PATTERN RECOGNITION OF METEOROLOGICAL ELEMENTS RELATED WITH HEAVY RAINFALL IN JAPAN USING SELF-ORGANIZING MAP (SOM) ALGORITHM

KOJI NISHIYAMA

Faculty of Engineering, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan

YOSHITAKA I

Kumamoto Prefecture, 6-18-1, Suizenji, Kumamoto, 86-28570, Japan

KENJI JINNO

Faculty of Engineering, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan

AKIRA KAWAMURA

Faculty of Engineering, Kyushu University, 6-10-1, Hakozaki, Higashi-ku, Fukuoka, 812-8581, Japan

In this study, using Self-Organizing Map (SOM) algorithm, which organizes complex nonlinear feature into simple two-dimensional relationship, the pattern of precipitable water and wind in designated large area including the Kyushu Island in Japan in a BAIU season was classified.

As a result, it was shown that locally distributed specific neurons in the map correspond to the category of heavy rainfall. Therefore, it may be expected that automatic selection of anticipated rainfall category in ahead using SOM technique without human selection can contribute to rainfall prediction. However, the neurons related with that of light rainfall are distributed at any weather situation. This trend indicates that any weather situation can cause light rainfall due to complicated atmospheric physical processes. It means that the clear-cut separation among rainfall categories requires further improvement of SOM technique.

INTRODUCTION

During a rainy season (BAIU) in Japan, many thunderstorms occur frequently and widely along a BAIU front, which maintains its strength and becomes stationary with repeating a slight movement along a latitudinal direction in the Japan Islands according to dynamical equilibrium between the 'warm' Pacific high pressure system and the 'cold' northern high pressure system. Heavy thunderstorms in this season occur as follows. Since supplied abundant warm and humid air continuously into the BAIU front from the south under the influence of Pacific high pressure contributes to the generation and maintenance of strong atmospheric instability as pointed out by Akiyama [1]. Consequently, heavy thunderstorms occur frequently along the BAIU front. Therefore, quantitative precipitation forecasts in the BAIU season are important for preventing serious disasters involving intense flood due to heavy rainfall.

The present weather prediction for providing useful precipitation guidance in many countries adopts a technique represented by MOS (Model Output Statistics), which means downscaling method of prediction model outputs by relating it to observed data of a specific area using statistical techniques including multiple linear regression and neural network. However, a relationship obtained by statistical techniques in ahead includes many kinds of rainfall categories. In this case, it is anticipated that the relationship strongly affects the accuracy of rainfall prediction. Therefore, pattern recognition techniques for the classification of rainfall category are required. It would be expected that the incorporation of pattern recognition technique into original prediction model as an example shown in Wong et al. [3] contribute to the improvement of the accuracy of rainfall prediction.

Here, Self-Organizing Map (SOM) algorithm is used as one of pattern recognition techniques. The SOM algorithm is a neural network model characterized by unsupervised learning for converting complicated high dimension data consisting of many variables into visually understandable low dimension geometrical relationship classified into some similar patterns of groups, developed by Kohonen [2]. This method is often used for the pattern recognition of complicated interrelationship among variables.

In this study, using SOM algorithm, which organizes complex nonlinear feature into simple two-dimensional relationship, the pattern of precipitable water and wind in designated large area including the Kyushu Island in Japan in a BAIU season are classified into some kinds of features characterizing a BAIU season.

OUTLINES OF GPV

The GPV (Grid Point Value) is three-dimensionally interpolated grid data, based on the sounding data set observed by meteorological balloons twice a day at the same time (00 UTC, 12 UTC) in the world and the other data sets (ground meteorological data, satellite data, e.g.). The GPV data is constructed by the assimilation of these observed data into three-dimensionally interpolated grid data set through the calibration using the predicted data by the numerical prediction model of Japan Meteorological Agency (JMA). In addition, the GPV is modified for satisfying some physical relationships included in an actual atmosphere represented by the specific mesh size of 20 km, which is taken for the construction of the GPV of the specific area including Japan. The GPV on the basis of the above-mentioned procedures consists of wind (velocity and direction), temperature, dew point depression, geo-potential height with the function of pressure. The structure of the GPV is represented by the horizontal mesh of 20 km and the irregular vertical meshes divided into 21 layers between the ground and 100 hPa level. The GPV is available for various purposes as well as an initial condition of the prediction model.

However, since the GPV strongly depends on predicted outputs by JMA model, it is unavailable for the diagnosis of heavy rainfall or squall line if the predicted outputs give inconsistent results with actual observed facts. In addition, the GPV has a drawback that it is unavailable for finding good correspondence of indices with heavy rainfall excepting the designated times, 0900 and 2100 JST. Therefore, the GPV should be treated carefully for the analysis of heavy rainfall, paying much attention to the above-mentioned features.

APPLICATION OF SOM ALGORITHM

In this study, many varieties of features in a BAIU season in Japan are classified and self-organized into 6×6 neurons of two-dimensional display map as shown in Figure 1a. $\hat{\mathbf{m}}_{i,j}$, an optimum model vector at a specific location (*i, j*), indicates one of 36 kinds of features in a BAIU season. On the other hand, $\mathbf{x}(t)$ is an input vector at regression step *t* for the learning of SOM and consists of GPV-based dominant variables characterized by a BAIU season. Here, $\mathbf{m}_{i,j}(t)$ is defined as model vector at a specific location of (*i, j*) at regression step *t*. If the number of input vectors is defined to be N, the modification of $\mathbf{m}_{i,j}(t)$ through N-times repetition of the SOM procedures given below yields an optimum model vector, $\mathbf{m}_{i,j}(N)=\hat{\mathbf{m}}_{i,j}$.

After the elements of $\mathbf{m}_{i,j}(1)$ are initialized randomly and normalized, the following procedures are repeated for finding all the optimum model vectors in the map. By comparing an input vector $\mathbf{x}(t)$ with all the model vectors, the first procedure is to find the location (*I, J*) in the map where $\| \mathbf{x}(t)-\mathbf{m}_{i,j}(t) \|$ shows minimum. Here, the neuron at the location (*I, J*) is called as 'winner neuron'. The next procedure is to modify the model vector, $\mathbf{m}_{i,j}(t)$ according to Eq. (1) and (2).

$$\mathbf{m}_{i,j}(t+1)=\mathbf{m}_{i,j}(t)+h(t,\| \mathbf{r}_{i,j}-\mathbf{r}_{I,J} \|)(\mathbf{x}(t)-\mathbf{m}_{i,j}(t)) \tag{1}$$

$$h(t,\| \mathbf{r}_{i,j}-\mathbf{r}_{I,J} \|)=\alpha(t)\exp\left(-\frac{\| \mathbf{r}_{i,j}-\mathbf{r}_{I,J} \|}{2\sigma^2(t)}\right) \tag{2}$$

$\mathbf{r}_{i,j}$ and $\mathbf{r}_{I,J}$ are vector location at any grid point in the map and at the grid point (*I, J*) of winner neuron, respectively. Since $\alpha(t)$ and $\sigma(t)$ are the learning-rate factor and the width of the neighborhood function, respectively, which decrease monotonically with an increase in the regression step *t*, $h(t,\| \mathbf{r}_{i,j}-\mathbf{r}_{I,J} \|)$ defined by the neighborhood function also decreases with monotonically with an increase in the regression step *t*. Therefore, $\mathbf{m}_{i,j}(t)$ of adjacent neurons to 'winner' are strongly affected and, consequently, modified into, $\hat{\mathbf{m}}_{i,j}$, an optimum model vector in the map after the repetition of these procedures.

In this study, it is assumed that precipitable water and wind vector characterize dominant features in a BAIU season, which indicate the intrusion of warm and humid air into the BAIU front with its location as discussed in the previous section. Here, as shown in Figure 1b, 9 sub-areas in the large square area including the Kyushu Islands are specified. In each sub-area, precipitable water (*PW*) and 2 components (*u, v*) of wind vector (850 hPa) are averaged and inputted into the components of $\mathbf{x}(t)$. These factors are treated as 9 dimensional vectors, expressed by $\mathbf{x}_{\mathrm{PW}}$, $\mathbf{x}_{\mathrm{u}}$, $\mathbf{x}_{\mathrm{v}}$. Therefore, the input vector, $\mathbf{x}(t)$, has 27 dimensions as confirmed Eq. by (3) and (4).

$$\mathbf{x}(t)=(\mathbf{x}_{\mathrm{PW}}\ \mathbf{x}_{\mathrm{u}}\ \mathbf{x}_{\mathrm{v}}) \tag{3}$$

$$\mathbf{x}_{\mathbf{PW}} = (PW_1(t),\cdots,PW_9(t))\ ,\ \mathbf{x}_{\mathbf{u}} = (u_1(t),\cdots,u_9(t))\ ,\ \mathbf{x}_{\mathbf{v}} = (v_1(t),\cdots,v_9(t)) \tag{4}$$

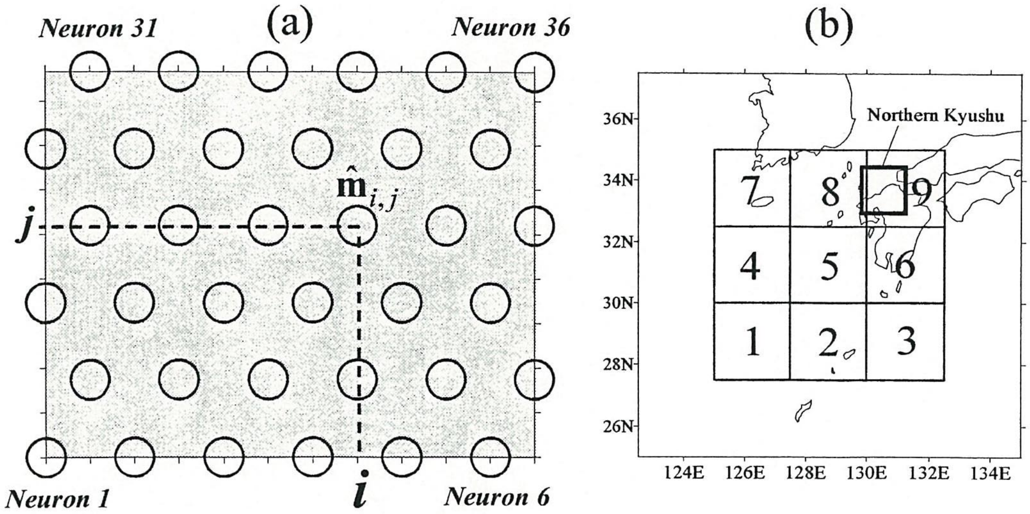

Figure 1. (a) Arrangement of 36 neurons on the SOM. (b) The specification of 9 sub areas in the large square area including the Kyushu Islands enclosed by bold square. In each sub-area, precipitable water (*PW*) and 2 components (*u, v*) of wind vector (850 hPa) are averaged and used as an input vector for the learning of SOM algorithm.

DOMINANT FEATURES IN A BAIU SEASON

From the procedures in the previous section, complicated relationship among 27 variables could be classified into 36 kinds of features characterized in a BAIU season in the display map using SOM algorithm. For the classification of the features, the unsupervised learning by SOM algorithm was done using input vectors obtained from the data set of GPV in June and July between 1996 and 1998. Here, it should be noted that rainfall data obtained in the Northern Kyushu enclosed by bold square in Figure 1b are not used for the learning of SOM algorithm.

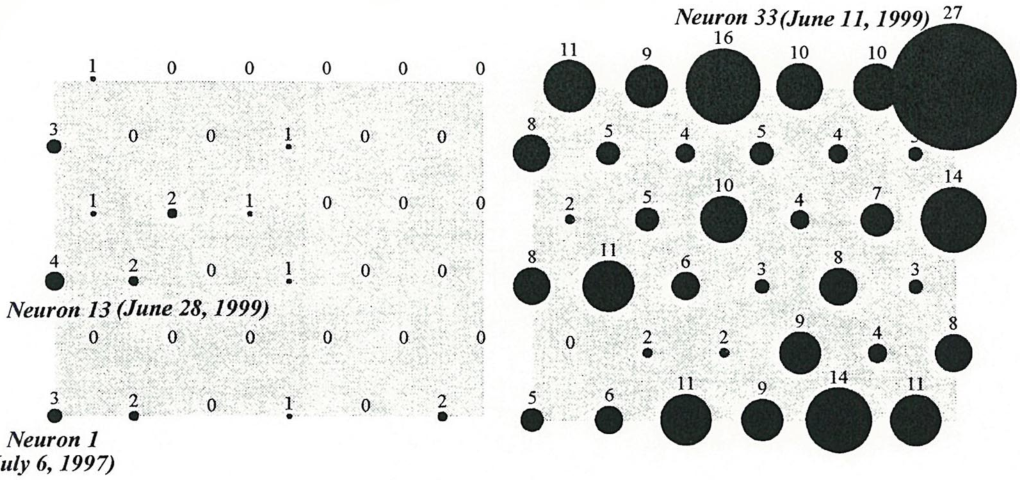

Figure 2. The left figure indicates the distribution of the neurons reacting to input vectors included in the category of heavy rainfall more than 30 mm/h. On the other hand, the right figure indicates that of the neurons reacting to those of light rainfall less than 10 mm/h.

In this study, notable features of a model vector in a neuron reacting to the category of heavy or light rainfall are compared with an actual weather situation on the day when heavy or light rainfall occurred, and some relationships between neurons and the field of the moisture and wind characterizing a BAIU season are discussed. Here, rainfall amount are selected as the radar-AMeDAS maximum hourly rainfall within 3 hours after 12 hours from the GPV time (0900, 2100JST) in the Northern Kyushu enclosed by bold square area in Figure 1b.

The left figure of Figure 2 indicates the distribution of the neurons reacting to input vectors included in the category of heavy rainfall more than 30 mm/h. On the other hand, the right figure indicates that of the neurons reacting to those of light rainfall less than 10 mm/h. As a result, the neurons related with category of heavy rainfall are localized to the left side of the map although heavy rainfall frequency is short. This trend indicates some specific weather situations characterized by the model vector in the neurons are required for the generation of heavy rainfall. On the other hand, the neurons related to that of light rainfall are distributed at any neuron. This trend indicates that any weather situation can cause light rainfall due to complicated atmospheric physical processes and, therefore, it is difficult to find key factors for the classification between heavy and light rainfall.

Next, our attention is focused on notable relationships between the elements of the model vector in the neuron related with heavy or light rainfall and the associated actual weather situations. Here, two heavy rainfall (July 6, 1997 and June 28, 1999) cases reacting to the neurons of No. 1 and 13 are shown in Figure 3. On the other hand, the above-mentioned light rainfall case reacting to the neuron of No.33 is shown in Figure 4.

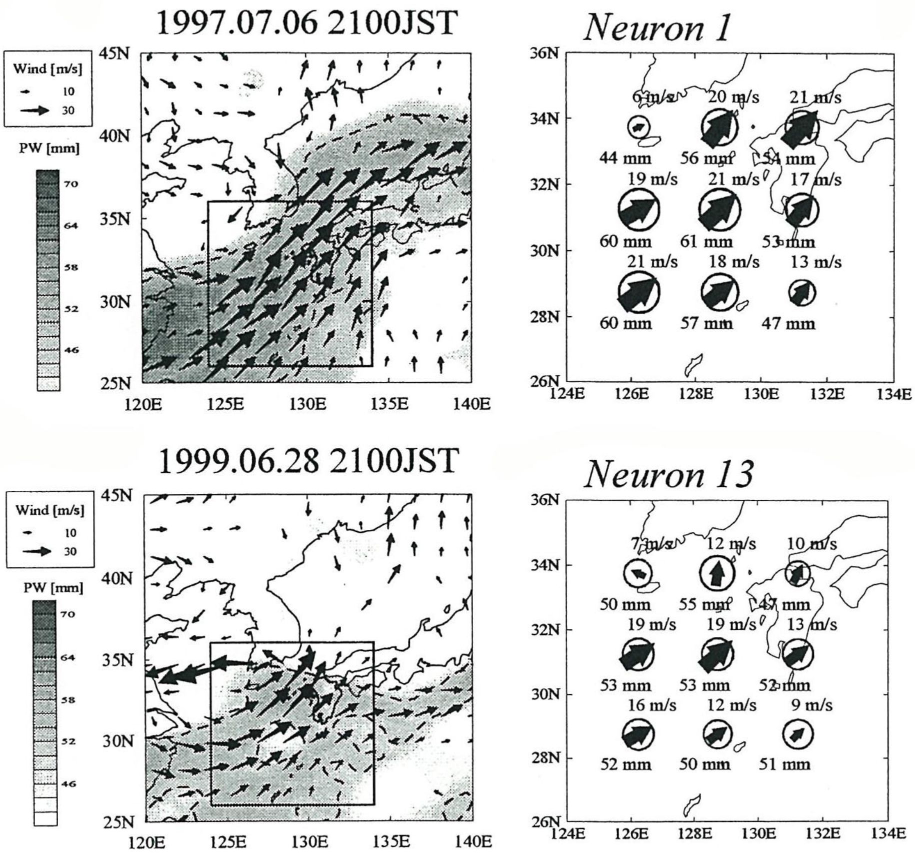

Figure 3. Two heavy rainfall (July 6, 1997 and June 28, 1999) cases reacting to the neurons of No. 1 and 13. The left figures indicate examples of actual weather situations adjacent to the feature of each neuron. The right figures indicate notable feature of heavy rainfall cases in a BAIU season in each neuron.

As a result, the model vectors in the neurons (No. 1 and 13) located in the left edge of the map and the corresponding weather situations have the common outstanding feature characterized by the flow and accumulation of a large amount of water vapor with strong south-western wind along the west side of the Pacific anti-cyclone. Particularly, the event of June 28, 1999 led to heavy rainfall more than 70 mm/h at some observational points in Fukuoka. The flow of a large amount of moisture along the Pacific high caused the formation of large PW can be confirmed along the lower jet stream and characterizes the field of heavy rainfall. Therefore, the localized neurons in the left edge are related with dominant weather situation of heavy rainfall.

On the other hand, the model vectors in the neuron of No. 33 located in the upper-middle side of the map. In this case, active and stationary BAIU front was located over the western area of the Pacific Ocean. No rainfall occurred in the Northern Kyushu because of no synoptic disturbance in the northern side of the front. Around the Kyushu

Islands, weather situation has a feature characterized by dry weather situation affected by dry northern wind in the northern side of the BAIU front.

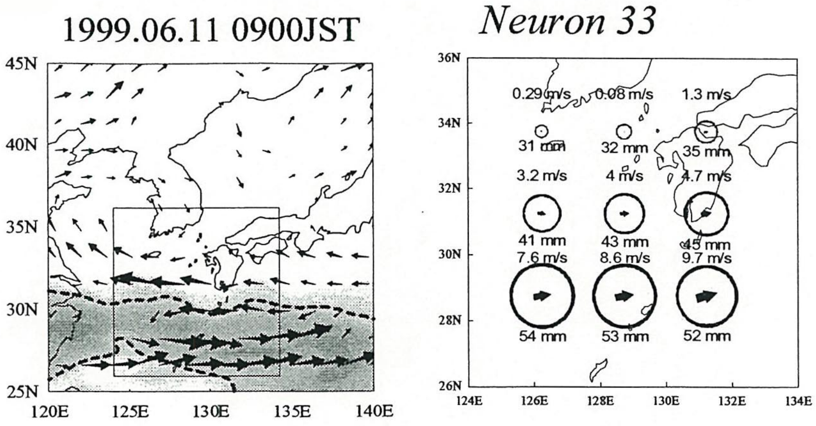

Figure 4. Light rainfall case reacting to the neuron of No.33. The left figure indicates an example of actual weather situation adjacent to the feature of the neuron. The right figure indicates notable feature of light rainfall case in a BAIU season in the neuron.

From the above-mentioned analysis, by further improvement of SOM technique, it can be expected that SOM model classifies dominant features in a BAIU season depending on the amount of precipitable water and the direction and strength of wind with the location of a BAIU front.

CONCLUSION

In this study, using Self-Organizing Map (SOM) algorithm, which organizes complex nonlinear feature into simple two-dimensional relationship, the pattern of precipitable water and wind in designated large area including the Kyushu Island in Japan in a BAIU season was classified. For the classification, complicated relationship among 27 variables consisting of precipitable water and 2 components of wind in 9 sub-areas, obtained from the data set of GPV in June and July between 1996 and 1998, could be classified into 36 kinds of features characterizing a BAIU season in the display map using the unsupervised learning by SOM algorithm.

As a result, it was shown that locally distributed specific neurons in the map correspond to the category of heavy rainfall. The outstanding feature of the neurons is characterized by the flow and accumulation of a large amount of water vapor with strong south-western wind along the west side of the Pacific anti-cyclone. Therefore, it may be expected that automatic selection of anticipated rainfall category in ahead using SOM technique without human selection can contribute to rainfall prediction.

However, the neurons related with that of light rainfall are distributed at any neuron. This trend indicates that any weather situation can cause light rainfall due to complicated atmospheric physical processes. It means that the clear-cut separation among rainfall categories requires further improvement of SOM technique.

It would be expected that the application of the techniques to a part of rainfall prediction procedures contribute to the improvement of the accuracy of rainfall prediction.

REFERENCES

1. Akiyama, T., "The large-scale aspects of the characteristic features of the Baiu front", *Papers in Meteor and Geophys*, Vol. 24, pp. 157-188, (1973).
2. Kohonen T., "Self-Organizing Maps", Springer Series in Information Sciences, Vol. 30, (1995).
3. Wong K W., Gedeon T D., Fung C C. and Wong P M., "Spatial Interpolation Using Neural Fuzzy Technique ", *Proceedings of 8th International Conference on Neural Information Processing (ICONIP)*, Shanghai , China, p.4 , (2001).