Merging of the observation sources

Sami Saarinen, Niko Sokka

ECMWF, Reading/UK

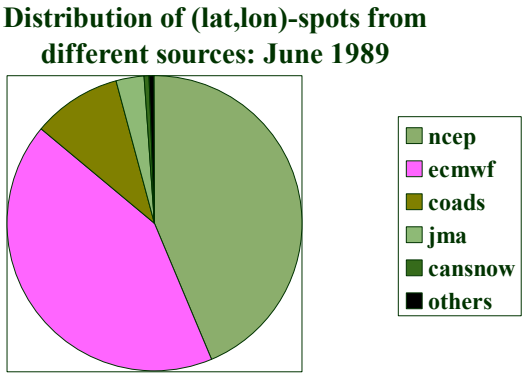
Outline

- **Conventional data sources**
- **Merging process into the PREODB**
  - **Also a short overview of the ODB software**
- **Data extraction for re-analysis**
  - **Duplicate checking procedure**
- **Case studies**
  - **Above 50hPa level winds in June 1966, -89 and -96, Tropics**
  - **January 1989 snow data from different sources**
  - **Comparison of ECMWF & NCEP/NCAR TEMP data**
  - **Peculiar SYNOP disappearance in December 1972**

Conventional data sources

- Data available from over 50 sources
- Often they complement each other, but also lots of duplicates are present
- Main sources from
  - ECMWF ("MARS"-data from late 1979 onwards)
  - NCEP/NCAR data mining exercises
  - JMA, COADS, GATE, BOM, TOGA, Antarctic,...
  - Canadian snow, USSR former snow

Conventional data sources (cont'd)



Merging process into PREODB

- Data arrives in a various BUFR formats
  - converted to ECMWF compliant BUFR
- Basic date & time cross-checks carried out
- BUFR multi-subsets get dismantled
- A simulated analysis preprocessing (MAKECMA) performed on a SGI-server
- Valid data is retained and appended to the "helper" database PREODB

Merging process (cont'd)

- PREODB-database is split into 6 hourly time windows and contains :
  - originating data source name
  - position information
    - lat&lon, date&time, obstype, ...
  - measurement information
    - pressure level, quantity, obsvalue, ...
  - the originated (full) BUFR-message

A few words about ODB ...

- ODB stands for Observational Data Base
- A database-software developed at ECMWF
  - Flexible data layout definition and fast data extraction using ODB/SQL data query language
  - Fortran90 user interface for data access
- Enables potentially more satellite obs.
- Operational at ECMWF since *27/6/2000*
- ERA-40 project has used it since 1999

ODB-applications at ECMWF

- **IFS 4D-Var/3D-Var/3D-Var FGAT**
  - closely coupled with ODB through its Fortran90 layer
- **ERA-40 Re-analysis**
  - **PREODB**-database for conventional *input* data
    - select *non-duplicate* BUFR-messages for ERA-40 analysis
    - sweep and scan data anomalies in advance when required
  - Split into 6 hourly data pools over period 1950-1999
  - Current size 180GB and is available ***ONLINE*** on SGI
- **Various data monitoring**
    - DA-system debugging, deriving statistics, plots

A typical analysis ODB contains

- Obs. ident (stat.id, lat, lon, st.alt, date, time, ...)
- Observed values for all obs. types, for example:
  - winds, temperatures, humidities, … per pressure
  - radiances per channel per instrument type
- Various flags: active, rejected, blacklisted, etc.
- Departures (obs - background, obs - analysis)
- Bias corrections
- Satellite specific info: zenith angle, field of view, etc.

Data extraction for analysis

- Assign the 6h time window for PREODB
- for years 1979-plus skip those SYNOPs and TEMP/PILOTs whose WMO-block number indicates different region than the (lat,lon)
- For station ids “99999” regenerate id using (lat,lon) value, otherwise IFS rejects them
- Perform duplicate check and pick BUFR messages for analysis

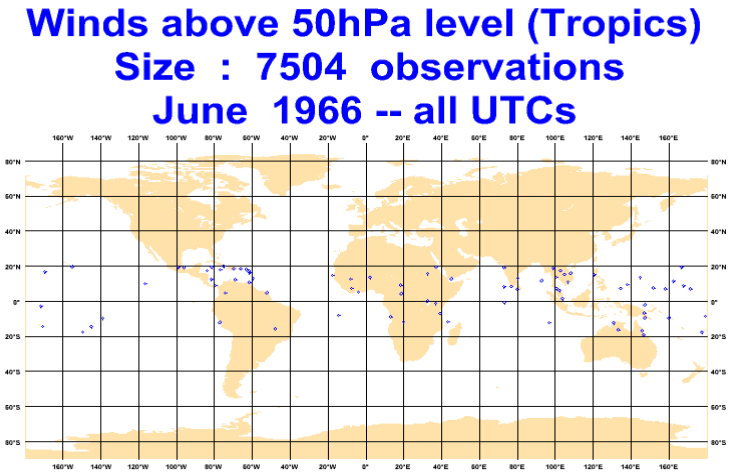
Duplicate elimination procedure

- Apply (lat,lon) tolerances of about 0,1..1 degrees depending on observation type and 1..10meters for altitude (if applicable)
  - *Still duplicate* ? Apply time tolerance (varies from 300sec for Aircraft, 1h for SYNOP, to 6h for TEMP/PILOT)
    - *Still duplicate* ? Retain those which contain most information in terms of no. of pressure levels or observed analysis variables (T,q,Ps,u,v,…)
      - *Still duplicate* ? Retain those with the highest precedence and most recently added to PREODB

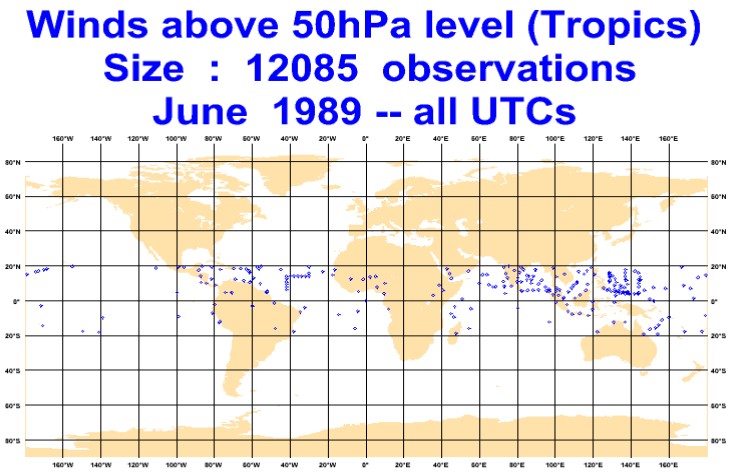
Case studies

- Above 50hPa level winds in Tropics
  - **Evolution of coverage in June of 1966, -89 and -96**
- Coverage of January 1989 snow data
  - **Canadian snow, USSR snow & the rest**
- Comparison of ECMWF & NCEP/NCAR TEMPs
  - **TEMP temperature levels between 1989-93**
- Near disappearance of SYNOPs in Dec'1972
  - **Almost empty AFSF-dataset**

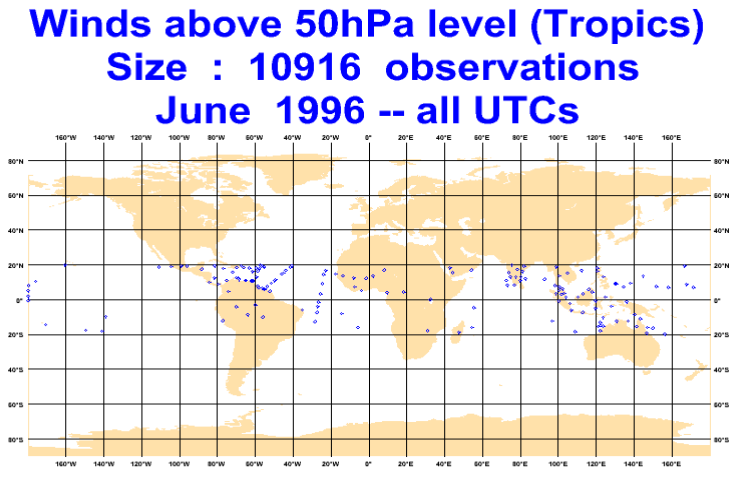
p<50hPa level winds in Tropics'66



p<50hPa level winds in Tropics'89



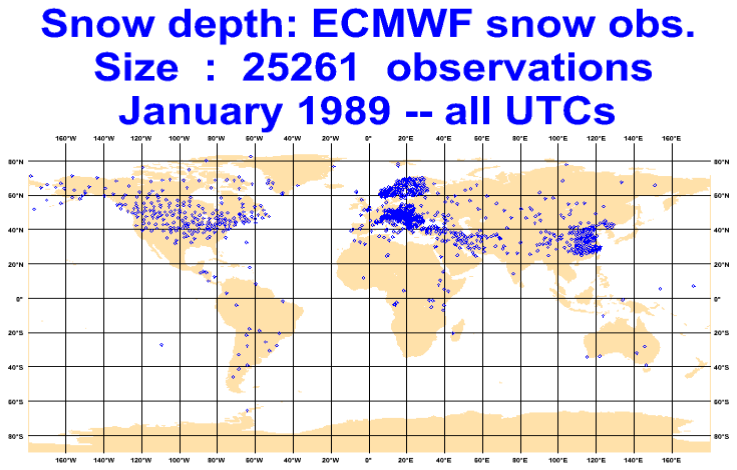
p<50hPa level winds in Tropics'96



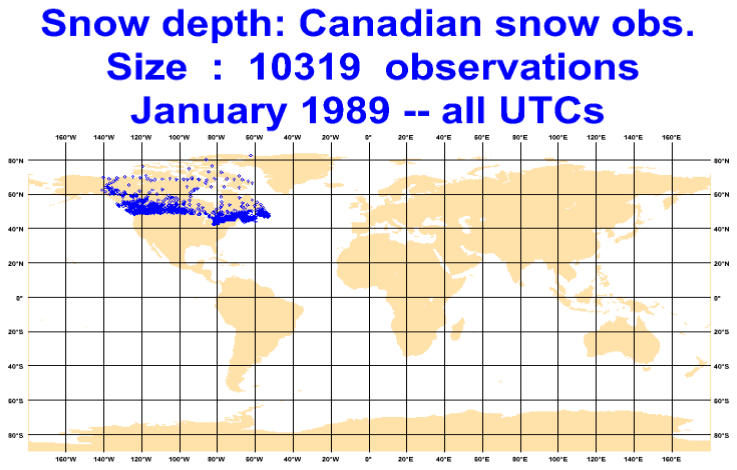
Snow coverage in January 1989

- Measured snow depths at distinct space/time locations which are greater than zero are shown
- Much better coverage thanks to supplied Canadian and Former USSR snow data

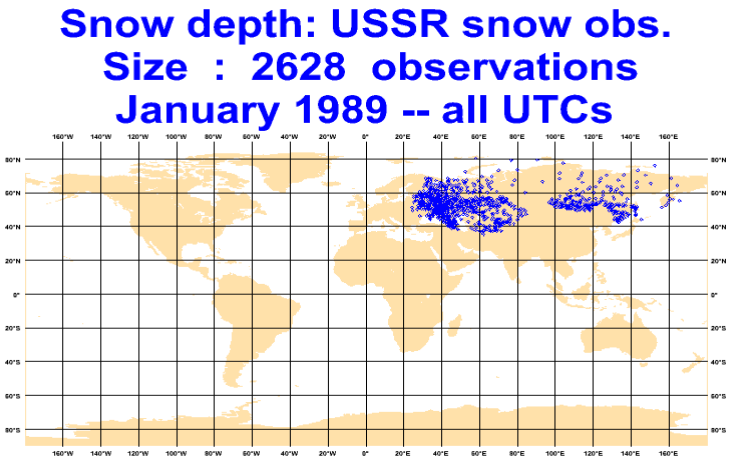
ECMWF snow depth > 0



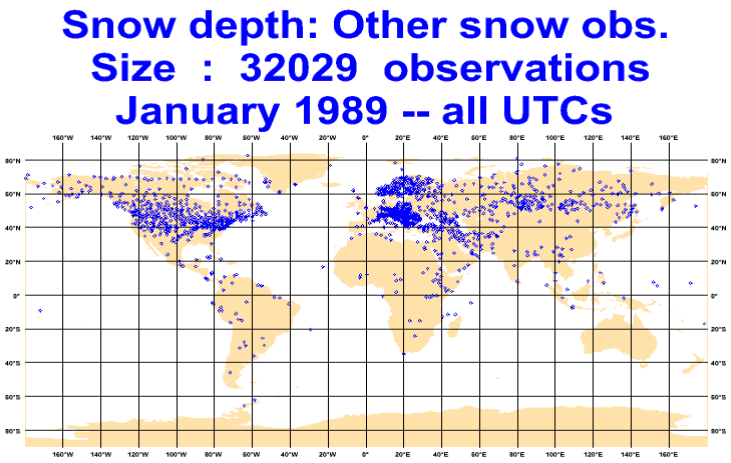
Canadian snow depth > 0



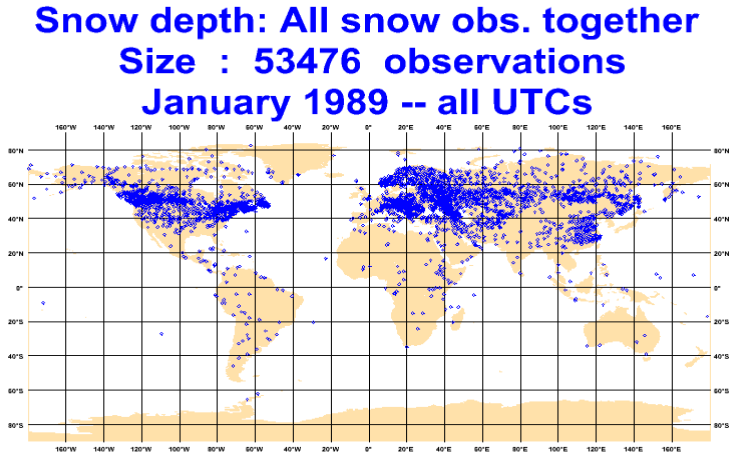
Former USSR snow depth > 0



All the other snow depths > 0



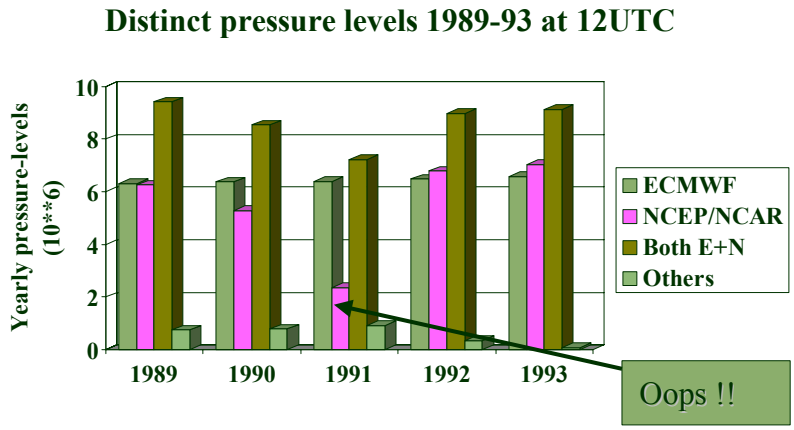
*All **combined** snow depth > 0*



Comparison of ECMWF & NCEP/NCAR

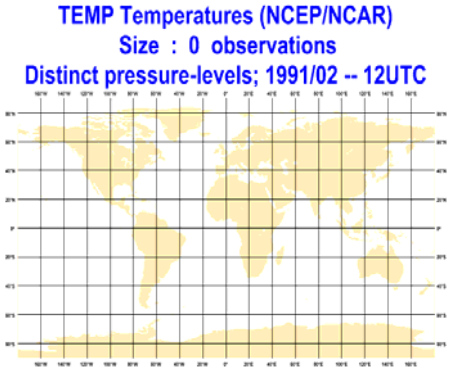
- TEMP temperatures over the years 1989-93
- Find the amount of TEMP temperatures at distinct space/time-locations (12UTC) for
  - ECMWF
  - NCEP/NCAR
  - combined ECMWF & NCEP/NCAR
  - others, but ECMWF & NCEP/NCAR
- Why there is much less data in 1991 ?

TEMP Temperatures

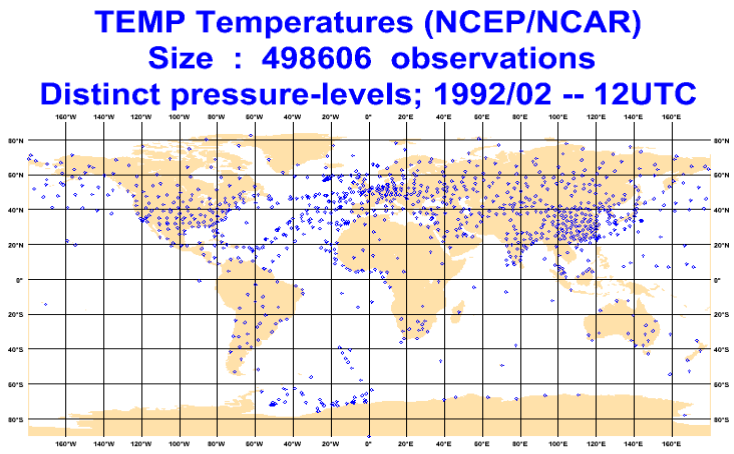


Loss of TEMP temperatures in 1991

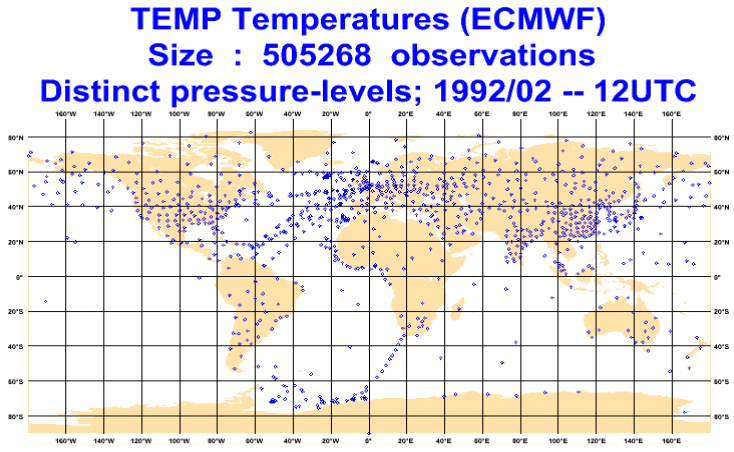
- **NCEP/NCAR first reported below normal TEMPs in Nov'1990**
- **The year 1991 seems to have been an ON/OFF situation**
- **What might be the reason for this ca. 20% loss of total data in that year ?**



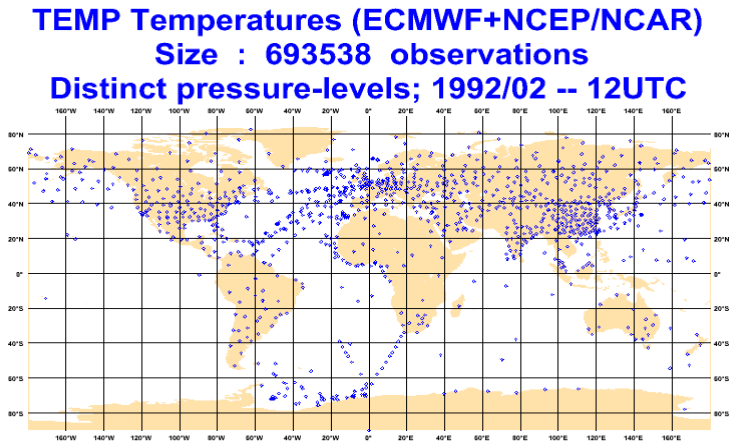
Normal month for NCEP/NCAR



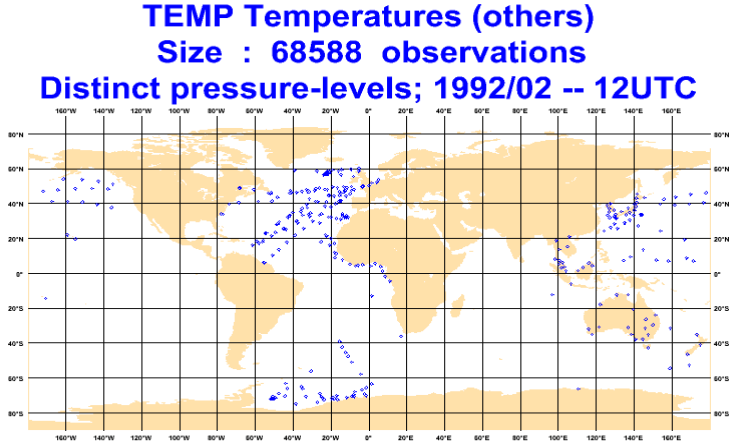
... and for ECMWF



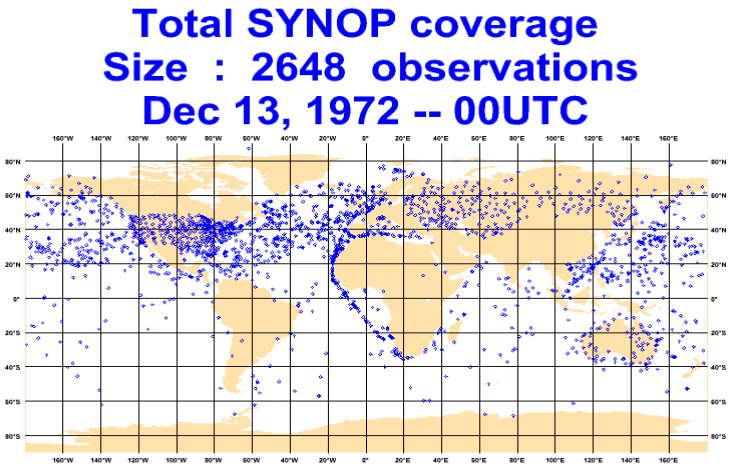
... and ECMWF + NCEP/NCAR



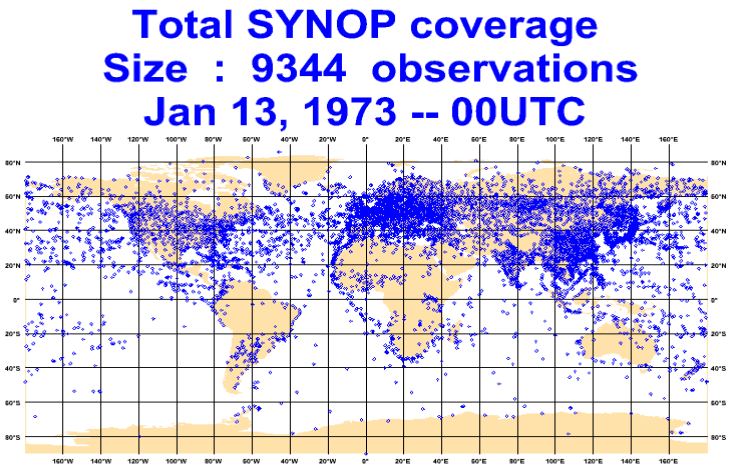
... and all other TEMP/temperatures



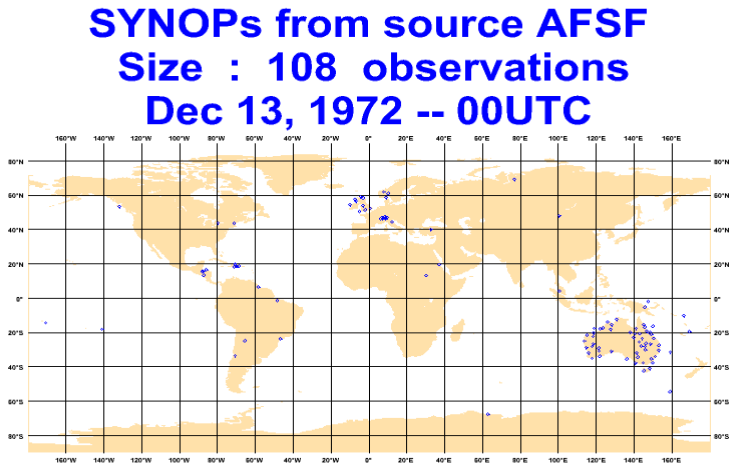
Where are SYNOPs in Dec'1972 ?



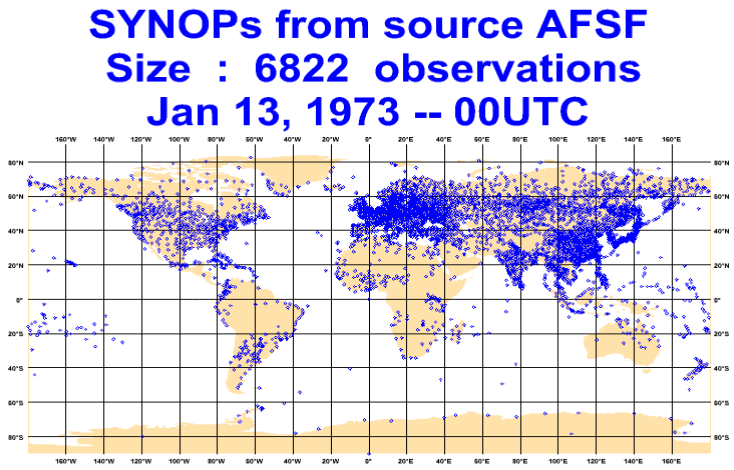
... compared to normal (Jan'73)



Is the culprit nearly missing AFSF ?



Compare: normal situation for AFSF



Conclusions

- ERA40 conventional observations are still being fed into the “helper” database PREODB
- Merging, duplicate checking and bad observation detection is a challenging task
- Despite careful checking there are still gaps and unexplained omissions in the data coverage
- Once merged, the ERA40 obs. dataset constitutes one of the best tested input datasets for subsequent re-analysis projects