The assimilation of past data in dynamical analysis. I[1]

By K. MIYAKODA and OLIVIER TALAGRAND,[2] *Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, New Jersey 08540*

(Manuscript received February 15, 1971; revised version April 2, 1971)

ABSTRACT

A method is presented to reduce the error in meteorological data analysis by using observations taken prior to analysis time, extrapolating them to the present time with the prediction equations, and mixing them with the present observation data. The equations can be non-linear, irreversible, and three-dimensional in space. The time span of the extrapolated data is several days. One-dimensional linear as well as two-dimensional non-linear vorticity equations are treated as examples. The best estimate of the analyzed field is the linear combination of a number of the predicted solutions from the past data with weights included that depend on predictability decay as functions of the data's age.

1. Introduction

Suppose that an atmospheric model is completely accurate, and that the exact solutions of the equations are obtainable. Then, new observations are not needed after an initial condition is given precisely once and for all.

In reality, of course, the situation is much different. The model is far from perfect; observations always suffer some error; the atmosphere is constantly receiving unpredictable perturbations from outside; a finite difference method with its inevitable error is required to treat the differential equations; and uncertainty is intrinsic as a dynamical property of the atmosphere.

However, an advantage can be taken, to a limited extent, from the above implication, by supplementing an insufficient data set and thus reducing the error in the initial condition. The prediction model is used to retrieve the past data and assimilate them into the present condition.

The idea is not new. The mixing of the output of the yesterday's data is being performed operationally in the "objective analysis" (see e.g. Gilchrist & Cressman, 1954; Bergthorsson & Döös, 1955), though it is used only as the "first guess" field. In 1961, Thompson studied the problem of reconstructing the meteorological condition in the "data-hole" region which is surrounded by a data-rich area. He proposed a method of combining the analysis and prediction; accurate data is advected into the "hole" area from outside (see also Richardson, 1961; Smith, 1962).

In the communication and control theory, Kalman (1960) discussed the prediction of signals and the separation of signals from noise. Jones (1965*a*, *b*) and Petersen (1968) introduced this approach into the meteorological field. They modified the deterministic dynamical system to a kind of stochastic-dynamical system, where the equation to predict the next time step is designed so as to combine the previous prediction and current observations with the proper weights. This is one of the promising methods for continual data assimilation (so far, at least, as the linear dynamical equation is concerned).

[1] Some part of these contents was presented by Miyakoda on the conference on the Plan for United States participation in GARP, held in Boulder, Colorado, October 15–17, 1969.

[2] On leave from le Centre National de la Recherche Scientifique, Paris, France.

On the other hand, Sasaki (1969) devised a technique of "objective" analysis, in which not only the observed data but also any complicated dynamical constraints can be included, and thereby the off-time data at several time levels apart can be assimilated. Tadjbakhsh (1969) proposed a scheme for updating the solution of a linear equation by using newly received data and avoiding local discontinuity between the existing solution and the new data.

The objective of the present article is to propose a method to compute the initial condition using the time sequential data for a period of one day or perhaps several days. The purpose is not to design a system for the stochastic-dynamic prediction, but is just to determine an initial condition. In that sense, our study is similar to those of Gilchrist & Cressman (1954), Gandin (1963), and Sasaki (1969).

But the data to be used in our analysis are scattered over a period of several days, and the skill will crucially depend on the long range performance of the model. The governing equations are not linearized, but non-linear, and could be irreversible and complex. In that respect, the approach resembles that of Thompson (1961).

Our main concern does not lie in "updating" of the meteorological information, which is so relevant to the management of "real-time" data. The increase of accuracy in the analysis is our purpose. It turns out that a most important aspect is the determination of weights for the prediction and observation, and in this way, the discussion is close to that of Kalman and Petersen.

2. The case of the linear vorticity equation

To explain some of the essential features of the method, we begin with a simple linear equation which is used to simulate the time variation of an atmospheric wave (Charney & Eliassen, 1949).

Dynamical system

The one-dimensional, linear vorticity equation is:

$$\frac{\partial \zeta}{\partial t} + U\frac{\partial \zeta}{\partial x} + \beta v = 0 \qquad (2.1)$$

where U is the basic eastward flow which is constant, v is the perturbed flow of y-component, ζ is the vorticity, and β is the latitudinal change of coriolis parameter taken as a constant. It follows that $\zeta = \partial v/\partial x$.

With a certain initial condition for v (in practice, the actual data of January, 1966) and under the assumption that v is cyclic at the lateral boundary for x, equation (2.1) is solved numerically for any time as a function of x approximating the time and space derivatives by centered finite difference. Let us call the solution thus obtained the "true solution", i.e., v_{true}.

Next we take a solution of v at an arbitrary time. Out of the total gridpoints for x, we select some points (80 out of 143 in this case) randomly, which are assumed to be the "observation stations". v at these stations is deliberately disturbed by small random amounts, which may correspond to "observation errors". Thus we have another set of values, i.e., the "observed data", v_{obs}.

The problem posed is as follows:

Take a certain time t_0. Assume that observed data v_{obs} are available at the stations for t_0 and also for the previous times t_{-1}, t_{-2}, ... Under this condition, compute v at t_0, so that the computed v will compare well to the true solution at t_0.

Spatial analysis and prediction

First with v_{obs} at a certain time, say t_{-n}, an "objective analysis" is made for v, so that values of v are determined at all gridpoints. v_{anal} denotes the "analyzed data". Gandin's (1963) analysis procedure for space was employed.

With v_{anal} at t_{-n} as the initial data, the prediction is carried out with equation (2.1) from t_{-n} to t_0. Let us denote the solution at t_0, $v_{\text{pred}}(t_0|t_{-n})$ or simply $v_{\text{pred}}(t_{-n})$

Synthesis method

Likewise we compute $v_{\text{pred}}(t-1)$, $v_{\text{pred}}(t-2)$ etc. (see Fig. 1). It is assumed for simplicity that the observation stations remain the same for all cases.

In order to get a supposedly better estimate, a synthesized solution is constructed as follows,

$$v_{\text{syn}}(t_0) = \frac{1}{N}\left[v_{\text{anal}}(t_0) + \sum_{n=1}^{N-1} v_{\text{pred}}(t_{-n})\right] \qquad (2.2)$$

If there is no systematic bias in the measurement error or in the application of Gandin's

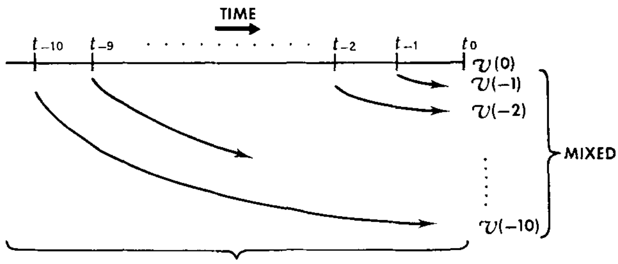

Fig. 1. Scheme of data assimilation. For instance, $v(-2) = v(t_{-2})$.

analysis, the errors in v_{pred} are random, and consequently they should cancel each other by averaging a sufficient number of solution v_{pred}, as in equation (2.2). Fig. 2 is an example of the solution thus obtained. $N = 11$ is taken in this particular case. v_{syn} is indicated by small crosses. v_{true} is the solid curve. v_{obs} at t_0 is shown by small dots. The agreement between v_{syn} and v_{true} is excellent.

Discussion

For the solution in Fig. 2, the root-mean-square error of v is evaluated, i.e.,

$$\text{rms} = \left[\sum_i (v_{\text{syn}} - v_{\text{true}})^2/n\right]^{\frac{1}{2}} \tag{2.3}$$

where $\sum_i$ is the summation over the entire gridpoints and n is the total number of gridpoints.

In a way similar to the above, the synthesized solutions of v are obtained for $N = 1, 2, 3 \ldots$, which means that one, two, three ... solutions of v_{pred}, respectively, are used for the synthesis. The rms error for each v_{syn} is computed. For each N, several values of v_{syn} have been computed. The results are shown in Fig. 3. The central curve is the ensemble mean E (rms) of all values of rms for N. The upper limit and the lower limit in our test cases are also displayed.

With increase of N, the error expressed by the rms is reduced. If the differences between the predictions and the true solutions at time t_0 were absolutely random, the curve would be proportional to $N^{-\frac{1}{2}}$ (dashed line), but in reality there is a slight deviation (solid line). Probably this is due to the bias in the spatial objective analysis or a biased selection of observation stations. The range marked by "Assumed range of observation error" in the figure is the range of the error between an "observed state" and a "true state".

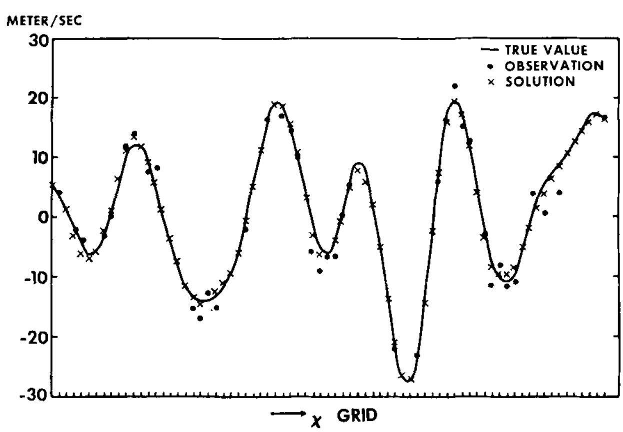

Fig. 2. A synthesized solution of v as a function of x in the one-dimensional case.

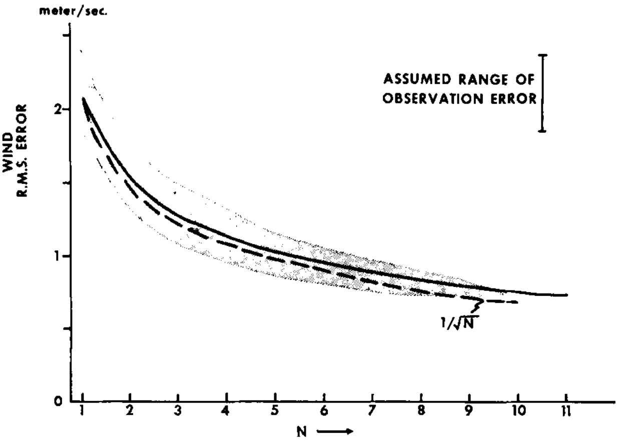

Fig. 3. The root-mean-square error of v in the one-dimensional case. The abscissa is the sample number N.

3. The case of the two-dimensional non-linear equation

In the prediction treated in the previous section, the magnitude of prediction error does not grow with time. Namely,

$$\frac{1}{n}\sum_{i}(v_{\text{pred}}(t\,|\,t_0)-v_{\text{true}}(t))^2/n = \text{const.} \qquad (3.1)$$

This is because the equation does not contain any process of dynamical instability, which mathematically corresponds to the fact that equation (2.1) is linear. As a consequence, all the different predictions at time t_0 contain as much useful information and the same weight $1/N$ was given to all of them.

We shall next take a more general example, where the error increases with time.

Dynamical system

The non-divergent inviscid vorticity equation on a rotating frame, with use of the stream function ψ, is written as

$$\frac{\partial \nabla^2 \psi}{\partial t} - J(\nabla^2\psi + f, \psi) = 0 \qquad (3.2)$$

where J is the Jacobian operator, and f is the coriolis parameter.

For the numerical computation, the finite difference approximation is used, and Arakawa's (1966) formulation is employed for spatial differentiation. The "leap-frog" scheme is used for time integration.

The domain is a zonal channel encircling the globe with the north-south edges at 65° N and 15° N latitudes. The boundary conditions are that ψ is cyclic in the x-direction, and that ψ is constant along the north or south boundary. The total number of gridpoints is 72×11 (the x- and y-directions).

Spatial analysis and prediction

The procedure is the same as for the one-dimensional case. A run for a given period was made from an initial state taken from a real situation of January, 1966, and the solution is used as the reference run, assumed to represent the time solution of the system. As for the simulation of the observations, 352 "observation stations" were randomly chosen from the two-dimensional array of 792 gridpoints. The root-mean-square of the "observation error" of geopotential height is assumed to be 20 m. (Geopotential height $= f\psi/g$, where g is the acceleration of gravity and f is taken as the value of 45° latitude.) An objective analysis is performed in the same way as in the previous case, according to Gandin's method. In order to evaluate the root-mean-square error of the analysis alone, a number of analyses were taken, independent of the observation error, and the analysis error was found to be 17.5 m.

The key time t_0 was chosen as the final time of the reference run. All the synthesized states were computed using predictions taken from a set of 25 different forecasts, which had been performed from states "observed" at regular

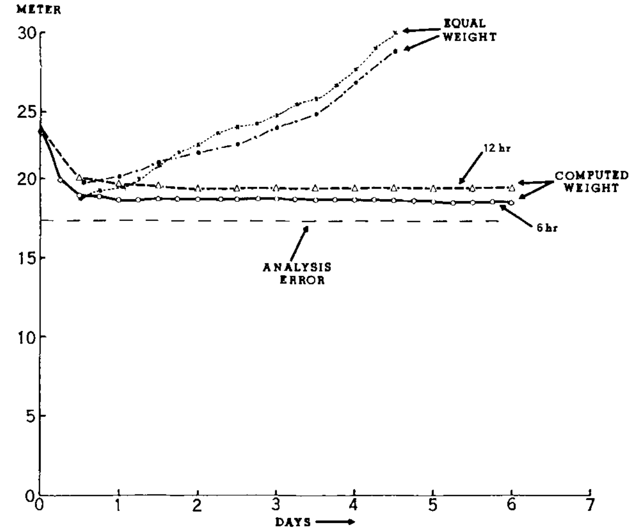

Fig. 4. The root-mean-square error of $f\psi/g$ in the case of the non-linear vorticity equation. The abscissa is sample number in terms of days.

intervals 6 hours apart for the period from $t_0 - 6$ days to t_0.

The streamfunction field ψ (not the velocity field, as in the one-dimensional case), was used to compute synthesized states and to determine an "optimal estimate", as will be seen below.

Preliminary study

The synthesized solution of ψ was first obtained, as in the one-dimensional case, by

$$\psi_{\text{syn}}(t_0) = \frac{1}{N} \sum_{n=0}^{N-1} \psi_{\text{pred}}(t-n), \tag{3.3}$$

where $\psi_{\text{pred}}\ (t_0) = \psi_{\text{anal}}\ (t_0)$. Two series of synthesized solutions were computed. In the first one, N varying from 1 to 25, all the forecasts, performed from initial times 6 hours apart, were used. (The addition in the synthesis of a new forecast for each new value of N is achieved antichronologically for example, for $N = 2$, $\psi_{\text{syn}} = 1/2\ (\psi_{\text{pred}}\ (t_0) + \psi_{\text{pred}}(t_{-1}))$. In the second series, N varying from 1 to 13, the forecasts performed from initial times 12 hours apart were used.

To compare the synthesized solutions with the true solution, the root-mean-square error was computed for each synthesized solution. The results for both series of solutions are shown in Fig. 4 (curves marked *equal weight*) which gives the rms error as a function of the time span over which the initial times are distributed for the forecasts used in the synthesis. It is interesting to note that the error first decreases with time until 0.5 days and then increases rapidly. This means that for large N the synthesis was not successful. Evidently the old aged data were given too large a weight and damaged the analysis rather than improved it.

The optimal estimate

Instead of equation (3.3) let us take the more general formula

$$\psi_{\text{syn}}(t_0) = \sum_{n=0}^{N-1} w_n \cdot \psi_{\text{pred}}(t_{-n}) \tag{3.4}$$

where the weights w_n satisfy the condition

$$\sum_{n=0}^{N-1} w_n = 1 \tag{3.5}$$

Next let us consider the rms error F

$$F = E\overline{(\psi_{\text{true}} - \psi_{\text{syn}})^2}^H \tag{3.6}$$

where $\overline{(\quad)}^H$ means an average over the two-dimensional area and E a statistical average.[1]

[1] Actually, due to the fact that the stream-function ψ is determined, up to an additive constant, ψ is to be replaced in formula (3.6) and the rest of the paragraph, by $\psi - \overline{\psi}^H$.

In order to make the optimal estimate of ψ_{syn}, we want to determine $w_n(n = 0, 1, \ldots N\text{-}1)$ such that F becomes a minimum.

Using (3.4) and (3.5), F is written

$$F = E \sum_{n,k} w_n \cdot w_k \overline{(\psi_{\text{true}} - \psi_{\text{pred}}(t_{-n}))\,(\psi_{\text{true}} - \psi_{\text{pred}}(t_{-k}))}^{H} \tag{3.7}$$

Let us define the "error covariance"

$$P(n, k) = E\overline{(\psi_{\text{true}} - \psi_{\text{pred}}(t_{-n}))\,(\psi_{\text{true}} - \psi_{\text{pred}}(t_{-k}))}^{H} \tag{3.8}$$

This quantity is similar to that defined under the same name by Kalman (1960). Hence,

$$F = \sum_{n,k} w_n \cdot w_k P(n, k) \tag{3.9}$$

Using the method of Lagrange's multiplication, we find w_n's which minimize the quadratic form F, i.e.,

$$\left.\begin{aligned} &\frac{\partial F}{\partial w_n} + \lambda = 0 \quad \text{for } n = 0, 1, \ldots \quad N-1 \\ &\sum_k w_k = 1 \end{aligned}\right\} \tag{3.10}$$

where λ is the multiplier. (3.10) is a system of $N+1$ simultaneous linear equations with $N+1$ unknowns (the N of w_n plus λ).

The quantities $P(n, k)$ were determined from statistical studies on a certain number of the runs. Six different statistical estimates of each of the quantities $\overline{(\psi_{\text{true}} - \psi(t_{-n}))\,(\psi_{\text{true}} - \psi(t_{-k}))}^{H}$ were obtained, and $P(n, k)$ were taken as the averages of these estimates. From these values of $P(n, k)$, the optimal weights w's were determined. The table gives the examples of values of w_k's both for a 6 hour interval ($k = 1, 2, 3 \ldots$ corresponding to the "age" of 0, 0.25, 0.5 … days) and a 12 hour interval ($k = 1, 2, 3 \ldots$ corresponding to 0, 0.5, 1 … days).

Beyond one and a half days, the w_k's turned out to be small values and to fluctuate in an apparently random way, showing a low statistical significance. So no useful information can be derived from forecasts older than 1.5 days. When N varies, the w_k's remain practically constant beyond certain values of N.

Table weight w_k

Time interval	"Age" of the forecast (days) 0	0.25	0.5	0.75	1	1.25	1.5
6 hours	.519	.196	.135	.083	.009	.046	.018
12 hours	.617		.251		.068		.067

Fig. 4 shows the variation with N of the rms error between the synthesized states and the true state, for both cases of 12 hour interval and 6 hour interval (marked *computed weight*). The error decreases with increase of N but levels off after about 1.5 days corresponding to the fact that a negligible weight is given to the additional forecasts introduced in the synthesis. The reduction from the original rms error ($N = 1$ or 0th day) is about 20%; the reduction is a little larger in the case of the 6 hour interval probably because more statistical samples are used. It must be noted, too, that in both cases (6 hour and 12 hour intervals) the final reduction of the error is not much more than that achieved after 0.5 days. The asymptotic level probably depends on the spatial and temporal distributions of the observed data, and the measurement errors, and also on the error of the spatial "objective analysis". This level, in the case of the present results, is close to the level of the systematic error occurring in the objective analysis process (marked "analysis error" on the diagram). Further studies (Part II of this paper) seem to indicate, however, that the closeness of the two levels is coincidental. Still the problem remains as to what factors really determine the final level of reduction of the error and by what means it can be reduced further (especially, modification of the data density and frequency, and of the objective analysis technique).

Error covariances

The forecasts to determine $P(n, k)$ were extended to study the property of the error growth in the model. Twelve samples were used and 30 day predictions were made. Fig. 5 shows $P(n, n)$, which are the diagonal elements in the matrix consisting of $P(n, k)$, and correspond to the variance or the mean-square-error of ψ. The upper limit and the lower limit of the error among 12 samples are also indicated.

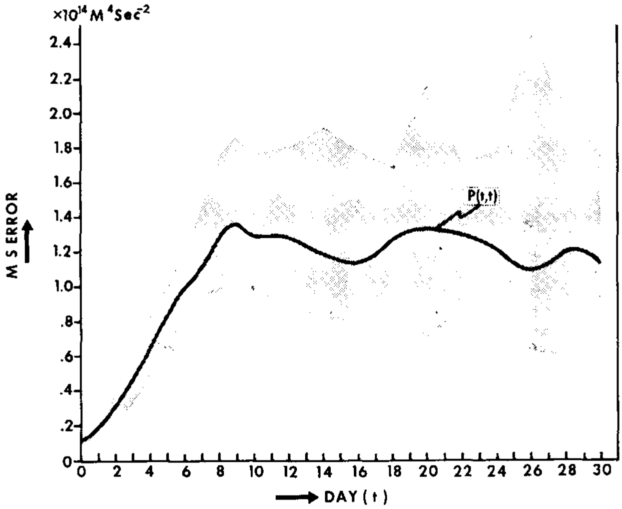

Fig. 5. The mean-square error of ψ with time for the case of the vorticity equation. The solid line is the ensemble average of the mean-square error. This curve indicates the "predictability decay".

This may remind us of the similar quantity computed by Lorenz (1965), Charney (1966) (National Academy of Sciences) and Smagorinsky (1969), Sundström (1969) in the study of the "predictability", though they are for the baroclinic equations except the last paper. The curve in the figure is the "Predictability decay". The prediction error grows due to the existing dynamical instability, however small the initial disturbance is.

Fig. 5 shows that the error starts at a small value, increases first exponentially (with a characteristic time of about 2.5 days), then linearly, and levels off after 9 days. The range of the variability is very small in the first 7 days, and it is then broadened. (Apparently the time of level-off in the barotropic case is shorter than that in the baroclinic case. It is later than 3 weeks in the latter case.)

4. Conclusions

The observation error can be curtailed by averaging a number of prediction results which are obtained from past data with a dynamical equation. For the case of a linear equation, the residual of error decreases as the sample number increases.

For a non-linear equation, a simple average of the prediction results did not improve the accuracy of analysis, but actually increased the error as the sample number increased. This is due to the continuous error growth with time in the prediction, which is called "predictability decay".

The optimal assimilation of the past data is achieved by ascribing proper weights according to the data's age. The weights are determined by the error covariance which reflects the predictability decay. However, the accuracy of the dynamical analysis was not able to be reduced beyond a certain limit.

REFERENCES

Arakawa, A. 1966. Computational design for long-term numerical integration of the equations of fluid motion: two dimensional incompressible flow. Part I. *Journ. Computational Physics 1*, 119–143.

Bergthorsson, P. & Döös, B. 1955. Numerical weather map analysis. *Tellus 7*, 329–340.

Charney, J. G. & Eliassen, A. 1949. A numerical method for predicting the perturbations of the middle latitude westerlies. *Tellus 1*, 38–54.

Cressman, G. P. 1959. An operational objective analysis system. *Mon. Wea. Rev. 87*, 367–374.

Gilchrist, B. & Cressman, G. P. 1954. An experiment in objective analysis. *Tellus 6*, 309–318.

Gandin, L. S. 1963. *Objective analysis of meteorlogical fields.* Gidrometeorologicheskoe Izdatel' stvo, Leningrad, 1963. Translated from Russian in English, Israel Program for Scientific Translations, Jerusalem, 1965. 242 pp.

Jones, R. H. 1965*a*. Optimal estimation of initial

conditions for numerical prediction. *J. Atm. Sci.* *22*, 658–663.
Jones, R. H. 1965*b*. An experiment in non-linear prediction. *Journ. Appl. Meteor.* *4*, 701–705.
Kalman, R. E. 1960. A new approach to linear filtering and prediction problems. *Journ. Basic Engineering* *82*, 35–45.
Lorenz, E. N. 1965. A study of the predictability of a 28-variable atmospheric model. *Tellus* *17*, 321–333.
National Academy of Sciences—National Research Council. 1966. *The feasibility of a global observation and analysis experiment.* Publication 1290, Washington, D.C. 172 pp.
Petersen, D. P. 1965. Recursion formulas for sequential estimation. *IEEE Trans. on Information Theory*, *IT–11*, 18–30.
Petersen, D. P. 1968. On the concept and implementation of sequential analysis for linear random fields. *Tellus* *20*, 673–686.
Richardson, N. N. 1961. Numerical tests of a method for dynamic analysis in regions of poor data coverage. *Tellus* *13*, 353–362.
Sasaki, Y. 1969. Proposed inclusion of time variation terms, observational and theoretical, in numerical variational objective analysis. *Journ. Meteor. Soc. of Japan* *47*, 115–124.
Smagorinsky, J. 1969. Problems and promises of deterministic extended range forecasting. Wexler Memorial Lecture, Jan. 20, 1969. *Bull. Amer. Meteor. Soc.* *50*, 286–311.
Smith, F. B. 1962. Objective analysis of the vorticity field with a region of no data. *Tellus* *14*, 281–289.
Sundström, A. 1969. Stability theorems for the barotropic vorticity equation. *Mon. Wea. Rev.* *97* 340–345.
Tadjbakhsh, I. G. 1969. Utilization of time-dependent data in running solution of initial value problems. *Jour. Appl. Meteor.* *8*, 389–391.
Thompson, P. D. 1961. A dynamical method of analyzing meteorological data. *Tellus* *13*, 334–349.

ИСПОЛЬЗОВАНИЕ ПРОШЛЫХ ДАННЫХ В МЕТЕОРОЛОГИЧЕСКОМ АНАЛИЗЕ ЧАСТЬ I

Представлен метод, уменьшающий ошибки при анализе метеорологических данных путем использования наблюдений за предыдущие сроки, экстраполяции их к моменту анализа с помощью прогностических уравнений и комбинирования полученных данных с данными текущих наблюдений. Уравнения могут быть нелинейными, необратимыми и трехмерными. Данные для экстраполяции используются за несколько дней. В качестве примеров рассматриваются одномерное линейное и двумерное нелинейное уравнения вихря. Наилучшее приближение к анализируемому полю дает линейная комбинация предсказанных решений с весами, зависящими от предсказуемости, которая является функцией прогностического интервала.