CHAPTER 1

INTRODUCTION

1.1 Background

The economy of Malawi is agricultural-based and this sector is almost entirely reliant on favourable climatic conditions for good agricultural production and therefore economic growth. The agricultural sector accounts for about 40% of the Gross Domestic Product (GDP) and also contributes over 90% of export earnings. It is also a reliable source of raw materials for the industrial sector (MEPD, 2002). For an agricultural-based economy like that of Malawi, reliable, accurate and timely information on the types of crops, their requirements and potential crop yield are vital components for the planners and policy makers. Timely dissemination of the information also assists in the distribution system and import/export policies of these commodities from time to time and in efficient management of natural resources.

Climatic hazards such as droughts have resulted in food shortages and consequential suffering of the majority of the population in the Southern African region. For example, in the 1991/92 season, Southern Africa experienced a severe drought. Eldridge (2002) suggested that with a credible, accurate forecast in early 1991 Southern Africa might not have experienced the costly emergency conditions that threatened millions of people with severe food shortages and famine. Governments had to divert funds meant for development into food assistance. The impacts of food shortages are worse when national governments do not take precautionary measures such as maintaining the strategic grain reserves, importing food and appealing for donor aid. However, these decisions should be based on information that has been made available by scientists with enough lead-time. This is where the process of yield estimation/ forecasting comes into play.

1.2 Yield Estimation

There are several crop-yield forecasting methods that have been developed and applied in different parts of the globe. Some of the methods are based on crop surveys just after planting, during crop development and during harvest (Rojas, 2004b). This whole process of surveys requires considerable human and financial resources that are not easily available, rendering the methods ineffective. In recent years, the Food and Agricultural Organization (FAO) devised a yield estimation method based on the crop water balance model-derived water requirement satisfaction indices (WRSIs).

The advent of remote sensing technology during the 1970s and its great potential in the field of agriculture has opened new opportunities for improving the agricultural system in many parts of the world. Space borne remotely sensed satellite data has been used in the field of Agriculture for the estimation of planted area. Remote sensing has also been used in crop yield estimation through development of regression models using historical yield data, meteorological data and remotely sensed satellite data (Singh, *et al.,* 2005).

Remote sensing techniques have been, and will continue to be, a very important factor in the improvement of the other systems of acquiring and generating agricultural and other environmental data. According to Balaselvakumar and Saravanan (2005) and Rasmussen (1992) the advantages of these remote sensing techniques in agricultural surveys include

- Vantage point - agricultural landscape depends on the sun as the energy source; therefore it is exposed to the aerial view, and therefore suited for remote sensing techniques.
- Coverage - it is now possible to record extensive areas on a single image due to the use of high altitude sensor platforms. Satellite data provides the actual synoptic view of a large area at a time.
- Permanent record - images serve as a permanent record of a landscape at a point in time against which agricultural changes can be monitored and evaluated.
- Real-time capability - the rapidity with which imagery can be obtained and interpreted may help to eliminate the lack of timeliness that plagues so many agricultural surveys.
- Easy data acquisition over inaccessible areas.
- Data acquisition at different scales and resolutions. The process of data acquisition and analysis is very fast through GIS. Due to the different resolutions of different satellites different agricultural surveys are possible depending on the accuracy required for each of them.
- Images are analyzed in the laboratory, thus reducing the amount of fieldwork.
- Remote sensing techniques have a unique capability of recording data in visible as well as invisible (ultra violet, reflected infrared, thermal infrared, microwave, etc) parts of the

electromagnetic spectrum thus making possible the observation of certain phenomena that cannot be seen by human eye.

1.3 Aim and Objectives

The overall aim of this study was to investigate the potential of forecasting maize yield in Malawi using climatological and satellite data-sets. This study focused on maize because it is the main food crop in Malawi and yield data was available.

The specific objectives of the study were:

- to assess the relationship between Normalized Difference Vegetation Index (NDVI) and rainfall at dekadal and seasonal time scales;
- to investigate the potential use of satellite derived vegetation indices for maize yield estimation for two different maize varieties at Rural Development Project (RDP) level;
- to evaluate the AgroMetShell (AMS) water balance model and
- to apply the model output parameters for maize yield estimation for two different maize varieties at RDP level.

1.4 Structure of the Thesis

This study concentrated on maize yield prediction in Malawi using a water balance model, the AgroMetShell (AMS) and NDVI. The motivation to undertake such a study was ignited by the need to provide near-real time, accurate and relatively cheap crop monitoring and yield estimation tools for early warning purposes. Chapter 1 of this study introduces the study and gives its objectives. Also different yield estimation / forecasting methods are briefly discussed, and the advantages of using remotely sensed techniques are outlined. Chapter 2 gives a background to the study in the form of literature review. Various studies relating to the current study were summarised. This also includes the yield estimation methods that are currently used in Malawi. In Chapter 3**,** the types of data and the methods used are outlined and the AMS model is also briefly described. In Chapter 4**,** the results are presented and discussed. These include the NDVI-rainfall relationships, NDVI-yield relationships and the AMS parameters-yield relationships. Conclusions and recommendations are set out in Chapter 5**.**

CHAPTER 2

LITERATURE REVIEW

2.1 The Normalized Difference Vegetation Index (NDVI)

The normalized difference vegetation index (NDVI) products are produced using measurements from the Advanced Very High Resolution Radiometer (AVHRR) on board the National Oceanic and Atmospheric Administration (NOAA) polar orbiting meteorological satellites. The reflectance measured from Channel 1 and Channel 2 is used to calculate the index which is a dimensionless parameter. These two wavelength regions are used because they provide a strong signal in NDVI, and secondly, they have a spectral contrast for most background materials and surfaces such as water and vegetation. NDVI is calculated using the relationship:

$$NDVI = \frac{CH_2 - CH_1}{CH_2 + CH_1} \tag{2.1}$$

where CH_1 and CH_2 represent the reflectance from channels 1 and 2 respectively (Teng, 1997).

Channel 1 is a visible wavelength (0.58 μm to 0.68 μm) and is in a part of the spectrum where chlorophyll causes considerable absorption of the incoming radiation while channel 2 is a near infrared wavelength (0.73 μm to 1.1 μm) and is in a spectral region where spongy mesophyll leaf structure leads to considerable reflectance. The differential reflectance in these bands provides a means of monitoring density and vigour of green vegetation growth using the spectral reflectivity of solar radiation. This contrast between responses of the two bands can be shown by ratio transforms. NDVI is one of the several proposed ratio transforms for studying different land surfaces. Other vegetation indices (ratios) derived from the same spectral channels include the Ratio Vegetation Index (RVI), the Difference Vegetation index (DVI) and the Vegetation Productivity Index (VPI). The use of a ratio between bands is important in reducing variations in radiance as a function of sun elevation for different parts of an image (Unganai and Kogan, 1998a, Teng, 1997). NDVI has been found to be highly correlated with vegetation parameters such as green leaf biomass and green leaf area (Teng, 1997). The ratio transforms have also been used for real time rangeland monitoring in different studies (Sannier *et al.,* 1998).

NDVI values potentially range from –1.0 to +1.0. Higher values (0.1 to 1.0) are associated with greater density and greenness of the plant canopy whereas clouds and snow on vegetation will cause very low or even negative NDVI values. Green leaves have larger reflectance in the near infrared (NIR) than in the visible (VIS) range. As the leaves come under water stress, become diseased or dieback, they become more yellow and reflect significantly less in the NIR range, thus reducing channel 2 reflectance and consequently resulting in low values of NDVI (Hutchinson, 1991, Teng, 1997).

The use of AVHRR-NDVI sensor system has advantages because of its capability in providing data with high frequency (Steven and Clark, 1990; Rasmussen, 1992). However, the low spatial (coarse) resolution (7.6 km by 7.6 km) appears too broad so that separation of crops from other vegetation types is not possible. As a result of this limitation, most studies just attempt to relate yield to NDVI for large areas and include a variety of vegetation types. This is done with the assumption that there is high correlation between the productivity of the dominant herbaceous vegetation and crops. However, this assumption is somewhat limited because phenologically there are significant differences between crops and herbaceous vegetation. For example, Rasmussen (1992) in a study of the phenology of Sahelian vegetation types showed that millet has a long growing season while herbaceous vegetation has shorter growing period, which ends after flowering. Therefore treating these two cases in a similar manner can significantly limit the accuracy of yield and biomass prediction. This limitation can be minimized by treating separately information from NDVI data related to herbaceous vegetation relative to the crops. The situation becomes more complicated if one crop is to be studied in areas where various types of crops are grown. This is still a challenge if the AVHRR-NDVI sensor system will continue to be used where the resolution does not allow separation between different types of vegetation. However with the advent of much higher resolution sensors systems like the Moderate Resolution Imaging Spectroradiometer (MODIS) and SPOT vegetation (with spatial resolution within hundreds of metres), this problem might be overcome. Some of the sources of errors that tend to increase CH_1 with respect to CH_2 thereby reducing the computed values of NDVI include scattering by dust and aerosols, Rayleigh scattering, sub-pixel sized clouds, large solar zenith angles and large scan angles (Maselli *et al.,* 1993).

2.1.1 Correction for atmospheric effects on NDVI images

Clouds obscure a significant portion of the land surface during most times of the year, especially as cloud cover increase during the rainy season when crops are grown. This complicates the observations of surface processes and phenomena from space. Clouds as well as snow interfere with the studies of vegetation and soils. The areas fully covered by thick clouds or snow can be easily distinguished but it is much harder to identify pixels covered by thin clouds, smoke or haze, or pixels that are partly covered and partly clear (Maselli *et al.,* 1993). There are also a lot of aerosols, dust even carbon dioxide and other gases that reflect, absorb and retransmit solar radiation. Another source of error in satellite data are what are called bi-directional effects. These include the sun-target-sensor geometry effects, whereby a target viewed at different times of the day or from different positions may appear to have different reflectance due to most surfaces being anisotropic reflectors and the atmosphere scattering also being anisotropic. These cause the sensors to pick wrong signals instead of the reflectance from the intended targets. Therefore, there is always need to identify and correct the contaminated pixels of satellite images.

Multi-temporal maximum value compositing (MVC) procedure has been devised and adopted to reduce atmospheric effects on the NDVI images. It involves retaining maximum NDVI value of several temporally adjacent and coregistred scenes as the least affected by atmospheric perturbations. In other words, the best NDVI value for a particular pixel is achieved by choosing the highest pixel value from NDVI multitemporal data. Generally, MVC is used to minimize the effects of cloud contamination, to reduce directional reflectance and off-nadir viewing effects, and also to minimize the sun's angle and shadow effects as well as other atmospheric effects (Maselli *et al.,* 1993). For dekadal images, taking the highest of the pixel value for the dekad from all the daily composites produces MVC NDVI maps. This minimizes data gaps in any particular composite due to cloud interference or missing data and overcomes some of the systematic errors that reduce the vegetation index (NDVI) value. Maselli *et al.* (1993) used this approach for cloud correction. The systematic errors that are overcome by MVC include: larger solar zenith angles, large scan angles, atmospheric effects and cloud shadows.

2.1.2 Interpretation of NDVI

According to Kanemasu *et al.* (1990) and Liping *et al.* (1994), NDVI can be used as an indicator of relative greenness and standing biomass of a crop. They also stated that if sufficient ground data is available, the NDVI could also be used to calculate and predict primary production, dominant species, grazing impacts and stocking rates. NDVI is also highly correlated with percentage cover of the ground, leaf area index (LAI) and intercepted or absorbed photosynthetically active radiation (PAR), hence photosynthesis (Groten, 1993). Since photosynthesis occurs in the green parts of plant material and it is also correlated to NDVI, the latter is normally used to estimate green vegetation. Other studies have also found the relationship between NDVI and climatic variables like ENSO and precipitation (Maselli *et al.,* 1993 and Sannier *et al.,* 1998). The significant relationships that have been established between climatic factors (precipitation, evapotranspiration and temperature) incited interest in scientists to investigate the relationship between NDVI and yield/ biomass production (Ji and Peters, 2004). The interpretation of NDVI images is difficult because of the following reasons: Firstly, an explicit relationship between NDVI and a variety of vegetation conditions is not yet available. Secondly, in one pixel there are usually different vegetation types which are not distinguished by the satellite sensors. Thirdly, geographical variations may have an influence on the satellite image (Sannier *et al.* 1998).

For most seasonal vegetation types the NDVI curve consists of three portions, as illustrated in figure 2.1 for Nsanje RDP. In the first portion, the NDVI will be increasing in association with the active growth of the vegetation (represented by the segment between dekad 28 and dekad 2 in figure 2.1); the second portion consists of the maximum part of the curve corresponding to the peak vegetation cover (represented by the segment between dekad 3 and dekad 9 or 10). In the third part of the curve the index values will be decreasing corresponding to the maturing (senescing) phase, represented by the segment from dekad 11 to the end of the profile, of the vegetation (Potdar, 1993)

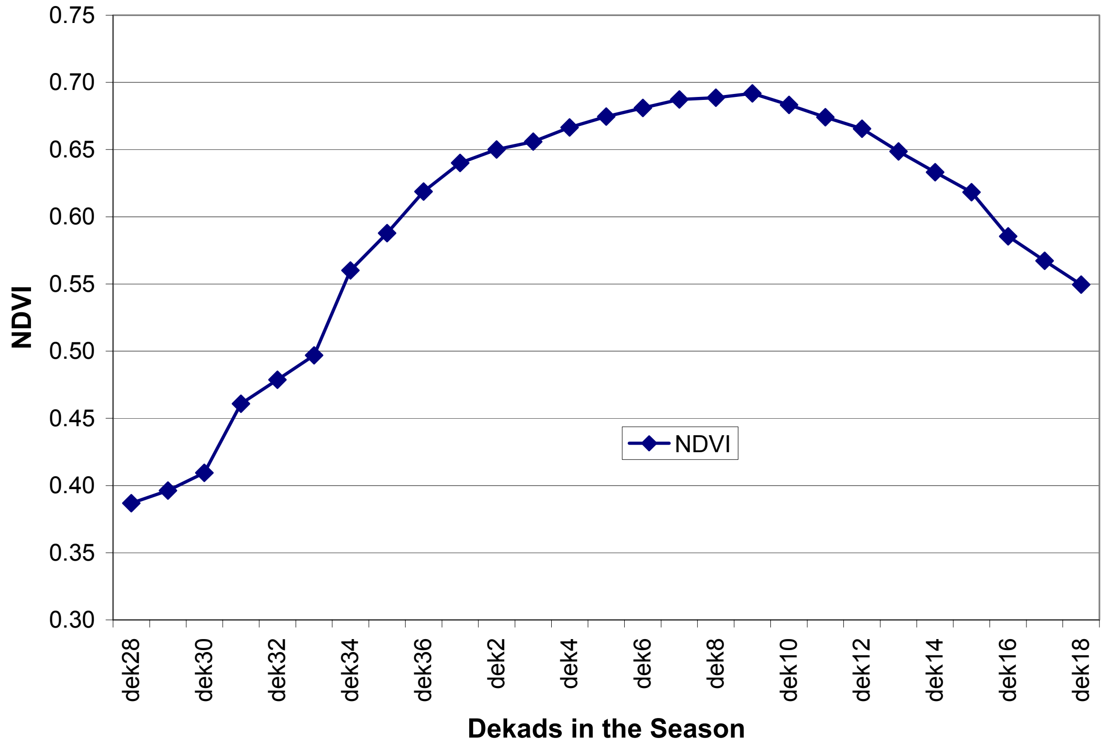


Figure 2.1: An illustration of a general profile of NDVI through the growing season, from Nsanje RDP

On the other hand, NDVI curves can be divided into two categories according to the time when maximum NDVI was attained and the temporal development of the curve. The first category represents NDVI curves with well-defined maxima that occur earlier in the season, followed by a significant decrease. This pattern is often attributed to the heterogeneity of vegetation in the pixels. In this case, the pixels have a mixture of natural vegetation (trees, bushes, and herbaceous vegetation) and agricultural crops. The second category represents those with less distinct maxima occurring later in the season and followed by a period with a very low rate of decrease in the NDVI values. This represents a situation in which there is a mixture of grass and herbaceous vegetation and agricultural crops that have different maximum NDVI values and with these maxima being attained at different times. This type of NDVI curve represents the most intensively farmed areas (Rasmussen, 1992).

Rasmussen (1992) stated the assumption that is made when relating a vegetation index like NDVI, to crop yield and / or production is that natural vegetation is affected by climatic and other factors that also affect crop growth. Hence the degree of greenness would reflect the growing conditions for crops as well. However, crops may be affected among others by hazards such as lack of nutrients, pests and diseases. In such a case, it is recommended to apply appropriate non-remote sensing corrections to the NDVI-yield model estimates if the information is to be meaningfully used.

2.2 Applications of NDVI

One of the uses of the NDVI is for the real-time monitoring of vegetation development. According to Sannier *et al.* (1998), several studies have been carried out to assess crop yields or savanna primary production using NDVI. An example is the work by Groten (1993) in which a statistical method was developed for forecasting crop yield per unit area from NDVI profiles for Burkina Faso. Sannier *et al.* (1998), using the 7.6 km resolution Africa Real Time Environmental Monitoring System (ARTEMIS) NDVI, found a direct relationship between the variation in annual development of vegetation in agricultural areas and the variation in production. They also found a strong relationship between NDVI and rainfall in semi-arid environments.

Rembold and Nègre (2004) in a study for the agricultural regions of Somalia obtained high correlations between vegetation index and production for both sorghum and maize. Four NDVI parameters were tested namely seasonal maximum NDVI, seasonal average NDVI, cumulative NDVI until the peak of the season and monthly NDVI average during the whole crop season. It was found that the NDVI average for June (the middle month of the growing season) gave the highest correlation in the Bay agricultural region for sorghum, and in the Lower Shabelle region for maize. For the second growing season no unique NDVI parameter was directly linked to production. The failure to find a significant relationship was attributed to the erratic character of the rainfall in the regions under study. Therefore the regression analysis was done only for the main growing season on the assumption that the mean monthly NDVI average for June plays a key role in the assessment of the final production. Unganai and Kogan (1998) correlated NDVI derived vegetation condition index (VCI) and temperature condition index (TCI) with maize yield in Zimbabwe. They found that district averaged VCI and TCI for each week correlated highly with corresponding crop yield (r^2 values between 0.70 and 0.95). These higher correlations were mainly in the principal maize growing regions of the country with weaker relationships in tea and coffee growing regions. Weekly Pearson correlation coefficients of maize yield anomaly on VCI and TCI were used to identify the critical periods of high sensitivity depending on the magnitude of the correlation coefficients. Through this process, two periods were identified, namely the phase of intense photosynthetically active and biomass accumulation and the reproductive phases. These represented the periods when maize had the highest sensitivity to thermal conditions and water stress (TCI and VCI, respectively). AVHRR based regression yield models were developed using VCI during the week of peak correlation and also the VCI integrated over a three-week period up to the peak. It was observed that consistently marginally lower value or r^2 were observed by using the latter approach. Using TCI and VCI at their respective weeks of peak correlation significantly improved the r^2 for the regions where neither of them explained a significant proportion of the maize yield variation. Rasmussen (1992) used AVHRR-NDVI data for the assessment of millet yield in Burkina Faso and singled out NDVI integrals for the vegetative, reproductive period and for the whole growing season as being very significantly related to crop yield. The integral over the reproductive period gave the highest correlation ($r^2 = 0.89$). Comparing the correlation of yield and biomass with integrated NDVI, the former gave better correlation.

Groten (1993) investigated the possibility of using NDVI for crop monitoring and early yield assessment of millet in different areas of Burkina Faso. She tested NDVI parameters such as maximum dekadal/monthly NDVI, NDVI integrated up to some point during the season, and dekadal/monthly NDVI increments. The NDVI-yield regression based on simple maximum NDVI values proved to be superior to the others. It was also found that multiple linear regression analysis led to significantly higher correlation coefficients, especially when done towards the end of the growing season. Regression results of monthly NDVI parameters showed that reasonable yield forecasts can be made with at most two months' lead time to harvest. Mutikani (1997) carried out statistical analysis of NDVI-maize yield relationships at communal area level in Zimbabwe. She found that generally maize yield was significantly correlated with the dekadal NDVI averages.

NDVI is also useful both in the mapping of the presence of vegetation on a pixel basis and provides measures of the amount or condition of vegetation within a pixel. For example, Menenti *et al.* (1993) used the AVHRR NDVI to classify land vegetation types, while Wan *et al.* (2004) used NDVI for monitoring vegetation growth conditions. In other cases NDVI has been used for mapping the start of the growing season and sowing dates for some crops (Groten, 1993). The importance of this is that results can be used for the definition of the starting time for dynamic crop simulation modelling. Groten (1993) also investigated the use of NDVI in monitoring the quality of the growing season. This was done by comparing observed current NDVI data with historical data, optimum or averaged reference curves. Negative deviations from reference curves imply that the season is not progressing normally. However, Groten suggested that if negative deviations from a reference curve are to be related to yields, critical stages of the crop development cycle when these were observed should be taken into consideration.

Studies by Maselli *et al.* (1993) have shown that in the Sahelian areas of the Niger, NDVI can be a good indicator of crop productivity especially in areas where crop productivity is highly variable due to environmental conditions. In this study final yield was regressed against relative NDVI values at the middle of the season in order to investigate the possibility of using NDVI as a tool for predicting agricultural production. The results showed that the correlations were relatively low suggesting that the yield-NDVI relationship was not well defined in the study area. Thus, in that particular study area, the use of NDVI at the middle of the rainy season as a yield predictor was not recommended. However NDVI at that point could provide information about the primary productivity of an area when all the other environmental parameters are taken into account. The other use of remotely sensed data in agricultural production is acreage estimation, and this needs high resolution instruments (Sridhar *et al.,* 1994). This is considered the first step of crop production forecasting using an all NDVI approach. The next step is yield estimation. The product of acreage and yield thereafter gives production. It was found that remote sensing methods underestimated acreage in the majority of the districts that were studied due the difficulty in identifying unirrigated wheat training signatures as well as omission errors during land classification. On yield and production forecast, the yield-spectral index relation that was developed generally gave an underestimation that was due to the use of a single regression equation for the whole of the study area (districts).

NDVI is also used for drought monitoring. For example, Wan *et al.* (2004) used MODIS NDVI and land surface temperature (LST) products for developing a near real-time drought monitoring approach in the southern Great Plains of the United States of America (USA). They found that NDVI was related to total monthly cumulative precipitation (TPCP) combinations for 3-, 6-, 9- and 12-month intervals for the whole study area, croplands and grasslands. The correlation coefficients for croplands were stronger (and positive) than for either grasslands or for the whole study area. It was also found that there was a month's time lag between precipitation and NDVI response to the precipitation event. The study also established that the vegetation temperature condition index (VTCI) is related to both recent rainfall events and past rainfall amounts. These results were interpreted to show that VTCI is a near-real time drought monitoring approach. Unganai and Kogan (1998) used two remotely sensed products namely Vegetation Condition Index (VCI) and Temperature Condition Index (TCI) for drought detection in the Southern African region. These parameters together provided a reliable additive drought detection and crop condition assessment scheme. The additive expression for vegetation condition assessment and drought mapping, VTI, is defined as

$$VTI = r_1 * VCI + r_2 * TCI \tag{2.2}$$

where VTI represents a combined vegetation and temperature index, r_1 and r_2 are the weighting factors of VCI and TCI respectively.

The three parameters, VTI, VCI and TCI were identified to be potentially powerful, cost-effective and reliable tools for more spatially complete and comprehensive regional drought detection and mapping in Southern Africa.

Precipitation and potential evapotranspiration have been identified as the most significant climate variables that control vegetation at an annual time scale (Ji and Peters, 2004). This implies that the soil water balance is an important factor in controlling vegetation condition. In a study that tried to establish the relationship between AVHRR-NDVI and climate in the northern Great Plains of America, it was found that climate accounted for 46% and 24% NDVI variation in grassland and cropland, respectively. It was therefore concluded that since the index was highly correlated to green-leaf density, it could be used as a proxy for aboveground biomass. Ji and Peters (2004) also suggested that NDVI can be used for the evaluation of environmental and climatic changes at regional and global scale, arguing that climate significantly contributes to vegetation condition hence the change in index over time must be related to the environmental and climatic changes.

According to Rojas (2004), crop condition is monitored by observing the difference of the degree of greenness between consecutive dekadal NDVI images. A case study was conducted in Malawi by Rojas during the 2003/04 growing/cropping season. Dekadal NDVI images were compared with a five-year average for respective dekads and for each RDP. The difference between these sets of images was used to give an inference of how the 2003/04 crop yield would compare with the five-year average yield. A negative difference observed for some RDPs in the southern region of the country implied that the yield would be lower than the five-year average yield. In the rest of the districts, average to above average yields were expected. However, these could not be validated since yield data for the 2003/04 season was not available. The advantage of using the vegetation index is that it allows monitoring of the crop season every ten days. Rojas suggested that if a series of yield data is available, there is possibility of developing regression models between NDVI and the crop yield. Because of the difficulty in separating crops from other vegetation types like trees and shrubs, and also separation of the various crops in different fields, the accuracy of this approach is still not the best.

Currently the use of NDVI for yield estimation is not popular in Malawi due to lack of equipment and qualified personnel for analysing and interpreting NDVI images in the country.

2.3 Other Yield Estimation Methods Currently Used in Malawi

2.3.1 Crop estimate survey (CES)

The crop estimate survey (CES) involves data collection from field surveys by extension officers from the Ministry of Agriculture and Irrigation of Malawi. There are three crop estimate surveys during the growing season (Rojas, 2004).

(a) First round estimate

This involves calculation/mapping of areas planted for each crop. The assessment is carried out as soon as possible after crop establishment for most areas in the country, that is, between second dekad of December (dekad 35) and the second dekad of January (dekad 2). The enumerators find out from farmers the status of the crop, in which case the farmers compare planted area and the current crop stand with that at the same time in the previous season(s). The farmer qualitatively assesses, with reasons, whether harvest will be lower, better or same as the previous season(s). The enumerator then puts a score of percentage change of yield for the farmer in comparison with the previous season. Ideally, all the planted areas are to be inspected but only a representative sample is used. These figures are compiled for administrative units namely the Extension Planning Areas (EPAs), Rural Development Projects (RDPs), and Agricultural Development Division (ADDs), and are discussed at the CES meeting for all the stakeholders that take place at the end of January every year. At this meeting national figures are compiled. The advantage of this method is that it takes place at a convenient time to survey planted areas due to early stages of the crops. However, this is considered too early to estimate yield because high yield variations occur due to what happens in the later parts of the season. Financial and human resource constraints limit the timeliness and accuracy of the procedure.

(b) Second round estimate

At this stage, the enumerators visit sample fields under the crop to assess the developments in terms of crop stand, comparing between the current situation and the time when the previous crop was at that stage. The assessment is similar to the first round, that is, crop stands are classified as being the same, better or lower than the previous season's crop stand. Figures are similarly compiled and discussed at CES meeting for all stakeholders at the end of March. The advantage of this assessment over the first round is that farmers have a better idea of the expected production. The procedure may be subjective since the farmers' judgment may depend on personal opinion. Financial and human resource constraints limit the yield estimation at this stage.

(c) Third round (post-harvest) estimate

The CES is carried out at or soon after harvest. Samples of harvest units are used to determine the quantity of harvests. Harvest units used include bags, baskets, ox-carts and others. Figures are once again compiled and discussed at the last CES meeting for the growing season that is held either in May or early June. The third survey is advantageous in that quantitative measurements of the production, which are very useful in the calibration of other approaches to yield or yield estimation, are done. However, the timing is considered too late for early warning purposes. Like the other methods, financial constraints limit the success of this approach.

This whole process of CESs requires considerable human and financial resources that are not easily available rendering the method ineffective. Therefore other methods of estimating yield need to be considered.

2.3.2 Agrometeorological yield estimation

In this approach, the FAO water requirement satisfaction index (WRSI) is calculated using crop water balance models (Nayava and Munthali, 1992). This index indicates crop performance based on the crop water satisfaction during the growing season. WRSI for a season is calculated as the percentage ratio of the seasonal actual crop evapotranspiration (AET_c) to the seasonal crop water requirement (WR), which is the same as potential crop evapotranspiration (PET_c). Mathematically WRSI can be expressed as

$$WRSI = \frac{\sum AET_c}{\sum PET_c} * 100 \tag{2.3}$$

PET_c denotes crop specific potential evapotranspiration after adjustment is made to the reference (potential) crop evapotranspiration (PET) by the use of appropriate crop coefficients (K_c). AET_c and PET_c are summations over the period concerned. These can be done on different time scales like daily, dekadal, or monthly, depending on what type of data is being used and the use to be made of the outputs. K_c values define the water use pattern of a crop. Allen et al. (1998) published the K_c values for critical points in a crop phenology and intermediate values are linearly interpolated. The K_c values for maize have been given as 0.3, 0.3, 1.20, 1.20, and 0.35 for the times corresponding to 0%, 16%, 44%, 76%, and 100% of LGP, respectively (see figure 2.2). The water requirement of the crop (PET_c) at a given time in the growing season is calculated by multiplying standard reference crop PET by the crop coefficient (K_c), that is,

$$PET_c = K_c PET \tag{2.4}$$

On the other hand, AET_c represents the actual amount of water withdrawn from the soil reservoir in a particular period (day, dekad or month) where the shortfall relative to potential crop evapotranspiration (PET_c) is calculated by a function that takes into account the amount of available soil water in the reservoir.

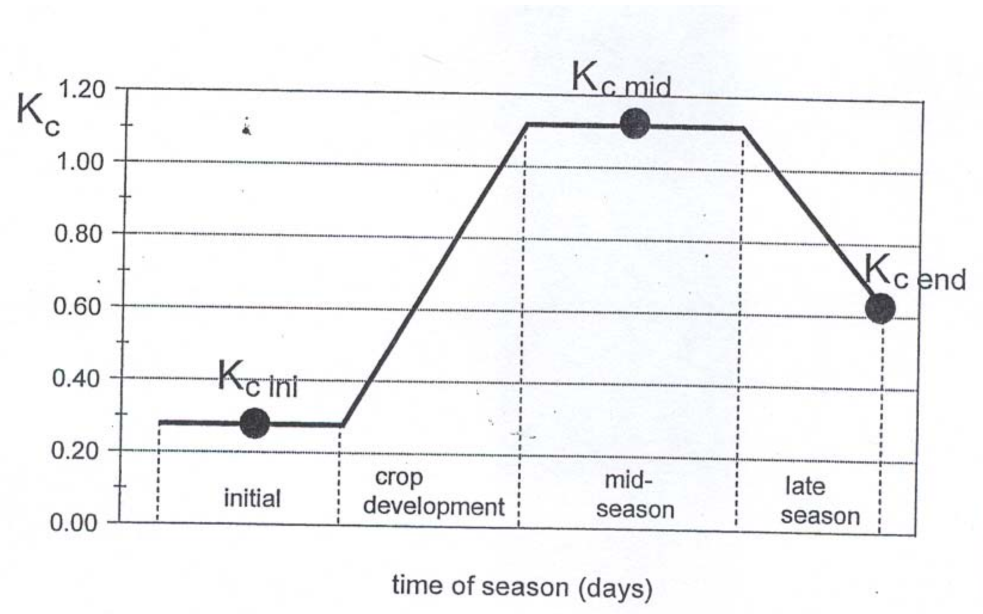


Figure 2.2: Crop coefficient curve (from Allen et al., 1998)

WRSI was found to be related to crop production using a linear yield-reduction function specific to a crop (Doorenbos and Pruitt, 1984). At any point in the growing season, the WRSI calculation continues (is forecast) till the end of the growing season using long term average potential evapotranspiration and rainfall data. Then these forecast final indices are fed into regression equations that were developed by Nayava and Munthali (1992). A yield projection is made for each Rural Development Project (RDP), which are later aggregated to Agricultural Development Divisions (ADDs) and the whole country. These results are discussed together with the ones from CES. In the recent past, the two approaches produced very similar results. For example, in the 2000/01 season the agrometeorological model maize production national estimate was 2.2 metric tons while that from the ministry of Agriculture was 2.1 million metric tons. However, this figure was reduced to 1.8 in the final estimate due to flooding that occurred in some parts of the country in early 2001. For the 2001/02 season, total maize production national estimate from the model was 1.8 million tons while the final outcome was 1.6 metric tonnes (Mukhala and Magadzire, 2004). However, FAO replaced the FAOWRSI as a tool for agrometeorological crop yield estimation in the Southern Africa Development Community (SADC) region by a new tool, the AgroMetShell (AMS). Part of this study investigated the possibility of using AMS water balance model and developed regression models for yield estimation at RDP level in Malawi.

2.4 Water Balance of a Cropped Soil

In semi-arid regions, water is the most important environmental factor that determines crop yield (Mukhala and Hoefsloot, 2004). The FAO method based on a water balance calculation scheme helps to determine how the agriculture season has performed and how far the crop's water requirements have been satisfied in order to achieve a potential yield. The index that is the output of a water balance calculation represents the percentage of needed water the plant has received during its cycle. If all other factors (crop diseases and other hazards) are kept constant, the index is closely related to yield in semi-arid countries. The new revised water balance model (AMS), apart from the water requirement satisfaction index (WRSI), produces some other parameters that may be used to estimate crop yields. These parameters include: actual evapotranspiration (AET), excess water (WEX) and water deficit (WDF) at various phenological stages.

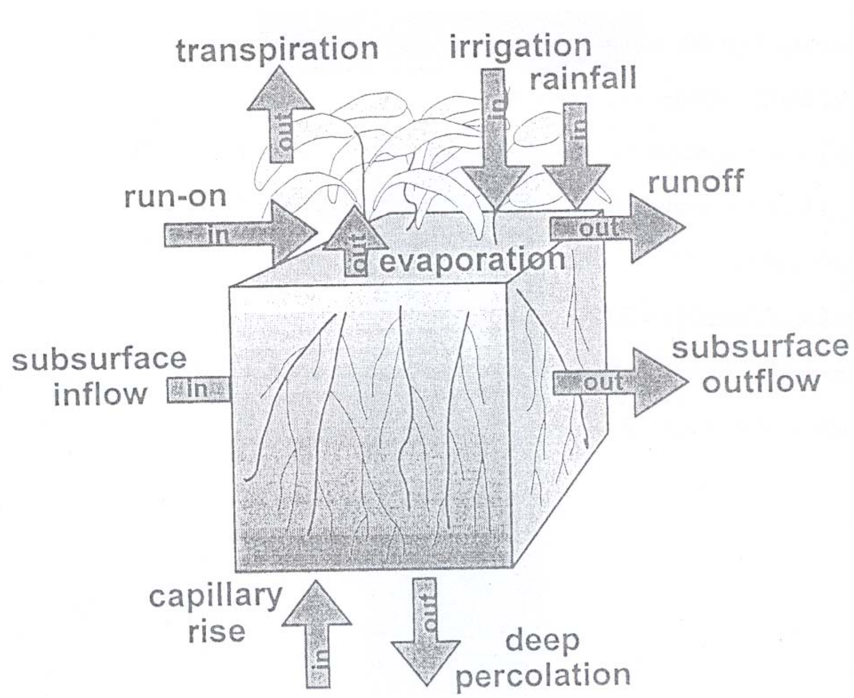


Figure 2.3: Soil water balance of the root zone (from Raes, 2001)

Figure 2.3 shows the different processes that either add or subtract water to and from the root zone. Water is added to the soil by rainfall (P), irrigation (I), surface run-on (RO_{in}), capillary rise (CR) and subsurface inflow (SF_{in}). On the other hand, water is lost from the root zone (soil reservoir) by soil evaporation (E), crop transpiration (T), surface runoff (RO_{out}) subsurface outflow (SF_{out}), and deep percolation (DP) from the soil reservoir. However some fluxes such as subsurface flow, deep percolation and capillary rise from a water table are difficult to assess over short periods and are ignored in most calculations (depending on the level of accuracy required for any calculation). Therefore, at any particular point in time the amount of soil water retained in the root zone can be calculated by the relationship given in equation 2.5 below and re-illustrated in figure 2.4.

$$W_{r,t} = W_{r,t-\Delta t} + \left(P_{\Delta t} - RO_{\Delta t}\right) + I_{\Delta t} + CR_{\Delta t} - ET_{\Delta t} - DP_{\Delta t} \qquad (2.5)$$

where $W_{r,\,t\text{-}\Delta t}$, $WR_{r,\,t}$, is the amount of soil water (mm) stored in the root zone at the beginning and end respectively of the considered time period, ($P_{\Delta t}$ - $RO_{\Delta t}$) is the precipitation minus runoff (effective rainfall) across the top boundary of the root zone during the given time period, $I_{\Delta t}$ is the irrigation (mm) across the top boundary of the root zone during the given time period, $CR_{\Delta t}$ is the capillary rise (mm) across the bottom boundary of the root zone during the given time period, $ET_{\Delta t}$ is the water extraction by plant roots and soil evaporation (mm) during the given time period, and $DP_{\Delta t}$ is the deep percolation (mm) across the bottom boundary of the root zone during the given time period (Raes, 2001).

Alternatively, the amount of soil water retained in the root zone can be expressed as a water shortage in the soil called root zone depletion (D_r), which is an inverse of W_r and the water balance computation can be expressed as

$$D_{r,t} = D_{r,t-\Delta t} - \left(P_{\Delta t} - RO_{\Delta t}\right) - I_{\Delta t} - CR_{\Delta t} + ET_{\Delta t} + DP_{\Delta t} \qquad (2.6)$$

where $D_{r,\,t\text{-}\Delta t}$, $D_{r,\,t}$ is the root zone depletion at the beginning and end of the considered time period (mm) (Raes, 2001).

The effective rainfall ($P_{\Delta t}$ -$RO_{\Delta t}$) depends on the rainfall depth ($P_{\Delta t}$), rainfall intensity, the slope of the land, soil type, its hydraulic conditions and the antecedent moisture content and the land use cover (Raes, 2001).

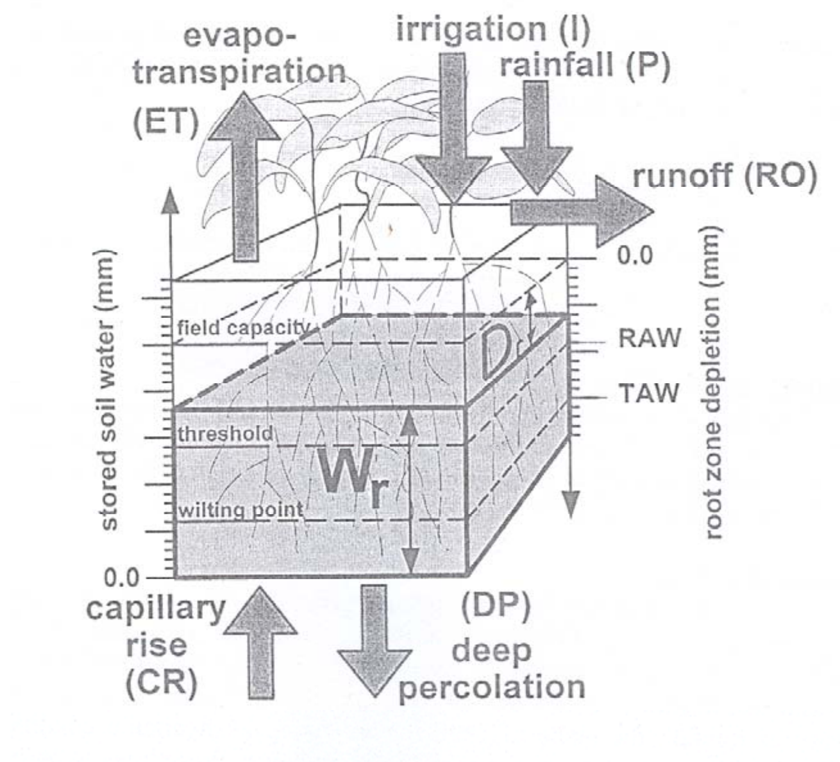


Figure 2.4: Root zone as a reservoir (from Raes, 2001)

From figure 2.4, field capacity (FC) determines the amount of water that the root zone can retain when the soil has drained after being wetted. It represents the upper limit of plant extractable water. The amount of soil water that the root zone retains at field capacity is given by the expression

$$W_{r,FC} = 1000\theta_{FC}Z_r \tag{2.7}$$

where $W_{r,\,FC}$ is the soil water depth in the root zone at field capacity (mm), θ_{FC} is volumetric soil water content at field capacity (m^3/m^3) and Z_r is the rooting depth (m) (Raes, 2001).

On the other hand, permanent wilting point (WP) is the soil water content at which plants stop extracting water and will permanently wilt. It represents the lower limit of the plant extractable water. The amount of water that may be still present in the root zone at permanent wilting point is given by the expression

$$W_{r,WP} = 1000\theta_{WP} Z_r \tag{2.8}$$

where $W_{r,\,WP}$ is soil water in the root zone at permanent wilting point (mm), θ_{WP} is volumetric soil water content at permanent wilting point (m^3/m^3) (Raes, 2001).

The difference between field capacity and permanent wilting point of the soil is termed the total available water (TAW). It is the amount of water a crop can theoretically extract from the soil and represents the amount of water held in the soil between FC and WP. TAW can be calculated from the values of FC and WP or by the expression

$$TAW = 1000(\theta_{FC} - \theta_{WP})Z_r \tag{2.9}$$

all terms have their meanings as defined above.

Readily available soil water (RAW) refers to the maximum amount of water that can be depleted below field capacity without inducing stress. RAW can be expressed as a fraction of the total available soil water using the expression:

$$RAW = pTAW = 1000p(\theta_{FC} - \theta_{WP})Z_r \tag{2.10}$$

where p is the depletion factor (the fraction of TAW that can be depleted from the root zone before moisture stress occurs) and Z_r is the effective rooting depth (m) (Raes, 2001).

2.5 Maize Crop Requirements

Maize is an important cereal for both human and livestock consumption in the Southern African region. In Malawi, it is largely grown at small-scale level. Large-scale commercial maize farming is not common in Malawi due to land limitation. The historical maize yield data that has been used for this study represents small-scale farmers' (post-harvest estimate) yield figures. In Malawi, for many years maize production was largely rain fed, but with the introduction of modern irrigation technology in some parts of the country, winter cropping is becoming popular thereby supplementing the rain fed maize production. This will have a positive impact on the food security of the country.

Maize is grown both in temperate and tropical climatic regions. The crop is tolerant to hot and dry atmospheric conditions (up to a temperature of 45°C) as long as sufficient water is available. The optimum mean daily temperature for maize germination is between 18°C and 20°C. The length of the growing period depends on both the variety and the environmental factors of the region. Doorenbos and Kassam (1986) stated that for daily mean temperature higher than 20°C during the growing season, the early maturing varieties take between 80 and 110 days while the medium maturing varieties take between 110 and 140 days. On the other hand, at mean daily temperatures of less than 20°C, time to mature takes 10 to 20 days longer for each 0.5°C decrease in temperature (depending on variety). Norman et al., (1984) stated that grown at same latitude, duration to maturity of maize increases at the mean rate of 7.6 days per 100 m increase in altitude. This can be explained by the fact that temperature decreases with altitude hence maize take a longer period to mature at higher altitude than at lower altitude. In terms of growing degree days (GDDs), maize require between 1800 and 3700 GDDs to maturity. It is a short day and/or day-neutral crop. Maize plant prefers well-drained and aerated loamy soils. The fertility demands of grain maize are relatively high and amounts of up to 200kg/ha Nitrogen, 50 to 80 kg/ha phosphorus and 60 to 80 kg/ha Potassium are recommended for high –yielding varieties.

2.5.1 Developmental phases of a maize plant

The maize plant undergoes several developmental phases with different durations. According to Doorenbos and Kassam (1986) the developmental phases are as follows: Establishment (initial) stage (15 to 25 days), vegetative stage (25 to 40 days), flowering stage which includes silking and tasseling (15 to 20 days) and the final stage that includes yield formation (35 to 45 days) and ripening (10 to 15 days). These are illustrated in the figure 2.5.

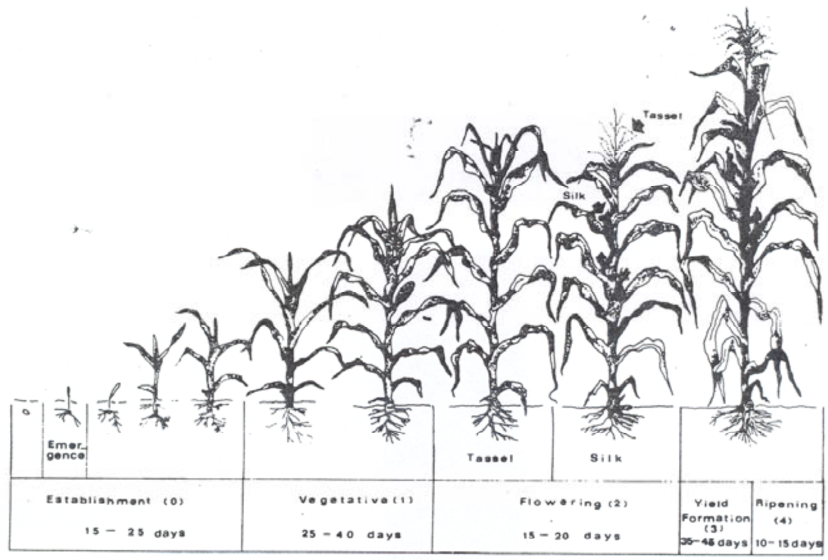


Figure 2.5: Growing periods of maize (from Doorenbos and Kassam, 1986)

2.5.2 Crop water requirement for maize and effects of water deficits on yield

According to Doorenbos and Kassam (1986) and Jackson (1989), the total water requirement for the crop depends on the length of the growing periods for a particular variety. For medium maturing varieties, between 500 to 800 mm of water is required (excluding wastages during conveyance). On the relationship between water supply and crop yield, it was found that the reproductive stages (flowering to yield formation) are critically affected by water shortages and these affect the final yield. The flowering stage includes tasseling, silking and pollination. Water deficits at these stages results in reduced grain number per cob. If the water deficit at flowering (especially during silking and pollination) is severe, it may result in little or no grain yield due to the dying of silks. The yield reduction due to water deficits during yield formation is because of the resultant grain size reduction. The effects of drought (water deficit) at flowering were also summarized by Norman *et al.* (1984) who outlined that both leaf area and leaf photosynthetic rate are reduced especially during the stress period, though leaves may recover. Furthermore, silking is delayed thereby reducing the grain yield components. The reduction in grain number has been attributed to increased asynchrony in flowering since water deficit reduces the rate of pollen production during the period the silks are receptive and also reduces the period when silks are exposed to pollen. However, pollen viability is never affected by water deficits (Norman *et al.,* 1984).

Soil moisture during flowering and early grain formation has been found to be critical in the determination of yield. However, in a study conducted in East Africa, Jackson (1989) identified the germination, fertilization and grain filling periods as being the most essential in the determination of the effects of water deficit on final yield, explaining that water stress in early growth delays flowering. Stress-induced delay in silking leads to loss of synchrony in the development of silks and tassels resulting in adverse effects on the final yield. Water stress at fertilization lowers yield since the nutrients fail to dissolve so that they might become available to the crop. Similarly, water logging conditions during the fertilization period have the effect of leaching the minerals out of the reach of the plant roots.

2.5.3 Water-yield relationships

The water stress effects on crop yield can be quantified by considering the relationship between relative yield decrease and relative evapotranspiration deficit. This relationship can be summarized by the equation below.

$$\left(1-\frac{Y_{act}}{Y_{\max}}\right)=k_y\left(1-\frac{ET_{act}}{ET_{\max}}\right) \tag{2.11a}$$

or

$$Y_{act}=Y_{\max}\left[1-k_y\left(1-\frac{ET_{act}}{ET_{\max}}\right)\right] \tag{2.11b}$$

where Y_{act} is the actual harvested yield (kg/ha), Y_{max} is the maximum (potential) harvested yield (kg/ha) - without environmental/ water stress, k_y is the yield response factor, ET_{act} is the actual evapotranspiration (mm), and ET_{max} represents maximum evapotranspiration (mm).

k_y is a factor that describes the reduction in relative yield according to the reduction in ET_{act} caused by soil water shortage; k_y varies with growth phase of the crop. Maximum yield (Y_{max}) is established by the genetic characteristics of the crop and by the degree of crop adaptation to the environment (Doorenbos and Kassam, 1986). Berka et al. (2003) adjusted Y_{max} (kg/ha) to the maximum yield achieved for a healthy crop without water and nutrient deficiencies.

CHAPTER 3

MATERIALS AND METHODS

3.1 Study Area

This study focussed on the whole of Malawi because there was need to evaluate the degree to which various yield estimation methods could be applied in various Rural Development Projects (RDPs) of the country so that in the end the models could be adopted for use to come up with national yield figures.

Malawi is located in the eastern side of the southern African region between latitudes 8.8°S and 17.5°S, and longitudes 32.0°E and 36.5°E. The country shares its borders with Mozambique to the south, southeast and southwest, Zambia to the west and Tanzania to the northwest, north and northeast. It covers a total area of 118483 km^2 of which 92275 km^2 is land. The country is divided into eight large agricultural administrative regions called Agricultural Development Divisions (ADDs), which are further divided into smaller units called RDPs. There were 30 RDPs prior to the 2002/03 season when they were revised to only 25. The RDPs are further divided into smaller units called Extension Planning Areas (EPAs). All the agricultural data collection is initially done at EPA level and then it is aggregated to RDP, ADD, and national levels depending on the use the data is meant for. Topographically, there are mountain ranges in various parts of the country especially to the north-western and south-eastern parts. There are also a few lakes, including Lake Malawi, the third largest in Africa, which runs from north to south on the eastern side of the country and lies along the great rift valley of Africa.

3.1.1 Climatological zones of Malawi

Malawi is divided into five climatological zones. The zoning was done using meteorological parameters like rainfall, temperatures and potential evapotranspiration, among others. Efforts by the author to get reference material on the zoning process and features/ characteristics of each zone proved futile. However, table 3.1 below gives the summary of the RDPs in each of the five zones.

Table 3.1: RDPs in each climatological zone of Malawi

ZONE	RDPs
Northern	Chitipa, Rumphi/Mzimba, Mzimba Central and Mzimba South
Central	Ntcheu, Bwanje Valley, Thiwi-Lifidzi, Dedza Hills, Lilongwe East, Lilongwe West, Dowa East, Mchinji, Dowa West, Ntchisi and Kasungu
Lakeshore	Karonga, Nkhata Bay, Nkhota kota, Salima and Mangochi
Southern	Namwera, Balaka, Zomba, Mwanza, Shire Highlands, Phalombe, Mulanje and Kawinga
Shire Valley	Chikwawa, Nsanje

3.2 Data

3.2.1 Climatic data

The climatic data used in the study was for various periods from 1961 to 2002. The data was for seventy-five (75) early warning stations in Malawi (Figure 3.1 and Table 3.3). The data was obtained from the Malawi Meteorological Services Department, FAO publications, and the SADC Regional Remote Sensing Unit (RRSU).

The dataset included:

- Actual (1983 to 2004) and long-term (1961 to 1990) average dekadal rainfall,
- Actual and normal dekadal potential evapotranspiration (PET),
- Soil water holding capacities (SWC), and
- Maize crop parameters published by FAO (Allen *et al.*, 1998).

3.2.2 Satellite data

The satellite data was obtained from the SADC Regional Remote Sensing Unit (RRSU), Harare. The RRSU receives NDVI images on a ten day basis from NASA Goddard space Flight Centre in Maryland, USA, via United States Agency for International Development (USAID) Famine Early Warning Systems Network (FEWSNET) project.

The data used constituted the dekadal NOAA-AVHRR NDVI images in Global Area Coverage (GAC) format and at a resolution of 7.6 km for the period 1981 to 2003. A new NDVI dataset was used instead of the old one that has been used in previous studies. This new dataset has never been tested in any studies, as far as the literature reviewed by the author of the current study is concerned. The images were rearranged so that each season comprised of 36 dekads from June of one calendar year to July of the next. Since the growing season in the Southern African region runs from October to April/ May, images for the same period were used. Conventionally the dekads are numbered from the first dekad of January as dekad 1 to the third dekad of December as dekad 36 (table 3.2). Using this naming convention, the growing season in the Southern African region is from dekad 28 of one year to dekad 12 of the following year. The NDVI value for each dekad represents the highest value (called the Maximum Value Composite, MVC) recorded during the ten-day period. The process of compositing daily NDVI data over the ten day period helps to minimize atmospheric interference (Townshend and Justice, 1986; Maselli et al., 1993).

Table 3.2: Dekad numbers during the year, the shaded portions indicate the growing season in the Southern African region.

Dates → Dekads ↘ Month ↓	**1 to 10**	**11 to 20**	**21 to 30/31**
January	1	2	3
February	4	5	6
March	7	8	9
April	10	11	12
May	13	14	15
June	16	17	18
July	19	20	21
August	22	23	24
September	25	26	27
October	28	29	30
November	31	32	33
December	34	35	36

3.2.3 Historical maize yield data

These were RDP-averaged yield data for 30 RDPs across the country (Figure 3.2). The dataset was for the 1983/84 to 2001/02 growing seasons and was obtained from the Planning Division of the Ministry of Agriculture and Irrigation (Malawi). The 2002/03 yield data was obtained separately and was used for testing the developed models. This data is compiled on a yearly basis by the National Statistical Office (NSO) based on the third round crop yield estimates from field reports. Since this data is obtained from post-harvest estimates that involves the actual measurements of the harvested crop, it is used the same as the measured/ actual yield. The same data was used for the development of WRSI regression equations for yield estimation in the early 1990s (Nayava and Munthali, 1992). The details of the crop yield estimation procedures were discussed in Section 2.6.

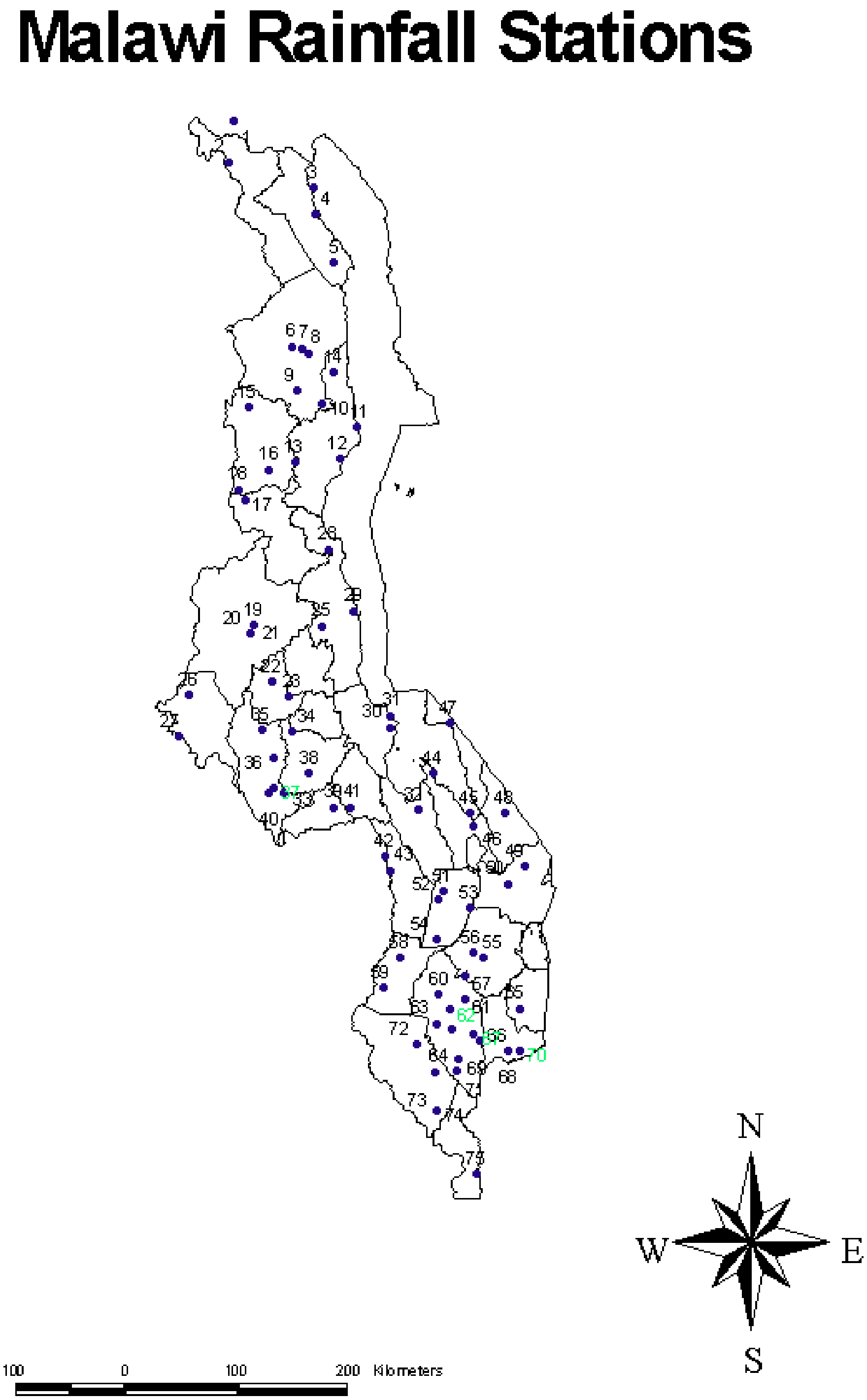


Figure 3.1: Malawi rainfall stations

Table 3.3: Early warning rainfall stations in Malawi

STATION CODE	STATION NAME	LONGITUDE	LATITUDE	ALTITUDE	RDP	DATA DURATION
1	Misuku	33.31	-9.4	1524	Chitipa	1983 to 1999
2	Chitipa	33.27	-9.7	1285	Chitipa	1983 to 2002
3	Karonga	33.95	-9.88	529	Karonga	1983 to 2002
4	Lupembe	33.97	-10.07	488	Karonga	1983 to 2002
5	Vinthukutu	34.12	-10.42	579	Karonga	1983 to 2002
6	Bolero	33.78	-11.02	1100	Rumphi/Mzimba	1983 to 2002
7	Rumphi	33.87	-11.03	1067	Rumphi/Mzimba	1983 to 2002
8	Bwengu	33.92	-11.07	1052	Rumphi/Mzimba	1983 to 2002
9	Zombwe	33.82	-11.33	1143	Rumphi/Mzimba	1983 to 2002
10	Mzuzu	34.02	-11.43	1254	Rumphi/Mzimba	1983 to 2002
11	Nkhata Bay	34.3	-11.6	500	Nkhata Bay	1983 to 2002
12	Chintheche	34.17	-11.83	495	Nkhata Bay	1983 to 2002
13	Chikangawa	33.8	-11.85	1728	Mzimba Central	1983 to 2002
14	Mbalachanda	34.12	-11.2	1318	Mzimba Central	1984 to 2002
15	Euthini	33.42	-11.45	1143	Mzimba Central	1983 to 2002
16	Mzimba	33.6	-11.9	1349	Mzimba Central	1983 to 2002
17	Mbawa	33.4	-12.12	1244	Mzimba South	1983 to 2002
18	Embangweni	33.35	-12.05	131	Mzimba South	1983 to 2002
19	Kasungu	33.47	-13.02	1058	Kasungu	1983 to 2002
20	Mwimba	33.45	-13.08	1097	Kasungu	1983 to 2002
21	KFCTA	33.47	-13.02	1085	Kasungu	1984 to 2002
22	Madisi	33.62	-13.42	1097	Dowa West	1984 to 2002
23	Mponela	33.75	-13.53	1036	Dowa West	1983 to 2002
24	Dowa	13.39	-33.56	1403	Dowa East	1983 to 2002
25	Ntchisi	34.02	-13.03	1350	Ntchisi	1983 to 2002
26	Mkanda	32.95	-13.52	1219	Mchinji	1983 to 2002
27	Mchinji	32.87	-13.82	1350	Mchinji	1983 to 2002
28	Dwangwa	34.08	-12.48	488	Nkhota kota	1983 to 2002
29	Nkhota kota	34.28	-12.92	500	Nkhota kota	1983 to 2002
30	Salima	34.58	-13.75	512	Salima	1983 to 2002
31	Lifuwu	34.58	-13.67	488	Salima	1983 to 2002
32	Nankumba	34.8	-14.35	518	Mangochi	1983 to 2002
33	Mtakataka	33.72	-14.22	624	Dedza Hills	1983 to 2002
34	Lilongwe	33.78	-13.78	1229	Lilongwe East	1983 to 2002
35	Kasiya	33.53	-13.77	1036	Lilongwe West	1983 to 2002
36	Chitedze	33.63	-13.97	1149	Lilongwe West	1983 to 2002
37	Sinyala	33.63	-14.18	1067	Lilongwe West	1983 to 2002
38	Nathenje	33.92	-14.08	1036	Lilongwe East	1983 to 2002
39	Thiwi	34.12	-14.32	1249	Thiwi-Lifidzi	1983 to 2002
40	Dzalanyama	33.6	-14.22	1219	Lilongwe West	1983 to 2002
41	Dedza	34.25	-14.32	1632	Dedza Hills	1983 to 2002
42	Mlangeni	34.53	-14.68	1067	Ntcheu	1983 to 2002
43	Ntcheu	34.58	-14.78	1245	Ntcheu	1983 to 2002
44	Monkey Bay	34.92	-14.08	482	Mangochi	1983 to 2002
45	Namiasi	35.22	-14.37	488	Mangochi	1983 to 2002
46	Mangochi	35.25	-14.47	482	Mangochi	1983 to 2002
47	Mpilipili	35.05	-13.72	678	Mangochi	1983 to 2002
48	Namwera	35.5	-14.37	899	Mangochi	1983 to 2002
49	Chikweo	35.67	-14.75	717	Kawinga	1983 to 2002
50	Ntaja	35.53	-14.87	731	Kawinga	1983 to 2002
51	Toleza	35	-14.93	689	Balaka	1983 to 2002
52	Balaka	34.97	-14.98	625	Balaka	1983 to 2002
53	Liwonde	35.22	-15.05	457	Balaka	1983 to 2002
54	Phalula	34.95	-15.27	579	Balaka	1983 to 2002
55	Zomba	35.32	-15.4	915	Zomba	1984 to 2002
56	Chingale	35.25	-15.37	610	Zomba	1983 to 2002
57	Makoka	35.18	-15.53	1029	Zomba	1983 to 2002
58	Neno	34.65	-15.4	655	Mwanza	1983 to 2002
59	Mwanza	34.52	-15.62	655	Zomba	1983 to 2002
60	Chileka	34.97	-15.67	767	Shire Highlands	1983 to 2002
61	Chiradzulu	35.18	-15.7	884	Shire Highlands	1983 to 2002

62	Chichiri	35.05	-15.78	1132	Shire Highlands	1983 to 2002
63	Mpemba	34.95	-15.88	900	Shire Highlands	1983 to 2002
64	Bvumbwe	35.07	-15.92	1146	Shire Highlands	1983 to 2002
65	Naminjiwa	35.62	-15.77	773	Phalombe	1983 to 2002
66	Luchenza	35.3	-16	678	Shire Highlands	1983 to 2002
67	Thuchila	35.25	-15.95	716	Phalombe	1983 to 2002
68	Mulanje	35.53	-16.07	686	Mulanje	1984 to 2002
69	Thyolo	35.13	-16.13	820	Shire Highlands	1983 to 2002
70	Mimosa	35.62	-16.07	652	Mulanje	1983 to 2002
71	Masambanjati	35.12	-16.22	959	Shire Highlands	1983 to 2002
72	Chikwawa	34.78	-16.03	107	Chikwawa	1983 to 2002
73	Nchalo	34.93	-16.23	64	Chikwawa	1983 to 2002
74	Ngabu	34.95	-16.5	102	Chikwawa	1983 to 2002
75	Nsanje	35.27	-16.95	60	Nsanje	1983 to 2002

Malawi RDPs

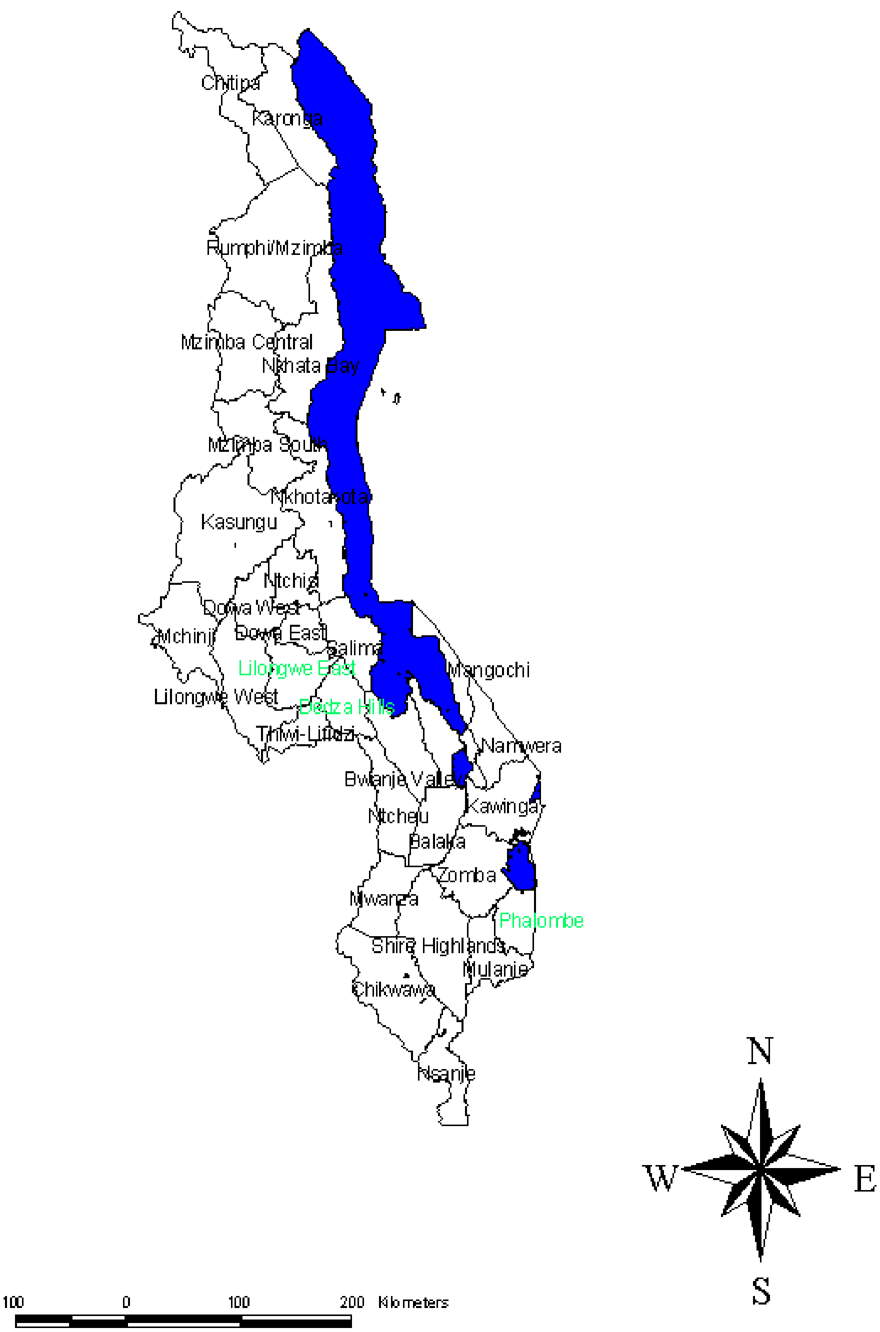


Figure 3.2: Rural Development Projects (RDPs)

3.3 Methodology

3.3.1 Extraction of statistics from NDVI images

WinDisp is a geographic information systems (GIS) software package that is used for display and analysis/interpretation of maps and images. The software was used to display and extract NDVI statistics from the satellite images. For the data to be at RDP level a blank (bna) land cover map for the country with all the RDP boundaries was used. This Malawi RDP bna map was produced by the Council for Scientific and Industrial Research (CSIR). Maps of cultivated areas were used in the analysis in an attempt to concentrate only on those areas in each RDP where cultivation is practiced. This would ensure that only NDVI relevant to agriculture would be considered in the analysis. In the typical scenario, cultivated areas can be quite scattered in a single RDPs, and several distinct and separate cultivated areas can be identified in a single RDP. To further simplify the analysis, only the biggest of these distinct cultivated areas (hereafter referred to as cultivated portions) in each RDP was selected and considered as being representative of the cultivation in the RDP. NDVI statistics were extracted for these "representative" cultivated portions.

The *Process →Series →Max/Avg* menu commands were used to analyse time series for pixels in a series of images (such as for the whole season). Either a maximum or an average value was obtained for each pixel and an image of the same was produced. This was an image in which each pixel is the result of the applied analysis to the pixels in the same location in each of a series of images.

The *Process →Stats →Max/Avg* menu commands were used to extract statistics (maximum or average) for areas within the images. The pixel values were averaged for the whole RDP for the particular dekad with the aid of a boundary (.bna) map for cultivated areas of the country's RDPs. The area averaged dekadal statistics were in the form of ASCII tables that could be displayed either as WinDisp or Excel spreadsheets.

The extracted statistics were reorganized/ reprocessed to come up with the required NDVI parameters that were later to be correlated and/or regressed with the seasonal maize yield data. The parameters that were extracted and computed included: seasonal maximum NDVI, cumulative (integrated) NDVI, seasonal NDVI averages, seasonal NDVI increments and monthly NDVI averages during the crop growing season.

- Cumulative NDVI values were obtained by adding the current dekadal value to the sum of all the previous ones from the starting point of the accumulation process.
- Monthly NDVI aggregates were calculated by the maximum value composite method whereby the highest of the three dekadal NDVI values was used to represent NDVI value for the whole month. Others have used averages of the three dekadal values to represent the monthly value.
- Seasonal averages were calculated from dekadal values by finding the arithmetic means of the NDVI values for the respective periods.
- NDVI increments were obtained by calculating the difference between consecutive NDVI values (whether dekadal or monthly). For example, the increment for the second dekad of January (dekad 2) was obtained by subtracting the NDVI value for the first dekad of January (dekad 1) from the NDVI value of dekad 2.

The NDVI parameters mentioned above were first correlated with maize yield to determine the degree of relationship between the two. Later those that gave positive and significant correlations were regressed with yield data. The same parameters were also correlated to seasonal rainfall.

3.3.2 Interpolation of rainfall data

Station rainfall was interpolated to RDP level in AgroMetShell (AMS) so that the rainfall data could be easily compared with NDVI and yield data. The interpolation procedure is as follows:

- An Excel spreadsheet file for each year was prepared that included columns for the following: station name, longitude, latitude, altitude, and dekadal rainfall values (for each of the 36 dekads in the year). The file was saved in an ASCII format.
- Each file (dekadal data) was imported into AMS.
- The interpolation itself involved two major steps whose series of commands in AMS are shown in italics below.

Step1: *Interpolation→ Make Input file→ Database…* This step specifies among other things the year, month and dekad whose data are to be interpolated and creates a data (.dat) output file.

Step2: *Interpolate→ SEDI→ Inverse Distance…* This step uses the satellite enhanced data interpolation (SEDI) method and a background layer to create an image from the data file from Step 1. SEDI is an inverse distance weighting (between the background and the parameter) method for assisted interpolation. The method can be applied to any parameter of which the values are available for a number of geographical locations, as long as a 'background' field is available that has either a negative or positive relationship to the parameter that needs to be interpolated. There are three requirements for the successful application of the SEDI method: Firstly, the availability of the parameter to interpolate as point data at different geographical locations (for example rainfall, potential evapotranspiration, crop yields), secondly the availability of a background parameter in the form of a regularly spaced grid (or field) for the same geographical area (for example CCD, NDVI, altitude) and thirdly a relation between the two parameters, either negative or positive. The SEDI method yields the parameter being interpolated as a field. In this study a flat file with constant values throughout the image, which effectively makes the SEDI operation a simple inverse distance weighting (IDW) interpolation (Charlier, 1999), was used. The final output was a series of interpolated dekadal rainfall across the country. Using WinDisp, the data were averaged at RDP level and then extracted and were displayed in Excel spreadsheets. Therefore for every RDP there were dekadal rainfall values which were then combined in different formats for further analysis.

3.3.3 Time Series Analysis

Time series analyses were performed in order to describe some features of, and relationships between climatic and satellite as well as historical yield data at RDP level. Two seasons (1988/89 and 2000/01) were identified to represent above normal and two others (1991/92 and 1994/95) represented below normal rainfall seasons respectively. The choice of these seasons was done after considering time series of seasonal rainfall totals for different RDPs. For example, time series plots for five representative RDPs of different climatological zones are given in figure 3.3. From that figure, the seasons listed above were identified as above normal and below normal rainfall seasons, respectively. Rainfall and NDVI profiles were plotted for some selected RDPs across the country. The use of the extreme seasons aimed at establishing if the rainfall performance (conditions) could be depicted in the NDVI profiles.

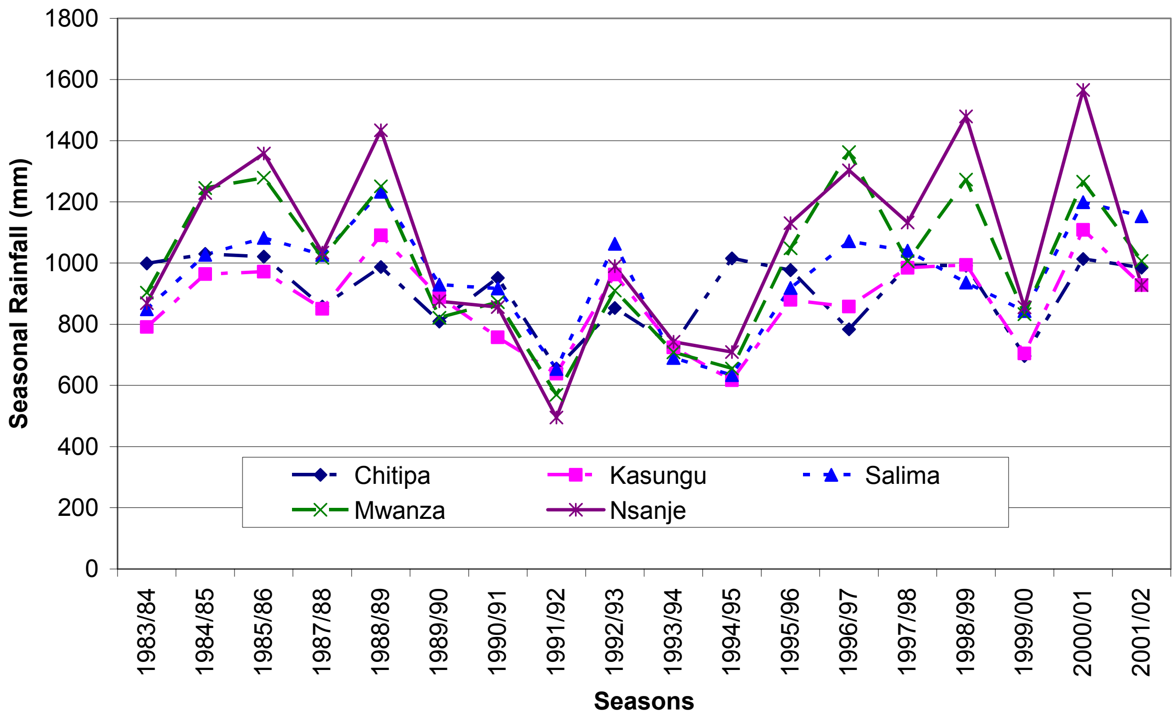


Figure 3.3: Time series for seasonal rainfall for climatological zone representative RDPs – Chitipa (North), Kasungu (Central), Salima (Lakeshore), Mwanza (South) and Nsanje (Shire Valley).

3.3.4 Brief description of AgroMetShell (AMS) water balance model

AgroMetShell (AMS) is a new crop specific water balance (CSWB) modelling software that is intended to replace the simple FAOINDEX model for agrometeorological crop monitoring and yield forecasting for early warning purposes. The FAOINDEX uses only current and extended Water Requirement Satisfaction Index (WRSI) as yield predictors. AMS gives the opportunity of choosing predictors from several of its output parameters that can be statistically related to historical yield to come up with models that may be used for yield and production estimation (Mukhala and Hoefsloot, 2004).

The AMS includes a database that holds all weather, climate and crop data that are needed to analyze the impact of weather on crops. Within AMS, the FAO crop specific soil water balance can be operated in two modes that include the monitoring mode in which analyses are made for one season for a number of stations together (for example in a province or the whole country). It is performed from the start of the growing season until harvest time. In this mode, the programme is operated in such a way that the early warning information is extracted with enough lead time before harvest to serve the purpose. The second operation mode, called the risk analysis involves analysis for one station over several years to give an idea of how a particular crop has been performing in that region in relation to water requirement satisfaction. This second mode helps to give an idea of the suitability of a particular crop to the area under study.

Inputs to the AMS water balance model include meteorological parameters such as actual and normal (long-term average) rainfall, actual and normal (long-term average) PET, the type of crops for which the water balance is calculated, percentage effective rainfall (which is a function of the terrain and type of soil), planting dekads (start of the season), soil water holding capacity, the length of the growing period (LGP), crop coefficients and irrigation amounts where applicable. The outputs from AMS include total water requirement (TWR_q) integrated at different time scales depending on the scale being used over the crop's life cycle assuming no water stress. The water requirement of a crop at any phenological stage can be calculated by multiplying the potential evapotranspiration (PET) by the crop coefficient associated with that phenological stage (equation 2.4). The other AMS output is water requirement satisfaction index (WRSI) both current and extended/forecast, calculated as the ratio of actual evapotranspiration (AET) to the crops water requirements (WR) - refer to Section 2.3.2; soil water excess (WEX) and soil water deficit (WDF) calculated from the overall water balance by comparing the total water inputs and outputs to the soil; and the actual evapotranspiration (AET) at various phenological stages. Actual evapotranspiration (which is the same as PET_c) is calculated by multiplying the Penman-Monteith calculated potential (reference) evapotranspiration by the crop coefficients at any phenological stage. Mathematically,

$$AET = K_c PET \ , \qquad (3.1)$$

where AET is the actual evapotranspiration, K_c is the crop coefficient and PET is the Penman-Monteith derived potential evapotranspiration.

As a tool for early warning, AgroMetShell (AMS) is used to monitor the growing season from onset of planting to harvest at dekadal time step. The outputs give an indication of areas in a province or country that are causes of concern in relation to the amounts of rainfall received, water deficits, water excess and others at various stages of crop development. Water balance calculation with AMS is done at station level. Then the results are averaged over larger administrative units like districts, provinces and RDPs or ADDs in Malawi to compare to crop yield data (Mukhala and Hoefsloot, 2004). The outputs are then regressed against historical yield data in order to derive yield functions (models) that may be used to estimate yields for early warning purposes. The outputs from AMS may be displayed both as SEDI images or ASCII tables. AMS is compatible with other software including WinDisp, Microsoft Excel and Instat.

3.3.5 AMS methodology

This involved running the AgroMetShell (AMS) model in order to come up with potential maize yield predictors that were later regressed with historical maize yield. Outlined below are the commands that were used for various tasks.

(a) Creation of station list file

Database→ Manage station list… (specified a new name for the list, for example, Malawi). All the data were imported into this file.

(b) Data importation

Data that imported into AMS was in spreadsheet ASCII format. The files must contain a column that can be matched with a corresponding AMS station list (prepared as described above) by either station name or code. The most common format is an Excel comma separated value (CSV) file. The basic data for each station include its name (and code), longitude, latitude and altitude. Data was imported for the whole year (dekads 1 to 36). In the Southern African region, the agricultural season overlaps two consecutive calendar years; therefore data for two consecutive years were required for one season's model run. The data importation procedure was as follows:

Database → Import → From ASCII file…

(c) Preparation of a water balance run

This was done for each season for which data was available in AMS (1983/84 to 2001/02). Two sets of monitoring runs were created, one for the hybrid maize variety and the other one for the local maize variety. The following steps were followed:

Water Balance→ Monitoring run→ New… Save

(d) Determination of planting dekad

In the absence of actual information, the model determines the SOS using onset of rains (planting dekads) based on a simple precipitation accounting. Planting dekads for each season and station were determined using a threshold amount and distribution of rainfall received in three consecutive dekads. In this study, the criterion used was that at least 25 mm of rainfall was received in one dekad followed by a total of at least 20 mm of rainfall in the next two dekads. The end of the season was determined by the length of the crop cycle (maxima of 12 and 14 dekads for the hybrid and local varieties, respectively). The procedure followed is given below:

Water Balance → Monitoring run → Calculate planting Dekads → Based on rainfall threshold (specify the thresholds)*…→ Save and run…*

(e) Running the water balance (monitoring runs) for each season

The following steps were followed to run a water balance:

Water Balance→ Monitoring run (specify monitoring run) *→Run …OK.*

The programme calculates each of output parameters at decadal time scale and at station level. Then the parameters are integrated for the different phenological stages. These outputs were automatically stored in a GIS compatible Satellite Enhanced Data Interpolation (SEDI) viewer summary file within AMS, ready for extraction.

(f) Extraction of statistics from AMS

This was done using two Excel macros written by T.T. Magadzire of the SADC RRSU (then based in Harare). The first macro takes the summary files produced by AMS (at station level) and interpolates them to produce grids in WinDisp, then extracts statistics at RDP level for each season. The files are saved in a temporary folder in readiness for the next step. The second macro post processes the outputs from the first. It involves rearranging the outputs according to RDPs and in a chronological order. A yield file is input into the temporary folder. The output from this macro is a combination of the AMS outputs and historical yield data at RDP level and is of the form shown in table 3.3. Such files (tables) were imported into **Instat plus** statistical programme for data analysis.

Table 3.4: Sample AMS output parameters for Salima RDP.

RDP	TWRq	WSIs	WSIn	WSIc	WEXi	WEXv	WEXf	WEXr	WEXt	WDFi	WDFv	WDFf	WDFr	WDFt	AETi	AETv	AETf	AETr	AETt	Yield	Season
Salima	323	93	99	93	79	130	148	41	410	0	0	-6	-9	-15	12	73	183	41	311	4000	1983/84
Salima	340	96	99	96	49	112	115	43	317	0	0	-6	-2	-8	13	79	196	46	334	2200	1984/85
Salima	327	96	99	96	16	65	267	69	423	0	0	-4	0	-3	12	75	187	51	325	3000	1987/88
Salima	333	92	98	92	33	102	378	110	621	0	-1	-7	0	-8	12	76	188	49	325	2600	1988/89
Salima	338	93	99	93	56	140	207	6	410	0	0	-4	-8	-12	13	79	193	40	325	2300	1989/90
Salima	414	94	99	94	16	158	120	12	304	-1	-7	-4	-2	-14	14	87	240	59	401	2500	1990/91

In table 3.4, TWR_q is the total water requirement for the crop throughout the growing season without experiencing stress, WSI_s, WSI_n and WSI_c are the end of growing season, normal (calculated using long-term averages of climatic data) and the current water requirement satisfaction indices, respectively; WEX_i, WEX_v, WEX_f, and WEX_r is the excess water (mm) at the initial (establishment), vegetative, flowering and ripening stages while WEX_t is the total water excess (mm) for all the stages; WDF_i, WDF_v, WDF_f, and WDF_r is the water deficit (mm) at the initial (establishment), vegetative, flowering and ripening stages while WDF_t is the total water deficit (mm) for all the stages; and finally, AET_i, AET_v, AET_f, and AET_r is the actual crop evapotranspiration (mm) at the initial (establishment), vegetative, flowering and ripening stages while AET_t is the total crop evapotranspiration (mm) for all the stages during the growing season.

3.4 Statistical Analyses

3.4.1 Correlation analysis

Simple correlation analysis is used to determine the degree of relationship between any pair of variables. The following equation is used to calculate the correlation coefficient between the two variables x and y

$$r_{xy} = \frac{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right)}{\left[\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2\right]^{\frac{1}{2}}} \qquad (3.2)$$

Where x_i and y_i are independent variables, $\bar{x}$ and $\bar{y}$ are the means of x and y respectively and n is the length of data record. The value of r_{xy} lies between –1 and +1. The negative value means the variables are inversely related while the positive value means that they are directly related. If $r_{xy} = 0$ it means that the variables are not related.

Correlation matrices were constructed in MS Excel that showed relationships between each parameter with all the others, the parameters are shown in table 3.3. This process was done for each RDP. Due to the thrust of this study, of much interest was the relationship between each of the parameters with historical maize yield. This helped to identify the potential yield predictors at the RDP level. Furthermore, the magnitude of the correlation coefficient between the parameters themselves helped in later selection of parameters, especially when it came to inclusion of various parameters in multiple regression analysis. If parameters were highly correlated to each other, and both qualified to be potential predictors, one of them would be used, depending on its contribution to the regression model that was assessed by the magnitude of the coefficient of determination (r^2) and the F-test values.

3.4.2 Regression analysis

Regression is a statistical method that is used to investigate the relationship between two or more variables in terms of the percentage of variation on the dependent variable that is being explained by the variation in explanatory variable(s). In this study step-wise regression was done using statistical software called Instat plus version 3.20. Each of the parameters that were highly correlated with historical yield data in section 3.4.1 above was regressed and the values of the coefficient of determination (r^2), the F-test values and the standard error analysis were used to determine the predicting capabilities of the variable. Since the regression had to be done for every RDP, the degree of correlation with maize yield for each RDP considered was the highest among all the possible parameters. However, in this study coefficients of determination (r^2) of at least 0.40 were considered moderate and / or high at the 95% significance level. Then each of the other variables with higher correlations would be added and the improvements on the r^2 and the F-values were assessed and decisions were made taking into consideration the physiology of the crop and agronomical practices of the region.

3.4.3 Zonal regression equations

Since there was no consistence in the predictors giving higher correlations among the RDPs, the RDPs were grouped according to the climatological zone in which each is found (Section 3.1.5, Table 3.1). Each combination of predictors for each RDP was tested with the other RDPs in that particular zone. The combination that gave relatively higher values of the r^2 in each case was selected for that zone and a regression equation/model was developed for each RDP using that particular combination of predictors.

CHAPTER 4

RESULTS AND DISCUSSION

4.1 Spatial and Temporal Patterns

The cumulative seasonal rainfall for selected seasons (wet and dry) in Malawi is shown in figure 4.1. Figure 4.2 shows the spatial rainfall pattern during wet and dry seasons.

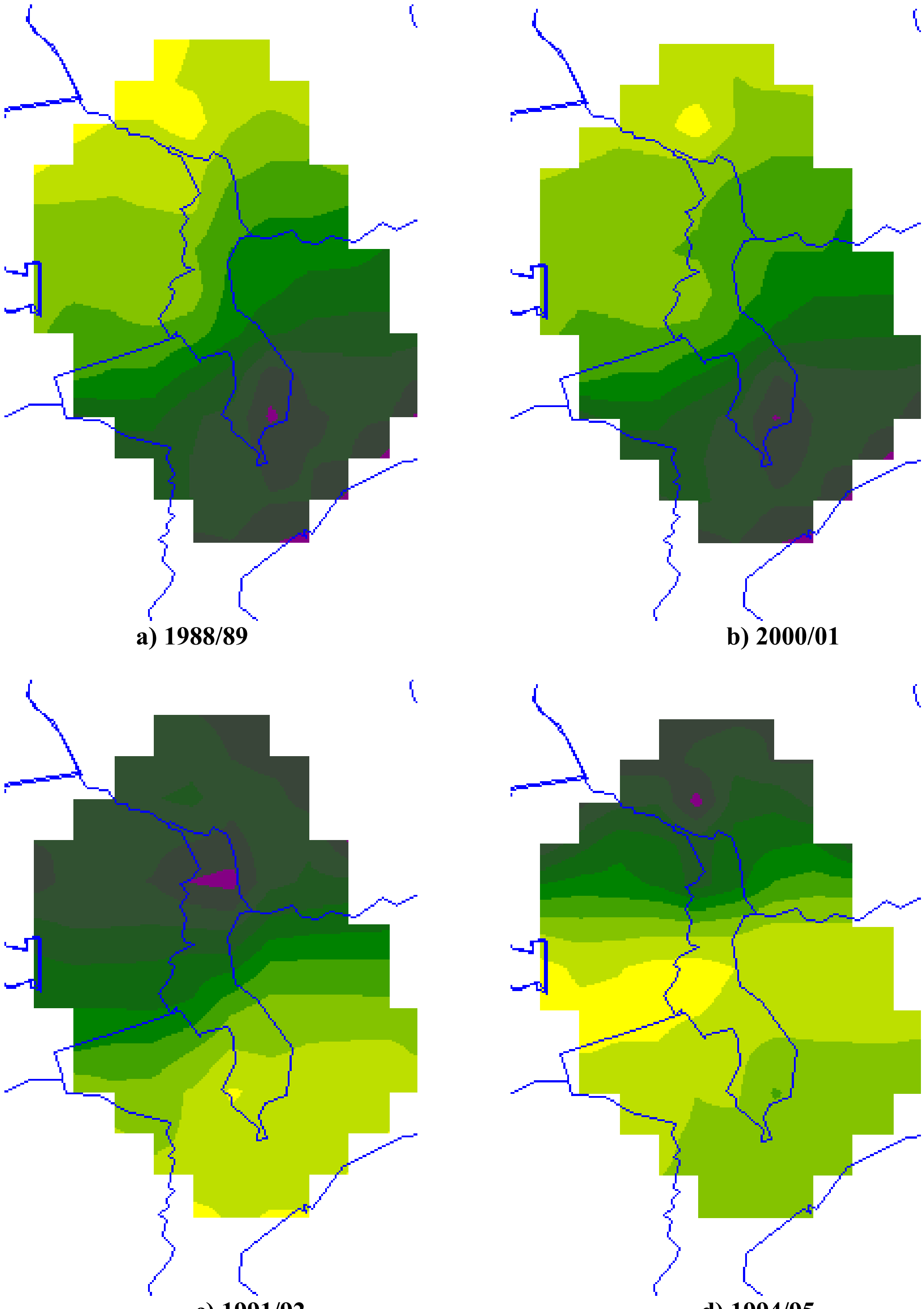


Figure 4.1: Cumulative seasonal rainfall for different seasons across Malawi (the darker the shades the higher the seasonal rainfall).

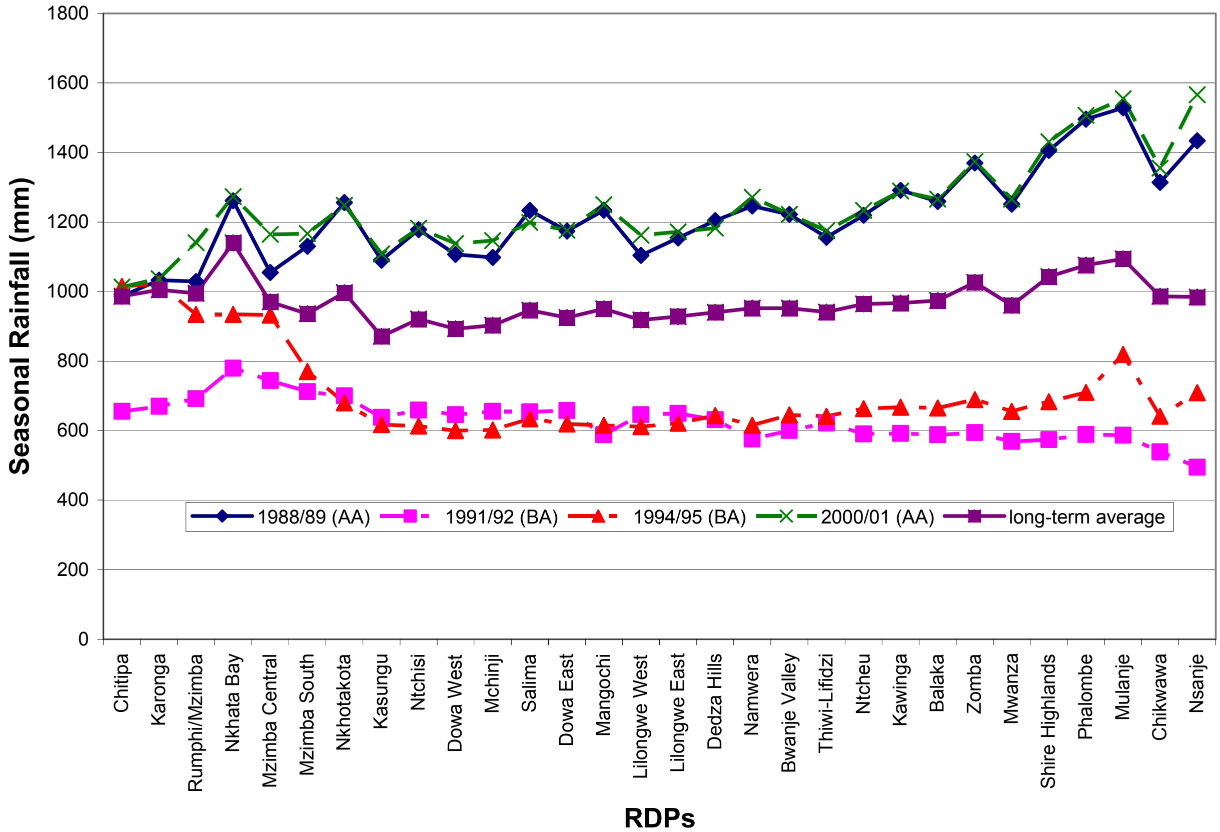
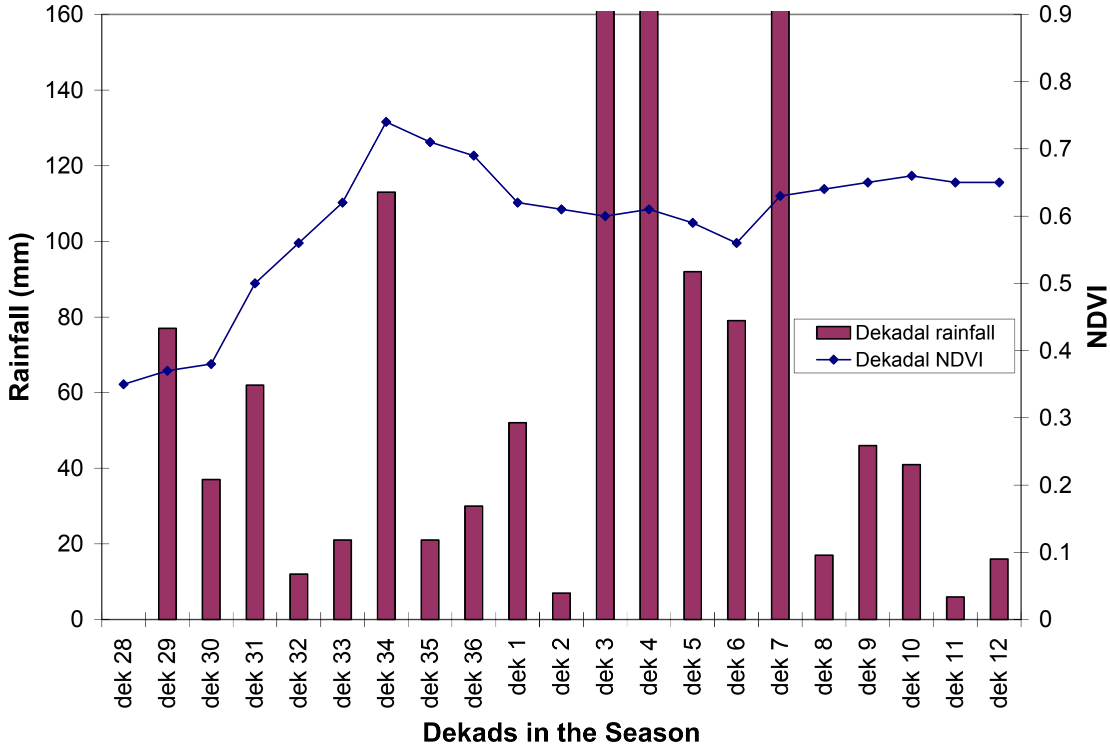


Figure 4.2: Spatial rainfall pattern during wet (above average (AA)), dry (below average (BA)) seasons and the long-term average (1971-2000) across Malawi (from left to right the RDPs are in North to South order).

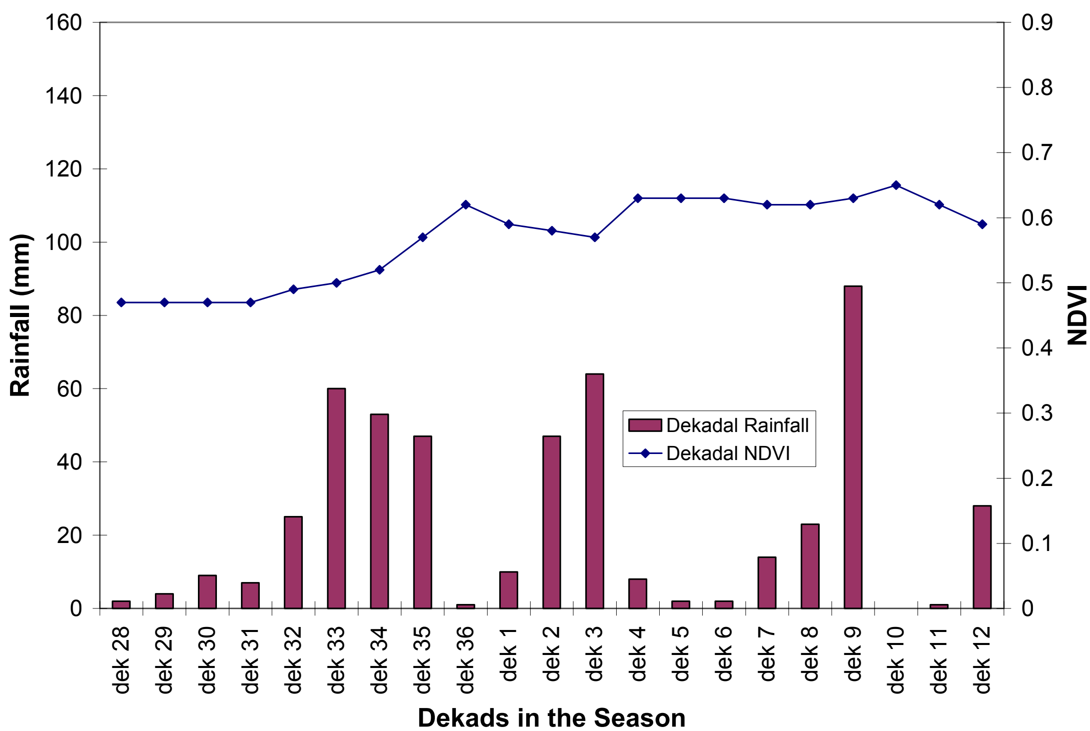
The rainfall patterns between selected wet (1988/89 and 2000/01) and dry (1991/92 and 1994/95) seasons across the country show that during wet seasons the cumulative rainfall amounts increase from north to south, ranging from about 1000 mm in the north to about 1600 mm in the southern tip (Figure 4.1a,b and Figure 4.2). On the other hand, during the selected dry seasons, cumulative amounts decrease from north to south of the country with the northern tip receiving between 600 and 1000 mm and the southern areas receiving between 500 and 625 mm annually (Figure 4.1c,d and Figure 4.2). Effects of ENSO episodes in Malawi may help to explain these different patterns. During a warm ENSO phase (El Niño such as the 1991/92 and the 1994/95 seasons), the Southern African region generally experiences drought while the Eastern African region experiences wet conditions. Since Malawi is located in the transition zone between these two regions, some parts of the northern tip of the country benefit from the Eastern African wet conditions hence the relatively high amounts during dry seasons. The southern parts of the country are affected by the drier conditions in the southern African region during El Niño seasons hence the lower rainfall amounts. During cold ENSO phases (La Niña such as the 1988/89), the conditions are reversed between Southern Africa and Eastern Africa. Therefore the Northern parts receive relatively low rainfall amounts due to drier conditions in the Eastern African region while the southern areas benefit from the much wetter conditions in the Southern Africa (http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/eln/def.rxml).

On the intra-seasonal patterns for rainfall and NDVI, the same extreme seasons were compared in various selected RDPs across the country, chosen on the basis of the climatological zones they belong to. These RDPs are namely: Chitipa (North), Kasungu (Central), Salima (Lakeshore), Mwanza (Southern) and Nsanje (Shire Valley). It was observed that apart from the amounts received during respective seasons in these RDPs, generally the wet seasons had earlier onsets and the seasons progressed without significant dry spells (see figures 4.3a; and Appendix I). As for the drier seasons, the onsets of the rains were delayed and cessation was relatively earlier. Dry spells were frequent and prolonged (see figures 4.3b; and Appendix I). These resulted in the growing seasons being short. As a result, the performance of crops was negatively affected during the drier growing seasons due to shortages of water to meet plant requirements. This was evident from the maize yield record, especially for the worst drought season in the series (1991/92), in which the lowest yield figures for almost all RDPs and for both varieties was registered. In some cases almost total crop failure was reported for that particular season. For example, in Balaka RDP, yields of 24 kg/ha and 2 kg/ha for hybrid and local maize varieties respectively while in Mangochi RDP yield of 29 kg/ha and 12 kg/ha (for hybrid and local maize, respectively) were registered during the 1991/92 season.

Seasonal NDVI profiles for wet and dry seasons were also explored for selected RDPs and the results are shown in Figure 4.3 for Nsanje RDP. For the other selected RDPs the results are given in Appendix I.



a) 1988/89

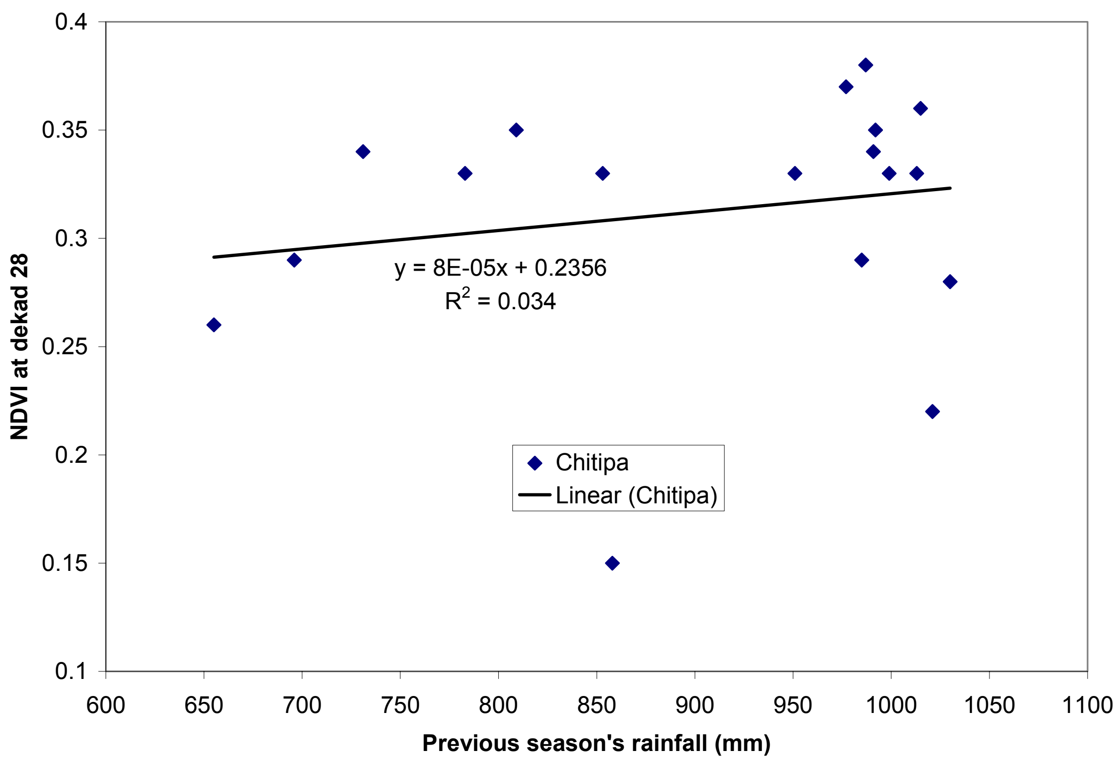


b) 1991/92

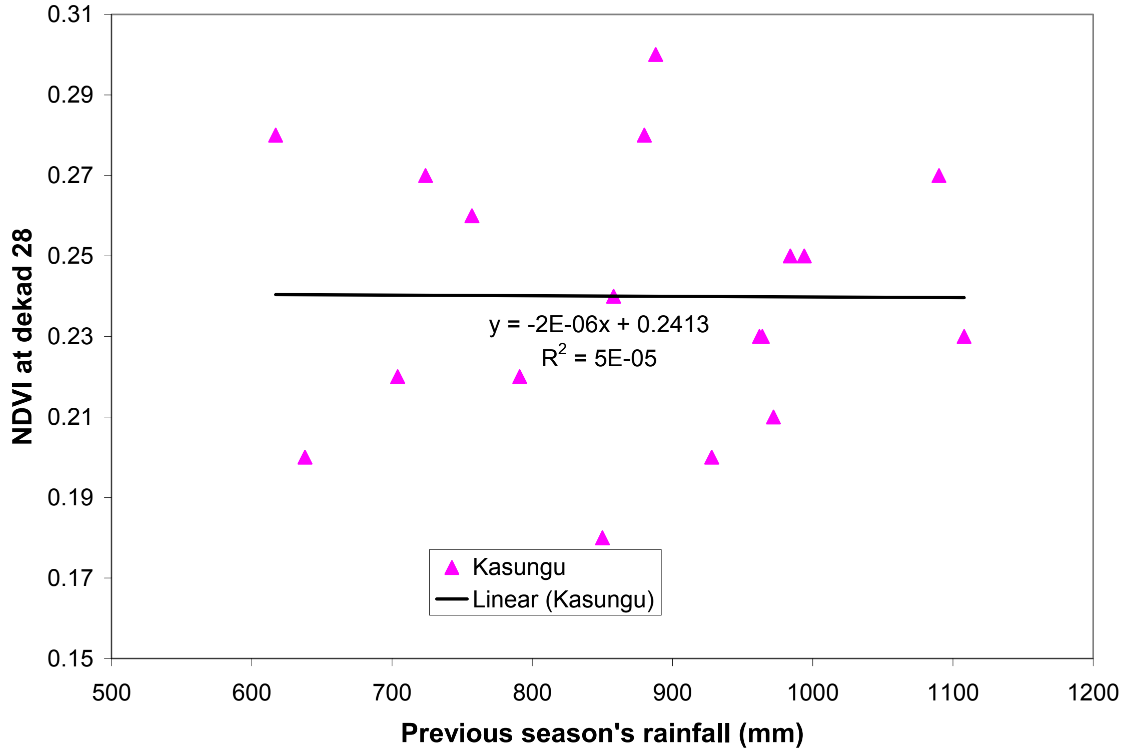
Figure 4.3: Rainfall and NDVI Series for Nsanje RDP for 1988/89 (above average (AA)), (and 1991/92 below average (BA)) rainfall seasons.

As regards seasonal NDVI profiles for wet and dry seasons, it was observed that generally the profiles for wet seasons had a steeper ascent up to the dekad of maximum NDVI. For drier season, though the maxima attained were not significantly different from those of the wet seasons, the ascent was gentle for the sample RDPs. When seasonal maximum NDVIs were inspected for the seasonal variability, no significant variations were identified. For example, the NDVI values for 1988/89 and 1991/92 seasons respectively were: Chitipa (0.62, 0.68), Salima (0.58, 0.60), Kasungu (0.56, 0.60), Mwanza (0.77, 0.76) and Nsanje (0.75, 0.68) (see figures 4.3a and 4.3b, and Appendix I).

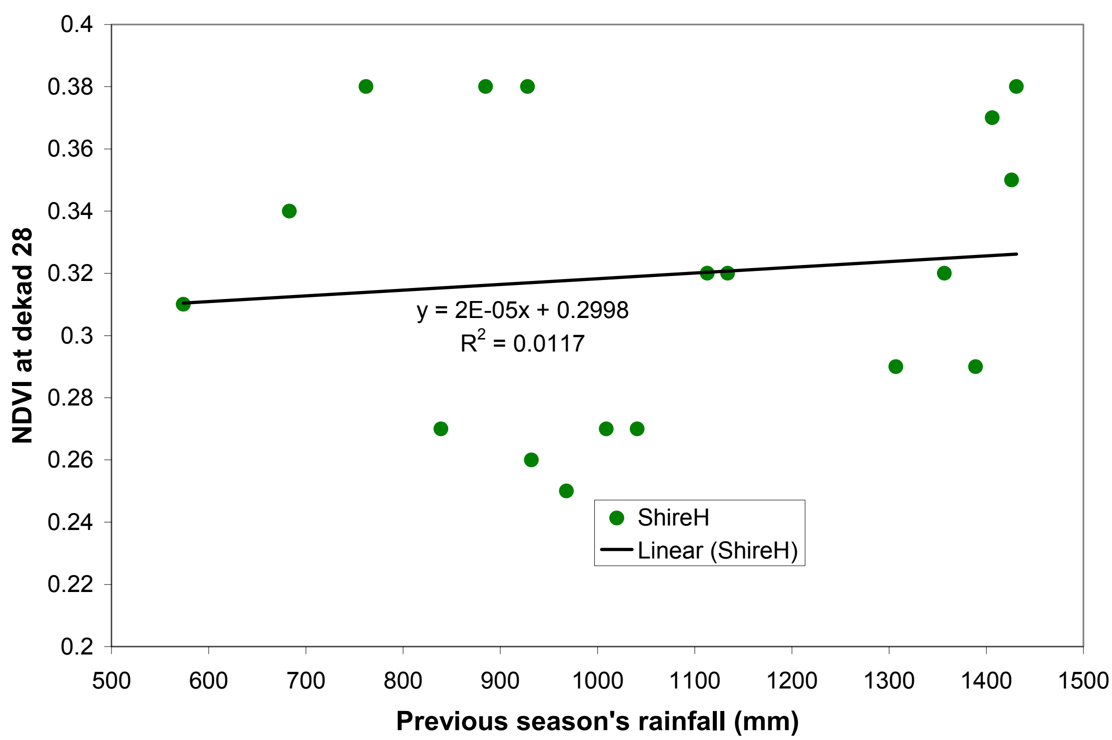
Another feature that was explored was whether the NDVI value at the start of the season (defined using the rainfall thresholds of at least 25 mm in one dekad and a total of at least 20 mm in the next consecutive two dekads, as used in the AgroMetShell (AMS) model evaluation) was related to the rainfall pattern for the previous season. This feature was explored because the NDVI starting value was necessary in the determination of the NDVI increments that were explored and are described in later sections. The results for the relationship between NDVI starting value for the season and the previous season's rainfall are given in figures 4.4a to 4.4d.



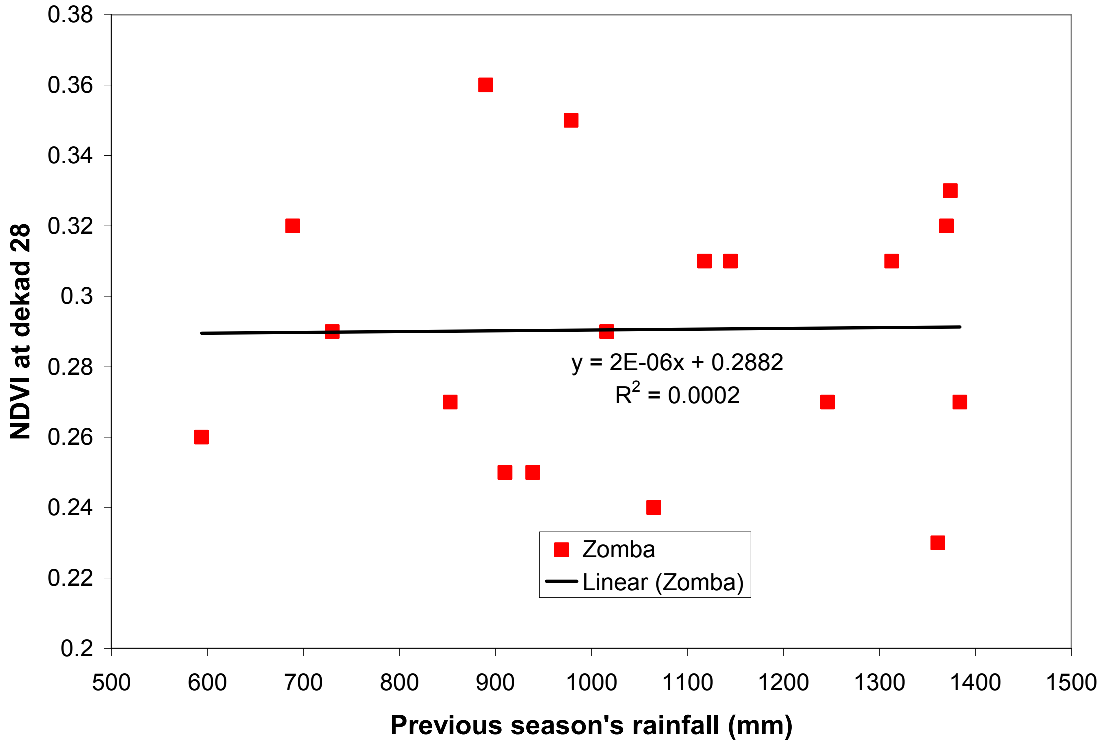
(a) Chitipa RDP



(b) Kasungu RDP



(c) Shire Highlands RDP

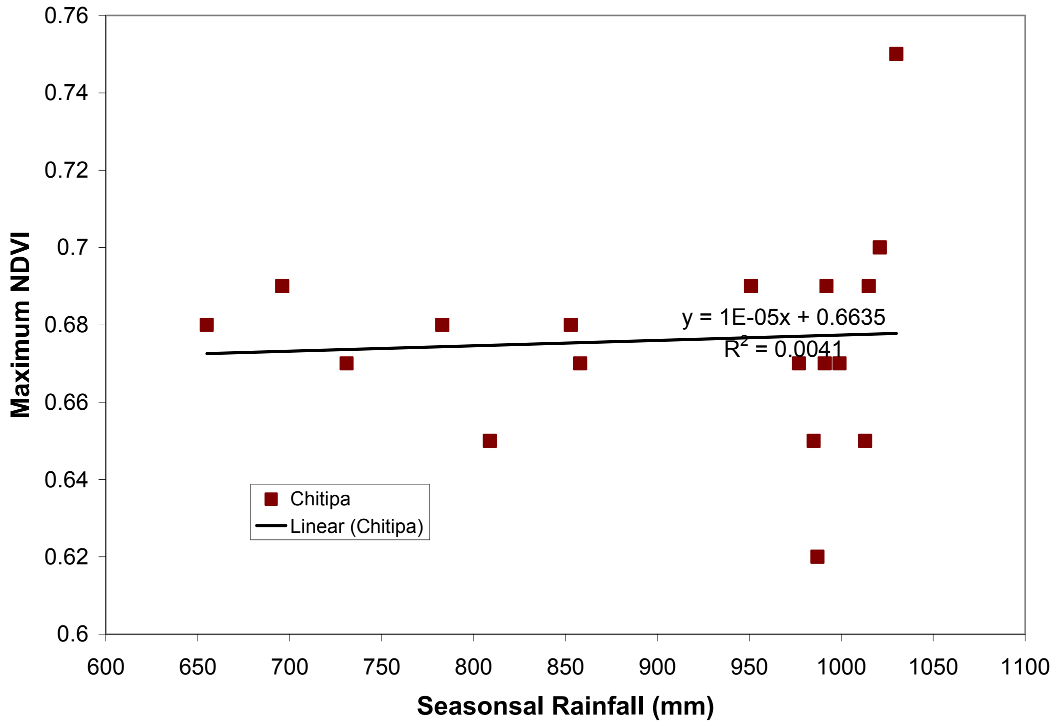


(d) Zomba RDP

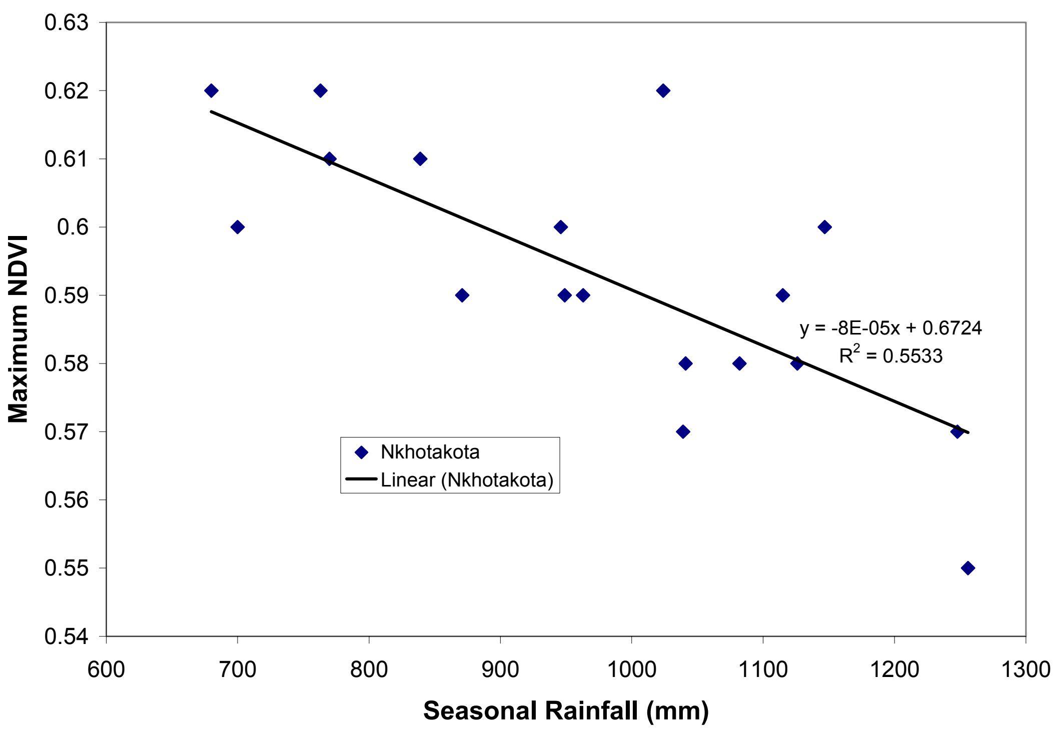
Figure 4.4: Scatter plots for NDVI at the start of the season (dekad 28) and rainfall for the previous season.

From the values of the coefficients of determination, the slope of the lines of best fit through the data points for all the selected RDPs as shown in figure 4.4, the two parameters are not directly related. Correlation analysis was also performed between the NDVI value at the start of the season and previous season's cumulative rainfall amounts. It was found that the correlation coefficients (r) were very weak and negative for most of the sampled RDPs. For example, Chitipa, Kasungu, Zomba and Shire Highlands gave the following values of r for the aforementioned relationship: 0.17, -0.03, -0.02 and 0.07, respectively. These values of r and the graphs in figure 4.4 show lack of significant relationship between the two parameters in question. Therefore, the current season's initial NDVI values cannot be attributed to the rainfall performance of the previous season. However, residual soil moisture was expected to have bearing on the season's initial value. Residual moisture is a function of previous season's cumulative rainfall and when the season ended. It can be argued that the performance of the winter season might have an impact on this relationship. Hence an analysis of the performance of the winter season can help to understand this relationship.

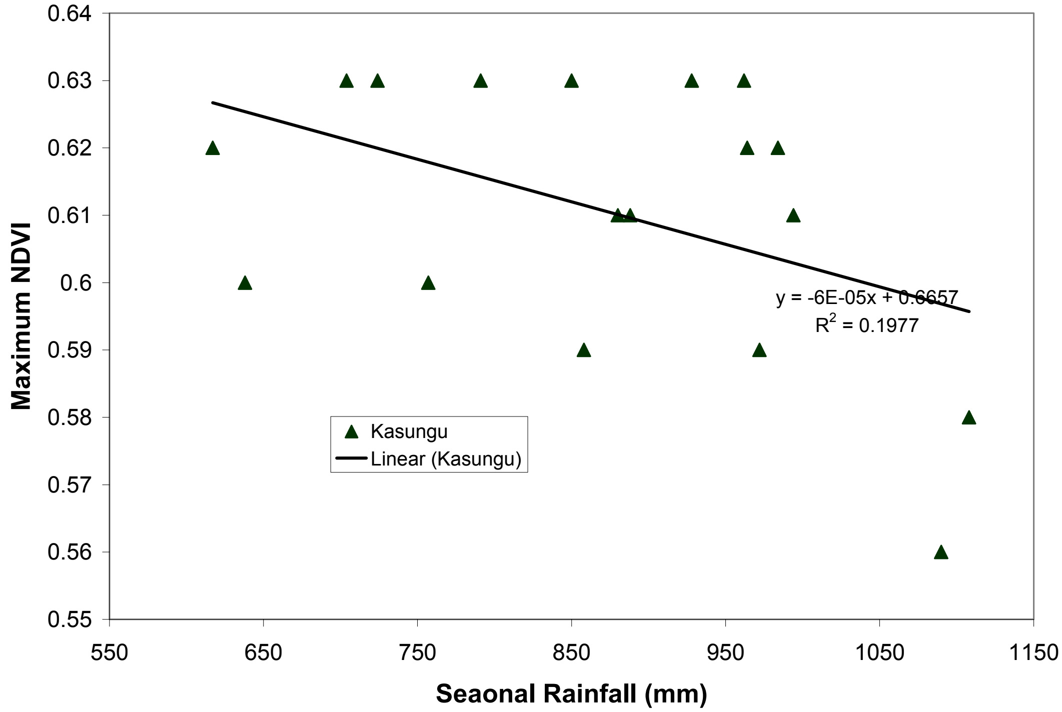
The relationship between maximum NDVI and seasonal rainfall was also investigated. The results are shown in figures 4.5a to 4.5e.



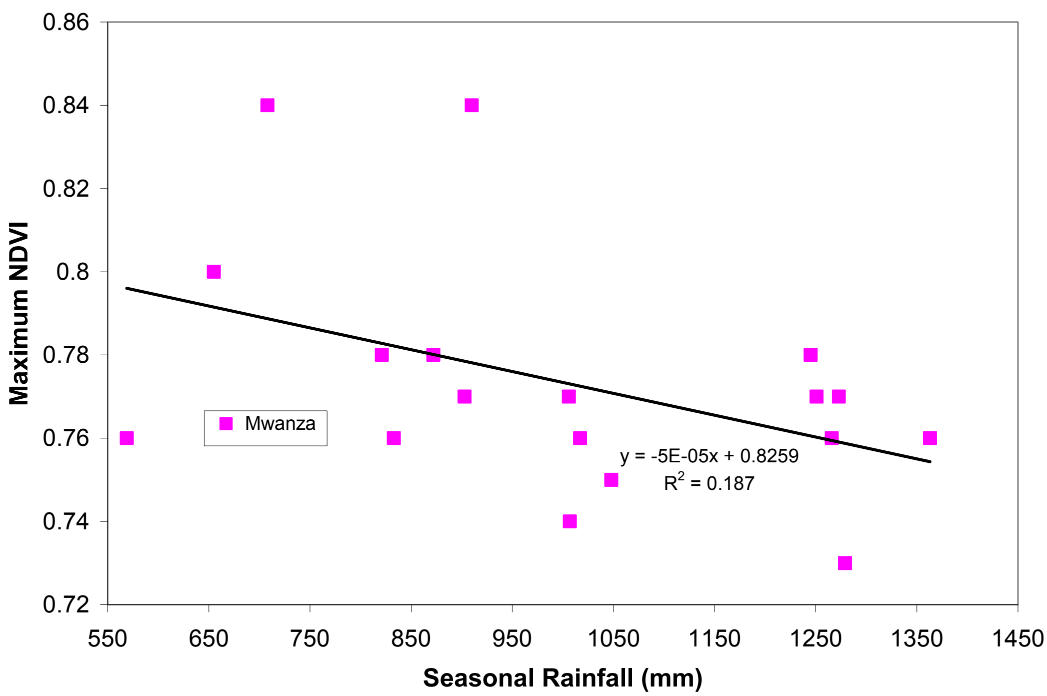
(a) Chitipa RDP



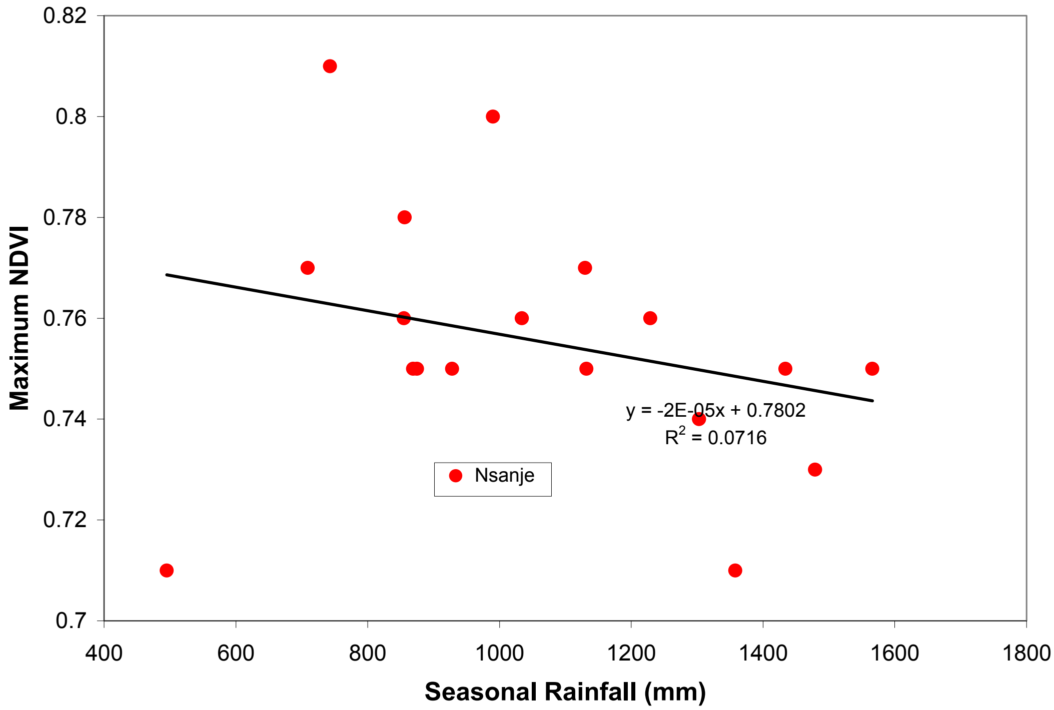
(b) Nkhota kota RDP



(c) Kasungu RDP



(d) Mwanza RDP



(e) Nsanje RDP

Figure 4.5: Scatter plots for the relationship between maximum NDVI and seasonal rainfall for Chitipa, Nkhota kota, Kasungu, Mwanza and Nsanje RDPs.

It was found that the seasonal maximum NDVI were in general inversely (negatively) related to the seasonal rainfall total mainly for the central parts of the country. However, in the northern parts of the country, the majority (four out of six) of RDPs gave positive correlations between seasonal rainfall and maximum NDVI (Figure 4.6). Though weak, figure 4.5a illustrates the positive correlations in Chitipa RDP, in the northern region of Malawi. Figure 4.5b clearly shows the inverse relationship in Nkhota kota, while figure 4.5c gives the scatter plot for Kasungu RDP; the relationship for Mwanza RDP is shown in figure 4.5d, and figure 4.5e gives the same relationship for Nsanje RDP. In the northern region, the two parameters are weakly but positively related to each other. For the rest of the country these parameters were negatively related to each other. This observation was confirmed when correlation analysis was performed for the two variables. Countrywide, stronger but negative correlations were obtained in the following RDPs: Nkhota kota, Kasungu, Ntchisi, Salima, Namwera, Bwanje Valley, Thiwi-Lifidzi, Ntcheu, Kawinga, Mwanza, Shire Highlands (r=-0.74, -0.44, -0.45, -0.53, -0.51, -0.44, -0.44, -0.44, -0.48, -0.43, and –0.51, respectively). Figure 4.6 shows the pattern (distribution) of the correlation coefficients (r) between maximum NDVI and seasonal rainfall in Malawi. From figure 4.6, spatially the correlations were negative across the country expect for a few RDPs in the northern parts of Malawi. Maximum NDVI is a function of the type of dominant vegetation in an area that does not vary from season to season unlike rainfall that shows significant interseasonal variation. This helps to explain the lack of better or the negative relationships observed.

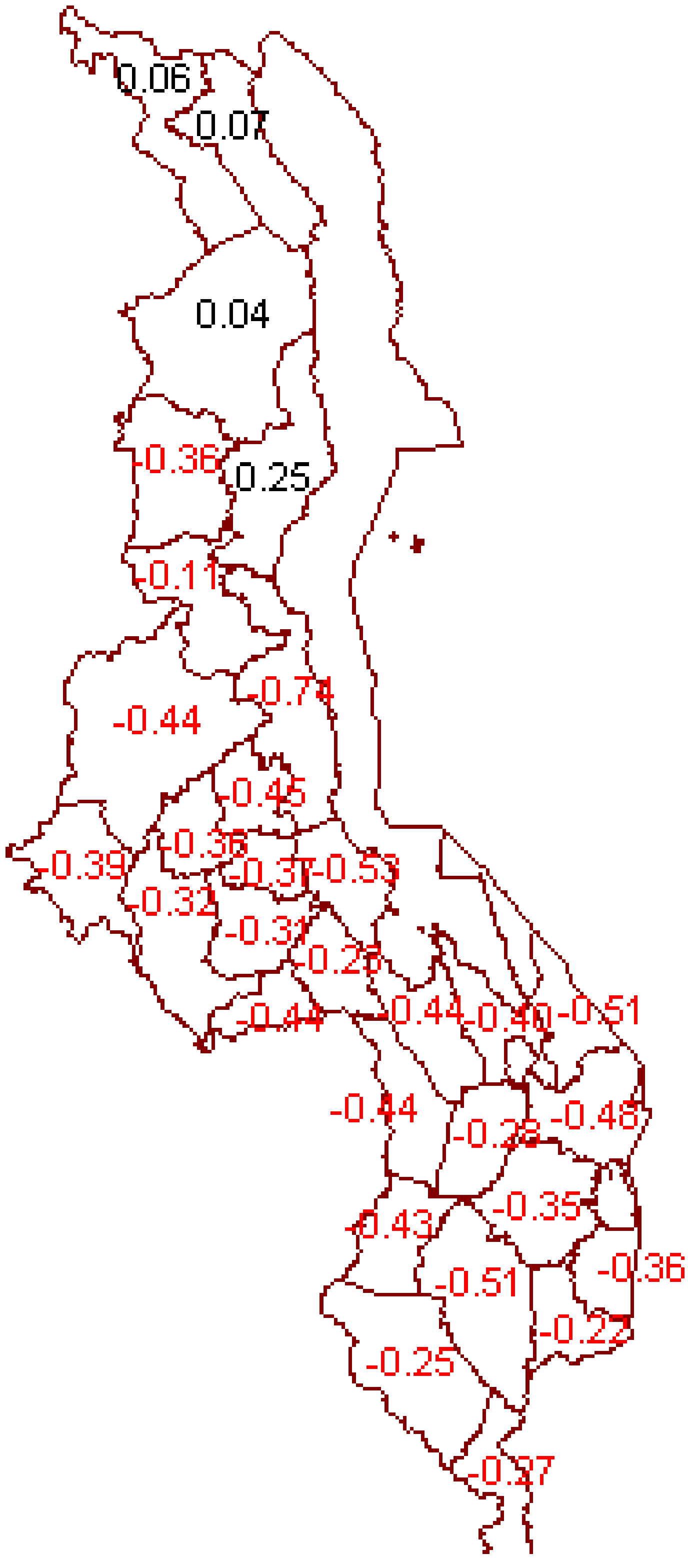


Figure 4.6: Map of Malawi showing the distribution of the correlation coefficients (r) between seasonal rainfall and maximum NDVI. The red labels represent RDPs with r < 0.40, while the black labels represent RDPs with r > 0.40. For the identification of the RDPs refer to figure 3.2.

4.2 Correlation Analysis between NDVI Parameters and Rainfall

4.2.1 Cumulative NDVI and seasonal rainfall

Correlation coefficients between NDVI cumulated from the first dekad of October (dekad 28) to various stages of the growing season and seasonal rainfall are shown in table 4.1.

Table 4.1: Correlation coefficients between cumulative NDVI (cumulated from dekad 28) and seasonal rainfall. The shaded values show moderate correlations (r > 0.40) between the two parameters (dekads are numbered from 1 January).

RDP	Dekads in which highest correlations were obtained	Highest correlation coefficients (r) and dekad attained	Comment
Chitipa	28 to 30	-0.19 (28)	Early season
Karonga	28 to 30	-0.25 (28)	Early season
Rumphi/Mzimba	28 to 30	-0.26 (28)	Early season
Nkhata Bay	28 to 31	-0.45 (29)	Early season
Mzimba Central	9 to 12	-0.33 (12)	Late season
Mzimba South	5 to 12	-0.53 (9)	Mid to late season
Nkhota kota	28 to 32	-0.52 (28)	Early season
Kasungu	28 to 32	-0.57 (28)	Early season
Ntchisi	28 to 32	-0.51 (29)	Early season
Dowa West	28 to 32	-0.66 (31)	Early season
Mchinji	28 to 32	-0.38 (31)	Early season
Salima	28 to 29	-0.55 (28)	Early season
Dowa East	34 to 35	0.45 (34)	Early season
Mangochi	28 to 32	-0.68 (28)	Early season
Lilongwe West	28 to 32	-0.58 (29)	Early season
Lilongwe East	28 to 32	-0.51 (31)	Early season
Dedza Hills	28	-0.42 (28)	Early season
Namwera	32 to 12	-0.60 (10)	Throughout season
Bwanje Valley	28 to 4	-0.50 (28)	Early to mid season
Thiwi-Lifidzi	28 to 4	-0.46 (29)	Early to mid season
Ntcheu	28 to 30	-0.51 (28)	Early season
Kawinga	28 to 33 6 to 12	-0.46 (33) -0.49 (7)	Early season Mid to late season
Balaka	28 to 33	-0.71 (29)	Early season
Zomba	28 to 32	-0.47 (28)	Early season
Mwanza	28 to 33 4 to 12	-0.69 (28) -0.53 (9)	Early season Mid to late season
Shire Highlands	31 to 32	0.43 (32)	Early season
Phalombe	32 to 3	0.27 (33)	Early to mid season
Mulanje	28 to 29	-0.56 (28)	Early season
Chikwawa	28 to 33	-0.71 (28)	Early season
Nsanje	28 to 30	-0.61 (28)	Early season

For the correlation between NDVI cumulated from the start of the season to different stages of the season, moderate to strong relationships were observed mainly between cumulative NDVI in the first six dekads of the season and cumulative seasonal rainfall, except for a few RDPs (Mzimba Central, Mzimba South, Namwera, Kawinga and Mwanza) in which the moderate relationships were realized towards the end of the season. In most of the RDPs, as shown in table 4.1, negative correlations between cumulative NDVI and seasonal rainfall were realized. NDVI cumulated from the first dekad of October up to around mid-season (January to February) showed that the correlations were almost zero for most RDPs, indicating no significant relationship between cumulative NDVI (up to that particular stage) and seasonal rainfall. The RDPs that gave moderate relationships ($r>0.40$) are shaded in table 4.1. The negative correlation between cumulative NDVI and seasonal rainfall can be explained by the fact that vegetation responds to a rainfall event after some time lag of between two to four dekads (Wan *et al.,* 2004). Since seasonal rainfall represents the overall rainfall effects for the whole season, it is hard to isolate its relationship with cumulative NDVI. The negative correlations, which occur mainly at the beginning of the season, can be explained by the fact that during early stages the vegetation is not well established and its development is not mainly in response to rainfall for the same season but might be due to residual moisture in the soil and other factors affecting its development. As for the stronger relationships that occurred towards the end of the season, it was found that these occurred in only four RDPs (Mzimba South, Namwera, Kawinga and Mwanza). On the other hand, only three RDPs, namely Dowa East, Shire Highlands and Phalombe, showed relatively high positive correlations between cumulative NDVI and seasonal rainfall, and these occurred at the start of the season. These RDPs with unique relationships were found to be areas of high seasonal rainfall. The weak and negative relationships between NDVI parameters and seasonal rainfall may be attributed to the rainfall patterns in the respective RDPs. In high rainfall areas NDVI remains high throughout the growing season hence minimal seasonal variations but rainfall shows higher seasonal variations. This explains the unique relationships between cumulative NDVI and seasonal rainfall in these higher rainfall RDPs (Mkhabela *et al.,* 2005).

4.2.2 Seasonal NDVI averages and seasonal rainfall

Various NDVI averages during the growing season namely October to April, November to January, November to February and December to February averages, which were also used by Mutikani (1997), were correlated with maize yield and the results are shown in table 4.2.

Table 4.2: Correlation coefficients between NDVI averages and total seasonal rainfall (the shaded values indicate moderate correlations {r>0.40} between the parameters).

RDP	Correlation coefficients (r) for October-April averages	Correlation coefficients (r) for November-January averages	Correlation coefficients (r) for November-February averages	Correlation coefficients (r) for December-February averages
Chitipa	-0.01	0.02	0.16	0.09
Karonga	0.07	-0.01	0.10	0.08
Rumphi/Mzimba	-0.08	0.10	0.05	-0.07
Nkhata Bay	-0.15	0.01	0.04	0.10
Mzimba Central	-0.33	0.07	-0.12	-0.19
Mzimba South	-0.49	-0.25	-0.37	-0.25
Nkhota kota	-0.10	0.33	0.23	0.25
Kasungu	-0.18	0.21	0.07	0.07
Ntchisi	0.10	0.39	0.28	0.31
Dowa West	-0.03	0.27	0.13	0.14
Mchinji	-0.29	0.11	-0.11	-0.24
Salima	-0.06	0.28	0.14	0.10
Dowa East	0.30	0.43	0.30	0.35
Mangochi	0.01	0.24	0.06	0.05
Lilongwe West	0.04	0.29	0.15	0.09
Lilongwe East	0.17	0.32	0.17	0.18
Dedza Hills	0.04	0.36	0.13	0.10
Namwera	-0.56	-0.31	-0.44	-0.47
Bwanje Valley	0.10	0.20	0.06	0.07
Thiwi-Lifidzi	-0.03	0.10	-0.08	-0.18
Ntchisi	0.03	0.09	-0.04	-0.03
Kawinga	-0.48	-0.16	-0.29	-0.28
Balaka	0.10	0.14	0.02	0.14
Zomba	-0.25	0.12	-0.10	-0.08
Mwanza	-0.44	-0.21	-0.32	-0.32
Shire Highlands	-0.44	0.09	-0.15	-0.13
Phalombe	-0.32	0.17	-0.04	-0.01
Mulanje	-0.28	0.08	-0.09	-0.15
Chikwawa	0.03	0.20	0.05	0.19
Nsanje	-0.12	0.21	0.02	-0.13

Seasonal NDVI, averaged from October to April, was found to be weakly and mostly negatively correlated to cumulative seasonal rainfall. A few RDPs that showed moderate (though negative) correlations include Mzimba South, Namwera, Kawinga, Mwanza and Shire Highlands (r = -0.49, -0.56, -0.48, -0.44 and -0.44, respectively). When the NDVI was averaged between November and January, and then correlated with seasonal rainfall, no significant relationships were observed. However, the correlation coefficients for the majority of the stations were positive except for Karonga, Mzimba South, Namwera, Kawinga, and Mwanza where the correlation coefficients were negative.

These results are shown in table 4.2. This relationship can further be explored since it is the only one that gave positive relationships between the vegetation indices (NDVI) and seasonal rainfall, though generally weak. When NDVI is averaged from October to April, there is a high possibility of incorporating the effects of residual soil water since the onset of the rains is not always as early as October.

It was established when working with the AMS that for the majority of the RDPs seasonal rainfall starts from late November into December and in other cases in January. For the other averages, that is, from November to February and December to February, the correlations were generally very weak and negative in most of the RDPs. Higher correlation coefficients were obtained only for Namwera RDP (-0.44 and -0.47, respectively for the two sets of averages).

From the results of this section it has been observed that Mzimba South, Namwera, Kawinga and Mwanza have been consistent in giving higher negative correlation coefficients for cumulative NDVI and seasonal NDVI with seasonal rainfall. The same explanation as that for relationship between cumulative NDVI and seasonal rainfall, given in section 4.2.1, applies in the present case.

4.3 Statistical Analyses of NDVI and Maize Yield

4.3.1 Dekadal and monthly NDVI increments against yield

(a) Dekadal NDVI increments and yield

For the correlation between NDVI increments (for the calculation of the NDVI increments, refer to section 3.3.1) and maize yield, both negative and positive moderate correlations were obtained for both the local and hybrid varieties. The increments from the first to the second dekad of October (dekad 28 to dekad 29) and from the second to the third dekad of October (dekad 29 to dekad 30) were observed to give mainly positive and moderate ($r > 0.40$) correlations for the highest number of RDPs as compared to the other dekadal increments during the season. Also the increments from the third dekad of January (dekad 3) to the first dekad of February (dekad 4) gave moderate but negative correlations with final yield (see Appendix IIa). However, these could not be used to develop regression models because firstly, October NDVI is too early to decide the final yield (unless the output is to be used for assessing the area planted), and it was observed that in most of the seasons the start of the rains would occur from the second dekad of November (dekad 32) onwards. Therefore, there must be other factors explaining these moderate correlations early in the season. Secondly, the NDVI increments from dekad 3 to dekad 4 generally gave negative correlation with maize yield therefore could not be meaningfully used for developing regression models for the purpose of yield prediction.

Temporally, it was observed that with the progression of the growing season towards the end, especially in March and April, the correlation coefficients between the yield and dekadal NDVI increments generally become positive though weak. This can be explained by the fact that in Malawi the peak of the rainfall season is in January to February. Since NDVI responds to rainfall within two to four dekads after the event (Wan *et al.* 2004) the positive correlations come about because of the significant biomass accumulation of the vegetation in response to the January or February rainfall peaks.

(b) Monthly NDVI increments and yield

The RDPs with high correlation coefficients between monthly NDVI increments and maize yield are shown in table 4.3a for the hybrid variety and in table 4.3b for the local variety. All the results on the correlation between monthly NDVI increments and maize yield are given in Appendix IIb.

Table 4.3: Correlation coefficients between maize yield and monthly NDVI increments for selected RDPs with moderate correlations

RDP	**Correlation coefficient (r)**	**Month**
Salima	0.67	April
Lilongwe West	0.42	April
Bwanje Valley	0.54	April
Mwanza	0.47	April

(a) hybrid variety

RDP	**Correlation coefficient (r)**	**Month**
Mzimba Central	0.54	April
Mzimba South	0.43	April
Salima	0.43	April
Lilongwe East	0.42	April
Bwanje Valley	0.48	April
Thiwi-Lifidzi	0.52	April
Mwanza	0.45	April
Phalombe	0.42	April
Mulanje	0.63	April

(b) local variety

As the season progresses from start to end, the number of RDPs in which there were moderate relationships between monthly NDVI increments (derived from the difference between the NDVI for the third dekad of the current month and that of the third dekad of the previous month) and final maize yield generally increased. During the early months of the season, very few RDPs showed moderate relationships between the two parameters while in later parts of the season, the relationships become stronger (with more RDPs giving r > 0.40 and also the values of r for the same RDP were getting larger). For a few RDPs the increase from January to February showed moderate but negative correlations for both the hybrid and local maize. March to April increments showed the moderate and positive correlations with historical yield data only in four and nine out of thirty RDPs for hybrid and local varieties respectively (tables 4.3a and 4.3b). Hence no regression was attempted due to the small number of RDPs in which monthly NDVI increments gave higher correlations with yields. Tables 4.3a and 4.3b show the RDPs where the higher correlations were obtained for the increase from March to April both the hybrid and local varieties. These results are similar to those of Dubey *et al.* (1994) who found that the significant parameters for wheat yield estimation for Punjab, India were the NDVI at heading and flowering stages. Mkhabela et al. (2005) found that in Swaziland, the period between early January and late March, that coincide with the flowering and grain filling stages gave higher and significant (r^2 ranging between 0.51 and 0.68) correlations between NDVI and maize yield. Unganai and Kogan (1998a) found that maize yield in Zimbabwe correlated well (r^2 ranging from 0.72 to 0.93 for different districts) with vegetation condition index (VCI) during the flowering and grain formation stages. For maize in Malawi, due to a stretch in planting dates (among seasons and also among RDPs) this critical period may run from February to March or April. Perhaps this helps to explain improved correlations during the second half of the growing season especially in March and April.

4.3.2 Dekadal NDVI and yield

Similar to the case of NDVI increments above, the degree of association between dekadal NDVI and final maize yield increased in significance during the second half of the growing season (from January to April). During this period, the correlation coefficients (r) became stronger and positive in most of the RDPs. For each RDP the dekads, values of the correlation coefficients and coefficient of determination are shown in tables 4.4a and 4.4b.

Table 4.4a: Correlation coefficients for dekadal NDVI and hybrid maize yield. Shaded cells indicate RDPs where r > 0.40; dekads are counted from 1 January.

RDP	Consecutive dekads of highest correlation coefficients (r)	Correlation coefficient (r)	Confident of determination (r^2)
Chitipa	11, 12	-0.54, -0.46	0.29, 0.21
Karonga	5, 6	0.35, 0.29	0.12, 0.08
Rumphi/Mzimba	31, 32, 33	0.50, 0.53, 0.48	0.25, 0.28, 0.24
Nkhata Bay	5, 6	-0.36, -0.34	0.13, 0.12
Mzimba Central	8, 9	-0.47, -0.46	0.22, 0.21
Mzimba South	4, 5, 6	-0.37, 0.10, 0.39	0.14, 0.01, 0.15
Nkhota kota	5, 6, 7	-0.31, -0.37, -0.35	0.10, 0.14, 0.12
Kasungu	2, 3	0.27, 0.34	0.07, 0.12
Ntchisi	7, 8	-0.51, -0.46	0.26, 0.21
Dowa West	31, 32	0.31, 0.26	0.10, 0.07
Mchinji	2, 3	0.33, 0.35	0.11, 0.12
Salima	10, 11, 12	0.62, 0.70, 0.69	0.38, 0.49, 0.48
Dowa East	2, 3	0.11, 0.43	0.01, 0.18
Mangochi	11, 12	0.52, 0.58	0.27, 0.34
Lilongwe West	10, 11	0.47, 0.46	0.22, 0.21
Lilongwe East	10, 11, 12	0.51, 0.52, 0.43	0.26, 0.27, 0.18
Dedza Hills	2, 3	0.26, 0.25	0.07, 0.06
Namwera	1, 2	-0.45, 0.40	0.20, 0.16
Bwanje Valley	10, 11, 12	0.70, 0.73, 0.65	0.49, 0.53, 0.42
Thiwi-Lifidzi	2, 3	0.43	0.18, 0.11
Ntcheu	10, 11	0.65, 0.56	0.42, 0.31
Kawinga	1	-0.43	0.12, 0.18, 0.12
Balaka	11, 12	0.58, 0.64	0.34, 0.41
Zomba	7, 8	0.41, 0.45	0.17, 0.20
Mwanza	11, 12	0.46, 0.63	0.21, 0.40
Shire Highlands	2, 3	0.40, 0.41	0.09, 0.16, 0.17
Phalombe	1, 2, 3	0.42, 0.53, 0.43	0.18, 0.28, 0.18
Mulanje	1	0.40	0.16, 0.07
Chikwawa	8, 9, 10, 11, 12	0.42, 0.54, 0.44, 0.50, 0.47	0.18, 0.29, 0.19, 0.25, 0.22
Nsanje	9, 10, 11, 12	0.46, 0.54, 0.71, 0.69	0.21, 0.29 0.50, 0.48

Table 4.4b: Correlation coefficients for dekadal NDVI and local maize yield. Shaded cells indicate RDPs where r > 0.40; dekads are counted from 1 January.

RDP	Consecutive Dekads of high correlation coefficient (r)	Correlation coefficient (r)	Coefficient of determination (r^2)
Chitipa	11, 12	-0.42, -0.33	0.18, 0.11
Karonga	2, 3	-0.52, -0.61	0.27, 0.37
Rumphi/Mzimba	12	0.54	0.29
Nkhata Bay	31, 32	-0.52, -0.52	0.27, 0.27
Mzimba Central	11, 12	0.38, 0.50	0.14, 0.25
Mzimba South	11, 12	0.50, 0.51	0.25, 0.26
Nkhota kota	31, 32	-0.39, -0.38	0.15, 0.14
Kasungu	2, 3	0.24, 0.38	0.06, 0.14
Ntchisi	8, 9	-0.39, -0.27	0.15, 0.07
Dowa West	31, 32	0.27, 0.31	0.07, 0.10
Mchinji	28, 29	-0.49, -0.34	0.24, 0.12
Salima	10, 29, 30	0.51, 0.61, 0.62	0.26, 0.37, 0.38
Dowa East	2, 3	0.35, 0.39	0.12, 0.15
Mangochi	11, 12	0.56, 0.61	0.31, 0.37
Lilongwe West	10, 11	0.34, 0.36	0.12, 0.13
Lilongwe East	10, 11, 12	0.64, 0.66, 0.57	0.41, 0.44, 0.32
Dedza Hills	10, 11	0.31, 0.29	0.10, 0.08
Namwera	29, 30, 31	0.31, 0.46, 0.30	0.10, 0.21, 0.09
Bwanje Valley	9, 10, 11, 12	0.57, 0.66, 0.71, 0.67	0.32, 0.44, 0.50, 0.45
Thiwi-Lifidzi	10, 11	0.58, 0.49	0.34, 0.24
Ntcheu	11, 12	0.49, 0.51	0.24, 0.26
Kawinga	30, 31	0.27, 0.22	0.07, 0.05
Balaka	11, 12	0.67, 0.72	0.45, 0.52
Zomba	2, 3	0.32, 0.42	0.10, 0.18
Mwanza	11, 12	0.32, 0.54	0.10, 0.29
Shire Highlands	1, 2, 3	0.34, 0.32, 0.21	0.12, 0.10, 0.04
Phalombe	2, 3	0.34, 0.44	0.12, 0.19
Mulanje	1, 2	0.60, 0.55	0.36, 0.30
Chikwawa	9, 10, 11, 12	0.51, 0.47, 0.49, 0.41	0.26, 0.22, 0.24, 0.17
Nsanje	10, 11	0.57, 0.52	0.32, 0.27

From tables 4.4a and 4.4b, eight RDPs were identified that give statistically significant relationship between consecutive dekadal NDVI and maize yield. These include Salima, Mangochi, Lilongwe East, Bwanje Valley, Ntcheu, Balaka, Mulanje and Nsanje (shaded in tables 4.4a and 4.4b). However, in most of these RDPs the correlations were significant for only one variety but they were still used for comparison (that is, table 4.4a for hybrid variety and table 4.4b for local variety). Only Bwanje Valley RDP gave significant correlations for both varieties.

From the multiple linear regression analysis (in the selected RDPs) it has been shown that dekadal NDVI values, especially towards the end of the growing season are positively and moderately related to the final yield (as shown tables 4.5a and 4.5b later). This is due to the fact that during this period the vegetation (crops inclusive) will be at the peak of its activity in terms of biomass accumulation. The vegetation condition at these late stages determines how much photosynthate will be converted to yield during the reproductive stage. Except for just a few isolated cases such as Nkhota kota, Ntchisi, Mchinji, Namwera, and Kawinga RDPs, positive correlations were observed between yield and the dekadal NDVI in both the Southern and Central regions of the country, but for most of the RDPs in the Northern region except for Karonga and Rumphi/Mzimba RDPs, either negative or insignificant correlations were obtained between dekadal NDVI and yield. Therefore no regression analyses were done for any of the RDPs in the North. The negative correlations in general for the northern parts of the country compared to the positive relationships in both the central and southern parts mean that in the northern region there are other factors that play a more significant role in the determination of yield than does the soil moisture balance that controls vegetation growth. Most RDPs in the northern parts of the country receive high seasonal rainfall (except for the ones showing positive correlations). This means low variations in NDVI. However seasonal yield variations are high therefore resulting in these negative or very weak relationships.

Considering the percentage of planted area in the North compared to the South and Central, more of the area in the South and Central is planted to crops compared to the North where a larger percentage of land cover comprise perennial forests. This can have different impacts on the relation between NDVI and crop observations, resulting in differences in the NDVI-yield relationships in the RDPs of these respective regions. The explanations for the higher correlations towards the end of the growing season have already been given in the previous sections. On another note, it was also observed that for the local variety there were more RDPs that gave positive correlations than was the case with the hybrid variety. This may be as a result of the differences in plant structures together with leaf sizes and orientation between hybrid and local maize varieties. The local variety's plants are large in size, as are the leaves, whereas the hybrid variety's plants are short and small with small and short leaves. This leaves a lot of bare ground exposed to the sensors, for the hybrid variety whereas there is maximum ground cover by the crop in the case of the local variety. Therefore when the satellite images are taken there is significant inclusion of the bare soil in hybrid maize compared to the local variety. NDVI values for bare soils are either very small or negative hence the negative correlations with hybrid maize yield. As for the local variety, ground cover is relatively high though not complete. This partly explains the presence of both positive and negative correlations between local yield and these NDVI parameters. Another explanation is that local maize is widely grown by the small-scale farmers compared to the hybrid variety in Malawi (from observation, though there are no statistics to support this). This means that a considerable portion of the satellite image pixel comprise reflectance from the local compared to the hybrid variety hence explaining the differences in correlations between NDVI and the two varieties' yield.

(c) Dekadal NDVI-yield regression models for selected RDPs

In order to model maize yield for a given RDP from dekadal NDVI, the set of dekads that gave statistically significant correlation with yield were considered in performing linear regression analysis. This procedure was also used by Mutikani (1997). For each of the years in the series, a new NDVI variable was formulated in which the individual dekads were weighted in proportion to the value of the coefficient of determination (r^2). The new set of variables was then regressed against final maize yield. The resulting linear multiple regression equations were of the form:

$$Yield = a + b_1(w_1x_1) + b_2(w_2x_2) + \ldots + b_n(w_nx_n), \qquad (4.1)$$

where a is the regression constant, x_i is the NDVI for dekad i, b_i is the x coefficient, and w_i is the weight of the coefficient of determination for dekad i calculated by $w_i = r_i^2 \Big/ \sum_{i=1}^{n}\left(r_{i,\ldots,n}^2\right)$, where r_i^2 is the coefficient of determination for dekad i and the denominator is the sum of the coefficients of determination for all the dekads used in the model (1 to n).

Considering the values of r^2 only eight RDPs for each variety were modelled. In the other RDPs the correlation coefficients were either negative or insignificant at the 95% significance level so could not be considered for regression analysis. Tables 4.5a and 4.5b show the selected RDPs, their respective models and the coefficients of determination for the models.

Table 4.5a: Dekadal NDVI-yield regression models (hybrid). Shaded cells indicate $r^2 > 0.40$.

RDP	Models	r^2-values	F-value	N
Salima	$Y = -5089.9 - 2281.86 * NDVI_{10} + 15291 * NDVI_{11} + 790.848 * NDVI_{12}$	0.47	4.93	19
Bwanje Valley	$Y = -6245.6 + 8592.82 * NDVI_{10} + 755.24 * NDVI_{11} + 4762.38 * NDVI_{12}$	0.53	6.09	19
Lilongwe East	$Y = -1026.3 - 6966.4 * NDVI_{10} + 24485 * NDVI_{11} - 11674 * NDVI_{12}$	0.29	2.03	19
Ntcheu	$Y = -3430.9 + 10013.7 * NDVI_{10} - 814.13 * NDVI_{11}$	0.43	3.21	19
Mangochi	$Y = -3156.3 + 766.11 * NDVI_{11} + 9050.9 * NDVI_{12}$	0.33	3.99	19
Balaka	$Y = 3365 + 1357.02 * NDVI_{11} + 7491.55 * NDVI_{12}$	0.40	4.20	15
Mulanje	$Y = 444.47 + 30223 * NDVI_1 - 53952 * NDVI_2 + 26108 * NDVI_3$	0.23	1.61	19
Nsanje	$Y = -4471.7 + 4491.88 * NDVI_{10} + 4423.23 * NDVI_{11}$	0.53	6.62	16

Table 4.5b: Dekadal NDVI-yield regression models (local). Shaded cells indicate $r^2 > 0.40$.

RDP	Models	r^2-values	F-value	N
Salima	$Y = -1675.1 + 666.89 * NDVI_{11} + 4287.47 * NDVI_{12}$	0.38	5.08	19
Bwanje Valley	$Y = -2077.6 + 752.46 * NDVI_9 + 4283.76 * NDVI_{10} - 2895.19 * NDVI_{11} + 2550.81 * NDVI_{12}$	0.54	3.82	19
Lilongwe East	$Y = -1663.3 + 1079.93 * NDVI_{10} + 2211.6 * NDVI_{11} + 1308.26 * NDVI_{12}$	0.43	4.09	19
Ntcheu	$Y = -466.34 + 892.11 * NDVI_{11} + 9050 * NDVI_{12}$	0.27	2.20	19
Mangochi	$Y = -928.98 + 261.24 * NDVI_{11} + 2865.67 * NDVI_{12}$	0.35	4.75	19
Balaka	$Y = 1567 + 2257.27 * NDVI_{11} + 1347.52 * NDVI_{12}$	0.54	7.48	15
Mulanje	$Y = -328.78 + 1346.13 * NDVI_1 + 540.59 * NDVI_2$	0.36	4.60	19
Nsanje	$Y = -1390.5 + 2112.2 * NDVI_{10} + 951.9 * NDVI_{11}$	0.33	3.97	19

For the modelled RDPs except for Mulanje, both the hybrid and local final yields can be estimated using NDVI values mainly from the three dekads of the month of April. For Mulanje RDP, the NDVI for the first and second dekads of January can be used for the yield prediction. It was found that in most of the cases shown in tables 4.5a and 4.5b, the multiple regression analysis did not show significant improvement on the predicting power of consecutive dekadal NDVI over that of single dekads included in the series. For example, for Salima RDP (hybrid variety) the highest r^2 for a single dekad (dekad 11) was 0.49 but after integrating NDVI for three consecutive dekads the value of r^2 decreased to 0.47. In the other RDPs the r^2 for the integrated NDVI with yield were either lower than or the same as those from the dekads used individually. Generally the r^2-values did not improve significantly after incorporating NDVI values for consecutive dekads over those where NDVI for just one dekad were used, especially for those that singly gave highest values of r^2. However, the use of more than one dekad in the regression and subsequent yield prediction is recommended because it helps to minimize the effect of noise in the dekadal NDVI data that may be as a result of using data from only one dekad (Mutikani, 1997).

NDVI at the flowering dekad (eight to ten dekads after the planting dekad as determined using rainfall thresholds) was correlated with yield for five selected RDPs across the country to find out if NDVI at such particular stages was related to maize yield in a particular way. The results for the five RDPs are shown below.

Table 4.6: Correlation coefficients for dekadal NDVI at flowering dekad with maize yield

RDP	Chitipa		Kasungu		Salima		Mwanza		Nsanje	
Variety	Hybrid	Local	Hybrid	Local	Hybrid	Local	Hybrid	Local	Hybrid	Local
Correlation Coefficients (r)	-0.44	-0.22	0.09	0.26	0.01	0.24	-0.14	-0.20	-0.14	-0.38

From table 4.6, almost all the correlation coefficients for the two varieties are insignificant at the 95% significance level. For the same RDPs the highest correlations for dekadal NDVI as shown in tables 4.4a and 4.4b were as follows for hybrid and local varieties respectively: Chitipa (-0.54, -0.42), Kasungu (0.34, 0.38), Salima (0.70, 0.62), Mwanza (0.63, 0.54) and Nsanje (0.71, 0.57). Therefore, the use of the NDVI at the flowering dekad for yield estimation was not found better compared to the case when selected dekadal NDVI were used.

4.3.3 Monthly NDVI and yield

Monthly NDVI values were correlated with maize yields and the results are shown in table 4.7.

Table 4.7: Correlation coefficients for monthly NDVI and maize yield (the shaded cells indicate months where r > 0.40, at 95% significance level, were obtained in different regions)

RDP	Variety	Month(s) of highest correlation	Correlation coefficient (r)	Coefficient of determination (r^2)
Chitipa	Hybrid	April	-0.52	0.27
	Local	April	-0.40	0.16
Karonga	Hybrid	February	0.33	0.11
	Local	January	-0.53	0.28
Rumphi/Mzimba	Hybrid	November	0.55	0.30
	Local	November	-0.27	0.07
Nkhata Bay	Hybrid	February	-0.35	0.12
	Local	November	-0.51	0.26
Mzimba Central	Hybrid	March	-0.47	0.22
	Local	April	0.39	0.15
Mzimba South	Hybrid	October	-0.21	0.04
	Local	April	0.50	0.25
Nkhota kota	Hybrid	February	-0.29	0.08
	Local	November	-0.39	0.15
Kasungu	Hybrid	January	0.29	0.08
	Local	April	0.27	0.07
Ntchisi	Hybrid	March	0.47	0.22
	Local	March	-0.26	0.07
Dowa West	Hybrid	November	0.24	0.06
	Local	April	0.27	0.07
Mchinji	Hybrid	January	0.33	0.11
	Local	March	0.28	0.08
Salima	Hybrid	April	0.70	0.49
	Local	April	0.61	0.37
Dowa East	Hybrid	November	-0.13	0.02
	Local	January	0.35	0.12
Mangochi	Hybrid	April	0.51	0.26
	Local	April	0.54	0.29
Lilongwe West	Hybrid	April	0.45	0.20
	Local	April	0.36	0.13
Lilongwe East	Hybrid	April	0.52	0.27
	Local	April	0.67	0.45
Dedza Hills	Hybrid	January	0.25	0.06
	Local	November	0.39	0.15
Namwera	Hybrid	January	-0.39	0.15
	Local	April	-0.32	0.10
Bwanje Valley	Hybrid	April	0.72	0.52
	Local	April	0.71	0.50
Thiwi-Lifidzi	Hybrid	January	0.41	0.17
	Local	January	0.41	0.17
Ntcheu	Hybrid	April	0.55	0.30
	Local	April	0.48	0.23
Kawinga	Hybrid	January	-0.34	0.12
	Local	April	0.21	0.04
Balaka	Hybrid	April	0.59	0.35
	Local	April	0.68	0.46

Zomba	Hybrid	March	0.46	0.21
	Local	January	0.32	0.10
Mwanza	Hybrid	April	0.46	0.21
	Local	April	0.32	0.10
Shire Highlands	Hybrid	November	-0.44	0.19
	Local	January	0.31	0.10
Phalombe	Hybrid	January	0.53	0.28
	Local	April	0.41	0.17
Mulanje	Hybrid	April	0.33	0.11
	Local	January	0.55	0.30
Chikwawa	Hybrid	April	0.51	0.26
	Local	April	0.50	0.25
Nsanje	Hybrid	April	0.72	0.52
	Local	April	0.55	0.30

On monthly NDVI values, it was found that April NDVI aggregates were the most strongly correlated with yield. As explained in section 3.3.1, the maximum composite value for dekadal NDVI was used to represent the monthly value. In all but one RDP (Chitipa) where stronger correlations were observed, the correlation coefficients were positive indicating direct relationship between the variables. However, in some RDPs the r-values were stronger in preceding months other than in April (see table 4.7). In cases where the latter case applied it was observed that the majority of the RDPs, especially those in the northern region, gave negative relationships between monthly NDVI and yield. For the hybrid variety, about 40% of the RDPs had their highest correlations between monthly NDVI and final yield in the month of April, 10% in the month of March, another 10% in the month of February, 24% in January, 13% in November and only 3% of the RDPs in October. As for the local maize, 60% of the RDPs showed their highest correlations between the two parameters in question in the month of April, about 7% in March, 20% in January and another 13% in November. This explanation just considered the relative magnitude of the correlation coefficients not their significance. Table 4.7 shows a summary of these results. It can also be observed from this table that coefficients of determination (r^2) for all but three (for each variety) RDPs were insignificant at the 95% significance level. The RDPs that gave significant regressions are namely Salima, Bwanje Valley and Nsanje (r^2= 0.49, 0.52 and 0.52, respectively) for hybrid maize; and Lilongwe East, Bwanje Valley and Balaka (r^2= 0.45, 0.50 and 0.46) for local maize. This suggests that monthly NDVI aggregates (especially for April), in some areas can be used qualitatively to indicate crop condition but their use in (quantitatively) forecasting crop yield in Malawi was found to be limited considering the variation in yield explained by the monthly NDVI aggregates and the number of RDPs that showed high correlations. The RDPs that gave stronger relationships are shaded in table 4.7. The same table also shows that for most of the RDPs in the Northern region of the country (the RDPs in the table are arranged from north to south), the relationships between monthly NDVI and yield were negative for both varieties. For the rest of the country's RDPs positive relationships were obtained. In some RDPs it was observed that the two varieties gave correlations with opposite signs. For example, Ntchisi RDP in the central zone the stronger correlations were obtained in March for both varieties, but the hybrid variety gave a correlation coefficient (r) of 0.47 while the local variety gave correlation coefficient (r) of -0.26. These discrepancies may be attributed to the quality of the yield data. Since the yield data used are the National Statistical Office (NSO) post-harvest estimates, it cannot be considered 100% accurate. However, the same type of data was used in earlier studies by Nayava and Munthali (1992) to produce regression models for the FAOINDEX for maize yield forecasting in Malawi.

4.3.4 Seasonal NDVI averages and maize yield

Four sets of averages during the growing period were respectively correlated with yield. These included averages from October to April, November to January, November to February, and December to February. The purpose for trying the last three options was to test whether these averages could be useful tools for early warning purposes. The results showed that none of the four averages were significantly correlated to final yield for either hybrid or local varieties. However, for each of the two sets of NDVI seasonal averages (October to April, and November to January) only three RDPs gave moderate correlations for hybrid and local maize varieties. For the hybrid variety using the NDVI averages from October to April, Rumphi/Mzimba, Lilongwe East and Bwanje Valley RDPs gave r = 0.49, 0.47 and 0.43, respectively while for the local variety using the same averages, only Nkhata Bay, Nkhota kota, and Lilongwe East gave r = -0.55, -0.47 and 0.44 respectively. Using the November to January averages, the hybrid variety showed relatively high correlations in Rumphi/Mzimba, Lilongwe West and Thiwi-Lifidzi (r = 0.62, 0.41 and 0.45, respectively), while for the local variety relatively high correlations were obtained for Nkhata Bay, Thiwi-Lifidzi and Mulanje (r = -0.45, 0.42, and 0.41, respectively). Though most of the other RDPs gave weaker correlations, it was observed that most of them were positively correlated to the final yield. No regression analyses were attempted for these parameters due to the weaker relationships obtained from the correlation analysis. Full results are given in Appendix III.

4.3.5 Seasonal maximum NDVI and maize yield

Seasonal maximum NDVI and yield were found to be very weakly correlated in almost all RDPs, except for only three and two RDPs respectively for hybrid and local maize that gave moderate correlations, namely Rumphi/ Mzimba, Mchinji and Balaka (r = 0.43, 0.45 and 0.53, respectively) for hybrid maize; and Nkhata Bay and Zomba RDPs (r = 0.48 and 0.55, respectively) for local maize. No attempt was also made to do a regression analysis between these two parameters due to the weak relationships between them in most RDPs. The table in Appendix IV shows full results.

4.3.6 Cumulative NDVI and yield

RDPs with correlation coefficients (higher than 0.40) between cumulative NDVI and yield and the dekads when the higher correlations were attained are shown in table 4.8. Full results of this analysis are given in Appendix V.

Table 4.8: Dekads and values of higher correlations between cumulative NDVI from first dekad of October (dekad 28) and maize yield

RDP	Variety	Dekads of higher r	Highest r and dekad attained
Rumphi/ Mzimba	Hybrid	33 to 12	0.65 (3)
Nkhata Bay	Local	34 to 12	-0.62 (9)
Nkhota kota	Local	32 to 12	-0.56 (34)
Ntchisi	Local	28	-0.40 (28)
Mchinji	Local	28 to 292	-0.47 (28)
Salima	Local	33 to 35	-0.51 (33)
Lilongwe West	Hybrid Local	28 28, 29	-0.45 (28) -0.56 (28)
Lilongwe East	Hybrid Local	11 to 12 11 to 12	0.48 (12) 0.45 (12)
Bwanje Valley	Hybrid	12	0.42 (12)
Thiwi-Lifidzi	Hybrid Local	28 to 29 28	-0.45 (28) -0.45 (28)
Balaka	Local	28	-0.40 (28)
Zomba	Hybrid	34 to 1	-0.46 (35)
Mwanza	Hybrid	33 to 10	-0.44 (34)
Shire highlands	Hybrid	33 to 36	-0.47 (34)
Mulanje	Local	3 to 6	0.41 (4)
Chikwawa	Hybrid	33 to 35	-0.42 (33)
Nsanje	Local	12	0.43 (12)

NDVI cumulated from different points at the beginning of the season like the first dekad of October (dekad 28), first dekad of November (dekad 31) and first dekad of December (dekad 34) to any point in the growing season were correlated, dekad after dekad, with yield in order to determine whether the total NDVI value up to some particular point would be a useful variable to estimate final maize yield. It was observed that only a few RDPs gave moderate relationships between the two variables for each of the two varieties. Furthermore, the time at which the higher correlations occurred varied from one RDP to another. It was also observed that most of the RDPs gave higher correlations for a number of consecutive dekads rather than just one dekad. Mostly, where stronger relationships occurred, the value of the correlation coefficient would be negative indicating indirect (opposite) association between cumulative NDVI and maize yield. For example, table 4.8 shows that in Rumphi/Mzimba RDP (for hybrid maize), the dekads of moderate correlation coefficient stretched from the third dekad of November (dekad 33) to the third dekad of April (dekad 12) with the highest of all the correlation coefficients being attained during the third dekad of January (dekad 3), r = 0.65. For other RDPs like Lilongwe East the peak correlations occurred during the same dekads for both varieties, and magnitudes and signs of the correlations were also similar. That is, during dekad 12 and r = 0.48, and r = 0.45 respectively for hybrid and local final maize yield. In other RDPs higher correlations between cumulative NDVI and maize yield occurred during different times and also the magnitudes were different for the two maize varieties. In some RDPs higher correlations occurred for long periods (for example, Rumphi/Mzimba, Nkhata Bay, Nkhota kota and Mwanza) while in others like Ntchisi, Mchinji, Lilongwe West, Bwanje Valley, Thiwi-Lifidzi, Balaka and Nsanje higher correlations occurred for just one or two (consecutive) dekads. Few RDPs gave moderate correlations between NDVI cumulated from either November or December and yield. However, in the few RDPs where the moderate correlations occurred, they were positive in contrast to the case when NDVI were cumulated from October (table 4.8), in which case the correlation coefficients were mainly negative. This trend resembles the other cases already discussed in this section where vegetation (NDVI) tends to establish some pattern later in the growing season hence the positive relationships with yield (though not consistently so). Regression analysis could not be performed due to the fact that just a few RDPs gave moderate correlations with yield and that most of these correlations were negative. Some of the RDPs in which the relationships were moderate, and when they occurred for the case in which the NDVI was cumulated from the first dekad of October (dekad 28) are given in table 4.8. The dekad that is shown in table 4.8 represents the point up to which the NDVI had been cumulated beginning from dekad 28. For example, dekad 33 represent NDVI cumulative from dekad 28 to the third dekad of November (dekad 33).

The explanations for the lack of significant and consistent relationships may be that, generally, the cultivated areas do not cover the whole RDP, and that areas planted to maize in some RDPs are about half or less the total cultivated areas. Unfortunately the actual proportions of the RDPs planted to maize could not be quantified due to lack of data on the sizes of each RDP. Since the resolution of the satellite sensors for the NDVI used for the study is coarse, the probability of missing the finer details of the relationships are very high hence the inconsistency in most RDPs. The other problem of the coarse resolution data is that even within the cultivated areas separation of crops is not possible. So the satellite-based NDVI observations included all the crops in the cultivated areas, where in some cases maize is not the dominant crop, the signal picked by the sensors might not have a significant contribution from maize crop. For example, in rice, tea and coffee growing areas maize is not very popular but the NDVI extraction included such areas to represent maize growing areas in general. These mixed satellite observations (signals) were correlated with maize yield. In a related observation, it was discovered much later during the course of the study that in some cases, in the cultivated area maps used, large cultivated portions in some of the RDPs had been left out during NDVI extraction, as only the largest cultivated portion was being selected from each RDP. For example, in Chitipa, about 33% of the cultivated area in the RDP was omitted from the extraction and analysis; in Karonga about 50% was left out; about 25% in Nkhata Bay; about 50% in Mzimba South; about 66% in Mangochi; about 33% in Mwanza RDP; about 20% in Chikwawa; and about 25% in Nsanje RDP were left out. While the cultivated portions gave limited representation of the cultivation in the RDP, the yield statistics that the NDVI was being correlated to was an average over the entire RDP. Therefore, the limited fit of areas being compared for NDVI and yield could have negatively affected the NDVI-yield relationship.

On trying to explain the inconsistency that was observed between NDVI and maize yield for the two varieties, the two sets of yields were correlated with each other. In most RDPs the yields from two varieties were moderately to highly correlate to each other. However, in the following four RDPs very weak and negative correlations were obtained as follows: Rumphi/Mzimba ($r = 0.26$), Nkhata Bay ($r = -0.01$), Mzimba Central ($r = 0.25$) and Mzimba South ($r = -0.20$). The correlation coefficients between the two varieties for each RDP are given in table 4.9. With such correlations between the two varieties, it was not surprising to obtain different results for each variety especially in the above RDPs. These low/ negative correlations between varieties explain the situation where in some cases the correlation between NDVI with one variety would be positive while that with the other variety it would be negative. For example, using monthly NDVI aggregates, in Ntchisi RDP, the month of highest correlation was March but the correlation coefficients (r) for NDVI-hybrid and NDVI-local varieties were 0.47, and -0.26, respectively; for Rumphi/ Mzimba the highest correlations were observed in November, and the values of r were 0.55 and -0.27 for hybrid and local maize respectively. Ntchisi and Rumphi/ Mzimba were used for this discussion because the relatively high correlations were observed to occur during the same month for the hybrid and local varieties respectively. In the other RDPs, the highest correlation coefficients were observed in different months.

Table 4.9: Correlations between hybrid and local maize yields at RDP level in Malawi (The shaded cells indicate the lowest values of r obtained)

RDP	Correlation coefficients (r)		RDP	Correlation coefficients (r)
Chitipa	0.59		Lilongwe East	0.71
Karonga	0.63		Dedza Hills	0.68
Rumphi/Mzimba	0.26		Namwera	0.66
Nkhata Bay	-0.01		Bwanje Valley	0.86
Mzimba Central	0.25		Thiwi-Lifidzi	0.91
Mzimba South	-0.20		Ntcheu	0.54
Nkhota kota	0.64		Kawinga	0.66
Kasungu	0.88		Balaka	0.83
Ntchisi	0.65		Zomba	0.66
Dowa West	0.67		Mwanza	0.90
Mchinji	0.50		Shire Highlands	0.79
Salima	0.83		Phalombe	0.63
Dowa East	0.52		Mulanje	0.67
Mangochi	0.83		Chikwawa	0.94
Lilongwe West	0.80		Nsanje	0.80

4.4 Results from the AMS Model

4.4.1 AMS output parameters

After running the AgroMetShell (AMS) model at station level, using the dekadal climatic variables, the following were the output parameters (and possible maize yield predictors).

- Total Water Requirement (TWR_q respectively).
- The current, normal and extended/seasonal or forecast Water Requirement Satisfaction Index (WSI_c, WSI_n, and WSI_s respectively).
- The initial, vegetative, flowering, ripening and total Water Excess (WEX_i, WEX_v, WEX_f, WEX_r and WEX_t respectively).
- The initial, vegetative, flowering, ripening and total Water Deficit (WDF_i, WDF_v, WDF_f, WDF_r and WDF_t respectively).
- The initial, vegetative, flowering, ripening and total actual evapotranspiration (AET_i, AET_v, AET_f, AET_r and AET_t respectively).

These were interpolated and averaged to RDP level for every phenological phase of the crop cycle (as described in Section 3.3.5). An example table for Salima RDP is shown in table 3.4 of Section 3.3.5.

The advantages of using more than one parameter in contrast with just one as in the FAOINDEX approach is that if only one parameter is used, there is a high probability that some important factors that contribute to the determination of maize yield may be left out. For example, the assumption in the FAOINDEX is that once crops are affected by hazardous conditions like severe water deficit (drought) during earlier parts of the season, they do not recover. But actually they can recover if favourable conditions return before permanent damage has been done. In using different parameters that account for crop growing conditions at different stages, the process becomes very specific and takes in the relative effects at each stage according to physiological and genetic characteristics of a particular crop. For example, water deficit is considered most critical during the flowering stage while water excess may have pronounced impacts at a different phase; say the vegetative, of crop development. Therefore if both are used in the modelling process it may assist to capture all the factors that may influence production throughout the life cycle of the crop.

4.4.2 Regression analyses between AMS parameters and maize yield

A step-wise linear multiple regression analysis procedure was carried out in ***Instat Plus*** statistical package Version 3.20 at RDP level. This involved identification of parameters that gave the strongest correlation with yield for each RDP. The selection of parameters to include in the regression models depended on a number of factors including the value of the coefficient of determination (r^2), the F-value, and physiological and agronomic characteristics. Firstly, simple linear correlation between each of the AMS output parameters and yield (by way of correlation matrices) was performed in order to identify potential yield predictors for each RDP. Then linear multiple regression analysis was performed only for the parameters that gave stronger correlations for each RDP taking into consideration physiological and agronomic characteristics. From these analyses it was found that there was no consistency from one RDP to another with respect to the parameters giving stronger correlations as well as the degree of r^2 when more than one RDP had the same parameters giving stronger relationships. Therefore, the RDPs were grouped according to the agroclimatological zones of the country as shown in table 3.1. Then different combinations of parameters were regressed against yield and the combination that gave the highest values of r^2 and made agronomic sense was used for a particular zone. Regression equations were developed for each RDP in that zone from the same combination of parameters. Table 4.10 shows the parameters, model coefficients, values of coefficients of determination (r^2) as well as the standard errors (SE) for the regression models for each RDP.

Table 4.10: Regression models for both hybrid and local varieties with AMS output parameters (and SE for each model). Shaded cells indicate $r^2 > 0.40$

NORTH							
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(hybrid)	$y=a+b_1*WEX_r+b_2*WDF_r+b_3*AET_r$						
Chitipa		1629.9	12.925	73.943	11.987	0.60	±414
Rumphi/Mzimba		11567	30.092	113.76	-166.5	0.68	±370
Mzimba Central		3316.2	17.45	-2.528	-30.85	0.47	±617
Mzimba South		3146.2	1.2757	14.308	-16.14	0.08	±525
(local)	$y=a+b_1*WEX_r+b_2*WDF_r+b_3*AET_r$						
Chitipa		5222.3	0.246	52.321	-41.42	0.30	±203
Rumphi/Mzimba		5641.8	-0.51	52.842	-42.02	0.39	±196
Mzimba Central		93.503	3.842	-9.69	4.813	0.39	±138
Mzimba South		3788.5	-2.496	28.689	-27.56	0.12	±151
CENTRAL							
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(hybrid)	$y=a+b_1*WEX_f+b_2*WDF_r+b_3*AET_i$						
Ntcheu		7193.7	-3.727	47.291	-249.6	0.38	±528
Bwanje Valley		11568	-3.999	96.36	-507.6	0.82	±401
Thiwi-Lifidzi		4935.1	0.216	65.113	-144.4	0.59	±327
Dedza Hills		3150.4	-1.884	29.644	-53.86	0.12	±430
Lilongwe East		5476.4	0.803	61.815	-193.0	0.54	±387
Lilongwe West		5976.9	1.817	76.973	-230.3	0.76	±321
Dowa East		7151.3	5.138	-0.233	-355.8	0.50	±483
Mchinji		7576.2	-3.046	56.538	-306.5	0.37	±472
Dowa West		5680.6	-1.544	-0.276	-218.9	0.22	±550
Ntchisi		1946.5	3.533	-0.58	17.643	0.26	±531
Kasungu		3687.5	-1.819	28.593	-53.59	0.25	±393
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(local)	$y=a+b_1*WEX_f+b_2*WDF_r+b_3*AET_i$						
Ntcheu		4034.5	0.211	11.171	-137.9	0.43	±139
Bwanje Valley		5313.1	-0.985	22.907	-150.7	0.79	±143
Thiwi-Lifidzi		1308.5	0.695	7.639	-11.68	0.46	±169
Dedza Hills		941.7	-0.424	5.16	-2.948	0.23	±141
Lilongwe East		596.7	1.912	7.351	9.344	0.31	±271
Lilongwe West		1576.8	1.893	9.466	-19.88	0.51	±252
Dowa East		3539.8	0.162	7.811	-85.52	0.45	±229
Mchinji		1476.4	2.316	-2.564	-23.45	0.29	±232
Dowa West		2402.5	0.896	4.882	-45.90	0.22	±277
Ntchisi		4546.8	0.625	6.899	-113.8	0.43	±225
Kasungu		410.43	-0.6773	2.143	25.92	0.05	±206

LAKESHORE							
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(hybrid)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Karonga		-5602	66.412	5.6505	20.327	0.51	±427
Nkhata Bay		6982.9	-0.8033	101.99	-74.84	0.29	±429
Nkhota kota		-9136	144.8	-24.08	-36.64	0.66	±334
Salima		-9791	161.18	81.185	-51.67	0.53	±582
Mangochi		-4255	65.27	26.93	8.4864	0.66	±544
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(local)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Karonga		540.19	51.309	36.007	-45.52	0.63	±160
Nkhata Bay		12027	-14.944	126.85	-104.4	0.42	±193
Nkhota kota		18.872	70.597	40.475	-58.08	0.36	±316
Salima		-53.35	48.598	31.975	-36.89	0.42	±268
Mangochi		111.31	31.599	24.754	-28.25	0.58	±162
SOUTH							
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(hybrid)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Kawinga		9107.6	-46.556	63.529	-33.06	0.05	±921
Namwera		-2426.5	45.358	-22.68	5.1566	0.20	±584
Balaka		-9455.9	9.7842	-113.3	154.47	0.46	±704
Zomba		-2376	58.203	3.46	-8.496	0.19	±707
Mwanza		-3555.2	36.709	-13.86	34.699	0.81	±308
Shire Highlands		-2440	33.434	-17.60	21.137	0.57	±354
Phalombe		1496.3	12.763	37.706	-6.327	0.27	±516
Mulanje		4328.3	-14.321	54.219	-17.24	0.44	±446
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(local)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Kawinga		-1896.2	-11.45	-33.96	47.688	0.41	±183
Namwera		-412.27	6.9716	-8.706	5.7919	0.08	±156
Balaka		111.75	5.0687	6.5337	1.95	0.37	±211
Zomba		134.11	14.748	7.259	-8.59	0.20	±178
Mwanza		821.42	24.084	24.49	27.513	0.67	±158
Shire Highlands		-161.79	8.543	-0.811	1.662	0.32	±158
Phalombe		-2122	13.964	-18.73	20.159	0.11	±246
Mulanje		-5232.8	8.913	-50.31	56.538	0.47	±149
SHIRE VALLEY							
RDP	**MODEL**	**a**	**b_1**	**b_2**	**b_3**	**r^2**	**SE**
(hybrid)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Chikwawa		2441.5	13.637	45.185	-22.67	0.77	±312
Nsanje		2510	6.469	32.873	-18.03	0.70	±254

RDP	MODEL	a	b_1	b_2	b_3	r^2	SE
(local)	$y=a+b_1*WSI_s+b_2*WDF_r+b_3*AET_r$						
Chikwawa		-1736.3	16.488	-12.27	13.746	0.75	±131
Nsanje		-3982.2	8.789	-45.34	45.718	0.52	±139

Since there are other local factors that affect crop yield other than the soil water balance, the different performance of the models in different RDPs even in the same zone is expected since these zones cover synoptically large areas therefore not expected to have exact features with each other despite being considered to fall within the same climatological regimes. Some of the other factors that may affect final crop yield apart from soil water balance include crop husbandry/ farming practices, pests and disease management; and availability and utilization of resources like fertilizers (Mkhabela *et al,* 2005).

For the RDPs in the Northern zone, soil water conditions during the reproductive stages were found to be important in the determination of final crop yield. This is in agreement with Jackson (1989) who stated that the soil moisture during flowering and early grain formation is very critical in the determination of yield. These two stages comprise the reproductive stage (Doorenbos and Kassam, 1979). Grain formation is part of ripening hence the sensitivity at the ripening stage is also considered to be in the reproductive stage. Research in Costa Rica established that the rain amount and distribution during the reproductive stages was the main meteorological factor that has a significant impact on yield (Lomas and Herrera, 1984). Just like in other parts of the country, it was found that the coefficients of determination (r^2) for the hybrid variety were generally higher than those of the local variety. For example, in Chitipa r^2_h=0.60 while r^2_l=0.30, where r^2_h is the coefficient of determination for the hybrid maize while r^2_l represents that for the local variety; and in Rumphi/ Mzimba r^2_h=0.68 while r^2_l=0.39. In Mzimba South RDP, the trend was reversed though the correlations were very weak, with only 8% and 12% variance in yield being explained for hybrid and local varieties by the combined variance in WEX_r, WDF_r and AET_r. This shows that in Mzimba South RDP, there are other factors that significantly affect yield other than the soil water balance, as is the case in the other RDPs in the same zone. Some of the factors that may affect final yield have been outlined in the previous paragraph. This might be applicable to most of other RDPs where the coefficients of determination were generally weak. It can also be argued that since the RDP borders with the Central Zone, there might be a mixture of agro-ecological characteristics from both zones which make it hard for its behaviour to be purely one or the other.

For RDPs in the central zone, soil water conditions during reproductive phase, especially soil water excess at flowering and water deficit at ripening, were observed to make significant contributions to the determination of yield (table 4.10). In addition, actual evapotranspiration during the initial stages was found to contribute significantly to final yield. This observation on the importance of soil water balance at initial stages concurs with the observation by Jackson (1989) who found that for maize, water stress in early growth delays flowering. This stress-induced delay in silking leads to the loss of synchrony in development of silks and tassels, thereby negatively affecting final yield. The use of water excess (though at various stages) for both the northern and the central zones is justified because in the work done in Malawi by Hearn and Wood (1964) as quoted by Jackson (1989), over-irrigation was found to significantly reduce maize yield. The same reasoning can also be applied when it comes to too much rainfall because Jackson (1989) quotes Benacchio (1983) who found that in Venezuela, excess water to plants especially during the grain filling period, when demand was reduced, resulted in negative effects on yield. It is also common knowledge that maize does not perform optimally in water-logged soils due to lack of enough air (oxygen) for the roots which results in chlorosis (yellowing of leaves) and reduces the photosynthate production. If prolonged, these conditions may lead to death of the whole plant and thus complete crop failure. On another note, high intensities and concentration of rain in a few large storms results in a large proportion of water being lost as surface runoff, considerable erosion and only a limited proportion of water becoming available to plants. This can have significant negative effects on the final crop yield. The regression coefficients for the parameters, for example, those of Bwanje Valley RDP show that the effect of water excess on the final yield is minimal and negative, whereas that of the actual evapotranspiration during the initial stages is larger and also the overall effect is to reduce yield (table 4.10). As for the effect of water deficit at the ripening stage, the equation (model) shows a positive contribution. Since the model output is negative, the positive sign on the model shows that the overall effect of this parameter is to reduce the final yield as expected. Similar to the case for the Northern zone, r^2 values for hybrid are generally higher than those for the local variety (table 4.10).

As for the Lakeshore, Southern and Shire valley zones, it was found that the combination of parameters that gave the most significant and meaningful contribution to final yield was that of Water Requirement Satisfaction Index at the end of the season (WSI_s), water deficit (WDF_r) and actual evapotranspiration at the ripening stage (AET_r). Lakeshore and Shire Valley zones have almost similar characteristics, in terms of temperature and rainfall regimes, so their discussion is the same. The use of the water requirement satisfaction index was appropriate since it takes into account the water satisfaction for the crop throughout its whole lifecycle. The only weakness of this parameter is the assumption that once the crop is affected by drought (water shortages) in earlier parts of the season, it does not recover (Mukhala and Hoefsloot, 2004). But observations have shown that in some cases when the drought/ dry spell was not severe the crops may recover when normal growing conditions return. In the models, the

WSI_s is a positive multiplier of the coefficients hence having a positive contribution towards final crop yield. This means that the value of WSI_s is directly related to the final yield observed. Except for Nkhata Bay, the coefficients for water satisfaction index are positive in all RDPs for the two zones under consideration (Lakeshore and Shire Valley). Just like in the other zones and RDPs already discussed, water deficits and actual evapotranspiration have the effect of reducing expected final yield since the crop is put under water stress. In these two zones, water deficit and actual evapotranspiration have pronounced effects because climatologically these are high temperature and relatively low rainfall areas. Therefore water shortages are regularly experienced, and in many cases the seasons are short due to late onset and early cessation of rains, coupled with pronounced dry spells. Nkhata Bay RDP receives higher amounts of rain distributed almost throughout the season due to the sharp change in topography from the lake to the nearby mountains. Therefore water excess may have a significant contribution to final yield than the parameters that were chosen and used that reflect the effects of soil water deficit.

For the southern agroecological zone, though the $WSI_s/WDF_r/AET_r$ parameter combination was considered the best, in most of the RDPs the percentages of yield variation explained were not significant at the 95% significance level, especially for the local variety, in which case only three (Kawinga, Mwanza and Mulanje) out of the eight RDPs over 40% of the yield variation was explained by the variation in the predictands. As for the hybrid variety, in only four out of the eight RDPs (Balaka, Mwanza, Shire Highlands and Mulanje) over 40% yield variance was explained by the variation in the predictands (table 4.10). Mwanza RDP shows the highest degree of variation in yield being explained by the combined variation in the three parameters used for both the hybrid and local varieties (81% and 67% respectively). These high correlations suggest that water is the limiting factor in the crop (maize) production in this RDP compared to the other factors. Namwera and Zomba RDPs, on the other hand show the lowest values of r^2. That may be explained by the types of land use practised in these RDPs. In both Namwera and Zomba, maize is not extensively cultivated. In Namwera as well as in parts of Zomba, rice cultivation is common due to the type of soil and availability of swampy areas. However, the water balance interpolation process considered whole RDPs. In this case the lower correlations between the AMS output parameters and maize yield should be expected. As for the other RDPs, most of the cultivated areas (especially at small holder level) include maize. Therefore, after the interpolation the water balance parameters were expected to be strongly correlated to maize yield.

On a another note, it was also observed from table 4.10 that in some cases the same AMS parameter had different (positive or negative) effects on the determination of yield in different RDPs and at times, between the two varieties in the same RDP. For example, in the central zone, for hybrid variety models,

WEX_f had a yield reducing effect in Ntcheu RDP while the same parameter had a positive effect on yield in Ntchisi RDP. In Mchinji RDP, WDF_r had opposite effects on hybrid and local maize varieties. Since these observations have not been consistent, they were considered chance events. Also the fact that in most cases the odd parameter contributions were less significant compared to the contributions from the other parameters, their effects on the overall performance of the model were negligible.

Table 4.11 below is a summary of the how the AMS models performed in each agroecological zone and at the national level.

Table 4.11: Summary of the AMS model results

ZONE	VARIETY	r^2 RANGE	RDPs WITH $r^2 > 0.40$	SE RANGE (kg/ha)
North	Hybrid	0.08-0.68	3 of 4	370-617
	Local	0.12-0.39	0 of 4	138-203
Central	Hybrid	0.12-0.82	6 of 11	321-550
	Local	0.05-0.79	6 of 11	139-277
Lakeshore	Hybrid	0.29-0.66	4 of 5	334-582
	Local	0.36-0.63	4 of 5	160-316
South	Hybrid	0.05-0.81	4 of 8	308-921
	Local	0.08-0.67	3 of 8	149-246
Shire Valley	Hybrid	0.70-0.77	2 of 2	254-312
	Local	0.52-0.75	2 of 2	131-139
NATIONAL	**Hybrid**	**0.05-0.82**	**19 of 30**	**254-921**
	Local	**0.05-0.79**	**15 of 30**	**131-316**

Table 4.11 shows that the AMS models generally perform well across the country. However, the performance of hybrid maize models was better than that of the local maize models. For example in the northern zone, three out of four of the RDPs showed significant coefficients of determination for the hybrid variety while none of the RDPs showed significant values of r^2 for the local variety in the zone. For the Southern Zone, four and three out of eight gave r^2 greater than 0.40 for hybrid and local maize models respectively. At national level, more than half of the RDPs' models showed that over 40% variation in yield could be explained by the variation in the AMS parameters.

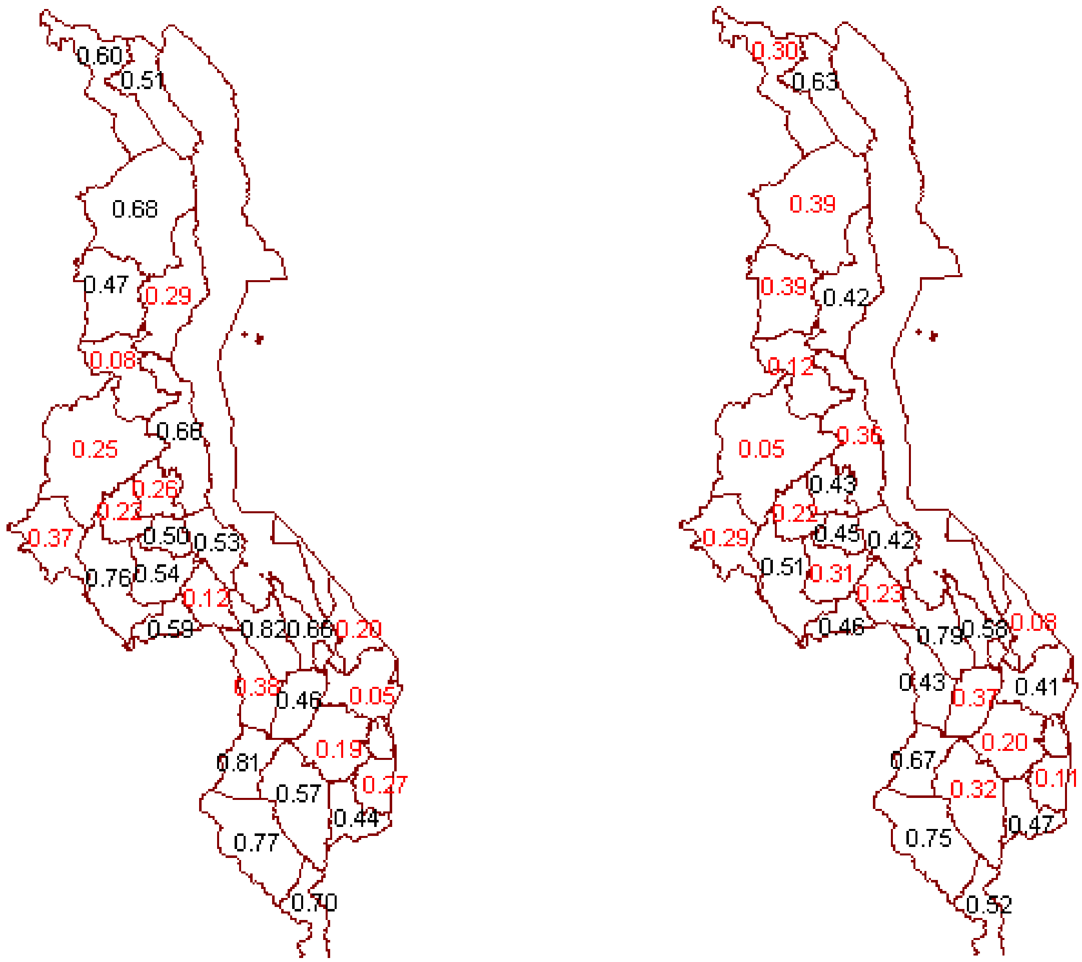
Standard errors for the models were calculated using the relationship given in equation 4.2:

$$SE = \sqrt{\frac{\sum (y_{act} - y_{est})^2}{n}} \qquad (4.2)$$

where SE is the standard error of the regression, y_{act}, and y_{est} are the actual and estimated yield (kg/ha), respectively, and n is the length of the data record. The results are incorporated in table 4.10 and the summary thereof is part of table 4.11.

Taking the upper limit of the standard errors to be to be ±475 kg/ha and ±200 kg/ha (these were national arithmetic means for the SE) for the hybrid and local maize regression models respectively, it was observed that in the Northern zone 50% of the RDPs had their SE above the threshold for the hybrid, while 25% of the RDPs had their SE above the threshold for the local variety. Using similar expressions, in the Central zone it was 36% and 64%, respectively; over the Lakeshore zone, it was 40% for both varieties; in the Southern zone, it was 63% and 25% respectively and finally in the Shire Valley zone none of the RDPs had their SE above the stated thresholds (see table 4.10). At national level, about 43% of the RDPs had the SEs outside the threshold for hybrid variety and about 40% of the RDPs were outside the threshold for the local variety. From this analysis, it can be said that the confidence interval for hybrid are wider than those for the local variety. This suggests that yield estimation/prediction can be done with more confidence for the hybrid than for the local variety. One of the reasons why hybrid maize performed better in many cases is that technologically hybrid is better developed than the local variety. Also the farmers who grow hybrid maize are consistent in their operations which are better than those for the local variety. These consistencies help the modelling process to be more reliable compared to the case of the local variety. Even at small scale level, advanced technologies have been practised resulting in the high performance of the hybrid variety.

Looking at the coefficients of determination (r^2) values from table 4.10, the RDPs with high r^2 were not located in one region but were distributed throughout the country. For example Rumphi/ Mzimba in the North, Bwanje Valley in the Centre, most of the RDPs in the Lakeshore, Mwanza RDP in the South and Nsanje and Chikwawa RDPs in the Shire Valley have quite high r^2 in this study using AMS parameters. Figure 4.7 shows the distribution of the r^2 across the country for hybrid (figure 4.7a) and local maize (figure 4.7b) varieties.

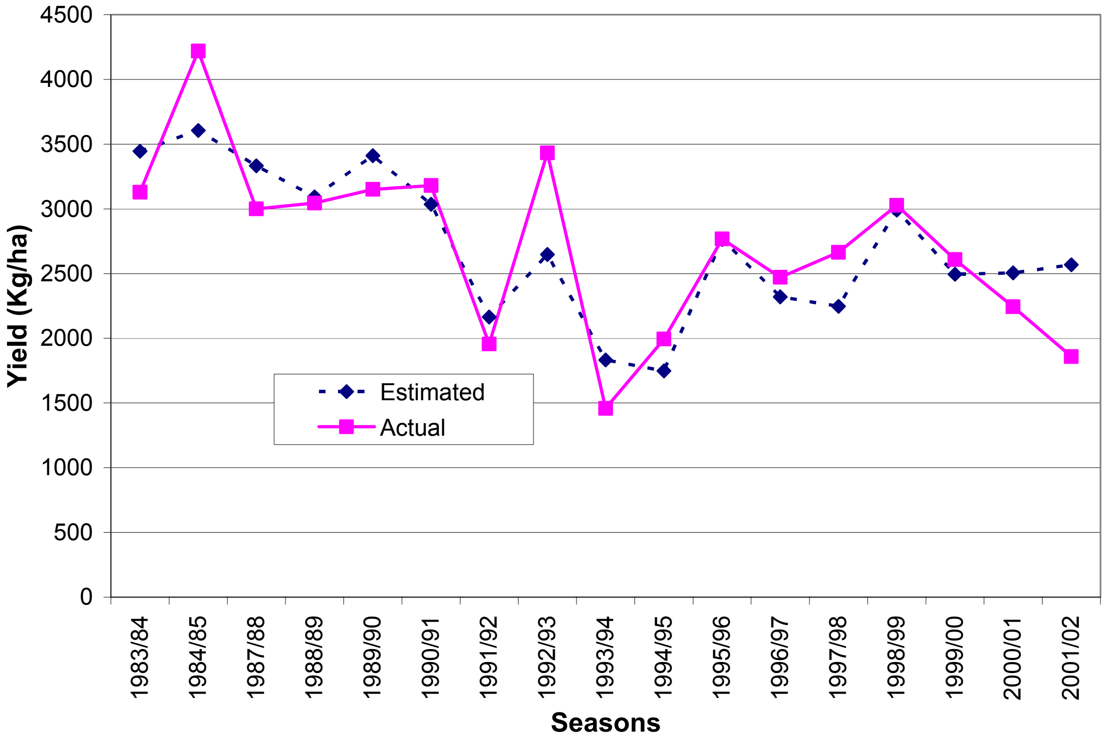


(a) Hybrid (b) Local

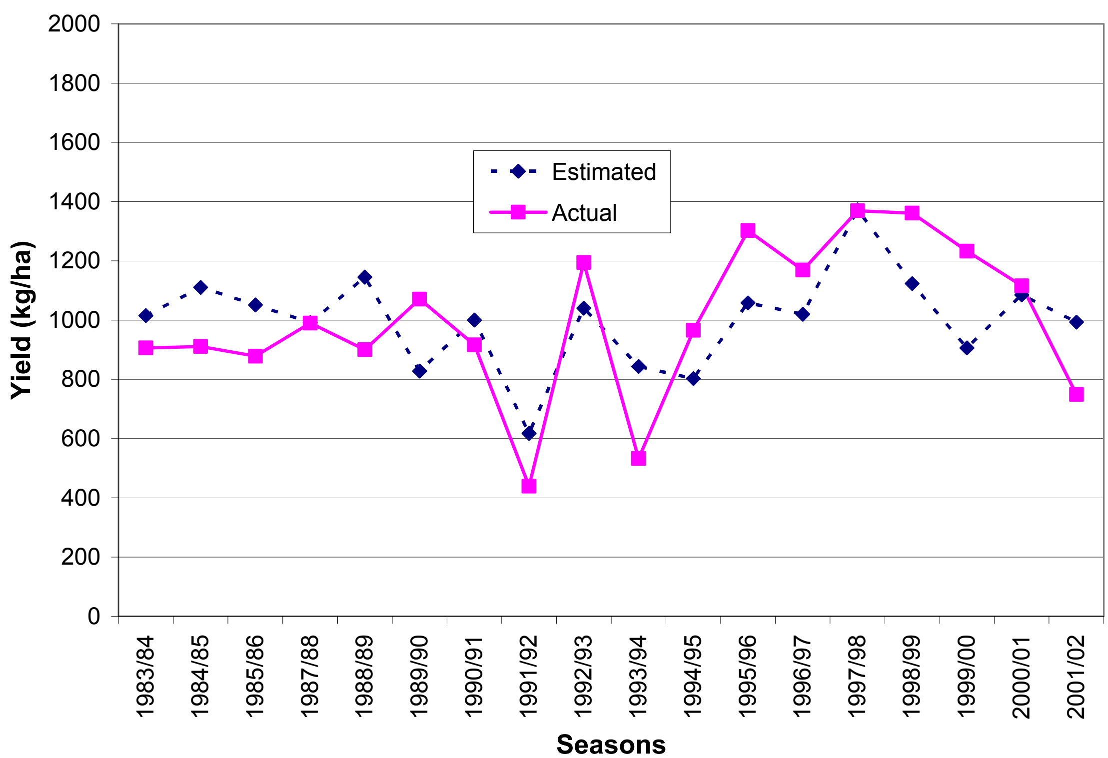
Figure 4.7: Map of Malawi showing the distribution of the coefficients of determination (r^2) for (a) hybrid maize variety and (b) local maize variety. The black labels indicate $r^2 >$ 0.40 while the red labels indicate $r^2 <$ 0.40.

4.4.3 Testing for goodness of fit of the regression models

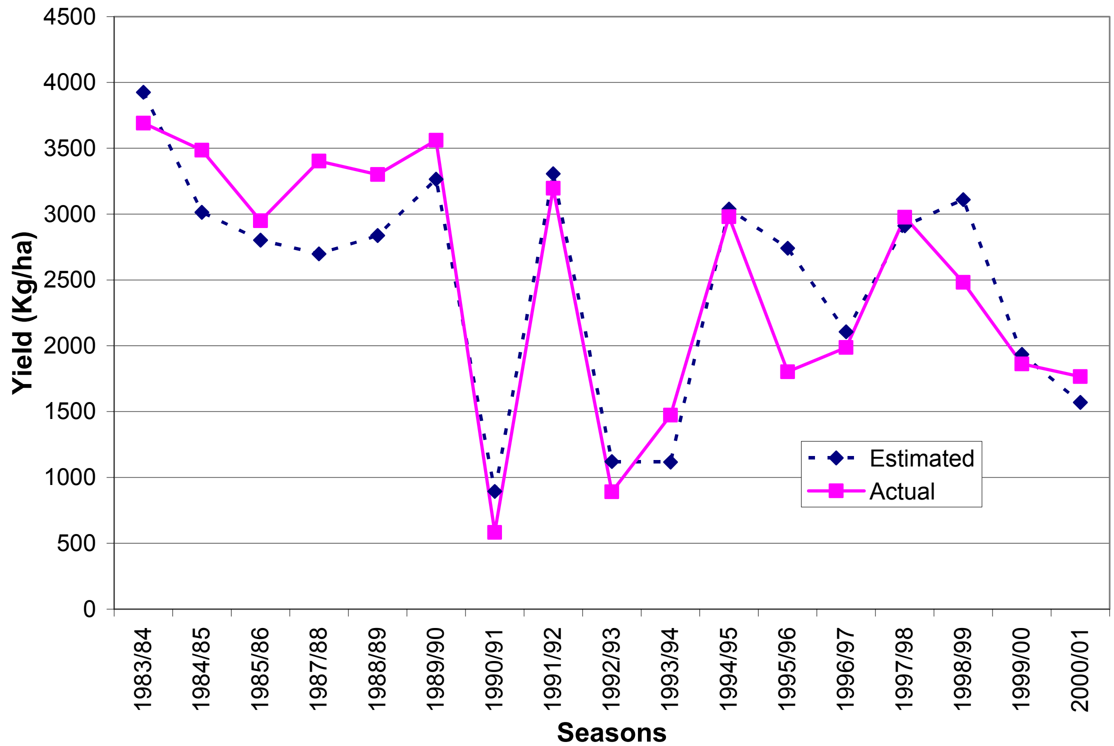
The regression equation models presented in table 4.10 were used to estimate yield for each RDP for the hybrid and local maize varieties for 19 seasons (1983/84 to 2001/02). The estimated yield and the actual yield for selected RDP are shown in Figures 4.8 (a) to (j). The graphs for the rest of the RDPs across the country are given in Appendix VI.



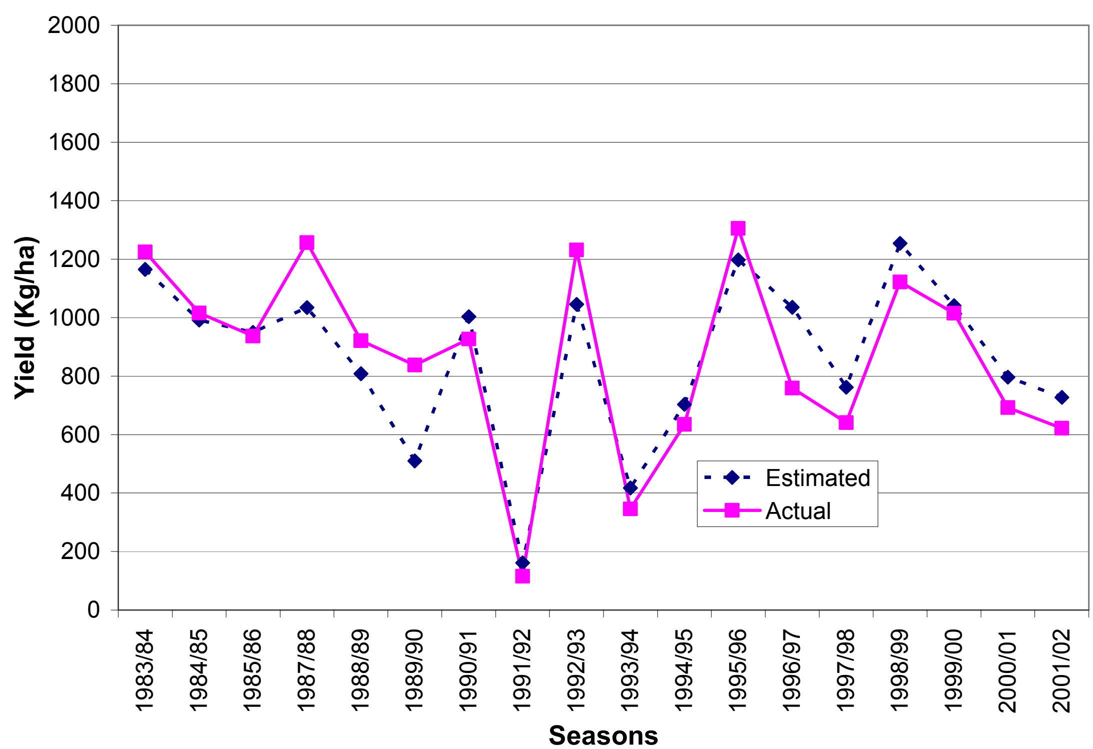
a) Rumphi/Mzimba RDP - Hybrid



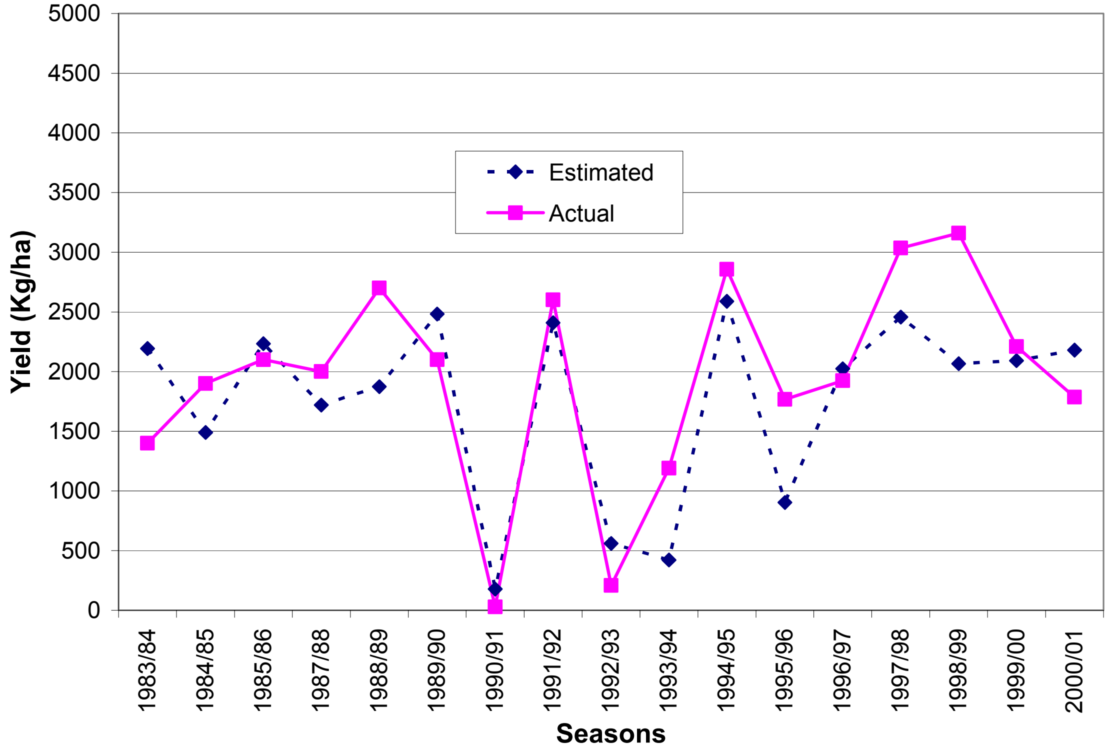
b) Rumphi/Mzimba RDP - Local



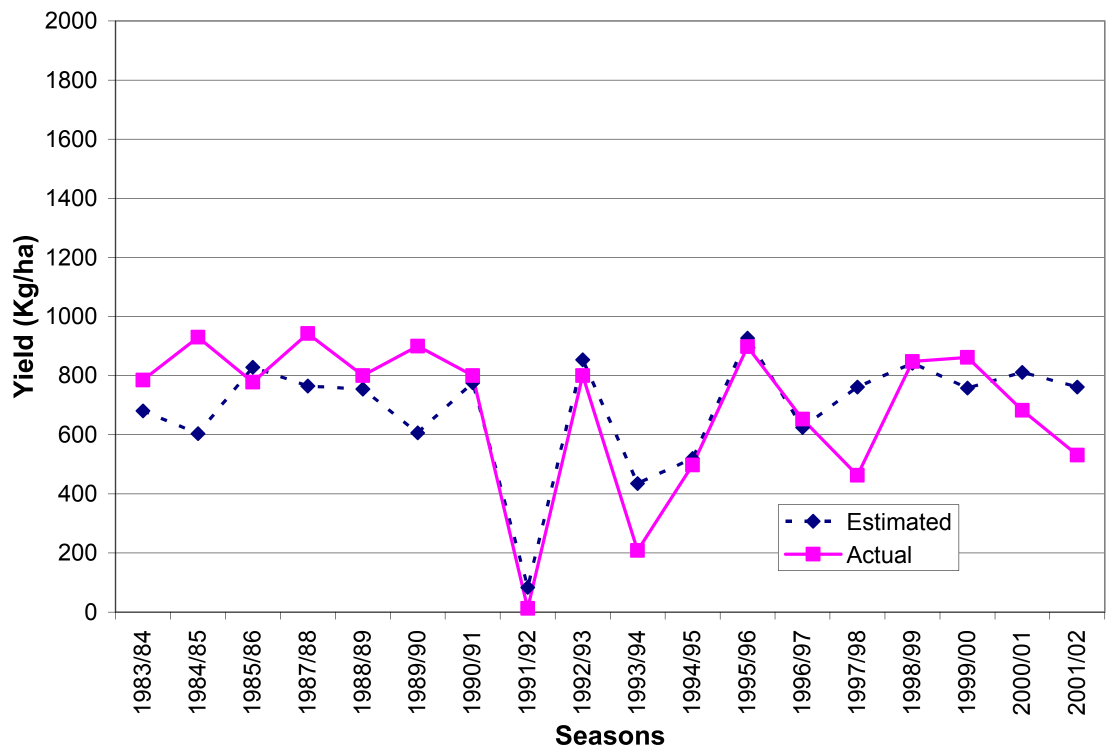
c) Bwanje Valley RDP – Hybrid



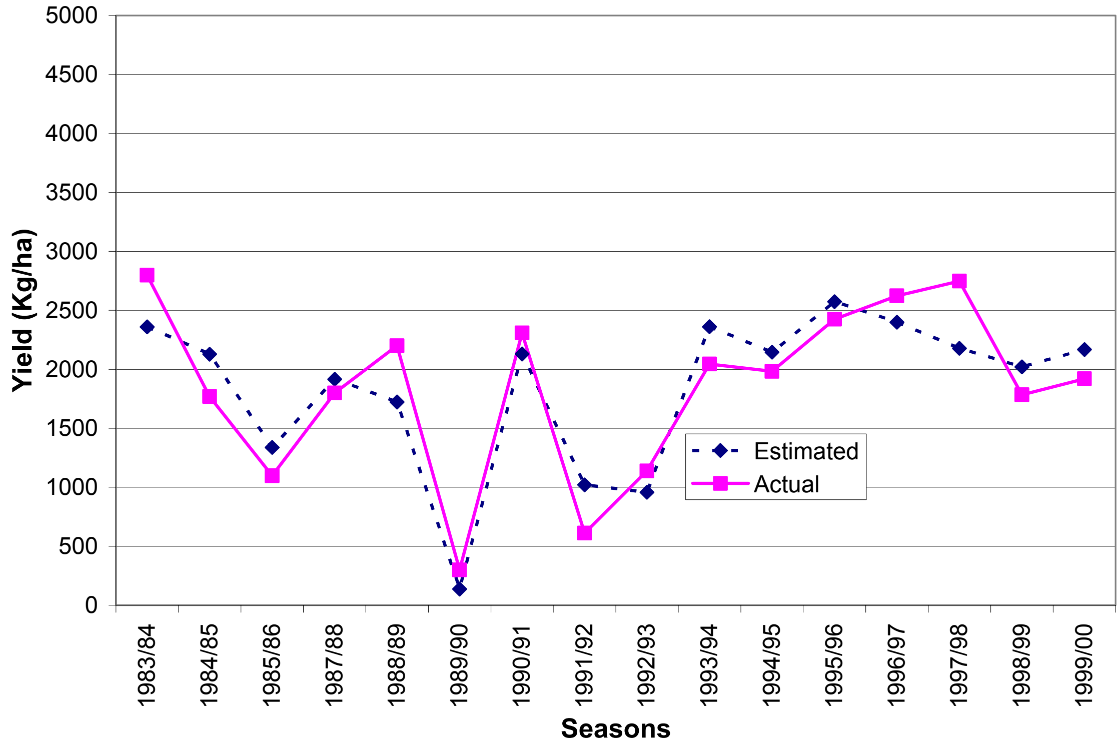
d) Bwanje Valley RDP – Local



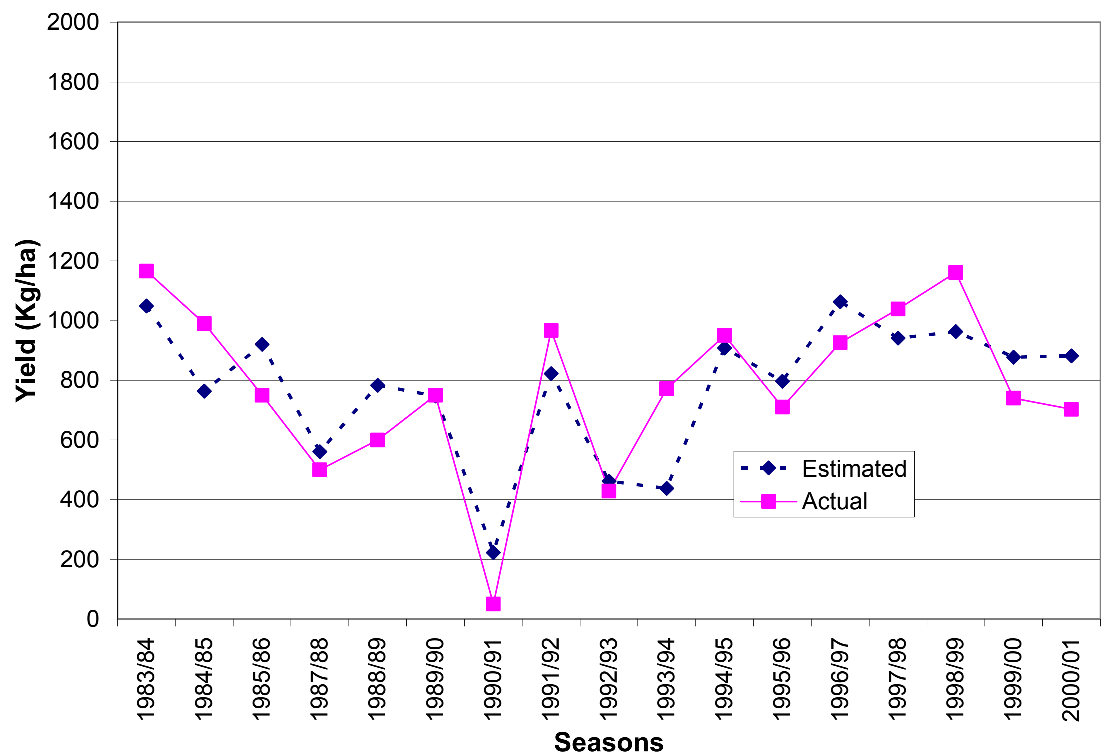
d) Mangochi RDP – Hybrid



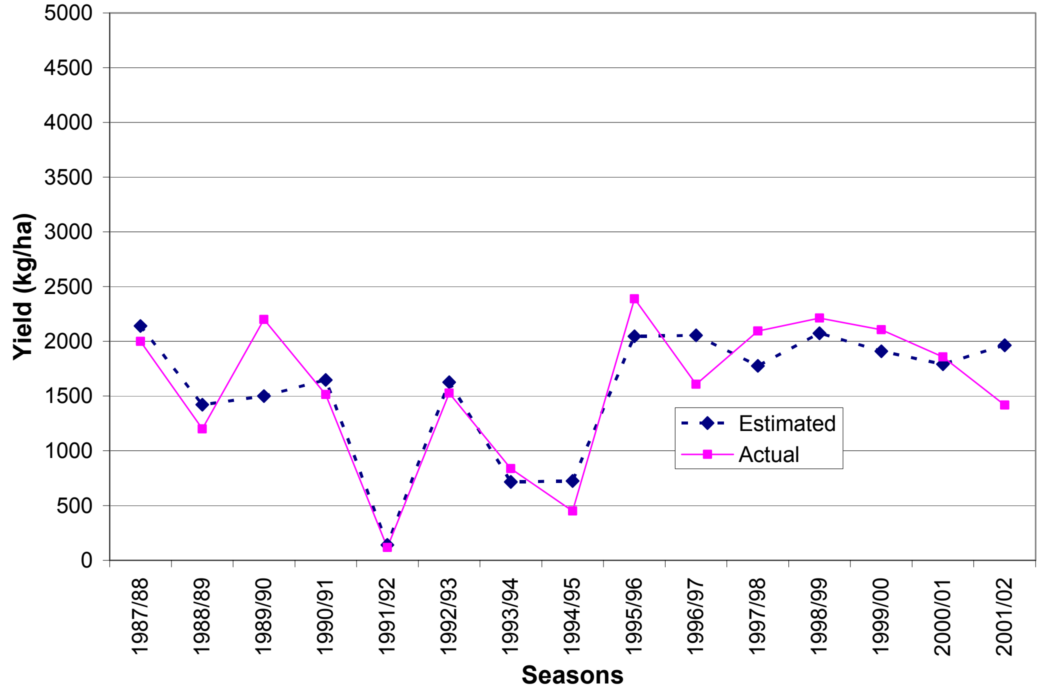
e) Mangochi RDP – Local



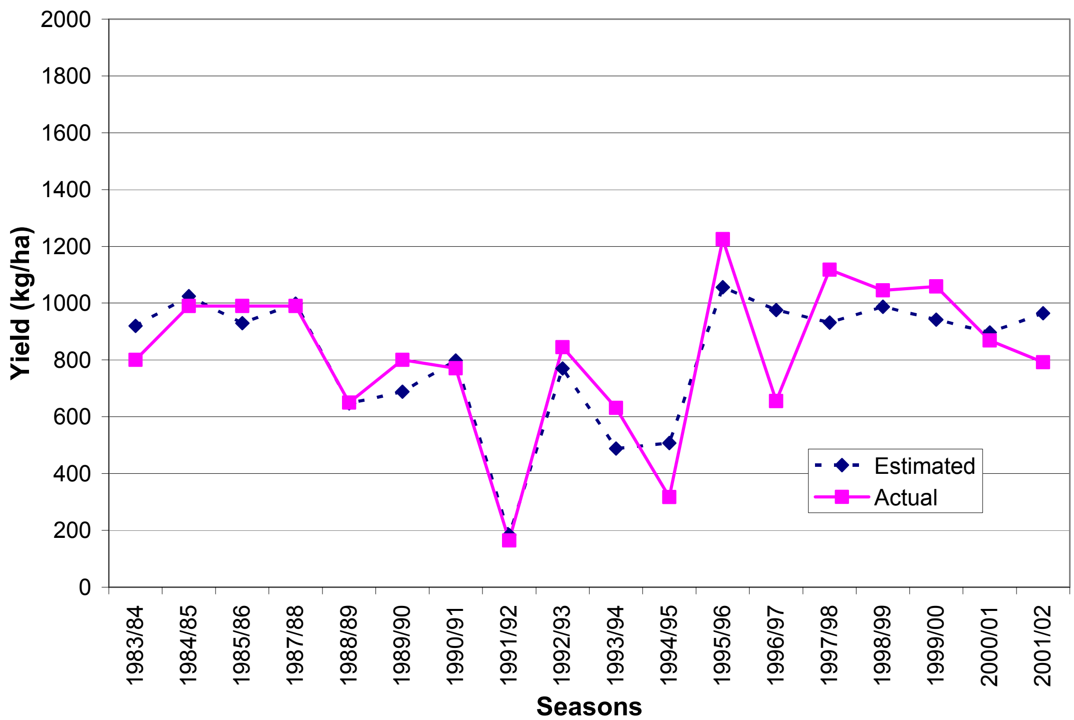
f) Mwanza RDP – Hybrid



g) Mwanza RDP - Local



h) Chikwawa RDP – Hybrid

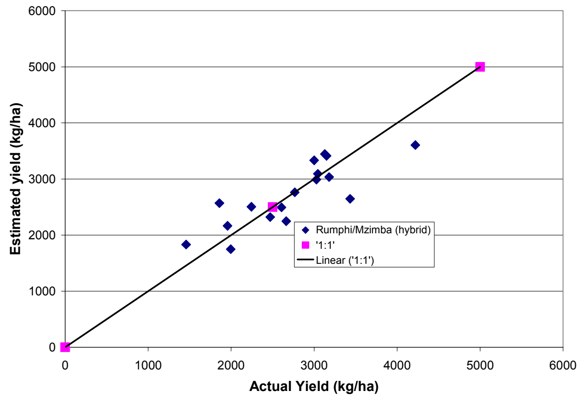


i) Chikwawa RDP – Local

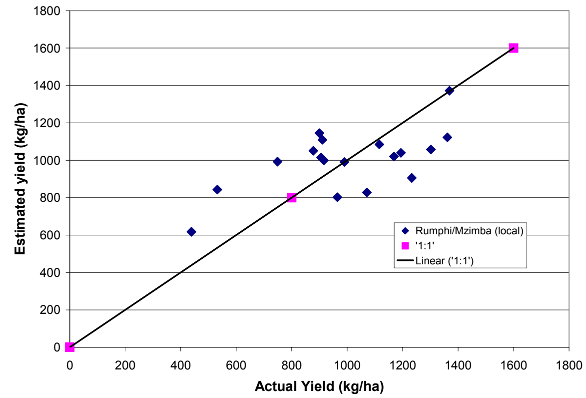
Figure 4.8: Model estimated and actual maize yield for both hybrid and local maize varieties for RDPs representing agroecological zones in Malawi [Rumphi/Mzimba (North); Bwanje Valley (Central); Mangochi (Lakeshore); Mwanza (South) and Chikwawa (Shire Valley)].

Figures 4.8(a) and 4.8(b) show that for Rumphi/Mzimba RDP, the AMS parameter regression model simulates better for hybrid maize compared to the local maize yield. This is in agreement with the values of the r^2 for the two varieties (0.68 and 0.39, respectively). Of all the RDPs in the Northern zone, Rumphi/Mzimba RDP gives the best correspondence between estimated and actual hybrid maize yields. For the local variety, the best correspondence between estimated and observed yield was realized in Mzimba Central RDP (shown in Appendix VI). For RDPs in the Central zone, Bwanje Valley showed the best correspondence between the estimated and observed yields for both hybrid and local maize varieties (figures 4.8c and 4.8d). In the Lakeshore areas, Mangochi was observed to give better correspondence between estimated and observed yields for both hybrid and local varieties and the plots for Mangochi are shown in figures 4.8e and 4.8f for the hybrid and local varieties respectively. Also the plots show that the hybrid models simulate yield better than those for the local variety. Considering the Southern areas, figures 4.8g and 4.8h show the plots for Mwanza RDP that gave both the highest values of r^2 and best correspondence between estimated and observed yields. Similar to the other cases already discussed, it can be seen from these plots that the model for hybrid predicts better than that for the local variety. Chikwawa RDP gave the best correspondence between estimated and observed yield as well as the highest values of r^2 in the Shire Valley zone (figures 4.8i and 4.8j). Another observation from figure 4.8 is that the models simulated better both hybrid and local maize yields during seasons of extremely low crop yields. Therefore, it can be concluded that the AMS-derived models are more accurate in years with low crop production. This aspect makes AMS a suitable tool for early warning, especially for seasons destined for low yields. When the means of the estimated and actual maize yields were tested for significant differences using the Student's t-test for each RDP and variety, it was found that the two sets were not statistically significantly different from each other for all the RDPs and both varieties at the 95% significance level.

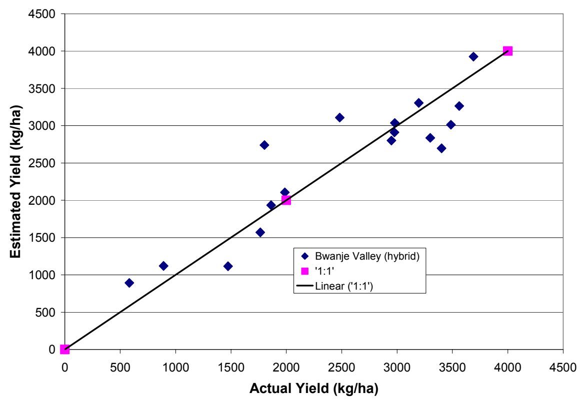
For the same RDPs shown in figure 4.8 above, scatter plots have been given in figure 4.9 (a) to (i).



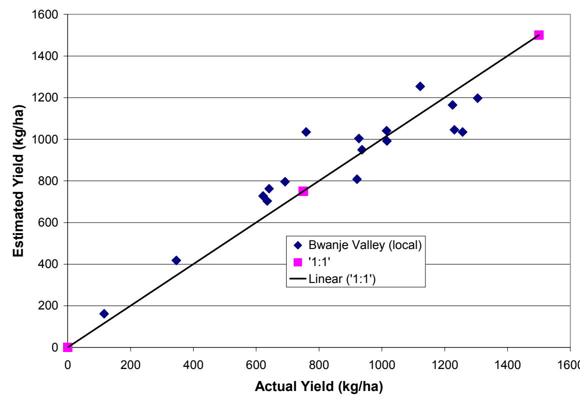
(a) Rumphi Mzimba RDP (hybrid) -North



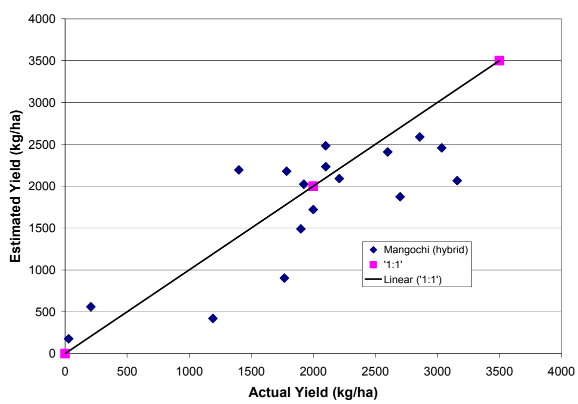
(b) Rumphi/Mzimba RDP (local)



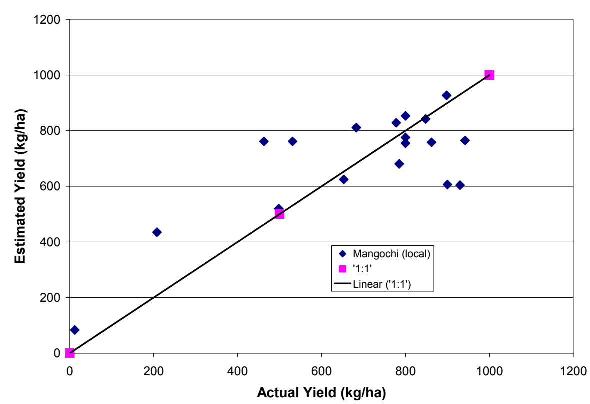
(c) Bwanje Valley RDP (hybrid) -Central



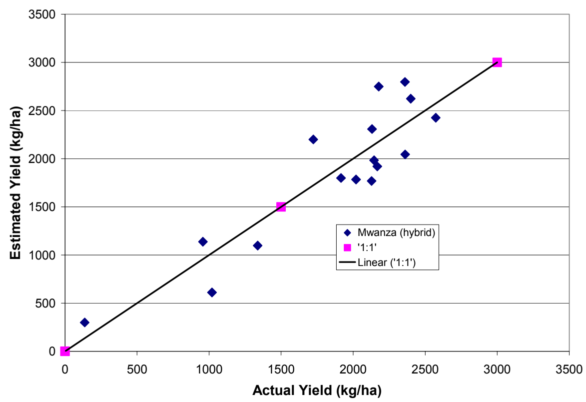
(d) Bwanje Valley RDP (local)



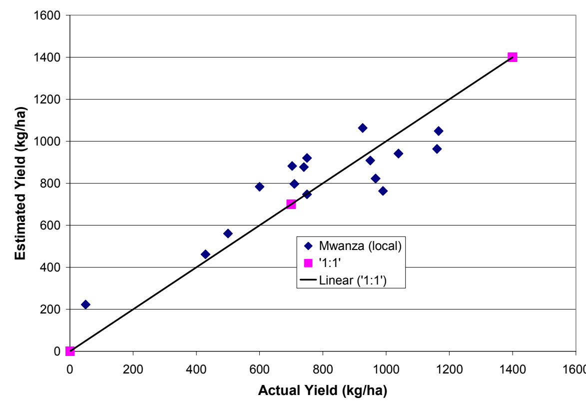
(e) Mangochi RDP (hybrid) -Lakeshore



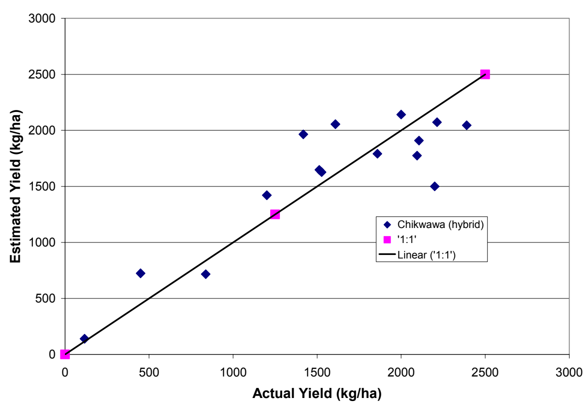
(f) Mangochi RDP (local)



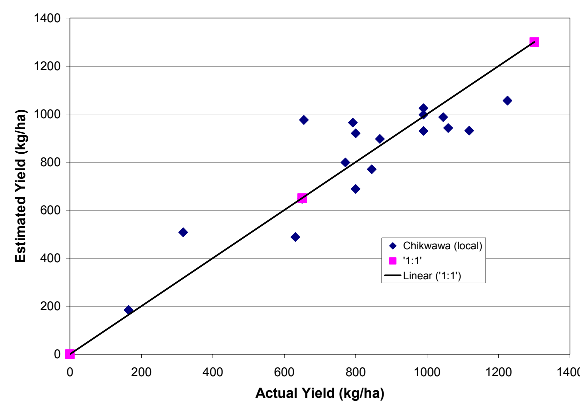
(g) Mwanza RDP (hybrid) -South



(h) Mwanza RDP (local)



(i) Chikwawa RDP (hybrid) -Shire Valley



(j) Chikwawa RDP (local)

Figure 4.9: Scatter plots for the comparison between the model-predicted maize yield and actual yield (NSO post-harvest figures) for selected RDPs in various agroecological zones (as in figure 4.8) for both hybrid and local varieties.

From the scatter plot display, the data points do not significantly deviate from the 1:1 line also shown in the same diagrams. This further shows that the models estimated well the yields for both the hybrid and local maize varieties.

4.4.4 Model testing using independent data (2002/03)

The same regression models in table 4.10 were used to estimate the 2002/03 yields at RDP level. This served to test the models with independent data since this dataset was not used in developing the regression models. However, model testing with the 2002/03 yield data was not done for all the RDPs because some of the RDPs' boundaries were revised from that season onwards, though others were not changed (and these were the ones used in the model testing). Some RDPs were either combined, or parts thereof assigned to others– such RDPs could not be used in this analysis. Because of these changes to the administrative boundaries, the dataset since 2002/03 was different from the historical yield dataset from 1983/84 to 2001/02 for some of the RDPs. Therefore some of these data could not be meaningfully applied in testing the developed regression models. For those RDPs where the models were tested, it was observed that in most of the RDPs the models overestimated yield for both varieties (see table 4.12).

Using percentage errors in table 4.12, for hybrid maize, in eight of the fifteen RDPs the percentage errors were above 10% of the actual yield, thus the models overestimated yield. For the rest of the RDPs the percentage errors were within ±10%, thus the models performed well in such RDPs. As for the local variety, using the same argument as that for the hybrid variety, yield was overestimated in nine of the fifteen RDPs, and there were underestimations in four of fifteen RDPs, while only in two RDPs the percentage errors were within ±10% of actual yield. It can therefore be said that the hybrid variety models performed better than the local variety models for the selected RDPs.

Table 4.12: Estimated, actual and the standard errors from the regression models for the 2002/03 season.

RDP	VARIETY	Y_{est} (kg/ha)	Y_{act} (kg/ha)	$\Delta Y = Y_{est} - Y_{act}$ (kg/ha)	Percentage error of actual yield (%)	SE (kg/ha)
Chitipa	Hybrid	2305	1943	362	19	±414
	Local	783	1074	-264	-25	±203
Karonga	Hybrid	1266	949	317	33	±427
	Local	648	580	68	12	±160
Nkhota kota	Hybrid	2396	1824	572	31	±334
	Local	1067	749	318	42	±316
Ntchisi	Hybrid	2619	2189	430	20	±531
	Local	1264	1076	188	17	±225
Mchinji	Hybrid	2893	2356	537	23	±472
	Local	1255	1073	182	17	±232
Salima	Hybrid	2023	2159	-136	-6	±582
	Local	1102	843	259	31	±268
Mangochi	Hybrid	2119	2218	-99	-4	±543
	Local	730	644	86	13	±162
Dedza Hills	Hybrid	2057	2176	-119	-5	±430
	Local	768	707	61	9	±141
Ntcheu	Hybrid	2856	1740	1116	64	±528
	Local	1297	750	547	73	±139
Balaka	Hybrid	2194	2024	170	8	±704
	Local	734	568	166	29	±211
Zomba	Hybrid	2644	2406	238	10	±707
	Local	798	570	228	40	±178
Phalombe	Hybrid	2196	2138	58	3	±516
	Local	573	780	-207	-27	±246
Mulanje	Hybrid	1816	1443	373	26	±446
	Local	635	607	28	5	±150
Chikwawa	Hybrid	1675	1629	46	3	±312
	Local	527	843	-316	-37	±131
Nsanje	Hybrid	1534	1334	200	15	±254
	Local	466	685	-219	-32	±139

As for the standard errors (SE) of the regression models, for each particular RDP, it was found that those for the hybrid were higher than those for the local variety. This observation may be due to the distribution characteristics of the hybrid yield populations, in particular the average and standard deviation. Both standard deviations and averages were observed to be higher for the hybrid than for the local variety. Therefore using the relationship $SE = \frac{SD}{\sqrt{n}}$, where SD is sample standard deviation, and n is the sample size (Becker, 1999), the higher SE for hybrid variety models can be explained. The highest standard errors were registered in the Southern zone RDPs where for example, Balaka and Zomba had standard errors of ±704 and ±707 kg/ha respectively. Table 4.12 shows the estimated yields, actual yields and the standard errors for the regression models in various RDPs. From this table it can be seen that except for nine out of thirty cases, the model errors fall within the standard error bands. Also considering the percentage errors of actual yield, the same number of RDPs (cases) produced errors larger that 30% (taking into account absolute values of the errors, that is, not considering the signs for the errors). This indicates that most of the models developed for different RDPs performed well in predicting the 2002/03 maize yields. However, in Ntcheu RDP, the model errors were very high for both hybrid and local varieties (64% and 73% of the actual yield, respectively). For more than two thirds of the cases considered the model errors were less than 30% of the actual yields, thus showing that the models predicted with higher accuracy the 2002/03 yield.

4.4.5 Comparison between estimated and actual production for the 2002/03 season

Model estimated maize production (given by: estimated yield {by the models} multiplied by the area planted {as estimated from the first stage of the CES discussed in Section 2.3.1}) and actual maize production (given by: actual yield {from the NSO post-harvest figures} multiplied by the area planted {as estimated from the first stage of the CES}) were compared for all the RDPs for which the 2002/03 yield estimates were made for both varieties. The results are shown in table 4.13.

Table 4.13: Comparison between estimated and actual production for the selected RDPs for the 2002/03 season.

RDP	VARIETY	$YIELD_{est}$ (T/ha)	AREA (ha)	$PROD_{est}$ (T)	$PROD_{act}$ (T)	$PROD_{est}-PROD_{act}$ (T)
Chitipa	Hybrid	2.305	3508	8086	6815	1271
	Local	0.783	11600	9083	12453	-3370
Karonga	Hybrid	1.266	7419	9392	7043	2349
	Local	0.648	8271	5360	4800	560
Nkhota kota	Hybrid	2.396	12224	29289	22294	6995
	Local	1.067	4942	5273	3703	1570
Ntchisi	Hybrid	2.619	7958	20842	17423	3419
	Local	1.264	18317	23153	19717	3436
Mchinji	Hybrid	2.893	25069	72525	68497	4028
	Local	1.255	39062	49023	45136	3887
Salima	Hybrid	2.023	18283	36987	39480	-2493
	Local	1.102	23968	26413	20209	6203
Mangochi	Hybrid	2.119	29103	61669	64545	-2876
	Local	0.730	54468	39762	35071	4691
Dedza Hills	Hybrid	2.057	16978	34924	36941	-2017
	Local	0.768	71256	54724	50393	4332
Ntcheu	Hybrid	2.856	20450	58405	35588	22817
	Local	1.297	47802	61999	35829	26170
Balaka	Hybrid	2.194	21679	47564	43881	3683
	Local	0.734	36998	27157	21027	6130
Zomba	Hybrid	2.644	32980	87199	79359	7840
	Local	0.798	35232	28115	20073	8042
Phalombe	Hybrid	2.196	16472	36173	35225	948
	Local	0.573	19433	11135	15165	-4030
Mulanje	Hybrid	1.816	17010	30890	24539	6351
	Local	0.635	27335	17358	16602	756
Chikwawa	Hybrid	1.675	20159	33766	32831	935
	Local	0.527	50924	26837	31143	-4306
Nsanje	Hybrid	1.534	9260	14205	12352	1853
	Local	0.466	8913	4153	6103	1950
TOTAL	**Hybrid**		**258552**	**581916**	**526813**	**55103**
	Local		**458521**	**389545**	**337424**	**52121**

The above comparison was done in trying to move towards producing nation production estimates (when all the RDPs are used). From the production estimates in table 4.13, in the majority of the RDPs production was overestimated. Similar for the yield estimation results, Ntcheu RDP gave the worst estimates for both varieties. The overestimations (errors) for the hybrid and local varieties were 22817 and 26170 metric tons. This suggests that the model developed for Ntcheu was not the best considering its local conditions. For the other RDPs the errors were considered to be within the acceptable range. Overall, total production was overestimated by about 55103 tons (about 10% of the actual production) for hybrid maize and 52121 tons (about 15% of the actual production) for local maize. Though these statistics are not for the whole country, the AMS models can be said to be performing well for most of the RDPs in the country. As already pointed in earlier sections, the hybrid models perform better than those for the local variety.

4.5 NDVI-yield models compared with AMS-yield models

For the eight RDPs in which both NDVI-yield and AMS-yield models were developed, it was observed that in five RDPs the values of the r^2 were higher in the latter for both hybrid and local varieties whereas in the other three RDPs the AMS approach was advantageous on one of the two varieties. Table 4.14 shows these results.

Table 4.14: Comparison between NDVI-yield models and AMS-yield models (the shaded cells indicate where $r^2 > 0.40$ were obtained.

RDP	NDVI-yield r^2		AMS-yield r^2	
	Hybrid	**Local**	**Hybrid**	**Local**
Salima		0.38	0.53	0.42
Bwanje Valley	0.53	0.54	0.82	0.79
Lilongwe East	0.29	0.43	0.54	0.31
Ntcheu	0.43	0.27	0.38	0.43
Mangochi	0.33	0.35	0.66	0.58
Balaka	0.40	0.54	0.46	0.37
Mulanje	0.23	0.36	0.44	0.47
Nsanje	0.53	0.33	0.70	0.52

Results shown in this table suggest that the use of the AMS models for yield simulation is better than the use of NDVI.

4.6 Overall Discussion

In this study, the potential of using NDVI for yield estimation was investigated. It was found that, though the method can be applicable in some parts of the country, a significant relationship between NDVI and yield was not observed in the majority of the RDPs investigated. Out of all the NDVI parameters that were tested with yield, only dekadal and monthly NDVI values showed a significant and positive relationship with yield especially towards the end of the season (March and April). However, the number of RDPs in which such relationships were observed was just a small proportion of the total number of RDPs tested. The timing of these positive relationships was observed to be useful for yield estimation but not for forecasting. Comparing the results for the hybrid and local varieties, no consistent pattern was observed as to which of the two gave better results. In some cases the results were much better for the local variety and vice versa. Similarly, the RDPs that gave moderate correlations were not the same for each of the parameters tested. There were only a few RDPs namely Mzimba South, Kawinga, Namwera and Mwanza that were consistent in giving results different from the rest of the RDPs. For example in a case where the majority of the RDPs showed a negative relationship between the NDVI parameter and yield, in these four RDPs, the relationship would be positive. This could be attributed to the land cover type and the geography of the RDPs. For example Mzimba South and Mwanza are mainly covered by natural vegetation and are both high altitude and rainfall areas. This lack of consistent relationships was also noticed in the NDVI-rainfall relationships where most of the NDVI parameters either showed lack of significant correlation or were negatively related to rainfall.

The major suspect reason to which the failure of the NDVI-rainfall and NDVI-yield relationships was that a newer version of NDVI dataset that was still untested was used in the study. However, the older NDVI version was tested for one or two analyses but no improvements on the results from such analyses was achieved. Due to limited time not all the analyses were tested. It is expected that if all the analyses were tested a concrete conclusion on the dataset could be reached. Along the same line of reasoning, the yield data was also another suspect reason for the failure of these relationships. As stated in earlier sections, the yield data used in the study is the one that is compiled by the National Statistical Office (NSO) after the third round of CES meetings. This process is not entirely error free, thus rendering the approach non-perfect but is the best available data at present. On the other hand, the problems associated with yield data may not be as pronounced since the same dataset was used with the AMS approach and it was somehow successful, it was also used in earlier studies to develop FAOINDEX regression models (Nayava and Munthali, 1992). Meanwhile the results of this study show that the use of NDVI for yield estimation and forecasting in Malawi is not a good approach.

The second part of the study involved investigating the potential of using AgroMetshell (AMS) output parameters for yield estimation. Regression models were developed for use in maize yield estimation. It was found that the parameters that were prevalent in the models were those towards the end of the growing season, that is, during the ripening stages. This compares well with the NDVI-yield results in which the few useful results were also obtained towards the end of the season. The r^2 values in most of the RDPs were higher for the hybrid variety than the local ones. The goodness of fit of the models was tested using yield data that was used to develop them. In the majority of the RDPs the plots for the estimated and observed coincided very well (figure 4.8, Appendix VI), showing high accuracy in using the models for yield prediction. The models were also observed to be very sensitive during seasons of low yields in each RDP. This property may imply that the AMS could become a very useful tool for yield estimation, especially during years of impending crop failure or low yields. This is an important aspect for early warning functions. The models were also tested with the 2002/03 independent data. But this was not done for all the thirty RDPs in the country since there was a reorganization of RDPs from that season. Only those whose boundaries remained unchanged were used for the validation. Only one season's yield data was available as the 2003/04 yield data compilation was still not complete at the time of the model validation. In most of the RDPs tested the errors of prediction were within the standard error bounds. The most conspicuous RDP where the model estimations were far outside the SE limits was Ntcheu. In this RDP the prediction errors were above 64% of the actual yield. The reason for this discrepancy was not established. In terms of percentage difference between the estimated and actual yield, in the majority of the RDPs for both hybrid and local varieties, the models overestimated the yield. For the hybrid variety, there was no underestimation by the models (using a threshold of 10% of the actual) while for the local variety, the models underestimated in four of the fifteen RDPs.

Since the NDVI-yield models were only developed in eight RDPs, a comparison with AMS-yield results was only done for the same number of RDPs. Overall, the AMS-yield models performed better than the NDVI-yield models. In five of the eight RDPs the r^2 values for AMS models were higher for both hybrid and local maize varieties, while in the remaining three, one variety showed higher r^2 than the other for the two approaches (see table 4.14).

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

The main objective of the study was to investigate the potential use of satellite data and a water balance approach for maize yield maize yield estimation/ forecasting at Rural Development Project (RDP) level in Malawi. In addition some NDVI-rainfall relationships were tested.

NDVI-rainfall relationships were observed to be inconsistent among RDPs and also from one parameter to the other in the same RDP. Generally for most of the NDVI parameters tested with rainfall, the relationships were either very weak or negative. For example, the following observations were made on NDVI-rainfall relationships:

- Maximum NDVI for the season was inversely related to seasonal rainfall for the majority of RDPs.
- Cumulative NDVI, especially in the first half of the growing season, was also inversely related to seasonal rainfall in most RDPs.
- Seasonal NDVI averages were weakly and inversely related to seasonal rainfall except for a few RDPs where stronger relationships were obtained.

Therefore it is suggested that rainfall estimates from Cold Cloud Duration (CCD) must be used in order to relate NDVI to rainfall. This analysis has been suggested because point data, though it was interpolated, is less representative compared to the CCDs which estimates rainfall at the same pixel level as the NDVI.

For NDVI-yield relationships, the following observations were made:

- Moderate positive correlations between dekadal and monthly NDVI towards the end of the rainy season and maize yield but only in few RDPs in the central and southern parts of the country.
- Correlating seasonal NDVI and maize yield resulted in weak but positive correlation coefficients.
- Maximum NDVI was weakly related to maize yield for both varieties.
- The RDPs in which higher correlations were realized were not the same. There was also lack of consistency in the variety that showed higher correlations for each RDP.

It can therefore be concluded that although Rojas (2004) suggested that there was a lot of potential in using NDVI parameters for maize yield prediction in Malawi, this study failed to identify the NDVI parameters that consistently and sigificantly correlate with maize yield and can be used for maize yield forecasting. Other methods must therefore be identified and tested before they can be recommended for use in yield forecasting. These may include the use of not only linear but other types of models like logarithmic, quadratic, exponential and other parameters. It is also suggested that larger areas be used for the analysis since the satellite data used had a coarse resolution. For example, Mkhabela *et al.* (2005) used larger areas and found significant correlations between NDVI and maize yield in three of the four agroecological zones of Swaziland. Also isolating only maize growing areas and extracting NDVI for only such areas might also improve these results. Though not statistically supported, in some parts of the country like the main rice, tea, coffee or tobacco growing areas, the signal captured by the satellite sensors contains a very small proportion from the maize crop. But all these areas were treated as entirely maize growing areas. This might have contributed to the weak relationships between maize yield and NDVI parameters in some RDPs. The use of the MODIS/ SPOT vegetation indices might be important in addressing this problem due to their capability to sample very small areas (finer resolution). The other suspect that may have caused much of the inconsistency in the results is the NDVI dataset that was used. It was a new set that has never been tested, according to the literature survey by the author. It is suspected that the data might have some serious problems with regards to its processing procedure, hence the resultant problems with the outputs. Therefore trying to do the same analyses but using older versions of NDVI data might help to eliminate some of the problems though in this study a few analyses were tried with the older version but not improvements were observed.

For the use of AMS output parameters as yield predictors, there was lack of consistency in the parameters that gave high correlations among the RDPs. For example in neighbouring RDPs where the climatic conditions were assumed to be similar, no two RDPs gave high correlations for similar parameters tested. Therefore the RDPs were grouped according to the agroclimatological zones to which they belong. Different combinations of potential predictors were tested by regression analysis with yield. The following combinations were found to give the best results in each zone: North- WEX_r, WDF_r and AET_r; Central- WEX_f, WDF_r and AET_i; and Lakeshore, Southern and Shire Valley- WSI_s, WDF_r and AET_r (acronyms have same meanings as defined earlier).

Overall, the results for the hybrid maize were better in terms of the percentage of maize yield variation explained compared to the local maize variety. On average, the r^2 value for the hybrid and local varieties are 0.46 and 0.39, respectively. Soil water balance conditions during the ripening stage were found to be critical in the determination of final yield compared to the other stages. The use in the model of parameters derived from conditions at the ripening stage indicates that yield forecasting/ estimation using the AMS-based model can only be done very late as the crop nears the end of its growing season. This is a limitation on the application of the AMS results for early warning purposes. Actual observations and research have shown that water balance (especially water deficit) conditions during the reproductive stages (flowering and grain filling) are critical to the determination of final yield. Since the models also use other parameters representing conditions earlier in the season (for example the initial and flowering phases) and also the WSI_s (which covers the whole life span of crop growth and development), these other parameters can detect whether poor conditions occurred during earlier stages of crop growth. The use of a number of parameters is an advantage of the AMS procedure over the initially used FAOINDEX method in which only one parameter is used for yield estimation and forecasting. For one thing, the FAOINDEX approach was very simple since it only involved one parameter compared to the AMS procedure.

The goodness of fit of the regression models developed was tested with the same data that was used to develop them and the results showed that the models performed well in most RDPs. It was also observed that the models performed much better during seasons with low yields than in the normal and above normal seasons. For these reasons it can be concluded that the AMS is a good tool for yield estimation especially for early warning during low yield seasons, which serves to warn for any need for food security measures.

The models were also used to estimate the 2002/03 yields for both hybrid and local varieties (independent data) for some RDPs. In eight out of the fifteen RDPs for which hybrid yields were simulated the models overestimated whereas in nine of the fifteen RDPs for which local yields were simulated the models overestimated yield. For the majority of the RDPs tested the error of prediction was found to be within the standard error limits. From the yield estimates, maize production at RDP level was estimated and compared with the observed production figures. For all the RDPs whose data was available hybrid and local productions were together overestimated by 10% and 15%, respectively. This shows that the models performed well for the 2002/03 season in most of the RDPs that were used.

It is recommended that the use of non-linear regression models like those suggested for the NDVI-yield relationships (logarithmic, power, exponential and quadratic models) be tested. Further research should be done also using a longer data set to find out if the results can be improved. Further work is also suggested to test the predicting power of the AMS parameters by using datasets cutoff at different points during the season. This would help to identify the most appropriate time to forecast maize yield and their accuracy in terms of the percentages of yield variation explained at those chosen cut-off points in the season.

The use of the AMS for yield estimation/ forecasting in Malawi was found to have more potential than the use of NDVI. Therefore, from this study it is suggested that yield estimation in Malawi can be done successfully using the AMS. As for the use of NDVI for yield prediction, the approach needs further refinement before it can be implemented.

REFERENCES

Allen, R.G., Pereira, L.S., Raes D. and Smith M. (1998): Crop evapotranspiration: guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 56. FAO, Rome.

Balaselvakumar, S. and Saravanan, S. (2005): Remote sensing techniques for agricultural survey GIS development.net (viewed on 21/1/2005). (http://www.gisdevelopment.net/application/agriculture/overview/agrio014pf.htm)

Benacchio, S. (1983): Agroclimate information and agricultural production in the western Llanos of Venezuela, Quoted by: Jackson, I.J. (1989): Climate, Water and Agriculture in the Tropics, Longman Group UK Limited, 2nd Edition, pp. 377.

Becker, L.A. (1999): Confidence Intervals (viewed on 20/4/2005). (http://web.uccs.edu/lbecker/SPSS/confintervals.htm).

Berka, L.M.S., Rudorff, B.F.T. and Shimabukuro, Y.E. (2003): Soybean yield estimation by an agrometeorological model in a GIS, *Sc. Agric.* **60**, 433-440.

Charlier, I. (1999): WinDisp 3.5, User's manual (unpublished).

Doorenbos, J., and Kassam A.H. (1986): Yield Response to Water. FAO Irrigation and Drainage Paper 33. FAO, Rome, pp. 193.

Doorenbos, J., and Pruitt, W.O. (1984): Crop Water Requirements: Guidelines for Predicting Crop Water Requirements. FAO Irrigation and Drainage Paper 24. FAO, Rome, pp. 144.

Dubey, R.P., Ajwani, N., Kalubarme, M.H., Sridhar, V.N., Navalgund, R.R., Mahey, R.K., Sidhu, S.S., Jhorar, O.P., Cheema, S.S., and Narang, R.S. (1994): Pre-harvest wheat yield and production estimation for the Punjab, India. *Int. J. Remote. Sens.* **15**, 2137-2144.

Eldridge, C. (2002): Why was there no famine following the 1992 Southern African drought? *IDS Bulletin,* **33**, 79-87.

Groten, S.M.E. (1993): NDVI-Crop monitoring and early yield assessment of Burkina Faso, *Int.J. Remote Sens.* **14**, 1495-1515.

Hayes, J.M. and Decker, W.L. (1996): Using NOAA AVHRR data to estimate maize production in the United States Maize Belt. Quoted by: Mkhabela, M.S., Mkhabela, M.S. and Mashinini, N.N. (2005): Early maize yield forecasting in four agro-ecological regions of Swaziland using NDVI Data derived from NOAA's-AVHRR. *Agric. For. Meteorol.* **129,** 1-9.

Hearn, A.B. and Wood, R.A. (1964): Irrigation control experiments on dry season crops in Nyasaland, *Emp. J. Exp. Agric.* **32,** 1-17.

Hutchinson, C.F., (1991): Uses of satellite data for famine early warning in Sub-Saharan Africa. *Int. J. Remote Sens.* **12**, 1405-1421.

Jackson, I.J. (1989): Climate, water and agriculture in the tropics, 2nd Edition. Longman Group UK Limited, England.

Ji, L. and Peters, A.J. (2004): A spatial regression procedure for evaluating the relationship between AVHRR-NDVI and climate in the Northern Great Plains, *Int. J. Remote Sens.* **25,** 297-311.

Kanemasu, E.T., Demetriades-Shah, T.H., Su, H. and Lang A.R.G. (1990). Estimating grassland biomass using remotely sensed data. Quoted by: Steven, M.D. and Clark, J.A. (Ed). Applications of Remote Sensing in Agriculture. Butterworths-Reed International P.L.C. Cambridge, Great Britain, pp 247.

Liping D., Rundiquist D.C., and Han L. (1994): Modelling relationships between NDVI and precipitation during vegetative growth cycles. *Int. J. Remote Sens.* **15,** 2121-2136.

Lomas, J. and Herrera, H. (1984): Weather and maize yield relationships in the tropical region of Guanacaste, Costa Rica. Quoted by: Jackson, I.J. (1989): Climate, Water and Agriculture in the Tropics, Longman Group UK Limited, 2nd Edition, pp. 377.

Ma, B.L., Dwyer, L.M., Costa, C., Cober, E.R., and Morrison, M.J. (2001): Early prediction of soybean yield from canopy reflectance measurements. Quoted by: Mkhabela, M.S., Mkhabela, M.S. and Mashinini, N.N. (2005): Early maize yield forecasting in four agro-ecological regions of Swaziland using NDVI Data derived from NOAA's-AVHRR. *Agric. For. Meteorol.* **129,** 1-9.

Maselli, F., Conese, C., Petkov, L. and Gilabert, M. (1993): Use of NOAA-AVHRR NDVI data for environmental monitoring and crop forecasting in the Sahel. Preliminary results. *Int. J. Remote Sens.* **13,** 2743-2749.

Menenti, M., Azzali, S.A., Verhoef, W., and van Swol, R. (1993): Mapping agroecological zones and time lag in vegetation growth by means of Fourier analysis of time series of NDVI images. *Adv. Space Res.* **13,** 233-237.

Ministry of Economic Planning and Development (MEPD). (2002): A Review of Malawian Economy and Key Socio-Economic Indicators. (Viewed on 7/2/05). (http://www.sdnp.org.mw/ict/framework-2002/chap_2.htm)

Mkhabela, M.S., Mkhabela, M.S. and Mashinini, N.N. (2005): Early maize yield forecasting in four agro-ecological regions of Swaziland using NDVI data derived from NOAA's-AVHRR. *Agric. For. Meteorol.* **129,** 1-9.

Mukhala, E. and Hoefsloot, P. (2004): AgroMetShell Manual (unpublished).

Mukhala, E. and Magadzire, T. (2004): Regional Agrometeorological 2004 workshop presentation, Harare.

Mutikani V. (1997): Use of satellite data for biomass and yield prediction in Zimbabwe. (Unpublished MSc. Thesis). Agricultural Meteorology, University of Zimbabwe.

Nayava, J.L. and Munthali, G.K. (1992): The crop specific soil water balance for estimating crop yield in Malawi. Revised Technical Paper No.4, FAO, Lilongwe.

Norman M.J.T., Pearson, C.J. and Searle, P.G.E. (1984): The ecology of tropical food crops, Cambridge University Press, Melbourne, Australia.

Potdar, M.B. (1993): Sorghum yield modelling based on crop growth parameters determined from visible and near-infra red channel NOAA AVHRR data. *Int. J. Remote Sens.* **14,** 895-905.

Raes D. (2001): Water balance of a cropped soil (unpublished draft). Leuven, Belgium.

Rasmussen, M.S. (1992): Assessment of millet yields and production in Northern Burkina Faso using integrated NDVI from the AVHRR. *Int. J. Remote Sens.* **18,** 3431-3442.

Rembold, F. and Nègre, T. (2004): Monitoring agricultural vegetation in Somalia with the VGT vegetation index, AFRICOVER and ECMWF Global Meteorological Modelling – Technical description of the quantitative production forecast for the main crop season in Somalia.

Rojas, O. (2004a): FAO Crop Specific Water Balance (CSWB) model for estimating crop yield in Malawi. Technical Note No.2 (Unpublished).

Rojas, O. (2004b): Vegetation index as crop yield indicator in Malawi. Technical Note, Lilongwe.

Sannier, C.A.D., Taylor J.C., du Plessis W. and Campbell K. (1998): Real-time vegetation monitoring with NOAA-AVHRR in Southern Africa for wildlife management and food security assessment. *Int. J. Remote Sens.* **19**, 621-639.

Singh, R., Sahoo, P.M. and Rai, A. (2005): Use of remote sensing and GIS technology in agricultural surveys. (Viewed 21/1/2005) (hhttp://www.gisdevelopment.net/application/agriculture/cropping/mi0368.htm)

Sridhar, V.N., Dadhwar, V.K., Choudhari, K.N., Sharma, R., Bairagi, G.D. and Sharma, A.K. (1994): Wheat production forecasting for a predominantly unirrigated region in Madhya Pradesh (India). *Int. J. Remote Sens.* **15**, 1307-1316.

Steven, M.D. and Clark, J.A. (Ed). (1990): Applications of remote sensing in agriculture. Butterworths-Reed International P.L.C. Cambridge, Great Britain pp. 247.

Teng, W. (1997): NDVI data from AVHRR land pathfinder (Unpublished). (Viewed 21/1/2005) (http://yang.gmu.edu/~yang/nasacd/www/pal.html)

Townshend, J.R.G., and Justice, C.O. (1986): Analysis of the dynamics of African vegetation using the normalized difference vegetation index. *Int. J. Remote Sens.* **7,** 1435-1445.

Unganai, L.S., and Kogan, F.N. (1998a): Drought monitoring and maize yield estimation in Southern Africa from AVHRR data. *Remote Sens. Environ.* **63**, 219-232.

Unganai, L.S., and Kogan, F.N. (1998b): Southern Africa's recent drought from space. *Adv. Space Res.* **21**, 507-511.

Wan, Z., Wang, P., and Li, X. (2004): Using MODIS land surface temperature and normalized difference vegetation index products for monitoring drought in the Southern Great Plains, USA. *Int. J. Remote Sens.* **25,** 61-72.

WW2010. (Viewed April, 2005) (http://ww2010.atmos.uiuc.edu/(Gh)/guides/mtr/eln/def.rxml)