GARP TOPICS

No. 69 January 1982

An occasional series reporting on U.S. and international GARP scientific, technical, and planning activities, developments, and programs, presented as a public service to the meteorological community by the American Meteorological Society through arrangements with the U.S. Committee on the Global Atmospheric Research Program of the National Academy of Sciences–National Research Council. Opinions expressed in "GARP Topics" do not necessarily reflect the point of view of the U.S. Committee.

FGGE 4-Dimensional Data Assimilation at ECMWF

L. Bengtsson, M. Kanamitsu,[1] P. Kållberg, and S. Uppala

European Centre for Medium Range Weather Forecasts
Reading, United Kingdom

Abstract

The 4-dimensional data-assimilation system used to produce the FGGE level III–b data set at the European Centre for Medium Range Weather Forecasts (ECMWF) is described. The system consists of a three-dimensional multivariate optimum interpolation, a nonlinear normal mode initialization, and associated automatic system for data checking. A 15-level model with a horizontal resolution of 1.875° is used for the dynamical assimilation. The observations are the Main level II–b data, as specified in the FGGE data management plan. The quality of the observations, and in particular, those from the special observing systems must be regarded as very good. Only very few observations are regarded as incorrect and discarded. The ECMWF level III–b production was completed in June 1981, and the analyses are available in archives of the World Data Centers (WDCs). Global analyses 0000 GMT and 1200 GMT have been produced for all standard levels up to and including 10 mb. During the special observing periods, analyses have also been archived at 0600 GMT and 1800 GMT.

[1]Electronic Computation Center, Japan Meteorological Agency, Tokyo, Japan.

0003–0007/82/010029–15$03.75

1. Introduction

The Global Weather Experiment will undoubtedly be marked as a milestone in the development of meteorology. By all standards, it has been a great experiment as well as an unparalleled example of scientific cooperation. Of the four phases (planning, observational, data processing, and research), we now have entered the last two, namely, data processing and research. In this paper, we will describe the data assimilation system and the data processing system that have been used to produce the FGGE analyses (the III–b data) at ECMWF. In an accompanying paper (Bengtsson *et al.*, 1982), we will present some of the preliminary research results. The data management plan that was put into operation when the observational experiment started in December 1979 is described in Fig. 1. The chart shows the flow of the data from the level of the instrument signals, *level I;* its transformation into basic meteorological parameters, *level II;* and its final merging into a complete global set of basic meteorological parameters, the so-called II–b data set. The label "b" denotes all the data collected within a three-month

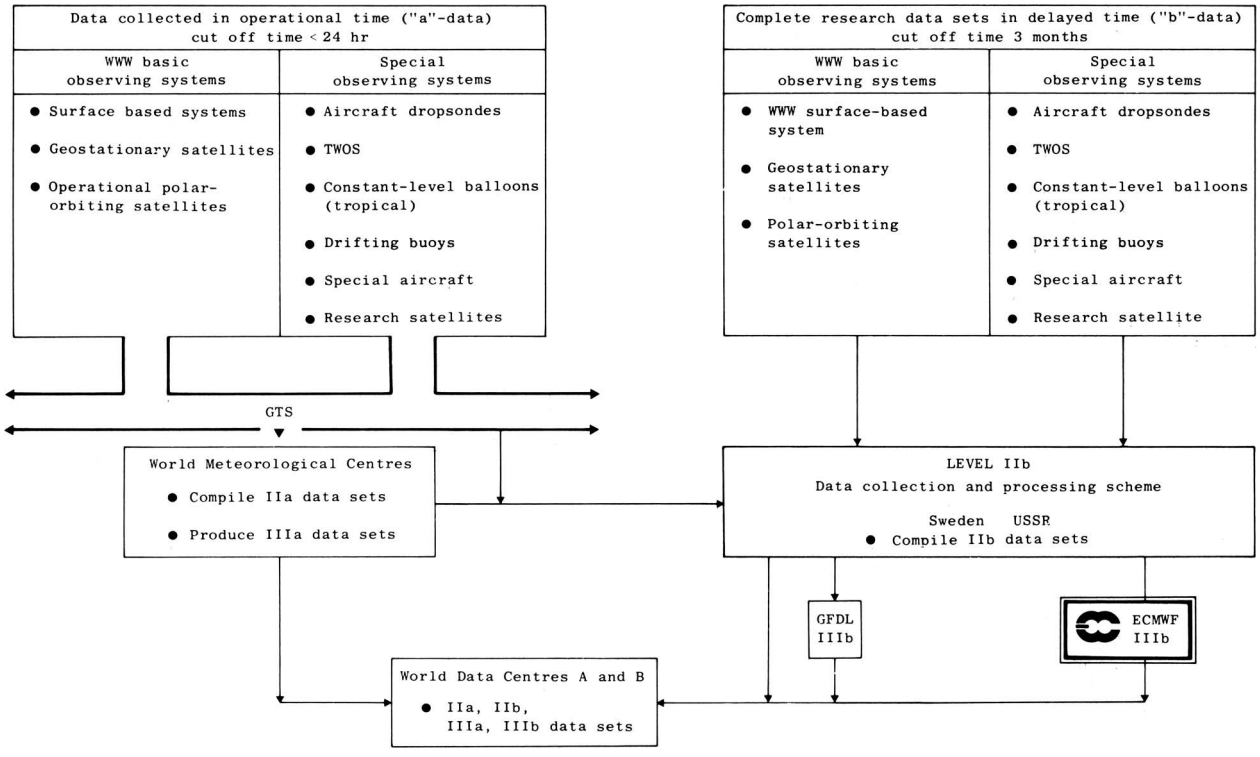


FIG. 1. The overall data management during FGGE.

cut-off period, and contains as a subset the operational data collected in real time. The real-time data are correspondingly indicated by label "a." The level II–b data constitute the basic meteorological product of the Global Weather Experiment. These data now have been checked, assembled, and delivered to the two WDCs, as well as to the two organizations, ECMWF and the Geophysical Fluid Dynamics Laboratory (GFDL), selected to undertake the task of analyzing the data and transforming them into meteorological fields, the so-called III–b data set. In Section 2, the data-assimilation system is presented. The corresponding data checking and data quality procedures are presented in Section 3.

The data processing of the FGGE data is a tremendous task. The total number of observations for the whole FGGE year (II–b data set) amounts to around 15×10^6 or 40000/day, where one complete radiosonde observation or one synoptic observation is counted as one observation. Stored on magnetic tapes, including all control information, this corresponds to around 4×10^9 bytes. After having transformed this data set into a set of analyzed and dynamically derived fields (III–b data set), the total amount of information is approximately the same, or 4×10^9 bytes. This information is stored on about 100 magnetic tapes with a packing density of 1600 bpi. During the course of the data processing, a large number of temporary files are being generated which are used in the control and validation of the data and in the research program at ECMWF. The essential elements of the data processing and its organization are described in Section 4. In Section 5, finally, we will give some examples of analyses selected in areas where, under normal conditions, very few observations are available.

2. The data assimilation system

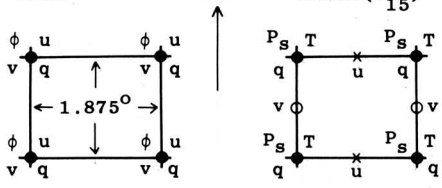
The data assimilation system used to produce the ECMWF FGGE analyses is practically identical to that used for the Centre's operations (Fig. 2). It is an intermittent data assimilation system using a multivariate optimum interpolation analysis, a nonlinear normal mode initialization, and a high-resolution forecast that produces a first estimate for the subsequent analysis. Data are assimilated in 6 h periods (Fig. 3). The analysis consists of two parts, one for a simultaneous analysis of surface pressure, geopotential, and wind, and another part for humidity.

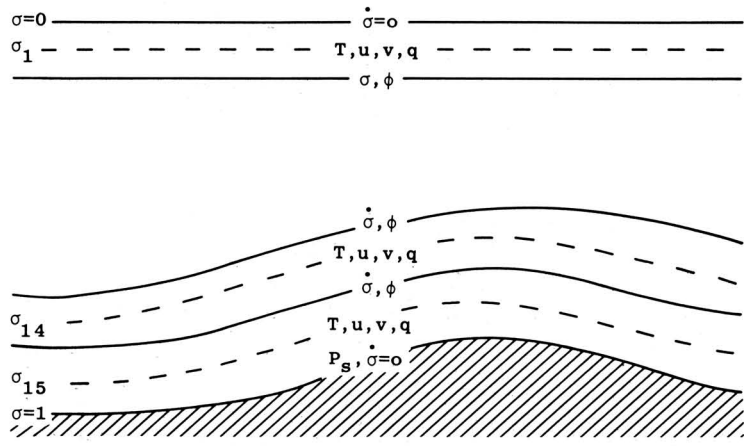
The basic idea of optimum interpolation (Eliassen, 1954; Gandin, 1963) is to determine, for a gridpoint, those interpolation weights that give the best fit of the analysis to the observations in a root-mean-square sense. These optimum weights not only will depend on the distribution of the observations, but also on the error characteristics of the observations and the first-guess forecast. In the ECMWF system, observed deviations from a first-guess forecast are analyzed (Rutherford, 1976). Using A to represent any scalar variable, and E to represent its estimated root-mean-square-error, the basic interpolation equation is

$$\frac{A_k^a - A_k^p}{E_k^p} = \sum_{i=1}^{N} W_{ki} \frac{A_i^o - A_i^p}{E_i^p}. \qquad (1)$$

Superscripts a, p, and o represent analysis, prediction, and observation, respectively. Subscript k denotes the gridpoint, and subscript i, $i = 1, \ldots N$ denote the N selected observations. W_{ki} are the unknown weights to be determined.

ANALYSIS	PREDICTION
ϕ, u, v, q (for $p \geq 300$)	T, u, v, q
p(mb)	σ
10	$0.025 (\sigma_1)$
20	0.077
30	0.132
50	0.193
70	0.260
100	0.334
150	0.415
200	0.500
250	0.589
300	0.678
400	0.765
500	0.845
700	0.914
850	0.967
1000	$0.996 (\sigma_{15})$





Vertical and horizontal (latitude-longitude) grids and dispositions of variables in the analysis (left) and prediction (right) coordinate systems.

ANALYSIS

Method	3 dimensional multi-variate (15-analysis levels, see above)
Independent variables	λ, φ, p, t
Dependent variables	ϕ, u, v, q
Grid	Non-staggered, standard pressure levels
First guess	6 hour forecast (complete prediction model)
Data assimilation frequency	6 hour ($\pm$ 3 hour window)

INITIALISATION

Method	Non-linear normal mode, 5 vertical modes, non-adiabatic

PREDICTION

Independent variables	λ, φ, σ, t
Dependent variables	T, u, v, q, p_s
Grid	Staggered in the horizontal (Arakawa C-grid). Uniform horizontal (regular lat/lon). Non-uniform vertical spacing of levels (see above).
Finite difference scheme	Second order accuracy
Time-integration	Leapfrog, semi-implicit (Δt = 15 min) (time filter ν = 0.05)
Horizontal diffusion	Linear, fourth order (diffusion coefficient = $4.5 \cdot 10^{15} m^4 s^{-1}$)
Earth surface	Albedo, roughness, soil moisture, snow and ice specified geographically. Albedo, soil moisture and snow time dependent.
Orography	Included, moderately smooth.
Physical parameterisation	(i) Boundary eddy fluxes dependent on roughness length and local stability (Monin-Obukov)
	(ii) Free-atmosphere turbulent fluxes dependent on mixing length and Richardson number
	(iii) Kuo convection scheme
	(iv) Full interaction between radiation and clouds
	(v) Full hydrological cycle
	(vi) Computed land temperature, no diurnal cycle
	(vii) Climatological sea-surface temperature

FIG. 2. ECMWF—Global Forecasting System, 15-level grid point model. Horizontal resolution 1.875° Lat./Long.

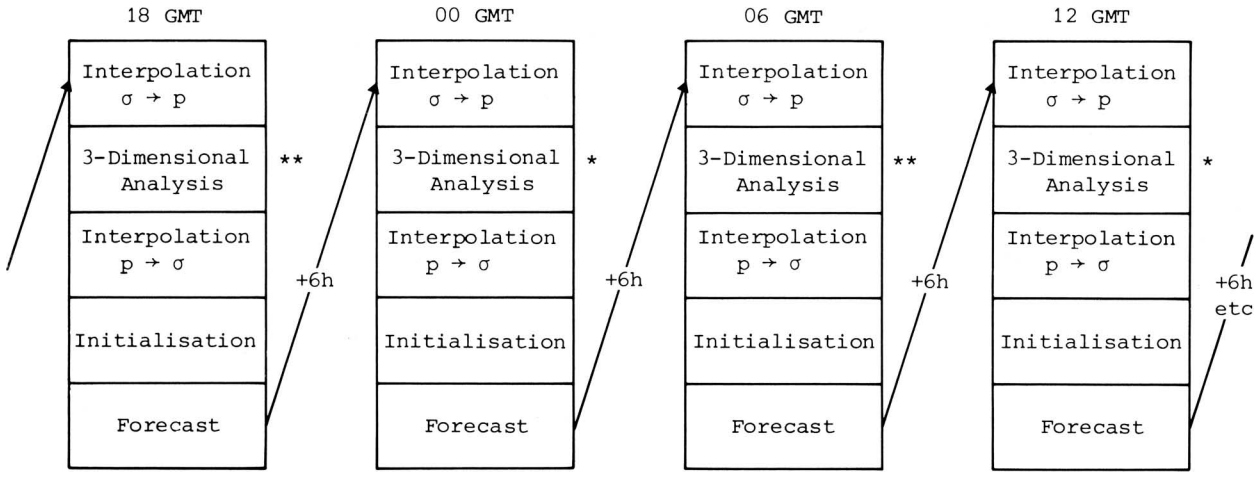


FIG. 3. The different stages in 4-dimensional data-assimilation at ECMWF (*archived through the whole FGGE year, **archived during the SOPs).

Introducing the normalized deviations from the so-called true value (superscript t)

$$\alpha_i^o = (A_i^o - A_i^t)/E_i^o$$
$$\alpha_i^p = (A_i^p - A_i^t)/E_i^p \qquad (2)$$
$$\alpha_k^a = (A_k^a - A_k^t)/E_k^a$$

and normalizing the observation and analysis errors

$$\epsilon_i^o = E_i^o/E_i^p$$
$$\epsilon_k^a = E_k^a/E_k^p \qquad (3)$$

we can derive an expression for the normalized analysis error for the gridpoint k

$$\epsilon_k^a \cdot \alpha_k^a = \alpha_k^p + \sum_{i=1}^{N} W_{ki}(\alpha_i^o \epsilon_i^o - \alpha_i^p). \qquad (4)$$

Squaring this expression, and taking the ensemble average over all k (denoted by $\langle\ \rangle$), the squared analysis error to be minimized can be written

$$\epsilon_k^{a^2} = 1 + 2\sum_{i=1}^{N} W_{ki}\{\langle \alpha_k^p \alpha_i^o \rangle \epsilon_i^o - \langle \alpha_k^p \alpha_i^p \rangle\} + \sum_{i=1}^{N}\sum_{j=1}^{N} W_{ki}\{\langle \alpha_i^p \alpha_j^p \rangle + \epsilon_i^o \langle \alpha_i^o \alpha_j^o \rangle \epsilon_j^o - \epsilon_i^o \langle \alpha_i^o \alpha_j^p \rangle - \langle \alpha_i^p \alpha_j^o \rangle \epsilon_j^o\} W_{kj}. \qquad (5)$$

The bracketed expressions, $\langle \alpha_k^p\ \alpha_i^p \rangle$, etc. denote the covariances between different types of errors (deviations from the truth) as defined by the sub- and superscripts.

It can be seen from Eq. (5) that the weights depend not only on the normalized observations errors, ϵ_i^o, but also on the covariances of the observation errors, $\langle \alpha_i^o\ \alpha_j^o \rangle$, the covariances for the prediction errors, $\langle \alpha_k^p\ \alpha_i^p \rangle$, and the cross covariances between observation and prediction errors, $\langle \alpha_i^o\ \alpha_j^p \rangle$, which are put equal to zero.

Minimizing the mean square error, Eq. (5), with respect to variations of the weights, W_{ki}, leads to a set of N linear equations for the N weights for each gridpoint, k.

$$\sum_{j=1}^{N} \{\langle \alpha_i^p \alpha_j^p \rangle + \epsilon_i^o \langle \alpha_i^o \alpha_j^o \rangle \epsilon_j^o\} W_{ki} = \langle \alpha_k^p \alpha_i^p \rangle \qquad i = 1, N \qquad (6)$$

The observational errors and error statistics used in the analysis scheme represent instrumental error, including radiation effect in the stratosphere, interpolation error in the radiosonde report, as well as sub-grid scale variations.

Table 1 shows the thickness errors (expressed as corresponding layer mean temperature errors) and wind errors for the different observing systems.

In the analysis of meteorological fields, it is of great importance to retain the meteorologically significant structures of the fields through assimilation—i.e., the new, observed, information has to be assimilated into the large and synoptic scale systems in a meteorologically realistic way (Daley and Puri, 1980). If this is not assured, a major part of the observed information may, when inserted, appear as high frequency gravity waves with no, or minor, prognostic value. In the ECMWF system, the observations are assimilated in a consistent way through the assumption of geostrophically balanced prediction error covariances between the geopotential height and the wind components. This causes the analyzed corrections to the first-guess forecast to be locally nondivergent and approximately geostrophic at high latitudes. Experiments using climatology as a first guess show that the large-scale features of the divergence field can be analyzed from the observations, although the detailed features have to be generated by the model itself. Hydrostatic balance is achieved through conversion of temperature observations into thicknesses, prior to the assimilation.

The prediction errors are estimated from the analysis errors of the previous analysis that are allowed to grow in time, according to a simple linear growth equation.

Equation (6) constitutes a set of linear algebraic equations that are solved by standard matrix inversion methods. In

TABLE 1. Observational errors for different observing systems.

Level (mb)	Temperature (°C)			Wind ($m\ s^{-1}$)			
		TIROS-N		Radiosonde pilot			Clouddrift wind
	Radiosonde	Clear/ partly cloudy	Microwave	ASDAR AIDS	NESS WISCONSIN	ESA LMD	HIMAWARI
10	4.5	2.8	2.8	6	8	8	13
20	3.8	2.6	2.7	6	8	8	13
30	3.2	2.5	2.6	6	8	8	13
50	2.7	2.4	2.5	6	8	8	13
70	2.3	2.2	1.4	6	8	8	13
100	2.1	2.0	1.6	6	8	8	13
150	2.1	2.0	1.7	6	8	8	13
200	2.0	1.9	1.8	6	8	8	13
250	1.8	1.9	1.9	6	8	8	13
300	1.6	1.8	2.0	6	8	8	13
400	1.5	1.8	2.2	5	7	8	10
500	1.2	1.7	2.2	4	7	8	10
700	1.1	1.8	2.5	3	5	8	6
850	1.1	2.0	3.9	2	4	7	6
1000	—	—	—	2	4	7	6

Sea surface pressure: SYNOP/SHIP 1.0 mb; buoy 2.0 mb; COLBA/DROPWINDSONDE/TWOS–NAVAID observation errors are calculated from the level II-b quality information. Temperatures given as layer means.

addition to the weights, W_{ki}, the solution provides a direct measure of the analysis error as follows from Eq. (4). In most optimum interpolation systems employed for operational analysis forecasts, the analysis is evaluated at every gridpoint and an elaborate data selection algorithm is used to select the N "best" observations for the gridpoint. N is usually of the order 5 to 10, due to computer limitations. At ECMWF, a different programming approach has been adopted to obtain the full benefit of the fast vector processing capability of the CRAY–1 computer. Hereby a set of gridpoints within a given atmospheric volume is analyzed simultaneously. The same weights then are used for all gridpoints within the volume, which is about 6° × 6° × 300 mb at the equator, and approximately the same at other latitudes. A detailed description of the ECMWF analysis scheme is found in Lorenc (1981).

For analyzing the layer mean water content, we use the correction method (Döös, 1967), and the observed deviations from the first guess are weighted according to the observation error and distance.

The basic purpose of an initialization procedure in the data assimilation is to ensure that the first-guess predictions are not contaminated with irrelevant gravity waves. Through the elimination of such waves, the observations are projected on to the slowly varying, meteorologically significant modes that constitute what has been named the "slow manifold" by Leith (1980). A very efficient way of eliminating high speed gravity waves, chosen for the ECMWF scheme, is the so-called nonlinear normal mode initialization, first proposed by Machenhauer (1977). In this method, the initial change of the gravity waves is put equal to zero, which gives an excellent, noise-free development. It also gives quite realistic initial divergences and surface pressure tendencies, at least at high latitudes. A separation of high frequency Rossby waves and low frequency gravity waves, which is impossible with traditional dynamical initialization schemes, now is done easily, and no specific time filter is needed to dampen gravity waves. Initialization is done in two iterations and with five vertical modes, using only adiabatic forcing. Dry adiabatic adjustment is performed before and after the initialization step. A description of the initialization scheme can be found in Temperton and Williamson (1981) and Williamson and Temperton (1981). This scheme performs well, except in the tropics, where the Hadley circulation is being partly suppressed, and where a five-mode pattern can be seen in the vertical profile of the divergence. Work is in progress to incorporate the nonadiabatic forcing that will reduce this error. It will be implemented in future FGGE data assimilation experiments at ECMWF.

The numerical model (Fig. 2), used for the first-guess predictions, is a semi-implicit gridpoint model using a finite difference scheme that conserves potential enstrophy during the vorticity advection by the nondivergent wind (Sadourny, 1975). The vertical and horizontal resolution (15 levels, 1.875° in both latitude and longitude) has been selected to describe barotropic and baroclinic instabilities, as well as the lower boundary conditions, with necessary accuracy. The parameterization of physical processes has been designed to describe the feedback loops assumed to be important for medium-range forecasting. It contains an interactive cloud-radiation coupling, surface hydrology, large-scale and convective precipitation, and a planetary boundary layer based on similarity modeling. A description of the model is found in Bengtsson (1980).

Since the initialization and the forecast model uses $\sigma = p/p_s$ as the vertical coordinate, while the analysis is carried out on pressure surfaces, vertical interpolation is necessary for the preparation of the first guess. The sigma to pressure level interpolation uses cubic splines. For pressure levels below the lowest σ-level, an extrapolation procedure is carried out. Observed mean sea level pressure deviations are converted to height deviations at the nearest analysis level, i.e., usually 1000 mb, using the first-guess temperatures. In the same way, the analyzed height increments are converted back to give the final p_{msl}. Elsewhere, the interpolation errors are usually negligible. However, recent experiments have

TABLE 2. The amount of data on 4 June 1979 (SOP II) and on 4 June 1980 as available at ECMWF from the Global Telecommunications System (GTS). The high number of pilot winds in the FGGE data is due to both TEMP and PILOT reports from the same ascent.

Datatypes used from II-b dataset			FGGE II-b 4 June 1979 SOP II GMT:0000	0600	1200	1800	GTS 4 June 1980 0000	0600	1200	1800
Rawinsonde data			767	81	766	153	549	53	540	71
Pilot wind data			515	440	583	320	184	281	179	209
TWOS NAVAID		data	8	9	7	10				
Aircraft Dropwindsonde		data	18	21	18	26				
Constant-level balloon		data	84	84	83	84				
		Total per 6 h	1392	635	1457	593	733	334	719	280
		Total per day				4077				2066
Aircraft data		(ASDAR)	164	150	83	186				
		(AIDS)	543	535	373	596				
		(AIREP)	859	945	658	876				
		Total per 6 h	1566	1630	1114	1658	356	563	252	295
		Total per day				5968				1466
Surface land	manual	(SYNOP)	2357	2428	2528	2398				
Ship	fixed	(SHIP)	27	26	23	19				
	mobile		1104	1074	1075	929				
		Total per 6 h	3488	3528	3626	3346	5042	5239	3762	3998
		Total per day				13988				18041*
Satellite sounding data										
cloudy (microwave)			424	347	433	425				
partly cloudy			282	195	181	209				
clear			1422	1321	1521	1480				
		Total per 6 h	2128	1863	2135	2114	493	348	322	502
		Total per day				8240				1665
PAOB on Southern Hemisphere			—	—	—	—				
		Total per 6 h	—	—	—	—	306	—	140	—
		Total per day				—				446
Satellite wind data										
NESS	high	level	170	—	85	124				
	low	level	349	—	432	428				
WISCONSIN	high	level	—	77	414	572				
	low	level	—	474	866	848				
LMD	high	level	—	—	107	—				
	low	level	—	—	311	—				
METEOSAT	high	level	194	—	250	—				
	low	level	91	—	269	—				
HIMAWARI	high	level	232	—	205	—				
	low	level	77	—	187	—				
	high	level per 6 h	596	77	1061	696	351	—	251	234
	low	level per 6 h	517	474	2065	1276	229	—	209	349
	high	level per day				2430				836
	low	level per day				4332				787
Drifting buoy data			445	510	530	504				
Environmental buoys			56	81	81	57				
Other			55	40	42	29				
		Total per 6 h	556	631	653	590	109	125	90	176
		Total per day				2430				500

*These numbers are misleading, since the II-b SYNOP data were selected according to a reduced stationlist in data-dense regions.

shown that some improvement, in particular in the humidity analysis, is obtained when only the observed deviations from the first guess, i.e., the analysis parameters themselves, are interpolated to the pressure levels, instead of to the full fields. In order to maintain a consistent set of FGGE analyses throughout the whole period, the recent modification has not been implemented in the III–b production.

3. Data-checking and data quality

a. Data-checking

For the level III–b production at ECMWF, the so-called Main FGGE Level II–b dataset has been used. This dataset contains all observations collected in accordance with the FGGE data management plan, and with a delayed cut-off

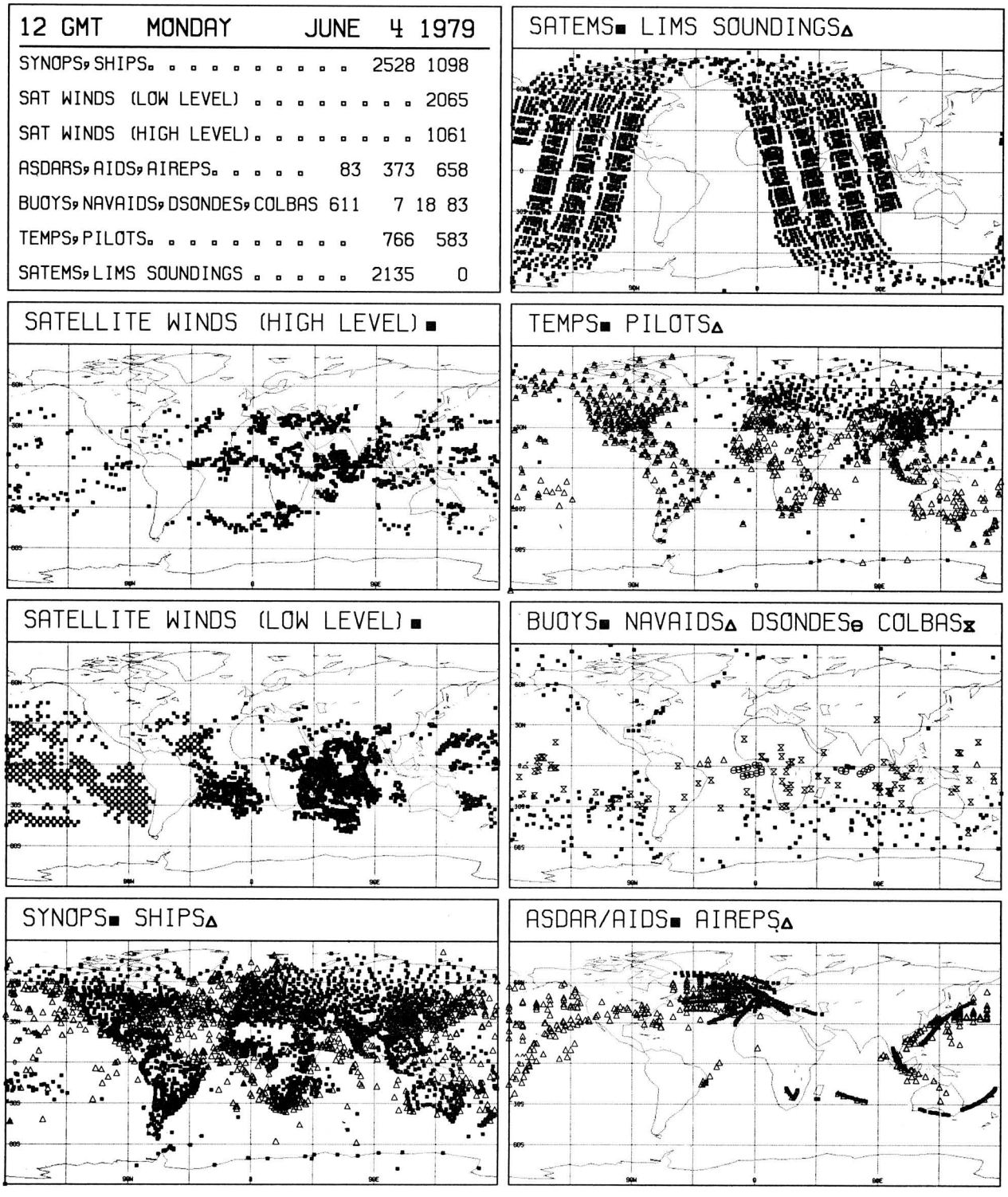


FIG. 4. FGGE level II–b data distribution. 4 June 1979 12 GMT ±3 h.

time of three months. Observations received at the level II–b centers later, as well as certain additional data, mainly from the regional experiments (Winter MONEX, Summer MONEX, and the West African Monsoon Experiment (WAMEX)), have not been used by ECMWF. This means that the observational data base used for the ECMWF III–b data assimilation is exactly defined, and that comparative analyses using other assimilation systems can be done easily.

The main level II–b data, received from the Special and

Space Based Observation Center in Sweden, is formally checked and converted to ECMWF-internal format. The data also are sorted into 6° × 6° boxes, in which the analysis equations are being solved. At the same time, data coverage maps are prepared, and the first quantitative check is made. Table 2 shows as an example the data available in the level II-b data base for 4 June 1979. As a comparison, the data available at ECMWF via the Global Telecommunications System (GTS) exactly one year later also are included in the table. The maps in Fig. 4 show the distribution of the II–b data in the 6 h period centered at 1200 GMT, 4 June 1979.

The observations have gone through quality control procedures in the level II–b centers. The quality control indicators provided by them are used as initial values for a quality indicator on each observed value, zero for correct, one for probably correct, two for probably incorrect, and three for an incorrect value. This indicator is modified in the data checking procedures of the analysis scheme, if necessary. At each stage in the analysis, data indicating three are discarded, data indicating two are rechecked (if possible) but not used, and data indicating one or zero are used.

Before the data enter the pre-analysis, they go through the following tests, in which data can be deleted or modified:

1) location check; (–)
2) duplicate check; (275)
3) elimination of ships over land; (5)
4) elimination of bad drifting buoys; (13)
5) elimination of microwave soundings over land and in tropics; (83)
6) elimination of bad microwave soundings through rainy clouds using clear radiance data; (46)
7) correction of 100 mb surface pressure errors using first-guess; (–)
8) modifying the assigned cloud wind heights (see Section 3b) (24)
   a) according to first-guess temperatures if temperature is reported;
   b) according to tropopause if only pressure is reported.

The numbers in parentheses show how many observations were deleted or modified in these tests at 0000 GMT 4 June 1979.

After finishing these preliminary checks and modifications, the main data checking procedures are performed in the data-assimilation cycle. The first check is a comparison against the first guess. Each value is assigned a quality indicator greater than zero if its squared deviation from the first guess is greater than a predetermined multiple of its estimated variance. The datum then is assigned the maximum of this value and the indicator from the preprocessing checks. Comparison with nearby observations is done in a neighbor check. Pairs of observations are searched within a certain distance, and a check of redundant information is performed level by level for each analysis variable. If the check fails, the quality indicator is increased by one. Neighboring observations containing redundant information are combined into averaged observations.

The mass and wind analysis program selects data and sets up and solves the analysis equations for analysis volumes. The data checking is done by calculating the difference between each observed value and a local analysis where the particular value has not been used. Values deviating more than a prescribed tolerance are successively rejected in descending order. In the example from 4 June, 0.28% were listed as probably incorrect (quality level 2), and 0.33% as incorrect (quality level 3).

After the data-checking procedures are completed, a quality control summary is made. It contains information on the final quality indicator, and in what stage of the data checking it was set. As an example, Table 3 shows how the indicators were set for the heights from all the rawindsondes at 0000 GMT 4 June on selected levels.

b. Data quality

During the processing of the FGGE at ECMWF, the quality of level II–b data has been found to be generally very good. This is true in particular for the special FGGE observing systems, which show excellent quality. In the following, the main quality features of the level II–b data are summarized. Upper-air data quality has been good throughout the production. The II–b quality information is used to give the initial value for the quality indicators on each observed value. Additionally, an algorithm was developed to estimate the observation error for those observing systems (aircraft, dropwindsonde, TWOS–NAVAID, and constant-level balloon) that make use of the OMEGA wind-finding system. Aircraft data, especially those of AIDS/ASDAR, have been of high quality. Occasionally, errors in date/time occur in the AIDS data.

Microwave soundings from TIROS–N have to be treated carefully in the assimilation system. In an evaluation of the TIROS–N soundings carried out at the National Meteorological Center (NMC); Phillips *et al.* (private communication) found that the root mean square (rms) deviations between derived temperatures from the microwave and infrared channels were quite large in the lower layers. The same also was found at ECMWF during the "FGGE End-to-End test." The problems occur mainly in ocean areas with heavy rain and over land. Following the practice at the NMC (Phillips, private communication), TIROS–N microwave data were eliminated from the assimilation in the following three areas: 1) over land, 2) between 20°N and 20°S, and 3) over ocean areas with heavy precipitation. The latter can be identified from the reported radiances in two microwave channels. The infrared soundings from TIROS–N were used everywhere. Humidity information from TIROS–N was not used.

Experience from the "FGGE End-to-End test," and from other centers (e.g., NMC) indicated that the vertical temperature profile radiometer (VTPR) temperatures suffered from aging instruments already at the beginning of FGGE. It was thus decided not to use the data from this satellite for the level III–b production. However, in December 1978, when TIROS–N data were not available, VTPR soundings were used at the southern hemisphere.

Cloud driftwinds from geostationary satellites suffer from bad height assignments, mainly due to the unavailability of good temperature/height profiles to the data producer. In most cases, climatology, or even a standard atmosphere, has

TABLE 3. Quality control summary for all heights from rawinsonde data after data checking on 0000 GMT 4 June 1979 on selected levels. Both the frequency and percentage of setting of the final quality indicator (0, 1, 2, or 3) in the preprocessing (q.c), the comparison against the first-guess (f-g), the comparison with nearby observations (obs), and the data-checking analysis (dca) are given. No frequency is given, for the data found erroneous in the preprocessing stage (NA).

	Level	Final flag set frequency				Final flag set %				
	(mb)	q.c	fg	obs	dca	q.c %	fg %	obs %	dca %	Sum
(0)	1000	81	—	394	87	13.85	—	67.35	14.87	96.07
Final	850	11	—	—	668	1.57	—	—	95.29	96.86
Quality	500	13	—	1	695	1.80	—	0.14	96.26	98.20
Indicator	300	12	—	—	674	1.71	—	—	95.87	97.58
	200	12	—	—	632	1.83	—	—	96.64	98.47
	50	6	—	—	490	1.19	—	—	96.84	98.02
(1)	1000	—	—	1	3	—	—	0.17	0.51	0.68
Final	850	—	—	—	8	—	—	—	1.14	1.14
Quality	500	—	—	—	4	—	—	—	0.55	0.55
Indicator	300	—	—	—	7	—	—	—	1.00	1.00
	200	—	—	—	3	—	—	—	0.46	0.46
	50	—	—	—	3	—	—	—	0.59	0.59
(2)	1000	—	—	—	2	—	—	—	0.34	0.34
Final	850	—	—	—	1	—	—	—	0.14	0.14
Quality	500	—	—	—	2	—	—	—	0.28	0.28
Indicator	300	—	—	—	1	—	—	—	0.14	0.14
	200	—	—	—	2	—	—	—	0.31	0.31
	50	—	—	—	2	—	—	—	0.40	0.40
(3)	1000	NA	1	6	10	NA	0.17	1.03	1.71	2.91
Final	850	NA	4	—	9	NA	0.57	—	1.28	1.85
Quality	500	NA	4	—	3	NA	0.55	—	0.42	0.97
Indicator	300	NA	6	—	3	NA	0.85	—	0.43	1.28
	200	NA	2	—	3	NA	0.31	—	0.46	0.76
	50	NA	3	—	2	NA	0.59	—	0.40	0.99

been used. With the availability, in the 4-dimensional data assimilation, of a high quality first-guess (and assuming that the reported cloud top temperatures are better than the reported cloud heights), an algorithm to reassign the cloud wind heights from all data producers was designed. The algorithm contains two steps:

1) Any cloud reported above the first-guess tropopause is moved to the reported cloud top temperature below the tropopause. If the cloud temperature is not available, the cloud is moved to the tropopause level.
2) Any cloud below the tropopause but above 600 mb, which has a reported cloud top temperature, is moved to the best fitting temperature level in the first guess. Displacements larger than 5000 m are not allowed.

In a pilot study at ECMWF, Julian and Kanamitsu (private communication) tested the procedure, using colocated (3° × 3°, ± 3 h) rawindsonde data as a control. In a small sample of six 6 h periods, containing 48 colocated reports, the mean vector wind difference between the rawinsonde and the cloud wind at its reported level was 15.4 m s^{-1}. After reassignment to the level of best temperature fit in the first guess, the vector difference was reduced to 12.5 m s^{-1}, corresponding to an improvement of about 20%.

In the example from 00 GMT 4 June, the following number of cloud wind height modifications were performed on a total of 596 reported cloud winds above 600 mb.

Height raised > 40 mb, > 30°N or > 30°S	(11)
Height raised > 40 mb, between 30°N and 30°S	(3)
Height lowered > 40 mb, > 30°N or > 30°S	(7)
Height lowered > 40 mb, between 30°N and 30°S	(3)

Certain fleets of cloud drift winds are sometimes assigned to completely erroneous heights. Also, many high-latitude cloud winds seem to be inaccurate when compared with aircraft and rawinsonde data.

Buoy data have made it possible to analyze small-scale intense weather systems, and both the quantity and quality are relatively high. Sometimes, however, a buoy can have a systematic pressure error of 20–30 mb but still have a normal daily variation.

During the assimilation of the Main Level II–b dataset, certain data problems were encountered. Most of them are easily detected and corrected by a careful user and should not cause any problems. Surface ship data suffer from a relatively high frequency of erroneous positions. Duplicates can occur within some data sources, particularly SYNOP, TEMP, and SATEM. Some other errors are wrong units in a few TEMP soundings and 100 mb errors in surface pressure reports. As has been mentioned earlier, the level II–b data are provided with quality indicators. Cases have been encountered when these indicators were probably incorrect, and users should not rely completely on the level II–b quality information.

In Table 4, the number of rejected height and wind data during the period 17 May 1979–13 September 1979 is shown for different data sources. Each single value rejected in the

TABLE 4. The data rejected (each single datum counted) in the pre-analysis and in the data checking analysis during the period 17 May 1979–13 September 1979. The data flagged erroneous in the level II-b data and the data discarded in the preprocessing phase (e.g., SHIPS over land) are not included.

	1000–700 mb	700–250 mb	250–70 mb	70–10 mb	1000–700 mb	700–250 mb	250–70 mb	70–10 mb
90°N–30°N								
	Height				Wind			
Rawinsonde	3753	3632	3310	2918	4165	2437	1514	725
Pilot wind	—	—	—	—	2080	664	324	172
TWOS-NAVAID	—	—	—	—	—	—	—	—
Dropsonde	—	—	—	—	—	—	—	—
COLBA	—	—	—	—	—	—	2	—
ASDAR	—	—	—	—	6	3	4	—
AIDS	—	—	—	—	—	876	94	—
AIREP	—	—	—	—	2	1309	674	1
SYNOP manual	18918	505	5	—	574	—	—	—
SYNOP autom.	—	—	—	—	—	—	—	—
Surface-SHIP fixed	285	—	—	—	9	—	—	—
Surface-SHIP mobil	17534	—	—	—	1102	—	—	—
Surface-environmental buoy	9	—	—	—	2	—	—	—
Cloud wind	—	—	—	—	14	42	26	—
Drifting buoy	132	—	—	—	—	—	—	—
	Thickness							
Rawinsonde	157	2076	1247	199				
SATEM	78	24	84	72				
30°N–30°S								
	Height				Wind			
Rawinsonde	2131	2605	2804	1160	1475	978	1128	370
Pilot wind	—	—	—	—	2939	730	494	98
TWOS-NAVAID	79	56	89	106	6	—	4	1
Dropsonde	—	—	—	—	79	7	—	—
COLBA	—	—	—	—	—	—	187	—
ASDAR	—	—	—	—	22	9	5	—
AIDS	—	—	—	—	—	59	31	—
AIREP	—	—	—	—	5	252	142	—
SYNOP manual	24667	14	15	1	1733	—	—	—
SYNOP autom.	2	—	—	—	—	—	—	—
Surface-SHIP fixed	—	—	—	—	—	—	—	—
Surface-SHIP mobil	12144	—	—	—	768	—	—	—
Surface-environmental buoy	61	—	—	—	1	—	—	—
Cloud wind	—	—	—	—	91	54	31	—
Drifting buoy	934	—	—	—	—	—	—	—
	Thickness							
Rawinsonde	340	3607	6117	422				
Dropsonde	8	18	0	0				
SATEM	57	25	45	18				
30°S–90°S								
	Height				Wind			
Rawinsonde	317	176	145	92	509	399	286	50
Pilot wind	—	—	—	—	431	260	125	3
TWOS-NAVAID	—	—	—	—	—	—	—	—
Dropsonde	—	—	—	—	—	—	—	—
COLBA	—	—	—	—	—	—	40	—
ASDAR	—	—	—	—	8	2	—	—
AIDS	—	—	—	—	—	26	—	—
AIREP	—	—	—	—	—	59	31	—
SYNOP manual	2169	2	—	—	427	—	—	—
SYNOP autom.	—	—	—	—	—	—	—	—

TABLE 4. *continued*

	1000–700 mb	700–250 mb	250–70 mb	70–10 mb	1000–700 mb	700–250 mb	250–70 mb	70–10 mb
Surface-SHIP fixed	—	—	—	—	—	—	—	—
Surface-SHIP mobil	1140	—	—	—	138	—	—	—
Surface-environmental buoy	—	—	—	—	—	—	—	—
Cloud wind	—	—	—	—	19	79	23	—
Drifting buoy	2371	—	—	—	—	—	—	—
	Thickness							
Rawinsonde	25	744	435	2				
SATEM	71	157	124	203				

pre-analysis and in the data-checking analysis is counted, excluding those data that are erroneous, according to the level II–b quality control information.

In spite of these problems, the general quality of the level II–b data is, in our opinion, very good, and the dataset constitutes an invaluable source of information for future scientific work.

4. The data processing system

The configuration of the computing system at ECMWF comprises three levels—the CRAY–1, a powerful vector processing computer, front-ended by a CDC CYBER 175, which has a wider range of software facilities to control the runs. This, in turn, is front-ended by smaller computers to control communications. There is a link between CRAY–1 and CYBER 175, which facilitates the submission of jobs and data to and from the CRAY–1. In the following, the main data processing features are described.

a. General

The development of the data processing system for the level III–b data production into this computer environment was done in parallel with the development of the ECMWF operational data-assimilation system. The major difference was that the level III–b system was run in "batch" environments, leaving the full control and responsibility of the runs to the FGGE scientists. To make the control easier and more flexible, many options that allow the monitoring functions to vary from one analysis time to another have been built into the job control language. The options include post-processing of files, plotting of maps, and different types of diagnostics. The system has been designed on the basis that the manual intervention by the monitoring scientists during the actual computer process should be minimized. Instead, the monitoring tasks have been concentrated on the quality control of the products, in the form of plotted maps on selected areas and in diagnostics. Three data bases are created during the production:

1) observational data base;
2) data base for post-processed files,
3) data base for raw σ- and p-coordinate files.

Most of the files are archived at each analysis time and, as archiving medium, magnetic tapes of 6250 bpi density are used. The number of tapes used is about 2000 for the FGGE year.

b. Jobs and files in the assimilation cycle

The analysis, initialization, and forecast steps in the data-assimilation are individual jobs with files as interfaces. These nonpacked purely model- or analysis-generated sigma, or pressure coordinate files, are available online during the cycle. For back-up and archiving purposes, they are transferred through the link to magnetic tapes. These tapes then can be used in possible restarts of the assimilation, and as initial conditions in forecast experiments.

Previous to the analysis, the σ-coordinate forecast is interpolated to the p-coordinate analysis levels and variables to prepare the first guess. The programs and files in the analysis step are briefly described in Table 5. At the end of the analysis, interpolation from the p-coordinate analysis levels and variables is done to the forecast σ-levels and variables.

The nonlinear normal mode initialization then is performed on the σ-level analysis. A stabilizing step is done before the initialization, since the analysis scheme does not check the vertical stability of the analyzed temperature and humidity profiles. The same also is done after the initialization, in case any unstable profiles have been reintroduced. The 6 h forecast then is produced with the grid-point model.

During one assimilation cycle, data are archived and post-processed according to Table 6. The diagnostics used in the monitoring of the quality of the analysis are calculated after the archiving process. The observations that are deleted or modified before pre-analysis, as well as those accepted, averaged, or rejected during the data-checking procedures,

TABLE 5. The main functions in the analysis step.

Preprocessing of observations	Input:	Level II-b observations
		Sorting into boxes Checking Modifying Deleting Plotting
	Output:	Modified II-b observations
Mass and wind analysis 10 mb 20 30 50 70 100 150 200 250 300 400 500 700 850 1000	Input:	First-guess on p-levels
	Preanalysis	Input: Modified II-b observations Estimated observation errors Estimated forecast errors
		Extracts all available data for the analysis input levels and variables from observations, first-guess and estim. first-guess errors in all the boxes. Converts observations to normalized increments, on which the preliminary data checking is done. Searching of pairs of observations with redundant information and forming of averaged observations.
		Output: **Random access workfile ****Normalized increments and observation errors. Information on observations, e.g., quality indicator
	Data-checking analysis	Input: Random access workfile from preanalysis Estimated forecast-error correlations Estimated observation error correlations
		Each datum selected for use in the final analysis is checked against an analysis using the other selected data. Workfile updated accordingly.
	Gridpoint analysis	Input: Random access workfile from pre- and data-checking analysis
		The normalized analysis increments and errors are calculated for all grid points and variables. The matrix inverses calculated during the data-checking analysis are used.
		Output: Normalized analysis increments Normalized analysis errors
	Postanalysis	The increments are added to the first guess to give the mass and wind analysis, which then is used as the first-guess in the humidity analysis.
		Output: Analysis Analysis increments
Humidity analysis 300 mb 400 500 700 850 1000		Input: Mass and wind analysis Modified II-b observations Observation errors Estimated first-guess humidity errors
	Preanalysis	What is done in preanalysis of mass and wind, is now done for humidity variables.
	Gridpoint Analysis	The integrated specific humidity of layers upto 300 mb are analysed two dimensionally.
	Output:	Analysis on p-levels

are plotted for each analysis time. The fit of observations to the analysis is followed by plotting of maps, especially in data-sparse areas and in the tropics. Also, maps from the post-processed initialized analyses are produced, and the derived surface pressure tendencies are compared with the synoptic development. Vertical cross sections from the uninitialized σ-coordinate analysis are produced daily to show the static stability and the zonal wind between selected grid points. A powerful tool in the monitoring has been the calculated rms-differences between the observations and first guess, analysis, and the post-processed initialized analysis. Only those observations that are used in the final grid

TABLE 6. The files archived internally at ECMWF and those sent to the World Data Centers A and B.

	Archived at ECMWF Unpacked at (p or σ)	Packed (p)	WDC–A WDC–B (p)
Analysis			
First-guess, p	x	x	
Analysis, p	x	x	z,u,v,p_{msl}
Analysis errors	x		x
Analysis, σ	x		
Quality information on observations deleted-modified observations		x	x
Initialization			
Initialized analysis, σ	x	x	T, ω, r
Forecast			
Forecast, σ	x		
Forecast errors	x		
Physics fields	x	x	

point analysis are included in the calculations. The quality of the products then can be examined against different observing systems and over different geographical areas.

c. *The ECMWF level III–b dataset*

The ECMWF level III–b dataset based on the Main level II–b data will be archived in the WDCs for the whole FGGE year. Within the Special Observing Periods (SOPs), archiving is done every 6 h, and outside SOPs, every 12 h. The analyses are available on 15 levels (1000 to 10 mb) and with a horizontal resolution of 1.875° latitude/longitude. The dataset contains the basic analysis fields: height, horizontal wind components, and mean sea level pressure. In addition to this, temperature, relative humidity (r), and vertical velocity are included (Table 7). The temperature field has been calculated from initialized heights and surface pressure, and the vertical velocity from the initialized winds. Finally, the relative humidity has been obtained from precipitable water analyses. The wind analyses are not initialized, and thus the height and wind fields are *not* dynamically balanced. A catalog showing a selection of daily global analysis maps for the FGGE year is being prepared at ECMWF (Bjørheim *et al.*, 1981).

TABLE 7. The fields archived in the ECMWF level III-b dataset.

	z	u	v	T	ω	r
10 mb	x	x	x	x	x	
20	x	x	x	x	x	
30	x	x	x	x	x	
50	x	x	x	x	x	
70	x	x	x	x	x	
100	x	x	x	x	x	
150	x	x	x	x	x	
200	x	x	x	x	x	
250	x	x	x	x	x	
300	x	x	x	x	x	x
400	x	x	x	x	x	x
500	x	x	x	x	x	x
700	x	x	x	x	x	x
850	x	x	x	x	x	x
1000	x	x	x	x	x	x
	p_{msl}					

5. Examples of analyses

In this section, we will show a few examples of synoptic analyses from the III–b production. The purpose is to give an impression of the ECMWF analyses and their fit to the observations. A more comprehensive discussion of the FGGE results, and examples of ongoing research activities using the FGGE data, can be found in Bengtsson *et al.*, 1982.

a. *Intense cyclonic activity around Antarctica*

The two maps in Fig. 5 show a typical intense polar front cyclone from July of 1979. In addition to the analyzed and observed surface pressure, Fig. 5a also shows the surface pressure tendency, calculated from the continuity equation using the initialized wind analysis as input. The initialization scheme is described in Section 2. High latitude divergences obtained in this way are very realistic, and the calculated isallobaric fields generally agree well with observations over data dense areas. (In low latitudes, however, the initialization tends to weaken the divergence fields).

All the surface pressure observations have not been accepted in the analysis. The rejected observations are, in most cases, obviously in error. The buoy at 32°E/66°S reporting 952 mb may be questionable, but is most likely wrong; it behaved erratically throughout the month of July.

A composite wind, temperature, and moisture analysis is shown in Fig. 5b. Since there are no wind observations in this area, the wind field is analyzed from the pressure and SATEM observations only by the multivariate analysis scheme. The humidity field is almost entirely carried through from the first-guess forecast, and the equivalent potential temperature shown in the map is calculated from the hydrostatic temperatures of the analyzed geopotential field and mixing ratios from the first guess forecast.

It can be seen that the maps, in spite of the fact that the analyzed weather system has been obtained almost entirely from drifting buoy surface pressure data and TIROS–N temperature profiles, show many of the features of a typical polar front cyclone, as originally described for the northern hemisphere by the Bergen school. This allows us to be fairly optimistic about the possibilities of an automated observing system like the one operated during FGGE, to provide reliable synoptic analyses for these areas in the future.

b. *Equatorial upper troposphere*

Animated films prepared from GOES-East and GOES-West infrared images indicate a large variability in the equatorial upper troposphere flow. There are many examples of intense, low latitude cold core vortices, and comparatively strong

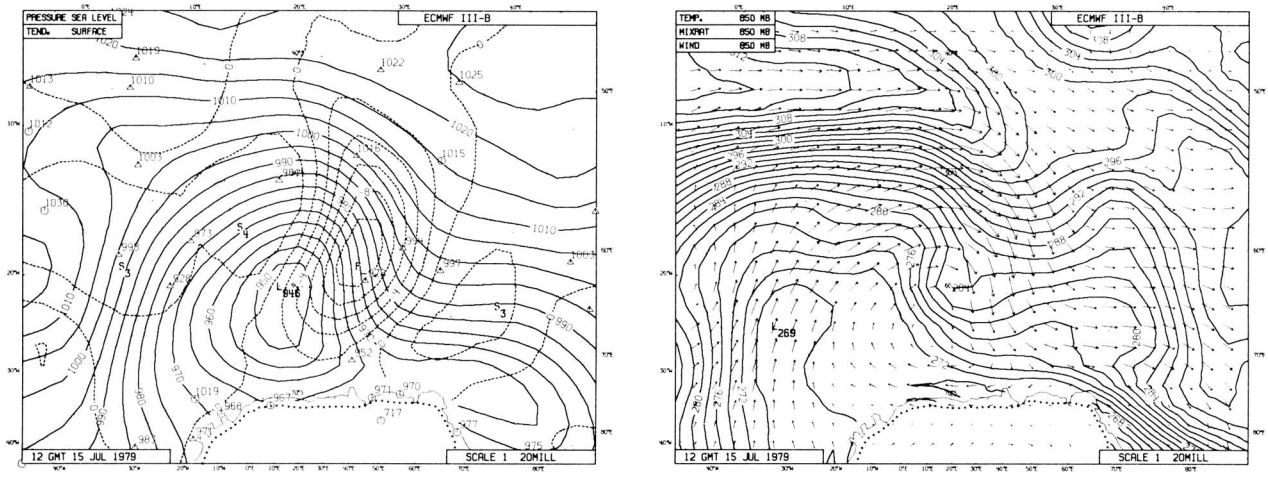


FIG. 6. 200 mb streamline analysis 00 GMT 24 January 1979. East Pacific Ocean. Dashed contours—isotachs. Areas with wind speeds above 20 m s^{-1} are stippled.

cross-equatorial flows. The good coverage of tropical upper air wind data in the FGGE level II–b dataset makes it possible to analyze these features with considerable detail. Figure 6 shows streamlines and isotachs over the eastern Pacific Ocean on 24 January. The comparatively strong southwesterly flow from 20°S/100°W to 20°N/60°W was a

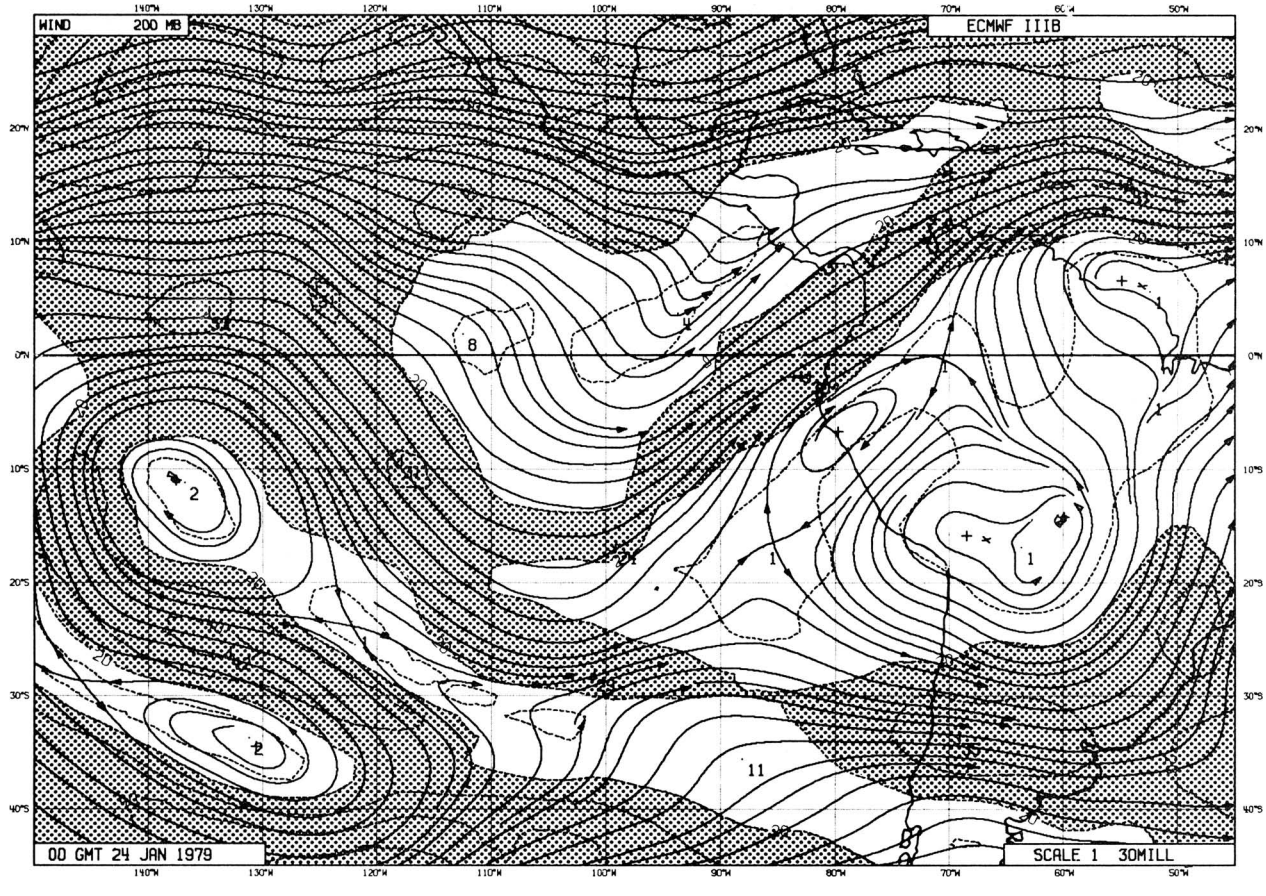


FIG. 5. a) Surface observations of pressure, and level III–b analysis. 12 GMT, 15 July 1979, south Atlantic and Indian oceans. Full contours—isobars, dashed contours—isallobars. △—drifting buoy, ⌀—SYNOP/SHIP. b) 850 mb wind (arrows) and equivalent potential temperature (full lines). Area and time as Fig. 5a.

persistent phenomenon during the latter part of January.

The strong vortex around 10°S/140°W was one in a series of such vortices cut off from the subtropical jet stream in the Southern Hemisphere in the summer of 1978–79.

6. Conclusions

The 4-dimensional data assimilation system developed for operational global weather prediction at ECMWF has been used for analyzing the observations from FGGE. It has been proven to be a very versatile and efficient system, and the data assimilation has been carried out in a quasi-operational mode. It thereby has been clearly demonstrated that an observing system of the kind used during FGGE, consisting of a mixture of conventional synoptic observations and nonsynoptic observations from satellites, buoys, aircraft, and drifting balloons, can be used for operational numerical weather prediction on an operational basis. The quality of the special observing systems was found to be up to expectations. Major difficulties in the analyses have occurred in situations with gross errors in the satellite observations and in areas where no other independent observations have been available. The height assignment of high-level satellite winds can be wrong by a considerable amount, and a systematic way of height correction has to be implemented for these observations.

The considerable improvement in the data coverage in the tropics and at the Southern Hemisphere during FGGE has made it possible to analyze in great detail weather systems that previously had been hardly possible. Further, a considerable synoptic activity could be demonstrated in the tropics, often connected to a strong interhemispheric exchange. In a second contribution to GARP Topics, the ongoing FGGE research work will be reported (Bengtsson *et al.*, 1982). This report will be comprised of diagnostic investigation, predictability studies, and evaluation of the different observing systems during FGGE.

References

Bengtsson, L., 1980: European Centre for Medium Range Weather Forecasts (ECMWF), WMO/JSC, Catalogue of numerical atmospheric models for the FIRST GARP GLOBAL EXPERIMENT, Part I: Analysis and forecast system, pp. 287–340.

——, M. Kanamitsu, P. Kållberg, and S. Uppala, 1982: FGGE research activities at ECMWF. To be published in *Bull. Am. Meteorol. Soc.*

Bjørheim, K., P. Julian, M. Kanamitsu, P. Kålberg, P. Price, S. Tracton, and S. Uppala, 1981: FGGE III–B Daily Global Analyses. European Centre for Medium Range Weather Forecasts, Reading, U.K.

Daley, R., and K. Puri, 1980: Four-dimensional data assimilation and the slow manifold. *Mon. Wea. Rev.*, **108,** 85–89.

Döös, B. R., 1967: Numerical analysis of meteorological data. *Lectures on Numerical Short Range Weather Prediction,* Regional Training Seminar, Moscow, 17 November–14 December 1965. WMO, Geneva.

Eliassen, A., 1954: Provisional report on calculation of spatial covariance and autocorrelation of the pressure field. *Acad. Sci.,* Institute for Weather and Climate Research, Report No. 5, Oslo, Norway.

Gandin, L. S., 1963: Objective analysis of meteorological fields. Translated to English by the Israel Program for Scientific Translations from Russian, *Gidrometeoizdat,* Leningrad, 1961.

Leith, C., 1980: Nonlinear normal mode initialization and quasi-geostrophic theory. *J. Atmos. Sci.,* **37,** 958–968.

Lorenc, A., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

Machenhauer, B., 1977: On the dynamics of gravity oscillations in a shallow-water model, with application to normal mode initialization. *Beitr. Phys. Atmos.,* **50,** 253–271.

Rutherford, I. D., 1976: An operational 3-dimensional multi-variate statistical objective analysis scheme. *The GARP Programme on Numerical Experimentation Rep. 11,* WMO/ICSU, Geneva.

Sadourny, R., 1975: The dynamics of finite-difference models of the shallow water equations. *J. Atmos. Sci.,* **32,** 680–689.

Temperton, C., and D. Williamson, 1981: Normal mode initialization for a multi-level gridpoint model. Part I: Linear aspects. *Mon. Wea. Rev.,* **109,** 729–743.

Williamson, D., and C. Temperton, 1981: Normal mode initialization for a multi-level gridpoint model. Part II: Nonlinear aspects. *Mon. Wea. Rev.,* **109,** 744–757.