Robust method for estimating grain yield in western Kenya during the growing seasons
Edward M. Mugalavai and Emmanuel C. Kipkorir

ABSTRACT

Uncertainties caused by climate change and population explosion require suitable methods for estimating grain yield during the growing seasons. This paper evaluates the applicability of the AquaCrop model in the region of western Kenya. The objectives of the study were to: simulate the long-term maize crop yields for the region using AquaCrop model for variable climate scenarios, and estimate the expected yield for the ongoing season. Climate was classified into below normal ($<x̅/C01/∂$), normal (between $x̅/C01/∂$ and $x̅+1/∂$) and above normal ($>x̅+1/∂$) conditions based on the Kenya Meteorological Department (KMD) convention. Simulation of grain yield was based on model calibration results, periodic KMD forecasts and the long-term mean for the seasons. The calibrated model is able to estimate both long-term seasonal grain yield and expected harvest for the ongoing season based on climatic conditions that are compared with the long-term seasonal characteristics and complemented by meteorological forecasts. The ongoing season yield simulation was based on persistence theory of Markov processes whose results strongly correlated ($r=0.9$) with actual seasonal observed yield.

Key words | AquaCrop, climate, crop yields, food security, Markov processes, simulation

INTRODUCTION

Studies aimed at planning food security requirements within the growing seasons are often faced with challenges arising from the scantly documentation of yield data. Two main types of crop growth simulation models can be exploited to estimate the expected yield for several growth conditions. Mechanistic models incorporate detailed mathematical descriptions of each growth controlling process as presently understood, and such models have been used in the following studies (Robertson et al. 2001; Batchelor et al. 2002; Hartkamp et al. 2002; Brisson et al. 2003; Wang et al. 2003; Ziaei & Sepaskhah 2003; Yang et al. 2004). Empirical models, on the other hand, represent the same processes based on empirically derived functions.

Empirical crop models that have been developed and used include EPIC (Williams et al. 1989), CropSyst (Stockle et al. 2003), the DSSAT cropping system model (Jones et al. 2003), the Wageningen models (Van Ittersum et al. 2003), the APSIM models (Keating et al. 2003) and the AquaCrop water productivity model (Raes et al. 2009a), among others. Crop growth simulation models usually differentiate between the effects of water stress on photosynthesis (or biomass production, in the more empirical model), leaf area growth and harvest index. The AquaCrop model, which is a water productivity model, is preferred due to its accuracy, simplicity, robustness and ease of use. The model converts daily transpiration directly to daily biomass production, using daily reference evapotranspiration (ETo) and normalized water productivity for biomass (Steduto et al. 2009).

AquaCrop software requires input data consisting of climatic data, crop data, soil data and management data to make simulations (Raes et al. 2009a). The model contains a complete set of characteristics that can be selected and adjusted for different soil or crop types. AquaCrop is able to: assess crop water stress under rainfed conditions; estimate yield response to water; determine crop water productivity that can be used to design irrigation schedules; and evaluate
irrigation strategies (Raes et al. 2006, 2009a, b; Steduto et al. 2006). Crop growth models simulate both yield and crop development throughout the growing cycle. Research has shown that yield predictions issued later in the growing season provide more accurate results compared to those issued earlier because they incorporate more actual weather conditions (Kipkorir et al. 2011). The AquaCrop model does its calculations through processes that are susceptible to water stress conditions (Hsiao et al. 2009).

The AquaCrop model simulates the change of water stored in the soil throughout the growing season (Steduto et al. 2009). The water content in the root zone determines the canopy development and the corresponding crop transpiration which is converted in AquaCrop into biomass production and yield formation. Canopy transpiration is modelled in a stepwise procedure using the Kc-ETo approach (Allen et al. 1998). The reference evapotranspiration-ETo is a model input obtained from weather data by means of the FAO Penman-Monteith procedure (Steduto et al. 2009).

The conceptual basis of AquaCrop model and its structure and algorithm are found in the companion papers (Steduto et al. 2009; Raes et al. 2009a, b). Calibration of the AquaCrop model has been done for some crops (Farahani et al. 2009; Geerts et al. 2009; Hsiao et al. 2009; Heng et al. 2009; Andarzian et al. 2011; Abedinpour et al. 2012; Mkhabela & Bullock 2012) using single location data sets. Attempts to calibrate AquaCrop on a global scale are still ongoing through data collection from locations of diverse climate and soil conditions, hence the choice of the model in the current study. This paper aimed at evaluating the applicability of the AquaCrop model for grain yield simulation during the growing seasons in the region of western Kenya.

STUDY AREA AND METHODOLOGY

Study area

This study was carried out in an area approximately of 48,000 km² which is bordered by Uganda to the west, Tanzania to the south and Lake Victoria basin boundaries to the north and the east (Figure 1). The Lake Victoria basin region lies in the western part of Kenya between 1°30’N and 2°00’S and between 34°00’E and 35°45’E. The region is an area of high agricultural potential for both subsistence and plantation farming systems mainly under rainfed conditions.

Climatic data requirements

Sampling of stations was based on the classification of the Lake Victoria basin of Kenya into eight homogeneous zones using seasonal climatic clusters (Figure 1). Sample stations for collection of both meteorological and agrometeorological parameters were spatially distributed in the basin based on homogeneous zonation of the region (Ogallo 1980).

Secondary data collection

Daily rainfall, daily minimum and maximum temperature, sunshine hours and wind speed records (2 m level) were collected from sampled stations in the lake basin. Most of the historical data were obtained from the Kenya Meteorological Department (KMD). For stations that collect most climatic parameters, the ETo calculator (FAO 2009), based on procedures outlined in FAO No. 56, was used to determine reference evapotranspiration (Allen et al. 1998). The climatic parameters used in this study included: sunshine hours, wind speed, maximum and minimum temperatures and mean temperature due to availability of data. It was observed that most stations lacked solar radiation data.

Primary data collection

Crop and soil data were collected for the entire period of field work that covered the rainy seasons’ long and short rains. Crop data (maize H614) was collected from the time of sowing to harvesting at intervals of 10 days (decadal) for Kakamega-KARI and Moi University, Eldoret stations where the experimental sites were located.

Experimental design for crop data collection

Experimental sites were set up at Moi University, Eldoret and Kenya Agricultural Research Institute (Kakamega-KARI). At Moi University site, a representative block was
selected for crop and soil data collection. One plot was identified and four replications marked for data collection; plant density was determined at the initial stage. Each replication had five rows of 60 cm each with a length of 4 m making an area of about 64 square metres (Figure 2). The centres for each replication were determined and plant characteristics were obtained around the marked centres.

At Kakamega-KARI an experimental site was also set up. The site comprised five rows with spacings of 75 cm each. The procedure for marking the centres of the plots was similar to that used at Moi University. For each of these experiments the characteristics determined at decadal intervals were canopy cover and above ground biomass. To determine dry biomass, destructive sampling was involved.
where at least two plants were identified from each plot for use in determining dry biomass. The grain yield (tons/ha) at 13% moisture level was determined after harvesting. Crop data collected were used as input parameters in the AquaCrop model calibration.

**Canopy cover**

The canopy cover was estimated using a metre rule which was placed at the determined centres of each replication at decadal intervals. The shaded portions of the (100 cm) metre rule indicated the canopy cover as a percentage proportion of the full length. For consistency in the observations, the time for taking the canopy cover records was maintained for each plot on every field visit. The value recorded was obtained by determining the average canopy cover along the diagonals of each replication. This was done for each of the four replications and the mean value taken to represent the canopy cover of a particular plot.

**Above ground biomass**

Destructive sampling that involved identification of two representative plants from each plot was done. To ensure that the yield from the experimental plot remained intact, the rows on either side of the plot were considered for destructive sampling. This procedure was repeated on decadal time steps up to the plant maturity stage. The samples were dried in the oven for at least 48 hours to obtain a moisture content of about 13%. The dry biomass was then weighed and computations done to determine the weight per plant for each plot. The dry biomass weight for each plot with known area was obtained by considering the plant density earlier recorded at the start of the observations. The weight was then converted into metric units (tons/ha) for use in the computations.

**Soil data**

**Soil particle size analysis**

To obtain soil data at the experimental sites, the particle size analysis (Bouyoucos 1962) method was used to estimate the percentage of sand, silt, and clay contents of the soil samples collected from all the experimental plots. Based on the proportions of different particle sizes, soil textural categories were assigned to the samples. The method of silt and clay measurement relies on the effect of particle size to different settling velocities within a water column. Settling velocity is a function of liquid temperature, viscosity and specific gravity of the falling particles. To ensure ideal conditions, corrections were made for the temperature of the liquid samples. Three samples were collected from each of the experimental plots at 10 cm, 20 cm and 30 cm depths, respectively.

**Soil texture determination**

The soil texture for each of the varying soil depths was determined using the sand, silt and clay distribution obtained. The soil at these depths was assigned a texture class based on the soil textural triangle. The textural triangle (Saxton et al. 1986) contains various soil textures which depend on the relative proportions of the soil particles.

**Soil-water characteristics**

The soil parameters required as inputs in the AquaCrop model are the soil water content in volume % at saturation ($\theta_{\text{SAT}}$), field capacity ($\theta_{\text{FC}}$) and at wilting point ($\theta_{\text{WP}}$). Indicative values of the water content for soil types identified in the study area were obtained by means of a pedotransfer function (Saxton et al. 1986, 2005) by considering the soil type.

**Calibration of AquaCrop**

Calibration of AquaCrop model was done based on data collected from the two experimental sites in the western Kenya region. Since the basin has two major zones with different cropping patterns, Kakamega-KARI was used for the western zone while Moi University was used for the Rift Valley zone. The crop characteristics used for calibration included canopy cover, above ground biomass and grain yield. Both experiments were set up during the year 2009 growing season. The harvest index adopted in simulating yield for the experimental sites and other stations in the
region was determined at the end of the growing season (2009) after harvesting and drying the maize grain to the recommended 13% moisture level using the relationship:

\[ Y = HI \times B \]  

(1)

where \( Y \) is harvested grain yield, \( HI \) is harvest index and \( B \) is dry biomass.

**Conservative and cultivar specific parameters**

The generalized model input parameters for maize reported by Hsiao *et al.* (2009) and validated by Heng *et al.* (2009) to be conservative were adopted and used for all simulations. Only parameters considered to vary with cultivar and environment were considered for adjustment depending on availability of data for the specific parameters. The parameters specified during calibration included soil parameters (soil water content at field capacity (\( \theta_{FC} \)), soil water content at permanent wilting point (\( \theta_{WP} \)) and soil depth), maximum canopy cover (CCx), plant density, maximum rooting depth (\( Z_r \)), length of growth cycle and reference harvest index (\( HIo \)).

**Validation of AquaCrop**

Validation was based on grain yield results by holding all model parameters determined in the calibration stage constant. To determine the data used for validating yield, simulations were first run based on the climate, crop and soil characteristics using all the available data for both Moi University and Kakamega-KARI experimental sites. The simulations covered the entire periods for which climate and soil data were available. For the Moi University site the period (2001–2009) was considered while for Kakamega-KARI simulations were done over the period (1982–2009). The simulations identified three seasons from each site during which actual harvested yield compared well with their simulated yields. Based on these simulations the seasons (1998, 2001, 2005) and (2001, 2003, 2006) for Kakamega-KARI and Moi University, respectively, were adopted together with the 2009 season for use in validating yield.

**Rainfall time series analysis**

The sequences for above normal, normal and below normal rainfall years were determined for eight representative stations in western Kenya based on the homogeneous zones using a simple statistical method (KMD 2008, 2009). Annual and seasonal statistics were derived by use of standard deviation. For the annual statistics, the long-term annual seasonal mean (\( \bar{x} \)), standard deviation (\( \sigma \)), (\( \bar{x} - \sigma \)) and (\( \bar{x} + \sigma \)), were computed for each station. Using the annual seasonal rainfall series established, years whose rainfall fell above (\( \bar{x} + \sigma \)) were classified as wet years (above normal rainfall), those between (\( \bar{x} - \sigma \)) and (\( \bar{x} + \sigma \)) were normal years and those below (\( \bar{x} - \sigma \)) were drought years (below normal). Varying the data about the mean by use of standard deviation ensures that most data fall within the normal distribution.

**YIELD ESTIMATION DURING THE GROWING SEASON**

The climatic conditions generated from the rainfall time series analysis done using previous seasonal climatic records is important in simulating yield during the growing season. The climate for the ongoing season depends on rainfall records collected at the time of releasing the quarterly forecast by KMD which is then applied based on persistence theory (Whiton 1977) to estimate the projected yield for that particular season. According to this theory, the climate used to simulate yield for the ongoing season is assumed to be similar to that of the immediate preceding season.

**RESULTS AND DISCUSSION**

**Soil particle analysis**

The results obtained from the analysis of soil samples collected from experimental plots are presented in Table 1. Table 1 shows the samples for each of the plots for the three levels analysed, the liquid room suspension temperature and its corrected value and the percentage proportions for sand, silt and clay, respectively. The corresponding soil type obtained using the soil textural triangle
is also provided in the last column. Soil types for stations not analysed were adopted from Kenya Soil Technical Report—KENSOTER (Batjes & Gicheru 2004).

**Soil-water characteristics**

The generalized soil-water characteristics obtained using the soil texture results assisted in determining the wilting point, field capacity and the soil saturation levels for the experimental plots (Saxton et al. 1986). Table 2 shows the soil water characteristics for the experimental plots.

**AquaCrop calibration (Kakamega-KARI)**

**Model calibration**

The model was calibrated based on canopy cover, dry biomass and grain yield (2009 season). The secondary data used in the simulations were obtained from relevant agricultural institutions (KARI, Agriculture Ministry) and Moi University Farm Department. The model calibration results for canopy cover and biomass are presented in the following sections.

**Canopy cover results**

Data for canopy cover (Kakamega-KARI) was collected from the development stage through the maturity stage. The canopy cover development curve was obtained using the data collected during the 2009 growing season. Maximum canopy cover of 78% (observed) was obtained and this was adopted to calibrate the model for the western zone. Statistical analysis of the canopy cover data gave an $R^2$ value of 65% (Figure 3). The other crop calibration parameters for Kakamega-KARI in western Kenya are shown (Table 3).

**Biomass results**

Dry biomass data for a single plant (Kakamega-KARI) were derived and converted into the total biomass for the experimental plot using plant density and the weight was given...
in tons/ha. The data collected were plotted against the simulated biomass attained by inputting the canopy cover obtained after canopy calibration, soil data, and the climate for Kakamega-KARI station (2009 growing season) into AquaCrop and it gave an $R^2$ value of 78% (Figure 4).

Model validation

The experiments for collecting data used in AquaCrop calibration and validation covered one growing season (2009). Due to this limitation, it was not possible to employ all the parameters (canopy cover, dry biomass and grain yield) used for calibration in the validation process. Therefore, only grain yield was used to validate the model since secondary data were obtainable from the relevant government agricultural institutions (Kakamega-KARI, Moi University farm and Agriculture Ministry) found in the region. The observed and simulated results for the years used in validating the model are shown (Table 4). The observed data were plotted versus the simulated data obtained after yield calibration and the results are shown (Figure 5). The statistical parameters used to establish model performance during calibration were: EF (Nash–Sutcliffe model efficiency coefficient (Nash & Sutcliffe 1970)); the root mean squared error (RMSE); and $R^2$ which is the goodness of fit coefficient and the slope (Figure 5).

AquaCrop calibration (Moi University)

Model calibration

A similar procedure used for model calibration for Kakamega-KARI was adopted and applied to this station. Calibration was based on canopy cover, dry biomass and grain yield then validation was done using grain yield alone. This was due to lack of dry biomass and canopy cover data for other seasons. Validation was based on grain yield data obtained from secondary data records from relevant agricultural institutions (KARI, Agriculture Ministry and Moi University farm).

Canopy cover results

Plant characteristics were obtained around the marked centres and the mean value from four replications recorded on every field visit. For consistency in data collection, the

<table>
<thead>
<tr>
<th>Description</th>
<th>Moi University (Primary)</th>
<th>Kakamega-KARI (KARI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial canopy cover</td>
<td>0.36</td>
<td>0.29</td>
</tr>
<tr>
<td>Canopy growth coefficient</td>
<td>8</td>
<td>8.5</td>
</tr>
<tr>
<td>CCx</td>
<td>65</td>
<td>78</td>
</tr>
<tr>
<td>Canopy decline coefficient at senescence</td>
<td>11.7</td>
<td>11.7</td>
</tr>
<tr>
<td>Harvest index (adopted for all simulations)</td>
<td>42</td>
<td>34</td>
</tr>
</tbody>
</table>
observations were taken at 11 am during every visit at decadal intervals from the development stage through the maturity stage. A canopy cover of 65% was adopted for calibrating AquaCrop model for the Rift Valley zone of western Kenya (Table 3).

Statistical analysis of the canopy cover data gave an $R^2$ value of 62% (Figure 3). This value could be attributed to the cloudiness of the region that interfered with the field observations.

**Biomass calibration**

A similar procedure to that applied at the Kakamega-KARI site was employed for the Moi University site. The data collected were plotted against the simulated biomass attained by inputting the canopy cover obtained after calibration, soil data and the climate for Moi University station into AquaCrop which gave an $R^2$ value of 96% (Figure 4).

**Model validation**

Validation of the model was based on grain yield results by holding constant model parameters determined in the calibration stage. To validate the model, 3 years (2001, 2003, 2006) where the actual harvested yield compared well with their simulated yields were used together with the 2009 harvest from the experimental site yield at Moi University (Table 4). The procedure for determining a suitable harvest index for simulations was similar to that used for Kakamega-KARI station. The Nash–Sutcliffe model efficiency coefficient (EF) (Nash & Sutcliffe 1970, Wglarczyk 1998) was used to assess the predictive power of the model while the per cent RMSE was used to indicate the error in the model estimates. A plot for the observed yield versus simulated yield after calibration gave a goodness of fit value $R^2 = 89\%$ (Figure 5).

**Time series analysis results**

Rainfall time series analysis was done for the maize growing seasons in western Kenya. This analysis was used to classify climate into different conditions including below normal, normal and above normal. A classification method for the climatic conditions is demonstrated using March–April–May (MAM) rainfall for Kakamega-KARI and Moi University stations (Figures 6 and 7). This analysis is done for each station and the results are complemented by the quarterly forecasts released by KMD and applied to predict grain yield during the growing season. Expected rainfall amounts for each classification were derived for western Kenya region as shown (Table 5). These statistics were obtained for both MAM and June–July–August (JJA) using (20–60) year long-term data series for the region.

Time series analysis is complemented by periodic quarterly forecasts released by KMD. The time series analysis focused on MAM and JJA seasons that cover the sensitive stages of maize crop in the region (Mugalavai & Kipkorir
Results for the two representative stations (Kakamega-KARI and Moi University) are shown for the seasons that had both simulated and observed data (Table 6). For Kakamega-KARI station 1998, 2001, 2005 and 2009 growing seasons are used. The results reveal that 1998 and 2009 seasons were normal rainfall years for both MAM and JJA, the year 2001 had below normal rainfall seasons while 2005 had normal rainfall for MAM and below normal rainfall for JJA (Table 6). The results indicate that below normal rainfall during the sensitive (germination and flowering) growth stages leads to low yields (Mugalavai & Kipkorir 2013); however, the model still simulates yield based on the prevailing conditions and therefore offers an early opportunity to plan for food reserves well in advance.

For Moi University station, the 2001, 2003, 2006 and 2009 growing seasons were used to validate the use of AquaCrop model by linking the climatic conditions to the yield obtained. This station had normal rainfall conditions for the seasons 2001, 2003 and 2006 and below normal rainfall conditions for the 2009 growing season. These results further reveal that occurrence of below rainfall conditions in one or both MAM and JJA seasons leads to low yields (Figures 6 and 7).

### Long-term yield simulation

The calibrated AquaCrop model was used to simulate maize yield for the two representative stations in western Kenya based on long-term seasonal rainfall characteristics (Figures 7 and 8). Point-specific analysis is required to obtain good results and avoid overgeneralizations for the region. The AquaCrop model provides a robust method for estimating grain yield during the growing season. Using the long-term simulated yields generated in AquaCrop, the mean simulated seasonal yields were obtained for each station. The results indicate that Kakamega-KARI and Moi University stations have mean simulated yields of 6.9 and 4.4 tons/ha, respectively. These results reveal that the...
climate observed during the growing season determines the crop yield. This was demonstrated using the KMD forecasts for the seasons 2008 and 2009 for both stations. Kakamega-KARI station, which received above average yield for both years (Figure 8), registered normal rainfall amount whereas Moi University station (Figure 9) had below normal to near normal rainfall and crop yield was consequently below the long-term seasonal mean. Once calibrated, the model is able to identify years with yields below the long-term mean and therefore provide a basis for planning food reserves before the end of the growing season.

Yield simulation based on persistence theory

The climate condition used as an input parameter in AquaCrop for a particular season depends on Markov processes (Whiton 1977). According to Markov processes, climate is considered to change from one state or condition to another with the condition observed at time step \( n \) depending exclusively on the condition observed at the time step \( n - 1 \) immediately preceding it. This theory is adopted in selecting the climatic condition to use as an input parameter in the AquaCrop model for simulating grain yield during the ongoing growing season. For any projected condition, the climate applied was based on a similar preceding event.

Assuming that the actual observed yield for Kakamega-KARI station during the seasons 1998, 2001, 2005 and 2009 were unknown and the climatic conditions for the station were as shown in Figure 8, Whiton’s (1977) theory can be applied to simulate the yield for these seasons. By applying this method, the forecasted yield for these seasons (Figure 6) would be 7.1, 7.2, 5.7 and 7.2, respectively. However, the actual observed yield for these seasons (Table 4) is 7.5, 7, 5.9 and 7.4 tons/ha, respectively, giving a strong correlation coefficient \( r \) of (0.9). Kakamega-KARI station in western Kenya was used to validate this technique since it had a longer length of climate data compared to Moi University.

Seasons that are projected to obtain below average mean yields are identified for issuing early warning for food security reserve arrangements. Accurate results will depend on closely monitoring the ongoing seasonal rainfall characteristics in comparison with the long-term seasonal means that should be complemented by the quarterly seasonal climate updates released by KMD.

Crop yield and observed rainfall

The amount of rainfall received during the growing season determines the quality of the season in terms of the yield obtained. The seasonal rainfall time series for the two experimental stations used were done and the mean seasonal rainfall amount obtained. Seasons that had actual observed yield were used to establish the relationship between yield and the actual rainfall received. For Kakamega-KARI station (MAM), seasons 1998, 2005 and 2009 (shown in Figure 10) were used while the total rainfall received for both MAM and JJA (2001, 2003, 2006 and 2009) was used for Moi University station which has a longer growing season (Figure 10). The results reveal that there exist strong
relationships \((\text{Kakamega-KARI } R^2 = 0.7 \text{ and Moi University } R^2 = 0.9)\) between the rainfall received and the yield obtained during the growing seasons.

**CONCLUSIONS**

The AquaCrop model was calibrated for the western Kenya region. Due to the diversity of the region, calibration parameters were based on two distinct zones (western and Rift valley) with different cropping patterns. The western zone was calibrated based on a canopy cover of 78% and a harvest index of 34% whereas the Rift valley zone performed well with a canopy cover of 65% and a harvest index of 42%. The soil and climate parameters were site specific. The calibrated model is able to estimate both long-term seasonal grain yield and expected harvest for the ongoing season based on climatic conditions that are compared with the long-term seasonal characteristics and complemented by KMD forecasts. The results reveal that Kakamega-KARI and Moi University stations have long-term mean simulated yields of 6.9 and 4.4 tons/ha, respectively. It is important to incorporate the ongoing season climate in the simulation since there exist strong relationships \((\text{Kakamega-KARI } R^2 = 0.7 \text{ and Moi University } R^2 = 0.9)\) between the rainfall received and the yield obtained during the growing seasons (Figure 10). The model is able to identify years with yield below the long-term seasonal means and therefore provides a basis for responsive planning of food reserves before the end of the season. The ongoing season yield simulation is based on persistence theory of Markov processes.

**REFERENCES**


Hsiao, T. C., Heng, L., Steduto, P., Rojas-Lara, B., Raes, D. & Elias Fereres, E. 2009 *Aquacrop – The FAO crop model to simulate...


First received 28 March 2014; accepted in revised form 31 August 2014. Available online 7 October 2014.