Project number		M103/15
Project title		Estimating methods for grain yield across South Africa using system/statistical analysis
Project manager		W Durand
Co-worker(s)	Internal	DL du Toit, MA Prinsloo
	External	DAFF, Farmers, Agribusiness, SIQ, Geoterra Image, University of KwaZulu-Natal
Project Status		Complete
Duration		01/04/2005 - 31/03/2012

Final abstract

More accurate and reliable crop estimate data is one of the prime requirements of all market players in the South African maize industry to minimize uncertainty in the market. One major source of information for all role players in the South African maize value chain is the official crop estimates released on a monthly basis by the Crop estimates Committee (CEC), with the Department of Agriculture, Forestry and Fisheries (DAFF) acting as secretariat. Most information regarding cropping area, yield and forecast production is gathered by input providers to the CEC in the form of expert opinion, postal or telephonic surveys, crop modelling or from actual field sampling. The main aim of this research project was to investigate facets and processes within the crop estimates system of the National Crop Statistics Consortium (NCSC) that acts in an advisory capacity to the CEC that could be improved to support more accurate yield forecasts. The facets within the crop estimates system that were investigated were: objective yield surveying (OYS), yield forecasts using crop models as input (IDSS-YES) and trend analysis. The accuracy of the South African OYS was increased by adapting the United States Department of Agriculture's OYS methodology to suit the South African cropping conditions through the development of the rapped assessment method, the introduction of the difference factor and the differentiation between dry-land and irrigation. Comparing the national production based on OYS with the national production of the final crop estimate for period 2002 to 2006 which, represents no research impact, with that of the period 2007 - 2011, which represents the period of gradual introduction of the research facets into the operational OYS system, indicated that the D-index increased by 3% from 0.96 to 0.99. The Mean Absolute Error (MAE) for the May production forecast using OYS decreased by 2% to 7.4% comparing the two periods (2002 - 2006 *vs.* 2007 - 2012). This is within the 8 % requested by the Crop estimates Liaison Committee (CELC). Facets within the Integrated Decision Support System for Yield Estimation in South Africa (IDSS-YES) were improved by adapting the newest version, introducing adaptations for South Africa and validating the model. Aspects of model inputs that were improved are the spatial representation based on fields that allows a weighted yield calculation, expansion to all provinces except the Western Cape, upgrading the soil, climate and management inputs using GIS, geographic weighted regressions and improved statistics. Like the OYS, the D-Index also increased (2 %) comparing the two study periods. The MAE for the May forecast using IDSS-YES decreased by 4.5 % to 8.2 % comparing the two study periods This may still drop further when the new climate database will be introduced to the system in the next season. Both yield and production components of the trend analysis were calculated and presented as an information aid to the CEC on a regular basis in May, June and July. The evidence of the impact of the research of all three facets researched can, however, be verified by the increased accuracy of the CEC comparing the two periods, where the MAE decreased by 4.2% to 6.4 %. This is within reach of the 5 % limit as set by CELC that is set as target for the final estimate and well within the 8 % set for the first four estimates.

Keywords

Crop modelling, CERES-maize, DSSAT4, Crop systems model, Trend analysis, Difference factor, Objective yield, Crop forecasting systems, Surveying

Introduction

Of the many crops grown in South Africa, maize (*Zea mays* L.) is the most important grain crop with a contribution of 76% to total of grains and legumes produced in South Africa. However, maize in South Africa is produced in a climate which is highly variable both in time and space, which accounts for the inter-annual variation of 36% found in production. In conjunction with this, maize is traded in a highly speculative open market, which is influenced by external factors such as the international grain price, stock levels and the exchange rate. Locally the market reacts nervously to information, confirmed or speculative.

In the run-up to a maize harvest, the production areas eventually realised, the productivity and last but not least the size of the crop that reaches the market are inseparably linked. Decision-making without one or more of these links can never be complete, is speculative and carries great risk. More accurate and reliable crop estimate data is one of the biggest requirements of all market players in the grain industry to minimize uncertainty in the market. The quality of the information is therefore decisive and quality again is the function of a large number of inputs that are continuously undergoing renewal.

The inception of this exclusively research-driven project concerning a crop yield forecasting system for South Africa, took place when comments were made by Mr JF de Villiers (NAMM), Mr WA Olivier (African Products - Wet Milling Industry), Mr WA Ferreira (GSA) and Dr JL Purchase (GSA) at the 2nd GCI Maize Planning Committee meeting in Potchefstroom on 20 October 2004. The proposal was also recommended by the chairman of the Crop Estimates Committee.

One major source of information for all role players in the South African maize value chain is the official crop estimates released on a monthly basis by the Crop estimates Committee (CEC), with the Department of Agriculture, Forestry and Fisheries (DAFF) acting as secretariat. The role of the CEC is to co-ordinate, control and release official, reliable, objective, accurate, timely, credible and unbiased forecasts of the areas planted to and production of selected summer grain and winter cereal crops on a national and provincial level. The objective of the crop forecasts is, firstly, to provide an indication of the expected area, secondly, an indication of the expected production and yield and lastly, a crop estimate at the end of the season in order to provide an indication of the total crop harvested. The Crop Estimates Liaison Committee (CELC), constituting of role players from industry, act as an advisory committee to the CEC. CELC, on behalf of the industry, has requested that the crop forecast not deviate by more than 8% in April to May and not more than 5% in June from the final crop (CELC meeting, 29 November 2007). Secondly, the forecasts should be based on scientific verifiable figures rather than farmer inputs (CELC, meeting 27 November 2008).

Most information regarding cropping area, yield and forecast production is gathered by input providers for the CEC in the form of expert opinion, postal or telephonic surveys, crop modelling or from actual field sampling. In postal and telephonic surveys farmers are questioned on their current experience and their opinion towards the outcome of the season. Depending on the market and the collective uncertainty levels these results are often biased. Actual field samples, as collected during an objective yield survey and modelling initiatives, are two methods that allow decision makers access to objective, producer independent, unbiased yield estimates during a production season.

The main aim of this research project was to investigate facets and processes within the crop estimates system of the National Crop Statistics Consortium (NCSC) that acts in an advisory capacity to the CEC that could be improved to support more accurate yield forecasts. In-time accurate yield estimates for use by the CEC were therefore not the focus of this study. The facets within the crop estimates system that were investigated were:

- objective yield surveying,
- yield forecasts using crop models as input and
- trend analysis.

Recommendations towards improving the whole methodology of the crop estimates system were suggested, based on on-going experience and literature research that was conveyed to members of the CEC, CELC, verbally through workshops and meetings attended. In the following section each of the three facets is individually discussed, conclusions drawn and recommendations given. At the end of the report an overview of the success of the project is given.

Objective yield surveying

Introduction

With the deregulation of the South African maize market in May 1997 new methodologies had to be developed to estimate crop yield and production, as the single channel marketing system that had delivered these figures was no longer available. To overcome this shortcoming the National Agricultural Statistics Service (NASS) of the United States (US) through the guidance of Ms T. Holland aided the South African Department of Agricultural Statistics in the implementation of a objective yield surveying system (OYS) based on NASS technology (anon, 2006). In such a survey actual field measurements are collected of the crop with the aim of deriving a yield.

In the US, NASS conducts several surveys to obtain the basic data needed to fulfil its obligation of estimating the production of most crops grown. These surveys, that are a mix of grower interviews and objective field visits, employing sophisticated survey sampling designs and application of probability sampling and statistical methodology such as regression models have been undertaken since 1961 (Waldhaus *et al*.). The challenge that however, confronted South African researchers that were targeted to conduct an objective yield survey in South Africa, was to implement the US survey methodology under South African conditions without having experience in such an operation within a very limited time frame with a high accuracy.

The first OYS for the 2001/2002 summer crop production year was conducted for six crops (maize, sorghum, sunflower, soya beans, ground nuts and dry beans) in five provinces (Free State, North West, Mpumalanga, Gauteng and Limpopo) on a monthly basis for three months (April, May and June). Already with the first report from the input providers of the NCSC there was some controversy about the high figures presented as these were “raw” figures with no bias or so-called reduction factors implemented (Du Toit, Smalberger and Pretorius, 2002). This is in contradiction to parts of the NASS system, where regression models that were developed over time were utilised. This could be attributed to lack of knowledge and misperception in the use and interpretation of yield figures derived from a fully functional OYS by both the CEC members as well as the NCSC personnel. In the US system, trends, analysis and interpretation thereof forms an integral part in the survey analysis, using historic data to identify trends between season and more importantly within season. The relative changes are more important than absolute figures. The South African Crop estimates system is, however, based on a convergence of evidence method and does not rely on trends. Thus, absolute figures are more important.

Researchers of the ARC-GCI were consequently challenged to adopt the OYS methodology that was successfully applied in the US and adapt it to South African conditions and expectations within budget constraints. In this respect the following inadequacies were identified:

- Allocation of samples within budget constraints
- Number of months to take samples
- Improvement to sampling methodology applicable to South African conditions
- Weighed kernel weight versus estimated kernel weight
- Number of samples per field
- In field sampling location
- Number of ears to be sampled
- Harvest difference
- Weighted average yields
- Outlier identification
- Mean versus median versus trimmed mean

Material and methods

In South Africa the OYS locations to be sampled are selected from the Subjective Area Frame Survey System (SAFSS) where crops of interest were found. Objective yield samples are selected within each province with a probability proportional to size, making the samples self-weighting. This allows fields, which are large and thus have a large expansion factor to be selected for more than one sample. Trained enumerators visited the selected fields and collect background information on planting dates, cultivars, row with, coordinates etc. followed up by measurements of plants and cobs within randomly selected sites within a field following a system of set procedures.

Results

In this section the the different inadequacies that were investigated to improve the OYS are discussed and solutions presented. Table 1 indicates the timeline of implementation of methods.

Table 1 Operational Objective yield survey and methods implemented for improvement.

Year	Crop	Provinces	Locations (Fields)	Months	Method	Samples per field	Factor	Stratifica-tion	In-field method	Outliers
2001/2002	6	5	250	April May June	Traditional	5				
2003/2004	2	3	360	April May June	Traditional	2				
2004/2005	2	4	325	April May June	Traditional	2				
2005/2006	1	3	500	May	Traditional	2	5%			
2006/2007	1	3	600	May - MP April - FS & NW	Rapid	2	5%			
2007/2008	1	3	700	1 (May - April)	Rapid	5	regression	For dry-land For irrigated Weighted average	4 methods	
2008/2009	1	3	830	May - MP April - FS & NW	Rapid	5	regression	For dry-land For irrigated Weighted average	4 methods	
2009/2010	1	3	830	May - MP April - FS & NW	Rapid	5	regression	For dry-land For irrigated Weighted average	4 methods	Outliers removed
2010/2011	1	3	830	May - MP April - FS & NW	Rapid	5	regression	For dry-land For irrigated Weighted average	4 methods	Outliers removed
2011/2012	1	3	700	May - MP April - FS & NW	Rapid	5	regression	For dry-land For irrigated Weighted average	4 methods	Outliers removed

Allocation of samples within budget constraints

The budget determines the number of fields that can be visited. The challenge was to identify how to allocate the fields that have to be sampled to the different provinces in a proportional manner. It was decided that the number of fields that could be sampled should be based on the proportion of the area planted either to white or yellow maize in each province. Either the February CEC or the PICES figure can be used with the latter being the preferred. After the first implementation of this method using a stratification for dry-land and irrigated fields, it was observed that the small sample prescribed by the method, for example North West Province for yellow maize, was less than two samples. This posed a problem when these samples were lost due to refusal or other obstructions. It was proposed that no less than ten samples for irrigation should be allocated and that the difference required would be deducted from the dry-land sample allocation. As to the number of provinces it was proposed that only three provinces would be sampled, i.e. Mpumalanga, Free State and North West as they contribute more than 85% to the total national maize production.

Number of months to take samples

As already stated the NASS system is based on trends and sampling, continuously and consecutively over five months. For the sampling seasons 2001/02, 2003/04 and 2004/05 this method was followed by sampling in April, May and June. However, the fluctuation in calculated yields caused confusion within the CEC, as yields were at the beginning of the season mostly calculated from on estimated kernel weight. It was therefore proposed to sample once a season, when the physiological maturity of the maize allows actual measurements of kernel weight. This however, should be undertaken at the earliest stage convenient. After the 2005/06 survey it was, however, established that May was too late for surveying in Mpumalanga as many farmers had already harvested. It was proposed to survey Mpumalanga in April and the Fee State and North West Province in May. This once-off sampling regime, allows more fields to be sampled. The number of fields that were surveyed has subsequently increased over the years from 250 in the beginning to 830 in the 2010/11 season. The fewer fields sampled in the 2011/12 season (700) was due to a smaller budget (Table 1).

Improvement to sampling methodology to be applicable to South African conditions

The NASS field sampling methodology calls for measurements to be collected with three replications at each of the two randomly selected site within a field. At each replicate, row width, number of plants per 10 m and number of ears per 10 m are captured. The first three ears per row are selected and kernel rows, number of kernels per row is counted and when the maize has reached physiological maturity kernels are separated from the cob and weighed. Therefore, one location sampled had six field measurements and eighteen ear measurements (Figure 1).

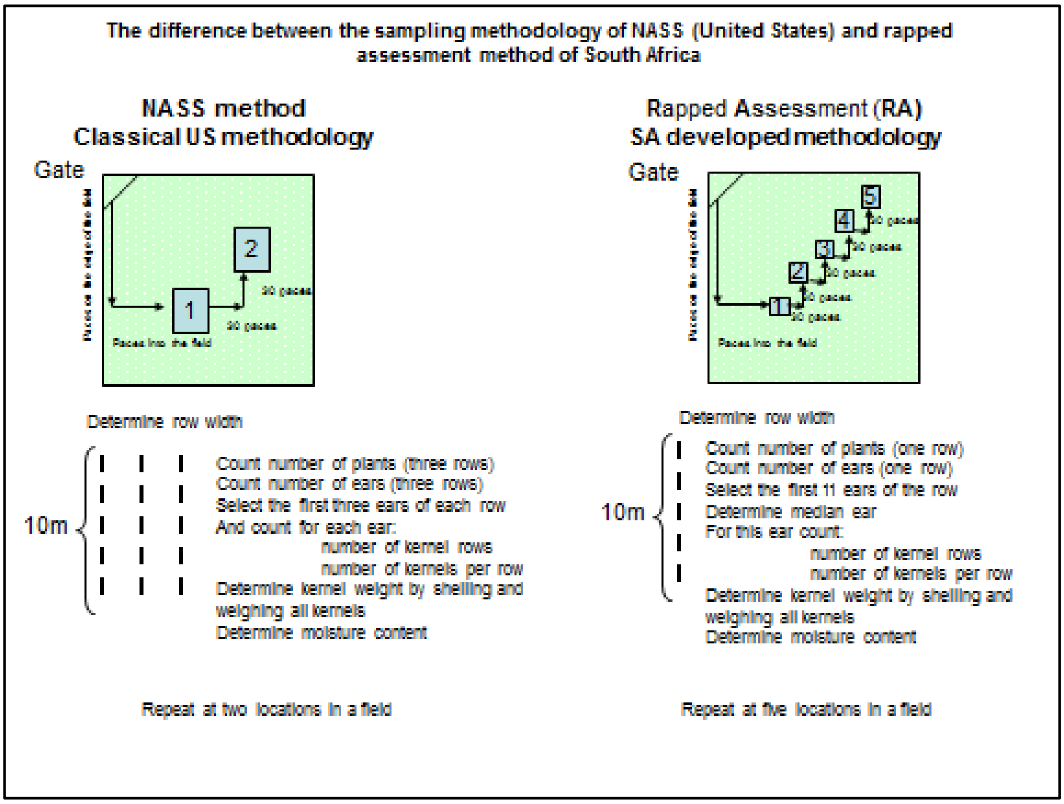


Figure 1 The difference between the sampling methodology of NASS (United States) and the Rapped Assessment (RA) method developed for South Africa.

Rapped assessment (RA), as a tool for crop estimations, has been used by crop insurers and agri-businesses for a long time. It was the first time adapted for crop yield estimations in a study by AS du Toit in cooperation with Sentraoes in 2000. Since then this methodology has been refined and adapted to suit the needs for objective crop yield estimation. Its procedures are very simple and easy to apply, making this ideal to estimate more points within a limited budget.

The main difference between RA and the NASS methodology is the number of replications sampled within a field, the method of selecting ears to be measured and the determination of kernel weight. With RA five replications per field are sampled but for each replication only one 10m row, whereas with the NASS method only two replications per field are sampled with three 10 m rows. For the RA method enumerators pick the first eleven cobs per row and determine the median ear. The RA methodology considerably cuts back on man-hours required to collect data and time can be spent to collect more spatially diverse data. The NASS method requires the sampling of the first 3 ears per row, thus 9 ears per replication.

In addition to sampling according to the methodology prescribed by the NASS in May 2006, 294 fields were sampled using both RA methodology and the traditional NASS method. The similarity between the two measurement techniques was calculated using the index of agreement (D-index) (Willmott, 1982). The higher the index (the nearer to 1) the better two models agree with each other. The root mean square error systematic (RMSEs) should approach 0 for good model agreement whilst the root mean square error unsystematic should approach the RMSE for good model agreement. The advantage of RMSEs is that it indicates the bias (deviation of the actual slope value from the 1:1 line) in a particular model, compared with the random variation (RMSEu) that may occur (Savage, 1993).

Results (Table 2) in the comparison between NASS sampling and RS sampling indicate that most of the error is unsystematic and that there is a high index of agreement between the two sampling techniques. The D-Index for the Free State and North West Province was above 0.90 whereas for Mpumalanga it was marginally lower at 0.88. The 1:1 graphs which depict the NASS (Observed) measurements with the RA (Simulated) measurements are a good indication of model agreement (Figure 2).

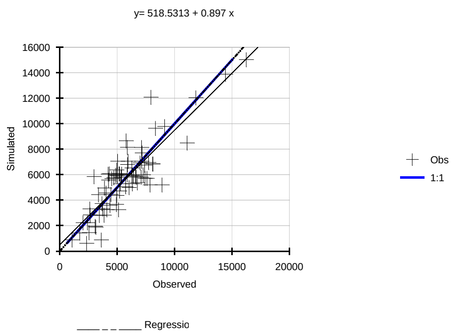
Table 2 Summary of statistics when comparing National Agricultural Statistics Service (NASS) sampling results to RA sampling results

	White Maize (kg/ha)			**Yellow Maize (kg/ha)**		
	Free State	**Mpumalanga**	**North West**	**Free State**	**Mpumalanga**	**North West**
MAE	1087	1467	1094	1081	1419	851
RMSE	1368	1844	1588	1359	1856	1089
RMSEs	276	803	364	241	432	529
RMSEu	1340	1660	1545	1337	1805	952
D-Index	0.93	0.88	0.91	0.96	0.88	0.94

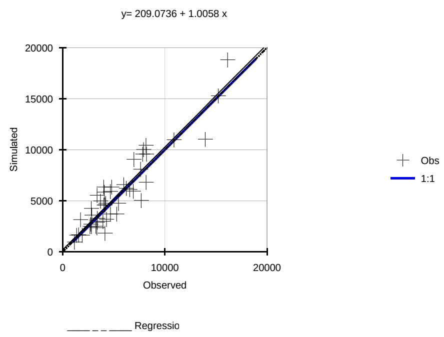
MAE Mean absolute error
RMSE Root mean square error
RMSEs Root mean square error systematic
RMSEu Root mean square error unsystematic

Free State

White maize

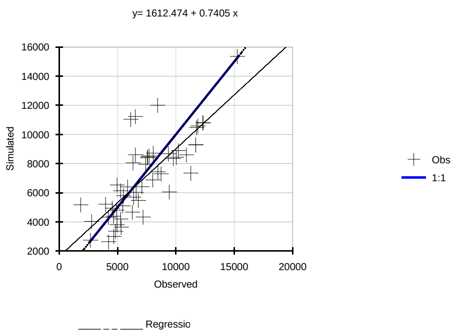


Yellow maize

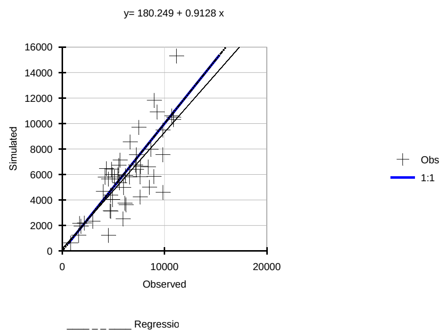


Mpumalanga

White maize

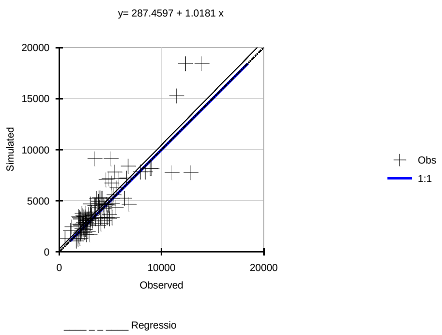


Yellow maize



North West

White maize



Yellow maize

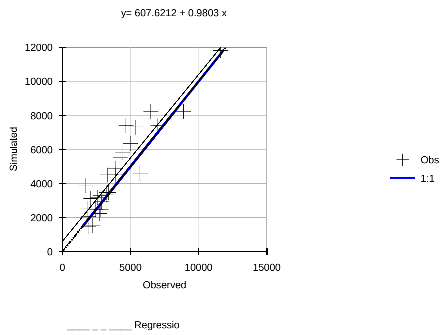


Figure 2 Validation graphs depicting the National Agricultural Statistics Service (NASS) observed versus the results of the Rapped assessment (RA) method simulated.

Results using both methods were compared (Table 3) and the conclusion was made that there was very little difference between the two sampling methods. However, the coefficient of variation (CV) for both techniques was high.

Table 3 Yield (kg/ha) for the three provinces as estimated using Rapped assessment (RA) and National Agricultural Statistics Service (NASS) methodologies

		Free State (kg/ha)	**CV**	**Mpumalanga (kg/ha)**	**CV**	**North West (kg/ha)**	**CV**
White maize	RA	5637	47	7047	38	4454	67
	NASS	5706	45	7339	39	4092	61
Yellow maize	RA	5661	64	5942	50	4728	52
	NASS	5420	62	6312	40	4154	55

From the research recommendations made to the CEC were that the RA methodology should be adopted as the preferred method for the OYS. The following were given as reasons:

- It is similar in accuracy as the traditional NASS method.
- It covers a larger in-field variation (5 samples).
- It is simple and easy to use by non-scientists who are employed as enumerators in South Africa.
- The time spent sampling a field is much shorter allowing for more fields to be sampled within a restricted time-span and allowing a larger spatial variation to be sampled.

The adoption of the RA method was however met with scepticism by the CEC and CELC but based on findings from a working group meeting 4 October 2007, it was decided to keep the RA method for OYS.

It must however, be remembered that each technique, NASS and RA, has its own limitations due to its own random systematic sampling error e.g. in moisture meter equipment and errors due to failure in including parameters in the sample (unsystematic) like to human nature.

Since its inception the RA methodology has been widely accepted and various workshops have been held nationally to demonstrate this methodology for both commercial and subsistence farming systems.

Weighed kernel weight versus estimated kernel weight

Depending on the timing of the OYS the yield can be calculated either on counts of kernels and the estimation of average kernel mass (equation 1) or it can be determined by physically removing the kernels from the cob and weighing them (equation 2).

Equation 1

$$\text{Yield (kg/ha)} = \text{Ears per 10m} * \left(\frac{10\,000}{10 * \text{Row width}}\right) * \left(\frac{\text{Kernels per row} * \text{Kernel rows} * \text{Kernel weight}}{1\,000}\right)$$

Equation 2

$$\text{Yield (kg/ha)} = \text{Ears per 10m} * \left(\frac{10\,000}{10 * \text{Row width}}\right) * \left(\frac{\text{Ear mass}}{1\,000}\right)$$

Both the NASS method and the RA technique in South Africa have one inherent problem early in the sampling season and that is the lack of a well established data-base on kernel mass. Early in the season kernel mass is not yet measurable and has to be derived subjectively.

Some preliminary data analysis was conducted to determine the effect of using a generic kernel weight on the estimation accuracy of RA method using data from the 2005/06 season. Examining this, an average kernel mass of 0.25g was taken as representing a worst-case scenario, 0.32g for an average case scenario and 0.38g for a best-case scenario. These were compared to the results obtained by enumerators who judged kernel mass according to example pictures for kernels with different masses. For each of these the index of agreement was calculated as presented in Table 4.

Table 4 Index of agreement between RA using judged kernel mass and kernel mass associated with worst, average and best case scenarios

	White maize			**Yellow maize**		
	Worst	**Average**	**Best**	**Worst**	**Average**	**Best**
Free State	0.83	0.95	0.97	0.73	0.84	0.88
Mpumalanga	0.84	0.94	0.89	0.87	0.95	0.91
North West	0.88	0.96	0.95	0.85	0.96	0.97

Indications are that for the current season, using a kernel mass of 0.38g the best-case scenario would have yielded the best results in the Free State (Table 5) whilst in Mpumalanga an average scenario would have been adequate. In the Northwest province either the average or best would have yielded good results. One indication could be that in the Free State better than average yield is expected whilst in Mpumalanga yields are average. With the knowledge that less maize was planted in the Free State this year and most of the maize was planted on better soils it could explain this observation.

Table 5 Average yield (kg/ha) as calculated using RA which uses judged kernel mass and yield for the three scenarios (worst, average and best cases) where kernel mass is 0.25, 0.32 and 0.38g respectively

	White maize (kg/ha)				**Yellow maize (kg/ha)**			
	RA	**Worst**	**Average**	**Best**	**RA**	**Worst**	**Average**	**Best**
Free State	5744	3874	4959	5889	6483	4336	5551	6591
Mpumalanga	6441	5075	6496	7714	5547	4457	5705	6774
North West	4391	3285	4205	4993	5109	3695	4729	5616

Unfortunately, as in this case, the information is not available early in the season. However, using sound judgment together with the knowledge of the people within an area it could be concluded which one of the scenarios best describes a relevant season.

In conclusion it was found that, conducting an OYS early in the season, using estimated kernel weight and the resulting variation in yield together with the inability of the CEC to incorporate trends rather than absolute figures from a OYS system does not justify research. OYS estimates based on measured kernels is preferable for the current South African system.

Number of samples per field

Due to South Africa's large variation in both soil and climate properties, even within-field, the coefficient of variation (CV) is over 30% for yield. The question was asked: what is the optimum number of samples to be taken in an OYS that would adequately cover this variation? During the 2004/05, 2005/06 and 2006/07 fields were sampled taking two, five and seven random samples per field. A regression analyses indicated five samples to be adequate in capturing most of the in-field variation (Table 6) (Previous report). Increasing the number of samples per field decreases the allowable error but increases the CV of the sample. It was proposed that in the conducting of the OYS five random samples should be taken per field. This has been implemented in OYS since the 2007/08 season.

Table 6 R^2 of the regression analysis between the number of samples per field and farmer yield

	2004/05	2005/06	2006/07
2 samples	0.6687	0.6351	0.7848
5 samples	0.8033	0.8183	0.8910
7 samples	0.8143	0.8258	0.8670

In field sampling location

With optimum number of samples per field determined, a further need arose to identify the minimum number of locations in conjunction with their spatial location in a field that best reproduces actual farmer yields of a field. As from the 2007/08 season fields located in the North West Province, Free State and Mpumalanga were selected and different sampling strategies were evaluated (Figure 3 a & b). The differences were based on number of samples and spatial stratification (Table 7).

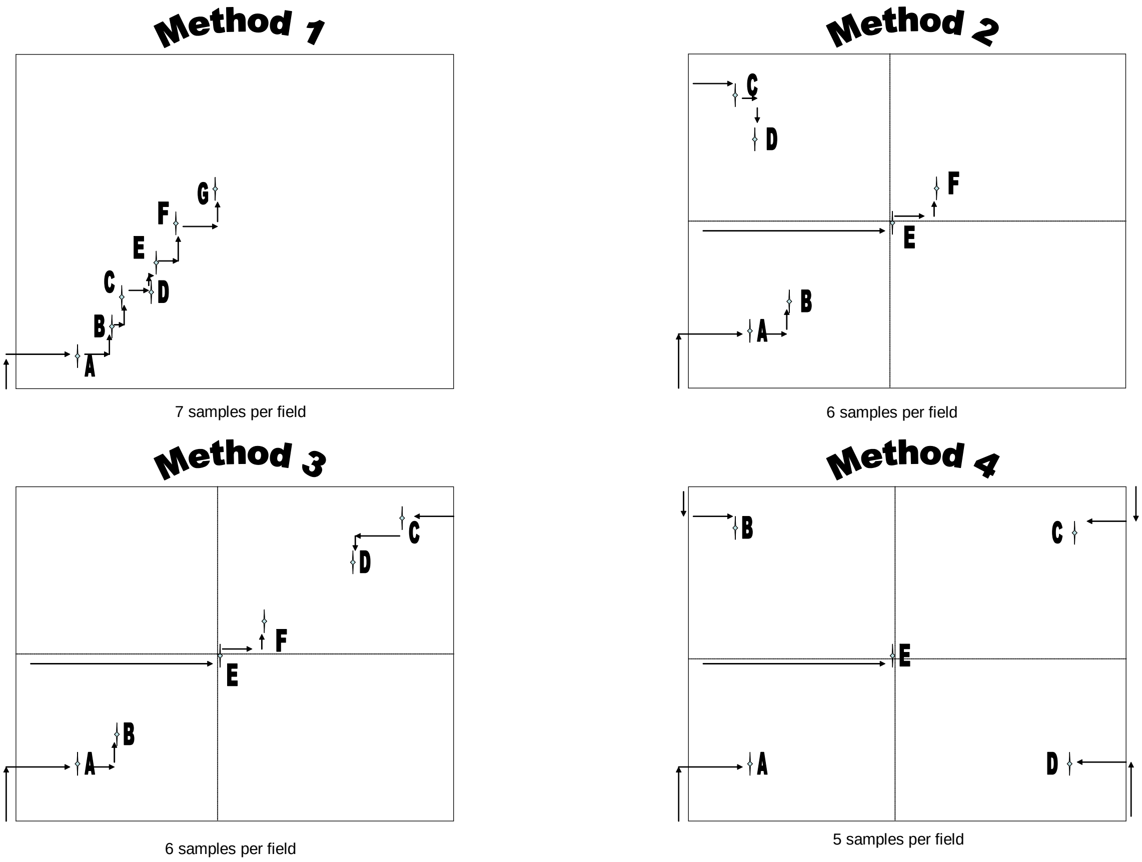


Figure 3a Overview of the four sampling methodologies 2007/2008 season.

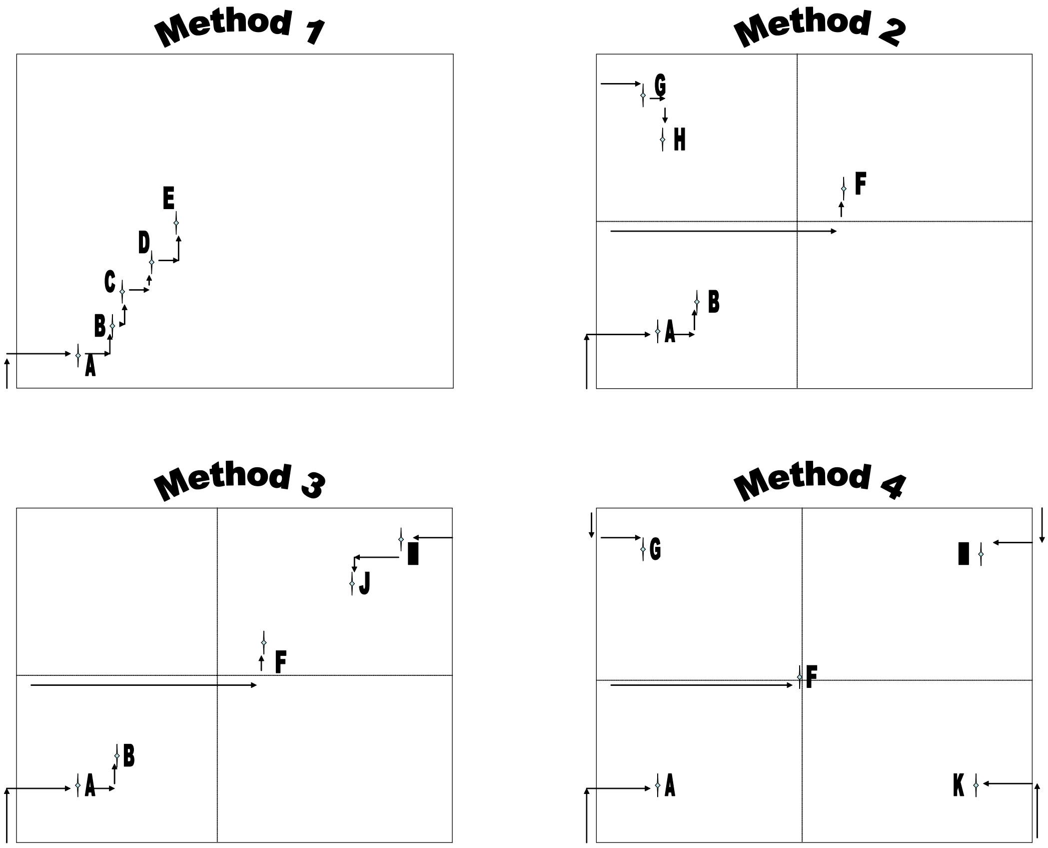


Figure 3b Overview of the four sampling methodologies from the 2008/2009 season onwards.

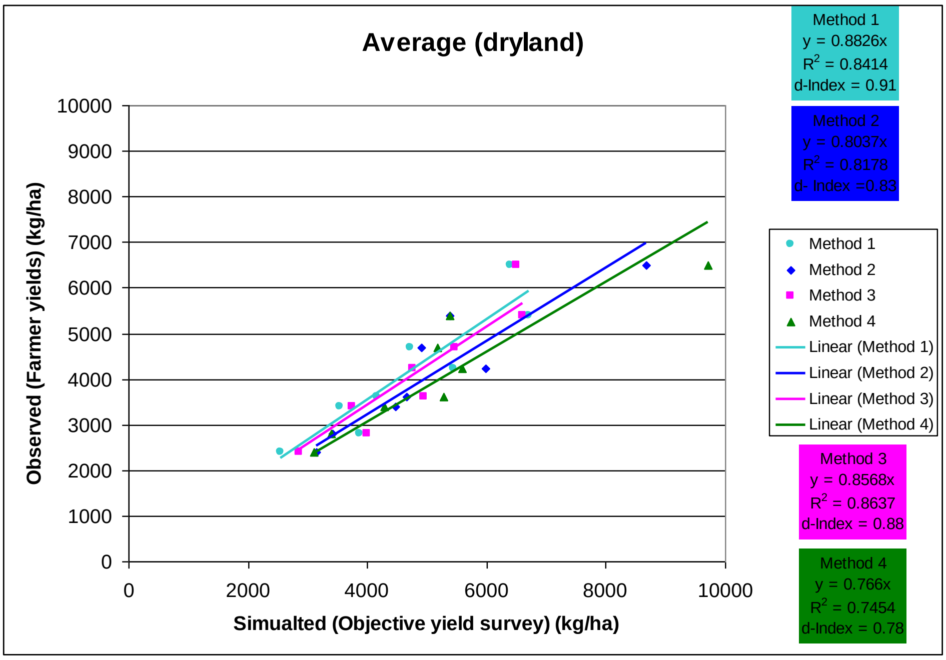
Table 7 Description of the 4 sampling methods evaluated

Method	**2007/08**	**2008/09 to 2011/12**
Method 1	Seven samples spread randomly across field.	Five samples spread randomly across field.
Method 2	Six samples, two samples in each corner of a field and two samples in the middle.	Five samples, two samples in each corner of a field and one sample in the middle
Method 3	Six samples, two samples in one corner of the field, two samples in the middle of the field and two near the adjacent corner of the field.	Five samples, two samples in one corner of the field, one sample in the middle of the field and two near the adjacent corner of the field.
Method 4	Five samples, one in each corner of the field and one in the middle.	Five samples, one in each corner of the field and one in the middle.

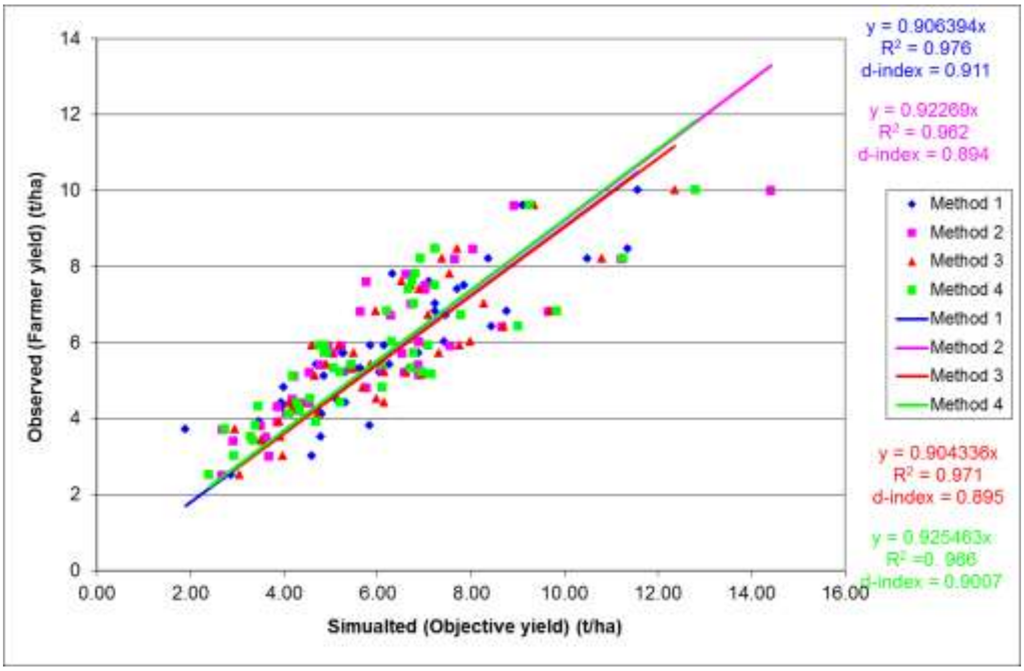
For the 2007/08 season, initial data analysis that included the irrigated sampling locations in the calculation, gave no conclusive results as to which method was more accurate. Using only the dry-land locations' data, regression trend lines through zero were fitted and the index of agreement was calculated for each season (Figure 4). R^2 is an indication of the variation of the samples from the regression line whilst the d-Index is an indication of the index of agreement between the sampled (objective yield yields) and observed (farmer yield) data.

Figure 4 Regression and index of agreement of each of the four in-field sampling methods evaluated.

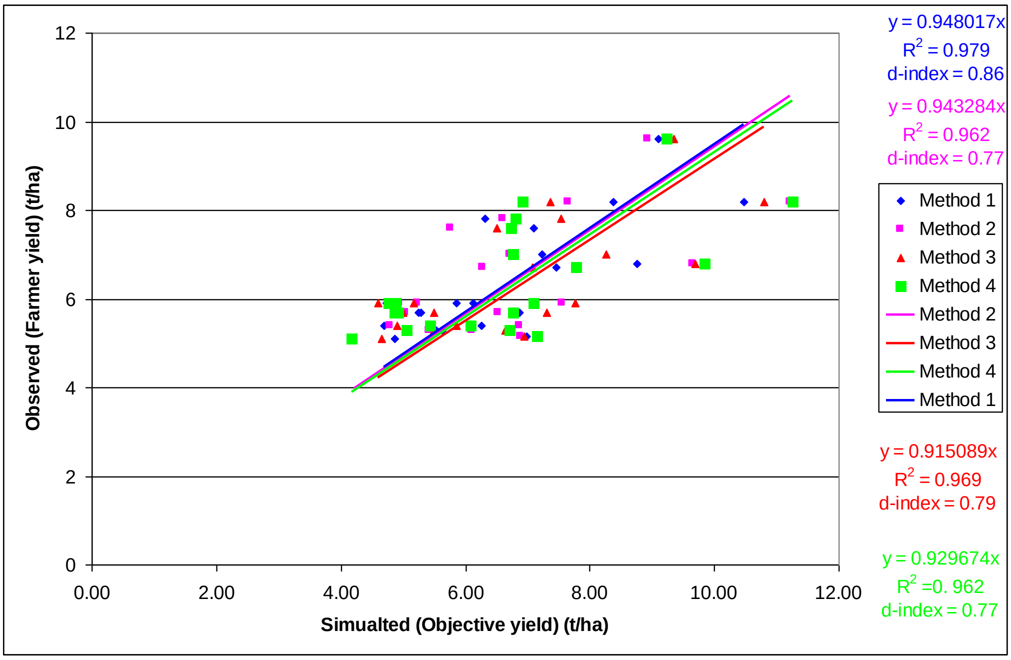
2007/08



2008/09



2009/10



2010/11

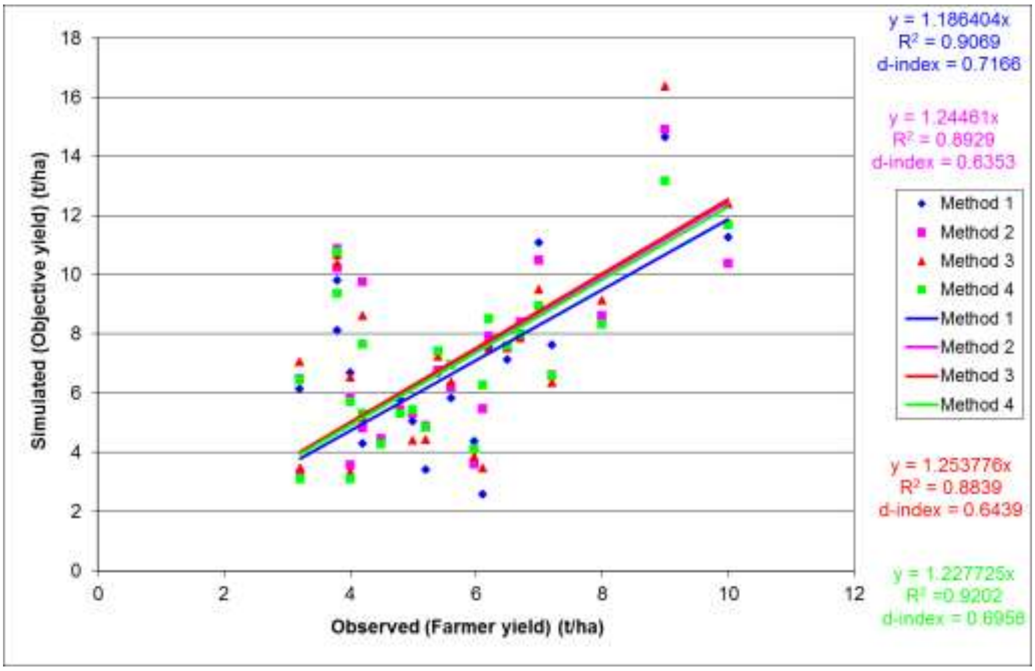


Table 8 Index of agreement between the sampling method and the actual farmer yield.

	2007/08	**2008/09**	**2009/10**	**2010/11**
Number of field sampled	**8**	**20**	**43**	**24**
Method 1	0.91	0.86	0.91	0.72
Method 2	0.83	0.77	0.89	0.64
Method 3	0.88	0.79	0.90	0.64
Method 4	0.78	0.77	0.90	0.70

Results indicate that taking up to five random samples per field, was marginally better than any of the other stratified sampling methods in all four seasons, best capturing in-field variation and explaining between 70 and 90% of the variation found in a field (Table 8). It is proposed that taking five random samples in a field is the best suited methodology to follow in an operational OYS in South Africa. This has been the preferred method for sampling from the beginning as it also is the simplest method to implement practically and takes the least time and effort to complete.

Number of ears to be sampled

In the NASS methodology the first three ears within each row are selected for measurements resulting in 18 ears sampled per field. This is a very timeous and cumbersome effort. It was proposed that taking many samples would not necessarily increase yield accuracy. It was therefore decided that the median would be more representative than the mean. The following ear selection methods were suggested: At each sampling location in the RA method 11 ears occurring in a 10m row should be sampled. These should be sorted and set out according to their length from short to long (Figure 5). From this configuration, the middle ear (median) should be selected. This ear should be shelled and the kernels weighed using a scale with an accuracy of two decimals to the gram. The moisture content of the kernels should be determined using a moisture meter. With the adoption of the RA methodology this is the preferred method and used in the operational OYS in South Africa since 2007/08 season.

Figure 5 Methodology for ear selection

Harvest difference

As already stated, in the beginning of objective yield surveying in 2001/2002 there was a misconception about the magnitude of the data of the OYS and a dispute regarding the application of a reduction factor started. The “raw” yield data as collected by an OYS will always be much higher than that of an official final estimate. NASS conducted corn validation studies from 1954 through 1983. A majority of these studies, conducted to examine relationships between objective survey estimates and actual yield of corn, showed an unexplained difference of between 2.0 and 4.8 percent. However, differences between the objective survey estimates and the official final estimated yields for a region of 10 major States generally were between 6 and 12 percent. The principal recommendation from these studies was that the official estimate be adopted from final average yield for the 10 State region, adjusted for non-sampling errors, as its final estimated yield of corn grain for that region. These adjustments by NASS are introduced in the form of models that are mathematical equations used to represent the relationship between two or more variables. NASS has developed five different models as yield indicators which are: Field level forecast that excludes harvest loss, the farmer reported field yield for a specific sample field, the field level indication regressed to final official estimate of yield, the farmer reported yield indication regressed to final official estimate of yield and the state average counts regressed to final official estimate of yield.

Semantics have become a problem in describing these yield indicator models in South Africa. The terms yield reduction factor, realization factor, harvest loss, yield loss, harvest difference, difference factor and bias adjustment have all been used to refer to one or other model used to adapt the "raw" yield figure to provincial level.

Reduction factor and realization factor

Du Toit et al. (2002) developed the first factor, termed by them reduction factor, from a regression at provincial level to CERES-maize prediction model. He suggested that a reduction factor of 20% should be used for maize. Using the national cultivar yield data as a filter to increase the accuracy of crop yield forecasting Durand and Du Toit (1999) developed a so-called realization factor of 0.76. Yield estimations of the 2003/2004 season were recalculated using this factor, however after the season it was found to be too high, resulting in an under prediction when the adjusted yield were compared to the reconciled yields at the end of the season.

Harvest difference or difference factor

One of the main objectives of this study was to determine the difference between OYS estimated yield for a field and that actually harvested by a farmer. For this study during the past seven seasons (2004/2005 to 2010/2011) data was collected from 207 fields where both the actual farmer yield and the yield estimated, using the RA method, were available for a specific field. Four models were developed: white maize dry-land, white maize irrigated, yellow maize dry-land and yellow maize irrigated. (Figure 6). From 2007/08 onward the regression model developed was used to adapt the "raw" yield data before being presented to the CEC. The regression model was updated annually as new information became available.

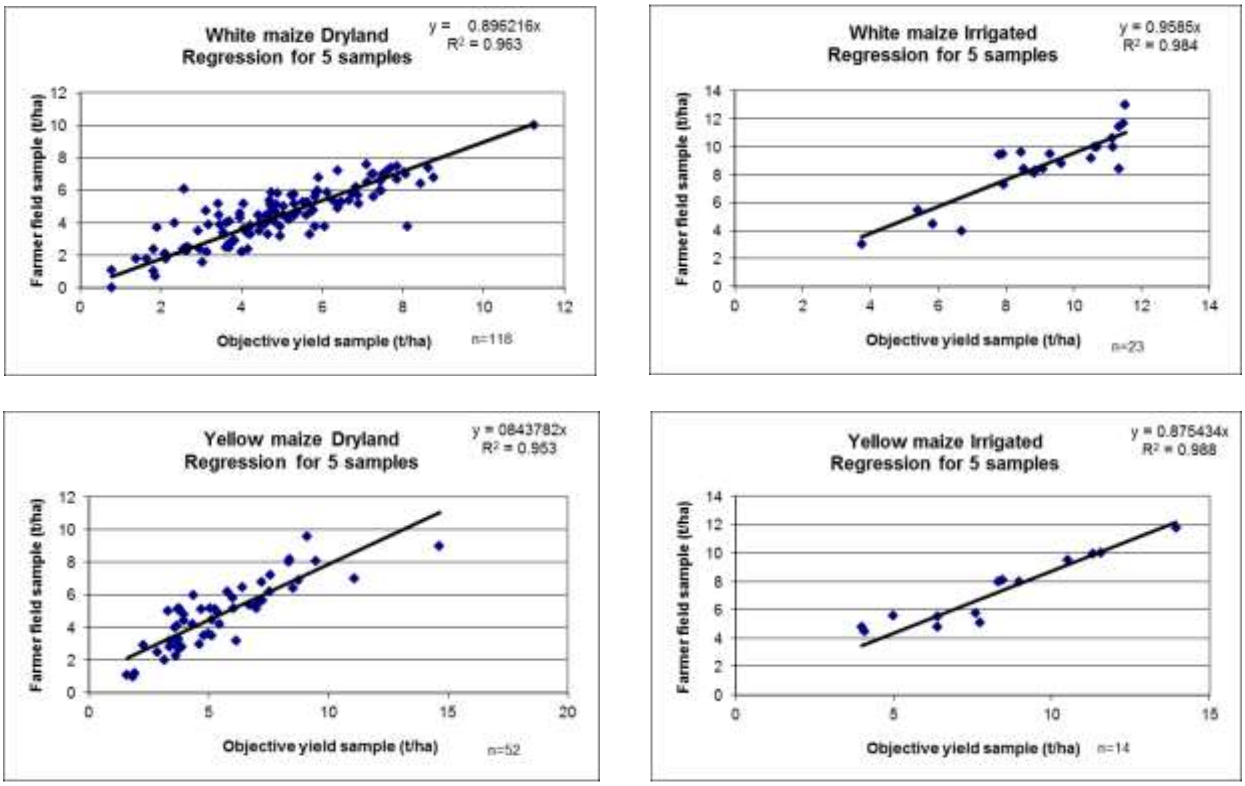


Figure 6 Regression equations for adaptation to farmer yields (Harvest difference model)

Harvest loss

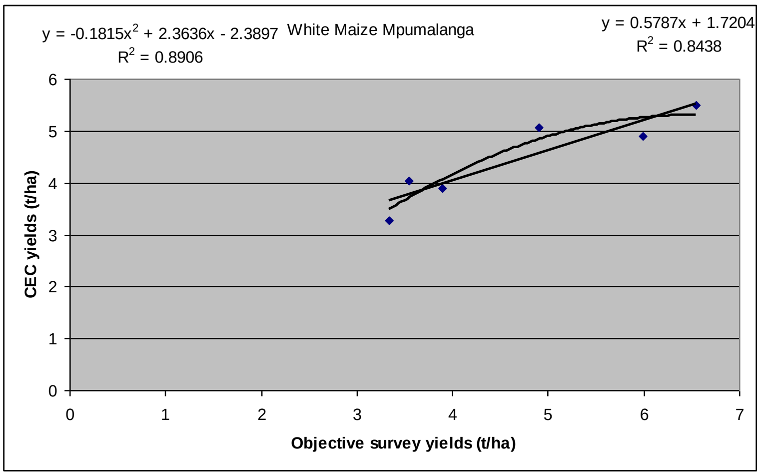
In the US data collected for harvest loss consists of gleaning samples of fruit left in field after harvest. This factor is subtracted from the yield. Except for the last sample a 5-year average is used. A harvest loss figure for South Africa will be difficult to determine due to the large variation in management practices and harvesting techniques adopted by farmers and may also not be a useful indicator for use over seasons, as farmers pick up cobs after harvest depending on the maize price and other external factors. The advice is that this factor can be ignored in South Africa as it is inherently part of the harvest difference or difference factor.

Difference factor or bias adjustment that describes the provincial average counts regressed to final CEC yield.

As mentioned earlier NASS has established that there is a difference of between 6 and 12 % between survey estimates and the final official yield estimates for a region. Table 9 presents the percentage difference between the objective survey estimates and the official final estimated yields for the three sampling provinces and for the seasons 2002 to 2008. The negative figures indicate that objective yield survey yields are lower than those estimated by the CEC. The individual regression models are presented in Figures 7 a to c. In most instances a second order polynomial regression showed a better correlation than a simple linear regression.

Table 9 Percentage difference between the objective survey estimates and the official final estimated yields for the three sampling provinces and for the seasons 2002 to 2008.

	White maize Difference (%)	**Yellow maize Difference (%)**
Mpumalanga	5	8
Free State	4	-2
North West	2	-1



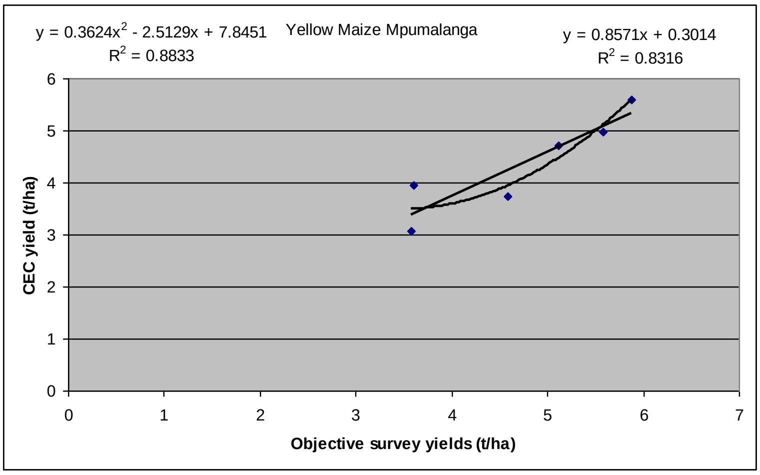


Figure 7 a Linear and polynomial regression models for maize yield estimates in Mpumalanga

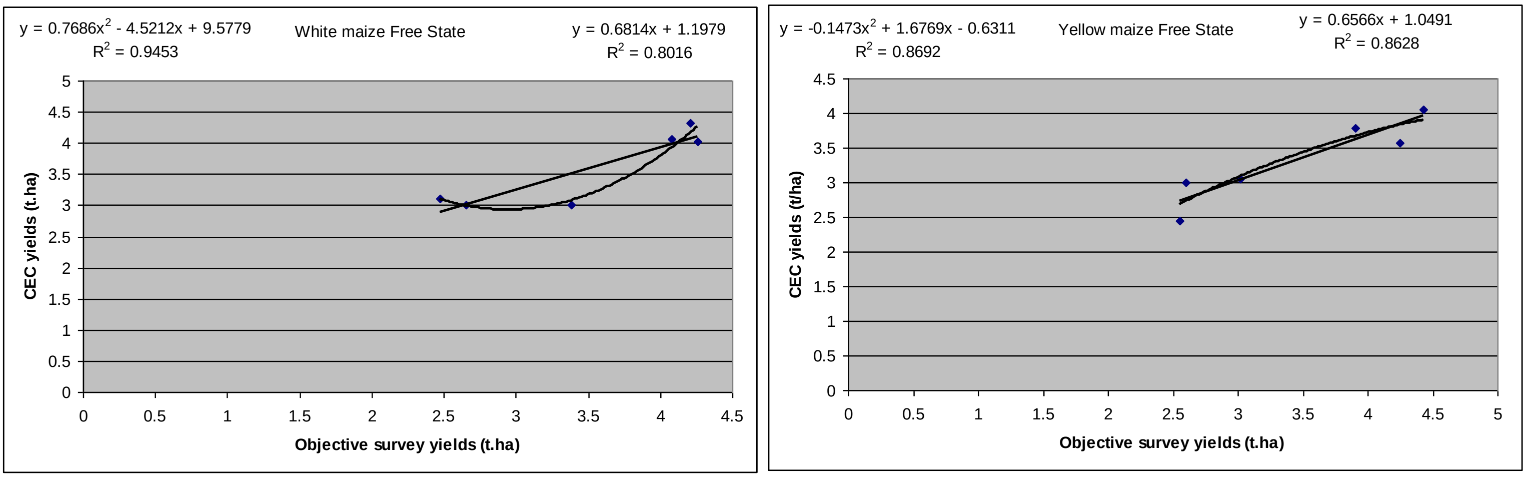


Figure 7 b Linear and polynomial regression models for maize yield estimates in the Free State

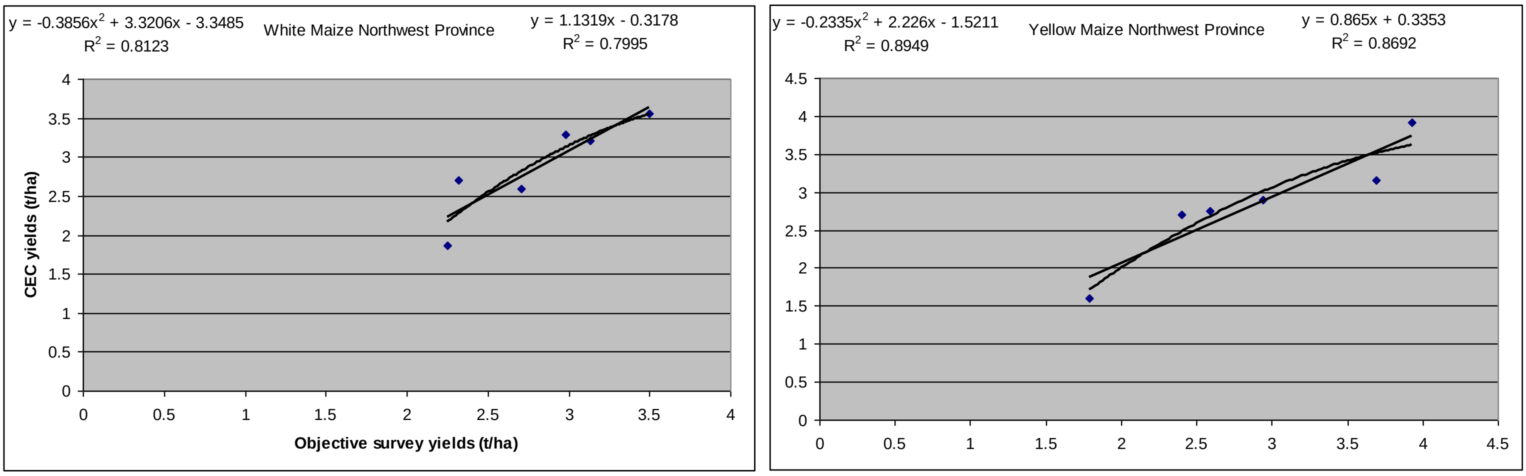


Figure 7 c Linear and polynomial regression models for maize yield estimates in the North West Province

Although these regression factors have been developed they have not been implemented in the operational OYS in South Africa, because they do not form part of the conceptual framework the CEC has of the OYS system. However the bigger problem is that the yield figures estimated by the CEC are often out of context as the CEC does not adjust its area estimates during the season and an error in the area estimate is compounded in the yield estimate. Yield figures might thus be equivalently lower or higher than they should be.

Weighted average yields

Although through the stratification procedure the average yield calculated for a province should be self-weighting, post analysis of 2006/07 yield figures indicated that the OYS methodology could be improved by better grouping area under dry-land and irrigation and using the area to calculate a weighted average. This method of calculation was first presented to the CEC in 2007/2008 season (Table 10). Using this method the percentage over or under production can be calculated.

Table 10 Example table of figures presented to the Crop estimates Committee (CEC)

North West 2011/12 production season						
		Area (ha) (PICES)	**Yield (t/ha)**	**Production (t)**	**CEC production (t) (April)**	**Difference (%)**
White Maize	Dry	(planted) 593 600	2.86	1 695 313		
	Irrigated	(planted) 21 200	10.64	225 575		
	Total/Average	(harvested) 621 500	3.09	1 920 887	1 952 000	-2%
Yellow Maize	Dry	(planted) 116 900	2.29	268 070		
	Irrigated	(planted) 25 500	9.88	251 965		
	Total/Average	(harvested) 145 000	3.59	520 035	496 000	+5%

Outlier identification

As in any survey, errors occur due to: human nature, measurement errors and calculation errors. Outliers have an influence on the mean. Since 2007/2008 in the analysis of the OYS data all outliers that are 2.5 standard deviations from the mean are identified (Figure 8). An inquiry about these observations is made, looking into measurement error by the enumerator (i.e. inches instead of cm for row width), erroneous calculations (wrong moisture percentage) and the spatial location (water table soils) of the sample field. If the mistake is revealed the correct value of the observation is used in the analysis. If, however, no reason of the high or low yield can be found this observation is omitted from the calculation.

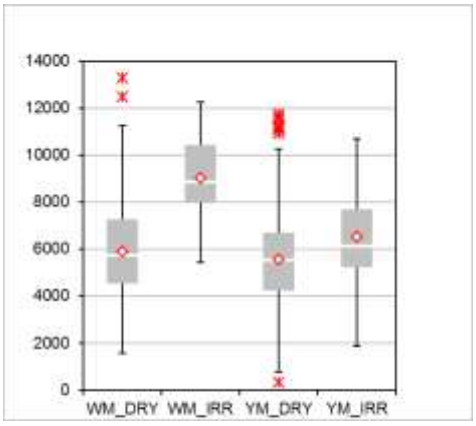


Figure 8 Example of a Box plot for outlier identification

Mean versus median versus trimmed mean

For the calculation of a yield from a survey the mean, median or trimmed are different calculation methods that can be used (Prof M. Kidd, 2011 personal communication). In a post analysis of the OYS data for the 2010/11 season all three methods were evaluated. The difference in the final yield between the three methods was only found in the second and third decimal of the yield. Regarding this, it was decided only to use the mean in the calculation of the operational OYS figures.

Discussion

Various diverse facets of the objective yield system have been investigated in the study and recommendations have been implemented into the operational OYS. The most significant of these were the implementation of the RA method that enables enumerators to survey more fields to a greater accuracy than the original method suggested by NASS. The second was the introduction of the harvest difference factor and the third the calculation of weighed average yield based on dry-land and irrigated production. The objective yield survey has been a great success, indicating to the CEC not only yields, but also areas where there are discrepancies in the yield versus area e.g. North West Province. On a national level the OYS has been within the 8% target as set by the CELC for May (Figure 9). There is however a tendency for the OYS combined with the PICES area to result in an over estimate of production at national level.

It must be kept in mind that the target of the research was to investigate methods to increase accuracy and not the delivery of yield figures to the CEC. The accuracy of the yield figures to the CEC from an operational OYS are influenced by the accuracy at which enumerators sample fields. Also in comparing yield figures directly to CEC figures is not optimal as yield figures may be skewed through incorrect areas and cross border deliveries.

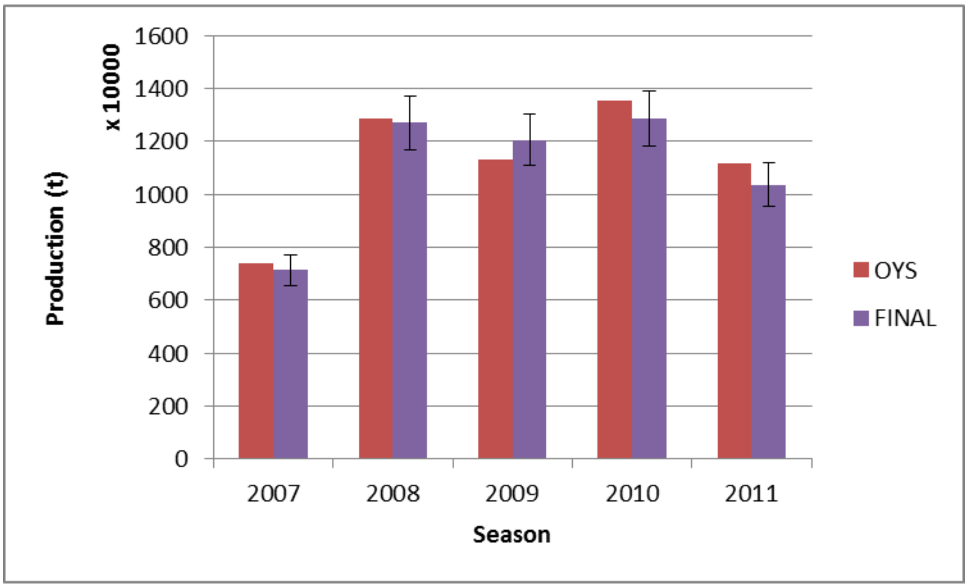


Figure 9 National production as forecast in May using OYS and PICES area for the seasons 2006/07 to 2010/11 compared to the final production estimate with a 8% error tolerance.

Crop Modelling

Introduction

There are many approaches in climate-based yield forecasting. In essence, the methods use climate data and crop models. The crop yield models provide a conceptual / mechanistic explanation of a crop's response to external factors such as climate, soil and crop management conditions (Egli and Bruening, 1992; Boote *et al*., 1996; Hoogenboom, 2000; Matthews, 2000; Palmer *et al*., 2004). The basic assumption in an agro-meteorological approach to yield forecasting is that climate conditions influence crop growth and yield formation and fluctuations in crop yield can be attributed to variation in climate. Crop yield models are thus becoming increasingly important in translating information on climate variability into forecasts and recommendations tailored to the needs of decision makers (Hansen and Jones 1999). Numerous crop yield forecasting efforts have made use and are currently using dynamic crop simulation models as tools 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).

Illustrative of a successful crop yield forecasting system using crop models is the Monitoring Agriculture with Remote Sensing - Crop Yield Forecasting System (MARS-CYFS). It has been used with success by the European Union (EU) since 1990 (http://mars.jrc.it/mars/About-us/AGRI4CAST/Crop-yield-forecast; G. Genovese, MARS-STAT, personal communication). MARS-CYFS is an international system that forecasts yields for eleven crops of 35 member states. Qualitative assessment has indicated that the system, for the main crops, has a mean absolute percentage error (MAPE), over the six year period 1996 - 2002, of only 3 - 5% at EU level (Genovese and Bettio, 2004). Genovese and Bettio (2004) have shown that at EU level, over the six year period for maize, the root mean square percentage error (RMSPE) was higher at the beginning (6.26%) and lower towards the end (2.77%) of the season. The average forecasted RMSPE error for maize over the eight months of forecasting and the six seasons was 4.46%.

The main aim of the study is to develop a maize yield forecasting tool for South Africa using crop models on a spatial and temporal scale which is able to deliver an objective, producer independent, unbiased yield forecast for white and yellow maize at, not only a provincial and national level but on a multi-scalar level for much wider application by the maize industry than what is currently the case.

This framework should be:

- Able to forecast yields for white and yellow maize within a correct spatial context.
- Multi-scalar, meaning that forecasts should be able to be generated for farm, district, province or national scale, enabling a much wider application by the maize industry than what is currently the case.

Inputs should be:

- Objective.
- Producer independent.
- Near real time or in-season.
- Readily available.

Outputs should be

- Unbiased.
- Within 5% of the realised production figure in May.

Numerous types of crop models are available, ranging from simple statistical rule based models i.e., Crafford and Nott, Smiths Climatic criteria to semi empirical models of intermediate complexity such as ACRU, WRSI and STIN, to complex simulation models such as PUTU, DSSAT, Aquacrop and Apsin as examples. Evaluating the PUTU and CERES maize models for yield estimations over a larger area De Vos and Mallett (1987) and Prinsloo and du Toit (1996) found the CERES- maize model to be more accurate at this scale. CERES maize has since 1995 been used as a model in drought monitoring and forecasting of maize yields in the Free State Province (Van den Berg and Potgieter, 1997; van den Berg and Manley, 2000). Since 2001, CERES maize (SA) is used to estimate maize yields for six provinces (Durand and Du Toit, 2007).

Both the ACRU-maize model and the CERES-maize (SA) seem to be appropriate for the development of a multi-scalar maize crop yield forecasting system. However, ACRU-maize is less known to industry stake holders and has not yet been tested for maize yield predictions at a regional level. An advantage of the CERES model is that it can incorporate different management regimes accounting for dry-land, irrigated, different maize varieties i.e., white or yellow maize, fertilisation, planting dates, plant population and row widths. A current disadvantage of CERES-maize is however, that it is no longer recognised as the latest “stand alone” model as it has been incorporated into the Crop Systems Model of DSSAT.

To achieve the aim of developing a maize yield forecasting tool for South Africa the following matters had to be addressed:

- The upgrading of CERES-maize (SA) and of DSSAT3.5 to the newer CSM of DSSAT4.0 through:
  - Crop model verification
  - Crop model calibration
- The reorganising of the input data from random locations (points) to be representative of maize production practices employed in different regions (field, district, province and national) through the:
  - Development of a multi-scalar approach
- Organising input data for climate, soils, planting date, plant population, row width and maize varieties that is representative of the fields within all regions in model required format through the:
  - Development of transfer functions for model data requirements
- Dry-land and irrigated white maize and yellow maize fields are found in certain locations. Identifying the exact spatial distribution will:
  - Increase spatial accuracy of yield forecasts using the crop model
- Climate changes during the season. Accurately predicting the climate early in the season will:
  - Increase temporal accuracy of yield forecasts using the crop model
- All the inputs have to be executed one after the other in an automatic manner - the so called model runs. The output of the model runs have to be offset to yield at the required scale. All this is done by developing a program (computer code) which is the:
  - Development of a systems approach using the crop model for yield forecasting
- After each season model outputs have to be verified for accuracy by comparison to other sources i.e. CEC yield figures, OYS. This process will:
  - Validate and scientific verify yield figures from a system using the crop model.

Figure 10 presents a schematically overview of the approach that is being followed to develop a maize yield forecasting system for South Africa using crop models.

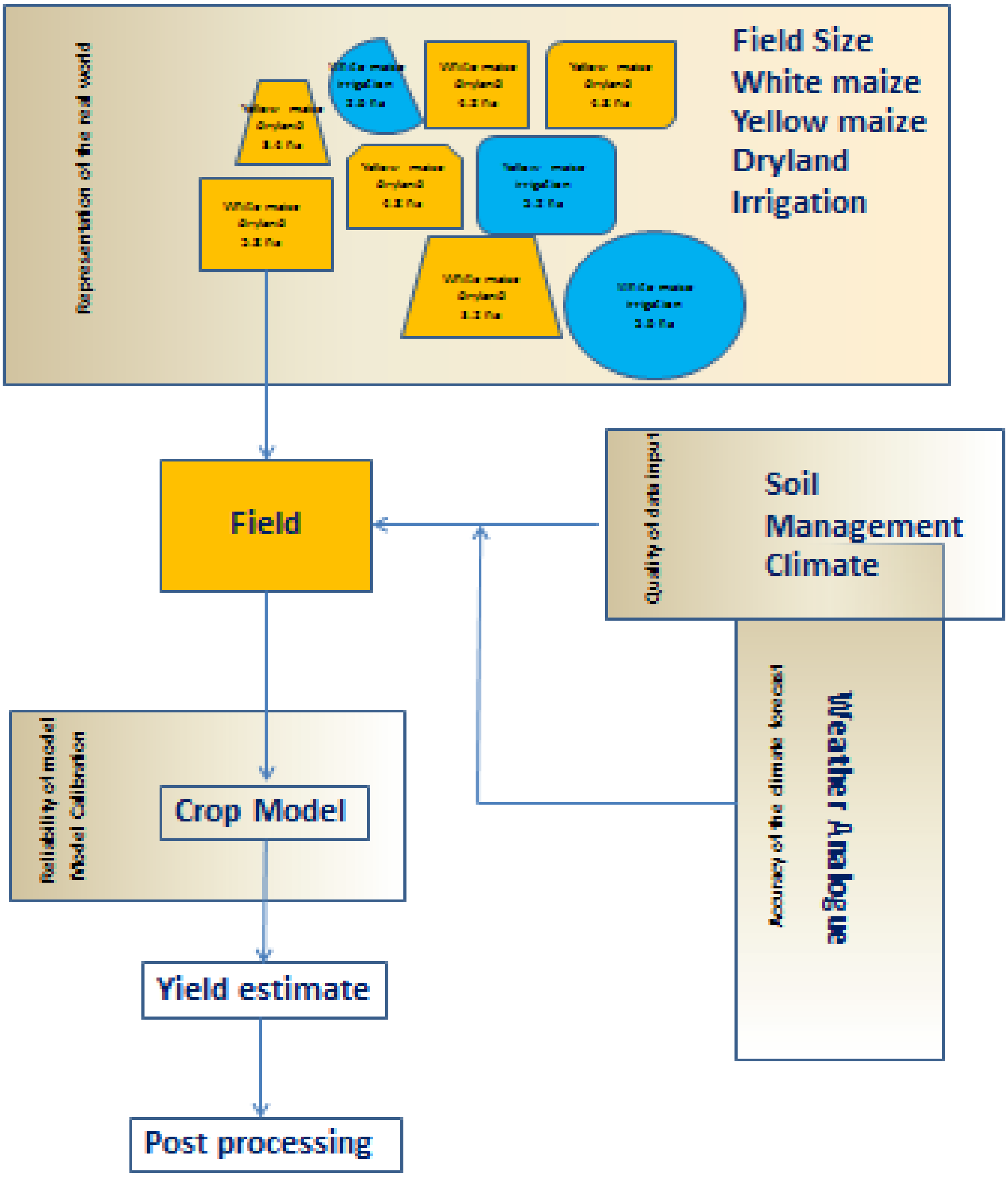


Figure 10 Schematic overview of the Integrated Decision Support system for Yield Estimation in South Africa (IDSS-YES).

Material and methods

The study relies on the accurate computation of yield figures that are representative for different scales, but with the main focus of delivering an accurate yield forecast early in the season. The tools necessary for such a study is computing hardware with specifications to process large volumes of data and relevant software. The standard packages used in this study are DSSAT version 3.5, DSSAT version 4.0.2.0, ArcGIS 9.3, ARCGIS 10, Spatial Analyst, 3D Analyst, Hawths tools, Spatial ecology, Microsoft Office Professional, Matlab, VisualFoxPro, VisualBasic.Net, VisualBasic, a Fortran compiler and GENSTAT. A number of codes have been purposefully compiled to automate various tasks.

Data used in the study was scoured from various studies, persons and institutions. Table 11 is a list of data sources and data providers. Much of the data received is “raw” data, i.e. data not in model usable format, not at the right scale, not in the right units, disparate etc. Much of the data requires extensive manipulation to achieve model ready input status.

Table 11 Data sources and data providers for crop modelling

Data type	**Data element**	**Data Provider**
Climate	Rainfall	ARC-ISCW
		University of Natal
		Johan van den Berg
		MGK
		Senwes
		Suidwes
		OTK
		VKB
		Afgri
		NWK
	Temperature	ARC-ISCW
		University of Natal
		Johan van den Berg
	Solar radiation	ARC-ISCW
		University of Natal
		Johan van den Berg
Climate	Cultivar traits	Maize cultivar trails
		Literature
Soil inputs	Soil depth Clay percentage Slope/run off Soil colour Rooting depth pH Bulk density Drainage rate	Land Type- ARC-ISCW
		Hein Beukes
		Modal soil profiles- ARC-ISCW
		Literature
		Astrid Hatting
		ACRU model
Crop management	Planting date Row width Planting density	Objective yield survey
		Maize cultivar trials
		Start of season model
Field crop boundaries	Delineation	Geoterra Image

Data type	Data element	Data Provider
Crop classification	Maize cultivation Dry land production Irrigation	Geoterra Image
		SIQ-PICES
		Objective yield survey
Verification data	Field scale	Objective yield survey
	District scale	DAFF-Sea postal survey
		Subjective yield survey
	Provincial scale	Crop Estimates Committee
		DAFF-Sea postal survey
		Subjective yield survey
		Objective yield survey
	National scale	Crop Estimates Committee

Results

In this section a brief overview is presented of the achievements up-to date in the development of a maize crop yield forecasting system for South Africa. The system was given the acronym of IDSS-YES standing for the Integrated Decision Support system for Yield Estimation in South Africa.

Crop model verification

Verification refers to the internal consistency of the model and the software implementation. This requires the source code and refers to the algorithms used. Validation is a sequence of tests and checks that convince the user that the model is valid for the intended purpose. A model must be validated at the same spatial scale and with the same type of data as those that will be available in operational work. The DSSAT's CERES-maize and DSSAT's CSM has been developed with the aim of global application (Jones et al., 2003). Thus model developers suggest only calibration of the cultivar and/or ecotype coefficients that are required for accurate yield simulations in different regions of the world (G. Hoogenboom, personnel communication, 2008). To enhance regional performance in South Africa, du Toit (du Toit, 1992, du Toit *et al*., 1994a; du Toit *et al*., 1994b; du Toit *et al*., 1994c; du Toit, 1995, du Toit *et al*., 1998; du Toit and Prinsloo, 2000; du Toit *et al*., 2002) however, added and changed some of the model algorithms. The changes were incorporated into the releases of CERES maize for DSSAT version 2.1, 3.0 and 3. From all the model modifications introduced into the CERES-maize to enhance its regional performance only certain elements, by means of a sensitivity analysis, were chosen to be included into the DSSAT v4.0 CSM. Included and verified were the prolificacy and tiller algorithms, calculation of P3, calculation of P9, change of the root, stem and leaf development rate in ISTAGE 3 and 4, the reduction of water stress on leaf size and weight and removing of the soil water deficient factor. Not included was the water-logging subroutine (du Toit and Prinsloo, 2000) as the DRAINE subroutine has been replaced by a new soil model component that now also simulates perched water tables. Changes to the PCARB subroutine were also not introduced as the algorithm in DSSAT v4 CSM addresses the deficiency that was identified by du Toit (1994a).

Figure 11 a to c illustrates the increase in accuracy achieved by including the algorithms into the CSM of version 4 of DSSAT.

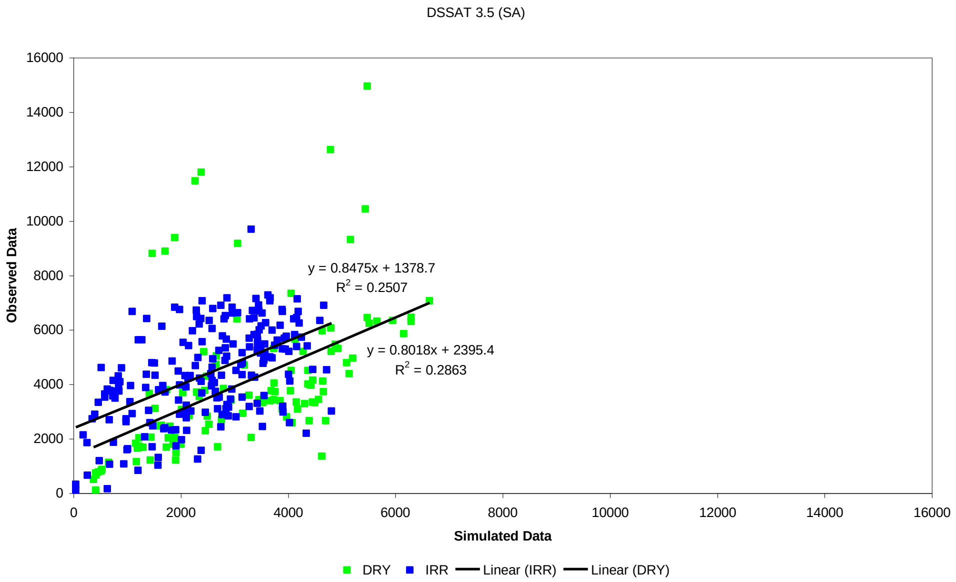


Figure 11a Model verification using DSSAT 3.5 South African version.

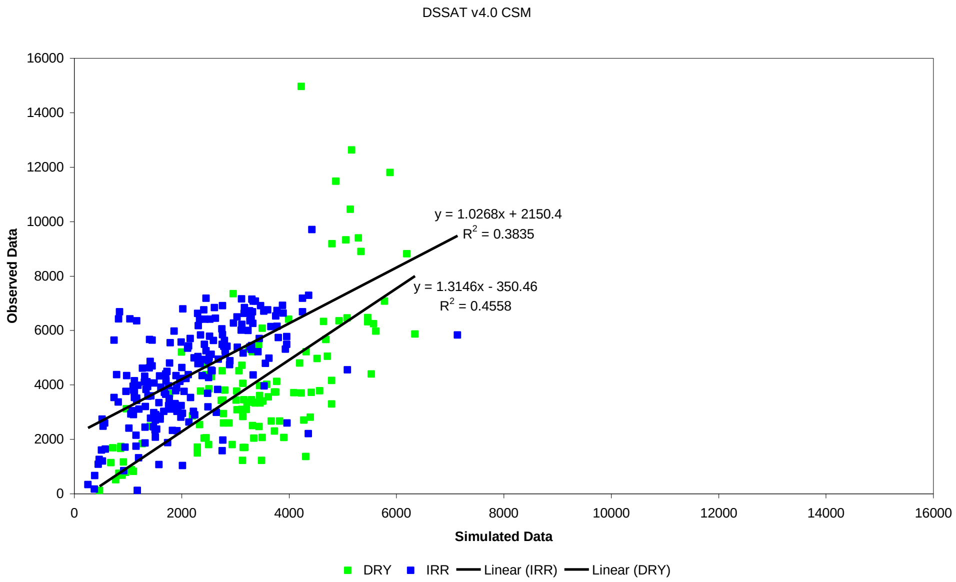


Figure 11b Model verification using DSSAT 4.0 Crop systems model.

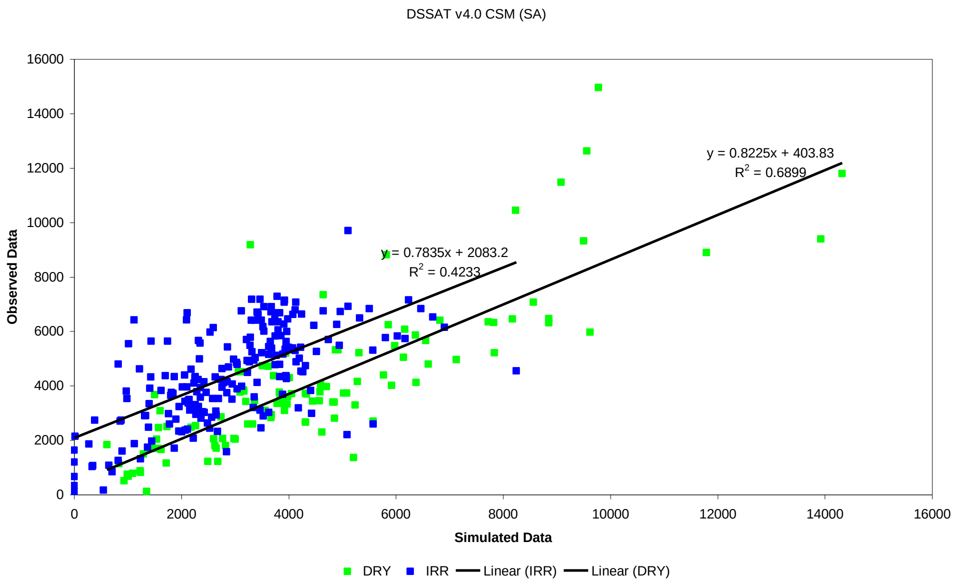


Figure 11c Model verification using DSSAT 4.0 Crop systems model South African version.

Crop model calibration

Calibration refers to the fine tuning of a model. Real-world data is used and the model is run to see if it mimics reality sufficiently well. Calibration is mostly done by trial and error until the "best" set of parameters is selected. The greater the variety of training data, the greater the chances that the model will perform satisfactory under new conditions. Within the CSM of DSSAT there are two sets of coefficients that require calibration i.e. ecotype and cultivar coefficients.

Ecotype coefficients

To make the DSSAT modelling system more adaptable to regional application model developers have taken certain coefficients, that were previously hard coded in CERES-maize, and made them accessible to crop modellers to enable better calibration of the model to different climatic regions, named ecotype coefficients. For the CSM simulating maize there are nine coefficients as presented in Table 12.

Table 12 A review of ecotype coefficients based on literature

Coefficient	Description	Default value	Suggested from literature	Suggested for SA ecotype
TBASE	Base temperature below which no development occurs (°C).	8	De Jager 10 Schulze 10 Eicker 10	8
TOPT	Temperature at which maximum development rate occurs during vegetative stages (°C).	34	Eicker 25 De Jager 30 Schulze 32	30
ROPT	Temperature at which maximum development rate occurs for reproductive stages (°C).	34	De Jager 30 Schulze 32	30
P2O	Day length below which day length does not affect development rate (hours).	12.5	Du Toi 14.2	14.2
DJTI	Minimum days from end of juvenile stage to tassel initiation if the cultivar is not photoperiod sensitive (days).	4.0		4.0
GDDE	Growing degree days per cm seed depth required for emergence (GDD/cm)	6.0		4.0
DSGFT	GDD from silking to effective grain filling period (°C).	170		170
RUE	Radiation use efficiency (g plant dry matter/MJ PAR).	4.2	Tsubo et al (2.26-2.85	4.5
KCAN	Canopy light extinction coefficient for daily photosynthetically active radiation or photon flux (PAR)	0.85	De Jager 0.7 Du Toit et al Row width 0.8m 1.0 Row width 1.0m 0.83 Row width 1.5m 0.77 Row width 2.0m 0.76 Row width 2.5m 0.76 Row width 3.0m 0.74	< 1.0m 0.85 >1.0 m 0.77

Cultivar coefficients

In South Africa over 230 white and yellow maize varieties are listed. Three types of maize cultivars are distinguished in terms of the number of days from planting to silking. Fields are planted either to a single cultivar, or a mixture to spread the risk should one cultivar fail. At field and farm level the accuracy of a simulation can be improved by using the genetic coefficients of the cultivars that were planted. However, at higher levels such as magisterial district or provincial the type of maize cultivars planted to each field is unknown. To overcome this, two options have been identified: either the genetic coefficients of a cultivar with good yield stability over a range of environments can be used or generic coefficients that are optimised within a certain environment can be developed. DSSAT 3.5 was released with a genetic coefficient calculator. However this tool is not available in the DSSAT 4 release. For this purpose software was developed and simulations for the development of cultivar coefficients, unique to the delineation of maize ecotypes, were set up. These coefficients are referred to as fitted whilst those derived from actual field measurements are observed.

Calibration

Using an independent set of trial data over different locations, years and management regimes confirms the better simulation ability of DSAAT v4.0 CSM (SA) as indicated in Table 13. Although the R^2 and D-index are quite similar, the smaller Mean Difference (MD), smaller Root Mean Square Error Systematic ($RMSE_S$) and the smaller difference between the Root Mean Square Error Unsystematic ($RMSE_U$) and Root Mean Square Error (RMSE) indicate a more accurate simulation by the DSSAT v4.0 CSM (SA).

Table 13 Independent evaluation of DSSAT v4.0 and DSSATv4.0 (SA)

	Observed	**DSSAT v4.0 CSM**	**DSSAT v4.0 CSM (SA)**
Minimum	663	0	0
Maximum	12400	13421	20041
Mean	4547	3352	4995
Standard deviation	2671	2948	4261
CV	59	88	85
Slope		0.81	1.23
Intercept		-348	-589
Mean Difference		-1196	448
Mean Absolute Difference		1725	1938
RMSE		2368	2809
$RMSE_S$		1295	753
$RMSE_U$		1983	2706
D-index		0.82	0.83
R^2		0.54	0.59

Another mean to calibrate a model across a region is to make use of benchmark sites. Three sites, one in each province of Mpumalanga, Free State and North West were identified and planted to maize (cultivar: PAN 6479, a medium season white maize cultivar) in the 2007/08 season using management practices appropriate to the region (Appendix H). Soil samples were collected and the climate data was derived from the nearest rainfall station. Yield was determined by averaging three samples taken using the rapid assessment method and obtained from the farmer. Six different model scenarios were run to evaluate the differences between the two models. The d-Index indicated that there is a relatively good agreement between modelled yield and actual yield using the adapted DSSAT4. Due to the high cost of maintaining such benchmark sites it was not repeated in the following season.

Development of a multi-scalar approach

DSSAT originally was developed at a plot or field scale. Scaling up is one of the great methodological challenges when attempting to apply a model that was developed at plot or field scale to a regional scale. Using field boundaries digitised from satellite images makes it possible to model at a regional scale, keeping true to the model integrity, by modelling singular fields. An area weighted yield can be calculated and summarized at any delineation or administrative unit, introducing a multi-scalar approach. When applying a model designed at plot or field scale to a regional scale some form of aggregation or averaging has to be implemented in input data. Using fields a basis for modelling allows the use of either spatial averages or actual field data as model inputs. The model inputs together with the extent to which the model outputs are summarised will determine the accuracy of the simulation for the region.

At the beginning of the study, real world representation was a point, representing locations where soil profile descriptions were available (total number of ±13 000. These were often not even representative of maize plantings. The current real world representation is a polygon coverage where each polygon represents a known maize field (±130 000) (Table 14). The advantage in moving from point to a vector based (polygon) and not to a raster, is, that the area can be associated with each yield. This allows for weighted aggregation to the required level such as average farm, district of provincial yields. Each field can also be associated with an unique input profile such as climate, soil and management.

Table 14 Average area planted to maize for the seasons 2003/04 to 2009/10 (7 seasons) compared to the area identified to have been panted to maize used for the study.

	Average area under maize 2003/04 to 2009/10 season	**Percentage of total area**	**Total area identified planted to maize used for crop yield modelling**	**Percentage representation**
Western Cape	2 904	0.11		
Northern Cape	49 424	1.95	26 421	53
Free State	984 429	38.77	1 312 396	133
Eastern Cape	16 343	0.64	3 036	19
KwaZulu-Natal	77 957	3.07	21 160	27
Mpumalanga	485 714	19.13	371 368	76
Limpopo	43 900	1.73	21 226	48
Gauteng	107 929	4.25	129 366	120
North West	770 571	30.35	607 668	79
Total	2 539 171		2 492 641	98

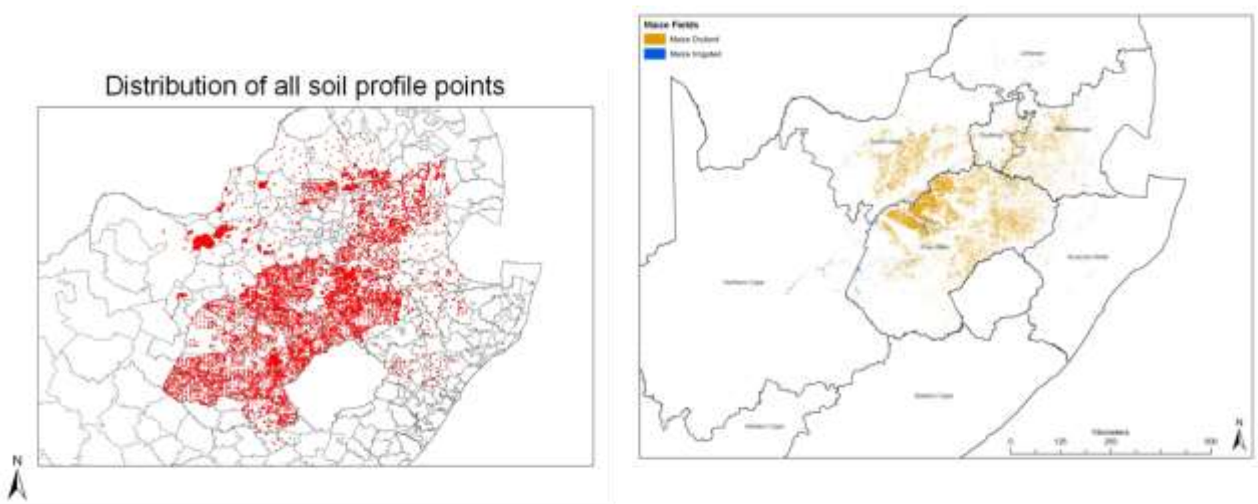


Figure 12 Spatial allocation based on points and field.

Development of transfer functions for model data requirements

In South Africa, as in many developing countries, availability of high quality input data to crop models is a major restriction. To bridge the gap between data needs of the model and data availability transfer functions were developed. This, to date, has proven to be a challenge especially with regard to soil and climate inputs as most is a very timeous effort due to the volume of data required for producing a forecasting system at a national scale. Only a very brief description without any technical detail follows for each of the main model input data elements.

Soil

DSSAT requires relatively detailed soil data variables. Soil information available for the study are the modal soil profile descriptions, own data, various other data providers and the spatial Land Types (LT) data base (person comm. D. Beukes, ARC-ISCW). For the spatial LT;s the original land type data base with a spatial reference to polygons were modified based on a 90m digital elevation model (DEM). This spatial LT data base formed the basis for the study. Soil series within terrain units (TU) unsuitable to maize production were eliminated i.e. al soil series with MB2 or higher, all soil series with a depth less than 400 mm and all soil series with 50% or higher clay content in any of the soil horizons. A weighted average of the characteristics (depth of the A horizon, depth of the B horizon, clay content of A horizon, clay content of the B horizon, slope, bulk density and drainage rate) for the remaining soil series was calculated for each TU. Using the zone function within a GIS a further weighted average for the individual TU's soil properties for each field were calculated. This information forms part of an access data base for model inputs. Where possible the exact information of soil properties for a specific field was sourced from other data.

Climate

Simulations at locations far from measured data, or where essential variables or periods are missing, must rely on estimated data which is often the case with climate inputs. The most common estimate is to simply use the nearest weather station as a proxy for unmeasured climate at the location of interest (Micale and Genovese, 2004). For regional applications spatial partitioning or Theissen polygons are methods for identifying the nearest station to any geographical point and for obtaining aerial weighting factors (Carbone et al., 1996; Rosenthal et al., 1998). Remote sensing offers a method for filling gaps in weather station networks (surface-measured) climate data (Gommes and Hoefsloot, 1998). Two other methods sometimes used to estimate daily weather data are spatial averaging and interpolation.

Although the study had access to many sources of climate data the following were problems:

- Spatial distribution of weather stations versus spatial partitioning required for model inputs
- Missing days in weather data
- Not all elements (rainfall, minimum and maximum temperature and solar radiation) required by the crop model available at all stations
- Errors in the data self
- Length in time of the data records
- Disparate data bases with not all the same data format

- Data not in model required format

Currently spatial partitioning is based on Theissen polygons for locations that measure rainfall. The missing elements such as minimum and maximum temperature and solar radiation are generated, based on monthly means. The length of the historic data base is relatively short (20 years). This implies that only rainfall is a correct driver from which yield is derived. However, if Theissen polygons are not drawn up representing some fields it may be associated with weather stations that are not representative because of distance, altitude changes etc.

Currently no institution hosts a climate data base with the four elements, namely precipitation, min and max temperature and solar radiation with a continual spatial coverage for South Africa on a daily basis (daily aerial climate) and with a time span of at least 50 years that is in-time and updatable. The nearest to this is the Quaternary Catchment database of the University of Natal (Schulze *et al*., 2003). This data base uses a so-called driver rainfall data approach and has daily climate data for all the elements from 1950 to 1999. A process is under way to use this data base as starting point to extend it to in-time.

Firstly all the available data has been compounded into one coherent database and most of the larger errors were eliminated. In the process, limitations include the sheer volume of data and the speed of both hard and software available. This necessitated the use of more powerful hard and software equipment at different institution. Rainfall has been interpolated and adjusted to long-term means following an adapted procedure described by Lynch (2004) (using both monthly and and annual averages). Rainfall for a quaternary catchment is calculated using zonal averaging. A process is underway to compile a temperature data base. This process is hindered by the fact that there are much less stations with available data. Use is being made of triangular irregular networks (TIN) for interpolation and Self organizing maps (SOMS), also known as Kohonen maps, to eliminate irregular transitions and form a continuum with the QC database. A undertaking of this facet of the study is to develop a method that will allow continual updating of the QCDB base using many different data sources that can contain missing values within a known software environment such as ArcGIS 10.

Management

Crop management inputs to crop models may include crop species and cultivars; planting date, planting spatial arrangement; irrigation, fertilizer and sometimes biocide applications and land preparation (mulching) and tillage. Spatial heterogenity of management can contribute to aggregation bias. Because management is seldom consistent from year to year, spatial representations of management variables are generally not available. Typical or recommended practices are therefore, often applied uniformly within a region (Stephans, 1995). The data from the objective yield survey and that of the maize cultivar trials are currently the only spatial representative data sources available. This data has been used to develop a spatial weighted regression based on rainfall within land types. In the development of the regression equation other properties such as soil depth, temperature and altitude were also included but they were found to have little influence.

Increasing spatial accuracy of yield forecasts using the crop model

When dealing with crop forecasts both the spatial and the temporal variation is important. Both aspects influence the yield, within and between seasons. Interfacing the crop model with a Geographic Information System (GIS) allows simulations analysis of the two aspects and a better understanding and interpretation of simulation results. In South Africa, national maize production constitutes the sum of white and yellow maize which is cultivated under either dry-land or irrigated management practices. Crop classification distinguishes land use to different crops using a set of satellite imageries covering a season. From an imagery it is, however, not possible to derive a distinction between white and yellow maize. Using the split between white and yellow maize at provincial level and the distribution of white and yellow maize of the OYS a method was devised to randomly allocate fields either to white or yellow maize keeping spatial proportions. Thus, if more of one type of maize is planted in a drier part it will be reflected in the spatial array. When averaging to a provincial scale these fields will have a higher influence due to weighted averaging.

Increasing temporal accuracy of yield forecasts using the crop model

In most crop estimate systems some assumptions are made concerning future weather. Most analogies are based either on the Southern Oscillation Index (SOI) or weather forecasts issued by meteorological institutions. Probabilities are calculated and simulations are run accordingly, resulting in a probabilistic outlook to the season. The South African maize market however does not react favourably towards probabilities, as it only gives ground to speculation. A statistical approach based on the D-index (Willmott, 1982) where no assumption towards future weather is made has the means of producing a yield figure that should still be accurate and reliable for industry use.

The tool that was developed and integrated into the system is the so-called weather analogue program (du Toit *et al*. 2001; du Tot and Du Toit, 2002). The closest analogue year from a historic database is identified positioning the current or forecasted period according to rainfall already received. This is one of the reasons why an extended historic climate data base is required for model input. The longer the historic component the higher the probability that a corresponding year will be found. Some minor changes were made to the code when a small bug was found.

Development of a systems approach using the crop model for yield forecasting

In the development of a systems approach all of the prior elements discussed are wrapped together. This is in the form of specialised code/computer program written to automate all the functions and where post modelling processing of data is prepared to deliver a yield at the required scale to the stakeholder. The part that the user sees of the software/program is called a shell. The part that communicates with the computer is the code, the calculations performed are functions and if these are grouped together they are processes. A programme can also call other programs to calculate inputs for further calculations. The IDSS-YES system that was developed using the programming language VisualBasic.Net contains all these elements. As is known all programs require continual revision as inputs change or more outputs are required. This necessitates changes to the code. During the course of the project the code was updated to accommodate:

- DSSAT 4.
- The difference in dry-land and irrigation calculations
- Better calculation of soil inputs
- Calculation of weighted averages for yield at provincial level
- Different variety types
- Over 700 different management regimes

Because of changes in the DSSAT4 structure, changes in the source code of the program were required to allow for automated simulations within the purpose written shell of IDSS-YES. A additional model was introduced using the PICES area to calculate the weighted average yield from model derived yields for irrigated and dry-land maize.

Validity and scientific verification of yield figures from a system using the crop model

For industry to be able to use the developed system with confidence as a source of information for crop yields, the model inputs but more importantly model outputs, have to be scientifically verified. The coefficient of variation of model outputs must be minimised. Hindcasts can be compared to actual yield or production, calculating the forecast error for different periods before harvest and the mean forecast error for the system intra and inter seasonal. Depending on the targeted accuracy the value of the system inputs can be evaluated. This target is set by CELC, on behalf of the industry. They have requested that the crop forecast not deviate by more than 8% in April and not more that 5% in June from the final crop (CELC meeting, 29 November 2007).

For yield a variety of verification figures are available at different scales. However, these are all associated with known accuracy errors:

- CEC figures: Yield is skewed due to error in area
- Subjective yield survey: Farmer bias only available once early in the season
- DAFF-Sea: Nonprobability sample

Production would be a better indicator, but area figures in turn, also have known errors. Currently there is no 100% reliable benchmarks available to accurately and scientifically verify yield figures at any scale if regional yields require validation. In the previous report and in the special report to the maize trust effort was made to validate yield figures.

Discussion

Since the beginning of the project in 2005, and especially after the successful study tour undertaken in November/December 2005, every effort has been made to develop a yield forecasting system using crop models similar to that of MARS of the Joint Research Commission (JRC) of the EU. However, a yield forecasting system's accuracy depends on the accuracy of four components: a) representation of the real world, b) reliability of the model used, c) quality of data inputs (i.e. soils, management and climate information and d) accuracy of the forecasted climate. Within the study most of these have been addressed. Evidence that research increased the accuracy of the crop yield modelling system is demonstrated in Figure 13 where only in the 2007/08 season production was underestimated by IDSS-YES, however, in this season the CEC also under estimated in May by over 10%.

Much attention has been given to the scouring, aggregation and assembly of quality data inputs to the model and is the component that will require continual up-grade as new data becomes available. An example would be incorporation of the crop classification of the Northwest province and Limpopo once this coverage becomes available. This will however necessitate up-grade in the soils database and the establishment of new links to the management and climate data bases. The IDSS-YES system was also designed to be applicable for in-season yield estimates, but can also be used for evaluating pre-season climate scenarios (drier/wetter season than the previous) and the impact of climate change (long term forecast using Global Circulation models (GCM)).

The establishment of benchmark sites, sites where season after season the development of the maize plant is monitored in relation to the interaction to climate, soils and management would not only have been beneficial to calibrating the crop model but would have added valuable benchmarking data to the development of a crop yield forecasting system using earth observation data.

Currently a simulation run for a field takes only 1 second and the database consists of 130 000 fields. With the current available computer (Core i7), in theory it would take 36 hours (1½ days) to complete a national scale yield estimate if no problems are encountered. This is, however, not ordinarily the case, as errors in the climate data are always present. Another limitation is that Windows 7 64-bit platform does not support the use of PIF files. This element has been dropped in Windows 7, because it is one of the main access points for computer viruses. The IDSS-YES shell however made use of this facility to call DSSAT. IDSS-YES can still be successfully run using windows XP. However, XP does not utilize all the speed of a 64 bit machine. If a solution to this problem can be found the speed of a simulation can be greatly improved.

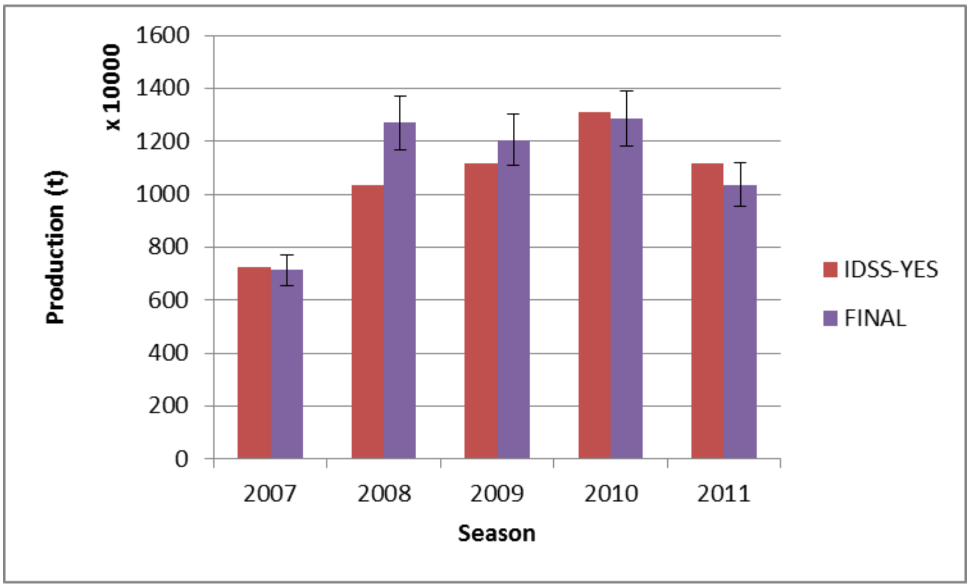


Figure 13 National production (t) as forecast in May using IDSS-YES yields and the previous months CEC area estimate for the seasons 2006/07 to 2010/2011 compared to the final production estimate with an 8% error tolerance.

Trend analysis

Material and methods

The trend analysis software, using analogue technology was developed by the ARC-GCI to assist the CEC to improve crop estimates for maize by indicating over or under estimation trends. The software requires at least three estimates for a season before a trend can be established. Using the analogue process two historic seasons following the same trend as the current season are identified using two seasons with the highest index of agreement from a historic database. For each of the two seasons respectively, the average is calculated using the current production value and the reconciled value of the historic season. An average between these two values is calculated and is the production figure estimated. Using the current seasons' area a yield can therefore be calculated.

Results

The trend analysis has annually been calculated since 2001/02 and production and yield figures have been reported to the CEC annually for May, June and July.

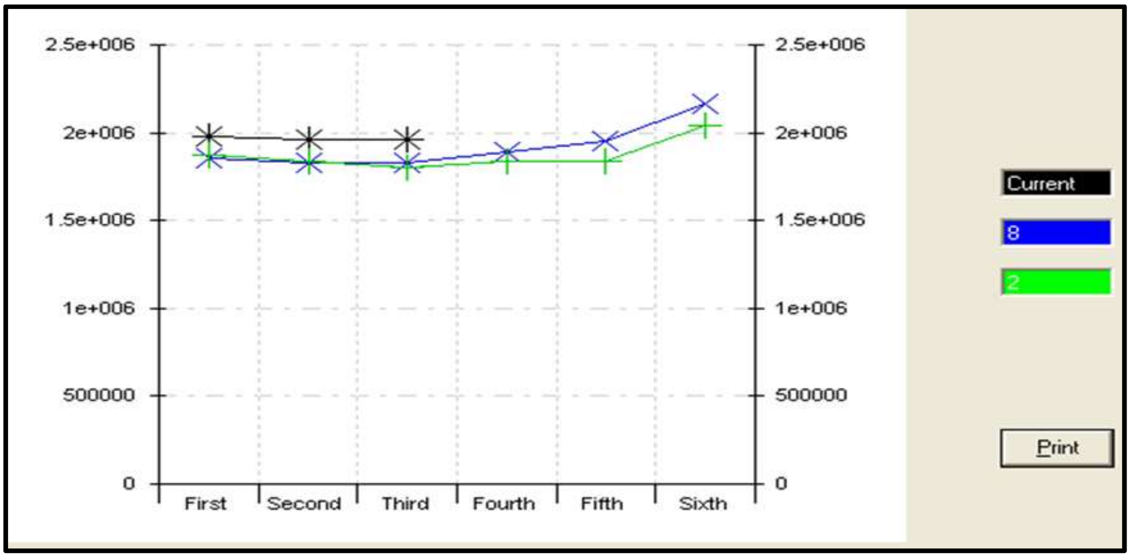


Figure 14 Example of yield trend observed using the trend analysis software

Table 15 Example of yield figures presented to the CEC

	White maize	Yellow Maize	Sun flower	Soy beans	Ground nuts	Sorghum	Dry beans
WC	10.64	10.00					1.90
NC	12.12	12.03	0.65	3.00			1.90
FS	3.95	4.14	1.35	1.27	2.83	2.96	1.09
EC	5.06	5.69		1.50	1.21		1.20
KZN	5.25	5.49		2.43		4.06	1.30
MP	5.05	4.70	1.28	1.35	1.00	4.69	1.28
LI	5.11	4.01	0.95	2.35		1.71	1.25
GA	4.90	3.94	1.30	1.50	1.24	3.27	1.17
NW	3.40	3.49	1.24	1.83		2.43	1.09

Table 16 Example of the over or under production as presented to the CEC

Crop	CEC-June (National production) tons	Trend analysis (National production) tons	Over/under estimation (%) July	Over/under estimation (%) June	Over/under estimation (%) May
White maize	6 369 250	6 751 263	-6%	- 7%	- 8%
Yellow maize	4 628 800	4 725 820	-2%	- 2%	- 2%
Total maize	10 998 050	11 486 552	-4.3%	- 5%	- 6%
Sunflower	780 470	792 551	-2%	- 6%	- 4%
Soy beans	699 250	700 445	0%	+ 2%	- 3%
Groundnuts	83 210	85 444	-3%	0%	- 2%
Sorghum	191 900	199898	-4%	- 4%	- 5%
Dry beans	45 250	45250	0%	+ 2%	+ 2%

Discussion

The trend analysis is an aid to the CEC to determine whether it is under, or over predicting. Generally, the data is acknowledged but not much emphasis is given to the figures and it is regarded as a so-called "c-line" data input, which means that the CELC has requested the CEC to use the figures only as an evaluation tool.

General discussion and recommendations

In hindsight this project should have been two projects, with one project addressing the OYS and trend analysis and the other the crop modeling component. An advantage of this project was that all the facets studied were improved and made part of the operational forecasting system as soon as they became available. To study the impact of the research on the improvement of the accuracy of the OYS and the crop modeling system two methods were used. In the first system, the index of agreement between the total production forecast for the three main provinces (Free State, Mpumalanga and Northwest) in May was used and compared to the official final estimate released at the end of the season for both white and yellow maize. This was done for the seasons 2001/02 to 2005/2006 grouped (5 seasons) and 2006/2007 to 2010/2011 (5 seasons) grouped. The rationale being that the earlier seasons represent yield estimates without the impact of research and the second with research applied. The second verification test implemented is the mean absolute error (MAE)[1] over the seasons. The three provinces were chosen because, firstly yield is not forecast for all provinces by all the methods and secondly they represent over 80% of the national production.

Table 17 Comparison of the D-index and mean absolute error for period 2002 to 2006, representing no research impact, with that of the period 2007-2011, which represents the period of gradual introduction of the research facets into the operational OYS system. (calculated using the May forecasts and the Final crop estimate).

		CEC	**OYS**	**IDSS-YES**
2002 - 2006	D-index	0.91	0.96	0.94
	Mean absolute error	10.6	9.4	12.7
2007 - 2011	D-index	0.99	0.99	0.96
	Mean absolute error	6.4	7.4	8.2

Both the D-index and the mean absolute error indicated that research increased the accuracy of the production forecast over the years. The D-index, the deviation from the 1:1 line, was better and the mean absolute error was smaller.

According to Wikipedia in decision theory, the expected value of sample information (EVSI) is the expected increase in utility that you can obtain from gaining access to a sample of

[1] In statistics, the mean absolute error (MAE) is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The mean absolute error is given by

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|f_i - y_i| = \frac{1}{n}\sum_{i=1}^{n}|e_i|.$$

additional observations before making a decision. The additional information obtained from the sample may allow you to make a more informed, and thus better, decision thus resulting in an increase in expected utility. EVSI attempts to estimate what this improvement would be before seeing actual sample data; hence, EVSI is a form of what is known as *preposterior analysis*. The value of the research and the more accurate yield figures presented to the CEC can also be evaluated in the light of these figures not being present when the decision on a forecast was fixed. The aid of the additional information and the confidence therein can be seen in the decrease of the mean absolute error tween the two periods that are compared. With this, the success of the project is illustrated.

Currently there is renewed international interest in the use of crop models to evaluate climate change, yield gap analysis, crop forecasting, management adaptation strategies and the use of generated data in mitigation processes. Even if the MAE of the IDSS-YES system is higher than that of the OYS (just out of the limit of 8% as set by CELC for the first four estimates) all indications are there that this can in future be lowered when the system uses better climate data, i.e. climate data where temperature is also measured and not simulated as has been the practice.

Through communication to stakeholders/funders the initiative, to rate the quality of the data inputs to the CEC based on scientific and evidence rating, was started. This lead to the establishment of the so-called A, B and C-line classification of data inputs that was suggested to be used as guideline to evaluate and verify inputs from the different data suppliers. This initiative can also be rated as an output of this project.

With this project ending the support of the Maize Trust is recognized for the modeling effort within the ARC-GCI.

References

ANON, 2006. The yield forecasting and estimation program of NASS, by the statistical Methods Branch, Estimates Division, national Agricultural Statistics Service, U.S. Department of Agriculture, Washington, D.C., May 2006. NASS Staff Report No. SMB 06-01.

BOOTE, K.J., JONES, J.W. & PICKERING, N.B., 1996. Potential uses and limitations of crop models. *Agronomy Journal*, 88, 704 - 716.

CHALLINOR, A.J., SLINGO, J.M., WHEELER, T.R. & DOBLAS-REYES, F.J., 2005. Probabilistic hindcasts of crop yield over western India using the DEMETER seasonal hindcast ensembles. *Tellus*, 57, 498 - 512.

DE VOS, R.N & MALLETT, J.B., 1987. Preliminary evaluation of two maize (*Zea Mays* L.) growth-simulation models. *South African Journal of Plant and Soil*, 4, 131 - 135.

DU TOIT, A.S., 1992. Die invloed van rywydte op die voorspellingswaarde van die CERES-maize mieliegroeisimulasiemodel in die Wes-Transvaal. Unpublished MSc thesis. University of the Free State.

DU TOIT, A.S., 1995. The quantification of the compensation ability of the maize plant. Unpublished PhD thesis. University of the Free State.

DU TOIT, A.S., BOOYSEN, J. & HUMAN, J.J., 1994a. Evaluation and calibration of CERES-Maize 1. Non-linear regression to determine genetic parameters. *South African Journal of Plant and Soil*, 11, 96 - 100.

DU TOIT, A.S., BOOYSEN, J. & HUMAN, J.J., 1994b. Evaluation and calibration of CERES-Maize 2. Phenology prediction values. *South African Journal of Plant and Soil*, 11, 121 - 125.

DU TOIT, A.S., BOOYSEN, J. & HUMAN, J.J., 1994c. Evaluation and calibratin of CERES-Maize 3. Row widths for the western Highveld. *South African Journal of Plant and Soil*, 11, 153 - 158.

DU TOIT, A.S., BOOYSEN, J. & HUMAN, J.J., 1998. Calibration of CERES3 (Maize) to improve silking date prediction values for South Africa. *South African Journal of Plant and Soil*, 15, 61 - 66.

DU TOIT, A.S. & PRINSLOO, M.A., 2000. Incorporating prolificacy into CERES-Maize prediction of kernel number. In: *Physiology and Modelling Kernel Set in Maize*. CSSA Special Publication No. 29. Crop Science Society of America and American Society of Agronomy, Madison, USA.

DU TOIT, D.L, SMALLBERGER, S.A. & DU TOIT A.S, 2001. The use of a weather analogue model to manage seasonal variability. INRM workshop, 28 - 31 August, 2001 CIAT, Cali, Colombia.

DU TOIT A.S., SMALBERGER S.A. & PRETORIOS A.J., 2002. Objective yield Survey Summer Grains 2002. Report to the Crop estimates Committee. ARC-Grain Crops Institute, Private Bag X1251, Potchefstroom. Unpublished.

DU TOIT A.S. & DU TOIT, D.L., 2002. Short description of the Model Statistical Package and Weather Analogue Program. In White, J.W. (Ed) Proceedings of a Workshop, CIMMYT, El Batan, Mexico, 2001. NRG-GIS Series 01-0. Mexico, D.F.

DU TOIT, A.S., PRINSLOO, M.A., WAFULA, B.M. & THORNTON, P.K., 2002. Incorporating a waterlogging routine into CERES-Maize, and some preliminary evaluations. *Water SA*, 28, 323 - 328.

DURAND, W., 2008. Personnel communication, 2008. Crop Estimates Liaison Committee meeting held in the Sefala Building, DAFF, Pretoria, 27 November 2008.

DURAND W. & DU TOIT A.S., 1999. Devloping methodology to use the nationsl cultivar trial yield data a filter to increase the accuracy of yield forecasting. Report to the Crop estimates Committee. ARC-Grain Crops Institute, Private Bag X1251, Potchefstroom. (Unpublished).

DURAND, W. & DU TOIT, D.L., 2007. Special report to the crop estimates committee on crop modelling and objective yield surveying. (Unpublished).

EGLI D.B. & BRUENING W.P., 1992. Planting date and soybean yield: evaluation of environmental effects with a crop simulation model: SOYGRO. Agric. Meteorol., 62: 19 - 29.

GENOVESE, G. & BETTIO, M., 2004. *Methodology of the MARS crop yield forecasting system Volume 4.* Statistical data collection, processing and analysis. Office for Official Publications of the European Communities, ISBN 92-894-8183-8.

GOMMES, R., & HOEFSLOOT, P.,1998. Gaps in maps, estimation of missing data in agricultural tatistics maps. In : *Proceedings of the EU/COST-79 Seminar on Data Spatial Distribution in Meteorology and Climatology*, Volterra, 28 Sep.-3 Oct. 1997, EU, Luxembourg, EUR18472, 155 - 168.

HANSEN, J.W. & JONES, J.W., 2000. Scaling-up crop models for climate variability applications. *Agricultural Systems*, 64, 43 - 72.

HANSEN, J.W., CHALLINOR, A., INES, A., WHEELER, T., & MORON, V., 2006. Translating climate forecasts into agricultural terms: advantages and challenges. *Climate Research*, 33, 27 - 41.

HOOGENBOOM, G., 2000. Contribution of agrometeorology to the simulation of crop production and its application. *Agricultural and Forest Meteorology*,103, 137 - 157.

JONES, J.W., HOOGENBOOM, G., PORTER, C.H., BOOTE, K.J., BATCHELA, W.D., HUNT, L.A., WILKENS, P.W., SINGH, U., GERSMAN, A.J. & RITCHIE, I.T., 2003. The DSSAT cropping system model. *European Journal of Agronomy*, 18, 235 - 265.

KIDD M., 2011 personal communication, University of Stellenbosch Centre for Statistical Consultation.

LYNCH, S.D., 2004. *Development of a Raster Database of Annual, Monthly and Daily Rainfall for Southern Africa*. Water Research Commission, Pretoria, RSA, WRC Report 1156/1/04. pp 78.

MATTHEWS, R. B., STEPHENS, W., HESS, T., MASON, T. & GRAVES, A. 2000. *Applications of crop/soil simulation models in developing countries*. PD82. 173p, Cranfield University Institute of water and environment, Silsoe, UK.

MICALE F. & GENOVESE G., 2004. Methodology of the MARS Crop yield forecasting system Volume 1. Meteorological data collection, processing and analysis. Office for Official Publications of the European Communities, ISBN 92-894-8183-8.

PALMER, T.N., ALESSANDRI, A., ANDERSEN, U., CANTELAUBE, P., DAVEY, M., DELECLUSE, P., DEQUE, M., DIEZ, E., DOBLAS-REYES, F.J., FEDDERSEN, H.,

GRAHAM, R., GUALDI, S., GUEREMY, J-F., HAGEDORN, R., HOSHEN, M., KEENLYSIDE, N., LATIF, M., LAZAR, A., MAISONNAVE, E., MARLETTO, V., MORSE, A.P., ORFILA, B., ROGEL, P., TERRES, J-M., & THONSPN, M.C., 2004. Development of a European Multi-Model Ensemble System for Seasonal to Inter-Annual Prediction (DEMETER). *Bulletin of the American Meteorological Society*, 85, 6, 853 - 872.

PRINSLOO, M.A. & DU TOIT A.S., 1996. Gebieds-oesskattings binne die Losdoringsgebied. Losdorings Oesskattingsforum. NAMPO.

SAVAGE, M.J., 1993. Statistical aspects of model validation. At Workshop on the field water balance in the modelling of cropping systems, University of Pretoria, South Africa.

SCHULZE, R.E., HALLOWES, L.A., HORAN, M.J.C., LUMSDEN, T.G., PIKE, A., THORNTON-DIBB, S. & WARBURTON, M.L., 2007. South African Quaternary Catchments Database. In: Schulze, R.E. (Ed). 2007. South African Atlas of Climatology and Agrohydrology. Water Research Commission, Pretoria, RSA, WRC Report 1489/1/06, Section 2.3.

STEPHENS, D.J., 1995. Crop yield forecasting over large areas in Australia. Unpublished Ph.D. Thesis. Murdoch University, Perth, Western Australia.

VAN DEN BERG, W.J. & MANLEY C.R., 2000. Maize: Framework for forecasting the extent and severity of drought, size of the crop and provide decision support information to maize farmers. In: *Proceedings of Highveld Ecoregion Workshop*. (Eds) Durand, W. and du Toit A.S., Potchefstroom, South Africa, 21 - 23 June 1999. ISBN: 1-86849-143-9.

VAN DEN BERG, W.J. & POTGIETER, A.B., 1997. Climatological and agricultural conditions in the Free State, January 1997. Monthly report, Free State Department of Agriculture, Section Information Technology.

WALDHAUS, E., OAKS, E. & STEINER, M., 1993. Unpublished. History and procedures of Objective yield Surveys in the United States. Washington Statistical Society Talk. [Internet:http://www.nass.usda.gov/Education_and_Outreach/Reports,_Presentations_and_Conferences/Yield_Reports/History%20and%20Procedures%20of%20Objective%20Yield%20Surveys%20in%20the%20United%20States.pdf]

WILLMOTT, C.J., 1982. Some comments on the evaluation of model performance. Bulletin of the American Meteorological Society, 63,1309-1313.

Reference list

ESTES, L.D., B.A. BRADLEY, H. BEUKES, W. DURAND, D. HOLE, M. LAU, M. OPPENHEIMER, R. SCHULZE, M.A. TADROSS, W. TURNER & D.S. WILCOVE, 2012. Comparing mechanistic and empirical models for assessing environmental suitability of a well understood crop species: Implications for ecological forecasting, Submitted to Global Ecology & Biogeography

ESTES L., OPPENHEIMER M., BRADLEY B.A, BEUKES H., DURAND W., SCHULZE R.E. & TADROSS M., 2010. A multi-model approach for assessing South African maize productivity and potential distribution: a comparison of empirical and process-based models, American Geophysical Union Fall Meeting, San Francisco, December 2010

A. SINGELS, J.G. ANNANDALE, J.M. DE JAGER, R.E. SCHULZE, N.G. INMAN-BAMBERL.D. VAN RENSBURG, P.S. VAN HEERDEN, C.T. CROSBY, G.C. GREEN, W. DURAND & J.M. STEYN, 2010. Modelling crop growth and crop water relations in South Africa: Past achievements and lessons for the future. South African Journal of Plant and Soil, 27(1): 25th Anniversary Edition 1983 - 2008.

DURAND W., 2010. Drought Adaptation Measures and Risk Tolerance of Commercial, Small-scale and Subsistence Corn Farmers in the Free State and Northwest Provinces of South Africa. Policy Note 0X-0910, September 2010 Water Science and Policy Center • 2228 Geology Bldg. Riverside CA 92521, 951-827-9772, http://wspc.ucr.edu.

DURAND W., 2010. Drought Adaptation Measures and Risk Tolerance of Commercial, Small-scale and Subsistence Corn Farmers in the Free State and Northwest Provinces of South Africa. International Drought Symposium - Integrating Science and Policy and Project Planning Workshop University of California, Riverside, 23 - 28 March 2010.

Crop Estimates Committee Media Releases

Summer Crops - Revised area and 1st to 7th , production forecast for the years 2005-2012. Directorate: Statistics and Economic Analysis, Private Bag X246, Pretoria, 0001. http://www.daff.gov.za/doaDev/cropsestimates/media. .