Validation of SWAT Model for Prediction of Water Yield and Water Balance: Case Study of Upstream Catchment of Jebba Dam in Nigeria

Adeniyi G. Adeogun, Bolaji F. Sule, Adebayo W. Salami, Michael O. Daramola

Abstract—Estimation of water yield and water balance in a river catchment is critical to the sustainable management of water resources at watershed level in any country. Therefore, in the present study, Soil and Water Assessment Tool (SWAT) interfaced with Geographical Information System (GIS) was applied as a tool to predict water balance and water yield of a catchment area in Nigeria. The catchment area, which was 12,992 km², is located upstream Jebba hydropower dam in North central part of Nigeria. In this study, data on the observed flow were collected and compared with simulated flow using SWAT. The correlation between the two data sets was evaluated using statistical measures, such as, Nash-Sucliffe Efficiency (NSE) and coefficient of determination (R²). The model output shows a good agreement between the observed flow and simulated flow as indicated by NSE and R², which were greater than 0.7 for both calibration and validation period. A total of 42,733 mm of water was predicted by the calibrated model as the water yield potential of the basin for a simulation period between 1985 to 2010. This interesting performance obtained with SWAT model suggests that SWAT model could be a promising tool to predict water balance and water yield in sustainable management of water resources. In addition, SWAT could be applied to other water resources in other basins in Nigeria as a decision support tool for sustainable water management in Nigeria.

Key words—GIS, Modeling, Sensitivity Analysis, SWAT, Water Yield, Watershed level.

I. INTRODUCTION

The knowledge of water balance and water yield in a river catchment is an indispensable prerequisite in the sustainable management of water resources at watershed and basin wide levels. Stehr et al. (2008) [1] reported that the study of water resources at river catchment level has been widely adopted as a better way of managing and assessing these important natural resources. At the decision making stage, models are usually employed for the purpose of selecting an optimal courses of action. Such models are often constructed to enable reasoning within an idealized logical framework about the processes [2]. Due to the complexity in the representation of these natural processes and conditions, models are usually calibrated prior to the application of the models to obtain a realistic description of the processes match with the reality.

In recent times, the use of Geographical Information System (GIS) and remote sensing technique has improved the application of hydrological modeling in many capabilities, such as, in the area of data management, parameter extraction and interpolation, visualization and interface development [1]. Hence, the use of modeling tool interfaced with GIS provides the platform to streamline GIS processes tailored towards hydrological modeling.

Various physically based hydrological models have been identified and used to simulate hydrological processes in a river catchment [3]. One of the promising candidates of the models is SWAT. SWAT has been adjudged by researches as computationally efficient in its prediction [4]. Against this background, the main objective of this study is to identify challenges and prospects of using SWAT in the prediction of water yield within a river catchment. The specific objective of the study was to model the hydrology and predict the water yield and balance of a selected catchment in Nigeria using SWAT. Through the application of SWAT for hydrological modeling and prediction of water yield and balance, some challenges and prospects of applying SWAT to predict basin water characteristics within a river catchment could be identified.

II. DESCRIPTION OF CASE STUDY AREA

The selected catchment area diagnosed in this study was the upstream watershed of Jebba lake located in the central area of Nigeria between Lat 10.31N and Long 5.01E and Lat 8.99N Long 4.79 E (see Fig. 1 for details). The watershed has a perimeter of about 567 km and an estimated area of 12,992 km². Major rivers and tributaries within the watershed are Rivers Niger, Awun, Moshi, Eku, Kotongora and Oli. The range of elevation of the watershed was between 114 m to 403 m above sea level, and the average monthly discharge at Jebba station situated at the outlet of the watershed was 1053 m³/s for the period 1984 to 2008. The minimum and maximum monthly discharge during the period were 378 m³/s and 26,664 m³/s, respectively [5]. The watershed area is...
sandwiched between two main hydropower reservoirs in Nigeria, namely Kainji and Jebba reservoir in the North-Central zone of Nigeria. Villages within the watershed area are Zugurma, Ibhi, Patiko, Felegi (custodian of Kainji Lake National Park) and Sabonpegi. The selection of the area for this study (using the SWAT model) was based on the availability of required input data at the hydrological stations established by Kainji and Jebba hydropower power stations and at Nigeria Metrological Agency (NIMET). In addition, the watershed plays a significant role in the national energy supply since it contributes significantly to the water flow into Jebba Lake where a power plant of 764 MW capacity is installed for electricity generation.

III. MATERIALS AND METHOD

A. SWAT Model Description

SWAT was originally developed by the United States Department of Agriculture (USDA) to predict the impact of land management practices on water, sediment and agricultural chemicals yields in large un-gauged basins [6]. The SWAT model is a catchment-scale continuous time model that operates on a daily time step with up to monthly or annual output frequency. The model operates by dividing a catchment into sub-catchments and each sub-catchment is connected through a stream channel and further divided into a Hydrological Response Unit (HRU). The HRU is a unique combination of a soil and vegetation types within the sub-catchment. The model calculation was performed on a HRU basis and flow and water quality variables were routed from HRU to sub-basin and subsequently to the catchment outlet. The simulation of hydrological cycle by SWAT is based on the water balance as in (1):

\[ SW_t = SW_i + \sum_{i=1}^{n} (R_{day} - Q_{surf} - E_a - W_{seep} - Q_{gw}) \]  

where \( SW_t \) is the final soil water content (mm water); \( SW_i \) is the initial soil water content in day \( i \) (mm water); \( R_{day} \) is the time of precipitation in day \( i \) (mm water); \( Q_{surf} \) is the amount of surface run-off in day \( i \) (mm water); \( E_a \) is the amount of evapo-transpiration in day \( i \) (mm water); \( W_{seep} \) is the amount of water entering the vadose zone from the soil profile in day \( i \) (mm water); and \( Q_{gw} \) is the amount of return flow in day \( i \) (mm water).

Water yield is the total amount of water leaving the HRU and entering main channel during the time step. It is one of the important parameters that need to be estimated for sustainable management of water resources of the study area. Water yield of a river catchment is estimated by the model using (2):

\[ WYLD = SURQ + LATQ + GWQ - TLOSS \]  

where WYLD is the amount of water yield (mm H₂O), SURQ is the surface runoff (mm H₂O), LATQ is the lateral flow contribution to stream flow (mm H₂O), GWQ is the groundwater contribution to stream flow (mm H₂O) and TLOSS is the transmission losses (mm H₂O) from tributary channels in the HRU via transmission through the bed. The estimation of surface runoff can be performed by the model using two methods. These are the SCS curve number procedure USDA Soil Conservation Service (3) and the Green & Ampt infiltration method [4].

\[ Q_{surf} = \frac{(R_{day} - 0.24)^2}{(R_{day} + 0.85)} \]  

In (3), \( Q_{surf} \) is the accumulated runoff or rainfall excess (mm), \( R_{day} \) is the rainfall depth for the day (mm), \( S \) is the retention parameter (mm). The retention parameter \( S \) and the prediction of lateral flow by SWAT model are defined in (4) and (5), respectively.

\[ S = 25.4 \left( \frac{100}{CN} - 10 \right) \]  
\[ q_{lat} = 0.024 \left( \frac{2SSC_s \sin \theta_a \cdot \theta_d}{\theta_d} \right) \]  

where, \( q_{lat} \) is lateral flow (mm/day); \( S \) is drainable volume of soil water per unit area of saturated thickness (mm/day); \( SC \) is saturated hydraulic conductivity (mm/hr); \( L \) is flow length, \( \alpha \) is slope of the land, \( \theta_d \) is drainable porosity.

The estimation of the base flow was done using (6):

\[ Q_{gwj} = Q_{gwj-1} \cdot e^{-\alpha_{gw} \cdot \Delta t} + w_{rchrg} \cdot (1 - e^{-\alpha_{gw} \cdot \Delta t}) \]  

where \( Q_{gwj} \) is groundwater flow into the main channel on day \( j \); \( \alpha_{gw} \) is base flow recession constant; \( \Delta t \) is time step.

B. Input Data Collection and Processing

Model inputs required to run SWAT include the Digital Elevation Model (DEM), land use map, soil map and weather data (see Table I). The Map window GIS [7] interface of the MWSWAT model was used to discretize the catchment area and extract the SWAT input files. The 90 m resolution topography data used for this study was extracted from the Shuttle Radar Topography Mission (SRTM) final version developed by [8]. The final DEM (see Fig. 2) of the watershed was used to delineate the catchment, and to provide topographical parameters, such as, overland slope, stream network and slope length for each basin. The upstream catchment area of Jebba Lake was delineated and discretized into 77 sub-basins and 107 Hydrological Response Units (HRU) (each with unique combination of land use, slope and...
soil). Division of sub-basins into areas having unique land use, soil and slope combinations makes it possible to study the differences in evapo-transpiration and other hydrological conditions for different land covers, soils and slopes [9].

### Table I

<table>
<thead>
<tr>
<th>Data type</th>
<th>Description</th>
<th>Resolution</th>
<th>Remark</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topography</td>
<td>Digital Elevation Model</td>
<td>90mx90m</td>
<td>Shuttle Radar Topographical Mission</td>
</tr>
<tr>
<td>Land Use Map</td>
<td>Land Use Classifications</td>
<td>1km</td>
<td>Global Land Cover Classification, Satellite Raster</td>
</tr>
<tr>
<td>Soil Map</td>
<td>Soil Types and Texture</td>
<td>10km</td>
<td>Digital Soil Map of the World</td>
</tr>
<tr>
<td>Weather</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table II

<table>
<thead>
<tr>
<th>S/N</th>
<th>SWAT Code</th>
<th>Description</th>
<th>Area(Ha)</th>
<th>% of Watershed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>URMD</td>
<td>Urban and Built-Up Land</td>
<td>129.92</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>CRDY</td>
<td>Dryland Cropland and Pasture</td>
<td>332.8</td>
<td>0.03</td>
</tr>
<tr>
<td>3</td>
<td>CRGR</td>
<td>Cropland/Grassland Mosaic</td>
<td>4885.48</td>
<td>0.38</td>
</tr>
<tr>
<td>4</td>
<td>CRWO</td>
<td>Cropland/Woodland Mosaic</td>
<td>2109.11</td>
<td>0.16</td>
</tr>
<tr>
<td>5</td>
<td>GRAS</td>
<td>Grassland</td>
<td>915.2</td>
<td>0.07</td>
</tr>
<tr>
<td>6</td>
<td>SHRB</td>
<td>Shrubland</td>
<td>1863.67</td>
<td>0.14</td>
</tr>
<tr>
<td>7</td>
<td>SAVA</td>
<td>Savannah</td>
<td>1257234.26</td>
<td>96.77</td>
</tr>
<tr>
<td>8</td>
<td>FOEB</td>
<td>Evergreen Broadleaf Forest</td>
<td>166.4</td>
<td>0.01</td>
</tr>
<tr>
<td>9</td>
<td>WATB</td>
<td>Water bodies</td>
<td>20164.26</td>
<td>1.55</td>
</tr>
<tr>
<td>10</td>
<td>BSVG</td>
<td>Barren or Sparsely Vegetated</td>
<td>11435.27</td>
<td>0.88</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>1299233.67</td>
<td>100.00</td>
</tr>
</tbody>
</table>

### Results and Discussion

#### Landuse map of the Global Land Cover Characterization (GLCC) database was used to estimate vegetation and other parameters representing the watershed area. The GLCC database was developed by [10] and has a spatial resolution of 1Km and 24 classes of land use representation. Also, the watershed area was visited to obtain on-site information about the land use and land cover of the area. Information gathered during the site visit was used to update the GLCC database in order to arrive at the final map for the study area. Table II shows the land use and land cover types and their approximate percentage area coverage for the upstream watershed of Jebba reservoir.

Digital soil data for the study was extracted from harmonized digital soil map of the world (HWSD v1.1) produced by Food and Agriculture Organization (FAO) of the United Nations [11]. The digitized soil map was completed in January 2003 and the database provides data for 16,000 different soil mapping units, containing two layers (0 - 30cm and 30 - 100cm depth). Soil units were then extracted from the database and completed by additional information gathered by analyzing 16 soil samples collected from different locations within the watershed area from two different layers (that is, (0 - 30cm and 30 - 100cm depth). The samples were analyzed and the results used to update the database of the soil map used for the modeling study.

Weather variables, such as, rainfall, temperature (maximum and minimum), solar radiation and humidity used for driving the hydrological balance within the watershed are from the period 1985 to 2010 for 3 meteorological stations distributed spatially over the basin. This data was obtained from Nigeria Meteorological Agency (NIMET), Jebba and Kainji hydroelectric stations, and employed in the simulation using the SWAT model. In the case of insufficient required data (e.g. weather conditions) for the simulation, a weather generator embedded in the SWAT model developed by [12] was employed to generate the required data.

#### C. Model Application

The configuration and set-up of MWSWAT model starts with the projection of all required spatial datasets to the same projection called UTM Zone 31N Northern Hemisphere for the selected catchment area in Nigeria. The configuration of the model involves the settings of the simulation period (start and finish date) and the selection of weather sources from the SWAT database. In addition, selection of the method for the estimation of surface run-off (Curve Number or Green and Ampt method), channel water routing (variable or Muskingum method), and potential evapo-transpiration (Priestley, Penman-Monteith and Hargreaves) is available. In the simulation study with SWAT presented in this study, Run-off Curve Number method was employed to estimate surface run-off from precipitation, the Hargreaves method for estimating potential evapo-transpiration generation, and the variable storage method to simulate channel water routing. The simulation period was for the case study was from 01 January 1985 to 31 Dec. 2010, and all necessary files needed to simulate SWAT were written at this level, and the appropriate selection of weather sources were done before running the SWAT executables.

### IV. Results and Discussion

#### A. Sensitivity Analysis of Modeling Parameters

During hydrological simulation via SWAT model, several parameters are refined via a calibration and validation
process, that is often cumbersome and time consuming. Therefore to save time and effort, sensitivity analysis of the model parameters is always carried out prior to simulation to identify and rank the parameters that have significant impact on specific model output [9]. In this study, the sensitivity analysis for 27 model parameters was carried out via auto calibration based on the parameter solution optimization method. The results showed that the most sensitive parameters for hydrological modeling of upstream watershed of Jebba dam are CN2, GWQMN, ESCO and SOL_AWC (see Fig. 3 for detailed explanation of these parameters). This result is in agreement with those found by many similar studies [1], [12], [13], confirming that these four parameters are the crucial sensitive parameters for water balance and stream flow.

![Fig. 3 Relative sensitivity and description of modeling parameters](image)

In Fig. 3, CN2: Initial SCS CN II value; GWQMN: Threshold water depth in shallow; ESCO: Soil evaporation compensation factor; SOL_AWC: Soil available water capacity; CH_K2: Channel effective hydraulic conductivity; Sol_z: Soil Depth; GW_REV: Groundwater “revap” coefficient; epco: Plant uptake compensation factor; Blai: Maximum potential leave area index; Alpha_Bf: Base flow Alpha factor.

### B. Performance Evaluation, Calibration and Validation of Model

Model performance evaluation is necessary for the verification of the robustness of the model. In this study, performance evaluation of the model was carried out based on the guidelines obtained elsewhere [14]. The guidelines, which is based on statistical methods, use parameters such as Coefficient of Determination (R\(^2\)) an written in (7) and Nash-Sutcliffe Efficiency (NSE) (8) to assess the goodness of fit between the observed and simulated parameters. According to the guidelines, the model output is considered satisfactory if NSE> 0.5. In addition, the closer the R\(^2\) to 1, the better is the goodness of fit between the observed and simulated parameters [15].

\[
R^2 = \frac{\sum_i (Q_{mi} - \bar{Q}_m)(Q_{si} - \bar{Q}_s)}{\sum_i (Q_{mi} - \bar{Q}_m)^2 \sum_i (Q_{si} - \bar{Q}_s)^2}
\]  

(7)

\[
NSE = 1 - \frac{\sum_i (Q_{mi} - Q_{si})^2}{\sum_i (Q_{mi} - \bar{Q}_m)^2}
\]  

(8)

where \(Q_m\) is the measured discharge, \(Q_s\) is the simulated discharge, is the average measured discharge and \(\bar{Q}_s\) is the average simulated discharge.

Evaluation of the performance of the model was done by comparing the observed and simulated monthly inflow at the Jebba gauge station for both the calibration and validation periods. In total, 14 parameters were selected to be calibrated through the Parasol optimization method. The model was calibrated with the observed monthly inflow of Jebba Lake from 1990 to 1992, and cross-validated with another set of independent data set from 1993 -1995. The results of the calibration and validation exercise are presented in Figs. 4 and 5 for calibration, while Figs. 6 and 7 are for the validation period. As it can be seen from the Figures, there is a good correlation between the observed flow and the simulated flow, indicated by NSE and R\(^2\) of 0.76 and 0.72, respectively, for calibration period, and NSE and R\(^2\) of 0.70 and 0.78, respectively, for the validation period. In addition, the correlation coefficient of 0.85 for calibration data and correlation coefficient of 0.88 for the validation data indicate that the experimental data are reliable. Furthermore, the data for the calibration and the validation exercises are within the confidence interval of 95% (see Figs. 4 and 6).

![Fig. 4 Simulated versus observed monthly flow during the calibration period (1990-1992)](image)
C. Prediction of Water Yield of the Basin

In SWAT modeling, water yield can be defined as the total amount of water leaving the HRU and entering main channel during the time step [16]. It is one of the important parameters to be estimated for efficient water management and planning of the case study area. The contributions of each sub-basins in the watershed area to water yield during the period of simulation period was examined using the calibrated SWAT model. It was noted that sub-basin 16 with catchment area of 169.3km² has the highest contribution of 650.26mm to water yield of the area during the simulation period. In addition, the lowest water yield value was received from sub-basin 24 with catchment area of 505 km². Further analysis of the results also revealed that the northern part of the study area, which has Niger, Kotongora and Oli as the major rivers, has contributed larger percentage of the water yield in the area. Furthermore, the results showed that the lowest contribution to the water yield came from the western part of the catchment which have rivers Awun and Moshi as the major rivers. A total of 42,733 mm of water was estimated by the model as the potential water yield of the basin between the simulation period of 1985-2010. Fig. 8 depicts the pictorial view of the contributions of each sub-basin to the average water yield of the area during the simulation period.

D. Estimation of Water Balance

Water balance is the driving force behind all the processes in SWAT because of its impacts on plant growth and the movement of sediments, nutrients, pesticides, and pathogen within the watershed area [17]. In order to deal with water management issues, it is ideal to analyze and quantify the different elements of hydrological processes occurring within the area of interest. Understanding the spatial and temporal variation and interaction of these hydrologic components could be instrumental to assisting water planners in the formulation of strategies for water conservation. Reference [18] asserted that the most important elements of water balance in a basin consist of precipitation, surface run-off, lateral flow, base flow and evapo-transpiration. All these elements, with the exception of precipitation, have to be predicted using appropriate modeling tool because their quantification by measurement is not easy Therefore, SWAT model was used to quantify each of the hydrological processes occurring in the study area considered in this study. Analysis of the results showed that evapo-transpiration has the highest
share of the water balance with values ranging from 34.85% (2007) to 71.15% in the year 1998. High evapo-transpiration rate predicted could be attributed to the type of vegetation cover (mostly savannah) and high temperature associated with the area. Lateral flow has the lowest percentage with values between 0.18% (1998) to 0.39% (2003). Deep aquifers recharge in all cases is very low with percentage variation of 2-10% of the total rainfall. The implication of this is that the water yielding potential of the deep aquifers for the study area will be quite minimum. See Fig. 9 for predicted water balance of selected years in the watershed.

**Fig. 9 Predicted Water Balance of selected years in the watershed area.** A. for 1985; B for 1989, C for 1995, D for 1998, E for 2003 and F for 2007

**V. CONCLUSION**

This study focused on the prediction of water yield and balance upstream catchment of Jebba Lake in North Central part of Nigeria with the use of SWAT 2009, a physically based semi-distributed hydrological model interfaced with MapWindow GIS software. The preparation of thematic maps and database necessary for the successful running of the model was done using the GIS components. The model was simulated daily for a period of 26 years (1985 to 2010) and the performance evaluation of the model was carried out using Nasch Sutcliffe Efficiency (NSE) and Coefficient of Determination ($R^2$). The Nash and Sutcliffe Efficiency (NSE) and Coefficient of Determination were 72% and 76%, respectively, for calibration, and 70% and 78%, respectively, for validation periods, indicating a satisfactory simulation.
Sensitivity analysis of modeling parameters was also examined. The most sensitive parameters for hydrologic modeling using SWAT are curve number CN2, threshold water depth GWQMN, soil evaporation compensation factor ESCO and soil available water capacity SOL_AWC.

Overall modeling results suggest that SWAT model is potentially useful in studying the hydrology and predicting water yield of river catchments. The results is also an indication that SWAT model embedded in GIS environment is highly prospective in its usage as a tool to support policies and decision making by relevant authorities in Nigeria for the sustainable water management at watershed level. The utility of the remote sensing data and GIS in the creation and generation of the required maps necessary to set up and simulate the hydrological models in Nigerian catchments is also demonstrated in this study.

However, one of the major issues encountered in the application of SWAT model in developing countries, like Nigeria, is the scarcity or unavailability of required data as model inputs. The use of hybrid data, which involves combining local and in-situ data gathered from local agencies and global data obtainable from global database, has been adopted to overcome these challenges in this study. For example, the land use map extracted from the Global Land use Classification (GLCC) cannot adequately represent the present land use scenario of the study area, due to reported cases of massive deforestation by the local habitants in the present land use classification of the study area. Unavailability of a land use map that could adequately represent the present situation of the study area could have an effect on the modeling results of the area if measures are not taken to properly factor it in to the land use map of the area.

Furthermore, this study has revealed that poor representation of catchment parameters, most especially the precipitation of study area as well as uneven rainfall station distribution, can contribute to poor performance of SWAT model. Extra effort is therefore required to ensure that the input weather data are of high quality while spatial data should be of high resolution for better hydrological prediction.

REFERENCES


