IMPROVED DROUGHT EARLY WARNING AND FORECASTING TO STRENGTHEN PREPAREDNESS AND ADAPTATION TO DROUGHTS IN AFRICA

DEWFORA

A 7th Framework Programme Collaborative Research Project

Local scale agricultural models for Limpopo and Oum-er-Rbia river basins

WP4-D4.10

February 2013

Coordinator: Deltares, The Netherlands

Project website: www.dewfora.net

FP7 Call ENV-2010-1.3.3.1

Contract no. 265454

Page intentionally left blank

DOCUMENT INFORMATION

Title	Local scale agricultural models for Limpopo and Oum-er-Rbia river basins
Lead Author	CSIR
Contributors	CSIR, ECMWF
Distribution	<Please select on of three below> PP: Restricted to other programme participants (including the Commission Services) RE: Restricted to a group specified by the consortium (including the Commission Services) CO: Confidential, only for members of the consortium (including the Commission Services)
Reference	WP4-D4.10

DOCUMENT HISTORY

Date	Revision	Prepared by	Organisation	Approved by	Notes
03/02/2013		WA Landman	CSIR		Draft version v1
05/02/2013		WA Landman	CSIR		FINAL version

ACKNOWLEDGEMENT

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement N°265454

Page intentionally left blank

Summary

Statistical methods are applied in Work Package (WP) 4.10 of the DEWFORA project in order to develop prediction models for crop yields of respectively the Limpopo (in southern Africa) and Oum-er-Rbia river basin (in Morocco). WP4 is responsible for drought forecasting at different temporal and geographical scales, and in this case implication of drought in agriculture. Part of the work includes the preparation of ECMWF System 4 (S4) global hindcasts into the formats required for the statistical modelling described here. The statistical downscaling methods are presented, followed by some verification statistics for both basins. The main conclusion to be drawn from this work is that the ECMWF S4 global coupled model can in some cases provide the output required for commodity-orientated forecast systems for application in agriculture.

Page intentionally left blank

TABLE OF CONTENTS

List of figures

LIST OF TABLES

Page intentionally left blank

1. INTRODUCTION

This deliverable report describes a downscaling modelling system to predict seasonal crop yields over the Limpopo (southern Africa) and Oum-er-Rbia (Morocco) river basins. The crops of both basins considered here are strongly rain-fed, so the assumption is made that if a global model is able to predict seasonal rainfall over an area of interest, then the same global model's output can also be used in a statistical forecast system to predict a rain-fed commodity such as crops. The low-level circulation fields of the ECMWF System 4 (S4) are used as predictors in a principal components regression (PCR) approach to test the predictability of seasonal crop yields over the two basins. The models are tested over a 26-year period to determine their deterministic skill levels, as well as over a 16-year retro-active forecast period to test their probabilistic skill capabilities. These downscaled hindcasts are available to all partners in the project.

2. DATA AND METHODS

2.1 ECMWF S4

The global model data sets used in the downscaling work are from the ECMWF (European Centre for Medium-Range Weather Forecasts) System 4 (S4) described in WP4-D4.2. The model data have been transformed from GRIB into the format required by the statistical software package used in the analysis. Ensemble mean data are used here and 3-month averaged sea-level pressure (SLP) and 850 hPa geopotential height data are the predictors considered.

2.2 CROP DATA

Crop yield data for the Limpopo basin are yellow and white maize obtained from the South African National Department of Agriculture, Directorate: Statistics and Economic Analysis. Yields since the early 1980's are estimated from data assimilated from producers and other co-workers of the Department in the maize-producing areas of South Africa. The production figures used here are provided for three districts for white and yellow maize combined and are representative of dry land agriculture. Irrigation cultivation comprises less than 10% of maize produced in South Africa and the influence of this should therefore be small in the event of contamination of yield data for dry land cultivation. Considering the Limpopo region (Figure 1), we focus on three agricultural districts in close proximity to the southern border of the Limpopo Province, but located inside the Limpopo river basin, namely the Witbank and Middelburg agricultural districts in Mpumalanga, and the Rustenburg agricultural district in the Northwest Province. Data are available from 1981 to 2011. The maize yields are first separately detrended by fitting a second-order polynomial to each series. Harvest time is during the austral autumn.

The Oum-er-Rbia basin (Figure 1) crop yield data (durum wheat) are obtained for three regions namely the coast (2 stations), the plains (4 stations) and over the mountains (3 stations). Data are available form 1979-80 through 2007-2008. The sowing and harvesting periods are:

	Coastal areas	Plains	Mountain areas
Sowing dates	October-November	November-December	February
Harvest dates	May	May-June	August

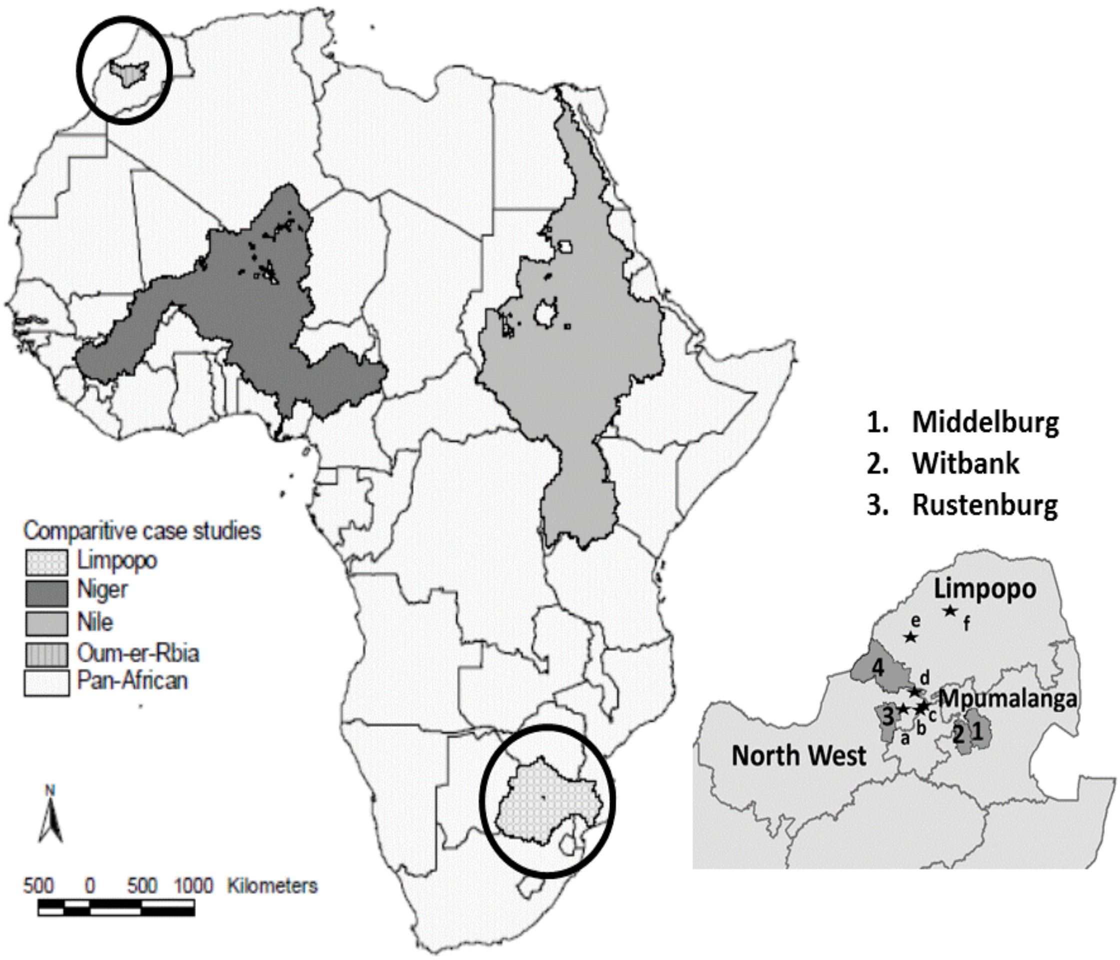

Figure 1. The location of the Limpopo and Oum-er-Rbia river basins, including a more detailed description of the location of the agricultural districts used in the Limpopo study.

2.3 STATISTICAL POST-PROCESSING: MODEL OUTPUT STATISTICS

Biases in global model seasonal outputs have been shown to be minimised through statistical post-processing (e.g. Landman and Goddard, 2002), and such processing also results in the production of forecast data directly applicable at a point of interest (Landman et al., 2012). Moreover, crop yields are not represented explicitly by global climate models, and so post-processing large-scale model output is warranted. The method used here to post-process ECMWF S4 data to crop yields is called model output statistics (MOS; Wilks, 2011). MOS equations are developed by using the principal component regression (PCR; Jolliffe, 2002) option of the Climate Predictability Tool (CPT) of the International Research Institute for Climate and Society (IRI; http://iri.columbia.edu). Prior to the PCR, yield values are transformed into approximate normal distributions. The ECMWF S4 outputs used are low-level circulation data for the three-month season prior to the period of harvesting. The assumption is that the seasonal averaged low-level circulation during that three-month period is associated with the rainfall over the region of interest and hence related to the production of dry land crops. For Limpopo DJF hindcast are used throughout, and for Oum-er-Rbia, both FMA (for coastal and plains) and MJJ (for mountains) hindcasts are used. ECMWF S4 850 hPa geopotential height fields are the Limpopo predictors, and sea-level pressure (SLP) fields the Oum-er-Rbia predictors. The hindcast fields used in the MOS equations are restricted over a domain that covers an area between the equator and 45°S and from 20°W to 60°E for the Limpopo downscaling, and from 40°N to 30°S and from 150°E to 20°W for Oum-er-Rbia downscaling (Figure 2). These domains are selected through sensitivity studies to obtain the best cross-validated skill. Moreover, the Limpopo domain has already been applied successfully in the statistical downscaling of seasonal rainfall over South Africa (Landman et al. 2012). The Oum-er-Rbia domain includes the equatorial Pacific Ocean and its Atlantic Ocean part is far enough to the north in order to include the influence from the North Atlantic Oscillation. The CPT's PCR starts by applying principal component analysis to the predictor fields, thereby eliminating the possibility of multi-colinearities in the model data and simplifying the regression calculations (Jolliffe, 2002).

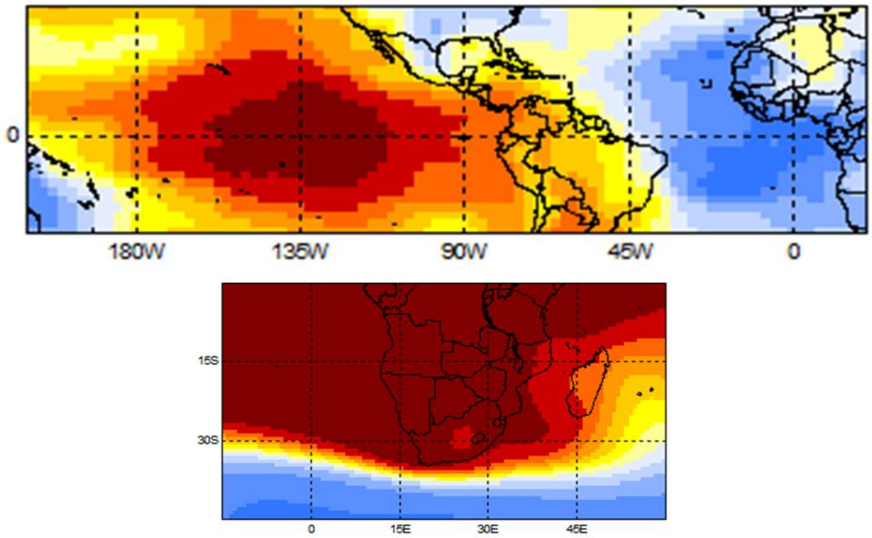

Figure 2. Top: The SLP domain used for Oum-er-Rbia yield predictions; Bottom: The 850 hPa geopotential height domain used for Limpopo yield predictions. The spatial patterns shown here are typical of the spatial loadings associated with the first principal component used in the PCR-MOS models.

Figure 3 is a schematic representation of the procedure involved in producing crop yield forecasts through MOS. The procedure involves the prediction of seasonal circulation fields by the ECMWF S4 coupled model, which are subsequently used as predictors in a MOS system that has seasonal crop yields as the predictand.

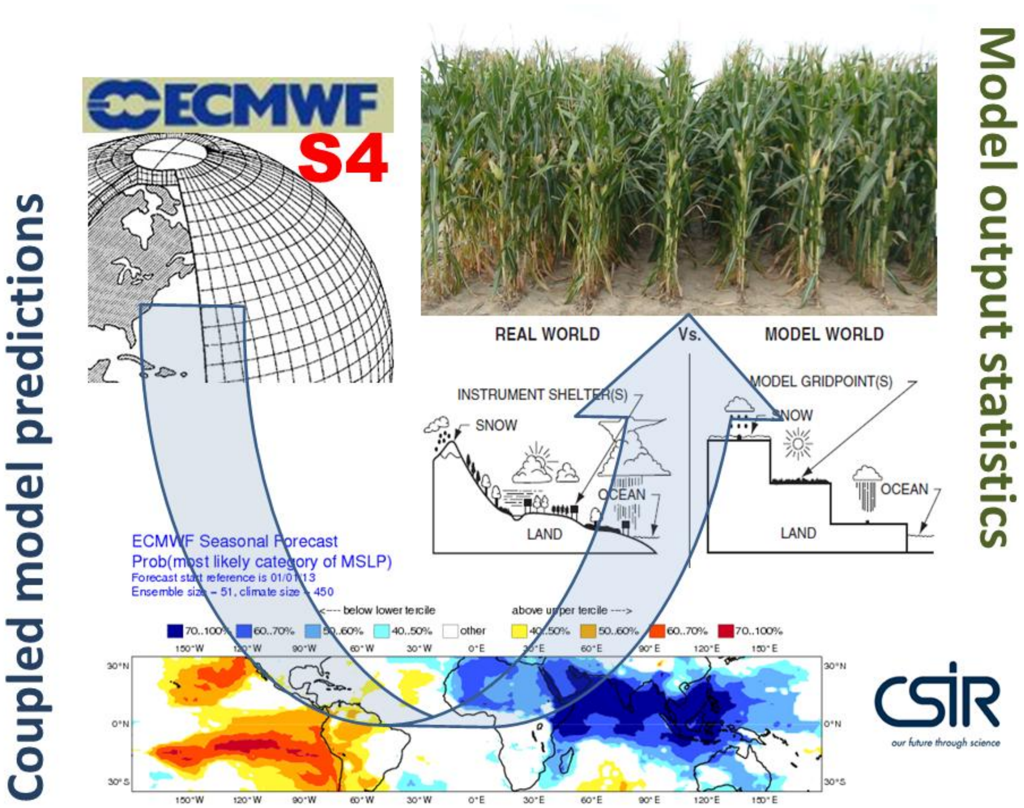

Figure 3. Schematic representation of the procedure to produce crop yield forecasts from output generated by the ECMWF S4 coupled model.

2.4 VERIFICATION

The MOS equations' ability to skilfully produce downscaled forecasts is tested both deterministically and probabilistically. Deterministic skill is determined over a 26-year period for the harvest years of 1983 to 2008. For this purpose the procedure of cross-validation is performed with a large 5-year-out design, which means that 2 years on either side of the predicted year are omitted in order to minimise the artificial inflation of skill. Deterministic forecast skill is calculated by considering Kendall's tau (Wilks, 2011), an alternative to the conventional Pearson correlation and has the additional attribute of discrimination (Jolliffe and Stephenson, 2011). In addition to cross-validation, the process of retro-active forecasting is applied over the 16-year period from 1993 to 2008 in order to produce a set of probabilistic downscaled hindcasts which are subsequently verified. In the retro-active process a cross-validation 5-year-out design is also used for the retro-active process, and the initial cross-validation period is progressively increased by 1 year at each downscaled hindcast step. Owing to the relatively small ensemble size of 15 members, the hindcasts distributions may be poorly sampled and so their uncertainties have to be estimated. Probabilistic hindcasts for the 16 years are subsequently obtained from the error variance of the 5-year-out cross-validated hindcasts using the ensemble mean (Troccoli et al., 2008). These hindcasts are tested for discrimination (to determine if the hindcasts are discernibly different given different outcomes – for example, is the forecast probability for a bumper harvest systematically higher when the event occurs than when it does not occur?) and for reliability (to determine if the confidence communicated in the hindcasts is appropriate). For calculating the former as a verification measure, the relative operating characteristic (ROC; Mason and Graham, 2002) is used, and for the latter the reliability diagram (Hamill, 1997) is used. If the area below ROC curves is ≤ 0.5, the model discriminates correctly only for less than half the time. For a maximum ROC score of 1.0, perfect discrimination has been obtained.

3. RESULTS

3.1 LIMPOPO

The cross-validated downscaled yields over the Limpopo basin are shown in Figure 4. Five lead-times, with one month in between consecutive lead-times, are considered: L0 is for forecasts initialised in December in order to produce DJF 850 hPa geopotential height fields, and L4 for forecasts initialised in August. Table 1 shows the Kendall's tau and associated statistical significance values for the three agricultural districts. The figure and table show that significant forecast skill is mainly restricted to the Rustenburg agricultural district.

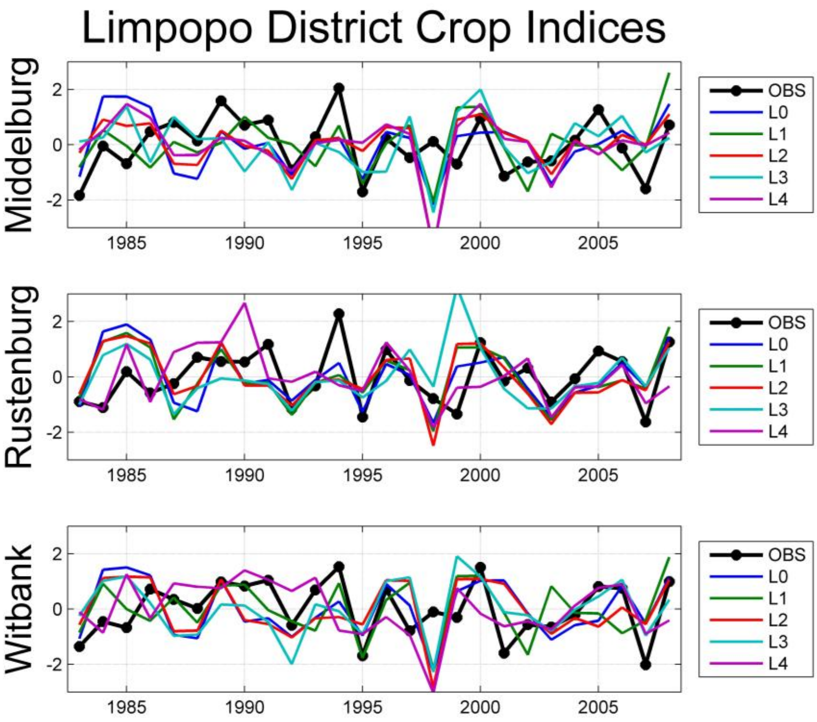

Figure 4. Agricultural district yield indices (normalised values) over the Limpopo basin vs. cross-validated (5-year-out) downscaled hindcasts for all lead times. The harvest years are shown along the x-axes.

Table 1. Kendall's tau correlations and associated levels of statistical significance for the Limpopo basin.

	Middelburg		Rustenburg		Witbank	
	Kendall τ	1 – p	Kendall's τ	1 – p	Kendall's τ	1 – p
Lead 0	0.17	0.88	0.27	0.97	0.19	0.91
Lead 1	0.28	0.98	0.24	0.96	0.26	0.96
Lead 2	0.13	0.82	0.23	0.95	0.12	0.80
Lead 3	0.11	0.78	0.22	0.94	0.18	0.89
Lead 4	0.11	0.77	0.37	0.99	0.26	0.97

The probabilistic verification outcomes for Limpopo are shown in Figures 5 (ROC) and 6 (reliability). Some discrimination has been achieved, especially for the Rustenburg agricultural district, but the ability to predict for high yields for this district is restricted to lead-times up to two months (Figure 5). Nonetheless, useful skill for Rustenburg can be seen for both high and low yields at short lead times. Figure 6 shows the reliability plots at a 1-month lead-time for the three agricultural districts. Good reliability for the prediction for low yields (a consequence of drought) can be seen for Rustenburg (middle panel), but over-confidence is found in predicting high yields for all three districts.

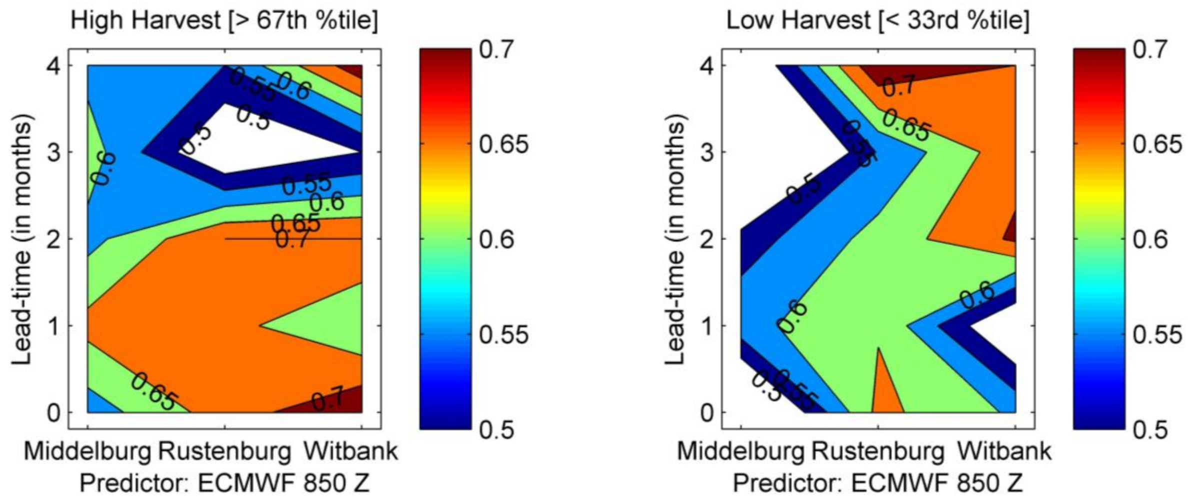

Figure 5. ROC values obtained for the Limpopo basin by retro-actively predicting high-yield (>67th percentile of the climatological yield record) and low-yield (<33rd percentile) seasons probabilistically over 16 years (1993 to 2008). The x-axes show the names of the agricultural districts of the basin considered here, and the y-axes show the forecast lead times in months.

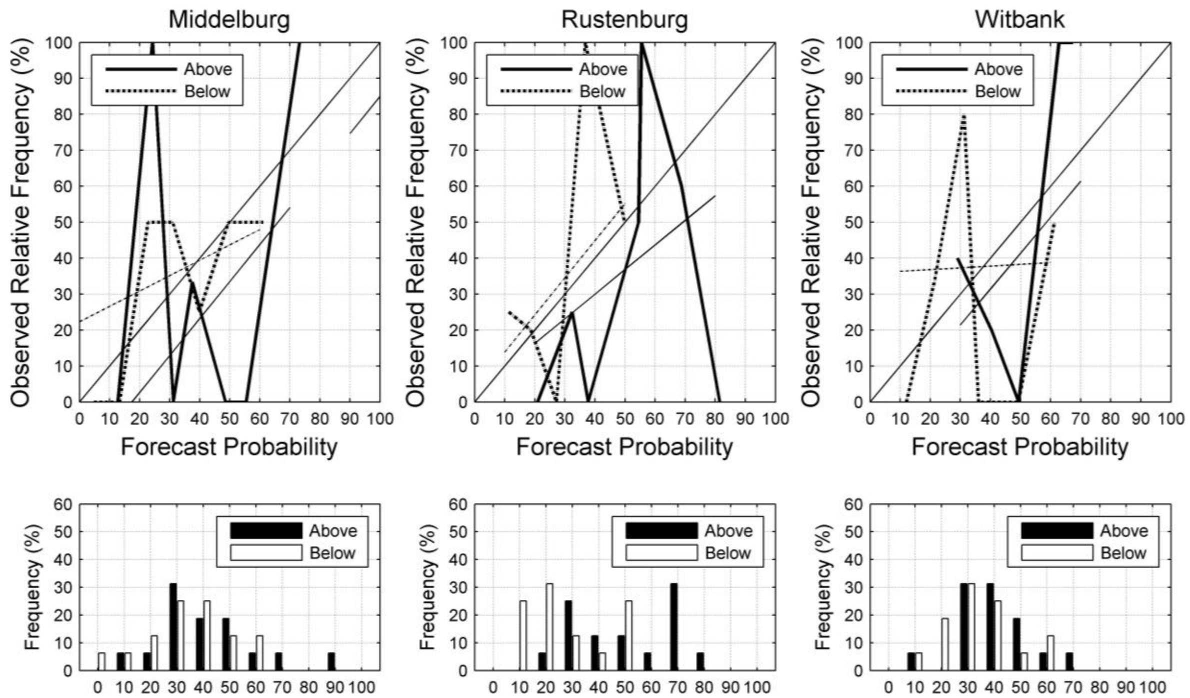

Figure 6. Reliability diagram and frequency histogram for high (>67th percentile) and low (<33rd percentile) yields for the three Limpopo agricultural districts at a 1-month lead-time obtained from 16 years of hindcasts produced by downscaling the ECMWF S4 coupled model. The thin solid (dashed) line is the weighted least squares regression line of the high (low) yield reliability curve. Black and white bars of the frequency histogram are respectively for high and for low yields.

3.2 OUM-ER-RBIA

The ECMEF S4 domain selected for the Oum-er-Rbia basin was chosen such that it considered forcings from both the Atlantic and Pacific Oceans. This selection allows the downscaling models to include influences from both oceans. Moreover, sensitivity tests performed with downscaled ECMWF S4 model data showed that this domain produced better results than considering each ocean domain on its own. The 5-year-out cross-validated downscaled yields over the Oum-er-Rbia basin are shown in Figure 7. The downscaled results for all five lead-times are presented, but take note that the initialization months are different since the predicted circulation fields is a function of the harvest times which vary for this basin. Table 2 shows the Kendall's tau and associated statistical significance values for the coastal, plains and mountain regions respectively. The figure and table show that some skill may be found for the mountains and coastal areas, but very low predictability is seen over the plains of the basin.

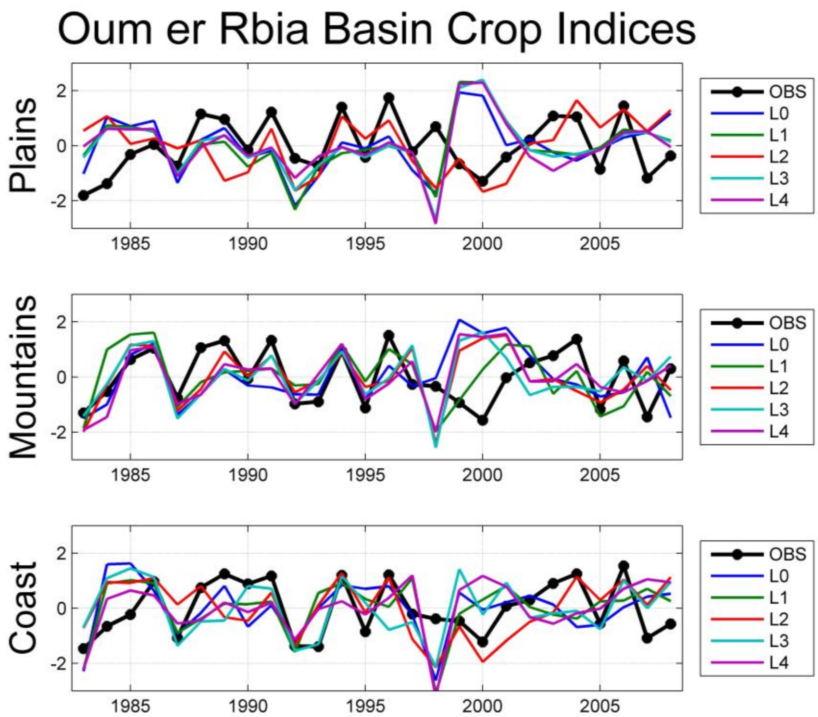

Figure 7. Regional yield indices (normalised values) over the Oum-er-Rbia basin vs. cross-validated (5-year-out) downscaled hindcasts for all lead-times. The harvest years are shown along the x-axes.

Table 2. Kendall's tau correlations and associated levels of statistical significance for the Oum-er-Rbia basin.

	Plains		Mountains		Coast	
	Kendall τ	1 – p	Kendall's τ	1 – p	Kendall's τ	1 – p
Lead 0	0.01	0.52	0.18	0.89	0.20	0.92
Lead 1	-0.11	0.22	0.29	0.98	0.05	0.62
Lead 2	0.18	0.89	0.12	0.80	0.34	0.99
Lead 3	-0.03	0.40	0.13	0.81	0.22	0.94
Lead 4	-0.14	0.16	0.15	0.85	0.08	0.70

Probabilistic skill estimates (Figures 8 and 9) for the Oum-er-Rbia further demonstrate the potential for making yield predictions over the mountains and over the coastal areas. Take note of the poor skill once again found over the plains. Good discrimination (Figure 8) is found for especially the mountain areas and for high yields in particular. At a 2-month lead-time, high and low yields are well discriminated for both the coastal and mountain areas, and so we will present a reliability analysis of the forecast system at this lead time only. Good reliability is shown in Figure 9 for predicting low yields over the mountain and coastal areas, but the prediction of high yields has been found to be over-confident.

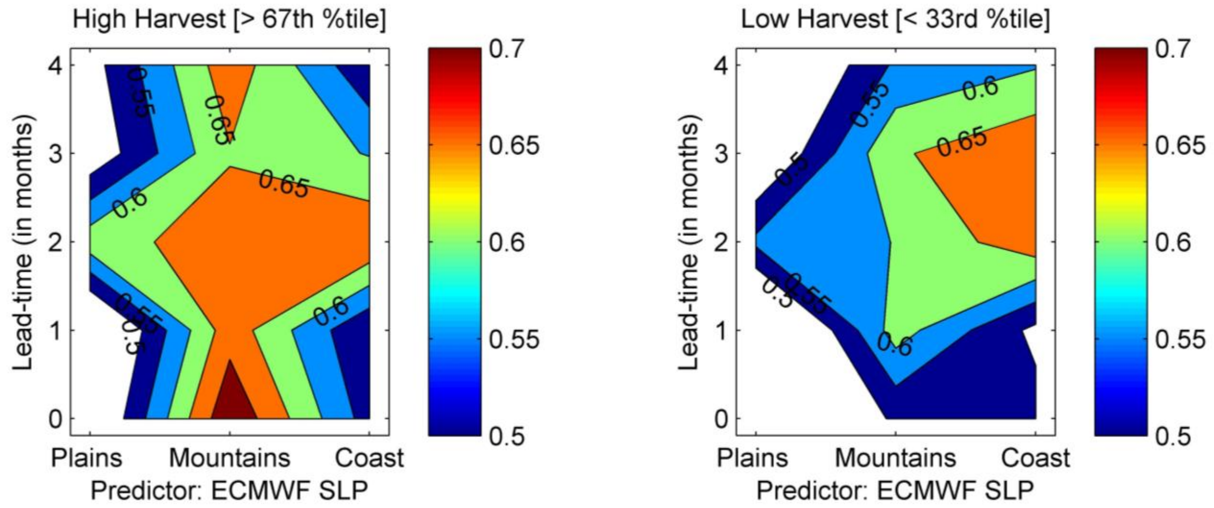

Figure 8. ROC values obtained for the Oum-er-Rbia basin by retro-actively predicting high-yield (>67th percentile of the climatological record) and low-yield (<33rd percentile) seasons probabilistically over 16 years (1993 to 2008). The x-axes show the areas considered, and the y-axes show the forecast lead times in months.

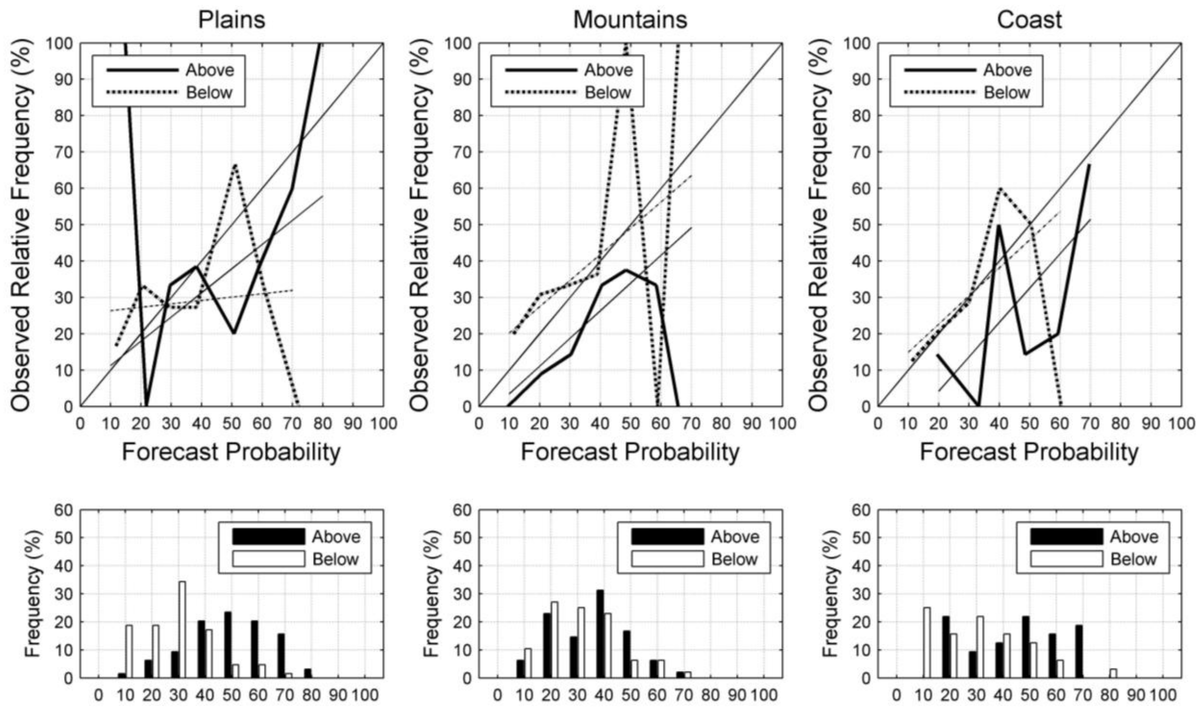

Figure 9. Reliability diagram and frequency histogram for high (>67th percentile) and low (<33rd percentile) yields for the three Oum-er-Rbia areas at a 2-month lead-time obtained from 16 years of hindcasts produced by downscaling the ECMWF S4 coupled model. The thin solid (dashed) line is the weighted least squares regression line of the high (low) yield reliability curve. Black and white bars of the frequency histogram are respectively for high and low yields.

4. CONCLUSIONS

This report introduced a statistical-dynamical approach to estimate crop yields over the Limpopo basin of southern Africa and the Oum-er-Rbia basin in Morocco. Low-level circulation data of the ECMWF S4, produced at lead-times up to 4 months, were statistically downscaled to archived seasonal yields. Both deterministic (over 26 years) and probabilistic (over 16 years) verification was performed at all available lead-times. The results are encouraging for certain areas, and so this report has demonstrated the ability to develop relatively simple commodity-orientated forecast systems for application in agriculture, especially for dry land farmers. However, more sophisticated methods to model crops remain warranted. The results presented here may also be considered as a baseline that needs to be outscored by such sophisticated approaches. Examples of these approaches include the use of physical crop models that assimilate output from global climate models on temporal and spatial scales reconcilable with their requirements.

More expansive verification results of this forecasting system are presented in D4.11.

5. REFERENCES

Hamill T. 1997. Reliability diagrams for multicategory probabilistic forecasts. Weather Forecasting 12: 736-741.

Jolliffe IT. 2002. Principal Component Analysis, Second Edition. Springer, New York.

Jolliffe, IT and Stephenson DB. 2011. Forecast Verification: A Practitioner's Guide in Atmospheric Sciences, Second Edition. Wiley, Amsterdam.

Landman WA, and Goddard L. 2002. Statistical recalibration of GCM forecasts over southern Africa using model output statistics. Journal of Climate 15: 2038-2055.

Landman WA, DeWitt D, Lee D-E, Beraki, A and Lötter D. 2012. Seasonal rainfall prediction skill over South Africa: 1- vs. 2-tiered forecasting systems. Weather and Forecasting 27: 489-501. doi: 10.1175/WAF-D-11-00078.1.

Mason SJ and Graham NE. 2002. Areas beneath the relative operating characteristics (ROC) and relative operating levels (ROL) curves: Statistical significance and interpretation. Quarterly Journal of the Royal Meteorological Society 128: 2145-2166.

Troccoli, A., Harrison, M., Anderson, D.L.T. and Mason S.J. (2008). Seasonal Climate: Forecasting and Managing Risk. NATO Science Series on Earth and Environmental Sciences, Vol 82. Springer, New York.

Wilks DS. 2011. Statistical Methods in the Atmospheric Sciences, Third Edition. Academic Press, Amsterdam.