Evaluation of the Use of Data Reanalysis for Climate Regionalization

Yon Sugiarto1*, Perdinan2, Tri Atmaja3, Shalsa Nurhasanah4

1,2,3,4Department of Geophysics and Meteorology, Bogor Agricultural University, Indonesia
3Generasi Hijau Indonesia
*E-mail: yons@ipb.ac.id

Abstract. Climate regionalization has an important role in providing information on profile and climate potential of the region, but its requires long time series data with equally spatial distribution. Climate data in Indonesia is still a constraint related to the limitations of spatial and temporal data availability. Therefore, the use of good quality of reanalysis data can be an alternative to obtain information on profile and climate potential in local or regional scale. This paper evaluate the use of daily rainfall reanalysis data based on WordClim and CHIRPS with spatial resolution of 1 km2 and 5 km2 respectively in the study area of Subang district, West Java. The climate regionalization of the region was done by three climate classification methods and statistical approach (PCA) using cluster analysis based on seven linkage methods. The results of the analysis show the similar pattern of seasonal and yearly spatial distribution of rainfall on both reanalysis climate data with annual rainfall range between 1330-3700 mm and 1200-3300 mm for WorldClim and CHIRPS data respectively. Climate regionalization produced seven clusters on both reanalysis data with monsoonal rainfall pattern, characterised by one peak of rainfall at the beginning of the year. The result also indicates the similarity of monthly rainfall pattern from north to south region following altitude distribution. Despite of some different patterns noticeably for monthly temporal distribution, these findings suggest that both climate reanalysis data is proper to be use as alternative to obtain information on profile and climate potential of the region.

Keywords: CHIRPS, climate, PCA, reanalysis, regionalization, WorldClim.

1. Introduction

Climate refers to the statistical properties of the atmosphere and is concerned with the long-term behavior, or expected (typical) conditions. Climate condition varies spatially over a region determined by topographical, geographical, and local condition. Generally, there is a pattern that underlies climate regions. Therefore, climatic classification is become common as an efficient method to identify and describe the climatic differences on a regional scale. The ideal climate classification should clearly differentiate among all the major type of climate that occur earth, show the relationships among these climate types, apply to the global scale of the world surface, provide a framework for subdivision to cover specific locale factors and demonstrate the controls that cause any particular climate such as causing the changing climate [1].
There have been many classifications developed for both global, regional or even local scales. Several climatic classifications have also been developed for special purposes, for examples are climate classification used for crop agriculture, plantation and forestry sectors. In global scale, the Koppen climate classification is the first quantitative classification [2-3] and now is probably the most widely used [4]. The Koppen system based on world distribution of natural vegetation boundaries and the combinations of monthly mean temperature and precipitation associated with those boundaries. Specifically in Indonesia, three climate classifications widely used are Koppen, Oldeman, and Schmidth – Ferguson. All the climate classifications are required rainfall data as input.

Climate classification requires long time series data with equally spatial distribution. The availability of observations of climate data depends on the coverage, in space and time, of global observing systems, which includes weather station, radiosonde and others. The main limitation of the climate data observation is coverage, where stations usually undistributed equal. Another constrain is related to data quality that may vary with time. These are become a serious problem where limitations of spatial and temporal climate station data are usually occur in Indonesia. Alternative data source derived from satellite, re-analysis, and models for employing climate classification techniques to define spatial variability of climate types over a region. Reanalysis data can be no better than the underlying observations, or than the models that process the data, but its still become an opportunity to solve the problem of climate data limitation.

WorldClim and CHIRPS are the samples of reanalysis climate data, contained a set of global climate layers (gridded climate data) with specific spatial and temporal resolution. These data can be used for mapping and spatial modeling. Both types of data have been widely used in various climate research at regional and global levels to strengthen spatial analysis of observational climate data. Both reanalysis data have different advantages and disadvantages related to spatial and temporal resolution. WorldClim has better spatial resolution with more limited temporal data, while CHIRPS has a lower temporal spatial but has a longer collection of data with temporal resolution of 5 and 10 daily. Both reanalysis data offer an opportunity to use in certain purpose of analysis. Therefore, it is important to know the character of the results of the spatial analysis of the climate using both the reanalysis data. This study aims to evaluate the use of gridded daily rainfall reanalysis data from WorldClim and CHIRPS specifically for climate regionalization and provides climate classification for crop production growing regions of Subang district, West Java – Indonesia.

2. Material and Methods
2.1. Material
Reanalysis climate data derived from WorldClim and CHIRPS with 1 km² dan 5 km² spatial resolution data respectively are used to produce climate regionalization. Wordclim data used is for 1971 - 2000 year period developed by [5] and it is available online at www.worldclim.org, includes data on rainfall and air temperature. CHIRPS is a global coverage at high temporal and spatial resolutions with time series data availability goes back to 1981 and available online at http://chg.geog.ucsb.edu/data/chirps [6]. The data used for this research is for 1981 - 2010 year period. The year period of 1971-2000 and 1981-2010 are considered to represent current or baseline of climate profile.

Furthermore, Season Zone information (ZOM) issued by BMKG is used as a reference for the amount of rainfall and season in the study area. The spatial resolution on WorldClim data is quite detailed and subtle so that it can be used to produce climate classification while CHIRPS has greater resolution but more detailed data because it is available up to daily data.

2.2. Material
Cluster analysis Reanalysis climate data from Worldclim and CHIRPS are used in the climate regionalization process based on Koppen, Oldeman and Schmidt Ferguson classification system. Furthermore, mapping of climate regionalization was done using hierarchy cluster method and non-hierarchy cluster analysis (K-means). Classification and mapping of climate sub-regions are aimed to understand the climate characteristics of the study area. Regionalization will be performed using a
combination of Principal Component Analysis (PCA) and cluster analysis using monthly rainfall data. Both approach has been widely used to create climate regionalization (e.g., [7-10]) The use of PCA is intended to eliminate the multicollinearity between climatic variables used in the zonation process (in this case the monthly data). In general, cluster analysis is aimed to maximizing group differences and minimizing differences in groups [11].

The hierarchy cluster technique chosen is by linkage ward and distance Euclidean based on the result of dendogram obtained. In the cluster hierarchy method, the number of clusters can be seen using dendrogram and scree plot as a result of statistical analysis in the processing of rainfall data. The ward method calculates the number of groups based on the total number of squares deviation. Determination of the number of groups using dendrogram that is by cutting the dendrogram of the conformity of the results obtained.

Another case uses the non-hierarchy method of K-means. K-means clustering method has a principle by entering the value of K as a constant of the desired number of clusters. Rainfall data grouped into the number of clusters sees similar data. The similarity of this data is measured by the distance measurement method to find out the error value of rainfall data distance to the centroid. After the clustering process, then their needs to be cluster validation as evaluation of cluster algorithm result, this is because cluster determination using K-means method can cause bias where certain cluster membership can move to another cluster. The optimal cluster is obtained by looking at the smallest number of variations within cluster by calculating using within-groups sum-square. And then the cluster validation can be determined by Beale's Pseudo F. The number of factors formed is used in the process of non-hierarchy and hierarchy cluster analysis. In the hierarchy cluster process, each data is assumed to be a cluster. If the amount of data (n) and the number of clusters (k) then the magnitude will be n = k, and the process of calculating the distance between clusters by using Euclidian distance based on the average distance between objects. Then the calculation result is selected minimum distance and combined so that yield value of n = n-1 and will continue repetition until fulfill condition number k = 1. At this stage, the climatic type based on hierarchy cluster method and non-hierarchy of K-means was obtained. Further, climate regionalization issued by BMKG or better known as Season Zone (ZOM) was used as comparison of climate patterns clustering.

3. Result and Discussion
In geographical perspective, the study area in Subang district is located at eastern part of Java Island and has a topographic profile with altitude that varies from 0 to 1500 meter above sea level. This topographic profile causes the study area to have a coastal climate and low land to highland climate profile. Subang district has an average total of precipitation around 3000 mm/year with monsoonal rain pattern (one peak rainy season) and has divided into 4 regions based on Season Zone (ZOM) used by BMKG. ZOM was produced to identify areas that have a clear climatological boundary between the period of the rainy season and the dry season. Distribution of 4 ZOM regions for Subang district are Southern region (ZOM74), West part region of central Subang (ZOM75), Northern region of Subang (ZOM76) and East part region of central Subang (ZOM80).

Figure 1 shown the result of climate regionalization based on three climate classification systems. For Koppen classification system, the study area indicates a tropical climate type, consist of three subtypes which are Af, Am and Aw. Both climate data reanalysis used for the Koppen classification gave similar results for the entire coastal area with the classification class of Aw. Aw or tropical wet and dry climate has a pronounced dry season. They undergo much greater variability in seasonal precipitation and temperature than do the tropical wet and the monsoonal climates. Based on the climatic classification, agricultural land in northern Subang is indeed more vulnerable to drought than other regions.

Regions with low to medium altitude areas from 100 - 500 meter above sea level have climate classification class of Am for both WorldClim and CHIRPS data. This area is centers of agricultural production of food crops in the study area. Am is monsoonal climate undergoes relative dryness for 1
to 3 months but receives sufficient moisture that vegetation need not be adapted to seasonal drought. Am climates usually occur along tropical, coastal areas subjected to predominant onshore winds that supply warm, moist air to the region throughout most of the year. In Subang district, the area of class Am has a quite close location to the coastal area with a distance around 10 km from along the coastline.

The third subtype tropical climate class found in the study area is Af or the tropical wet climate region that has significant rainfall every month of the year. Tropical wet climates have no dry period and precipitation is almost always convectional with strong solar heating of the surface triggering the brief but heavy thundershowers in the mid to late afternoon. In this climatic class, Af is found in areas with an altitude from 500 - 1500 meter above sea level. Horticultural crops and forest in highest altitude regions dominate land use in most areas.

In general, WorldClim and CHIRPS data have the same class distribution results for the Koppen classification. The difference in spatial resolution for WorldClim and CHIRPS that are 1 km² and 5 km² respectively only generate small differences to the classification result. This may occur because Koppen classification uses monthly rainfall and air temperatures data, while both reanalysis of climate data performed similar patterns and variability for monthly climate data.

The use of WorldClim and CHIRPS data for Oldeman classifications gave slightly different results between both data sources. In determining its climatic classification, Oldeman only uses average monthly rainfall date period of the wet months respectively from the average monthly rainfall during a particular observation period. Classification with WorldClim data produced 7 climatic classes, which
are C1, C2, D2, D3, D4, E3 and E4. This illustrates that the study area has a climatic class that tends to have only a fairly short wet season period of 5-6 wet months with CH> 200 per month. This causes the region as a vulnerable area of drought, and if it does not have irrigation support then this area only have one season planting opportunities for rice crops. Classification results with CHIRPS data also showed relatively different with 6 classes, which are B1, B2, C2, D2, E2 and E3. Class B showed a fairly long wet season period with wet months with CH> 200 mm reaching 7 - 9 months.

The Schmidth Ferguson classification that is widely used in the field of forestry and plantation determines classification class based on the value of Q, which is the ratio between the mean dry months and wet months. The Q value can be expressed as a drought index, where the greater the value indicates the drier region. The Schmidth classification analysis using Wordclim and CHIRPS data shows a similar pattern in which each produces classes A, B, C and D. The entire north coastal area has class D of classification for analysis with Wordclim data, but only produces class D on a small part of the northern west coast of the study area while using CHIRPS data. The use of WorldClim data also produced a classification pattern of successive classes are C, B and A for more southerners region with higher topography. This result also shows the consistency of the classification class with the higher rainfall pattern in the higher plateau region. The result of analysis using CHIRPS data also shows the same pattern where there is influence of topographic distribution. The classification results of the use of both reanalysis data also indicate similar pattern of area where classification class is dominated by area for class A and C.

The climate regionalization using hierarchical technique revealed that the climate regions for the Subang District could be distinguished into seven clusters that expand from south to north. The application of non-hierarchical (i.e., the K-means) also shows the same numbers of clusters for the district (Figure 2). The comparison of the two regionalization methods showed that the variation of climate types, i.e., the distribution and the zonal extent of each cluster, are only slightly different as the general pattern of south to north gradient still holds.

Figure 2. The results of climate regionalization using hierarchical technique for reanalysis climate data from WorldClim (1971-2000) and CHIRPS (1981-2010)
The results of climate regionalization based on principal component analysis (PCA) using the reanalysis rainfall data from WorldClim in the study area produced seven clusters that following the distribution of topographic pattern in the region. This result indicates a possibility effect of the movement of air masses between the sea in the north and the highland in the south to rainfall pattern in the study area. The use of CHIRPS data provides the same cluster distribution with the WorldClim data that follows the topographic pattern of the region but with a smaller diversity. Figure 2 above shown that clusters 1 and 2 on the results of WorldClim data overlapped with cluster 1 on the results of analysis from CHIRPS data. Meanwhile, clusters 5 and 6 on the results of WorldClim data also overlapped with cluster 5 from the results of CHIRPS data. The WorldClim data also has higher spatial resolution, it is implied in the result of the cluster shape that is more variable with smoother cluster boundaries line. This is particularly important for local climate analysis in smaller region.

Monthly rainfall pattern of 7 clusters based on WorldClim and CHIRPS data can be seen in Figure 3. The result of climatic regonalization performed using cluster analysis for study area produced 7 clusters outputs in both reanalysis data with monsoonal rainfall pattern. Cluster from 1 to 4 have most consistent monsoon rainfall patterns characterized by one peak of rainfall at the beginning of the year. In this pattern, rainfall peaks occur at the beginning of the year and continue to decline until the middle of the year. After that rainfall again increased until the end of the year in a gradual manner. This result is consistent to [12] that stated most of southern region of Indonesia, specially Java island have one peak and one trough and experiences strong influences of two monsoons, namely the wet northwest (NW) monsoon from November to March (NDJFM) and the dry southeast (SE) monsoon from May to September (MJJAS).

There are also similar pattern of seasonal and yearly spatial distribution of rainfall in both reanalysis climate data. On Java Island, most of the rainfall patterns from weather stations show very distinctive wet and dry seasonal variations due to the Asian-Australian monsoon [13], which affects Southeast Asia and Australia via the winds from the southeast and northwest during the cooler and warmer months, respectively [14]. Zonal pattern of rainfall are captured by regional climate profile from both data, although, WorldClim provide more detailed variation considering the higher spatial resolution of the data than CHIRPS.
Figure 3. Monthly rainfall pattern of seven clusters based on hierarchical technique using reanalysis climate data from WorldClim (1971–2000) and CHIRPS (1981–2010)

4. Conclusion
Despite of some different patterns noticeably for monthly temporal distribution, these findings suggest that WorldClim and CHIRPS as gridded climate reanalysis data are proper to be used as alternative to obtain information on climate potential of the region and provide an alternative source to perform climate classification when the site observations are often limited. The spatial variation of climate types may also be in line with topographical condition of the region. Future works can further employ the
regional climate classification to study land-climate suitable areas for many purposes such as distribution of agricultural commodities over crop production regions.

**Acknowledgement**

This research is supported by the Indonesia Climate Change Trust Fund (ICCTF-Bappenas) Project No. 016/SPK/PPK-06.14/SGP/04/2016. The authors highly appreciate the ICCTF team and the officers of Subang district government involved in this research for their valuable contribution.

**References**

[1]. de Blij HJ, Peter OM, Richard SW Jr 2004 *Physical Geography* The Global Environment 3th edition Oxford University Press New York 702 ISBN 0-19-516022-3

[2]. Essenwanger OM 2001 Classification of Climates World Survey of Climatology 1C General Climatology Elsevier Amsterdam 126 ISBN-13:978-0444882783

[3]. Kottek M, Jürgen Grieser, Christoph Beck, Bruno Rudolf and Franz Rubel 2006 World Map of the Köppen-Geiger climate classification updated *Meteorologische Zeitschrift* 15 (3) 259-263

[4]. Peel MC, Finlayson BL, and McMahon TA 2007 Updated world map of the Köppen-Geiger climate classification, *Hydrol Earth Syst. Sci.* 11 1633-1644 https://doi.org/10.5194/hess-11-1633-2007

[5]. Hijmans RJ, Cameron SE, Parra JL, Jones PG, and Jarvis A 2005 Very high resolution interpolated climate surfaces for global land areas *International Journal of Climatology* 25 (15) 1965-1978 doi: 10.1002/joc.1276

[6]. Funk, Chris, Pete Peterson, Martin Landsfeld, Diego Pedreros, James Verdin, Shrefnanand Shukla, Gregory Husak, James Rowland, Laura Harrison, Andrew Hoell and Joel Michaelsen The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes *Scientific Data* 2 150066 doi:10.1038/sdata.2015.66 2015

[7]. Gong XF and Richman MB 1995 On the application of cluster-analysis to growing-season precipitation data in North-America East of the Rockies *Journal of Climate* 8(4) 897-931

[8]. Stooksbury DE and Michaels PJ 1991 Cluster-analysis of southeastern United-States climate stations *Theoretical and Applied Climatology* 44(3-4) 143-150

[9]. Unal Y et al 2003 Redefining the climate zones of Turkey using cluster analysis *International Journal of Climatology* 23(9) 1045-1055

[10]. Winkler JA 1992 Regional patterns of the diurnal properties of heavy hourly precipitation *The Professional Geographer* 44(2) 127-146

[11]. Kalkstein LS et al. 1987 An evaluation of three clustering procedures for use in synoptic climatological classification *Journal of Climate and Applied Meteorology* 26(6) 717–730

[12]. Aldrian E and Dwi SR 2003 Identification of three dominant rainfall regions within Indonesia and their relationship to sea surface temperature *Int. J. Climatol* 23 1435–1452 doi:10.1002/joc.950

[13]. Lee HS 2015 General Rainfall Patterns in Indonesia and the Potential Impacts of Local Seas on Rainfall Intensity *Water* 2015 7 1751-1768 doi:10.3390/w7041751.

[14]. Rafferty JP 2011 *Winds and windstorms. In Storms, Violent Winds, and Earth’s Atmosphere* Rafferty JP Ed Britannica Educational Publishing New York USA Chapter 2