A Study of Typhoon Intensity Change Using Satellite Remote Sensing Data

Chung-Ru Ho*, Yu-Ching Chsng, Yung-Hsiang Lee, Zhe-Wen Zheng, Shih-Jen Huang, Nan-Jung Kuo

Department of Marine Environmental Informatics, National Taiwan Ocean University, Keelung, 20224, Taiwan – (b0211, X0168, D93840002, D95840004, huangsj, c0021)@mail.ntou.edu.tw

Abstract – Mechanisms of influence on typhoon intensity change in the northwestern Pacific are studied by using satellite remote sensing data. A decision tree of data mining technique is applied to analyze the possible influence of geophysical parameters on typhoon intensity change. The related geophysical parameters include sea surface temperature, atmospheric water vapor, rain rate, sea surface height anomaly, and sea-air temperature difference. The total number of 14 Category-5 typhoons occurred between 2003 and 2007 in the northwestern Pacific is employed for this study. The results indicate that the major mechanism of influence on typhoon intensity change is sea-air temperature difference and the second one is sea surface temperature. About 88% typhoon intensity is enhanced when a typhoon passes over the ocean where its sea surface temperature is larger than air temperature. The model is further validated by typhoon JANGMI. The accuracy and precision of this model are 82.3% and 85.7%, respectively.

Keywords: Typhoon, Northwestern Pacific, Data mining, Decision tree, Sea surface temperature, Air-sea interaction.

1. INTRODUCTION

The western North Pacific is an area of the most frequent tropical cyclones strikes over the world (Figure 1). There are 6 to 10 typhoons of Category 4 or 5 in the Saffir-Simpson hurricane scale emerge in the western North Pacific every year (Lin et al., 2005). These severe typhoons bring drastic impact on the coastal area through powerful winds and torrential rain. The ocean response to the typhoon is one of the most important components of air-sea interaction. Previous studies have shown that the changes of geophysical parameters of ocean and atmosphere may affect the typhoon intensity during its lifetime. Alliss et al. (1992) used the rainfall data from SSM/I (Special Sensor Microwave/Imager) onboard the DMSP (Defense Meteorological Satellite Program) satellites to calculate the total latent heat release from Typhoon Hugo (1989) and found that the typhoon intensity is enhanced with the increase of total latent heat release. Schade and Emanuel (1999) reported that the surface cooling response directly affects the transfer of heat from the ocean to the atmosphere and thus the typhoon intensity changes. Shay et al. (2000), and Goni and Trinanes (2003) estimated the tropical cyclone heat potential (TCHP) from Microwave Imager (TMI) and sea surface height anomaly (SSHA) derived from satellite altimetry and pointed out the relationship between typhoon intensity and TCHP. When a typhoon passes by a warm eddy, it is possible to enhance its intensity. Lin et al. (2005), Wu et al. (2007), and Lin et al. (2009) also indicated that warm upper ocean may provide more heat content to typhoon to increase its intensity. Thus, getting more understanding about the behavior of upper ocean response in response to a typhoon passage prove to be the key for further improving the understanding and prediction of the typhoon intensity change.

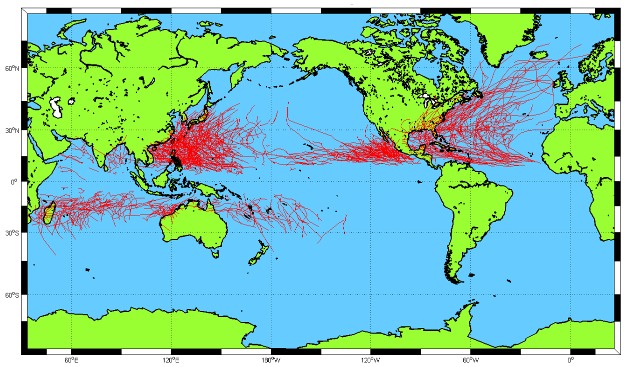

Figure 1. Tracks of tropical cyclones from 2000 to 2007. The most intensity occurred area of tropical cyclones is in the Northwestern Pacific.

Because typhoons are such transient, violent atmospheric processes with great variations in trajectory and strength that upper ocean response are hardly captured by ship-borne observation and cruise tracks. Therefore, satellite observations and reanalysis data are used to characterize and quantify the upper ocean dynamic existing prior to the typhoon passage, as well as the location and magnitude of an upper ocean response to a typhoon after typhoon passage. Satellite remote sensing has the potential to be an efficient and reliable way to provide continuous measurements for quantifying the ocean response to a typhoon. In this paper, we use data mining technique to explore mechanisms of influence on typhoon intensity change from extensive remote sensing data. The study area is in the western North Pacific from the Equator to 40°N in latitude and from 120°E to the International Date Line in longitude (Figure 2).

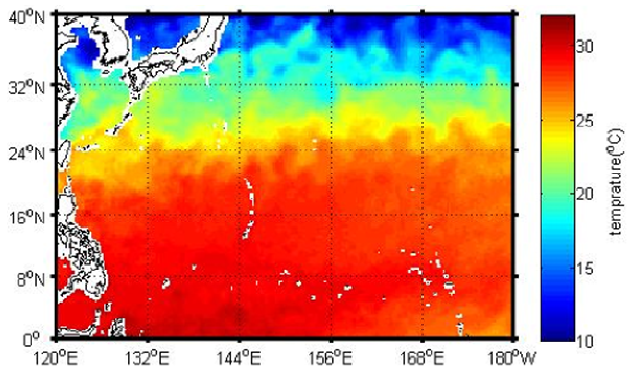

Figure 2. The Study area. The color shows the distribution of sea surface temperature in 2007.

* Corresponding author.

** This work is supported by National Science Council of Taiwan through NSC98-2611-019-017-MY3.

2. DATA

2.1 Typhoon Data

The total number of 14 typhoons with category 5 in Saffir-Simpson scale, that is, central pressure lower than 920 hPa in the northwestern Pacific from 2003 to 2007 is used for this study. The data of typhoon's central pressure, latitude and longitude in every six hours are obtained from Japan Meteorological Agency (JMA).

2.2 Satellite Data

Four kinds of remote sensing measurements area used in this work including sea surface temperature (SST), sea surface height anomaly (SSHA), water vapor (WV), and rain rate (RR). The SST data is derived from the Tropical Rainfall Measuring Mission/Microwave Imager (TRMM/TMI) and the Advanced Microwave Sounding Radiometer for the Earth Observing System (AMSR-E). The SSHA data is derived from altimeters onboard the TOPEX/POSEIDON, Jason-1, ERS-1/2, and ENVISAT satellites. SSHA is defined as the difference between the observed sea-surface height and the seven-year (1993-1999) mean of sea-surface height data. The WV and RR data is derived from TRMM/TMI. All remote sensing data is interpolated into a spatial resolution of 0.25º in latitude and longitude and a temporal resolution on daily.

2.3 Re-analysis Data

Since air temperature is not easy to obtain from satellite remote sensing, the air temperature data is re-analysis data obtained from the National Center for Environmental Prediction (NCEP) at National Center for Atmospheric Research. The data has a spatial resolution of 2º and daily temporal resolution.

3. METHOD

3.1 Data Mining and Decision Tree

Data mining, also called knowledge discovery, is the process of automatically or semi-automatically analyzing data from different perspectives to discover useful information (Berry and Linoff, 1997). It is also the process of finding correlations or patterns in a group of data (Frawley et al., 1992; Berson et al., 1999; Ronald, 2001). Many algorithms are used in data mining. One of the most frequently used algorithms is decision tree. Han and Kamber (2006) pointed out that decision tree is a tree-form structure consisting of nodes and branches. Each internal node represents a test of the nature of the data and the branch indicates the result of the test. The leaf nodes (i.e., the final nodes) stand for the categories or the category distribution. A chart of decision tree is shown in Figure 3. There is a specific path that dictates the decision tree classification from root to leaf nodes in decision tree. To avoid the over-fitting problem in decision tree, appropriate reduction, or pruning, must be conducted to improve decision tree accuracy. There are two types of pruning-pre-pruning and post-pruning. In pre-pruning, the preset threshold keeps decision tree from growing as the nodes on the rear turn into leaves of the tree, and the label of the leaves becomes the overwhelming category in the training set of the node. In post-pruning, a complete number is established before removing the branches. The branches are removed based on the calculation of the error rate of the branch, and the branch nodes on the rear that are not removed become leaves. Frequently used algorithms in decision tree include ID3 (Iterative Dichotomiser 3), C4.5, CART (Classification and Regression Trees) and CHAID (Chi-Squared Automatic Interaction Detector). C4.5 is the algorithm to be used in this study.

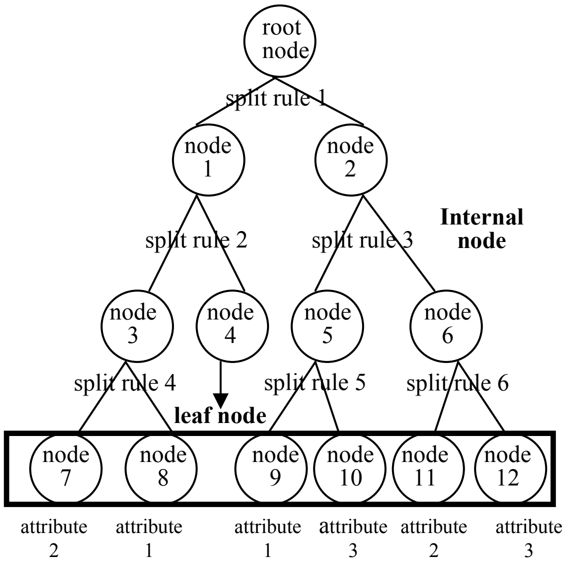

Figure 3. A diagram of decision tree algorithm.

3.2 C4.5 Algorithm

C4.5 is an algorithm modified from ID3 algorithm (Quinlan, 1993). It uses an extension to information gain as gain ratio to overcome bias of ID3. At each node of the tree, C4.5 algorithm chooses one attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. It can handle both continuous and discrete attributes. In order to handle continuous attributes, C4.5 algorithm creates a threshold and then splits the list into those whose attribute value is above the threshold and those that are less than or equal to it. The algorithm builds decision tree from training data using the concept of information entropy. The smaller the information entropy required, the greater the purity of the partitions. The information entropy is given by (Han and Kamber, 2006)

$$I(S) = -\sum_{i=1}^{m} p_i \log_2(p_i)\,, \tag{1}$$

where p_i is the probability that an arbitrary tuple in *S* belongs Class C_i. Assume we were to partition the tuples in *S* on some attribute *A* having *n* distinct values, $\{a_1, a_2, \ldots, a_n\}$, as observed from the training data. Attribute A can be used to split *S* into *n* partitions or subsets, $\{D_1, D_2, \ldots, D_n\}$, where D_j contains those tuples in *D* that have outcomes a_j of *A*. These partitions would correspond to the branches grown from Node *N*. The expected information required to classify a tuple from *S* based on the partitioning by *A* is measured by

$$I_A(S) = \sum_{j=1}^{n} \frac{|D_j|}{|D|} \times I(D_j)\,. \tag{2}$$

Information gain is defined as

$$G(A) = I(D) - I_A(D)\,. \tag{3}$$

The information gain measure is biased towards tests with many outcomes. To overcome this bias, C4.5 uses an extension to information gain as gain ratio. It applies a kind of normalization to information gain using a "split information" value defined as

$$SI_A(D) = -\sum_{j=1}^{n} \frac{|D_j|}{|D|} \times \log_2\left(\frac{|D_j|}{|D|}\right). \tag{4}$$

The gain ratio is then defined as

$$GR(A) = \frac{G(A)}{SI(A)} \tag{5}$$

The attribute with the maximum gain ratio is selected as the splitting attribute.

4. DATA MINING MODE

4.1 Data Preprocessing

The central pressure of typhoon is assigned to be "N" if the pressure difference between the one and previous one is positive (the intensity weakens), to be "Y" if the difference is negative (the intensity strengthens), as well as to be "S" if the difference is zero (the intensity sustains).

To identify the difference of SST before and after typhoon pass by, the SST anomaly is computed by subtraction of 7-day running average from daily SST. The reason that we use 7-day running average instead of monthly average is to avoid the influence of seasonal variation. The anomalies of WV and RR are computed by subtraction of monthly mean from daily data. For the different spatial resolution of air temperature, it has been re-gridded into 0.25º, the same as that of SST and then calculates the temperature difference between sea surface and air (SATD). If the value is positive of above-mentioned data, we assign it as "Y". If the value is negative, we assign it as "N" and as "S" if the value does not change.

4.2 Data Sampling

After remove the null data, we finally have the total number of 378 data in which there are 186 data during the strength period of typhoon intensity, 62 data during the weakness period, and 130 data during the unchanged phase. Since the cases of weakness are fewer than others and the number of cases would influence the accuracy of predication result, to reduce the effect we adapt the "force sampling" technique to increase the number of data during weakness period from 62 to 92. Therefore, we have the total number of 408 data to be analyzed.

4.3 Model Construction

There are five input attributes including SST, WV, RR, SSHA, and SATD. The output attribute is the intensity of typhoon. For the training data, which is usually the bigger part of data, is used for constructing the tree. The more training data collected, the higher the accuracy of the results. The other group of data, testing, is used to get the accuracy rate and misclassification rate of the decision tree. In this study, we randomly choose 70% of the data (286) for training and 30% of the data (122) for testing. The minimum support of split is 40 to avoid the tree too complicate.

5. RESULTS AND DISCUSSION

5.1 Results

Using the decision tree model with C4.5 algorithm, we have the result. The major factor to affect the typhoon intensity is the temperature difference between sea surface and air and the second factor is sea surface temperature. As shown in Figure 4, the first split point of the attribute is SATD > 0 and SATD ≤ 0. The second split point is SST > 0 and SST ≤ 0.

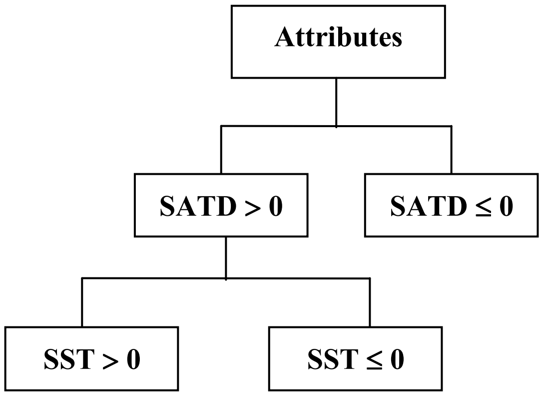

Figure 4. The tree results from decision tree model.

The SATD > 0 means that the sea surface temperature is larger than the air temperature, which indicates the heat flux is upward from sea surface to atmosphere. The more heat flux atmosphere receives, the more typhoon intensity strengthens. The SST is also a major factor to affect typhoon intensity because higher SST can provide more heat flux. This prediction model indicates that when a typhoon passes over the ocean where SATD > 0, about 88% the typhoon intensity is enhanced.

5.2 Effective Assessment

To assess whether the model is effective at making predictions, we use the typhoon JANGMI (2008) data set which does not apply to the training and the testing stages for constructing the model. Table A shows the model predicted results and the actual data of typhoon JANGMI. The classification matrix (Table B) is also created for effective assessment. Because there are only three possible conditions (Y, N, and S) for this predictable attribute, it is easy to know how the model correctly makes a prediction. From Table B we can compute the accuracy of predicted which is 14/17=82.3% and the precision of typhoon enhanced is 12/14=85.7%.

Table A. The Typhoon Intensity Change from Predicted Results (Pred) and Actual Data (Actu)

Pred	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	S	Y	Y	S	Y	N
Actu	Y	Y	Y	S	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	S	N	N

Table B. Classification Matrix of typhoon JANGMI

	N (Actual)	S (Actual)	Y (Actual)	Total
N (Predicted)	1	0	0	1
S (Predicted)	0	1	1	2
Y (Predicted)	1	1	12	14
Total	2	2	13	17

6. CONCLUSIONS

The typhoon intensity changes of category 5 typhoons from 2003 to 2007 in the western North Pacific have been studied using data mining technique. The oceanic and atmospheric parameters in the study area are retrieved from satellite remotely-sensed data and re-analysis data. These parameters include sea surface temperature, water vapor, rain rate, sea surface height anomaly, and air temperature. The algorithm of the data mining technique is C4.5 decision tree model. The results are summarized as follows.

1. The major mechanism of influence on typhoon intensity change is the sea-air temperature difference and the second one is sea surface temperature.
2. About 88% typhoon intensity is enhanced when a typhoon passes over the ocean where its sea surface temperature is larger than air temperature.
3. The predicted model is further validated by using typhoon JANGMI. The accuracy of this model is 82.3% and its precision is 85.7%.

REFERENCES

Alliss, R. L., S. Raman, and S. W. Chang, "Special Sensor Microwave/Imager (SSM/I) observations of hurricane Hugo (1989)," Mon. Weather Rev., vol 120, p.p. 2723-2737, 1992.

Berry, M. J. A. and G. Linoff, "Data mining technique for marketing, sale, and customer support," John Wiley & Sons Inc., p.p. 243-285, 1997.

Berson, A., S. Smith, and K. Thearling, "Building data mining applications for CRM," McGraw-Hill, p.p.45-46, 1999.

Frawley, W., G. Piatetsky-Shapiro, and C. Matheus, "Knowledge discovery in database: An overview," AI Magazine, p.p. 213-228, 1992.

Goni, G. J., and J. A. Trinanes, "Ocean thermal structure monitoring could aid in the intensity forecast of tropical cyclones," Eos, Trans. Amer. Geophys. Union, vol 84, p.p. 573-580, 2003.

Han, J., and M. Kamber, "Data mining: concepts and techniques," 2nd Ed., Morgan Kaufmann, 2006.

Lin, I. I., C. C. Wu, K. A. Emanuel, I. H. Lee, C. R. Wu, and I. F. Pun, "The interaction of supertyphoon Maemi (2003) with a warm ocean eddy," Mon. Weather Rev., vol 133, p.p. 2635-2649, 2005.

Lin, I.-I., C. H. Chen, I. F. Pun, W. T. Liu, and C. C. Wu, "Warm ocean anomaly, air sea fluxes, and the rapid intensification of tropical cyclone Nargis (2008)." Geophys. Res. Lett., vol 36, L03817, 2009.

Quinlan, J. R., "C4.5: Programs for machine learning," Morgan Kaufmann, 1993.

Ronald, S. S. "Accelerating customer relationships - using CRM and relationship technologies, Prentice-Hall, p.p. 51-56, 2001.

Schade, L. R., and K. A. Emanuel, "The ocean's effect on the intensity of tropical cyclones: Results from a simple coupled atmosphere – ocean model," J. Atmos. Sci., vol 56, p.p. 642-651, 1999.

Shay, L. K., G. J. Goni, and P. G. Black, "Effects of a warm oceanic feature on Hurricane Opal," Mon. Weather Rev., vol 128, p.p. 1366-1383, 2000.

Wu, C. C., C. Y. Lee and I. I. Lin, "The effect of the ocean eddy on tropical cyclone intensity," J. Atmos. Sci., vol 64, p.p. 3562-578, 2007.

ACKNOWLEDGEMENTS

The authors appreciate the Japan Meteorological Agency (http://www.jma.go.jp/) for providing the typhoon data, the remote sensing system (http://www.ssmi.com/) for providing sea surface temperature, atmospheric water vapor, and rain rate data, the Archiving Validation and Interpretation of Satellite Oceanography (AVISO) (http://www.aviso.oceanobs.com/) for providing the sea surface height anomaly data, as well as the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA) for providing the re-analysis air temperature data.