A Spatial Downscaling of Soil Moisture from Rainfall, Temperature, and AMSR2 Using a Gaussian-Mixture Nonstationary Hidden Markov Model

Moonhyuk Kwon[1], Hyun-Han Kwon[2]* and Dawei Han[1]

12/11/2017

*Corresponding Author: Hyun-Han Kwon, hkwon@jbnu.ac.kr

[1] Civil and Environmental Engineering, University of Bristol, United Kingdom

[2] Department of Civil Engineering, Chonbuk National University, Jeonju-si, Jeollabuk-do, South Korea

Abstract

A multivariate stochastic soil moisture estimation approach based on a Gaussian-mixture nonstationary hidden Markov model (GM-NHMM) is introduced in this study to spatially disaggregate the AMSR2 soil moisture data for multiple locations in the Yongdam dam watershed in South Korea. Rainfall and air temperature are considered as additional predictors in the proposed modeling framework. In GM-NHMM, a six-state model is constructed with three predictors representing an unobserved state associated with soil moisture. It is clearly seen that the rainfall predictor plays a substantial role in achieving the overall predictability. Using weather variables (i.e., rainfall and temperature) can be effective in picking up some of the predictability of local soil moisture that is not captured by the AMSR2 data. On the other hand, larger scale dynamic features identified from the AMSR2 data seem to facilitate the identification of regional spatial patterns of soil moisture. The efficiency of the proposed model is compared with that of an ordinary regression model (OLR) using the same predictors. The mean correlation coefficient of the proposed model is about 0.78, which is significantly greater than that of the OLR at about 0.49. The proposed GM-NHMM method not only provides a better representation of the observed SM than the OLR model but also preserves the spatial coherence across all stations reasonably well.

Keywords: Soil moisture, stochastic model, AMSR2, spatial downscaling, Gaussian mixture model, and nonstationary hidden Markov model

1. Introduction

Soil moisture (SM) is a key hydrologic state variable for understanding hydrologic processes, including runoff, infiltration, drought, crop growth, and many other phenomena closely related to soil conditions (Albergel et al., 2008; Barrett and Petropoulos, 2013; Brocca et al., 2011; Zhao and Li, 2013), even though the amount of water in the soil profile accounts for less than 0.001 % of the total global water budget (Barrett and Petropoulos, 2013). Thus, acquiring accurate SM information has been a priority in hydrology, meteorology, and climatology. SM data can be obtained in several ways, including in-situ measurements, remote sensing techniques, and soil moisture accounting models. However, each approach has its own advantages and limitations, so different data sources are often integrated to mitigate individual limitations. For more details, the reader is kindly referred to, e.g., Brocca et al., (2017a), Owe et al., (2008), Parajka et al., (2006), and Zhuo and Han, (2016).

In-situ SM observations are generally regarded as the most reliable measurement to validate remotely sensed soil moisture products. The reason for using in-situ SMs is their robustness with respect to the SM retrieved through either remote sensing techniques or soil moisture accounting models. However, in many parts of the world, it remains challenging to collect spatially and temporally suitable ground-based soil moisture data (Brocca et al., 2017b; Peng et al., 2017; Zhuo and Han, 2016). Another issue is that in-situ SM observations are rarely representative of large-scale SM (Griesfeller et al., 2016; Merlin et al., 2012; Reichle et al., 2007), and hydrological analysis is typically conducted on a catchment scale. Considering the limitations of using point-based SM measurements, satellite remote sensing has become an alternative way to monitor SM conditions on a regional scale (Brocca et al., 2011), providing more comprehensive and coherent coverage both spatially and temporally to better understand soil moisture variability in the context of water resource management (Zhao and Li, 2013).

Satellite-based active and passive microwave sensors have the potential advantage of estimating SM spatial fields. Specifically, microwave remote sensing techniques use a longer wavelength than visible and infrared radiation, so they are less affected by cloud coverage, haze, rainfall, and many other weather conditions (Barrett and Petropoulos, 2013; Zhao and Li, 2013). SM data retrieved from various remote sensing sensors, such as Advanced Microwave Scanning Radiometer 2 (AMSR2; JAXA, 2013) , the Soil Moisture Ocean Salinity Satellite (SMOS; Kerr et al., 2012), Soil Moisture Active Passive (SMAP; Das et al., 2011), and the Advanced Scatterometer (ASCAT; Albergel et al., 2008), have become widely available in recent years, providing reasonable accuracy over a wide area with relatively high spatial–temporal resolution. In the past few decades, many studies have explored the accuracy of microwave sensors and improved their applicability to hydrology (Brocca et al., 2017a; Cenci et al., 2016; Parajka et al., 2006; Zhuo and Han, 2016). The challenges associated with these efforts have in turn led to the introduction of new methods to facilitate the suitable use of satellite-based SM measurements with a reasonable degree of accuracy. One major challenge in using satellite SM data for practical applications is their coarse spatial resolution and uncertainties stemming from an inability to resolve sub-grid scale variability. To overcome those limitations, various statistical approaches have used a downscaling framework to achieve a higher spatial resolution for microwave SM data (Merlin et al., 2012; Peng et al., 2016; Piles et al., 2014; Ranney et al., 2015; Zhao and Li, 2013). Those techniques can be divided into two categories: statistical and dynamic downscaling approaches. The downscaling methods also vary depending on the type of data being studied, such as radar, optical/thermal, topography, or soil information data (Peng et al., 2017). Optical/thermal sensor data (generally vegetation index, surface temperature, albedo, etc.) have been widely used to disaggregate the original satellite SM products into fine-scale estimates because they not only provide land surface parameters at higher spatial resolution

(Peng et al., 2016; Piles et al., 2011; Zhao and Li, 2013) but also have a significant correlation with soil moisture (Fang and Lakshmi, 2014; Peng et al., 2015; Srivastava et al., 2013). The basic idea behind these approaches is to build a statistical model (based on the relationship between the satellite SM products and surface parameters) that can simulate SM sequences using given surface parameters as predictors. The most frequently reported practical limitation of this approach is that optical and thermal properties can be obtained only under clear-sky conditions (Djamai et al., 2016; Park et al., 2017). Geo-information data, such as topography, soil attributes, and vegetation, have also been used to disaggregate coarse-scale SM values into fine-scale ones using a regression framework (Busch et al., 2012; Ranney et al., 2015).

During the past few decades, machine learning techniques have been used to spatially downscale satellite-based SM data for enhanced spatial resolution (Im et al., 2016; Park et al., 2017; Srivastava et al., 2013; Xing et al., 2017). For example, Srivastava et al. (2013) tested and compared several machine learning techniques, including an artificial neural network, a support vector machine, and a relevance vector machine, to spatially downscale the SMOS SM data sets. Specifically, they used Moderate Resolution Imaging Spectro-radiometer (MODIS) land surface temperature as auxiliary information in disaggregating the SMOS SM products. Park et al. (2017) developed a downscaling scheme based on a modified regression tree model that combined multiple sensors (AMSR2 and ASCAT) with four other predictors: MODIS land surface temperature, the normalized difference vegetation index, land cover, and a digital elevation model.

However, the existing approaches all largely depend on a linear or nonlinear regression model to spatially downscale the satellite SM products without considering the stochastic nature of soil moisture dynamics. The spatiotemporal dynamics of soil moisture content result from complicated and mutually related processes of hydro-meteorological elements, such as subsurface flow, lateral flow, infiltration, precipitation, climate, and soil (Botter et al., 2007; Ridolfi et al., 2003). The influence of spatiotemporal variability in precipitation and temperature on the slow-varying behavior of basin-scale SM can be better represented within a stochastic modeling framework (Botter et al., 2007). Recently, a stochastic downscaling technique, a nonstationary Markov model with a gamma (or exponential) distribution, has been widely used in both hydrology and meteorology (Cioffi et al., 2017; Khalil et al., 2010; Mehrotra and Sharma, 2005; Robertson et al., 2004). The stochastic downscaling approaches have been mainly used for rainfall simulation at multiple locations (Cioffi et al., 2017; Khalil et al., 2010; Kwon et al., 2011, 2009, Mehrotra and Sharma, 2010, 2006; Robertson et al., 2004; Stehlík and Bárdossy, 2002); they have rarely been applied to SM data by means of a multivariate downscaling framework (no literature regarding SM has been found).

Given this background, we here investigate the following questions:

(1) Can daily soil moisture sequences conditional on intraseasonal variability in climate be effectively clustered and discretized as a small set of states? In addition, can the identified states of daily soil moisture and their transition probability be explicitly considered to better characterize soil moisture dynamics?

(2) Is it desirable to use a nonstationary stochastic model that considers climate variables such as precipitation, temperature, and satellite-based soil moisture products as predictors? Does a combination of climate variables and satellite-based soil moisture better inform simulations?

(3) Can the proposed stochastic modeling framework be applied to simultaneously simulate the daily sequences of soil moisture at multiple locations on a watershed scale?

We here propose a multivariate Gaussian mixture nonstationary hidden Markov model (GM-NHMM), which is primarily based on Hughes et al., (1999) and Yoo et al., (2015), to investigate those questions, with the intention of providing a practical tool for the estimation of daily soil moisture on the watershed scale for use in agricultural drought monitoring and hydrologic modeling. In-situ SM observations at multiple stations are here used as a dependent variable, and both air temperature and rainfall, as well as the AMSR2 data, are considered as predictors. The proposed downscaling approach is applied to the Yongdam dam watershed in South Korea. The performance of the proposed downscaling scheme is then validated with 6 in-situ observations through a cross-validation procedure.

2. Study Area and Data

2.1 Site description and observation data

In this study, we apply the spatial downscaling approach to satellite SM measurements for multiple stations in the Yongdam dam watershed in southwestern Korea (35.6°–36.0°N latitude and 127.3°–127.7°E longitude). Most of the in-situ SM observation stations in this catchment are in the forest, and the dominant soil type consists of sand (62.1 %), loam (20.7 %), and silt (17.0 %). The average annual precipitation and air temperature during the investigation period (2014–2016) were 1,147 mm and 11.4˚C, respectively. Figure 1 shows the study area and six in-situ soil moisture stations where precipitation data were also measured (http://www.ydew.or.kr/kdrum/main/main.do). Here, precipitation data are averaged over the entire region. Additionally, air temperature (available for download from https://data.kma.go.kr/cmmn/main.do) was measured at the Jangsu weather station operated by the Korea Meteorological Administration (https://web.kma.go.kr/eng/). The soil moisture observation network covers a drainage area of 930 km^2 with elevation ranging from 209 to 1,588 m a.s.l. The Korea Water Resources Corporation has continuously recorded in-situ SM observations measured at half-hourly time interval since 2014 using a time domain reflectometer (TDR; Topp et al., 1980). The specifications for the observation sites used in this study are given in Table 1. Depth-averaged SM representing the mean soil moisture content in the soil layer 0-60cm were used for subsequent study.

[Insert Figure 1 and Table 1]

2.2 Satellite data

AMSR2 is on the GCOM-W1 satellite launched by the Japan Aerospace Exploration Agency (JAXA) in May 2012. As a follow-on instrument to AMSR-E, which was operated from 2002 to 2012, the AMSR2 is a passive microwave sensor that measures the brightness temperature at seven different frequencies between 6.9 GHz and 89.0 GHz (Imaoka et al., 2010). It is widely acknowledged that microwaves measured from space are severely contaminated by radio frequency interference (RFI) effects (Liu et al., 2011; Njoku et al., 2005; Zeng et al., 2015). Therefore, a new 7.3-GHz channel was added to the AMSR2 to identify and address RFI signals. Additionally, the AMSR2 has a larger antenna (2.0 m) than the AMSR-E (1.6 m) to provide a higher spatial resolution. The AMSR2 provides geophysical products such as integrated water vapor, integrated cloud liquid water, precipitation, sea surface temperature, sea surface wind speed, sea ice concentration, snow depth, and soil moisture content (Imaoka et al., 2010). For this study, we obtained the AMSR2 L3 SM products, derived from the JAXA algorithm with 10 km spatial resolution, from the distributor’s website (https://gcom-w1.jaxa.jp/auth.html). Readers are referred to Koike (2013) for a detailed description of the retrieval algorithm. The AMSR2 sensor provides volumetric SM content from 0 to 60 % with 1–2 day revisit frequency. The daily AMSR2 SM data are extracted by averaging the ascending (1:30 pm) plus descending (1:30 am) overpasses over a three-year period (2014–2016).

3. Methodology

3.1 Multivariate Gaussian-Mixture Nonstationary Hidden Markov Model

In this study, we propose a novel approach to stochastic modeling of soil moisture at multiple locations that takes into account a set of exogenous variables: rainfall, temperature, and satellite information. Here, we briefly present only the relevant details of a multivariate hidden Markov model described elsewhere (Khalil et al., 2010; Kwon et al., 2011, 2009; Robertson et al., 2004; Yoo et al., 2015) and primarily based on Hughes et al. (1999). Figure 2 shows schematically the procedure of this study.

[Insert Figure 2]

A hidden Markov model (HMM) describes a process in which part of the system dynamics is hidden, and some other part of the system can be partially explained by other observations. The HMM uses a Markovian process and a set of stochastic functions to generate plausible sequences for a given time series based on stochastic sampling from probability distributions conditioned on different hidden states (Daniel and Martin, 2017; Gharhramani, 2001).

Let $\mathbf{SM}_t$ be an M-dimensional vector of in-situ soil moisture measurements corresponding to M-stations at time *t*. Let $\mathbf{SM}_{1:T} = \left(\mathbf{SM}_1, \ldots, \mathbf{SM}_T\right)$ denote a sequence of soil moisture with length *T*. The sequence of observed soil moisture measurements $\mathbf{SM}_{1:T}$ is presumed to be governed by a Markov property with the corresponding sequence $\mathbf{S}_{1:T} = (S_1, ..., S_T)$ of a finite number of hidden states, taking on values *k* in {1, *K*}. A joint distribution of $\mathbf{SM}_{1:T}$ and $\mathbf{S}_{1:T}$ can be explicitly defined by taking the two conditional independence (CI) assumptions (Bishop, 2006; Smyth et al., 1997), as formulated below.

First, assume that the sequence of hidden states $\mathbf{S}_{1:T}$ follows the stationary Markovian process that relies only on the values of the previous *k*-th order states. Obviously, the probability distribution for the current hidden state with a first-order model ($k = 1$) can be represented as equation (1) (Rabiner, 1989).

$$p(S_1,...,S_T) = p(S_1)\prod_{t=2}^{T} p(S_t|S_{t-1}) \tag{1}$$

For a stationary HMM, $p(S_1)$ is the initial-state probability vector, and the state-transition probability matrix of a hidden state can be denoted as $p(S_t \mid S_{t-1}) = \{\gamma_{ij}\}$, $1 \le i, j \le K$. Second, assume that individual in-situ observations $\mathbf{SM}_t$ are conditionally independent of all other variables in the model given the current state $\mathbf{S}_t$ (Robertson et al., 2006; Smyth et al., 1997).

$$p(\mathbf{SM}_{1:T}|\mathbf{S}_{1:T}) = \prod_{t=1}^{T} p(\mathbf{SM}_t|S_t) \tag{2}$$

The joint probability of the soil moisture data $\mathbf{SM}_{1:T}$ and the hidden states can then be formulated as equation (3) (Kwon et al., 2011, 2009; Robertson et al., 2006).

$$p(\mathbf{SM}_{1:T}, \mathbf{S}_{1:T}) = \left[p(S_1)\prod_{t=2}^{T} p(S_t|S_{t-1}) \right]\left[\prod_{t=1}^{T} p(\mathbf{SM}_t|S_t) \right] \tag{3}$$

Soil moisture values, $\mathbf{SM}_t^M$, at time *t* for M stations are assumed to be conditionally independent of one another given the hidden state S_t . Here, spatial dependencies across multiple stations are indirectly modeled by the hidden state variable, as described in equation (4). Note that a more advanced approach to modeling the spatial structure of $\mathbf{SM}_t$ across M sites could be of particular interest in situations with high spatial correlation. More specifically, the spatial coherence across stations is considered by assigning a state to each day, representing the spatial structure of soil moisture (Kwon et al., 2011, 2009; Robertson et al., 2006).

$$p(\mathbf{SM}_t | S_t) = \prod_{m=1}^{M} p(\mathbf{SM}_t^m | S_t) \tag{4}$$

The probability density function for the emission distribution at an individual soil moisture station $\mathbf{SM}_t^m$ is assumed to be approximated by a Gaussian mixture function of C components for non-zero soil moisture , with $p_{i,m,c} \geq 0$ and $\sum_{c=1}^{C} p_{i,m,c} = 1$ for all $m = 1,...,M$ and $i = 1,...,K$, as follows:

$$p(\mathbf{SM}_t^m = r | S_t = i) = \sum_{c=1}^{C} p_{i,m,c} N(\mu_{i,m,c}, \sigma_{i,m,c}) \tag{5}$$

Here, μ and σ are the mean and variance of the Gaussian distribution, respectively, and the set of parameters associated with the transition matrix, the initial states, and the parameters of emission distribution are simultaneously estimated from the observed soil moisture data using the expectation-maximization (EM) algorithm in an optimization context. Gaussian mixture models are a statistical tool for multimodal density estimation (Bilmes, 1998; Gauvain and Lee, 1994). Gaussian mixture models have been used for soil moisture modeling (Ryu and Famiglietti, 2005; Verhoest et al., 2015; Vilasa et al., 2017), and have also been used extensively in hydrologic field (Carreau et al., 2009; Lakshmanan and Kain, 2010; Rings et al., 2012; Yoo et al., 2015). Unlike the HMM, the underlying assumption of the GM-NHMM is that soil moisture is generated in a stochastic process that sequentially depends on a set of predictors represented by rainfall, temperature, and the satellite product. Specifically, NHMMs can be constructed by imposing a non-stationarity assumption on the probability distribution of the response variables, which in turn depends on observed independent variables (Hughes et al., 1999; Hughes and Guttorp, 1994; Kwon et al., 2011). This soil moisture model can be substantially expanded by introducing a mixture model for soil moisture content into the existing HMM. In this study, we use a mixture of Gaussians to describe soil moisture at multiple stations in a stochastic framework to account for soil moisture variability. Again, we use the EM algorithm to estimate the parameters (Dempster et al., 1977).

The concept of CI can be illustrated as edges in a directed acyclic graph of the GM-NHMM, as shown in Figure 3.Suppose $\mathbf{X}_{1:T} = (\mathbf{X}_1,...,\mathbf{X}_T)$ is a set of predictors representing soil moisture, such as rainfall, temperature, and AMSR2 soil moisture data. In a GM-NHMM, the state-transition matrix is assumed to be nonstationary, and therefore, the dynamic evolution of transition probability is a function of multivariate exogenous variables, $\mathbf{X}_{1:T}$. The GM-NHMM is then written as equation (6) (Khalil et al., 2010; Kirshner, 2005; Kwon et al., 2011, 2009).

$$p(\mathbf{SM}_{1:T}, \mathbf{S}_{1:T} | \mathbf{X}_{1:T}) = \left[p(S_1 | \mathbf{X}_1) \prod_{t=2}^{T} p(S_t | S_{t-1}, \mathbf{X}_t) \right] \left[\prod_{t=1}^{T} p(\mathbf{SM}_t | S_t) \right] \quad (6)$$

[Insert Figure 3]

In this study, we consider uniform priors, thus leading to the maximum likelihood approach to estimating a set of model parameters, $\arg\max_{\Theta} P(\mathbf{SM}|\mathbf{X}, \Theta)$. Again, note that the proposed model assumes that the observed soil moisture sequences from different years are conditionally independent. Under the GM-NHMM, the log-likelihood function $LL(\Theta)$ of the observed soil moisture data at multiple locations can be written as follows (Khalil et al., 2010):

$$\begin{aligned} LL(\Theta) &= \ln p(\mathbf{SM}_{1:T} | \mathbf{X}_{1:T}, \Theta) \\ &= \sum \ln \sum_{S_{1:T} \in [1,\dots,K]^T} \left[p(S_1 | \mathbf{X}_1, \Theta) \prod_{t=2}^{T} p(S_t | S_{t-1}, \mathbf{X}_t, \Theta) \right] \left[\prod_{t=1}^{T} p(\mathbf{SM}_t | S_t, \Theta) \right] \end{aligned} \quad (7)$$

The parameter values cannot be obtained analytically, so we use the EM algorithm to estimate the value of the parameter vector Θ by maximizing equation (7). The EM algorithm is an iterative method for maximizing the likelihood function in a parameter space

Θ. Finally, the state evolutions over time in equation (6) are simulated by a multinomial logistic regression as follows (Kirshner, 2005; Kwon et al., 2011):

:

$$p(S_t = \beta | S_{t-1} = \alpha, \mathbf{X}_t = \mathbf{x}) = \frac{\exp(\omega_{\alpha\beta} + \xi'_\beta \mathbf{x})}{\sum_{k=1}^{K} \exp(\omega_{\alpha\beta} + \xi'_k \mathbf{x})} \quad (8)$$

All the parameters ω are real, and ξ is a vector in a multi-dimensional parameter space. Here, the prime denotes the transpose of the vector. Parameterization and prediction using NHMM are well documented in the statistical literature and, thus, need not be elaborated here. For more detailed description of the NHMM algorithm the reader is referred to Daniel and Martin, (2017), Gharhramani, (2001), Rabiner, (1989), and Robertson et al., (2003).

3.2 Ordinary Linear Regression (OLR)

As a comparison to the GM-NHMM, we applied a linear regression model with the same input variables used in the GM-NHMM to downscale the AMSR2 SM product for each station *m*. Here, each parameter (**β**) is obtained from the least squares method. The linear combination of predictors for estimating soil moisture can be written as follows:

$$SM_t^m = (\beta_0^m + \beta_1^m \times R_t + \beta_2^m \times Tp_t + \beta_3^m \times ST_t) \quad (9)$$

where *SM, R*, and Tp are in-situ SM, rainfall, and temperature data, respectively, and *ST* is 10km AMSR2 SM data. Again note that predictor variables used here are averaged over the entire region.

4. Results and Discussion

4.1 Quantile Mapping for Bias Correction

The mismatch in spatial-temporal resolution between AMSR2 SM products and in-situ observations causes inevitable systematic biases. Therefore, a statistical bias correction approach is commonly applied to remove the systematic bias from the satellite SM data for subsequent use in either downscaling or SM modeling (Kornelsen and Coulibaly, 2015). We used a quantile mapping method in which the cumulative density function of the AMSR2 data is matched with that of the in-situ SM observations. In this study, t location-scale (eq. (10)) and gamma (eq. (11)) distributions were selected to fit the AMSR2 and in-situ soil moisture data, respectively, based on the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), respectively, as summarized in Table 2. As shown in Figure 4, the bias-corrected AMSR2 SM data exhibit enhanced variability and match well with the in-situ observations. We used these bias-corrected AMSR2 SM products for our subsequent analyses.

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\sigma\sqrt{\nu\pi}\Gamma(\nu/2)}\left[\frac{\nu+(\frac{x-\mu}{\sigma})^2}{\nu}\right]^{-(\frac{\nu+1}{2})} \tag{10}$$

$$f(y) = \frac{y^{\theta-1}e^{-\tau y}}{\tau^{\theta}\Gamma(\theta)} \tag{11}$$

where μ, σ, and ν are the location, scale, and shape parameters of the t location-scale distribution, respectively, and $\Gamma(\bullet)$ is the gamma function. θ and τ are the shape and scale parameters of the gamma distribution, respectively.

[Insert Figure 4 and Table 2]

4.2. Predictor Selection

It is important to identify a suitable set of predictors that consistently influences the response variables. However, in a regression model, using several predictors can cause serious overfitting, which results in unrealistic predictions (Khalil et al., 2010). For a parsimonious model, we consider only three predictors, daily rainfall, air temperature, and AMSR2 data, and we initially evaluate the cross-correlations for all lagged orders. The correlations are statistically significant and strongly persistent, as illustrated in Figure 5. Note that here the values are averaged over the entire watershed for a representation. The lag-1 correlation is high for daily rainfall, and the correlations appear to be consistent with the lag in the temperature and AMSR2 data. Therefore, we retained a set of 1 day time-lagged values for the three predictors to simulate soil moisture content in the proposed GM-NHMM.

[Insert Figure 5]

4.3 Stochastic Modeling of Soil Moisture Using GM-NHMM

The performance of the GM-NHMM is greatly influenced by the number of hidden states used to represent an unobserved SM state. In this study, we estimated the number of hidden states by recursively maximizing the log-likelihood (or minimizing the BIC) in the context of optimization. The maximized log-likelihoods for each state are shown in Figure 6, together with the minimized BIC. As shown in Figure 6(a), the log-likelihoods gradually increase with the number of hidden states, but we could not clearly identify an inflection point on the curve to determine the optimal number of hidden states. On the other hand, the BIC decreases rapidly at 4 states, and the degree of reduction beyond 6 hidden states is negligible. Therefore, we used 6 hidden states to build our stochastic soil moisture model at multiple locations.

[Insert Figure 6]

For the selected 6 hidden states, the most likely temporal sequences can be efficiently determined using the Viterbi algorithm (Viterbi, 1967), which calculates the probability of that a hidden state will occur as well as the probability that it will transition to another state at a certain date. The estimated temporal sequences of observed SM are illustrated in Figure 7, and considerable inter-annual and intraseasonal variability are clearly identified. The Viterbi analysis is a useful tool not only to capture intra- and inter-annual variability but also to quantify its intensity. More specifically, changes in the intra-annual sequence of observed SM states are shown along a horizontal line, and inter-annual variability is represented by a vertical line.

[Insert Figure 7]

The degree of soil wetness and the frequencies associated with hidden states are presented in Figure 8. Figure 8 (a) shows boxplots representing station-averaged SM data corresponding to each state in 2014–2016. Clearly, the lower states are closely related to drier soil conditions, and vice versa. Moreover, the median SM value increases largely as a function of the number of states (i.e., from 21% (state 1) to 29.3 % (state 6)). The percentage of days falling into the 6 hidden states for SM data across 6 stations are 14.4, 14.8, 19.5, 19.8, 20.3, and 11.1 %. States 3–5 occur dominantly during the entire period, accounting for 59.6 %, whereas state 6, representing the wettest soil condition, has the lowest frequency, as shown in Figure 8(b). The estimated transition probabilities of the NHMM are shown in Table 3. Note that the state-transition in the GM-NHMM is assumed to be nonstationary and informed by exogenous variables, such as rainfall and temperature. As expected, the self-transition probability (more likely to stay in the current state than to transition to a new state) is noticeably high, with state 1 being the most persistent (0.93) and state 6 being the least persistent (0.70).

[Insert Figure 8 and Table 3]

The temporal patterns of the simulated SM and the in-situ observations at 6 stations are illustrated in Figure 9. To verify the potential of the model to reproduce the variability observed in the SM data, we conducted 100 simulations. The results show a fairly good agreement with the in-situ observations. Here, the proposed GM-NHMM is illustrated across the entire period (2014–2016), along with the OLR model, in Figure 10. The GM-NHMM comprises the vector of observed SM data from 6 stations (as dependent variables) given a vector of observed covariates (as independent variables). For comparison, we built an OLR model for each station using the ordinary least square method for the best-fit model of SM data. Summary statistics for the comparison between the GM-NHMM and OLR are presented in Table 4, and the GM-NHMM outperforms the OLR model. More specifically, the SM data simulated through the GM-NHMM agree well with the in-situ observations, with correlation coefficients (*r*) ranging from 0.73 to 0.81 (mean: 0.78), and a root mean square error (RMSE) ranging from 1.47 % to 2.62 % (mean: 2.06 %), whereas the OLR has much lower performance (mean *r*: 0.49 and mean RMSE: 2.58 %).

[Insert Figure 9-10 and Table 4]

To further ensure that the proposed modeling scheme can predict SM, we subdivided the SM data into different groups and then validated the proposed GM-NHMM using a cross-validation scheme. We partitioned a sample of SM data into three different subsets corresponding to the year of interest, trained the model on one subset, and then validated the model with the remaining data. In other words, a set of parameters for the GM-NHMM is estimated in the training period, and the identified parameters are then used to simulate SM for the validation. We performed 100 simulations for each cross-validation partition for both the training and validation periods. As a representative case, the simulated SM values for 6 stations are compared with the values observed at those stations for the training period (2014–2015) and the testing period (2016) in Figure 11. The SM data are reasonably well reproduced by the proposed GM-NHMM for both the training and testing phases. The results of the cross-validation using the GM-NHMM for the different partitions are summarized in Table 5. We considered three goodness- of- fit measures, correlation coefficient (*r*), RMSE, and bias, in evaluating the models. During the training periods, the 6-station averaged correlation coefficient values range from 0.72 to 0.80, whereas during the validation period, the *r* values show slightly lower correlations than during the training period. However, the GM-NHMM can clearly generate the intraseasonal sequence of daily SM fairly well, and other measures also show reasonable performance at multiple locations, leading to higher correlations with the observed SM data. The RMSE and bias values are also generally better for the training period than the validation period.

[Insert Figure 11 and Table 5]

For a multisite SM simulator, it is of particular importance to correctly reproduce the spatial coherence of daily SM across multiple stations. Therefore, we estimated the spatial correlations of the sequence of daily SM and compared them with the observed values. As shown in Figure 12, the spatial correlations across the stations are reasonably well reproduced by proposed GM-NHMM model.

[Insert Figure 12]

Table 6 shows the results of applying the GM-NHMM with different combinations of predictors to examine the contribution of the AMSR2 SM data to the proposed model. The use of rainfall and temperature without the AMSR2 data (case-1) led to a slightly lower correlation coefficient of 0.73, compared to the results obtained with all three predictors shown in Table 5. On the other hand, there was no significant change in the correlation coefficient of 0.63 when we used rainfall alone as a predictor (case-2). Furthermore, we found a similar trend in our cross-validation analysis. Therefore, the 1 day time-lagged rainfall data might be the main factor in properly reproducing SM dynamics. Nonetheless, combining rainfall with temperature and AMSR2 still yielded the highest correlation with the in-situ observations.

[Insert Table 6]

5. Concluding Remarks

We have here presented a stochastic soil moisture estimation model based on a GM-NHMM to spatially disaggregate AMSR2 SM data at multiple locations in the context of downscaling. Given the close relationship with SM, we considered both rainfall and air temperature as potential predictors in the proposed stochastic downscaling model. We used 1 day time-lagged values for the three predictors to simulate SM in the proposed GM-NHMM model. Before applying the proposed downscaling scheme, we used the quantile mapping approach to reduce the systematic bias in the AMSR2 SM products, and we then used those bias-corrected AMSR2 SM products for subsequent analyses. In GM-NHMM terms, we formulated a six-state model with three predictors representing an unobserved SM state based on the BIC. The temporal sequences of unobserved hidden states and the dynamic evolution of transition probability were estimated by the Viterbi algorithm. Consequently, the proposed GM-NHMM was applied to simulate fine-resolution SM products in a multivariate framework. We compared our results with in-situ observations from the Yongdam dam watershed in South Korea. The key results obtained are summarized as follows.

1. The estimated small set of hidden states that most likely corresponds to localized soil moisture dynamics is effectively captured and accounts for a certain fraction of the soil moisture process, which improves understanding of the intraseasonal and inter-annual variability of SM dynamics. Based on the identified state transition-probability matrix, self-transitions are more significant than the probability of transitioning to other states, indicating that the states seem to be persistent over time due to the slow-varying behavior of basin-scale SM (Botter et al., 2007).
2. Given the relatively short length of the in-situ SM time series data, we considered a cross-validation performance assessment of the simulations. The rainfall predictor plays a substantial role in achieving overall predictability. Adding temperature and AMSR2 data as predictors improves the fit to the SM data. Therefore, weather variables (i.e., rainfall and temperature) could be effective in picking up some of the predictability of local SM that is not captured by AMSR2 data. On the other hand, large-scale dynamic features identified in remote-sensed SM data seem to facilitate the identification of other SM states with well-defined regional spatial patterns. The results presented here illustrate the potential of a stochastic model with a climate-predictor-based forecast. However, the relatively small improvement in forecast skill that the AMSR2 SM products offer in the model suggests that the AMSR2 data might not sufficiently reflect the regional or seasonal characteristics of this study area.
3. We compared the efficiency of the proposed model with that of an ordinary regression model using the same predictors. The mean correlation coefficient for the GM-NHMM obtained by averaging over all the stations is about 0.78, which is significantly greater than that of the OLR, about 0.23. The proposed model also yields a noticeable reduction

in RMSE. Moreover, the proposed GM-NHMM method not only provides a better representation of the observed SM than the OLR model but also preserves spatial coherence across all the stations, which is a fundamentally important property in describing the spatial pattern of soil moisture and its association with runoff on a catchment scale.

Our main contributions in this study are our insights into the soil moisture process and its potential predictability, leading to the way for more applications in hydrologic studies. We expect that future work will address this study's shortcomings with respect to the use of satellite-based products and predictor selection and further investigate cross-validation assessment of forecasts for different regions over a longer period of record, which are required to support these applications.

Appendix A

List of Abbreviations	
AIC	Akaike information criterion
AMSR2	Advanced Microwave Scanning Radiometer 2
ASCAT	Advanced Scatterometer
BIC	Bayesian information criterion
CI	Conditional independence
EM	Expectation-maximization
GM-NHMM	Gaussian mixture nonstationary hidden Markov model
HMM	Hidden Markov model
JAXA	Japan Aerospace Exploration Agency
MODIS	Moderate Resolution Imaging Spectroradiometer
OLR	Ordinary regression model
r	Correlation coefficient
RMSE	Root mean square error
SM	Soil moisture
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture Ocean Salinity Satellite

Appendix B

List of Symbols	
Tp	Temperature
R	Rainfall
ST	AMSR2 SM data
$\Gamma(\bullet)$	Gamma function
θ	Shape parameter of the gamma distribution
ν	Shape parameter of the t location-scale distribution
τ	Scale parameter of the gamma distribution
$\mathbf{SM}_t^M$	M-dimensional vector of in-situ soil moisture measurements at time t.
$\mathbf{S}_{1:T}$	Finite number of hidden states
X	A set of predictors
$LL(\Theta)$	Log-likelihood function
μ	Location parameter of the t location-scale distribution
σ	Scale parameter of the t location-scale distribution

Acknowledgement

This research was supported by a grant (17AWMP-B121100-02) from the Water Management Research Program funded by Ministry of Land, Infrastructure and Transport of Korean government.

References

Albergel, C., Rüdiger, C., Carrer, D., Calvet, J.-C., Fritz, N., Naeimi, V., Bartalis, Z., Hasenauer, S., 2008. An evaluation of ASCAT surface soil moisture products with in-situ observations in southwestern France. Hydrol. Earth Syst. Sci. Discuss. 5, 2221–2250. doi:10.5194/hessd-5-2221-2008

Barrett, B., Petropoulos, G., 2013. Satellite Remote Sensing of Surface Soil Moisture. Remote Sens. Energy Fluxes Soil Moisture Content 85–120. doi:doi:10.1201/b15610-6

Bilmes, J.A., 1998. A Gentle Tutorial of the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models, International Computer Science Institute. doi:10.1016/S0550-3213(97)00753-0

Bishop, C.M., 2006. Pattern Recognition and Machine Learning. Springer. doi:10.1117/1.2819119

Botter, G., Porporato, A., Rodriguez-Iturbe, I., Rinaldo, A., 2007. Basin-scale soil moisture dynamics and the probabilistic characterization of carrier hydrologic flows: Slow, leaching-prone components of the hydrologic response. Water Resour. Res. 43, 1–14. doi:10.1029/2006WR005043

Brocca, L., Ciabatta, L., Massari, C., Camici, S., Tarpanelli, A., 2017a. Soil Moisture for Hydrological Applications: Open Questions and New Opportunities. Water 9, 140. doi:10.3390/w9020140

Brocca, L., Crow, W.T., Ciabatta, L., Massari, C., De Rosnay, P., Enenkel, M., Hahn, S., Amarnath, G., Camici, S., Tarpanelli, A., Wagner, W., 2017b. A Review of the Applications of ASCAT Soil Moisture Products. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 10, 2285–2306. doi:10.1109/JSTARS.2017.2651140

Brocca, L., Hasenauer, S., Lacava, T., Melone, F., Moramarco, T., Wagner, W., Dorigo, W., Matgen, P., Martínez-Fernández, J., Llorens, P., Latron, J., Martin, C., Bittelli, M., 2011. Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. Remote Sens. Environ. 115, 3390–3408. doi:10.1016/j.rse.2011.08.003

Busch, F.A., Niemann, J.D., Coleman, M., 2012. Evaluation of an empirical orthogonal function-based method to downscale soil moisture patterns based on topographical attributes. Hydrol. Process. 26, 2696–2709. doi:10.1002/hyp.8363

Carreau, J., Naveau, P., Sauquet, E., 2009. A statistical rainfall-runoff mixture model with heavy-tailed components. Water Resour. Res. 45. doi:10.1029/2009WR007880

Cenci, L., Laiolo, P., Gabellani, S., Campo, L., Silvestro, F., Delogu, F., Boni, G., Rudari, R., 2016. Assimilation of H-SAF Soil Moisture Products for Flash Flood Early Warning Systems. Case Study: Mediterranean Catchments. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. PP, 5634–5646. doi:10.1109/JSTARS.2016.2598475

Cioffi, F., Conticello, F., Lall, U., Marotta, L., Telesca, V., 2017. Large scale climate and rainfall seasonality in a Mediterranean Area: Insights from a non-homogeneous Markov model applied to the Agro-Pontino plain. Hydrol. Process. 31, 668–686. doi:10.1002/hyp.11061

Daniel, J., Martin, H., 2017. Speech and Language Processing, in: Ch 9. Stanford University. doi:10.1007/978-1-61779-400-1_22

Das, N.N., Entekhabi, D., Njoku, E.G., 2011. An Algorithm for Merging SMAP Radiometer and Radar Data for High Resolution Soil Moisture Retrieval. IEEE Trans. Geosci. Remote Sens. in press, 1–9. doi:10.1109/TGRS.2010.2089526

Dempster, A.P., Laird, N.M., Rubin, D.B., 1977. Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc. Ser. B Methodol. 39, 1–38. doi:http://dx.doi.org/10.2307/2984875

Djamai, N., Magagi, R., Goïta, K., Merlin, O., Kerr, Y., Roy, A., 2016. A combination of DISPATCH downscaling algorithm with CLASS land surface scheme for soil moisture estimation at fine scale during cloudy days. Remote Sens. Environ. 184, 1–14. doi:10.1016/j.rse.2016.06.010

Fang, B., Lakshmi, V., 2014. Soil moisture at watershed scale: Remote sensing techniques. J. Hydrol. 516, 258–272. doi:10.1016/j.jhydrol.2013.12.008

Gauvain, J.-L., Lee, C.-H., 1994. Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains. IEEE Trans. Speech Audio Process. 2, 291–298. doi:10.1109/89.279278

Gharhramani, Z., 2001. An Introduction to Hidden Markov Models and Bayesian Networks. J. Pattern Recognit. Artif. Intell. 15, 9–42.

Griesfeller, A., Lahoz, W. a., Jeu, R.A.M. d., Dorigo, W., Haugen, L.E., Svendby, T.M., Wagner, W., 2016. Evaluation of satellite soil moisture products over Norway using ground-based observations. Int. J. Appl. Earth Obs. Geoinf. 45, 155–164. doi:10.1016/j.jag.2015.04.016

Hughes, J.P., Guttorp, P., 1994. A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. Water Resour. Res. 30, 1535–1546. doi:10.1029/93WR02983

Hughes, J.P., Guttorp, P., Charles, S.P., 1999. A non-homogeneous hidden Markov model for precipitation occurrence. J. R. Stat. Soc. Ser. C (Applied Stat. 48, 15–30. doi:10.1111/1467-9876.00136

Im, J., Park, S., Rhee, J., Baik, J., Choi, M., 2016. Downscaling of AMSR-E soil moisture with MODIS products using machine learning approaches. Environ. Earth Sci. 75, 1–19. doi:10.1007/s12665-016-5917-6

Imaoka, K., Kachi, M., Fujii, H., Murakami, H., Hori, M., Ono, A., Igarashi, T., Nakagawa, K., Oki, T., Honda, Y., Shimoda, H., 2010. Global change observation mission (GCOM) for monitoring carbon, water cycles, and climate change. Proc. IEEE 98, 717–734. doi:10.1109/JPROC.2009.2036869

Kerr, Y.H., Waldteufel, P., Richaume, P., Wigneron, J.P., Ferrazzoli, P., Mahmoodi, A., Bitar, A. Al, Cabot, F., Gruhier, C., Juglea, S.E., Leroux, D., Mialon, A., Delwart, S., 2012. The SMOS Soil Moisture Retrieval Algorithm. Geosci. Remote Sens. 50, 1384–1403. doi:10.1109/TGRS.2012.2184548

Khalil, A.F., Kwon, H.-H., Lall, U., Kaheil, Y.H., 2010. Predictive downscaling based on non-homogeneous hidden Markov models. Hydrol. Sci. J. 55, 333–350. doi:10.1080/02626661003780342

Kirshner, S., 2005. Modeling of multivariate time series using hidden Markov models. University of California, Irvine.

Koike, T., 2013. Description of GCOM-W1 AMSR2 Soil Moisture Algorithm, in: Descriptions of GCOM-W1 AMSR2 Level 1R and Level 2 Algorithms. Japan Aerospace Exploration Agency Earth Observation Research Center, p. 8.1-8.13.

Kornelsen, K.C., Coulibaly, P., 2015. Reducing multiplicative bias of satellite soil moisture retrievals. Remote Sens. Environ. 165, 109–122. doi:10.1016/j.rse.2015.04.031

Kwon, H.H., Lall, U., Obeysekera, J., 2009. Simulation of daily rainfall scenarios with interannual and multidecadal climate cycles for South Florida. Stoch. Environ. Res. Risk Assess. 23, 879–896. doi:10.1007/s00477-008-0270-2

Kwon, H.H., Sivakumar, B., Moon, Y. Il, Kim, B.S., 2011. Assessment of change in design flood frequency under climate change using a multivariate downscaling model and a precipitation-runoff model. Stoch. Environ. Res. Risk Assess. 25, 567–581. doi:10.1007/s00477-010-0422-z

Lakshmanan, V., Kain, J.S., 2010. A Gaussian Mixture Model Approach to Forecast Verification. Weather Forecast. 25, 908–920. doi:10.1175/2010WAF2222355.1

Liu, Y.Y., Parinussa, R.M., Dorigo, W.A., De Jeu, R.A.M., Wagner, W., M. Van Dijk, A.I.J., McCabe, M.F., Evans, J.P., 2011. Developing an improved soil moisture dataset by blending passive and active microwave satellite-based retrievals. Hydrol. Earth Syst. Sci. 15, 425–436. doi:10.5194/hess-15-425-2011

Mehrotra, R., Sharma, A., 2010. Development and application of a multisite rainfall stochastic downscaling framework for climate change impact assessment. Water Resour. Res. 46, 1–17. doi:10.1029/2009WR008423

Mehrotra, R., Sharma, A., 2006. A nonparametric stochastic downscaling framework for daily rainfall at multiple locations. J. Geophys. Res. Atmos. 111, 1–16. doi:10.1029/2005JD006637

Mehrotra, R., Sharma, A., 2005. A nonparametric nonhomogeneous hidden Markov model for downscaling of multisite daily rainfall occurrences. J. Geophys. Res. D Atmos. 110, 1–13. doi:10.1029/2004JD005677

Merlin, O., Rüdiger, C., Al Bitar, A., Richaume, P., Walker, J.P., Kerr, Y.H., 2012. Disaggregation of SMOS soil moisture in Southeastern Australia. IEEE Trans. Geosci. Remote Sens. 50, 1556–1571. doi:10.1109/TGRS.2011.2175000

Njoku, E.G., Ashcroft, P., Chan, T.K., Li, L., 2005. Global survey and statistics of radio-frequency interference in AMSR-E land observations. IEEE Trans. Geosci. Remote Sens. 43, 938–946. doi:10.1109/TGRS.2004.837507

Owe, M., de Jeu, R., Holmes, T., 2008. Multisensor historical climatology of satellite-derived global land surface moisture. J. Geophys. Res. Earth Surf. 113, 1–17. doi:10.1029/2007JF000769

Parajka, J., Naeimi, V., Blöschl, G., Wagner, W., Merz, R., Scipal, K., 2006. Assimilating scatterometer soil moisture data into conceptual hydrologic models at the regional scale. Hydrol. Earth Syst. Sci 10, 353–368. doi:10.5194/hessd-2-2739-2005

Park, S., Park, S., Im, J., Rhee, J., Shin, J., Park, J., 2017. Downscaling GLDAS Soil Moisture Data in East Asia through Fusion of Multi-Sensors by Optimizing Modified Regression Trees. Water 9, 332. doi:10.3390/w9050332

Peng, J., Loew, A., Merlin, O., Verhoest, N.E.C., 2017. A review of spatial downscaling of satellite remotely sensed soil moisture. Rev. Geophys. 1–26. doi:10.1002/2016RG000543

Peng, J., Loew, A., Zhang, S., Wang, J., 2016. Spatial downscaling of global satellite soil moisture data using temperature vegetation dryness index. IEEE Trans. Geosci. Remote Sens. 1, 558–566.

Peng, J., Niesel, J., Loew, A., 2015. Evaluation of soil moisture downscaling using a simple thermal-based proxy-the REMEDHUS network (Spain) example. Hydrol. Earth Syst. Sci. 19, 4765–4782. doi:10.5194/hess-19-4765-2015

Piles, M., Camps, A., Vall-llossera, M., Corbella, I., Panciera, R., Rüdiger, C., Kerr, Y.H., Walker, J., 2011. Downscaling SMOS-Derived Soil Moisture Using MODIS Visible / Infrared Data. IEEE Trans. Geosci. Remote Sens. 49, 3156–3166.

Piles, M., Sánchez, N., Vall-Llossera, M., Camps, A., Martínez-Fernandez, J., Martinez, J., Gonzalez-Gambau, V., 2014. A downscaling approach for SMOS land observations: Evaluation of high-resolution soil moisture maps over the Iberian peninsula. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 3845–3857. doi:10.1109/JSTARS.2014.2325398

Rabiner, L.R., 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE. doi:10.1109/5.18626

Ranney, K.J., Niemann, J.D., Lehman, B.M., Green, T.R., Jones, A.S., 2015. A method to downscale soil moisture to fine resolutions using topographic, vegetation, and soil data. Adv. Water Resour. 76, 81–96. doi:10.1016/j.advwatres.2014.12.003

Reichle, R.H., Koster, R.D., Liu, P., Mahanama, S.P.P., Njoku, E.G., Owe, M., 2007. Comparison and assimilation of global soil moisture retrievals from the Advanced Microwave Scanning Radiometer for the Earth Observing System (AMSR-E) and the Scanning Multichannel Microwave Radiometer (SMMR). J. Geophys. Res. Atmos. 112, 1–14. doi:10.1029/2006JD008033

Ridolfi, L., D’Odorico, P., Porporato, A., Rodriguez-Iturbe, I., 2003. Stochastic soil moisture dynamics along a hillslope. J. Hydrol. 272, 264–275. doi:10.1016/S0022-1694(02)00270-6

Rings, J., Vrugt, J.A., Schoups, G., Huisman, J.A., Vereecken, H., 2012. Bayesian model averaging using particle filtering and Gaussian mixture modeling: Theory, concepts, and simulation experiments. Water Resour. Res. 48. doi:10.1029/2011WR011607

Robertson, A.W., Kirshner, S., Smyth, P., 2004. Downscaling of daily rainfall occurrence over Northeast Brazil using a hidden Markov model. J. Clim. 17, 4407–4424. doi:10.1175/JCLI-3216.1

Robertson, A.W., Kirshner, S., Smyth, P., Charles, S.P., Bates, B.C., 2006. Subseasonal-to-interdecadal variability of the Australian monsoon over North Queensland. Q. J. R. Meteorol. Soc. 132, 519–542. doi:10.1256/qj.05.75

Robertson, A.W., Sergey, K., Padhraic, S., 2003. Hidden Markov models for modeling daily rainfall occurrence over Brazil, Technical Report UCI-ICS 03-27. Information and Computer Science University of Califonia, Irvine.

Ryu, D., Famiglietti, J.S., 2005. Characterization of footprint-scale surface soil moisture variability using Gaussian and beta distribution functions during the Southern Great Plains 1997 (SGP97) hydrology experiment. Water Resour. Res. 41, 1–13. doi:10.1029/2004WR003835

Smyth, P., Heckerman, D., Jordan, M.I., 1997. Probabilistic Independence Networks for Hidden Markov Probability Models. Neural Comput. 9, 227–269. doi:10.1162/neco.1997.9.2.227

Srivastava, P.K., Han, D., Ramirez, M.R., Islam, T., 2013. Machine Learning Techniques for Downscaling SMOS Satellite Soil Moisture Using MODIS Land Surface Temperature for Hydrological Application. Water Resour. Manag. 27, 3127–3144. doi:10.1007/s11269-013-0337-9

Stehlík, J., Bárdossy, A., 2002. Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation. J. Hydrol. 256, 120–141. doi:10.1016/S0022-1694(01)00529-7

Topp, G.C., Davis, J.L., Annan, A.P., 1980. Electromagnetic Determination of Soil Water Content: Measruements in Coaxial Transmission Lines. Water Resour. Res. 16, 574–582. doi:10.1029/WR016i003p00574

Verhoest, N.E.C., Van Den Berg, M.J., Martens, B., Lievens, H., Wood, E.F., Pan, M., Kerr, Y.H., Al Bitar, A., Tomer, S.K., Drusch, M., Vernieuwe, H., De Baets, B., Walker, J.P., Dumedah, G., Pauwels, V.R.N., 2015. Copula-based downscaling of coarse-scale soil moisture observations with implicit bias correction. IEEE Trans. Geosci. Remote Sens. 53, 3507–3521. doi:10.1109/TGRS.2014.2378913

Vilasa, L., Miralles, D.G., de Jeu, R.A.M., Dolman, A.J., 2017. Global soil moisture bimodality in satellite observations and climate models. J. Geophys. Res. Atmos. 122, 4299–4311. doi:10.1002/2016JD026099

Viterbi, A., 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. IEEE Trans. Inf. Theory 13, 260–269. doi:10.1109/TIT.1967.1054010

Xing, C., Chen, N., Zhang, X., Gong, J., 2017. A Machine Learning Based Reconstruction Method for Satellite Remote Sensing of Soil Moisture Images with In Situ Observations. Remote Sens. 9, 484. doi:10.3390/rs9050484

Yoo, J., Kwon, H.H., So, B.J., Rajagopalan, B., Kim, T.W., 2015. Identifying the role of typhoons as drought busters in South Korea based on hidden Markov chain models. Geophys. Res. Lett. 42, 2797–2804. doi:10.1002/2015GL063753

Zeng, J., Li, Z., Chen, Q., Bi, H., Qiu, J., Zou, P., 2015. Evaluation of remotely sensed and reanalysis soil moisture products over the Tibetan Plateau using in-situ observations. Remote Sens. Environ. 163, 91–110. doi:10.1016/j.rse.2015.03.008

Zhao, W., Li, A., 2013. A downscaling method for improving the spatial resolution of AMSR-E derived soil moisture product based on MSG-SEVIRI data. Remote Sens. 5, 6790–6811. doi:10.3390/rs5126790

Zhuo, L., Han, D., 2016. Could operational hydrological models be made compatible with satellite soil moisture observations? Hydrol. Process. 30, 1637–1648. doi:10.1002/hyp.10804

Table 1. Specification and characteristics of soil observation sites in the Yongdam dam watershed.

Site	Elevation (m a.s.l)	Longitude (º)	Latitude (º)	Annual rainfall (mm/yr)	Observation depth (cm)	Land Cover
SM & Rainfall						
Station 1	313	127.55	35.87	1,107	10, 20, 40, 60	Forest
Station 2	330	127.43	35.97	1,224	10, 20, 40, 60	Forest
Station 3	396	127.4	35.86	1,191	10, 20, 40, 60	Forest
Station 4	334	127.49	35.8	1,120	10, 20, 40, 60	Agriculture
Station 5	453	127.63	35.81	1,049	10, 20, 40, 60	Agriculture
Station 6	409	127.51	35.68	1,193	10, 20, 40, 60	Forest
Temperature						
Jangsu	406	127.52	35.66	-	-	-

Table 2. BIC and AIC scores with respect to distribution models.

In-situ			AMSR2		
Distribution	BIC	AIC	Distribution	BIC	AIC
Gamma	44,677	44,663	t-location scale	31,445	31,425
Log-logistic	45,051	45,037	Log-logistic	32,316	32,303
Normal	45,128	45,114	Gamma	36,550	36,536
t-location scale	45,137	45,116	Weibull	38,680	38,666
Weibull	45,259	45,246	Normal	43,660	43,646

Table 3. Transition probability matrix of 6 hidden states for soil moisture at 6 stations in the Yongdam watershed.

Site	Station 1	Station 2	Station 3	Station 4	Station 5	Station 6
Station 1	0.93	0.01	0.05	0.00	0.00	0.01
Station 2	0.02	0.90	0.01	0.04	0.00	0.04
Station 3	0.04	0.03	0.88	0.00	0.03	0.02
Station 4	0.00	0.04	0.00	0.92	0.02	0.02
Station 5	0.00	0.00	0.07	0.05	0.79	0.09
Station 6	0.00	0.00	0.00	0.00	0.30	0.70

Table 4. Comparison between in-situ and simulated SM.

Site	BC AMSR2		GM-NHMM		OLR	
	r	RMSE (%)	*r*	RMSE (%)	*r*	RMSE (%)
Station 1	0.34	4.55	0.79	2.62	0.49	3.36
Station 2	0.10	4.07	0.78	2.02	0.55	2.42
Station 3	0.31	2.55	0.73	1.52	0.49	1.83
Station 4	0.38	2.54	0.81	1.47	0.54	1.86
Station 5	0.17	4.34	0.79	2.22	0.41	2.95
Station 6	0.10	4.93	0.79	2.50	0.48	3.06
Average	0.23	3.83	0.78	2.06	0.49	2.58

Table 5. Comparison between in-situ and simulated SM.

Site	Training (2014–2015)			Validation (2016)			Training (2015–2016)			Validation (2014)			Training (2014, 2016)			Validation (2015)		
	r	RMSE (%)	Bias	*r*	RMSE (%)	Bias	*r*	RMSE (%)	Bias	*r*	RMSE (%)	Bias	*r*	RMSE (%)	Bias	*r*	RMSE (%)	Bias
Station 1	0.79	2.69	0.43	0.80	2.47	0.32	0.83	2.14	0.32	0.62	3.34	0.26	0.77	2.79	0.72	0.68	3.14	1.53
Station 2	0.79	2.10	0.66	0.75	1.85	0.19	0.86	1.65	0.37	0.63	2.57	1.36	0.73	2.25	0.83	0.86	2.26	0.69
Station 3	0.76	1.49	0.28	0.67	1.57	0.10	0.75	1.45	0.25	0.69	1.65	0.00	0.68	1.70	0.44	0.74	1.80	1.02
Station 4	0.80	1.57	0.24	0.83	1.24	0.06	0.73	1.44	0.20	0.74	1.86	0.06	0.76	1.57	0.35	0.59	1.91	0.99
Station 5	0.79	2.37	0.67	0.78	1.89	0.23	0.76	2.18	0.48	0.60	2.63	0.19	0.65	2.54	0.87	0.66	3.47	2.11
Station 6	0.83	2.23	0.68	0.73	2.96	1.04	0.88	1.88	0.41	0.68	2.43	0.30	0.71	2.80	1.01	0.86	2.60	1.27
Average	0.79	2.08	0.49	0.76	2.00	0.33	0.80	1.79	0.34	0.66	2.41	0.36	0.72	2.28	0.70	0.73	2.53	1.27

Table 6. Comparison of *r* values with respect to different combinations of predictors.

Sta. No	Modeling	Cross Validation					
	Entire period (2014–2016)	Training (2014–2015)	Validation (2016)	Training (2015–2016)	Validation (2014)	Training (2014, 2016)	Validation (2015)
(Case 1) Predictors: Rainfall, Temperature							
Station 1	0.75	0.76	0.72	0.76	0.64	0.71	0.59
Station 2	0.73	0.74	0.70	0.84	0.60	0.56	0.74
Station 3	0.63	0.70	0.48	0.69	0.66	0.52	0.65
Station 4	0.78	0.78	0.79	0.70	0.71	0.81	0.66
Station 5	0.73	0.75	0.68	0.70	0.54	0.64	0.42
Station 6	0.75	0.77	0.72	0.87	0.60	0.55	0.66
Average	0.73	0.75	0.68	0.76	0.63	0.63	0.62
(Case 2) Predictor: Rainfall							
Station 1	0.78	0.78	0.79	0.72	0.81	0.80	0.70
Station 2	0.39	0.45	0.22	0.23	0.51	0.38	0.47
Station 3	0.62	0.66	0.53	0.57	0.67	0.56	0.62
Station 4	0.81	0.80	0.84	0.79	0.83	0.84	0.70
Station 5	0.62	0.61	0.64	0.49	0.75	0.67	0.53
Station 6	0.57	0.61	0.50	0.49	0.67	0.59	0.63
Average	0.63	0.65	0.58	0.55	0.71	0.64	0.61

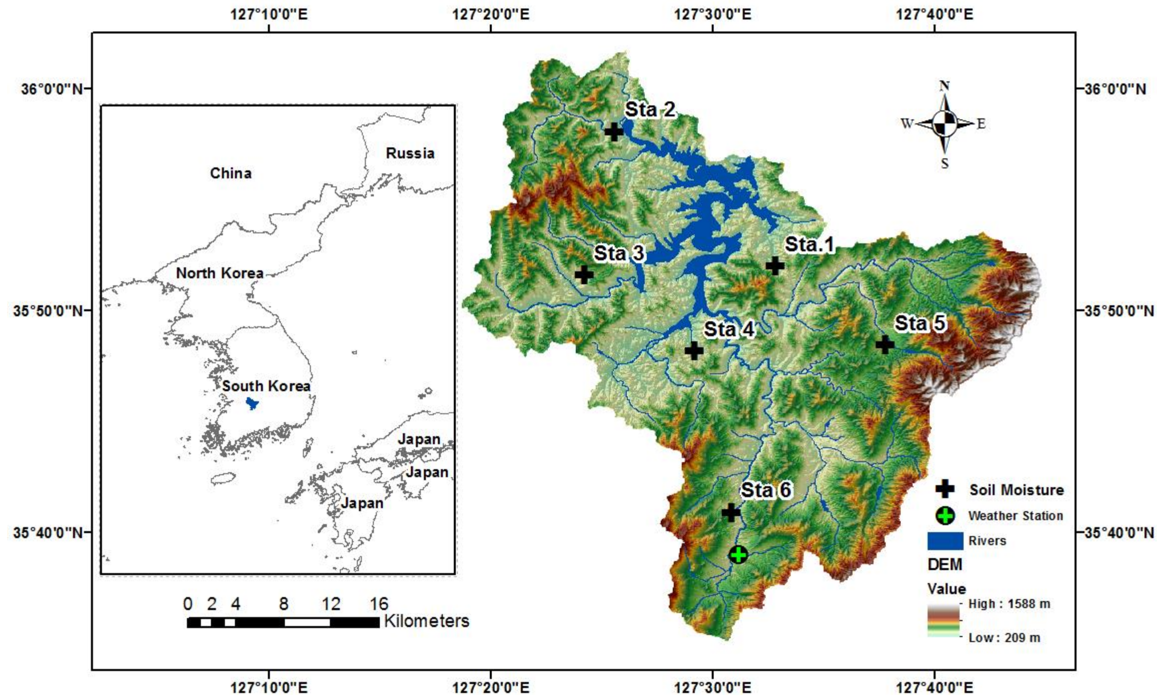


Figure 1. The study site with topography and observation stations.

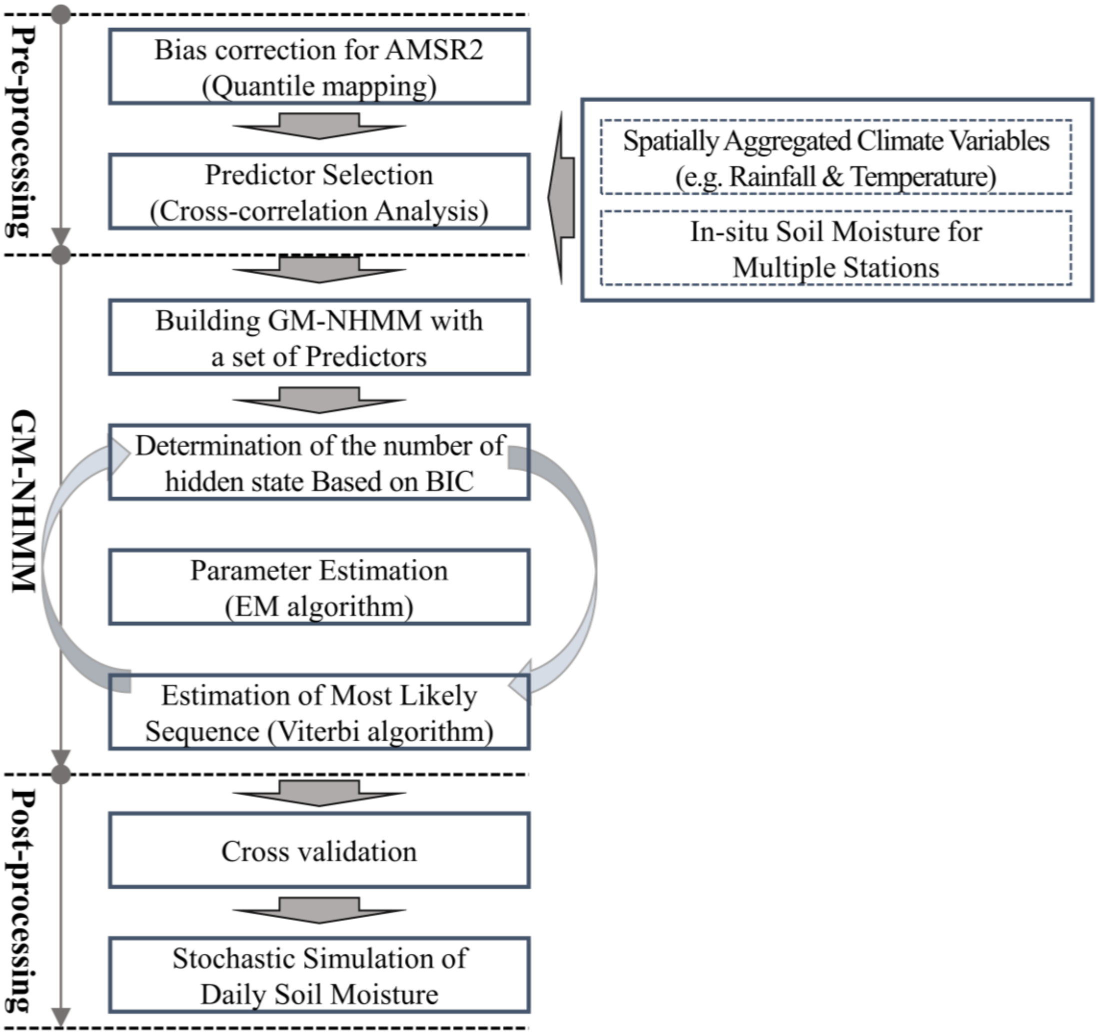


Figure 2. Schematic diagram representing the processing steps.

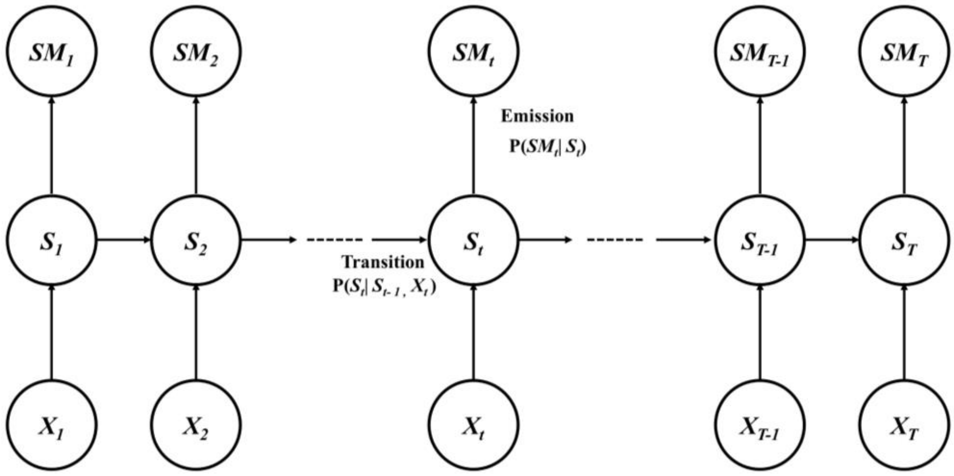


Figure 3. Graphical model representation of nonhomogeneous hidden Markov model. Here, *SM*, *S*, *X* indicate soil moisture, hidden state and exogenous variable (i.e., rainfall, temperature, and AMSR2), respectively.

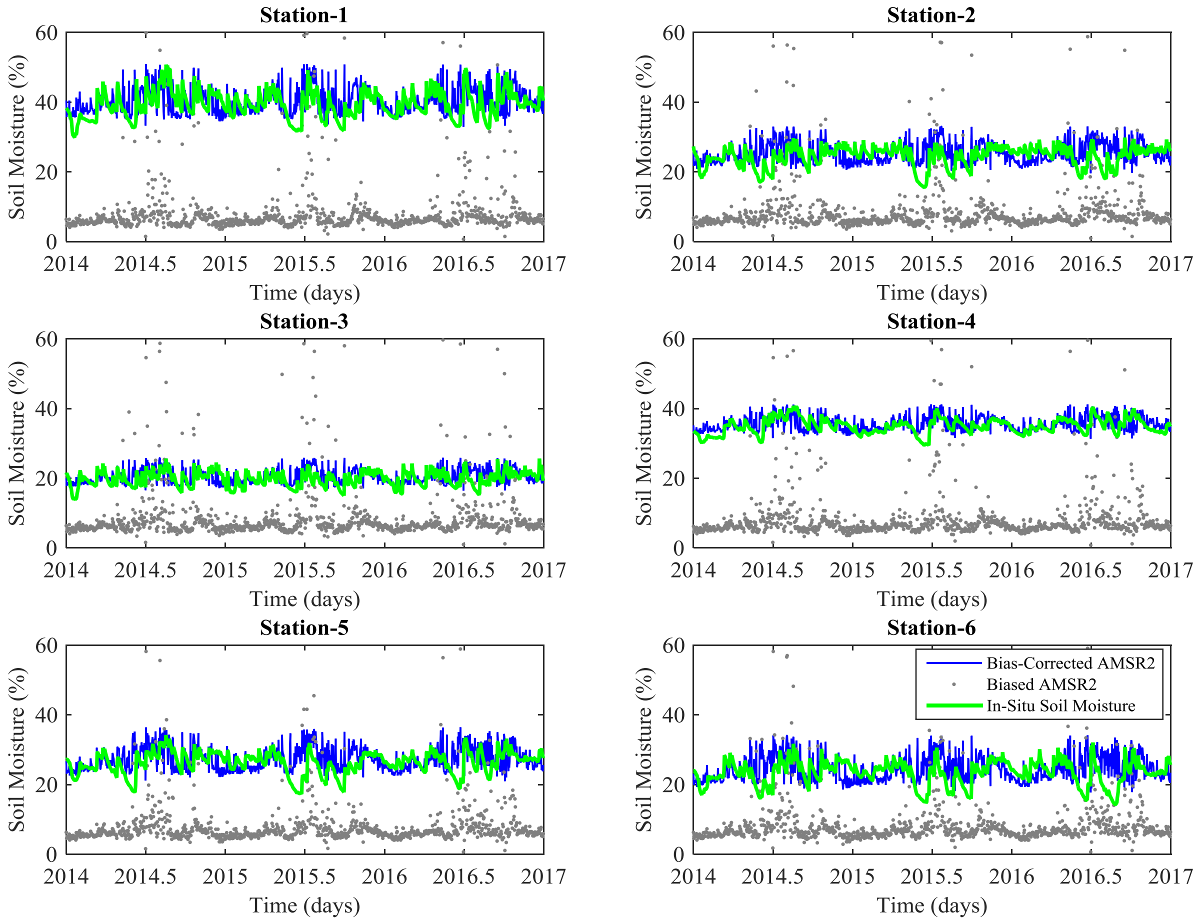


Figure 4. Bias-uncorrected and bias-corrected AMSR2 SM time series data with in-situ observations during the study period, 2014–2016.

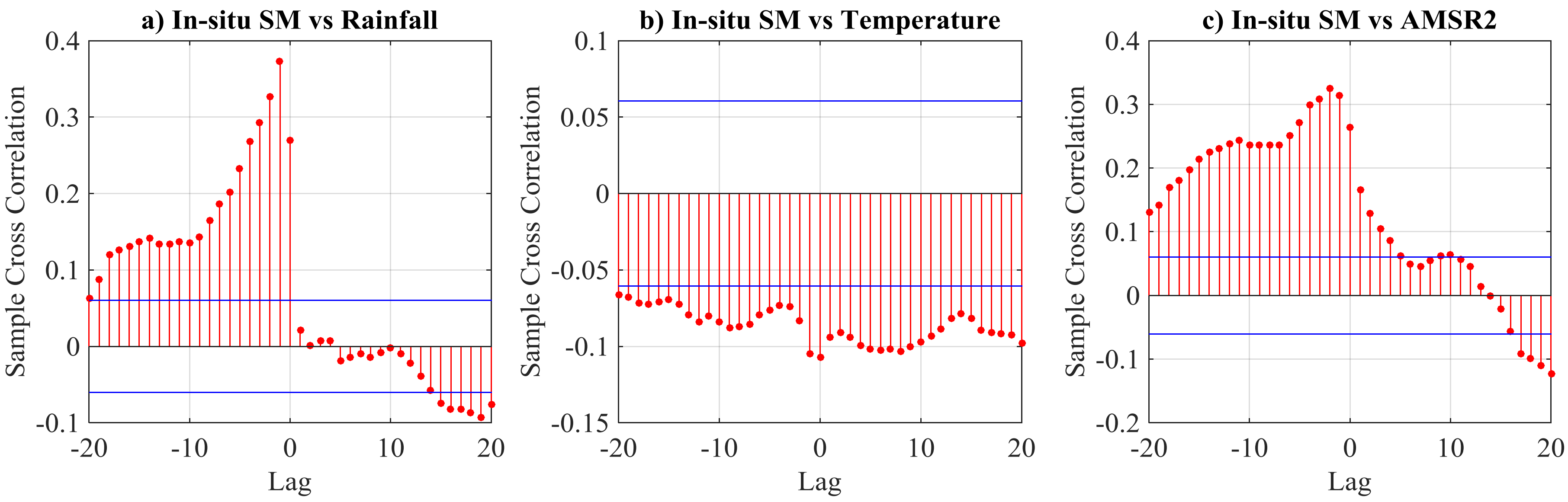


Figure 5 Sample cross correlation between the in-situ soil moisture and a set of predictors: a) rainfall, b) temperature, and c) AMSR2 soil moisture data. All values are averaged over the entire watershed.

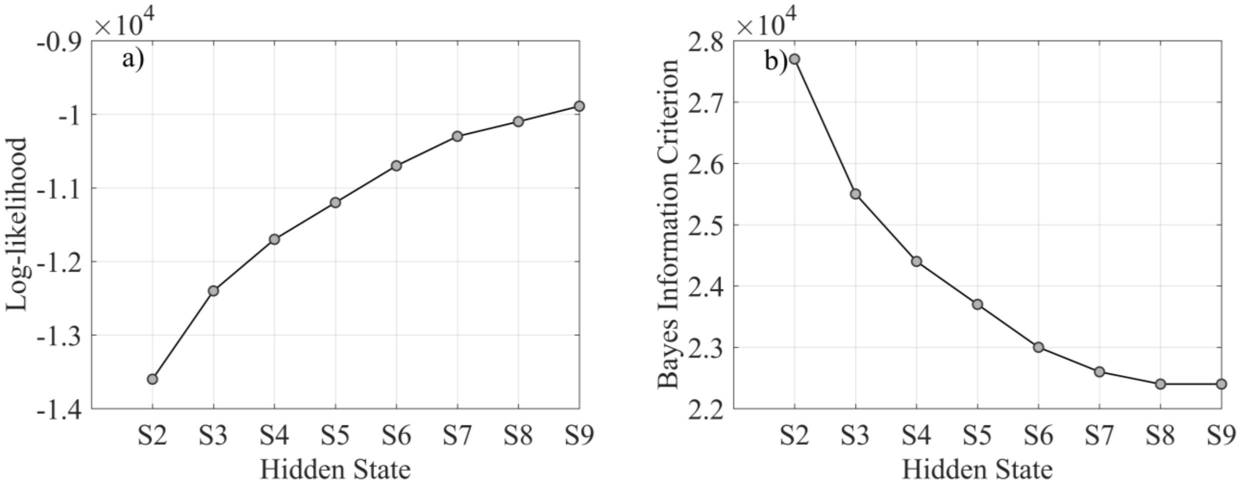


Figure 6. Log-likelihood and BIC values in terms of hidden states.

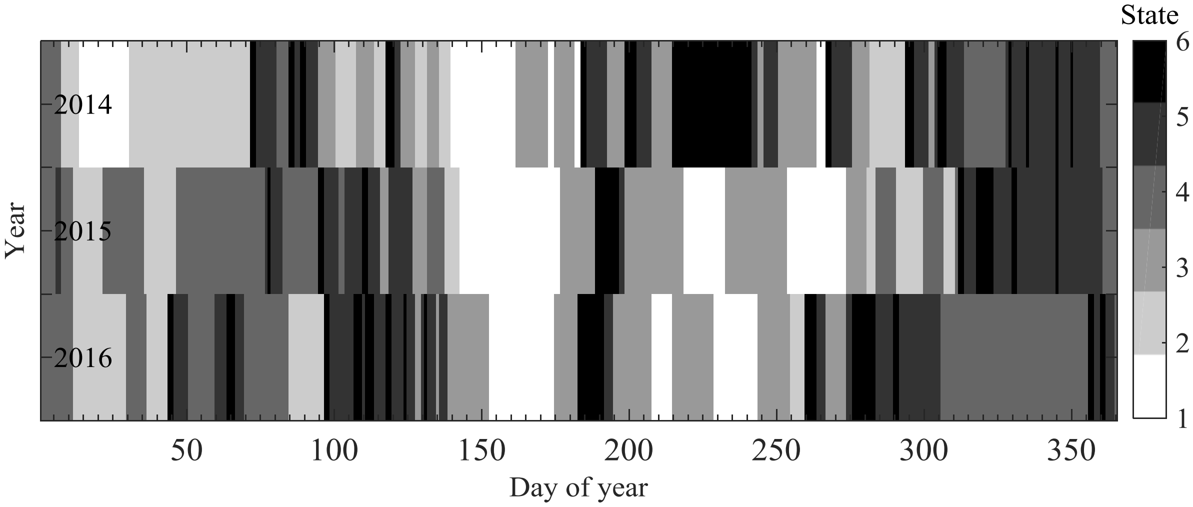
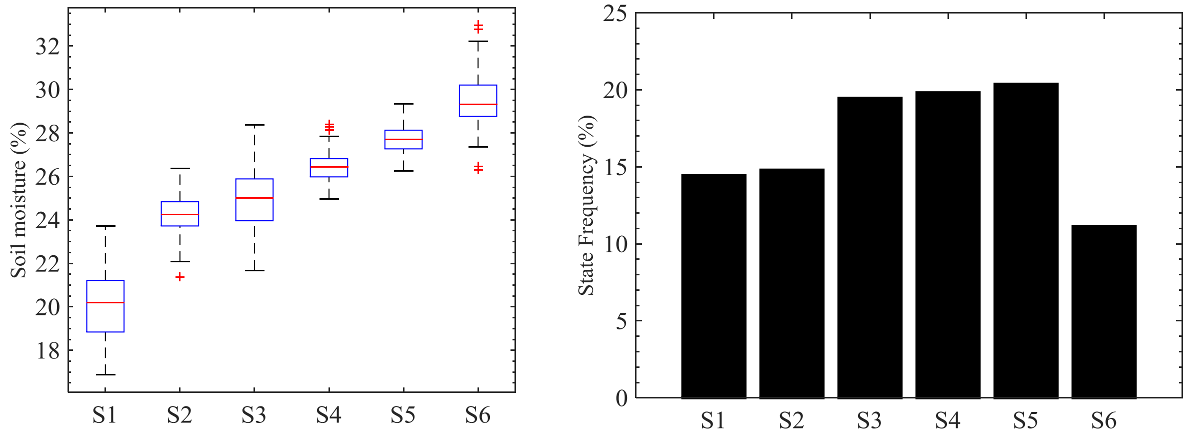


Figure 7. Estimated hidden state sequence for a 3-year period (2014–2016).



(a) SM per state (b) Frequency of SM per state

Figure 8. The estimated distribution and frequency of soil moisture in each state.

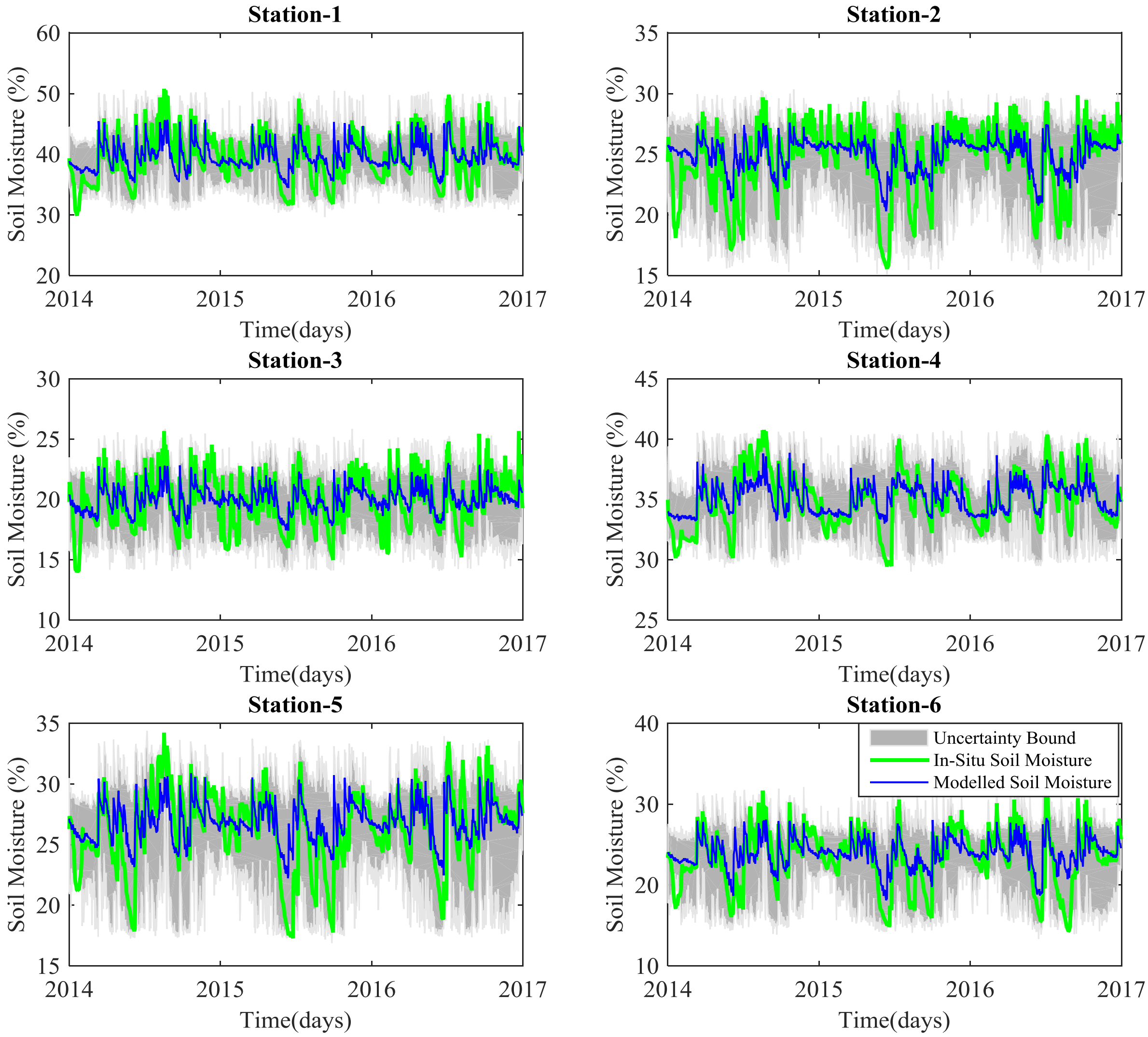


Figure 9. A comparison of time series data between the in-situ and GM-NHMM-simulated SM data for 2014–2016: the green line indicates the in-situ observations, and the blue line represents the median of 100 simulations. The shaded area represents the uncertainty bound of simulations (between 2.5% and 97.5%).

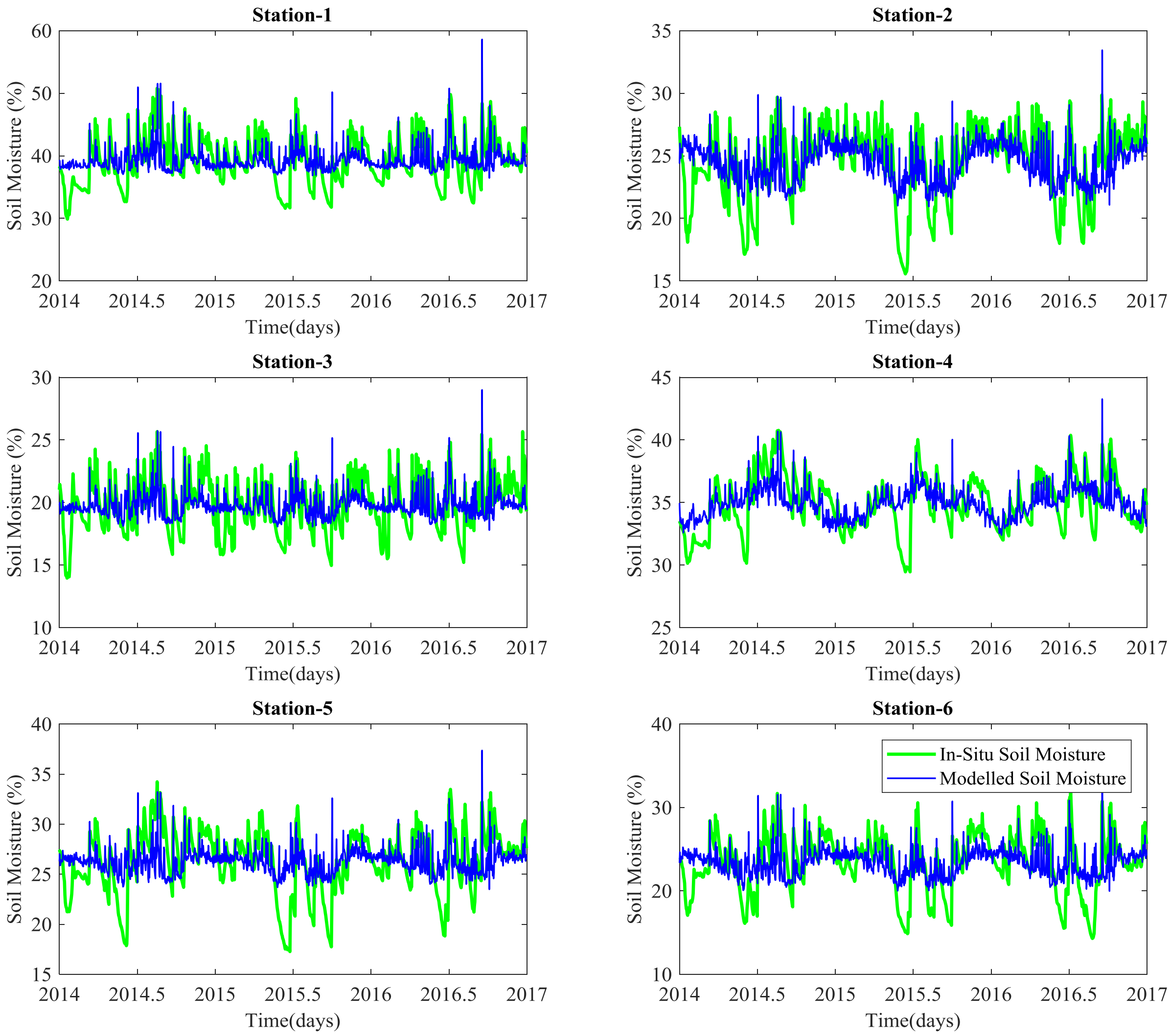
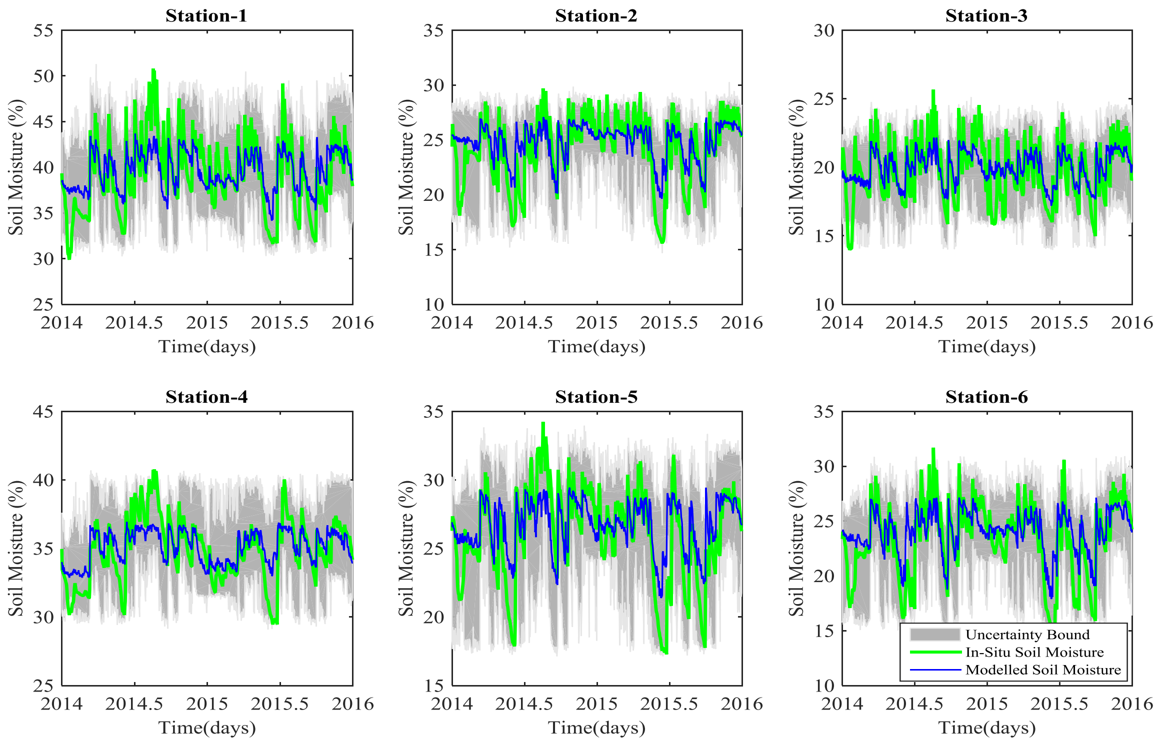


Figure 10. A comparison of time series data between the in-situ and OLR-simulated SM products for 2014–2016: the green line indicates in-situ observations, and the blue line represents OLR-simulated SM.

(a)



(b)

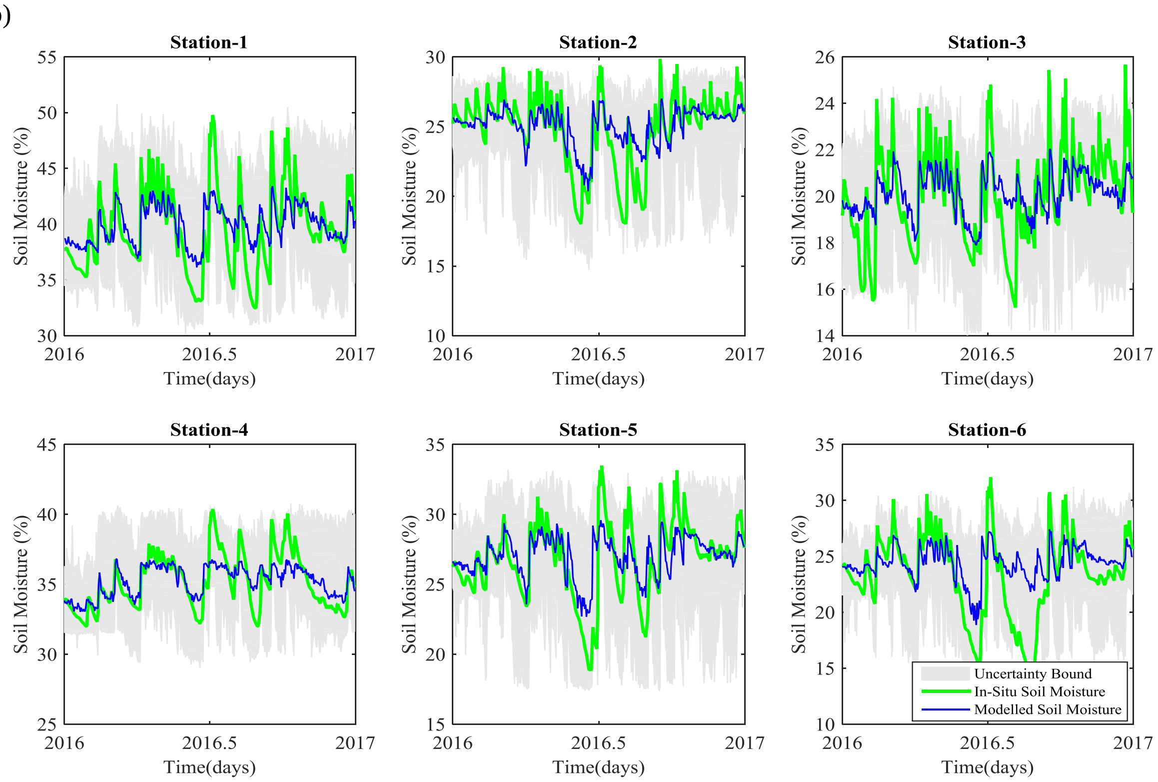


Figure 11. Comparisons between the sequences of simulated soil moisture and that observed at multiple locations in the Yongdam watershed for a) the training period (2014–2015) and b) the validation period (2016).

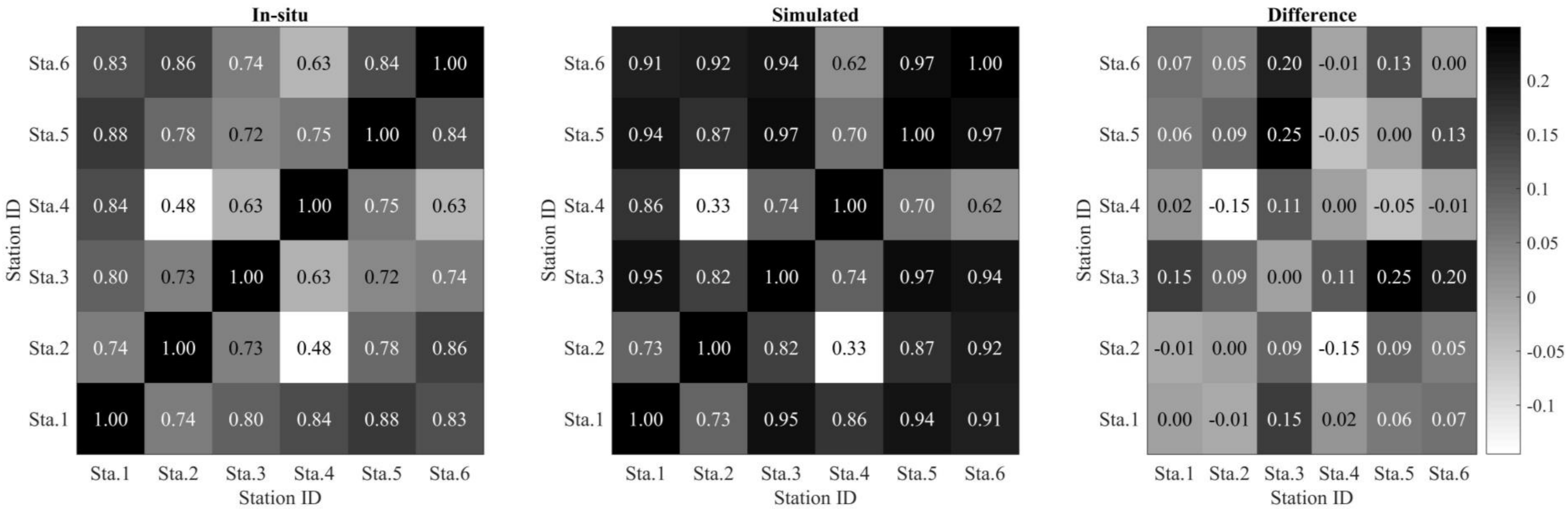


Figure 12. Comparison of the spatial correlation matrices between the observations and simulations of daily soil moisture sequences across 6 stations.