Comments on "Statistical Single-Station Short-Term Forecasting of Temperature and Probability of Precipitation: Area Interpolation and NWP Combination"

LANCE M. LESLIE

School of Mathematics, University of New South Wales, Sydney, Australia

MILTON S. SPEER

Bureau of Meteorology, Sydney, Australia

14 March 2001 and 4 July 2001

We offer the following comments on the recent paper by Raible et al. (1999), hereafter referred to as RB99. Their paper argues the case for implementing short-term (defined by RB99 as 0–24 h) statistical methods for single-station and areal forecasting. RB99 describe how they set up the statistical models, combine them with the relevant output from a regional NWP model, and how they introduce an areal interpolation procedure for operational usage on a Web site. Before beginning our comments, we wish to state that although we will be making some criticisms of RB99, we strongly endorse the general approach and recognize the pioneering efforts over the past two decades in this field of at least one of the authors (KF). Our motivation for writing these comments is to use RB99 as a springboard. We wish to generate an exchange of ideas in the literature, as an essential part of the health and vitality of this important area of applied research.

We begin with a point that often seems to be ignored by the advocates of purely deterministic procedures. Statistical techniques have a necessary role in very short-term forecasting and in our opinion will continue to do so for the foreseeable future. At whatever resolution, or grid scale, a deterministic model is run at, there is always a subgrid-scale. The ultimate subgrid-scale is, of course, a point and a point forecast can only be attained from a deterministic model by employing either an intelligent downscaling technique (e.g., one of the various forms of regression) or by interpolating in some manner. All procedures for obtaining subgrid-scale, or point, forecasts are statistical methods.

Focusing now on RB99, we make a series of comments to which we look forward to a response from RB99. First, the approach of RB99 is one of extreme simplicity and as a consequence has a very limited range of applicability. Simplicity is a laudable objective and we agree that it is an excellent guiding principle. However, we do not endorse oversimplification and wish to hear their views on this aspect. The RB99 method predicts just two (temperature anomalies and probability of precipitation, or PoP) of the very many key meteorological variables that are expected and required of a truly useful modern forecasting system. Moreover, by concentrating only on temperature anomalies and PoP at 6-hourly intervals, only a tiny subset of the wide range of possible predictands and their associated timescales is present in the forecasts. It is now very easy to generate a large number of meteorological variables at high temporal frequency, such as 15-min time intervals, compared with the 6-hourly output of RB99. We suggest that such a comprehensive set of meteorological variables output at high frequency is necessary for useful present-day applications. For example, to define what a given day's weather is truly like requires frequent updates of variables such as those defining an accurate diurnal cycle, maximum and minimum temperatures, relative humidity, wind strength, and precipitation amount, type, duration, and intensity, and so on. There are many applications of high temporal and spatial frequency model output of the major atmospheric variables. One example will suffice here. Public utilities, such as energy suppliers, require detailed meteorological information to estimate power load demand, or to alert maintenance crews to possible damage to power lines and other equipment. To obtain a suitable degree of accuracy, they also require these variables at high spatial and temporal resolutions. The limited, simple, and infrequent output from the RB99 approach is an

Corresponding author address: Lance M. Leslie, School of Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: L.Leslie@unsw.edu.au

inadequate basis for a power utility in their key functions of computing their demand, load, and cost estimate algorithms; risk management; and policy strategies.

As we see it, a second limitation highlighted by the RB99 approach is a failure to provide forecasts of the most destructive meteorological events of all, namely, severe and extreme weather. We recognize that RB99 did not claim to produce such a forecast system, but we are disappointed that they did not place their own forecast system in its very limited context, or comment on severe events. By severe and extreme weather we include heavy rain, flash flooding or prolonged flooding, thunderstorms with or without hail, wind strengths ranging from gale force through to storm and hurricane force, extreme temperatures and relative humidity, or sudden changes in these conditions. The above-mentioned list, which is by no means comprehensive, can generate great problems of many kinds. Such problems include loss of life and property through destructive weather; meteorologically related environmental stress on humans, animals, and crops; the formation of inversions that can produce fog and air pollution episodes; and so on. The RB99 approach, as it stands, cannot capture such events, with just three very simple predictive categories of cirrus cloud, stratus cloud, and whether it is raining. Moreover, their forecasts are at 6-hourly intervals, which often is longer than the life cycle of the severe weather events that actually produce the temperature anomalies or rainfall they are forecasting. So, vital information, such as prediction of duration and amount of precipitation is not present, only the probability of measurable precipitation. Finally, it is worthwhile noting that statistical procedures are inherently conservative and it is again well known that statistical procedures have problems in predicting extreme events.

A third feature of RB99 to which we would like their response concerns the categories they employ to represent their three mutually exclusive weather states. These categories are highly limited in that they do not apply to a large number of important weather events and to many regions of the earth. Over much of the globe, the RB99 weather states and predictands seem to be inadequate. For example, in the Tropics and subtropics, which represent considerably more than 50% of the earth's surface, and the majority of its population, a description of rainfall classes alone requires at least three to four categories, not just one as in RB99. For example, here in Sydney, rainfall studies involve light (typically less than 1 mm h^{-1}), moderate (1–5 mm h^{-1}), and heavy (greater than 5 mm h^{-1}). The heavy rainfall category is often further subdivided. In the deep Tropics, further categories are definitely needed. Even in the midlatitudes, there usually is a seasonal dependence with severe and extreme rainfall amounts produced by transition and warm season thunderstorms. Finally, turning to the World Meteorological Organization's official observational code categories, there are at least three rainfall categories reported by observing stations.

Our fourth comment concerns the apparent lack of originality in RB99. In fact, their approaches mostly date back almost 20 years, with some of the techniques taken directly from those developed jointly by one of their authors (KF), with one of the commentators (LML). The procedure developed by Miller and Leslie (1984, 1985) owes much to the earlier work of Fraedrich and Mueller (1983), but went considerably beyond it and was operational for a period of many years. The RB99 approach reintroduces the Miller and Leslie (1984, 1985) procedure with no apparent difference other than a reduction in weather categories and the introduction of some different covariates.

A fifth question that we are interested in hearing RB99 comment upon is the verification procedures that are appropriate for short-range single-station (point) forecasting of weather elements. Our procedure is to compute the *accuracy* of the predictions by direct comparison with observations. We then assess skill using a range of measures, depending on the predictand. However, RB99 still use persistence and climatology as benchmarks of skill. This is surprising, as we suggest that the days of using climate and persistence as benchmarks, relative to which *skill* is measured, have long gone for short-range predictions. They simply are too low in skill. For most events of interest, particularly severe or extreme weather, they are not of value as standards for a new scheme to be measured against.

The sixth, and final, point we wish to raise with RB99 is the concept of a ''potentially'' good operational areal prediction system. Unless we have misunderstood the approach outlined in RB99, they are simply using an interpolation procedure that might be economical, as claimed by RB99, but it cannot add much new information to the areal predictions. As such it is an example of what we referred to above as a mathematical procedure. To enhance the predictions, a range of interpolation procedures exist that introduce new information because they employ knowledge of the forecast errors at the stations, which can quickly be reconstructed with modern computers, even when there is a model change. In recent years, the level of computing power has grown dramatically. It is no longer a major impediment to recalculating the weights for statistically corrected models. The time taken even for a large number of stations and a period of several years to obtain the relevant corrections is no longer a major exercise.

We do not wish to act purely as critics of RB99, but to use their work as a basis for generating a discussion of possible limitations of and alternatives to such approaches and in encouraging the development of more useful alternatives. As such, it is appropriate for us to finish with a very brief description of one of our own large range of available statistical methods. A very effective statistical approach has been running operationally, twice daily at the University of New South Wales, for four years and has provided forecasts via the Web to a range of users. It produces point (i.e., single station) forecasts of temperature, dewpoint, wind, precipitation, and sea level pressure out to 7 days ahead, to a power utility, a fire service, and the Sydney Environmental Protection Agency. The Web site is accessed by the users and used by them as the meteorological input to their own programs. The horizontal resolution is 15 km and the meteorological variables are provided at the designated points every 15 min, as a time series. There are other shorter-term forecasts out to 24- and 48-h forecasts at even higher horizontal resolutions of 5 km and still at 15-min intervals. This range of variables, numbers of stations (over 800 for the fire service), and the 15-min frequency of the forecasts far exceed that described by RB99. The procedure used to generate the set of single station forecasts is to run the University of New South Wales NWP model, referred to as a high-resolution model, or HIRES (Leslie and Speer 1998), with a 1.5-h data cutoff. The point data are output at the required frequency, location, and for the requested range of variables. We note here that the HIRES model has been designed to have excellent phase and conservation properties (Leslie and Purser 1997). As such, the major errors are amplitude errors, which are highly suited to correction using linear regression techniques. In practice, the HIRES predictions are corrected using the model error obtained from the stations, over a period of many years. The corrections have a very dramatic impact on the forecast quality.

To finish on a positive note, underlining our wish for an exchange in the literature, we reiterate our belief that the methodologies of statistical and statistical–dynamical prediction remain alive and thriving. In our opinion, such techniques represent the best and possibly only option at present for many forecasts, particularly for extremely short-range, very long-range, and for point predictions. Given that situation, we encourage active researchers such as RB99 to expand their ideas, not restrict them to minor extensions of techniques developed essentially in the 1980s and early 1990s.

REFERENCES

Fraedrich, K., and K. Mueller, 1983: On single station forecasting: Sunshine and rainfall Markov chains. *Contrib. Atmos. Phys.,* **56,** 108–134.

Leslie, L. M., and R. J. Purser, 1997: A new semi-Lagrangian semi-implicit NWP model for operations and research: Formulation and performance in single and multi-processor computing environments. *Atmos.–Ocean*, Numerical Methods in Atmospheric and Oceanic Modeling: The André J. Robert Memorial Volume, 75–101.

——, and M. S. Speer, 1998: Short range ensemble forecasting of explosive Australian east coast cyclogenesis. *Wea. and Forecasting,* **13,** 822–832.

Miller, A. J., and L. M. Leslie, 1984: Short-term single-station forecasting of precipitation. *Mon. Wea. Rev.,* **112,** 1198–1205.

——, and ——, 1985: Short-term, single-station forecasting using linear and logistic models. *Contrib. Atmos. Phys.,* **58,** 517–527.

Raible, C. C., G. Bischof, K. Fraedrich, and E. Kirk, 1999: Statistical single-station short-term forecasting of temperature of temperature and probability of precipitation: Area interpolation and NWP combination. *Wea. Forecasting,* **14,** 203–214.