables such as skin temperature, precipitation, and vegetation index are found to be highly correlated to the soil water content (Gillies and Carson 1995).

Using IR heating rate, normalized difference vegetation index (NDVI), and surface precipitation as the primary inputs to the ANN model, we examined the applicability and potentials of the ANN in soil moisture estimation using data from LDAS outputs. The calibration and validation are based on two months of data over the continental United States. Good agreement is demonstrated between the ANN model estimate and the LDAS soil moisture values. Soil moisture information from LDAS model output is used as the target data to adjust the ANN parameters.

2 ARTIFICIAL NEURAL NETWORKS

The artificial neural network (ANN) model used in this study is developed by a group of scientists in the University of Arizona. Readers are referred to Hsu et al. (1997,1999) for detailed description, evaluation, and discussion on the overall performance of the model. A brief description is given below.

The architecture of the neural network is a modified Counter Propagation Network (CPN) (Hecht-Nielsen 1988). The CPN structure consists of two functional components. Fig. 1 shows the schematic diagram of the ANN model. The input-hidden component detects and classifies features and nonlinearity in the input data, and then maps them into clusters by using a self-organizing feature map (SOFM) algorithm. The SOFM is trained through an unsupervised learning using only the input data in the training process.

The hidden-output component performs linear mapping (LM) that relates the classified inputs into the output variables. The linear mapping algorithm is trained using a negative gradient algorithm in which the mean square error between the ANN output and the target data is minimized. Once the network error has decreased to less than a specified threshold value, the network has converged and is considered to be trained.

The two components are trained separately. Ideally, one would use a data set that contains seasonal and topographic changes to train the ANN model. When there is only limited number of input data, the ANN model allows continuously training, namely sequential training, if additional observations become available after the model is initially trained. The training process is also considered as the model calibration, and the sequential training is treated as the model validation, testing, or simulation. During the sequential training, the model parameters are adjusted and the model output will be improved.

3 SELECTION OF THE INPUT VARIABLES

A neural network learns the input-output relationship through a training process. One of the critical issues in training the ANN model is to select input variables that are highly correlated to the soil moisture. Previous studies have showed good correlation between the soil moisture and the infrared skin temperature and a normalized difference vegetation index (NDVI) (Gillies and Carlson 1995), and the infrared heating rate and the antecedent precipitation index (API) (Wetzel and Woodward 1987). These variables are sought from two independent data sources.

Although satellite data are readily available, it is still very time consuming to process a huge amount of satellite data. To focus our attention on understanding and evaluation of the ANN model without spending a large effort to deal with data processing, we obtained the skin temperature, surface precipitation, and root zone soil moisture from the available LDAS data (available online: http://ldas.gsfc.nasa.gov/), and used the LDAS data as the "ground truth" for soil moisture. The LDAS data contains the hourly gridded data over the US (125W – 67W, 25N – 53N) with the resolution of 0.125^o by 0.125^o lat-long.

The NDVI data are composites biweekly gridded data over the US (128.52W – 75.41W, 22.48N – 48.4N) from U.S Geological Survey's Earth Resources Observations Systems (USGS EROS) data center. This dataset is projected and linearly interpolated onto the LDAS data grid. The NDVI data provides seasonal variations in the vegetation pattern.

4 EVALUATION OF TRAINING RESULTS

Following Gillies and Carlson (1995), the ANN model is first trained with the normalized infrared skin temperature ($T1^*$) and NDVI (N^*) as the input variables (see Table 1). Note that the definition of N^* in this study is slightly different from that of Gillies and Carlson (1995) for simplicity. N^* ranges from 0 to 1. To select the input variables that produce the best correlation, several different input variables are tested (Table 1). Correlation coefficients between the input variables and soil moisture are calculated (Table 2). The performance of the ANN model is evaluated by examining the correlation between the ANN model estimate and LDAS soil moisture after initial training (Table 3).

EXP	Input Variables
exp1	$N^* = \frac{(NDVI-NDVI_{min})}{(NDVI_{max}-NDVI_{min})}$, $T1^* = \frac{(T_s-T_a)}{R_{net}}$
exp2	N^*, $T2^* = T_s/R_{net}$
exp3	N^*, $T3^* = T_s$
exp4	N^*, $T4^* = (T_{smax}-T_{smin})/(0.5*(T_{smax}+T_{smin}))$
exp5	N^*, $T4^*$, DRR
exp6	N^*, $T4^*$, ARR

Table 1: T_s is the infrared skin temperature, T_a is the air temperature measured at 2m above surface, R_{net} is the net radiative fluxes, T_{smax} and T_{smin} are calculated over 6 h time period during the morning hours, DRR is averaged daily rain, and ARR is the previous 30-day accumulated rain.

Input variables	Corr(input,sm)
$T1^*$	0.0155
$T2^*$	-0.2374
$T3^*$	-0.1079
$T4^*$	-0.5449
N^*	0.4639
DRR	0.3068
ARR	0.6706

Table 2: Correlation coefficient between input variables and soil moisture. The second column lists the correlation coefficient between the corresponding input variables and soil moisture.

Data from June 1998 are used to train the ANN model. As shown in Table 2, the correlation coefficients are low for $T1^*$, $T2^*$, and $T3^*$, but increased (negatively) significantly when $T1^*$, $T2^*$, and $T3^*$ used in exp1, exp2, and exp3, respectively, are replaced with the normalized heating rate ($T4^*$) in exp4. The daily rain (DRR) or the previous 30-day accumulated rain (ARR) is added as the third input variable. The correlation coefficient between soil moisture and the accumulated rain is more than twice of that between the soil moisture and daily rain. The statistical results (Table 3) of the initial training show a steady decrease in RMSE as the correlation between the

ANN model estimate and the LDAS soil moisture improves. Given the results in Table 2 and 3, the input variables used in exp6 are clearly the winner. With this level of correlation, useful soil moisture data are obtainable from the ANN model estimate.

EXP/Stat	CORR	RMSE	BIAS
exp1	0.4929	0.07740	-0.12289e-5
exp2	0.5513	0.07423	0.45050e-6
exp3	0.6084	0.07061	-0.40703e-6
exp4	0.6052	0.07081	0.26237e-6
exp5	0.6140	0.07021	-0.34857e-6
exp6	0.7347	0.06038	-0.39903e-6

Table 3: Statistics of training results

5 ANN SOIL MOISTURE ESTIMATION RESULTS

The trained ANN was then applied to independent cases to see if it could produce the soil moisture close to the observations. The performance of the ANN model in soil moisture estimation is tested with data from July 1998.

Correlations between the ANN model estimate and LDAS soil moisture (Fig. 1a) are all higher than those training results shown, respectively, in Table 3, while the root mean square error (RMSE, Fig. 1b) are lower after the sequential training with the data from July 1998. The correlation is as high as 0.879 for the exp6. The positive bias (model estimate minus the observational values) indicates that the ANN model is overestimating the soil moisture (Fig. 1c). Comparison among the three experiments (Fig. 1) shows that the exp6 produces the best correlation and least RMSE and bias.

Spatial distributions from testing results for 1 July 1998 is plotted along with the observation (LDAS data) in Fig. 2 to find out where the model has missed. As indicated in the error plot (Fig. 2c), the model overestimated the soil moisture the most in the western region of Nebraska. There is a generally overestimate of the soil moisture in the dryer regions and underestimate in the moister regions by the model.

Based on the soil moisture plot (Fig. 2a) the model is applied to three different regions ranging from dry to moist, and the results are plotted in Fig. 3. The three regions are New Mexico (dry), mid-west (moist), and Nebraska which is the most overestimated region by the model. The averaging volumetric soil moisture is below 0.2 in New Mexico (Fig. 3a), and close to 0.4 in the mid-west (36N – 41N, 90W – 85W). The overestimation (3a) and underestimation (3c) are small when comparing

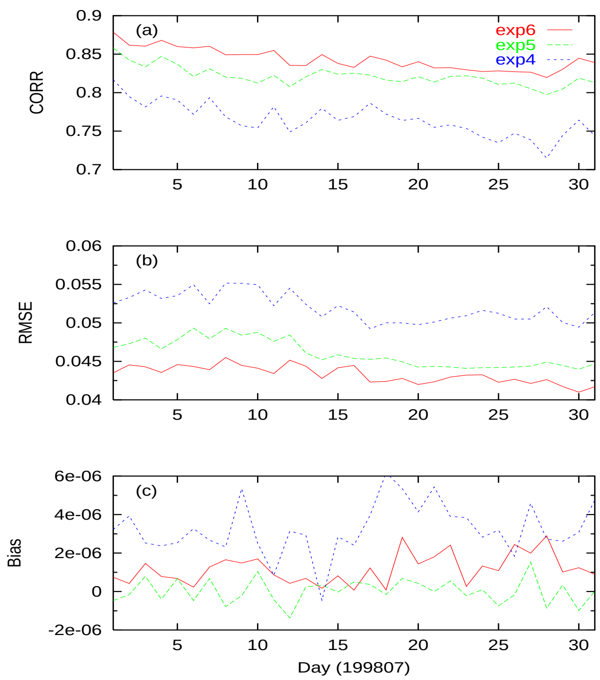


Figure 1: Statistical results of the ANN soil moisture estimation from July 1998. Variables are (a) correlation between the ANN model estimate and LDAS soil moisture, (b) Root mean square error, and (c) bias defined as the difference between the model estimate and LDAS soil moisture with line types as labeled. LDAS data is used as the "ground truth" for soil moisture in this study.

with the results of Nebraska. When all the data points (Fig. 3d) are used to compute the correlation coefficient between the estimated and observed soil moisture, R is much higher (0.879) whereas R for individual regions is much lower as labeled in each panel. These results show that sufficient amount of data are necessary to adequately represent the full range of soil moisture distributions.

6 SUMMARY AND DISCUSSION

A neural network based methodology is tested in soil moisture estimation. We showed that the ANN model can be trained to retrieve soil moisture information. The biggest challenge is whether the ANN model is capable of retrieving soil moisture directly from remotely sensed data since the LDAS is still model output. If it can, then ANN has the following advantages over LDAS:

- Soil moisture retrieval is not restricted to the ETA model domain.
- Resolution can be as high as the input data (skin temperature, precipitation).
- Soil moisture can be retrieved for historical periods as long as the input data records.

The research is ongoing, and more results will be presented at the conference, and appear in the publication (Jiang and Cotton, 2002).

Acknowledgments. The authors wish to thank Dr. Kuo-lin Hsu of the University of Arizona for providing us the ANN model and answering questions related to the ANN model. We also thank Dr. Brian Cosgrove of NASA LDAS group for his assistance in obtaining the LDAS data. This research was funded by NOAA under Grant NA67-RJ0152.

7 REFERENCES

Burke, E. J., W. J. Shuttleworth, K-h, Lee, and L. A. Bastidas, 2001: Using area-averaged remotely sensed surface soil moisture in multipatch land data assimilation systems. *IEEE Trans. Geosci. Remote Sensing*, **39**, 2091–2100.

Clark, C. A. and R. W. Arritt, 1995: Numerical simulations of the effect of soil moisture and vegetation cover on the development of deep convection. *J. Appl. Meteorol.*, **34**, 2029–2045.

Fast, J. D. and M. D. McCorkle, 1991: The effects of heterogeneous soil moisture on a summer baroclinic circulation in the central United States. *Mon. Wea. Rev.*, **119**, 2140–2167.

Gallus, W. A. Jr. and M. Segal, 2000: Sensitivity of forecast rainfall in a Texas convective system to soil moisture and convective parameterization. *Wea. Forecasting*, **14**, 509–525.

Gillies, R. R. and T. N. Carlson, 1995: Thermal remote sensing of surface soil water content with partial vegetation cover for incorporation into climate models. *J. Appl. Meteorol.*, **34**, 745–756.

Golaz, J-C, H. Jiang, and W. R. Cotton, 2002: A Large-eddy simulation study of cumulus clouds over land and sensitivity to soil moisture. *Atmos. Res.*, **59-60**, 373–392.

Hecht-Nielsen, R, 1988: Applications of counterpropagation networks. *Neural Networks*, **1**, 131-139.

Hsu, K. L, X. Gao, S. Sorooshian, and H. V. Gupta, 1997: Precipitation estimation from remotely sensed information using artificial neural networks. *J. Appl. Meteor.*, **36**, 1176–1190.

Hsu, K. L, H. V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of physical variables from multichannel remotely sensed imagery using a neural network: application to rainfall estimation. *Water Resour. Res.*, **35**, 1605–1618.

Jiang, H and W. R. Cotton, 2002: Soil moisture estimation using an artificial neural network. in preparation.

Jones, A. S, I. G., Guch, and T. H. Vonder Haar, 1998a: Data assimilation of satellite-derived heating rates as proxy surface wetness data into a regional atmospheric mesoscale model. Part I: methodology. *Mon. Wea. Rev.*, **126**, 634–645.

Jones, A. S, I. G., Guch, and T. H. Vonder Haar, 1998b: Data assimilation of satellite-derived heating rates as prosy surface wetness data into a regional atmospheric mesoscale model. Part II: a case study. *Mon. Wea. Rev.*, **126**, 646–667.

Leese, J., T. Jackson, A. Pitman, and P. Dirmeyer, 2001: GEWEX/BAHC international workshop on soil moisture monitoring, analysis, and prediction for hydrometeorological and hydroclimatological applications. *Bull. Amer. Meteor. Soc.*, **82**, 1423–1430.

McNider, R. T., A. J., Song, D. M. Casey, P. J. Wetzel, and W. L. Crosson, 1994: Toward a dynamic-thermodynamic assimilation of satellite surface temperature. *Mon. Wea. Rev.*, **122**, 2784–2803.

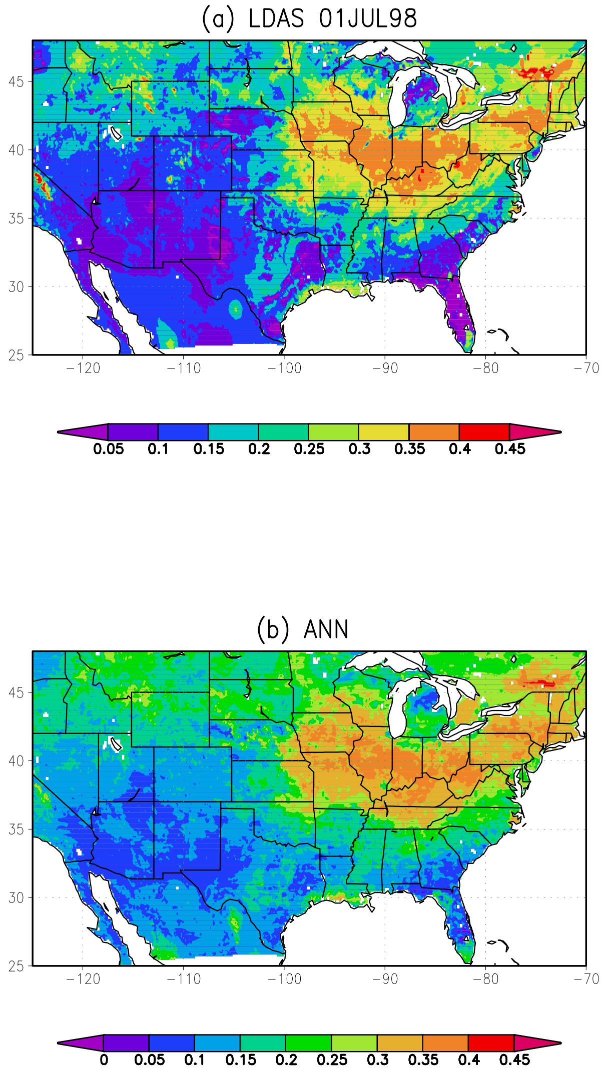
Owe, M., A. A., Van de Griend, R. de Jeu, J. J. de Vries, E. Seyhan, and E. T. Engman, 1999: Estimating soil moisture from satellite microwave observations: past and ongoing projects and relevance to GCIP. *J. Geophys. Res.*, **104**, 19735–19742.

Sorooshian, S., K-L, Hsu, X. Gao, H. V. Gupta, B. Imam, and D. Braithwaite, 2000: Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Amer. Meteor. Soc.*, **81**, 2035–2046.

Vinnikov, K. Y., A. Robock, S. Qiu, J. K. Entin, M. Owe, B. J. Choudhury, S. E. Hollinger, and E. G. Njoku, 1999: Satellite remote sensing of soil moisture in Illinois, United States. *J. Geophys. Res.*, **104**, 4145–4168.

Wetzel, P. J., D. Atlas, and R. H. Woodward, 1984: Determining soil moisture from geosynchronous satellite infrared data: a feasibility study. *J. climate Appl. Meteorol.*, **23**, 375–391.

Wetzel, P. J. and R. H. Woodward, 1987: Soil moisture estimation using GOES-VISSR infrared data: A case study with a simple statistical method. *J. climate Appl. Meteorol.*, **26**, 107–117.



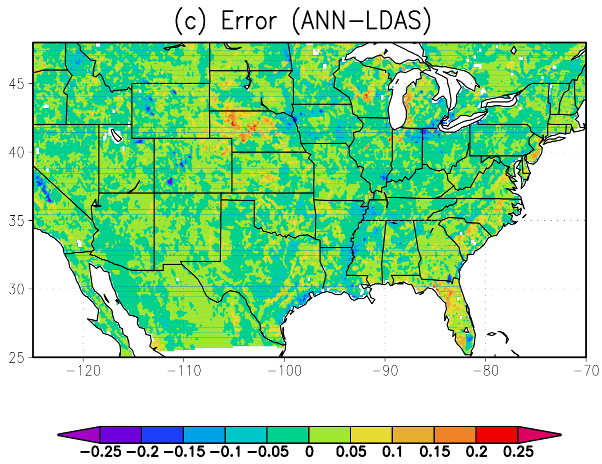


Figure 2: Statistical results of the ANN soil moisture estimation from July 1998. Variables are (a) correlation between the ANN model estimate and LDAS soil moisture, (b) Root mean square error, and (c) bias defined as the difference between the model estimate and LDAS soil moisture with line types as labeled. LDAS data is used as the "ground truth" for soil moisture in this study.

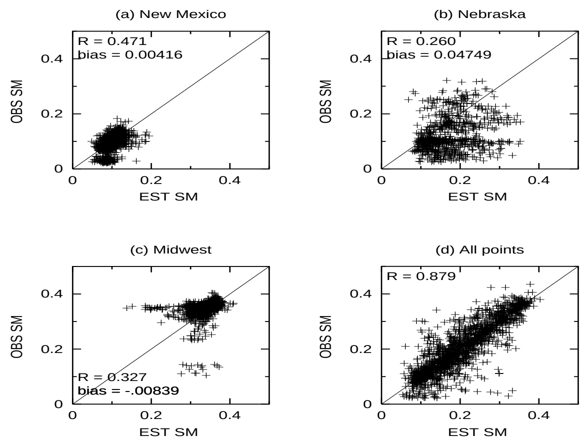


Figure 3: Scatter plots of estimated and observed soil moisture for different regions as labeled for 1 July 1998. Solid line represents points where the estimated soil moisture equals to that of observed.