



GLOBAL JOURNAL OF ADVANCED RESEARCH
(Scholarly Peer Review Publishing System)

ESTIMATION OF CRITICAL STREAMFLOW DISCHARGE LEVEL USING NONPARAMETRIC QUANTILE REGRESSION MODEL

Dr. Francis Kiarie

PhD,
Department of Management
Science, Kenyatta University,
P.O Box 43844-00100 Nairobi,
Kenya.
spssnairobi@gmail.com

ABSTRACT

Various parametric models have been designed to analyze volatility in river flow time series data. For maximum likelihood estimation these parametric methods assume a known conditional distribution. This paper considers the problem of nonparametric estimation of critical streamflow discharge levels of a river regime based on quantile regression methodology of Koenker and Basset (1978). In particular, the paper demonstrates the use of kernel estimators for conditional quantiles resulting from a kernel estimation of conditional distribution function. It is finally proved that the estimate of the nonparametric quantile function is consistent and asymptotically normally distributed and under suitable conditions, the estimator converges uniformly with an appropriate rate.

Key words: Conditional quantile, kernel estimate, quantile autoregression, consistency, asymptotic normality, critical discharge level.

1. INTRODUCTION

Globally, floods impact an estimated 520+ million people per year, resulting in estimates of up to 25,000 annual deaths, extensive homelessness, disaster-induced diseases, crop and livestock damage and other serious harm (UNU, 2004). One way of reducing losses due to floods is by use of flood early warning systems (FEWS). Such a system consists of streamflow monitoring and forecasting as well as public information system. Various methods are employed in hydrological stream flow monitoring and forecasting. For hydrological applications, models are usually based on regression relationships derived from paired catchment and catchment treatment experiments, Muthusi (2004).

The conventional method to estimate the conditional distribution function $F(Y_t|X_t)$ is to assume that $Y_t|X_t$ follows a particular distribution and then estimate its parameters. However, such specification may lead to wrong inferences and therefore, in this paper, we focus on quantile regression methodology introduced in Koenker and Basset (1978). This is a flexible method that requires no strict assumptions on moments and distribution of the underlying process. Instead of assuming that $m(X)$ is a conditional mean, it is assumed to be θ -th conditional quantile and denoted as $m_\theta(X_t)$.

In this paper, I use nonparametric approach to model $m_\theta(X_t)$. I first estimate, non-parametrically, the conditional distribution of Y_t given X_t and then invert it at θ -level of probability to get conditional θ -th quantile estimate as in Franke and Mwita (2003).

2. THE STUDY MODEL (CRITICAL STREAMFLOW DISCHARGE LEVEL)

Assume that the underlying hydrological process of interest is of the form

$$Y_t = m_\theta(X_t) + \epsilon_t \quad (2.1)$$

where Y_t is the streamflow discharge at time t measured in cubic meters per second (cubecs). The variable $X_t = (Y_{t-1}, \dots, Y_{t-d})$ is a d -dimensional vector consisting of the past observations of Y_t . The conditional quantile function $m_\theta(X_t)$ is the streamflow discharge level at $\theta \in (0, 1)$. The errors, ϵ_t , are assumed to be zero quantile with some scale function σ_θ . Therefore the errors may also include heteroscedastic cases, see Franke and Mwita (2003), Mwita (2005) among others.

Here, model (2.1) can be viewed as a robust generalization of Autoregressive (AR) – Autoregression Conditional Heteroscedastic(ARCH) models introduced in Weiss(1984) and their nonparametric generalizations reviewed by Hardle (1989), see Franke and Mwita (2003) and Mwita (2005) for more details.

If we choose $X_t = (Y_{t-1}, \dots, Y_{t-d}, U_{t-1})$ where the random vector U_t consists of observations from other time series such as soil moisture budget(SMB), precipitation, evapotranspiration, El Nino Southern Oscillations(ENSO), Pacific Decadal Oscillations(PDO), then model (2.1) would become a quantile autoregressive model with exogenous components.

2.1 Estimation of critical streamflow discharge level

We consider the model (2.1), and define a true conditional distribution function $F_x(y)$ of Y_t given $X_t = x$ as

$$F_x(y) = P(Y_t \leq y \mid X_t = x) = E[I_{t,y} \mid X_t = x] \quad (2.2)$$

where $I_{t,y} = I\{Y_t \leq y\}$ is an indicator function with $\Pr(Y_t \leq y \mid X_t = x) = 1$ and 0 otherwise.

For any $\theta \in (0, 1)$, we define the true critical streamflow discharge level as

$$m_\theta(x) = \inf\{y \in \mathbb{R} \mid F_x(y) \geq \theta\} \quad (2.3)$$

The distribution function in (2.2) can be estimated by the Nadaraya (1964) and Watson (1964) estimator as

$$\hat{F}_x(y) = \frac{\sum_{t=1}^n K_h(x - X_t) I_{t,y}}{\sum_{t=1}^n K_h(x - X_t)} \quad (2.4)$$

where $K(u)$ is a d -dimensional kernel and $K_h(u) = h^{-d} K(u/h)$ is the rescaled kernel, see Franke and Mwita (2003) and Mwita(2005). Therefore the kernel estimator for the critical streamflow discharge level is given by

$$\hat{m}_\theta(x) = \inf\{\mathbb{R} \mid \hat{F}_x(y) \geq \theta\} \equiv \hat{F}_x^{-1}(\theta) \quad (2.5)$$

where $\hat{F}_x^{-1}(\theta)$ denotes the usual generalized inverse of the distribution function $\hat{F}_x(y)$ which is a pure jump function of y .

2.2 Asymptotic normality

Assume that the time series (Y_t, X_t) satisfies α -mixing conditions. According to Masry and Tjøstheim (1995, 1997), both ARCH processes and nonlinear additive autoregressive models with exogenous variables are stationary and α -mixing under some mild conditions. As Franke and Mwita (2003) demonstrated, if we choose $X_t = Y_{t-d}$ in (2.1) and assuming the time series Y_t is α -mixing, we get an example of a quantile autoregressive process for which (Y_t, X_t) and $I_{t,y}$ in (2.4) are α -mixing as well.

The following assumptions are necessary for proving asymptotic normality of

$$\hat{m}_\theta(x)$$

Henceforth, $g(x)$ denotes the stationary probability density of X_t at point x .

(A1) For all $u \in \mathbb{R}$

- (i) $K(u) \geq 0$
- (ii) K is Lipschitz continuous i.e. $|K(u) - K(v)| \leq C_k |u - v|$, for all $C_k, u, v \in \mathbb{R}$ and $C_k > 0$
- (iii) $|K(u)| \leq K_\infty$, with K_∞ being a constant
- (iv) $\int K(u) du = 1$, $\int u K(u) du = 0$ and $\int \|u\|^2 k(u) du < \infty$

(A2) For all y, x satisfying $0 < F_x(y) < 1$, $g(x) > 0$

- (i) $F_x(y)$ and $g(x)$ are twice continuously differentiable and bounded in y, x
- (ii) $f_x(m_\theta(x)) > 0$, for all x .

(A3) The process (Y_t, X_t) is stationary and α -mixing with mixing coefficients satisfying $\alpha(s) = O(s^{-(2+\delta)})$ for some $\delta > 0$, $n \geq 1$, and $\{s_n\}$ is an increasing sequence of positive integers.

The consistency and asymptotic normality properties of $\hat{m}_\theta(x)$, and their proofs can be found in Franke and Mwita (2003).

Here, we only state the theorems.

Theorem 3.1

Assume that (A1)- (A3) hold. As $n \rightarrow \infty$, let the sequence of bandwidths $h > 0$ converge to 0 such that $nh^d \rightarrow \infty$. Then the conditional quantile estimator is consistent, $\hat{m}_\theta(x) \xrightarrow{p} m_\theta(x)$

that is

$$E[\hat{m}_\theta(x)] - m_\theta(x) = h^2 B_m(m_\theta(x)) + O(h^2) \text{ where } B_m(y) = -\frac{B(Y)}{f_x(Y)} \tag{2.6}$$

Further if, the bandwidths are chosen such that nh^{d+4} is either 1 or converges to 0, then $\hat{m}_\theta(x)$ is asymptotically normal,

$$\sqrt{nh^d}(\hat{m}_\theta(x) - m(x) - h^2 B_m(m(x))) \xrightarrow{D} N\left(0, \frac{V(\hat{m}_\theta(x))}{f_x^2(m_\theta(x))}\right) \tag{2.7}$$

where, $B(y)$ and $V2(y)$ are the bias and variance expansion for the conditional distribution estimator in (2.4)

2.3 Uniform consistency and uniform convergence

For uniform consistency and uniform convergence of the quantile autoregressive estimate, Franke and Mwita(2003) first establishes the uniform consistency of the Nadaraya-Watson kernel estimate (2.4). For this purpose, the following conditions are imposed.

- (B1) for some compact set G , there are $\varepsilon > 0, \gamma > 0$, such that $g(x) \geq \gamma$ for all x in the ε -neighborhood $\{x; \|x-u\| < \varepsilon \text{ for some } u \in G\}$ of G .
 - (B2) (Y_t, X_t) is stationary and α -mixing with mixing coefficients $\alpha(n), n \geq 1$, and there is an increasing sequence $s_n, n \geq 1$, of positive integers such that for some finite A

$$(n/s_n) \alpha(2s_n/(3n)(s_n)) \leq A, 1 \leq s_n \leq n/2 \text{ for all } n \geq 1.$$
- Uniform consistency and uniform rate of convergence properties of the estimator under the regularity conditions in Franke and Mwita, (2003) are given in Theorem 3.2.

Theorem 3.2

Assume (A1), (A2), (B1), and (B2). If, as $n \rightarrow \infty$, the bandwidth $h \rightarrow 0$ such that

$$\hat{S}_n = nh^d (s_n \log n)^{-1} \rightarrow \infty$$

then (3.2.4) is uniformly consistent on G in the strong sense. That is, for $x \in G$

$$\sup_{x \in G} |\hat{F}_x(y) - F_x(y)| \rightarrow 0 \text{ a.s.}$$

2.4 Summary

In this section, we have shown that the estimate of our nonparametric quantile function is consistent and asymptotically normally distributed, and under suitable conditions, the estimator converges uniformly with an appropriate rate. The asymptotic normality property is used to construct the required confidence intervals for our estimator. These are strong properties that significantly imply sufficiency of our estimator is accurate estimation of the critical streamflow discharge level.

3. REAL DATA RESULTS

The application of our estimator was performed with data from the gauge at River Nyando, in Western Kenya, (River Station No. IGD03) in the wider Nyando Basin, located at 35.2 oE longitude and -0.1oS latitude and covering an area of 3,587 km². The drainage area downstream of the outlet of the catchment (IGD03) was found to accommodate all the discharge in the river channel. Flooding is experienced starting from Ahero plains, down to Lake Victoria through KUSA swamps. For this reason, monthly maximum streamflow data from gauging station IGD03 for the period 1970 – 1997 was used for calibrating the model. Also, the twenty-seven year period was considered long enough to capture diverse weather conditions, thus making the model to be a good representative of the basin.

Figure 3.1 gives the daily streamflow hydrograph of ground station gauging data for twenty-seven years, from 1970 – 1997. From the hydrograph, it is clear that the river regime experiences both peak and extremely high flows which are responsible for flood inundations experienced in flood plain areas of the Nyando basin.

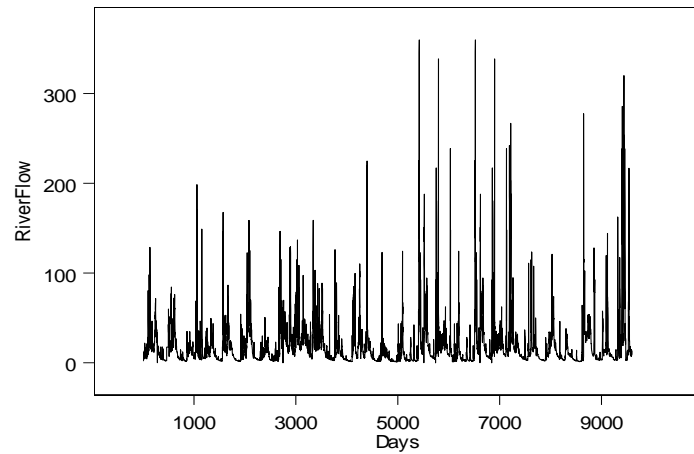


Figure 3.1 Daily streamflow discharges for River Nyando (1970 – 1997) Station (IGD03)

Considering the critical streamflow discharge level to be our target variable, we first present hydrograph for monthly maximum streamflow for the period 1970 – 1997 in Figure 3.2.

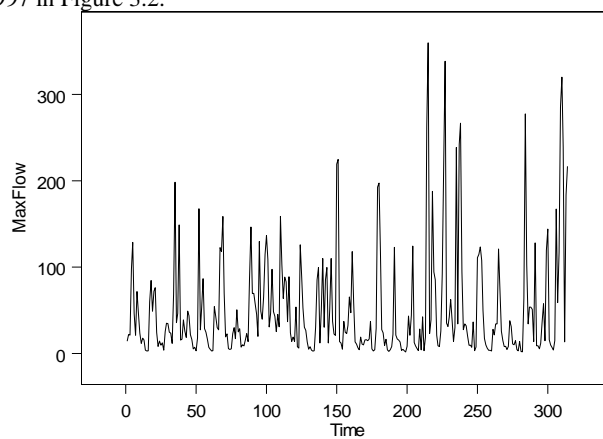


Figure 3.2 Monthly maximum streamflow discharges for River Nyando (1970 – 1997) for Station (IGD03)

The hydrograph of figure 5.2 shows that the river pattern of low flows, peak flows and extremely high flows is preserved by the monthly maximum streamflow time series of our ground station gauging data.

Figure 3.3 gives the volatility of the monthly maximum streamflow discharges. The hydrograph of these deviations depict the turbulence experienced by the Nyando River regime with an observable increase in trend.

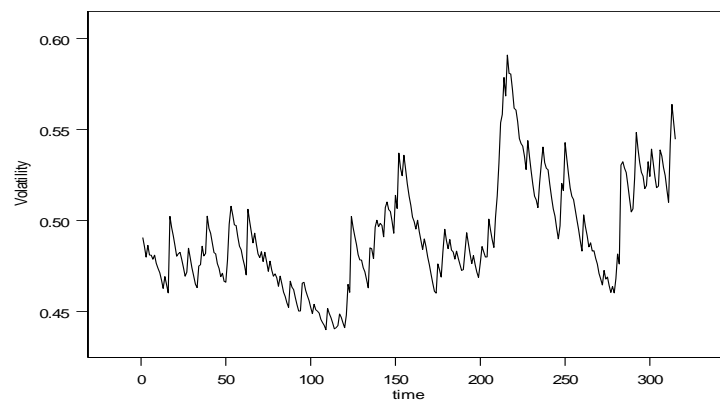


Figure 3.3 Volatility of monthly maximum streamflow discharge for River Nyando.

Figure 3.4 gives the monthly maximum streamflow discharge levels together with 0.95 and 0.99 conditional quantiles respectively.

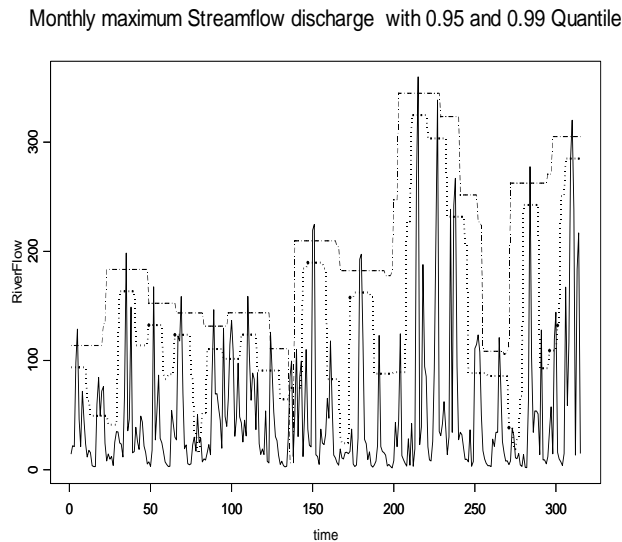


Figure 3.4 Monthly maximum streamflow discharges with 0.95 and 0.99 quantiles.

The dotted curve represents the 0.95 conditional quantile while the dashed curve represents the 0.99 conditional quantile. Streamflow discharge levels above the 0.95-quantile curve represent critical streamflow discharge levels responsible for flood inundations at 95% confidence level. Such a level calls for some site-specific operational instructions to be issued by authorities monitoring the river catchment. The instructions may include shutting of floodgates and other engineering measures. Discharges above the 0.99 quantile curve represent extreme river flow levels. Such levels call for flood control teams to respond to imminent flood conditions and operate a warning system for the public as well as industries.

3.1 Model validation

To demonstrate that the study model produced good estimates, a model validity test using the Basler Amplel method of Backtesting was performed. This method is mainly applied in financial modeling. However, its basic principles do apply to other applications as well.

The Basler Ampel method suggests that we define a Bernoulli- distributed series of random variables B_i , such that $B_i = 1$ if $Y_i > m_\theta(x_i)$ or $B_i = 0$ if $Y_i < m_\theta(x_i)$ $i \in \{1, 2, \dots, n\}$ where n is the number of backtesting points. Y_i is the streamflow discharge level for the i -th month. $m_\theta(x_i)$ is the i -th month θ -th quantile level. This number is usually set at 250 or 500, but can be selected arbitrarily.

Using this model validation method, the study model was tested using a sample of size 250 data points. At both 99% and 95% confidence levels, the model results (4 data points at 99% and 16 data points at 95%) lied within the green zone of acceptance. Consequently, the study model was considered adequate in the estimations of critical stream flow discharge levels.

4. REFERENCES

- [1] Abberger, K. (1997). Quantile smoothing in Financial Time Series. Statistical papers Vol38, 125-148.
- [2] Acreman, M. (1994). The Role of Artificial Flooding in Integrated Development of River basins in Africa. Journal of Hydrology, Vol78, 23-39.
- [3] Box, G.E.P. and Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control. J. Clim., Vol14, 2105-2128.
- [4] East African Standard, (15th May, 2003). <http://www.eastandard.net/>
- [5] Franke, J. and Mwita, P. (2003). Nonparametric estimates for Conditional Quantiles of Time Series. Report in Economic Mathematics, Nr 87. Technical University of Kaiserslautern.
- [6] Goswami, D.C. (2001). Flood Forecasting in the Brahmaputra River, India: A Case study, Department of Environmental science, Nr 222. Guahati University, Guahati-781014 Assam, India.
- [7] Hardle, W. (1989). Applied nonparametric regression. Cambridge University press, Cambridge.

- [8] International Centre for Research in Agroforestry (ICRAF, 2001). ArcView shapefiles of contours and rivers for Nyando basin. Map No 2001(5).
- [9] Jain, S. and Lall, U. (2001). Floods in a changing climate: Does the past represent the future? *Water Resour. Res.*, Vol37(12), 3193-3205.
- [10] Japan International Co-operation Agency (JICA) (1992). Lake Victoria Basin Catchment Development. River Profile Studies Vol1.
- [11] Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica* Vol46, 33-50.
- [12] Kongo, V.M. (2001). Development of a Simulative Computerized Flood Warning System for River Nyando Flood Plain in Kenya. Msc Thesis, University of Dar es Salaam, Tanzania.
- [13] Masry E., and Tjostheim, O. (1995). Nonparametric Estimation and Identification of Nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econometric Theory* Vol. 11,258-289.
- [14] Masry, E., and Tjostheim, O. (1997). Additive nonlinear ARX time series and projection estimates. *Econometric Theory* Vol. 13,214-252
- [15] Ministry of Environment and Natural Resources (MENR, 1980) Annual Report on Flood Control in Kano Plains. Report No. 55(1980).
- [16] Muthusi, F.M. (2004). Evaluation of the USGS Streamflow Model for Flood Simulation. M.Sc Thesis, Jomo Kenyatta University of Agriculture and Technology, Kenya.
- [17] Mwita, P. (2003). Semi-parametric Estimation of Conditional Quantiles for Time Series with Application in Finance. PhD Thesis, University of Kaiserslautern, Germany.
- [18] Mwita, P. (2005). On conditional scale function: Estimate and asymptotic properties. *African Diaspora Journal of Mathematics* Vol2(2). In press.
- [19] Nadaraya, E. A. (1964). On estimating regression. *Theory Probab. Appl.* Vol9, 141-142.
- [20] Njogu, A.K. (2000). An integrated River Basin Planning Approach – Nyando Case Study in Kenya. Proceedings of the 1st WARFSA/Water Net Symposium on Sustainable Use of Water Resources; University of Zimbabwe, 1- 2 November 2000.
- [21] Office for the Co-ordination of Humanitarian Affairs (OCHA) of the UN, (2002, May, 13). Situation Report No. 1, Kenya – Floods.
- [22] Openshaw, S., Dougherty, M., Corne, S., See, L. (1997). Some initial experiments with neural network models of flood forecasting on the River Ouse. Proceedings of the 2nd annual conference of GeoComputation, University of Otago, New Zealand, 26-29 August 1997.
- [23] Ruppert, D., Sheather, S.J and Wand, M.P. (1995). An effective bandwidth selector for local least squares regression. *J. Amer. Statist. Assoc.*, Vol90, 1257-1270.
- [24] Sankarasubramanian, A. and Upman, L. (2003). Flood quantiles in a changing climate: seasonal forecasts and casual relations. *Water resources research*, Vol39 (5), 1-12.
- [25] Scudder, G. (1980). River Basin Development and Local Initiative in African Savanna Environments. Academic Press, London.
- [26] Stronsaka, K. and Borowicz, A. (2002). MIKE II as Flood Management and Flood Forecasting Tool for the Odra River, Poland. *Journal of Hydrology*, Vol15, 65-70
- [27] Tukey, J.W (1977). Exploratory data analysis. *Theory prob. Appl.* Vol10, 200-205.
- [28] United Nations University (UNU) (2004). The June- August newsletter of United Nations University and its international network of research and training centers/programmes. Accessed from <http://www.update.unu.edu/issue322.htm> .
- [29] Vertessy, R.A., T.J. Hatton, P.J. O'shaughnessy (1993). Predicting Water Yield From a Mountain Ash Forest Catchment using Terrain Analysis Based Catchment Model. *Journal of Hydrology*, Vol150, 665-700.
- [30] Watson, G. S. (1964). Smooth regression analysis. *Sankya ser. A*26, 359-372.
- [31] Weiss, A. (1984). ARMA models with ARCH errors. *Journal of Time Series Analysis* Vol.3, 129-143.
- [32] Woolhiser, D.A. and D.L. Brakensiek (1982). Hydrological System Synthesis. ASAE Monograph No. 5, 3-16.
- [33] Yakowitz, S.J. (1985). Markov Flow Models and the Flood warning Problem. *Water Resources Research*, Vol21, 81-88.
- Yapo, P., Sorooshian, S., Gupta, V. (1993). A Markov Chain flow Model for flood Forecasting. *Water Resources Research*, Vol29, 2427-2436.