**FLOOD ANALYSIS AND FORECASTING BY SPATIO-TEMPORAL DATA MINING BASED ON HISTORICAL SATELLITE IMAGE DATABASE**

LI Xi1 and Wataru TAKEUCHI

1Institute of Industrial Science, University of Tokyo, 2-28-4, Tomigaya, Shibuya-k, Tokyo 151-0063, Japan,

Email: lixi@iis.u-tokyo.ac.jp

**KEY WORDS:** AMSR-E, LSWC, image similarity, BCP

**ABSTRACT:** In this paper, AMSR-E data were used to map land surface water coverage (LSWC), which can fast reveal large-scale flood patterns. And the daily LSWC database in time series from 2002 to 2011 was built. Then, PALSAR data (ScanSAR mode) was used to compare to AMSR-E, it is indicated that the availability and importance of LSWC database for flooding detection, moreover, by using PALSAR data (Fine mode) to map water coverage, showed its superiority in the accurate flood evaluation based on a known flood event. Finally, focused on one flood event to analyse and then conducted image similarity calculation, gave a ranking and lined up all the historical images from highest to lowest of LSWC to discover hidden regularities and useful information from large collection of LSWC images.

**1. INTRODUCTION**

Due to the global warming combined with excessive human activities, the flood have been one of the most recurrent, widespread, and disastrous hazards worldwide which caused enormous damages both in terms of loss of life and economics (Singh et al., 2013). Therefore, it’s necessary to grasp accurate and rich information in flood forecasting for the strategy of government and business continuity planning (BCP) of enterprises.

In recent years, the research of flood forecast and a series of methodologies based on remote sensing have been proposed. Among them, AMSR-E, which based on microwave observations have shown superiority since the capability of the signal to penetrate through clouds, their large spatial coverage. And furthermore, since flood events are dynamic processes, higher temporal resolutions are required (Watts, J. D. et al., 2012; Pandey, R. K. et al., 2014).

The objective of this paper is to detect flood event by AMSR-E and PALSAR and forecast flood by spatio-temporal data mining approaches, which based on viewpoint of retrieval of historical similar patterns from image databases to perform instance-based flood analysis and forecasting.

**2. METHODOLOGY**

**2.1 A flowchart and data used in this study**

Figure1 shows a framework of this research. I utilized the platform of International Charter as my disaster information source. The International Charter “Space and Major Disasters” was initiated by European and French space agencies (ESA and CNES). Aims to provide a unified system of space data acquisition and delivery to those affected by natural or man-made disasters. Table1 shows the basic information of flood events worldwide, which were selected from the International Charter. AMSR-E data, which has a spatial resolution of 10 kilometers, a temporal resolution of 0.5 day were used to map LSWC and to build LSWC database. PALSAR data (ScanSAR mode), which has a spatial resolution of 100 meters, was used to compare and verify the results. Moreover, PALSAR data (Fine mode), which has a spatial resolution of 6.25 meters, was used to prove its superiority in the flood evaluation. Finally, conduct the spatio-temporal analysis by image similarity calculation for the LSWC map set derived from LSWC database and probability density function in order to forecast flooding.

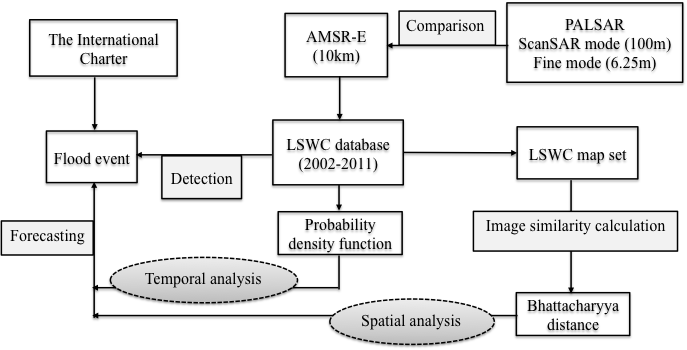


Figure 1. A flowchart of this study

Table 1. Basic information of selected flood events occurred in the worldwide

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Country** | **Location of Event** | **Date of Charter Activation** | **Lat.** | **Lon.** |
| China | Huai River Basin | 2007-7-19 | 32.5N | 115.8E |
| Vietnam | North and Central provinces | 2008-11-5 | 20.9N | 105.8E |
| Pakistan | North West Pakistan | 2010-8-19 | 28.2N | 69.4E |
| Senegal | Senegal | 2007-9-18 | 16.1N | 13.8W |
| Namibian | Northern Namibia | 2011-4-5 | 18.2S | 15.7E |
| Argentina | Santa Fe and Entre Rios provinces | 2007-3-30 | 31.2S | 60.6W |
| Bolivia | Cochabamba, Santa Cruz | 2008-2-9 | 14.6S | 65.1W |
| Colombia | Bolivar province | 2011-5-23 | 8.3N | 73.9W |
| Mexico | Tabasco | 2007-11-3 | 18.1N | 92.7W |
| USA | Iowa | 2008-6-13 | 42.5N | 93.2W |
| Australia | Queensland | 2011-1-9 | 27.3S | 151.3E |

**2.2** **Land surface water coverage mapping**

AMSR-E daily mosaics including 18.7 and 23.8GHz are used to compute normalized difference frequent index (NDFI). When atmospheric transmission is near to 1, we can obtain NDFI as follows (Takeuchi et al., 2006);

(1)

where TB18.7V and TB23.8V are the brightness temperature of vertical (V) polarization at 18.7GHz and 23.8GHz. NDFI is less affected by atmospheric conditions and is not dependent on the soil temperature. NDFI provides a sensitive indicator of the presence of surface water and it has a good capability to distinguish the water surface and land surface (Takeuchi et al., 2009). Land surface water coverage is one of the critical parameters for agriculture and flood monitoring (Mori et al., 2009).

**2.3 Image similarity calculation**

**Histogram**

A histogram is a graphical representation of the distribution of data. It is an estimate of the probability distribution of a continuous variable. Histograms are used to plot the density of data, and often for density estimation.

In discrete form, on behalf of discrete gray levels by , there is the following formula (Pearson, K. 1895):

(2)

where: is number of pixels in the image appears as a gray level , is the total number of image pixels, and is the frequency.

**Bhattacharyya distance**

The Bhattacharyya distance measures the similarity of two discrete or continuous probability distributions. It is often used to determine the relative closeness of the two samples or separability between classes in classification being considered (Bhattacharyya, A. 1943). For histogram similarity calculation, BD obtained the best effect.

For discrete probability distributions p and q over the same [domain](http://en.wikipedia.org/wiki/Domain_of_a_function) X, it is defined as:

(3)

where: (4) is the Bhattacharyya coefficient.

where: is the Bhattacharyya distance between p and q distributions, are two different distributions.

**3. RESULTS AND DISCUSSIONS**

**3.1 Flood event detection by AMSR-E and PALSAR on global scale**

I mapped the LSWC of all flood events in table 1, which was derived from AMSR-E. And then used PALSAR data acquired almost at the same satellite pass time with AMSR-E data, which has a higher spatial resolution, to compare and verify. The floods happened in Colombia in 2011 and in Vietnam in 2008 as an example for analysis (Figure 2,3), figure a shows LSWC at flood area in Colombia and Vietnam. Brighter area indicates high abundance of water coverage at that pixel. In order to access the reliability of the LSWC derived from AMSR-E so that investigated the potential of AMSR-E for flood detection, as for PALSAR data, after calculated backscattering, conducted masking、stretching and got PALSAR water map of original resolution as figure b shows, then re-defined the image and got PALSAR water map of new resolution as figure c shows.

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a. AMSR-E LSWC (10km) | b. PALSAR water map (100m) | c. PALSAR water map (10km) |

Figure 2: Two kinds of remote sensing image in Colombia (2011/05/23)

|  |  |  |
| --- | --- | --- |
|  |  |  |
| a. AMSR-E LSWC (10km) | b. PALSAR water map (6.25m) | c. PALSAR water map (100m) |

Figure 3: Two kinds of remote sensing image in Vietnam (2008/11/05)

Comparing figure a and c of figure 2, we knew that the area was absolutely covered by waterlogging, and a good agreement could be found between AMSR-E LSWC and PALSAR water map. So we could conclude that the flood area displayed in PALSAR image and in AMSR-E image could be basically matched. LSWC database derived from AMSR-E was available to detect flood events.

Meanwhile, we could see from figure 3 that even be resized into 100m, also could clearly distinguish the water cover area for the spatial resolution is 6.25m. So we knew that the PALSAR (fine mode) data could better evaluate certain flood events and distinct slight difference.

|  |  |
| --- | --- |
|  |  |
| a. Colombia (2011/05/23) | b. Vietnam (2008/11/05) |

Figure 4: Scatter plot between dB and WC

Moreover, we also analyzed the relationship between dB derived from PALSAR and water coverage as figure 4 shows. Comparing with figure a, which X axle represents PALSAR water coverage of Colombia with spatial resolution of 100m, figure b in Vietnam has the spatial resolution of PALSAR data of 6.25m.

By comparing the Figure a and b of figure 4, we could find that, AMSR-E data slightly overestimated flood because the emitted microwave signal is sensitive to both the water and the soil moisture, the fractional water coverage estimated by using AMSR-E is a combination of water and wet soil area. It is impossible to distinguish the wetness and flooding effects solely by AMSR-E coarse resolution (Takeuchi et al., 2009). On the other hand, PALSAR data (6.25m) could better detect and evaluate the flood event for its high spatial resolution. It could read detail information, and tell small difference.

**3.2 Spatio-temporal analysis of flood events**

For AMSR-E has a strong advantage of high temporal resolution, which is very significant in research of time dynamic flooding problems, we made use of AMSR-E data to do this part of research. According to the principle that the higher AMSR-E LSWC values will be associated to a relative increase of both water area and soil moisture, I have calculated the LSWC to indicate the water level. Figure 5 shows daily changes of LSWC of one pixel (32.5N 115.8E) in Huai River Basin from 2002 to 2011, we could see that in every year around July, LSWC increased obviously, which was most evident in 2003 and 2007, more than 80%. In fact, the flood happened around July 19th, 2007. Moreover we could initially conclude that the flood happened in Huai River belongs to the seasonal flooding.

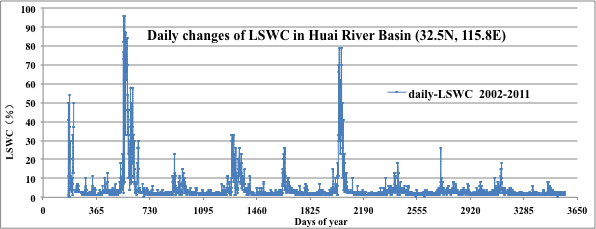


Figure 5: Daily changes of LSWC in Huai River Basin (32.5N, 115.8E) from 2002 to 2011

In order to know the frequency of each LSWC, I calculated the number of days corresponding to each LSWC value and then got the histogram (Figure 6).

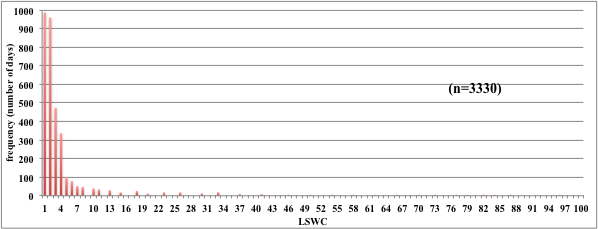


Figure 6: Histogram graph

Then I tried to estimate the probability density of the LSWC. The total area of a histogram used for probability density is always normalized to 1. I divided the result in figure 6 by the number of all days in 10 years (3330 days), and got the probability density curve (Figure 7). From this curve we could see that it was a natural logarithmic distribution, and determination coefficient reached 0.64. Also we found that when LSWC was 10, the probability will reach to 90%, and then gradually approaching to 1.

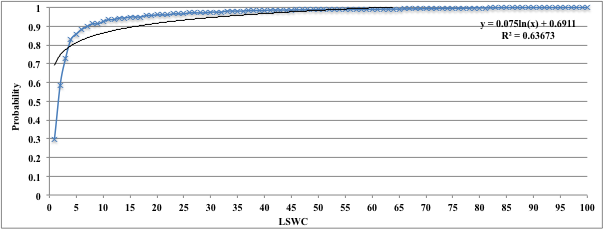


Figure 7: Probability density curve of LSWC in Huai River Basin (32.5N, 115.8E)

In order to visually see the daily change of land surface water coverage, I built a LSWC database of one pixel in Huai River in China and mapped the continuous daily LSWC from 2002 to 2011. Among a total 3330 images, I used images in January in 2007, which belong to dry season and images in July, which belong to rainy season to show the result. The result is showed as figure 8 and figure 9 respectively. In which we could clearly see the big difference of LSWC between dry season and rainy season. Moreover, in addition to the specific water area such as rivers and lakes, the water area expanded significantly in July, the amount of water in rivers and lakes also increased significantly. According to the research results above, there was indeed a flood happening around July 19 in 2007.

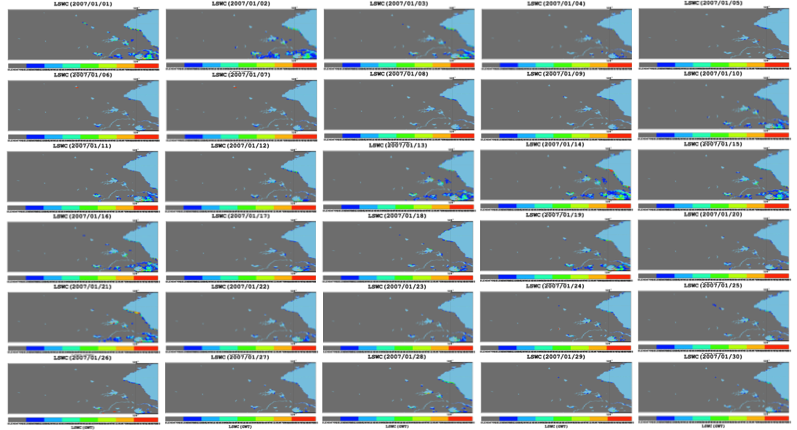


Figure 8: Daily LSWC map in Dry season (Jan. 2007) in Huai River Basin

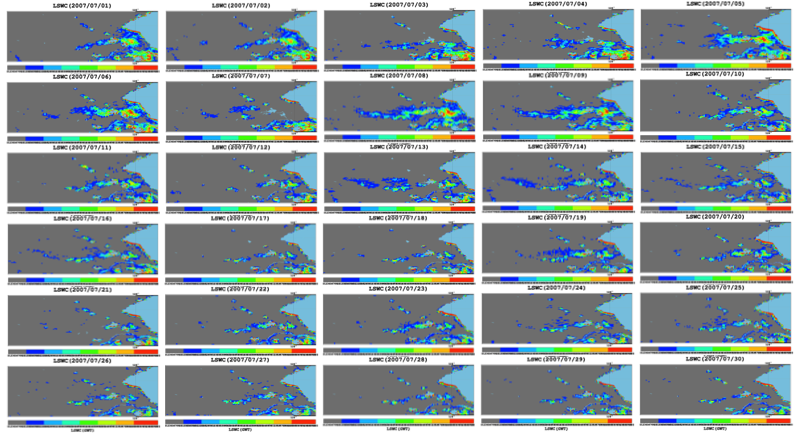


Figure 9: Daily LSWC map in Rainy season (Jul. 2007) in Huai River Basin

**3.3 Image similarity calculation**

In order to get more information and hidden regularities of a certain flood event, consequently to make clear the flood pattern and forecast the development of flooding from large LSWC image collections. I conducted image similarity calculation and then gave a ranking and lined up all the historical images from highest to lowest of LSWC.

I treated the image in 2007/07/19 as a standard. First, I generated the histogram of each image, and calculated the Bhattacharyya distance between the standard and each target, which is the distance between two histograms, the smaller the distance is the better they are matched. And finally got the ranking result as table 2 shows. From which we knew that the image of 2003/07/18 was the most similar with the standard, 2007/07/19. We could also see from figure 10, two images were very similar.

So we could conclude that the flooding happened around 2007/07/19 might be the same pattern with the flooding happened around 2003/07/18. We could get some useful information and hidden regularities according to the historical information and development of flooding in 2003/07/18.

Table2. Ranking result of image similarity calculation

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date list | Similarity | Rank | Date list | Similarity | Rank | Date list | Similarity | Rank |
| 20030718 | 0.986173 | 1 | 20110720 | 0.980722 | 11 | 20050609 | 0.977885 | 21 |
| 20070716 | 0.985977 | 2 | 20050712 | 0.980637 | 12 | 20100418 | 0.977743 | 22 |
| 20030715 | 0.98325 | 3 | 20070713 | 0.980468 | 13 | 20030722 | 0.977543 | 23 |
| 20090616 | 0.983151 | 4 | 20110719 | 0.980265 | 14 | 20100613 | 0.977541 | 24 |
| 20030717 | 0.982222 | 5 | 20030721 | 0.979767 | 15 | 20090617 | 0.977408 | 25 |
| 20070715 | 0.981469 | 6 | 20050707 | 0.979642 | 16 | 20050930 | 0.977162 | 26 |
| 20070710 | 0.981099 | 7 | 20030421 | 0.979502 | 17 | 20060710 | 0.977039 | 27 |
| 20050711 | 0.981075 | 8 | 20020630 | 0.978822 | 18 | 20070629 | 0.97702 | 28 |
| 20030720 | 0.980838 | 9 | 20020704 | 0.978149 | 19 | 20090613 | 0.976988 | 29 |
| 20030331 | 0.980794 | 10 | 20050928 | 0.977937 | 20 | **…** | **…** | **…** |

|  |  |
| --- | --- |
| Macintosh HD:Users:rigi:Desktop:学习天堂:东大:ACRS-2014:研究:AMSR－E历史图集:淮河流域LSWC:LSWC20070719_Huaihe.jpg | **Macintosh HD:Users:rigi:Desktop:学习天堂:东大:ACRS-2014:研究:AMSR－E历史图集:淮河流域LSWC:LSWC20030718_Huaihe.jpg** |

Figure 10: Comparison of LSWC map between standard (2007/07/19) and target (2003/07/18) in Huai River Basin

**4. CONCLUSIONS AND FUTURE WORK**

In this paper, we demonstrated the potential of AMSR-E for flood detection on a global scale. The daily LSWC database in time series from 2002 to 2011, which was derived from AMSR-E was built and PALSAR data was used to access the reliability of AMSR-E, it is indicated that the availability and importance of LSWC database for flooding detection. And we focused on a flood event happened in Huai River Basin in China to analyse, then conducted image similarity calculation, gave a ranking and lined up all the historical images from highest to lowest of LSWC to discover hidden regularities and useful information from large collection of LSWC images.

The future work is to use other means of data mining algorithms like Bayesian inference and Hidden Markov Model in order to dig deep information and make clear the flood pattern, possibility of flooding, as well as the reliability of information for flood analysis and forecasting. And by using distributed hydrological model to investigate and understand the processed and mechanism of complex flooding and to simulate flooding in order to solve water problems.

**References:**

Bhattacharyya, A. 1943. “On a measure of divergence between two statistical populations defined by their probability distributions”. Bulletin of the Calcutta Mathematical Society 35, pp: 99–109.

Mori, S., Takeuchi, W and Sawada, H, 2009. Estimation of land surface water coverage (LSWC) with AMSR-E and MODIS. Proceedings of 2nd joint student seminar at AIT: Bangkok, Thailand.

Pearson, K. 1895. “Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material”. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 186, pp. 343–414.

Pandey, R. K., Crétaux, J. F., et al, 2014. Water level estimation by remote sensing for the 2008 flooding of the Kosi River, International Journal of Remote Sensing, 35(2), pp: 424–440.

Singh, Y., et al, 2013. Flood monitoring using microwave passive remote sensing (AMSR-E) in part of the Brahmaputra basin, India International Journal of Remote Sensing, 34(14): 4967–4985.

Takeuchi, W., Komori, D., Oki, T. and Yasuoka, Y., 2006. An integrated approach on rice paddy irrigation pattern monitoring over Asia with MODIS and AMSR-E. Proceedings of the AGU: San francisco, USA.

Takeuchi, W and Louis Gonzalez, 2009, Blending MODIS and AMSR-E to predict daily land surface water coverage, International Remote Sensing Symposium.

Watts, J. D., Kimball, J. S., et al, 2012. Satellite Microwave remote sensing of contrasting surface water inundation changes within the Arctic–Boreal Region Remote Sensing of Environment, 127, pp: 223–236.