How good are NIWA's 15-day weather forecasts[1]?

In Summary

Daily values from NIWA's 15-day weather forecasts exhibit useful levels of skill through approximately the first week of the forecast interval, as expected from general considerations of atmospheric predictability. In addition, five-day averaged forecasts generally exhibit some skill out to the day 6-10 average. Forecasts of whether the week will be generally wet (more than half the days being rain-days) or dry were found to be marginally skilful through week two of the forecast. These results suggest that while the skill of daily rainfall forecasts beyond five days may be low, there is skill in predicting periods of rain, even though the day to day timing and amounts may not be skilful.

The 15-day forecasts exhibited greater skill compared with predictions based only on knowledge of the weather over the last few days together with knowledge of the average conditions for the time of year. This was true in almost all cases out to day eight. At day three, the average improvement in explained variance (over knowledge of the time of year) was around 27%. At day-eight, the average improvement was around 6%.

Graphical displays of the forecasts (as illustrated in Fig. 1 below, but not including the observed data) were made available to a small group of potential users from late 2004. There was general agreement that the visual display used, clearly illustrating the forecast trend in comparison to climatology, would be useful in a qualitative sense as input to decision-making in the agriculture and other climate-sensitive sectors. For example, decisions around the timing of fertilizer application, or about the deployment of a work force for outdoor maintenance work, could be informed by qualitative indications of the likelihood of rain, or of a period of high or low temperatures over the coming 10-15 days. Some users are presently trialling the use of 15-day forecasts for operational decision-making.

The forecast model

A forecast for the next 15 days is potentially of great interest to user communities as it is the time scale of much operational planning for weather-affected industries (e.g., construction, renewable energy generation, agriculture). Generally, climate prediction focuses on the seasonal (three-month) scale or longer and deals in averages. In contrast, weather predictions describe the daily sequence of weather over the coming few days to one week, after which chaotic effects come to dominate the day-to-day forecasts.

The 15-day forecast model developed by NIWA is based on a set of statistical relationships between observed large-scale (2.5° latitude/longitude grid)

[1] This synopsis has been drawn from the published article: Renwick, J., Mullan, B., Thompson, C. and Porteous, A., 2009: Downscaling 15-day ensemble weather forecasts and extension to short-term climate outlooks. *Weather and Climate*, 29, p 45-69.

meteorological fields (temperature, winds, etc) over the New Zealand region and local surface weather observations measured at climate stations throughout the country. These relationships are then applied to the meteorological fields from a weather prediction model (at the same 2.5° latitude/longitude grid scale) to specify the local surface weather expected to occur, given the forecast large-scale pattern.

The statistical relationships (and hence the 15-day forecasts) were defined seasonally for six climate parameters: precipitation amount, maximum and minimum temperature, 10cm earth temperature, solar radiation, and wind run; and for eight event probabilities: 24h precipitation greater than 0, 1, 5, 10, 20 mm, 24h precipitation greater than median amount for the time of year, maximum temperature greater than 25°C, and minimum temperature less than 0°C.

Currently, the NIWA system uses 21 forecast model runs for every day (each with slightly different initial "perturbations") produced by the National Centers for Environmental Prediction (USA), to produce the median, lower quartile and upper quartile forecasts as shown in Fig. 1. Use of this "ensemble" of 21 forecast runs gives a direct estimate of the uncertainty of the forecast, i.e. the predictability of the weather over the coming two weeks.

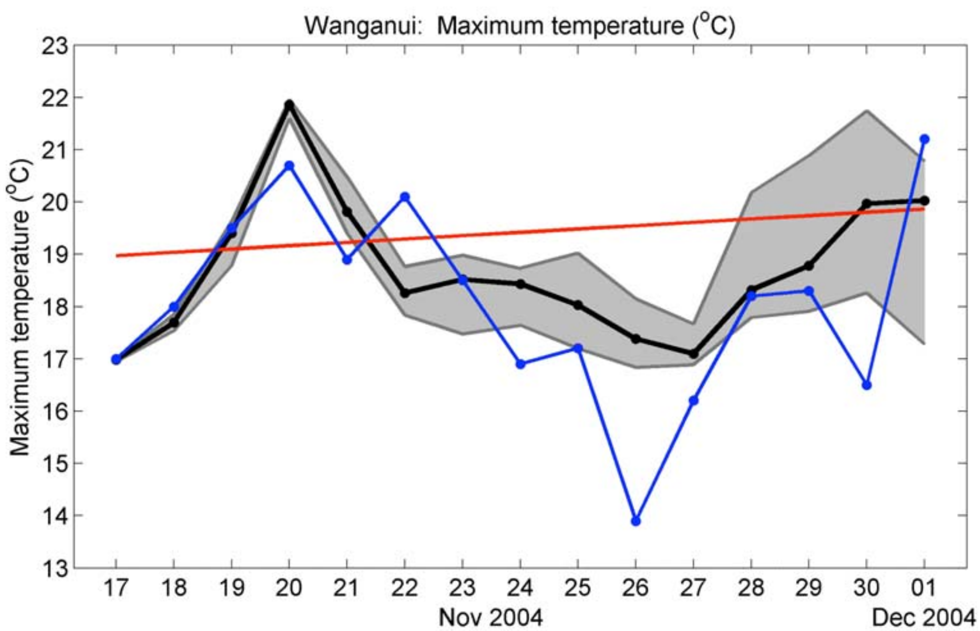

Figure 1: Example plot of a 15-day downscaled maximum temperature forecast from mid-November 2004, for Wanganui. The heavy black line indicates the median of the forecast ensemble, the grey shaded area encompasses the inter-quartile range of the ensemble distribution, the red line is the daily climatological maximum temperature for Wanganui, and the blue line is the observed maximum temperature for the period indicated.

Fig. 1 illustrates a number of typical features of the forecast system:

- The inter-quartile range (IQR) increases with forecast interval, as the ensemble spread increases;

- About 50% of the observed values fall within the IQR envelope, as expected, although the IQR appears to be underestimated over the first few days of the forecast period;
- The median forecast is quite accurate over the first 5 or so days of the forecast interval;
- The trend in the forecast (e.g. above climatology on days 3-5 but tending below climatology during week two) is representative of the observations, although errors on individual days may be large, beyond day 5.

Evaluation of the forecast skill

Skill statistics were calculated for "persistence" forecasts, using current observations as the forecast for each day over the coming fifteen (i.e. armed only with the knowledge of the weather over the last few days and the time of year, how well can you predict the weather over the next two weeks?) The 15-day forecasts exhibited greater skill (explained variance) than persistence forecasts in almost all cases out to day eight. At day three, the average improvement in explained variance, compared to persistence forecasts, was around 27%. At day-eight, the average improvement was around 6%.

To further gauge the statistical significance of forecast skill, the forecasts were compared to a large set of trials where observations from the historical record were selected at random and used as "forecasts". This gives us a distribution of skill scores for "random" forecasts. Skill from the NIWA 15-day forecasts was greater than the 95th percentile of the distribution of random forecasts in almost all cases out to day eight (indicating good skill). Beyond day 10, all measures of forecast skill were less than the 95th percentile of the distribution of random forecasts at more than half of the sites used. At day three, the skill of the forecasts was higher than the 95th percentile of the random distribution for all variables and all sites, apart from three of the 114 sites for predictions of the probability of precipitation greater than 10mm, and at three sites for the probability of maximum temperature greater than 25°C.

By the day-8 forecast, the skill of the forecasts was less than the 95th percentile of the random distribution at around one third of sites for the above two parameters. For the non-probabilistic forecasts (i.e. precipitation amount, maximum and minimum temperature, 10cm earth temperature, solar radiation, and wind run), the day-8 explained variance exceeded the 95th percentile from the random trials at all sites for all temperature variables (max, min, 10cm Earth), at all but two for wind run, at all but seven for solar radiation, and at 90 sites (all but 24) for precipitation amount.

Five-day mean forecasts show somewhat more skill than forecasts for individual days, but the skill of the 5-day mean forecast was often similar to the mean of the skill of the forecasts for each of the five days. The exception is for precipitation, where the 5-day mean forecasts are noticeably more skilful than the individual day forecasts that go into the mean. This may be due to the more "noisy" nature of daily precipitation, compared to temperature. Multi-day averages smooth out some of the random variability and leave a more obvious slowly-varying signal. In other words, the results suggest that while the skill of daily rainfall forecasts may be low, there is skill in predicting periods of rain, even though the day to day timing and amounts may not be skilful.

To investigate the above idea further, a series of trials were carried out with the precipitation probability forecasts. For each forecast day, the prediction was classed as "wet" if the ensemble median probability of more than 1 mm of precipitation was greater than the climatological average probability for the time of year. Otherwise, the prediction for that day was classed as "dry". Forecasts were then grouped into 5- or 7-day periods predicted to be "wet" if more than half the days (i.e. at least 3 or 5, respectively) were predicted to be "wet". The observations were similarly grouped and labelled wet or dry. The hit rate (percentage of forecasts correctly predicting a "wet" or "dry" period) for the days 3-7 5-day period, and the days 8-14 7-day period ("week two") is illustrated in Fig. 2, for a two year trial period (2004-2006).

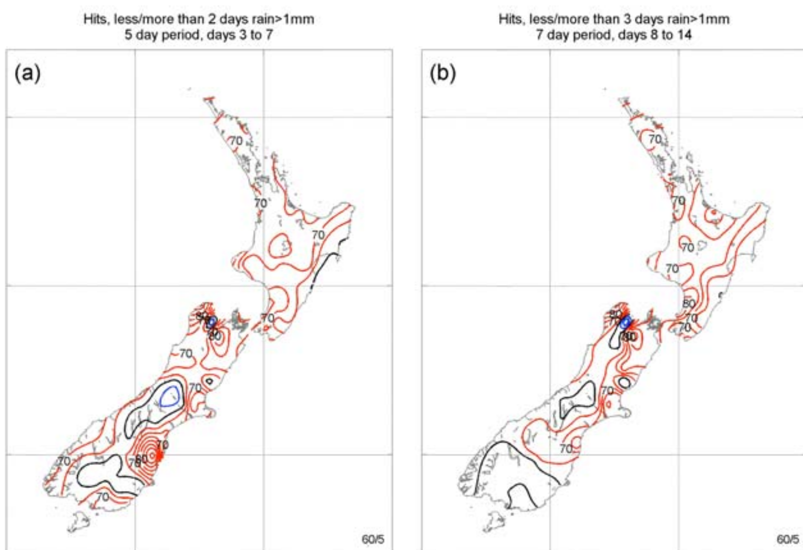

Figure 2: Hit rate (percent) for predictions of wet spells (more than half the period being rain days); (a) of days 3-7 having more than 2 wet days, and (b) of days 8-14 having more than 3 wet days. The contour interval is 5%, with values less than 60% in blue, values greater than 60% in red, and the 60% contour in black.

The forecasts score a "hit" on about 70% of occasions, on average. However, the hit rates illustrated in Fig. 2, and the associated skill scores, are only a small average improvement on those for persistence forecasts (using the frequency of wet days in the past week as the forecast for days 8-14). Forecast skill was less than that for persistence forecasts, and below the 95th percentile from the set of random forecasts, at around one third of stations. The stations where the forecasts performed poorly (compared to persistence and random predictions) were almost all in the east of the country, where conditions are driest on average.