Reply

R. ATLAS, M. GHIL[1] AND M. HALEM

Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD 20771

(20 May 1980)

The comments presented by Druyan (1980) deal primarily with the automated forecasting method (AFM) and not with time-continuous assimilation or the impact of satellite temperature sounding data on weather prediction. Therefore, we will limit our reply to a discussion of the development of the AFM and its role in forecast impact evaluation.

Numerical measures of forecast accuracy such as the rms error or S1 score have been the main tool used for the purpose of assessing the impact of remote sounding data on weather forecasting. However, impacts in rms error or S1 score are often difficult to interpret in terms of impact on local weather forecasting (Druyan *et al.*, 1975; Houghton and Irvine, 1976; Sanders, 1973). This is due to the large-scale averaging inherent in this type of verification procedure as well as to the methodology by which weather forecasts are actually produced. Local weather forecasting is not based solely on the one or two large-scale fields for which numerical scores are routinely computed; a variety of prognostic primary variables and derived quantities are

[1] Courant Institute for Mathematical Sciences, New York University, New York, NY 10012, and Goddard Laboratory for Atmospheric Sciences.

TABLE 1. Verification of model and AFM local precipitation forecasts. The results refer to the summation of comparisons of 12 h intervals at 128 U.S. cities, for each 72 h forecast from 3–21 February 1976.

February	Both correct	Both incorrect	Model better	AFM better
3	372	112	96	188
5	376	163	79	150
7	480	100	83	105
9	475	108	52	133
11	533	66	57	112
13	437	96	57	178
15	416	123	87	142
17	358	150	80	180
19	473	80	56	159
21	485	70	85	128
Totals	4405	1068	732	1475

considered in combination (George, 1960; Naval Education and Training Command, 1974). Quasi-geostrophic theory forms the underlying framework for much of subjective weather forecasting through the omega and geopotential tendency equations (Holton, 1979, Chap. 6).

Recognizing the need for evaluating the impact of satellite data on local weather forecasting, we initiated research to develop improved measures for such an evaluation, in preparation for assimilation experiments with data from the January–March 1976 Data Systems Test (DST-6). We conducted a subjective forecast experiment and, utilizing the results of this experiment, developed the AFM. In the subjective forecast experiment, a series of numerical forecasts were generated from initial conditions which included satellite temperature soundings (referred to as SAT) and initial conditions which excluded satellite soundings (referred to as NOSAT). For each forecast case, a set of SAT and NOSAT prognostic packages, consisting of most of the charts available at local weather stations, was delivered to two experienced forecasters. These forecasters would then compare prognostic charts, note changes in meteorological variables, and then issue a weather forecast at a few preselected cities scattered throughout the United States. Each forecast was based completely on the subjective interpretation of the SAT or NOSAT prognostic charts. This experiment revealed that the impact of satellite sounding data was large enough to influence local weather forecasts (Atlas and Sakal, 1978). The experiment's approach, however, was limited in utility because of the large amount of time required to produce weather forecasts at even a small number of cities and because of the subjectivity involved. An attempt therefore was made to automate the forecaster's interpretation of the complete prognostic package.

Objective algorithms were developed which gave results correlating well with the subjective judgment of the magnitude and importance of each of the prognostic quantities that were utilized. To formulate the AFM, the objective quantities were combined sequentially in approximately the same order as had been employed by the forecasters. Minor modifications to the AFM were then incorporated to minimize bias and to account for local topographical differences. The AFM is thus based primarily on subjective forecasting experience which is to a large extent in agreement with quasi-geostrophic theory.

As mentioned in our paper, the AFM was verified in terms of its accuracy when applied to a perfect prognosis (i.e., an analysis) and in terms of its agreement with experienced forecasters. Only after the AFM proved satisfactory with respect to both these criteria was it applied to model prognoses. The AFM was found to be 93% accurate in determining whether or not precipitation would occur; it agreed with the subjective judgments of each of three forecasters (two of which did not participate in the original experiment on which the AFM was based) 94% of the time (on the average). The latter agreement is comparable to that between experienced forecasters. We concluded that the AFM was comparable in performance to the average skilled weather forecaster. The AFM was then utilized to evaluate the impact of satellite sounding data assimilation on local precipitation forecasts at 128 cities in the United States as reported by Ghil *et al.* (1979).

Druyan has questioned how well the AFM would compete with unmodified model precipitation forecasts. We did not present such comparisons in our original paper for reasons of brevity and also because we felt that the AFM's utility as a verification tool was not dependent on its skill relative to the model's own physical parameterization of precipitation. Its main utility results from comparing the AFM applied to model prognoses, with the AFM applied to analyses. This provides a very useful measure of the accuracy, deficiencies, and practical utility of the model prognoses. No attempt is made to correct model bias so that as the model's forecast accuracy improves, whether due to changes in physical parameterizations or initial data specification, the AFM forecasts will also improve and approach that obtained with a perfect prognosis.

Recently, GLAS model forecasts of measurable precipitation and AFM forecasts of precipitation occurrence have been compared as suggested by Druyan. A summary of the results of this evaluation appears in Table 1. Ten different 72 h forecasts from February 1976 were verified at 12 h intervals at 128 cities, giving a total of 7680 precipitation forecast comparisons. Of these, the AFM and model forecasts differed 29% of the time. Where they were different, the AFM was more accurate in its determination of precipitation occurrence 67% of the time. The greater accuracy of the AFM is related to

TABLE 2. Verification of AFM and single parameter predictions of precipitation occurrence for February 1976.

Forecast scheme	Percent correct	Bias	Pre-figurance	Post-agreement	Threat score
Precipitation occurs when 500 mb vorticity advection:					
$>0 \times 10^{-9}$ s^{-2}	55	3.03	0.70	0.23	0.21
$>0.5 \times 10^{-9}$ s^{-2}	79	0.92	0.35	0.37	0.22
$>1 \times 10^{-9}$ s^{-2}	83	0.30	0.17	0.55	0.15
Precipitation occurs when 850 mb temperature advection:					
$>0 \times 10^{-9}$ s^{-2}	49	2.91	0.49	0.17	0.14
$>0.5 \times 10^{-9}$ s^{-2}	74	0.99	0.24	0.24	0.14
$>1 \times 10^{-9}$ s^{-2}	79	0.39	0.10	0.25	0.08
Precipitation occurs when 850 mb dew-point depression:					
<5°C	83	1.12	0.57	0.51	0.37
<2.5°C	83	0.16	0.16	0.63	0.10
AFM	93	0.99	0.76	0.76	0.60

the deficiencies of the model's vertical motion predictions, as described by Ghil *et al.* (1979), and to the inclusion in the AFM of procedures to account for frontal precipitation.

Druyan also has questioned the utility of vorticity advection as a predictor for precipitation; he cites his limited study of the relationship of vorticity advection and moisture to rainfall occurrence, and statistical testing by NMC's Techniques Development Laboratory (TDL). We do not believe that Druyan's study for Israel is particularly relevant to the choice of parameters in the AFM, which has only been applied to cities in the United States: the difference between Israel's subtropical Mediterranean climate and the temperate climate predominant in the United States involves both the AFM's predictors, the large-scale prognostic field quantities, and its predictands, the local weather events. In addition, the statistical testing of vorticity advection by TDL may not be representative of its actual prognostic utility. Vorticity advection may not have been chosen as a predictor by the screening process because its effects are indirectly incorporated within predictors which were selected, or because it was not effectively combined with other parameters. Previous empirical studies (Dunn, 1962; Jenrette, 1960; Williams, 1963) have shown vorticity advection to be a useful predictor when considered in conjunction with low-level moisture, thermal advection and vertical motion associated with fronts. The AFM does not rely solely on vorticity advection: it simulates the subjective combination of several parameters from model prognoses or from analyses to arrive at its prediction of precipitation occurrence.

To determine the effectiveness of the AFM's combination of parameters, we have compared its accuracy when applied to perfect prognoses with simple forecast schemes in which the prediction is based on a single parameter. Table 2 presents the results of this evaluation for the AFM versus selected cutoff values of 500 mb vorticity advection, 850 mb temperature advection, or 850 mb dew-point depression. Comparisons with observed precipitation events were made every 12 h for the entire month of February 1976 for the same cities that were utilized in our original verification of the AFM. Statistics verified included percent correct, bias, prefigurance, post agreement and threat score; for a definition of these statistics see, for instance, Ghil *et al.* (1979) and Tracton and Stackpole (in Murphy and Williamson, 1976, p. 736).

Table 2 shows the relationship of each of the single predictors to precipitation occurrence, and the relative utility of each prediction scheme. From the prefigurances, we see that 70% of all precipitation events occurred in conjunction with positive vorticity advection, as compared to 49% occurring with warm advection and 57% with dewpoint depressions < 5°C. However, the forecasts based on weak vorticity or thermal advection have higher bias and correspondingly lower values of percentual correctness, post-agreement, and threat score than those based on dew-point depressions < 5°C. The AFM, with near perfect bias, accounts for 76% of all precipitation events and verifies 76% of all positive precipitation forecasts. This results in a 10% increase in the total number of correct forecasts, and an increase in threat score of 0.23, relative to the best single parameter prediction. We conclude that 1) vorticity advection is a useful predictor of precipitation occurrence, 2) dew-point depression is the most accurate single predictor and 3) the combination of parameters in the AFM yields a substantial increase in accuracy over any single predictor.

In summary, the AFM was developed in response to the need for improved automated measures for assessing forecasting impact. It simulates subjective interpretation of an ensemble of prognostic and diagnostic quantities. The AFM has been verified from the standpoint of its accuracy when applied to analyses and of its agreement with experienced forecasters. Its actual prognostic accuracy was tested for a relatively large number of cases; these cases, however, were drawn from a limited number of model forecasts. The AFM appears to be comparable in performance to the average skilled weather forecaster's interpretations of large-scale numerical guidance and is thus a useful tool for evaluating the accuracy and utility of large-scale numerical forecasts. Further improvements and adaptations are under development and testing.

REFERENCES

Atlas, R., and D. Sakal, 1978: Evaluation and verification tests. *GISS Sounding Temperature Impact Test*, NASA Tech. Memo. 78063, Goddard Space Flight Center, 5.1–5.83.

Druyan, L. M., 1981: Comments on "Time-continuous assimilation of remote-sounding data and its effect on weather forecasting." *Mon. Wea. Rev.*, **109,** 200–201.

——, R. C. J. Somerville and W. J. Quirk, 1975: Extended-range forecasts with the GISS model of the global atmosphere. *Mon. Wea. Rev.*, **103,** 779–795.

Dunn, C. R., 1962: Forecasting precipitation with the aid of vorticity advection at the 500 mb level. Tech. Note No. 9, Office of Forecast Development, U.S. Weather Bureau, Washington, DC, 9 pp.

George, J. J., 1960: *Weather Forecasting for Aeronautics*. Academic Press, 673 pp.

Ghil, M., M. Halem and R. Atlas, 1979: Time-continuous assimilation of remote-sounding data and its effect on weather forecasting. *Mon. Wea. Rev.*, **107,** 140–171.

Holton, J. R., 1979: *An Introduction to Dynamic Meteorology*, 2nd ed. Academic Press, 391 pp.

Houghton, D. D., and W. S. Irvine, 1976: A case study comparison of the performance of operational prediction models used in the United States. *Mon. Wea. Rev.*, **104,** 817–827.

Jenrette, J. P., 1960: An objective application of vorticity principles to precipitation forecasting. *Bull. Amer. Meteor. Soc.*, **41,** 317–323.

Murphy, A., and D. Williamson, Coordinators, 1976: Weather forecasting and weather forecasts: Models, systems, and users. Notes from a Colloquium held at the National Center for Atmospheric Research, Summer 1976. NCAR/CQ-5 + 1976-ASP, Vols. 1 and 2, 900 pp. [NTIS PB-268 913/914].

Naval Education and Training Command, 1974: *Aerographer's Mate 1 & C.* NAVEDTRA 10362-B, U.S. Govt. Printing Office, Washington, DC, 659 pp.

Sanders, F., 1973: Skill in forecasting daily temperature and precipitation: some experimental results. *Bull. Amer. Meteor. Soc.*, **54,** 1171–1179.

Williams, P., 1963: Relationship of precipitation to vorticity and vertical motion at Salt Lake City, Utah. *Mon. Wea. Rev.*, **91,** 215–219.