# Flood Forecasting over Lower Nzoia Sub-Basin in Kenya

Ngaina Joshua Ndiwa[*1 a], Njoroge James Mwaura[2 b], Mutua Francis[1 c], Mutai Bethwel Kipkoech[1 d] and Opere Alfred Owuor[1 e]
[1] Department of Meteorology, University of Nairobi, P. O. Box 30197 Nairobi (K)
[2]Department of Electrical and Information Engineering, University of Nairobi, P. O. Box 30197 Nairobi (K)
[a] jngaina@gmail.com[*], [b] semajamo@gmail.com, [c] fmutua@uonbi.ac.ke, [d] berth@uonbi.ac.ke, [e] aopere@uonbi.ac.ke

**ABSTRACT**
Real time flood forecasting is one of the most effective non- structural measures for flood management. In this study, Ensemble Kalman Filter (EnKF) is used with Probability Distributed Moisture model (PDM) to forecast flood events over Nzoia sub-basin. The performance of four variations of EnKF (state updating, parameter updating, dual (state parameter) and dual (parameter-state) updating) were evaluated using the Root Mean Square (RMSE) and Coefficient of Efficiency (CoE) for 1, 3, 6, 9 and 12-hour lead time forecasts. In 1EE01 gauging station, RMSE and CoE values was 30m$^3$/s and 0.70 while 1EF01 station had RMSE and CoE values of 50m$^3$/s and 0.82 respectively. For the state variables, standard deviation of 1.1, 0.32, 0.21 and 0.05 were found for recharge, surface storage, groundwater storage and storage respectively while for the PDM parameters, standard deviation of 4.0, 0.2 and 0.2 were found for maximum store capacity, exponent of recharge function and ground recharge time respectively. Parameter updating performed better in terms of RMSE and CoE and thus potential of improving flood forecasting to enable management of flood related risk on real time basis over the sub-basin.

**Key Words:** Discharge, rainfall, Ensemble Kalman Filter, flood forecasting, rainfall-runoff model**.**

## 1.0 INTRODUCTION

Water availability and its use form fundamental components for economic, social and cultural development in Kenya [1]. Kenya's record of flood disasters indicates the worst floods that were recorded in 1961-62 and 1997-98 are associated with a dipole reversal in atmospheric circulation (Indian Ocean Dipole) and Indian Ocean sea surface temperatures [2]. This event caused widespread flooding, rapid and prolonged increases in the levels of many lakes in East Africa and significant economic disruption [3]. Notably, building larger structures alone to cope up with the extremely low probability flood events cannot completely circumvent risk of flood hazards.

Flood forecasting with sufficient lead-time could be used as nonstructural measure for flood hazard mitigation and for minimizing flood related losses. There exists a number of ways to quantify uncertainty in real-time flood forecasting such as sequential data assimilation techniques that provide a means of explicitly taking account of input, model and output uncertainties. One of the earliest data assimilation techniques is the Kalman filter developed for linear systems [4]. For use with nonlinear models, it was later extended resulting in the extended Kalman filter (EKF). These two filters have been widely used in hydrologic modeling [4]. If the nonlinearities in the model are strong, the linearization becomes very inaccurate. This has led to the development of the EnKF where the errors are allowed to evolve with the nonlinear model equations by performing an ensemble of model runs [5]. Feasibility of applying EnKF to real-time flood forecasting by comparing it with EKF for the Sobek River in Netherlands showed that the EnKF gave similar results to those of the already operating EKF model with ten or more ensemble members [6].

Real time flood forecasting systems are aimed at issuing the flood warning in real time in order to prepare the evacuation plan during the flood. The effectiveness of real time flood forecasting systems in reducing flood damage would depend upon among other factors, on how accurately the estimation of future stages or flow of incoming flood and its time sequence at selected points along the river could be predicted. Therefore, this study aims at evaluating the performance of EnKF in real time flood forecasting of River Nzoia in Lower Nzoia sub-basin.

## 2.0 MATERIALS AND METHODS

The study area is the Nzoia River located at latitudes 34º–36ºE and longitudes 0º–1.25ºN in East Africa [7]. It drains into the Lake Victoria and Nile river basins. Nzoia, a sub-basin of Lake Victoria, is chosen as the study area because of its regional importance as it is a flood-prone and also one of the major tributaries to Lake Victoria as shown in Figure 1 [7]. The Nzoia sub-basin covers approximately 12 900 $km^2$ of area with an elevation ranging between 1100 to 3000m. The Nzoia River originates in the southern part of the Mt. Elgon and Western slopes of Cherangani Hills [8]

Rainfall and discharge data were used as shown in Table 1. Daily rainfall records were obtained from the Kenya Meteorological Department for the period 1980-2010 while daily discharge records for the year 2003 were sourced from Kenya's Water Resources Management Authority in order to cover the flood event of April May 2003 used for real time flood forecasting.

The Probability Distributed Moisture (PDM) model is a conceptual rainfall-runoff model which transforms rainfall and evaporation data to flow at the catchment outlet [9]. The runoff production at a point in the catchment is controlled by the absorption capacity of the soil (treated together with canopy interception and surface detention) to take up water. In the PDM model formulation, the surface runoff is calculated from the previous values of the surface runoff and net rainfall. Several studies [4, 10] have shown that Ensemble Kalman filter (EnKF) and its four variations can be used as a sub-optimal estimator, where the error statistics are predicted by using Monte Carlo integration methods.

In this study, error variances of input rainfall and output discharge are specified priori and not updated. Further, a preliminary estimate of input error term [6] was used while a preliminary estimate of discharge measurement error was assumed to be equal to 0.1 times the standard deviation of the measured discharge [12]. A flood event of April–May 2003 [1] was used to evaluate the performance of EnKF over River Nzoia sub-basin. For this flooding event, stream flow data from the gauging stations selected and corresponding rainfall data from stations located in the River Nzoia sub-basin were identified. The PDM model was calibrated for the event data using the Shuffled Complex Evolutionary (SCE) algorithm. Based on the calibrated model parameters, the PDM model was then run using observed daily rainfall data until the beginning of May 2003. Hourly data from TRMM satellite was then used to run the model for River Nzoia sub-basin. The study assumed that the variance of noises introduced to the input forcing and flow measurements were proportional to their magnitudes based on uncertainties in input and output terms as measured [11, 12]. The EnKF and its four variations were then applied to the flood event. The four variations of EnKF considered in this study were the state updating, parameter updating, dual (state-parameter) and dual (parameter-state) updating. The probability distributed moisture (PDM) model was used to transform the rainfall to discharge. The resulting discharge from the PDM model was updated using the EnKF

Using EnKF with state updating, four storage in the PDM model which included recharge ($S_1$), surface storage ($S_2$), groundwater storage ($S_3$) and storage ($S_t$) were considered as state variables and were updated sequentially as new measurements became available. The standard deviation of the four state variables was selected by sensitivity analysis. To obtain the lead time and peak forecasts, a perfect knowledge of the future rainfall was assumed to avoid the error in forecasting rainfall. In reality, the uncertainty in the rainfall forecasts adds to the other uncertainties in the forecasting process. The time interval used for the flood hydrograph was one hour. Forecasts were made at 1-, 3-, 6-, 9- and 12-hour lead times at every forecast time. At each forecast time, the magnitude of the forecast peak was also obtained. For EnKF, with parameter updating, three of the PDM model parameters, namely, maximum store capacity ($C_{max}$), exponent of recharge function ($b_g$) and ground recharge time ($k_2$) were updated. These are the most sensitive parameters of the PDM model. The updating procedure was initialized by defining prior uncertainty range associated with the three parameters. As the initial ensemble of parameters had to be specified, these three parameters were randomly sampled from a normal distribution with the standard deviations and obtained by sensitivity analysis. For dual EnKF, state parameter updating was considered where both state variables and parameters were sequentially updated. The adequacy of the EnKF was evaluated by using the root mean square (RMSE) and the coefficient of efficiency for 1, 3, 6, 9 and 12 hour forecasts. The RMSE is defined by equation (1) as

$$RMSE_L = \sqrt{\frac{\sum_{i=1}^{n}\{Q_{f,L}(i) - Q_{obs}(i)\}^2}{n}} \qquad 1$$

Where $Q_{f,L}(i)$ is the forecasted discharge for lead time L for forecast i and $Q_{obs}$ corresponding observed discharge. The coefficient of efficiency (CoE) of a model is defined as the proportion of the variance of the observed discharge accounted by the model [13]:

$$CoE_L = 1 - \frac{S}{S_{obs}} \qquad 2$$

Where $S$ is the simulated model discharge while $S_{obs}$ is observed discharge. In addition, the error in the peak discharge magnitude and the timing were also used in the evaluation. To make objective comparison of the performance of EnKF with different options, a perfect knowledge of future observed rainfall was assumed in obtaining lead time forecasts and peak discharges. Computations were determined only for the period where the observed discharge was greater than $30m^3/s$ to avoid the small discharge values which are not significant in flood forecasting [4].

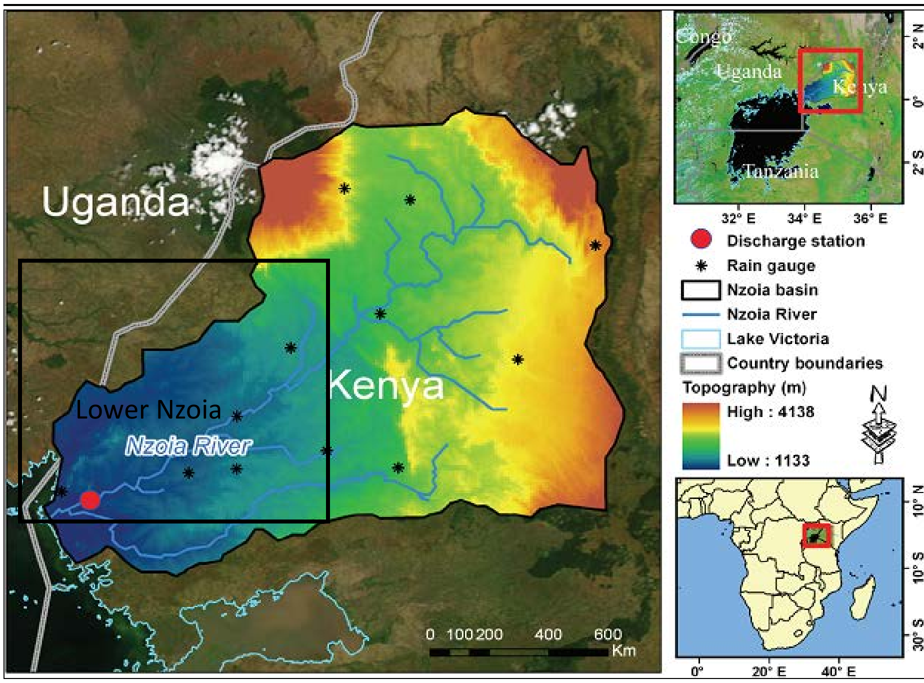


Figure 1: Map of Nzoia river sub-basin in Lake Victoria region showing the location of Lower Nzoia sub-basin [7]

Table 1: Selected rainfall stations

| | Station Name | Coordinates | Record Period | Record Length (Years) |
|---|---|---|---|---|
| 1. | Kitale meteorological station | 1.01°N 35.0°E | 1980-2010 | 31 |
| 2. | Bunyala irrigation scheme | 1.30°S 36.8°E | 1980-2010 | 31 |

Table 2: Selected River Gauging Stations (RGS)

| | Station ID | Location | Sub-basin Name | Period | Record Length (Years) |
|---|---|---|---|---|---|
| 1. | 1EE01 | 0.18 N, 34.22 E | Nzoia | 2003 | 1 |
| 2. | 1EF01 | 0.12 N, 34.09 E | Nzoia | 2003 | 1 |

## 3.0 RESULTS AND DISCUSSION

### 3.1 Model Calibration

Results of calibration of PDM model based on daily and hourly rainfall data are presented in Table 3. Table 4 and 5 shows the standard deviation of the four state variables and the standard deviation and range of PDM parameters updated.

Table 3: The calibrated parameters of the PDM model

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| $C_{max}$ | 436 | $K_1$ | 1 |
| $C_{min}$ | 0 | $K_2$ | 4.9 |
| b | 1.42 | $K_g$ | 900 |
| $b_e$ | 5.0 | $S_t$ | 3.28 |
| $b_g$ | 1.26 | $t_d$ | 0 |
| $K_b$ | 20 | | |

Table 4: State variable standard deviations

| State variable | $S_1$ | $S_2$ | $S_3$ | $S_t$ |
|---|---|---|---|---|
| Standard deviation | 1.1 | 0.32 | 0.21 | 0.05 |

Table 5: Standard deviation and range of the PDM parameters updated

| Parameter | Minimum | Maximum | Standard deviation |
|---|---|---|---|
| $C_{max}$ | 90 | 500 | 4.0 |
| $b_g$ | 0.5 | 2.5 | 0.2 |
| $k_2$ | 8.0 | 16.0 | 0.2 |

Based on the calibrated parameters in table 2 RMSE and CoE values in 1EE01 gauging station, was $30m^3/s$ and 0.70 while 1EF01 station had RMSE and CoE values of $50m^3/s$ and 0.82 respectively. This indicated that the PBM model output (discharge) corresponded to the observed discharge data over River Nzoia sub-basin.

### 3.2 Model Evaluation

For each EnKF method applied to Nzoia River, the average values of the RMSE and CoE for the selected flood event with a lead times of 1, 3, 6, 9, 12 hours were presented in Table 6 and 7 respectively. The 1-h, 3-h, 6-h, 9-h and 12-h lead time forecasts with 95% forecast limits are shown in Figure 2 to Figure 4 for the flood event of may 2003

Table 6: Comparison of RMSE

| Variations of EnKF | Lead time (hours) | | | | |
|---|---|---|---|---|---|
| | 1 | 3 | 6 | 9 | 12 |
| State | 9.2 | 12.2 | 16.1 | 18.7 | 20.1 |
| Parameter | 6.5 | 9.2 | 12.6 | 15.2 | 17.0 |
| State-Parameter | 7.6 | 10.7 | 14.5 | 16.9 | 18.3 |
| Parameter-State | 10.1 | 13.1 | 16.8 | 19.3 | 20.6 |

Table 7: Comparison of the Coefficient of efficiency

| Variations of EnKF | Lead time (h) | | | | |
|---|---|---|---|---|---|
| | 1 | 3 | 6 | 9 | 12 |
| State | 0.80 | 0.74 | 0.65 | 0.57 | 0.53 |
| Parameter | 0.82 | 0.80 | 0.73 | 0.67 | 0.62 |
| State-Parameter | 0.83 | 0.78 | 0.69 | 0.63 | 0.59 |
| Parameter-State | 0.79 | 0.73 | 0.63 | 0.56 | 0.52 |

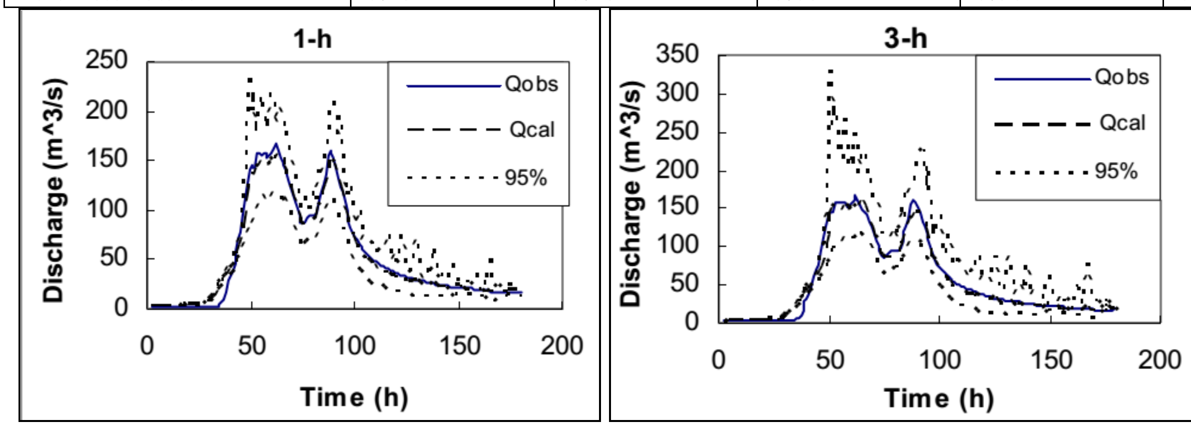

Figure 2: Comparison of flood forecast for 1-h and 3-h hr lead time

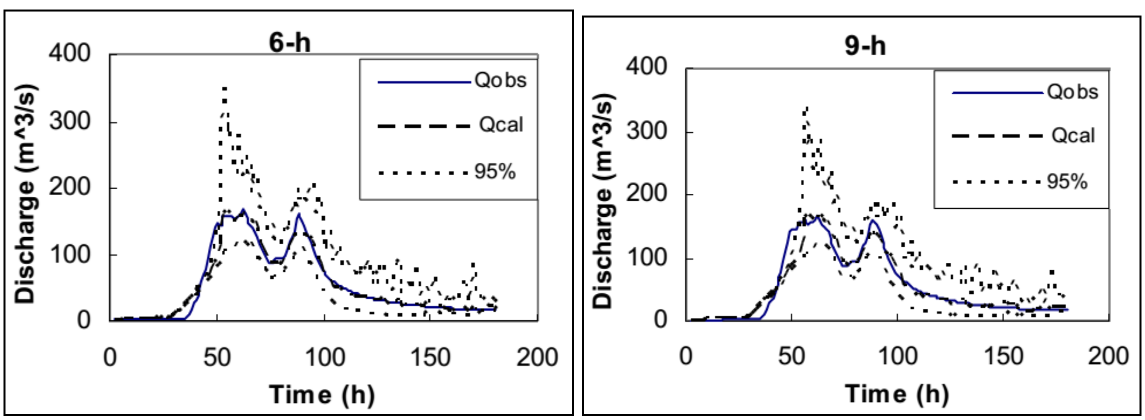

Figure 3: Comparison of flood forecast for 6-h and 9-h lead time

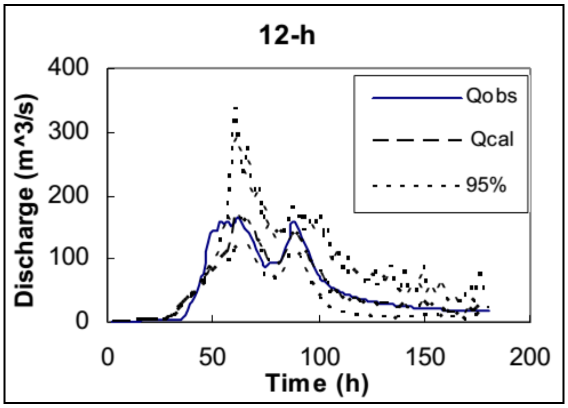

Figure 4: Comparison of flood forecast for 12-h lead times

It can be seen from Table 6 that the EnKF with parameter updating gave the smallest RMSE for all the lead time. Similarly, it gave the largest coefficient of efficiency for all the lead time (Table 7). Table 6 and Table 7 shows superior performance of parameter updating over the other variations. This indicated that the EnKF with parameter updating performed better compared to the other three variations. Based on Figure 2 to Figure 4, the study showed that the quality of the forecasts deteriorated with increase in the lead time from 1-h to 12-h.

## 4.0 CONCLUSION

The ability of the EnKF with the PDM model to forecast discharge was evaluated by using flood event of April-May 2003 in the Nzoia River sub-basin. Four variations of the EnKF, namely, the state, parameter and dual (state-parameter and parameter-state) were considered. Based on the calibrated parameters, RMSE and CoE values in 1EE01 and 1EF01 gauge stations indicated that the PDM model output (discharge) corresponded to the observed discharge data for Nzoia River Basin. A standard deviation of 1.1, 0.32, 0.21 and 0.05 were found for the state variables $S_1$, $S_2$, $S_3$ and $S_t$ respectively while a standard deviation of 4.0, 0.2 and 0.2 were found for the PDM updated parameters which included $C_{max}$, $b_g$ and $k_2$ respectively. The study noted EnKF with parameter updating gave the smallest RMSE for all the lead time. Likewise, it gave the largest coefficient of efficiency for all the lead time. This indicate that the EnKF with parameter updating performed better compared to the other three variations and thus could be used for real time flood forecasting to manage flood risk over the sub-basin.

**ACKNOWLEDGEMENT**

The authors would like to thank the Department of Meteorology, University of Nairobi for providing support during the study, the Kenya Meteorological Department for provision of rainfall datasets and Water Resource and Management Authority for provision of river discharge data used for the study.

## REFERENCES

1 Ministry of Water and Irrigation (MoWI), 2009: Flood Mitigation Strategy

2 Saji, N. H, Goswami B. N, Vinayachandran P. N., and Yamagata T. (1999): A dipole mode in

the tropical Indian Ocean. Nature, 401, 360-363

3 Birkett, C.M., Murtugudde, R. and Allan, J.A. (1999): Indian Ocean climate event brings floods to East Africa's lakes and Sudd Marsh. Geophys. Res. Lett. 26, 1031-1034.

4 Srikanthan, R, Amirthanathan, G. E. and Kuczera, G. (2007): Real-time flood forecasting using ensemble Kalman filter. In Oxley, L. and Kulasiri, D. (eds) MODSIM 2007 International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, pp.1789-1795

5 Burgess G, Van Leeuwen P.J, Evensen G, (1998): Analysis Scheme in the Ensemble Kalman Filter. Monthly Weather Review 126, 1719-1724

6 El Serafy G. Y. and Mynett A.E. (2004): Comparison of EKF and EnKF in Sobek River: Case study Maxau-Ijssel. In: Liong S-Y, Phoon K-K and Baovic V. (Eds.) Proc. 6th International Conference on HydroInformatics, World Scientific, pp. 513-520.

7 Khan S. I, Adhikari P, Hong Y, Vergara H, Adler R. F, Policelli F, Irwin D, Korme T, and Okello L, (2011): Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. Hydrol. Earth Syst. Sci., 15, 107–117

8 Li, L., Hong, Y., Wang, J. H., Adler, R. F., Policelli, F. S., Habib, S., Irwn, D., Korme, T., and Okello, L (2009): Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational flood prediction system in Nzoia Basin, Lake Victoria, Africa, Nat Hazards, 50, 109–123.

9 Moore, R. J. (2007): The PDM rainfall-runoff model, Hydrology and Earth System Sciences, 11(1), 483-499.

10 Khan S. I, Adhikari P, Hong Y, Vergara H, Adler R. F, Policelli F, Irwin D, Korme T, and Okello L, (2011): Hydroclimatology of Lake Victoria region using hydrologic model and satellite remote sensing data. Hydrol. Earth Syst. Sci., 15, 107–117

11 Weerts A.H. and El Serafy G. (2006): Particle filtering and ensemble Kalman filtering for State updating with hydrological conceptual Rainfall Runoff Models, Water Resources Research, 42, W09403

12 Georgakakos, K.P. (1986): A generalized stochastic hydrometeorological model for flood and flash flood forecasting: 2. Case studies, Water Resources Research, 22(13), 2096-2106.

13 Nash, J. E. and Sutcliff, J. V. (1970): River flow forecasting through conceptual models, 1. Discussion of principles", Journal of Hydrology, 10, 282-290