Construction and Experiment of Hierarchical Bayesian Network in Data Assimilation

To cite this article: B R Gudu et al 2014 IOP Conf. Ser.: Earth Environ. Sci. 17 012129

View the article online for updates and enhancements.

Related content

- Assimilation of microwave brightness temperatures for soil moisture estimation using particle filter
  H Y Bi, J W Ma, S X Qin et al.

- Data assimilation: Particle filter and artificial neural networks
  Helaine Cristina Morais Furtado, Haroldo Fraga de Campos Velho and Elbert Einstein Nehrer Macau

- Study on the relationship between soil moisture and its dielectric constant obtained by space-borne microwave radiometers and scatterometers
  Chen Quan, Liu Jiuli, Tang Zhihua et al.
Construction and Experiment of Hierarchical Bayesian Network in Data Assimilation

B R Gudu\textsuperscript{1,3}, S X Qin\textsuperscript{2,3} and J W Ma\textsuperscript{2}

\textsuperscript{1}Institute of Remote Sensing Application, Chinese Academy of Sciences, Beijing, China

\textsuperscript{2}Center for Earth Observation and Digital Earth, Chinese Academy of Sciences, Beijing, China.

\textsuperscript{3}University of Chinese Academy of Sciences, Beijing, China.

sxqin@ceode.ac.cn

Abstract. A Hierarchical Bayesian Network Algorithm (HBN) is developed for data assimilation and tested with an instance of soil moisture assimilation from hydrological model and ground observations. In this work, data assimilation separates into data level, process level and parameter level, and conditional probability models are defined for each level. The data model mainly deals with the scale differences between multiple data, while the process model is designed to take account of non-stationary process. Soil moisture from Soil Moisture Experiment in 2003 and Variable Infiltration Capacity Model is sequentially assimilated with HBN. The result shows that the assimilation with HBN provides spatial and temporal distribution information of soil moisture and the assimilation result agrees well with the ground observations.

1. Introduction

Soil moisture is one of the key environmental variables and it plays a significant role in the terrestrial water cycle [1]. Reliable and accurate information of soil moisture is of paramount importance due to the strong influence on many water resources applications. Merging information from different data sources in a systematic and effective way would result in an improvement of prediction accuracy. Data assimilation (DA) provides a mechanism to combine observations and model simulations, thus yields superior soil moisture retrievals.

The theory of DA rests on the mathematical framework of estimation theory. However, DA always involves nonlinear, highly complex, and exceedingly large systems with complicated error structures that defy the straightforward application of classical optimization methods. Thus a judicious selection of algorithms and approximations based on physical insights is necessary. Many researchers are devoted to improving the existing algorithms and models and introducing newly reported mathematics achievements into data assimilation, such as particle filter [2], Bayesian Network [3]. In this work, a hierarchical Bayesian Network (HBN) algorithm based on hierarchical Bayesian theory is developed to sequentially assimilate data from multiple sources.
Hierarchical Bayesian theory has enjoyed broad scientific applications in ecology, evolution biology, meteorology, oceanography, and so on. As initially described by Berliner [4,5], data assimilation problem is convenient to consider a general three-stage factorization of \( p(\text{process, parameters} \mid \text{data}) \).

\[
p(\text{process, parameters} \mid \text{data}) \propto p(\text{data} \mid \text{process, parameters}) \times p(\text{process} \mid \text{parameters}) \times p(\text{parameters})
\]

(1)
p is the probability distribution and \( p(a \mid b) \) is the conditional probability distribution of \( a \) given \( b \). As long as the data model, process model and parameter model are defined, the posterior and marginal distribution of process and parameters can be estimated from equation (1).

In this work, the northern Alabama State in the USA is chosen as the study area. The intension of this work is to assimilate soil moisture observations from Soil Moisture Experiment in 2003 (SMEX03) and output from Variable Infiltration Capacity Model (VIC) with HBN. The result of DA shows that the spatiotemporal distribution and variation of soil moisture is effectively captured by the created HBN. The DA results are well consistent with the ground observations.

2. Data assimilation scheme

2.1 Study area
The study area is located at the intersection of Alabama State and Tennessee State in USA and covers a flat area of 120km×120km. Figure 1 shows the study area and the distribution of stations. There are 21 stations in the study area with 10 stations from Soil Climate Analysis Network (SCAN) and 11 stations from Alabama Mesonet (ALMNet).

![Figure 1. Location of study area and distribution of observation stations. Observation station is marked with five-pointed star and validation station is marked with solid five-pointed star.](image)

2.2 Land surface variable infiltration capacity model
Another data source was daily soil moisture output from VIC model. VIC model balances both the water and surface energy budgets within the grid cell and has been well calibrated and applied in a number of large river basins over the continental US and globe. In this work, the daily soil moisture output from VIC model was available on 3km×3km grid and there are 1600 grid cells in the study area. Soil moisture from VIC model was obtained at the depth of 10 cm, 60cm and 100cm. In order to get rid of the spin-up period, VIC model was forced from 1 January 2002 to 15 July 2003 but only output between 5 June 2003 and 15 July 2003 was assimilated.
2.3 Development of Hierarchical Bayesian Network algorithm

Let \( Z_t = [Z(S_1,t), \ldots, Z(S_N,t)]^T \) denote observations, \( X_t = [X(B_1,t), \ldots, X(B_M,t)]^T \) is VIC model output, and \( Z^s_t = [Z(S_1,t), \ldots, Z(S_{Num},t)]^T \) is prediction for time \( t \). \( S_1, S_2, \ldots, S_N \) are observation stations, \( B_1, B_2, \ldots, B_M \) are grid cells of VIC model, \( S'_1, S'_2, \ldots, S'_{Num} \) are prediction locations. The goal of this work is to estimate \( Z^s_t \) given \( Z_t \) and \( X_t \).

2.3.1 Data model

Assume there is a latent true process \( O(S_i,t) \) at any point \( S_i \) at time \( t \). The data model for ground measurements can be defined as equation (2).

\[
Z_t = O_t + e_t
\]

For \( t = 1, 2, \ldots, 41 \), \( O_t = (O(S_1,t), O(S_2,t), \ldots, O(S_N,t))^T \) is the true value vector of all observation stations and \( e_t = (e(S_1,t), e(S_2,t), \ldots, e(S_N,t))^T \) is the measurement error. \( e_t \) follows \( N(0, \sigma^2 \Sigma_N) \), where \( \Sigma_N \) is the identity matrix and \( \sigma^2 \) is constant across space and time.

2.3.2 Process model

The process model is defined by equation (3).

\[
O_t = \xi + \rho O_{t-1} + (\beta_0 + \beta)X_t + \eta_t
\]

\( \xi \) is a constant vector across space and time and it represents the trend surface of soil moisture in the study area. \( \rho O_{t-1} \) is an auto-regressive term with \( 0 < \rho < 1 \). \( (\beta_0 + \beta)X_t \) is a spatially varying regression term. \( X_t \) is a vector with elements be the VIC output of the grid cells containing observation stations. \( \beta_0 \) is constant in space and time, while \( \beta = [\beta(S_1), \beta(S_2), \ldots, \beta(S_N)]^T \) with \( \beta(S_i) \) varies in space but keeps constant across time.

Assume \( \eta_t \sim N(0, \Sigma_\eta) \), where \( \Sigma_\eta \) has elements \( \sigma^2_\eta (i, j) = \sigma^2_\eta \rho_\eta (S_i - S_j; \phi_\eta) \) and \( \rho_\eta (S_i - S_j; \phi_\eta) = \exp(-\phi_\eta d(S_i, S_j)) \), \( d(S_i, S_j) \) is distance between observation stations \( S_i \) and \( S_j \).

Assume \( \beta \sim N(0, \Sigma_\beta) \), where \( \Sigma_\beta \) has elements \( \sigma^2_\beta (i, j) = \sigma^2_\beta \rho_\beta (S_i - S_j; \phi_\beta) \) and \( \rho_\beta (S_i - S_j; \phi_\beta) = \exp(-\phi_\beta d(S_i, S_j)) \).

2.3.3 Parameter model

Parameter model defines the prior distribution of all parameters. \( \rho \) follows \( N(0, 10000) \) but restricted in the interval \( I(0 < \rho < 1) \), \( \sigma^2_\beta \), \( \sigma^2_\eta \) and \( \sigma^2_\epsilon \) are assumed to follow inverse gamma distribution independently. \( \xi \) is obtained through the AR(1) analysis. \( \phi_\beta \) and \( \phi_\eta \) are determined through the validation of HBN.

3. Prediction of HBN

We first develop methods for spatial prediction of soil moisture \( Z(S'_i, t) \) at a new location \( S'_i \) and any time \( t \). According to equation (2), \( Z(S'_i, t) \) follows the distribution \( p(Z(S'_i, t) \mid z) \sim N(O(S'_i, t), \sigma^2_\epsilon) \). \( O(S'_i, t) \) is computed using equation (7)

\[
O(S'_i, t) = \xi + \rho O(S'_i, t-1) + (\beta_0 + \beta(S_i))X(S_i, t) + \eta(S_i, t)
\]

(3)
It is clear that \( O(S_i^t, t) \) can only be sequentially determined using all the previous \( O(S_i^t, t) \) up to time \( t \). The posterior distribution of \( \beta(S_i^t) \) given parameters is written as equation (12):

\[
\beta(S_i^t|\theta) \sim N(S_{i,12}S_{i}^{-1}\beta, \sigma^2(1-S_{i,12}S_{i}^{-1}S_{i,21}))
\]  

(4)

The posterior distribution of prediction \( O(S_i^t, t) \) conditional on the observations, VIC output, parameters is defined in equation (21).

\[
p(O(S_i^t, t)|\beta(S_i^t), O_t, \theta, w) \sim N(\chi, \Lambda)
\]  

(5)

\[
\chi = \xi + \rho O(S_i^t, t-1) + \beta_iX(S_i^t) + \beta_iS_{i,12}S_{i}^{-1}(O_t - \rho O_t - \beta_iX - X_i) + S_{i,21}
\]  

(6)

\[
\Lambda = \sigma^2(1-S_{i,12}S_{i}^{-1}S_{i,21})
\]  

(7)

4. Validation and parameter estimation of HBN

In this work, we first trained HBN with ground measurements and the corresponding VIC model output. Then the HBN was validated with data from the validation stations to assess the performance and parameters were adjusted until we got the minimum validation error. The validation mean-square error (VMSE) was given by equation (8).

\[
\text{VMSE} = \frac{1}{n_t} \sum_{i=1}^{n_t} \sum_{t=1}^{T} |Z(S_i^t, t) - Z^*(S_i^t, t)|
\]

(8)

where \( Z^*(S_i^t, t) \) is the mode from HBN, \( Z(S_i^t, t) \) is ground measurement and \( n_t \) is the total number of available observations at all validation stations during the assimilation period.

Table 1 gives \( (\phi_\theta, \phi_\eta) \) and corresponding VMSE values of all layers in the experiment. The HBN performed best when \( (\phi_\theta = 0.1, \phi_\eta = 0.10) \) for layer 1 and layer 3. For layer 2, the VMSE minimized when \( (\phi_\theta = 0.1, \phi_\eta = 5) \), but we still take \( (\phi_\theta = 0.1, \phi_\eta = 10) \). That’s because the correlation between HBN results and ground measurements was better. Detailed comparisons among ground measurements, VIC model output and HBN results at 3 layers are demonstrated in Figure 2. Precipitation and 95% confidence interval of HBN results are also represented by black bars and light gray in these figures respectively.

<table>
<thead>
<tr>
<th>Layer Depth</th>
<th>( \phi_\theta )</th>
<th>( \phi_\eta )</th>
<th>( \phi_\theta )</th>
<th>( \phi_\eta )</th>
<th>( \phi_\theta )</th>
<th>( \phi_\eta )</th>
<th>( \phi_\theta )</th>
<th>( \phi_\eta )</th>
</tr>
</thead>
<tbody>
<tr>
<td>10cm</td>
<td>0.03592</td>
<td>0.02680</td>
<td>0.03592</td>
<td>0.02691</td>
<td>0.02668</td>
<td>0.02319</td>
<td>0.02724</td>
<td>0.03039</td>
</tr>
<tr>
<td>60cm</td>
<td>0.03208</td>
<td>0.02102</td>
<td>0.03208</td>
<td>0.01772</td>
<td>0.01701</td>
<td>0.01754</td>
<td>0.01802</td>
<td>0.02141</td>
</tr>
<tr>
<td>100cm</td>
<td>0.03000</td>
<td>0.02800</td>
<td>0.0357</td>
<td>0.02689</td>
<td>0.02551</td>
<td>0.02516</td>
<td>0.04971</td>
<td>0.05210</td>
</tr>
</tbody>
</table>

For SCAN2055 at layer 1, R is improved from 0.5838 to 0.7542 while VMSE drops from 0.0366 to 0.0117. Compared with the VIC model output, the tendency of HBN is more consistent with the ground measurements and most of the ground measurements are in the 95% confidence interval. Similar case is also demonstrated in figure 2(b) and 2(c). In Figure 3, from top to down are HBN prediction results at depth of 10cm, 60 cm and 100cm, while from left to right are HBN prediction results on 6 June, 21 June and 2 July, respectively. Soil moisture shows diversity and variation across space and time. During the assimilation period, variation of soil moisture across time of layer 1 is the most obvious, layer 2 comes second and layer 3 is the last. Since the near-surface soil moisture is affected greatly by precipitation, while soil moisture at the deeper depth is determined by the vegetation root and soil texture. Soil moisture at (34°56’N, 86°20’W) and (34°40’N, 86°45’W) at 10 cm are relatively lower than that of other location in the study area (figure 3(a)-(c)). That’s because these two locations are very close to ALMNet07 and ALMNet13. Observations of soil moisture at these 2 stations are low, making soil moisture at these location are relatively low. While for layer 3, soil moisture at (35°00’N, 86°30’W), (34°32’N, 86°55’W) and
(35°00′N, 85°45′W) is higher than soil moisture at other locations in the study area (figure 3 (g)-(i)). That’s because the VIC model output at layer 3 is averagely higher than the ground measurements. Although the ground measurements have been assimilated to correct the trajectory of the VIC model, the HBN doesn’t perform so satisfying at the locations where the observation stations are very sparse. In such situation, the performance of HBN is greatly depends on the VIC model output.

5. Conclusion and Discussion

Data assimilation, an effective method to obtain multi-parameter large-scale spatial-temporal distribution and variation information with high precision, draws more and more attention in the recent years. In this work, we develop a hierarchical Bayesian network algorithm to assimilate sequential soil moisture data. Soil moisture from SMEX03 ground observation as well as VIC model is assimilated.
with HBN. A general technical procedure is also extracted based on the experiment. The result of DA shows that HBN result can reflect the spatial-temporal distribution and variation of soil moisture. Compared with the VIC model, results after assimilation are better consistent with the ground measurements and the accuracy is also improved.

References