EPIDEMIC RISK ASSESSMENT OF ACUTE WATERY DIARRHEA FOR THE 2011 AYUTTHAYA FLOOD DISASTER USING REMOTE SENSING AND WATER QUALITY

Peera YOMWAN[a,c,d], Chunxiang CAO[b*], Jian ZHAO[a,c], Rong TIAN[a,c] and Warawut SUPHAMITMONGKOL[e,f]

[a] Ph.D. Candidate, State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications of Chinese Academy of Sciences, Beijing 100101, China
E-mail: peerayom@hotmail.com, zhaojian2009@irsa.ac.cn

[b] Professor, State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications of Chinese Academy of Sciences,
Postbox 9718-13, Datun Rd., Chaoyang, Beijing 100101, China;
Tel: +86-139-1161-0226;
E-mail: cao143@irsa.ac.cn

[c] Graduate School of Chinese Academy of Sciences, Beijing, 100049, China

[d] Professional Survey Engineer, Department of Lands, Bangkok, 10210, Thailand

[e] Research Center on Fictitious Economy and Data Science, Chinese Academy of Sciences, Beijing, 100190, China
E-mail: suphamitmongkol@yahoo.com

[f] Kasetsart Agricultural and Agro-Industrial Product Improvement Institute, Kasetsart University, Bangkok, 10900, Thailand

KEY WORDS: Flood parameters, Remote Sensing, Acute watery diarrhea, Multilayer Perceptron, Epidemic risk

Abstract: The flood disaster generally leads to infectious disease outbreaks especially diarrhea frequently having a maximum rate of patients during flood occurrence. Flood water contaminated with pathogen closely relates to the epidemic risk of waterborne infectious diseases, therefore measuring water quality in inundated area can lead to estimate the epidemic risk due to flood. We focus to utilize dissolved oxygen (DO), which is a well-known chemical indicator of environmental water quality. Remote sensing especially radar imagery has an effective capability to detect water variables and also widely uses for determining flood parameters. This study aim to assess the epidemic risk of acute watery diarrhea by using remote sensing and the water quality sampling based on multilayer perceptron (MLP) function. The multiple temporal Radarsat2 images are used to determine flood area and duration during September to December 2011 in Ayutthaya province located on Chao Phraya river basin most severely affected by the 2011 major flood in South East Asia. The reference of epidemic risk derives from the weekly report of diarrhea morbidity from entire hospitals in study area when assume that the more than 50th percentile of weekly morbidity rate are epidemic period. To define input layer of MLP, the main parameters composing of flood duration, population density and dissolved oxygen are simulated to 50-m resolution and divided by each district area. The input and output layer are randomly divided into two parts: 66% for training and 34% for testing. After model training, the predicted epidemic risk of each district has correlation coefficient between 0.70 to 0.91 and RMS error between 0.012 to 0.045. Therefore, we conclude that based MLP algorithm we can assess the epidemic risk of acute watery diarrhea due to flood and the output risk map should also assist the decision maker to prevent and relieve the epidemic.

1. INTRODUCTION

The 2011 monsoon season, Southeast Asia region faced a major flood event particularly in Thailand, having more than 700 reported deaths with over 12.8 million people were affected, and the World Bank estimated damages reached $US 45 billion. When flash floods occur, people in inundated area frequently suffer from waterborne infectious disease during and after flood occurrence (Schwartz et al., 2006). Bureau of Epidemiology of Thailand reported most severe infectious disease outbreaks such as diarrhea, fever, pneumonia, conjunctivitis, dengue fever, leptospirosis and hand-foot-mouth disease, and the report showed that the morbidity rate of diarrhea had the maximum increment. The main cause of such epidemics is dirty flood water which people suffered from flood become infected through drinking, cooking and contact including via indirect way such as unclean food or drink container. Therefore, if we can identify the spatial information of the polluted flood water, it should be lead to assess waterborne disease risks in inundation area.

To determine flood parameters, remote sensing has been used for water detection and characterization for decades, as well as radar data is very effective for extracting flood area, duration and depth in many papers. There are a number of researches attempt to utilize remote sensing so as to assess the disease risk. Beck et al. (2000) illustrated an increasing number of health studies have used remotely sensed data for monitoring, surveillance, or risk mapping and also expressed two case studies such as Lyme disease in the Northeastern United States and cholera in Bangladesh. Lleo et al. (2008) reviewed that remote sensing and waterborne diseases can predict cholera epidemics by monitoring sea surface temperature and sea surface height via earth observing satellites, and can also predict Rift Valley fever (RVF) outbreaks up to five months in advance in East Africa using Pacific and Indian Ocean sea surface temperature anomalies coupled with satellite normalized difference vegetation index (NDVI) data. Kazama et al. (2011) simulated the coliform bacteria distribution in Cambodia using a hydraulic model and estimated the impact of inundation on public health using a dose–response model. However, no studies have tried to model the epidemic risk of diarrhea due to flood disaster in term of spatial-temporal prediction.

The multilayer perceptron (MLP) with a supervised learning technique called back propagation neural network (BPNN) has been widely used to estimate the risk in disaster field. For example, Bai et al. (2005) provided the method of predicting BP neural networks for SARS epidemic to improve the existing computational methods. Cao et al. (2010) investigated the potential epidemic disease risk after earthquake based on BPNN with normalized difference vegetation index (NDVI) and normalized difference water index (NDWI) of remote sensing images. This study aims to utilize MLP or BPNN function with flood parameters for modeling the epidemic risk of diarrhea due to flood.

2. STUDY AREA AND MATERIALS

2.1 STUDY AREA

Our study area covers 8 districts of Ayutthaya province locating on Chaophraya river basin which most affected by the 2011 major floods in Thailand. The districts, called "amphoe" in Thai, of the study area compose of Phra Nakhon Si Ayudhya, Bang Chai, Bang Pa-in, Bang Pahan, Phak Hai, Sena, Bang Sai and Uthai shown in Figure 1 (right). Bureau of Epidemiology of Thailand reported that these districts faced epidemic infectious disease during flood disaster especially diarrhea. Being full of river and canal expressed in Figure 1 (middle) is the main cause that these areas frequently are affected from flood almost every year. The study area covers 1,423.28 square kilometers with population currently estimated at around 484,439 people.

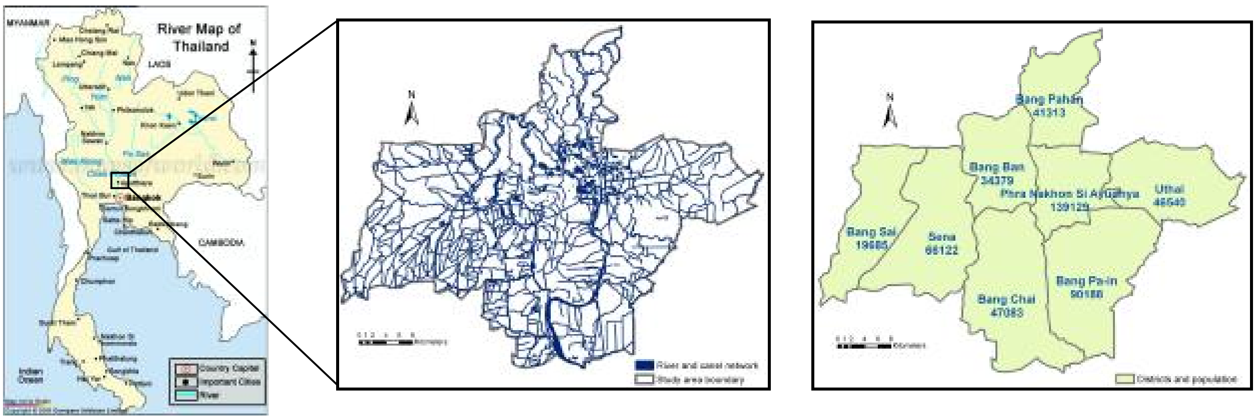

Figure 1. Left: the study area locates in Ayutthaya province at Chao Phraya river basin (www.mapsofworld.com); Middle: the river canal network in our study area; Right: the 8 districts (amphoe) of the study area with its population

2.2 MATERIALS

Six Radarsat2 scenes ScanSAR narrow mode with 50-m resolution acquired on (1) 9 September 2011, (2) 3 October 2011, (3) 21 October 2011, (4) 14 November 2011, (5) 4 December 2011, and (6) 28 December 2011 were used to determine flood area and duration. We applied 186 water quality samplings of Pollution Control Department of Thailand during flood occurrence for estimating dissolved oxygen value in inundated area. The census block of each commune, called "tambol" in Thai, was utilized to calculate population density shown in Figure 2 (left). The surveillance weekly reported of Bureau of Epidemiology of Thailand was used for estimating the diarrhea morbidity of each district during flood expressed in Figure 2 (right).

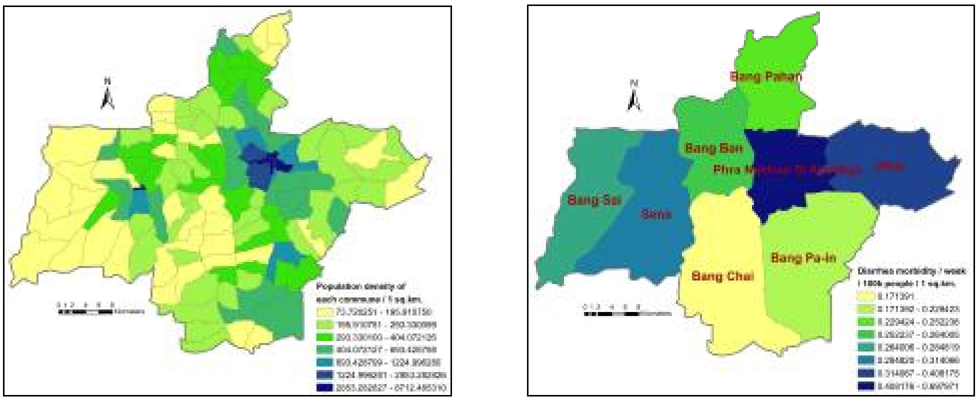

Figure 2. Left: Population density per 1 sq.km. estimated by the census block of each commune (tambol); Right: the diarrhea morbidity of each district (amphoe) during flood disaster (week 36-52, 2011)

3. METHEODOLOGY

3.1 FLOOD PARAMETERS DETERMINATION

To extract flood area, the maximum likelihood with supervised classification was used to extract water of six Radarsat2 scenes, and then we therefore define inundated area by overlaid with surface water layer. Figure 3 expresses that flooding began approximately in week 36 (left), peaked in week 46 (middle) and abated in week 52 (right).

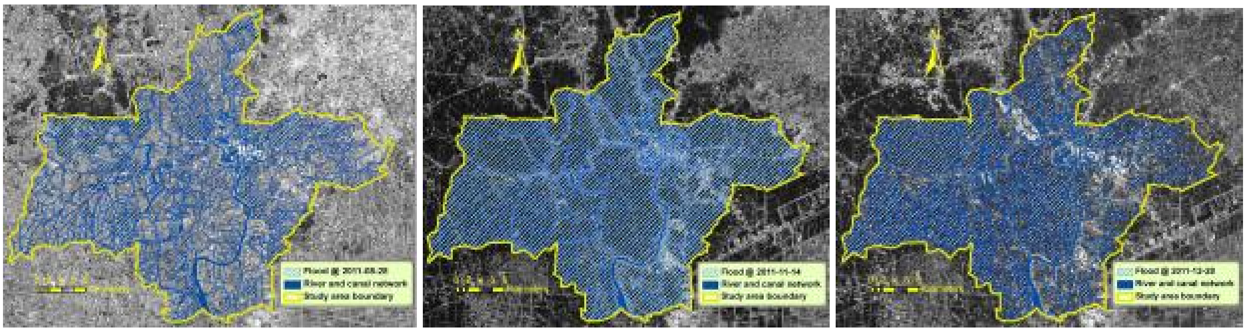

Figure 3. Flood area extracted from Radarsat2 scenes on October 28 (left), November 14 (middle) and December 28 (right), 2011

Subsequently, we estimate flood duration of each area by comparing flood areas of six Radarsat2 scenes. The flood duration map expressed in Figure 4 (left) illustrates the entire study area suffered from flood during 1-121 days.

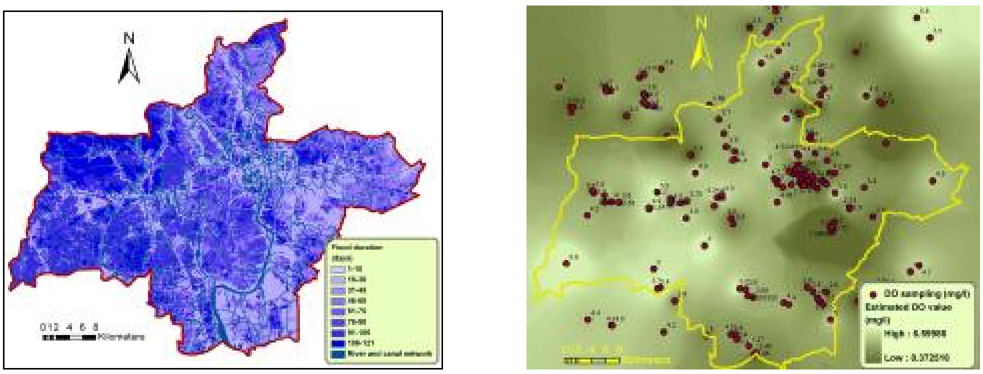

Figure 4. Left: flood duration map of our study area between week 36-50 in 2011; Right: Dissolved Oxygen map in study area estimated from 186 flood water samplings by using inverse distance weight (IDW) function

For estimating flood water quality in the study area, we focus to utilize dissolved oxygen (DO) value as the water quality indicator of this study. We applied 186 flood water sampling collected in Ayutthaya province to evaluate DO map in the study area by using inverse distance weight (IDW) interpolation shown in Figure 4.

3.2 EPIDEMIC INVESTIGATION

As the weekly diarrhea morbidity of entire hospitals in the study area in 2011, we can plot the weekly morbidity rate per 100,000 people as shown in Figure 5 (left). In our case, we define the onset of epidemic as the weekly morbidity rate per 100,000 populations exceed the median or 50th percentile of weekly morbidity rate per 100,000 populations in 2011 in the study area (Schwartz et al. 2006, U.S. Department of Health and Human Services 2006). As the graph in Figure 5 (left), we found that in 2011 the study area has two main periods of the diarrhea outbreaks that are during week1-11 and during week 38-51. After questioning the officers of the ministry of public health, we found that the first period of the outbreaks at the beginning of the year is probably a result of the long celebrations on New Year festival, frequently drinking alcohol or eating unhygienic food. Therefore, we only focus to the second period of outbreak occurring during the flood period.

Moreover, we attempted to find the temporal relationship between flood and epidemic by comparing the second period of epidemic with flood area of radarsat2 scenes. The graph of flood area in figure 5 (right) expressed a very close relationship of flood area and epidemic intensity. While flooding started, peaked and abated in week 36, 46 and 52 respectively, the second epidemic period also started at week 38 although in week 39, 40 and 43 were not in epidemic period but it became peak in week 46 and also abated in week 52.

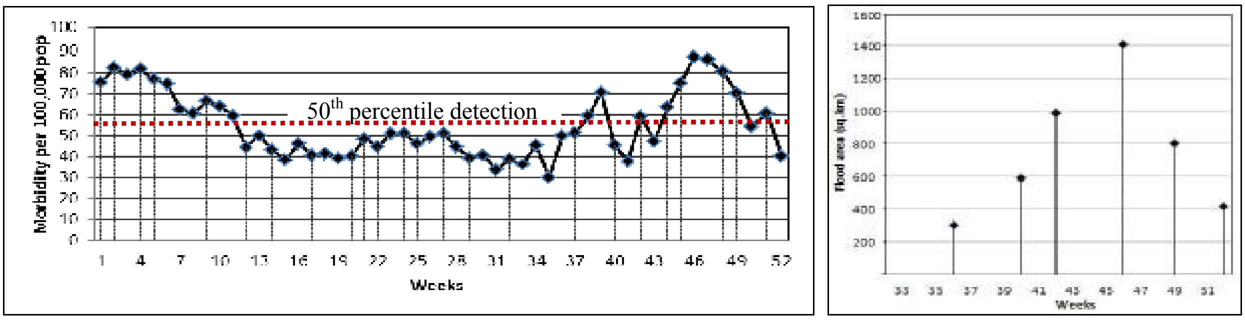

Figure 5. The epidemic investigation of the weekly morbidity rate of diarrhea per 100,000 populations at 50th percentile (left) compared with flood area and duration in the study area (right)

Subsequently, we define the reference epidemic risk by averaging the weekly morbidity rate during flood (week 36-52) and comparing with the 50th percentile of morbidity rate per week in 2011. Nevertheless, these risks are only representative one value per district as the location of its hospital. Therefore, we used this risk set to interpolate the epidemic risk of entire area by using inverse distance weight (IDW) function as shown on figure 8 (left).

3.3 MULTI-LAYER PERCEPTRON AGORITHMS

Multilayer perceptron (MLP) is an artificial neural network model that maps sets of input data onto a set of appropriate output. MLP utilizes a supervised learning technique called backpropagation neural network (BPNN) which is the most popular neural network algorithm (Han et al. 2006). It consists of an input layer, one or more hidden layers, and an output layer as shown in Figure 6. The key component of MLP is the formal neuron, which sums the inputs, and performs a non-linear transform via the activation function (Kanevski et al. 2004).

For training phase, we used three main flood parameters such as flood duration, dissolved oxygen and population density as the input layer and we randomly selected 66% of reference epidemic risk as the output layer. The appropriate number of hidden layer in each district is the most important on this step. The synaptic weights of the feedforward network are adjusted by using supervised learning to minimize the error between the network output and the desired output (Chang et al. 2010). After training, the weights are fixed and the input value is propagated through the network until it reaches the output layer. Then the testing phase was carried out by using the trained model with remain input data comparing with remain 34% of reference epidemic risk data.

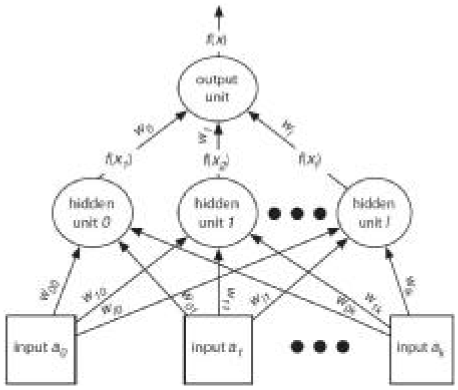

Figure 6. The basic layers of multi-layer perceptron such as input, hidden and output layer (Witten et al., 2005)

4. RESULTS

Table 1 demonstrates the number of hidden layer, correlation coefficient and RMS error of MLP testing data divided by each district. The proper hidden layer number we carried out is between 3-8, while correlation coefficient and RMS error are between 0.70–0.91 and 0.012-0.045 respectively.

District (Amphoe)	Population	Area(km^2)	Hidden layer	Correlation coefficient	RMS error
Phra Nakhon Si Ayudhya	139,129	117.82	7	0.74	0.028
Bang Chai	47,083	250.03	4	0.79	0.040
Bang Ban	34,379	136.75	8	0.83	0.025
Bang Pa-in	90,188	237.10	9	0.86	0.012
Bang Pahan	41,313	130.96	7	0.89	0.045
Sena	66,122	215.28	3	0.83	0.035
Bang Sai	19,685	164.72	5	0.91	0.020
Uthai	46,540	170.62	8	0.70	0.033

Table 1. The statistical error and precision of MLP output divided by each district

We examined the precision of MLP prediction by plotting graph between the predicted MLP risk and reference risk. Figure 7 represent that the scatter plot of Bangpahan, Bang sai and Bangban have R value of 0.89, 0.91 and 0.83 respectively.

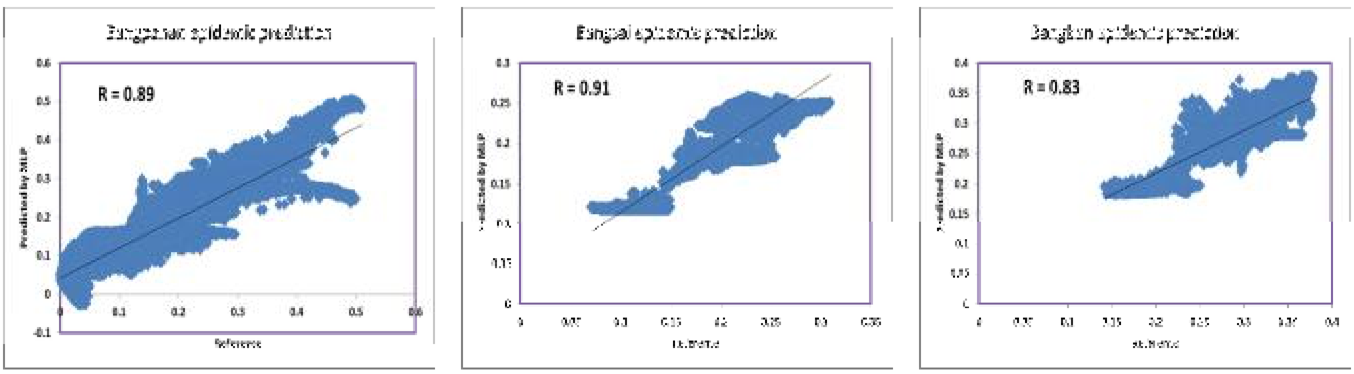

Figure 7. The scatter plot between the diarrhea epidemic risk derived from hospital morbidity report (reference) and MLP prediction risk in Bangpahan (left), Bangsai (middle) and Bangban (right) district

Finally, we determine the prediction map of diarrhea epidemic risk from the output of MLP testing phase as shown in figure 8 (right). There are high risk in Bangban and Sena, low risk in Bang chai and Bang pa-in and moderate in other district. This trend has a good relation with flood duration and DO map in Figure 4.

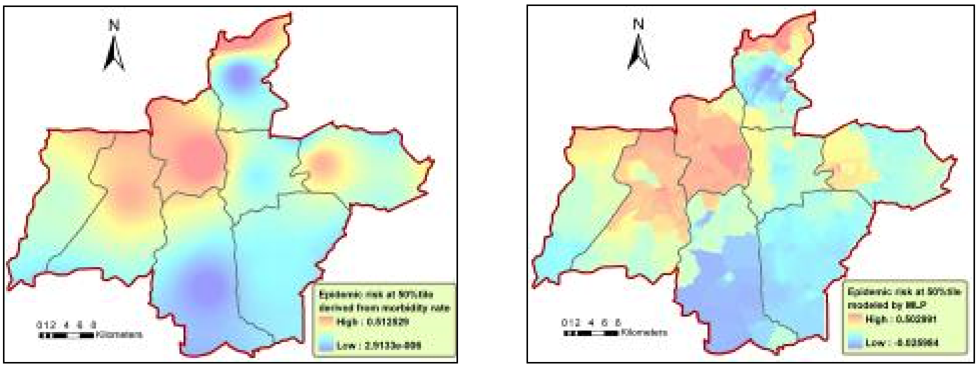

Figure 8. The MLP predicted map of diarrhea epidemic risk (right) compared with the risk calculated by morbidity rate of hospitals in study area during flood occurrence

5. DISCUSSION AND CONCLUSION

Based on MLP approach, we have gained agreeable accuracy and precision to assess the epidemic risk by using remote sensing data and dissolved oxygen sampling of diarrhea. The most advantage of this approach is using only three input factors to calculate the risk. This means we can predict the epidemic risk in advance on any local area, and decision-maker is also able to use this tool to prevent, response and relieve the epidemic. However, in flood disaster also have many other factors affecting epidemic such as emigration, hygienic condition of local people, flood intensity and etc. Moreover, the limitation of morbidity data or reference data should affect the accuracy of model. Therefore, in further research we can more study on the other factors relating to epidemic due to flood disaster.

6. REFERENCES

Bai, Y. P. and Jin, Z., 2005. Prediction of SARS epidemic by BP neural networks with online prediction strategy. Chaos, Solitons and Fractals, 26, pp. 559–569.

Beck, L. R., Lobitz, B. M. and Wood, B. L., 2000. Remote Sensing and Human Health: New Sensors and New Opportunities. Emerging Infectious Diseases, 6 (3), pp. 217-227.

Cao, C. X., Chang, C. Y., Xu, M., Zhao, J., Gao, M. X., Zhang, H., Guo, J. P., Guo, J. G., Dong, L., He, Q. S., Bai, L. Y., Bao, Y. F., Chen, W., Zheng, S., Tian, Y. F., Li, W. X. and Li, X. W., 2010. Epidemic risk analysis after the Wenchuan Earthquake using remote sensing. International Journal of Remote Sensing, 31 (13), pp. 3631–3642.

Chang, L. C., Shen, H. Y., Wang, Y. F., Huang, J. Y. and Lin, Y. T., 2010. Clustering-based hybrid inundation model for forecasting flood inundation depths. Journal of Hydrolog, 385, pp. 257–268.

Han, J. W. and Kamber, M., 2006. Data Mining: Concepts and Techniques. Elsevier, the United States of America, pp.328-333.

Kanevski, M., Parkin, R., Pozdnukhov, A., Timonin, V., Maignan, M., Demyanov, V. and Canu, S., 2004. Environmental data mining and modeling based on machine learning algorithms and geostatistics. Environmental Modelling & Software, 19, pp. 845–855.

Kazama, S., Aizawa, T., Watanabe, T., Ranjan, P., Gunawardhana, L. and Amano, A., 2011. A quantitative risk assessment of waterborne infectious disease in the inundation area of a tropical monsoon region. Sustainability Science, doi 10.1007/s11625-011-0141-5.

Lleo, M. M., Lafaye, M. and Guell, A., 2008. Application of space technologies to the surveillance and modelling of waterborne diseases. Current Opinion in Biotechnology, 19, pp. 307-312.

Qi, H. and Altinakar, M. S., 2011. Simulation-based decision support system for flood damage assessment under uncertainty using remote sensing and census block information. Natural Hazards and Earth System Sciences, 59, pp. 1125–1143.

Schwartz, B. S., Harris, J. B., Khan, A. I., Larocque, R. C., Sack, D. A., Malek, M. A., Faruque, A. G., Qadri, F., Calderwood, S. B., Luby, S. P. and Ryan E. T., 2006. Diarrheal epidemics in Dhaka, Bangladesh, during three consecutive floods: 1988, 1998, and 2004. The American Society of Tropical Medicine and Hygiene, 74 (6), pp. 1067–1073.

U.S. Department of Health and Human Services, 2006. Principles of Epidemiology in Public Health Practice. SELF-STUDY, Course SS1978.

Witten, I. H. and Frank, E., 2005. Data Mining: Practical Machine Learning Tools and Techniques. Elsevier, United States of America, pp. 223-233.