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

DEWFORA

A 7th Framework Programme Collaborative Research Project

# Skill of developed regional hydrological models to forecast hydrological drought

**WP4-D4.8**

**June 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 | Skill of developed regional hydrological models to forecast hydrological drought |
| Lead Author | UNESCO-IHE |
| Contributors | PIK, NFC, GFZ |
| Distribution | <Please select on of three below><br>PP: Restricted to other programme participants (including the Commission Services)<br>RE: Restricted to a group specified by the consortium (including the Commission Services)<br>CO: Confidential, only for members of the consortium (including the Commission Services) |
| Reference | WP4-D4.8 |

## DOCUMENT HISTORY

| Date | Revision | Prepared by | Organisation | Approved by | Notes |
|---|---|---|---|---|---|
| <27/03/2013> | 1 | P. Trambauer | UNESCO-IHE | | |
| <14/05/2013> | 2 | S. Fournet | PIK | | |
| <17/05/2013> | 3 | P. Trambauer | UNESCO-IHE | | |
| <22/05/2013> | 4 | D. Amin | NFC | | |
| <29/05/2013> | 5 | P. Trambauer | UNESCO-IHE | | |
| <30/05/2013> | 6 | M. Seibert | GFZ | | |
| <30/05/2013> | 7 | P. Trambauer | UNESCO-IHE | | |
| <30/05/2013> | 8 | S. Maskey | UNESCO-IHE | | |
| <06/06/2013> | 9 | E. Dutra | ECMWF | | |
| <11/06/2013> | 10 | M. Werner | Deltares | | |
| <13/06/2013> | 11 | S. Maskey | UNESCO-IHE | | |
| <13/06/2013> | 12 | P. Trambauer | UNESCO-IHE | | |

## 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

## SUMMARY

This study explores the potential to improve and supplement hydrological drought early warning systems for the Limpopo, Niger and Nile case studies. For the Limpopo case study we present two different approaches for hydrological drought forecasting: a statistical seasonal forecast and a dynamic multi ensemble seasonal forecast. The statistical models are set up for a three month lead time forecasting based on teleconnected ocean regions. Multiple linear models and an artificial neural network model are compared for all stations. The artificial neural network model shows far better performance in fitting the calibration data. However, by cross-validation it is shown that the multiple linear models are more robust and hence more suitable for prediction. However, the prediction skill of these models is still low. The best results are achieved for the station at Combomume in Mozambique which is drains a large catchment of the upper Limpopo. The dynamic multi ensemble seasonal forecast for the Limpopo river basin, which is based on a process-based distributed hydrological model forced with ECMWF seasonal forecasts, show higher skill in predicting low flows and high flows than in predicting near normal conditions. The predictive skills of these forecasts, which worsen with lead time, are higher than using the climatology. The evaluation of the forecasts for selected seasons when a drought or flood actually occurred suggests that the hydrological forecast using ECMWF S4 seasonal meteorological forecast is improved and offers a clearer signal compared to the use of the ensemble streamflow prediction (ESP). These selected seasons suggest that seasonal forecasts based on numerical weather prediction models can offer better seasonal hydrological forecasts than ESP.

For the Niger case study, this report presents OPIDIN (Outil de Prédiction de l´inondation dans le Delta Intérieur du Niger), a tool product of simple statistical relationship based on historical water height time series monitored at different gauged stations along the Inner Niger Delta. OPIDIN demonstrates less skill to predict extremes and tends then to overestimate (for dry episodes) or underestimate (wet periods) flood peak water level. It presents better skill in capturing extreme events for shorter lead times. Moreover, the tool shows to better capture the flood peak water level than the flood peak timing.

The Nile Forecast System (NFS) is a near real-time distributed hydro-meteorological forecast system designed for forecasting Nile flows at designated key points along the Nile; of major interest is the inflow of the Nile into the High Aswan Dam, Egypt. The NFS is applied within this study with ECMWF meteorological seasonal forecast. The results show that the ensemble mean is on average better than any individual ensemble member.

# TABLE OF CONTENTS

**1. INTRODUCTION ..... 1**

**2. METEOROLOGICAL FORECASTS DATASETS ..... 3**

2.1 ECMWF-SEASONAL FORECASTS ..... 3

2.2 CLIMATOLOGICAL BIAS CORRECTION OF SEASONAL FORECASTS OF PRECIPITATION ..... 3

**3. SEASONAL FORECAST OF HYDROLOGICAL DROUGHTS FOR THE LIMPOPO CASE STUDY BASIN ..... 5**

3.1 STATISTICAL FORECASTING ..... 5
3.1.1 Model and simulations set-up ..... 5
3.1.2 Results ..... 10
3.1.3 Conclusions ..... 24

3.2 DYNAMIC MULTI-MODEL ENSEMBLE FORECAST ..... 25
3.2.1 Model and simulations set-up ..... 25
3.2.2 Results ..... 28
3.2.3 Concluding remarks ..... 40

**4. FORECAST OF HYDROLOGICAL DROUGHTS FOR THE NIGER CASE STUDY BASIN 41**

4.1 MODEL AND SIMULATIONS SET-UP ..... 41
4.1.1 Description of OPIDIN platform ..... 41
4.1.2 Description of the OPIDIN tool (initial version) ..... 47
4.1.3 Forecast skill assessment method ..... 54

4.2 RESULTS ..... 56
4.2.1 Retroactive forecast skill ..... 56

**5. FORECAST OF HYDROLOGICAL DROUGHTS FOR THE NILE CASE STUDY BASIN... 63**

5.1 MODELS AND SIMULATIONS SET-UP ..... 63
5.1.1 Satellite Precipitation Estimation ..... 65
5.1.2 Description of the regional hydrological model ..... 67
5.1.3 Meteorological forecasts used ..... 67
5.1.4 Initial conditions for the forecasts ..... 67
5.1.5 Time period of the simulations ..... 69
5.1.6 Bias correction methods of seasonal forecasts ..... 70

5.2 RESULTS ..... 71

5.3 RETROACTIVE FORECAST SKILL ..... 73

5.3.1 NFS Evaluation ..... 75
5.3.2 Skill scores of the results ..... 76

**6. REFERENCES ..... 81**

**7. APPENDICES ..... 85**

7.1 NIGER CASE STUDY ..... 85

## LIST OF FIGURES

Figure 3-1: Priorities of runoff gauges (GRDC, DWAF and ARA-Sul) based on length of record and catchment size. Priorities 1,2 and 3 are marked by the green, orange and red triangles, respectively. Grey marked gauges had no priority due to very short records. ..... 10
Figure 3-2: Köppen-Geiger Climate classification and Limpopo basin with main tributaries (Kottek et al.,2006). ..... 10
Figure 3-3 Composite 2 of consecutive sea surface temperature anomalies during droughts in the Limpopo ..... 11
Figure 3-4 Composite 3 with sea surface temperature anomalies of the 12 months preceding a drought in October to December (left figure) and January to March (right figure). ..... 12
Figure 3-5: Teleconnected regions established by correlation and composites analysis. ..... 13
Figure 3-6: Multiple linear models: Estimated predictors for the AIC selected models at the four runoff stations. ..... 16
Figure 3-7: Multiple linear models: Estimated explicative contribution of predictors for the AIC selected models at the runoff stations. ..... 17
Figure 3-8: Botswana: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10). ..... 18
Figure 3-9: Loskop Noord (Olifants): Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10). ..... 18
Figure 3-10: Combomume: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10). ..... 19
Figure 3-11: Chókwè: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10). ..... 19
Figure 3-12 Botswana: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange). ..... 21
Figure 3-13 Loskop Noord: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange). ..... 21
Figure 3-14 Combomume: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest

probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange). ..... 22
Figure 3-15 Relative operating characteristics plot of the MLM model for stations Combomume (left) and Chókwè (right). ..... 23
Figure 3-16 Chókwè: Early warning with the multiple linear model. Above: simulation (grey line) and observed values (black line) with early warnings three months ahead for the class of highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange). ..... 23
Figure 3-17 Approach followed in this forecasting system for the Limpopo river basin ..... 25
Figure 3-18 Customized Delft-FEWS for the Limpopo river basin ..... 26
Figure 3-19 Seasonal meteorological forecast data imported in Delft-FEWS ..... 27
Figure 3-20 Location of hydrometric station 24 in the Limpopo basin ..... 28
Figure 3-21 ROC diagram for the average seasonal runoff (Oct- March) in station 24 for a lead time of zero months. Panel a) shows the original ROC diagram and panel b) shows a "binormal approximation" whereby the POD and POFD are assumed to be bivariate normally distributed29
Figure 3-22 ROC diagrams for the average 3-months runoff (Dec- Feb) in station 24 for a lead time of: a) zero months, b) 1 month, c) 2 months, and d) 3 months ..... 30
Figure 3-23 ROC scores for the average 3-months runoff (Dec- Feb) in station 24 for different thresholds and different lead times ..... 31
Figure 3-24 a)Brier Score and b) Brier Skill score for the average 3-months runoff (Dec- Feb) in station 24 for different thresholds and lead times ..... 31
Figure 3-25 Correlation coefficient (CC) and Relative Mean Error (RME) for each forecast month as a function of lead time ..... 32
Figure 3-26 ROC diagrams for monthly runoff (verification period: October to March) in station 24 for a lead time of: a) zero months, b) 1 month, c) 2 months, d) 3 months, e) 4 months, f) 5 months, and g) 6 months ..... 33
Figure 3-27 Mean rank of the forecasted water level anomaly ..... 35
Figure 3-28 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1991/92. The shaded areas represent the ensemble distribution between the percentiles 30 and 70. ..... 37
Figure 3-29 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1999/2000. The shaded areas represent the ensemble distribution between the percentiles 30 and 70. ..... 38
Figure 3-30 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1982/83. The shaded areas represent the ensemble distribution between the percentiles 30 and 70 ..... 39
Figure 4-1 The floodplains (light blue) and permanent water bodies (dark blue) of the Inner Niger Delta, as indicated on the topographical maps of the Institut Géographique National (IGN). The maps are from 1956, and based on aerial photographs and field work in the early 1950s, a period with very high flood; source: Zwarts, 2009. ..... 41
Figure 4-2: cattle passing the Niger river @WetlandsInternationalSévaré/A&W ..... 43
Figure 4-3: DEWFORA/AFROMAISON workshop in July 2012 grouping some of the key regional and local stakeholders of the OPIDIN platform @WetlandsInternationalSévaré ..... 44
Figure 4-4: *The scale at Akka* @WetlandsInternationalSévaré/A&W ..... 48

Figures 4-5 and 4-6: Water level time series in cm.at Mopti and Akka from 1956 to 2007 (from darkblue to lightblue); source: DNH ..... 49
Figures 4-7 and 4-8: Average, minimum and maximum daily change in water level at Mopti and Akka in cm. (based on 1956-2007 time series); source: DNH ..... 49
Figures 4-9 and 4-10: At Mopti from Mopti ..... 50
Figures 4-11 and 4-12: At Akka from Akka ..... 50
Figures 4-13 and 4-14: At Akka from Mopti ..... 51
Figure 4-15 Frequency distribution expressed in percentage of the annual flood peak water level (left column) and of the annual flood peak timing (right column) ..... 57
Figure 4-16 Quantitative deviation errors frequencies expressed in percentage between the retrospective OPIDIN predictions of the annual flood peak water level (left column) and of the annual flood peak timing (right column) ..... 58
Figure 4-17 Quantitative deviation errors frequencies expressed in percentage ..... 59
Figure 4-18 Scores between the retrospective OPIDIN predictions of the annual flood peak water level (right column) and of the annual flood peak timing (left column) ..... 60
Figure 4-19 Skill of OPIDIN when predicting peak water level in the range “Really Low” and peak timing in the range “Really Early” ..... 61
Figure 4-20 skill of the initial version of OPIDIN using $R^2$ calculation ..... 62
Figure 5-1 Schematic of the Nile Forecast System ..... 64
Figure 5-2 Hydrological Models at the Pixel Scale ..... 67
Figure 5-3 Differences between the Corrected precipitation over the Eastren Nile by the bias correction method and the Original forecasted Precipitation ..... 71
Figure 5-4 PET calculated by the forecasted ensembles compared to the NFS fixed PET ..... 72
Figure 5-5 Simulations Flow in the Eastern Nile using the forecasted PET comparing to the simulations flow using NFS_PET ..... 73
Figure 5-6 Skill Score criteria at Atbara Basin ..... 77
Figure 5-7 Skill Score criteria at Diem Basin ..... 78
Figure 5-8 Skill Score criteria at Khartoum Basin ..... 79
Figure 7-1 Annual flood peak prediction for dry and wet episode at Akka from Akka using classes of quantitative deviation ..... 85
Figure 7-2 Annual flood peak prediction for dry and wet episode at Akka from Mopti using classes of quantitative deviation ..... 86
Figure 7-3: Annual flood peak prediction at Mopti from Mopti using classes of quantitative deviation 87
Figure 7-4: Annual flood peak prediction at Akka from Akka using classes of quantitative deviation 87
Figure 7-5: Annual flood peak prediction at Akka from Mopti using classes of quantitative deviation 87
Figure 7-6: Annual flood peak prediction at Mopti from Mopti using early warning classes ..... 88
Figure 7-7: Annual flood peak prediction at Akka from Akka using early warning classes ..... 88
Figure 7-8: Annual flood peak prediction at Akka from Mopti using early warning classes ..... 88
Figure 7-9 Annual deflood (200cm.) prediction for all, dry and wet episode at Akka from Akka using classes of quantitative deviation ..... 89
Figure 7-10 Annual deflood (200cm.) prediction for all, dry and wet episode at Akka from Mopti using classes of quantitative deviation ..... 90

Figure 7-11: Annual deflood timing (200cm.) at Akka from Akka using classes of quantitative deviation ..... 91
Figure 7-12: Annual deflood timing (200cm.) at Akka from Mopti using classes of quantitative deviation ..... 91

## LIST OF TABLES

Table 3-1: Selection criteria for the composites: the Limpopo SPI threshold regarded as drought (SPI), the temporal resolution of a composite (time resolution), the relation of composite time to drought occurrence time (time relation), which can be contemporary or with lead time. ..... 9
Table 3-2: Priority definition criteria for runoff gauges in the Limpopo basin. ..... 9
Table 3-3: Climate indexes considered for prediction. ..... 12
Table 3-4: Potential predictors for precipitation variability in the Limpopo basin, detected by composites and correlation analysis. Parameters are sea surface temperature (SST) and sea level pressure (SLP), the types of the reference composites (see Table 3-1) are indicated in the last column with the months and respective lead time. ..... 13
Table 3-5: Coefficients of determination ($R^2$) for multiple linear models and artificial neural networks. Linear models were established by forward selection with Akaike's information criterion (AIC) and Bayes' information criterion (BIC) for June predictors (J) and predictors aggregated over April-June (AMJ). Neural networks were established using the same predictors as the AIC selected AMJ model using two degrees of complexity with 10 neurons (ANN-10) or 3 neurons (ANN-3) in the hidden layer. ..... 14
Table 3-6: Multiple linear models: Significances of selected predictors for the AIC selected model with AMJ predictors. ..... 15
Table 3-7: Root mean squared errors from calibration data and from leave-one-out cross validation. 20
Table 4-1 Early warning classes used to test the initial version of OPIDIN ..... 52
Table 4-2 Annual time series used to assess the retroactive forecast skill for dry and wet episods 55
Table 4-3 Classes used to assess the skill of OPIDIN in term of quantitative deviation ..... 55
Table 4-4 Standard deviation of Mopti and Akka time series (1956-2007) in days for the peak flood and deflood (200cm.) timing and in cm. for the peak flood water level ..... 55
Table5-1 Performance Criteria Definitions ..... 75
Table 5-2 Summary Statistics for Key Locations along the Eastern Nile, *NFS Evaluation Report, 2011.* ..... 76

## LIST OF ACRONYMS

| | |
|---|---|
| A&W | Altenburg & Wymenga Ecological Consultants |
| ABFN | Agence régionale du Bassin du Fleuve Niger |
| AIC | Akaike's Information criterion |
| ANN-GA | Artificial neural network – genetic algorithm |
| BIC | Bayesian information criterion |
| Delft-FEWS | Delft-Flood Early Warning System |
| DRH | Direction Nationale de l´Hydraulique du Mali |
| DNH | Direction Régionale de l´Hydraulique |
| DWAF | Department of Water Affairs |
| ECMWF | European Centre for Medium-Range Weather Forecasts |
| ERAI | ERA-Interim |
| ESP | Ensemble Streamflow Prediction |
| EVS | Ensemble Verification System |
| GIRE | Gestion integrée de resources naturelles |
| GPCP | Global Precipitation Climatology Project |
| GRDC | Global Runoff Data Centre |
| IGN | Institut Géographique National |
| IND | Inner Niger Delta |
| JFM | January, February, March |
| MLM | Multiple linear model |
| OND | October, November, December |
| OPIDIN | Outil de Prédiction de l´inondation dans le Delta Intérieur du Niger |
| PCR-GLOBWB | PCRaster GLOBal Water Balance Model |
| POD | Probability of detection |
| POFD | Probability of false detection |
| RMSE | Root mean squared error |
| RMSECV | Cross validated root mean squared error |
| ROC | Relative operating characteristic |
| ROCS | ROC score |
| R2 | Coefficient of determination |
| SPPT | Stochastically perturbed parameterized tendency |
| SPS | Stochastic back-scatter scheme |

| | |
|---|---|
| SSTs | Sea Surface Temperatures |
| UICN | International Union for Conservation of Nature |
| WLA | Water level anomaly |

# 1. INTRODUCTION

Preparedness to droughts can be enhanced by providing timely warnings based on medium range to seasonal drought forecasts. Warnings issued several months ahead of a drought event can trigger action plans to mitigate drought and to enhance preparedness. An early warning system relies upon forecasting systems which provide the necessary information. The information required by the decision makers varies from case to case depending on the type of mitigations or preparedness required. For example, in the Limpopo river basin irrigation is the major water use. Hence, forecasts of seasonal water level anomalies in its rivers (that can be expressed in terms of a standardized runoff index) and reservoirs can give an indication to the water managers of the percentage of irrigation demand that can be supplied in the summer season. In the Niger basin, the regular floods in the inner Niger delta are of prime importance as the economy and the ecosystem of the delta are strongly driven by the dynamics of these floods. A seasonal forecast of the flood extent is therefore key information which can support stakeholders to anticipate their economic activities for the coming season and possibly reduce their livelihood risks in the delta. For the Nile case study, a skilful seasonal prediction of the inflow of the Nile into the High Aswan Dam, Egypt, is highly valuable.

This report presents the evaluation of the performance of the hydrological models in seasonal hydrological forecasts for the Limpopo (Chapter 3), Niger (Chapter 4) and Nile (Chapter 5) case studies. The hydrological models used in these case studies use the seasonal meteorological forecasts from the European Centre for Medium Range Weather forecasting (ECMWF) as inputs.

# 2. METEOROLOGICAL FORECASTS DATASETS

Hydrological predictions are generally carried out by forcing hydrological models with ensembles of atmospheric forecasts produced by climate models. This section describes briefly the meteorological forecast product that is used in the hydrological forecasts of the Limpopo and the Nile case studies presented in the subsequent sections. This meteorological product is provided by the project partner ECMWF and is described at length in the previous deliverable 4.1 (DEWFORA 2011c).

## 2.1 ECMWF-SEASONAL FORECASTS

The most recent seasonal forecasting system at ECMWF (system 4 - S4) was made available to the different modelling groups (this system is an updated version of the forecast system described in (DEWFORA 2011c)). S4 became operational in November 2011 issuing 51 ensemble members with six months lead time. The atmospheric resolution is about 79 km with 91 vertical levels, and is fully coupled with an ocean model with a horizontal resolution of 1°. The initial perturbations are defined with a combination of atmospheric singular vectors and an ensemble of ocean analysis. Atmosphere model uncertainties are simulated using a 3-time level stochastically perturbed parameterized tendency (SPPT) scheme and the stochastic back-scatter scheme (SPS) that are also operational in the current ECMWF medium-range ensemble prediction system. The hindcast set is provided for calibration, covering a period of 30 years (1981 to 2010) with the same configuration as the operational forecasts but only with 15 ensemble members. Molteni et al. (2011) presents an overview of S4 model biases and forecasts performance, and Dutra et al. (2012) (see also DEWFORA deliverable D4.3) presents an evaluation of S4 in seasonal forecasts of meteorological droughts in several African Basins.

## 2.2 CLIMATOLOGICAL BIAS CORRECTION OF SEASONAL FORECASTS OF PRECIPITATION

Mean biases and drifts in the seasonal forecasts of precipitation can have a detrimental influence in the hydrological forecasts. Therefore, a simple climatological bias correction, based on monthly means, was applied to the seasonal forecasts in the form:

$$P'_{m,l} = \alpha_{m,l} P_{m,l} \quad (1)$$

where *P and P'* are the original and corrected seasonal forecasts of precipitation, respectively, $\alpha$ is a multiplicative correction factor and the subscripts *m* and *l* are the calendar month (1 to 12, of the initial forecast date) and lead time (0 to 5 months), respectively. The correction factor is given by the ratio:

$$\alpha_{m,l} = \overline{P}^{base}_{m*} / \overline{P}_{m,l} \quad (2)$$

where $\overline{P}_{m*}^{base}$ is the multi-annual mean of precipitation of the base dataset (observations, or reanalysis) for a particular calendar month *m** (*m*=m+l)*, and $\overline{P}_{m,l}$ is the multi-annual and ensemble mean of the forecasts for a particular month *m* and lead time *l*. The correction factor *α* is limited between a reasonable range (0.1 and 10), and is linearly interpolated from monthly values to daily assuming it corresponds to day 15 of the particular month, and equation (1) is applied to daily values of precipitation. This is a simple bias correction that only guarantees that the mean forecast climate is similar to the base dataset, not addressing other problems of the forecasts such as inter-annual variability, ensemble spread or daily variability.

# 3. SEASONAL FORECAST OF HYDROLOGICAL DROUGHTS FOR THE LIMPOPO CASE STUDY BASIN

Various methods are available for hydrological drought forecasting. For the Limpopo case study we hereby present two different approaches: the first is a statistical seasonal forecast and the second is a dynamic multi ensemble seasonal forecast.

## 3.1 STATISTICAL FORECASTING

### 3.1.1 Model and simulations set-up

The statistical forecasting approach followed here employs two different methods: multiple linear models and artificial neural networks (ANN). All these methods have in common that they are able to relate multiple predictors (input) to one predictand (output) variable[1]. Hence, the first step was to identify potential predictors for drought in the Limpopo basin. Second, best predictors were chosen. Third, the models were set up with these predictors and drought forecasting performance was compared.

For the identification of potential predictors a number of climate and circulation anomaly indices were compiled. In addition, teleconnected sea surface regions were selected based on preliminary work in deliverable D4.6 (Seibert and Apel 2013). In the deliverable 4.6 on "Statistical analysis of drought variability and hydrological parameters" the relationship of drought variability and sea surface temperature was analysed. Regions that showed high correlations were selected and the spatial fields were averaged to obtain time series of SST anomalies. The correlation analysis highlighted regions which show a general relationship with precipitation, but are not specific to drought. This more drought specific focus of analysis was addressed by composite analysis based on ERA-Interim data of SST anomalies, wind anomalies, moisture anomalies and sea level pressure anomalies. First, consecutive monthly composites were calculated. Second, composites were calculated for conditions during the 12 months preceding drought in early rainy season (OND) and late rainy season (JFM). See Table 3-1 for an overview of the applied composite definitions. Here, the analysis was focused on the rainy season. Droughts were defined as moments in time, where the all Limpopo SPI was below a threshold of -0.8, which equals the 21 % lowest periods ($P_{standardnormal}(q< -0.8)$). The compilation of all potential predictors is listed in Table 3-1.

The set of predictors contained several non-independent predictors. Hence it was affected by collinearity. For example, this was the case for the different predictors related to the El-Nino phenomenon. Hence, predictors had to be selected carefully and models had to be validated to check for overfitting. We followed a widely established approach which has also been applied by Diro, Black & Grimes (2008)for seasonal forecasting of Ethiopian rainfall. The approach is a stepwise selection of predictors from the set of potential predictors for a

[1] ANN can also be adapted to more than one output variable

multiple linear model. The algorithm as it is implemented in the “stats” package for the statistical programming environment (R Core Team 2012) was applied in the following manner. In the first step all potential predictors were added to individual models. For every model the Akaike information criterion (AIC) was calculated. It evaluates the model depending on the number predictors in use. Of all the models the best is chosen based on AIC and then in the second step the remaining potential predictors are added individually to the model. Again, the predictor which gives the best AIC score is kept for the next round. This selection procedure is continued until no potential predictor leads to an improvement to the model according to AIC. The resulting linear model was then further validated by analysing significance of predictors and the calculation of the root mean squared error (RMSE). The selection of predictors was analysed using a routine by Lindeman et al. (1980). Hereby the importance of the different predictors is compared by estimation of the contributed explained variance. When this is calculated from a linear model the explained variance for each predictor is biased by the order of the predictors. The method by Lindeman et al. (1980) does approach this issue by averaging over different orders of the predictors. This gives an estimation of the explicative contribution by all the predictors.

The general purpose of the seasonal forecasting is to improve drought early warning by supplying additional information to managers and decision makers involved in water resource management. Ideally, operational early warnings systems would issue a warning when a forecast is likely to fall below a specific threshold. The warning would then trigger mitigation actions or other action, improving preparedness. Therefore, the thresholds have to be set carefully in advance by local early warning experts in respect to the envisaged actions. Hence, the information needed for the early warning system consists of a warning threshold and a related probabilistic forecast. In this study we set the drought warning threshold of the standardized runoff index to -0.5. In order to make use of a multiple linear model forecast for early warning, we estimated the average error from the residuals of the calibration data set. Following the assumption inherent to multiple linear models we assumed normally distributed residuals with a standard deviation equalling the residual standard error. This normal distribution was applied to calculate the probability for droughts for the forecasted values. In this way it is possible to transform the deterministic discrete MLM forecast to a probabilistic categorical forecast (Diro, Grimes and Black 2011).

Multiple linear models and linear discriminant analysis have one major shortcoming, which is their limitation to model non-linear processes. However, atmospheric processes are very likely to be non-linear (Mwale et al. 2007). Mwale et al. (2007) successfully applied artificial neural networks for the prediction of summer rainfall in southern Africa, since ANN are capable of representing more complex non-linear relationships between variables. Here, we applied artificial neural networks, which were trained with the genetic algorithm (ANN-GA). As input layer the predictors from the MLM were chosen. This step of the model development

shows if the inclusion of non-linearity can improve the forecast. The artificial neural networks as they were employed here consist of three layers: one input layer, one hidden layer and one output layer. Each layer contains a number of neurons, which has to be predefined before the training procedure. The input layer contains as many neurons as there are variables in the predictor data set. Every neuron handles the data of one of these variables. The number of neurons in the hidden layer determines the complexity of the model. When networks are designed with a high number of neurons, they also have many parameters, which generally increases the risk of overfitting. Lawrence et al. (1997) found that ANN are not very prone to overfitting even when set up with a higher degree than necessary. However, here the available data is small, so chances for overfitting are higher. In order to check the influence of the number of neurons, two different designs were set up: one ANN with three and one ANN with 10 neurons in the hidden layer. The output layer contained only one neuron which is the predictand. Prior to the training procedure all input and output data was scaled to the range zero to one.

The algorithm employed for network training is called “genetic algorithm” and it applies the principles in genetic evolution for training of neural networks. The algorithm used here is implemented in the R package “ANN” (Roy-Desrosiers 2012). The principle is that a population of neural networks moves from one generation to the next, a process during which survival of the fittest, mutation and crossover (gene transfer) apply. Initially, a population of neural networks is created, where all parameters (weights and biases) are set at random. Here, the population size was set to 3000. With the genetic algorithm the network training is a selection process, where each generation experiences survival of the fittest models, mutation of some individuals in the population and gene transfer. The fitness of the networks is measured by the sum of squares of the residuals. Mutation rate was set to 5% and crossover rate was set to 0.8.

From each generation 5% were kept unaltered for the next generation. The least fittest 20% of a generation are removed and replaced by mutated –randomized – individuals in the next generation. For the rest of the individuals a one-point crossover representing a gene transfer is processed by randomly selecting a parameter within two networks which is exchanged between the two parents to create the two children networks for the next generation.

Following the calibration and training, the models were validated. The fit of any given model by the coefficient of determination $R^2$, the mean squared error (MSE) and the root mean squared error (RMSE) were the employed indices. When calculated based on the training data, these measures indicate the quality of fit of a model. However, the ability to represent the training data does not lead directly to good generalisation by the model. In order to test the generalisation ability a 10-fold cross validation scheme was applied. With this approach, both MLM and ANN-GA were analysed, for it was an applicable way to validation. Alternatives, like the leave-one-out validation would be very time consuming with neural

networks training. The split of the dataset for validation is disadvantageous with small datasets and only provides a biased estimate of the generalisation properties.

10-fold cross validation provides an estimate of the model error for independent data. It does so by splitting the calibration data set in ten equally sized parts and train the model ten times by excluding one part in turn, which is then used for validation. The resulting RMSE is then averaged for all runs to obtain $RMSE_{cv}$. It is possible then to compare the RMSE based on the complete calibration data with the $RMSE_{cv}$ from the cross validation. A perfect model would result in no difference between the two errors. Though normally, there is a difference. And the more $RMSE_{cv}$ exceeds RMSE, the weaker is the generalisation performance of a model. Reasons for this can be overfitting or unsuitable predictors.

The models compared here were evaluated in an early warning setting. As already mentioned (see above) this involves the selection of a threshold for a warning. For demonstration purpose in the report the threshold was set to SRI < -0.5, which approximates 31 % of the driest values. By application of the threshold the observation time series was transformed to a binary variable, which indicated drought and non-drought events. The forecast probability is then used to issue a warning, when the forecasted probability for the value is high to fall below the threshold. If warnings are issued at very high probabilities only, then on the one hand most warnings will be correct, on the other hand it is likely that some drought events are missed. In contrast, if warnings are issued at low probabilities already, then more events would be detected early but at the price of a higher false alarm rate. Missed events and false alarms both lower the credibility of an early warning system. Decision makers can set the alarm triggering probability depending on the importance of hits versus false alarms. Depending on the circumstances of the decision maker or the early warning institution false alarms are regarded differently. Here, we analysed the early warning systems performance of models with receiver operating characteristics (ROC), which is a plot that displays the hit rate vs the false alarm rate depending on different levels of forecast probability. Hereby, the plot eases the definition of the early warning trigger. The diagonal resembles a random forecast of no skill since hit rate and false alarm rate are always equal independent of the probability trigger. If a forecasting model has any value then the curve is higher than the diagonal. ROC analysis also offers the so called ROC score to compare different models. The ROC score is the area below the curve, hence it can reach values between 0 and 1. The no skill diagonal already reaches an area of 0.5 so that models have skill only when they have ROC scores higher than 0.5.

**Table 3-1: Selection criteria for the composites: the Limpopo SPI threshold regarded as drought (SPI), the temporal resolution of a composite (time resolution), the relation of composite time to drought occurrence time (time relation), which can be contemporary or with lead time.**

| N◦ | SPI | time resolution | time relation |
|---|---|---|---|
| **C1** | $SPI_1 < -0.8$ | all | contemporary |
| **C2** | $SPI_1 < -0.8$ | monthly | contemporary |
| **C3** | $SPI^{OND,JFM} < -0.8$ | OND, JFM | preceding twelve months |

The analyses in this report used hydrological and sea surface temperature data of the HADSST2 data set (Rayner et al. 2006). Runoff data sources analysed here are from the Global Runoff Data Centre (GRDC, http://www.bafg.de/GRDC), the regional water administration in Mozambique (ARA-SUL) and from the Department of Water Affairs (DWAF) of the Republic of South Africa (http://www.dwaf.gov.za/Hydrology/). Furthermore, metadata such as catchment boundaries and rivers were taken from the free HydroSHEDS data set (http://hydrosheds.cr.usgs.gov/). Country boundaries were used from the Global Administrative Areas database version 2.0 (GADM, http://www.gadm.org/) or as available in the R package “maps” which builds upon the CIA World Data Bank II (http://www.evl.uic.edu/pape/data/WDB/).

The available runoff data covered a high number of stations. Many of these had small catchments. Data preprocessing covered three steps. First, a quality control was performed, second, completeness was checked, third and last, priorities were assigned to the stations. Then, the Standardized Runoff Index (Shukla and Wood 2008) was calculated for the runoff time series. See Table 3-2 and deliverable 4.6 (Seibert and Apel, 2012) for details on the preprocessing procedure.

Four stations of priority one were chosen for analysis in this study. These are the following stations: Botswana (DWAF A5H006), Loskop Noord (DWAF B3H001, GRDC 1196102), Combomume (ARA-Sul A33) and Chókwè (ARA-Sul A35), which are the biggest catchments (see also Figure 3-1).

**Table 3-2: Priority definition criteria for runoff gauges in the Limpopo basin.**

| Criteria | Prioritiy 1 | Priority 2 | Priority 3 |
|---|---|---|---|
| 90 % data coverage after 1979 | Yes (&) | No | No |
| More than 30 years observations | Yes (&) | No | No |
| Catchment area > 4000 km² | Yes (&) | No | No |
| 50 % data coverage after 1979 AND more than 30 years observations AND catchment with > 4000 km² | No | Yes ( \| ) | No |
| 90 % data coverage after 1979 AND missing catchment area | No | Yes ( \| ) | No |
| Catchment area > 1000 km² | No | No | Yes (&) |
| More than 20 years of observations | No | No | Yes (&) |
| Number of stations falling in category (total 454) | 13 | 37 | 65 |

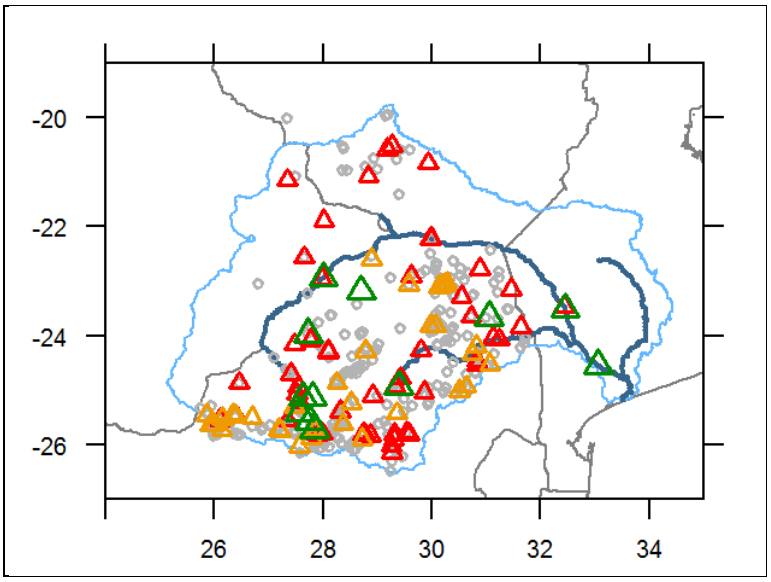



**Figure 3-1: Priorities of runoff gauges (GRDC, DWAF and ARA-Sul) based on length of record and catchment size. Priorities 1,2 and 3 are marked by the green, orange and red triangles, respectively. Grey marked gauges had no priority due to very short records.**

### 3.1.2 Results

The Limpopo basin lays in Southern Africa and has a catchment area of approx.415000 km². The mean annual flow is 7330 Mm³. The river drains from regions in Botswana, Zimbabwe, Mozambique and South Africa, which are characterized by the Köppen-Geiger climate classification as hot arid steppe (Bsh), warm temperate with dry winter and hot summer (Cwa) or warm summer (Cwb) (see Figure 3-2).

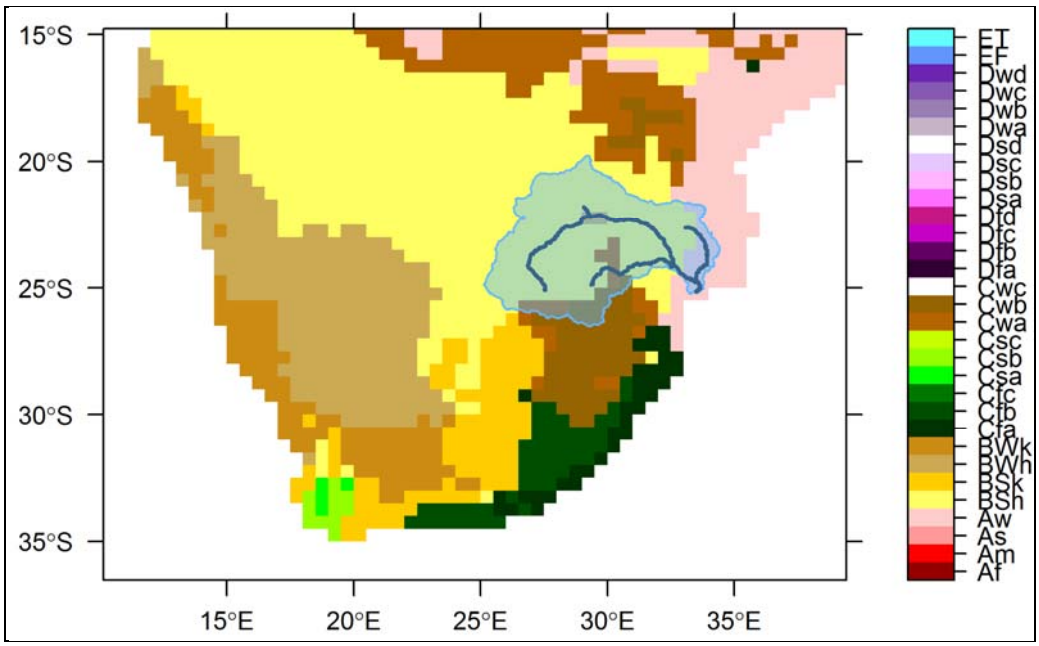



**Figure 3-2: Köppen-Geiger Climate classification and Limpopo basin with main tributaries (Kottek et al.,2006).**

## SETUP OF STATISTICAL MODELS

### *Potential predictors*

Statistical methods rely on the selection of a reasonable set of predictors. The main intention in this study was to identify teleconnected ocean regions which show potential to predict drought in the Limpopo basin. It is known that droughts in Southern Africa are connected to the El Nino phenomenon, but Manatsa et al. (2007) have shown that the sea level pressure in Darwin (Australia) has even better predictive potential. Hence, a number of indices related to the El Nino/southern oscillation (ENSO) were selected (see Table 3-3). The main difference is the length of the underlying SST data set caused by the measuring technique. Additionally, the Northern Atlantic Oscillation was added to the list as well as the Indian Ocean dipole mode index (DMI), which describes a SST anomaly of the equatorial Indian Ocean. During negative phases the western Indian Ocean has lower than normal temperatures can affect monsoon dynamics (see Table 3-3). These established ocean anomaly indices were then complemented by predictors based on Ocean regions in closer proximity to southern Africa in the Indian Ocean and the Atlantic.

Teleconnected Ocean regions were identified by composites analysis. The composites consecutive to drought in the Limpopo basin are presented in Figure 3-3. During the rainy season the Ocean regions in the Southern Indian Ocean showed lower than normal temperatures during drought events. Most of the composites showed only small regions with significant anomalies. Three Ocean regions were chosen from the November and December composites (Table 3-4 and Figure 3-5).

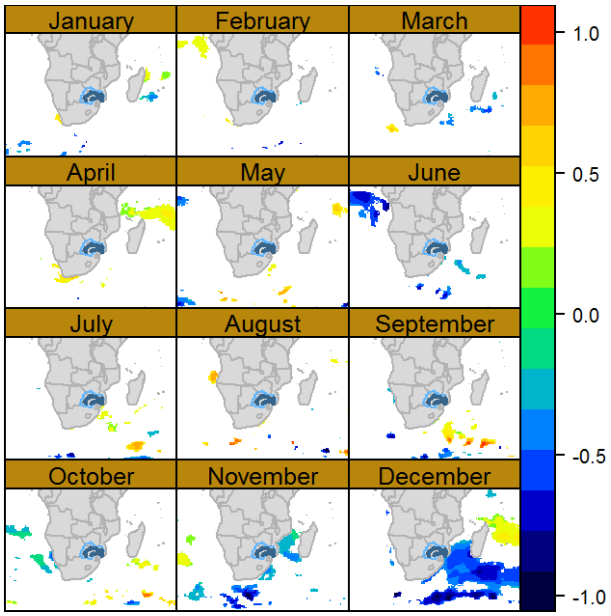



**Figure 3-3 Composite 2 of consecutive sea surface temperature anomalies during droughts in the Limpopo**

For the purpose of identifying regions specifically suitable for drought prediction, most regions were selected based on composites that show significant SST anomalies preceding drought. This was performed separately for the early (October to December, OND) and late (January to March, JFM) rainy season (C3 in Table 3-1). At longer lead times of more than three months (during March to June) very few Ocean regions showed significant anomalies (see Figure 3-4). However, the Southern Atlantic was cooler than normal from July to September preceding early rainy season droughts (OND). Ocean regions close to the coast of Angola and Namibia were significantly colder as well for both early and late rainy season, although these anomalies did

not occur in the same preceding months. In the Indian Ocean positive anomalies were present north of Madagascar before droughts occurred early in the rainy season. The region southeast of Madagascar was colder than normal during a drought in OND while the regions Southwest of Madagascar were warmer preceding an early rainy season drought. The Ocean region at the Eastern coast of Southern Africa is related to the Agulhas current. The region of the Agulhas current is one of the important sources for atmospheric moisture in the region for Southern Africa (Gimeno et al., 2010). Hence, it is very likely that the sea surface temperature of the region influences droughts in the Limpopo.

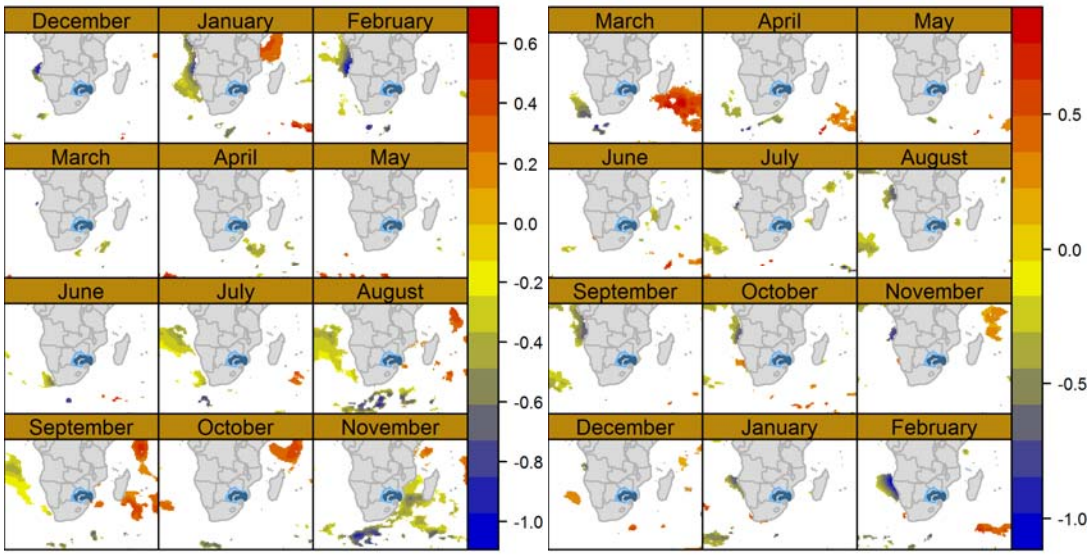



**Figure 3-4 Composite 3 with sea surface temperature anomalies of the 12 months preceding a drought in October to December (left figure) and January to March (right figure).**

**Table 3-3: Climate indexes considered for prediction.**

| Variable / Data set | Covered period | Source |
|---|---|---|
| Southern Oscillation Index (SOI) | 01.1951 - now | Climate Prediction Center of NOAA: ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/soi |
| ENSO indexes (ERSST) | 01.1950 - now | Climate Prediction Center of NOAA: http://www.cpc.ncep.noaa.gov/data/indices/ersst3b.nino.mth.ascii |
| ENSO indexes (OISST) | 01.1982 - now | Climate Prediction Center of NOAA: http://www.cpc.ncep.noaa.gov/data/indices/sstoi.indices |
| Darwin sea level pressure (SLP) | 01.1951 - now | Climate Prediction Center of NOAA: http://www.cpc.ncep.noaa.gov/data/indices/darwin |
| Tahiti SLP | 01.1951 - now | Climate Prediction Center of NOAA: http://www.cpc.ncep.noaa.gov/data/indices/tahiti |
| North Atlantic Oscillation (NAO) | 01.1950 - now | Climate Prediction Center of NOAA: ftp://ftp.cpc.ncep.noaa.gov/wd52dg/data/indices/nao_index.tim |
| Indian Ocean Dipole Mode Index (DMI) | 11.1981 - now | Based on NOAA OISST Ver.2 http://www.jamstec.go.jp/frcgc/research/d1/iod/DATA/dmi.monthly.ascii |
| Oceanic Nino Index (ONI) | 02.1950 - now | Based on ERSST.v3b of Climate Prediction Center of NOAA: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml |
| Trans Nino Index (TNI) | 03.1870 - now | HadSST1.1 until Nov 1981 and NCEP NOAA OI after: http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/tni.long.data |
| NINO3.4 (HadSST) | 1871 -2011 | http://www.esrl.noaa.gov/psd/gcos_wgsp/Timeseries/Data/nino34.long.data |

Further potential predictor regions were selected from correlation analysis (see deliverable DEWFORA 4.6, Seibert and Apel, 2012). The complete extended list of potential predictors is compiled in Table 3-4. The exact outlines of the selected regions is shown in Figure 3-5.

**Table 3-4: Potential predictors for precipitation variability in the Limpopo basin, detected by composites and correlation analysis. Parameters are sea surface temperature (SST) and sea level pressure (SLP), the types of the reference composites (see Table 3-1) are indicated in the last column with the months and respective lead time.**

| Ocean | region | Latitude (min, max) | Longitude (min, max) | Parameter | Composite (lead time) | Abbreviation |
|---|---|---|---|---|---|---|
| Atlantic | Namibian coast | -26 -11 | 5, 15 | SST | $C3_{OND}(11)$, $C3_{JFM}(12,2,1)$ | NamibCoast |
| | Angolan coast | -5 -17 | 8, 13 | SST | $C3_{OND}(11,10)$, $C3_{JFM}(6,5)$ | AngolCoast |
| | South | -4, 4 | -27 -36 | SST, SLP | $C3_{OND}(8,7,6,4)$ | SAtl |
| | South African coast | 12, 14 | -28 -34 | SST | $C3_{OND}(5,4,3)$ | SACoast |
| | SW | -4, 8 | -10 -20 | SST | $C3_{OND}(5,4,3)$ | SWAtl |
| | Central South | -41 -30 | -15 -1 | SST | (CCF) | CSAtl |
| | E equatorial | 5, 16 | -35 -17 | SST | (CCF) | eEqAtl |
| | W equatorial | 12, 22 | -68 -45 | SST | (CCF) | wEqAtl |
| Indian Ocean | E equatorial | -10, 0 | 51, 59 | SST | $C3_{OND}(2,3)$, $C3_{JFM}(4)$ | eqIO |
| | W equatorial | -10, 0 | 41, 49 | SST, SLP | $C3^{SST}{}_{OND}(11,1)$, $C3^{SLP}{}_{OND}(12,3,2)$ | weqIO |
| | Southeast | -42 -32 | 50, 63 | SST | $C3^{SST}{}_{JFM}(1)$, $C3^{SST}{}_{JFM}(4)$ | seIO |
| | Western | -25-12 | 93, 115 | SST | (CCF) | weIO |
| | Mozambique shore | -28 -15 | 32, 45 | SST | $C2_N$ | Moz |
| | S Agulhas current | -43 -36 | 17, 30 | SST | $C2_N$, $C2_D$ | sAGU |
| | N Agulhas current | -34 -28 | 28, 34 | SST | $C2_N$, $C2_D$ | nAGU |

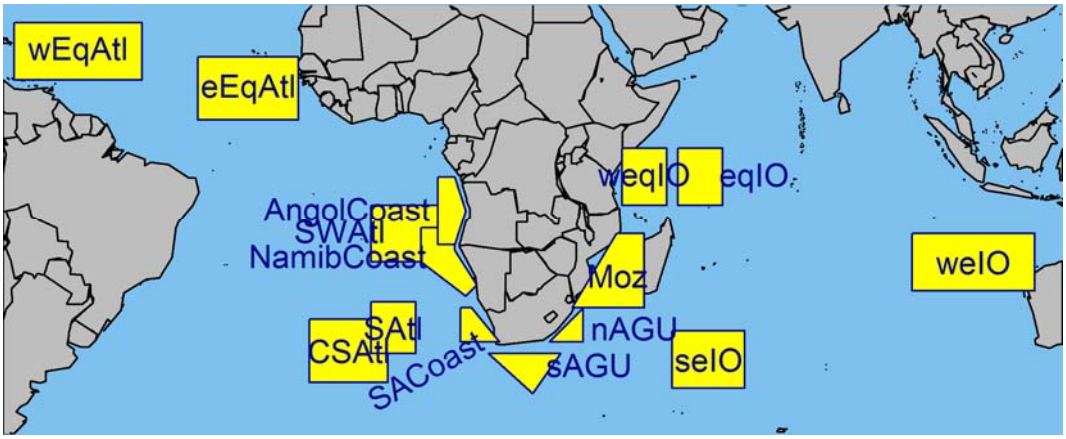



**Figure 3-5: Teleconnected regions established by correlation and composites analysis.**

**Table 3-5: Coefficients of determination ($R^2$) for multiple linear models and artificial neural networks. Linear models were established by forward selection with Akaike's information criterion (AIC) and Bayes' information criterion (BIC) for June predictors (J) and predictors aggregated over April-June (AMJ). Neural networks were established using the same predictors as the AIC selected AMJ model using two degrees of complexity with 10 neurons (ANN-10) or 3 neurons (ANN-3) in the hidden layer.**

| | AIC | | BIC | | ANN-10 | ANN-3 |
|---|---|---|---|---|---|---|
| Station | J | AMJ | J | AMJ | AMJ | AMJ |
| Botswana | 0.32 | 0.19 | 0.23 | 0.18 | 0.62 | 0.44 |
| Loskop Noord | 0.38 | 0.52 | 0.28 | 0.26 | 0.83 | 0.69 |
| Combomume | 0.31 | 0.59 | 0.18 | 0.33 | 0.91 | 0.82 |
| Chókwè | 0.21 | 0.38 | 0.21 | 0.28 | 0.65 | 0.57 |

### *Selected predictors in multiple linear models*

Here, the results of the parameter selection and model setup is presented for four stations in the Limpopo. The selection of predictors from the list of potential predictors was performed with stepwise selection for multiple linear models. The selection is a process which is sensitive to the selection criteria. When selection was performed with the Akaike's information criterion (AIC) the resulting models had a higher coefficient of determination than with the Bayes information criterion (BIC). The BIC is more conservative than the AIC since the penalty for the number of parameters is higher. As a result, less parameters are selected and the coefficient of determination is lower. In cases where this effect is strong it is likely, that some of the parameters only contributed little additional model skill.

The given set of potential predictors was tested in two ways which differed in the sea surface temperature (SST) aggregation. One dataset offered SST data which was aggregated at a monthly scale, the second was aggregated over three months. For the prediction of the rainy season runoff at a three-month lead time this resulted in one dataset of June (J) and another data set of values aggregated over April – June (AMJ). The selection was performed independently for both data sets.

The resulting multiple linear models had coefficients of determination which ranged from 0.18 to 0.59 (Table 3-5). AIC selected models were up to 0.26 higher than BIC selected models. The highest explained variance was reached with the AIC selected AMJ model for the Combomume station. With the exception of the Station "Botswana" coefficients of determination of the models with AMJ predictors were always higher than June only models. Hence, we gave a general preference to AMJ models, which will be analysed in more detail.

Based on the AIC selected linear models with AMJ predictors artificial neural networks were defined using the same selection of predictors. Hereby it was possible to check if non-linear relationships limit the model in regard to these predictors. The artificial neural networks yield a much improved coefficient of determination, which reached a maximum of 0.91 in the ANN with 10 neurons in the hidden layer. The more complex ANN with 10 neurons in the hidden

layer was always able to explain a higher portion of the total variance. These result indicate that the variability within the dataset could be represented much better with the artificial neural network. However, it is important to note that from these findings it can be concluded that the artificial neural networks are also more powerful as prediction models. This question is analysed and discussed in the next section on model validation.

**Table 3-6: Multiple linear models: Significances of selected predictors for the AIC selected model with AMJ predictors.**

| Parameter | Botswana | Loskop Noord | Combomume | Chókwè |
|---|---|---|---|---|
| SRI3_NOW | | 0,000 | 0,014 | |
| AngolCoast_f3 | | 0,096 | 0,012 | |
| SWAtl_f3 | 0,428 | | | |
| weqIO_f3 | | 0,170 | | |
| seIO_f3 | | | 0,002 | |
| eEqAtl_f3 | | | 0,000 | |
| wEqAtl_f3 | 0,062 | 0,017 | | |
| Moz_f3 | | 0,001 | | 0,033 |
| sAGU_f3 | | | | 0,414 |
| nAGU_f3 | | 0,007 | | 0,012 |
| SOI.STAND | | | 0,040 | |
| ERSST.ENSO3 | | 0,166 | | |
| DARWIN.STAND | | | 0,003 | |
| NAO | | | 0,079 | 0,004 |
| DMI | | | | 0,130 |
| ONI | 0,005 | 0,063 | 0,006 | 0,011 |

The linear models which were selected based on AIC and had AMJ predictors had highest coefficients of determination in most of the cases and will be compared in more detail here. Every model did have a unique selection of parameters but some parameters were selected more frequently (see Figure 3-6, Figure 3-7, Table 3-6). The most prominent one is the Oceanic Nino Index which is an index for El Nino phenomenon. The predictors were significant at all four stations (Table 3-6). The estimated values were in a similar range of -0.5 to -0.8. This results in droughts (negative SRI) during El Nino events (positive ONI anomalies). The explicative contribution by ONI was highest at the station “Botswana” with 0.13 and lowest at the station “Loskop Noord” with 0.05 (Figure 3-7). The remaining predictors were only selected in one or two of the models presented here. Among those, some predictors achieved high importance since the realised a high explicative contribution. The standardised runoff for April to June (SRI3_NOW) was able to contribute approximately 0.18 in the model for the station “Loskop Noord”. A similarly high contribution was reached by the eastern equatorial Atlantic region (eEqAtl_f3) in the model for “Combomume”. Very few parameters were selected from the Southern Atlantic and the western coast of Southern Africa. Two parameters were selected from this ocean region and had explicative

contribution of at max 0.05. Parameters related to the Agulhas current at the eastern coast of Southern Africa were all selected at least once and contributed up to 0.09 to a model's explained variance.

In some of the models the selected parameters were not significant at a level of 0.05, which means that the estimated parameters were not significantly different from zero (Table 3-6). However, the selection of these predictors is caused by the nature of the AIC selection which is less strict than BIC. In the BIC selection less non-significant predictors were selected (not shown). The non-significant predictors only contributed 0.01 to 0.05 to the explained variation (Figure 3-7). The Indian Ocean Dipole mode index (DMI) contributed approximately 0.05 to the model of station "Chókwè" even though it was not significant.

In conclusion, the parameter selection that resulted for the four presented stations was very diverse. With exception of an El-Nino index no parameter was present in all models. It is possible, that this is caused by real differences in teleconnections and atmospheric relationships. The stations differ in size and location of the catchment. At same time, "Combomume" contains the station "Botswana" and "Chokwe" integrates all three stations. Hence, it can be expected that the atmospheric influences on these stations are similar. Since this is not the case there are different potential reasons. The parameters from the same region are likely to show colinearity and hence it can happen easily that one parameter is preferred to another. Considering the low contribution of a several stations it is also likely that the models suffer from overfitting. In the following section on model validation this will be addressed and the generalisation properties will be compared.

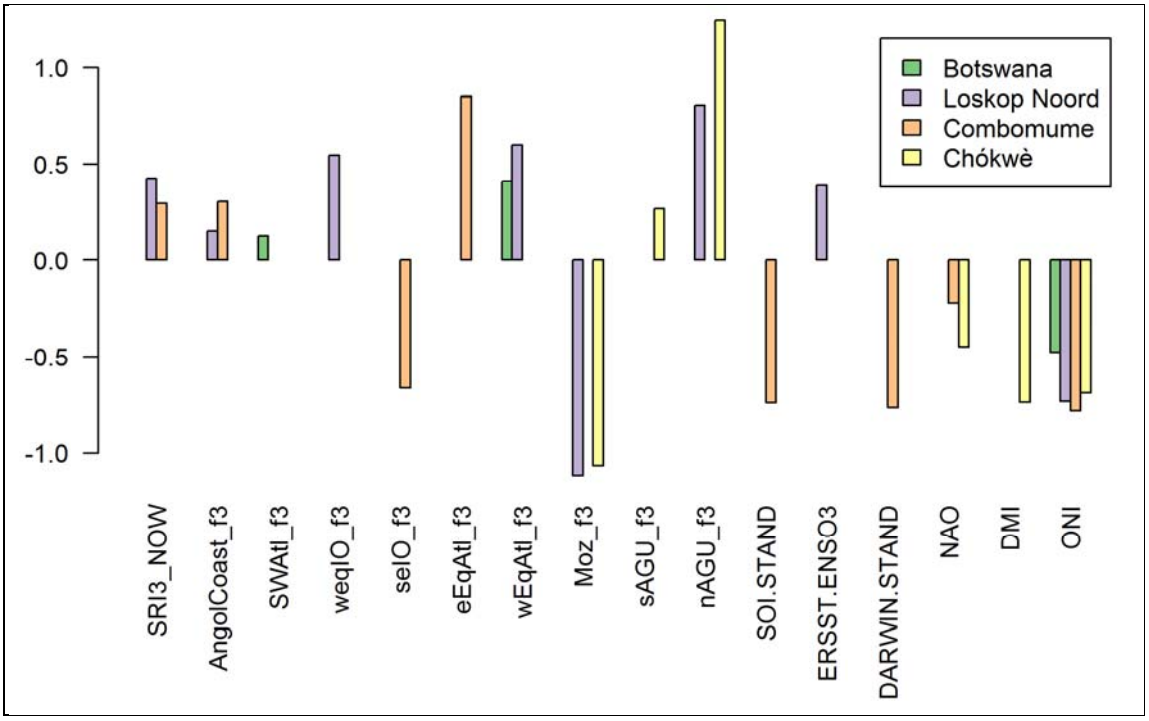



**Figure 3-6: Multiple linear models: Estimated predictors for the AIC selected models at the four runoff stations.**

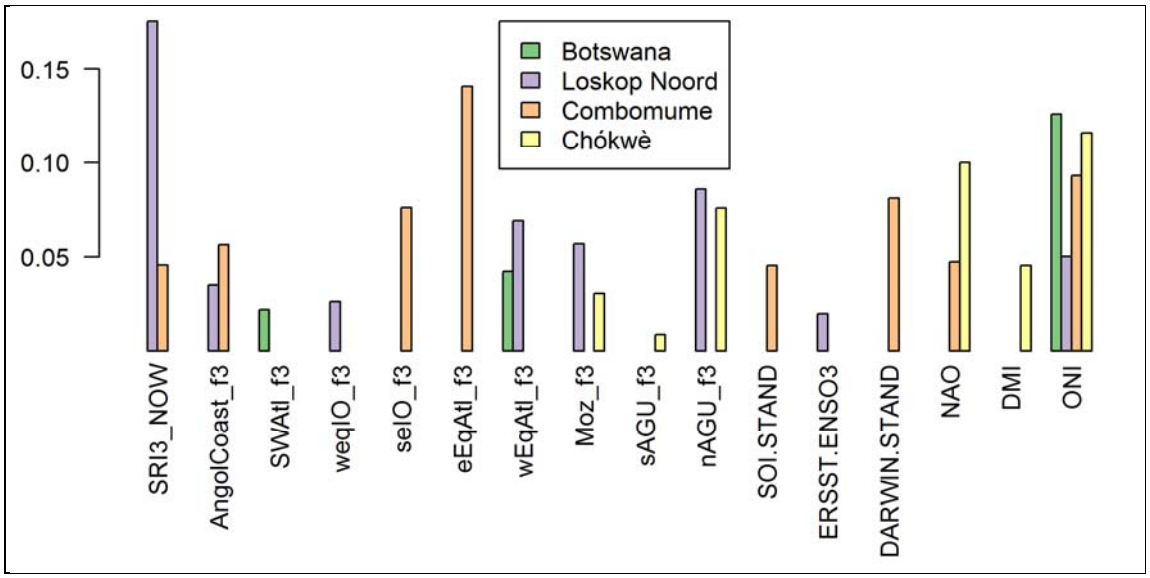



**Figure 3-7: Multiple linear models: Estimated explicative contribution of predictors for the AIC selected models at the runoff stations.**

## MODEL VALIDATION

In the previous section the predictor selection results and model fit performances were presented. In Figure 3-11 to Figure 3-9 results of the fitted models are presented. For all the stations the artificial neural network with 10 neurons in the hidden layer reached the best fit. Particularly, extreme events were captured much better by the ANN than with the linear model (AIC selection, AMJ predictors). The forecasts of most stations show some gaps of the forecasted values. This is caused by gaps in the input data sets.

At the Station “Botswana” the fit to the model is largely increased in the artificial neural network in comparison to the linear network (Figure 3-8). However, several extreme events are missed completely by the model. The only extreme droughts that are simulated well by the ANNs are the years 92 and 2003. With the given set of predictors the remaining droughts could not be explained.

The simulation of station “Loskop Noord” also showed the closest fit with the ANN10, which matched extreme values in the best way of the three models presented here (Figure 3-9). However, the MLM performed better for some droughts than the ANN3.

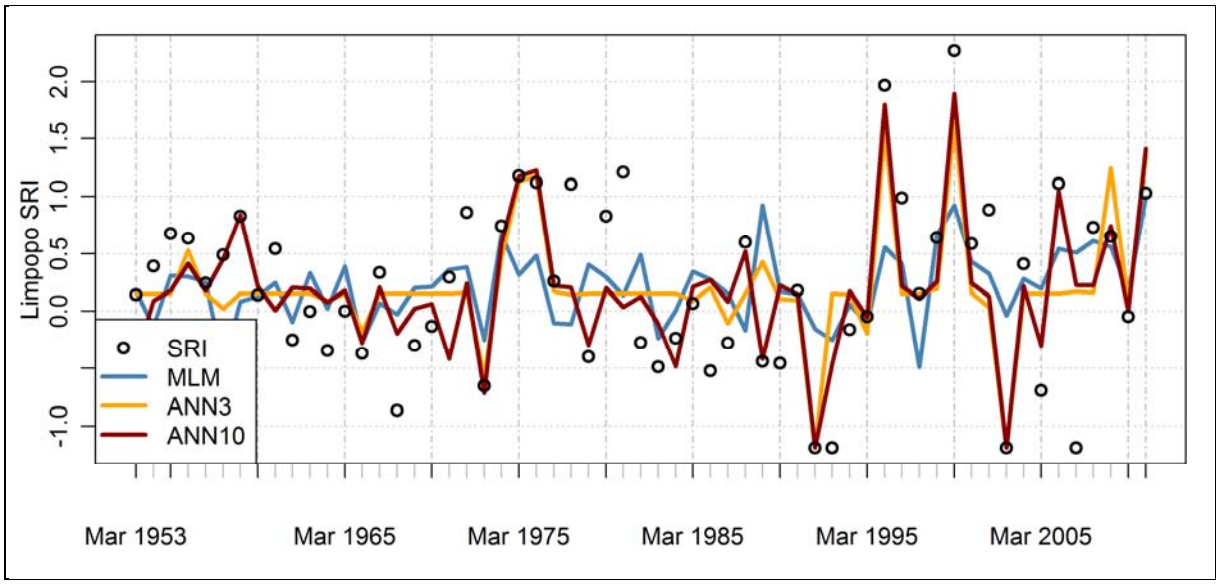



**Figure 3-8: Botswana: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10).**

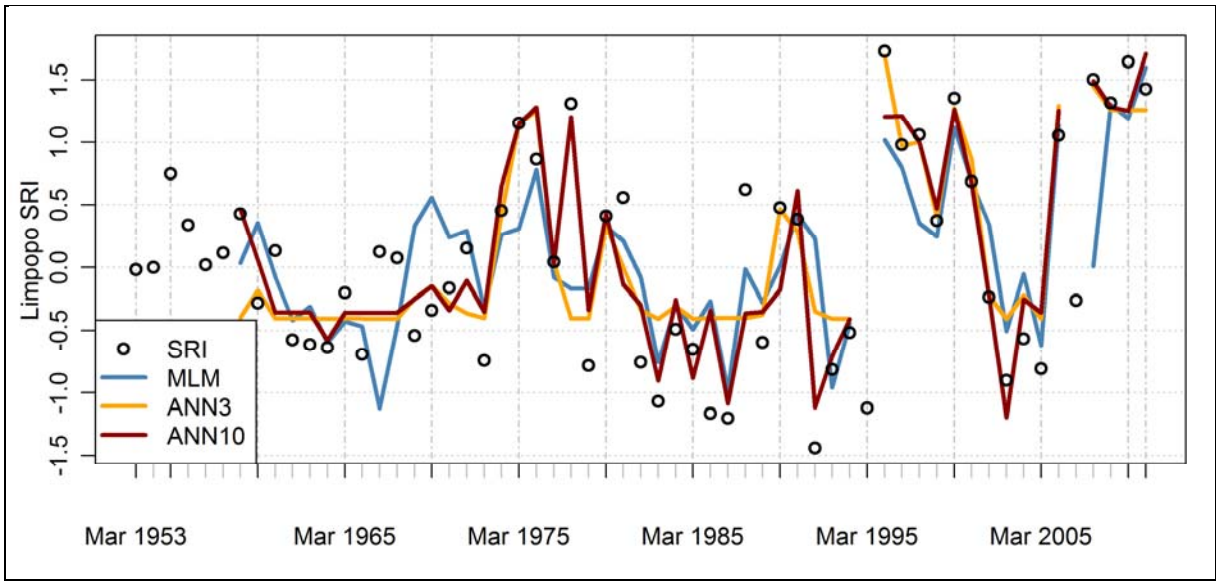



**Figure 3-9: Loskop Noord (Olifants): Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10).**

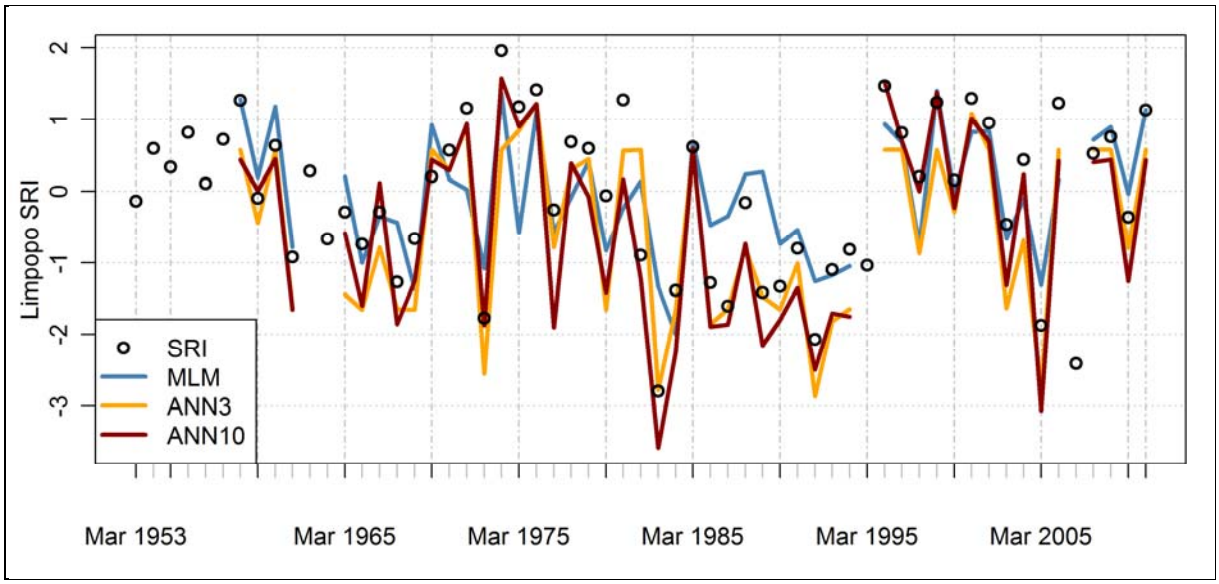



**Figure 3-10: Combomume: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10).**

The station “Combomume” was simulated best from all the presented stations (Figure 3-10). Here, all droughts were captured well by the ANN10 and even the simpler ANN3. The MLM exhibited less power for extreme values. With the selected predictors the models are able to explain most of droughts. The drought period from 1986 to 1995 is represented better by the artificial neural networks.

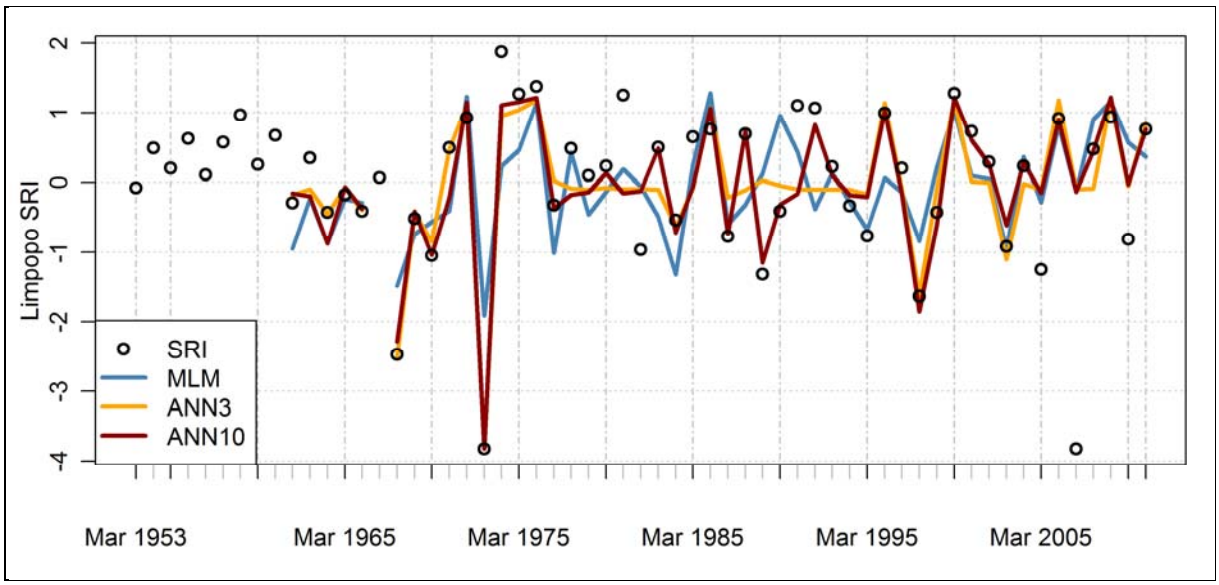



**Figure 3-11: Chókwè: Fitted statistical seasonal forecasts models of standardised rainy season runoff ($SRI_{ONDJFM}$) in June by a multiple linear model and artificial neural networks (ANN) with two degrees of complexity, one with 3 neurons (ANN3) and the other one with 10 neurons (ANN10).**

At the station “Chókwè” runoff observation showed anomalies with unusually low runoff during some years. Whether these are errors, we were not able to resolve. The extreme

drought in 1973 is represented by all three models. In contrast, the extreme drought in 2007 is not captured by the models.

The simulations at all stations were always improved with the artificial neural networks. The capability for non-linear relationships was capable of representing the extreme values much better than the linear model. However, these results do not give any reliable indication of the generalisation performance of the models. This was analysed by applying leave-one-out cross validation. Using cross validation the root mean squared error (RMSE) for independent data can be estimated. Models for which the RMSE of the crossvalidation is much higher than the RMSE of the calibration have only low skill and have low generalisation performance. As expected from the discussed results so far, the RMSE of the MLM is higher than for both ANNs. However, the crossvalidation showed that for all stations the crossvalidated RMSE increases by far more than for the MLM. The resulting RMSE for the MLM was lowest at the station “Loskop Noord” with 0.69 and as high as 1.06 at the station “Chókwè”. This leads to the conclusion that the MLM method is more robust than the ANN method for this application, but the skill is generally very low.

**Table 3-7: Root mean squared errors from calibration data and from leave-one-out cross validation.**

| Station | run | MLM | ANN-10 | ANN-3 |
|---|---|---|---|---|
| Botswana | ALL | 0,68 | 0,45 | 0,55 |
| | CV | 0,72 | 0,76 | 0,87 |
| Loskop Noord | ALL | 0,64 | 0,34 | 0,46 |
| | CV | 0,69 | 0,90 | 0,88 |
| Combomume | ALL | 0,78 | 0,33 | 0,47 |
| | CV | 0,88 | 0,85 | 1,15 |
| Chókwè | ALL | 1,00 | 0,69 | 0,77 |
| | CV | 1,06 | 1,05 | 1,32 |

## MODEL PERFORMANCE FOR DROUGHT EARLY WARNING

Two types statistical methods were developed, of which the multiple linear model exhibited the best generalisation performance. For drought early warning it was necessary to calculate probabilities for drought and flood based on the forecasting models. The thresholds for drought and flood were set to -0.5 and 0.5. In Figure 3-12 to Figure 3-14 the results are presented with a simulated warning for the class with highest probability.

At the station “Botswana” the model had no skill for drought early warning (Figure 3-12). No drought that occurred was warned ahead. Instead, drought warnings were issued when no droughts occurred. At this station the Early warning system has no value.

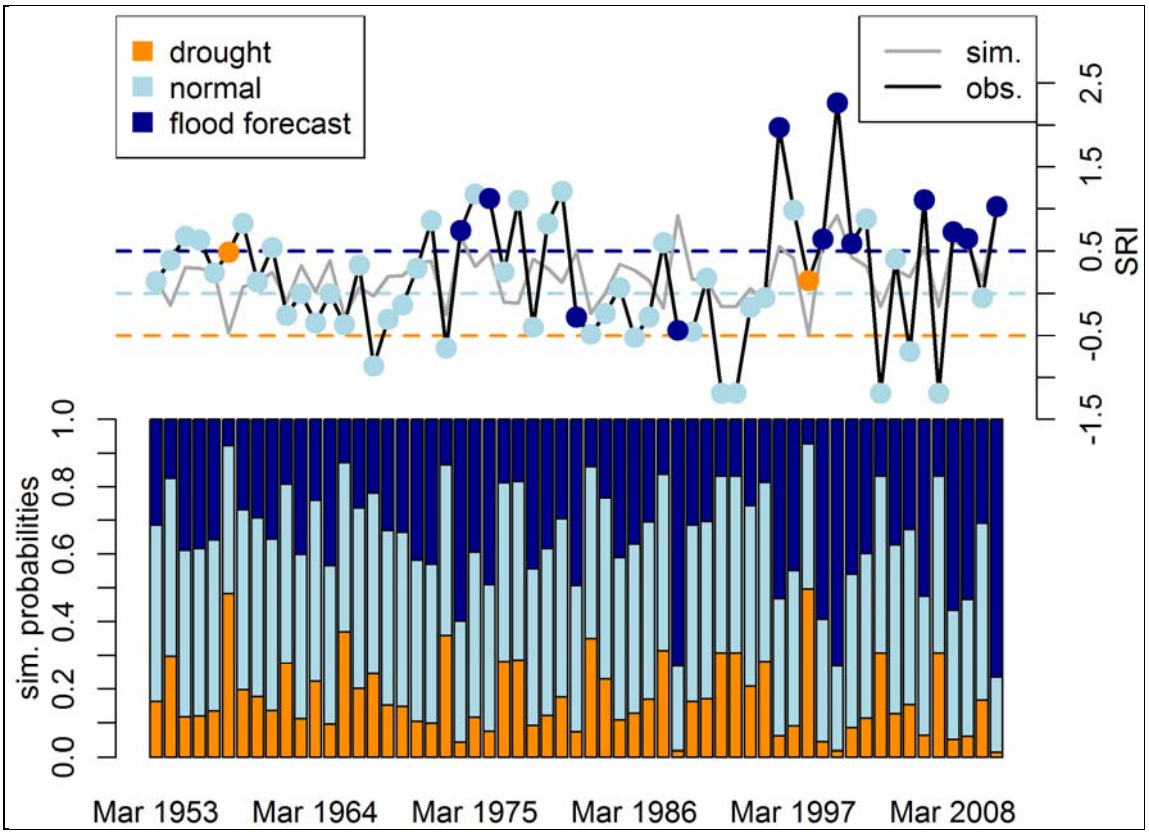



**Figure 3-12 Botswana: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange).**

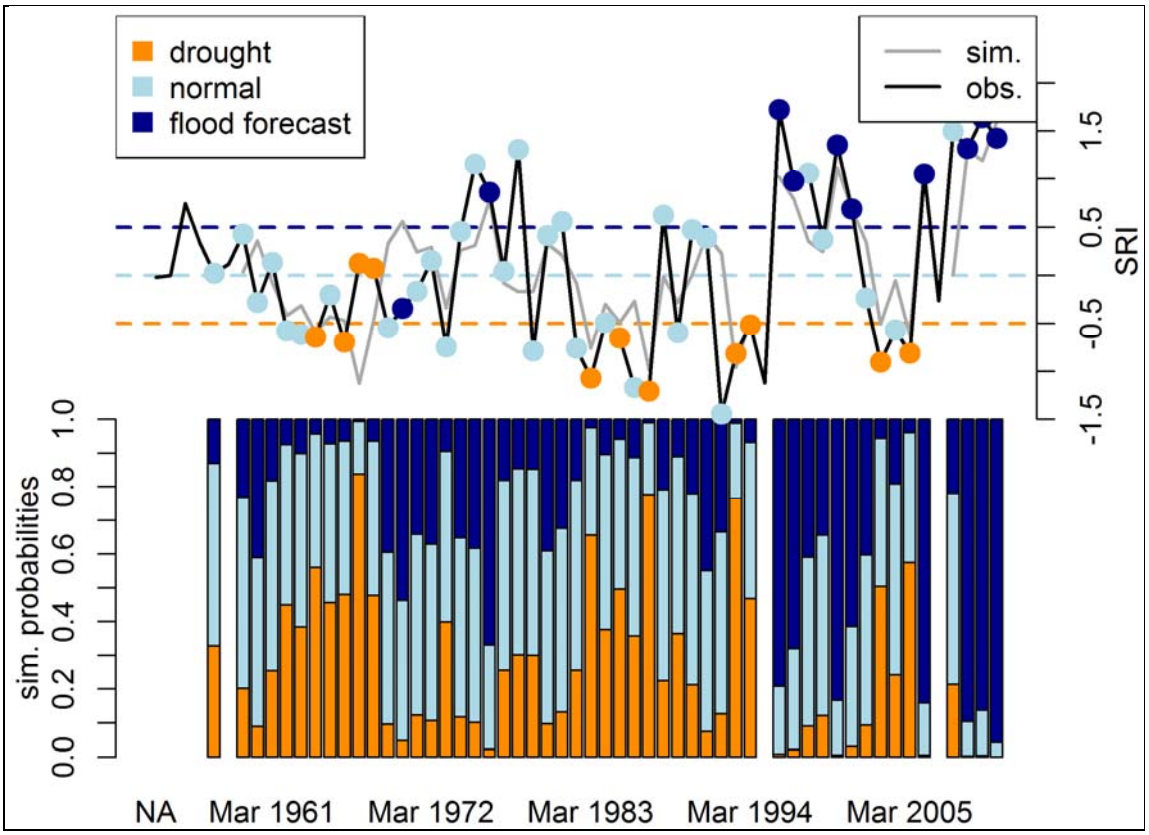



**Figure 3-13 Loskop Noord: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange).**

An early warning system for station “Loskop Noord” based on the multiple linear model showed little skill. The ROC score was 0.77 and many droughts would have been warned ahead. However, the majority of droughts (52%) was missed and 18 % of the drought warnings failed. For flood conditions the model performance was even better with only one failed warning and 4 of the major drought missed.

The early warnings for Combomume were the most skillfull ones from all the stations (Figure 3-14). Almost all droughts were warned ahead and only three were missed. The ROC score was 0.85 and an analysis of a ROC plot showed that the early warning system can be improved by setting an early warning probability threshold to 0.5, which reduces the false alarm rate (Figure 3-15, left). Warnings for flood were similarly successful with many correct warnings. However, one drought warning was issued and a flood was happening. These kind of false warning are critical for early warning institution, since they can lower the credibility of the institution.

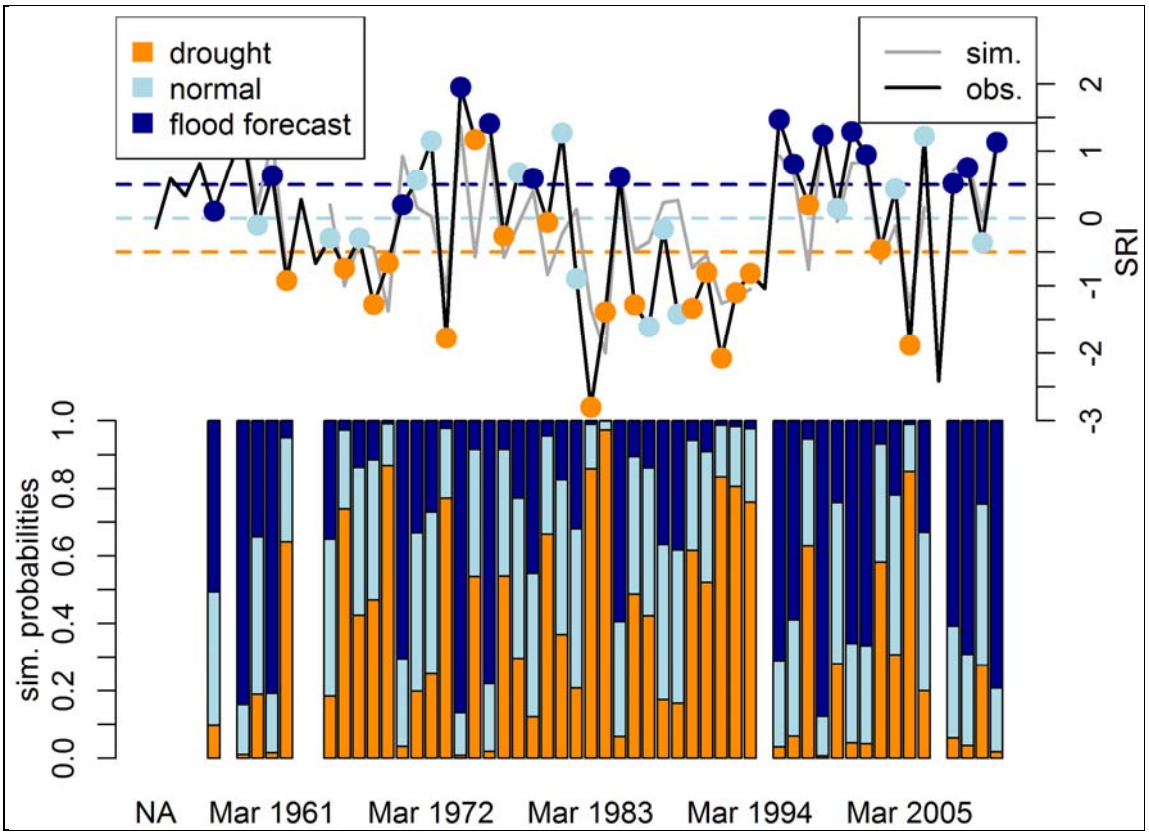



**Figure 3-14 Combomume: Early warning with the multiple linear model. Above:simulation (grey line) and observed values (black line) with early warnings three months ahead for class with highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange).**

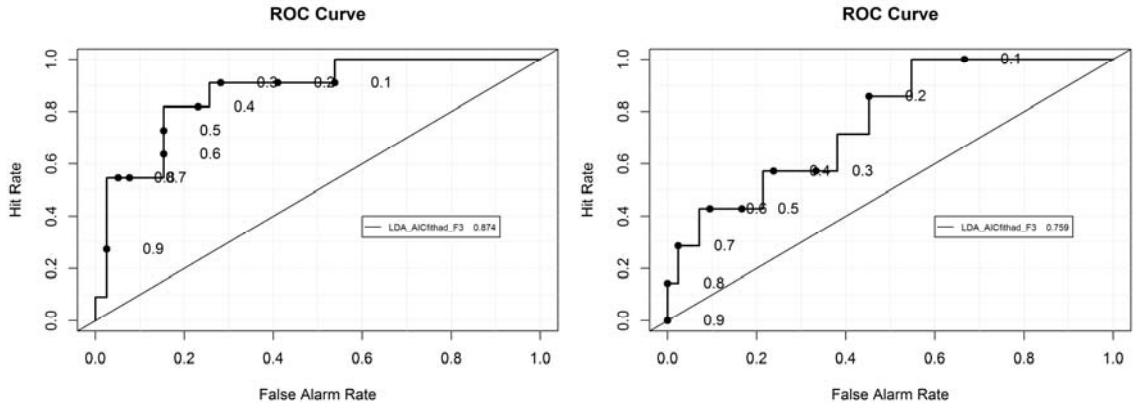



**Figure 3-15 Relative operating characteristics plot of the MLM model for stations Combomume (left) and Chókwè (right).**

At the station Chókwè the early warning system had less skill than at Combomume (Figure 3-15). Many drought warnings were issued when normal conditions and even floods occurred. The ROC score was a low value of 0.77 and an adjustment of the early warning probability threshold would not lead to much improvement (Figure 3-15, right).

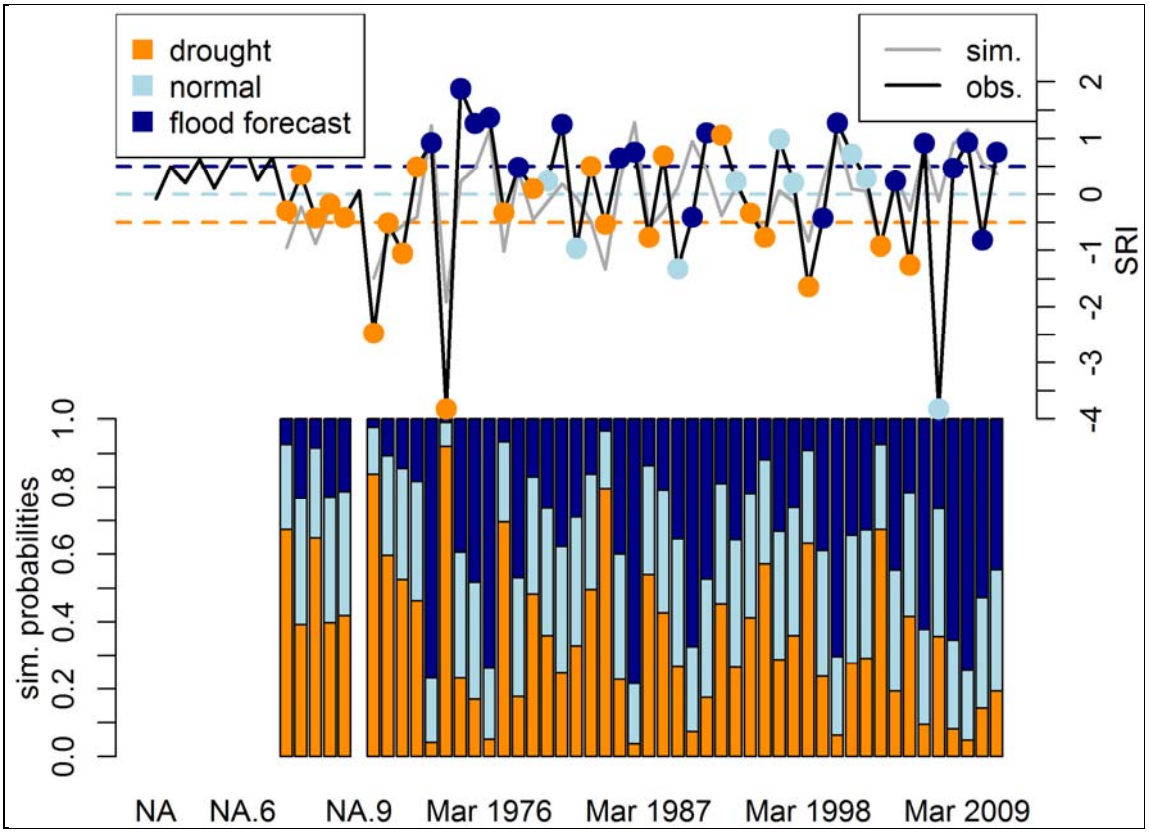



**Figure 3-16 Chókwè: Early warning with the multiple linear model. Above: simulation (grey line) and observed values (black line) with early warnings three months ahead for the class of highest probability (coloured points). Below: Probabilities for flood (dark blue), normal (bright blue) and drought (orange).**

### 3.1.3 Conclusions

Within this study a thorough analysis of potential predictors for hydrological drought was conducted. From an extensive list of teleconnected ocean regions, skillfull predictors were selected for multiple linear models. Hereby, it was taken into account how the automatic selection procedure affects the choice of parameters. The linear models were outperformed by artificial neural networks when it comes to fitting the calibration data set. However, it was shown that the linear models provide more robust results for independent data. The artificial networks probably suffers from overfitting for a small dataset as it was available for the analysed stations. The linear model underestimated the variability in the runoff time series for most stations.

The early warning properties were not sufficient in most stations. Only the station "Combomume" was predictable to a degree which delivered an early warning with value. There are many potential reasons for the weak performance. First, the predictability of rainfall variability in southern Africa is low (Mwale et al., 2007) due to non-linear processes. Second, runoff measurements are sometimes unreliable and some stations' time series had extreme values that may be affected. Third, runoff is affected by management. The runoff time series contains and is affected by much more than just the meteorological signal. Catchment properties, memory of the system and management strongly affect the runoff response. For example, a memory effect of past drought might introduced by dam operation. After extreme drought dam storage have to be replenished so that runoff may be reduced and appear as a drought even when the meteorological condition is normal. These long term effects of drought can have an effect on the runoff. These factors are not taken into account by the statistical approach presented here.

For a station such as Combomume, which was simulated with the best results, the coupling to the meteorological anomaly might be stronger. The catchment is large and individual effects of reservoirs were less pronounced.

### 3.2 DYNAMIC MULTI-MODEL ENSEMBLE FORECAST

This section presents a dynamic forecasting system based on a distributed hydrological model developed within DEWFORA for the seasonal prediction of hydrological droughts for the semi-arid Limpopo river basin in Southern Africa. The PCR-GLOBWB hydrological model was coupled with the output of the global atmospheric model ECMWF-System 4. The system is tested in a hindcast mode for the period 1981 to 2010 to allow for testing prior to its use in pre-operational real-time forecasts. This system seeks to provide operational guidance to farmers and water mangers within the basin at the seasonal time scale by predicting the availability of water for irrigation and water supply, which are the most important water uses in the basin.

In this section we will not describe the Limpopo river basin as it was already thoroughly described in previous deliverables. Deliverable D6.1 (DEWFORA 2012b) presents a description of the basin, as well as a summary of the existing drought management practices in the basin. Deliverable D2.1 (DEWFORA 2011a) and Deliverable 2.2 (DEWFORA 2011b) present a detailed description of the current drought monitoring and forecasting systems, networks, and institutions in the Limpopo region, as well as the institutional frameworks for drought mitigation and adaptation, and a review of the historical droughts in the region. The tailor-made hydrological model for the Limpopo river basin was presented in Deliverable D4.7 (DEWFORA 2012a).

The approach followed in this forecasting system is summarized in Figure 3-17. It starts with obtaining the meteorological seasonal forecast from ECMWF together with some necessary "pre-processing" of the data. The hydrological model (embedded in the Delft-FEWS forecasting shell) then simulates the forecasted seasonal hydrology obtaining predicted streamflows and other hydrological variables.

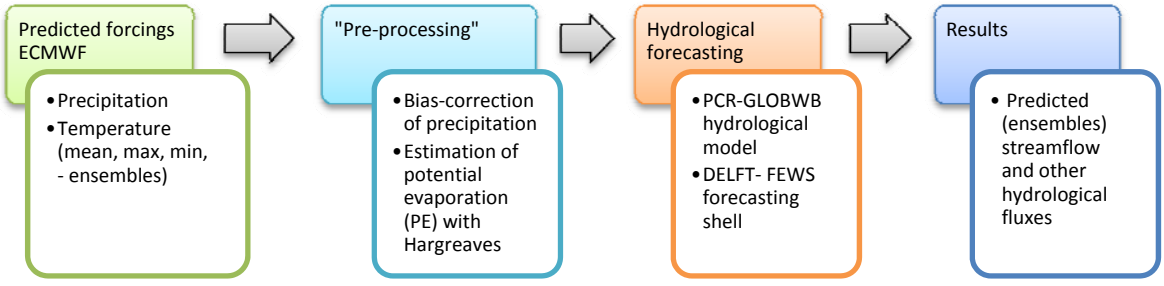



**Figure 3-17 Approach followed in this forecasting system for the Limpopo river basin**

#### 3.2.1 Model and simulations set-up

## DESCRIPTION OF THE REGIONAL HYDROLOGICAL MODEL

For the Limpopo river basin case study, a downscaled version of the global PCR-GLOBWB hydrological model is used. This is a continuous-time simulation, process based distributed model applied on a cell-by-cell basis. PCR-GLOBWB is in many ways similar to other global

hydrological models, but it has several improved features, such as improved schemes for sub-grid parameterization of surface runoff, interflow and baseflow, a kinematic wave based routing for the surface water flow, dynamic inundation of floodplains and a reservoir scheme (van Beek and Bierkens 2009, van Beek 2008). The model was set up for the Limpopo basin with a spatial resolution of 0.05 x 0.05˚ and the simulation is carried out on a daily time step. For a thorough description of the model set up for the Limpopo river basin the reader is referred to Deliverable 4.7 (DEWFORA 2012a).

## DESCRIPTION OF THE FORECASTING SHELL

The hydrological model was embedded in the Delft-FEWS (Flood Early Warning System) open shell for forecasting purposes. The system is a sophisticated collection of modules designed for building a hydrological forecasting system customised to the specific requirements of an individual organisation. The philosophy of the system is to provide an open shell for managing the data handling and forecasting process. This shell incorporates a comprehensive library of general data handling utilities, allowing a wide range of external models to be integrated in the system through a published open interface. This allows existing simulation models and data streams to be turned into a comprehensive and reliable forecasting system (Werner et al. 2013). Figure 3-18 shows a window of the customized Delft- FEWS shell for the Limpopo river basin.

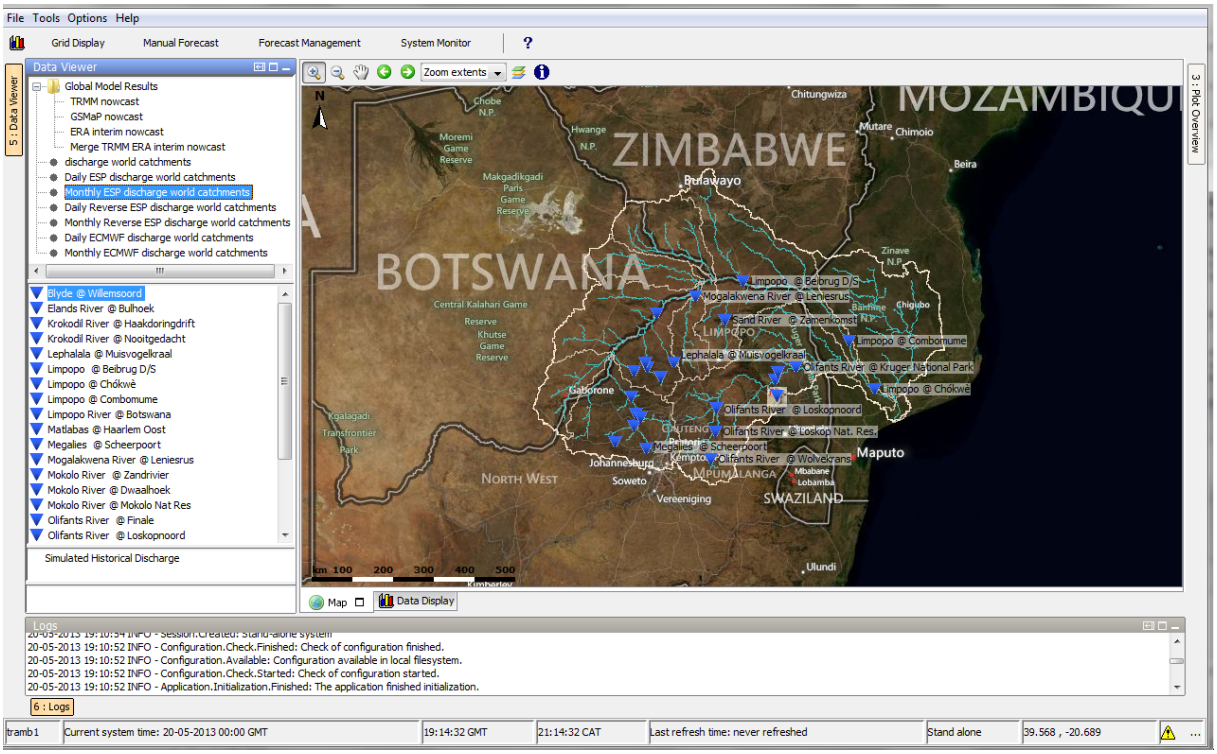



**Figure 3-18 Customized Delft-FEWS for the Limpopo river basin**

## METEOROLOGICAL FORECASTS USED

Seasonal meteorological forecasts from ECMWF system 4 - S4 are used for the Limpopo case study (see section 2.1). An ensemble of hindcasts was first obtained for the summer rainfall season over southern Africa for the period 1981 to 2010. Each year, the forecasts are issued at the start of the month from August up to February included, which means that the first forecast of the season is issued in August and include the seasonal forecast from August to January. Likewise, seasonal forecasts are issued in September, October, November, December, January and the last forecast of the season is issued in February, including the seasonal forecast from February to July. Figure 3-19 shows the imported seasonal forecast in the Delft-FEWS where one daily precipitation forecast map is displayed.

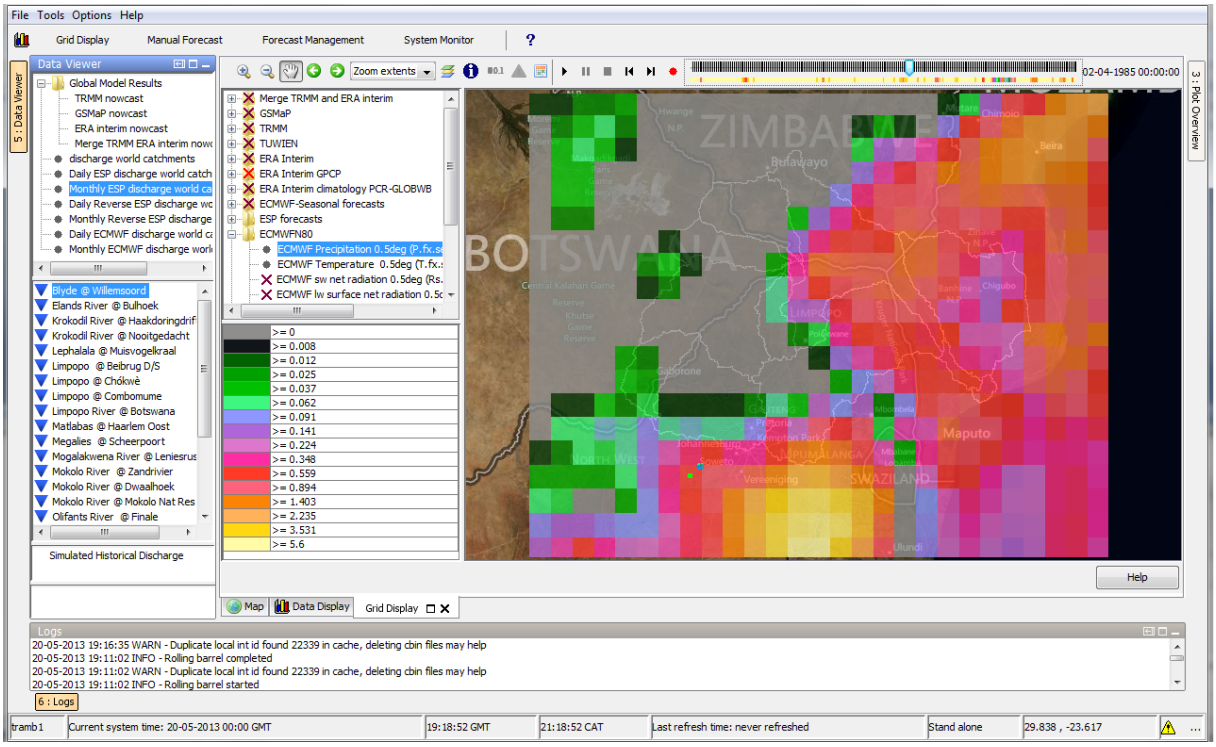



**Figure 3-19 Seasonal meteorological forecast data imported in Delft-FEWS**

## INITIAL CONDITIONS FOR THE FORECASTS

The hydrological model was run in the simulation mode for the 32 year period (1979-2010) with the ERA-Interim (GPCP corrected) forcing meteorological data. The initial conditions or states at the beginning of each month of the simulation were saved in the Delft-FEWS so as to have "warm states" or "realistic conditions" to start the forecasts when doing the retroactive forecast or hindcasts.

## TIME PERIOD OF THE SIMULATIONS

During each year of the hindcast period the seasonal forecasts are issued at the start of the month (from August to February) and results in the hydrological forecast for the following six

months. The simulations are done in a daily time step and subsequently accumulated to monthly values.

## BIAS CORRECTION METHODS OF SEASONAL FORECASTS

The precipitation dataset from the ECMWF S4 seasonal forecast was bias corrected as explained in detail in section 2.2. For the bias correction the "base" rainfall dataset used was ERA-Interim +GPCP to be consistent with the baseline simulation period.

## VERIFICATION SKILL SCORES

General verification skill scores (see Section 3.2.2) are selected to measure the skill or quality of the different forecast ensembles to predict drought indices and streamflow (as simulated by the hydrological model). The verification is performed against the simulations and the resulting skills are relative to sample climatology. The metrics were computed using the Ensemble Verification System (EVS) software (Brown et al. 2010). Moreover, the skill of the models in predicting indicators that are meaningful to the local end users in the basin is also assessed. The skill scores used for the Limpopo case study are explained in the Results section (Section 3.2.2) together with the corresponding graphs to ease the explanation.

### 3.2.2 Results

We start this section by presenting some general skill scores of the seasonal streamflow prediction, and continue by presenting the skill of the forecast for a specific decision variable in the basin. Finally we illustrate the seasonal streamflow prediction for a particular dry season (1991/92), a particular wet season (1999/2000), and a relatively dry season (1982/83). All the streamflow predictions presented in this report are computed for Station 24 (ARA-SUL E35 at Chókwe) in the Limpopo river basin. This station is the one with the largest drainage area in the basin with available discharge data and is highlighted in Figure 3-20.

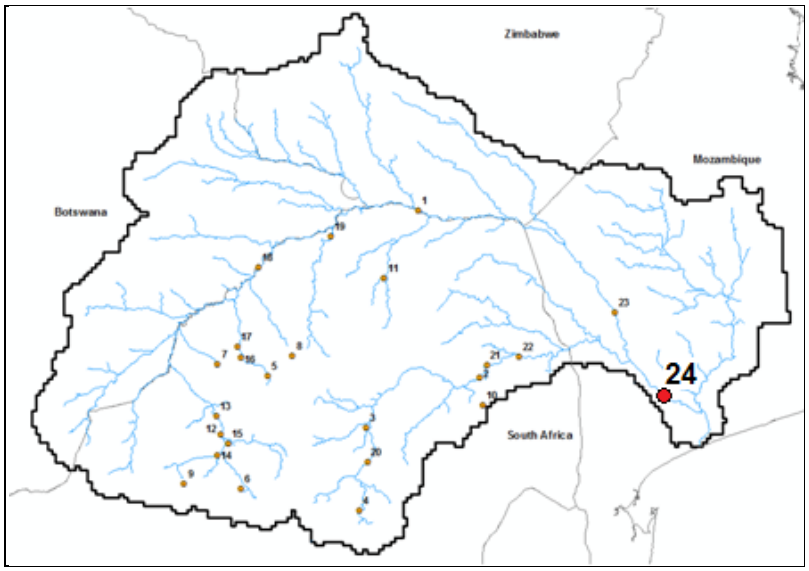



**Figure 3-20 Location of hydrometric station 24 in the Limpopo basin**

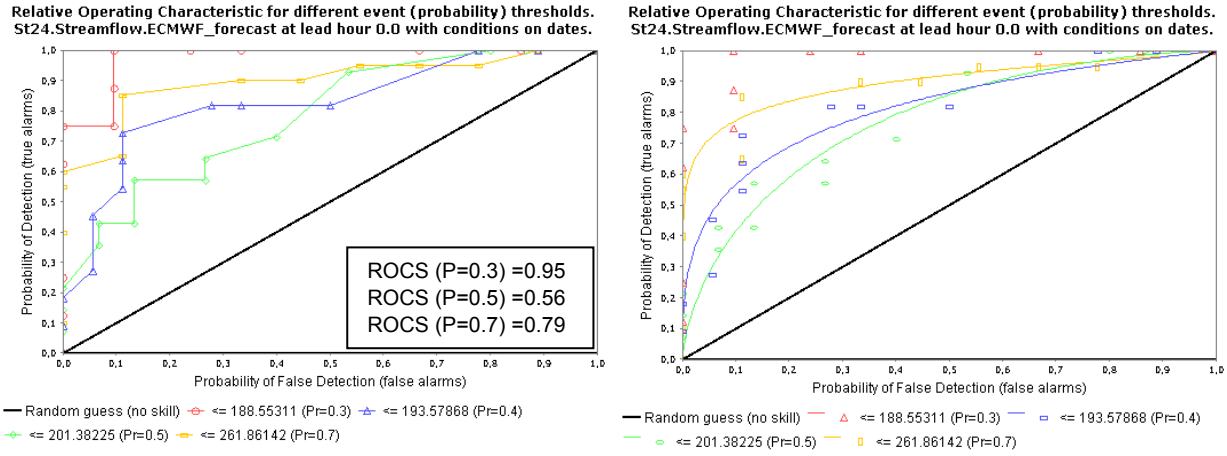
## RETROSPECTIVE FORECAST SKILL - 30 YEAR HINDCAST

The following section outlines the findings when applying the different atmospheric forecasts into the hydrological model over the 30 year retroactive period from 1981 to 2010. The analysis is carried out for different verification periods and lead times.

The ROC (relative operating characteristic, or receiver operating characteristic) diagram measures the ability of the forecast to discriminate between two alternative outcomes. It plots the hit rate or Probability of detection (POD) versus the false alarm rate or probability of false detection (POFD). It is not sensitive to bias in the forecast, so says nothing about the reliability. It is conditioned to the observations. In summary, it indicates the ability of the forecast to discriminate between events and non-events (WWRP/WGNE updated: 26 January 2012).

The ROC diagram was used here as a tool for drought detection and determination of the skill of categorical forecasts. Figure 3-21 presents the ROC diagram for the averaged seasonal runoff (verification period from October to March) in station 24, for different percentile thresholds (30%, 40% 50% and 70%) and for a lead time of zero month (the forecast is issued in October). October is the only start time of the forecast that captures the whole 6 months rainy season (from October to March) in the seasonal forecast.

In order to try to reduce the impact of the small sample size of the forecasts (to some extent) we used the "binormal approximation" whereby the POD and POFD are assumed to be bivariate normally distributed. This approximation (presented in Figure 3-21 b) together with the raw data) has a sound theoretical basis and it smoothes out some of the sampling variability, although it breaks down for very small sample sizes (e.g. for Pr=0.3 in Figure 3-21 b)) (James Brown, personal communication, May 2013; Metz and Pan (1999); Hanley (1996)).

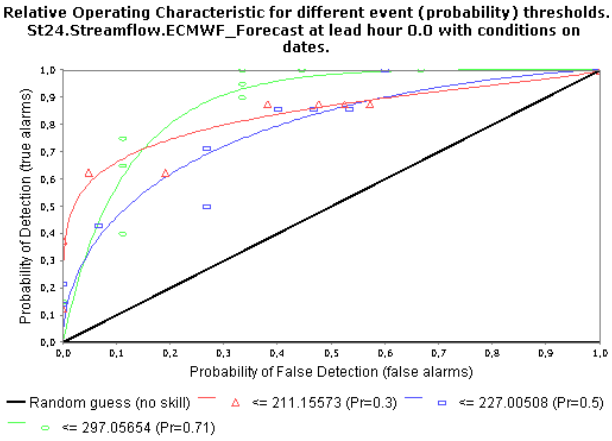



a) Original ROC diagram

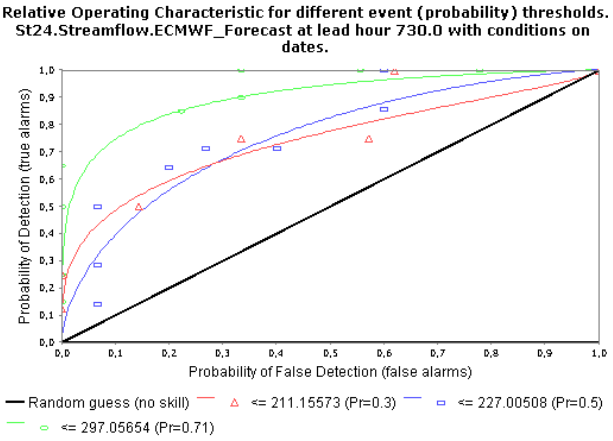
b) "Binormal approximation"of the ROC

**Figure 3-21 ROC diagram for the average seasonal runoff (Oct- March) in station 24 for a lead time of zero months. Panel a) shows the original ROC diagram and panel b) shows a "binormal approximation" whereby the POD and POFD are assumed to be bivariate normally distributed**

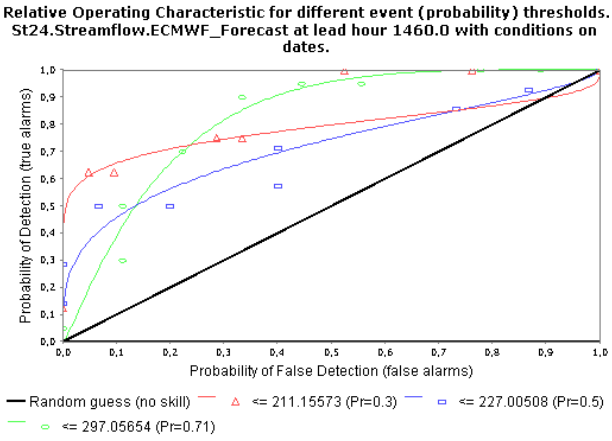
The verification was also done for forecasts issued for the same 3-montly period Dec-Feb, with forecasts issued from September to December respectively. Figure 3-22 present the ROC diagrams for the average streamflow in the 3-months period for four different lead times (zero to three months). Figure 3-23 summarizes the ROC scores (ROCS) for the different lead times and shows that in general the score tends to decrease with lead time. For the percentile 30, which indicates a drought in this study, there seems to be a slightly higher score for two months lead time than for one month lead time, but this could be a result of the small sample sizes.



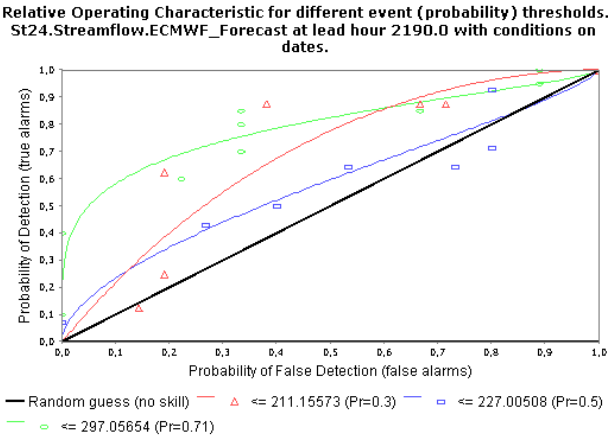

a) Forecast issued in December




b) Forecast issued in November




c) Forecast issued in October




d) Forecast issued in September

**Figure 3-22 ROC diagrams for the average 3-months runoff (Dec- Feb) in station 24 for a lead time of: a) zero months, b) 1 month, c) 2 months, and d) 3 months**

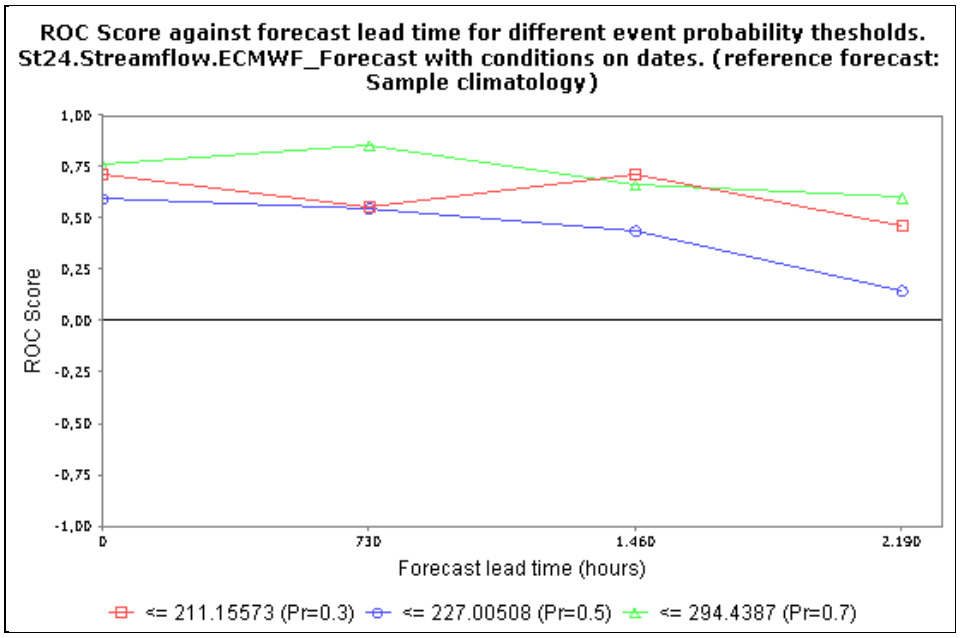
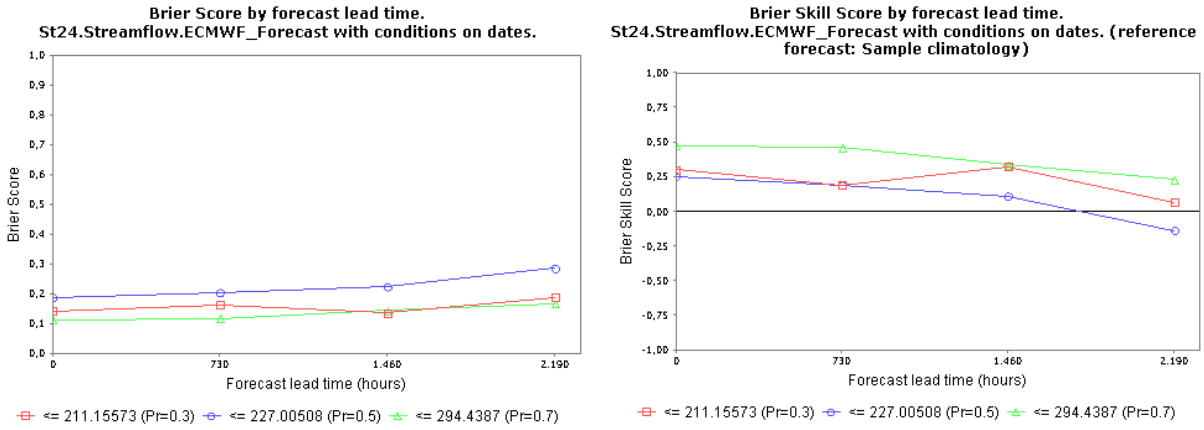



**Figure 3-23 ROC scores for the average 3-months runoff (Dec- Feb) in station 24 for different thresholds and different lead times**

For small sample size of forecasts, "lumped" metrics or scores that are less sensitive to sample size, such as the Brier Score are preferred. The Brier Score (BS [0-1]) measures the mean squared probability error and represents the magnitude of the probability forecast errors, with a perfect score of zero. The Brier Skill Score (BSS [-∞ to 1]) measures the improvement of the probabilistic forecast relative to sample climatology and indicates what the relative skill of the probabilistic forecast is over that of the climatology, in terms of predicting whether or not an event occurred (WWRP/WGNE updated: 26 January 2012). Figure 3-24 present the BS and the BSS for the average 3-months runoff (Dec- Feb) in station 24 for different thresholds and lead times. The figure shows higher skill in the forecast of droughts (Pr=0.3) and the 70 percentile flows than the mean streamflow conditions (Pr=0.5). For the case of droughts (Pr=0.3), lead times up to two months present some skill, while a lead time of three times present almost no skill.




a) b)

**Figure 3-24 a)Brier Score and b) Brier Skill score for the average 3-months runoff (Dec- Feb) in station 24 for different thresholds and lead times**

Monthly forecasts are also assessed independently without accumulation. We analyzed the skill of the seasonal streamflow forecast in terms of the Correlation Coefficient (CC) and the Relative Mean Error (RME) for the ensemble mean for Station 24. The CC and RME for each forecast month as a function of lead time is represented in Figure 3-25. In the figure, there are some white tiles (no data) as the forecast was not issued in every month of the year but only from August to February. The CC [-1 to 1] represents how well the forecast values correspond to the observed values. The CC show almost perfect score for the months of August to November, in January presents some skill only for short lead times, and it is not very skillfull from February onwards.

The RME is a measure or the relative bias in the forecasts, where a RME of zero denotes the absence of relative bias in the centre of the ensemble forecast, a positive value indicates overforecasting and a negative value indicates underforecasting. Figure 3-25 show almost no bias for the months of August to November, a slightly overforecast for the months of December and January and some underforecast from February onwards.

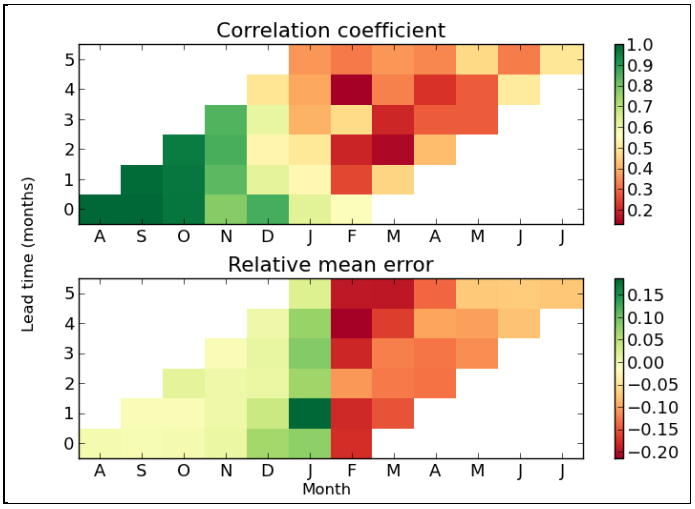
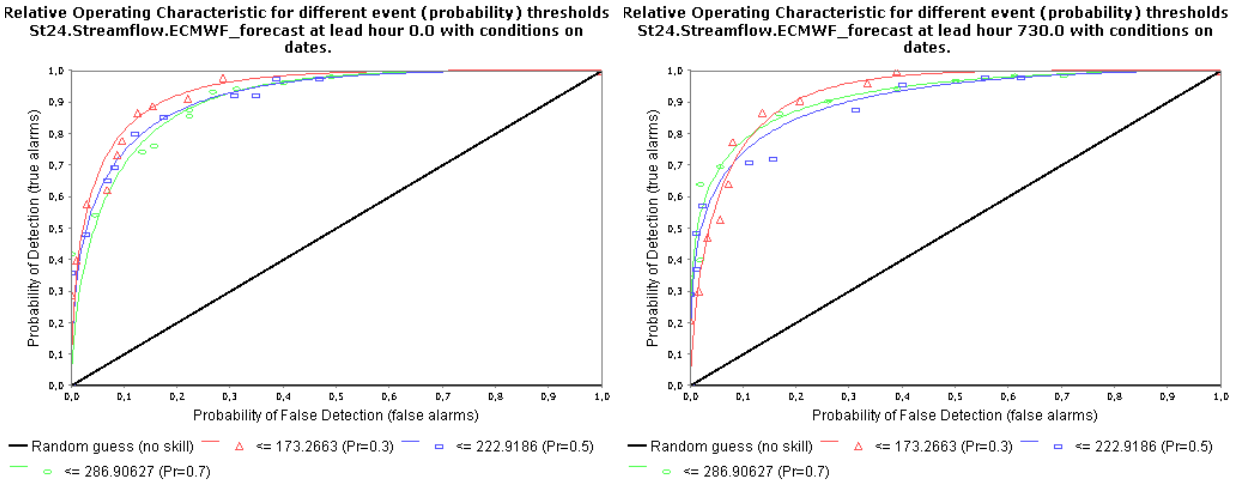



**Figure 3-25 Correlation coefficient (CC) and Relative Mean Error (RME) for each forecast month as a function of lead time**

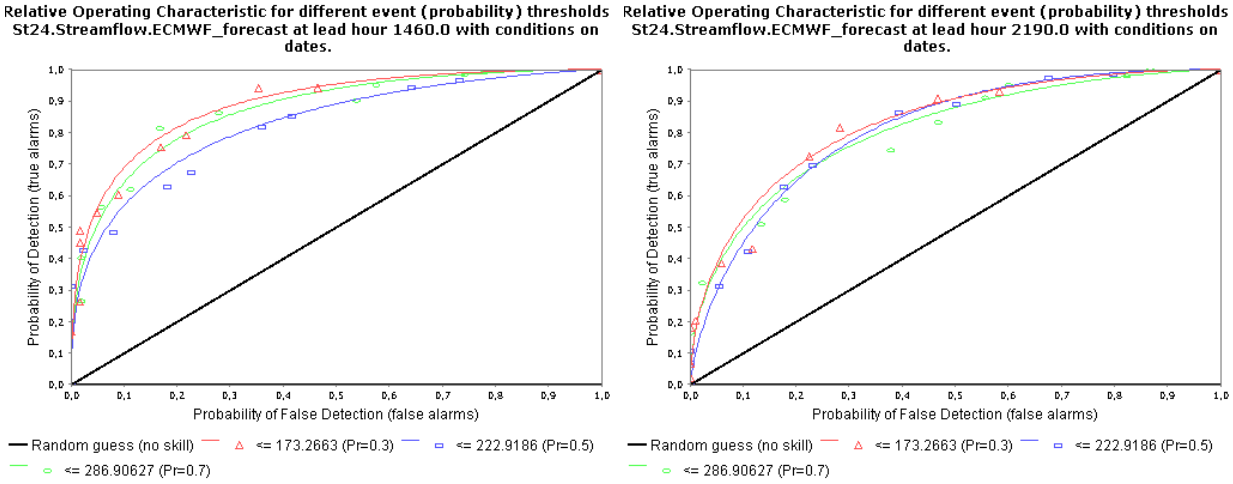
Moreover, we restricted the verification period to the rainy season months (October to March) within the overall verification period and we assessed the monthly forecasts during the rainy season as a function of lead time. The resulting ROC diagrams are presented in Figure 3-26. The figure shows a good skill for the monthly forecast during the rainy season and shorter lead times, and it clearly shows how the skill of the monthly forecasts deteriorates with increasing lead time.




a) Zero month lead time



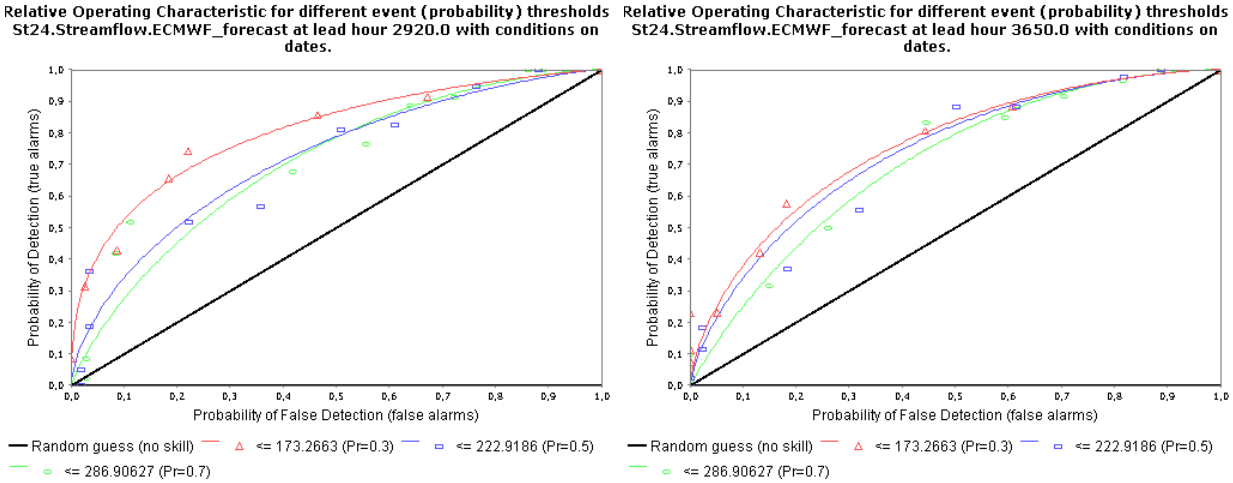

b) One month lead time




c) Two months lead time




d) Three months lead time




e) Four months lead time




f) Five months lead time

**Figure 3-26 ROC diagrams for monthly runoff (verification period: October to March) in station 24 for a lead time of: a) zero months, b) 1 month, c) 2 months, d) 3 months, e) 4 months, f) 5 months, and g) 6 months**

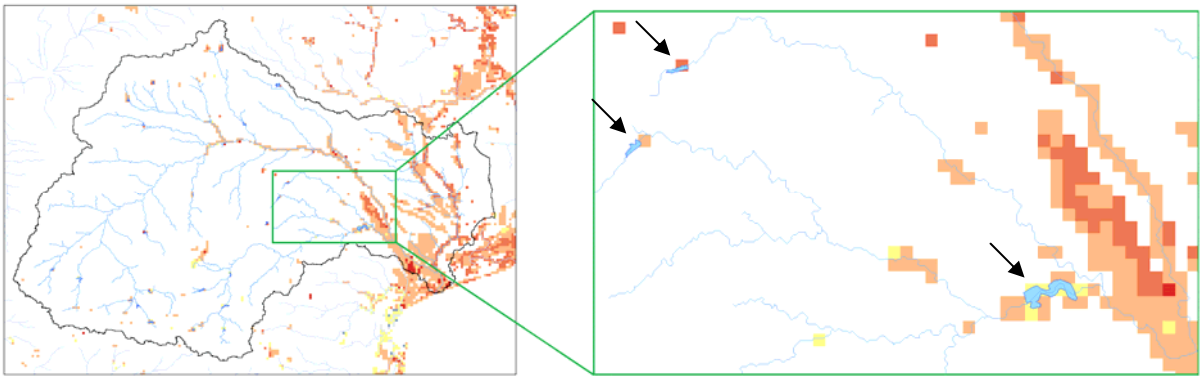
## SKILL OF SPECIFIC DECISION VARIABLE IN THE LIMPOPO RIVER BASIN

Irrigation is the major water use in the Limpopo basin accounting for 50% of the total water demand. The Tzaneen Dam, on the Groot Letaba River, currently allocates a total of 128 million $m^3$ of water per year from which 105 million $m^3$ are allocated to irrigation. However, the actual availability of water for irrigation sector is restricted depending on the water level in the dam in percentage of the full capacity (DWA 2013). The forecasted anomaly of the water level in a reservoir is a "decision variable" that can give an indication to the water managers of the percentage of irrigation demand that can be covered in the summer season. This indicator results from this forecasting system as the reservoir operation is included in the hydrological model.

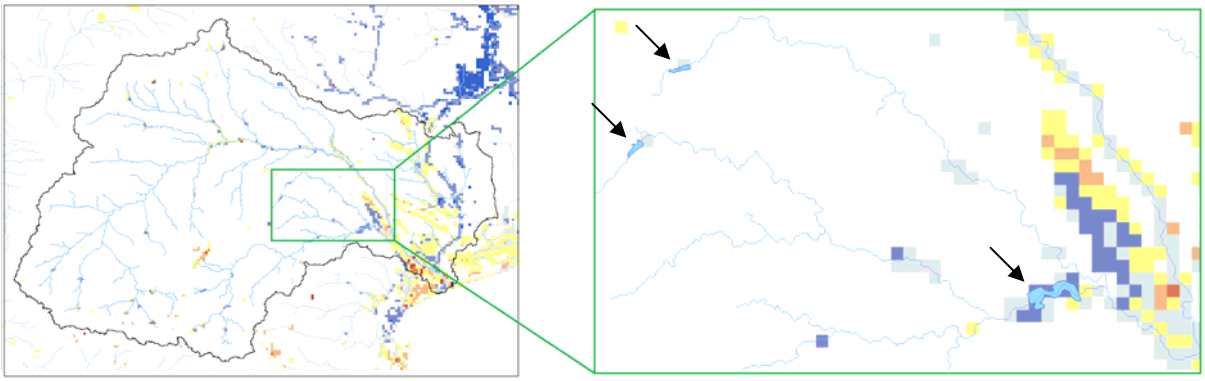
The forecasted water levels are compared with the climatology, and the anomalies of the forecasted water levels with the climatology (WLA) are ranked from 1 to 8. A rank of 1 indicates that the water level is forecasted to be within the percentile 5 and a rank of 8 indicates that the water level is forecasted to be higher than the percentile 95. In between, a rank of 2, 3, 4, 5, 6 and 7 correspond to values within percentiles 25, 40, 50, 60, 75, and 95 respectively.

A preliminary assessment of the forecasted WLA was undertaken in this study to check the forecast for a couple of years. A more in depth assessment will be completed in a subsequent stage of this study. In this preliminary assessment we computed a rough mean seasonal WLA rank for a forecast issued in October (and therefore covering the forecast from October to March). To compute the "mean seasonal WLA rank" we first computed the mean seasonal rank (mean monthly WLA rank from October to March) for each ensemble member and then we computed the mean value over the 15 ensemble members. These values are presented in Figure 3-27 for the dry season of 1991/92 (a) and for the wet season 1999/2000 (b). The figure shows that the forecasted average rank for the season of 1991/92 is always on the "red" side of the scale, indicating drier than normal conditions and the season 1999/2000 is generally on the "blue" side of the scale indicating wetter than normal conditions. A zoom is made to show the rank of the seasonal water level anomaly in three of the reservoirs in the basin (which are marked with a black arrow).

It seems that the forecasted water level anomaly in the reservoirs could be a potential indicator with significant value for the decision makers in the Limpopo river basin concerning available water for irrigation during the season. This indicator can be simulated and assessed to verify the skill of the seasonal forecast over the basin.



a) Mean WLA rank for the season 1991/92



b) Mean WLA rank for the season 1999/2000


**Figure 3-27 Mean rank of the forecasted water level anomaly**

## ENSEMBLE STREAMFLOW PREDICTION FOR SELECTED SEASONS

Figure 3-28, Figure 3-29, and Figure 3-30 present the streamflow anomaly forecast for three particular seasons for the Limpopo river basin. The graphs show the forecasted runoff anomaly with respect to the median[2] Ensemble Streamflow Prediction (ESP) which is used as reference forecast. The ESP predicts future streamflow from the current initial conditions (warm state) in the hydrological model with historical meteorological data (observed meteorology from the last 30 years). The procedure assumes that meteorological events that occurred in the past are representative of events that may occur in the future (Day 1985). This forecasting procedure is normally used in the absence of a seasonal forecast. In this section we measure the improvement of the ECMWF seasonal forecast S4 compared to the ESP forecast for selected seasons.

Negative predicted runoff anomalies indicate that the seasonal forecast points to a drier season than the forecast based on the ESP. In the same way, positive predicted runoff

[2] We chose to use the median instead of the average to avoid the "mean" value to be highly influenced by one anomalous season (1999/2000) with extremely high precipitation/discharge values.

anomalies shows that the seasonal forecast predicts a wetter season than the forecast based on the ESP.

Figure 3-28 presents the streamflow anomaly forecast for the dry season 1991/92 where each panel corresponds to a different initial forecast time (from August 1991 to February 1992). The runoff baseline simulation and the 30% percentile of the ESP are also shown in the graph. We defined this 30% percentile threshold as an indication of a moderate to severe drought and created 2x2 contingency tables to verify the ability of the forecast in predicting moderate to severe droughts. The Probability of the event occurring in the forecast (FP [0-1] = hits) / (hits + misses)) was computed from the contingency tables and is presented on the upper right corner of every panel. This score represents the fraction of the observed "yes" events that where correctly forecasted from all the ensembles. This score seems to give more information than others in this extreme event case as we already know that this event occurred. For this reason other scores such as the false alarm ratio will have a perfect score. The figure shows that all the forecasts point to a drier season than the ESP prediction given by the negative runoff anomaly values for the ensemble median. Moreover, for every panel the shaded areas representing the ensemble distribution between the percentiles 30 to 70 is mostly covering negative discharge anomalies. The seasonal forecast seems to offer a clearer signal than the ESP forecast, given that we know that this is one of the driest seasons in the recent history. From the FP values it can be inferred that even though the seasonal forecast is good in predicting drier conditions than the ESP forecast, the ability in predicting the severity of the drought is average.

Similarly, Figure 3-29 presents the streamflow forecast for the very wet season 1999/2000. The figure shows that the seasonal forecast point to a wetter season than the ESP forecast, given by the positive runoff anomaly values for the ensemble median and the ensemble distribution between the percentiles 30 and 70.

The runoff baseline simulation is shown in the graphs but in almost every case it is trimmed due to a scale issue given that the observed discharge for this year was exceptionally high. Panel h) of Figure 3-29 shows the ensemble discharge anomaly for the forecast issued in February (same as the panel g)) but in a scale that allows the simulated discharge to be visible in the plot.

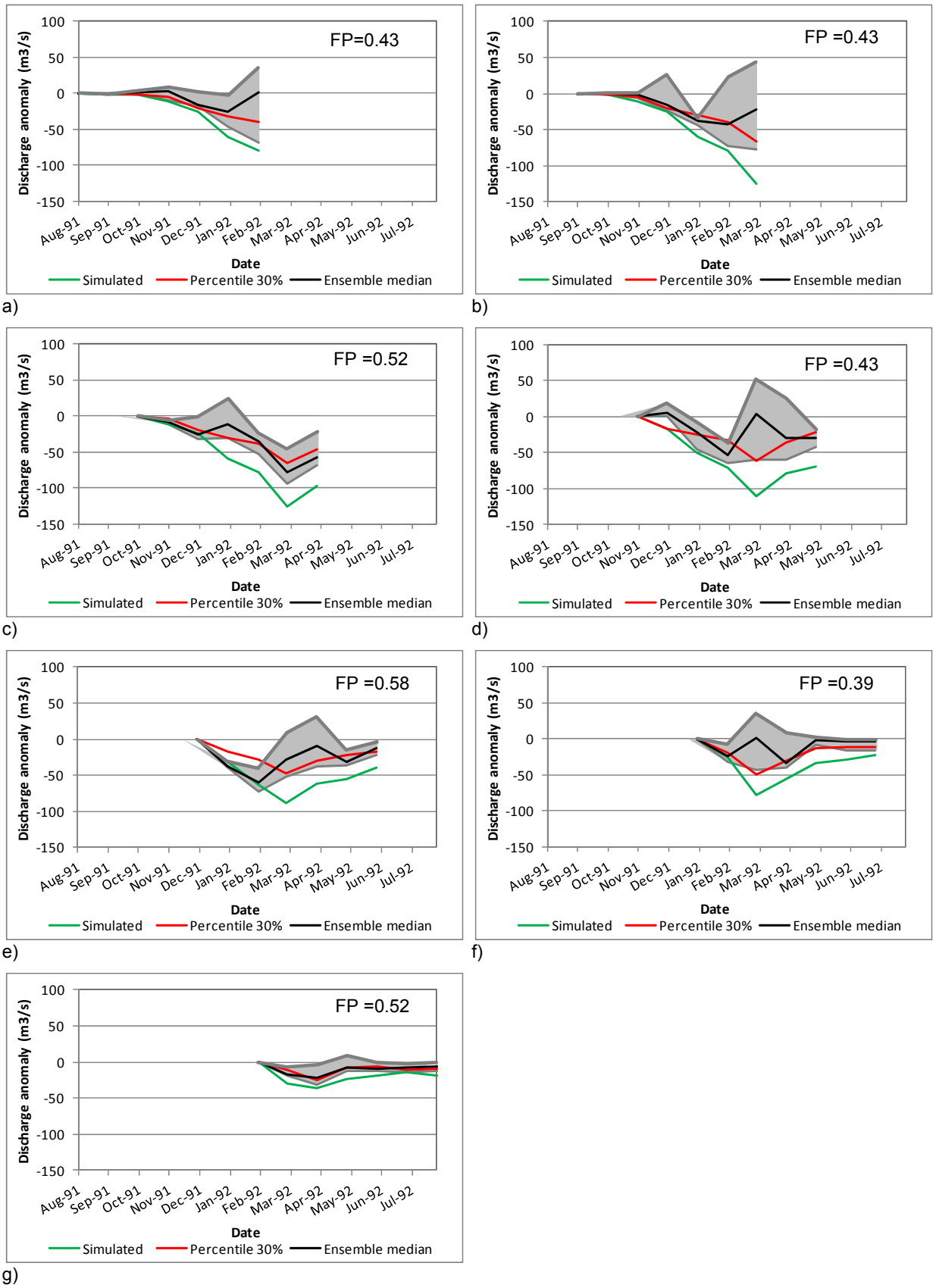



**Figure 3-28 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1991/92. The shaded areas represent the ensemble distribution between the percentiles 30 and 70.**

Figure 3-29 also presents the percentile 70 of the ESP (30% probability of exceedance) as an indication of a very wet season. We created 2x2 contingency tables to verify the ability of the forecast in predicting this very wet year. The FP is presented on the upper right corner of

every panel. From the FP values it can be inferred that even though the seasonal forecast is good in predicting wetter conditions than the ESP forecast, the ability to predict the severity of the flood is average, with slightly higher FP for the forecasts starting in November to January.

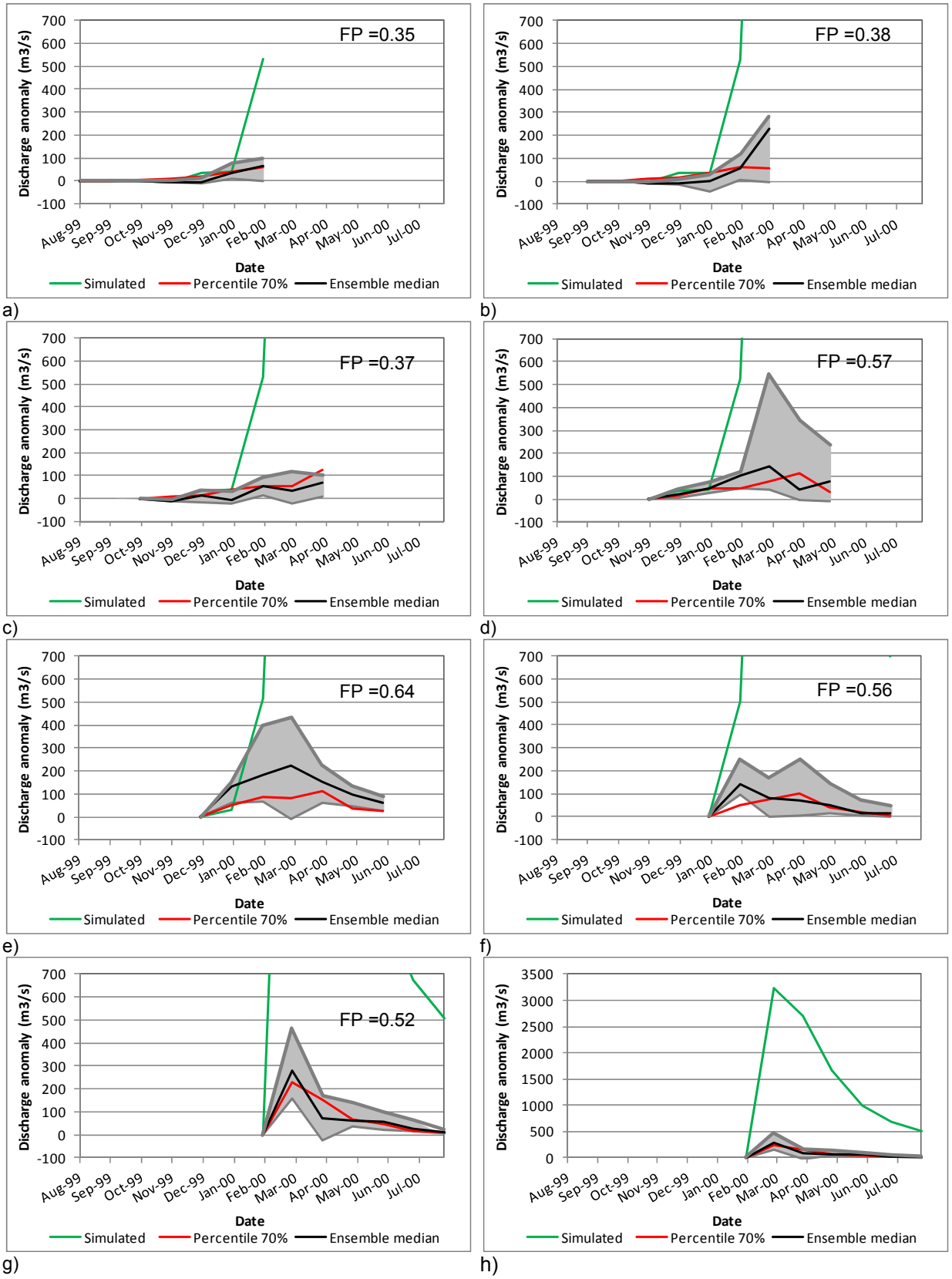



**Figure 3-29 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1999/2000. The shaded areas represent the ensemble distribution between the percentiles 30 and 70.**

Finally, Figure 3-30 presents the streamflow anomaly forecast for the dry season 1982/1983. This is to illustrate the seasonal streamflow forecast for another dry season but not as extreme as the drought of 1991/92.

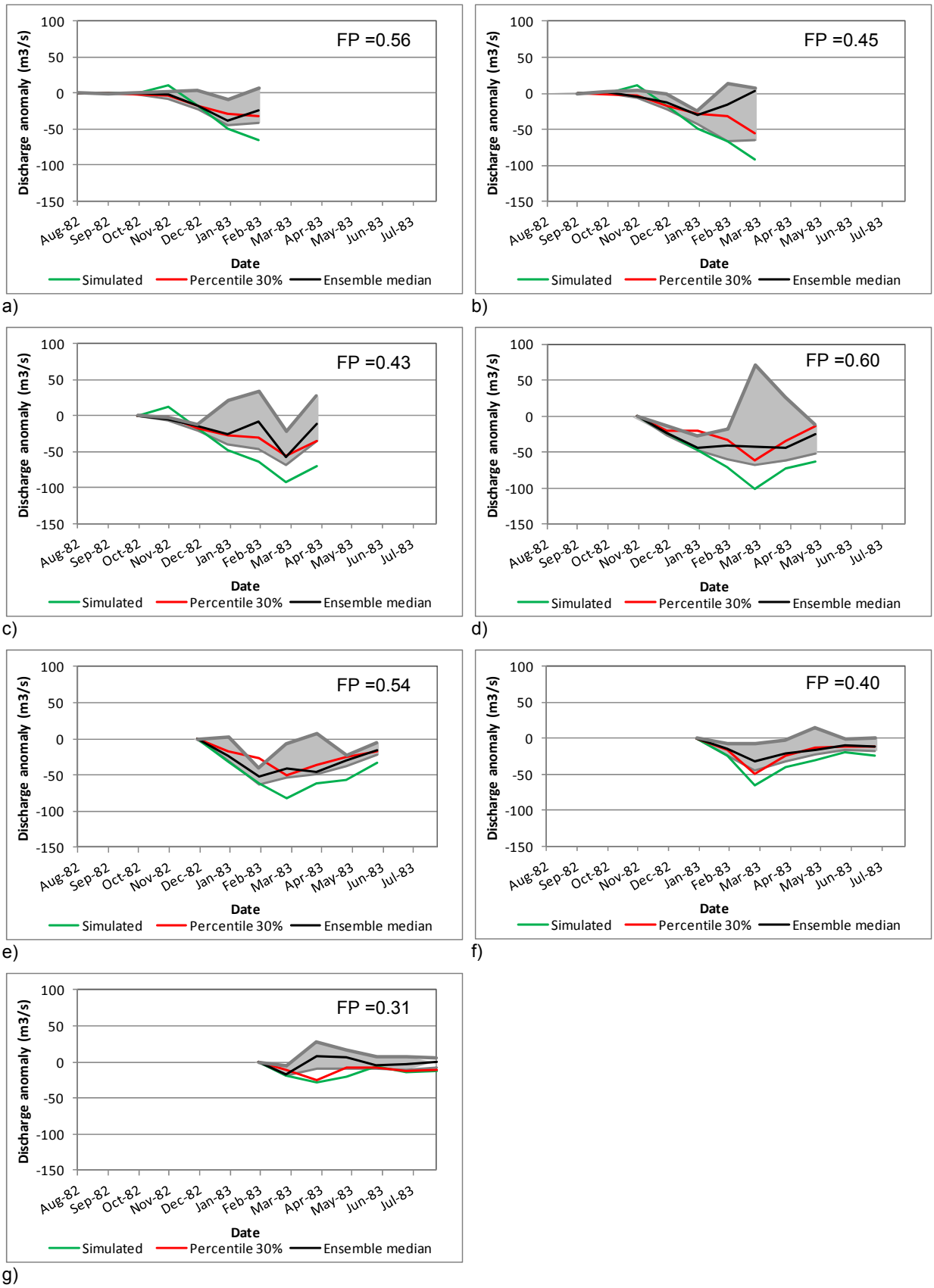



**Figure 3-30 Seasonal streamflow anomaly forecasts for the Limpopo for the season 1982/83. The shaded areas represent the ensemble distribution between the percentiles 30 and 70**

The seasonal forecast predicts a drier season than the ESP forecast (negative median of the ensemble streamflow anomaly) (Figure 3-30) but the ability to forecast a moderate to extreme drought is average.

### 3.2.3 Concluding remarks

The dynamic multi ensemble seasonal forecast for the Limpopo river basin forced with ECMWF S4 seasonal forecasts show higher skill in predicting low flows and high flows than in predicting near normal conditions. The predictive skills of these forecasts, which worsen with lead time, are higher than using the climatology. The evaluation of the forecasts for some selected seasons when a drought or flood actually occurred suggests that the hydrological forecast using ECMWF S4 seasonal meteorological forecast is improved and offers a clearer signal compared to the use of the ensemble streamflow prediction (ESP). These selected seasons suggest that seasonal forecasts based on numerical weather prediction models can offer better seasonal hydrological forecasts than ESP.

# 4. FORECAST OF HYDROLOGICAL DROUGHTS FOR THE NIGER CASE STUDY BASIN

## 4.1 MODEL AND SIMULATIONS SET-UP

### 4.1.1 Description of OPIDIN platform

**THE INNER NIGER DELTA CASE STUDY**

The Inner Niger Delta (IND) is one of the largest floodplains in Africa. The topographical maps of the Institut Géographique National (IGN) reveal that the inundation area measures 36 470 km$^2$, including 5340 km$^2$ of levees, dunes and other islands within that area (Figure 4-1). It also shows that water coverage declines from 31130 km$^2$ in wet periods to 3840 km$^2$ in the dry period. The entire floodplain area is included in the 41195 km$^2$ designated as a Ramsar Wetland Site of International Importance in January 2004 (Zwarts 2009).

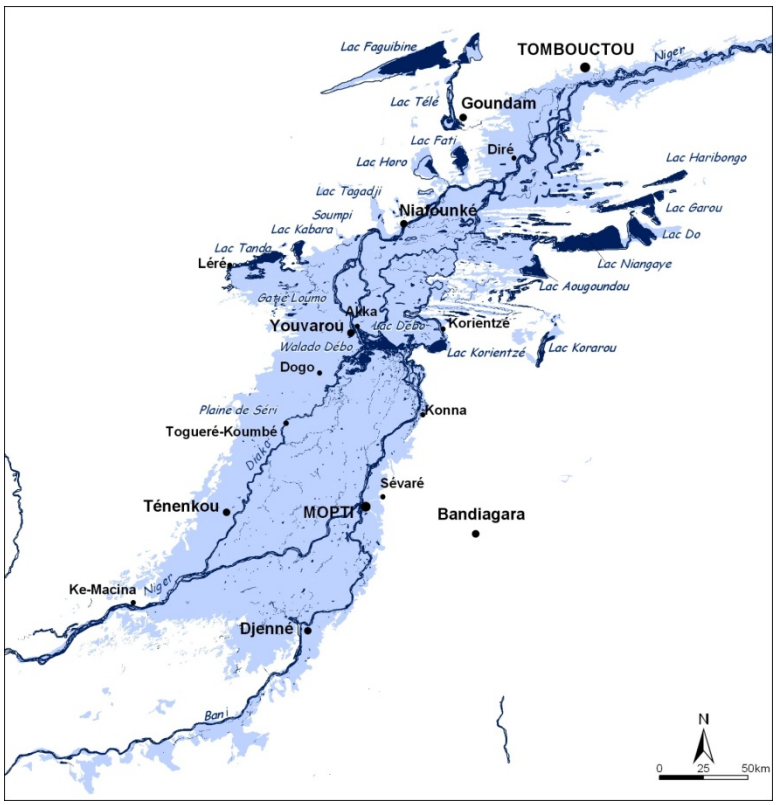



**Figure 4-1 The floodplains (light blue) and permanent water bodies (dark blue) of the Inner Niger Delta, as indicated on the topographical maps of the Institut Géographique National (IGN). The maps are from 1956, and based on aerial photographs and field work in the early 1950s, a period with very high flood; source: Zwarts, 2009.**

The economy and the ecosystem of the IND are strongly driven by these flood dynamics; in other words the IND is a flood-dependent ecosystem and economy. Accordingly, changes in its dynamics (level, discharge and timing) have significant social, economic and environmental consequences. As an example, rice production, the staple food in the IND, fails when water levels are too shallow. Given recent changes in the water regime, as a consequence of upstream dams and diversions and of climate change and variability,

farmers may lack the knowledge to choose the best suited areas for their production. Fish catches are determined by the area flooded which vary with flooding levels. Catches vary between 50,000 to 100,000 tons for minor and major flood years, respectively (Zwarts et al. 2005). The uncertainty of this fish production places the livelihoods of the 300,000 fishermen in this area under continuous pressure. Also, cattle grazing are planned since historic times in conformity to cultivated areas and flood levels. Again, a limited flooding, or changes in the flood regime through upstream interventions and global climate forcing, affects the levels of livestock productivity.

Understanding the flood dynamics better and being able to predict the flood dynamics significantly may support stakeholders in anticipating their economic activities and by such reduces their livelihood risks in the delta. The life of (rural) people in the IND is governed by the flooding. It might help them in several ways to know some months before what would be the peak flood level and when this peak will be reached, at what time the flood level will have been declined to a certain level. Given the changes to be expected in the Inner Niger Delta due to the ongoing climate change, predicting the flood becomes also gradually more important. In this context, OPIDIN (Outil de Prédiction de l´inondation dans le Delta Intérieur du Niger) was developed in the IND to predict the height and the timing of flood peak and to function as an early warning system and a decision support tool for the people, being either fishermen or farmers, to achieve food security.

## OPIDIN AN EXPERIMENTAL STAKEHOLDER PLATFORM FOR EARLY WARNING SYSTEM

OPIDIN (Outil de Prédiction de l´inondation dans le Delta Intérieur du Niger) tool is the product of simple statistical relationship based on historical water height time series monitored at different gauged stations by DNH (Direction Nationale de l´Hydraulique du Mali) along the IND. The water level and timing of the annual peak flood as well as the timing of annual deflood are projected at different dates some months in advance. The simplicity of the approach, requiring only the participation of technical and local services to read the monitored water level, enables to generalize the use and the dissemination of the tool in many locations of the IND for a relative low cost of implementation which can, in return, remain very useful for the planning of various activities such as agriculture, livestock herding, fishing and navigation sectors. For example, projecting the timing of annual deflood permits to know when floodplains will become potentially accessible for grazing cattle during receding water. This information also supports the decision of local stakeholders and authorities to select the appropriate date regulating each year the access of rich and highly nutritive pastureland from the floodplain to the livestock (Figure 4-2). The flood prediction function of OPIDIN can enable various stakeholders in the IND to plan their livelihood activities and to make no-regret investments. Fishers can select fishing sites, modify their

practices, and optimize preparatory activities (investments to purchase appropriate fishing gears, risk free investment). Herders can plan to use different regional transhumance routes (temporary migration to other parts of the country) and trails in the IND to adapt to the timing of grazing in bourgou pastures. Timing and rural planning of rice production sites and agricultural practices are subject to differ according to forecast and might encourage early starting of rice farming in lowland areas or to use dry resistant type of rice (field preparation, removal of animals). Timing of socio-cultural activities and festivities often determined by the flooding pattern of the IND could be anticipated. Last but not least, people adopting as adaptation strategies temporary migration (leaving the Inner Niger Delta in years with a poor flood to work elsewhere in Mali or even abroad), may plan to leave their village when they know already in August that the flood will be low.

**Figure 4-2: cattle passing the Niger river @WetlandsInternationalSévaré/A&W**

An initial version of OPIDIN was developed in 2009 within a study carried out by Royal Haskoning (lead), Altenburg & Wymenga Ecological Consultants and Wetland International. This study was financed by “Partners for Water”, a joint initiative of six departments of the Government of the Netherlands. In the framework of the Wetland & Livelihoods Project GIRE (Gestion integrée de resources naturelles), the Malian antenna of Wetlands International took the initiative to further explore the possibilities to extend the model and asked A&W to investigate and list potential improvement of the tool.

In this context, DEWFORA project aims to support the continuous improvement and the development of OPIDIN tool. In collaboration with AFROMAISON fp7 project, workshops were organized with key regional and local stakeholders of the OPIDIN stakeholder platform (Figure 4-3). Beside the experimentation of other approaches (use of SWIM eco-hydrological model coupled with seasonal weather forecast and assessment of local rainfall forecast), DEWFORA developed with this report extended methods to improve the skill assessment of the initial version of OPIDIN tool. The report highlights the remaining uncertainties by presenting the general deviation of retroactive forecast from two representative gauge stations in the IND using OPIDIN methods against observed water height and related timing.

By defining two types of classes, errors were expressed as pure quantitative deviation (in cm. or days) and as communication failure when using the operating early warning forecasting system communication protocol. With this work, DEWFORA project aims to bridge the gap between two major phases of OPIDIN project which are now the object of further details in the following paragraph.

**Figure 4-3: DEWFORA/AFROMAISON workshop in July 2012 grouping some of the key regional and local stakeholders of the OPIDIN platform @WetlandsInternationalSévaré**

The development of OPIDIN tool is marked by series of initiatives, major phases in the development of the OPIDIN stakeholder and early warning system platform can be as well dissociated:

1. In 2008-2009, OPIDIN first started to involve national, regional and local actors. The cornerstone's structures were the DNH (Direction Nationale de l´Hydraulique) and the DRH of Mopti (Direction Régionale de l´Hydraulique). At the national level, the others stakeholders during this developing phase were composed from all the ministries linked to water ressources management and the related decentralized administration units (Ministères de l'Agriculture, Elevage et Pêche, Environnement et Assainissement, Eau et Energie), representatives from the main offices diverting water for irrigation (Office du Niger, Office de Développement Rural de Selingue, Office Riz Ségou) and some national and international NGO´s (UICN). At the regional level, the authorities, representatives of the interest of sectorial economic activities were represented by decentralized organization (Gouvernorat de Mopti (Présidence), Direction Régionale de la Pêche, Office Riz Mopti, Chambre Régionale de la Chambre d'Agriculture). The initiative integrated some social organizations from the community level representing fishermen, herders and farmers in order to sound their interest in the use of the forecast information but also to collect knowledge and experiences on traditional forecast methods. The dissemination of information from OPIDIN was at this stage processed through scheduled stakeholder meetings at national, regional and local levels where the results and the performance of the tool

for the previous year were presented. Based on this first meeting, information were diffused via local radio or via leaders of socio-professional groups. At the gauges stations in Mopti and Akka, simple tables were published enabling any interested persons to know the timing and the water level for Mopti and Akka station based on the present water level at the station. In this frame, the dissemination of inundation maps of the southern and central part of the delta with different level of floods helped to interpret spatially the annual propagation of the flood based on the present evolution of the water level measured at Mopti. Finally, an effort to project specific annual deflood threshold was done to deliver crucial information to select dates for the traditional crossing of the livestock.

2. In 2009-2012, a new developing phase permitted to structure the regional and local representatives of the Mopti region into a permanent stakeholder platform with the DRH of Mopti as a leading structure and Wetlands International as a technical support. An effort to further integrate key regional and community level representatives permitted to better associate rural development services of the regional authorities (agriculture, herding, fisheries), the ABFN (regional Niger river Basin Agency), the FODESA (Fonds de Développement en Zone Sahélienne) and more important open a dialogue with the coordination of the Dioros (the “water masters” which are the traditional pastureland manager in the IND). For the logistics of the platform, Wetlands International is now responsible to develop information to project flood with a pre-validaton of the information from the DRH and A&W consultancy. First , the DRH transmits the information to the Gouvernorat from the region of Mopti and then to the other stakeholders of the platform via a meeting or via a bulletin. Each stakeholder use their network and partners in the IND to disseminate the information of the projected flooding pattern for the coming season. The Gouvernorat calls the conference of the Bourgoutière of the Delta in which the information is used to select the crossing date of the livestock at Diafarabé. In 2009, based on OPIDIN method and on the flood peak water level, the crossing date was the 21 of November
3. A new project funded by the Dutsch Embassy is about to be launched this year. The objective of the planned improvements under the Dutch Embassy funded project is to communicate OPIDIN predictions to a much wider audience and to bring it directly to grass-root level decision-making. Following results have to be achieved to reach this objective
    a. An inundation atlas of the IND, being widely distributed and easy to understand, explaining how the peak flood level and flood extent may be derived from the actual water level in Mopti and Akka in August/September. The Atlas gives tables allowing the reader to extract the predicted flood level and flood extent from the

water level measured in Akka and Mopti between 1 August and 15 October. The report also gives detailed maps showing the areas being flooded at different flood levels. The maps will form together an atlas. To make the maps easier to use, detailed topographical information will be added (Cf. section New joint functionalities).

b. The same information will be distributed digitally (like on CDs), allowing people to safe this information to their own computer. The same information may also be downloaded from the OPIDIN site (see next point).

c. A dedicated OPIDIN website will be made giving access to an online version of OPIDIN. This online version will allow users to get requested information in a few mouse clicks in a graphical user interface. Such requested may be, for instance, the predicted flood curve, the predicted peak flood level compared to the peak flood levels during previous years, the predicted flooding of the IND. The user will be given the possibility to get spatial information in larger details, i.e. zoom-able maps showing the flood extent. The internet version will not only give the predicted level of the peak flood, but also the expected inundation (showing with different colours whether an area will be flooded for sure or likely be flooded). The map will be detailed (resolution 30 m) and easy to read (by adding topographical data: Cf. section New joint functionalities).

d. Actual water levels in Mopti and other hydrological stations will be updated in the online OPIDIN tool as frequently as possible between 1 Augustus and 15 October. This will allow people in the IND with internet access to use the tool itself and to predict flooding based on the actual or in any case most recently measured water levels.

e. Local people will be regularly informed with latest flood prediction updates in three ways:

   i. Key persons in the IND will be informed by mobile phones about the actual water levels and the consequences regarding the predicted flooding and given recent information regarding the seasonal rainfall forecasting tools of the "Commission Nationale d'Hydro-Agro-forestry-pastoral".

   ii. The same information will be regularly broadcasted local radio stations to inform the people in this time of the year about the expected flood level in an non-technical way, e.g. “flood will be about as high as two year ago and thus not as low as last year”.

iii. A digital weekly bulletin will be made available on the OPIDIN-site in August and September giving an updated summary about the actual hydrological situation and the predicted flooding. These bulletins will also refer seasonal rainfall forecasting tools. These weekly bulletins will be sent automatically to whom having indicated they want to be informed.

### 4.1.2 Description of the OPIDIN tool (initial version)

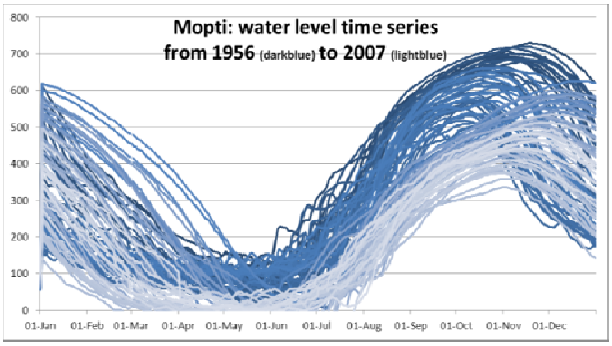
## INTRODUCTION

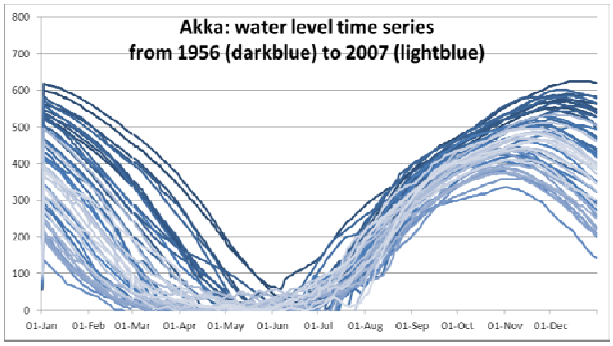
The daily water levels in the Malian portion of the basin of the Niger have been measured by the National Direction of Hydrology (DNH) at different locations for decades. This resulted in time series of great value. On the basis of an analysis of these daily water level measurements at Akka and Mopti (Figures 4-5 and Figure 4-4) between 1956 and 2007, the first version of OPIDIN tool derived clear statistical functions, amongst which are functions for the maximal flood levels and timing at these locations based on the observed water levels in August and later. For each year the water level and the timing of the annual flood peak and of specific annual deflood were correlated with linear regression functions to the water level measured at Mopti or Akka the same year at specific dates ($10^{th}$ of July, $20^{th}$ of July, $30^{th}$ of July, $10^{th}$ of August, $20^{th}$ of August, $30^{th}$ of August, $10^{th}$ of September, $20^{th}$ of September and $30^{th}$ of September). Since it only refers to statistic and do not require to represent and compute in depth the complex flood propagation and altered hydrological regimes of the delta, OPIDIN remains a user friendly tool for stakeholder. The only input requested to the stakeholder is a single lecture of water level at a single gauge which is sufficient to assess based on a table the expected annual flood peak height and timing. Based on the same linear regression functions, OPIDIN already published and make public at Akka and Mopti sites a first version of tables relating the present measurement the user has to read from the gauges stations at the specific dates mentioned above to the predicted peak flood (date and level) calculated following the OPIDIN method.

**Figure 4-4:** ***The scale at Akka*** **@WetlandsInternationalSévaré/A&W**

## HYDROLOGICAL MONITORING USED

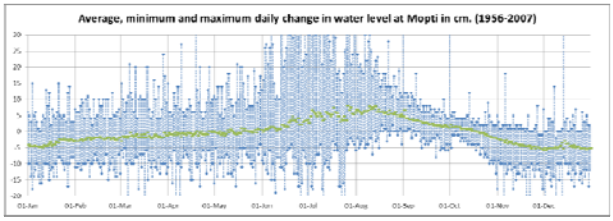
Fortuitously, the seasonal rise and retreat of the flood in the Inner Niger Delta has been measured daily by the DNH at several hydrological stations over many decades, producing a time series of great value. In average, flood height reaches 6.1 m. the 21th of October at Mopti and 4.8 m. the 16th of November at Akka, slowly engulfing an area of 400 by 100 km (Cf. next two figures). The flood peak is then in average delayed of 26 days and decreases of 1.3 m. between Mopti and Akka due to the propagation and water losses by infiltration and evapotranspiration processes. With the use of other stations, we can see that the flooding in the northern Delta is delayed by two months due to the shallow northward slope of the floodplains. As a consequence, the measured flood levels in the southern Delta in September can be used to predict the peak flood level in the northern Delta (usually in December or January). The water level at Mopti is low from April to June, and begins to rise in July when the same processes are observed at Akka with a shift of one month. At Mopti, in years of low river flow, the water reaches a height of about 3.5 m, peaking in late October. At high river flows, although the water level rises at the same daily rate, it does so over a longer period, peaking at 6 to 7 m by late December. In the figures, the gradual colors from dark to light blue characterize the chronology of the water level time series, darkblue colors being associated with the older years from the time series. We can clearly dissociate observe a gradual decrease of the hydrological regime at both stations with particular high and extended water level in the late fifties and sixties. Usually, lower-level floods cover floodplains for four months only (October-February), but high floods inundate them for twice as long (September-April). During the flooding (*crue*) and the deflooding (*decrue*) the water level rises and declines 3-5 cm per day.

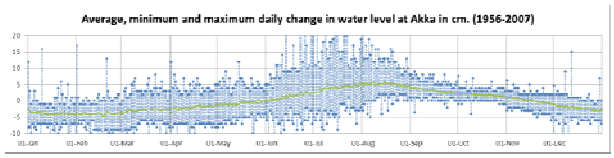



**Figures 4-5 and 4-6: Water level time series in cm.at Mopti and Akka from 1956 to 2007 (from darkblue to lightblue); source: DNH**

The daily flood curves at Mopti and Akka were also used to calculate the daily rise of the water level. On average, the water level starts to rise in early June (Cf. next two Figures). The rate of rising gradually increases to reach a maximum of 4 to 6 cm per day between mid-July and late August. Later in the season the rise of the water level continues more slowly. The figure shows also a rather large variation in the rate of flooding between years, especially in the early season. A part of the variation in the daily increase of the flood level may be attributed to the peak flood level. Obviously, the water level rises in wet and dry years at the same rate of 5-6 cm at a maximum, on average, but the period during which this maximal rising takes place is in wet years twice as long as in dry years. The difference is still larger when the dry years are compared to the extremely wet years before 1968 (*average* peak flood level: 572 cm, thus 50 cm above both recent wet years). There is one complication in such an analysis, however. Since 1982, the Selingue reservoir is filled with 2.1 km3, as a consequence of which the river flow is severely reduced in July and to a lesser degree later in the season. In those wet years before 1982 the flood rose in July at the same rate as in August. For the present analysis, we may conclude that the degree of flooding shows a clear seasonal variation. It is also obvious that if the flood rises more than average in July, this may indicate that the flood will reach a high peak level.




**Figures 4-7 and 4-8: Average, minimum and maximum daily change in water level at Mopti and Akka in cm. (based on 1956-2007 time series); source: DNH**

## STATISTICAL METHOD APPLIED

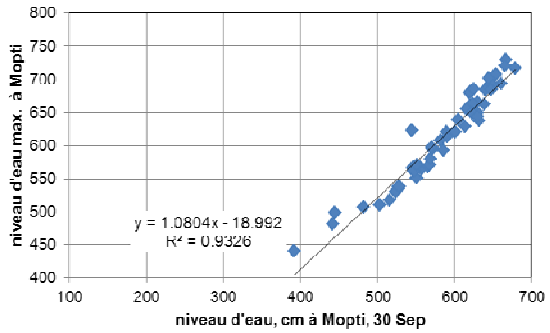
The initial version of OPIDIN is based on the two gauge stations Mopti and Akka presented in the section above from which five different statistical relationships were built:

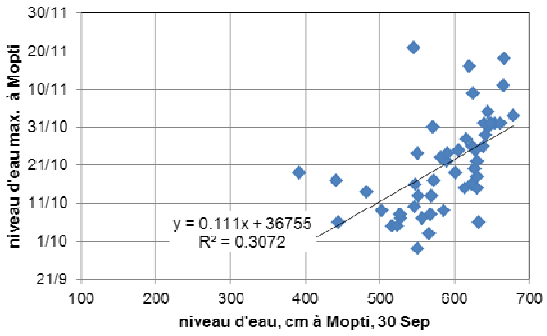
- Measured water level in July-September in Mopti to predict the timing and the level of the peak flooding also in Mopti.

- Measured water level in July-September in Akka to predict the timing and the level of the peak flooding also in Akka,
- Measured water level in July-September in Mopti to predict the timing and the level of the peak flooding in Akka.
- Measured water level in July-September in Akka to predict the timing of the deflooding also in Akka,
- Measured water level in July-September in Mopti to predict the timing of the deflooding in Akka

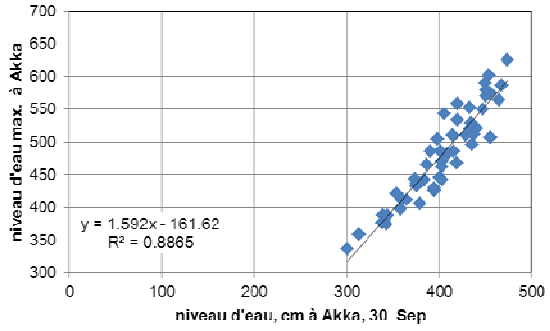
For illustration, three statistical linear regressions are presented in the next six following figures. We only show the statistical linear regressions of OPIDIN for the flood peak prediction at one issuing date forecast (the 30th of September). More issuings date and the linear for the timing of deflooding are presented in detail in Zwarts, 2009 and Zwarts, 2010 from which the given figures were extracted. From these figure, we can observe a wider scattering around the regression lines for the prediction of the flood peak timing compared to the flood peak water level.

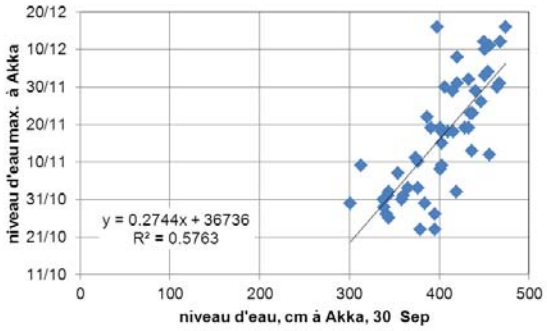
**Statistical linear regression for the Annual Flood Peak Water level and Timing in OPIDIN (Issue: 30th of September)**



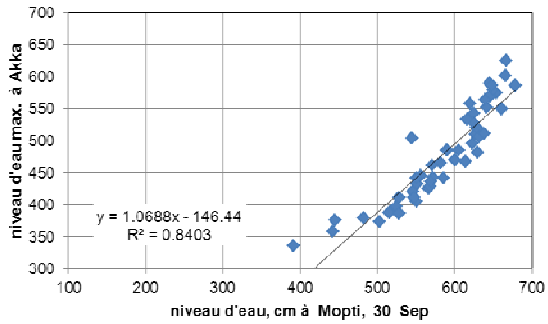



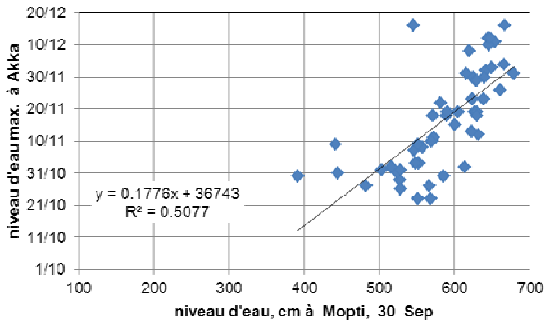

**Figures 4-9 and 4-10: At Mopti from Mopti**




**Figures 4-11 and 4-12: At Akka from Akka**




**Figures 4-13 and 4-14: At Akka from Mopti**

After presenting the statistical method, the advantage of the OPIDIN approach becomes more evident to enunce:

- The prediction can be summarized within comprehensive and user friendly tables.
- The prediction only needs one measure of water level in a given station at specific dates which facilitate the logistics of data collection to prepare prediction.
- The statistical method enables to broad the lead time of the prediction but longer the scope of the prediction period is, more the uncertainty will increase.

However, the use of this simplified statistical approach involved also different assumptions and constraints which can be source of uncertainties in the final forecast.

- OPIDIN prediction is based on a single water level measure in one station; the prediction is then subject to uncertainty from local disturbances or local punctual peaks in the water levels.
- The statistical method in OPIDIN is based on the correlation of one water level measurement upstream at a single date to a specific occurrence of a given water height in the downstream time series. Therefore, the prediction is based only on tiny numbers of correlations equal to the number of years in the upstream and downstream time series. Moreover, the transformations (by linearization or other technics) of the statistical correlation partly truncate the correlation.
- As the statistical correlation do not account for time of concentration according to water levels, the prediction is also subject to large uncertainty concerning the variability of the flood onset and the inner flood cycle.
- The robustness of the statistical prediction of the initial version of OPIDIN is altered when the patternof the hydrological regime changed due for example to upstream water management like the implementation of Sélingué dam or the increase of water uptake from the Office du Niger
- Finally, extreme low water or high flood annual events that do not fit to the records might be not well predicted as it is then beyond the scope of the statistical correlations

## FROM FORECAST TO EARLY WARNING SYSTEM

After the prediction being produced, the main problem still remains to convert technical information, such as “peak flood will be 511 cm (confidence limit: +/-23 cm) at the scale of Akka” into really meaningful and practical information for people. So far, the solution adopted in OPIDIN is to categorize the flooding prediction into a five-point early warning scale (very high, high, normal, low and very low) and use systematically five classes in the transfer of information. To define the range of the five classes, Zwarts (2010) suggested ordering the annual peak water level of a given station into a cumulative frequency distribution. Afterwards, the years were split up into five representative classes having more or less the similar frequencies. Following this method, the table below presents the limits between the early warning classes for Akka and Mopti station:

**Table 4-1 Early warning classes used to test the initial version of OPIDIN**

| Annual flood peak water level | Mopti | | Akka | | Annual flood peak water timing | Mopti | | Akka | |
|---|---|---|---|---|---|---|---|---|---|
| Classes | Range | Freq. | Range | Freq. | Classes | Range | Freq. | Range | Freq. |
| Really low (80´s to 00´s) | 440-550 cm. | 10 | 330-410 cm. | 10 | Really early (70´s to 00´s) | before 9 Oct. | 11 | in Oct. | 10 |
| Low (70´s to 00´s) | 551-590 cm. | 9 | 411-450 cm. | 11 | Early (70´s to 00´s) | 9 to 16 Oct. | 10 | 1 to 9 Nov. | 10 |
| Normal (70´s to 00´s) | 591-640 cm. | 11 | 451-500 cm. | 9 | Normal (70´s to 00´s) | 17 to 24 Oct. | 11 | 10 to 19 Nov. | 12 |
| High (80´s) | 641-680 cm. | 10 | 501-550 cm. | 10 | Late (80´s) | 24 to 31 Oct. | 8 | 20 Nov. to 30 Dec. | 8 |
| Really high (50´s to 60´s) | 681-730 cm. | 12 | 551-625 cm. | 12 | Really late (50´s to 60´s) | in Nov. | 12 | in Dec. | 12 |

Some limitations of the early warning classes become evident when we study the referring periods. Indeed, the class “Really high” and “Really low” occurred only during the late fifties and sixties. Also, none of the years in the eighties belong to the class “High” and “Late”. The fifties and the sixties are never ranged in “Really Low”, “Really Late”, “Low”, “Late” classes. As a consequence, the information might Also, it should be noted that the classification depends on the span of period over which the frequency distribution has been calculated. Fortunately, surveys launched in OPIDIN stakeholder workshops and among the population of the iND show that people have a „collective memory“ regarding to the flood levels in the past and, for instance, know very well that there were high flood in 1994 and 1999 and a very low one in 1984 (Zwarts, 2010). This will then facilitate the transfer of information from classes to relevant and historical information, e.g. “the flood level this year will be Low, - about as high as in 2002, 2004 and 2005”. Each year of one class will be communicated to give significant elements of comparison. This information will be used by the local radio and key relevant local stakeholders to inform the people in this time of the year about the expected flood level in an non-technical way, e.g. “flood will be about as high as two year ago and thus not as low as last year”.

## CURRENT AND FUTURE WORK FOR OPIDIN TOOL IMPROVEMENT

The content of this section comes mainly from Zwarts, 2010 and from one shared document of work from the same author and Bakary Koné.

i. Statistical predictive skill improvement

- **Improve statistical regression** with the application of curvilinear statistical relationship (generally the polynomial curve (second order approach)) to better fit for extremes
- **Extend historical time series:** the initial version of the model is based on the time series 1956-2007 at Mopti and Akka. In the version under development, recent years are added to the time series to increase the reliability of the predictions. The DRH provide also longer time series for Mopti and Akka and more water level stations will be incorporated to the OPIDIN tool.
- **Extend locations for OPIDIN predictions:** the extension of the gauging station coverage in the improved version of OPIDIN is now based on the full data set since 1921 for Douna, Beleni-kegny, Sofara, Ke-Macina, Mopti, Akka, Niafunké, Goundam and Dire, so much more data (and much longer series) as in the initial version of OPIDIN.
- **Add confidence interval:** In the improved model, the uncertainty of the predictions is communicated through the creation of a confidence interval. The work in this report will then contribute to this aspect.
- **Extend OPIDIN prediction** to any issued forecasting date and to any deflooding water level threshold
- **Statistical bias of upstream water management:** study the impact in the statistical prediction from the alteration of flow regime due to the implementation of new river basin infrastructures (construction of dams and channel diversion for irrigated schemes). As the regression equations used in the initial version of OPIDIN are based on the hydrological data from 1956 onwards, the implicit assumption of the model is that the curve for the flooding and the deflooding varies according to the river discharge and to nothing else. However, the flood curve in the Inner Niger Delta is altered especially since 1982 with the construction of the Sélingué dam. In the improved model, the first solution considered is to recalculate solely for 1982 onwards. However, all years with a high flood are lost for the model which will then significantly reduce the statistical strength of the regression equations on which the model is built. The second solution currently under investigation is to apply a statistical bias at the stage of 1982 the creation of the Sélingué dam.

ii. New joint functionalities

- **Inundation maps:** on the basis of OPIDIN statistical prediction and of analysis of satellite images, an archive of inundation maps for different water levels was produced. Zwarts et al. 2005 converted satellite images of the Inner Niger Delta available for 24 different dates into maps showing the actual coverage by the flood at a certain moment. The water levels in Mopti, Akka and Diré have been measured on those days, so the relation between the actual flooding and the measured water level

could be ascertained. These water maps were combined to construct a composite water map which was done separately for incoming and receding water. For receding water it was even necessary to make a separate flooding model for years with a high and a low peak level since there were more temporary, isolated water bodies after the peak water level had been high. The disadvantage of the given composite water maps was that the intervals between the different water levels were unequal. To make a water map with equal intervals, the water line was interpolated at a water level of each additional 10 cm (i.e. 10, 20, 30 cm, etc.), As OPIDIN predicts the water level within the same flooding cycle, these predictions may be converted into water maps, using the just described digital flooding model. The digital flooding model was based on Landsat satellite images with a resolution of 30 x 30 m. This appeared to be accurate enough to describe, even at a very local level, the flooding and deflooding process. The digital flooding model was extensively used during the field work between 2003 and 2010. The water maps appeared to be surprisingly precise, given the fact that the selected satellite images (being used to construct the water maps) refer to the period 1984 – 2003. Apparently, there have been amply spatial changes in the flooding process during the last 20-30 years. The maps were combined with maps designed community groups representing their own spatial interpretation of their territory allowing users to understand and analyse the inundation map. This information permits to obtain reliable information on flooding pattern some months in advance and better manage the potential of the territory. During the “Partners for Water” project, the Dutch consortium has developed combined flooding maps (natural resource maps elaborated by stakeholders + satellite images)in collaboration with stakeholders of the IND (farmers, fishers and herders), national NGOs and Government institutions. The developed maps for Akka (Rural district of Deboye) and Kakagna (rural district of Dialloube) have helped villagers to understand and get interest about the importance of predicting of the flood on their farming systems. Extrapolate the prediction to the estimation of the flood extent with the conversion into digital water maps based on satellite imagery from LANDSAT during flooding and deflooding

#### 4.1.3 Forecast skill assessment method

**INITIAL CONDITIONS FOR THE FORECAST AND TIME PERIOD OF THE SIMULATIONS**

First, the skill assessment presented in the next subsection was based solely on the initial version of OPIDIN (version from 2010). Second, the initial conditions of each prediction are based on a single daily water level information read from the time series of Akka or Mopti gauge stations.

The general retrospective forecast skill, presented in the following sub-sections, summarizes 52 predictions applied each year from 1956 to 2007 using OPIDIN statistical method.
The table summarizes the annual time series used when assessing the retroactive forecast skill for dry and wet episods using classes of quantitative deviations. The scoring is then only based on ten retrospective predictions:

**Table 4-2 Annual time series used to assess the retroactive forecast skill for dry and wet episods**

| Annual time series used to assess skill for extreme episods | Mopti | | | | | | | | | | Akka | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lowest Annual Discharge | 1973 | 1983 | 1984 | 1985 | 1986 | 1987 | 1989 | 1990 | 1992 | 1993 | 1983 | 1984 | 1985 | 1986 | 1987 | 1989 | 1990 | 1991 | 1992 | 1993 |
| Highest Annual Discharge | 1956 | 1957 | 1958 | 1959 | 1960 | 1962 | 1964 | 1965 | 1967 | 1969 | 1956 | 1957 | 1958 | 1959 | 1960 | 1962 | 1963 | 1965 | 1967 | 1969 |
| Earliest Annual Flood peak timing | 1957 | 1958 | 1960 | 1961 | 1962 | 1963 | 1964 | 1966 | 1967 | 1976 | 1956 | 1957 | 1958 | 1959 | 1960 | 1962 | 1963 | 1964 | 1969 | 1976 |
| Latest Annual Flood peak timing | 1972 | 1973 | 1979 | 1982 | 1986 | 1989 | 1991 | 1992 | 1993 | 2002 | 1973 | 1980 | 1982 | 1983 | 1984 | 1986 | 1988 | 1991 | 1992 | 1993 |
| Lowest Annual flood peak water level | 1983 | 1984 | 1986 | 1987 | 1989 | 1990 | 1991 | 1992 | 1993 | 2002 | 1982 | 1983 | 1984 | 1986 | 1987 | 1989 | 1990 | 1991 | 1992 | 1993 |
| Highest Annual flood peak water level | 1956 | 1957 | 1958 | 1959 | 1960 | 1961 | 1962 | 1964 | 1966 | 1967 | 1957 | 1958 | 1959 | 1960 | 1962 | 1963 | 1964 | 1965 | 1967 | 1969 |

## INDICATORS

The figures used to present OPIDIN skill assessment enable to see the evolution of the hit score from the 52 retrospective OPIDIN predictions for seven different issuing dates and according to two different types of classes. These classes will be now further detailed.
The first classes used enable to assess the hit score according to different threshold of quantitative deviation. The quantitative deviation intervals were defined with Wetlands International in order to fit different needs from the local stakeholders in term of accuracy (a fisherman having less need to get an accurate prediction than a rice farmer for instance). The classes to assess the quantitative deviations of OPIDIN are expressed in metric for the peak water level and in time for peak flood timing. The following table summarizes the classes adopted for the rest of the report.

**Table 4-3 Classes used to assess the skill of OPIDIN in term of quantitative deviation**

| Classes Flood peak Water Level | Classes Flood peak Timing | Classes Deflood peak Timing |
|---|---|---|
| more than 1m. lower | more than 2 weeks earlier | more than 4 weeks earlier |
| from 0.5 to 1m. Lower | from 2 to 1 week earlier | from 4 to 2 weeks earlier |
| from 0.25 to 0.5 m. Lower | from 1 week to 2 days earlier | from 2 weeks to 4 days earlier |
| .+/- 0.25 m. | .+/- 2 days | .+/- 4 days |
| from 0.25 to 0.5 m. higher | from 2 days to 1 week later | from 4 days to 2 weeks later |
| from 0.5 to 1m. Higher | from 1 to 2 weeks later | from 2 to 4 weeks later |
| more than 1 m. higher | more than 2 weeks later | more than 4 weeks later |

The classes presented in the table were also defined also according to the standard deviation of the annual flood peak water level and timing as well as the deflood timing (200cm.) as it is summarized in the following table:

**Table 4-4 Standard deviation of Mopti and Akka time series (1956-2007) in days for the peak flood and deflood (200cm.) timing and in cm. for the peak flood water level**

| Standard deviation 1956-2007 time series | Mopti | Akka |
|---|---|---|
| Peak Flood Water level in cm. | 69 | 72 |
| Peak flood timing in days | 12 | 9 |
| Deflood timing (200cm.) in days | | 35 |

The second type of classes was the object of detailed description in a previous sub-section as it is the current classes used in the operational mode of the early warning system of OPIDIN to communicate the flood peak. All the classes and their intervals are presented in Table 4-1. In this frame, the OPIDIN prediction ranged at the lowest classes (“Really Low”. And “Really Late”) were the object of a standalone skill assessment in order to characterize the predictive skill for dry episode.

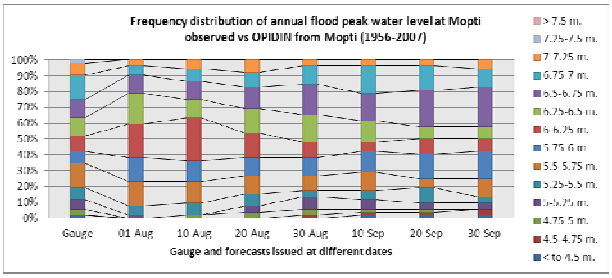
## 4.2 RESULTS

### 4.2.1 Retroactive forecast skill

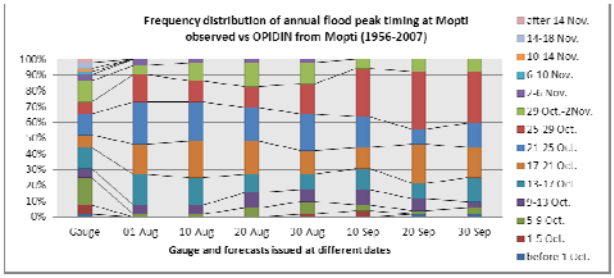
**ANNUAL FLOOD PEAK**

**Frequency distribution**

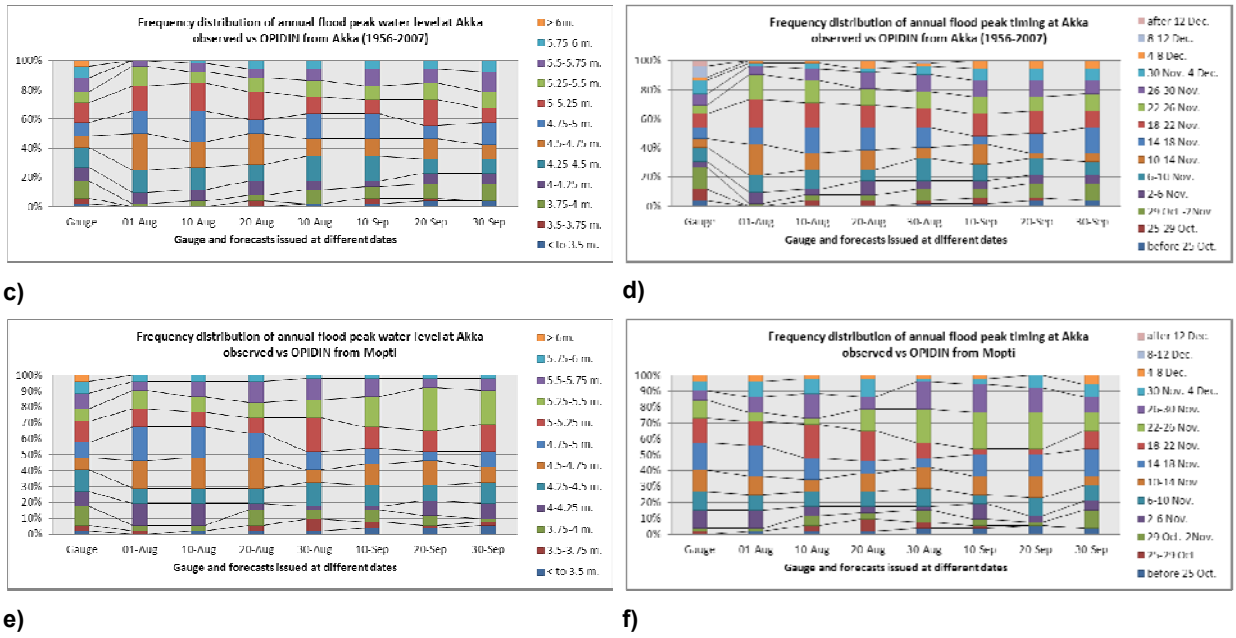
The six panels in the figure below compare the frequency distribution expressed in percentage of the annual flood peak water level (left column) and of the annual flood peak timing (right column) from 1956 to 2007 observed at Mopti (first row) and Akka (second and third row) and predicted with OPIDIN based on Akka (second row) or Mopti (first and third rows) time series at seven different issuing dates. The first bar of each chart represents the observed frequency distribution and the lecture of the following bars shows the evolution of the predicted distribution further we decreased the lead time of the forecast (from 1th of August to 30$^{th}$ of September). Each color represents a specific interval of peak flood water level (left column) and peak flood timing (right column).In all the figures, the observed frequency distribution are better distributed among all classes compared to OPIDIN prediction. Even if the frequency of the peak water level and timing in OPIDIN prediction tends to capture higher and lower classes when decreasing the lead time, the frequency distribution of really extreme low and high classes are generally underrepresented. This improvement when decreasing lead time is also really more significant for peak water level prediction than peak timing. The prediction of Akka from Akka presents a better frequency distribution than Akka from Mopti. As well, the frequency distribution of the prediction of Mopti from Mopti remains more coagulated into the centered classes than the ones performed in Akka.




a)




b)




c) d)

e) f)

**Figure 4-15 Frequency distribution expressed in percentage of the annual flood peak water level (left column) and of the annual flood peak timing (right column)**

## General retrospective forecast skill using quantitative deviation errors

The six panels in the figure below compare the quantitative deviation errors frequencies expressed in percentage between the retrospective OPIDIN predictions of the annual flood peak water level (left column) and of the annual flood peak timing (right column) at seven different issuing dates (each bar representing a issuing date) against observed time series at Mopti (first row) and Akka (second and third rows) from 1956 to 2007. OPIDIN prediction is based on Mopti time series for the first and third rows and on Akka for the second rows. Each colour represents a specific interval error expressed In meter for the peak flood water level (left column) and expressed in days and weeks for the peak flood timing (right column). For the prediction of the peak water level (left column), OPIDIN increased largely its prediction skills when decreasing lead time. The peak water level prediction of Mopti demonstrates better skill than the predictions of Akka. Also, the skill of OPIDIN at Akka from Akka is better when issuing dates of the prediction are in September. In August, OPIDIN at Akka from Mopti shows better score. The probability to get an error superior to 1 meter in the OPIDIN predictions gets really low when the issuing date is later than the10th of September. On the right column, we can now consider the skill of OPIDIN to deliver appropriate peak timing. The results show a very low probability to match the timing for plus or minus 2 days. Moreover, the skills of OPIDIN do not present any sensible improvement when the lead time is decreased. Although OPIDIN predictions at the initial version show skill to predict the range of peak water level at Mopti and Akka, the flood timing seems to be hardly captured with the statistical method.

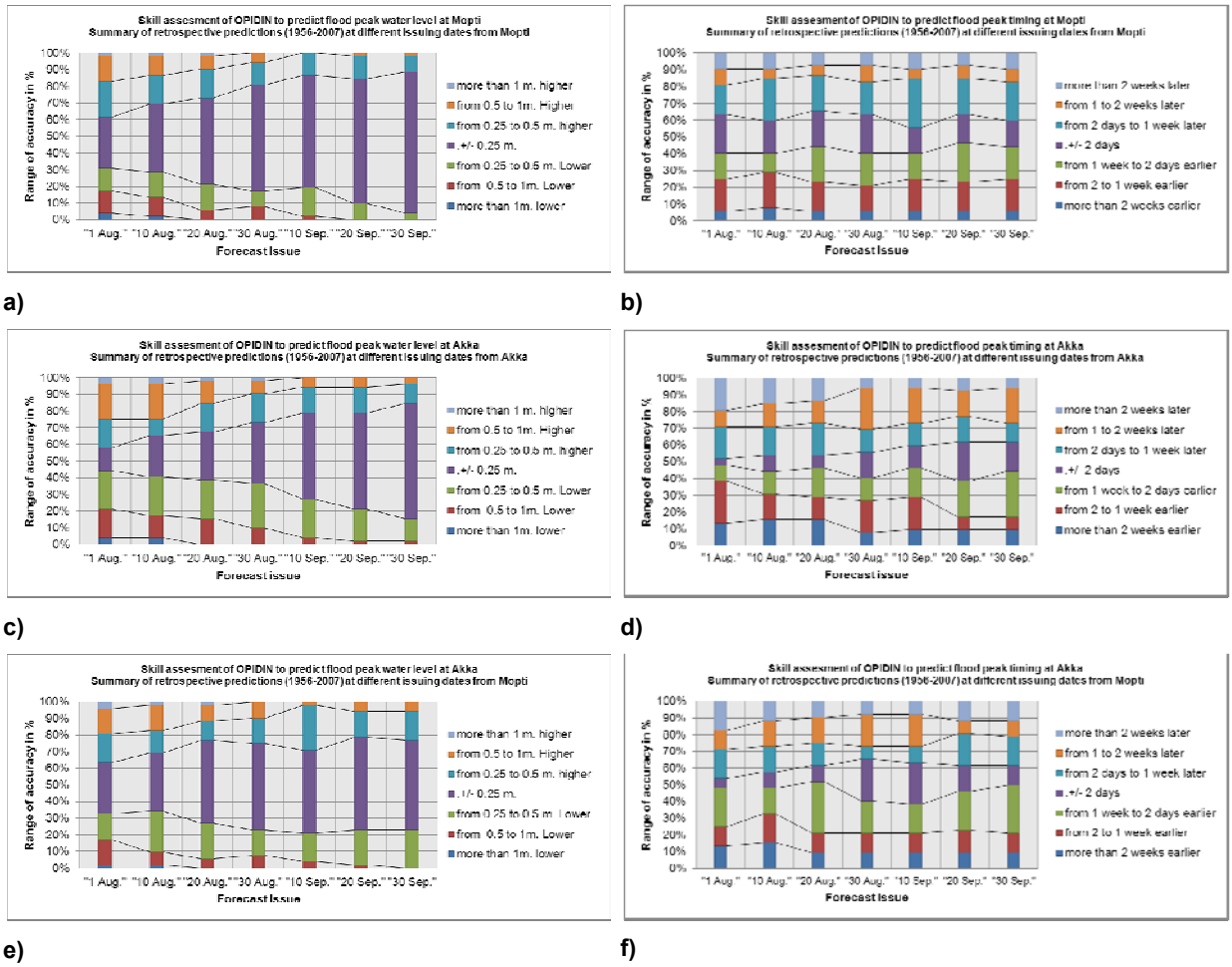



**Figure 4-16 Quantitative deviation errors frequencies expressed in percentage between the retrospective OPIDIN predictions of the annual flood peak water level (left column) and of the annual flood peak timing (right column)**

## Retroactive forecast skill for dry and wet episods using quantitative deviation errors (Mopti)

Same as previous paragraph, the six panels in the figure below compare the quantitative deviation errors frequencies expressed in percentage between the retrospective OPIDIN predictions of the annual flood peak water level at seven different issuing dates (barplot) against observed time series at Mopti from 1956 to 2007. OPIDIN prediction is based this time only on Mopti time series (for the predictions of Akka based on Akka and Mopti, please refer to the annex). In order this time to capture the skill of OPIDIN to predict extremes (dry or wet episods), the plots present then similar results but only applied to the 10 years of the time series with the lowest and highest annual discharge (first row), to the 10 years of the time series with the earliest and latest annual flood peak timing (second row) and to the 10 years of the time series with the lowest and highest annual flood peak water level. If we compare the plots from row to row, the results look relatively similar. The deviation error remains slightly higher when applying early and late flood peak timing. As early flood peak timing is highly correlated to low annual discharge and low peak water level, the results then remain similar because the 10 years selected were finally almost similar for the three

methods. Same remark can be addressed for extreme wet periods. OPIDIN increases largely it prediction skill for dry and wet episods when decreasing lead time. For dry episods, the probability to get an error superior to 0.5 meter in the OPIDIN predictions gets low (<10%) when the issuing date is in September and using any of the three methods. However, the probability to get an error superior to 0.25 meter in the OPIDIN predictions still remains high. For prediction of peak water level during wet episods, OPIDIN prediction shows slightly more discrepancies in average.

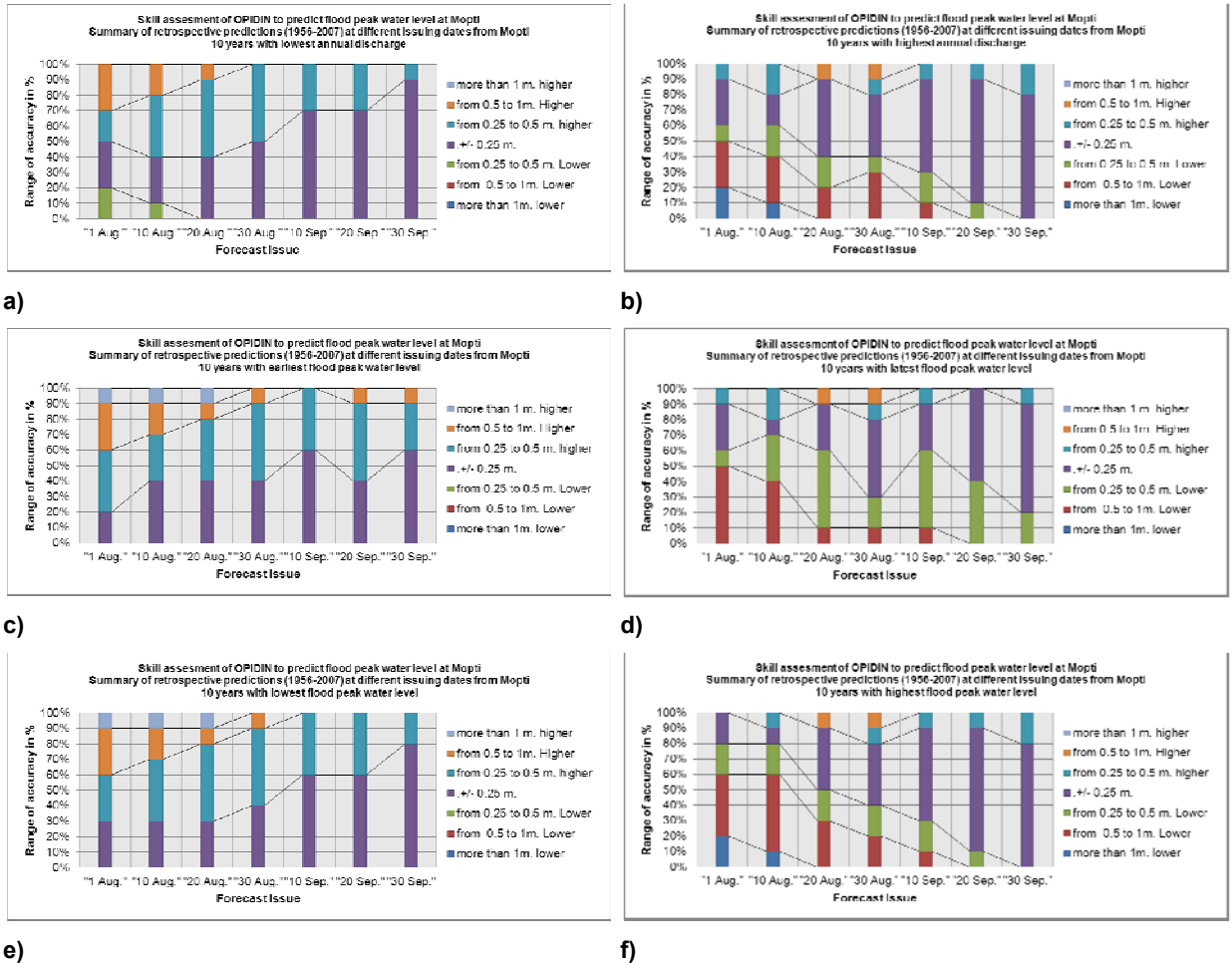



**Figure 4-17 Quantitative deviation errors frequencies expressed in percentage**

## General retroactive forecast skill using early warning classes

The six panels in the figure below presents the scores between the retrospective OPIDIN predictions of the annual flood peak water level (right column) and of the annual flood peak timing (left column) at seven different issuing dates (each bar representing a issuing date) against observed time series at Mopti (first row) and Akka (second and third rows) from 1956 to 2007 using the classes of the operating early warning forecasting system communication protocol. OPIDIN prediction is based on Mopti time series for the first and third rows and on Akka for the second rows. Each colour represents the match or the deviation of the prediction against the observation according to the early warning classes. For the prediction of the peak water level (right column), OPIDIN increased largely its prediction skills when decreasing lead time. For example, OPIDIN predictions at Mopti from Mopti match the good

early warning class at around 40 % the 1st of August and increased gradually the score to 75 % the 30th of September. Also, the probability to stand two classes lower or higher is low in August and really low in September. As already stated when presenting the skill with the use of quantitative deviation classes, the OPIDIN prediction for the flood peak timing do not improve when the lead time decreases. The probability to get an error of 1 or 2 classes lower and higher remain at around 70% for any issuing date.

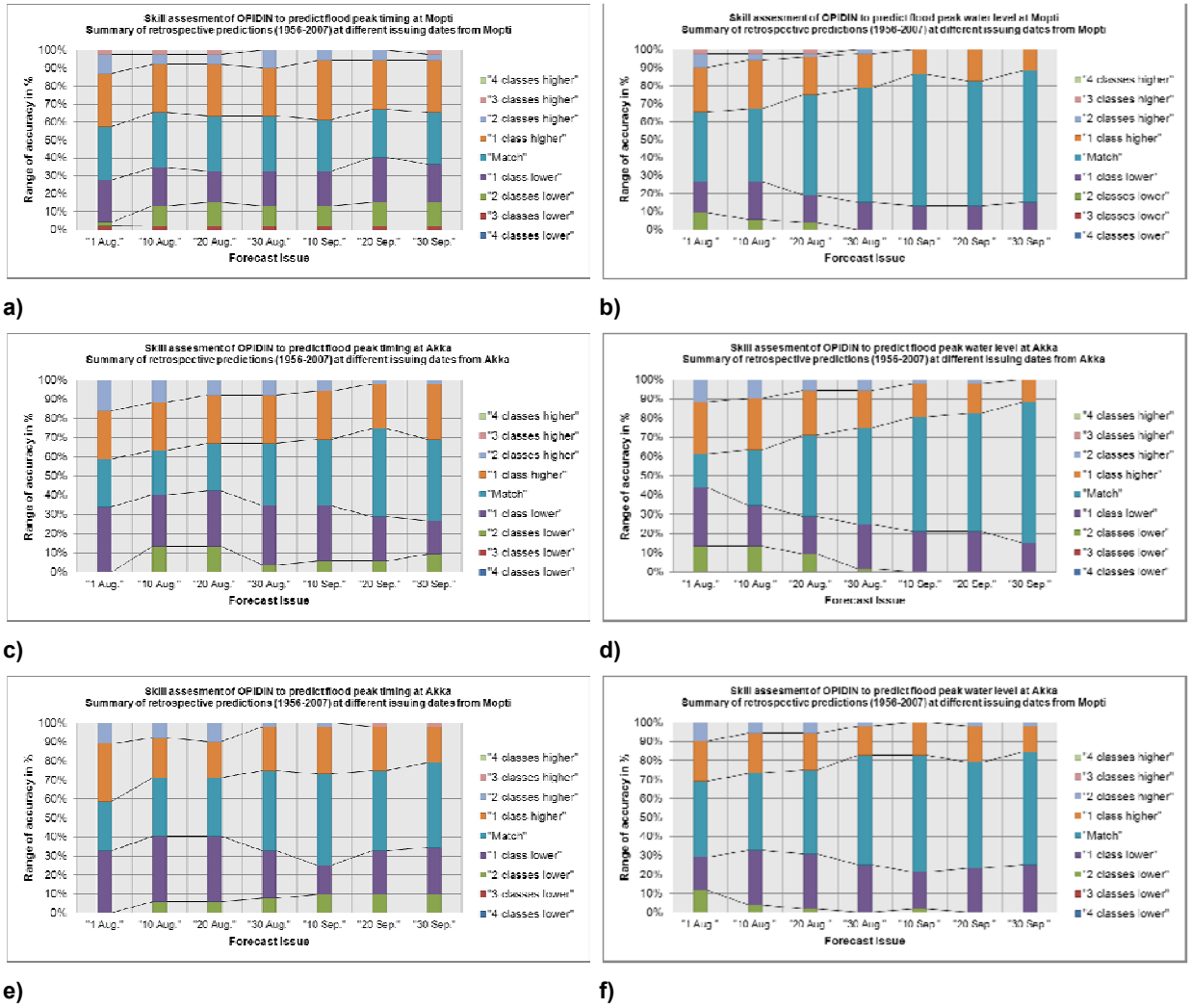
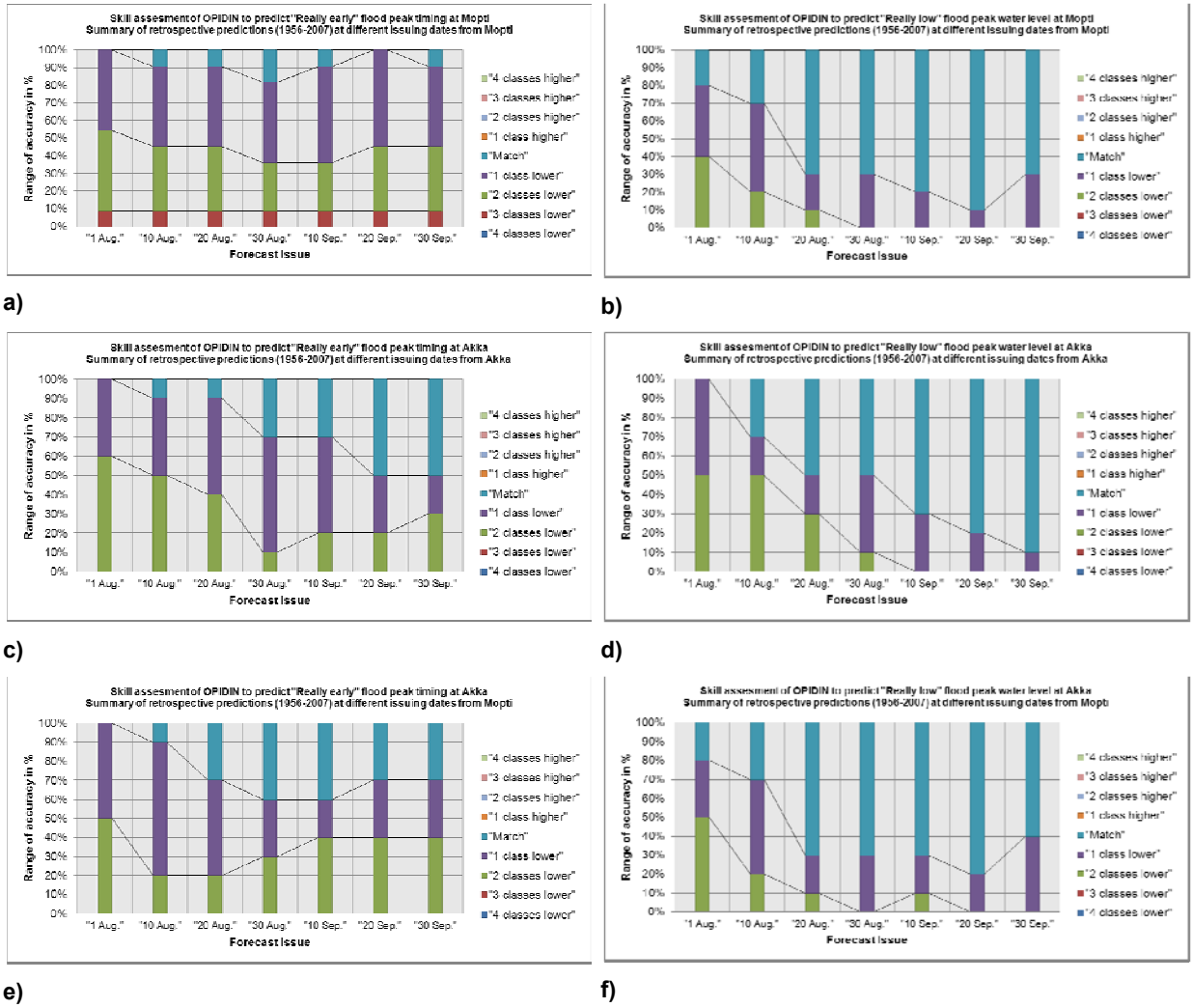



**Figure 4-18 Scores between the retrospective OPIDIN predictions of the annual flood peak water level (right column) and of the annual flood peak timing (left column)**

## Retroactive forecast skill for dry episods using early warning classes

The six panels in the figure below present similar results than previous paragraph focusing this time on the skill of OPIDIN when predicting peak water level in the range “Really Low” and peak timing in the range “Really Early” (Table 4-1). When OPIDIN is predicting a “Really Early” flood peak at different issuing date, only a really low percentage of the retrospective predictions are at the end correct especially at Mopti. However, prediction at Akka shows sensibly better results. Also the prediction at Akka does not deviate more than 2 classes. For the peak water level prediction (right colum), the OPIDIN skill improves when the issuing dates is closer to the events. The deviation is also not more than 2 classes. As seen in the

section analysing the overall frequency distribution. OPIDIN demonstrates less skill to predict extremes and tends then to overestimate (for dry episods) or underestimate (wet periods) flood peak water level but the closest the issuing date of the forecast is, better the skill to capture extreme events is.




a) b)

c) d)

e) f)

**Figure 4-19 Skill of OPIDIN when predicting peak water level in the range "Really Low" and peak timing in the range "Really Early"**

## Additional analysis

For more detailed year to year analysis about the skill assessment of OPIDIN prediction (using either quantitative deviation or early warning classes) for flood peak water level and timing at/from Akka/Mopti, please refer to the tables presented in the Appendix 7.1.

## DEFLOOD TIMING (200M.)

The same skill assessment was performed for the prediction of the deflood timing. Please may refer to the tables and figures presented in the Appendix using the same way to read and interpret the tables and figures as we proceeded for the annual flood peak in the section above.

**Summary**

Figure 4-20 summarizes the skill of the initial version of OPIDIN using R² calculation. The linear regressions at/from Akka and Mopti for the flood peak water level and timing are grouped into a single plot. In this figure, we can observe a steady improvement of the R² for the flood peak water level when lead time decreases. For the flood peak timing, the R² score tends to decline for the last three issuing date except for “at Akka from Akka”. The best R² for the flood peak water level is reached for Mopti based on Mopti but the flood peak timing prediction is the weakest (less than 0.4). The best R² for the flood peak timing do not reach 0.6 (Mopti from Mopti). The figure shows that OPIDIN better captures the flood peak water level than the flood peak timing.

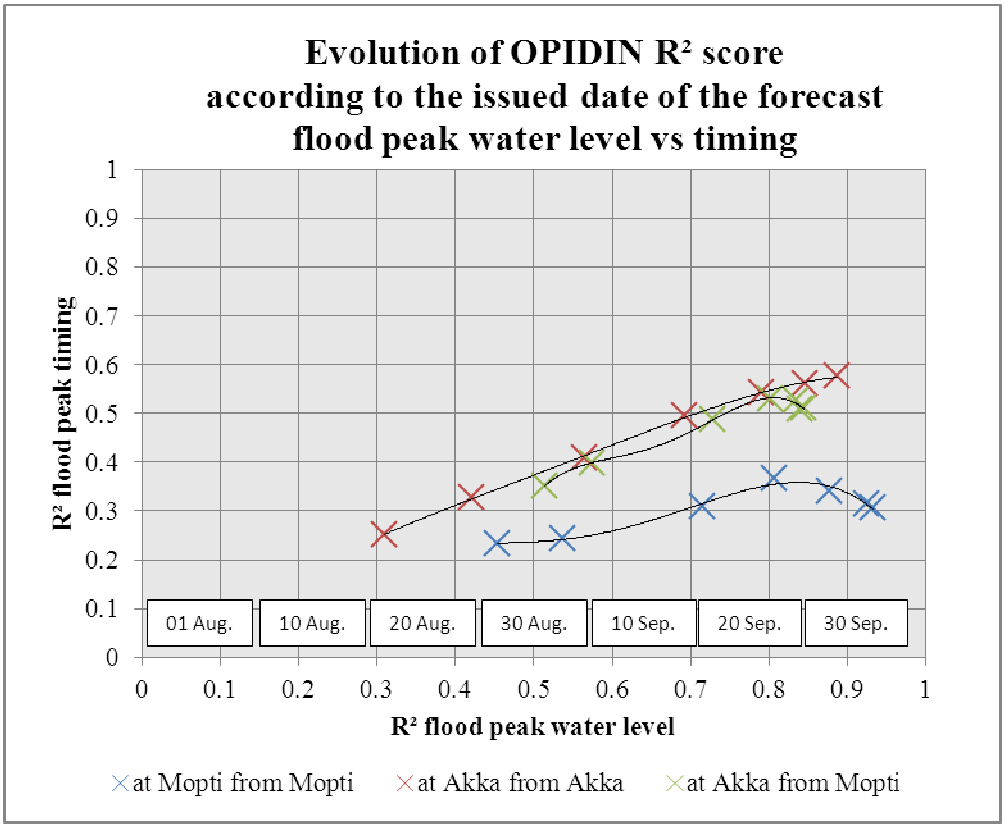



**Figure 4-20 skill of the initial version of OPIDIN using R² calculation**

# 5. FORECAST OF HYDROLOGICAL DROUGHTS FOR THE NILE CASE STUDY BASIN

## 5.1 MODELS AND SIMULATIONS SET-UP

The Nile Forecast System (NFS) is a near real-time distributed hydro-meteorological forecast system designed for forecasting Nile flows at designated key points along the Nile; of major interest is the inflow of the Nile into the High Aswan Dam, Egypt. The core of the NFS is a conceptual distributed hydrological model of the Nile system including soil moisture accounting, hillslope and river routing, lakes, wetlands, and man-made reservoirs within the basin. The main inputs to this model are the rainfall and potential evapotranspiration. The current version of the NFS is 6.08, which is still under testing. The latest stable release of the system is 5.1.

From the outset of NFS development, it was decided to utilize satellite remote sensing technology in estimating rainfall over the basin. This was motivated by the scarcity and discontinuity of rainfall records within the basin in addition to the lack of direct monitoring control over rain gauges, as all basin rainfall occurs outside the borders of Egypt where the system is hosted. Therefore, the system grid was designed to match that of the satellite rainfall. In addition, the system includes a large database of rain gauge data, as part of the Nile Basin Hydrometeorlogical Information System (NBHIS) which also holds flow records at all key river gauges.

Geostationary satellite imagery is received using a dish antenna on the roof of the ministry building. This imagery is used to generate daily high resolution (20km) gridded estimates of rainfall over the basin, that is then blended with gauge-only rainfall analyses. The combined daily rainfall product is used to drive a gridded distributed hydrological model of the basin.

When used for forecasting as intended, a short NFS simulation (a few weeks) is performed using observed rainfall (merged satellite and gauge estimates) to define the model status (soil moisture storage, reach storage) on the current date. Subsequently, an ensemble of historical rainfall (for as many years as available) for the 3 months following the current date is applied to the model to simulate possible inflow series to Lake Nasser, called Extended (and more recently Ensemble) Streamflow Predictions (ESP). Once a week observed flows at some key points (e.g. Diem) are assimilated to update the model states. It implies that the rainfall estimates are adjusted for the last 4 weeks to minimise the difference between the simulated and the observed flows. When input data are missing or felt unreliable the deterministic model can be replaced by a first order Markov model for three locations (Malakal, Roseires and Khartoum). This model produces possible monthly runoff traces, which are subsequently disaggregated to daily data. The stochastic model is often applied to replace the deterministic forecast for the White Nile at Malakal, as the quality of the latter is considered less reliable. The system relies on a GIS database to represent the connectivity

of the different pixels as well the different streams, rivers, and sub-basins associated with the designated forecast points. In summary the NFS (Figure 5-1) is composed of 6 main components that perform the following functions:

1. Rainfall Estimation
2. Hydrological Simulation
3. River Flow Forecasting
4. Assimilation
5. Data Collection and Management (NBHIS)
6. GIS Functions

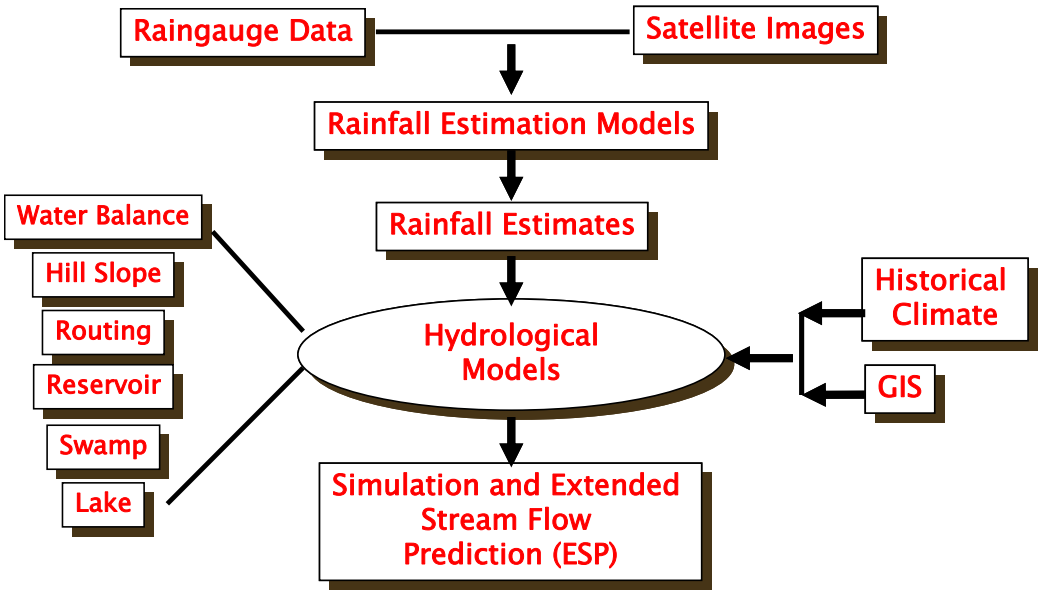



**Figure 5-1 Schematic of the Nile Forecast System**

As mentioned above, the NFS requires rainfall and potential evapotranspiration data for the whole Nile basin as inputs. These are required as gridded rainfall and potential evapotranspiration data (i.e. maps) with a daily time step although monthly data can be also used as the system contains routines for disaggregating these to the daily time step. NFS also requires discharge data at key stations for assimilation, calibration, and performance evaluation. Discharge is usually obtained from stage measurements using rating curves that are updated annually based on concurrent stage and discharge measurements.

Since its development started in 1992, the NFS has been enhanced several times. The latest enhancement of the NFS was undertaken in 2009 by the Hull University under the remit of the Nile Basin Initiative Flood Preparedness and Early Warning (FPEW) project. The enhancement focused, amongst other things, on the rainfall estimation methodology through developing a multi-spectral technique that takes advantage of the new generation of the European METEOSAT satellite (METEOSAT Second Generation – MSG). The NFC obtained a new reception system from VCS Engineering, Germany in 2007 to replace the old Primary Data User System (PDUS) unit that was in operation since 1992. The MSG

spacecraft provide considerably enhanced datasets over their predecessors, both in the spatial and temporal resolutions of available imagery and in the number of separate wavebands for which images are available. The following section describes the old and new rainfall estimation techniques while Section 5.1.2 gives more details about the hydrological simulation model.

### 5.1.1 Satellite Precipitation Estimation

The NFS 5.1 satellite rainfall-estimation system is based on the principle of cold cloud duration (CCD). CCD techniques are based on the assumption that high cloud tops (which show cold in infra-red imagery) are frequently associated with strongly precipitating cumulonimbus systems. Thus, there is a statistical relationship between the duration of a probably-precipitating cloud in a given time period and the amount of rainfall within that period. The rainfall estimation equation is:

$$\text{Rainfall} = \text{Number of Cloudy Hours} \times \text{Mean rainfall per cloudy hour} \quad (3)$$

A probably precipitating cloud is identified from thermal infrared (TIR) imagery using a brightness temperature threshold. TIR brightness temperatures correspond closely to cloud-top heights and clouds with high tops may thus be identified by locating image pixels with brightness temperatures less than a given threshold (set to 233°K in the NFS). As noted above, high tops are commonly associated with vertically extended cumulonimbus clouds that generate the most significant element of tropical precipitation. Unfortunately, high tops are also associated with non-precipitating cirrus clouds, whose presence introduces considerable uncertainty into the statistical relationship. It is also possible for low-topped stratiform clouds to be associated with significant rainfall. An additional cause of uncertainty is that not all precipitating clouds rain at the same rate. However, these uncertainties average out over large space and time scales. In terms of impact on hydrological model outputs, they are therefore more likely to affect flow timings rather than total flow volumes and these uncertainties will average out over larger catchments.

The MSG satellite provides five channels in the thermal infra-red bands. The availability of a ‘split IR window’ 10.8μm and 12.0μm is particularly useful since it facilitates screening for thin cirrus clouds which contaminate single-window infra-red based products. A number of multispectral algorithms have been developed (Bellerby et al. 2000, Capacci and Conway 2005, Turk and Miller 2005). These algorithms vary in the methodologies used to derive the functions combining the spectral elements, although all show a distinct advantage in using multi-spectral data. The choice of image data to incorporate into an enhanced NFS satellite precipitation estimation system involves a trade-off between the additional precipitation information provided by viewing the Nile Basin in a new wavelength against the substantially increased data archiving requirements presented by the use of each new MSG image band.

In addition, it must be noted that the precise utility of multispectral image data for rainfall estimation may only be determined through an extensive study containing many months, or preferably several years of data. Such an extensive run of archived data was not, of course, available at the time of enhancing NFS and as a result a pragmatic decision had to be made as to the additional channels to archive. These channels are thermal infra-red channels (9.7, 10.8, 12.0 and 13.4 μm). The evaluation of the rainfall estimation technique presented in Section 5.3.1 is a first attempt towards assessing the added value of the additional channels in the latest version of the NFS (6.08).

Many satellite rainfall algorithms rely on the statistical relationship between rainfall rates and cloud top brightness temperatures (Todd et al. 2001). In addition, some multi-spectral approaches examine the differences between pairs of spectral channels to identify certain types of precipitating clouds, and in particular excluding non-precipitating cirrus clouds (Inoue 1987). Combining and generalising the two approaches, it is reasonable to express the rainfall rate (RR) associated with a raining pixel location as a linear combination of pixel brightness temperatures from the available image bands:

$$RR = a + b\,T1 + c\,T2 + d\,T3 + \ldots \quad (4)$$

Since not all pixel locations are raining, this estimate of the positive rainfall rate must be combined with a function identifying raining areas. If a function $\delta$ is defined to be equal to one for rain and zero for rain-free areas then the complete estimate of rainfall rate (R) at a given image pixel location becomes:

$$R = \delta\,RR = \delta\,(a + b\,T1 + c\,T2 + d\,T3 + \ldots) \quad (5)$$

Since MSG satellite images are received every 15 minutes while both the desired rainfall product and the calibration data have a daily time step, the individual rainfall estimates for each image time must be aggregated to produce daily rainfall Rd:

$$\begin{aligned} Rd &= \sum \delta\,(a + b\,T1 + c\,T2 + d\,T3 + \ldots) \\ &= a\sum\delta + b\sum\delta\,T1 + c\sum\delta\,T2 + d\sum\delta\,T3 + \cdots \end{aligned} \quad (6)$$

where $\sum$ is the sum over the images in a given day. The function $\delta$ must also be estimated from the satellite data. The simplest approach for this (and the one most compatible with the original NFS (5.1) procedures) is to identify raining pixels as those for which a single thermal infra-red channel brightness temperature ($T_b$) is colder than a given threshold (TTHRESH):

$$\delta = \begin{cases} 1 & T_b < T_{THRESH} \\ 0 & T_b \geq T_{THRESH} \end{cases} \quad (7)$$

In this case the first term in the equation for daily rainfall, ($a \sum \delta$) becomes equal to the cold cloud duration (CCD).

### 5.1.2 Description of the regional hydrological model

The hydrological component of the NFS is defined on the quasi-rectangular grid of the METEOSAT. Each grid cell (pixel) imitates a small basin with generalized hillslopes and stream channels. Input to each grid cell is precipitation and potential evaporation/evapotranspiration. This input is applied to the water balance model of the grid cell. Based on the moisture deficit in the cell, the water balance model computes the actual evapotranspiration and the surface and subsurface runoff components from the pixel. Surface and subsurface runoffs are subsequently input to the pixel's hillslope routing model, simulating the transfer of water towards the main channel. Then water is routed through this channel to the downstream pixel according to a pre-defined connectivity sequence (established via GIS). Figure 5-2 shows a flowchart of the simulation process in a grid cell. For computational efficiency, the current implementation of the NFS groups 4 × 4 pixels (≈ 484 km$^2$) as one computational unit. All three components are computed for each unit, and through a predefined channel network, the water is routed through the various tributaries to the outlet points of the Nile basin (Georgakakos et al., 2001) taking into account the effect of direct evaporation from river reaches, abstractions for irrigation, and man-made reservoirs. Special sub-models for lakes and wetlands are called for those pixels designated to fall into these water bodies. Then the accumulated runoff into a lake or a wetland from the surrounding basin and the lake/wetland area is treated using a lumped model. The lake model is applied to Lakes Victoria, Kyoga, and Albert while the swamp model is applied to the Sudd and Machar swamps in southern Sudan.

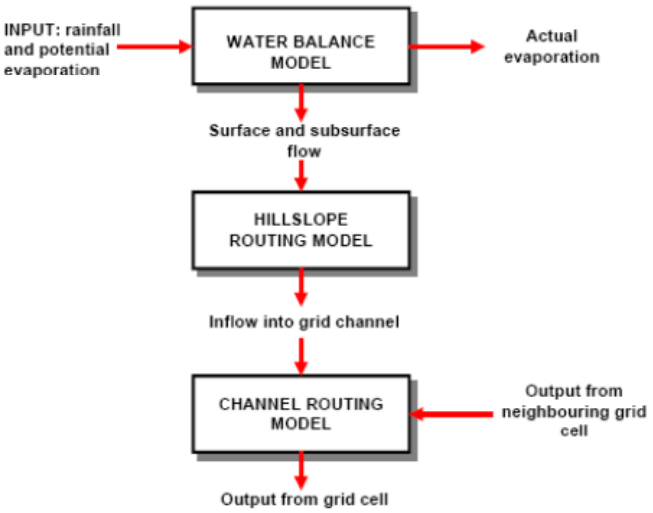



**Figure 5-2 Hydrological Models at the Pixel Scale**

### 5.1.3 Meteorological forecasts used

The ECMWF seasonal forecast daily data start from May till November from 2005 till 2012.

### 5.1.4 Initial conditions for the forecasts

At every grid point, three models are executed:

1) Water balance model.
2) Hillslope model.
3) Channel routing model.

Each model has driving input, parameters, state variables, model equations, and output.

Parameters are constants. State variables are dynamic (changing) variables that represent the “state” of the system and which make the model continuous.
Each model contains nonlinear model equations to compute the outflows and uses a form of the storage equation to compute new states at the end of each time step. The use of state variables and storage equation ensures that the volume is preserved. Each model has the following generic logic:

***Outflow = function of (Inflow, Parameters, Prior State)***

***New States = function(Prior States, Inflows, Outflows, Parameters)***

The water balance models the rainfall-runoff process in the soil. It has two layers which represent a shallow, fast-flowing upper (surface) soil zone and a slower lower (groundwater) soil zone. The outflow is the total runoff (in mm) which is a combination of the surface and groundwater runoff. The water balance model has five parameters, which are conceptual and represent the maximum soil moisture capacities and percolation rates. There are two state variables, which represent the soil moisture deficit in each layer. The deficit is the difference between the soil moisture capacity and actual soil moisture content. Thus, a deficit of zero is completely wet.
Outflow from the upper layer is proportional between coming inflow and reducing the actual deficit of this layer. The primary inflow is rainfall (in mm) which decreases the soil moisture deficit. A secondary input is the potential ET (or PET). The PET is used to calculate the actual ET which is an output layer or can be thought of as negative inflow sink since it increases the deficit. The actual ET is a function of the potential ET and the current soil moisture.
The hillslope model serves to rout the runoff from the water balance runoff over land to the channel, effectively producing a time lag and some attenuation for the local runoff within a pixel. The hillslope model consists of a single, non-linear storage equation. It has three parameters which are the slope, roughness, and length. In nature, each pixel is a similar to a small watershed and may contain several small hillslopes and channels, even though the model assumes each pixel has a simple hillslope leading to one main channel. The parameters represent some average hillslope and channel since there may be several actual hillslopes in any one pixel depending on the pixel size and topography. There is one state variable, which is the depth of water (in mm) across the hillslope. The inflow is the total runoff from the water balance model. The outflow is the flow rate (in mm/day) into the channel. The channel routing model routes water from pixel to pixel and cause both a time lag and attenuation of the flow as it moves downstream.
The channel model consists of a single, non-linear storage equation. It has six parameters which are the slope, roughness, length, pixel area, and two channel shape variables. The pixel area is used to transform the hillslope water into cubic meters from square millimeters.

The METEOSAT grid is not constant so that the area of each pixel varies across the grid, such that pixels furthest away from the satellite are distorted the most and have the largest area. There is one state variable which represents the amount (cross section area) of water in the channel. The inflow is combined flow of the hillslope model and any upstream channel routing. The outflow is the streamflow (in cms) that gets routed downstream.

In the channel model, water is “dumped” from pixel to pixel within the same time step. This does not mean that all of the water reaches the outlet in one day. Instead, each pixel will release a certain amount of water downstream based on its prior state and addition of the current inflows. Thus, a pixel with dry state will not release much water in the current time step even if there are inflows coming in. The amount of water released is a nonlinear, numerical function of the state, the inflows, and the parameters. Because of the nonlinear nature of channels, dry pixels will release less water and at a slower rate, assuming all other variables (like parameters) are equal. The outflow rate is based on Mannings’ equation. Thus, if all other variables equal, water comes out more slowly as the roughness increases, the slope decreases, the length increases, or the channel shape gives a lower hydraulic radius.

The hillslope and channel routings have equations that require numerical solutions instead of (analytical solutions). Therefore, an iterative technique is used to achieve an optimal solution. The iterations will stop when a minimum error difference is achieved or the number of iterations exceeds some limit.

#### 5.1.5 Time period of the simulations

The simulation program can be run for any time period. At the start of each run, the program reads the parameter and state files which should be full resolution, gridded files. Parameter files are static and are not for any specific data. The state files are dynamic, and there is a different file for each day. The initial state file for any run is one day prior to the start of the run. So, there should be state variable files at the start of any run period selected. If states are missing, default values are assumed.

After loading, the program loops through each day and reads a gridded file for both the precipitation and PET on that day. Once the gridded input are read for that day, the program loops through every model unit in the order prescribed by the flow sequence and does the model computations. Internally, the program has a gridded (2-D) array for each parameter, state value, input, and output used in the model. At the end of each calculation, the array values for the state and output values are updated. The outflow for each model serves as the input for the next model (in the same model unit) or, in the case of the channel model, the outflow is added to the next downstream model unit.

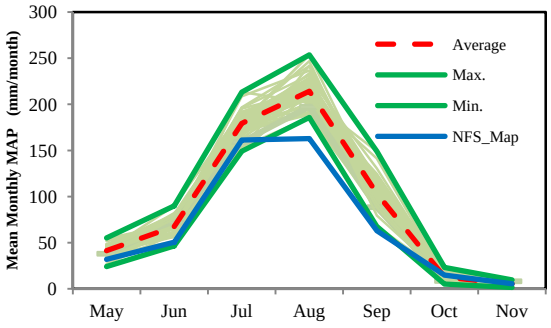
In this study, the simulation program run for each ensemble (0-50), daily time step from the first of May till the end of November for each year according to the rainfall forecasting time

series. The number of runs is one run for each year and 7 runs for each ensemble, so the total numbers of the runs is 357 runs. A script is written to loop over all ensembles and run the simulation program for each year.

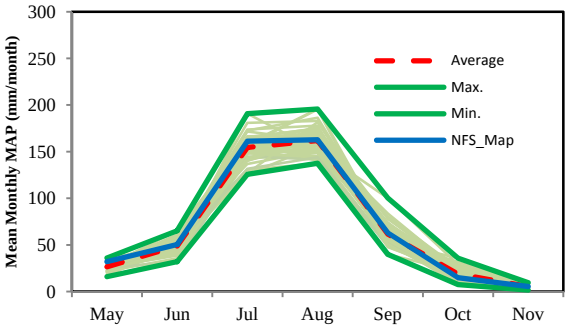
### 5.1.6 Bias correction methods of seasonal forecasts

The method to bias correct precipitation is based on monthly means, by applying a multiplicative correction factor. The correction factor was calculated by comparing the mean climate of the forecasts and an observed mean climatology of the real time rainfall data in NFS (for each calendar month).

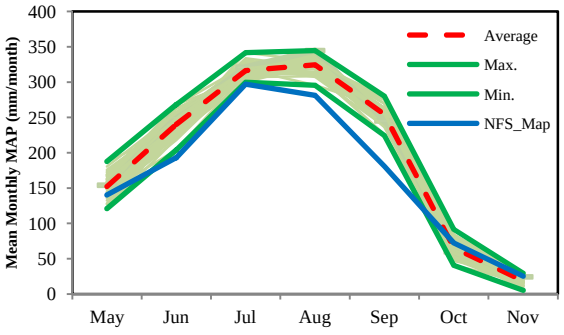
Figure 5-3 shows the change in the precipitation over the eastern Nile basins. The rainfall in the three basins (Atbara, Diem, and Khartoum) was over estimated which affect the simulation flow. This region (Eastern Nile basins) is very highly sensitive to the change of rainfall; however the small change in the rainfall in this area leads to more change in the flow. After using the bias correction Method, the average rainfall for the 51 ensembles became close to the real time rainfall estimation time series in NFC.



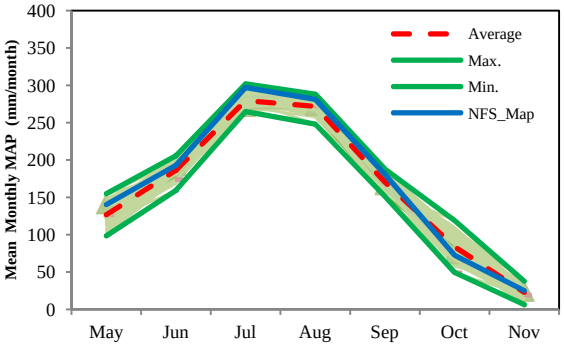

Original MAP over Atbara Basin



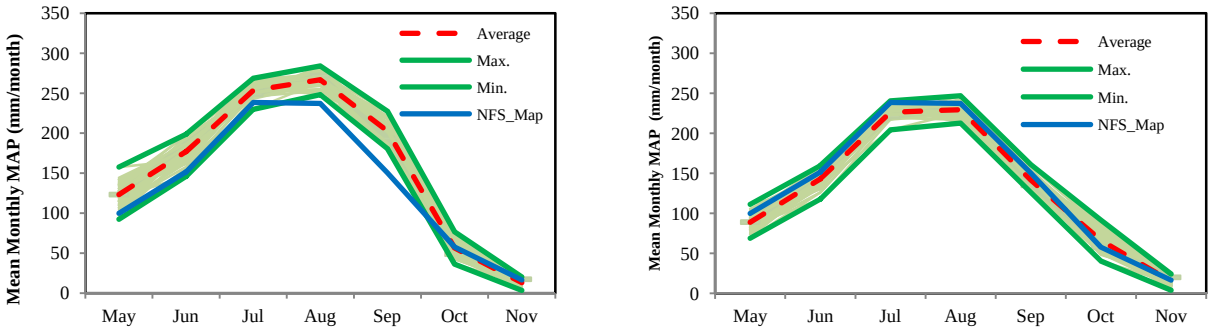

Corrected MAP over Atbara Basin




Original MAP over Diem Basin




Corrected MAP over Diem Basin




Original MAP over Khartoum Basin Corrected MAP over Khartoum Basin

**Figure 5-3 Differences between the Corrected precipitation over the Eastren Nile by the bias correction method and the Original forecasted Precipitation**

## 5.2 RESULTS

The results of this study are divided to two parts: part 1 using the rainfall forecast which corrected by the bias correction method with the PET which calculated from the outputs of the seasonal forecast for all ensembles. Part 2 using the same rainfall forecast with the Nile forecast System (NFS) PET. The PET in NFS are twelve grid files, each file expresses the Mean monthly PET for one month (Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec).

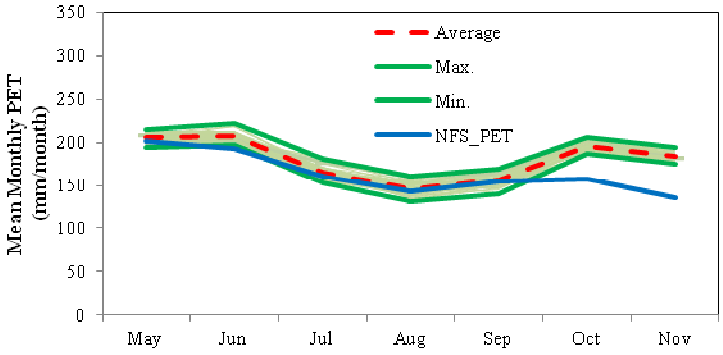
The range of the forecasted PET is close to the NFS_PET except October and November which are far from the NFS_PET (over estimate) this is clear in the Atbara basin, Figure 5-4 a), and the differences is decrease in the Khartoum basin, Figure 5-4 c), but in the Diem basin the NFS_PET located in the range of the 51 ensembles, Figure 5-4 b). It is expected that the differences in PET in October and November will affect simulation flow in these two months in Atbara and Khartoum sub basins.

Figure 5-5 shows the simulation flow range for all 51 ensembles comparing to the observed flow and NFS simulation. For each sub basin, there are two cases; case 1 using the forecasted PET and case 2 using NFS_PET. It is found that when using the NFS_PET the simulation flow range reduced; however the Upper limit and the lower limit for Atbara basin reduced around 1.5 BCM and 0.3 BCM respectively and reduce the rang width from 5 BCM to 4.5 BCM, for Diem the upper limit and lower limit reduced around 0.8 BCM and 0.4 BCM respectively and reduce the rang width from 5.25 BCM to 5 BCM, While in Khartoum basin the reduction in the rang width was 1.5 BCM, and upper limit and lower limit reduced around 2 BCM and 0.9 BCM respectively.

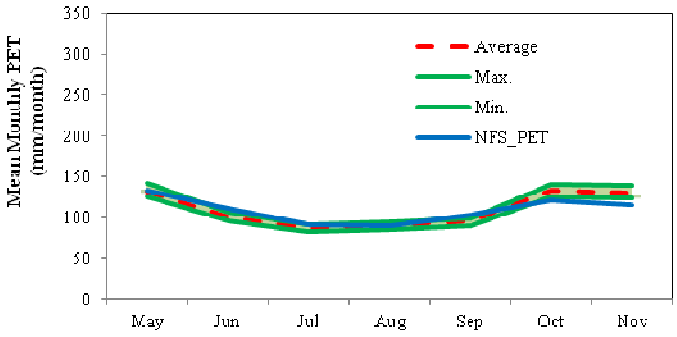
The results shows that, In case of using NFS_PET in Atbara basin the NFS simulation flow is close to the average simulated flow of the 51 ensembles and this is not occur in case of using the forecasted PET, the NFS flow become close to the lower limit of the range of the 51 ensembles. The results in the Blue Nile basins (Diem basin and Khartoum basin) shows some changes in the trend of the range of the 51 ensembles than NFS simulation flow in

both cases especially in October and November, while in the rest of months it is seen that the NFS simulation is closer to the average range of the 51 ensembles in case of using the forecasted PET than in case of using NFS_PET. These give indication that at the months (May, June, July, August, and September the forecasted weather elements which used in calculate the PET over the Blue Nile basin is successes in simulated the PET which is fixed in NFS.

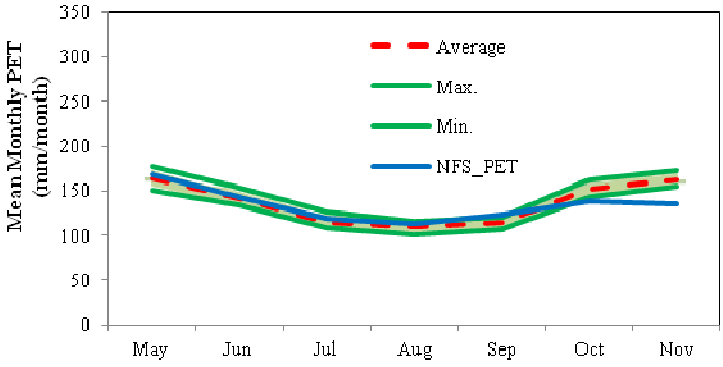
It was expected to get reduction in October and November flow simulation in both cases, because of the raise of PET in these months, but the rainfall which was slightly above the NFS flow simulation in October and November affects the results and raises the flow simulations range than the NFS flow.




a) Atbara Station

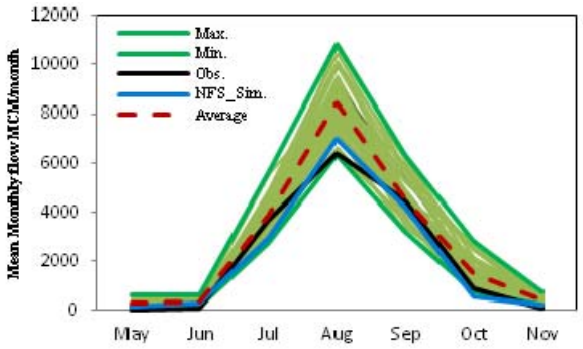



b) Diem Station

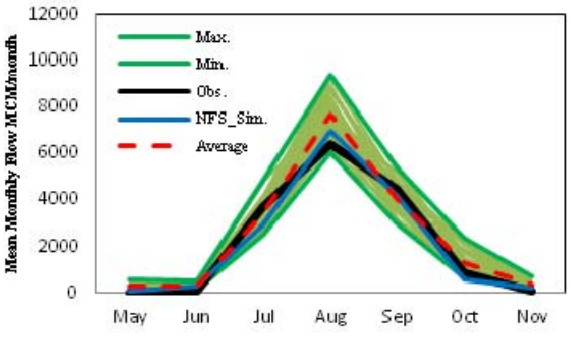



c) Khartoum Station

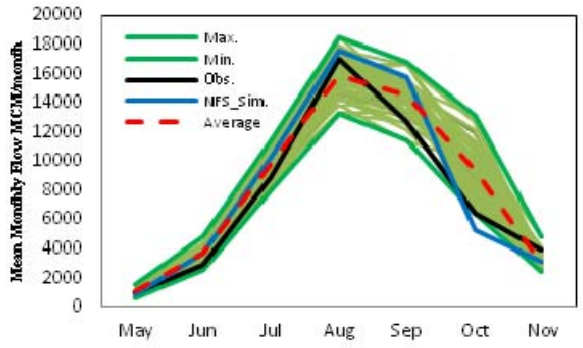
**Figure 5-4 PET calculated by the forecasted ensembles compared to the NFS fixed PET**



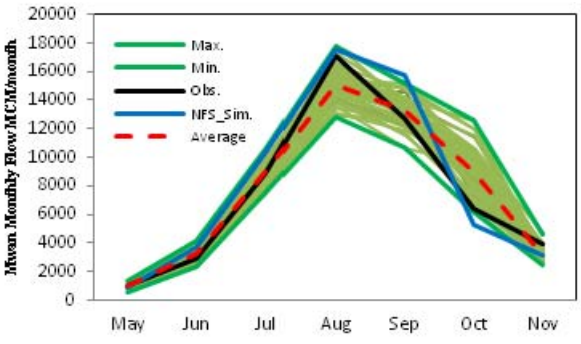

Simulation Flow at Atbara using forecasted PET



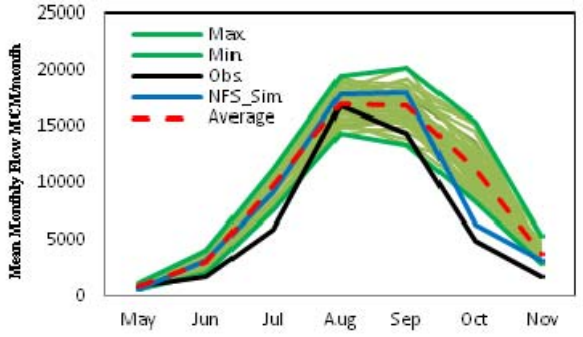

Simulation Flow at Atbara using NFS PET



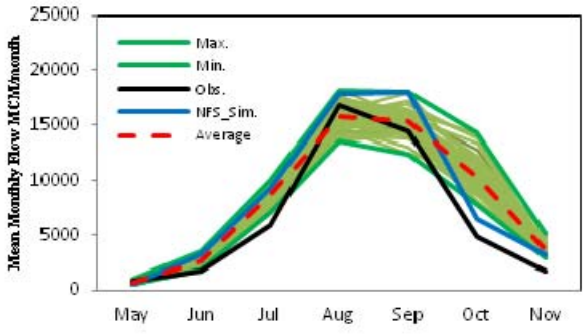

Simulation Flow at Diem using forecasted PET




Simulation Flow at Diem using NFS PET




Simulation Flow at Khartoum using forecasted PET




Simulation Flow at Khartoum using NFS PET

**Figure 5-5 Simulations Flow in the Eastern Nile using the forecasted PET comparing to the simulations flow using NFS_PET**

## 5.3 RETROACTIVE FORECAST SKILL

Performance of mathematical models in general and hydrological models in particular, is evaluated statistically by calculating one or more performance criteria. These measures are also used for model calibration by finding optimal parameters that maximize or minimize a selected measure or a set of measures in case of multi-criteria methods. The most commonly used measures in hydrologic related applications are the Root Mean Squared Error (RMSE) and the Nash-Sutcliffe (coefficient of) Efficiency (Nash and Sutcliffe 1970) calculated for either daily or monthly stream flow (as the evaluation variable). The Nash-Sutcliffe efficiency (R2) is particularly appealing because it is dimensionless, and thus can be

used to compare results for different locations/variables irrespective of the actual values of the simulated quantity. It is also easy to interpret as the percentage of the observed variance explained by the model and has the special zero value at which the mean of the observations is as good as the used model.

However, R2 and RMSE give more weight to peaks and peak timing (Gupta et al. 2003) because they involve squaring of the errors in addition to being sensitive to outliers (Legates and McCabe 1999). Therefore, other measures have been developed by many researches. These include the Heteroscedastic Maximum Likelihood Estimator (HMLE) (Sorooshian and Dracup 1980) and the index of agreement (Willmott 1981). In addition, Legates and McCabe (1999) presented modifications of the coefficient of efficiency and the index of agreement by using absolute error values instead of squared ones to give equal weights to high and low flows, and by using a time-varying mean for observations to account for seasonality. Boyle et al. (2000) developed another approach to assess model simulations of the different parts of the hydrograph (peak, rising and recession limbs, baseflow) by partitioning the hydrograph into three time segments and calculating the RMSE separately for each segment from all time steps belonging to that segment. This attempts to extract more information from the same data set (Wagener et al. 2003). Hogue et al. (2003) used a squared log error criterion to give more weight to errors in low flows.

Other commonly used criteria include the correlation coefficient, the bias, and the absolute bias. The coefficient of correlation and its squared value, the coefficient of determination, were criticized by Legates and McCabe (1999) because they only measure the linear association between the simulated and observed series and cannot detect consistent overestimation or underestimation errors. The two bias measures focus on the overall volume balance over the simulation period and can either be calculated as averages over the simulation period or as percentages of the annual observed volume. The bias, in particular, can be misleading if used alone because errors may cancel each other so that the sum of errors is close to zero. That is the reason for using a sign cancelling transformation such as squaring or absolute value transformations but other transformations were used such as the power transformation of Box and Cox (1964). Legates and McCabe (1999) investigated the use of the coefficient of determination, the coefficient of efficiency, and the index of agreement for assessing simulations of evapotranspiration and runoff and concluded that the coefficient of efficiency and the index of agreement should be used instead of correlation-based measures (i.e. the correlation and determination coefficients). They also recommended that at least one of the absolute error measures (i.e. those having the dimensions as the simulated variable such as the RMSE, absolute bias, etc.) be used in conjunction with relative (i.e. dimensionless) measures like the R2 or the index of agreement, or their modified counterparts, for a more meaningful assessment of model results. Visual inspection of hydrographs/distributions and scatter plots in addition to direct comparisons of

modelled and observed statistics (e.g. the mean and standard deviations) all aid the assessment. Table5-1 gives definitions and information about the various criteria.

**Table5-1 Performance Criteria Definitions**

| Criterion | Symbol | Equation | Range & Best Value | Emphasis |
|---|---|---|---|---|
| Root Mean Squared Error | *RMSE* | $\sqrt{\frac{1}{N_p}\sum_{N_p}(Z_s - Z_o)^2}$ | $0 - +\infty$ 0 (Minimize) | Peak values Timing of peaks |
| Nash Sutcliffe Efficiency | $R^2$ | $1 - \frac{\sum_{N_p}(Z_s - Z_o)^2}{\sum_{N_p}(Z_o - \overline{Z_o})^2}$ | $-\infty - 1$ 1 (Maximize) | Peak values Timing of peaks |
| Mean Error | *BIAS* | $\frac{1}{N_p}\sum_{N_p}(Z_s - Z_o)$ | $-\infty - +\infty$ 0 | Volume balance |
| Mean Absolute Error | *MAE* | $\frac{1}{N_p}\sum_{N_p}\lvert Z_s - Z_o\rvert$ | $0 - +\infty$ 0 (Minimize) | Volume balance |
| Mean Log Error | *MLE* | $\frac{1}{N_p}\sum_{N_p}(Ln(Z_s) - Ln(Z_o))$ | $-\infty - +\infty$ 0 | Baseflow |
| Mean Squared Log Error | *MSLE* | $\frac{1}{N_p}\sum_{N_p}(Ln(Z_s) - Ln(Z_o))^2$ | $0 - +\infty$ 0 (Minimize) | Baseflow |
| Correlation Coefficient | *r* | $\frac{\sum_{N_p}(Z_s - \overline{Z_s})(Z_o - \overline{Z_o})}{\sqrt{\sum_{N_p}(Z_s - \overline{Z_s})^2 (Z_o - \overline{Z_o})^2}}$ | $0 - 1$ 1 (Maximize) | Overall (linear) association between observed and simulated runoff |

$Z_s$ and $Z_o$ : Simulated and Observed Values; $\overline{Z_s}$ and $\overline{Z_o}$ : Mean Simulated and Observed Values

$N_p$: Number of observation points

### 5.3.1 NFS Evaluation

The last evaluation of the NFS was in 2011. In this evaluation, NFC used the above criteria. The NFS performance is generally satisfactory for the Blue Nile given the uncertainties in rainfall estimates and irrigation abstractions in addition to the fact that Lake Tana and the small swamps in some sub-basins are not explicitly modelled.

In Atbara basin, as the Atbara runs nearly dry outside the flood season, errors in the annual flow are mainly due to errors in simulating the peak flow in the different years. The association between the simulated annual series and the MAP series is quite high (0.76) compared to that for the observed flow (0.62) indicating that rainfall estimates are the main error source. Table 5-2 summarizes the results of the comparisons for the Eastern Nile at the designated key locations for the monthly series. These results are based on the 1989-2009 period. The NFS performance is satisfactory for the Blue Nile which reflects the NFS historical development and design objectives. The main objective of the NFS was to forecast

the inflows of the Nile basin to Lake Nasser which are calculated at Dongola rather than performing long-term simulations. As most of the flow originates from the Blue Nile, the development and calibration of the NFS focused on the Blue Nile. As a result the performance for this catchment is better than other areas. The performance for the Atbara is reasonable because of its similarity with the Blue Nile.

**Table 5-2 Summary Statistics for Key Locations along the Eastern Nile,** *NFS Evaluation Report, 2011.*

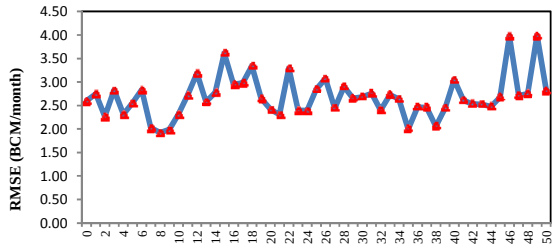
| Station | r | $R^2$ | MAE (BCM) | RMAE % | RMSE (BCM) | RRMSE % | Mean Annual Volume (BCM) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | OBS | NFS |
| Diem | 0.96 | 0.92 | 0.77 | 1.52 | 1.38 | 2.72 | 50.74 | 50.80 |
| Khartoum | 0.94 | 0.88 | 1.01 | 2.25 | 1.70 | 3.80 | 44.80 | 42.23 |
| Atbara | 0.82 | 0.66 | 0.52 | 4.12 | 1.24 | 9.82 | 12.63 | 10.83 |

*MAE: Mean Absolute Error; RMSE: Root Mean Squared Error; RMAE: Relative MAE to Mean Observed Annual Volume; RRMSE: Relative RMSE to Mean Observed Annual Volume; $R^2$: Nash-Sutcliffe Efficiency; r: coefficient of correlation*

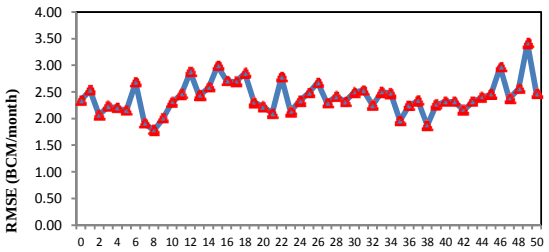
### 5.3.2 Skill scores of the results

Figure 5-6, Figure 5-7and Figure 5-8 show the Skill Scores criteria which study the performance of each simulation flow ensemble from (0-51), references to the Observed flow at the main key locations over the Eastern Nile (Atbara, at Atbara river, and Diem, and Khartoum, at Blue Nile). From the results it found that the best ensemble simulate the performance of the Atbara River at river station was ensemble No. 8, while the worst ensemble was ensemble No. 49. In the Blue Nile the best ensemble in both Diem and Khartoum was ensemble No. 43, and the worst one in both locations also was ensemble No.18.

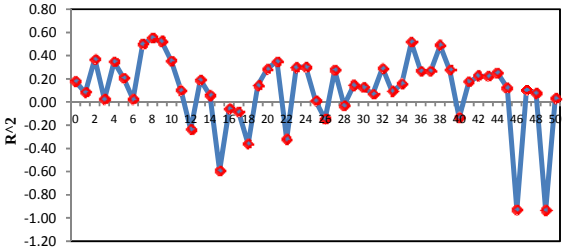
The skill scores of the ensembles in case of using the NFS_PET give better indication on the performance in case of using forecasted PET.



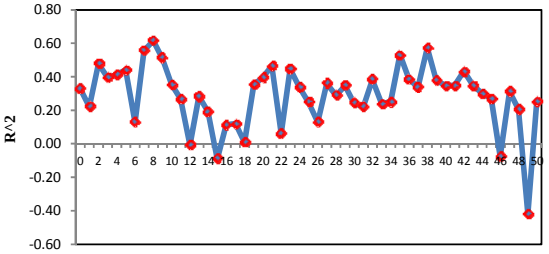

RMSE for the Simulation using Forecasted PET



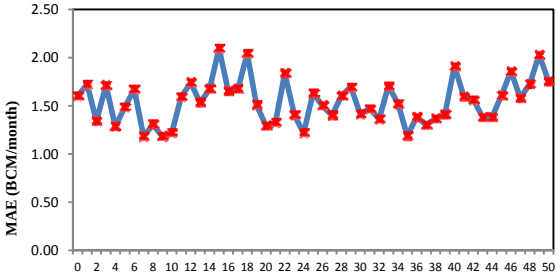

RMSE for the Simulation using NFS_PET



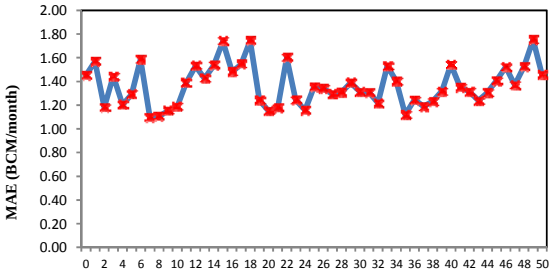

R^2 for the Simulation using Forecasted PET



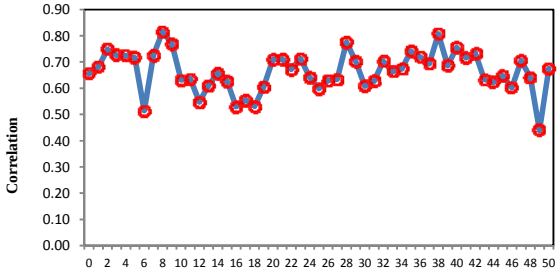

R^2 for the Simulation using NFS_PET



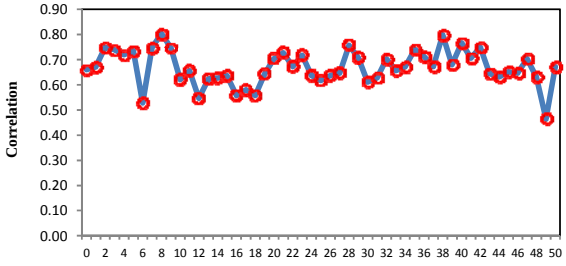

MAE for the Simulation using Forecasted PET




MAE for the Simulation using NFS_PET

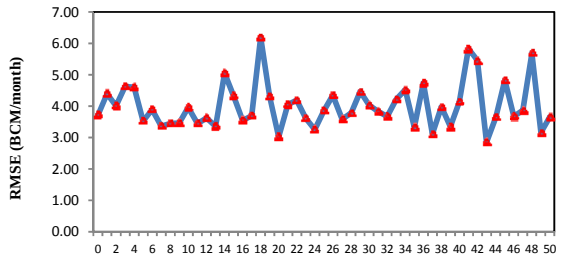



Corr. for the Simulation using Forecasted PET

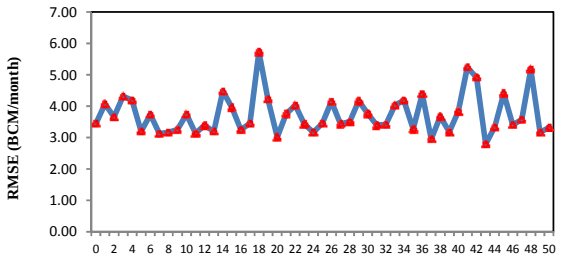



Corr. for the Simulation using NFS_PET

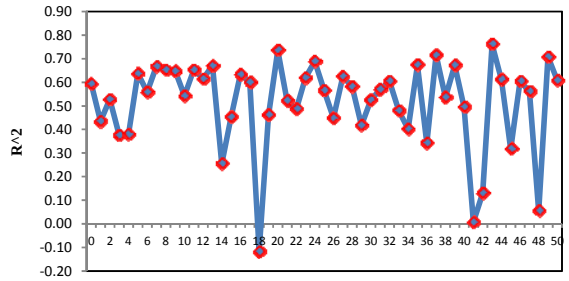
**Figure 5-6 Skill Score criteria at Atbara Basin**



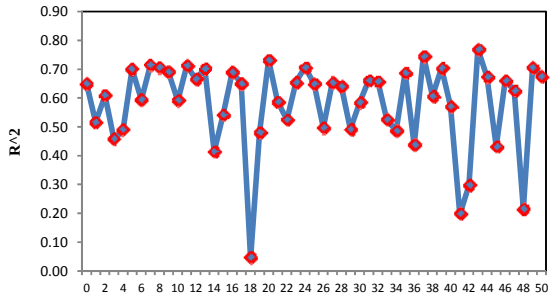

RMSE for the Simulation using Forecasted PET



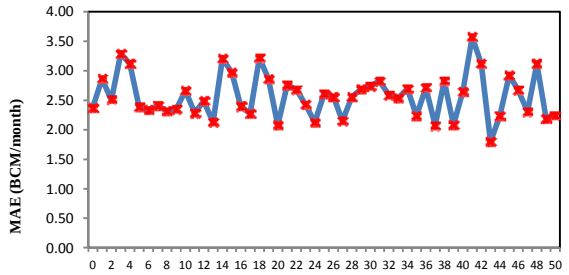

RMSE for the Simulation using NFS_PET



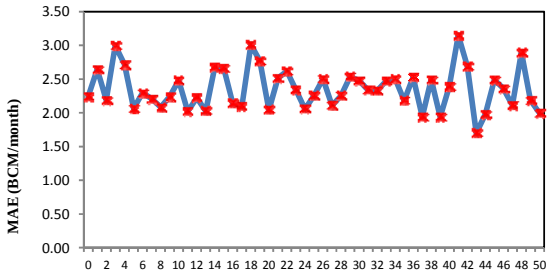

R^2 for the Simulation using Forecasted PET



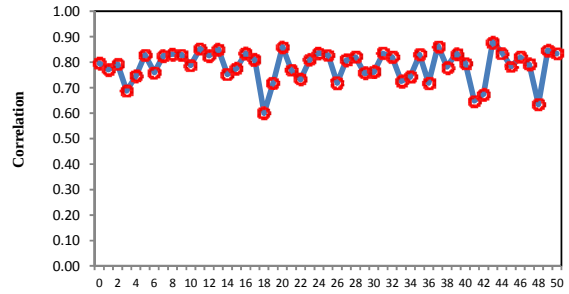

R^2 for the Simulation using NFS_PET



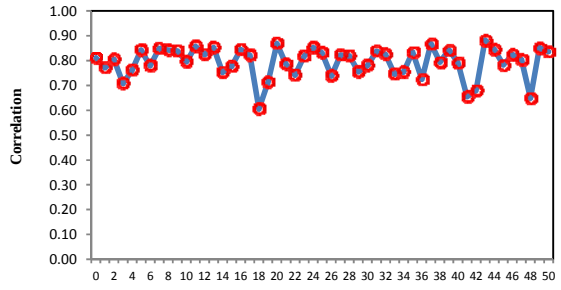

MAE for the Simulation using Forecasted PET




MAE for the Simulation using NFS_PET

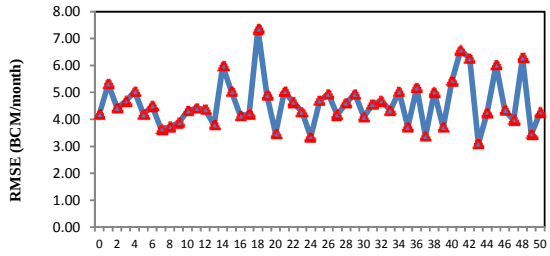



Corr. for the Simulation using Forecasted PET

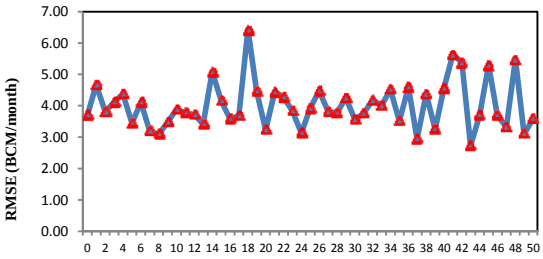



Corr. for the Simulation using NFS_PET

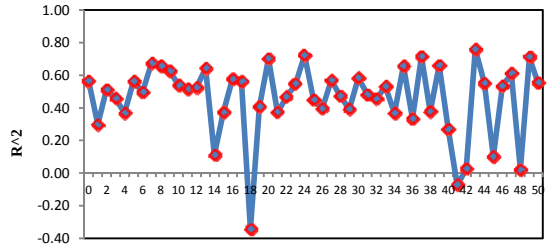
**Figure 5-7 Skill Score criteria at Diem Basin**



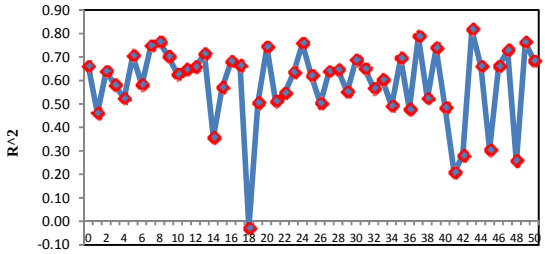

RMSE for the Simulation using Forecasted PET



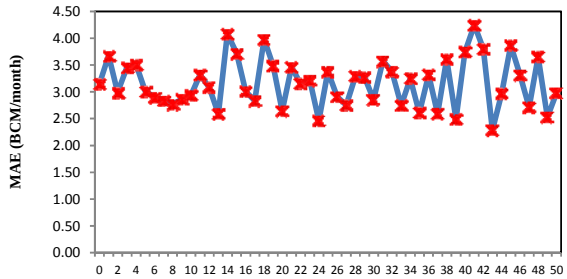

RMSE for the Simulation using NFS_PET



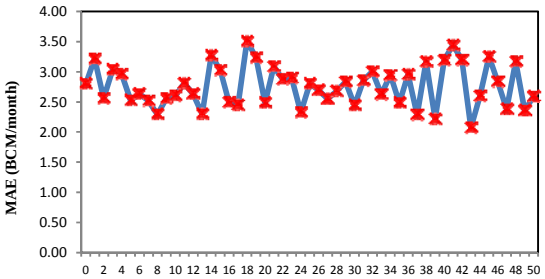

R^2 for the Simulation using Forecasted PET



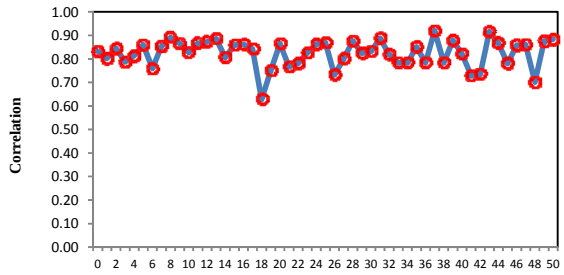

R^2 for the Simulation using NFS_PET



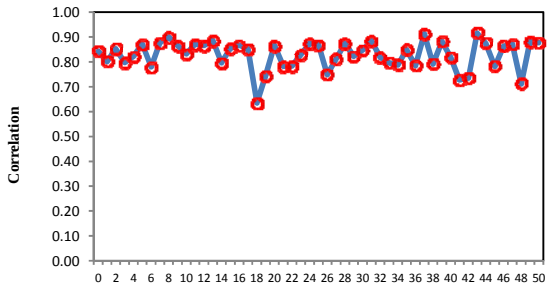

MAE for the Simulation using Forecasted PET




MAE for the Simulation using NFS_PET




Corr. for the Simulation using Forecasted PET




Corr. for the Simulation using NFS_PET

**Figure 5-8 Skill Score criteria at Khartoum Basin**

# 6. REFERENCES

Bellerby, T., M. Todd, D. Kniveton & C. Kidd (2000) Rainfall Estimation from a Combination of TRMM Precipitation Radar and GOES Multispectral Satellite Imagery through the Use of an Artificial Neural Network. *Journal of Applied Meteorology,* 39**,** 2115-2128.

Box, G. E. P. & D. R. Cox (1964) An Analysis of Transformations. *Journal of the Royal Statistical Society. Series B (Methodological),* 26**,** 211-252.

Boyle, D. P., H. V. Gupta & S. Sorooshian (2000) Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research,* 36**,** 3663-3674.

Brown, J. D., J. Demargne, D.-J. Seo & Y. Liu (2010) The Ensemble Verifcation System (EVS): A software tool for verifying ensemble forecasts of hydrometeorological and hydrologic variables at discrete locations. *Environmental Modelling & Software,* 25**,** 854- 872.

Capacci, D. & B. J. Conway (2005) Delineation of precipitation areas from MODIS visible and infrared imagery with artificial neural networks. *Meteorological Applications,* 12**,** 291-305.

Day, G. (1985) Extended Streamflow Forecasting Using NWSRFS. *Journal of Water Resources Planning and Management,* 111**,** 157-170.

DEWFORA. 2011a. WP2-D2.1 - Inventory of drought monitoring and forecasting systems in Africa. DEWFORA project - EU FP7.

---. 2011b. WP2-D2.2 - Inventory of institutial frameworks and drought mitigation and adaptation practices in Africa. DEWFORA project - EU FP7.

---. 2011c. WP4-D4.1 - Meteorological drought forecasting (monthly to seasonal forecasting) at regional and continental scale - ECMWF and CCAM meteorological data description DEWFORA project - EU FP7.

---. 2012a. WP4-D4.7 - Downscaled and tailor made hydrological models for the Limpopo and Niger case study basins. DEWFORA project - EU FP7

---. 2012b. WP6-D6.1 - Implementation of improved methodologies in comparative case studies - Inception report for each case study. DEWFORA project - EU FP7.

Diro, G. T., E. Black & D. I. F. Grimes (2008) Seasonal Forecasting of Ethiopian Spring Rains. *Meteorological Applications* 83**,** 73–83.

Diro, G. T., D. I. F. Grimes & E. Black (2011) Teleconnections Between Ethiopian Summer Rainfall and Sea Surface Temperature: Part II. Seasonal Forecasting. *Climate Dynamics 37 (1-2) (September 14): 121–131*.

Dutra, E., F. Di Giuseppe, F. Wetterhall & F. Pappenberger (2012) Seasonal forecasts of drought indices in African basins. *Hydrol. Earth Syst. Sci. Discuss.,* 9(9), 11093-11129.

DWA. 2013. Tzaneen Dam. Department of Water Affairs, South Africa.

Gupta, H. V., S. Sorooshian, T. S. Hogue & D. P. Boyle. 2003. Advances in Automatic Calibration of Watershed Models. In *Calibration of Watershed Models,* eds. Q. Duan, H. V. Gupta, S.

Sorooshian, A. N. Rousseau & R. Turcottte, 9-28. Washington D.C.: American Geophysical Union.

Hanley, J. A. (1996) The use of the 'binormal' model for parametric ROC analysis of quantitative diagnostic tests. *Statistics in Medicine,* 15**,** 1575-1585.

Hogue, T. S., H. V. Gupta, S. Sorooshian & C. D. Tomkins. 2003. A Multi-Step Automatic Calibration Scheme for Watershed Models. In *Calibration of Watershed Models,* eds. Q. Duan, H. V. Gupta, S. Sorooshian, A. N. Rousseau & R. Turcottte, 165-174. Washington D.C.: American Geophysical Union.

Inoue, T. (1987) An instantaneous delineation of convective rainfall areas using split window data of NOAA-7 AVHRR. *Journal of the Meteorological Society of Japan,* 65**,** 469-481.

Legates, D. R. & G. J. McCabe (1999) Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research,* 35**,** 233-241.

Lindeman, R. H., P. F. Merenda & R. Z. Gold. 1980. *Introduction to bivariate and multivariate analysis*. Scott, Foresman Glenview, IL.

Metz, C. E. & X. Pan (1999) “Proper” Binormal ROC Curves: Theory and Maximum-Likelihood Estimation. *Journal of Mathematical Psychology,* 43**,** 1-33.

Molteni, F., T. Stockdale, M. A. Balmaseda, G. BALSAMO, R. Buizza, L. Ferranti, L. Magnunson, K. Mogensen, T. Palmer & F. Vitart. 2011. The new ECMWF seasonal forecast system (System 4), ECMWF Tech. Memo., 656, 49 pp.

Mwale, D., T. Y. Gan, S. S. Shen, T. T. Shu & K.-M. Kim (2007) Wavelet empirical orthogonal functions of space-time-frequency regimes and predictability of southern Africa summer rainfall. *Journal of Hydrologic Engineering,* 12**,** 513-523.

Nash, J. E. & J. V. Sutcliffe (1970) River flow forecasting through conceptual models part I -- A discussion of principles. *Journal of Hydrology,* 10**,** 282-290.

R Core Team. 2012. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

Rayner, N., P. Brohan, D. Parker, C. Folland, J. Kennedy, M. Vanicek, T. Ansell & S. Tett (2006) Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. *Journal of Climate,* 19**,** 446-469.

Roy-Desrosiers, F. 2012. ANN: Feedforward Artificial Neural Network optimized by Genetic Algorithm. In *R package version 0.1.4. http://CRAN.R-project.org/package=ANN*.

Seibert, M. & H. Apel. 2013. Statistical Analysis of drought variability and hydrological parameters. DEWFORA, Deliverable 4.6.

Shukla, S. & A. W. Wood (2008) Use of a standardized runoff index for characterizing hydrologic drought. *Geophysical research letters,* 35.

Sorooshian, S. & J. A. Dracup (1980) Stochastic Parameter-Estimation Procedures for Hydrologic Rainfall-Runoff Models - Correlated and Heteroscedastic Error Cases. *Water Resources Research,* 16**,** 430-442.

Todd, M. C., C. Kidd, D. Kniveton & T. J. Bellerby (2001) A Combined Satellite Infrared and Passive Microwave Technique for Estimation of Small-Scale Rainfall. *Journal of Atmospheric and Oceanic Technology,* 18**,** 742-755.

Turk, F. J. & S. D. Miller (2005) Toward improving estimates of remotely-sensed precipitation with MODIS/AMSR-E blended data techniques. *IEEE Transactions on Geoscience and Remote Sensing,* 43**,** 1059-1069.

van Beek, L. P. H. 2008. Forcing PCR-GLOBWB with CRU data. Utrecht University.

van Beek, L. P. H. & M. F. P. Bierkens. 2009. The Global Hydrological Model PCR-GLOBWB: Conceptualization, Parameterization and Verification. Utrecht, The Netherlands: Utrecht University, Faculty of Earth Sciences, Department of Physical Geography.

Wagener, T., N. McIntyre, M. J. Lees, H. S. Wheater & H. V. Gupta (2003) Towards reduced uncertainty in conceptual rainfall-runoff modelling: Dynamic identifiability analysis. *Hydrological Processes,* 17**,** 455-476.

Werner, M., J. Schellekens, P. Gijsbers, M. van Dijk, O. van den Akker & K. Heynert (2013) The Delft-FEWS flow forecasting system. *Environmental Modelling & Software,* 40**,** 65-77.

Willmott, C. J. (1981) On the validation of models. *Physical Geography,* 2**,** 184-194.

WWRP/WGNE. updated: 26 January 2012. Forecast verification: Issues, Methods and FAQ. http://www.cawcr.gov.au/projects/verification/#Methods_for_probabilistic_forecasts.

Zwarts, L. 2009. Predicting the Annual Peak Flood Level in the Inner Niger Delta. A&W-rapport 1254. Altenburg & Wymenga, Ecological Consultants, Feanwâlden." A&W-report 1254. Altenburg & Wymenga ecologisch onderzoek.

---. 2010. Towards a Further Extension of the OPIDIN tool/Vers Une Nouvelle Extension de L'outil OPIDIN. A&W-report 1514/1528. Altenburg & Wymenga ecologisch onderzoek.

Zwarts, L., P. van Beukering, B. Kone & E. Wymenga. 2005. The Niger, a Lifeline, Effective Water Management in the Upper Niger Basin. RIZA IVM / WISO / A&W. http://www.altwym.nl/uploads/file/133Executive%20summary%20-%20The%20Niger,%20a%20lifeline.pdf.

# 7. APPENDICES

## 7.1 NIGER CASE STUDY

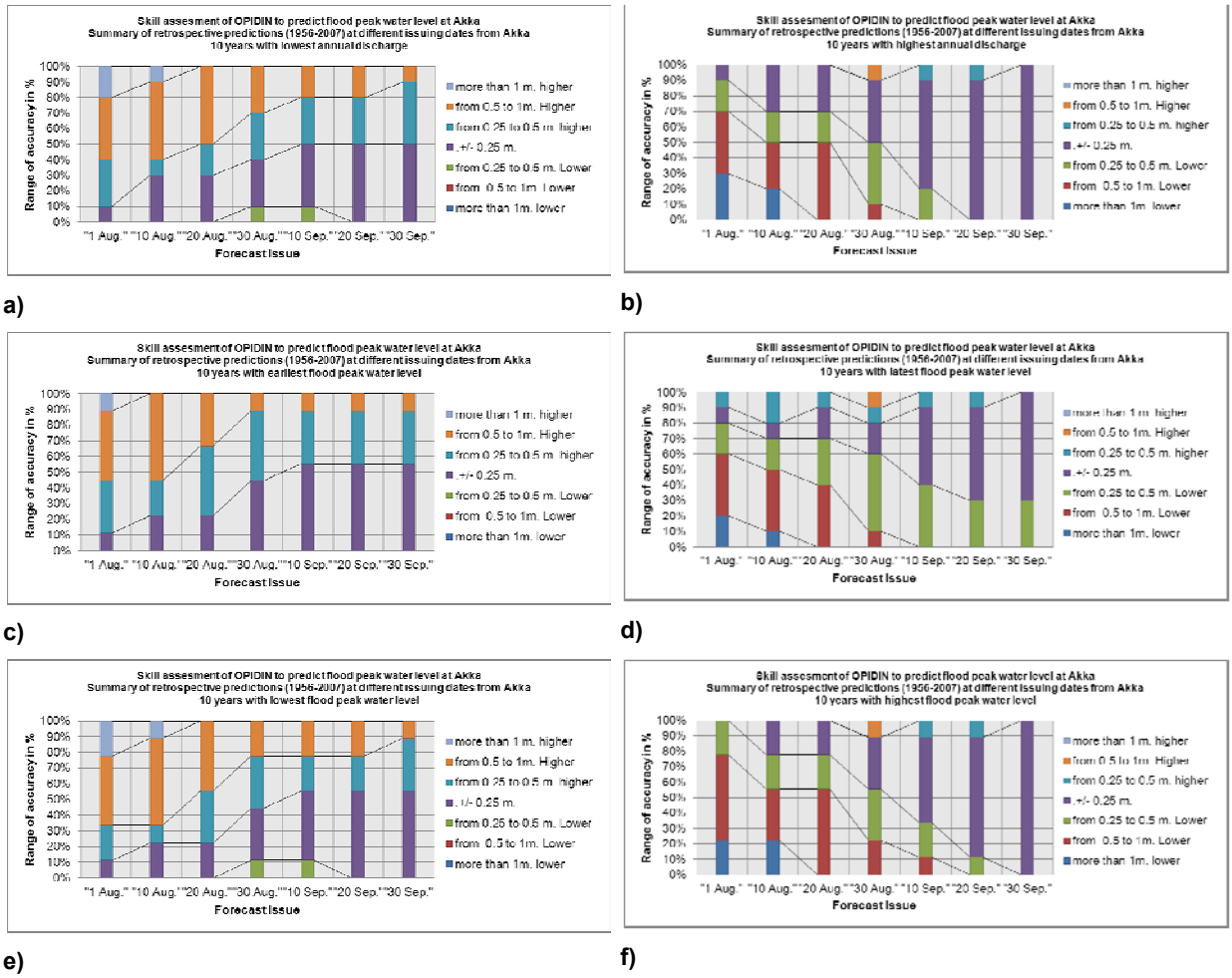
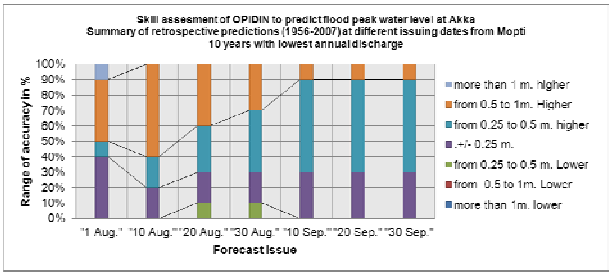
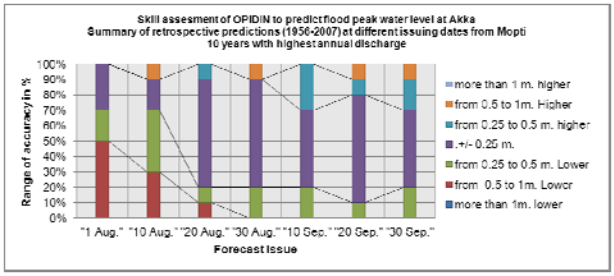


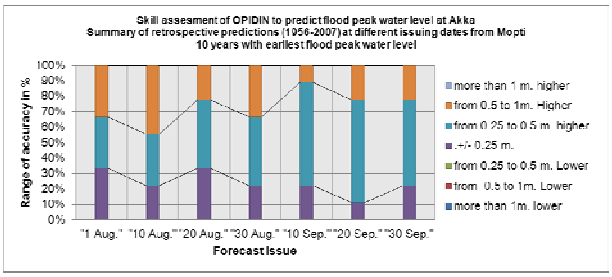

**Figure 7-1 Annual flood peak prediction for dry and wet episode at Akka from Akka using classes of quantitative deviation**



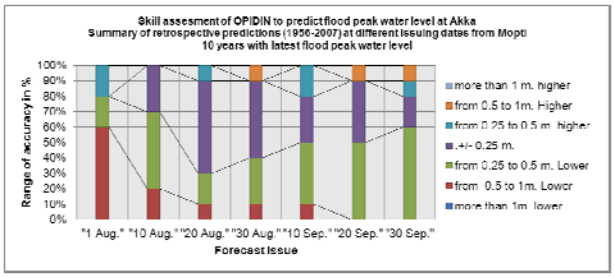

a)



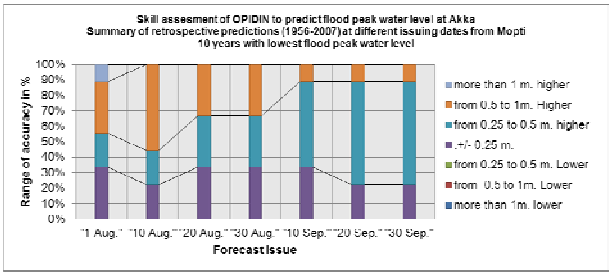

b)



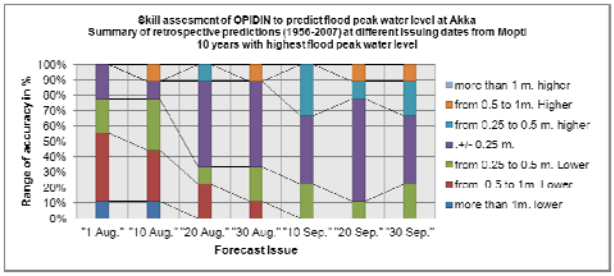

c)




d)




e)




f)

**Figure 7-2 Annual flood peak prediction for dry and wet episode at Akka from Mopti using classes of quantitative deviation**

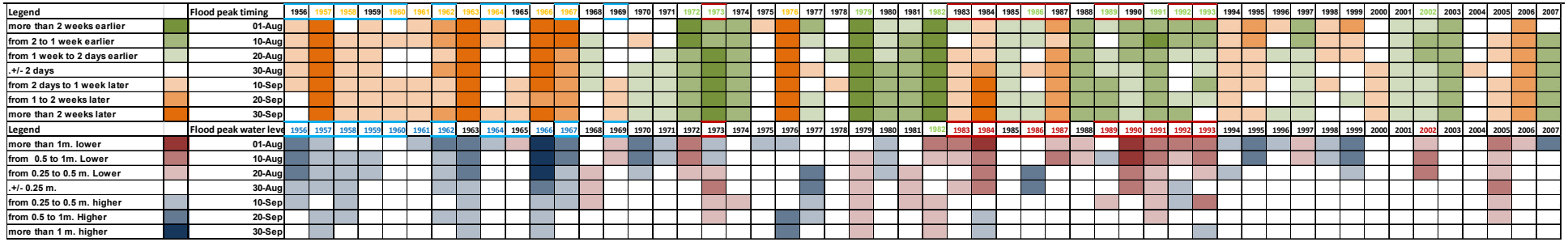



**Figure 7-3: Annual flood peak prediction at Mopti from Mopti using classes of quantitative deviation**

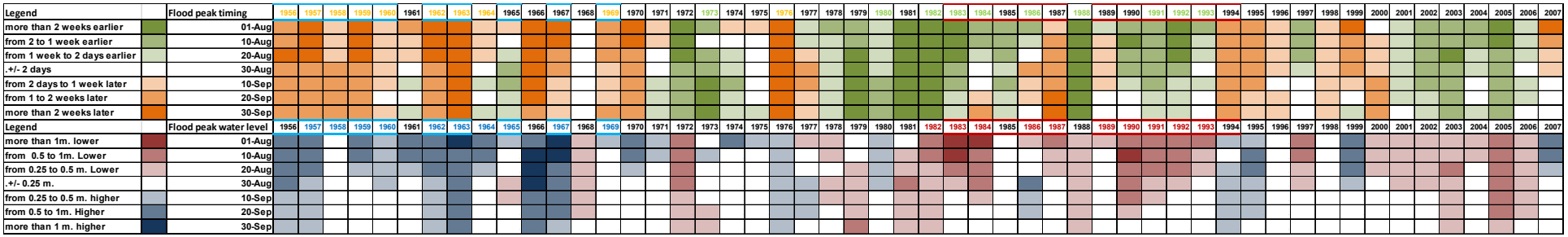



**Figure 7-4: Annual flood peak prediction at Akka from Akka using classes of quantitative deviation**

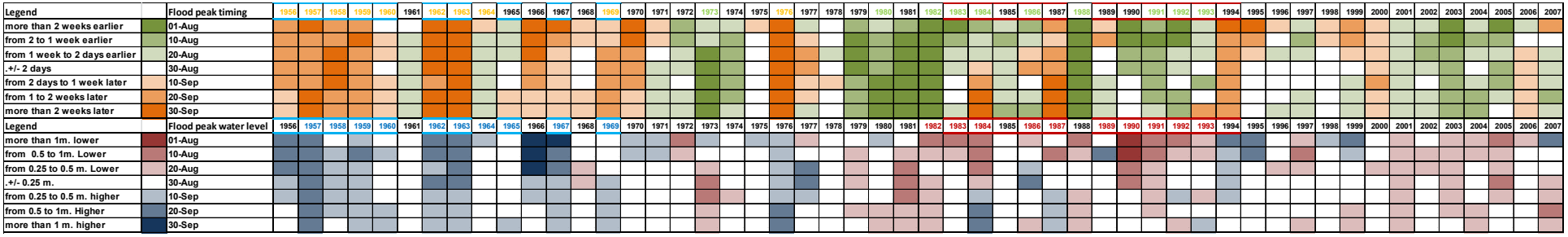



**Figure 7-5: Annual flood peak prediction at Akka from Mopti using classes of quantitative deviation**

**Additional legends to interpret the three figures above**

Based on 1956 to 2007, the 10 years with:

- the lowest annual discharge have blue top/and bottom borders_
- the highest annual discharge have red top/bottom borders_
- the earliest flood peak timing have **text in orange**
- the latest peak timing have **text in green**
- the lowest peak flood water level have **text in blue**
- the highest peak flood water level have **text in red**

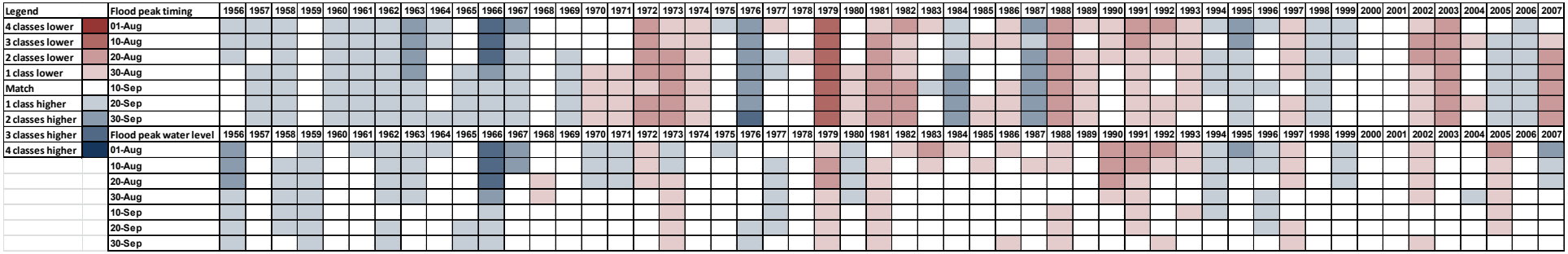



**Figure 7-6: Annual flood peak prediction at Mopti from Mopti using early warning classes**

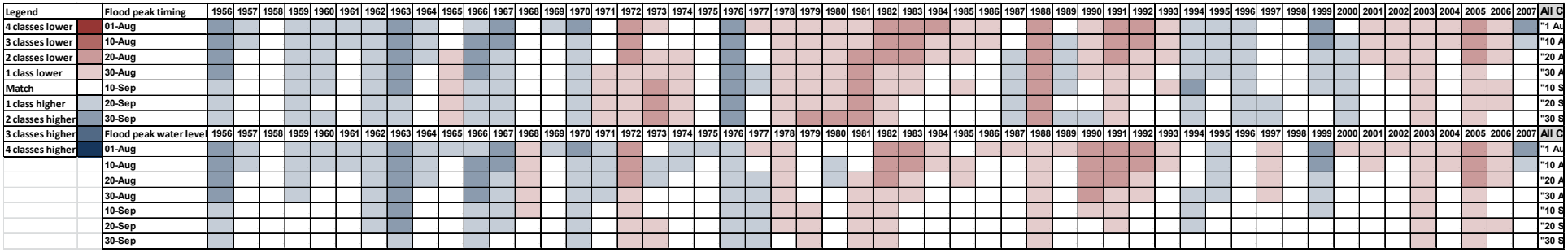



**Figure 7-7: Annual flood peak prediction at Akka from Akka using early warning classes**

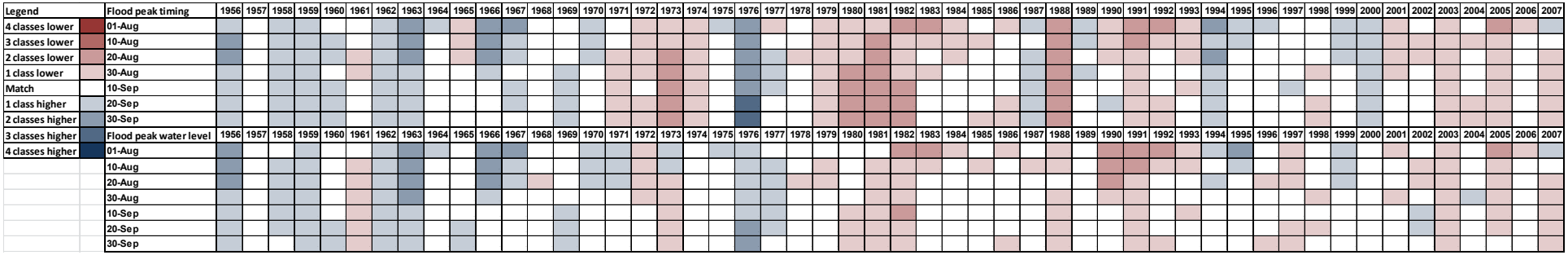
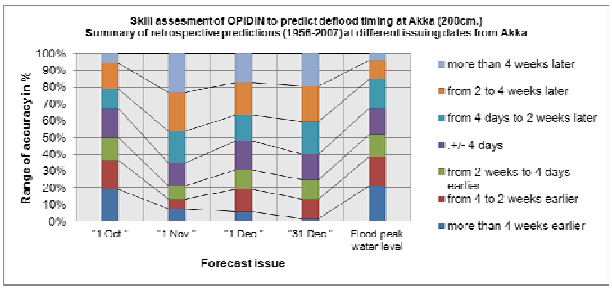
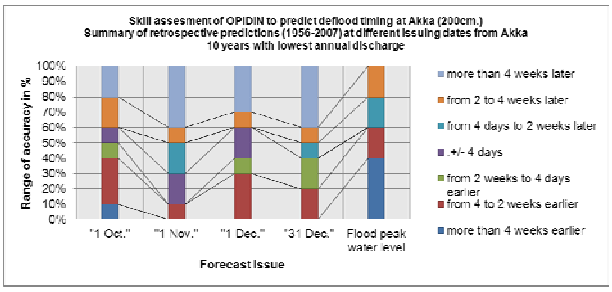


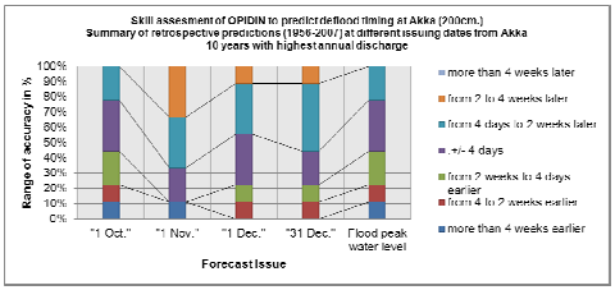

**Figure 7-8: Annual flood peak prediction at Akka from Mopti using early warning classes**



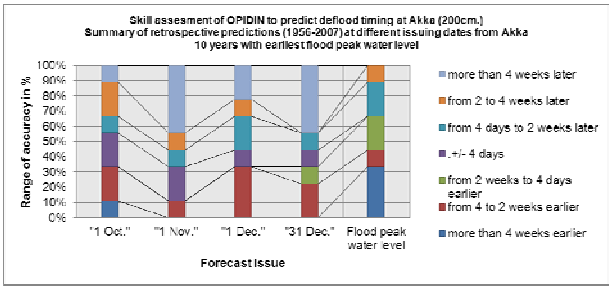

a)



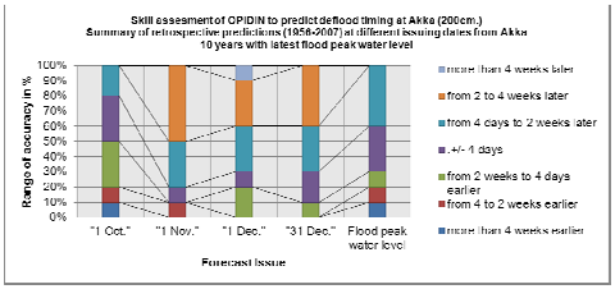

b)



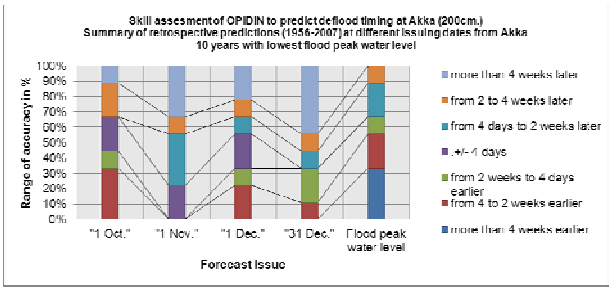

c)



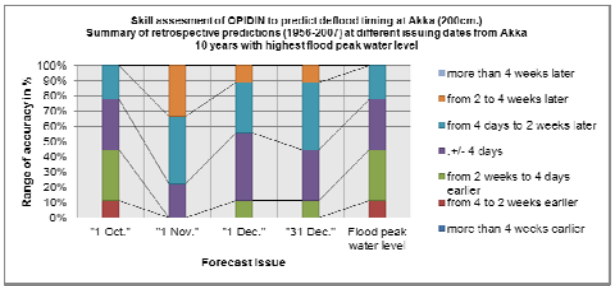

d)




e)

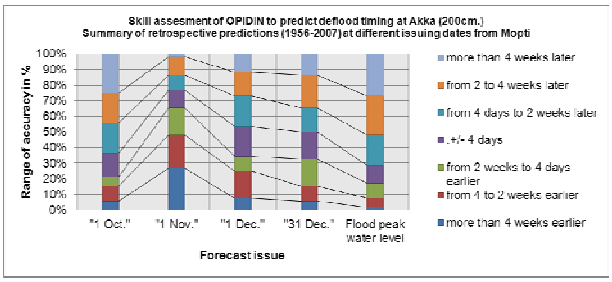



f)

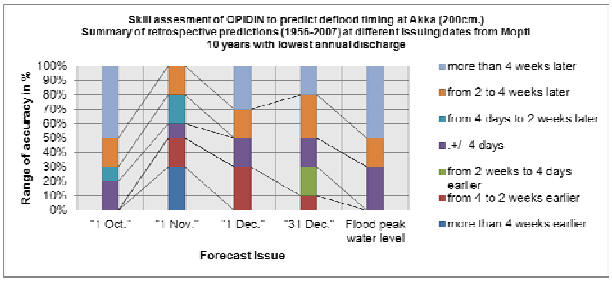



g)

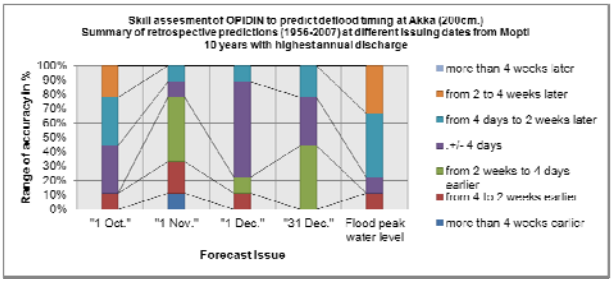
**Figure 7-9 Annual deflood (200cm.) prediction for all, dry and wet episode at Akka from Akka using classes of quantitative deviation**



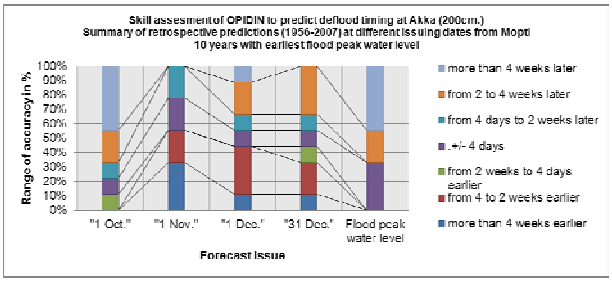

a)



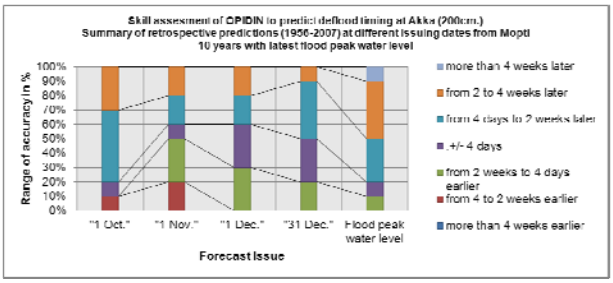

b)



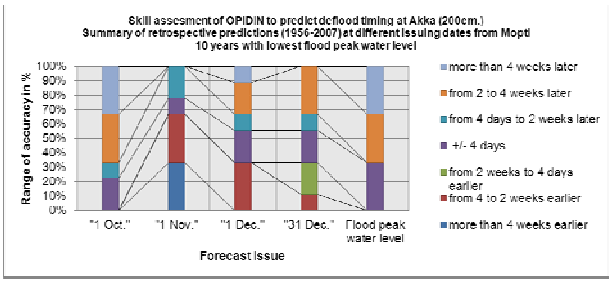

c)



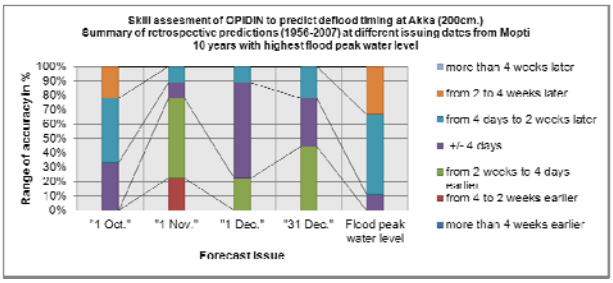

d)




e)




f)




g)

**Figure 7-10 Annual deflood (200cm.) prediction for all, dry and wet episode at Akka from Mopti using classes of quantitative deviation**

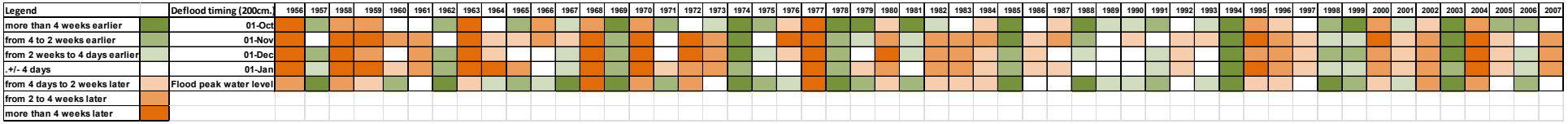



**Figure 7-11: Annual deflood timing (200cm.) at Akka from Akka using classes of quantitative deviation**

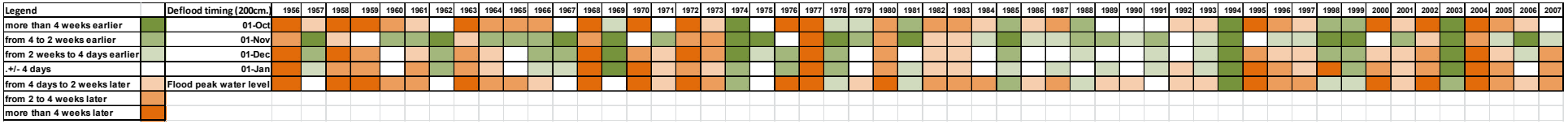



**Figure 7-12: Annual deflood timing (200cm.) at Akka from Mopti using classes of quantitative deviation**