STRUCTURE, APPLICATIONS AND EXTENSIONS OF THE DATA ASSIMILATION DATABASE

1. OBJECTIVES

The objective for introducing this data base was to provide information on the use of data in the analysis. Initially only TEMP data was studied, but as data accumulated it was clear that other observation types should be included. The statistics file contains the following:-

(a) observed values

(b) departures from analysis, first guess, and initialisation

(c) analysis and pre-processing flags.

Information from this data base provides a basis for a detailed examination of the flow of observational data through the analysis, the departures found, and the decisions taken concerning acceptance or rejection. Since all analysed observational types are included, it now represents in extensive detail the results of observational assessment within the analysis.

2. CURRENT APPLICATIONS

The data are currently used to determine structure functions, to investigate error characteristics of observational systems, and as raw material for data quality monitoring. Several presentations during this workshop have expanded on some of these applications. For example, F. Delsol has described many of the monitoring processes that have been undertaken, and A. Lorenc has compared the distributions of accepted and rejected SHIP data.

3. POTENTIAL APPLICATIONS

The current applications outlined above represent a small subset of the data monitoring and error investigation possibilities such a data base presents. Work is in progress to:

(a) monitor in more detail the performance of TEMP systems

(b) determine the possibility of modelling TEMP biases

(c) extend such investigations to other observation types.

It would also be possible to use the data as a basis for MOS type models, and to apply collocation type methods of error assessment. Also it should be noted that the departures from the first guess could become departures from any forecast presented in lieu of the first guess. Thus, by small modifications to the pre-analysis pass, it would be possible to incorporate departures from a forecast requiring verification.

A potential application of the Data Assimilation Data Base is the investigation of the flow dependence of forecast error statistics, as they could be of considerable value both in quality control and in the analysis algorithms themselves. Two alternative approaches are possible. One is to assume that the analysis increment fields give a good estimate of the forecast error, and base the statistics on the increment fields. This of course presupposes a great deal about the validity of the assumptions. The alternative is to extend the Data Assimilation Data Base to include some

indicators of flow type such as analysed vorticity or observed pressure-tendency. This would remove some of the short-comings of the first approach, and the results could be directly compared with currently available results on structure functions.

4. TECHNICAL PROBLEMS

Some technical problems concerning the size of the statistics data base need to be addressed. As the data base evolved, its usefulness was realised to be far beyond the scope of its original design. Since the inclusion of TOVS data, the size of one month of statistics has increased from about 120 mbyte to 450 mbyte. This size is not excessive in terms of total data, but requires some partitioning of the data base for ease of access. If additional data is to be added, then the above problems would become even more serious.

Consideration needs to be given to the integration of the real-time black list and bias corrections with the archive. The Centre ought to consider using the archive for real-time maintenance of the black list. Adequate documentation of changes to the present system needs to be available to users of the archive.

5. TECHNICAL SOLUTIONS

A simple way of overcoming the size problem would be to simply split the statistics file into two or more parts. In the longer term, there is a need to completely re-assess the data base contents and structure. It is probable that additional information should seriously be considered for inclusion, such as vorticity, pressure change, static stability, and data related to positional errors. There is also a duplication of data in that the observed

values are retained. While the current format is useful in the current working conditions at ECMWF, faster access to observational data in the future may render this duplication unnecessary. It is proposed that the structure of the statistics data base be reviewed in line with the future review of observational data required for MARS archiving and the introduction of the new analysis system. Eventually a close coupling between the reports data base and the statistics data base would assist in accessing associated information. Such a coupling, together with a sound data base design, and enhanced by additional requirements covered later, would lead to a comprehensive data assimilation data base.