Theor. Appl. Climatol. 60, 21–36 (1998)

Theoretical and Applied Climatology

National Institute of Water and Atmospheric Research Ltd., Wellington, New Zealand

A Regression-based Assessment of the Predictability of New Zealand Climate Anomalies

R. I. C. C. Francis and **J. A. Renwick**

With 8 Figures

Received July 18, 1997
Revised April 2, 1998

Summary

Estimates of the predictability of New Zealand monthly and seasonal temperature and rainfall anomalies are calculated using a cross-validated linear regression procedure. Predictors are indices of the large scale circulation, sea-surface temperatures, the Southern Oscillation Index and persistence. Statistical significance is estimated through a series of Monte Carlo trials. No significant forecast relationships are found for rainfall anomalies at either the monthly or seasonal time scale. Temperature forecasts are however considered to exhibit significant skill, with variance reductions of the order of 10–20% in independent trials. Temperature anomalies are most skilfully predicted over the North Island, and skill is greatest in Spring and Summer in most areas. At the monthly time scale, predictors local to the New Zealand region account for most of the forecast skill, while at the seasonal time scale, skill depends strongly upon "remote" predictors defined over regions of the southern hemisphere distant from New Zealand. Indices of meridional flow over the Tasman Sea/New Zealand region are found to be useful predictors, especially for monthly forecasts, perhaps as a proxy for atmospherically-forced sea surface temperature anomalies. Sea surface temperature anomalies to the west of New Zealand and in the tropical Indian Ocean are also useful, especially for seasonal predictions. Forecast skill is more reliably estimated at the monthly time scale than at the seasonal time scale, as a result of the larger sample size of monthly mean data. While long-term mean levels of skill may be estimated reliably over the whole data set, statistically significant decadal-scale variations are found in the predictability of temperature anomalies. Therefore, even if long-term forecast skill levels are reliably estimated, it may be impossible to predict the short-term skill of operational seasonal climate forecasts. Implications for operational climate predictions in mid-latitudes are discussed.

1. Introduction

Recent advances in understanding the physics of the El Niño/Southern Oscillation (ENSO) phenomenon and its effect upon the global circulation have generated considerable interest in the prospect of routine seasonal to interannual "climate predictions" (e.g., predictions of monthly- or seasonal-mean temperature and rainfall anomalies), for at least some regions of the globe. Much international effort is now going into the development of operational seasonal forecasting products (e.g., World Meteorological Organization, 1995). The ultimate goal of research into short-term climate variability and predictability is model-based prediction of climate on intraseasonal to interannual time scales (CLIVAR Scientific Steering Group, 1995). However, there is presently much scope for empirical prediction procedures, since the physical basis for much of the observed global climate variability has yet to be fully understood (Lau, 1997).

Empirical studies of seasonal climate prediction generally focus on relating surface climate parameters to indices of the large-scale circulation and boundary forcings (e.g., upper-level geopotential height fields, sea-surface temperature fields, etc.). The underlying assumption is that there are features of the mean large scale circulation which are predictable for periods much longer than those of individual weather events (Palmer and Anderson, 1994). Such circulation features include: ENSO forcing, sea-surface temperature (SST) forcing, low-frequency variability of the atmospheric circulation, and persistence effects in local climate variables. Seasonal predictability is strongly related to the level of persistence (serial correlation) of climate anomalies. Persistence is partly a manifestation of local boundary forcings, such as soil moisture content, snow cover and local SST anomalies, and is also related to much larger-scale low-frequency forcings, typically associated with ENSO and other tropical SST variability.

A number of studies have investigated the persistence, predictability, or potential predictability, of the New Zealand (NZ) climate, usually defined in terms of monthly temperature and rainfall anomalies. The most extensive study of the persistence of NZ temperatures (Goulter, 1984) found the highest serial correlations in the warm half of the year, over northern regions, at the seasonal (120d) time scale. Indices of the strength of the zonal flow in the NZ region have also been found to exhibit seasonal persistence between Spring and Autumn (Kidson and Barnes, 1984). ENSO-related SST and circulation variability is known to be significantly correlated with mean sea-level pressure (MSLP) anomalies in the NZ region and to monthly mean temperature and rainfall in some regions of the country, notably in Spring and Autumn (Gordon, 1986; Mullan, 1995). Analysis of the influence of SST on NZ climate (Trenberth, 1975; Basher and Thompson, 1995; Mullan, 1998) suggests that local SST variability is generally forced by the atmosphere, as is typical of mid-latitudes. Despite this, the possibility exists for remote SST forcing of the region's circulation anomalies, as reported by Nicholls (1989), Mullan and Renwick (1996) and Mullan (1998).

Based on the research described above, the generation of useful monthly or seasonal forecasts of NZ climate anomalies seem possible in principle. Studies of potential predictability lend support to this notion. Frequency-domain analysis suggests that, at many locations, up to 50% of the variance in NZ mean temperatures is potentially predictable on the seasonal time scale and approximately 30% is potentially predictable on the monthly scale (Madden and Kidson, 1997). However, studies attempting to synthesise the effects on NZ climate of many of the forcings discussed above (Kidson and Gordon, 1986; Mullan and Renwick, 1996) have concluded that no more than around 20% of the variance of monthly or seasonal temperature and rainfall anomalies is predictable by linear statistical means. Kidson (1988) found little month-to-month predictability in time series of low-frequency modes of the Southern Hemisphere circulation, concluding that the prospects for long-range forecasting in the Southern Hemisphere are not bright.

This paper estimates the skill of prediction of NZ climate anomalies on monthly and seasonal time scales, through the use of linear statistical techniques and a limited set of predictor variables. The main purpose is to identify robust statistical relationships useful for monthly or seasonal prediction in an operational environment. Both local and large-scale predictors are considered, accepting significant large-scale predictors only after the inclusion of local predictors. This approach illustrates the relative importance of different spatial scales for prediction at different time scales (monthly or seasonal). A cross-validation procedure is employed to estimate forecast skill in independent trials. Results of the cross-validation show up apparent decadal-scale variations in predictability which may have implications for seasonal predictability in other mid-latitude locations.

2. Data

2.1 Station Data

The station records used are monthly average temperature and monthly total precipitation for a wide range of locations around New Zealand (Mullan and Renwick, 1996), taken from the New Zealand Climate Database (operated by NIWA). While there are over 100 years of data available from some stations, the total period used here was limited to the years 1957 to 1991

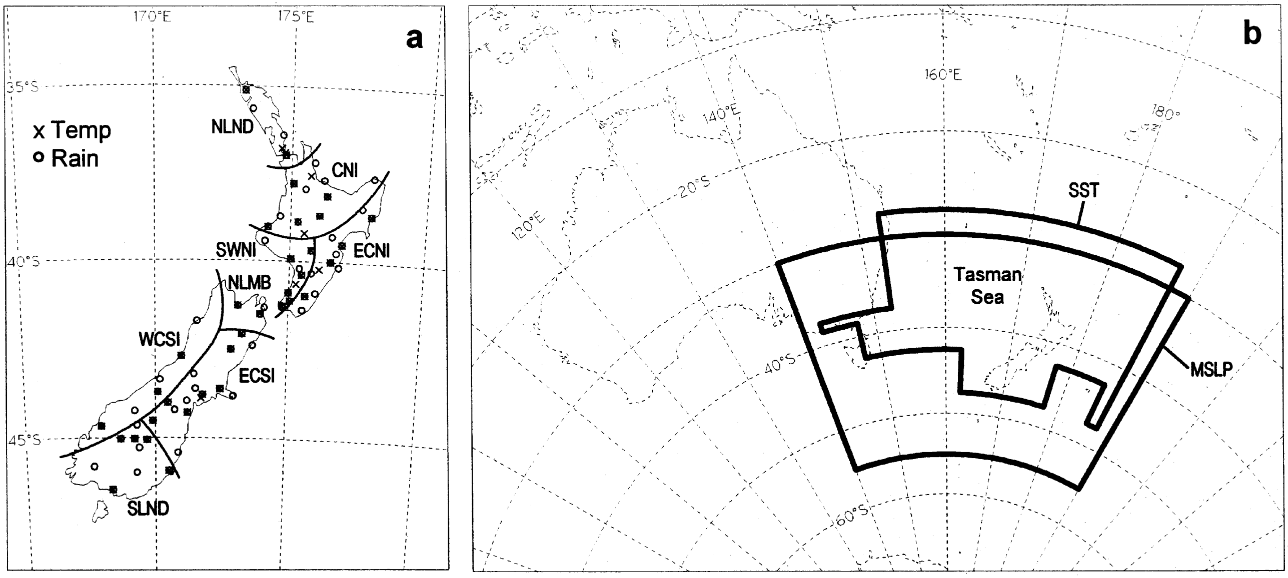

Fig. 1. (a) Map of New Zealand showing location of climate stations used in development of regression equations. Temperature stations are indicated by crosses are rainfall by open circles. Solid lines indicate the regional divisions used for calculation of station averages. Regions are: Northland (NLND), Central North Island (CNI), Southwest North Island (SWNI), East Coast North Island (ECNI), Nelson-Marlborough (NLMB), West Coast South Island (WCSI), East Coast South Island (ECSI) and Southland (SLND). (b) Boundaries of "local" MSLP and SST regions used in EOF analysis

inclusive, to match the record of the available larger-scale circulations indices.

Sites were selected on the basis of the length and continuity of record. In all cases, the record was required to be at least 70% complete, with no break longer than six consecutive months. Breaks of duration up to three months were filled in by linear interpolation and those between four and six months in length were filled using 1957–1991 mean monthly values. Breaks of between four and six months occurred at only a handful of selected stations, and only once at each of those stations. On average, less than 2% of observations were missing at each of the stations selected (65 rainfall stations and 41 temperature stations). Figure 1a shows the geographical spread of stations chosen and the divisions used for generating regionally-averaged time series.

The annual cycle was removed from the temperatures by subtraction of the 1957–1991 mean for each calendar month, at each station. Rainfalls were normalised by conversion to fraction of the mean, i.e. each monthly total was divided by the 1957–1991 monthly mean rainfall, for each station. Seasonal statistics were formed by the combination of monthly statistics, where March, April and May are taken as Autumn; June, July and August as Winter, and so on.

2.2 Large-scale Data

The large-scale data sets comprise monthly averages of gridded mean sea-level pressure (MSLP) and SST, and a monthly time series of the Southern Oscillation Index (SOI), for the period July 1957 through December 1991, this being the maximum period for which all time series were available. This is a period of 34.5 years, or 414 months in total. There were no missing values in these data sets for the period chosen.

The MSLP fields were taken from the regional gridded data set maintained by NIWA (Kidson and Barnes, 1984) up to December 1979, and thereafter from analyses issued by the European Centre for Medium-range Weather Forecasts (ECMWF). MSLP data are defined on a $5^\circ \times 10^\circ$ latitude-longitude grid, from 5° S to 80° S and 60° E to 150° W. The annual cycle was removed by subtraction of the 1957–1991 mean pressure for each month at each grid point. For much of the analysis, we used data over only the local NZ area (30° S to 55° S, 140° E to

170° W, Fig. 1b), but tests were also carried out using MSLP information over many other regions of the grid, as will be described below. Values of the SOI were calculated by "Troup's method" (e.g., McBride and Nicholls, 1983) as the monthly mean Tahiti minus Darwin MSL pressure difference, normalised by the monthly means and standard deviations over the 1941–1980 base period.

SST data were taken from the U.K. Meteorological Office Historical Sea Surface Temperature data set (MOHSST4), described in detail by Bottomley et al. (1990). The data are defined on a 5° latitude-longitude grid and are expressed as anomalies from 1951–1980 monthly means. There is a great deal of missing data prior to about 1950, and also south of about 50° S, which has influenced our choice of study area (12.5° N to 57.5° S and 47.5° E to 77.5 °W). Again, much of the analysis concentrates on the local NZ area, but SST information from many other Southern Hemisphere and Equatorial regions were examined in trials of remote "teleconnectivity".

Missing SST values were handled in much the same way as with station records (see above). SST grid points were used only if there were no breaks of more than 6 months and if no more than 30% of the total record was missing. Gaps were filled in first by linear interpolation in space, treating each monthly field separately, then by linear interpolation in time, treating each grid point separately (for breaks of no more than 3 months), and finally by substitution with long term mean values at remaining missing locations. For the NZ region, 40 of the possible 60 points were retained, mostly north of 50° S (Fig. 1b).

3. Methods and Preliminary Analysis

The main statistical tool used here is multiple linear regression (Draper and Smith, 1981). Many of the MSLP and SST predictors were defined in terms of Empirical Orthogonal Function (EOF) time series (Jolliffe, 1986), to reduce the number of potential predictors. A forward stepwise selection procedure was used to choose predictors from the available set. The cross-validation approach used to estimate forecast skill involved dividing the complete data record into ten-sub-periods. Each sub-period was taken in turn as an independent subset upon which the regression equations were tested. Forecast skill was averaged across all ten independent sets. To estimate significance, forecast skill was compared to that obtained in a series of randomisation trials, as described below.

Results are presented for individual stations and for regional averages (Fig. 1). As will be shown later, there is no gain in mean forecast skill from averaging station time series over geographical regions, which is consistent with the findings of the potential predictability study of Madden and Kidson (1997). However, there are pragmatic advantages to the use of regional averages, in terms of the ease of carrying out repeated trials (e.g., Monte Carlo tests) and in the ease of implementation of the forecast system.

Because of the limited sizes of the samples used and the tendency for large variability between dependent samples in the cross validation, no seasonal stratification was used, at the possible expense of under-estimating forecast skill at some times of year. While earlier results find seasonal variations in predictability (Goulter, 1984; Gordon, 1986), limited trials that we have conducted based on seasonally-stratified data (not shown) suggest that it is the skill of the best prediction equation, rather than its coefficients, that vary seasonally.

3.1 EOF Analysis

An EOF analysis was applied to the monthly mean MSLP and SST anomaly data sets, to reduce the number of potential predictors made available to the regression procedure. In both cases, EOFs were calculated as eigenvectors of the matrix of covariances between pairs of grid points (Jolliffe, 1986). No seasonal stratification was employed, but the mean seasonal cycle was removed at each grid point, as described above. EOF analysis separates the total variance associated with a multivariate data set into a set of orthogonal spatial patterns and an associated set of uncorrelated amplitude time series. Here, the spatial patterns are referred to as the EOFs and their amplitude time series as the principal components (PCs). Most emphasis was given to the EOFs of MSLP and SST over the New Zealand region, but EOFs were also calculated for MSLP in other more distant regions. As the purpose of the EOF analyses is solely to reduce

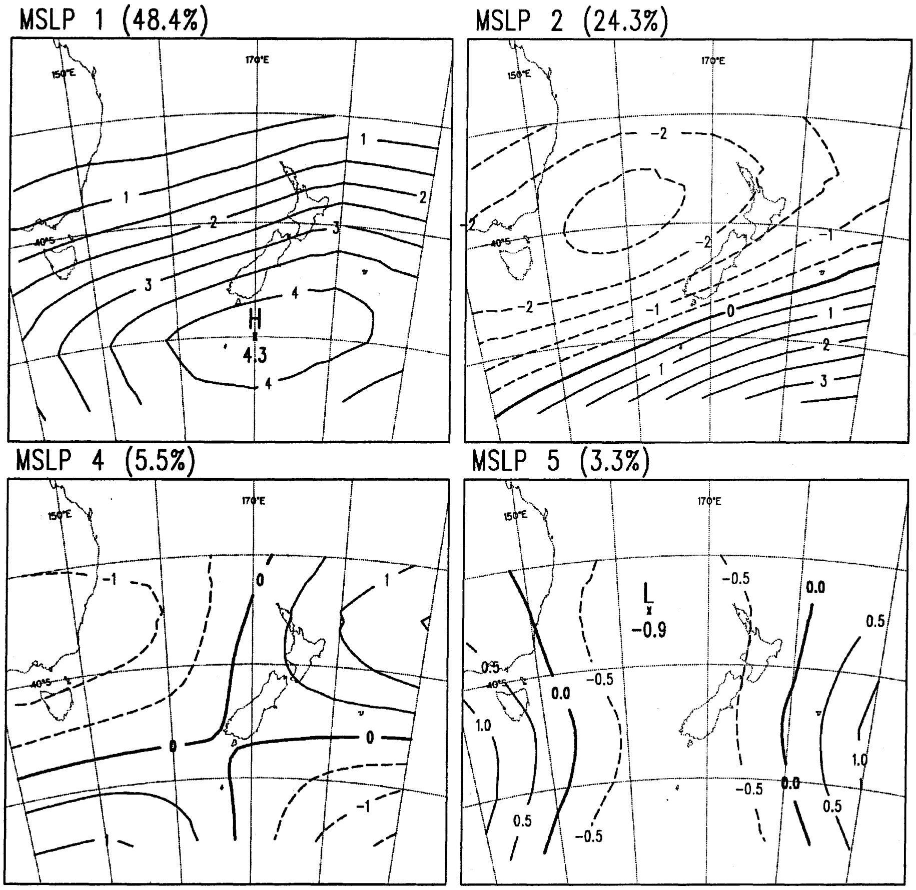

Fig. 2. EOFs 1, 2, 4, and 5 of monthly mean MSLP anomalies. Contours are scaled to show the mean amplitude (hPa) associated with a principal component value of +1 standard deviation. The contour interval is 0.5 hPa with negative contours dashed. Explained variance figures are in brackets

the dimensionality of the predictor data sets, only very brief results are presented.

For the NZ region, standard tests (Craddock and Flood, 1969; North et al., 1982) suggest that at least the leading 6 modes are statistically significant for both MSLP and SST. We retained only the leading five EOFs, which account for 94% of the total MSLP variance and 59% of the SST variance. Figure 2 shows MSLP EOFs 1, 2, 4, and 5, which will be shown later to be the most useful of the MSLP predictors (for brevity, SST patterns are not shown). As is typical of EOF analyses of spatially coherent fields over small domains (Richman, 1981), the leading pattern in both cases describes an in-phase variation over the whole grid, followed by dipoles representing southwest and northwest gradients. The fourth and fifth patterns represent smaller-scale features, being narrow ridges/troughs across NZ with opposing centres to the east and west. The PC time series used as predictors in the regression analyses are referred to below as MSLP.PC1, SST.PC1, and so on.

3.2 Definition of Predictors and Predictands

In all cases, regression equations were developed for "one-time-period-ahead" forecasts, either one month or one three-month season. For any predictand, predictors were defined for several prior time periods, between 1 and 6 months and between 1 and 4 seasons. Four predictands were considered: monthly and seasonal values of temperature and of rainfall. Results are presented for predictands at individual stations and for regional predictands, the latter constructed by averaging the station predictands over each of the eight regions shown in Fig. 1a. As noted in the previous section, no seasonal stratification was used.

The predictors made available to the regression analyses were categorised into two groups, the "standard" predictors and the "additional" predictors. The standard set has thirteen members: the leading five PCs of MSLP and SST anomalies over the NZ region, the SOI for the current month/season and for the month/

season immediately prior to the current time period (referred to as SOI and SOI.L1), and the current value of the predictand (PERS, representing persistence). The additional predictor sets were used to test additional hypotheses, such as bilinear relationships with the SOI (Mullan, 1995, 1996) and remote SST and MSLP teleconnections (Mullan, 1998). For convenience, additional predictors were evaluated only for regional predictands. The additional predictors fall into three distinct groups, as follows.

The first additional group consists of three predictors derived from the SOI. Mullan (1996) found that the effect of the Southern Oscillation on New Zealand climate may be non-linear. In particular, he established that a bilinear (or "broken-stick") form of relationship may exist between the SOI and some circulation variables. Thus, the absolute values of SOI and SOI.L1 (denoted ABS.SOI and ABS.SOI.L1) were considered, since all bilinear functions of SOI that have their break-points at zero may be constructed as a linear combination of SOI and ABS.SOI. The third SOI-related predictor is SOI.P1, the value of the SOI at the future validity time of the predictand. Of course this predictor cannot ever be available for operational forecasting. However, increasingly good model-based ENSO forecasts are becoming available (Chen et al., 1995; Ji et al., 1996) so the aim in evaluating SOI.P1 is to put an upper limit on the utility of a forecast SOI as a predictor for temperature or rainfall.

The second group is made up of MSLP predictors from the entire area covered by the NIWA MSLP data set (5° S to 80° S, 60° E to 150° W). Fifty-four of the 256 grid points were omitted because they contained a large percentage of missing data and the remaining points were grouped in 29 overlapping sub-grids of 5×5 points each. For each grid, the first five principal components from an EOF analysis were used. This makes a total of 145 predictors, referred to as MSLPi.PCj, for grid region $i=1,\ldots,29$ and PC $j=1,\ldots,5$.

The final group of predictors is a set of SST variables derived from the work of Mullan (1998). He defined 8 SST "key areas" that appear to exert some influence on New Zealand climate (see his Fig. 13). For both monthly and seasonal forecasts, a basic set of 8 predictors was constructed: SST1, ..., SST8, the monthly or seasonal mean SST anomaly from each of the eight key areas. Further, the 8 SST anomalies were lagged over the previous 5 months for the monthly forecasts, and over the previous 4 seasons for the seasonal forecasts. Lagged SST predictors are denoted as SSTi.Lj where i denoted the key area number and j the lag. This makes a total of 48 SST predictors for monthly forecasts and 40 for seasonal forecasts.

3.3 Regression Procedure

It is well known that the "true" skill of a forecast system in usually over-estimated if the data set on which the forecast system is constructed is the same as that on which its skill is measured (e.g., Murphy and Katz, 1985). To reduce this over-estimation, a cross-validation procedure was used. The data period was divided into ten parts, each containing approximately 3.5 years of data. Ten sets of regression equations were then developed using nine tenths of the full data set and were tested on the remaining tenth. An overall estimate of forecast skill was taken as the mean skill over the ten independent subsets.

For each predictand, the selection of predictors from the "standard" set of 13 was carried out using the following forward stepwise regression procedure (Draper and Smith, 1981). Predictors were selected one at a time; at each step, the next predictor to be selected was that having the highest partial correlation with the predictand. Selection continued until the next predictor failed one of the two following conditions:

- the predictor must account for at least 1% of the total variance of the predictand, and
- the F-statistic for the newly-included predictor must be significant at the 5% level.

Since the additional predictor sets are so large (nearly 200 variables), making all additional predictors available to the stepwise regression is almost guaranteed to result in a number of spurious relationships being judged significant. An initial "pre-screening" step was therefore carried out, using the full data set. For every predictand, each additional predictor was considered in isolation, after inclusion of all the selected standard predictors. To be selected for a particular region, an additional predictor was required to not only satisfy the above variance reduction and significance criteria, but also to increase the forecast skill by at least 0.01 and to be selected in at least two other regions. The additional predictors which survived the pre-screening were then made available to the regression/cross-validation system in the same way as the standard predictors.

3.4 Estimation of Skill

The skill, s, of a regression forecast of temperature or rainfall was measured, relative to a baseline or "zero-skill" forecast (usually climatology), as a ratio of sums of squares:

$$s = \sum_i (b_i - o_i)^2 / \sum_i (r_i - o_i)^2$$

were o_i is the observed value of the predictand (temperature or rainfall) for the ith month (or season) and r_i and b_i are the values predicted using regression and the baseline forecast, respectively. A skill measurement of s=1 indicates that the regression is no more accurate on average than the baseline forecast, s>1 indicates increased accuracy over the baseline forecast, and s<1 indicates decreased accuracy.

The above skill score was chosen over the more "standard" R^2 statistic (the square of the correlation between the observed and forecast values), since R^2 is not in general a good measure of skill for a forecast (for example, adding a constant to a forecast does not change R^2 but will change the skill). The percentage of variance explained (which is equivalent to R^2 for regression forecasts over the dependent data set) was not used either because, with the cross-validation procedure, it can take negative values, which are difficult to interpret.

Unless otherwise stated, the results presented below use a baseline forecast of climatology, the mean observed value for the given location and time of year. Some results are given (for temperature predictands only) with a baseline of "damped persistence", a first-order autoregressive model (AR(1)) where the only predictor is the lagged predictand (PERS). When the baseline forecast is climatology, skill s is related to percentage explained variance V_{expl} on the dependent data by

$$V_{expl} = 100 \cdot \left(1 - \frac{1}{s}\right)$$

In the cross-validation procedure, the forecast skill s_j is calculated for each of the independent samples $D_j, j = 1, \ldots, 10$. The overall estimate of skill, s, is taken as the mean of the ten s_j. To calculate s_j a forecast equation is constructed by regression using only the data in D_j', the complement of D_j. The equation is then applied to the independent sample D_j and a sum of the squares of the differences between predictions and observations is calculated. That is,

$$\begin{aligned} s &= \frac{1}{10}\sum_j s_j \\ &= \frac{1}{10}\sum_j \left[\sum_i (b_{ij} - o_{ij})^2 / \sum_i (r_{ij} - o_{ij})^2\right] \end{aligned}$$

Where o_{ij} is the observed value of the predictand for the ith month (or season) of D_j, and r_{ij} and b_{ij} are predictions of o_{ij} (from the regression and baseline predictors, respectively) based on forecast equations constructed in D_j' and applied to D_j. Note that the periods used in the cross validation apply equally to both the regression and baseline predictors (i.e., climatology and damped persistence are defined differently for each D_j).

The significance of any estimate of skill for a given predictand was assessed using the following randomisation procedure. First, a dummy predictand was created by choosing at random a predictand of the same averaging period (i.e., monthly or seasonal) as the given predictand, and randomly reordering it in time. An estimate of skill was then calculated for the dummy predictand using the same procedure as for the given predictand. Two hundred such trials were performed for each predictand and the original estimate of skill was deemed significant if it was greater than at least 95% of the dummy values.

4. Results

4.1 Standard Predictors

Using the standard predictor set, the estimated skill was found to be modest for temperature stations (median values were 1.12 and 1.09 for

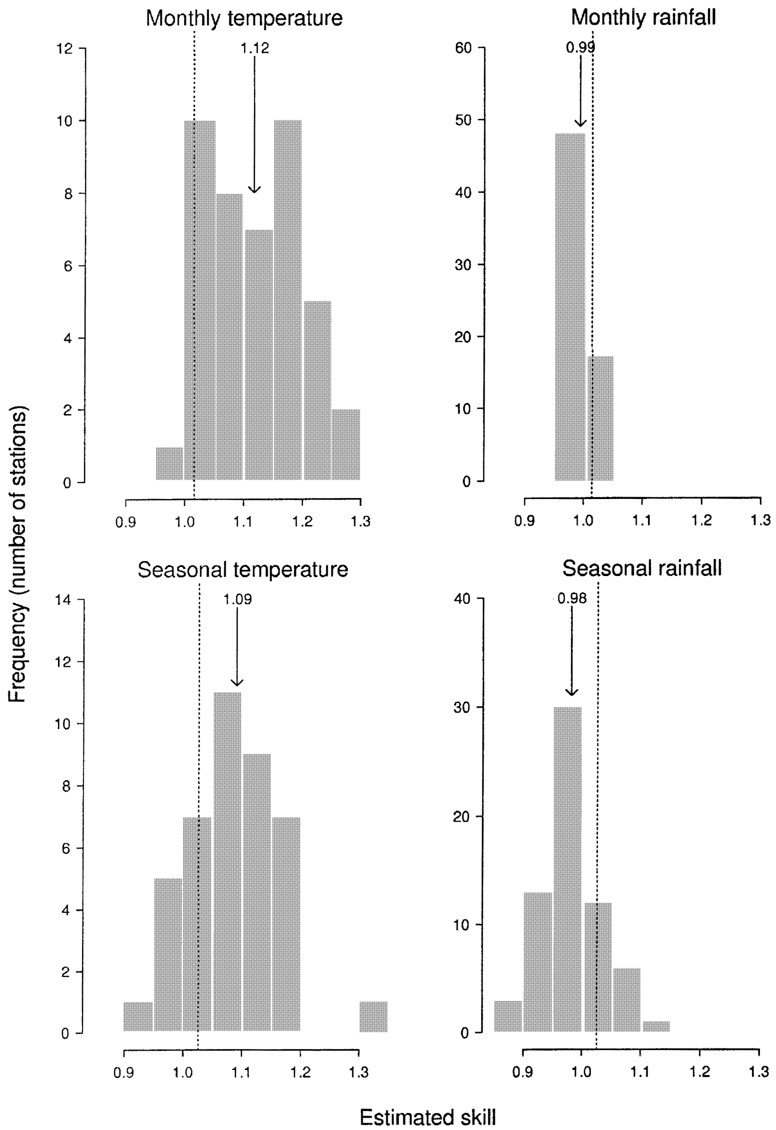

Fig. 3. Distribution of estimated skill of forecasts using standard predictors, for each of four predictand types: (a) monthly temperature; (b) seasonal temperature; (c) monthly rainfall; and (d) seasonal rainfall. The dashed lines indicate 95% significance levels calculated from randomised trials. Printed figures (above arrows) indicate the median skill for each predictand type

monthly and seasonal forecasts, respectively) and poor for rainfall stations (medians were 0.99 and 0.98, Fig. 3). The randomisation tests resulted in threshold values of 1.014 for monthly forecasts and 1.026 for seasonal forecasts. Based on these values forecast skill was significantly greater than 1 for most temperature stations (40/41 monthly and 32/41 seasonal), but for only a minority of rainfall stations (7/65 and 14/65, respectively). For regional forecasts, estimated skill was typically close to, but slightly less than, the mean skill over the associated individual station predictands. Hence, regional forecast skill was significantly greater than 1 for almost all temperature predictands (except ECNI and NLMB seasonal temperatures) but for only 3 of the 16 rainfall predictands. On the basis of the standard predictors alone, monthly forecasts were more skilful than seasonal forecasts. As will be shown later, the situation is reversed once the additional predictors are considered.

Mean levels of skill were positively correlated with latitude and with the mean number of predictors selected. For monthly temperatures, forecast skill was significantly correlated with the mean number of predictors ($R=0.73$) and with latitude ($R=0.70$, Fig. 4a), reflecting the greater predictability of monthly temperature anomalies in the north of the country. However, southern stations fared better at the seasonal time scale, where there is a relatively uniform latitudinal distribution of forecast skill using

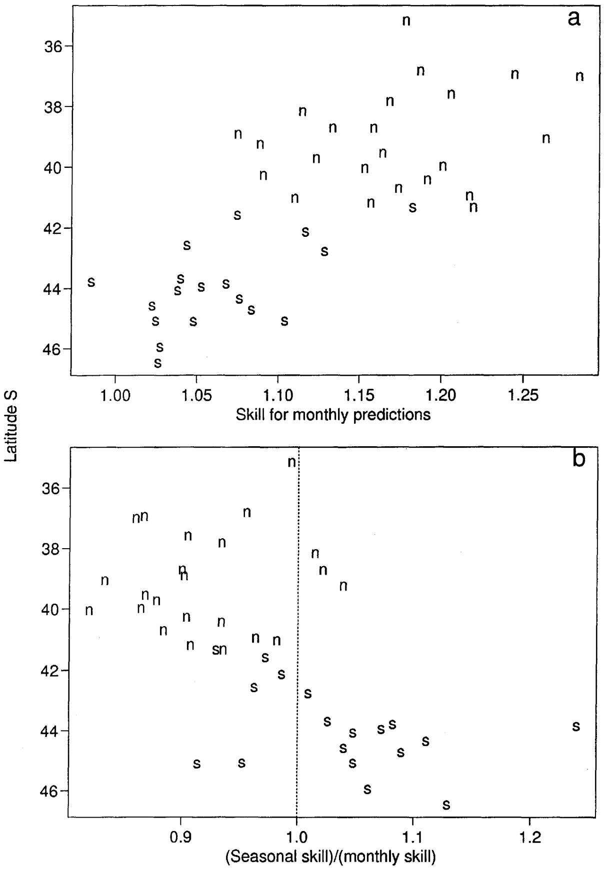

Fig. 4. Skill of Predictions of temperature one month or one season ahead, based on standard predictors. Each point in the plots represents one temperature station ('n' and 's' represent North Island and South Island stations, respectively). (a) Monthly skill plotted against latitude; (b) The ratio of seasonal skill to monthly skill, plotted against latitude

the standard predictors. Seasonal skill is generally slightly lower than monthly skill for North Island stations, while the reverse is true in the South (Fig. 4b). The above results suggest that intraseasonal variability becomes less predictable at higher latitudes, presumably because it becomes more dominated by the unpredictable internal variability in the higher-latitude atmospheric circulation. Seasonal predictability is more uniform as it is more strongly controlled by more remote larger-scale SST forcing (Rowell et al., 1995).

The skill of monthly temperature forecasts relative to damped persistence ranged from 0.96 to 1.12 (median 1.03). Skill was greater than 1 at all but one North Island station and at two thirds of South Island stations (18 of 24). For seasonal temperature forecasts however, skill relative to damped persistence was less than 1 at most stations (31 of 41). As mentioned above, inclusion of the additional predictors noticeably improves the skill of seasonal temperature forecasts.

For some temperature stations there was a clear pattern of variation of skill through the year. This was strongest at the six northern-most stations, where skill was highest in spring and was lowest in autumn (Fig. 5a). A slightly different pattern occurred at the western-most stations (in the south and west of the South Island), where skill was lowest in winter (Fig. 5b). At most of the remaining stations, there was either no clear pattern, or skill was lowest in the middle of the year. An exception was a single station on the west coast of the South Island (Hokitika), where skill was highest in the middle of the year (Fig. 5c).

The importance of particular predictors (in terms of how often they were selected) varied strongly with predictand type (Fig. 6). For monthly temperature forecasts, the predictor most commonly selected was MSLP.PC5, which was chosen in 96% of the regressions. This 5th MSLP EOF (Fig. 2) represents a ridge/trough just west of NZ and is associated with anomalous meridional flow. As defined here, a positive value of MSLP.PC5 is associated with anomalous northerly (poleward) flow over New Zealand and anomalous southerly (equatorward) flow over the western Tasman Sea. Its regression coefficient was always negative, implying that equatorward flow anomalies to the west this month/season are associated with below normal temperatures next month/season. The connection may arise through the effects of the circulation anomaly upon local SSTs, since circulation variability appears to lead SST variations in the NZ region by 1–2 months (Basher and Thompson, 1995). Correlations are strongest in the southeast of the North Island, the region of the country where the anomalous surface flow associated with the EOF is strongest.

The five most frequently selected predictors (MSLP.PC5, SST.PC1, PERS, MSLP.PC1, and MSLP.PC4) accounted for almost 90% of predictor selections for monthly mean temperatures. The same five predictors were also important for seasonal temperature forecasts

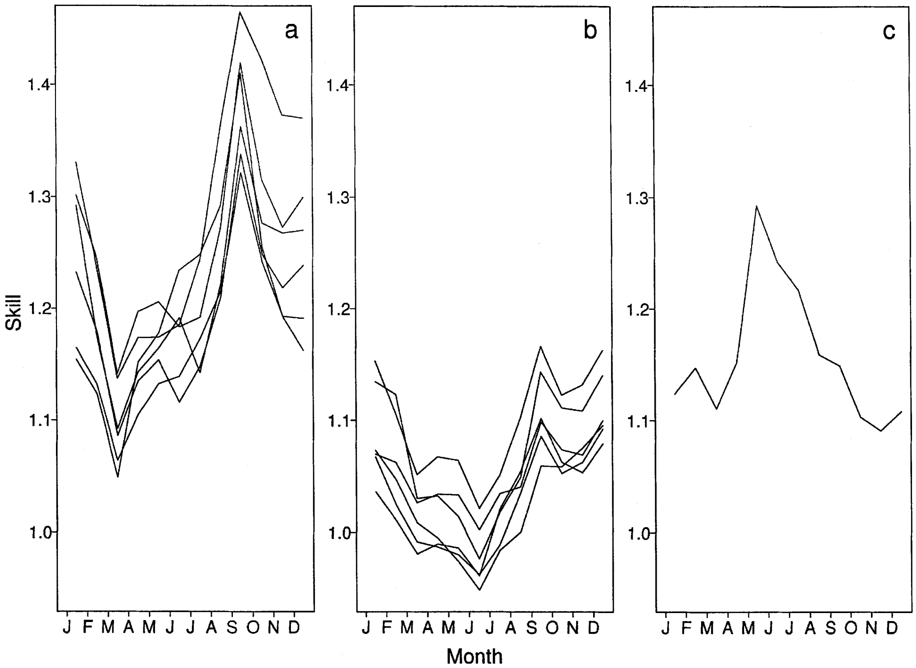

Fig. 5. Within-year variation in skill in predicting temperature one month ahead, for selected temperature stations: (a) the six northern-most stations; (b) the six western-most stations; (c) one isolated station at Hokitika, on the west coast of the South Island. The data were smoothed: each line represents one temperature station and joins points which are the mean skill for an adjacent group of six months plotted against the mid-point of the six-month period

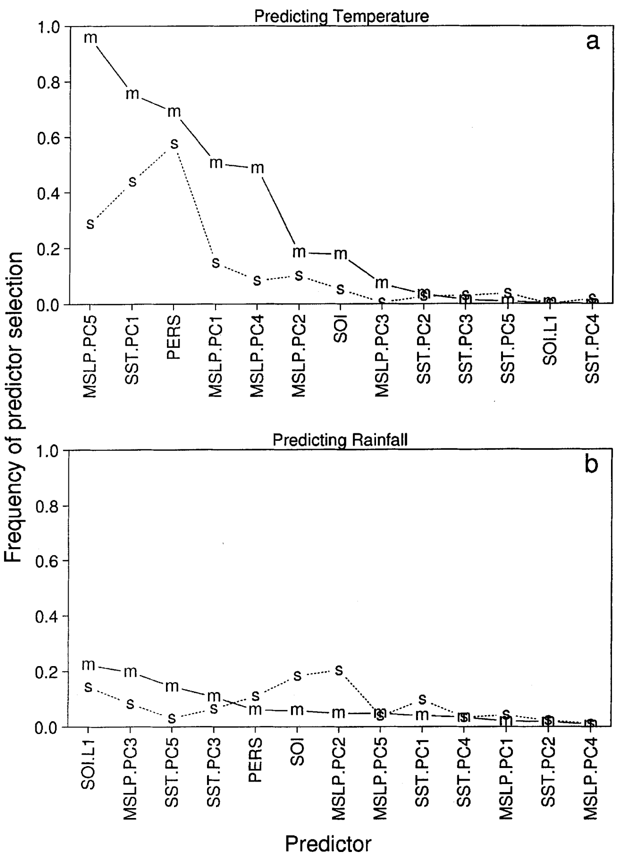

Fig. 6. Frequency of predictor selection for monthly ('m') and seasonal ('s') forecasts of (a) temperature; and (b) rainfall. The y-axis in both panels shows the proportion of predictands for which each predictor is selected

(accounting for 85% of selections) though their ranking differed. Amongst the SST predictors, only the first principal component, SST.PC1, was selected with any frequency for temperature forecasts, always with a positive regression coefficient. For temperature, persistence is relatively more important at the seasonal time scale than at the monthly time scale, which is consistent with the results of Goulter (1984). In contrast to temperature, for rainfall the predictors important for monthly forecasts were usually not so for seasonal forecasts, and vice versa. The one exception was SOI.L1, which was ranked 1st and 3rd for monthly and seasonal forecasts, respectively.

There were some strong geographical trends in the frequency of selection of some predictors (Table 1). Some predictors (notably the two SOI predictors, and the first two MSLP principal components) were more often selected in the North Island; others (e.g., SST.PC3) were more commonly chosen in the South Island. The preference for the SOI at northern stations, especially for rainfall (where regression coefficients were always positive), agrees with the earlier findings of Gordon (1986), who found a preference for moist northeast flow over the North Island in La Niña conditions.

Table 1. *Predictors that were Selected more than Twice as Frequently in the North Island as in the South Island (or vice versa). Predictors for which the Frequency of Selection was less than 0.1 in Both Islands are omitted*

Predictand type	Predictor	Frequency of selection (proportion of stations) North Is.	South Is.	Island in which frequency is highest
Monthly temperature				
	SOI	0.26	0.07	North
	MSLP.PCI	0.70	0.26	North
	MSLP.PC2	0.32	0.01	North
	MSLP.PC3	0.00	0.16	South
Seasonal temperature				
	SST.PC1	0.64	0.18	North
	MSLP.PC1	0.23	0.03	North
	MSLP.PC2	0.17	0.01	North
	PERS	0.38	0.82	South
	MSLP.PC4	0.03	0.14	South
Monthly rainfall				
	SOI.L1	0.38	0.06	North
	SST.PC5	0.25	0.03	North
	SST.PC3	0.04	0.18	South
	MSLP.PC3	0.07	0.32	South
Seasonal rainfall				
	SOI.L1	0.25	0.03	North
	MSLP.PC3	0.14	0.02	North

4.2 Additional Predictors

For brevity, the additional predictors are considered only in regressions upon regionally averaged temperature and rainfall anomalies, rather than at individual stations. Given that the skill of prediction of regionally averaged anomalies is roughly equal to the average of the skill of prediction of each region's individual station anomalies, the use of regional averages here is expected to give a realistic picture of the utility of new predictor variables.

The pre-screening procedure described in section 3.3 selected 7 additional predictors for monthly temperature, 11 for seasonal temperature, and 1 for monthly rainfall (Table 2). Mean skill and changes in skill after inclusion of additional predictors are listed in Table 3. The additional predictors had very little impact on monthly forecast skill but showed a large positive impact on the skill of seasonal forecasts. The median gain in skill for seasonal temperature forecasts was 0.18 (0.04 for monthly) and was negligible for rainfall. The greater impact of the ''remote' predictors at the longer time scale seems physically reasonable in terms of larger spatial scales being associated with longer time scales in the atmospheric and oceanic circulation. With the inclusion of additional predictors, seasonal temperature forecasts are more skilful on average than monthly forecasts, as might be expected a priori (Rowell et al., 1995). As will be discussed in Section 4.3, there is considerable scatter in the skill estimates between different

Table 2. *Predictors Selected from the set of Additional Predictors for each of the four Predictand Types*

Predictand Type	Selected Predictors
Monthly Temperature	SST2.L1, MSLP16.PC1, MSLP17.PC2, MSLP22.PC1, MSLP23.PC3, MSLP24.PC4, MSLP25.PC4
Seasonal Temperature	SST2.L2, SST6.L3, SST7.L1, SST8.L2, MSLP1.PC1, MSLP2.PC1, MSLP3.PC2, MSLP8.PC2, MSLP11.PC4, MSLP12.PC4,MSLP25.PC4
Monthly Rainfall	MSLP11.PC4

Table 3. *Estimated Skill by Predictand Type and Region, after the Inclusion of Predictors Selected from the Additional Predictor set, and estimated Gain in Skill over that from the Standard Predictors Alone*

Predictand Type	Region								
	NLND	CNI	ECNI	SWNI	NLMB	ECSI	WCSI	SLND	Median
Skill									
Monthly temperature	1.27	1.17	1.18	1.13	1.14	1.09	1.14	1.12	1.12
Seasonal temperature	1.38	1.19	1.16	1.49	1.22	0.98	1.13	1.35	1.20
Monthly Rainfall	0.98	1.03	1.03	1.03	0.98	0.99	0.97	0.97	0.98
Gain in skill									
Monthly temperature	0.04	0.00	0.06	−0.06	−0.02	0.04	0.04	0.04	0.04
Seasonal temperature	0.24	0.15	0.19	0.33	0.26	−0.07	0.10	0.16	0.18
Monthly Rainfall	0.01	0.02	0.00	0.04	0.00	0.00	0.00	0.00	0.00

independent samples. The occasional decreases in skill after inclusion of additional predictors may be attributed to sampling variability, but may also reflect the presence of a few "chance" relationships.

Most of the selected predictors came from the set of wide-area MSLP grids but the most consistently useful were from the SST key area sets. For monthly temperatures, MSLP regions over the southern ocean south of Australia and the Tasman Sea were often selected. The most useful wide-area MSLP predictor was the index of a trough/ridge EOF pattern over a region south of the Tasman Sea, which may be acting as an indicator of the phase of the dominant pattern of planetary wave activity in the westerlies. For seasonal temperatures, an index of meridional flow over the western Tasman Sea was most commonly selected. Lag correlation maps (not shown) indicate this index is associated with SST anomalies near NZ one season later, presumably through a combination of surface heat fluxes and oceanic advection.

At both time scales, the most commonly selected SST key area was area 2, in the Australian Bight/Tasmania region. Variations in SST2 are positively correlated with monthly temperatures over northern NZ two months later and with seasonal temperatures over most of NZ two seasons later, suggestive of oceanic advection. Indian Ocean SST variations (key area 7) were also important for seasonal temperature, where positive anomalies of SST7 were related to above-average temperatures in most regions of New Zealand two seasons later, perhaps related to the formation of "northwest cloud bands" (Mullan, 1998), but not directly to advection.

Somewhat surprisingly, none of the additional SOI predictors were selected. Additional predictors are considered only if they add significant information beyond that given by the standard predictors. Hence, the additional "future" value SOI.P1 was perhaps not selected as it is very highly correlated with the standard predictor SOI from the previous time period.

4.3 Reliability of Skill Estimates

The predictor selection procedures and cross-validation approaches used here aim to maximise the reliability of the results, with a view to operational use. However, when forecast skill is modest, as it is here, the fine detail of the method of measuring skill becomes important. Note only is there considerable variability in skill between independent samples, but also slight changes in the method of estimating skill may be sufficient to shift the estimated skill from less than to greater than 1.

For each forecast evaluated above, the estimated skill s the mean of 10 estimates s_j, one for each of the independent sets D_j. An examination of some patterns amongst these s_j is useful in interpreting the skill of these forecasts. Two patterns are apparent in Fig. 7. First, as the estimated skill of a forecast increases, so does the variability amongst the s_j. This means that the larger the estimated skill for a predictand the less certain that estimate is. Note also that even when s is high there are always some of the s_j that are less than one. This has implications for the level of skill that can be expected from operational forecasting of these predictands using the regression technique. Although we can be

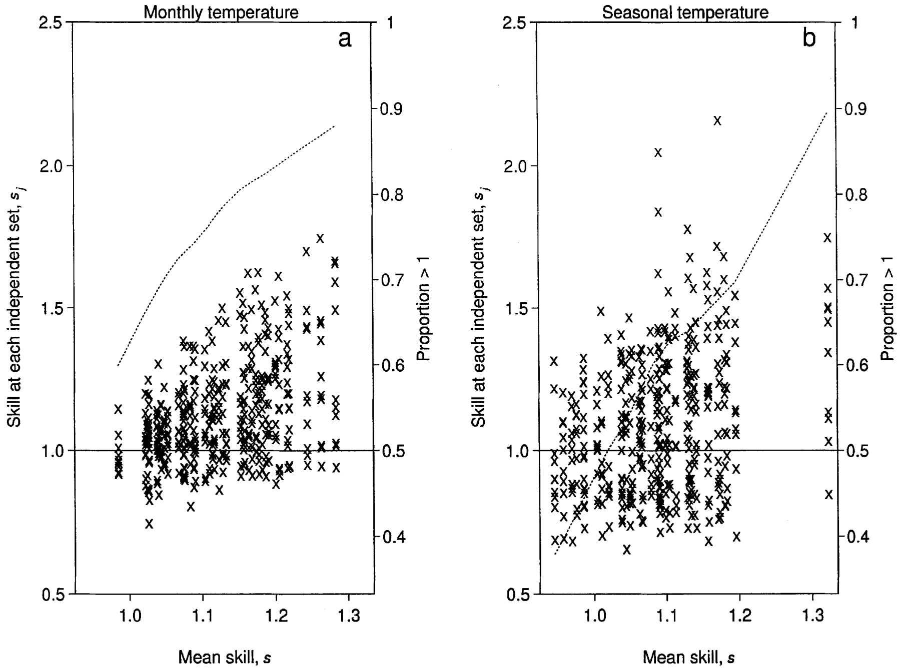

Fig. 7. The relationship between the skill of a forecast, s, and the 10 skill estimates, s_j, of which s is a mean: (a) for monthly temperature forecasts; and (b) for seasonal temperature forecasts. For each predictand (monthly or seasonal temperature at a particular station) ten s_j values are plotted against their mean, s. The broken lines are smooth curves fitted to the data showing the proportion of the s_j for each station that are greater than 1 (values shown on the right-hand axis)

reasonably confident that, in the long run, these forecasts will be more skilful on average than climatology, this may not be so in the short term.

The second obvious pattern in Fig. 7, which follows from the above observation, is that there is much more variability in skill estimates for seasonal than for monthly forecasts. One consequence of this is that the proportion of the s_j that are greater than 1 is typically less for seasonal than for monthly predictands (compare the height of the dotted lines in the two panels). Thus, for a seasonal forecast we may have less confidence that it will be more skilful than climatology in the short term than we can for a monthly forecast of the same estimated skill. This result will be partly due to the much smaller samples available for the development of seasonal forecasting equations.

When the data in Fig. 7 are plotted by independent sample number rather than by mean skill (Fig. 8) it becomes clear that skill varies substantially amongst independent sets. The differences between sets are statistically significant: using Mann-Whitney tests to make all pairwise comparisons of skill estimates in different independent sets, 34 of 45 results were significant ($p < 0.05$) for monthly temperature, and 35 of 45 were significant for seasonal temperature. Further, the temporal variations seen in Fig. 8 are not a result of the somewhat arbitrary size chosen for the independent sets. Since each set is not an integer number of years, the number of relatively unpredictable winters varies with the set number, which may influence the results. However, there is little correlation between the distribution of seasons in each independent set and the mean forecast skill, and trials with independent sets covering exactly three years each produced essentially the same results.

Thus it is clear that the vertical scatter in Fig. 7 is caused partly by sampling error and partly by temporal variation in skill. In other words, there are periods (like the late 1950s – the first independent set) when regression forecasts of temperature could not be expected to be better than climatology, and other periods (e.g., 1988–1991) when they would have been substantially better. The observed decadal-scale variations in skill are not obviously ENSO-related, but nevertheless are reminiscent of decadal-scale variations in the skill of ENSO predictions over the last 15–20 years (Ji et al., 1996). The forecast skill of statistical and dynamical models of ENSO was relatively high in the 1980's but was considerably lower through the extended "warm event" of the early-mid 1990's. Such variations in predictability may be related to

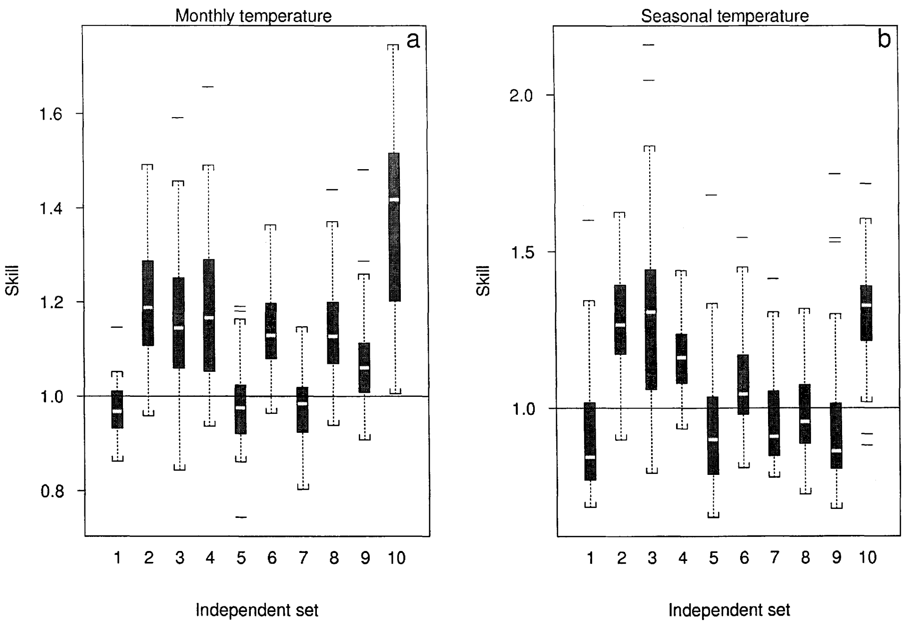

Fig. 8. Boxplots of skill estimates, s_j, grouped by independent set for forecasts of (a) monthly temperature, and (b) seasonal temperature. Note that the data plotted are exactly the same as those in Fig. 7, the difference being that the horizontal grouping is by independent set, rather than by station. In each boxplot the black box represents the middle half of the data, the median is shown as a white line, and the "whiskers" extend to 1.5 times the interquartile range. Outliers are indicated by short horizontal bars

decadal and longer-scale modulation of the amplitude of ENSO and other low-frequency elements of the general circulation.

5. Discussion and Conclusions

The main motivation for this work was to provide tools for, and estimate a lower bound to the skill of, operational short-term climate predictions for New Zealand. The average skill of regression-based rainfall predictions is found to be insignificantly different from that of a pure climatological forecast. Temperature forecasts do however appear to be significantly more skilful than climatology, with median skill of 1.12 for monthly anomalies and 1.20 for seasonal anomalies (roughly equivalent to explained variances of 10% and 17% over climatology, respectively). In broad agreement with earlier studies (Gordon, 1986; Kidson and Gordon, 1986; Mullan and Renwick, 1996), we find that skill is correlated with latitude, variability at northern stations being the most predictable. Skill is generally highest in Spring and Summer.

Beyond persistence and ENSO-related effects, the most useful predictors of temperature variations appear to be SST anomalies to the west (upstream) of New Zealand, and indices of meridional flow anomalies around and to the west of the country. The latter relationships are suggestive of the role of atmospherically-forced local SST anomalies in modulating local climate. Concomitant changes in land surface conditions such as soil moisture and snow cover may also play a role. In terms of seasonal prediction, atmosphere-ocean effects may be more useful than the pure persistence of atmospheric anomalies alone (e.g., Kidson and Barnes, 1984). Seasonal temperature anomalies are found to be predicted more skilfully than monthly anomalies, but only after the inclusion of MSLP and SST predictors defined over regions remote from NZ. Slowly-varying features of the global atmospheric and oceanic circulation (ENSO-related and otherwise) therefore confer a small but significant amount of predictability upon seasonal-mean NZ climate.

The levels of skill found here are broadly consistent with those found for a number of other extra-tropical regions of the globe (e.g., Barnston and Smith, 1996). The greater predictability of temperature over rainfall is largely due to the stronger spatial and temporal coherence of temperature compared to the more localised, episodic nature of rainfall. Even for seasonal temperature anomalies, forecast skill is modest using the techniques employed here. However, given the potential value of climate predictions to the energy, agricultural and other sectors, even a small level of skill may be translated into significant benefits in dollar terms. Possible improvements in skill are presently being investigated, through the use of fully hemispheric data sets and through the analysis of ensembles of seasonal climate model integrations, though it is unlikely that large improvements in skill will emerge.

Although this study concerns only one relatively small region of the globe, the methodology is applicable to any region. The more general findings, regarding the linking of longer time-scale local climate variations with larger spatial-scale circulation variability, and the temporal variations found in predictability, are likely to apply to other regions of the extra-tropics.

On the basis of the above results, it appears feasible to generate operational monthly and seasonal temperature predictions, at least for the warm half of the year. One caveat for operational forecasting is the observed decadal variability in overall predictability (Fig. 8), which suggests that although we may be able to estimate long-term mean forecast skill, we do not know whether the near future (next 3–5 years) will be a period of high or low predictability. Further research into the processes involved may help elucidate the reasons for such variations in skill.

Acknowledgements

We wish to thank Craig Thompson for assistance with provision of data sets. Brett Mullan and John Kidson provided useful comments on an earlier version of this manuscript. The Contributing Editor, Reid Basher, and two anonymous reviewers provided many valuable comments which helped to clarify and focus the paper. The research described here was funded by the New Zealand Foundation for Research, Science and Technology under contract CO1628.

References

Barnston, A. G., Smith, T. M., 1996: Specification and prediction of global sea surface temperature and precipitation from global SST using CCA. *J. Climate,* **9**, 2660–2697.

Basher, R. E., Thompson, C. S., 1995: Relationship of air temperature in New Zealand to regional anomalies in sea surface temperature and atmospheric circulation. *Int. J. Climatol.,* **16**, 405–425.

Bottomley, M., Folland, C. K., Hsiung, J., Newell, R. E., Parker, D. E., 1990: Global ocean surface temperature atlas (GOSTA). *Joint U.K. Meteorological Office/Massachusetts Institute of Technology Project*. HMSO, London. 20+iv pp and 313 plates.

Chen, D., Zebiak, S. E., Busalacchi, A. J., Cane, M. A., 1995: An improved procedure for El Niño forecasting. *Science,* **269**, 1699–1702.

CLIVAR Scientific Steering Group, W. C. R. P., 1995: *CLIVAR Science Plan: A Study of Climate Variability and Predictability.* World Meteorological Organization, 157 pp.

Craddock, J. M., Flood, C. R., 1969: Eigenvectors for representing the 500 mb geopotential surface over the Northern Hemisphere. *Quart. J. Roy. Meteor. Soc.,* **95**, 576–593.

Draper, N. R., Smith, H., 1981: *Applied Regression Analysis,* 2nd edn. Probability and Mathematical Statistics. New York, USA: Wiley, 709 pp.

Gordon, N. D., 1986: The Southern Oscillation and New Zealand weather. *Mon. Wea. Rev.,* **114**, 371–387.

Goulter, S. W., 1984: Persistent mean air temperatures in the New Zealand sector of the South Pacific. *N. Z. J. Science,* **27**, 221–236.

Ji, M., Leetmaa, A., Kousky, V. E., 1996: Coupled model predictions of ENSO during the 1980s and the 1990s at the National Centers for Environmental Prediction. *J. Climate,* **9**, 3105–3120.

Jolliffe, I. T., 1986: *Principal Component Analysis.* Springer-Verlag, 271 pp.

Kidson, J. W., 1988: Interannual variations in the Southern Hemispehre circulation. *J. Climate,* **1**, 1177–1198.

Kidson, J. W., Barnes, B. G., 1984: Indices of the atmospheric circulation in the Australasian region. *New Zealand Journal of Science,* **27**, 355–364.

Kidson, J. W., Gordon, N. D., 1986: Interannual variations in New Zealand temperature and precipitation patterns. *N. Z. J. Geol. Geophys.,* **29**, 363–375.

Lau, N.-C., 1997: Interactions between global SST anomalies and the midlatitude atmospheric circulation. *Bull. Amer Meteor. Soc.,* **78**, 21–33.

Madden, R. A., Kidson, J. W., 1997: The potential long range predictability of temperature over New Zealand. *Int. J. Climatol.,* **17**, 483–495.

McBride, J. L., Nicholls, N., 1983: Seasonal relationships between Australian rainfall and the Southern Oscillation. *Mon. Wea. Rev.,* **111**, 1998–2004.

Mullan, A. B., 1995: On the linearity and stability of Southern Oscillation-climate relationships for New Zealand. *Int. J. Climatol.,* **15**, 1365–1386.

Mullan, A. B., 1996: Non-linear effects of the Southern Oscillation in the New Zealand region. *Aust. Met. Mag.,* **45**, 83–99.

Mullan, A. B., 1998: Southern Hemisphere sea surface temperatures and their contemporary and lag association with New Zealand temperature and precipitation. *Int. J. Climatol.,* (submitted).

Mullan, A. B., Renwick, J. A., 1996: Predictability of New Zealand Climate on monthly and seasonal time scales. *NIWA Science and Technology Series,* **40**. Wellington. 52 pp.

Murphy, A. H., Katz, R. W., (eds.), 1985: *Probability, Statistics and Decision Making in the Atmospheric Sciences.* Boulder, Colorado, USA. Westview Press, 545 pp.

Nicholls, N., 1989: Sea surface temperatures and Australian winter rainfall. *J. Climate,* **2**, 965–973.

North, G. R., Bell, T. L., Cahalan, R. F., 1982: Sampling errors in the estimation of empirical orthogonal functions. *Mon. Wea. Rev.,* **110**, 699–706.

Palmer, T. N., Anderson, D. L. T., 1994: The prospects for seasonal forecasting – A review paper. *Quart. J. Roy. Meteor. Soc.,* **122**, 755–793.

Richman, M. B., 1981: Obliquely rotated principal components: an improved meteorological map typing technique? *J. Appl. Meteor.,* **20**, 1145–1159.

Rowell, D. P., Folland, C. K., Maskell, K., Ward, N., 1995: Variability of summer rainfall over tropical north Africa (1906–92): Observations and modelling. *Quart. J. Roy. Meteor. Soc.,* **121**, 669–704.

Trenberth, K. E., 1975: A quasi-biennial standing wave in the Southern Hemisphere and interrelations with sea surface temperature. *Quart. J. Roy. Meteor. Soc.,* **101**, 55–74.

World Meteorological Organization, 1995: Climate Information and Prediction Services. *World Meteorological Organization Publication,* **832**. Geneva. 16 pp.

Authors' address: James A. Renwick and R. I. C. Chris Francis, NIWA Wellington, P.O. Box 14–901, Kilbirnie, Wellington, New Zealand, (e-mail: J. Renwick@niwa.cri.nz).