The development of Articulated Weather Generator model and its application in simulating future climate variability

A B Sekaranom$^{1,2}$, E Nurjani$^1$

$^1$Graduate School of Earth and Environmental Studies, Nagoya University-Japan
$^2$Faculty of Geography, Universitas Gadjah Mada, Yogyakarta, Indonesia
andungbayu@geo.ugm.ac.id

Abstract. Major areas over global tropics are vulnerable to the effects of climate change, particularly due to the high population and wide agricultural areas. The impacts of climate change have been predicted to affect higher number of hydrometeorological disasters. Over some areas, it is predicted to generate severe droughts and floods. In relation to the above issue, software called 'Articulated Weather Generator for Seasonal Climate Prediction' (AWGenSCP) has been developed to simulate the daily weather condition for the leading 6-8 months. This paper is intended to briefly explain the internal structure of the AWGenSCP algorithms. Basically, the program is a statistical medium range prediction system which is developed based on markov chain models. In addition to the general model, an articulation scheme was developed and implemented in the systems by utilizing General Circulation Model (GCM) output. This provides ability in simulating the following next seasonal condition, particularly in term of precipitation and temperature. A sample validation of the model is presented in the paper to show the model performance.

1. Introduction

Hydrometeorological dynamics that occurred over the past 1400 years show drought intensity in the past few decades due to the increasing intensity of El Nino [1], although in the long time until mid Holocene there were no significant changes [2]. Based on this reason, the impact of increasing climate variability in recent decades affects seasonal rainfall, especially with the increasing intensity of El Nino and La Nina [3]. Although the changes are significant, efforts to adapt and mitigate the impacts of climate change at the local level are currently limited, especially in developing countries [4].

Information related to weather, particularly precipitations and temperatures, play an important role as input in surface water budgets, agricultural production, and hydrometeorological disasters [4-6]. In the agriculture sector, the above variables play an important role in determining crop yields and failures [7]. Information related to precipitations and temperatures thus are very important in the decision-making process related to the crop management carried out by farmers [8]. Weather prediction systems therefore become very important to be utilized in the agricultural adaptation, particularly due to adverse impact of climate change [7,9]. Nevertheless, rainfall prediction information is very limited to be accessed at the local level because the system developed generally requires a complex computation to be carried out at the local level.

In general, weather and climate prediction system can be classified into three groups, including the short, medium and long term [10,11]. It means that there are differences in each system in addressing the specific period of the prediction whether it is hourly, daily, seasonally or annually. In short-term predictions, numerical weather prediction (NWP) can accurately predict weather condition in the next...
few days, while in the long run, general circulation models (GCM) can be used to predict climatic conditions due to increased greenhouse gases. The seasonal prediction system, when simulated using short term complex models, might requires a large computational capacity, which is difficult to implement. In contrast to complex prediction models that require very large computational resources, statistical prediction models might provide a more simplistic calculation that could be implemented in average computer.

2. Revisiting the Weather Generator algorithms

Weather Generators (WG) are statistical models that utilized to produce synthetic weather datasets based on observed data over a specific location. The weather generator models are useful in simulating weather condition over an area with limited measurements. Using the observed statistical characteristics of the weather observed at that location, a simulated data series can be produced and utilized for various purposes. In general, the WG produces stochastic weather data where the process which can be separated in two stages [12]. In the first step, the statistical properties of the distribution of specific weather variables (e.g. probability of rainfall and distribution of mean temperature) are calculated. The statistical distribution can vary from a simple normal distribution to gamma distributions, depending on the characteristics of the data. In the second step, a simulation is conducted based on Markov Chain by considering probability of occurrence of an event when a certain condition occurs after a sequence of event.

Markov Chains is a mathematical technique commonly used to perform modeling of various systems introduced by A.A. Markov in 1906 [13]. This technique can be used to estimate changes for specific variables in the future based on how it changed in the past. Therefore, this technique can be used to analyze future events by using mathematical approach, where for a given time t, the occurrence of an event called K(t) is influenced by the sequence of the previous events of K(t-1),K(t-2),...,K(t-n). Where n is number of the maximum sequence that gives probability to the current event. For a single Markov Chain (n=1) for example, it mathematically can be expressed as:

\[ K(t) = f_{\text{rand}}(P(K(t-1))) \]  

where \(f_{\text{rand}}\) is a random number generator function, equal to 1 for raining days and 0 for non-raining days. \(K(t)\) is the probability of event occurrence for time t, and \(P\) is the transitional probability from two different state (e.g. probabilities of rain days after non-rain days).

One of the earliest approaches in the WG development, the Markov Chain was implemented by [14], which is often referred as "Richardson WG". This model involves simulation of the wet and dry days using the Markov procedure, and then estimation of the amount of rainfall for each given wet day in the simulation period by using functional estimates of precipitation frequency distribution. The remaining variables (e.g. temperatures) are then calculated based on their correlation with each other and with wet or dry condition for each day. Richardson WG has been implemented in various applications in the water management, agriculture, and disasters [15].

3. Implementation of the articulation scheme to Weather Generator Model

Synthetic data generated by WG are widely implemented particularly for the following conditions; a) Simulating weather condition at locations with lack of observation; b) hydrological analysis and modeling; c) investigating weather characteristic and variability over a specific area. Although the WG appears to be useful for the above conditions, the simulated data itself generated upon assumption that the statistical properties of the data do not change over time. However, it is common that climate dynamics could give influences toward the above statistical properties. For example, El-Nino events could contribute to lower probability of rainfall over regions in Western Pacific [5,16], thus produce a shift to the transitional probability in the Markov Chain.
By adding some parameters to simulate the change of probability inside the WG algorithms, more realistic data that close to the observation could be generate. This can be obtained by, for example, by considering large scale climate dynamics that gives and influence towards the local climatology. The above backgrounds suggest the development of Articulated Weather Generator for Seasonal Climate Prediction (AWGenSCP) software. The software is intended to produce synthetic daily rainfall and temperature data as similar as general weather generator models, but with additional articulation scheme to estimate the change of probability inside the Markov Model. By implementing the above approach, a larger scope of application could be conducted, for example in simulating the seasonal changes of precipitation and temperature due to El-Nino and La-Nina events. The AWGenSCP algorithm code has been developed in JAVA programming language as well as its Graphical User Interface (GUI) by the Lab of Environmental Hydrology and Climatology, Faculty of Geography Universitas Gadjah Mada (Figure 1).

Mathematically speaking, AWGenSCP uses different algorithms for daily rainfall and temperature simulations. For the rainfall simulation a conditional probability is implemented to the rain/non-rain days using 1-chain Markov. When rain days are simulated, it will synthetically produce rainfall data (in mm/day) based on gamma distribution of daily rain rates. The basic calculation of rain probability $P_{rain}$ is as follow:

$$P_{rain}(t) = \sum_{i=1}^{n} dr(i, t) \times \left(\sum_{i=1}^{n} dr(i, t) + \sum_{i=1}^{n} dd(i, t)\right)^{-1}$$  \hspace{1cm} (2)

where

$$dr(i) = \begin{cases} 
1 & \text{if } rain > 1 \text{ mm/day} \\
0 & \text{if } rain \leq 1 \text{ mm/day} 
\end{cases}$$  \hspace{1cm} (3)

$$dd(i) = \begin{cases} 
1 & \text{if } rain \leq 1 \text{ mm/day} \\
0 & \text{if } rain > 1 \text{ mm/day} 
\end{cases}$$  \hspace{1cm} (4)

where $rain$ is daily rainfall for a given period $t$. In the AWGenSCP model, the length of period $t$ can be specified to monthly or seasonal (three monthly) climatology. By default, total daily
precipitations of less than 1 mm are considered very low and not considered as rain days rather than using absolute value of 0 mm/day.

Once the rain probability for each period i have been identified, the changes of probability $P_{\text{rain}}$ for a given time $t$ can be generated by using linear regression by a known large scale forcing $X$ (e.g. sea surface temperature (SST) anomaly) at the same time windows to the observed data constructed to build the model:

$$P_{\text{rain}}(t) = a + b \times X(t)$$

where $P_{\text{rain}}(t)$ is the transitional probability for a given period $t$, and $a$ and $b$ are intercept obtained from linear regression.

Once the probability for each time step (monthly/seasonal) has been calculated, a Markov Chain can be implemented to simulate whether the current time $t$ is raining or not, therefore rain-rate (RR) can be simulated as follow:

$$RR = \begin{cases} f_{\text{rand}}(\text{gamma}(\alpha(t), \beta(t))) & \text{if } f_{\text{rand}}(P_{\text{rain}}(t)) = 1 \\ 0 & \text{if } f_{\text{rand}}(P_{\text{rain}}(t)) = 0 \end{cases}$$

and,

$$\alpha(t) = c + d \times X(t)$$

$$\beta(t) = e + f \times X(t)$$

where $f_{\text{rand}}$ is random number generator function to pick one rain-rate value in the gamma distribution, $c, d, e, and f$ are intercept in linear regression model, and $\alpha$ and $\beta$ are shape and scale parameters in gamma distribution.

For temperature model, a similar equation to the above is implemented, except for the unconditional probability and assumption of normal distribution rather than gamma distribution.

$$temp(t) = f_{\text{rand}}(\text{normaldist}(\text{avg}(t), \text{std}(t)))$$

and,

$$\text{avg}(t) = g + h \times X(t)$$

$$\text{std}(t) = i + j \times X(t)$$

where $\text{avg}$ and $\text{std}$ are mean temperature and standard deviation for each time step $t$ (monthly/seasonal), and $g, h, i, and j$ are linear regression intercept for large scale variable $X$.

4. Example implementation of AWGenSCP program

As an example of the application and validation of the AWGenSCP program, a comparison of the model simulation and observation is conducted. In this paper, a case study analysis has been conducted over Kebumen Area, Central Java-Indonesia, located at 7°27’ - 7°50’ S and 109°22’ - 109°50’ E. Kebumen has a total area of 1581.11 km². The land cover in this area is dominated by agricultural land, to more than 60% of the total area. Therefore, the community in general still depend their livelihoods on agricultural land. In 2005, drought struck Kebumen and causing damage around
10,838 hectares of rice field that are spread throughout the district. In 2008, drought also occurred in Kebumen. The decrease in the discharge of water from river such as Luk Ulo, Kalibanda and Telomoyo River, resulting difficulties of the residents to access clean water. Due to this condition, about 2,000 acres of farmland also affected and crop failure cannot be avoided. From the above description, it can be concluded that the area is prone to the changes on climate variability. Therefore, information about the future climatology variables becomes important in implementing climate adaptation.

To simulate the daily precipitation, daily precipitation data from Indonesian Bureau of Meteorology and Geophysics (BMKG) from 1960-2010 was utilized from Kebumen Weather Station. In this analysis, the influence of El-Nino/La-Nina is addressed. In line to the above purpose, long term monthly observation of Sea Surface Temperature (SST) data over Nino 3.4 region has been obtained from NOAA ESRL database (https://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/). The SST data is used as large-scale predictor variable in the AWGenSCP input. By comparing the monthly/seasonal SST anomaly data and the changes of rain probability, a linear regression model of probability changes could be conducted. All the above process could be simply conducted in AWGenSCP program, including the SST data download. In the AWGenSCP program, a function to generate future simulation, monthly SST prediction from NCEP coupled forecast system model version 2 (CFSv2) also available from (http://www.cpc.ncep.noaa.gov/products/CFSv2/CFSv2seasonal.shtml). However, for the analysis presented in this paper, only result from the mean ensemble model is shown. Furthermore, only observed SST data from NOAA ESRL are utilized in back testing the program.

To compare the simulation result with the observed data, comparison of mean monthly rainfall is conducted for dry season and rainy season from 1998 to 2010. Comparison of the dry season consists of March-April-May (MAM) and June-July-August (JJA). The result of the simulation and the observed value for dry season is shown in Figure 3. The result shows that the simulated rainfall data follows the pattern of the observed precipitation data, although sometimes it overestimate/underestimate the actual precipitation. Based on statistical analysis, the correlation between simulated and observed data is 0.852, while the $R^2$ is 0.746. The standard error of the average monthly rainfall is 144 mm/month.

![Image](image_url)
To compare the simulation result with the observed data during rainy season, a similar comparison of mean monthly rainfall is conducted from 1998 to 2010. Comparison of the rainy season consists of September-October-November (SON) and December-January-February (DJF). The result of the simulation and the observed value for the rainy season is shown in Figure 4. The result shows that the simulated rainfall data also match with the actual data, although the result is less consistent compared to dry season simulation. This condition implies less robust statistical value for rainy season than dry season. The correlation between simulated and observed data is 0.842, while the R² is 0.710. The standard error of the average monthly rainfall is 284 mm/month, almost twice compared to the dry season rainfall.

In addition to mean monthly rainfall comparison for each season, the simulated precipitation data is also compared with the actual data in term of rainy season onset from 1998-2010. The onset analysis marks the number of days to indicate beginning of rainy season. It is calculated based on 100 mm cumulative rainfall threshold by adding up daily rain-rates since the beginning of dry season. The result is shown in Figure 5. This plot indicates that the model could produce a similar pattern compared to the actual onset days with a couple of day biases. It can be observed that earlier onset is simulated in 1998 and 2010, which remarks La-Nina years. The late onset is also simulated in 2002 and 2006; with give remarks of higher Southern Oscillation Index (SOI) than the averages.
5. Summary

This paper briefly explains about the development of AWGenSCP program and algorithm behind it. An example simulation result over the case study area has been presented in this paper. The result indicates that the model could produce simulated data that are comparable to the actual data. This is particularly due to the characteristic of local condition of the study area that is sensitive to El-Nino/La-Nina events. In general, the simulation over rainy season is less robust than dry season. Although appear as less robust, in general the precipitation in rainy season less varied than the dry season. This condition yields difficult estimation of the probability of rainfall and its distribution. On another hand, dry season rainfall are highly influenced by El-Nino events, where total monthly precipitation reaching less than 50 mm/month during the events and could be higher than 500 mm/month during La-Nina.

Results of this study have shown the statistical robustness of the program in simulating daily precipitation data, which is based on the past rainfall observation and large-scale circulations that gives an influence on the statistical properties of the rainfall over a given area of study. This is particularly useful in estimating how much rainfall and when the large amount of rainfall will occur in term of probability. The above information is crucial for the local community in the case study area, particularly for those who live in agricultural sector. Currently, the agricultural sector in the study area are dominated by rain-fed systems and only small portions of them have irrigation channels. Crop productions in the study area highly depends on the amount of rainfall. Delayed precipitation onset due to El-Nino often produce crop failure in the study area. Therefore, the development of this system could help predicting the beginning of rainy season, and could help reduce crop failures. Information of the amount of rainfall can also be simulated by this system, which are useful in reducing crop failures due to flood inundation if excessive rain events occurred. The above information also could give influences on farmer decision regarding the type of crop and planting/harvesting time, which is particularly important due to increasing climate variability.

In the current state, we are looking forward for the development of possible implementation of the program in the decision-making process as a form of climate change mitigation and adaptation strategies. There are many ways in the AWGenSCP program to be implemented in hydrometeorological disaster risk reduction. For example, it can be utilized to simulating heavy rainfall probabilities over a specific area. It is useful particularly over the tropics due to high intensity of heavy rain events and often associated with disaster [17]. Another example application of the AWGenSCP program is for producing inputs for flood modelling. The AWGenSCP program could produce synthetic data for estimating area of inundation based on probability, while the observation
data are very limited. This could be very useful for delta cities in developing countries, in which most of them are prone to floods.

References

[1] Rodysill J R et al 2012 A paleolimnological record of rainfall and drought from East Java, Indonesia during the last 1,400 years. *J. Paleolimnol.* 47: 125–139.
[2] Azmy K Edinger E Lundberg J and Diegor W 2010 Sea level and paleotemperature records from a mid-Holocene reef on the North coast of Java, Indonesia. *Int. J. Earth Sci* 9: 231–244
[3] Kripalani R H and Kulkarni A 1997 Rainfall variability over south-east Asia-connections with Indian monsoon and ENSO extremes: New perspectives. *Int. J. Clim.* 17: 1155–1168.
[4] Marfai M A Sekaranom A B and Cahyadi 2015 A Profiles of marine notches in the Baron coastal area—Indonesia. *Arab. J. Geosci.* 8
[5] Ward P J et al 2014 Strong influence of El Nino Southern Oscillation on flood risk around the world *Proc. Natl. Acad. Sci.* 111: 15659–15664.
[6] Sekaranom A B and Masunaga H 2017 Comparison of TRMM-derived rainfall products for general and extreme rains over the maritime continent. *J. Appl. Meteorol. Climatol.* 56: 1867–1881.
[7] Sudibyakto H A Gunawan D Nurjani E and Sekaranom 2016 A B A Projection on Climate Change Impact towards Meteorological Droughts over Java Island, Indonesia. in *International Conference on Agriculture, Food Science, Dynamics, Natural Resource Management and Environmental The Technology, People and Sustainable Development* 246–249.
[8] Sekaranom A B Nurjani E Hadi M P and Marfai M A 2018 Comparison of TRMM Precipitation Satellite Data over Central Java Region-Indonesia. *Quaest. Geogr.* 37: 55–72.
[9] Howden S M et al 2007 Adapting agriculture to climate change. *Proc. Natl. Acad. Sci.* 104: 19691–19696.
[10] Hong S Y and Lee J W 2009 Assessment of the WRF model in reproducing a flash-flood heavy rainfall event over Korea. *Atmos. Res.* 93: 818–831.
[11] Chen J Brissette F P and Leconte R 2010 A daily stochastic weather generator for preserving low-frequency of climate variability. *J. Hydrol.* 388: 480–490.
[12] Hutchinson M F 1995 Stochastic space-time weather models from ground-based data. *Agric. For. Meteorol.* 73: 237–264.
[13] Gilks W R Richardson S and Spiegelhalter D 1995 *Markov chain Monte Carlo in practice* (CRC press).
[14] Richardson C W 1981 Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resour. Res.* 17: 182–190.
[15] Racsko P Szeidl L and Semenov M 1991 A serial approach to local stochastic weather models. *Ecol. Modell.* 57: 27–41.
[16] Sekaranom A B Nurjani E and Puijastuti I 2018 Cloud structure evolution of heavy rain events from the East-West Pacific Ocean: a combined global observation analysis. in *IOP Conference Series: Earth and Environmental Science* 148: 12006.
[17] Sekaranom A B and Masunaga H 2018 Origins of heavy precipitation biases in the TRMM PR and TMI products assessed with CloudSat and reanalysis data. *J. Appl. Meteorol. Climatol.* DOI: 10.1175/JAMC-D-18-0011.1.
[18] Marfai M A Sekaranom A B Ward P 2015 Community responses and adaptation strategies toward flood hazard in Jakarta, Indonesia. *Natural hazards* 75 (2): 1127-1144.