COMPARATIVE CASE STUDY OF RAINFALL - RUNOFF MODELS OVER THE NYANDO RIVER BASIN

Stephen K. Rwigi, Alfred O Opere and Francis M. Mutua

Department of Meteorology, University of Nairobi

CORRESPONDING AUTHOR

Stephen K. Rwigi
Department of Meteorology, University of Nairobi
P.O. Box 30197-00100, Nairobi, Kenya

(Manuscript received 9 June 2011, in final form 10 July 2012)

ABSTRACT

The performance of some existing rainfall-runoff linear systems models; the Simple Linear Model (SLM), the Linear Perturbation Model (LPM), the Linearly Varying Gain Factor Model (LVGFM) and a conceptual model (the Soil Moisture Accounting and Routing (SMAR) are tested on data from the Nyando catchment. The linear systems models were applied to the Nyando catchment in both non-parametric and also under the constraint of the gamma function impulse responses. There was no meaningful loss of generality associated with the constraint of the gamma function.

The data used in this study are: daily rainfall, evaporation and runoff. Arithmetic mean method was used to convert point rainfall measurements to areal average rainfall. Runoff and evaporation data were tested for consistency before application to the models. It was found that daily rainfall, runoff, and evaporation data over the Nyando catchment were of good quality.

The results obtained from the application of the models to the catchment indicated that among the four models, the conceptual SMAR model had the best performance for Nyando. It was followed by the LVGFM, the LPM and the SLM in that order. Compared with the linear systems models, the SMAR model performed consistently better in both calibration and verification periods; an indication that the model is apparently superior compared to the linear systems models. We therefore concluded that among the models considered, the SMAR model is the best model suited for the Nyando catchment. Among the linear systems models, the LVGFM performed better than the other two. Hence in the absence of the SMAR model, LVGFM, which does almost as well as the SMAR model, may be applied.

Key words: Linear systems models, Conceptual models, Simple Linear Model, Linear Perturbation Model, Linearly Varying Gain Factor Model, Nyando Catchment, Calibration period, Verification period, Rainfall, Evaporation, Runoff

1. Background

Sustainable water resources management depends on reliable hydrological forecasts. The basic management process of water resources is to collect and assess hydrological data so that informed decisions can be taken and future strategies would be based on the assessment of the available facts. Modern managers and designers of hydrological projects are accepting models as interactive decision support tools since model forecasts of hydrological processes permit decisions to be made with more confidence leading to a more efficient use of the water resources (Falconer, 1992). This makes model forecasts a major factor in economic development since water is a major player in almost all economic activities.

Hydrological models are sets of mathematical abstractions describing relevant phases of the hydrologic cycle, and are probably the
most powerful tools available for clarifying the significant processes occurring in a natural hydrological system and for forecasting the effects of any changes, such as the level of runoff, from the system. The level of runoff in the rivers is a determining factor in the development of many hydrological projects. This level is determined by the proportion of rainfall that finally transforms to runoff. This proportion is always a function of the catchment’s physical properties, which determine the level of abstractions such as evaporation, infiltration and percolation. Since runoff is a function of rainfall, a rainfall-runoff model, with rainfall data as input can be used to forecast discharge from a catchment to enable water resource managers to provide reliable information for better water resources management. The relatively slow response of hydrological processes to the changes in meteorological processes makes it possible to model runoff using meteorological data (WMO 1975).

Since no single model structure is capable of producing consistently accurate forecasts for a wide range of catchments and climatic conditions, and given that most hydrological models have been developed and applied in temperate zones, there is need to evaluate the performance of various rainfall runoff models in tropical regions in order to select the one that best fits our needs (Adegú, 1999). This is normally done through recalibration of the models using data from the specific tropical regions.

Nyando River basin experiences frequent flood events which exposes the people living within the catchment to flood related risks. There is need, therefore, to develop a comprehensive hydrological modelling system for timely warning of floods and efficient management of the available water resources in the catchment. Such a system would be particularly helpful in making decisions related to flood protection measures within the catchment as well as the rational use of water for irrigation and other purposes. This paper aims at providing a basis for an early warning system for floods in Nyando.

1.1 Area of Study

The study area is located in the western side of Kenya and includes the whole of that area drained by the Nyando River and its tributaries. The area is located within Lake Victoria South Drainage Basin and covers an area of about 3580 square kilometers. The area of interest is found within latitudes 0°7’/N and 0°25’/S and longitudes 34°34’E and 35°43’E (Figures 1).

![Location of Nyando Catchment](image)

Figure 1: Location of Nyando Catchment within Kenya and rain gauge stations. (Source: Adegú, 1999)

The landform of Nyando basin varies from low plains near the shores of Lake Victoria to plateaus and mountains to the east with elevations varying from 1134 m to over 3000 m above sea level. The geology of the basin is the same that of the main Lake Victoria basin which has geological formations varying from recent quaternary sediments to old rocks of Archean age. The soils in the Nyando basin are predominantly clays but vary greatly in texture, composition and structure. Soils derived from the quaternary volcanic
rocks, which are generally fertile, are found in the higher rainfall areas on the eastern side of the catchment. Those soils derived from very ancient granite are reasonably fertile and tend to be in areas of low rainfall within the catchment (Opere, 1998).

The dominant land use activities in the Nyando Catchment include farming and settlement with the main crop being sugar cane, which doubles as the main cash crop for the inhabitants and is dominant in the northern parts of the catchment. Rice and cotton are other crops in the catchment and are dominant in the southern parts. Maize, beans and sorghum are the main subsistence crops in the area (Wanjohi, 1999).

The Nyando river basin is in the equatorial zone of low pressure where winds are generally light and variable. Two monsoons generally prevail over the basin in the course of the year, Northeast and Southeast monsoons, which have contrasting thermodynamic characteristics. Spasmodic outbreaks of westerlies often intrude into the established wind systems to cause marked changes on the normal rainfall pattern. These westerlies are known to cause above-normal rainfall whose source is the moisture inflow from the tropical rain forests of the Congo. Rainfall is normally used as the descriptor of climate within the tropics. The variation of rainfall from January through December is usually considered a sufficient descriptor of tropical climate. Other meteorological variables do not exhibit significant variation throughout the year. The basin experiences a trimodal pattern of rainfall with peaks in the months of April, August and November with magnitudes decreasing as we move towards November (Figure 2).

The basin experiences an average annual rainfall of about 1400 mm ranging from well below 1100 mm per annum around the lakeshores to over 1800 mm towards eastern highlands. The basin has no distinct dry period and thus it can be regarded as a moderately humid catchment.

![Variation of Monthly Rainfall over Nyando](image)

Figure 2: Average monthly rainfall at seven typical rainfall stations in the Nyando catchment

2. Data and Methods

The data sets required for this study are daily rainfall, evaporation and runoff. Daily rainfall and evaporation data sets were obtained from the Kenya Meteorological Department (KMD) while stream flow data sets were obtained from the Ministry of Water and Irrigation. Availability and adequacy of data were the main limitations and filling in the missing data records for all the three data sets had to be done using suitable methods in order to assure continuity of data which is a vital requirement in any research. The number of years with data varied for the three
data sets had to be done using suitable methods in order to assure continuity of data which is a vital requirement in any research. The number of years with data varied for the three data sets but all the three were available from 1983 to 1995. Six years of data, 1985 to 1990, were used as this was found to be the period with the highest quality data.

Rainfall measurements, obtained from rainfall stations as point measurements, were converted to areal rainfall using the arithmetic mean method to represent the average depth of rainfall in the catchment which is considered a better representation of the catchment rainfall than the point observations. The areal average was calculated using 16 rain gauge stations (Table 1).

Table 1: Rainfall stations used for the purpose of this study; their location, years of records and the percentage of missing data. The station coding, Gi, is used here only for the purpose of convenience.

<table>
<thead>
<tr>
<th>Station Code</th>
<th>Station Name</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Years of records</th>
<th>Missing Data (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8935033</td>
<td>Nandi Hills Savani</td>
<td>3° N</td>
<td>35°6'E</td>
<td>1960–2000</td>
<td>8.0%</td>
</tr>
<tr>
<td>8935148</td>
<td>Kipkurere Forest</td>
<td>5° N</td>
<td>35°25'E</td>
<td>1960–2000</td>
<td>6.4%</td>
</tr>
<tr>
<td>8935161</td>
<td>Nandi Hills Tea Est.</td>
<td>5° N</td>
<td>35°9'E</td>
<td>1962–2000</td>
<td>1.0%</td>
</tr>
<tr>
<td>9034086</td>
<td>Ahero Irrigation</td>
<td>9° S</td>
<td>34°56'E</td>
<td>1960–2000</td>
<td>1.9%</td>
</tr>
<tr>
<td>9035020</td>
<td>Kipkelion Railway</td>
<td>12° S</td>
<td>35°28'E</td>
<td>1960–2000</td>
<td>7.6%</td>
</tr>
<tr>
<td>9035075</td>
<td>Kaisugu House</td>
<td>20° S</td>
<td>35°23'E</td>
<td>1962–2000</td>
<td>3.7%</td>
</tr>
<tr>
<td>9035148</td>
<td>Koru Bible School</td>
<td>12° S</td>
<td>35°16'E</td>
<td>1962–2000</td>
<td>4.3%</td>
</tr>
<tr>
<td>9035151</td>
<td>Londiani Entomology</td>
<td>8° S</td>
<td>35°35'E</td>
<td>1962–2000</td>
<td>3.6%</td>
</tr>
<tr>
<td>9035155</td>
<td>Londiani Forest Stn</td>
<td>3° S</td>
<td>35°37'E</td>
<td>1962–2000</td>
<td>3.7%</td>
</tr>
<tr>
<td>9035188</td>
<td>Tinga Monastery</td>
<td>12° S</td>
<td>35°27'E</td>
<td>1964–2000</td>
<td>4.4%</td>
</tr>
<tr>
<td>9035201</td>
<td>Kipkorem Estate</td>
<td>19° S</td>
<td>35°20'E</td>
<td>1968–2000</td>
<td>8.4%</td>
</tr>
<tr>
<td>9035220</td>
<td>Koru Homaline co.</td>
<td>10° S</td>
<td>35°17'E</td>
<td>1962–2000</td>
<td>3.2%</td>
</tr>
<tr>
<td>9035226</td>
<td>Londiani Forest</td>
<td>5° S</td>
<td>35°21'E</td>
<td>1960–2000</td>
<td>8.4%</td>
</tr>
<tr>
<td>9035240</td>
<td>Keresoi forest</td>
<td>17° S</td>
<td>35°32'E</td>
<td>1961–2000</td>
<td>3.6%</td>
</tr>
<tr>
<td>9035256</td>
<td>Malagat Forest</td>
<td>5° S</td>
<td>35°32'E</td>
<td>1962–2000</td>
<td>6.5%</td>
</tr>
<tr>
<td>9035263</td>
<td>Tinderet Tea Est.</td>
<td>21° S</td>
<td>35°21'E</td>
<td>1964–2000</td>
<td>4.6%</td>
</tr>
</tbody>
</table>

Many statistical analyses require that data being used be homogeneous in order that the research results be considered satisfactory. Homogeneity of data was tested using mass and double mass curve analyses, the two most popular methods in hydrology (Ogallo, 1981). The models that were tested using homogeneous data from the Nyando basin are: the Simple Linear Model (SLM), the Linearly Varying Gain Factor Model (LVGFM), the Linear Perturbation Model (LPM) and the Soil Moisture Accounting and Routing (SMAR) model. The first three are linear systems models which presume that rainfall is the cause of runoff without making any attempt to account for the physical nature of the transformation of rainfall into runoff (O’Connor, 1995) while the fourth one is a conceptual model which is more physically based as it attempts to account for the physical nature of the transformation of rainfall into runoff. Though simple in nature, linear systems models have been shown to perform equally well as the higher physically based conceptual models. They have the advantage that their data requirements are readily met. Only a brief highlight of these models is given in this paper. Details are available in Rwigi (2004) among others.

2.1 Simple Linear Model

The Simple Linear Model (SLM) is the simplest of all the linear systems models and usually serves as a convenient starting point in rainfall-runoff modeling. Nash and Sutcliffe (1982) introduced the model as a basis of effici-
efficiency comparison with other more elaborate models. The intrinsic hypothesis of the SLM is the assumption of a linear time-invariant relationship between the total rainfall $x_i$ and the total discharge $y_i$. The convolution summation relation of the form shown below expresses the SLM in its discrete form (Kachroo and Liang, 1991)

$$y_i = \sum_{j=1}^{m} x_{k,j+1}h_j + e_i \quad \text{................................(1)}$$

Where $h_j$ refers to the $j^{th}$ ordinate of the pulse response function, $e_i$ refers to the $i^{th}$ forecast error term, $x_i$ refers to the $i^{th}$ rainfall input series, $y_i$ refers to the $i^{th}$ runoff output series and $m$ refers to the memory length of the system.

### 2.2 Linear Perturbation Model

This model lays more emphasis on the observed seasonal behaviour inherent in the observed rainfall and discharge data series. The Linear Perturbation Model (LPM) structure reduces reliance on the linearity assumption of the SLM and gives substantial weight to the observed seasonal behaviour of the catchment (Nash and Barso 1983). It is therefore considered suitable for catchments that exhibit marked seasonal variations as it involves the assumption of a linear relationship between departures from seasonal expectations of both input and output series.

The relation between the departure series of the LPM given by the convolution summation of the form:

$$Q_i = \sum_{j=1}^{m} R_{i-j+1}h_j + e_i \quad \text{................................(2)}$$

where, $Q_i$ refers to the departure series of outflow (runoff), $R_i$ refers to the departure series of inflow (rainfall) from their seasonal means, $m$ refers to the memory length with $j = 1, 2, 3...m$, $h_j$ refers to the $j^{th}$ ordinate of the discrete pulse response relating the departure series of input and output, $e_i$ refers to the $i^{th}$ output error term.

### 2.3 Linearly Varying Gain Factor Model

The Linearly Varying Gain Factor Model (LVGFM) is based on the variation of the gain factor with the selected index of the prevailing catchment wetness index. Surface runoff is directly related to effective rainfall, which is inversely related to the hydrologic abstractions which vary widely both in time and in space, depending on the initial level of the soil moisture. The main assumption of the LVGFM model is that the amount of rainfall that transforms to runoff is a function of the state of the soil moisture. It involves only the variation of the gain factor with the selected index of the prevailing catchment wetness, without varying the shape of the response function. Using a time varying gain factor $G_i$ the model output has the structure of the form:

$$y_i = G_i \sum_{j=1}^{m} w_j x_{i-j+1} + e_i \quad \text{................................(3)}$$

with $\sum_{j=1}^{m} w_j = 1 \quad \text{................................(4)}$

Where $x_i$ refers to the magnitude of the pulsed input during the $i^{th}$ time interval, $y_i$ refers to the magnitude of the pulsed output at the end of the $i^{th}$ time interval; $G_i$ refers to the magnitude of the gain factor at the end of the $i^{th}$ time interval. This is the time varying coefficient of runoff that accounts for the short-term time dependence of runoff, which is due to the effect of variable soil moisture conditions. $w_j$ refers to the $j^{th}$ ordinate of the non-parametric discrete weighting and $e_i$ function, $m$ refers to the memory length of the system and $e_i$ refers to the $i^{th}$ output error term.

### 2.4 Conceptual Soil Moisture Accounting and Routing Model

The Conceptual Soil Moisture Accounting and Routing (SMAR) Model is a lumped quasi-physical conceptual rainfall-evaporation-runoff model, with quite a distinct water-balance component in which the rainfall and evaporation interact to produce runoff and a linear routing co-
mponent which transforms the generated runoff into discharge. This is achieved through a series of steps, which represent some of the known physical processes in a hydrological system. The structure of the model emphasizes the distinct roles of the linear water balance component and the linear routing component.

The GFVS version presented here, has three two-parameter distribution options available for routing the generated ‘surface runoff’ component of the SMAR model, namely, the classic gamma distribution model (Nash, 1957), its discrete counterpart, the Negative Binomial distribution (O’Connor, 1976), and the sharply peaked Inverse Gaussian distribution (Bardsley, 1983) for flashy catchments. The version of SMAR used in the present study has nine parameters as shown below.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z</td>
<td>The combined water storage depth capacity of the layers (mm)</td>
</tr>
<tr>
<td>T</td>
<td>A parameter (less than unity) that converts the given evaporation series to the model-estimated potential evaporation series.</td>
</tr>
<tr>
<td>C</td>
<td>The evaporation decay parameter, facilitating lower evaporation storage rates from the deeper soil moisture storage layers</td>
</tr>
<tr>
<td>H</td>
<td>The generated ‘direct runoff’ coefficient</td>
</tr>
<tr>
<td>Y</td>
<td>The maximum infiltration capacity depth (mm)</td>
</tr>
<tr>
<td>n</td>
<td>The shape parameter of the Nash gamma function ‘surface runoff’ routing element; a routing parameter</td>
</tr>
<tr>
<td>nK</td>
<td>The scale (lag) parameter of the Nash gamma function ‘surface runoff’ routing element; a routing parameter</td>
</tr>
<tr>
<td>g</td>
<td>The weighting parameter, determining the amount of generated ‘groundwater’ used as input to the ‘groundwater’ routing element.</td>
</tr>
<tr>
<td>K_g</td>
<td>The storage coefficient of the ‘groundwater’ (linear reservoir) routing element; a routing parameter</td>
</tr>
</tbody>
</table>


2.5 Updating Model Forecasts

Model updating refers to the process of modifying the output from a model by correcting them in accordance with the behaviour of errors just before issuing a forecast. In the updated mode, the model errors in the previous forecasts are used to improve the model outputs. Updating is particularly important when a model is used for real-time forecasting. One of the commonly used methods in forecasting improvements is the autoregressive (AR) updating where errors in the most recent forecast are used to correct the future model forecasts. In the updated mode, the model output accuracies results are improved during calibration and verification periods. Figure 7 and Table 3 present the results of the third order autoregressive (AR) updating. This order was adopted since it gave the best efficiency values compared to the other orders. Orders lower than or higher than the third resulted in lower efficiencies than those of the third order.

2.6 Model Efficiency Criteria

The efficiency criteria that express the accuracy of each model are generally linked with the objective function used in calibration for optimizing its parameters. A commonly used objective function is the sum of squares of differences, F, between the observed and estimated discharges, where the summation is taken over the whole calibration period (Kachroo, 1992).

\[ F = \sum (y - \hat{y})^2 \]  

(5)

Where F represents the sum of squares of differences that reflect the extent to which a model is successful in reproducing the observed discharges, the lower the value of F, the better the performance of the model. The \( y \) represents the measured or observed output and \( \hat{y} \) represents the model output estimates.
Nash and Sutcliffe (1970) corrected the shortcomings of Equation 5 by defining model efficiency $R^2$ that is analogous to the coefficient of determination in linear regression. By so doing different models could be compared even in different catchments. They defined the efficiency $R^2$ as the proportion of the initial variance accounted for by $F$.

$$R^2 = \frac{F_0 - F}{F_0} \quad \text{(6)}$$

Where initial variance, $F_0$, was defined as

$$F_0 = \sum (y - \bar{y})^2 \quad \text{(7)}$$

The efficiency criterion, $R^2$, measures the percentage improvement of the applied substantive model over the naïve model situation.

After the models were calibrated and verified, they were then fitted on daily data from the Nyando catchment. The model performance criteria used in each of the models include the coefficient of efficiency ($R^2$), the proportion of improvement ($r^2$), the index of volumetric fit (IVF), and the mean square error (MSE). The GFFS software package was used in the application of each model.

3. Results and Discussion

3.1 Rainfall, Runoff and Evaporation Data

Results of rainfall, runoff and evaporation data quality control show that all the three data sets were consistent. Generally only a single line could be fitted in the mass curves of all the data sets. The percentage $R^2$ values were quite high for the three data sets, indicating that the trend lines fitted the observed data sets quite well. This means that observed data were quite consistent. The daily rainfall, runoff and evaporation data over the Nyando catchment were therefore declared homogeneous.

3.2 Model Comparisons

The results obtained with each individual model and their comparisons are presented in this section. As a basis for comparison, it was found necessary to start with a situation where no model was available. This is the seasonal model situation where, in the absence of any model, seasonal mean values are normally used as future forecasts of discharge. Such a model is the most naïve of any known model. The model merely represents daily seasonal patterns of areal rainfall and discharge.

3.2.1 Simple Linear Model

Table 1, gives a summary of the results obtained from the application of the simple linear model (SLM) in both parametric and non-parametric form while Figures 3a and 3b give the corresponding hydrographs. The model efficiencies in non-parametric mode are about 47 % and 29 % for calibration and verification periods, respectively. In parametric mode, the model efficiencies are about 43 % and 29 % in the calibration and verification periods respectively. This indicates a decline in the performance during calibration in parametric mode while no change is observed in the verification period. There is a slight discrepancy in the volume match as given by the index of volumetric fit (IVF) values. The values differ slightly from the desired value of unity in both calibration and verification periods. The value of 1.26 in the calibration period suggests that the model tends to overestimate the flow volume during the calibration period while underestimating it, as given by the value of 0.89, during the verification period.

The mean square error values are about 3.5 and 7.7 in calibration and verification periods respectively. This is about 7.4 % and 26.7 % of the mean values during the calibration and verification periods, respectively. This indicates a standard deviation of less than 10% in calibration and less than 30% in verification.
Table 1: Summary of the results of application of the three linear systems models: SLM, LPM and the LVGFМ and the conceptual SMAR model showing $R^2$ which is the measure of model efficiency; $r^2$ which is the proportion of improvement of the model, MSE which is the mean square error and, IVF which is the index of volumet-

<table>
<thead>
<tr>
<th>Calibration period</th>
<th>Verification period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Model</td>
<td>SLM</td>
</tr>
<tr>
<td>$R^2$ (%)</td>
<td>47.0</td>
</tr>
<tr>
<td>$r^2$ (%)</td>
<td>-</td>
</tr>
<tr>
<td>MSE</td>
<td>3.49</td>
</tr>
<tr>
<td>IVF</td>
<td>1.26</td>
</tr>
</tbody>
</table>

Figures 3a and 3b present the pulse response ordinates in both the constrained and unconstrained form. It is observed from Figure 3a that the unconstrained pulse response ordinate estimates generally decrease in value from left to right. There is a clear recession when the shape is constrained as shown by the parametric simple linear model (PSLM) hydrograph (Figure 3b). Imposing the constraint on the shape of the derived pulse response ordinates normally produces little loss in model efficiency as shown in Table 1 columns 2 and 6.

Figures 6a shows rainfall together with the hydrograph of observed and SLM’s estimated discharges for the year 1985. It is observed from the figure that the SLM captures the extreme flows very well, as displayed by the peaks and troughs in the hydrograph where observed and estimated discharge peaks and troughs coincide.

However, it is also observed from the figure that the model overestimates discharge during low flow (July - November) while underestimating it during high flow (April -June) seasons.

![3(a)](image1.png)  ![3(b)](image2.png)

Figures 3a and 3b: Graphical representation of the pulse response ordinates over the Nyando catchment for parametric, (a), and Non-parametric, (b), forms of SLM
3.2.2 Linear Perturbation Model

The results of the model’s performance in both simulation and parametric modes are presented in three formats. These are: a table of summary of results in terms of $R^2$, MSE, IVF and $r^2$ as presented in Table 1 columns 3 and 7, pulse response ordinates as presented in Figures 4a and 4b, and hydrographs of estimated and observed discharges as given in Figure 6b.

From Table 1 columns 3 and 7, it is observed that the model efficiencies as measured by the $R^2$ in non-parametric mode are about 55% and 25% during the calibration and verification periods respectively. In parametric mode the $R^2$ values are about 59% and 20% in calibration and verification periods respectively.

Comparison of the results of the non-parametric LPM with those of the SLM shows that in the calibration period there is some marked improvement in efficiency, of about 15% in non-parametric and about 29% in parametric modes respectively, over the SLM during the calibration period as shown by the $r^2$ values which show the initial variance unaccounted for by SLM but is subsequently accounted for by the LPM.

The seasonal component of LPM seems to account for a larger percentage of the initial variance during the calibration period than the SLM. This is about 15% of the initial variance during calibration but fails to account for 5% of initial variance during verification. Even though the SLM performed fairly well in the catchment, the combination of the seasonal and the linear component of LPM appear to have improved the results significantly during the calibration period. During the verification period however there is virtually no difference between the results of SLM and LPM both of which may be considered poor.

The IVF is 1.0 during the calibration period. This is the ideal value and is an indication that the volume of the computed output is about the same as the volume of the observed output. During verification the IVF is 0.72 indicating underestimation during this period.

Figures 4a and 4b present the unconstrained and constrained pulse response ordinate estimates respectively as obtained from the LPM.

The observed and estimated discharges (Figure 6b) for the year 1985 follow a similar pattern to that of the SLM; where it overestimates discharge during low flow and underestimates it during high flow seasons. However the deviations between the observed and estimated discharge values are generally smaller than those of SLM. The highest observed discharge value, in mid April, is about 252 cumecs and corresponds to the estimated discharge value of about 150 cumecs. This is a major improvement over the SLM.

Though the deviations from the observed values are still large, they are generally smaller than those obtained with the SLM, an indication that simulated flows from the LPM match the observed flows much better than those from the SLM.

---

**Figure 4a**

Graphical representation for parametric, (a), and non-parametric, (b), forms of the pulse response ordinates over the Nyando catchment for LPM.
3.2.3 Linearly Varying Gain Factor Model

The summary of results obtained from this model is given in Table 1 columns 4 and 8. The $R^2$ efficiency value is about 69% and 28% during calibration and verification periods respectively. The 69% is a much better efficiency during calibration than those obtained from either SLM or the LPM during the same period. This is indicated by the $r^2$ value of about 41% which is an indication of the improvement of the Linearly Varying Gain Factor Model (LVGFM) over the SLM; a clear indication that the model is more suited to the catchment than the other two. This improvement could possibly be attributed to the model’s ability to incorporate a variable gain factor as opposed to the SLM, which assumes a constant gain factor. In the verification period, however, there is a decline in performance over the SLM as indicated by the negative $r^2$ (-1.3) value. It appears that the SLM performs better than the other linear systems models during the verification period.

The Index of Volumetric Fit (IVF) values are 0.94 and 0.60 during calibration and verification periods respectively. The large deviation during the verification period reflects the model’s tendency to estimate less volume than the observed volume during the verification period.

The MSE during calibration (1.83) is less than those of the other two models. This suggests an improvement over the SLM and LPM respectively during the calibration period. However the MSE during the verification (7.8) is slightly higher than in the case of the other two models suggesting a decline in the model’s performance level.

Figure 5a presents the pulse response ordinates of the LVGFM. From Figure 6c which shows rainfall together with the observed and estimated discharges for the year 1985, we observe that the LVGFM fits the data much better than either the SLM or the LPM. The maximum observed discharge value is about 252 cumecs and it corresponds to the estimated value of about 248 cumecs. These results show that the LVGFM estimated values are much closer to the observed values than in any of the other two models; further suggesting that this model could be more suited for the Nyando catchment than the other two.

![Graphical representation of pulse response ordinates for LVGFM and SMAR models](image)

3.2.4 Soil Moisture Accounting and Routing Model

The results of the chosen model configuration and the optimum water balance parameter values are presented in Table 2. It was observed from the table that the unconstrained value of the objective function was about 0.31 mm$^2$day$^{-1}$. The model efficiency was about 71.2 % during the calibration and 44.0 % during the verification periods respectively. There is a marked improvement of the Soil Moisture Accounting and Routing (SMAR) model over the SLM as shown by the $r^2$ values; about 46 % and 21 % in calibration and verification periods respectively. This improvement, which is obviously larger than that of either the LPM or the LVGFM,
% during the calibration and 44.0 % during the verification periods respectively. There is a marked improvement of the Soil Moisture Accounting and Routing (SMAR) model over the SLM as shown by the $r^2$ values; about 46 % and 21 % in calibration and verification periods respectively. This improvement, which is obviously larger than that of either the LPM or the LVGF.

Figure 6d shows rainfall together with the observed and the SMAR model's estimated discharges for the year 1985. The SMAR model fits the data much better than the linear systems models. The extremes are captured very well. Estimated discharge follows that of observed discharge very closely.

The deviations between the observed and the estimated discharge values are generally not quite serious with this model compared to those of linear systems models; suggesting further that the model has a higher potential than the linear systems models to become the model of choice for the Nyando catchment. It was observed from these results that the conceptual SMAR model performed consistently well for both calibration and verification periods. It performed much better than the linear systems models; showing a marked improvement in both calibration and verification periods, respectively, unlike the linear systems models whose performance were lower than that of SLM in verification.

Figures 6 (a – d): Graphical representation of rainfall; observed and estimated discharges over the Nyando catchment from SLM, LPM, LVGF and SMAR models for the year 1985
A comparison of the results of the SMAR model with those of the linear systems models, as presented in Table 1 columns 5 and 9, show a substantial improvement in the SMAR model performance over that of any of the linear systems models considered in both calibration and verification periods. This is shown by the $r^2$ values, which show an improvement of about 46 % and 21 % over the SLM in calibration and verification periods respectively. These values are higher than those observed with the linear models, especially in verification where the substantive linear models performed worse than the SLM as indicated by the negative values of $r^2$ in columns 7 and 8.

Table 2: Summary of the chosen model configuration parameters under volumetric constraint including the results of the SMAR Model: $R^2$ is the measure of model efficiency; $r^2$ which is the proportion of improvement of the model C is the parameter that controls potential evaporation; H is the parameter that represents the available soil moisture content of the first five layers IVF is the index of volumetric ratio and MSE is the mean square error.

<table>
<thead>
<tr>
<th>Model parameters</th>
<th>Calibration</th>
<th>Verification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Configuration</td>
<td>C</td>
<td>H</td>
</tr>
<tr>
<td>CH</td>
<td>0.8</td>
<td>0.6</td>
</tr>
</tbody>
</table>

3.2.5 Autoregressive Updating

It is observed from the Table 3 that a lead-time of one day provides the highest efficiency in all the three models. The efficiency decreases progressively as we move towards lead-times of more than one day both in calibration and verification periods.

Comparing the simulation and updated mode efficiencies, it is observed that there is a marked improvement in updated mode efficiencies over the simulated ones both in calibration and verification periods for all the linear models with the most notable improvement being that of SLM where the efficiency improved from 47.0 % to 73.1 % during calibration and from 28.8 % to 42.0 % during verification periods respectively.

Table 3: Summary of results of the AR updating results in both calibration (Cal.) and verification (Ver.) periods.

<table>
<thead>
<tr>
<th>Model</th>
<th>Non-updated R^2 %</th>
<th>IVF Cal</th>
<th>IVF Ver</th>
<th>R^2 % updated 1day lead Cal</th>
<th>1day lead Ver</th>
<th>2day lead Cal</th>
<th>2day lead Ver</th>
<th>3day lead Cal</th>
<th>3day lead Ver</th>
<th>IVF Cal</th>
<th>IVF Ver</th>
</tr>
</thead>
<tbody>
<tr>
<td>SLM</td>
<td>47.0</td>
<td>28.8</td>
<td>1.2</td>
<td>6.0</td>
<td>9.0</td>
<td>42.0</td>
<td>0.0</td>
<td>67.0</td>
<td>0.0</td>
<td>6.3</td>
<td>21.7</td>
</tr>
<tr>
<td>LPM</td>
<td>55.0</td>
<td>24.7</td>
<td>1.0</td>
<td>2.0</td>
<td>9.0</td>
<td>37.0</td>
<td>0.4</td>
<td>66.0</td>
<td>0.5</td>
<td>6.2</td>
<td>20.3</td>
</tr>
<tr>
<td>LVGFN</td>
<td>68.7</td>
<td>27.9</td>
<td>0.9</td>
<td>4.0</td>
<td>0.0</td>
<td>39.0</td>
<td>0.9</td>
<td>77.0</td>
<td>0.8</td>
<td>7.5</td>
<td>26.5</td>
</tr>
</tbody>
</table>

Only the results of one-day lead-time are discussed in this paper in comparison to the simulation mode results since this is the lead-time that has the highest efficiency. Moving from one day lead time through to three days lead time, the efficiency decreases progressively from about 73 % to 63 % for SLM during calibration and from about 42 % to 22 % during verification. For the LPM, the efficiency decreases from about 71 % to 63 % during calibration and from about 38 % to 20 % during verification. In the case of LVGFN, the decrease in efficiency is from only about 79 % to 75 % during calibration and from about 40 % to 27 % during verification.

In Figures 7 (a–f) examples of results of estimated and observed discharge in updated
mode for lead times of one and two days are present. The first row represents results of the simple linear model (SLM), the second row represents results of the linear perturbation model (LPM) and the third row represents results of the linearly varying gain factor model (LVGF). It is observed from this figure that the maximum estimated discharge for SLM in updated mode is about 316 cumecs and 219 cumecs at one and two days lead times respectively. The maximum estimated discharge for LPM in this mode is about 334 cumecs and 232 cumecs at one and two days lead times respectively. The corresponding estimates for LVGF are about 377 cumecs and 302 cumecs at one and two days lead times respectively.

Figures 7 (a – f): Updated forecasts for the SLM, LPM and the LVGF over the Nyando catchment at lead times of one and two days respectively.
The corresponding estimates for LVGFMs are about 377 cumecs and 302 cumecs at one and two days lead times respectively.

It is clear from these observations that the deviations from the observed discharge are smaller in updating mode compared to those observed in simulation mode with the smallest deviations occurring in one-day lead-time. Hence the updating seems to have improved the forecasts quite significantly, as they are closer to what is observed. These results show that the three models gave better forecasts in updated mode for all the lead times considered than the simulation mode forecasts.

4. Summary and Conclusion

By and large, the application of the SMAR model shows improved efficiency compared to the linear systems model. The model efficiencies increased from about 47 % to 71 % for the SLM and from about 55 % to 71 % for the LPM, and from about 69 % to 71 % in calibration.

The improvement of the SMAR model over the LVGFMs is not notable in calibration. The improvement was from about 69 % to 71 %. In verification however, there is a marked improvement; from about 28 % to 44 %. The results of the LVGFMs and those of the SMAR model do not differ much during calibration. This is an indication that the two models are equally good in simulating flow during calibration over the Nyando catchment. However the SMAR model may be considered superior on account of its better performance than the LVGFMs during the verification period. In both calibration and verification, it shows a marked improvement over the SLM.

In terms of efficiency, the SMAR model is the best for the Nyando catchment followed closely by the LVGFMs, then the LPM and the SLM respectively. It should also be noted that, whereas the efficiencies of the linear systems models are given in unconstrained mode, the efficiency of the conceptual SMAR model is given in constrained mode. This means that the conceptual model would have performed even better if it were not constrained.

Among the linear systems models, the LVGFMs performed better than the other two models with R² value of about 69 % followed by the LPM at R² value of about 55 % and SLM at R² value of about 47 % in that order during calibration. The SLM, as indicated by the results of the model application, is the most inferior of all the models applied in simulation mode during calibration period. During verification period however, the results show that the LPM is inferior to all the other models. In updating mode, all of the linear models did better than in simulation mode with LVGFMs approaching an efficiency of about 80%.

In conclusion this study has identified two important models for the Nyando catchment; one linear and the other conceptual. The two models, the conceptual Soil Moisture Accounting and Routing (SMAR) model and the linear systems analysis Linearly Varying Gain Factor Model (LVGFMs), can give good forecasts of the flow for the Nyando River using rainfall as input. In particular the LVGFMs is seemingly superior to all other models in its updating mode. Operationalizing this model requires that a forecasting system be put in place by stake holders who include users of rainfall and discharge products and for them to work in partnership. A policy to adopt this model as a flood early warning tool needs to be put in place. This can be useful for the planning and rational use of water supply from the Nyando River. In addition the models can be used as tools for early warning systems for the perennial flooding problem in nyando.
REFERENCES
WMO (1975): Intercomparison of conceptual models used in operational hydrology. Operational hydrology report No. 7 WMO No. 429, 1-25