User Guide to ECMWF forecast products

Anders Persson

Designed, edited and printed by ECMWF

Preface

This meteorological User Guide is not like most other “user guides”, which provide clear and straightforward instructions how to “plug in”, “get started”, “execute” and “switch off”. Nor is this Guide a handbook in NWP, dynamic meteorology or weather forecasting; these objectives are also already served by existing literature. The aim of the Guide is to facilitate the use of conventional ECMWF medium range forecast products and encourage the use of newer, more advanced products such as the wave forecasts, seasonal forecasts and forecasts from the Ensemble Prediction System. It tries to do this by providing:

1. A brief description of the forecast system. This will include numerical techniques, the characteristics of the forecast model, the data assimilation and analysis and the Ensemble Predictions System (EPS). If possible, the strong and weak sides of the components are addressed already at this stage. In a separate chapter, the full list of disseminated products is given, together with suggestions how to solve meteorological problems related to the retrieval and plotting of meteorological fields.

2. *An introduction to forecast verifications and their interpretation*. Although the statistical verifications as such are objective, any final conclusion drawn from them is in the end subjective in the sense that it depends on what one is trying to achieve. What looks like “good statistics” might sometimes be due to a rather bad forecast performance; a genuine improvement of a forecast system might show up as an increase in errors. It is therefore also important to realize the distinction between “error”, “skill” and “usefulness”.

3. *Recommendations on how to interpret and make use of the NWP products.* Experience shows that unawareness of basic principles of interpreting NWP products, in particular medium-range forecasts, can seriously worsen the meteorological value of the final forecast. Experience shows that forecasters *can* add substantial value, provided they know where to direct their efforts. For automatic use of the forecasts products, methods for statistical interpretation or adaptation are briefly presented.

In this User Guide much emphasis is laid on the Ensemble Prediction System. It is in the literature sometimes portrayed as a new, revolutionary way of making weather forecasts, when it is rather a logical development of traditional weather forecasting, which has always been about trying to tell what is *most likely* to happen, what *might* happen and what will *probably not* happen. What is new with the

EPS is that for the first time this uncertainty can be evaluated in an objective and consistent way.

What is also new with the EPS is that it provides an overwhelming amount of information and offers an almost unlimited combinations of products. While the forecasters in the past had to make sense out of thousands of SYNOP, SHIP, TEMP and PILOT reports, coming out from the telecommunication lines to provide the best weather forecast service, they will in the future have to make sense out of 10-100 times this amount of information, coming out from the computers. It is a challenge to the forecaster to convey the relevant parts of the EPS information to the end-customers or the public, taking the system’s shortcomings into account.

Even if the forecasters cannot always improve on the predictive information, they can take comfort from the fact that however much the NWP systems improves, the public will always credit *them* with all the good forecasts.

Acknowledgement: This Guide is the fruit of several years of discussions with scientists, working at EMWF as Staff Members, Consultant or visiting scientists and, most importantly, meteorological forecasters on the ECMWF Training Courses, workshops and during Member State visits. It has been the interaction with these two specialized groups, trying to draw the scientific consequences of the forecasters’ experiences, trying to see the practical implications of the scientific exploration, which has been the main driving force and inspiration for this work.

1. The European Centre for Medium-Range Weather Forecasts - an historical background

The European Centre for Medium Range Weather Forecasts (ECMWF) is the consequence of 100 years of development in dynamic and synoptic meteorology, and fifty years of development in numerical weather prediction (NWP). This time has seen periods of optimism vary with pessimism, during which a slow but steady improvement of the quality of the weather forecasts has been made

1.1 The early history of Numerical Weather Prediction

A century ago, in 1904, the Norwegian hydrodynamist V. Bjerknes suggested that the weather could be quantitatively predicted by applying the complete set of hydrodynamic and thermodynamic equations to carefully analysed initial atmospheric states. Lacking both the theoretical and practical means to make any quantitative predictions he initiated instead the qualitative approach that has became known as the "Bergen School".

After the Second World War two technological developments appeared to make mathematical weather forecast along the lines suggested by Bjerknes possible: the establishment of a hemispheric network of upper–air stations and the development of the first electronic computers. In 1948 a young meteorological theoretician, Jule Charney, succeeded to derive simplified mathematical models of the atmospheric motions, based on the quasi–geostrophic approximations. These equations would be able to forecast the large scale flow in spite of minor inaccuracies in the initial analyses.

When the first NWP experiments were conducted in 1950, due to the limited computer capacity, only the most simple of Charney’s models could be used, the barotropic equation of atmospheric motion. The results were surprisingly successful: the general 500 hPa flow pattern over North America was forecast 24 hours in advance with greater skill than previous subjective methods.

From this successful start two different strategies developed: countries with limited computer resources, preferred to explore the potential of the barotropic model, whereas countries like the US and Britain took a more ambitious approach by developing baroclinic models where forecasts of vertical motion were possible. It soon turned out that the nature of the problem was much more complicated than envisaged. That is why during the 50’s the first operationally useful NWP forecasts were barotropic: in Sweden in 1954, in the US in 1958 and Japan in 1959. Only in 1962 could the US launch the first operational quasi–geostrophic baroclinic model, followed by Britain in 1965. By that time, work was already under way, to introduce more realistic numerical models, based on the primitive equations (PE).

In a PE-model changes in wind and geopotential fields are not restricted by any quasi-geostrophic constraint, but are allowed to interact freely. The physical parametrizations such as convection, which are difficult to handle in the quasi–geostrophic model, could now be realistically incorporated, so that the tropical regions, essential for forecasts over Europe beyond two or three days, can be included. The first global PE model began operating in 1966 at NMC Washington, with a 300 km grid and six–layer vertical resolution. During the 70’s several other PE models were implemented, global, hemispheric or as Limited Area Models, which ran with a higher resolution over a smaller area and took boundary values from a larger hemispheric or global model.

Interest in ocean wave forecasting started during the Second World War when it was realised that information on the sea state could be of vital importance. The first operational predictions were based on the use of empirical wind sea and swell laws. An important advance was the introduction of the concept of a wave spectrum in the mid 1950’s, followed by a dynamical equation describing the evolution of the wave spectrum, the energy balance equation.

During the 1980’s it became evident that wave forecasts did not only have an intrinsic value, but that they also provide a means for increased realism of the atmospheric system through incorporating the friction the waves except on the wind, which on its turn affects the ocean circulation and the storm surge. Ultimately, it is expected to have a model consisting of the atmosphere and the oceans where the ocean waves are the agent that transfer energy and momentum across the interface in accordance with the energy balance equation. Presently, we have taken the first step by coupling the IFS atmospheric model with the wave model in a two-way interaction mode.

This coupled model provides the 10 day weather and wave forecast since the 29th of June 1998. As a next step ECMWF is developing a coupled atmosphere, ocean-wave, ocean-circulation model. This coupled model will be used in seasonal forecasting and monthly forecasting in the near future.

With the increasing number of satellites providing observations also from the upper atmosphere, the atmospheric models have been extended to ever higher altitudes. One of the major breakthroughs in the last 15 years in NWP came from an enormous improvement in data assimilation techniques together with the availability of an increasing number of remotely sensed observations from satellites, providing a global and high frequency data coverage. The development of variational techniques has progressively allowed for a direct assimilation of infrared and microwave sounder radiances which impact on analysed temperature and humidity fields. This technique also ensures that the information coming from satellites is dynamically consistent.

Recent studies have shown that in terms of NWP performance, satellite observations are now equally important as radiosondes, not only in the Southern Hemisphere (void of conventional observations), but also in the Northern Hemisphere

1.2 The creation of ECMWF

From the experience gathered with short–range and climatological simulations, there was, in the late 60’s, enough know–how to motivate an attack on the medium–range forecast problem, defined as the interval from 3 to 10 days ahead. The scientific and technical problems were still formidable, and only few countries had enough expertise to tackle them. This made medium–range forecasting an ideal candidate for multi–national co–operation. When a PE model began operating in the USA in 1966, there were moves in Europe to build up a similar system.

In October 1967 the Council of Ministers of the European Communities adopted a resolution to implement a programme to promote joint scientific and technical research. A proposal for a "European Meteorological Computer Centre for Research and Operations" occupied the first place on a list of meteorological projects submitted by an expert group in April 1969. The proposal was accepted and other European nations were invited to participate. In April 1970 an expanded expert group initiated two study groups to look into the economic and scientific motivations for the project.

The reports from the two groups were completed in August 1971, and at the conference of ministers in the same year it was decided to create the European Centre for Medium–Range Weather Forecasts. The ambition, laid out in the plans, was to produce forecasts ten days ahead with the five–day forecasts having the same accuracy as subjective two–day forecasts in the 50’s.

The ECMWF convention was signed in October 1973. Seventeen European States are currently members: Belgium, Denmark, Germany, Spain, France, Greece, Ireland, Italy, the Netherlands, Norway, Austria, Portugal, Switzerland, Finland, Sweden, Turkey and the United Kingdom. The objectives of the Centre were laid down as follows:

- To develop dynamic models of the atmosphere with a view to preparing medium–range weather forecasts by means of numerical methods;
- To prepare, on a regular basis, the data necessary for the production of medium–range weather forecasts;
- To carry out scientific and technical research directed towards the improvement of these forecasts;
- To collect and store appropriate meteorological data;
- To make available to the meteorological offices of the Member States, in the most appropriate form, the results of the studies and research provided for in the first and third objectives above and the data referred to in the second and fourth objectives;
- To make available a sufficient proportion of its computing capacity to the meteorological offices of the Member States for their research, priority being given to the field of numerical forecasting. The allocation of the proportions would be determined by Council;
- To assist in implementing the programmes of the World Meteorological Organization;
- To assist in advanced training for the scientific staff of the meteorological offices of the Member States in the field of numerical weather forecasting.

Since 1979 cooperation agreements have been concluded with Iceland, Hungary, Croatia, Slovenia, Czech Republic, WMO, EUMETSAT and ACMAD.

The first operational forecast was produced on 1 August 1979. Every day ECMWF makes a forecast to ten days ahead, and distributes it from its computer system to the systems of the national meteorological services of its Member States via a dedicated telecommunication network.

1.3 The ECMWF forecasting model since 1979 – an overview

The ECMWF forecasting system consists of three components: a general circulation model (coupled with an ocean wave model), a data assimilation system and, since 1992, an ensemble forecast system.

The first ECMWF numerical model was a grid–point model with 15 levels in the vertical up to 10 hPa. The horizontal resolution was 1.875 degrees of latitude and longitude, corresponding to a grid length of 200 km on a great circle. In April 1983 this grid–point model was replaced by a model with a **spectral representation** in the horizontal with a triangular truncation at wave–number 63. The spectral technique was more accurate than the grid point model for the same computational cost. With today’s high resolutions with both grid point and spectral models there is no longer any significant difference in accuracy. The semi-lagrangian technique (see 2.1.3) removes one source of difference between the two methods. The number of levels in the vertical was increased to 16. In May 1985 the spectral truncation was extended to wave–number 106. The number of levels was increased to 19 in 1986.

In September 1991, a **high resolution spectral model** was put into operations, where the spectral truncation was extended to wave–number 213 and the number of levels increased to 31 (Simmons et al, 1991). The model used a computational grid with a resolution of about 60 km. In 1998 the horizontal resolution was increased to wave number 319 linear truncation and in the following year the number of levels increased to 50 with the highest at 0.1 hPa. Later in 1999 the vertical resolution in the PBL was increased, giving a total of 60 levels in the model. In autumn 2000 the resolution was increased to T511.

The spectral technique was introduced operationally The global grid contains 8,300,760 points in all three dimensions. At each of these grid points, the meteorological variables are re–calculated every 20 minutes out to ten days ahead. The total number of computations amounts to about 52 x 10^{12} and with the current Fujitsu VPP 700 takes approximately 1 hour 35 minutes for the forecast.

Until 1995 the ECMWF model did not contain any explicit clouds, only interpretations from other fields like relative humidity, precipitation, vertical motion and vertical temperature gradients. A new **cloud scheme** was introduced in April 1995 with clouds as prognostic parameters, defined through the cloud fraction and the content of cloud liquid water and cloud ice. Ozone was added as a predicted variable in 1999.

Up to 1996 the **analysis system** was based on optimum interpolation. That year it was replaced by a three dimensional variational system (3DVAR), which was upgraded to a four dimensional variational system (4DVAR) in 1997.

The **wave model** that is used for ocean wave forecasting at ECMWF is the WAM model, developed during the 1980's. The WAM model is the first model that solves the complete energy balance equation, including the computationally expensive non linear interactions. A global version of the model became operational at ECMWF in 1992, followed after a few months by a Mediterranean implementation. In June 1998 the wave model was integrated into the atmospheric model allowing two–way interaction of wind and waves. At the same time ensemble prediction of ocean waves started. A big stimulus for developing the WAM model was provided by the advent of remote sensing techniques for measurements of the ocean surface by means of microwave instruments. Assimilation of altimeter data was introduced in the global version of the wave model in August 1993.

In 1992 the ECMWF started its **Ensemble Prediction System**. In autumn 1996 the number of members was extended from 32 to 50 members and the model was upgraded from T63 to T_L159, in autumn 2000 to T_L255. The vertical resolution was increased from 31 to 40 levels in 1999. Crude allowance for the uncertainty of physical processes was made in autumn 1998 with the introduction of stochastic physics.

From 1998 the first **seasonal forecasts** were issued on an experimental basis

Improvements in medium range forecasts

Since the ECMWF was established there has been an almost doubling of the range of useful forecasts. Whereas the limits of useful deterministic forecasts for the hemispheric surface wind and pressure is almost a week, it increases to one or two days for in the troposphere, in the stratosphere beyond ten days.

Table 11: Summary of the development in large scale NWP, 1950-2000

	Type of Model	Computer performance (MIPS)	Dynamic skill (days)	Numerical technique	Resolution	Parametrization	Model output
1950’s	Baro-tropic, regional	0.01	1-2 (baro-tropic develop-ments)	Finite difference	300 km, 1 level		500 hPa height
1960’s	Baro-clinic, quasi-geos-trophic, hemi-spheric	1	2-3 (baro-clinic develop-ments)		150-300 km, 2-5 levels	Simple topography, land/sea, moisture	1000 & 500 hPa height and thickness
1970’s		10	4-5 (large scale flow)	Semi-implicit, Finite difference Spectral methods Semi-Lagrang-ian	100-150 km, 6-10 levels	Convection, cloud, radia-tion, fric-tion, diffusion Real clouds Fog	Most atmos-pheric param-eters, incl. 2m T, 10m wind, clouds, rain, snow, show-ers Ozone Ocean waves
1980	Primitive equations Coupled models, EPS	50-100	5-6 (block-ings and cut-offs)		50-100 km, 10-20 levels 15-50 km 30-50 lev-els		
1990’s 2000		> 500	6-7 (up to 10 days in the strato-sphere)				

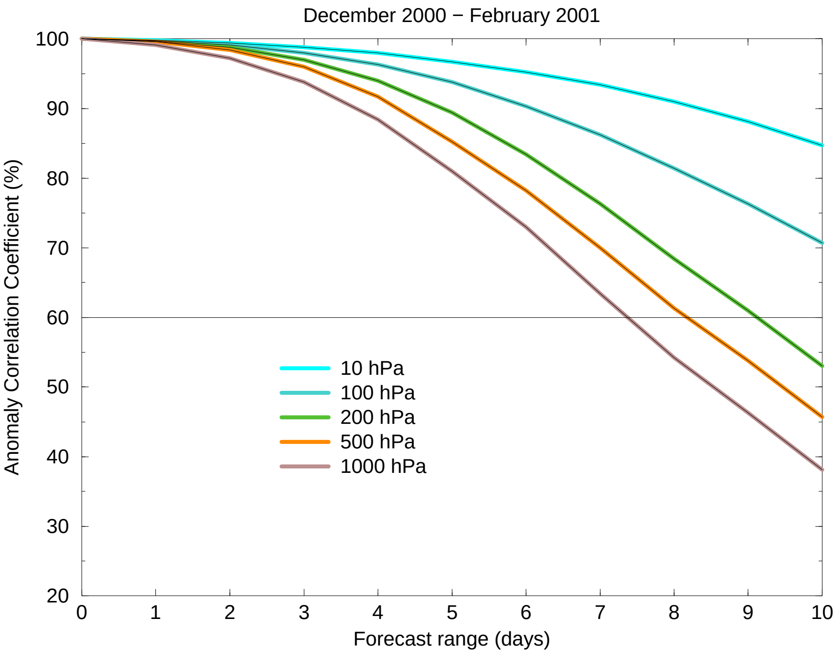


Figure 1 : Anomaly Correlation Coefficient for ECMWF forecast for different levels over the Northern Hemisphere winter 2000-2001.

Even when the stratospheric flow occasionally undergo sudden transitions the ECMWF forecasts this accurately ten days in advance.

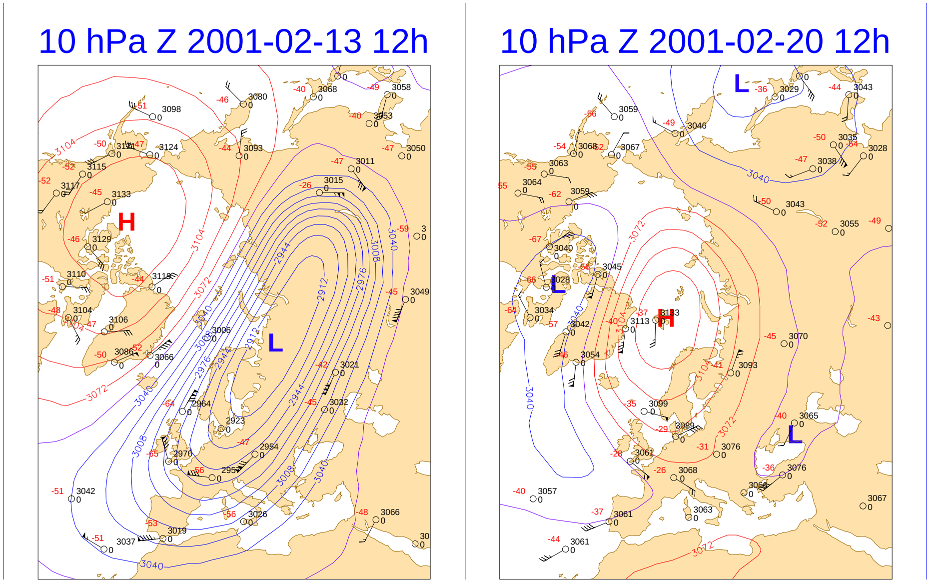


Figure 2 : During a week in February 2001 a sudden “stratospheric warming” took place on the Northern Hemisphere when the polar vortex was replaced by a anticyclonic flow with temperatures increasing by 25-35 degrees.

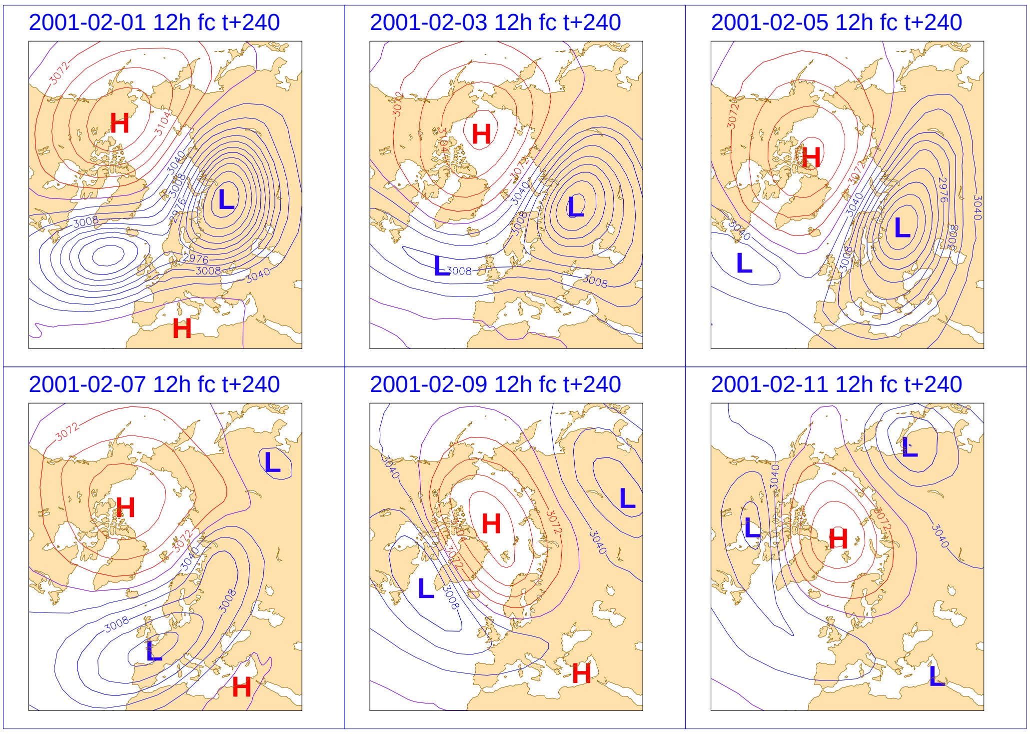


Figure 3 : D+10 ECMWF forecast at level 10 hPa from 1, 3, 5, 7, 9 and 11 February 2001 valid 11, 13, 15, 17, 19 and 21 February.

The late 1990's has seen a slow, but steady improvement of the ECMWF deterministic forecasts. Complemented with the Ensemble Prediction System there are now resources available to the forecasters to issue useful forecasts well up to a week in advance, with further indications of possible evolution up to ten days.

2. The ECMWF global atmospheric model

The ECMWF general circulation model, T_L511L60, consists of a dynamical component, a physical component and a coupled ocean wave component.

2.1 The model formulation

The model formulation can be summarized by six basic physical equations, the resolution in time and space and the way the numerical computations are carried out.

2.1.1 The model equations

Of the six equations governing the ECMWF primitive equation atmospheric model, two are diagnostic and tell us about the static relation between different parameters:

- The GAS LAW gives the relation between pressure, density and temperature.
- The HYDROSTATIC EQUATION shows the relationship between the density of the air and the change of pressure with height.

The other four equations are prognostic and describe the changes with time of the horizontal wind components, temperature and water vapour content of an air parcel, and of the surface pressure.

- The EQUATION OF CONTINUITY expresses the mass conservation and determines the vertical velocity and change in the surface pressure.
- The EQUATION OF MOTION describes how the momentum of an air parcel changes due to the pressure gradient and the Coriolis force. Included are also the effects of turbulent drag and gravity wave breaking
- The THERMODYNAMIC EQUATION expresses how a change in an air parcel temperature is brought about by adiabatic cooling or warming due to vertical displacements. Other physical processes like condensation, evaporation, turbulent transport and radiative effects are also included
- The CONSERVATION EQUATION FOR MOISTURE assumes that the moisture content of an air parcel is constant, except for losses due to precipitation and condensation or gains by evaporation from clouds and rain or from the oceans and continents. Adding to this there are specific prognostic equations for the cloud fraction, water, ice content and ozone.

Latent heat release, radiation from the sun and the earth’s surface and frictional or turbulent processes (diffusion) which are governed by the basic equations are, due to their small scale, described in a statistical way as a parametrization process (see 2.4).

2.1.2 The resolution in time and space

The present system uses a **temporal resolution** of 15 minutes. The computational time step has to be chosen with care in order to avoid numerical instabilities and ensure enough accuracy. The **vertical resolution** (measured in geometric height) is highest in the planetary boundary layer and lowest in the stratosphere and lower mesosphere.

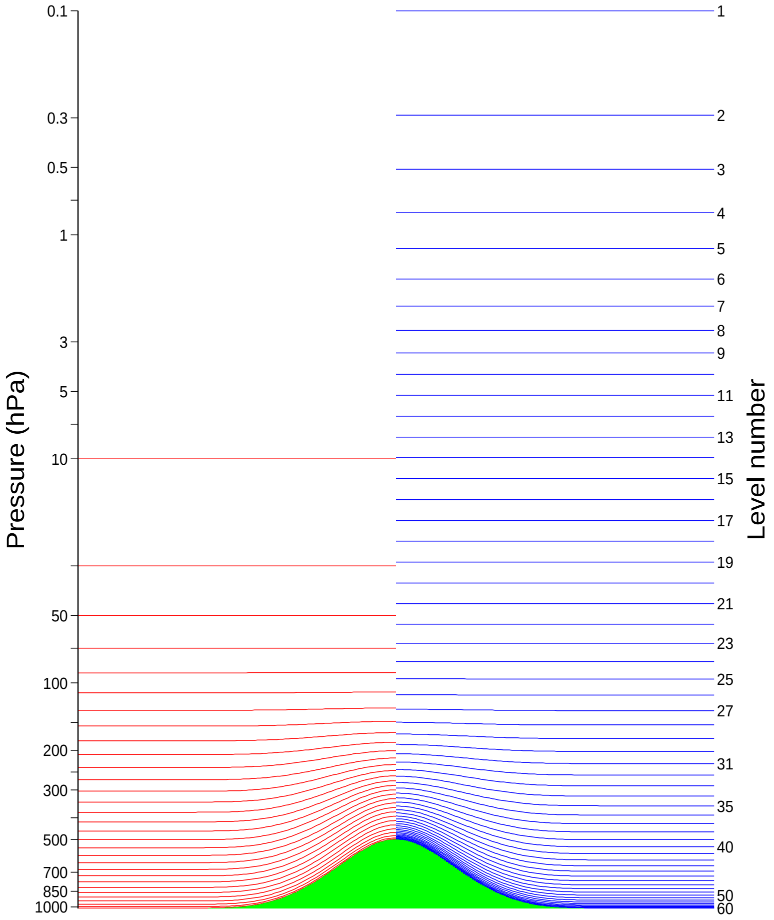


Figure 4 : To the left the vertical resolution before 1999 (31 level to 10 hPa), and to the right after1999 (60 levels to 0.1 hPa)

The atmosphere is divided into 60 layers up to 0.1 hPa (about 64 km). These so called σ–levels which follow the earth's surface in the lower and mid-troposphere are used as vertical coordinates but are surfaces of constant pressure in the upper stratosphere and mesosphere. A smooth transition between these types of levels is ensured (Untch et al, 1999)

For its **horizontal resolution** the ECMWF model uses two different numerical representations:

A *spectral method*, based on a spherical harmonic expansion, truncated at total wave number 511, for the representation of upper air fields and the computation of the horizontal derivatives. Apart from the operational T_L511L60 model (511 spectral components and 60 levels), a T_L255L40 is run for ensemble predictions (only up to 10 hPa), a T_L159L40 for the 4DVAR assimilations and T63L31 for seasonal forecasts.

Table 12: Pressure of model levels when the surface pressure is 1015 hPa

	(hPa)		(hPa)		(hPa)
1	0.1	21	44	41	577
2	0.3	22	55	42	616
3	0.5	23	67	43	654
4	0.8	24	80	44	691
5	1.2	25	96	45	728
6	1.6	26	113	46	763
7	2.1	27	133	47	797
8	2.7	28	154	48	828
9	3.4	29	177	49	857
10	4.2	30	202	50	884
11	5.2	31	229	51	908
12	6.4	32	257	52	930
13	8.0	33	288	53	949
14	9.8	34	320	54	965
15	12	35	353	55	979
16	15	36	388	56	989
17	19	37	425	57	998
18	23	38	462	58	1004
19	29	39	500	59	1009
20	36	40	538	60	1012

In addition there is a *grid point representation* used for computing dynamic tendencies and the diabatic physical parametrization. This so–called Gaussian grid, is regular in longitude and almost regular in latitude (Hortal and Simmons, 1991). Due to the convergence of the longitudes toward the poles, the east–west distance between the grid points decreases poleward. To avoid some numerical problems around the poles, but most importantly to save computing time, a reduced Gaussian grid was introduced in 1991 by reducing the number of grid points along the shorter latitude lines near the poles, so as to keep the east–west separation between points on different latitudes almost constant. With the current resolution the grid is identical to a regular Gaussian grid between 24N and 24S.

The model surface is logically divided into sea and land points, by using a **land–sea mask**. A grid point is defined as a land point if more than 50% of the actual surface of the grid-box is land. With a T_L511 resolution, islands like Corsica, Crete and Cyprus are represented by around five land grid points, Mallorca and Gotland by only two. The Faeroe Islands, the Shetland Island and Rhodos are not represented by any land point. The only minor inland lake which is represented by sea points is Vanern in southern Sweden.

2.1.3 The numerical formulation

The choice of a semi-Lagrangian numerical scheme instead of an Eulerian is the result of partly the need to save computer time and speed up the forecast. The basic difference between an Eulerian and a Lagrangian formulation can be seen from the equation (in a one-dimensional space):

$$\frac{dQ}{dt} = \frac{\partial Q}{\partial t} + U\frac{\partial Q}{\partial x} = 0$$

which in an Eulerian way expresses that the local changes in Q are due to the advection of Q by the wind U:

$$\frac{\partial Q}{\partial t} = -U\frac{\partial Q}{\partial x}$$

or in a Lagrangian way that Q is conserved for a given parcel:

$$\frac{dQ}{dt} = 0$$

In a pure Lagrangian framework (following a set of marked fluid parcels) shear and stretching deformations tend to concentrate parcels inhomogeneously, so that it is difficult to maintain uniform resolution over the forecast region. A *semi–Lagrangian* scheme is used to overcome this difficulty. In this version, the grid points are stationary and at each time step the scheme computes a backward trajectory from every grid point. The point reached defines where the air parcel was at the beginning of the time step. The interpolated value of the variable in that point is then carried forward to the grid point, applying the various physical processes.

Whereas all Eulerian schemes require small time steps to avoid numerical instability, (the quantity Q must not be advected by more than one grid length per time step), the semi-Lagrangian scheme allows longer time steps. The limitation for stability is that the trajectories do not cross (a parcel can not "overtake" another one). In the present two–time–level scheme the movement of the parcels is assumed of constant acceleration, not straight lines. Tests have shown that a semi–Lagrangian timestep can be at least fifteen times longer than the Eulerian without becoming unstable.

2.2 Parametrization of physical processes

The primary function of the forecast weather parameters in the ECMWF model, lies in their impact on the overall atmospheric flow. A ten–day integration makes it absolutely necessary to include effects with relatively long time scale, even as subtle as the evaporation by vegetation, in order to handle the flow pattern more accurately. The different time scales and feed–back mechanisms between the various processes makes the computations extremely complex and expensive.

These processes are mainly related to small scale disturbances in space and time, smaller than the scales explicitly resolved by the model, from convective clouds down to molecular processes. The effect that these subgrid scale processes have on the larger scales can be computed only by parametrization, i.e. formulating indirectly their overall effect in terms of known grid scale variables.

2.2.1 The model orography

The representation of the **orography** uses the mean orography and four additional fields describing the standard deviation, orientation, anisotropy and slope of the sub–grid orography. This takes some account of the orographic variability, but does not change the fact that for the usefulness of the weather parameters, the model orography is still significantly smoother than reality (see 5.7.3).

However, the parametrization allows a realistic representation of the mountain drag, which is important for the creation of large scale atmospheric eddies. A novel and important part of the scheme is that, depending on dynamical criteria, it can block the low level flow rather than make the air go over the orography.

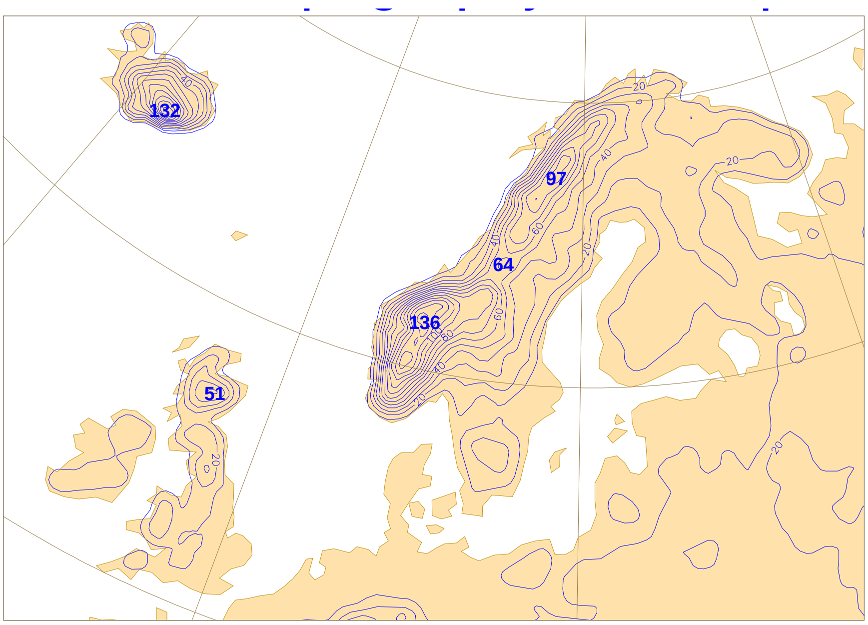


Figure 5 : The model height (in dekameters) for northern Europe

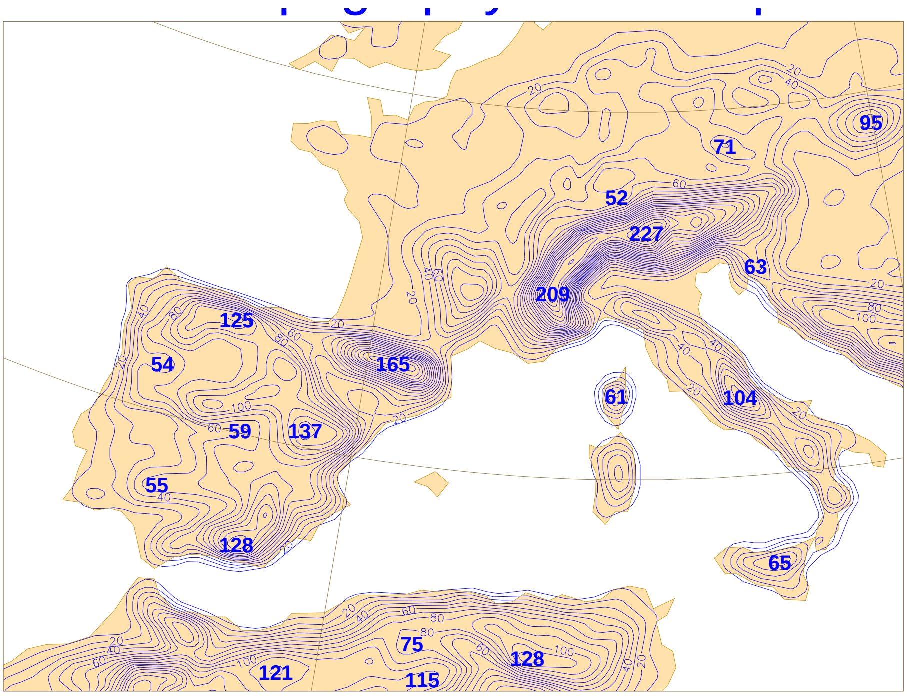


Figure 6 : The model height (in dekameters) for southwestern Europe

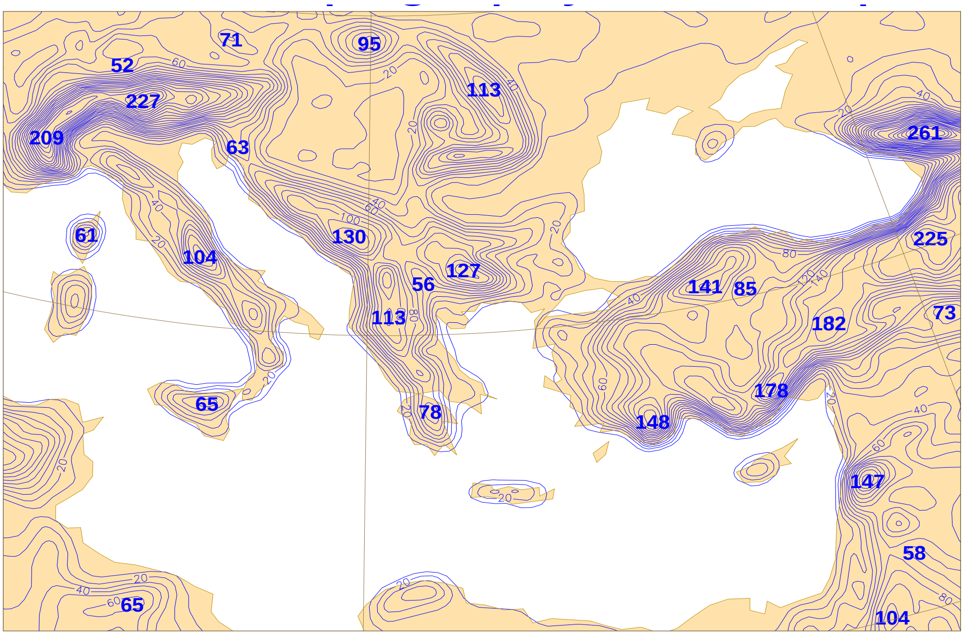


Figure 7 : The model height (in dekameters) for southeastern Europe

2.2.2 The Planetary Boundary Layer

The treatment of the **Planetary Boundary Layer (PBL)**, plays a fundamental role for the whole atmosphere–earth system. It is through the surface exchanges of momentum, heat and moisture that the atmosphere "feels" that it moves over a rough land surface or a wet smooth sea (Beljaars and Viterbo, 1993).

The lowest 13 levels are at around 10, 30, 60, 100, 160, 240, 340, 460, 600, 760, 950, 1170 and 1400 m above the model surface. Even with this fairly high resolution the vertical gradients of temperature, wind, moisture etc. in the PBL cannot be described very accurately, let alone the turbulent transports of momentum, heat and moisture. For the estimation of these parameters the model uses the larger scale variables such as wind, temperature and specific humidity, with the assumption that the transports are proportional to the vertical gradients.

At the earth’s surface, the turbulent transports of momentum, heat and moisture are computed as a function of air–surface differences and surface characteristics. Over land areas, snow depth, soil temperature and wetness are forecast variables, calculated by a model of the soil with four layers with respective depths of 7, 21, 72 and 189 cm.

The **sea surface temperature (SST)** is based on analyses received daily from NCEP, Washington. It is based on ship, buoy and satellite observations. In small waters like the Baltic Sea where rapid changes in SST can take place during the cold season, the real SST can sometimes differ by as much as 5° from the analysis.

The **sea–ice** fraction is based on satellite observations. The temperature at the surface of the ice is variable, according to a simple energy balance/heat budget scheme. The SST over ice–free water and the distribution of sea and sea–ice points is kept constant during the forecast; no freezing of the water or melting of the ice is allowed.

For the **albedo** a background monthly climate field is used over land. Over sea-ice the albedo is set to 0.7 and 0.5 for the two spectral bands. Open water has an albedo of 0.06 for diffuse radiation and a functional dependence of solar radiation for direct radiation. Over land the forecast albedo depends on the background albedo and the snow depth. It has a minimum of 0.07 and can go up to 0.80 for exposed snow and 0.20 for snow in forest.

The thermal properties of **snow covered ground** depend only on the snow mass per unit area. The snow depth evolves through the combined effect of snowfall, evaporation and melting (Beljaars and Viterbo, 1996). As the snow ages, the albedo decreases and the density increases.

The **soil moisture** is divided into skin and soil reservoirs. The skin reservoir (which mainly is moisture on vegetation) evolves under the action of its own evaporation and its ability to collect dew and intercept precipitation. The soil reservoir takes into account precipitation and snow melt, as well as vertical transfer of water due to drainage and capillarity, evaporation over bare ground and root uptake by vegetation.

The **vegetation ratio** is separated into low and high vegetation fractions and the corresponding dominant types of vegetation are specified in each grid point and used by the model to estimate the evaporation.

The **orographic drag** scheme represents the momentum transport due to sub–grid gravity waves and the blocking effect of orography in relatively stable conditions. When stably stratified air flow crosses a mountain ridge, gravity waves are excited into the flow. Depending on the static stability and vertical wind shear, these gravity waves can propagate vertically until they have sufficiently large amplitude to break. The scheme has a certain impact on the large scale flow; it makes it slightly less zonal and contributes to the formation of blocking highs and cut–off lows.

2.2.3 Radiation

In view of the importance of **cloud–radiation interaction** in both long and short term processes, ECMWF has placed high emphasis on the treatment of the absorption and scattering by clouds of solar and terrestrial radiation. About 15 percent of the overall computational time is devoted to the radiation scheme.

The **radiation spectrum** is divided into eight frequency bands: two in the short wave spectrum (direct from the sun and diffuse radiation), and 15 in the long wave spectrum (from the earth and within the atmosphere). The upward and downward diffused radiation is computed for each of the 16 spectral bands. The parameters influencing the emission and absorption are pressure, temperature, moisture, cloud cover and cloud water content, and carbon dioxide, ozone, methane, nitrous oxide, CFC–11 and CFC–12. Assumed parameters are the solar constant, the concentration of CO_2, O_3, and other trace gases, the distribution and optical properties of aerosols and ground albedo, this last one modified according to the snow cover.

The radiation scheme is designed to take the **cloud–radiation interactions** into account in considerable detail. It allows partial cloud cover in any layer of the model. For cloudy grid points, computations are made both for clear and overcast conditions, and the total amount weighted together according to the forecast cloud amount. Provision is made to have the radiative effects of various types of aerosols (oceanic, continental, desert, urban, and stratospheric background) taken into account. The carbon dioxide has a constant mass mixing ratio over the whole globe corresponding to a volume concentration of 353 ppmv.

2.2.4 Clouds

The main purpose of the **cloud scheme** is to provide input to the radiation computations and to calculate precipitation. The clouds are generated by large–scale ascent, cumulus convection, boundary layer turbulence and radiative cooling. They are dissipated through evaporation due to large–scale descent, cumulus–induced subsidence, radiative heating and turbulence at both cloud tops and sides, as through precipitation processes.

The cloud scheme is unique in treating the main cloud–related processes in a consistent way by forecasting both cloud fraction and cloud water/ice content with their own prognostic equations. In the scheme the cloud processes are strongly coupled to other parametrized processes.

Convective clouds are computed in parallel with the convective scheme which in the model fulfils five objectives:

- compute the cloud amount and cloud water/ice to be passed on to the cloud scheme,
- compute the convective precipitation,
- compute the vertical transport of moisture,
- compute the vertical momentum fluxes,
- compute temperature changes in the atmosphere due to release of latent heat or cooling in connection with evaporation.

Sub–grid vertical fluxes of mass, heat, water vapour and momentum are computed at each model level with the help of a simple mass flux model interacting with its environment. The scheme is applied to penetrative convection, shallow convection and mid–level convection. They are mutually exclusive, so only when the scheme fails to create cloud of one type, does it try the next.

Deep convection predominantly occurs in disturbed situations with a deep layer of conditional instability and large–scale moisture convergence. The downdraught mass flux is assumed proportional to the updraught mass flux.

Shallow convection predominantly occurs in undisturbed flow, in the absence of large scale convergent flow. The moisture supply is from surface evaporation. It does not normally produce precipitation.

Mid–level convection describes convective cells which originate at levels above the boundary layer, like Altocumulus castellanus floccus. Less clearly visible, but frequent, are rain bands connected to extratropical cyclones.

Stratocumulus clouds are linked to the boundary layer moisture flux produced by the vertical diffusion scheme.

Stratiform clouds (e.g. low level stratus and medium level nimbostratus types) are determined by the rate at which the saturation specific humidity decreases due to upward vertical motion and radiative cooling.

Evaporation processes in connection with clouds are accounted for in several ways: large–scale and cumulus–induced subsidence and radiative heating, evaporation at the cloud sides due to turbulent processes and turbulent motion at the cloud tops.

2.2.5 The hydrological cycle

Precipitation processes do not only take into account the local water/ice content, but also different precipitation enhancement processes. The effect of evaporation of falling precipitation is also included. Two mechanisms to generate precipitation are included in the ECMWF model, for convective and for stratiform (frontal or dynamical) precipitation:

Convective precipitation: the condensate formed in the updrafts of the convection parametrization is water above 0˚C, ice below –23˚C and a mixture of the two in between. If the amount of condensate formed exceeds the value that can be sustained by the vertical velocity, precipitation is formed in the form of snow or water.

Stratiform precipitation: cloud water and ice from the cloud scheme are converted into precipitation dependent on the water/ice content. Precipitation enhancement processes, such as collection of cloud water by precipitation and the Bergeron process are also taken into account.

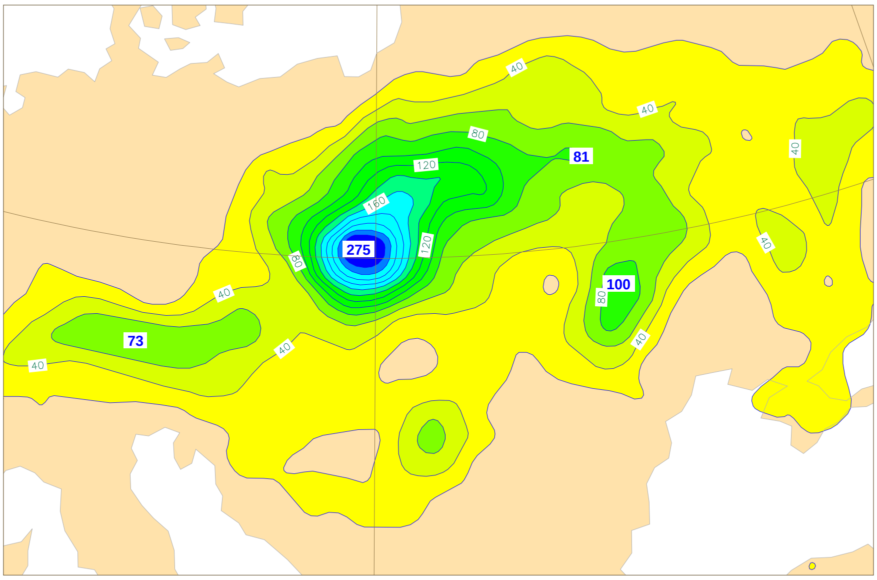


Figure 8 : Accumulated rainfall during the first 96 hours of the T213 operational forecast 5 July 1997 12 UTC. The floodings in eastern Europe summer 1997 were well forecast by the ECMWF model. However, the maximum rainfall of 400 mm in southeastern Poland was slightly underestimated.

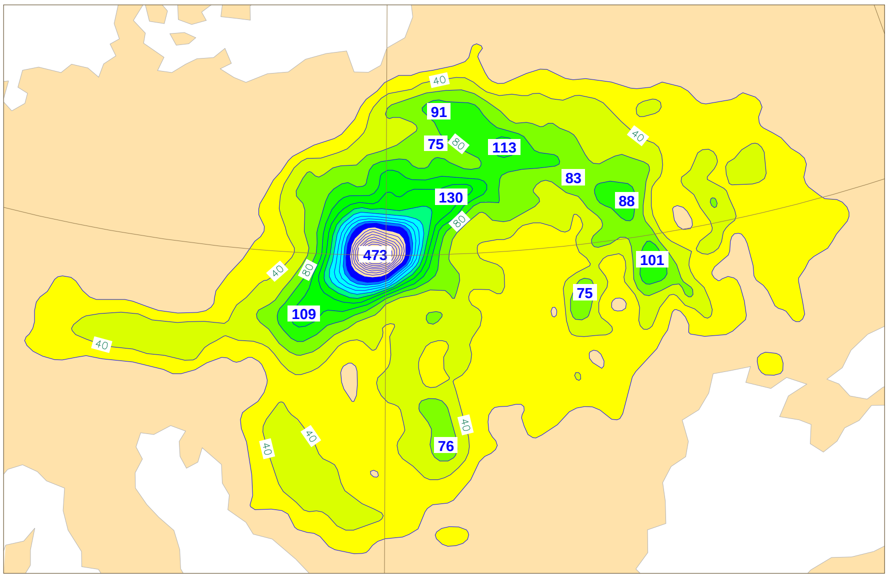


Figure 9 : Accumulated rainfall during the first 96 hours of a T639 experimental model forecast 5 July 1997 12 UTC. With a higher resolution model the correct level of intensity is achieved, and the orographic effects more realistically treated.

Evaporation: it is assumed that falling precipitation evaporates in non–saturated layers before reaching the ground. This may substantially reduce the surface precipitation. Evaporation of the precipitation is not assumed to take place within the cloud, but only in cloud free air besides or below the model clouds.

Melting: melting of falling snow occurs in a thin layer of a few hundreds of metres below the freezing level. It is assumed that snow can melt in each layer whenever the temperature exceeds 0˚C. The melting is limited not only by the snow amount, but also by keeping the induced cooling of the layer such that the temperature of the layer after melting is not below 0˚C.

2.3 The ocean wave forecast

The wave model that is used for ocean wave forecasting at ECMWF is the WAM model (WAMDIG, 1988, Komen et al, 1994). It describes the rate of change of the wave spectrum due to advection, wind input, dissipation due to white capping and non linear wave–wave interactions. The wave spectrum gives the distribution of wave energy over frequency and direction and gives a complete specification of the sea state. The WAM model is the first model that solves the complete energy balance equation, including the computationally expensive non linear interactions.

Table 13: Wave forecast products

2D-spectra: Peak period of 1D-spectra 2D-spectra for total sea, wind: Significant wave height, Mean wave direction, mean wave period
Global model: 0.5° x 0.5° latitude/longitude T+0 to T+240 every 6 hours Baltic and Mediterranean model: 0.25° x 0.25° latitude/longitude T+0 to T+120 every 6 hours

In June 1998 the wave model was integrated into the atmospheric model allowing two–way interaction of wind and waves. At the same time ensemble prediction of ocean waves started. Assimilation of altimeter data was introduced in the global version of the wave model in August 1993. Buoy data are not assimilated, instead, they serve as an independent check of the quality of modelled wave height (Janssen, 1997; Janssen et al, 1997).

Two versions of the WAM model are running at ECMWF: the global model has an irregular latitude longitude grid with a resolution of 55 km. The advection time step is 20 minutes, the same as for the source term integration. The wave spectrum has 30 frequency bins and 24 directions; the limited area model covers the North Atlantic, Norwegian Sea, North Sea, Baltic Sea, Mediterranean and the Black Sea. It has a resolution of 28 km, shallow water effects are included and the advection and source time step are 10 minutes. The wave spectrum has 30 frequency bins and 24 directions (Janssen, 1998).

The wave forecast is to a considerable extent determined by the quality of the surface wind forecast. This implies that there may be underestimation of the wave forecast near the coasts and in enclosed basins such as the Baltic and the Mediterranean. Furthermore, in rapidly varying circumstances such as occur near fronts or at the peak of the storms the resolution of the atmospheric and wave model may limit a realistic representation of the sea state.

Finally, there may also a number of problems with the wave model itself. For example, the propagation of swell is handled by a simple first-order upwinding scheme which may give rise to a smoothing of the wave field. Errors due to this are, however, fairly small as they are on average of the order of 10-20 cm in the significant wave height.

Nevertheless, it should be remarked that since the introduction of the T_L511 model, verification of significant wave height and peak period against Northern Hemisphere buoy data has shown an outstanding performance of wave analysis and forecast, in particular near the coast and during extreme events.

Coupled to the ensemble prediction system (see ch.4) fifty wave forecasts are generated from the fifty ensemble forecasts. From this a wide range of information can be derived, for example the probability of significant wave heights of more than 4 metres.

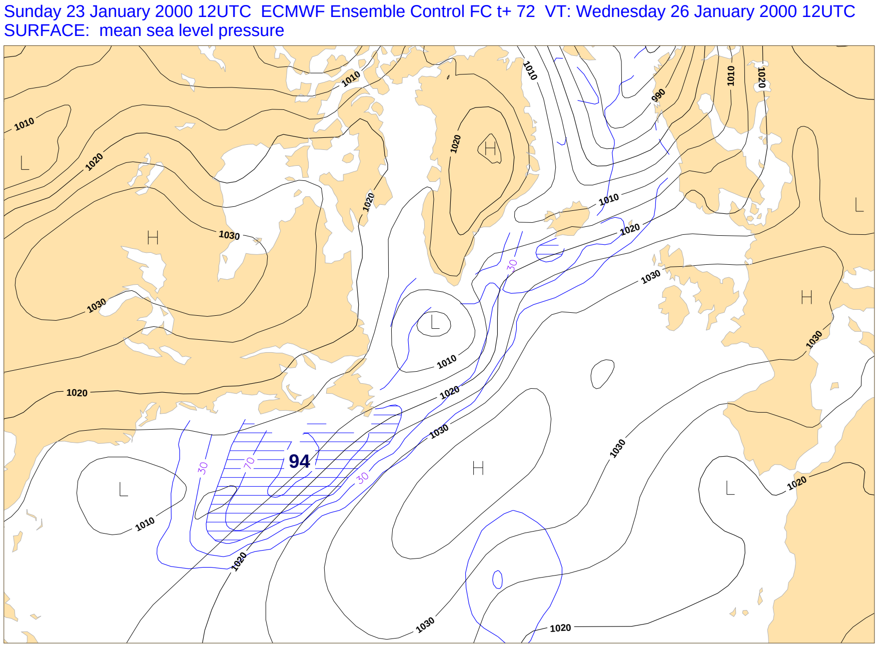

Sunday 23 January 2000 12UTC ECMWF Ensemble Control FC t+ 72 VT: Wednesday 26 January 2000 12UTC
SURFACE: mean sea level pressure

Figure 10 : Probabilities of wave heights exceeding four metres on 26 January 2000 12 UTC according to the wave ensemble forecast three days earlier.

2.4 The seasonal forecast system

Seasonal forecasting is the attempt to predict the probability distribution for weather several months or more into the future. The emphasis is on averages over a month or season, and how the probability distribution differs from "climatology". Seasonal forecasting is possible, because although the details of individual weather systems are not predictable on these time scales, the statistics of them are determined by various factors, some of which can be predicted. The most important factor is sea surface temperature, especially in the tropics. Other factors include soil moisture and snow cover.

In many parts of the tropics, where changes such as those associated with El Nino can have a large impact on global weather patterns, a substantial part of the year–to–year variation in seasonal mean rainfall and temperature is predictable. In mid–latitudes, the level of predictability is lower, and Europe in particular is a difficult area to predict. It is essential to understand that the forecasts are necessarily probabilistic, and that the range of possible values predicted may often differ from climatology by only a modest amount.

A basic seasonal forecasting system has been running at ECMWF since early 1997, and a limited set of forecast products has been supplied since December 1997. A comprehensive "Seasonal Forecast User Guide" is available from the ECWMF web pages. Information on the system presently running can also be found there, together with some assessments of the skill of the seasonal forecasts. Initial applications of seasonal forecasting are expected in energy, agriculture, insurance, health and emergency relief; many of these applications will be most valuable beyond the geographical borders of Europe.

Using numerical models of the ocean and atmosphere to calculate seasonal forecasts is a challenging problem, and although a useful level of skill now exists, a mature seasonal forecasting capability will take many years to develop. Note that while an operational schedule may be maintained, the science of seasonal forecasting is still under development. From a scientific perspective our systems are still experimental.

The value of seasonal forecasting is expected to increase in the coming years, both because of improvements in the forecast skill and from better use being made of the information that the forecasts contain.

3. The data assimilation and analysis system

3.1 Introduction

In the early years of NWP initial conditions for the simulations were obtained from manually analysed meteorological charts, laboriously interpolated to pre–defined grid points. It was not until the mid 50's that the current concept of fitting a prognostic first–guess field to observations was suggested and successfully tried.

During the 80's the increasing availability of asynoptic data, in particular over the oceans, stimulated the research into more advanced analysis procedures like the "four–dimensional variational data assimilation" where the concept of a continuous feedback between observations and model was put on a mathematical foundation in the so–called Kalman filter (Bouttier and Courtier, 1998). This technique has become even more important due to the last decades' gradual reduction of the radio sounding network and increase in satellite data.

3.2 The ECMWF analysis cycle

At ECMWF global analyses are made in two twelve hour intervals a day, centred around the main analysis times, 00 and 12 UTC. The analysis is performed by comparing the observations directly with a very short (3 to 15 hour) forecast using exactly the same model as the operational medium-range forecast. The differences between the observed values and the equivalent values predicted by the short–range forecast are used to make a correction to the first–guess field in order to produce the atmospheric analysis (Courtier et al, 1998). For remotely sensed radiances (currently ATOVS and SSM/I) the comparison involves complex radiative calculations (Andersson et al., 1998). The upper-air analysis is then combined with the surface analyses of snow, ice, SST, soil wetness and ocean waves, in order to produce the initial state for the next short-range forecast.

3.3 Data availability

As a forecasting centre with the emphasis on the medium–range, ECMWF operates with long data collection times, varying between 15 hours for the 00 UTC analysis and 8 hours for the 12 UTC final analysis. A preliminary 3DVAR analysis is made of the 00 UTC analysis cycle, based on slightly reduced data (shorter cut–

off time) to allow an early forecast to be run to provide boundary conditions for limited area models.

Real–time operational constraints mean that the data can only be used if it is received in time for each analysis. Approximate cut–off times at the beginning of year 2000 are given below for each of the three daily operational analyses.

Table 14: Cut-off times for different analysis cycles

Nominal analysis time	00 UTC short cut-off	00 UTC	12 UTC
Earliest data time	21:00 UTC	15:00 UTC	03:00 UTC
Latest data time	03:00 UTC	03:00 UTC	15:00 UTC
Cut-off time	03:05 UTC	16:20 UTC	19:50 UTC

This schedule ensures the most comprehensive global data coverage, including the Southern Hemisphere surface data and global satellite sounding data.

3.4 Data pre-selection

A very large amount of observed data is available for use by the assimilation and forecast system. In a particular 12–hour period there is a total of more than 500 000 pieces of data that are used. All the available data is considered for use in the analysis. Exceptions are satellite observations or locally dense data like aircraft reports where some thinning is applied. Some data are monitored for future use, others may appear in several different forms and some are redundant. Below are shown the average percentage of used observations:

- > **90%** Pressure from SYNOP and SHIP, humidity from SHIP
- **70–90%** Pressure from DRIBU, temperature from TEMP, wind from DRIBU, PILOT and TEMP.
- **50–70%** Humidity from land SYNOP and TEMP, pressure from PAOB (bogus pressure observations from Melbourne), winds from SHIP and AIREP, temperatures from AIREP.
- < **15%** satellite observations from ATOVS (radiances), SCAT (scatteometer winds), SSM/I (humidity observations) and SATOB (cloud wind observations)

For some data types, there is a transformation of the received data into physical variables that are better suited to the analysis procedure. This is the case for SCAT (the numbers given are for ambiguous wind vectors) and for SSM/I (the numbers given are for retrieved total column water vapour). Also, humidity observations (originally relative or specific humidity, or dew–point temperature) are transformed using the reported temperature. Other data types are used to analyse snow, ice, SST (sea surface temperature), soil wetness and ocean waves.

The various data types have very different characteristics in terms of geographical coverage, vertical structure, temporal distribution and ability to correct the model state. Whereas in situ observations tend to be more informative than remotely–sensed observations, the latter have a better coverage. Many conventional observing stations report at 12–, 6– or 3–hourly intervals, but some report every hour. Aircraft, buoy and (orbiting) satellite data production tends to be continuous in time (Kelly, 1997; Järvinen et al, 1999).

3.5 Quality control of observations

An intricate quality control procedure is applied in order to ensure that only good quality data are used in the analysis. The data is **permanently blacklisted** if

* it cannot be compared reliably to the model state, usually because the model is not realistic enough, or not enough is known about the physical nature of the data. This is true for example with SCAT observations over land, or SYNOPs for which the model orography is very far from the actual station height.

* its quality has not yet been properly assessed. All experimental data is carefully monitored before it is allowed into the system.

* it could not be decoded satisfactorily, for instance because it does not follow WMO conventions.

A particular form of blacklisting is the thinning produce applied to a station, platform or satellite reporting too densely in time or space. This is the case for all satellite data and some AIREPs. Multiple reports from the same (geographically fixed) station are used at up to one–hour frequency. All reported radiosonde levels are used.

The data is **temporarily blacklisted** if

* the monitoring has found problems in the data. The blacklist is updated manually once a month; it involves the comparison of the data with model fields, neighbouring observations and other data types.

Data which is not blacklisted can still be **rejected by the quality control** if

* it is not realistic according to basic tests that follow WMO recommendations: hydrostatic check of TEMP data, check of the displacement of ships and drifting buoys, ship data over land, very large deviations from climatology, duplicate report, etc.

* it is very different from the background fields of the model (i.e. the previous short–range forecast). This test is more stringent for areas and parameters that are believed to be more accurately forecast (Järvinen and Undén, 1997). It can be

combined with the thinning procedure to ensure that the best data is not thinned out.

* it disagrees significantly with its neighbours. This test is done within the variational data assimilation procedure (Andersson and Järvinen, 1999).

All quality control decisions are recorded for diagnostic purposes. The final decision to use or to reject the data can involve more than one of the above steps.

3.6 The 4DVAR analysis procedure

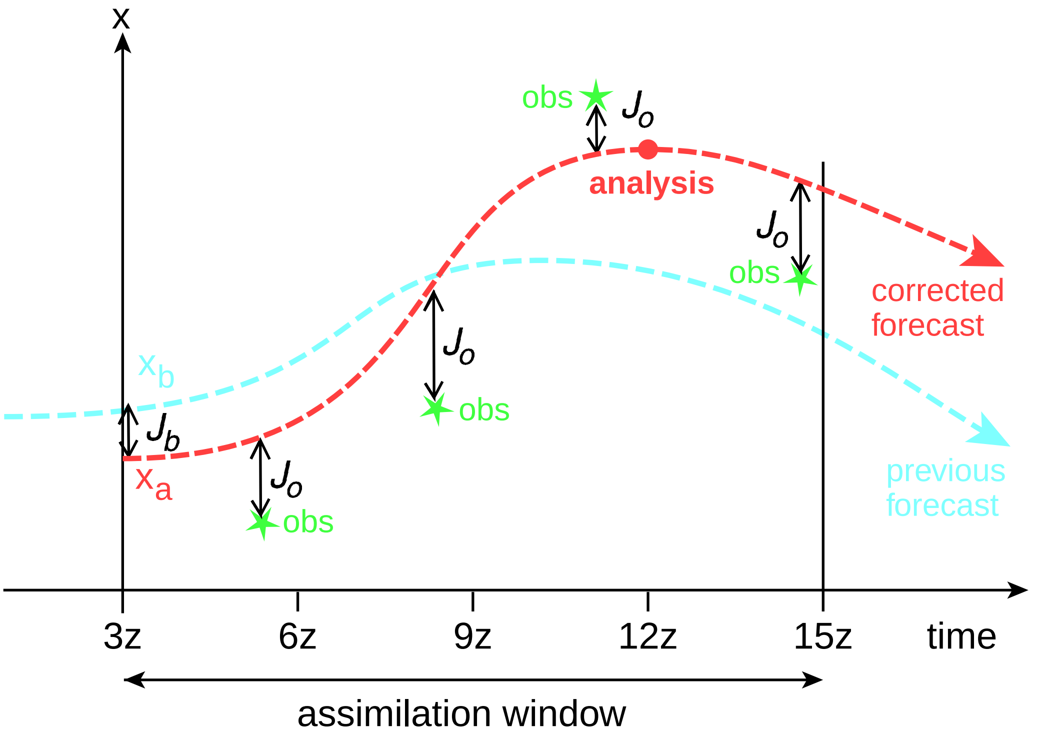


Figure 11 : A schematic illustration of the 4DVAR analysis

With the four-dimensional variational analysis (4DVAR) system introduced in November 1997 the influence of an observation in space and time is controlled by the model dynamics which increases its realism of the spreading out of the information (Bouttier and Rabier, 1998, Saunders et al, 1998, Rabier et al, 2000, Klinker et al, 2000, Mahfouf et al, 2000). This is achieved by having the background errors modified by the model dynamics over the assimilation period in a flow dependent way. Observations are thereby given larger weights near rapidly moving or deepening cyclones where the forecast uncertainty is larger.

The atmospheric analysis procedure is a four–dimensional variational data assimilation over a time window from 9 hours before to 3 hours after the nominal analysis time. The correction to the first–guess fields are made with a T159L60 model. The algorithm is designed to find a compromise between the previous forecast at the beginning of the time window, the observations, and the model evolution inside the time window.

The result can either be seen as a sequence of analyses, or one short–range T511L60 forecast, that spans the 12–hour time window. For instance, the 12:00 atmospheric analysis is actually an optimal forecast at 9–hour range, which was started at 3:00 using observations between 3:00 and 15:00. Theoretical studies indicate that the most realistic model state is obtained near the middle of the time window.

3.7 Structure functions

In 4DVAR the compromise between the information in the background and in the observations is determined by the combination of the model dynamics with weights given to the background (or first–guess) field and to the observations. These weights are based on a priori estimates that describe the errors one expects to find in the background fields (Derber and Bouttier, 1999) and in the observed values.

A correction of one model variable will generate corrections of other variables whenever this is consistent with the dynamics. For instance, a sequence of observation of humidity that show a displacement of atmospheric structures, will entail a correction not only of the moisture but also of the wind field that is consistent with the advection of these structures.

The relative weight given to the observations and to the background depends on the local predictability of the flow: an observation inside a baroclinically unstable jet or a growing storm will be fitted more than the same observation in an anticyclone, for instance. This makes 4DVAR able to correct the phase, the tilt and the deepening rate of developing storms (Rabier et al 1997) - but it may also make 4DVAR sensitive to erroneous observations, individual or in a group, which have not been detected by the quality control.

3.8 No analysis is perfect

There are different reasons why errors in the initial conditions (the analysis), can negatively affect the forecast:

- *No data over considerable times and areas.* This can be the case when cyclonic developments have their origin in the arctic and, in particular, in subtropical areas where the presence of warm and moist air make errors have larger than normal consequences.
- *Bad data have been accepted.* This is often the case when the First Guess (FG) is bad and agrees with erroneous data, or when lack of surrounding data makes any internal check impossible.
- *Good but unrepresentative data have been accepted.* Isolated observations reflecting extreme or small scale weather conditions are interpreted as related to more large scale dynamical systems.
- *Good data have been rejected.* Either the FG is so wrong that the observations have not been accepted by the system, or other erroneous observations in the neighbourhood have gained more weight.
- *Good data has influenced the analysis in a wrong way.* A combination of mainly correct observations can yield bad forecasts because the way the structure functions spread out the information. This can happen for example when a weather system is covered only partially by observations.

An analysis error does not necessarily lead to any forecast failure (except in the immediate environment). To have a more wide spread impact it must occur in a dynamically sensitive region, in particular where young baroclinic systems develop. During a cyclogenesis huge amounts of kinetic energy feed into the upper mid–latitude westerlies, in particular the jet streams, which will transport the energy out of the area. In a short time the conditions downstream will be affected. If a new cyclone is developing there, it might start to develop erroneously. In a few days time the errors from this system will spread from one system to the next in a process, reminiscent of a "domino effect”.

3.9 Monitoring of the data assimilation system

The ECMWF puts great emphasize on monitoring the quality of the observational network. It is done through statistical monitoring of all platforms to establish any systematic biases or erratic variations, through daily inspections of the analysis material and case studies of bad or very inconsistent forecasts.

The investigation of a bad forecast is made in three steps to answer when did the error enter into the analysis, where did it happen and what caused the error?

To trace the geographical origin of the error, different methods, empirical as well as objective, have been developed: forecast error maps, sensitivity analyses and the EPS perturbations. The speed of this transport is roughly determined by the upper tropospheric flow. For a velocity of 30 m/s this corresponds to 30˚/day at 45˚ latitude, which agrees well with the theoretical calculations of the "group velocity" of dispersive Rossby waves. On an error map this appears as an initial error which will move slowly downstream while it generates a wave train with increasing longitudinal extension.

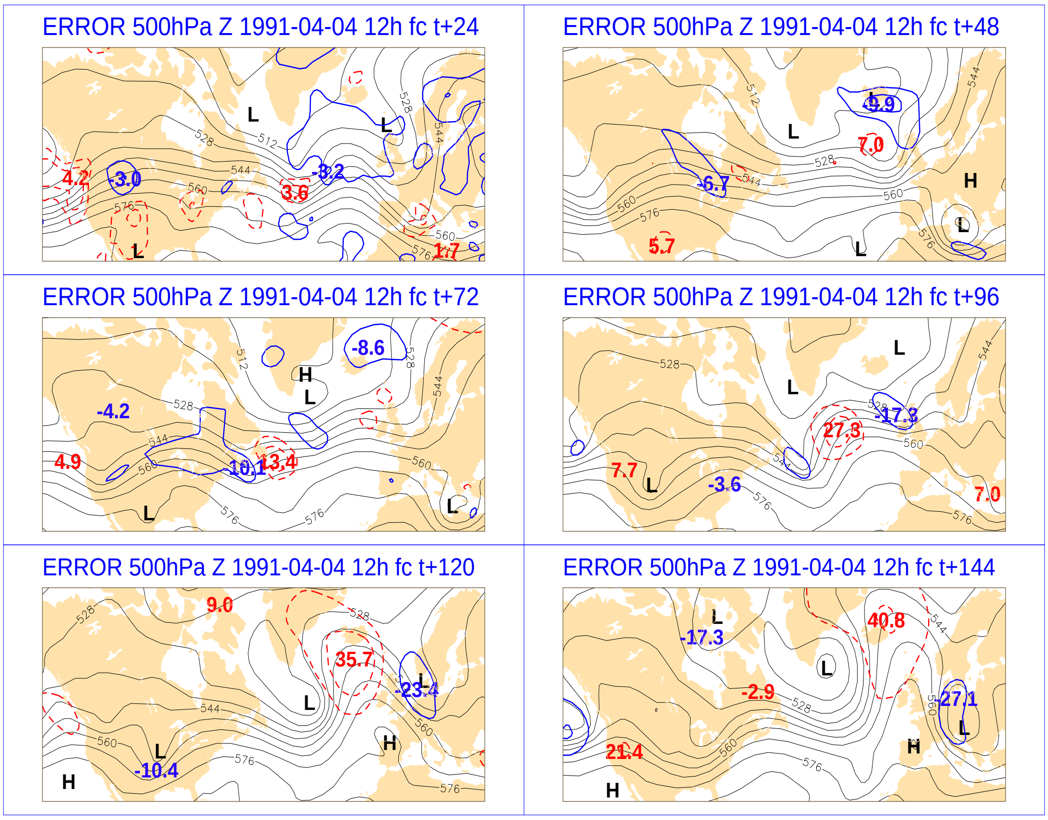


Figure 12 : The forecast 500 hPa flow (thin lines) and errors (thick dashed lines positive, thick full lines negative errors) from 4 April 1991. A TEMP SHIP on the northwest Pacific, west of Vancouver had been wrongly decoded and caused a 20–30 gpm positive analysis error (+4.2 dgpm at +12h). Early in the forecast a new negative error (-3.0 dgpm) is created downstream over W USA. While it moves eastward under amplification (amplifying from -6.7 to -18.1 dgpm) another error of opposite sign (amplifying from +13.4 to+27.8 and +35.7 dgpm) is followed by a fourth error (amplifying from -17.3 to -23.4 and -27.1 dgpm) which was associated with a spurious cut-off over NW Europe.

The fast speed of the influence means that a three day forecast for Europe is dependent to a large extent on the initial conditions over the whole N Atlantic, a five day forecast on the initial conditions over North America, and a seven day forecast on the initial conditions over the North Pacific. In summertime the spread of influence is slightly slower.

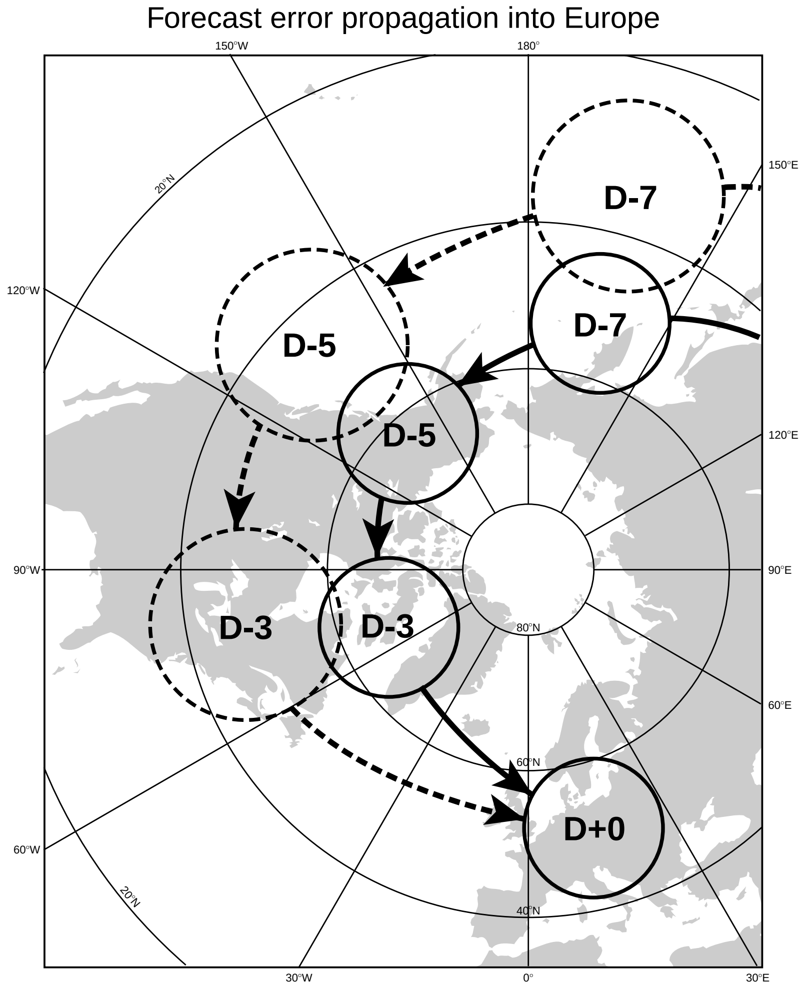


Figure 13 : The areas in the NH where analysis errors D–N days back in time will have most effect on the forecasts over Europe at D+0. During winter the zonal flow is stronger and passes over the dense network over the US. At summertime when the US is covered by a subtropical ridge, the main flow is slower and passes over Alaska, Canada and Greenland with their coarse network of observing stations.

In the Southern Hemisphere the typical group velocity is 40°/day due to the predominant zonal flow and low frequency of blocked patterns. (Persson, 2000)

Finally, the search for the actual observation or observations that caused the error is not trivial. As mentioned above, even correct observations can give rise to analysis error if the analysis spreads out their influence in an erroneous way. Rerunning the analysis and forecast with the suspected observations left out is one way to establish the cause of the failure. But this method is not 100% conclusive since a change in one or a few observations might change the assimilation of other neighbouring platforms.

3.10 Future developments of the data assimilation system

A main future directions for the evolution of the analysis are to resolve smaller–scale structures by using more data (particularly from satellites) and by increasing the model resolution, to make the analysis more consistent with the model dynamics and the water cycle (clouds and precipitation), and to assimilate new variables such as ozone.

The sensitivity of the analysis to the dynamics has recently been improved by an algorithm called the Reduced Rank Kalman Filter (RRKF), which will mean that the recent history of the atmospheric flow will be taken into account, even at the time of the background field (Fisher, 1998).

4 The Ensemble Prediction System (EPS)

4.1 Introduction

Twenty-thirty years ago the introduction of primitive equation models, the rapid advance in computer technology, remote sensing from satellites and an ever increased sophistication of numerical methods fostered a sense of great optimism. But progress in predictive skill remained slow and gradually the question arose if there was an ultimate limit to atmospheric predictability.

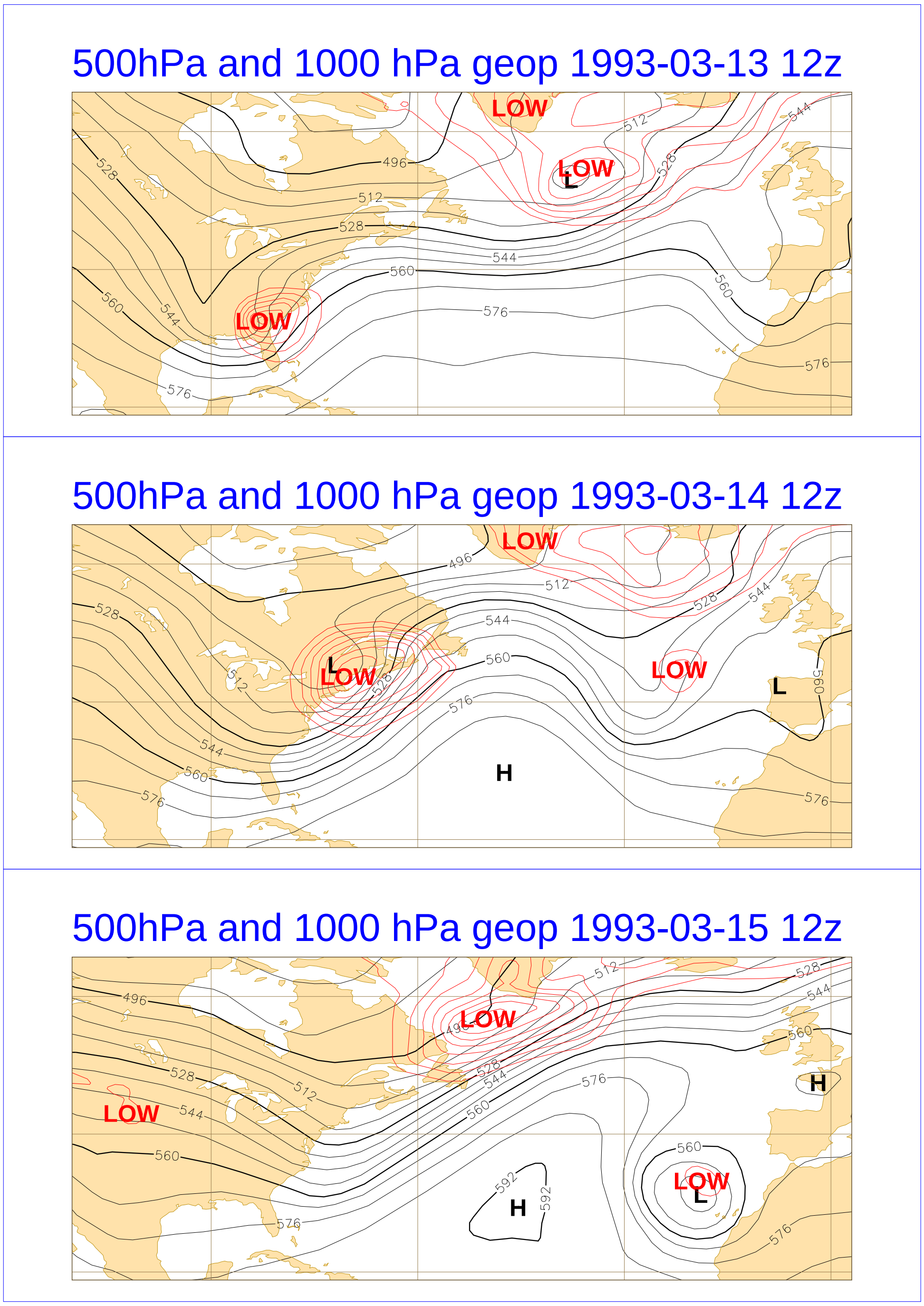
The interest came to focus on a strange result, first reported at a NWP meeting in Tokyo in 1960 by Edward Lorenz at the MIT. He had investigated if calculations based on non–linear differential equations could be replaced by statistical methods. The answer was “no”, but during one of his computational simulations he noticed how very small differences in the initial conditions could affect his extremely sensitive non–linear differential equations. It looked as if the old saying that “a sneeze in China may set people to shovelling snow in New York" indeed was true (Lorenz, 1993, p.15,130ff).

The consequences for NWP was that the limitations in the observational network and measurement accuracy would impose an upper limit in weather forecast quality. However, it was realized that in spite of this, the value of the NWP would be highly enhanced if the quality of the forecasts could be assessed a priori. The idea of including a stochastic element in NWP was born, but it had to wait until the late 1980’s until sufficient computer power made experiments possible.

4.2 The spread of forecast errors

As mentioned earlier, for an analysis error to have a more wide spread impact it must occur in a dynamically sensitive region, in particular where young baroclinic systems develop. The errors from this weather system will in a few days time spread to the next, in a process reminiscent of a "domino effect”.

The 2-4 day forecasts for the European area are therefore sensitive to the analysis over W Atlantic and eastern North America. The intensity and positio of the cut-offs frequently forming in the eastern Atlantic, is highly influenced by the presence of a strong cyclogenesis over the Mexican Gulf and southern USA. Any error in the forecast of this upstream feature will be crucial for the success of the downstream cut-off.

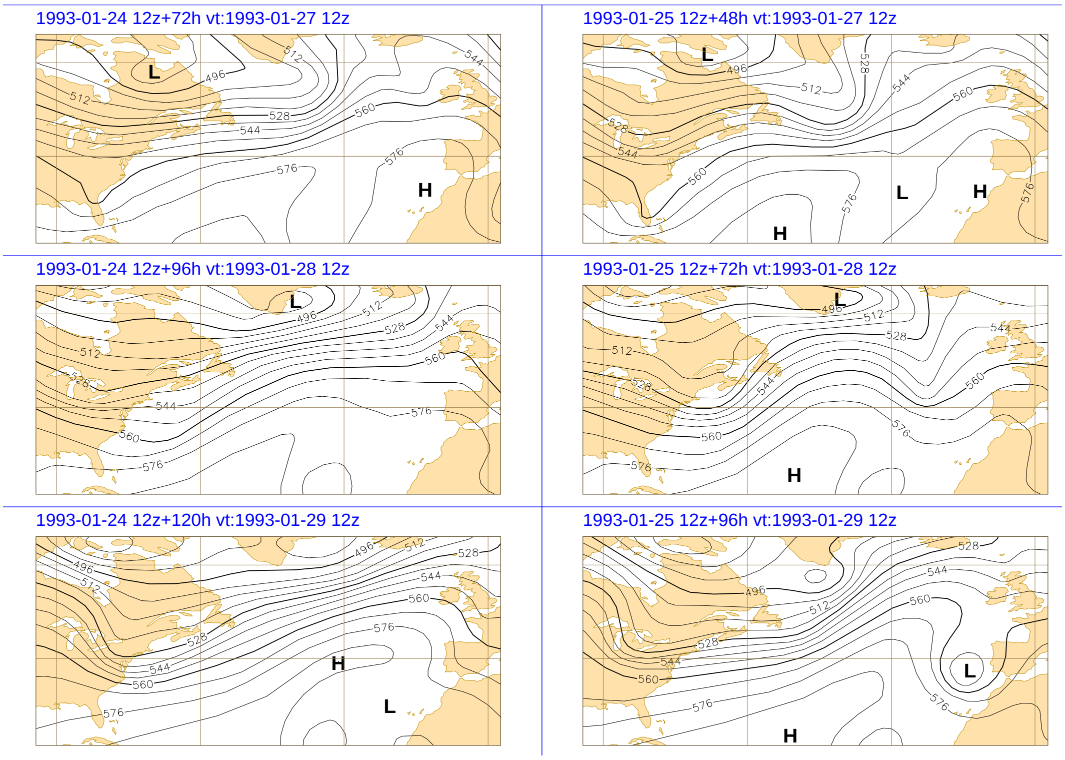

500hPa and 1000 hPa geop 1993-03-13 12z

500hPa and 1000 hPa geop 1993-03-14 12z

500hPa and 1000 hPa geop 1993-03-15 12z

Figure 14 : The ECMWF analyses of the 500 and 1000 hPa flow during the March 1993 American “Super Storm” show how the deepening of the strong upper trough over S USA in connection with the intensification of the surface low (thin lines only for every 4 dgpm 1000 geop. below 0), was followed by a cut-off some days later west of Portugal. Note that the origin of the trough that “cut off” was a weak trough southeast of Newfoundland.

The 5–7 day forecasts for the European area are sensitive to the initial conditions over the central and western part of North America, and the eastern part of the Pacific. Forecasts beyond a week are influenced by the initial conditions over central and western parts of the Pacific, and at day 10 from eastern Asia.


1993-01-24 12z+72h vt:1993-01-27 12z

1993-01-25 12z+48h vt:1993-01-27 12z

1993-01-24 12z+96h vt:1993-01-28 12z

1993-01-25 12z+72h vt:1993-01-28 12z

1993-01-24 12z+120h vt:1993-01-29 12z

1993-01-25 12z+96h vt:1993-01-29 12z

Figure 15 : A similar chain of developments as in the previous figure occurred some time earlier on 27-29 January 1993. This was well forecast by the run from 25 January (right column), whereas the forecast from the day before (left column) missed the formation of the cut-off. The decisive analysis differences were not associated with the trough which ultimately “cut off”, but with a upstream trough and baroclinic development over southeastern USA and northern Mexico. This system was forecast to pass over Florida on the 27 January (upper row) and was weaker in the earlier forecast than in the later. The stronger development caused the downstream ridge to amplify which in turn affected the trough moving eastward over the Atlantic.

Since the initial state of the atmosphere is known with a limited accuracy, even small analysis errors in sensitive parts of the atmosphere may affect the very large scale flow during the course of the ten day forecast period (Palmer, 2000). Another, equally accurate analysis with a slightly different geographical distribution of the initial errors, might yield a different forecast. *The deterministic forecast is just one possible development of a number of alternatives, not necessarily the most likely.*

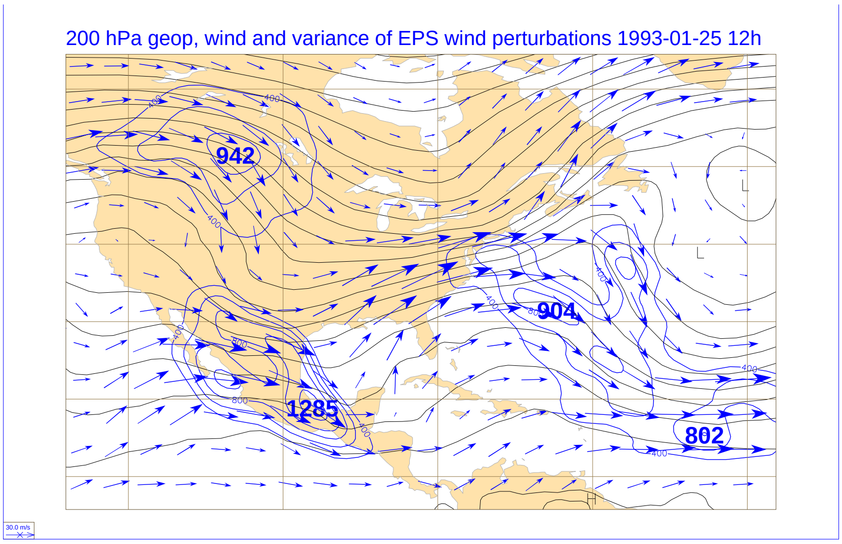


Figure 16 : The 200 hPa geopotential and wind field 25 January 1993 12 UTC with the variance of the EPS wind perturbations. The trough over southeastern USA and northern Mexico is coupled to a major baroclinic development. The EPS has identified this trough as potentially unstable and suggested large variations in its intensity. Due to the smallness of the Coriolis parameter at lower latitudes a more correct representation of the perturbations is achieved by the wind than by the pressure or geopotential field. Note the strong ageostrophic winds over southern USA

4.3 The ECMWF ensemble prediction system - an overview

The ECMWF Ensemble Prediction System (EPS) has been a part of the operational production since 1992. The EPS simulates possible initial uncertainties by adding, to the unperturbed analysis, small perturbations within the limits of uncertainty of the analysis. From these, a number of different forecasts are produced (Mureau et al, 1993; Molteni et al, 1996).

At the time of writing (Spring 2001), 50 perturbations are computed over the Northern Hemisphere. A set of perturbations is computed separately over the Southern Hemisphere. This forms the basis for 50 alternative forecasts, run with approximately half the horizontal resolution (T_L255) compared with the deterministic forecast (T_L511) and with 40 levels in the vertical instead of 60 (Buizza et al, 1999). The different initial states are a priori assumed to be equally likely.

Fields from the free atmosphere are archived every 12 hours. Weather parameters like 2 m temperature, 10 m wind, precipitation and cloudiness are archived every 6 hours. Also the maximum and minimum temperatures are archived (see ch.5). One forecast, *Control*, is run at $T_L255L40$ with a non-perturbed analysis.

4.4 Perturbations

The success of any ensemble system depends on its ability to identify regions where small uncertainties in the analysis are likely to have significant impact on the forecast, and to create structures which will simulate these uncertainties.

4.4.1 The calculations of perturbations in the mid-latitudes

The EPS perturbation technique, based on a mathematical method called singular vector analysis, tries to identify the dynamically most unstable regions of the atmosphere by calculating where small initial uncertainties would affect a *48 hour* forecast most rapidly, i.e. both increasing or dampening the forecast amplification of a developing baroclinic system or unstable ridge (Buizza and Palmer, 1995).

The first 25 of these singular vectors, chosen not to overlap too much, are combined in a linear way to calculate hemispheric structures (separately for each hemisphere) which are able to have a significant effect on the forecast after 48 hours.

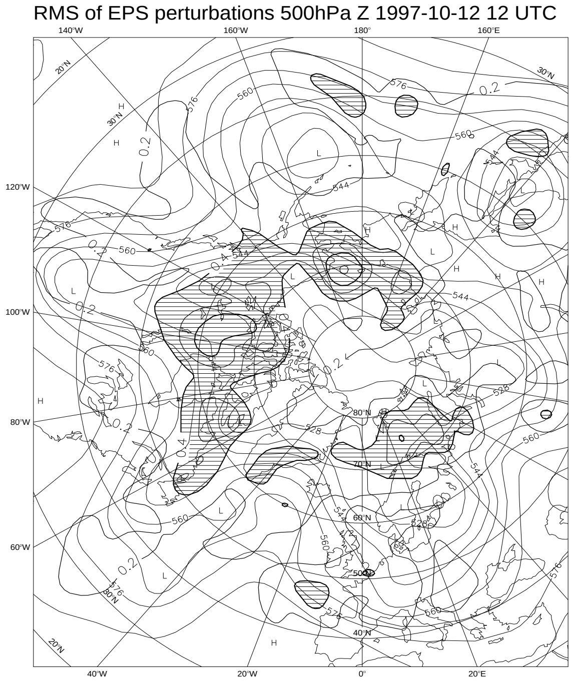
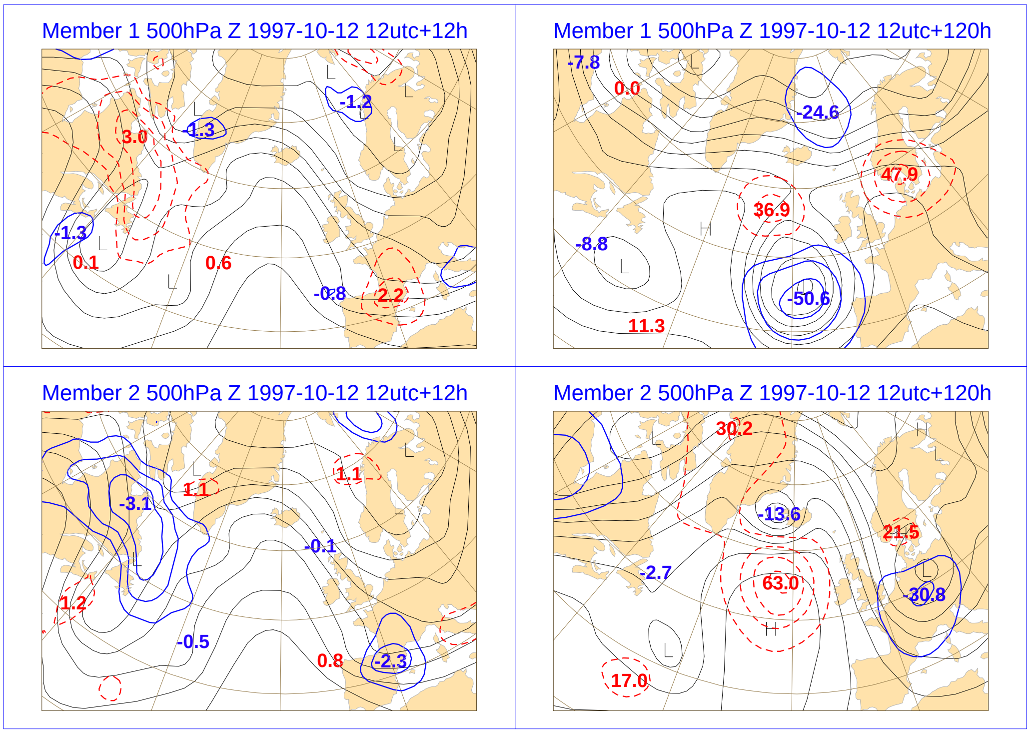


Figure 17 : The rms of the EPS perturbations at 500 hPa 12 October 1997 12 UTC. The system has identified the regions over Canada and Alaska as particularly sensitive to possible uncertainties in the analysis. Lower down in the troposphere (not shown) the conditions over western USA is also identified as sensitive.

By reversing the signs, 25 “mirrored” perturbations are produced, yielding a total of 50 global perturbation fields. These initial perturbations are scaled so that their local maxima are comparable to local analysis errors, and to have a realistic ensemble spread after 48 hours. The final perturbations are spatially uncorrelated. They are also considered a priori to be equally likely.



Member 1 500hPa Z 1997-10-12 12utc+12h

Member 1 500hPa Z 1997-10-12 12utc+120h

Member 2 500hPa Z 1997-10-12 12utc+12h

Member 2 500hPa Z 1997-10-12 12utc+120h

Figure 18 : The EPS perturbations at 500 hPa 12 Oct 1997 12 UTC at initial time, at +12 h and at +120 h for members 1 and 2. For these pair wise members the initial perturbations have opposite signs and mirror each other, but gradually the differences become random.

To further improve the spread and realism, perturbations from two days back are computed and added to the T_L255L40 (Barkmeijer et al, 1999). There are no perturbations within 30 latitude degrees on either side of the Equator. The forecasts for the Tropics are affected by influences propagating in from the extra-tropics.

The success of the EPS over Europe is determined to what degree it can correctly account for the uncertainties and alternative developments of an upstream baroclinic development. Most of the EPS perturbations which are of importance for the medium range forecast over Europe on a weeks range are inserted in the analysis of baroclinic systems over the north Pacific.

4.4.2 The simulation of model errors

Although the main approach has been to simulate the effects of possible errors in the initial conditions, increasing research work is devoted to simulate the effect the finite resolution of the model grid or simplified representation of the physical processes. These will have importance in connection with strong physical forcing, for example when tropical cyclones enter the mid-latitudes and interact with the baroclinic development in the westerlies. The source of such errors has been addressed by the introduction of *stochastic physics* (Buizza et al, 1999). For each ensemble member, the stochastic physics perturbs grid point tendencies by up to 50%, with a spatial correlation radius of 10 latitude degrees and a time correlation interval of 6 hours. The whole globe is perturbed, including the Tropics. The non-perturbed Control forecast is run without stochastic perturbations.

4.4.3 Simulating uncertainties in the Tropics

To improve the ensemble forecasts of developments which are greatly influenced by physical processes typical of lower latitudes, in particular tropical cyclones, work is under way to introduce a scheme for creating perturbations specially designed for the Tropics (Puri et al, 1999). Recent experiments have shown that they also can improve also extra-tropical developments

4.5 Spread–skill

Depending on the particular hemispheric flow pattern, forecasts originating from perturbed analyses develop more or less differently during the course of a ten day forecast.

4.5.1 Basic interpretation of the spread

If model errors played no role, and if initial uncertainties were fully included in the EPS initial perturbations, a small spread among the EPS members would be an indication of a very predictable situation. In other words, whatever small errors there might be in the initial conditions, they would not seriously affect the deterministic forecast. In these cases extended and/or detailed forecast interpretations are possible. By contrast, a large spread indicates a large uncertainty of the deterministic forecast, which prevents any extended or detailed forecast interpretation.

4.5.2 The provision of alternatives

But the EPS does not limit the interpretation of the spread just as a measure of uncertainties. The information will also suggest possible *alternative* developments and their respective likelihood. Last but not least, it will also indicate what is *not* likely to happen, which at times might be as important as knowing what is likely to happen. When, on some rare occasions, the spread might cover most of the climatological range, then nothing can be deduced from the forecast about any significant deviations from climate.

4.5.3 The spread is not a unique value

The spread-skill interpretation of the EPS is complicated by the fact that in one and the same forecast the spread often varies considerably from one parameter to another. A small spread in the 500 hPa geopotential forecasts does not necessarily imply a small spread in for example the forecast precipitation, and vice verse (see chapter 8 for further discussion).

4.6 EPS clustering

To compress the amount of information being produced by the EPS and highlight the predictable and thus relevant parts, individual EPS forecasts, which are “similar” according to some norm, are grouped together and averaged to constitute new forecast fields, so called *clusters*. The norm for judging this “similarity” can be the correlations between the fields or, as in the ECMWF system, their RMS differences.

4.6.1 The operational clustering

The ECMWF operational clustering algorithm is based on the RMS differences between the 500 hPa geopotential height ensemble forecasts, averaged from +120h to +168h taking the synoptic continuity into account. It is always the same members which make up the contents of each cluster

For two EPS members to join the same cluster they must display more or less the same synoptic 500 hPa development from +120 to +168 hours whereby a synoptic consistency is thereby obtained. This would not be the case if the clustering was made separately for each day in the forecast.

There are occasions when two members in the same cluster can be rather different at the beginning or end of the period, but sufficiently similar during the rest of the time interval to be placed in the same cluster. On the other hand, two members, being similar during a part of the period, may be placed in different clusters if they are sufficiently different during most of the period.

The number of clusters depends on three factors:

- The spread of the day, i.e. the EPS standard deviation,. It is varying from day to day, but follows a seasonal trend as the forecast errors, with higher values in winter than in summer
- The clustering threshold used to limit the clusters standard deviation. It follows the same seasonal trend as the spread and errors
- The degree of “multi modality”, the tendency of the forecasts to form discrete alternatives. For the same spread and threshold a multi modal distribution might lead to a smaller number of clusters than a mono modal distribution.

A large spread in the ensemble does therefore not necessarily lead to more clusters, nor does a small spread necessarily lead to fewer.

The clustering is performed separately for the whole of Europe plus four European sub–domain. The clustering over one sub-domain with respect to the position and intensity of a dominating feature will then be made without considerations of the uncertainties in the forecast of a blocking over another sub-domain.

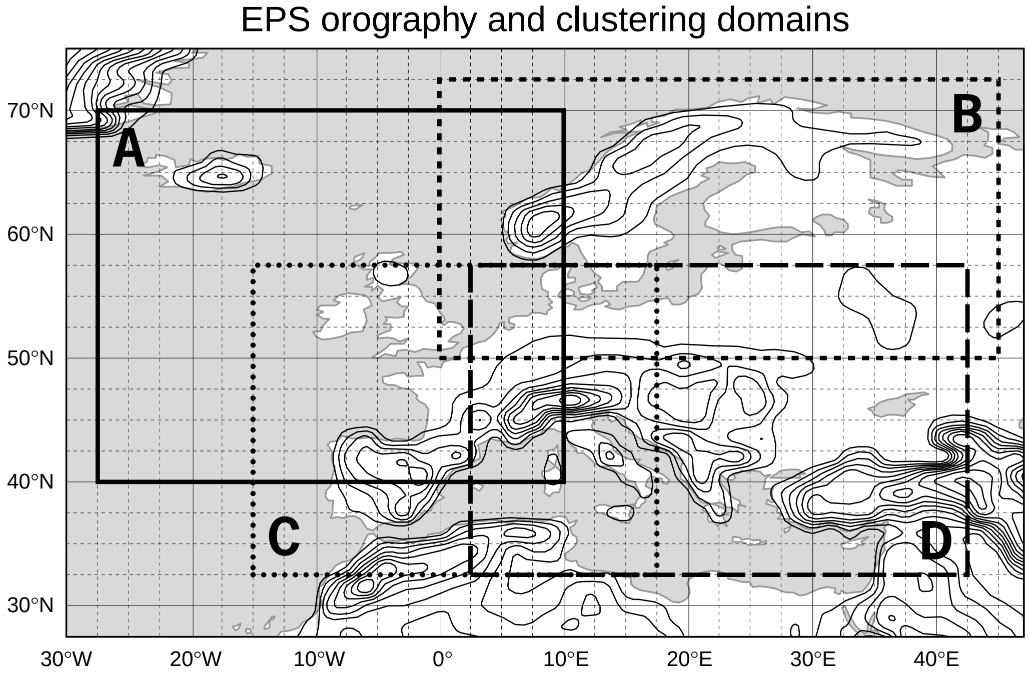


Figure 19 : The five main clustering areas, the European and four sub-areas..

4.6.2 The “tubing” clustering

Another clustering method, called *tubing*, averages all ensemble members which are similar, on a RMS basis, to the ensemble mean and excludes members which are significantly different. The average of all these “similar” members provides a more refined ensemble mean, *the central cluster mean*. The excluded members are grouped together in a number of "tubes" (maximum 9) each represented by their most extreme member allowing to better visualize the different scenarios in the ensemble.

The central cluster mean and the tubes are computed for the whole forecast range. For each tubing reference step (+96h, +144h, +168h, +192h and +240h), tubing products are generated over a 48-hour sequence finishing on the reference step (for example +48/+72/+96h for the +96h tubing), allowing a sequential view of the different tendencies. In the case of the +168h tubing, the sequence is over 96 hours, from +72h to +168h. The results are then applied to the 1000 hPa geopotential and 850 hPa and 500 hPa temperature. Tubes are computed over each of the five geographical domains Europe, NW Europe, NE Europe, SW Europe and SE Europe. They do not intended to serve as probability alternatives, only to give an indication of what is not included in the central cluster.

4.6.3 No ideal clustering

Every possible clustering is a compromise; the advantage of condensing information has to be paid by the risk of losing information which on some occasions, in hindsight, might have been important. There is really no superior or objective measure of which type of clustering is "best". Clustering can be performed over larger or smaller geographical areas, on different parameters, it can be done for each forecast time or for a longer period. Other possible ways:

clustering using correlation measures will highlight similarities in the patterns but may group together forecasts which differ in the overall level of temperature and geopotential heights.

clustering according to the 500 hPa flow might in a zonal situation give one cluster, whereas if the clustering had been performed on the MSLP pattern, the differences in the position and intensity of zonally moving baroclinic waves might have created 3–4 clusters.

clustering on individual forecast days will have the advantage that each day can be judged by its own merits; the disadvantage is that the temporal continuity and synoptical consistency will be lost.

Ideally the forecaster should have access to more than one clustering method, since what is the "best" clustering will vary according to the weather situation.

5 The forecast products

5.1 The operational schedule

ECMWF produces global analyses for the four main synoptic hours 00, 06, 12 and 18UTC, and global 10–day forecasts based on 12 UTC analysis. An additional analysis is run for 00 UTC with a cut–off time of around 3 hours, followed by a global 3–day forecast to provide some Member States with boundary conditions to their limited area models.

5.2 Direct model output

The model variables for the computation of the forecasts are temperature, wind and specific humidity. These primary parameters are converted into other atmospheric parameters. Tables 5 and 6 summarize the main output of the forecast model. These parameters are computed at 3–hourly intervals from 6 to 240 hours, based on 12 UTC data.

Table 15: Upper air parameters

Geopotential height (not on model levels) Temperature Vorticity and Divergence Wind (U and V components) Vertical Velocity Specific Humidity cloud ice/water content on model levels
Upper-air parameters are produced on the original model levels and on standard pressure levels (1000, 925, 850, 700, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 7, 5, 3, 2 and 1 hPa).

Table 16: Surface and single level parameters

Mean sea level pressure
10 metre wind
2 metre temperature
2 metre dew point
Maximum and minimum 2m temperature since previous post-processing
Large scale and convective precipitation
Snowfall
Surface temperature and soil wetness
Snowdepth
Total cloud cover
Low, medium, high and convective cloud cover
Surface fluxes, surface stress, surface roughness, albedo
Solar and thermal radiation

The 2 metre temperature and dew point and the 10 metre wind are computed from the values at the lowest model level (approx. 30 metres above ground) and at the surface, taking into account a prescribed state of the surface (albedo, roughness etc.). Analysis fields for 00, 06, 12 and 18 UTC including additional fields such as model orography, land sea mask, percentage of vegetation, etc. are also available. It should be borne in mind that surface parameters and cloud and radiation parameters are not analysed in the present system. The analysis and forecast output is archived into MARS (the ECMWF archiving system of meteorological data, cf. Meteorological Bulletin M1.9/2).

5.3 Dissemination products

A subset of parameters is available to ECMWF Member States through the operational dissemination system (table 17; cf. Meteorological Bulletin M 3.1 (2) for a description of the system). All parameters are available in lat-lon grid form, upper air parameters (except humidity) are also available in spectral form.

Table 17: ECMWF dissemination products

Operational products	Additional experimental products
Upper air parameters (on pressure levels and model levels) Mean sea level pressure to day 7	Upper air parameters (on pressure levels and model levels) Mean sea level pressure from day 71/2 to day 10
2 metre temperature 2 metre dew point 10 metre wind total precipitation total cloud cover to day 7 (every 3 h up to day 3, every 6 h beyond	2 metre temperature 2 metre dew point 10 metre wind total precipitation total cloud cover from day 71/2 to day 10 Additional weather parameters: large scale precipitation convective precipitation low cloud cover medium cloud cover high cloud cover snowfall snowdepth throughout the forecast range

Table 18: ECMWF dissemination EPS products

Control and Perturbed forecast products: Levels and validity: 1000, 850, 700,500,200 hPa for +0 hour to +240 hour forecasts at 12 h interval Geopotential, temperature, u- and v-velocities, specific humidity, vertical velocity, vorticity, MSL pressure, divergence Surface products: Large scale precipitation, convective precipitation, snow fall, total cloud cover, 10 m u- and v-components, 2m temperature and dew point temperature, two metre maximum and minimum temperature, all for +0 to 240 h at 6 h interval

Table 18: ECMWF dissemination EPS products

Cluster and ensemble means and standard deviations:

Geopotential 1000 and 500 hPa

Temperature 850 and 500 hPa

Validity: for means +72 hour to +168 hour forecasts at 12 h interval, for ensemble standard deviations +0 hour to +240 hour forecasts at 12 h interval

Tubes

Geopotential 1000 and 500 hPa

Temperature 850 and 500 hPa

+48 to +96h for the 96 h tubing, +96 to +144 h for the 144 h tubing, +72 to +168h for the 168 h tubing, +144 to +192 h for the 192h tubing and +192h to +240h for the 240h tubing.

Forecast probability products:

850 hPa anomaly probabilities

- cold anomaly of at least -8K
- cold anomaly of at least -4K
- warm anomaly of least +4K
- warm anomaly of at least +8K

850 hPa anomaly probabilities from day 6 to10, day 6 to 7 and day 8 to 10

- average temperature at 12 UTC more than 2K below climate
- average temperature at 12 UTC more than 2 K above climate

Precipitation probabilities over 24 hours

- at least 1 mm
- at least 5 mm
- at least 10 mm
- at least 20 mm

Precipitation probabilities from day 6 to10, day 6 to 7 and day 8 to 10

- less than 0.1 mm over the period
- mean precipitation rate less than 1 mm/day
- mean precipitation rate greater than 3 mm/day
- mean precipitation rate greater than 5 mm/day

Wind probabilities

- at least 10 m/s
- at least 15 m/s

5.4 Products on the GTS

A limited quantity of ECMWF analysis and forecast products is disseminated via the GTS. The product range is summarized in table 19.

Table 19: ECMWF products on the GTS

Northern and southern hemisphere: MSL pressure 850 hPa temperature 500 hPa geopotential Validity: 12 UTC analysis, 24, 48, 72, 96, 120, 144 and 168 hour forecasts
Tropics (35N - 35S): 850 hPa winds 200 hPa winds Validity: 12 UTC analysis, 24, 48, 72, 96 and 120 hour forecasts
Code form: FM47-V GRID (5^{o}x5^{o} resolution) FM92-Ext. GRIB (2.5^{o} x 2.5^{o} resolution)

5.5 Data archives

Weather forecasting makes use and generates very large volumes of data that need to be stored for long periods. ECMWF operates a comprehensive data service from its archives. In particular, it maintains an archive of level III-A atmospheric data in support of projects associated with the WMO World Climate Research Programme.

This includes observations, analysis, forecast and also research experiments. ECMWF has accumulated 150 Tbytes (150,000,000,000,000 characters). These data represent a valuable asset, providing a detailed record of worldwide weather and weather forecasts over the past 25 years. To accommodate these data, ECMWF has a dedicated Data Handling System. In order to manage and access this large archive, ECMWF has developed a dedicated software: the Meteorological Archive and Retrieval System (MARS). Data is stored in standard formats agreed with the World Meteorological Organisation, namely GRIB format for meteorological fields and BUFR format for meteorological observations. Retrievals can be easily prepared making use of a pseudo-meteorological language.

5.6 Access to archived data

All authorised users within Member States and Co-operating States can access ECMWF's archive and retrieve data. This can be done either through the lines between ECMWF and Member States or through the Internet. For research or education purposes from states which are not Member States or Co-operating States. This service is provided at handling cost. Access for commercial purposes is also possible, but then the request must go via one of the Member States.

5.7 Retrieving data from the ECMWF archives

Analysis and forecast values are available for every 6 hours, for surface parameters every 3 hours up to the 72 hour forecast range. The exact value of these parameters might be affected by the way data is selected, interpolated and presented.

5.7.1 Temporal resolution

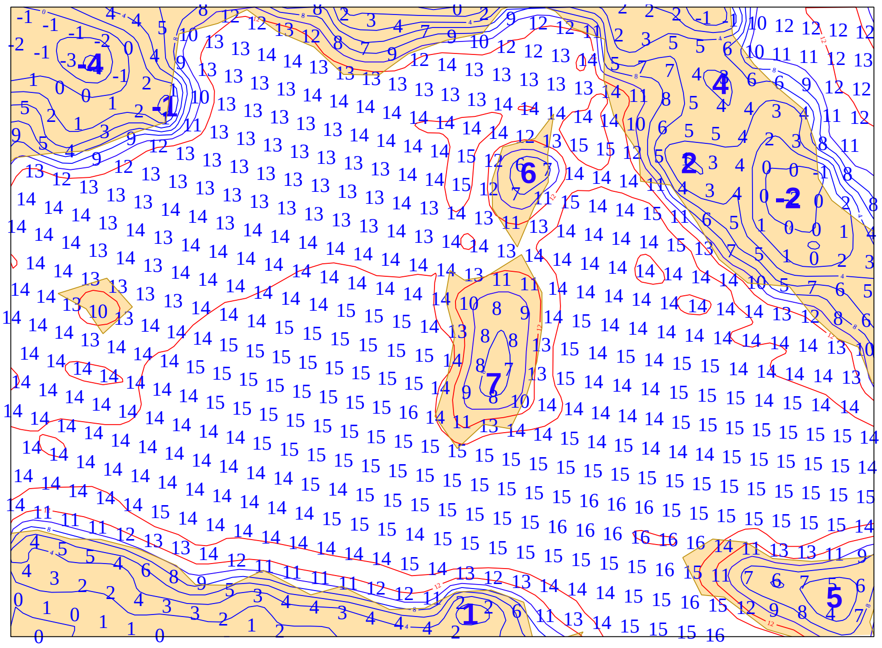
The range of the daily variation of the 2 metre temperature is best estimated by retrieving the forecast daily maximum and minimum values since the forecasts valid times do not necessarily coincide with the times of the local diurnal extremes.

Precipitation forecasts are time integrated values for the last three hours, after 72 hours every 6 hours. No information about the occurrence of precipitation at the specific UTC times can therefore be deduced. The same applies to the other parameters which are accumulated, such as evaporation.

5.7.2 Horizontal and vertical resolution

The ECMWF forecast products can be retrieved at a wide range of resolutions, from coarse lat-lon grids to the original reduced Gaussian grid of about 60 km. The data can be retrieved both from model, pressure and isentropic levels.

Two meter temperature forecast T511 Thursday 1 February 2001 12 UTC+60h



Two meter temperature forecast T255 Thursday 1 February 2001 12 UTC+60h

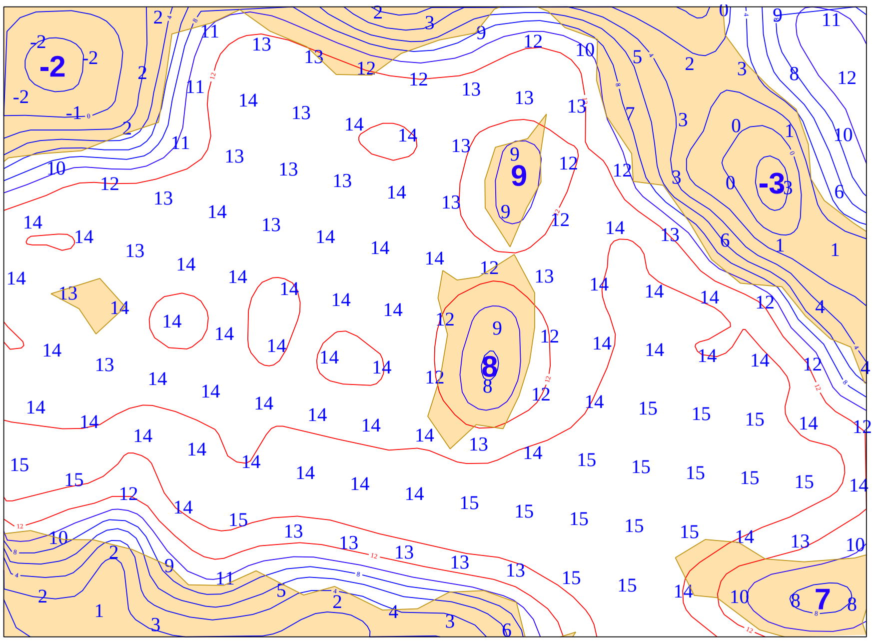


Figure 20 : The a) T319 and b) T159 gaussian grid data. There are 4 grid points over Corsica in the T319, none in the T159. The same for Sardinia is 8 and 2 respectively.

The reduced Gaussian grid values should not be considered as representing the exact location of the grid point, but as a mean within a two– or three–dimensional grid box. This is particularly important for precipitation forecasts where the variance of the observations within the grid area can be as large as the area average. Any comparison or verification should then be against some spatial average around the grid point. For near surface parameters the distinction between land and sea points may be crucial, for example for 2 m temperature, precipitation or 10 m wind. Note that the 1.5° x 1.5° lat-lon grid value are point values interpolated from the reduced Gaussian grid and do not represent a mean over the 1.5° x 1.5° lat-lon area.

5.7.3 Orography

As mentioned in ch. 2.2.1, valleys and mountain peaks are smoothed out by the model orography. Due to this difference the direct model output of 2m temperature represents an altitude significantly different from the real one. A more representative height might be found in one of the nearby grid points. Any remaining discrepancy can be overcome by a correction using the Standard Atmosphere lapse rate, or statistical adaptation (see ch.7.4)

5.7.4 Islands and peninsulas

Some small islands and narrow peninsulas are not represented in the model with land points (see 21.2). Also here statistical interpretation schemes might be beneficial.

5.7.5 Interpolation

Repeated interpolations, horizontally or vertically, will smooth the fields and dampen the extreme values. Graphical systems also introduce a slight smoothing. This might in some applications, like upper air fields, have a positive effect on the forecast quality, but for surface fields it might give unrealistic values. The use of the model's own reduced Gaussian grid is highly recommended. If, due to lack of archive storage or limitation on the telecommunications links, compromises have to be made with the retrieval of other fields, it is suggested that upper air fields, in particular from the EPS are retrieved with a coarse resolution, for example 5 x 5 deg, to allow for high resolutions of the near surface weather parameters.

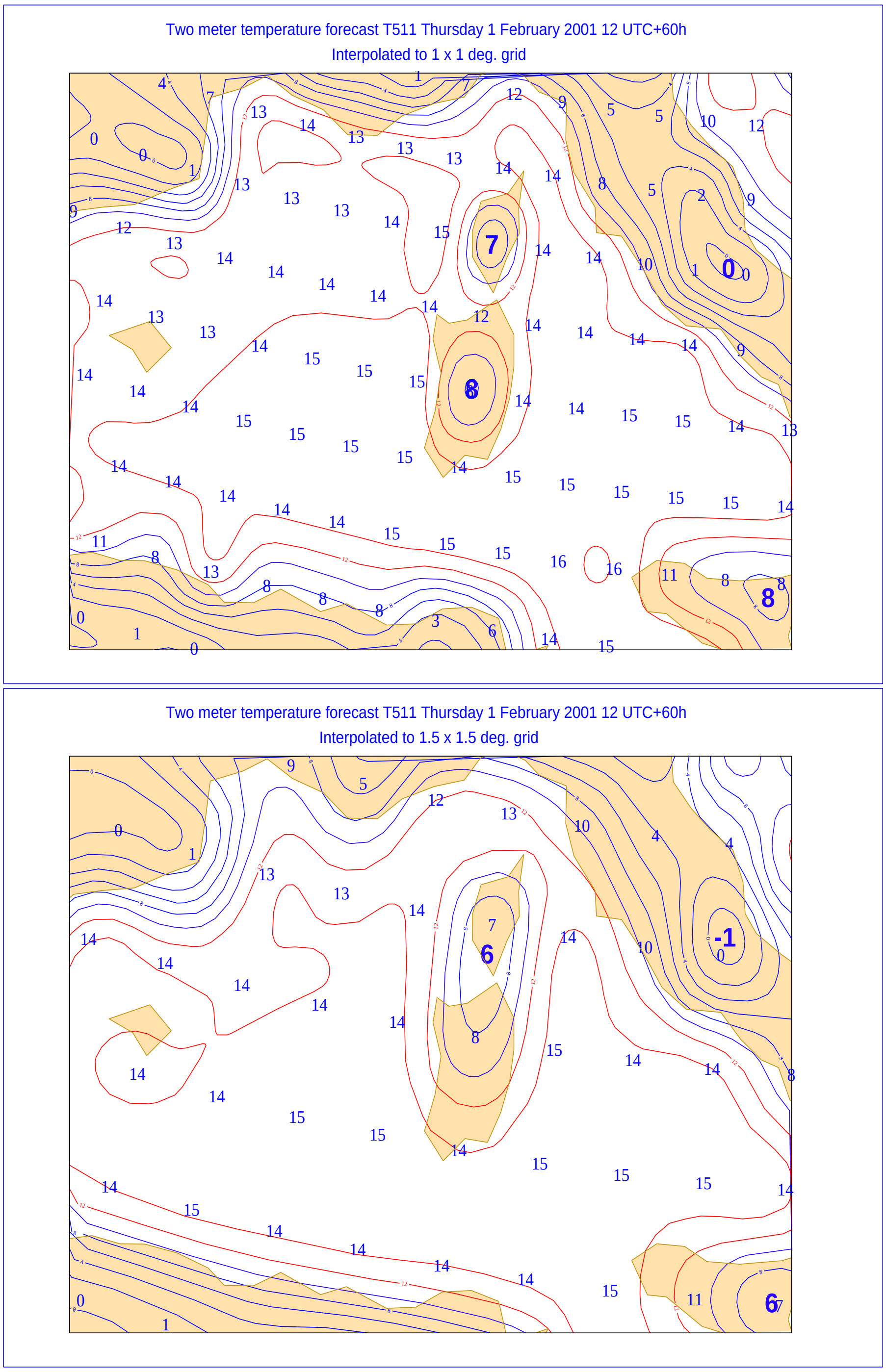


Figure 21 : The impact of interpolation to a) 1.5 deg. and b) 2.5 deg. latitude-longitude grids. Note than in contrast to the previous figure b, the 1.5 deg. grid makes the values from Sardinia “spill over” into Corsica due to the interpolation.

6 The verification of ECMWF forecasts

As part of the monitoring and validation of the model performance, a wide range of verification statistics is produced at the ECMWF. Some are mathematical and measure the accuracy of the forecasts or their correlations, others measure their skill by using reference forecasts and others still focus on the utility aspects.

6.1 The standard verifications of deterministic forecasts

At ECMWF several types of statistical verification scores are computed for a number of areas and parameters, and stored in a historical data base. Most of them have been defined in agreement with other NWP centres and are regularly exchanged following WMO/CBS recommendations. The most common are the Root Mean Square Error (RMSE), here denoted **E**, the root mean square difference between forecast and analysis

$$E^2 = \overline{(f-a)^2}$$

and the Anomaly Correlation Coefficient (ACC), the correlation between the forecast and analysed anomalies.

$$ACC = \frac{\overline{(f-c)(a-c)}}{A_f A_a}$$

Both scores are highly flow dependent. When the flow is zonal the RMSE will take small “good” values, during highly anomalous situations the quite large and “bad” values. The opposite is true for the ACC which can score quite badly in zonal situations when the positions of shallow troughs are out of phase. In blocked situation the high degree of anomaly can compensate for rather significant errors. The RMSE takes higher values during winter than summer, the ACC shows less of seasonal variations.

The interpretation of these two scores is not trivial; for medium range forecasts, in contrast to short-range forecasts, it is not always the case that the lower the RMSE and the higher the ACC, the better the forecast. To understand why, let us decompose the RMSE into terms which measure different aspects of what determines the value of RMSE.

6.1.1 RMSE – a simplified analysis

If **f** and **a** are the forecasts and the verifying analyses, and **c** the climatological value of the day of verification, then assuming there is no bias in the forecasts, the square of RMSE can be written

$$E^2 = \overline{(f-a)^2} = \overline{(f+c-c-a)^2} = \overline{(f-c)^2} + \overline{(a-c)^2} - 2\overline{(f-c)(a-c)}$$

where the overbar denotes averages in time, i.e. over a large number of forecasts, and space. This can be written:

$$E^2 = A_f^2 + A_a^2 - 2cov((f-c)(a-c))$$

The first term on the r.h.s. $\mathbf{A_f^2}$ represents the forecast variance around climate. It depends on the realism of the atmospheric model. The second term $\mathbf{A_a^2}$ is the analysis variance around climate. It is set by the character of the atmospheric flow during the period.

The same analysis can be conducted in graphical form in a phase–space, using vector algebra where the observed anomaly **a-c** and the forecast anomaly **f-c** are represented by vectors of length $\mathbf{A_f}$ and $\mathbf{A_a}$ separated by an angle **ß**. It can be shown that the ACC=cos(**ß**) and that the distance **a–f** is proportional to RMSE (Persson, 1996)

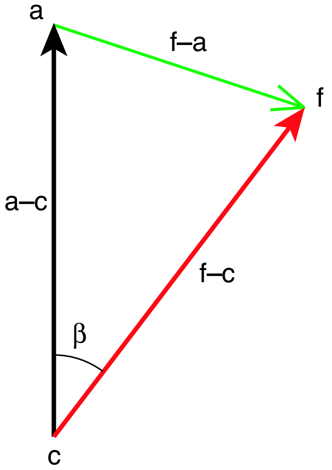


Figure 22 : A schematic representation of the relation between the magnitude of RMSE and ACC and the level of anomaly (a-c respective f-c). The vectors **f-c** and **a-c** represent the average level of dynamic variability. The value **E** will be affected by the length of these vectors.

For a realistic model these two terms should be equal, i.e. the model’s variance around climate should be the same as the atmosphere’s or at least the analysis. Since both $\mathbf{A_f}^2$ and $\mathbf{A_a}^2$ vary strongly with seasons the RMSE tends to have a large seasonal variation with maximum in winter and minimum in summer.

The third term represents, in some sense, the “skill” of the forecasts. It measures the covariance between forecast and observed anomalies. For forecast ranges without any predictive skill, this last term is zero. With $\mathbf{A_f}=\mathbf{A_a}$ we get an expression for an upper limit of the average RMSE values, the so called “error saturation level, the limit where any informative value in the forecast is lost. This also follow from the vector diagram above for the case when β=90

$$E_{saturation} = A_a\sqrt{2}$$

6.1.2 Understanding RMSE verifications

Before the NWP subjective forecasts tended to be cleverly quasi-linearly extrapolated persistence forecasts which, while preserving a realistic variation around climate, with increasing range approached a climatological statement. Whereas a persistence forecast, which always will display the full variability of the atmosphere, will approach the higher saturation level, the subjective forecasts tended to approach an error level well below.

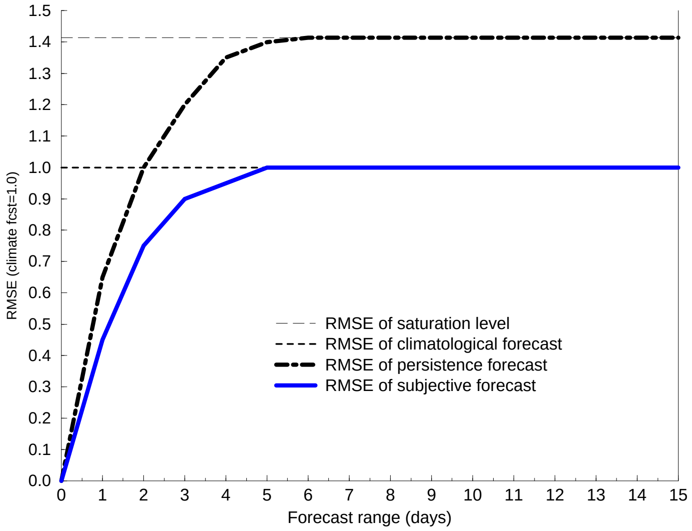


Figure 23 : A schematic representation of the pre-NWP accuracy of weather forecast measured in RMSE with the climatological variance as norm.

Like a persistence forecast, a NWP forecast will with increasing range, approach the error saturation level. A forecaster, just following the NWP, would then be less accurate than a a pre-NWP forecast and even worse than a climatological forecast. When the RMSE=$\mathbf{A_a}$ the forecast has the same accuracy as a climatological statement (and the ACC=50%)

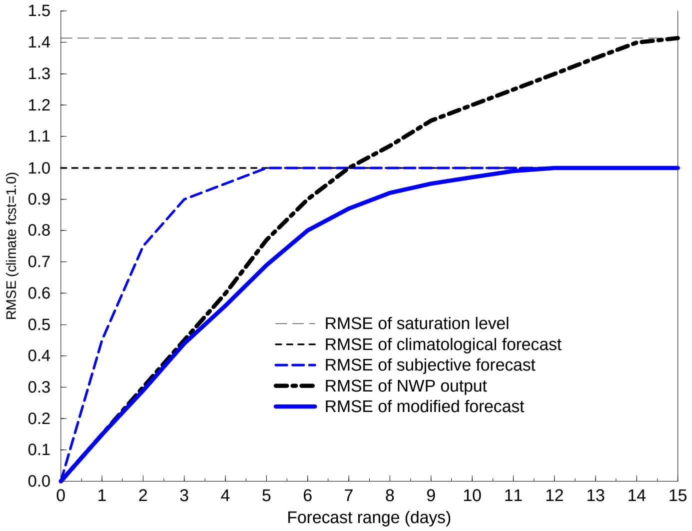


Figure 24 : Schematic representation of the accuracy of NWP based weather forecasts with the variance around climate as norm.

The art of forecasting is to relay on the latest observation in the very short range, approach climatology for very long ranges and in between judge which flow patterns are more likely to verify, which often means those of the largest scales.

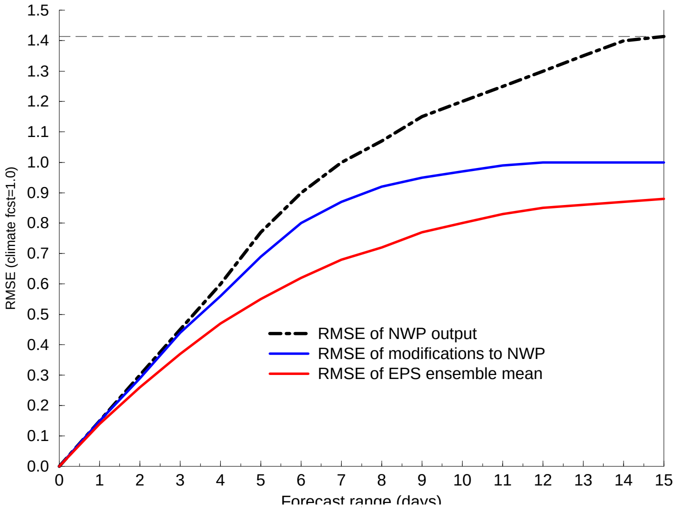


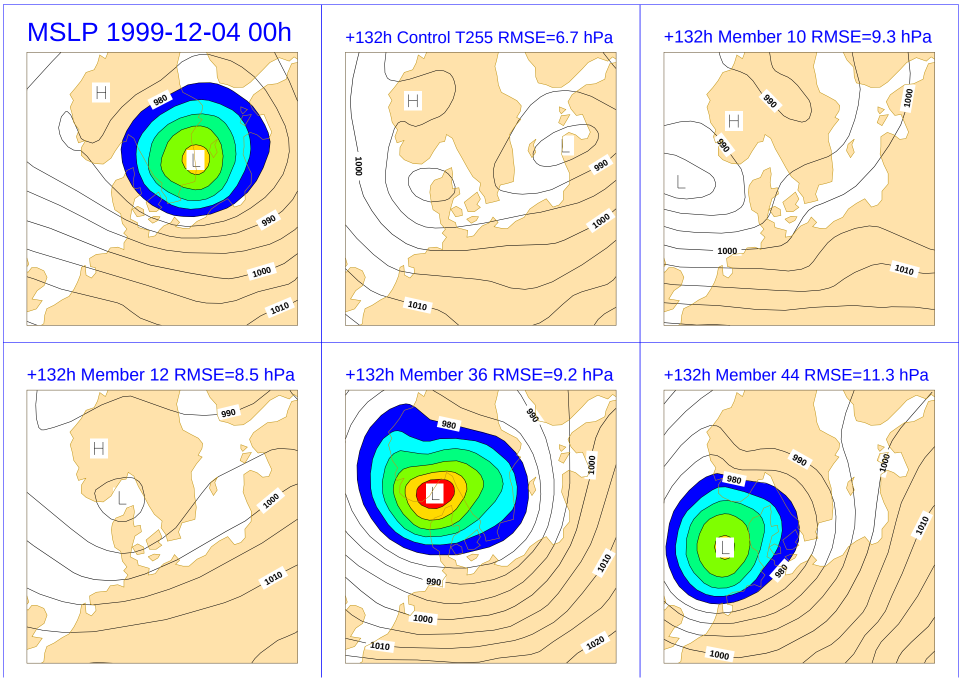
Figure 25 : A schematic representation of the accuracy of weather forecasting using the Ensemble Prediction System.

Whereas a forecaster can only apply his experience which scales are predictable at a certain range, the ensemble mean from the EPS is able, at every instant, to determine which particular atmospheric features are predictable at different locations. The EPS can also, although will be outside the ability of the ensemble mean to appreciate, suggest which features have a certain risk of occurring, something the purely statistical-empirical methods are unable to do.

6.1.3 Interpretation of the RMSE

Since the general level of RMSE also depends on the range of atmospheric variability, changes in the RMSE level, from one season or year to another, might therefore not necessarily be due to changes in the model characteristics, but to the nature of the atmospheric flow. Comparing verifications is therefore really only possible when the general variability of the atmosphere has been on the same level.

The same is true for individual forecasts. A phase error of half a wave length or more will score worse than if the system had not been forecast at all. The RMSE will punish the former forecast twice: for *not* having a low where there is one, and for *having* a low where there is none. .



MSLP 1999-12-04 00h

+132h Control T255 RMSE=6.7 hPa

+132h Member 10 RMSE=9.3 hPa

+132h Member 12 RMSE=8.5 hPa

+132h Member 36 RMSE=9.2 hPa

+132h Member 44 RMSE=11.3 hPa

Figure 26 : On 4 December 1999 Denmark was severely hit by an extremly forceful storm (upper left). An experimental EPS was run with higher resolution, T255. This is a selection from the +132 h forecasts. The Control is in upper centre followed by four members. The RMSE score is noted to the right in the heading. Not having the storm yielded lower RMSE, than having the storm with minor position or phase speed errors.

6.1.4 Interpretation of the ACC

Since the ACC measures the correlation between anomalies, it is sensitive to similarities in forecasts and analysed patterns, rather than their absolute values. In contrast to RMSE the ACC has a tendency to score higher (small **ß**) during meridional flow situations (large $\mathbf{A_a}$ and $\mathbf{A_f}$) and lower (large **ß**) during periods of predominantly zonal flow (small $\mathbf{A_a}$ and $\mathbf{A_f}$). While the RMSE is rather sensitive to errors in the position and intensity of strong vortices, the ACC might easily score badly in a generally shallow field, if the forecast and observed positions vortices are out of phase. ACC displays a weaker seasonal and annual variability than RMSE. From synoptic-empirical investigations it has been found that the level of ACC=60% corresponds to the limit where the forecast does not exhibit any significant skill. It can be shown mathematically that the limit ACC=50% corresponds with the forecast range where a climatological forecast would score as well in RMSE as a categorical forecast.

6.1.5 Measures of dynamic activity

Comparing verifications between different models or different versions of the same model, is really only possible when their general variability is on the same level. It is therefore important to compare the statistical scores with the variance of the forecast and analysed fields. What looks statistically "good" might be synoptically “bad”, what looks statistically "bad" might be synoptically “good”.

Since intense cyclogeneses and blockings become more difficult to forecast with increasing lead time, any forecast system which gradually, during the forecast, decrease their frequency or amplitude, will therefore display better statistics, in particular using the RMSE.

Different variance measures are used to make sure that the dynamical activity in the ECMWF forecasts is the same as the one observed throughout the forecast period. One method calculates the daily variance of the field over a certain area, another the monthly or quarterly variance for every grid point in an area, and a third method calculates the RMS of the 12 or 24 h change in the forecasts. All these variance measures are compared with the corresponding variances from analysed fields.

Because the definitions of RMSE and ACC do not involve any threshold such as a level of error that makes the difference between a useful and a useless forecast, they belong to the family of *continuous* scores. Example of *categorical* scores will be discussed in section 6.2.

6.2 Hit rate and False alarm rate

Verification measures like the RMSE and the ACC will value equally the case of an event being forecast, but not observed, as an event being observed but not forecast. But in real life the failure to forecast a storm that occurred will have much more dramatic consequences than forecasting a storm that did not occur. To assess the forecast skill under these conditions another type of verifications must be used.

For any threshold (like frost/no frost, rain/dry or gale/no gale) a forecast can be simplified to a *yes/no* statement (categorical forecast). The observation itself can be put in one of two categories (event observed/not observed). Let **H** denote “hits”, i.e. all correct yes-forecasts - the event is predicted to occur and it does occur, **F** false alarms, i.e. all incorrect yes-forecasts, **M** missed forecasts (all incorrect no-forecasts that the event would not occur) and **Z** all correct no-forecasts. Assume altogether 100 forecasts of this type with **H+F+M+W=100.** A perfect forecast sample is when **F** and **M** are zero. The ratio of observed to non-observed cases **(H+M)/(F+Z)** is determined by the verification sample (sample climate) and is independent of the forecast quality.

Table 20: A forecast/verification table

forecast\obs	observed	not obs
forecast	H	F
not forecast	M	Z

The Proportion of Perfect Forecasts is **PPF=(H+Z)/100.**

The Hit Rate **HR=H/(H+M)**, the proportion of perfect forecasts when the weather occurred.

The False Alarm Rate **FAR=F/(F+Z)**, the proportion of forecasts of the event when it did not occur. HR and FAR can be combined into a diagram, the Relative Operating Characteristics (ROC) to compare deterministic and probabilistic forecasts (see 6.2.4 below)

The Probability Of Detection, POD=H/(H+M), is the proportion of perfect yes-forecasts.

The Frequency Bias Index, **FBI=(H+F)/(H+M)**, measures the relative frequency of occurrence in the forecast with respect to observations that is not measured by the True Skill Score. It serves the same function as the dynamic activity in the forecast verification using the RMSE or the ACC

The False Alarm Rate **FAR=F/(F+Z**) is the proportion of yes-forecast of the event when it did not occur.

A very simple measure of success of categorical forecasts is the difference **POD-FAR** which is known as the Hansen-Kuiper or True Skill Score. Among other properties, it can be easily generalised for the verification of probabilistic forecast (see 6.3 below)**.**

6.3 Verification of probabilistic forecasts

In contrast to a deterministic forecast an individual probabilistic forecast can never be "right" or "wrong", except when 100 or 0% have been stated. Due to its nature the performance of the EPS can therefore *only* be evaluated from large samples of forecasts. However, like the deterministic forecast system the performance is determined not only by its predictive accuracy but also its ability to account for the variability of the atmosphere.

6.3.1 The Brier score

The most common verification method for probabilistic forecasts, the Brier score **BS** is similar to the RMSE, measuring the difference between a forecast probability of an event **p** and its occurrence **o,** expressed as 0 or 1 depending on if the event has occurred or not. As with RMSE, the lower the Brier score the “better”

$$BS = \overline{(p-o)^2}$$

A Brier Skill Score (BSS) is conventionally defined as the relative probability score compared with the probability score of a reference forecast

$$BSS = (BS_{ref} - BS)/(BS_{ref})$$

The BS score can be decomposed in a similar way as the RMSE, yielding three terms which help to explain different aspects of the scoring system

$$BS = \overline{(p_k - c)^2} + \overline{(c_k - c)^2} + (1-c) \cdot c$$

where a sample of N forecasts has been divided into categories (k=1,2,3...T) each comprising $\mathbf{n_k}$ forecasts of average probability $\mathbf{p_k}$. The observed frequency in each frequency is $\mathbf{c_k}$ and **c** the observed frequencey of the whole sample. The first term expresses the degree of *reliability*, the second the *resolution* and the third the *uncertainty* of the forecasts (Atger, 1999)

6.3.2 The reliability, resolution and uncertainty

The *reliability* term measures the ability of the system to forecast accurate probabilities. After a large number of, for example 20% probability forecasts, the predicted event should verify for 20% of the forecasts, not more, not less. The degree of reliability is displayed in a diagram where the x-axis is the forecasted probability **p** and the y-axis the frequency it occurs on those occasions.

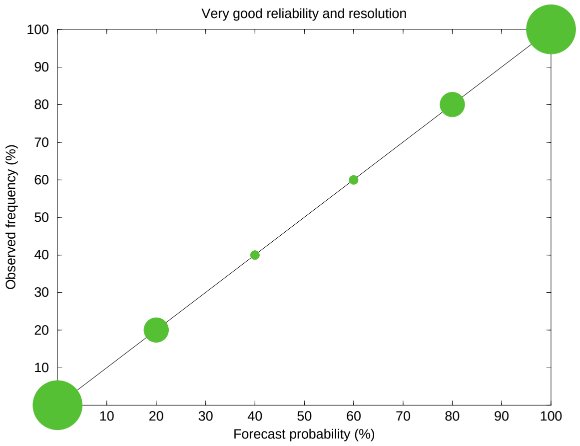


Figure 27 : An example of very good reliability and resolution. The area of the circles are proportional to the number of forecasts.

Ideally the distribution should lie along the 45˚ diagonal. When the low risks are underestimated and the high risks overestimated, the forecasts have been *over-confident* and the distribution is flatter than 45˚. On the other hand, if the low risks have been overestimated and the high risk underestimated the forecasts are *under–confident* and the distribution is steeper than 45˚

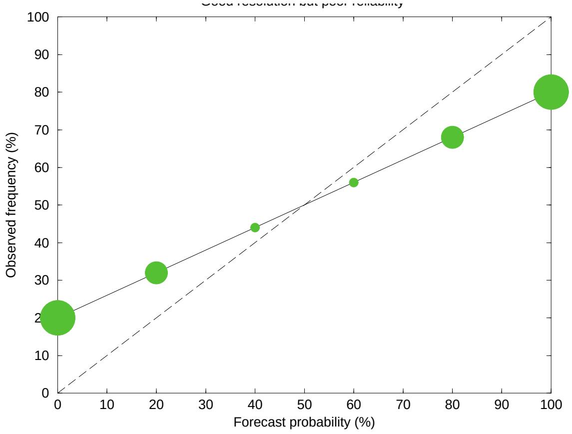


Figure 28 : An example with good resolution but poor reliability since the high and low probabilities are under respective overforecast.

The *resolution* indicates the average square difference between the observed frequency in each category and the mean frequency observed in the whole sample. It indicates the ability of the forecast system to correctly separate the different categories, whatever the forecast probability. For a given reliability, the resolution thus indicates the sharpness of the forecast. The maximum resolution corresponds to a deterministic forecast (only 0% and 100% are forecast), the minum resolution corresponds to a climatological forecast (only one probability is forecast).

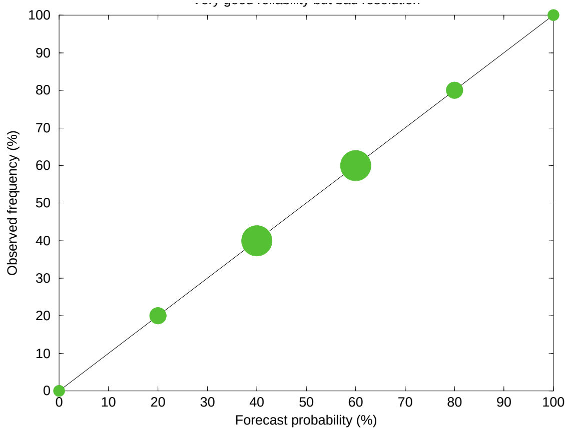


Figure 29 : An example of good reliability but poor resolution. The high and low probabilities are rarely forecast, whilst the near climate probabilities are forecast at a majority of times.

For operational purposes, the resolution term is the most relevant, since the reliability, as any bias, can generally be improved by a calibration. However, this is obtained at the expense of sharpness. The resolution is not modified by the calibration if the number of categories remains the same and the EPS error characteristics remain stable from one season to another.

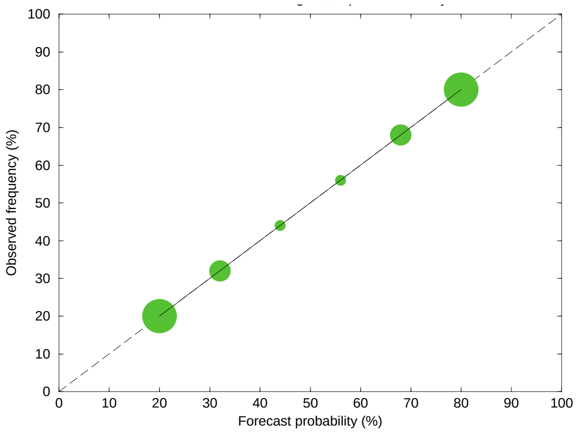


Figure 30 : By multiplying the forecast probabilities in figure 27 by 0.6 and adding 20 the forecasts will be calibrated to yield a very good reliability, but with a reduction in resolution.

The *uncertainty* is the variance of the observations, indicating the intrinsic difficulty in forecasting the event during the period. It is also the probability score of the sample climatology forecast. The uncertainty is obviously independent of the forecast system: eing the same for the reference forecast and the forecast under evaluation, it plays no role in the skill score - but it can be shown to be an upper bound for the resolution.

The Probability skill Score can thus be decomposed into two terms, positively orientated, indicating (i) the skill due to the reliability and (ii) the skill due to the resolution:

$$PSS = (Reliability_{ref} - Reliability)/PS_{ref} + (Resolution_{ref} - Resolution)/PS_{ref}$$

Reliability and resolution are independent. For example, if the observed frequency is 90% in the 10% probability category, and 10% in the 0% probability category, the resolution is high but the reliability is poor.

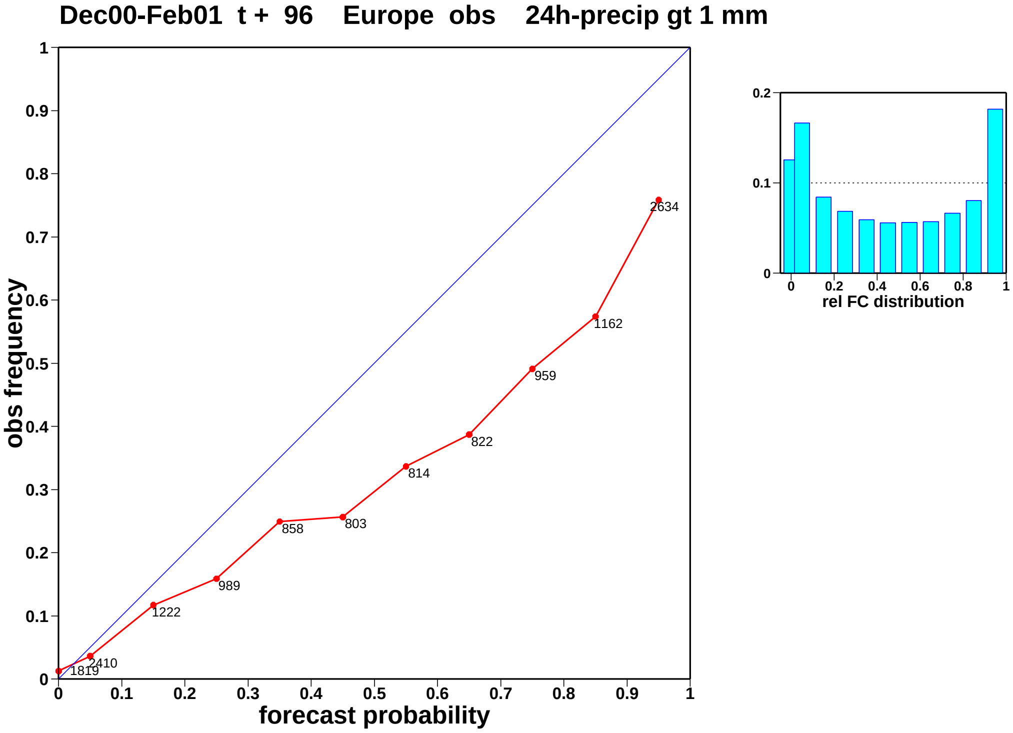


Figure 31 : An example of a reliability diagram of the probability that the 72 to 96 hour forecast precipitation will exceed 1 mm during winter 2000-2001. The reliability is not perfect since high probabilities are over-estimated. The diagram to the right shows the frequency of forecast probabilities, indicates a fair resolution. By a calibration, multiplying all forecast probabilities by 0.7, the reliability curve can be made to approximately coincide with the diagonal. But then no probabilities exceeding 70% would be forecast

6.3.3 Talagrand diagram

In an ideal EPS system the verifying analysis is equally likely to lie between any two ordered adjacent members, including the cases when the analysis will be outside the ensemble range on either side of the distribution. In such an ideal system the long term distribution should be flat with equally many verification in each interval. A measure of the degree this is satisfied can be made by a so called *Talagrand diagram*.

In reality the distribution in the Talagrande diagram is slightly U–shaped with around 10% of the analyses verifying outside the ensemble. This is an indication that there is not enough spread in the EPS.

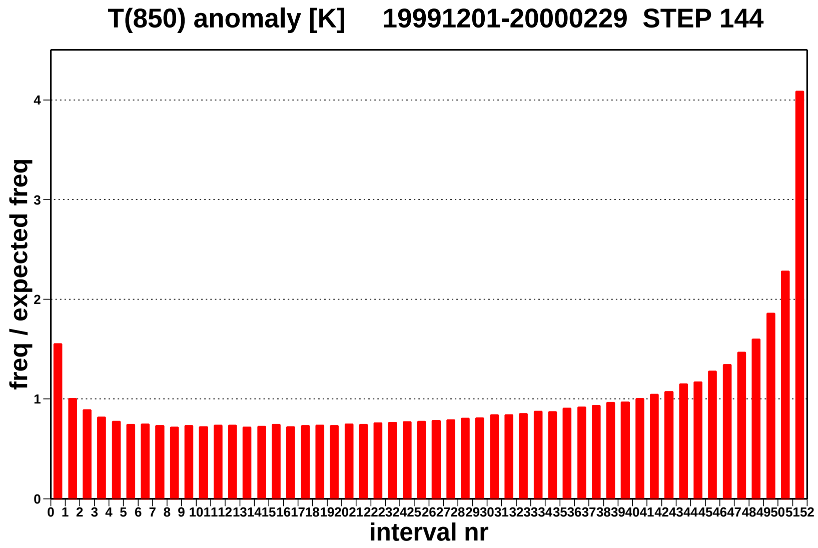


Figure 32 : The x-axis in a Talagrand diagram constitutes all the members in the ensemble, ordered according to some norm, for example the 850 hPa temperature in a certain point. For an EPS with 50 members there are 51 intervals including the unlimited intervals at the edges of the distribution. The y–axis indicates the number of cases the analysis is found between two ordered adjacent members on the x-axis. The U-shape indicate that the ensemble does not spread out sufficiently, the tendency to have a J-shape that the system has a slight cold temperature bias.

Improving the spread is a necessary, but not sufficient condition; a random sampling of weather parameters from the same season from the last 50 years would provide a flat distribution in the Talagrand diagram, but of course with no predictive skill. Any improvement in spread in EPS forecasts must come through improving the system’s ability to identify all the possible uncertainties in the initial conditions. Due to the limited number of EPS members, the verifying analysis will be outside the ensemble range. For a system with 50 members this will happen 2/51 (~4%) of the time. At present the proportion of outliers is 2–3 times this value.

6.4 Skill scores

While all the previous scores have been absolute, in the sense that they are not related to any comparison, the following ones measure the skill relative another method, which can be climate or an alternative model.

6.4.1 The Brier Skill Score

The Brier Skill Score (BSS) is computed with reference to the Brier Score of another probabilistic forecast, like for example a climatological distribution.

$$BSS = \left(1 - \frac{BS}{BS_{cl}}\right)$$

6.4.2 The ROC diagram

Probabilistic forecasts can be transformed into a categorical yes/no forecasts defined by some probability threshold. The corresponding hit rates **H** and false alarm rates **F** can be computed and entered into a ROC–diagram with **H** defining the y-axis, **F** the x-axis. The closer the **F, H** is to the upper left corner (low value of F and high of H) the higher the skill. A perfect forecast system would have all its points on the top left corner, with H=100% and F=0. .

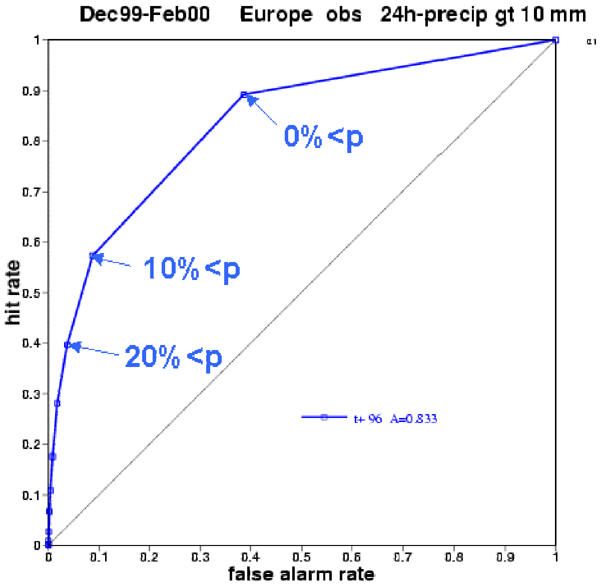


Figure 33 : An example of a ROC diagram. The highlighted points on the curve illustrate the relation between the probabilities and the rate of hits and false alarms. If a yes/no decision is made on a 20% level, then the hit rate is likely to be 0.4, the false alarm rate less than 0.1.

In reality a non-perfect system will have its values on a long convex curve pointing to the upper-left corner (the “ROC curve”). The area between ROC curve and the y-axis and the y=100% axis measures the skill of the forecasts

The ROC curve enables a comparison between a probabilistic and a deterministic forecast system. If the deterministic **F,H** value lies over the ROC curve, the deterministic system is more skilful. However, in terms of utility, greater advantages might be gained from the probabilistic information in the ROC curve. It takes very good deterministic forecasts to be more useful than probabilistic ones.

6.4.3 The Rank Probability Score (RPS) and the Continuous Rank Probability Score (CRPS)

With the output from the EPS the probability of any event (according to some threshold) can be calculated and subsequently verified by the Brier Score. The *mean* of a range of such verifications by itself defines a new score, the Rank Probability Score (RPS). If the intervals between the consecutive thresholds become infinitesimal small (or the number of thresholds infinitely large) an *integrated value* of the Brier Score can be defined, the Continuous Rank Probability Score (CRPS).

6.5 Decision making from meteorological information

Ultimately the motivation for weather forecasts are the guidance they give in decision makings, their *utility*. A forecast system that provides good scores normally also provide good guidance for a wide range of needs. But the exceptions are numerous. A forecast system that over-forecasts the occurrence of rain will score badly, but will be very useful for anyone who is sensitive for rain. On the other hand, a forecast system which might provide useful guidance for most applications, can be without value for other applications because the skill is *not be high enough*.

6.5.1 Decision making from categorical forecasts

A common situation in decision making is to weight the cost **C** of taking a protective action against the risk of making a loss **L** when no protection is made. Thanks to the rich statistical data meteorological decision making can easily be mathematically analysed. If the climatological risk of adverse weather is **c** then constant protection is favourable if the economic risk **cL** is larger than the cost of protection **C**. The breaking point occurs when **cL=C** or **c=C/L**.

The "cost–loss ratio", **C/L**, is an important indicator of the sensitivity to weather forecast information. If **C/L < c** a permanent protection will be favourable, if on the other hand **C/L > c** it might be economic to take the risk that hazardous weather might occur.

6.5.2 A clarifying example

At a certain location it rains two days a week. The climatological risk is then 2/7. Someone organizing an outdoor public event hesitates between having to pay for rain protection, costing $200, or taking a chance, in which case if there is rain, his loss will be $1000. With a climate with two rain days out of seven, his *expected* daily loss would be slightly above $280 (=2000/7) so he is well advised to invest in protection.

However, if the protection cost had been $400, he would have been wise to take a chance since this cost exceeds the expected loss. The same would have been the case if the loss in case of rain would only have amounted to $500. His expected loss would then only be $140 (=1000/7).

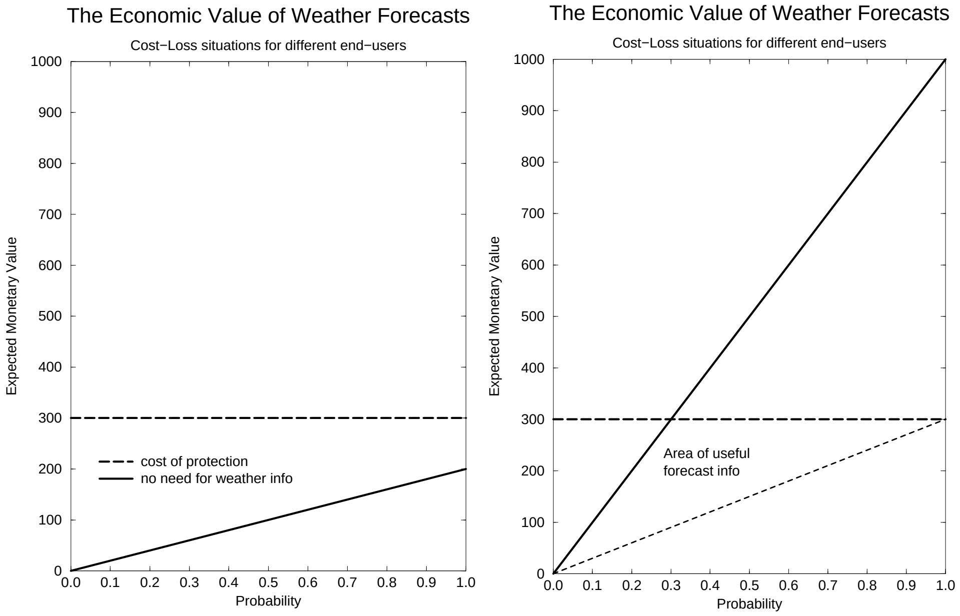


Figure 34 : A schematic diagram of the cost/loss situation for a decision maker facing a possible loss of $200 in case of adverse weather (left), a loss of $1000 (right). The cost of protection is $300 in all cases (horizontal dashed line). The x-axis indicate the risk of adverse weather, the y-axis the economic expenses. In a situation when the cost of protection is always higher than the potential loss (left) there is obviously no need for weather information. Note that even perfect forecasts will be associated with losses, since the decision maker has to take protective actions (inclined dashed line or the right)

The cost/loss concept also explains why some end-users are not served by weather forecasts, even if these are of high quality. In the case the cost for protection is low compared with the value of what should be protected, permanent protection becomes beneficial, and there is no need for weather forecast information. In case the protection is so expensive compared to the value, protection can only be considered, if at all, if the forecast probabilities of bad weather are sufficiently high. The closer **C/L** is to **c** the more important weather *forecast* information becomes. For the extreme cases when **C/L** << **c** or **C/L>>c** there might be no need for weather forecast information.

6.5.3 The need of forecast information

When the loss might exceed the cost of protection, meteorological information will reduce the losses in two ways: pure climatological information will prompt the user to always take protection if the region has a climatological risk larger than 30%. Forecast information will tell when the synoptic risk is different from the climatological and reduce his losses.

7 The use of deterministic medium range forecasts

7.1 Introduction

The many meteorological parameters which is produced by the medium range forecast system do not seem to need any "interpretation". Time series of the forecast temperature, cloud cover, wind, rain etc. presented graphically for a specific location (“meteogram”), can be read off by meteorologists as well as laymen. However, care must be observed in the interpretation. The figure below shows two consecutive forecasts of the 2 meter temperature for a location in the Netherlands.

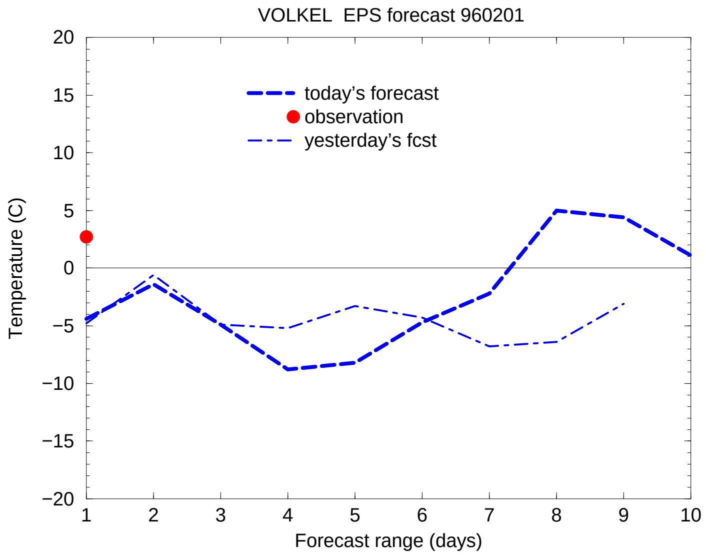


Figure 35 : The forecast 2–meter temperature development for a location in Holland in winter 1995 according to two consecutive ECMWF forecasts. The filled circle at the start of the meteogram indicates the observed 12 UTC temperature.

–The last forecast indicates a return to milder conditions a week into the forecast whereas the forecast from the day before indicated continued cold conditions. *Does this indicate that today’s forecast is less reliable?* A further potential problem might be that the temperature on the first day does not correspond with the forecast. *Does this mean that the rest of the forecast be trusted?*

7.2 What can the forecaster do?

Whereas in short range forecasting it is a useful technique to modify the NWP products quasi-linearly in light of later information, this is not possible in the medium range. The impact of an analysis change remains approximately linear only up to 48 hours. Beyond this range it is normally impossible without computer based calculations to deduce how later information ought to modify the forecast.

The forecasters have possibilities to make positive correction for *systematic* model deficiencies due to poor representativeness, limitations in the horizontal resolution or in connection to some physical processes. This is in particular true for mountainous regions where the model orography differs from the real. However, the forecasts of the large scale flow do not exhibit any significant systematic errors which can be taken into account by the forecasters.

Paradoxically, the forecasters best opportunity to add value to the forecast rests with addressing the *non-systematic* errors, in particular of the movements, positions and intensifies of synoptic features. They can add substantial value to the NWP not by adding information, but by *removing* information.

7.3 Scale and predictability

The range of meteorological scales is the same throughout the forecast; a ten day forecasts *looks* like a analysis of the atmosphere. But the predictability of these scales decrease rapidly throughout the forecast, starting with the smallest scales.

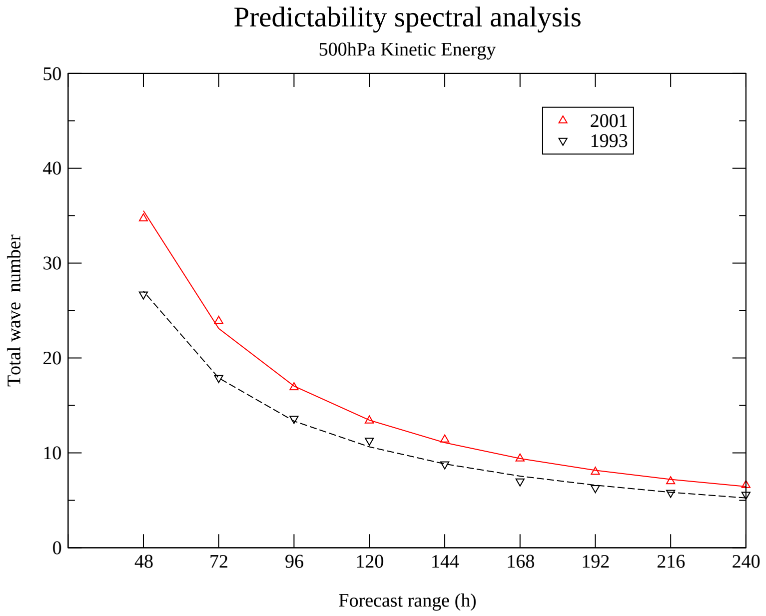
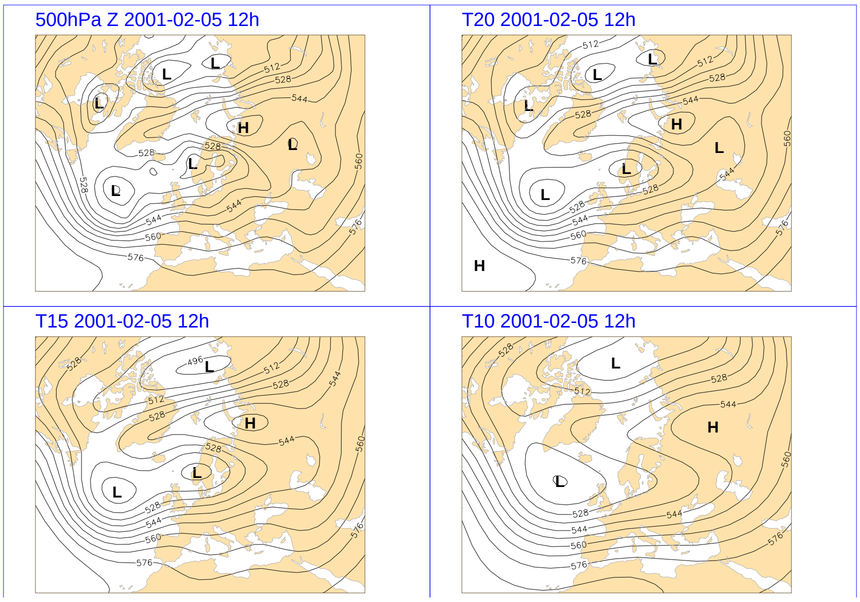


Figure 36 : The relation between the scale of an atmospheric feature at 500 hPa, measured by the number of spectral components needed to describe it, and its predictability. Forecasts beyond five days normally contain useful information only in the first 10-15 spectral components. Since 1993 the predictability, measured in this way, has increased by one day.

By relying on his experience of what is normally predictable at a certain range, the forecaster can disregard the small and unpredictable scales, but concentrate on the large scale.

Just reading off the maps or the NWP output would convey the lack of predictability in the smaller scales and inconsistency in full to the public. They might therefore lose confidence in the forecasts well before they have verified. It is of little use to issue a very confidently sounding forecast if it is not likely to verify.

There are different techniques to highlight the larger, more predictable scales and by filtering eliminate small scale and less predictable features.The most consistent way to filter away the non-predictable synoptic scales is the Ensemble Prediction System (see ch. 8)


500hPa Z 2001-02-05 12h

T20 2001-02-05 12h

T15 2001-02-05 12h

T10 2001-02-05 12h

Figure 37 : The 500 hPa flow 5 February 2001 12 UTC with different spectral truncations.

Rainfall can for example be more skilfully forecast beyond D+3 if accumulated values over two or more days are calculated. The advantage of condensing information has to be paid by the risk of losing information which on some occasion, in hindsight, might have been important.

The larger the scale an atmospheric system, the more predictable it normally is. Small baroclinic systems or fronts are well forecasted up to around D+2, cyclonic systems around D+4 and the long planetary waves around D+6. The later relate to the general weather type. The forecaster using the ECMWF deterministic forecasts will most of the time be able to make useful forecasts up to a week ahead, although the timing of cyclones or fronts might be in error.

The skill of the forecast for different phenomenon versus range as follows.

Table 21: The current skill in NWP

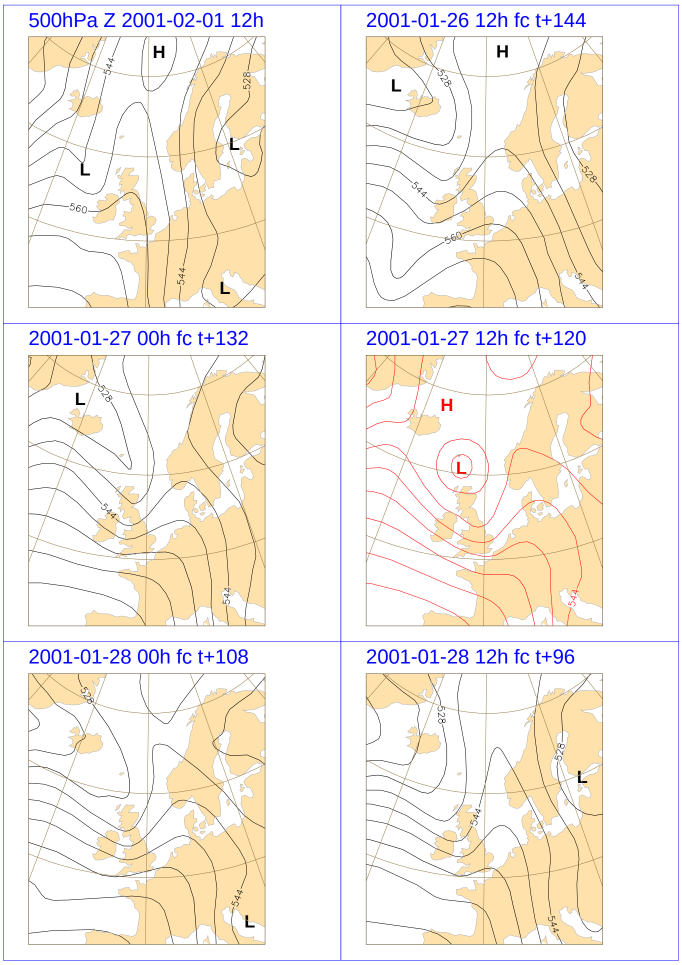
Feature	<D+3	D+3 to D+5	D+5 to D+7	D+7 to D+10
Hemispheric flow transitions	Excellent	Excellent	Good	Some skill
Blocking creation and breakdown	Perfect	Good	Fair	Low skill
Cyclones' life cycle	Perfect	Fair	Low skill	——
Fronts and 2nd developments	Good	Fair	——	——
Temperature/ wind	Very good	Skill in daily extremes	Skill in 5–10 day mean	
Acc.precip./ mean clouds	Good	Some skill	Some skill in precipitation 5–10 day acc. values	

The predictability can vary considerably from situation to situation. Sometimes a D+4 forecast can be wrong even in the large scale, occasionally the D+7 can be perfect even in details.

Considering the speed by which atmospheric systems influence each other charts plotted to be used in medium–range forecasting should have a wide geographical coverage. Whereas an European–Atlantic coverage is suitable only for a two day forecasts, three to five day forecast are best understood when also the western part of the Atlantic and the easternmost part of North America are included. For forecasts beyond five days the whole of the North American continent and easternmost Pacific ought to be included.

7.4 The day–to–day inconsistency

Changes in the forecast from one day to the other are necessary to enable a forecast system that take full benefit of new observations and modify previous analyses of the atmospheric state. Since the latest forecast is based on more recent data than the previous forecast, it is on average better. Although in most cases the changes in successive NWP forecasts are quite small, at least for the first five or six days, occasionally there might be large differences. This occurs when new observations have appeared in dynamically sensitive areas, or when these areas have moved into regions with observations.


500hPa Z 2001-02-01 12h

2001-01-26 12h fc t+144

2001-01-27 00h fc t+132

2001-01-27 12h fc t+120

2001-01-28 00h fc t+108

2001-01-28 12h fc t+96

Figure 38 : Five ECMWF 500 hPa forecasts verifying on 1 February 2001 12 UTC. All are fairly consistent and skilful, except the 27 January 12 UTC forecast, which brings in a cyclone over Scotland. This was later found to be caused by complex analysis problem of a baroclinic zone in the NW Pacific.

The fact that the model can develop synoptic features with the same overall frequency, means that there are no constraints to prevent forecast changes. In contrast to a numerical model, a forecaster should not be inconsistent since this will have a psychologically adverse effect on the public. It is normally almost impossible to determine in each individual case which one of the inconsistent forecasts is the best, if any of them.

Since non–predictable scales are those which will first be affected by any inconsistency the best advice is to try to avoid coming in the situation by having over-interpreted details in a previous NWP, scales which are normally not predictable. Cases when the inconsistency affects even the largest scales are rare, but sufficiently common to be of concern. These can only be dealt with in an EPS context (see chapter 8)

7.4.1 Consistency and skill?

There is no significant difference in quality between consistent and inconsistent forecasts. During several ECMWF Training Courses the attendees have as excercises made subjective evaluations of forecast consistency and skill and found-correlation between 29% and -20%. The reason why objective verifications show correlation around 30% is due to a statistical artefact.

To illustrate this imagine a realistic but completely *skill-less* forecasts system, for example forecasts 50 days ahead. The 37 day forecast from today compared with a 38 day forecast from yesterday both lack skill and are not correlated. Still their errors correlate 50% since they verify on the same analysis and their difference correlate also 50% with the error (see figures below). For shorter lead times, when both the skill of the forecasts increase, the consistency/error correlation drops paradoxically to typical values of 20-30% for day 5 and 6 forecasts. This is because the forecasts, with their increasing skill, also start to become mutually correlated.

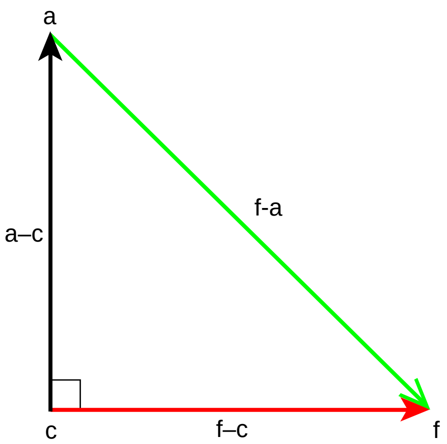


Figure 39 : With the notation used in figure 21 it is easy to understand why the maximum RMSE level for a realistic but completely skill-less forecasts f (orthogonal to the verification a) is 1.414 times the variance around climate.

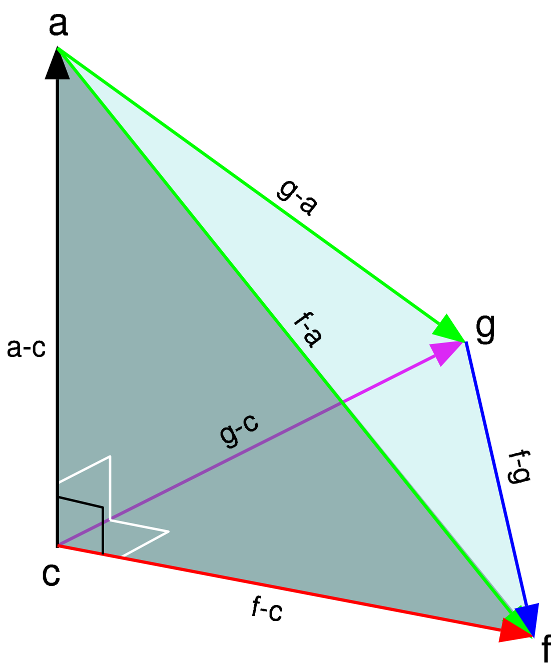


Figure 40 : The case with two realistic, but completely skill-less forecasts f and g, can be illustrated in the same way, as a three-dimensional vector diagram. The consistency then is the vector difference between the two forecasts f-g. This vector difference, together with the error vectors f-a and g-a form an equidistant triangle.

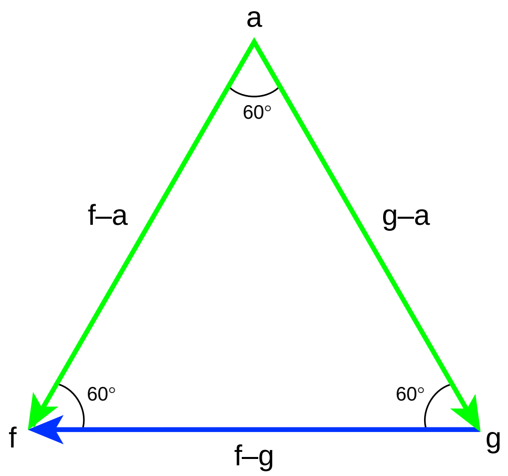


Figure 41 : The angles in an equidistant triangle are 60 deg, and since cosine(60)=0.5 the correlation between the RMS errors as well as between the consistency and errors are 50%. If we from the previous figure imagine that the skill improves, this will lead to some correlation between f-a and g-a. The angle will be >60 and consequently the two other angles will together be >120, with at least one >60. If f-a has lower RMSE than g-a, it will be the angle between f-a and f-g which becomes >60 and thus yielding correlation <50% whilst the angle between g-a and f-g will be <60% and the correation > 50%.

The level of inconsistency depend on the quality of the *preceding* forecast, not the current one (Persson and Strauss, 1995, Persson, 1997). Indeed, the correlation between the D+5/D+6 consistency and the error of the D+6 is 70-80%! Unfortunately, this can not be used for operational purposes since a consistent D+6, although of higher quality will not be better than the last D+5 - which it resembles anyhow.

7.4.2 Beware of consistent forecasts!

Experience shows that forecasters, in spite of all the difficulties, manage to handle inconsistent forecast situations well. It is in cases with several days of *consistent* forecasts, when the forecasters can find themselves in great difficulty when the NWP suddenly changes direction. In these situations, the lack of forecast alternatives might give the forecasters a false feeling of reliability, which does not prepare them for possible new developments.

7.4.3 The poor man’s ensemble approach

A look at the last 2-3 forecast will help to identify those scales which, in spite of all the “jumps” remain consistently forecast and therefore ought to be more predictable; the inconsistent parts will provide information about possible alternative developments. This “poor man’s ensemble forecast” approach, which also can involve other models, provides further possibilities to add value to the forecast.

7.4.4 Misleading “systematic errors”

It is important to realize that some non-systematic errors can easily be misinterpreted as being “systematic”. Verifications over a long period of time will show that more cut-offs are forecasted at D+6 west of the Iberian Peninsula than actually verify. This does not necessarily mean that the model is over-forecasting these cut-offs. There are cases when the D+6 fail to forecast a cut-off, and statistical studies show that approximately the same number of cut-offs are forecast as are observed - they just do not always match at a D+6 range!

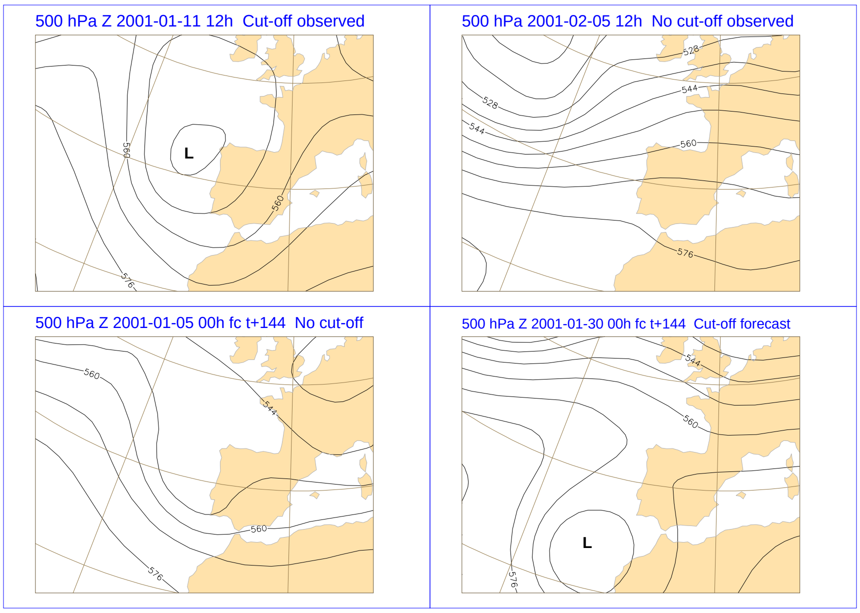


Figure 42 : The lack of skill forecasting cut-offs west of Portugal does not only involve 'misses", failures to forecast a specific cut-off (left column), but also “false alarms”, cases when a cut-off is forecast but does not verify (right column).

This misinterpretation of “systematic errors” comes back in many disguises. If, for a certain location, the number of cases with heavy rain (more than 20 mm/24 h) *forecast* on equals the number of cases *observed*, there is obviously no systematic model error. But since the heavy rain forecasts do not always verify it might give an impression that the model is forecasting heavy rain “too often”. Judging from the number of cases of obsered heavy rain, the model instead appears to forecast these event “not often enough”. Forecasts of extreme events

Extreme weather events are coupled both to the small and large atmospheric scales. The large scale extremes can be long periods of anomalous temperatures or rainfall over large areas. These situation, often coupled to persistent blocked flow, are skilfully forecast, five days or more in advance. For small scale extreme events, like heavy rainfall, strong winds and rapid changes in temperature the forecast skill decreases from day 3 onwards (see table 10).

The difficulty with extreme weather events is that the forecasting system should on one hand detect as many of these as possible, without too many false alarms. What is “too many” is ultimately a matter of subjective choice (See. ch. 6.2 and 6.5). The ECMWF has chosen to maintain a uniform variability in the model over the ten day forecast range with neither over- nor under-forecasting (see. ch.

6.1.4). This is important to make it possible for the EPS to forecast extreme developments. It also the EPS that provides the most consistent treatment of extreme event forecasting.

7.5 Statistical interpretation of deterministic forecasts

A dynamical–statistical interpretation can be produced for any particular weather parameter (predictand) e.g. precipitation, cloud, visibility, temperature, provided that historical data for the location exists. Some techniques also partly compensate for the model's systematic errors.

7.5.1 Non-adaptive interpretations

There are two traditional statistical interpretation methods: the Perfect Prog Method (PPM) and the Model Output Statistics (MOS) technique. In the PPM a statistical relationship is established between observed values of the predictand and *analysed* predictors from the free atmosphere; in the MOS a statistical relationship is established between observed values of the predictand and *forecast* predictors, both from the surface and the free atmosphere. If the model has a tendency to under– or over–forecast any predictor, this will be compensated for by the MOS technique Murphy and Katz, 1985, Glahn et al, 1991).

7.5.2 Adaptive interpretative methods

Adaptive methods, in particular the Kalman filter. It shares MOS' advantage of being able to compensate for model errors while at the same time being able to continue to work despite changes in the model characteristics. In contrast to MOS and PPM, the adaptive filter does not need any long historical data base, statistical interpretations can provide skilful interpretations 2 or 3 weeks after the start. If the model changes in any significant way, the filter will notice it and gradually, normally within 5–10 days, adjust the statistical relationship (Persson, 1991, WMO, 1992, Cattani, 1994).

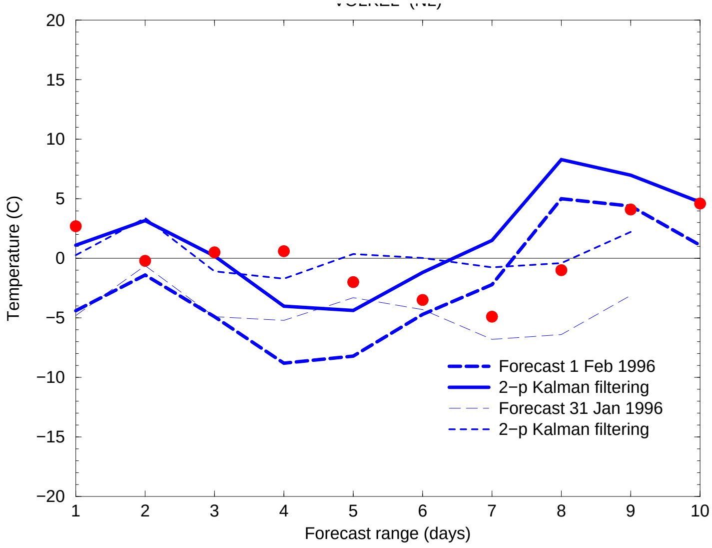


Figure 43 : Kalman filtering of the ECMWF 2 meter temperature forecasts for Vokel in winter 1996.Relying on past verification a two-parameter filter has detected a tendency to under-forecast the temperature by 1-5 degrees, depending on the forecast itself.

The most simple task is to modify the 2 meter temperature or the 10 meter wind speed which mostly have convenient statistical structures. Weather parameters like rainfall, clouds and humidity have a more complex statistical structures.

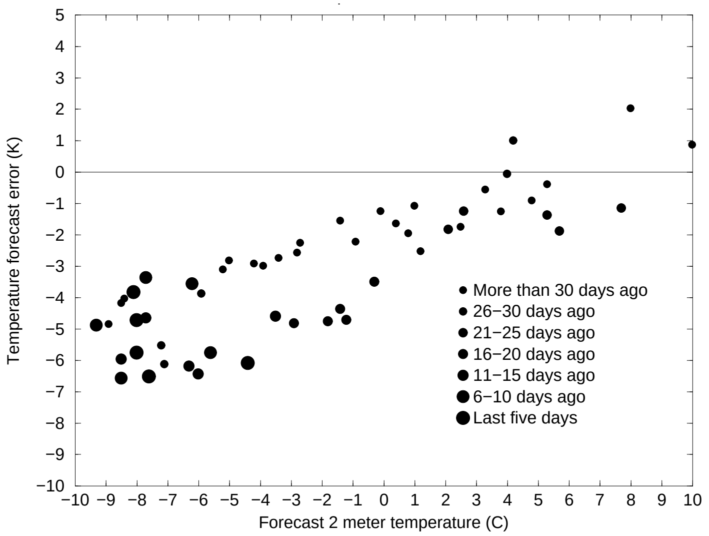


Figure 44 : The relation between the forecast 2 meter temperature value and the +24 hour forecast error for Volkel during January 1996. The larger the circle, the more recent the verification. Although there is an over-all bias of about -3 K, there is also a linear relationship which a two-parameter filter can take advantage of and making small or no correction for forecasts between 5 and 10 C, whereas forecast between -5 and -10 C are corrected with about +5 K.

It is essential to realize that it is only a 1-parameter Kalman filter which removes plain biases, i.e. mean errors independent of the forecast values. For most weather forecast parameters the mean error depends on one or several factors, in particular the forecast itself. By using 2 or more parameters different corrections can be applied in different weather regimes.

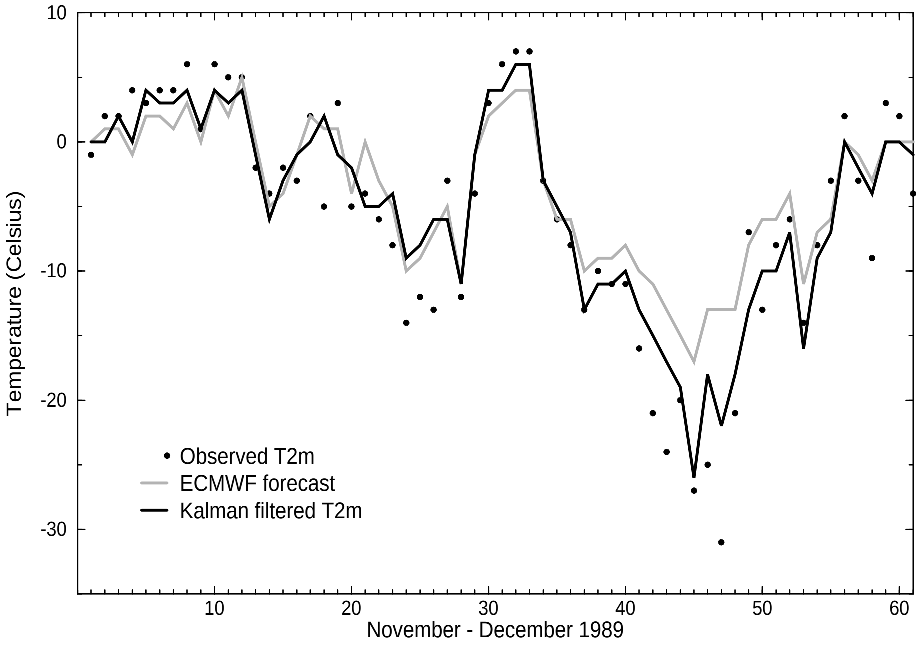


Figure 45 : Direct model output (shaded line) and 2-parameter Kalman filtered 12 UTC t+42h temperature forecast (full line) valid at 06 UTC for Lulea, November-December 1988. The 2-dimensional filter makes a start from ”scratch” on 1 November and after some weeks manages to identify a relation between forecast error and forecasted temperature, according to which the model underforecasted warm temperatures and underforecasted cold. When a cold spell sets in mid-December the filter manages to make useful corrections, in the order of 10 degrees.

With the increasing level of output from the NWP systems, in particular the EPS, the use of automatic statistical interpretation schemes, will become more necessary. The use of adaptive techniques minimizes any inconvenience due to changes in the NWP forecast system.

8 The use of the Ensemble Prediction forecasts

The information from the Ensemble Prediction System (EPS) can be used to formulate deterministic as well as probabilistic forecasts, both in a qualitative and quantitative way. It is also possible to increase the usefulness and skill of the EPS products by applying different kind of statistical post–processing.

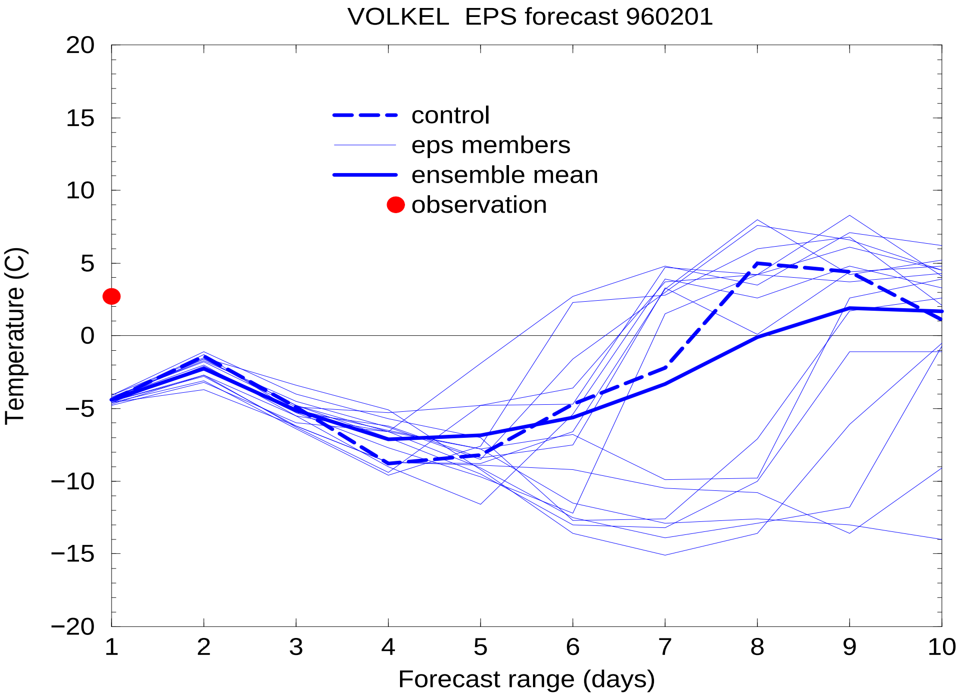


Figure 46 : The same situation as in figure 14 but now with the output from the EPS. The Control is supported by more than half of the EPS members, indicating a return to milder conditions after about five days. A substantial minority support the forecast from the day before that it will stay cold during most of the ten-day period.

There are mainly three strategies in formulating weather forecasts using EPS information:

- a categorical forecast with decreasing degree of details and/or confidence in periods of low predictability;
- a categorical forecast with more details and higher confidence, supplemented with an alternative for the less predictable part of the forecast range;
- a probabilistic forecast, indicating all possible developments with an indication of their likely hood, in percentage.

8.1 Deterministic use of EPS

Averages of the whole ensemble information, or different forms of classifications, like clustering or tubing, can be used for deterministic forecasting.

8.1.1 The ensemble mean

The *ensemble mean* is obtained by averaging all ensemble forecasts. This leads to a smoothing of the forecast fields. This is also, but to a lesser extent, the case with the central cluster in the tubing. The degree of smoothing depends on the spread of the ensemble: when the spread is small it will be possible to follow individual synoptic systems, sometimes even fronts, into the medium range. When the spread is large, only the largest atmospheric scales, the planetary (Rossby) waves remain. Ensemble mean maps display a higher degree of skill and day–to–day consistency than the deterministic forecast.

8.1.2 The ensemble spread

The ensemble spread relates to the differences between ensemble forecasts. Small spread indicates *low* forecast uncertainty, large spread *high* forecast uncertainty. It indicates how far into the forecast the ensemble mean forecast can carry informative value. It helps the forecaster to express appropriate uncertainties. Although the predictability decreases with forecast time, there are many occasions when this is not the case. The developing of a cyclone early in the forecast might be very uncertain, but not the formation of a blocking high some days later.

8.1.3 The combined use of mean fields and spread

The ensemble standard deviation field, 500 hPa geopotential height, superposed on the ensemble mean field of the same parameter, allows to identify the meteorological features which are most affected by forecast uncertainty.

8.2 Mixed deterministic-probabilistic use of EPS

There is no sharp dividing line between products for deterministic or probabilistic forecasts information. Information extracted from probability distributions can serve as a deterministic statement like the mean (the best estimate), the most likely and the median. A confidence interval around the ensemble mean value adds information on the uncertainty of the forecast.

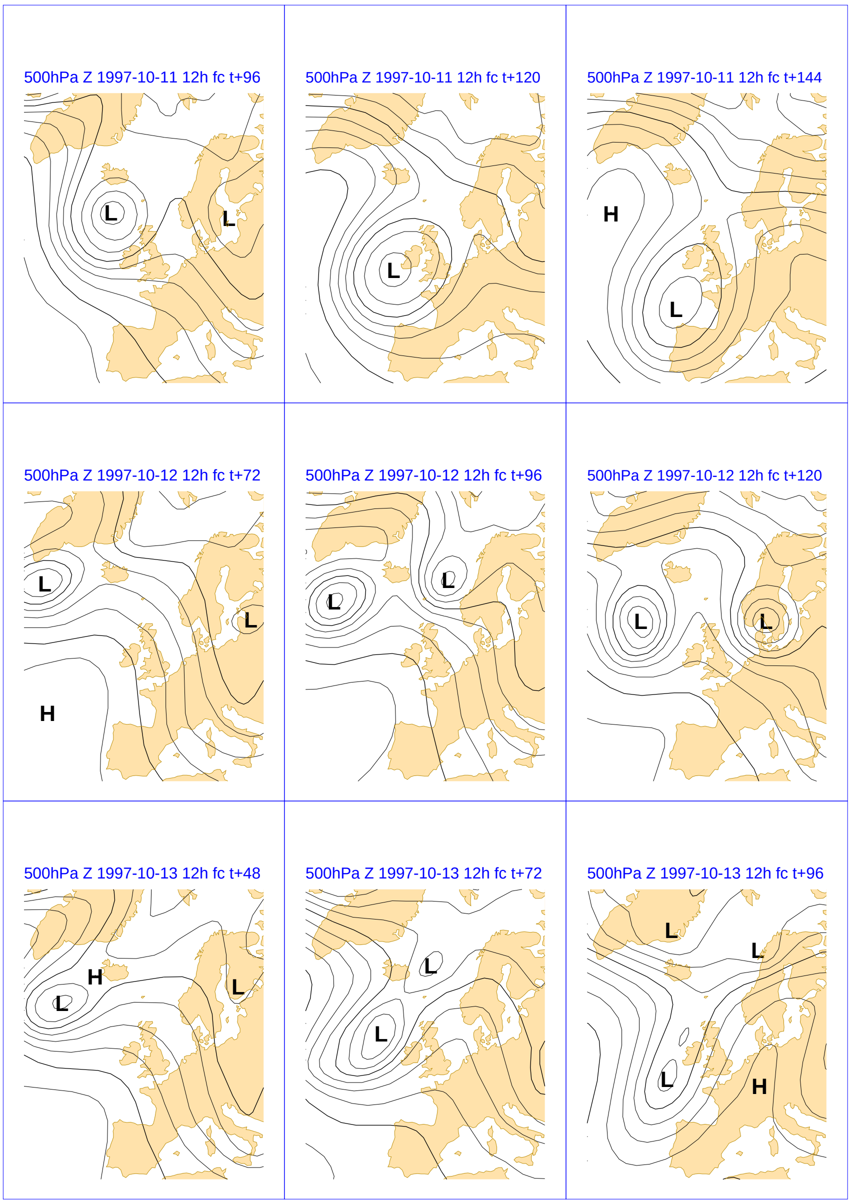
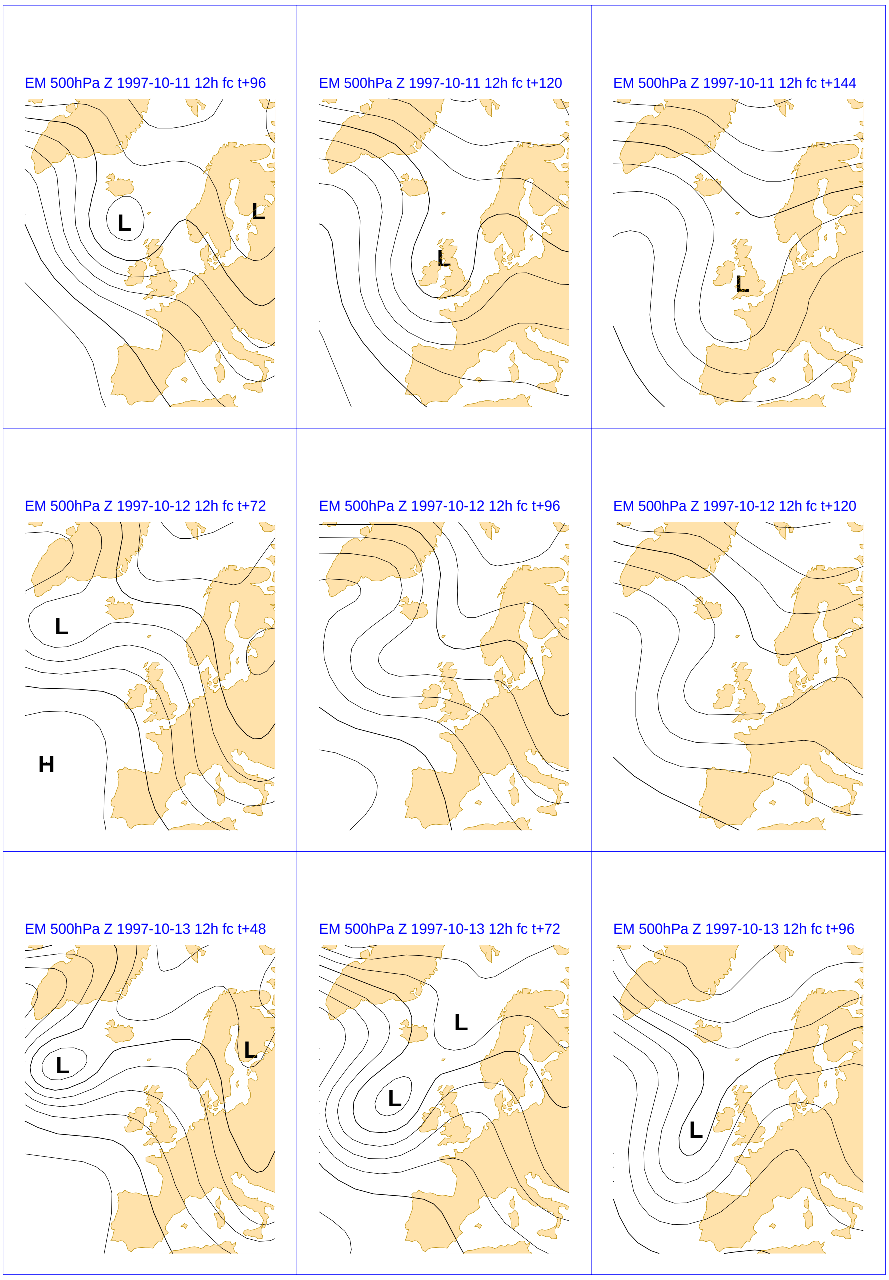


Figure 47 : A case of severe forecast inconsistency (or “jumpiness”): The upper row shows the D+4 to D+6 500 hPa flow pattern in the 11 Oct 1997 forecast; the middle row the D+3 to D+5 from 12 Oct and the bottom row the D+2 to D+4 from 13 Oct. The forecasts in the left, middle and right columns verify on 15, 16 and 17 Oct respectively. The forecasts from 11 and 13 Oct both indicate that a vortex will move down from Iceland to the Bay of Biscay, with a ridge over central Europe. The 12 Oct forecast has a different solution. The vortex is now forecast to stay south of Iceland while a new trough is developing north of Iceland and moves to southern Scandinavia.



EM 500hPa Z 1997-10-11 12h fc t+96

EM 500hPa Z 1997-10-11 12h fc t+120

EM 500hPa Z 1997-10-11 12h fc t+144

EM 500hPa Z 1997-10-12 12h fc t+72

EM 500hPa Z 1997-10-12 12h fc t+96

EM 500hPa Z 1997-10-12 12h fc t+120

EM 500hPa Z 1997-10-13 12h fc t+48

EM 500hPa Z 1997-10-13 12h fc t+72

EM 500hPa Z 1997-10-13 12h fc t+96

Figure 48 : The 500 hPa ensemble mean of the EPS forecasts 11-13 October 1997. The ensemble mean chart from 11 October is similar to the deterministic forecast. The ensemble mean chart the following day, the 12 October, did not support the operational forecast, since most members were consistent with the previous days forecast. The following day, the 13 October, the ensemble mean still supported the one-trough development. Interestingly, a closer examination of the 11 October members (not shown), showed that more than a handful of them suggested a two-trough solution, like the one that the deterministic forecast displayed the next day, on 12 October.

8.2.1 Interpreting clusters and tubes

Clusters, to some degree also tubes, are useful for qualitative synoptic risk assessments. The number of members in each cluster give its “weight” or probability, since all members are a priori regarded as equally likely. These probabilities depend to some extent on the area. The forecasters are recommended to put more emphasis on the sub-area clusters than the “European” cluster.

For one and the same weather situation, the number of clusters in the European area can differ from the numbers of clusters in each of the sub-areas. For example, in a “European” perspective a blocking event might figure prominently in the clusters. However, in a sub-area, not affected by the blocking, the clustering might focus on differences in the forecast of a certain cut-off development.

8.2.2 “Similar” clusters not similar

Similar looking clusters might differ in the overall level of geopotential or temperature: a *cold* westerly flow in one cluster will not give the same weather as a *warm* westerly flow in another. This is an effect of clustering according to the RMS differences. When estimating risks from the clusters, forecasters should take into account that, for a specific location, different clusters might have the same consequences in terms of weather, temperature and wind.

8.2.3 The effect of averaging

During the late medium range, well defined synoptic system in the individual ensemble member forecasts might, when averaged in a mean cluster, indicate a weakening of the system. This paradoxical results occurs when the position of the system displays a geographical spread.

8.2.4 Interpretations of spread measures

There is no high correlation between standard deviation and synoptic uncertainty. The standard deviation is sensitive to the geographical positions of gradients. Large standard deviations are therefore often associated with strong gradients, whereas areas with weak gradients are likely to exhibit small standard deviations. Two similarly looking forecast maps might display large differences if they contain strong gradients, whereas two completely different maps with small gradients will display small differences.

A similar problem becomes important when plotting a certain isoline for each of the members on a map, *“spagetti diagrams”*. It is a nice is way to summarize the information in a striking way. But spagetti diagrams are sensitive to the gradients of the field. They easily show large “spread” in situations when the forecast is not uncertain, but the gradient weak, small spread when there are important uncertainties in a strong zonal flow.

8.2.5 No strong correlation between spread and skill

There can be no 100% correlation between the spread and the forecast error. Although low spread should imply that the control or ensemble mean forecasts will be good, *the opposite is not necessarily true*. When the spread is large and the confidence in the forecast therefore should be low, there is still a chance that the control or the ensemble mean might verify. Large spread should therefore not be taken as a reason not to issue a forecast. The best strategy is to issue a forecast based on the ensemble mean or the dominant cluster(s), be careful in the formulations and try to indicate possible alternatives as displayed by the clusters.

8.2.6 There is mostly something to be certain about

Even in cases with large spread and great uncertainties, there is often something to be certain about. The EPS might not be able to give confident information if a blocking will be followed by a normal zonal flow or an outbreak of cold air. Yet, anybody who is sensitive to warm and dry conditions can be given a very confident forecast that this will not occur.

It is also worth remembering that even if the forecast can be uncertain in absolute terms, it can be quite accurate in relative. The actual temperature a week or so ahead might be difficult to specify; but for many applications a confident forecast of the *trend* is quite useful.

8.2.7 Different parameters have different spreads

Even in situation with synoptic high predictability in the large scale, there might be uncertainties in a smaller scale. The uncertainty by 100 kilometres of the passage of a small baroclinic wave with its associated warm and cold airmasses is not normally taken into account by the clustering. However, this small scale uncertainty has the consequence that regions close to where the centre passes have a 50-50% chance of being affected by the cold or warm air.

The level of predictability might also vary between the weather parameters. A low confidence in the temperature forecast does not exclude that the confidence in the precipitation forecast might be high.

8.3 Probabilistic use of EPS

The EPS provides guidance for quantitative probabilistic forecasts of weather parameters.

8.3.1 Probability maps

If all ensemble members are assumed equally likely, the probability of a weather event is simply the proportion of EPS members forecasting this event. From this set of distributions the probability of virtually any parameter, which are forecast by the model, can be computed. Maps of such probability distributions normally show great consistency from one forecast to the next.

When consulting the probability maps it is important to be aware of the time interval; is it a instantaneous probability like the risk of winds more than 15 m/s at 00 or 12 UTC, or does the probability refer to more than 5 mm over 24 hours? A 10% probability of 15 m/s winds or more has quite different significance if the time interval is 10 minutes, 12 hours or one day.

As mentioned earlier, the “tubing” does not provide any explicit probabailities. Synoptic experience suggest, however, that every “tube” has a 10% chance of verifying, which leave the central cluster with a typical probability of 60-90% to verify, depending on the number of “tubes”. See ch. 8.3.4 for more detail about probability maps for longer time intervals

8.3.2 “Alarm bell” maps

To facilitate *warnings of extreme weather* a new experimental product, the “alarm bell” map, has been developed. Based on the history of the last three years of D+5 and D+10 ensemble forecasts (all at T_L255 resolution) a climatology of the model frequency distribution of temperature, precipitation amount and wind speed at each grid point has been established as a function of the month of the year.

Each EPS run is compared with this climatology and a shift in the predicted frequency distribution away from the climatological norm is used to indicate the forecast of extreme events. The assumptions is that what is an “extreme” event in the models climate also should be an extreme event in the real atmosphere.

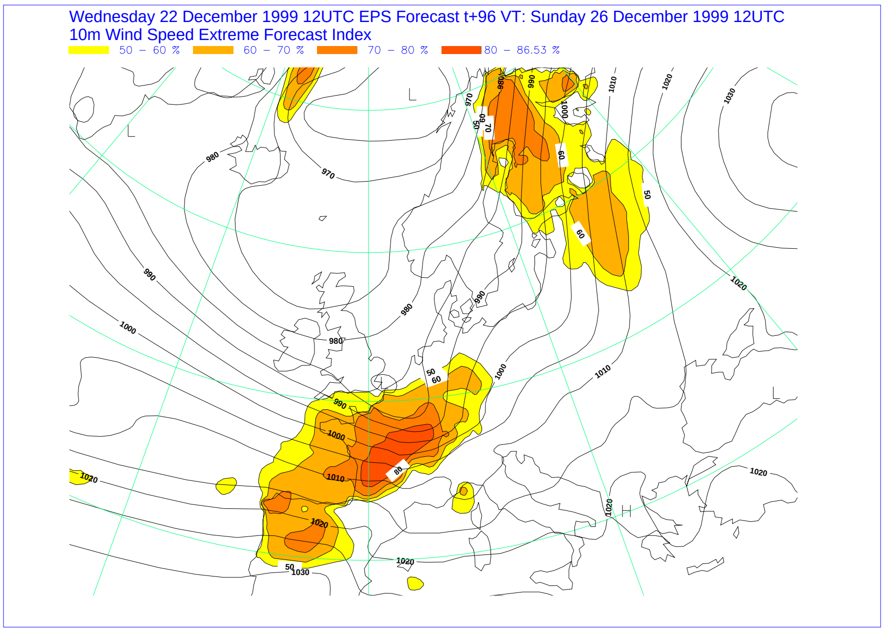


Figure 49 : An alarm bell map from the French storms 26 December 1999. Unusually strong winds are forecast for northern and central France, which indeed turned out to be true. Also northeastern Europe was pointed out but here the winds were only expected to reach 10 m/s, which is much more than the normal of 2-3 m/s.

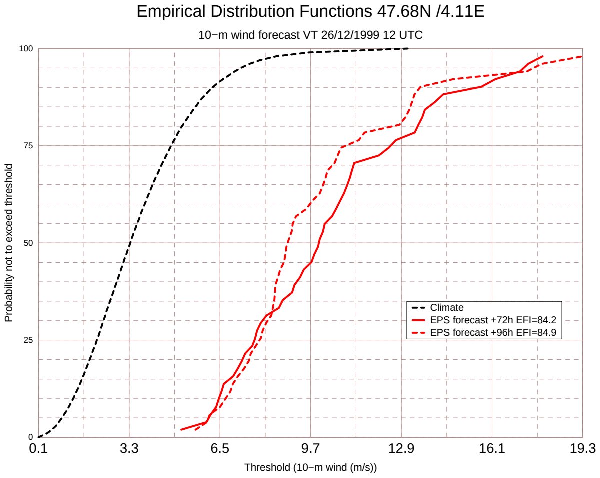


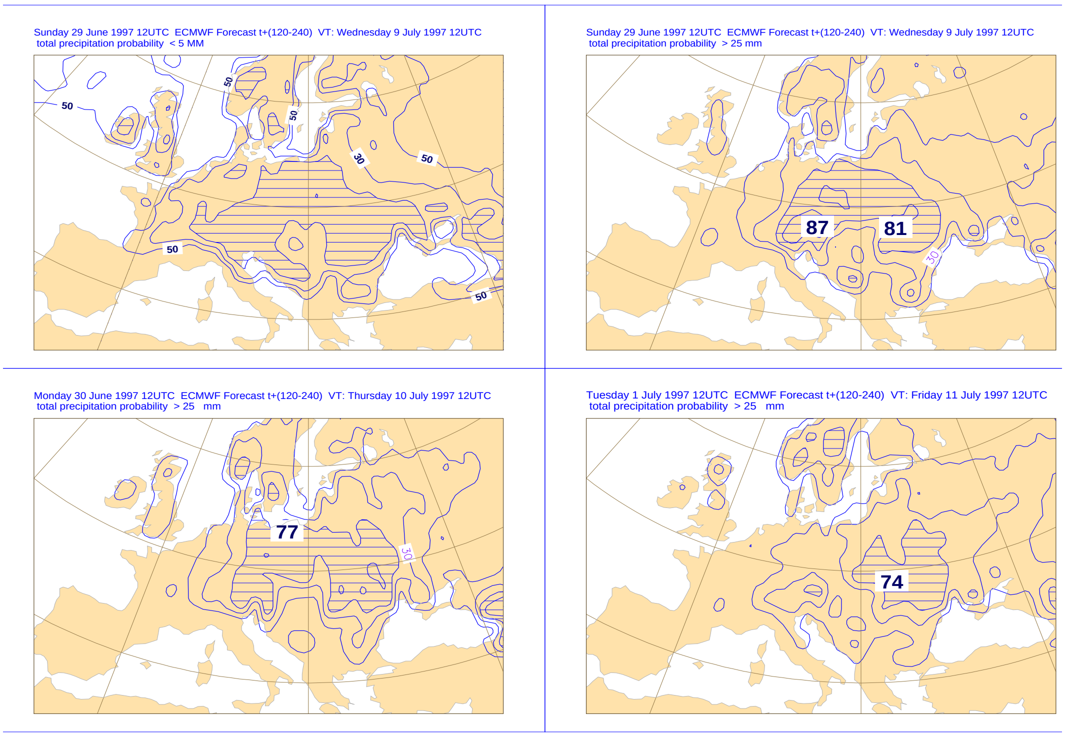
Figure 50 : An EDF diagram for Paris 26 December 1999 12 UTC. Wind speeds along the x-axis, accumulated probabilities along the y-axis. The dashed curved to the left indicates the accumulated probabilities of the climate, which for example tells that there is 50% probability to have more than 3.3 m/s. The curves to the right are two consecutive EPS forecasts which indicate that the risk of more than 9.7 m/s increased from 40% in the D+4 from 22 December to 55% in the D+3 from 23 December. This should be compared with a climatological risk of a 1-2%. Note that the probabilities refer to 12 UTC sharp; the forecast probabilities over 12 or 24 hours would necessarily be higher.

8.3.3 Combined events

The EPS is also suitable to calculate probabilities of combined events like less than 6/8 cloud cover *and* temperatures above 20 deg. Combinations of temperature and wind can define a “wind chill” index, temperature and humidity a “comfort index”, both examples of products that can be derived from the EPS output and given probability formulations.

8.3.4 Probabilities over longer time intervals

The longer the time period over which the probabilities are calculated, the more skilful they are. The confidence in the individual rain forecasts for days 5, 6 and 7 separately, is always lower than for the whole three day period.



Sunday 29 June 1997 12UTC ECMWF Forecast t+(120-240) VT: Wednesday 9 July 1997 12UTC
total precipitation probability < 5 MM

Sunday 29 June 1997 12UTC ECMWF Forecast t+(120-240) VT: Wednesday 9 July 1997 12UTC
total precipitation probability > 25 mm

Monday 30 June 1997 12UTC ECMWF Forecast t+(120-240) VT: Thursday 10 July 1997 12UTC
total precipitation probability > 25 mm

Tuesday 1 July 1997 12UTC ECMWF Forecast t+(120-240) VT: Friday 11 July 1997 12UTC
total precipitation probability > 25 mm

Figure 51 : 2-5 day probability maps: Probability of less than 5 mm during 5-9 July 1997 from the 6-10 day forecast 29 June 1997 (upper left), probability of more than 50 mm during 5-9 July from the same forecast (upper right), the same for 6-10 July from the 30 June forecast (lower left) and for 7-11 July from the 1 July forecast (lower right). Central and eastern Europe became affected by very heavy rain fall during the period 5-10 July 1997 (see fig.2 p.23)

Rain which over a short period of time may not pose any threat of severe weather may give rise to flooding etc. if persisting over a sufficiently long period.

8.3.5 The EPS meteogram

The ensemble information at one grid–point location may be displayed through a probabilistic meteogram, which indicates the time evolution of a given parameter for all ensemble members. The spread is indicated by the range of forecast values. Note the discrepancy between the T_L511 temperature (full line) and the ensemble. The reason for the difference is the coarse resolution in the

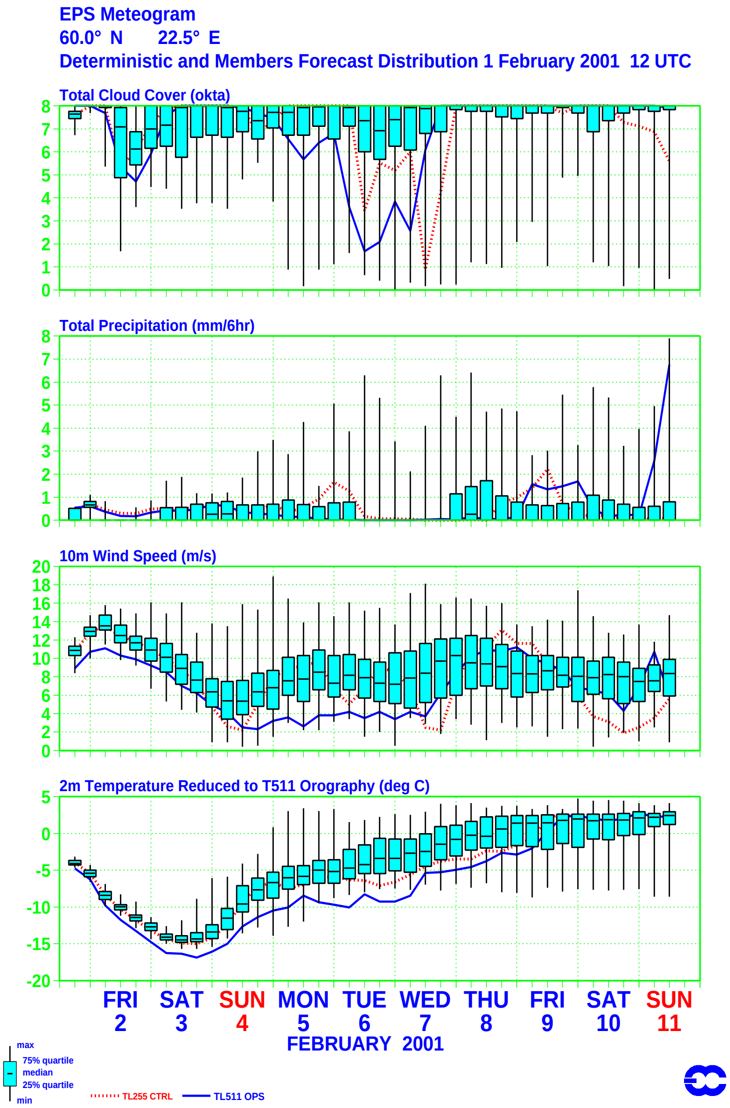
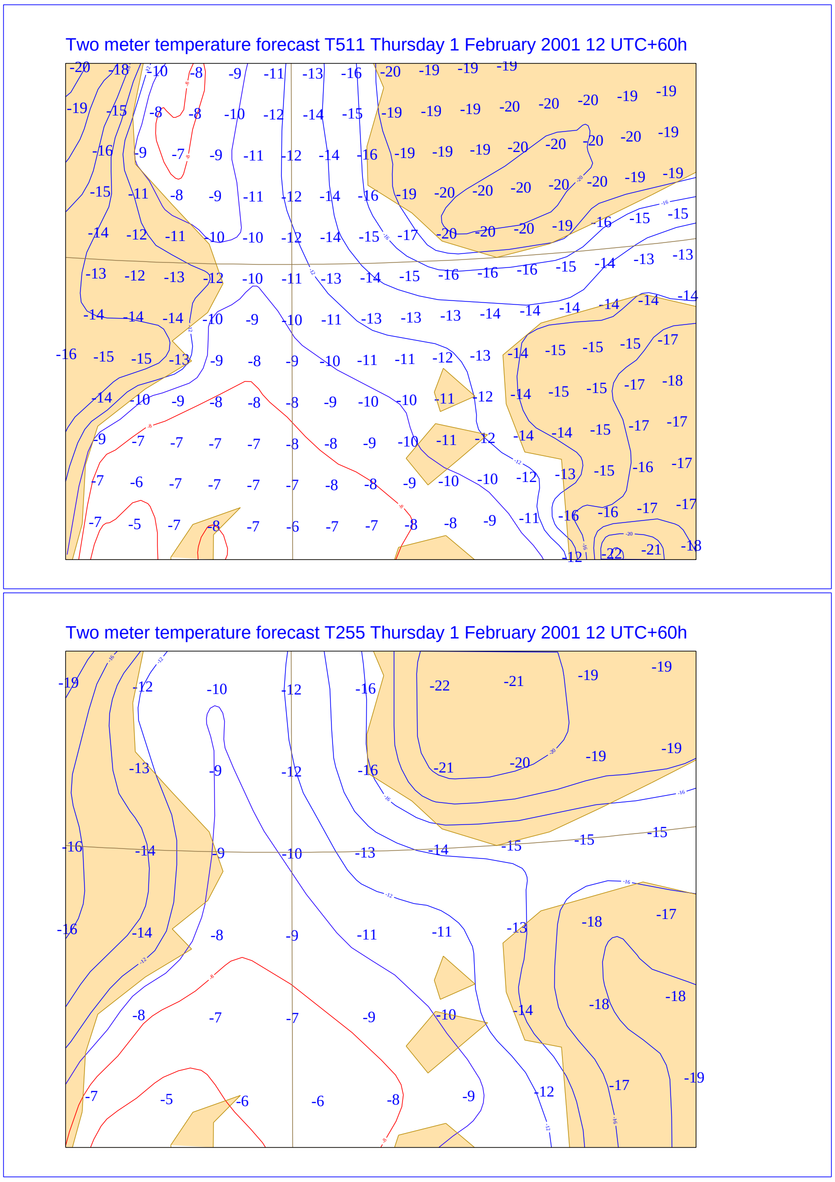


Figure 52 : EPS meteogram for Turku, southwestern Finland, 1 February 2001. The spikes indicate the full range of the ensemble values, the rectangles the interval around the median (indicated by -) of 50% of the ensemble values. The discrepancy between the ensemble values and the T_L511 is due to different resolution and the location along coastline with sharp temperature gradients.

resolution in the T_L255 model, previously mentioned in ch. 5.6. It leaves islands and exposed coastal areas with 2 metre temperatures which in reality are sea surface temperatures, or greatly influenced by sea surface temperatures. If values are interpolated the effect will spread even further into land.



Two meter temperature forecast T511 Thursday 1 February 2001 12 UTC+60h

Two meter temperature forecast T255 Thursday 1 February 2001 12 UTC+60h

Figure 53 : 2 metre temperature in a +60 h forecast in the T_L511 model (above) and in the T_L255 (below) from 1 February 2001 12 UTC from the same period as the epsogram above. Isotherms for every second degree and forecast values for every Gaussian grid point. In the coarse resolution model forecast for Eastern Sweden and southwestern Finland the temperature is 5-10 degrees higher than in the high resolution due to the sharp thermal contrasts between land and sea.

8.4 Statistical post–processing of EPS products

The output of the EPS represents an enormous quantity of forecast meteorological fields and an important task is to find methods to condense the information. Doing so we must keep in mind that the advantage of condensing information has to be paid by the risk of losing some information.

8.4.1 Calibration

The probabilities can be artificially improved statistically. If the verifications show that the forecasts are over-confident (low probabilities verify too frequently, high probabilities verify with lower frequency) a correction can be imposed which will “upgrade” low probabilities and “downgrade” high probabilities. This kind of calibration tends to decrease the range of forecast probabilities.

8.4.2 Statistical interpretation

Since the EPS products are the same as from a NWP model it can easily be applied to any statistical post–processing like MOS, PPM or Kalman filtering. The MOS or PPM equations are applied separately for each EPS member, after which plumes, histograms or probability charts can be made.

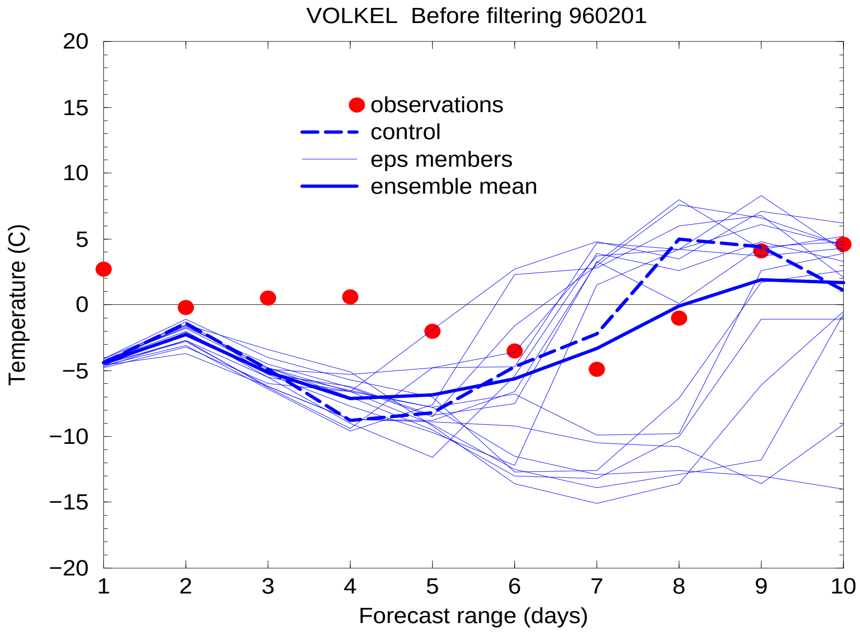


Figure 54 : The same as figure 25 but now with the verifying observations, showing a systematic underestimation of the temperatures and not a very good description of the synoptic development.

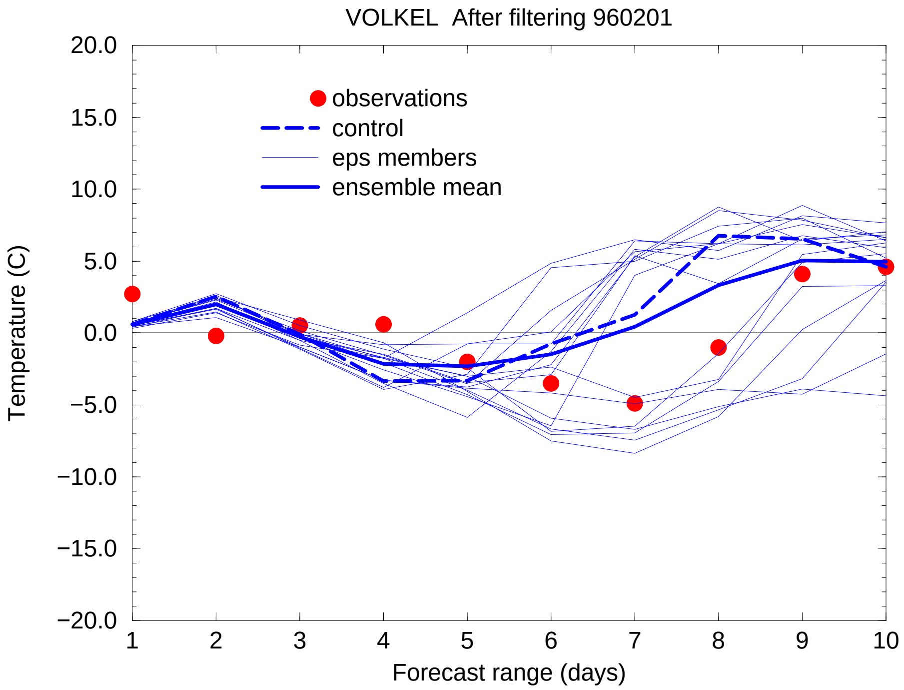


Figure 55 : By Kalman filtering the +24 h unperturbed forecasts assuming that the error depends linearly on the forecast value, the system “learns” how to correct the forecasts from both the unperturbed and perturbed members. Since the correction only aims at correcting for systematic errors in the physical parametrization or geographical location, the same correction formula applies irrespective if the forecasts are perturbed or not. After correction the EPS forecasts appear very realistic and it becomes clear that much of the misfit between forecast and verification was not due to the dynamics. Note that the kalman filtering has reduced the spread by correcting more for cold temperatures than for warm.

8.4.3 Climatological weather type classification

It is possible to cluster on climatologically predefined flow patterns, where each pattern also provides statistical information about the probability of weather events, providing this pattern verifies. This is in particular valuable in mountainous regions where there is low skill in the weather parameters.

The statistical values for each station in the area under different flow regime conditions can be evaluated and thus provide a supplement for the probabilities derived from the weather parameters. At a certain location, the probability for an event then becomes the combination of the probability that the event will take place provided the flow type will occur, multiplied with the probability, derived from the EPS, that it will occur.

8.5 What value can the forecaster add to the EPS?

As with deterministic forecasts, the usefulness of the EPS cannot be judged only on the statistical value of the predicted parameters, but how it can guide the forecaster into adding extra value. Since the EPS is rather new, we still lack sufficient experience to give detailed advice, but the following guidelines might be useful.

8.5.1 Systematic errors

The EPS has a small tendency to underestimate the atmospheric variability, i.e. to identify all possible weather regimes that can occur in a given situation. This can sometimes be seen when the deterministic T_L511 provides a forecast that is not covered by the ensemble. It is difficult to estimate how this should be interpreted, but a tentative suggestion is that the T_L511 should be regarded as likely as a handful members. *It should not, unless there are strong reasons, be regarded as the most likely solution.* Remember that the T_L511, for statistical reasons, should be outside the ensemble around 4% of the time.

8.5.2 Non-systematic errors

There should, in principle, not be any inconsistent “jumps” in the EPS forecasts from one day to the other. But sometimes, when there is dynamic activity over many regions on the Northern Hemisphere, it might not be enough with 50 analysis alternatives. It can in those cases happen that the perturbations one day are more concentrated in one sensitive region than another, only to change the distribution slightly the next day. In such cases the forecaster is wise to consider *both* days’ ensemble output.

8.5.3 Spread and predictability

As mentioned earlier, for one and the same ensemble, some parameters can display unusual high spread, whereas others can have quite small spread. It is therefore misleading to judge the “predictability” only after one parameter, for example the 500 hPa geopotential.

During a blocking event there can be a large spread in the wind, temperature and pressure forecasts, but small in the precipitation. In a zonal regime, with small spread in the upper air fields, the exact location and track of a baroclinic wave might yield large spread in areas where the exact trajectory is uncertain.

8.5.4 Interpreting probabilities

It is always assumed that just because the forecasters have no possibility to modify the probability values coming out from the EPS (or some post-processing with correction for systematic error or adjustment to local conditions) that they cannot add extra value. There is a lot to add to the raw output:

-Which probability is relevant in a particular situation; more than 1 mm/24h or more than 5 mm/24? Or more than 20 mm over 72h or the probability of winds above 5 m/s and temperatures below +5 C?

-Why is the forecast uncertain? A 25% probability of more than 5 mm/24h can be due to different uncertainties. It can mean that only 25% of the area will have more than 5 mm, or it can be the uncertainty if an approaching frontal system will reach the area?

A 25% risk forecast for temperatures below zero can in the same way mean that the clouds in a cold air mass are difficult to forecast and consequently the temperature, or that arrival of cold air in connection with a possible change to meridional flow is uncertain. Probabilities of gale force winds can both be related to variations within a broad zone of strong winds, or the uncertainty associated with one specific baroclinic development.

8.5.5 Forecasts of extreme events

Extreme or anomalous events can be of mainly two types: large or medium scale persistent anomalies like cold outbreaks or heat waves lasting for more than a week, or small scale events with heavy rain or strong winds. The EPS is well equipped to forecast the large scale and with increased higher resolutions, also the smaller scales, responsible for many extreme weather events, have improved.

The cascading process for providing guidance on severe events, long as well as small scale, proposed by WMO/CBS, aligns well with traditional forecast practises:

> 72 hours and earlier: Preliminary indicative guidance, based on probabilistic EPS material;
>
> 24 to 72 hours in advance: more specific warning guidance based on a mixture of probabilistic EPS and deterministic material
>
> less than 24 hours: warnings issued by the responsible centre, based on detection and tracking of the severe weather system;

The forecaster is advised to consider deviating from the EPS in the rare event of the T_L511 operational model has during the last 2-3 days indicated a risk of a severe event, which has not been included in the EPS. It should also be considered when the EPS has forecasted a synoptic flow regime where the forecaster, from his experience, knows that severe weather, not explicitly presented by the EPS, might occur. This is in particular true if meteorological scales are involved which it is not possible for a T_L255 resolution to describe.

8.6 Some general recommendations for realtime use of EPS

There are at least two principle ways of working with the ensemble forecasts.

The most common is to start by inspecting the last days' T_L511 forecasts to establish for how long into the forecast there is a reasonable consistency, and what alternative developments are indicated after that. The last days' EPS clusters are examined to determine if the inconsistencies in the operational model are reflected in the clusters, but also to determine if the EPS itself appears to be reasonably consistent. *This will establish which is the most likely synoptic development, and the main alternatives.* When this is done the probability information is consulted in order to establish if the weather parameter information for different locations, in particular to establish if the spread in the weather parameter forecasts correspond with the synoptic spread in the EPS clusters.

A more direct approach is to do the opposite: to start with the weather parameter information, both in the Epsograms and the probability maps, and then consult the clusters and deterministic forecasts of the general atmospheric flow, to find the synoptic background to the forecast variations in the weather and their probabilities. The advantage with this method is that when there are weak relations between the spread in the synoptic pattern and the weather parameters the forecaster does not necessarily have to spend time finding out which flow scenarios are more or less likely.

9 Epilogue

The ECMWF was set up in 1975 with the aim of providing 10 day forecasts of economic value for the European area. The first target was to provide 5 day forecasts which had the same skill as 2 day forecasts before the “computer age”. This has been achieved and the deterministic forecast now have a skill up to 8 days. The skill varies considerably with sometimes useful forecasts up to 10 days, sometimes hardly beyond 4days. The EPS provides a measure of when a deterministic forecast can be relied on and when not; in those cases the EPS provides a skilful estimation of the likely alternatives.

Used in this way the forecasts, either as an ensemble mean or in a probabilistic sense, already has useful skill up to day 10. The continued work at the ECMWF is to develop this skill further. The resolution of the deterministic model will increase to T_L511, the ensemble system will be run on T_L255 with more members, perhaps twice a day. The use of satellite data will increase in quantity and quality. The 4DVAR will increase its window from 12 hours, later perhaps 24 hours. The consequences will not only be a continued increase in skill of large scale weather systems, but also of small scale, in particular in the EPS.

This provides the meteorological services with an even more increased potential to serve a wide range of needs in the society, since in particular the EPS is suited for tailor suited forecast production. The challenge for today’s meteorologists is not only to make use of the current and future skill of the ECMWF forecasts, but also to develop new products and reach new sectors of society and satisfy new demands.

This will unavoidably involve an increased proportion of automatic or computer to computer generated products. The best experts to do this work are meteorologists with good forecast experience, and with skill in computer based systems. The design, maintenance and upgrade of computer based post-processing software is already and will increasingly become an important task for meteorologically skilled staff.

Forecasts generated in this way, perhaps with computer-to-computer access, will free the forecasters from routine work and enable them to concentrate on situations where their personal intervention and interpretation is needed. To fulfil their task as presenters of information, the forecasters must not only be familiar with the way the atmosphere works, but also how the numerical prediction scheme functions works in theory and practise. Hopefully this User Guide has provided a useful basis for this with respect to the ECMWF forecast system.

10 References and further literature

10.1 ECMWF documentation and publications

10.1.1 Newsletter

A quarterly ECMWF Newsletter is distributed to national weather services in the Member States and users of the GTS products worldwide. It deals with topics in meteorology and the operational activities at the Centre and provides short descriptions of operational changes to the analysis and forecasting system. The newsletter also deals with computing topics.

A collection of important articles related to the development of the ECMWF analysis and forecast system, together with relevant references, can be found in *ECMWF Data Services, 1999: ECMWF/WCRP Level III-A Global Atmospheric Data Archive, The description of the evolution of the ECMWF forecasting system and corresponding archive.*

10.1.2 Bulletins and memoranda

Comprehensive documentation of the analysis and forecasting system, the archiving and dissemination is given in the Meteorological Bulletins. The Computer Bulletins provide the guidance to the Centre's computing facilities. Scientific and technical aspects of the Centre's work are discussed in informal ECMWF Technical Memoranda. A limited distribution within the ECMWF Member States applies to these three types of documentation. Individual copies are available from the Centre's library on request.

10.1.3 Proceedings and reports

Proceedings from the Centre's annual seminar and workshops are distributed widely to the national weather services and scientific institutions of the meteorological community.

ECMWF publishes reviewed papers of results in its own series of Technical Reports, available in the libraries of most national weather services and scientific institutions.

10.1.4 Documentation

A documentation of the analysis and forecast model can be found in the ECMWF Research Manuals:

Data assimilation - scientific documentation (Meteorological Bulletin 1.5/1)
Forecast model - adiabatic part (Meteorological Bulletin 1.6/3)
Forecast model - physical parametrization (Meteorological Bulletin 1.6/2)

10.2 User Guide references

10.2.1 Analysis system

Introductory note: The ECMWF implementation of four dimensional variational assimilation is covered by three papers (Rabier et al, 2000; Mahfouf et al, 2000 and Klinker et al, 2000)

Andersson, E. and H. Järvinen, 1999: Variational quality control, Q. J. R. Meteorol. Soc.,1999, vol 125, pp. 697-722

Bouttier, F. and P. Courtier, 1998: Data assimilation concepts and methods. ECMWF training course lecture notes, Mar 98, 64pp.

Courtier, P., J-N. Thépaut and A. Hollingsworth, 1994: A strategy for implementation of 4DVAR using an incremantal approach, QJRMS, vol, 120, pp.1367-88.

Bouttier F. and F. Rabier, 1998: The operational implemenation of 4DVAR, ECMWF Newsletter Number 78, reprinted in ECMWF Data Services, pp.72-76.

Derber,J. and F.Bouttier, 1999; A reformulation of the background error covariance in the ECMWF global data assimilation system.Tellus 51A,195--221

Fisher, M. 1998: Development of a simplified Kalman filter. ECMWF Tech. Memo. 260.

Gérard, E, and R. Saunders, 1999: 4D-Var assimilation of SSM/I total column water vapour in the ECMWF model. Q. J. Roy. Meteor. Soc.vol. 125, pp.3077-3101.

Järvinen, H., E. Andersson and F. Bouttier, 1999: Variational assimilation of time sequences of surface observations with serially correlated errors, Tellus, 51A, 469-488)

Järvinen, H. and P. Undén, 1997: Observation screening and background quality control in the ECMWF 3D-Var data assimilation system. ECMWF Research department technical memorandum no.236.

Kelly, G., 1997: Influence of observations on the operational ECMWF system, ECMWF Newsletter Number 76, reprinted in ECMWF Data Services pp.65-71.

Klinker, E., F.Rabier, G.Kelly and J.F. Mahfouf, 2000:The ECMWF operational implementation of four dimensional variational assimilation. Part III: Experimental results and diagnostics with operational configuration.Q. J. R. Meteorol. Soc., vol. 126, pp.1191-1215.

Mahfouf, J. F. and F.Rabier, 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part II: Experimental results with improved physics. Q. J. R. Meteorol. Soc., vol. 126, pp.1171-1190.

Persson, A, 2000: Synoptic-dynamic diagnosis of medium range weather forecast systems, ECMWF Seminar on diagnosis of models and data assimiliation systems, 6-10 September 1999.pp.123-137.

Rabier et al., 1997: Recent experimentation on 4D-Var and first results from simplified Kalman filter. ECMWF Research department technical memorandum no.240.

Rabier,F., H. Järvinen, E. Klinker, J-F. Mahfouf and A. Simmons, 2000: The ECMWF operational implementation of four dimensional variational assimilation. Part I: Experimental results with simplified physics.Q. J. R. Meteorol. Soc. vol.119, pp.845-880,

Saunders R., E.Andersson, H.Järvinen, E. Gérard, M. Rohn and L.Isaksen, 1998: Recent improvements to the ECMWF 4D-Var data assimilation, ECMWF Newsletter no 81, pp.2-7, reprinted in ECMWF Data Services, pp. 90-97.

Thépaut, J.-N., and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. Q.J.Roy.Meteor.Soc., 117, 1225-1254.

10.2.2 Forecast model

Beljaars, A.C.M., and A.A.M. Holtslag, 1991: Flux Parametrization over land surfaces for atmospheric models. J.Appl.Meteor., 30, 327-341.

Beljaars, A. and P. Viterbo, 1993: A new surface/boundary layer formulation at ECMWF, ECMWF Newsletter September 1993, reprinted in ECMWF Data Services pp.31-41.

Beljaars, A. and P. Viterbo, 1998: The role of the boundary layer in a numerical weather prediction model, in: *Clear and cloudy boundary layers,. A.M.M. Holtslag and P.G.Duynkerke, Eds*, pp. 287-403. Royal Netherlands Academy of Arts and Sciences.

Beljaars, A., J-F. Mahfouf, J. Teixeira and P.Viterbo, 1996: Improvements to the 2 m temperature forecast, ECMWF Newsletter Number 73, reprinted in ECMWF Data Services pp.56-60.

Gregory, D., J-J. Morcrette, C.Jakob and A.Beljaars, 1998: Introduction of revised parametrization of physical processes into the IFS, ECMWF Newsletter Number 79, reprinted in ECMWF Data Services, pp.77-83.

Gregory, D., J.-J. Morcrette, C. Jakob, A. C. M. Beljaars, and T. Stockdale, 2000: Revision of convection, radiation and cloud schemes in the ECMWF Integrated Forecasting System. {\qj}, {•f 126}, 1685-1710.

Hortal, M., and A.J. Simmons, 1991: Use of reduced Gaussian grids in spectral models. Mon.Wea.Rev., 119, 1057-1074.

Jakob, C., and S.A. Klein, 1999: The role of vertically varying cloud fraction in the parametrization of microphysical processes in the ECMWF model. Quart. J. Roy. Meteor. Soc., 125A, 941-965.

Jakob, C., and S. A. Klein, 2000: A parametrization of the effects of cloud and precipitation overlap for use in general-circulation models. {\qj}, {•f 126}, 2525-2544.

Jakob, C., E. Andersson, A. Beljaars, R. Buizza, M. Fisher, E. Gérard, A. Ghelli, P. Janssen, G. Kelly, A. P. McNally, M. Miller, A. Simmons, J. Teixeira, and P. Viterbo: The IFS cycle CY21r4 made operational in October 1999. {\it ECMWF Newsletter}, {•f 87}, 2-9.

Lott, F., and M.J. Miller, 1997: A new sub-grid scale orographic drag parametrization: Its formulation and testing, Quart. Journ. Roy.Met.Soc. Vol.123, pp.101-27.

Miller, M.J., T.N. Palmer, and R. Swinbank, 1989: Orographic gravity-wave drag: its parametrization and influence in general circulation and numerical weather prediction models. Meteor.Atmos.Phys., 40, 84-109.

Miller, M.J., A.C.M. Beljaars, and T.N. Palmer, 1992: The sensitivity of the ECMWF model to the Parametrization of evaporation from the tropical oceans. J.Climate, 5, 418-434.

Miller, M.J., 1993: The modelling of hydrological processes in the atmosphere. In: Modelling Oceanic Climate Interactions, Ed. J. Willebrand and D.L.T. Anderson. NATO ASI Series I, Vol.11. Springer-Verlag, 1-33.

Miller M., M. Hortal and C.Jakob, 1995: A major operational forecast change, ECMWF Newsletter Number 70, reprinted in ECMWF Data Services, pp.42-49.

Morcrette, J.-J., 1991: Radiation and cloud radiative properties in the ECMWF operational weather forecast model. J.Geophys.Res., 96D, 9121-9132.

Morcrette, J.-J., 1993: Revision of the clear-sky and cloud radiative properties in the ECMWF model. ECMWF Newsletter, 61, 3-14.

Ritchie, H., C. Temperton, A. Simmons, M. Hortal, T. Davies, D. Dent, and M. Hamrud, 1995: Implementation of the semi-Lagrangian method in a high-resolution version of the ECMWF forecast model. Mon.Wea.Rev., 123, 489-514.

Tiedtke, M., 1989: A comprehensive mass flux scheme for cumulus parametrization in large-scale models. Mon.Wea.Rev., 117, 1779-1800.

Tiedtke, M., 1993: Representation of clouds in large-scale models Mon. Wea. Rev., 121, 3040-3061.

Tiedtke, M., 1996: An extension of cloud-radiation parameterization in the ECMWF model: The representation of sub-grid scale variations of optical depth. Mon. Wea. Rev., 124, 745-750.

Untch, A., A. Simmons et al., 1999: Increased stratospheric resolution in the ECMWF forecasting system, ECMWF Newsletter Number 82, repirnted in ECMWF Data Services pp. 98-105.

10.2.3 Ensemble prediction

Barkmeijer, J, R. Buizza and T.N. Palmer, 1999. 3D-Var Hessian singular vectors and their potential use in the ECMWF Ensemble Prediction System. Q. J.R. Meteorol. Soc., 125,2333-2351.

Buizza, R, and T.N.Palmer, 1995. The singular-vector structure of the atmospheric generalcirculation. J. Atmos. Sci., 52, 9, 1434-1456.

Buizza, R., T. Petroliagis, T.N.Palmer, J.Barkmeijer, M.Hamrud, A.Hollingsworth, A.Simmons and N.Wedi, 1998. Impact of model resolution and ensemble size on the performance of an ensemble prediction system. Q.J.R. Met. Soc. 124, 1935-1960.

Buizza, R, M.Miller and T.N.Palmer, 1999. Stochastic representation of model uncertantiesin the ECMWF Ensemble Prediction System. Q. J. R. Meteorol Soc., 125, 2887-2908.

Molteni, F, R. Buizza, T.N. Palmer and Petroliagis, T, 1996. The new ECMWF ensemble prediction system: methodology and validation. Q. J. R. Meteorol. Soc., 122, 73-119.

Palmer, T.N., 2000: Predicting uncertainty in forecasts of weather and climate, Rep. Prog. Phys. **63,** pp. 71-116.

Puri, K., J. Barkmeijer and T.N. Palmer, 1999: Ensemble prediction of tropical cyclones using targeted diabatic singular vectors. ECMWF, Technical Memorandum No 298.

10.2.4 Ocean wave modelling

Hasselmann S., K. Hasselmann, E. Bauer, P. Janssen, G.J. Komen, L. Bertotti, P. Lionello, A. Guillaume, V.C. Cardone, J.A. Greenwood, M. Reistad, L. Zambresky and J.A. Ewing, 1988. The WAM model-a third generation ocean wave prediction model, J. Phys. Oceanogr. 18, 1775-1810.

Komen G.J. , L. Cavaleri, M. Donelan, K. Hasselmann, S. Hasselmann and P. Janssen, 1994. Dynamics and modelling of ocean waves. Cambridge University Press, 532p.

Janssen P., B. Hansen, and J.-R. Bidlot, 1997. Verification of the ECMWF Wave Forecasting system against Buoy and Altimeter Data. Weather and Forecasting, 12, 763-784.

Janssen P., 1997: Operational implementation of the high resolution Ocean Wave model, ECMWF Newsletter Number 74, reprinted in ECMWF Data Services pp.61-64.

Janssen P. 1998: Progress with wind-wave interaction, ECMWF Newsletter Number 80, reprinted in ECMWF Data Services, pp.84-89.

10.2.5 Use of products

Atger, F., 1999: The skill of ensemble prediction systems, Month. Wea. Rev., Vol. 127, pp. 1941-53.

Cattani, D., 1994: Application d’un filtre de Kalman pour adapter les températures à 2 mètres fournies par le modèle ECMWF aux stations météorologiques de la Suisse. Arbeitsbericht der Schweizerischen Meteorologischen Anstalts, Zürich.

Glahn, H.R., Murphy, A.H., Wilson, L.J., Jensensius, J.S., 1991: Lectures presented at the WMO Training Workshop on the Interpretation of NWP Products in terms of Local Weather Phenomena and their Verification. WMO Programme on Short and Medium-Range Weather Prediction Research, PSMP Report Series No. 34, WMO/TF no. 421.

Murphy, A., and R.W.Katz (eds), 1985: Probability, statistics and decision making in atmospheric sciences. Westview Press.

Persson, A., 1991: Kalman filtering - a new approach to adaptive statistical interpretation of numerical meteorological forecasts. Lecture presented at the WMO Training Workshop on the Interpretation of NWP Products, Wageningen, The Netherlands, 29 July-9 August 1991. WMO/TD. No. 421.

Persson, A. and B. Strauss, 1995:On the skill and consistency in medium range weather forecasting, ECMWF Newsletter, No 70, Summer 1995, p.12-15..

WMO, 1992: Methods of interpreting numerical weather prediction output for aeronautical meteorology. Technical note No. 19