Weather and Climate, **23** (2004), 43-60

Are NCEP Seasonal Forecasts Useful in New Zealand?

R.I.C. Chris Francis* and James A. Renwick

NIWA, Wellington

Abstract

There is growing interest, and activity, in seasonal climate forecasts, on both global and local scales. In New Zealand, seasonal forecasts have been produced by NIWA and issued to the public regularly since mid 1999. Such forecasts have obvious potential economic value in industries like agriculture and energy generation. A possible source of guidance for local forecasts is the set of global extended range forecasts produced by major international agencies. This paper evaluates one such source: seasonal forecasts of 200 hPa heights, surface temperature, and rainfall produced by the National Centers for Environmental Prediction (USA) for the three-year period 1997 to 1999, inclusive. Two classes of forecasts are found to be more skilful than both persistence and climatology: 200 hPa heights in the tropics outside of the Pacific, and surface temperatures in the eastern tropical Pacific. Over much of the middle latitudes, including New Zealand, forecasts are found to have little skill compared to use of a climatological mean state. Model forecasts do display skill compared to persistence forecasts. For New Zealand, forecast utility seems limited to predictions of the state of ENSO, which can be combined with local statistical information in the generation of climate outlooks.

**Corresponding author: Dr Chris Francis, NIWA, PO Box 14-901, Wellington, New Zealand.*
email: c.francis@niwa.co.nz

1. Introduction

In a country like New Zealand, where agriculture makes a substantial contribution to the economy, there is clearly potential value in extended-range forecasts of climate anomalies. Since June 1999, the National Climate Centre of NIWA has been producing monthly reports that include such forecasts for the following three-month period (see http://www.niwa.cri.nz/NCC/). These forecasts are semi-objective, deriving from a mixture of inputs (Renwick et al. 1999), including statistical predictions based on current conditions (e.g., Francis and Renwick 1998). Validations indicate that issued forecasts show positive skill compared to using the climatological mean as a "forecast", but skill levels are modest (Renwick et al. 2000).

New Zealand climate predictability appears to be reliant largely on remote sea-surface temperature (SST) forcing, especially from the tropical Pacific and Indian Oceans, combined with an element of persistence (Mullan 1998, Zheng and Renwick 2003). The potential predictability of seasonal mean climate over New Zealand is around 50% for temperature, but closer to 30% for rainfall (Madden and Kidson 1997, Madden et al 1999). Statistical forecast techniques for New Zealand use predictors based predominantly on tropical and subtropical SST anomalies, plus local SSTs and recent local climate anomalies.

There are two ways of improving real-time seasonal forecasts: by getting more information from existing inputs, and by using new inputs. Francis et al. (2003) examined one new input: one-month ensemble forecasts of mean sea-level pressure (MSLP) in the Southern Hemisphere from the United Kingdom Meteorological Office (UKMO). In the present study we examine another potential input: seasonal forecasts of global geopotential height, temperature, and rainfall from the U.S. National Centers for Environmental Prediction (NCEP). The NCEP approach uses predicted tropical SSTs combined with observed extra-tropical SSTs that are damped towards climatology to force at atmosphere model. From this approach, one might expect some skill for forecasts over New Zealand, in light of the dependence of New Zealand seasonal climate variations on tropical SST forcing. As in the earlier study, we do not directly address the problem of downscaling. That is, we do not try to infer local New Zealand temperatures and rainfalls from the contemporaneous global forecast fields. Instead, we simply examine the global forecasts to see when and where they appear to have skill. Only if these show clear skill in regions that influence New Zealand climate will it be worthwhile to attempt downscaling.

2. Data And Methods

Two data sets were used: forecasts produced by the Coupled Modelling Branch of NCEP, and corresponding analyses derived from the NCEP/NCAR reanalysis data set (Kalnay et al. 1996), the latter being interpreted as the "truth" against which the former were compared. Both data sets are for three-month periods and are expressed as anomalies from 1961–1990 climatologies (30-year means for each three-month period).

The forecasts cover the nearly three-year period from when the forecasting system "went live" (December 1996), up until a break in transmission caused by a computer room fire at the end of September 1999. They were generated in two steps: first, tropical Pacific sea surface temperatures (SSTs) were forecast using a coupled atmosphere/ocean model; then atmospheric climate was forecast with an atmospheric general circulation model forced by SSTs. At the second step, SSTs outside the tropical Pacific were started with observed anomalies and then damped towards climatology with an e-folding time of 90 days (for other tropical regions) or 30 days (for mid latitudes).

Natural variability	*nvar*	$\sqrt{\text{mean}\left(a_{is}^2\right)}$
Bias	*bias*	$\dfrac{\text{mean}\left(f_{isl}\right)}{nvar}$
Skill	*skill*	$100 - \dfrac{100}{nvar}\sqrt{\text{mean}\left(f_{isl} - a_{is}\right)^2}$
Anomaly correlation	*acor*	$\dfrac{\text{mean}\ (f_{isl} a_{is})}{nvar\sqrt{\text{mean}\left(f_{isl}^2\right)}}$

Table 1: Scores used to evaluate the forecasts (f_{isl} refers to the forecast anomaly with lead l made for the ith three-month period at grid-point s, and a_{is} refers to the associated analysis anomaly). Means are calculated over time, i, and/or space, s.

Forecasts are means of a 20-member ensemble whose individual members have different observational initial conditions. (For further information on the modelling system, see http://www.emc.ncep.noaa.gov/research/cmb/atm_forecast/ and associated links).

Each month, starting in December 1996, four forecasts were made for overlapping three-month periods, with the first four forecasts covering the months JFM, FMA, MAM, and AMJ of 1997 (where JFM refers to January-February-March, etc). These forecasts will be referred to as having leads 1, 2, 3, and 4 months, respectively. The last forecast in the present data set was made in September 1999 and there are no forecasts for any period past OND 1999. Thus, there are a total of 130 forecasts, made over 34 months (specifically, there are 4 forecasts made in each month between December 1996 and June 1999, then 3 forecasts made in July 1999, 2 in August 1999, and 1 in September 1999). The analyses cover the 34 overlapping periods from JFM 1997 to OND 1999, inclusive.

	Mid-point of latitude band							
	78° S	56° S	33° S	11° S	11° N	33° N	56° N	78° N
z	60	54	49	36	36	45	52	68
t	2.9	1.2	0.7	0.8	0.7	1.0	1.5	2.6
r	22	18	26	53	52	28	23	12

Table 2: Natural variability, nvar, for z (200 hPa height, m), t (temperature, °C), and r (rainfall rate, mm/month), calculated for each of eight bands of latitude.

Three forecast variables were considered:

z = 200 hPa height anomaly (m)

t = temperature anomaly at 2 m (°C)

r = rainfall rate anomaly (mm/month)

and these are global, on a 64 (latitude) x 128 (longitude) grid (roughly 2.8° x 2.8°). Anomalies are from the 1961–90 NCEP/NCAR climatology.

For some of the results presented below, forecasts were grouped by “season forecast”. In this grouping, the season of each forecast was determined according to which of the usual austral seasons it most overlaps. Thus, forecasts for the period NDJ, DJF, and JFM are all labelled as summer forecasts; FMA, MAJ, and AJJ are labelled as autumn, etc.

Four types of scores were calculated to evaluate the forecasts (Table 1). The main scores were *skill* and *acor*. The former is based on root-mean-square (rms) error and both are relative to climatology (so that a positive score implies a forecast better than climatology). A perfect forecast would have *skill* 100 and *acor* 1. The third score is *bias* which can be useful in

	Global						Near New Zealand					
	Model				Persistence		Model				Persistence	
lead:	1	2	3	4	1	4	1	2	3	4	1	4
skill												
z	-3	-5	-6	-5	-4	-20	-13	-19	-20	-20	33	23
t	1	-1	-1	-1	-10	-28	-0	-5	-7	-8	-4	-23
r	-11	-12	-13	-14	-6	-24	-17	-17	-16	-15	-25	-35
acor												
z	0.27	0.23	0.22	0.22	0.47	0.33	0.08	-0.03	-0.07	-0.10	0.82	0.78
t	0.19	0.17	0.17	0.18	0.40	0.22	0.21	0.13	0.08	0.08	0.50	0.32
r	0.28	0.25	0.22	0.20	0.46	0.27	0.01	0.00	0.00	-0.01	0.24	0.12
bias												
z	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.8	-0.9	-0.9	-0.9	-0.1	-0.1
t	-0.2	-0.2	-0.2	-0.2	0.0	0.0	-0.2	-0.3	-0.3	-0.3	-0.1	-0.2
r	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	-0.1	-0.1	-0.1	0.0	0.1

Table 3: Global and near-New Zealand scores (skill, acor, and bias) for model and persistence forecasts of z (200 hPa height), t (temperature), and r (rainfall rate), with various lead times

determining whether a forecast with low *skill* is inaccurate (high *bias*, of either sign) or simply imprecise (low *bias*). *nvar* is the fourth score and is calculated simply as a step in the calculation of the other three scores. The fact that natural variability shows strong latitudinal variation (Table 2) underlines the impotance of scaling scores by *nvar.*

All scores can be calculated in different versions by averaging across different subsets of the set of forecasts. The following versions are used below:

global	averaged across all values of *i* (time) and *s* (space)
near NZ	averaged across all *i*, for *s* within latitudes 30° S to 50° S and longitudes 142° E to 172° W
by latitude band	averaged across all *i*, and all *s* in one of eight equal latitudes bands
by season forecast	averaged across all *s*, for *i* in a given season
by grid point	averaged across all *i*

Scores were also calculated for persistence forecasts with leads 1 and 4 months. In these forecasts the analysis anomaly for months JFM, say, is taken as a forecast of the anomaly for AMJ (for lead 1) or JAS (for lead 4), etc.

No attempt was made to calculate statistical significance for scores. Such calculations are not very reliable for these sorts of scores because they depend on distributional assumptions (particularly independence) that are not correct. Also, our interest was more in the patterns amongst these scores (e.g., are model scores usually positive or negative, and are they usually greater than or less than those for persistence) than in their actual values.

3. Results

The global scores (averaged across all seasons and grid points) for the model forecasts are not promising (Table 3). *Skill* is almost always negative (the only exception being for *t* in lead 1, where it is barely positive). Thus, according to this measure, the model forecasts have no more skill than climatology (on a global scale), though they are almost always better than persistence. The anomaly correlations for all model forecasts are positive, but the scores (between 0.17 and 0.28) are all substantially less than those for persistence (0.22 to 0.47), particularly at lead 1. This is in contrast with the *skill* scores.

Scores for the model forecasts in the area near New Zealand are almost all worse than the global scores (despite the fact that, for *z* and *t*, persistence scores better during this period than it does globally). For *z*, there was substantial negative bias near NZ.

The *acor* scores for persistence may be misleading because of the short period covered by the data. In periods where the analysis anomalies are predominantly of the same sign, *acor* will be inflated. This is seen most strongly for *z* near New Zealand, where the mean analysis anomaly was 54 m, comparable with its rms amplitude of 67 m, which caused very high *acor* scores for persistence (0.82 and 0.78, for leads 1 and 4). In comparison the mean forecast anomalies for this region were close to zero, so the forecasts show substantial negative bias and poor *acor* scores.

An interesting feature of these (and other) results concerns the rate at which forecast performance (as measured by *skill* and *acor*) degrades with increasing lead time. Clearly some degradation is to be expected. What stands out is that the estimated performance of the model forecasts degrades markedly more slowly than does that of persistence forecasts. For example, the global *skill* of persistence forecasts of *z* declines from –4, at lead 1, to –20, at lead 4, whereas the model forecasts only decline from

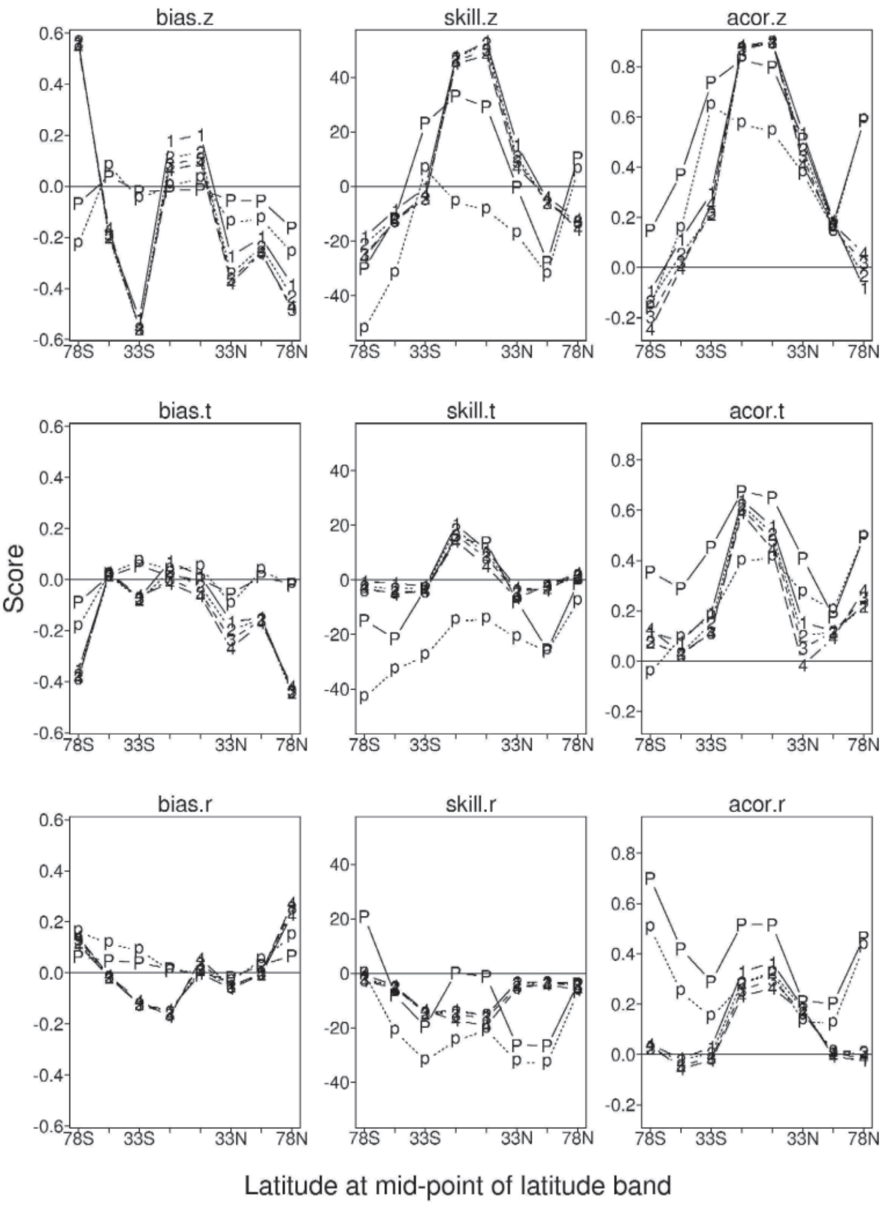


Figure 1: Scores (bias, skill, and acor) calculated for each of eight latitude bands for model forecasts with leads 1 to 4 (plotting symbols '1' to '4'), and also for persistence forecasts with leads 1 ('P') and 4 ('p'). The quantities forecast are z (200 hPa height), t (tempretaure), and r (rainfall rate).

	Model				persistence	
lead:	1	2	3	4	1	4
z in 21° S to 21° N, west of 150° E or east of 120° W						
skill	57	54	51	48	36	-4
acor	0.9	0.91	0.91	0.92	0.84	0.59
r in 14° S to 10° N, 80° W to 120° W						
skill	61	57	53	48	42	7
acor	0.92	0.92	0.92	0.90	0.88	0.62

Table 4: Forecast scores for regions of highest skill for z(200 hPa height) and t (temperature).

3 to 5. Similar differences are apparent for the other variables, *t* and *r*, and also for *acor*. This contrast in behaviour is probably caused by the generally much lower amplitude of the forecast anomalies, compared to those of the analysis. For example, the rms amplitude, averaged over space and time, of the analysis for *z* is 52 m, compared to 32 m for the forecasts (the latter figure varies little with lead).

All scores show strong variation with latitude, with most being better in the tropics than elsewhere (Figure 1), which is consistent with results from other studies (Rowell 1998, Frederiksen 2001). The most promising model results are for *z* in the tropics. Here, *skill* exceeds 40 and *acor* exceeds 0.85 for all leads, and these results are better than those for persistence. Elsewhere, values of *skill* and *acor* are either modest (or negative), or are not clearly better than those for persistence. Model forecasts were worst for *r*, with skill being negative at all latitudes. Figure 1 also shows very clearly how much more the persistence forecasts degrade from lead 1 to lead 4 than do the model forecasts.

Some seasonal variation in scores was found, but there is no combination of variable and season for which skill is substantial. For both *z* and *t*, *skill* is positive in winter, but only just. These scores are not shown here because it

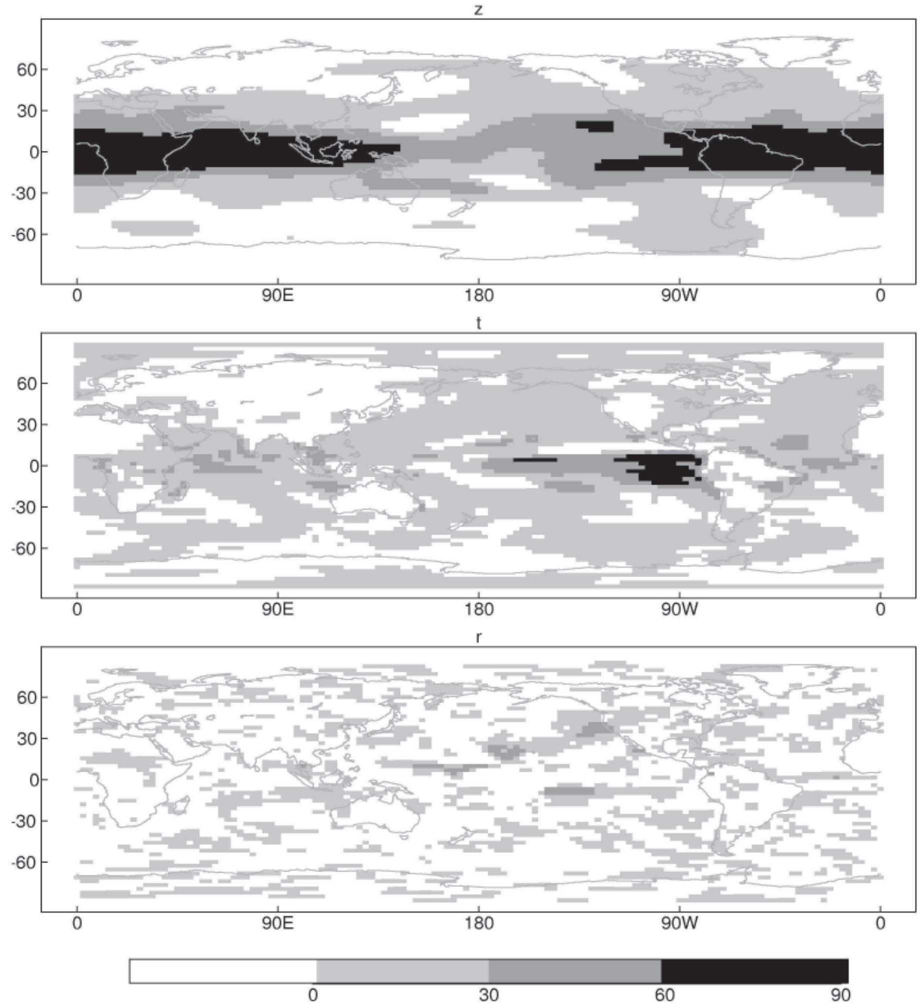


Figure 2. Geographical disdtribution of skill for model forecasts, with lead 1, of z (200 hPa height), t (temperature), and r (rainfall rate). Unshaded regions are those with negative skill.

is probably unrealistic to expect to define clear seasonal patterns from a set of forecasts covering only three years. Francis et al (2003) were unable to demonstrate any statistically significant seasonal variation in skill with a four-year set of forecasts.

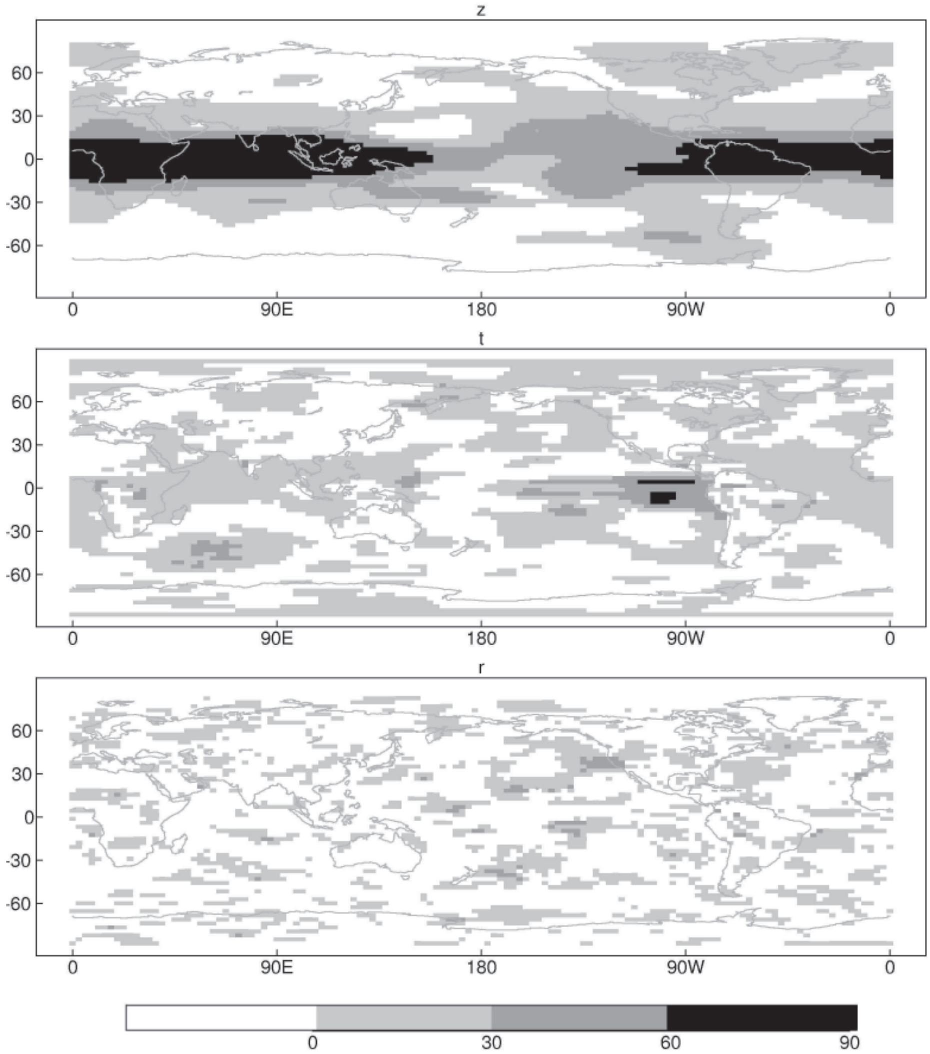


Figure 3: As for Figure 2, but for forecast lead 4.

The geographical areas of substantial *skill* are: for *z*, all of the tropics except the western Pacific; and for *t*, the eastern tropical Pacific (an area much affected by El Niño/La Niña oscillations) (Figure 2). In these two areas the model forecasts

		Model				Persistence	
	lead:	1	2	3	4	1	4
z							
	skill	62	60	59	58	40	-8
	acor	0.93	0.94	0.95	0.96	0.86	0.56
t							
	skill	48	43	39	34	42	-3
	acor	0.89	0.88	0.87	0.86	0.88	0.55
r							
	skill	34	28	24	19	31	-15
	acor	0.79	0.75	0.72	0.68	0.80	0.43

Table 5: Scores (skill and acor) for model and persistence forecasts of first principal components of z (200 hPa height), t (temperature), and r (rainfall rate) in the southern hemisphere to 60° S.

show more skill than persistence (Table 4). The areas of apparent high skill for r are so small and scattered that they are probably modelling artefacts. The geographical patterns of skill are very similar for all forecast leads (compare Figures 2 and 3).

Predicting principal components

It might be expected that that large-scale patterns (as shown by the first few principal components, or by empirical orthogonal functions (Wilks 1995)) would be more predictable than values at most individual grid points. To test this, we calculated scores for the first five principal components (PCs, calculated using the covariance matrix) of (three-month means of) the analysis data sets for all three variables (*z*, *t*, and *r*). Because of our interest in regions that might affect New Zealand, the PCs were calculated just for the southern hemisphere north of 60° S.

It was only for the first PCs that the model forecasts showed substantial skill, and only for *z* was this markedly greater than that shown by persistence (Table 5). For *z* (but not for *t*) scores for the first PC were better than those found in the high-skill region of Table 4. The first PC for *z* accounts for 35% of the southern hemisphere variation (analogous figures for *t* and *r* are 40% and

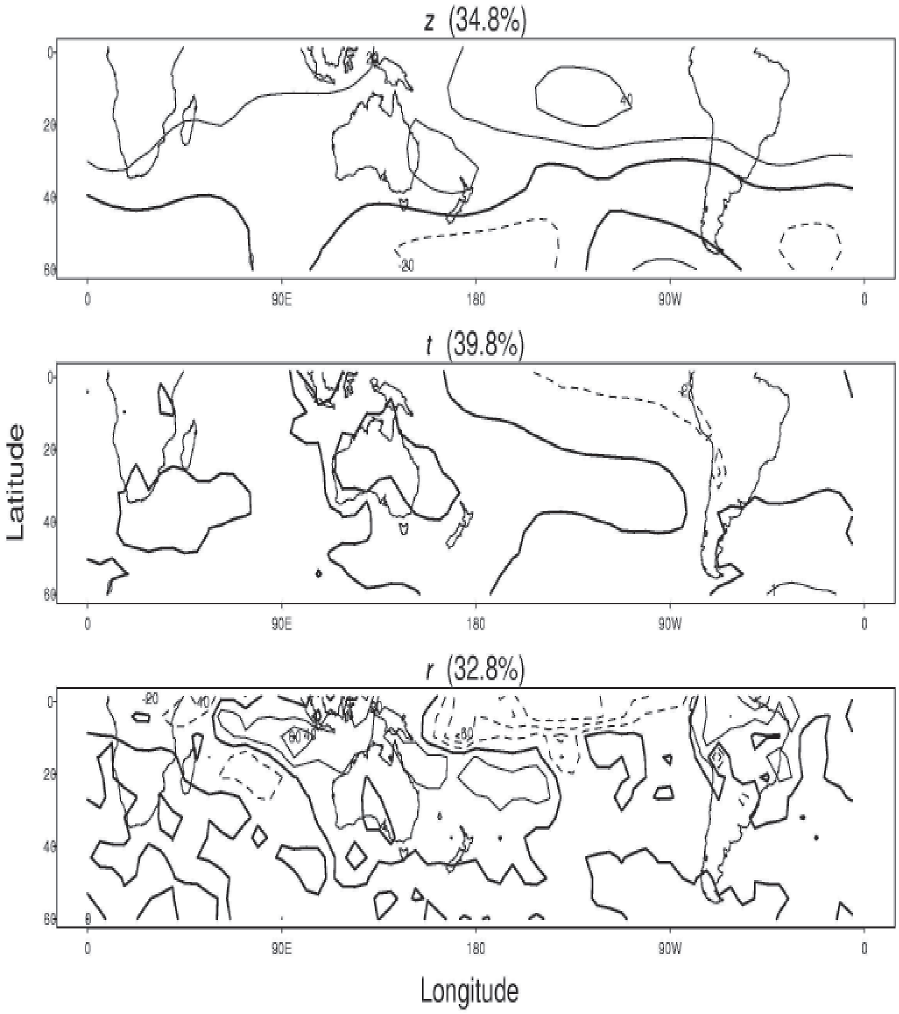
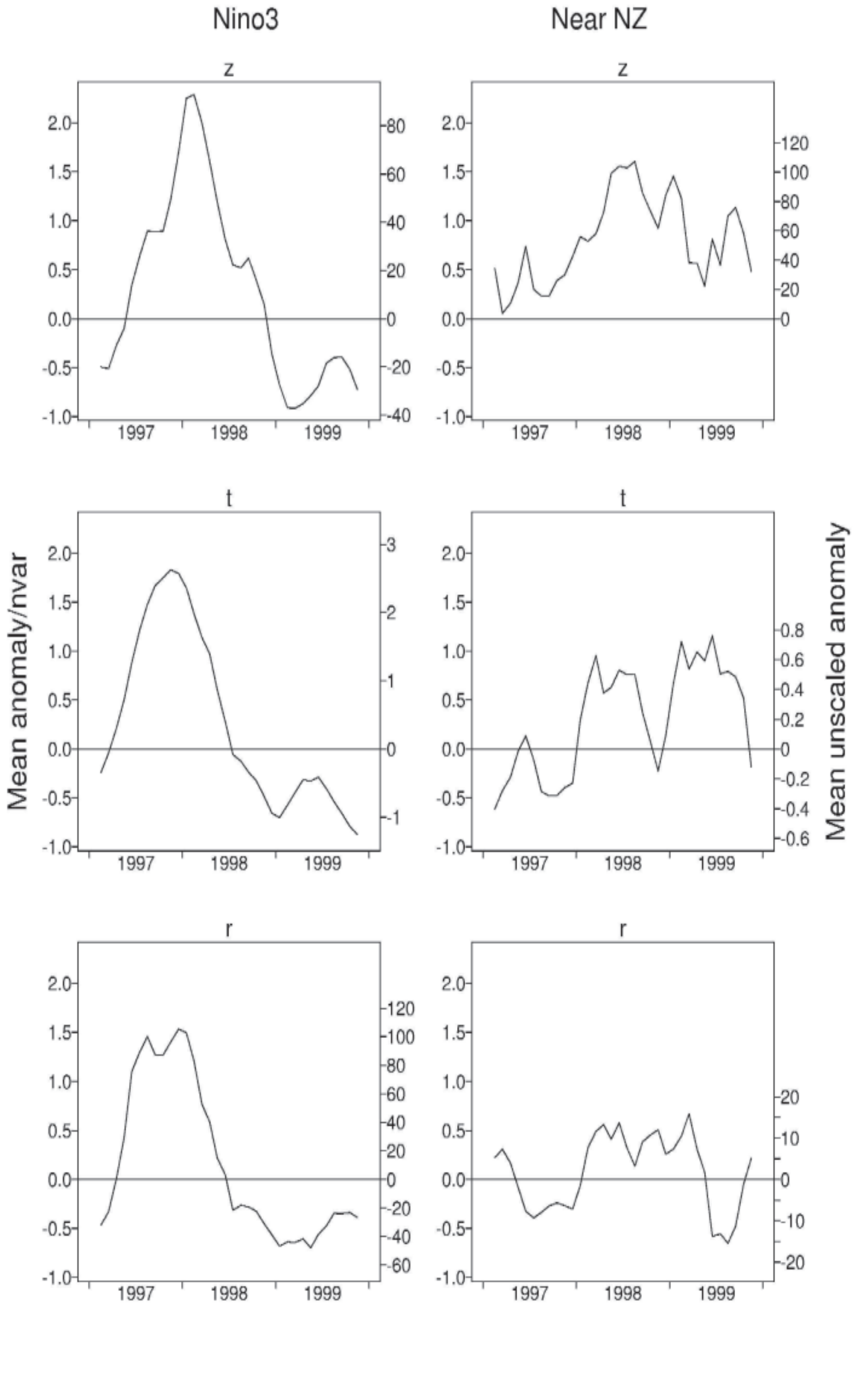


Figure 4. Leading empirical orthogonal functoions (EOFs) for z (200 hPa height), t (temperature), and r (rainfall rate). The percentage of variance explained is given above each panel. Zero isopleths are plotted as thick lines: positive isopleths as thin solid lines, and negative isopleths as thin broken lines.

Figure 5: Mean anomalies of z (200 hPa height), t (temperature), and r (rainfall rate) in two areas: Niño3 (5° S to 5°N, 90° W to 150° W; left panels) and near New Zealand (right panels). In each panel the left-hand scale shows the "scaled" anomalies (mean anomaly divided by nvar); the right-hand scale shows the natural, unscaled, units (i.e. m for z, °C for t, and mm/month for r).


Nino3
Near NZ
z
t
r
Mean anomaly/nvar
Mean unscaled anomaly
1997
1998
1999

33%). The leading principal component patterns (Figure 4) all appear to be ENSO-related. The leading pattern for z describes variations in the zonal wind in the region of the subtropical jet, and is associated with modulation of zonal winds over New Zealand. The leading *t* pattern shows the typical ENSO cold/warm tongue in the eastern tropical Pacific and anomalies of opposing sign over the southwest Pacific. The leading *r* pattern is mostly associated with east-west variations across the equatorial Pacific and in the region of the South Pacific Convergence Zone.

This analysis was repeated with PCs calculated for the restricted area near New Zealand (latitudes 30° S to 50° S and longitudes 142° E to 172° W). For these PCs, the skill of the model forecasts was never large and often negative.

4. Discussion

Compared to our benchmarks of climatology and persistence, we found only limited skill in the NCEP forecasts. The best results are for tropical 200 hPa heights outside the Pacific, and temperatures in the eastern tropical Pacific. Rainfall is not forecast well anywhere, apart from ENSO-related variations across the tropical Pacific. For the New Zealand region, prospects for direct use of global model output (for local "downscaling")are not bright. Given that forecast tropical SSTs (and associated air tempera-tures) do show skill, the best approach may be a combination of dynamically forecast tropical sea temperatures, coupled to statistical relationships with New Zealand climate parameters (a variant of the kind of approach used in Francis and Renwick 1998).

It is perhaps still too soon to expect much skill from extended-range global forecasts like those examined here, because these have been produced in an operational way for less than a decade. It seems likely that skill will increase somewhat as model resolution is increased and as more experience is gained.

However, the important role of internal (chaotic) variability in the extratropical circulation suggests that forecast skill in the New Zealand region may never be very high (e.g., Zheng et al. 2000). It is at least encouraging that the present results are more positive than those of Francis et al. (2003). They found that UKMO forecasts of southern hemisphere MSLP showed some skill (greatest in the tropics) up to 15 days, but little skill thereafter.

Our analyses may slightly overstate the skill to be expected in forecasts from this source. There are grounds to believe that forecast skill will be greater when the SOI is most extreme (Brankovic and Palmer 1997, Francis et al.2003). The period covered here is one dominated by extremes of the ENSO cycle, as can be seen from the mean anomalies in the region Niño3 (Figure 5, left panels). Forecast skill may be less in times of more moderate SOI values. The high scores for persistence forecasts of *z* in the near-NZ region (Table 3) perhaps relate to the fact that the analysis for this variable was positive throughout the period (Figure 5, top right panel).

Acknowledgments

We are grateful to Wanqui Wang of the Coupled Modeling Branch, Environ mental Modeling Center, NCEP for providing the forecasts and supporting information, and to John Kidson and two anonymous referees, for helpful comments. This work was funded by the Foundation for Research, Science and Technology under contract C01X0030.

References

Brankovic, C. and T. N. Palmer 1997: Atmospheric seasonal predictability and estimates of ensemble size. *Monthly Weather Review*, **125,** 859-874.

Francis, R. I. C. C., A. B. Mullan and J. A. Renwick 2003: An evaluation of UKMO one-month ensemble forecasts of MSLP in the Southern Hemisphere. *Theoretical and Applied Climatology*, **75**, 1-14.

Francis, R. I. C. C. and J. A. Renwick 1998: A regression-based assessment of the predictability of New Zealand climate anomalies. *Theoretical and Applied Climatology*, **60**, 21-36.

Frederiksen, C. S. 2001: Dynamical seasonal forecasts during the 1997/98 ENSO using persisted SST anomalies. *Journal of Climate*, **14,** 2675-2695.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, M. Chelliah, W. Ebisuzaki, W. Higgins, J. Janowiak, K. C. Mo, C. Ropelewski, J. Wang, A. Leetmaa, R. Reynolds, R. Jenne and D. Joseph 1996: The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, **77**, 437-471.

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

Madden, R. A., D. J. Shea, R. W. Katz, and J. W. Kidson, 1999: The potential long-range predictability of precipitation over New Zealand. *Int. J. Climatol.* **19,** 405-421.

Mullan, A. B., 1998: Southern Hemisphere sea surface temperatures and their contemporary and lag association with New Zealand tempera-ture and precipitation. *Int. J. Climatol.* **18,** 817-840.

Renwick, J. A., R. E. Basher, A. B. Mullan, M. J. Salinger, C. S. Thompson and R. I. C. C. Francis 1999. Procedures for the production of monthly and seasonal climate outlooks. NIWA Internal Report. 56. (Unpublished report held in the NIWA library, Wellington.). 38 p.

Renwick, J. A., A. B. Mullan and C. S. Thompson 2000. Objective climate prediction at NIWA. Proceedings, ANZ Climate Forum, CSIRO, Hobart, 10-12 April 2000

Rowell, D. P. 1998: Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. *Journal of Climate*, **11**, 109-120.

Wilks, D. S. 1995: *Statistical Methods in the Atmospheric Sciences: an Introduction*. Academic Press, San Diego. 467 pp.

Zheng, X., H. Nakamura and J. A. Renwick 2000: Potential predictability of seasonal means based on monthly time series of meteorological variables. *Journal of Climate*, **13**, 2591-2604.

Zheng, X. and J. A. Renwick, 2003: A regression-based scheme for seasonal forecasting of New Zealand temperature. *J. Climate*, **16,** 1843-1853.

Submitted to Weather and Climate September 2003, Revised September 2004