TOWARDS A FRAMEWORK FOR MAIZE YIELD FORECASTING; A COMPARISON OF TWO PROMISING METHODS

The Case of Mashonaland West Province, Zimbabwe

* V. Venus

* International Institute for Aerospace Survey and Earth Sciences (ITC)
P.O. Box 6, Enschede, The Netherlands

Preface

Thesis submitted to the International Institute for Aerospace Survey and Earth Sciences in partial fulfillment of the requirements for the degree of Master of Science in Rural Land Ecology Survey for Natural Resource Management purposes. Views or opinions expressed herein do not necessarily represent those of the Institute or the division. The Degree Assessment Board consisted of the following members:
Prof. Dr A.K. Skidmore, Head of the Division, Professor in Vegetation and Agricultural Land Use Survey; and, Dr. J. de Leeuw, Associate Professor, and Director of Studies; and, Prof. H. van Keulen, External Supervisor. The supervisors of this study were:
Prof. Dr Ir. P.M. Driessen, Professor in Modelling in Quantified Land Evaluation; and,
Dr. M.A. Sharifi, Associate Professor.

Abstract

Several alternative approaches for estimating maize yields have been proposed in the last decade, to assist food (security) planners in Zimbabwe. These were mainly aimed at improving crop status reports and used various techniques, each with its merits and limitations. In view of the country's limited resources, development of an accurate yield-forecasting framework that makes use of readily available data and technology poses a formidable challenge.

The first yield-forecasting framework considered in this study elaborates on the Crop Growth Monitoring System (CGMS) developed by the Joint Research Center of the European Union (EU). This system is currently used in crop yield forecasting at European level. The computed (water-limited) potential production can be reached, in theory, under optimum conditions of nutrient supply and weeds, pest and disease control, and is therefore not yet an accurate indicator of actual future yield levels. After aggregation to a smaller spatial scale, the model's output is entered in a regression equation of historical real-world and modeled yields to forecast the current season's yield level. The procedure regresses historical yield as a dependent variable with historical time trend and a model generated crop growth indicator as independent variables. The time trend is introduced to account for gradual, structural changes in yield that are brought about by technological or socio-economic changes. The model generated crop growth indicator accounts for year-specific weather effects. The indicator can be the standing biomass, the Leaf Area Index (LAI), etc. The yield-forecast relation obtained is then evaluated using a relative error technique. The equation with the highest accuracy for a specific period is used for forecasting purposes.

The second yield-forecasting framework considered in this study is largely based on the same principles. It uses the same crop growth model as the first one, but it incorporates also the links between El Niño/Southern Oscillation (ENSO) and rainfall anomalies in its crop yield predictions. To better estimate the further course of the current season's weather after the moment of prediction, it uses current

SOI patterns to improve on the straightforward synoptic weather patterns as used in short-term forecasts. The ENSO index employed are the SOI Phases as developed by Roger Stone. Non-parametric testing of twenty-three weather stations suggested that rainfall probability distributions for different Phases differ significantly. To limit the workload, only the 25, 50, and 75% percentiles of the cumulative rainfall distribution functions are retained as surrogate rainfall data. This tacitly assumes that the forecast part of the growing season is characterized by a weather pattern that lies between the lower and upper percentile. The thus derived surrogate weather data permit to complete crop growth simulation for the remainder of the season.

Evaluating the relationship by matching records of reported and predicted yields for a period of 9 years made it possible to successfully perform statistical correction of the predicted yields. The good correlation ($R^2_{adj.}$ = 0.84) between observed and modeled yields is proof of this. Although the present exercise availed of insufficient data to test this statistically, it is tentatively concluded that the second approach is outperformed by the first one in terms of accuracy. Under relative normal conditions, simulation results based on CGMS showed very high prediction accuracy, with a long-term average error of 9.4 %.

Acknowledgements

The author acknowledges the department of Agricultural Technical and Extension services (AGRITEX), Ministry of Agriculture, Zimbabwe for providing the yield statistics. Specifically, in the persons of Mr. Chitsiko, Deputy Director, Mr. Japi, S.A.E.O. district Hurungwe West, and Mr. Shepard deployed at the provincial HQ Chinhoyi. I would also like to thank the Zimbabwe Meteorological Services, Ministry of Transport, and Energy, for kindly providing the necessary weather data. Mr. Unganai, head Advisory Services of the same institute, is acknowledged for inviting me to present at their Seasonal Climate Forecast Workshop 1999, Harare.

Mr. Stone, Principal Research Scientist QCCA, Toowoomba, Australia is respected for his useful remarks regarding the El Niño modelling. My gratitude goes to Mr. Day, Principal Scientist (Drought Research) Queensland Department of Natural Resources and the Department of Primary Industries, Australia, for kindly providing the Southern Oscillation Index Phases and for being a “virtual teacher” in El Niño rainfall modelling. In this respect, Mr. van den Berg, Free State Department of Agriculture, Section Information Technology, Glen, South Africa is also thanked for his remarks.

My gratitude also goes to Mr. Van der Groot, CGMS development team Joint Research Center, Ispra, Italy, for his useful suggestions and for adding research specific functionalities and removing bugs from the crop growth modelling software.

My gratitude and thanks go to my supervisor, Mr. Driessen, for his advice, humor, and assistance in setting up both the hardware as the software to process all the data. Mr. Sharifi, my second supervisor, is acknowledged for providing some very useful literature. Mr. De Bie, fieldwork supervisor, is thanked for his help in “meeting all the right people” in Zimbabwe and for his friendship. I thank the ITC in general, and the Agriculture, Conservation, and Environment Division specifically, for providing the opportunity to learn how to conduct research according to scientific standards.

Contents

List of Chapters

List of Figures

List of Appendices

List of Abbreviations

ACMP	Agromet and Crop Monitoring Project
AGRITEX	Agricultural Technical and Extension services
ALES	Automated Land Evaluation Systems
CDF	Cumulative Distribution Functions
CGMS	Crop Growth Monitoring System
CSO	Central Statistical Office, Harare, Zimbabwe
ENSO	El Niño/Southern Oscillation
ETO	Potential transpiration from a crop canopy
FAO	Food and Agricultural Organization
FP	Forecast Period
FPAR	Photo-Synthetically Active Radiation
GIS	Geographic Information Systems
ILWIS	Integrated Land and Water Information System
IQR	Inter-quartile Range
ITCZ	Inter-Tropical Convergence Zone
JRC	Joint Research Center
LAI	Leaf Area Index
LPM	Locking Phase Month
NDVI	Normalized vegetation index
NEWU	National Early Warning Unit
POC	AGRITEX Provincial Office Chinhoyi, Zimbabwe
PREC	Precipitation
QX	Percentile X
RS	Remote Sensing
RTD	Total Rainfall Difference
SOI	Southern Oscillation Index
YFF1	Yield-Forecasting Framework Based on Actual Rainfall Data
YFF2	Yield-Forecasting Framework Based on Surrogate Rainfall Data

1 Introduction

Since the 1960s, growing demand for food and fibber products have been met through substantial increases in both area and per hectare yield (FAO, 1992). Agricultural planners face huge challenges, since nearly all land suitable for agricultural use has already been allocated. Moreover, the total area is expected to decrease in size due to depletion of land resources beyond their recovery. In addition, agricultural production especially in underdeveloped countries is variable due to inconstant availability of natural resources such as water. Since it is hard to keep food availability up with current and expected population increase and per capita consumption level, food security management is becoming increasingly important.

To cater for these problems many attempts have been made to evaluate and finally model land use and environmental processes that determine the suitability of land for defined uses (FAO, 1978). These models ultimately serve decision-making processes, notably the allocation of land to uses that provide the greatest sustainable benefits. Although there remains doubt whether processes of such complexity can be understood and modeled, progress has undeniably been made over the last decade. Before the '90s, the predictive strength of such models was limited since comprehensive evaluations of varying climatic-soil-crop conditions were lacking and processes were incompletely understood. Now they are better understood and can be partly modeled. As interest in automated data processing and analysis increased during the last two decades, tools such as Geographic Information Systems (GIS), Remote Sensing (RS), automated land evaluation systems (ALES) and digital databases have become generally available. This development has altered land use planning since simulation models facilitate multiple-objective 'what if?' scenario building. A most promising development is that of the quantified models. Many of such models are under development or have already been developed and simulate expected plant or animal growth and yields under varying circumstances. Examples are models developed by the FAO, or the PS12(3) model described by Driessen and Konijn (1992).

If confidence in simulation modelling increases, i.e. if conditions can be adequately modeled, crop growth simulation holds promise for early-warning applications. A reliable forecast system would reduce risks in potentially 'bad' years and maximize returns in potentially 'good years'. Such a framework for monitoring and forecasting yield would be useful to a broad range of users. Possible beneficiaries are grain producers and traders; particularly those involved in import and export, governments, financiers, farmers or farming co-operatives and production input suppliers, milling companies, extension services, and state planners, policy- and decision-makers. Decision-makers in South Africa already make use of maize yield forecasts to fix the maize price in March for bad seasons or farmers are advised to change their crops to more water-stress tolerant ones. In cases of 'good' seasons, milling activities are planned accordingly as well as arrangements for transport and silo storage and production loans. This has resulted in decreased risks and overall greater profits (de Jager et al., 1998).

Although attempts are currently under development, a reliable framework for crop yield forecasting is not available for Zimbabwe. Some of the questions that may arise when such a system will be developed are: "For what crop?" and; "What framework should be used?". The answer to the first question is relatively simple. Although not particularity suitable to the Zimbabwean climate, maize is the main staple crop and is used for a variety of uses, e.g. for human consumption (raw grain, crop residues for fuel or further processed to e.g. beverages), as a base for industrial products (oils, syrup and starch) or animal consumption (fodder, bedding). It contributes on average 40% of the calories needed in peoples' diets.

Since there are strong reasons to believe that a maize forecasting framework could yield valuable information, the second question "What framework should be used?" deserves more attention. There are many different approaches to forecasting yield, each having its merits and weak points. Different approaches classified according to the information they use are given in Appendix IV. Most promising are methods that integrate weather forecasting with a crop growth model. The methods described vary in complexity but have in common that they forecast a season's climatic patterns with the implicit assumption that the variability of future climate will be similar to that of historical records. At best such an analysis of anomalies in historical climatic conditions can provide an envelope in which season forecasts can be fitted; the direction and extend of the variance cannot be predicted. Hence, modeled yields may prove inaccurate.

At the same time, Vossen and Rijks (1995) proposed to combine a linear time trend with the results of crop growth simulation to explain the annual variation of yield per hectare. The Crop Growth Monitoring System (CGMS) at the Joint Research Center of the European Union is based on this premise. Statistical analysis is used to select the most robust predictor of yield for different stages in the growing season. To accomplish this, four indicators of modeled yields are regressed against historical yields, and the most significant one is used to forecast yield. These model results reflect the compound effect of soil-weather conditions throughout the growing season on crop growth. To account for the influence of increasing farmers' skill and use of technology a fifth predictor, the so-called 'time trend', is used. This trend of rising yields can be observed in official yields (Hooijer and van der Wall, 1994).

Amidst all the complexities, soil and crop input variables could be considered relatively constant; the determining, variable force in crop growth simulation for early-warning applications remains the weather (Sharifi et al., 1997). As described above, CGMS is not based on weather forecasts. Instead, it uses model predictors that differ in nature and accuracy as the season develops. In this manner, it avoids some of the difficulties as experienced in the traditional methods are described in Appendix IV. At approximately the same time, but very different in nature, another framework for predicting yield was developed. This approach does not attempt to 'avoid' forecasting weather by using statistics and interpreting indirectly weather variables, but it attempts to improve the 'robustness' of weather forecasting itself. Indeed, better results have been reported by researchers basing their forecast weather data on the findings of Stone et al. (1996) who showed that southern oscillation index (SOI) phase system provides an accurate predictor of rainfall in certain regions of the world. The SOI method considers 'phases' of the SOI; that is, the method uses both change and value of the SOI to derive cumulative rainfall probability distributions for any location. Because every month for every year can be placed into a particular analogue of months, those months can be placed together to take out daily rainfall, evaporation, temperature, and radiation. These data can be used as input for a crop simulation model. Consequently, the system provides both probability distributions of rainfall and potential yield. Justification of this technique can be found in Meinke and Hammer (1998) who demonstrated that highly significant differences in peanut yields in Australia exist among seasons grouped according to the SOI phases of Stone. De Jager et al. based their weather forecast component of a calibrated CERES—maize model on this same principle, with the intention to forecast the extent and severity of drought in maize in the Free State Province of South Africa one month before the growing season started. The accuracy of this type of forecasting system is yet uncertain but the high correlation value (r^2=0.86) for simulated versus actual yields is an encouraging sign (de Jager et al. 1998). A formidable challenge lies in applying both of these promising methods to a study area in Zimbabwe and comparing their results.

2 Objectives

The main aim of this research is to contribute to the development of a framework for maize yield forecasting by setting up and testing the relative usefulness of two approaches. This is done for a pilot area in Zimbabwe detailed in the sub-section *Study area,* Chapter Research Approach.

The two yield-forecasting frameworks are briefly introduced here and are further detailed under sub-section *Methods and materials,* Chapter Research Approach. With soil and crop input variables relatively constant, the determining factor for yield forecasting is the climate-soil interaction (Sharifi et al., 1997). Therefore, the frameworks should differ significantly, in how they address the interaction between climate and soil.

The first yield-forecasting framework, referred to as YFF1, uses CGMS as introduced in Chapter 1. This method was selected because its output is a reflection; *inter alia*, effect of climate conditions throughout a growing season on crop growth. CGMS uses a linear regression equation to forecast maize yield that may vary for different stages in the growing season for different regions. The regression equation includes the average actual (statistical) yield as the dependent variable and a model indicator to account for the year-specific weather effects, with or without a time trend for yearly increases or decreases of yield observed as a result of technological or possible socio-economical changes. This can be biomass, Leaf Area Index (LAI), etc., as the independent variables. After this, the prediction errors for all these equations are compared quantitatively. The elementary predictor with the highest accuracy is used for the forecasting purposes. The method has the potential of giving accurate (maize) yield predictions and is therefore selected for comparison.

The second yield-forecasting framework, referred to as YFF2 hereafter, is largely based on the same premises. It uses the same crop growth model and regression equation as YFF1, but for yield forecasting it also uses the alleged links between El Niño/Southern Oscillation (ENSO) and rainfall as introduced in Chapter 1. Note that the forecasting method does not look at synoptic weather patterns to give short-term forecasts. Instead, it uses current SOI patterns and suggests future rainfall probabilities based on lag-relationships derived from historical SOI patterns to predict the further course of the current season's weather. With this surrogate weather data, the crop simulation can be completed for a season. Both systems are further detailed in the following Chapters.

Since the usefulness of yield indications increases as they come available earlier and with more accuracy, the prediction error and the time of forecasting are used to compare the two frameworks. The prediction error is the relative error in the prediction of a yield forecast expressed in percentages by taking the difference between simulated and actual maize yield relative to the actual maize yield. This is done on a monthly basis for the forecasting period. The forecasting period is defined as the growing season, which is generally equivalent to the period November to May.

Although it cannot be proven *a priori*, it is expected that the relative usefulness of YFF1 is less than that of YFF2. This is because accurate predictions of maize yield earlier in the season are valued higher and the accuracy of YFF1 is expected to increase only halfway the growing season as more accurate model indicators of climate-soil interaction come available, e.g. water-limited storage organ yield. The weather forecasting of YFF2 allows regression of more trustworthy model indicators of a further or even fully developed crop, even though at the time of forecasting not physically present. In addition, the weather forecasting power itself is also expected to improve as the season progresses as it benefit from updated information on the state of El Niño.

2.1 Assumptions

For the feasibility of the research, assumptions have to be made. The following assumptions hold for both frameworks in general.

(a) Literature indicates that cultivation in the communal lands of Zimbabwe is characterized by small, irregularly shaped fields, which are inter-cropped and scattered across a vast expanse of bush and trees (Rugege, 1998). The area under maize is therefore estimated by a land classification module as part of CGMS, which is based on an evaluation of soil properties. Assumed is that this generalization does not affect the accuracy of the production estimates.

(b) The reliability of the official maize yield statistics is an important factor. The accuracy of official yields is unknown. This makes it difficult to separate the effects of simulation results from errors in the official statistics. Therefore, historical maize yield statistics are assumed accurate.

(c) The (water-limited) production maximum that will be obtained using the crop growth model can be reached under optimum conditions of nutrient supply, weed, pest and disease control, etc. and is therefore not yet an accurate indication of future yield. After aggregation to higher spatial scale, the model indicator is entered in a regression equation based on historical relationships (nine-year trend) between actual and modeled yield to arrive at a yield forecast for the current season. The assumption made here is that future farming practices only vary according to this trend.

Assumptions made with respect to the yield-forecast procedure of YFF2:

(a) To limit the workload, the testing of this yield-forecast framework will be limited to three weather scenarios per season. From the analogue years based on the current SOI phase the 25, 50 and 75% percentiles are selected to extrapolate the current season's weather. Therefore, an implicit assumption is that the forecast part of the growing season receives weather lying between the lower and upper rainfall totals head from the cumulative probability distribution.

(b) Since differences in rainfall probability distribution are tested for significance for only twenty-three locations, there is of course a possibility that significant differences identified are there by chance. It is an arguable assumption that the sample, i.e. the meteorological stations tested, is representative of the whole study area.

3 Research approach

The following Chapter describes the approach taken to test the two yield-forecasting frameworks for their relative usefulness.

3.1 Research questions

Research questions have been formulated to guide the research and permit a differentiation between primary and secondary issues. They relate to yet unknown or not fully understood underlying mechanisms of the two yield-forecasting frameworks under comparison.

3.1.1 Common issues

The first research question relates to the regression analyses of observed yields against model outputs. When actual yields do not show good correlation with modeled yields, changing the degree of aggregation up to levels with more reliable yield statistics does not always solve this problem. Possible introduction of new variables could improve the regression analyses.

At present, a limited number of model output variables, e.g. crop development stage, biomass, and modeled grain dry matter ('yield') are used with or without a possible existing time trend to relate yields of annual crops to the crop simulation model outputs. This rigid approach for selecting variables is imposed by the fact that most of the crop model inputs and outputs that may significantly account for inter-annual yield variability are strongly cross-correlated. It is on the other hand possible that some variables not used to avoid data redundancy may still, from a common sense point of view, significantly contribute to an improved explanation of yield variability. Rainfall is for example strongly related to soil moisture reserve, itself related to dry matter and grain production. But rainfall more than the simulated water requirements during part of the growing cycle might be useful to explain unexpectedly low yields of certain crops in certain regions or during certain years (Vossen, 1994). The possible number of variables here is almost unlimited and the selection must be careful and based upon objective considerations reflecting real constraints to crop production.

For example, excess rainfall during the flowering stage and/or water logging may be relevant to include, considering the agro-meteorological problems of the region, also observed in the last couple of years. The rainy season in Zimbabwe, especially in the northern areas, is not continuous from November to April. In general, it consists of alternating periods of relatively wet and dry spells lasting about 6 days which increases the change of water logging.

3.1.2 Framework-specific issues

The following research questions are listed separately, since mutually exclusive for both yield-forecast frameworks.

3.1.2.1 YFF2

Research has shown that the behavior of the Southern Oscillation is a useful indicator of summer rainfall over Zimbabwe (ZIMMET, 1998). However, to justify feeding a cop growth model with surrogate weather data generated and based on this premise an in-depth analysis of the impact of El Niño on the region is needed first. The temporal and spatial characterization of El Niño for rainfall estimation detailed in the sub-section *Surrogate meteorological data and their processing: YFF2* was required to address the following research questions:

(1) YFF2, which is based on the alleged links between ENSO and rainfall anomalies, assumes that significant differences in seasonal rainfall exist in the study area and affect maize crop seasons grouped according to Stone's SOI phases (Stone et al., 1996). Since this assumption needs verification, a non-parametric test (Kruskal-Wallis) has been applied for several seasons and for several meteorological stations in the area. The relating sub-hypothesis can be extracted as follows:

The null hypothesis (H_o) for research is that there is no systematic difference in forecast period (FP) rainfall totals between the seasons grouped according to the different SOI phases; versus the alternative hypothesis (H_a) for research, that there is systematic difference in FP rainfall totals between the seasons grouped according to the different SOI phases.

(2) El Niño/Southern Oscillation (ENSO) derived weather outlooks are based on the assumption that the atmosphere is the best model of itself (Unganai, 1998). In view of the profound role of historical rainfall data in these models and the vast amounts of data required, a question to be answered is whether rainfall patterns in the recent past are as strong directed by ENSO as observed in earlier times. I.e., when historical rainfall data are stronger correlated with the state of El Niño than recent rainfall data, it is also not likely that El Niño will have a major impact on future rainfall patterns either. Thus, it becomes less suitable for forecasting purposes. Furthermore, it would greatly complicate the technical applicability of the indicator as the number of relevant observations would then become limited to those observed in the time-period of strong impact. The robustness of the statistical analysis would suffer correspondingly.

(3) The nature of weather forecasts currently provided by the Meteorological Department of Zimbabwe does not fully comply with user demands. For regional priority management in malaria control or, as in our case, crop growth modelling, quantified rainfall estimates are required. Preferably at a high spatial detail, rather than the national estimates that are expressed in such terms as " 70% chance of receiving above average rainfall". This confronts us with another question for research: "Are there differences between weather stations in how they are affected by ENSO and does this result in spatially different weather outlooks?" Another question in this context is whether it is possible to produce quantified rainfall forecasts within a reasonable error margin using a model based on the ENSO principle. Although statistical significance, i.e. a comparison of medians, is a useful measure to justify or deny an indicator to be used for modelling, it is not conclusive as such as a component of forecasting. Ultimately, the error observed in forecasts is of greater value since it indicates the amount of variation in rainfall totals that can be explained.

(4) The scale at which crop performance is monitored is fit for the analysis of a region whereas the ENSO rainfall estimation is on a point basis. To justify interpolation of these point estimates, geostatistical analysis of the phenomena involved is required. Only if the significance of the predictor is sufficient and spatially structured, regionalisation of ENSO weather outlooks are justified.

3.2 Methods and materials

This Chapter describes the methods used to test the hypothesis and to answer the research questions stated in the preceding Chapter. Firstly, the study area and the criteria that have been used to select this area are summarized. This is followed by an explanation of the research methods and materials used in this comparative analysis.

3.2.1 Study area

In this sub-section, the area for study is briefly described and an explanation is provided for its selection. Availability of historical weather data and yield statistics were the main criteria for selecting an appropriate study area for this research. Zimbabwe is endowed with a relatively dense and well-maintained network of meteorological stations. Zimbabwe also has a National Early Warning Unit (NEWU) in the department of Agricultural Technical and Extension services (AGRITEX) that collects yield statistics in a relative objective manner since 1989, when the Food and Agricultural Organization (FAO) provided technical assistance in setting up an improved crop monitoring methodology. Therefore, the study area was selected within this country. The study area itself is Mashonaland West Province, located in the northern part of Zimbabwe extending from the communal area of the Zambezi Valley in the north to the (semi) commercial areas in the south.

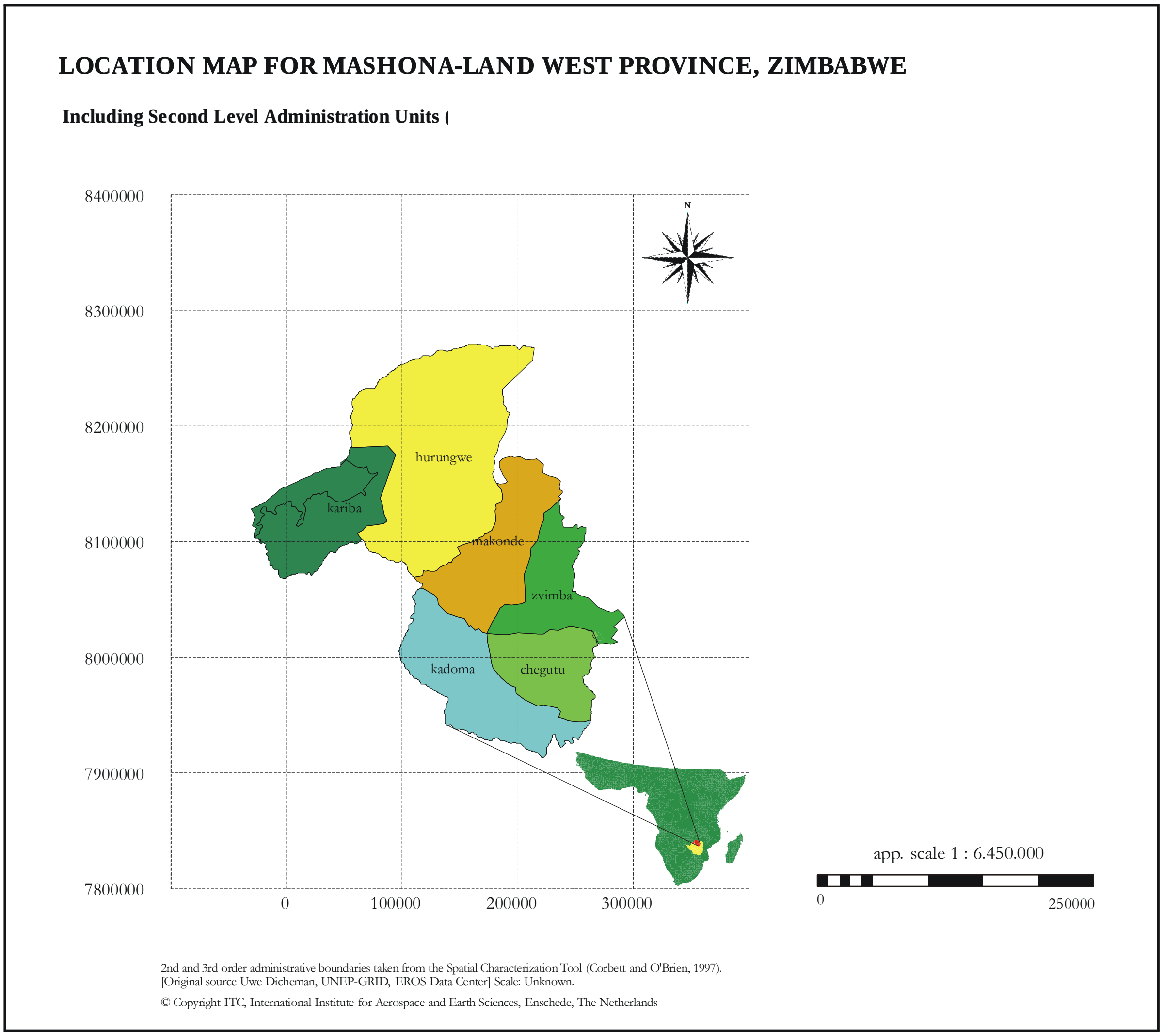


Figure 1, Location Map Study Area

Geographically this area extends from 15º 30' - 18º 50' S, 28º 00' -31º 00' E. Physiographically, most of the area belongs to the Middle-veld relief region with altitudes varying from 450 to 1200 m.

3.2.2 Research methods

In this sub-section, the methods and materials used to test the hypothesis and to answer the research question have been described.

3.2.2.1 Yield forecasting

Since actual land use systems are very complex, modelling them in a realistic manner is not possible or not practicable (Driessen and Konijn, 1992). Therefore, the research reported here addresses singular land-use systems. The following system components are more or less common for both yield-forecast frameworks under comparison.

Basic data input at the first hierarchical level comprises weather data, soil data, farm management data, and crop information. Predefined crop growth simulation and forecasting procedures generate crop yield per land unit. Basic data are stored in a GIS and the structure of this system is presented in Figure 2.

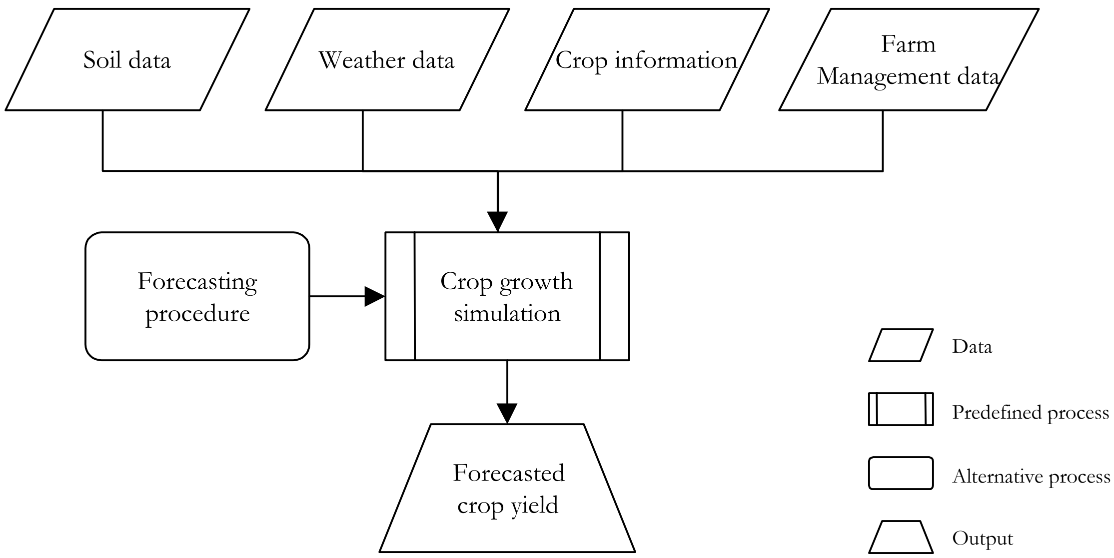


Figure 2, Flowchart General Yield Forecasting Pathway

To improve the objectiveness of the comparison of the two yield-forecasting frameworks discussed in this study, its components are processed in a possibly similar manner. The main difference between the two approaches is in the forecast procedure itself. Therefore, only when system components related to YFF1 and YFF2 differ significantly they are treated separately.

3.2.2.2 YFF1 and YFF2

The following section describes the two yield-forecast frameworks under comparison. Both are based on the CGMS model, developed in the framework of the MARS (Monitoring of Agriculture by Remote Sensing) project of the Joint Research Center (JRC) of the Commission of the European Communities. CGMS is developed around the model WOFOST 6.0, a crop growth model developed by the DLO–Winand Staring Center (SC–DLO) in collaboration with the Research Institute for Agro biology and Soil Fertility (AB-DLO), both located in Wageningen, The Netherlands. WOFOST is a dynamic explanatory model. It simulates crop growth based on different sets of crop parameters at a lower level of integration. It takes account of certain soil characteristics and uses daily meteorological data that can be assumed to be homogeneous within the land mapping unit area. WOFOST could be described as a 'point' model in the sense that it performs calculations for one single point in space/time. WOFOST is incorporated in

CGMS and to allow the application of the model on a larger scale, there is a need for the identification of areas where the meteorological data and the soil characteristics can be assumed to be homogenous.

The CGMS aims to monitor agricultural season conditions and make the yield forecasts at sub-regional, regional, and national level. To arrive at these forecasts, the system combines crop growth simulation with interpolation algorithms, data handling functions and statistical procedures. Operationally, the model is run approximately every ten days to simulate crop yields for several crop varieties in various farming environments, but uses daily meteorological data as input.

The CGMS system is founded on the notion that agricultural production is the result of interaction between weather, crop, soil and farm management. Within a given region, crop and soil characteristics and farm management are assumed relatively constant over years. Variation in agricultural yield can be explained by variation in weather conditions.

CGMS is being used at JRC for regional crop state monitoring and yield forecasting at regional and national scales on individual years. The area for which these calculations are being performed covers all of Europe. The extension beyond the members' states of the EU is a recent one, and most experience has actually been gathered with the operation of the system for the countries of the EU.

The meteorological data (and their pre-processing), the soils database, the crop knowledge and the model themselves are integrated in one single system, composed of 2 main modules:

(1) The module processing daily meteorological data: quality control, formatting and patching of data gaps/missing values; calculation of derived parameters such as solar radiation (from cloud cover or sunshine duration), vapor pressure and potential evapotranspiration; all interpolated to a regular grid.

(2) The agro-meteorological crop growth simulation. Since various soil types coexist in a grid, outputs for a basic square are produced for each of the major soil types and profile available water capacities, to reach a representativity of approx. 80% of the suitable soil coverage.

The following module has been added by the author of this research:

(3) A statistical module, relating the model outputs, through a regression analysis and possibly in combination with a technological time trend function based on historical yield data, to series of observed (sub) regional yields

Module One: weather data

The generation/identification of weather data follows the same procedure for YFF1 and YFF2, with this difference, that YFF2 uses not only available actual meteorological data, but accepts and processes surrogate weather data as well.

Actual meteorological data and their processing: YFF1 and YFF2

Figure 3 gives the network of the major meteorological stations for Mashonaland West Province and neighboring regions from which both the historical and current data are used for interpolation to 50 x 50 km grid cells. Only those stations are depicted from which on a daily basis, on or more of the input parameters required for running the models are available. In practice, it represents the WMO network managed by the Zimbabwean Meteorological Services. In addition, stations recording only rainfall data are included to improve the interpolation of this parameter. Although the number of stations that has data to calculate additional parameters, e.g. radiation, is not evaluated by the model, there is a need for a geographically well-distributed network of these critical weather stations. In total 40 stations are used for crop growth modelling as listed in Appendix I.

Stations recording one or more of the necessary parameters are used, after decoding, filing, quality evaluation, and replacement of missing values for calculation of derived parameters (e.g., potential evapotranspiration and estimated global radiation).

The interpolation of daily meteorological data from the synoptic network of CGMS is largely left unmodified, with only a few adjustments to suit the use of the model to the less favorable conditions as observed in the study area. The adjustment relates to the fact that not all weather stations have complete sets.

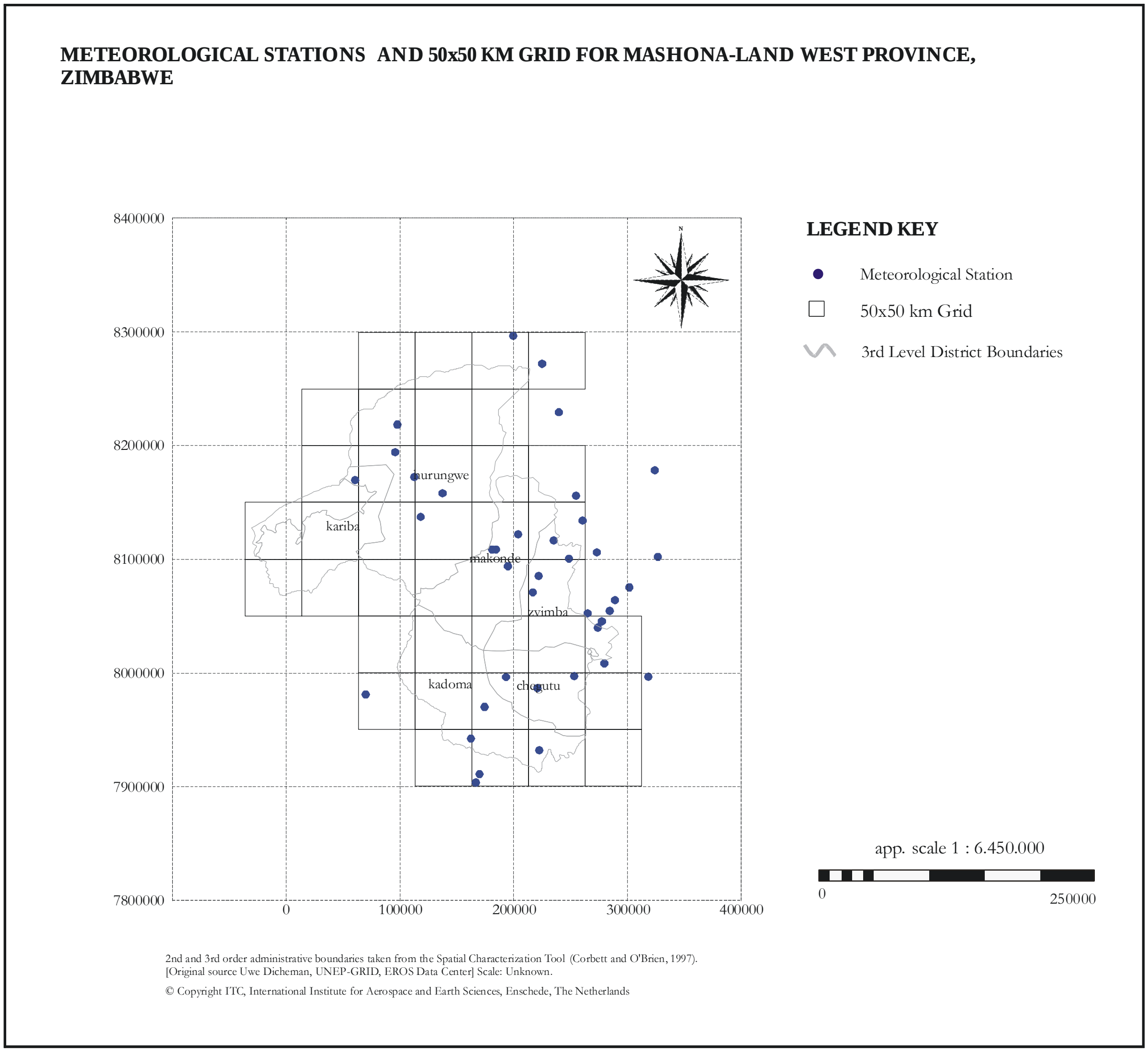


Figure 3, Network of Meteorological Station and 50 X 50 km Grid

Often, days are missing. Some stations have for certain years very few data. These stations should be avoided in further calculations. Therefore, a check is performed on the number of daily values in a year. Each relevant variable is checked against the total number available records in that year. Seven variables are taken into account for this analysis: wind speed at 10 meters height, sunshine duration and cloud cover, vapor pressure, minimum temperature, maximum temperature and rainfall. The variables sunshine and cloud covers are cross-linked. If one of them is available, the other one can be calculated. The first check is based on a classification with respect to the type of data that station can deliver. Three classes are distinguished: rainfall data, temperature data, and all other variables. The second check is based on the temporal availability of the data in these classes. The selection procedure determines for each weather station if the availability for a class of data falls above a certain threshold. If so, then the station will be marked as a valid station for that type of data. The threshold value can be selected per station, but is applied to all three categories. Appendix II gives an example of the scoring of weather stations per grid for this research.

A typical example would be a threshold value of 80%, i.e. if the station data in a particular class are available for more than 80% of the time, the station will be used for the interpolation of the data in that class. The timeframe taken into account for the check is the total number of days in the year for past years (365 days for normal years and 366 days for leap years) and the number of days up to the 'end-of-simulation' day in the current year. This is where the adjustment is required. The threshold determining the exclusion of stations in the interpolation procedure required downward modification to arrive at an acceptable compromise solution, where the quality of interpolation procedure benefits from more stations with less accurate records than very few stations with good records. The modified threshold is set to 50%. Missing daily weather data on stations can be replaced by reference weather. Reference weather is defined as daily long-term average values.

The method for interpolation of the weather data is used to perform a point-to-point interpolation from a geographically irregularly distributed set of weather stations to a regular grid of 50 x 50 km. This is based on a knowledge-based method that consists of selection of the optimum set of at most four stations used in the interpolation. To determine which stations to use, and indeed how many stations to use, a combination score is used. The interpolation is performed using the best possible station configuration for each grid and each year, based on the algorithm developed by Van der Voet et al. (1993). A set of so-called 'best' stations is selected from the set of all weather stations having sufficient data for a given year as explained before. Thus, for different years different sets of stations can be selected for the calculation of the grid weather. The actual interpolation, once this selection has been made, is in fact a simple average for most of the meteorological parameters, corrected for an altitude difference in case of temperature and vapour pressure. The exception is rainfall data, which are not interpolated. The rainfall for a grid is taken from the weather station(s) that is the most similar to the grid center in terms of altitude and distance to the coast. Presence of natural barriers is also evaluated. The other weather variables are used for the interpolation, which is done by averaging the observed daily data from the optimum (set of) weather stations, surrounding the center of the grid cell. The interpolation algorithm used for the estimation of daily weather data on a regular grid is described in Appendix C of the *User Manual for the CGMS Model* (Mahalder and Sharifi, 1998).

Surrogate meteorological data and their processing: YFF2

In the following sub-section the method used to feed the crop growth model with surrogate weather data is detailed. Since no software is available for temporal and spatial characterization of El Niño for rainfall estimation to answer the research questions as stated in the section *Research Questions*, sub-section *YFF2*, the following model was prepared by the author.

Agro meteorological ENSO Rainfall Analysis and Forecast Model

The El Niño/Southern Oscillation (ENSO) Model prepared for this study uses a statistically based approach in producing seasonal weather outlooks based on the assumption that the atmosphere is the best model of itself. The relationship between ocean surface temperature anomalies in the eastern Pacific and the occurrence of rainfall over Zimbabwe is one of them, also referred to as a tele-connection. Teleconnections represent a statistical relationship (i.e. correlation) from which causality cannot necessarily be inferred. In analogue models, the predictand and predictor are determined first, to which a close analogue is sort from historical cases (Unganai, 1998). El Niño is the ocean component and slight anomalies herein

(+/- 1 to 3 °C) can cause significant changes in the Southern Oscillation, which is the atmospheric component.

Since no tight boundaries between different regions exist in the atmosphere, changes in one of them can have noticeable effects in another, even if far apart. However, El Niño does have different impacts in different parts of the world and at different times of the year. During the northern hemisphere winter, El Niño's expected impacts include drought in southern Africa, continuing drought in northern Australia and Indonesia, high rainfall in three continents and unseasonably warm weather in parts of North America and eastern China. Although the ENSO model was prepared specifically for this research, it can be applied to any of the regions depicted in Figure 4.

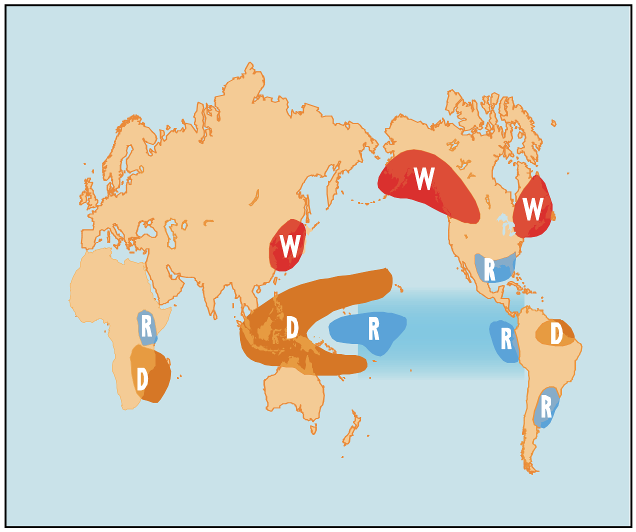


Figure 4, Climatic impacts of warm El Niño events (Oct-Mar) Source: FAO website on El Nino.

The predictor

Literature suggests to use SOI phases to calculate future seasonal rainfall probabilities since it gives a more accurate result than using SOI averages (QDPI, 1995). Therefore, recent trends in the SOI have been used to calculate the probabilities of receiving particular amounts of rainfall over a particular location. A comparison with other monthly SOI values can be found in Allan et al. (1996b).

The original index is calculated as follows (Troup et al., 1965):

$$\mathrm{SOI} = \frac{\mathrm{PA(Tahiti)} - \mathrm{PA(Darwin)}}{\mathrm{St.Dev.Diff}} \times 10 \qquad [1]$$

Where:

PA() is the pressure anomaly = monthly mean minus long - term mean, and;

St.Dev.Diff. is the Standard deviation of the difference.

A SOI of -10 means the SOI is 1 standard deviation on the negative side of the long-term mean for that month. Troup's monthly SOI (from the year 1876 onwards) is derived from normalized Tahiti minus Darwin mean sea level pressure. The SOI Phase is determined by the change in average monthly SOI over the two previous months. The phases of the SOI were defined by Dr Roger Stone, QDPI, who used cluster analysis to group all sequential two-month pairs of the SOI (from 1882 to 1999) into five clusters. The five SOI phases are defined as follows: Phase 1, consistently negative; Phase 2, consistently positive; Phase 3, rapidly falling; Phase 4, rapidly rising; and Phase 5, consistently near zero. The boundaries between phases (clusters) were further developed by Dr Jeff Clewett, QDPI, who plotted the distribution of the clusters and mathematically defined boundary curves to minimize errors. The SOI Phases were kindly provided by Queensland Department of Natural Resources and the Department of Primary Industries, Australia.

The predictand

The predictand is formed by summed rainfall figures that cover part or all of the growing season of a specific summer crop with a limit of six consecutive months. Within this boundary, the period for which rainfall totals are summed, i.e. the forecast period (FP), is dependent on the current date and the latest possible month within a crop specific growing season. This is defined by the computer model, as more rainfall outlooks are possible starting from and within the actual growing season.

Computer Model

The computer model itself is a Microsoft Excel 2000™ Template augmented with functionality and wizards programmed in Microsoft Visual Basic 6™. The added functionality permits in-depth analysis and consists mainly of "SQL like" querying and filter commands and wizards. The "open source" character of the template permits users to understand and question the procedures or to further tailor-make it to answer specific research questions. This is not possible with any of the existing software such as RAINMAN. The user is guided through the steps as outlined in the model structure in (semi) automated and user-friendly way (see Figure 5).

Model Structure

The model exist of the following two main components:

(1) Statistical Component, consisting of:
  i. Data Exploration Tool, and;
  ii. Kruskal-Wallis Rank Test, and;
  iii. Tool to assess usefulness of Kruskal-Wallace results.

(2) Seasonal Weather Forecast Component
  i. Quantified Rainfall Forecast, and;
  ii. Season Similarity Tool.

The Data Exploration Tool was developed mainly to provide information about the populations under comparison, i.e. historical FP rainfall totals grouped by SOI Phase similarity, by means of Box & Whisker Diagrams. The Kruskal-Wallis Rank Test permits a first look at the impact that ENSO has on the province, whereas the strength of the probability to receive a specific amount of rainfall in a FP is more conclusive in this matter. This is provided by the tool under 1.iii. In summary, the Statistical Component facilitates analysis of possible differences in the statistical correlation, in the probability to receive a specific amount of rainfall, and in absolute difference between a typical positive ENSO year and a typical negative ENSO year over different periods in history. Based on these results, periods in the year, i.e. the months, which are well correlated with El Niño, can be isolated.

There are two different model versions. The first model version takes the predictor, i.e. the SOI phases, as they were calculated by Roger Stone and is called "Single Phased". The advantage of defining the predictor is this manner is that the characterization can be more specific. Thus, the odds of being able to forecast more specific are greater. However, this only holds if enough years with rainfall figures are on file. The second model version, called "Double Phased", was introduced since for some rainfall stations less than 40 years of recorded data are available, resulting in cumulative distribution functions (CDF) based on only two or three observations. Statistically, this is insufficient and may result in very inaccurate weather outlooks. Therefore, Phases of similar nature were merged. Phase 1, consistently negative, and Phase 3, rapidly falling were merged to Phase 1/3, negative. Phase 2, consistently positive, and Phase 4, rapidly rising were merged to Phase 2/4, positive and Phase 5, consistently near zero, remained unmodified. **Note that** the predictive capability of this model version may suffer from this generalization.

Once the results of the Statistical Component are analyzed from which the forecast strategy is determined for a specific rainfall station, the actual forecast can be prepared. The Quantified Rainfall Forecast component does this as follows. Firstly, probability distributions, i.e. Cumulative probability Distribution Function (CDF), for each of the candidate FP rainfall totals have to be established for each SOI phase for each of stations in the study area based on the data available. Each of the selected stations, 23 in total, has data stretching for at least 30 years (up to 108 years). When forecasts are made for a given station, the CDF corresponding to the current SOI phase is selected. From this CDF the lower, upper and middle percentile are taken out and used as forecast rainfall totals (see Figure 7). The years that correspond to these figures within a user-defined deviation are identified by the Season Similarity Tool to permit judicious selection of surrogate daily rainfall figures. This functionality is added since a specific season may have an intra-seasonal distributed rainfall that is rare and therefore less suitable to be used as input for crop growth modelling purposes.

The input of the model consists mainly of rainfall data, SOI Phases, a crop calendar, and some user-preferred model options prompted for by the included wizards (see the *User Input* symbols, Figure 5). For one of the stations a detailed example is provided in the next sub-section to illustrate the model. The processes and inputs/outputs of the Agrometeorological ENSO Rainfall Analysis and Forecast Model are described in Figure 5.

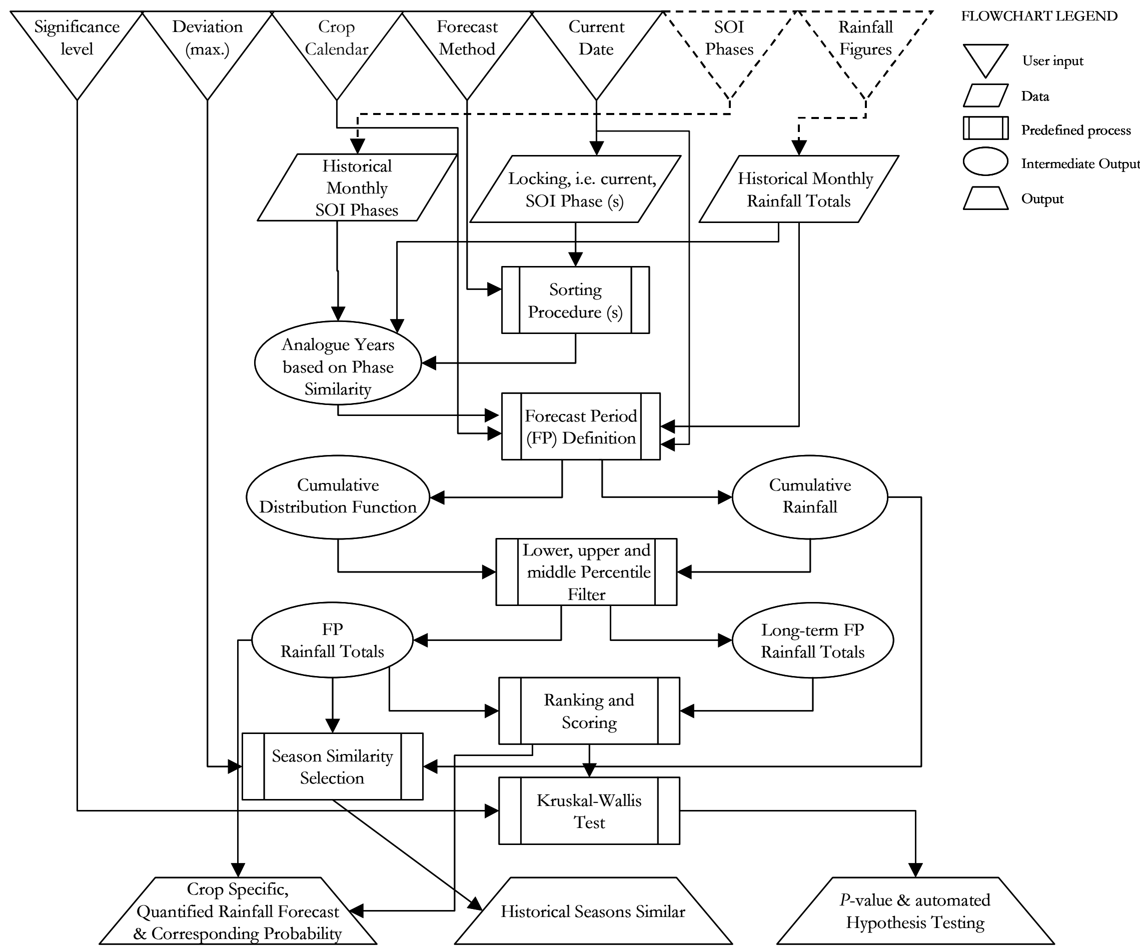


Figure 5, Structure of Agrometeorological ENSO Rainfall Analysis and Forecast Model

The case of Karoi

For the following station an example is worked out in this sub section:

TABLE 3-1

Station Characteristics

Name	Karoi
LAT Degree	-16.500
LONG Degree	29.370
Elevation	1344

Internal #	2
WMO_nr	67765020

Forecast Strategy

The forecast strategy for Karoi was determined as follows. The model was initialized with input data comprising amongst others the month maize generally attains maturity (assumed similar for whole the province in this procedure) and by setting the significance level to a user-defined level for the Automated Hypothesis Test Tool. In our case, this was set to 95% for all stations analyzed. Hereafter, for each of the candidate months preceding their corresponding forecast periods the forecasting strength was identified. This was done for both model versions separately. "The forecasting strength" was defined by the statistical significance of the relationship and the observed long-term difference in FP rainfall totals between a Phase 1(/3) and Phase 2(/4) classified season. Note that if absolute differences are large, chances of having a strong probability to receive a specific amount of rainfall are great as well. In practice, "the forecasting strength" of a specific month is identified by means of Tables 3-2, 3-3, and Figures 6, and 7.

TABLE 3-2

CDFs - Analogue Years Selected for Karoi Based upon Stone's SOI Phase of October

PHASE 1		PHASE 2		PHASE 3		PHASE 4		PHASE 5	
FP_TOTAL	A_YEAR	FP_TOTAL	A_YEAR	FP_TOTAL	A_YEAR	FP_TOTAL	A_YEAR	FP_TOTAL	A_YEAR
401	Nov-94	623	Nov-83	566	Nov-81	440	Nov-86	491	Nov-78
444	Nov-72	666	Nov-74	676	Nov-63	562	Nov-29	639	Nov-67
463	Nov-91	758	Nov-64	728	Nov-41	580	Nov-56	646	Nov-27
506	Nov-82	771	Nov-45	838	Nov-44	618	Nov-50	661	Nov-37
673	Nov-87	800	Nov-28	842	Nov-25	676	Nov-53	676	Nov-58
677	Nov-69	811	Nov-35	927	Nov-92	696	Nov-76	694	Nov-36
685	Nov-40	814	Nov-70	1002	Nov-47	716	Nov-48	719	Nov-90
699	Nov-46	837	Nov-71			739	Nov-30	735	Nov-95
734	Nov-32	867	Nov-89			914	Nov-34	745	Nov-49
761	Nov-93	942	Nov-55			949	Nov-57	754	Nov-59
771	Nov-97	1008	Nov-73			1016	Nov-52	786	Nov-85
798	Nov-65	1016	Nov-98					796	Nov-61
883	Nov-51	1019	Nov-43					832	Nov-79
946	Nov-77	1047	Nov-62					843	Nov-66
1173	Nov-39	1056	Nov-75					855	Nov-68
		1085	Nov-96					862	Nov-26
		1097	Nov-42					863	Nov-31
		1142	Nov-88					905	Nov-33
		1326	Nov-38					918	Nov-84
								934	Nov-80
								1003	Nov-60
								1114	Nov-54

From these CDFs all four percentiles are selected and used for further analysis. An example of such a selection is given in Table 3-3. In addition, absolute differences between a typical positive and a negative season, i.e. Phase 1(/3) and Phase 2(/4), are prepared also.

Percentiles for Karoi for LPM October

PERCENTILE	PHASE	PHASE	PHASE	PHASE	PHASE	Δ
	1	2	3	4	5	1 - 2
q0	401	623	566	440	491	
q1	590	805	702	599	700	215
q2	699	942	838	696	791	243
q3	785	1051	885	827	863	267
q4	1173	1326	1002	1016	1114	

TABLE 3-3

Before further analysis is done, possible outliers in this dataset are identified as follows.

TABLE 3-4

Suspected Outliers for Karoi for LPM October

	PHASE	FP_TOTAL	PHASE	FP_TOTAL	PHASE	FP_TOTAL	PHASE	FP_TOTAL	PHASE	FP_TOTAL
IQR+q3	1	>979.25	2	>1297.4	3	>1067.05	4	>1054.05	5	>1025.65
IQR-q1	1	<395.15	2	<558.95	3	<519.7	4	<371.55	5	<537.475
Outliers (#,value)	1	1173	1	1326	0		0		2	

Remarks: The Inter-quartile Range (IQR) is defined as: IQR= q3 - q1. Suspected outliers are defined as: < 1.5 x IQR - q1 or >1.5 x IQR + q3.

Although several outliers were identified, there is no reason to believe these are the result of measurement or recording errors and are therefore not excluded from further analysis as described below.

When displayed in a Box & Whisker Diagram, a better exploration of the spread within each Phase population and the variability between these populations is possible.

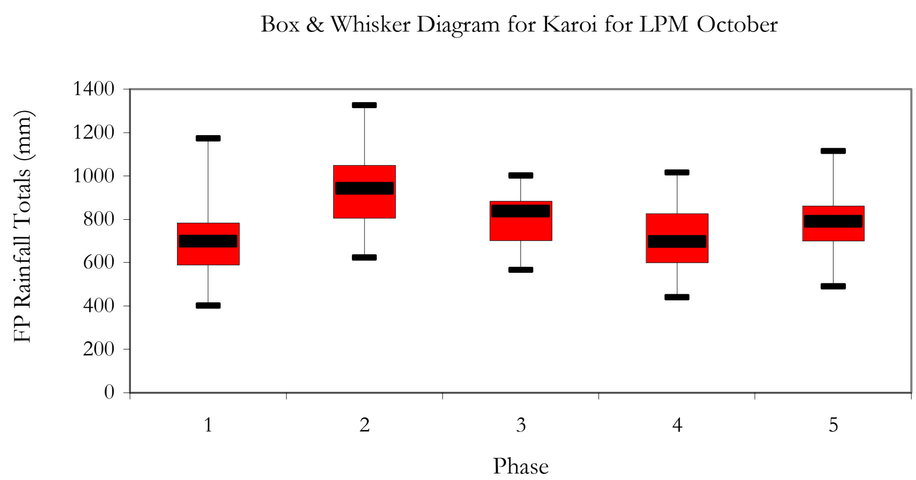


Figure 6, Box & Whisker Diagram for Karoi for LPM October

When displayed in a graph, the upper, lower, and middle percentiles can be identified graphically by arrows. The corresponding figures can then be taken out as forecast figures.

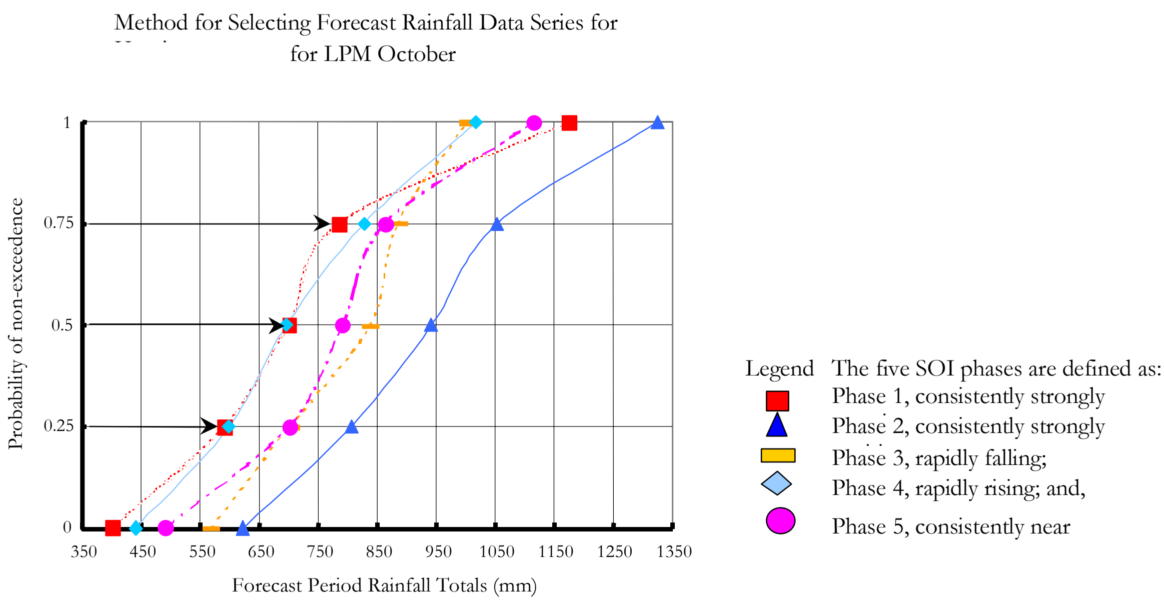


Figure 7, Method for Selecting Forecast Rainfall Data Series

Preferably, each line in the graph displaying the distribution of FP rainfall totals within a population should be vertical. This would indicate that the characterization is specific and that a distinct pattern of rainfall follows a certain Phase. Secondly, the lines should be wide apart indicating that each Phase represents different amounts of rainfall. Also when reading from left to right, the line for Phase 1 should preferably be positioned left, followed by the line of Phase 3, followed by that of Phase 5, followed by that of Phase 4 and the line of Phase 2 should be positioned on the most right side of the graph. A preliminary conclusion from the above two figures is that these requirements are only partially met and that the odds of being able to predict weather with any degree of accuracy is not clear yet. The disadvantage of any graphical method for exploring data sets is that these observations tend to be subjective. Therefore, further analysis is necessary.

To quantify the phenomena a non-parametric test, i.e. Kruskal-Wallis Rank Test, was introduced. The results for this case are presented in Table 3-5.

TABLE 3-5

H-Test (Kruskal-Wallis) for Karoi for LPM October

	PHASE 1	PHASE 2	PHASE 3	PHASE 4	PHASE 5	TOTAL
N	15	19	7	11	22	74
Mean Rank	27.7	51.8	37.1	27.5	37.0	
Degrees of Freedom (df)	4					
Squared Sum of Ranks	172225	968256	67600	91809	660969	
H	13.897					
P-value	0.008	99.2%				
a' specified significance level	0.05					

This data permits us to test the hypothesis as stated under the sub section *Research Questions*.

Hypothesis Testing for Karoi

$\chi^2_{(0.05)}(4) = 9.488$

Now H = 13.897 is greater than 9.488

Therefore, I reject Ho at $\alpha = 0.05$ and conclude there is systematic difference in FP rainfall totals between the SOI phases for this station for the Locking Phase Month (LPM):

October

The high probability of 99.2% leads us to believe that the initial conclusion against the predictive strength of El Niño based on the graphical exploration may be incorrect, or, that at least further analysis is required to be conclusive. **Note that** the disadvantage of a non-parametric test such as the one used here is that only measures of central tendencies of populations, i.e. medians, are compared. The practical value is therefore limited, since forecast figures should be taken from the whole population to be more objective, taking into account the intra-distribution also. Thus, a (even highly) significant statistical correlation does not necessarily imply that the prediction error also is low.

As mentioned before, the probability associated with receiving a specific amount of rainfall is a useful indicator in this matter. A probability is the chance of an event happening expressed as a percentage. A probability of 70% means the event can be expected to occur in 7 out of 10 years. If for a dry year, with Phase 1 (or 3), a low probability of exceeding the long-term medium corresponds and for a typical wet year, with a Phase 2 (or 4), a high probability of exceeding the long-term medium corresponds, forecasts are expected to make more sense than otherwise. For this case, an example is given in Table 3-6.

TABLE 3-6

November Forecast - FP Rainfall Totals in Relation to Long-Term Indicators for Karoi
November to April

LONG-TERM		THIS FORECAST	
INDICATOR	(mm)	Phase 1, consistently strongly negative; Probability of receiving rainfall;	
q0 rainfall FP	401	within 1st quartile	33%
q1 rainfall FP	677	within 2nd quartile	40%
q2 rainfall FP	791	within 3rd quartile	13%
q3 rainfall FP	925	within 4th quartile	13%
q4 rainfall FP	1326	> median	27%

With high probabilities in the first en second quartile and the low probabilities in the third and fourth (the former two amounting to a total of 73%) this seems to be an argument in favor of the predictive strength of ENSO for this station.

For reasons explained earlier in this sub-section, attention was also given to the total rainfall difference (RTD) between a typical negative year, with Phase 1 (or 3), and a typical positive year, with a Phase 2 (or 4). For the locking Phase month (LPM) October, this is listed in Table 3-3. Since no general criteria are available to evaluate these figures, this procedure was followed for all candidate months to permit a relative comparison. Figure 8 displays these averaged absolute differences (q1, q2 and q3) for Phase 1 vs. Phase 2 seasons per LPM. The second y-axes displays the statistical relationship for each of these LPMs against their corresponding FP rainfall totals, which permits further comparative analysis.

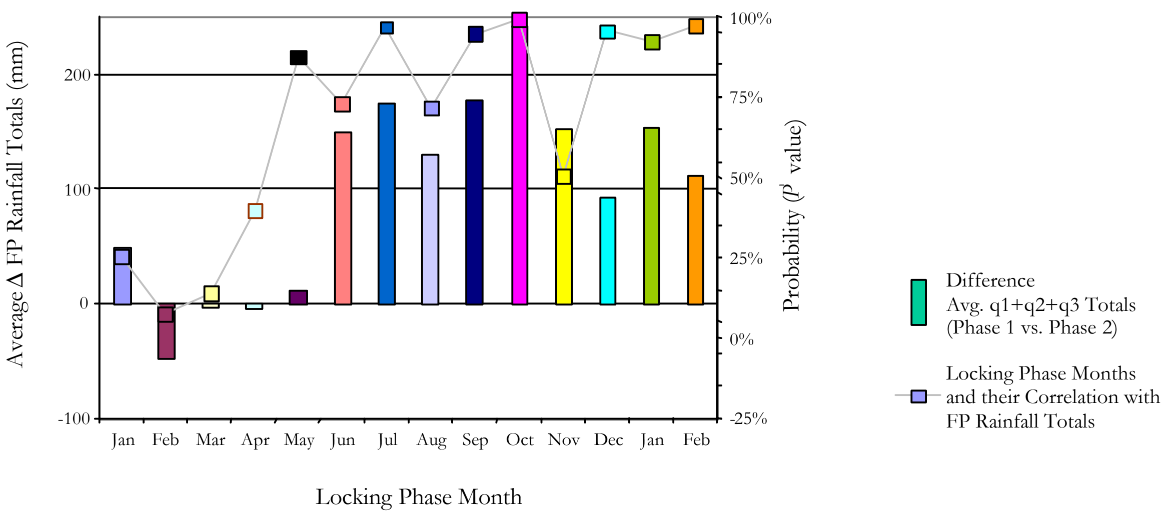


Figure 8, Average FP RTD for Phase 1 and Phase 2 Seasons for Karoi

Clear is now that the LPM October is not only statistically strongly correlated with its FP rainfall totals, i.e. November to March, but that at least relative to the other LPM months, the absolute difference is considerable too. **Note that** the graph also provides a more comprehensive look at the relationship between the two measures. The *P*–value of LPM January is for instance 92.3 - 97.4 = 5.1% lower than that of February, whereas the absolute difference in rainfall totals between a typical negative year vs. a typical positive is 153 –111 = 42 mm higher for LPM January. In these contradictory situations, preference is given to the absolute difference observed in rainfall totals since it implies a stronger characterization and is therefore of more practical value for forecasting purposes. Finally, the performance of the model versions "Single Phase" vs. "Double Phased" was assessed. In addition, it was also assessed whether rainfall patterns in the recent past are as strong influenced by ENSO as in earlier times.

Tables 3-7 and 3-8 and Figure 9 and 10 enable us to do this.

Single Phased Model Version

TABLE 3-7

Locking Phase Months and their Correlation with FP Rainfall Totals for Karoi

MONTH	Single Phase *P*-values (Probability)		
	ALL YEARS	< 1958	> 1958
Jan	25.5%		
Feb	7.2%		
Mar	14.0%		
Apr	39.7%		
May	87.3%		
Jun	73.0%		
Jul	96.6%		
Aug	71.4%		
Sep	94.6%	47%	91%
Oct	99.2%	87%	96%
Nov	50.1%		
Dec	95.5%		
Jan	92.3%		
Feb	97.4%		

This is shown graphically in Figure 9.

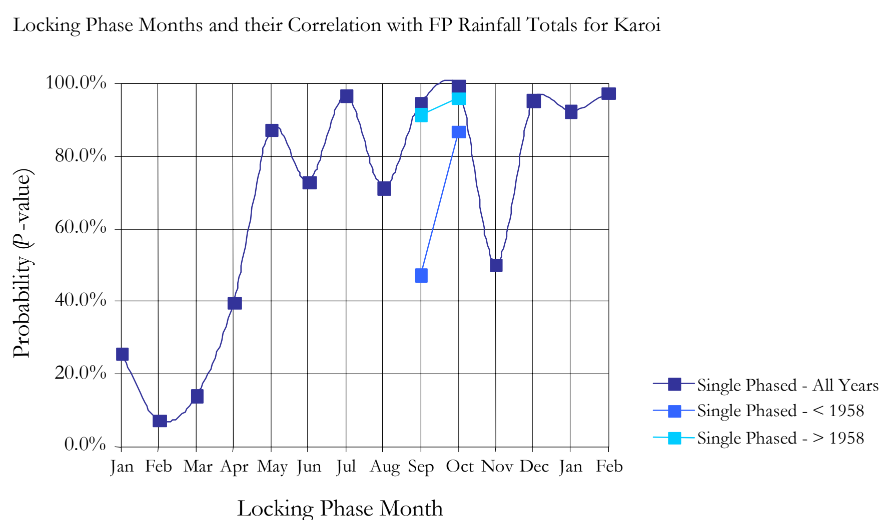


Figure 9, LPMs and their Correlation with FP RT for Karoi – Single Phased

Concluded can be that rainfall patterns in the recent past are directed stronger by ENSO than in earlier times, and hence, that the indicator is expected to be valid in the near future as well. This view is strongly supported by comparison of absolute difference in rainfall totals (Figure 10).

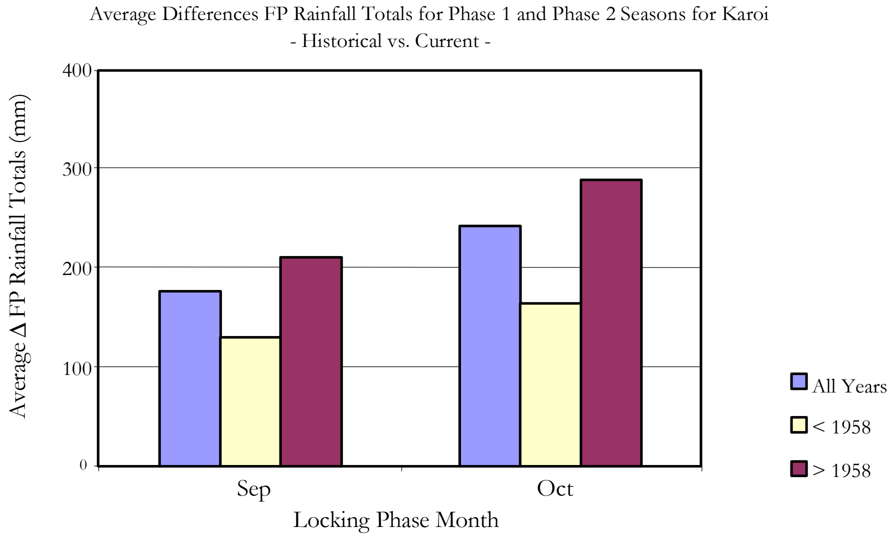


Figure 10, FP RTD for Phase 1 & Phase 2 Seasons for Karoi - Historical vs. Current

Double Phased Model Version

TABLE 3-8

Locking Phase Months and their Correlation with FP Rainfall Totals for Karoi

MONTH	Double Phase *P* -values (Probability)		
	ALL YEARS	< 1958	> 1958
Jan	27.6%		
Feb	0.5%		
Mar	29.1%		
Apr	42.4%		
May	81.9%		
Jun	78.7%		
Jul	99.0%		
Aug	92.1%		
Sep	97.4%	53%	98%
Oct	87.2%	6%	96%
Nov	68.6%		
Dec	99.0%		
Jan	84.9%		
Feb	95.7%		

This is graphically presented in Figure 11.

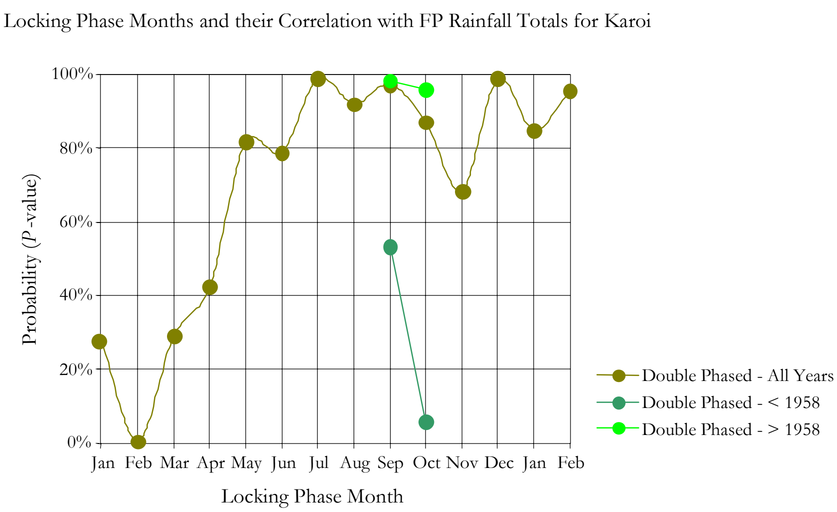


Figure 11, LPMs and their Correlation with FP RTs for Karoi – Double Phased

For the "Double Phased" model version can also be concluded that rainfall patterns in the recent past are more strongly affected by ENSO than in earlier times, and hence, that the indicator is expected to be valid in the near future as well. The same holds for the absolute difference observed in rainfall totals before 1958 and after 1958 (Figure 12).

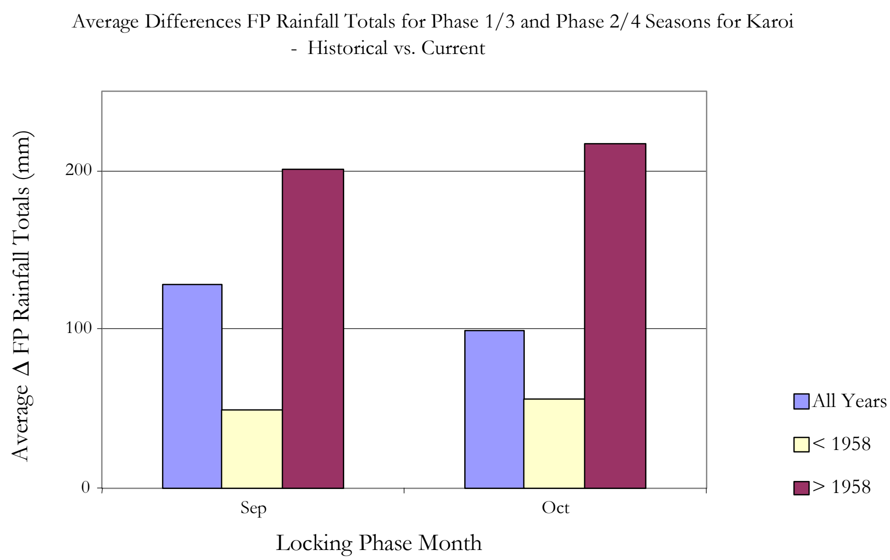


Figure 12, FP RTD for Phase 1/3 and Phase 2/4 Seasons for Karoi - Historical vs. Current

Single versus Double Phased Model Version

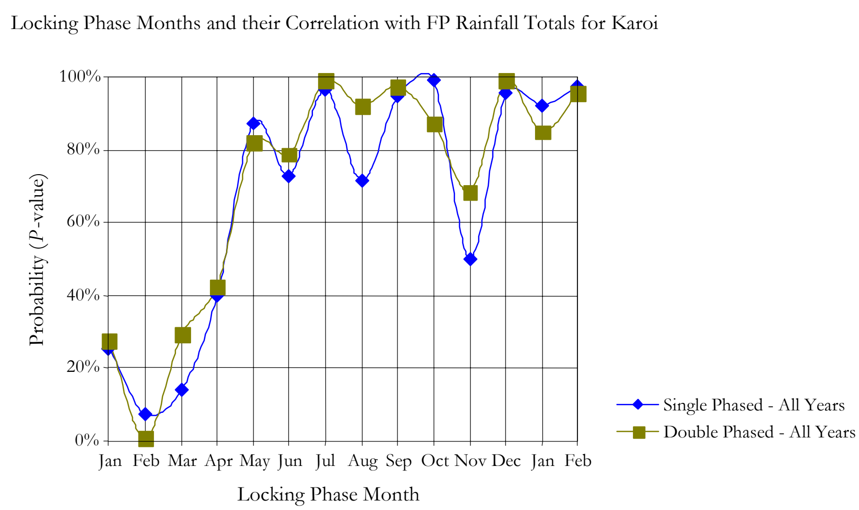


Figure 13, Correlation with FP RTs for Karoi – Single vs. Double Phased

Concluded can be that model version "Single Phased" yields better results than model version "Double Phased". This view is confirmed by comparison of absolute difference in rainfall totals (Figure 14).

Average Differences FP Rainfall Totals for Karoi
Comparison of Single vs. Double Phased model Versions

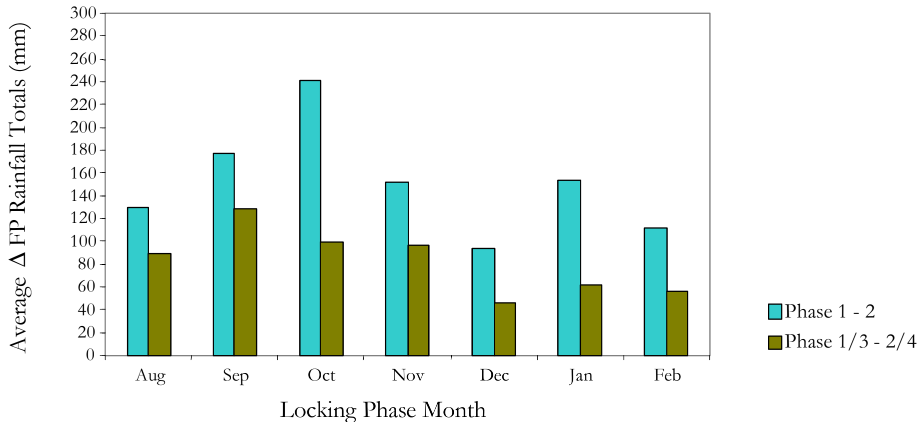


Figure 14, FP RTDs for Karoi - Comparison of Single vs. Double Phased Model

Forecast Strategy Conclusions

In conclusion, it has been shown that to define the forecast strategy for a specific weather station a well balanced, combined look at the following elements was required:

(1) The statistical significance, as preliminary, semi-quantified exploration of the influence of ENSO. The statistical significance explains the relative difference between the various populations, i.e. the ENSO Phases and their corresponding rainfall totals, without explaining the direction and extend of the observed differences.

(2) The qualitative Phase distributions. The direction and extend of these distributions can be graphical displayed by plotting the cumulative distribution functions.

(3) The quantitative Phase distributions. The direction and extend of these distributions can be quantified by taking the absolute rainfall difference between typical positive and negative ENSO years, and, by analysis of the probability to receive an ENSO based rainfall forecast relative to station-specific long-term trend indicators (below or above median).

To ascertain the relevance of ENSO on future rainfall patterns to some extend, historical trends in the phenomenon were compared to recent trends, i.e. the impact of ENSO before 1958 vs. after 1958. As mentioned earlier, for this particular weather station it is now clear that rainfall patterns in the recent past are affected more strongly by ENSO than in earlier times, and hence, that the indicator is expected to be valid in the near future as well. In addition, the strength of the correlation of model version "Double Phased" before 1958 is so low that observations from this period should be excluded from the actual forecasting process. Compared to the "Single Phased" model version this is a clear disadvantage, since the remaining observations may not be representative of the event. As far as the predictive strength of the two is concerned, the analysis made clear that model version "Single Phased" yields better results than model version "Double Phased". Absolute difference in rainfall totals was of high importance in this matter for reasons explained earlier.

Using the same criteria to identify temporal variability in ENSOs predictive strength, concluded can be that the most useful relationship is found between its pattern around October and January and the following rainfall forecast period. LPM September is also relatively strong but is out-performed by October. October is convenient because at that time it is generally early enough to adapt management decisions such as whether, when and how dense to plant and what fertilizer to apply, etc. The February forecast permits to account for any changes in the predictor. Hence, intra-seasonal dynamics can be modeled to provide updated forecast as the season progresses.

Quantified Rainfall Forecast

Surrogate rainfall data assessment commences on 1 November using the SOI phase of the preceding month, i.e. October, to decide which past climate analogue to use as surrogate weather data. Only when the SOI phase has changed, a new weather data analogue is selected; rainfall predictions are then repeated for each of the proceeding LPM months identified during the analysis (in this case only February). Earlier predicted weather data are replaced with actual data up to the current date. In this manner, the quality of the predictions during the season is expected to increase in accuracy. In our case, the forecasts were made on November based on Phase 1, consistently strongly negative. As mentioned earlier, the October forecast was skipped since the analysis revealed that this LPM is slightly less predictive. Table 3-9 and 3-10 below list the forecasts in more detail.

TABLE 3-9

November Forecast - Analogue Years and Rainfall Totals for Karoi
November to April

LOW		MIDDLE		HIGH		(AVERAGE)
FP Rainfall Total	Based on Year	FP Rainfall Total	Based on Year	FP Rainfall Total	Based on Year	FP Rainfall Total
590	1982	699	1946	785	1997	691
Season Similarity figures below are based on a deviation of 10 mm						
580	1956	694	1936	786	1985	(ACTUAL)
		696	1976			FP Rainfall Total
						771

The mean of the low, middle, and high rainfall forecast based on the state of ENSO in October is 691 – 771 = 80 mm of from the actual total that was observed. This is equivalent to a relative error of 11.6 %. In the course of the crop season, the state of ENSO altered from Phase 1, consistently negative, to Phase 3, rapidly falling, which demanded for an update in February.

TABLE 3-10

February Forecast - Analogue Years and Rainfall Totals for Karoi
February to April

LOW		MIDDLE		HIGH		(AVERAGE)
FP Rainfall Total	Based on Year	FP Rainfall Total	Based on Year	FP Rainfall Total	Based on Year	FP Rainfall Total
272	1998	333	1971	400	1955	335
Season Similarity figures below are based on a deviation of 5 mm						
270	1947	329	1931	397	1980	(ACTUAL)
271	1988	335	1963	396	1981	FP Rainfall Total
		334	1969	403	1990	270

The mean of the low, middle, and high rainfall forecast based on the state of ENSO in January is 335 – 270 = 65 mm of from the actual total. This is equivalent to a relative error of 19.4 %.

Note that the above described analysis and forecasting was done for 23 rainfall recording meteorological stations in and near the province, resulting in two or three forecasts per year for the period 1991/92 to 1998/1999. In addition, 5 of these 23 analyses also investigated future relevance of the indicator; historical trends in the phenomena were compared to recent trends, i.e. the impact of ENSO before 1958 and after 1958. This resulted in app. 23 x 3 = 69 statistical analyses and in 23 x 3 x 9 = 621 pre-forecasts to define the station-specific forecast strategies. This provided data to be conclusive about the strength of ENSO as indicator for seasonal weather forecasting and to permit further spatial analysis as described in the following sub-section.

Spatial analysis of ENSO

The spatial scale at which crop performance is monitored and forecasted is regional, whereas the ENSO rainfall estimation is performed on point basis. Therefore, to justify interpolation of these estimates geo-statistical analysis of the phenomena is required first. I.e., only when spatially evenly or structurally distributed regionalization of the ENSO based rainfall outlooks is permitted. This is assessed by means of map analysis. The maps have been produced with the Integrated Land and Water Information System

(ILWIS) GIS using the moving surface method for interpolating point data. Moving surface calculates a pixel value by fitting a surface for each output pixel through weighted point values (ITC, 1998). A 2nd degree function was used to calculate the surfaces since this is advanced enough and yielded the best results. The weight factors for the points are calculated by the linear weight function since ENSO is rather a regional than a local phenomenon. The inverse distance function was rejected since it assigns relatively larger weights to points close to an output pixel and this would over-estimate the level of detail in the phenomenon. Furthermore, the weight functions are implemented in such a way that points farther away from an output pixel than a user-defined limiting distance, obtain weight zero; this speeds up the calculation and prevents artifacts. In view of the same regional aspect, the value for the limiting distance is set to a tolerant value, 185 km for all maps. All maps are prepared in the UTM Coordinate System for Zone 36, since this projection permits relatively accurate area and distance estimation for Zimbabwe.

Again, statistical significance and an analysis of the probability to receive relative high rainfall totals in typical positive ENSO years vs. relative low rainfall totals in typical negative ENSO years were used as performance indicators. Model precision, stability, and the trustworthiness of the forecasts can also be evaluated with the technique of independent estimates. Therefore, the long-term forecast error is also introduced as measure of the performance. This is done at province level for all possible forecasts identified during the *Forecast Strategy*, i.e. the November forecast and the February forecast.

November Forecast

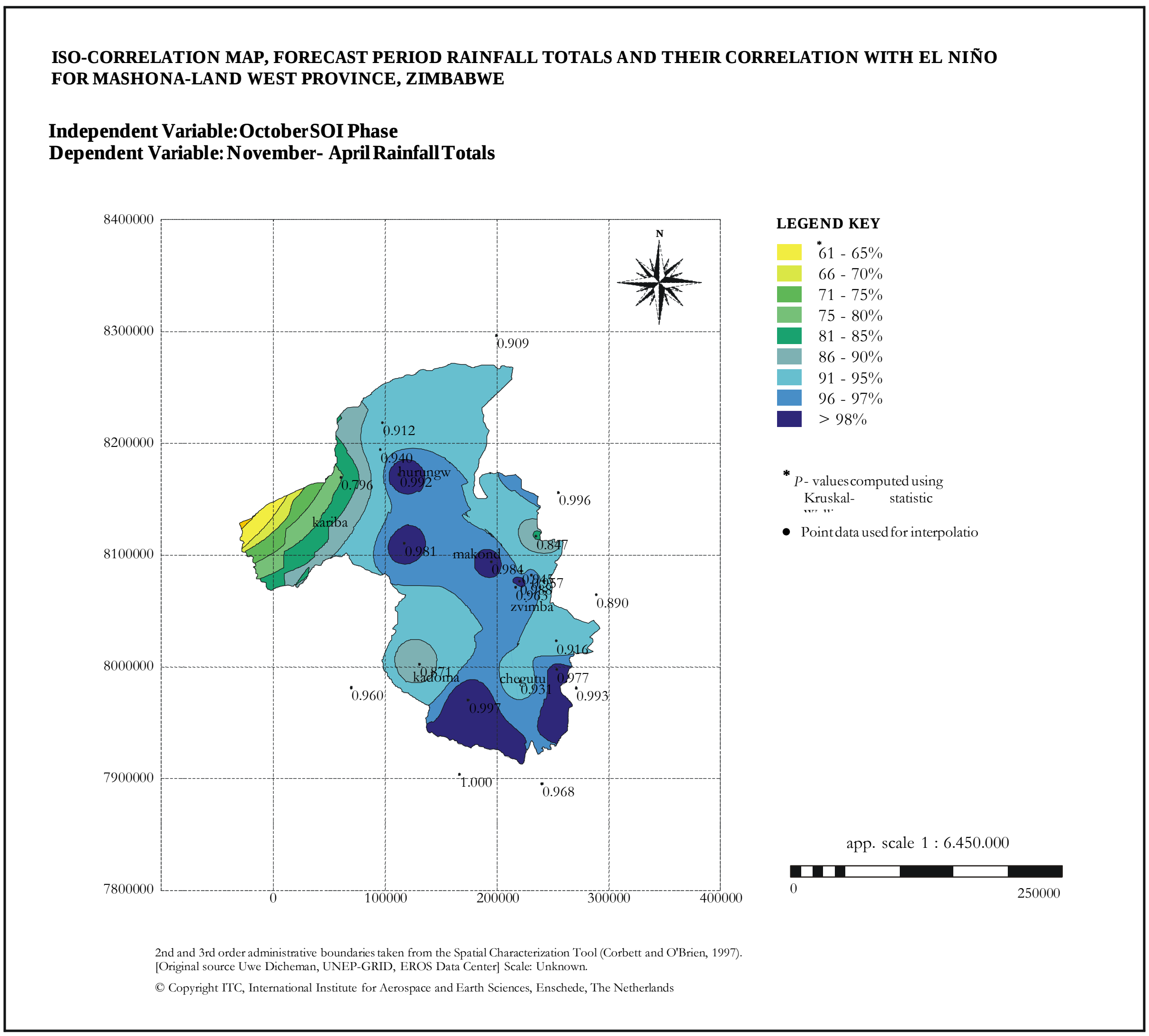


Figure 15, Iso-Correlation Map for LPM October

Probabilities, i.e. *P*-values computed using Kruskal-Wallis statistics, are given for LPM October in Figure 15. This type of maps will be referred to as Iso-Correlation Map. Polygons having more than 91% probability make up more than 80% of the total province for this LPM. Table 3-11 lists the map statistics in more detail.

TABLE 3-11

Map Statistics for Iso-Correlation Map for LPM October

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 50%	0	0%
51 - 55%	0	0%
56 - 60%	42674.25	0%
61 - 65%	585620.11	1%
66 - 70%	570276.62	1%
71 - 75%	1094834.69	2%
75 - 80%	1591349.17	3%
81 - 85%	2333933.18	4%
86 - 90%	4433338.91	8%
91 - 95%	23352777.2	41%
96 - 97%	15472353.1	27%
> 98%	7023283.92	12%
	56500441.15	100%

Note that areas lying in the upper-left corner of the province, i.e. north-western part of the district Kariba
covering mainly Lake Kariba, should be excluded from this and any of the following map analyses since the artifacts observed here are the result of an error in the interpolation due to lack of meteorological stations. The areas outside this region, having a probability lower than 91% are few, account for only one or two percent of the total.

The areas having more than 91% probability consist of three classes: 91 – 95%, and 96 – 97%, and >98%. Covering more than 80% of the province, these three classes could provide us with more information on the intra-distribution. However, from the number of unique island polygons can be seen that slight differences in the point data may be overvalued if not recurrent and, hence, no specific intra-distribution or pattern may be identifiable. If so, a weighted average of these three classes characterizes the situation better. This would amount to a probability of 95.1%, which implies that for more than 80% of Mashonaland West Province the following hypothesis as stated under the sub section *Research Questions,* can be answered positively.

Hypothesis Testing for LPM October

$\chi^2_{(0.05)}$ (4) = 9.448

Now H = 9.533 is greater than 9.448

Therefore, I reject Ho at a' = 0.05 and conclude there is systematic difference in FP rainfall totals between the SOI phases for this area for the Locking Phase Month (LPM):

October

Based on the Iso-correlation map for LPM October, a preliminary conclusion would be that regionalization of the rainfall forecasts based on this LPM is permitted and seems useful.

The probability map (Figure 16) depicts areas having similar likelihood of receiving below median rainfall totals in November to April for the same LPM.

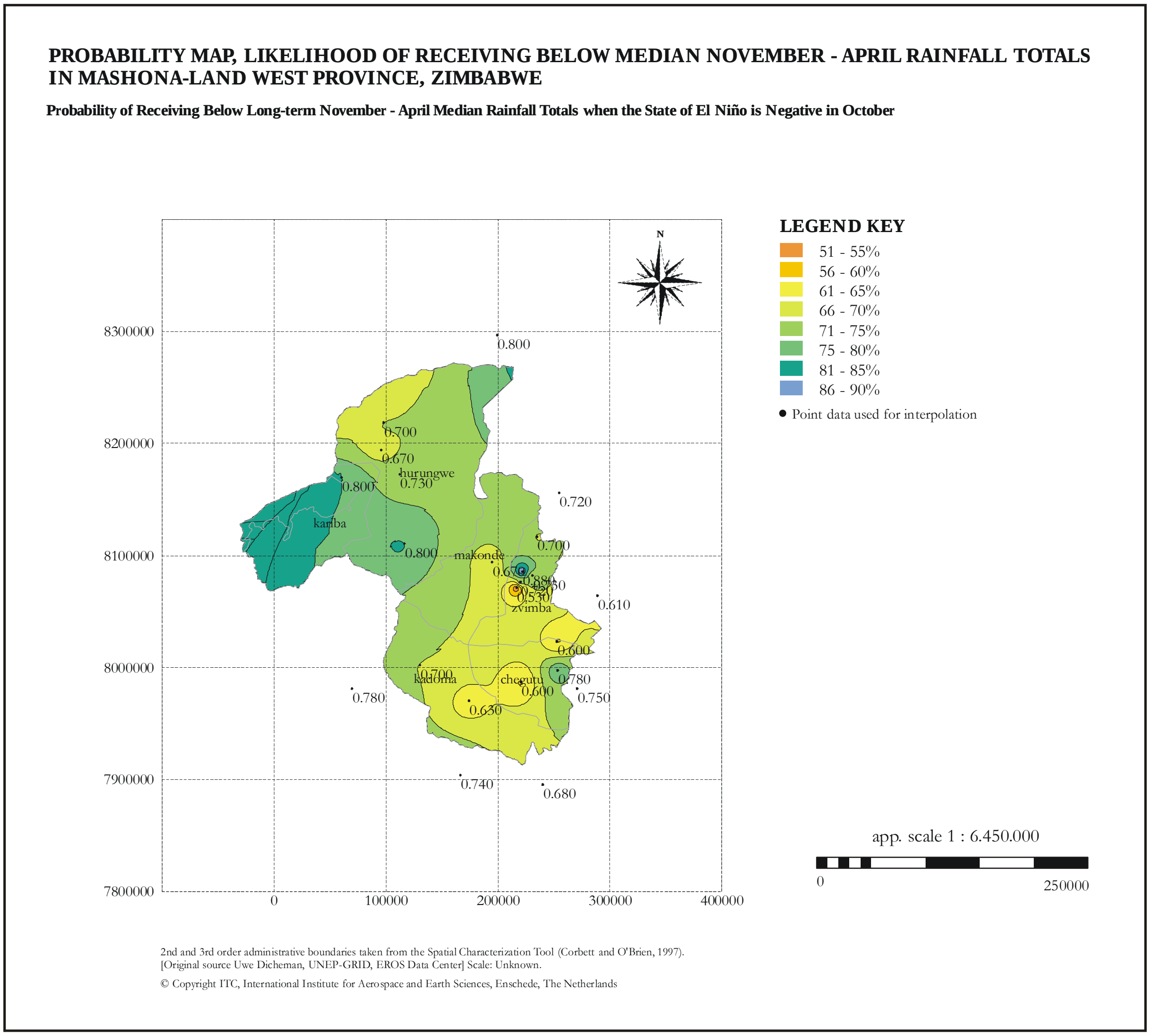


Figure 16, Probability Map - Below Median RTs in Nov - Apr for LPM October

Again, from the number of unique island polygons can be seen that slight difference in the point data may be overvalued if not recurrent and, hence, no specific inner-distribution or pattern could be identified. A weighted average of the map classes may characterize the likelihood of receiving below median rainfall totals in November to April better (based on the state of ENSO in October). This has been computed from the map statistics given in Table 3-12.

TABLE 3-12

Map Statistics for Probability Map for LPM October
Likelihood of receiving below median rainfall totals

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 50%	0	0.00%
51 - 55%	15346.1	0.03%
56 - 60%	90804.19	0.16%
61 - 65%	3344495.44	5.92%
66 - 70%	16231498.2	28.73%
71 - 75%	22279598	39.43%
76 - 80%	8769472.89	15.52%
81 - 85%	5744351.12	10.17%
86 - 90%	24728.63	0.04%
> 91%	0	0.00%
	56500294.57	100%

From the weighted average of 72.7% can be concluded that roughly 7 out of 10 times below median rainfall can be expected. Although this is not an extremely strong chance, it is reasonable. In this respect, the likelihood of receiving above median rainfall totals in November to April should be evaluated also.

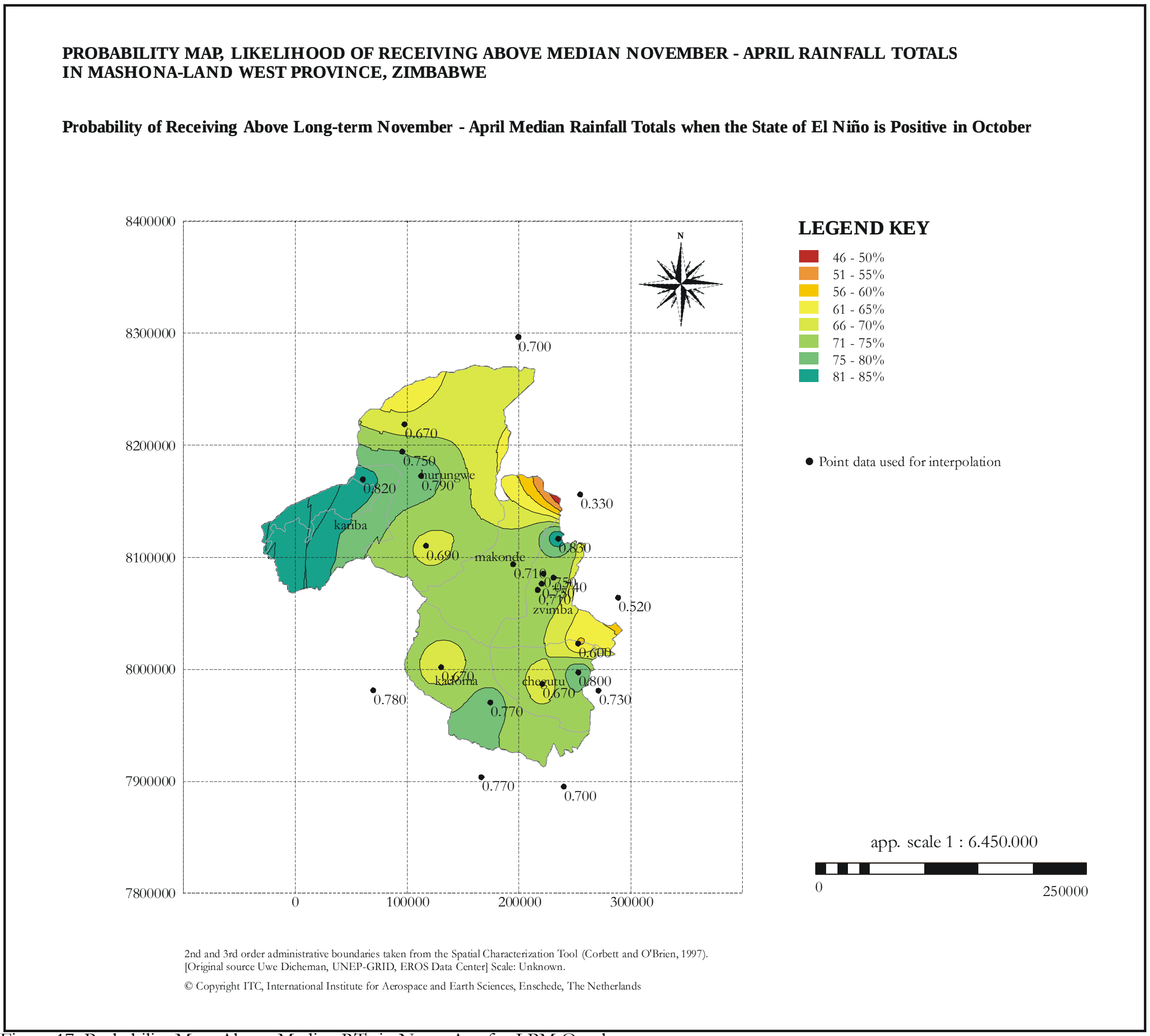


Figure 17, Probability Map, Above Median RTs in Nov - Apr for LPM October

In this case, a weighted average of the map classes could also characterize the likelihood of receiving above median rainfall totals in November to April better (based on the state of ENSO in October). This can be computed from the map statistics given in Table 3-13.

TABLE 3-13

Map Statistics for Probability Map for LPM October
- likelihood of receiving above median rainfall totals -

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 45%	0	0.00%
46 - 50%	38051.41	0.07%
51 - 55%	216433.37	0.38%
56 - 60%	466615.46	0.83%
61 - 65%	3067942.87	5.43%
66 - 70%	13078131.7	23.15%
71 - 75%	26395322.1	46.72%
76 - 80%	7409734.1	13.11%
81 - 85%	5828765.59	10.32%
86 - 90%	0	0.00%
91 - 95%	57.3	0.0001%
96 - 100%	0	0%
	56501053.9	100%

From the weighted average of 72.8% can be concluded that roughly 7 out of 10 times above median rainfalls can be expected. Although this is not a very strong likelihood, it is still reasonable.

Finally, the relative error in the rainfall forecast is averaged over the seasons 90/91 to 98/99 as indicator of the long-term accuracy of the system at provincial scale. **Note that** this includes extreme weather conditions such as the droughts of 1991/92 and 1994/95 as well as extreme wet years such as 1997/98.

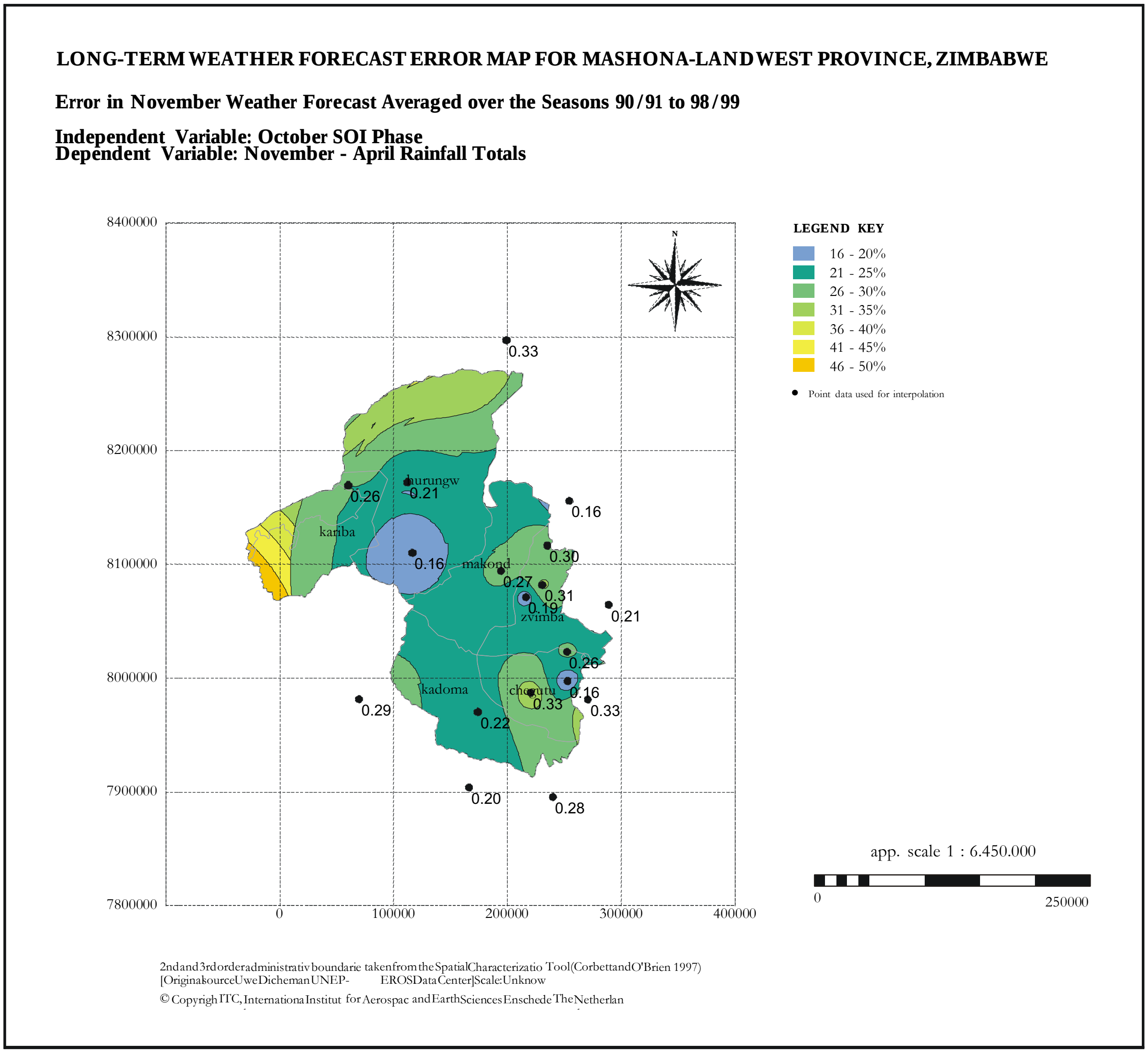


Figure 18, Relative Error in November Forecast

In this case, a weighted average of the map classes could characterize the relative error in the November rainfall forecast also better (based on the state of ENSO in October). This can be computed from the [1]map statistics given in Table 3-14.

TABLE 3-14

Map Statistics for Forecast Error Map for LPM October

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 15%	0	0%
16 - 20%	4261415.98	7.5%
21 - 25%	28236146.3	50.0%
26 - 30%	16411609.7	29.0%
31 - 35%	5309678.44	9.4%
	54218850.42	96%

[1] **Note that** areas lying in the upper-left corner of the province are excluded from this map analyses since the artifiactsobserved here are the result of an error in the interpolation due to a lack of meteorological stations.

From the weighted average of 25.8% can be concluded that roughly 75% of the variability in November to April rainfall totals can be explained when analyzed over a period of nine years.

Conclusion November Forecast

Taking into account the conclusions of the other map analyses described above we may state that regionalization of the rainfall forecasts based on the state of El Niño in October is permitted. Moreover, use of rainfall outlooks with this accuracy is expected to prove useful for maize yield forecasting. However, from the number of unique island polygons and the fact that they differ for each of the three indicators investigated, we can conclude that slight differences in the point data should not be overvalued. Hence, no specific structural regional differences can be identified within the region and a weighted average of map classes proved sufficient.

February Forecast

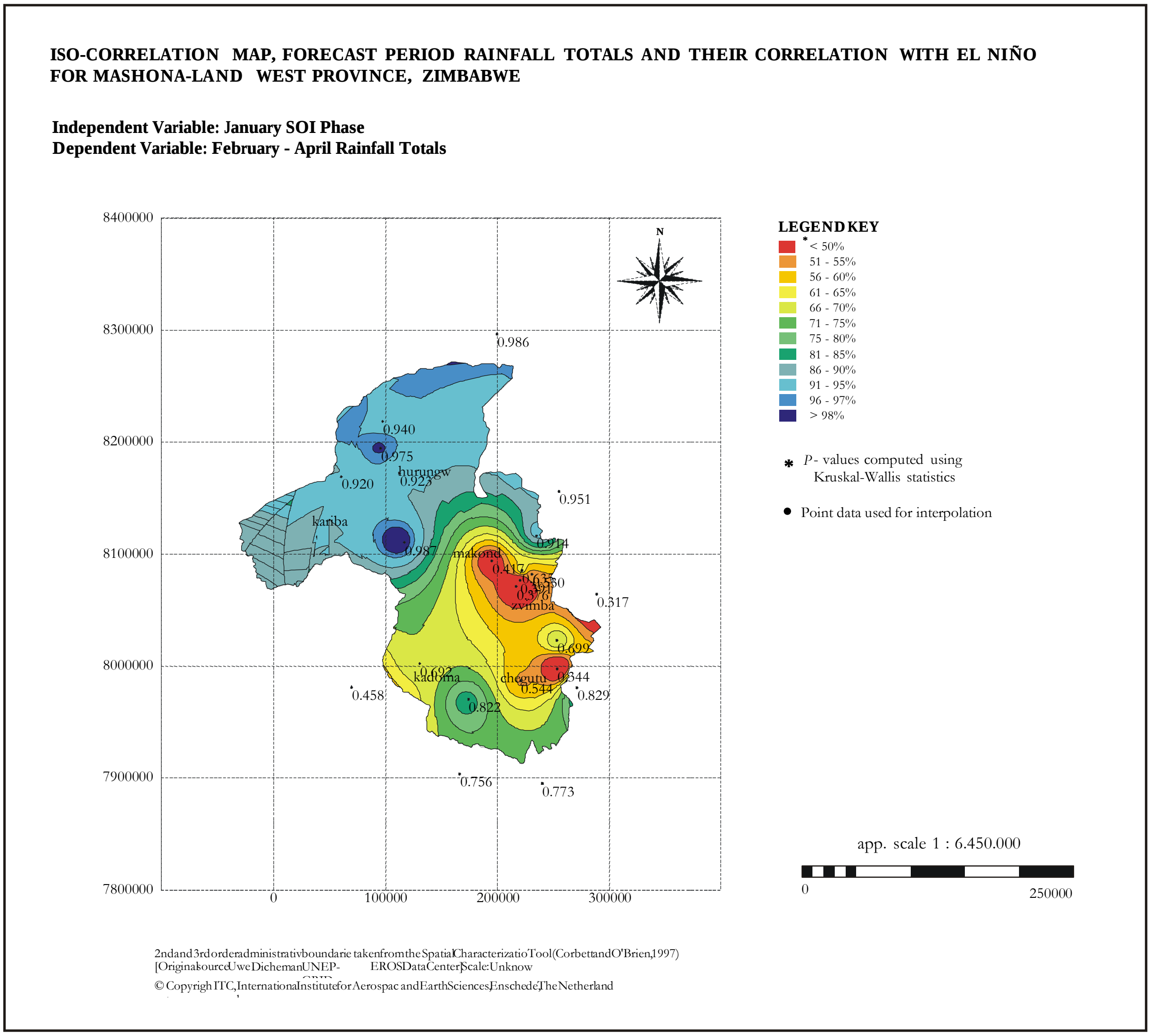


Figure 19, Iso-Correlation Map for LPM January

In Figure 19 the probabilities, i.e. *P*-values computed using Kruskal-Wallis statistics, are given for LPM January.

Table 3-15 lists the map statistics in more detail.

TABLE 3-15

Map Statistics for Iso-Correlation Map for LPM January

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 45%	1949258.89	3.4%
46 - 50%	2621613.11	4.6%
51 - 55%	3947827.97	7.0%
56 - 60%	3804247.96	6.7%
61 - 65%	6592668.5	11.7%
66 - 70%	5646034.86	10.0%
71 - 75%	2745172.69	4.9%
76 - 80%	2176696.81	3.9%
81 - 85%	7560900.71	13.4%
86 - 90%	15334433.8	27.1%
91 - 95%	3571342.33	6.3%
96 - 100%	549531.72	1.0%
	56499729.35	100%

Again, from the number of unique island polygons may be concluded that slight differences in the point data are overvalued. Especially if non-recurrent in the indicators assessed below, no specific intra-distribution or spatial pattern can be identified. A weighted average of these classes characterizes the situation better. This amounts to a probability of 78.7%, which implies that for Mashonaland West Province the hypothesis as stated under the sub section *Research Questions,* should be answered negatively.

Hypothesis Testing for LPM January

$\chi^2_{(0.05)}$ (4) = 9.448

Now H = 5.821 is greater than 9.448

Therefore, I reject H_o at $\alpha = 0.05$ and conclude there is systematic difference in FP rainfall totals between the SOI phases for this area for the Locking Phase Month (LPM):

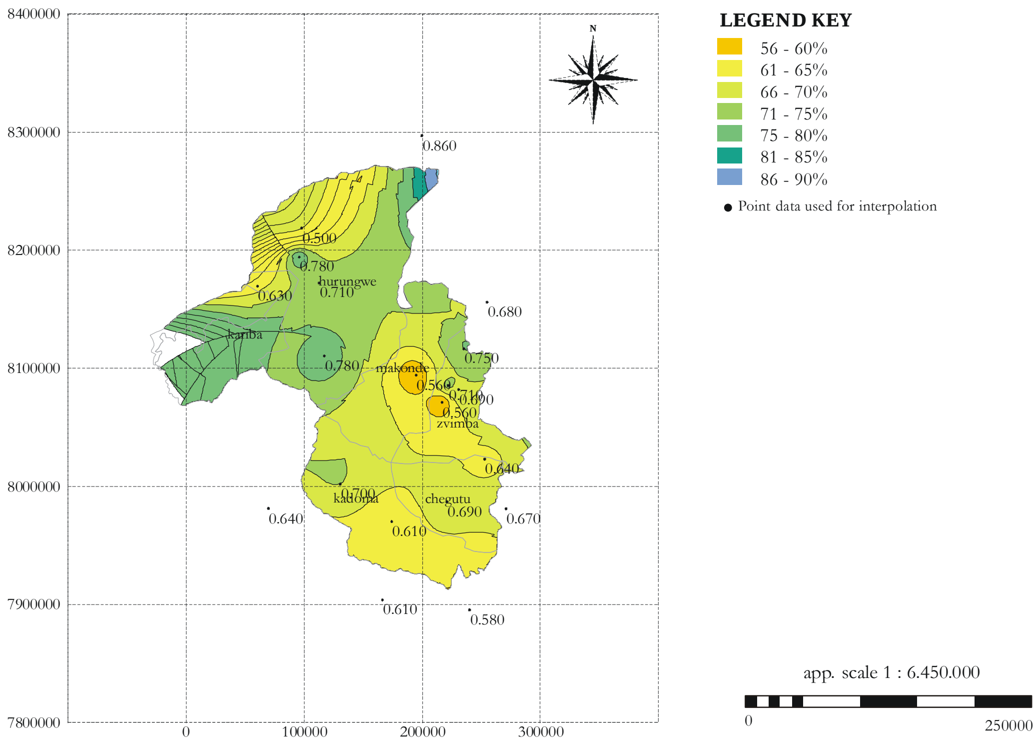
January

Based on the Iso-correlation map for LPM January, a preliminary conclusion would be that regionalization of the rainfall forecasts based on this LPM is permitted but may not prove useful considering the relatively low weighted average probability.

Below the probability map depicts areas having similar likelihood of receiving below median rainfall totals in February to April for the same LPM.

PROBABILITY MAP, LIKELIHOOD OF RECEIVING BELOW MEDIAN FEBRUARY - APRIL RAINFALL TOTALS IN MASHONA-LAND WEST PROVINCE, ZIMBABWE

Probability of Receiving Below Long-term February - April Median Rainfall Totals when the State of El Niño is Negative in January



2nd and 3rd order administrative boundaries taken from the Spatial Characterization Tool (Corbett and O'Brien, 1997).
[Original source Uwe Dicheman, UNEP-GRID, EROS Data Center] Scale: Unknown.

© Copyright ITC, International Institute for Aerospace and Earth Sciences, Enschede, The Netherlands

Figure 20, Probability Map, Below Median RTs in Feb - Apr for LPM January

A weighted average of the map classes may characterize the likelihood of receiving below median rainfall totals in February to April better (based on the state of ENSO in January). This can be computed from the map statistics given in Table 3-16.

TABLE 3-16

Map Statistics for Probability Map for LPM January
Likelihood of receiving below median rainfall totals

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 35%	0	0%
36 - 40%	89.16	0.0002%
41 - 45%	265.66	0.0005%
46 - 50%	2740.57	0.0049%
51 - 55%	0	0%
56 - 60%	814415.79	1.47%
61 - 65%	14015797.3	25.23%
66 - 70%	19007546	34.22%
71 - 75%	13989392.8	25.18%
76 - 80%	7302876.38	13.15%
81 - 85%	256849.29	0.46%
86 - 90%	157513.46	0.28%
91 - 95%	10897.16	0.0196%
> 96%	0	0%
	55547486.41	100%

From the weighted average of 69.3% can be concluded that roughly 7 out of 10 times below median rainfall can be expected. Although this is not a very strong likelihood, it is still reasonable. In this respect, the likelihood of receiving above median rainfall totals in February to April should be evaluated also (Figure 21).

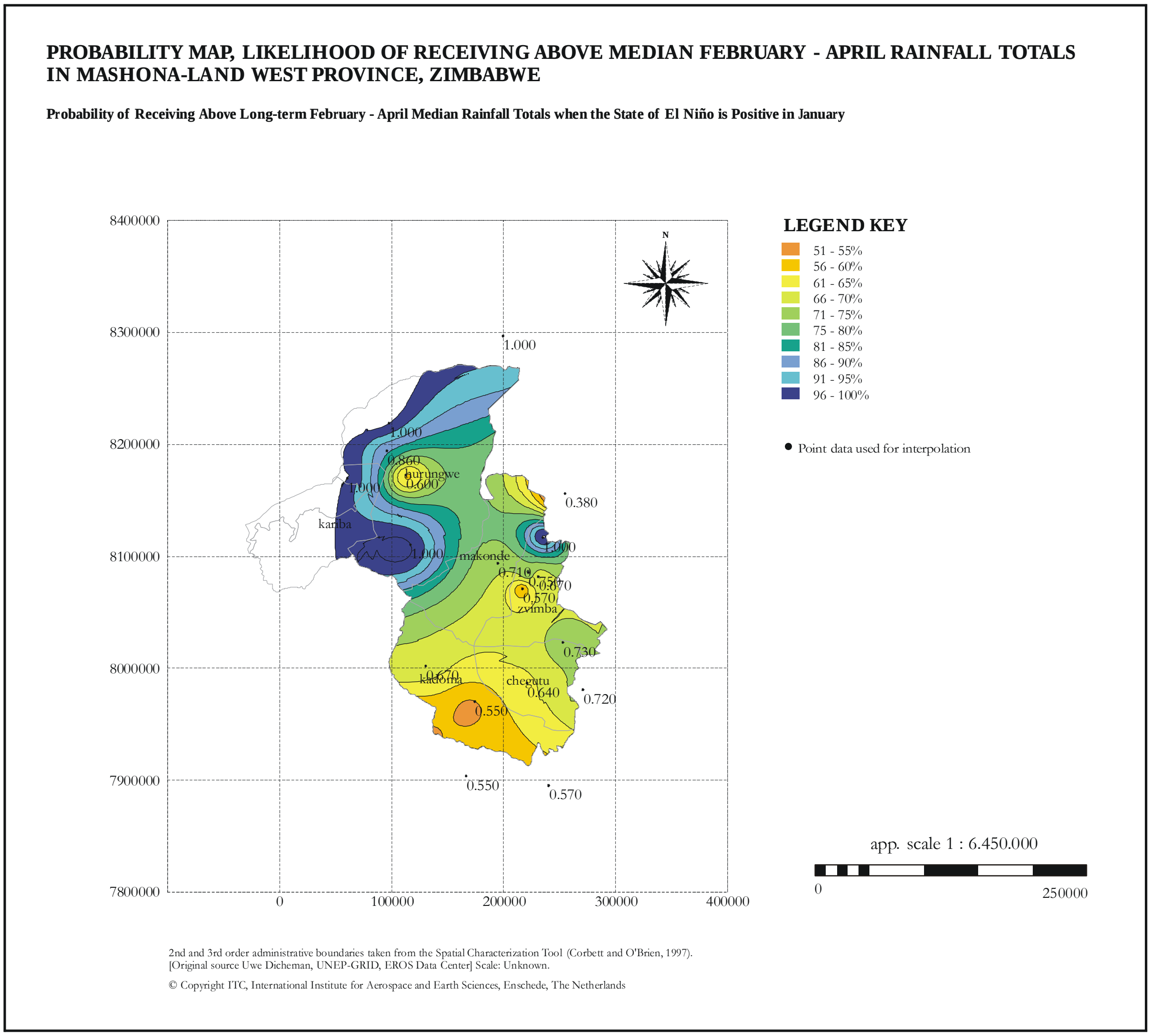


Figure 21, Probability Map, Above Median RTs in Feb - Apr for LPM January

Again, a weighted average of the map classes may characterize the likelihood of receiving above median rainfall totals in February to April better (based on the state of ENSO in January). This can be computed from the map statistics given in Table 3-17.

TABLE 3-17

Map Statistics for Probability Map for LPM January
Likelihood of receiving above median rainfall totals

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 50%	0	0%
51 - 55%	502104.4	0.9%
56 - 60%	3718523.41	6.6%
61 - 65%	5426470.04	9.6%
66 - 70%	9714611.52	17.2%
71 - 75%	6771292.48	12.0%
76 - 80%	6348481.64	11.2%
81 - 85%	3395826.95	6.0%
86 - 90%	3224423.53	5.7%
91 - 95%	4136581.38	7.3%
96 - 100%	5848599.9	10.4%
	49086915.25	87%

From the weighted average of 76.6% can be concluded that roughly 7 to 8 out of 10 times above median rainfall can be expected. Although this is not a very strong likelihood, it is reasonable.

Finally, the relative error in the rainfall forecast is averaged over the seasons 90/91 to 98/99 as indicator of the long-term accuracy of the system at provincial scale (Figure 22). **Note that** this includes extreme weather conditions such as the droughts of 1991/92 and 1994/95 as well as extremely wet years such as 1997/98.

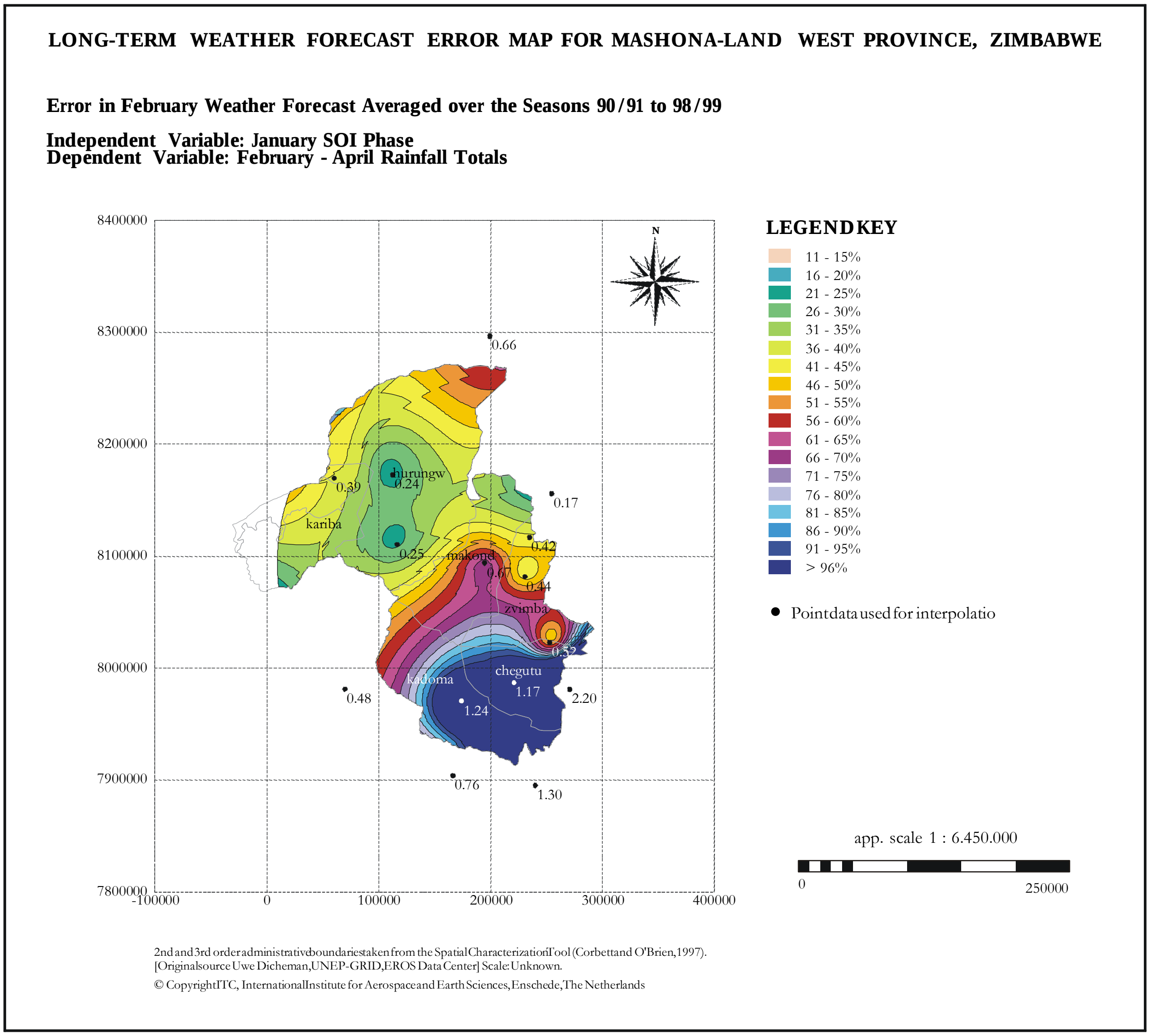


Figure 22, Relative Error in February Forecast

In this case, a weighted average of the map classes characterizes the relative error in the February rainfall forecast also better (based on the state of ENSO in January). This can be computed from the [2]map statistics given in Table 3-18.

[2] **Note that** areas lying in the upper-left corner of the province are excluded from this map analyses since the probabilities observed here are the result of an error in the interpolation due to a lack of meteorological stations.

TABLE 3-18

Map Statistics for Forecast Error Map for LPM February

PROBABILITY (%)	AREA (km^2)	AREA (%)
< 10%	0	0.0%
11 - 15%	27729.65	0.1%
16 - 20%	28482.16	0.1%
21 - 25%	829460.37	1.5%
26 - 30%	4609553.3	8.5%
31 - 35%	7377580.15	13.6%
36 - 40%	7688191.18	14.2%
41 - 45%	5288670.69	9.8%
46 - 50%	3586896.99	6.6%
51 - 55%	2702957.84	5.0%
56 - 60%	2886412.23	5.3%
61 - 65%	2600501.98	4.8%
66 - 70%	2228313.61	4.1%
71 - 75%	1383636.57	2.6%
76 - 80%	1180010.68	2.2%
81 - 85%	1066980.42	2.0%
86 - 90%	998026.25	1.8%
91 - 95%	877967.36	1.6%
> 96%	8687408.31	16.1%
	54048779.74	70%

From the weighted average of 55.7% can be concluded that roughly 45% of the variability in February to April rainfall totals can be explained when analyzed over a period of nine years.

Conclusion February forecast

Taking into account the conclusions of the other map analyses described above, we conclude that regionalization of the rainfall forecasts based on the state of El Niño in January is permitted. However, from the three indicators investigated, we can also observe that below the 8.100.000 (UTM) latitude, the statistical significance, probability analysis and the long-term forecast error indicated that ENSO is less strongly correlated with rainfall patterns than in areas lying above this latitude. Therefore, if Hurungwe and Kariba district were treated separately, another 10% of the variability in seasonal rainfall totals could be explained here. Even then, use of the rainfall outlook with this accuracy may not prove useful for maize yield forecasting. However, since the relationship between weather forecast accuracy and maize forecast accuracy is unknown, maize yield is still forecasted using the February weather outlook.

Surrogate Rainfall Data as input for the Crop Growth Model

To limit the workload, the mean of the low, middle, and high forecast figure is computed. These figures are not yet used as surrogate rainfall data. The Season Similarity Tool first assists in the task of finding years on file that have similar FP rainfall totals within a user-specified range for reasons explained earlier (page 14). The deviation introduces an error and should therefore be carefully selected, resulting in sufficient similar seasons from which to head appropriate weather data with an as small as possible deviation. Weather data from one of these years are appended to the current season weather data available up to the current date. **Note that** other parameters than rainfall required for the crop growth model should be derived from these proposed years as well since some of them are cross-linked. I.e., cloud cover and maximum temperature are related to the amount of rainfall a specific day receives. Due to time constraints, it was decided that other parameters except rainfall could be taken from the actual data series for the periods rainfall was predicted without jeopardizing the objectiveness of the research.

Weather Data as input for the Crop Growth Model

Daily weather data for each weather element for each GIS grid-point are interpolated from the station values using the method fully described in the common weather section of YYF1 and YFF2. These data are then used as input for the WOFOST model.

In conclusion, the weather module of CGMS for YFF1 and YFF2 can be described graphically. The major processes and important inputs and outputs of the weather data interpolation are described in Figure 23.

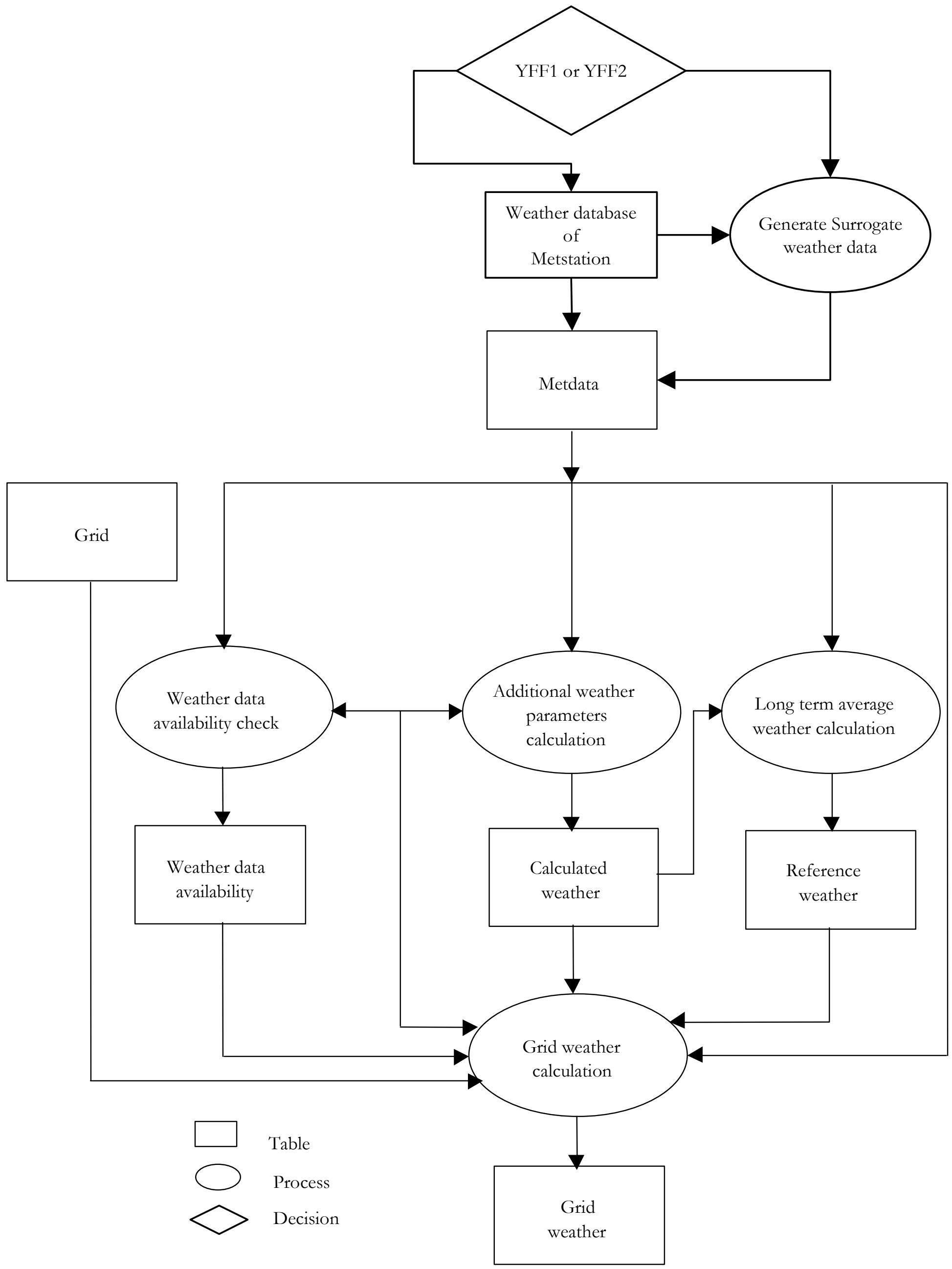


Figure 23, Flow Diagram of Weather Data Processing and Interpolation

Source: Modified from User Manual for the CGMS Model (Mahalder and Sharifi, 1998)

For the crop growth simulation, a regionalized version of the WOFOST model as described by Supit et al. (1994), as part of the Crop Growth Monitoring System (CGMS) has been used. The heart of the system is the geographical information system, around which the various databases, software modules and user-interfaces are constructed and which is the driver of crop growth simulation.

The model is driven by a combined energy balance/water balance module which compares actual transpiration with calculated potential transpiration through a light interception/CO_2 -assimilation/water requirements/water availability module. The model uses only those daily meteorological data that can be made available: rainfall, temperature, including maximum and minimum temperature, vapor pressure (or relative humidity), 24-hour mean wind speed, sunshine duration or cloud cover (to estimate radiation, potential evapo-transpiration, etc.), and if available, measured radiation.

The basic frame of the model was adapted to accommodate the maize variety of interest and is made specific for the study area, based on crop knowledge obtained from maize research at Chibero. This research was executed by AGRITEX with technical assistance from Wageningen University, the Netherlands. Thus, the model is calibrated (calibration of the length of phenological stages as a function of sums of temperatures) based on site-specific field data. In addition, a district specific crop calendar was extracted from the Fortnightly Crop Forecast Reports issued by AGRITEX covering the nineties. The model is run once every 10 days. Model outputs are then available a few days after the end of a 10-day period.

Crop information

The crop knowledge bases refer to literature and expert-knowledge based information available on:

(a) Suitable soil types;

(b) Planting (sowing), flowering and harvest dates;

(c) Crop cycle length and relations between phenology and temperature and day length;

(d) Initial dry matter after emergence (and, indirectly, crop spacing);

(e) Crop specific parameters such as light interception as a function of leaf area index, energy conversion and the partitioning of dry matter into the various plant parts.

The available soils data

The soils information consists of FAO soil types derived from the FAO Digital Soil Map of the World and Derived Soil Properties (ver 3.5, November 1995) on small scale (1:5.000.000). The soil database, consisting of texture classes and soil depth classes, is mainly used in conjunction with the crop knowledge bases, to identify the areas where a given crop can possibly grow (soil suitability). A mean value per grid is obtained by weighting the basic results of the model by the relative area occupied by these soil types.

In Figure 24 the soil map is given for Mashonaland West Province.

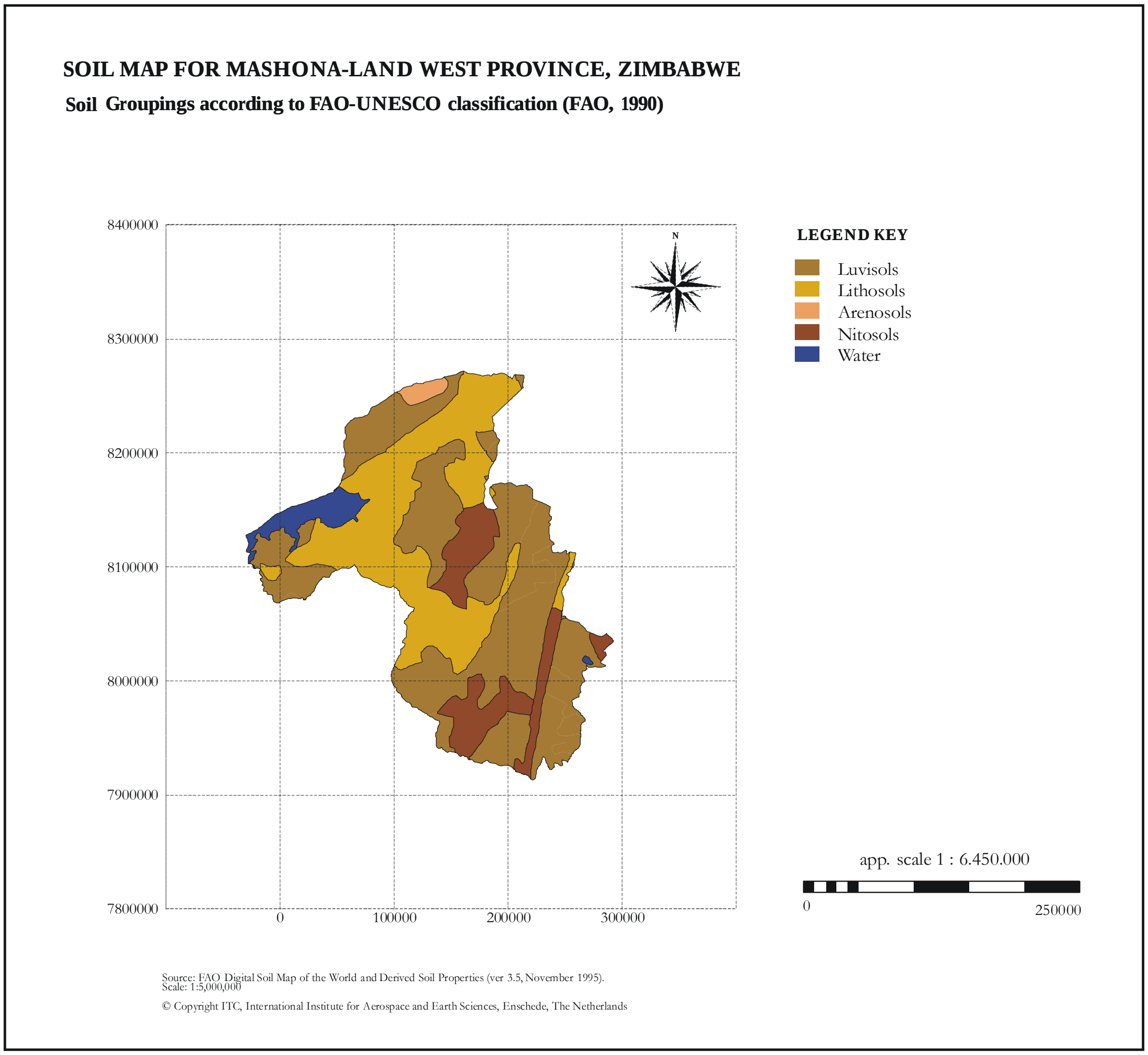


Figure 24, Soil Map for Mashonaland West Province, Zimbabwe

The output of CGMS

The outputs of the system are outputs of agricultural season quality indicators. The main available modeled crop preformance indicators are:

- Biomass, under the actual rainfall conditions, and as if all required moisture was available;
- Grain production, under the actual rainfall conditions, and as if all required moisture was available;
- Estimated actual soil moisture reserve; differences as compared to the previous decade or month;
- State of advancement of the cycle during a given decade; and in addition,

District crop state assessments are aggregated to province level where yield forecasts are produced using an added statistical procedure at 10-day intervals.

Module Three: statistical analysis

In the statistical analysis, a linear regression equation is produced and used to forecast maize yield, which may vary for different stages in the growing season. The regression equation includes the average actual (statistical) yield and the technological time trend over the years, with or without a model indicator to account for the year-specific weather effects. The statistical module was programmed as a spreadsheet, relating the model outputs, through a regression analysis and possibly in combination with a technological time trend function drawn from historical yield data, to the series of historical (sub) regional yields available. The time trend is only used provided it gave satisfactory results in terms of the one-year-ahead and two-year-ahead error analysis; if not, the model function is used singly.

The general procedure passes through the following phases to select model indicators describing the regression equation.

In the beginning of the growing season:

Phase 1. Identification of crop yield/production information, at the level of the smallest region for which series of several years of reliable statistics are available. In our case this is at provincial level.

*At the end of each 10-day period **j**:*

Phase 2. Running of the models using daily meteorological input data at the level of a grid cell within the region. The size of a grid cell is 50 x 50 km.

Phase 3. For each grid cell, calculation of average model outputs.

Phase 4. Aggregation of the weighted average model outputs per grid cell to average values at province level for which yield statistics are available.

Phase 5. Regression analysis of the series of regional crop yields P_i of a region or country against the model outputs and (possible) time trend:

$$P_i = constant + \Psi\{MODEL\ INDICATOR\}_{ji} + \Theta\{TREND\}_i + error_{ji} \qquad [2]$$

Since this phase is computational intensive, the algorithms used were programmed using Microsoft® Visual Basic 6 in order to automate the process.

Model indicators

The candidate model indicators are: a time trend, potential grain yield (storage organ), water-limited grain yield, potential total biomass, and water-limited total biomass. The latter two are used because these are more robust, being less sensitive to modelling errors in the distribution of assimilates. Furthermore, biomass indicators allow 10-day yield predictions during the growing season, when grain filling has not yet started or grains are still very small (Hooijer and van der Wal, 1994). These model indicators reflect the compound effect of soil-weather conditions throughout the growing season on crop growth. Their contribution in the equation is calculated over a large number of years, i.e. six or seven in total, to improve accuracy.

The possible variables that enter in the regression analysis are then:

a) The trend (1 variable)

b) One of the following outputs, which are interdependent: biomass or grain (1 variable). Water-limited yield storage organ (wlyld), potential yield storage organ (ptyld), water-limited yield biomass and potential yield biomass (ptbio) simulation variables were selected as the model indicators. Originally, it was intended to predict yields by solely using the water limited yield (wlyld) as the model indicator. Later on, the other indicators were also added. Water limited yield, for instance, is inappropriate for a region with lot of irrigation. Total dry matter is expected to be a more robust predictor than the grain weight since it is less sensitive to modelling errors in the distribution of assimilates (de Koning et. al., 1993). Thus, biomass was also added to avoid the error coming from yet not fully understood mechanism in the distribution of assimilates.

c) Or the following variable: rainfall more (or less) than the simulated water requirements as defined by the ratio rainfall over total water requirement on decadal basis in the form of:

$$PREC_{sum} / ETO_{sum} \qquad [3]$$

Where :

$PREC_{sum}$ is the summed precipitation, and;

ETO_{sum} is the summed potential transpiration from a crop canopy (according to Penman).

all on decadal basis.

This can affect crop yield, but is not reflected or only indirectly by the direct model outputs. Considering the agro-meteorological problems observed in the recent past of Zimbabwe, especially in the northern areas, this variable was regressed additionally as proposed under the section Research Question, Common Issues.

Simulation results

The accuracy of the simulation results was assessed by analyzing the correlation coefficient and coefficient of determination (R^2) of a simple-linear regression between observed and simulated data for decade 20 of the crop cycle. Since yield statistics were provided by two independent sources, namely by AGRITEX, provincial office Chinhoyi (POC) and by the Central Statistical Office (CSO), Harare, the analyses is made for both cases.

TABLE 3-19

CORRELATION MATRIX CSO-OBSERVED VS. SIMULATED

	CSO_observed	*ptyld*	*wlyld*	*ptbio*	*wlbio*
CSO_observed	**1**				
ptyld	**0.062659885**	1			
wlyld	**0.907575262**	0.1603	1		
ptbio	**0.100771651**	0.9423	0.1244	1	
wlbio	**0.918181296**	-0.259	0.8148	-0.16	1

TABLE 3-20

CORRELATION MATRIX POC-OBSERVED VS. SIMULATED

	PHQ_observed	*ptyld*	*wlyld*	*ptbio*	*wlbio*
PHQ_observed	**1**				
ptyld	**-0.228091805**	1			
wlyld	**0.613112534**	0.1603	1		
ptbio	**-0.186908675**	0.9423	0.1244	1	
wlbio	**0.841848529**	-0.259	0.8148	-0.16	1

For simulation variables with the highest correlation coefficients with observed data, the coefficient of determination (R^2) are depicted in the scatter plots below.

CSO

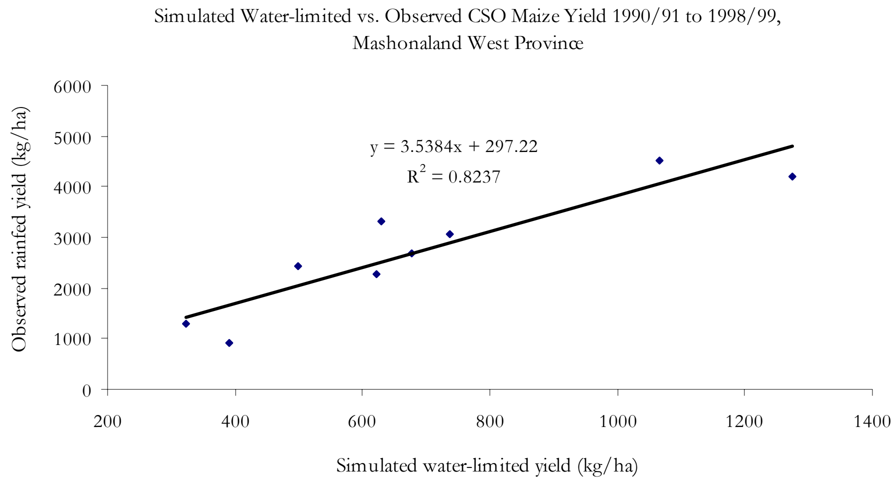


Figure 25, Scatter plot simulated water-limited vs. observed CSO Maize Yield 1990/91 to 1998/99, Mashonaland West Province, Zimbabwe

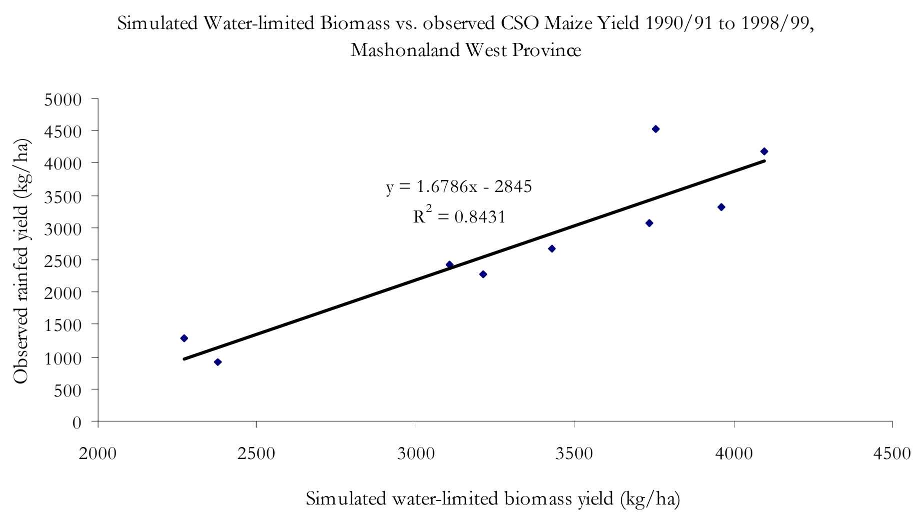


Figure 26, Scatter plot simulated water-limited biomass yield vs. observed CSO Maize Yield 1990/91 to 1998/99, Mashonaland West Province, Zimbabwe

POC

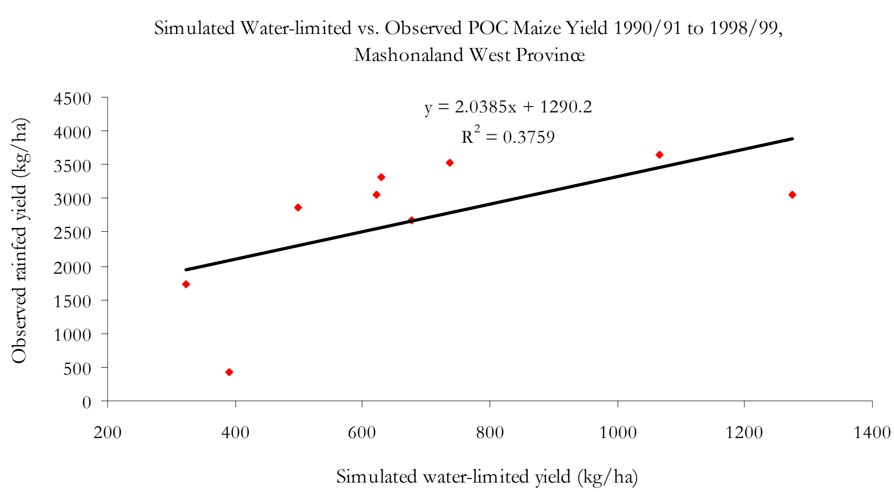


Figure 27, Scatter plot simulated water-limited vs. observed POC Maize Yield 1990/91 to 1998/99, Mashonaland West Province, Zimbabwe

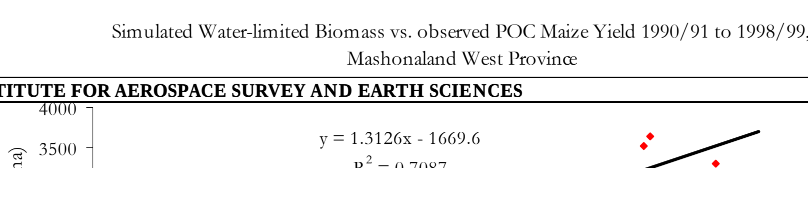


Figure 28, Scatter plot simulated water-limited biomass yield vs. observed CSO Maize Yield 1990/91 to 1998/99, Mashonaland West Province, Zimbabwe

As the regression analysis reveals, simulated crop performance (water-limited biomass) correlates better with the observed if we use yield statistics provided by CSO instead of those provided by AGRITEX (POC) (respectively a adjusted R^2=.82 versus a R^2 of .67).

The time trend

Simulated yields cannot directly be considered as actual yields, even when corrected for sub-optimal cultivation practices using linear regression. This can partly be subscribed to a trend (of rising yields) in official yields not yet included in the model (Hooijer and van der Wall, 1994). Therefore, to account for the influence of increasing farmers' skill, increasing use of technology to crop maize on yield, a fifth indicator, the so-called 'time trend', is tested as well. Literature suggests that a simple linear model to describe this trend is sufficient in most cases (Swanson and Nyankori, 1979 cited by Hooijer and van der Wal, 1994). A smooth trend of any type over a large number of years assumes a continuity that might be unrealistic. For that reason, Hooijer and van der Wal (1994) suggest to base this indicator only on data from the recent past. Its length should nevertheless be long enough to give a sufficient number of degrees of freedom in a regression analysis.

In practice, the length of the time series used for the statistical model validation has been set to k = 9 years (if the total length n of the available series is <k, then k=n). If the tested trend is zero, there is no yield increase or decrease because of technological or possible socio-economical change or no time trend. In Figure 25 the time trend for Mashonaland West is given based on yield statistics provided by AGRITEX, provincial office Chinhoyi.

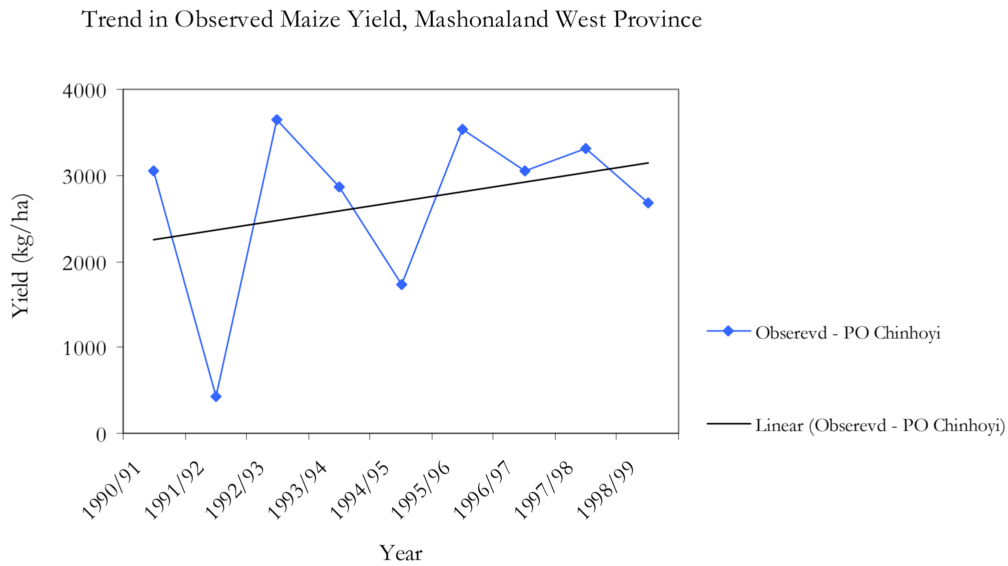


Figure 29, Time Trend in Observed Maize Yield according to PO Chinhoyi

In **Error! Not a valid link.** the time trend for Mashonaland West is depicted based on yield statistics provided by the Central Statistical Office (CSO), Harare.

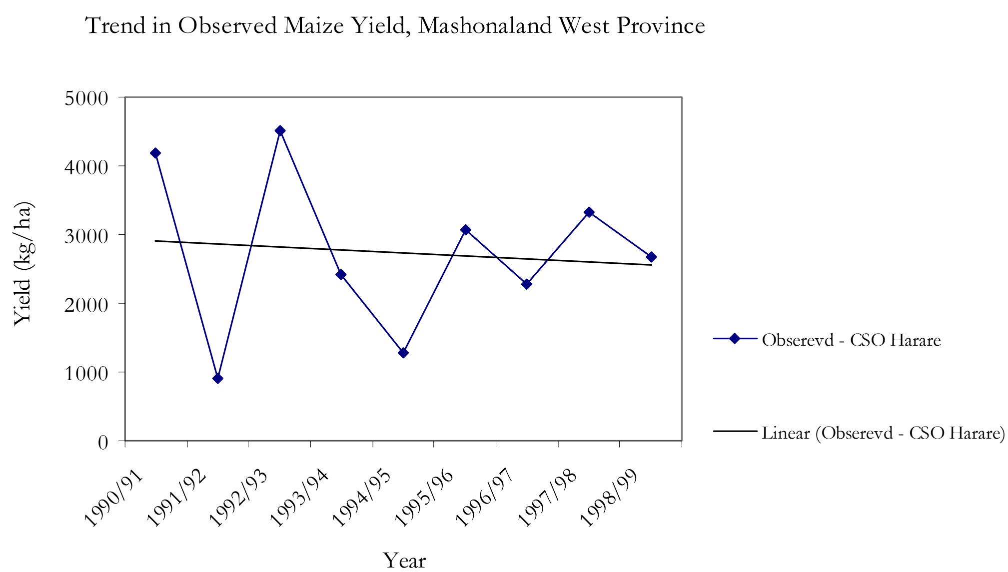


Figure 30, Time Trend in Observed Maize Yield according to CSO Harare

The results show that the trend observed in yield statistics provided by AGRITEX, provincial office Chinhoyi is positive, whereas that based on yield statistics provided by the Central Statistical Office (CSO), Harare, is negative. In addition, the deviations from the general trend are significant for both cases, meaning that the trend functions are expected to have little predictive power on their own.

Elementary predictors

A (even highly) significant statistical regression does not necessarily imply that the prediction error also decreases. To assess the usefulness of the statistical validations for yield forecasting, the following indicators of the quality and trustworthiness of the predictions are determined during the procedure.

The elementary predictor, based on simulation results with time trend function, with the lowest jackknife root mean squared prediction error was selected as full prediction rule. The jackknife method (also referred to as leave-one-out method or Allen's PRESS method) works as follows (Allen, 1971 and Wallach & Goffinet, 1989). The yield observations of all years, except one, are used to construct a predictor which is applied to the year kept out of sight, in order to evaluate the prediction error. This is done for each year in turn.

$$e_{jac(i)} = Y_i - P_i \qquad [4]$$

We then calculate the jackknife root mean square of the prediction errors for the *k* years in question:

$$Jac_k = \sqrt{(e^2_{jac(1)} + e^2_{jac(2)} + e^2_{jac(3)} + ... + e^2_{jac(k)})/k} \qquad [5]$$

Where *k* is the number of predictions made from information of *k*-1 years (i.e. *k* = 1991-1999=9). In this way, the estimates for each year are not directly linked to the observation to the year in question.

A one-year-ahead (OYA) prediction error was used to evaluate the elementary predictor rule accuracy using a dynamic data-window. The sums of squares of the differences between the observed values and the values predicted by this elementary predictor was calculated and divided by *n*, now taking account of only the six most recent years (*n*=6) to formulate the elementary predictor. The mean square MS_{oya} obtained in this way was compared with the variance of the residuals of the yields in relation to the general trend, S^2_z, to characterize the part of that variance that is explained by the simulation data:

$$R^2_{oya} = \frac{S^2_z - MS_{oya}}{S^2_z} \qquad [6]$$

Similar to R^2_{oya}, also a jackknife prediction error was calculated, R^2_{jac}, where the mean square, MS_{jac}, was now obtained based on all data present for the region (*k*=12).

If R^2_{oya} is positive, then the predictions using agrometeorological model outputs are an improvement as compared to the use of the time trend alone. However, if it is negative, the time trend function should be used alone, making the simulation effort futile.

The procedure eventually results, at the end of each 10-day period j, in a set of model indicators, their constants and corresponding regression coefficients. For decade 7 to 20 of the crop cycle the statistical indicators are tabulated (**Error! Not a valid link.**).

TABLE 21

ELEMENTARY PREDICTORS AND THEIR PERFORMANCE

	Statistical indicators					
Predictand	PO Chinhoyi			CSO Harare		
Decade	MI	R^2_{jac}	R^2_{oya}	MI	R^2_{jac}	R^2_{oya}
7	wlbio	-0.11	-2.54	wlbio	-0.97	-3.63
8	prec/twr	0.35	0.11	prec/twr	0.14	-1.33
9	wlbio	-0.14	0.38	wlbio	-0.82	-0.26
10	wlbio	-0.15	0.58	wlbio	-0.81	-0.10
11	wlbio	-0.15	0.51	wlbio	-0.78	-0.19
12	ptyld	-0.10	0.62	wlbio	-0.82	-0.15
13	ptyld	-0.14	0.69	wlyld	-0.81	-0.09
14	wlbio	0.06	0.47	wlbio	-0.36	-0.10
15	wlbio	0.36	0.30	wlbio	0.19	0.15
16	wlyld	0.55	0.64	wlbio	0.51	0.45
17	wlyld	0.56	0.32	wlyld	0.77	0.47
18	wlbio	0.52	0.05	wlyld	0.79	-0.04
19	wlbio	0.49	-0.05	wlbio	0.72	0.64
20	wlbio	0.48	-0.06	wlbio	0.72	0.63

Where,
MI = Model indicator

Examination of the statistical indicators for predictand CSO Harare shows that R^2_{oya} is negative up to decade 14. This means that the sum of squares of the forecasting errors is greater than the sum of squares of the trend residuals, which reveals that the simulation models have no predictive ability for this particular period. From decade 15 onwards, however, all R^2_{oya} values are positive (except decade twelve). Based on these results, we have every reason to be satisfied with models based on simulation results that, on average, explain 50% of the variability of the deviation of the yields from the general trend.

When we examine the R^2_{oya} values for predictand PO Chinhoyi, we immediately find that from decade 8 up to 18 they are positive, which means that models with simulation results are an improvement compared to the time trend function alone. On average, more than 40% of the variability of the deviation of the yields from the general trend is explained during this period. The R^2_{oya} for the first and last two decades are negative, meaning that the simulation models have no predictive ability for this particular period.

When we look more closely to the model indicators used by the prediction models, the hypothesis that prec/twr can successfully be used to model the performance of maize crops during waterlogged periods does not hold. As expected, “water-limited” predictors predominantly outperformed their “potential” variants, and were predominantly selected to predict maize yield.

Full prediction rules

Again, a one-year-ahead (OYA) relative prediction error was used (algorithm **Error! Not a valid link.**) to evaluate the full prediction rule accuracy using a dynamic data-window. The OYA relative prediction errors are calculated based on models developed with data of the most recent six years. If 9 years are available in the database, we have 9-6=3 occasions to compare predictions.

$$e_{rel(i)} = \frac{Y_i - P_i}{\bar{Y}_i} \times 100 \qquad [7]$$

The procedure eventually results, at the end of each 10-day period j, in a set of relative errors. The attractiveness of this method lies in the number of observations; even with a very limited mount of data, the dynamic data-window technique allows for an in-depth analysis of the prediction errors with three observations available per decade. The mean relative error of these three observations is an indication of the performance of the framework (YFF1) under study.

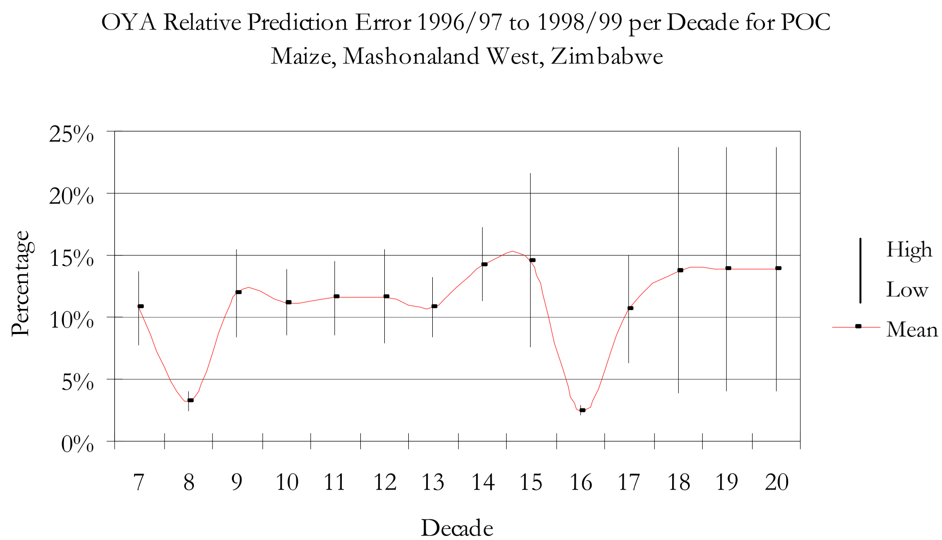


Figure 31, OYA Relative Prediction Error 1996/97 to 1998/99 per Decade for POC Maize, Mashonaland West, Zimbabwe

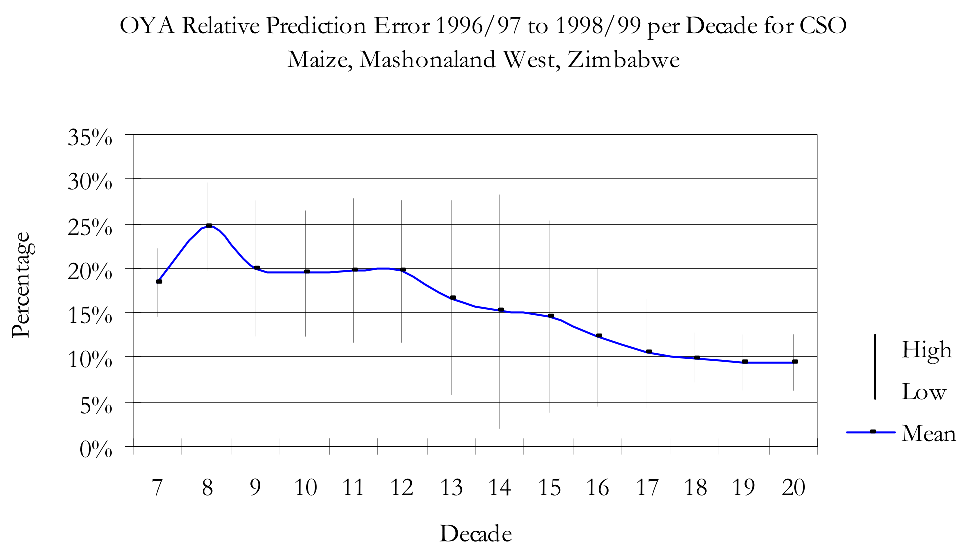


Figure 32, OYA Relative Prediction Error 1996/97 to 1998/99 per Decade for CSO Maize, Mashonaland West, Zimbabwe

The results show that the prediction error of CSO maize yields averages around 10% close to harvest, whereas that of POC yields around 14%. This points to a reliable yield forecast framework.

Based on expert knowledge and common sense, one would expect lower accuracy with less information. Indeed, from **Error! Not a valid link.** it can be concluded that this is the case for predicting CSO yield. However, the forecast on the 2nd decade of POC predictions are unexpectedly low (**Error! Not a valid link.**). This forecast is prepared using the full prediction rule based on the time trend and the amount of rainfall below/above crop water requirement, expressed in the ratio of rainfall over crop water requirement ($PREC_{sum}/ETO_{sum}$). Apparently, it was useful as an indicator of rainfall availability during the beginning of the vegetative development stage of the maize crop for POC yield statistics. Remember that during this period, exponential growth is observed and all energy is prioritized to leaf production for the plant to intercept more light. Thus, as the specific leaf area increases the transpiration increases also and the demand for moisture increases accordingly. Being partly dependent on weather data and less on the performance of the simulation model, moisture in excess or shortage of the projected water requirement proved useful for early crop monitoring, and consequently, for estimating the final yield. Caution is required in making hard conclusions from these results since they are based on models developed with only six observations, whereas nine is a (statistically) sufficient number. The number nine is linked to the degree of freedom to keep a minimum number of observations, considering that we are using two predictors (trend + indicator).

Another observation is that the deviation from the mean error are quite significant for the CSO yield up to decade 17; from decade 18 and onwards these deviations decrease to a couple of percent points. For the POC yield however, this is the other way around. During the first 14 decades, relative small deviation from the mean can be observed (app. 5%), whereas these deviations increase during the last 6 decades to approximately 10% (except for decade 16). This is in agreement with what the analysis of the R^2_{oya} values revealed (see **Error! Not a valid link.**). For CSO yield, only at the end models with simulation variables started to explain the variability of the deviation of the yields from the general trend. The R^2_{oya} values of models for POC yield prediction, however, remained positive for practically the first half of the crop cycle, pointing to more stability and explanatory power for that period.

YFF1 vs. YFF2

Figures 29 allow analysis of the maize yield frameworks under comparison, i.e. YFF1 based on actual rainfall data and YFF2 based on ENSO predicted rainfall. Note that only preliminary conclusions are permitted from this comparison, since the data for YFF2 covers just one season.

The relative yield forecast error of YFF2 - November seems to stabilize at app. 20% from decade 16 onwards which is equivalent to the weather forecast error of app. 20% (long-term: app. 25%) on which this yield forecast is based. Although not representative for the whole population, a relative forecast error of 20% is an encouraging sign, since this forecast is available as early as November. Note that although YFF1 out performs YFF2 in accuracy, it is dependent on current weather data, which may come available at a much later point in time. In such occasions, YFF2 is expected to be useful since its timeliness is assured, being dependent on, often, available historical weather data. The above may also indicate a 1:1 relationship between the accuracy of the ENSO rainfall prediction (for November) and the accuracy of the forecasted maize yield (based on that rainfall prediction). Therefore, there is reason to expect that the long-term error in the ENSO based rainfall outlook is also a good indicator for the overall performance of YYF2. For the November ENSO based rainfall outlook, this would amount to 24.5%.

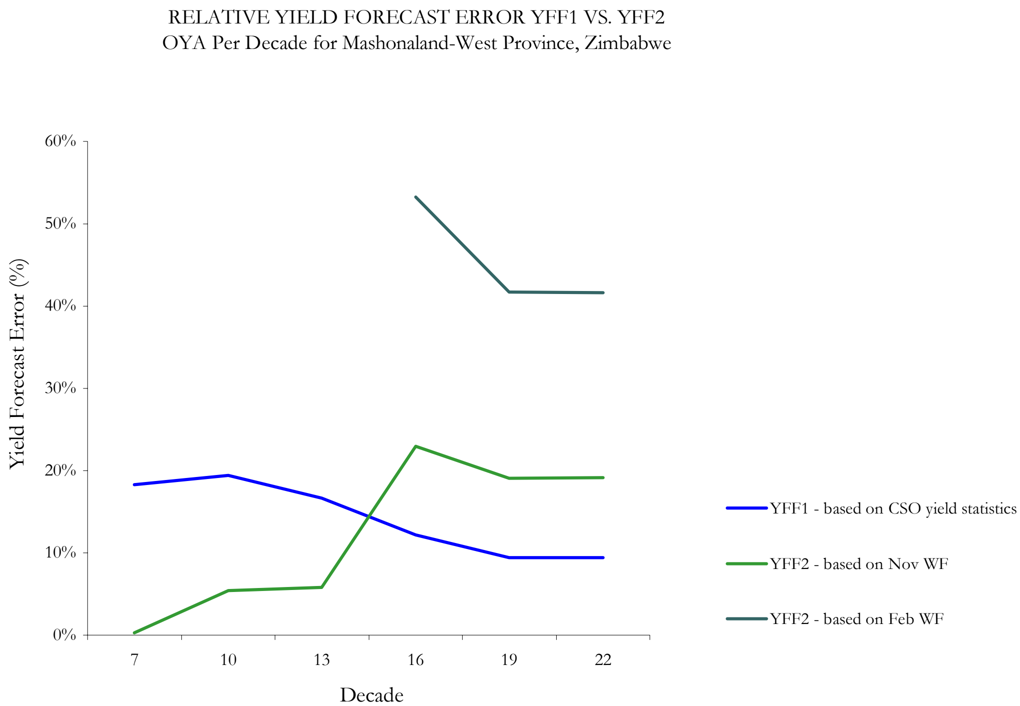


Figure 33, Relative YF Error for OYD Yield Statistics Scenario YFF1 vs. YFF2

The error of YFF2 - February yield forecast seems to stabilize at roughly 40% from decade 19 onwards which is not far from to the weather forecast error of 30% (long-term: app. 75%) on which this yield forecast is based. However, it does not give reason to expect a 1:1 relationship between the accuracy of the ENSO rainfall prediction (for February) and the accuracy of the forecasted maize yield (based on that rainfall prediction). One may conclude that considering the unexpected high accuracy of the WF (30% vs. long-term of app. 75%), the error observed in this yield forecast (40%) is not representative for its long-term performance and that worse accuracy may be expected.

4 Discussion

Elements common to both yield-forecasting frameworks are discussed in the following.

4.1 Common issues

Waterlogging and cob rot problems observed in Zimbabwe in the recent past prompted to introduce a variable $PREC_{sum}/ETO_{sum}$ in *Module Three: Statistical Analysis* of this thesis, where this variable was regressed against observed yields as detailed under *Research Question, Common Issues*. The variable is an indicator of rainfall excess or shortage projected onto the simulated water requirements computed on a decadal basis. Rather unexpectedly, it was found that simulated yields were not lower than average in years in which problems of this nature are known to have occurred. 'Fortnightly Crop Forecast Reports' indicated that problems of this nature happened indeed, scattered over (parts of) the province. This does not necessarily imply that the indicator could not detect this. It is possibly obscured by the averaging technique applied to all model outputs in the process to arrive at provincial estimates. Moreover, these outputs are normalized per hectare. Averaging of data tends may result in overestimating phenomenon; land units with very low per-hectare values may be obscured during the aggregation process if averaging involves other, high per-hectare values.

Early forecasts, i.e. forecasts in the 8th decade, are based on the same variable selected for its relatively low forecast error; it was identified as discussed under *Module Three: Statistical Analysis*. Although not particularly useful to explain problems induced by excess rainfall, the indicator was successfully used to model the performance of the maize crop during early vegetative development, when maize plants exhibit exponential growth and all absorbed energy is focused on leaf production to maximize light interception. As the leaf area increases, relative transpiration increases also, and moisture uptake by the root system must increase accordingly. Therefore, moisture availability is an essential factor in this stage and a parameter that describes this is essential for accurate forecasting of crop performance. Note that any forecast technique will be hard pressed to predict accurately early in the season; with traditional model indicators being the only information available one must expect low accuracy results.

The dependability of official maize yield statistics on regional scale is a further concern. The assumption that historical maize yield statistics are 'accurate' remains doubtful, being confirmed by our comparison between those issued by AGRITEX and those by the CSO. Even if it holds, they remain normalized; it does not hold in the absolute sense since the total acreage under maize is not accurately known. This makes it impossible to arrive at provincial production totals. This implies that for operational use of the approaches evaluated in this research, acreage under maize must be established first and with acceptable accuracy.

4.2 Framework-specific issues

The following issues are mutually exclusive for the YFF1 and YFF2 approaches.

4.2.1 YFF2

The ENSO weather phenomenon is argued to be strongly region-orientated. Most probably, similar results to the one described here can be achieved for the whole SADC region. This is in agreement with the experience of De Jager et al. (1998) who attempted to forecast maize yields in the Free State Province of South Africa, using weather forecasts according to a calibrated CERES-MAIZE model based on this same principle. However, the finding that (reasonably) accurate weather forecasts could only be made one month before the start of the growing was unexpected. De Jager and co-workers maintained that the correlation of rainfall and ENSO over South Africa is such that forecasting could not just be done one month before the growing season, but could continue in the season, all the way up to the end (De Jager et al., 1998). Hence, within-season dynamics can be modeled to provide updated forecasts as the season progresses. At first, this point of view seemed to hold for Mashonaland West province also, but spatial analysis of the ENSO based weather forecast revealed that the long-term relative error for the February forecasts was larger than expected from the statistical analysis. Thus, it became clear that model precision should be evaluated with a well-balanced, comprehensive evaluation of the statistical significance, the absolute rainfall difference observed between typical positive and negative ENSO years, the probability to receive a specific rainfall forecast, and the long-term relative prediction error. It was hoped at first that the February forecast would permit to incorporate the effects of changes in the predictor; now it is known that for this sub-region such dynamics cannot be modeled in a way that would permit to update forecasts as the season progresses. The low usefulness of the February weather forecast was confirmed by the low maize yield forecast accuracy based on these surrogate rainfall figures.

Spatial analysis of the February weather forecasts also showed that below app. the 8.100.000 (UTM) latitude, ENSO is less strongly correlated with specifically late rains than above this latitude. This may to some extent be connected with the fact that the inter-tropical convergence zone (ITCZ), which is responsible for most rainfall over Zimbabwe, comes from the north and is less hindered by other atmospheric developments, e.g. subtropical high-pressure belts, at higher latitudes (Unganai, 1998).

Although El Niño does have different impacts in different parts of the world, the agrometeorological ENSO Rainfall Analysis and Forecast Model prepared for this research can be used for any of the regions of the world depicted in Figure 4: Climatic impacts of warm El Niño events (October-March).

5 Conclusions

Research questions were formulated to guide the research and permit to distinguish between issues of primary and secondary importance (Chapter 2, section *Research Questions*). They addressed the yet unknown or incompletely understood aspects of the two yield-forecasting frameworks under comparison. In the following, conclusions are drawn from the results of the research initiated to answer these questions.

The main aim of this research was to contribute to the development of a framework for maize yield forecasting by setting up and testing the relative merits of two approaches. It is generally true that the relevance of yield indicators increases as they become available earlier and/or have greater accuracy. Hence, the prediction error and the decade of the forecast were used to compare the two frameworks.

It is tentatively concluded that YFF2 is outperformed by YFF1 in terms of accuracy, however the aspect timeliness is an important issue especially if the forecasts should have any value for farm management. The timeliness of YFF1, which uses weather data of the current season, is obviously affected by the speed at which this weather data comes available for crop growth simulation. If this data is quickly collected, processed and distributed to provincial level, it seems a defendable statement that YFF2 is outperformed by YFF1. However, if there exists a significant delay or complete absence of weather data, YFF2 outperforms YFF1 since it is not dependent on current weather data but on historical, readily available records.

The author was invited to present the Agrometeorological ENSO Rainfall Analysis and Forecast Model at the Seasonal Climate Forecast Workshop 1999 in Harare, where several users indicated a need for site-specific, quantified rainfall forecasts. Based on this thesis, the methodology and model have been set up and its results analyzed over a long period of time and under extreme weather conditions.

The recommendations of the "Agromet and Crop Monitoring Project (ACMP) in the SADC region – 3" report specifically mention (Roebeling et al., 1999):

(1) "Investigate the improvements of the ACMS products contained in the use of annually derived planting date files."

Although improvements were not specifically investigated, the results reported in this thesis are based on district-specific crop calendars, extracted from the 1990's Fortnightly Crop Forecast Reports issued by AGRITEX.

(2) "Application of the relationship between records of reported and predicted yields, during a period of at least 8 years, to perform a statistical correction of the predicted yields."

In this thesis, the relationship between reported and predicted yield records, for a period of 9 years, has been used to perform statistical correction of yields predicted for the study area. The good correlation ($R^2_{adj.}$ = 0.84) between observed and modeled yield is encouraging.

(3) In the same context, elsewhere in the report it was stated (page 61): "…In fact, this method has been applied in the framework of the JRC MARS Project, MARIE-C. It turned out that the prediction accuracy with this method is very high, with errors below 10%. It is tempting to use a similar methodology in the SADC."

Hereby, this methodology has been applied to a small part of the SADC region. This study confirmed that the prediction accuracy of the method is indeed very high, with an average relative prediction error of 9.4 % (last decades). However, under extreme weather conditions, such as the droughts during the year 1994/95, the relative error of the forecasts is expected to increase

Considering the above, it is justified to believe that this research contributed to the development of a framework for maize yield forecasting for Zimbabwe.

5.1.1 Common issues

Conclusions that apply to both approaches are given below.

The first research question concerned the regression analyses of observed yields against model outputs. When actual yields do not correlate well with model-generated yield levels, changing the degree of aggregation to levels with more reliable yield statistics does not always solve the problem. This thesis suggests that introduction of 'new' variables could improve this regression. In our case, $PREC_{sum}/ETO_{sum}$ was introduced as an indicator of rainfall excess or shortage in relation with the simulated water requirements computed on a decadal basis. Either this variable does not perform as expected, or, the methodology did not permit to test this variable for what it was intended, possibly due to the level of aggregation as explained in sub-section *Maize Yield Forecast Component,* Chapter 4.

5.1.2 Framework-specific issues

The following conclusions to research questions are listed separately for YFF2 and YFF1, as they are mutually exclusive for both yield-forecast frameworks.

5.1.2.1 YFF2

Research has shown that the Southern Oscillation Index is a useful indicator of the amount of summer rainfall over Zimbabwe (ZIMMET, 1997). However, to justify feeding a cop growth model with surrogate weather data generated and based on this premise, an in-depth analysis of the impact of El Niño on the region was needed. The temporal and spatial characterization of El Niño for rainfall revealed the following (detailed in the sub section *Surrogate meteorological data and their processing: YFF2).*

(1) YFF2 is based on alleged links between ENSO and rainfall anomalies, and assumes that significant differences in seasonal rainfall exist in the study area and affect maize crop seasons grouped according to Stone's SOI phases (Stone et al., 1996). Since this hypothesis needs verification, a non-parametric test (Kruskal-Wallis) was executed for twenty-three meteorological stations in the study area, based on data varying from 30 years up to 108 years. The null hypothesis (H_o) was that there is no systematic difference in forecast period (FP) rainfall totals between seasons grouped according to the different SOI phases; versus the alternative hypothesis (H_a) for research, that there is systematic difference in FP rainfall totals between the seasons grouped according to the different SOI phases. The thesis that systematic difference in FP rainfall totals exists between seasons grouped according to the different SOI phases (with a confidence level of 95%) holds for 14 stations in the study region.

(2) El Niño/Southern Oscillation (ENSO) derived rainfall outlooks are based on the assumption that the atmosphere is the best model of itself (Unganai, 1998). In view of the prominent role of historical and huge data needs of models based on this premise, a question to be addressed was whether rainfall patterns in the recent past are as strongly conditioned or influenced by ENSO as observed in earlier times. If historical rainfall data are stronger correlated with the state of El Niño than recent rainfall data, it is less likely that El Niño will have less impact on rainfall patterns in the future as well. That would make it less relevant for forecasting purposes. Furthermore, it would greatly weaken the weather prediction, as the number of relevant observation data would then reduce to those observed in the time-period of strong impact only. The robustness of the statistical analysis would suffer correspondingly.

To answer this question, five of the twenty-three stations were analyzed for differences in statistical significance for different periods in time. For Karoi, an example has been included in the section entitled "*Surrogate meteorological data and their processing: YFF",* Chapter Three. Table 5-1 summarizes the results of these five meteorological stations analyzed.

TABLE 5-1

LPM October and its Correlation with FP Rainfall Totals over Different Periods

STATION	PROBABILITY ALL YEARS	PROBABILITY < 1958	PROBABILITY > 1958
Harare/Belvedere	94.6%	94.8%	77.1%
Guruve	99.6%	89.1%	94.1%
Karoi	99.2%	86.9%	96.0%
Gokwe	96.0%	64.9%	89.4%
Chinhoyi	98.4%	85.4%	86.0%
mean:	97.6%	84.2%	88.5%

Considering the extensive amount of data analyzed it is a defendable statement that rainfall patterns in the recent past are affected more strongly by ENSO than in earlier times, and hence, that the indicator is expected to be valid in the near future as well. This view was also confirmed by a comparison of the absolute difference in rainfall totals.

(3) The nature of the weather forecasts currently provided by the Meteorological Department of Zimbabwe does not fully meet user demands. For regional priority management in malaria control or, as in our case, crop growth modelling, quantified rainfall estimates are required, at sufficient spatial detail, rather than national estimates that are expressed in such terms as "70% chance of receiving above

average rainfall". This raises the following question for research: "Are there differences between weather records in how they are affected by ENSO and does this result in spatially differentiated weather outlooks?"

The answer to this question is affirmative: (slight) between-station differences were observed in the performance indicators used in the spatial analysis of ENSO. However, this differs for different forecasts. The spatial analysis of the November forecast showed that a slight difference in the point data must not be overvalued since it does not recur in all performance indicators and hence, a specific intra-distribution or pattern could not be identified. It can be concluded that for the November forecast a weighted average of map classes would spatially characterize the impact of ENSO on rainfall totals in November to April best fro Mashonaland West (based on the state of ENSO in October). Hence, no improvement can be made by a region-to-region approach for this forecast.

The spatial analysis of the February weather forecast, however, showed that below app. the 8.100.000 (UTM) latitude, ENSO is less strongly correlated with rainfall patterns than for areas above this latitude. Possible explanations for this phenomenon are offered in Chapter Four. It can be concluded that an improvement can be made by a region-to-region approach for this forecast, i.e. another 10% of the variability in seasonal rainfall totals could be explained in Hurungwe and Kariba districts if these districts were handled separately. This is a defendable statement, because it appears to hold for all performance indicators analyzed.

The question whether it is possible to produce quantified rainfall forecasts within a reasonable error margin based on the ENSO principle can be answered positively. However, again this differs for different forecasts. The long-term mean error in the ENSO based rainfall outlook of November would amount to 24.5%, whereas that of the rainfall outlook of February would be as much as 74.5%. Clearly, the latter is unacceptably high and would not improve any of the existing methods introduced in Chapter One. The long-term mean error in the ENSO based rainfall outlook of November is acceptable in that it would be an improvement compared with the results of existing methods. This is confirmed by the error of the November yield forecast (YFF2), which seems to stabilize at app. 20% from decade 5 onwards, of the same order as the weather forecast error of app. 20% (long-term: app. 25%) on which this yield forecast was based. Although yield forecast error data are insufficient to permit hard conclusions, a relative yield forecast error of 20% is an encouraging sign, considering the fact that it is available at the start of the growing season.

(4) The scale at which crop performance is monitored aims at analysis of a region, whereas the ENSO rainfall estimation is on point basis. To justify interpolation of these point estimates, geo-statistical analysis of the phenomenon is required. Only if the significance of the predictor is sufficient and spatially structured, is regionalisation of ENSO-based weather outlooks justified.

Spatial analysis of the significance level of all 23 stations revealed that for 80% of Mashonaland West Province the theses stated under (1) hold. Taking into account the conclusions of the other map analyses, described in section *Surrogate meteorological data and their processing: YFF,* Chapter Three, we may conclude that regionalisation of rainfall forecasts based on the state of El Niño is permitted in this case.

An overall conclusion would be that, although limited, ENSO-based November weather forecasts remain useful for preliminary estimates of regional corn production since the forecast is available well in advance of actual production and with reasonable accuracy. Although yield forecast error data are insufficient to permit hard conclusions, a relative yield forecast error of 20% for the ENSO based crop simulation is still an encouraging sign since the weather forecasts can be improved when lurking variables would be introduced.

6 Recommendations

In this section, recommendations are given that can contribute to further improve the yield forecast frameworks presented in this report.

6.1 Common issues

(1) As mentioned earlier, the reliability of the official maize yield statistics is an important factor. To permit absolute yield forecasting at (sub) national scale, the acreage under maize has to be established first and with sufficient accuracy.

(2) Verify the automated selection procedure used to select the model indicator as described under *Phase 6,* section *Module Three: statistical analysis.* As mentioned earlier, careful and case-by-case analysis would improve forecasting.

(3) To permit optimal use of the frameworks, it is necessary to correct and enhance the existing soil information.

(4) Investigate the accuracy of yield forecasts at district level instead of provincial level.

(5) Investigate the implications for (results of) crop modelling of the use of cold cloud duration images to interpolate measured rainfall.

(6) Investigate possibilities to calibrate the crop growth model, or replace entire sub-routines of the crop growth model, using estimated leaf area index (LAI), photo-synthetically active radiation (FPAR), or estimated actual evapo-transpiration from remote sensed images. This would alleviate the need for accurate Land Unit or soil information.

6.2 Framework-specific issues

The following recommendations are listed separately for YFF2 and YFF1 as they are mutually exclusive for both yield-forecast frameworks.

6.2.1 YFF2

(1) To limit the workload, the low, middle, and high rainfall forecasts were averaged before being input into the crop growth model. Feed the crop growth model with all three datasets and average the surrogate yields afterwards.

(2) To limit the workload, other weather parameters than rainfall were not forecast. Assess the implications on the accuracy of maize yield forecasts if other crop growth input parameters are also derived from historical datasets.

(3) Introduce additional parameters than the SOI Phases to improve the explanation of rainfall variability.

(4) Detailed validation of this forecast system based on long-term forecast error, preferably based on at least 9 seasons under varying weather conditions.

Dehghan, A., 1998. *Aggregation Levels in Yield Forecasting System; Experience of Hamadan Irrigated Wheat*
Institute for Aerospace and Earth Science, Enschede, The Netherlands.

Driessen, P.M., and Konijn, N.T., 1992. *Land Use Systems Analysis.*
Wageningen Agricultural University, Wageningen, The Netherlands.

Driessen, P.M., 1997. *Biophysical Sustainability of Land-Use Systems*
Institute for Aerospace Survey and Earth Sciences, Enschede, The Netherlands.

FAO, 1978. *Agro-Ecological Zoning*
Food and Agricultural Organization, Rome, Italy.

FAO, 1992. *Land Evaluation and Farming System Analysis for Land Use Planning.*
Food and Agricultural Organization, Rome, Italy.

Fouché, H.J.1992. *Simulering van die produksiepotensiaal van veld en die kwantifisering van droogte in die sentrale Oranje-Vrystaat,* Ph.D. thesis
University of the Orange Free State Province of South Africa, South Africa.

Hammer, G., and Nicholls, N.N., 1996. *Managing Climate Variability – The Role of Seasonal Climate Forecasting in Improving Agricultural Systems*
University of Queensland Printer, Brisbane, Australia.

Hodges, T., Botner, D., Sakamoto, C., and Hays Haug, J., 1987. *Using the CERES Maize Model to Estimate Production for the U.S.*
Agricultural and Forest Methodology, Cornbelt, United States of America.

Hooijer, A.A., and Van der Wal, T, 1994. *Technical Documents 15.1, CGMS ver. 3.1.*
JRC-Joint research Center, Ispra, Italy, DLO Winand Staring Center, Wageningen, The Netherlands.

ITC, 1998. *ILWIS ver. 2.1 User's Guide*
Institute for Aerospace Survey and Earth Sciences, Enschede, The Netherlands.

Jager, J.M. de, Potgieter, A.B., and van den berg, W.J., 1998. *Framework for Forecasting the Extent and Severity of Drought in Maize in the Free State Province of South Africa*, Elsevier Scientific Publications No. 57 pp. 351-359
Elsevier Science Ltd., London, Great Britain.

Jager, J.M. de, and Singels, A., 1990. *Using Expected Gross-margin Uncertainty to Delimit Good and Poor Areas of Maize in the Semi-arid Regions of South Africa*
Applied Plant Science No. 4, pp. 25-29.

Lourens, U.W., and de Jager, J.M., 1997. *A Computerized Crop-Specific Drought Monitoring System*
Agricultural Systems No. 53, pp. 303-315.

Mahalder, B.K., and Sharifi, M.A., 1998. *User Manual for the CGMS Model*
Institute for Aerospace Survey and Earth Sciences, Enschede, The Netherlands.

McKeon, G.M., 1996. *Development of a National Drought Alert Strategic Information System,* in Proceedings of a Workshop on: Indicators of Drought Exceptional Circumstances, pp. 63-65.
Bureau of Sciences Australia, Canberra, Australia.

Meinke, H., and Hammer, G.L., 1998. *Forecasting Regional Crop Production and Climatic Risk During SOI Phases; a Case Study for the Australian Peanut Industry Australian*
Journal of Agricultural Research, Australia.

QDPI, 1995. *AUSTRALIAN RAINMAN Version 2.1 Users Guide.*
Queensland Department of Natural Resources and the Department of Primary Industries, Australia

Roebeling R.A., Rosema, A., Kashasha, D.A., Masamvu, K., and van der Harten, C.A.J., 1999. *Agromet and Crop Monitoring Project (ACMP) in the SADC region – 3.*
SADC Regional Remote Sensing Unit, Harare, Zimbabwe

Rugege, D., 1998. M.S.c. thesis
Institute for Aerospace Survey and Earth Sciences, Enschede, The Netherlands.

Sharifi, M.A., Driessen, P.M., van Keulen, H., Bronsveld M.C., Clavaux M., and Leenaars, J., 1997. *Development of a Crop Inventory and Forecasting System for the Major Agricultural Commodities in Hamadan Province, Iran, Mid Term Report*
Institute for Aerospace Survey and Earth Sciences, Enschede, The Netherlands, and ASID, Tehran, Iran.

Singh, G., and Pariyar, M.P., 1994. *Crop Yield Forecasting; a Review of Methods used in Developing Countries of Asia,* Crop Yield Forecasting Methods, Proceedings of the Seminar 24-27 Oct. 1994, pp. 115.
Office for Official Publications of the EU, Luxembourg, Luxembourg.

Stadler, R., 1997. *Harvest Preview and Yield Forecasting in Baden Wuttenberg,*
Crop Yield Forecasting Methods, Proceedings of the Seminar 24-27 Oct. 1994, pp. 205.
Office for Official Publications of the EU, Luxembourg, Luxembourg.

Stone, R.C., Hammer, G., and Marcussen, T., 1996. *Prediction of Global Rainfall Probabilities Using Phases of the Southern Oscillation Index*
Nature 384, pp. 252-255.

Supit, I., Hooijer, A.A., and Van Diepen, C.A., 1994. *System Description of the Wofost 6.0 Crop Simulation Model Implemented in CGMS.*
The Winand Staring Centre for Integrated Land, Soil and Water Research (SC– DLO), Wageningen, The Netherlands.

Swanson, E.R., and Nyankori, J.C., 1979. *Influence of Weather and Technology on Corn and Soy Bean Yield Trends.*
Agricultural Meteorology 20, pp. 327-342.

Troup, A.J. Quart. J., 1965. Roy. Meteor. Soc. 91, pp. 490-506.

Unganai S.L., 1998. *Seasonal Climate Forecast for Farm Management*, In Proceedings of the Training
Zimbabwe Meteorological Services, Ministry of Transport, and Energy, Harare, Zimbabwe

Voet, P. van der, van Diepen, C.A., and Oude Voshaar, J., 1993. *Spatial Interpolation of Daily Meteorological Data; a Knowledge based Procedure for the Region of the European Communities Report 53.3*
The Winand Staring Centre, Wageningen, The Netherlands

Vossen, P., and Rijks, D., 1995. *Early Crop Yield Assessment of the E.C. countries; the System Implementation by the Joint Research Center,* pp. 120.
Office for Official Publications of the EU, Luxembourg, Luxembourg.

Vossen P., 1994. *Early Crop Yield Assessment of National Crop Yields; the Approach Developed by the MARS-STAT Project on behalf of the European Commission*, in Crop Yield Forecasting Methods, Proceedings of the Seminar, 24-27 Oct. 1994
Office for Official Publications of the EU, Luxembourg, Luxembourg.

ZIMMET, 1998. *Proceedings of the Post Season National Climate Stakeholders Seminar*, *Apr. 1998*
Zimbabwe Meteorological Services, Ministry of Transport, and Energy, Harare, Zimbabwe

Zucchini, W., and Adamson, P.J., 1984. *The Occurrence and Severity of Droughts in South Africa*
WRC Report No. 91/1/84. Publisher unknown.

El Niño — El Niño (Spanish for Christ Child) is the name given by Peruvian fisher folk to the warming of the surface waters of the Pacific Ocean that tends to occur around Christmas. A natural event that recurs in more or less regular cycles (on average every four to five years), El Niño affects southern Africa and the Pacific from Peru to Indonesia. The local warming of the world's largest ocean also has repercussions for global atmospheric circulation of winds and waters.

ITCZ — The Inter-tropical Convergence Zone is where the moist southeast trade winds meet the northeast trades of the northern hemisphere. It is a zone of heavy rain and thunderstorms, and constitutes a main source of tropical rain.

Appendix I List of Weather Stations

WMO_NO	WMO_NAME	LATDD	LONGDD	ELEV
21797183	Alabama	-18.8667	29.8667	1140
22806651	Ballineety	-17.6	30.7833	1440
67769010	Banket Res. Stn.	-17.3	30.3833	1244
21796441	Battlefields	-18.5833	29.8	1115
22808206	Carnock	-18	30.9167	1400
21799793	Chegutu Rail/Hartley	-18.1	30.1	1190
67893020	Chibero	-18.1	30.6667	1335
67771030	Chinhoyi	-17.22	30.13	1143
67871020	Chivhu Met	-19.02	30.53	1460
22804040	Darwendale Rail	-17.43	30.333	1360
67861030	Gokwe	-18.2167	28.9333	1282
67773020	Guruve	-16.6667	30.7	1158
67775050	Harare Airport	-17.55	31.06	1497
67795010	Harare Res. Stn.	-17.48	31.03	1506
67774010	Harare/Belvedere	-17.5	31.01	1472
67785020	Henderson Res. Stn.	-17.5833	30.9667	1292
22808656	Henderson Weed East	-17.5833	30.9667	1292
22801082	Hunyani Mission	-16	30.5667	350
22816731	Impingi Ranch	-16.8667	30.75	1420
67869050	Kadoma Res. Inst.	-18.3333	29.9167	1188
67767020	Kanyemba	-15.39	30.2	340
67761060	Kariba Airport	-16.5167	28.8833	518
67765020	Karoi	-16.5	29.37	1344
67865030	Kwekwe	-18.9333	29.8333	1215
22805099	Lone Cow Estate	-17.1667	30.6333	1330
21818910	Long Valley	-17.0833	30.0333	1130
21811535	Magunje	-16.8167	29.4167	850
21810095	Makuti Tsetse	-16.3	29.2167	1070
67877070	Marondera Res. Stn.	-18.11	31.28	1631
67891010	Mhondoro	-18.19	30.36	1260
67779030	Mount Darwin	-16.47	31.35	965
21813858	Moy	-16.6333	29.6	1280
67772010	Muzarabani	-16.25	31.01	450
67789010	Mvurwi	-17.02	30.51	1481
1	Ngezi Dam	-18.6833	30.3667	1240
67793010	Rattray Arnolds Res. Stn.	-17.4	31.13	1341
67778020	Shamva Panmure	-17.16	31.37	881
22807639	Stapleford Farm	-17.7167	30.8667	1465
21818107	Yeanling	-17.0833	30	1220
20819421	Zvipani	-15.6167	30.4333	340

Appendix II Weather Stations Per Grid

Grid	Rain	Temperature	Rest
11003	67767020 67761060 22801082 21811535 67778020 67779030 21813858	67761060	67761060
11004	67767020 22801082 67761060 21811535 67779030 67778020 67773020	67761060	67761060
11005	67767020 22801082 67761060 21811535 67779030 67778020 67773020	67761060 67779030	67761060
11006	22801082 67767020 21811535 67779030 67778020 67761060 67773020	67779030 67778020 67761060	67761060 67765020 67774010
21002	67761060 67767020 21811535 22801082 67778020 21813858 67765020	67761060	67761060
21003	67761060 21811535 67767020 22801082 67778020 67779030 21813858	67761060	67761060
21004	21811535 67761060 67778020 67779030 67767020 22801082 21813858	67761060 67779030 67765020	67761060 67765020
21005	21811535 22801082 67761060 67767020 67778020 67779030 67773020	67761060 67779030 67773020	67761060
31002	67761060 21811535 67767020 22801082 67778020 21813858 67765020	67761060	67761060
31003	67761060 21811535 22801082 67767020 67778020 21813858 67765020	67761060	67761060
31004	21811535 67761060 67778020 21813858 67779030 67771030 67773020	67761060 67773020 67765020	67761060 67765020
31005	21811535 67761060 22801082 67778020 67779030 67767020 67773020	67761060 67779030 67773020	67761060 67765020 67774010
31006	67773020 67779030 67771030 67778020 67769010 21811535 22805099	67773020 67779030 67771030	67765020 67774010
41001	67761060 21811535 22801082 67767020 67778020 67771030 21813858	67761060	67761060
41002	67761060 21811535 67778020 67771030 21813858 22801082 67765020	67761060	67761060
41003	21811535 67761060 67778020 67771030 21813858 67779030 67765020	67761060 67771030 67765020	67761060 67765020
41004	21813858 67765020 67769010 67771030 22804040 22805099 67861030	67765020 67771030	67765020
41005	67771030 67773020 67769010 21813858 22805099 22804040 67869050	67771030	67765020 67774010
41006	67773020 67771030 67769010 22805099 67785020 22808656 22816731	67773020 67769010	67765020 67774010
51001	67761060 21811535 67861030 21796441 67771030 67778020 22801082	67761060	67761060
51002	21811535 67761060 67771030 67778020 21796441 67861030 67869050	67761060 67861030	67761060 67765020 67865030
51003	21811535 67761060 67771030 67778020 21796441 67869050 21813858	67761060 67771030 67861030	67761060 67765020 67865030
51004	67771030 21811535 21796441 67869050 67769010 67773020 21813858	67771030 67769010 67861030	67765020 67865030 67761060
51005	67771030 67769010 67869050 67773020 22804040 67891010 22805099	67771030	67865030 67765020 67774010
51006	22804040 22805099 22806651 22816731 67793010 22808206 67769010	67769010 67789010 67893020	67774010 67765020 67877070
61003	21811535 21796441 67861030 67778020 67869050 67771030 67761060	67861030 67761060 67865030	67761060 67865030
61004	21796441 67869050 67771030 21797183 21811535 67891010 67861030	67771030 67891010 67861030	67865030 67765020
61005	67869050 67771030 67891010 67769010 21796441 1 21797183	67771030 67769010 67865030	67865030 67765020 67774010
61006	22808206 67893020 22804040 22806651 22807639 67785020 22808656	67893020 67891010 67769010	67774010 67877070 67765020 67865030
61007	22807639 67774010 22806651 22808206 67789010 67893020 67877070	67774010 67893020 67877070	67774010
71003	21796441 67861030 21797183 67869050 21811535 67865030 67771030	67861030 67865030 67771030	67865030 67761060
71004	21796441 67869050 21797183 67865030 67771030 67861030 67891010	67865030 67771030 67861030	67865030
71005	67869050 21796441 21797183 67891010 1 67865030 67771030	67891010 67865030 67771030	67865030
71006	67893020 67891010 22808206 1 22804040 22808656 67785020	67893020 67891010	67865030 67774010 67877070
71007	22808206 67893020 22807639 22806651 67774010 67793010 22808656	67893020 67877070	67774010 67877070 67865030
81004	21796441 21797183 67865030 67869050 1 67861030 67891010	67865030	67865030
81005	67865030 67869050 21797183 1 21796441 67891010 67893020	67865030	67865030
81006	1 67893020 67891010 67865030 22808206 67869050 22808656	67893020 67891010 67865030	67865030 67877070
81007	22808206 67893020 22807639 22806651 1 67891010 67774010	67893020 67891010 67877070	67774010 67877070 67865030

Appendix III Overview Yield Forecasting Methods

Different approaches to crop forecasting according to the information they use are (Dehghan, 1998):

(a) Air (or space) - based information.

Although remote sensing techniques exist, they do not permit yet, for various reasons, the quantitative prediction, and assessments of crop yield. Using remotely sensed vegetation indices (LAI, NDVI, etc.), information about evapotranspiration, and water stress, they have the potential of being independent of any other data source and provide real-time information over vast regions.

(b) Ground-based information

There are many different methods successfully used in this approach which can be classified as:

i. Forecasts based on pre-harvest crop reports.
Infiltration, systematically, flows from villages to provincial level, reporting crop conditions. Many examples could be given here, to name just a few: crop forecasting in India, paddy yield forecasting in Indonesia (Singh & Pariyar, 1994) and crop forecasting in Germany (Stadler, 1994).

ii. Time trend analysis.
This is the most commonly used approach, and crop-weather time trend analysis models are the most successful methods used in most of the continents. The trend extrapolations adjusted for weather conditions of European Statistical Office and various methods applied in the former USSR and FAO's Early Warning System GIEWS and etc. are based on statistical regressions between weather or agro-meteorological conditions and crop yield (Vossen and Rijks 1995).

iii. Crop growth simulation models.
This is the most advanced method introduced for yield forecasting in the last decade. A yield gap is calculated by regression analysis between simulation results as potential yield and actual yield observations from the field. These models are either used to modify time trend yields, experience of CGMS (ibid.) or field observable land quality indicators (above ground mass or LAI) are collected during the growing season to relate the simulation results to actual land-use systems (Driessen, 1997), or simulation results and actual yields are directly regressed to build forecasting model.

Under the latest approach to yield forecasting, a sub-research started that primarily focused on the weather component to provide input for the crop growth simulation models. Preparing adequate surrogate seasonal weather data for crop growth modelling offers a formidable challenge (Hammer and Nicholls, 1996).

Hodges et al. (1987) cited by de Jager et al. (1998) selected appropriate analogue historical weather data series, depending upon the 90-day weather outlook (below, above or normal). Randomized weather data series generation (e.g. the climate model, Weathergen) is a possibility, as is the use of the daily rainfall data series generator of Zucchini and Adamson (1984) as cited by de Jager et al. (1998). Lourens and de Jager (1997) forecast weather data within a growing season with historical data series that had delivered lower quartile, median and upper quartile seasonal rainfall (de Jager et al., 1998). Fouché (1992) cited by de Jager et al. (1998) constructed seasonal rainfall scenarios of composite monthly rainfall data from historical meteorological records, assuming that each month received median monthly rainfall. De Jager and Singels

(1990) used combinations of daily sunshine, maximum, and minimum temperature and daily rainfall data selected randomly from historical data series (de Jager et al., 1998). McKeon (1996) sited by de Jager et al. (1998) simulated forage yields by completing the season with 5-10 analogue years of weather data from which he determined the mean and coefficients of variation.