Soil moisture estimation using classification and regression trees and neural networks

Gwangseob Kim, Jung-a Park
Department of Civil Engineering, Department of Spatial Information
Kyungpook National University
1370 Sangyeok-dong, Buk-gu, Daegu, 702-701
Korea (South)
kimgs@knu.ac.kr, parkk@knu.ac.kr

Abstract: In this paper, a soil moisture estimation model was developed to calculate the nationwide soil moisture fields using on site soil moisture observation, precipitation, surface temperature, MOD IS NDVI and a data mining technique, Classification And Regression Tree (CART) algorithm and neural networks. The model was applied to the Yong-dam dam basin since the soil moisture observations of the Yong-dam dam basin were reliable. Soil moisture observations of 4 sites were used for the model calibration and that of a site was used for the model validation. Results showed that the soil moisture estimation using a data mining technique and ancillary data allow us to get reasonable soil moisture fields which are suitable for the hydrologic model application.

Key-Words: Soil moisture, Neural Networks, CART, MODIS, Precipitation, NDVI, Surface Temperature

1 Introduction

Soil moisture is a fundamental and important variable to understand hydrological processes and the water-energy balance between land surface and atmosphere [6]. Despite of its importance, there is no long-term measurement of soil moisture in South Korea. Because of such limitations of the lack of in-situ measurements, a downscaling methodology is required to estimate the areal soil moisture based on the field measurements [6].

Neural networks can be applied several hydrologic problems such as precipitation or runoff forecasts, soil moisture retrieval etc. and showed the good applicability for representing the nonlinear systems. Kim and Barros (2001) showed the applicability of neural networks in a runoff forecasting model for ungauged watersheds using a data driven approach [5]. Frate et al. (2003) demonstrated that the NN model is very effective to retrieve soil moisture [3]. It is useful to explain a complex process such as retrieving soil moisture using a simple physical equation. Jiang and Cotton (2004) also suggested that the NN approach can be employed for estimating soil moisture with global scale using remote sensing data [4]. Elshorbagy and Parasuraman (2008) said that they could obtain relatively higher correlation coefficient when using the Higher-Order Neural Networks (HONNs) model than common NN model [2].

In this paper, we tried to estimate the areal soil moisture using a neural networks (NN) model and classification and regression tree (CART) algorithm. The model input variables are consisted of in-situ soil moisture measurements from Korea Water Resources Corporation (KWRC), precipitation, surface temperature were provided by Korean Meteorological Administration (KMA), Normalized Difference Vegetation Index (NDVI) from the MODerate resolution Imaging Spectroradiometer (MODIS).

2 Methodology

Fig.1 represents the conceptual diagram of the soil moisture estimation process. We propose that the spatial downscaling process (1km x 1km) by using precipitation, soil moisture, surface temperature and MODIS NDVI. The CART algorithm builds classification and regression trees for the classification continuous dependent variables. The application of neural networks produced to estimate the soil moisture about the data group of same substance.
2.1 CART

The advantage of CART algorithm is that user understands the rules easily. The Gini index is used to classify the categorical predictor variables and the decrease of variance is used to divide the continuous dependent variables [1]. The algorithm makes two under data groups and each group proceeds again. In this study, a soil moisture estimation method is developed to calculate the nationwide soil moisture fields using CART algorithm. The results using CART algorithm are 5 classification rules and 5 regression equation. Results show that the soil moisture estimates are suitable for the hydrologic model application.

2.2 Neural networks

The neural network methodology is an information processing paradigm that is suitable for the represents the nonlinear relationships between variables. The key element of this paradigm is the novel structure of the information processing system [7]. It is composed of a large number of highly interconnected processing elements working in unison to solve specific problems. Neural networks learn by examples. It is configured for a specific application, such as pattern recognition or data classification, through a learning process. Learning in hydrologic systems involves adjustments to the synaptic connections that exist between the neurons. The model learning is accomplished by training processes between input and output data.

We used the neural networks for soil moisture estimation using classification by CART. The model is 5 nodes of 1 input layer, 5 nodes of 1 hidden layer, 1 node of 1 output layer.

3 Study area

The Yongdam dam basin is located between 36° 00’ - 35° 35’ north latitude and 127° 20’ - 127° 45’ east longitude and its area is about 930.43 km² (Fig.2). Soil moisture networks, Ju-cheon, Bu-gui, Sang-jeon, Ahn-cheon, Cheon-cheon2 sites, are located in the Yongdam dam basin. The resolution of ancillary data is converted to the same spatial resolution of MODIS NDVI (1km x 1km). The observation data were collected between May 16th, 2008 and August 19th, 2008.

4 Result and conclusion

Soil moisture observations of 4 sites, Ju-cheon, Bugui, Sangjeon, Ahncheon, were used for the model calibration (Fig.3) and that of the Cheoncheon2 site was used for the model validation (Fig.4). The average correlation coefficient between soil moisture observation data and model estimates using neural
networks model and CART algorithm are about 0.937 (Table 1).

Fig.3 Estimation of soil moisture in Ju-cheon, Bu-gui, Sang-jeon, Ahn-cheon

Table 1 Test statistics of the soil moisture estimation for each area

<table>
<thead>
<tr>
<th>Area</th>
<th>R</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calibration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jucheon</td>
<td>0.9598</td>
<td>1.3925</td>
<td>1.1528</td>
</tr>
<tr>
<td>Bugui</td>
<td>0.9607</td>
<td>1.0254</td>
<td>0.7462</td>
</tr>
<tr>
<td>Sangjeon</td>
<td>0.9153</td>
<td>1.8783</td>
<td>1.4490</td>
</tr>
<tr>
<td>Ahncheon</td>
<td>0.9417</td>
<td>1.0037</td>
<td>0.7688</td>
</tr>
<tr>
<td>Validation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cheoncheon2</td>
<td>0.9114</td>
<td>3.1918</td>
<td>2.7203</td>
</tr>
</tbody>
</table>

The result shows the applicability of the soil moisture estimation model. Results demonstrates that even though there are limitations of the lack of reliable soil moisture according to the different land use conditions, the soil moisture estimation method using ancillary data, CART and neural networks model should be a reasonable approach to generate 2D soil moisture fields for hydrologic applications since the model provided proper soil moisture field estimations and represented soil moisture behavior well.

Fig.4 Verification of soil moisture model in Cheoncheon2

Fig.4 Sample estimation of soil moisture in Korea (2009)
Fig. 5 Monthly mean of soil moisture in Korea (2009)

References: