DYNAMICAL AND EMPIRICAL METHODS OF WEATHER FORECASTING

11/79

Edward N. Lorenz
Massachusetts Institute of Technology

1. Introduction

It was somewhat over a century ago that Admiral Fitzroy began publishing daily weather forecasts in the newspapers in London. This action brought him considerable criticism from his scientific colleagues. Their attitude appeared to be that he was publishing scientific results for which there was no scientific basis. Following Fitzroy's death a few years later the publication of weather forecasts was discontinued.

Today we not only take it for granted that the weather forecasts will appear in the daily newspapers, but we recognize that there is a firm scientific basis for forecasting. Nevertheless, the forecasts have yet to attain the quality which one might expect of something whose scientific basis is understood. Forecasts more than two days ahead are especially uncertain, but major weather occurrences, such as heavy snowstorms, are sometimes unanticipated twenty-four hours in advance. If we are to choose judiciously among various possible forecasting procedures, we should appreciate the reasons for today's failures.

The atmosphere is a fluid system. It is an inhomogeneous fluid, since one of its principal constituents, water, occurs in each of three phases in varying concentrations. It is nevertheless subject to the laws of fluid dynamics. These, together with the influence of the sun and the underlying ocean and land surfaces, determine how the atmosphere will evolve from one state to another as time progresses.

Before we conclude that weather forecasting must consist in effect of solving the dynamic equations as an initial value problem, we should consider another phenomenon which we are accustomed to predict, namely, the oceanic tides. Like the atmosphere, the oceans are a fluid system, and the evolution of the ocean from one state to another is likewise governed by the laws of fluid dynamics. Yet we may predict the height of the ocean surface at a specified time a few months or years in advance without even examining the present state. We simply determine the positions of the moon and sun relative to the earth at the time in question.

To a certain extent we can predict the atmosphere by similar procedures. If we know that the time for which we are predicting is a winter night, we can forecast cold weather; if it is a summer day, we can forecast warm weather. Yet such predictions are unlikely to satisfy our clients. The distinction between tidal forecasting and weather forecasting is that in the case of the tides the forced variations - those variations resulting directly from variable external influences - constitute nearly all of the signal which we wish to forecast. In the case of the weather they omit much of the signal, including such things as the expected and unexpected snowstorms.

The vortices with diameters of hundreds or thousands of kilometers which travel across the oceans and continents, and often bring sudden temperature changes and heavy rain or snow as they pass, are examples of free variations. They would exhibit nearly the same behavior if the external forcing did not vary at all. These free variations, which must be forecast if our forecasts are to be acceptable, cannot be predicted simply by consulting the calendar and the clock. To predict their state tomorrow, we must know their state today. Similar free variations occur in the oceans; among them are large-scale meanders of the Gulf Stream. They are no more predictable without a knowledge of the present state than are atmospheric free variations, but they have very little influence on the tides.

Why do we have free variations? The atmospheric equations admit a solution in which only forced variations are present. Each day's weather nearly repeats the previous day's, being modulated slowly by the advance of the seasons. Each year nearly repeats the previous year, and would be an exact repetition if the number of days in a year were an integer. This solution is not the one which characterizes the real atmosphere because it is unstable with respect to perturbations of small amplitude. Among the systems which will appear when the inevitable perturbations have amplified are the large-scale migratory vortices.

A further property of free variations in the atmosphere, but not in all fluid systems, is that even when such variations are present, they are unstable with respect to still further small-amplitude perturbations. It follows that any error in estimating the present state of the atmosphere will amplify during the period of the forecast, regardless of whether the forecasting procedure is based on the equations which are responsible for the instability. There is therefore a limit to the range at which specific weather features can be acceptably predicted. Present estimates place this limit at more than a week, which is well beyond the range at which acceptable forecasts are made by current procedures. Nevertheless, in attempting to formulate new procedures we should bear this limit in mind.

2. Pure methods

Procedures for forecasting the weather may be classified as dynamical or empirical. Most methods in use involve a certain amount of both dynamics and empiricism. Our ensuing remarks will apply mainly to prediction one or two days ahead.

The most purely dynamical method would consist of solving the equations governing the atmosphere, using the observed atmospheric state for initial conditions. Since the equations are highly nonlinear, numerical procedures appear necessary. Hence some approximations would be immediately demanded. If no empirical information is used, the approximations might consist of representing each atmospheric field - temperature, wind components, etc. - by its values at the largest three-dimensional grid of points which the computer can accomodate. It turns out that such an approximation would not adequately represent the atmosphere for extrapolation purposes, and the method would f il completely.

At the other extreme, one purely empirical method would be the analogue method, consisting of searching past records until an atmospheric state just like the present state is found, and predicting the present state to evolve just as its analogue did. Since recorded past history is not infinite, identical states cannot be found, and we would have to settle for an approximation to the present state. The approximations which occur in the records are not close enough to make the method successful.

A modification which shows some forecasting skill consists of choosing as an analogue a state resembling the present state over a limited region, perhaps of subcontinental size. The probability of finding a good analogue is thereby greatly increased, although it is less certain that states which are similar over only limited regions will evolve similarly.

In effect the limited-area analogue method is what the subjective weather forecaster generally uses. He analyzes the current set of weather observations into systems, such as high and low pressure centers and fronts. Usually he will have seen other arrangements of systems which look like the ones he is presently dealing with, and he will assume that the present ones will behave like their analogues. When he encounters an rrangement which does not resemble one which he has seen before, his forecast is likely to go bad.

Another method which can be purely empirical is the use of linear regression. For each quantity to be predicted, a set of predictors is chosen, and the linear combination of the predictors which most closely resembles the predictand, within a selected sample of data, is chosen as the prediction. A limitation is that the number of observations of each predictor must be much larger than the number of predictors, since otherwise chance relations is the selected sample, which need not repeat themselves at the time of the forecast, are likely to occur. When the selected sample cannot be made large, techniques for reducing the number of predictors while retaining most of the useful information in them are available. These include principal component analysis and a stepwise regression scheme.

The predictors may be the values of weather elements at the present time; they may instead be values at the present and one or more past times, say 12 and 24 hours ago. Use of present predictors exhibits positive skill, while addition of the past predictors yields considerable improvement. Yet something is lacking. Comparisons have shown that use of present predictors alone yields a forecast which closely resembles the one ich would be produced by displacing every weather system through its climatological normal displacement. Use of present and past predictors is equivalent to using an average of the normal and immediate past displacements. Subjective forecasters have been displacing their systems in this manner for years, but they also take other factors into account.

3. Mixed methods

Procedures which are ordinarily categorized as dynamical usually involve some empiricism as well. For example, among the motions which are governed by the exact dynamic equations are sound waves of all sizes, including some with wave lengths comparable to the typical diameters of weather systems. These waves appear to have very little effect upon the weather. It would be highly wasteful, and perhaps impossible, to retain every large-scale sound wave in a numerical solution of the equations. Replacement of the vertical equation of motion by the empirically verified hydrostatic equation, which expresses a balance between the vertical pressure forces and gravity, reduces the system to a far more readily handled one which does not describe the bothersome vertically traveling sound waves, but leaves the description of the important weather systems virtually unaltered. This new system, the so-called primitive equations, serves as the basis for most of today's operational numerical forecasts.

The primitive equations admit another type of motion, namely gravity-wave oscillations, whose influence on the weather has been questioned. These oscillations may be eliminated by replacing the divergence equation, derived from the horizontal equations of motion, by the empirically established geostrophic equation, which expresses a balance between the divergences of the horizontal pressure force and the Coriolis force. The resulting simple system, the so-called geostrophic model, served as the basis for the first operational numerical weather forecasts, in the 1950's. The geostrophic equations distort the important weather systems to some extent, and, now that greater computer power is available, the primitive equations, combined with an initialization procedure for removing gravity waves from the initial conditions, have been found to give better results.

Empirical findings are also useful in deciding how to replace the continuous dynamic equations, whether they be exact, primitive, or geostrophic, by systems which a computer can handle. If the observations told us that the spatial spectra of temperature, wind, and other atmospheric variables were confined to large scales, we could without loss replace each variable by a terminating trigonometric series, and use the coefficients in these series as dependent variables in a numerical scheme. Alternatively, we could use a restricted grid of points with an appropriate space-differencing scheme. What the observations actually tell us is that virtually all scales, from thousands of kilometers to fractions of a kilometer, are present. We must therefore retain as many terms in the series as we can, or use as many grid points as possible, and in addition we must introduce new terms in the equations to represent the expected effect of the unresolved on the resolved scales. Absence of such terms was one of the inadequacies of early numerical weather forecasts; even today the optimum formulation of these terms has eluded us. Use of high-resolution grids over limited areas, which present-day computers allow, affords a partial solution.

A regular operational procedure which is more obviously mixed is known as Model Output Statistics (MOS). The procedure begins with the conventional operational numerical forecast. The output of this forecast consists of fields of temperature, wind, and humidity. Quantities which one may be more interested in forecasting, such as daily rainfall or snowfall amounts, are then predicted by linear regression, with predictors chosen from the output of the numerical model. The procedure appears to be competitive with a good subjective forecaster.

The MOS procedure follows a dynamical step by an empirical step. Some methods which have been examined but not yet put into operational use mix the dynamics and empiricism more thoroughly. Leith has observed that when a certain dynamical forecasting model is applied to an extensive set of initial conditions, the climate of the predicted weather patterns differs from that of the initial weather patterns. That is, the means, variances, etc. are different. This should not happen with a perfect forecasting model. Accordingly, he suggests adding a term to the dynamic equations to make them produce the correct predicted mean value. Other modifications can perhaps be made to correct the variances.

The above procedure would produce no corrections if the means and other statistics were correct, whether or not the forecasts were good. Faller and Schemm have proposed a method in which additional terms of a specified form are to be added to the dynamical equations to minimize the mean square forecast errors one time step in advance. The coefficients in these terms are to be determined by linear regression.

The method works well on artificial data, where the "true" weather pattern is available at every time step. To apply the method to real data, where the true pattern is generally available only at 12 - hour intervals, it would appear necessary to interpolate the true patterns. This may be impractical, because of the considerable high-frequency variance which characterizes the weather.

In methods which are based on empirical procedures, such as linear regression, the contribution of dynamics is likely to be in the choice of predictors. The dynamical equations are nonlinear, and they suggest that some nonlinear empirical scheme might produce good results. One method which is effectively nonlinear consists of choosing nonlinear functions of the values of present and past weather elements as predictors in a linear scheme. Since it is impossible to include all nonlinear functions, we may base our choice on the form of the dynamic equations.

A procedure which we have investigated begins by making a complete dynamic forecast, and then choosing dynamically predicted values of the weather elements, together with observed present and past values, as predictors in a linear regression scheme. Formally, it is somewhat like the MOS procedure. The dynamically predicted values are of course complicated nonlinear functions of the present values, and their use in the scheme yields decided improvement.

A variant of this scheme uses as predictors the first few time derivatives of the initial pattern, as given by the dynamic equations, in place of a single pattern several steps in advance. The time derivatives are easily obtained by numerically integrating the equations through a few uncentered forward time steps. Use of two or three time derivatives as predictors produces results at least as good as those obtained by the original procedure.

Experience with these procedures suggests that certain other procedures, which do not appear to have been tested, might prove useful. The most important nonlinear terms in the dynamic equations may be approximated by quadratic terms. Use of all linear and quadratic functions of a moderately large number of present weather elements does not appear possible, because it would be difficult to obtain a data sample whose size would greatly exceed the number of predictors. Use of linear and quadratic functions of a small number of weather elements is feasible, and it might be worth investigating in order to assess the quadratic contribution. However, in the dynamic equations the quadratic functions of the initial state are only time derivatives; they are not finite forward differences.

If the first time derivatives are quadratic, the second time derivatives are cubic. Use of all linear, quadratic, and cubic functions of some initial variables should yield a marked improvement over use of just linear and quadratic functions, except that it would intensify the problem of securing a large enough data sample.

A possible way to reduce the predictors might be to admit only those cubic functions which would appear in the second time derivative if the chosen quadratic functions appeared in the first. In view of our earlier results with time derivatives, this procedure might be equivalent to, but more easily implemented than, determining the quadratic function which, when iterated in a numerical forecasting scheme, would produce the best forecast a finite time in advance.

We suspect that if any real breakthrough in empirical forecasting is to occur, it will involve a procedure differing radically from any which are now being considered. Discovering and implementing the procedure should offer a challenging problem to mathematicians.