

Visual Exploration of Climate Variability in the Philippines

Wendell Q. Campano
School of Statistics
University of the Philippines

Rona May U. Tadtas
Department of Computer Science
University of the Philippines

In this paper, a data visualization framework for investigating and exploring climate time series data is introduced. This method utilizes the results obtained from performing series of cluster analysis based on a particular multivariate dataset for each defined subset in the time series. The said approach is implemented to the climate data in the Philippines. The data image results obtained from the procedure revealed the expected overall climate pattern in the Philippines as well as some localized segments of climate changes in the time series which deviate from the overall pattern. A wavelet analysis which is a well-established method in analyzing climate data is also done to validate the results shown by the proposed visualization method.

Keywords: *Information visualization, data image, cluster analysis, wavelet, climate change, climate variability, time series, multivariate data*

1. Introduction

People have evolved ways on how to live their lives and have adapted to the climates in which they lived. However, the climate is ever changing and large deviations from the norm with increasing frequency can be very disturbing, and even hazardous. Due to its potential large impact on ecology, economy and future developments, climate change increasingly draws the attention of scientific researchers around the world.

According to Intergovernmental Panel on Climate Change (IPCC), the biggest challenge in confronting the negative impacts of climate change lies in the developing world, where people and systems are most vulnerable. Thus, the

impact is most serious in subtropics and tropics where most developing countries reside because the coping or adaptive mechanisms in these regions are limited. Furthermore, research on how to respond and deal with the impacts of climate change are also limited in developing countries (Leary et al., 2008).

In the Philippines, it is evident that the effect of climate change is far ranging affecting areas such as food production, health, safety, economy, etc. Improved understanding of climate variability, especially during extreme cases which brings the greatest environmental hazards, is relevant in achieving sustainable development.

Even though, climate change is a very important and widely discussed subject, only few visualization papers are directly concerned with climate applications (Janicke et al., 2009). That is why we are hoping to give additional insights about the said topic through visualization.

In reality, climate variability analysis should involve multivariate data analysis because climate systems and conditions are associated not only to one but to several factors (Janicke et al., 2009). The proposed visualization allows us to utilize multivariate time series data simultaneously since cluster analysis is a part of the method which is a multivariate technique itself.

In this paper, we analyze the Philippine time series climate using our method. The main goal of this empirical analysis is to assess whether the proposed visualization approach can clearly present a logical overall picture of the climate pattern in the Philippines. We say logical if the results will somehow conform to what the weather administration (PAGASA) in the Philippines has already defined. The results are also compared to the results obtained from wavelet analysis adapted from Torrence and Compo (1998) so that we can assess the validity of the proposed framework. Wavelet analysis is already a popular choice and a well-established method in analyzing climate variability (Janicke et al., 2009; Santos et al., 2001; Torrence and Compo, 1998; Yueqing et al., 2005).

The main purpose of this paper is to introduce a simple yet effective alternative approach in exploration of and analyzing climate data.

In Section 2, we present the description of the data used for the analysis as well as the concepts involved in this study. Then, the procedure for the proposed visualization method and for the implementation of wavelet analysis is discussed in Section 3. Finally, the results and discussion pertaining to the application of the methods to Philippine climate data and conclusion are presented in Sections 4 and 5, respectively.

2. Description of Terms Concepts

Here, the details about the empirical data used in the study and an overview of the wavelet analysis and clustering method considered for the proposed procedure are presented.

2.1 Data

The data sets used in this study are multivariate time series climate measurements from Philippine Atmospheric, Geophysical, and Astronomical Services Administration (PAGASA). The data are average of the measurements from Vigan, Ilocos Sur and Hinatuan, Surigao Del Sur stations. Ilocos Sur and Surigao Del Sur are situated in Northern and Southern Philippines, respectively. These measurements include monthly average mean sea level pressure or MSLP (Pa), temperature ($^{\circ}\text{C}$), and relative humidity or RH (%) from 1951 to 2002. Note that relative humidity can also be a proxy for rainfall since high humidity is directly associated with the formation of water in the air. Though, of course, relative humidity usually attains its maximum value during the warmest month because rain is formed when the relative humidity or the amount of water vapor in the air is very high. Note that, in contrast to rainfall data, relative humidity is not greatly influenced by extreme or outlying disturbance such as typhoons which is common throughout the year. And thus, we can say that it a robust basis for measuring the “natural” climate pattern in the Philippines.

In order to have a clearer picture of the variability, the measurement anomalies were also used in the analysis. Anomaly series are derived by removing mean annual cycle. This is done by building the difference between the monthly mean fields and the means of the corresponding 52 monthly mean fields of the years 1951-2002 (Janicke et al., 2009).

2.2 Wavelet analysis

Wavelet analysis is a method used for investigating signal structures that change over time. It is a useful tool for analyzing time series with many different time scales or changes in variance. By visualizing the time-frequency space power spectrum derived from this method, one can identify time periods on which the frequencies of a particular signal are dominant. It is now becoming a popular tool in analyzing climate data (Janicke et al., 2009; Torrence and Compo, 1958; Yueqing et al., 2005). The next subsection describes the concepts used in wavelet analysis as discussed by Torrence and Compo (1958).

2.2.1 Wavelet transform

Mathematical transformations are applied to raw series to obtain further information which is not readily available in the original series. Assume that one has a time series, y_n , consisting of $n=0 \dots, N-1$ time steps with equal spacing δt . Furthermore, assume that one has a wavelet function, $\psi_0(\eta)$ that depends on a nondimensional “time” parameter η . To be admissible as a wavelet, this function must have 0 mean and be localized in both time and frequency space. A wavelet widely used in climate research is the Morlet wavelet (Janicke et al., 2009), consisting of a plane wave modulated by a Gaussian:

$$\psi_0(\eta) = \pi^{-1/4} e^{i\omega_0\eta} e^{-\eta^2/2} \quad (1)$$

where $i^2 = -1$, and ω_0 is the nondimensional frequency, here taken to be 6 to satisfy the admissibility condition (Torrence and Compo, 1998). In this study, the procedure is adapted from the one described by Janicke et al. (2009). Hence, first the original series y_n and the wavelet is Fourier transformed. The discrete Fourier transformation (DFT) of y_n is given by:

$$\hat{y}_k = \frac{1}{N} \sum_{n=0}^{N-1} y_n e^{-2\pi i k n / N}, \quad (2)$$

where $k = 0, \dots, N-1$ is the frequency index. And the Fourier transform of the Morlet wavelet is

$$\hat{\psi}_0(s\omega) = \pi^{-1/4} H(\omega) e^{-(s\omega - \omega_0)^2} \quad (3)$$

where s is the scale and ω_k the angular frequency

$$\omega_k = \begin{cases} \frac{2\pi k}{N\delta t} & \text{for } k \leq \frac{N}{2} \\ -\frac{2\pi k}{N\delta t} & \text{for } k > \frac{N}{2} \end{cases} \quad (4)$$

and is the Heaviside step function

$$H(\omega) = \begin{cases} 1 & \text{if } \omega > 0 \\ 0 & \text{otherwise} \end{cases}$$

To ensure that the wavelet transforms at different scales are comparable, the wavelet function at each scale s is normalized to have unit energy:

$$\hat{\psi}(s\omega_k) = \sqrt{\frac{2\pi s}{\delta t}} \hat{\psi}_0(s\omega_k) \quad (5)$$

After the two transformations above, they are then multiplied pointwise. The wavelet transform is the inverse of the Fourier transform of the product:

$$W_n(s) = y_n * \psi_0 = \frac{1}{N} \sum_{k=0}^{N-1} \hat{y}_k \hat{\psi}^*(s\omega_k) e^{i\omega_k n \delta t} \quad (6)$$

where (*) indicates the complex conjugate.

2.2.2. Power Spectrum

Visualization of the power spectrum can give insights about what part of the time-frequency the series is dominating. From (6), we define a wavelet power spectrum as $|W_n(s)|^2$ and that

$$E\left[|W_n(s)|^2\right] = NE\left[|\hat{y}_k|^2\right]$$

Thus, for a white-noise time series,

$$E\left[|\hat{y}_k|^2\right] = \frac{\sigma^2}{N}$$

where σ^2 is the constant variance, so

$$E\left[|W_n(s)|^2\right] = \sigma^2$$

at all n and s . Therefore, the normalized wavelet power spectrum for the climate anomalies studied here is given by $|W_n(s)|^2/\sigma^2$. The normalization by $1/\sigma^2$ gives a measure of the power relative to white noise.

To compute the scale parameter s , Torrence and Compo (1998) suggest the following formula:

$$s_j = s_0 2^{j\delta j} \quad j = 0, \dots, J \quad s_j = s_0 2^{j\delta j}$$

$$J = \frac{\log_2(N\delta t / s_0)}{\delta j}$$

where s_0 is the smallest resolvable scale and J determines the largest scale. Note that δj depends on the width in spectral-space and wavelet function. For the Morlet wavelet $\delta j = 0.5$ is the largest value for which δj gives adequate sampling.

The power spectra comprise a large variety of subtle structures and not all of them encode significant changes. Torrence and Compo (1998) showed that if the original Fourier components are normally distributed then the power spectrum $|W_n(s)|^2$ has a chi-square distribution with 2 degrees of freedom. See Torrence and Compo (1998) for more details.

2.3 Cluster analysis by Ward's method

Cluster analysis is a data exploration method of classifying observations into groups (called clusters) according to some similarity or distance measure. There are many procedures and algorithms developed for cluster analysis such as Average Linkage (Sokal and Michener, 1958), Centroid (Sokal and Michener, 1958), Median Method (Gover, 1967) and Ward's Minimum Variance (Ward, 1963). In this study, the Ward's Minimum Variance Method is utilized.

In Ward's minimum-variance method, the distance between two clusters is the analysis of variance (ANOVA) sum of squares error between the two clusters added up over all the variables. At each step, the within-cluster sum of squares error is minimized over all partitions obtainable by merging two clusters from the previous generation. The sums of squares are easier to interpret when they are divided by the total sum of squares to give proportions of variance called the squared semipartial correlations. For more details about Ward's Method, see Ward (1963).

Ward's method tends to join clusters with a small number of observations. Moreover, it is strongly biased toward producing clusters with roughly the same number of observations. These are also the reason why this method was used since only clustering of 12 months for each year is done.

2.4 *Philippine climate*

According to PAGASA, the general climate of the Philippines can be divided two major season:

- rainy/wet season, occurring from June to November; and
- dry season, occurring from December to May.

Furthermore, the dry season may be subdivided further into:

- cool dry season, which occurs from December to February; and
- hot dry season, which occurs from March to May.

However, these defined seasons can be altered by some disturbance such as typhoons and global climate changes.

3. **Methodology**

Now, we present the procedure on how to create data images to aid in investigating climate variability. In Section 3.1, the steps in creating data images based on the new framework is summarized. The wavelet analysis implementation is discussed in Section 3.2.

3.1 *Data image of clusters*

This section describes the procedure for the creation of the overall data image of cluster analysis done for each year of the entire time series. The cluster numbers used are 2 and 3. These numbers are chosen so that each can be a relatively close representation of the general climate pattern in the Philippines – Dry, Cool or Moderate, and Wet Season. The steps are enumerated below.

1. For each year, perform cluster analysis to determine the group membership of each month based on the mean sea level pressure and relative humidity measurements. Keep in mind that the cluster labeling or naming (such as “dry” or “wet”) should be consistent in terms of MSLP and RH property across years. A plot of the cluster analysis result for the year 1954 is shown

in Figure 1. Note that only the cluster or group membership of the months is of interest in this step however, a plot like Figure 1 could help in determining the right color to be associated for a particular cluster.

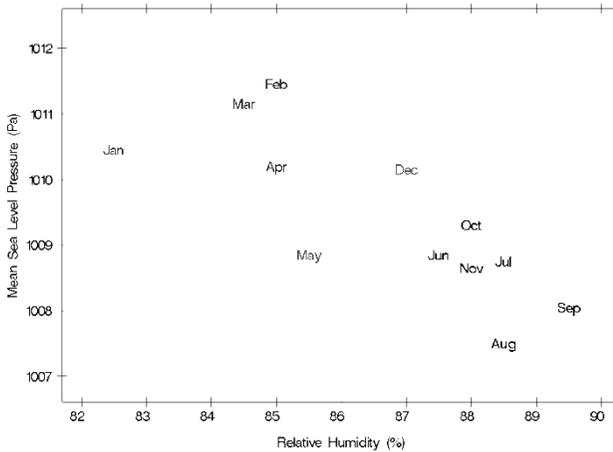


Figure 1. Plot of cluster membership for year 1954

2. Concatenate all the obtained grouping in (1) to form the 3 fields for year, month, cluster where year = 1951,...,2002; month = J an,...,Dec; and cluster = 1 ,2, and 3.
3. The data image of the whole series based on the cluster analysis is obtained by plotting the years in the x-axis and month in the y-axis while the groupings are represented by 3 different colors (e.g. 1 = red, 2 = green, 3=blue) of the points in the plot.

3.2 Wavelet power spectrum contour plot

The visualization of the power spectrum is adapted from the wavelet analysis procedure introduced by Torrence and Compo (1958). The outline of the method is given below.

1. Find the Fourier transform of the time series.
2. Choose a wavelet function and a set of scales to analyze. In this study, Morlet is used. The parameters are: $s_0=2\delta t$, $\delta t=0.25$ to do 4 suboctaves per octave with $j_1=7/\delta j$ to do 7 power-of-two with δj suboctaves each.
3. For each scale, construct the normalized wave let function using (6).
4. Find the wavelet transform at that scale using (4);
5. Determine the cone of influence and the Fourier wavelength at that scale. After repeating steps 3–5 for all scales, remove any padding and contour plot the wavelet power spectrum.

- Assume a background Fourier power spectrum (e.g., white or red noise) at each scale, then use the chi-squared distribution to find the 95% confidence (5% significance) contour.

Software programs for FORTRAN, IDL, and MATLAB for implementing the above procedure were also developed by the said authors. In this paper, we used their program for MATLAB. The use of the wavelet analysis here is to provide a more in-depth insight about the visualization of the data image of the climate variability described in Section 3.1.

4. Results and Discussion

Figures 2 and 3 show the data visualization output of the procedure described in Section 3.1. These are the data images of the cluster analysis results. Note that the clusters shown in Figure 1 for year 1954 are also represented in Figure 2 through a colored film strip on the corresponding year.



Figure 2. Data image based on the yearly 2-cluster results

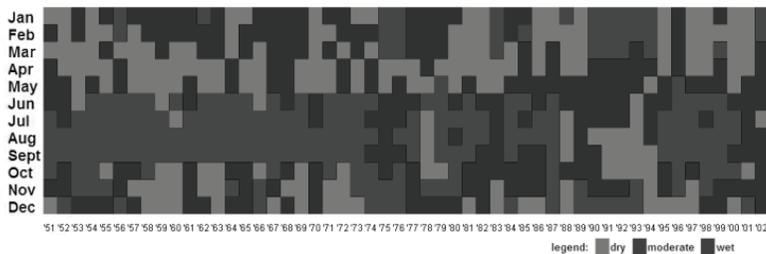


Figure 3. Data image based on the yearly 3-cluster results

As seen in Figure 2, in most time points, the data image shows consistency with the general climate pattern in the Philippines which consists of wet season for the months June to somewhere near September or October and dry season for the rest of the year. There are, however, years where the harmony is distorted as it can be seen, for example, in the years from 1973 to 1976. But in these time points, it is known that there were three occurrence of la niña and one el niño phenomena.

Another example is during 1978 which also deviated from the “usual” climate pattern however, it turned out that there was also an occurrence of an el niño in 1977-1978. This was also the case for the year 2002 because of the reported 2002-2003 el niño.

In general, the data image in Figure 2 shows that during the years 1974-1978 and 1988-1993, there are localized pattern variations while the rest of the time points exhibits the “usual” global climate pattern.

Note that PAGASA also established that the dry season can be subdivided further into cool season, which we call here moderate (December-February) and the hot dry season or simply dry season (March to May). Figure 3 shows the data image based on performing cluster analysis and forming 3 clusters per year to represent 3 seasons.

Evidently, based from the 3-cluster data image, there has been a big change in the climate pattern starting late 1980s to early 1990s and towards the end of the series where the majority of the months were in the moderate to dry condition. These findings were also evident in the visualization of the power spectrum of temperature and of temperature anomaly.

In Figure 4, we can observe that concentration of power is within the 2-4 years band which shows that the series oscillates every 2-4 years, in general. Moreover, within the 2-4 years band, the wavelet power of the temperature data is significant in the years 1970-1975, 1980-1990, and 1999 until early 2000s. On the other hand, the power spectrum of the temperature anomaly in Figure 5 indicates strong energy starting 1985 until the end of the series.

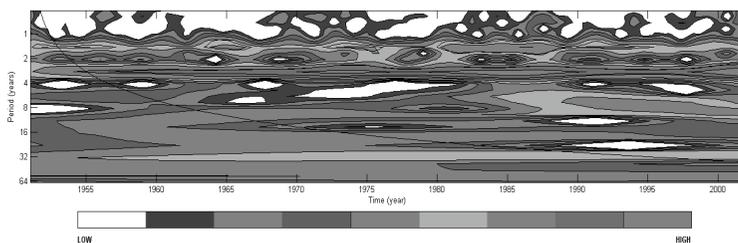


Figure 4. Contour plot of power spectrum of temperature (°C)

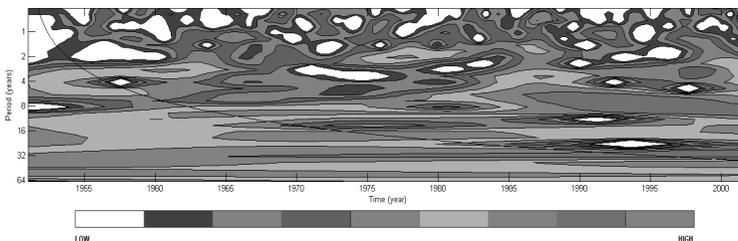


Figure 5. Contour plot of power spectrum of temperature anomaly (°C)

The power spectrum of the MSLP anomaly also shows a relatively high concentration of power starting late 1980s which also suggests that the climate pattern in these time segments changed as also revealed through the data images in Figures 2 and 3. Meanwhile, for the relative humidity in Figure 7, although there is somewhat structural change in the power spectrum during 1980s, the values are not significant. Hence, giving evidence about our claim earlier that relative humidity is somewhat robust or insensitive to influential fluctuations of the climate.

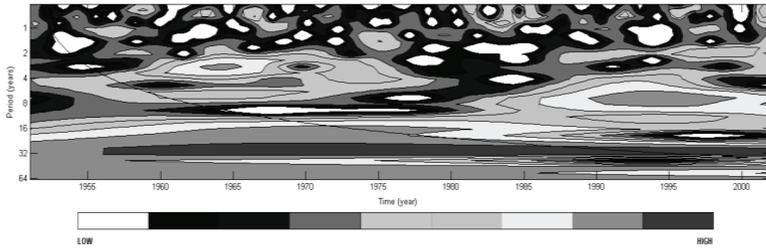


Figure 6. Contour plot of power spectrum of mean sea level pressure anomaly (Pa)

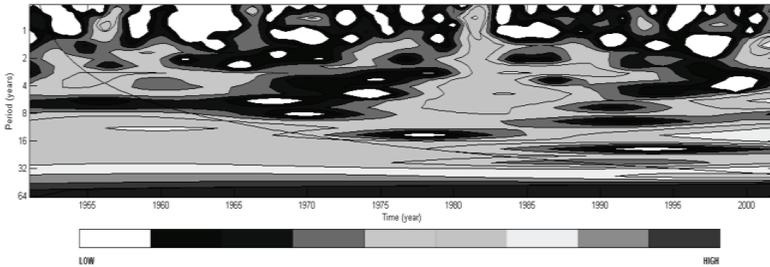


Figure 7. Contour plot of power spectrum of relative humidity anomaly (%)

5. Conclusions

In this paper, we introduced and described a wavelet- and visualization-based frame work in investigating climate patterns and variability based on visualization of series of cluster analysis. Based on the empirical study of the climate in the Philippines, the data images obtained from the proposed framework effectively captured the underlying general climate pattern as well as revealing localized variations or climate changes that departed from the overall pattern. Since unusual behaviors can easily be pinpointed through visual inspections, the method can also serve as a preliminary exploratory data analysis for studies that would consist of in-depth, localized, and detailed analysis of time series data. An assessment

of the results from the data images were also cross validated with the results of the widely used wavelet analysis and it turned out that the two methods showed consistent results.

REFERENCES

- GOWER, J.C., 1967, A Comparison of some methods of cluster analysis, *Biometrics* 23: 623-637.
- JANICKE, H., M. BOTTINGER, U. MIKOLAJEWICZ, and G. CHEUERMANN, 2009, Visual exploration of climate variability changes using wavelet analysis, *IEEE Transactions on Visualization and Computer Graphics* 15(6): 1375-1382.
- LEARY, N., et. al. (ed), 2008, *Climate Change and Vulnerability*, Earthscan, London.
- SANTOS, C., C. GALVAO, K. SUZUKI, and R. TRIGO, 2001, Matsuyama City rainfall data analysis using wavelet transform, *Annual Journal of Hydraulic Engineering*, JSCE, 45.
- SOKAL, R.R., and C.D. MICENER, 1958, A statistical method for evaluating systematic relationships, *University of Kansas Science Bulletin* 38:1409-1438.
- TORRENCE, C. and G.P. COMPO, 1998, A practical guide to wavelet analysis, *Bulletin of the American Meteorological Society* 79:61-78.
- WARD, J.H., 1963, Hierarchical grouping to optimize an objective function, *Journal of the American Statistical Association*, 58, 236-244.
- YUEQING, X., L. SHUANGCHENG & C. YUNLONG, 2005, Wavelet analysis of rainfall variation in the Hebei Plain, *Science in China Ser. D Earth Sciences*, 48(12): 2241-2250.

ACKNOWLEDGEMENTS

The authors would like to thank their professor in Data Visualization, Jasmine A. Malinao, for insightful class discussions and help to improve this manuscript. Special thanks to Dr. Erniel B. Barrios for kindly providing the needed climate data. Computing software access was made available through the UP School of Statistics. Scholarship support for Ms. Tadlas is given by Engineering Research & Development for Technology (ERDT). And of course, thanks to the Creator for giving wisdom and strength to write this paper.