THE VERIFICATION OF WEATHER PARAMETERS

Andreas Lanzinger
European Centre for Medium-Range Weather Forecasts
Shinfield Park, Reading, UK

1. INTRODUCTION

At ECMWF extensive verification of direct model output of weather parameters is performed. Operationally, the model performance with respect to the parameters precipitation, total cloud cover, 2 metre temperature and specific humidity, and 10 metre wind speed and direction is monitored on a daily, weekly and monthly basis. The aim of this paper is not to discuss in detail model deficiencies and possible reasons, but to give an overview of methods used to monitor the model's weather parameter performance.

Section 2 outlines the purposes of weather parameter verification at ECMWF. Some general considerations for verification of large scale models against conventional surface observations are presented in section 3. Section 4 gives an overview over procedures applied. A few results are presented in section 5.

2. PURPOSES OF VERIFICATION

There is a variety of possible purposes for verifying numerical model products or forecasts in general. For a list of these purposes see e.g. *Murphy* (1985).

At ECMWF, the main motivation is monitoring and diagnostics of operational model behaviour of the parameters mentioned in the introduction. All these parameters are not actual prognostic variables in the current formulation of the model, but are derived or diagnosed from "real" model variables. Still, these fields are products which are disseminated to member states and are used in one or the other way as forecast guidance. They are the atmospheric parameters that describe the actual weather most closely. Therefore, despite the fact that the skill in predicting these weather parameters does not necessarily reflect the skill in forecasting the "free atmosphere" fields, it is important to assess their quality.

Another important aspect is the indirect monitoring of the behaviour of model soil processes and variables. 2 metre specific humidity errors, for instance, are often a good indicator of deficiencies in soil moisture. The 2 metre values can be verified easily against SYNOP observations, whereas soil water content observations are not available on a regular basis.

A third, important application is the evaluation of experiments. Especially when changes in the model physics are tested is it useful and often crucial to assess the effects of these changes upon weather parameters.

3. REPRESENTATIVENESS OF VERIFYING OBSERVATIONS

Fig. 1 schematically depicts the general "scale problem" in verification of large scale

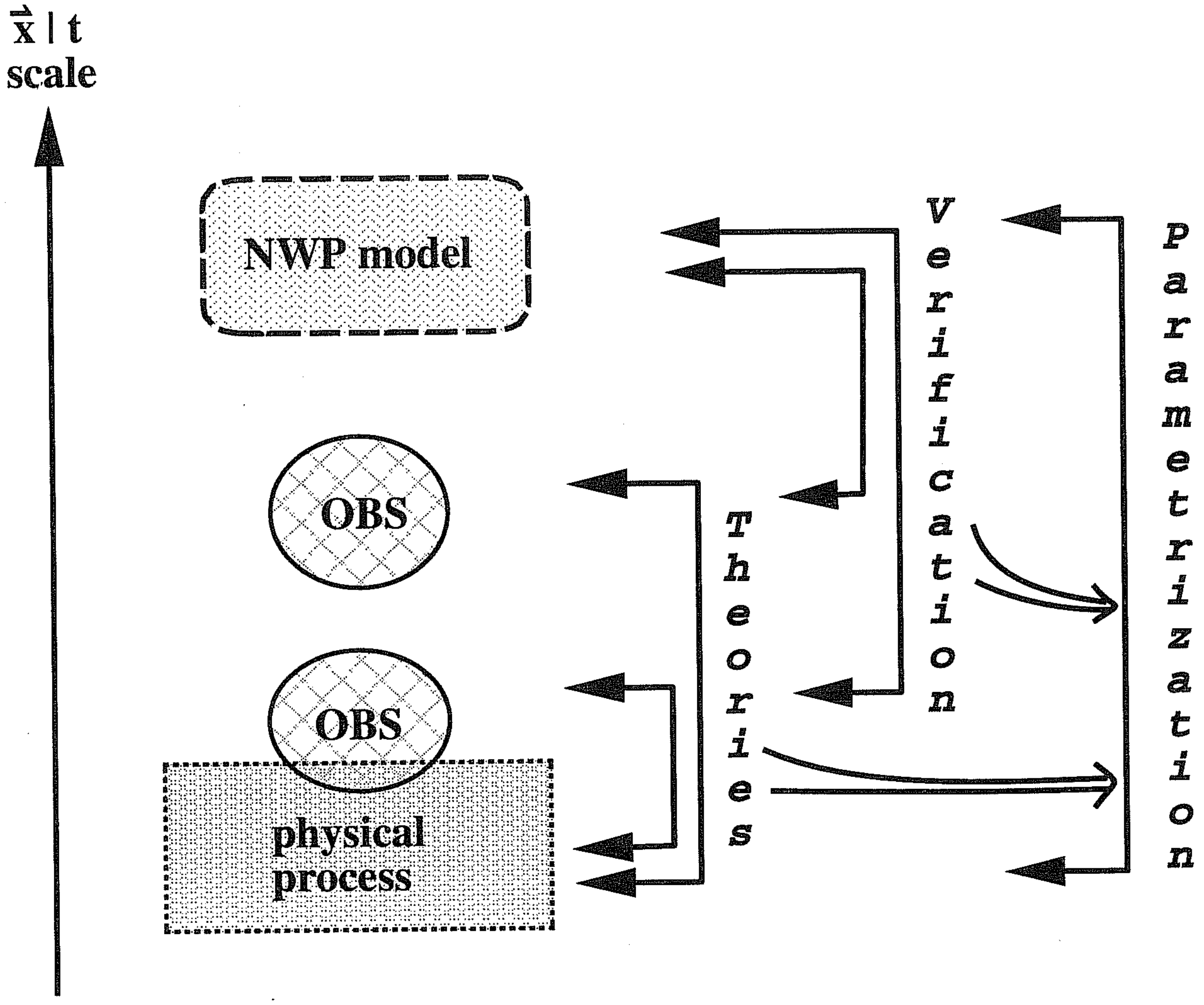


Fig 1 Sketch of the scale problem in verification.

atmospheric models against observations. It also hints the feedback of verification to the parametrization of physical processes in such models.

Observations generally contain atmospheric processes on a smaller spatial and temporal scale than can be represented in a numerical weather prediction model. E.g., mesoscale circulations in mountain valleys which are not resolved in the model orography cannot be captured, although they might have substantial effects on all weather parameters. Similarly, heat and momentum fluxes due to land-sea contrasts can often not be simulated well enough to give realistic coastal temperatures. The model may well have a distinction of land and sea grid points, but the scale of actual processes is mis-represented. Soil and vegetation conditions determine the energy and moisture transfers at the surface. These conditions have to be prescribed at grid points, and it is clear that they can not be representative of all the observation sites in their particular grid box.

This scale defect, or in other words, the models inability to describe processes at the same scale as they are measured, has to be considered when interpreting verification results. This leads directly to the issue of statistical correction of direct model output. Especially for parameters like the ones considered here, it is useful to apply statistically derived corrections (e.g. "Model Output Statistics - MOS", "Kalman filter") to compensate for known scale (and other) deficiencies in the direct model output. These filtered products will usually have better skill and can serve better as forecast guidance for weather parameters.

There are other important aspects in verification which are also related to the above scale problem. The differences in model orography height and observation height have to be taken into account. Firstly, corrections have to be applied for 2 metre temperatures. (In our case a simple correction with the standard lapse rate is added and observations with too large orography differences are excluded). For other parameters, like precipitation, these differences are also critical: comparing model precipitation over a large, smooth mountain to observations in complex terrain - valley, slope, mountain top locations - is generally not very representative.

When deriving error statistics over areas or sets of observation locations the sampling can play a significant role. The overall statistics, e.g. the model bias over a larger area, should not be dominated by errors in a small sub-area where relatively too many verification points are located. Meaningful, balanced statistics can be achieved quite easily by assuring fairly even spatial distribution of verifying observations, without having to apply area weighting factors.

Finally, observational errors are a difficult problem in verification. With no manual intervention these errors are hard to detect and rejection procedures are usually complicated and sometimes unreliable. Observational errors are not treated in this paper.

4. PROCEDURES

In the continuous monitoring of weather parameters the idea is to diagnose errors in a forecast range where synoptic errors are generally small enough not to contaminate too much or even dominate verification results. On the other hand, the results should not be influenced by spin-up effects, which excludes a too short forecast range. Two and three day forecasts seem to serve this purpose best.

All verification results presented here are derived from direct model output, with no corrections applied other than the above mentioned temperature corrections, to account for station - model height differences. This follows the main purpose of assessing the models ability to represent these parameters and to detect physics related problems. Statistical corrections can be applied locally to enhance the skill of weather parameter forecasts, but are not part of our verification.

The parameters listed in the introduction are measured regularly at SYNOP stations, usually with high accuracy. More than 1000 European SYNOP land stations and around 2000 additional stations in other regions around the globe are used for verification reference. The reported values are only checked for gross errors, bearing in mind that usually only a very small percentage of SYNOP observations - depending on parameter and also on region - is erroneous. Over the oceans SYNOP ship and buoy observations are also used for verification.

To compare to observation values the model values have to be interpolated horizontally from grid point to observation locations. There are numerous possibilities to do this. We apply linear interpolation between four surrounding grid points, but for coastal locations the nearest land grid point value is used for temperatures and 10 metre winds. The latter method tries to minimize the influence from sea grid points which often have boundary layer structures completely different from land points. Of course, in situations when there actually is large sea influence at coastal stations, taking the nearest land grid point might result in larger errors.

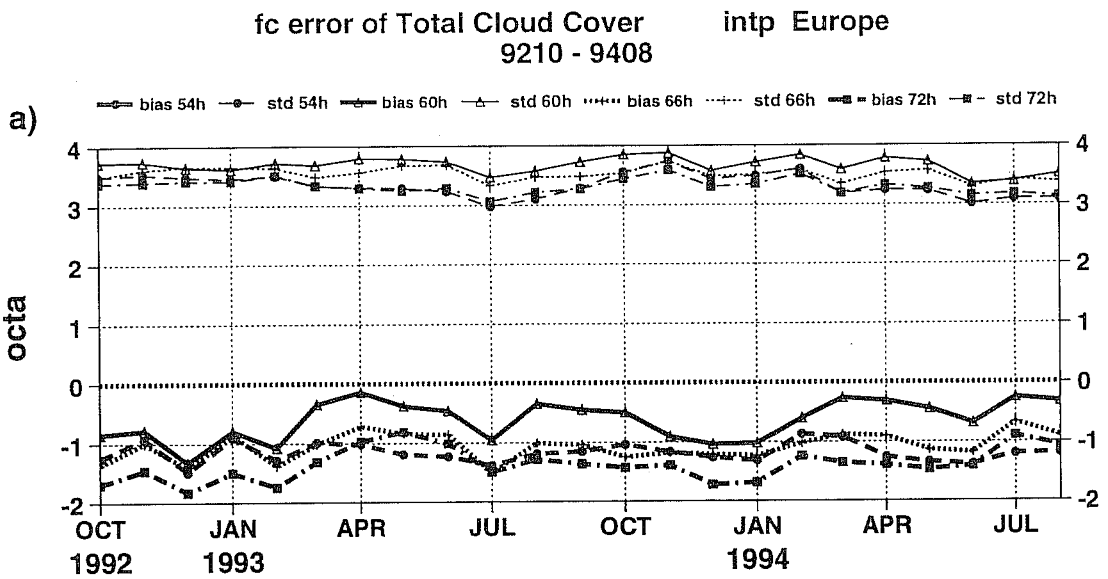
A further remark on interpolation should be made here: the coarser the grid from which one interpolates the more smoothing is applied to model values. Also, with lower grid resolution, the aforementioned representativeness problems become more critical. Especially interpolation from a grid that is already the result of an interpolation from the actual model grid can lead to biases and distort the results. Here, values are interpolated from the "reduced Gaussian grid" which yields fairly uniform grid distances around 60 km over the globe.

5. A SELECTION OF VERIFICATION RESULTS

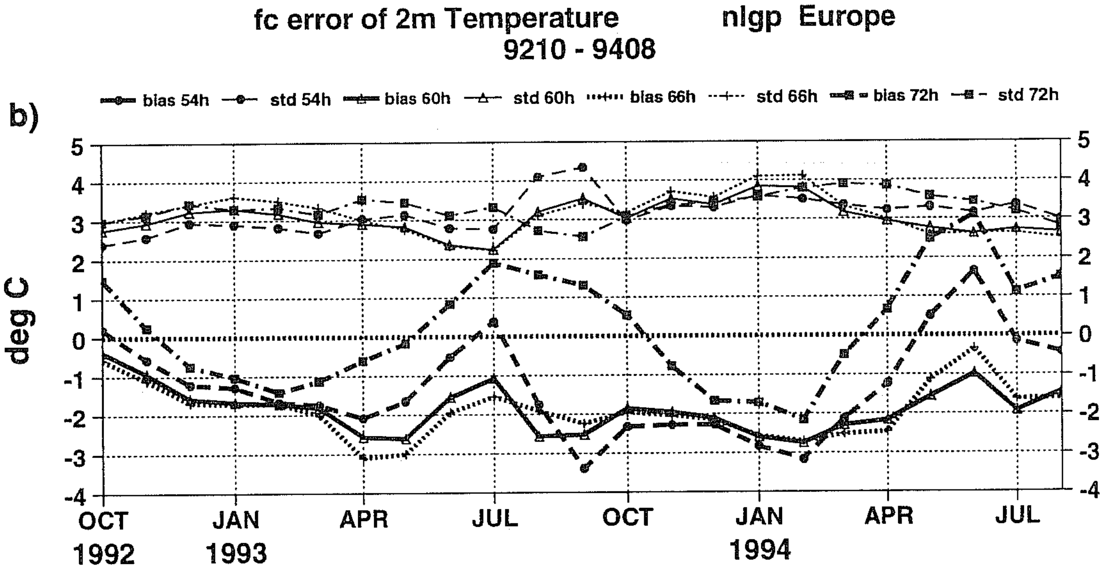
Some ECMWF direct model output verification results for the parameters total cloud cover, 2 metre temperature and precipitation over Europe are presented in Strauss and Lanzinger, 1993. In the paper presented here the results are extended to 2 metre specific humidity and 10 metre winds. Also, a few selected results for other global areas will be presented.

Fig. 2 illustrates monthly means and standard deviations of forecast errors for six weather parameters from October 1992 to August 1994 for Europe. In the following, the seasonal and diurnal characteristics of these errors are briefly summarized.

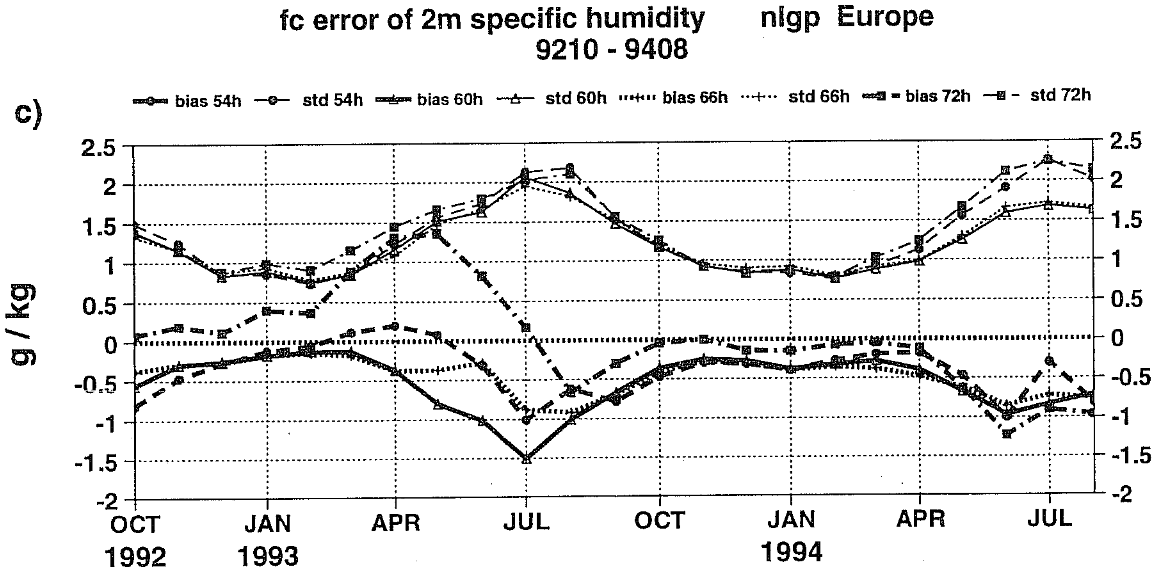
Total cloud cover is negatively biased, esp. at day time during the cold season. Record monthly biases reach almost -2 octa. The level of random errors, expressed as standard deviation of the forecast errors, is fairly constant throughout the year. **2 metre temperature** is negatively biased, around -2 to -3 degrees, at night time throughout the year, and there is a distinct warm bias at day time during the warm season. **2 metre specific humidity** is slightly negatively biased since the change in the surface and boundary layer scheme on 4 August 1993. The +1 g/kg day time bias of spring 1993 changed sign the following spring season 1994. Random errors are larger in the warm season, when the average moisture content



fc error of Total Cloud Cover intp Europe
9210 - 9408

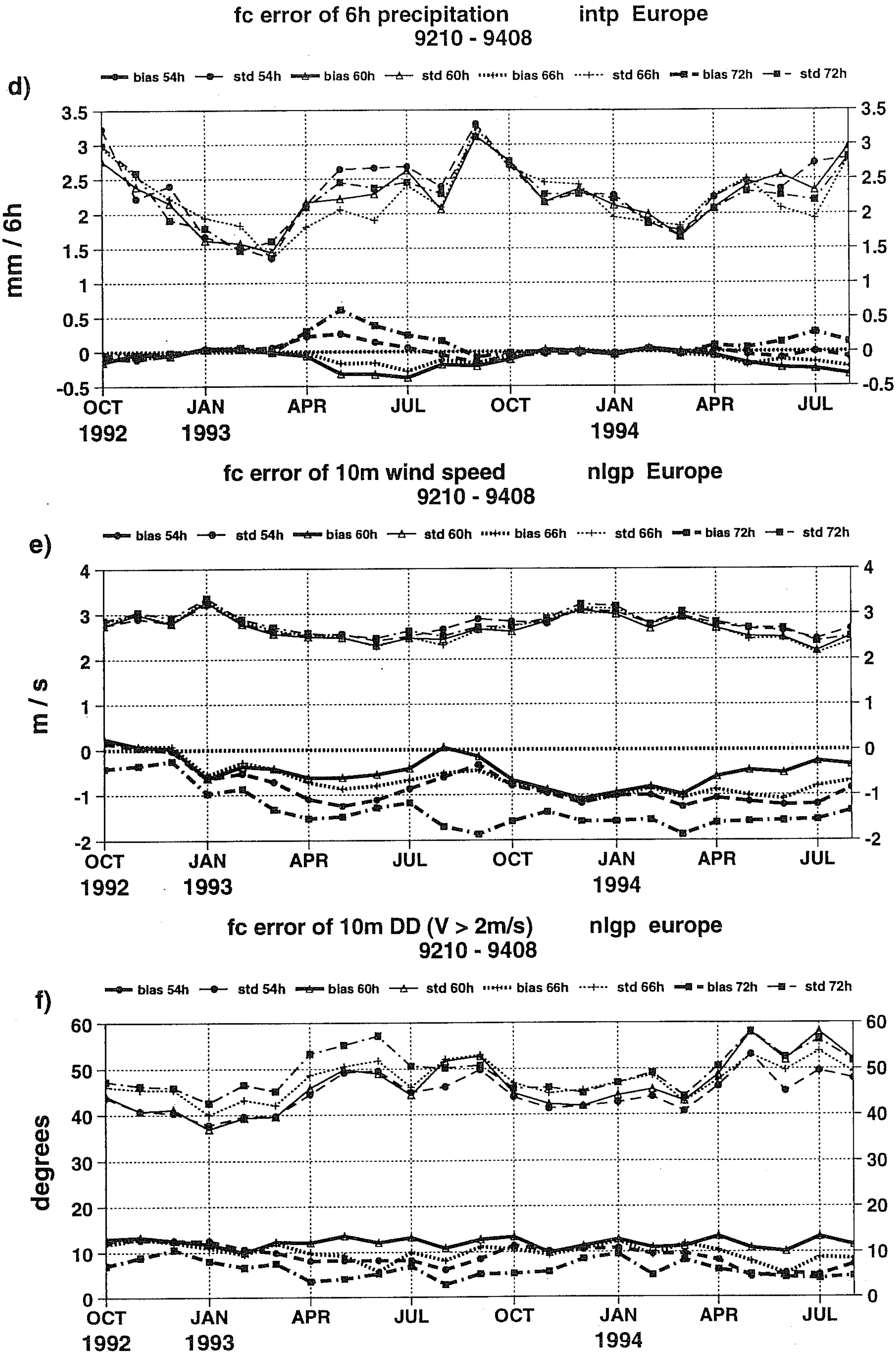


fc error of 2m Temperature nlgp Europe
9210 - 9408



fc error of 2m specific humidity nlgp Europe
9210 - 9408

Fig 2 Monthly values of bias and standard deviation of forecast errors for t+54, +60, +66 and +72 hours, as denoted in the legend, for Europe. a) Total cloud cover. b) 2 metre temperature. c) 2 metre specific humidity. d) 6 hour precipitation. e) 10 metre wind speed. f) 10 metre wind direction (for wind speeds exceeding 2 m/s).


fc error of 6h precipitation intp Europe
9210 - 9408
d)
bias 54h
std 54h
bias 60h
std 60h
bias 66h
std 66h
bias 72h
std 72h
mm / 6h
3.5
3
2.5
2
1.5
1
0.5
0
-0.5
OCT 1992
JAN 1993
APR
JUL
OCT
JAN 1994
APR
JUL
fc error of 10m wind speed nlgp Europe
9210 - 9408
e)
bias 54h
std 54h
bias 60h
std 60h
bias 66h
std 66h
bias 72h
std 72h
m / s
4
3
2
1
0
-1
-2
OCT 1992
JAN 1993
APR
JUL
OCT
JAN 1994
APR
JUL
fc error of 10m DD (V > 2m/s) nlgp europe
9210 - 9408
f)
bias 54h
std 54h
bias 60h
std 60h
bias 66h
std 66h
bias 72h
std 72h
degrees
60
50
40
30
20
10
0
OCT 1992
JAN 1993
APR
JUL
OCT
JAN 1994
APR
JUL

is higher. **Total precipitation** has very distinct seasonal behaviour, with practically no mean errors from October to April, and positive day time and negative night time biases during the summer half year. Note that the 1992 day time biases were also reduced after the above mentioned model changes. The standard deviation of precipitation errors is highest in autumn. **10 metre winds** have negative systematic errors in wind speed all year round. The wind direction is positively biased which means that model winds are too geostrophic.

The evolution of verification measures with forecast range is illustrated for the examples of temperature and specific humidity at 2 metres over the area of North America, valid for the verification period of January 1994. In figs. 3 a and b mean errors, mean absolute error and standard deviation of errors are plotted against forecast range up to t+120 hours. Whereas only little drift can be seen in the bias curves, standard deviation and mean absolute errors grow by about 50% from forecast day 1 to day 5 for both parameters.

The skill score with respect to mean absolute error

$$SS=1-\frac{MAE}{MAE_{ref}}$$

with MAE_{ref} denoting the mean absolute error of a trivial reference forecast, like in this case a persistency forecast, for the same verification set as before, is shown in figs. 4 a and b. A skill score of 0 means that the forecast is useless ($MAE = MAE_{ref}$), a value of 1 means a perfect forecast ($MAE = 0$). We see that there is considerable skill in the short range, peaking at day 2 - at day 1 persistency is often quite a good forecast. The curves drop towards day 5, but there is still significant skill. For cloud cover and 10 metre wind speed (not shown) for the same period and area, SS is much smaller (maximum values at day 2 of 0.20).

In a last example biases of daily precipitation over the extra-tropical northern hemisphere are compared to the tropical belt. Figs. 5 a and b depict time series of 24 hour precipitation, observed against forecast values for three summer months. Clearly, precipitation is overestimated over the tropics, at least over land areas, whereas in the extra-tropics negative biases dominate heavily. To illustrate the geographical distribution of verification sites, data coverage maps are included in the figures.

The examples presented are only a small selection of weather parameter monitoring activity at ECMWF. It is clear that for global assessment many different climatic regions have to be examined. Also, a lot of emphasis is being put on the diurnal behaviour of model errors, to detect possible shortcomings in the parametrization of physical processes in the soil, at and near the earth's surface and processes associated with (convective) precipitation.

Once more it should be stressed that the results relate to direct model output and do not represent the full potential accuracy and skill of weather parameter forecasting with a global numerical model.

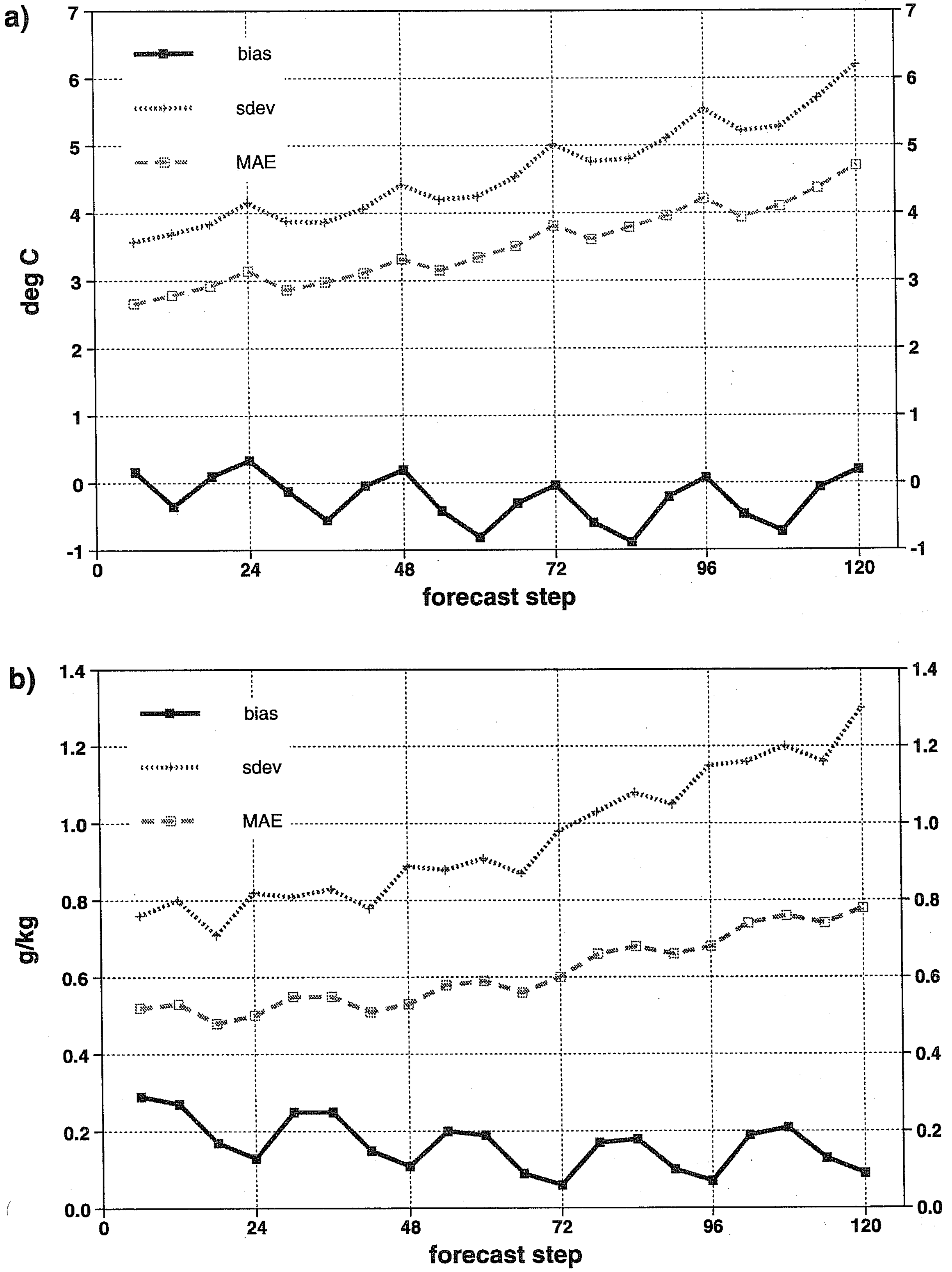


Fig 3 Bias, standard deviation of forecast error and mean absolute forecast error plotted against forecast range t+6 to t+120 hours, for January 1994 over North America. a) 2 metre temperature. b) 2 metre specific humidity.

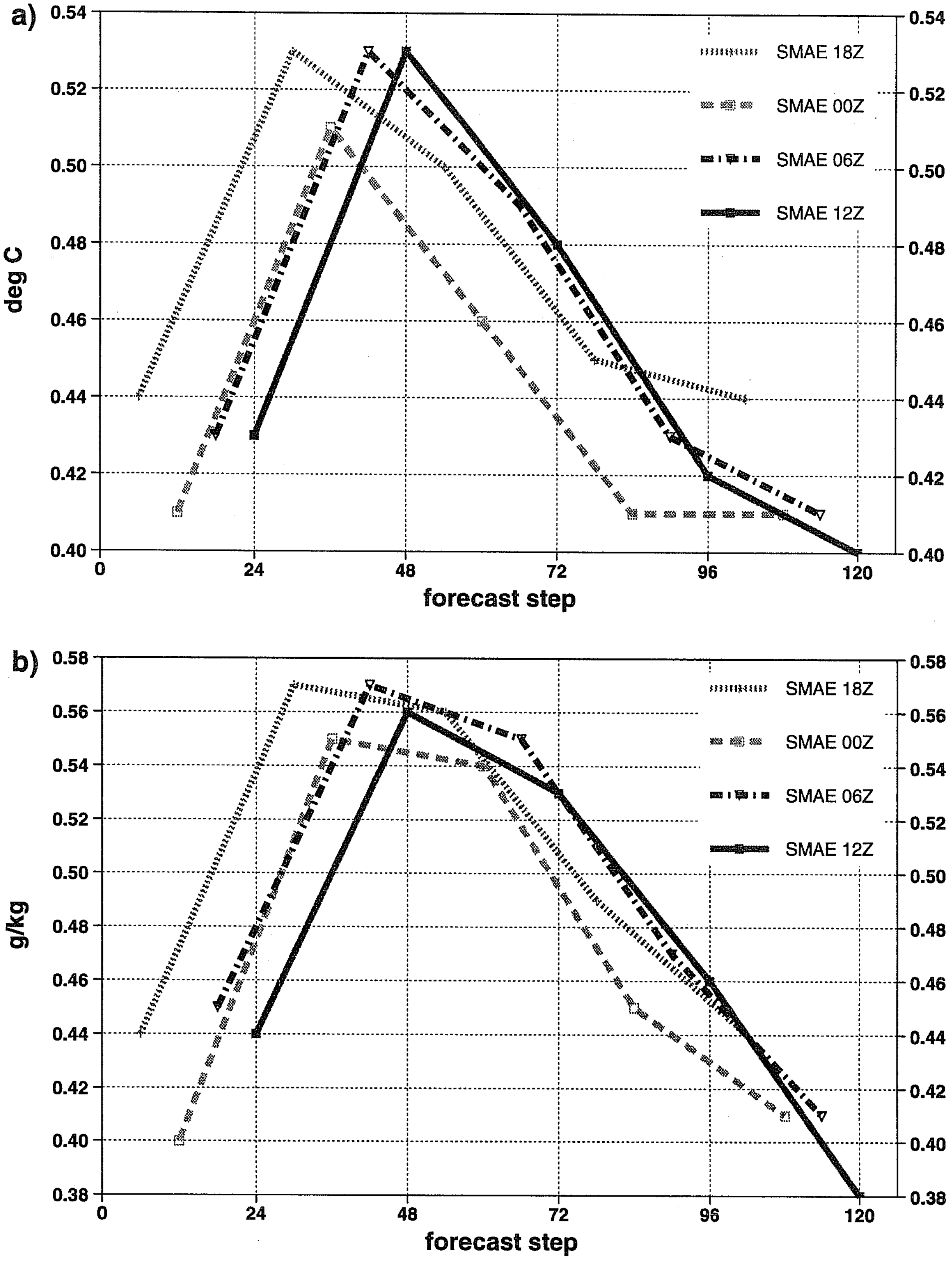


Fig 4 Skill score with respect to mean absolute error, as for Fig 3.

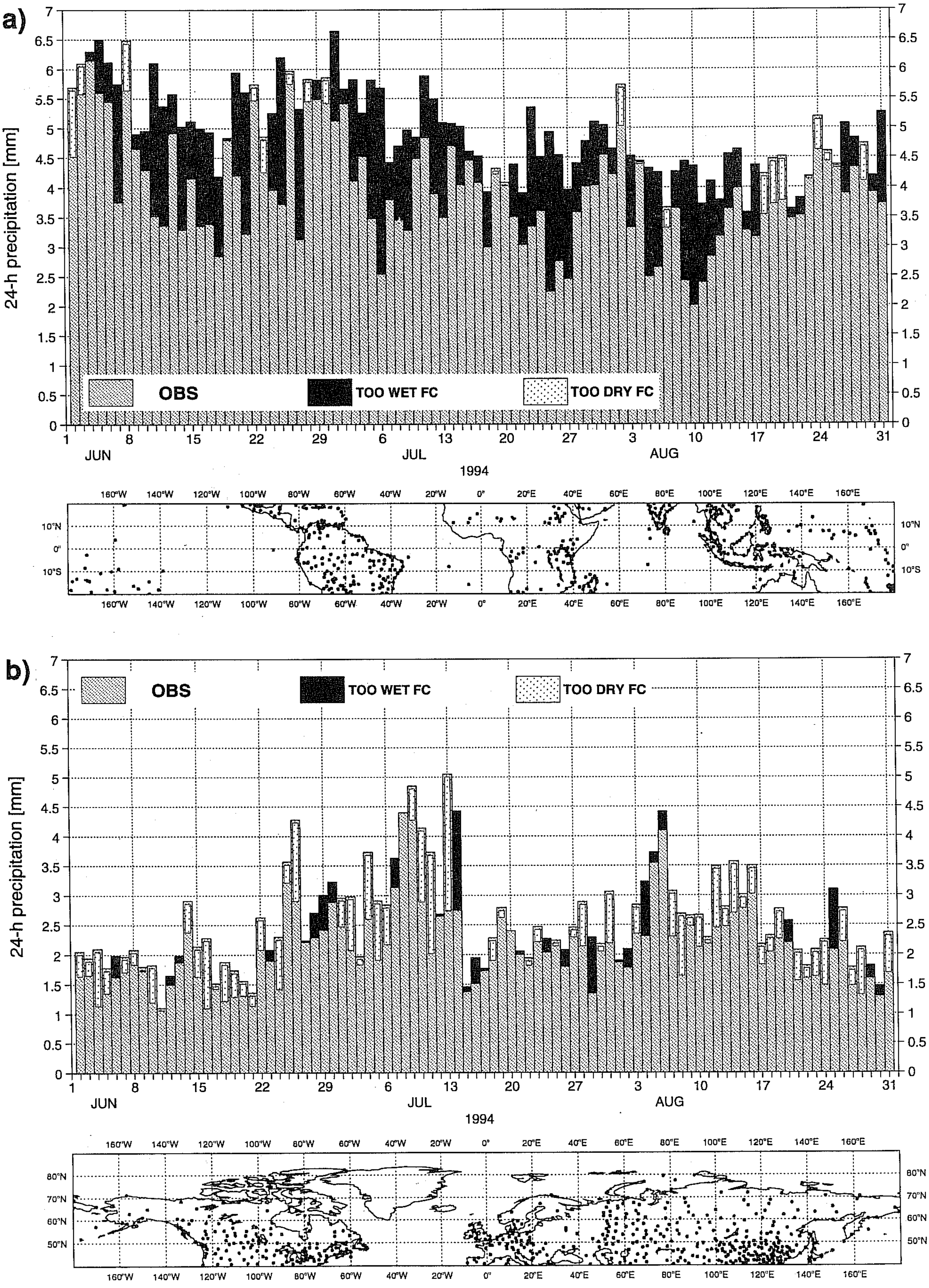


Fig 5 Time series of total precipitation accumulated over 24 hours, for June, July and August 1994. Mean observed values (heavily shaded bars), positive bias (solid on top of obsered), and negative bias (light shading in upper portion of observed value bars). a) Mean over tropical belt stations. b) Extra-tropical Northern Hemishpere north of 40 degrees latitude. All used observation stations are indicated in the attached maps. (European stations have been "thinned out")

REFERENCES:

Murphy A., 1985: Proposed Standard Procedures for Verification of Local Weather Forecasts. WMO PSMP Report No. 15.

Strauss B., A. Lanzinger, 1993: Overview of Validation of Direct Model Output. ECMWF Seminar 1992, Seminar Proceedings Vol. II, 93-104.