Generating quantitative precipitation forecasts using a knowledge based system

Harvey Stern

Bureau of Meteorology, Australia

Preface

'Consider mechanically integrating judgmental and statistical forecasts instead of making judgmental adjustments to statistical forecasts … Judgmental adjustment (by humans) of (automatically generated statistical forecasts) is actually the least effective way to combine statistical and judgmental forecasts … (because) judgmental adjustment can introduce bias[1] (Mathews and Diamantopoulos, 1986) …The most effective way to use (human) judgment is as an input to the statistical process … Cleman (1989) reviewed over 200 empirical studies on combining and found that mechanical combining helps eliminate biases and enables full disclosure of the forecasting process. The resulting record keeping, feedback, and enhanced learning can improve forecast quality' (Sanders and Ritzman, 2001).

Introduction

Sanders and Ritzman (2001) highlight the difficulty associated with utilising (human) judgment as an input to the statistical process 'when the (human) forecaster gets information at the last minute'. In generating the predictions presented here, the strategy is therefore to take judgmental (human) forecasts (derived with the benefit of knowledge of all available computer generated forecast guidance), and to input these forecasts into a system that incorporates a statistical process to mechanically combine the judgmental (human) forecasts and the computer generated forecast guidance, thereby immediately yielding a new set of forecasts.

In this context, the purpose of the present work is to evaluate the new set of forecasts, and to document the increase in accuracy achieved by that new set of forecasts over that of the judgmental (human) forecasts.

Some 30 years ago, Snellman (1977) lamented that whereas the initial impact of guidance material was to increase the accuracy of predictions on account of a healthy human/machine 'mix', operational meteorologists were losing interest and that the gains would eventually be eroded by what he termed the 'meteorological cancer'. Snellman suggested that producing automated guidance and feeding it to the forecaster who 'modifies it or passes it on', encourages forecasters 'to follow guidance blindly' and concluded by predicting an erosion of recent gains. Hindsight informs us from forecast verification statistics that the erosion of gains did not take place. In fact, the accuracy of forecasts continued to increase - see, for example, Stern (2005a, 2005c). Nevertheless, evidence is emerging that the increasing skill displayed by the guidance material is rendering it increasingly difficult for human forecasters to improve upon that guidance (Mass and Baars, 2005; Ryan, 2005).

A knowledge based system

Over recent years, the present author has been involved in the development of a knowledge based weather forecasting system (Stern, 2002, 2003, 2004a, 2004b, 2005a, 2005b, 2005c, 2006). Various components of the system may be used to automatically generate worded weather forecasts for the general public, terminal aerodrome forecasts (TAFs) for aviation interests, and marine forecasts for the boating fraternity. The knowledge based system

---

[1] Stern (1996) documents forecaster over-compensation for previous temperature errors.

generates these products by using a range of forecasting aids to interpret NWP model output in terms of such weather parameters as precipitation amount and probability, maximum and minimum temperature, fog and low cloud probability (Stern and Parkyn, 2001), thunderstorm probability (Stern, 2004b), wind direction and speed, and swell (Dawkins, 2002). For example, Stern's 2005b forecasts in weather graphic format are generated from an algorithm that has a logical process to yield HTML code by combining predictions of temperature, precipitation, wind, morning and afternoon weather, and special phenomena (thunderstorm, fog), with features of the forecast synoptic type (strength, direction, and cyclonicity of the surface flow).

Stern (2005b) conducted a 100-day trial (Feb 14, 2005 to May 24, 2005) of the performance of the knowledge based system, with twice-daily forecasts being generated out to seven days in advance. During the trial, the overall percentage variance of observed weather explained by the forecasts so generated (the system's forecasts) was 43.24% compared with 42.31% for the official forecasts. That the knowledge based system achieved some success in its attempt to replicate the cognitive decision making processes in forecasting is confirmed by the closeness of the overall percentage variances explained by the two sets of forecasts. Specifically for precipitation, the percentage variance explained by the quantitative precipitation forecasts and probability of precipitation forecasts so generated was 26.78% compared with 25.07% explained by the official forecasts. On a rain/no rain basis, the percentage of correct forecasts so generated was 78.82% compared with 77.64% of the official forecasts.

However, the overall percentage variance of official forecasts explained by the system's forecasts was only 45.91%. This was made up of 63.59% of the variance of officially forecast temperature, and 28.23% of the variance of officially forecast precipitation. This indicates, that, on a day-to-day basis, there are significant aspects of the processes employed in deriving the official forecasts that are not taken into account by the system's forecasts (in all likelihood what Sanders and Ritzman (2001) refer to as 'domain knowledge'[2]), and vice versa.

Combining forecasts by mathematically aggregating a number of individual forecasts increases the reliability of forecasts (Kelley, 1925; Stroop, 1932) and averages out unsystematic errors (but not systematic biases) in cue utilization. A common method for combining individual forecasts is to calculate an equal weighted average of individual forecasts' (Stewart, 2001). However, under some conditions unequal weights make sense 'if you have strong evidence to support unequal weighting' (Armstrong, 2001b)[3]. Regarding the two sets of forecasts as partially independent and utilising linear regression to optimally combine the estimates of minimum temperature, maximum temperature, precipitation amount, and precipitation probability, Stern (2005b) demonstrated a lift in the overall percentage variance of observed weather explained. This result suggested that adopting such a strategy of optimally combining the official and system predictions has the potential to deliver a set of forecasts that are substantially more accurate than those currently issued officially. Indeed, the overall percentage variance of observed weather explained was lifted (by the consensus forecasts) to 50.21% from 43.24% (system) and 42.31% (official). Specifically for precipitation, the percentage variance explained was lifted (by the consensus forecasts) to 34.09% from 26.78% (system) and 25.07% (official),

[2] Sanders and Ritzman (2001) define 'domain knowledge' as 'knowledge practitioners gain through experience as part of their jobs' and make particular reference to that component of domain knowledge named 'contextual knowledge, which is the type of knowledge one develops by working in a particular environment.' 'The quality of domain knowledge is affected by the forecaster's ability to derive the appropriate meaning from the contextual (or environmental) information' (Webby et al., 2001).

[3] Krishnamurti et al. (1999) found that weather forecasts based on a combined forecast using weights based on regression were more accurate than combined forecasts with equal weights.

whilst on a rain/no rain basis, the percentage of correct forecasts was lifted to 83.55% from 78.82% (system) and 77.64% (official)[4].

Ongoing work

The knowledge based system has been modified so that it now automatically integrates judgmental (human) forecasts and the computer generated guidance, thereby incorporating the forecasters' valuable contextual knowledge into the process[5]. It is undergoing a 'real-time' trial, the results of which are being evaluated.

In conclusion, there is an increasing interest in the question of what might be the appropriate future role for the human in the forecast process (Stewart, 2005). The answer may be that the future role of human forecasts is as an input to a system that mechanically combines the human forecasts with the computer generated guidance

References

Armstrong, J. S., 2001a: Principles of forecasting: a handbook for researchers and practitioners. *Kluwer Academic Publishers*.

Armstrong, J. S., 2001b: Combining forecasts (refer to *Armstrong, 2001a, 417-439).*

Cleman, R.T., 1989: Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting, 5,* 559-583 (refer to *Armstrong, 2001a, 411).*

Dawkins, S. S., 2002: A web-based swell and wind forecasting tool. *9th Conference of the Australian Meteorological and Oceanographic Society, Melbourne, Australia, 18-20 Feb., 2002.*

Kelley, T. L., 1925: The applicability of the Spearman-Brown formula for the measurement of reliability. *Journal of Educational Psychology, 16,* 300-303 (refer to *Armstrong, 2001a, 95).*

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhan Zhang, Williford, C. E., Gadgil, S., and Surendran, S., 1999: Improved multi-modal weather and seasonal climate forecasts from multimodel "superensemble". Science, 285, 1548-1550 (refer to *Armstrong, 2001a, 423)*.

Mass, C. F. and Baars, J., 2005: The performance of National Weather Service forecasts compared to operational, consensus, and weighted model output statistics. *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*

Mathews, D. P. and Diamantopoulos, A., 1990: Judgmental revision of sales forecasts: effectiveness of forecast selection. *Journal of Forecasting, 6,* 407-415 (refer to *Armstrong, 2001a, 411).*

Ryan, C., 2005: Implemenation of operational consensus forecasts. *Bureau of Meteorology Analysis and Prediction Operations Bulletin No. 60.*

Sanders, N. R. and Ritzman, L. P., 1992: The need for contextual and technical knowledge in judgmental forecasting (refer to *Armstrong, 2001a, 406).*

Sanders, N. R. and Ritzman, L. P., 2001: Judgmental adjustment of statistical forecasts (refer to *Armstrong, 2001, 405-416*).

---

[4] The accuracy increases because 'Combining is most effective when the forecasts combined are not correlated and bring different kinds of information to the forecasting process' (Sanders and Ritzman, 2001) and that although 'both (human) intuitive and (computer) analytic processes can be unreliable … different kinds of errors will produce that unreliability' (Stewart, 2001).

[5] Sanders and Ritzman (2001) refer to their 1992 study, in which they demonstrated that judgmental forecasts based on contextual knowledge were significantly more accurate than those based on technical knowledge (and) … were even superior to (a) … statistical model.'

Snellman, L. W., 1977: Operational forecasting using automated guidance. *Bulletin of the American Meteorology Society, 58,* 1036-1044.

Stern, H., 1996: Statistically based weather forecast guidance. Ph. D. Thesis, School of Earth Sciences, University of Melbourne, subsequently published in 1999 as *Meteorological Study 43, Bureau of Meteorology, Australia.*

Stern, H., 2002: A knowledge-based system to generate internet weather forecasts. *18th Conference on Interactive Information and Processing Systems, Orlando, Florida, USA 13-17 Jan., 2002.*

Stern, H., 2003: Progress on a knowledge-based internet forecasting system. *19th Conference on Interactive Information and Processing Systems, Long Beach, California, USA 9-13 Feb., 2003.*

Stern, H., 2004a: Incorporating an ensemble forecasting proxy into a knowledge based system. *20th Conference on Interactive Information and Processing Systems, Seattle, Washington, USA 11-15 Jan., 2004.*

Stern, H., 2004b: Using a knowledge based system to predict thunderstorms. *International Conference on Storms, Storms Science to Disaster Mitigation, Brisbane, Queensland, Australia 5-9 Jul., 2004.*

Stern, H., 2005a: Using a knowledge based forecasting system to establish the limits of predictability. *21st Conference on Interactive Information and Processing Systems, San Diego, California, USA 9-13 Jan., 2005.*

Stern, H., 2005b: Defining cognitive decision making processes in forecasting: a knowledge based system to generate weather graphics. *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*

Stern, H., 2005c: Establishing the limits of predictability at Melbourne, Australia using a knowledge based forecasting system and NOAA's long range NWP model. *Submitted to Australian Meteorological Magazine.*

Stern, H., 2006: Combining human and computer generated weather forecasts using a knowledge based system. *Submitted to 22nd Conference on Interactive Information and Processing Systems, Atlanta, Georgia, USA 27 Jan. - 3 Feb., 2006.*

Stern, H. and Parkyn, K., 2001: A web-based Melbourne Airport fog and low cloud forecasting technique. *2nd Conference on Fog and Fog Collection, St John's, New Foundland, Canada 15-20 Jul.,2001.*

Stewart, N. A., 2005: Forum on the future role of the human in the forecast process. Part 2: Cognitive psychological aspects of expert forecasters (Chairperson: N. A. Stewart). *21st Conference on Weather Analysis and Forecasting; 17th Conference on Numerical Weather Prediction. American Meteorological Society, Washington, DC, 1-5 August, 2005.*

Stewart, T. R., 2001: Improving reliability of judgmental forecasts (refer to *Armstrong, 2001a, 81-106*).

Stroop, J. R., 1932: Is the judgment of the group better than the average member of the group? *Journal of Experimental Psychology, 15,* 550-560 (refer to *Armstrong, 2001a, 95*).

Webby, R., O'Connor, M, and Lawrence, M., 2001: Judgmental time-series forecasting using domain knowledge (refer to *Armstrong, 2001a, 81-106).*

Acknowledgments. To Stuart Coombs, for drawing attention to the skill displayed by the NOAA Global Forecasting System (used in the trial), to Marco Berghege for his suggestion regarding the programming utilised herein, to Neville Nicholls and Frank Woodcock, for their comments on combining forecasts, to Noel Davidson, for his comments on the NWP output, and to my other Bureau of Meteorology colleagues for their helpful discussions and feedback.