OPERATIONAL MAIZE YIELD MODEL DEVELOPMENT AND VALIDATION BASED ON REMOTE SENSING AND AGRO-METEOROLOGICAL DATA IN KENYA*

O. Rojas

Joint Research Centre (JRC), MARS-FOOD-IPSC, EC, TP 266, 21020 Ispra (VA), Italy; email: oscar.rojas@jrc.it

Commission VIII, WG VIII/10

ABSTRACT:

Remote sensing (RS) data acquired by satellite have wide scope for agricultural applications owing to their synoptic and repetitive coverage. On the one hand, spectral indices deduced from visible and near-infrared RS data have been extensively used for crop characterization, biomass estimation and crop yield monitoring and forecasting. On the other hand, extensive research has been conducted using agrometerological models to estimate soil moisture to produce indicators of plant-water stress. This paper reports the development of an operational spectro-agrometeorological yield model for maize using a spectral index, the Normalized Difference Vegetation Index (NDVI) derived from SPOT-VEGETATION, meteorological data obtained from the European Centre for Medium-Range Weather Forecast (ECMWF) model and crop-water status indicators estimated by the Crop Specific Water Balance model (CSWB). Official figures produced by the Government of Kenya (GoK) on crop yield, area planted and production were used in the model. The statistical multiple regression linear model has been developed for six large maize-growing provinces in Kenya. The spectro-agrometeorological yield model was validated by comparing the predicted province-level yields with those estimated by GoK. The performance of the NDVI and land cover weighted NDVI (CNDVI) on the yield model was tested. Using CNDVI instead of NDVI in the model reduces 26% of the unknown variance. Of the output indicators of the CSWB model, the actual evapotranspiration (ETA) performs best. CNDVI and ETA in the model explain 83% of the maize crop yield variance with a root square mean error (RMSE) of 0.3298 t/ha. Very encouraging results were obtained when the Jack-knife re-sampling technique was applied proving the validity of the forecast capability of the model (r^2= 0.81 and RMSE= 0.359 t/ha). The optimal prediction capability of the independent variables is 20 days and 30 days for the short and long maize crop cycles respectively. The national maize production during the first crop season for the years 1998 to 2003 was estimated with a RMSE of 185 060 tons and coefficient of variation of 9 %.

1. INTRODUCTION

Crop-weather models had long been used for crop monitoring and yield forecasting before the advent of remote sensing products, like the Normalized Difference Vegetation Index (NDVI). More than 50 years have passed since the first paper on mathematical modeling of photosynthesis and productivity in plant communities was published in Japan (Monsi and Saeki, 1953) and these kinds of studies were later continued by research groups formed in Netherlands (de Wit et al., 1970; de Wit and Goudriaan, 1974). In the USA, McCree (1970), Curry (1971), and Loomis et al. (1979) published outstanding papers along the same lines. Interesting research was undertaken in Poland by using statistical empirical models (Górski et al., 1994).

In 1975, during the major world food crisis produced by climatic events, including the Sahelian droughts of 1972 and 1973, FAO established the Global Information and Early Warning System for Food and Agriculture (GIEWS). The absence of low-cost methods applicable to large regions oriented FAO to use the Crop Specific Water Balance (CSWB) model (Frère and Popov, 1979; Gommes, 1993) as a tool for monitoring and yield forecasting in African countries. In those countries, the insufficient information on weather and crops precluded applying more complex models.

In Europe, the crop-weather model, WOFOST, was adapted for monitoring and yield forecasting in European countries; the model represents the engine of the Crop Growth Monitoring System established in 1998 by the Monitoring Agriculture with Remote Sensing (MARS) project at the Joint Research Centre (JRC) (Meyer-roux and Vossen, 1994). WOFOST is a member of the family of models developed in Wageningen by the C.T. de Wit school (de Wit et al., 1970; de Wit and Goudriaan, 1974).

The introduction of remote sensing and the derived vegetation indices in the early 80's was considered a potential tool to improve simulations by objective observations in real-time. NDVI has been used as an indicator of the vigour of vegetative activity as represented by indirectly observable chlorophyll activity (Hastings and Emery, 1992). Low values of NDVI have been associated with the lack of vegetation, dormant states of existing vegetation or stress caused by drought, over-irrigation, or diseases (Hastings, 2005). Remote sensing products alone

---

have been used in different parts of the world to estimate crop yield (Lewis, et al, 1998; Hochheim and Barber, 1998, Wang, et al, 2005).

Potdar et al (1999) observed for some cereal crops grown in rain-fed conditions that rainfall distribution parameters in space and time need to be incorporated into crop yield models in addition to vegetation indices deduced from remote sensing data. Such hybrid models show higher correlation and predictive capability than the simple models (Manjunath and Potdar, 2002). The agro-meteorological models introduce information about solar radiation, temperature, air humidity and soil water availability while the spectral component introduces information about crop management, varieties and stresses not taken into consideration by the agro-meteorological models (Rudorff and Batista; 1990). The purpose of this research is to improve the spatial estimation of yield by combining crop-weather models and satellite observations.

Using Kenya as an Eastern Africa case study, this paper presents the methodological approach employed to build and validate a maize yield model using remote sensing data from SPOT VEGETATION and the outputs of the FAO-CSWB model. Kenya was chosen as a trial study area for developing the model because Kenya is relatively rich in data, and agriculture is practiced in coastal, low land, and high land areas which have diverse climates and are representative of most regions in Eastern Africa. Maize is a major food crop cultivated in Kenya. It represents 90% of national cereal production. Between 1998 and 2003, the average area cultivated with maize was 1 574 370 hectares with a total national production of above 2 475 947 tons and a national average yield of 1.57 tons by hectare. Maize is mainly cultivated in the South-Western part of the country, in the provinces of Rift Valley, Nyanza and Western. The three provinces together produce more than 80% of the national maize production. Nyanza, Western and Rift Valley provinces have a mono-modal rainfall distribution while Central, Eastern and Coast provinces exhibit on average a bi-modal distribution of rainfall with the possibility of having two crop seasons each year. The first crop season historically extends from February to August, producing more than 82% of the national maize production, while the second crop season, from September to January, represents 18%.

2. MATERIALS AND METHODS

2.1 Real-time input data

2.1.1 Meteorological data
The rainfall and potential evapotranspiration (PET) data used in this study are products of the European Centre for Medium-Range Weather Forecast (ECMWF model) at Reading in the UK. The data were interpolated from the original 1-degree grid to a final resolution of 0.5 degree (approximately 55 km). Dekadal rainfall and ETP were then spatially averaged for each area comprised in the maize crop mask using ArcMap GIS tools.

2.2.2 Remote-sensing data
The products of SPOT VEGETATION acquired by MARS are 10-day NDVI (Normalized Difference Vegetation Index) synthesis (S10) images, obtained through Maximum Value Compositing (MVC). The images are corrected for radiometry, geometry and atmospheric effects. The 10-day images are delivered to the JRC with a delay of around 2-3 days.

2.2. The CSWB model
The FAO CSWB is a very simple but physically sound soil water balance model which is used to assess the impact of weather conditions on crops (Frère and Popov, 1979; Gommes, 1993, Rojas et al., 2005). The water balance of the specific crop is calculated in time increments, usually 10-days. The equation of the water balance is:

$$W_t = W_{t-1} + R - ETA - (r + i) \qquad (1)$$

Where,

W_t: amount of water stored in the soil at the time t.
W_{t-1}: amount of water stored in the soil at the end of the previous period (t-1).
R: cumulated rainfall during the dekad or t-period of time.
ETA: actual evapotranspiration in the t-period time
r: represents the water losses due to runoff in the t-period time
i: represents the water losses due to deep percolation in the t-period time

Two main outputs of the CSWB model are demonstrated to be positively correlated with the crop yield: the Actual evapotranspiration (ETA) and the Water Satisfaction Index (WSI). ETA has the advantage to include the radiation, which is an important climatic variable susceptible to influence the crop yield in the region. The influence of factors other than water stress which can reduce crop yields such as water logging, mechanical damage produced by strong winds, or biological factors, such as locusts, birds, insects or plant diseases are not considered by the CSWB model. The WSI is an index of the CSWB model to assess the amount of water received by the crop during any time of the season. Normally, the WSI is used for defining qualitative yield classes (i.e. good, average, and poor) or in relative figures (percent of an optimal yield crop). When the WSI is equal to 100, it indicates no water stress and good crop yields, while a WSI of 50 corresponds to poor crop yield or crop failures. The estimation of the actual evapotranspiration (ETA) was done using Agromet-Shell[2] (Hoefsloot, 2005). The crop information needed to run the water balance model (water holding capacity and cycle length) was taken from the Crop production system zones database (CPSZ) (Van Velthuizen et al, 1995).

2.3 Planting date estimation model

To start the simulation, the CSWB model requires the current planting date of each crop season. The criterion followed to define the planting dekad was the 1st dekad with at least 20 mm of rainfall followed by two dekads with at least 20 mm of total rain. The same planting date was used to start accumulating NDVI values up to the end of the crop cycle (Table 1).

[2] AgrometShell is a software that integrates the main tools used in the Early Warning System such as SUIVI (agrometeorological database), FAOINDEX (Crop specific water balance), and the most common tools of data interpolation. Agromet-Shell was designed for storing agrometeorological information on meteorological stations base. In the present study the information from the crop mask units (spatially averaged of the polygon) was inserted in lieu of meteorological stations with the objective to run the water balance model. AgrometShell has been developed by the Agrometeorological Group at FAO and programmed by Peter Hoelsloot.

2.4 Crop statistics

The Kenyan government started collecting disaggregated agricultural statistics by crop season in 1997. However, since the SPOT VEGETATION sensor was launched on board the SPOT 4 satellite later, in 1998, crop data was analyzed between 1998 and 2003. The statistics are collected at district level and aggregated by province[*].

Table 1. Maize crop cycle length and phenological phases in dekads

Province	crop cycle dekads*	Initial dekads	Vegetative dekads	Flowering dekads	Ripening dekads
Central	16	3	3	7	3
Coast	11	2	2	5	2
Eastern	9	2	2	3	2
Nyanza	13	3	2	5	3
Rift Valley	16	3	3	7	3
Western	16	3	3	7	3

*10-days period

2.5 Maize crop mask

In this study, two levels of maize crop mask were defined. The first level, the 'general' maize crop mask was created using only statistical information; the second one is the result of intersecting the first level of crop mask with the Africover land cover information (Di Gregorio et al, 2000). The first level of crop mask was used for area-averaging of the meteorological and NDVI information. The second one was used for extracting the land cover weighted NDVI (CNDVI) (see point 2.7). To define the first level of crop mask the statistical information about the area planted with maize at district level was used. For each district, we calculated the percentage of the total provincial area planted with maize. The districts with no maize planted and those with less than 6% of the area planted with maize were masked-out from each province. As a result a general maize crop mask was used, constituting all the districts with more than 6% of area planted with maize for each province. The final resolution of this crop mask is at province level. To obtain a more precise maize crop mask we used two classes from the Africover land cover database: the isolated small fields and continuous small fields that were considered to better represent the traditional maize farms of Kenya. The first maize crop mask at province level was intersected by AFRICOVER classes. The result is a better delimitation of the areas cultivated with maize in each province. Unfortunately, the polygons resulting are too small to be used for extracting meteorological information at 0.5 degrees resolution.

2.6 NDVI

The Normalized Difference Vegetation Index (NDVI) has been the most frequently used vegetation index within agrometeorological analysis. It is defined as:

$$NDVI= (NIR–RED)/(NIR+RED) \qquad (2)$$

NIR and RED are, respectively, the reflectance (%) in the near-infrared and in the red channels. It is easy to understand the index when the characteristics of absorption and reflection of the radiation by green leaves is studied. The chlorophyll of the plant absorbs the majority of the radiation in the visible part of the spectrum, principally the red portion (0.6-07 μm), and is highly reflective in the near-infrared. Thanks to this property of green vegetation, NDVI is a direct indicator of the plant's photosynthetic activity. Therefore parameters such as water stress can be monitored successfully by analysing the NDVI values. The NDVI values were spatially averaged for each area comprised in the maize crop mask. Three variables were created when aggregating the NDVI values on a temporal scale: cumulated NDVI values starting from planting date up to the end of the length of the crop cycle (NDVIc), maximum NDVI during the crop cycle (NDVIx) and 3 dekad-averages around the maximum NDVI (NDVIa) to smooth the curve when an isolated peak represents the maximum.

2.7 CNDVI methodology

Despite the fact that it is not possible to eliminate all spectral responses from non-agricultural vegetation in the African parcels, any improvement of the crop mask will reduce the influence of natural vegetation and show a higher correlation of the remote sensing indices with crop yield. Therefore, it was decided to include the Land cover weighted NDVI method (CNDVI[*]) using Africover land cover (Di Gregorio et al, 2000). The CNDVI method has been developed to extract NDVI profiles from low resolution satellite imagery. It is currently in use for agricultural monitoring in Europe, with two main objectives: a) to aggregate NDVI information by administrative regions in order to give synthetic and manageable information; b) to focus on agricultural land only, owing to the integration of land cover information. The CNDVI method is fully documented by Genovese et al, 2001. The method was originally designed and tested for NOAA-AVHRR (with a 4.4 km resolution) and CO-ordination of Information on the Environment (CORINE) land cover data (Perdigao and Annoni, 1997), but can theoretically be applied to all combinations of low resolution images and higher resolution land cover data. Negre et al, 2001, have adopted the methodology to work in Africa using SPOT-VEGETATION instead of NOAA-AVHRR and AFRICOVER land cover classes (Di Gregorio et al, 2000) instead of CORINE Land cover. For the CNDVI extraction itself, the regrouped agricultural classes are re-scaled to the same resolution as the VGT images (1km), creating so called abundance images. In these images the value of each 1 km pixel expresses the percentage which is covered by an Africover class. The NDVI of each 10-day image pixel is weighted following the abundance image and the final NDVI profiles are class specific. Aggregation is done at a regional level to obtain a single CNDVI value per region, through a weighted average of NDVI values. The agricultural AFRICOVER classes are not crop specific but provide information about field size and field distribution. In this study two classes were selected: the isolated small fields and continuous small fields that were considered to better represent the traditional maize farms of Kenya.

---

[*] MARS-FOOD received the Kenya statistics of area planted, yield and production aggregated at national level of maize and sorghum for the period 1985-2003, and disaggregated by crop season ('Long rains' and "Short rains") at district level for the period 1997-2003 from Nancy Mutunga, FEWS-NET Country Representative of Kenya.

[*] CNDVI: 'C' for Land cover information that in the case of Europe represents CORINE land cover and in Africa mainly AFRICOVER land cover and 'NDVI' for Normalised Difference Vegetation Index.

AFRICOVER classes were used to refine the crop mask by province, by selecting specific agricultural areas, supposed to be maize areas within each district. As done with the NDVI, three variables were created aggregating the CNDVI values on a temporal scale: using cumulative CNDVI values starting from planting date up to the end of the crop cycle (CNDVIc), maximum CNDVI during the crop cycle (CNDVIx) and 3 dekad-averages around the maximum CNDVI (CNDVIa) to smooth the curve when an isolated peak represents the maximum.

2.8 Crop yield model development and validation

A multiple linear regression analysis was used in the development of the crop yield model testing the following independent variables: WSI, cumulated ETA during the whole maize cycle, ETA cumulated by phenological phase (initial, vegetative, flowering and ripening), cumulated soil water deficit and surplus, NDVIc, NDVIx, NDVIa, CNDVIc, CNDVIx, CNDVIa and total cumulated rainfall during the crop cycle. To increase the number of observations and hence the net degree of freedom, the model was developed considering all the observations from all regions together. The Jack-knife re-sampling technique (leaving one data value out each time) was applied to test the forecast capability of the model. To avoid a strong influence of climatic conditions given by a specific year each time it excluded a set of observations belonging to the same year. To assess the prediction capacity of the model, a correlation matrix with the independent variables accumulated during the phenological phase of maize was tested. To study the evolution of the r-square and RMSE, 4 multiple linear regression models were built at provincial level whereby each model represents a

phenological phase

of maize (initial, vegetative, flowering and ripening) using the most correlated variables. The Jack-knife technique was applied to each model to validate its forecasting capability.

The methodology flow chart describing briefly the steps involved in digital data analysis, the agrometeorological model outputs and the development of the spectro-agrometeorological yield model is given in Figure 1.

2.9 Estimation of national maize production during the first crop season

Although our main scope was the development of a crop yield forecasting model, due to the fact that the area planted with maize has a strong time-trend in Kenya, it is possible to obtain an estimate of the national maize production during the first crop season. The maize crop yield was estimated by province using the spectro-agrometeorological yield model. To obtain the national maize yield average, the provincial yields obtained by the spectro-agrometeorological model were weighted by the percentage of contribution of each province to the total national area planted with maize. The national area planted with maize was estimated using the time-trend equation of Figure 3 *(b)*. Between 1998 and 2003, the area planted with maize during the first crop season represented above 71% of the national area planted with maize. A correction factor from the statistics of 71% was applied to estimate the area planted during the first crop season. The weighted national yield was multiplied by the estimated area planted during the first crop season with maize to obtain the national production figures for the first crop season. Finally, comparison was made between the estimated production and the observed national production figures of Kenya.

3. Results
3.1 Maize crop masks

Figure 2 *(a)* shows the districts in each province that represent more than 6% of the area planted with maize. The districts with
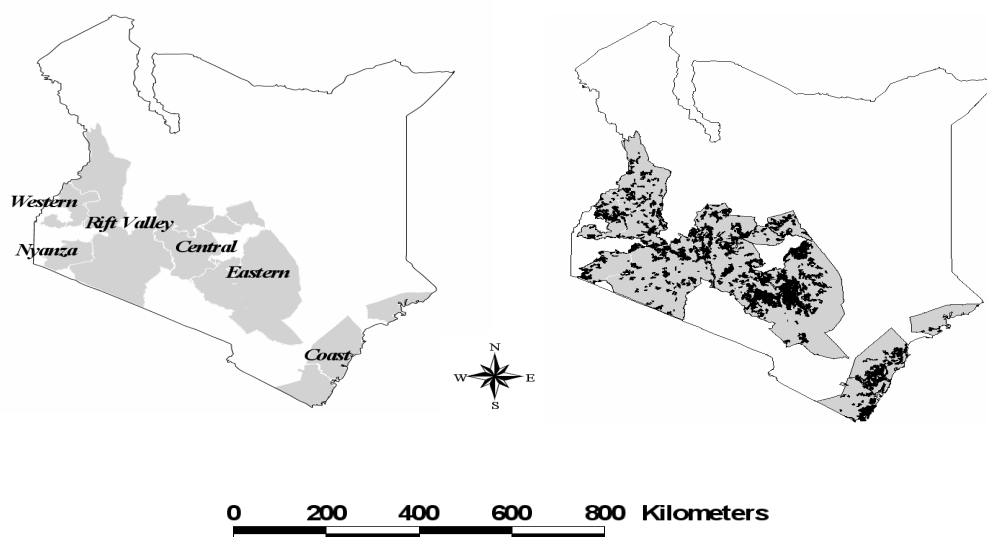

Figure 2. *(a)* Maize crop mask done based on the percentage of area planted with maize at district level (districts with less than 6% of area planted with maize have been masked-out). *(b)* In black the isolated small fields and continuous small fields from the Africover database; in gray the maize crop mask done based on the percentage of area planted with maize at district level. The Africover classes outside the general maize crop mask have been masked-out.
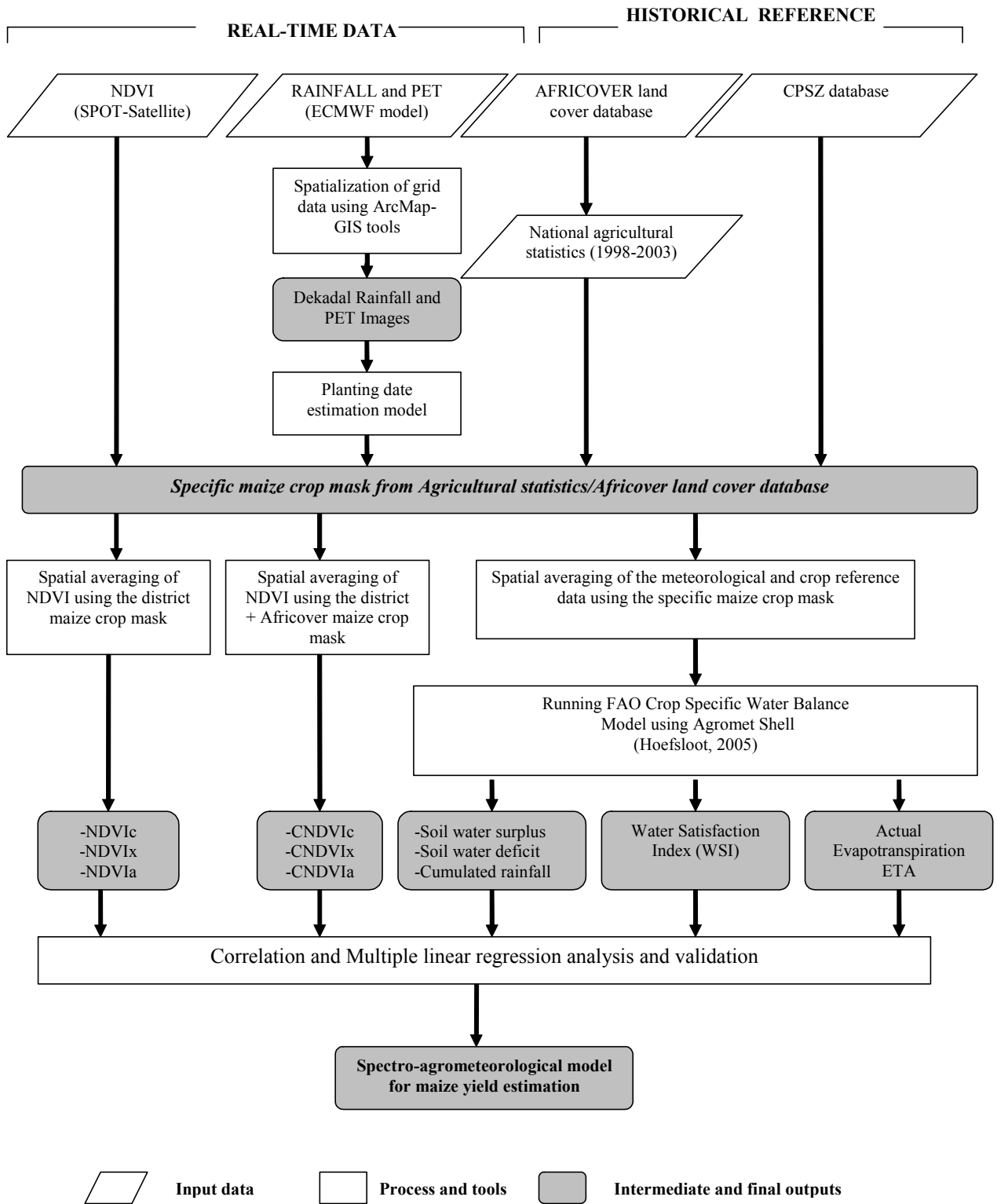
Figure 1 Methodology flow chart describing input data, process and tools for the development of the spectro-agrometerological yield model.

less than 6% have been masked-out. This crop mask was used in order to extract the NDVI and the meteorological values needed to run the CSWB model. Figure 2 *(b)* shows the two classes of Africover land cover considered in this study: isolated small fields and continuous small fields. The Africover classes outside the general crop mask were not considered during the extraction of the CNDVI values.

3.2 Trend analysis

The trend in rainfall, area planted, yield and production of maize during the first crop season was studied. The results of the analysis carried out at the province level are presented in Table 2. The maize yield exhibits a negative trend in Coast, Nyanza and Western provinces; the data was not de-trended due to the fact that this tendency can be explained by the trend in rainfall. Considering the objective of the study, we de-trend when the tendency is explained by variables other than climate, such as technological improvements. Nyanza and Rift Valley have a very positive trend in area planted during the 'Long rains' crop season. Due to the fact that 6-years is a short series to have a conclusive trend analysis, we used the longest series of national aggregated data of Kenya (1985-2003) and analyzed the production, area planted and yield of maize at national level.

The statistics shown in Figure 3 *(a)* illustrate that maize production has no trend. The average production is above 2.5 million tons with a minimum production of 1.7 million tons which occurred in 1993 followed by a maximum production of 3.0 million tons in 1994. Figure 3 *(b)* shows that the area planted has a strong positive trend while maize yield has a negative one. Kenya has increased the area planted to compensate for decreased productivity and the growing demand for maize. The results of our trend analyses undertaken during the first crop season suggest that the increase in area planted has been concentrated mainly in the Nyanza and Rift Valley provinces, and less extensively in the Coast province. Eastern is the only province that shows a negative trend in area planted during 1998-2003 period (Table 2). Figure 4 shows the trend in annual rainfall (1989-2005), first crop season (1989-2005) and maize yield aggregated at national level (1985-2003).

Figure 3 *(a)* Maize production for the period 1985-2003. *(b)* Area planted and yield for maize during the period 1985-2003. Data from Government of Kenya.

3.3 NDVI and CNDVI

Figure 5 shows the difference between the spatially-averaged NDVI and CNDVI for the different provinces and years. The differences are smaller in Coast and Eastern province, suggesting little impact on the model. Meanwhile, the rest of the provinces have large differences. Negative differences could be explained by the addition of dry areas with low NDVI values in the general maize crop mask; meanwhile positive differences indicate that the general crop mask includes very dense natural vegetation with high NVDI values within the agricultural areas. We conclude that there is a difference between NDVI and CNDVI that spans from 0.01 to 1.08 when the variables are accumulated for the whole crop cycle.

Table 2. Trend analysis in rainfall, area planted and yield (1998-2003) and rainfall (1989-2005) by provinces during the first crop season in Kenya

Province	1998-2003						1989-2005	
	Rainfall	r^2	area plante	r^2	yield	r^2	Rainfall	r^2
Central	no trend	-	no trend	-	no trend	-	no trend	-
Coast	negative	0.56	positive	0.30	negative	0.87	no trend	-
Eastern	no trend	-	negative	0.23	no trend	-	no trend	-
Nyanza	negative	0.61	positive	0.84	negative	0.52	negative	0.49
Rift Valley	no trend	-	positive	0.79	no trend	-	negative	0.20
Western	negative	0.58	no trend	-	negative	0.52	negative	0.37

The trends of all three figures are negative. The trend in annual rainfall shows a small coefficient of determination ($r^2 = 0.13$) when compared with the coefficient of the first crop season ($r^2=0.34$). The accumulated rainfall from September to January, second crop season (1989-2005) has no trend, signifying that the decrease in maize yields is due to reduced water availability during the main cropping season in Kenya.
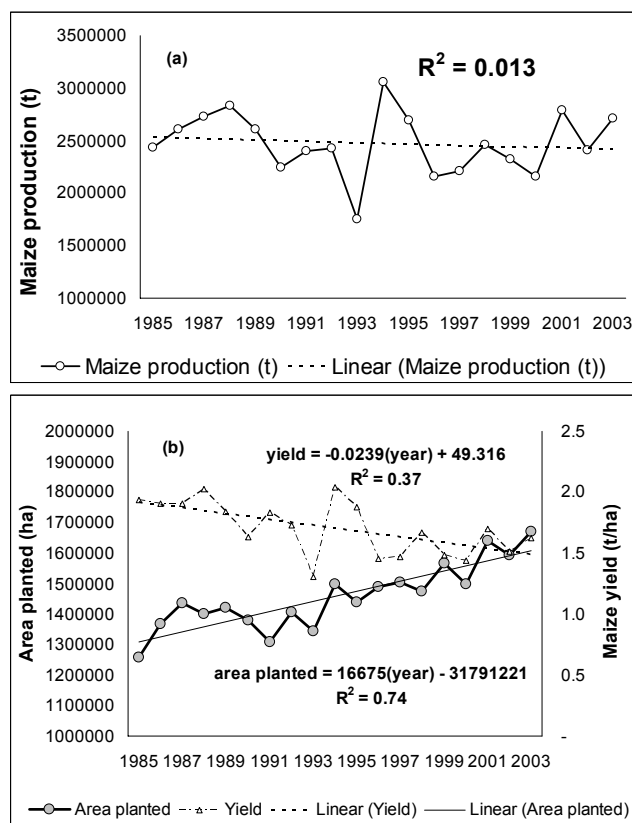
To assess the impact of such differences on the statistics of the model we calculated the reduction in per cent of the unknown variance using $1-r^2$ (r from Table 3). The unknown-variance has been reduced by 26% when CNDVI is used in the model instead of NDVI. We conclude that the CNDVI gives a better spectral signal of maize crop areas than NDVI spatially averaged by the general crop mask.
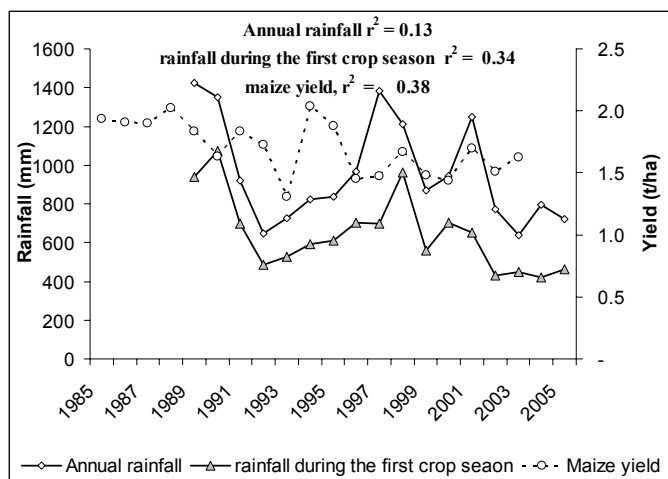
Figure 4 Negative trend in annual rainfall (1989-2005) and in the cumulated rainfall (February-August), first crop season, compared with the negative trend of the national maize yield.
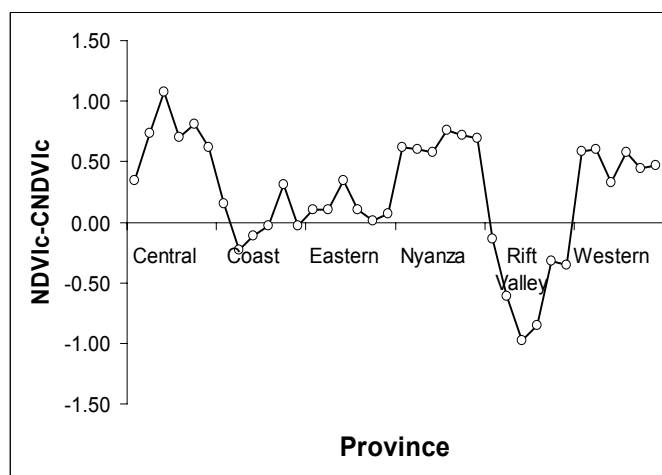

Figure 5 Difference between the spatially-averaged cumulated NDVI from planting dekad up to end of the crop cycle (NDVIc) and cumulated CNDVI for the same period (CNDVIc) by province.

Table 3 Correlation Matrix of maize yield and the independent variables

	WSI	WEXt	WDEFt	ETAi	ETAv	ETAf	ETAr	ETAt	Rain	CNDVIc	CNDVIx	CNDVIa	NDVIc	NDVIx	NDVIa
WSI	1														
WEXt	0.130	1													
WDEFt	0.921	0.283	1												
ETAi	0.298	-0.278	0.164	1											
ETAv	-0.402	-0.142	-0.382	0.107	1										
ETAf	0.852	0.265	0.795	0.309	-0.285	1									
ETAr	0.818	0.339	0.812	0.328	-0.278	0.764	1								
ETAt	0.860	0.275	0.811	0.386	-0.175	0.976	0.858	1							
Rain	0.381	0.947	0.494	-0.135	-0.215	0.502	0.551	0.517	1						
CNDVIc	0.237	0.290	0.122	0.585	0.144	0.517	0.408	0.560	0.405	1					
CNDVIx	0.395	0.320	0.361	0.613	0.148	0.433	0.505	0.523	0.423	0.700	1				
CNDVIa	0.383	0.268	0.338	0.610	0.081	0.381	0.463	0.462	0.358	0.675	0.954	1			
NDVIc	0.176	0.312	0.060	0.582	0.125	0.437	0.340	0.477	0.403	0.979	0.700	0.670	1		
NDVIx	0.186	0.253	0.161	0.560	0.092	0.170	0.278	0.244	0.288	0.582	0.901	0.868	0.656	1	
NDVIa	0.216	0.205	0.190	0.540	-0.011	0.141	0.248	0.200	0.238	0.512	0.825	0.829	0.602	0.960	1
Yield	0.511	0.350	0.463	0.574	0.057	0.663	0.649	**0.731**	0.525	**0.869**	0.782	0.771	0.818	0.594	0.525

Table 4. Multiple regression coefficients, Standard Error, t-Stat, P-value and 95% Confidence Inte

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upp
Intercept	-1.4429	0.2324	-6.2094	0.00000052164	-1.9157	
ETAt	0.0030	0.0007	4.2611	0.00015945560	0.0016	
NDVIc	0.2498	0.0312	8.0022	0.00000000312	0.1863	

3.4 Correlation matrix

Table 3 presents the correlation matrix of maize yield and the independent variables. There are three groups of independent variables: variables derived from remote sensing, climatic variables derived from the ECMWF model and variables derived from the CSWB model. The CNDVIc shows the highest value of correlation coefficients amongst remote sensing variables (r = 0.87). The average of 3-dekads around the maximum NDVI doesn't give a better correlation than the maximum itself, which means that smoothing the peaks of the NDVI curve does not improve the correlation with yield. Rainfall accumulated from planting date up to the end of crop cycle gives a correlation coefficient of 0.52 which is higher than some of the indicators produced by CSWB model such as the Water Satisfaction Index (WSI) and accumulated water excess (WEXt) and deficits (WDEFt). Amongst the CSWB model variables, the ETA shows a high correlation coefficient (r = 0.73). It is interesting to highlight that the results of ETA by phenological phases show good correlation during the initial, flowering and ripening phases and low correlation during the vegetative one. These results are in agreement with international research about impact of water stress on crop yield during different phenological phases (Doorenbos and Pruitt, 1977). Finally, we selected the two most correlated variables, ETA (total) and CNDVIc, to create the multiple linear regression model.

3.5 Spectro-agrometerological model

Figure 6, shows the comparison between the estimated maize yields by the model and the observed ones. Table 4 shows the multiple regression coefficients of the intercept and independent variables, the t-Stat and the interval of confidence of the coefficients at 95% of probability of occurrence. The adjusted r-squared is 0.83. The root mean square error (RMSE) of the model is 0.3298 t/ha and the coefficient of variation is 21%[*]. [Insert table 4 about here] The following equation of the spectro-agrometeorological model was found:

---

[*] Root mean square error (RMSE) examines the size of our forecast error. This measure assumes that larger forecast errors are of greater importance than smaller ones; hence they are given a more than proportionate penalty. The root mean square error (RMSE) is defined as:

$$\left(\frac{1}{T} \sum_{t=1}^{T} (F_t - A_t)^2 \right)^{1/2}$$

where: T = number of observations, F_t = forecast of component A_t = actual outturn

The coefficient of variation (CV) is a measure of relative dispersion and is given by (CV=standard deviation/mean). It is generally expressed as a percentage. In this study it was calculated using the standard deviation of the residual divided by the mean of the observed variable.

$$\text{Yield} = -1.4429 + 0.2498 \sum_{i=PD}^{t=EOCC} CNDVI(t) + 0.0030 \sum_{i=PD}^{t=EOCC} ETA(t) \quad (3)$$

Adjusted r-square = 0.83, n = 36, where t is dekad number;
EOCC = End of maize crop cycle;
PD = Planting dekad;
Yield = Maize crop yield expressed in tons by hectare;
CNDVI= Weighted NDVI using Africover land cover by dekad;
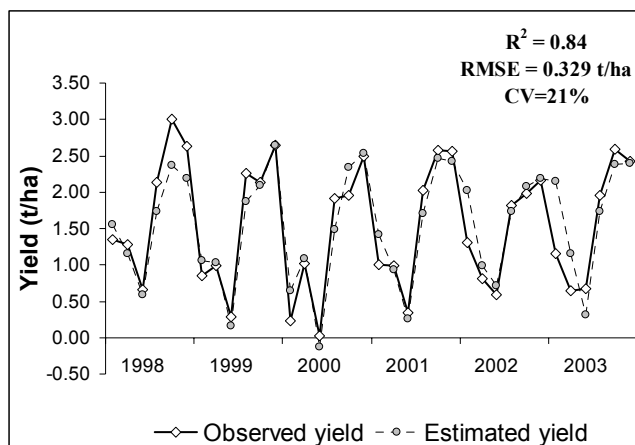ETA= Actual Evapotranspiration in millimetres by dekad

Figure 6 Comparison between the maize yield estimated by the spectro-agrometerological model and the observed yields for the different provinces.

3.6 Jack-knife re-sampling technique

To validate the forecast capability of the model the Jack-knife re-sampling technique was used. The impact of the difference on climatic conditions of each province was reduced leaving out each time a set of observations belonging to the same year. Figure 7 shows the comparison between the maize yields'
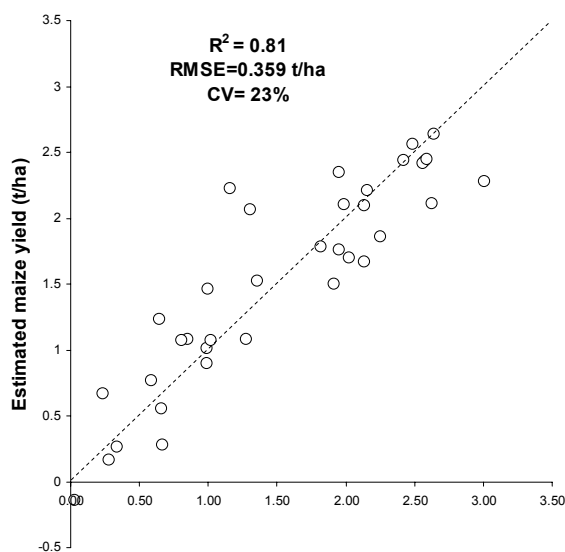

Figure 7 Comparison between maize yield estimated by the model using the Jack-knife re-sampling technique and the observed yield. The Root mean square error (RMSE) = 0.359 t/ha, coefficient of determination (0.81) and coefficient of variation of 23%.

estimates done by the model using the Jack-knife re-sampling technique and the observed yields. The r-square is 0.81, the root mean square error (RMSE) of the model is 0.359 t/ha and the coefficient of variation is 23%. Our results are encouraging when compared with those reported by Lewis *et al* (1998). They used a simple regression model with NDVI from NOAA-AVHRR for estimating maize production in Kenya and they obtained a Jack-knife r-square of 0.56.

3.7 Prediction capability of the independent variables

To study the prediction capability of the independent variables the correlation coefficient of CNDVI and ETA with maize yield was calculated. Figure 8 *(a)* shows the evolution of the correlation coefficient accumulated during the whole cycle and accumulated by phenological phases. CNDVI has a very strong correlation during the whole cycle. ETA shows low correlation only during the vegetative phase. The high correlation found in both variables requires further study. During the initial phase CNDVI has higher correlation than ETA, which can be explained by the fact that CNDVI integrates information about the pre-planting condition ("long memory") and is therefore better than simulations by the CSWB model. Also the spectral signal contains information about the characteristics of different soils that is difficult to introduce into the CSWB model. During the vegetative phase, the correlation of both variables by phenological phase decreases, this confirms the well-known low sensibility of yield when some stress happens during this phase (Doorenbos and Pruitt, 1977). The flowering phase shows a high correlation followed by the ripening phase. It was decided to build a multiple regression model using the CNDVI and ETA accumulated from the initial to ripening phases (that means by phenological phases). The Jack-knife re-sampling technique was used to avoid any strong influence of climatic conditions of a specific year. Figure 8 *(b)* shows the evolution of the adjusted r-square and RMSE after the Jack-knife technique had been applied. Even if the correlation during the initial phase is high in both variables, the adjusted r-square for this phase is 0.59. The variables at the early stage of the crop explain 59% of the variability of maize yields. Since the adjusted r-square is very high at the beginning of the crop season it should be tested once a longer time series is available to see if r-square remains high. Using the model at this early stage has a lot of uncertainty. Uncertainty decreases during the flowering period in which the adjusted r-square increases to 0.74 with a RMSE of 0.42 t/ha. We suggest that using the variables CNDVI and ETA accumulated from planting up to the end of flowering as a preliminary forecast and to refine it when the crop cycle reaches the end. The CNDVI and ETA accumulated for the whole crop cycle explains 81% of the maize yield variance with a RMSE of 0.36 t/ha when the Jack-knife technique is applied.

3.8 Estimation of national production during the first crop season

Although our main scope was the development of a crop yield forecasting model, due to the fact that the area planted with maize has a strong time-trend in Kenya, it is possible to obtain an estimate of the national maize production during the first crop season. We estimated the total area planted using the equation of time-trend (Figure 3 *(b)*). Using the spectro-agrometeorological model, the maize-yield was estimated at province level. Figure 9 shows the comparison of the observed production with the estimated production for the years

1998 to 2003. The RMSE is 185 096 tons with a coefficient of variation of 9%.
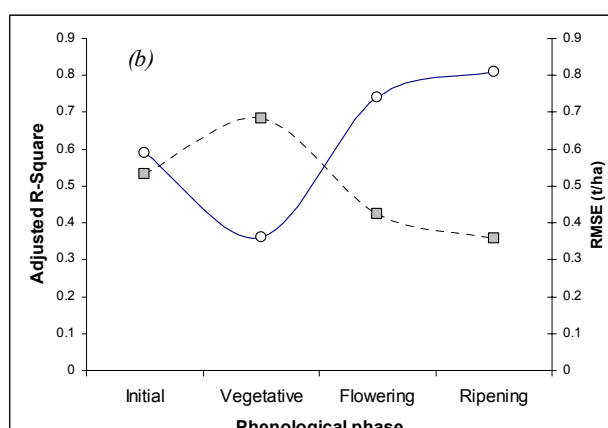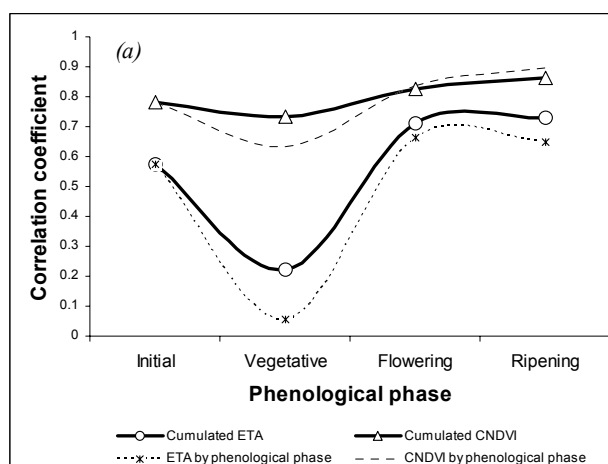

Figure 8 *(a)* Variation of the correlation coefficient (r) with the independent variables cumulated during the whole cycle and cumulated by phenological phases. *(b)* Variation of adjusted r-square and root mean square error in t/ha of the spectral-agrometeorological model using the cumulated CNDVI and ETA when applying the Jack-knife re-sampling technique.
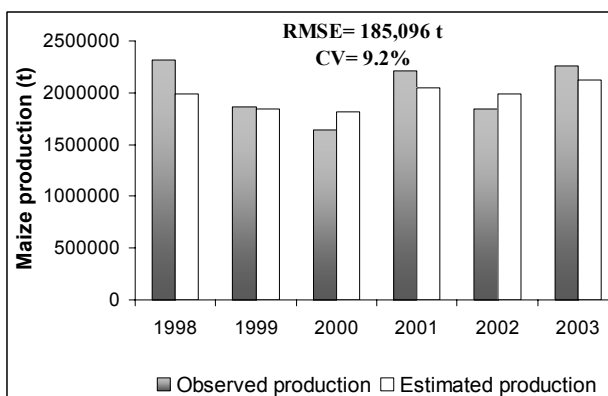

Figure 1 Comparison of the observed national maize production and the estimated maize production during the first crop season.

4. Conclusions and recommendations

It has been shown that it is possible to conduct operational maize yield forecasts using CNDVI derived from SPOT VEGETATION and ETA from the FAO CSWB model.

CNDVI showed to improve the spectral signal of the maize crop areas when compared with the simple spatially averaged NDVI using the general crop mask. CNDVI proved to be a simple and valid method for NDVI extraction with low resolution satellite images and highly fragmented high resolution land cover classes. However, significant improvements in extracting pure agricultural time profiles were primarily due to spatial refinements of the crop masks. The model showed a suitable prediction capability of 20 and 30 days before harvest for the short and long maize crop cycles, respectively. Thanks to this prediction capacity it is possible to obtain an early forecast using the CNDVI and ETA accumulated from planting dekad to the end of the flowering phenological phase. A more accurate estimate will be possible when the maize crop cycle reaches the end using the CNDVI and ETA accumulated for the whole length of the maize crop cycle. Even the second forecast using the variables accumulated up to the end of the crop cycle makes it possible to have reliable predictions 3 to 4 months earlier than the official estimates provided by national authorities and based on traditional field sampling surveys. As the time-series of the yield data was limited, some reservations for the model must be made, until a longer series of yield data will become available. The simplicity of the proposed regression yield model should allow an operational implementation in developing countries. Based on these encouraging results, regression models could be developed by MARS-FOOD for other geographical areas in Eastern Africa.

References

CURRY, R.B., 1971. Dynamic simulation of plant growth. I. Development of model. Trans. ASAE 14 (5), 946-949.

de WIT, C.T., BROUWER, R., PENNING de VRIES, F.W.T., 1970. The simulation of photosynthetic systems. In: Šetlík, I.(Ed.), Prediction and measurement of photosynthetic productivity. Pudoc, Wageningen, pp. 47-70.

de WIT, C.T., GOUDRIAAN, J., 1974. Simulation of Ecological Processes. Centre for Agricultural Publishing and Documentation, Wageningen.

DI GREGORIO A., JANSEN L., 2000, Land Cover Classification System (LCCS). Classification Concepts and User Manual. Publication of the FAO Environment and Natural Resources Service-AFRICOVER Project, 179 pp, ISBN 92-5-104216-0, FAO, Rome.

DOORENBOS J., PRUITT W., 1977, Guidelines for predicting crop water requirements, FAO Irrigation and Drainage Paper No. 24, FAO, Rome.

FRERE, M. and POPOV, G., 1979, Agrometeorological crop monitoring and forecasting. FAO Plant Production and Protection, Paper No.17, Rome.

GENOVESE, G., VIGNOLLES, C., NEGRE, T. and PASSERA, G., 2001, A methodology for a combined use of Normalized Difference Vegetation Index and CORINE land cover data for crop yield monitoring and forecasting. A case study on Spain. Agronomie, 21, 1, pp. 91-111.

GOMMES, R., 1993, FAOINDEX, Version 2.1. Agrometeorology Group. FAO Rome.

GORSKI, T., DEMIDOWICZ, G., DEPUTAT, T. GORSKA, K., MARCINKOWSKA, I., SPOZ-PAC, W., KRAKOWIAK, A., 1994. Agrometeorological quantification of agricultural year in Poland. Zesz. Probl. Post. Nauk Roln. 405, 81-87.

HASTINGS, D.A. and EMERY, W.J. 1992, The Advanced Very High Resolution Radiometer (AVHRR): A brief reference guide. Photogram. Eng. Remote Sensing, 58, 1183-1888.

HASTINGS, D.A., 2005, Africa's climate observed: perspectives on monitoring and management of floods, drought and desertification. In: Climate Change and Africa, Ed. by Paul Sum Low. Cambridge University Press. UK.

HOCHHEIM, K.P. and BARBER, D.G. 1998. Spring wheat yield estimation for western Canada using NOAA NDVI data. Canadian Journal of Remote Sensing. Vol. 24. No. 1, March 1998. pp. 17-27.

HOEFSLOOT, P., 2005, Agrometshell version 1.5. FAO, Rome. 23 p.

LEWIS, J.E., ROWLAND, J. and NADEAU, A., 1998, Estimating maize production in Kenya using NDVI: some statistical considerations. International Journal of Remote Sensing, 13, 2609-2617.

LOOMIS, R.S., 1970. Dynamics of development of photosynthetic systems. I.(Ed.), Prediction and measurement of photosynthetic productivity. Pudoc, Wageningen, pp. 137-141.

McCREE, K.J., 1970. An equation for the rate of respiration of white clover plants grown under controlled conditions. In: Šetlík, I.(Ed.), Prediction and measurement of photosynthetic productivity. Pudoc, Wageningen, pp. 221-230.

MANJUNATH, K.R. AND POTDAR, M.B., 2002, Large area operational wheat yield model development and validation based on spectral and meteorological data. Int. J. Remote Sensing, 2002, vol. 23 No. 15, 3023-3038.

MEYER-ROUX, J., VOSSEN P., 1994. The first phase of the MARS project, 1988-1993. Overview, methods and results. In proceedings of the Conference on: 'The MARS project, overview and perspectives', Belgirate, November 1993. EUR Publication No. 15599 EN, of the Office for the Official Publications of the E.C., Luxembourg, Space Applications Institute, JRC Ispra, pp 33-81.

MONSI, M. and SAEKI, T., 1953. Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduction. Jpn. J.Bot. 14, 22-52.

NEGRE, T., REMBOLD, F. and di GREGORIO, A., 2001, AFRICNDVI: An advanced tool for real-time crop monitoring in Africa. In proceedings of: "International Symposium Arid Regions Monitored by Satellites Marrakech 12-15 Nov. 2001.

PERDIGAO, V., ANNONI, A., 1997, Technical and methodological guide for updating CORINE land cover data base. Publication of the European Commission Joint Research Centre, Environmental European Agency, 124 pp. EUR 172888 EN, Brussels.

POTDAR, M.B., MANJUNATH, K.R., and PUROHIT, N.L., 1999, Multi-season atmospheric normalization of NOAA AVHRR derived NDVI for crop yield modeling. Geocarto International, 14, 51-56.

ROJAS, O., REMBOLD, F., ROYER, A. and NEGRE, T. 2005, Real-time agrometeorological crop yield monitoring in Eastern Africa. Agronomie 25 (2005) 1-15

RUDORFF, B.F.T.; BATISTA, G.T., 1990, Yield estimation of sugarcane based on agrometeorological-spectral models. Remote Sensing of Environment. 33, 183-192.

VAN VELTHUIZEN H., VERELST L., SANTACROCE P., 1995, Crop production system zones of the IGADD sub-region, Agrometeorology Working Paper Series No. 10, FAO, Rome.

WANG, J., RICH, P.M., PRICE, K.P., KETTLE, W.D., 2005. Relations between NDVI, grassland production, and crop yield in the Central Great Plains. Geocarto International, Vol. 20, No. 3, September 2005. pp. 1-7.