Multiplicative Methods for Assessment of Variation in Rainwater Chemical Content with the Time of Rain Events

Sci. Rep. Grad. Sch. Agric. & Biol. Sci. Osaka Pref. Univ. **57**:1-8 (2005) Original

Multiplicative Methods for Assessment of Variation in Rainwater Chemical Content with the Time of Rain Events

Annas SUWARDI, Takenori KANAI and Shuhei KOYAMA

(Laboratory of Environmental Information Science and Application Engineering, Graduate School of Agriculture and Biological Sciences, Osaka Prefecture University, Sakai, Osaka 599-8531, Japan)

Abstract

Periodic measurement of the concentration of chemicals dissolved in rainwater to differing time of rain events was important in detecting rainwater contamination. Multiplicative methods such as the AMMI (additive main effects and multiplicative interaction) analysis and the SOM (self-organizing map) algorithm were useful for studying patterns of the variation of the rainwater chemical content relative to the month in which the time of rain occurred. The AMMI incorporated both additive main and multiplicative interaction effects into an ANOVA (analysis of variance) of AMMI in testing the significances of the analyzed factors. The multiplicative normal distribution in term of the confidence interval 95% on the AMMI was employed to assess the stability of the chemical content in rainwater during the rainy season. Overlapping of the two first PCI (principal component of interaction) into a biplot was usually used for interpreting the components of interaction. However, in the present study the AMMI biplot not eligible to be used for displaying the interaction patterns since the second PCI was not significant according to F – test (Fisher's test) at the level of $\alpha = 0.05$. We then used the SOM to overcome this deficiency of AMMI. This method therefore offered benefits in inspecting the possible correlations of the multi-parameter of the vector components in the input data. A plane map of SOM has effectively revealed the interaction patterns as well as, visualizing the relations between the concentrations of chemicals in rainwater to the times of rain events.

Key Words: AMMI, ANOVA, chemical content, PCI, rainwater, SOM

Introduction

Rainwater pollution in the rainy season is within the scope of the BMG (Bureau of Meteorology and Geophysics) of Indonesia. Monitoring of rainwater quality by BMG is performed to measure the concentration of chemicals of rainwater and to assess the environmental impact. The measured parameters include the degree of acidity, the conductivity, and the concentrations of Cation and Anion (BMG of Indonesia, 2004). Sampling was performed periodically during the rainy season, so that the dataset is large.

This study analyzed the variation of the chemical content of rainwater as the time of rain events varies based on the datasets. One possible approach is to integrate the use of the AMMI (additive main effects and multiplicative interaction) analysis and the SOM (self-organizing map) algorithm. The AMMI model incorporates both additive main effects and multiplicative components of the two-way data structure, and can therefore account effectively for the underlying interaction patterns (Shafii and Price, 1998). It assists in the identification of interaction patterns and thereby aids interpretation (Smith et al., 1998). Ebdon et al. (1998) reported that compared with ordinary ANOVA (analysis of variance), AMMI was more effective in detecting and interpreting interaction in case to assess the potential for water conservation among turfgrass germplasm.

The biplot graphic can be used for interpreting the correlation between the chemical content in rainwater relative to the time of rain event by detecting their interaction pattern. It is done by overlapping of the two first PCI (principal component of interaction) into a biplot. However, if one or both of the two first PCI are not significant according to $F-$ test (Fisher's-test), estimation of the interaction pattern from the AMMI biplot may be biased (Suwardi et al., 2001). To overcome this problem, a SOM algorithm can be used in conjunction with AMMI.

The SOM is based on unsupervised training of a neural-network with competitive learning in which effectively used for transforming and visualizing of multi-dimensional data into a lower-dimensional map (Kohonen, 1998; Mancuso, 2001). It is also a useful tool in inspecting the possible correlations between vector components in the input data (Vesanto et al., 1998). The application of SOM was used by Aoki et al. (2002) for classification and evaluation of the water quality of irrigation ponds. They found that the effects of complicated and nonlinear environmental components could be classified using SOM.

The objectives of this study were to develop ANOVA of AMMI so as to clarify the significances of analyzed factors effects, and to assess the stability of the chemical content of rainwater using the multiplicative normal distribution on the AMMI. The SOM further detected how the distribution and the concentration level of chemicals were related with time of rain events.

Materials and Methods

The study covered areas in North Sulawesi during the rainy season in 2002. Rainwater quality was monitored over nine months at Winangun and over seven months at Samratulangi Station. Sampling was performed weekly under the supervision of BMG District IV Makassar. The concentration level of the chemicals dissolved in rainwater was then analyzed at the BMG Laboratory in Jakarta. The resulted parameters involved the pH, conductivity (Con), and concentration of Magnesium (Mg), Calcium (Ca), Ammonia (NH_4), Sodium (Na), Potassium (K), Sulfate (SO_4), Nitrate (NO_3), and Chloride (Cl). After analyzing the concentration of chemicals, the datasets were then passed to the BMG District IV Makassar, which supported the acquisition of data.

In analyzing the datasets, we first employed ANOVA for testing the main and interaction effects between both analyzed factors. We assumed the following linear model for a two-way classification with interaction:

$$y_{ij} = \mu + \alpha_i + \tau_j + (\alpha\tau)_{ij} + \varepsilon_{ij},\ (i = 1,\ldots,c;\ j = 1,\ldots,t) \quad (1)$$

y_{ij} = mean of $i-$th chemical content in $j-$th time of rain event
μ = overall mean
α_i = effect of $i-$th chemical content
τ_j = effect of $j-$th time of rain event
$(\alpha\tau)_{ij}$ = interaction of $i-$th chemical content with $j-$th time of rain event
ε_{ij} = error of $i-$th chemical content in $j-$th time of rain event

If the interaction effect on ANOVA was significantly different in the $F-$test, the AMMI analysis was then done. The principle of AMMI is to first fit additive main effects for treatment factors and then to apply PCA (principal component analysis) to the matrix of residuals that remain after the fitting of main effects (Piepho, 1994). The interaction plus mean error $(\alpha\tau)_{ij} + \varepsilon_{ij}$ from (1) can be composed into $k-$PCI axes:

$$y_{ij} = \mu + \alpha_i + \tau_j + \sum_{n}^{m} \sqrt{\lambda_n}\,\varphi_{in}\,\rho_{jn} + \delta_{ij} \quad (2)$$

where $\sqrt{\lambda_n}$ is the $n-$ th singular for PCI axis where $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m)$, φ_{in} is the $n-$th eigenvector of chemical content, ρ_{jn} is the n-th eigenvector of time of rain event, and δ_{ij} is a residual if not all axes are used. The number of axes is $m \leq \min(c-1, t-1)$. The ANOVA of AMMI is developed according to the number of PCI axes that significant in the $F-$test (Gauch, 1990).

For assessing the stability of chemical content of rainwater in terms of the times of rain, we used the multiplicative normal distribution (confidence interval 95%) by constructing the ellipse of the PCI-1 and PCI-2 scores. The nearest distance of the coordinate point of the parameters describing the chemical content to the center of coordinate ellipse is the most stable situation

(Suwardi et al., 2001). The ellipse can be specified by

$$R_1 = \sqrt{\lambda_1}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} \quad \text{and}$$

$$R_2 = \sqrt{\lambda_2}\sqrt{\frac{p(n-1)}{n(n-p)}F_{p,n-p}(\alpha)} \tag{3}$$

where R_1 and R_2 are the major and minor radii of the ellipse, p is the number of parameters, n is the number of samples, λ_1 and λ_2 are respectively the eigenvalues of PCI-1 and PCI-2, and $F_{p,n-p}(\alpha)$ is the table value from the percentile of F distribution with degree of freedom (df), and $df_1 = p$ and $df_2 = n - p$ at the level of significance α = 0.05, where the first df is p and the second is $n - p$.

We finally studied the distribution and the concentration level of chemicals in relation to the time of rain events using the SOM. This involved visualization all chemical contents via the plane map and the grid map of month of rain events. Each plane illustrated the distribution of the chemicals in each node and the colorbar represented the concentration level of chemicals. The following discuss the used shapes in the SOM algorithm.

The structure of the SOM network consists of an input layer and competitive layer that have n and N units, respectively (Yamakawa et al., 2001). Further, every unit from the input layer a parametric model vector $\mathbf{m}_i = [m_{i1}, m_{i2}, \ldots, m_{in}]^T$, $(i = 1,2,\ldots,N)$ is associated (Kohonen, 2001). Application of the SOM algorithm in this study takes the interaction matrix $\mathbf{x} = (\alpha\tau)_{ij}$ from (1) as the input data. The matrix vectors $\mathbf{x} = [x_1, x_2, \ldots, x_n]^T \in \Re^n$ from the input data are then applied to the input layer. For each set of input vectors, the best-matching node or the winner node c is found using the criterion of the minimum Euclidean distance,

$$\|\mathbf{x} - \mathbf{m}_c\| = \min_i \{\|\mathbf{x} - \mathbf{m}_i\|\} \tag{4}$$

The weights of the winner node c are then updated in accordance with the rule,

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(t)[\mathbf{x}(t) - \mathbf{m}_i(t)] \tag{5}$$

where t indicates the iteration, $\mathbf{x}(t)$ is the input supplied in random form at the iteration t, and $h_{ci}(t)$ is the neighborhood around the winning unit c (Barbalho et al., 2001). We used the SOM Toolbox of MATLAB environment in the simulation.

Results and Discussion

ANOVA of AMMI Modeling

ANOVA (Table 1) showed that the main and interaction effects of the analyzed factors from the datasets at both stations were significantly

Table 1. ANOVA of AMMI modeling.

Source	Winangun Station			Samratulangi Station		
	df	SS	MS	*df*	SS	MS
Chemicals (C)	9	5625	625.0 *	9	3187	354.1 *
Times (T)	8	237.7	29.71 *	6	107.1	17.85 *
Interactions (C x T)	72	1221	16.96 *	54	699.4	12.95 *
PCI-1	16	1197	74.80 *	14	658.1	47.01 *
PCI-2	14	12.57	0.898 ns	12	29.77	2.481 ns
PCI-3	12	4.448	0.371 ns	10	6.209	0.621 ns
PCI-4	10	4.115	0.412 ns	8	3.172	0.397 ns
PCI-5	8	2.302	0.288 ns	6	1.92	0.320 ns
PCI-6	6	0.306	0.051 ns	4	0.288	0.072 ns
PCI-7	4	0.18	0.045 ns	na	na	na
PCI-8	2	0.067	0.034 ns	na	na	na
Errors	270	2515	9.314	140	998.9	7.135
Total	359	9599		209	4992	

Degrees of freedom (df) associated with a PCI were calculated as $df = c + t - 1 - 2k$ for PCI axis k.
Sum of squares (SS) of PCI = $r\lambda_n$, r = replication, λ_n = eigenvalues ($\lambda_1 \geq \ldots \geq \lambda_m > 0$).
* Mean square (MS) significant at the level of α = 0.05, ns = non-significant, na = not available.

different in the F – test at the level of $\alpha = 0.05$. When comparing the contribution of the variance for both datasets, the chemicals have more contribution of variance with sum of squares (SS) of (5625) and (3187) than time of rain events with SS of (237.7) and (107.1), respectively. These results implicated that the chemical content of rainwater more vary during the rainy season. The significantly different of interaction also indicated the concentration of chemicals great dependent of the kind of chemical content at the differing time of rain events which means the interaction pattern needs to be analyzed further. This study therefore focused to analyze the stability of the chemical contents of rainwater and then to detect their variations relative to the time of rain events.

The first step of analyses, we developed the ANOVA of AMMI in which incorporates both additive main factors and multiplicative components of interaction. The multiplicative components of interaction were estimated by singular value decomposition (SVD) of the matrix of residuals remaining after fitting the main effects. Process of SVD for each dataset decomposed the SS of interaction into 8 PCI and 6 PCI of Winangun and Samratulangi Station as shown in Table 1.

Multiplicative normal distribution of chemical contents

We then assessed the stability of the chemical content during the rainy season using confidence interval (95%) of the multiplicative normal distribution. This method performed an ellipse for the multiplicative interaction based on their first two-PCI scores. From the analysis of both datasets we obtained two ellipse, the major and minor radii of which were (a) $r_1 = 3.91$, $r_2 = 0.45$ and (b) $r_1 = 4.47$, $r_2 = 0.95$ at the level of significance $\alpha = 0.05$

The chemical content corresponding to points outside the ellipse represented instability. Whereas coordinate points of the chemicals inside of the ellipse indicated their concentration values were stable with respect to chargers in time of rain events. As an example, Fig. 1a showed that the rainwater parameters: pH, conductivity (Con), and ammonia (NH4) varied during the rainy season of Winangun dataset. Likewise the dataset of Samratulangi (Fig. 1b) the pH, Con, and NH_4 were also at large distances from the center of the ellipse in which indicated high variance, but only NH_4 is unstable.

Analysis of interaction

The best model of AMMI was employed based on the number of PCI that significant in the F – test. ANOVA (Table 1) showed that only PCI-1 was significant at the level of $\alpha = 0.05$ so that this analysis performed AMMI-1 as the best model for both datasets. In the ANOVA of AMMI-1 (Table 2) the SS of analyzed factors were divided into two partitions such as models

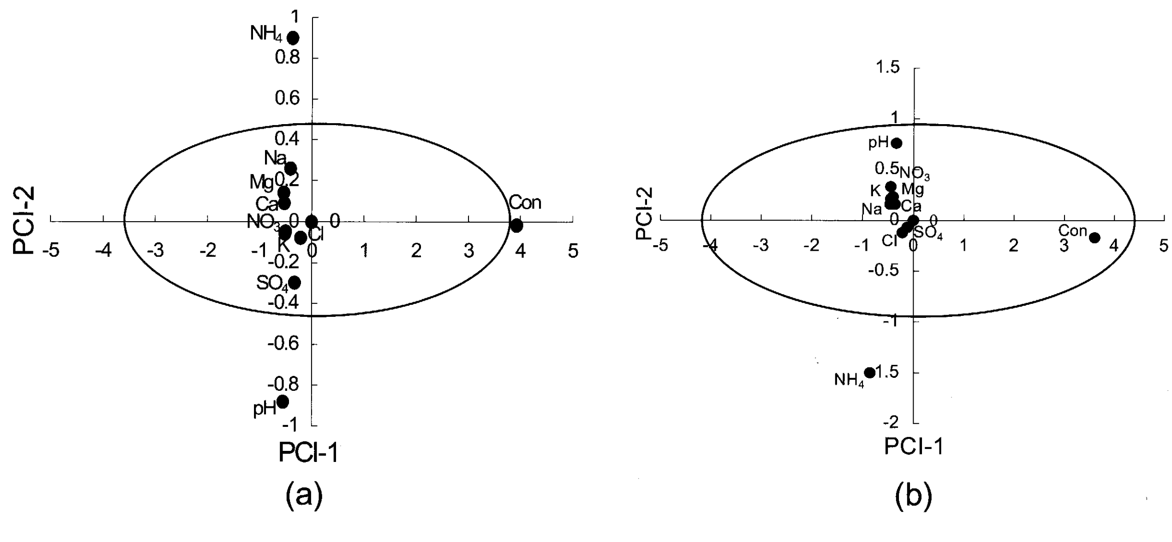

Fig. 1. The multiplicative normal distribution (confidence interval 95%) for chemical contents datasets Winangun (a) and Samratulangi (b).

Table 2. ANOVA of AMMI-1 model.

Source	Winangun Station			Samratulangi Station		
	df	SS	MS	*df*	SS	MS
Chemicals (C)	9	5625	625	9	3187	354.1
Times (T)	8	237.7	29.71	6	107.1	17.85
PCI-1	16	1197	74.8	14	658.1	47.01
Residual	56	23.99	0.428	40	41.36	1.034
Errors	270	2515	9.314	140	998.9	7.135
Total	359	9599		209	4992	

PCI-1 was significant at the level of $\alpha = 0.05$.
Residual was a combination of some PCIs and not significant at the level of $\alpha = 0.05$.

component and residual. The model component of both datasets involved the accumulation of the main, PCI-1, and error effects contributed to the total of SS with the SS of 99.75% (9575) and 99.18% (4951) with $df = 303$ and $df = 169$, respectively. These values mean that the interaction effect for each dataset can be revealed using PCI-1 in which called AMMI-1 model with only losing the information according to the component of residuals with the SS of 0.25% (23.99) and 0.82% (41.36).

The interaction pattern between the analyzed factors in the AMMI model was usually interpreted using graphic biplot. However, in the current analysis the PCI-2 was not significant, so that interpreting interactions by the AMMI biplot can not be used. Further, studies of interactions were therefore continued using the SOM to interpret the correlation pattern between the chemical content in rainwater relative to the month in which rain event occurred. Figs. 2 and 3 showed the two-dimensional map of the ten-dimensional input data for the chemical content dissolved in rainwater. The SOM training organized the maps such that neighboring neurons on the grid have similar weight vector: neighboring objects in the SOM have similar properties.

Correlation of chemicals contents to the observed month

Visualization of the SOM plane maps of the dataset of Winangun Station can be seen in Fig. 2. It showed the distribution for the pH, conductivity, and the concentration of Anions and Cations dissolved in rainwater. The concentration of Anions (Cl, SO_4, and NO_3) had a similar distribution. The concentration of Cations (Mg, Ca, and Na) had a similar distribution to the conductivity (except for NH_4 and K). Although some chemicals had similar distributions, the concentration level varies with rain time, as shown by the colorbar values (d) in the right maps.

The colorbar for pH map (Fig. 2) indicated high values from 4.74 to 5.27 (yellow to red color). Those values correlated with the grid position of rain times from May to June and October to December on the map of months. The low values of pH, from 4.21 to 4.74 (blue to yellow color) correlated with January to April. The colorbar for the Con map showed that the conductivity of rainwater is high (from 13 to 19.8 mho/cm), showing yellow to red, for February, April to May, and October to November. Low values of the conductivity (from 6.1 to 13mho/cm), showing blue to yellow occurred in January, March, June, and December. Further, the groups of Anions (Cl, SO_4, and NO_3) and Cations (Mg, Ca, Na, NH_4 and K) showed two clusters in the concentration level, having low values in January, March (somewhat high for Mg and NH_4), June and December and high values, at others times.

The SOM map of Samratulangi Station (Fig. 3) showed essentially no correlation between distributions of the concentration values of the chemicals. However, their colorbars indicated high values of pH, from 4.25 to 4.62 (yellow to red color) in May to June and December, and low values from 3.88 to 4.25 (blue to yellow color) at others times. Likewise, high values of the conductivity, from 15.1 to 21.8 mho/cm (yellow to red color), occurred only in December, and

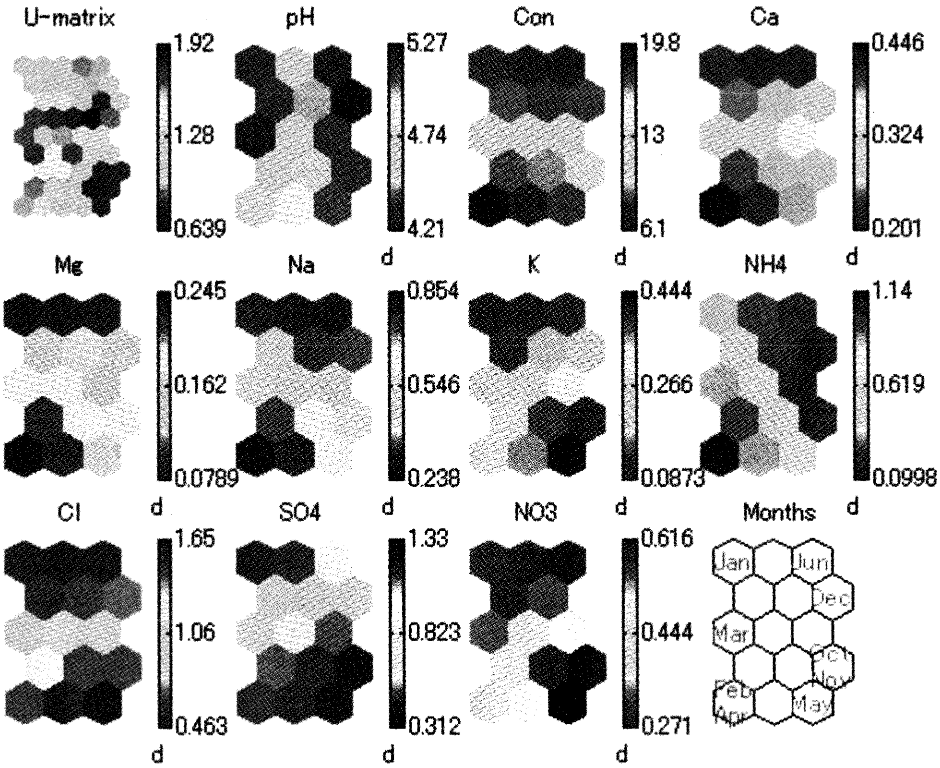

Fig. 2. Correlation maps for concentrations of ten chemicals versus the months of the rainy season at Winangun Station.

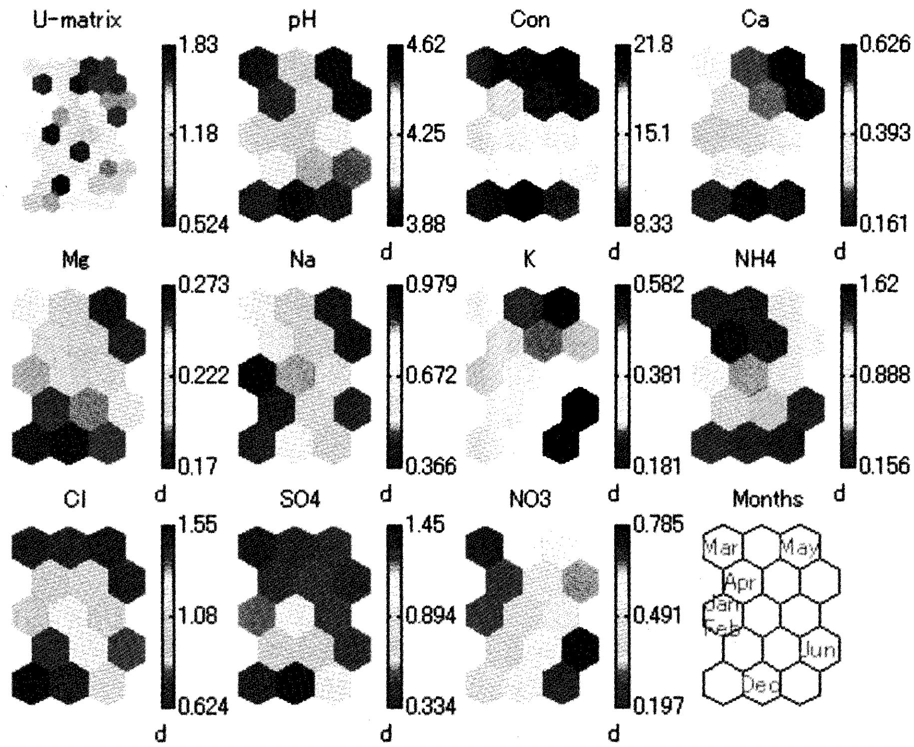

Fig. 3. Correlation maps for concentrations of ten chemicals versus the months of the rainy season at Samratulangi Station.

most times the rainwater had low conductivity, from 8.33 to 15.1 mho/cm (blue to yellow color). The distribution of Anions (Cl, SO_4, and NO_3) and Cations (Mg, Ca, Na, NH_4 and K) can be clearly classified according to their colorbar values.

The quality of rainwater from both datasets can be also predicted based on pH value (BMG of Indonesia, 2004). The datasets of Winangun and Samratulangi Station respectively had pHs in the range from 4.21 to 5.27 and from 3.88 to 4.62 (see the colorbar of pH map in Fig. 2 and 3). These values were below the range from 6.5 to 8.5 of the standard quality of rainwater (Coombes et al., 2004).

Another indicator of rainwater contamination was based on the conductivity content of rainwater. The colorbar of the Con plane from the datasets showed the conductivity of rainwater of Winangun and Samratulangi Station were in the ranges from 6.1 to 19.8 and from 8.33 to 21.8 mho/cm, respectively. In this case, no single water quality standard was detected from the data, though higher the conductivity indicated more material was dissolved in the water (Ipswatch, 2004).

Conclusion

The multiplicative normal distribution in term of the confidence interval 95% on the AMMI clearly showed that the parameter specifying the chemicals (pH, conductivity, and ammonia) more vary than all the other parameters during rain event. The significantly different of interaction of both datasets in the ANOVA of AMMI mean that the variation of concentration of the chemicals great depending on the time of rain events. It was revealed in the correlation of chemical contents relative to the observed month in which rain occurred from the SOM plane map. The colorbar of pH map showed that the pH values of rainwater samples from both datasets fall outside the range of standard quality rainwater so that distressed poor quality during the rainy season.

This study recommended the use of the SOM algorithm for detailed investigation of the interaction pattern on the AMMI model, when the AMMI biplot not eligible to be used. The visualization of the SOM component plane was therefore more effectively in inspecting the possible correlations between vector parameters of rainwater datasets.

Acknowledgments

The authors wish to acknowledge the Bureau of Meteorology and Geophysics (BMG) District IV Macassar, Indonesia for the support of the rainwater quality data.

References

Aoki, R., Minamida, T., Tachibana, J., Thanh, H. L., and Koyama, S. 2002. Classification and evaluation of water quality at irrigation ponds by using self-organization maps. *In* "6th Symposium of Water Resource," 437-442, Tokyo.

Barbalho, J.M., Neto, A.D.D., Costa, J.A.F., and Neto, M.L.A. 2001. Hierarchical SOM applied to image compression. IEEE, 442-447.

BMG of Indonesia. 2004. The Rainwater Quality in Indonesia. http://iklim.bmg.go.id/ku/kimiaairhujan.jsp.

Coombes, P.J., Kuczera, G., Jetse, D., Kalma, J.D., and Dunstan, H.R. 2004. Rainwater Quality from Roofs, Tanks and Hot Water Systems at Figtree Place. Department of Civil, Surveying and Environmental Engineering, and School of Biological and Chemical Sciences, University of Newcastle, New South Wales.

Ebdon, B., Petrovic, A.M., and Zobel, R.W. 1998. Stability of evapotranpiration rates in Kentucky bluegrass cultivars across low and high evaporative environments. Crop Sci. Soc. Amer. **38**, 135-142.

Gauch, H.G. Jr. 1990. Matmodel version 2.0. AMMI and Related Analyses for Two-way Data Matrices. Microcomputer Power, Ithaca.

Ipswatch, 2004. Understanding Water Flow and Water Quality. United States Environmental Protection Agency. http://www.ipswatch.sr.unh.edu/parmexplanation.html.

Kohonen, T. 1998. Self-organizing map. Neurocomputing **21**, 1-6.

Kohonen, T. 2001. Self-organizing maps. 3rd Ed, Springer-Verlag, Berlin Heidelberg.

Mancuso, S. 2001. Clustering of grapevine (*Vitis vinivera* L.) genotypes with Kohonen neural networks. Visits **40**, 59-63.

Piepho, H.P. 1994. Linear unbiased prediction

(BLUP) for regional yield trials: comparison to additive main effects and multiplicative interaction (AMMI) analysis. Theor. Appl. Genet. **89**, 647-654.

Shafii, B. and Price, W.J. 1998. Analysis of genotype-by-interaction using the AMMI model and stability estimates. Jour. of Agric., Biol., and Envir. Stat. **3**, 335-345.

Smith, A., Cullis, B., Gilmour, A., and Thompson, R. 1998. Multiplicative model for interaction in spatial mixed model analyses of multi-environment trial data. International Biometric Conference (IBC98), Cape Town.

Suwardi, A., Mattjik, A.A., and Susetyo, B. 2001. AMMI methods for mixed models. Jour. of Stat. and Comp. Forum, Bogor Agriculture University. **6**, 32-34.

Vesanto, J., Himberg, J., Siponen, M., and Simula, O. 1998. Enhancing SOM based data visualization. http://www.cis.hut.fi/projects/ide/publications/html/iizuka98/.

Yamakawa, T., Horio, K., and Kubota, R. 2001. A SOM association network. *In* "Advances in Self-Organising Maps," (eds. Allinson, A., Yin, H., Allinson, L., Slack, J.), 15-20, Springer, London.

(Received Dec.2, 2004; Accepted Jan.12,2005)