Pattern Recognition of Meteorological fields Using Self-Organizing Map (SOM)

Koji Nishiyama

Research associate, Institute of Environmental Systems, Faculty of Engineering, Kyushu University, Fukuoka, Japan

Shinichi Endo

Master student, Faculty of Engineering, Kyushu University, Fukuoka, Japan

Kenji Jinno

Professor, Institute of Environmental Systems, Faculty of Engineering, Kyushu University, Fukuoka, Japan

Abstract: In order to systematically and visually understand well-known but qualitative and relatively complicated relationships between synoptic fields in the BAIU season and heavy rainfall events in Japan, these synoptic fields were classified using the Self-Organizing Map (SOM) algorithm. This algorithm can convert complex nonlinear features into simple two-dimensional relationships, and was followed by the application of the clustering techniques of the U-matrix and the K-means. It was assumed that the meteorological field patterns be simply expressed by the spatial distribution of wind components at the 850 hPa level and Precipitable Water (PW) in the southwestern area including Kyushu in Japan. Consequently, the synoptic fields could be divided into eight kinds of patterns (clusters). One of the clusters has the notable spatial feature represented by high PW accompanied by strong wind components known as Low-Level Jet (LLJ). The features of this cluster indicate a typical meteorological field pattern that frequently causes disastrous heavy rainfall in Kyushu in the rainy season. From these results, the SOM technique may be an effective tool for the classification of complicated non-linear synoptic fields.

Keywords: Self-Organizing Map (SOM); Pattern recognition; Low Level Jet; Precipitable Water; Heavy rainfall

1. INTRODUCTION

During the rainy season (BAIU) in Japan, heavy rainfall events frequently occur due to the intrusion of warm and humid air into a stationary front known as the ‘BAIU front’. Many heavy rainfall events have caused flooding and serious damages to human life and properties all around Japan, especially, in Western Japan. However, it is difficult to diagnose whether heavy rainfall events occurs in a specific area because they are non-linearly connecting with many meteorological factors. Therefore, it will be desirable to systematically identify the complicated non-linear interrelationships among the factors associated with heavy rainfall events using pattern recognition techniques.

A pattern recognition technique called the Self-Organizing Map (SOM) suggested by Kohonen (1995) is a potential method for systematically investigating the complicated non-linearity. The SOM represents an unsupervised ANN and has the powerful ability to project high-dimensionally non-linear features onto visually understandable two-dimensional map. Therefore, the SOM technique has been widely used in the recent years in various research fields for recognizing non-linear properties inherent in each study field. For example, Nikkila et al. (2002) applied the SOM for the analysis and visualization of gene expression. Park et al. (2003) applied it for the recognition of aquatic insect species richness in running water. In the field of meteorology and climate, the SOM was applied for the classification of monsoon climate modes in Southeastern Arizona (Cavazos et al., 2002) and the classification of winter large-scale atmospheric circulation and humidity fields related with extreme rainfall events in Northeastern Mexico and Southeastern Texas (Cavazos, 1999). In these climate studies, separation of visually clear-cut climate modes from complicated non-linear relationships was achieved.

Therefore, the aim of this study is to systematically and visually classify well-known but qualitative and relatively complicated relationships between meteorological fields and heavy rainfall events in the BAIU season using the SOM algorithm.

2. SELF-ORGANIZING MAP

This study involves the classification of complicated synoptic field in the BAIU season according to the procedures shown in Fig. 1. For the classification, this study utilizes using the SOM method in combination with the clustering technique based on Park et al. (2003), Lopez and Machon (2004), and Beccali et al. (2004).

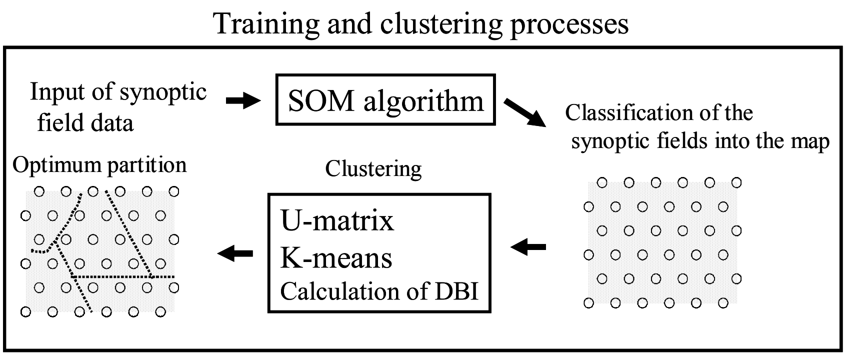

Fig. 1 Training process for the classification of synoptic fields and validation process for the identification of the heavy rainfall events.

First, the meteorological fields are classified into two-dimensional units through the training of the SOM.

Next, all the synoptic field patterns are divided into several clusters using the clustering techniques of the U-matrix method suggested by Ultsch and Siemon (1990) and the following K-means method, in order to visualize and better understand the features of synoptic fields more visually. The optimal number of clusters is determined based on the Davies-Bouldin Index (DBI) suggested by Davis et al (1979).

The SOM is an ANN characterized by unsupervised training. It is an algorithm for projecting high dimensional, complex target data (input data) onto two-dimensional regularly-arranged units. Therefore, it is a useful tool for the visualization and classification of the complex input data space or for data mining. As shown in Fig. 2, the basic structure of the SOM algorithm consists of an input layer and a competition layer. The two-dimensional arrangement of the units in the competition layer (hereafter, named map) takes a hexagonal form having the same distance between neighboring units. Each unit i has a weight vector $\boldsymbol{m}_i$, which is updated through the training process of the SOM explained in the next subsection.

Finally, the updated weight vector represents common features of the input data classified into each unit. The wei ght vector $\boldsymbol{m}_i$ (t) has the same dimension as an input vector x. Prior of the SOM training, the elements of a weigh t vector are initialized on a random basis, and the input vector elements are normalized within the extent of 0 to 1 using the maximum and minimum of each element.

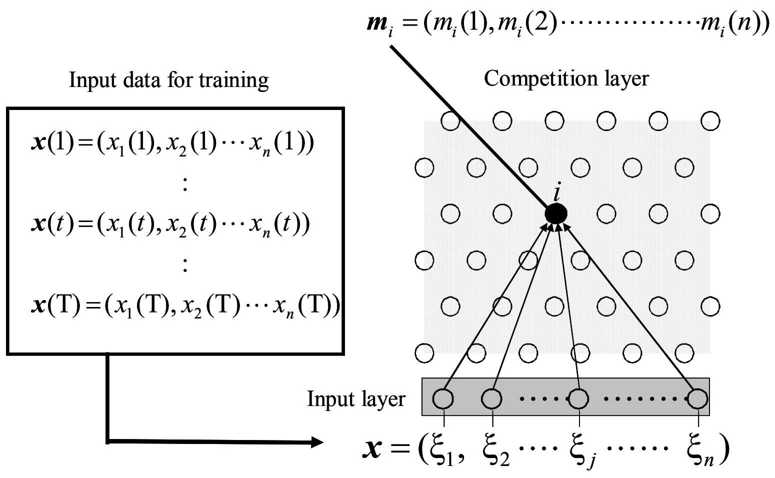

Fig. 2 Architecture of the SOM network consisting of input and competition layers.

The first step is to calculate the Euclidean distance between an input vector and all the weight vectors and, subsequently, to find the 'winner' unit c including the closest weight vector to the input vector as shown by Eq.(1). The 'winner' unit c is called as the Best Matching Unit (BMU).

$$\mathrm{c} = \arg\min_{\mathrm{i}} \left\{ \left\| \boldsymbol{x}(t) - \boldsymbol{m}_{\mathrm{i}}(t) \right\| \right\} \tag{1}$$

The next step is to update all the weight vectors against the presentation of the input vector according to Eqs.(2) and (3).

$$\boldsymbol{m}_i(t+1)=\boldsymbol{m}_i(t)+\mathrm{h}_{\mathrm{ci}}\left(t,\|\boldsymbol{r}_c-\boldsymbol{r}_i\|\right)\left[\boldsymbol{x}(t)-\boldsymbol{m}_i(t)\right] \tag{2}$$

$$\mathrm{h}_{\mathrm{ci}}\left(t,\|\boldsymbol{r}_c-\boldsymbol{r}_i\|\right)=\alpha(t)\cdot\exp\left(-\frac{\|\boldsymbol{r}_c-\boldsymbol{r}_i\|^2}{2\sigma^2(t)}\right) \tag{3}$$

The modification of the weight vectors depends on the neighborhood function represented by $h_{ci}(t,\|\boldsymbol{r}_\mathrm{c}-\boldsymbol{r}_\mathrm{i}\|)$ that takes a Gaussian form, which decreases with the distance from the BMU. Therefore, weight vectors in a unit close to the BMU are modified more strongly. In addition, the modification of the weights decrease with the iteration step t because of a monotonic decrease in the learning-rate and the width of the neighborhood function $\sigma(t)$, which determines the Gaussian form, according to Eqs.(4) and (5).

$$\alpha(t)=\alpha(0)\frac{T-t}{T} \tag{4}$$

$$\sigma(t+1)=1+(\sigma(t)-1)\frac{T-t}{T} \tag{5}$$

Here, T is the total number of iteration steps. $\alpha(0)$ and $\sigma(0)$ are set 0.2 and 5, respectively.

In this case, it should be noted that a series of computation procedures from (1) to (5) must be repeated at least 500 times the number of units in the domain so as to keep the stability of the SOM training, as shown by Kohonen (1995). In this study, input data are used repeatedly to satisfy the condition. The iterative calculation proceeds towards the projection of similar data samples in the high dimensional, complex input data space to an identical unit area in the map. Consequently, neighboring units in the map are similar to each other and, on the other hand, distant units are dissimilar.

3. SPECIFICATION OF INPUT VECTOR

The input vector for the SOM training is determined using objectively analyzed meteorological Grid Point Value (GPV), routinely issued by the JMA (Japan Meteorological Agency) for providing the initial conditions of the numerical prediction models. The GPV is three-dimensionally interpolated grid data, based on upper air soundings obtained by meteorological balloons twice a day (00 UTC, 12 UTC) as well as other kind of data sets (e.g., ground meteorological data, satellite data). The structure of the GPV is represented by a horizontal mesh of 20km × 20km and 21 vertical layers between the ground and the 100 hPa level. The GPV data consist of wind (velocity and direction), temperature, dew point depression, and geo-potential height as a function of pressure.

Considering that the remarkable synoptic features associated with heavy rainfall in the rainy season implies the intrusion of a large amount of water vapor into the BAIU front with the Low-Level Jet (LLJ) by shown by Akiyama (1973) and Ninomiya (2000), these features are simply represented by three variables, Precipitable Water (*PW*) and the two components of the wind vector at the 850 hPa level (*U*, *V*).

In this study, the selected area for extracting target heavy rainfall amounts is Northern Kyushu enclosed by the bold square depicted in Fig. 3, which has an extension of approximately 10^4km^2. In addition, the southwestern area of the Japan Islands is selected as the area for identifying meteorological patterns in the rainy season using the SOM algorithm. Further, the area is divided into nine sub-areas, as shown in Fig. 3.

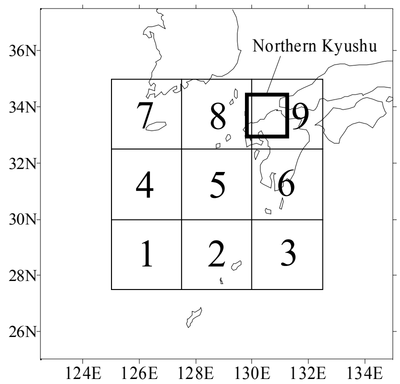

Fig. 3 The target area for extracting meteorological patterns consists of nine sub-areas including the Kyushu Islands. On the other hand, Northern Kyushu enclosed by the bold square is the area used for the extraction of rainfall amounts.

Therefore, the input vector $\boldsymbol{x}(t)$ utilized for the SOM training is represented by 27 elements consisting of *PW*, *U* , and *V* as follows.

$$\boldsymbol{x}(t) = (PW_1 \cdots PW_9 \ U_1 \cdots U_9 \ V_1 \cdots V_9) \tag{6}$$

Here, the subscript of each element represents the number of the sub-area. Each element in (6) is calculated as an area-averaged value in a sub-area including data from a number of GPV nodes, and is treated as a normalized value between 0 and 1 according to Eq. (7).

$$\widetilde{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{7}$$

Here, X denotes an original non-normalized value. $\widetilde{X}$ denotes normalized values of PW, U, and V using the maximum X_{max} and minimum X_{min} within the range of X values.

4. CLASSIFICATION OF METEOROLOGICAL FIELDS

For the SOM training, 366 input vectors obtained in the rainy season (June and July) during the period between 1996 and 1998 were used. Input vectors presented into the SOM algorithm were classified into the units arranged in a map consisting of 10×10 hexagonal grid points (i.e., 100 types of meteorological patterns). In addition, the application of the clustering techniques based on the U-matrix and K-means methods to the trained SOM yielded the optimum configuration consisting of eight clusters as shown in Fig. 4.

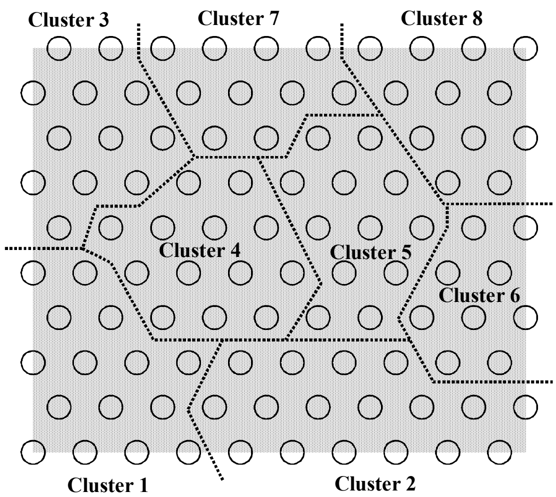

Fig. 4 Optimum configuration determined based on the training and clustering processes.

In this section, we investigate the meteorological features of the eight clusters in the configuration, and relate t he features of rainfall occurring in Northern Kyushu to the properties of each cluster. The average features of a m eteorological field characterizing each cluster are shown in Fig. 5. Each characteristic meteorological field expre ssed by PW, U and V was calculated as the average of the weight vectors in each cluster’s units.

The results of the classification roughly indicate five main different meteorological patterns: an entirely dominant dry situation, a typical summertime situation affected by the Pacific high pressure system, a low pressure system, a movement of the BAIU front, and an intrusion of a large amount of water vapor with LLJ.

Dry air masses originating from the north and the existence of a BAIU front in the far south of Kyushu represents the synoptic features of cluster 1.

A typical summertime synoptic patterns represents cluster 2, which is characterized by a large value of PW and clockwise flows affected by a high-pressure system in the Pacific Ocean, although only a part of the clockwise flow can be recognized. Meteorological features of the cluster generally lead to the generation of local thunderstorms in an unstable atmospheric condition. This is because a local atmospheric circulation occurs due to the sunshine effect increasing towards the afternoon in no remarkable synoptic scale disturbance. Therefore, the meteorological features of cluster 2 have a potential to cause local afternoon thunderstorms.

In clusters 3 and 7, the actual meteorological fields of the events seem to be affected by a low pressure system. However, the existence of a low pressure system cannot be clearly verified due to the coarse resolution provided by only nine sub-areas and the location of predetermined target area shown in Fig. 3.

Clusters 4 and 5 are closely related with the movement of the BAIU front. In cluster 4, a steep gradient of PW corresponding to the BAIU front can be confirmed in 26~30 N. On the other hand, in cluster 5, the BAIU front moves furthermore towards the north than in cluster 4 and, therefore, LLJ flows accompanying the transportation of a large amount of water vapor can be confirmed south of the area of strong PW.

Clusters 6 and 8 are characterized by the intrusion of a large amount of water vapor with LLJ into Northern Kyushu. In cluster 8, the axis of the LLJ, extending from the southwest to the northeast, is located just over Northern Kyushu. In this case, Northern Kyushu experiences a strongly unstable atmospheric condition due to the effect of a large amount of water vapor transported by the LLJ. Therefore, the features of this cluster correspond to the typical synoptic situation causing disastrous heavy rainfall in the BAIU season. On the other hand, in cluster 6, the influence of the LLJ shifts further towards the north of Northern Kyushu, compared with cluster 8. Therefore, cluster 6 would represent an intermediate phase between the situations in clusters 2 and 8.

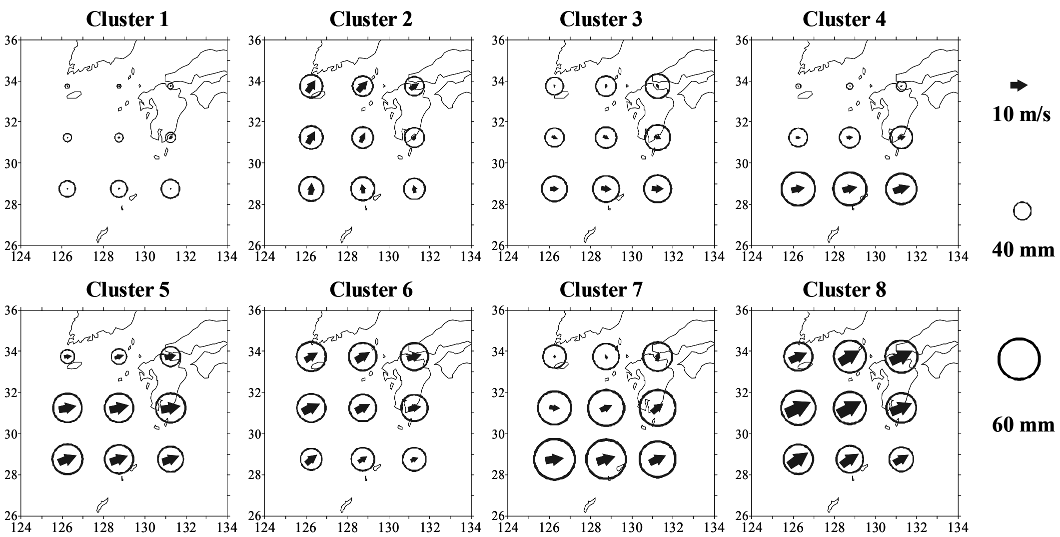

Fig.5 The average feature of each meteorological field calculated as the average of the weight vectors in each cluster's units. The size of a circle in the figures represents PW values.

5. RAINFALL PROPERTIES OF CLASSIFIED FIELDS

In this study, RADAR-AMeDAS rainfall data issued by the JMA are utilized for the selection of rainfall events in comparison with the trained SOM. The RADAR-AMeDAS shows the spatial distribution of hourly rainfall with a 5 km horizontal fine resolution.

By analyzing the spatial distributions of the RADAR-AMeDAS rainfall events occurring in Kyushu in Japan between 1996 and 1999, most of the events causing heavy rainfall could be recognized as having an organized band-like rainfall distribution (e.g., squall line), classified as the Mesoscale Convective Systems (MCS). Since the convective portion of the MCS is the source of flood-producing rainfall, the rainfall related to the strength of the convective activity of the MCS is represented by the maximum hourly rainfall observed during the passage of the MCS within three hours of the GPV data. This amount of rainfall will be utilized for the comparison with the trained SOM in the next section.

Table.1 indicates the frequency of rainfall occurrence corresponding to each cluster within a categorized rainfall range. Here, weak, moderate, and heavy rainfall are defined as 0<R=<10 mm/h, 10<R=<30 mm/h, and R>=30 mm/h, respectively.

Table. 1 Frequency of rainfall events included in each cluster in the trained SOM.

	C1	C2	C3	C4	C5	C6	C7	**C8**	TOTAL
R=0	52	51	11	30	21	13	8	**1**	187
0<R=<10	5	5	18	4	19	14	13	**13**	91
10<R=<20	1	4	6	1	2	3	9	**10**	36
20<R=<30	0	0	5	1	1	5	2	**14**	28
30<R=<40	0	0	1	0	1	3	4	**9**	18
40<R=<50	0	1	1	0	0	0	0	**2**	4
R>50	0	0	0	0	0	0	0	**2**	2
TOTAL	58	61	42	36	44	38	36	**51**	366

There were only few rainfall occurrences in the events corresponding to cluster 1 because the atmosphere had not enough instability to cause rainfall due to the dry atmospheric condition.

However, most events did not cause heavy rainfall because the atmospheric conditions were not always unstable and the detection of target rainfall data was limited to within three hours (0900-1200LST and 2100-2400LST) from the GPV time (0900, 2100JST). Cluster 6, which corresponds to in the intermediate phase between the situations in clusters 2 and 8, includes heavy rainfall events because of the continuous intrusion of water vapor.

Clusters 3 and 7 include many weak and moderate rainfall events and a few heavy rainfall events caused by a low pressure system and relatively large PW. Especially, cluster 7 includes four heavy rainfall cases because of larger PW in cluster 7 than in cluster 3.

Many of the events allocated in clusters 4 and 5 are included in the category of no rainfall because Northern Kyushu was characterized by dry atmospheric conditions and the BAIU front was located away from Northern Kyushu. However, the frequency of weak rainfall in cluster 5 is higher compared with cluster 4 because of larger PW in cluster 5. The situation indicates that the BAIU front approaches towards Northern Kyushu.

Here, in particular, it should be noted that the rainfall events classified into cluster 8 included a higher frequency of moderate and heavy rainfalls than that of the other clusters. This implies strong activity of the BAIU front induced by the intrusion of a large amount of water vapor accompanied by LLJ as shown by Fig. 6, which indicates the rainfall event included in unit 97 (Cluster 8).

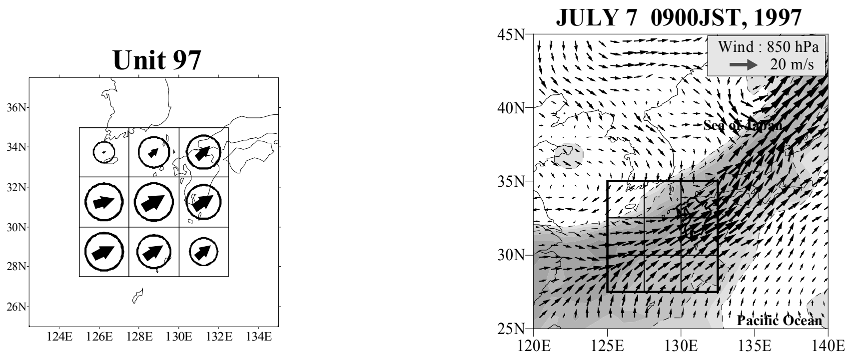

Fig. 6 The left figure indicates the meteorological field characterizing unit 97. The right figure shows a case of the meteorological fields of heavy rainfall events classified into unit 97.

6. CONCLUSIONS

In this study, in order to systematically and visually investigate and understand well-known but qualitative and relatively complicated relationships between meteorological fields in the BAIU season and heavy rainfall events in Japan, these meteorological fields were classified using the Self-Organizing Map (SOM) algorithm. This algorithm can convert complex nonlinear features into simple two-dimensional relationships, and followed by the application of the clustering techniques of the U-matrix and the K-means. It was assumed that the meteorological field patterns be simply expressed by the spatial distribution of wind components at the 850 hPa level and Precipitable Water (PW) in the southwestern area including Kyushu in Japan. Consequently, the synoptic fields could be divided into eight kinds of patterns (clusters), which basically revealed five representative meteorological situations characterized by (1) dry air masses, (2) anti-cyclonic flows due to the Pacific high pressure system, (3) the existence of the BAIU front, (4) the intrusion of a large amount of water vapor with LLJ, and (5) the effect of low pressure system.

One of the clusters has the notable spatial feature represented by high PW accompanied by strong wind components (LLJ). The pattern indicated a typical meteorological field pattern that frequently causes heavy rainfall in Kyushu in the rainy season. From these results, it can be expected that the SOM technique is availa ble as an effective tool for the classification of complicated non-linear synoptic fields and for the identification of the occurrence of heavy rainfall events.

7. REFERENCES

Akiyama, T., 1973. The large-scale aspects of the characteristic features of the Baiu front. *Papers in Meteor and Geophys*. 24, 157-188.

Cavazos, T. 1999. Large-scale circulation anomalies conducive to extreme events and simulation of daily precipitation in northeastern Mexico and southeastern Texas. *J. Climate*. 12, 1506-1523.

Cavazos, T., A. C. Comrie, and D. M. Liverman, 2002. Intraseasonal anomalies associated with wet monsoons in SE Arizona. *J. Climate*. 15, 2477-2490.

Kohonen T., 1995. Self-Organizing Maps. Springer Series in Information Sciences. 30, 362pp.

Ninomiya, K., 2000. Large- and meso-α-scale characteristics of Meiyu and Baiu front associated with intense rainfalls in 1-10 July 1991. *J. Met. Soc. Japan*. 78, 141-157.

Park, Y. S., Cereghino, R., Compin, A., Lek, S., 2003. Application of artificial neural networks for patterning and predicting aquatic insect species richness in running water. *Ecological Modeling.* 160, 265-280.

Ultsch, A., Siemon, H. P., 1990. Kohonen's self organizing feature maps for exploratory data analysis. Proceedin gs of INNC'90, International Neural Network Conference, Kluwer Academic Publisher, Dordrecht, Netherlan ds, 305-307.