INTERPRETATION OF RAINFALL DATA USING ANALYSIS FACTOR METHOD

Falahah[1] and S. Suprapto[2]

[1]Faculty of Engineering, Universitas Widyatama, Bandung, Indonesia
[2]Department of Meteorology, ITB, Bandung, Indonesia

Abstract. Rainfalls are one of important climate elements. The rainfalls distribution across time and space would impact human life, not only contribute on human wealth but also restricted human life. For specific inspection area, the map of average rainfall isoline (isohyet) could not describe the phenomenon undercover the rain. Using analysis factor method, the research then conduct to analysis the rainfall for 30 years collected from 50 weather stations across Yogyakarta and Central Java try to uncover the specific phenomenon that would impact the rain characteristics. The result then plotted on factor pattern to identify the dominant factor for each region and each inspection period. Based on analysis of factor pattern, it was found 5 dominant factors which influence the rainfall. The result show that first factor which dominated the rain distribution is monsoon circulation, both Asia and Australia monsoon. This factor then reduced by mountain range in the middle of Central Java and it was identified as second factor, sea and valley breeze. The third factor is topographic related factor (mountain range area) and the other factor was related to local condition. The separation of factor which are influenced the rainfall can explain the fact why the south coastal area relatively humid than north coast. The impact of first factor on south coastal area and contribution of second factor gave rainfall accumulation on rainy season.

Keywords: rainfall, interpretation, factor analysis, spatial, central java

1 Introduction

Rain is an important natural phenomenon which can influence the human life. In fact, the rain that falls into certain area can viewed as a result of many factors, which can divided into three segments: space, time and other factor. For Indonesia area, there are many factors which influence the rainfall, such as geographic position, Monsoon, topographic, and other factor, periodic or non periodic such as ENSO, QBO and tropical disturbance event. These factors could collaborate in complex manner to contribute the rainfall.

To determine the rainfall pattern in certain area, it needs the rainfall data from observation station which spread in observed area. The rainfall data which measured by rain gauge only can show the quantity of water caused by rain, but not the factor caused the rain. These data help us to predict the quantity of rainfall, but could not help us to uncover the hidden factor beside the rainfall. By identify the dominant factors caused the rainfall, we can interpret the rainfall data better and can help us to make the better prediction in the future.

2 Theoritical Background

2.1 Monsoon System

Rain is liquid precipitation, as opposed to non-liquid kinds of precipitation such as snow, hail and sleet. Rain requires the presence of a thick layer of the atmosphere to have temperatures above the melting point of water near and above the Earth's surface. On Earth, it is the condensation of atmospheric water vapor into drops of water heavy enough to fall, often making it to the surface. Two processes, possibly acting together, can lead to air becoming saturated leading to rainfall: cooling the air or adding water vapor to the air. Virga is precipitation that begins falling to the earth but evaporates before reaching the surface; it is one of the ways air can become saturated. Precipitation forms via collision with other rain drops or ice crystals within a cloud. Rain drops range in size from oblate, pancake-like shapes for larger drops, to small spheres for smaller drops. Rain can presence by many ways such as: Frontal activity, convection, and orographic effects. In global scale, the rainfalls also influenced by global monsoon system [7].

Monsoon is traditionally defined as a seasonal reversing wind accompanied by seasonal changes in precipitation, but is now used to describe seasonal changes in atmospheric circulation and precipitation associated with the asymmetric heating of land and sea[7]. Usually, the term monsoon is used to refer to the rainy phase of a seasonally-changing pattern, although technically there is also a dry phase[15].

The major monsoon systems of the world consist of the West African and Asia-Australian monsoons[16]. The Asian monsoons may be classified into a few sub-systems, such as the South Asian Monsoon which affects the Indian subcontinent and surrounding regions, and the East Asian Monsoon which affects southern China, Korea and parts of Japan. The East Asian monsoon affects large parts of Indo-China, Philippines, China, Korea and Japan. It is characterized by a warm, rainy summer monsoon and a cold, dry winter monsoon. The rain occurs in a concentrated belt that stretches east-west except in East China where it is tilted east-northeast over Korea and Japan. The seasonal rain is known as *Meiyu* in China, *Changma* in Korea, and *Bai-u* in Japan, with the latter two resembling frontal rain.

The onset of the summer monsoon is marked by a period of premonsoonal rain over South China and Taiwan in early May. From May through August, the summer monsoon shifts through a series of dry and rainy phases as the rain belt moves northward, beginning over Indochina and the South China Sea (May), to the Yangtze River Basin and Japan (June) and finally to North China and Korea (July). When the monsoon ends in August, the rain belt moves back to South China.

Australia monsoon also known as the Indo-Australian Monsoon. The rainy season occurs from September to February and it is a major source of energy for the Hadley circulation during boreal winter. The *Maritime Continent Monsoon* and the

Australian Monsoon may be considered to be the same system, the Indo-Australian Monsoon. It is associated with the development of the Siberian High and the movement of the heating maxima from the Northern Hemisphere to the Southern Hemisphere. North-easterly winds flow down Southeast Asia, are turned north-westerly/westerly by Borneo topography towards Australia. The onset of the monsoon over the Maritime Continent tends to follow the heating maxima down Vietnam and the Malay Peninsula (September), to Sumatra, Borneo and the Philippines (October), to Java, Sulawesi (November), Irian Jaya and Northern Australia (December, January). However, the monsoon is not a simple response to heating but a more complex interaction topography, wind and sea, as demonstrated by its abrupt rather than gradual withdrawal from the region. The Australian monsoon or rainy season occurs in the austral summer when the monsoon trough develops over Northern Australia. Over three-quarters of annual rainfall in Northern Australia fall during this time. Figure 1 and figure 2 show the main wind pattern during January and July Monsoon.

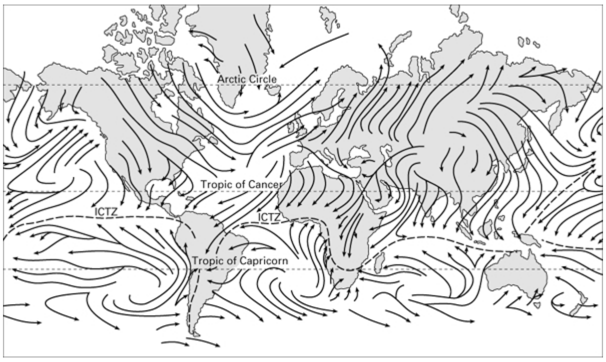


Figure 1. Typical monsoon pattern on January [17]

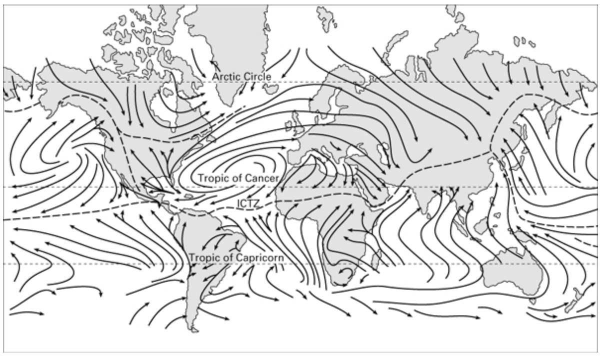


Figure 2. Typical monsoon pattern on July [17]

2.2 Rainfall in the Indonesian Monsoon

Rainfall in the Indonesian monsoon can be trigger by some causes such as monsoon which is reinforced by sea and valley breezes, orographics effect and convectional rainfall from local factor[12].

Monthly rainfall distribution can characterized by three type of pattern, namely: 1. Monsoonal type, the distribution of monthly rainfall is influenced by monsoon Asia. 2. Equatorial type, the distribution of monthly rainfall is affected by equinoxes, so the distribution show double maxima, 3> local type, rainfall pattern is influenced by local condition and monthly rainfall distribution is the opposite of monsoonal type. Figure 3 show the distribution of montly rainfall in Semarang, Pontianak and Ambon as monsoonal, equatorial and local types respectively.

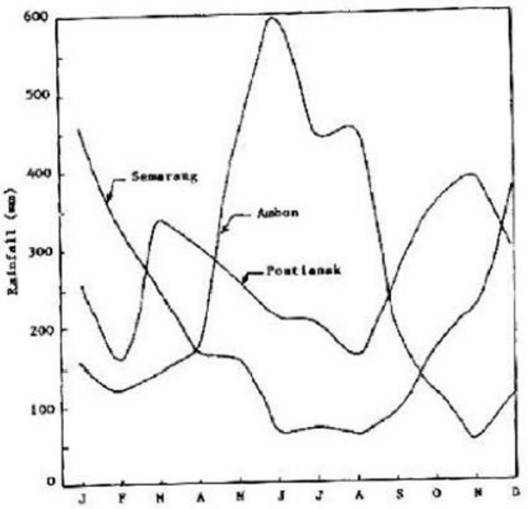


Figure 3. Monthly rainfall distribution for Semarang, Pontianak and Ambon [12]

2.3 Factor Analysis

Factor analysis is a collection of methods used to examine how underlying constructs influence the responses on a number of measured variables[8]. There are basically two types of factor analysis: exploratory and confirmatory. Exploratory factor analysis (EFA) attempts to discover the nature of the constructs influencing a set of responses. Confirmatory factor analysis (CFA) tests whether a specified set of constructs is influencing responses in a predicted way. Both types of factor analyses are based on the *Common Factor Model*.

Factor analyses are performed by examining the pattern of correlations (or covariances) between the observed measures. Measures that are highly correlated (either positively or negatively) are likely influenced by the same factors, while those that are relatively uncorrelated are likely influenced by different factors.

Factor analysis searches for such joint variations in response to unobserved latent variables. The observed variables are modeled as linear combinations of the potential factors, plus "error" terms. The information gained about the interdependencies between observed variables can be used later to reduce the set of variables in a dataset[13].

Types of factoring [14]

- Principal component analysis (PCA): The most common form of factor analysis, PCA seeks a linear combination of variables such that the maximum variance is extracted from the variables. It then removes this variance and seeks a second linear combination which explains the maximum proportion of the remaining variance, and so on. This is called the principal axis method and results in orthogonal (uncorrelated) factors.
- Canonical factor analysis, also called Rao's canonical factoring, is a different method of computing the same model as PCA, which uses the principal axis method. CFA seeks factors which have the highest canonical correlation with the observed variables. CFA is unaffected by arbitrary rescaling of the data.
- Common factor analysis, also called principal factor analysis (PFA) or principal axis factoring (PAF), seeks the least number of factors which can account for the common variance (correlation) of a set of variables.
- Image factoring: based on the correlation matrix of predicted variables rather than actual variables, where each variable is predicted from the others using multiple regression.
- Alpha factoring: based on maximizing the reliability of factors, assuming variables are randomly sampled from a universe of variables. All other methods assume cases to be sampled and variables fixed.

The type of factor analysis using on this research is CFA. There are six basic steps to performing an CFA:

1. Define the factor model. This involves selecting the number of factors, and defining the nature of the loadings between the factors and the measures. These loadings can be fixed at zero, fixed at another constant value, allowed to vary freely, or be allowed to vary under specified constraints (such as being equal to another loading in the model).
2. Collect measurements.
3. Obtain the correlation matrix.
4. Fit the model to the data. The most common model fitting procedure is Maximum likelihood estimation, which should probably be used unless the measures seriously lack multivariate normality.
5. Evaluate model adequacy. When the factor model is fit to the data, the factor loadings are chosen to minimize the discrepancy between the correlation matrix implied by the model and the actual observed matrix. The amount of discrepancy after the best parameters are chosen can be used as a measure of how consistent the model is with the data. The most commonly used test of model adequacy is the Â2 goodness-of-fit test. The null hypothesis for this test is that the model adequately accounts for the data, while the alternative is that there is a significant amount of discrepancy. Unfortunately, this test is highly sensitive to the size of data sample, such that tests involving large samples will generally lead to a rejection of the null hypothesis, even when the factor model is appropriate.

3 Research Procedure

Inspection Area for this research include central java province and Yogyakarta which are limited by geographic boundary 109.5E and 111E and Java Sea in North and Hindia Ocean in South. This area was chosen based on data availability and distribution of weather stations which are relatively consistent and the variation of topography.

The data collected from 50 weather stations which are spread across inspection area. The inspection period is for 30 years starting from 1951 until 1980, as requested by climatology data analysis that needs at least 30 years data.

The data then was divided into three periods which are wet period, dry period and yearly period. Wet period start from November until January, and dry period start from May, June, and July. For each period first it will calculate the monthly average for each weather station.

After calculate monthly average, the data become input for factor analysis. Analysis then conduct on spatial and temporal dimension to explore the pattern of rainfall across area and time.

Preliminary input for factor analysis is matrix of observation data for each period. The data processing with factor analysis will done by 2 dimension, spatial to observe the rainfall pattern within the region and temporal to identify rainfall pattern across the time.

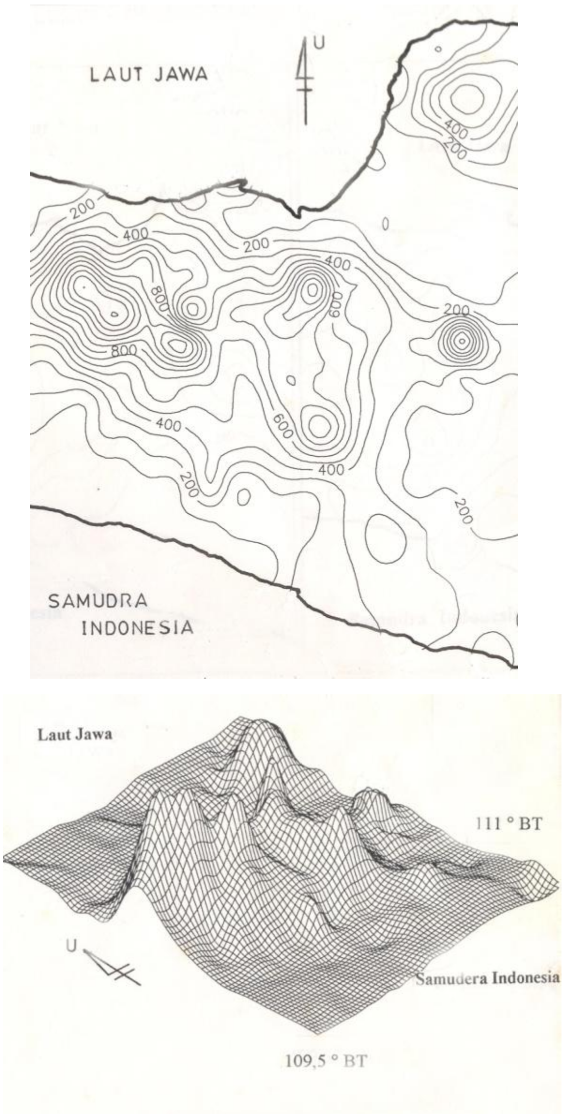


Figure 4. The countour and topography of observed area.

Spatial analysis use the stations as variable and years as unit of observation. It produces the 50x30 matrix size as the input. Some step to calculate the data are:

1. Calculate the average, so it would produce the vector of sample average (m)

The format of observation data matrix is:

	1st year	*2nd year*	***.....***	*30th year*
Obs1	$RF_{1\,1}$	$RF_{1\,2}$	.....	$RF_{1\,30}$
Obs2	$RF_{2\,1}$	$RF_{2\,2}$	.....	$RF_{2\,30}$
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
Obs50		$RF_{50\,2}$	.....	$RF_{50\,30}$

$RF_{i\,j}$ = rainfall at observer i and at year j

If X_i is observer station and n as total year of observation, the next steps are:

b. Calculate the sample average

– (1)

i = 1…50, j = 1…30

Then we will have the vector of sample average

(2)

c. Identify variant-covariant matrix (V)

(3)

with

–

i,j = 1,2,3,....50

Diagonal element is a sample variance and the others element is a sample covariance

d. Identify correlation matrix (R)

To identify the correlation matrix (R), firstly it needs to calculate the correlation coefficient for each variable, r_{ij} (correlation coefficient between variable i to variable j), from the equation:

(4)

Where is :

V_{ij} = covariance i to j

V_{ii} = variance i, i = 1, ..., 50

V_{jj} = variance j, j = 1, ..., 50

Correlation matrix is a matrix which the element consists of the correlation coefficient value from observed variable, which is :

(5)

$r_{11} = r_{22} = r_{33} = \ldots = r_{5050} = 1$

With r_{ij} = coefficient correlation between variable i and j

e. Calculate characteristic value of covariant matrix and correlation matrix(λ) which can calculate from matrix determinant (A – λI) from the equation:

$$| A - \lambda I | = 0 \qquad (6)$$

A = correlation / covariance matrix

I = Identity matrix

λ = eigen value

now we can calculate $\lambda_1, \lambda_2, \ldots, \lambda_{50}$, where is $\lambda_1 > \lambda_2 > \ldots > \lambda_{50}$

f. Calculate proportion (λ_j)

$$\text{——} \qquad (7)$$

$j = 1..\ 50$

g. Calculate characteristic vector, covariance / correlation matrix from the equation bellow :

$$AK = \lambda K \qquad (8)$$

A = correlation / covariance matrix,

K = characteristic vector

λ = characteristic value

Now we have 50 characteristic vector for each eigen value

h. Define loading factor for correlation matrix, with the equation :

$$a_{ij} = K_j \sqrt{\lambda_i} \qquad (9)$$

i. Calculate communality

$$(10)$$

j. Rotate the factor

$$– \quad \text{—} \quad \text{—} \quad \text{—} \qquad (6)$$

The similar procedure for factor analysis data processing as describe above also implement for temporal analysis, but, the inputs are year as variable and weather stations as observation units, and then the data matrix would pass the same step as spatial analysis.

The final result for each factor then plotted on isoline map into specific inspection area. Analysing of pattern on isoline map tend to uncover the natural phenomena that can give significant contribution on rainfall data.

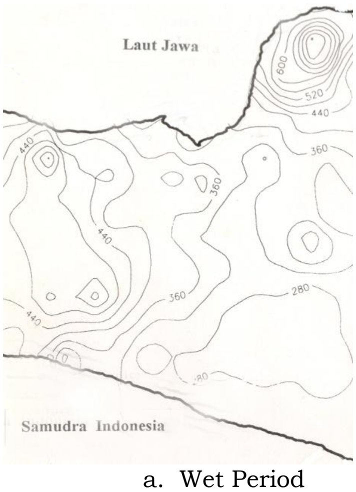
4 Data Analysis

The first step to analyze and interpret the data from factor analysis result, we identify the top 5 factor value as dominant factor. To identify the pattern we divide the data into 3 observation period, wet period, dry period and yearly period (average over 1 year). Each factor then plotted into the map to determine the pattern and to see the coverage of each factor into observed area.

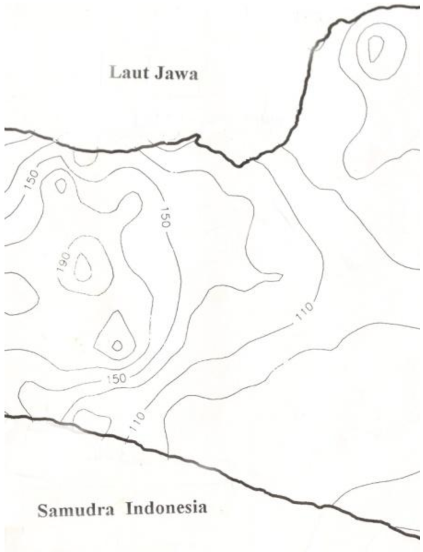
4.1 Analysis of Average Rainfall Data

Figure 5 show the isohyet of average rainfall for each period. The distribution of average rainfall for dry period as depict by isohyet does not show the significant change on zonal or meridional path. The rainfall that collected from stations in the North area had shown the similar value as in the South area.

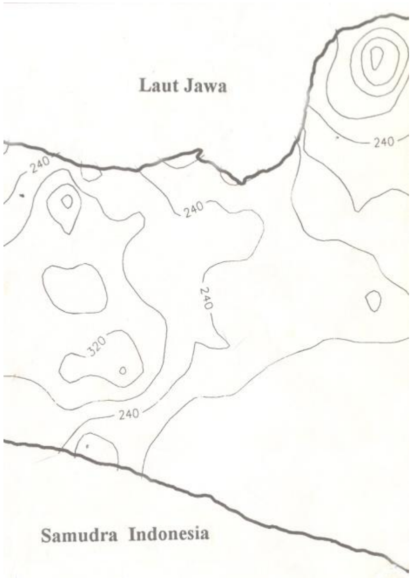
The isohyet map for each period (wet, dry and yearly period) show the similar pattern in the North and South coastal area. In the central area of observation, around Jepara, it has relatively higher value for each observation period.



a. Wet Period



b. Dry period



c. Yearly Period

Figure 5. Isohyet of Average rainfall

4.2 Factor Analysis

Analysing and interpretating the rainfall data using factor analysis will done by this steps:

a. Choose the 5 top eigen value for each period, to determine the quality of data representation on showing characteristic variance of rainfall data for each period.
b. Determine the dominant factor for each period and evaluate the factor distribution on observed area.
c. Spatial Analysis:

   The top 5 eigen value for each observation period then plotted into a isoline map. Distribution of dominant factor can identify by create the map which consists the plot of biggest factor for each observed stations, and then calculated the percentage for each area covered by each factor.
d. Temporal Analysis:

   The value of each factor as result of data processing then plotted into a graphic to show the yearly fluctuation. The spectrum of power for each factor plotted into graphics to identify the periodic of tropical disturbance
e. Identify the potential natural caused that might related with each factor.

Table Ia, Ib, and Ic show the top 5 eigen value of correlation matrix for each period (wet period, dry period and yearly average), both temporal and spatial

Table Ia. Top 5 Eigen Value for wet period

No.	Spatial		Temporal	
	Eigen Value	Proportion (%)	Eigen value	Proportion (%)
1	14.7	29.4	16.9	58.4
2	7.1	14.1	2.8	9.8
3	3.9	7.8	1.7	5.7
4	2.9	5.7	1.1	3.9
5	2.7	5.4	1.0	3.5
Total cummulative		62.4		81.26

Table Ib. Top 5 Eigen Value for dry period

No.	Spatial		Temporal	
	Eigen Value	Proportion (%)	Eigen value	Proportion (%)
1	33.3	66.5	13.8	46.1
2	3.2	6.5	3.9	13.2
3	1.9	3.9	1.7	5.8
4	1.7	3.4	1.3	4.4
5	1.4	2.8	1.1	3.7
Total cummulative		83.1		73.15

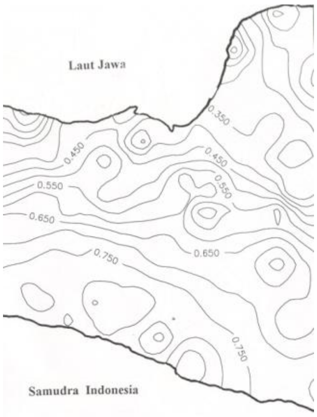
Table Ic. Top 5 Eigen Value for yearly period

No.	Spatial		Temporal	
	Eigen Value	Proportion (%)	Eigen value	Proportion (%)
1	22.9	45.9	19.3	64.8
2	5.2	10.4	2.7	8.9
3	3.2	6.4	1.6	5.3
4	2.9	5.8	1.2	4.1
5	1.9	3.7	0.8	2.6
Total cummulative		72.2		85.70

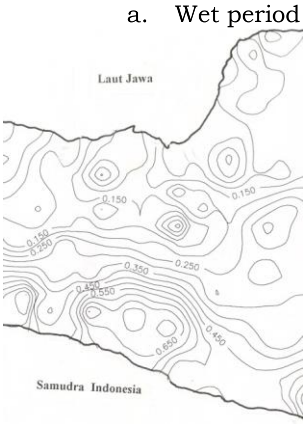
Total cummulative of top 5 highest eigen value for each observation period has shown the average over 60%, which can concluded that the rainfall data using on this research can represent the characteristic of rainfall data variance on each period.

The difference of total cummullative value for spatial and temporal can be explain as the effect of choosing variabel for calculationg the correlation matrix, and on this research, it was not deeply analyzed.

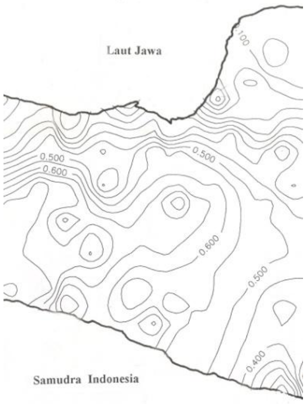
Spatial Analysis of isoline map shows the significant difference pattern for first factor (factor-1).



a. Wet period



b. dry period



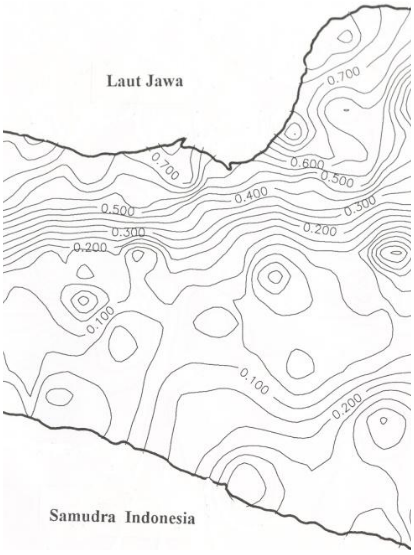
c. Yearly period

Figure 6. Isoline of Factor-1

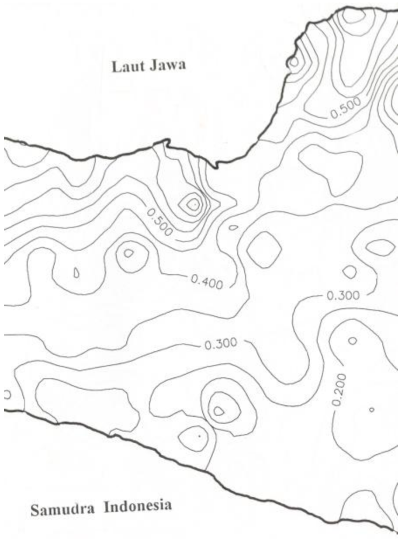
During wet period, factor-1 has the highest value in northern area, along the coasted line to Laut Jawa. The value of isoline tend to smaller into meridional path (North-South). The countour pattern shows rapid changes when it accross the mountain area and become wide apart and smaller in the South area.

The pattern for dry period show the inverse result. The high value of factor-1 can found in southern area and it spreads widely almost covered half of area. The pattern of yearly period show the similar result as dry period. Mostly highest value of factor-1 appears in southern area.

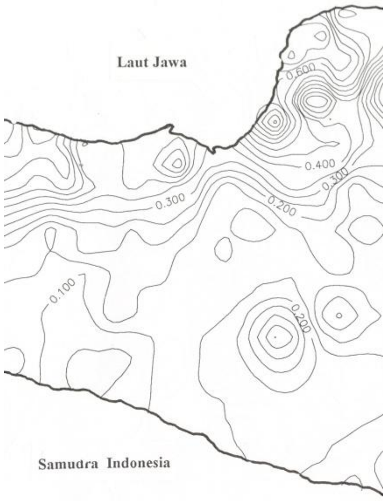
Figure 7 show the distribution of factor-2. Pattern of second factor (factor-2) show the highest value is in the South during wet period. The countour line show rapid changes on the south area and the value become low in the North area. During dry period, the value of factor-2 show the highest value in the middle of observed area, especially near the mountain area, but map of yearly period show the range of value is relative small accross the observed area.



a. Wet period



b. Dry period



c. Yearly period

Figure 7. Isoline of factor 2

Figure 8 show the example of distribution the dominant factor accross the observed area. It can calculate that the factor-1 has spread dominantly on 52% of area, then followed by factor-2 (18%), factor-3(16%), factor-4(6%) and factor-5(8%).

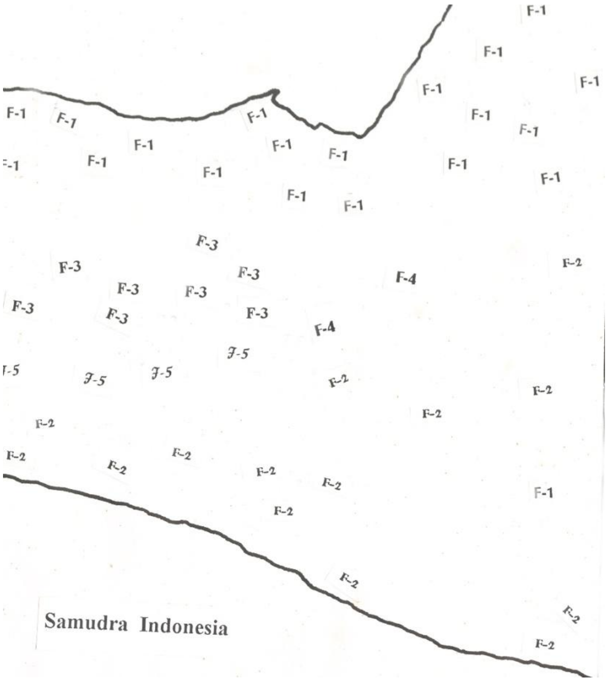


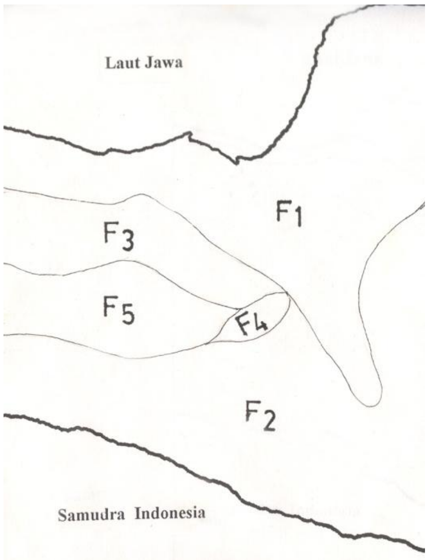
Figure 8. Distribution of Dominant Factor for Wet Period

For easy interpretation, we draw the line to create the border of factor distribution for each period. Figure 9 show the distribution of each factor during each period (wet, dry and yearly). This map can help us to analyze the natural phenomena that has strong relation with these factors.

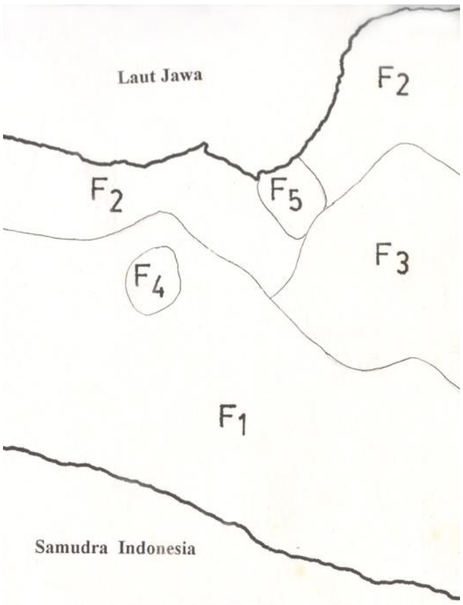
As depicted on the map, factor-1 dominantly appear on the most of observed area during wet and dry period. On the wet period, factor-1 dominant appear in the North area and factor-2 in the South area, and vice viersa on the dry season. In the middle, we can find the appearance of factor-3, factor-4 and factor-5. Map of yearly period show the factor-1 appears in almost of South area and factor2 dominant in North area.

Based on this pattern, we then try to adjust the pattern into correlated natural phenomena, as the caused of rainfall in this observed area. Factor-1, can cleary justified as monsoon factor. It can shown by the area which is covered by this factor during wet and dry period. As we can show on the figure 9, the domination of factor-1 during wet area may has strong relationship with monsoon which is mainly westerly, and vice viersa, in dry season, when the monsoon mainly come from Australia. Factor-2 can identify as sea and valley breeze factor, because as shown on the map, this factor significantly appear on the coast and has decreased when meet the mountain area in the middle. Factor-3 can identify as orographic factor which can caused wind climb the mountain. Factor-4 and factor-5 can identify as local factor, which can caused by local topographic or condition which need more data to uncover it. Generally factor-5 can identify as mountain area or the valley between 2 mountain. Contribution of factor-1 and factor-2 on the observed area can describe the fact that the climate on South coastal areal is relatively wet compare with North coastal area, as joined contribution between factor-1 and factor-2 into rainfall on this area.

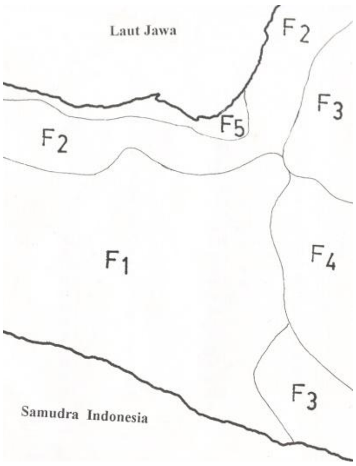
The temporal analysis was not deeply analyzed on this research, although the results also produce some graphics which show some fluctuation. The highest score of factor-1 show the domination of monsoon phenomena, followed by another factor.



a. wet period



b. dry period



c. Yearly period

Figure 9. Distribution of dominant factor accross observed area

5 Conclusions

As the final conclusion of this research, we can find some interesting result, such as:

1. Factor analysis method can explain and simplified the rainfall data interpretation adn can use to identify the hidden factors that contribute the rainfall amount on specific area.
2. The data processing of factor analysis method, spatial and temporal, has produced 5 dominant factor which can assumed has significant contribution on rainfall data.
3. Spatial analysis of factor distribution can show that the biggest factor that caused most of rainfall is monsoon circulation which is dominant over Indonesia area. The other factor has strong relationship with sea-valley wind circulation (factor-2), orographic (factor-3), local (still unidentify, factor-4), and between mountain or valley area (factor-5). Pattern of factor distribution during each period also can explain the fact that South coastal area is relatively humid and wet compare with North coastal area, because this area was dominated by two factor consistenly during each period (factor-1 and factor-2).

This research tends to be a preliminary research which introduce the alternative tools to analyze the rainfall data. For better result,it needs some additional data such as wind, satelite, topography, plantation or area usage (rural, city, industry area, etc), so it can help us to make better analysis and representation of each factor, especially for local factor such as factor-4 and factor-5.

References

1. Akhmad Jamaludin, Z, 1991, ***Identifikasi Pola-pola Cuaca Ortogonal melalui Analisis Komponen Utama dan Spektral***, Final Project, Department of Geophysics and Meteorology, FMIPA, ITB
2. Barry, R.G, and Perry, A.H, 1973, ***Synoptic Climatology: Methods and Applications***, Methuen & Co.Ltd.
3. Cody, Ronald P., and Smith, Jeffrey K. ***Applied Statistic and the SAS Programming Language***, 2nd Edition, Elsevier Science Pub. Co. Inc.
4. Dillon, William R, and Goldstein, 1984, ***Multivariate Analysis : Methods and Application***, John Willey and Sons, Inc.
5. Gedzelman, Stanley David, 1980, ***The Science and Wonders of the Atmosphere***, John Willey & Sons, Inc
6. Lamb, H.H. ***Climate : Present, Past and Future, Vol 1 : Fundamentals and Climate Now***, Methuen & Co. Ltd.
7. Ramage C.S. ***Monsoon Meteorology***, Academic Press, N.Y. London
8. Yuheni, 1993, ***Penggunaan Metoda Faktor Utama dalam Analisis Faktor***, Final Project, Department of Mathematics, ITB.
9. Dyer, T.G.Jl, 1975, ***On The Components of Time Series : The Removal of Spatial Dependenc***e, Quart.J.Roy Meteo.Soc, 101.
10. Lyons, Steven W, 1982, ***Empirical Orthogonal Function Analysis of Hawaiian Rainfall***, J.Applied Meteorology, 21

11. Richman, Michael B, 1980, ***Obliquely Rotated Principal Components : On Improved Meteorological Map Typing Technique***, J.Applied Meteorology, 20.
12. Tjasyono, Bayong HK, Gernowo, R, B.H, Sriworo, and J, Ina, ***The Character of Rainfall in The Indonesian Monsoon***, International Symposium on Equatorial MOnsoon System, September 16th-18th, 2008, Yogyakarta, Indonesia
13. http://en.wikipedia.org/wiki/Factor_analysis
14. http://www.stat-help.com/factor.pdf
15. http://en.wikipedia.org/wiki/Rain
16. http://en.wikipedia.org/wiki/**Monsoon**
17. http://science.jrank.org

Details of author(s)

FALAHAH

Faculty of Engineering,

Universitas Widyatama,

Bandung, West Java, Indonesia

Email: falahah@widyatama.ac.id

S. SUPRAPTO

Department of Meteorology,

Institut Teknologi Bandung,

Bandung, West Java, Indonesia

Email: cirrusfoto@gmail.com