15.2

COMBINING OBJECTIVE AND SUBJECTIVE INFORMATION TO IMPROVE FORECAST EVALUATION

Michael P. Kay*[1] and Michael E. Baldwin[2]
[1]Cooperative Institute for Research in Environmental Sciences (CIRES) University of Colorado/
NOAA Research-Forecast Systems Laboratory Boulder, Colorado
[2]Cooperative Institute for Mesoscale Meteorological Studies (CIMMS) University of Oklahoma/
Storm Prediction Center/National Severe Storms Laboratory Norman, Oklahoma

1. INTRODUCTION

Traditionally, forecast verification has been used for a wide variety of purposes such as providing baseline statistics against which to judge other forecasts, justifying increased numerical model resolution, facilitating forecaster and forecast-system intercomparisons, and measuring the economic value of forecasts. Arguably the most important of all reasons for performing forecast verification is to determine the weaknesses, and also the strengths, of a set of forecasts and *provide meaningful feedback to those responsible for producing the forecasts*.

Perhaps the most overlooked component of the verification process is that of the forecaster. Realizing that the goal of verification is to improve forecasts it becomes obvious that the forecasters themselves must be an integral part of any verification exercise. In the case of computer-generated forecasts, the forecaster can be considered to be the model developer. All too often this is not the case and forecasters are only considered in hindsight and left to wonder how to interpret the verification results as well as how to relate these results to the forecasts that they produced. Clearly, a more integrated system involving the forecasts, observations, statistics, and the forecasters themselves can help to provide such a framework to improve forecasts.

The purpose of this paper is to discuss our experience combining subjective methods of forecast evaluation with objective methods of forecast verification to provide a more complete diagnosis of forecast performance. This work focuses on methods that may be applied to the verification of meteorological forecasts, particularly those produced by human forecasters rather than by automated means.

2. CURRENT PERSPECTIVE

Despite the extensive history of forecast verification, there remain numerous difficulties and frailties with the process such that results are often viewed with distrust and disdain from the users of the information. Even with significant advances in forecast verification the long-standing biases and misunderstandings have hindered its adoption as an important part of the forecast process. In most cases, verification is not used systematically in the forecast process but instead is used informally to a varying degree by some forecasters (Fig. 1). An important point to be made here is that not all forecasts produced are verified operationally or are even formulated such that they can be verified! Perhaps the most significant hurdle that will always remain is how to relate objective statistical information back to the meteorology and the forecasts themselves.

Forecast verification is, in objective form, intimately tied to statistics. Statistics provide the objectivity necessary to compare data, draw conclusions, test hypotheses, and make rational decisions. The link between verification and statistics cannot be removed and represents a significant challenge to be considered when including the persons responsible for the forecasts in the verification process. Operational meteorologists, who may have had some basic statistical training, cannot and should not be considered statisticians. Therefore, one of the challenges of forecast verification is to provide meaningful information that can be understood and applied by those responsible for the forecasts. To this end, recent work has been done in an attempt to improve the verification information given to users (Brooks et al. 1998, Baldwin et al. 2001, Brown et al. 2002).

One aspect of the forecast process that is often not considered in verification studies is the set of opinions and judgments that a forecaster formulates when creating a forecast but that are never explicitly stated or quantified. This is simply due to the fact that a forecaster weighs numerous factors and makes internal judgments concerning

---

* Corresponding author address: Michael P. Kay, NOAA/R/FS5 Boulder, CO 80305; e-mail: kay@fsl.noaa.gov

likely outcomes and the result of these thoughts is the forecast product itself. The forecast products however, generally contain only a very limited amount of those judgments and therefore a significant amount of information that may be used to gain insight into the forecasts themselves is lost. A more complete view of verification can be achieved by specifying and recording certain information when forecasts are being produced, and using that information in conjunction with objective verification techniques.

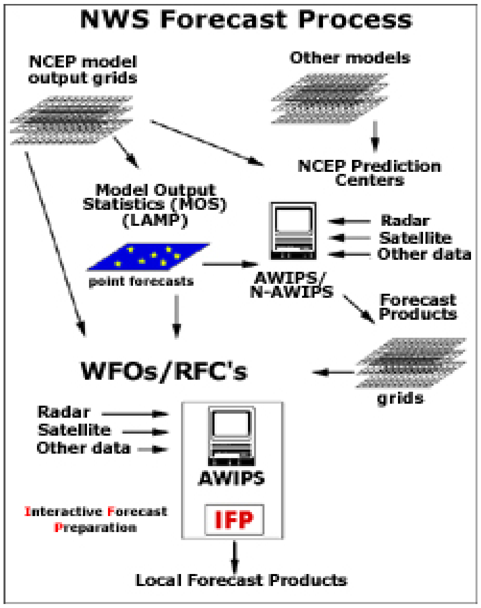


Figure 1. Schematic of the modernized National Weather Service forecast process. Verification is not shown on the figure. (Adapted from www.nws.noaa.gov/om/images/cartoon.gif)

The inclusion of subjective evaluation information with objective techniques is a non-trivial task for a number of reasons, such as specifying the relevant information to collect a priori, respecting time constraints of forecasters, and managing the data. Because of the difficulties involved, such activities have generally been undertaken only in conjunction with forecast experiments (Doswell and Flueck 1989, Jincai et al. 1992) and not in operational environments. The natural flexibility of forecasting experiments represents a perfect testbed for not only new types of forecasts, but also techniques and approaches. A point that should not be overlooked regarding the difference in experimental forecasts and operational forecasts is that experimental forecasts are often designed with verification in mind and as such are designed to be easily verifiable (Kay and Brooks 2000, Vislocky et al. 1995). There is no reason that such approaches cannot be applied in an operational setting.

At organizations such as the Storm Prediction Center (SPC), forecasters are highly motivated to understand their successes and failures. They routinely evaluate their forecasts in a subjective, visual manner as a part of the forecast process. Persons responsible for issuing areal forecasts can gain great insight into the strengths and weaknesses of individual forecasts by simply visually comparing the forecasts to the verifying observations. By taking such an approach forecasters are able to learn about displacement errors, over- or underforecasting biases, etc. and more important, they relate them to the meteorological conditions. Despite the apparent simplicity of such an approach, great insight can be gained, which is difficult to accomplish through purely objective means (Ebert and McBride 2000).

Historically, verification has been done as a postmortem exercise on a limited number of forecast products by members of the research community with results published in technical memorandums and journal articles. There appears to be a relative dearth of verification systems that verify forecasts operationally and are meant to provide immediate feedback to the forecasters themselves (Mahoney et al. 1999). This situation, while not necessarily intentional, nonetheless limits the ability of the verification to improve the forecasts.

The greatest benefit of collecting ancillary information about forecaster thoughts at production time is that this information can be used to stratify the resulting forecasts and observations (Murphy 1995). Stratification refers to separating forecasts into various categories based on user-specific criteria. These categories which need not be specified a priori, include synoptic flow regime, wind direction, amount of precipitation forecast, or forecast difficulty ("easy" vs. "hard"). Murphy and Winkler (1987) developed a general framework for verification based upon the joint distribution of forecasts and observations that completely specifies all non-time-specific aspects of the forecasts and observations. The inclusion of stratification information into the general framework greatly improves its utility. This type of approach mimics the method undertaken by forecasters who mentally catalog events and patterns and use these patterns and memories when formulating future forecasts.

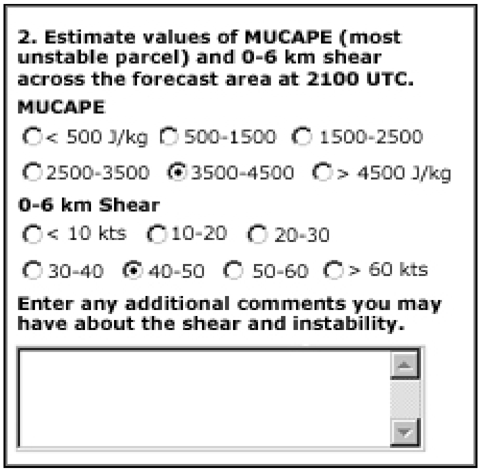
3. A NEW WAY OF THINKING

The current state of the art in meteorology is such that forecasting and verification are considered independent processes with no formal overlap. By keeping the processes separate, both parties suffer in that those verifying the forecasts are not in touch with the forecast process, and those generating the forecasts are left with little understanding of the verification. Clearly, a synergistic relationship between the two groups would prove beneficial and represents the central focus of an improved approach to verification. Is is important to realize that this approach does not advocate forecasters verifying their own forecasts which is clearly non-desirable, but instead advocates forecaster involvement in the overall verification process.

A more robust approach to forecast verification can be summarized in the following manner. First and foremost, forecasts need to be designed in such a way that they are scientifically robust, usable, and *verifiable*. Forecasts are often implemented without the verification procedures being well-posed, or even discussed. Information should formally be collected on the forecast process by those creating the forecasts at the time of forecast creation. The information collected should be of direct relevance to the forecast product and the forecast process and allow for stratification. Forecasters are already faced with numerous time constraints so any approach to collect data should be as concise as possible. Real-time collection of data should also be done using unambiguous wording and include as much error-checking as possible to limit mistakes and confusion. Questions should be formulated such that the evaluator need only specify one or more predefined options, or provide a confidence level. By choosing such an approach, statistical compilation and evaluation is possible in a systematic manner. General comments and free-form questions should be kept to a minimum.

An example of such an approach is shown in Fig. 2. This example illustrates the collection of information about instability and wind shear through the Most-Unstable parcel Convective Available Potential Energy (MUCAPE) and surface to 6 km vertical wind shear, respectively. This example shows an interface which allows the forecaster to choose between all possible, or relevant, choices in an efficient, intuitive manner. Further, forecasters are allowed to provide additional, free-form comments if they desire.

Additionally, the data should be stored and organized in such a way as to provide users with efficient means to interrogate and segregate results in a highly flexible manner. The use of relational database management systems for such tasks is ideal. One must also consider how to present the verification information to the forecasters in an easy-to-use manner. The use of the Web as the communication medium is ideally suited for such purposes. The popularity of the Web and surrounding technologies has exploded in recent years, and thus allows them to assume a central role in any future operational verification system that serves a variety of purposes and users.



2. Estimate values of MUCAPE (most unstable parcel) and 0-6 km shear across the forecast area at 2100 UTC.

MUCAPE

○< 500 J/kg ○500-1500 ○1500-2500

○2500-3500 ●3500-4500 ○> 4500 J/kg

0-6 km Shear

○< 10 kts ○10-20 ○20-30

○30-40 ●40-50 ○50-60 ○> 60 kts

Enter any additional comments you may have about the shear and instability.

Figure 2. Sample question from the 2001 NSSL/SPC Spring Program showing the collection of additional data about wind shear and instability to help characterize the forecast for later use in evaluation. Note how the forecaster has distinct selections to characterize all possible situations in addition to a text area to provide additional comments.

An example system using the concepts discussed above was tested during the 2001 Spring Experiment conducted by the National Severe Storms Laboratory (NSSL) and the SPC. The interested reader is urged to go to the Program's Web site (www.spc.noaa.gov/exper/Spring_2001/) to explore in greater detail the approach that we have advocated toward the collection of additional evaluation information to improve forecast verification. Similar design decisions have recently been used for the evaluation of turbulence diagnosis and forecast algorithms at the Aviation Weather Center and also as part of the Forecast Systems Laboratory's involvement in numerical model evaluation during the International H_2O Project (IHOP). Previous projects such as Coach (Nelson et al. 1999) have also been created with the stated goal of improving forecast performance, albeit without the focus on verification.

4. SUMMARY

An approach for improving forecast verification through the collection of subjective evaluation information has been presented. The goal of this methodology is to include information from those responsible for creating the forecasts that help to describe the forecasts using information not contained in the forecasts themselves. This information can then be used to stratify the forecasts and corresponding verification in a variety of ways, as suggested by Murphy (1995). Such an approach brings verification into the forecast process in a much more meaningful way than is currently done.

5. ACKNOWLEDGMENTS

Beth Sigren graciously helped format the two figures. The authors would also like to thank Jack Kain, Paul Janish, Steve Weiss and Greg Carbin for numerous insightful discussions related to the NSSL/SPC Spring Experiment and their enthusiasm for including research in operations and operations in research.

6. REFERENCES

Baldwin, M.E., S. Lakshmivarahan, and J.S. Kain, 2001: Verification of mesoscale features in NWP models. *Preprints, 9th Conf. On Mesoscale Processes*, Amer. Meteor. Soc., Ft. Lauderdale, FL, 255-258.

Brooks, H.E., M.P. Kay, and J.A. Hart, 1998: Objective limits on forecasting skill of rare events. *Preprints, 19th Conf. On Severe Local Storms*, Amer. Meteor. Soc., Minneapolis, MN, 552-555.

Brown, B.G., C. Davis, and R. Bullock, 2002: Verification methods for objects and fields: Standard methods, issues, and enhanced approaches. Report to the National Weather Service Radar Operations Center (available from the author at bgb@ucar.edu), 31 pp.

Doswell, C.A. III and J.A. Flueck, 1989: Forecasting and verifying in a field research project: DOPLIGHT '87. *Wea. Forecasting*, **4**, 97-109.

Ebert, E.E., and J.L. McBride, 2000: Verification of precipitation in weather systems: Determination of systematic errors. *J. Hydrol.*, **239**, 179-202.

Janish, P.R., S.J. Weiss, J.S. Kain, and M.E. Baldwin, 2001: Advancing operational forecasting through applied research programs at the Storm Prediction Center and National Severe Storms Laboratory. *Preprints, 18th Conf. on Weather Analysis and Forecasting*, Amer. Meteor. Soc., Ft. Lauderdale, FL, 245-249.

Jincai, D., C.A. Doswell III, D.W. Burgess, M.P. Foster, and M.L. Branick, 1992: Verification of mesoscale forecasts made during MAP '88 and MAP '89. *Wea. Forecasting*, **7**, 468-479.

Kay, M.P., and H.E. Brooks, 2000: Verification of probabilistic severe storm forecasts at the SPC. *Preprints, 20th Conf. Severe Local Storms*, Amer. Meteor. Soc., Orlando, FL, 285-288.

Mahoney, J.L., J.K. Henderson, B.G. Brown, C.S. Hartsough, 1999: Benefits of using FSL's Real-Time Verification System (RTVS) at the NWS/NCEP Aviation Weather Center. *Preprints, 8th Conf. on Aviation, Range, and Aerospace Meteorology*, Amer. Meteor. Soc., Dallas, TX, 95-99.

Murphy, A.H., 1995: A coherent method of stratification within a general framework for forecast verification. *Mon. Wea. Rev.*, **123**, 1582-1588.

Murphy, A.H., and R.L. Winkler, 1987: A general framework for forecast verification. *Mon. Wea. Rev.*, **115**, 1330-1338.

Nelson, S.E., L.P. Rothfusz, C. Jacobson, and J.T. Johnson, 1999: Coach - Design and implementation of an automated performance support system at the Tulsa WFO. *Preprints, 15th Conf. On IIPS for Met., Ocean., and Hyd.*, Amer. Meteor. Soc., Dallas, TX, 91-94.

Vislocky, R.L., J.M. Fritsch, and S.N. DiRienzo, 1995: Operational omission and misuse of numerical precipitation probabilities. *Bull. Amer. Meteor. Soc.*, **76**, 49-52.