Deep Sea Research Part I: Oceanographic Research Papers

Yukiharu Hisaki

Classification of surface current maps

This is an Author's Accepted Manuscript of an article published in

Deep Sea Research Part I

Volume 73, March 2013, Pages 117-126

Available online at: https://doi.org/10.1016/j.dsr.2012.12.001

Classification of surface current maps

Yukiharu Hisaki[a,*]

[a] *University of the Ryukyus, 1-Aza Senbaru, Nishihara-cho, Okinawa, 903-0213 Japan. Tel. +81-98-895-8515; Fax: +81-98-895-8552*

Abstract

Classification of ocean data maps is important for analysis of ocean data. Here, we compare Self-Organizing Map (SOM) analysis with cluster methods such as the Ward method and K-means method. The HF (high-frequency) radar surface current data east of Okinawa Island, Japan were used for the comparison. There are two typical current patterns in the observation area: a strong southward current and a clockwise eddy-like current pattern. The classification results by the Ward method was similar to that by the SOM analysis. SOM analysis was insensitive to the cut-off Empirical Orthogonal Function (EOF) mode number for reducing the data dimensions and noise, while the K-means method was the most sensitive to the EOF mode number.

Keywords: Cluster analysis; Ward method; K-means method; SOM; EOF

1. Introduction

Ocean data such as currents, temperatures and salinities are functions of time and space. One of the important analyses for ocean data is the classification of physical features according to contour or vector patterns in

*Corresponding author
Email address: `hisaki@sci.u-ryukyu.ac.jp` (Yukiharu Hisaki)

maps. Analysis of pattern classifications of maps is less common in physical oceanography than in with meteorology, although there are some studies of pattern classification in physical oceanography. For example, Harms and Winant (1998) classified surface current maps in the Santa Barbara Channel manually.

It is useful to classify maps objectively. One of the methods of objective classification is cluster analysis. The applications of cluster analysis to physical oceanography are fewer than those to meteorology. Cluster analysis is used in physical oceanography to divide two- or three-dimensional graphs or maps from multivariate data (e.g., Freeman et al. (2012); Hasegawa and Hanawa (2003); You (1997)) such as water type identification or remotely sensed data. However, the dimensions of the data are not as high as those of the data for map classification.

Another method of objective classification is Self-Organizing Map (SOM) analysis (Kohonen (2001)). We also conducted SOM analysis. The philosophy behind the SOM and its application to remotely sensed oceanographic data are described in Richardson et al. (2003). SOM application to in-situ or remotely sensed ocean current data was reported in Liu and Weisberg (2005), Liu et al. (2006) and Liu et al. (2007). A review of SOM applications in oceanography and meteorology is provided in Liu and Weisberg (2011).

One of the methods to extract spatial patterns in ocean data, which is often used in physical oceanography, is Empirical Orthogonal Function (EOF) analysis. The EOF method extracts identical spatial patterns in data, and time series of weights describe their evolution in time. It is possible to classify the spatial pattern from the time series of weights .

There are some oceanographic studies that compare EOF with SOM (Liu and Weisberg (2005); Liu et al. (2006); Mau et al. (2007)). There are also some oceanographic studies which compare SOM analysis with cluster analysis. For example, Camus et al. (2010) compared non-hierarchical cluster analysis with SOM. The analyzed data were wave parameters such as significant wave height, mean period, and mean wave direction at a single observation point, and the dimensions of the data were not very high. It is better to compare hierarchical cluster analysis with other classification methods if possible, because the dissimilarities of the maps from the dendrograms are more apparent.

The dimensions of data for the classification of maps high: The dimensions of the data in each map are twice the number of data positions in the case of two-dimensional ocean currents. However, it is well known that doing cluster analysis in higher-dimensional space is more difficult because of the so-called "curse of dimensionality" (e.g., Frédérique and Aires (2009)).

EOF analysis is a simple method to reduce the dimensions of a data map. The observed data are reconstructed by the leading EOF time coefficients and eigenvectors. Therefore, the dimensions of the data map are the number of leading EOF modes.

We have used surface current data obtained by HF (high-frequency: 3 – 30 MHz) radar for analysis. The observation area is the open ocean, and currents in the area are affected by mesoscale eddies. Some studies have used EOF analysis for HF radar surface current data (e.g., Kaihatu et al. (1998); Marmorino et al. (1999); Hisaki (2006)). A few studies have used SOM analysis for HF radar surface current data (e.g., Liu et al. (2007); Mau et al. (2007)). There are no studies that have used cluster analysis for HF radar surface current data.

The objective of this paper is to classify surface current patterns observed by HF ocean radar and to compare the various classification methods.

Section 2 describes the current measurement by HF radar. The methods of the classification are also described in Section 2. Section 3 presents the results of the comparison. The results are discussed and conclusions are given in Section 4.

2. Methods

2.1. EOF and reduction of dimensions

The EOF method is the same as that described in Kaihatu et al. (1998). The EOF method is also described in Hisaki (2006). We did not remove the mean current field as Kaihatu et al. (1998) did, because we also compare EOF eigenfunctions with classified currents by other methods.

The reconstructed current at the time t and the position $\mathbf{x}$ is

$$V_{(m)}(\mathbf{x}, t) = \sum_{k=1}^{N_E} b_k(t)\Psi_{(m)}, \qquad m = 1, 2 \tag{1}$$

where $(V_{(1)}, V_{(2)})$ is the reconstructed current, $b_k(t)$ is the time coefficient for the EOF mode k, and $\Psi_{(m)} = \Psi_{(m)}(\mathbf{x})$ is the eigenfunction. The cut-off EOF mode number N_E is less than $2N_g$, where N_g is the number of positions. The dimensions of the data map were reduced from $2N_g$ to N_E.

2.2. Cluster analysis and SOM

The Ward method was applied as a hierarchical cluster analysis, and the K-means method was applied as a non-hierarchical cluster analysis. The International Mathematical Statistical Library (IMSL) was used for the cluster analysis.

K-means clustering depends on the initial guess of the cluster centers. The initial guess of the cluster groups is based on the time series number, because the current maps are changed gradually. The initial guess of the cluster centers for each cluster group is based on the groups clustered according to time.

SOM analysis evaluates the weight vectors $\mathbf{m}_i$ and the best-matching unit (BMU) c_k (Equation (1) in Liu and Weisberg (2005)), where i is the unit number and k is the time series number. The SOM method is described in Liu and Weisberg (2005) and Liu et al. (2006). The algorithm and parameters of the SOM method are described in Appendix A.

2.3. Observation of surface current by HF radar

The HF surface current data observed in 1998 were used for analysis. Figure 1 is the location of the radars and the observation area. The HF radars were located on the east coast of Okinawa (Ryukyu) Island, mapping the surface currents east of Okinawa Island.

The observation area is close to the Ryukyu trench, where the water depth deepens rapidly with distance offshore. Therefore, the HF radar observation area can be characterized as the open ocean The water depths in most of the observation area are greater than 200 m (Figure 1). The mean currents in the observation area are small, and their variabilities are affected by mesoscale eddies, while currents in the west of the island are affected by the recirculation current of the Kuroshio (Hisaki and Imadu (2009)).

The radar frequency was 24.515 MHz, and the Bragg frequency was $f_B =$

0.505 Hz. The radar system was a phased array system. The range resolution of the radar was 1.5 km, and the beam resolution was 7.5°. The HF ocean radar measured surface currents every 2 hours from shore sites at locations A (26° 7.19' N, 127° 45.78′ E) and B (26° 18.63′ N, 127° 50.25′ E) in Figure 1. The details of the HF radar observation are described in Hisaki et al. (2001). The radial currents are interpolated on the grid points in Figure 1 with respect to time and space. The number of grid points was $N_g = 355$. The daily-averaged HF radar currents are used for the analysis.

The EOF analysis for 2-hourly currents was conducted in Hisaki (2006). The effect of the tide on currents was out of the scope of the present study. It is possible to draw a dendrogram by reducing the number of time series. The period of the analysis was from April 16 to May 14 in 1998. The number of time series (days) was $N = 29$. We referred to the time series by day number. For example, the current map in April 16 , 1998, is day number 1 and the current map in May 14, 1998 is day number 29.

3. Results

3.1. Ward method and dendrogram

Figure 2 shows the dendrogram of the cluster analysis by the Ward method in 1998. The numbers below the horizontal axis show the day numbers, and the vertical axis shows the distance or dissimilarity between clusters. If the current patterns are divided into 6 groups, the groups are (29, 28, 27), (2, 1, 26, 25, 24, 23), (18, 17), (20, 19 , 9, 8, 14, 7, 10), (12, 5, 6, 13, 22), and (16, 15, 21, 11, 4, 3), which are referred as W-1-6, W-2-6,..., W-6-6, respectively. The group W-i-M means the $i-$th group in the dendrogram clustered into M groups by the Ward method. the notation (29, 28, 27) means that the current maps for May 14 (day number= 29), May 13 (day number= 28) and May 12 (day number= 27) were categorized in the same group, i.e., as having similar current patterns. If the current patterns are clustered into 12 groups, the groups are (29, 28, 27), (2, 1), (26, 25, 24, 23), (18), (17), (20, 19), (9, 8, 14, 7, 10), (12, 5, 6, 13), (22), (16, 15, 21), (11, 4) and (3).

The levels at which the clusters are joined are written as $c_d(i)$, $i = 1.., N-1$, where N is the number of data points to be clustered, and $c_d(i) \geq c_d(i+1)$. Figure 3 shows a schematic illustration of $c_d(i)$. The number N is equal to the time series number, and $N = 29$ in 1998. For example, the values are $c_d(1) = 2594.9$, $c_d(2) = 1964.0$, $c_d(3) = 1760.2$ and $c_d(28) = 79.7$ from Figure 2.

If d is the threshold distance to divide the data into M groups, d must satisfy $c_d(M) < d < c_d(M-1)$. Therefore, the value of $c_d(M-1) - c_d(M)$ is a reference to assess the validity of dividing the data into M groups. The value of $c_d(M-1) - c_d(M)$ decreases as M increases for most of M. The value of $c_d(M-1) - c_d(M)$ for $M = 6$ is 133.4 in Figure 2, and it is the local maximum values of $c_d(M-1) - c_d(M)$. Therefore, the clustering into 6 groups is reasonable.

Figure 4 shows the time series of group number clustered by the Ward method. The vertical axis of Figure 4 shows the total number of groups M. The group numbers are from 1 to M for the total number M. The groups are numbered in order of the number of daily surface current maps in the group; group 1 has the most surface current maps, and group M has the fewest surface current maps.

For example, group 1 is (7, 8, 9, 10, 14), group 2 is (5, 6, 12, 13), and group 3 is (23, 24, 25, 26) for the total group number $M = 9$. Group 9 is (22) for $M = 9$. If the numbers of groups are the same for different groups, the first days are compared. The group with the earlier first day is assigned the smaller group number, so the group of (5, 6, 12, 13) is assigned group number 2.

In the case of hierarchical cluster analysis, groups are split into smaller groups as the total number of groups increases. Although the information on similarities among groups is not included in Figure 4, it is possible to show groups for different numbers of longer data, while the dendrogram cannot show groups for longer time series.

3.2. K-means method

Figure 5 also shows the time series of group number as Figure 4 but divided by the K-means method. Group formation by the K-means method is inadequate. For example, if the surface current maps are separated into 2 groups, the surface current map at day= 2 is included in only one of the groups. On the other hand, the surface current map at day= 2 is in the same group as that at day= 1 in all of the three methods for 12 clusters. This is known as the "curse of dimensionality", and it is impossible to use the K-means method for classifying the surface current maps without reducing the dimensions of the data.

3.3. SOM analysis

Figure 6 shows 2×3 ($K = 2$, $L = 3$ in Appendix A) SOM arrays, which shows weight vectors $\mathbf{m}_i$, defined in Appendix A. Figure 7 shows time series of BMUs (Best Matching Units), which are defined in Appendix A. The 6 groups are (7, 8, 9, 14, 17, 19), (3, 4, 5, 6, 10, 11, 12, 15, 16, 21, 22), (2, 18, 20), (13), (24, 25, 26, 27, 28, 29) and (1, 23) for BMU= 1, 2, 3, 4, 5, and 6, respectively.

One of the typical current patterns shows that strong southward currents flow east of the 128° E line and weak northeastward currents flow near the coast, as shown for BMU= 1 and 2 (Figure 6a, b). The difference of current patterns for BMU= 1 and BMU= 2 is that the southward flows in the eastern observation area are stronger for BMU= 2, while the northeastward currents for BMU= 2 are weaker.

The other typical current pattern shows that strong northeastward currents flow west of 128.1° E, and eastward or southeastward currents flow east of that longitude, as shown in BMU= 5 (Figure 6e). This current pattern is a clockwise eddy-like pattern. The current patterns for BMU= 3 and BMU= 4 are mixed patterns of surface current patterns for BMU= 1, BMU= 2, BMU= 5 and BMU= 6. In all of the SOM arrays, the currents near the coast are northeastward,

Figure 8 shows the time series of group numbers formed by the SOM analysis. The vertical axis of Figure 8 shows the total number of groups $M = KL$. For a given M, the natural number K is the largest natural number satisfying $K \leq L$. The numbers $K = L = 3$ for $M = 9$, $K = 3$ and $L = 4$ for $M = 12$, and $K = 1$ and $L = M$, if M is a prime number. The group numbers are from 1 to M for the total number M. As in the case of the Ward method, the groups are numbered in order of the number of daily surface current maps in the group, with group 1 having the most surface current maps, and the group M having the fewest. This group number is different from the BMU number and information about the similarities among groups is not included in the group number.

If the total number of groups is $M = 6$, the group numbers are from 1 to 6. The daily surface current maps at day numbers (3, 4, 5, 6, 10, 11, 12, 15, 16, 21, 22) constitute group 1, which corresponds to BMU= 2 (Figure 7), (7, 8, 9, 14, 17, 19) constitute group 2, which corresponds to BMU= 1, and (24, 25, 26, 27, 28, 29) are group 3, which corresponds to BMU= 5, for $M = 6$.

The separation into smaller groups by increasing the total grouping number M is not systematic. For example, the day numbers 19 and 20 are in different groups for the total grouping number $M = 2$, $K = 1$ and $L = M$. However, they are in the same group for $M = 3$, $K = 1$ and $L = M$.

3.4. Comparison between Ward method and SOM

Figure 9 shows mean current maps for each group divided into 6 groups by the Ward method. Figure 9a is the mean current in the group W-1-6, Figure 9b is the mean current in the group W-2-6, and Figure 9f is the mean current in the group W-6-6. The mean current patterns are similar to some SOM array patterns in Figure 6. For example, the current pattern of W-2-6 (Figure 9b) is similar to BMU= 5 (Figure 6e). The day numbers of W-2-6 are (2, 1, 26, 25, 24, 23), while the day numbers of BMU= 5 are (24, 25, 26, 27, 28, 29). On the other hand, all of the currents near the coast are not northeastward; see, for example, Figures 9e and f. The formation of groups by SOM is similar to that by the Ward method.

The SOM array patterns are similar to the mean currents of each group. However, the magnitudes of the vectors in the SOM array are smaller than the mean current vectors in each groups. The neighborhood function is a Gaussian function (Eq. (A.3)). Liu et al. (2006) showed that the Gaussian neighborhood function results in more smoothed patterns and smaller vectors. On the other hand, the "ep" function (equation (3) in Liu and Weisberg (2005)) gives more accurate mapping, in which case the magnitudes of the vectors are larger. The SOM array patterns group the days according to features: The current patterns near the coast are not related with the grouping, and are almost always the same in the SOM arrays.

3.5. EOF analysis

Figure 10 and Figure 11 show results of the EOF analysis. Figure 10 shows the cumulative variances and eigenvalues plotted against EOF mode number K. Mode 1 accounts for 47.9%, mode 2 accounts for 34.5%, and mode 3 accounts for 8.65% of the total variance. The first three EOF modes together account for 91.0%. The first three modes are above the line of higher modes (Figure 10b), which shows that the first three modes are significant.

The eigenvectors of the first three modes are similar to those in Hisaki (2006), in which EOF analysis was conducted on 2-hourly data. The first EOF mode is related with the change of mesoscale eddies (Hisaki (2006)).

Figure 12 shows the relationship of time coefficients for the EOF mode 1 ($b_1(t)$ in Eq. (1)) and 2 ($b_2(t)$) for $t = 1, ..., 29$ days. If we separated the groups from the position of $(b_1(t), b_2(t))$ manually, the number of groups would be 6, and the groups would be (1, 2, 23, 24, 25, 26), (3, 4, 10, 11, 15, 16, 21), (5, 6, 12, 13, 22), (7, 8, 9, 14, 19, 20), and (17, 18).

3.6. Compression by EOF

Figure 13 shows the grouping after compressing the dimensions of the data by EOF as explained in section 2.1. Figure 13 shows the time series of the group number for the cut-off EOF mode number N_E in Eq. (1). The number of divisions is 6, and the group number is assigned as Figure 4: The number of daily current maps is the largest in group 1, and the second-largest in group 2 in Figure 13.

The grouping by the K-means method is sensitive to the the cut-off EOF mode number N_E, while the groupings by the SOM and Ward methods are not so sensitive to the cut-off EOF mode number. In particular, the SOM grouping dependency on the cut-off EOF mode number is the smallest, while the Ward method groupings for $N_E = 7$ and $N_E = 8$ differ. The grouping by the Ward method for $N_E = 12$ is different from that without EOF (Figure 4). The grouping by the Ward method for $N_E \geq 4$ is same as that without EOF (Figure 8). Therefore, the reduction of the dimensionality is unnecessary in the case of SOM analysis, if we do not need to reduce the noise in a dataset.

The groups for $N_E = 2$ by SOM analysis are (3, 4, 5, 6, 10, 11, 12, 15), (1, 24, 25, 26, 27, 28, 29), (7, 8, 14, 16, 17, 19), (2, 18, 20), (9, 13, 22). and (26). The grouping are similar to the manual grouping from Figure 12 as in section 3.5. It is shown that the scatter plot of the fist- and second-mode EOF coefficients as shown in Figure 12 can be used as a reference to decide the number of divisions for the grouping in the case of SOM analysis.

4. Discussion and Conclusions

This paper compares SOM analysis with cluster analyses for classifying surface current maps. In physical oceanography, there are few studies that use cluster analysis for classification of ocean data maps.

The Ward method and K-means methods are compared as cluster analyses. EOF is also compared with SOM analysis and is used to reduce the dimension of the data. The time series is not large, so it is possible to draw a dendrogram for the comparison.

The classification by SOMs reveals the current patterns. One of the patterns is that strong southward currents flow east of the 128° E line and weak northeastward currents flow near the coast. The other current pattern is a clockwise eddy-like pattern in the HF radar observation area: The current is northeastward in the western part of the observation area, and southeastward in the eastern part of the area. Other patterns are mixtures of the two typical patterns.

Nakano et al. (1998) classified the distributions of sea surface dynamic height (SSDH) east of Okinawa Island into three patterns: The first type is an area of high SSDH near Okinawa Island, which is related to BMU= 5 (Figure 6e). The second type is an area of low SSDH near Okinawa Island, which is related to BMU= 2 (Figure 6b). The third type is an area of high SSDH around Okinawa Island and an area of low SSDH offshore, which is related to BMU= 3 (Figure 6c). The classification of SSDH is related to the classification in the present study.

The classification by the Ward method is also similar to that by SOM. The SOM patterns show only the features for the grouping. On the other hand, the averaged data in the grouped data show features that are not related with the grouping.

The dendrogram can be a reference to decide the number of divisions of the grouping not only by the Ward method but also by the SOM. Although it is impossible to draw a dendrogram for a larger data set, we can decide the number of groupings by estimating $c_d(M-1) - c_d(M)$ defined as Figure 3.

The scatter plot of the first- and second-mode EOF coefficients is related with the classification by the SOM as shown by Mau et al. (2007). In addition, the scatter plot can be a reference to decide the number of groupings. In this case, they are divided into 6 groups.

The K-means method cannot be applied to grouping without compressing the dimensions. EOF is a simple method to reduce the dimensions of the data. However, the classification by the K-means method is sensitive to the cut-off EOF mode number N_E in Eq. (1), and we cannot see the relationship between the eigenvalues (Figure 10) and the optimal cut-off EOF mode number.

The classifications by the SOM and Ward method are not sensitive to the cut-off EOF mode number. The SOM is especially insensitive to the cut-off EOF mode number N_E. The grouping by SOM from EOF time coefficients $b_k(t)$ $(k = 1, \ldots, N_E)$ was identical to the grouping without EOF for $N_E \geq 4$, while the groupings by the Ward method with and without EOF differed for $N_E = 12$.

EOF analysis is frequently applied to reduce noise in a dataset. If the dataset is noisy, and the dimensions of the data are reduced by EOF, the SOM is the best method for classification due to its insensitivity to the cut-off EOF mode number N_E. It is difficult to decide the optimal cut-off EOF mode number for both the Ward method and the K-means method.

If the dataset is not noisy, we do not need the reduction by EOF for SOM classification. However, it is not always true that the grouping by the Ward method with EOF is better than that without EOF.

The surface current data are interpolated to fill the data gaps, and the number of grids of data are the same for the entire period. It is possible to apply the SOM even when there are data gaps, without having to fill the gaps.

The SOM and Ward method are better than the K-means method for the classification. If the number of divisions into groups is a prime number such as 3, 5, 7, and 11, it may be better to use the Ward method. In other cases, the SOM is better for the grouping.

Although we could demonstrate the advantage of the SOM for this short time-series data (29 maps in total), the short time series was not sufficient to reveal the true power of the SOM method. The SOM method can deal with very long time series, and it is useful to apply the SOM to the classification of longer time-series data maps.

Acknowledgments

This study was financially supported by a Grant-in-Aid for Scientific Research (C-2) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (20540429).

The GFD DENNOU Library (available online at http://www.gfd-dennou.org /arch/dcl/) was used for drawing the figures. Comments from anonymous

reviewers were helpful in improving the manuscript.

References

Camus P., J. Fernando, J. Mendez, R. Medina, A. S. Cofino, 2010. Analysis of clustering and selection algorithms for the study of multivariate wave climate. Coastal Engineering, 58, 453-462.

Frédérique C. and F. Aires, 2009. Cluster analysis of cloud properties over the Southern European Mediterranean area in observations and a model. Mon. Wea. Rev., 137, 3161-3176.

Freeman, L. A., A. J. Miller, R. D. Norris, and J. E. Smith, 2012. Classification of remote Pacific coral reefs by physical oceanographic environment, J. Geophys. Res., 117, C02007, doi:10.1029/2011JC007099.

Harms, S. and C. D. Winant, 1998. Characteristic patterns of the circulation in the Santa Barbara Channel. J. Geophys. Res., 103, 3041-3065.

Hasegawa, T. and K. Hanawa, 2003. Heat content variability related to ENSO events in the Pacific. J. Phys. Oceanogr., 33, 407-421.

Hisaki Y., 2006: Analysis of decomposed surface currents in a limited area. IEEE J. Oceanic Eng., 31, 768-778.

Hisaki, Y., and C. Imadu 2009, The southward recirculation of the East China Sea Kuroshio west of the Okinawa Island, J. Geophys. Res., doi:10.1029/2008JC004943.

Hisaki Y., W. Fujiie, T. Tokeshi, K. Sato, and S. Fujii, 2001. Surface current variability east of Okinawa Island obtained from remotely sensed and in-situ observational data, J. Geophys. Res., 106, 31057-31073.

Kaihatu, J. M., Handler, R. A., Marmorino, G. O., Shay, L. K., 1998. Empirical orthogonal function analysis of ocean surface currents using complex and real-vector methods. J. Atmos. and Oceanic Technol., 15, 927-941.

Kohonen, T., 2001. Self-Organizing Maps. Springer-Verlag, ISBN 3-540-67921-9, New York, Berlin, Heidelberg.

Liu, Y., and R. H. Weisberg, 2005, Patterns of ocean current variability on the West Florida Shelf using the selforganizing map, J. Geophys. Res., 110, C06003, doi:10.1029/2004JC002786.

Liu, Y., R. H. Weisberg, and C. N. K. Mooers, 2006. Performance evaluation of the Self-Organizing Map for feature extraction. J. Geophys. Res., 111, C05018, doi:10.1029/2005JC003117.

Liu, Y., R. H. Weisberg, and L. K. Shay, 2007. Current patterns on the West Florida Shelf from joint Self-Organizing Map analyses of HF radar and ADCP data. J. Atmos. and Oceanic Technol., 24, 702-712.

Liu, Y., and R. H. Weisberg, 2011: A review of Self-Organizing Map applications in meteorology and oceanography. In: Self-Organizing Maps - Applications and Novel Algorithm Design, Edited by J. I. Mwasiagi, In-Tech, Rijeka, Croatia, ISBN 978-953-307-546-4, 253-272.

Marmorino, G. O., L. K. Shay, B. K. Haus, R. A. Handler, H. C. Graber and M. P. Horne, 1999. An EOF analysis of HF Doppler radar current measurements of the Chesapeake Bay buoyant outflow. Cont. Shelf Res., 19, 271-288.

Mau, J-C., D.-P. Wang, D.S. Ullman and D.L. Codiga, 2007. Characterizing Long Island Sound outflows from HF radar using self-organizing maps. Estuarine, Coastal, Shelf Sci., 74, 155-165.

Nakano, T., T. Kuragano, and Y. Liu, 1998. Variations of oceanic conditions east of the Ryukyu Islands, in Proceedings of Japan and China Joint Symposium on Cooperative Study of Subtropical Circulation System, 1-4 December 1997, Nagasaki, Japan, Seikai National Fisheries Research Institute, Nagasaki, 129-140.

Richardson, A. J., C. Risien, and F. A. Shillington, 2003. Using self-organizing maps to identify patterns in satellite imagery, Prog. Oceanogr., 59, 223-239.

You, Y., 1997. Seasonal variations of thermocline circulation and ventilation in the Indian Ocean, J. Geophys. Res., 102 , 10,391–10,422, doi:10.1029/96JC03600.

Appendix A. Algorithm and parameters of SOM

The procedure of an SOM size of $K \times L$ is as follows:

1. The weight vectors $\mathbf{m}_i$ $(i = 1, ..., KL)$, which are N_d-dimensional vectors, are initialized by generating random numbers, where N_d is the number of data per time (for example, if two-dimensional currents (u, v) are observed, $N_d = 2N_g$, where N_g is the number of observation points), and i is the unit number.

2. Set $l = 0$ (iteration number)
3. Set $k = 1$ (time series number)
4. Find the best matching unit (BMU), $i = c = c_k$ $(1 \leq c \leq KL)$ to minimize the value of $|\mathbf{x}_k - \mathbf{m}_i|$, where $\mathbf{x}_k$ is the observation data at the time series number k, and $\mathbf{x}_k$ is the N_d-dimensional vector.
5. Update $\mathbf{m}_i$

$$\mathbf{m}_i = \mathbf{m}_i + \alpha h_{ic}(\mathbf{x}_k - \mathbf{m}_i) \tag{A.1}$$

6. Update time series number: $k + 1 \longrightarrow k$
7. If $k \leq N$, repeat from 4, where N is the total time series number.
8. If $k = N + 1$, update the iteration number: $l + 1 \longrightarrow l$.
9. If $l \leq T$, repeat from 3. The maximum iteration number T is called the training length.
10. If $l = T + 1$, stop the iteration.

The time-decreasing learning rate α in Eq. (A.1) is

$$\alpha = 0.5(1 - \frac{l}{T}). \tag{A.2}$$

The neighborhood function h_{ic} is

$$h_{ic} = \exp(-\frac{d_{ci}^2}{2\sigma_l^2}), \tag{A.3}$$

where d_{ci} is the distance between map unit number c and i on the map grid (e.g., Liu and Weisberg (2005)).

The neighborhood radius σ_l decreases as a function of l, and it is estimated as

$$\sigma_l = \sigma_a + \frac{l(\sigma_b - \sigma_a)}{T}. \tag{A.4}$$

The values of σ_a and σ_b are $\sigma_a = K$ and $\sigma_b = 1$.

430 ## Figure caption

Figure 1: HF radar observation area.

Figure 2: Dendrogram by the Ward method.

Figure 3: Schematic illustration of $C_d(i)$ defined in section 3.1.

Figure 4: Group number as a function of day number and total group number M found by the Ward method.

Figure 5: Same as Figure 4 but for the K-means method.

Figure 6: 2×3 SOM arrays, i. e., weight vectors $\mathbf{m}_i$, defined in Appendix A.

Figure 7: Same as Figure 6 but for time series of BMU.

Figure 8: Time series of group number divided by SOM.

Figure 9: Mean current maps for each of 6 groups formed by the Ward method. (a) W-1-6, (b) W-2-6, (c) W-3-6, (d) W-4-6, (e) W-5-6, and (f) W-6-6.

Figure 10: (a) Cumulative variances and (b) eigenvalues plotted against EOF mode number.

Figure 11: (a) Eigenvector for the first EOF mode, (b) time coefficients for the first EOF mode, (c) same as (a) but for the second EOF mode, (d) same as (b) but for the second EOF mode, (e) same as (a) but for the third EOF mode, and (f) same as (b) but for the third EOF mode.

Figure 12: Relationship of time coefficients for the EOF mode 1 ($b_1(t)$ in Eq. (1)) and 2 ($b_2(t)$) for $t = 1, ..., 29$ days. The numbers are day numbers.

Figure 13: (a) Number of groups formed by SOM analysis as a function of time and cut-off EOF mode number. (b) Same as (a) but by the Ward method. (c) Same as (a) but by the K-means method.

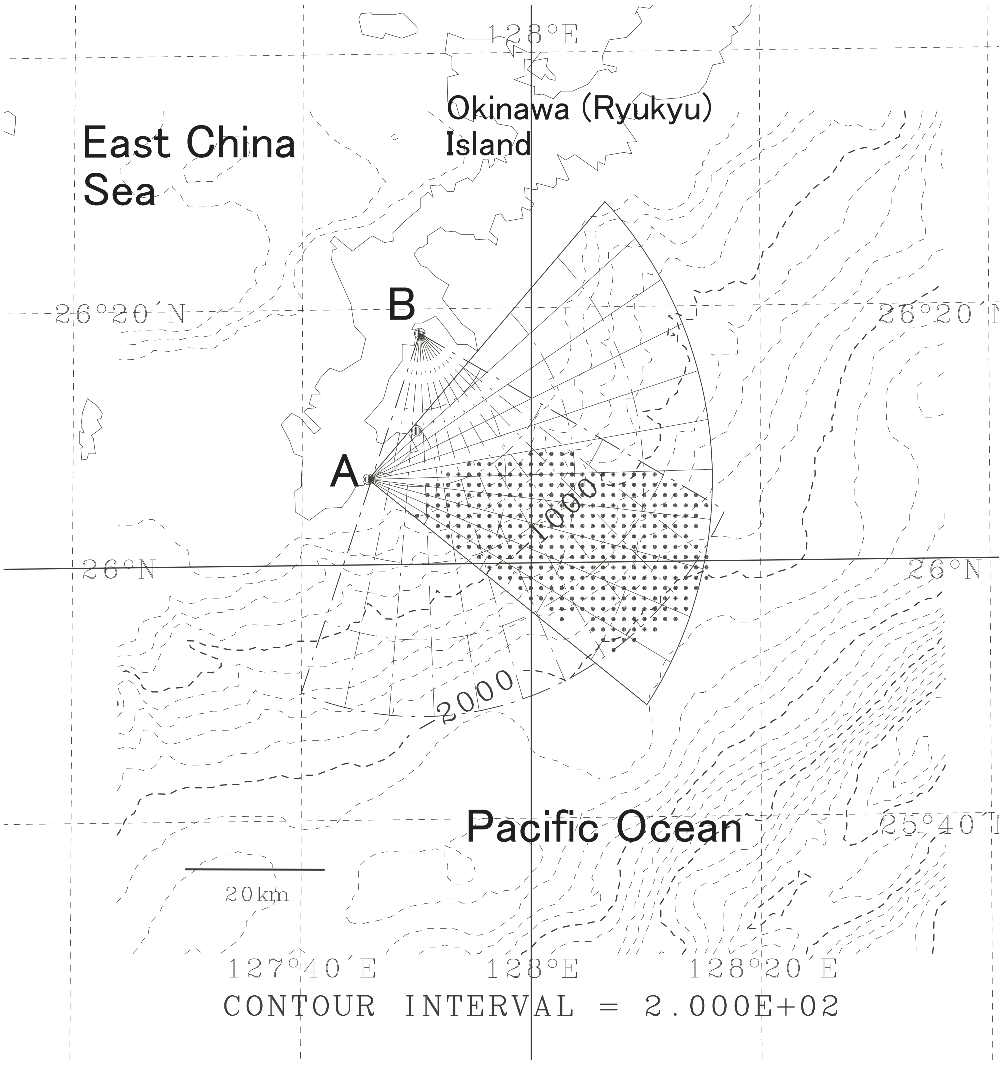

128°E
Okinawa (Ryukyu)
Island
East China
Sea
B
26°20´N
26°20´N
A
1000
26°N
26°N
2000
Pacific Ocean
25°40´N
20km
127°40´E
128°E
128°20´E
CONTOUR INTERVAL = 2.000E+02

Figure 1

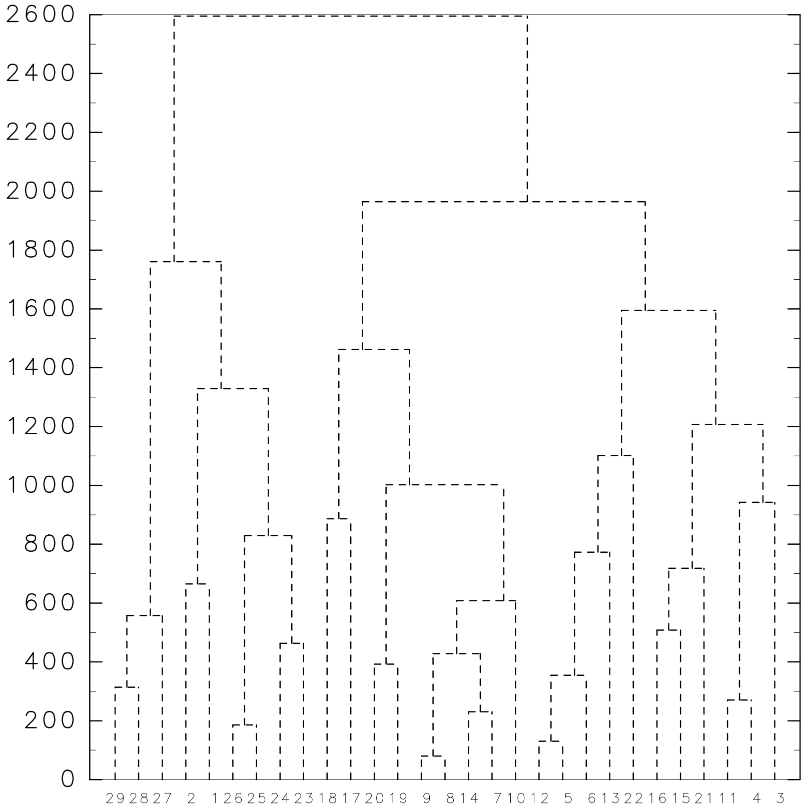

2600
2400
2200
2000
1800
1600
1400
1200
1000
800
600
400
200
0
29 28 27 2 1 26 25 24 23 18 17 20 19 9 8 14 7 10 12 5 6 13 22 16 15 21 11 4 3

Figure 2

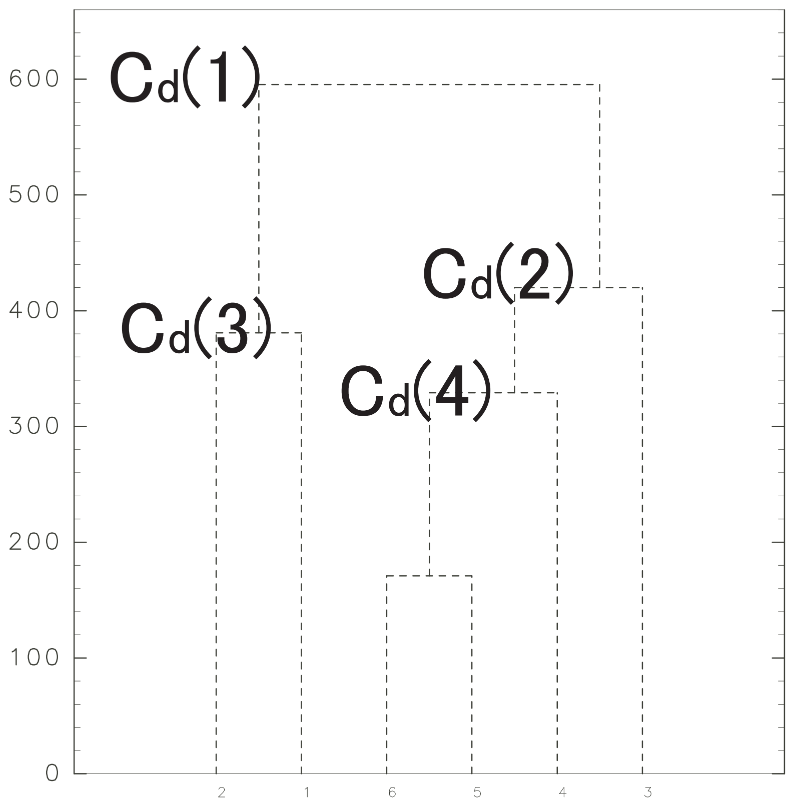


Figure 3

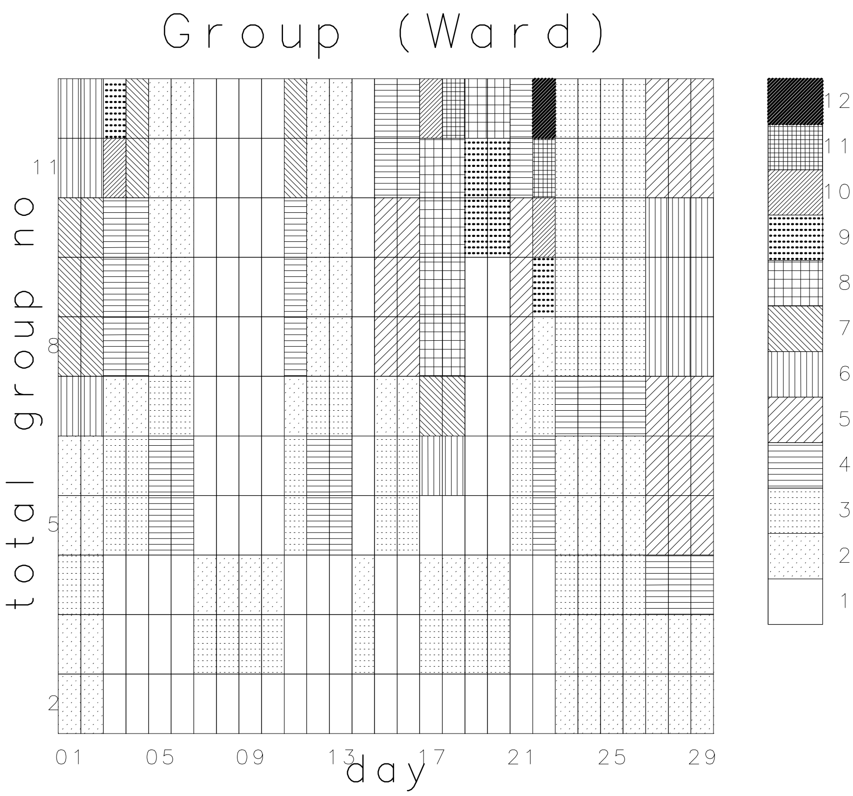

Group (Ward)
total group no
11
8
5
2
01
05
09
13
17
21
25
29
day
12
11
10
9
8
7
6
5
4
3
2
1

Figure 4

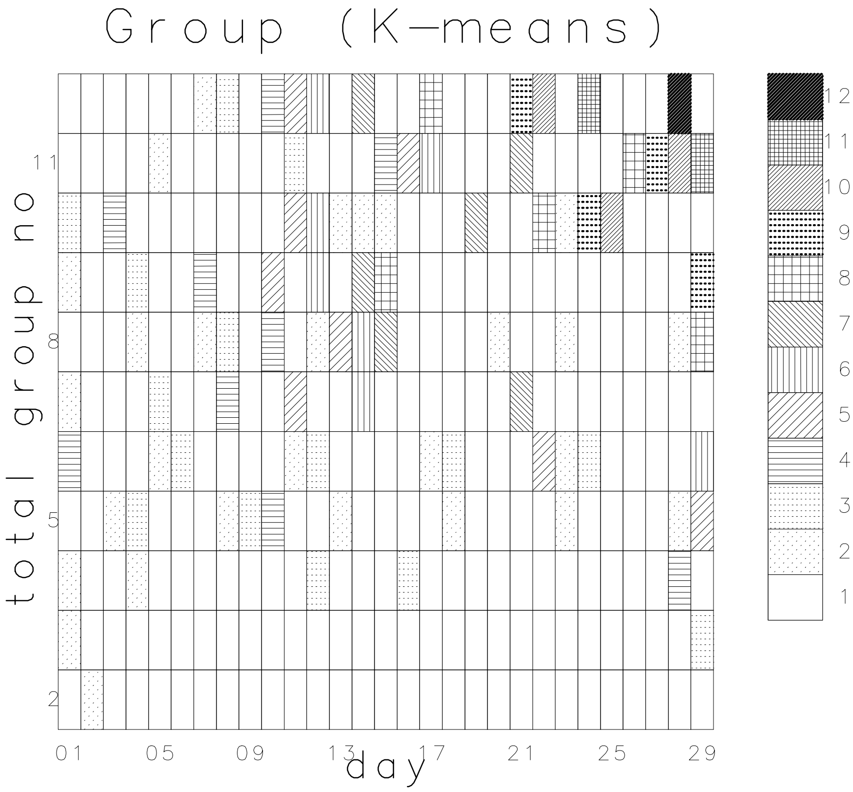

Group (K−means)
total group no
11
8
5
2
day
01
05
09
13
17
21
25
29
12
11
10
9
8
7
6
5
4
3
2
1

Figure 5

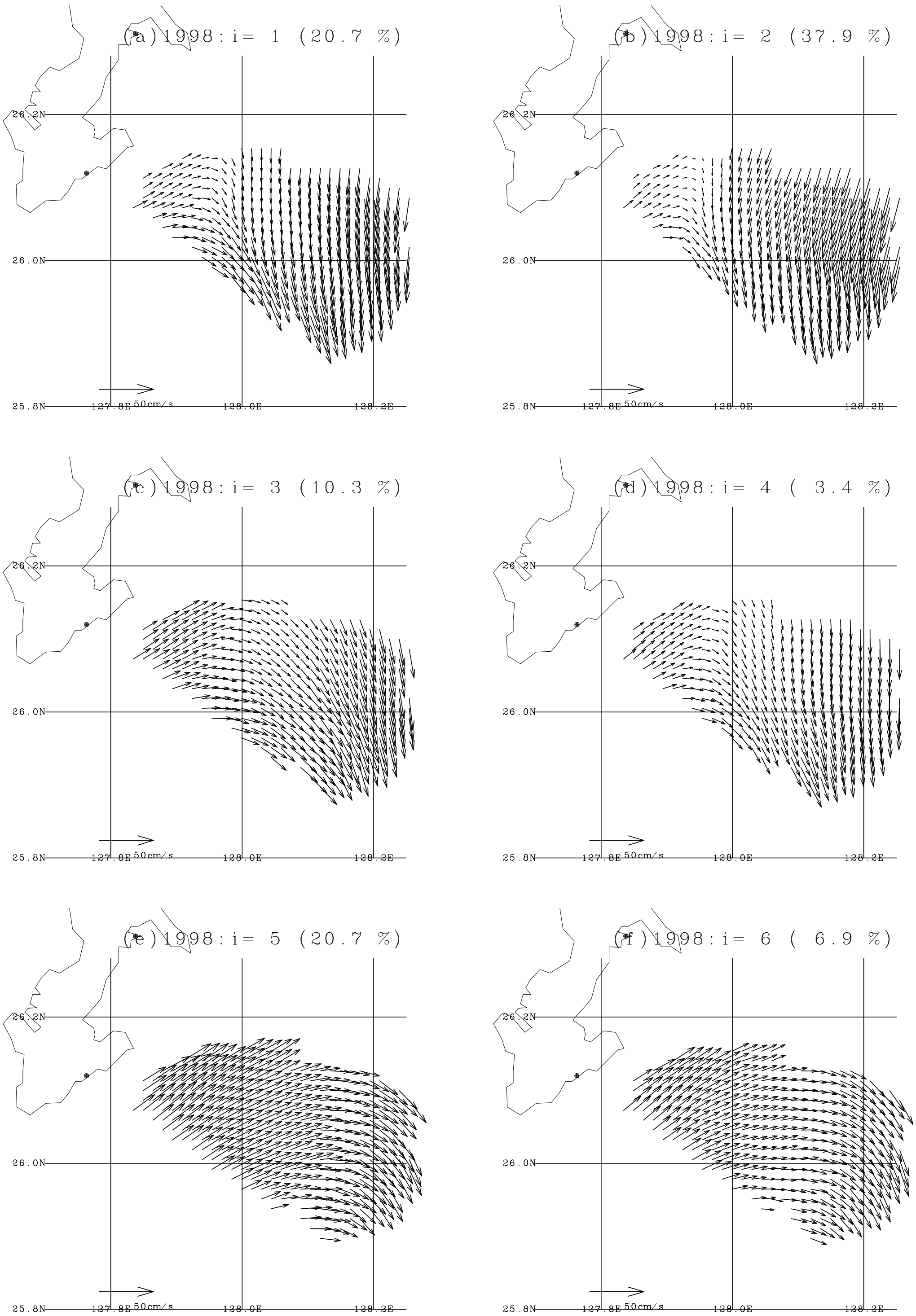

(a)1998:i= 1 (20.7 %)
(b)1998:i= 2 (37.9 %)
(c)1998:i= 3 (10.3 %)
(d)1998:i= 4 (3.4 %)
(e)1998:i= 5 (20.7 %)
(f)1998:i= 6 (6.9 %)
26.2N
26.0N
25.8N
127.8E
128.0E
128.2E
50cm/s

Figure 6

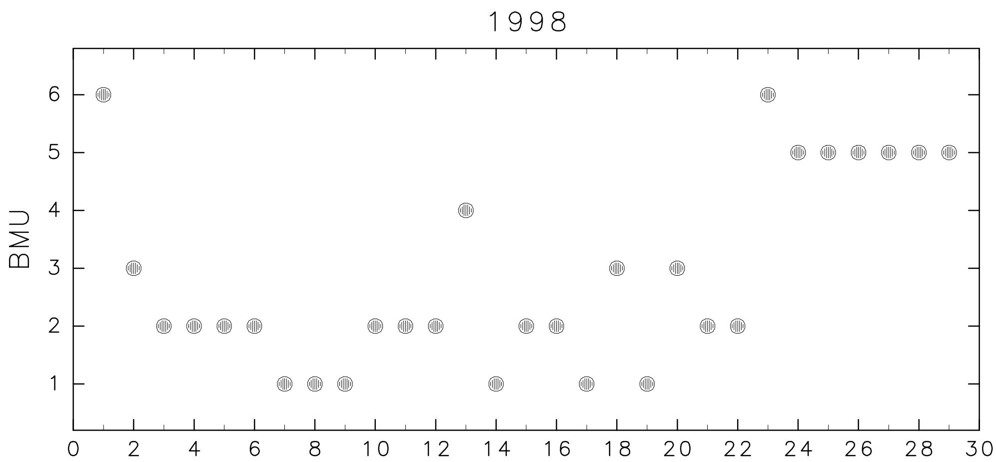
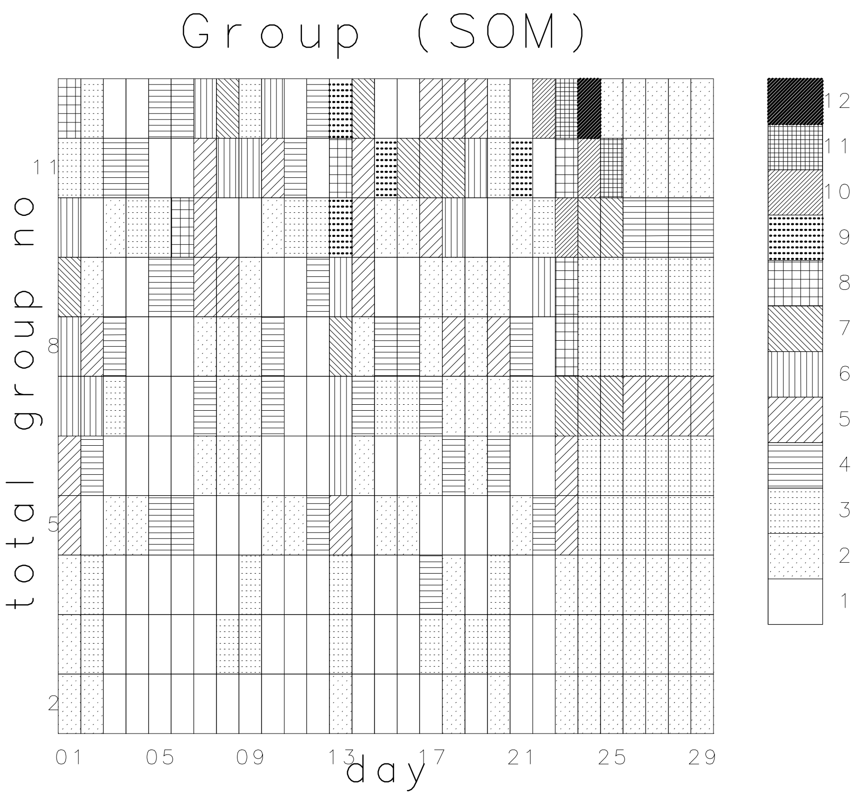


Figure 7



Group (SOM)
total group no
day
01
05
09
13
17
21
25
29
2
5
8
11
12
11
10
9
8
7
6
5
4
3
2
1

Figure 8

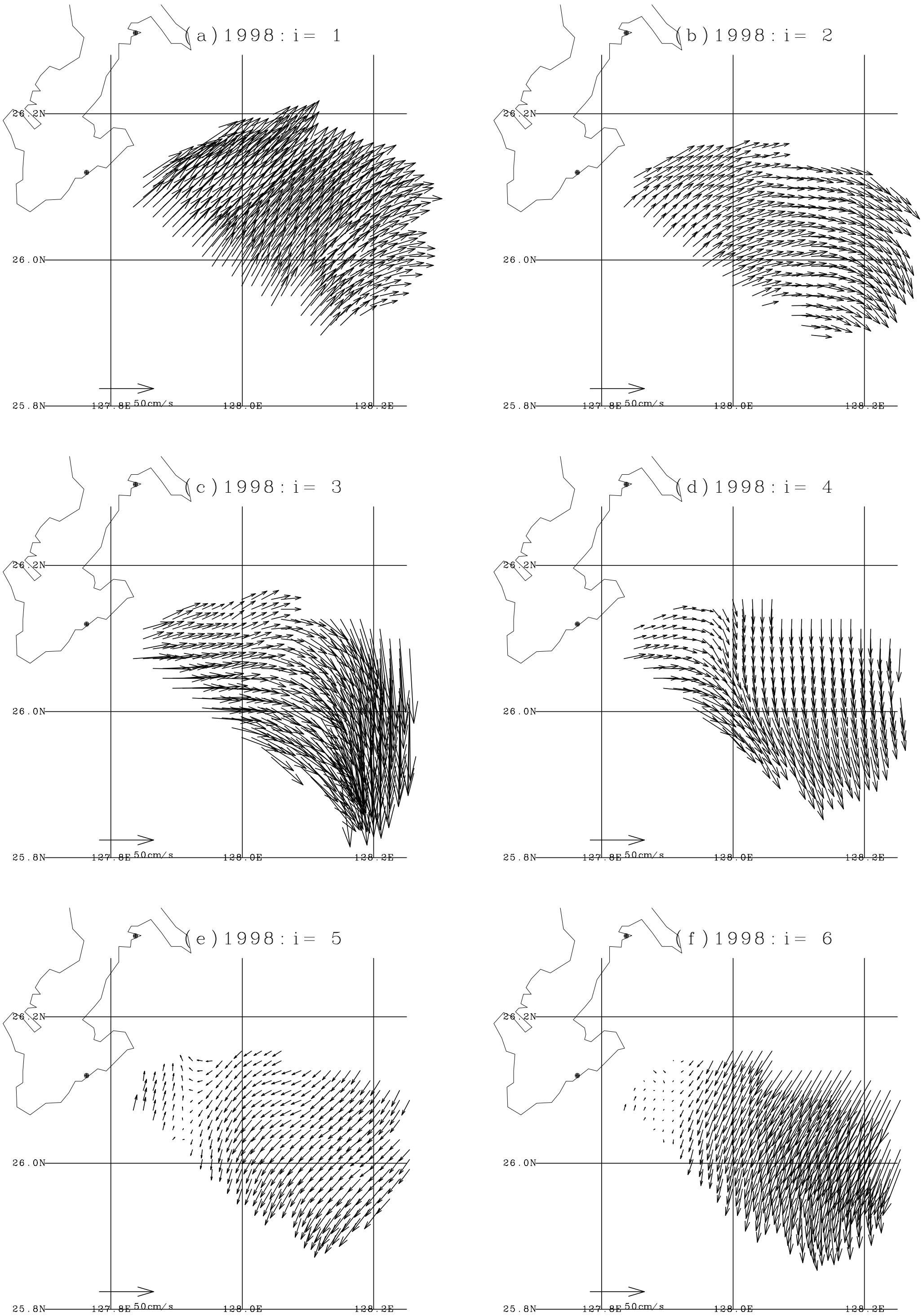

(a)1998:i= 1
(b)1998:i= 2
(c)1998:i= 3
(d)1998:i= 4
(e)1998:i= 5
(f)1998:i= 6
26.2N
26.0N
25.8N
127.8E
128.0E
128.2E
50cm/s

Figure 9

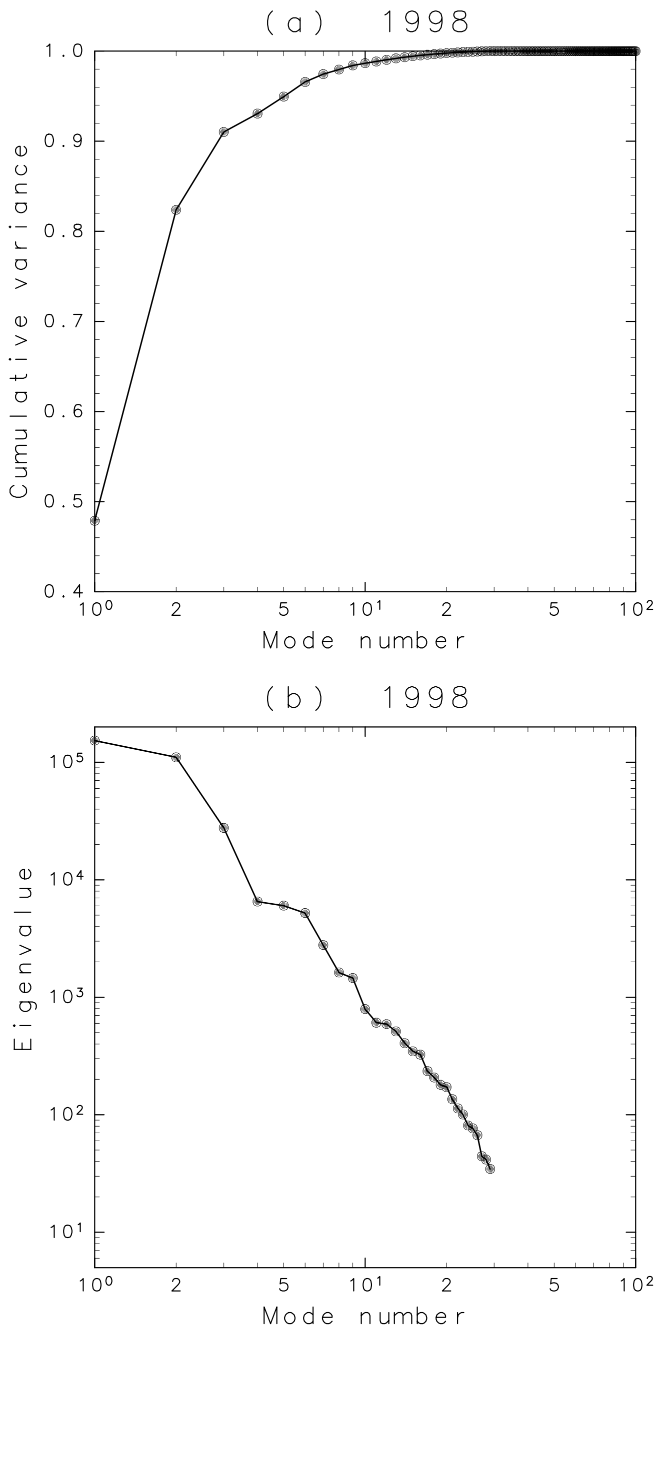

(a) 1998
Cumulative variance
1.0
0.9
0.8
0.7
0.6
0.5
0.4
10^0
2
5
10^1
2
5
10^2
Mode number
(b) 1998
Eigenvalue
10^5
10^4
10^3
10^2
10^1
10^0
2
5
10^1
2
5
10^2
Mode number

Figure 10

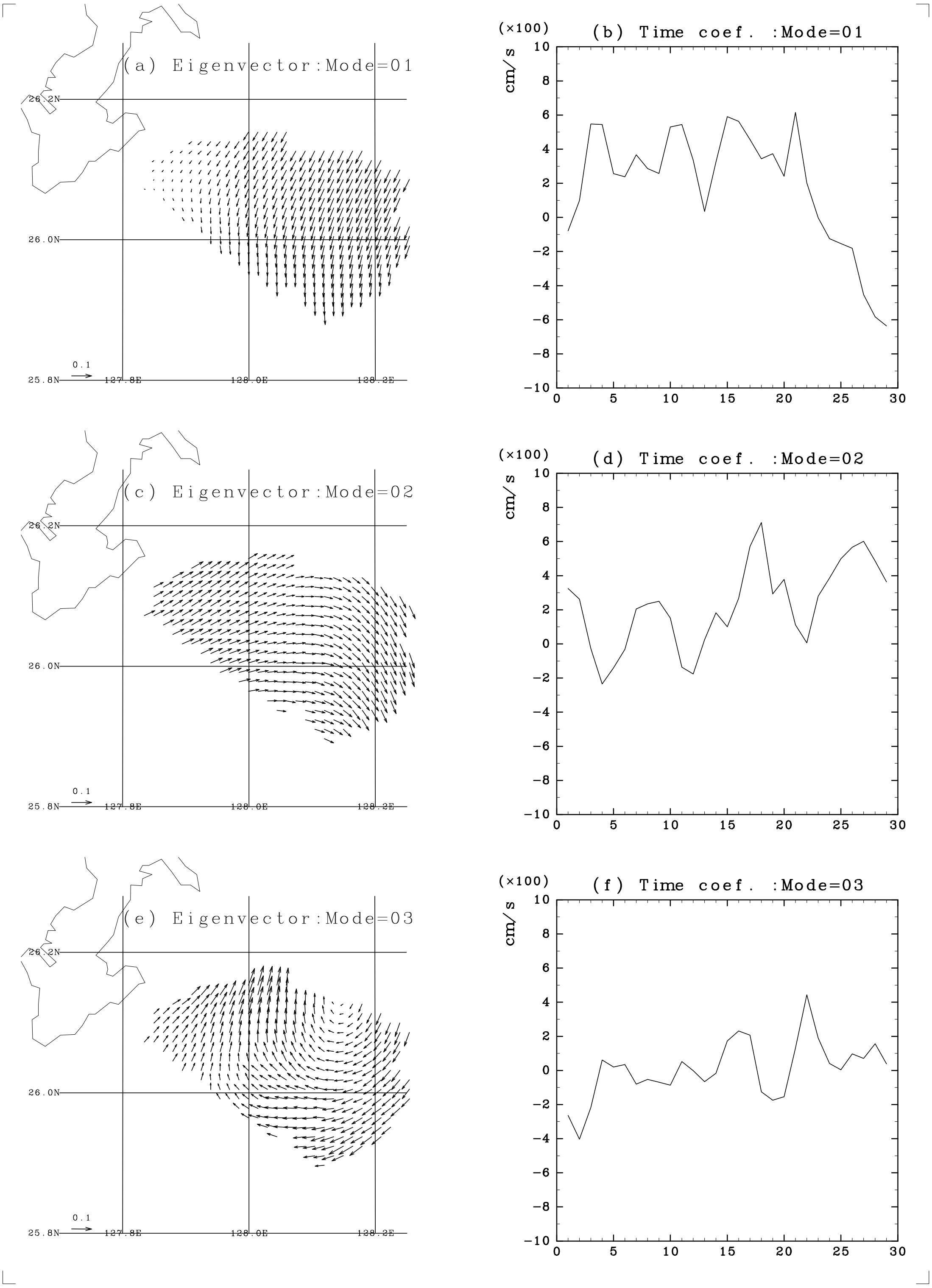

(a) Eigenvector:Mode=01
(b) Time coef. :Mode=01
(c) Eigenvector:Mode=02
(d) Time coef. :Mode=02
(e) Eigenvector:Mode=03
(f) Time coef. :Mode=03
26.2N
26.0N
25.8N
127.8E
128.0E
128.2E
0.1
(×100)
cm/s

Figure 11

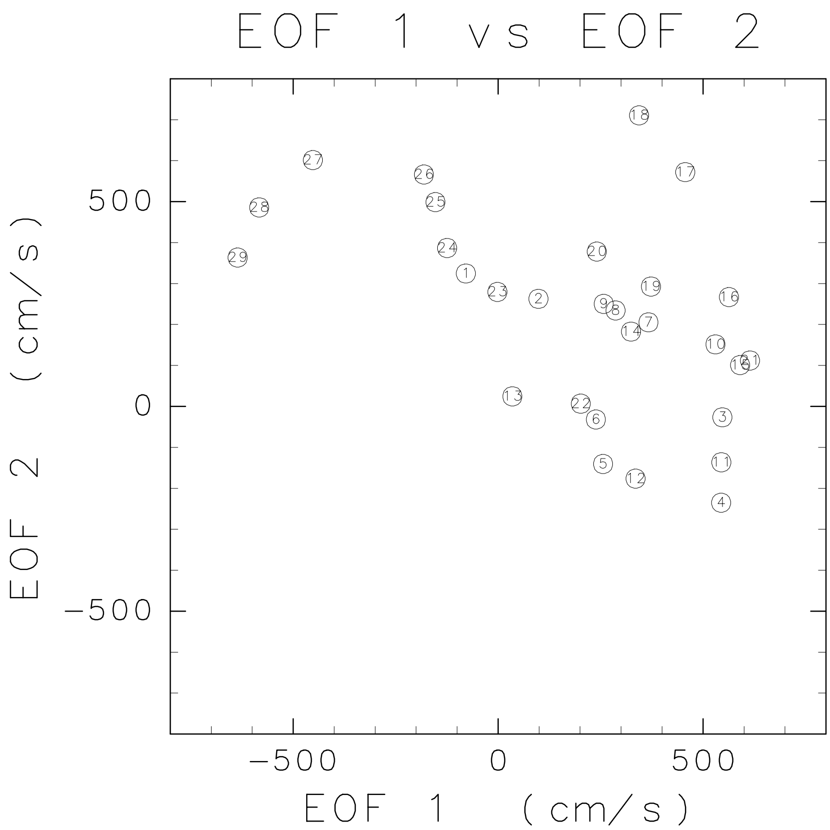

EOF 1 vs EOF 2
EOF 2 (cm/s)
EOF 1 (cm/s)
500
0
−500
−500
0
500
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29

Figure 12

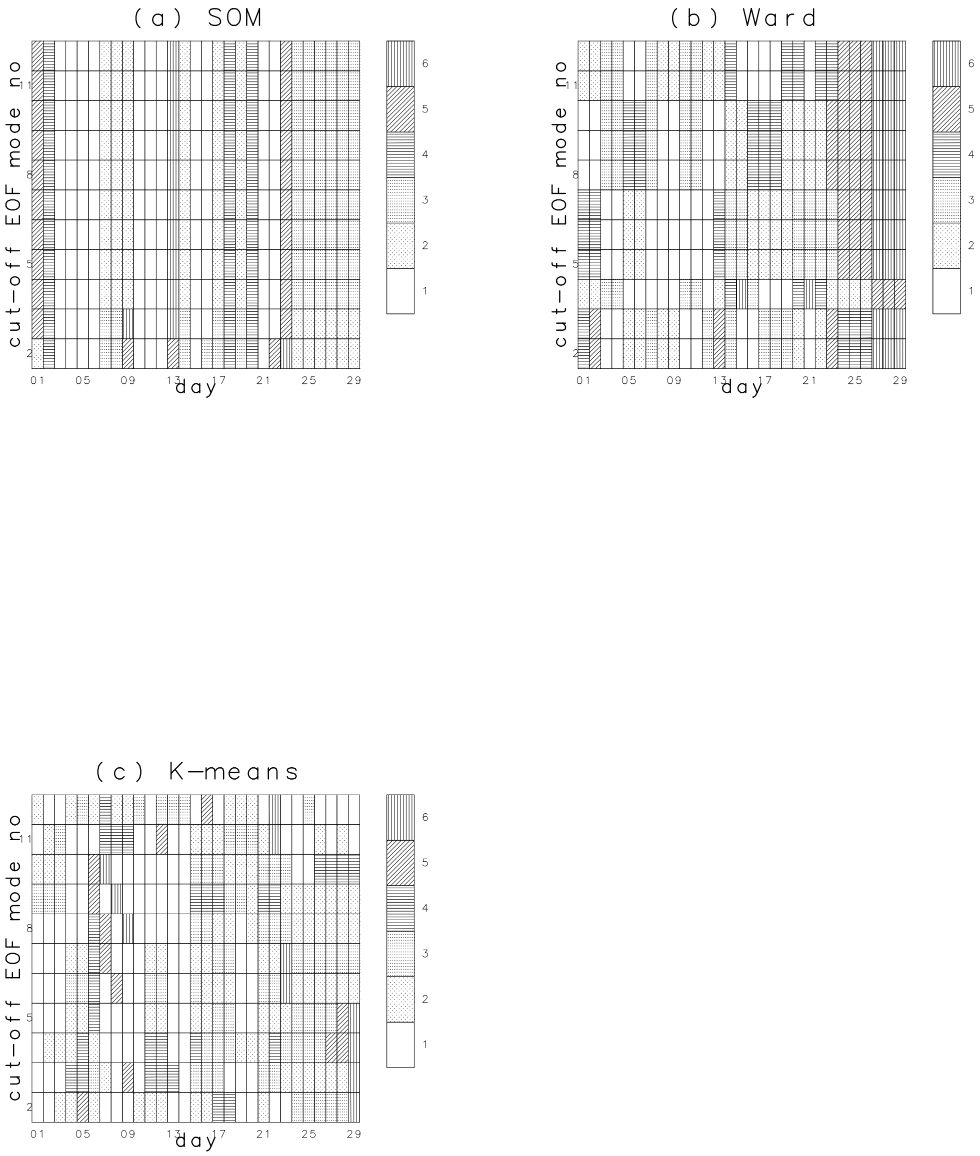


Figure 13