Reprinted from Preprint Volume, Eleventh Conference on Weather Forecasting and Analysis, June 17-30, 1986. Published by the American Meteorological Society, Boston, Mass.

12.5

Fundamental Considerations in Forecasting for Field Experiments

C.A. Doswell III, R.A. Maddox and C.F. Chappell
NOAA/Environmental Research Laboratories
Weather Research Program
Boulder, CO 80303

I. Introduction

During the summer of 1984 and spring of 1985, the Weather Research Program (WRP) participated in data-gathering experiments (AIMCS and OK-Pre-STORM, respectively) concerned with mesoscale convective systems, or MCSs.

It was felt that specialized forecasts, keyed to MCSs, were essential for coordinating the data collection phase of both experiments. Moreover, WRP wanted to have forecasting itself be a part of the total research goals. During the course of these two programs, it became clear that forecasting for field programs had several aspects which have never been treated in the literature, to our knowledge. Hence, we felt it appropriate to discuss these issues in advance of the large commitment of resources embodied in such experiments as the National STORM Program.

II. Specification of Program Objectives

It is almost a reflexive response today within the meteorological community to preface any planning for field programs with statements of concern for clear expression of the program's scientific objectives. An experiment's goals have a major impact on the forecasting element of the program. We are particularly concerned with the relationship between the scientific objectives and the observing systems to be deployed for achieving those goals.

The concept of the *Testable Hypothesis* is one which has been given some attention (see e.g., Zipser, 1984). Those wishing an in-depth discussion of scientific hypothesis testing might consult Popper (1962). Within the context of this paper, we consider a *testable hypothesis to be one in which a specific prediction is made about the results of an experiment, and the data collected during that experiment is likely to yield an unambiguous statement about the validity of the hypothesis.* A rather trivial example would be the hypothesis, "Cumulonimbus clouds are characterized by up- and downdrafts having magnitudes comparable to those predicted by parcel buoyancy theory." It is easy to envision experiments which might allow this to be tested -- e.g., flying instrumented aircraft into cumulonimbus clouds, sampling such clouds with high-resolution Doppler radars, etc.

If the field program seeks resolution of more than one such hypothesis, conflicts might arise because collection of data to test one hypothesis could preclude obtaining the data for testing another hypothesis. During operations, the anticipation of weather events may have a large influence on the actual priority given to a particular experiment on a particular day. This means that the weather forecast can become crucial in determination of which objective(s) is (are) most important for that day.

As discussed in Popper (1962), experimental data can never provide an *absolute* confirmation of any hypothesis. Accumulated positive experimental results can lead to a growing certainty of the idea's validity (e.g., the Second Law of Thermodynamics), but only one verifiable counterexample is sufficient to call the hypothesis into question. This suggests that an experiment should seek to test hypotheses under a variety of circumstances. One aim of the data collection is to establish with some confidence that one has, indeed, given the hypothesis a rigorous test. If, for instance, it was hypothesized that an MCS requires a large-scale environment characterized by near-saturation through a deep, surface-based layer, collecting data for one case in which the hypothesis is confirmed does not test that hypothesis very rigorously. Certain hypotheses are more difficult to test than others because the data to test them must span an extensive space of natural variability.

The limited time and space domain of a field experiment, along with the available data collection resources, define the limits on the sorts of hypotheses that can be tested. Participants in a field program know quite well that the vagaries of the weather are a real factor in determining the success of the program. If all the hypotheses under test concern supercell thunderstorms, then supercell storms must develop within the spatial and temporal boundaries of the experiment. Since virtually all field programs have limits on the type of events under consideration, forecasting for the program can have a substantial impact on the data collection operations.

Another sort of field program is the "fishing expedition", in which testable hypotheses are not required. The experiment simply goes about collecting data without any testable hypotheses. There are two implicit assumptions behind fishing expeditions. First, whether by accident or design, the data collection is done as if little or nothing is known about the phenomena to be sampled. Presumably, if something were known about the phenomena to be sampled, it would be possible to construct testable hypotheses and the data collection would be tailored to ensure the rigorous testing desired.

Second, it is assumed that simply collecting data is likely to lead to the development of testable hypotheses. Even when virtually nothing is known *a priori*, a likely outcome of the experiment is that the data analysis will lead to ideas. Those ideas may become the testable hypotheses of future field programs, or the data collected may suffice (fortuitously) to test those notions.

This is not to say that meteorology should *not* indulge in fishing expeditions. There is certainly much we do not know, so it may be of great importance to search for good ideas. When the fishing expedition is geared to certain phenomena (e.g., MCSs), forecasting may still have an important role. If the goal is simply to document with instrumented aircraft the processes operating in, say, thunderstorm anvils, one must still be able to anticipate the storms in order to get the planes in the air on time. We shall explore this further in what is to follow.

Another important aspect of the program's objectives is whether or not *forecasting* is a subject of research during the experiment. Forecasting need not be a direct objective of a particular project, but when it is, there are constraints imposed on the manner in which data are collected. These constraints will be discussed below.

III. Data Sources for the Program

For field programs, the types of data collection systems to be deployed are of paramount importance in defining the forecast team's responsibilities. There are several sorts of observing "systems", all of which have characteristics for which the team must account.

A. Operational Data Sets

Here, by "operational", we mean those which are available every day, whether or not an experiment is underway. In the U.S., the primary source for operational data is the National Weather Service, which oversees the surface observations, upper-air soundings, and routine radar and satellite data collection. These data are relatively coarse, and generally do not meet the needs of a field program. If they were satisfactory, why have a field program at all? Nevertheless, such data usually form an important framework for the special data sets of the experiment.

As in AIMCS and OK-Pre-STORM, it is possible to enhance the experimental data base simply by increasing the temporal frequency of soundings at *operational* sites. However, this adds substantially to the workload of the staff, whose numbers may be barely sufficient for routine duties. If it is anticipated that the field program will want to take advantage of operational sites for special data collection (be it rawinsondes or whatever), this calls for careful planning and considerable advance notice. Note that the advance notice (as to what is needed and expected from the operational staff) should be given well before the program begins, and on each day of anticipated data collection (or the day *before* data collection). Day-by-day notice is the responsibility of the experimenters and depends on input from the special program's forecast team.

B. Experimental Data Sets

1. Fixed Observing Systems

Fixed systems often run more or less continuously through many field programs. A typical exception to this is the special rawinsonde data. Part of the explanation for this exception is that the rawinsonde data are relatively expensive, owing to "labor" costs -- the teams needed to launch the sondes and process the data -- and to the cost of expendables. Thus, it is common for the soundings to be "turned on" in anticipation of a desired event and "turned off" at the close of the event (or if it fails to materialize).

In our experience, the sporadic acquisition of special soundings puts a great deal of stress on all parties involved, often leading to confusion, ill will, and gaps in the data set. While the cost of soundings is fairly high (OK-Pre-STORM soundings averaged about $320 apiece -- J. Cunning, personal communication), their scientific value can be compromised severely by this episodic acquisition scheme. The specific scientific value of concern is that associated with forecasting research, to be detailed below.

2. Mobile Observing Systems

Mobile systems (e.g. aircraft and mobile surface teams) require forecasting support by their very nature. Not only do they need to anticipate events in order to be in proper position during the phenomena, but they may require guidance during the event in order to maintain the desired observing positions. Such guidance is often referred to as "nowcasting", and differs from forecasting in that it deals with very short range (zero to a few hours) issues. This creates the need for close interaction between the forecast team and the mobile team(s), often through the intermediary of a "coordinator" of mobile operations.

For mobile teams, an important factor is their *response time* -- i.e., how long they need to prepare for operation, how long they take to get into position, how quickly they change position when necessary, and how long it takes to complete a field observing operation (e.g., a flight leg). This response time can have a major role in formulating forecast strategies and responsibilities.

3. Combined Fixed and Mobile Systems

Most field experiments incorporate *both* fixed and mobile systems (as in AIMCS and OK-Pre-STORM), so the coordination of the total observing system creates special problems for the forecast teams. The needs of each component of the whole system must be considered throughout the data collection phases of the program, and events can arise where serving one element's requirements causes the operation of another to suffer. It may require great effort to deal with the immediate problems faced by mobile teams, while the longer-range forecast issues associated with a fixed system go untreated. In such a situation, the time for a decision regarding the fixed system may arrive suddenly, with little or no thought having been given to it.

C. Verification Data

In one sense, the total data set collected becomes the source for verification of scientific hypotheses. *Forecast* verification may be considered by some to be a separate issue from the hypothesis testing associated with the experiment. This misconception owes its existence to the perceived dichotomy between forecasting and research.

The very same constraints about the data collection for hypothesis testing apply to forecasting: the data must be capable of an unambiguous evaluation of the forecast. If one has made a forecast for the occurrence (or non-occurrence) of a weather event, one should collect data in such a way that the forecast can be evaluated properly. For example, tornado forecasts can only be verified with tornado occurrence data. If there is a large chance of unobserved tornado occurrences and/or events erroneously identified as tornadoes, the verification may be *technically* rigorous, but has little value.

When forecasting research is to be included among the scientific objectives, the portion of the forecasts subject to rigorous verification must be tuned to the data set available within the program. To do otherwise defeats the purpose of verification: to evaluate the quality of the model(s) upon which the predictions are based, and the techniques by which those models are applied.

IV. Employment of Experimental Data Sets

The deployment of non-operational data systems in the field offers the potential to explore using the new data for forecasting. To use *operational* data, the forecast team has a base requirement for access to the operational data streams (facsimile and teletype, or AFOS-compatible hardware). If it is desirable to gain real-time access to the *experimental* data, this requires additional communication, processing, and display facilities (which must be maintained) at the forecast team operations center. It must be recognized that all this input may cause information overload. It is characteristic of many of the new observing technologies that they have a high "baud rate". This flood of data can overwhelm the forecast team as they provide support to field operations. This is an issue with implications about future operational forecasting, as well as within the limited context of field experiments.

In addition to the basic forecasting duties of the team, the new data sets could be used as nowcasting tools in support of actual field operations. However, the new data may not be an unalloyed blessing in this regard. There may not be enough understanding of how to use the new data for forecasting and/or nowcasting. Also, some or all of the experimental data may not constitute a "critical mass" of information. A few special soundings scattered over a mesoscale area do not provide a reliable information base for understanding what is happening on the mesoscale. In fact, it is possible to misinterpret a situation badly because one assumes that the special data are representative of a mesoscale region when, in fact, they are characteristic only of a very localized environment.

Having *real-time* access to the new data may not involve specific hypothesis testing until the new technologies for data collection have become more mature. In our experience, the enhanced data sets can give tantalizing hints of atmospheric processes to be studied in depth, perhaps in future field experiments (e.g., see Maddox, 1985)

V. Predictability and Scale

The science of meteorology is on its firmest ground when considering the large-scale (of order 1000 km) aspects of the atmosphere, in mid-latitudes. On this scale, the extratropical cyclone dominates events and our understanding of this dominant process is relatively advanced. For other scales (and in polar and tropical regions) our understanding is nowhere near as complete. This implies that our forecasting ability for other phenomena is correspondingly less advanced. The long-range impact of field programs in these areas can be improved weather forecasts. Indeed, this facet of scientific research has been a major thrust in efforts to underwrite the National STORM Program. However, the fact that our scientific base for problems other than large-scale meteorology is rather shaky indicates that forecasting for field experiments can be difficult.

We have already suggested that the response time of the observing systems is a factor which the forecast team must consider. Should that response time be comparable to the time scale of the phenomena under consideration in the experiment, deployment of that system to sample that event can become quite difficult. *If external programmatic constraints force the employment of slow-response observing systems to sample short time scale events, one must expect a lot of failure in attempting to collect a satisfactory data set.*

Also, it is commonly assumed that as weather events unfold, it is possible to define the event with increasing accuracy. This assumption underlies the operational concept for forecasting hazardous weather phenomena (the "outlook-watch-warning system).

To assume that each successive step is easier to do is scientifically unsound, because while we may be obtaining information about the event as time proceeds, this information increase is at a much slower rate than the *decrease* in scientific understanding as space and time scales shrink. Our data generally are obtained at a fixed sampling rate (defined by the observing system), so the sampling density *relative to the decreasing scale of the event* falls rapidly. Therefore, the real ability to pinpoint events as they unfold is considerably less that what is implied. This is just as true for field program forecast teams as it is for operational forecasters.

VI. The Psychology of Decision-Making

To this point, we have stressed the technical difficulties encountered by field program forecasters. Apart from situations where forecasting research is a stated scientific objective of the program, we believe that forecasting can be of great value to decision-making.

It is critical to understand the distinction between forecasting and decision-making. It is the responsibility of the forecaster(s) to pass on a meteorological assessment of the weather situation to the program decision maker(s). This should include a description of what is considered the most likely evolution of the situation and a forthright statement of the uncertainties. It may also include alternative scenarios, with some indication of their likelihood.

Then, it is the responsibility of the decision-maker(s) to incorporate this assessment with all other information relevant to field operations (status of observing systems, personnel, budget, etc.). The forecast input is but part of what is needed to render a decision about the day's field operations. Forecasts can improve the chances for making correct decisions, but it must be understood that there is a non-vanishing chance the weather will not match the forecast perfectly. These apparently obvious statements have, in our experience, proven to be a source of confusion and conflict in the heat of a real operation.

Perhaps the most difficult challenge in decision-making is the fear of making the wrong choice. There certainly are large responsibilities thrust on the decision maker(s), particularly the burden of expending the program's precious resources as fruitfully as possible. One apparent way to avoid being wrong is to not make a decision. However, this is an illusion -- no decision is really a decision! In fact, it seems probable that a "no decision" decision has about the same chance of being wrong as one which is formally made.

Since a "no decision" decision is not really a means of escaping the decision process, the next easiest path of apparent safety is to *postpone* decisions. This may take a middle ground between "go" and "no-go" decisions, referred to as "stand-by" status. A standby condition is one in which the operational program elements are in a state of readiness, but remain short of commencing operations. This can reduce the response time if a decision to go ahead is made. Unfortunately, this is an illusion if the commitment to operate is not likely to be made, and made rather soon.

Part of the problem with standby status is that some components cannot be put in that mode, such as rapid-scan satellite images. They are either on or off, with nothing in between. If operations commence after being on standby, without having made a *prior* commitment to operating these special components of the total system, then the operation must go ahead without them. If they are turned on by prior commitments, and the standby period goes by without commencing operations, the data collected by the special system are the *only* data.

Further, standby status may cost almost as much as the operation itself. Support staff on standby are normally conidered to be on duty and must be paid acordingly. One cannot be on standby for six hours and then work a ten hour operational day too many times without reducing one's capacity to perform at peak levels. Staying on standby for extended periods is tiring and frustrating, increasing the pressure on the decision-maker(s) to choose between go and no-go, a situation conducive to making bad choices.

Another apparent refuge which appears to avoid shouldering the burden for making decisions is to decide by consensus. It is well-documented that forecasting by consensus has the best verification. Also, no one person has to bear the responsibility for bad (or good) decisions. As with the other seemingly safe havens, there are real reasons why this is not necessarily the best approach to field program leadership. If one looks at the "forecasts by consensus" argument, it must be realized that forecasting and decision-making are distinct processes. One may be willing to use consensus forecasts, but decision-making by committee is notoriously ineffective.

Beyond this important distinction, consider how and why consensus produces the best verification. Since consensus always lies somewhere between extreme views, it has the best chance of success *in the long run*. However, it is not hard to find examples where forecasts near the extremes of opinion beat consensus. By basing decisions on consensus, one avoids the long range accumulation of errors, but simultaneously loses the opportunity for the spectacular success. In other words, consensus minimizes failure by minimizing success.

We want to emphasize that this does not imply that decisions (and forecasts) should be made without all possible input by leaders (and forecasters) in isolation. In fact, program decision-makers need to seek diverse opinions, since no individual is invariably correct in the assessment of a situation. But field programs are akin to military operations -- their success hinges on decisive leadership and the willingness of the participants to work toward a fruitful operation even when they disagree with the decision.

Most of the problem in decision-making springs ultimately from limitations on resources. It is not hard to empathize with someone who hesitates to commit the limited resources in the face of forecasting uncertainties. Even forecasters expressing a great deal of confidence in their forecasts have been wrong, so it is easy to understand why decision-makers may be reluctant to "go" on a forecast.

Nevertheless, we feel that decisions must be based on forecasts in most, if not all, field programs. Only when one has observing systems with negligible response time, or when one is not interested in the pre-event environment (in which case forecasting is most certainly not an objective of the experiment) can the go/no-go decision be postponed until the desired events are underway. If there are no limits to program resources, the non-occurrence of an event is without penalty and the experiment can run continuously.

In real programs, the point of decision-making is to avoid wasting the limited opportunities and resources. It makes little sense to design a program where all decisions would have to be right to achieve the objectives, but there is always pressure to maximize results.

VII. Some Suggestions

Based on our experience, we have some suggestions to offer for the design of future field programs which have a foreasting element. Since forecasting and program objectives are so intertwined, much of what follows involves the scientific objectives of a field experiment.

When one has testable hypotheses as a part of the total set of objectives, these should have priority in designing the program. It is far easier to accept compromises for objectives stated in terms of "to understand such-and-such" or "to document so-and-so" than when seeking specific information to evaluate clearly defined objectives. Let serendipity have its way, but one should try to come out of the program with something concrete, rather than simply collecting data in the hope that something good will turn up.

A curious phenomenon which arises in virtually all field programs is that the role of forecasting seems to diminish as the program proceeds through its alloted time frame. Early in the experiment, the decision-makers tend to be quite choosy about the situations to which they will commit resources. At some point, this economy may lead to a fear that the resources will not get used up by end of the program. Thus, as time passes, it becomes increasingly likely that a go decision will be made in a given situation.

It seems more sensible to be as consistent as possible with the treatment of forecasting input throughout the time alloted. This is difficult to do, and requires some fairly specific guidelines to be defined *in advance* about the relationship between the forecast team's input and the likelihood of a go decision.

If one must accept that forecasts can go awry, it is just as inevitable that some decisions may fail as well. Moreover, when the program focuses on a specific phenomenon, it is possible that a really good example of the event will not occur within the space and time limits of the program. To maximize the chances for something good to come from the experiment, it seems wise to provide contingency plans. This means that the scientific objectives should be designed so that if a go decision is made based on the forecast, and events do not evolve as anticipated, there is still something of scientific value which can be done. For example, if the forecast is for an MCS to develop late in the afternoon and nothing of that sort happens, the data collected can be used to serve another purpose, like documenting the evolution of the low-level jet stream.

If the scientific objectives cannot be met without a certain type of evolution within the boundaries of the experiment, that evolution had better have a very high climatological probability inside those time and space boundaries! We think a more reasonable approach is to broaden the objectives and have a series of contingency plans for most foreseeable circumstances. Doing this reduces the pressure on the forecasters and decision makers.

Whenever forecasting is a research element during the experiment, such research depends on being able to discriminate between environments which are favorable for the events under consideration and those which are unfavorable. This means that the "basic" data systems (special soundings, mesonetworks, etc.) should operate more or less continuously through the program. Forecasting the non-occurrence of an event can be just as difficult as forecasting its occurrence, so basic data collected on non-occurrence days is essential to forecasting research. It also guarantees that otherwise unanticipated occurrences will have at least the basic data collection available for study.

There is an obvious drawback to continuous operation: it costs resources. In our view, *when forecasting improvement is among the research objectives, it is better for that research to run a relatively short program with* ***continous*** *basic data collection than to operate the basic systems sporadically for a relatively long program.* It is the sophisticated mobile systems (e.g., instrumented aircraft) which are most sensibly operated on an *ad hoc* basis, but these should not drive the entire data collection operation.

In the same vein, the total observing system's response time on any given operational day is set by that component of the system with the slowest response time. The most ponderous form of data collection drives the forecasting and forecast-related decision process. If the objectives dependent on slow-response sytems do not require the *total* system to be operating in concert with them, it would be possible to be more flexible in deployment of those sluggish data collection tools. For example, the objectives served by rapid-scan satellite data may not require that mobile field teams be operating, but do require the basic data sets (which we assume to be available at all times). In such a case, if the situation develops that the field teams are not deployed for one reason or another, meaningful objectives are not compromised.

Such a decoupling of observing systems eases the stress on the forecasters and decision makers in one sense, but introduces other complexities. In effect, the total data collection operation is broken down into smaller units which are more or less independent. Daily operational decisions must account for a variable mix of observations in any given situation. While this seems potentially confusing, our experience with field programs suggests that something of this sort is almost inevitable. That is, various observing systems serving the objectives of different scientists arrive, leave, go down for maintenance, etc., more or less independently during extended field operations. What we propose is to have founation of continuous basic data to serve the broadest range of scientific objecives, with specialized systems designed to serve rather narrow objectives operating *within* that foundation.

VIII. Concluding Remarks

While this discussion is intended nominally to address forecasting for field programs, our experience leads us to believe that that forecasting process is influenced strongly by the program objectives and data systems. The success of field programs may be dependent on that forecasting, but the value of the forecast input is determined by the care expended during the planning stages for the field data collection phase of the experiment. Flexibility and contingency planning make it possible to achieve program objectives without asking the forecast team to do the impossible, and can dramatically reduce the impact of incorrect decisions based on forecast input.

Field programs provide a research scientist with an opportunity to experience something which approximates operational forecasting, and perhaps the chance to interact with operational forecasters. Operational forecasters can use this chance to see for themselves how the science of meteorology relates to their job, and perhaps can become acquainted with new technologies. If one agrees that the research-operations dichotomy is detrimental to the science as a whole, this opportunity can help alleviate the lack of communication which maintains the research-operations gap.

Acknowledgements. We would like to thank all the participants for their efforts in our two field programs, and especially those in the National Weather Service, for their contributions to the success of our experiments.

References

Zipser, E.J., 1984: *STORM-Central Phase: Preliminary Porgram Design.* National Center for Atmospheric Research, Boulder, Colorado, 147 pp.

Popper, K.R., 1962: **Conjectures and Refutations: The Growth of Scientific Knowledge.** Harper Torchbooks, New York, 417 pp.

Maddox, R.A., 1985: The relation of diurnal, low-level wind variations to summertime severe thunderstorms. *Preprints, 14th Conf. Severe Local Storms* (Indianapolis, IN), Amer. Meteor. Soc., 202-207.