USING SPATIAL RAINFALL AND PRODUCTS FROM THE MODIS SENSOR TO IMPROVE AN EXISTING MAIZE YIELD ESTIMATION SYSTEM

Celéste Frost

Student number 04-05400/P

Supervisor: Prof. Harold Annegarn

Co-supervisor: Mr Terry Newby

Dissertation submitted
in partial fulfilment of the requirements
for the degree of Master of Science
in the School of Geography, Archaeology and Environmental Studies,
Faculty of Science, University of the Witwatersrand,
Johannesburg

October 2006

UNIVERSITY OF THE WITWATERSRAND

SCHOOL OF GEOGRAPHY, ARCHAEOLOGY AND ENVIRONMENTAL STUDIES

DECLARATION

I, **Celéste Frost** (student number 04-05400/P), am a student registered for MSc in the year 2006.

I hereby declare:

- I am aware that plagiarism (the use of someone else's work without their permission and/or without acknowledging the original source) is wrong.
- I confirm that the work submitted for assessment for the above course is my own unaided work except where I have explicitly indicated otherwise.
- I have followed the required convention in referencing the thoughts and ideas of others.
- I understand that the University of the Witwatersrand may take disciplinary action against me if there is a belief that this is not my own unaided work or that I have failed to acknowledge the source of the ideas or words in my writing.
- This dissertation is being submitted in partial fulfilment of the degree of Master of Science in the School of Geography, Archaeology and Environmental Studies in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other university.

Celéste Frost

13 October 2006

Abstract

After deregulation of the agricultural markets in South Africa in 1997, the estimated maize crop could no longer be verified against the actual crop, due to the lack of control data from the Maize Control Board. This drove the need to explore remotely sensed data as a supplement to the current crop estimation methodology to improve crop estimations.

Input data for the development of a Geographic Information System (GIS)-based model consisted of objective yield point data, Moderate Resolution Imaging Spectroradiometer (MODIS) Normalised Difference Vegetation Index (NDVI) images and rainfall grids. Rainfall grids were interpolated from weather station data. NDVI values were obtained from the MODIS sensor aboard the Terra platform. Objective yield point field survey data for the 2001/2002 growing season were utilised since dry-land or irrigated conditions were recorded for that season.

MODIS NDVI values corresponded well with the growing stages and age of the maize plants after being adjusted to reflect the crop's age rather than the Julian date. Rainfall values were extracted from rainfall grids and also aligned with the age of the maize plants. This is a suggested alternative to the traditional method of using the mean NDVI for several districts in a region over a Julian growing period of 11 months according to Julian dates. South African maize production areas extend over seven (7) provinces with eight (8) different temperature and rainfall zones (du Plessis, 2004).

Planting-date zones based on the uniform age of the maize plants were developed from objective yield Global Positioning System (GPS) points for the 2001/2002 growing season and compared with the 2004/2005 growing season (Frost and Kneen, 2006). Planting dates were interpolated from these planting zones for objective yield GPS points which were missing planting dates in the survey database. MODIS imagery is affordable (free) and four (4) images cover the whole of South Africa daily, while one (1) image covers the study area daily. Several recommendations, such as establishing yield equations for a normal, dry, and wet season were made. It is also suggested that dry-land and irrigated areas continue to be evaluated separately in future.

Keywords: MODIS, maize crop yield estimation, GIS, remote sensing.

Acknowledgements

I wish to thank the following persons and institutions for their assistance:

- Mr Terry Newby, Programme Manager, Agricultural Research Council's Institute for Soil, Climate and Water (ARC-ISCW) in Pretoria, for the opportunity to do this project as part of the ARC-ISCW Crop Consortium initiative.
- NASA and USGS for the MODIS products.
- Mr Deon du Toit, ARC Grain Crops Institute (ARC-GCI), Potchefstroom, for providing the objective yield figures.
- Mr Johan Malherbe, Junior Researcher, ARC-ISCW, Agromet Division, Pretoria, for preparation of the rainfall grids and technical assistance.
- Ms Talita Germishuyse, Researcher, ARC-ISCW, Geo-Info Division, Pretoria.
- Mrs Melanie Kneen, Research Assistant, University of Johannesburg.
- My supervisor**,** Professor Harold Annegarn, of the Department of Geography and Environmental Management, University of Johannesburg, for his encouragement and guidance.
- Mrs Jeannette Menasce for proof-reading and correcting the text.
- Almighty God, who gave me the time, energy and resources to complete this work to glorify Him.

"Challenges:

New challenges will arise. New questions will be asked.

New solutions must be sought."

(John Maxwell, 2005)

CONTENTS

List of Tables

List of Figures

Abbreviations, Acronyms and Glossary

ADDS	African Data Dissemination Service
AOI	Area of Interest
Aqua	Polar orbiting satellite
ARC	Agricultural Research Council
ARC-GCI	Agricultural Research Council-Grain Crops Institute, Potchefstroom, South Africa
ARC-ISCW	Agricultural Research Council-Institute for Soil, Climate and Water, Pretoria, South Africa
AVHRR	Advanced Very High Resolution Radiometer
BT	Brightness Temperature
CC	Crop Consortium consisting of the Agricultural Research Council, GeoTerraImage (Pty) Ltd, AgrImage (Pty) Ltd and SpatialIntel (Pty) Ltd (formerly Agrista)
CEC	Crop Estimates Committee This body releases estimates of maize area planted and expected maize yields leading to expected total maize production estimates
Cereal pixels	Pixels which are covered by cereal (grain) crops over not less than 60 % of the pixel area
CROPMON	Operational Crop Monitoring and Production Forecast Program
CSIR-SAC	Council for Scientific Research-Satellite Application Centre
Decade	A period of ten years
Dekad	The cycle of each crop is subdivided into successive 10-day periods (dekads) taken ... as is the case - by definition - in most semi-arid areas of the world (www.fao.org/sd/EIdirect/AGROMET/model.htm). Using this definition, there are three dekads in a month.
DN	Digital Numbers
DoA	Department of Agriculture (formerly known as the National Department of Agriculture)
EDC	EROS Data Center of the United States Geological Survey (USGS)
EMR	Electro Magnetic Radiation
EOS	Earth Observation System
EOS-AM1	Earth Observation System AM1

EROS	Earth Resources Observation and Science
ERS-1SAR	European Remote Sensing satellite – 1 Synthetic Aperture Radar
ESA	European Space Agency
EVI	Enhanced Vegetation Index
FPAR	Fraction of Photosynthetically Available Radiation is an essential parameter relating the available visible solar radiation to its absorption by chlorophyll for plant photosynthesis. It is one of the surface parameters quantifying the CO_2 uptake by plants and release of water through evapotranspiration (as defined in http://www.wmo.ch/web/gcos/terre/variable/radfra.html).
FPAR	Fraction of Photosynthetically Active Radiation between 400 and 700 nm used by the green canopy in the photosynthetic process (as defined in http://ccrs.nrcan.gc.ca/glossary/index_e.php?id=2083).
GIS	Geographic Information System
GLAI	Green Leaf Area Index
GMFS	Global Monitoring for Food Security
GN	Greenness Value
GPS	Global Positioning System
GSE	Global Monitoring for Environment and Security (GMES) Services Element
GYURI	General Yield Unified Reference Index
H/E	The soil moisture index as the ratio of sensible heat to latent heat (Dabrowska-Zielinska *et al*., 1998:140)
IDW	Inverse Distance Weighted
IRS-1C/1D	Indian Remote Sensing satellites 1C and the identical IRS-1D
Kriging	A statistical interpolation method
LAI	Leaf Area Index
Landsat	NASA's Landsat satellite
Landsat TM	NASA's Landsat Thematic Mapper
LST	Land Surface Temperature
MD	Magisterial District
METEOSAT	European weather observation satellite
MIR	Mid Infra-Red part of the electromagnetic spectrum
MODIS	Moderate Resolution Imaging Spectroradiometer
NASA	National Aeronautics and Space Administration, USA

NCEC	National Crop Estimates Consortium. This body reports to the Crop Estimates Committee (CEC) which releases maize production figures. The NCEC comprises: SpatialIntel, GeoTerraImage, ARC-ISCW, ARC-SGI (the Small Grain Crops Institute, based in Bethlehem, Free State), ARC-GCI (the Grain Crops Institute, based in Potchefstroom, North-West Province)
NDA	National Department of Agriculture (now known as the Department of Agriculture (DoA))
NDVI	Normalised Difference Vegetation Index. Result of (NIR-Red)/(NIR+Red).
NIR	Near Infra-Red part of the electromagnetic spectrum
NLC	National Land Cover
NOAA	National Oceanic and Atmospheric Administration
NPOESS	National Polar-orbiting Operational Environmental Satellite System
NPP	Net Primary Production
PAR	Photosynthetically Available Radiation
PECAD	Production Estimates and Crop assessment Division of the Foreign Agricultural Service of the U.S. Department of agriculture
PEOI	Polar Earth Orbit Inclination
R^2	Correlation coefficient
RRSU	Regional Remote Sensing Unit
SA	South Africa(n)
SADC	Southern African Development Community
SAFEX	South African Futures Exchange
SAGIS	South African Grain Information Service
SA-NLC	South African National Land-Cover (that is, Thompson's (1995) database)
SAR	Synthetic Aperture Radar
SAWS	South African Weather Service
SPOT4/ VEGETATION	The French Satellite Pour l'Observation de la Terre. SPOT4 was launched on 24 March 1998. The VEGETATION instrument onboard SPOT4 is funded by the European Union, Belgium, France, Italy and Sweden and led by French space agency CNES.
SRP	Spectral Response Pattern
TCI	Temperature Condition Index
Terra	Polar orbiting satellite
TIR	Thermal Infra-Red region of the spectrum (~8 to 14 µm)
TM	Thematic Mapper (of Landsat)

Ton	A metric unit of mass equal to 1000 kg
USGS	United States Geological Survey
VCI	Vegetation Condition Index
VIS	Visible part of the electromagnetic spectrum
WDI	Water Stress Index (as described by Dabrowska-Zielinska *et al*., 1998:140)
WRSI	Water Requirement Satisfaction Index

1 INTRODUCTION

1.1 Background

There is a need to supplement and improve the maize production estimation system currently in use by the Crop Estimates Committee (CEC) in South Africa. The Maize Control Board ceased to exist in 1997. Prior to 1997, maize crop estimate data were tested against the size of the full maize crop produced the previous year, which data were obtained from the former Maize Control Board. This resulted in a more accurate figure than is currently generated.

As recently as 2005 there was a lack of check-data, as South African Grain Information System (SAGIS) deliveries only become available five (5) to six (6) months after the start of the marketing year and these figures are available only on a provincial level. This lack of check-data, together with the need to improve current crop yield estimates, has driven the need to explore remotely sensed data as a supplement to the current crop yield estimation methods.

The Department of Agriculture (DoA) and the National Crop Estimates Consortium (NCEC) supply estimate data to the Crop Estimates Committee (CEC). The DoA uses a non-probabilistic co-operator (farmer) survey to obtain their information. The DoA sends questionnaires to a non-probable sample of co-operators, resulting in a biased sample (as farmer participation and completion is voluntary). The estimated production by the respondents in the survey is taken to arrive at an average yield for all the respondents in a province. Field and farm information are derived from the subjective yield data. This information is supplied to the CEC at their monthly meetings.

In contrast, the NCEC uses a point frame sampling system to obtain their data (resulting in an unbiased sample). The NCEC uses both subjective surveys and objective measurement in their system. A schematic showing how the CEC and NCEC operate is given in Appendix 1.

1.1.1 Structure of the research report

In order to take the reader on a journey of systematically testing the **hypothesis (stated in section 1.5.2)**, this research report is structured into **six (6) chapters** and **an Appendix (1)**.

Chapter 1 explains some of the current maize yield estimation methods used by the NCEC members **(section 1.1)**. The importance of timely and accurate maize yield **estimates (section 1.2)** and the role of the research done in this dissertation to supplement the current methods are addressed **(section 1.3)**. The **advantages** of an improved maize yield estimation system are discussed in **section1.4**.

Section 1.5 discusses the importance of developing a **remotely sensed data** based system in conjunction with the current systems. **Section 1.5.1** states the **objectives** needed to achieve the **aim**, which is to find out whether MODIS 16-day NDVI composite data and spatial rainfall data can be used to **supplement** the currently used (objective yield) maize yield estimation system. This would lead to the verification and improvement of the currently used objective yield system. The final section **(section 1.6)** of the introductory chapter **(Chapter one (1))** is used to familiarise the reader with the **study area** for which the MODIS imagery and the objective yield points were acquired.

Chapter 2 gives a **literature overview** of some of the methods involving remotely sensed data and crop yield estimation. This review first explores methodologies used around the world, then in Africa, and lastly in South Africa, up to the year 2004.

Chapter 3 explains how the MODIS 16-day NDVI, spatial rainfall and objective yield data were **collected** and **extracted** to **prepare** for the **analysis**, which takes place in **Chapter 4**.

The development and discussion of three (3) types of **products** take place in **Chapter 4**; namely planting date zone maps, averaged MODIS NDVI curves and averaged spatial rainfall curves.

Section 4.1.2, the **Developmental stages of the maize plant**, was included specifically because it is vital to **interpretation** and **understanding** as well as the development of planting date maps. Without understanding each growth stage and the critical factors that occur within each associated rainfall or NDVI dekad, which relate to identification of stress factors that in its turn can directly be related to yield, the importance of this research will only be partly understood.

Sections 4.3.1 to 4.3.3 and sections 4.4.1 to 4.4.3 include discussions on each of the small (less than 3 ton/ha), medium (from 3ton/ha to 4 ton/ha) and large (above 4 ton/ha) yield **standard deviations**. This was done to point out during which part of the growing season there is greater or lesser overlap between the averaged NDVI and rainfall yield classes, and is therefore an optimal or less optimal dekad in the growing season to use the average NDVI and/or rainfall graphs for maize yield predictions.

Sections 4.2, 4.3 and 4.4 of **Chapter 4**, **Results and Discussions,** are devoted to **average NDVI** and **average rainfall graphs** of the three yield classes (low-, medium- and high yield). The averages are also compared with the averages of the **test data points** (which consisted of 1/3 of the data points) to illustrate the difference that sample size makes.

Of importance in **Chapter 4,** is the **start** and **end dekads** during which the average values of **rainfall** and **NDVI** can be used for **yield prediction**. This includes the identification of **window periods**. These are periods during the growing season in which average NDVI and/or rainfall values can be used to predict low, medium or high yields and/or in ton/ha. Some window periods can be used for predicting whether end yield will fall in the low, medium or high end yield class, while another type of window period can only be used to predict whether the end yields are going to be high or not (medium/low), or low or not (medium/high).

Three dekadal moving average graphs were statistically calculated to simulate the effects to be expected if monthly (three dekadal) instead of dekadal data were used due to data, time, or budget constraints.

Chapter 5 contains the **Conclusions**, which were compiled from **summarising** and explaining the reasons for the phenomena in **Chapter 4, Results and Discussion.**

The **Recommendations** made in **Chapter 6** were made from information that came to light during the research done for this dissertation. The reason why NDVI values are so widely used with such promising success rates, is that an **NDVI** value **inherently is a summary of many factors** such as the soil properties, rainfall, temperature and management practises. If it is too costly or impossible to obtain all of these values, especially on a large scale like in South Africa, the NDVI is a cost effective alternative.

For further research to develop a more complete **maize yield estimation** equation, other products from the MODIS sensor should also be considered for inclusion together with NDVI, i.e. LAI, FPAR, LST and NPP, over a **homogenous planting date zone**, i.e. for plants that are of the same age.

1.1.2 Subjective yield surveys

A point frame system is used and the points to be surveyed are selected on a stratified random basis per province per stratum. The strata used are derived according to cultivation and usage patterns. The stratification is done using the South African National Land-cover (NLC) database (Thompson, 1995) dataset and overlaying it with more current Landsat imagery. Four (4) stratification classes were distinguished for the 2001/2002 growing season.

More points are allocated to strata with the highest probability of containing grain crops than to strata with a smaller probability of containing grain crops.

All farms have an equal probability of being chosen on the basis of probability being directly proportional to farm size. Sample size per stratum is determined by factors such as budget and time constraints. Statistical methods of sample size are revised each year. Every year 20 % of the points are changed to keep the experimental design statistically sound.

After the strata and the number of points per stratum have been determined, a random sample is taken from a (225 m x 225 m) grid of points, covering the whole of South Africa. The points within this random sample are then visited. The number of points that were found to be located on a grain farm are recorded. These points are referred to as "hits". "Subjective yield" surveys are undertaken at these hit points, while "objective yield" surveys are undertaken using a random sample taken from the hit points.

The statistical design is still in the experimental phase and optimal methods of determining subjective and objective yield figures have yet to be established.

At each hit point the farmer is interviewed according to a set questionnaire (ARC-GCI, 2002; 2004). Data concerning expected yields and areas planted to maize for the farm are collected. These figures refer to what the farmer expects to harvest as at the date of the survey. The data obtained for the farm are used to obtain an estimate for planted area, harvested area and yield for the province.

1.1.3 Objective yield surveys

From the hit points, a sub-sample is taken, from which a second survey is done, called an "objective yield survey". A different questionnaire is used for this survey (ARC-GCI, 2002).

Field workers generated five (5) new Global Positioning System (GPS) points according to a statistical method, in a maize field on the chosen sub-sample of farms containing the hit points. Field workers visit these GPS points three (3) times during, and once at the end of, the growing season.

On the first three visits, field workers determine the number of plants per row, the number of rows per hectare, measure the height of the plants, count the cobs and assess the number of kernels per cob. Plant density is calculated. The measurements and counts obtained from these surveys are used in formulae to derive a predicted objective yield figure for each of the GPS points created in the maize field. On the last (fourth) visit, the field worker asks the farmer what yield was achieved from that maize field at harvest.

1.2 Problem Statement

The agricultural sector requires "timely, reliable, accurate and independent" estimates (NCEC Proposal, 2002) of crop production if it is to be internationally competitive and be in a position to manage national food security.

Traditional methods of crop estimation in South Africa are perceived as being biased, scientifically unsound and unreliable and no longer provide acceptable figures (NCEC Proposal, 2002). The removal of the Maize Control Board structures, the conversion of co-operatives into profit-motivated companies and producers having the freedom to sell products on an open, deregulated market have contributed to these perceptions.

According to the NCEC Proposal (2002), the NCEC would like to achieve improvements to the current crop estimation methods to improve the accuracy of the crop estimates to within a 10 % variance.

Subjective yield surveys are dependent on accurate farmer responses and the willingness of farmers to respond to surveys. Both the DoA and NCEC systems are designed for collecting provincial statistics. Statistics for districts are not regarded as reliable (Beukes, 2004a: Personal communication).

1.3 Justification and Importance

Valuable in-field data at GPS points are collected during the existing objective yield efforts, which can be used as ground-truthing information. It is important to investigate whether remote-sensing imagery could provide a source of check-data for the existing objective yield survey data efforts.

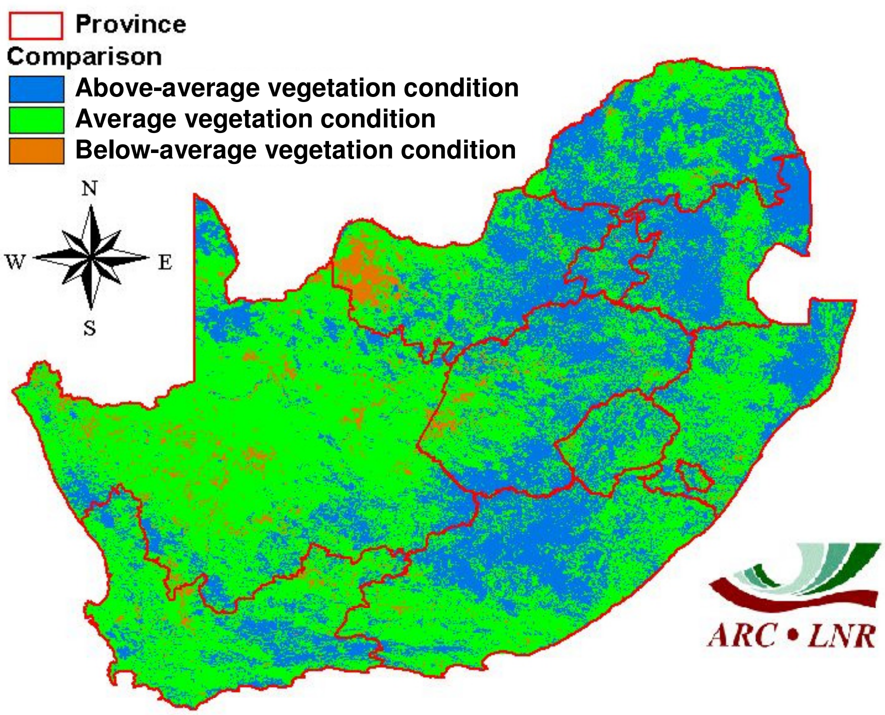
Currently, long-term average Normalised Difference Vegetation Index (NDVI) images, created from 17 years of National Oceanic and Atmospheric Administration Advanced Very High Resolution Radiometer (NOAA-AVHRR) data are used, from which a vegetation condition map is produced. Every 10 days a dekadal* NOAA-AVHRR NDVI image is used to produce a difference map from this long-term average image. The legend of the difference map contains three sections: average vegetation condition, above-average vegetation condition and below-average vegetation condition. This difference map (Figure 1) is used at monthly CEC meetings to inform decisions about anticipated maize yields.

The Umlindi Newsletter (Umlindi Report 04051, 2004) primarily consists of reports generated for the DoA, CEC and National Agro-meteorology Committee. The Umlindi system attempts to inform decision-makers of the current vegetation condition based on interpreted NOAA NDVI- and climate data.

A NOAA-AVHRR NDVI image from 21 to 30 April 2004 was compared to the long-term average for 21 to 30 April of 1985 to 2003 to produce a map (Figure 1) indicating change in vegetation conditions as compared to the long term.

* The cycle of each crop is subdivided into successive 10-day periods (dekads) taken ... as is the case - by definition - in most semi-arid areas of the world (www.fao.org/sd/EIdirect/AGROMET/model.htm).

The resulting difference map (Figure 1) shows healthy (average) vegetation conditions for most parts of South Africa. The eastern part of the country shows widespread above-average vegetation conditions mixed with average vegetation conditions. The western part of the country reflects average vegetation growth with isolated patches of above-average vegetation conditions. Patches of below-average vegetation conditions were experienced mostly in the North West Province and the Northern Cape Province, dominated, however, by the average vegetation conditions (Umlindi Report 04051, 2004).



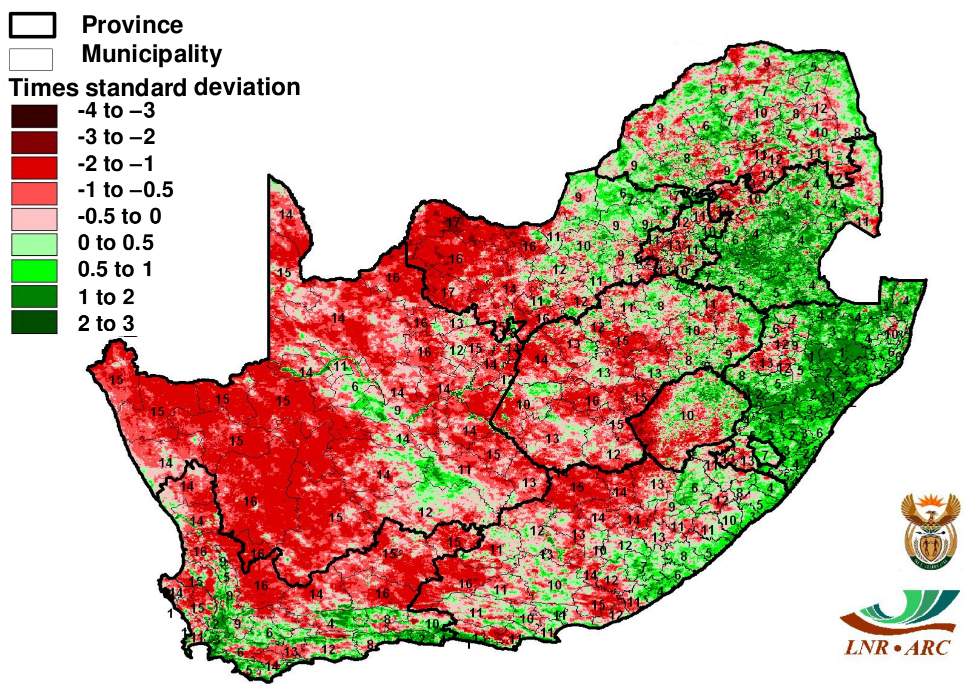
(Source: Umlindi Report 04051, 2004)

Figure 1 NDVI difference map for 21 to 30 April 2004 compared to the long-term average for 21 to 30 April of 1985 to 2003

The values of the NDVI for all the years of available data are collected per district municipality. The value for a particular year of interest is compared with the other years and ranked according to its position as to whether it is better or worse than other years (Figure 2). A value of one (1) would indicate that, for the municipality as a whole, the NDVI for the current period is the highest when compared to the same periods over other years. A higher ranking, like 17, means that it was one of the seasons with a lower NDVI value – the 17th year when ranked from high to low NDVI.

The DoA requires specific rainfall per magisterial district to assist in interpreting a map such as the one in Figure 2. It is, however, suggested from this research that specific rainfall per planting zone is required for yield prediction. This information can be related to magisterial district level at any later stage.

For the period from 11 December 2003 to 30 April 2004 the NDVIs were accumulated and compared with the accumulated NDVI values over the same period for all the other years (Figure 2). The standard deviation per pixel was calculated and this period's deviation from the mean was expressed as a multiple of the standard deviation of the NDVI for the pixel.

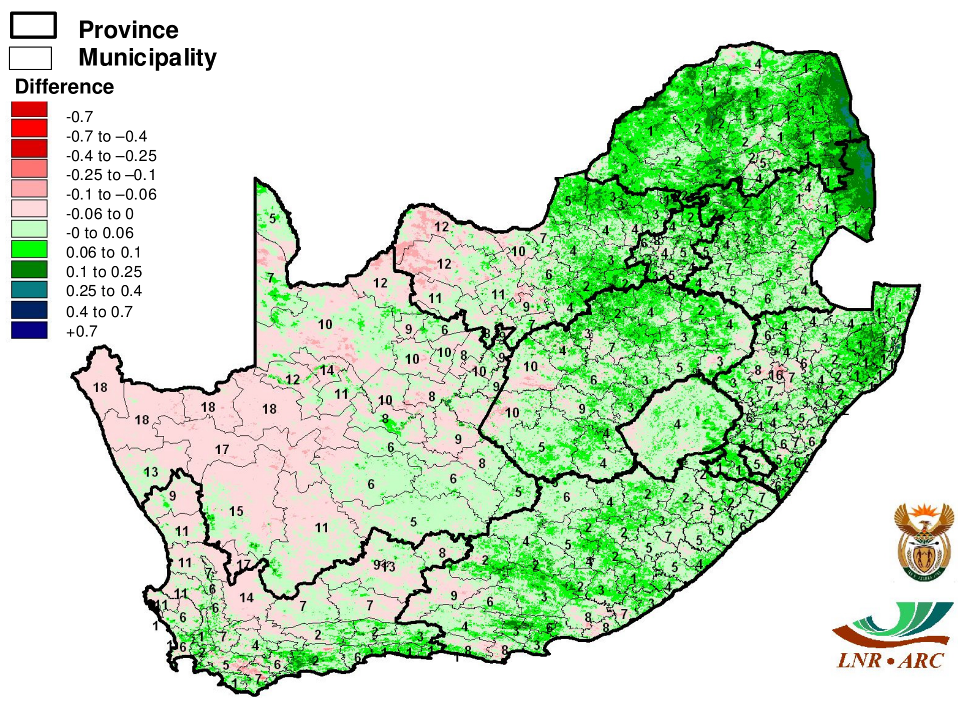


(Source: Umlindi Report 04051, 2004)

Figure 2 NDVI for 11 December 2003 to 30 April 2004 compared to the long-term mean

The long term NDVI period used in Figure 2 was from 1985 to 2004, excluding 1994 and 1995. One (1) is the highest NDVI in the 17 year period, while seventeen (17) refers to the lowest NDVI over the same period.

Figure 3 is an expression of the difference in NDVI value for the period 21 to 30 April 2004, and the long term mean NDVI for the same period (21 to 30 April). The long term NDVI period used in Figure 3 was from 1985 to 2004, excluding 1994. One (1) refers to the highest NDVI in the 18 year period, while eighteen (18) refers to the lowest NDVI over the same period (Umlindi Report 04051, 2004).



(Source: Umlindi Report 04051, 2004)

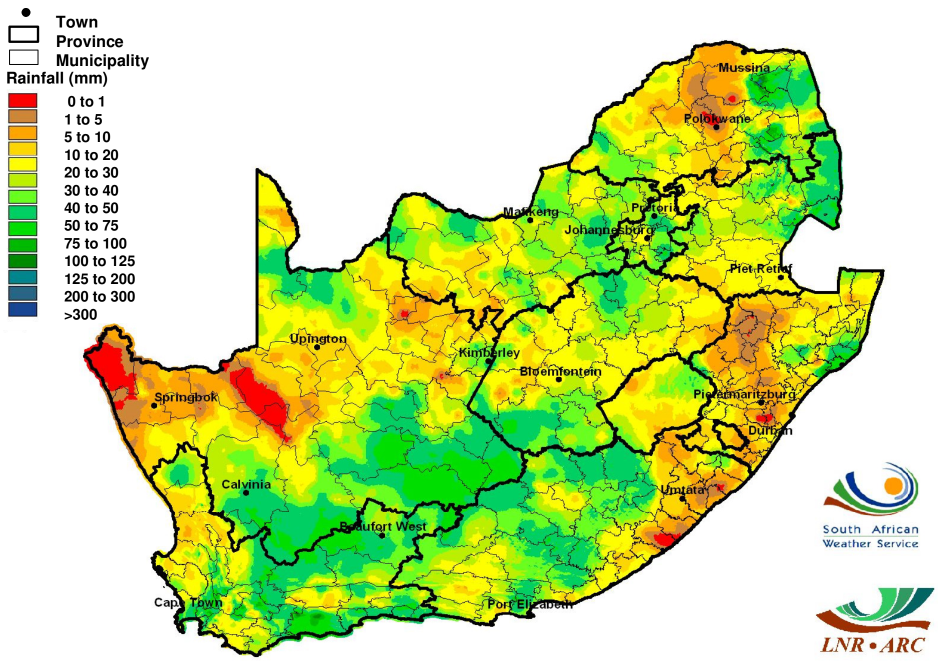
Figure 3 NDVI for 21 to 30 April 2004 compared to the long-term mean

The monthly Umlindi Report also contains rainfall reports, which are compiled by interpreting the current monthly rainfall map (Figure 4). The monthly rainfall map is created by combining rainfall station data with satellite rainfall estimates from the African Data Dissemination Service (ADDS) and long-term trend surfaces.

The monthly rainfall map (Figure 4) expresses rainfall in mm. Figure 5 expresses rainfall as percentage of the long-term mean for the month of interest.

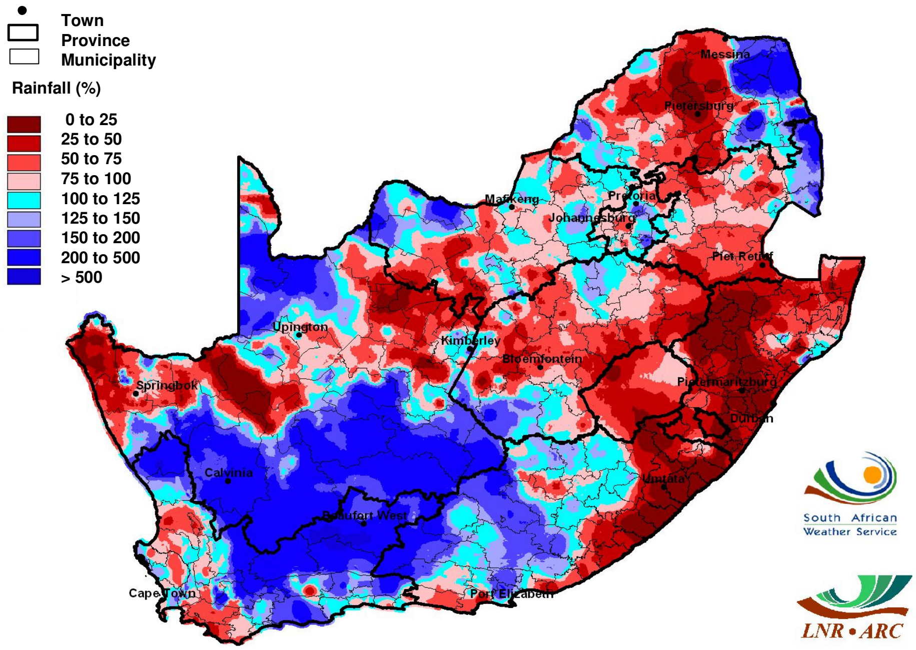
From the rainfall maps and discussions of the rainfall per province, decision-makers try to form an idea of the rainfall conditions that prevailed in a certain period. But, as can be seen from these reports, they do not know at which critical growing stage the maize was during this period for the different regions.

The function of the Umlindi Report is to discuss the monthly rainfall and percentage of the mean rainfall per province, and sometimes for certain regions within the province.



(Source: Umlindi Report 04051, 2004)

Figure 4 Rainfall in mm for April 2004

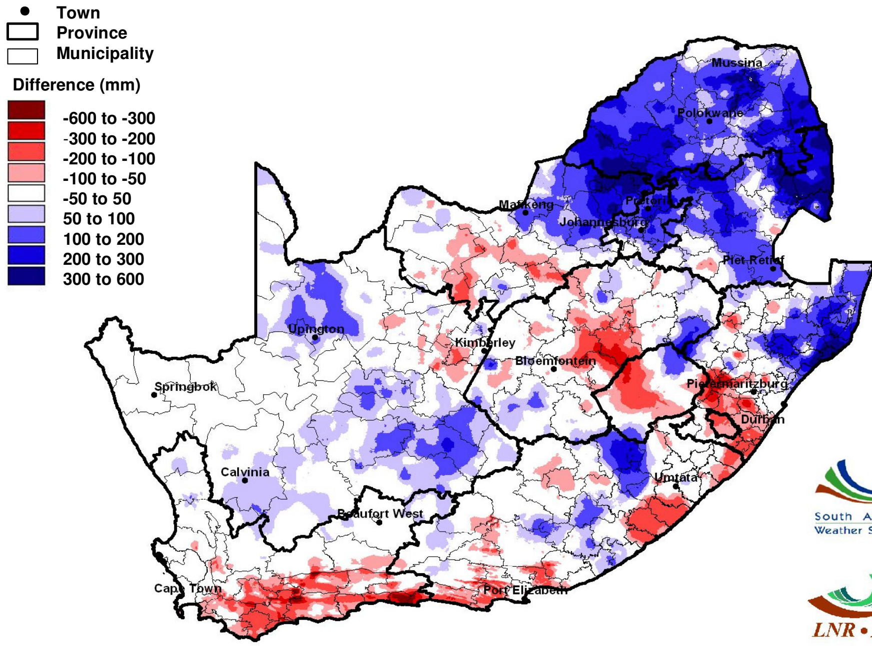


(Source: Umlindi Report 04051, 2004)

Figure 5 Percentage of the long-term mean rainfall for April 2004

From Umlindi Report 04051 (2004), Figure 6 was interpreted to describe the difference in rainfall between the period 1 January to 30 April 2003 and the same period for 2004.

Descriptions such as “wetter than normal conditions over the north-eastern third of the country” (Umlindi Report 04051, 2004) cannot quantitatively contribute to decision-making about specific (ton/ha) maize yield to be expected in specific planting date zones (areas).



South African Weather Service

LNR • ARC

(Source: Umlindi Report 04051, 2004)

Figure 6 Rainfall difference map between 1 January to 30 April 2003 and 1 January to 30 April 2004

As planting seasons differ from year-to-year, accurate final yield values cannot be directly linked to the NDVI or rainfall maps from the Umlindi Reports.

The proposed imagery for verification of objective yield data should be inexpensive and have a better spectral, spatial and temporal resolution than the currently used remote sensing data (NOAA-AVHRR) to ensure minimal additional costs to the existing maize estimation efforts whilst maximising the benefits from it.

The rainfall and NDVI values used for yield estimation purposes should be a reflection of (and be analysed according to) the growing stage and age of the maize plant at each objective yield GPS point, and not be compared with a long-term average per province as in current decision support reports.

The NDVI and rainfall values should be compared over areas of uniform maize growth stage or planting date zone. To achieve this, rainfall and NDVI values should be extracted per objective yield point and be linked to age, growing stage and an expected yield over dry or irrigated fields separately for different regions.

1.4 Economic Value

If the above criteria are met, an improved decision support system will result. South Africa has a total surface area of 121 957 787 ha, of which 14 million ha are classified as cultivated, temporary land cover (Thompson, 1995). The monitoring of maize crops using Terra's Moderate Resolution Imaging Spectroradiometer (MODIS) requires four (4) images to cover the whole of South Africa and one (1) MODIS image to cover the entire study area daily (http://edcimswww.cr.usgs.gov/pub/ims welcome/).

Terra MODIS images are free (except for a small courier charge) and suitable for the size of commercial maize farms in South Africa. One (1) MODIS pixel measures approximately 250 m x 250 m, depending on the chosen projection.

Estimating crop yield is of economic importance and helps with strategic planning. While time and effort are put into the objective survey data, it makes economic sense to utilise these datasets for ground-truthing of MODIS and other remotely sensed images in the possession of the Agricultural Research Council-Institute for Soil, Climate and Water, Pretoria, South Africa (ARC-ISCW).

More accurate, timely and cost-effectively obtained yield estimate figures will have the following advantages:

- Providing intermediate verification of objective yield efforts.
- Providing better decision support at municipal and provincial levels, for example, improving food security planning and allocation of financial aid.
- Having a stabilising effect on the South African Futures Exchange (SAFEX) market and grain trading factors, for example, price and quantity in the agricultural sector.

1.5 Problem to be Solved

The objective yield analyst should be able to verify the field calculations using a current dekadal MODIS NDVI and rainfall value. As at October 2004, no such verification procedure is in place.

There is a need for a system that is unbiased, independent and statistically sound for the estimation of maize yield, that can be used throughout South Africa. This system needs to be sustainable within the available financial and human resources constraints prevailing in South Africa and needs to be applicable to extensive commercial farming areas and to traditional subsistence farming areas.

As at January 2003, no GIS-based systems are incorporated into current maize yield estimation systems that account for the differences in planting dates that exist across South Africa's maize production areas, which extend over seven (7) provinces (Figure 7) with eight (8) different temperature and rainfall zones (du Plessis, 2004) (Figure 11). Analytical tools need to be developed to use the MODIS data available for crop prediction and monitoring purposes.

Funding bodies need to be made aware of the importance of continuing to capture the date of field visits and whether the field is cultivated under dry-land or irrigated conditions. The age of the maize plant should be noted by the field worker to assist the remote-sensing analyst to verify the calculated intermediate yields using current MODIS NDVI imagery.

1.5.1 Aim and objectives

The aim is to investigate the possibility of using MODIS NDVI and spatial rainfall data to supplement the current objective yield crop estimation methods. To achieve this aim, the objectives are:

- Defining the study area.
- Collecting objective yield data for the 2001/2002 growing season.
- Obtaining MODIS NDVI 16-day composite images and creating dekadal images from them.
- Obtaining dekadal rainfall grids for the 2001/2002 growing season.
- Extracting dekadal NDVI and dekadal rainfall values per objective yield point.
- Adjusting the dekadal NDVI and dekadal rainfall values from the Julian date to age of maize plant.
- Analysing NDVI curves for dry-land and irrigated fields corresponding to yield.
- Analysing rainfall curves for dry-land fields corresponding to yield.

- Quantifying the accuracy of the predictions made from average NDVI and rainfall curves for the 2001/2002 growing season.
- Comparing the average NDVI and rainfall curves from the 2001/2002 growing season with that of the 2004/2005 growing season.
- Comparing the accuracy of the predictions made from the average NDVI and average rainfall curves with that of the 2004/2005 growing season.

1.5.2 Hypotheses to be tested

Is it possible to use MODIS 250 m NDVI and spatial rainfall grid data to supplement the intermediate and final objective yield crop estimations?

1.6 Study Area

South Africa is situated on the most southern part of the African continent between 22° South and 35° South and 16° East and 33° East. One (1) MODIS image covers the entire area containing all the summer grain objective yield survey points. The experimental site lies between 25° and 32° South and 23° and 31° East (Figure 7). It includes the "maize triangle" as well as other major maize production areas in Mpumalanga, Limpopo, Free State and North West Province.

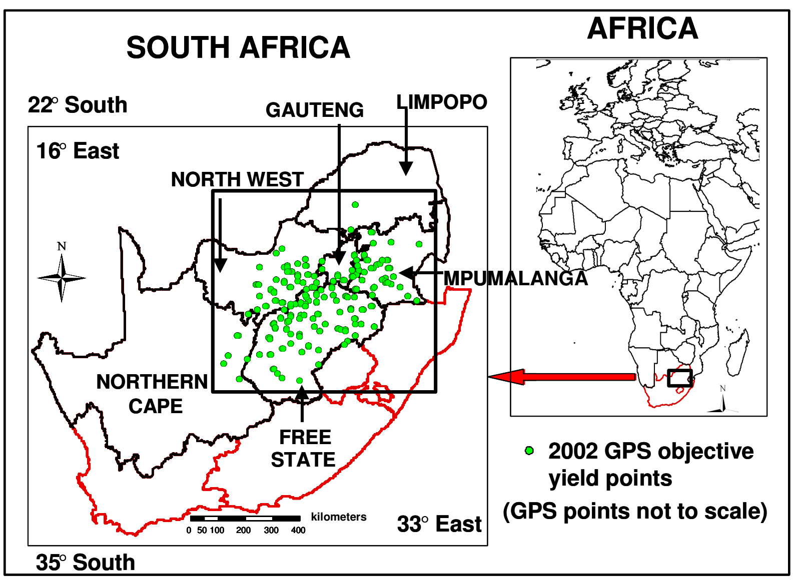


Figure 7 Orientation map of South Africa on the African continent showing 2001/2002 growing season objective yield points

2 LITERATURE REVIEW

Before the 1950s conventional crop yield estimation methods were based on field data collection and plant sampling. Since the launch of the first weather and communication satellites, these sensors have been investigated for their environmental monitoring potential. As early as the 1970s, attempts have been made to use remote sensors to distinguish between crops and/or calculate the area under a specific crop (Pinter *et al.,* 2003).

Dabrowska-Zielinska *et al.* (2002) modelled crop growth conditions and crop yield in Poland using satellite-derived indices. Using these indices, they estimated cereal yield for 49 regions. Newby (NCEC Proposal, 2002) recommended that the method be applied together with other methods to develop a decision support system for the estimation of crop yields in South Africa.

In Hungary, Ferencz *et al.* (2004) researched two methods for estimating the yield of different crops using Landsat and NOAA data. A correlation coefficient of $R^2 = 0.75$ was achieved for the field level yield for maize for three years using the General Yield Unified Reference Index (GYURI) and fitting a double Gaussian curve to the NOAA data. The county [region] average yield data showed higher correlation ($R^2 = 0.93$).

Since 1995 many factors have contributed to increasing pressure on researchers and governmental organisations to use remote sensing for crop monitoring. Scientists at agricultural research institutions, various government organisations and private institutions have gathered a lot of fundamental information relating to spectral reflectance and thermal emittance properties of soils and crops relating to their agronomic and biophysical characteristics. This knowledge has facilitated the development and use of various remote-sensing methods for non-destructive monitoring of plant growth and development and for detection of many environmental stresses which limit plant productivity (Pinter *et al.,* 2003).

Coupled with rapid advances in computing and position-locating technology (for example, GPS), remote sensing from ground-, air- and space-based platforms is now capable of providing detailed spatial and temporal information on plant responses to their local environment that is needed for site-specific agricultural management approaches.

The biophysical basis for agricultural remote sensing relies on the spectral reflectance properties of leaves, soils and crop canopies which, in turn, influence vegetation indices. Green plant leaves typically display low reflectance and transmittance in visible regions of the spectrum (that is, 400 to 700 nm) due to strong absorbance by photosynthetic and accessory plant pigments (Chappelle *et al*., 1992). In contrast, reflectance and transmittance are both usually high in the near infra-red (NIR) (700 to 1300 nm) regions of the spectrum because there is little absorbance by sub-cellular particles or pigments and also because there is considerable scattering at mesophyll cell wall interfaces (Gausman *et al*., 1975).

Yield is an important end-of-season observation that integrates the cumulative effect of weather and management practices over the entire season. There are two general approaches to using remote sensing for yield assessment. The first is a direct method in which predictions are derived totally from the remote measurements. The second is indirect, whereby remotely-sensed parameters are incorporated into computer simulations of crop growth and development, either as within-season calibration checks of model output (for example, biomass or Green Leaf Area Index (GLAI)) or in a feedback loop used to adjust model starting conditions or processes (Maas, 1993).

Currently many crop observation projects around the world use data from sensors aboard numerous platforms (for example, NOAA (Csornai *et al*., 1999) and MODIS (Ferencz *et al.,* 2004)). In many crop observation projects, satellite data products (for example, NDVI, Brightness Temperature (BT)), together with historical meteorological data (rainfall, temperature and evaporation) and yield data, are used to investigate time-series analysis (Dabrowska-Zielinska *et al.,* 2002).

In the Hungarian Agricultural Remote Sensing Programme, attempts to use Landsat with NOAA (Csornai *et al.,* 1999), to make crop yield estimations, led to a concise methodology that could be applied operationally. Crop area assessment, through the processing of multi-temporal Landsat and Indian Remote Sensing IRS-1C/1D data, proved to be efficient at county level because of the accuracy of thematic classification (Csornai *et al*., 1999). This crop yield forecast methodology performed well in Hungary for eight (8) major crops at county level.

A novel, robust method that combines land-use information with NOAA-AVHRR time-series for yield prediction was also introduced. Experiences of the first three (3) operational years (from 1997), as well as a general evaluation using the Operational Crop Monitoring and Production Forecast Program (CROPMON) are shared (Csornai *et al*., 1999).

Polish researchers Dabrowska-Zielinska *et al*. (1998) describe how optical data from NOAA-AVHRR and radar data from ERS-1SAR were used to monitor conditions of agricultural areas in Poland. The "cereal" pixels from the NOAA images were selected using supervised classification on Landsat Thematic Mapper (TM) images. For each of the pixels NDVI and soil moisture indices (H/E, WDI (defined in Glossary)) were calculated and the relationship between Leaf Area Index (LAI) and vegetation soil moisture indices was established. Information about vegetation conditions from Synthetic Aperture Radar (SAR) is essential for a region often covered by clouds.

The Southern African Development Community (SADC) Regional Remote Sensing Unit (RRSU) generates early warnings for food security. In conjunction with ground observations, satellite-derived vegetation, rainfall and modelling products are routinely used at the RRSU for annual monitoring of crops' performance during the main crop-growing period from September to April.

The RRSU uses SPOT4/VEGETATION, NOAA and METEOSAT imagery to form a convergent picture of the status of the crop-growing period. Visualisation of the SPOT4/VEGETATION Normalised Difference Vegetation Index (NDVI) images and their comparison with short-term (5-year) average conditions enable analyses of crop, pasture and general vegetation development.

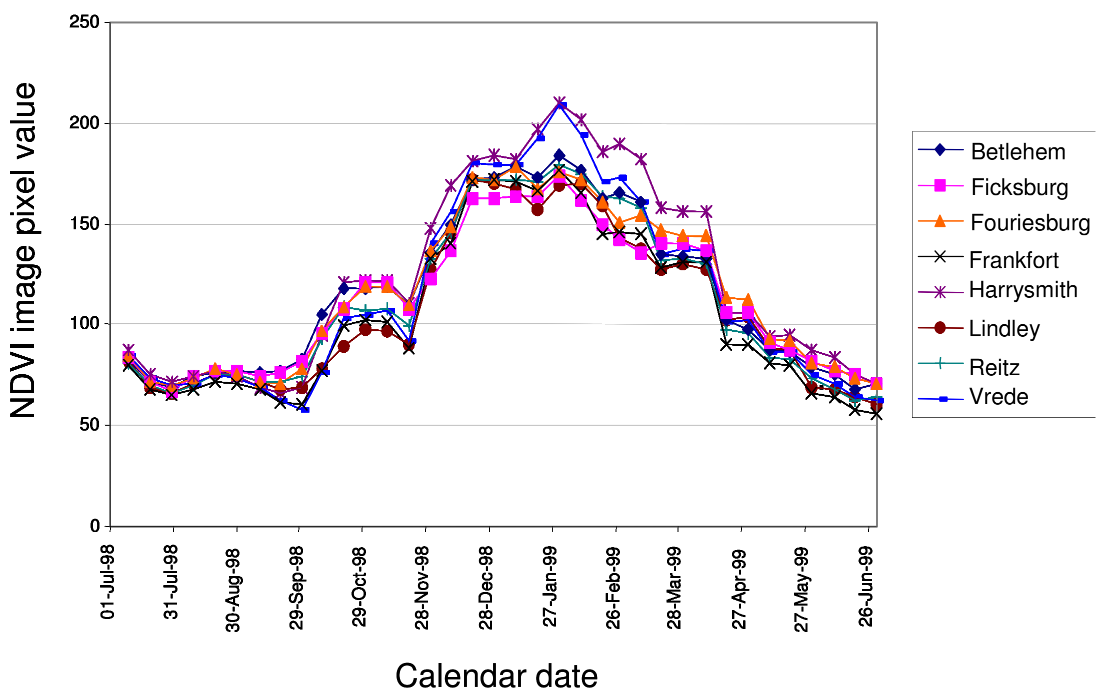
Simplified graphics of vegetation maps are used to illustrate the situation described by RRSU analysts in ten-day, monthly and special early-warning bulletins that are aimed at decision makers in SADC member states and their development partners.

To this extent RRSU has evaluated the ESA GSE Global Monitoring for Food Security (GMFS) prototype service for 2003. It is the intention of the RRSU to pursue further applications of SPOT4/VEGETATION and associated GMFS services, particularly in the quantitative estimation of crop yield and rangeland productivity (Masamvu and Siwela, 2004).

Current research initiatives involve the calculation of NDVI and BT values from NOAA (and later from MODIS) images. A Vegetation Condition Index (VCI) and a Temperature Condition Index (TCI) are also used in crop studies. The ARC-ISCW is also researching these indices (Newby, 2005: Personal communication).

In Poland, researchers such as Dabrowska-Zielinska *et al.* (1998) have implemented classical time-series analysis. In other parts of the world, researchers such as Ferencz *et al.* (2004) studied AVHRR greenness values (GN) corresponding to day of the year.

Present research trends at ARC-ISCW involve using the South African NLC (Thompson, 1995) data to determine which pixels in the NOAA images are represented entirely by natural grassland (Newby, 2005: Personal communication). Natural grassland pixels are used to determine crop growth over a season, because it is assumed that natural grassland is less variable than cultivated land which may be planted with different crops each year.



(Source: NCEC Proposal, 2002)

Figure 8 Mean NOAA NDVI for districts in the North-East Free State for the period July 1998 to June 1999

In South Africa, NOAA-AVHRR NDVI values are compared with historical yield data in the same way as done in Poland. The area under the curve in Figure 8 corresponds to yields obtained in magisterial districts of the North-East Free State.

Magisterial districts may differ greatly in terms of size, percentage of area used for agriculture, access to water for irrigation purposes, production region and agro-climatological properties.

Topographical, climatic and production areas and conditions in South Africa are unique and the implementation of methods developed in other countries might not be applicable. Financial constraints, the large area of the country and the many diverse topographic and agro-climatological regions pose a unique challenge to existing remote-sensing crop yield methodologies. As at 2004, a complete set of maps of crop types does not exist for South Africa.

The hypothesis; "Is it possible to use MODIS 250 m NDVI and spatial rainfall grid data to supplement the intermediate and final objective yield crop estimations?" stated in section 1.5.2, implies that the remotely sensed data should be able to give an indication of end and intermediate maize yields. But as the literature reviews in this chapter investigate the same phenomena, the hypothesis also inherently implies that the MODIS data can be used to verify (and/or improve) the objective yield ground data. In fact, it was hoped that the NDVI data would validate what the objective yield data estimations expected the end yield to be at each of the three intermediate readings. Although this falls outside the scope of the research for this particular report, it is well worth considering for subsequent research on objective yield point data.

A number of authors used surface measurements through field campaigns to validate and calibrate several MODIS products for the NASA Earth Observing System (EOS) program at a resolution of 250 to 100 m. What is important to remember, also for this research report, is the mismatch in scale between ground point measurements and the MODIS resolutions (Liang *et al*., 2002).

3 DATA COLLECTION AND METHODOLOGY

3.1 Data Sources

3.1.1 Terra MODIS

The Earth Observation System's Moderate Resolution Imaging Spectroradiometer (EOS MODIS) is a passive remote sensor, meaning that it makes use of the reflected sunlight and electro-magnetic radiation (EMR) radiated from the earth's surface to observe vegetation (Pinter *et al*., 2003). Terra MODIS 250 m resolution images were ordered directly from United States Geological Survey (USGS). This minimised in-house data processing and data storage requirements; ensured consistent quality of the processed products, and consistency of the procedure by which the NDVI images were produced. Products received included pre-produced Enhanced Vegetation Index (EVI) and NDVI images. The NDVI images were used to extract NDVI values at objective yield GPS points.

NASA's EOS-AM1 Terra satellite was launched on 18 December 1999. At an altitude of 705 km, and a descending Polar Earth Orbit Inclination of 98.2°, the MODIS sensor has a swathe of 2330 km. MODIS boasts 36 wave bands (VIS, NIR, MIR, TIR). These bands are narrower than those of NOAA and Landsat, providing a better spectral resolution. The entire surface of the earth is viewed every one (1) to two (2) days. Spatial resolution ranges from 250 m, 500 m to 1000 m. Currently, NOAA NDVI data are used monthly to help the CEC to predict expected maize yield for the end of the growing season. MODIS NDVI images have a spatial resolution of 250 m, which is an improvement on the NOAA-AVHRR NDVI product which has a resolution of 1.1 km.

Obtaining frequent updates on data is essential for crop monitoring and yield prediction (Doraiswamy *et al*., 2003). Terra MODIS delivers two (2) day-time images per area per day. The MODIS sensor aboard the Aqua satellite platform, launched on 4 May 2002, delivers early morning and night-time images, not suitable for crop monitoring.

Research that was done in the past using NOAA data (of which the ARC-ISCW has an 18 year archive) could be continued using MODIS data in future. Using MODIS data for future research by the ARC-ISCW is sustainable, as structures and funding are place for the continuation of a similar sensor via the National Polar-orbiting Operational Environmental Satellite System (NPOESS) (King, 2005: Personal communication). The ARC-ISCW is contracted to receive MODIS products from the Council for Scientific and Industrial Research-Satellite Application Centre (CSIR-SAC), which has its own MODIS receiving station, situated at Hartebeesthoek in South Africa.

MODIS red band (620 to 670 nm) and NIR band (841 to 867 nm) provide superior spectral resolution for the identification of maize. NDVI is the result of (NIR-Red)/(NIR+Red). Band 1 of MODIS is narrow and, therefore, is able to separate vegetation from rocks and soil well. Considerations in terms of the radiometric quality of the 16-day Terra/MODIS datasets used need to be taken into account.

Four (4) Terra MODIS land surface products have been examined in a study by Liang *et al.* in 2002. These included bidirectional reflectance product from atmospheric correction (MOD09), bidirectional reflectance distribution function (BRDF) (MOD43B1) broadband abedo's (MOD43B3) and nadir BRDF-adjusted reflectance (MOD43B4). The initial validation results showed that these products are reasonably accurate (less than 5% absolute error). The 5% absolute error of the MOD09 is applicable to reflective bands such as those used in the 16-day NDVI product (Terra MODIS band 1 RED and band 2 NIR). The MODIS products used in their study were not the final ones and they recommended that the final conclusion about the uncertainties of these MODIS products should be made after MODIS data processing (Liang *et al.*, 2002).

The Terra/MODIS 16-day 250 m NDVI data have been processed and according to the Terra L1B Product Disclaimer, the MODIS/Terra L1B Version 003 and higher products are considered validated. From the GES DAAC Terra Level 1B Data Quality Summary Statement, the sensor operational configuration, detector biases, and lookup table parameters are time-dependent quantities that have been changed to optimize sensor performance. Changes are documented in the metadata and/or through links available on the MODIS Characterization Support Team (MCST) main page http://www.mcst.ssai.biz/mcstweb/

The MODIS instrument experienced a Power Supply 2 (PS2) shutdown anomaly and did not take science data during the time period June 15, 2001 to July 2, 2001. The cause of the failure is consistent with an over-voltage shutdown most likely initiated by a high-energy radiation event that caused the Metal-oxide Semiconductor Field Effect Transistor (MOSFET) within the down-regulator of PS2 to fail. Immediately prior to the anomaly, the MODIS instrument had been acquiring data using electronics side-B.

When the MODIS instrument recovered, it was commanded to take science mode data using Power Supply 1 and electronics side A. Science data collected since recovery show that the instrument is performing as expected. However, the then-existing lookup tables used for L1B processing of post-anomaly data were from calibrations using electronics side B, so L1B and downstream products from the post-anomaly period processed in July 2001 were imprecise and will remain so until new lookup tables are constructed and implemented using post-anomaly calibrations.

The error introduced can amount to 5% for the reflective solar bands. The Solid State Recorder (SSR) error detected on May 20, 2001 has been corrected. The net result of the anomaly is an increase of one "superset" of memory being allocated to the MODIS instrument. This has increased MODIS' SSR buffer allocation by about 3%.

On August 15, 2001, the Terra MODIS instrument experienced the first of many formatter errors. Over the following weeks, the error rate increased from one every few days to several thousand per day. Memory dumps of the formatter patch locations have been performed multiple times and the error address buffer has been cleared once to aid in analysis of the faults. The error rate decreases when the instrument temperature increases, such as during blackbody and SRCA calibrations. Currently, the formatter errors have not degraded the MODIS instrument science data.

Data was processed using B-side electronics starting on October 30, 2000 through June 15, 2001. Subsequently, the measurement quality has improved dramatically. A new set of quality flags pinpoints the few remaining noisy detectors. Several previous areas of concern have been adequately addressed in this release and the subsequent post-anomaly A-side data release (MODIS Characterization Support Team(MCST)http://www.mcst.ssai.biz/mcstweb/).

The terms of the 250m 16-day NDVI Data Quality Summary statements found on the MODIS Data Support site, but also included in the .XML document received with every .HDF file, list the % cloud cover, aerosol and other general quality assessments. This might differ for every NDVI data set received. Whilst for most of the scenes, the Operational Quality was "Passed", the Science Quality was "Being Investigated". The quality summary also stated that the "Product assessment and validation continues. Users are advised to use caution applying these data to project-applications" (.XML documents). This should be kept in mind when the quality of the 16-day Terra/MODIS NDVI data sets is considered.

3.1.2 Rainfall data

Final rainfall maps for the 2001/2002 growing season were produced from rainfall data extracted from ARC-ISCW and South African Weather Service (SAWS) stations. Altogether information from 1500 weather stations was used. These surfaces were interpolated using the long-term rainfall surface as a trend for interpolation between stations. The Inverse Distance Weighted (IDW) interpolation technique was used (ARC-ISCW, 2004). Rainfall maps are available on www.agis.agric.za/umlindi/umlindiweb.

ARC-ISCW and SAWS stations with more than 20 years of reliable rainfall data were used to create long-term rainfall surfaces with the help of regression modelling between rainfall and surface parameters (such as elevation, distance from the sea, rain-shadow effects of mountains and large-scale roughness of the surface). For all 10-day periods (dekads) since 1985, 10-day rainfall totals for rainfall recording stations were expressed as a percentage of the long-term mean for that specific period.

The IDW method of interpolation was used to interpolate the rainfall received as a percentage of the long-term mean between stations for which rainfall data were available. This interpolated "percentage of long-term mean rainfall" surface was combined with the long-term mean rainfall surface (for which the long-term rainfall surface serves as a trend surface) in order to create an interpolated rainfall surface. The temporal resolution of the rainfall surfaces is dekadal (10-day period), while the spatial resolution is 1 km (ARC-ISCW, 2004).

3.1.3 Objective yield data

The ARC Grain Crops Institute in Potchefstroom (ARC-GCI) provided the objective yield data for the planting season 2001/2002. Only the dry-land points situated in the analysis area were used (539 points). Thirty-three percent (33.3 %) of the dry-land data points were randomly selected and removed to serve as a test dataset (179 points), which left 360 points for analysis purposes. Objective yield data surveys were not performed for the 2002/2003 growing season. Objective yield surveys for the 2003/2004 growing season did not differentiate whether fields were operated under dry-land or irrigated conditions.

The fieldworkers visit the objective yield points for the first, second and third times usually around March, April and May, respectively. The fourth visit is to record the final yield obtained from that specific field (or the farm as a whole) after the farmer has completed harvesting.

Mass per cob, expected maize yield and variable area at each GPS point in the field are calculated by substituting information from field measurements (Section 1.1.3) into the formulae below (du Toit, 2002):

Mass per cob = kernels per row x rows of kernels x average mass of a kernel

Maize yield (ton/ha) = area x (kernel mass per cob/1000) x (average cobs
x total plants / average number of plants)

Variable area (rows/ha) = 10 000 / sample area x [(row width x 10 metres)
= sample area)]

Harvest losses for maize are estimated at around 15 % due to handling, transport and harvesting methods. This excludes hail and pest damage.

Traditional crop yield prediction methods tended to be destructive: on field visits, the field worker sometimes removed cobs and counted the rows of kernels and kernel rows at the office, where the kernels were also weighed. This method of assessment was destructive and distressing to participating farmers. New, non-destructive objective yield methodologies, implemented in the 2004/2005 objective yield survey, include not removing cobs or kernels from the field.

During the 2004/2005 survey collected field data were captured directly using a web-based system, the Web Based Data Capture Wizard. This software has built-in quality control mechanisms (for example, the area of a field cannot be bigger than that of the farm on which it is located). The Web Based Data Capture Wizard can be linked to a central storage facility, and field information downloaded. From there, the information can be disseminated to the other crop consortium members.

3.2 Maize Yield

The two components of maize production forecasting are yield (ton/ha) and area planted to maize (ha). If both of these components are available, total maize production for a planting zone, magisterial district, province or the whole of South Africa can be calculated as follows:

Total maize production (ton) = Maize yield (ton/ha) x Area planted under maize (ha)

3.2.1 MODIS NDVI value extraction

The characteristic spectral pattern of the maize plants at different stages is due, in part, to the chlorophyll pigment in the leaves and stems of the plants. Absorption of bands near 0.45 μm and 0.68 μm, in the blue and red parts of the visible spectrum, gives healthy leaves their green appearance. In the shorter part of the infra-red spectrum, most of the energy absorbed is re-emitted to maintain the energy balance. Different vegetation types have different spectral response patterns (SRP) in the NIR region and this produces good results when used for distinguishing vegetation. The NDVI indicates the greenness of the plants on the soil and also distinguishes the soil surface from the plant-covered surfaces.

The developmental stages of the maize plant are such that every 10 days marks a significant stage in the plant's development. The rainfall grids were created from dekadal (10-day intervals) information while the MODIS NDVI images were 16-day composites. To reflect the stages in the growth of the maize as well as being in synchronisation with the rainfall grids, the 16-day MODIS NDVI values were converted to 10-day (dekadal) images by using interpolation to create an average second dekad image. The NDVI for each Julian date was moved to produce NDVI values for every dekad at the objective yield points corresponding to the age of the maize plants.

3.2.2 Rainfall value extraction

Dekadal rainfall values were extracted from the ArcMap rainfall grids in a GIS system for each of the 2001/2002 growing season objective yield points. These dekadal rainfall values were extracted using ArcGIS Spatial Analyst and transferred to Microsoft Excel (2000) spreadsheets. In the spreadsheets, the dekadal rainfall values were aligned to coincide with the dekadal age of the maize plants. These data tables were then analysed.

4 RESULTS AND DISCUSSION

4.1 Importance of homogeneous plant age

4.1.1 Planting date zones

From the data captured at the objective yield points, maize planting dates were used to generate “planting date maps”. The recommended future use of objective yield data for crop estimation is to firstly generate homogeneous planting date zones like those illustrated in Figure 9 and Figure 10. These maps illustrate the products that were created from objective yield planting dates. The planting dates zones range from the earliest planted maize (green) to the latest planted maize (red).

Five different planting date zones were identified and missing planting dates could be interpolated from the planting date maps. The planting date maps were created using two types of interpolation methods. The Kriging statistical method (Childs, 2004) was used because it results in a smoother surface than that of planting date maps generated using the Inverse Distance Weighted (IDW) method (Childs, 2004). Planting date zone maps should be used in conjunction with analysing NDVI and rainfall values, so that values can be associated with maize plants of the same age.

The planting date zones were chosen arbitrary and can be displayed in a number of ways depending on the interpolation techniques and number of planting date categories chosen for display. Figure 9 was created using the IDW method with 35 points. Quantile date zone classification was used for display.

The basic difference between Krige’s technique and the IDW method is that the grids created by Kriging (Figure 10) do not flow through the points of origin but resonate above and below them to create a smooth surface. The IDW grid, however, flows directly through the points creating accurate planting dates at each point.

The advantage of using the Krige method is that the smoother planting zones are generated which make it easier to distinguish large areas of homogeneous planting dates. The disadvantage of using Kriging is that the resulting planting dates generated could be slightly above or below the original dates.

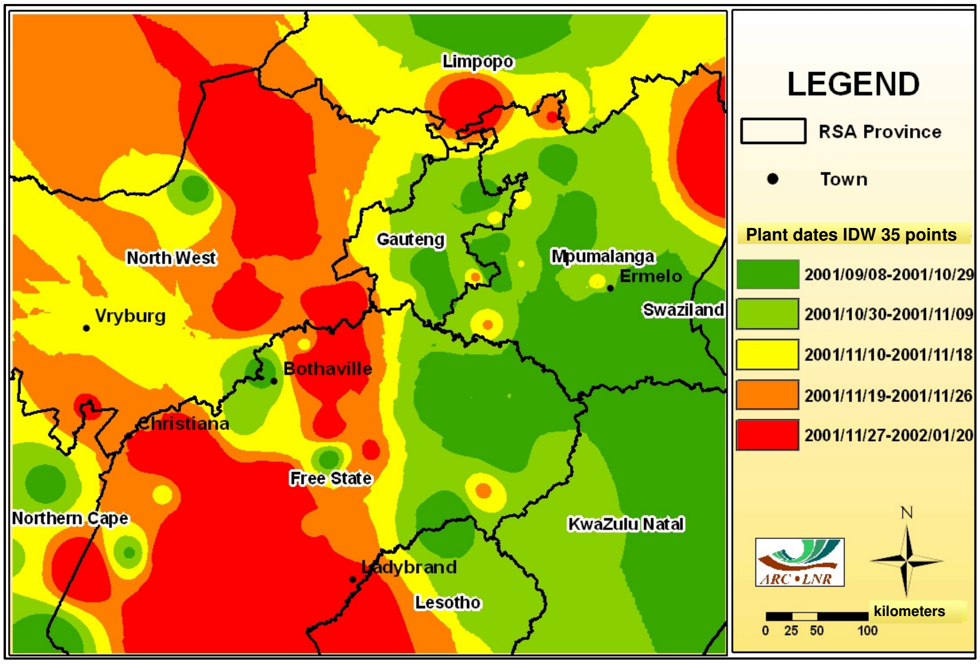


Figure 9 Planting date zone map of the 2001/2002 growing season. Maize planting date zones created using the IDW method with 35 points and quantile classification

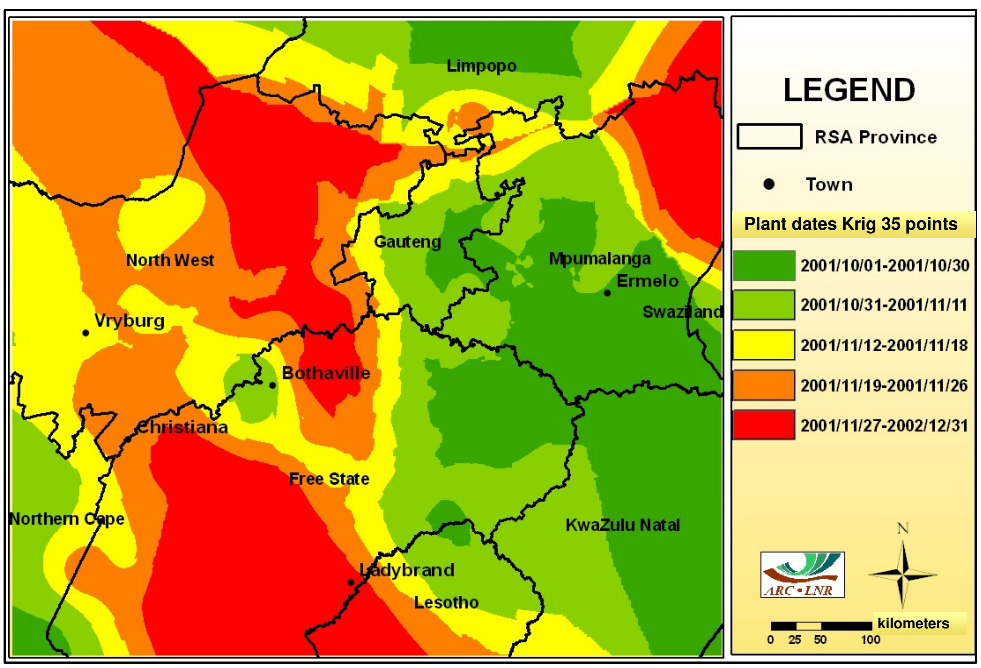


Figure 10 Planting date zone map of the 2001/2002 growing season. Maize planting date zones created using the Krige method with 35 points and quantile classification

Planting date depends on temperature and the amount and timing of the rains required for planting. In 2002 normal conditions prevailed. This means that the rains were neither late nor early and that it wasn't a particularly dry or wet year. Planting date maps created for the 2001/2002 growing season were compared with planting date maps created from objective yield data from the 2004/2005 growing season, because it was a year in which the rains were late but plentiful (Frost and Kneen, 2006).

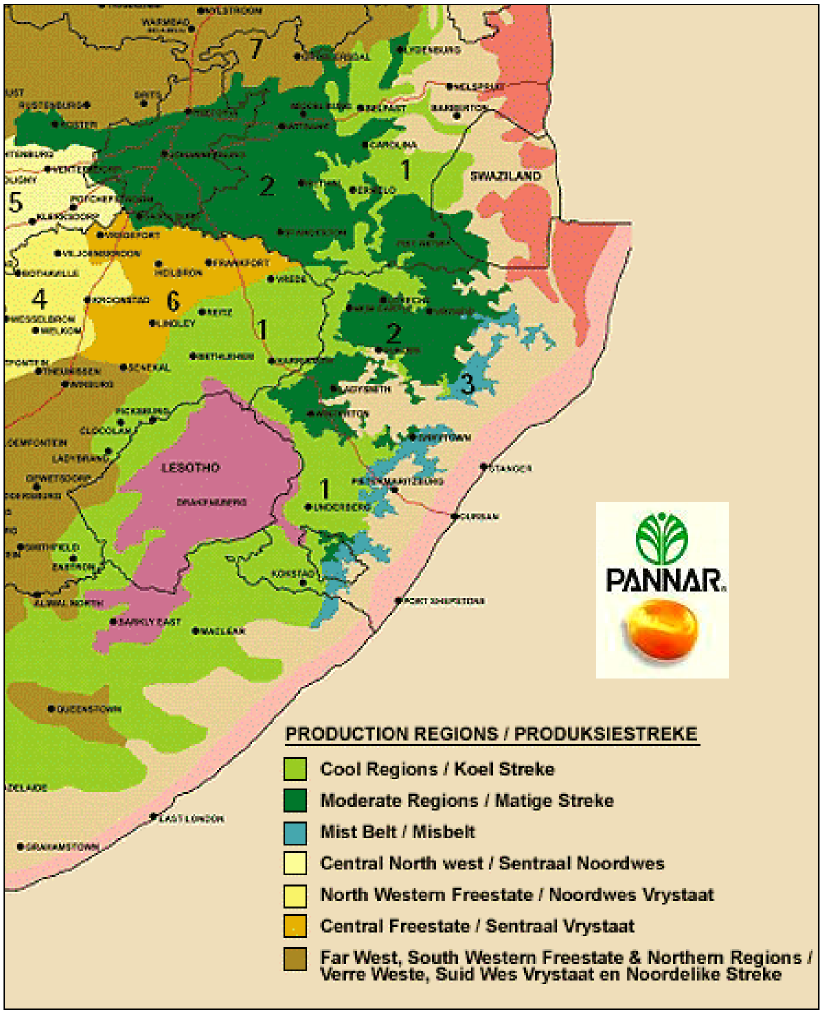
The planting date zone maps created from the 2004/2005 data had five planting zones that were more gradual, and no late planting in the northeast. The latest planting date was 28 January 2005. In 2004 the rains were late and it was perceived initially as being a dry season. However, at the end of the summer maize-growing season, record harvests were achieved (Frost and Kneen, 2006).

Further study is needed to determine the optimal number of planting zones for every growing season. This could include testing for optimum statistical, GIS and qualitative (visual) methods. These might differ from year to year (depending on whether the season is perceived to be "wet", "dry" or "normal") and depend on the total number of points available or the size of the study area. An optimal method of interpolation also requires further research.

Rain and temperature play an important role in the development of the maize plant. According to Hanway (1966) and Du Plessis (2004), maize prefers warm temperatures in which to grow. Table 1 indicates how growing periods vary for different cultivars in different temperature zones. Frost-free periods of 120 to 140 days are needed for the plant to develop and to produce grain. Figure 11 illustrates production regions based on temperature zones (PANNAR, 2004).

Using a long-term average temperature zone map (similar to the map in Figure 11) for NDVI analysis cannot accommodate different growing seasons in the way that annual "planting date zone maps" can (Figures 9 and 10). Visual inspection of one (1) province at a time reveals that, for example, the Free State has four (4) temperature zones but five (5) planting date zones, and that the position of these planting data zones does not fully coincide with the long-term average temperature zone map, although there are similarities.

The length of each growing stage is a guideline only and will differ slightly for short-, medium- and long growing cultivars in cool and warm regions as described in Table 1.



(Source: PANNAR, 2004)

Figure 11 PANNAR production regions based on temperature zones

Table 1 Length of the maize growing season

	SHORT	**MEDIUM**	**LONG**
	Days to pollen shed	**Days to pollen shed**	**Days to pollen shed**
Cool regions	70 to 75	75 to 80	80 to 85
Warm regions	60 to 65	65 to 70	70 to 75
	Maturing	**Maturing**	**Maturing**
Cool regions	130 to 140	140 to 155	155 to 165
Warm regions	120 to 125	130 to 140	145 to 150
Source: ARC-GCI Interviewer's manual, 2004:17.			

4.1.2 Developmental stages of the maize plant

In order to interpret the graphs in Section 4.2 one needs to understand the growth stages of the maize plant. NDVI behaviour can be linked to these stages (the age of the maize plants) at each objective yield point. From Figure 16 to Figure 40 (Section 4.3) and Table 1 (Section 4.2) it should be remembered that the days given in Section 4.1.2 are averages and might vary depending upon the objective yield point being in a warm or cool region and the growing season being long, medium or short.

According to Hanway (1966), there are 10 distinct growing stages for maize plants. Roughly every 10 days (dekad), the maize plant enters a significant stage of development. On the graphs, each stage (or dekad), expressed in terms of "days from planting date" can be evaluated against the MODIS NDVI and rainfall grids that prevailed at that stage, because those grids are also dekadal. This aids in the monitoring of stress factors linked to yield loss, experienced at any of the crucial growth stages or dekads.

Stage 1

If farmers plant under dry-land conditions, they have to wait for rain before the seeds can germinate and then it takes five (5) to eight (8) days before the seedlings emerge. Under optimal conditions, the average number of days that it takes from planting day to emergence is 10 days. But, if it does not rain, the seeds cannot germinate and it might be more than two (2) weeks before shoots appear above the ground.

That is another reason dry-land and irrigated maize objective yield points should be analysed separately, because, under irrigated conditions, farmers do not have to wait for rains before germination can take place and it takes one (1) dekad (10 days) from planting date before the seedlings appear. The difference between maize plant behaviour under dry-land or irrigated conditions should be borne in mind throughout yield analysis.

From the objective yield fieldwork, the planting date was acquired. This day is marked as zero (0) or is the point of origin on the time axis (X-axis) of the graphs. The occurrence of any of the 10 growing stages can now be easily identified on the X-axis.

By the end of Stage 1, four (4) leaves have completely unfolded 14 days after emergence (24 days after planting date), that is, by the middle of dekad 3. The number of internodes, leaves and ears has been determined and the growth point is under the soil surface. The tassel has been determined.

Stage 2

At this stage, eight (8) leaves have unfolded completely at 30 days after emergence (40 days after planting day, that is, at the end of dekad 4). The internodes are lengthening. The developing tassel is eight (8) cm above the soil surface. The apical meristem (top growing point) is now above ground. The tiller roots begin to develop from nodes below the soil surface.

Plant height can vary between 0.6 m and 3.0 m. At each internode a leaf develops, which is arranged spirally on the stem, and occurs alternately in two opposite rows on the stem. The number of leaves can vary between 8 and 22, depending on the maize cultivar.

Soil properties are still prominent at this stage of analysis. That is why monitoring is done using NDVI values instead of just NIR images.

The environment plays an important role in the height of the plant. Under stress, the internodes do not lengthen fully and shorter plants result. Around the nadir of the image, plant height is not easily discernible but stress in plants still influences the spectral signature.

Stage 3

This is 45 days after emergence (or 55 days after planting, that is, dekad 6) and 12 leaves have fully unfolded. The stem is thickening and the lowest four leaves are dying off. Ears are developing on nodes 6 to 8 and the tassel is developing more rapidly. The number of ears is genetically determined, but the environment plays an important role in the final number of ears that will develop.

This is a critical stage in the development of the plant, as the potential number of kernels and the ear size is determined. The terminal ear develops first and, depending on the environment, the rest of the ears will develop. Loss of leaves at this stage will result in yield reductions.

Stage 4

At this stage 16 leaves should have fully unfolded 60 days after emergence (70 days after planting, that is, dekad 7). Tassel development is almost complete and is pushed higher up in the plant. Husk leaves protect the ears. Silks lengthen and protrude from the husks. Silks at the bottom end of the ear start to lengthen. Damage to leaves can cause yield loss.

NDVI values continue to rise. Less soil is visible and, depending on the row width, make up a lesser part of the NDVI value. As in Stage 4, a low NDVI value might be linked to lower than expected yields. Sufficient rainfall is crucial during Stage 1 to Stage 4.

Stage 5

By the end of stage 5, all the leaves are fully unfolded, 70 days after emergence (80 days after planting day, that is, dekad 8) (Figure 12). Pollen starts shedding. The silks become visible, and are pollinated by pollen produced by the tassels. Heat and moisture stress during pollen shed and silk emergence can lead to yield reductions, as can leaf loss.

(Source: Monnik *et al*., 2002)

Figure 12 Maize plants at Stage 5

Too much rain at this stage can also lead to yield loss. Above-optimal temperatures can be detrimental during the pollination stage. Pollination is inhibited when the air moisture values are too high. The pollen becomes swollen and bursts, which can hinder pollination and subsequent kernel counts. To identify too much rain or other plant stress incidents, monitoring around dekad 8 will reveal possible factors responsible for the stress.

A prolonged hot, dry period can cause the pollen to be burnt, resulting in poor pollination and yield loss. It could be wise to examine temperature and air moisture (or rainfall) grids to detect any of these possible pollination-inhibiting circumstances.

NDVI values alone are not always an absolute indication of the potential yield at this stage in the maize crop development.

The kernels begin to develop and the increase in mass is approximately 9 mg/day.

Stage 6

Pollination took place in Stage 5 (Figure 12) when the maize plant was at 70 days after emergence (80 days after planting). The kernels are in the "milk stage" at Stage 6 (Figure 13). From 12 days after pollination (82 days after emergence), the kernels grow in size. The maize plant is now at 92 days after planting day, that is, between dekad 9 and dekad 10. Kernel moisture content at this stage is 80 %.

Leaf loss can result in the absence of kernels at the tip of the ear. Results from this research indicate that using observed NDVI values and changes in NDVI values to form an idea of future yield, should be effective around Stage 6 (dekad 10).

Note the difference in row width between Figure 12 and Figure 13. The amount of red or dark brown soil can affect the reflectance in the red channel of MODIS. Row width could have an impact on the NDVI value (and yield).

(Source: Monnik *et al*., 2002)

Figure 13 Maize plants at Stage 6

Stage 7

It is now 24 days after pollination and the kernels are in the "soft dough stage". The maize plant is 94 days old (104 days after planting, that is, dekad 11) (Figure 14). The kernel can still easily be broken with the thumbnail and contains about 70 % moisture. Kernel mass is 50 % of the final mass. NDVI around dekad 11 could indicate the vigour of the maize plant and the efficiency with which the kernels are fed via photosynthesis.

(Source: Monnik *et al.*, 2002)

Figure 14 Maize ears showing dried silks

Stage 8

Kernels are at the "hard dough" stage 36 days after pollination. The maize plant is now 106 days old (116 days after planting, that is, can be compared with dekad 12 of NDVI and rainfall). The sugars in the kernels are being converted into starch and the dent becomes visible on the crown of the kernel. The mass of the kernels is still increasing and their moisture content is between 50 % and 55 %.

Stage 9

Kernels are physiologically mature at 48 days after pollination. The maize plant is now 118 days old (128 days after planting, that is, dekad 13). Kernel mass does not increase further and the familiar black layer has developed at the base of the kernels. Nutrients are now prevented from reaching the kernel and plant growth has stopped. The black layer is observed on 90 % to 95 % of the kernels at the base of the ear and the moisture content is between 30 % and 40 %.

When inspecting the NDVI around day 128 (dekad 13) it is evident that it is no longer increasing.

Stage 10

Kernels are now biologically mature (60 days after pollination). The maize plant is now 130 days old (140 days after planting, that is, can be compared with dekad 14 of rainfall and NDVI grids). Plant leaves and husk leaves are changing colour (Figure 15). Drying down of the kernels is associated with the environmental conditions. Moisture loss is 5 % per week until the kernel moisture reaches 20 % after which moisture loss becomes much slower.

The reflectance of a field of maize plants in this stage is representative of dry vegetation. More light in the red part of the spectrum is reflected while less light in the green part is reflected. More infra-red is absorbed and less infra-red is reflected as would have been in the case of a healthy, growing plant (Drury, 2001).

(Source: Monnik *et al.*, 2002)

Figure 15 Mature maize ears

4.2 Data Analysis Methods

4.2.1 Yield classes

The high-, medium- and low yield class boundaries were selected after a series of trials with seven (7), five (5), four (4) and three (3) yield classes. The system had to be robust and easy to use and understand as the end users of the model might vary from farmers to decision makers and scientists. Using, low-, medium-, high- and very high maize yield classes did not result in a visual decrease in overlap (standard deviation between the classes) and did not deliver better yield prediction results (Frost, 2006).

When it was established that three (3) yield classes were optimal, the selection of class boundaries between the three classes were so chosen as to produce the smallest amount of standard deviation overlap between the classes. The yield boundaries between the three classes were chosen on a visual basis of greatest separation between the average NDVI and average rainfall value curves of the yield classes.

This least inherent overlap was visually established by producing graphs of the three classes with different class boundaries. Eventually it was decided at which boundaries the three curves were furthest apart or seemed to be visually optimally separated and these boundaries were chosen. The dry-land objective yield point data were classified into the following three yield classes for analysis purposes: "Low yield" refers to objective yield points with a final yield (farmer reported final harvest) below 3 ton/ha. "Medium yield" refers to objective yield points with a final yield from 3 ton/ha up to 4 ton/ha. "High yield" curves represent objective yield points with a final yield of above 4 ton/ha.

Irrigated fields were analysed using different class boundaries for low- (2 ton/ha to 5 ton/ha), medium- (9 ton/ha to 10.4 ton/ha) and high yield (11 ton/ha to 14 ton/ha) and can thus not be compared directly with estimates made for dry-land fields. This is due to the fact that the spread of yield values and the minimum and maximum yields achieved under irrigation conditions is much greater than those achieved under dry land conditions.

Several statistical methods of separation could also have been used to select the class boundaries and is recommended.

4.2.2 Yield prediction windows

Yield prediction windows are dekads or a collection of dekads after planting which seem to have a larger probability than other dekads to be used for predicting maize yield. A distinction was made between two types of window periods. The most significant type is a window period that seems to be useful for predicting the three maize yield classes and maize yield in ton/ha (from the average NDVI or average rainfall curves). The second type of window period is a collection of dekads that visually seem to be useful for rough food security predictions during the growing season by being able to distinguish between high or Medium/low yields or medium/high (black window in Figure 23) or low yields (i.e. the red window in Figure 27).

The prediction windows (Figure 23, Figure 24, Figure 25, Figure 27, Figure 35, Figure 36, and Figure 41) were visually identified. This was done on a qualitative basis by visually detecting where the average NDVI and average rainfall curves seem to be separated far enough to be able to be inherently different and could possibly be used for maize yield prediction.

Apart from the robust, qualitative method of selecting the window periods visually, other more quantitative and/or statistical methods could also have been considered for this purpose. One such a quantitative method that is worth mentioning is the food security early warning system developed by Jeremy Freund, 2006, for the SADC region. The predicted yield is based on the seasonal maximum NDVI at day 110 to day 160 after planting (onset of rains). As sufficient point data is not available in most of these countries, the MODIS NDVI time series data is smoothed and the SADC Landover map is used to eliminate pixels that are not crops. This maize yield estimation system is suited for district, province and national level, but not field level. A pixel level regression coefficient proportional to crop area is used to directly relate NDVI to yield. The yield values predicted correlate well to published data from the Production Estimates and Crop assessment Division (PECAD) of the Foreign Agricultural Service of the U.S. Department of agriculture at district level. The model does not need the published data to function as it has already been validated (Freund, 2006).

Several statistical methods of determining at what stages the average curves are significantly dissimilar to be able to predict yield from, at a significant confidence level, should be considered for future analysis.

4.2.3 Model verification

In section 3.1.3 it was discussed that thirty-three percent (33.3 %) of the dry-land data points (179 points) were randomly selected and removed prior to analysis to serve as a test dataset, which left 360 points for analysis purposes. Two methods are described to implement the test data to verify and test the model illustrated in Figure 41.

The first method is a visual testing method (described in section 4.3.4) and the second is a statistical testing method (described in section 4.3.4, 4.4.4 and Chapter 5 Conclusions.

The visual testing method involved firstly calculating the average yield value for each yield class curve. These values are given in section 4.3.4. The high yield curve average yield is 5.12 ton/ha, the medium yield curve average yield is 3.54 ton/ha and the low yield curve average yield is 2.16 ton/ha. The NDVI and rainfall values that were assigned by the GIS system to the objective yield GPS points were then applied to the model by pinpointing the values on the graphs on Figure 41 and making a visual estimate of which yield class the GPS point could belong to and what the yield (in ton/ha and or high-, medium-, low yield) could be expected.

The second method of verification is a statistical method. Two types of questions were answered for the test data. It is calculated whether the NDVI and/or rainfall values assigned by the GIS system to the test data point was within (+/-) 1 standard deviation of that of the class for that specific dekad or not. The other method was to test whether the NDVI and/or rainfall values assigned by the GIS system to the objective yield GPS point was within the 95% confidence level of that of the class: (average value for the class for that dekad +/- (1.96*(standard deviation of that class for that dekad)/square root (of the number of original observations)).

In Frost and Kneen, 2006, a statistical method was developed to test the accuracy of the predicted yield values within 0.5 ton/ha, 1.0 ton/ha, 1.5ton/ha etc. This however involved removing all objective yield points with farmer observed yield above 6 ton/ha. Since these values are achievable in some parts of the maize producing areas in South Africa, this method is not preferred. Many alternative statistical methods of model verification could also have been used on the data.

4.3 MODIS NDVI Data Analysis

4.3.1 Standard deviation in high yield dry land average NDVI data

Before one can use the graphs of the average NDVI and rainfall values (Figure 22 to Figure 26) for maize yield prediction, the overlap between the three yield classes in the specific dekad used to make the final maize yield prediction has to be considered. The standard deviations and the minimum and maximum values depicted on the graphs in Figure 16 to Figure 21 (NDVI) and Figure 28 to Figure 33 (rainfall) give an indication of the possibilities of the overlap in NDVI and rainfall values between the three classes.

By using the graphs of the standard deviations a visual assessment can be made to establish the amount of overlap that may occur in the dekad that the user has chosen to make the yield prediction and thus use the data with caution.

Objective yield points in the dry-land high yield class had final yield figures of up to 12 ton/ha. Some of these high figures could be due to recording errors, but are not impossible. The standard deviation of high yield analysis data is highest around day 10 and day 170 after planting (Figure 16). The standard deviation in the high yield NDVI data is greatest in the beginning and at the end of the growing season.

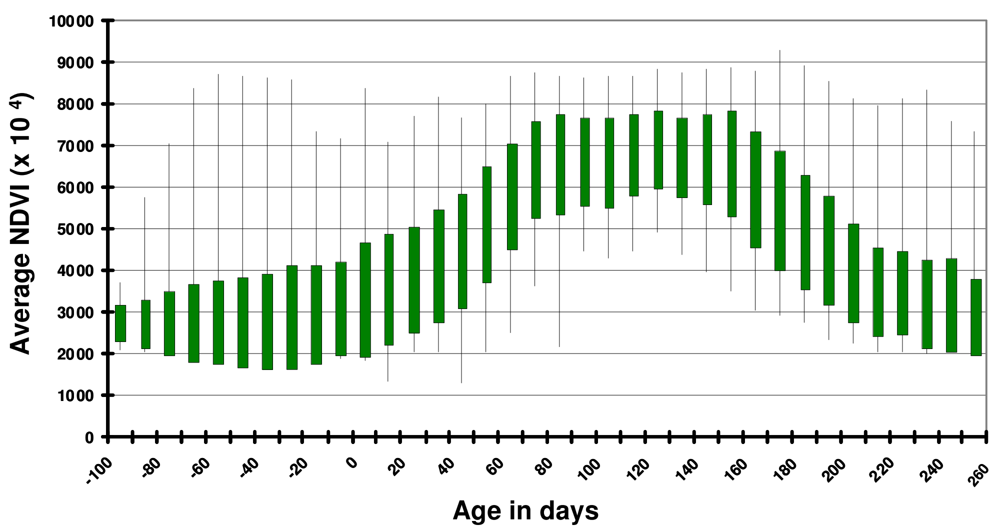


Figure 16 2001/2002 growing season high yield analysis data: Standard deviation

High yield test data show highest standard deviation around day 50 and day 150 (Figure 17). High standard deviation in NDVI values at these stages could be due to interpolation of missing planting dates, low average minus standard deviation-, high average plus standard deviation-, high maximum- (dekad 15) and low minimum (dekad 5) values.

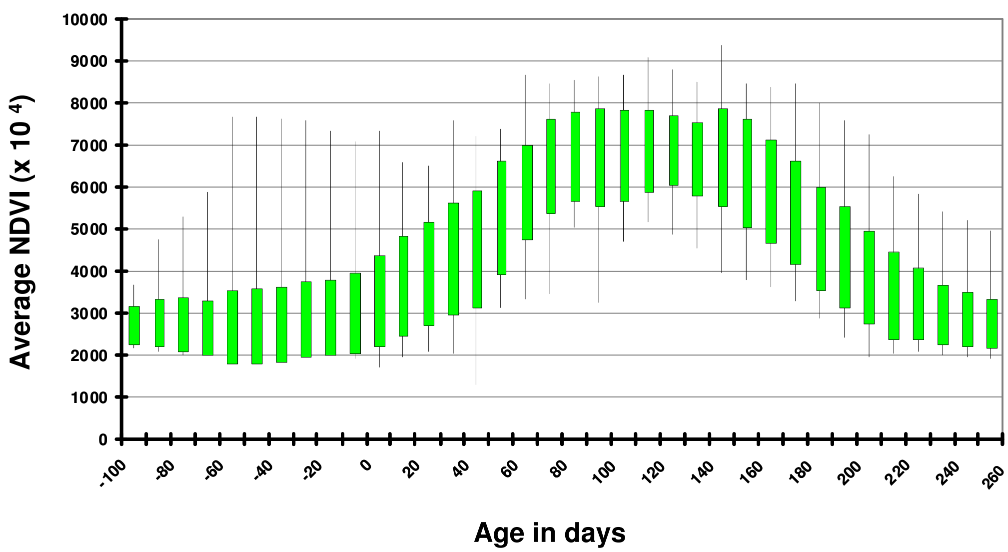


Figure 17 2001/2002 growing season high yield test data: Standard deviation

NDVI generally starts to decrease around day 130 after planting for high yield (perhaps longer growing) cultivars (Figure 23). The smallest standard deviation occurred around day 130 when the kernels begin to mature physiologically. Representing different (short-, medium- or long growing season) cultivars could be the reason the NDVI values differ greatly around day 160 after planting.

4.3.2 Standard deviation in medium yield dry land average NDVI data

The observed standard deviation in the medium yield analysis data (Figure 18) seem to differ less from one (1) dekad to the next than that of the high yield analysis data. The standard deviation in the medium yield analysis data is larger in the beginning (dekads 1 to 5) and end (dekads 12 to 18) of the growing season. Fewer points fall into the medium yield class than the low- and high yield classes. Yield class classification was done arbitrarily and different final yield boundaries could be explored which might yield slightly different results.

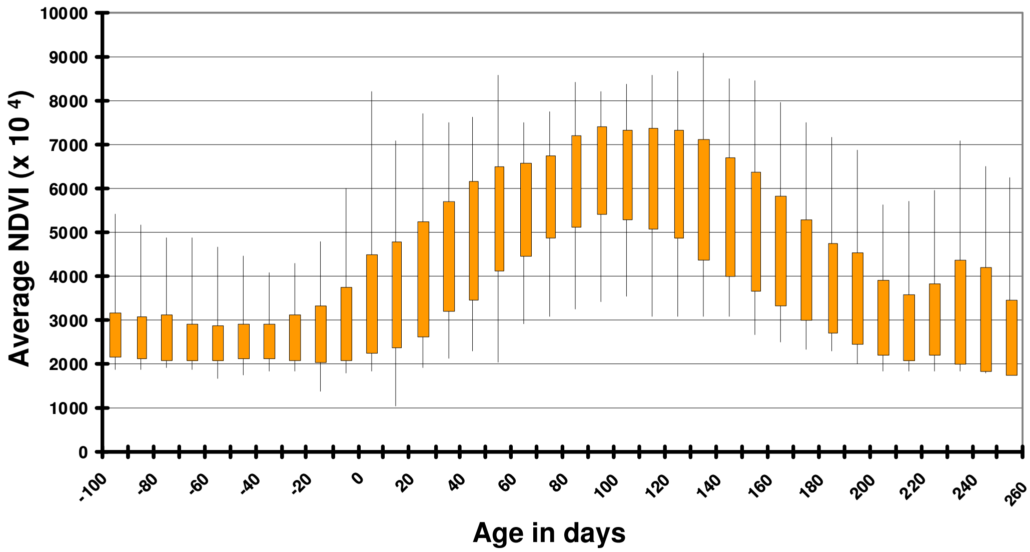


Figure 18 2001/2002 growing season medium yield analysis data: Standard deviation

The standard deviation for the medium yield test data (Figure 19) follows a similar distribution to that of the analysis data (Figure 18). From 30 days prior to the growing season (medium yield analysis data) and from 50 days prior to the growing season (medium yield test data) the standard deviation in the data gradually increases up to about 50 days (medium yield analysis data) and 60 days (medium yield test data) after planting. This trend in the medium yield data is different to that of the high- and low yield datasets where standard deviation in the data is high from about 60 days prior to planting day right up to about 60 days after planting date.

The standard deviation in the NDVI data is the smallest around day 70 after planting in both the medium yield analysis and test data. The standard deviation in the data is the highest at the end (day 120 to 180 after planting) of the growing season around day 160 after planting in both the medium yield analysis and test datasets.

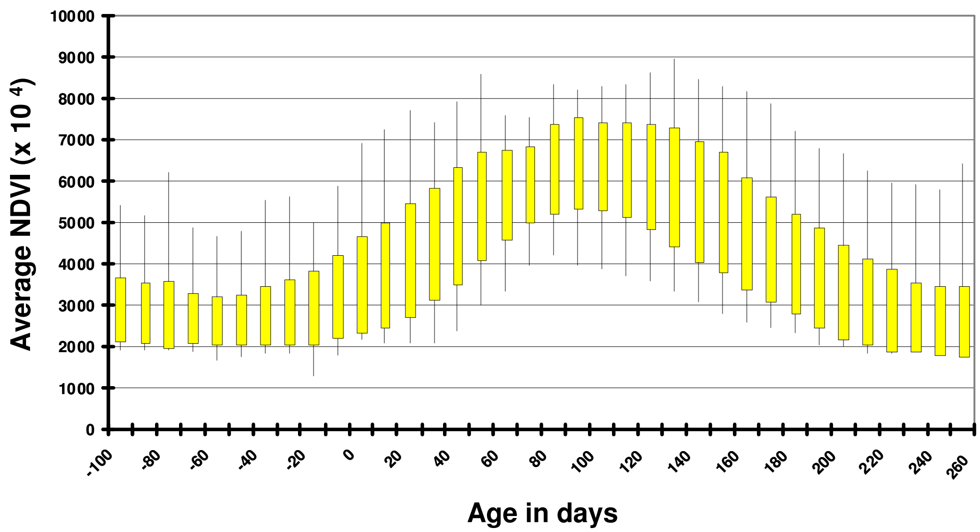


Figure 19 2001/2002 growing season medium yield test data: Standard deviation

4.3.3 Standard deviation in low yield dry land average NDVI data

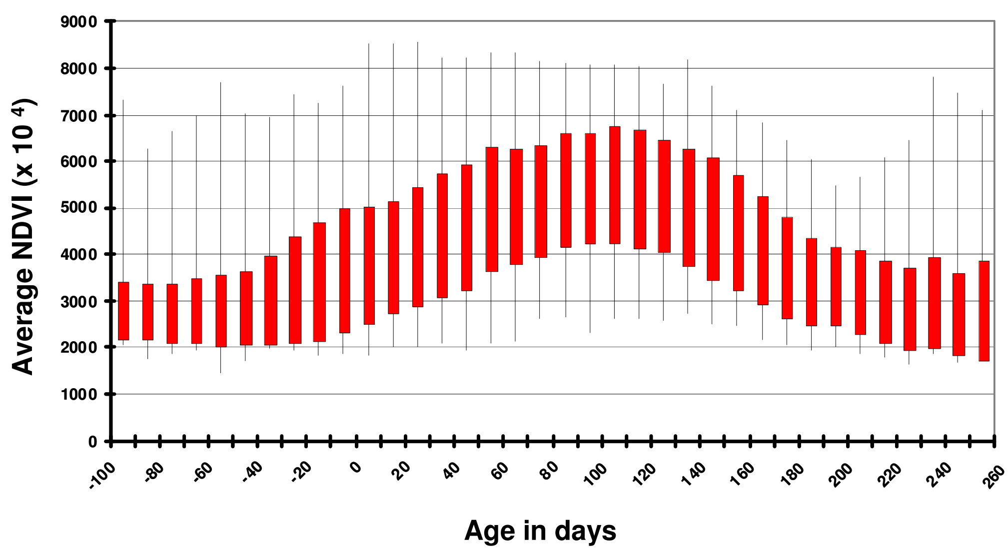


Figure 20 2001/2002 growing season low yield analysis data: Standard deviation

Low yield analysis data (Figure 20) depict an average standard deviation of around 3000 NDVI x 10^4 units throughout the growing season. The standard deviation pattern is different from that of the low yield test data (Figure 21), which have small (1000 NDVI x 10^4) to large (3000 NDVI x 10^4) standard deviations in the data.

The small standard deviations observed in the low yield test data standard deviation dataset occur outside the growing season (day -100 to day -30 and day 210 to day 260 after planting).

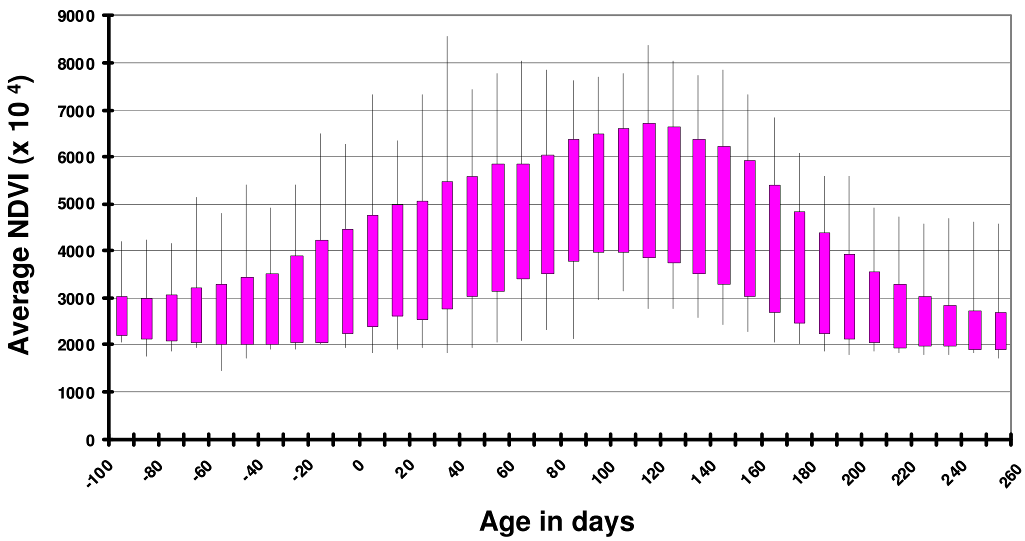


Figure 21 2001/2002 growing season low yield test data: Standard deviation

The low yield test data show a more varied standard deviation throughout the growing season than the low yield analysis dataset and the standard deviation varies more from one dekad to the next than that of the low yield analysis dataset.

Low-, medium- and high yield analysis and test datasets have a standard deviation in their NDVI values of up to 3000 NDVI x 10^4 units. This high standard deviation causes overlap of average values between the three classes, which accounts for some of the inaccurate predictions.

4.3.4 Using average NDVI data for maize yield prediction under dry land conditions

Using average NDVI curves for yield prediction is 69 % accurate within one (1) standard deviation of the predicted class. Using average NDVI curves to predict yields correctly within a 95 % confidence interval will be achieved only 12.64 % of the time. The average NDVI test and NDVI analysis data curves (Figure 22, Figure 23 and Figure 24) follow similar trends, although their standard deviations differ slightly as discussed.

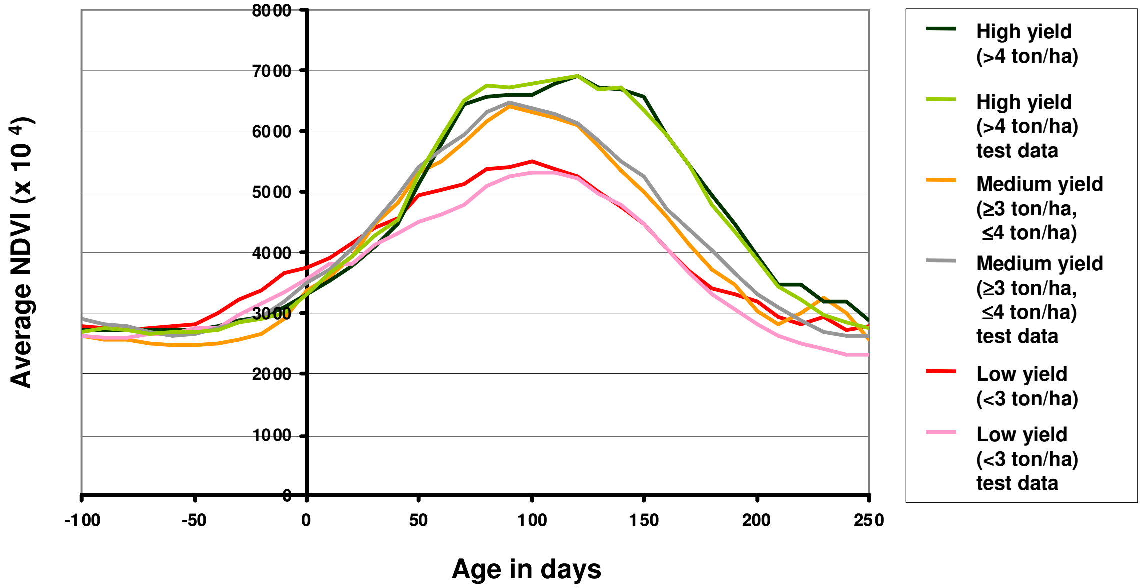


Figure 22 2001/2002 growing season low-, medium- and high yield: Average NDVI dataset and test dataset

Visual inspection of Figure 23 (section 4.3.4) seems to indicate that the average NDVI values between dekad 7 and dekad 15 could be used to predict the final yield class. But upon visual inspection of the six (6) graphs that depict standard deviation in NDVI data (Figure 16 to Figure 21) and the six (6) graphs that depict standard deviation in the rainfall data (Figure 28 to Figure 33) there seems to be an inverted trend between the NDVI and rainfall data sets. Where the standard deviation in the NDVI data sets increase as the yield decreases, the standard deviation in the rainfall data sets decrease as the yield decreases. This has the implication that the time period (dekad) used for maize yield prediction and the yield class (low, medium or high) will influence which average NDVI and rainfall graphs should be better suited for the specific prediction purposes.

Keeping in mind the standard deviation in each dataset, the average NDVI curves seem to be able to predict low yield from medium-/high yield as early as 50 days from planting day (line at day 50, Figure 23). From as early as 70 days after planting, the average NDVI curves are able to predict and discriminate between low-, medium- and high yield effectively up to 180 days after planting (magenta window period 1). Fields might already have been harvested at 180 days after planting, depending on the kernel moisture content at which the farmer wishes to harvest the maize.

Day 180 is roughly 40 days after the maize kernels have matured and lost about 20 % of their moisture content. The black prediction window period (Figure 23) can be used to predict high- from medium-/low yield from dekad 19 to dekad 23.

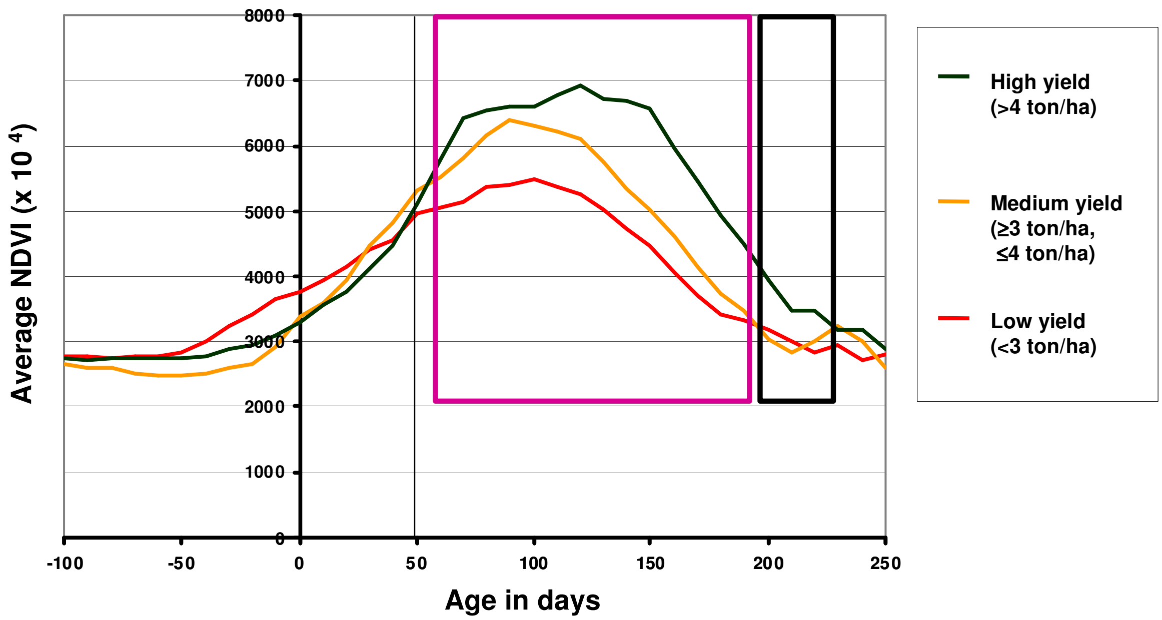


Figure 23 2001/2002 growing season low-, medium- and high yield: Average NDVI dataset

NDVI curves can be analysed from day 50 to around day 220 after planting for monitoring purposes. Between day 60 and day 180 after planting, these average NDVI curves are sufficiently accurate for yield predictions in the three different categories. After day 150, the accuracy of predicting correctly within one (1) standard deviation of the category, drops gradually.

Using the test data objective yield points (Figure 24), the three average NDVI curves indicate that one could distinguish between high-/medium- or low yields as early as 20 days after planting to 50 days after planting (black prediction window 1), after which the three classes of yield data seem to be predictable from day 50 to day 250.

Average NDVI curves created from the test dataset appear different to the average NDVI curves created from the analysis dataset, because fewer points were used (179 points) to create the test data curves than the analysis data curves (360 points).

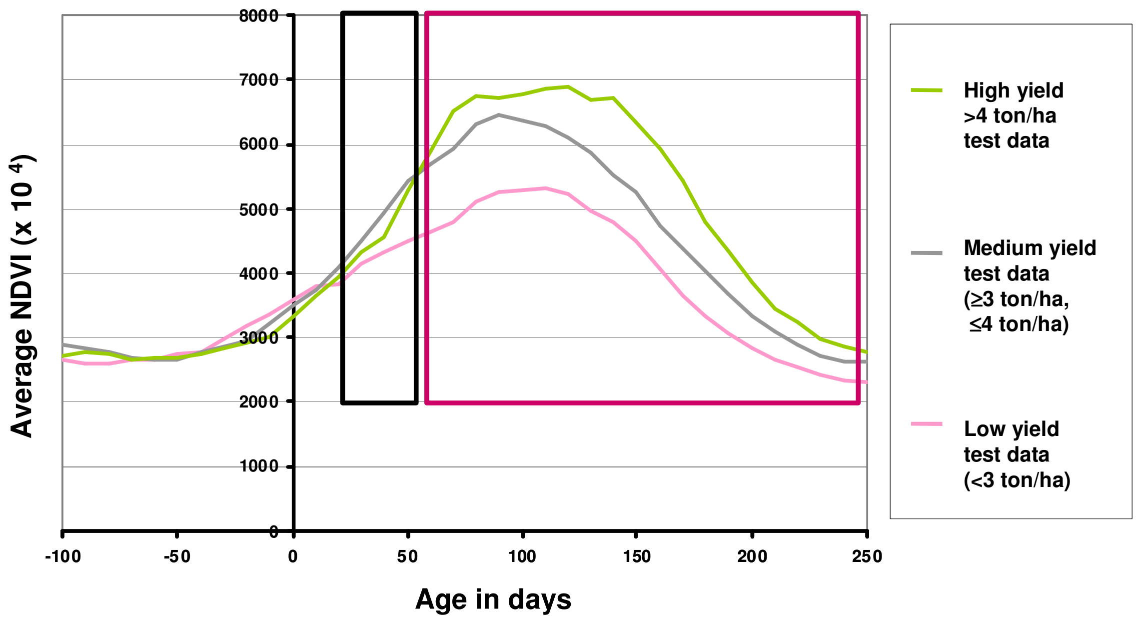


Figure 24 2001/2002 growing season low-, medium- and high yield test data: Average NDVI

The model verification methods mentioned in section 4.2.3 include the visual verification method illustrated below. The prediction windows and the average NDVI curves in Figure 23 are used. Three (3) of the objective yield points from the test data set of 179 test points are used to test whether the yields predicted from the graph were correct or close to the predicted value. Alternatively, several statistical methods could also have been applied to test the difference in predicted vs. actual yield involving residuals etc.

The NDVI value extracted at a specific dekad, for example, dekad 8 in these three test cases, was pinpointed on the average NDVI graph in Figure 23, and the expected end yield was read from the average curves. To assist in reading yields off the average NDVI curves, the average yield for each class was calculated (high yield curve average yield: 5.12 ton/ha, medium yield curve average yield: 3.54 ton/ha and low yield curve average yield: 2.16 ton/ha).

- Point 558, day 80, NDVI = 5397, recorded objective yield = 2.0 ton/ha. Predicted yield: 2.10 ton/ha
- Point 52, day 80, NDVI = 6205, recorded objective yield = 3.5 ton/ha. Predicted yield: 3.54 ton/ha

- Point 380, day 80, NDVI = 6504, recorded objective yield = 5.0 ton/ha. Predicted yield: 4.85 ton/ha

The results seem promising but, due to standard deviations previously discussed, not all test points will give such close results to the final farmer-observed yields. From section 4.3.1 to section 4.3.3, it is evident from the standard deviations in the data (average +/- one (1) standard deviation) and the outlier values present, that the average NDVI graphs must be used with caution when making final maize yield predictions in dekad 8 after planting.

4.3.5 Using three dekadal moving average NDVI data for maize yield prediction under dry land conditions

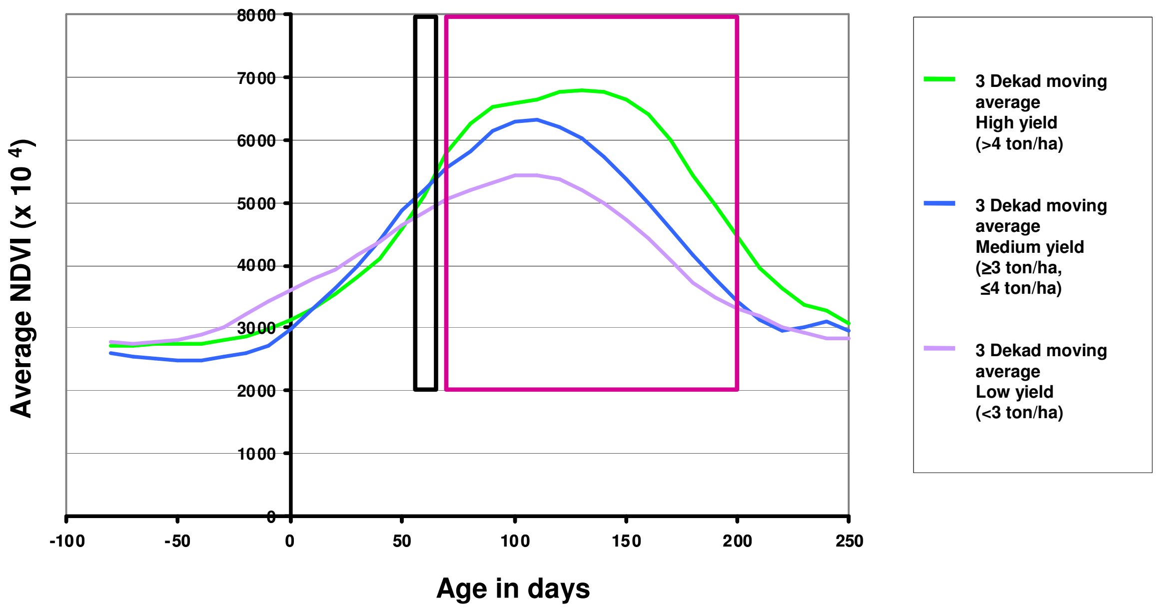


Figure 25 2001/2002 growing season three dekadal moving average: NDVI curves

To investigate the effect that the use of monthly NDVI data would have on maize yield predictions, incidentally a smoothing effect on the curves result. Three dekadal moving NDVI averages (which can be related to monthly data) were used to create the curves in Figure 25. The first (black) window period in which a low-or medium-/high yield class can be identified is from day 60 to day 70. The second (magenta) window period is from day 80 to day 200. During the second window period, low-, medium- or high yield classes can be distinguished.

These results indicate that using monthly NDVI data to distinguish between low-, medium-or high yield classes would result in similar NDVI window periods to those observed for 10-day (dekadal) predictions. Some experts advise that monthly NDVI composites would be more reliable for summer crop prediction as the number of cloud-free days should be greater.

A simultaneous view of the average and moving average NDVI curves (Figure 26) illustrates that the moving average curves tend to be lower than the average NDVI curves up to a point around the maximum average NDVI value. From that point onwards, the three dekadal moving average curves tend to be higher than the average curves. For high yields, this point is around day 130. For medium yields, the turning point is around day 100, whilst the turning point for low yield data points seems to be around day 110.

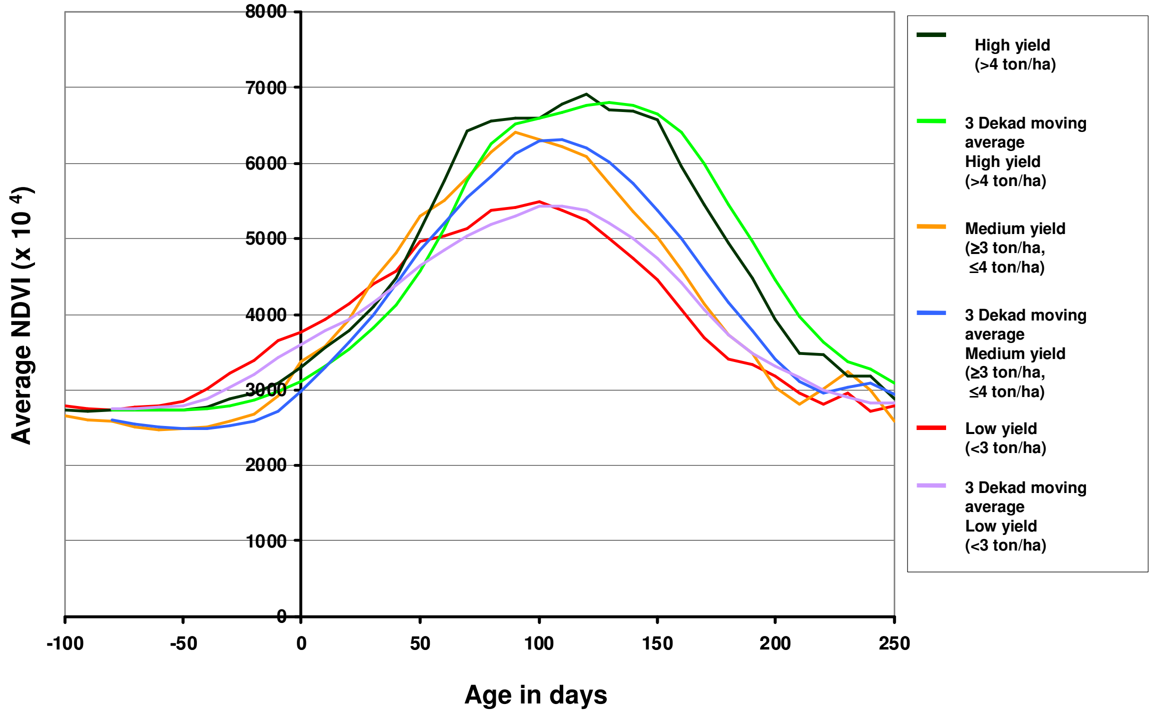


Figure 26 2001/2002 growing season: Three dekadal moving average NDVI curves and average NDVI curves

4.3.6 Using average NDVI data for yield prediction under irrigation conditions

Results obtained by analysing the data from irrigated fields are as a consequence of interference in the infra-red band causing bias, due to imagery acquired over recently irrigated fields. It is proposed and found that using the MODIS Enhanced Vegetation Product (EVI) instead of MODIS NDVI gives much better results (Huete *et al.*, 2005). From this research, it is suggested that for yield analysis the EVI product, which was received with the NDVI product, be processed and the EVI values extracted for irrigated objective yield data points.

Mixed NDVI pixels might also have been the cause of the pattern in Figure 27. These mixed NDVI pixels could be due to the size and orientation of the irrigated maize fields and other irrigated crops or land-cover classes (Thompson, 1995) in the vicinity relative to the 250 m MODIS pixels. Other aspects relating to the patterns in Figure 27 are discussed in the Conclusions (Chapter 5).

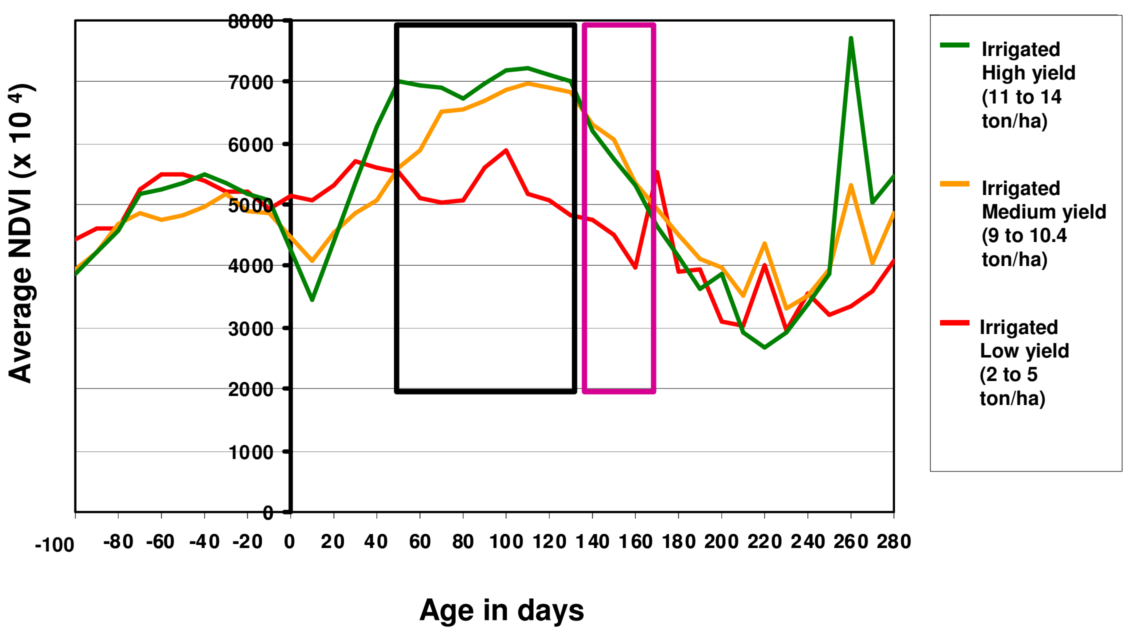


Figure 27 2001/2002 growing season irrigated objective yield points: Average NDVI curves for low-, medium- and high yields

The average NDVI curve of the low yield irrigated class was highest at planting day up to 20 days after planting, followed by the second highest average NDVI value for the medium yield class and the lowest average NDVI value for the high yield class at planting day. This prevails up to dekad 2 after which the average NDVI value of the high yield class gradually starts to surpass the medium class at day 20 and low yield class at day 30.

The average NDVI curves of the low yield dry-land analysis data (Figure 23) follow a similar trend to that of the low yield irrigated average NDVI curve (Figure 27) in that these average NDVI values were higher at planting day up to day 30 after planting than those of the medium- and high yield classes. After day 30 the average NDVI value of the dry-land medium yield class, starts to become larger than that of the dry-land low yield class. At day 40 the low- and high yield class average NDVI values are indistinguishable and only by day 50 the average NDVI value of the high yield class starts to surpass that of the low yield class for the dry-land data (Figure 23).

These phenomena raise the question whether higher yielding cultivars grow longer (have a longer growing season and/or mature later) than medium- and low yield cultivars. This question requires further investigation.

4.4 Rainfall Data Analysis

4.4.1 Standard deviation in high yield dry land average rainfall data

Rainfall has a high variability in South Africa and differs greatly depending on where in the country the objective yield GPS point lies. Generally, rainfall decreases as one moves from the north-eastern to south-western parts of the country. It is recommended that average rainfall curves, which relate to a specific yield class, be developed for each of the planting date zones. The standard deviation per planting zone might be less than when objective yield points are evaluated together as if homogeneous.

The standard deviation in the rainfall data (Figure 28 to Figure 33) should be kept in mind when making predictions by using average rainfall curves created from high-, medium- and low yield data. It should also be remembered that average rainfall minus the standard deviation totalling <0 mm is not displayed in Figure 28 to Figure 33.

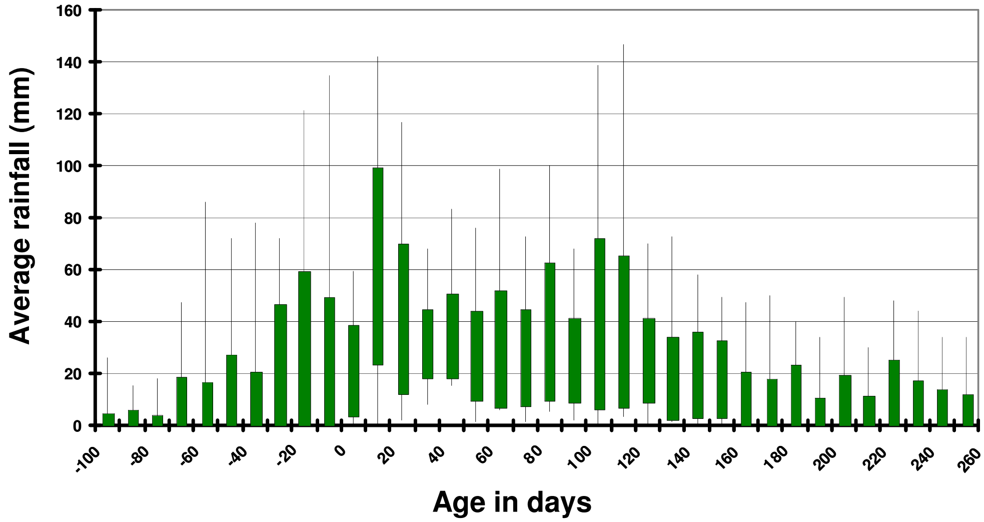


Figure 28 2001/2002 growing season high yield analysis data: Standard deviation in rainfall data

The results from Figure 28 indicate that the high maximum rainfall values have contributed to the higher than average rainfall for high yields, and the greater standard deviations associated with these dekads (dekad 2, dekad 11 and dekad 12).

Using real-time rainfall values to make accurate yield predictions are, for this reason, less successful than using real-time NDVI values to predict maize yield.

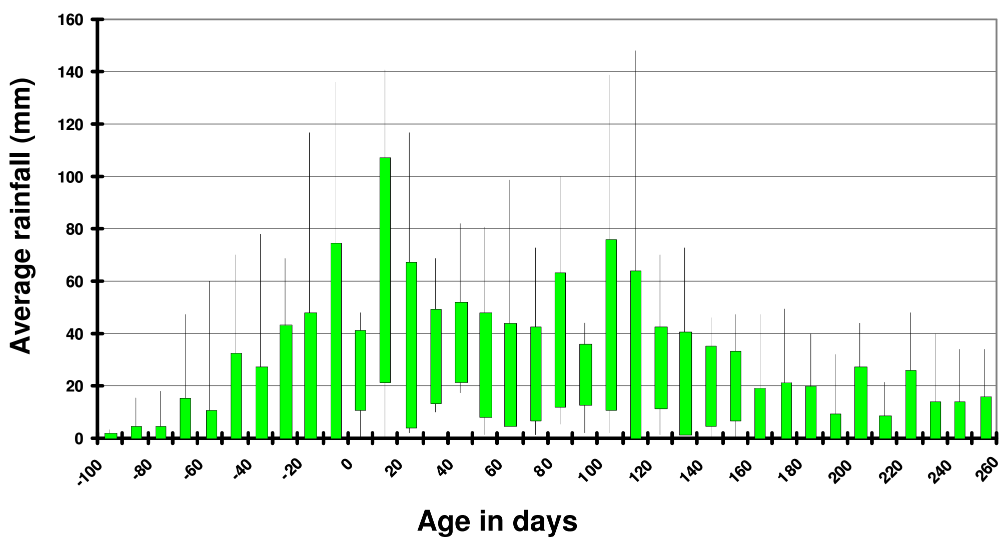


Figure 29 2001/2002 growing season high yield test data: Standard deviation in rainfall data

The high yield test data (Figure 29) display a similar pattern of standard deviation, and periods of higher rainfall (around dekad 0, dekad 2, dekad 11 and dekad 12) and lower rainfall (dekad 1 and dekad 9) are clearly visible.

4.4.2 Standard deviation in medium yield dry land average rainfall data

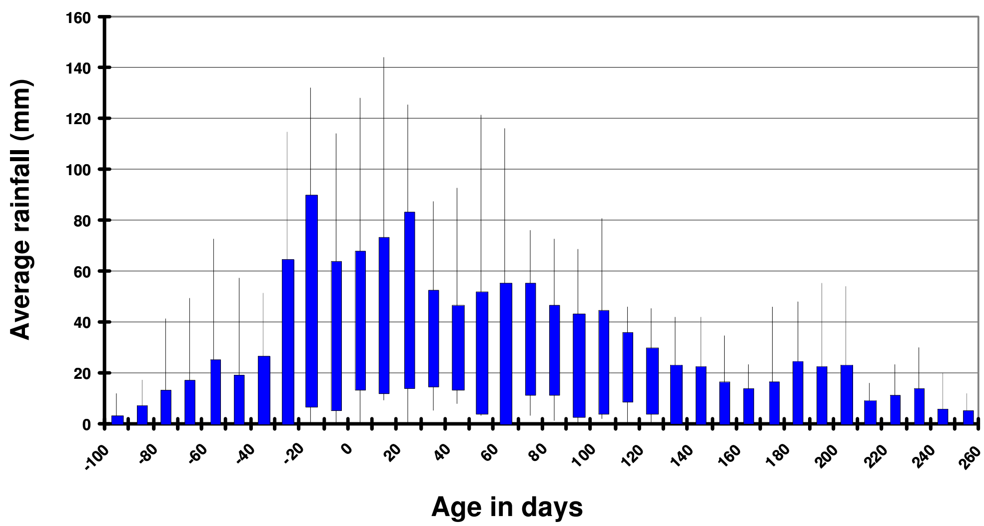


Figure 30 2001/2002 growing season medium yield analysis data: Standard deviation in rainfall data

The standard deviation in rainfall of the medium yield dry-land analysis and test data (Figure 30 and Figure 31) is less variable than that of high yields. The period known as "midsummer drought", which occurred in the 2001/2002 growing season between day 30 and day 80, are visible on the low-, medium- and high rainfall standard deviation plots.

The medium yield test data standard deviation follows a similar pattern to that of the medium yield analysis data.

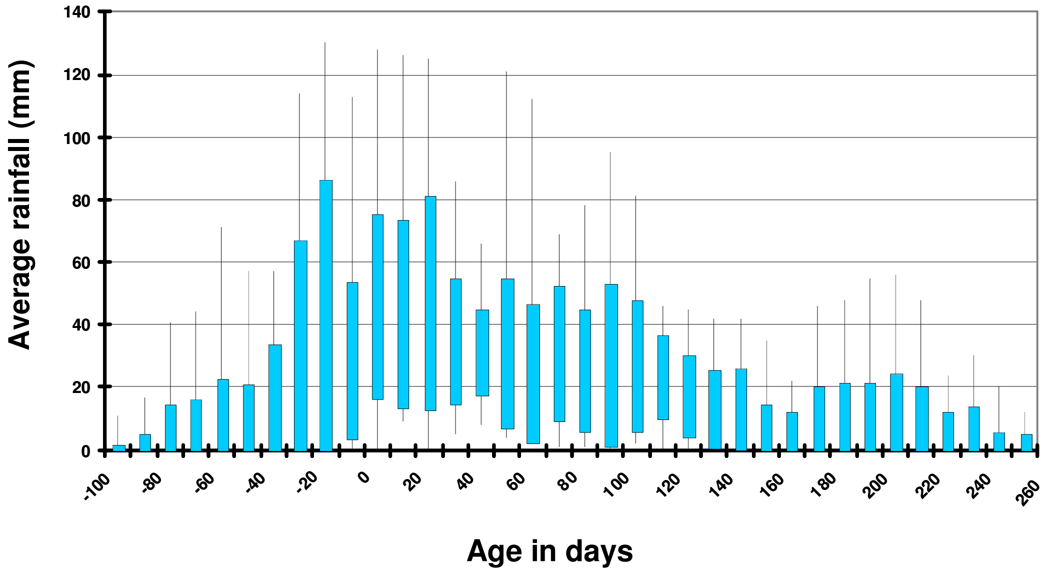


Figure 31 2001/2002 growing season medium yield test data: Standard deviation in rainfall data

4.4.3 Standard deviation in low yield dry land average rainfall data

Closer inspection of the standard deviation in medium- and low yield average rainfall analysis and test data, starts to reveal the trends for the different yield classes visible during the midsummer drought season. Between day 30 and day 80 after planting, the high yield standard deviation data graphs (Figure 28 and Figure 29) reveal that average rainfall values were higher than for the medium yield (Figure 30 and Figure 31) and low yield (Figure 32 and Figure 33) rainfall data in the same period. This phenomenon is evident also from the average rainfall data analysis and test plots (Figure 34 to Figure 40). The difference in the average dekadal rainfall between day 75 and day 130 reveals the reason for the difference in final yield.

Low yield standard deviation curves (Figure 32 and Figure 33) show high standard deviation when rainfall was high, and lower standard deviations during dekads with lower rainfall. On inspection, low yield average rainfall plus the standard deviation and average rainfall minus the standard deviation, reveal bars that are more equal in size than is the case for high- and medium yield bars. Thus the standard deviation for low yield average rainfall analysis and test data is less variable than that of the medium- and high yield classes.

The same phenomenon is also visible in the average plus- and average minus the standard deviation for test rainfall data in Figure 33 and the standard deviation in average NDVI data between the different yield classes (Figure 16 to Figure 21).

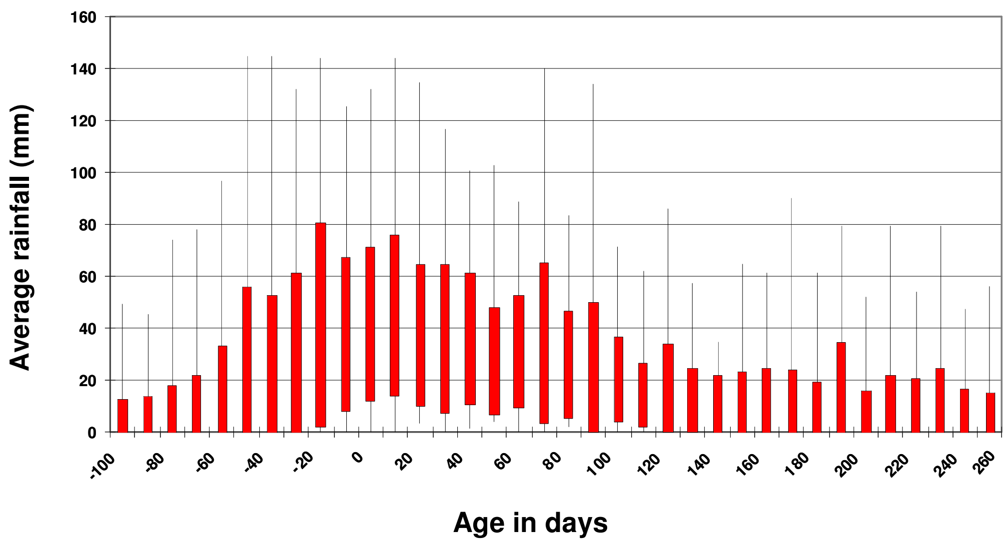


Figure 32 2001/2002 growing season low yield analysis data: Standard deviation in rainfall data

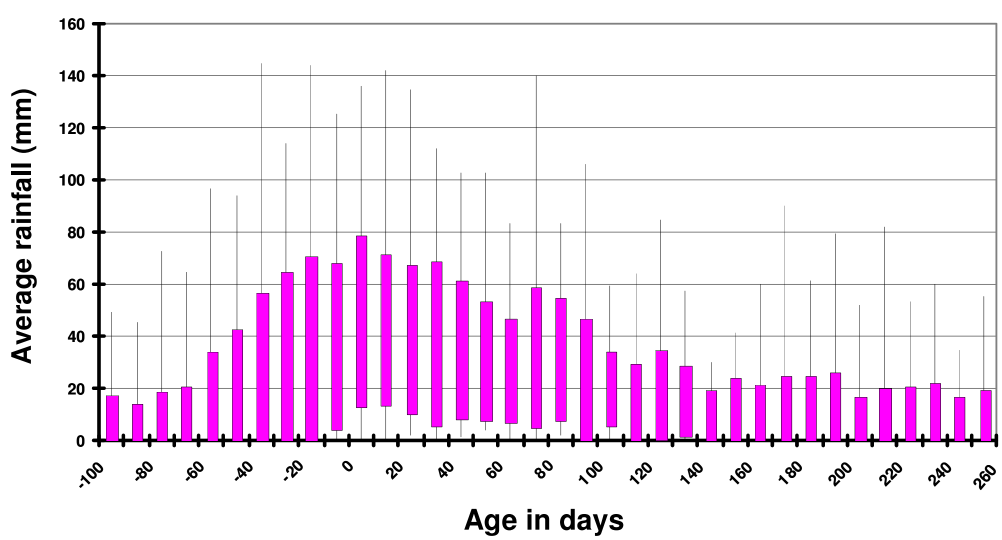


Figure 33 2001/2002 growing season low yield test data: Standard deviation in rainfall data

The low yield rainfall test data (Figure 33) also display a similar trend to the standard deviation in analysis data, in that there is a steady decline in the amount of maximum rainfall received per dekad from day 30 to day 70. This can be seen by the steady decline in the average minus- and average plus the standard deviation bars. The medium yield rainfall data displayed a steady average rainfall in this period, while the high yield data showed more increases than decreases during the same period (day 30 to day 70).

4.4.4 Using average rainfall data for maize yield prediction under dry land conditions

Visual interpretation from Figure 35 and Figure 41 seem to indicate that between dekad 9 and dekad 12 (day 90 to day 120 after planting), the average rainfall curves can be used to predict yield. From the three dekadal moving average graphs (Figure 34 and Figure 36), dekads 9 to 14 could be used to predict high yields from medium and low yields or visually make predictions in ton/ha. However, between dekads 8 and 14, the standard deviation graphs (Figures 28, 30 and 32) should be used in conjunction with the average rainfall curves to establish the likelihood of overlap in values between the classes.

For example, an objective yield point from the low yield test data set happened to have had 60 mm of rainfall between dekad 10 and dekad 11 (Figure 33). Although the average rainfall received between dekad 10 and dekad 11 in the low yielding maize class was 15 mm (Figure35), rainfall values as high 70 mm were experienced for low yield objective yield points between dekad 10 and dekad 11 (Figure 32). The statistical quantification method described in the next two paragraphs, revealed that 78.3% of the low yield data points had rainfall values within the standard deviation range (avg. +/- 1 standard deviation). The standard deviation graph reveals that those points experienced a rainfall of between 4 mm and 38 mm (avg. +/- 1 standard deviation.). Thus 22% of the test data points could have on average experienced rainfall outside the 'average +/- 1 standard deviation' range. The average and standard deviation graphs should thus never be used in isolation from each other to predict maize yield during any dekad.

After statistical quantification was done on the low-, medium and high yield test data NDVI and rainfall values, the following results were obtained. When using NDVI to predict within one (1) standard deviation in the low yield class, a 65.86 % success rate was achieved while a 70.83 % and 70.24 % success rate was achieved in the medium- and high yield classes respectively.

Using spatial rainfall values to predict final yield was 78.30 %, 80.18 % and 79.76 % successful within one (1) standard deviation of the low, medium and high yield classes overall. The high yield classes were cut off at >6 ton/ha, so the predictions in the high yield classes were more successful than the figures convey.

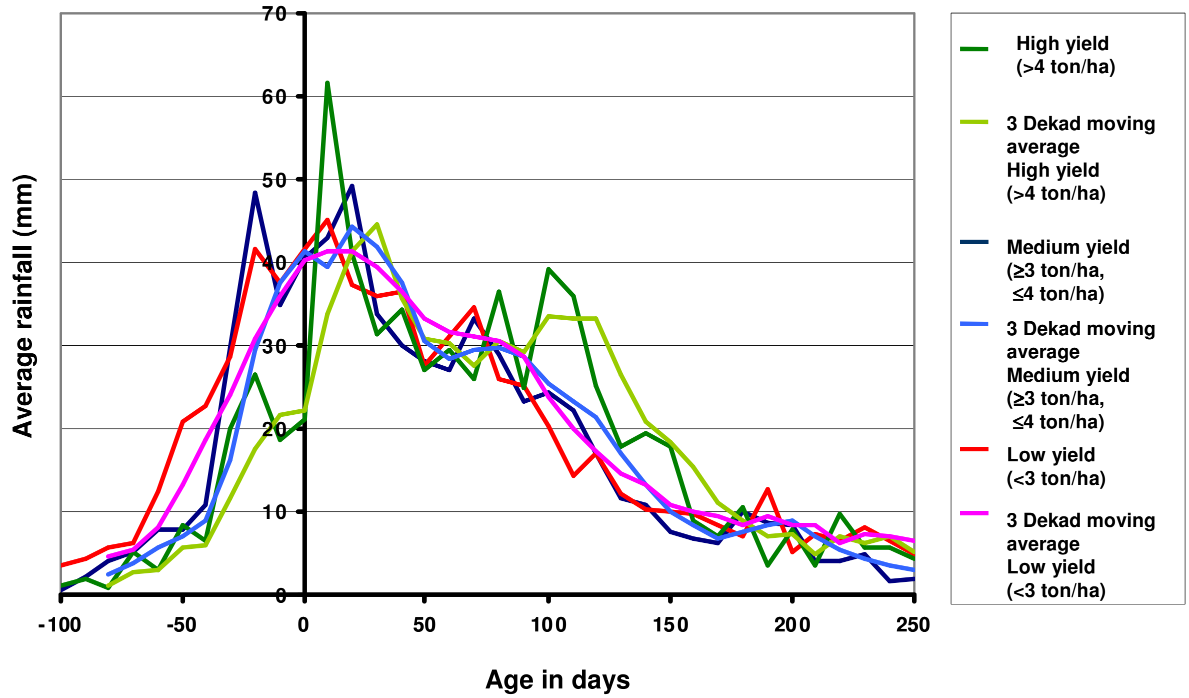


Figure 34 2001/2002 growing season low-, medium- and high yield: Average rainfall data and three decadal moving average rainfall data curves

It is difficult to follow the individual trend lines of the high-, medium- and low yield average and three dekadal (monthly) moving average rainfall curves, when displayed on the same graph as in Figure 34. By displaying average and moving average curves on separate graphs (Figure 35 and Figure 36 respectively) it is possible to see how the prediction windows differ. From dekad 2 to dekad 8, rainfall curves cannot be used to predict high-, medium- or low yield. Points reflected in the high yield rainfall data curves seem to have received less rain on average during this period than low- or medium yield rainfall curves.

The higher amount of rainfall received during dekad 8 to dekad 14 after planting (Figure 34), seemed to make all the difference between a high yield and a medium-/low yield. Three crucial prediction window periods fall in this period (Figure 35).

The first black window period (from dekad 8 to dekad 9) in Figure 35 is a crucial period in the development of the maize plant. All the leaves have just fully unfolded and pollen shed starts. During the second magenta window period (3rd prediction window) in Figure 35, (Stage 5, section 4.2) in dekad 9, the kernels begin to develop and reach maturity (to dekad 13). It is understandable that this main yield prediction window from average rainfall might occur in this period.

The fourth (4th) prediction window (second black window), from dekad 12 to dekad 15, makes use of the average rainfall curves to predict final yield in the high yield or low-/medium yield categories. It seems a less influential time period to use for prediction, as rainfall is no longer utilised for biological growth of the kernels during this stage, but the curves seem representative, nevertheless. This might not be the case for dry- or wet years.

The conclusion reached is that the crucial rainfall periods (between day 80 and day 130 after planting) are more significant than the cumulative rainfall over the growing period and also more significant than the rainfall received between dekad 3 and dekad 6 after planting. Figure 35 and Figure 36 each show the second magenta prediction window that can be used for yield prediction during the crucial period of rainfall (dekad 8 to dekad 12).

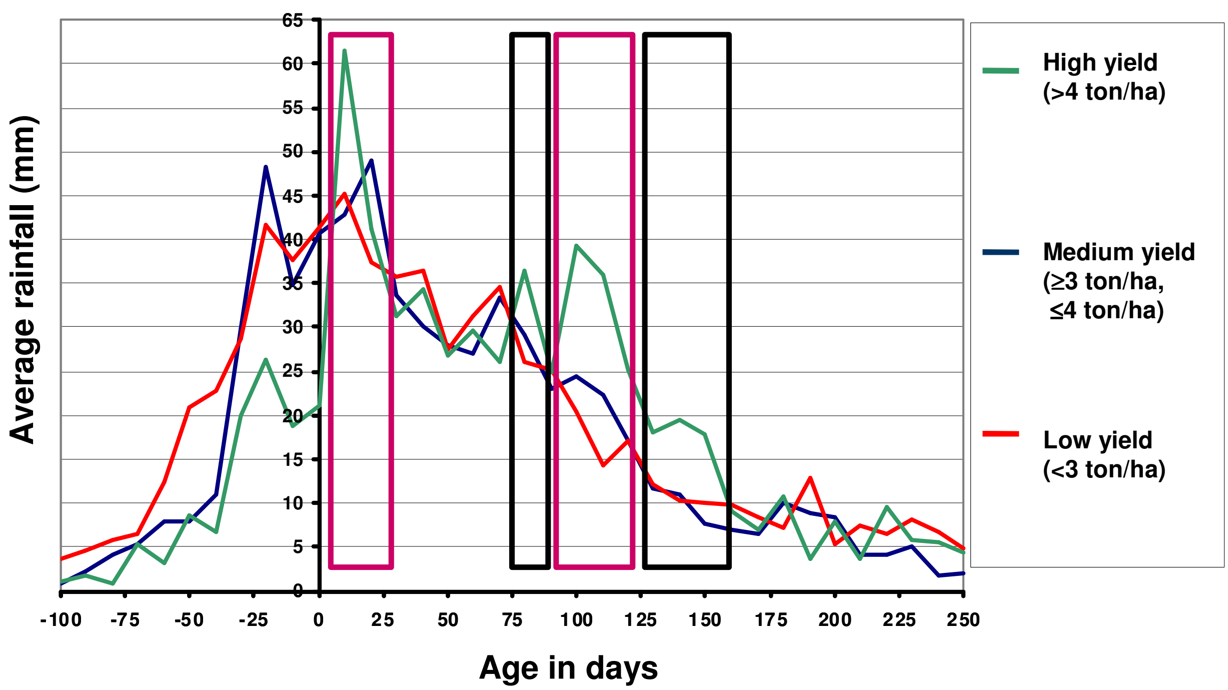


Figure 35 2001/2002 growing season low-, medium- and high yield: Average rainfall data curves

The pattern that the three yield classes follow for the average MODIS NDVI values in Figure 35, is different to the pattern that the three (3) yield classes follow in Figure 23, where average rainfall values are depicted. The average NDVI curves follow a smoother almost normal curve, and for a great deal of the growing season (dekad 6 to dekad 19) the high, medium and low yield curves are in that ranking order (Figure 23). The average rainfall curves in Figure 35, are less smooth, and in a small time span, the ranking order could change and even reverse with regard to the high, medium and low yield classes.

During dekad 3 to dekad 7 (Figure 35), the lower yield class actually received a higher average rainfall than the medium and high yield classes. Whilst the average NDVI curves reflect this phenomena in average rainfall during dekads three (3) to five (5), by dekad seven (7) the average NDVI curves have clearly started to take on the correct ranking pattern (Figure 23 and Figure 41).

There seems to be a distinct pattern for each of the yield classes from day 0 (planting day) up to day 30 after planting (dekad 3). During this period, it seems possible to predict high yields from medium- or low yields from the average rainfall curves (magenta prediction window 1, Figure 35). However, during this first month after planting, the average rainfall received by the three classes seem almost volatile, in that it changes from dekad to dekad, and no real clear decision can be reached by visual interpretation only.

4.4.5 Using three decadal moving average rainfall data for maize yield prediction under dry land conditions

The moving average curves (Figure 36) can be used from planting day up to dekad 3 to predict high- or medium-/low yield under dry-land conditions, and between dekad 3 and dekad 4 to predict high-, medium- or low yield.

Using either dekadal or monthly rainfall data to make final yield predictions makes a difference to which yield classes can be predicted and to the position of the prediction window period. So this window period (first magenta prediction window (Figure 35) and dekad 1 to dekad 4 (Figure 36)), should be treated with extreme caution for prediction purposes.

Dekads 1 to 4 seem to indicate that high yield classes received less rainfall during the first two dekads than the medium- and low yield class points (from Figure 36) and less rainfall than the low-/medium yield classes between day 0 and day 10 (Figure 35), but more rainfall than the medium- and low yield classes between day 10 and day 20 (Figure 35).

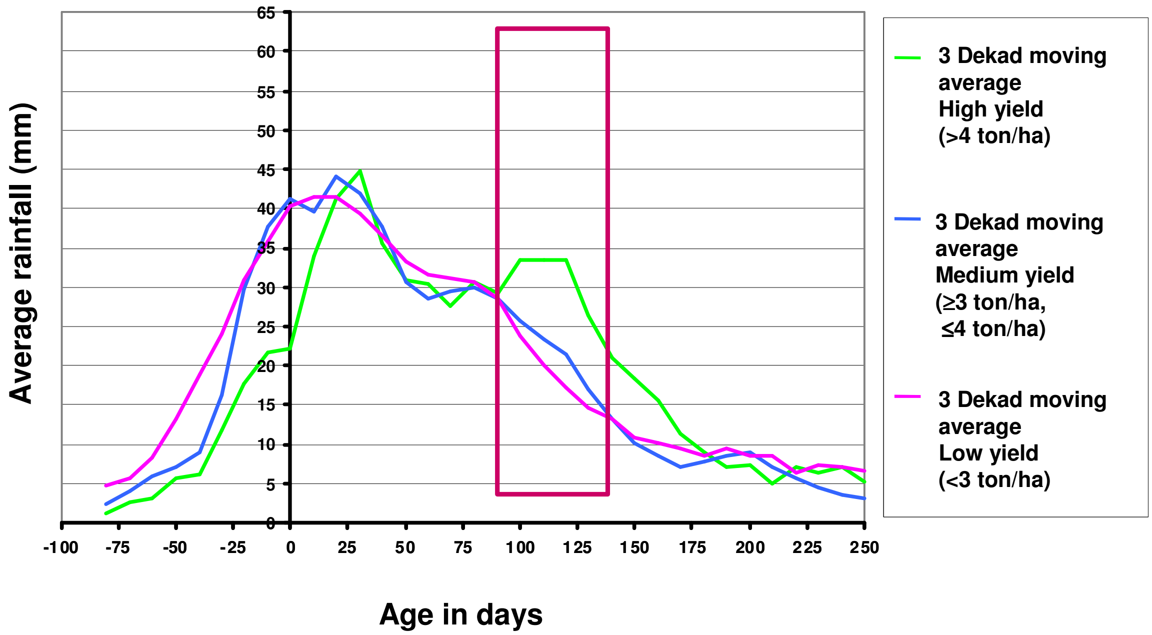


Figure 36 2001/2002 growing season low-, medium- and high yield: Three dekadal moving average rainfall data curves

The second prediction "window period" consists of two (2) prediction windows (on the average rainfall data curves) between day 80 to day 120 (Figure 35). This prediction "window period" is represented by the magenta prediction window (on the moving average rainfall data curves) between day 80 to day 140 of Figure 36. This "window period" can be used to predict low-, medium- or high yield.

The information that is lost by using monthly average (that is, three dekadal moving average) rainfall data (Figure 36) rather than average dekadal rainfall data (Figure 35), is the high amount of rainfall that defined high yield from medium- and low yield maize crops during dekad 7 and dekad 8 (prediction window 2 in Figure 35).

4.4.6 Comparing average rainfall analysis data and test data under dry land conditions

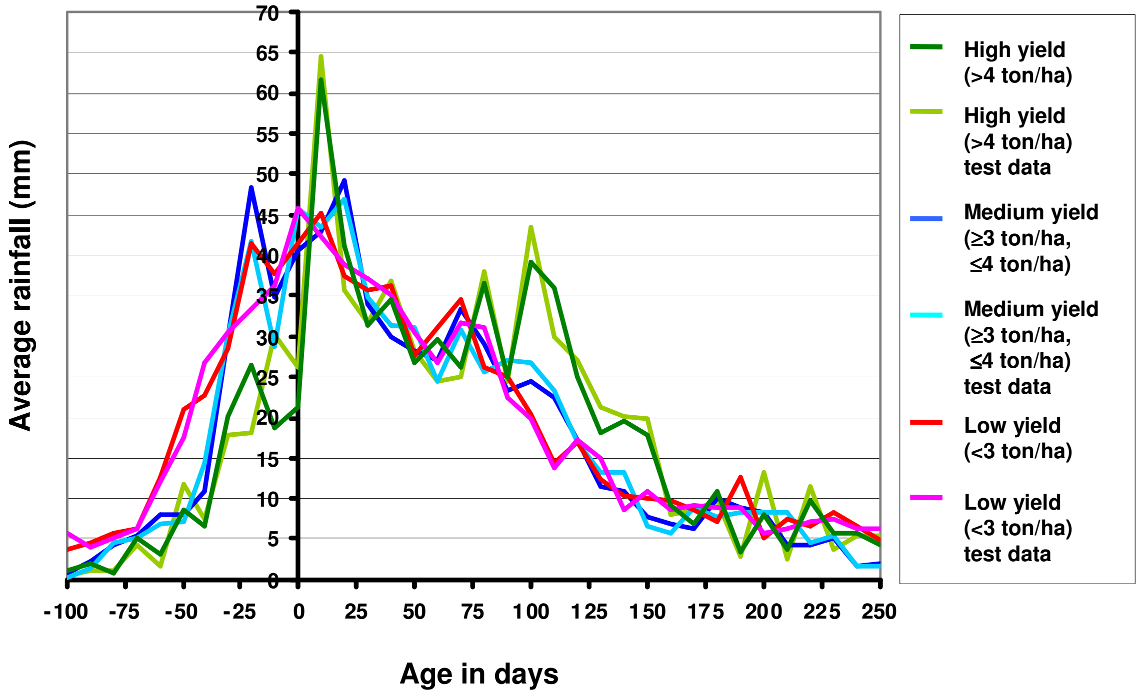


Figure 37 2001/2002 growing season low-, medium- and high yield: Average rainfall data and test data curves

Because of the difficulty experienced in discriminating between high-, medium- and low yield average and test rainfall data curves on the same graph (Figure 37), these curves have been illustrated on separate graphs. Figure 38, Figure 39 and Figure 40 illustrate how closely the test data curves follow the analysis data curves.

From Figure 37, however, three (3) distinct periods of high rainfall depicted by the high yield test and analysis data curves are identifiable. These periods were during dekad 2, dekad 8 and dekad 10 to dekad 15.

4.4.7 Crucial periods in which differences in average rainfall determine the final maize yield under dry land conditions

It is evident from Figure 35, Figure 38, Figure 39 and Figure 40, that the high -, medium - and low yield average rainfall curves displayed distinctive characteristics at the start and during dekads 7 to 14 of the growing season.

Figure 38, Figure 39 and Figure 40 prove that with a sample size that was half as large as that of the original sample size, the same peaks and troughs were visible on the graphs on the analysis- and test data sets. By displaying the analysis and test data graphs of the three classes separately it is easier to visually distinguish and recognise the differences in rainfall during these crucial periods that made a difference to the final yield.

The low yield average rainfall curve (Figure 38) peaked at day 10 after planting, while the test data curve peaked at day 0 (planting day) at ~45 mm of rainfall.

The medium yield average rainfall curves peaked at around day 20 after planting at ~48 mm of average rainfall (Figure 39). The high yield average rainfall analysis and test data curves peaked at day 10 after planting at ~64 mm of rainfall (Figure 40).

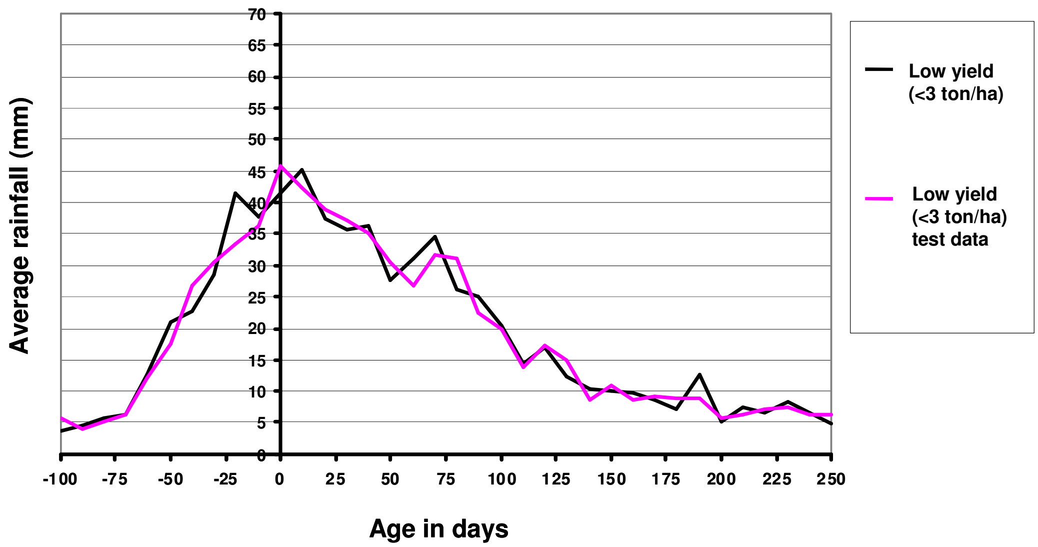


Figure 38 2001/2002 growing season low yield: Average rainfall data and test data curves

The low yield average rainfall test data curve (Figure 38) starts to decline from day 0 (planting day) to the end of the growing season (day 180 after planting). The low yield average rainfall analysis data curve (Figure 38) started to decline from day 10 after planting to the end of the growing season (day 180 after planting). Both the low yield average rainfall analysis and test data curves have experienced four (4) small peaks in average rainfall during the growing season around days 40, 70, 90 and 120 after planting.

The average rainfall at the beginning of the growing period received by the low yield objective yield analysis and test data points was ~45 mm. By the end of the growing season (day 180 after planting) the average rainfall received by the low yield analysis and test data objective yield points was ~8 mm of rainfall (Figure 38).

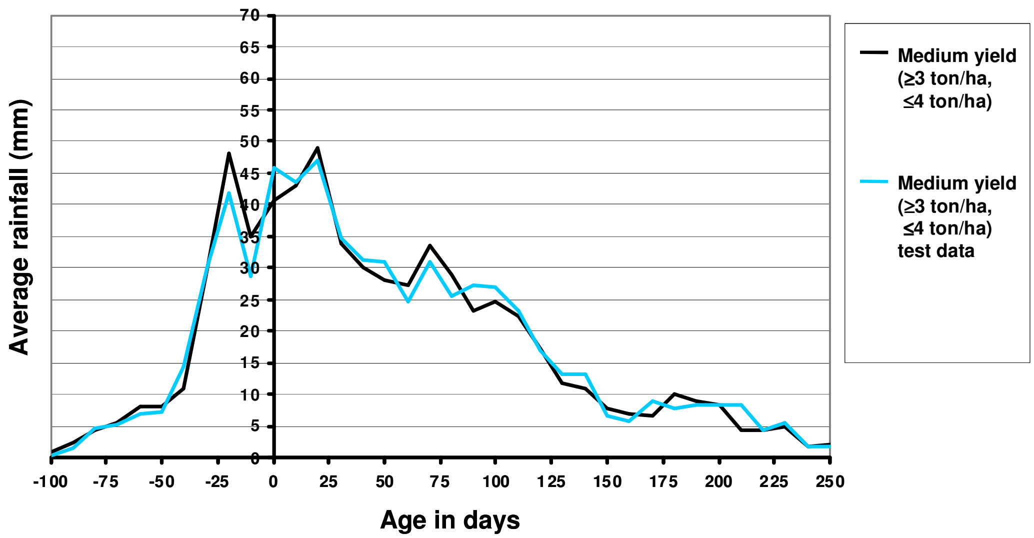


Figure 39 2001/2002 growing season medium yield: Average rainfall data and test data curves

The medium yield average rainfall analysis- and test data curves started to decline around 20 days after planting to around 160 days after planting (Figure 39). Between day 160 and day 180 after planting, both the medium yield average rainfall analysis - and test data curves experienced a peak in average rainfall (of ~10 mm of rainfall).

At the start of the growing season, from dekad 0 to dekad 3, the medium yield average rainfall analysis and test data objective yield points received ~48 mm of rainfall. By the end of the growing season, (dekad 18 after planting), the average rainfall received by the medium yield analysis and test data object yield points was down to ~8 mm of rainfall (Figure 39).

The medium yield rainfall analysis - and test data curves (Figure 39) have experienced five (5) peaks in average rainfall between day 0 (planting day) and day 180 after planting. These peaks occurred at days 20, 50, 100 and day 170 after planting.

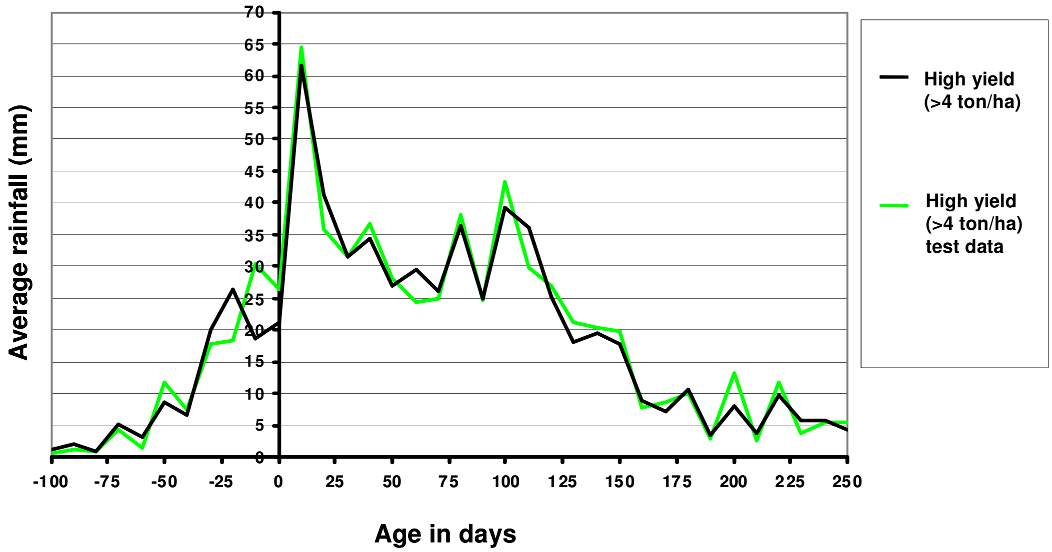


Figure 40 2001/2002 growing season high yield: Average rainfall data and test data curves

The high yield average rainfall test- and analysis data curves (Figure 40) started to decline in the beginning of the growing season at around day 10 after planting and steadily decreased to the end of the growing season (180 days after planting) with the exception of six (6) peaks in average rainfall experienced during this period. These six (6) peaks in average rainfall occurred around dekads 4, 6, 8, 10, 15 and 18 after planting (Figure 40).

The high yield average rainfall analysis- and test data objective yield points received ~64 mm of rainfall in the beginning of the growing season (day 10) after which both curves started to decline. By the end of the growing season (day 180 after planting) the average rainfall experienced by the high yield analysis and tests data objective yield points was down to ~10 mm of rainfall (Figure 40).

The average rainfall that occurs during dekads 5 to 11 has a significant impact on yield as pollination and grain fill take place during this period. Dekad 5 is a critical stage in the development of the maize plant as the potential number of kernels and the ear size is determined (Section 4.2).

Figure 38 shows that the average rainfall received by the low yield analysis and test data objective yield points had significantly decreased from ~30 mm of rainfall (at day 50 after planting) to ~14 mm of rainfall (at day 110 after planting).

Figure 39 shows that the average rainfall received by the medium yield analysis and test data objective yield points had only slightly decreased from ~30 mm of rainfall (at day 50 after planting) to ~25 mm of rainfall (at day 110 after planting).

Figure 40 shows that the average rainfall received by the high yield analysis and test data objective yield points had actually increased from ~27 mm of rainfall (at day 50 after planting) to ~39 mm of rainfall (at day 110 after planting). Although the average rainfall experienced during dekads 5 to 11 after planting was more variable for the high yield classes than the average rainfall received by the low- and medium yield classes, the number of peaks in average rainfall during this period experienced by each class should be noted.

The high yield class experienced three (3) peaks, the medium yield class experienced two (2) peaks and the low yield class experienced only one (1) peak in average rainfall during this crucial period (dekad 5 to dekad 11 after planting).

The important increase in average rainfall around days 40, 60, 80 and 100 after planting (four (4) peaks), that was received by the high yield objective yield points (Figure 40), made all the difference in the final yield (low, medium or high) that was harvested at the end of the growing season under dry-land conditions.

4.4.8 Using average rainfall data for yield prediction under irrigation conditions

Predicting yield from rainfall under irrigated conditions does not take into account the amount and quality of the irrigation received by the maize and is, therefore, not suited for this type of analysis. This reiterates the importance of analysing dry-land and irrigated fields separately as was not the case with the 2003/2004 growing season objective maize yield recordings.

5 CONCLUSIONS

NDVI and rainfall prediction “window periods” were identified (Figures 23, 24, 25, 27, 35, 36 and Figure 41). Using average NDVI graphs, the first window period (Figure 23) can be used to distinguish between high-, medium- or low yield from dekad 6 to dekad 19 after planting. The second window period (from dekad 20 to dekad 23) can be used to predict maize yield distinguishing between high- or medium-/low yields (Figure 23).

Using average rainfall graphs (Figure 35 and Figure 41), dekad 7 to dekad 8 and dekad 9 to dekad 12, can be used to determine whether maize yield predictions will fall into the low-, medium- or high yield category. Average rainfall curves can be used to predict low-/medium- or high final yields between day 130 and day 160 after planting.

These average graphs of rainfall and NDVI can be used to predict maize yields within a specific yield class (low-, medium- or high yield) and in ton/ha (quantitative yield prediction).

From section 4.3.1 to section 4.3.3 and section 4.4.1 to section 4.4.3 it is evident from the standard deviations in the data (average +/- one (1) standard deviation) and the outlier values present, that the average NDVI graphs must be used with caution when making final maize yield predictions in the window period dekads depicted in Figure 41. In any event, these prediction windows should always be used in conjunction with the standard deviation graphs for maize yield prediction.

Planting date maps, created from objective yield data points, will differ for each growing season (Frost and Kneen, 2006) and should be recreated annually (Section 4.1.1). As climate change predictions are becoming of greater concern to the maize producing communities (and around the world), maize production areas and time of planting will gradually shift. Producing these objective yield planting date maps and building up an archive could prove valuable to future maize yield estimate efforts.

Analysis of dekadal NDVI and dekadal rainfall data, done according to the planting date of the maize, provides an opportunity to identify vigour and stress factors during specific growth stages.

Predicted maize yield maps can be created from NDVI and rainfall values for decision makers from as early as 70 days after planting.

Prediction windows (Figure 41) identified periods in which the amount of rainfall received proved more significant to yield prediction than the total rainfall received throughout the growing period.

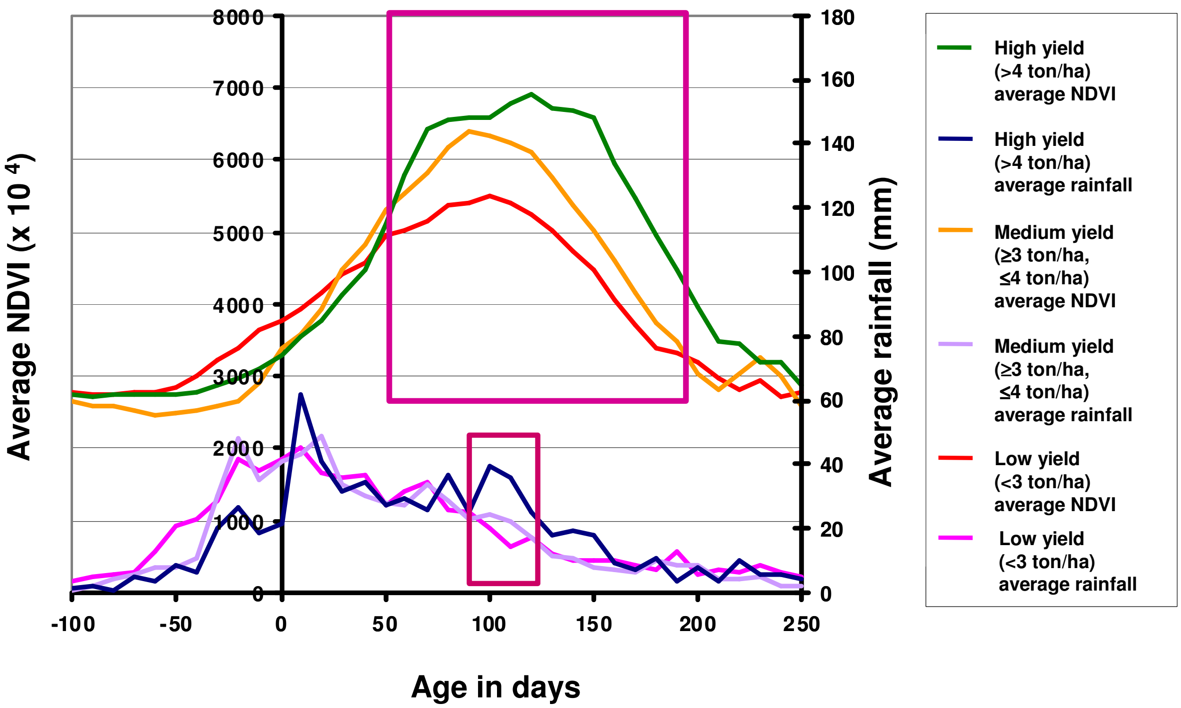


Figure 41 2001/2002 growing season prediction windows: Average NDVI curves and average rainfall curves for prediction purposes

The high-, medium- and low yield class boundaries are arbitrarily selected, but were optimal in that this selection of class boundaries produced the smallest amount of standard deviation overlap between the classes. Using, low-, medium-, high- and very high maize yield classes did not result in a decrease in overlap (standard deviation between the classes) and did not deliver better yield results.

From the qualitative and quantitative methods discussed in sections 4.2 to 4.4, it is evident that the size and scope of the rainfall prediction windows are very limited for maize yield prediction.

Although it seems that rainfall can only be used over a period of 40 days to predict the final yield very late in the season (from dekad 9 to dekad 13) the average rainfall graphs hold many other clues to the final yield that can be expected. Past dekad 13, rainfall cannot influence the yield (section 4.1.2).

The angles and positions of the average NDVI and average rainfall curves during the growing season but also prior to and after day 0 (planting date) should be investigated. These factors might contain information about the rate of increase of the NDVI curves or the rate at which the rainfall curves are declining, which could be linked to higher or lower yields. Measuring the length of the curve between the 0 axis up to the maximum NDVI value (greatest peak in the NDVI curve) could contain information about the rate at which the plants grew and the time period in which the plants had the opportunity to grow.

The individual amounts of peaks on the average rainfall curves, between day 70 and day 120 after planting, might also be useful for predicting the yield class. For example, the high yield average rainfall curve (figure 40) had two (2) peaks in that period, the medium yield average rainfall curve (Figure 39) experienced one (1) peak in that period while the low yield average rainfall curve (Figure 38) just declined steadily. The rate at which the average rainfall graphs descent from planting day as well as monthly average rainfall curves also provide clues to the expected yield that need to be investigated.

Irrigated fields were analysed using different class boundaries for low- (2 ton/ha to 5 ton/ha), medium- (9 ton/ha to 10.4 ton/ha) and high yield (11 ton/ha to 14 ton/ha) and can thus not be compared directly with estimates made for dry-land fields. The first prediction window period that can be used to predict maize yield over irrigated fields using average NDVI values, is from day 50 to day 130 after planting (Figure 27). The second window period (from day 130 to day 160) can be used to distinguish whether the expected yield will be either high/medium or low respectively.

NDVI is a measure of plant vigour and growth stage, but is not always an indication of the amount of pollination, seed development or grain fill. Outliers in the data (high NDVI value, but low final yield, or *vice versa*) could be due to intercropping, the presence of weeds or inaccurate final yield figures received from farmers (Du Toit, 2005: Personal communication).

Insufficient pollination could be responsible for low yields. Too high an air-moisture content causes pollen to burst during conditions of persistent rains (for longer than a week) over the pollination period. Below-optimal air moisture and above-optimal temperatures can scorch the plume, leading to inadequate pollination. Temperature, air moisture and rainfall values should therefore be monitored around day 60 to day 90 after planting.

Throughout the growing season, monitoring and yield prediction should be done with knowledge about the particular biological growth stage that is represented by every dekad.

Many variables that cannot be controlled (Galpin, 2004) contribute to the objective yield value during, as well as at the end of, the growing season. These variables include tillage practices; fertilisation quantity (and quality); soil type; soil moisture availability; quality of the water; cultivar differences; pest control practices; harvest loss (due to harvest- and transport methods).

Exogenous factors affecting remote observations include stage of growth (Hatfield *et al.* 1984), illumination and viewing angles, row orientation, topography, meteorological phenomena and other factors not directly related to agronomic or biophysical plant properties (Pinter *et al.*, 2003).

Compared with plants, the spectral signatures of most agricultural soils are relatively simple. They usually exhibit monotonic increases in reflectance throughout visible and NIR regions (Pinter *et al.*, 2003). High soil water content and high organic matter content generally cause lower reflectance, while dry, smooth surfaced soils tend to be brighter (Daughtry *et al.*, 1983).

As one moves from southwest to northeast in South Africa, there seems to be a trend of encountering more red soils in the west and darker soils to the northeast (Ferreira, 2005: Personal communication). This could not be confirmed with soil scientists but it is agreed that there is a gradient in the carbon content of soils from west to east. The carbon content plays a role in soil colour and fertility (Schoeman, 2005: Personal communication).

Climate in South Africa can be described as drier in the west and wetter in the east. These factors influence the infra-red and NIR bands used to create NDVIs and should be investigated for possible homogeneity for yield analysis purposes.

Under dry-land conditions, yields are on average higher in the north-eastern parts of the country than in the southwest. To be economically viable, farmers in the northeast require higher yields than farmers in the south-western parts, because input costs in the northeast are higher.

When plotting 2001/2002 objective yield 1 against final harvest, an $R^2 = 0.56$ is achieved when fitting a linear curve to the data. When objective yield 2 is plotted against final harvest, an $R^2 = 0.02$ is achieved when fitting a linear curve to the data. This low value was due to two (2) outliers (having very high values which could be faulty) in the objective yield 2 dataset.

For comparison, the first, second and third objective yield predictions of the 2004/2005 growing season were plotted against the final objective harvest as received from farmers. Results for dry-land fields delivered an $R^2 = 0.35$ for objective yield 1, an $R^2 = 0.4$ for objective yield 2 and an $R^2 = 0.49$ for objective yield 3.

When using NDVI values from test data for the 2001/2002 growing season, a 69 % accuracy in predicting the final yield within one (1) standard deviation was achieved. Using the spatial rainfall data to predict final yield was 79.4 % accurate within one (1) standard deviation.

When using 2001/2002 NDVI values alone, only 12.64 % of the time could one predict correctly within a 95 % confidence interval. Using only the spatial rainfall values, it was possible to predict correctly within a 95 % confidence interval 11.31 % of the time.

Using a statistical model (Frost and Kneen, 2006), a 39 % and 43 % accuracy were achieved using NDVI values to predict final maize yield at day 100 and day 130 respectively. Data used in the statistical model to predict yield directly from NDVI values included only yields up to 6 ton/ha (Frost and Kneen, 2006).

Predictions from NDVI and rainfall values that were correct within 1 ton/ha were achieved 14.5 % of the time, while in 52.2 % of the time, predictions were correct to within 2 ton/ha when using the statistical model (Frost and Kneen, 2006).

Maize yield predictions, using NDVI only, for the 2004/2005 growing season were 73.9 % accurate within one (1) standard deviation and using rainfall only were 79 % accurate within one (1) standard deviation.

When using 2004/2005 NDVI values alone, only 12.6 % of the time could one predict correctly within a 95 % confidence interval. Using only the spatial rainfall values, it was possible to predict correctly within a 95 % confidence interval 11.1 % of the time.

By comparing the very similar accuracies of the predictions between the 2001/2002 and 2004/2005 growing season data models from the average graphs, one expects that the results from a wet, dry and early rainfall year might yield similar accuracy levels. This raises the confidence to use and test the two existing growing season (2001/2002 and 2004/2005) models for yield prediction until data have been gathered for a wet, dry and early rainfall year (Frost and Kneen, 2006).

Analysis of the difference maps created from yield maps deducted from the final harvest objective yield (as reported by farmers) revealed that, for the 2004/2005 season, objective yields 1, 2 and 3 were very close to predictions made using only NDVI values to create maize yield maps for dekads 9 and 12 after planting.

Judging by the histograms of the difference maps, the means of these five (5) difference maps were all very close to 0 ton/ha (between 0.19 ton/ha and -1.58 ton/ha) and the standard deviations were all close to 1 ton/ha (0.76 ton/ha to 1.09 ton/ha) (Frost and Kneen, 2006).

Besides the overlap of standard deviations between the classes (low-, medium-, high yield), other reasons for incorrect predictions are MODIS pixel and field orientations, the size of the maize field and the position of the GPS point in the field. Another reason for incorrect predictions from NDVI values is that often farmers report final yields for their farm (as an average of all their fields), rather than for that specific field.

If the GPS point is not taken in the middle of a field with a size of at least 6.25 ha, the chances that the MODIS pixel is representative of a maize field become lower.

Maize fields smaller than 250 m x 250 m will result in MODIS pixels with NDVI values that are not representative of maize field values only. Vegetation surrounding small fields will influence the pixel value (Figure 42).

Because the MODIS pixel has an area of 6.25 ha, any field with an area <6.25 ha reduces the confidence in the MODIS NDVI value being representative of the field (Figure 42). Any field with an area >6.25 ha, causes the confidence in that pixel to be representative of that field, to increase.

When the GPS point is in the middle of the field, there is a good probability that the field contributes >50 % to the satellite pixel value.

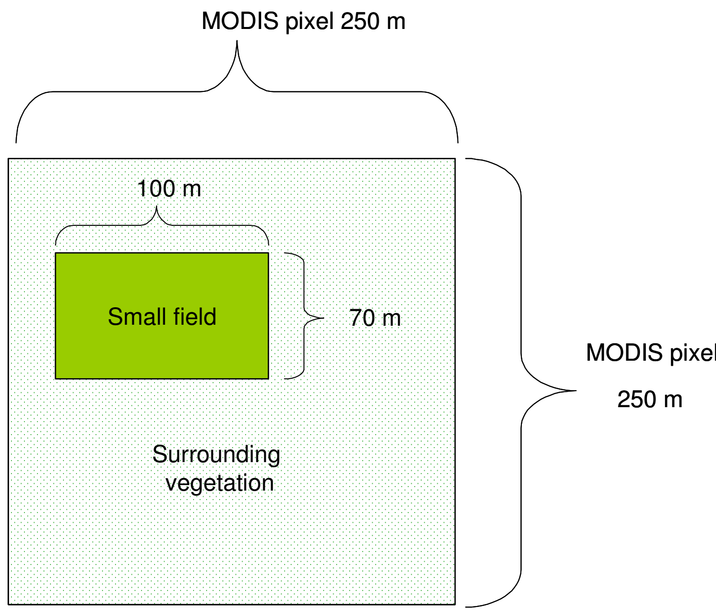


Figure 42 Schematic of the position of a small maize field in relation to a MODIS pixel

Figure 43 illustrates the probability of a MODIS pixel being representative of a maize field with an area of 6.25 ha (illustrated with four (4) similar maize fields).

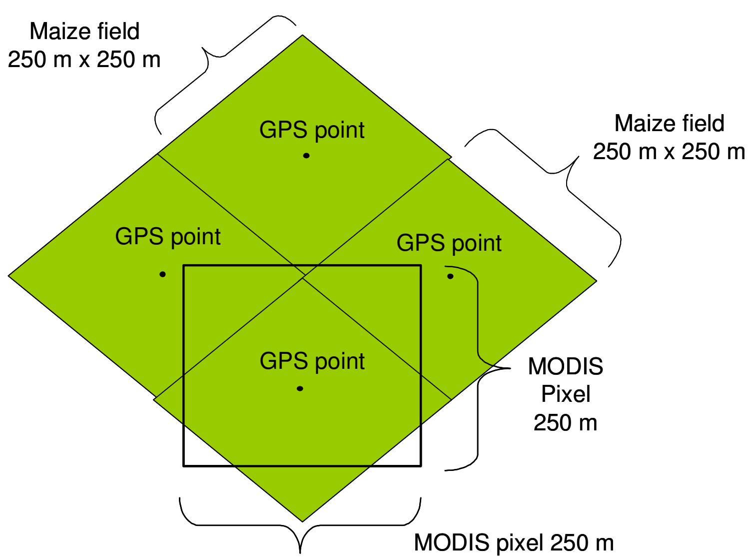


Figure 43 Schematic illustration of a 250 m x 250 m MODIS pixel and four (4) similar maize fields

Figure 44 illustrates the probability of the maize pixel being representative of a maize field. If a field has an area >25 ha, and the GPS point was taken in the middle, it would be nearly impossible for the pixel not to reflect the field accurately.

Similarly, if the field has an area of >500 m x 500 m (250 000 m²) 25 ha, and the GPS point was taken in the middle of the field, the chances increase that the pixel is representative of a maize field. As at January 2006, this concept has not been incorporated into fieldwork efforts.

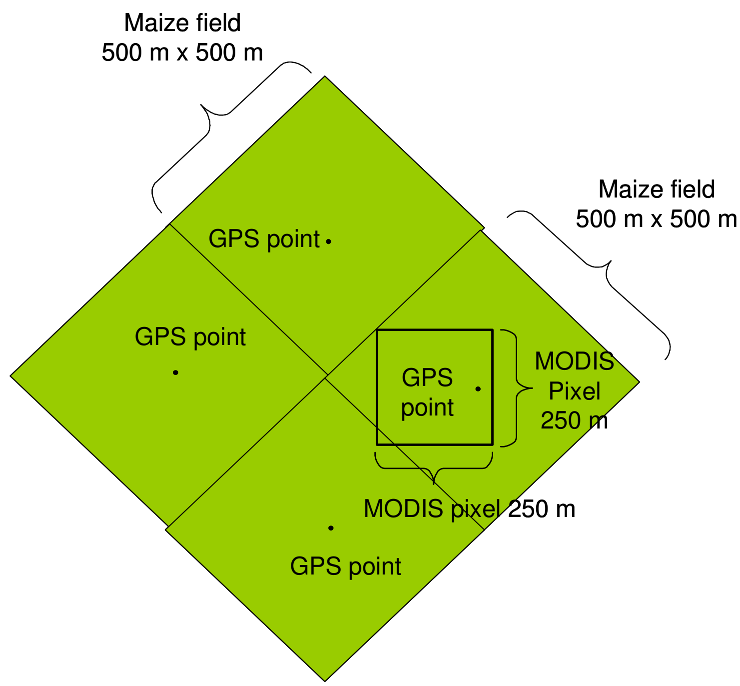


Figure 44 Schematic illustration of a 250 m x 250 m MODIS pixel and possible 500 m x 500 m maize field orientation

Maize yield estimations done by using NDVI and rainfall data in statistical models can supplement the objective yield predictions in areas where objective yield 1, 2 or 3 data are missing.

Objective yield observations produce valuable ground-truthing data for remote-sensing systems and can serve as a calibration system.

Using MODIS NDVI and spatial rainfall data in a GIS system is an effective way of verifying, contributing to and supplementing objective yield estimates.

6 RECOMMENDATIONS

Analysis according to planting date zone enables decision makers to monitor specific critical factors in specific areas during specific growth stages of the maize crop. Utilising spatial information, factors such as below- and above optimal air and soil moisture values and below- and above optimal temperature, occurrences of hail and frost, floods and continuous rains (du Plessis, 2004) can be identified as potential influences on expected yield. Utilising planting date maps and spatial information in this fashion could improve planning in terms of food security on a national and provincial level.

Analysing spatial rainfall and NDVI values at growth Stage 7 (dekad 11) (using rainfall) and growth Stage 6, Stage 7 and Stage 8 (dekad 7 to dekad 12) of the maize plant (using NDVI), proved most indicative of final maize yield and warrants further research.

The ARC-ISCW is able to produce near real-time MODIS NDVI- and spatial rainfall grids. Monitoring and yield prediction utilising these datasets together with planting date maps (conveying the age of the maize crop in each specific area) can thus readily be implemented to produce timely information for maize yield decision support.

Now that NDVI is analysed according to age, NDVI difference values (the difference in NDVI value at a specific point from one dekad to the next) plotted against age, at each specific objective yield GPS point, should be investigated for future yield analysis and to derive planting dates, growth stage and reaction to stress factors.

Analysing the rainfall dekadal grids along with the NDVI and NDVI difference grids, will provide insight into maize plant condition at each dekadal stage, especially in dry-land situations. However, this lies beyond the scope of the present research.

It is recommended that for a complete maize yield estimation system, data for a wet and dry year as well as an early rain year needs to be collected. The 2001/2002 growing season was representative of a normal year, while the 2004/2005 growing season was representative of a year with late rains but record yields.

Objective yield data collection is a crucial part of a successful maize yield estimation system and should be continued.

Ideally, GPS points should be taken as near as possible to the middle of fields and fields should have a size of at least 100 000 m² or 10 ha.

It is proposed that objective yield point information, for example, row width, planting date, dry-land or irrigated, combined with MODIS NDVI and spatial rainfall values at specific stages in the growing cycle under specific conditions (normal, wet, dry year) should be used in multivariate- and neural network analysis, as the next step in designing more sophisticated maize yield estimation systems.

The use of MODIS NDVI at objective yield points in irrigated fields produced graphs that could be used to predict expected yield over irrigated fields. It was suggested (Huete *et al.*, 2005) that for the analysis of irrigated data, MODIS Enhanced Vegetation Index (EVI) be used. This is deemed more viable for use in yield prediction and monitoring, as this ratio is less sensitive to the presence of water in irrigated fields (Huete *et al.*, 2005).

Ideally, a complete yield estimation system should include not only NDVI and rainfall, but also factors such as clay percentage and soil depth (both contained in Table 2), and temperature, cultivar and row width.

Table 2 Yield potential (shortened by du Plessis, 2004) from the Ceres maize model

Rainfall (mm)	% Clay	Soil depth 60 cm	Soil depth 90 cm	Soil depth 120 cm	Soil depth 150 cm
400	5	1.1	1.6	1.9	2.2
400	15	1.7	2.1	2.5	2.8
400	20	1.9	2.2	2.7	2.9
600	5	1.9	2.7	3.2	3.6
600	15	2.9	3.5	3.9	4.3
600	20	3.0	3.6	4.3	4.4
Source: ARC-GCI Interviewer's manual, 2004:18.					

Future objective yield analysis exercises should be done separately for dry-land and irrigated conditions.

The following influential factors as well as products from remote sensors are recommended for consideration during further research in the development of a more complete maize yield estimation equation:

- cultivated land-cover classes (Thompson, 1995)
- heat units
- agro-climatological zones
- evapo-transpiration
- rainfall received during the month prior to planting
- WRSI (Water Requirement Satisfaction Index)
- land slope
- soil type
- soil moisture availability
- land type database (to identify the soil colour, slope and % clay)
- air and ground temperature
- ecotope, included in the maize potential grids by Beukes (2004b)
- other MODIS products such as LAI, FPAR, LST and NPP should be investigated.

The resulting equation should be applied separately over homogenous planting date zones, i.e. to plants that are of the same age.

The final objective yield estimates should be tested against the final yield estimates as derived by the DoA after every growth season, as well as against SAGIS data, for benchmarking purposes.

It is suspected that some farmers intentionally do not reveal their true final yield for fear of jeopardising their market advantage. Standardisation in reporting final yields is essential. Not all farmers use scientific, accurate estimation techniques for reporting final yields. Precision farming equipment could be helpful to develop such a standardised methodology of final yield reporting. Also building relationships that involve mutual trust between the field worker and the farmer could improve the accuracy of the objective yield estimate efforts.

Field workers require improved training to produce more trustworthy objective yield data. Accurate recordings of in-field GPS point locations, maize cultivar, dry-land or irrigated conditions, planting date, field corner GPS positions and row width are vital to ensuring reliable data. A tendency amongst field workers to sample fields closer to roads (causing statistical bias), farmers denying access and incorrectly recorded GPS co-ordinates are some of the reasons for objective yield point information being inaccurate and not suitable for analysis purposes.

As at February 2006, analysis and quantification of the 2001/2002 and 2004/2005 growing season maize objective yield data have been completed (Frost and Kneen, 2006). The model accuracy should be quantifiable after objective yield analysis of this kind has been performed on data representative of a wet, dry and early rainfall year.

To complete the crop yield estimation model, a user-friendly interface is required. This interface should produce information that can be used in an interactive real-time Internet-based system like AGIS (www.agis.agric.za). Such an interface will enable the decision maker to enter a rainfall- or NDVI value at a particular point in space and time, and the system should produce a yield estimate and have it interpolated into a yield map. The interface should also contain a clause about the accuracy of the predictions, which is subject to the factors listed in Chapters 5 and 6.

REFERENCES

Agricultural Research Council - Grain Crops Institute (ARC-GCI). 2002. *2002 Crop Objective Yield Survey, Interviewer's Manual National Crop Estimates Project*. February - May. Potchefstroom, South Africa.

Agricultural Research Council - Grain Crops Institute (ARC-GCI). 2004. *Interviewer's Manual National Crop Estimates Project*. February - May. Potchefstroom, South Africa.

Agricultural Research Council - Institute for Soil, Climate and Water (ARC-ISCW). 2004. *2002, 2003, 2004 Rainfall data: Interpolated by J. Malherbe.* Pretoria, South Africa. Sourced from: www.agis.agric.za.umlindi/umlindi_report
website accessed: November 2004.

Beukes, H. 2004b. *Maize Potential Grids*. ARC-Infruitech Nietvoorbij. Stellenbosch, South Africa.

Beukes, R. 2004a. *Personal communication: Discussions on improving reliability of CEC yield statistics*. Department of Agriculture. Pretoria, South Africa.

Chappelle, E.W., Kim, M.S. and McMurtrey III, J.E. 1992. Ratio analysis of reflectance spectra (RARS) – An algorithm for the remote estimation of the concentrations of chlorophyll-a, chlorophyll-b, and carotenoids in soybean leaves. *Remote Sensing of the Environment,* 39(3) 239-247.

Childs, C. 2004. Interpolating Surfaces in ArcGIS Spatial Analyst. *ArcUser Newsletter,* July-Sept, 32-35.

Csornai, G., Wirnhardt, Cs., Suba, Zs., Somogyi, P., Nádor, G., Martinovich, L., Tikász, L., Kocsis, A., Zelei, Gy. and Lelkes, M. 1999. *Crop Monitoring by Remote Sensing*. Paper presented at the FIG Commission 3rd Annual Meeting and Seminar. Budapest, Hungary. 21-23 October.

Dabrowska-Zielinska, K., Gruszczynska, M., Stankiewicz, K. and Janowska, M. 1998. Application of visible, infrared and microwave data for crop yield assessment in Poland. *Proceedings of the International Symposium on Applied Agro-meteorology and Agro-climatology*. Volos, Greece. 24-26 April. 137-144.

Dabrowska-Zielinska, K., Kogan, F., Ciolkosz, A., Gruszczynska, M. and Kowalik, W. 2002. Modelling of crop growth conditions and crop yield in Poland using AVHRR-based indices. *International Journal of Remote Sensing,* 23(6) 1109-1123.

Daughtry, C.S.T., Gallo, K.P. and Bauer, M.E. 1983. Spectral estimation of solar radiation intercepted by corn canopies. *Agronomy Journal,* 75, 527-531.

Department of Agriculture (DoA). 2002. *Maize Crop Estimates per District*, *2002*. Pretoria, South Africa.

Department of Agriculture (DoA). 2004. *Maize Crop Estimates per District, 2004*. Pretoria, South Africa.

Doraiswamy, P.C., Moulin, S., Cook, P.W. and Stern, A. 2003. Crop yield assessment from remote sensing. *Photogrammetric Engineering and Remote Sensing*, 69(6) 665-674.

Drury, S.A. 2001. *Image Interpretation in Geology*. 2nd Edition. London: Blackwell Science Inc.

du Plessis, J.G. 2004. *Stages in the Development of the Maize Plant*. Slideshow. Agricultural Research Council - Grain Crops Institute. Potchefstroom, South Africa.

du Toit, D. 2002. *Objective Maize Yield Data: 2001/2002*. Agricultural Research Council - Grain Crops Institute. Potchefstroom, South Africa.

du Toit, D. 2005. *Objective Maize Yield Data: 2004/2005*. Agricultural Research Council - Grain Crops Institute. Potchefstroom, South Africa.

du Toit, D. 2005. *Personal communication*: *Discussions on factors influencing the maize yield*. Agricultural Research Council - Grain Crops Institute. Potchefstroom, South Africa.

Ferencz, Cs., Bognár, P., Lichtenberger, J., Hamar, D., Tarcsai, Gy., Timár, G., Molnár, G., Pásztor, Sz., Steinbach, P., Székely, B., Ferencz, O.E. and Ferencz-Árkos, I. 2004. Crop yield estimation by satellite remote sensing. *International Journal of Remote Sensing*, 25(20) 4113-4149.

Ferreira, S. 2005. *Personal communication: Discussions on research and dissertation during 2005*. ARC-Crop Estimates Committee. GeoTerraImage (Pty) Limited. Pretoria, South Africa.

Freund, J. 2006 *Personal communication: Discussions on the use of MODIS NDVI data in a direct yield prediction model.* Mr Friend is contactable at Jeremy.freund@gmail.com

Frost, C. and Kneen, M.A. 2006 The evaluation of including MODIS and spatial rainfall data into an Internet-based risk management support system. *Risk Management Report for Project 53/027/02*. ARC-ISCW. Pretoria, South Africa.

Galpin, J. 2004. *Statistical research and analysis: STAT 511*. Course notes. School of Statistics and Actuarial Sciences, University of the Witwatersrand. Johannesburg, South Africa.

Gausman, H.W., Gerbermann, A.H., Wiegand, R.W., Leamer, R.W., Rodriguez, R.R. and Noriega, J.R. 1975. Reflectance differences between crop residues and bare soils. *Soil Science Society of America Journal,* 39(4) 752-755.

Hanway, J.J. 1966. *How a Corn Plant Develops.* Iowa State University. Ames, Iowa, United States of America. Special Report, p48.

Hatfield, J.L., Pinter, P.J., Chasseray, E., Ezra, C.E., Reginato, R.J., Idso, S.B. and Jackson, R.D. 1984. Effects of particles on infrared thermometer measurements of canopy temperature in wheat. *Agricultural and Forest Meteorology,* 32(2) 97-105.

Huete, A., Kim, H., Miura, T. and Didan, K. 2005. *Scaling Dependencies and Uncertainties in Vegetation Index – Biophysical Retrievals in Heterogeneous Environments*. Paper presented at the IEEE IGARSS Conference. Korea. 25-29 July.

King, R. 2005. *Personal communication: Discussions on the future of the MODIS and NOAA.* Dr King is contactable at rking@ece.msstate.edu.

Liang, S., Fang, H., Chen, M., Shuey, C.J., walthall, C., Daughtry, C., Morisette, J., Schaaf, C., and Strahler, A. 2002. Validating MODIS land surface reflectance and albedo products: methods and preliminary results. *Remote sensing of Environment*, 83, 149-162.

Maas, S.J. 1993. Within-season calibration of modelled wheat growth using remote sensing and field sampling. *Agronomy Journal*, 85(3) 669-672.

Masamvu, K. and Siwela, B. (2004). Enhancing Food Security Early Warning Systems in the SADC region through the use of SPOT/VEGETATION and related yield indicators. 2nd VEGETATION International Users' Conference. Antwerp, Belgium. 24-26 March. SADC Regional Remote Sensing Unit, 43 Robson Manyika, Box 4046, Harare, Zimbabwe.

MODIS Characterization Support Team (MCST) http://www.mcst.ssai.biz/mcstweb/. Website accessed: 11 October 2006.

MODIS Data Support site, (http://daac.gsfc.nasa.gov/MODIS/Terra/rad_geo/MOD02QKM_disclaimer.shtml). Website accessed: 11 October 2006.

Monnik, K.A., Newby, T.S., Pretorius, A.J., du Preez, E., Ferreira, S. and Schiller, J. 2002. Development of a system to accurately estimate planted areas and forecast yields for certain summer and winter field crops in South Africa. *ARC- ISCW Report No. GW/A/2002/121*. Pretoria, South Africa.

National Crop Estimates Consortium (NCEC) Proposal, 2002. *Proposal: A crop estimates system for South Africa*. 2002. Agricultural Research Council Tender Proposal to Department of Agriculture (Section 7). Document available from ARC-ISCW.
Pretoria, South Africa.

National Department of Agriculture. 2002. *2002 Crop Objective Yield Survey Statistics*. Pretoria, South Africa.

National Department of Agriculture. 2003. *Summer Grain Crops Survey March/April 2003*. Pretoria, South Africa.

National Department of Agriculture. 2004. *Summer Grain Crops Survey February/May 2004*. Pretoria, South Africa.

Newby, T. 2005. *Personal communication: Discussion on the structure of the National Crop Estimates Consortium and Crop Estimate Committee stakeholders and data flow*. Agricultural Research Council's Institute for Soil, Climate and Water. Pretoria, South Africa.

PANNAR. 2004. *Production Regions Map, PANNAR Seed (Pty) Ltd.* Sourced from:
http://www.pannarseed.co.za
website accessed: 17 October 2004.

Pinter, P.J., Hatfield, J.L., Schepers, J.S., Barnes, E.M., Moran, M.S., Daughtry, C.S.T. and Upchurch, D.R. 2003. Remote sensing for crop management. *Photogrammetric Engineering and Remote Sensing*, 69(6) 647-664.

Schoeman, J. 2005. *Personal communication: Discussion on soil mineral content colour variation.* Agricultural Research Council's Institute for Soil, Climate and Water.
Pretoria, South Africa.

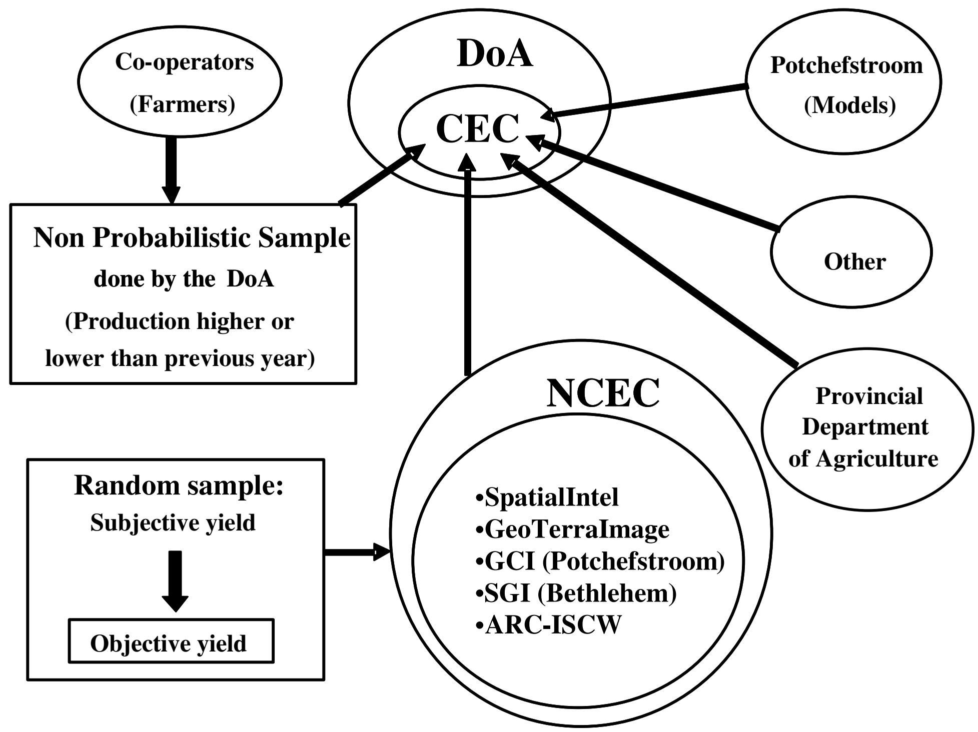
Terra MODIS
http://edcimswww.cr.usgs.gov/pub/ims welcome/
Website accessed: May, 2004.

Thompson, M.W. 1995. *Standard land-cover classification for remote sensing applications in South Africa*. Illustrated field guide sourced from: www.sac.co.za publications subdirectory website accessed: January 2004.

Umlindi Newsletter. 2004. *Umlindi Report 04051*. Sourced from: www.agis.agric.za.umlindi/umlindi_report
website accessed: June 2004.

APPENDIX 1

Schematic Showing the Structure of the CEC and the NCEC and their Sampling Systems



According to Newby, 2005.