Conversion of the NCEP re-analysis data to the netCDF format and quality control

R.E. Benestad

DNMI, October 6, 1999

Contents

1 Introduction

The NCEP re-analysis data were converted from GRIB format (*Dey*, 1998) to UNIDATA's netCDF format (*Rew* et al., 1996) in order to facilitate easier access to the data. Although there are analytical packages that can deal with the GRIB data such as GrADS, netCDF is by far more user-friendly and belongs to a newer generation of data format. NetCDF is in the author's opinion, far superior to GRIB in terms of usability. There are furthermore more applications which can deal with the netCDF format, and the netCDF format is much better documented.

The main purpose of this brief report is to serve as a documentation of the conversion program grib2ncdf.sh[1], and ascertain that conversion process has been correct. However, the report will also try to convey to the reader the ease and efficiency of working with the netCDF and the NOAA's Ferret (*S.Hankin & Denham*, 1994) graphics package. The quality control described here is by no means exhaustive, but is only meant to give a rough reassurance that the model data are realistic. This report also aims to give a brief overview over main features of the model data. This technical report is the third in a short series dedicated to the documentation of data format conversion, of which the first reports include *Benestad* (1998) and *Benestad* (1999).

2 Data and Method

Similar methods as described in *Benestad* (1999) was used here for quality control, but here the converted data plotted by Ferret were also compared with the original GRIB data displayed using GrADS (*Doty*, 1995) (not shown). Furthermore, the interpolated values of the NCEP data were compared to observations, such as the North Atlantic Oscillation Index (NAOI) and temperature and precipitation records. Although *Matlab* is inferior to Ferret when it comes to making high quality figures, *Matlab* was nevertheless used to produce figures 2 and 3 (The NAOI comparisons *ncep_naoi.m*) because the making of these figures involved some analysis and processing which were more easily done with *Matlab*.

The script *grib2ncdf.sh* compiles the FORTRAN code that reads the GRIB files and writes the netCDF files. This script both produces and

[1]The version of this code reading the NCEP data is slightly different to the one described in *Benestad* (1999). For instance, the dimensions used here (144×73) are different to those described in the previous report (128×64) and the GRIB codes are different, i.e. SLP is 2 here, whereas in the GSDIO data SLP is 151.

Table 1: *Examples of configuration types*

rdgrib /CDROM/data/monthly/prmsl.grb ncep_slp.nc > test_slp
rdgrib /CDROM/data/monthly/hgt.grb ncep_hgt1000.nc 1000 > test_hgt1000

runs an executable called *rdgrib*. The FORTRAN code *grib2ncdf.sh* and the executable *rdgrib* can run in two different configuration types (table 2).

For extraction of data with more than one level, a third argument was given to specify the height level. Following example shows the syntax for geopotential heights. The former *(a)* is used for GRIB files containing only one level (such as surface data), whereas the latter *(b)* picks the specified (here 1000hPa) level for a file containing more than one vertical level. If the file contains more than one level and no third argument (level) is specified, then the code will only extract those data at the same level as encountered during the first data slice read.

A listing of the *grib2ncdf.sh* code is given in the appendix of *Benestad* (1999). The original GRIB formatted data set on the NCEP re-analysis CD-ROM is briefly described in the appendix of this report. The version of *grib2ncdf.sh* used here was modified in order to give correct time information and correct details in the netCDF file header (*cr_ncdf.f* was also modified). Furthermore, the 180° longitudinal offset used in correcting the GSDIO longitudes (*Benestad*, 1999) was removed for the NCEP conversion. A UNIX script called *conv.sh* converts all the data files discussed in this report.

An example of a netCDF header is shown in the appendix (ncep_slp.nc), giving all the details about the data set dimensions and data attributes[2]. The converted data are global and are 10-13Mb in size, depending on how long the data records are. The GRIB files are slightly smaller: i.e. the SLP data require 10805184 bytes in the GRIB file whereas the netCDF file takes 12870736 bytes (about 20% larger). The large amount of data may result in memory problems when reading data. For instance, *Matlab* cannot manage all the data from one file on a computer with only 128Mb RAM due to insufficient memory. The solution is to specify a subregion, which the script manages to read. Alternatively, *Ferret* can extract a subset of the data and store in another file[3].

[2]This listing was produced using following UNIX command *ncdump ncep_slp.nc | more* (the netCDF library must be installed). The SGI *ncdump* executable has a bug which results in a mix-up between the time axis and the actual data (other UNIX versions do not have this bug).

[3]See *Benestad* (1999).

3 Results

Figure 1 shows the mean 1948-1994 NCEP sea level pressure (SLP) and SLP variance. A permanent North Atlantic Oscillation (NAO) resembling feature, as well as an almost symmetrical zonal band of high pressure in the mid-latitude can be seen. The polar low is also evident. The variance plot indicates strongest variability over Antarctica (hidden) and central Asia. The same features are evident in both the Ferret plots of netCDF and GrADS plots of the GRIB files (not shown).

Figure 2, upper panel, shows the estimated NAO index[4] (NAOI) from the NCEP gridded SLP compared with the observed NAOI obtained from the University of East Anglia (*Jones* et al., 1997). Similar low-frequency features can be identified in both records, but it is evident that the NCEP estimates are not identical to the observed NAOI. The figure shows indications of systematic errors over various sub-periods: From 1955 to 1968, there is a negative bias to the NCEP NAOI, but the NAOI from the NCEP data is systematically too high after 1970.

The correlation map between NAOI and the SLP in the lower panel of figure 2 provides a rough check of whether the NCEP SLP time series from Iceland and Lisbon were correct. This panel shows a spatial NAO structure similar to those reported in *Benestad* (1999), and therefore demonstrates that the SLP time series were obtained from the right locations and that the spatial structure of the NAO is well reproduced in the NCEP SLP.

Although any in-depth quality control is beyond the scope of this report, the results from the NAOI comparison are examined further in figure 3 in the attempt to uncover an explanation for the strange errors. It is evident from the upper panel of this figure that the NCEP NAOI excursions before 1970 often went too far down and that several positive anomalies seen in the observations were not seen in the NCEP data. There is a sudden systematic shift in the NCEP behaviour (with respect to the *Jones* et al. (1997) NAOI) around 1970, after which the NCEP NAOI describe too strong positive NAOI excursions and to weak NAO anomalies. The lower panel shows the SLP differences over Iceland between NCEP and a number of different data sets. There are no obvious systematic shifts around 1970 in these estimated SLP differences. The low-pass filtered differences are small in comparison to the individual errors. NAOI estimates from Matlab were compared with those estimated with Ferret, which is a completely different numerical tool, to reduce the chances of errors in the analysis. The upper panel in figure 4 shows

[4]Estimated according to: [SLP(9.10°W,38.70°N) - SLP(22.44°W,65.50°N) - mean(SLP diff)]/stdv(SLP diff).

the comparison between the 2 estimates together with the Jones NAOI, indicating small differences between the two approaches and little risk of errors in the analysis. Finally, GrADS was used to estimate the NAOI directly from the GRIB data in order to check to possibility of conversion errors, and the results from GrADS are shown in the lower panel of figure 4. The NAOI record from GrADS is not identical to the red or green curves in the upper panel, although there are many similarities. The curves in the two plots have been produced using different types of filtering techniques and constant offsets. The GrADS curve is reconciled with the Matlab results if a boxcar type filter is used (not shown), and it is clear that the errors in the NAOI is not a result of the conversion to the netCDF format.

The NCEP data were estimated from interpolation of $2.5° \times 2.5°$ gridded data, and the SLP values in the re-analysis data set therefore represent the large scale geographical structures (average SLP over each grid box). There are furthermore probably errors in both the reanalysed SLPs from NCEP and the measured SLPs used in the calculation of the NAOI. The lower panel of figure 3 shows SLP differences of 4-8hPa between NCEP and various data sets. Low-pass filtering of random white noise may introduce low-frequency fluctuations (red noise character) in a similar fashion as a random walk process *Pendlebury* (1985), which is demonstrated in figure 5. In both cases, the index is the (small) difference between two large noisy numbers, and small errors in each SLP observation may lead to relative large errors in the NAOI index. If the NCEP data are subject to systematic model errors then these are expected to be constant with respect to time, implying that the "systematic shifts" in the NAOI are unlikely to be caused by model shortcomings.

Figure 6 shows the mean surface temperatures (upper panel) and the geographical distribution of the strength of the temperature variability (lower panel - variance). As the annual cycle was not removed prior to this analysis, the variance map is expected to be dominated by the seasonal variations. A meridional temperature gradient is evident and the west Pacific warm pool is also visible. Most of the temperature variability is seen over land and in high latitudes, and especially over Antarctica and Siberia. Figure 7 shows mean values (upper) and variance (lower) of the 2-meter temperatures. The 2-meter temperatures tend to follow the surface temperatures in general. The impact of the warm gulf stream extension, The Norwegian current, is clearly visible in the mean temperature map.

The lower panel indicates strong variability over Antarctica. Figure 8 shows comparisons between interpolated NCEP temperatures and observed time series from Oksøy fyr (a lighthouse at 58.07°N - 8.05°E) and Kjøremsgrendi at Dombås (62.06°N - 9.03°E), both long homogeneous high quality temperature records. The NCEP re-analysis temperatures estimated for Oksøy fyr

were on average higher than the station data (upper panel). This offset was mainly due to the fact that the NCEP winter temperatures were not as cold as the station data. It is possible that the differences between the two data sets are due to the low resolution in the NCEP data and that the spatial interpolation gives too strong weight to a maritime climate type. According to this explanation, one would also expect lower maximum temperatures, so there seems to be a slight mean bias towards higher temperatures. The comparison between Kjøremsgrendi and NCEP (lower), on the other hand, suggested a better correspondence between the two data sets.

Figure 9 shows the mean precipitation map and the strength in the precipitation variability. Most of the rain falls in the tropics along the inter tropical convergence zone (ITCZ) and the south Pacific convergence zone (SPCZ). The rain-forests over Brazil also receives a significant amount of rain and the storm tracks are associated with local precipitation maxima. The mean precipitation map shows the well-known subsidence regions with little precipitation off the west coasts of the sub-tropical continents: off California/Mexico, Peru/Northern Chile, Western Sahara/Mauritania, Namibia, and Australia. The strongest variability in the precipitation is seen over India, associated with the South Asian monsoon system, and over Panama. Changes in the rainfall patterns associated with El Niño Southern Oscillation (ENSO) are also discernible as high variance over western Pacific and the SPCZ. There is furthermore high variance in the precipitation over the ITCZ and the storm tracks. Most of the variance shown here are associated with seasonal differences in the rain patterns (Monsoon, ITCZ), but interannual variability is also visible.

Figure 10 shows time series comparisons between interpolated NCEP precipitation and observed station values from Oslo-Blindern (upper, urban site at 59.57°N - 10.43°E) and Tromsø (lower, 69.39°N - 18.56°E). The NCEP data were multiplied with 30 in order to get an approximate estimate for the monthly accumulated rain fall instead of daily mean values. The observed station values in Oslo are of similar magnitude to the reanalysis, and there is a good agreement between the two data sets on most of the precipitation events. There are, however, differences between the peak values in the precipitation in both locations. Whereas there was no systematic bias in the maximum magnitudes in Oslo, the NCEP data consistently underestimated the amount of rain fall over Tromsø. The peak NCEP precipitation values were smaller than those of the station observations. The discrepancies may be due to spatial smoothing in the reanalysis, topographical biases or errors (but unlikely) in the station measurements, but may also be a result of systematic model errors. The NCEP data consists of precipitation averages for large grid boxes, and may therefore not capture local geographical variability

in the rainfall patterns. Large differences have been found between different reanalysis data sets (*Serreze & Hurst*, 1999; *Arpe & Rhodin*, 1999), and it is notoriously difficult to get a good description of local rain fall in global GCMs. The precipitation data from the NCEP data set must therefore be used with care.

Figure 11 shows the mean 1000hPa geopotential heights (Φ_{1000}) and the variance of this field. The 1000hPa surface is below the mean sea level over the southern seas. There are local permanent maxima over the sub-tropical seas, with the exception over the western Pacific. Local maxima are also seen over the Antarctica and the Himalayas (both where there are few *in situ* observations). Strong variability is seen over the Antarctica, the Himalayas, the Middle-East, North Pacific and the North Atlantic. The Φ_{500} field, shown in figure 12 is permanently higher in the tropics than elsewhere. The tropical variability is small, whereas Φ_{500} has high variance in the mid-latitudes.

Figure 13 shows the mean and variance maps for the 850hPa temperatures. The highest temperatures are found over the tropics and the strongest variance is seen over land in the mid-latitudes and over Antarctica. The Antarctic continent stands out as much colder than the rest of the world (-30 to -40°C).

The mean spectral specific humidity fields for 850hPa and 500hPa heights are shown in figure 14. Although these results seem reasonable, it is well known that humidity data from weather and climate models are highly uncertain and may be subject to significant systematic errors.

4 Summary

A quality control on a subset of the converted NCEP reanalysis GRIB files suggests that the converted (netCDF) files contain sensible data. The mean values and variance maps correspond to our expectations. There are, however, a few concerns regarding the NCEP data which probably are not related to the conversion itself, but more likely the re-analysis process of the raw data: *i)* There are differences between the estimated NAOI from the NCEP data and the NAOI from station data; *ii)* Figure 8 suggests a warm bias in the interpolated 2-meter NCEP temperatures with respect to 58.07°N - 8.05°E; *iii)* A low bias in the NCEP interpolated precipitation for Tromsø and differences between the peak values of the rain events.

The differences between the NAOI estimates from the NCEP data and the *Jones* et al. (1997) NAOI were scrutinized, but no obvious cause for these differences were found. An in-depth analysis of the SLP time series are outside the scope of this report, whose aim is to provide an overview over

the data and give a quick and crude reassurance that the converted data are the same as the original data. The NAOI discrepancies may be a result of minor "errors"[5] in the SLPs over Iceland and Lisbon respectively.

The interpolated NCEP temperatures seemed matched the observed station values closely, and did not show indications of reduced variance like the *Jones* et al. (1998) temperatures do (*Benestad* et al., 1999). Station observations may be biased because of local geographical effects, which are not described in the NCEP reanalysis because they are smaller than the grid scale. However, the warm bias in the NCEP 2-meter temperatures at Oksøy fyr suggests that the NCEP data may not always be appropriate for direct use in local climate studies. The NCEP data nevertheless seem to be a promising candidate for empirical downscaling studies because of similar variance as the station observations.

Although the local mid-latitudinal precipitation may be biased, the large scale rain fall pattern seemed realistic in the NCEP data set. Part of the differences between the local observations and the interpolated values from the gridded data set may be due to local topographical effects unresolved by the NCEP grid.

Despite shortcomings with respect to local climate variables, the NCEP data set may be extremely useful in empirical downscaling studies. The fact that the NCEP data may contain systematic model biases may not necessarily be a problem if the climate data from the GCM predictions contain similar biases and if these errors are unimportant with regards for the regional climate.

[5]NCEP gives area averaged SLP over large grid boxes, which means that the NCEP data are spatially smoothed. The station observations and the gridded data set are therefore strictly not directly comparable, and some of the differences between the two data sets may be due to this spatial smoothening.

5 Acknowledgments

Most of the plots in this paper were made with the NOAA/PMEL/TMAP graphics tool, Ferret. Many thanks to the NCEP/NCAR group who have produced these re-analysed data and made them available on CD-ROM and freely distributing these. Credit should also be given to Unidata who developed the netCDF library.

References

Arpe, K., & Rhodin, A. 1999. Differences in the hydrological cycle from different reanalyses - which one shall we believe? *Page 42 of:* ECMWF (ed), *Second International Conference on Reanalysis*ECMWF, for ECMWF/WCRP.

Benestad, R.E. 1998. *Description and Evaluation of the Predictor Data sets used for Statistical Downscaling in the RegClim.* Klima 24/98. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E. 1999. *Conversion and quality control of the ECHAM4/OPYC3 GSDIO data to the netCDF format and a brief introduction to Ferret.* KLIMA 27/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Benestad, R.E., Hanssen-Bauer, I., Førland, E.J., Tveito, O.E., & Iden, K. 1999. *Evaluation of monthly mean data fields from the ECHAM4/OPYC3 control integration.* Klima 14/99. DNMI, PO Box 43 Blindern, 0313 Oslo, Norway.

Dey, C.H. 1998. *GRIB.* 1 edn. NCEP Central Operations/NOAA. Office note 388.

Doty, B. 1995 (September). *GrADS.* COLA. V 1.5.1.12.

Jones, P. D., Raper, S. C. B., Bradley, R. S., Diaz, H. F., Kelly, P. M., & Wigley, T. M. L. 1998. Southern Hemisphere surface air temperature variations, 1851–1984. *J. Clim. Appl. Met.*, **25**, 1213–1230.

Jones, P.D., Jonsson, T., & Wheeler, D. 1997. Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and South-West Iceland. *International Journal of Climatology*, **17**, 1433–1450.

Pendlebury, J.M. 1985. *Kinetic theory.* Adam Hilger Ltd.

Rew, R., Davis, G., Emmerson, S., & Davies, H. 1996. *NetCDF User's Guide.* Unidata Program Center.

Serreze, M.C., & Hurst, C.P. 1999. Arctic precipitation in the NCEP/NCAR and ERA reanalyses. *Page 50 of:* ECMWF (ed), *Second International Conference on Reanalysis*ECMWF, for ECMWF/WCRP.

S.Hankin, & Denham, M. 1994. *Ferret Users Guide.* NOAA/PMEL/TMAP.

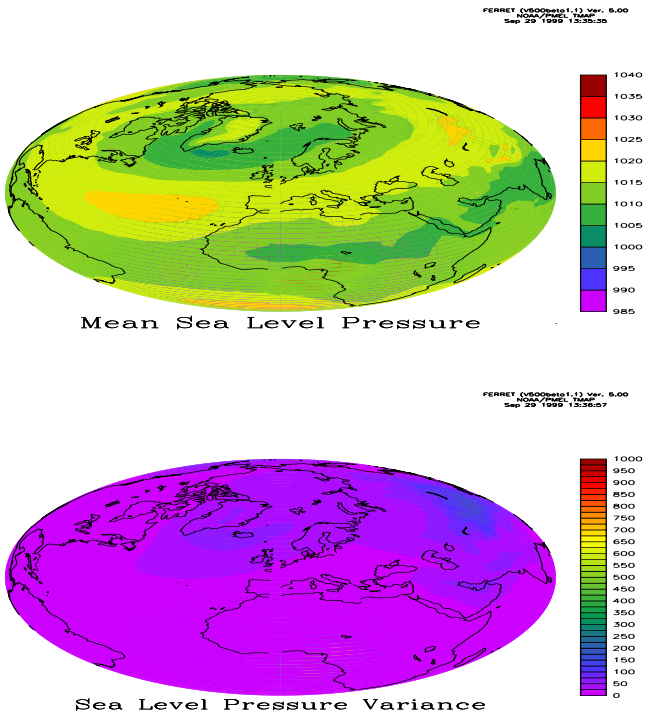


Figure 1: Map of mean SLP from the netCDF data file.

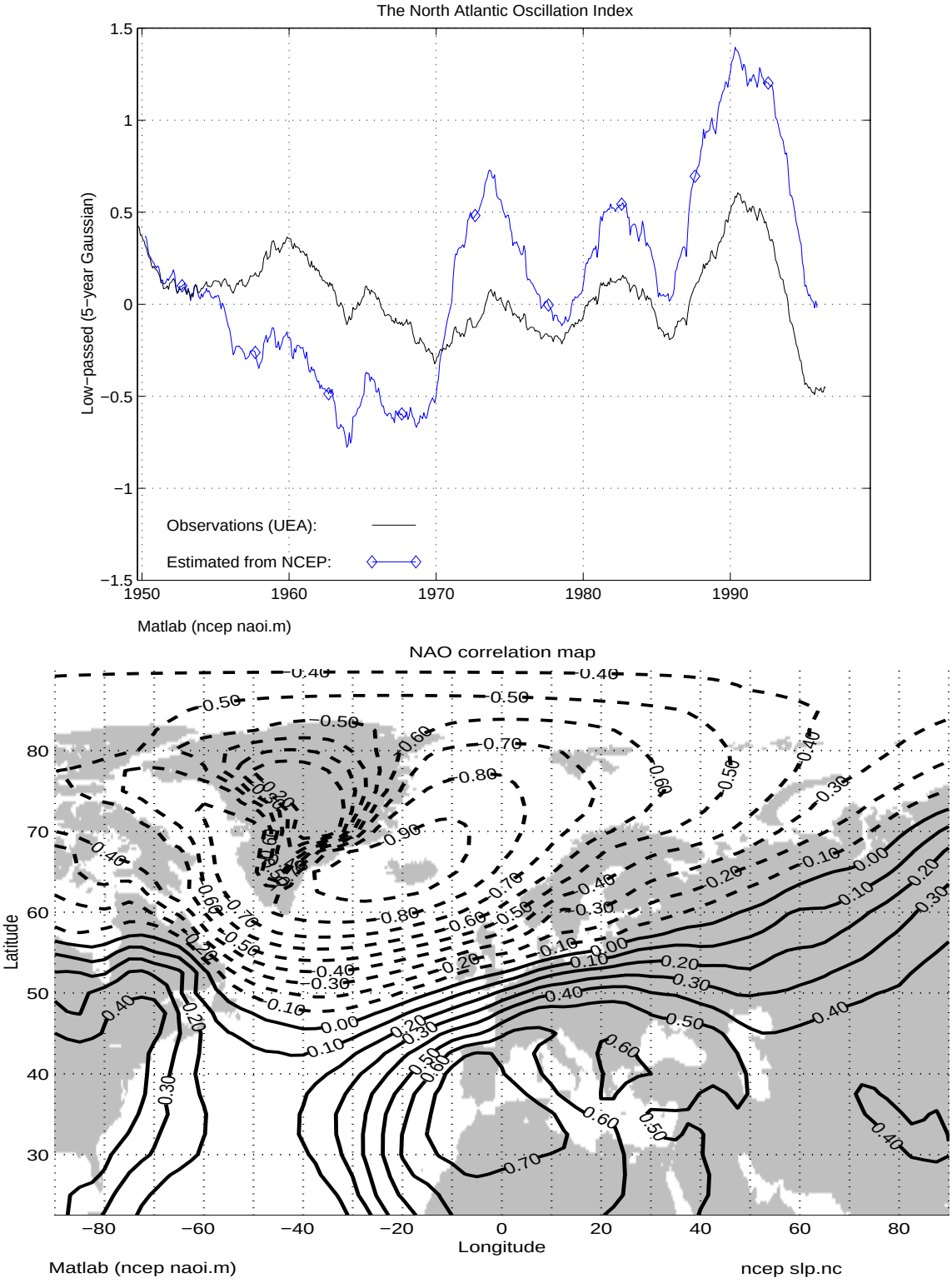


Figure 2: Upper: The North Atlantic Oscillation Index from the NCEP reanalysis (blue) and observations from UEA web site (black). Lower: Correlation maps SLP-NAOI.

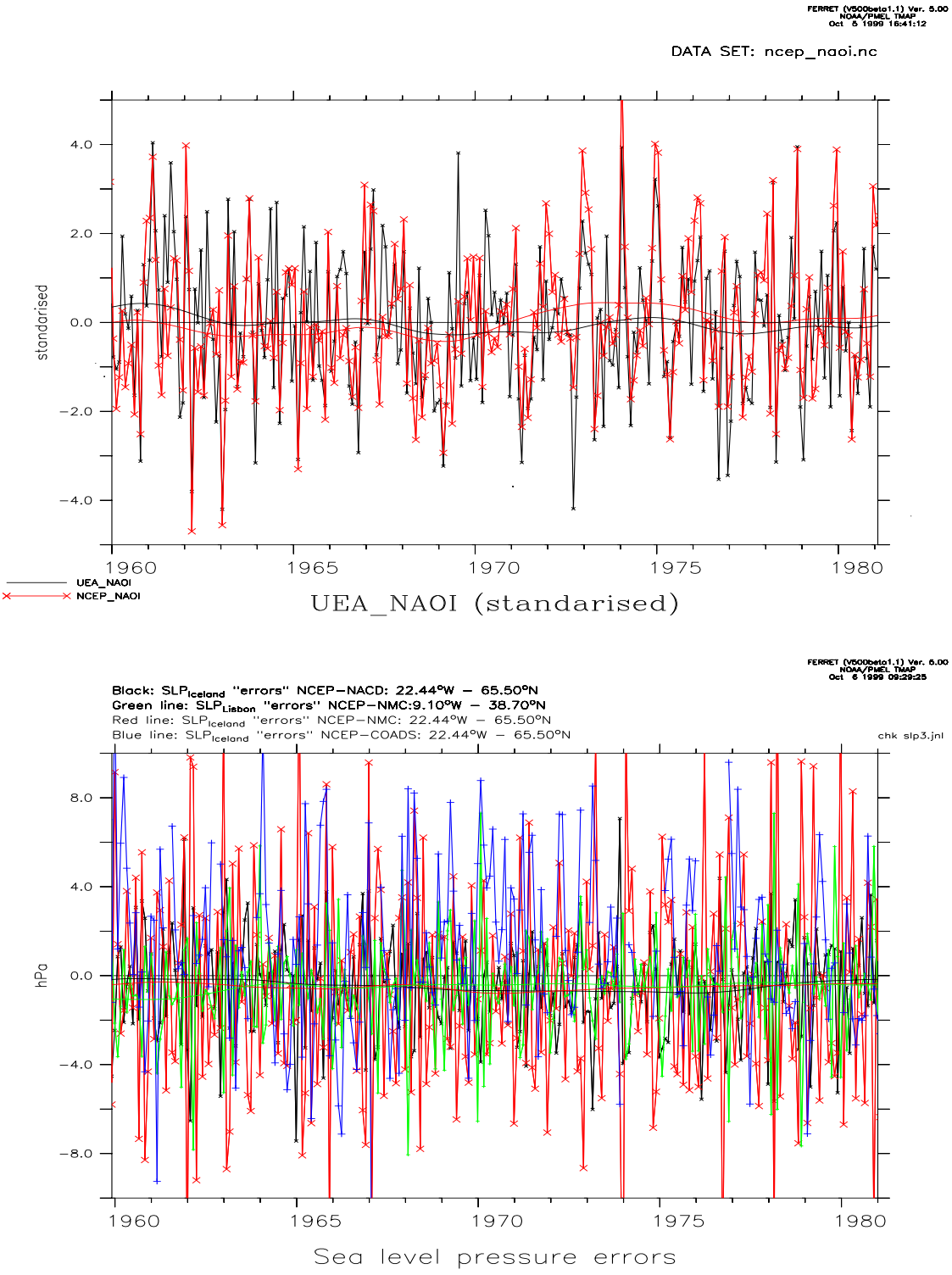
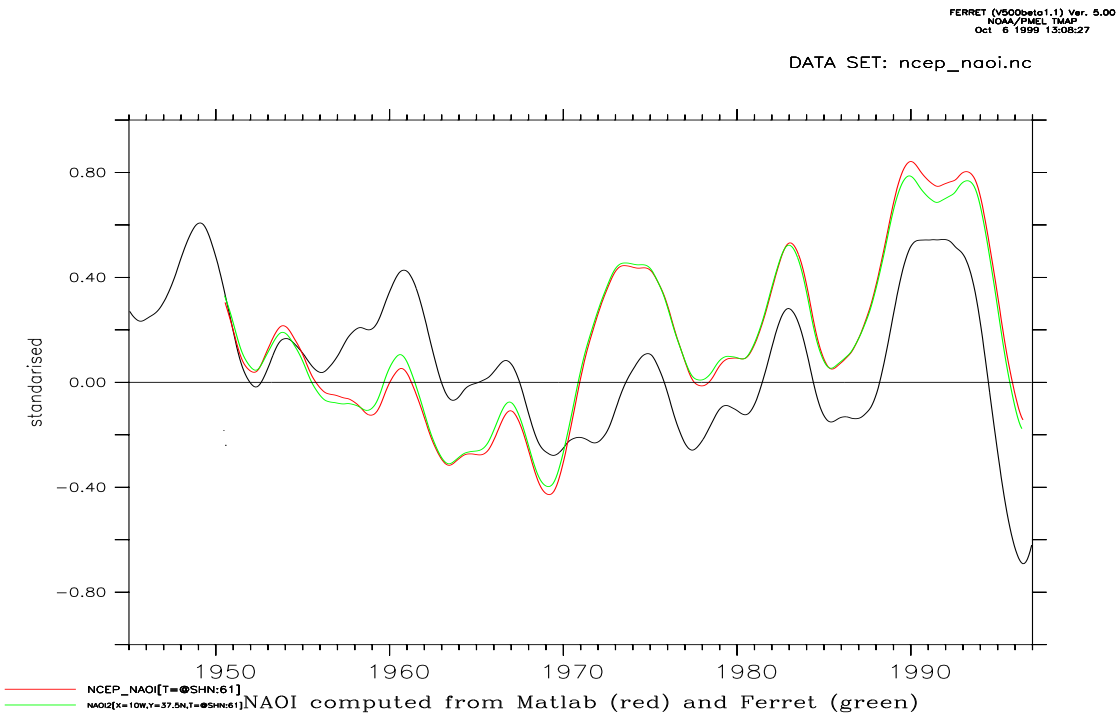


Figure 3: Upper: Unfiltered and low-passed NAOI records from *Jones* et al. (1997) (black) and the NCEP reanalysis (red). Lower: SLP differences between NCEP and NACD (black), COADS (blue) and NMC (red) over Iceland and between NCEP and NMC over Lisbon (green).



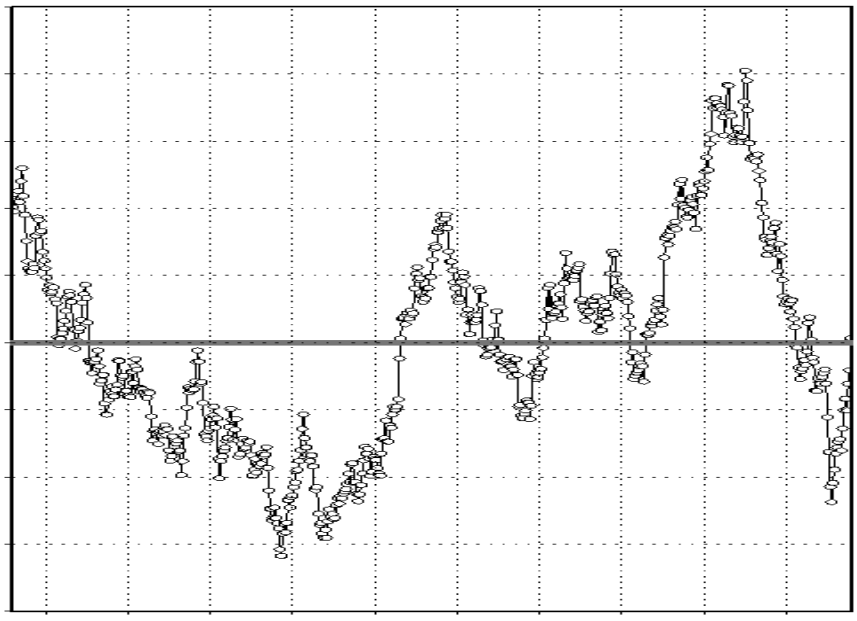


Figure 4: Upper: Low-passed NAOI records from *Jones* et al. (1997) (black) and the NCEP reanalysis estimated with Matlab (red) and Ferret (green). Lower: The estimated NCEP NAO index using GrADS reading from the GRIB formatted files directly. The figure has been produced using a 61-month boxcar filter (moving average).

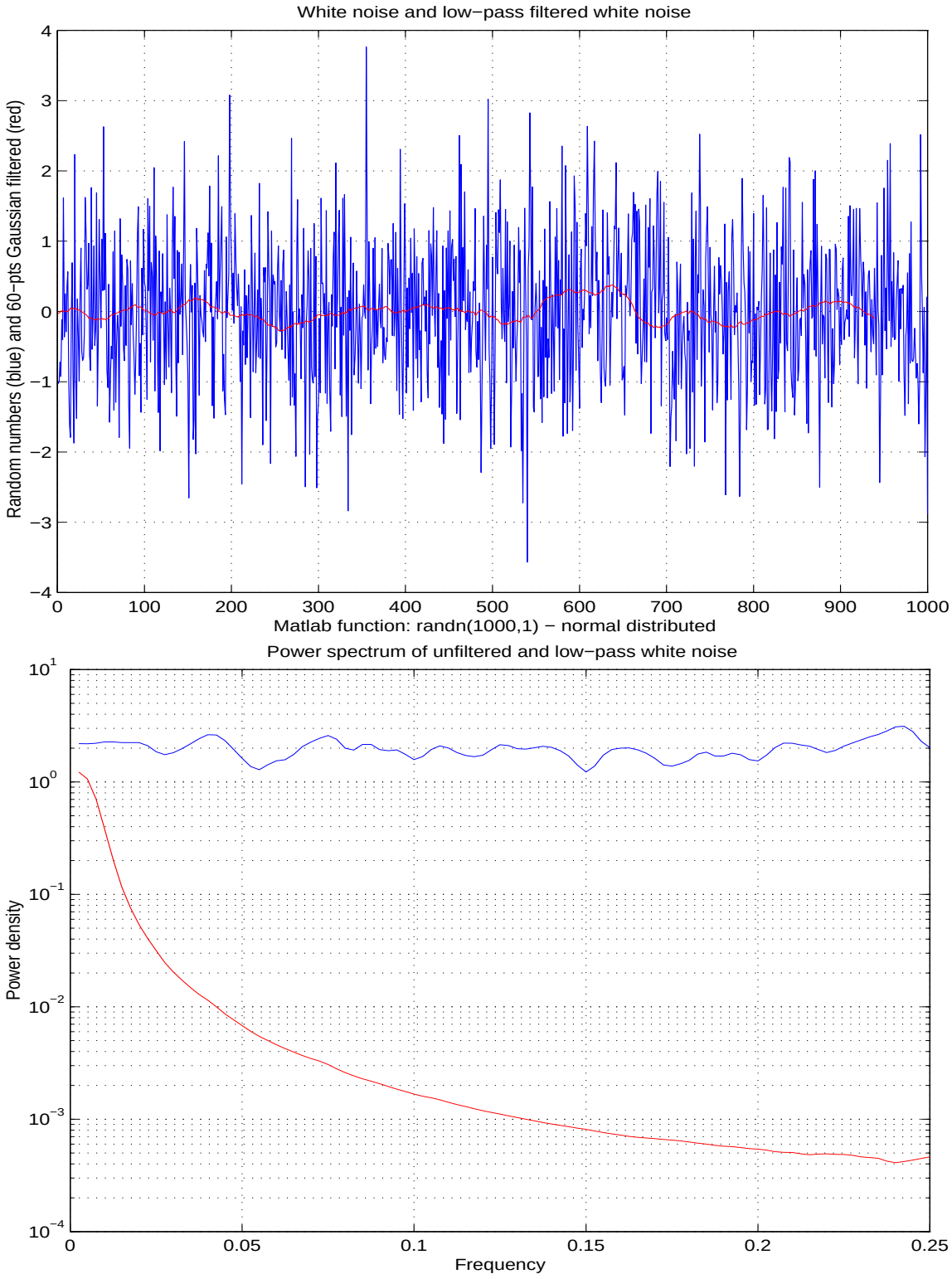
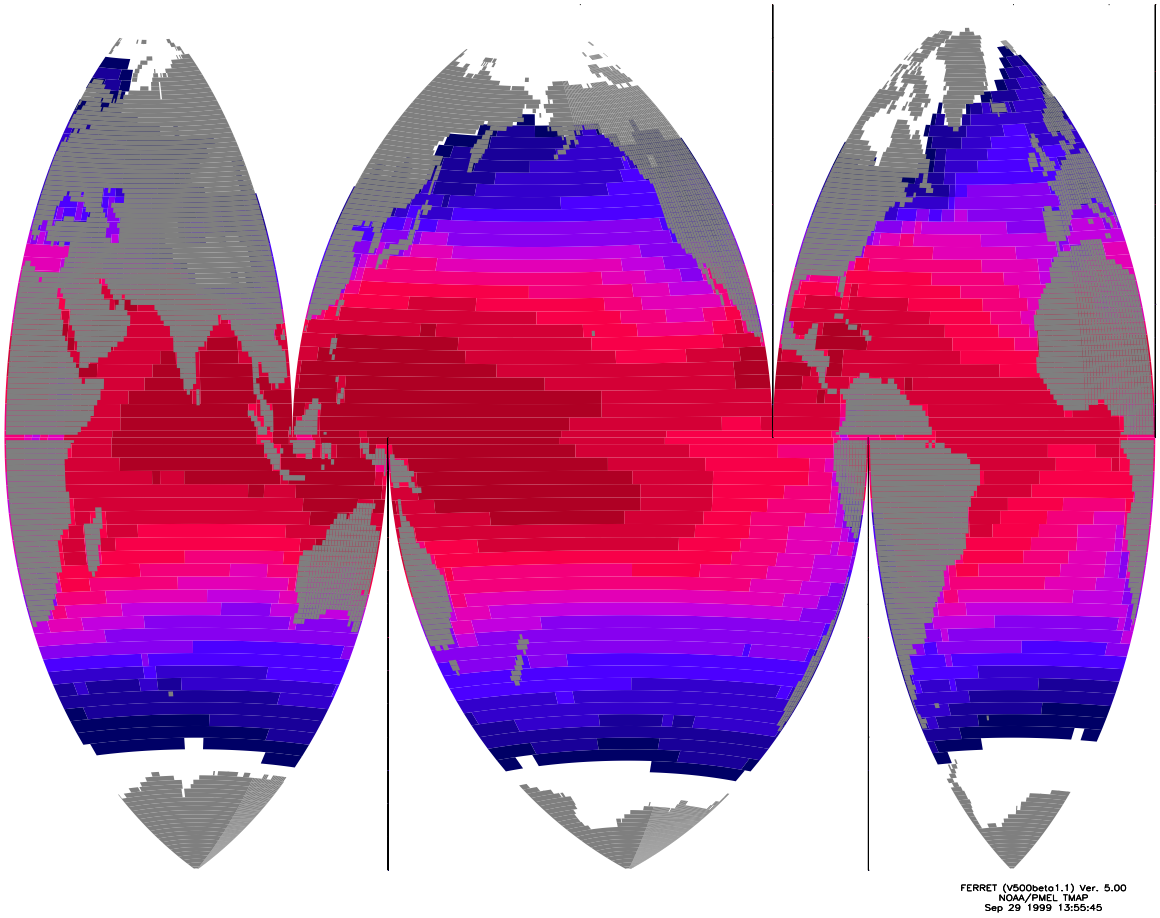


Figure 5: Upper: Unfiltered and low-passed white noise generated using Matlab's normally distributed random number generator. The white noise record is 1000pts long, and the low-passed results were obtained using a 60-pt Gaussian filter. Lower: The power spectra of the time records shown in the upper panel. The window width is 100.



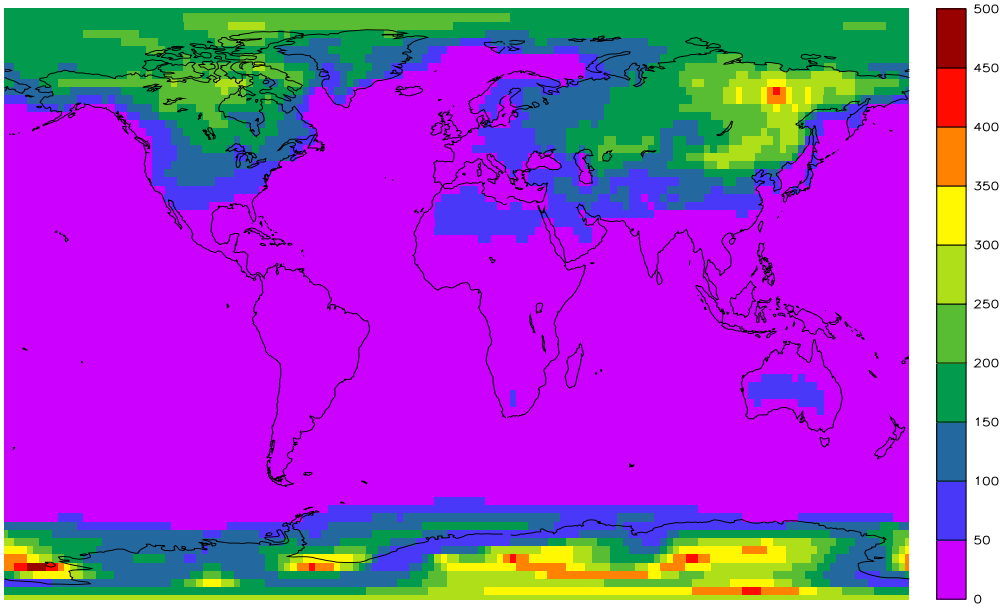


Figure 6: Map showing mean surface temperatures from the netCDF data file (upper). Lower panel shows the geographical distribution of variance.

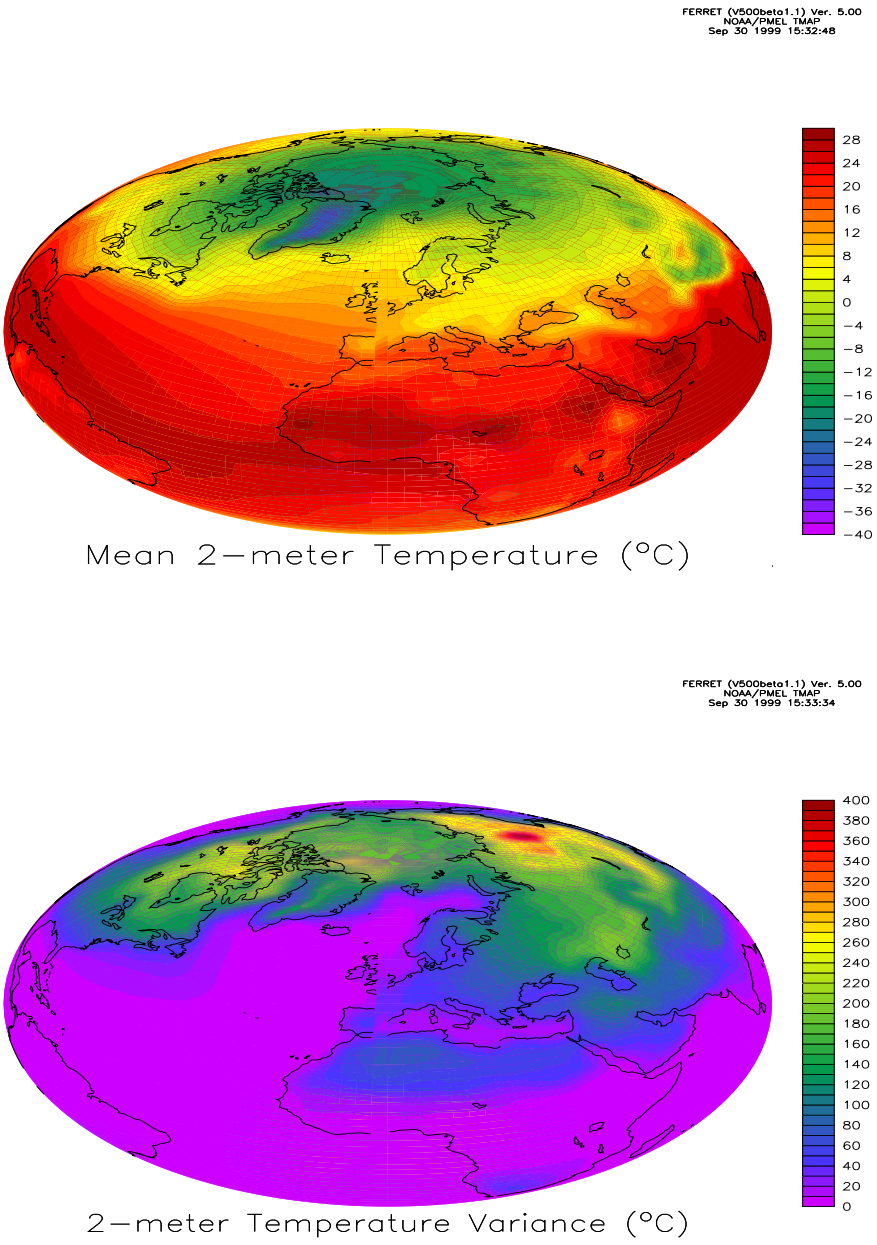


Figure 7: Map showing mean 2-meter temperatures from the netCDF data file (upper). The lower panel shows how the variance varies geographically.

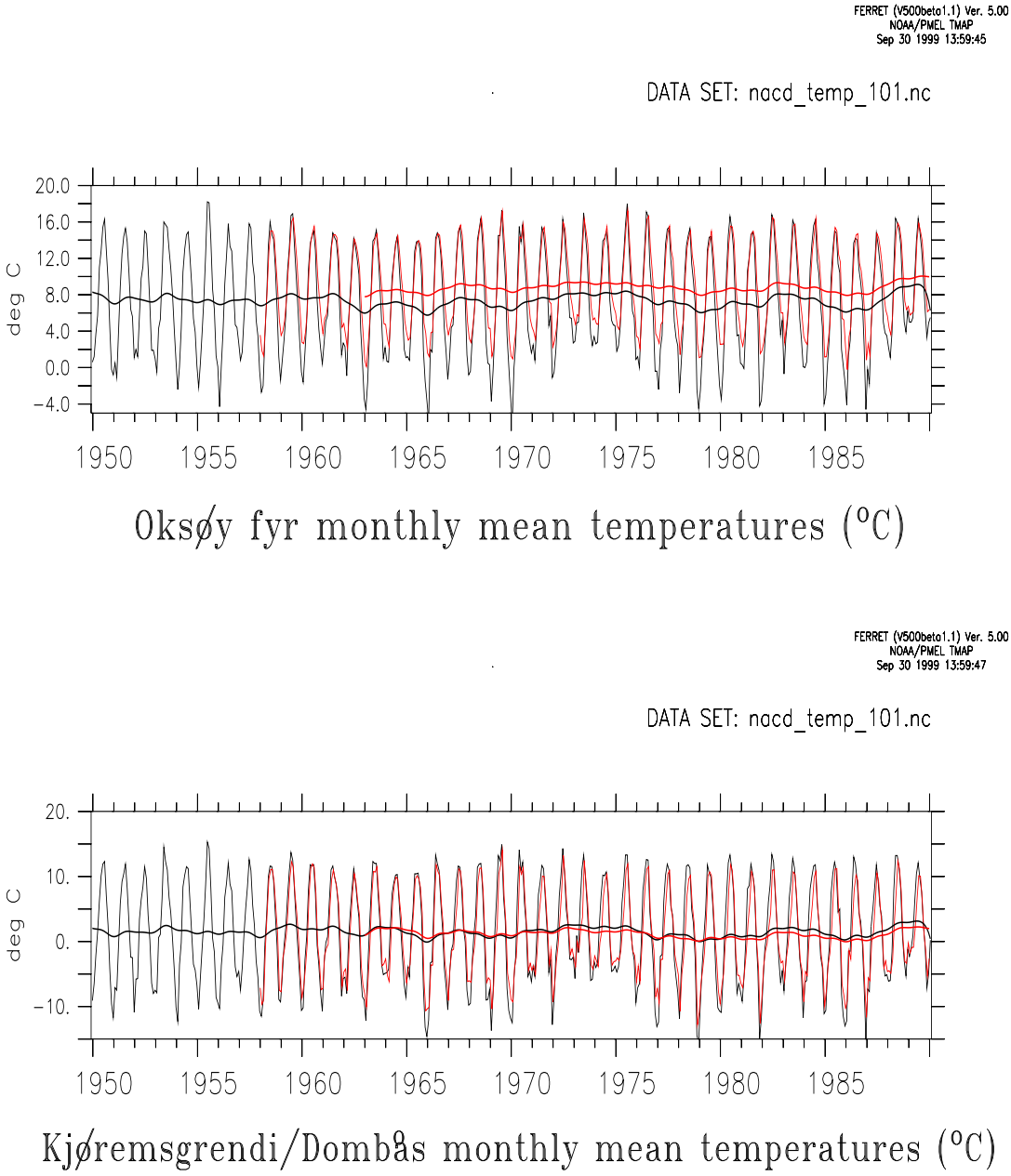


Figure 8: Comparison between observed temperatures from Oksøy fyr and interpolated values from NCEP (upper) and Kjøremsgrendi and NCEP (lower). The station observations are shown in black and the interpolated NCEP values are represented by read lines. The thick curves are 10-year low-pass filtered data (Ferret's Binomial filter).

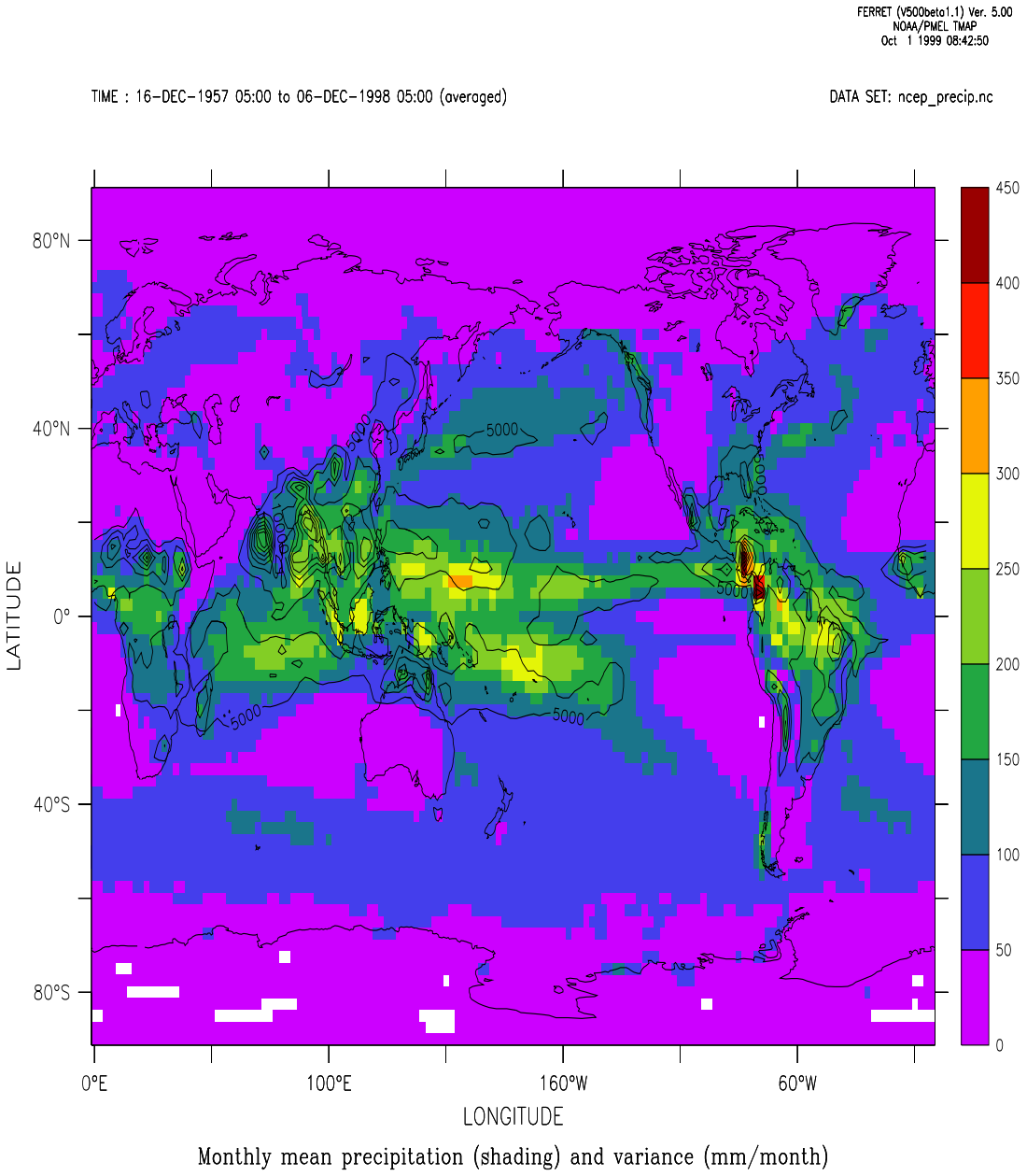
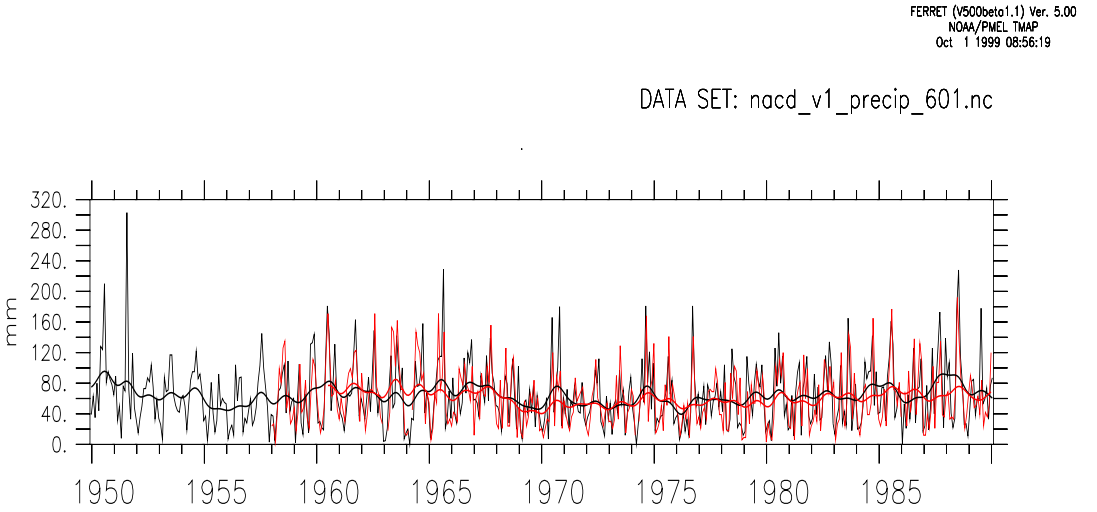


Figure 9: Map showing mean precipitation from the netCDF data file (shading). Contours indicate the spatial patterns of variance.



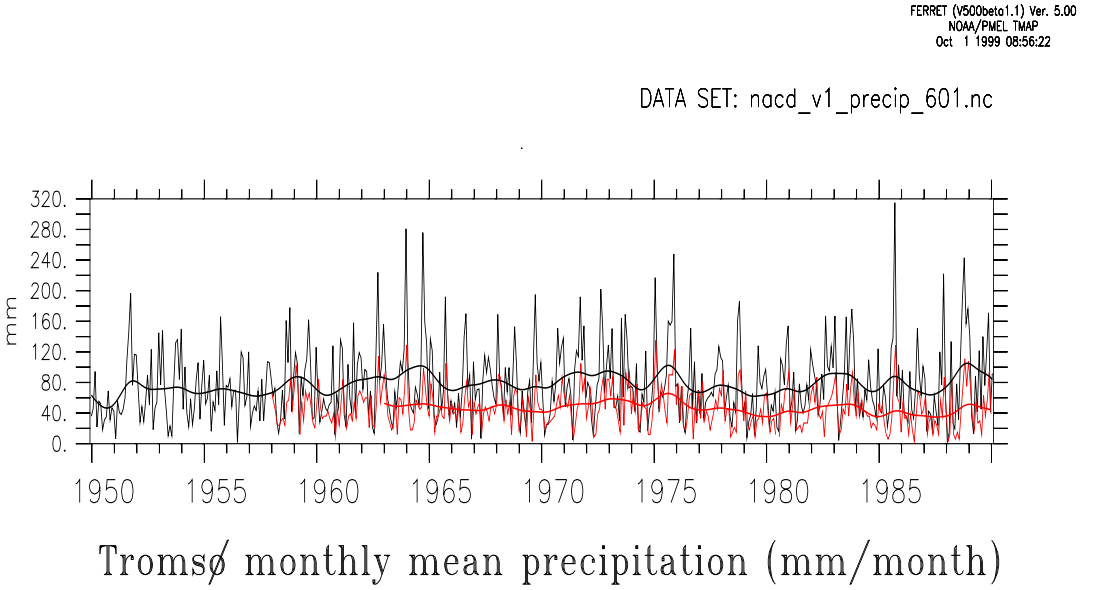
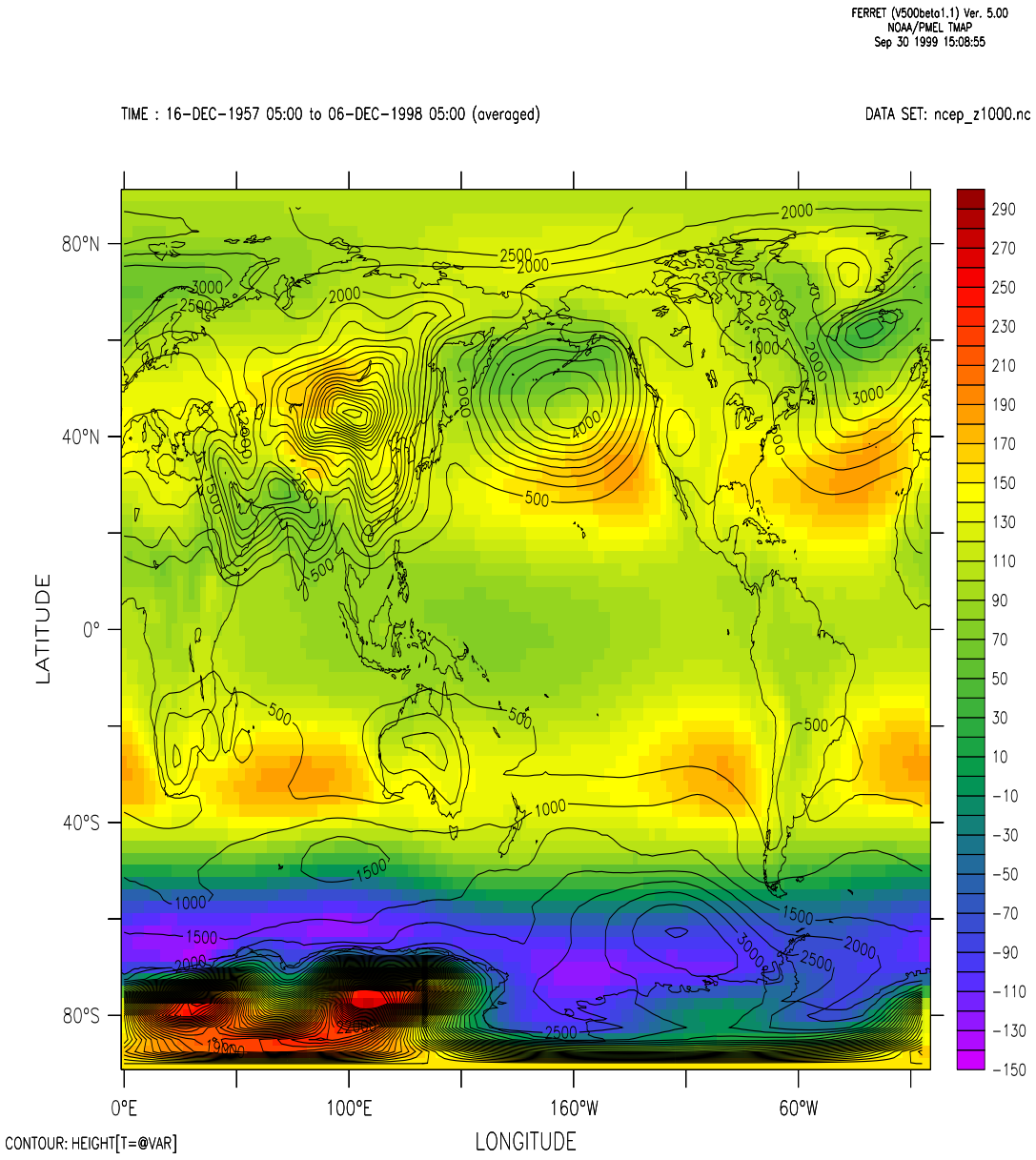


Figure 10: Comparison between observed precipitation from Oslo-Blindern and interpolated values from NCEP (upper) and from Tromsø and NCEP (lower).



Geopotential heights (grib 7) (m)

Figure 11: Map showing mean 1000hPa Φ from the netCDF data file. Contours indicate the spatial patterns of variance.

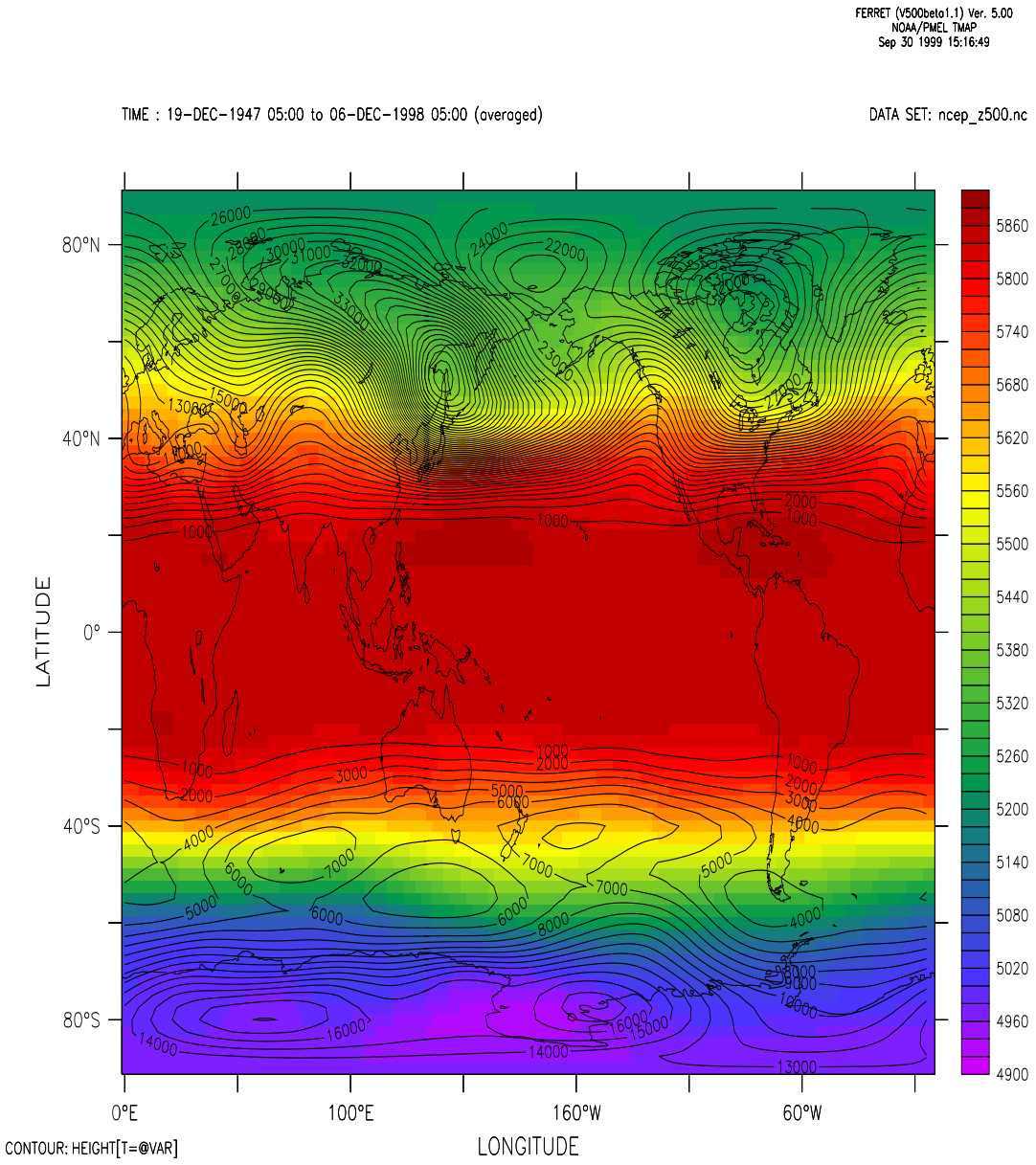


Figure 12: Map showing mean 500hPa Φ from the netCDF data file. Contours indicate the spatial patterns of variance.

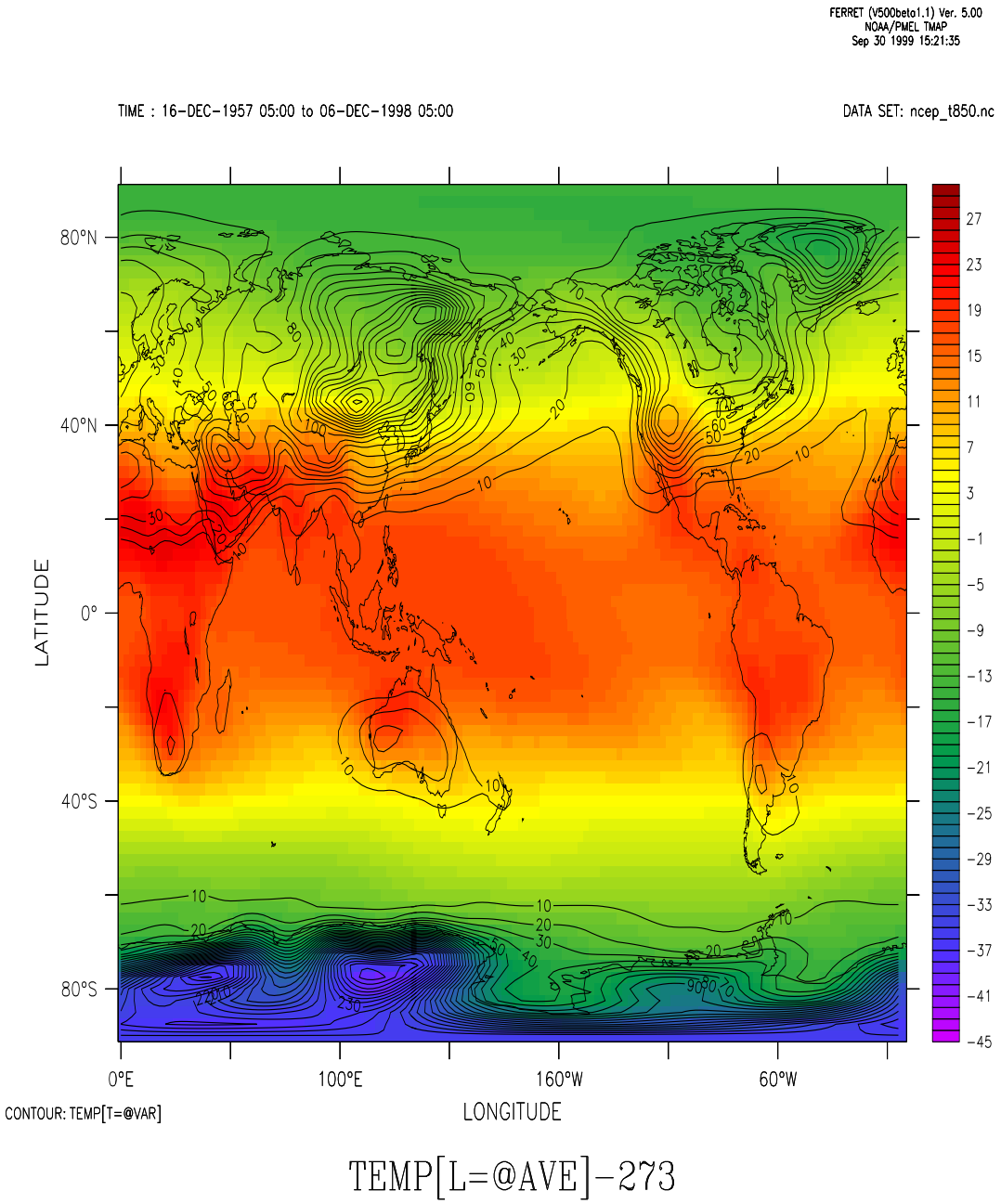
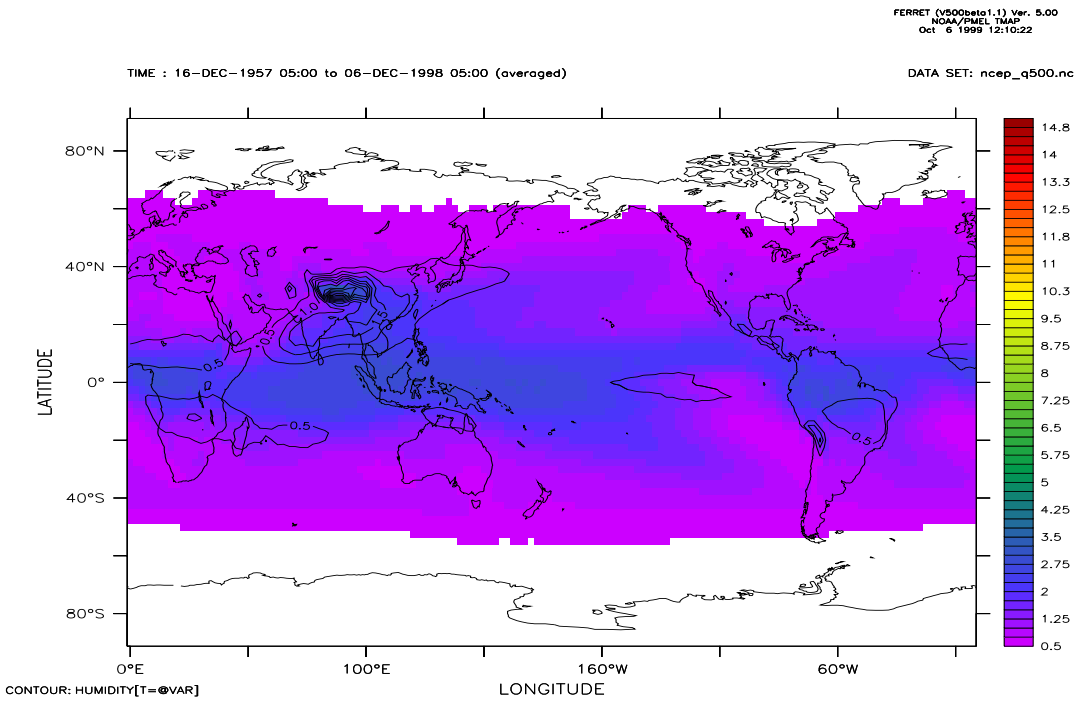


Figure 13: Map showing mean 850hPa temperatures from the netCDF data file. Contours indicate the spatial patterns of variance.



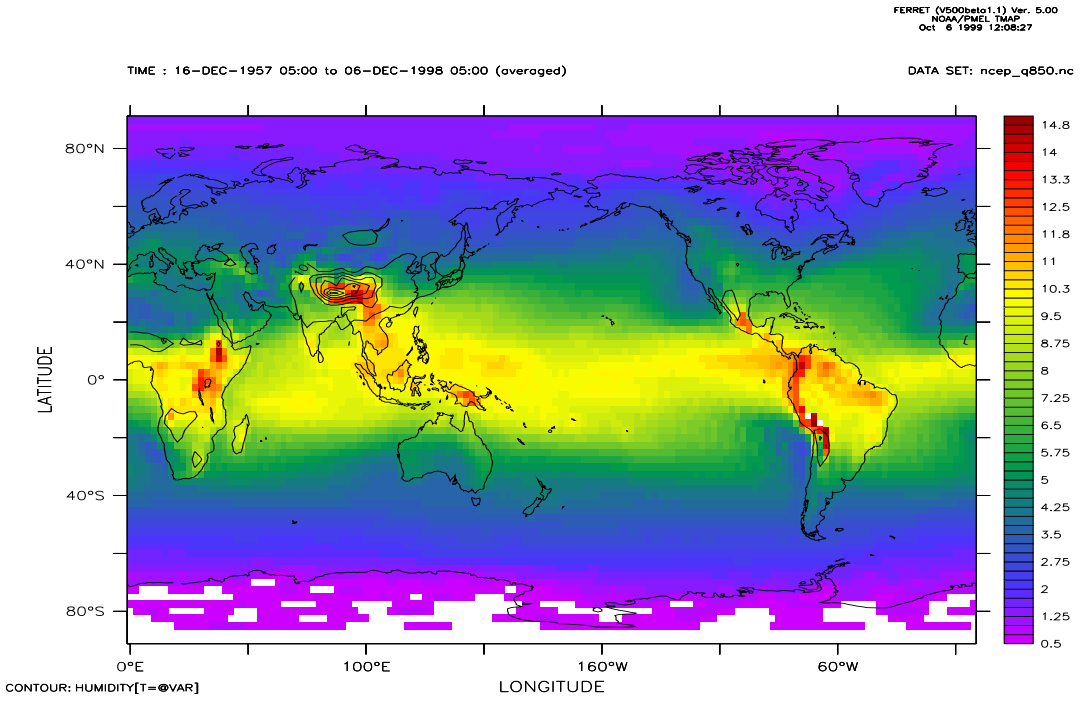


Figure 14: Map showing mean 850hPa (upper) and 500hPa (lower) specific humidity from the netCDF data file. Contours indicate the spatial patterns of variance.

6 Appendix