CHAPTER 1

INTRODUCTION

1.1 AGRICULTURE IN SOUTH AFRICA

The main crops cultivated in South Africa include maize, barley, wheat, sunflower, potatoes, sugarcane, soybeans and sorghum. Figure 1.1 shows the area that was harvested in 2008 for each of these crops (BFAP, 2008). It can be seen from Figure 1.1 that maize is the primary crop grown in South Africa, with 2.8 million hectares of land harvested. The second largest cultivated crop is barley (746 000 ha) and the third largest cultivated crop is wheat (718 000 ha). As maize is the main crop cultivated in South Africa and significantly contributes to the country's economy, this study focuses on maize.

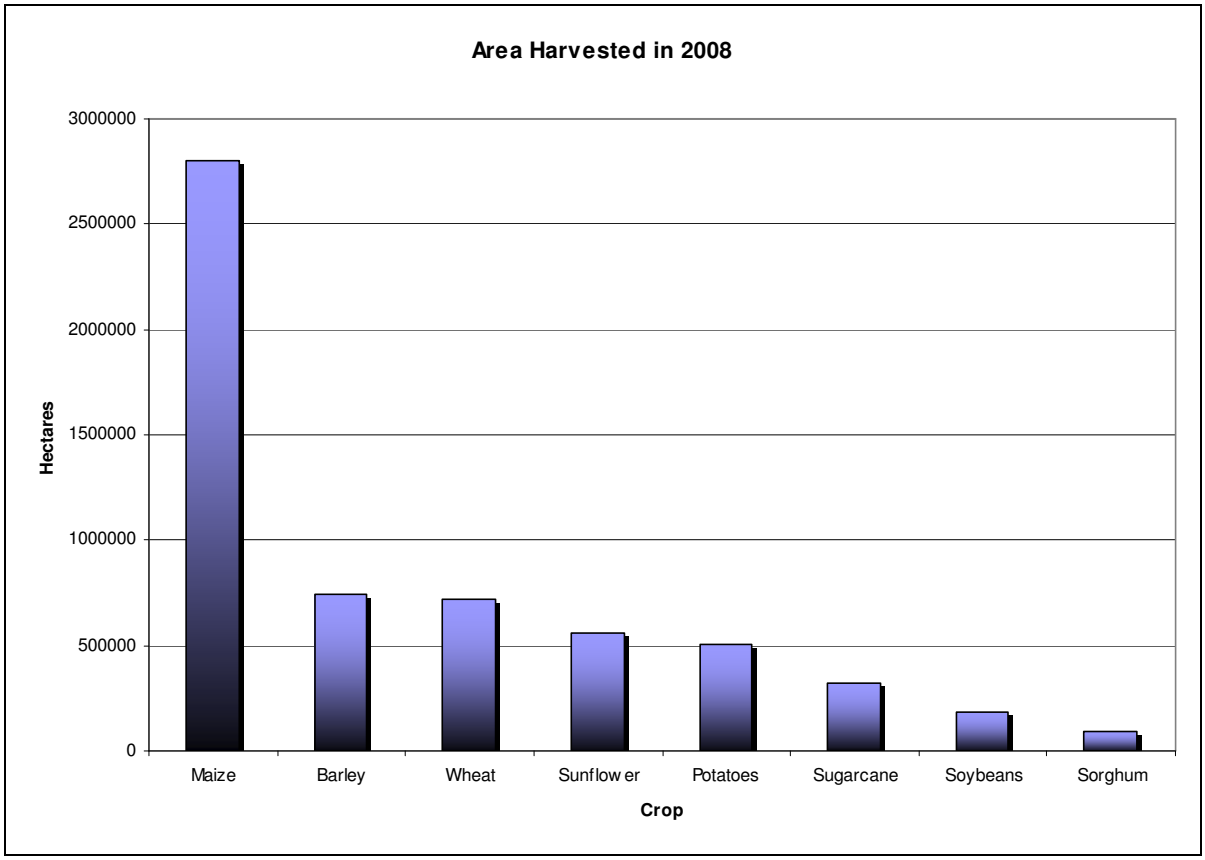

Figure 1.1: The area harvested in 2008 for each of the main crops cultivated in South Africa (BFAP, 2008).

1.2 THE GROWTH AND DEVELOPMENT OF THE MAIZE PLANT

The maize plant flourishes under warm conditions, and is only grown in areas where the average daily temperature is greater than 19℃ or where the average temperature of summer is greater than 23℃ (Du Toit, 1997). However, a minimum temperature of 10℃ is needed for germination. The critical temperature that can negatively influence the yield is approximately 32℃, and water availability normally limits maize yields. To obtain a yield of 3.2 t/ha a minimum rainfall of 350 mm is required per annum. Approximately 10 kg to 16 kg of grain are produced per hectare for every millimetre of water used. In the absence of moisture stress a total of 250 ℓ of water would have been used by each maize plant by the time it reaches maturity.

To understand the vulnerability of maize to climate variability, in particular related to rainfall, it is necessary to know the development stages of the maize plant and how each of these stages are influenced by unfavourable climatic conditions. High maize yields can only be obtained when the soil and climatic conditions are favourable during all the development stages of the plant (Sun *et al.*, 2007). The ten development stages of the maize plant, as stipulated by Hanway (1966) and Du Toit (1997), are described in the following subsections.

1.2.1 Stage 0 – From planting to seed emergence

When maize production takes place under rain-fed conditions, farmers need to wait for the first rains before planting can take place as the seed requires sufficient moisture to germinate. Under optimal warm and moist conditions the seedlings will emerge within 6 to 10 days, while unfavourable cold and dry conditions can lead to the seedlings only emerging after 2 weeks. The optimal temperatures for germination is between 20℃ and 30℃ and the plant available water in the soil should optimally be 60% of the capacity of the soil.

1.2.2 Stage 1 – Four leaves completely unfolded

The maximum number of leaves and lateral shoots are by now already determined, and approximately every 3rd day a new leave unfolds. The parts of the plant at this stage visible above the soil surface are limited to the leaf sheaths and blades. Tasseling is also initiated during this stage. The rate of development of the primary root system decreases rapidly until almost no further development takes place. If the nodes of the plant below the soil surface experience extremely dry or wet conditions, the adventitious roots will not develop

which makes the plant solely dependant on the primary root system for nutrient and water uptake. Dry climatic conditions can also lead to the roots shooting deeper. Unfavourable climatic conditions in the early development stages of the maize plant can limit the size of the leaves (Sun *et al.,* 2007).

1.2.3 Stage 2 – Eight leaves completely unfolded

During this stage the leave area of the plant increases 5 to 10 times and the mass of the stem increases 50 to 100 times. Thus, the size of the leaves and thickness of the stem are determined. The 9th, 10th and 11th leaves have reached their final size. The development of ears also commence during this stage.

1.2.4 Stage 3 – Twelve leaves completely unfolded

The leaves have now reached their final size. The stem is thickening and the lowest four leaves are dying off. The tassel in the growth point is now starting to develop rapidly. Lateral shoots bearing ears are developing from nodes 6 to 8 above the soil surface and the potential number of seed buds of the ear is genetically predetermined. During this stage plant nutrients are absorbed at a very fast rate. Prop roots are now also developing out of the first few nodes above the soil surface. At this stage the root system can extend up to 0.8 m in the horizontal and 1.2 m in the vertical.

1.2.5 Stage 4 – Sixteen leaves completely unfolded

The stem is lengthening rapidly and the tassel is almost completely developed. The tassel is pushed up higher in the plant and starts to emerge at the top. Silks begin to develop and lengthen from the base of the upper ear. Prop roots also develop out of the 7th node above the soil surface.

1.2.6 Stage 5 – Silk appearance and pollen shedding

All the leaves are completely unfolded and the tassel has been visible for 2 to 3 days. Pollen starts shedding. By now the plant has also reached its maximum height. The environment plays an important role in the determination of the height of the plant. Under unfavourable conditions, shorter plants may occur. The lateral shoot bearing the main ear have almost reached maturity. The seed buds are enlarging, while the silks are still

lengthening in the preparation for pollination. At this point leaf loss, high temperatures and too much rain can reduce the number of silks produced, cause poor pollination, and limit the size or number of kernels which will have detrimental effects on the yield (Ritchie *et al.*, 1993; Frost, 2006). The demand for water and nutrients are very high during this stage.

1.2.7 Stage 6 – Green maize stage

Pollination has now taken place. The ear and lateral shoots are now fully developed and starch begins to accumulate in the endosperm. The kernels are growing in size and rapidly increasing in mass. The kernels are filled with a milky fluid which contains a high concentration of sugar.

1.2.8 Stage 7 – Soft dough stage

The kernels are still relatively soft, and can be broken easily (Frost, 2006). The mass of the kernels are still increasing rapidly and the sugar is being converted into starch. Moisture stress can negatively influence the mass of the kernels at this stage.

1.2.9 Stage 8 – Hard dough stage

The sugar in the kernels is disappearing quickly. Starch accumulates in the crown of the kernels and extends downward. Moisture stress can also influence the mass of the kernels at this stage.

1.2.10 Stage 9 – Physiological maturity

The kernels have reached their maximum dry mass and a layer of black cells have developed at the base of the kernels. The kernels are now physiologically mature, but the moisture content still needs to be reduced. As soon as 90% to 95% of the kernels at the base of the ear appear black, the moisture content of the kernels should be in the 35% to 40% range.

1.2.11 Stage 10 – Biological maturity

Even though the kernels are physiologically mature, they have to dry out before reaching biological maturity. The drying of the kernels depends on the climatic conditions. Under

favourable conditions drying takes place at a rate of 5% per week until it reaches 20% where after the drying goes much slower.

The yield obtained at any give point is a direct product of the soil conditions and climate that prevailed during that specific season (Du Toit, 1997). Although, non-climatic factors such as crop genetic and management technique improvements also influence maize yields over time (Podesta *et al.*, 1999; Sun *et al.*, 2007). An upward trend has been found in South African maize production between 1951 and 1981, where after a downward trend followed (Du Toit *et al.*, 2001). The use of high yielding cultivars, improved fertilizer strategies, the availability of chemical weed control and improved management practices all contributed to the upward trend, while a combination of extreme weather events, rising input costs and the unstable maize price may have led to the downward trend (Du Toit *et al.*, 2001).

1.3 SOUTH AFRICAN CLIMATE

South Africa (situated at the southern tip of Africa) lies in the subtropical high pressure belt, an atmospheric zone dominated by dry descending air (Preston-Whyte and Tyson, 1993). The country has a hot and dry climate, with an average annual rainfall less than 500 mm (DoA, 2007). The distribution of rainfall is uneven, with only 35% of the country receiving more than 500 mm of rainfall per annum (DoA, 2007). The eastern parts of the country experience humid, subtropical conditions, while the western parts of the country experience dry, desert like conditions (DoA, 2007). Thus, the summer is characterized by a decrease in rainfall from east to west across the country (Preston-Whyte and Tyson, 1993; Schulze and Lynch, 2007). Three distinct regions are evident in South Africa, namely the summer rainfall region, winter rainfall region and region receiving rainfall throughout the year. The winter rainfall region is confined to the Western Cape Province and the western parts of the Northern Cape Province, the coastline of the Eastern Cape Province and parts of the Western Cape Province receives rainfall throughout the year and the remainder of the country receives summer rainfall (Schulze and Maharaj, 2007).

In terms of agriculture, a minimum annual rainfall of 500 mm is required for rain-fed cropping (DoA, 2007). The rainfall of South Africa is highly variable from year to year (Cook *et al.*, 2004) and to some extent insufficient. The country is subject to very high potential evapotranspiration that often exceeds the rainfall (DoA, 2007). The extreme irregularity of South Africa's rainfall largely influences the water resources available for agriculture (Cook *et al.*, 2004).

Maize production in South Africa is predominantly rain-fed (Martin *et al.*, 2000) and largely takes place in the central and northern interior regions of the country. This area receives summer rainfall with seasonal rainfall totals varying between 134 mm and 446 mm in the central interior parts, and between 209 mm and 584 mm in the northern interior parts (Tennant and Hewitson, 2002). Another characteristic of this region is the occurrence of mid-summer droughts, which normally takes place during January (DoA, 2007). These mid-summer droughts often coincide with the tasseling stage of the maize crop (DoA, 2007), a critical stage in its development. Moisture stress during the vital growth stages can have damaging effects on the maize plant, as a result limiting the growth and reducing the yield (Sun *et al.*, 2007).

1.4 OCEAN-ATMOSPHERE INTERACTIONS AND THE VARIABILITY IN SUMMER RAINFALL OVER SOUTH AFRICA

As rainfall is the most important factor essentially regulating maize production in South Africa, it is crucial to understand the range of factors responsible for the variability in South African summer rainfall. During the summer, temperature differences develop between the land surface of South Africa and the neighbouring oceans. This results in low surface pressure over the continent and higher surface pressure over the oceans (Preston-Whyte and Tyson, 1993). These pressure differences regularly initiate the development of subtropical troughs that are normally situated over the west coast of the country (Preston-Whyte and Tyson, 1993). Subtropical troughs, mid-latitude frontal systems and tropical circulation perturbations are the main forces responsible for moisture advection and summer rainfall over southern Africa (Hattle, 1968; Preston-Whyte and Tyson, 1993). Southern Africa can be seen as the section of Africa south of the equator. Early-summer moisture transport is mostly influenced by mid-latitude circulation patterns and mid- to late-summer moisture transport by tropical circulation patterns, causing a change in the synoptic flow during December/January (D'Abreton and Tyson, 1995). An association between wet conditions early in the summer and moisture from the tropical southeast Atlantic and south-west Indian Ocean (SWIO) converging north of South Africa have also been found (D'Abreton and Tyson, 1995). Late summer wet conditions were found to be associated with an anomalous Hadley cell, resulting in an increased flow of moisture from the north, and with the Inter-Tropical Convergence Zone (ITCZ) shifting to the south. Except for the neighbouring oceans, tropical Africa (D'Abreton and Tyson, 1995) and the Agulhas current (Jury *et al.*, 1993) were also investigated as possible sources of moisture, inducing summer rainfall over South Africa (Cook *et al.*, 2004). Furthermore, the heat and moisture fluxes in

the Agulhas retroflection region possibly influence the atmospheric pressure over the continent which may contribute to summer rainfall patterns over southern Africa (Walker, 1990; Crimp *et al.*, 1998).

Thus, many factors influence the inter-annual variability of rainfall over South Africa. A number of studies have demonstrated that some variability is remotely forced by ENSO, a phenomenon in the Equatorial Pacific Ocean (Nicholson and Entekhabi, 1986; Ropelewski and Halpert, 1987; Allan *et al.*, 1996; Mason and Jury, 1997; Reason and Rouault, 2002). ENSO is the most dominant and best defined inter-annual mode in the tropical Southern Hemisphere (Goddard *et al.*, 2001; Cane, 2004), and most previous work on the variability of the climate of southern Africa have focused on ENSO (Reason *et al.*, 2006a). The acronym ENSO originated from its oceanic component El Niño and its atmospheric component the Southern Oscillation (Cane, 2004). The correlation between ENSO and summer rainfall over southern Africa has been found to be strongest for the central continental parts (Lindesay and Vogel, 1990). However, a large amount of spatial variation occurs in ENSO rainfall impacts over southern Africa from one event to another (Reason and Jagadheesha, 2005). For example, one of the strongest El Niño events on record (1997/98) caused less intense dry conditions over southern Africa than the relatively weak 1991/92 and 2002/03 events that caused severe drought over the region (Reason and Jagadheesha, 2005). The high-phase of the atmospheric part of ENSO, the Southern Oscillation, has also been found to correlate with an increase in rainfall and the low-phase with a decrease in rainfall over southern Africa (Van Heerden *et al.*, 1988; Mason and Jury, 1997).

Even though the physical mechanisms related to ENSO are much better understood as those responsible for SST variability in the tropical Atlantic and Indian Oceans (Goddard *et al.*, 2001), several studies have confirmed that the neighbouring Indian and Atlantic Oceans also contribute to the variability in rainfall over South Africa (Walker, 1990; Jury and Pathack, 1991; Mason, 1995; Todd and Washington, 1998; Tennant and Hewitson, 2002; Reason *et al.*, 2006a). During wet events over South Africa, warm SSTs can be expected to the east and cooler SSTs to the west of the country (Tyson, 1986). Warm SSTs in the SWIO and cooler SSTs in the tropical Indian Ocean to the east of Madagascar were demonstrated to result in wet summer conditions over the continental parts of South Africa (Walker, 1990; Reason and Mulenga, 1999). Above-normal rainfall conditions over South Africa frequently relate to warm SSTs in the tropical western Indian Ocean (Landman and Mason, 1999a). SSTs south-east of South Africa also appear to be related to rainfall fluctuations over the country (Rautenbach and Smith, 2001). In addition, increasing

evidence exists that variability in the Atlantic Ocean is important for the climate of southern Africa (Reason *et al.*, 2006a). The influence of the Atlantic Ocean on the climate of southern Africa is largely related to the changing position of the ITCZ and variability in the South Atlantic anticyclone and mid-latitude westerlies (Reason *et al.*, 2006a). The link between the Benguela Niño in the Atlantic Ocean and zonal winds over the Equatorial Atlantic Ocean has also been found to be important in rainfall variability over southern Africa (Shannon *et al.*, 1986; Reason *et al.*, 2006a). Moisture fluxes, cloudband development and rainfall seasonality over some parts of southern Africa are influenced by the cycle in winds and SST anomalies over the southeast Atlantic Ocean (Reason *et al.*, 2006a). El Niño events and anomalies over the southeast Atlantic Ocean seem to be the two main factors responsible for severe drought conditions over southern Africa (Reason *et al.*, 2006a). In has been found that the upper ocean circulation and SST evolution of the South Atlantic responds to ENSO induced changes with a one-season lag (Colberg *et al.*, 2004). The impact of ENSO on the SSTs in the South Atlantic and South Indian Oceans and on the South Atlantic anticyclone have also made it evident that ENSO influences the onset of the summer rainfall season as well as dry spell frequencies within the summer season of southern Africa (Reason *et al.*, 2006a). Although it is clear that both the Indian and Atlantic Oceans potentially influences the climate of southern Africa, and some of these definitely have a relationship with ENSO, it must be kept in mind that the SSTs of the South Indian Ocean are thought to have a greater impact than the SSTs of the South Atlantic Ocean (Nicholson and Entekhabi, 1986; Reason, 2002).

1.5 SOUTH AFRICAN MAIZE YIELDS AND CLIMATE VARIABILITY

Before initiating the development of a maize yield forecast system that will aim to improve agricultural management systems by preparing the farmer for the climatic conditions of the forthcoming season, there should be a confirmed relationship between climate variability and South African maize yields. In many countries the variability in the climate accounts for as much as 80% of the year-to-year variability in crop yields (Petr, 1991; Fageria, 1992; Sivakumar, 2006). To confirm that ENSO-related climate variability largely contributes towards the variability in South African maize yields, historical maize yield figures for South Africa (GrainSA, 2007) are compared to Oceanic Nino Index (ONI) values (http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml) averaged over the growing season (OND, NDJ, DJF, JFM and FMA) (Figure 1.2).

The ONI is a principal measure for monitoring, assessing and predicting the El Niño Southern Oscillation (ENSO) phenomenon. ENSO results from an interaction between the

atmosphere and underlying ocean in the Equatorial Pacific Ocean and is seen as one of the key mechanisms responsible for climate variability in many parts of the world (Podesta *et al.*, 1999). This phenomenon involves two phases, a warm phase also known as an El Niño event and a cold phase also known as a La Niña event. The ONI is based on sea surface temperature (SST) departures from the average in the Nino 3.4 region in the Equatorial Pacific Ocean. If the SST anomaly is greater than or equal to +0.5°C for five consecutive seasons, it indicates an El Niño event, whereas a SST anomaly of less than or equal to -0.5°C for five consecutive seasons indicates a La Niña event.

ENSO is recognized as a phenomenon that significantly impacts the entire southern Africa region (Nicholson and Entekhabi, 1986; Mason and Jury, 1997), with El Niño events normally coinciding with below-normal and La Niña events normally coinciding with above-normal summer rainfall totals over the region (Ropelewski and Halpert, 1987; Mason and Jury, 1997; Reason *et al.*, 2006a). These exact ENSO impacts were observed over the central and western interior of South Africa (Rautenbach and Smith, 2001). The influence of ENSO on southern African rainfall was also found to have significant spatial and inter-event variations (Reason and Jagadheesha, 2005). Many studies have described the relationship between SST anomalies in the Equatorial Pacific Ocean and rainfall over southern Africa (van Heerden *et al.*, 1988; Allan *et al.*, 1996).

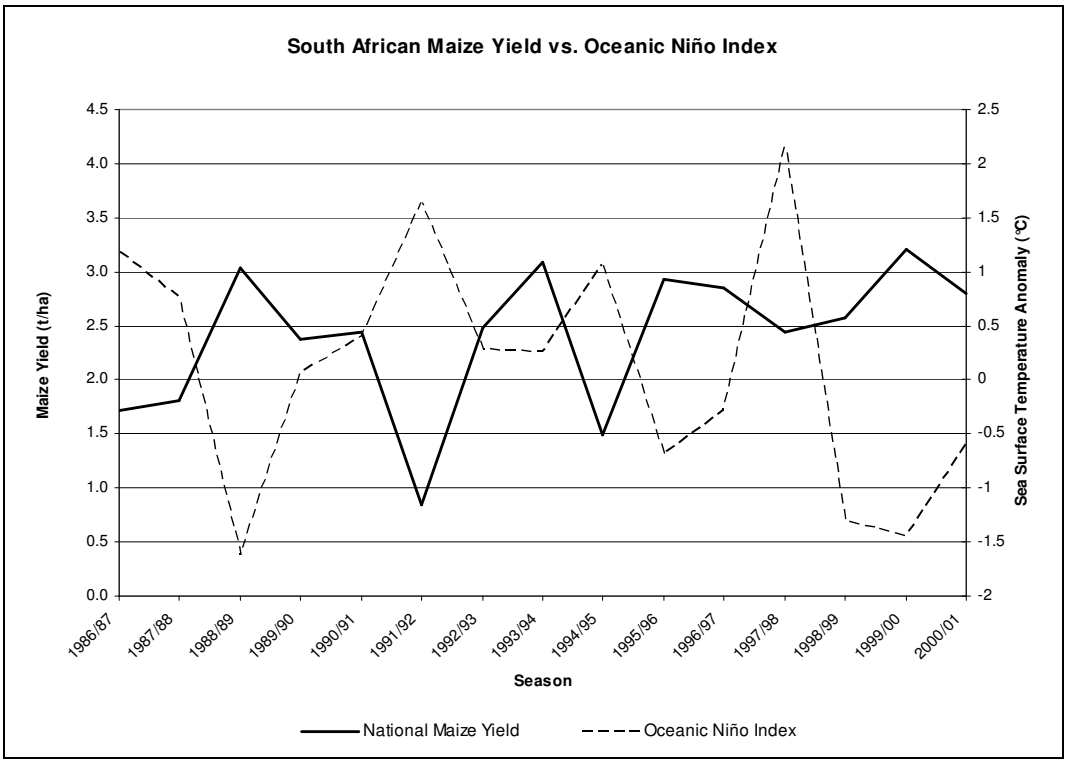

Figure 1.2: Historical maize yield figures for South Africa compared to Oceanic Nino Index values averaged over the growing season (OND, NDJ, DJF, JFM and FMA).

As ENSO directly affects the rainfall of South Africa, it should also have an effect on South African maize yields, seeing that rainfall is the sole source of water for maize production under rain-fed conditions. From Figure 1.2 it can be seen that the maize yield and ONI, are inversely proportional to each other. This means that if the ONI is positive, indicating an El Niño season and more often than not below-normal summer rainfall totals, the maize yield is low, while if the ONI is negative, indicating a La Niña season and more often than not above-normal summer rainfall totals, the maize yield is high. This is evidence that climate variability associated with ENSO contributes significantly towards the variability in South African maize yields.

1.6 THE SEASONAL PREDICTABILITY OF SOUTH AFRICAN RAINFALL

Numerical weather predictions of the exact state of the atmosphere is limited to a lead-time of about 2 weeks (1963, Lorenz). This is a result of the uncertainty associated with describing the initial state of the atmosphere, and the sensitivity of the dynamics of Global Circulation Models (GCMs) to the initial conditions. Lorenz's finding leads to the question; how is a seasonal climate forecast that provides insight into future averaged weather evolution with a lead time of, for example, 3 months then possible? Seasonal climate forecasts are possible due to an increased understanding of air-sea interactions, and in particular the interactions related to ENSO (Barnston *et al.*, 1994; Chen *et al.*, 1995; Neelin *et al.*, 1998; Palmer *et al.*, 2004; Hansen *et al.*, 2006; Vogel and O'Brien, 2006). At seasonal time scales the ocean plays a vital role (Doblas-Reyes *et al.*, 2006). The atmosphere reacts to changes in the SSTs within a few weeks, while the ocean takes 3 months or longer to react to changes in the atmosphere (Sivakumar, 2006). This slow evolution of the ocean offers the opportunity for making seasonal climate forecasts (CSIRO, 1998). Seasonal climate forecasts are far from perfect, but offer some predictability in terms of future temperatures and rainfall amounts (Ziervogel *et al.*, 2005).

Research done over the last few decades has revealed that the seasonal climate of many countries in the world is potentially predictable (Goddard *et al.*, 2001), with seasonal forecasts proving to be skilful for a number of regions particularly important for agricultural production (Challinor *et al.*, 2005). The highly variable nature of southern African rainfall, the key factor in rain-fed maize production, emphasizes the need of accurate and reliable seasonal rainfall forecasts prior to the summer season (Klopper, 1999). Skilful seasonal forecasts will allow farmers to alter management practices in light of expected weather conditions (Hollinger, 1988; Hammer, 1996; Hansen *et al.*, 2006). Even though major improvements have been made in understanding the seasonal predictability of rainfall over

southern Africa (e.g. Landman and Goddard, 2002), the region responds to a large number of factors, and this makes seasonal forecasting a challenging task (Reason *et al.*, 2006a). SST anomalies provide the main source of predictability of atmospheric developments at seasonal time-scales, since changes in the oceans result in changes in the atmosphere (Palmer and Anderson, 1994; Goddard *et al.*, 2001; Gong *et al.*, 2003; Sivakumar, 2006). Many predictability studies as well as projects involving the development of forecast models have focused on establishing relationships between global SSTs and seasonal rainfall anomalies over various regions including South Africa (Rautenbach and Smith, 2001;Tennant and Hewitson, 2002; Landman *et al.*, 2008). Thus, if a relationship is found between SST anomalies and rainfall over southern Africa, skilful seasonal rainfall forecasts will be possible if the SSTs responsible for the rainfall variability over land are predictable (Goddard *et al.*, 2001; Gong *et al.*, 2003). SST forcings account for a major portion of rainfall variability over southern Africa during the austral summer (Landman and Mason, 1999b), the most important season in maize production.

High predictability has been found for the Tropics, but most of southern Africa is situated outside of the Tropics where in general the seasonal predictability is lower (Palmer and Anderson, 1994; Landman and Goddard, 2002). Tropical atmospheric circulation patterns, which result from a direct response to SST changes, were observed to be the main source of seasonal predictability for the southern Africa region (Walker, 1990; Mason, 1995). As the peak summer rainfall period from December to February is dominated by tropical disturbances in the atmosphere (Harrison, 1984), the highest forecast skill has been obtained for this period (Barston *et al.*, 1996; Mason *et al.*, 1996). One of the most predictable tropical disturbances that impact the entire globe is the ENSO phenomenon (Allan, 2000; Goddard *et al.*, 2001). Much of the skill in predicting southern African rainfall is derived from ENSO, but definitely not dominated by it (Mason and Jury, 1997). ENSO events interact with other features in the global oceans and atmosphere which may also influence rainfall variability over southern Africa (Reason *et al.*, 2000). Therefore, the SST anomalies of the oceans surrounding southern Africa should also be taken into account when predicting southern African seasonal rainfall (Walker and Lidesay, 1989). It should also be emphasized that at seasonal time-scales there are no skill in predicting on which day a specific region will receive rainfall, but there is usable skill in predicting seasonal mean rainfall totals and intra-seasonal weather characteristics which forms part of large-scale patterns (Jury, 2002; Sun *et al.*, 2005).

1.6.1 Statistical Forecasting

Over the past few years, the South African Weather Service (SAWS) as well as a number of local universities started to issue operational seasonal rainfall forecasts for the southern Africa region (Mason *et al.*, 1996; Klopper, 1999; Landman and Mason, 1999b). Many of the seasonal forecasts were produced using statistical based techniques that include regression analysis, discriminate analysis, canonical correlation analysis, cluster analysis, time series analysis, period analysis and analogue methods (Landman and Mason, 1999b). The Climatology Research Group at the University of the Witwatersrand used a quadratic discriminant analysis model to produce seasonal rainfall outlooks for South Africa (Mason, 1998). The rainfall forecasts were produced for regions with similar inter-annual rainfall variability by relating the rainfall of each region to principal components of SST anomalies in the Indian, Atlantic and Pacific Oceans. 3-month forecasts and 6-month forecasts were produced. Skill in the 3-month forecasts for the summer rainfall region was limited to late-spring and early-summer, while the 6-month forecasts showed skill for early- to mid-summer. Even though forecast skill of the 6-month forecasts was found to be significantly higher that that of the 3-month forecasts, the 3-month forecasts revealed a high level of skill in predicting 'very dry' and 'very wet' conditions. In general, high forecast skill was obtained for the largest part of the country throughout the year, with the most reliable forecasts evident shortly before or after the start of the summer rainfall season.

Canonical correlation analysis (CCA) has also been used to investigate the variability and predictability of summer rainfall over South Africa (Landman and Mason, 1999b). CCA is a statistical method normally used to identify linear relationships between two highly correlated variables (Landman and Mason, 1999b). South African summer rainfall was the predictand and SST data of the global oceans the predictor. For each of the homogeneous regions over the country a forecast of total precipitation was produced for the October-November-December (OND) and January-February-March (JFM) seasons. The CCA model demonstrated low to moderate skill, with correlations higher than 0.5. Greater rainfall predictability was found for the JFM season than that for the OND season, which makes it extremely difficult to predict the onset of rainfall. OND predictability was restricted to the north-eastern regions and JFM predictability to the central and western regions. In general, predictions with high skill can only be expected for El Niño and La Niña years, as the Equatorial Pacific Ocean is the main source of predictability with weaker signals from the Equatorial Indian and Atlantic Oceans (Landman and Mason, 1999b).

A statistical model has also been used to investigate the climate signals around southern Africa and to predict area-average rainfall (Jury *et al.*, 2004). The investigation was based on the hypothesis that climate signals important in the prediction of southern Africa rainfall originate from slowly varying waves. Sea level pressure (SLP) and SSTs of the Atlantic and Indian Oceans as well as southern Africa rainfall were considered. The multi-variate model demonstrated useful skill in predicting southern Africa rainfall at 1-year lead time and is particularly skilful in predicting extreme events. The strongest SST and SLP signals are evident 6 to 12 months before the rainfall season which indicates potential predictability (Jury *et al.*, 2004).

1.6.2 Multi-tiered Forecasting

Even though most statistical forecast methods make use of linear relationships, many climate processes show strong non-linearities (Landman *et al.*, 2001). This ultimately limits the forecast skill of statistical models (Carson, 1998). However, these non-linearities can possibly be simulated with GCMs (Landman and Mason, 1999a). A GCM represents a simplification of the climate system through the equations of motion but suffer from initial condition and inherent model uncertainties which may lead to model output not reflecting the real system accurately (Holton, 1979; Hollinger, 1988; Doblas-Reyes *et al.*, 2006). Even so, the use of GCMs offers great opportunities for improving the seasonal predictability of summer rainfall over South Africa. The skill of a statistical model and that of a GCM has been compared over a 10-year retro-active period when predicting December-January-February (DJF) summer rainfall for southern Africa (Landman *et al.,* 2001). CCA was used as the statistical model and the GCM used was the Centre for Ocean-Land-Atmosphere Studies (COLA) T30 model with a horizontal resolution of approximately 400 km. The lower boundary conditions used to force the GCM was SSTs predicted with the CCA model. The GCM output was downscaled using the perfect prognosis approach (Wilks, 2006). This combination of statistical and dynamical forecasting techniques is known as a multi-tiered scheme. The results found suggested that the multi-tiered approach produce more skilful forecasts than that produced by the CCA statistical model. Even though skilful 1- and 3-month lead time predictions of Equatorial Pacific and Indian Ocean SSTs anomalies were obtained through CCA (Landman and Mason, 2001), improved SST forecasts can result in the multi-tiered approach increasing the skill of seasonal rainfall forecasts for South Africa. GCMs will most probably form the centre of seasonal forecasting in the years ahead, and no longer statistical models (Landman *et al.,* 2001).

1.6.3 Dynamical Forecasting

The capability of the CSIRO-9 GCM to model the major global SST forcings that contribute to the inter-annual variability in rainfall over South Africa and Namibia has been investigated (Rautenbach and Smith, 2001). The GCM was forced with observed global SST anomalies for the 30-year period from 1961 – 1990, and an ensemble of five simulations was produced, each initialized with different initial conditions. Skilful model results that correlate strongly with the observations were obtained for the dominant austral summer season (October-March). It was also demonstrated that the model simulated rainfall variability during the austral summer season compare well with SST perturbations in the Equatorial Pacific and tropical western Indian Oceans.

The ability of a GCM to simulate the impact of five ENSO events on southern Africa rainfall has been tested (Reason and Jagadheesa, 2005). The GCM used was the UKMO HadAM3 model which was forced in hindcast mode for a period from 1990 to 2003 with observed SSTs. The model was implemented at the University of Cape Town as part of a dynamical seasonal forecasting project. The investigation focused on the OND and JFM seasons. The GCM showed highest skill for the 1997/8 El Niño event, with lower skill for the 1991/2 and 2002/3 El Niño events and 1995/6 and 1999/00 La Niña events. The GCM was found to experience difficulty in capturing changes in the Angola low, a centre of tropical convection often associated with rainfall impacts over southern Africa during ENSO events. Therefore, as the GCM did not represent the Angola low properly, it struggled to simulate the ENSO rainfall impacts over southern Africa. In a study done on wet and dry spells over South Africa, Cook *et al.* (2004) also highlighted the importance of the Angola low in seasonal rainfall over South Africa.

Due to current computational capabilities, the spatial resolution of seasonal forecasts obtained from GCMs is often limited, in the range of 100 km x 100 km (Palmer *et al.*, 2004; Hansen *et al.*, 2006) or even courser. The GCMs focus on large-scale weather systems and are less skilful in representing local weather conditions, especially precipitation (Cantelaube and Terres, 2005). To obtain a higher spatial resolution over a specific area (downscaling) one can make use of a Regional Climate Model (RCM), which is normally nested within the GCM (Kgatuke *et al.*, 2008). A number of RCMs are currently used for seasonal simulations over southern Africa, with the MM5 RCM being used in Ghana, Nigeria, Zambia and Zimbabwe (Tadross *et al.*, 2006). The MM5 RCM has been used to simulate rainfall for a wet DJF season (1988/9) and a dry DJF season (1991/2) over southern Africa, and at the same time also investigated the influence of two different

planetary boundary layer and two different cumulus convection parameterization schemes on the model output (Tadross *et al.*, 2006). The simulated rainfall results were compared to observed precipitation (seasonal and diurnal), number of rain days, diurnal short-wave fluxes and optical depth. All four model configurations simulated the total precipitation for the wet DJF season well, but it was found that the model underestimates the inter-annual change. It was also demonstrated that the biases in the simulated DJF rainfall are largely related to biases in the number of rain days and the diurnal moisture and energy cycles.

From the research described above it appears as if the region and season for which models show skill vary from one model to another. It is also clear that significant progress has been made in seasonal forecasting techniques over the last decade. The majority of seasonal forecasting systems currently in use all make use of GCMs, several in multi-tiered or two-tiered approaches. Even though seasonal forecasts produced by GCMs will never be perfect, GCMs have proved to be skilful in many regions and in particular the Tropics (Hunt *et al.*, 1994; Mason *et al.*, 1999). GCMs are capable of simulating much of the large-scale atmospheric circulation, but often struggle to capture local sub-grid-scale variability (Goddard *et al.*, 2001). Even though GCMs tend to overestimate and spatially distort rainfall over southern Africa (Joubert and Hewitson, 1997), when forced with observed SSTs these models have seem to capture the main austral summer seasonal rainfall variability over the region (Rautenbach and Smith, 2001; Goddard and Mason, 2002; Reason and Jagadheesha, 2005), but skill is limited in non-ENSO years (Landman and Mason, 1999b). Even in ENSO years when the seasonal predictability is relatively high, it must be kept in mind that inter – El Niño differences exist (Hoerling and Kumar, 1997) which influences the confidence in the expected conditions predicted during these years (Mason and Goddard, 2001). The predictability of rainfall over the summer rainfall season of South Africa, from October to March, varies significantly when using SSTs as precursor (Landman and Mason, 1999b). October rainfall was found to be the least predictable when using only SSTs, while November rainfall was predictable using central-south Atlantic SSTs. The Equatorial Pacific Ocean and Arabian Sea were found to be important in predicting December rainfall, but January rainfall related poorly to SSTs. Most predictability for February and March rainfall were found to originate from the central Equatorial Indian Ocean (Pathack *et al.*, 1993).

The GCM fields that are used in this study as input into the crop model are not actual forecasts, but simulation data. Simulation data are obtained by forcing a GCM with simultaneous observed SSTs, while a real-time forecast is obtained by forcing a GCM with predicted SSTs. To recalibrate the circulation patterns generated by a GCM, a model

output statistics (MOS) method has been applied and results presented for the DJF rainfall season over southern Africa (Landman and Goddard, 2002). Two datasets were used, the first dataset was obtained by forcing the GCM with simultaneous observed SSTs for the DJF season (simulation data) and the second dataset was obtained by forcing the GCM with persisted November SSTs through the DJF season (hindcast data). The second dataset in effect has a lead-time of one month and could therefore be associated with a real-time forecast issued early December for the upcoming DJF season. Both the simulation-MOS and hindcast-MOS forecasts agreed significantly with the observations. Thus, similarly skilful seasonal rainfall forecasts could be produced using both the simulation data and hindcast data which correspond to a real-time forecast with a 1-month lead-time. Therefore, it can be assumed that the maize yield simulations produced in this study using the GCM-simulated fields as input into the crop model will yield similar results to maize yield forecasts produced using actual seasonal forecasts with a 1-month lead-time as input into the crop model.

1.7 ENSEMBLE AND MULTI-MODEL FORECASTING

Model ensembles have become an essential part of forecasting over the last few years (Krishnamurti *et al.*, 1999). Due to the atmosphere behaving in a chaotic manner, the initial state of the atmosphere is not certain (Palmer *et al.*, 2004; Sivakumar, 2006). The initial state of the atmosphere is used to initialize a GCM, which then integrates that initial conditions into the future to obtain a forecast. Thus, with an increase in lead time even the most sophisticated forecast model will diverge further and further away from reality. To address this source of uncertainty, the GCM is initialized from a number of possible atmospheric initial states (Sivakumar, 2006). In other words, several forecasts are made by each time introducing slightly different initial conditions into the GCM (Palmer *et al.*, 2004; Doblas-Reyes *et al.*, 2006). This is called an ensemble, with each forecast representing a member of the ensemble. This approach provides a range of possible outcomes (Sivakumar, 2006) and by investigating the ensemble spread the uncertainty in the forecasts associated with the initial conditions can be estimated (Barnston *et al.*, 2003; Palmer *et al.*, 2004; Hansen *et al.*, 2006). From the ensemble of forecasts a probability distribution function can be obtained (Doblas-Reyes *et al.*, 2006; Hansen *et al.*, 2006) by calculating the percentage of ensemble members that fall either within the below-normal, near-normal or above-normal categories (Reason *et al.*, 2006b). When an ensemble of seasonal forecasts is produced, the set of forecasts represent the probability distribution of climate in its response to SST forcings. When two or more skilful but independent forecasts of the same event are combined, the final forecast will be more accurate than any of the

individual forecasts by itself (Leith, 1974). Thus, ensemble forecasting improves the skill of forecasts (Barnston *et al.,* 2003; Palmer *et al.*, 2004).

A number of other factors also contribute to the uncertainty in GCM forecasts. The main factors include the way in which small-scale features are represented within the model (parameterization) (Palmer *et al.*, 2005), the way in which data is introduced into the model (assimilation), assumptions made within the model and model equation errors (Palmer *et al.*, 2005). To account for these uncertainties more than one GCM can be used, as each GCM makes use of different parameterization schemes, data assimilation procedures and assumptions and may even have different inherent model equation inaccuracies. Thus, each GCM will probably perform different due to these differences (Landman and Goddard, 2003). In effect each individual model runs its own ensemble, which can be combined to form a multi-model ensemble or super-ensemble (Barnston *et al.*, 2003; Palmer *et al.*, 2004; Sivakumar, 2006).

A number of studies have shown evidence that a multi-model forecast system provides more skilful forecasts than any individual model. In the PROVOST (Prediction of Climate Variations on Seasonal to Interannual Timescales) project (Palmer *et al.*, 2004) carried out in Europe, a number of GCMs were used to perform 4-month forecasts when forced with SSTs. Each model was initialized 9 times from slightly different atmosphere initial conditions, while the same boundary conditions (SSTs) were used to force all the GCMs. Results from the PROVOST project showed that regardless of identical SSTs, the ensembles from the individual models varied considerably in the seasonal-mean signal from the SSTs. Despite this, the multi-model ensemble system still proved to produce more reliable forecasts than any of the single-model ensembles.

In a more complex and more well know study also done in Europe, the DEMETER (Development of a European Multi-model Ensemble system for seasonal to inTERannual prediction) project (Palmer *et al.*, 2004), 7 coupled ocean-atmosphere global circulation models (CGCMs), each running its own ensemble of 9 simulations from different ocean initial conditions, were used to perform a series of six-month hindcasts. Thus, a multi-model ensemble of 7 x 9 was produced. In CGCMs the atmosphere and oceans can evolve freely and are consequently allowed to influence each other (Goddard *et al.*, 2001). The DEMETER results indicated that the multi-model forecasting technique is feasible to represent model uncertainty on seasonal and inter-annual time-scales. It was also found that on average the multi-model system provides more skilful seasonal forecasts than that

produced by a single-model system. In the USA similar attempts have been made under the Dynamic Seasonal Prediction (DPS) projects (Sivakumar, 2006).

An investigation has been done to assess whether the advantage of the multi-model system over the single-model system is only due to an increase in ensemble size (Hagedorn *et al.,* 2005). A single-model ensemble and a multi-model ensemble of the same size (54-members) were compared. Results indicated that even with the same ensemble size, the overall performance of the multi-model system is better.

Sufficient evidence has been presented that multi-model systems can improve on the skill of both weather and seasonal forecasts produced by a single-model system (Krishnamurti *et al*., 1999; Harrison, 2003; Cantelaube and Terres, 2005; Doblas-Reyes *et al*., 2006). The question arises whether this is also true for South Africa. Multi-model summer rainfall forecasts for southern Africa have shown to be more skilful than single-model forecasts produced for this region (Reason *et al*., 2006b). The skill of a multi-model system in predicting DJF rainfall for southern Africa has been investigated (Klopper and Landman, 2003). The three models used included two statistical models (CCA and quadratic discriminate analysis) and one GCM (ECHAM 3.6). Each model produced forecasts with different levels of skill, which means that a combined forecast will incorporate the strengths of each model. The forecasts were combined through simple unweighted averaging. The results showed that the multi-model forecast improved on the skill of the individual model forecasts, and that on average the combined forecast showed higher skill, at least for the majority of summer rainfall regions over southern Africa.

The skill in predicting DJF rainfall over southern Africa using a multi-model ensemble system has been evaluated (Landman and Goddard, 2003). The analysis used simulation data obtained by forcing five GCMs (CCM3.2, ECHAM4.5, NCEP-MRF9, COLA T63, NASA-NSIPP1) with simultaneous observed SST anomalies. Thus, DJF rainfall was predicted using observed DJF SST anomalies. The simulations performed by each GCM were recalibrated to homogeneous rainfall regions over southern Africa using the statistical method called MOS (Model Output Statistics). The simulations from the individual models were then combined by averaging the downscaled results. Results suggested that the combination of models improve on the performance of the best single-model ensemble (ECHAM4.5 – MOS).

1.8 OPERATIONAL SEASONAL FORECASTING AT SAWS AND UP

Routine seasonal climate forecasts using the Conformal-Cubic Atmospheric model (CCAM) has been produced at the University of Pretoria (UP) since August 2007. The model is also being applied at UP in the fields of climate simulation (Engelbrecht *et al.*, 2009) and short-range weather forecasting (Potgieter, 2007). In operational seasonal forecasting mode, CCAM is initialized using the 0Z analysis fields obtained from the Global Forecasting System (GFS). A three-month seasonal forecast (having 12 ensemble members initialized on 12 consecutive days) is issued on a monthly basis. Lower boundary forcing is prescribed from persisted SSTs, as obtained from the GFS. The model runs globally at C48 (approximately 200 km) horizontal resolution on a quasi-uniform grid. Output for a number of variables is available on a global 1° latitude-longitude grid. The forecasts are performed on the Velocity-cluster at the University of Pretoria and feeds into the multi-model seasonal forecast system of SAWS (Landman *et al.*, 2008). This dissertation reports on the use of CCAM-simulated fields as part of a maize yield forecast system for South Africa.

The latest development at the South African Weather Service (SAWS) is described in Landman *et al.* (2008). Here, the use of multi-model ensembles in operational predictability of seasonal to inter-annual rainfall over South Africa has been investigated. The models that were used include the ECHAM4.5 GCM, the CCAM GCM, the UKMO CGCM and a statistical CCA SST-rainfall model. The ECHAM4.5 model output were obtained from the International Research Institute (IRI), CCAM output data from UP and UKMO CGCM output form the European Centre for Medium-range Weather Forecasts (ECMWF). The ECHAM4.5 model has also been installed on the supercomputer of the SAWS and a 6-member ensemble of multi-decadal simulations, forcing the model with simultaneous observed SSTs, has been performed. Each model ran its own ensemble of simulations. Before combining the ensembles from the individual models, the ensemble mean was obtained for each model and MOS was applied to it for downscaling purposes. It was found that the model combinations did not always outscore the individual models, but the use of longer training periods and by combining only the best models are necessary requirements to improve on forecast skill. Inclusion of the statistical model that only uses antecedent SSTs as predictors in the multi-model systems made the results worse. It also appeared as if seasonal rainfall predictability is limited to mid-summer months which coincide with ENSO events and the highest skill during these seasons occur over the north-east and central-western regions of South Africa. Overall, useful skill was obtained from the multi-model systems for the DJF season. As a result of this project the first operational multi-model forecast, which made use of an 8-member ensemble CCAM forecast and a 24-member

ensemble ECHAM4.5 forecast for the April-May-June season, was issued for South Africa on 31 March 2008.

In the mean time, the multi-model system has been finalized, and operational rainfall and temperature forecasts are currently made routinely for the Southern African Development Community (SADC) region. SAWS has also obtained Global Producing Centre for Long-Range Forecasting status from the World Meteorological Organisation. The ECHAM4.5 GCM, which runs on the NAC SX-8 supercomputer of SAWS, is used for this purpose. ECHAM4.5 forecasts are also used in the operational multi-model forecasts mentioned above. This dissertation reports on the use of ECHAM4.5-simulated fields as part of a maize yield forecast system for South Africa.

1.9 CROP YIELD FORECASTING

Several studies have emphasized the sensitivity of agricultural production to weather (Hollinger, 1988). The vulnerability of field grown crops to fluctuations in the weather on a daily, monthly and seasonal time-scale (Doblas-Reyes *et al.*, 2006) affects the welfare of farmers due to the irregularity in crop yields from one year to another (Sivakumar, 2006). Therefore, many scientists have attempted to reduce the uncertainties associated with the growing season so that famers can make more informed decisions and take advantage of good seasons. The growth, development and yield of a crop are the function of interactions between the plant, weather, soil and management practices (Hansen *et al.*, 2006). The recent advances in the ability to predict fluctuations in the climate several months in advance have increased the opportunities of seasonal forecasts to alter management decisions and reduce the negative impacts of climate variability on crops (Hammer *et al.*, 2001; Mason, 2001; Sivakumar, 2006). However, farmers can benefit more from information when it is presented in terms of production outcomes than from a seasonal climate forecast by itself (Hansen and Indeje, 2004; Hansen *et al.*, 2006). Crop yield forecasts at an early enough lead-time can warn the farmer of a probable poor season and consequently allow the farmer to change the planting date, cultivar type as well as management and planning activities like the necessity of fertilizer and irrigation, in light of the expected conditions (Martin *et al.*, 2000).

Some of the first crop yield forecasting attempts were based on empirical relationships between variables in the environment and crop yield (Isard *et al.*, 1995). An advance in agricultural science took place in the 1970's with the development of the first crop simulation model (Fodor and Kovacs, 2003) and since then several simulation models have

been developed for a range of different crops (Whisler *et al.*, 1986; Ritchie, 1994). These models are by no means perfect, but can assist in understanding how cropping systems function (Bannayan and Court, 1999; Matthews, 2002). The main aim of the development of these crop simulation models was for application in agricultural research (Hoogenboom *et al.*, 1992), in particular the possibility of application in yield forecasting (Bannayan and Court, 1999). A crop simulation model is a mathematical representation of the complex real-world system (Fodor and Kovacs, 2003) and can simulate crop growth and estimate crop yield as a function of weather, soil and crop management conditions (Egli and Bruening, 1992; Boote *et al.*, 1996; Hoogenboom, 2000; Matthews, 2002; Palmer *et al.*, 2004). Numerous crop yield forecasting efforts have made use and are currently using dynamic crop simulation models as a tool to convert weather and climate forecasts into an estimation of the production in response to predicted future conditions (Challinor *et al.*, 2005; Hansen *et al.*, 2006).

Using a crop simulation model to predict crop yield for the forthcoming season or the season in progress requires weather input data for the entire growing season (Bannayan and Court, 1999; Lawless and Semenov, 2005). The key weather input variables are precipitation, temperature and solar radiation (Hoogenboom, 2000; Doblas-Reyes *et al.*, 2006). Thus, a forecast of these weather variables will need to be made weeks or even months prior to the specific season of interest (Doblas-Reyes *et al.*, 2006). The accuracy of the weather input data will influence the yield output produced by the crop simulation model (Nonhebel, 1994). Thus, improved weather forecasts will translate into more accurate yield forecasts (Challinor *et al.*, 2005). Many crop yield forecasting studies have used stochastic weather generators to construct synthetic weather for the growing season. These weather generators require some form of historical data for each weather variable as input in order to generate synthetic weather for a specific site (Bannayan and Court, 1999; Hoogenboom, 2000; Lawless and Semenov, 2005). Other crop yield forecasting studies have used daily GCM output for the growing season as input to a crop simulation model (Hansen *et al.*, 2006). Even though GCMs tend to distort daily variability, particularly precipitation, many of these studies were very successful by either calibrating the simulated yields produced with the raw GCM data, rescaling to GCM mean bias or by correcting the rainfall frequency and intensity of the GCM output (Hansen *et al.*, 2006). The set up of the crop model-GCM based simulation system used in this study is done in such a way to establish the skill of the system without any additional GCM output manipulation, therefore setting a baseline against which newly developed systems can be compared.

1.9.1 Yield Predictions in Europe

A method which uses the SUCROS crop simulation model in combination with the SIMMETEO stochastic weather generator has been tested for real-time winter wheat biomass and yield predictions at four sites in the UK (Bannayan and Court, 1999). Observed monthly mean values were used by SIMMETEO to generate representative weather data for each site. The simulations performed with SUCROS were updated throughout the growing season, by combining generated weather data with observed weather data. SUCROS simulated crop biomass, and yields were calculated by multiplying the simulated biomass with the measured harvest index at each site. The correlation between simulated and observed biomass and yield were found to increase as the growing season progresses, due to model updating. The forecasts showed reasonable skill which would provide an opportunity for the farmer to alter or adapt management before harvest.

In a similar study, the Sirius crop simulation model was used in combination with the LARS-WG weather generator for predicting within-season wheat yields at five sites in Europe and one site in New Zealand (Lawless and Semenov, 2007). The study aimed to assess lead-time for making skilful predictions before crop maturity. LARS-WG produced an ensemble of "artificial" weather datasets for each site. Each of the "artificial" weather ensemble members were combined with observed weather data and used to force the Sirius crop model. The combined datasets contained observed weather data for the initial part of the season and "artificial" weather data for the remainder of the season. As the season progressed, at 10-day increments, the observed data was increased and the "artificial" weather data reduced and new runs performed with the Sirius crop model. The results indicated that the uncertainty in the predictions decreases as the season evolves and that a usable level of skill is reached before crop maturity, but that the lead-time of skilful predictions varies significantly from one location to another.

Another objective of the DEMETER project was to demonstrate the value of seasonal climate forecasts by coupling the multi-model ensemble to application models (Palmer *et al.*, 2004). The WOFOST (WOrld FOod STudies) crop model from the Joint Research Centre (JRC) in Europe was used as the application model in this study. Output from the 7 CGCMs used in the DEMETER project was on a low spatial (1.5° x 1.5°) and temporal (monthly mean values) scale and had to be downscaled in both space and time before the data could be used to force the crop model. Hindcasts for each of the February-July seasons within the period from 1995 to 1998 were downscaled using singular value decomposition, MOS and a weather generator. Thereafter, the downscaled meteorological

data from each individual ensemble member was used to run the crop model for the four year period (1995 - 1998), which resulted in a wheat yield prediction ensemble and could consequently be used to derive a probability distribution function of wheat yield for Europe. The results were verified against actual wheat yield figures as well as yield simulations performed by the operational yield forecasting system of the JRC. The results were found to vary from country to country, with the highest correlation (0.73) between DEMETER-based yield predictions and actual yield figures found for the main wheat producing countries in Europe (France, Germany and the United Kingdom). It was concluded that reliable crop yield predictions is possible with the use of a multi-model ensemble of seasonal climate forecasts.

Crop simulation models have also been used in climate change studies. An investigation on the response of winter wheat production in France to climate change has been done, by forcing the CERES-Wheat model with raw daily output from the HadCM2 GCM (Mavromatis and Jones, 1998). The HadCM2-based yield predictions were found to correlate well with average yields simulated for the past century using observed weather data as input into the crop model. Furthermore, it was also observed from the results that the CERES-Wheat model captured the trend in yield associated with the trend in observed temperature well, but that the model did not capture the inter-annual variability in yield very well.

1.9.2 Yield Predictions in the United States of America

The United States Department of Agriculture uses a statistical model, which relates weather to yield, for their yield predictions (Lawless and Semenov, 2005). Long-term mean weather data are used as input into this statistical model, but this approach has been demonstrated to be inappropriate for yield predictions, due to the non-linear response of crops to their environment (Porter and Semenov, 1999).

The crop simulation model CORN-CROPS has been used to simulate the interactions of management practices and weather on maize yields in east central Illinois (Hollinger, 1988). Five plant dates, three plant populations, three cultivars, each with a different maturity rating, and weather data for a 14 year period (1970 - 1983) were considered for the analysis. The crop model was set up and run for the 14 year period and for all possible combinations of management inputs. The simulated yields were verified against actual yield estimates for Champaign County. Results revealed a strong agreement between the simulated yields and actual yield estimates, which is evidence that the crop model succeeds in representing the real world. It was concluded that CORN-CROPS has the potential to

influence crop management, by proposing different techniques that could lead to increased yields under the given weather conditions.

An examination on the possibility of using monthly weather projections for soybean yield estimates has been done for the Mississippi Delta (Reddy and Pachepsky, 2000). The crop simulation model GLYCIM and weather projections from three GCMs (GFDL R30, UKMO 89 and NCAR) were used. As crop simulation models normally use daily weather data as input, the monthly projections from the three GCMs were downscaled to daily weather data. Two methods were used to simulate soybean yields. The first method used the downscaled GCM data as input into the GLYCIM crop model and the second method employed a group method of data handling (GMDH) network with monthly weather data as input, to relate soybean yields to CO_2 levels, total solar radiation, average maximum and minimum temperature and rainfall for five months of the growing season. It was found that the GMDH network reproduced the GLYCIM simulated yields with a reasonably high level of accuracy and that this method could be used to obtain general relationships between crop yields and combinations of GCM projected temperature, precipitation and CO_2 concentrations.

1.9.3 Yield Predictions in Asia

To investigate the potential effect of climate change of rain-fed and irrigated maize yields in eastern China, three GCMs were coupled to the CERES-Maize model (Jinghua and Erda, 1996). The three GCMs that were used to produce the climate change projections included the Geophysical Fluid Dynamics Laboratory (GFDL) model, the high-resolution United Kingdom Meteorological Office (UKMO) model and the Max Planck Institute (MPI) model. The seasonal mean changes in temperature and precipitation evident from the output of the GCMs were applied to monthly temperature and precipitation of the baseline climate created by the Chinese Weather Generator, and this climate change data was then used to force the CERES-Maize model. Output from the CERES-Maize model under the climate change scenarios was compared to output from the simulations performed with the baseline climate data. Both rain-fed and irrigated maize yields were found to decrease under climate change conditions. This is a realistic result as an increase in temperature shortens maize growth, particularly the grain filling stage, and consequently results in lower yields.

As so many studies have focused on presenting crop yield forecasts deterministically, crop yield predictability using a probabilistic method has also been explored (Challinor *et al.*, 2005). The use of weather ensembles provides an opportunity for examining crop yield predictability probabilistically. A multi-model ensemble and the GLAM (General Large-Area

Model) crop simulation model were used to predict groundnut yields in western India. Daily output from the 7 CGCMs of the DEMETER project, each running their own ensemble of 9 seasonal hindcasts, were used as input into the GLAM crop model. GLAM was forced with the data of each individual ensemble member to produce both ensemble mean and probabilistic groundnut yield forecasts per district. The ensemble mean yield predictions were found to capture the inter-annual variability in yield relatively well, while predictive skill was found in predicting crop failure probabilistically.

1.9.4 Yield Predictions in Australia

One of the first studies for Australia, attempted to forecast crop yields by relating historical crop yields to the Southern Oscillation Index (SOI) (Nicholls, 1985). A number of years later, an investigation was done on the possibility of producing reliable sorghum yield predictions for the shires in Queensland by combining crop simulation and geographical information system technologies (Rosenthal *et al.*, 1998). Spatial rainfall, temperature and solar radiation data were overlaid and utilized in driving the QSORG sorghum simulation model for 300 locations in Queensland for the period from 1977 – 1988. Linear regression was used to find a relationship between the historical yields and simulated yields to obtain calibration equations for each shire. The predicted yields, at shire and state level, were verified against historical yields using regression analysis. The predicted yields were found to correlate exceptionally well with the historical yields at both shire and state level ($r = 0.96$). When comparing maps of predicted yield to maps of historical yield, it was found that this combined technique captures the spatial distribution of the yield among shires. It was concluded that this system can produce reliable sorghum production estimates on both shire and state level and that there exist good prospects for real-time use, especially in terms of the significance of seasonal climate forecasts on decision making at shire scale.

In Queensland the Agricultural Systems Research Unit developed “Whopper Chopper” software to predict production risk faced by farmers (Cox *et al.*, 2004). The software combined seasonal climate forecasts with crop modelling to assist farmers in selecting the management options that would result in the highest yields under the climatic conditions of the upcoming season. This system allowed farmers to investigate the effect of different plant dates, plant populations, nitrogen fertilizer application rates and many other variables on the expected yield.

1.9.5 Yield Predictions in South America

By using the information contained within seasonal climate forecasts, the potential predictability of maize yields in Ceara, Brazil has been investigated (Sun *et al.*, 2007). The predictability was analyzed for a period from 1971 to 2000 through the use of a 10-member ensemble of seasonal hindcasts produced by a RCM (the Regional Spectral model) nested within a GCM (the ECHAM4.5 model). The RCM integrations were for the main rainy season February-March-April-May. Two variables were considered to estimate the maize yield by means of linear regression in a cross-validated mode. The variables included seasonal mean rainfall and a weather index. In the maize yield simulations performed with the RCM data, the weather index showed superiority over the seasonal mean rainfall. When simulating maize yield with observed weather index values, it was found that the weather index accounts for almost 50% of the variance in maize yield in Ceara. It was concluded that the hindcasts correlate well with the observations and that the nested RCM is skilful in simulating seasonal mean rainfall and within-season weather statistics over the Ceara region.

1.9.6 Yield Predictions in Southern Africa

In the past, research mainly focused on relating crop yields to predictors such as rainfall. Numerous seasonal crop yield predictions for southern Africa have been derived from rainfall forecasts alone, but these forecasts do not account for the response of yields to other climatic variables like temperature, radiation, humidity and wind (Martin *et al.*, 2000).

A method for assessing the impact of drought on maize production in South Africa has been developed, by predicting the response of maize to drought that might occur during the course of the growing season (Du Pisani, 1987). The proposed method was tested for five locations, and historical climate records were used. For each of the variables in the historical climate records, median values were calculated for each month over the entire record. Thus, 12 median values were calculated (January to December) per variable and per location. Thereafter, the months in the historical records that yielded median values closest to the median values calculated over the entire record were used to construct a “median year” for each location (i.e. rainfall for location 1 can be made up of January 1939, February 1969, March 1981, April 1970 etc.). It was investigated whether it is possible to predict climatic impact at the end of December, January and February, by replacing the current season’s weather data with the median year’s data for the remainder of the growing season (January to May, February to May and March to May) and then forcing the CERES-

Maize model with these combined sets of genuine and "median year" data. The yield predictions based on the combined set of genuine and "median year" data were found to correlate well with the yield predictions based on a full set of genuine data. It was concluded that with the use of this method the impact of drought on maize yield should be predictable with a usable level of skill up to four months prior to harvest.

The impact of the ENSO phenomenon on rainfall variability over Zimbabwe and the potential for using ENSO predictions for maize management at site level were investigated (Phillips *et al*., 1998). A period from 1951 to 1991 was selected for the study. The mean NDJ SST anomaly in the Nino-3 region in the Equatorial Pacific Ocean were calculated for each of the 40 seasons and then used to group them into El Niño, La Niña and neutral years. Daily weather data, for the period under investigation, for four sites in Zimbabwe were used to drive the CERES-Maize model. The CERES-Maize model ran with different management strategies (2 nitrogen fertilizer treatments and 3 plant dates) to test for differences in yields between ENSO phases. The simulated yields were compared to observed yields and used as indicator of the potential usefulness of ENSO predictions. High variability in rainfall and high variability in the standard deviation of the simulated yields at each of the sites were found during all the ENSO phases. Thus, it was concluded that although ENSO is one of the most dominant sources of inter-annual climate variability at the four sites under investigation in this study, forecasts based on ENSO categories alone will probably not provide information with a level of skill high enough to be used in maize production decision making. In a similar and very successful study the Nino-3 index was used as the predictor to forecast both rainfall and maize yield for Zimbabwe (Cane *et al*., 1994). The results of this study indicated a stronger relationship between Zimbabwean maize yields and SSTs in the Nino-3 region than between rainfall over Zimbabwe and the SSTs in the Nino-3 region.

Seasonal maize water-stress forecasts for the primary maize-growing regions of South Africa and Zimbabwe have been prepared using a crop water-balance model (Martin *et al*., 2000). Historical gridded climate data for the period from 1961 to 1994 were used to force the simulations with the crop water-balance model. The model calculates water stress on a 0.5° x 0.5° grid from gridded rainfall, temperature, soil water holding capacity, plant date, monthly average wind speed, monthly average sunshine hours, cloud cover, and vapour pressure. Linear regression was then used to relate the output from the model (water-stress) to ENSO indices (SOI and Nino-3) with a 4-month lead-time to harvest. It was found that water-stress forecasts relate more strongly to ENSO than seasonal rainfall alone, but that the water-stress forecasts may provide a useful indication of climate fluctuations.

Exceptionally good results were found when forecasting water-stress for the main maize-growing regions in South Africa using the SOI.

In a more recent study, it was attempted to forecast maize yield for the Highveld region of South Africa using a weather analogue program (WAP) (Du Toit *et al.*, 2001). WAP identifies historical seasons with similar weather characteristics as the current season by considering the up-to-date weather conditions of the current season. Weather data of the five best-fitting seasons were used to force the CERES-Maize model to obtain a maize yield forecast. The impacts of climate variability on the economy of South Africa have also been examined (Jury, 2002). As maize production largely contributes to South Africa's gross domestic product (GDP), one of the objectives was to investigate the impact of rainfall variability on maize yields. July-August-September and September-October-November were identified as the two key seasons. The statistical method used intended to predict fluctuations in the maize yields at a lead-time of 3 to 6 months. A total of 18 predictors were considered. The predictors were selected based on principal component analysis of the maize yield and by correlation and composite mapping with respect to summer rainfall. An adjusted fit of 38% was found for the maize simulations when outgoing long-wave radiation in the central Indian Ocean and the stratospheric quasi-biennial oscillation were used as predictors. It was concluded that more that one-third of the variability in maize yields can be predicted at a lead-time of 6 months (a maize yield forecast issued in November for April).

1.10 AIMS AND APPROACH OF RESEARCH

The principal aim of this dissertation is to investigate the possibility of producing usable maize yield predictions for South Africa by using seasonal climate forecasts from a multi-model ensemble system as input into a crop model, thus simulating the response of the maize to potential climatic conditions.

The skill of the crop model is tested by firstly forcing it with observed weather data. These crop model integrations are performed for each of the magisterial districts in the main maize producing area of South Africa for the period 1979/80 to 1998/99. This simulation system sets the target skill level for the other simulation systems. The simulated maize yields are compared to actual maize yields.

Two crop model-GCM based maize yield simulation systems are described, in which GCM-simulated fields are used to force the crop model for each of the magisterial districts in the main maize producing area of South Africa. The skill of the two crop model-GCM based

simulation systems are tested over the same period used in the target simulations, by comparing the simulated maize yields to actual maize yields.

By combining the simulated maize yields produced by the two crop model-GCM based simulation systems, it can be tested whether the skill of a multi-model system outscores that of the best crop model-GCM based simulation system.

1.11 SUMMARY

To understand the effect of weather on maize yields, the growth stages of the maize plant have been described. The climate of South Africa and factors influencing South Africa's climate and the seasonal predictability of South African rainfall has also been discussed. The status of crop yield forecasting, globally as well as locally, has been summarized. Finally, the chapter concluded with the aim and approach of the research of this dissertation. In the next chapter the data, methods and models used to conduct the research are described in detail.

CHAPTER 2

DESIGN OF THE YIELD SIMULATIONS OVER SOUTH AFRICA

2.1 INTRODUCTION

In this chapter the research methodology applied and data used to obtain the maize yield simulations for South Africa are described in detail. This chapter is divided into five sections. In the first section the area of interest is depicted. The second section discusses the different weather datasets that are used in this study. The crop model that is used to simulate maize growth and development and to estimate the yield is described in section 3 and in section 4 all other data, other than weather data, required by the crop model in order to successfully simulate maize yield as well as the set up of the experiments, are described. The last section discusses the actual maize yields that are used as verification data in this study and the methods used to verify the simulated maize yields produced by the crop model.

2.2 AREA OF INTEREST

Maize is the primary grain crop grown in South Africa (Du Plessis, 2003). Approximately 8 million tons of maize grain is produced annually (Du Toit, 1997). Although maize production takes place across the country under various terrain, soil and climatic conditions (Du Plessis, 2003), the Free State and North-West Province constitute 65% of the total area under maize production in South Africa and 58% of the national maize yield are obtained from these two provinces (Du Toit, 1997). Based on their enormous contribution towards the national maize yield, the Free State and North-West Province are consequently selected as the combined area of interest for this study. Each magisterial district in the Free State and North-West Province are considered in this study, but unfortunately a number of districts are excluded from the investigation due to the lack of actual maize yields.

Three maize production regions are evident within the two provinces under investigation (ARC-GCI, 2008). These regions are based on spatial rainfall, temperature and heat unit differences and are known as the dry/warm western region, temperate eastern region and wet/cool eastern region (Figure 2.1). Different management practices are applied in each of the three regions in order to adapt to the climatic conditions (Du Toit *et al.*, 2000).

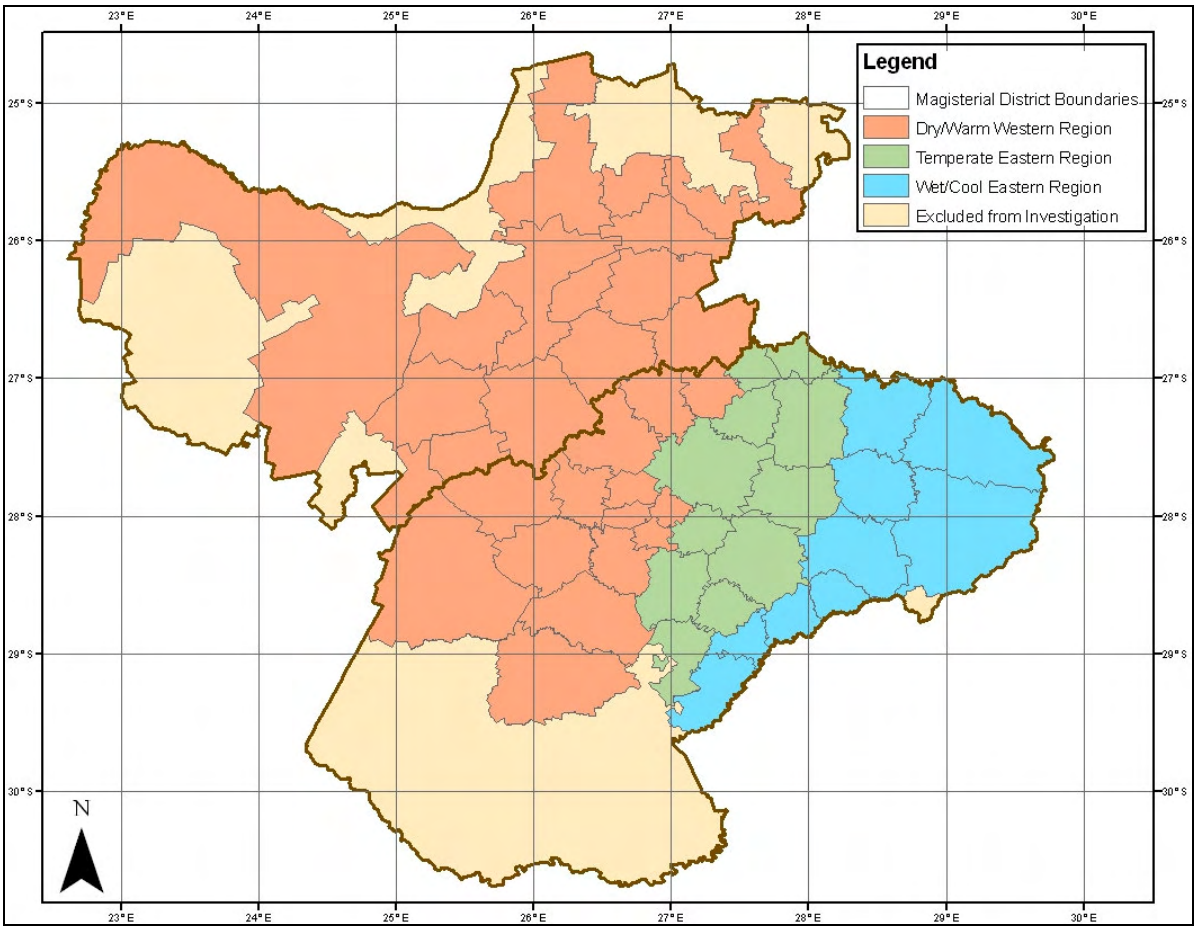

Figure 2.1: The study area and the three maize production regions.

2.3 WEATHER DATASETS

2.3.1 Observed Data

In order to simulate maize yield, daily rainfall, maximum temperature, minimum temperature and solar radiation data are required. Initially, weather station data was selected as the observed weather data in this study, but only 22 weather stations, both Agricultural Research Council and South African Weather Service stations, were identified within the study area that recorded the required variables for the entire period under investigation (1979 to 1999). From the uneven spatial distribution of the identified weather stations over the study area it was clear that the weather station data is not representative of the entire study area and therefore it was decided to use the weather data contained within the Southern African Quaternary Catchment Database (Schulze *et al.,* 2005) as an alternative.

This database comprises of a dense rain gauge network and originated out of a Water Research Commission project that was intended to be used for research purposes (Schulze

et al., 2005). The database contains daily hydroclimatic data for each quaternary catchment in South Africa and for a 50-year period from 1950 to 1999. The hydroclimatic variables that are included in this database are rainfall, maximum temperature, minimum temperature, vapour pressure deficit, minimum relative humidity, maximum relative humidity, solar radiation, Penman-Monteith reference evapotranspiration, soil water content in the A-horizon of the soil, soil water content in the B-horizon of the soil, soil moisture deficit in the A-horizon of the soil, soil moisture deficit in the B-horizon of the soil, saturated drainage from the A-horizon of the soil to the B-horizon of the soil and saturated drainage from the B-horizon of the soil to the groundwater zone.

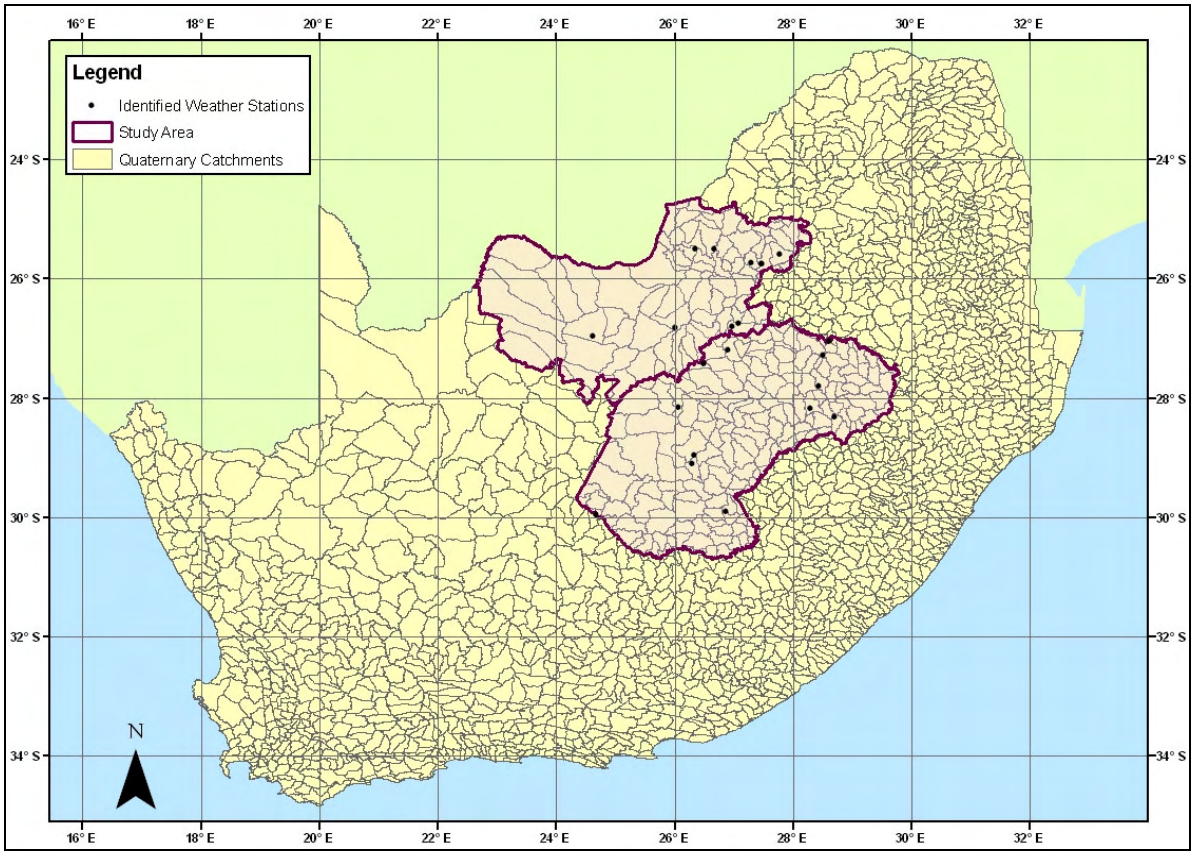

Figure 2.2: The distribution of the identified weather stations (black dots) within the study area compared to the distribution of quaternary catchments within the study area.

Figure 2.2 shows the quaternary catchments for the entire country as well as the positions of the 22 originally identified weather stations in the study area. From this figure it is evident that the quaternary catchment weather data is a much better geographical representation of the study area than the weather station data. Another advantage of the quaternary catchment database is that the data contained within the database closely fit the specific needs of this study. In other words, daily rainfall, temperature and solar radiation data are

available for each quaternary catchment in the study area for the entire period under investigation (1979 to 1999).

2.3.2 Simulated Fields

The numerically simulated fields that are used in this study were produced by two Global Circulation Models (GCMs). They are the Conformal-Cubic Atmospheric Model (CCAM) (McGregor and Dix, 2001) and the ECHAM4.5 model (Roeckner *et al.*, 1996). In order to account for the uncertainty in the initial state of the atmosphere, each model produced an ensemble of simulations from varying initial conditions. These two GCMs form a multi-model system by statistically combining the ensembles from the individual models.

2.3.2.1 CCAM

2.3.2.1.1 Model Description

CCAM is a GCM developed by the CSIRO Marine and Atmospheric Research in Australia (McGregor, 2005a). The model may be integrated in variable-resolution mode with high resolution over an area of interest using the Schmidt stretching factor, thereby allowing it to function as a regional climate model (Engelbrecht *et al.*, 2009). CCAM replaced the limited-area nested climate model DARLAM that was used for regional climate modelling applications (McGregor and Nguyen, 1999; Engelbrecht *et al.*, 2002). Variable-resolution global modelling offers vast flexibility for dynamic downscaling from other GCMs or reanalysis data, effectively requiring only sea surface temperatures (SSTs) and, optionally, far-field winds from the global model in which it is nudged (McGregor and Dix, 2001; Wang *et al.*, 2004).

The model uses a quasi-uniform grid, which is obtained by projecting the six panels of a cube onto the spherical surface of the earth (McGregor and Nguyen, 1999). Since the grid has a fairly uniform resolution over the globe it avoids problems associated with normal latitude-longitude grid projections that require filtering in the vicinity of the poles due to the clustering of grid points (McGregor, 2005a).

CCAM employs a two-time-level, semi-implicit discretization of the hydrostatic primitive equations (McGregor, 1996; McGregor, 2005a). The model also makes use of a semi-Lagrangian scheme for horizontal advection, which in combination with the semi-implicit procedure ensures numerical stability when using large time steps (McGregor, 2005a).

Total-variation-diminishing vertical advection is employed. An unstaggered grid is used, with winds transformed to/from C-staggered locations before/after gravity wave calculations (McGregor, 2005b). More details on the geometrical aspects and dynamic formulation of CCAM can be found in McGregor (2005a).

Furthermore, CCAM comprises of a comprehensive set of physical parameterization schemes. These include the CSIRO mass-flux cumulus convection scheme that takes downdrafts and the evaporation of rainfall into account, the long and shortwave radiation scheme of GFDL (Schwarzkopf and Fels, 1991) with interactive diagnosed cloud distributions (Rotstayn, 1997), a gravity wave drag scheme, a stability-dependent boundary layer scheme with non-local vertical mixing and a soil and canopy scheme describing six soil layers of temperature and moisture as well as three layers for snow (Gordon *et al.*, 2002).

2.3.2.1.2 Design of Simulations

CCAM performed five 25-year (1979 to 2003) integrations for the entire globe on a horizontal grid of approximately 2.1° x 2.1° degrees, with 18 σ-levels in the vertical. The simulations were initialised using a lagged average forecast approach (Hoffman and Kalnay, 1983). Each simulation was forced with observed monthly sea surface temperatures (SSTs), obtained from the Atmospheric Model Intercomparison Project (AMIP) dataset, at its lower boundary. Model output is available for a number of variables on a 1° latitude-longitude grid, at daily time intervals starting on 1 January 1979 and ending on 31 December 2003. These simulations were performed on the Velocity-cluster at the University of Pretoria.

2.3.2.2 ECHAM4.5

2.3.2.2.1 Model Description

The ECHAM4.5 GCM is a primitive equation model that was developed by the Max Planck Institute for Meteorology in Hamburg, Germany (Roeckner *et al.*, 1996). Many features of this model were adopted from the spectral weather prediction model of the European Centre for Medium Range Forecasts (ECMWF). However, a different set of parameterization schemes were employed for ECHAM4.5 than for the ECMWF model.

ECHAM4.5 uses a Gaussian transform grid on which the nonlinear terms and most of the parameterized physics is calculated. Furthermore, a semi-implicit time stepping scheme and weak time filter are used. This avoids the decoupling of the solutions at the two time levels in the time stepping scheme. The model employs a 19 – level hybrid sigma-pressure coordinate system which extends up to 10 hPa in the vertical.

The model occupies a semi-Lagrangian scheme to calculate the transport of water vapour, cloud water and trace constituents (Williams and Rasch, 1994). Land surface data that are supplied to the model include orography, albedo, roughness length, vegetation type, leaf area index, soil water holding capacity, soil heat capacity and soil thermal conductivity.

The physical parameterizations that are incorporated in the model includes a horizontal diffusion scheme which uses a high-order for diffusion in the troposphere, a surface flux and vertical diffusion scheme using the Monin-Obukhov theory to calculate turbulent fluxes at the surface, a land surface processes scheme which comprises of water and heat in the soil, snow pack over land, the heat budget of land ice, interception of rainfall and evapotranspiration, a gravity wave drag scheme using the McFarlane (1987) and Palmer *et al.* (1986) method, a cumulus convection scheme based on the bulk mass flux concept, a stratiform clouds scheme and a radiation scheme (Roeckner *et al.,* 1996).

2.3.2.2.2 Design of Simulations

A 6-member ensemble of simulations was produced by ECHAM4.5 for the period 1979 to 2003. The model ran globally at a horizontal resolution of approximately 2.8° x 2.8° with 19 levels in the vertical. The model was forced at its lower boundary using AMIP SSTs, equivalent to the lower boundary forcing applied in the CCAM simulations. Lagged Average Forecasting was likewise used to initialize the different ECHAM4.5 ensemble members. Model output, in daily time steps, is available on the 2.8° latitude-longitude grid and for a number of variables. These model runs were performed at the South African Weather Service (SAWS).

2.4 CROP GROWTH SIMULATION MODEL

The CERES-Maize model that forms part of the Decision Support System for Agrotechnology Transfer (DSSAT) is used in this study to simulate maize growth and estimate maize yield. A total of 25 model configurations are built into DSSAT for a number

of different crops such as cereals, legumes, root crops, oil crops, vegetables, forages and fruits. The DSSAT cropping system model was selected for this study based on it being an internationally recognized model which is used by researchers worldwide (Jones *et al.*, 2003) including researchers in South Africa (e.g. Du Toit, 1997; Du Toit *et al.*, 2000). In southern Africa the CERES-Maize model has been used to asses the impact of drought on maize production in South Africa (Du Pisani, 1987), to investigate the potential of ENSO predictions for maize management in Zimbabwe (Phillips *et al.*, 1998) and to forecast the maize yield of the Highveld region of South Africa (Du Toit *et al.*, 2001).

2.4.1 Model Description

DSSAT was developed through the collaboration of a number of researchers across the globe (Jones *et al.*, 2003). The development of this cropping system model formed part of the International Benchmark Sites Network for Agrotechnology Transfer (IBSNAT) project (Tsuji, 1998). The main drive behind the development of the DSSAT system was the need to make better decisions about transferring production technology from one site to another, where completely different soil and climate conditions prevail (Uehara and Tsuji, 1998).

In DSSAT a group of independent programs are joined together in order to predict the behaviour of a certain crop under specified conditions (Jones *et al.*, 2003). In other words, the growth, development and yield of a crop is simulated based on prescribed climatic conditions, soil conditions, cultivar specific genetic inputs and management information. The system allows the user to investigate the effect of different management practices on a particular crop in a specified environment (Du Toit *et al.*, 1994).

The CERES-Maize model within DSSAT was designed to simulate maize growth with a minimum set of data (Du Toit *et al.*, 2002). CERES-Maize is a daily time step model and therefore requires daily weather data that includes rainfall, maximum air temperature, minimum air temperature and solar radiation. The model uses the weather data supplied to compute the rate at which the plant progresses from one growth stage to another, daily plant growth, dry matter production, water stress and temperature stress (Jones *et al.*, 2003). A detailed description of the soil, in the form of a one dimensional profile, is also required by the model. Soil input data are used to compute daily changes in soil water content due to the infiltration of rainfall and irrigation, vertical drainage, unsaturated flow, soil evaporation, plant transpiration as well as root water uptake (Jones *et al.*, 2003). Another input required by the model is cultivar specific genetic coefficients (Table 2.1) which describes the phenology of each maize cultivar. Lastly, the CERES-Maize model requires

information on the management practices applied to the specific maize cultivar cultivated under the prescribed climatic and soil conditions. The management information includes data on planting, irrigation, fertilizer, organic amendments, chemical application, tillage and harvest.

Variable	Description
P1	Thermal time from seedling emergence to end of juvenile phase, during which the plant is not responsive to changes in photoperiod
P2	Extent to which development is delayed for each hour increase in photoperiod above the longest photoperiod at which development proceeds at a maximum rate.
P5	Thermal time from silking to physiological maturity
G2	Maximum possible number of kernels per plant.
G3	Kernel filling rate during the linear grain filling stage and under optimum conditions
PHINT	Interval in thermal time between excessive leaf tip appearances

Table 2.1: Genetic coefficients for maize as required by the CERES-Maize model (Jones *et al.*, 2003).

2.4.2 Agricultural Inputs

CERES-Maize model simulations are performed for each magisterial districts in the study area. Consequently, it is necessary to obtain the agricultural inputs required by the model to perform successful simulations for each magisterial district under investigation.

2.4.2.1 Soil Inputs

To simulate maize yield, the CERES-Maize model needs a detailed description of the soil in each magisterial district. The Soil Profile Information System of the Agricultural Research Council, Institute for Soil, Climate and Water contains descriptions and analyses data of soil profiles taken at numerous points all over South Africa. The data of all the soil profiles taken within the study area are extracted from this database. The number of soil profiles in each magisterial district varies. Six magisterial districts are eliminated from the study, as no soil profile data are available for them. These districts include Parys, Odendaalsrus, Fouriesburg, Clocolan, Frankfort and Ventersburg. Figure 2.3 shows the eliminated magisterial districts (light purple colour).

A number of different soils can occur in a single magisterial district, but it is unknown on which of these soils maize is cultivated. To take this uncertainty into account a range of

soils are selected for each magisterial district. The soils are categorized into four classes; high agricultural potential soils, medium agricultural potential soils, low agricultural potential soils and soils not suitable for agriculture. This categorization is based purely on the characteristics (soil depth and texture) of each soil. Out of the several soils that occur within a given magisterial district, a soil in each of the classes is selected for use in this study. Thus, three soils are selected per magisterial district, one with high potential, one with medium potential and one with low potential. Not all of the magisterial districts have a soil in each of the classes. This results in several districts with only two selected soils as well as districts with only one selected soil. Table 2.2 shows the soils that are selected for each of the magisterial districts as well as the soil profile data which the CERES-Maize model requires.

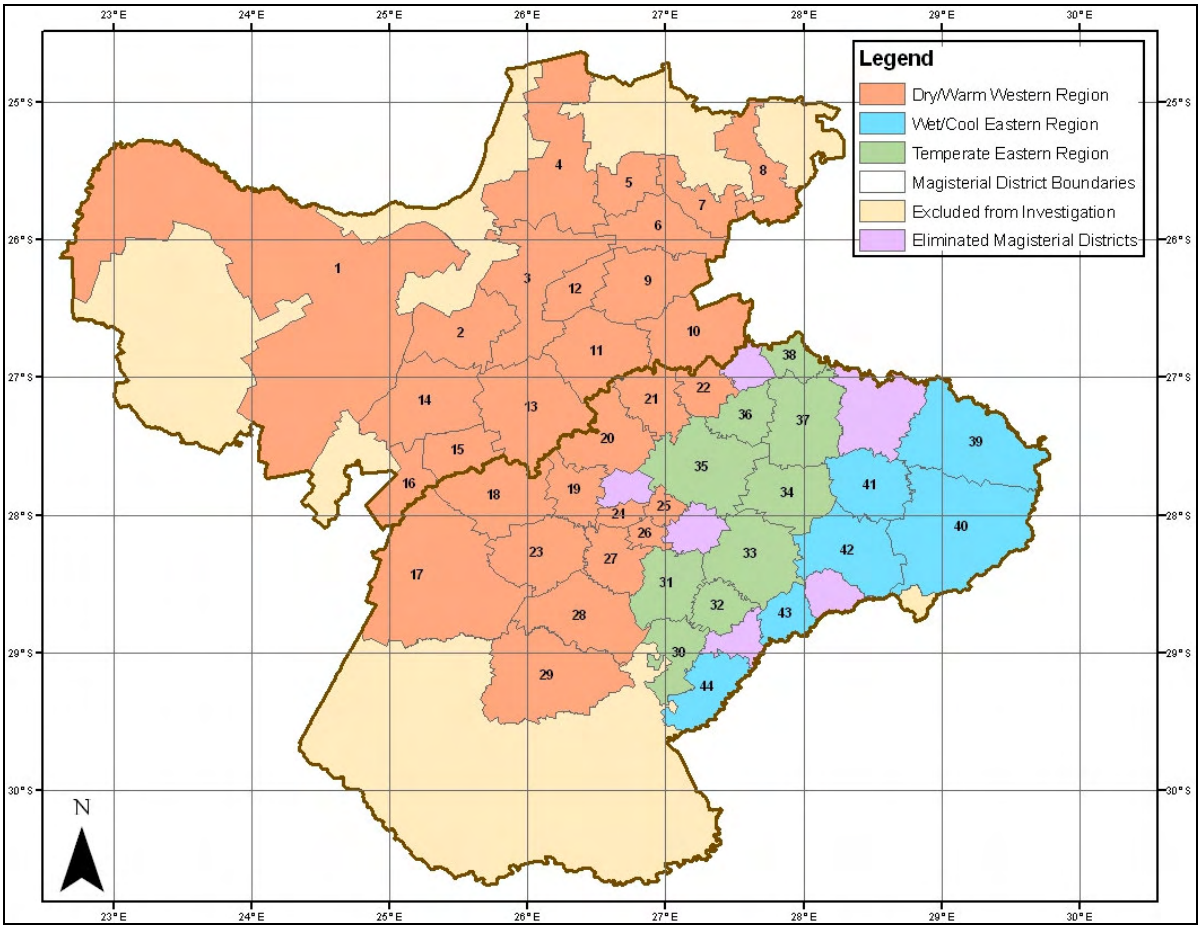

Figure 2.3: The magisterial districts eliminated from the study due to the lack of soil data (light purple colour).

Except for the information in Table 2.2, the CERES-Maize model also requires the colour, drainage, runoff potential, percentage stones, total nitrogen percentage of each soil as well as the slope of the site where the soil profile was taken. The colour of the soils is

determined from Macvicar *et al.* (1977) and Soil Classification Working Group (1991). Drainage are obtained from Schoeman *et al.* (2000) and runoff potential from Schulze (1985). The percentage stones in each of the selected soils are assumed to be zero based on information in the Land Types of South Africa, Soil Inventory Database (Land Type Survey Staff, 1972 – 2008). As the total nitrogen percentage of 500 soil profiles all range between 0.05% and 0.1% (Soil Survey Staff, 2008) an average value of 0.075% is assumed for each of the selected soils in Table 2.2. The slope of the sites are assumed to be perfectly flat (slope = 0%). It is important to keep these assumptions in mind when evaluating the model's performance.

Apart from the soil profile data entered into the CERES-Maize model the model also calculates a number of variables which include the lower limit (LL), drained upper limit (DUL), saturation, bulk density, saturated hydraulic conduct and root growth factor of each soil. These model calculated values can be edited. The LL and DUL is the water holding properties of the soil (Botha and Eisenberg, 1992). The LL can be interpreted as the soil water content at which the development and growth of a plant stops and the DUL as the soil-water content at which drainage from a pre-wetted soil comes to an end (Gebregiorgis and Savage, 2006). Thus, the LL and DUL of the soil plays an important role in crop growth in terms of the amount of water available to the plant. Therefore it was decided to replace the model calculated LL and DUL values (Rawls *et al.,* 1982) with values calculated using the methods described in Botha and Eisenberg (1992). These methods use clay content and cation exchange capacity to calculate soil water retention. It was tested for South Africa for a range of matric potentials, and excellent results were obtained when comparing the calculated water content values with observed values. The Botha and Eisenberg (1992) methods that are used to calculate the new LL and DUL values are as follows:

Lower limit (LL): $LL = 0.393 + 0.2556 * clay\% + 0.04043 * CEC$

Drained Upper Limit (DUL): $DUL = 2.315 + 0.2796 * clay\% + 0.07383 * CEC$

where CEC is the Cation Exchange Capacity in me kg^{-1} and LL and DUL is in %.

Number	Magisterial District	Classification	Soil Name	Master Horizon	Depth bottom (cm)	Clay (%)	Silt (%)	Organic Carbon (%)	pH in Water	Cation Exchange Capacity (cmol/kg)
1	Vryburg	High potential	Clovelly Setlagole (Cv3100)	A	30	9.6	1.4	0.2	6.8	5.0
				B	120	13.8	2.0	0.2	6.1	7.5
		Medium potential	Clovelly Annandale (Cv33)	A1	35	16.7	0.9	0.2	6.1	3.9
				B21	70	11.8	1.3	0.2	5.8	3
		Low potential	Hutton Mangano (Hu33)	A1	25	5.7	0.1	0.1	6.9	2.3
				B2	54	11.9	2.7	0.2	6.3	4.9
2	Delaryville	High potential	Avalon Soetmelk (Av36)	A1	20	11.3	2.2	0.5	5.8	3.2
				B21	90	20.2	3.0	0.2	6.2	4.0
				B22	140	26.4	3.7	0.2	6.2	4.5
		Medium potential	Avalon Mafikeng (Av3200)	A	30	12.1	1.6	0.3	7.0	5.8
				B1	75	27.8	2.5	0.3	7.2	10.3
		Low potential	Glencoe Beatrix (Gc33)	A1	27	10.8	2.0	0.3	6.3	7.4
				B21	60	12.9	1.9	0.1	7.6	3.3
3	Lichtenburg	High potential	Hutton Ventersdorp (Hu3200)	A	35	10.4	2.8	0.5	6.6	6.0
				B	120	19.9	1.8	0.3	6.7	7.2
		Medium potential	Avalon Mafikeng (Av3200)	A	30	11.9	1.7	0.4	7.0	7.3
				B1	90	24.4	2.1	0.3	7.1	13.6
		Low potential	Preiska Hougham (Pr1110)	A	20	19.8	3.9	0.8	7.8	7.5
				B	60	22.8	3.4	0.5	7.7	9.4

4	Marico	High potential	Hutton Msinga (Hu26)	A1	18	13.3	13.2	0.5	5.8	3.8
				B21	62	19.0	9.1	0.4	5.9	4.5
				B22	120	23.7	10.2	0.2	5.9	3.5
		Medium otential	Shortlands Glendale (Sd21)	A1	35	27.9	11.5	0.6	5.9	10.4
				B21	85	42.9	14.1	0.5	6.3	13.6
		Low potential	Hutton Roodepoort (Hu30)	A1	34	5.7	53.1	0.5	5.7	3.5
				B21	80	5.6	3.2	0.3	6.0	3.1
5	Swartruggens	High potential	Hutton Clansthal (Hu24)	A1	23	11.6	3.3	0.4	5.9	2.7
				B2	120	12.7	3.0	0.1	5.8	2.0
6	Koster	High potential	Shortlands Glendale (Sd21)	A1	32	25.0	20.5	0.8	6.7	9.5
				B21	120	37.6	26.4	0.4	6.5	11.1
		Medium potential	Hutton Suurbekom (Hu2200)	A	20	32.5	9.9	1.4	5.8	9.0
				B1	80	40.9	11.9	0.6	6.2	7.5
		Low potential	Clovelly Southwold (Cv26)	A1	27	15.7	4.9	0.6	5.5	3.5
				B2	50	20.0	5.2	0.4	5.3	3.3
7	Rustenburg	High potential	Hutton Doveton (Hu27)	Ap	32	10.2	3.1	0.4	7.2	4.5
				B21	83	38.4	5.8	0.3	7.2	7.3
				B22	120	28.2	2.4	0.3	7.1	6.8
		Medium potential	Hutton Hayfield (Hu2100)	A	30	30.4	5.1	1.3	6.0	6.7
				B	120	34.9	5.6	0.8	5.8	5.4
		Low potential	Westleigh Mareetsane (We2000)	A	25	6.9	4.2	0.5	5.8	1.2
				B1	50	21.8	5.8	0.3	5.5	5.0

8	Brits	High potential	Hutton Ventersdorp (Hu3200)	A	30	15.9	7.8	0.8	7.1	6.2
				B	120	22.8	6.9	0.4	7.9	7.0
		Medium potential	Bloemdal Roodeplaat (Bd3200)	A	30	14.7	6.1	0.7	7.9	5.8
				B1	80	23.6	7.0	0.4	8.0	5.8
		Low potential	Kroonstad Morgendal (Kd1000)	A	20	7.8	8.1	0.8	6.8	5.4
				G	70	9.2	11.9	0.3	7.5	5.4
9	Ventersdorp	Medium potential	Hutton Msinga (Hu26)	A1	20	23.4	7.4	0.7	5.6	6.8
				B2	70	26.6	7.3	0.3	5.9	5.1
		Low potential	Hutton Suurbekom (Hu2200)	A	20	14.6	3.6	0.5	5.1	4.3
				B1	60	23.7	5.3	0.4	6.2	5.7
10	Potchefstroom	High potential	Hutton Shorrocks (Hu36)	A1	58	21.5	15.8	0.6	6.7	7.8
				B2	110	18.1	19.1	0.2	6.9	5.0
11	Klerksdorp	High potential	Hutton Suurbekom (Hu2200)	A	30	14.9	5.5	0.6	5.9	4.8
				B	120	20.0	5.3	0.3	6.3	7.3
		Medium potential	Bainsvlei Amalia (Bv3200)	A	30	15.2	4.1	0.6	6.2	6.0
				B1	90	25.8	4.1	0.3	6.5	8.2
		Low potential	Hutton Stella (Hu3100)	A	25	9.3	2.8	0.3	6.7	3.7
				B	60	13.4	3.3	0.3	6.6	3.4
12	Coligny	High potential	Hutton Shorrocks (Hu36)	A1	20	24.9	8.0	0.6	6.3	6.5
				B21	56	33.5	13.4	0.6	6.2	8.4
				B22	100	32.7	19.2	0.4	6.0	8.0
		Medium potential	Avalon Mafikeng (Av3200)	A	30	18.4	5.0	0.5	6.2	6.8
				B1	90	31.5	5.4	0.6	6.1	11.9
		Low potential	Bloemdal Vrede (Bd3100)	A	30	6.9	2.7	0.3	7.9	3.1
				B1	70	11.5	3.3	0.2	6.8	3.5

13	Wolmaransstad	High potential	Avalon Mafikeng (Av3200)	A	30	13.5	2.9	0.4	5.2	5.5
				B1	100	21.2	2.6	0.3	6.8	6.7
		Medium potential	Hutton Shorrocks (Hu36)	A1	20	10.3	5.3	0.4	6.2	5.0
				B2	80	17.2	4.7	0.3	6.4	5.4
		Low potential	Westleigh Mareetsane (We2000)	A	30	9.5	1.5	0.3	7.1	6.0
				B2	80	25.9	1.5	0.3	6.5	12.2
14	Schweizer Reneke	High potential	Hutton Portsmouth (Hu35)	A1	24	10.9	2.4	0.2	6.3	5.7
				B2	130	6.9	3.1	0.2	6.2	3.8
		Medium potential	Hutton Ventersdorp (Hu3200)	A	20	14.1	3.3	0.3	6.1	6.9
				B	70	26.2	3.2	0.3	6.7	10.7
		Low potential	Glencoe Vlakput (Gc3200)	A	30	8.3	1.9	0.3	6.8	6.1
				B1	80	16.1	3.1	0.3	6.8	12.2
15	Bloemhof	Medium potential	Hutton Shorrocks (Hu36)	A1	20	14.9	5.4	0.3	6.1	6.6
				B2	80	25.6	5.3	0.3	6.3	8.6
16	Christiana	Low potential	Hutton Shorrocks (Hu36)	A1	25	11.4	1.3	0.4	7.8	6.6
				B2	55	24.5	1.6	0.4	7.2	12.3
17	Boshof	High potential	Hutton Mangona (Hu33)	A1	35	7.4	1.4	0.3	7.6	3.2
				B21	75	11.2	0.8	0.2	7.3	3.2
				B22	120	14.5	0.6	0.1	6.8	4.3
		Medium potential	Shortlands Kinross (Sd20)	A1	32	17.8	5.4	0.6	6.8	8.1
				B21	95	28.4	4.7	0.4	7.1	13.9
		Low potential	Sterkspruit Bakklysdrif (Ss13)	A1	12	12.5	2.0	0.3	8.1	6.2
				B21	40	28.2	4.8	0.7	7.9	15.3

18	Hoopstad	High potential	Clovelly Makuya (Cv34)	A1	40	8.0	0.2	0.2	8.5	4.7
				B2	120	10.6	1.3	0.2	8.4	5.7
		Medium potential	Oaklea Limpopo (Oa46)	B2	90	23.3	6.0	0.2	8.8	8.4
				C	130	22.5	36.5	0.1	9.0	7.6
19	Wesselsbron	High potential	Hutton Mangano (Hu33)	A1	56	8.4	2.5	0.2	8.5	2.9
				B21	92	12.6	2.4	0.2	6.5	3.6
				B22	130	18.7	2.2	0.2	6.4	5.5
		Low potential	Sterkspruit Stanfort (Ss23)	A1	18	8.3	2.4	0.4	7.4	3.6
				B2	36	25.6	4.5	0.6	9.1	13.7
20	Bothaville	High potential	Avalon Heidelberg (Av34)	A1	35	5.9	4.9	0.2	7.0	2.8
				B21	100	15.0	4.2	0.2	6.7	4.0
				B22	240	18.2	1.5	0.2	7.6	6.3
		Medium potential	Hutton Shorrocks (Hu36)	A1	25	24.7	3.1	0.3	6.7	9.1
				B2	75	25.4	3.8	0.2	6.8	9.1
		Low potential	Bonhein Weenen (Bo40)	A1	31	17.5	3.9	0.6	8.7	11.0
				B2	70	25.6	4.2	0.2	7.9	13.0
21	Viljoenskroon	Medium potential	Glencoe Weltevrede (Gc14)	A1	55	5.7	0.6	0.3	5.1	1.7
				B21	100	8.8	4.8	0.2	4.8	2.3
				B22	110	9.0	3.3	0.2	5.8	3.0
		Low potential	Westleigh Sibasa (We13)	A11	20	11.6	1.4	0.2	6.2	3.7
				A12	30	21.7	2.3	0.3	5.7	6.5
				B2	60	36.8	4.4	0.3	6.1	11.8
22	Vredefort	High potential	Hutton Msinga (Hu26)	A1	35	20.7	5.7	0.5	6.1	4.6
				B2	87	30.2	8.5	0.3	6.0	5.1
				B3	115	23.3	9.7	0.2	5.7	5.1

23	Bultfontein	High potential	Clovelly Annandale (Cv33)	A1	34	6.2	1.2	0.3	7.1	2.7
				B21	67	8.5	1.1	0.2	7.4	2.4
				B22	115	16.7	1.0	0.1	7.2	3.8
		Low potential	Hutton Shorrocks (Hu36)	A1	20	12.6	4.9	0.5	6.5	6.6
				B2	60	26.1	4.6	0.4	6.1	11.4
24	Welkom	High potential	Avalon Soetmelk (Av36)	A1	40	33.6	2.8	0.4	7.5	3.8
				B21	80	23.3	4.4	0.4	7.2	6.2
				B22	120	25.7	2.4	0.2	7.2	5.3
25	Henneman	Medium potential	Hutton Shorrocks (Hu36)	A1	28	17.8	1.5	0.6	7.2	7.3
				B2	72	28.2	2.4	0.5	6.9	12.6
26	Virginia	Medium potential	Clovelly Blinkklip (Cv36)	A1	35	11.3	1.2	0.4	7.7	3.5
				B21	95	22.7	1.2	0.3	7.3	6.6
27	Theunissen	High potential	Clovelly Blinkklip (Cv36)	A1	35	12.5	1.9	0.3	6.9	3.2
				B21	120	25.0	2.2	0.3	6.9	5.8
		Low potential	Valsrivier Waterval (Va11)	A1	25	13.8	3.1	0.4	6.7	5.1
				B21	65	43.0	3.7	0.4	7.1	11.3
28	Brandfort	High potential	Oaklea Limpopo (Oa46)	Ap	20	20.0	9.4	0.5	7.9	10.7
				B21	50	26.3	10.1	0.2	8.8	13.9
				B22	120	27.0	17.0	0.2	9.6	14.0
		Medium potential	Shortlands Glendale (Sd21)	A1	28	19.9	3.8	0.6	6.9	5.8
				B21	68	40.7	2.6	0.3	7.0	12.8

29	Bloemfontein	High potential	Hutton Shorrocks (Hu36)	A1	30	12.9	1.4	0.4	8.2	5.4
				B21	65	24.8	0.6	0.4	7.3	9.5
				B22	120	31.8	1.9	0.3	7.1	11.5
		Medium potential	Bainsvlei Bainsvlei (Bv36)	A1	30	10.8	2.1	0.3	7.6	3.9
				B21	65	27.1	1.8	0.3	6.9	6.9
		Low potential	Valsrivier Waterval (Va11)	A1	33	9.1	2.4	0.4	6.9	4.5
				B21	65	37.9	0.6	0.5	6.6	11.6
				B22	90	33.3	1.7	0.3	6.9	12.0
30	Excelsior	Medium potential	Westleigh Rietvlei (We12)	A1	35	16.4	9.2	0.5	5.7	5.7
				B21	60	33.1	10.7	0.4	6.3	8.4
				B22	90	70.5	6.1	0.1	6.5	18.1
		Low potential	Valsrivier Arniston (Va31)	A1	30	12.3	8.1	0.7	6.0	5.1
				B2	60	47.6	6.7	0.7	6.9	16.6
31	Winburg	Medium potential	Westleigh Sibasa (We13)	A1	40	18.1	3.5	0.4	7.8	5.3
				B21	85	33.6	2.8	0.4	6.8	8.4
		Low potential	Valsrivier Arniston (Va31)	A1	30	14.6	11.5	0.4	5.7	4.9
				B21	55	45.7	10.7	0.6	7.1	14.0
32	Marquard	Medium potential	Oaklea Leeufontain (Oa16)	A1	28	14.3	7.6	0.6	5.9	5.3
				B2	60	24.3	8.3	0.3	6.4	6.9
		Low potential	Westleigh Sibasa (We13)	A1	34	11.2	9.9	0.3	6.5	3.6
				B2	60	53.2	7.2	0.6	6.3	13.1
33	Senekal	Medium potential	Westleigh Sibasa (We13)	A1	58	12.8	2.3	0.3	6.2	5.0
				B2	84	55.5	4.0	0.3	6.2	11.7
		Low potential	Valsrivier Sheppardvale (Va42)	A1	23	22.1	8.0	0.7	6.8	7.3
				B2	54	55.2	16.1	0.7	7.6	18.6

34	Lindley	Medium potential	Hutton Msinga (Hu26)	A1	55	11.5	3.8	0.5	5.9	3.5
				B2	90	15.9	4.9	0.4	5.6	3.6
35	Kroonstad	High potential	Avalon Soetmelk (Av36)	A1	32	17.2	3.9	0.4	5.9	4.6
				B21	63	31.5	5.1	0.4	5.8	7.0
				B22	115	39.8	6.3	0.2	6.0	8.3
		Medium potential	Bainsvlei Bainsvlei (Bv36)	A1	30	12.7	2.2	0.3	6.4	3.7
				B21	60	20.7	2.7	0.3	6.4	5.3
				B22	100	23.0	4.6	0.3	6.6	8.2
36	Koppies	High potential	Hutton Schorrocks (Hu36)	A1	49	9.7	2.1	0.2	6.3	4.1
				B21	86	16.0	2.3	0.2	5.8	5.7
				B22	120	18.5	2.3	0.1	6.1	5.1
37	Heilbron	Medium potential	Clovelly Southwold (Cv26)	A1	33	23.9	10.4	0.7	6.0	6.5
				B2	67	30.4	10.1	0.5	5.8	6.9
38	Sasolburg	High potential	Avalon Mooiveld (Av31)	A1	40	3.3	1.3	0.3	5.1	1.5
				B21	110	3.1	0.6	0.1	4.9	1.0
				B22	120	9.5	0.8	0.1	5.1	1.7
39	Vrede	High potential	Avalon Bezuidenhout (Av37)	A1	60	22.9	4.8	0.9	6.3	6.1
				B21	92	36.0	5.0	0.6	5.8	9.2
				B22	110	43.8	8.2	0.2	6.3	9.9
		Medium potential	Clovelly Clovelly (Cv17)	A1	28	34.1	16.7	1.1	5.2	7.3
				B2	70	35.2	17.4	0.6	5.1	5.5
		Low potential	Clovelly Southwold (Cv26)	A1	40	16.3	4.2	0.7	5.6	4.2
				B2	65	18.1	1.8	0.5	5.3	3.8

40	Harrismith	High potential	Avalon Bezuidenhout (Av37)	A1	45	30.5	9.3	1.2	6.0	10.3
				B21	90	45.9	11.7	0.6	6.6	13.5
				B22	120	55.1	13.7	0.5	6.8	19.3
		Medium potential	Oaklea Jozini (Oa36)	A1	40	25.8	13.1	1.1	5.7	10.0
				B2	85	28.7	14.8	0.5	5.8	9.5
		Low potential	Longlands Waaisand (Lo11)	A1	35	17.5	10.7	0.7	5.9	5.3
				E	55	11.8	10.3	0.2	6.9	2.4
				B2	70	21.8	11.7	0.1	7.4	6.2
41	Reitz	High potential	Avalon Bleeksand (Av33)	A1	50	8.0	3.0	0.9	5.8	4.1
				B21	95	8.0	2.0	0.2	5.5	2.4
				B22	140	7.0	3.0	0.1	6.2	1.5
		Medium potential	Clovelly Blinkklip (Cv36)	A1	30	19.4	9.0	1.3	6.3	7.9
				B21	60	16.6	5.7	0.7	6.4	4.7
				B22	92	23.5	5.5	0.8	6.7	5.9
		Low potential	Longlands Waaisand (Lo11)	A1	35	17.5	10.7	0.7	5.9	5.3
				E	55	11.8	10.3	0.2	6.9	2.4
				B2	70	21.8	11.7	0.1	7.4	6.2
42	Bethlehem	High potential	Avalon Bezuidenhout (Av37)	A1	28	21.4	3.0	1.0	5.8	7.3
				B21	52	35.2	3.2	0.5	5.9	7.8
				B22	80	43.2	4.7	0.2	5.2	11.7
				B23	106	39.9	5.7	0.2	5.5	14.3
		Medium potential	Oaklea Jozini (Oa36)	A1	40	11.0	3.0	0.6	6.9	8.2
				B2	87	17.3	8.2	0.9	7.8	9.4
		Low potential	Sterkspruit Sterkspruit (Ss26)	A1	22	27.8	10.7	0.7	7.5	9.1
				B2	44	60.2	10.5	0.6	7.7	22.5

43	Ficksburg	High potential	Avalon Avalon (Av26)	A1	60	13.3	4.8	0.5	6.0	4.3
				B21	110	19.5	5.3	0.2	6.3	4.4
				B22	140	45.5	6.6	0.2	6.0	8.7
44	Ladybrand	Medium potential	Avalon Soetmelk (Av36)	A1	58	12.4	0.2	0.4	5.9	4.0
				B21	90	17.9	5.9	0.3	6.2	5.1
				B22	130	43.8	6.8	0.3	6.0	9.6
		Low potential	Estcourt Dohne (Es13)	A1	34	11.8	11.7	0.4	6.2	4.0
				E	56	9.8	13.1	0.2	6.5	2.3
				B2	90	31.2	13.4	0.3	7.0	7.8

Table 2.2: The soils selected for each magisterial district and the corresponding soil profile data (Soil Survey Staff, 2008) as used as input for the CERES-Maize model.

2.4.2.2 Cultivar Inputs

It is unknown which maize cultivars were planted in each of the magisterial districts over the period under investigation (1979 to 1999). Due to the different climatic conditions, the maize cultivars planted in each of the production regions in the study area (Figure 2.1) differ. Cultivar coefficients for a number of maize cultivars that were planted in each of the regions are obtained from the Agricultural Research Council, Grain Crops Institute. In terms of the number of days from planting to maturity, three types of maize cultivars can generally be distinguished. They are short season maize (60 - 75 days to flowering), medium season maize (65 - 80 days to flowering) and long season maize (70 - 85 days to flowering). It was decided to select three cultivars, a short season maize cultivar, a medium season maize cultivar and a long season maize cultivar, for each of the production regions in the study area. The three cultivars selected for each of the respective production regions, are used in the CERES-Maize model simulations performed for the magisterial districts that fall within each region. Table 2.3 shows the cultivar coefficients, as needed by the CERES-Maize model (Table 2.1), for the three cultivars that are selected for each of the production regions. Thus, as the selected cultivars remain the same over the entire period under investigation, it must be kept in mind that this study does not account for the change in hardiness of maize cultivars over time.

Production Region	Growing season length	Cultivar Name	P1	P2	P5	G2	G3	PHINT
Dry/Warm Western Region (Magisterial districts 1 – 29)	Short	DK618	198.7	0.659	871.0	945.8	14.13	82.34
	Medium	CRN4526	218.0	0.660	999.9	618.0	7.11	75.00
	Long	SR52	281.0	1.000	999.9	422.1	7.78	75.00
Temperate Eastern Region (Magisterial districts 30 – 38)	Short	DK61_24	283.2	0.957	979.0	999.0	18.28	99.00
	Medium	RO411	271.0	0.999	999.9	505.0	7.28	75.00
	Long	PAN6528	241.0	0.660	999.9	734.0	6.80	75.00
Wet/Cool Eastern Region (Magisterial districts 39 – 44)	Short	KK8202	211.0	0.659	999.0	839.9	14.71	82.34
	Medium	PAN6552	221.0	0.660	999.9	546.5	7.55	75.00
	Long	Tx24	220.0	0.990	999.0	592.0	7.40	75.00

Table 2.3: Cultivar coefficients for the selected cultivars used in the CERES-Maize model simulations performed for the magisterial districts under investigation.

2.4.2.3 Management Inputs

Different management practices are applied in each of the three production regions in the study area. Thus, the following management inputs, described per production region, are used in the CERES-Maize model simulations performed for the magisterial districts that fall within each of the respective regions.

2.4.2.3.1 Plant Dates

Broad optimal maize plant dates for the three respective production regions can be summarized as follows; the dry/warm western region from the second week in November to middle December, the temperate eastern region from the last week in October to middle November and the wet/cool eastern region from the beginning of October to the first week in November (Du Toit, 1997). As sufficient soil water is needed before planting can take place (Walter, 1967), these optimal plant dates for maize are probably based on the onset of the first rains. The inter-annual variability of the onset of the maize growing season over South Africa has been investigated (Tadross *et al.*, 2003). From the data used to analyse the mean onset, it could be seen that the eastern parts of the country receive earlier rains than the western parts of the country. In a study which investigated the rainfall seasonality over South Africa, the exact same results were found (Schulze and Maharaj, 2007).

Production Region	Growing season length	Plant date 1	Plant date 2	Plant date 3
Dry/Warm Western Region (2nd week in Nov. – middle Dec.)	Short	30-Nov	05-Dec	10-Dec
	Medium	20-Nov	25-Nov	30-Nov
	Long	10-Nov	15-Nov	20-Nov
Temperate Eastern Region (last week in Oct. – middle Nov.)	Short	10-Nov	15-Nov	20-Nov
	Medium	30-Oct	05-Nov	10-Nov
	Long	20-Oct	25-Oct	30-Oct
Wet/Cool eastern Region (beginning of Oct. – 1st week in Nov.)	Short	25-Oct	30-Oct	05-Nov
	Medium	15-Oct	20-Oct	25-Oct
	Long	05-Oct	10-Oct	15-Oct

Table 2.4: The range of possible plant dates selected for the cultivars in Table 2.3.

Maize cultivars with different growing season lengths are planted on different plant dates. As it is unknown which of the growing season length cultivars (short, medium or long) were planted and on which plant dates, a range of possible plant dates (within the optimal

planting period for each of the production regions) are selected for each of the cultivars in Table 2.3. Table 2.4 shows the plant dates that are selected, based on the fact that the days to flowering only increase with a maximum of 10 days between a short and medium season maize cultivar and 10 days between a medium and long season maize cultivar (ARC-GCI, 2008).

2.4.2.3.2 Planting Depth

The CERES-Maize model also requires the depth at which the dry maize seed are planted. Planting depths for maize normally range between 5 cm and 10 cm depending on the plant date and depth of the soil (Du Toit, 1997). Early plantings can be planted shallower (Du Toit, 1997). Only the plant dates are considered to obtain planting depths for this study. Table 2.5 shows the planting depths selected based on the plant dates in Table 2.4. The long season maize cultivars are planted early in the season and therefore planted shallower than the other cultivars, while the short season maize cultivars are planted later in the season and therefore planted deeper than the other cultivars.

Production Region	Growing season length	Planting depth on Plant date 1	Planting depth on Plant date 2	Planting depth on Plant date 3
Dry/Warm Western Region	Short	10 cm	10 cm	10 cm
	Medium	7.5 cm	7.5 cm	7.5 cm
	Long	5 cm	5 cm	5 cm
Temperate Eastern Region	Short	10 cm	10 cm	10 cm
	Medium	7.5 cm	7.5 cm	7.5 cm
	Long	5 cm	5 cm	5 cm
Wet/Cool eastern Region	Short	10 cm	10 cm	10 cm
	Medium	7.5 cm	7.5 cm	7.5 cm
	Long	5 cm	5 cm	5 cm

Table 2.5: The planting depths selected for each of the selected cultivars in Table 2.3.

2.4.2.3.3 Row Spacing & Plant Population

In South Africa wide rows (150 cm – 210 cm) are used in low to medium rainfall areas, whereas narrow rows (90 cm – 100 cm) are used in medium to high rainfall areas (Du Toit, 1997). The wider rows in the drier areas are normally accompanied by low plant populations. This production method eliminates competition between the plants by ensuring

that each plant has access to soil water (Du Toit *et al.*, 2000). Narrower rows and more dense plant populations are allowed in the wetter parts, as production are not as water-limited as in the drier parts (Du Toit *et al.*, 2000).

As the dry/warm western production region (see Figure 2.1) receives low to medium rainfall, it is assumed that wide rows and low plant populations are used. The temperate and wet/cool eastern production regions receive medium to high rainfall and are therefore assumed to utilize narrow rows and more dense plant populations.

Since it is unknown where in the dry/warm western production region 150 cm and where 210 cm rows are used, a middle value of 180 cm is chosen for the entire region. A row spacing of 90 cm is selected for the wet/cool eastern production region and 100 cm for the temperate eastern production region.

To acquire a maize yield of 3 t/ha in each of the production regions in the study area, different plant populations needs to be used. A plant population of 14 000 plants/ha are assumed for the dry/warm western region, 16 000 plants/ha for the temperate eastern region and 19 000 plants/ha for the wet/cool eastern region (Du Toit, 1997).

2.4.2.3.4 Irrigation

As this study focuses on rain-fed maize, no irrigation is applied to the experiments done for each of the magisterial districts under investigation.

2.4.2.3.5 Fertilizer

The CERES-Maize model simulations for each of the magisterial districts are nitrogen non-limited. Thus, the plants do not experience any N-stress. The CERES-Maize model supplies the plant with Nitrogen as needed.

2.4.2.3.6 Harvest

The CERES-Maize model harvests the maize at maturity.

2.4.2.3.7 Other Assumptions

Furthermore it is assumed that weeds, pests and diseases are controlled and that nutrients are not limited. Over the period modelled in this study technology is kept constant. These assumptions may all contribute to an overestimation in simulated yields (Rosenzweig and Iglesias, 1994).

2.4.3 Incorporating the Weather Data into the CERES-Maize Model

2.4.3.1 Observed Data

Daily rainfall, maximum temperature, minimum temperature and solar radiation data for the period 1979 to 1999 are extracted from the Southern African Quaternary Catchment Database (Schulze *et al.,* 2005) for each of the quaternary catchments in the study area. By superimposing the magisterial district boundaries over the quaternary catchment boundaries in the study area, the quaternary catchments that fall either completely or partially within each magisterial district are identified. Table 2.6 lists the quaternary catchments included in each magisterial district. The daily data of the quaternary catchments in a given magisterial district are then averaged in order to obtain a single weather dataset per magisterial district. These average daily values per district are then used as input for the CERES-Maize model to perform the observed weather data simulations.

Number	Magisterial District	Quaternary Catchments
1	Vryburg	C32A, C32B, C32D, C33A, D41B, D41C, D41D, D41E, D41F, D41H
2	Delaryville	C31B, C31D, C31E, C32C, C41B
3	Lichtenburg	C31A, C31B, C31C, D41A
4	Marico	A10C, A31A, A31B, A31C, A31E, A31F, A31J, A32C, A32D, D41A
5	Swartruggens	A22A, A22D, A22E, A31B, A31G
6	Koster	A22A, A22B, A22C, A22G, C23F, C24C
7	Rustenburg	A21K, A22C, A22D, A22F, A22G, A22H
8	Brits	A21H, A21J, A21K, A21L, A22J, A23L, A23A
9	Ventersdorp	C23F, C23G, C24C, C24D, C24E
10	Potchefstroom	C23C, C23G, C23H, C23J, C23K, C23L
11	Klerksdorp	C24A, C24G, C24H, C24J
12	Coligny	C24F
13	Wolmaransstad	C25A, C25C, C25D, C25E
14	Schweizer Reneke	C31F, C31E, C32C, C32D, C91A
15	Bloemhof	C25F, C91A

16	Christiana	C91B
17	Boshof	C52H, C52K, C91B, C91C, C91D
18	Hoopstad	C25F, C43C, C43D, C91A, C91B
19	Wesselsbron	C25B, C25F, C43B
20	Bothaville	C24J, C25B, C25C, C60J
21	Viljoenskroon	C24B, C24J, C70J, C70K
22	Vredefort	C23C, C23L, C70E, C70F, C70J
23	Bultfontein	C43A, C43C, C43D, C52H, C91C
24	Welkom	C42J, C43B
25	Henneman	C42J, C60H
26	Virginia	C42H, C42J, C42K
27	Theunissen	C41G, C41J, C42K, C42L
28	Brandfort	C41F, C41H, C52C, C52E, C52G, C52H
29	Bloemfontein	C51D, C51E, C52B, C52C, C52D, C52E, C52F, C52G, C52H, C52J
30	Excelsior	C41C, C41D, D23C
31	Winburg	C41A, C41B, C41D, C41E, C42E, C42G, C42K
32	Marquard	C41A, C41B, C42E
33	Senekal	C42A, C42B, C42C, C42D, C42E, C42F, C60E, D22A
34	Lindley	C60A, C60B, C60C, C70A, C83F
35	Kroonstad	C60C, C60D, C60F, C60G, C60H, C70D, C70G, C70H
36	Koppies	C70C, C70D, C70E, C70F, C70G
37	Heilbron	C22G, C23A, C70A, C70B, C70C, C83K, C83L, C83M
38	Sasolburg	C22F, C22G, C22K
39	Vrede	C12A, C12B, C12C, C12D, C13C, C13E, C13F, C13G, C13H, C82E, C82F, C82H
40	Harrismith	C13C, C81A, C81B, C81C, C81D, C81E, C81F, C81G, C81H, C81J, C81K, C81L, C81M, C82A, C82B, C82C, C82E
41	Reitz	C82D, C82G, C83E, C83F, C83G, C83H
42	Bethlehem	C42A, C60A, C81G, C83A, C83B, C83C, C83D, D21D
43	Ficksburg	D21H, D22A, D22B, D22D
44	Ladybrand	D22G, D22H, D22L, D23A, D23C, D23D, D23E

Table 2.6: The quaternary catchments either completely or partially within the magisterial districts under investigation.

2.4.3.2 Simulated Fields

2.4.3.2.1 CCAM

CCAM simulated daily rainfall, maximum temperature and minimum temperature are available on a 1° x 1° grid for each of the five ensemble members. The daily weather information at the CCAM grid point closest to the centre of a magisterial district is

objectively selected to represent the CCAM-simulated weather for that district (nearest neighbour approach). This is then used in the CERES-Maize model simulations performed for each district. Solar radiation was not available from the CCAM output, but is needed by the CERES-Maize model to simulate maize yield. To overcome this limitation, a daily solar radiation climatology is calculated for each of the magisterial districts using the observed weather data. This solar radiation data then makes up the complete dataset needed by the CERES-Maize model to simulate the yield.

2.4.3.2.2 ECHAM4.5

ECHAM4.5 simulated data are extracted for the domain 16°E to 34°E and 22°S to 35°S. Daily simulated rainfall, maximum temperature and minimum temperature data are available for each of the six ensemble members. The daily weather information at the ECHAM4.5 grid point closest to the centre of a magisterial district is objectively selected to represent the ECHAM4.5-simulated weather for that district (nearest neighbour approach). The simulated data of that specific ECHAM4.5 point are then used to run the CERES-Maize model for the corresponding magisterial district. As no ECHAM4.5 simulated solar radiation data are available, the same daily solar radiation climatology used in the CERES-Maize model runs performed with the CCAM-simulated fields are also used in the CERES-Maize model runs performed with the ECHAM4.5-simualted fields.

2.4.4 Set up of CERES-Maize Model Experiments

Due to the uncertainty in the soil on which maize was planted, the cultivar that was planted and when the maize was planted, a number of options are considered for each of these inputs (Table 2.2, 2.3 and 2.4). To investigate each soil option, with each cultivar option and with each plant date option, a number of scenarios resulted. A total of 27 scenarios can be set up for magisterial districts with three selected soils, 18 scenarios for districts with two selected soils and 9 scenarios for districts with only one selected soil. Table 2.7 shows all the possible scenarios that can be set up, while keeping the other management inputs (planting depth, plant population and row spacing) and weather inputs constant.

Scenario	Description
1	Short season maize planted on plant date 1 and on a high potential soil
2	Short season maize planted on plant date 1 and on a medium potential soil
3	Short season maize planted on plant date 1 and on a low potential soil
4	Short season maize planted on plant date 2 and on a high potential soil

5	Short season maize planted on plant date 2 and on a medium potential soil
6	Short season maize planted on plant date 2 and on a low potential soil
7	Short season maize planted on plant date 3 and on a high potential soil
8	Short season maize planted on plant date 3 and on a medium potential soil
9	Short season maize planted on plant date 3 and on a low potential soil
10	Medium season maize planted on plant date 1 and on a high potential soil
11	Medium season maize planted on plant date 1 and on a medium potential soil
12	Medium season maize planted on plant date 1 and on a low potential soil
13	Medium season maize planted on plant date 2 and on a high potential soil
14	Medium season maize planted on plant date 2 and on a medium potential soil
15	Medium season maize planted on plant date 2 and on a low potential soil
16	Medium season maize planted on plant date 3 and on a high potential soil
17	Medium season maize planted on plant date 3 and on a medium potential soil
18	Medium season maize planted on plant date 3 and on a low potential soil
19	Long season maize planted on plant date 1 and on a high potential soil
20	Long season maize planted on plant date 1 and on a medium potential soil
21	Long season maize planted on plant date 1 and on a low potential soil
22	Long season maize planted on plant date 2 and on a high potential soil
23	Long season maize planted on plant date 2 and on a medium potential soil
24	Long season maize planted on plant date 2 and on a low potential soil
25	Long season maize planted on plant date 3 and on a high potential soil
26	Long season maize planted on plant date 3 and on a medium potential soil
27	Long season maize planted on plant date 3 and on a low potential soil

Table 2.7: The range of possible scenarios resulting from different combinations of input data.

After setting up the scenarios for each of the magisterial districts, three different sets of maize yield simulations are performed by the CERES-Maize model. Firstly, the CERES-Maize model is forced with observed weather data (Figure 2.4), secondly with CCAM-simulated fields (Figure 2.5) and thirdly with ECHAM4.5-simulated fields (Figure 2.6). One observed weather data run, five CCAM-simulated field runs (one for each ensemble member) and six ECHAM4.5-simulated field runs (one for each ensemble member) are performed by the CERES-Maize model for each magisterial district. Thus, maize yield is simulated for each scenario under observed, CCAM-simulated and ECHAM4.5-simulated weather conditions. A total of 10 044 simulations are performed with the CERES-Maize model. The Multi-Model system is obtained by combining the ensemble of CERES-CCAM simulated maize yields and the ensemble of CERES-ECHAM4.5 simulated maize yields through a simple un-weighted averaging approach. This is good first approach to follow, as other combination methods struggle to beat the high standard of the simple averaging method (Mason, 2008).

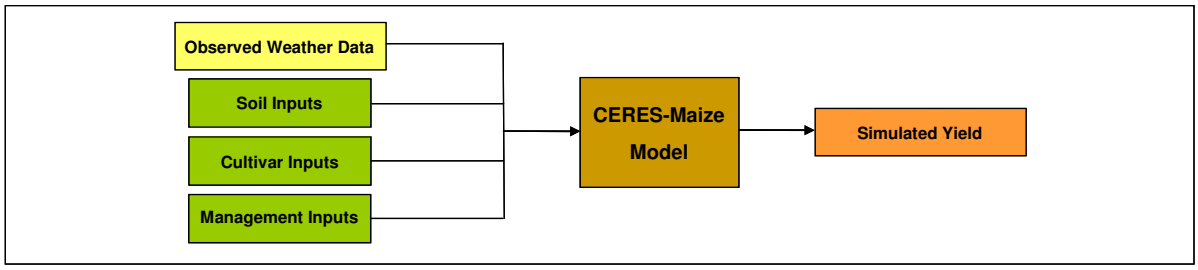

Figure 2.4: Flow diagram of the simulations performed with the CERES-Maize model.

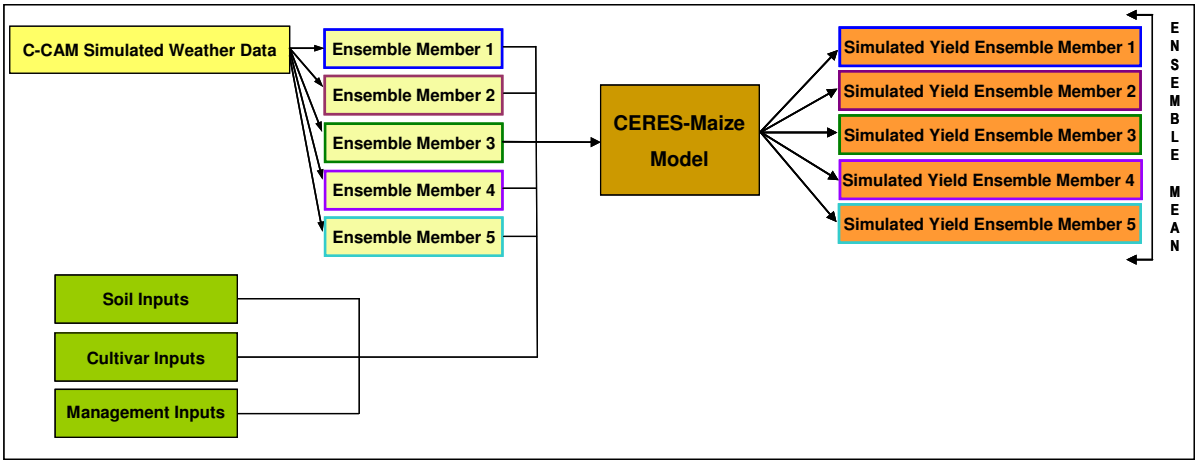

Figure 2.5: Flow diagram of the simulations performed with the CERES-Maize model when forced with each of the 5 ensemble members of the CCAM-simulated fields.

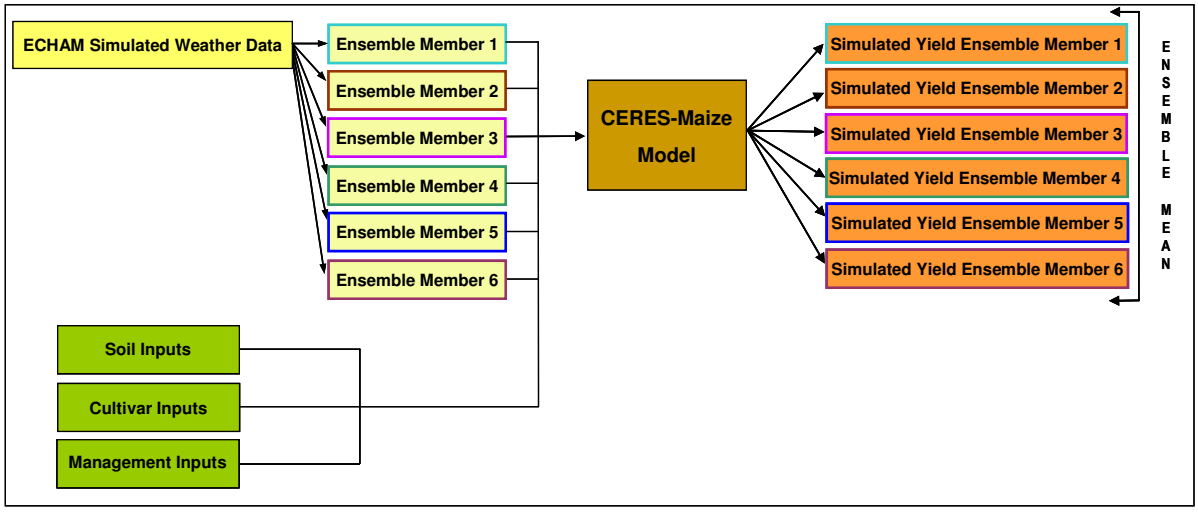

Figure 2.6: Flow diagram of the simulations performed with the CERES-Maize model when forced with each of the 6 ensemble members of the ECHAM4.5-simulated fields.

2.5 VERIFICATION OF THE SIMULATED MAIZE YIELDS

2.5.1 Verification Data

A number of maize yield datasets are available for South Africa. GrainSA is the custodian of a maize yield dataset that contains provincial maize yield figures for a period from 1980 to 2008. This dataset is compiled from information gathered from the silos in each province with regards to maize intake at the end of the season. Except for the fact that farmers often retain some maize for feed, the transportation process from the farm to the silo also results in some maize loss. Thus, a discrepancy may be evident between the maize yield figures in this dataset and the maize yield that was present on the land at the end of the season.

Another provincial maize yield dataset can be obtained from the Crop Estimate Committee of South Africa. This Committee meets once a month, starting at the beginning of the growing season, and uses a number of inputs to estimate the expected maize yield for that season. Thus, the yield estimates in this dataset are revised on a monthly basis until the final estimate is released. This data is available for a period from 1980 to 2008.

A subjective yield survey dataset that contains maize yield figures for 2000 random points over South Africa is also available from 2001. These figures are based on a questionnaire completed at each point before the start of the growing season. In these questionnaires the farmers state the plant density and area that is going to be planted and then also estimate the maize yield for that season.

For the objective yield survey dataset 200 out of the 2000 points in the subjective yield survey dataset are visited three times during the growing season. During these visitations the plant population is determined, number of cobs per plant is counted, the cobs are weighed and a maize yield is estimated accordingly. This dataset contains the three maize yield estimates for each of the 200 points for a period from 2001 to 2008.

The fifth maize yield dataset that is available for South Africa is the Co-operators yield dataset. This dataset contains maize yield figures per magisterial district for seasons 1980/1981 to 2007/2008. On a monthly basis from the start of the season, the Department of Agriculture sends questionnaires to the co-operating farmers in each magisterial district in which the expected maize yield for that season is stated. At the end of the season, after harvest, the co-operating farmers complete a final questionnaire in which the maize yield that was obtained is specified and this information then makes up the Co-operators yield dataset. The number of co-operating farmers in each magisterial district differs and

therefore these figures are not necessarily representative. The data in this dataset is confidential, but historical data were made available for research purposes.

In this study the CERES-Maize model simulates maize yield for 20 seasons in the period 1979 to 1999. In terms of verification the ideal would be to compare the CERES-Maize model results of maize yield to actual maize yield figures for the exact same period. From the five maize yield datasets described above, the Co-operators yield dataset is the only dataset suitable to be used as verification data in this study, because the data contained in the Co-operators yield dataset are for 19 out of the 20 seasons under investigation and on a higher spatial resolution (per magisterial district) than the other datasets that also have data for these seasons. Thus, CERES-Maize model output per magisterial district for the seasons 1980/1981 to 1998/1999 are verified against the Co-operators yield.

2.5.2 Verification Methods

Even though CERES-Maize model runs are performed for each of the 5 ensemble members of the CCAM-simulated fields and each of the 6 ensemble members of the ECHAM4.5-simulated fields, only the ensemble mean maize yield results of the CERES-CCAM runs, the ensemble mean maize yield results of the CERES-ECHAM4.5 runs and the ensemble mean maize yields results of the Multi-Model system are discussed. Although, the full ensemble of each simulation system is considered in the estimation of the skill when the maize yield simulations are expressed probabilistically.

From here onwards the maize yield results obtained from the CERES-Maize model runs performed with the observed weather data will be referred to as the CERES-Observed weather yield, the ensemble mean of the maize yield results obtained from the CERES-Maize model runs performed with the CCAM-simulated fields will be referred to as the CERES-CCAM ensemble mean yield, the ensemble mean of the maize yield results obtained from the CERES-Maize model runs performed with the ECHAM4.5-simulated fields will be referred to as the CERES-ECHAM4.5 ensemble mean yield and the combination between the CERES-CCAM and CERES-ECHAM4.5 maize yield results will be referred to as the Multi-Model ensemble mean yield.

Due to the limitation in the availability of soil data the number of soils that are selected for each magisterial district range between 1 and 3 (see Table 2.2). As the scenarios for which the CERES-Maize model is run for each magisterial district results from all the possible combinations of input data, 3 soils in combination with the other input data make up 27

scenarios, 2 soils make up 18 scenarios and 1 soil make up 9 scenarios. Thus, in the case of 3 selected soils, one scenario is duplicated for 3 soils and in the case of 2 selected soils, one scenario is duplicated for 2 soils. To simplify the verification process and obtain uniformity between the results obtained for the magisterial districts, the maize yield results of the duplicated scenarios are averaged across the different soils. Consequently each magisterial district ends up with maize yield results for only 9 scenarios (Table 2.8) and these results are presented in this study.

Scenario	Description
1	Short season maize planted on plant date 1
2	Short season maize planted on plant date 2
3	Short season maize planted on plant date 3
4	Medium season maize planted on plant date 1
5	Medium season maize planted on plant date 2
6	Medium season maize planted on plant date 3
7	Long season maize planted on plant date 1
8	Long season maize planted on plant date 2
9	Long season maize planted on plant date 3

Table 2.8: The 9 scenarios for which results are discussed.

2.5.2.1 Spatial Verification

The distribution of the simulated maize yields among magisterial districts is investigated through spatial verification. The aim is to determine whether the different simulation systems (CERES-Maize model integrations performed with observed weather data, performed with CCAM-simulated fields, performed with ECHAM4.5-simulated fields as well as the Multi-Model system) are able to capture the spatial distribution in maize yield across the study area.

To compare the simulated maize yields obtained from each of the different simulation systems to the actual maize yields, each of the simulated maize yield datasets and the actual maize yield dataset are normalized to a standard deviation of one and a mean of zero. Each dataset (simulated and actual) is normalized independently as follows:

$$NormalizedValue = \frac{x - \bar{x}}{\sigma} \tag{2.1}$$

where x represents the yield in t/ha of a specific magisterial district for a specific season, $\bar{x}$ represents the average yield in t/ha calculated for that specific season across all the magisterial districts in the study area and σ represents the standard deviation in t/ha calculated for that specific season across all the magisterial districts in the study area. Note that these normalized values are unit-less and consequently expressed as an index. The standard deviation is calculated as follows (Steyn *et al.*, 1998):

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N - 1}} \tag{2.2}$$

where x is the yield in t/ha of a specific magisterial district for a specific season, $\bar{x}$ the average yield in t/ha calculated for that specific season across all the magisterial districts in the study area and N the number of magisterial districts in the study area.

From each of these normalized datasets an average is calculated for each magisterial district across the 19 seasons, to obtain a single index value per district. Spatial maps of these average maize yield index values are displayed for the actual and simulated maize yields for each of the 9 scenarios in Table 2.8. The actual maize yield index map indicates which magisterial districts normally produce higher yields and which districts normally produce lower yields with respect to the entire study area. The maize yield index maps for each of the different simulation systems are visually compared to the actual maize yield index map in section 3.2.

2.5.2.2 Inter-Seasonal Variability Verification

This verification procedure examines the season-to-season variability in the simulated maize yields over the 19 seasons from 1980/81 to 1998/99. The aim is to determine whether each of the different maize yield simulation systems are able to capture the inter-seasonal variability in maize yield.

2.5.2.2.1 Subjective Validation

Once again the simulated maize yields obtained from each of the different simulation systems and the actual maize yields are normalized before any verification is done. The normalization is done similar to what is done for the spatial verification, although here the $\bar{x}$

in equation 2.1 and equation 2.2 represents the average yield in t/ha calculated for a specific magisterial district across the 19 seasons, σ in equation 2.1 represents the standard deviation in t/ha calculated for that specific magisterial district across the 19 seasons and N in equation 2.2 represents the number of seasons. It must be kept in mind that these normalized values are unit-less and therefore expressed as an index.

Figure 2.3 shows the magisterial districts that fall in each of the three production regions in the study area. The inter-seasonal verification is done per production region and therefore averages are calculated across the normalized values of the magisterial districts that fall in one production region to obtain one index value per season for each of the production regions. Thus, the normalized values of magisterial districts 1 to 29 are averaged for the Dry/Warm Western Region, 30 to 38 for the Temperate Eastern Region and 39 to 44 for the Wet/Cool Eastern Region. This is done for the simulated maize yield datasets of each of the different simulation systems as well as the actual maize yield dataset.

Time series graphs which depicts these average maize yield index values for the actual yield and simulated yields for each of the 9 scenarios in Table 2.8 are displayed for each of the 3 production regions. The actual maize yield index time series indicates which seasons had the highest maize yields and which seasons had the lowest maize yields with respect to the entire 20 year period under investigation. The maize yield index time series of the different simulation systems are visually compared to the actual maize yield index time series in section 3.3.

2.5.2.2.2 Objective Validation

Spearman's Rank Correlation Coefficient

Robust and resistant alternatives to the Pearson product-moment correlation are available. The first of these is known as the Spearman rank correlation. The normalized values previously calculated for each magisterial district are also used here. Spearman rank correlations are calculated between the actual maize yield index and each of the simulated maize yield indices of the different simulation systems for each magisterial district in the study area. Before calculating the correlations between the actual maize yield index and one of the simulated maize yield indices, the datasets are ranked, independently from each other, from the highest to the lowest value. The correlations are then calculated as follows (Steyn *et al.*, 1998):

$$R_S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n^3 - n} \tag{2.3}$$

where D_i is the difference between the ranks of the actual maize yield index for a specific season and the simulated maize yield index for that corresponding season and n is the number of seasons. R_S values range between -1 and 1, where -1 is a perfect negative correlation, 0 is no correlation and 1 is a perfect positive correlation (Steyn *et al.*, 1998).

These Spearman rank correlations which indicate the direction and strength of the relationship between the actual maize yield index and each of the simulated maize yield indices over the 19 seasons are displayed spatially and discussed in section 3.3.

Significance Testing

Here, it is tested if the number of magisterial districts with statistically significant local Spearman rank correlations between each of the simulated maize yield datasets and the actual maize yield dataset is significantly high (Wilks, 2006). A Monte Carlo test is performed to establish local significance at the 95% level. Since the magisterial districts are correlated spatially, re-randomization of each magisterial district's data are done by re-sampling random seasons, i.e., if the third season (1982/83) is the first season selected, then the first data vector of the re-randomized data will be the third season for all magisterial districts.

The re-randomized datasets for each magisterial district are then correlated with the actual maize yields, where after the re-randomization process is repeated for a large number of times (e.g., 1000 times). The subsequent Spearman rank correlations for each magisterial district are sorted and the 95th percentile identified. Thus, a set of Spearman rank correlations which represent the critical Spearman rank correlations at the 95% confidence level is available. From the unsorted Spearman rank correlations associated with each magisterial district, for each of the 1000 iterations, it is determined if the correlations are greater than or equal to its corresponding 95% confidence level. The number of times the 95% level is exceeded is counted. This is done for each magisterial district and for each of the 1000 iterations. Thus, in the end there is a 1000 counts ranging between and including

0 and the number of magisterial districts (44). This whole procedure is done for the data obtained from each of the different simulation systems.

If from the actual maize yield data it is established that 21 out of the 44 magisterial districts have significant local Spearman rank correlations, the number of the times the counts from the above explained procedure are greater than 21, are counted. This number then gives an indication of the probability of getting the actual maize yield results or better by chance. The magisterial districts with local significance at the 95% confidence level are described together with the Spearman rank correlations in section 3.3.

2.5.2.3 Probability Distributions

In this study the use of weather ensembles provides the opportunity to examine the predictability of maize yield probabilistically. The simulated maize yields, of those simulation systems that use weather ensembles, can be expressed probabilistically by calculating the percentage of the simulated maize yield ensemble members that fall in the below-normal, near-normal and above-normal categories. The aim is to determine the operational potential of this maize yield forecast system, as operational maize yield forecasts will most likely be expressed in terms of probabilities.

2.5.2.3.1 Subjective Validation

As the three production regions divide the study area into zones with similar climate and zones in which similar maize production methods are used, it would make sense to issue operational maize yield forecasts for each of these three production regions. Therefore, this analysis is performed for each of the three production regions. It is possible to calculate probabilities for the CERES-CCAM maize yield integrations (which has 5 ensemble members), for the CERES-ECHAM4.5 maize yield integrations (which has 6 ensemble members) and for the Multi-Model system (which is a combination between the CERES-CCAM and CERES-ECHAM4.5 simulations and consequently has 11 ensemble members), while the actual maize yields and the CERES-Observed weather yields are expressed deterministically.

Averages are calculated across the magisterial districts to obtain maize yield values per production region. This is done for the actual maize yields, the CERES-Observed weather maize yields as well as for each of the CERES-CCAM maize yield ensemble members and

each of the CERES-ECHAM4.5 maize yield ensemble members. These average maize yield values (actual and simulated) for each production region are then normalized to a standard deviation of one and a mean of zero, where $\bar{x}$ in equation 2.1 and equation 2.2 represents the average yield in t/ha calculated for a specific production region across the 19 seasons, σ in equation 2.1 represents the standard deviation in t/ha calculated for that specific production region across the 19 seasons and N in equation 2.2 represents the number of seasons. Once again, these normalized maize yield values are unit-less and therefore referred to as a maize yield index.

Before organising the maize yield index values into the below-normal, near-normal and above-normal categories, it is necessary to establish what 'near-normal' refers to. In this study, 'near-normal' refers to a range of values. The 19 maize yield index values of all 5 ensemble members of the CERES-CCAM simulations are combined (95 maize yield index values), sorted ascending and the third of the values in the middle of the arranged dataset are used as the 'near-normal' range. This exact procedure is followed to obtain the 'near-normal' range of the CERES-ECHAM4.5 simulated maize yields, but this time for 6 ensemble members (114 maize yield index values). From this the CERES-CCAM yield probabilities and the CERES-ECHAM4.5 yield probabilities are calculated for each of the 19 seasons. The yield probabilities of the Multi-Model system on the other hand are calculated by averaging the CERES-CCAM probabilities obtained for a specific season and the CERES-ECHAM4.5 probabilities for that same season. To obtain the 'near-normal' range of the actual maize yields, the 19 maize yield index values are sorted ascending and the values in the middle of the arranged dataset are used. Based on this, it is determined whether the actual maize yield of each of the 19 seasons was below-normal, near-normal or above-normal. This is repeated for the CERES-Observed weather yields. This whole procedure is done for each of the production regions.

Time series graphs of these probabilities are displayed for each of the three production regions and for 3 scenarios per region. Graphs were prepared for all 9 scenarios in Table 2.8, but as the changes in the probabilities from the one plant date to the next were insignificant, it was decided to average across the plant dates to obtain only 3 scenarios (Table 2.9). The CERES-CCAM yield probabilities, CERES-ECHAM4.5 yield probabilities and the Multi-Model yield probabilities are compared to the actual maize yields and CERES-Observed weather yields in section 3.4.

Scenario	Description
1	Short season maize
2	Medium season maize
3	Long season maize

Table 2.9: The resultant 3 scenarios obtained when averaging over the different plant dates.

2.5.2.3.2 Objective Validation

In probabilistic forecasts, the probability of occurrence of a certain event is estimated (Stefanova and Krishnamurti, 2001). Probabilistic forecasts are verified through the combined distribution of forecasts and observations (Stefanova and Krishnamurti, 2001). The relative operating characteristic (ROC) is one of the most commonly used methods to assess the skill of a forecasting system, by comparing the hit rate and the false-alarm rate (Mason, 1982). ROC scores are used to evaluate the value of probabilistic forecasts (Stanski *et al.*, 1989; Mason and Graham, 1999).

The probabilities displayed in the time series graphs of the subjective validation are also used here. ROC curves are constructed for each of the three equiprobable categories (below-normal, near-normal and above-normal) of the CERES-CCAM maize yield probabilities, CERES-ECHAM4.5 maize yield probabilities and the Multi-Model maize yield probabilities. The ROC curves are obtained by plotting the hit rates (HRs) and false alarm rates (FARs) against each other. HR and FAR are calculated as follows (Kharin and Zwiers, 2003):

$$HR(P_{cr}) = \frac{1}{\Pr(E=1)} \int_{\Omega_\beta} f_\beta(\beta) f(E=1 \mid \beta) d\beta \tag{2.4}$$

$$FAR(P_{cr}) = \frac{1}{\Pr(E=0)} \int_{\Omega_\beta} f_\beta(\beta) f(E=0 \mid \beta) d\beta \tag{2.5}$$

where P_{cr} is a critical threshold probability, $\Pr(E=1)$ denote a predictand when the event occurs and $\Pr(E=0)$ when the event does not occur and Ω_β is all the values of β (a potential predictable signal) for which $P > P_{cr}$. The HRs and FARs of a forecasting system with no skill are equal. The HR of a perfect forecast is 1 and the FAR is 0 (Kharin and Zwiers, 2003). The area under the ROC curve can be defined as the ROC score.

Forecasts with no skill have ROC scores of 0.5 and perfect forecasts have ROC scores of 1 (Kharin and Zwiers, 2003).

The ROC curves and ROC scores which give and indication of the skill of the different simulation systems and the value of the simulated maize yield probabilities are displayed and discussed in section 3.4 for each of the three production regions and each of the three scenarios in Table 2.9.

2.6 SUMMARY

The data, models and methods that are used to construct the different maize yield simulation systems have been described. The properties of the crop model and two GCMs used have been highlighted. The relevant input data required by the crop model and the set up of the experiments have been detailed. Finally, the data and methods used to verify the simulated maize yields have been discussed. In the next chapter the maize yield results obtained from each of the simulation systems for the main maize producing region of South Africa are discussed.