Water Resources Management

Supplementary material for

Time-varying Discrete Hedging Rules for Drought Contingency Plan Considering Long-Range Dependency in Streamflow

Seung Beom Seo[1*,2]

Young-Oh Kim[3]

Shin-Uk Kang[4]

[1]Institute of Engineering Research, Seoul National Univ., Seoul 08826, South Korea (ORCiD: 0000-0001-9819-6555)

[2]Korea Environment Institute, Sejong 30147, South Korea

[3]Professor, Department of Civil and Environmental Engineering, Seoul National Univ., Seoul 08826, South Korea

[4]Researcher, K-water Convergence Institute, K-Water, Daejeon 34045, Korea, South Korea

* Corresponding author: Seung Beom Seo (sbseo7@snu.ac.kr)

Section 2.2 Stochastic streamflow generation

Figure S1 illustrates a flow diagram for the generation of a synthetic monthly streamflow scheme, which was developed in this study.

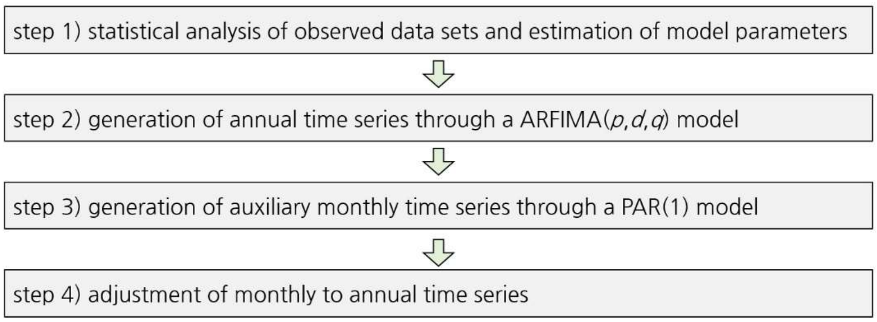

Figure S1 A flow diagram of stochastic streamflow generation approach.

Annual series generation: ARFIMA model

Many studies have suggested that the use of the fractionally differenced autoregressive integrated moving average (ARIMA) model, which uses a fractional difference operator rather than an integer operator, can better account for the long-range dependency (LRD) phenomenon (Granger and Joyeux 1980). Hosking (1981) defined an extension of the ARIMA model which allows for the possibility of stationary long-memory models. In Hosking and Granger's ARFIMA(p,q,d) process, X_t is defined as eq. (S1) (Beran, 1994; Beran et al. 2013)

$$\Phi(B)(1-B)^d X_t = \Psi(B)\varepsilon_t \tag{S1}$$

where $d \in (-0.5, 0.5)$, ε_t is a white noise term, and the backshift operator (B) is defined as $BX_t = X_{t-1}$. The polynomials $\Phi = 1 - \sum_{k=1}^{p} \emptyset_k z^k$ and $\Psi = 1 + \sum_{k=1}^{q} \psi_k z^k$ describe the autoregressive and moving average terms, respectively (Graves et al. 2017).

Thus, the autoregressive fractionally integrated moving average (ARFIMA) process, ARFIMA(p,q,d), is a naturally generalized form of the standard ARIMA or autoregressive moving-average (ARMA) processes. The difference coefficient, d, is a parameter that should be estimated first (Reisen et al. 2001). The value of d has a closed relationship with the Hurst parameter, H (Reisen et al. 2001; Sheng et al. 2011) as shown in equation (S2):

$$d = H - 1/2 \tag{S2}$$

The level of the LRD process can be measured or indicated by the estimated Hurst parameter, also known as the Hurst exponent (Hurst, 1951). The value of H varies between 0 and 1. When $H = 0.5$, the time series does not have statistical dependence. When $H < 0.5$, the time series contains an anti-persistent process. On the other hand, when $H > 0.5$, the time series is a positively correlated so that the time series has the LRD process. In this regard, when the value of d varies between 0 and 0.5, the time series is considered to follow the LRD process. Note that the closer the d value is to 0.5, the stronger the LRD is realized.

For annual streamflow series generation, the values of parameter p and q are first obtained from the observed annual streamflow series. By perturbing the value of d from 0 to 0.5, a wide range of LRD can then be realized using the ARFIMA(p,q,d) model. Since the ARFIMA(p,q,d) process is widely used in modeling LRD time series, most statistical analysis programs have been embedded with ARFIMA models (Liu et al. 2017). In this study, the ARFIMA_SIM function included in Statistics and Machine Learning Toolbox of MATLAB® are used for the simulation of ARFIMA(p,q,d) process.

Temporal disaggregation: PAR model and linear adjustment

In order to temporally disaggregate the annual series to a monthly scale, auxiliary monthly series are initially generated using the periodic autoregressive (PAR) model with no reference to the pre-generated annual series. The PAR(1) model was adopted in this study since PAR(1) is the most parsimonious among linear stochastic models (Efstratiadis et al. 2014). The PAR(1) model is given by the recursive equation:

$$\tilde{y}_{t,s} = \emptyset_s \tilde{y}_{t,s-1} + v_{t,s} \tag{S3}$$

where $\tilde{y}_{t,s}$ is stochastic process in year t and month s (s = 1…12), which represents the auxiliary variable that is to be adjusted to the annual synthetic data next (section 2.2.1). For each month s, this process is normally distributed with zero mean and variance $\sigma_s^2(Y)$. The $v_{t,s}$ is the uncorrelated noise term for which each month is normally distributed with zero mean and variance $\sigma_s^2(\varepsilon)$. The $\emptyset_s$ is the periodic autoregressive parameter for each month s.

Before calibrating model parameters, the original historic data sets were transformed to fit a normal distribution. In this study, the Box-Cox transformation technique (Box and Cox 1964; Sakia 1992) was employed to remove the skewness inherent in streamflow data sets. After synthetic monthly time series were generated from the transformed data sets, they were transformed back to the original space.

The PAR(1) model defined by the equation (S3) is proper for sequential generation of the auto-correlated monthly series $\tilde{y}_{t,s}$, but it cannot account for the annual series x_t, which were generated in advance using the ARFIMA model. Two time series data sets are not consistent since the annual sum of $\tilde{y}_{t,s}$, denoted as $\tilde{x}_t$, is not equal to the corresponding vector of annual variables, x_t. In order to show consistency, the adjustment procedure introduced by Koutsoyiannis and Manetas (1996) was employed in this study. The transforming equation is presented as equation (S4):

$$y_{t,s} = \tilde{y}_{t,s} + H_s(x_t - \tilde{x}_t) \tag{S4}$$

where H_s is a matrix of monthly parameters as estimated by the following equation:

$$H_s = \mathrm{Cov}[y_{t,s}, x_t]\{\mathrm{Cov}[x_t, x_t]\}^{-1} \quad \text{(S5)}$$

In case of a single variable, a linear transformation is employed in which the departure $\Delta x_t = (x_t - \tilde{x}_t)$ of the additive property is distributed to each monthly variable in which the covariance of this monthly variable is proportional with the annual variable (Efstratiadis et al. 2014). However, there is a limitation to this adjustment approach wherein highly negative departures, Δx_t, may result in negative values in the adjusted variables. In order to remedy this limitation, a simple repetitive process based on conditional sampling, as proposed by Koutsoyiannis and Manetas (1996) can be used. This procedure aims at minimizing the distance, Δx_t, by repeating the generation for each year's variable until Δx_t becomes lower than the accepted limit, which is 1% of the annual standard deviation of the associated variable.

Section 3.1 Application site and data sets

Since 2014, the Chungcheongnam-do province suffered from a severe drought that has persisted for several years. In 2015, the total rainfall was 800 mm, which amounts to just 60% of the annual mean rainfall of this region, and Boryeong Dam's water level remained around the dead storage level for 135 days. This was the most extreme drought since the dam had been built. Consequently, the government imposed a restriction on public water supply from October 2015 to February 2016. Hence, approximately 480,000 people suffered from a drink water shortage while an increase in the salinity of the reclaimed land due to lack of agricultural water caused crop damage in this region. Hence, the Boryeong Dam region and its drought contingency plan are selected for this case study. The observed streamflow (Boryung Dam inflow) data sets from 1971 to 2016 were provided from the National Drought Information Analysis Centre of K-water. The original data sets formed a daily series, so they were aggregated into monthly and annual series in order to calibrate the model parameters for PAR and ARFIMA models, respectively. Figure S2 depicts the monthly mean streamflow and the standard deviation derived from the observed monthly inflow data sets concerning Boryung Dam.

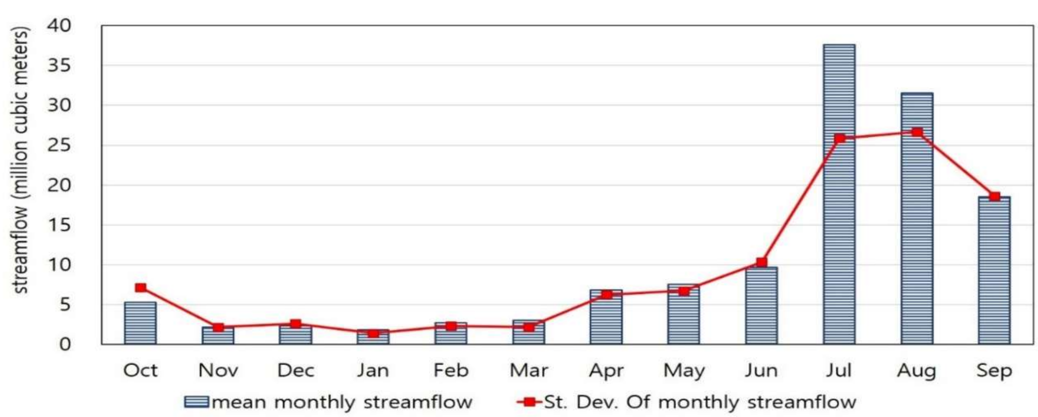

Figure S2 Monthly mean streamflow and standard deviation of Boryung Dam's inflow series.

Section 4.1 Synthetic streamflow generation

Figure S3 presents the four annual statistics of the synthetic streamflow series, which had an annual mean value decrease of 20%. As shown in figure S3a, the annual mean values were generally well matched with 80% of the historical statistics. On the other hand, the values of standard deviation of annual mean streamflow were slightly underestimated compared to historical statistics due to the overall decrease in synthetic streamflow values. Nonetheless, as shown in Figure S3c and d, the estimated H values were well-matched with the target values, and the lag-1 correlation values were also well-realized.

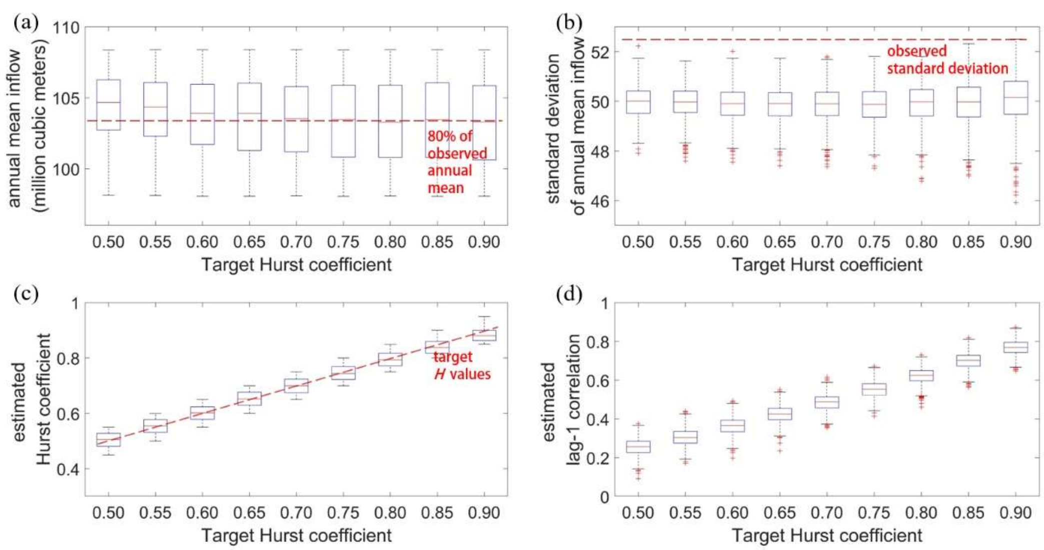

Figure S3 Annual statistics of the synthetically generated streamflow series: 20% decrease in the annual mean value: (a) annual mean, (b) standard deviation, (c) Hurst coefficient, and (d) lag-1 correlation.

Figure S4 depicts the monthly statistics of the synthetic streamflow series of the two cases, $H = 0.5$ and 0.8, with no changes in the annual mean streamflow. As shown in Figure S4, the three elementary statistical properties in both cases were well-matched with the historical statistics. Nonetheless, values of the monthly lag-1 correlation in October were underestimated as close to zero. This is because auxiliary monthly streamflow were generated from October to September. However, its impact would be negligible since the actual lag-1 correlation, which is estimated from historical data sets, was statistically not significant. Except this minor concern, all of the elementary statistical properties for both annual and monthly values were successfully realized across all 81 cases using the proposed methodology.

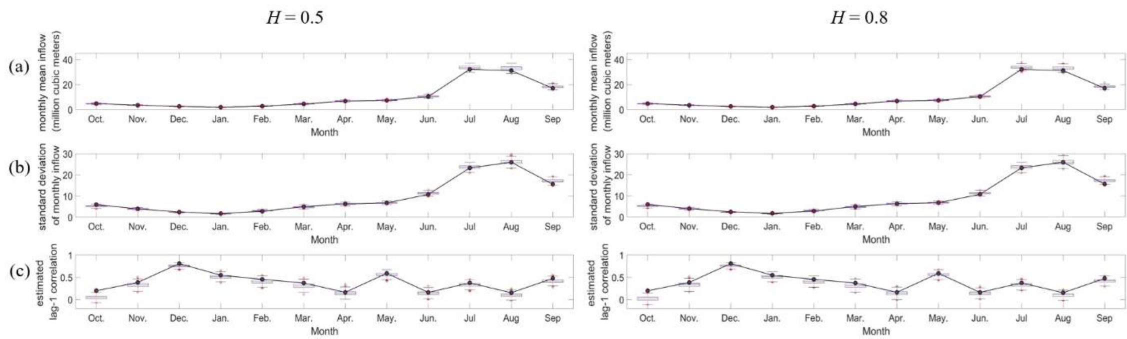

Figure S4 Monthly statistics of the synthetically generated streamflow series: two cases of *H* = 0.5 and *H* = 0.8 with no changes in the annual mean value: (a) monthly mean, (b) standard deviation, and (c) lag-1 correlation.

References

Beran, J., *Statistics for Long Memory Processes*, New York, USA: Chapman & Hall, 1994.

Beran, J., Feng, Y., Ghosh, S. and Kulik, R., *Long Memory Processes*, Heidelberg, Germany: Springer, 2013.

Box, G. E. P., and Cox, D. R., An analysis of transformations, *Journal of the Royal Statistical Society*, 1964, vol. 26, pp. 211–252.

Efstratiadis, A., Dialynas, Y. G., Kozanis, S., and Koutsoyiannis, D., A multivariate stochastic model for the generation of synthetic time series at multiple time scales reproducing long-term persistence, *Environmental Modelling and Software*, 2014, vol. 62, pp. 139–152. https://doi.org/10.1016/j.envsoft.2014.08.017

Granger, C. W., and Joyeux, R., An introduction to long-memory time series models and fractional differencing, *Journal of Time Series Analysis*, 1980, vol. 1, no. 1, pp. 15–29.

Graves, T., Gramacy, R., Watkins, N., and Franzke, C., A brief history of long memory: Hurst, Mandelbrot and the road to ARFIMA, 1951-1980, *Entropy*, 2017, vol. 19, no. 9, pp. 1–21. https://doi.org/10.3390/e19090437

Hosking, J. R., Fractional differencing, *Biometrika*, 1981, vol. 68, no. 1, pp. 165–176.

Hurst, H.E., Long-term storage capacity of reservoirs, *Transactions of the American Society of Civil Engineers*, 1951, 116, pp. 770–808.

Koutsoyiannis, D., and Manetas, A., Simple disaggregation by accurate adjusting procedures, *Water Resources Research*, 1996, vol. 32, no. 7, pp. 2105–2117.

Liu, K., Chen, Y., and Zhang, X. An Evaluation of ARFIMA (Autoregressive Fractional Integral Moving Average) Programs. *Axioms*, 2017, vol. 6, no. 4, p. 16.

https://doi.org/10.3390/axioms6020016

Reisen, V., and Abraham, B., and Lopes, S., Estimation of parameters in ARFIMA processes: A simulation study, *Communication in Statistics: Simulation and Computation*, 2001, vol. 30, pp. 787–803.

Sakia, R. M., The Box-Cox transformation technique: a review, *The statistician*, 1992, pp. 169–178.

Sheng, H., Chen, Y., and Qiu, T., *Fractional Processes and Fractional-Order Signal Processing: Techniques and Applications*, New York, USA: Springer Science & Business Media, 2011.