DEVELOPMENT OF A DISCRETE TIME-SERIES MODEL FOR VEGETATION MONITORING

Yoshito Sawada* and Haruo Sawada

Institute of Industrial Science, The university of Tokyo, 4-6-1 Komaba, Meguro, Tokyo 153-8505 Japan
(yoshitos, sawada)@iis.u-tokyo.ac.jp

KEY WORDS: Anomaly detection, Multi-temporal data, Hidden Markov model, Self-organizing map, Phenology, Land cover changes

ABSTRACT:

This paper describes the development of a novel time series modelling and spectrum anomaly-detection method, which takes into consideration wide-area seasonal changes. By taking advantage of both the high temporal resolution and the wide swath mode of multi-temporal satellite data, such as NOAA/AVHRR, MODIS, and SPOT/vegetation, it is possible to perform high-frequency monitoring of wide-area, land cover changes. However, since the multi-temporal satellite data are influenced by clouds and system noise, in many cases, they must be processed in order to accurately represent the actual surface conditions. We engineered a discrete time-series model using a self-organizing map (SOM) and a hidden Markov Model (HMM) to reduce the influence of clouds in order to improve the accuracy of the products. The spectral information of the pixels was first converted to nominal scale values, and the influence of clouds was eliminated through a time-series modelling using HMM. Since the anomaly-detection method requires a clustering of nominal vectors, dedicated software based on SOM algorithm was also developed. The data for anomaly detection is not dependent on the information of neighbouring pixels, and it is possible to detect an anomaly even if there is only one pixel.

1. INTRODUCTION

In order to address the issues arising due to the various environmental problems that are currently attracting attention, it is necessary to devise a method that enables wide-area monitoring of fluctuations in vegetation conditions such as variations in moisture and temperature and land cover changes.
Various research projects on the global environment utilize the characteristics of the cyclic nature of multi-temporal satellite data (Lhermitte et al., 2008). In these research projects, "the *n*-day composite imagery"(ex. *n*=8, 10), which is created by selecting the best data in 8 or 10 days for every pixel, is often used for characterised seasonal changes. However, the influence of clouds and haze remain, even in these ten-day composite data, and this complicates the monitoring of phenology with a 10 days interval (Sawada et al., 2005). Monthly composite data are not appropriate to monitor phenology dynamics (Alexandridis et al., 2008) because the seasonal changes of vegetation are phenomena taking place in a few weeks for most of cases.
This paper describes a novel method for time series modelling and spectrum anomaly detection by using SPOT/vegetation and MODIS data. Since this methodology can extract a seasonal change model with pixel by pixel, it is useful to monitoring land-cover change and ecological disasters such as large-scale forest fires.

2. METHOD

In this chapter, the algorithm of discrete time-series model is described. This model consists of four modules. (Figure 1)

2.1 Generating of spectral codebook and encoding

The batch-learning SOM algorithm (Kohonen, 2000 and Yamakawa et al., 2005) was used in order to generate a spectral codebook which encodes multispectral data. In each pixel, "pure components" (endmembers) are extracted throughout the data collection period by OPA (Orthogonal Projection Analysis, Cuesta Sánchez et al., 1996) for removing influence of cloud, haze and other noise. “Pure components” are classified by the batch-learning SOM algorithm. Generated SOM is used as a codebook in spectral encoding step. In addition, SOM node which are assigned as cloud and haze contaminated are masked. Then, mutitemporal and multispectral data such as MODIS band1~7 data (MOD09A1) and SPOT/vegetation S10 produts are encoded by the nearest-neighbour (NN) method. If a pixel was “data missing” and contaminated by cloud, this pixel is assigned “NULL” code. We set size of SOM to 20×20 (including “NULL” code, the total number of spectral codes is 401).

2.2 Time-series modelling by HMM

The observation data vector and the state vector of (x,y) pixel are denoted by $\mathbf{o}^{(x,y)}$ and $\mathbf{q}^{(x,y)}$, respectively. Both $\mathbf{o}^{(x,y)}$ and $\mathbf{q}^{(x,y)}$ are nominal scale vectors.

$$\mathbf{o}^{(x,y)}=(o_1^{(x,y)}, o_2^{(x,y)},..., o_T^{(x,y)}) \quad (1)$$
$$\mathbf{q}^{(x,y)}=(q_1^{(x,y)}, q_2^{(x,y)},..., q_T^{(x,y)}) \quad (2)$$

where T = number of scenes
(x,y) = image coordinates
Let assume $\mathbf{q}^{(x,y)}$ follows Markov process, we get the conditional probability of generating $\mathbf{q}^{(x,y)}$ (Rabiner, 1989)

$$P\left(\mathbf{q}^{(x,y)}\middle|\mathbf{o}^{(x,y)}\right)=\pi_{q1}b_{q1}\left(o_1^{(x,y)}\right)\cdot\prod_{t=2}^{T}a_{q(t-1),qt}b_{qt}\left(o_t^{(x,y)}\right) \quad (3)$$

where N = total number of states
N_c = total number of spectral code.

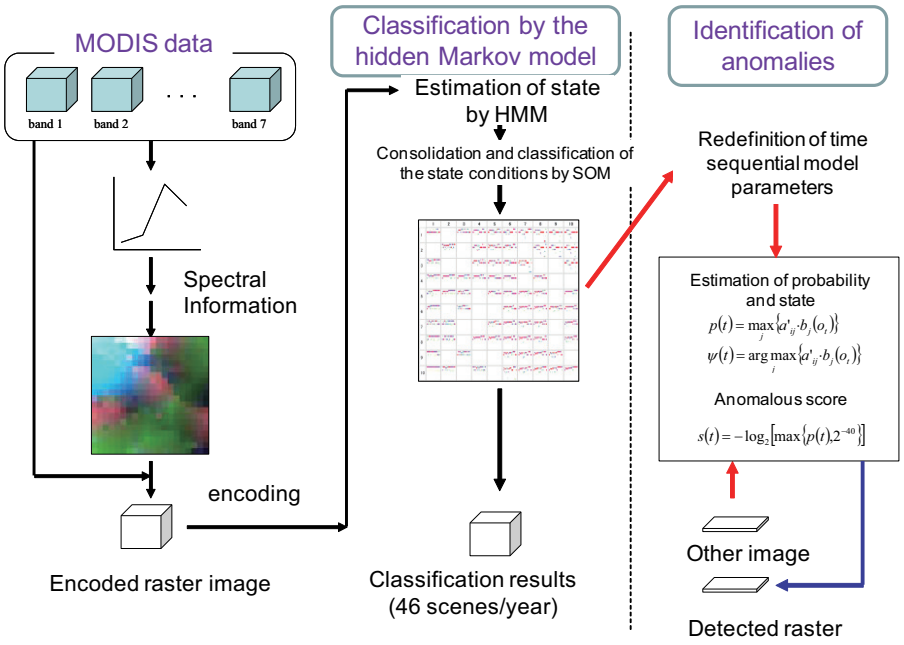

Figure 1. Flow of our discrete time-series model processing

Parameters $\boldsymbol{\pi}$, $\boldsymbol{a}$ and $\boldsymbol{b}$ are defined by following formulas

$$\pi_i = \frac{\#(q_1 = i)}{\sum_{i=1}^{N} \#(q_1 = i)}$$

$$a_{ij} = \frac{\sum_{t=2}^{T} \#(q_{t-1} = i, q_t = j)}{\sum_{j=1}^{N} \sum_{t=2}^{T} \#(q_{t-1} = i, q_t = j)} \quad (4)$$

$$b_i(k) = \frac{\sum_{t=1}^{T} \#(q_t = i, o_t = k)}{\sum_{k=1}^{Nc} \sum_{t=1}^{T} \#(q_t = i, o_t = k)}$$

"#()" is a function that returns occurrence frequency of phenomena in parentheses.
If parameters $\boldsymbol{\pi}$, $\boldsymbol{a}$ and $\boldsymbol{b}$ are known, $\mathbf{q}^{(x,y)}$ is obtained by Viterbi algorithm. Actually we calculate parameters $\boldsymbol{\pi}$, $\boldsymbol{a}$ and $\boldsymbol{b}$ iteratively.

2.3 Classification of seasonal change profile

Since the element of the state vector $\mathbf{q}^{(x,y)}$ is a nominal scale value, we developed nominal scale vectors (and/or strings data) classification method.
Firstly, we define $\mathbf{D}(\mathbf{q}^{(x,y)})$ as a characteristic matrix of the state vector $\mathbf{q}^{(x,y)}$. The (i,j) element of $\mathbf{D}(\mathbf{q}^{(x,y)})$ is one when $q_j^{(x,y)}=i$ and otherwise zero. The size of the matrix $\mathbf{D}(\mathbf{q}^{(x,y)})$ is $N \times T$. For example, when $\mathbf{q}^{(x,y)}$="*124*"

$$\mathbf{D}(\text{"}124\text{"}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

In bioinformatics researches, $\mathbf{D}$ is called 'Position Specific Scoring Matrix' (Gribskov et al., 1987).
Next, $\mathbf{D}_{\text{cent}}(\mathbf{Q})$, the centroid of $\mathbf{D}$, is defined when $\mathbf{Q}$ is a set that consists of state vectors.

$$\mathbf{D}_{\text{cent}}(\mathbf{Q}) = \frac{\sum_{\mathbf{q}^{(x,y)} \in \mathbf{Q}} \alpha^{(x,y)} \cdot \mathbf{D}(\mathbf{q}^{(x,y)})}{\sum_{\mathbf{q}^{(x,y)} \in \mathbf{Q}} \alpha^{(x,y)}} \quad (6)$$

where $\alpha^{(x,y)}$ = weight for $\mathbf{q}^{(x,y)}$
(x,y) = image coordinates

For instance

$$\begin{aligned} &\mathbf{D}_{\text{cent}}(\text{"}121\text{"},\text{"}314\text{"},\text{"}124\text{"}) \\ &= \frac{1}{3}\{\mathbf{D}(\text{"}121\text{"}) + \mathbf{D}(\text{"}314\text{"}) + \mathbf{D}(\text{"}124\text{"})\} \\ &= \frac{1}{3}\begin{pmatrix} 1+0+1 & 0+1+0 & 1+0+0 \\ 0+0+0 & 1+0+1 & 0+0+0 \\ 1+0+0 & 0+0+0 & 0+0+0 \\ 0+0+0 & 0+0+0 & 0+1+1 \end{pmatrix} \\ &= \frac{1}{3}\begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \end{aligned} \quad (7)$$

Then, we define $\mathbf{q}_{\text{cent}}(\mathbf{Q})$, the centroid of **Q**, and *consensus* function which returns consensus sequence.

$$\mathbf{q}_{\text{cent}}(\mathbf{Q}) = consensus\left(\frac{\sum_{\mathbf{q}^{(x,y)} \in \mathbf{Q}} \alpha^{(x,y)} \cdot \mathbf{D}\left(\mathbf{q}^{(x,y)}\right)}{\sum_{\mathbf{q}^{(x,y)} \in \mathbf{Q}} \alpha^{(x,y)}} \right) \tag{8}$$

For example, from eq. (7)

$$\begin{aligned} &\mathbf{q}_{\text{cent}}("121","314","124") \\ &= consensus\left(\mathbf{D}_{\text{cent}}("121","314","124")\right) \\ &= consensus\left(\frac{1}{3} \begin{pmatrix} 2 & 1 & 1 \\ 0 & 2 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 2 \end{pmatrix} \right) \\ &="124" \end{aligned} \tag{9}$$

And we define $f_{ij}(\boldsymbol{X})$ as a function that returns (i,j) element of a matrix $\boldsymbol{X}$ for further convenience.
Similarity of **Q** and another state vector **q**' is defined following formula:

$$s_i(\mathbf{Q}, \mathbf{q}') = \frac{1}{L} \sum_{j}^{L} \sum_{i}^{Nc} f_{ij}\left(\mathbf{D}_{cent}(\mathbf{Q})\right) \cdot f_{ij}\left(\mathbf{D}(\mathbf{q}')\right) \tag{10}$$

Consequently, the algorithm of our nominal scale vectors classification method is described below:

i) to initializing weight vectors $\mathbf{w}_{IJ}$
ii) to set iteration counter k_t to zero
iii) to assign $\mathbf{q}^{(x,y)}$ to a node by eq. (11)

$$I', J' \leftarrow \arg\max_{IJ} \left\{ \frac{1}{L} \sum_{j}^{L} \sum_{i}^{Nc} f_{ij}\left(\mathbf{w}_{IJ}\right) \cdot f_{ij}\left(\mathbf{D}\left(\mathbf{q}^{(x,y)}\right)\right) \right\} \tag{11}$$

where L = number of scenes
(I,J) and (I',J')= SOM node coordinates
$\mathbf{q}^{(x,y)}$ belongs to the node (I',J')

iv) to calculate coefficient h by eq. (12) and eq. (13)

$$h(I, J, x, y) = \exp\left\{ -\frac{(I - I')^2 + (J - J')^2}{2\sigma(k_t)^2} \right\} \tag{12}$$

$$\sigma(k_t) = \max\left\{ 0.01, \sigma_{\text{init}}\left(1 - \frac{k_t}{K_{\max}}\right) \right\} \tag{13}$$

where $K_{\max}$ = number of maximum iterations
σ_{init} = constant ($\sigma_{\text{init}} > 0$)

v) to update weight vector $\mathbf{w}_{IJ}$

$$\mathbf{w}_{IJ} = \frac{\sum_{x,y} h(I, J, x, y) \cdot \mathbf{D}\left(\mathbf{q}^{(x,y)}\right)}{\sum_{x,y} h(I, J, x, y)} \tag{14}$$

vi) $k_t \leftarrow k_t + 1$
vii) to go to step iii) until $k_t = K_{\max}$
viii) to determine final clustering vector by eq. (11)
ix) to calculate the centroid in each node by eq. (15)

$$\mathbf{q}_{IJ} = consensus\left(\mathbf{w}_{IJ}\right) \tag{15}$$

x) to write weight vector $\mathbf{w}_{IJ}$ and final clustering vector (x,y,I',J') to file

2.4 Spectral anomaly detection

In this section, we describe our algorithm about spectral anomaly detection. In general, an anomaly detection methodology is required to clearly distinguish spectral anomaly (i.e. land cover changes) and phenological changes.
If parameters **b**, seasonal change profile of reference year $\mathbf{q}^{(x,y)}$ and clustering results **Q** are known, we can define observation probability of encoded spectrum $p(v_t^{(x,y)})$ in another scene t at (x,y) pixel:

$$p\left(v_t^{(x,y)}\right) = \max_{l} \left\{ a_{kl}^{(t)} b_l\left(v_t^{(x,y)}\right) \right\} \tag{16}$$

where $v_t^{(x,y)}$ = encoded spectrum of another scene
(x,y) = image coordinates
t = scence ID (time)
$a_{kl}^{(t)} = (k \rightarrow l)$ state transition probability at t
$k = q_{t-1}^{(x,y)}$ (the state of reference year at t-1)
l = indices of HMM state

From eq. (16) we get inferential state $\psi_t^{(x,y)}$

$$\psi_t^{(x,y)} = \arg\max_{l} \left\{ a_{kl}^{(t)} b_l\left(v_t^{(x,y)}\right) \right\} \tag{17}$$

$a_{kl}^{(t)}$ is obtained by eq.(18)

$$a_{kl}^{(t)} = \frac{\#\left(q_{t-1} = k, q_t = l, \mathbf{q} \in \mathbf{Q}_{IJ}\right)}{\sum_{l=1}^{N} \#\left(q_{t-1} = k, q_t = l, \mathbf{q} \in \mathbf{Q}_{IJ}\right)} \tag{18}$$

where $\mathbf{Q}_{IJ}$ =a set consists of state vectors which assigned to the SOM node (I,J).
Then, according to hanon's theorem (Shanon, 1948), we define anomaly score s.

$$s = -\log_2 p\left(v_t^{(x,y)}\right) \tag{19}$$

For setting threshold value of s, we assume that encoded spectra are generated by random process. From eq. (16), generating probability p_{random} is

$$p_{random} = \frac{1}{N} \cdot \frac{1}{N_c} \tag{20}$$

where N = total number of states
N_c = total number of spectral code

From eqs. (19) and (20), we get threshold of anomaly score s_{tht}

$$s_{\text{tht}} = -\log_2(p_{random}) = -\log_2\left(\frac{1}{N} \cdot \frac{1}{N_c}\right) = \log_2 N + \log_2 N_c \tag{21}$$

We look upon pixels which is anomaly socre $s > s_{\text{tht}}$ as occurrence of anomaly phenomena.

3. RESULTS AND DISCUSSION

3.1 Water coverage monitoring

An example of our model processing of MODIS (MOD09A1) data is shown in Figure 2. 414 scenes from the early January 2001 till the end of December 2009 were used. The targeted area is Amazon River basin (10N~20S, 80W~40W, Figure 2). Figure 2a, 2b and 2c are raw composite, time-series modelling results (by HMM) and cluster centroid, respectively. As for the processed data, almost all the influences of cloud and noises disappear. 23 states were obtained and "state 23" is assigned to water spectra which is clearly distinguished from other states in the spectral shape (Figure 3). We obtained the water coverage period map by occurrence frequency of state 23 (Figure 4).

3.2 Burned area detection

In order to detect of burned area with 10 days interval, SPOT/vegetation S10 products were processed. The targeted area is far-east Russia (50N~54N, 127E~133E, Figure 5). The data collection period as reference year is from Apr. 1999 to Mar. 2000 (one year, 36scenes). We set SOM node size to 10×10 for clustering of seasonal change profile.
As results, 15 states were obtained. Figure 6 shows 4bands spectra of each state. There are mainly four categories "vegetation", "soil", "snow" and "water" from visual interpretation (Table 1).
The data collection period for anomaly detection is from Jun. to Aug. 2000 (12scenes). Original data (S10 product), inferred state and anomaly score are shown in Figure 7. For comparison,

(a) (b) (c)

Figure 2. Examples of our discrete time-series processed MOD09A1 data (Jan. 12 2009)
(a) original MODIS data (b) HMM processed image (c) cluster centroid image

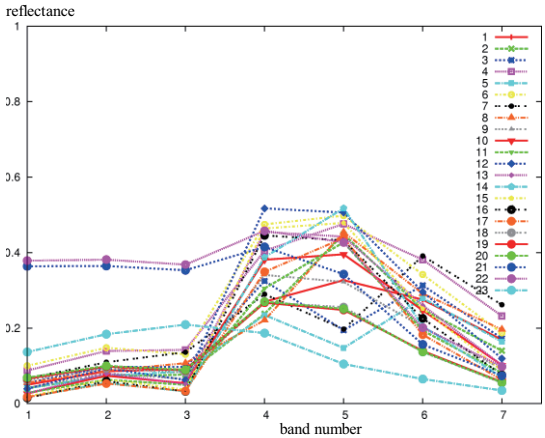

Figure 3. Spectra of states

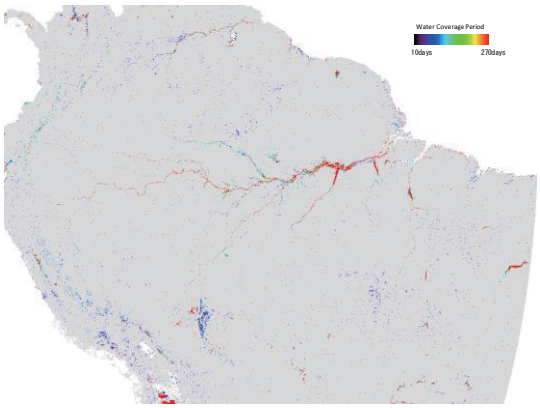

Figure 4. Water coverage period in 2008

same area of the GBA2000 (Global Burned Area, Grégoire et al., 2006) product is also shown in Figure 7. Our method was found effective to identify burned area with pixel by pixel.

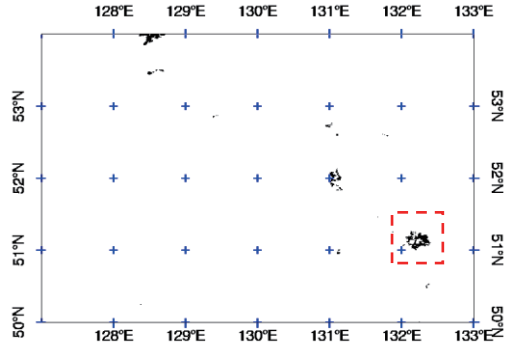

Figure 5. Targeted area of burned area detection

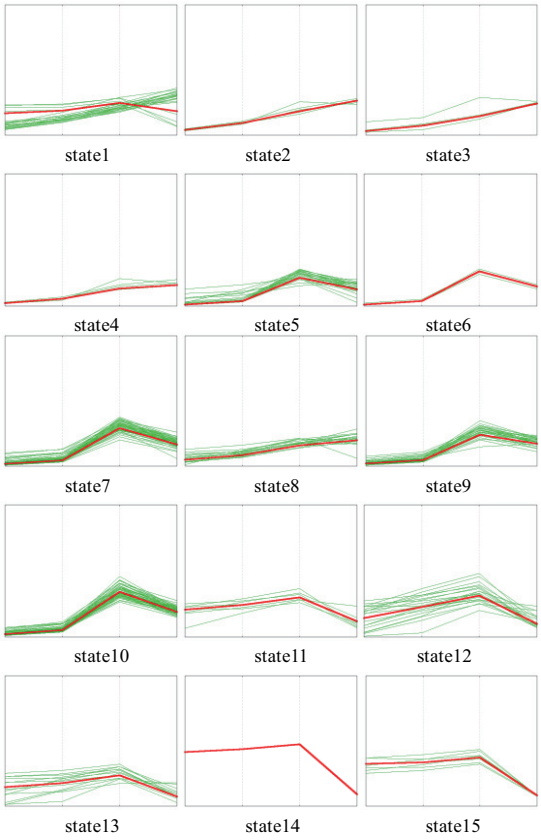

Figure 6. Spectra of each state
(red: averaged spectra)

Table 1. States and Categories

category	state ID
Vegetation	5,6,7,9,10
Soil	1,2,3,4,8
Snow	11,12,13,14,15
Water	1

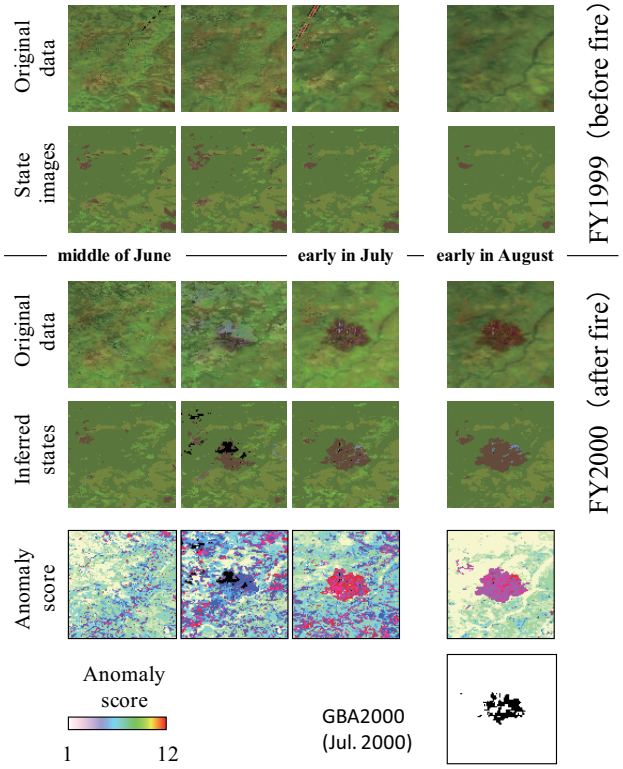

Figure 7. Identification of burned area

3.3 Forest development area detection

In order to identify deforestation trend with 10 days interval, 10-days composite MODIS data were processed. MODIS data were obtained from WebMODIS system (Takeuchi et al., 2005). The result shows to identify deforestation trend with pixel by pixel (Figure 8). ASTER VNIR images are also shown in Figure 9 for comparison.

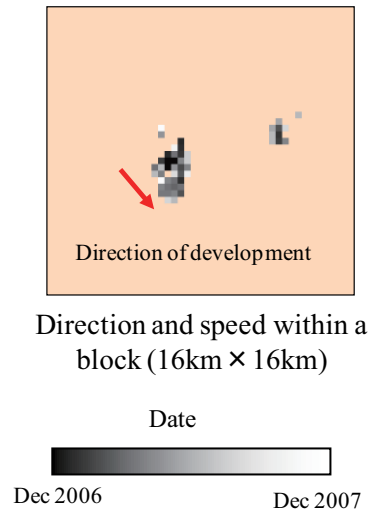

Figure 8. Deforestation direction and trend.
(13.417N, 106.112E)

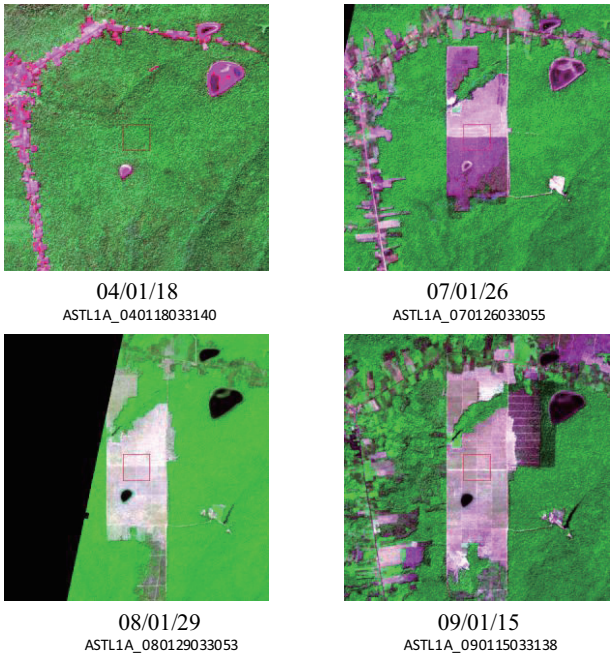

Figure 9. ASTER VNIR image (13.417N, 106.112E, 9km× 9km)

4. CONCLUSIONS

We developed a discrete time-series model using a self-organizing map (SOM) and a hidden Markov Model (HMM) to reduce the influence of clouds in order to improve the accuracy of the products. Our method is able to clearly distinguish spectral anomaly and phenological changes. Our method is found effective to identify land cover changes (i.e. forest fire and deforestation) with pixel by pixel, but further study will be required to clarify its limitation on areas of land cover changes. A small land cover changes will not be identified.

ACKNOWLEDGEMENT

This research used ASTER data beta processed by the AIST GEO Grid from ASTER data owned by the Ministry of Economy, Trade and Industry.
This study was partially supported by JST.

REFERENCES

Alexandridis, T. K., Gitas, I. Z. and Silleos, N. G., 2008. An estimation of the optimum temporal resolution for monitoring vegetation condition on a nationwide scale using MODIS/Terra data. *Int. J. Remote Sens.*, 29(12), pp. 3589-3607.

Cuesta Sánchez, F., Toft, J., van den Bogaert, B. and Massart, D. L., 1996. Orthogonal Projection Approach Applied to Peak Purity Assessment. *Anal. Chem.*, 68 (1), pp. 79-85.

Grégoire, J. M., Tansey, K. and Silva, J. M. N., 2003. The GBA2000 initiative: developing a global burnt area database from SPOT-VEGETATION imagery. *Int. J. Remote Sens.*, 24(6), pp. 1369-1376.

Gribskov, M., McLachlan, A. D. and Eisenberg, D., 1987, Profile analysis: detection of distantly related proteins. *Proc. Natl. Acad. Sci. USA.*, 84(13), pp. 4355-4358.

Kohonen, T., 2000. *Self-Organizing Maps (2nd ed.).* Springer, Berlin, chapter 3.
Lhermitte, S., K. Nackaerts, I. Jonckheere, J. van Aardt, W. Verstraeten, and P. Coppin. 2008. Hierarchical image segmentation based on similarity of NDVI time-series. *Remote Sens. Environ.*, 112 (2), pp. 506-521.

Rabiner, L.R., 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. of the IEEE*, 77(2), pp. 257-286.

Sawada, Y., Mitsuzuka, N. and Sawada, H., 2005. Development of a time-series model filter for high revisit satellite data, *Proceedings of the Second International VEGETATION User Conference (Veroustraete, F., Bartholome, E. and Verstraeten, W.W. ed.)*, pp. 83-89, Office for Official Publications of the European Committees, Luxembourg.

Shannon, C., 1948. A Mathematical Theory of Communication. Bell Syst. Tech. J., 27, pp. 379-423.

Takeuchi, W., Nemoto, T., Baruah, P. J. and Yasuoka, Y., 2005. Online satellite data distribution system for the monitoring of environment and disaster over Asia (in Japanese), *Photogrammetry and Remote Sensing*, 44(2), pp. 68-72.

Yamakawa, T., Horio, K. and Sonoh, S., 2005. Batch Learning of the Self-Organizing Relationship (SOR) Network. Neural Information Processing - Letters and Reviews, 8(3), pp. 31-38.