

Statistical downscaling of GCM simulations to Streamflow

Willem A. Landman^{a,c,*}, Simon J. Mason^b, Peter D. Tyson^a, Warren J. Tennant^c

^a*Climatology Research Group, University of the Witwatersrand, Johannesburg 2050, South Africa*

^b*International Research Institute for Climate Prediction, Scripps Institute of Oceanography, University of California San Diego, La Jolla, CA 92093-0235, USA*

^c*South African Weather Bureau, Private Bag X097, Pretoria 0001, South Africa*

Received 22 August 2000; revised 1 June 2001; accepted 4 June 2001

Abstract

A multi-tiered forecast procedure is employed to simulate real-time operational seasonal forecasts of categorized (below-normal, near-normal and above-normal) streamflow at the inlets of twelve dams of the Vaal and upper Tugela river catchments in South Africa. Forecasts are made for the December to February (DJF) season over an 8-year independent period from 1987/1988 to 1994/1995. A physically based model of the atmosphere system, known as a general circulation model (GCM), is used to simulate atmospheric variability over southern Africa, the output of which is statistically downscaled to streamflow. The GCM used is the COLA T30, and is forced at the boundary with predicted monthly-mean global sea-surface temperatures. The monthly-mean sea-surface temperature fields are first predicted over lead-times of several months using a canonical correlation analysis (CCA) model. GCM simulations are then obtained for an area including most of southern Africa and adjacent oceans. The GCM simulations are downscaled to catchment level from coarse resolution gridded climate variables, using a perfect prognosis approach: bias-corrected GCM simulations are substituted into the perfect prognosis equations to provide the downscaled categorized streamflow forecasts. Although surface characteristics of each catchment that affect the variability of streamflow are not considered in the proposed downscaling system, successful forecasts of streamflow categories were obtained for some of the years forecast independently. The scheme's operational utility is thus demonstrated, albeit over short lead-times. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Streamflow; General circulation model; Downscaling; Canonical correlation analysis; Perfect prognosis; South Africa

1. Introduction

The drought over southern Africa during the 1982/1983 El Niño event caused estimated damages in the vicinity of US\$1 billion (Moura et al., 1992). The impact of this event was made worse by a preceding drought over most of the region. Similarly, the El Niño event of the 1991/1992 season caused widespread drought. In contrast to this, a major flooding

event that lasted several days during February of 1988 (Triegaardt et al., 1991) also caused damages with the loss of lives and of livestock. Climate variations such as these have an important impact on water supply in the southern African region, to name but one impact. An ever-increasing population and an associated increase in demand for fresh water resources have made effective water management in the region essential. The need for forecast guidance for an approaching season has thus become necessary, and because understanding of the predictability of the atmosphere at seasonal-to-interannual time-scales has improved considerably during the past decade, the provision

* Corresponding author. Address: South African Weather Bureau, Private Bag X097, Pretoria, 001, South Africa.

E-mail address: willem@iri.columbia.edu (W.A. Landman).

of such forecasts has become a possibility (Carson, 1998).

Variability in sea-surface temperatures provides the main source of atmospheric predictability at seasonal time-scales (Palmer and Anderson, 1994), and has been shown to be related to southern African climate variability (e.g. Mason and Jury, 1997; Reason and Lutjeharms, 1998). Inclusion of anomalous sea-surface temperatures in the areas surrounding southern Africa into seasonal prediction models, has enabled predictions of climate variability over the region (Makarau and Jury, 1997; Mason, 1998; Landman and Mason, 1999). Great effort has gone into constructing seasonal forecast algorithms for southern Africa, and seasonal forecasts have been issued since the early 1990s by groups both locally and internationally (Mason et al., 1999; Jury et al., 1999; Landman and Mason, 1999). These efforts lead to a rapid increase in the availability of seasonal forecast products, albeit mostly for seasonal rainfall.

Most algorithms for predicting southern African inter-seasonal climate variability have made extensive use of linear statistics, relating seasonal rainfall to global-scale sea-surface temperatures, outgoing long-wave radiation and various atmospheric pressure and wind indices (Mason and Tyson, 1999). However, many important climate processes demonstrate strong non-linearities, and the forecast skill of statistical models is restricted owing to the exclusion of these important processes (Barnston et al., 1994; Carson, 1998). Furthermore, the inherent assumptions of most statistical models, such as that of Gaussian probability distributions, frequently are invalid. If the ocean–atmosphere system contains sufficient inherent predictability, physically based dynamical models should outscore statistical models (Barnston et al., 1994). Predictability using dynamical models of the atmosphere can be achieved by first predicting sea-surface temperatures for regions known to be partly responsible for the rainfall variability over land at a level of skill that is at least comparable with persistence. These predicted fields are then incorporated in a two-tiered forecasting system to force an atmospheric GCM (Bengtsson et al., 1993; Hunt et al., 1994; Mason et al., 1999). Research using general circulation models (GCMs) as forecast tools over southern Africa is currently underway (Mason et al., 1999; Landman et al., 2001), but the seasonal climate

modelling capacity of the southern African region has to be improved locally.

Accepting that rainfall over southern Africa during the main austral rainfall months of December to February (DJF) is affected by sea-surface temperature variability (Mason and Jury, 1997), seasonal streamflow variability may also be related to phenomena such as the El Niño/Southern Oscillation (ENSO) (Schulze, 1997; Compagnucci and Vargas, 1998). Given a sufficiently strong climate signal in streamflow, there is the potential to provide seasonal forecasts of streamflow. Streamflow for a given catchment, excluding the effects of dams or interbasin water transfers, is directly affected by precipitation, evaporation and changes in soil water storage (Schulze, 1997). However, non-climatic factors, such as vegetation cover and the soil surface characteristics of catchments, can play at least an important role in determining streamflow (Schulze, 1997). The rainfall/streamflow association is therefore complex, and depends on several factors mostly unrelated to atmospheric variability. These other factors are not directly incorporated into the linear relationships between atmospheric variability and streamflow that are described in this study. Notwithstanding, it is important to begin developing application forecasts, such as seasonal streamflow for important catchments, because of its potential value to decision makers.

In this paper a forecast system for streamflow categories (below-normal, near-normal and above normal) constructed using a GCM that considers the non-linearities in the climate system is described. The lower boundary conditions of the GCM are monthly-mean sea-surface temperatures, predicted using canonical correlation analysis (CCA) (Landman and Mason, 2001). The GCM-simulated circulation and moisture fields are bias-corrected, and subsequently used as predictors in prior constructed CCA-based perfect prognosis equations, based on the relationship between observed circulation and moisture, and naturalised streamflow. The GCM simulations are thus downscaled (Cui et al., 1995) because, although GCMs demonstrate significant skill at global or even continental scale, they are unable to represent local subgrid-scale features (Joubert and Hewitson, 1997). This multi-tiered forecast system is used to define reliable skill estimates, so that operational seasonal

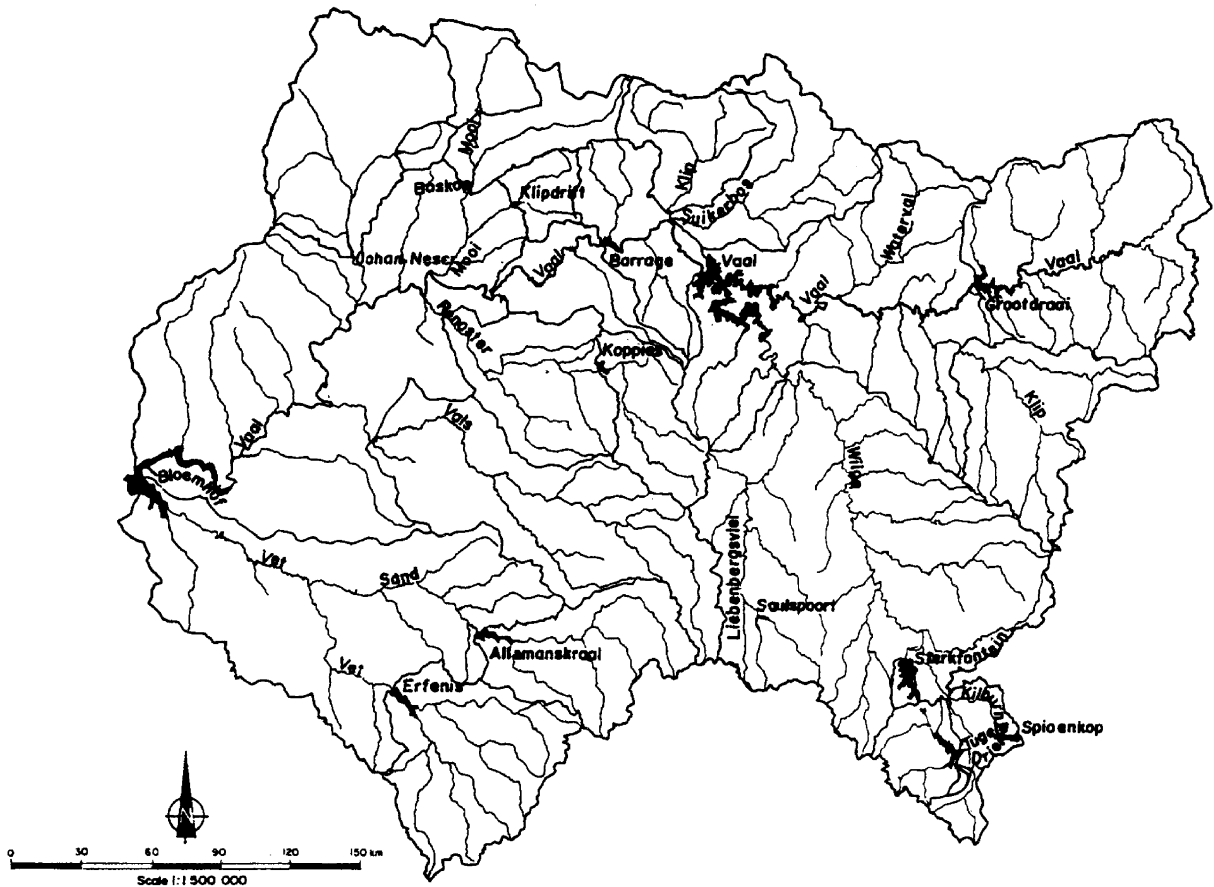


Fig. 1. Location map of the twelve dams and the catchments of the Vaal and upper Tugela river systems. Heavy lines depict the borders of the different catchments, and light lines the rivers, except for the Vaal river.

categorized streamflow forecasts can be made with realistic levels of confidence.

2. Data

2.1. Sea-surface temperatures

Sea-surface temperatures were used as the only predictors in the CCA sea-surface temperature forecasting model. Reconstructed sea-surface temperature fields using empirical orthogonal function (EOF) analysis (Jackson, 1991) are available for the period 1950–1995 (Smith et al., 1996). Optimum Interpolation (OI) sea-surface temperature data (Reynolds and Smith, 1994) were obtained from 1996 to date to

form a set of 3-month means. During operational predictions, the latest OI data are used to update the sea-surface temperature data set. Both the reconstructed (2° latitude \times longitude) and the OI (1° latitude \times longitude) sets were reduced to a 6° latitude \times 4° longitude grid (1078 points) to reduce the large matrices involved in the statistical forecast model. However, large oceanic features, such as El Niño or La Niña events, are still adequately represented by the coarser grid.

2.2. Streamflow

Naturalised streamflow data were obtained from BKS Engineering and Management. These streamflow data are available from 1920 to the first few

months of 1995 and are derived by eliminating the influence of developments within a catchment from the observed streamflow of that catchment. The catchment developments considered are irrigation demand on return flows, the effect of small dams, abstractions by and return flows from urban and industrial users, runoff or release from upstream catchments, and interbasin transfers. Naturalised streamflow data are available as monthly means measured in cubic meters per second, and from which 3-month mean DJF values were calculated. The streamflow at the inflow of twelve dams in the Vaal and upper Tugela catchments were considered (Fig. 1), and they are Allemanskraal, Barrage, Bloemhof, Boskop, Driel, Erfenis, Grootdraai, Koppies, Johan Nesor, Sterkfontein, Spioenkop and Vaal Dams. It is possible to use catchments where streamflow has not been affected significantly by developments, but forecasts have to be made for rivers where the forecasts may have societal economic value. Forecasts for the larger catchments, which tend to be most affected by development, are therefore of greatest interest.

2.3. Reanalysis data

GCM initial conditions at 24-h intervals were derived from the National Centers for Environmental Prediction (NCEP) reanalysis data (Kalnay et al., 1996). Optimal equations relating streamflow and atmospheric variables such as circulation and moisture were constructed using monthly-mean NCEP reanalysis data from 1950/1951 to 1994/1995 that spanned a window covering most of southern Africa and adjacent oceans (10–45°S; 10°W–60°E). The data have a horizontal resolution of 2.5° latitude X longitude and seventeen constant pressure levels in the vertical.

3. Methods

3.1. Steps of the downscaling methodology

A forecast procedure is employed to simulate seasonal forecasts of streamflow at the inlets of dams located within a large catchment in South Africa. Predictability of the streamflow using an atmospheric GCM can be estimated by first predicting global sea-surface temperature anomalies, which are

subsequently incorporated in a two-tiered forecasting system to force the atmospheric GCM (Bengtsson et al., 1993; Hunt et al., 1994; Mason et al., 1999; Landman et al., 2001). The GCM in turn produces atmospheric variables such as circulation and moisture fields as a result of the forcing. Statistical relationships between some of these atmospheric variables and the desired forecast parameter can be identified. In this paper, these relationships are utilized through optimally designed linear equations, making use of the perfect prognosis technique (Wilks, 1995). Streamflow forecasts are subsequently produced for three categories using the equations and tested over an independent period of 8 years. To calculate the categories, each of the observed and predicted fields is separated into its own three equiprobable classes or terciles for each dam of the predictand set. These terciles are referred to as above-normal (A), near-normal (N), or below-normal (B) categories.

3.2. The sea-surface temperature model

The statistically based technique used here to predict the near-global (45°N–45°S) sea-surface temperature anomalies is called canonical correlation analysis (CCA), a technique that has been used before as a seasonal forecast model of equatorial sea-surface temperatures in the Pacific Ocean (Barnston and Ropelewski, 1992). Canonical variates were used to make monthly sea-surface temperature anomaly forecasts, using evolutionary and steady-state features of antecedent sea-surface temperatures as predictors: the predictands were monthly near-global sea-surface temperatures, preceded by four seasonal (3-month) mean near-global sea-surface temperatures as predictors. For DJF streamflow forecasts issued at the beginning of September, sea-surface temperature forecasts of the 7 months leading up to March are required. For forecasts made in early November, the sea-surface temperature of 5 consecutive months are required. The following schematic illustrates the lead-times involved in making forecasts during early September and early November, respectively for the 1982/1983 season:

1981	1982	1983
<u>S O N</u>	<u>D J F M A M</u>	<u>J J A</u>
⇒ S, O, N, D, J, F, M		
(3-month lead)		

NDJFMAMJJASO ⇒ N, D, J, F, M (1-month lead)

Although the lead-times refer to the number of months leading up to the target season of DJF, much longer leads are involved in predicting the sea-surface temperatures at the end of the predictand period, i.e. forecasts made in early September for February have a 5-month lead.

In constructing the CCA model, pre-orthogonalisation using standard EOF analysis (Barnston, 1994) was performed on the predictor and the predictand field because of the large number of highly correlated variables and few observations contained in these fields. The predictor and predictand data sets were first standardised, so that the EOF pre-orthogonalisation was performed using the correlation matrices. The number of EOF modes to be retained in the CCA eigenanalysis was determined such that about 60% of the variance of both the predictand and predictor field is explained. This value of 60% is justified because 70% is the recommended threshold by the Guttman–Kaiser criterion (Jackson 1991), which normally over-selects the number of modes, and Jolliffe (1972) suggested a fraction of the number of modes suggested by this criterion. The truncation for the number of CCA modes retained was determined by using the Guttman–Kaiser criterion. The predictand fields, which are a combination of several 1-month fields, were separated after the prediction to obtain forecasts for each 1-month period contained in the combined predictand field. The predicted fields were interpolated to the GCM grid resolution of $3.75 \times 3.75^\circ$. The anomaly fields were smoothed using a grid–spectral-grid conversion to triangular truncation at 15 waves. Anomalies over land grid-points were set to zero for this purpose and areas of sea-ice were set to 271.4 K. Sea-surface temperature anomalies poleward of 45° were decayed exponentially to 0.5 K or less at 2 months. A comprehensive description of the sea-surface temperature model is found in Landman and Mason (2001).

3.3. The GCM

The GCM used is the T30 version of the Centre for Ocean–Land–Atmosphere Studies (COLA) spectral model with a resolution of roughly 400 km over the

study area (Kirtman et al., 1997). This corresponds to a horizontal grid of 96 by 48 points. The model has 18 layers in the vertical of unevenly spaced sigma-coordinate levels. Physical processes simulated by the model are deep and shallow convection, large-scale precipitation, radiation, surface physics, vertical diffusion and gravity wave drag. A simple biosphere model is included over land. Initial conditions for the GCM forecasts were derived from the NCEP re-analysis data (Kalnay et al., 1996) and boundary conditions consisted of predicted monthly sea-surface temperatures. The average of an ensemble of forecasts has better skill than a single control run (Brankovic et al., 1990), and each case in this study consisted of the average of five ensemble members generated using the lagged (24 h) average forecasting technique (Hoffman and Kalnay, 1983). In the absence of an independent model climatology, GCM mean biases were removed by subtracting the 3-month mean forecast errors in the remaining 7 years of the independent test period. The ideal would be to have a longer and completely independent GCM model climatology of 20–30 years, but for computational reasons it was not practical to build a sufficiently long climatology forced with hindcast sea-surface temperatures. Although Atmospheric Model Inter-comparison Project (AMIP)-style (Gates, 1992) model climatologies are available, because of the relatively weak variances of forecasts compared to observed sea-surface temperatures, the forecast atmospheric anomalies are likewise likely to be weak. While prognostic fields over the 8-year period cannot therefore be considered as truly independent, they do not use any information of the target season when the correction is made for that season. By removing only the GCM bias, it has been assumed that the variances of the GCM-simulated fields are correct. Because the target season is excluded when calculating the GCM's mean bias, the effect of not using an earlier independent climatology on forecast skill is expected to be minimal. Probably the strongest impact will be to underestimate the skill of the GCM because of the removal of any potentially predictable mean bias over the independent period.

3.4. Downscaling approach

Statistical interpretation of dynamical model output

to arrive at parameters such as rainfall or streamflow is essential (von Storch et al., 1993; Hewitson and Crane, 1996). Dynamical models represent the world as an array of grid-points, necessitating the parameterisation of important processes, such as sub-grid convection, that are important to local climate, but which may not be adequately simulated by GCM processes. Variables such as rainfall that are not accurately represented by these models, can be derived using statistical approaches to construct relationships between the desired forecast parameter and variables that are simulated more accurately. Owing to model imperfections, systematic errors may occur, but the statistical forecasts based on the dynamical model information can compensate for such biases. Furthermore, the GCM does not explicitly simulate streamflow, necessitating the statistical link between GCM-simulated fields and streamflow.

The perfect prognosis approach (Wilks, 1995) for downscaling dynamical model quantities was used in this paper. Using CCA (von Storch et al., 1993), observed (NCEP reanalysis) 850, 700, 500 and 200 hPa geopotential heights, 700 hPa relative humidity, 850–500 hPa thickness and 500–200 hPa thickness were related to the DJF seasonal streamflow at the inlets of the twelve dams of the Vaal and upper Tugela river systems in Fig. 1. These downscaling equations can be used to produce forecasts by applying them to the GCM-simulated atmospheric fields obtained by forcing the GCM with the predicted sea-surface temperatures at long (3-month) and short (1-month) lead-times. Usually, a perfect prognosis approach makes no attempt to compensate for systematic errors of the dynamical model and assumes that the dynamical model quantities are perfect. Notwithstanding, bias corrections can be, and were, applied before use in the perfect prognosis equations.

In designing the optimal CCA downscaling model, EOF analysis was performed on the predictor and predictand sets. The number of modes to be retained in the CCA eigenanalysis problem was determined using cross-validated (Barnston and van den Dool, 1993) forecast skill sensitivity tests. The sensitivity tests were conducted with a varying number of retained predictor and predictand EOF modes. The optimisation of the equations was done in such a way that the combination of predictor and predictand

EOF modes produced the highest mean cross-validation correlations in the CCA model. The truncation for the number of CCA modes retained was determined by using the Guttman–Kaiser criterion (Jackson, 1991). Two CCA modes were retained.

With high cross-validated and statistically significant correlations, it should prove possible to achieve good forecast skill, provided that the GCM-simulated circulation and moisture fields are accurate. Due to model deficiencies in both the COLA T30 GCM (Xue and Shukla, 1998) and the sea-surface temperature model (Landman and Mason, 2001), perfect prognostic fields are unlikely. Notwithstanding, if there is a sea-surface temperature signal in the climate, a well designed GCM forced with skilful sea-surface temperature forecasts should lead to significant reproducibility of climate anomalies among ensemble members and hence predictability (Ward and Navarra, 1997). It is this notion that remains to be tested for the Vaal and upper Tugela catchments.

3.5. Estimating forecast skill

There are several methods for determining model output performance (Wilks, 1995). To estimate the expected forecast performance of the perfect prognosis equations, the method of cross-validation was employed. For cross-validation, the value that is predicted, is omitted from the training period. Cross-validation was performed by removing one year at a time from the training period, and was employed to:

1. estimate the optimal number of combined DJF circulation and moisture fields empirical orthogonal function (EOF) (Jackson, 1991) modes and DJF naturalised streamflow EOF modes for the equations relating circulation and moisture to naturalised streamflow by maximizing the cross-validated skill; and
2. bias-correct the GCM prognostic fields by removing the mean of the GCM simulations during 7 years of the 8-year independent period from the simulation of the remaining year.

However, cross-validation may still produce biased skill levels (Barnston and van den Dool, 1993). In order to obtain skill levels that are unbiased, model

Table 1

Cross-validated correlations between NCEP reanalysis circulation and moisture data at certain standard pressure levels and naturalised streamflow, using a CCA model trained over each of the respective training periods indicated in the table. Correlations significant at the 95% level of confidence are marked with an asterisk

Dam	1951/1952–1986/1987	1951/1952–1989/1990	1951/1952–1992/1993
Allemanskraal	0.45*	0.43*	0.45*
Bloemhof	0.44*	0.50*	0.53*
Boskop	0.53*	0.37*	0.40*
Erfenis	0.39*	0.41*	0.44*
Grootdraai	0.24	0.36*	0.37*
Koppies	0.46*	0.41*	0.44*
Johan Nesor	0.51*	0.51*	0.53*
Sterkfontein	0.25	0.27	0.32*
Spioenkop	0.40*	0.45*	0.49*
Driel	0.40*	0.45*	0.49*
Vaal	0.38*	0.48*	0.51*
Barrage	0.38*	0.47*	0.50*

validation should be conducted over a test period that is independent of the training period. This method of model validation is referred to as retro-active forecasting and involves the evaluation of predictions compared to observations excluding any information following the target year. The procedure is designed to reproduce forecasts as they would have been in an operational environment, and can be explained by considering a retro-active forecast procedure for DJF streamflow during the 8-year period 1987/1988 to 1994/1995. The perfect prognosis model was first trained considering information leading up to and including 1986/1987. Three consecutive years were predicted using this trained model, and predictions of DJF streamflow for 1987/1988, 1988/1989 and 1989/1990 were made. The model was subsequently retrained using data up to and including 1989/1990 to predict 1990/1991, 1991/1992 and 1992/1993. This procedure was continued until the 1994/1995 DJF streamflow was predicted using a model trained with the data up to 1992/1993. For each dam, the categorized forecast was compared to the observed to calculate the model skill. The scoring methods used for the retro-active categorized streamflow forecasts were the Linear Error in Probability Space (LEPS) scores (Ward and Folland, 1991), and the hit scores (the number of times a correct category is forecast).

3.6. Significance testing

In order to see if the LEPS scores are statistically

significant, a Monte Carlo test (Wilks, 1995) was performed. Above-normal, near-normal and below-normal streamflow categories for both a predicted and observed streamflow set were generated, and the LEPS score was calculated. This procedure was repeated 1000 times. For the LEPS scores, the 90% confidence level is associated with about 38%, the 95% level with about 49% and the 99% level with about 68%. The significance of the hits scores was calculated from a binomial distribution (von Storch and Zwiers, 1999). For the hit scores the 90% confidence level is associated with 4 hits, the 95% level with 5 hits and the 99% level with 6 hits out of a possible 8.

4. Hydrology forecasts

Before assessing the skill of the retro-active GCM-derived categorized streamflow forecasts, statistical relationships between streamflow and observed atmospheric variability were identified to assess the feasibility of generating categorized streamflow forecasts directly from atmospheric variables simulated by the GCM. Table 1 contains the cross-validation correlations of the optimal models between the observed and predicted naturalised streamflows at the inlets of the 12 dams within the Vaal and upper Tugela catchments. Most of the correlations are statistically significant at the 95% level of confidence. Fig. 2 illustrates the cross-validation forecast time

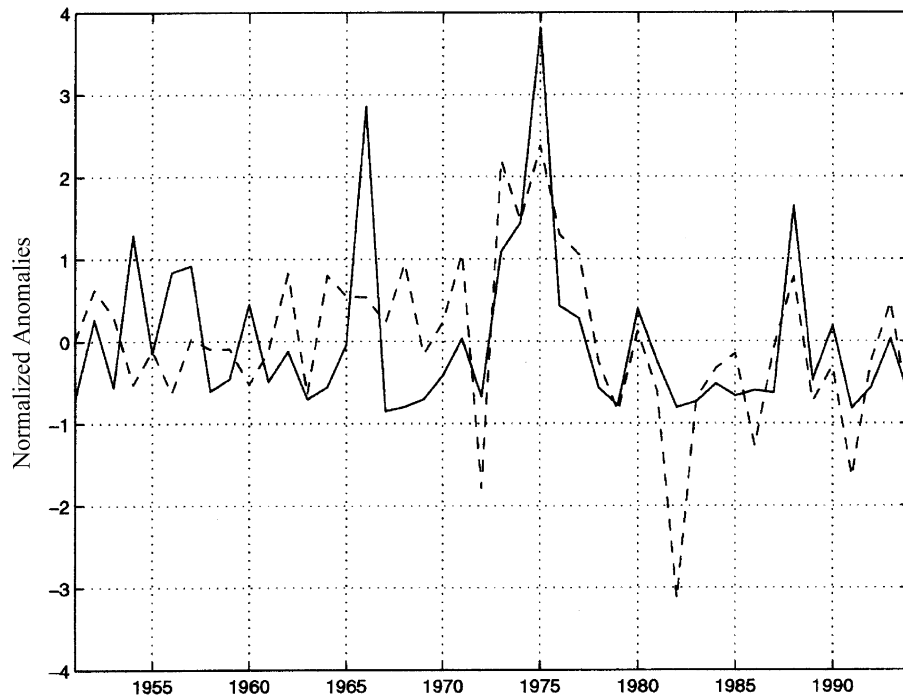


Fig. 2. Cross-validation prediction of naturalised streamflow of the Johan Nesper Dam catchment. The solid line represents the normalized streamflow anomalies, and the dashed line the predicted normalized anomalies.

series for streamflow to the Johan Nesper Dam using a 44-year training period. Extreme anomalies are evident. Some high streamflow years are associated with La Niña events, and some low streamflow years with El Niño events. Some years of high (low) streamflow are not associated with La Niña (El Niño) events. Seemingly poor forecast skill is found during the early parts of the training period leading up to about the late-1960s and early 1970s. The predictability improved towards the second half of the training period. The reason for this improvement in predictability is not an artefact of the naturalisation of streamflows, because most of the naturalised streamflows were verified against observations made during the early years of the training period.

4.1. Limitations of categorized streamflow atmosphere-driven forecast skill

Although most of the cross-validation correlations of Table 1 are significant at the 95% level of confidence, the skill of the perfect prognosis equations

described in the previous section has to be tested over a retro-active period to verify its operational utility. Assuming that NCEP reanalysis data are equal to observed data, using NCEP data in the perfect prognosis equations instead of GCM-simulated fields should indicate the upper limit to categorized streamflow prediction skill by incorporating only atmospheric variability in the perfect prognosis equations. By being able to assess the maximum obtainable skill level from this particular downscaling approach, a maximum skill level is set against which other prediction approaches can be benchmarked over the same retro-active forecast period. Fig. 3 represents the LEPS scores calculated from the categorized (below-normal, near-normal and above-normal) naturalised streamflow predictions over the 8-year retro-active period using NCEP reanalysis data as predictors. High skill is found for Grootdraai, Johan Nesper, Spioenkop, Driel and Bloemhof Dams, with LEPS scores significant at the 95% level of confidence and hit scores of at least 4 out of 8. A LEPS score significant at the 99% level of confidence is found for

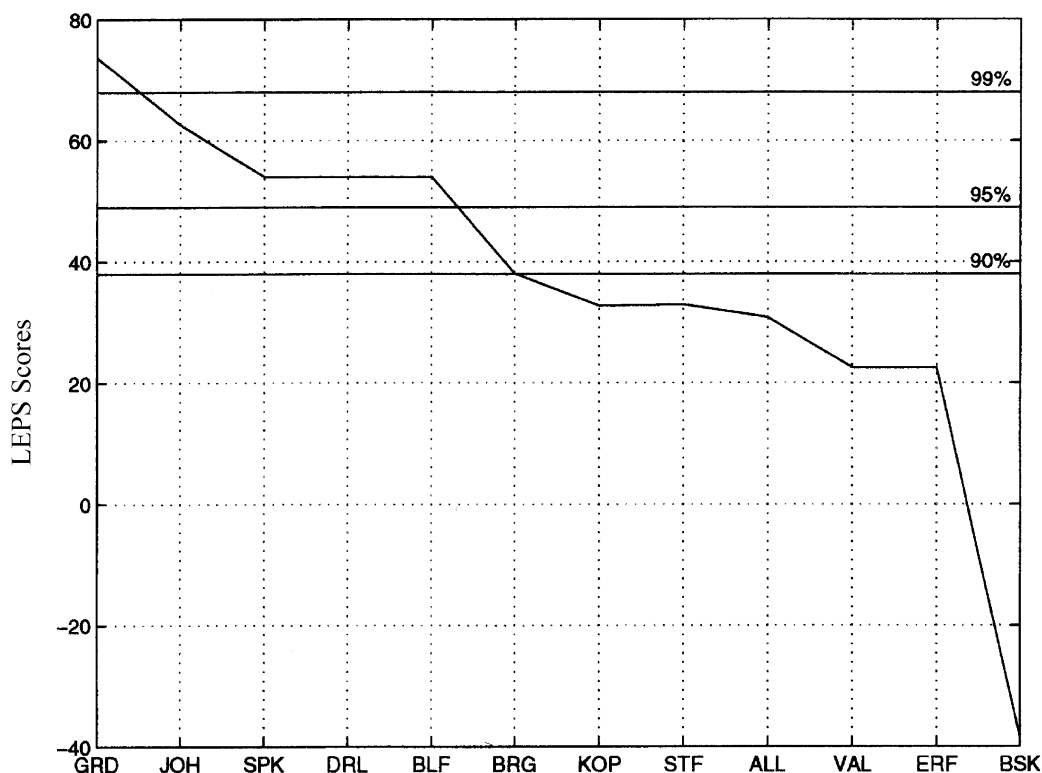


Fig. 3. LEPS scores of the retro-active forecasts of naturalised categorized streamflow for the twelve dams (GRD: Grootdraai; JOH: Johan Naser; SPK: Spioenkop; DRL: Driel; BLF: Bloemhof; BRG: Barrage; KOP: Koppies; STF: Sterkfontein; ALL: Allemanskraal; VAL: Vaal; ERF: Erfenis; BSK: Boskop) using NCEP reanalysis data as predictors in the perfect prognosis equations. The horizontal lines are the 90, 95 and 99% confidence levels, respectively.

Grootdraai Dam with a hit score of 5 out of 8. Poor forecast skill is found for the categorized streamflows of the catchments of Boskop Dam.

A significant difference in skill is found between forecasts for Johan Naser and Boskop Dams (Fig. 3), which are in close proximity of each other (Fig. 1). Because of their closeness, the same large-scale atmospheric variability should contribute to the predictability of the categorized streamflows into these dams. However, the association between the naturalised streamflows in the catchments of these dams is weak (correlation of 0.39, as opposed to the correlation of 0.91 between Johan Naser and Bloemhof Dams, for example). Streamflow is not only affected by precipitation, but also by water storage of soil moisture and groundwater, which is influenced by vegetation and soil surface characteristics of catchments (Schulze, 1997). Ignoring

possible data quality problems of the naturalised streamflows, the reason for the different forecast skill levels of Boskop and Johan Naser Dams is a result of differences in the surface characteristics of the two catchments. The canonical vectors (Fig. 4) of the predictors (circulation and moisture at standard pressure levels) of the first CCA mode associated with the three training periods considered in designing the perfect prognosis equations, show time series reminiscent of ENSO. Maximum time scores are found during the La Niña events of 1973/1974, 1975/1976 and 1988/1989, and minimum scores associated with the El Niño events of 1972/1973, 1982/1983, 1986/1987 and 1991/1992. The association between the canonical vectors and the streamflow of the Johan Naser Dam catchment (about 0.7) is higher than found for Boskop Dam (about 0.3). The streamflow variability within the Johan Naser Dam

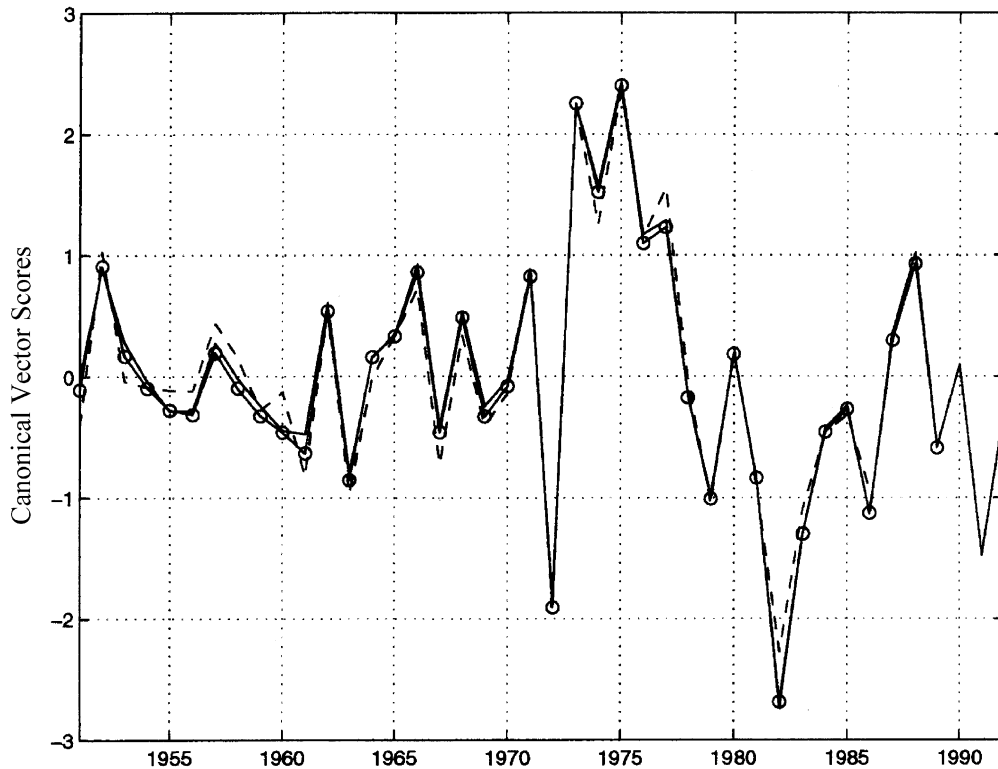


Fig. 4. Canonical vectors of the first CCA mode for the 36- (dashed line), 39- (solid circled line) and 42-year (solid line) perfect prognosis training periods.

Table 2

Cross-validated correlations between NCEP reanalysis circulation and moisture data at certain standard pressure levels and naturalised stream-flow, using a CCA model trained over each of the respective *shortened* training periods indicated in the table. Correlations significant at the 95% level of confidence are marked with an asterisk

Dam	1967/1968–1986/1987	1967/1968–1989/1990	1967/1968–1992/1993
Allemankraal	0.73*	0.64*	0.66*
Bloemhof	0.70*	0.75*	0.74*
Boskop	0.66*	0.45*	0.45*
Erfenis	0.76*	0.68*	0.69*
Grootdraai	0.49*	0.59*	0.53*
Koppies	0.63*	0.55*	0.57*
Johan Nesper	0.80*	0.76*	0.79*
Sterkfontein	0.72*	0.62*	0.66*
Spioenkop	0.74*	0.61*	0.68*
Driel	0.74*	0.61*	0.68*
Vaal	0.61*	0.68*	0.66*
Barrage	0.60*	0.69*	0.67*

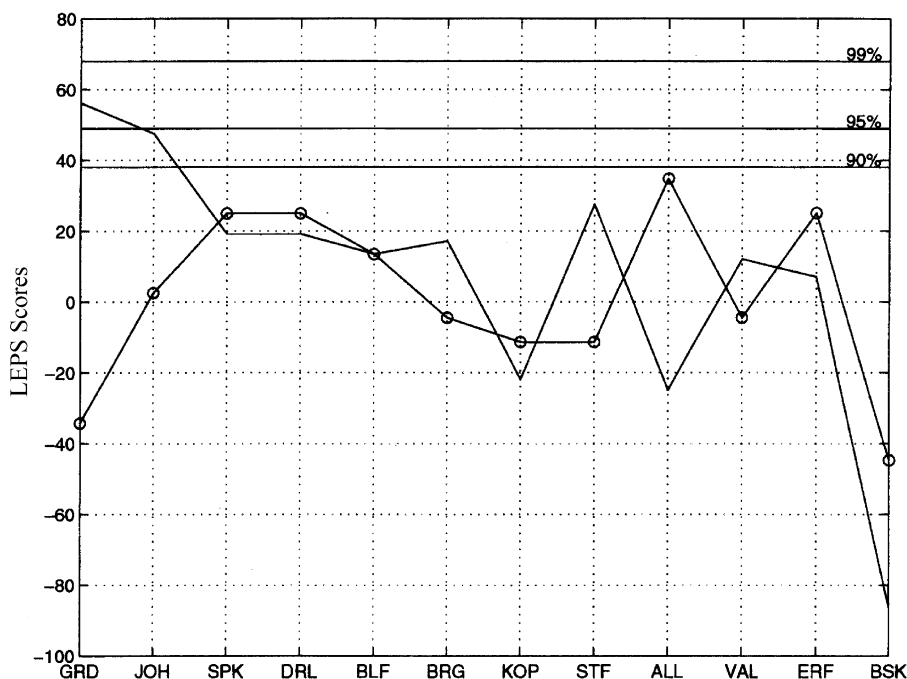


Fig. 5. LEPS scores of the retro-active naturalised categorized streamflow forecasts for the twelve dams (GRD: Grootdraai; JOH: Johan Neser; SPK: Spioenkop; DRL: Driel; BLF: Bloemhof; BRG: Barrage; KOP: Koppies; STF: Sterkfontein; ALL: Allemanskraal; VAL: Vaal; ERF: Erfenis; BSK: Boskop) with a 1-month lead-time. The solid line represents the scores associated with the longer training period, and the circled line represents the scores associated with the shorter training period. The horizontal lines are the 90, 95 and 99% confidence levels respectively.

catchment is thus best explained by ocean-atmospheric interactions, as opposed to the streamflow variability of the Boskop Dam catchment that is more affected by land surface conditions. Although further research is needed to obtain a better understanding of the relationships between streamflow and atmospheric variability, the differences of the streamflows of these two catchments are typical examples where streamflows of two adjacent catchments have been affected by factors not only associated with large-scale atmospheric variability. These differences emphasise the need to include additional variables such as evaporation and groundwater storage into the perfect prognosis equations when downscaling to streamflow.

4.2. Retro-active categorized streamflow forecasts

A weaker association exists between the first half (up to the 1970s) of the cross-validated perfect prognosis forecasts in comparison to the second half

(from the early 1970s to the end of the training period). The weakening is illustrated for the Johan Neser Dam catchment (Fig. 2). Because of the weaker predictability during the first half of the training period, it was decided to construct prognostic equations for a shortened training period from 1967/1968 to 1994/1995 to exclude most of the first half of the original training period. New perfect prognosis equations were set up over the remaining 20, 23 and 26 years respectively. Although the shortened training periods are possibly too short for statistical model training (Barnston and Ropelewski, 1992), optimal models were designed and the cross-validation correlations are shown in Table 2. All of these correlation values are statistically significant at the 95% level of confidence. LEPS scores graphs, obtained from considering categorized streamflow predictions are shown in Figs. 5 and 6. The scores for 1-month lead-times (Fig. 5) are shown for both the longer (36, 39 and 42 years) and the shorter (20, 23 and 26 years) training periods. Scores using the shorter

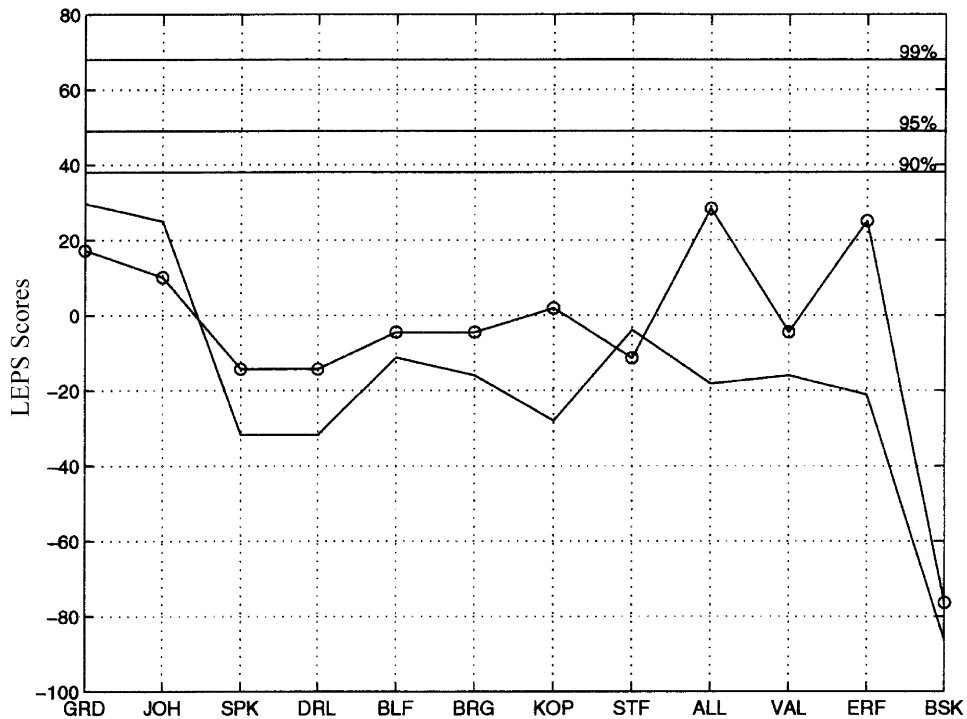


Fig. 6. As for Fig. 5, but for a 3-month lead-time.

training period (line and circle) are marginally better for six catchments, but none of these scores for the shorter training period are statistically significant at the 90% level of confidence. For the 3-month lead

forecasts (Fig. 6), none of the LEPS scores are significant at the 90% level of confidence. The cross-validation correlations over the shorter training periods are higher than those obtained for the longer periods.

Table 3

Categorized retro-active 1-month lead streamflow forecasts (lower case) and observations (upper case) for the twelve catchments in the Vaal and upper Tugela river systems (A and a refer to above-normal; N and n refer to near-normal; B and b refer to below-normal). Numbers in brackets are the number of hits associated with neither an El Niño nor a La Niña event. Correctly forecast categories are shown in bold

Dam	La Niña			El Niño			El Niño		Hit score
	1987/1988	1988/1989	1989/1990	1990/1991	1991/1992	1992/1993	1993/1994	1994/1995	
Allemandskraal	b A	a A	a B	n A	n B	b N	b N	n B	1(0)
Bloemhof	b N	n A	n B	n N	n B	b B	b A	b B	3(2)
Boskop	a B	a B	a N	a N	a B	n B	a B	a B	0(0)
Erfenis	b A	a A	n B	n A	n B	b B	b N	n B	2(1)
Grootdraai	b N	b N	b N	b N	b B	b B	b N	b B	3(1)
Koppies	b A	a A	n B	n A	n B	b N	b A	n B	1(0)
Johan Nesper	b B	a A	n N	n N	n B	n N	b N	n B	5(4)
Sterkfontein	b B	a A	a B	a A	a B	b B	b N	n B	4(3)
Spioenkop	b N	a A	a B	a A	a B	b B	n N	n B	4(3)
Driel	b N	a A	a B	a A	a B	b B	n N	n B	4(3)
Vaal	b A	n A	b B	b N	n B	b B	b A	b B	3(2)
Barrage	b A	n A	b B	n N	n B	b B	b A	b B	4(3)
Hits	2	7	3	6	1	9	2	4	

Notwithstanding, the higher correlations are not reflected in the skill levels that can be expected in an operational forecast environment. The longer period LEPS scores are significant for certain catchments while none are for the shorter period, which shows the importance of having a longer training set for statistical models, probably in excess of 30 years (Barnston and Ropelewski, 1992) to improve on the confidence of operational forecasts.

Table 3 represents a summary of the 1-month lead-time categorized naturalised retro-active streamflow forecasts for the period 1987/1988 to 1994/1995. The categorized streamflows during DJF of 1988/1989 (La Niña) and 1992/1993 were forecast accurately. Categorized streamflows into half of the dams were predicted accurately during 1990/1991. However, poor forecasts were produced for the categorized streamflows during the El Niño events, especially for 1991/1992 when only one hit is obtained, and the forecasts for four catchments are two categories out. No two-category misses are found for the forecast of categorized streamflow during the 1994/1995 El Niño event, when, apart from Boskop Dam, the forecasts were for either near-normal or below-normal streamflow.

Using the NCEP data as predictors in the perfect prognosis equations for the two El Niño events (1991/1992 and 1994/1995) produced high skill levels. For all the streamflows, below-normal was observed and below-normal was forecast. Forcing the GCM with observed sea-surface temperatures (reminiscent of perfect forecast fields) instead of the forecast fields, produced below-normal forecasts throughout for the streamflow during the 1994/1995 El Niño event when below-normal streamflow are observed. However, forecasts made for the El Niño event of 1991/1992 using observed sea-surface temperatures in the GCM, produced only three hits, albeit higher than the single hit obtained from the GCM forced with the predicted sea-surface temperature fields. These results suggest, as was found in downscaling to rainfall (Landman et al., 2001), that the prediction skill should improve if the GCM is forced with highly skilful sea-surface temperature forecasts that have the ability to capture the large sea-surface temperature anomalies observed during strong ENSO events. The perfect predictors (NCEP data) produced higher accuracy forecasts for 1991/1992 than what was

obtained using observed sea-surface temperatures, which may suggest that the GCM has failed in producing skilful circulation and moisture during that particular year.

For 3-month lead forecasts, skill levels are lower than those of the 1-month lead forecasts (compare Figs. 5 and 6). The categorized streamflow forecasts during the La Niña event of 1988/1989 had five hits, as did the forecasts of the following year (1989/1990), and poor forecast accuracy is found for the El Niño events, with three hits and one hit for the respective El Niño events of 1991/1992 and 1994/1995. However, forcing the GCM with persisted August sea-surface temperature anomalies instead of the forecast fields, produced accurate categorized streamflow forecasts for the El Niño event of 1994/1995 with six hits and just one forecast with two categories out, but poor categorized streamflow forecasts for 1991/1992 are produced.

5. Summary and discussion

A multi-tiered forecast scheme for predicting categorized streamflow has been constructed and tested for use in an operational environment. This approach to forecasting for three categories is becoming a commonly used method in seasonal forecasting. Considering the skill level of these forecasts, categorical and continuous forecasts are not very different from one another in their ultimate utility, because non-categorical forecasts require error bars. Thus, when skill is only moderate, changing the format of the forecast does not overcome the problem of uncertainty (Barnston, personal communications).

Altogether 12 dams are identified in the Vaal and upper Tugela river systems that have a sufficiently long training period on which the scheme could be implemented and tested. A retro-active forecast period of eight years from 1987/1988 to 1994/1995 has been used. The method is a multi-tiered scheme involving statistical forecasts of near-global sea-surface temperatures, which are used to provide boundary conditions for a GCM. Simulations of geopotential heights and moisture fields over southern Africa from an ensemble of integrations of the GCM are then used to derive categorized streamflow forecasts using a perfect prognosis technique.

Before forcing the GCM with the sea-surface temperature forecasts, and then downscaling the simulated fields to categorized streamflow, the potential predictability of categorized streamflow forecasts using this procedure has been estimated, by assessing the skill achieved given perfect forecasts of atmospheric variability. A potential for highly skilful forecasts was found for the categorized streamflow of five of the twelve catchments. This high potential shows the strong relationship between streamflow and atmospheric variability, and hence suggests that the multi-tiered forecasting system described in this paper may be able to provide useful skill. Low skill was found for the Vaal Dam categorized streamflow, but high skill scores were found for Grootdraai Dam upstream of the Vaal Dam. Grootdraai Dam has a much smaller catchment than Vaal Dam, and so is presumably affected less by non-atmospheric variables such as groundwater storage. However, high skill scores were found for Bloemhof Dam categorized streamflow, which has the largest catchment of all the dams considered in this study. Evidently no simple relationship exists between the strength of a climate signal on streamflow and catchment size. Larger catchments may have a stronger climate signal because they are less affected by local climate noise, but at the same time the effects of surface characteristics on streamflow may be greater than in smaller catchments. Other non-atmospheric factors, which are related to the surface characteristics of a particular catchment, affect the variability of streamflows, but are not considered in the downscaling system presented here. Notwithstanding, useful categorized streamflow forecasts were obtained for some of the catchments. A more involved forecast scheme that uses GCM predictions of climate variables to generate estimates of the climate inputs into a calibrated rainfall-runoff model to estimate likely streamflow should be tested. Such a scheme would then implicitly allow for different soil, geological and land cover characteristics. However, the forecast scheme presented here represents a simpler downscaling algorithm that is relatively cheap to run and could also be used to set a baseline skill level that must be exceeded to demonstrate the usefulness of more complex downscaling models.

When using the GCM-derived forecasts of atmospheric fields over southern Africa, and downscaling

these to categorized streamflow, accurate forecasts for three years of the retro-active period have been made (1988/1989, 1990/1991 and 1992/1993), so demonstrating the scheme's potential operational utility for short lead-times. However, to improve on the confidence of operational categorized streamflow forecasts, a longer training set is required, and more rigorous skill assessments must be conducted. By improving GCM simulations, the potential to make highly skilful categorized streamflow forecasts over relatively short lead-times has been demonstrated, especially if the GCM is supplied with highly skilful sea-surface temperature forecasts that capture the large sea-surface temperature anomalies during strong ENSO events. Notwithstanding the reasonably accurate forecasts for some of the years in the retro-active period and at some of the inflows, there is at present still a gap between demonstrated forecast skill and the application of operational forecasts by dam managers. In addition, forecasts for decision-making purposes should be presented in a probabilistic manner, and not deterministically as was done here.

Further research into the predictability of categorized streamflow in South Africa using a multi-tiered scheme is warranted. The skill of the scheme could be increased by improving the skill of the sea-surface temperature forecasts and by using improved hydrological parameterisations in the GCMs to account for surface characteristics, soil moisture conditions and other factors affecting streamflow. Further efforts to understand how seasonal predictability is related to spatial-scale, to improve GCMs and their forcing fields, and to use non-linear downscaling equations such as Neural Nets (Shamseldin, 1997) or dynamical downscaling through limited area models (Mason et al., 1999), should be investigated to improve streamflow forecast skill over lead-times of several months.

Acknowledgements

This research project was sponsored by Eskom Technology Research and Investigations and administered by the Water Research Commission. Additional funding was provided by a grant/cooperative agreement from the National Oceanic and Atmospheric Administration (NOAA). The views expressed herein are those of the authors and do not

necessarily reflect the views of NOAA or any of its sub-agencies. The utilization of the South African Weather Bureau's computer resources is acknowledged with gratitude

References

- Barnston, A.G., 1994. Linear statistical short-term climate predictive skill in the Northern Hemisphere. *J. Clim.* 7, 1513–1564.
- Barnston, A.G., Ropelewski, C.F., 1992. Prediction of ENSO using canonical correlation analysis. *J. Clim.* 5, 1316–1345.
- Barnston, A.G., van den Dool, H.M., 1993. A degeneracy in cross-validated skill in regression-based forecasts. *J. Clim.* 6, 963–977.
- Barnston, A.G., Van den Dool, H.M., Zebiak, S.E., Barnett, T.P., Ji, M., Rodenhuis, D.R., Cane, M.A., Leetmaa, A., Graham, N.E., Ropelewski, C.R., Kousky, V.E., O'Lenic, E.A., Livezey, R.E., 1994. Long-lead seasonal forecasts — where do we stand? *Bull. Am. Meteorol. Soc.* 66, 2097–2114.
- Bengtsson, L., Schlese, U., Roeckner, E., Latif, M., Barnett, T.P., Graham, N.E., 1993. A two-tiered approach to long-range climate forecasting. *Science* 261, 1026–1029.
- Brankovic, C., Palmer, T.N., Molenti, F., Tibaldi, S., Cubacsh, U., 1990. Extended range predictions with the ECMWF models: time lagged ensemble forecasting. *Q. J. R. Meteorol. Soc.* 116, 867–912.
- Carson, D.J., 1998. Seasonal forecasting. *Q. J. R. Meteorol. Soc.* 124, 1–26.
- Compagnucci, R.H., Vargas, W.M., 1998. Inter-annual variability of the Cuyo ivers' streamflow in the Argentinean Andean mountains and ENSO events. *Int. J. Climatol.* 18, 1593–1609.
- Cui, M., Von Storch, H., Zorita, E., 1995. Coastal sea level and the large-scale climate state A downscaling exercise for Japanese Islands. *Tellus* 47A, 132–144.
- Gates, L.W., 1992. AMIP: the Atmospheric Model Intercomparison Project. *Bull. Am. Meteorol. Soc.* 73, 1962–1970.
- Hewitson, B.C., Crane, R.G., 1996. Climate downscaling: techniques and application. *Clim. Res.* 7, 85–95.
- Hoffman, R.N., Kalnay, E., 1983. Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus* 35A, 100–118.
- Hunt, B.G., Zebiak, S.E., Cane, M.E., 1994. Experimental predictions of climate variability for lead times of twelve months. *Int. J. Climatol.* 14, 507–526.
- Jackson, J.E., 1991. *A User's Guide to Principal Components*. Wiley, New York 569pp.
- Jolliffe, I.T., 1972. Discarding variables in principal component analysis. I: Artificial data. *Appl. Statist.* 21, 160–173.
- Joubert, A.M., Hewitson, B.C., 1997. Simulating present and future climates of southern Africa using general circulation models. *Prog. Phys. Geogr.* 21, 51–78.
- Jury, M.R., Mulenga, H.M., Mason, S.J., 1999. Exploratory long-range models to estimate summer climate variability over southern Africa. *J. Clim.* 12, 1892–1899.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D., 1996. The NCEP/NCAR 40-year re-analysis project. *Bull. Am. Meteorol. Soc.* 77, 437–471.
- Kirtman, B.P., Shukla, J., Huang, B., Zhu, Z., Schneider, E.K., 1997. Multiseasonal predictions with a coupled tropical ocean–global atmosphere system. *Mon. Weath. Rev.* 125, 789–808.
- Landman, W.A., Mason, S.J., 1999. Operational long-lead prediction of South African rainfall using canonical correlation analysis. *Int. J. Climatol.* 19, 1073–1090.
- Landman, W.A., Mason, S.J., 2001. Forecasts of near-global sea surface temperatures using canonical correlation analysis. *J. Clim.* (in press).
- Landman, W.A., Mason, S.J., Tyson, P.D., Tennant, W.J., 2001. Retro-active skill of multi-tiered forecasts of summer rainfall over southern Africa. *Int. J. Climatol.* 21, 1–19.
- Makarau, A., Jury, M.R., 1997. Predictability of Zimbabwe summer rainfall. *Int. J. Climatol.* 17, 1421–1432.
- Mason, S.J., 1998. Seasonal forecasting of South African rainfall using a non-linear discriminant analysis model. *Int. J. Climatol.* 18, 147–164.
- Mason, S.J., Jury, M.R., 1997. Climate variability and change over southern Africa: a reflection on underlying processes. *Prog. Phys. Geogr.* 21, 23–50.
- Mason, S.J., Tyson, P.D., 1999. The occurrence and predictability of drought over southern Africa. In: Wilhite, D.A. (Ed.). *Drought: A Global Assessment*. Routledge, New York, pp. 113–134.
- Mason, S.J., Goddard, L., Graham, N.E., Yelavea, E., Sun, L., Arkin, P.A., 1999. The IRI seasonal climate prediction system and the 1998 El Niño event. *Bull. Am. Meteorol. Soc.* 80, 1853–1873.
- Moura, A. D., Bengtsson, L., Buizer, J., Busalacchi, A., Cane, M. A., Lagos, P., Leetmaa, A., Matsuno, T., Mooney, K., Morel, P., Sarachik, E. S., Shukla, J., Sumi, A., Patterson, M., 1992. International Research Institute for Climate Prediction: A Proposal. 1100 Wayne Avenue, Suite 1225, Silver Spring, MD 20910, USA, 51pp.
- Palmer, T.N., Anderson, D.L.T., 1994. The prospects for seasonal forecasting. *Q. J. R. Meteorol. Soc.* 120, 755–793.
- Reason, C.J.C., Lutjeharms, J.R.E., 1998. Variability of the south Indian Ocean and implications for southern African rainfall. *S. Afr. J. Sci.* 94, 115–123.
- Reynolds, R.W., Smith, T.M., 1994. Improved global sea surface temperature analysis using optimum interpolation. *J. Clim.* 7, 929–948.
- Schulze, R.E., 1997. Impacts of global climate change in a hydrologically vulnerable region: challenges to South African hydrologists. *Prog. Phys. Geogr.* 21, 113–136.
- Shamseldin, A.Y., 1997. Application of a neural network technique to rainfall-runoff modelling. *J. Hydrol.* 199, 272–293.
- Smith, T.M., Reynolds, R.W., Livezey, R.E., Stokes, D.C., 1996. Reconstruction of historical sea surface temperatures using empirical orthogonal functions. *J. Clim.* 9, 1403–1420.
- Triegaardt, D. O., Van Heerden, J. and Steyn, P. C. L., 1991.

- Anomalous precipitation and floods during February 1988. South African Weather Bureau Technical Paper, No 23, 25pp.
- Von Storch, H., Zwiers, F.W., 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge 484pp.
- Von Storch, H., Zorita, E., Cubasch, U., 1993. Downscaling of global climate change estimates to regional scales: An application to Iberian rainfall in wintertime. *J. Clim.* 6, 1161–1171.
- Ward, M.N., Folland, C.K., 1991. Prediction of seasonal rainfall in the North Nordeste of Brazil using eigenvectors of sea-surface temperature. *Int. J. Climatol.* 11, 711–743.
- Ward, M.N., Navarra, A., 1997. Pattern analysis of SST-forced variability in ensemble GCM simulations: examples over Europe and the tropical Pacific. *J. Clim.* 10, 2210–2220.
- Wilks, D.S., 1995. *Statistical Methods in the Atmospheric Sciences*. Academic Press, San Diego 467pp.
- Xue, Y., Shukla, J., 1998. Model simulation of the influence of global SST anomalies on Sahel rainfall. *Mon. Weath. Rev.* 126, 2782–2792.