U.S.DEPARTMENT OF COMMERCE
NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION
NATIONAL WEATHER SERVICE
NATIONAL METEOROLOGICAL CENTER

OFFICE NOTE 363

COMPREHENSIVE HYDROSTATIC QUALITY CONTROL AT THE NATIONAL METEOROLOGICAL CENTER

WILLIAM G. COLLINS
LEV S. GANDIN

DEVELOPMENT DIVISION

NOVEMBER 1989

ABSTRACT

The Comprehensive Hydrostatic Quality Control (CHQC) of rawinsonde data of height and temperature at mandatory isobaric surfaces designed and implemented at the National Meteorological Center in Washington is described in detail. Main principles of the complex, or comprehensive, quality control are discussed, followed by a brief description of the CHQC design and implementation at NMC. The CHQC algorithm is presented, with particular emphasis on the Decision Making Algorithm. Numerous examples taken from the operational CHQC outputs illustrate the CHQC performance in general, as well as its reaction to errors of various types and to their combinations.

1. Introduction

The term "quality control of meteorological data" is often understood in a wide sense, encompassing all actions connected with the quality, and often also with the quantity, of the data. Alongside with this understanding, or maybe even instead of it, it is worthwhile to understand the term in a narrow sense, as a set of actions directed against so-called rough errors in meteorological information. Unlike random errors, which influence all meteorological data but are usually comparatively small, rough errors may be, and often are, large enough, but they are present in a small part of all data. Each rough error is due to some definite cause, which may be a deficiency in observation, processing, or communication. The aim of the quality control is to detect the rough errors and then to correct every erroneous datum, or, if this proves to be impossible, to reject it.

The necessity to perform the quality control of meteorological data had been recognized long ago. This task has become much more important during the last decades, particularly in connection with numerical weather prediction. There exists numerous evidence that retention of erroneous data, or even rejection of too many correctable data, may substantially distort the objective analysis results and lead therefore to large errors in predicted fields. The relative importance of the quality control is permanently increasing alongside with improvement of the analysis and prediction models, the deficiencies in initial data thus becoming the major source of erroneous forecasts. It is to be stressed in this respect that the more advanced is the prediction model, the more sensitive it is to the errors in initial data.

At the same time, the quality control problem has become much more complicated nowadays due mainly to two interconnected factors, (1) a dramatic increase in the amount of operationally available data, and (2) the development and implementation of new kinds of meteorological observations, particularly of satellite soundings. Due to the huge amount of data, it is absolutely impossible to perform their quality control manually, particularly under operational conditions.

The necessity to have an automated, computerized, quality control had been recognized at the beginning of the NWP era, and some methods of such quality control were proposed at that time. Nevertheless, there still exists an opinion, or rather a superstition, that the quality control is to be based on human intuition and experience, and that it is impossible to program for a computer the complicated ways of judgement performed by a specialist. The role of a computer is often thought of as that of a means to display the information in a form convenient for a visual inspection and quality control by a human being.

This opinion is, of course, wrong. Any chain of judgements, however complicated it is, may be easily coded for a computer, provided that it is precisely formulated. The code has to contain not only various procedures for detecting suspicious data, the so-called checks, but also a decision making algorithm (DMA) designed to analyze the results of checks in order to conclude which of the suspected data are wrong and, if possible, to correct them.

There are two main advantages of the automatic quality control over the subjective one: the speed and the objectiveness. The speed even of a moderate computer makes it capable of performing the quality control of the whole amount of operational data, the aim not achievable even by a huge team of human beings. The objectiveness of the automatic quality control is also very important. There existed many cases when the same situation with suspected data was differently treated by different specialists. (Of course, such cases never occur with the automatic quality control.) It is important to understand, however, that the automatic quality control is objective only in the same sense as this is valid for objective analysis, numerical prediction, and so on. The results of the quality control do depend on the coded algorithm and may substantially change due to even slight modifications of the algorithm.

What has been said about the advantages of an automated quality control does not imply that there is, or will be, no place for human activity connected with it. The monitoring of the quality control performance is very important for improving both the algorithms applied and the data quality itself. Although many stages of this quality control monitoring may, and should, also be performed automatically, there still exists much room for subjective considerations and decisions made by specialists. Moreover, experience shows that there exist some cases, though very rare ones, when the data under check are distorted in such a complicated way, that it is necessary to devote a special, separate, part of the DMA to every or almost every such combination of rough errors. From a purely theoretical point of view, this may be done, but that would be impractical. For such very rare cases, it is much better to leave the final decision for a specialist to make, requiring from the code only to provide the specialist with all information obtained during the attempt to make the decision automatically.

The possibility to have data displayed on a screen is very useful for both quality control monitoring and solving complicated cases. Let us stress again, however, that these displays have to complement the automatic quality control, but by no means to be used instead of it.

Speaking about some superstitions connected with the quality control, it is necessary to mention the requirement that every datum received at a prognostic center has to be retained, however erroneous it is. Even if it is absolutely clear that the error has originated not in the course of observation but on a communication line, and even if the error may be corrected with absolute confidence, - even in these cases it is presently required that the erroneous datum should be preserved, if not instead the corrected one then, at least, alongside with it. Undoubtedly, this requirement, like almost every requirement expressed in categorical terms, is unreasonable, and one has to be flexible when deciding whether or not to preserve an erroneous data. It is true that very often it is difficult to decide confidently, whether a suspected datum is wrong or correct, and as long as such a decision cannot be made, every suspected datum has to be retained. Also often, however, it happens that there is no slightest doubt that the quality controlled datum is wrong and it is absolutely clear how to correct it. In every such case, there exists no reason to retain the erroneous datum, even alongside with the corrected value. Our experience shows that no one at NMC or elsewhere ever tried to access any datum after it had been confidently corrected at NMC by the Comprehensive Hydrostatic Quality Control described below, although such data were retained because we were obliged to follow this requirement. It seems just improper for a prognostic center to apply confidently corrected data for its internal use, while concealing this from other users of the information. Analogously, it also happens sometimes, though more seldom, that the datum under quality control is found to be definitely wrong and, also definitely, uncorrectable. Once again, there is no slightest reason to preserve such data, they have just to be rejected.

May we tolerate that a quality control procedure results sometimes in rejection of a datum which is actually correct or even introduces an erroneous correction? The common opinion is that situations of this kind have to be completely avoided at any price. In reality, however, even if this aim is achievable, the price would be too high: a quality control system obeying this requirement would be capable of managing only a small percentage of actually occurring errors. There exist, in principle, two kinds of erroneous behavior of any quality control: the errors of the first kind, when it does not reject or correct wrong data, and the errors of the second kind, when it rejects correct data or introduces wrong corrections. The task of every reasonable quality control design is to make the numbers of errors of both kinds as small as possible, preserving some balance between these two numbers.

Although the general statements above seem evident for us, we felt it desirable to discuss them in some detail, because they differ from views expressed by many specialists, particularly, though not exclusively, by those not involved directly in the quality control design or application.

As mentioned above, methods of automatic quality control began to be developed and implemented soon after the first successes of numerical weather prediction, about four decades ago. Nowadays, such methods are in operational use at every center producing numerical weather forecasts. Analysis shows, however, that the quality control methods now in use do not differ much from those proposed long ago, which may be caused by the fact that the quality control is still considered by many specialists as a purely technical problem. In any case, the presently used quality control systems still contain some more or less evident shortcomings. So, the so-called flagging is widely used in connection with various quality control procedures. A special digit, a "flag", is assigned to every datum suspected by one or another check. One of the purposes of this is not to take erroneous data into account in the forthcoming objective analysis, not losing, at the same time, these data at all. One may, certainly, argue that at least an attempt to correct erroneous data has to be undertaken, and also, whether it is really desirable to retain definitely wrong data. There exists, however, another, more advanced kind of flagging based on the fact that several more or less independent quality control checks have to be applied to each suspected datum in order to decide whether it is correct or wrong. The usually applied procedure for that is based on flags assigned to the datum in question by various checks. This means, however, that the quantitative information achieved by using each quality control check is replaced by a qualitative, or semi- qualitative information - by a flag. This loss of information results in a substantial decrease of the quality control possibilities, particularly of the possibility to estimate, and thus to correct, the error. It may even happen that, according to one of the checks, the value in question was too high, while another check diagnosed the same value as being too low. This fact will not be, however, reflected by the flags, and the datum will be rejected, although some other datum or data must definitely be wrong in such a case, either alongside with the datum in question or even instead of it. Even if the effect will be less dramatic, it is absolutely clear that by replacing the quantitative results of various checks by flags we can only lose important information while gaining practically nothing. Nevertheless, flagging of data is continuing to be widely used almost everywhere, and particularly at NMC.

The application of flagging procedures is, perhaps, the major shortcoming of quality control systems now in use, but unfortunately, far from the only one. For example, there is no reason for data containing very large errors to be rejected before other tests are applied. In fact, the larger an error, the higher is the probability that the error has originated not in the course of observation but later, particularly on a communication line. In many cases, such data may be confidently corrected and used after that for many purposes, including the quality control of other data.

In general, the opinion shared by many specialists is that the existing quality control systems, or at least many properties of them, are due to historical reasons rather than to logical ones. In any case, there exist numerous examples where objective analysis and subsequent forecast showed deficiency in one or another area as a direct result of incorrect quality control.

Based on the decision by W. Bonner, NMC Director, some work has begun at NMC directed towards design of the new NMC data quality system from scratch, rather than improvement of the existing system. The design includes, among other things, the application of the so-called Complex, or Comprehensive, Quality Control (CQC) approach (Gandin, 1988, 1989). The main CQC idea is that several checks, actually as many of them as possible, have first to be applied to the data under the quality control, and any decision regarding correction or rejection of some of the data has to be made only after the application of all available checks and to be based on results of all of them. This means that the CQC algorithm has to consist of two major parts, the first of them being the application of all checks, and the second one the Decision Making Algorithm (DMA).

The first stage of the work at NMC in this direction, the design and implementation of the Comprehensive Hydrostatic Quality Control (CHQC), is described in this paper.

Only one check, the hydrostatic one, is used within the CHQC, and this may seem to contradict the CQC idea. In fact, however, the hydrostatic check is applied to many layers for every report, and if decisions are based on the analysis of its results for several layers, as really is the case with the CHQC, then we may consider it as a kind of CQC. It does contain a comparatively advanced DMA, which is also characteristic for the CQC approach.

We have to admit, of course, that the CHQC design was only the first step in the design and implementation of the new NMC data quality control system. Among various kinds of observations, it deals only with rawinsonde data. Only heights and temperatures are subjected to the quality control, leaving aside wind and humidity. Only mandatory level data are quality controlled as yet, not the significant level data. Finally, only one check, the hydrostatic one, is included.

Despite all these limitations, the CHQC proves to be very productive, which is due to the fact that the hydrostatic redundancy, caused by the presence of data on both heights and temperatures of mandatory surfaces in rawinsonde reports, is the most pronounced redundancy in meteorological information available at prognostic centers. It may be mentioned in this respect that this redundancy would be even higher if height, temperature and pressure were measured independently, in other words, if the hydrostatic equation were not used to derive one of these parameters.

2. Design, testing, and implementation

It is highly desirable, when beginning the quality control design for any data, to have information on rough errors in the data: what are the major causes of these errors, to what extent may it be possible to detect and even to correct erroneous data, and most important, how often the errors occur? Unfortunately, such information is almost never available at this stage, just because the only way to obtain reliable information of this kind is to apply the quality control procedure to the data in question. It is particularly so because an overwhelming majority of data do not contain rough errors, and in order to obtain more or less reliable statistics about the errors it is necessary therefore to apply the quality control procedure to very large amount of data. The situation in this respect was even worse than it could be if existing quality control methods paid any attention to the causes of rough errors.

We have found ourselves in this situation at the beginning of the CHQC design. We did know, that not very long ago, there were many rough errors in rawinsonde data on height and temperature of mandatory isobaric surfaces and that the errors originated mainly on communication lines due to human errors (Gandin, 1988). At the same time, we knew that the numbers of these errors had permanently decreased because of the computerization of both data processing at rawinsonde stations and communication procedures, and it was not clear whether the errors of this kind continue to exist nowadays, and if they do, how often they occur.

We decided therefore to proceed sequentially: to design first a comparatively simple quality control algorithm, to apply it to a sufficiently large amount of arriving data, and then, depending on results of this test, to decide what to do next.

This work was begun in January 1988, and the overall situation became clear after several months. The main conclusion is that the geographical distribution of the errors in question is highly non-homogeneous. There exist countries - USA, Canada, some West-European countries - where the processing and communication procedures are completely, or almost completely, computerized, and rough "hydrostatic" errors, (errors detectable by the hydrostatic quality control) occur therefore very seldom. At the same time, there still exist large areas where the computerization did not take place, or at least was not complete enough, so that there still remain a substantional number of hydrostatic errors in reports coming from these areas. For example, USSR, India and continental China produce together about a half of all hydrostatic errors. In general, about 7 or 8% of all rawinsonde reports received at NMC contain at least one hydrostatic error each.

We have also found, at this stage, that many of these errors belong to the category of what may be called "simple" errors, like an error in only one digit expressing, e.g., the number of hectometers, or an error in the temperature sign. This demonstrates that the majority of the errors are caused by human mistakes happening in the course of the data processing at stations and particularly in the course of communicating them. Many such errors, if they are large enough, may be univaluedly diagnosed, so that data containing them may be confidently corrected. In order to be capable of doing so, the Decision Making Algorithm has been substantially improved. The present DMA recognizes most often occuring types of simple errors, namely, one-digit errors, errors resulted in transposition of two or more digits, sign errors (in temperature), and combinations of sign errors with one-digit or transposition ones. By doing so, the DMA is also capable, in many cases, of correcting so-called "shifting errors", when one digit is missing, all others being correct.

After extensive testing, the Comprehensive Hydrostatic Quality Control has been implemented operationally at NMC on December 14, 1988, first for two global data assimilation times a day and soon after that it has begun to be applied for all data dumps. It has not replaced, as yet, any of existing checking procedures, but is applied before them. Experience shows that data confidently corrected by the CHQC are practically never rejected by subsequent checks.

The CHQC performance is being carefully monitored both by the NMC Meteorological Operations Division specialists and by the designers, by means of printed outputs. In addition to them, monthly summaries are also produced by the computer at the end of each month. They are also sent to some other centers, both within the United States and outside.

There also exists a kind of human-machine interaction in the course of the CHQC. If an error happens to be at the lowest reporting level (a so-called Type 4 error, see section 3c), then the DMA proposes two possible corrections, either of height or of temperature. The same is true with the suspected errors at the highest reporting level (Type 5 errors). Having corresponding outputs at hand, a MOD staff member decides in each such case, which of the two corrections, if any, to make.

Operational use of this procedure allows to correct more erroneous data, than it would be possible without it. We believe, however, that the main achievement connected with this procedure was the demonstration of rational ways of the human-machine interaction in the course of quality control, when almost everything is performed automatically, so that specialists have enough opportunity and time to make final decisions in rare cases when that cannot be done univaluedly by the algorithm.

There is no doubt that in the near future, when the hydrostatic check will be used in a complex with other, statistical, checks, an overwhelming majority of Type 4 and 5 errors will be dealt with entirely automatically, so that the MOD specialists will be able to devote their efforts to decision making in much more complicated cases. Experience gained with the present interaction procedure will be very useful in this respect.

The design of a new, improved version of the CHQC has begun soon after the previous one had been implemented. The aim was threefold. First, it has been found that errors at two adjacent levels occur more often than we expected, and it was desirable to make the Decision Making Algorithm capable of detecting and correcting them. Secondly, it was possible to make the DMA criteria for hydrostatic errors of various types more consistent with each other. Finally, a reordering of the DMA has been performed in order to facilitate its further generalization when statistical quality control checks will be included. Detailed testing of the new algorithm in parallel with the operational one and corresponding improvements of the new algorithm required several months. On July 12, 1989, the new version was implemented operationally instead of the previous one. Only this version will be described in the next Section.

3. The method

The comprehensive hydrostatic quality control is based upon an examination of the pattern of hydrostatic residuals (to be defined below) caused by errors. It is a rather easy problem to determine what the residuals would be for a particular error or pattern of errors. The actual problem consists in finding the errors which caused a particular pattern of residuals. This "backward" problem is much more complex, especially in view of our examination of mandatory level data only and the resultant approximate agreement of the data with the hydrostatic equation. These issues will be made more clear in the discussion below.

The hydrostatic equation integrated through a layer between pressures p_i and p_{i+1} may be written in the form

$$z_{i+1} - z_i = -(R/g) \int_{p_i}^{p_{i+1}} T_v \, d(\ln p) \qquad (1)$$

where z is height, p is pressure, R is the gas constant for dry air, and T_v is the virtual temperature. A sample of data over the globe shows that the effect of humidity, that is, the mean difference between mandatory level heights solved for using virtual temperature and "dry" temperature, is about 5 meters in the 1000-850 hPa layer, 2 meters in the 850-700 hPa layer, and negligible above. We want to isolate the examination of heights and temperatures from any possible errors in the dew-point temperature depression. One approach would be to account for humidity by using a standard or climatological profile, but since we are always capable of detecting only those errors in the data that are significantly larger than the humidity effect, we have chosen not to explicitly take into account the humidity influence. Between mandatory levels, the hydrostatic equation is written as

$$z_{i+1} - z_i = \frac{R}{g} \left[\frac{T_i + T_{i+1}}{2} + t_i^{i+1} \right] \cdot \ln \left[\frac{p_i}{p_{i+1}} \right] \qquad (2)$$

where T is the dry temperature and t_i^{i+1} is the adjustment to make Eqn. (2) exact; it combines any random effects of humidity errors and nonlinearity of temperature as a function of ln(p), since the average temperature at the top and bottom of the layer is used to represent an integrated mean temperature. (It does not include the small consistent effects of humidity, which are merely ignored.) In a large ensemble of cases, the average of t_i^{i+1} is assumed to be zero. It is noted that at this point, it is still assumed that the temperatures and heights contain no errors.

By shifting the temperatures to Celsius, Eqn. (2) may be written in the form:

$$z_{i+1} - z_i = A_i^{i+1} + 2B_i^{i+1} \left[\frac{T_i + T_{i+1}}{2} + t_i^{i+1} \right] \tag{3}$$

where

$$\begin{aligned} A_i^{i+1} &= (RT_0/g) \ln(p_i/p_{i+1}) \\ B_i^{i+1} &= (R/2g) \ln(p_i/p_{i+1}) \end{aligned} \tag{4}$$

and T_0 = 273.15 K. The hydrostatic residual is defined (see Fig. 1) as

$$s_i^{i+1} = z_{i+1} - z_i - A_i^{i+1} - B_i^{i+1} (T_i + T_{i+1}) \tag{5}$$

Examination of the pattern of hydrostatic residuals forms the essence of the method of hydrostatic quality control. In the absence of a rough error:

$$s_i^{i+1} = 2\, B_i^{i+1}\, t_i^{i+1} \tag{6}$$

Now assume that the heights and temperatures are composed of an exact value (subscript o) plus a rough error (primed):

$$\begin{aligned} z_i &= z_{oi} + z'_i \\ T_i &= T_{oi} + T'_i \end{aligned} \tag{7}$$

With rough errors, the residuals are:

$$s_i^{i+1} = z'_{i+1} - z'_i - B_i^{i+1}(T'_i + T'_{i+1} - 2t_i^{i+1}) \tag{8}$$

Eqn. (8) forms the basis for further development of the method of hydrostatic checking and correction.

The most general case to be considered will involve the residuals in three adjacent layers with at most two errors at the interior levels. The equations are

$$s_1^2 = z'_2 - B_1^2 \bullet (T'_2 - 2t_1^2)$$
$$s_2^3 = z'_3 - z'_2 - B_2^3 \bullet (T'_2 + T'_3 - 2t_2^3) \qquad (9)$$
$$s_3^4 = -z'_3 - B_3^4 \bullet (T'_3 - 2t_3^4)$$

where it is assumed that there are no errors at the outer levels, i.e. $z'_1 = z'_4 = T'_1 = T'_4 = 0$. Only special cases of this system of equations will be considered. The general principles of the development may be illustrated by considering the correction to a single height value or single temperature value. Appendix A considers some more complicated cases.

a. Special Case -- Single Height Error: (Type 1 error)

A hydrostatic error is suspected only when at least one residual in a profile is "large", i.e., its absolute value exceeds the admissible value that has been determined empirically for each layer. Table 1 shows the admissible values in present use. These values are about 7 standard deviations of the residuals when no errors are present. When there are missing data, the admissible residual must be specified over two or more mandatory pressure layers. It is calculated as the square root of the sum of squares of the individual layer admissible residuals.

In the case of a single height error it is assumed that the interior residuals, s_2^3 and s_3^4 (see Fig. 1) are large and

$$z'_2 = T'_2 = T'_3 = 0, \text{ and } z'_3 \text{ is not zero.}$$

Eqns. (11) thus become

$$z'_3 = s_2{}^3 - 2B_2{}^3 \bullet t_2{}^3$$
$$-z'_3 = s_3{}^4 - 2B_2{}^3 \bullet t_3{}^4 \qquad (10)$$

Adding the equations gives

$$(s_2{}^3 + s_3{}^4) = 2(B_2{}^3 \bullet t_2{}^3 + B_3{}^4 \bullet t_3{}^4) \qquad (11)$$

This Eqn. will hold whether there is an error z'_3 or not (but does not hold in the presence of a temperature error). Squaring and averaging this equation over many realizations, assuming that the $t_i{}^{i+1}$'s are independent of each other, gives

$$\overline{(s_2{}^3 + s_3{}^4)^2} = 4\left[(B_2{}^3)^2\overline{(t_2{}^3)^2} + (B_3{}^4)^2\overline{(t_3{}^4)^2}\right] \qquad (12)$$

From Eqn. (6):

$$\overline{(t_i{}^{i+1})^2} = \overline{(s_i{}^{i+1}/2B_i{}^{i+1})^2} \qquad (13)$$

It is found from statistics of the residuals that $\bar{t}_i{}^{i+1} \equiv (\overline{(t_i{}^{i+1})^2})^{1/2}$ is nearly independent of height (equal to about 2 degrees), and will be assigned the constant $\bar{t}$. Eqn. (13) can therefore be written as

$$|s_2{}^3 + s_3{}^4| = 2\left[(B_2{}^3)^2\overline{(t_2{}^3)^2} + (B_3{}^4)^2\overline{(t_3{}^4)^2}\right]^{1/2}$$
$$= 2\,\bar{t}\left[(B_2{}^3)^2 + (B_3{}^4)^2\right]^{1/2} \qquad (14)$$

Eqn. (14) holds even in the presence of a single height error, but for other kinds of errors, the sum on the left-hand-side of Eqn. (14) will have a larger value, so the equation can be used to determine when a single height error is present.

The condition used is:

$$|s_2{}^3 + s_3{}^4| < 2\,\bar{t}_{all}\left[(B_2{}^3)^2 + (B_3{}^4)^2\right]^{1/2} \equiv Z_3{}^* \qquad (15)$$

where $\overline{t}_{all}$ is related to $\overline{t}$, but empirically determined. Presently, a value of 3.5 degrees is used. The magnitude of the error is determined from (12):

$$z'_3 = \tfrac{1}{2}(s_2{}^3 - s_3{}^4 - 2B_2{}^3 \bullet t_2{}^3 + 2B_3{}^4 \bullet t_3{}^4) \approx \tfrac{1}{2}(s_2{}^3 - s_3{}^4) \quad (16)$$

The correction, δz_3, is the negative of the error. It is applied only if it satisfies the magnitude condition:

$$|\delta z_3{}^{(1)}| > Z_3{}^* \quad (17)$$

The superscript (1) on δz_3 is used to signify the correction to an error of type 1: a correction to a single height error. Examples will be given in section 6 of the various error types and the corrections.

b. Special Case -- Single Temperature Error: (Type 2 error)

In this case, the residuals $s_2{}^3$ and $s_3{}^4$ are assumed to be large, $z'_2 = T'_2 = z'_3 = 0$, and T'_3 is not zero. Eqns. (11) become

$$-B_2{}^3 \bullet T'_3 = s_2{}^3 - 2B_2{}^3 \bullet t_2{}^3$$
$$-B_3{}^4 \bullet T'_3 = s_3{}^4 - 2B_3{}^4 \bullet t_3{}^4 \quad (18)$$

It is seen that a single temperature error will cause residuals which, when divided by the layer B's, will be closely equal. Subtracting shows this more clearly:

$$s_2{}^3/B_2{}^3 - s_3{}^4/B_3{}^4 = 2(t_2{}^3 - t_3{}^4) \quad (19)$$

Define the hydrostatic residual in terms of temperature instead of height:

$$X_i{}^{i+1} = s_i{}^{i+1} / B_i{}^{i+1} \quad (20)$$

Eqn. (190) becomes

$$X_2{}^3 - X_3{}^4 = 2(t_2{}^3 - t_3{}^4) \quad (21)$$

This equation will hold whether there is a single temperature error or not. Consider its average for a large ensemble of cases, remembering that the t's are assumed to be independent and have zero mean.

$$\overline{(X_2^3 - X_3^4)^2} = 2\overline{((t_2^3)^2 + (t_3^4)^2)}$$
$$= 4\ \overline{t^2} \qquad (22)$$

Therefore

$$|X_2^3 - X_3^4| = 2\ \overline{t} \qquad (23)$$

It has been assumed that the two layer residuals s_2^3 and s_3^4 are "large". Therefore, there is an error present. However, the errors may be any magnitude and Eqn. (22) is still valid, yet for other types of errors, the left-hand-side of Eqn. (22) will be larger. Therefore, in a particular case, we diagnose a single temperature error when

$$|X_2^3 - X_3^4| < 2\ \overline{t}_{all} \equiv T_3^* \qquad (24)$$

The error is obtained from Eqns. (18):

$$T'_3 = -½(X_2^3 + X_3^4 - 2t_2^3 - 2t_3^4)$$
$$\approx -½(X_2^3 + X_3^4) \qquad (25)$$

The correction is the negative of the error. It is applied only when the correction satisfies the magnitude condition:

$$|\delta T_3^{(2)}| > T_3^* \qquad (26)$$

The superscript (2) on δT_3 is used to signify the correction to an error of type 2: a correction to a single temperature error.

c. Large Residuals in the Bottom or Top Layer (Type 4 and 5 errors)

When large residuals occur in either the top or bottom layers, it is not possible from this information alone to determine the cause of the error. For a large residual in the lowest layer, the cause could be a temperature error at the lowest level, a height error at the lowest level, or an error in the computation of the thickness of the lowest layer, leading to all heights above the lowest being in error. For a large residual in the top layer, the cause could be either an error in the top level temperature or height. An error in the top level height could be either due to a communication error for this height or a computation error for the layer thickness, in which case the height error is equal to the thickness error. For these cases, we suggest height and temperature corrections, <u>either</u> of which would lead to zero residual. For convenience, errors at the bottom are called Type 4 and errors at the top are called Type 5.

e. Multiple Errors (Error Types 3,7,8,9,10)

If a profile of temperatures and heights contains more than one error and the errors are separated by at least one level of correct data, then the correction is no different than for an isolated error; each error is considered separately. If there are two errors at adjacent levels, then equations (10) are appropriate to consider. A derivation is given of the existence and magnitude conditions for two height errors (Type 7), two temperature errors (Type 8), or lower height and upper temperature errors (Type 9) in the Appendix A. If there are two errors at the same level, then sometimes the two passes of the decision-making algorithm will make the necessary corrections (one during each pass), but more usually a pair of corrections that lead to zero residual are only suggested. (Type 3 errors.)

Table 2 shows the corrections, Table 3 the existence conditions, and Table 4 the magnitude conditions for single-layer or double-layer errors. The suggested corrections for cases where confident corrections may not be made are also given in Table 2.

4. The Decision-Making Algorithm

The Hydrostatic Complex Quality Control at NMC represents the first stage in the development of a new Comprehensive Quality Control system. And the Decision-Making Algorithm (DMA) within the CHQC represents the first of several DMA's to be developed. It might be supposed that a general purpose artificial intelligence program could be used to determine the hydrostatic errors and make the corrections. And indeed it might be possible for some of the functions of the DMA to be performed in another way. But we believe that the description of the DMA will make it clear that the artificial intelligence that it contains is very particular to this problem.

Because of the very specialized logic that is necessary for hydrostatic error correction, there will be a detailed description of the DMA. The logic is complicated and yet the required computer time is minimal since only suspected reports are examined. It will be clear that the DMA is conservative in the sense that only confident corrections are actually applied.

a. Overall strategy

The Decision-Making Algorithm (DMA) was designed with the objective of making the maximum number of confident corrections possible. Most of these corrections are simple height or temperature corrections (types 1 and 2). More complicated confident corrections form only about 5 percent of the total. The strategy that was developed begins by consideration of a rawinsonde profile upward from the lowest reported level. A set of three layers is considered at a time. First, confident height or temperature corrections are considered for the upper two of the three layers, and failing to find any, then all three layers are considered for more complicated confident corrections. Errors at the top, bottom and other types of errors are considered along the way. Three layers are considered progressively from the bottom to the top, and the process is repeated a second time, since at times a first-pass correction will allow the algorithm to recognize an additional correction on the second pass. A more detailed description of the procedure follows. The methods used to try to find not only a good correction but, in many cases, the actual correct value, are described.

b. Steps of the Decision-Making Algorithm

This section will describe the steps of the Decision-Making Algorithm. They begin with acquiring the necessary data, continue with the calculation of the hydrostatic residuals and then proceed to the determination of error types and necessary corrections. The process is repeated a second time with slightly altered parameters, allowing additional corrections to be made ocassionally.

1) Get sufficient layers of data

For each pass through the data, the layers are considered from the lowest to the highest, with three layers of data considered at a time. As each layer is completed, it is necessary to determine whether more layers need to be considered.

2) Calculate the hydrostatic residuals

The hydrostatic residuals are calculated according to Eqn. (5) for the three layers.

3) If the top of the three layers is the top layer:

a) Test for "holes". "Holes" are the occurrence of missing layers of data. It is useful to keep information on holes. They come in two types: those that may occur almost anywhere and those with pressures exceeding and including 100 hPa. For the latter type there is often a coding error in Part A of the ravinsonde message which prevents its complete decoding at NMC, but there is no problem with Part C.

b) Test for non-confident corrections at the top. Non-confident or uncertain errors in a layer result in a pair of corrections being suggested. This pair represents those corrections that would lead to zero residuals. In some cases, this pair is the proper correction to make. More usually, there may be additional problems that can only be diagnosed with the help of other checks.

c) Test for error at the top. A large residual in the top layer can be caused by, among other things, an error in the top level temperature or height, it being impossible to decide the cause by the hydrostatic check alone. In these cases, a choice of corrections is suggested to be examined by an analyst. Either suggested correction by itself would lead to zero residual for the layer.

d) Proceed to step 6.

4) Otherwise: Calculate non-dimensional ratios to be used to determine the most probable error type.

The ratios of the right-hand-sides to the left-hand-sides of the existence conditions (existence condition ratios) are calculated. The most probable error type is considered to be the type which has the largest ratio, while satisfying the magnitude condition. In some rare cases, the residuals will be consistent with both types 7 and 8 or types 9 and 10. In these cases, which occur when the central residual is small, it can be shown that a distinction of types cannot be made by the hydrostatic check alone, and no correction is made. The reasons for this ambiguity will be discussed further in Appendix B.

During the first pass through the data, there is a small preference given to confident single corrections to height or temperature, compared to multiple corrections. This helps to prevent some small changes to the data which do not seem warranted.

When a residual pair is very large, it is clear that there is an error, and in many situations it is clear what correction is needed. In some of these cases, the conditions for a confident correction would nevertheless not be quite satisfied. For this reason, the value of the existence condition ratio is inflated for large residuals.

5) Determine the most probable error type

As stated above, generally the error type is determined to be the type associated with the maximum existence condition ratio among the types satisfying the magnitude condition. In some more detail, these conditions are summarized below:

[1 - confident height correction, and
2 - confident temperature correction:]

a) maximum existence condition ratio

b) magnitude condition satisfied

c) no more than one adjacent layer of information missing

[7 - two confident height corrections,
8 - two confident temperature corrections,
9 - lower height and upper temperature corrections, and
10 - lower temperature and upper height corrections:]

a) three layers do not include top or bottom layers

b) no more than one missing level between any levels

c) maximum existence condition ratio

d) magnitude condition satisfied

13 - Part A hole, missing data at pressure(s) including 100 hPa and possibly greater pressures, but not 70 hPa:at least one missing level, 100 hPa and at possibly greater pressures

14 - general hole, missing at any level, but not type 13: does not satisfy conditions for types 1,2,7-10 or 13 at least one missing level

4 - error at bottom:

$|s_2^3| >$ admissible and $|s_3^4| <$ ½ admissible, or

$|s_2^3| >$ admissible and $|s_3^4/s_2^3| < 1/3$

does not satisfy another type

3 - correction pair suggested:

$|s_2^3| >$ admissible and $|s_3^4| >$ ½ admissible, or

$|s_3^4| >$ admissible and $|s_2^3| >$ ½ admissible

does not satisfy another type

6 - isolated large residual:

$|s_2^3| >$ admissible

not bottom or top layer

adjacent residuals small enough

6) Make corrections of the appropriate type

a) Confident height corrections (as per Table 2). The starting point for the corrections is the value given in the table, but this is modified as described below to attempt to find the likely correct value. The original height corrections are first rounded to the nearest 10 m for mandatory pressure levels of 500 hPa and lower pressures and rounded to the nearest meter for greater mandatory level pressures. Then a correction is sought within 15 to 20 meters of this value which is consistent with the idea that either a single digit is in error or the error results from a permutation of digits. If such a "simple" error is found, it is accepted; otherwise the provisional value is accepted. The magnitude of the correction is at least equal to the limiting value that satisfies the magnitude condition for type 1. The value varies from 26 to 99meters as seen in Table 4.

b) Confident temperature corrections (as per Table 2). As with heights, the starting point for the corrections is the value given in the table. Temperature corrections are rounded to the nearest 1/10 degree. For temperatures also "simple" corrections are sought, which include: a sign correction, and/or a single digit correction or permutation of digits correction. Checks are made to ensure that no temperature correction leads either to large dry static instability in the layers above or below the correction nor to excessive curvature among the temperatures in the layer and adjacent layers. The magnitude of the temperature correction is at least equal to the limiting value satisfying the magnitude condition for type 2; it is set to 7.0 deg K.

c) Suggested correction pair. The pair of corrections that leads to zero residuals is given in Table 2. The range is also depicted in Fig. 2.

d) Suggested corrections at the top or bottom. In this case, one of the suggested corrections may be appropriate, <u>but not both</u>. Either correction would lead to zero residual. The values of the suggested corrections are given in Table 2.

7) Repeat the pass through the data for a second time.

As explained earlier, the second pass will sometimes allow corrections for layers that were partially corrected during the first pass.

5. Baseline Check

The baseline check is a simple test which checks the consistency of the reported surface pressure and (supposedly) known station elevation with the 1000 and 850 hPa reported heights. The procedure uses the two mandatory level heights to define the mean virtual temperature for the 1000-850 hPa layer. A standard lapse rate (-6.5 deg/km) is assumed to apply downward to the reported surface pressure. These assumptions are used to solve for the station elevation, which is compared with the station elevation in NMC's station dictionary. When there is a large dicrepency, there are several possible sources. If the heights are accurate, then either the station elevation or the surface pressure must be in error. We find cases of both kinds. When the station elevation is wrong in the dictionary, this results in a permanent baseline check error. On the other hand, errors in the surface pressure or computation of 1000 mb height lead to sporadic errors.

When the station elevation and surface pressure are accurate, then the baseline check can help to corroborate the inference from the CHQC. An example will be shown in the next section.

The value of the station height which is consistent with the 1000 and 850 hPa heights and the reported surface pressure is calculated from the following equations.

$$z_s^C = z_{1000} + [(\alpha-1)/b] * [T_{lay} + b(z_{1000} - z_{lay})]$$

where

$$\alpha = (p_s/p_{1000})^{-Rb/g}$$

$$T_{lay} = -\tfrac{1}{2}b(z_{850}-z_{1000})\left[\frac{1+\alpha_{lay}}{1-\alpha_{lay}}\right]$$

$$\alpha_{lay} = (850/1000)^{-Rb/g}$$

$$z_{lay} = \tfrac{1}{2}(z_{1000}+z_{850})$$

$$b = -.0065$$

R is dry air gas constant and g is the acceleration of gravity.

We routinely record all baseline check errors of 30 m and greater. The results are operationally available to MOD and are summarized periodically for determination of station elevation and other consistent problems.

6. Examples of hydrostatic corrections

This section will show examples of several errors and corrections. Each example will be discussed in detail as they illustrate the sources of error, the logic necessary to determine the error, the determination of simple corrections and when simple corrections are not appropriate, and some of the difficulties that remain in making corrections. Some cases are especially complicated; some can be corrected automatically, while others cannot. There is a well-stocked "zoo" of "animals" from which to select these complicated examples. Some errors are particularly due to a lack of qualification of the person(s) involved with the data--totally unreasonable data which is nevertheless accepted, multiple errors, etc. It is noted that the likelihood of the occurrence of an error is not statistically independent of the presence of any other errors for the same profile. Therefore, the probability of the presence of two errors in a profile is distinctly larger than the square of the probability of one error. An example will be shown that illustrates the value of two passes.

Statistics have been collected since January 1989 on the performance of the CHQC. The examples which follow in this section are designed to show the kinds of problems which occur within individual rawinsonde profiles. They are of interest when considered individually since they illustrate the great variety of difficulties that can occur, but these features may not be the most important when considering the overall performance of the CHQC.

The examples will clearly show the need for a flexible and powerful Decision-Making Algorithm, and will show the scope of the present algorithm as well as some of its limitations. The examples will be grouped as: a) confident, simple corrections, b) errors where other checks are needed for a decision, c) complicated corrections, d) special cases.

The information shown in the examples is: the pressure (hPa), height (m), temperature (deg-C), layer residual (m), residual after correction (m), height correction (m) (if any), temperature correction (deg) (if any), and error type. Any digits in error and corrected values are indicated by bold numbers. And any values that are calculated manually are surrounded in parentheses.

a. Confident, Simple Corrections

1) Confident height correction, single digit correction Example 1 shows the pattern of residuals which signify a single height error. The layer residuals are large, with nearly the same value and opposite signs. In this case the error is a simple one: a single digit needs correction. A provisional correction of -86.4 m was made. This value was rounded to the nearest 10 m, giving -90 m, and then a simple correction was sought near this value, finding a single-digit correction at 100 m. Note that after the correction, the layer residuals are all small.

Example 1: Type 1 error, single digit correction

89/09/27/12 08594

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			7.4	7.4			
150	14250	-69.1					0
	14250	-69.1					
			84.6	-15.4			
100	16720	-75.5			-100	0.0	1
	16620	-75.5					
			-88.2	11.8			
70	18750	-65.3					0
	18750	-65.3					
			-8.9	-8.9			

2) Confident height correction, transposition of digits correction.

Example 2 also shows a pattern of two large residuals of about the same value and opposite sign, indicating a height correction. The suggested correction is 2680.7 m, which is rounded to 2680 m. Adding this value to the original height gives 14120 m, which is not a simple correction. However, a correction of 2700 m leads to a corrected height of 14140 m, which results from a transposition of digits. This solution is found by explicitly looking in a range, beginning from the proposed correction and moving outward, for a correction that is first a single digit, or second a transposition of digits. The correction is acceptable as can be seen from the fact that the new residuals are reduced to values smaller than the admissible values (see Table 1).

Example 2: Confident height correction, transposition of digits correction

89/09/26/00 47158

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			0.2	0.2			
200	12340	-55.7					0
	12340	-55.7					
			-2701.1	-1.1			
150	11440	-63.1			2700	0.0	1
	14140	-63.1					
			2660.3	-39.7			
100	16570	-67.3					0
	16570	-67.3					
			-6.8	-6.8			

3) Confident height correction, general correction

A more general type of correction is illustrated by Example 3. In this case there is a kind of shift of digits: 5300 should be 5530. The correction is made as the average of the difference of the residuals, rounded to the nearest 10 m. The resulting residuals are small.

Example 3: Type 1 error, general - type correction

89/07/27/00 23933

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			16.2	16.2			
700	2910	-2.5					0
	2910	-2.5					
			-214.3	15.7			
500	5300	-15.3			230	0.0	1
	5530	-15.3					
			239.1	9.1			
400	7180	-28.9					0
	7180	-28.9					
			-10.2	-10.2			

4) Confident temperature correction, single digit correction

Cursory examination shows that the 200 hPa temperature in Example 4 is in error. The suggested correction is -19.2 deg. A simple correction is sought in the vicinity of this value, finding one at -20.0 deg. The corrected value of the temperature is thus -48.9 C.

Example 4: Type 2 error, single digit correction

89/07/26/12 62053

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			-4.5	-4.5			
250	10980	-38.1					0
	10980	-38.1					
			-66.4	-1.3			
200	12480	-28.9			0	-20.0	2
	12480	-48.9					
			-76.6	7.3			
150	14320	-62.5					0
	14320	-62.5					
			-12.9	-12.9			

5) Confident temperature correction, sign correction

The single most frequent kind of temperature error is a sign error. Example 5 illustrates a correction of temperature sign.

The new residuals are small, confirming the correction. It is noted that we have not insisted that the details of the meteorological code be followed--there is no reason why we should--, as can be seen by the fact that the resulting value is both negative and has even tenths of a degree. There is no basis upon which to choose between a value of -13.9 and -14.1.

Example 5: Type 2 error, sign correction

89/07/26/12 46747

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			17.0	17.0			
500	5840	-3.7					0
	5840	-3.7					
			-99.0	-7.5			
400	7560	14.0			0	-28.0	2
	7560	-14.0					
			-112.3	5.7			
300	9690	-28.1					0
	9690	-28.1					
			-3.5	-3.5			

6) Confident temperature correction, transposition of digits

The two large negative residuals in Example 6 indicate the need for a temperature correction. The suggested correction is -27.7 deg. Since there is no single-digit correction of smaller difference from the suggested value, the correction of -27.0 deg, which gives a transposition of digits, is accepted. This value is both aesthetically good and results in small residuals.

Example 6: Type 2 error, transposition correction

89/07/24/12 32389

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			5.1	5.1			
70	19000	-54.3					0
	19000	-54.3					
			-138.8	-5.8			
50	21160	-25.5			0	-27.0	2
	21160	-52.5					
			-204.4	-2.4			
30	24480	-49.7					0
	24480	-49.7					
			-2.1	-2.1			

7) Confident temperature correction, correction of sign plus one digit

Example 7 illustrates another type of error that can be successfully corrected by the CHQC. The suggested correction is -84.7 deg. The closest simple correction gives both a sign and single digit correction. As before, the tenths of degree is not adjusted to agree with the coding practice, this being a conscious decision.

Example 7: Type 2 error, sign plus one digit correction

89/07/26/12 94527

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			-6.8	-6.8			
700	3135	1.4					0
	3135	1.4					
			-428.0	-6.2			
500	5740	67.8			0	-85.6	2
	5740	-17.8					
			-269.8	9.9			
400	7380	-29.7					0
	7380	-29.7					
			-3.4	-3.4			

b. Errors for Which Additional Checks are Needed for Correction

1) Error of Type 3--pair of large residuals, but conditions not satisfying confident height or temperature corrections

In Example 8, there is a pair of large residuals; however, they do not satisfy the existence and magnitude conditions for confident corrections to either height or temperature. Therefore, a pair of corrections is suggested which when taken together would lead to zero residuals. The suggested corrections are -1000 m and -29.2 deg. A height correction of -1000 m is most likely correct. A temperature sign correction leads to a temperature correction of -26.8 deg, which looks correct. Therefore, in this case but not generally, the proposed correction pair appears to be close to good corrections. Since there can be reasons other than a pair of incorrect values at the same level for a pair of large residuals, it is necessary to combine the result of this check with other checks to make a final decision.

Example 8: Type 3 error

89/02/11/12 87047

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			12.9	12.9			
500	5880	-4.9					0
	5880	-4.9					
			906.9	(-5.5)			
400	8600	13.4			-1000	-29.2	3
	(7600	-13.4)			(-1000	-26.8)	
			-1125.4	(-12.5)			
300	9700	-31.5					0
	9700	-31.5					
			-1.7	-1.7			

2) Large residual in the lowest layer--height correction needed

When there is a large residual in the lowest layer, it is impossible to tell whether the lowest level height is in error, the lowest level temperature is in error, or there is an error in the computation of the lowest layer thickness, leading to all the heights above the lowest being in error. It is essential to combine the CHQC with other checks to definitely determine the nature of the error and hopefully correct it. However, by use of a small amount of additional information it is sometimes possible to make a determination. For that reason, these errors are submitted to NMC's Meteorological Operations Division (MOD) personnel for examination. Simple reasoning can be used to determine the likely error in Example 9. It is easily seen that a temperature correction of -38.4 would be unreasonable. Therefore, the likely correction is a height correction in the vicinity of -91.5 m by default. A correction of -100 m is suggested.

Example 9: Type 4 error, height correction needed

89/02/11/00 46747

p	z	T	RES	NRES	ZCOR	TCOR	TYP
1000	180	15.6			-91.5	-38.4	4
	(80)	15.6			(-100)		
			-91.5	(8.5)			
850	1444	7.6					0
	1444	7.6					
			-6.0	-6.0			

3) Large residual in the lowest layer, temperature correction needed

Example 10 is similar to Example 9 except that in this case a correction to the temperature appears necessary. The suggested correction is reasonable, but closer examination shows that a value within a few degrees would give a kind of shift of digits correction and may be the proper correction. This particular kind of correction is not proposed automatically, but illustrates the rich variety of errors that are made.

Example 10: Type 4 error, temperature to be corrected

89/07/26/12 50527

p	z	T	RES	NRES	ZCOR	TCOR	TYP
850	1421	36.0			-69	-24.3	4
	1421	(13.6)				(-22.4)	
			-69.1	-69.1			
				(-5.4)			
700	3000	-2.7					0
	3000	-2.7					
			7.5	7.5			
500	5600	-17.5					0
	5600	-17.5					

4) Large residual in lowest layer, difficult to decide

In Example 11, there is not enough information to decide what the error is, the magnitude of the error, or actually whether there definitely is an error. First, the residual only slightly exceeds the allowable value. And second, either suggested correction is not unreasonable.

Example 11: Type 4 error, difficult to decide

89/07/24/12 35394

p	z	T	RES	NRES	ZCOR	TCOR	TYP
850	1583	18.8			-37	-13.1	4
	1583	18.8					
			-37.3	-37.3			
700	3173	7.2					0
	3173	7.2					
			-3.6	-3.6			

5) Large residual in the highest layer, height to be corrected

Example 12 shows an error leading to a large residual in the highest layer. The proposed temperature correction of -305.0 deg is ridiculous, so the required correction, by negative reasoning, is a height correction of approximately -996.6 m. It is very likely that the exact correction that is required is -1000 m. As for errors at the lowest level, it is not possible to determine what happened without other checks (in this case a crude climatological or gross check). Errors at the highest level are given to NMC's MOD for examination.

Example 12: Type 5 error, height to be corrected

89/07/24/12 59981

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			1.9	1.9			
250	11010	-40.5					0
	11010	-40.5					
			-996.6	(3.4)			
200	11500	-50.9			1000	-305.0	5
	(12500)	-50.9					

6) Large residual in the highest layer, temperature correction needed

There is a large residual in the 300-250 hPa layer for Example 13 and the 250 hPa temperature is clearly in error. The proposed correction of -151.1 deg is close to the sign plus one digit correction of -152.0 deg, which changes the temperature from 86.0 to -66.0. This value would likely be accepted by an analyst with additional information, or by a complex quality control using vertical and horizontal statistical interpolation checks.

Example 13: Type 5 error, temperature correction needed

89/07/31/00 68906

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			-0.3	-0.3			
300	9120	-50.5					0
	9120	-50.5					
			-403.5	(2.4)			
250	10270	86.0			400	-151.1	5
	10270	(-66.0)				(-152.0)	

7) Large residual in the highest layer, both height and temperature need correction

Example 14 also shows a case with a large residual in the uppermost layer. That both the height and temperature are in error may be seen as follows. First try to apply the suggested temperature correction. The revised temperature would be -87.1, a value that is too low, and yet the original value is too high. Therefore, a part of the residual only is due to the error in temperature, showing that the height must also be wrong. No correction can be made without a complete set of checks and DMA.

Example 14: Type 5 error, both height and temperature need correction

89/07/30/12 43333

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			9.4	9.4			
150	14410	-63.5					0
	14410	-63.5					
			-320.4	-320.4			
100	16760	-33.1			320	-54.0	5
	16760	-33.1					

8) Isolated large residual (thickness computation error)

When there is a single large residual it is risky to decide upon a correction based upon the hydrostatic check alone. The most likely cause is an error in the computation of the thickness of the layer, but it may be that there is a high linearity of the temperature within the layer leading to a relatively large residual. Other possibilities have been encountered. In a few cases, there has been a compensating large residual several layers removed, so that only the intervening heights needed correction. The present computation and coding practice at some stations also leads to a possible cause of an isolated large residual. At some stations the Parts A and C of the report are computed independently. The use of an incorrect temperature for Part A, which is discovered before Part C is computed will then lead to an isolated residual in the 100-70 hPa layer. In Example 15 the most usual cause of an isolated large residual--a thickness computation error--has occurred. All heights above the layer need to be corrected by 100 m.

Example 15: Type 6 error

89/07/27/00 43371

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			7.8	7.8			
700	3101	7.2					0
	3101	7.2					
			-103.4	(-3.4)			
500	5700	-5.1			0	0.0	6
	(5800)	-5.1			(100)		
			10.7	10.7			
400	7430	-15.1					0
	(7530)	-15.1			(100)		
			14.7	14.7			
300	9550	-31.5					0
	(9650)	-31.5			(100)		
	....						
	....						

c. Complicated Corrections

1) Height corrections to adjacent levels

Example 16 shows a type 7 correction, a correction to heights at two adjacent layers. This case may be identified by the fact that the sum of the three layer residuals is small (=-10.9). A simple correction is found for the 850 hPa height, but not for the 700 hPa height.

Example 16: Type 7 error, one digit correction plus general-type correction

89/07/24/12 02365

p	z	T	RES	NRES	ZCOR	TCOR	TYP
1000	225	21.6					0
	225	21.6					
			56.7	-3.3			
850	1671	11.6			-70	0.0	7
	1601	11.6					
			-460.1	2.9			
700	2799	0.6			393	0.0	7
	3192	0.6					
			392.5	-0.5			
500	5820	-13.5					0
	5820	-13.5					
			7.6	7.6			

2) Temperature corrections to two adjacent levels

Example 17 shows corrections to two adjacent levels. The existence and magnitude conditions are satisfied for a type 8 error and so the correction is automatically performed. A sign correction is determined to be the simple correction at 400 hPa and a single digit correction is made at 300 hPa. The smallness of the new residuals confirms the corrections.

Example 17: Type 8 error, sign correction plus one digit correction

89/07/26/12 55591

p	z	T	RES	NRES	ZCOR	TCOR	TYP
500	5840	-0.7					0
	5840	-0.7					
			-53.1	7.0			
400	7600	9.2			0	-18.4	8
	7600	-9.2					
			135.0	1.9			
300	9770	-72.5			0	50.0	8
	9770	-22.5					
			130.6	-2.9			
250	11080	-32.1					0
	11080	-32.1					
			2.4	2.4			

3) Corrections to adjacent layers, one height and one temperature

Examples 18 and 19 show corrections to two adjacent levels: one temperature and one height. These corrections were determined automatically, and incidentally both involve a sign correction for temperature and a single-digit correction for height. All residuals become small after the corrections are made.

Example 18: Type 9 error, one-digit height correction plus temperature sign correction

89/07/28/00 36259

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			0.7	0.7			
250	10770	-36.1					0
	10770	-36.1					
			-203.2	-3.2			
200	12090	-44.1			200	0.0	19
	12290	-44.1					
			-266.8	-5.0			
150	14170	54.8			0	-109.6	29
	14170	-54.8					
			-659.0	-8.2			
100	16720	-60.7					0
	16720	-60.7					

Example 19: Type 10 error, temperature-sign correction plus one-digit height correction

89/07/23/00 46747

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			2.7	2.7			
200	12570	-53.1					0
	12570	-53.1					
			-572.8	-3.1			
150	14300	67.6			0	-135.2	20
	14300	-67.6					
			-1105.4	-2.6			
100	16460	-74.1			300	0.0	10
	16760	-74.1					
			292.3	-7.7			
70	18860	-68.7					0
	18860	-68.7					
			13.6	13.6			

d. Cases of Special Interest

There is a wide variety of errors that occasionally occur, illustrating particular points about characteristic human errors, observation code deficiencies, or the abilities and limitations of the CHQC. Some of these points will be illustrated by use of the following examples.

1) Cases where a simple correction is not appropriate

All modifications to the data look for a simple correction, i.e. a sign correction to temperature, a single-digit correction or a transposition correction. Simple corrections are most likely proper when the error has resulted from human intervention in the data in decimal form, but not in all cases. In any case where the error has occurred to the data generated automatically and in particular for data in other than decimal form, a simple correction is not appropriate. The following two examples show cases when a simple correction is not appropriate (but yet a simple correction is better than no correction).

a) Data repeated from one layer to another

In Example 20 the 700 hPa temperature was copied to 500 hPa. The CHQC correctly identifies the error and gives a preliminary correction of -14.1 deg, which should be accepted. However, a sign error is considered more likely by the code since the new residuals would be acceptable and the code looks first for an acceptable sign correction. The decision in this case is not proper, but, with the present algorithm, can only be seen by manual inspection. It would be possible to explicitly look for this particular kind of error, but it occurs so rarely that it does not seem efficient or profitable to build into the code all such special cases. The CHQC arrives at an acceptable solution, even if it is not the best one.

Example 20: Type 2 repetition error, no simple correction needed

89/08/21/00 35746

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			0.6	0.6			
700	3130	4.8					0
	3130	4.8					
			-69.3	-22.6			
				(0.1)			
500	5800	4.8			0	-9.6	2
	5800	-4.8					
		(-9.3)				(-14.1)	
			-46.8	-15.5			
				(-0.7)			
400	7480	-22.7					0
	7480	-22.7					
			-2.1	-2.1			

b) Errors to automatically generated (binary) data

There is a peculiarity of the practice in dealing with mandatory levels, which continues from days when rawinsondes did not regularly reach high levels, that has led to hydrostatic error for many Canadian stations. When a significant level is within ±0.5 hPa of a mandatory level, then the information is reported as if it were at the mandatory level. This can lead to errors of several hundred meters at the uppermost levels. The reporting process is entirely automatic, so that the error is certainly not a simple one. Yet, as seen in Example 21, a simple correction was proposed. It is more proper to accept the provisional correction of 124 m.

This is just one example of an error caused by lack of foresight into the possible causes for error. Example 25 will show another instance where the code itself was designed without the necessary foresight. And it can indeed be said that the design of NMC's decoding algorithm itself should have foreseen the possibility of some of the problems that were largely unknown until the CHQC was run. (Probably the CHQC itself has missed some important points, but that will not be known until the future, of course.)

Example 21: "Canadian 1" error, no simple correction needed

89/07/30/00 71909

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			-3.8	-3.8			
50	20990	-45.9					0
	20990	-45.9					
			128.3	28.3			
30	24520	-45.7			-100	0.0	1
	24420	-45.7					
			-119.1	-19.1			
20	27120	-42.7					0
	27120	-42.7					
			15.8	15.8			

2) Example to illustrate the need for two scans

In the following Example 22 there are two successive corrections at the same mandatory pressure level. This is made possible by the fact that the residuals are so large. When they are large, the requirement for existence for error types 1 and 2 is somewhat relaxed to allow confident corrections. In the second scan then, for this case, the temperature correction was correctly discovered. Without relaxing the conditions for a type 1 correction on the first scan, the CHQC would have found a type 3 error. Note that there is a successive reduction in both residuals following each scan.

Example 22: Errors in both height and temperature at the same level, successfully dealt with by first correcting height (Type 1) and then temperature (Type 2), instead merely diagnosing the Type 3 error.

89/08/17/00 44259

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			1.9	1.9			
500	5690	-21.1					0
	5690	-21.1					
			-1033.0	-33.0			
			-33.0	-0.4			
400	6290	-26.6			1000	0.0	1
	7290	-26.6			0	-10.0	2
	7290	-36.6					
			957.7	-42.3			
			-42.3	-0.2			
300	9230	-50.3					0
	9230	-50.3					

3) Complicated errors

An example will be shown that illustrates that rather complex error structures can at times be successfully handled by the CHQC. In Example 23 there are 7 large residuals. After corrections are made, 8 residuals are made smaller, all now less than the admissible layer values. Beginning from the lowest levels, there is a height error 200 m at 400 hPa. This is diagnosed from the pair of large residuals surrounding this level. The next corrections needed are a pair of sign corrections to the temperatures at 250 and 200 hPa. A triplet of large residuals is examined to make these corrections (type 8). Above these a pair of height corrections is needed at 70 and 50 hPa, one general correction and one single-digit correction. Again, a triplet of large residuals is used to determine the necessary corrections (type 7). And finally, there is missing data at 100 hPa but not at 70 hPa. This situation is diagnosed as a "hole"--a type 13 error.

This case is remarkable in that so many problems occurred simultaneously, but the fact that the CHQC found good corrections is not. This is because all the errors were "simple" and because they were sufficiently separated in the vertical for the analysis to work.

Example 23: Multiple corrections: what the present, improved, Decision Making Algorithm can do

89/09/05/00   46747

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			15.9	15.9			
500	5850	-4.3					0
	5850	-4.3					
			207.4	7.4			
400	7780	-14.9			-200		1
	7580	-14.9					
			-202.1	-2.1			
300	9690	-30.1					0
	9690	-30.1					
			-204.1	7.4			
250	10970	39.6				-79.2	8
	10970	-39.6					
			-609.9	-0.7			
200	12450	53.6				-107.2	8
	12450	-53.6					
			-471.5	-19.9			
150	14230	-65.5					0
	14230	-65.5					
			-131.8	-1.8			
100	99999	9999.9					
	99999	9999.9					
70	18660	-72.1			130		7
	18790	-72.1					
			43.2	13.2			
50	20660	-77.1			100		7
	20760	-77.1					13
			109.6	9.6			
30	23890	-52.1					0
30	23890	-52.1					
			-54.8	-54.8			

4) Use of the baseline to correct height

The baseline check determines the discrepancy between the station elevation at the reported station pressure as determined from the 1000 and 850 hPa heights and the assumption of a standard lapse rate, and the known station elevation. A history is kept of those stations whose discrepancy is greater than 30 m. Some stations have permanent baseline discrepancies, while others are sporadic. The baseline check has found its greatest utility in determining those stations for which NMC might have an inconsistent station elevation in its dictionary, corresponding to the stations with permanent baseline discrepancies. For the remaining cases, examination has shown that most are likely caused by a communication error in the station pressure or computation error in the 1000 mb height since the hydrostatic check does not confirm another reason. However, there are cases in which the baseline error does confirm the results from the hydrostatic check and could be used by our code if we chose to be a little less cautious, or by an analyst to correct the 1000 hPa height or temperature. Example 24 shows such a case. The baseline error is -125 m and the lowest layer residual is 119.3 m. A correction of about 120 m seems to be needed to greatly reduce both errors. Perhaps the exact correction is 117 m since that value would result from the 9 having been dropped in the 1000 hPa height of 129 m.

Example 24: Type 4 error, application of baseline check results to correct height

89/07/30/12 23552

p	z	T	RES	NRES	ZCOR	TCOR	TYP
	Baseline error: -125						
1000	12	17.6			119	50.1	4
	(129)	17.6			(117)		
			119.3	(2.3)			
850	1483	4.0					0
	1483	4.0					
			8.5	8.5			

5) Correction needed because of present WMO observation code deficiency

Example 25 shows an error of 1000 m at 700 hPa for an Antarctic station. The temperatures are low and the 700 hPa height has a value which is coded as 350, dropping the leading 2. Present code convention demands that upon decoding the report, the leading digit should be a 3, resulting in the error. The CHQC makes the necessary correction.

Example 25: "Antarctic type" Type 1 error

89/07/23/00 89592

p	z	T	RES	NRES	ZCOR	TCOR	TYP
850	906	-14.3					0
	906	-14.3					
			998.7	-1.3			
700	3350	-23.7			1000	0.0	1
	2350	-23.7					
			-989.9	10.1			
500	4720	-43.7					0
	4720	-43.7					
			-3.6	-3.6			

6) Errors too complicated to be corrected by the CHQC alone

In the Example 26 there are two large negative residuals which are consistent with a negative temperature correction at 100 hPa. However, such a correction would violate dry static stability and must be rejected (as the CHQC does). From the information available it is not clear what has actually happened. There are a small number of cases each observation time that appear to have a good correction, but it must be rejected on internal evidence. Other components of the complex quality control are needed to make these situations more clear.

Example 26: Impossible to understand what has happened

89/08/02/12 61641

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			-9.5	-9.5			
150	14240	-65.7					0
	14240	-65.7					
			-189.6	-189.6			
100	16450	-76.5			0	-32.8	12
	16450	-76.5					
			-175.3	-175.3			
70	18420	-59.1					0
	18420	-59.1					

Another situation is shown in Example 27. There are four large residuals in a row, but neither the upper nor lower pair by itself leads to a correction. Nor do three of the residuals taken at a time lead to a correction. But notice that the sum of the 1st and 3rd is nearly the negative of the sum of the 2nd and 4th. Therefore, height corrections are necessary, which may be worked out as indicated by values in parentheses. All new residuals would then become small.

This case is not handled by the present CHQC because we have chosen not to look for more than 2 errors in a row. It would probably be possible to extend the algorithm to include successively more layers, but the rarity of these cases, and the statistical nature of the existence and magnitude conditions would make the possibility for correction severely limited.

Example 27: Height error at three consecutive levels, not correctable by CHQC

89/02/03/12 83971

p	z	T	RES	NRES	ZCOR	TCOR	TYP
			7.5	7.5			
700	3120	6.2					0
	3120	6.2					
			2205.2	(5.2)			
500	8000	-9.7			-3760	-314.7	3
	(5800)	-9.7			(-2200)		
			-4784.7	(5.3)			
400	4900	-21.1			6470	514.5	3
	(7490)	-21.1			(2590)		
			8633.4	(3.5)			
300	15600	-34.7			-7050	376.7	3
	(9560)	-34.7			(-6040)		
			-6040.4	(-0.4)			
250	10810	-43.3					0
	10810	-43.3					
			4.9	4.9			

The final Example 28 shows a hopeless case (from the point of view of the CHQC). Similar cases are not as rare as they should be. In this example, there are 5 sign errors and 2 height errors! The CHQC suggests type 4 corrections at 80 hPa and pairs of type 3 corrections at other levels, all of which are bad. Human intervention could make the required corrections, but it is questionable whether the time taken would be well-spent.

Example 28: Too many corrections needed: even the present, improved, Decision Making Algorithm is not capable of making them (difficulties caused by the existing coding system of rawinsonde data are demonstrated by this example as well).

89/09/01/12 62721

p	z	T	RES	NRES	ZCOR	TCOR	TYP
1000	13	9999.9					0
	13	9999.9					
850	1483	-26.5			252	88.6	4
	1483	(26.5)				(53.0)	(2)
			252.0	(12.9)			
700	3169	-15.5					0
	3169	(15.5)				(31.0)	(2)
			72.0	(5.4)			
500	5900	8.8			-870	-162.2	3
	5900	(-8.8)				(17.6)	(2)
			-1401.5	(26.1)			
400	6260	-16.1			1820	127.9	3
	(7620)	-16.1			(1360)		(1)
			2358.2	(4.1)			
300	10970	28.0			-1870	114.9	3
	(9750)	(-28.0)			(-1220)	(-56.0)	(3)
			-1567.6	(10.2)			
250	11040	39.0			750	-306.8	3
	11040	(-39.0)				(-78.0)	(2)
			-254.5	(0.4)			
200	12530	-51.5					0
	12530	-51.5					

7. Summary

As the first stage of the design of the new data quality control system at the National Meteorological Center, the Comprehensive Hydrostatic Quality Control of rawinsonde data on heights and temperatures on mandatory isobaric surfaces has been designed, tested and implemented operationally. The paper contains a detailed description of the leading principles of the new approach, as well as of the CHQC algorithms themselves, with particular emphasis on the Decision Making Algorithm. The DMA is a completely new algorithm which not only diagnoses the error causes, but also computes their values and, subsequently, introduces confident corrections of erroneous data whenever this is possible, based only on the hydrostatic checks for several layers. For other cases, the DMA proposes several possible corrections for further analysis by specialists.

The performance of the CHQC, and particularly of its DMA, is illustrated by numerous examples, taken from the automatically produced operational outputs reporting the CHQC actions. The examples also give some information on the present status of the data quality.

It would be hardly possible to substantially improve the CHQC version now in operational use at NMC. Further progress may be achieved only after some other, statistical checks have been developed and added to the hydrostatic one. Work in this direction is underway.

Acknowledgments. We wish to extend our thanks to our many collegues at the National Meteorological Center and elsewhere, and especially to Dr. William Bonner, Dr. Eugenia Kalnay and Dr. Ronald McPherson for their encouragement, support, discussion and critical observations during the development of this work. We wish to thank two forecasters in NMC's Meteorological Operations Division, Valerie Thompson and William Whitmore, especially for their eagerness to use diagnosis provided by the CHQC in their work and provide important feedback to the CHQC development.

APPENDIX A

EXISTENCE AND MAGNITUDE CONDITIONS FOR COMPLEX ERROR TYPES

Two Height Corrections (Type 7)

In the case of one height error at two adjacent levels we have

$$t'_2 = t'_3 = 0, \text{ and } z'_2, z'_3 \text{ are not zero,}$$

while at least $s_1{}^2$ and/or $s_3{}^4$ is large.

$$\begin{aligned} z'_2 &= s_1{}^2 - 2B_1{}^2 \bullet t_1{}^2 \\ -z'_2 + z'_3 &= s_2{}^3 - 2B_2{}^3 \bullet t_2{}^3 \\ -z'_3 &= s_3{}^4 - 2B_3{}^4 \bullet t_3{}^4 \end{aligned} \qquad (A1)$$

Adding the equations gives

$$s_1{}^2 + s_2{}^3 + s_3{}^4 = 2(B_1{}^2 \bullet t_1{}^2 + B_2{}^3 \bullet t_2{}^3 + B_3{}^4 \bullet t_3{}^4) \qquad (A2)$$

This equation, which does not have any terms involving the height errors, nevertheless is valid in their presence. Squaring the equation and averaging over many realizations and assuming the $t_i{}^{i+1}$ to be independent of each other gives

$$\begin{aligned} \overline{(s_1{}^2+s_2{}^3+s_3{}^4)^2} &= 4[(B_1{}^2)^2\overline{(t_1{}^2)^2} + (B_2{}^3)^2\overline{(t_2{}^3)^2} + (B_3{}^4)^2\overline{(t_3{}^4)^2}] \\ &= 4[(B_1{}^2)^2 + (B_2{}^3)^2 + (B_3{}^4)^2]\,(\overline{t})^2 \end{aligned} \qquad (A3)$$

Therefore

$$\overline{|s_1{}^2+s_2{}^3+s_3{}^4|} = 2[(B_1{}^2)^2 + (B_2{}^3)^2 + (B_3{}^4)^2]^{1/2}\,\overline{t} \qquad (A4)$$

Two height errors are diagnosed when

$$\overline{(s_1{}^2+s_2{}^3+s_3{}^4)} < 2[(B_1{}^2)^2 + (B_2{}^3)^2 + (B_3{}^4)^2]^{1/2}\,\overline{t}_{all} \qquad (A5)$$

where $\overline{t}_{all}$ has the same value used for other error types. The magnitude of the errors is determined from Eq. (A1), i.e.

$$z'_2 = s_1^2 - 2B_1^2 \cdot t_1^2 \approx s_1^2$$
$$z'_3 = -s_3^4 + 2B_3^4 \cdot t_3^4 \approx -s_3^4 \quad \text{(A6)}$$

The corrections are the negative of the errors. These corrections are only applied if the magnitude conditions are satisfied:

$$|\delta z_2^{(7)}| > 2 \cdot t_{all} [(B_1^2)^2 + (B_2^3)^2]^{½}$$
and
$$|\delta z_3^{(7)}| > 2 \cdot t_{all} [(B_2^3)^2 + (B_3^4)^2]^{½} \quad \text{(A7)}$$

The superscript (7) on the corrections indicates that they are of type 7--two height corrections at adjacent levels.

Two Temperature Corrections (Type 8)

In the case of one temperature error at two adjacent levels we have

$$z'_2 = z'_3 = 0, \text{ and } t'_2, t'_3 \text{ are not zero,}$$

while at least s_1^2 and/or s_3^4 is large.

$$-t'_2 = X_1^2 - 2t_1^2$$
$$-t'_2 - t'_3 = X_2^3 - 2t_2^3 \quad \text{(A8)}$$
$$-t'_3 = X_3^4 - 2t_3^4$$

From these equations

$$X_1^2 - X_2^3 + X_3^4 = 2(t_1^2 - t_2^3 + t_3^4) \quad \text{(A9)}$$

The temperature errors have been eliminated from this equation. Squaring the equation, averaging over many realizations and assuming the t_i^{i+1} to be independent of each other gives

$$\overline{(X_1^2 - X_2^3 + X_3^4)^2} = 4[\overline{(t_1^2)^2} + \overline{(t_2^3)^2} + \overline{(t_3^4)^2}$$
$$= 12(\overline{t})^2$$

Therefore

$$\overline{|X_1{}^2 - X_2{}^3 + X_3{}^4|} = 2\sqrt{3}\ \bar{t} \tag{A10}$$

Two temperature errors are diagnosed when

$$\overline{(s_1{}^2 + s_2{}^3 + s_3{}^4)} < 2\sqrt{3}\ \bar{t}_{all} \tag{A12}$$

where $\bar{t}_{all}$ has the same value used for other error types. The magnitude of the errors is determined from Eq. (A8, i.e.

$$\begin{aligned} z'_2 &= s_1{}^2 + 2B_1{}^2 \bullet t_1{}^2 \approx s_1{}^2 \\ z'_3 &= -s_3{}^4 - 2B_3{}^4 \bullet t_3{}^4 \approx -s_3{}^4 \end{aligned} \tag{A13}$$

The corrections are the negative of the errors. These corrections are only applied if the magnitude conditions are satisfied:

$$\text{and} \quad \begin{aligned} |\delta T_2{}^{(8)}| &> 2 \bullet t_{all} \\ |\delta T_3{}^{(8)}| &> 2 \bullet t_{all} \end{aligned} \tag{A14}$$

The superscript (8) on the corrections indicates that the corrections are of type 8--two temperature corrections at adjacent levels.

Lower Height Correction, Upper Temperature Correction (Type 9)

The conditions that must be satisfied in order to diagnose a height correction at the lower level and a temperature correction at the upper level will be derived. For all the correction types which involve three layers, it is necessary that the outer two residuals have sufficent size. This will be seen in the derivation that follows.

Under the assumption of only errors to the lower height and upper temperature:

$T'_2 = z'_3 = 0$ while z'_2 and T'_3 are not zero.

The residual equations for the three layers are

$$
\begin{aligned}
z'_2 &= s_1^2 - 2B_1^2 \bullet t_1^2 \\
-z'_2 - B_2^3 \bullet T'_3 &= s_2^3 - 2B_2^3 \bullet t_2^3 \qquad (A15) \\
-B_3^4 \bullet T'_3 &= s_3^4 - 2B_3^4 \bullet t_3^4
\end{aligned}
$$

Adding the first two equations and substituting into the third gives

$$s_1^2 + s_2^3 - \frac{B_2^3}{B_3^4} \bullet s_3^4 = -2(B_2^3 \bullet t_3^4 - B_1^2 \bullet t_1^2 - B_2^3 \bullet t_2^3) \qquad (A16)$$

Squaring and averaging over many realizations, with the assumption of independence of the t_i^{i+1}'s gives

$$
\begin{aligned}
\overline{\left[s_1^2+s_2^3-\frac{B_2^3}{B_3^4}\bullet s_3^4\right]^2} &= 4\left[(B_1^2)^2\overline{(t_1^2)^2}+(B_2^3)^2\overline{(t_2^3)^2}+(B_2^3)^2\overline{(t_3^4)^2}\right] \\
&= 4[(B_1^2)^2 + 2(B_2^3)^2]\ (\bar{t})^2 \qquad (A17)
\end{aligned}
$$

Therefore

$$\overline{\left[s_1^2+s_2^3-\frac{B_2^3}{B_3^4}\bullet s_3^4\right]} = 2[(B_1^2)^2 + 2(B_2^3)^2]^{1/2}\ \bar{t} \qquad (A18)$$

Errors to the lower height and upper temperature are diagnosed when

$$\overline{\left[s_1^2+s_2^3-\frac{B_2^3}{B_3^4}\bullet s_3^4\right]} < 2[(B_1^2)^2 + 2(B_2^3)^2]^{1/2}\ \bar{t}_{all} \qquad (A19)$$

and the errors are determined from Eqn. (A15). They are

$$
\begin{aligned}
z'_2 &= s_1^2 + 2B_1^2 \bullet t_1^2 \approx s_1^2 \\
T'_3 &= -s_3^4/B_3^4 - 2t_3^4 \approx -s_3^4/B_3^4 = X_3^4
\end{aligned}
\qquad (A20)
$$

The corrections are the negative of the errors, but are not applied unless they satisfy the magnitude conditions:

$$|\delta z_2^{(9)}| > 2 \cdot t_{all} \, [(B_1^2)^2 + (B_2^3)^2]^{1/2}$$

and (A21)

$$|\delta T_3^{(9)}| > 2 \cdot t_{all}$$

The derivation of the conditions for a lower temperature correction and upper height correction is analogous.

APPENDIX B

ERROR-TYPE CONFUSION PARADOX

There are arrangements of residuals for which it is impossible to distinguish between various possible errors. These cases are distinguished by the simultaneous satisfaction of the existence and magnitude conditions for more than one error type, and only occur for error types 7 & 8 or 9 & 10. When such an ambiguity exists, the DMA makes no correction. The following section will illustrate the reasons for this ambiguity.

Two height errors or two temperature errors

For simplicity, it is assumed that the layer B's are identical and that the small nonlinear terms in the residual equations may be neglected, leading to the following equations.

$$s_1{}^2 = z'_2 - B \bullet T'_2$$
$$s_2{}^3 = z'_3 - z'_2 - B \bullet (T'_2 + T'_3) \qquad (B1)$$
$$s_3{}^4 = -z'_3 - B \bullet T'_3$$

It is easily seen then, that two height errors of the same value and sign,

$$z'_2 = z'_3 = z',$$

will lead to residuals

$$s_1{}^2 = z'; \quad s_2{}^3 = 0; \quad s_3{}^4 = -z',$$

while, two temperature errors of the same value and opposite signs,

$$T'_2 = -T'_3 = T',$$

will produce the residuals

$$s_1{}^2 = -B \bullet T'; \quad s_2{}^3 = 0; \quad s_3{}^4 = B \bullet T'.$$

It is thus impossible to distinguish between these two quite different corrections, based only upon the hydrostatic residuals. In both cases the middle residual is zero while the two others are of the same value and opposite signs.

The existence conditions for error types 7 and 8 can also be used to investigate the conditions for error-type confusion. With identical B's for the layers, the existence conditions for error types 7 and 8 are

$$|s_1^2 + s_2^3 + s_3^4| < 2\sqrt{3}\ B \bullet \bar{t}_{all}$$

and

$$|s_1^2/B - s_2^3/B + s_3^4/B| < 2\sqrt{3}\ \bar{t}_{all}. \qquad \text{(B2)}$$

When s_2^3 is small, these conditions are indistinguishable.

Lower height error and upper temperature error or lower temperature error and upper height error

Analogously, it is easy to show, based upon the same equations (B1) that if the middle residual is zero but the two others are of the same value and of the same sign, then the cause could be either a height error, z'_2, below and a temperature error, T'_3, above it, the two connected by the equation

$$B \bullet T'_3 = -z'_2$$

or, quite differently, a temperature error, T'_2, below and a height error, z'_3, above it, obeying the equation

$$z'_3 = -B \bullet T'_2,$$

so that, again, it is impossible to decide between these two combinations, using only the hydrostatic residuals.

Again, the existence conditions can be used to see the potential difficulty. For identical layer B's, the existence conditions for error types 9 and 10 are

$$|s_1{}^2 + s_2{}^3 - s_3{}^4| < 2\sqrt{3}\ B\cdot\bar{t}_{all}$$

and (B3)

$$|s_1{}^2 - s_2{}^3 - s_3{}^4| < 2\sqrt{3}\ B\cdot\bar{t}_{all}.$$

These conditions clearly show that the difficulty is present only when $s_2{}^3$ is small. For a more general variation of B between the layers, the general existence and magnitude conditions must be used to determine cases where the error types cannot be distinguished.

REFERENCES

Gandin, Lev S., 1988: Complex quality control of meteorological observations. Mon. Wea. Rev., 116 (5), 1138-1156.

Gandin, Lev S., 1989: Comprehensive quality control. Seminar on Data Assimilation and the Use of Satellite Data, Reading, U.K., European Centre for Medium Range Weather Forecasts, XXX-XXX.

List of Figures

Table 1 - Admissible Residuals

pressures (hPa)	admissible residuals (meter)
1000-850	65.
850-700	35.
700-500	50.
500-400	35.
400-300	40.
300-250	35.
250-200	40.
200-150	50.
150-100	85.
100-70	70.
70-50	70.
50-30	80.
30-20	70.
20-10	100.

Table 2 - Corrections

Error Type	Corrections
1 - single height error	$\delta z_3^{(1)} = -\frac{1}{2}(s_2^3 - s_3^4)$
2 - single temperature error	$\delta T_3^{(2)} = \frac{1}{2}(X_2^3 + X_3^4)$
3 - pair at the same level	$\delta z_2^{(3)} = (B_2^3 \cdot s_3^4 - B_3^4 \cdot s_2^3)/(B_2^3 + B_3^4)$ $\delta T_2^{(3)} = (s_2^3 + s_3^4) / (B_2^3 + B_3^4)$
4 - error at bottom	$\delta z_2^{(4)} = s_2^3$ $\delta T_2^{(4)} = X_2^3$
5 - error at top	$\delta z_2^{(5)} = -s_2^3$ $\delta T_2^{(5)} = X_2^3$
7 - two height errors	$\delta z_2^{(7)} = -s_1^2$ and $\delta z_3^{(7)} = s_3^4$
8 - two temperature errors	$\delta T_2^{(8)} = -X_1^2$ and $\delta T_3^{(8)} = X_3^4$
9 - lower height upper temperature	$\delta z_2^{(9)} = -s_1^2$ and $\delta T_3^{(9)} = X_3^4$
10 - lower temperature upper height	$\delta T_2^{(10)} = -X_1^2$ and $\delta z_3^{(10)} = s_3^4$

Table 3 - Existence Conditions

Error Type	Existence Condition
1 - single height error	$\lvert s_2^3 + s_3^4 \rvert < 2\,\bar{t}_{all}\left[(B_2^3)^2 + (B_3^4)^2\right]^{1/2}$
2 - single temperature error	$\lvert X_2^3 - X_3^4 \rvert < 2\,\bar{t}_{all}$
7 - two height errors	$\lvert s_1^2 + s_2^3 + s_3^4 \rvert < \left[(B_1^2)^2 + (B_2^3)^2 + (B_3^4)^2\right]^{1/2}$
8 - two temperature errors	$\lvert X_1^2 - X_2^3 + X_3^4 \rvert < 2\sqrt{3}\,\bar{t}_{all}$
9 - lower height upper temperature errors	$\left\lvert s_1^2 + s_2^3 - \frac{B_2^3}{B_3^4}\cdot s_3^4 \right\rvert < 2\left[(B_1^2)^2 + 2(B_2^3)^2\right]\bar{t}_{all}$
10 - lower temperature upper height errors	$\left\lvert s_2^3 + s_3^4 - \frac{B_2^3}{B_1^2}\cdot s_1^2 \right\rvert < 2\left[(B_3^4)^2 + 2(B_2^3)^2\right]\bar{t}_{all}$

Table 4 - Magnitude Conditions

Error Type	Magnitude Condition
1 - single height error	$\lvert \delta z_3^{(1)} \rvert > 2 \cdot \bar{t}_{all} \left[(B_2^3)^2 + (B_3^4)^2 \right]^{1/2} \equiv Z^*_3$
2 - single temperature error	$\lvert \delta T_3^{(2)} \rvert > 2 \cdot \bar{t}_{all} \equiv T^*_3$
7 - two height errors	$\lvert s_1^2 \rvert > Z^*_1$ and $\lvert s_3^4 \rvert > Z^*_3$
8 - two temperature errors	$\lvert X_1^2 \rvert > T^*_1$ and $\lvert X_3^4 \rvert > T^*_3$
9 - lower height upper temperature	$\lvert s_1^2 \rvert > Z^*_1$ and $\lvert X_3^4 \rvert > T^*_3$
10 - lower temperature upper height	$\lvert X_1^2 \rvert > T^*_1$ and $\lvert s_3^4 \rvert > Z^*_3$

Table 5 - Magnitude Conditions

pressure (hPa)	temperature (Deg K)	height (meter)
1000	7.	35.
850	7.	26.
700	7.	40.
500	7.	41.
400	7.	37.
300	7.	35.
250	7.	30.
200	7.	37.
150	7.	51.
100	7.	55.
70	7.	60.
50	7.	63.
30	7.	67.
20	7.	82.
10	7.	99.

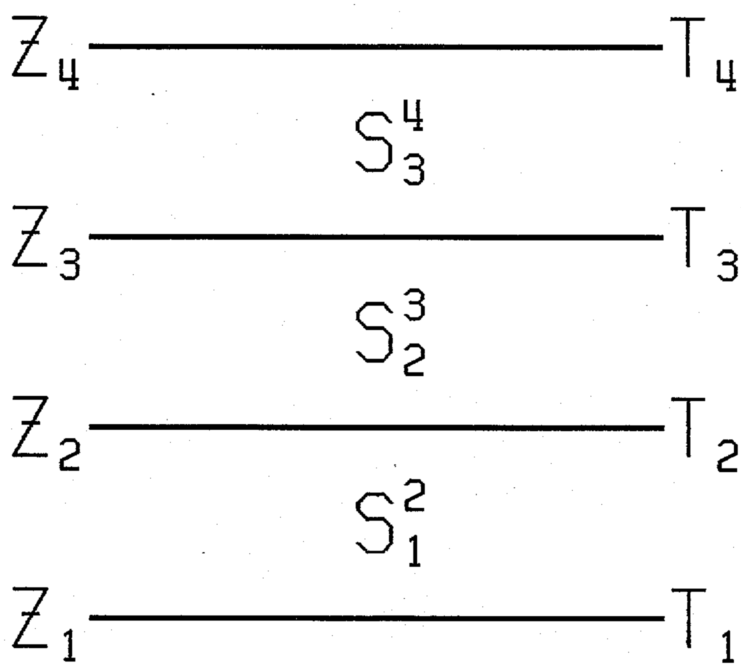

Fig. 1. Schema for pressure, height, temperature and residuals.

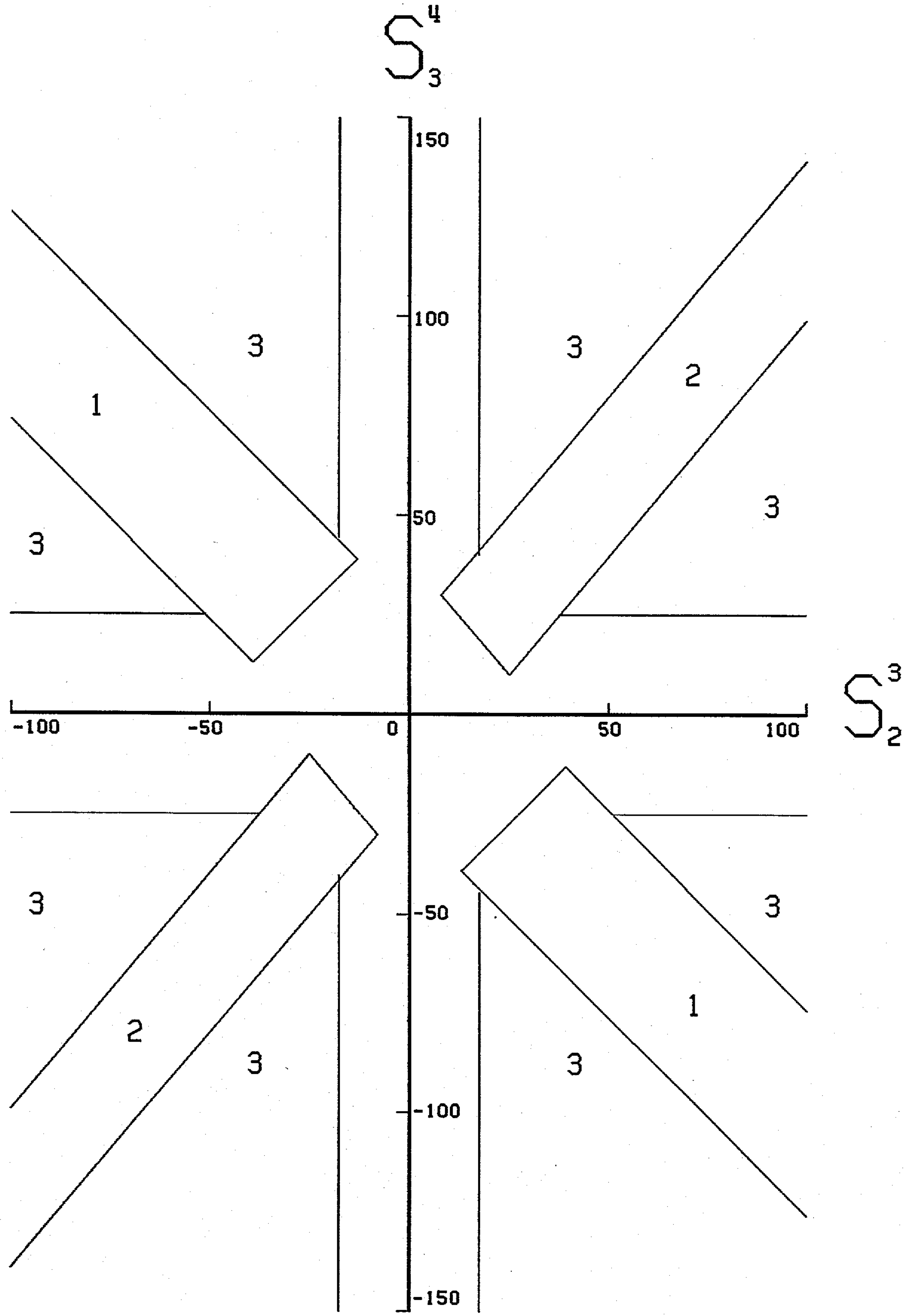

Fig. 2. Existence conditions for error types 1, 2, and 3.