A Short- and Long-Term Memory Model to Estimate Soil Moisture

Nazario D. Ramirez-Beltran[1], Ramon Vasquez[2], Harold Cruzado[2], and Eric Harmsen[3]

[1]Department of Industrial Engineering, University of Puerto Rico, Mayaguez Puerto Rico, nazario@ece.uprm.edu
[2]Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez Puerto Rico
[3]Department of Agricultural Engineering, University of Puerto Rico, Mayaguez Puerto Rico.

Abstract - Soil moisture exhibits spatial and temporal variability. Spatial variability is mostly associated to climatology patterns, topography, soil texture, and vegetation index; where as, temporal variability is primarily associated to rainfall and air temperature events. The spatial variability is modeled by a long-term memory model whereas the temporal variability by a short-term memory model. The long-term memory model is a regression model that expresses the expected monthly soil moisture for a specific grid, where as the short-term memory model is a stochastic transfer function model that estimates the soil moisture response in hourly basis for a given grid.

Keywords: soil moisture, transfer function model, remote sensing, MODIS, NEXRAD, volumetric water content.

1. INTRODUCTION

It is well established that performances of atmospheric numerical models are very sensitive to initial and boundary conditions. The regional atmospheric modeling system that is being used to simulate the climate dynamics in Puerto Rico is highly sensitive to soil moisture initial conditions (Comarazamy, 2001). The soil moisture over land is a key component of the surface water and energy budget. The soil water content regulates the partition of latent and sensible heat fluxes at the surface, affecting a large number of boundary layer (Balsamo, et al., 2004). Incorrect initial conditions will generate misleading modeling results. For instance Balsamo, et al., (2004) reported that erroneous estimate of total soil moisture affects the quality of the forecast for several days in using the numerical weather prediction scheme. This article attempts to present a high resolution and a reliable methodology for estimating soil moisture. The main purpose of the current research is to use remote sensing information and statistical tools to estimate soil moisture.

The second section of this paper describes the data used in this research. Section three presents the proposed methodology to estimate soil moisture. Preliminary results are shown in section fourth, and conclusions and recommendations are exhibits in the last section.

2. DATA COLLECTION

Data are divided in long-term records and short-term record. The long-term records include monthly records of: rainfall, air temperature, topography, vegetation and soil moisture; where as, short-term records includes hourly records of: rainfall, temperature, and soil moisture. Rainfall observations were obtained by using NEXRAD and include observation from 2002 to 2004. Air temperature and vegetation index were obtained from MODIS and the records include information from 2000 to 2004. Soil texture, elevation and average slope with a 1 km spatial resolution for Puerto Rico (PR) were obtained from the Natural Resources Conservation Service (NRCS). *In-situ* observations provide simultaneous data for soil moisture, rainfall and air temperature in hourly basis. *In-situ* observations come from 15 soil moisture stations located in the western part of PR. The soil sensor measures the dielectric constant of the soil to estimate its volumetric water content. It does this by finding the rate of change of voltage applied to the sensor once it is buried in the soil. A soil station has 3 soil sensors, an air temperature sensor, and a rain gauge. Three soil core samples were obtained at each station with the purpose of measuring in the laboratory the volumetric water content and calibrate the soil moisture sensors. These soil cores are also used to identify the soil texture. A theta probe was also used to measure the spatial variability of the soil moisture in a nearby area of the soil moisture. Figure 1 shows the instruments that integrate a soil moisture station.

Figure 1. Instruments of a soil moisture station.

3. METHODOLOGY.

3.1. Modeling spatial variability.

The long-term memory model that expresses the spatial variability is a regression model that represents the expected soil moisture for a specific area and time. The proposed model can be expressed as follows:

$$\bar{h}_{j,m} = d_{0,j,m} + d_{1,j,m}\bar{r}_{j,m} + d_{2,j,m}\bar{g}_{j,m} + d_{3,j,m}v_j + d_{4,j,m}s_j + d_{5,j,m}e_j + \varepsilon_{j,m} \quad (1)$$

where $\bar{r}_{j,m}$ and $\bar{g}_{j,m}$ are the average values of rainfall, and gradient air temperature at the j^{th} grid and the m^{th} month, respectively. The variables v_j, s_j and e_j are the vegetation index, the average slope and the elevation of the j^{th} grid. $\varepsilon_{j,m}$ is the random error with mean zero and constant variance for the j^{th} grid and the m^{th} month. $d_{i,j,m}$ is the i^{th} regression constant associated to the j^{th} grid and the m^{th} month. The average monthly gradient air temperature is defined as follows:

$$\bar{g}_{j,m} = \bar{T}\max_{j,m} - \bar{T}\min_{j,m} \tag{2}$$

Where $\bar{T}\max_{j,m}$ and $\bar{T}\min_{j,m}$ are the average maximum and minimum air temperature for the j^{th} grid and the m^{th} month, respectively. The coefficients of equation (1) will be obtained by using *in-situ* observations that come from 15 fixed stations and about 35 random samples from portable stations.

3.2 Modeling temporal variability.

The short-term stochastic interaction among soil moisture with rainfall and air temperature can be represented by a Transfer Function (TF) model. The concept of a TF model derives from the idea of cause and effect relationship among the input and output variables of a dynamic system. The input variables of the soil moisture system transfer into variations to the output variable of the system. Thus, the input variables are the air temperature and rainfall, and the system response (output) is the soil moisture. Sampling field observations show that the soil moisture is driven by the cumulated rainfall and air temperature when instantaneous rainfall event is not present. However, if a large spell of no rainfall has occurred a significant response on the soil moisture is observed under the presence of an instantaneous precipitation event (see Figure 2). On the other hand, if the soil is almost saturated or if it reaches its hold capacity with large spell of cumulated rainfall then the response of the soil moisture to the next rainfall event is marginal, even if a large rainfall event has occurred (see Figure 3). Figure 2 shows sampling observations obtained at Mayaguez station during the period of June 16-23, 2004 where the sequence of dry spell and an instantaneous rainfall of 0.25 in/hour generates a large response on soil moisture. Figure 3 shows observation obtained from the same station but during the period of July 3-16, 2004 when the cumulated rainfall generates that the soil almost reaches its hold capacity and the response of the soil moisture is very small, even though the instantaneous rainfall event is very large, 1.25 in/hour. These causal relationships show evidence that the dynamics of the soil moisture system can properly be modeled by a TF model.

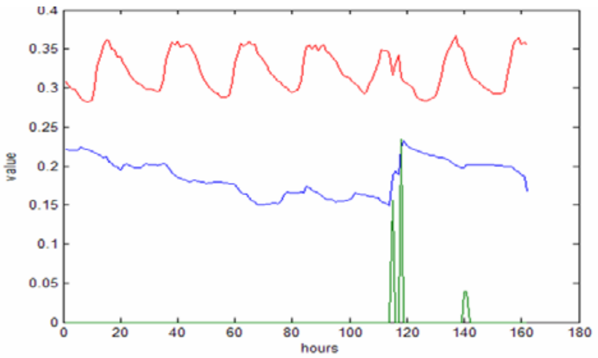

Figure 2. Sampling at Mayaguez station from June 16-23, 2004. (green is rainfall, blue is soil moisture and read is air temperature).

The TF model that represents the soil moisture system for Puerto Rico has several input variables in addition of the noise component. However, because of data limitations only two input variables were considered: rainfall and air temperature. Rainfall was provided in hourly basis and it was noted that the air temperature is also required in hourly basis; however, the available information is in a daily basis. Thus, two TF models are proposed in this work, the first one is used to estimate air temperature and the second one is used to estimate soil moisture. Because of space limitations the air temperature model is not included in this paper.

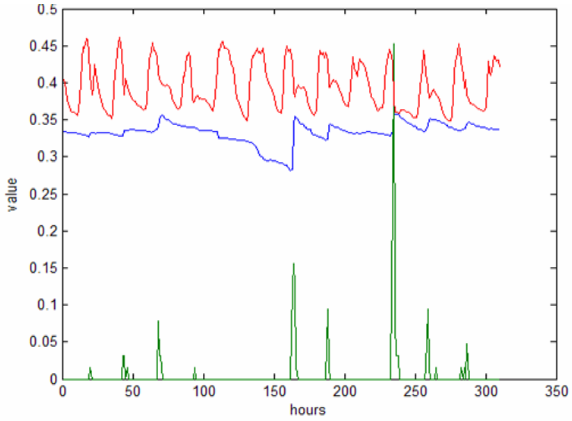

Figure 3. Sampling at Mayaguez station from July 3-16, 2004. (green is rainfall, blue is soil moisture and read is air temperature).

3.3. Soil Moisture Model.

The soil moisture process exhibits short- and long-term memory responses. In this research the long-term memory response is the soil moisture reaction to events that lasted more than 30 days, where as the short-term memory is the response to events that occurred in an interval that is less than 30 days and typically are instantaneous events that seriously impact the soil moisture.

In this research a model is proposed to estimate soil moisture and has four major components: the trend component, rainfall intervention, temperature effects, and noise component. The model can be written as follows:

$$h_{t,j} = n_{t,j} + \frac{\omega_{01,j} - \omega_{11,j}B}{1-\delta_{11}B} R_{t,j} e^{-y_{t,j}} + \frac{\omega_{02,j} - \omega_{12,j}B}{1-\delta_{12}B-\delta_{22}B^2} T_{t,j} + \frac{1}{1-\phi B}\varepsilon_{t,j} \tag{3}$$

where

$$n_{t,j} = \bar{h}_{j,m} + b_{1,j} d_{(i)t.j} + b_{2,j} Q_{t,j} + b_{3,j}\tau_{t,j} \tag{4}$$

$$Q_{t,j} = \begin{cases} \ln(y_{t,j}), & y_{t,j} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{5}$$

$$y_{t,j} = \sum_{i=t-\rho}^{t} R_{i,j} \tag{6}$$

$$\tau_{t,j} = \frac{1}{\rho}\sum_{i=t-\rho}^{t} T_{i,j} \tag{7}$$

The variable $h_{t,j}$ is the soil moisture at time t over the j^{th} grid. The variable $n_{t,j}$ is the trend component and represents the dynamic mean of the soil moisture. The trend component has four major sources: the average monthly soil moisture (computed from equation (2)), the 24-hours gradient air temperature, the cumulative rainfall during the last 4 days, and the average air temperature during the last 4 days. The variable $d_{(i)t,j}$ is the gradient air temperature that occurred during the last 24 hours and was defined by equation (2). The 24-hours gradient temperature is essentially the result of the sun energy affected by relative humidity, cloud coverage and wind dynamics. This gradient provides a linear effect to the soil moisture, the larger the gradient the smaller the soil moisture. Equation (4) tries to express this linear relationship. The variable $Q_{t,j}$ is the cumulative rainfall during the last 4 days and increases the soil moisture in a logarithmic manner. Equations (4) and (5) try to express these relationships. The variable $\tau_{t,j}$ represents the moving averages air temperature during the last 4 days, and it is computed at time *t* and at the location *j*. This moving average varies linearly with the soil moisture, the larger the average the smaller the soil moisture. Equation (4) tries to represent this relationship. The parameter ρ represents the moving window size of the memory and in this case $\rho = 96$ hours. The ρ value was estimated by minimizing the sum of square errors. The variable $\varepsilon_{t,j}$ is the noise component of the model.

Equation (3) represents the impulse response function of the soil moisture to an instantaneous rainfall occurrence, i.e., the soil moisture short-term memory response. It should be noted that instantaneous rainfall is multiplied by and exponential term which indicates the history of rainfall process. It can be noted that when the soil reaches its hold capacity the soil moisture response is marginal. On the other hand, when the cumulated rainfall is small then the responds of the soil moisture is large. The overshooting response can be controlled by multiplying the instantaneous rainfall by an exponential function as is shown in equation (3). The parameter $\omega_{01,j}$ represents the proportional increment of soil moisture as a response to an instantaneous occurrence of rainfall. The parameter $\delta_{11,j}$ expresses the exponential decay of the soil moisture effect as a result of instantaneous rainfall event. Although, the rainfall occurs at time *t* its effect remains in the soil and starts disappearing in an exponential fashion. The smaller the delta, $\delta_{11,j}$, the faster the elimination of soil moisture, as shown by equation (8). It should be noted that the absolute value of delta is restricted to be less than one in order to obtain a causal relationship. The exponential decay can be observed in the expansion of the following rational form:

$$\frac{\omega_{01,j}}{1-\delta_{11,j}B} = 1 + \delta_{11,j}B + \delta_{11,j}^2 B^2 + \delta_{11,j}^3 B^3 + \cdots \qquad (8)$$

The third term of equation (3) represents the response of the soil moisture to the impulse of a given temperature value at time *t*. The soil moisture is responding to the current and the last two temperature values. The soil moisture response is dissipating by the influence of the second order system defined by the delta values. In order to be a causal response, the roots of the polynomial in *B* of the denominator on the third term of equation (3) must fall outside of the unit circle. Where the $\phi_{1,j}$ is the autoregressive parameter of the noise component and should be estimated by using observations. Again in order to be a causal process the absolute value of phi must be less than one.

4. PRELIMINARY RESULTS

4.1. Parameter estimation.

The identification of a TF model is a trivial process when there are one input and one output variable. The typical procedure consists of prewhitening the input variable and using the identified filter to process the output variable throughout the filter and then the cross-correlation function between the filtered variables is used to identify the time delay and the polynomial size of numerator and denominator of the impulse response function (Box and Jenkins, 1976; Brockwell and Davis 2002). When two or more input variables are considered the cross-correlation function is inaccurate to identify the time delays and to determine the size of polynomials of impulse response functions. Thus, it was noted that it is more efficient to use an optimization algorithm to identify the model structure and then estimate the parameters of the model. In this work a nonlinear optimization tool was used to identify and estimate the elements of the soil moisture and temperature models.

The implemented procedure is flexible in the sense that a model with a no trivial structure can be estimated by using a direct search in a nonlinear computational approach. The procedure is based on performing the long-term division of the involved polynomials and the series are truncated to include only the significant values. Since the structure of the model is simple and the range of the parameters is small, then an efficient procedure can be implemented to estimate the parameters. The simplex search algorithm has been selected to perform parameter estimation. The simplex search does not require computing derivatives to perform optimization; this algorithm uses a function evaluation procedure taking advantage of the geometry of the optimization surface. This procedure is very efficient especially when the long division and optimization algorithms are programmed in the same software. Matlab provide a robust set of optimization algorithm in a single computer package.

Model identification and parameter estimation was performed and results are summarized in Table 1. This table shows the estimates for the soil moisture model.

Table 1. Parameter estimates for soil moisture model for the Mayaguez station.

Parameter	$\bar{h}_j$	b_1	b_2	b_3	ω_{01}	δ_{11}
Estimate	0.09898	0.0049237	0.0011812	0.0000843	0.2895	0.86923
Parameter	ω_{02}	ω_{12}	ω_{22}	δ_{12}	δ_{22}	ϕ
Estimate	0.00040729	0.00025094	0.00016458	0.37444	0.58163	0.9106

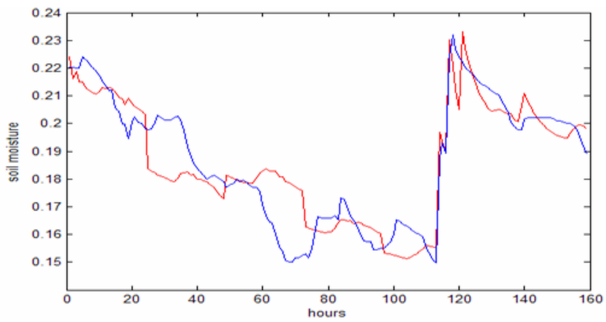

Figure 4. Soil moisture model fitting
(blue the observed and read the estimated soil moisture)

Model fitting performances are shown in Figures 4 and 5. Figure 4 shows the observed and the estimated soil moisture at Mayaguez station. Figure 5 shows the observed and estimated air temperature at Mayaguez station.

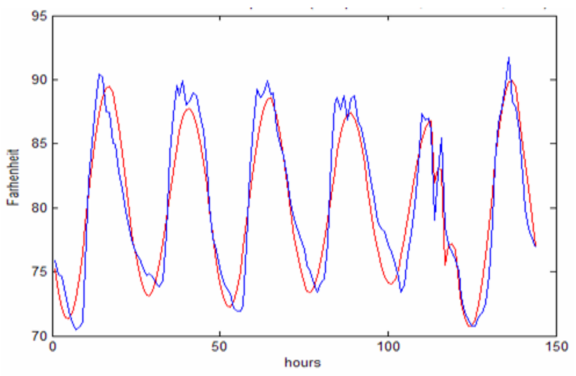

Figure 5. Air temperature model fitting.
(blue the observed and read the estimated air temperature)

4.2. Model Validation

The actual model testing procedure is known as the cross-validation. The performed cross-validation exercise consists of evaluating the fitted model with an independent data set. The selected data set for cross-validation consists of data collected at the same station from the period of July 3-16, 2004. It should be noted that climatic conditions from these two data sets are different and the average level of soil moisture changes from 22% to about 35%. The cross-validation results can be observed in Figures 6 and 7.

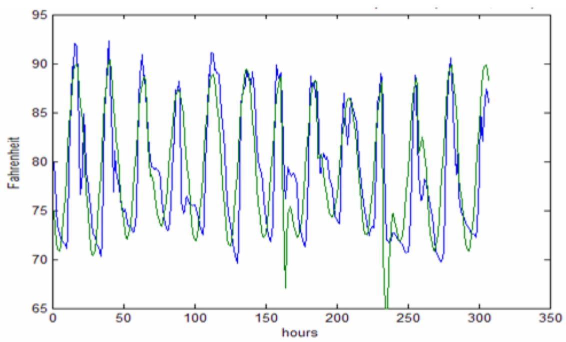

Figure 6. Observed and estimated air temperature model.
(blue the observe and green the estimated air temperature)

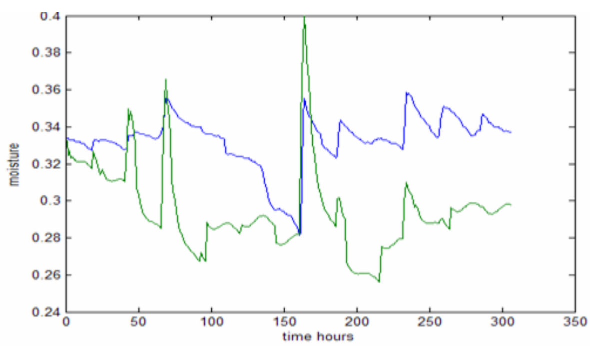

Figure 7. Observed and estimated soil moisture model.
(blue the observe and green the estimated soil moisture)

The average absolute errors of the cross-validation exercise for air temperature and soil moisture models are: 3.21 ^{o}F, and 0.0394 cm^3/cm^3, respectively. The correlation coefficient between the observed and the estimated values are: 0.91 and 0.72. These results show that proposed models are appropriate statistical tools to estimate air temperature and soil moisture in hourly basis. These models can easily be calibrated to other tropical regions

5. COCLUSIONS

A new method is proposed to estimate soil moisture in hourly basis. These estimates can be used to generate the initial condition to run a regional atmosphere model. This methodology can be easily implementing in other climatic conditions, after properly adjusting the parameters of the TF model.

The proposed soil moisture model exhibits short and long term memory. The long term memory is modeled by using climatological patterns, as well as gradient temperature that occurs during the last 24 hours and also using the last 4 days of cumulated rainfall, and the average temperature during the last four days. The instantaneous rainfall and air temperature are used to model the short term memory. Elevation, soil and vegetation classes are inherent into the spatial variability of soil moisture, while precipitation and air temperature are mostly associated to time variability.

The air temperature model generates hourly estimates based on daily air temperature gradient and hourly rainfall.

Preliminary results of model fitting and cross-validation techniques show that the proposed model is a potential tool to estimate the soil moisture at high resolution.

6. ACKNOWLEDGEMENTS

This research has partially been supported by the NOAA-CREST grant NA17AE1625 and NASA-EPSCoR grant NCC5-595, and also by the University of Puerto Rico at Mayaguez. Authors want to appreciate and recognize the funding support from these institutions.

7. REFERENCES

Balsamo, G., Bouyssel. F., and Noilhan, J. (2004). A Simplified bi-dimensional varaitional analysis of soil moisture from screen-level observations in a mesoscale numerical weather-prediction model. *Q.J.R. Meteorol. Soc.* 130, pp 895-915.

Box, G.EP., and Jenkins, G.M., (1976). *Time Series Analysis: Forecasting and Control*, Holden-Day, Oakland, CA.

Brockwell, P.J., and Davis, R.A., (2002). *Introduction to Time Series and Forecasting*, Second Ed., Springer, New York.

Comarazamy, D.E., (2001). *Atmospheric Modeling of the Caribbean Region: Precipitation and Wind Analysis in Puerto Rico for April 1998*, Master Thesis, Department of Mechanical Engineering, University of Puerto Rico, Mayaguez, PR.