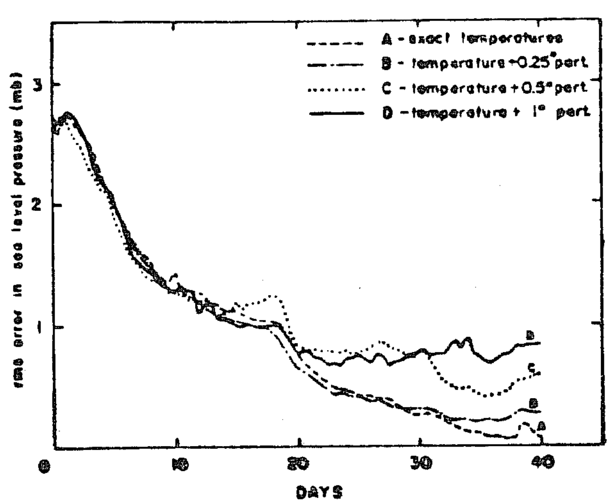
OBSERVING SYSTEM EXPERIMENTS - REVIEW AND OUTLOOK

A. Gilchrist
Meteorological Office, Bracknell, Berkshire, U.K.

1. REVIEW OF OBSERVING SYSTEM EXPERIMENTS

1.1 The first phase of activity on observing system experiments may be considered to start with the work of Charney and others (e.g. Charney et al, 1969) in the mid-sixties. With the prospect of an operational system of meteorological satellites, they were concerned to demonstrate the usefulness of satellite-derived temperatures. A typical result of the experiments is shown in Fig. 1. Temperatures, extracted from a run of a general circulation model were converted into 'pseudo-observations' by the addition of a random error. The model was then started from an arbitrary initial state and the temperatures were inserted into it during an integration in a manner which mimicked the likely availability of temperatures from a sat-

FIGURE 1



The rms error in sea level pressure anomaly (mb) for the Northern Hemisphere, in cases where temperature with 0, 0.25, 0.5 and 1C random error perturbations are inserted every 12 hr at all grid points.

ellite. Initially, the r.m.s. temperature difference from the original run was large (Fig. 1) but as the temperatures were inserted over a period of time, it decreased, eventually reaching a value approximating to the assumed observational error, at which time the atmospheric situation of the original run had, effectively, been recovered.

1.2 Other investigations around the same time, for example Gordon et al (1972), Williamson and Kasahara (1971) and Kasahara and Williamson (1972), used a similar approach to investigate the need for observations, particularly of wind observations in the tropics. They too used simulated observations, derived from the same model that was used for the assimilation, a technique now identified as the 'identical twin' approach. The technique was seized upon by investigators because it seemed to offer a valid way of answering questions concerning observing systems without becoming involved in the complexities of operational forecasting systems and the results pointed to fairly clear-cut conclusions. It was of course realised that it had limitations. It took a comparatively long time to reach a reasonable approximation to the 'atmospheric' situation, the implication being that much of the information in the data presented to the model was lost as 'noise'. There were nevertheless grounds for hoping that the asymptotic situation could be taken to give an accurate representation of what could, with more refined assimilation, be achieved practically in a much shorter time. Investigators were at pains to point out that the results might be model-dependent and that consequently the results could not be accepted without further corroboration. It is doubtful however if it was realised how serious the shortcomings of the results were because of the use of the 'identical twin' approach. The validity of investigations such as that of

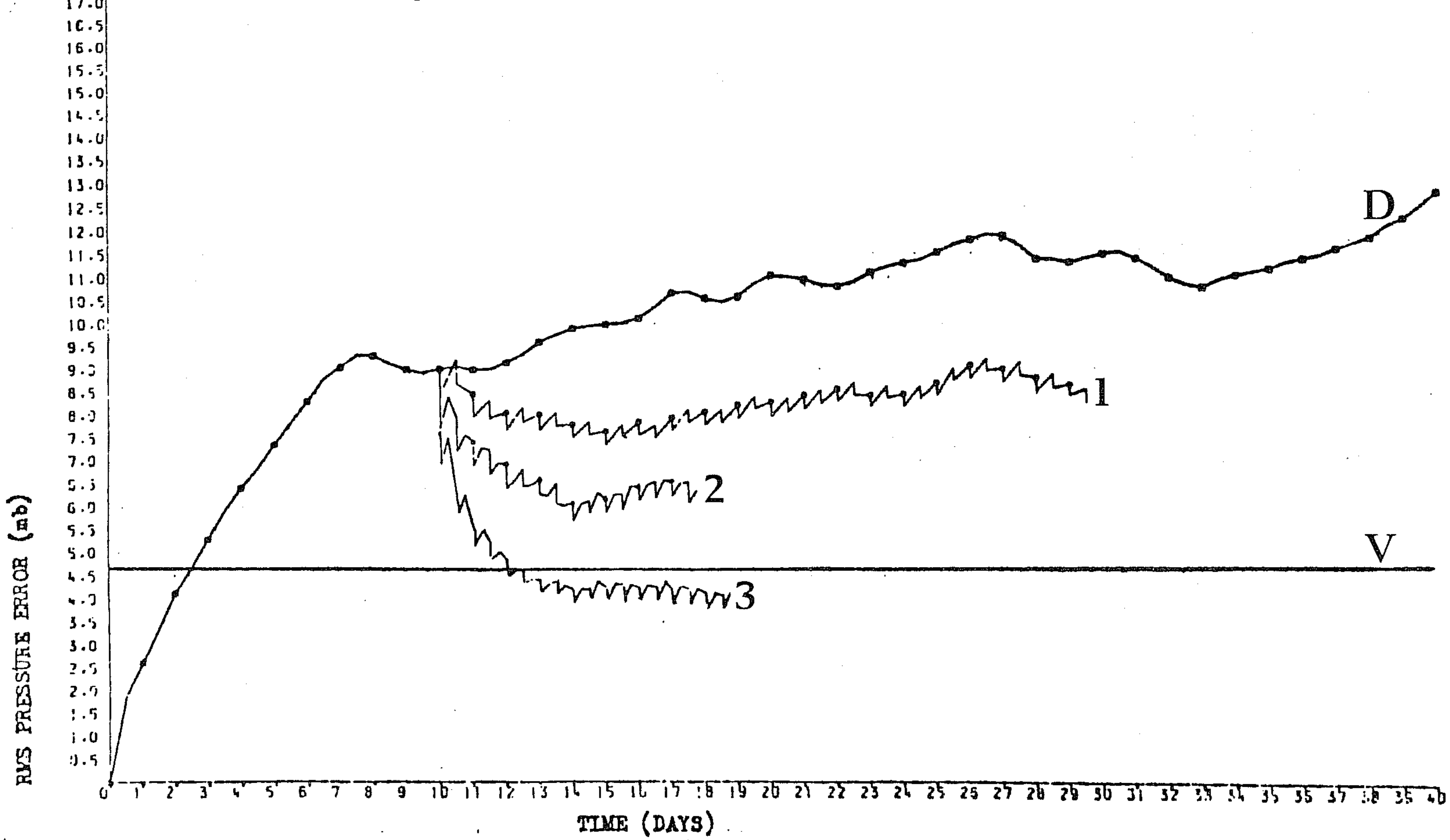
Smagorinsky et al (1970) which indicated "that some variables such as surface pressure, tropospheric humidity, boundary layer wind and boundary layer temperature are not really essential as initial data for prediction of extra-tropical cyclone-scale weather for a few days ahead" may be similarly limited by the fact that some of these variables are implied to an unrealistic extent by the other model variables. Compared with the real atmosphere the number of degrees of freedom in a model is small, and data which conform exactly to the implied model constraints therefore appear to the model to have a high information content. Observations from the real atmosphere, or from an independent model, however, contain information which conflicts with the constraints, and therefore when they are inserted a model may not be able to extract information effectively. For example, Miyakoda and Talagrand (1971) found that if geopotential information was inserted continuously into a barotropic model without modification there was an 'intolerable error growth'. Morel et al (1971) compared the insertion of model data and independent data into a simple 1-layer model and concluded that great care was needed in interpreting an identical twin type experiment.

1.3 The problem is illustrated in Fig. 2, taken from Lorenc (1975). The curve of the root mean square surface pressure differences when simple insertion of data from a second model was used levels off to values considerably above the model's natural variability; when a form of optimum interpolation was used to spread the information the errors were substantially reduced (curve 2); they were further reduced when, additionally, measures were taken to correct the model's climatology (curve 3).

GLOBAL RMS PRESSURE ERRORS FOR DIFFERENT ANALYSIS SCHEMES

FIGURE 2

1 Simple updating of OSL
2 Interpolation of OSL
3 Interpolation of OSL with 'climatological' data
D Difference between assimilating and 'truth' models
V Variability of 'truth' model



1.4 Following this initial phase of Observing System Experiments there was a second, directed towards aiding the process of planning for the FGGE. It was started by the preparation of the paper "The performance of space observing systems for the first GARP global experiment" by Bengtsson and Morel (1974) and the meeting of the JOC at its 9th session in Canberra at which it was considered and where recommendations for the experiments were made. The main experiments are reported in WGNE Reports 10 and 16. Importantly they departed from the procedure of using an 'identical approach' methodology. Experiments were carried out at the Meteorological Office and at NCAR; a network-type study was also undertaken at ECMWF. In the Meteorologial Office experiments, Lorenc (1975) the 'truth' and 'assimilation' models differed not only in respect of their resolution, but also in their parametrizations, the effect being such that the 'truth' model had a climatology that was closer to reality. Williamson (1975) used the same basic model for both aspects of the experiment, but the assimilating model had a lower resolution. No specific recommendations were made about the techniques to be used for assimilating the data, but it quickly became clear that this was one of the crucial aspects of the experiments. Fig. 2 is taken from Lorenc's paper (1975) and illustrates the differences in his system of using different assimilation techniques. The NCAR experiments were based on direct insertion of the 'pseudo-observations.

1.5 In all the experiments there were serious problems of data acceptance particularly as regards surface pressure observations, whose effect could all too readily be dissipated in gravity waves. The problems remained significant even when the influence of the data was spread to nearby gridpoints by univariate interpolation techniques. Judicious interpretation of the results was therefore necessary to interpret the results of the experiments for the purpose of planning observing system dispositions for the FGGE.

1.6 Thus, the early experiments and these pre-FGGE investigations may be seen as leading progressively to the realisation that the initial assumption that there were shortcuts, which would enable the assessment of observing systems to avoid using a full operational forecasting system, was illusory. It became clear that for an assessment to be realistic it was essential for the system to extract the maximum amount of information from the observations and results from systems which do not do so are suspect. Equally, the characteristics of real observations have to be represented in the information presented to the assimilating system, and none of the simulation experiments achieved this satisfactorily. The period during and since the FGGE has been marked by an abandonment, at least temporarily, of observing simulation experiments in favour of experiments with real observations and operational forecasting systems. The implications of the change are very substantial. New problems such as those concerned with data selection, quality control and data redundancy which it was hoped could be side-stepped have to be faced, and they introduce new elements of uncertainty into the interpretation of the results. As the results can only be considered credible if 'state-of-the-art' systems are used, only the larger forecasting centres can provide the essential basis for the experiments, and the contributions of the individual scientist is likely to be through the improvement, or increased understanding, of the operational system. This loss of control of the whole experiment by the individual scientist may lead to a lack of commitment by some of those most able to contribute at a fundamental level. Further the experimentts are inevitably performed for a particular limited period of time, usually fairly short, and both because of the characteristics of the atmospheric circulation prevailing, and details of the availability of observations, at the time, the results may lack the generality that is needed for planning decisions. On the positive side of the balance, it has become clear that the performance of assimilation systems in observing system experiments has proved to be one of the most effective ways of testing and indicating improvements to them. Their characteristics are dissected and their weaknesses in treating specific observation types are highlighted.

1.7 Among the earliest experiments of this kind were those carried out before the FGGE to examine the need for Atlantic Ocean Weather Ships, Atkins and Jones (1974), Barnes and Ireland (1979), and the FGGE Data Systems Tests conducted by GISS, GFDL, and NMC (Tracton and McPherson (1977), Halem et al. (1978), Ghil et al. (1979)). The problems of selection emerge rather clearly from the results of these experiments.

1.8 The period since 1980 has been marked by a series of experiments using the full observational data set available for the FGGE. Comparisons have been made of analyses and forecasts made using the full data set and then only particular subsets of the data. The experiments up till 1982 have been reviewed in the Exeter Conference Report (WGNE Report No. 4 in the GARP/WCRP series). Additional experiments carried out since that time are being reviewed at the present meeting.

1.9 The essential points from the conclusions of the Exeter Conference can be summarized very briefly as the following:

1. Satellites are essential to the Global Observing System in both their observing and telecommunication roles, SATEMs being their most useful product.

2. 'Single level' wind observations from satellites and aircraft make a highly significant contribution to analyses.

3. During the FGGE, buoys made a substantial impact on analyses and forecasts for the southern hemisphere, and could improve analyses and forecasts for the North Atlantic if deployed in existing data gaps.

4. Atlantic radiosondes from ships or coastal stations are important for forecasting downstream.

There are two further conclusions related to the assimilation techniques:

5. Assimilation of 'single level' observations needs to be improved for further testing of their impact, particularly in forecasts.

6. Quality control procedures must be improved for results from observing systems experiments to become more uniform.

1.10 One of the hopes for the present assessment is that as a result of the new experiments carried out since 1982, it will be possible to remove some of the limitations implied in the above statements, especially with respect to 2., the use of single level observations.

The most important conclusion in the report is the first, concerning the essential nature of satellite observations to the global observing system. The GARP was undertaken precisely because the new technologies of remote sensing and computers had reached the point at which there were good grounds for believing that they could make a major impact on the practice of weather forecasting, and the period from the early sixties to the early eighties can be seen as one in which the essential nature of the satellite system to the global observing network was progressively established. During that time the quality of the infra-red soundings and wind measurements from satellites was improved to the point where their usefulness, in a global sense, was not in doubt.

1.11 However the emergence of satellites as the basis of the global system has occurred despite the fact that, at least for the northern hemisphere, considerable work remains to be done to ensure that satellites provide a useful addition to the conventional network on all occasions. Temperatures retrieved using the normal procedures show characteristic errors which are correlated spatially, frequently making them difficult to detect and correct.

2. EXAMPLE OF 'HERMES' SATELLITE RETRIEVALS

2.1 As an illustration of an occasion when the retrieved satellite temperatures could be misleading, and where therefore improvements are necessary in retrieval techniques if satellite temperatures are to make a reliable addition to the information already available in the northern hemisphere, the case of 2nd March, 1984 is presented. A more detailed description is available in an internal Meteorological Office Note by Adams (1984).

2.2 The temperatures were obtained from the Meteorological Office 'Hermes' system of local retrievals from the TIROS-N series of NOAA satel-

FIGURE 3

12Z UPDATE ANALYSIS 2/3/84
THICKNESS(DAM), WIND SHEAR(M/S) & VERIFYING OBSERVATIONS
VALID AT 12Z ON 2/3/1984 DAY 62
LEVEL: 500 MB - 1000 MB

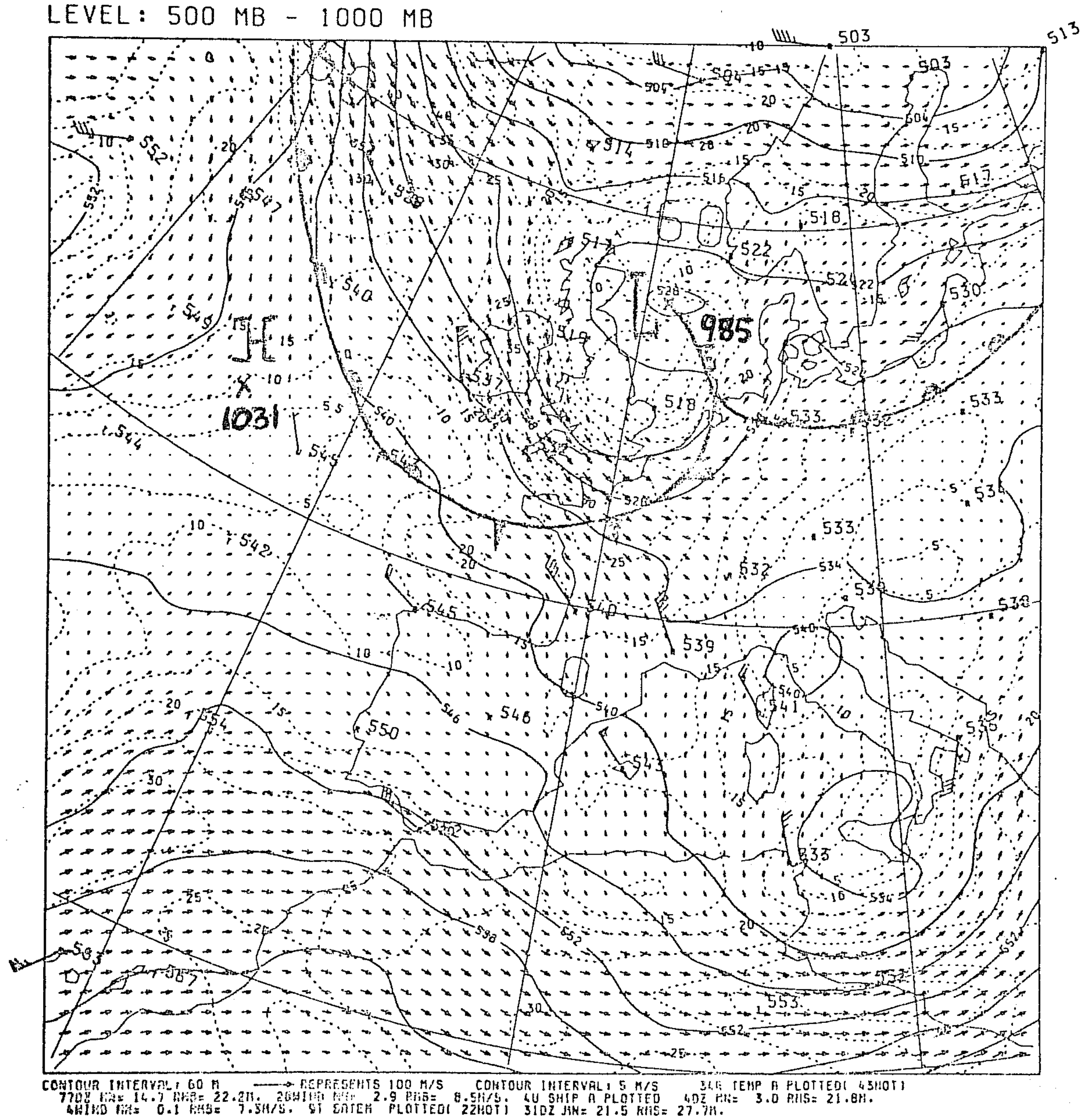
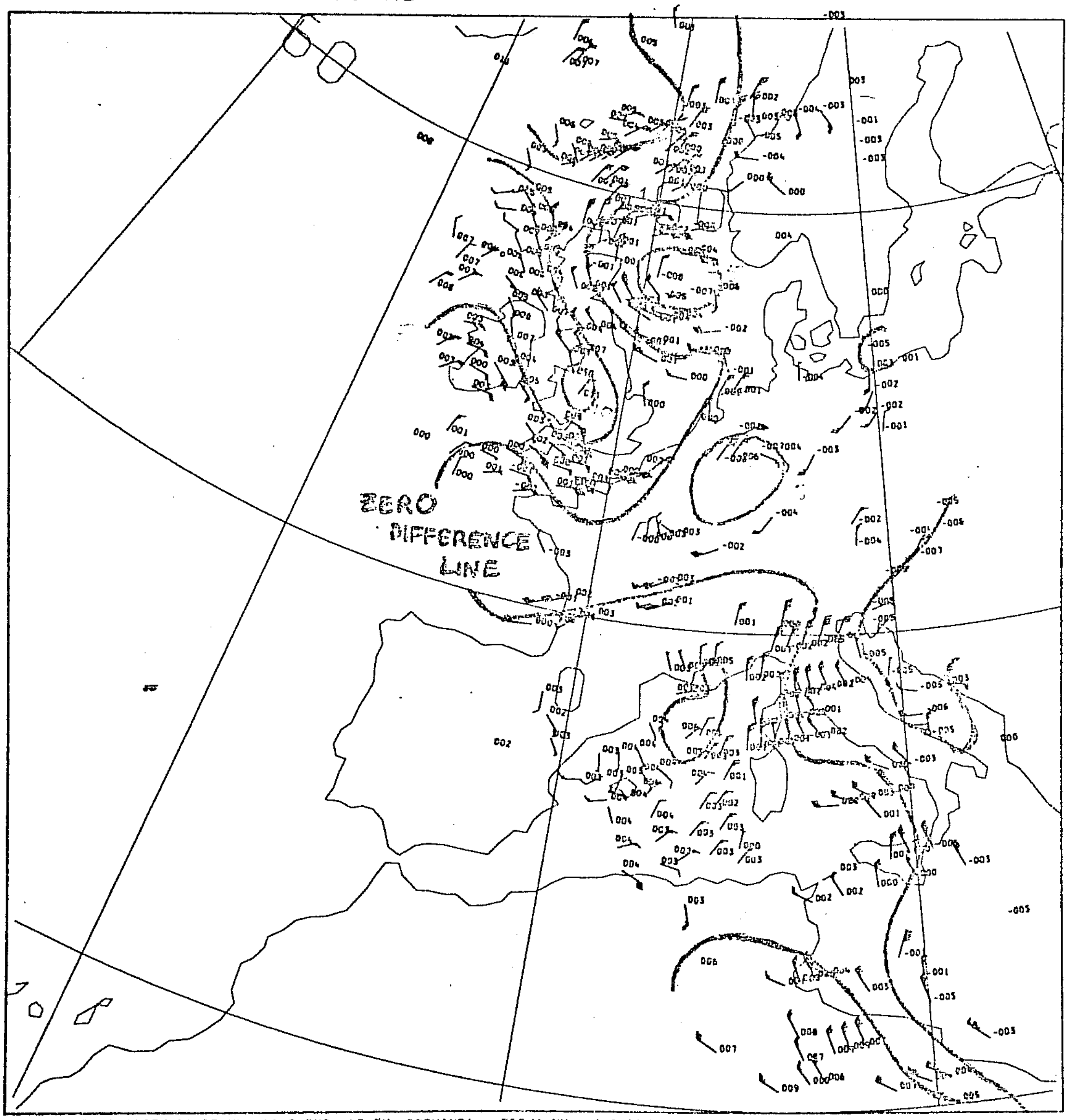


FIGURE 4

12Z UPDATE ANALYSIS 2/3/84
OBSERVED - FIELD VALUES
VALID AT 12Z ON 2/3/1984 DAY 62
LEVEL: 500 MB - 1000 MB

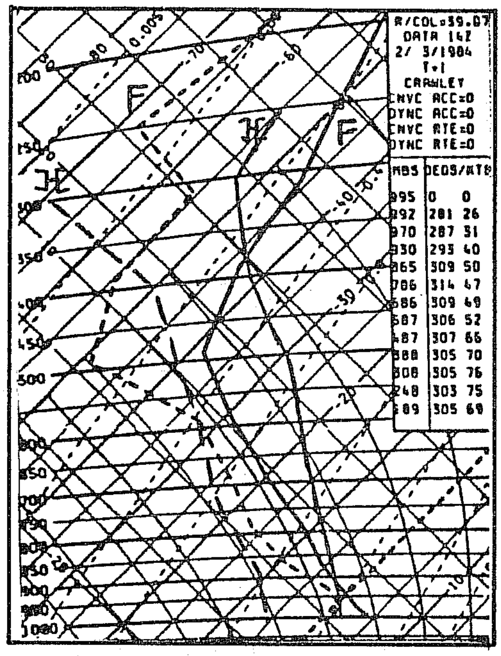


504H HERMES PLOTTED 504OZ HH= 11.6 RMS= 40.3M. 2GOHINDI 3REJ) HH= -1.2 RMS= 14.0M/S.

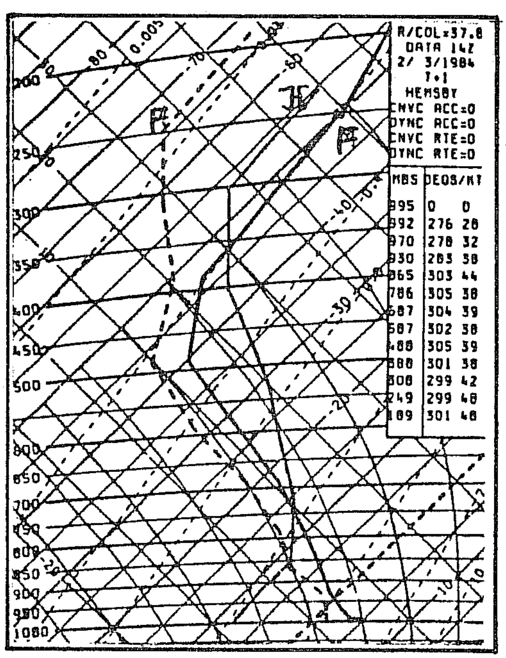
FIGURE 5 COARSE-MESH FORECAST 1430 (F)

1430 HERMES DATA (H)

CRAWLEY



HEMSBY



lites. The retrieval technique for clear sky radiances is the same as that used for SATEMs. The cloud-clearing algorithm is not the same as that used by NESS but comparison of SATEM and Hermes temperatures for this occasion indicated no significant differences.

2.3 At 12 Z, 2 March, there was a deep trough over the British Isles. Fig. 3 shows the field of 500-1000 mb thickness with radiosonde values below 520 dm at a number of stations. The Hermes retrievals gave values that were too high over the U.K., and too low over the North Sea, (Fig. 4) where there was relatively warm air. Careful examination has confirmed that the differences were not accounted for by small differences in the times of the measurements. In Fig. 5 are shown the retrievals and the forecast temperature values from the 12Z forecast (which did not use the Hermes data) for the same time both for the data available closest to Crawley and Hemsby, respectively. The model forecast ascents retained most of the information inserted 2½ hr before from the 12Z ascent. It is evident that the main problem with the Hermes retrievals is that they did not reproduce the very low tropopause heights which occurred on this occasion, and consequently did not indicate a sufficiently deep trough.

2.4 Because these errors are spatially correlated and affect a large area, they are potentially damaging to a forecast. In this example they affected an area which is relatively well covered by radiosondes and the error could be detected. In more sparsely observed areas this might not be possible.

2.5 This particular study did not set out to be an observing system experiment. Nevertheless, it is one in effect, with results no less significant than those from experiments formally set up as such. It is, I believe going to become increasingly important in the future to gather together the results of case studies of this kind in order to supplement results that may come from more normal routes.

3. OUTLOOK

3.1 I believe we have now reached the end of a phase in observing system experiments and that the problems which are about to be confronted are of a different nature from those that have been dealt with in the past. Up till now the dominant problem has been that of assessing the part of satellites in the global system. That is now secure in the sense that there is universal acceptance that satellites make an essential contribution which cannot be provided by any ground-based systems. The question of the security of resources to maintain the system is of course separate and subject to the same uncertainties that affect all resource allocations.

3.2 The spotlight of concern and therefore to some extent of scientific interest has moved to other areas, particularly weather ships and the conventional land-based network of upper air observations. The search is on for economies that might compensate at least in part for the additional expenditure on space-based systems and for the improvement of the conventional network in limited areas where it is very deficient. Important changes in the role and perhaps even the methodology of observing system experiments may be necessary.

3.3 Comparing the situation that obtained during the previous phase with that we are now in, there are some very obvious differences. Firstly, the global satellite system was introduced in a period which was, for most of the time at least, one of expansion, and when the tide generally was running swiftly in favour of remote sounding from space. Secondly, the satellite information has been additional to what was already available. Generally speaking, though with some important exceptions, the other systems have remained intact and mostly the response of Meteorological Services has been one of great interest, but at the same time of considerable relaxation because no part of the systems on which they have come to depend was being withdrawn. In these circumstances, the impact on the services provided could only be beneficial. Thirdly, because the emphasis was on the possibilities of using global observing systems, it was natural to direct attention to global forecasting and, concomitantly, forecasts over extended time ranges. It will be recalled that the purpose of the GARP was explicitly related to the possibility of extending global predictions for longer into the future. A result of this emphasis, combined with the general acceptance in the late seventies that observing systems experiments had to involve the best operational forecasting systems, was that the number of centres taking part in the experiments was few. This could be part of the reason for the great majority of meteorological services taking only a passing interest in what was being done.

3.4 The situation today is very different. The observing systems that are now being questioned are those that forecasters and forecasting systems generally have come to rely on over the last thirty to forty years, and resources for the expansion of meteorological observing systems are much more difficult to come by than they were. The changes in observing systems now being contemplated are much closer to the vital concerns of meteorological services, whose main requirement is to produce forecasts for fairly short periods ahead - up to 24 or 36 hours, containing as much detail as possible about the surface weather conditions. It is sometimes overlooked that the ability to provide foreacasts of this kind has grown steadily over the last 20 years, but particularly over the last decade, partly as a result of the improved parametrizations it has been possible to include in forecasting models. In some respects this improvement in forecasting is no less striking than the improvement seen in longer period forecasts and it is certainly possible, and perhaps even likely, that, of the two, its impact on forecasting services has been the greater as far as the majority of customers is concerned. Attention is now being focussed on even shorter period forecasts with even greater detail and precision, it being clear that the market for forecasts of this kind is very great indeed. Now, the requirements for short period forecasts differ significantly from those for longer periods. For example, the need for detail in the vertical structure of the atmosphere and for accurate representation of the humidity fields is much greater for short than for medium range forecasts, and this is the kind of information that is provided by conventional upper air sounding techniques, but generally lacking in space-based observations.

3.5 For observing system experiments, the need to explore shorter range forecasts and therefore smaller scale meteorological systems poses additional and more fundamental problems. Whereas on the large scale, atmospheric systems are rather well understood and on the whole well forecast by modern numerical weather prediction systems, understanding is much less for smaller scale meteorological features such as fronts and meso-scale developments. It is by no means clear that we have the most appropriate techniques for forecasting them (for example, the handling of scale-collapse in a frontal region may call for a new approach to grid representation), and the concept of 'balance', which is so useful in dealing with large scale features, remains to be explored in the context of meso-scale forecasting. Nevertheless it is certainly conceivable that further improvement of forecasting, even for extended forecasting may depend critically on understanding and reproducing more realistically the dynamics of these smaller systems. In these circumstances, our ability to specify what the observational requirements are with any finality must be severely limited.

3.6 Alongside this factor of increasing difficulty in carrying out the observing system experiments, two other relevant factors remain unchanged. The first is that the total effort worldwide devoted to the problem is pitifully small. Only a handful of scientists are involved, and not all of them for the major portion of their time. The second is that the complexity of the questions being posed by those concerned with planning a global observing system shows no sign of lessening. Thus, planners wish to have information on the 'best mix' of observing systems. Initially, this may appear a simple concept but on consideration it is highly complex. For what purpose is the mix to be best? - for all forecasting activities or only global forecasts for medium range prediction? What is the meaning of 'best'? is it irrespective of cost or is the concept of cost incorporated in it? and if the latter by what mechanism can observing system experiments take costs into account in any realistic way?

3.7 I wish to make the point, and to make it very strongly that there is no realistic prospect of observing system experiments providing the kind of information that planners are asking for unless the effort put into them is increased by a very substantial amount.

3.8 For the future, there are clear indications that the need to test and demonstrate the performance of observing systems is likely to increase. New types of instrument are being proposed; as are alternative dispositions of existing observing systems; improved techniques for deriving meteorological quantities from satellites are being considered. Operational WWW System Evaluations (OWSE) are being planned, initially for the Atlantic, and their general feasibility is being investigated.

3.9 It is desirable to avoid the shortcomings of past Observing System Experiments. This requires two changes in particular. Firstly, an increased effort in the subject as a whole. In this respect, however, it should be noted that many results that are relevant to the evaluation of observing systems come from investigations that were not set up formally as OSEs. The satellite results from the 'Hermes' system, quoted earlier, are but one example. In essence, it matters little where the results come from; OSEs, and studies which compare analyses and forecasts starting from different observational bases which did not set out to be OSEs, are equally valid. Perhaps, therefore, one of the main thrusts has to be in setting up a system to ensure that all relevant results are gathered together to be part of the final evaluation.

3.10 The second major change which I believe is required is that those concerned with planning the global observing system should take more account of what is feasible in this field. This will undoubtedly call for more cooperation between them and modellers engaged in evaluating observing systems. Particularly, this is likely to happen in the context of the Atlantic OWSE, now being planned. Because the scientific assessment must be an integral part of the OWSE as a whole, there will have to be continuing and continuous interaction between modellers and those involved in the operational implemementation of the Evaluation. It is, in my view, to be welcomed that the next phase of the development of the global observing system is likely to be dominated by this type of combined deployment-assessment operational exercise, undertaken on a regional basis, rather than by the more idealistic and holistic approach through the attempt to specify the global system in terms of a hierarchy of observations suitable for different purposes.

4. REFERENCES

Atkins, M.J. and M.V. Jones, 1974: Report on an experiment to determine the value of ocean weather ship and satellite data in numerical forecasting. Met. O. 11 Tech. Note No. 47, Bracknell, U.K.

Barnes, G.J. and D. Ireland, 1979: The Ocean Weather Ships trial of December 1978. Met. O. 11 Tech. Note No. 124, Bracknell, U.K.

Bengtsson, L. and P. Morel, 1974: The programme of space observing systems for the First GARP Global Experiment. WGNE Report No. 6, WMO, Geneva.

Charney, J. et al., 1969: Use of incomplete historical data to infer the present state of the atmosphere. J. Atmos. Sci., 26, 1160-1163.

Ghil, M. et al., 1979: Time-continuous assimilation of remote-sounding data and its effect on weather forecasting. Mon. Weath. Rev., 107, 140-171.

Gordon, C.T. et al., 1972: Simulation experiments for determining wind data requirements in the tropics. J. Atmos. Sci., 29, 1064-1075.

Halem, M. et al., 1978: The GISS sounding temperature impact test. NASA Tech. Memo. 78063, Goddard Space Flight Center.

Kasahara, A. and D.Williamson, 1972: Evaluation of tropical wind and reference pressure measurements: Numerical experiments for observing systems. Tellus, 24, 100-115.

Lorenc, A.C., 1975: Results of observing systems simulations for the First GARP Global Experiment. WGNE Report No. 6, WMO. Geneva.

Miyakoda, K. and O. Talagrand, 1971: The assimilation of past data in dynamical analysis. Part I. Tellus, 23, 310-317.

Miyakoda, K. and O. Talagrand, 1971: The assimilation of past data in dynamical analysis. Part II. Tellus, 23, 318-327.

Morel, P. et al., 1971: On initialization and non-synoptic data assimilation. Tellus, 23, 197-206.

Smagorinsky, J. et al., 1970: The relative importance of variables in initial conditions for dynamical weather prediction. Tellus, 22, 141-157.

Tracton, M.S. and R.D. McPherson, 1977: On the impact of radiometric sounding data upon operational numerical weather prediction at NMC. Bull. Amer. Met. Soc., 58, 1201-1209.

Williamson, D.L., 1975: Observing System Experiments for the First GARP Global Experiment. WGNE Report No. 10, WMO, Geneva.

Williamson D.L. and A. Kasahara, 1971: Adaptation of meteorological variables forced by updating. J. Atmos. Sci., 28, 1313-1324.