TECHNICAL NOTES

OBSTAT User Guide (Draft version)

M.L. DAHOUI[1]

Operations Department

[1]Meteorological Operations Section

May 2011

European Centre for Medium-Range Weather Forecasts
Europäisches Zentrum für mittelfristige Wettervorhersage
Centre européen pour les prévisions météorologiques à moyen terme

Contents

1 Introduction

The OBSTAT program computes and plots statistics on quantities recorded at observation points in the ODB (Observations Data Base) and feedback files of the ECMWF data assimilation suites. Being systematically enabled in all IFS (Integrated Forecasting System) related experiments, OBSTAT it is currently considered as a key component of the ECMWF data diagnostic system. Since its first implementation at ECMWF, OBSTAT was mainly used to produce and plot area and time averaged statistics. Recently, OBSTAT has been significantly upgraded to allow more flexibility in the production and plotting of statistics from many perspectives: temporal, geographical, vertical column, land and sea, data usage flags, observation angles, etc
This user guide aims to provide a comprehensive technical description of OBSTAT. Examples are provided when relevant. The document is devided into 5 main sections covering the following items:

1. Overview of OBSTAT
2. OBSTAT code organization and data structure
3. Statistics production
4. Statistics plotting
5. Plots interpretation guide

2 Overview of OBSTAT

Observation based diagnostics (e.g. departures statistics) plays a role upstream and downstream of the data assimilation process. Its use can be broadly grouped into three areas:

- Characterisation of observations and their observation operators, including observation biases (Geer et al., 2010), performance of quality control (Krzeminski et al., 2009), bias correction (Auligne and McNally, 2007), and the specification of observation errors (Desroziers et al., 2005; Bormann and Bauer, 2010; Bormann et al., 2010), etc. ECMWF frequently provides important input to the calibration/validation of new satellite missions (Lu et al., 2010).
- Monitoring of the temporal stability of the observations to ensure that only consistently good quality data is used in operations. Data anomalies are detected automatically, and these may trigger corrective action. The routine monitoring statistics are of great interest also externally for data providers and other NWP centres.
- Highlighting of model problems which will be apparent, for instance, through consistent systematic non-zero FG departures for several observations (Healy, 2008). In the future, this will be extended to comparisons between all assimilated observations and forecasts at different ranges.

The main purpose of observation monitoring tools is to produce statistical products that help diagnosing the data assimilation from various perspectives. This includes:

- Production of statistics for various data assimilation quantities (e.g. FG and analysis departures, bias correction, ...)
- Production of statistics for various data selection criteria (e.g. Used, Clear, All, Active, ...)
- Production of statistics according to Land-Sea mask

- Production of statistics according to the Field Of View or observation angles (e.g. for SMOS)
- Production of statistics according to the vertical (e.g. channels, pressure levels, pressure layers, impact heights, ...)
- Production of statistics for different geographical areas and over targeted sites
- Production of time averaged statistics
- Production of time series statistics

To help evaluating and understanding the impact of the forecasting system technical changes, it's mandatory for a monitoring tool to:

- Produce generic products
- Allow an easy comparison of statistics between model versions
- Allow the customization of products for special investigations

OBSTAT was designed and several times upgraded to satisfy the monitoring requirements mentioned above. It has shown a high level of robustness within the framework of the ECMWF extensively evolving forecasting systems. The main characteristics of OBSTAT are

- Fully modular fortran code supported by Korn shell based scripts.
- Almost observations definitions are provided via an ASCII based configuration files. This is major feature allowing an easy maintenance of the software.
- Support of the existing Feedback observations files: ODB and BUFR. The ODB software has evolved significantly during the last IFS cycles. The OBSTAT code has been made backward compatible especially regarding the ODB columns names. Although compatible with BUFR feedback files, OBSTAT is not being used in the BUFR mode. Therefore, this facility is not properly maintained.
- The statistics calculation is done separately for each analysis cycle. The merging and plotting modules are used to gather and plot statistics.
- The statistics calculation is driven by a statistics definition file (called hereafter STATDEF file) in ASCII format. The STATDEF file is based on a language specific to OBSTAT and it contains a user defined number of Stat Blocks. Each Stat block is generally used to produce statistics for a specific dataset.
- The statistics plotting is driven, depending on the plot type, by the STATDEF file or a plotting option files produced by OBSTAT but customizable by users.
- Fully integrated in prepIFS and automatically enabled for all IFS experiments.
- The plotting programs are based on Magics++.
- OBSTAT produces statistics in ASCII and GRIB formats. Both formats are self-contained and can be used by other applications (e.g. Metview 4)
- Availability of a maintained OBSTAT off-line version allowing users to merge and plot statistics previously produced in the context of prepIFS. The off-line version is also able to produce, merge and plot statistics from Feedback observations files (ODB, BUFR, ...). This facility called the "slow mode" is generally used to produce customized products.

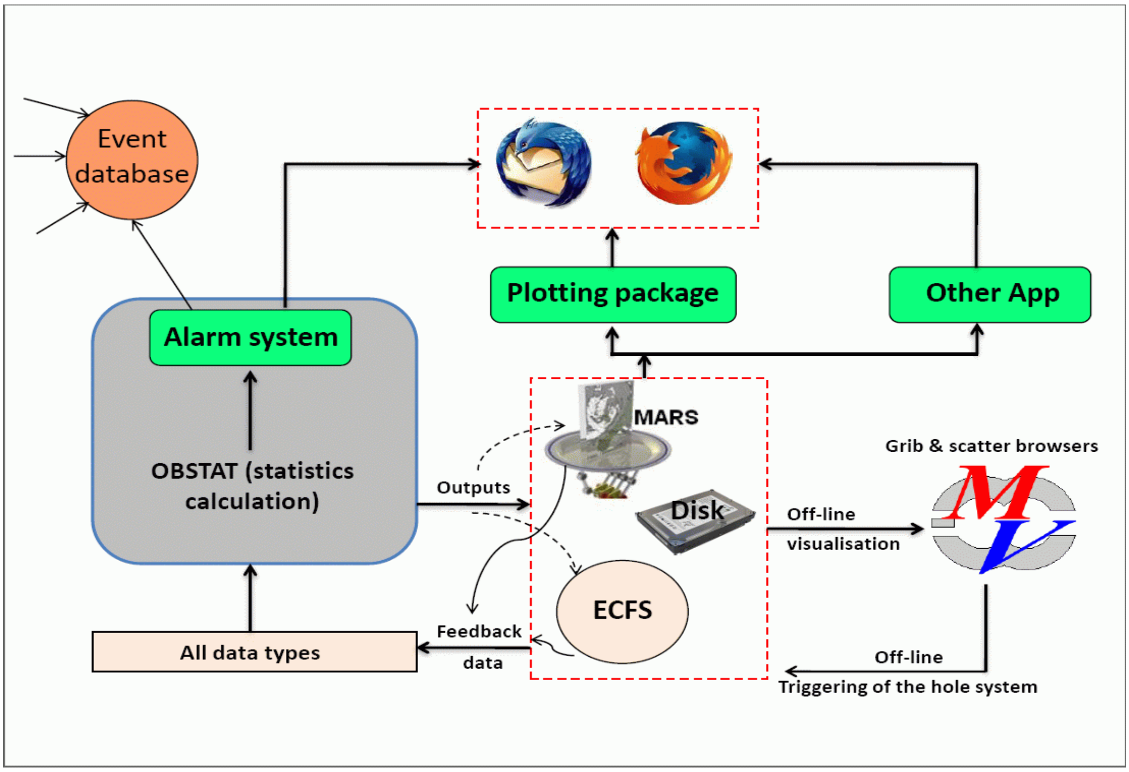

Figure 1: `OBSTAT` *implementation at ECMWF*

`OBSTAT` was initially used by the ECMWF Research Department to evaluate the impact of scientific changes on the analysis and model fit to observations. It was, to a less extent, used by Operations mainly to check daily data counts or to compare the operational data assimilation statistics against the pre-operational suites. Recently `OBSTAT` has been upgraded to satisfy almost monitoring requests for both departments. This includes, the operational satellite data monitoring products available on the ECMWF website.
The operational implementation of `OBSTAT` at ECMWF is described in Fig. 1. `OBSTAT` retrievs Feedback files from the Data Handling System (ECFS, MARS and local disks). Based on the content of the STATDEF file, `OBSTAT` produces statistics if the requested data are available in the input Feedback file. The generated statistics are then stored in the Data Handling system (DHS). Based on the plotting configuration files (defining what to plot: data, periods and plot types) the plotting modules of `OBSTAT` retrieve statistics from the DHS. The generated products are used to feed the monitoring websites and the automatic alarm system receipies. Users have also the possibility to design tools (e.g. Metview macros) to customize plots based on pre-computed statistics. Metview 4 offers tools to manipulate, browse and plot `OBSTAT` gridded statistics encoded in GRIB-2 format. Metview 4 is also being upgraded to trigger the statistics calculation. Users will have the possibility to customize the input STATDEF file.

3 OBSTAT code organization and data structure

The `OBSTAT` package architecture is presented in Fig. 2. The package is composed of a scripting part, ASCII configuration files, ODB sql requests, fortran library (generated from a fortran code) and a set of plotting and data manipulation binary tools.

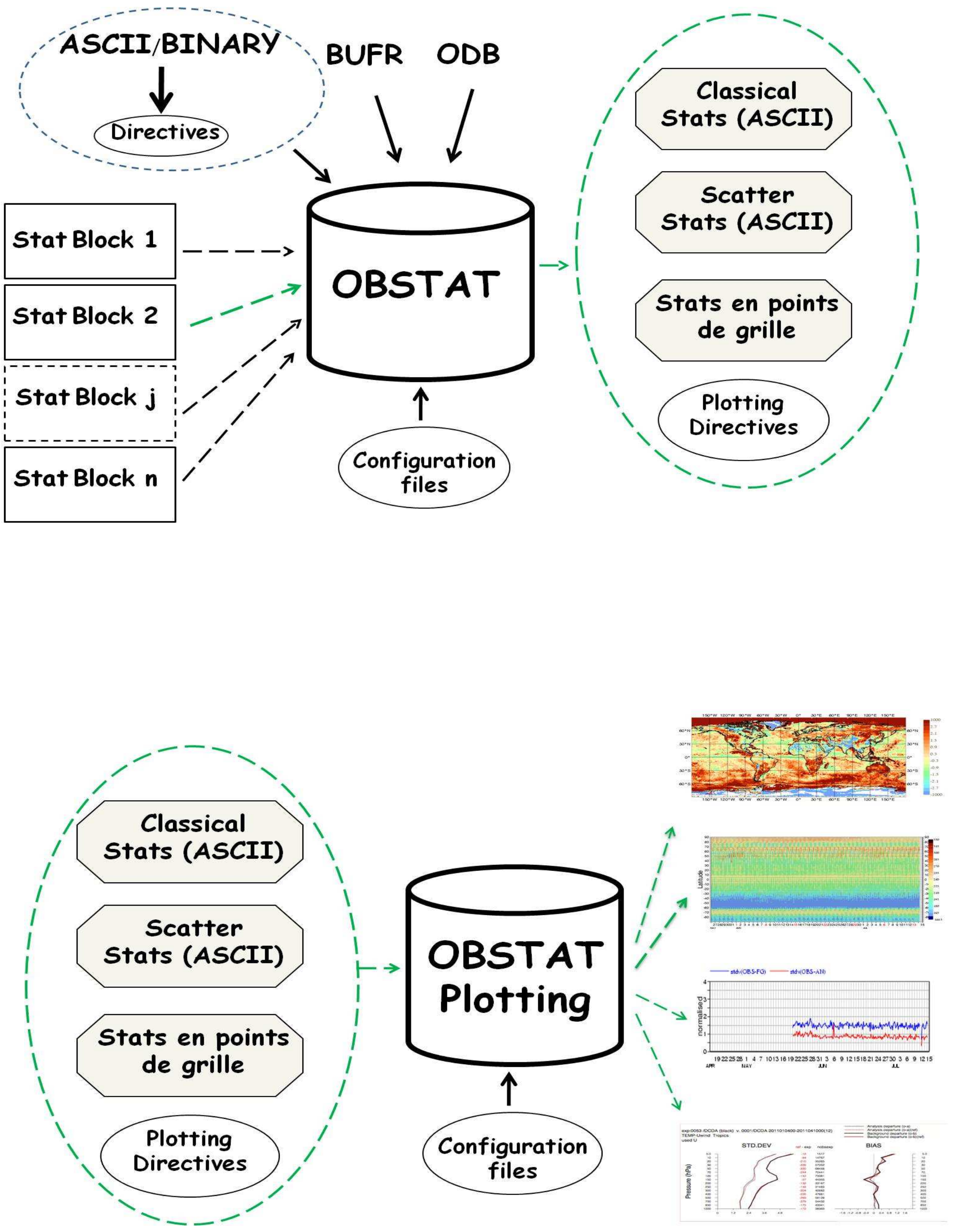

Figure 2: `OBSTAT` *architecture*

3.1 OBSTAT scripts

The OBSTAT scripting part is mainly composed of the obstat script which is used in the different configuration (except for the gridded statistics plotting) and in all platforms. The configuration settings are provided to obstat by the script dobstat for the standalone version and by the scripts pobstat.sms and obstat_init for the prepIFS version.
There are two main configurations of OBSTAT: The standalone version and the prepIFS version. The prepIFS version (see Fig. 3) is designed to run on-the-fly while the IFS experiment is progressing using the online ODB files. Statistics are produced and archived on ECFS for a later use (merge and plotting) by the standalone version. This version produces and publishes, on the intranet website, a set of plots (including time series of statistics). The standalone version (see Fig. 4) is designed to produce and plot statistics in off-line mode given that the IFS experiment to monitor had run and the associated feedback information saved in ODB/BUFR. If statistics have been already calculated within prepIFS and the user is satisfied with the default statistics specifications (default STATDEF file), the standalone version can then be used to perform a post mortem diagnosis of the assimilation suites by extracting and agglomerating the pre-computed statistics over long periods (given that statistics have been already produced for these periods). The standalone version allows also the extraction of the analysis log files for the plotting of the VARBC predictors.

OBSTAT works in three main steps: Statistics calculation, statistics merging and statistics plotting.
The statistics calculation (STACLAC=true) performs the following tasks:

- Checks the obstat parameters. Defaults are used if nothing is specified by the user.
- Retrieval of ODB/BUFR files. In case of prepIFS version, the online ODB files are used.
- Decodes and produces statistics based on the specified STATDEF file. An ascii file ($EXPVERyyyymmddhh$STREAMobstat.txt) is produced for each analysis cycle. If requested (in the STATDEF file) gridded statistics, in GRIB2 format, are also produced.
- Saves statistcs on ECFS for the prepIFS version.

It is worth mentioning that with the standalone OBSTAT, the reading of ODB/BUFR is time consuming and this depends of the ODB type and so it is best performed in a batch job.

The statistics merging task (STAMERGE=true) reads all present OBSTAT output ascii statistics files and merge them into one single ascii file called stat$EXPVER.txt which will be used by the plotting task. The merging task is systematically performed prior the the plotting if more than one cycle is requested.
Prior to the implementation of gridded statistics support in OBSTAT, the statistics plotting task was limited to the production of statistics present in the OBSTAT ascii file. The statistics plotting task (STAPLOT=true) read the OBSTAT ascii file (produced by the statistics merging task) to produce a multi-page postscript file. Each page contains up to three panels. Each panel shows statistics associated to one stat block (after areas, flags, instruments and streams expansion) of the STATDEF file. At a second stage, the plotting task has been upgraded to produce timeseries, 2D maps and scatter plots to look like the operational satellite data monitoring plots. This intermediate upgraded version of the plotting was based on the use of the OBSTAT ascii file processed by several perl scripts (see Fig. 3 and Fig. 4). With the introduction of gridded statistics, a new plotting package has been developped to produce a large number of plotting products using gridded statistics in GRIB-2 and ascii scatter statistics. At the time of writing this documentation, prepIFS is still not using the new developped plotting package but work is progress to add full support to this package in prepIFS. The scripting part of the new plotting package (decsribed in Fig. 5) performs the following tasks:

- Checks parameters. Defaults are used if nothing is specified by the user

- Reads the plotting option files
- Retrieves the requested gridded statistics based on the content of the option files and the user specified date and periods.
- Produces geographical maps of statistics (if requested)
- Produces timeseries of area averages (if requested)
- Produces Hovmoeller plots (if requested)
- Produces scatter plots (if requested)
- Produces overview timeseries plots (if requested)
- Saves all produced plots in the user sepecified location

To run the standalone `OBSTAT`, the user needs to copy the script `dobstat` from the directory "/home/mo/obstat/OBSTAT". The user should edit the `dobstat` script and change the variables described in the text below. The most useful ones are near the beginning — the defaults for the others are usually right for most users. Parameters `FIRSTCYCLE` and `LASTCYCLE` specify the assimilation dates and times to loop over. `obstat` has a default setting from the time increment between two analysis times, in hours, called TSTEP. It is 12 for ODB files and 6 for BUFR feedback and ERA-40 ODB. The user can specify another value for TSTEP, this will overrule the default setting.

`EXPVER1` defines the assimilation experiment identifier, e.g. 0001 for oper, 0053 for e-suite, eixd for rd and ER40 for ERA-40. `CLASS1` defines the class; rd, od, e4 and ei are supported. `STREAM1` defines the stream: DA, DCDA and SCDA are supported.

It is possible to calculate or extract statistics for two experiments and overlay them on the plots - and excellent way to compare two experiments. This is achieved by activating the following three lines in the script:

```
#EXPVER2=0053
#CLASS2=od
#STREAM2=DCDA
```

It is possible to compare an experiment with any type of other experiment (od, rd, e4,ei). But comparison can of course only be done for the observing systems the two experiments have in common.

`OBSTAT` is able to produce a range of observation diagnostic related plots. This is selected by the user from the following list of applications:

```
OBSTATPLOT=yes       # yes: make obstat plot
TIMESERIES=no        # yes: make time series plot
VARBC_PARAMS=no      # yes: make varbc time series plot
GEOWRITE=no          # yes: produce 2D geo plots
HOVWRITE=no          # yes: produce hovmoeller plots
SCATTERWRITE=no      # yes: produce scatter plots
CONDITION_NUMBER=no # yes: produce condition number plot
```

The setting `OBSTATPLOT=yes` produce classical obstat, typically vertical profile plots and histograms of departures for a range of observing systems. The user can produce more specialized plots, as will be described further in the subsequent sections.

The other options above relate to time series plots and 2D-plots and will documented later. With the implementation of gridded statistics in OBSTAT, it is advised to keep these variables disabled (=no) and relay on gridded statistics (in GRIB2 format) to produce timeseries plots and 2D-plots.

OBSTAT is used in operations, for e-suites and in research prepIFS experiments to calculate departure related statistics for predefined regions and for a specified range of observing systems, typically the ones used in the operational cycle and some new observing systems that are being monitored passively. The results of these obstat calculations are stored in ascii text files on ECFS. This makes it possible to produce obstat plots quickly from these ascii files without the need to calculate statistics from ODB/BUFR feedback files. This makes the comparison of order 1000 times faster, but also restrict the statistics to the predefined areas and instruments. It is controlled by the variable:

```
USE_ECFSLOG=yes      # yes: the fast mode obstat plot where the obstat statistics
                     #   files are retrieved from ECFS
```

When USE_ECFSLOG=no ODB-files or BUFR feedback files are retrieved from ECFS or MARS, respectively. The feedback information related to observations are extracted and statistical calculations of RMS and bias are calculated. To use BUFR feedback we set OBSFORMAT=BUFR, for ODB we set OBSFORMAT=ODB. For ODB is is also necessary to specify the ODB file type: CCMA, ECMA, ODBCMP, MONDB, OFB or MFB (OFB and MFB are the new ODB formats stored in MARS). The datalist described below can be used to reduce the number of observing system to consider.

COMMENT and LAYOUT are only used for the plotting. FBKCLEAN should be kept to ‘yes’ unless you want to cache the ODB/feedback files on STOREDIR (typically $SCRATCH or /bigtmp) for later use; these files are big, so this should be avoided.

The definition of statistics (to be produced) can be changed by the user by editing and updating the STATDEF file (see).

The full list of dobstat script parameters is described below. Not all of these parameters are pre-defined in the script dobstat. However the scripts obstat contains defaults for all these parameters. The user is invited, to set these paramters in dobstat if the default is not what she/he needs.

EXPVER1/EXPVER2 the 4-character expver string that defines the assimilation, e.g. abcd for a prepIFS experiment, 0001 for the operational suite, or 0011 for an e-suite.

CLASS1/CLASS2 must be rd for a prepIFS experiment or od for an operational suite. e4 for ERA-40 and ei for ERA-interim. This parameter is used to define the MARS request for the feedback files to retrieve and to identify where in ECFS to fetch the ODB-files.

FIRSTCYCLE the initial yyyymmddhh date of the feedback file to process. The time must be 00, 06, 12 or 18. It is not possible to filter observations inside a file according to their time with this parameter. The software is year-2000-compliant.

LASTCYCLE the final date to process. Same remarks as above. Set it equal to FIRSTCYCLE to process one single analysis.

TSTEP the time increment, in hours (e.g. 12 or 06), used to loop from FIRSTCYCLE to LASTCYCLE. Feedback files are currently produced every 6 hours regardless of the 4D-Var cycling period.

COMMENT a character string to display on each panel of the statistics plot. It should be a short mnemonic, ideally less than 10 characters. Note that EXPVER1 and the dates are automatically displayed anyway. This string shall be used to document the reference dataset on the plots when superimposing statistics from 2 experiments (see OVERFILE below).

LAYOUT a string equal to A3 or A4 to specify the size of the Postscript plotting page. The default operational plots are A3. A third option, EPS, is an attempt to improve the output for publications, with one panel per page, no frame and a mimimum of printed material.

OBSTAT_LIB is the Unix path of the relocatable library of program `obstat`. It is always relinked with the EMOS and MAGICS libraries on execution. A reasonable default is supplied with the script. The program works with single precision reals.

EXEDIR is the path of the directory where temporary files are created and the output data files are stored. The default is `$SCRATCH/tmp$EXPVER`. In order to run several instances of `dobstat` in parallel, you must ensure that these directories are all different. The value `$SCRATCH/tmp$EXPVER` allows the simultaneous processing of several experiments, but not several dates of the same experiment, nor starting a plot before the feedback decoding is complete, unless `EXEDIR` is suitably redefined.

DATADIR is the path of a directory that contains some ascii configuration files used by the program (see 3.3). These files can be customized as long as their (obvious) layout is adhered to.

STATDEF is the path of the *statistics definition* file. This ascii file, described below, describes all the statistics that are to be computed and/or plotted by `obstat`. Following a simple syntax described below, one can redefine the computations in terms of type of statistics, observation parameter, type, intrument, geographical domain, etc. The default file is the one used for the operational monitoring.

GENERALFILE Usually users need to update or customize the content of the the file `general.cfg` containing the definition of standard pressure levels, areas, colours and curves attributes. With GENERALFILE variable, users have the possibility to use their own `general.cfg` without having to have their own version of all the configuration files located in DATADIR.

STOREDIR is the path of the directory where the ODB files/BUFR feedback files are kept on disk. Up to 4000Mb are required for AIRS/IASI ODB, other observing system data files require much less 1000Mb per analysis, and no more is necessary `FBKCLEAN=yes` (see below).

OBSFORMAT specifies the obs input files. Three choices are possible: ODB, BUFR and USER. ODB is the format used at ECMWF. BUFR support is not any more maintained but it is most likely that `OBSTAT` will continue to work for this option if a valid EMOS_CYCLE is specified. However, it is important to mention that BUFR feedback files will cease to be produced and archived beyond the IFS cycle 37R3. The option USER can be used to feed `OBSTAT` from a user defined ascii/binary files. To use this facility, the user has to indicate to `OBSTAT` how to access the file and how the data is organized inside the file. This is to be done by editing and changing the file `data_interface` that can be found in "/home/mo/obstat/OBSTAT/data"

DATA_INTERACE Defines the location of the `data_interface` file required when OBSFORMAT=USER.

USERDATA This variable is mandatory when OBSFORMAT=USER. It is indicating to `OBSTAT` where to access input obs files.

datalist is a list of ODB/BUFR feedback filenames that will be processed for each analysis. Currently, to get all standard active observations from BUFR feedback or ODB the user must ask for:

```
 #for BUFR feedback:
datalist="fbconv fbscat fbssmi fbrad1c fbgeos fbtovs fbreo3 fbairs"
 #for ODB (ECMA and ODBCMP) files:
datalist="conv satob scatt ssmi msu ssu amsua amsub hirs geos reo3 airs iasi ssmis
mhs tmi gpsro meris"
 #for ODB (CCMA) files:
```

```
datalist="oneodb grid"
 #for ODB (MONDB) files:
datalist="oneodb"
 #for ODB (OFB and OMB) files:
datalist="conv satob scatt amsua amsub hirs geos reo3 airs iasi
  gpsro meris mhs ssmi_as windsat_as ssmis_as tmi_as amsre_as mwts mwhs"
```

This changes in time. Less files may be available on old dates or in periods of data coverage problems. Asking for a subset of these files will speed up the processing substantially. Unfortunately there is (yet) no safe way to tell if an ECFS retrieval fails because of the feedback file is really missing, or because there is a technical problem, so if large differences are found in the observation numbers between two suites, it is worth checking the listings for such problems.

FBKCLEAN (`yes` [recommended] or `no`) Switch to save disk space. For each requested analysis, the files specified in `datalist` are retrieved one at a time from MARS or ECFS onto `STOREDIR` unless they are already present, one by one they are processed, and then deleted from `STOREDIR` if `FBKCLEAN=yes`. If the user is performing different kind of statistics on the same ODB datafile(s) and has got sufficient amount of disk space it is possible to set FBKCLEAN=no and set RETRIEVEODB=no. Then the ODB files are kept on the disk in un-tarred form ready for use.

WRKCLEAN (`true` [recommended] or `false`) Switch to clean the working directory before starting the statistics calculation. Setting WRKCLEAN to false is particularly usefull in case a problem occurs while `OBSTAT` is running. By not cleaning the working directory, the user can continue (after understanding to cause of the problem) the statistics calculation without having to start from the beginning of the period.

STACLEAN (`yes` or `no` [recommended]) Switch to save even more disk space by deleting the intermediate ascii statistics files that were generated, after they have been merged over the requested assimilation period. These files are stored in `EXEDIR` under the name `statyyyymmddhh` where `yyyymmddhh` is the date of the corresponding analysis. They can be used to remake statistics over subsets of the assimilation period, or to see the evolution of statistics over time; they take about 300Kb per analysis. Their internal formatting is self-explanatory, they follow the same ascii template as the initial `STATDEF` file, plus the actual values of the statistics.

STACALC, STAMERGE, STAPLOT (`yes` or `no` each) Switches to activate any of the three successive step of scripts `obstat`. `STACALC` fetches the BUFR feedback files, decodes them and produces one ascii statistics file `statyyyymmddhh` per analysis ; it takes about 15mn of real time per analysis, so it is best performed in a batch job. `STAMERGE` rereads all the ascii statistics files and merges them into one single ascii file called `stat$EXPVER.txt` ; this only takes a few seconds. `STAPLOT` reads the latter file and plots it into a Postscript file before printing and previewing. The merging and plotting steps are usually run in interactive mode, although in principle it is possible to do all three steps in a single job.

GEOGRIB_AVE_PERIOD (`yes` [recommended] or `no`) Switches to produce weekly/monthly normalized high resolution gridded statistics. Otherwise, high resolution gridded statistics will be kept for each analysis cycle (usefull for special investigation over a short period).

SCAT_PERIOD (`yes` [recommended] or `no`) Switches to produce weekly/monthly scatter (bined populations) statistics. Otherwise, scatter statistics will be kept for each analysis cycle (usefull for special investigation over a short period).

LALLDATA (`yes` or `no` [recommended]) Switches the production of statistics associated to all observations present in the ODB/BUFR file. For data types not specified in the `STATDEF` file, generic statistics will be produced. The user has the possibility to define the content of ”generic” statistics. Otherwise, a default definition of ”generic” will be used.

PARAMS (optional) is an ascii file containing up to 5 user defined diagnostics (outside the list pre-defined in OBSTAT). The file has to comply with a specific format (see 4.3).

FLAGS (optional) is an ascii file containing up to 5 user defined data selection criteria (outside the list pre-defined in OBSTAT). The file has to comply with a specific format (see 4.4).

GRIBDIR (commented by default) is the location where to store output gridded statistics. The location can be on ECFS or on local discs. By default GRIBDIR="ec:/$USER/backup/$STREAM/an/$EXPVER/obstat"

ODBDIR (optional) specifies the location where the retrieved ODB files can be stored on disc to allow a multi-use of the retrieved files. If ODBDIR is present, OBSTAT searches for the input ODB first in ODBDIR. If the file is not present, OBSTAT does the retrieval from ECFS and saves the file on ODBDIR. Under ODBDIR, the ODB files should have the name $ODBFILETYPE.tar.$EXPVER.yyyymmddhh. As the files can be huge, this option should be used only during tests.

ARCHIVE_ECFS (yes or no [default]) this variable applies only to gridded statistics. When GRIBDIR is pointing to a location on local discs, the user still have the possibility to save a copy on ECFS by setting ARCHIVE_ECFS=yes. In this case, the default location is used "ec:/$USER/backup/$STREAM/an/$EXPVER/obstat"

GENOPT (yes [default] or no) this variable applies only to gridded statistics. When GENOPT=yes, activated, OBSTAT produces the plotting option files associated to the generated gridded statistics. Users can use directly these option files to produce the plots. Users can also customize the option files before use. It is highly recommended to always activate this option to ensure a complete consistency between produced gridded statistics and the option files espcially in cases where the user changes the definition of statistics in STATDEF.

ALARM (yes or no [default]) OBSTAT has the possibility to produce small ascii files containing global statistics of selected diagnostics. These ascii files are currently used as input of the automatic alarm system. For each stat block (within the STATDEF file) the user can switch on the production of such ascii files. To enable the production of these files, the global variable ALARM should be equal to "yes".

SQLDIR (optional) specifies the location of ODB SQL queries to be used for the building of the OBSTAT statistics calculation binary `obstat_cacl.x`. Starting from the IFS cycle 33R3, the ODB has been subject to important changes in the context of ODB cleaning process. The changes affected, the ODB columns names and tables. Each update of ODB requires different SQL queries. The standalone version of OBSTAT is maintained to work with all ODB versions. By changing the SQLDIR, the user needs to be sure that the SQL queries present in SQLDIR are compatibles with the input ODB. Otherwise, OBSTAT will abort. Care should be taken when comparing two experiments from different cycles using different versions of ODB. OBSTAT is able to handle both version when the default is used for SQLDIR. This may not be the case with a user defined requests.

SQLSECTION (optional) specifies the name of ODB SQL view to be considered by `obstat_cacl.x` when reading the data from the ODB input file.

HSRIS_PLOT defines the plotting mode for AIRS/IASI by indicating to OBSTAT the vertical coordinate to consider for the plotting of statistics. There are five options: 1 for channels peaking pressure, 2 for channels index, 3 for channels number, 4 for channels wavenumber and 5 for channels wavelenght.

PLOTAISRBIAS (yes or no [default]) switches to include AIRS/IASI bias correction curves on AIRS/IASI data plots

PLOTBIAS (yes [default] or no) Switches to include bias correction curves on satellite data plots

STATFILE (optional) is the name of an ascii statistics file (in EXEDIR) to be plotted directly by the STAPLOT step, instead of the stat$EXPVER.txt default. Useful to switch between different sets of statistics in the same directory.

OVERFILE (optional) is the name of a 'reference' statistics file to be superimposed (by default, in red) with the existing statistics (in black). This file must have been computed by a previous execution of the STACALC step of dobstat, and its contents must have exactly the same layout as in the current statistics, usually as defined by the STATDEF file, otherwise the program will abort. If necessary, statistics can be added, deleted or moved by hand in the ascii files with a little practice... The superimposed statistics produce some extra legend, printouts, and in the plots of vertical profiles the signed difference between the two sets of populations is printed on the left of the normal numbers (mnemonic: reference **plus difference** equals **population**).

STDEV (yes or no) used to plot standard deviation vertical profiles instead of rms, which is the normal practice.

FIXEDAXES (yes [default] or no) Switches to get fixed preselected axis limits on obstat plots

WHISKERS (yes [default] or no) Switches to get st.dev box and min/max whiskers on histogram plots

IVERB integer number (default is zero) to specify the level of verbosity in the output. 1 produces a few diagnostics, 2 or more provide many details, 5 produces such a big output that it aborts execution after the next message is processed; in addition, 5 causes the contents of the latest BUFR box to be printed out to ascii file prtbox.out . 7 lists all the observation data decoded, without aborting voluntarily.

UPDHARD (.true. [default] or .false.) Switches to print summury statistics for all observations types present in the input observations file. It is highly recommended to keep the default (.true.). It is common to make mistakes during the first attemp of adding a new stat block (e.g. wrong obs codetype or instrument ID). The availability of summary statistics helps in getting the right code numbers.

UPDSOFT (.true. [default] or .false.) when UPDSOFT=.false., OBSTAT will not produce statistics but allow the production of summury statistics for all observations types present in the input observations file (when UPDHARD=.true.).

SCALING (.true. [default] or .false.) when SCALING=.false., the rescaling applied to ODB variables is disabled.

STATFILTER (.true. or .false. [default]) by default, OBSTAT plots all statistics available in the output ascii file. However, in some situations the users are interested only by a subset of statistics. In such situations, the user has the possibility to filter statistics to plot. This is achievable by setting STATFILTER to .true. and by pointing the variable STATDEF to a file containing only the definition of the desired statistics.

IMAXREAD integer number used to stop processing prematurely after that many BUFR products have been read. Used for debugging; the default is not to stop until the end of each file.

GZIPSTAT (yes or no (default)) when equal to "yes", OBSTAT will compress (using gzip) the output ascii files before saving them on ECFS.

EMOS_CYCLE the identifier of the EMOS library to use. Normally it is the same as the latest default in prepIFS .

GRIB_API_VERSION (default is current) Specifies the GRIB_API version to be used for the GRIB2 encoding of gridded statistics. It is recommended to keep the default.

Figure 3: OBSTAT *standalone version scripts*

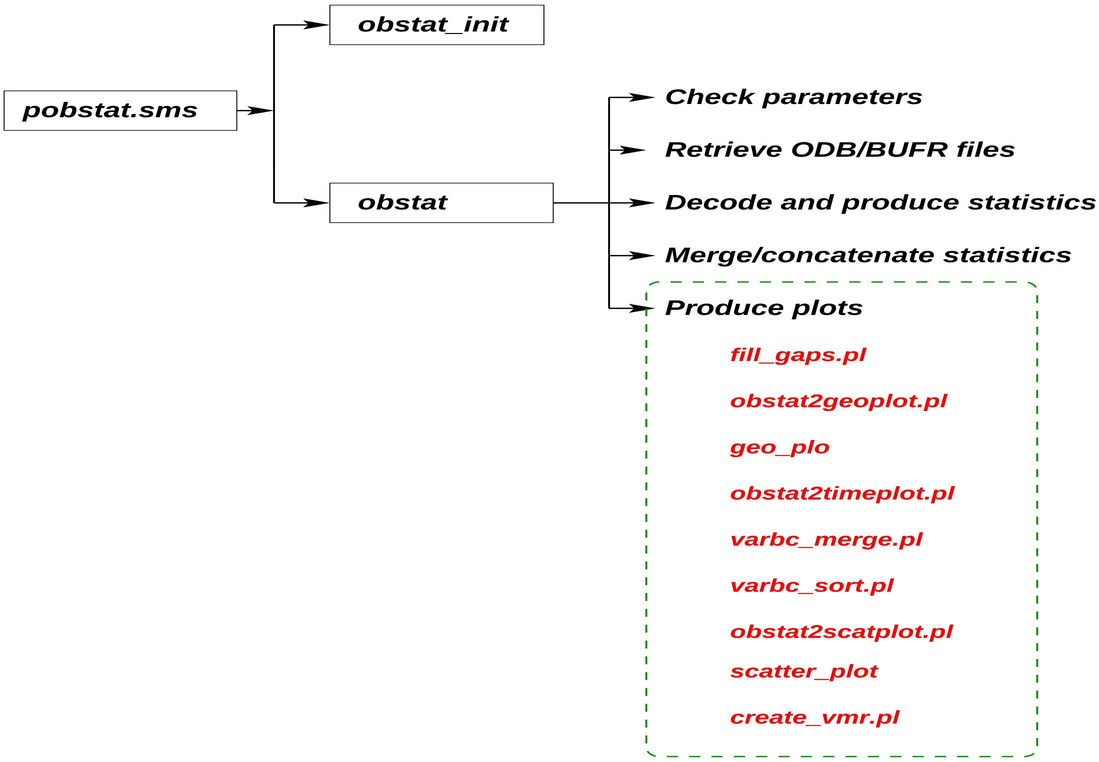

Figure 4: OBSTAT *prepIFS version scripts*

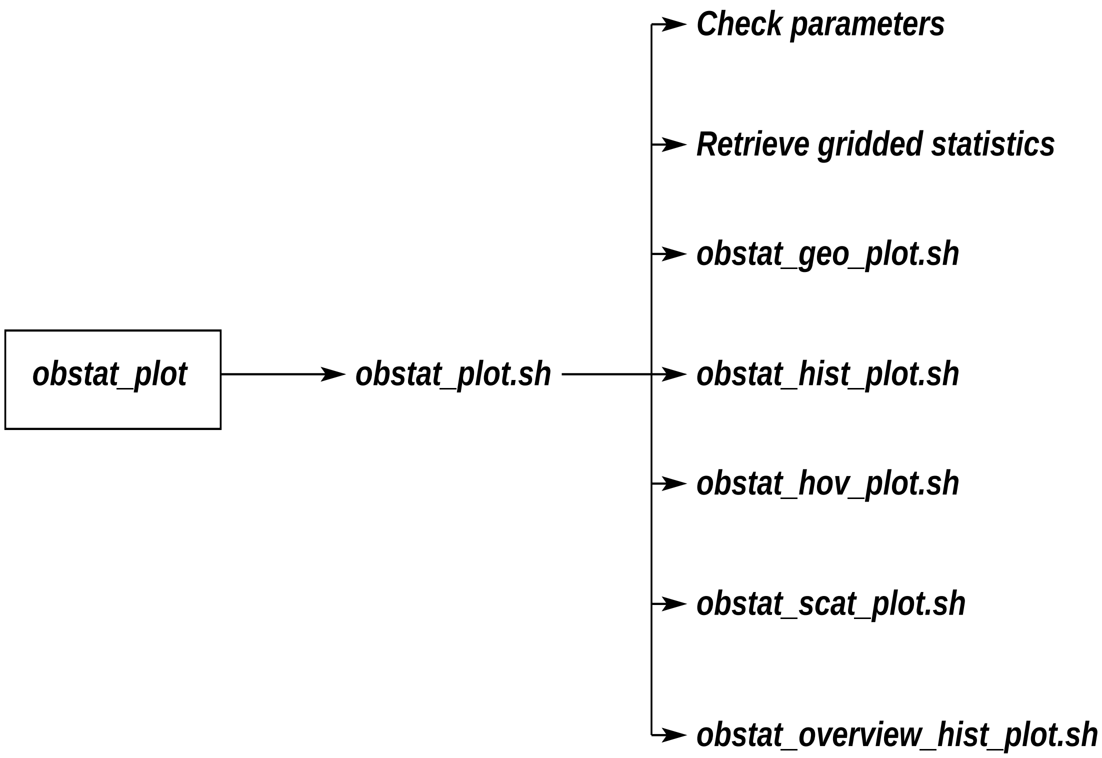

Figure 5: OBSTAT scripts for the gridded statistics plotting

3.2 OBSTAT fortran code

The architecture of the Fortran code is straightforward: initializations, loop to read, decode the ODB/BUFR file and update the statistics, output of the results as printouts, ascii file, gridded statistics, scatter statistics and plots. As little code as possible depends on the OBSFORMAT: as soon as it is extracted the relevant feedback information is stored in an array in module `obsdata` which acts as an interface between the decoding and the statistics update process. The residual dependencies to the BUFR codes are limited to cosmetic features in the plotting part.

The code has been made as independent as possible from the observation type. It means that the behaviour may not be optimal for data such as TOVS or SATOB. This is a choice, in order to keep the code simple and easy to maintain.

The information in module `obsdata` is kept in a 2-D real array, with as many lines as data that have been decoded (in practice, all the data from each BUFR product), and as many columns as feedback *items* that are expected to be used later on (missing item values are set to constant `rindef` defined in module `globvar`). This means that the structure of the original message or report s lost ; items which are the same for all data in a report, such as observation code type, are duplicated for all data in `obsdata`. The subsequent processing is supposed to act independently on each scalar datum, with the only exception of wind components (where V is supposed to follow U if wind speed statistics are needed). Some items are 'virtual' in that they are not computed until the latest minute, because they are simple functions of other feedback items (e.g. the background departure before bias correction is inferred as the sum of the bias correction and of the departure found in the ODB/BUFR file). This saves CPU time and memory.

The items defined for each datum are identified in routine `iniglob.F90`. All the items are initialised to indefinite value. An item is only activated (associated to a number as indicated below) if the associated information can be accessible from the input OBS files (ODB/BUFR/USERDATA). This has two main benifits:

- Avoids the static mapping between the ODB SQL requests and the order of items. The static mapping causes a lot of failure if a change in the order of the SQL request is done independently of the `OBSTAT code` (in iniitemloc.F90 routine).
- Prevents failures caused by a removal of an ODB word from the ODB SQL request.

The decisions of enabling the items are performed at the level of iniitemloc.F90 routine. Items are enabled automatically for BUFR and dynamically for ODB/USER obs formats.
At the time of writing the items defined for each datum are the following: (refer to the comments in `obsdata` for up-to-date information)

```
 name                index    description

irtyp_item          =1    ; cobit(1)='report type'
idtyp_item          =2    ; cobit(2)='datum parameter code'
irins_item          =3    ; cobit(3)='report instrument/sensor type'
irlat_item          =4    ; cobit(4)='report latitude'
irlon_item          =5    ; cobit(5)='report longitude'
idpre_item          =6    ; cobit(6)='datum vertical coordinate'
irtim_item          =7    ; cobit(7)='report time (hours)'
idval_item          =8    ; cobit(8)='datum observed value'
ifgdp_item          =9    ; cobit(9)='background departure o-b'
iandp_item          =10   ; cobit(10)='analysis departure o-a'
iober_item          =11   ; cobit(11)='observation standard error'
ifger_item          =12   ; cobit(12)='background standard error'
il1dp_item          =13   ; cobit(13)='lo-res departure update 1'
ih1dp_item          =14   ; cobit(14)='hi-res departure update 1'
il2dp_item          =15   ; cobit(15)='lo-res departure update 2'
ipbgr_item          =16   ; cobit(16)='probability of gross error'
ipval_item          =17   ; cobit(17)='range of possible values'
ibcor_item          =18   ; cobit(18)='bias correction of obs'
idate_item          =19   ; cobit(19)='date yyyymmdd'
istid_item          =20   ; cobit(20)='integer station ID'
ibcctrl_item        =21   ; cobit(21)='Variational bias correction of obs'
isoe_item           =34   ; cobit(34)='Solar elevation'
ifov_item           =38   ; cobit(38)='Field of view'
icldcover_item      =42   ; cobit(42)='Cloud cover'
iqi1_item           =46   ; cobit(46)='SATOB QI1'
iqi3_item           =47   ; cobit(47)='SATOB QI3'
ipclear_item        =48   ; cobit(48)='% clear pixels'
ioskint_item        =49   ; cobit(49)='Obs Skin Temp'
imskint_item        =50   ; cobit(50)='Model Skin Temp'
iansens_item        =51   ; cobit(51)='An sensitivity'
ifcsens_item        =52   ; cobit(52)='Fc sensitivity'
ireo3qc_item        =53   ; cobit(53)='Reo3 quality'
iobsnorm_item       =56   ; cobit(56)='Normalised observation'
ifcdp1_item         =57   ; cobit(57)='Forecast departure 1'
ifcdp2_item         =58   ; cobit(58)='Forecast departure 2'
ifc1fit_item        =59   ; cobit(59)='Normalised Forecast departure 1 fit'
ifc2fit_item        =60   ; cobit(60)='Normalised Forecast departure 2 fit'
iobtyp_item         =71   ; cobit(71)='Obs type'
iobsubtyp_item      =72   ; cobit(72)='Obs subtype'
imodlsm_item        =84   ; cobit(84)='Land Sea Mask'
ireporttype_item    =85   ; cobit(85)='Datum Report type'
```

```
iscanline_item     =63  ; cobit(63)='Scan line'
ipolarisation_item =62  ; cobit(62)='SMOS polarisation'
itbcloud_item      =75  ; cobit(75)='Cloud/rain status'
ielev_item         =81  ; cobit(81)='Radar elevation'
iazim_item         =82  ; cobit(82)='Radar azimuth'
idprl_item         =83  ; cobit(83)='Datum 2nd vertical coordinate'
iproduct_type_item =310 ; cobit(310)='reo3 product type'

! Flag words

idsta_item         =22  ; cobit(22)='datum status (bits 27-30)'
irsta_item         =23  ; cobit(23)='report status (bits 27-30)'
irevt_item         =24  ; cobit(24)='report/blk events (17+15bits)'
irspe_item         =25  ; cobit(25)='report-type specific events'
idflg_item         =26  ; cobit(26)='datum flags (4+4+4+4bits 16-31)'
idevt_item         =27  ; cobit(27)='datum events (bits 4-30)'

! These are virtual items, computed only in updsoft. They are useful quantities
! derived from the extracted ODB/feedback variables
iubfg_item         =28  ; cobit(28)='background dept (no bias cor.) o-b'
iwind_item         =29  ; cobit(29)='wind speed'
inccv_item         =30  ; cobit(30)='incremental convergence'
ibent_item         =31  ; cobit(31)='Bennett''s diagnostic'
iafit_item         =32  ; cobit(32)='normalised analysis fit'
ihbhr_item         =33  ; cobit(33)='sqrt(HBH+R)'
ihbhr_item         =35  ; cobit(35)='normalised background fit'
ihbhr_item         =36  ; cobit(36)='relative analysis fit (%)'
ihbhr_item         =37  ; cobit(37)='relative background fit (%)'
isigoest_item      =39  ; cobit(39)='Est obs std error'
isigbest_item      =40  ; cobit(40)='Est bkg std error'
isigaest_item      =41  ; cobit(41)='Est ana std error'
irsigoest_item     =43  ; cobit(43)='Est rel obs std error'
irsigbest_item     =44  ; cobit(44)='Est rel bkg std error'
irsigaest_item     =45  ; cobit(45)='Est rel ana std error'
```

In addition to the items defined in the above list the users have the possibility to define up to five extra items without need of any Fortran code change. The items to add can be filled directly from the input ODB or virtually from the existing (or newly defined) items. This possibility is activated using the variable PARAMS in the `dobstat` script and it is limited to ODB and works only on linux workstation, ecgate or any other platform where the standalone version can run.

The station ID is also stored in a separate string array. Some bit items are packed values from several ODB/BUFR status or flag words. The meaning of the codes and bits in the above items is defined in configuration file `bufrodbcodes.cfg`. Note that item 3 is an ad hoc composite of the ODB/BUFR instrument code, retrieval method, cloudiness indicator, satellite ID, etc, whose precise definition is a function of the report type (particularly TOVS and SATOB satellite data). This definition is explicited in `bufrodbcodes.cfg` and it must be kept consistent with the relevant code in routine `buxtract.F90` for BUFR and in routine `odbread.F90` for ODB. This allows data selection at a finer level than BUFR/ODB subtype, while hiding the inconsistencies of BUFR encoding between the numerous data types.

The 'vertical coordinate' is usually the pressure (Pa), except for surface pressure observations (altitude), for satellite retrievals (index of layer), and satellite multichannel data (channel index).

Apart from these peculiarities and the scaling performed to diagnostics for some variable numbers (e.g. Ozone, GPSRO) all values in `obsdata` are the raw ODB/BUFR values, and no data conversion is performed, so that the user knows what is plotted from the BUFR feedback documentation.

The statistics are computed sequentially from the data in `obsdata`. There are two kinds of statistics. The 'hard' statistics are always computed as a diagnostic of the actual contents of the ODB/BUFR file, regardless of what the user expects. They contain global summaries of the amount of data available, which items are actually defined, etc., and they are only printed, not plotted nor output to a file. The 'hard' statistics provide a short summary of data usage and departures for each TPI (type, parameter, instrument codes) triplet that was found in the BUFR file. These statistics are held in module `stathard`.

The other kind of statistics is the 'soft' statistics that are requested by the user; their internal representation is a derived Fortran type in module `statsoft`, which contains both the *definition* of the statistics and their *values*. Their external representation is the `STATDEF` file, or, equivalently, the ascii statistics files managed by script `obstat`. The syntax for statistics specification is explained in section 4. The soft statistics are computed and accumulated at the level of the routine `updsoft.F90`. In this routine, the following actions are performed:

- Loop over data
- Filter the data according to the user request (content of `STATDEF` file): The filtering includes the obs type, code type, instrument, stream, flag, etc.
- compute the virtual items
- Accumulate statistics and fill the data structures associated to the different types of statistics: ascii `OBSTAT` stats, gridded statistics and scatter statistics.

With the implementation of gridded statistics support in `OBSTAT` users should be aware that they have more flexibility and possibilities to customize their statistics but this may impact the memory use and the computation time. The user can for instance request for all IASI channels to produce 2D ($0.5°x0.5°$) statistics for several diagnostics (e.g. FG and AN departures, bias correction,...), for several data selection (e.g. used, clear, active, ...) each 3 hours. The way `OBSTAT` is working implies that more we request more memory is needed: `OBSTAT` reads all the data and fills memory allocated arrays. The memory is released only after looping over all data.

After looping over all data, `OBSTAT` performs I/O operations to store the produced statistcs in ascii and binary formats.

3.3 `OBSTAT` configuration files

One of the main advantages of `OBSTAT` is its design allowing the perform almost configurations, customisation and data definitions through ascii configuration files. Fortran code changes are only required to add new features and products or to define specific treatment for novel observation types. Some configuration files are mandatory but few are optional. The list of `OBSTAT` configuration files is:

general.cfg This file defines the global properties used in the statistics definition & plotting. It is read at runtime by `OBSTAT`. The definitions controlled by this file are:

- Standard pressure levels used mainly by conventional observations
- Geographical areas (coordinates and names)
- Pressure levels specific to certain satellite instruments
- Plotting attributes: colours, curves, ...
- Definition of polar satellites orbits to be used for the production of ascending/descending statistics. The orbits definition is simply composed of the equator crossing time and the sign of mouvement at that moment (ascending/descending).

```
        | iniglob -> inititemloc
        | inigene -> loadcfg
        | inibufr -> loadcfg
        | inihard
        |                                         | defsensor
        |                        | initsoftdef  ->| matcharea
        |         | initsoft -> | wrsoftdef       | countlist
        | LOVER -> |             | allocsoft
        |         |             | initsoftstat
        |         |             | initsoftarea ->| allocsoft
        |         |             |                | matcharea
        |         |             |
        |         |             | inisoftinstr ->| allocsoft
        |         |             |                | defsensor
        |         |             |
        |         |             | inisoftflag  ->| allocsoft
        |         |             |                | defsensor
        |         |             |
        |         |             | inisoftstream->| allocsoft
        |         | copysoft -> allocsoft
        |
        | LMERGE -> mergesoft(T)
        |
        | LOVER  -> mergesoft(F)
        |
        |         | LBUF -> buopen
        |         |
        | LUPD  -> | LODB
OBSTAT->|         |
        |         | LCMA -> inicma
        |         |
        |         | LODB -> |odbread       |  | odbscatamb
        |         |         |odb2read      |->| odbscaling
        |         |         |user_data_read|  | updhard
        |         |                           | updsoft ->| LALLDATA -> addstat ->| defsensor
        |         |                                       |                       | allocsoft
        |         |                                       | winditem -> uv2sd
        |         |                                       | relhum
        |         |                                       | flags_check
        |         |                                       | param_val
        |         |                                       | utc_to_solartime
        |         |                                       | gridpos->sucalcgauw->gauaw->bsslzr
        |         |                                       | iniscat
        |         |                                       | enlstatarray
        |         |                                       | dataqc routines
        |         |
        |         |                                 | buprtdebug
        |         | LBUF ->| buread  ->| buxtract ->| prtobs
        |         |        |           | buprtdebug | buprintbox
        |         |        |                        | bucoord
        |         |        |                        | bustatid
        |         |        |                        | satobfreq_obstat
        |         |        | updhard
        |         |        | updsoft
        |         |
        |         | LCMA ->| cmaread
        |         |        | updhard
        |         |        | updsoft
        |         |
        |         |
        |         | PE>1 -> mpsoft -> comm_stat
        |
        |           | wrsoftdef
        | writesoft->| writealarm
        |           | writegribs ->| defsensor
        |           |              | genopt
        |           |              | GRIB_API
        |           | writescat
        |                                             | calcairspop
        |                                | plorrms  ->| defaxis
        | LPLOT -> plotsoft -> plothis ->| plotrmsbias
        |                                | plotusage
        |                                | plotime
        |                                | plotcov
```

Figure 6: OBSTAT Fortran code diagram

This mandatory file is located in DATADIR directory. Users have the possibility to point to their own version of this file via the variable GENERALFILE (in the script `dobstat`).

bufrodbcodes.cfg This mandatory file defines all codes used in OBSTAT to map the content of the ODB/BUFR with the data definition used by OBSTAT. The file is read at runtime by OBSTAT. The definitions controlled by this file are:

- BUFR and ODB codes of RDB report subtypes used at ECMWF
- Positions of BUFR and ODB report event bits in 32-bit integer
- Positions of BUFR and odb report/data status bits in 32-bit integer
- BUFR and ODB codes of observed parameters
- Values of BUFR datum flags
- Positions of BUFR datum event bits in 32-bit integer
- Definition of composite instrument/sensor. The formula used in the computation of the composite instrument/sensor is provided in `odbread.F90` routine and it is dependent on the observation type.
- Definition of the reportype used to identify observations in the ODB files archived in MARS

radiosonde_mask file containing the mask of radiosondes to collocate the data with for verification purposes

iasi_channels file containing the metadata associated to the set of IASI channels being screened by the IFS. The metadata provided is: channel index, channel number, channel wavenumber, channel wavelengh and the peaking level.

airs_channels file containing the metadata associated to the set of AIRS channels being screened by the IFS. The metadata provided is: channel index, channel number, channel wavenumber, channel wavelengh and the peaking level.

rtablel_xxx files containing the definition of gaussian grids for various resolutions.

params.txt this file is a template that can be used to define user defined diagnostics

flags.txt this file is a template that can be used to define user defined data selection criteria.

data_interface this file is a template to be used to define user defined obs input files.

3.4 OBSTAT statistics manipulation and plotting tools

After the production of statistics, there are tools designed to be used for the manipulation and plotting of statistics. The list of tools is provided below. Detailes will be provided in section 4.

dobstat_quick : script known as quick `obstat` is used to merge and plot the already computed statistics.

obstat_plot.x : binary file used by the scripts `dobstat` and `dobstat_quick` to produce the classical obstat plots (RMS plots, histograms, data usage diagrams, ...)

obstat_timeseries.x : binary to plot the old fashion OBSTAT timeseries. This binary will continue to be maintained even though the implementation of new tools to produce statistics from gridded statistics.

obstat_geo_plot : binary used to plot geographical mean statistics based on gridded statistics

obstat_hist_plot : binary used to plot timeseries of statistics based on gridded statistics

obstat_hov_plot : binary used to plot Hovmoeller diagrams based on gridded statistics

obstat_scat_plot : binary used to plot scatter statistics

obstat_overview_plot : binary used to produce, on the same plot, an overview of statistics timeseries for several channels and instruments

obstat_add_grib : binary used to accumulate non-normalized statistics. Non-normalised gridded statistics are useful to produce time averaged gridded statistics. Before use, non-normalised statistics should be normalised.

obstat_normalize_grib : binary used to normalise non-normalised statistics

normalize_obstatgrib.sh : script to wrap the binary obstat_normalize_grib

obstat_normalize_scat : binary used to Gather scatter statistics from different dates/times

obstat_grib_tools : perform arithmetic operations between gridded statistics.

obstat_grib_merge.sh : merge obstat grib fields according to one or several data selection (stream, time, surface, sensor, channel, satellite, phase, orbit)

obstat_grib_split.sh : splits `OBSTAT` grib files according to a splitting criteria: by cycle (default), by diagnostic, by flag, by date, by scan position and by land sea mask.

obstat_grib_dump.sh : dump the content of `OBSTAT` grib files in a simple ascii format. The tool allows the user to select what to dump and the geographical area of interest.

4 Statistics production

4.1 Types of statistical products

`OBSTAT` is designed to produce a wide range of end user products in the form of plots and warning messages. These products are based on intermediate ascii and binary statistics generated by `OBSTAT`. Although mainly designed to be used by the plotting package, ascii and binary files can be manipulated by expert users to generate custom products. In this section we focus on the description of the intermediate statistical products. The end user products will be described in 5.
The list of ascii and binary statistics is:

- Classical `OBSTAT` acsii statistics designed to produce the "**RMS profiles**", "**Histograms**" and "**usage charts**".
- High resolution gridded statistics (in GRIB2) designed to produce geo maps of statistics.
- Low resolution gridded statistics (in GRIB2) designed to produce all types of timeseries: area averages timeseries and Hovmoeller diagrams.
- Scan dependent statistics (in GRIB2) designed to produce area averages of scan dependent statistics timeseries.
- Solar time dependent statistics (in GRIB2) designed to produce area averges of solar time statistics timeseries.
- Targeted sites statistics

- Scatter statistics (in ascii) designed to produce scatter plots.
- Summary statistics (in ascii) designed to produce automatic warnings.

Before describing the above products, we start by the presentation of the grib template used by OBSTAT to encode gridded statistics.

4.1.1 GRIB2 statistics template

During the process of upgrading OBSTAT to produce gridded statistics, it has been decided to use the GRIB-2 format to encode the statistics. This format has several advantages:

- The format is self-contained allowing an easy access to statistics and metadata.
- Offers a good compression rate.
- Supported by ECMWF Archiving system (MARS). This means that statistics can be archived in MARS. The statistics retrievals can be done using filters (based on defined indexes).
- Well handled by ECMWF graphical tools (Magics, Magics++ and METVIEW). This means that users can use their own tools to produce plots from gridded statistics.
- Suitable for plotting on demand. Gridded statistics can be indexed and placed on-line on highly available discs. A web interface (such as EcCharts) can access them and produce statistical plots on demand.
- Convertible to other formats (e.g. NetCDF). GRIB-API has tools to access the data and even dump it in ascii format.

The template used for the encoding of gridded statistics is local to ECMWF but can be used by external users as it is part of the GRIB-API distribution. The template is described below using the GRIB-API keywords:

editionNumber [=2]: GRIB-2 edition

discipline [=0]: Meteorological products

typeOfProcessedData [=0]: Analysis products

marsClass [=1]: Operational archive (this will be changed to vary with the type of experiment)

marsType [=2]: Analysis

marsStream [=1025]: Atmospheric model

experimentVersionNumber [=$EXPVER]: Experiment version

observationType [=Observation type]: Although this keyword is part of the template but currently it is not used by the plotting package and we do not expect to use it for MARS indexation. The user can just ignore it.

codeType [=Observation code type]

varno [=variable number]

reportType [=Report type]: It is a unique number used at ECMWF to identify observations. This number will be part of the MARS indexation of gridded statistics archive.

phase [=0/1/2] used to identify setting/rising RO occultations or ascending/descending nodes

platform [=$SATID]: satellite identifier. For conventional observation the number 999 is used

instrument [=$SENSOR]: instrument identifier. For conventional observation the number 999 is used

dataStream [=1/2/255]: data reception stream. the value 255 is used for the global stream

typeOfFirstFixedSurface [=100/0]: 100 for pressure and 0 for channels, impact heights and the surface

scaledValueOfFirstFixedSurface [=`channel/level/layer`]:channel/level/layer(top) value

scaledValueOfSecondFixedSurface [=255/level]: pressute level/layer bottom pressure or missing for channels, impact heights and the surface

observationDiagnostic [=observation diagnostics]: quantities to produce statistics for (e.g FG departure)

dataSelection [=flag]: data selection criteria (e.g. "used" data)

scanPosition [=Field of View]: Binned Field of View or observation angles. Zero is used in case sacn dependent statistics are not requested.

mask [=surface mask]: land sea mask. The value 4 is used to identify "all surface types combined" is used by default when the surface mask dependence is not requested.

gridType [=regular_ll/regular_gg/reduced_gg]: specifies the grid type. For low resolution statistics, only regular grids are allowed (=regular_ll). The grid keywords and their values depend on the grid type.

dataDate [=`yyyymmdd`]: At the beginning of the time interval

dataTime [=`hh`]: At the beginning of the time interval

indicatorOfUnitOfTimeRange [=1]: the unit of time is hour

yearOfEndOfOverallTimeInterval [=`yyyy`]: year of the end time

monthOfEndOfOverallTimeInterval [=`mm`]: month of the end time

dayOfEndOfOverallTimeInterval [=`dd`]: day of the end time

hourOfEndOfOverallTimeInterval [=`hh`]: hour of the end time

lengthOfTimeRange [=`time step`]: Time step

The metadata are translated using the grib local definitions maintainted within GRIB-API. An up to date version of the `OBSTAT` grib tables is presented in appendix 1.

The use of GRIB2 format for the decoding of gridded statistics suffers from a slower access compared to other formats. This is due to the lack of a native built-in indexing. The indexing is performed for each execution of the plotting package. this operation is time consuming when there is a need to generate long time series especially for Advanced sounders. This problem will be solved in the future with the expected improvements of GRIB-API indexing. Although it is highly recommended to use GRIB format, `OBSTAT` can be easily adapted to use other formats such as NetCDF. The encoding of gridded statistics is done at the level of one routine (see writegribs.F90 in Fig. 6)
`OBSTAT` gridded statistics (in GRIB2) are designed to be archived in MARS. However more work is still to be done to achieve this. Currently, users have the possibility to specify the output storage location via the parameter GRIBDIR in `dobstat` script or in `obstat_init` for the prepIFS version. At ECMWF this can be done

on ECFS or local disks.
Unlike MARS, the current way of storage (on ECFS/discs) does not offer the filtering of statistics before use. To cope with the slower access issue, OBSTAT produces individual files per week and per ”channel/impact height/pressure level/pressure layer”. By doing so, users can access effeciently the requested statistics only without involving huge files each time. The produced grib files (belonging the the same product type) are then grouped in monthly tar files and stored. For the prepIFS version, one tar file is produced for each OBSTAT task. For the standalone version, a tar file is produced per stat block. This will be more consistent in the future versions of OBSTAT.
The naming rule used for individual grib files is:

PRODgrib_**EXP_STRM_OBS_VAR_SAT_SENS_STR_LEVTYP_CH1_CH2_YYYYMMWW**

where:

PROD product type :

- **GEO** for high resolution gridded statistics
- **HOV** for low resolution gridded statistics
- **FOV** for FOV dependent gridded statistics
- **LOC** for targeted sites statistics

EXP experiment version in four digits

STRM data assimilation stream (e.g. DCDA)

OBS string representing the code type. The value is derived from the file `bufrodbcodes.cfg`

VAR string representing the observed variable. The value is set at the level of the routine writegribs.F90

SAT satellite identifier in 3 digits. For conventional obs the value is 999

SENS sensor identifier in 3 digits. For conventional obs the value is 999

STR data stream in 1 digit. If the data stream is not requested the string **STR** is not present

LEVTYP level type: PR is used for pressure levels/layers. Otherwise the value CH is used

CH1 level/channel/layer(top) value in 6 digits

CH2 level/channel/layer(bottom) value in 6 digits. For channels and height impacts CH2 is the same as CH1

YYYYMMWW 8 digit number composed of the year, month and week number.

4.1.2 Classical OBSTAT *ascii statistics*

The classical OBSTAT ascii statistics are the default output of OBSTAT since its first implementation. The structure of the output file is an extension to the input STATDEF file (detailed in 4.2). Each stat block, is followed by the produced statistics for each quantity. The classical OBSTAT ascii statistics ar categorized in three main types/plots:

RMS profiles : There are three main types of plots (see Fig. 7). The ‘**RMS profile**’ plots display vertical profiles of rms average and arithmetic average (‘BIAS’) departures for conventional data as a function of datum pressure. Actually they are histograms, the values are only linked by dashes because they happen to have some vertical consistency. The **binning** is represented on the y-axis by the nominal pressure of the bin, in hPa. The bin limits are the average (in pressure) of the nominal pressures of the two adjacent bins. The top bin includes all data above its lower bin limit. For satellite radiances the bins are defined by the channel indices and tagged by the index or a channel acronym (for TOVS: S=SSU, M=MSU, H=HIRS). Do not always assume that higher-index channels see a lower part of the atmosphere, refer to the documentation supplied by the satellite section to know their meteorological significance.For thicknesses there is one bin per layer, tagged by the *middle pressure* of the layer. The layer boundaries are easy to guess, e.g. 150 (between 85 and 250) is the 100-200hPa layer. Finally, the x-axis labels give the (fixed) horizontal scaling, in the same units as the data.

In the RMS profile plots, consecutive values are connected by dashes if the bin populations are large enough to give meaning to the statistics. The populations are displayed in a column. Nothing is plotted if the population is below a fixed threshold (currently 5) of if the values are outside the plot. A meaningful statistic that is isolated between two consecutive levels is denoted by a symbol. The distinction between the departures is as follows:

a plus + or a continuous line for background departures

a cross x or a dashed line for analysis departures

Usually, the rms analysis departure is smaller than the rms background departure.

Histograms : displays simple statistics (total population, minimum, maximum, average and standard deviation of the values) and a population histogram for each departure. The population histogram represents the probability density function of the departures if the population is large enough (several hundreds of data are required to get a reliable picture); if the plot is jagged it is normally because the population is very small. The x-axis gives the horizontal scaling, in the same units as the data; data that fall outside the plot are gathered in the extremal bins.

Usage charts : are a small bar chart that indicates the amount of data that is used, blacklisted and rejected, plus some statistics (rms, bias, standard deviation) for the whole data and for the subset that is used. It is mainly used to check the data coverage and changes to the blacklisting and quality control practices.

4.1.3 High resolution gridded statistics

High resolution gridded statistics are mainly designed to produce geographical maps of statistics. Such statistics are meaningful only when averaged over a period of time to accumulate a representative number of observations over each grid box. Conequently the default, is to produce **time averaged** high resolution gridded statistics. `OBSTAT` also offers the possibility to produce high resolution gridded statistics per each analysis cycle. This option is costly but it is usefull for special investigations or for the production of daily data coverages without compromising the possibility to time average statistics for any chosen period.
The `OBSTAT` support of high resolution gridded statistics is characterised by:

- The production of the high resolution GRIBs is optional (default yes)
- Three grid types supported : REGULAR LAT/LON, REGULAR GAUSSIAN and REDUCED GAUSSIAN
- For REGULAR LAT/LON grids, free specification of the resolution (default $1^{\circ}x1^{\circ}$)
- For GAUSSIAN and REDUCED GAUSSIAN : N16, N24, N32, N48, N64, N80, N108, N128, N160, N200, N256, N320, N400, N512, N576 and N640

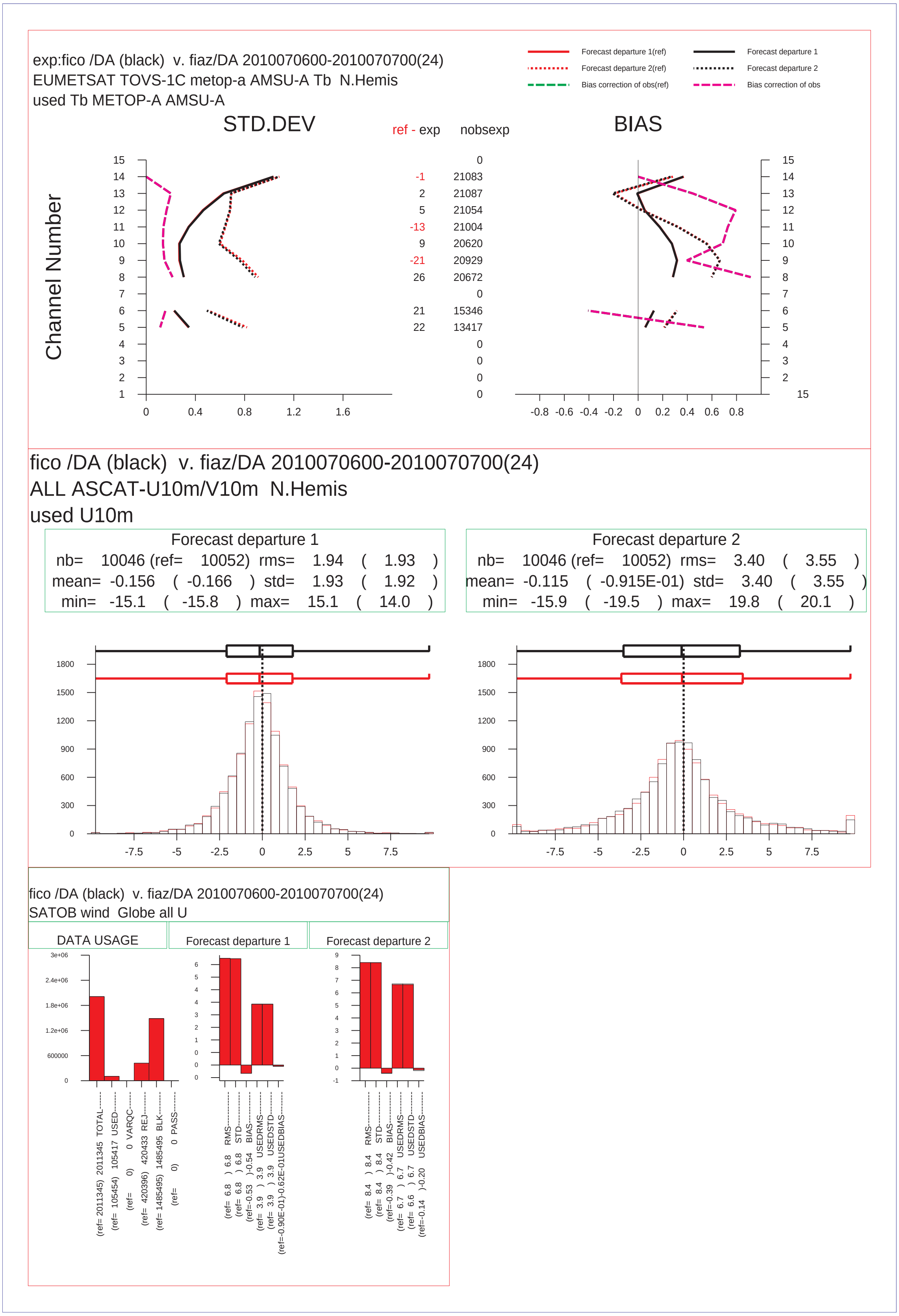

Figure 7: Example of classical OBSTAT *plots: RMS profile (top), histogram (medium panel) and usage chart (bottom)*

- A bitmap is used to handle missing values

High resolution gridded statistics can be produced for:

- Channels/levels/layers/impact heights
- Diagnostics (e.g. FG departure)
- Time (average or explicit)
- Data selection flags
- Field Of Views
- Ascending/Descending nodes
- Setting/Rising GPSRO

For each diagnostic, `OBSTAT` generates by default the average and standard deviation. The data counts are systematically produced.
The high resolution gridded statistics are encoded differntly depending on the time averaging.

- If statistics are not averaged in time, `OBSTAT` produces (for each time slot) statistics normalised and ready for use: The encoded values are the final ones. A GRIB file belongs to this category if the data counts grib message is encoded using the code **199**.
- If statistics are averaged in time, `OBSTAT` produces accumulated values over each grid box. Together with the data counts, the user (and the plotting tools) can at any time, produce the final (normalised) statistics (average and stdev/rms). By keeping the accumulated values, `OBSTAT` can resume an experiment just by contunuing the accumulation process. A GRIB file belongs to this category if the grib code of the data counts is **198**. The averaging in time can be done weekly or monthly. The output of the normalisation process (see 3.4), has **197** and the grib code of the data counts.

In both methods, `OBSTAT` keeps statistics for different time slots during the day. The time slots depend on the time step used (by default it is the analysis window). For instance, with a time step of 6 hours, `OBSTAT` keeps statistics for 00, 06, 12 and 18. The plotting package can then gather statistics from all time slots or produce statistics from selected time slots only.

4.1.4 Low resolution gridded statistics

Low resolution gridded statistics are mainly designed to produce time series statistics of area averages. This product is very valuable as it is the one used in a daily basis to check the overall availability and quality of the data. Furthermore, a wide range of end user products can be derived from these statistics.
The `OBSTAT` support of low resolution gridded statistics is characterised by:

- The production of the low resolution gridded statistics is optional (default yes)
- Only REGULAR latitude/longitue grids are supported with a free user specification of the resolution (default $2.5°x120°$)
- Capacity to produce statistics according to the surface mask: land, sea, sea-ice and all surface types

- A bitmap is used to handle missing values

Low resolution gridded statistics can be produced for:

- Channels/levels/layers
- Diagnostics (e.g. FG departure)
- Time
- Data selection flags
- Land-Sea mask
- Field Of Views
- Ascending/Descending nodes
- Setting/Rising GPSRO

As mentioned above, this product can be used to produce a wide range of products. However what can be produced depends strongly on the selected resolution (The user has control over this for each stat block). Below some guideline on how to specify the resolution:

- The resolution is composed of two values: latitudinal and longitudinal resolutions in degrees
- Higher resolutions satisfy all needs (by downscaling) but require more disk space and also more time and memory to produce.
- If the only need is to produce time series of global averages, the user can simply decide to produce one value for the whole globe by defining the resolution as $360° X 180°$
- If the interest is to produce statistics for large areas over the globe, the user can split the globe in grid boxes large enough to satisfy the need. e.g. $360° X 20°$ is suitable to produce statistics over large latitude bands.
- If the interest is for area averages statistics in addition to latitude bands against time Hovmoeller diagrams, the user can, for instance, define a small latitudinal resolution. e.g. $30° X 2.5°$
- If the interest is for area averages statistics in addition to time against longitude Hovmoeller digrams, the user can, for example, define a small longitudinal resolution. e.g. $5° X 20°$

4.1.5 Scan dependent gridded statistics

Scan dependent statistics are designed to produce area and time averages of statistics for binned Field of Views. This product can be categorised as low resolution gridded statistics as they are based on the same regular lat/lon grid. The main difference with normal low resolution gridded statistics is related to the time dimension. The production of statistics for a large number of Field of Views (the number depends on the applied FOV binning) increases the memory and disk space requirements. Such requirements are not affordable without averaging in time. This makes sense as the focus is generally not to have the time evolution of statistics per FOVs. If there are needs, users have still the possibility to produce low resolution gridded statistics with both time and FOV dimensions enabled.
In addition to the time dimension, the land-sea mask dependency is not included. The ascending/descending separation is also ignored.
In summary, FOV dependent statistics can be produced for:

- Channels/levels/layers
- Diagnostics (e.g. FG departure)
- Time
- Data selection flags
- Field Of Views

This product is optional and not produced by default.

4.1.6 Solar time dependent gridded statistics

For highly inclined low orbiting satellites (e.g. TRMM/TMI), each orbit covers areas with different sun viewings. For such satellites, it is usefull to assess the effect of the sun on statistics. This is achievable by the production of the **solar** time evolution of global averages. If there is evidence of a significant influence of the sun on the data quality, appropriate measures can be taken to remove the sun effect (through the bias correction) or to blacklist the data. Although considered as gridded statistics, one point is produced over the globe. Statistics are encoded in GRIB2.
In summary, this product produces statistics for

- Channels/levels/layers
- Diagnostics (e.g. FG departure)
- Solar Time
- Data selection flags
- Field Of Views

This product is optional and not produced by default. Currently this product is generated routinely for TRMM/TMI only.

4.1.7 Targeted sites statistics

This product is designed to produce time series statistics for targeted sites or small areas. Such product is usefull for CAL/VAL operations.
The user has a generic way to specify, for each stat block, the list of sites names for which the statistics are to be produced (see Fig. 8). The sites names are defined in `general.cfg` file. `OBSTAT` uses a regular lat/lon grid with a resolution of $2°x2°$. All observations (fulfilling the stat block conditions) belonging to a site are included in the statistics. At the level of the plotting, the user can request to produce statistics for the different sites. This product is optional and not produced by default.

4.1.8 Scatter statistics

Scatter statistics are a valuable tool to easily evaluate the relationship between two variables. The product indicates the degree of correlation between two physically comparable quantities. `OBSTAT` offers a generic way to produce scatter statistics. The product is in ascii format and it is self-contained. In addition to the

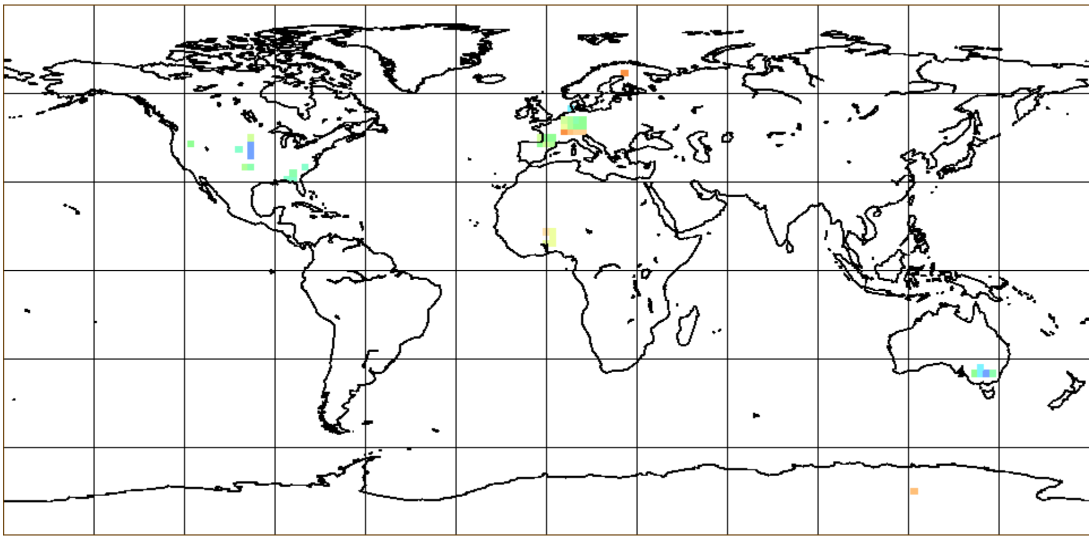

Figure 8: Example of targeted sites statistics. The product is designed to produce time series (This map is only a demo)

metadata, the file contains the population (number of observations) distribution for binned X and Y. Global statistics are also included (such as maximum, minimum, correlation, ...).
OBSTAT support of scatter statistics is characterised by:

- The production of scatter data is optional (default no)
- Up to 18 pairs of variables can be requested for comparison for each stat block
- Easy procedure to define the binning to be used for the generation of the populations
- Scatter data can be produced, in one go, for different channels/levels/layers/areas and for different data selection criteria
- Populations are stored in a self-documented ASCII file.

OBSTAT offers the possibility to produce scatter statistics separately for each analysis cycle or gather statistics over a period. The first options requires obviously more disk space but offers more flexibilty for users to produce on demand plots for any covered period. Tools are available to gather statistics.

4.1.9 Summary statistics

Summary statistics are a very compact ascii files containing global statistics. Each line is composed of the time, obs definition, flag, and global statistics for the requested quantities. This product is designed mainly to provide input to the alarm system. Its lightness and easy accessibility may be attractive for users.

4.2 Handling of the vertical coordinate

OBSTAT is able to handle all existing vertical coordinate types: Channels, Pressure levels, pressure layers, impact heights and surface. For each observation, OBSTAT reads the vertical coordinate value and assign it to the appropriate position depending on the statistics definitions. The following rules are applied:

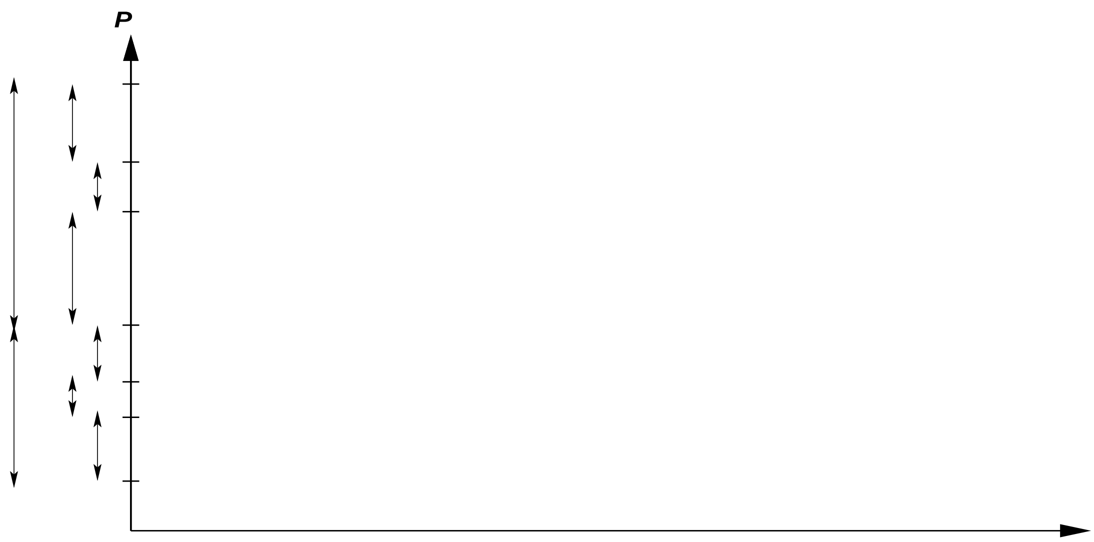

Figure 9: Example of user defined pressure binning

- For channels and impact heights, it is necessary to sepcify the number of levels to consider (via keyword nbbin: see 4.7) for the statistics computation. For RMS profiles, the number of considered vertical bins is exactly the one specified in the stat block. To consider all channels and impact heights the user needs to specify a number equal or heigher than the real number of channels. For gridded statistics, it is highly recommended to specify (in the stat block) the exact number of channels. If the specified number is higher than 0 but less than the real number, `OBSTAT` adjusts dynamically the value of nbbin each time an undefined level is found. However this operation is time consuming as `OBSTAT` needs to deallocate and allocate statistics arrays several times.
- For pressure levels and layers, the user can, by ignoring to set nbbin, uses a default vertical pressure binning defined in `general.cfg`. If nbbin is set to a value higher than 0, `OBSTAT` uses for non ozone data the actual pressure of the observation. For ozone data (codetype=206) with $nbbin > 0$, `OBSTAT` first verifies if there is a pre-defined pressure binning for the data in `general.cfg` (the definition of pressure binning in `general.cfg` is done by instrument and parameter). In case the pre-definned binning is not found, `OBSTAT` uses the actual pressure of the observation.
- For pressure levels and layers, the user can ignore, for a stat block, any pre-defined binning by specifying her/his own vertical binning (works only with gridded statistics). The user is free to specify the layers (number and thickness). The overlapping of layers is not restricted (see Fig. 9).
- For surface observations, it is recommended to specify nbbin equal to 1 for gridded statistics.

4.3 User defined quantities

As a generic diagnostic tool `OBSTAT` should allow users to generate statistics for all observation related quantities. Currently, most of qunatities are already defined (hardly coded) in `OBSTAT` (see 3.2). If the user is interested by a non pre-defined quantity the standalone `OBSTAT` offers a generic procedure to include it in the generation of statistics. This is achievable by the activation of the parameter PARAMS in `dobstat` script. The parameter PARAMS points to a user modified ascii file containing the definition of the desired new quantities. There is a mandatory rule to define quantities. Users are advised to get a copy of the file `params.txt` (available in DATADIR) and change it according to their needs. The procedure used by `OBSTAT` to produce statistics for user defined quantities is:

- Reads the file provided by PARAMS
- Performs consistency checks of the file.
- Application of a parsing and syntax analyser tool (based on yacc and lex) to generate a fortran routine (param_val.F90) providing the definition of the virtual quantities (computed from other quantities)
- Updates the ODB SQL queries to include the retrieval of the new non-virtual quantities.
- Genrate a new obstat_calc.x and obstat_plot.x binaries taking into account the updated SQL requests and the new fortran routine.

Only five new quantities are allowed. Once defined, the new quantities can be used by just requesting them within the STATDEF file.

The structure of the configuration file is:

```
#------------------------------------------------------------------------------#
# Param      SQLflag       Number     GribNumber      CompMethod          Datalist #
#                          GE 100      GE 200
#------------------------------------------------------------------------------#
  tausfc  tausfc@body      105         228               0     amsua,amsub,hirs,mhs
  stalt    stalt@hdr       101         226               0                 conv
  modoro    modoro@hdr     102         230               0                 conv
 stalt-modoro   0          103         232      (itm_101-itm_102)          conv
  htcort2       0          104         233     (itm_9+0.0065*itm_103)      conv
  fgvalue       0          100         230       (itm_8-itm_9)             all
#------------------------------------------------------------------------------#
```

where:

- Param is the name provided by the user to identify the quantity. This name is spreaded to the product
- SQLflag is the ODB word to access the quantity from the ODB. For virtual items, SQLflag should be zero.
- Number is the number identifying the quantity inside OBSTAT. This number is the one used in STATDEF. This number should be higher than 100 as the number below are already used for the hard coded quantities (see 3.2).
- GribNumber is the number used to identify the quantity inside the grib (for gridded statistics). For each quantity the number provided is used for the average. The standard deviation is coded using the provided number incremented by 1. To avoid using the same number for different quantities, it is highly recommended to use only Even numbers.
- CompMethod is used to provide the computing rule for virtual quantities. It should be equal to zero for non virtual items (derived from the ODB). The computation method is fortran like arithmetic operation involving already defined quantities. Already defined quantities are identified by itm_OBSTATnumber. User defined quantities can be used also to define virtual items. If CompMethod is provided, the SQLflag should be zero. If SQLflag is provided the CompMethod should be zero. OBSTAT translates the computation method to a fortran routine that is plugged to the OBSTAT code during the generation of obstat_calc.x and obstat_plot.x.
- Datalist restricts the use of the new defined quantity to certain data types.

Before defining a new quantity it is important to check that is not already hard coded. The list of the currently hard coded items is presented in 3.2.

4.4 User defined flags

As a generic diagnostic tool OBSTAT should allow users to produce statistics for all possible data selection criteria. Currently, most usefull data selection criteria are defined (hardly coded) in OBSTAT. If the user is interested by a non pre-defined data selection criteria the standalone OBSTAT offers a generic procedure to take it into account in the generation of statistics. This is achievable by the activation of the parameter FLAGS in dobstat script. The parameter FLAGS points to a user modified ascii file containing the definition of the new data selection criteria. There is a mandatory rule to define such criteria. Users are advised to get a copy of the file flags.txt (available in DATADIR) and change it according to their needs. The procedure used by OBSTAT to handle users defined data selection criteria is:

- Reads the file provided by FLAGS
- Performs consistency checks of the file
- Application of a parsing and syntax analyser tool to generate fortran routine (flags_check.F90) providing the definition of the new flags.
- Generate a new obstat_calc.x and obstat_plot.x binaries by plugging flags_check.F90 to the existing code.

Only five new flags are allowed. Once defined, the new flags become ready for use by using them within the STATDEF file.

The structure of the configuration file is:

```
#-------------------------------------------------------------------------------#
# flag              Number             Definition                   Datalist
#                   GE 200
#-------------------------------------------------------------------------------#
  daytime            205               (itm_34>0)                       reo3
  Night              206               (itm_34<=0)                      reo3
#-------------------------------------------------------------------------------#
```

where:

- flag is the flag name provided by the user to identify the data selection criteria. This name is spreaded to the products
- Number is the flag number identifying the new flag inside OBSTAT. This number is the one used in STATDEF. This number should be higher than 200. Numbers below 200 are reserved for hard coded data selection criteria.
- Definition is the arithmetic relation used to define the new flag. Quantities involved in the definition are identified by itm_OBSTATnumber.
- Datalist restricts the use of the new flag to certain data types.

Before defining a new flag, it is important to check if it is not hard coded. Current hard coded flags are listed in 4.7.

4.5 User defined input feedback files

For tests, validation and comparison purposes, it is usefull to be able to use the OBSTAT framework without having to have the feedback information in ODB or BUFR. To do so the user needs to :

- Have access to the feedback information from ascii/binary files. The data should be organised by columns. Each column contains an observation related quantity or a metadata. As for ODB, some information are mandatory for OBSTAT to be able to process the data (e.g. position, time, ...).
- In the script dobstat the user need to set the parameter OBSFORMAT to USER. The user needs also to specify the datalist as for ODB.
- The user needs to activate the parameter DATA_INTERFACE (at the level of dobstat script. This parameter points to a configuration file necessary to ensure the mapping between the input file columns and the internal OBSTAT items (see 3.2). This file should follow a defined format. Users are advised to get and modify a copy of the file data_interface from the DATADIR folder. The data interface file contains the list of OBSTAT hard coded items. Each item is assigned to a column number. The column number is zero if the item is not available in the data input file. Otherwise, this number is the number of the column containing the item values.
- Finaly, the user needs to specify where the data files are located. This is done via the parameter USERDATA in the script dobstat.

An example of data_interface file is :

```
form= 'ascii'
format= 'none'
missing= 2147483647
#--------------------------------------------
# Attribute          Col            Definition
#--------------------------------------------
BEGIN ATTRIBUTES
varno                001             Variable number
lat                  002             Latitude
lon                  003             Longitude
modlsm               000             Land Sea mask
press                004             Vertical coordinate
obsvalue             005             Observation value
fg_depar             006             FG departure
an_depar             007             Analysis departure
obs_error            008             Observation error
fg_error             009             FG error
final@update_1       000             lo-res ana departure update 001
hires@update_2       000             hi-res bg departure update 2
final@update_2       000             lo-res ana departure update 2
fg_check_1           000             probability of gross error
fg_check_2           000             range of possible values
biascorr             010             Bias correction of Obs
statid               011             integer station ID
biasctrl             012             Variational bias correction of obs
solar_elevation      013             Solar elevation
quality_retrieval    015            reo3 quality
fov                  000            Field of view
cldcover             000            Cloud cover
fov_reo3             016            Field of view for reo3
mf_stddev            000            Obs Skin Temp
predictor_11         000            Model Skin Temp
```

```
qi_1                000             SATOB QI1
qi_3                000             SATOB QI3
csr_pclear          000             % clear pixels
an_sens_obs         000             analysis sensitivity
fc_sens_obs         000             forecast sensitivity
press_rl            014             datum 2nd vertical coordinate
elevation           000             radar elevation
azimuth             000             radar azimuth
codetype            017             report type
sensor              018             report instrument/sensor type
satinst             000             report instrument/sensor type
datastream          000             datastream
time                019             Report time
date                020             Report date
tbcloud             000             Cloud/rain status
active              021             Obs Active status
passive             022             Obs Passive status
rejected            023             Obs rejection status
blacklisted         024             obs blacklist status
rdb_rejected        025             odb rdb rejection status
datun_redundant     026             datum redundunt
level_redundant     027             level_redundant
duplicate           028             duplication status
redundant           029             Redundant
varqc_flag          030             VarQC-Flag
fgcheck             031             fg check status
product_type        032             reo3 product type
param1              000             user's additional atribute 1
param2              000             user's additional atribute 2
param3              000             user's additional atribute 3
param4              000             user's additional atribute 4
param5              000             user's additional atribute 5
END ATTRIBUTES
```

4.6 Input database auto browsing

By default OBSTAT produces output statistics only for the stat blocks defined in the STATDEF file (Although hard statistics are produced for all data present in the ODB/BUFR). For users interested by systematically producing statistics for all data present in the ODB, OBSTAT offers an option to do so. This is achievable by activating the option LALLDATA (=true) in the dobstat script for the standalone version and in obstat_init for the prepIFS version.
By activating LALLDATA, OBSTAT builds a stat block for each observation not belonging to one of the predefined stat blocks. The decision to generate a new stat block does not include any data selection criteria and it is based only on the codetype, observed parameter, instrument and stream. To build a new stat block, OBSTAT needs to know what to specify as data selection criteria, land-sea mask, diagnostics, type of statistics to produce, grids resolution, ... The following rules are applied:

- OBSTAT produces "generic" statistics for all data not defined in one of the present stat blocks (in the STATDEF file).

- The user has the possibility to define the content of "Generic" statistics (grid resolution, flags, diagnostics, time slot, surface mask, etc.). To do so, the user can add one special stat block to the STATDEF file. This special stat block should have 77777 as instrument ID. The statistics definitions specified in this special stat block applies automatically to all new defined stat block. The existing stat block keep their statistics definitions.

- If the special stat block is not present, OBSTAT has its definition of ”Generic” statistics. The default ”generic” definition is :

  - production of high resolution lat/lon gridded statistics with a resolution of 1.5°x1.5°
  - production of low resolution gridded statistics with a resolution of 360°x2.5°

  production of statistics only for ”used” data

  production of statistics only for Obs values, FG and AN departures and bias correction. for GPSRO data, the items considered are normalised obs, FG and AN departures.

  Again, this default definition of ”Generic” applies only to new defined stat blocks.

4.7 Statistics definition

The definition of statistics is fully controlled by the STATDEF ascii file. The user has control over the definition of statistics by providing his/her own version of the STATDEF ascii file. This file is composed of up to 1000 statistcs blocks. Each block is independent of the others and it contains the definition of specific statistics. Each block (see example below) is supposed to produce one plot (one MAGICS subpage), gridded statistics (in GRIB2) or scatter statistics in ascii format.

```
BEGIN STATDEF
comment= 'TEMP-Uwind'
statkind= 1
areaNSEW= NH.Tr.SH
types= 101 102 106
params= 11003
items=      9    10
instrument=       999
flagfilter= 1
END STATDEF
```

For each run OBSTAT reads the content of the STATDEF file and produces the available statistics for each block. A block is left unchanged if the requested statistics are not available (e.g. the requested data not present in the input ODB/BUFR).
The first line of the STATDEF ascii file should be `#obstat version n` (where n=0,1,2 or 3) to cater for future evolutions in the specification language, and that you may incorporate comment lines after a # sign.

The syntax inside each block is `keywords= values[s]` where it is important to leave the spaces as separators (for a Fortran list-directed read) and to keep the **same line ordering**. Some keywords are mandatory. Non mandatory keywords may have dependencies with other keywords. If dependencies are not satisfied, an abort occurs together with an explanatory message. The up-to-date documentation of the keywords and values is in module `statsoft`. The reading of values is done at the level of `inisoftdef` routine (see 6). The list of available keywords, at the time of writing, is presented below following the necessary ordering. Mandatory keywords are highlighted. The dependencies are also highlighted when possible.

comment [mandatory]: a quoted string that will be displayed only inside the MAGICS panel for this statistic. Useful to describe the observation type, paramter and instrument when more than one code is used in the definition.

statkind [mandatory]:the type of statistic, as documented in module `statsoft.F90`. Currently supported choices:

1. vertical rms and bias profiles binned according to vertical coordinate satellite channel index. The binning for pressure levels is defined in file `general.cfg`. Clear-text channel index names are displayed for some satellite data types.
2. histograms binned according to value, with numeric statistics.
3. bar-chart of number of used and rejected data, and of statistics.
4. output a listing of data in the style of Metview geopoints, into an ascii file in your EXEDIR.
5. histogram as a function of the time of the day. It wraps every nbbin hours and the side bins are halved just like in 4D-Var.
6. left for future use, to compute autocovariance histograms for the estimation of true obs and background errors.
7. dummy definition used to force changing page in the plot at this point.
8. gridded and scatter statistics

lowresgrib [optional and only when statkind= 8]: enable(y)/disable(n) the production of low resolution gridded statistics (in GRIB2) designed mainly to produce timeseries. The low resolution gridded statistics are produced by default.

highresgrib [optional and only when statkind= 8]: enable(y)/disable(n) the production of high resolution gridded statistics (in GRIB2) designed mainly to produce geograpical maps of statistics. The production of these statistics is enabled by default.

solartimegrib [optional and only when statkind= 8]: enable(y)/disable(n) the production of gridded solar time dependent statistics (in GRIB2). The production of these statistics is disabled by default.

fovgrib [optional and only when statkind= 8]: enable(y)/disable(n) the production of Field Of View dependent statistics (in GRIB2). The production of these statistics is disabled by default.

orbitstat [optional and only when statkind= 8]: enable(y)/disable(n) the production of ascending/descending statistics (in GRIB2) for low orbit satellites. The production of these statistics is disabled by default.

locgrib [optional and only when statkind= 8]: enable(y)/disable(n) the production of targeted area statistics (in GRIB2). The production of these statistics is disabled by default.

alarm [optional and only when statkind= 8]: enable(y)/disable(n) the production of summary statistics (in ascii) to be used mainly as input of the automated alarm system. The production of these summary statistics is enabled by default. However, the global parameter **ALARM** should be set to "yes" in the master script `dobstat` (for the standalone version) and `obstat_init` (for the prepIFS version).

The order of the previous keywords (from **lowresgrib** to **alarm**) is not important but they should be just after the **statkind** line.

areaNSEW definition of the geographical area from where to select the data. A number of areas are predefined in file `general.cfg`, in which case the name of the area can be specified directly. A special name is `NH.Tr.SH` which expands into three statistics on `N.Hemis`, `Tropics` and `S.Hemis`. Areas for which there is no name can be specified as 4 reals: N and S latitudes, E and W longitudes (in degrees). Two additional reals can optionally be specified to select data from a layer: the minimum (inclusive) and maximum (exclusive) value of the vertical coordinate. For radiance data the vertical coordinate is the channel index, so specify e.g. 5.9 and 6.1 to get channel 6 (to allow for integer rounding). If you want to reuse a custom area definition the easiest is to put is name into file general.cfg and give just the name in the statdef file.
In case of statkind= 8, the areaNSEW should always be equal to GLOBE. This restriction may be removed in the next versions of `OBSTAT`

grid [optional but works only with statkind= 8]: Defines the specifications of the grids to be used for the production of gridded statistics. A minimum of 2 and a maximum of 4 entries need to be provided for this keyword. The first entry can have the following values:

- REGULAR for regular grids
- REGGAUSSIAN for regular gaussian grids
- REDGAUSSIAN for reduced gaussian grids

the grid types listed above concern only the high resolution gridded statistics. The grid used for the low resolution gridded statistics is always regular. The second entry defines the resolution to be used for the high resolution statistics. The resolution depends on the grid type:

- Real values comprised between 0.1 and 60
- An integer value from the list: $31, 47, 63, 95, 127, 159, 215, 255, 319, 399, 511, 639, 799$. These values correspond to the predefined gaussian grids. For more information see rtablel_xxx in DATADIR or http://www.ecmwf.int/products/data/technical/gaussian/index.html

The third and fourth entries (optional) define the longitudinal and latitudinal resolutions for the low resolution grids. The default resolution is $120°x2.5°$. Later we will give guidelines on how to choose the resolutions according to the needs.

locareas [required for locgrib= y]: defines the target sites for which the user wants to produce statistics. The user has the possibility to define as many sites as far as the list of sites names can be hold in a string of 180 characters. The sites names are to be defined in the file `general.cf`. If a site is not defined in `general.cfg` the user has the possibility to use its own `general.cfg` file via the variable GENERALFILE (at the level of OBSTAT script). The concept of targeted sites will be described later in this documentation.

surfaces [optional and works only with statkind= 8]: defines the list of surface types for which the user needs to produce gridded statistics separately. By default, OBSTAT produces statistics for ”All surface types combined”. The possible values are:

1. Land
2. Sea
3. Sea-ice
4. All surface types combined

The production of surface type dependent statistics relies on the availability of the land sea mask information which is usually present in the ODB for all observation types. If the land sea mask information is not accessible (e.g. not specofied in the ODB SQL request) OBSTAT notifies the user and produces statistics for all surface types combined. The production of surface type dependent statistics is only available for low resolution gridded statistics.

tslot [optional but works only with statkind=8]: defines the time step to be considered for the production of gridded statistics. The possible values are from 1 hours to the data assimilation window (default). At the end of each time step, OBSTAT produces an image of statistics. Users are free to use this facility according to their needs but they should be aware of the memory requirements increase.

types [mandatory] list of observation (sub)types allowed to enter the statistic, as defined in file `bufrodbcodes.cfg`

params [mandatory] list of datum parameter codes allowed to enter the statistic, as defined in file `bufrodbcodes.cfg`

items [mandatory] list of item indices (see 3.2 for details) on which to compute statistics. Popular choices are 9 and 10, the background and analysis departures. Each item generates a curve (and a legend) in the vertical rms profile plot, and a maximum of two items are plotted in each histogram plot. For gridded statistics, the mean and standard deviation statistics are produced for each item. Data for which the item is not defined simply do not enter the statistics. If an item is not enabled (at the level of iniitemloc) `OBSTAT` will abort and produce an explanatory message.

instrument [mandatory] list of instrument codes as defined in file bufrodbcodes.cfg, or 999 to match any instrument. These codes are usually a composite of the satellite ID and some other interesting characteristic of the measurement, defined in an ad hoc way by routine `odbread` (for ODB), `odb2read` (for ODB2) or `buxtract.F90` (for BUFR).

wildcard [optional]: Specifies whether or not `OBSTAT` should merge statistics from the listed instruments. By default `OBSTAT` produces separate statistics for each instrument in the list (**wildcard**= n)

datastream [optional]: list of data streams for which `OBSTAT` should produce statistics. The valid values of the data stream are those present in the ODB (odb word datastream@sat). The current possible values are

1 EARS (for ATOVS and SCATTEROMETER) or Direct Broadcast MODIS AMVs

2 PAC-RARS (for ATOVS)

0 Global stream

If the keyword **datastream** is not specified, `OBSTAT` will produce statistics for all data stream combined. If datastream dependent statistics are requested but the datastream information is not accessible (e.g. not relevant for the data type or not specified in the ODB SQL request) `OBSTAT` will notify the user and produces statistics for all streams combined.

flagfilter [mandatory] code of data selection, as documented in module `statsoft.F90`. Currently available choices:

1 data with status=active

2 all data processed

3 all data with status not active (i.e. rejected, blacklisted or passive)

4 best active SCAT (closest to analysis)

5 used data (active and not rejected by the variational QC) i.e. data that has a significant weight in the analysis

6 rejected by the variational quality control

7 blacklisted

8 failed (non blacklisted and rejected still)

9 data passed first guess check

10 data that is not rejected

11 active and passive data (i.e. data that is varBC bias corrected but has very low weight in the analysis

12 Data failed FG check but not blacklisted

13 Data failed FG check and VARQC rejected

20 AMVs with QI < 20

21 AMVs with $20 < \mathrm{QI} \leq 65$

22 AMVs with QI > 65

23 AMVs with QI $>$ 80

24 AMVs with QI $>$ 90

30 GEOS data with clear fraction $<$ 70 % (WV) and $<$ 80 % (IR)

31 GEOS data with clear fraction $\leq$ 70 % (WV) and $\leq$ 80 % (IR)

32 GEOS data completely clear (according to IR window channel)

33 GEOS data with clear fraction $\leq$ 40 % (WV)

34 GEOS data with clear fraction $\leq$ 70 % (WV)

35 GEOS data completely clear (according to WV channel)

40 Clear

41 [for microwave obs only] Used clear data. The production of this flag requires the availability of tbcloud@hdr or report_tbcloud@allsky valid info in the ODB.

42 [for microwave obs only] Used cloudy and rainy data. The production of this flag requires the availability of tbcloud@hdr or report_tbcloud@allsky valid info in the ODB.

43 [for microwave obs only] All cloudy and rainy data. The production of this flag requires the availability of tbcloud@hdr or report_tbcloud@allsky valid info in the ODB.

44 [for microwave obs only] Used Obs cloudy and FG clear. The production of this flag requires the availability of tbcloud@hdr or report_tbcloud@allsky valid info in the ODB.

45 [for microwave obs only] Used Obs clear and FG cloudy. The production of this flag requires the availability of tbcloud@hdr or report_tbcloud@allsky valid info in the ODB.

50 Good ozone data

51 Day time data

52 Night time data

The user has the possibility to request several flags in the same stat block. In case of classical products (**statkind** $\leq$ 7) `OBSTAT` starts by splitting the concerned stat block to several stat blocks (times the number of flags) to end up with one flag for each stat block. In case of gridded statistics (**statkind**= 8) no expansion is needed as the final products will be stored in individual GRIB messages even if they are part of the same GRIB file.

qi1limits [optional] Limits to production of AMVs statistics to a user specified range of QI1. The range should be provided as 2 real values (lower and higher limits) between 0 and 100.

qi3limits [optional] Limits to production of AMVs statistics to a user specified range of QI3. The range should be provided as 2 real values (lower and higher limits) between 0 and 100.

pclear [optional] Limits to production of GEOS radiances to a user specified range of percentage of clear pixels. The range should be provided as 2 real values (lower and higher limits) between 0 and 100.

sizebin (optional) real to define the histogram bin diameter (classical products), Field of View binning or Observation angle binning (for gridded statistics).

refval (optional) real to define the center value to be used with coorditem; the default is zero.

nbbin [optional] integer to define the number of bins. Required when `statkind=2` ; then the bins are symmetric around zero except for wind speed which starts at zero. Optional when `statkind=1` in order to predefine the expected channel number for satellite data.
For satellite data binned by channel, it is mandatory to set this keyword to a value higher than 0 if gridded statistics are to be produced. It is also highly recommanded to set this keyword to the right number of

channels to optimize the computation time. By setting the keyword to a smaller value, `OBSTAT` will perform several deallocation/allocation of statistics arrays each time an increase of channels dimension is detected.

pres_t [optional and effective only with **statkind**=8] allow users to define their own vertical pressure binning. This is performed via the tendem **pres_t** and **pres_b** providing the pressure top and bottom. The number of pressure levels should be the same as **nbbin**. The pressure should be provided in pascal (Pa). There are no restrictions in the definition of pressure layers. Uers can even define overlapping layers. Currently this option is restricted to conventional observations and ozone data (codetype=206).

pres_b [optional and effective only for **statkind**=8] should follow **pres_t**.

channels [optional and effective only for **statkind**=8] Restricts the production of statistics to a subset of channels.

coorditem [optional] integer to define the item that is used for the histogram binning with statkind 1 or 2. Meaningless for bitmaps and gridded statistics. Can be used e.g. to stratify departures against observed value.

coorditem2 [optional] integer to define the item that can be used for the second dimension. The second dimension can be used to produce 2D maps from the ascii files (instead of the recommendant method based on gridded statistics)

sizebin2 [optional but required when **coorditem2** is activated] real to define the binning to be used with coorditem2.

refval2 [optional but required when **coorditem2** is activated] real to define the central value to be used for coorditem2. The default value is zero.

scatwrite [optional and works only with **statkind**=8] allows (**scatwrite**= y) the production scatter statistics.

scatitems [mandatory when **scatwrite**= y] specifies the list of items couples for which `OBSTAT` should provide scatter statistics. Up to 18 pairs of items can be specified. The items are those described in 3.2. If an item is not enabled (e.g. not specified in the ODB SQL request), `OBSTAT` will abort and provide an explanatory message.

scatareas [optional and works only when **scatwrite**= y] Specifies the list of geographical areas for which `OBSTAT` whould produce scatter statistics. The area names to use are those defined in `general.cfg` configuration file. Users can define their own area names by providing their own `general.cfg` via the variable GENERALFILE at the level of `dobstat` script

For each stat block requesting the production of scatter statistics it is mandatory to add, immediately after `END STATDEF`, a certain number of binning blocks to define the necessary binnings for the production of populations. To built the binning blocks it is important to know:

- The binning should be defined for all items involved in the scatter statistics production.
- Each binning block can apply to several items if the user thinks that this makes sense. It is advised to use the same binning for items with the same order of magnitude.

Based on the above principales the number of additional blocks varies from one to up to the number of the different items involved in the scatter statistics production.
Each binning block should start with `BEGIN SCATITEM` and end with `END SCATITEM`. In the same line containing `BEGIN SCATITEM` the user should add the list of items concerned with the binning being defined.

After the line containing `BEGIN SCATITEM`, the user should specify for each channel/level/layer the channel number, the minimum value of the item, the maximum value and the binning value. If the same binning applies to all channels/levels/layers, the user can simplify things by using `9999` before the binning definition. Below, an example of binning block definition:

```
BEGIN SCATITEM 9 10
#bin      min        max        incr
 1         -3.        3.          0.1
 2        -3.        3.          0.1
 3        -5.0       5.0         0.25
 4        -8.0       8.0         0.5
 5       -25.0      25.0         2.5
 6      -140.0     140.0        10.0
END SCATITEM
BEGIN SCATITEM 4
#bin      min        max        incr
9999     -90.0      90.0         2.5
END SCATITEM
```

When writing a new statistics specification, it is useful to refer to the hard statistics printout to check the actual values of observation types, parameters and instruments. Be careful about the ordering of the parameters, it is fixed.

4.8 Examples of statistics definition

In this section, several examples are provided to guide users on how to build the stat definition files

4.8.1 RMS profiles

This example shows the production of RMS profiles (statkind= 1) for AMSU-A radiances from several satellites (produced separately) for ”all” and ”used” data over the Northern Hemisphere, the Tropics and the Southern Hemisphere.

```
BEGIN STATDEF
comment= 'EUMETSAT TOVS-1C metop-a AMSU-A Tb'
statkind= 1
areaNSEW= NH.Tr.SH
types= 54 55
params= 12062
items=  9 10 18
instrument= 403 20603 20903 22303 78403
flagfilter= 2 5
nbbin= 15
END STATDEF
```

4.8.2 Histograms

This example shows the production of Histograms (statkind= 2) for ozone data from two instruments (GOME-2 and SCIAMACHY) for ”used” data over the GLOBE. 40 bins are to be considered with a size bin of 2.

```
BEGIN STATDEF
```

```
comment= 'Ozone'
statkind= 2
areaNSEW= Globe
types= 206
params= 15020
items=  9 10
instrument= 4220001 60175011
flagfilter= 2
sizebin= 2.0
nbbin= 40
END STATDEF
```

4.8.3 Usage charts

This example shows the production of usage charts (statkind= 3) for the U component of AIREP winds over the Globe.

```
BEGIN STATDEF
comment= 'AIREP-Uwind'
statkind= 3
areaNSEW= Globe
types= 142 143 144 145
params= 11003
items=     9    10
instrument=       999
flagfilter= 2
END STATDEF
```

4.8.4 High and low resolution resolution gridded statistics

This example shows the production of high and low resolution gridded statistics (statkind= 8) for radiances from AMSUA on-board several satellites (considered separately) and for several data streams (global, ears and pacrars). Statistics are produced for four flags (all, used, active+passive and clear) and for four quantities (obs, FG and AN departures and bias correction). Statistics are produced over sea and for "all surface types combined" and with a time step of 6 hours. $2°x2°$ is the resolution considered for the high res grid and $360°x2.5°$ for low res grid.

```
BEGIN STATDEF
comment= 'AMSUA Tb'
statkind= 8
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
surfaces= 2 4
tslot= 6
types= 54 55
params= 12062
items=  8 9 10 18
instrument= 20603 20703 20803 20903 22203 78403 403 22303
datastream= 0 1 2
flagfilter= 2 5 40 11
nbbin= 15
END STATDEF
```

4.8.5 Low resolution gridded statistics and scatter statistics only

This example shows the production of low resolution gridded statistics and scatter statistics only. High resolution gridded statistics are disabled. Statistics are for ”used” total column ozone from several instruments.

```
BEGIN STATDEF
comment= 'Total Column ozone instruments'
statkind= 8
highresgrib= n
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
tslot= 6
types= 206
params= 15020
items=  8 9 10 18
instrument= 4220001 785394001 60175011 56207001
flagfilter= 2 5
nbbin= 1
scatwrite= y
scatitems= 8 80  8 90  80 90  80 4  90 4  8 4  8 34  9 4  10 4  9 34  10 34
scatareas= GLOBE N.Hemis S.Hemis
END STATDEF
BEGIN SCATITEM 8 80 90
#bin     min       max       incr
9999      0.0     750.0      25.0
END SCATITEM
BEGIN SCATITEM 9 10
#bin     min       max       incr
9999    -140.0    140.0      10.0
END SCATITEM
BEGIN SCATITEM 4
#bin     min       max       incr
9999    -90.0      90.0       2.5
END SCATITEM
BEGIN SCATITEM 34
#bin     min       max       incr
9999      0.0      90.0       2.5
END SCATITEM
```

4.8.6 User defined pressure layers

This example shows user defined pressure layers.

```
BEGIN STATDEF
comment= 'GOMOS OZONE'
statkind= 8
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
tslot= 6
types= 206
params= 15020
items=  8 9 10 18
instrument= 60172004
flagfilter= 2
nbbin= 12
```

```
pres_t= 10. 20. 40. 70. 110. 170.  260. 390. 610. 1000.  2000.  4000.
pres_b= 20. 40. 70. 110.  170.  260.  390.  610. 1000.  2000. 4000.  6000.
END STATDEF
```

4.8.7 Field of View gridded statistics included

This example shows the production of FOV dependent statistics in addition to high and low res gridded statistics. The FOV bin size is 6 and the number of FOVs bins is 15.

```
BEGIN STATDEF
comment= 'MHS/AMSUB Tb'
statkind= 8
fovgrib= y
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
surfaces= 2 4
types= 54 55
params= 12062
items=  8 9 10 18
instrument= 20604 20704 20804 20904 415 22315
datastream= 0 1 2
flagfilter= 2 5 40
nbbin= 5
sizebin2= 6.
nbbin2= 15
coorditem2= 38
END STATDEF
```

4.8.8 Solar time gridded statistics

This example show the production of solar time dependent statistics in addition to low res gridded statistics.

```
BEGIN STATDEF
comment= 'TMI Tb'
statkind= 8
solartimegrib= y
highresgrib= n
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
tslot= 6
types= 129
params= 12062
items=  8 9 10 18
instrument= 282009
flagfilter= 2 5 40 41 42
nbbin= 6
END STATDEF
```

4.8.9 Low gridded statistics with FOV dimension

The following example shows the production of high and low resolution gridded statistics with activation of the FOV dimension. The fact that coorditem2 is not zero and fovgrib is not set to 'y' means for `OBSTAT` that the FOV dimension should be represented in high and low resolution gridded products.

```
BEGIN STATDEF
comment= 'MWTS Tb'
statkind= 8
alarm= n
areaNSEW= GLOBE
grid= REGULAR 2. 360. 2.5
types= 54 55
params= 12062
items=  8 9 10 18
instrument= 52040
flagfilter= 2 5 40
nbbin= 4
sizebin2= 1.
nbbin2= 15
coorditem2= 38
END STATDEF
```

4.8.10 Targeted sites statistics

The following example shows the production of targeted sites statistics (for 5 sites) in addition to high and low resolution gridded statistics.

```
BEGIN STATDEF
comment= 'SMOS Tb'
statkind= 8
locgrib= y
areaNSEW= GLOBE
grid= REGULAR 0.25
locareas= SMOSmania Lancaster Chase Nemaha Darlingto
surfaces= 1
types=    400
params= 190
items=     8 9 10 18
instrument= 46018
flagfilter=  2
nbbin=  2
sizebin2= 9.
nbbin2= 7
coorditem2= 38
END STATDEF
```

5 Statistics plotting

The present section describes in detail the plotting procedure of OBSTAT statistics. The plotting is done differently for the classical statistics and for gridded statistics. Classical products (RMS plots, histograms and usage charts) are plotted using the same code (with a different configuration) used for the statistics calculation. For gridded statistics a separate plotting package has been developped. This package is considered as part of OBSTAT but it is operating differently.

5.1 Plotting of classical products

The plotting of OBSTAT classical products is performed:

- By the standalone version immediately after the completion of statistics calculation and merging. dobstat is used, in off-line mode, when more customisation of statistics is needed by users.
- By running the fast version of OBSTAT (script dobstat_quick) which relies on the pre-computed statistics. Statistics are retrieved, merged and plotted. dobstat_quick is used to plot statistics produced by all RD prepIFS experiments in addition to OD suites.

As described in 3.1, the two scripts dobstat and dobstat_quick are similar and differs only at the level of settings.
The plotting of the classical OBSTAT statistics offers the following facilities :

- Possibility to superimpose statistics from two different experiments. This is achieved by uncommenting the parameters EXPVER2, CLASS2 and STREAM2. The quick version of OBSTAT supposes that statistics from both experiments are already computed and stored in the default location. If for some reason statistics are stored in a different location (on disk or ECFS) the user has to provide this location using the parameter GRIBDIR.
- Users have the possibility, via the parameter OVERFILE, to plot statistics associated to an experiment against a reference statistics file. This file must have been computed by a previous execution of the STACALC step of dobstat, and its contents must have exactly the same layout as in the current statistics, usually as defined by the STATDEF file, otherwise the program will abort. If necessary, statistics can be added, deleted or moved by hand in the ascii files with a little practice. . . The superimposed statistics produce some extra legend, printouts, and in the plots of vertical profiles the signed difference between the two sets of populations is printed on the left of the normal numbers
- The plotting program plots by default all the statistics present in the ascii file. The result is dozens of postscript pages. If the need is only to plot a subset of of these statistics, the user can set the parameter STATFILTER to yes and provide a STATDEF file containing only the stat blocks of interest. In that case, the plotting program will restrict the plotting to the stat blocks present in the STATDEF. STATFILTER does not work when superimposing results from two experiments.
- For AIRS and IASI, users have the possibility to define the plotting mode by indicating to OBSTAT the vertical coordinate to consider for the plotting of statistics. This is done via the parameter HSRIS_PLOT : 1 for channels peaking pressure, 2 for channels index, 3 for channels number, 4 for channels wavenumber and 5 for channels wavelenght.
- For AIRS and IASI, the parameter PLOTAISRBIAS enable(yes)/disbale(no) the plotting of the bias correction.
- For other observation types, the paramter PLOTBIAS enable(yes)/disbale(no) the plotting of the bias correction.
- The parameter STDEV (yes/no) is used to plot standard deviation vertical profiles instead of rms, which is the normal practice.
- The parameter WHISKERS (yes/no) is used to get st.dev box and min/max whiskers on histogram plots

5.2 Plotting of gridded and scatter statistics

The plotting of gridded and scatter statistics is done by a separate plotting package. It is based on the interpretation of the so-called option files containing the plotting directives. They inform the plotting script about the exact content of statistics files. These files are produced directly by `OBSTAT` during the statistics calculation. They are usuable in their original format but users have a lot of customisation possibilities.
To start the plotting process, the user has to get a copy of the script `obstat_plot` from "/home/mo/obstat/OBSTAT". Few parameters need to be changed inside the script to identify the experiment, period and statistics location. The user need then to get the option files produced by `OBSTAT` during the statistics calculation. These files are stored in the same location as gridded statistics:

- If the parameter GRIBDIR has been set during the statistics calculation, the plotting options are stored under GRIBDIR (on ECFS or local disks).
- If GRIBDIR was not set, the default location for the off-line use of `OBSTAT` is "ec:/$USER/backup/$STREAM/an/$EXPVER"
- For the prepIFS version of `OBSTAT`, the plotting options are stored on ECFS under the location "ec:/$FDBASE/obstat". The variable FDBASE is defined at the level of prepIFS.

Option files are gathered in one tar file that needs to be un-tared. Each stat block (in `STATDEF` after the possible expansions by instrument and data stream) requisting gridded statistics is represented by an option file. As the option file maps the content of gridded statistics it is advised to use the new generated option files each time a change is made at the level of statistics definitions. This avoids to request a data that is may be not present in the statistics anymore.
The arguments of the script `obstat_plot` are:

EXPVER the experiment identifier (4-character string)

CLASS the experiment class. Must be `rd` for prepIFS experiments and `od` for operational suites. Users need to use `e4` for ERA-40 and `ei` for ERA-interim.

STREAM defines the analysis stream: DA and DCDA are supported.

EXPVER2 [optional] defines the ID of the experiment to compare against. It should be commented if not comparison is needed.

CLASS2 [optional] defines the class of the experiment to compare against. It should be commented if not comparison is needed.

STREAM2 [optional] defines the stream of the experiment to compare against. It should be commented if not comparison is needed.

LASTTIME defines the end of the period to plot. The format to be used is `yyyymmddhh`

GEO_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of geographical maps of statistics. The default value is 30 days.

HIST_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting area averages time series. The default value is 30 days.

HOV_period [optional] defines the period back in time (from LASTTIME) to be used for the plotting of Hovmoeller diagrams. The default value is the value of HIST_PERIOD.

HOV_LEV_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of channels vs time Hovmoeller diagrams. The default value is the value of 30 days

FOV_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of Scan dependent statistics. The default value is 30 days.

SCAT_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of scatter statistics. The default value is 30 days.

OVIEW_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of overview plots. The default value is the value of HIST_PERIOD.

HIST_SOLT_PERIOD [optional] defines the period back in time (from LASTTIME) to be used for the plotting of solar time dependent statistics. The default value is 10 days.

PLOT_OPTIONS is a mandatory parameter defining the list of plotting option files.

PLOTDIR is a mandatory parameter defining where the option files are located.

type defines the code type name. It is needed only for conventional observations. Typical values are: synop, airep, dribu, temp, metar, pilot. For the rest see the file `bufrodbcodes.cfg`.

USER [optional] defines the user owning the statistics. This is necessary to use other's statistics.

OBSMON_SCRATCH [opional] defines where to run the plotting.

PLOTDIR [optional] defines where to store the produced plots. The default location is OBSMON_SCRATCH/plots_$EX

GRIBDIR define the location of statistics. This should be consistent with the value used during the statistics calculation. It is commented by default.

GRIBDIR2 [optional] defines the location of statistics of the second experiments.

HIST_TEMPLATES [optional] defines the templates for area averages time series plotting: number and content of panels.

SCALINGS_FILE [optional] defines the file containing the scales to be used for the production of plots in case the autoscaling is off.

COLOURS_FILE [option] defines the number and list of colours to be used for the production of geo maps and Hovmoeller products.

The parameters described above defines only the global settings for the plotting. The list of products to be plotted is specified at the level of the plotting option files.

5.2.1 List of products

A wide range of products can be produced based on gridded statistics. The list of products is:

- Geographical maps of statistics.
- Area averages of time series.
- Latitude versus time hovmoeller diagrams
- Channels versus time hovmoeller diagrams

- Time versus longitude hovmoeller diagrams
- vertical levels versus latitude hovmoeller diagrams
- Scatter plots
- Area averaged overview time series. The Overview product contains three separate plots: bias+/-stdev, stdev and data counts.
- Area and time averages of scan depended statistics
- Area averaged solar-time time series
- Targeted sites time series

5.2.2 Plotting option files

The products above and their content is fully prescribed by the plotting options. As mentioned above, the plotting option files are produced by OBSTAT during the statistics calculation. However, the raw files contain only generic plots (e.g only Latitude versus time Hovmoeller). Users are invited to edit the option file and customise them. Below a full description of the content of the plotting option files. Not all the parameters listed below are present by default in the raw option files and it is up to the users to add them according to their needs. An option file is an ascii file containing the initialisation of unix variables (based on the Korn shell standards). Before the user go through the description of the plotting directives, it is important to stress that OBSTAT produces a file ready to use. The modification of an option file is only needed to customise the products.

MODE [mandatory] present by default and should not be modified. It is part of the GRIB file naming and therefore it is used to locate the statistics file.

SATIDS [mandatory] present by default and it defines the list of satellites. This variable is set using the command ”set -A SATIDS” to use it as an array. By default there is only one satellite in the list but users can add other satellites if the associated statistics exist. For conventional observations, the value of SATIDS is 999.

SATIDS2 [optional] used in case of comparison with another dataset from another satellite(s). By default SATIDS2 is set to SATIDS if the comparison is needed but the satellite involved is the same (e.g. comparison of the same data set from two diferent experiments). SATIDS and SATIDS2 should have the same dimension.

SENSORS [mandatory] present by default and it defines the list of instruments. This variable is set using the command ”set -A SENSORS” to use it as an array. By default there is only one instrument in the list but users can add other instruments if the associated statistics exist. For conventional observations, the value of SENSORS is 999. The total number of instruments should be the same as the total number of satellites.

SENSORS2 [optional] used in case of comparison with another dataset from another instrument(s). by default SENSORS2 is set to SENSORS if the comparison is needed but the instruments involved are tye same (e.g. comparison of the same data set from two diferent experiments). SENSORS and SENSORS2 should have the same dimension.

DATASTREAM [mandatory when statistics are produced by data stream] set by OBSTAT to indicate the data reception streams considered during the statistics calculation. Users are free to remove streams if there is no interest to plot statistics from.

NSELECT [mandatory] present by default and it defines the number of channel/levels/layers/heights for which `OBSTAT` generated statistics based on the statistics definition in `STATDEF`. Users are free to change (lower it only) this variable to plot a subset of channels.

ISEL_CHAN [mandatory] present by default and it defines the list of channel/levels/layers/heights for which `OBSTAT` generated statistics based on the statistics definition in `STATDEF`. Users are free to change (remove only) this variable to plot a subset of channels. The number of channels should be the same as NSELECT.

ISEL_CHAN2 [optional] used in case the user is interested by comparing statistics from different channels (from the same or different source). by default ISEL_CHAN2 is set to ISEL_CHAN if the comparison is needed but using the same channels/levels. ISEL_CHAN and ISEL_CHAN2 should have the same dimension.

PRES_T [mandatory if the vertical dimension is pressute] set by `OBSTAT` to indicate to the plotting program the top of the pressure layer/level. This variable is set using the command "set -A SENSORS" to use it as an array. If present, this variable should not be removed or modified.

PRES_B [mandatory if the vertical dimension is pressute] set by `OBSTAT` to indicate to the plotting program the bottom of the pressure layer/level. This variable is set using the command "set -A SENSORS" to use it as an array. If present, this variable should not be removed or modified.

OVIEW_SEL_CHAN [mandatory only when overview statistics are requested] Used when overview statistics are requested (see L_OVIEW or L_SCAN). In sepecifies the list of channels to include in the overview. Chennels from different sources (satellites or instruments) are separated by -99. When OVIEW_SEL_CHAN is present, NSELECT and ISEL_CHAN are not needed but there is no harm in keeping them.

FOVSIZE [mandatory if the FOV dimension is present] set by `OBSTAT` and should not be modified by users. It indicates to the plotting package the value of the sizebin used for the production of statistics when the FOV dimension is enabled.

ISEL_FOV [mandatory if the FOV dimension is present] set by `OBSTAT` to indicate to the plotting package the list of FOVs present in the statistics. Users are free to change (remove only) it if the interest is to plot only a subset of statistics.

CHANNELS [optional] if set to 'all_channels' the plotting program will produce statistics for each level separately but all plots will be in the same postscript file. This is usefull for files managment.

GROUPSTAT [optional] if set to 'yes' the plotting package will merge (in a consistent way) statistics from the provided list of instruments, sensors and datastreams. A further merging of statistic from all channels depends on the value of CHANNELS above. The combination of CHANNELS='all_channels' and GROUPSTAT='yes' means that only one plot per product type is to be produced. This has a particular interest when generating statistics for parameters like the forecast or analysis sensitivity to observations. This option allows to have an overview of the behaviour of a group of data sets. Details can be requested by disabling GROUPSTAT

ORBITSTAT [optional] used to plot separately (=true) statistics for ascending/descending parts of the orbit. This parameter has sense only when statsitics were produce separately for ascending/descending nodes. ORBITSTAT=false means that the need is for the full orbit (a merging of statistics is performed by the plotting package).

TSLOT_H [mandatory] set by `OBSTAT` to indicate the time step used for the producion of statistics. Users should not modify this parameter.

L_HIST [optional] enable(true)/disbale(false) the plotting of area averages time series. If low gridded statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_HOV [optional] enable(true)/disbale(false) the plotting of Hovmoeller diagrams. If low gridded statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_GEO [optional] enable(true)/disbale(false) the plotting of geographical maps of statistics. If high resolution gridded statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_SCAT [optional] enable(true)/disbale(false) the plotting of scatter statistics. If scatter statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_LOC [optional] enable(true)/disbale(false) the plotting of targeted sites statistics. If targeted sites statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_SOLT [optional] enable(true)/disbale(false) the plotting of solar time statistics. If solar time statistics are not present, OBSTAT will not put this keyword in the option file. In that case it is useless to add it manually.

L_OVIEW [optional] enable(true)/disbale(false) the plotting of overview statistics. This keyword is not set by OBSTAT but requires the presence of low resolution gridded statistics. With this option on the list of instruments in SENSORS are all plotted in the same overview plot.

L_SCAN [optional] enable(true)/disbale(false) the plotting of overview scan dependent statistics. This keyword is not set by OBSTAT but requires the presence of low resolution gridded statistics. With this option on the list of instruments in SENSORS are all plotted in the same overview plot.

HIST_CHOICE [mandatory] indicates to the plotting program, the template to be used for the plotting of area averages timeseries: number and content of panels to plot in addition to the specification of each curve attributes (colour, line style, legend, ...). The definition of templates is present the file `hist_templates.sh` maintained by OBSTAT admin. However the users can provide their customised version via the variable HIST_TEMPLATES at the level of the plotting script (see 5.2).

NAREA [mandatory] Number of geographical areas defined. By default NAREA is 1 (globe). The areas defined can be used by optionnaly used by the different products as described below. Each area is defined by its position within the definition list.

LON_WEST [mandatory] provides the west longitude for all the areas in the list. LON_WEST is a string with quotes. The list of values should be delimited by a comma.

LON_EAST [mandatory] provides the east longitude for all the areas in the list. LON_EAST is a string with quotes. The list of values should be delimited by a comma.

LAT_NORTH [mandatory] provides the north latitude for all the areas in the list. LAT_NORTH is a string with quotes. The list of values should be delimited by a comma.

LAT_SOUTH [mandatory] provides the west longitude for all the areas in the list. LAT_SOUTH is a string with quotes. The list of values should be delimited by a comma.

AREANAMES [optiona] the user can different her/his own names of the defined areas.

ISEL_AREA [mandatory] provides the list of areas for which area averages time series are to be produced. by default is 1

HOV_AREA [optional] provides the list of areas to consider for the production of 'time vs longitude' and 'level vs time' Hovmoeller diagrams.

GEO_AREA [optional] provides the list of areas to consider for the production of geographical maps of statistics.

OVIEW_SEL_AREA [mandatory with L_OVIEW and L_SCAN] provides the list of areas to consider for the production of overview plots.

NSCATAREA [mandatory when L_SCAT is enabled] set by `OBSTAT` to indicate to the plotting package the number of geographical areas considered for the production of scatter statistics.

ISEL_SCATAREA [mandatory when L_SCAT is enabled] set by `OBSTAT` to indicate to the plotting package the explicit list of geographical areas names considered for the production of scatter statistics and for which the scatter plots are required. This list is independent of the areas definition done for the other products.

AREAS [optional] if set to 'all_areas' tells the plotting package to put all related areas plots in one postscript file (although the plots are produced separately for each area). This is usefull for files managment.

NSURF [mandaroy] set by `OBSTAT` to indicate to the plotting package, the number of land-sea masks considered for the production of statistics.

ISEL_SURF [mandatory] set by `OBSTAT` to indicate to the plotting package, the list of land-sea masks considered for the production of statistics. Users can change this by removing the surface masks unwanted for the plotting. This variable is used only for the plotting of area averages time series.

OVIEW_SEL_SURF [mandatory for L_OVIEW and L_SCAN] specifies the list of surface types (per channel) to be used for the production of overview plots. Surfaces for different sources (satellites or instruments) are separated by -99

SURFACES [optional] allows users to put all land-sea mask dependent plots (per plot type) in one postscript file (The plots are still separate inside the multi-pages postscript).

NFLAGHIST [mandatory] set by `OBSTAT` to indicate to the plotting package, the number of flags for which low resolution gridded statistics have been produced and for which area averages time series will be produced.

ISEL_FLAGHIST [mandatory] set by `OBSTAT` to indicate to the plotting package, the list of flags for which low resolution gridded statistics have been produced and for which area averages time series will be produced. Users have the possibility to remove items from the list if the interest is for a subset of flags.

NFLAGHOV [mandatory] set by `OBSTAT` to indicate to the plotting package, the number of flags for which low resolution gridded statistics have been produced and for which hovmoeller diagrams are to be produced.

ISEL_FLAGHOV set by `OBSTAT` to indicate to the plotting package, the list of flags for which low resolution gridded statistics have been produced and for which hovmoeller diagrams are to be produced. Users have the possibility to remove items from the list if the interest is for a subset of flags.

NFLAGGEO [mandatory] set by `OBSTAT` to indicate to the plotting package, the number of flags for which high resolution gridded statistics have been produced and for which geo maps are to be produced.

ISEL_FLAGGEO set by OBSTAT to indicate to the plotting package, the list of flags for which high resolution gridded statistics have been produced and for which geo maps are to be produced. Users have the possibility to remove items from the list if the interest is for a subset of flags.

OVIEW_SEL_FLAG [mandatory when L_OVIEW and L_SCAN] specify the list of flags (one per channel) to consider for the plotting of overview plots. Flags for different sources (satellites or instruments) are separated by -99

NFLAGSCAT set by OBSTAT to indicate to the plotting package, the number of flags for which scatter statistics have been produced and for which scatter plots are to be produced.

SCAT_FLAGS set by OBSTAT to indicate to the plotting package, the list of flags for which scatter statistics have been produced and for which scatter plots are to be produced. Users have the possibility to remove items from the list if the interest is for a subset of flags.

HOVMODE [mandatory] tells the plotting programs what type of hovmoeller to produce. Four options are available:lat_time (default), lev_time, time_lon and lev_lat.

LHOV_NUMBER, LHOV_OBS, LHOV_OBS_RMS, LHOV_FG_DEP,.. when producing the statistics, OBSTAT specifies here the list of diagnostics available for hovmoeller plotting. By default the plotting is enabled for all of them but the user can decide to siwtch off the plotting of one or several diagnostics. The average and stdev are considered separately.

HOV_FIELDS [optional] put plots associated to all diagnostics in one multipages postscript file.

LGEO_NUMBER, LGEO_OBS, LGEO_OBS_RMS, LGEO_FG_DEP,.. when producing the statistics, OBSTAT specifies here the list of diagnostics available for geo maps plotting. By default the plotting is enabled for all of them but the user can decide to siwtch off the plotting of one or several diagnostics. The average and stdev are considered separately.

GEO_FIELDS [optional] put plots associated to all diagnostics in one multipages postscript file.

GEO_HOURS [optional] allows the plotting of a subset of time slots per day. This is useful for instance to evaluate the diurnal effect. By default GEO_HOURS='all'.

NSCATITEMS [mandatory when L_SCAT is enabled] tells to the plotting package the number of item couples for which scatter statistics have been generated and for which we desire to produce scatter plots.

ISEL_SCATITEMS_X [mandatory when L_SCAT is enabled] lists the X items used for the production of scatter statistics. The number listed here should be equal to NSCATITEMS. The user can freely remove items from the list. If so this should be done also for ISEL_SCATITEMS_Y

ISEL_SCATITEMS_Y [mandatory when L_SCAT is enabled] lists the Y items used for the production of scatter statistics. The number listed here should be equal to NSCATITEMS

SCAT_ITEMS [optional] put plots associated to all scat items in one multipages postscript file.

PLOT_CORR_XY [optional] adds correlation info in the scatter plots for each scat item

PLOT_MEAN_X [optional] plots a line showing the mean for X values

PLOT_MEAN_Y [optional] plots a line showing the mean for Y values

LGRIDSHADING [optional] tells the plotting package to use the grid shading for the plotting of countours by magics++

SMOOTH [optional] tells the plotting package to smooth high resolution statistics if they are noisy. Currently the smooting method is fixed and hard coded.

AUTOSCALE [optional] enables(true)/disbales(false) the autoscaling. By the default the autoscaling is on which is very usefull to get an idea about the scales when preparing the scaling files.

PLOT_OUTL [optional] when the autoscaling is off and user defined scales are to be used, the user can decide to show(true)/hide(false) the outliers. By default PLOT_OUTL is enabled.

PLOT_TITLE [optional] enable(true)/disbale(false) the plotting of plot titles. PLOT_TITLE is enabled by default.

UNIT [optional] allows users to provide the parameter unit to used for the plotting.

title1 [optional] allows users to provide the first line of the title.

title2 [optional] allows users to provide the second line of the title.

title3 [optional] allows users to provide the third line of the title.

title4 [optional] allows users to provide the fourth line of the title.

5.2.3 Area averages timeseries templates

Area averages time series are a valuable tool to appreciate the evolution in time of statistics. The parameters to show in such product depends on each data type and also on user needs. The plotting package offers a generic way to customise this product via the use of an ascii configuration file called `hist_templates`. This file contains a pre-defined templates. Each template contains the definition of the number of pannels, content of each pannel an the plotting attributes for each curve. A template is used by simply using its name inside the plotting option. The variable HIST_TEMPLATES in `obstat_plot` allow users to define their own templates or just modify the existing ones. An example of a template config file is:

```
#----------------------- Definition of hist plot templates ------------------------------
# The user can use the pre-defined templates or make use of them after modification
# The user can also define a new template
# CX_NBC : number of curves for panel X
# CX_VAR : list of variables to plot for panel X
# (choose among : fgdep, andep, fgdep_bcor, andep_bcor, fgdep_stdv, andep_stdv, fg, an, obs,
#   obs_stdv, all, used, passive, active)
# CX_TIT : list of titles for variables to plot in panel X
# CX_CLR : list of culours to be used to plot variables in panel X
# CX_STY : list of styles to be used to plot variables in panel X
#-----------------------------------------------------------------------------------------
C1_NBC=1
C2_NBC=1
C3_NBC=1
C4_NBC=1
C1_VAR=""
C1_TIT=""
C1_CLR=""
C1_STY=""
C2_VAR=""
C2_TIT=""
C2_CLR=""
C2_STY=""
C3_VAR=""
C3_TIT=""
C3_CLR=""
```

```
C3_STY=""
C4_VAR=""
C4_TIT=""
C4_CLR=""
C4_STY=""
case $HIST_CHOICE in
  'obs_fg_an_bcor')

     NBPANELS=4

     C1_NBC=4
     C1_VAR="'fgdep', 'andep', 'fgdep_bcor', 'andep_bcor'"
     C1_TIT="'OBS-FG', 'OBS-AN', 'OBS-FG(bcor)', 'OBS-AN(bcor)'"
     C1_CLR="'BLUE', 'RED', 'BLUE', 'RED'"
     C1_STY="'SOLID', 'SOLID', 'DOT', 'DOT'"

     C2_NBC=2
     C2_VAR="'fgdep_stdv', 'andep_stdv'"
     C2_TIT="'stdv(OBS-FG)', 'stdv(OBS-AN)'"
     C2_CLR="'BLUE', 'RED'"
     C2_STY="'SOLID', 'SOLID'"

     C3_NBC=3
     C3_VAR="'obs', 'fg', 'an'"
     C3_TIT="'OBS', 'FG', 'ANA'"
     C3_CLR="'GREEN', 'BLUE', 'RED'"
     C3_STY="'SOLID', 'SOLID', 'SOLID'"

     C4_NBC=3
     C4_VAR="'count_displayed', 'count_all', 'count_used'"
     C4_TIT="'n_displayed', 'n_all', 'n_used'"
     C4_CLR="'MAGENTA', 'BLUE', 'KELLY_GREEN'"
     C4_STY="'SOLID', 'DASH', 'DOT'"
     ;;
  'fso')

     NBPANELS=3

     C1_NBC=1
     C1_VAR="'fso'"
     C1_TIT="'FSO'"
     C1_CLR="'BLUE'"
     C1_STY="'SOLID'"

     C2_NBC=2
     C2_VAR="'fgdep', 'andep'"
     C2_TIT="'OBS-FG', 'OBS-AN'"
     C2_CLR="'BLUE', 'RED'"
     C2_STY="'SOLID', 'SOLID'"

     C3_NBC=1
     C3_VAR="'count_used'"
     C3_TIT="'n_used'"
     C3_CLR="'BLUE'"
     C3_STY="'SOLID'"
     ;;
  *)

     NBPANELS=4

     C1_NBC=4
     C1_VAR="'fgdep', 'andep', 'fgdep_bcor', 'andep_bcor'"
```

```
    C1_TIT="'OBS-FG', 'OBS-AN', 'OBS-FG(bcor)', 'OBS-AN(bcor)'"
    C1_CLR="'BLUE', 'RED', 'BLUE', 'RED'"
    C1_STY="'SOLID', 'SOLID', 'DOT', 'DOT'"

    C2_NBC=2
    2_VAR="'fgdep_stdv', 'andep_stdv'"
    C2_TIT="'stdv(OBS-FG)', 'stdv(OBS-AN)'"
    C2_CLR="'BLUE', 'RED'"
    C2_STY="'SOLID', 'SOLID'"

    C3_NBC=3
    C3_VAR="'obs', 'fg', 'an'"
    C3_TIT="'OBS', 'FG', 'ANA'"
    C3_CLR="'GREEN', 'BLUE', 'RED'"
    C3_STY="'SOLID', 'SOLID', 'SOLID'"

    C4_NBC=3
    C4_VAR="'count_displayed', 'count_all', 'count_used'"
    C4_TIT="'n_displayed', 'n_all', 'n_used'"
    C4_CLR="'MAGENTA', 'BLUE', 'KELLY_GREEN'"
    C4_STY="'SOLID', 'DASH', 'DOT'"
    ;;
esac
```

The result of the template 'fso' is presented in Fig. 10.

5.2.4 *Plots scaling*

OBSTAT offers several ways to handle the scaling used for the different plots:

- Automatic scaling based on the distribution of 90% of values to plot. Good resuls are obtained but it is not suitable for plots comparison. This option is activated by setting AUTOSCALE=true in the plotting option file.

- The user can disable the autoscaling and instead provides a user defined scaling. This is done via an ascii file. OBSTAT admin is maintaining a centralized version containing scales adapted to almost data types. Users are invited to take a copy of this file (`scaling.cfg`) and customize it. The parameter SCALINGS_FILE in `obstat_plot` indicates to OBSTAT to use the customized file. User have also the possibility to customize the scaling differently for the three main products : area averages time series, Hovmoeller and geo maps.

- If the autoscaling is disabled and no scaling is supplied by the user or found in the centralized file, OBSTAT builds a scaling based on the minimum and maximum values of the data. The result is not very diffrent than the autoscaling.

The scaling definition file provides for all channels associated to an instrument, the minimum value of the scale, the scale step and the scaling factor (always equal to 1) for the most used diagnostics. When the minimum value is set to -99, OBSTAT starts from the real minimum value. Diagnostics considered are:

- Observation averages. The scaling defined applies to FG and AN averages

- FG and AN departures. One scaling is defined for both quantities

- Bias correction.

Figure 10: Example of use of hist templates

- Standard deviation of Obs values, FG and AN values. One scaling is defined for the three quantities.
- Standard deviation of FG and AN departures. One scaling is defined for both quantities
- Standard deviation of bias correction.
- data counts.

Each line of the file starts with the observation identifier composed of a minimum of 7 digits: The first three digits represents the satellite ID, The sensor ID is represented by the following three digits. The channel number follows. The scaling definition for each diagnostic is delimited by vertical bars (one on the left and the other one on the right). The vertical bars should be separated to the digits by at least one space character
An example of the scaling definition file is:

```
#Inst         Obs          fgdep        bcor       Obstdv     fgdStdv    BcorSTDV    Count    Unit
#SMOS
 0460181 | 050. 12. 1 | -20 2. 1 | -99 1.5 1 | 0. 1.4 1 | 0. 1.4 1 | 0. 1.5 1 | 0 20 1 | K
 0460182 | 050. 12. 1 | -20 2. 1 | -99 2.  1 | 0. 1.4 1 | 0. 1.4 1 | 0. 2.  1 | 0 20 1 | K
#NOAA-19/MHS
 2230151 | 100. 10. 1 | -99 .5 1 | -99 1.5 1 | 0. 2. 1  | 0. .3 1  | 0. 1.5 1 | 0 20 1 | K
 2230152 | 100. 10. 1 | -99 .5 1 | -99 2.  1 | 0. 2. 1  | 0. .3 1  | 0. 2.  1 | 0 20 1 | K
 2230153 | 100. 10. 1 | -99 .5 1 | -99 1.5 1 | 0. 2. 1  | 0. .3 1  | 0. 1.5 1 | 0 20 1 | K
 2230154 | 100. 10. 1 | -99 .5 1 | -99 1.5 1 | 0. 2. 1  | 0. .3 1  | 0. 1.5 1 | 0 20 1 | K
 2230155 | 100. 10. 1 | -99 .5 1 | -99 1.5 1 | 0. 2. 1  | 0. .3 1  | 0. 1.5 1 | 0 20 1 | K
```

5.2.5 *Colours*

The plotting package uses a set of 22 colours to produce geo maps and hovmoeller diagrams. Users have the possibility to choose their own colours via a configuration file. The user can get a copy of the default file `colours_22.cfg` and customize it by setting the number of colours (and consequently the number of scales) and the list of colours. Colours can be sepcified using the colour names of Magics++ or simply the RGB codes. The parameter COLOURS_FILE in the script `obstat_plot` is the one to be used to make the user defined colours file effective.
An example of the colours definition file is:

```
# NCOL is the number of colours to be used effectively (Maximum 22 colours)
# COLOURS is the explicit list of colours to be used.
NCOL=15
COLOURS="'LAVENDER','NAVY','BLUE','GREENISH_BLUE','SKY','CYAN','BLUISH_GREEN','YELLOW',
'ORANGE_YELLOW','YELLOWISH_ORANGE','ORANGE','REDDISH_ORANGE','RED_ORANGE','RED','BRICK'"
```

or using the RGB concept:

```
# NCOL is the number of colours to be used effectively (Maximum 22 colours)
# COLOURS is the explicit list of colours to be used.
NCOL=22
 COLOURS="'RGB(0.45,0.64,1.)','RGB(0.47,0.80,1.)','RGB(0.49,0.95,1.)',
'RGB(0.51,1.,0.92)','RGB(0.53,1.,0.8)','RGB(0.55,1.,0.68)', 'RGB(0.57,1.,0.58)',
'RGB(0.69,1.,0.59)','RGB(0.82,1.,0.61)', 'RGB(0.93,1.,0.63)','RGB(  1.,0.97,0.65)',
'RGB(  1.,0.89,0.67)', 'RGB(1.,0.82,0.57)','RGB(1.,0.75,0.47)','RGB(0.99,0.67,0.37)',
'RGB(0.98,0.58,0.27)','RGB(0.98,0.47,0.17)','RGB(0.97,0.36,0.07)','RGB(0.91,0.26,0.03)',
'RGB(0.81,0.19,0.03)','RGB(0.70,0.13,0.03)', 'RGB(0.6,0.08,0.03)'"
```

5.2.6 *Statistics comparison*

One of the new features of the plotting package is the complete freedom to compare statistics. The software is able the compare two datasets and it is up to the user to define the datasets to compare. It also the responsibility

of the user to compare what is physically comparable.
A dataset is defined by :

- Experiment
- Satellite
- Sensor
- Data stream
- Channel

The user is free to define the two datasets. The only condition is to have statistics already computed for the two datasets to compare. The user is referred to the description of `obstat_plot` and plotting options parameters to see how to configure the plotting script for comparison purposes.

5.3 Examples of gridded statistis plotting

In this section several examples are provided: option files are presented and followed by examples of plots.

5.3.1 Example 1 (area averages timeseries, Hovmoeller and geo maps)

This example shows the plotting of:

- Area average time series of 6 'used' ozone channels from SBUV-2 on-board several satellites. Six geographical areas are considered.
- Latitude vs time Hovmoeller diagram. Only Obs values, Data counts, FG departures and Stdev of FG departures are plotted for "used" data.
- Geographical maps of statistics. Only Data counts, FG departures and Stdev of FG departures are plotted for "used" data.
- Scatter plots for several scat items. Only used data over two areas are plotted.

The plotting option file is:

```
#
 # Generic Plotting options file
 #
 #====================================================================

 #====================================================================
 #   Define mode, satellites and sensors
 #   to monitor.
 #====================================================================
 MODE=o3
 set -A SATIDS   207 208 209 223
 set -A SENSORS  624 624 624 624
 #--------------------------------------------------------------------
 #  Define channels/levels to monitor
 #  (namelist sel_lev)
 #  only activate the DATASTREAM when it's available
```

```
#-------------------------------------------------------------------
NSELECT=6
ISEL_CHAN='1    2    3    4    5    6'
set -A PRES_T     10    100    160    410    640   1600
set -A PRES_B    100    160    410    640   1600 101325
#-------------------------------------------------------------------
#  Define time averaging parameters
#  (namelist time_average)
#  TSLOT_H should be equal to statistics time step (in the GRIB file)
#-------------------------------------------------------------------
TSLOT_H=6

#-------------------------------------------------------------------
#  Define monitoring files to output
#-------------------------------------------------------------------
L_GEO=true
L_HOV=true
L_HIST=true
L_SCAT=true

#-------------------------------------------------------------------
#  Define output for history files
#  variables HIST_CHOICE to customize the output
#  select HIST_CHOICE from obs_fg/obs_fg_an/obs_fg_bcor/obs_fg_an_bcor
#  or customize a new one in $SCRIPTDIR/hist_templates.sh
#  Default is obs_fg_an_bcor (see $SCRIPTDIR/hist_templates.sh)
#-------------------------------------------------------------------
HIST_CHOICE='obs_fg_an_bcor'
#-------------------------------------------------------------------
#  areas (namelist history_area)
#-------------------------------------------------------------------
NAREA=6
LON_WEST='0. , 0. ,  0. ,  0. ,  0. ,  0.'
LON_EAST='360. , 360. , 360. , 360. , 360. , 360.'
LAT_NORTH='90. , 90. , 70. , 20. ,-20. , -70.'
LAT_SOUTH='-90. , 70. , 20. , -20. , -70. , -90.'
ISEL_AREA='1 2 3 4 5 6'
AREAS='all_areas'


#-------------------------------------------------------------------
#  surface types (namelist surface_type)
#  1=land, 2=sea, 3=sea-ice, 4=all
#-------------------------------------------------------------------
NSURF=1
ISEL_SURF='4'

#-------------------------------------------------------------------
#  usage flags (namelist sel_flag)
#-------------------------------------------------------------------
NFLAGHIST=1
ISEL_FLAGHIST='5'
HIST_FLAGS='all_flags'

#-------------------------------------------------------------------
#  Define output for geo mean files
#-------------------------------------------------------------------
LGEO_NUMBER=true
LGEO_OBS=false
LGEO_FG_DEP=true
LGEO_AN_DEP=false
LGEO_BCOR=false
```

```
LGEO_OBS_RMS=false
LGEO_FG_RMS=true
LGEO_AN_RMS=false
LGEO_BCOR_RMS=false
GEO_HOURS='all'
GEO_FIELDS='all_fields'
NFLAGGEO=1
ISEL_FLAGGEO='5'
GEO_FLAGS='all_flags'
#--------------------------------------------------------------------
#  Define output for hovmoeller files
#--------------------------------------------------------------------
LHOV_NUMBER=true
LHOV_OBS=true
LHOV_FG_DEP=true
LHOV_AN_DEP=false
LHOV_BCOR=false
LHOV_OBS_RMS=false
LHOV_FG_RMS=true
LHOV_AN_RMS=false
LHOV_BCOR_RMS=false
HOV_FIELDS='all_fields'
NFLAGHOV=1
ISEL_FLAGHOV='5'
HOV_FLAGS='all_flags'
HOVMODE=lat_time
AUTOSCALE=true

 #--------------------------------------------------------------------
#  Define output for scatter files
#--------------------------------------------------------------------
NSCATITEMS=9
ISEL_SCATITEMS_X=' 8 8  80 8  8 9 10  9 10'
ISEL_SCATITEMS_Y='80 90 90 4 34 4  4 34 34'
SCAT_ITEMS='all_scats'
PLOT_CORR_XY='T T T F F F F F F'
PLOT_MEAN_X='T T T T T T T T T'
PLOT_MEAN_Y='T T T T T T T T T'
NFLAGSCAT=1
ISEL_FLAGSCAT='5'
SCAT_FLAGS='all_flags'
NSCATAREA=2
ISEL_SCATAREA='GLOBE N.Hemis'
SCAT_AREAS='all_areas'

LGRIDSHADING=true
SMOOTH=false

UNIT='[DU]'
```

A subset of the plots produced is presented in the following figures 11, 12, 13 and 14

5.3.2 *Example 2 (Channels versus time Hovmoeller diagram)*

The following example shows the plotting of impact heights versus time Hovmoeller for normalised obs values and departures for METOP-A/GRAS Setting RO. Data passing the first guess check are plotted.
The plotting option file is:

```
#
```

```
# Generic Plotting options file
#
#=====================================================================

#=====================================================================
#   Define mode, satellites and sensors
#   to monitor.
#=====================================================================
MODE=gpsroS
set -A SATIDS      4
set -A SENSORS   202
#---------------------------------------------------------------------
#  Define channels/levels to monitor
#  (namelist sel_lev)
#  only activate the DATASTREAM when it's available
#---------------------------------------------------------------------
NSELECT=50
ISEL_CHAN='1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
  26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50'
#---------------------------------------------------------------------
#  Define time averaging parameters
#  (namelist time_average)
#  TSLOT_H should be equal to statistics time step (in the GRIB file)
#---------------------------------------------------------------------
TSLOT_H=6

#---------------------------------------------------------------------
#  Define monitoring files to output
#---------------------------------------------------------------------
L_HOV=true

#---------------------------------------------------------------------
#  Define output for history files
#  variables HIST_CHOICE to customize the output
#  select HIST_CHOICE from obs_fg/obs_fg_an/obs_fg_bcor/obs_fg_an_bcor
#  or customize a new one in $SCRIPTDIR/hist_templates.sh
#  Default is obs_fg_an_bcor (see $SCRIPTDIR/hist_templates.sh)
#---------------------------------------------------------------------
HIST_CHOICE='obs_fg_an_dep_stdv'
#---------------------------------------------------------------------
#  areas (namelist history_area)
#---------------------------------------------------------------------
NAREA=1
LON_WEST='0.'
LON_EAST='360.'
LAT_NORTH='90.'
LAT_SOUTH='-90.'
ISEL_AREA='1'

#---------------------------------------------------------------------
#  surface types (namelist surface_type)
#  1=land, 2=sea, 3=sea-ice, 4=all
#---------------------------------------------------------------------
NSURF=1
ISEL_SURF='4'

#---------------------------------------------------------------------
#  Define output for hovmoeller files
#---------------------------------------------------------------------
LHOV_NUMBER=true
LHOV_OBS=true
LHOV_OBS_RMS=true
```

```
LHOV_FG_DEP=true
LHOV_FG_RMS=true
LHOV_AN_DEP=true
LHOV_AN_RMS=true
HOV_FIELDS='all_fields'
NFLAGHOV=1
ISEL_FLAGHOV='9'
HOVMODE=lev_time
AUTOSCALE=false
LGRIDSHADING=true
```

The associated plot is presented in Fig. 15.

5.3.3 Example 3 (Levels vs latitude and Time vs longitude Hovmoeller)

This example shows two Hovmoeller diagrams: Levels versus latitudes (see Fig. 16) and time versus longitude (see Fig. 17). The option file is similar to the previous ones. The difference is for the HOVMODE variable which is set to lev_lat for the first diagram and time_lon for the second diagram.

5.3.4 Example 4 (Statistics merging)

This example shows how to group statistics from all AMSU-A channels from several satellite. This is particularily usefull to evaluate the impact of AMSU-A observations on the reduction of the Forecast errors. Grouping statistics for this case has sense because we group quantities with the same order and meaning.
The plotting option used here produces area averages timeseries, latitude versus time hovmoeller and geographical map of statistics.

```
#
# Generic Plotting options file
#
#=====================================================================

#=====================================================================
#   Define mode, satellites and sensors
#   to monitor.
#=====================================================================
MODE=tb
set -A SATIDS   4 206 207 209 222 223
set -A SENSORS  3   3   3   3   3   3
#---------------------------------------------------------------------
#  Define channels/levels to monitor
#  (namelist sel_lev)
#  only activate the DATASTREAM when it's available
#---------------------------------------------------------------------
NSELECT=9
ISEL_CHAN='5    6    8    9   10   11   12   13   14'
CHANNELS='all_channels'
GROUPSTAT=yes
#---------------------------------------------------------------------
#  Define time averaging parameters
#  (namelist time_average)
#  TSLOT_H should be equal to statistics time step (in the GRIB file)
#---------------------------------------------------------------------
TSLOT_H=12

#---------------------------------------------------------------------
#  Define monitoring files to output
```

```
#------------------------------------------------------------------
L_HIST=true
L_HOV=true
L_GEO=true

#------------------------------------------------------------------
#  Define output for history files
#  variables HIST_CHOICE to customize the output
#  select HIST_CHOICE from obs_fg/obs_fg_an/obs_fg_bcor/obs_fg_an_bcor
#  or customize a new one in $SCRIPTDIR/hist_templates.sh
#  Default is obs_fg_an_bcor (see $SCRIPTDIR/hist_templates.sh)
#------------------------------------------------------------------
HIST_CHOICE='fso'
#------------------------------------------------------------------
#  areas (namelist history_area)
#------------------------------------------------------------------
NAREA=1
LON_WEST='0.'
LON_EAST='360.'
LAT_NORTH='90.'
LAT_SOUTH='-90.'
ISEL_AREA='1'

#------------------------------------------------------------------
#  surface types (namelist surface_type)
#  1=land, 2=sea, 3=sea-ice, 4=all
#------------------------------------------------------------------
NSURF=1
ISEL_SURF='4'

#------------------------------------------------------------------
#  usage flags (namelist sel_flag)
#------------------------------------------------------------------
NFLAGHIST=1
ISEL_FLAGHIST='5'
#------------------------------------------------------------------
#  Define output for geo mean files
#------------------------------------------------------------------
LGEO_FSO=true
GEO_HOURS='all'
NFLAGGEO=1
ISEL_FLAGGEO='5'

#------------------------------------------------------------------
#  Define output for hovmoeller files
#------------------------------------------------------------------
LHOV_FSO=true
NFLAGHOV=1
ISEL_FLAGHOV='5'
HOVMODE=lat_time
AUTOSCALE=true
SMOOTH=true
```

The result is three plots presented in Figs. 19, 18 and 20.

5.3.5 *Example 5 (Overview plots)*

The following example shows the plotting of overview plots for area averages timeseries and Scan dependent statistics. In this example plotted statistics are from 'Global' stream METOP-A AMSU-A and MHS.
The option file used is:

```
#
#   Monitoring option file for METOP-A/AMSU A/B
#        OVERVIEW PLOTS ALL CHANNELS
#
#======================================================================

#----------------------------------------------------------------------
#  Define mode, satellites and sensors
#  to monitor.
#----------------------------------------------------------------------
MODE='tb'
set -A SATIDS   004  004
set -A SENSORS  003  015

DATASTREAM=0

#----------------------------------------------------------------------
#  Define channels/levels to monitor
#  (namelist sel_lev)
#   Separate by commas or spaces
#   with -99 after each satellite set
#----------------------------------------------------------------------
OVIEW_SEL_CHAN='1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 -99 1 2 3 4 5 -99'

#----------------------------------------------------------------------
#  Define time averaging parameters
#  (namelist time_average)
#----------------------------------------------------------------------
TSLOT_H=6

#----------------------------------------------------------------------
#  Define monitoring files to output
#  (namelist output_type)
#----------------------------------------------------------------------
L_OVIEW=true
L_SCAN=false
#----------------------------------------------------------------------
# 2. areas (namelist history_area)
#----------------------------------------------------------------------
# global,
NOVIEWAREA=6
LON_WEST='0. , 0. ,  0. ,  0. ,  0. ,  0.'
LON_EAST='360. , 360. , 360. , 360. , 360. , 360.'
LAT_NORTH='90. , 90. , 70. , 20. ,-20. , -70.'
LAT_SOUTH='-90. , 70. , 20. , -20. , -70. , -90.'
OVIEW_SEL_AREA='1 2 3 4 5 6'
AREAS='all_areas'



#----------------------------------------------------------------------
# 3. surface types (namelist surface_type)
#  must be same number of surfaces as channels and satids
#   with -99 after each satellite set
#  1=land, 2=sea, 3=sea-ice, 4=all
#  for each surface type please ensure that associated statistics
#  were requested during the statistics calculation part
#----------------------------------------------------------------------
OVIEW_SEL_SURF='2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 -99 2 2 2 2 2 -99'
#----------------------------------------------------------------------
# 4. usage flags (namelist sel_flag)
#  must be same number of flags as channels and satids
```

```
#  with -99 after each satellite set
#-----------------------------------------------------------------------
OVIEW_SEL_FLAG='40 40 40 40 40 40 40 40 40 40 40 40 40 40 40 -99 40 40 40 40 40 -99'
```

The result of this option file is several plots. Two are presented here Figs. 21 and 22.

5.3.6 *Example 6 (Solar timeseries)*

In this example we show a solar time series plot (see Fig. 23). The option file is not presented as the only difference with the first one is the use of 1 hour as a time step and the activation of solar time production by setting L_SOLT=true. As can be noted in the plot, statistics show a high dependency on the solar time. The bias correction used is taking this into account and the result is a successful dynamical bias correction.

5.3.7 *Example 7 (Statistics comparison)*

In this last example we present two plots showing the comparison between two datasets. For the first plot (see Fig. 24) we compare the same channel from the same instrument from two different satellites. In the second example (see Fig. 25) the datasets differ only in the experiment version.

5.3.8 *Example 8 (Targeted sites statatistics)*

The following example shows the production of SMOS statistics over a targeted site (Nemaha). The associated plot is shown in Fig. 26.

```
#
# Generic Plotting options file
#
#======================================================================

#======================================================================
#   Define mode, satellites and sensors
#   to monitor.
#======================================================================
MODE=tb
set -A SATIDS  46
set -A SENSORS  18
#----------------------------------------------------------------------
#  Define channels/levels to monitor
#  (namelist sel_lev)
#  only activate the DATASTREAM when it's available
#----------------------------------------------------------------------
NSELECT=2
ISEL_CHAN='1    2'
#
#
FOVSIZE=9.000000000
ISEL_FOV='1    2    3    4    5    6    7'


#----------------------------------------------------------------------
#  Define time averaging parameters
#  (namelist time_average)
#  TSLOT_H should be equal to statistics time step (in the GRIB file)
```

```
#-------------------------------------------------------------------
TSLOT_H=12

#-------------------------------------------------------------------
#  Define monitoring files to output
#-------------------------------------------------------------------

#-------------------------------------------------------------------
#  Define monitoring files to output
#-------------------------------------------------------------------
L_LOC=true

#-------------------------------------------------------------------
#  Define output for history files
#  variables HIST_CHOICE to customize the output
#  select HIST_CHOICE from obs_fg/obs_fg_an/obs_fg_bcor/obs_fg_an_bcor
#  or customize a new one in $SCRIPTDIR/hist_templates.sh
#  Default is obs_fg_an_bcor (see $SCRIPTDIR/hist_templates.sh)
#-------------------------------------------------------------------
HIST_CHOICE='obs_fg_an_bcor'

#-------------------------------------------------------------------
#  areas (namelist history_area)
#-------------------------------------------------------------------
NAREA=6
LON_WEST='   0. , 358.3, 263.4, 258.2, 263.3'
LON_EAST=' 360. ,   3.9, 263.6, 258.4, 263.9'
LAT_NORTH=' 90. ,  45.7,  45.0,  40.5,  39.8'
LAT_SOUTH='-90. ,  42.3,  40.8,  40.3,  39.6'
AREANAMES='GLOBE , SMOSmania, Lancaster, Chase ,Nemaha'
LOC_AREA='5'

#-------------------------------------------------------------------
#  surface types (namelist surface_type)
#  1=land, 2=sea, 3=sea-ice, 4=all
#-------------------------------------------------------------------
NSURF=1
ISEL_SURF='1'

#-------------------------------------------------------------------
#  usage flags (namelist sel_flag)
#-------------------------------------------------------------------
NFLAGHIST=1
ISEL_FLAGHIST='2'
AUTOSCALE=false
```

6 Making modifications in OBSTAT

For users wishing to upgrade/modify the OBSTAT code, the appropriate method is to create a perforce branch of a recent IFS release using the following command:

```
q2 create_branch -r CY371R2 -b test -p obstat
cd obstat/src
```

obstat/src contains all obstat related fortran code (including the plotting package). obstat/module contains the associated modules. The prepIFS make familes will produce libraries for both supercomputer and

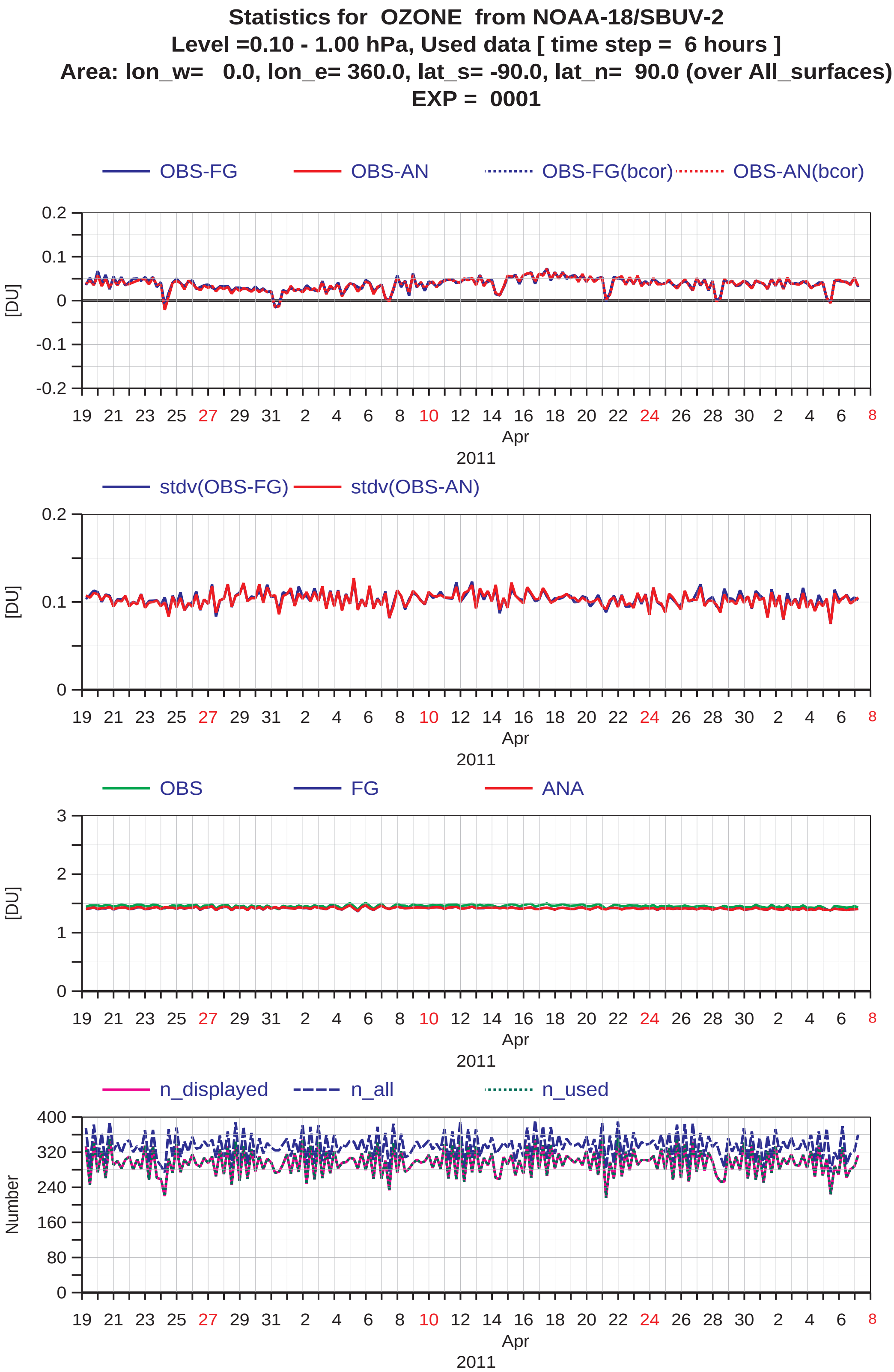

Figure 11: Area averages timeseries for layer 0.1-1.0 hPa used ozone from NOAA-18/SBUV-2 over the Globe

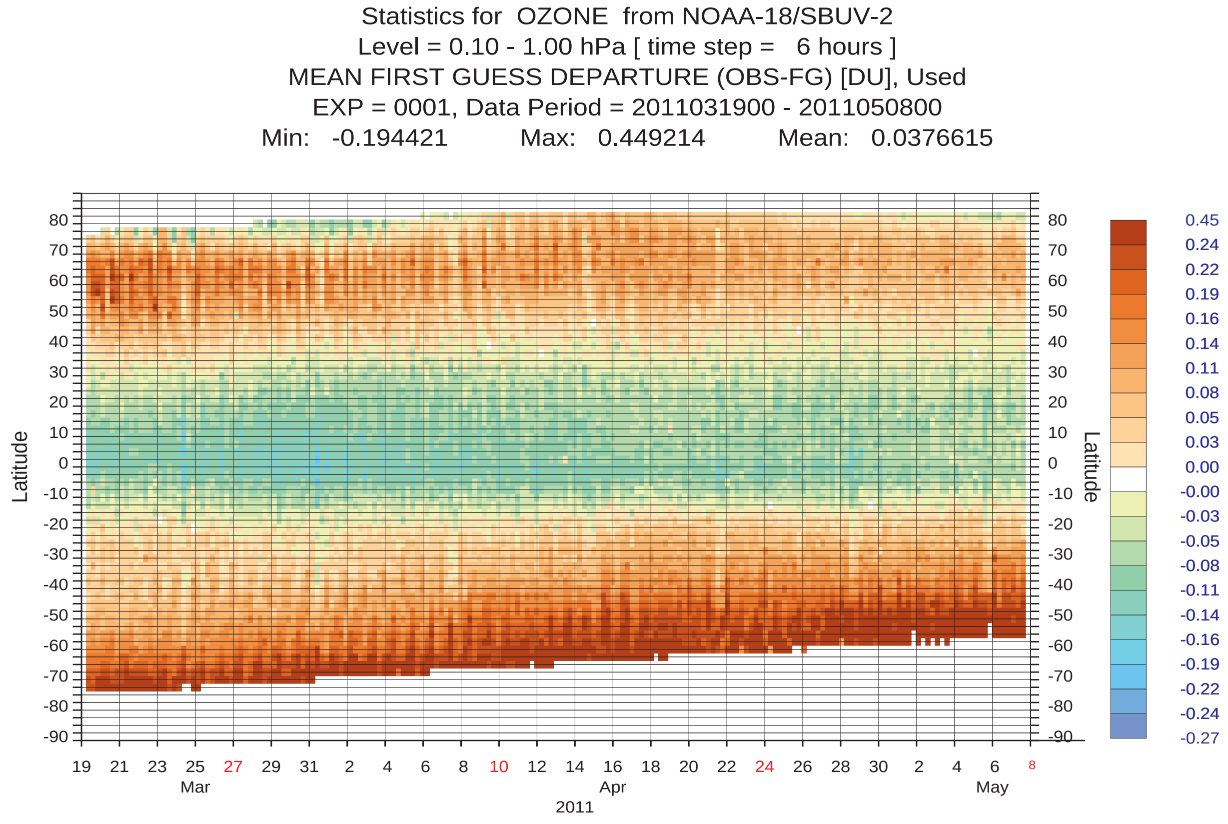

Figure 12: Latitude versus time Hovmoeller of FG departure for layer 0.1-1.0 hPa used ozone from NOAA-18/SBUV-2

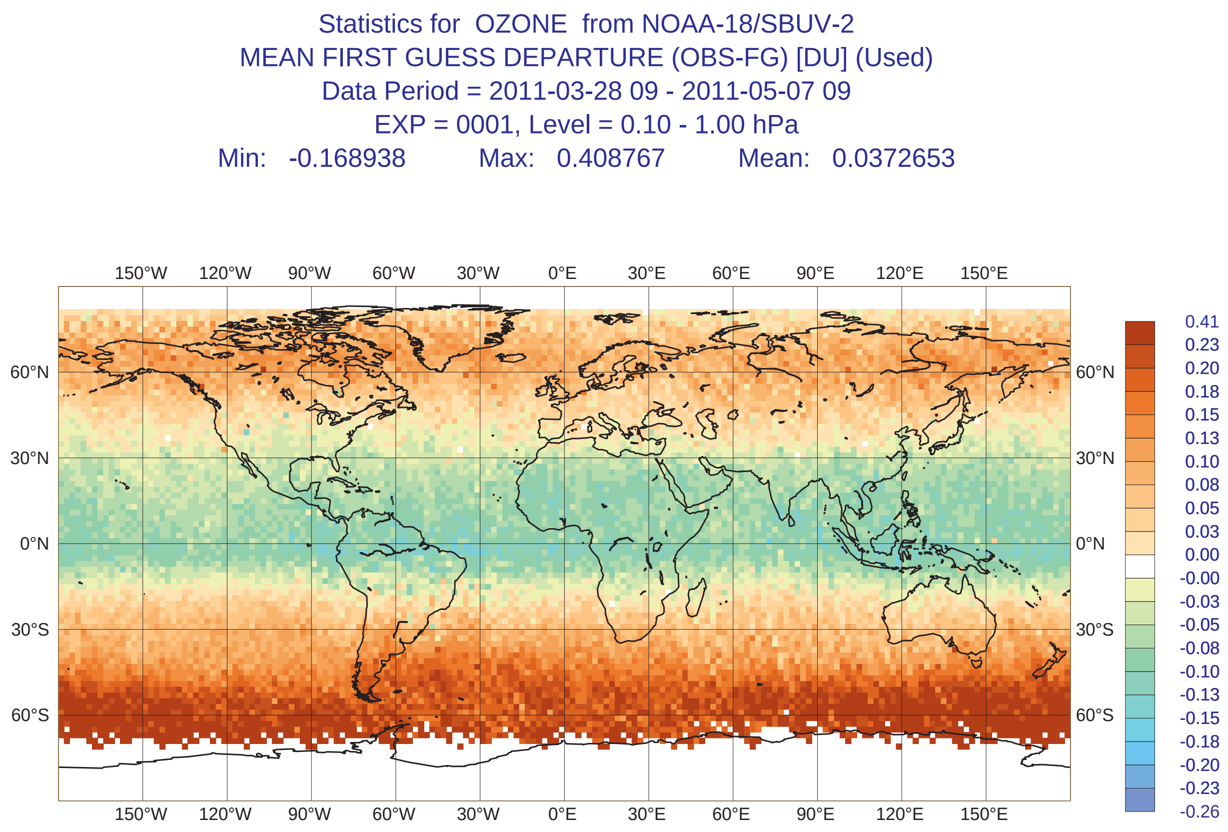

Figure 13: goegraphical map of FG departure for layer 0.1-1.0 hPa used ozone from NOAA-18/SBUV-2

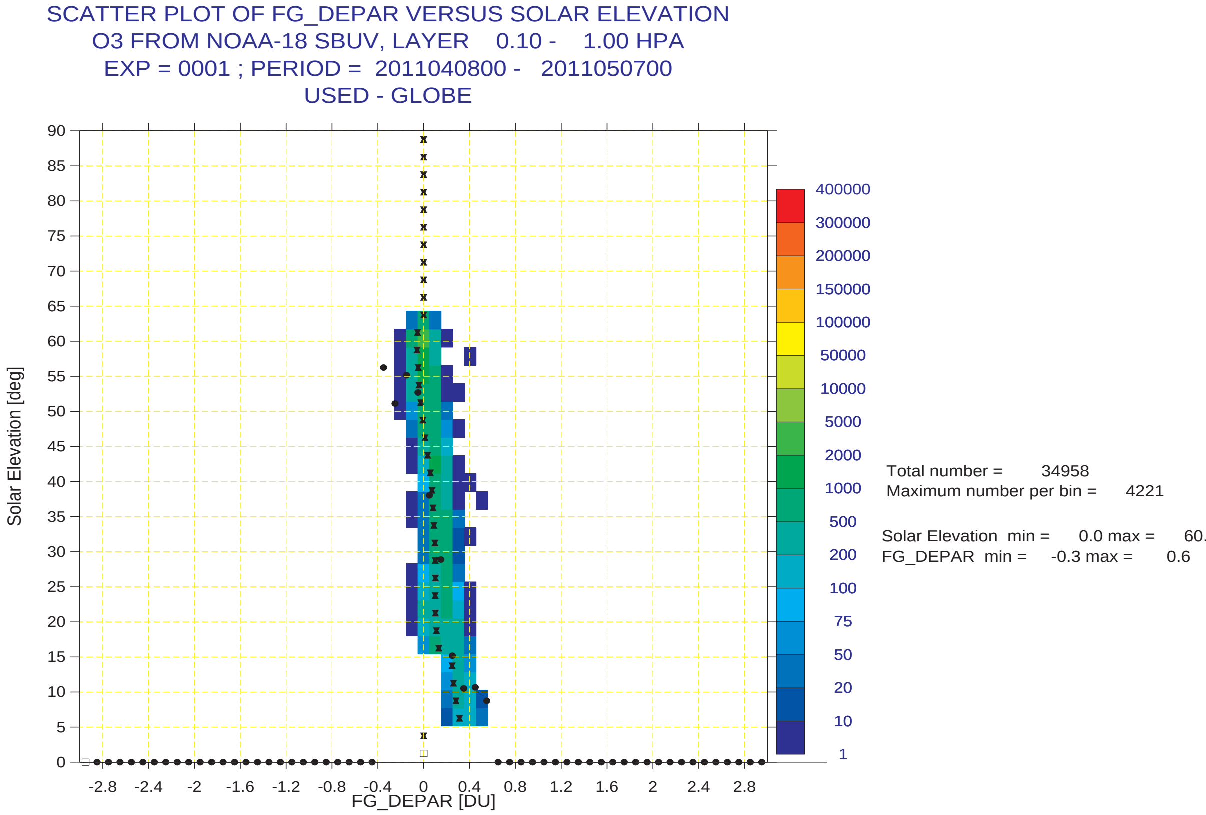

Figure 14: scatter plot FG departure against Solar elevation for layer 0.1-1.0 hPa used ozone from NOAA-18/SBUV-2

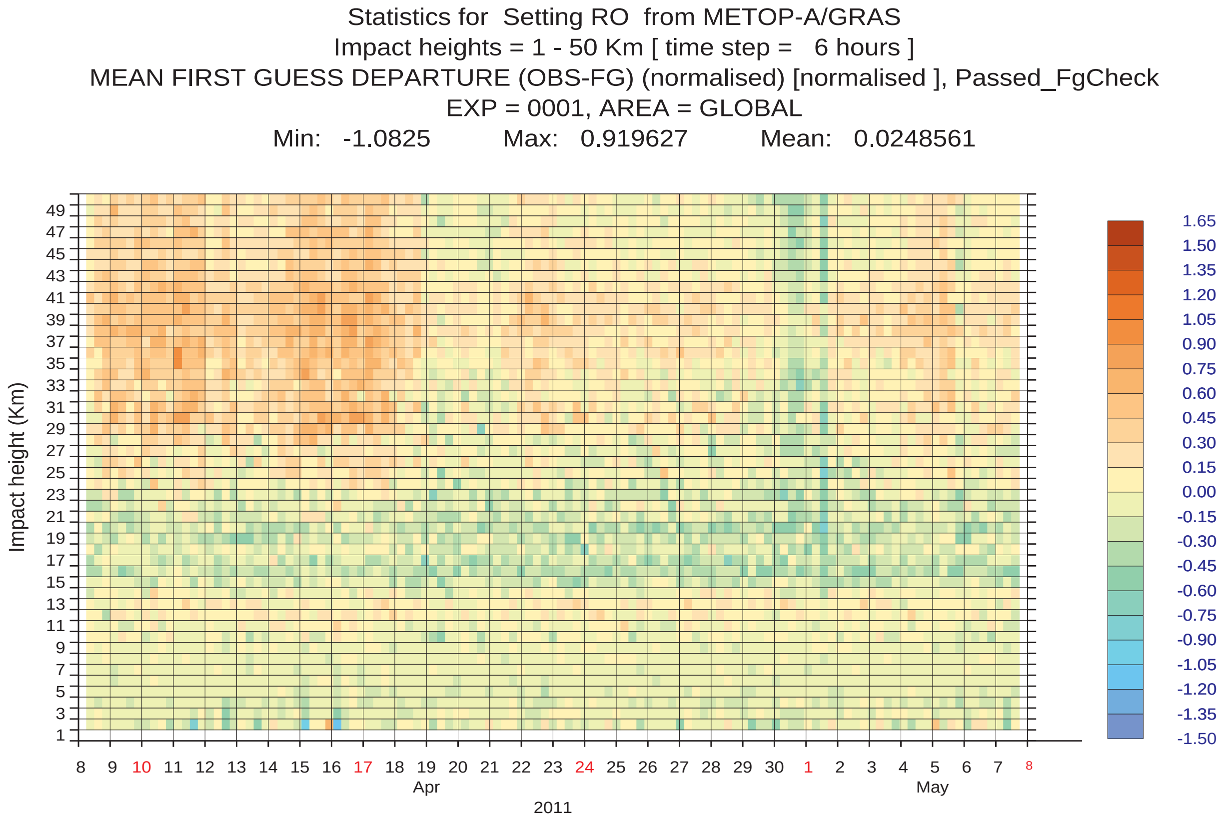

Figure 15: Imapct heights versus time of Normalised FG departures for METOP-A/GRAS Setting RO Data passing FG check

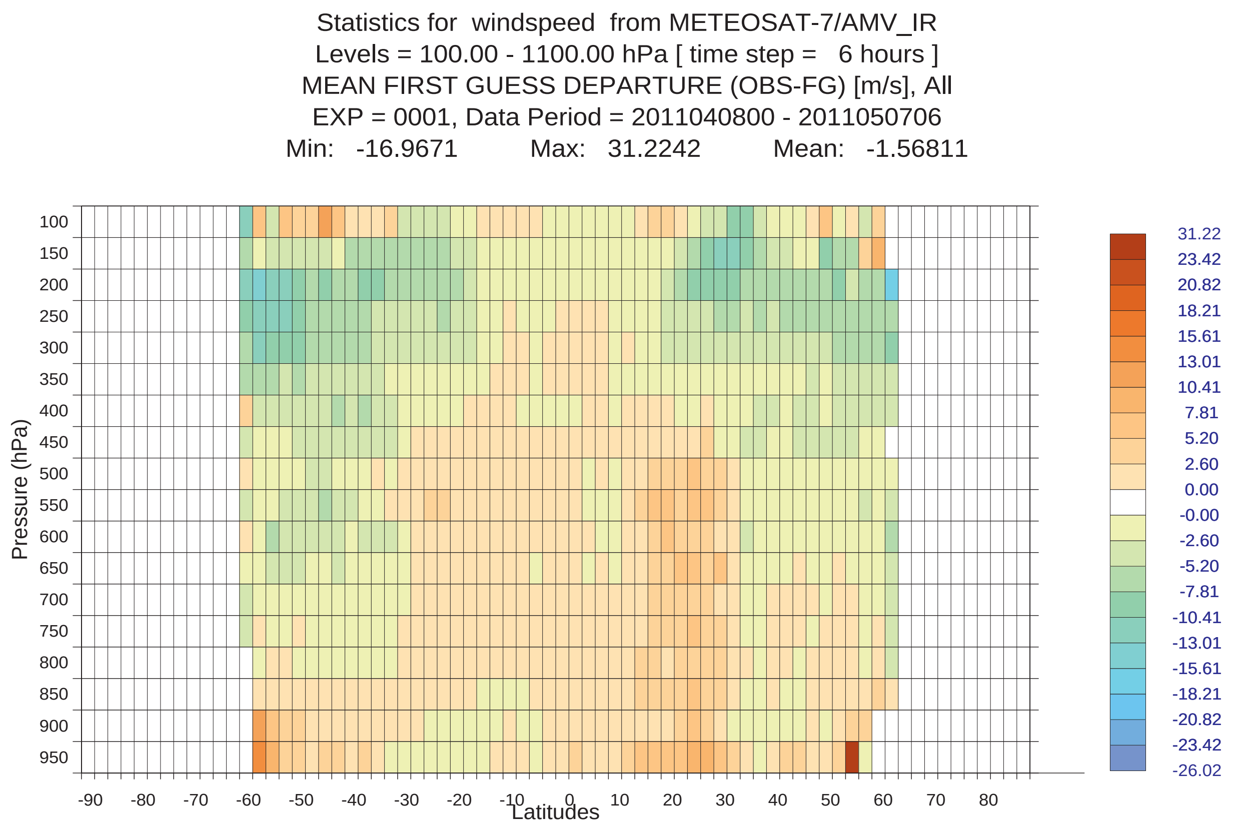

Figure 16: Example of pressure levels versus latitude bands overaged over a period of time

the linux-cluster/desktop workstation. An offline set of makefiles does exist (contact the OBSTAT admin for details).

To modify the scripting part of OBSTAT the user needs to add the scripts project using the following command:

```
q2 add_projects
and then select and past the project 'scripts'
```

Users wishing to modify the scripting part of the plotting packge are invited to take a copy of the folder /home/mo/obstat/OBSTAT/scripts. This directory contains all scripts involved in the plotting.

The standalone statistics calculation and the plotting scripts allow users to plug their modifications easily:

- The parameters LIBS and OBSTAT_LIB in `dobstat` allows users to point the their version of the OBSTAT library
- DATADIR `dobstat` allows users to point the their version of OBSTAT scripts and configuration files.
- BINDIR in `obstat_plot` allows users to use their generated plotting binaries
- SCRIPTDIR in `obstat_plot` allows users to use their modified plotting scripts

To contribute to the improvment of the software, users are kindly invited to make aware the OBSTAT admin about technical modifications they have made if they think that such modifications are relevant for others.

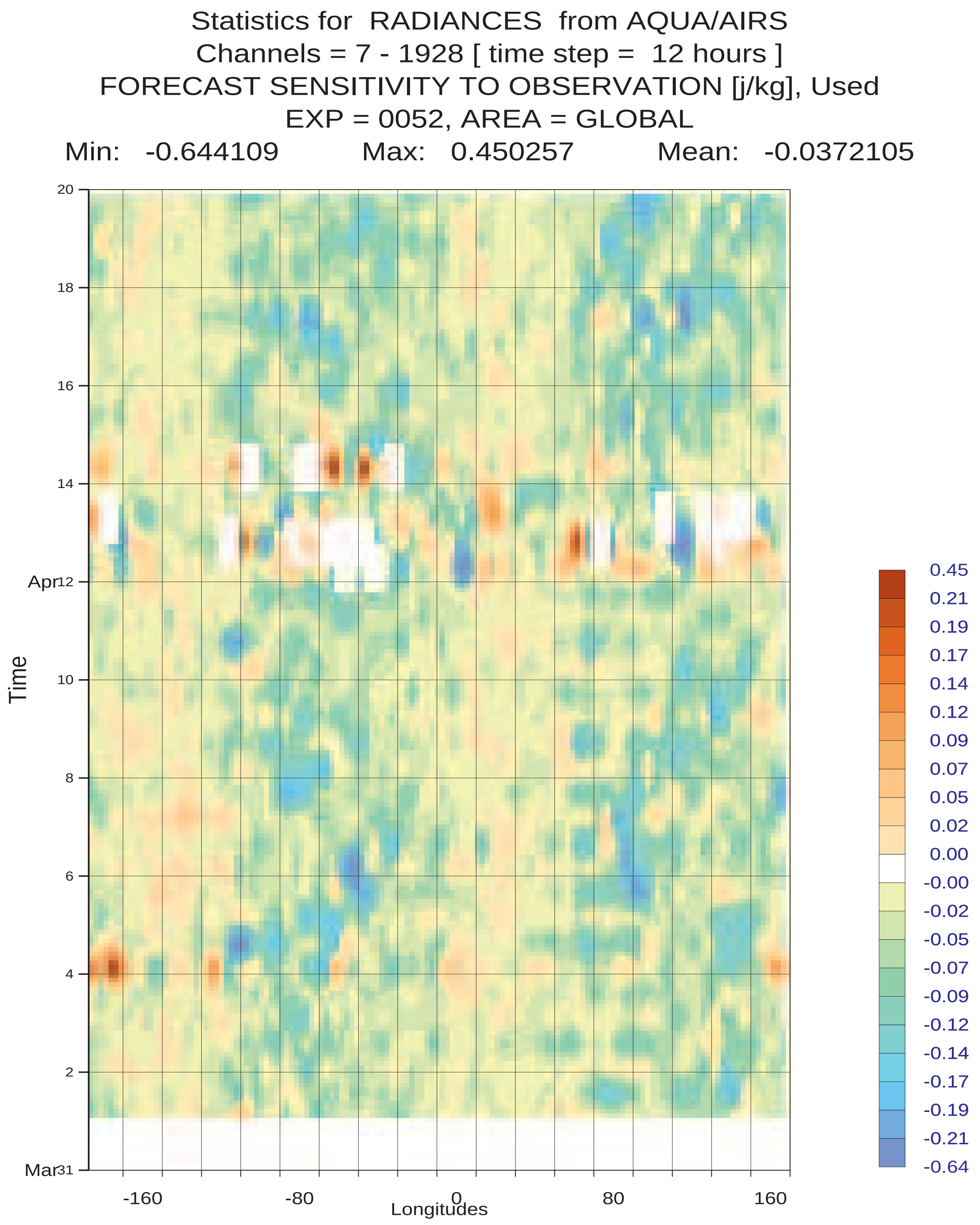

Figure 17: Example of time versus longitude bands

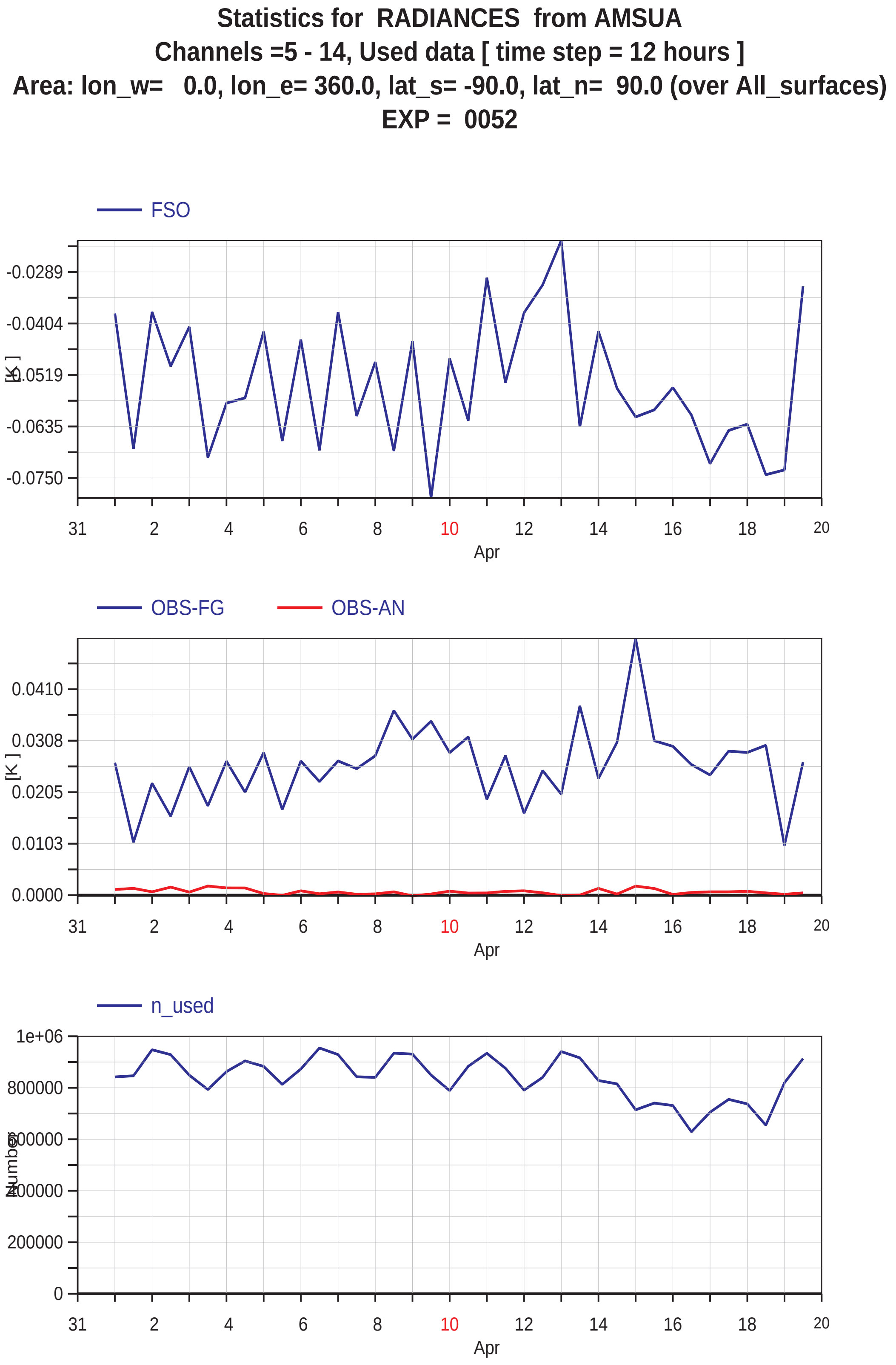

Figure 18: Area averages timeseries for 'fso' template. AMSUA data are plotted from several satellites

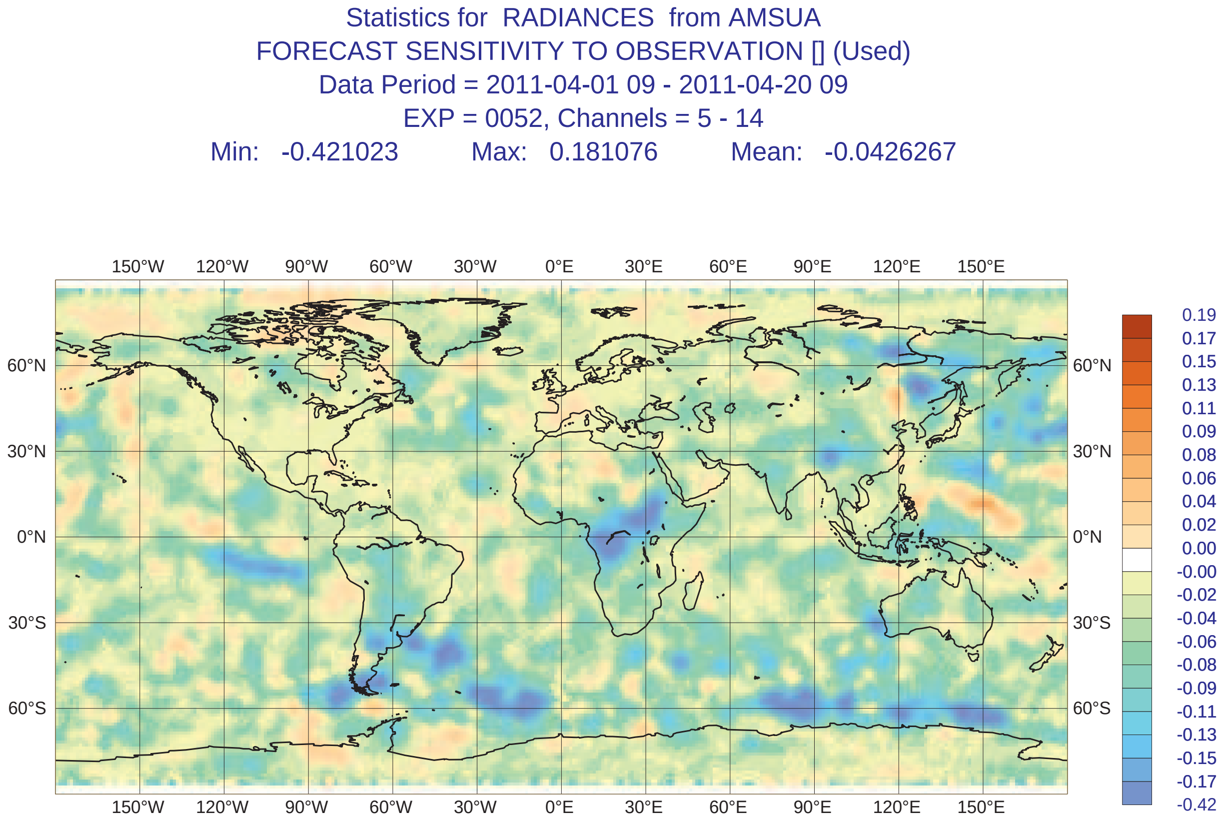

Figure 19: Geographical maps of FSO for AMSUA data from all satellites

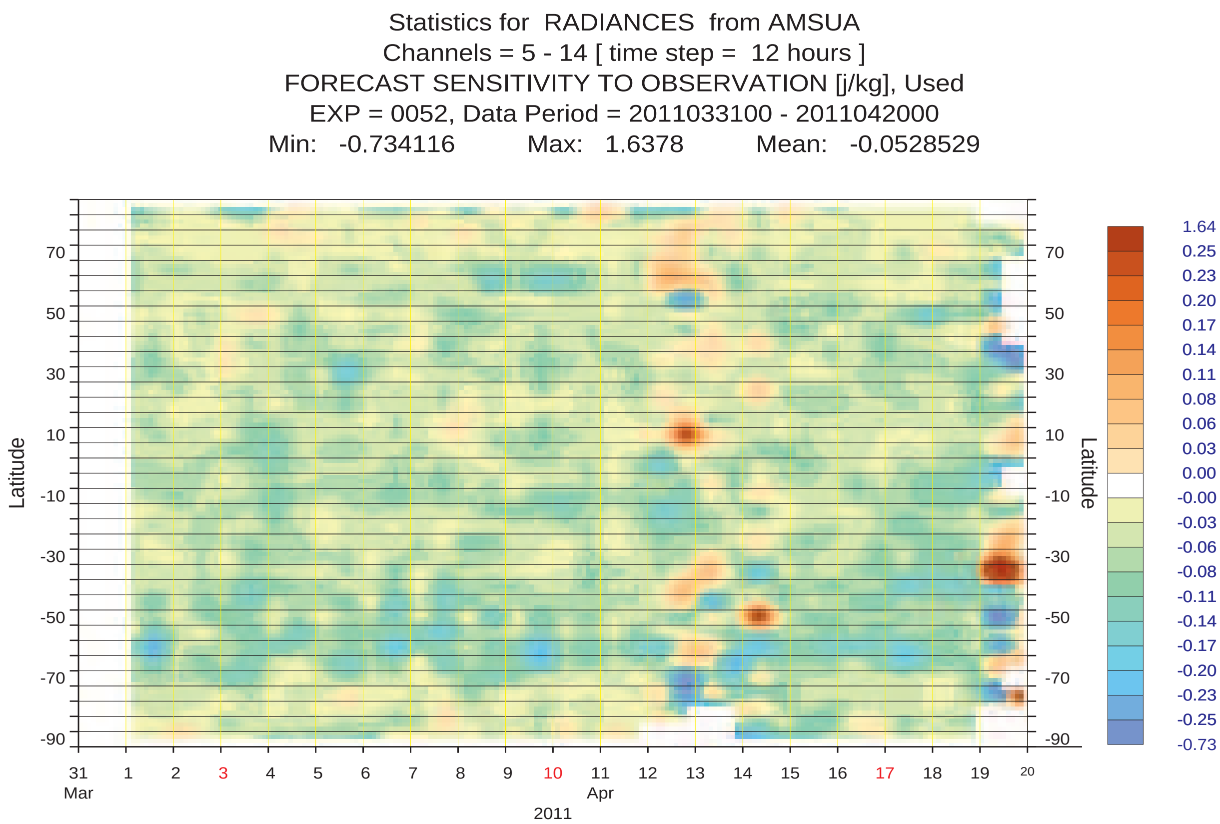

Figure 20: Latitude versus time FSO for AMSUA data from all satellites

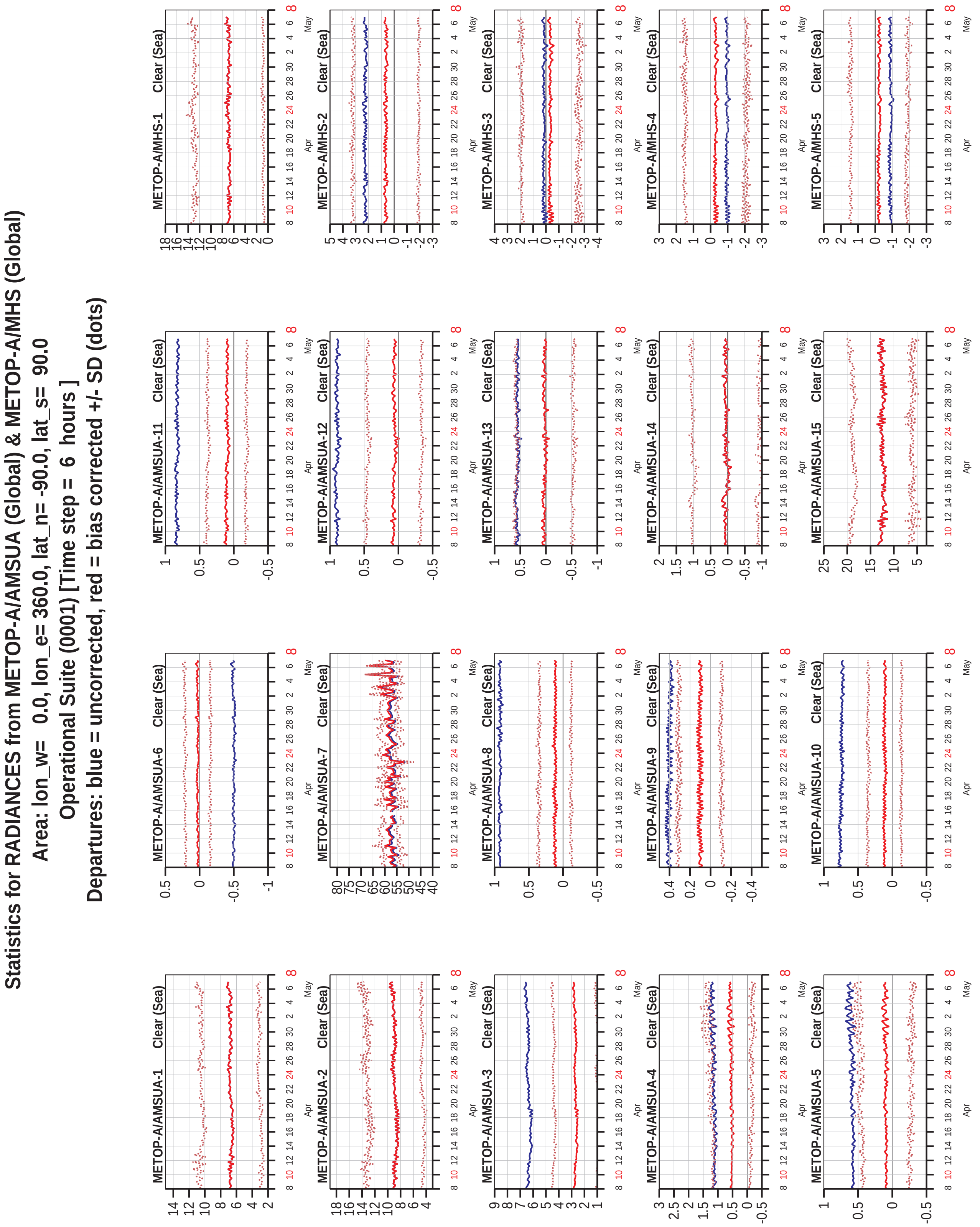

Figure 21: Overview plot showing area averages timeseries for clear radiances from METOP-A/AMSUA and METOP-A/AMSUB

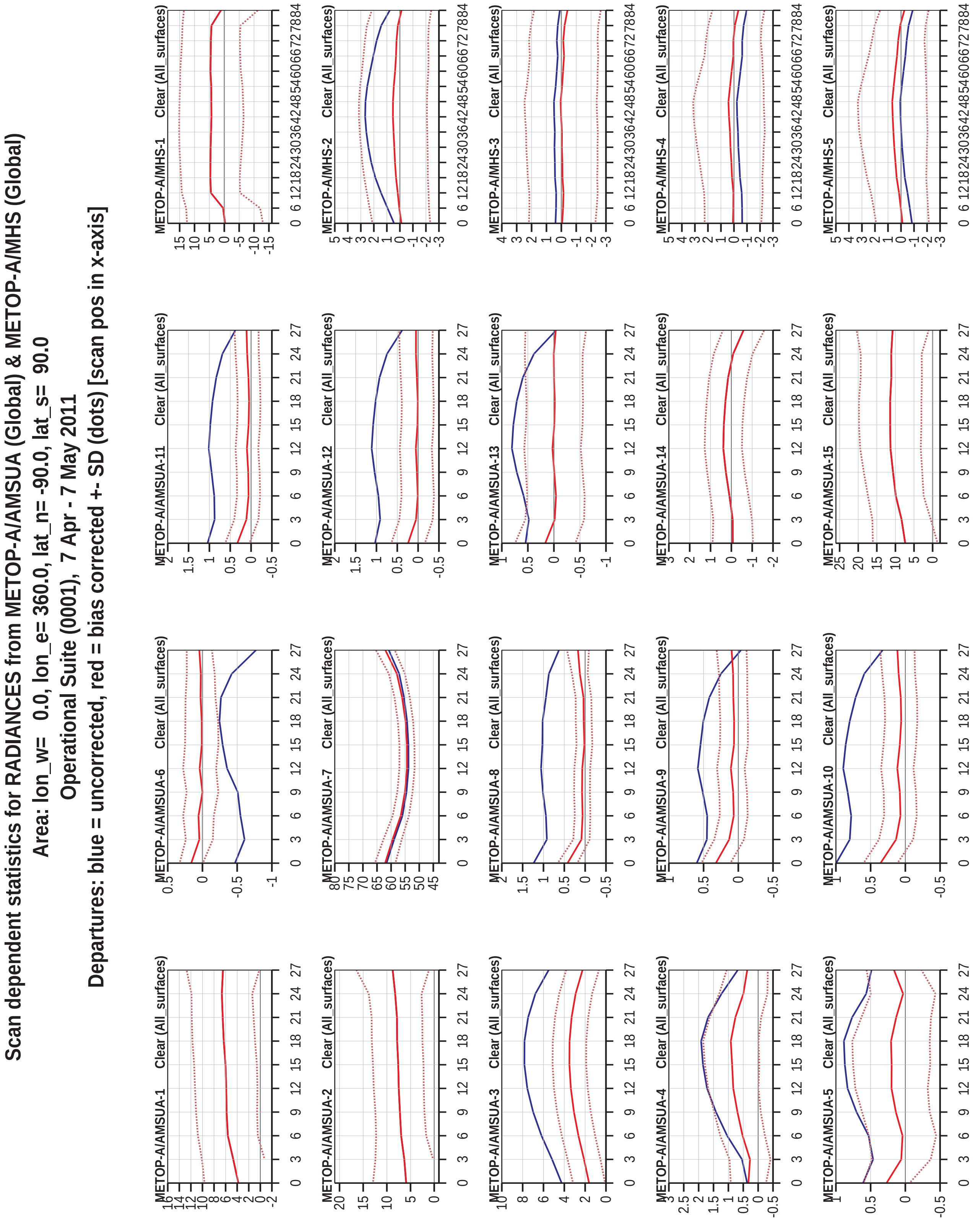

Figure 22: Overview plot showing time averaged scan dependend clear radiances from METOP-A/AMSUA and METOP-A/AMSUB

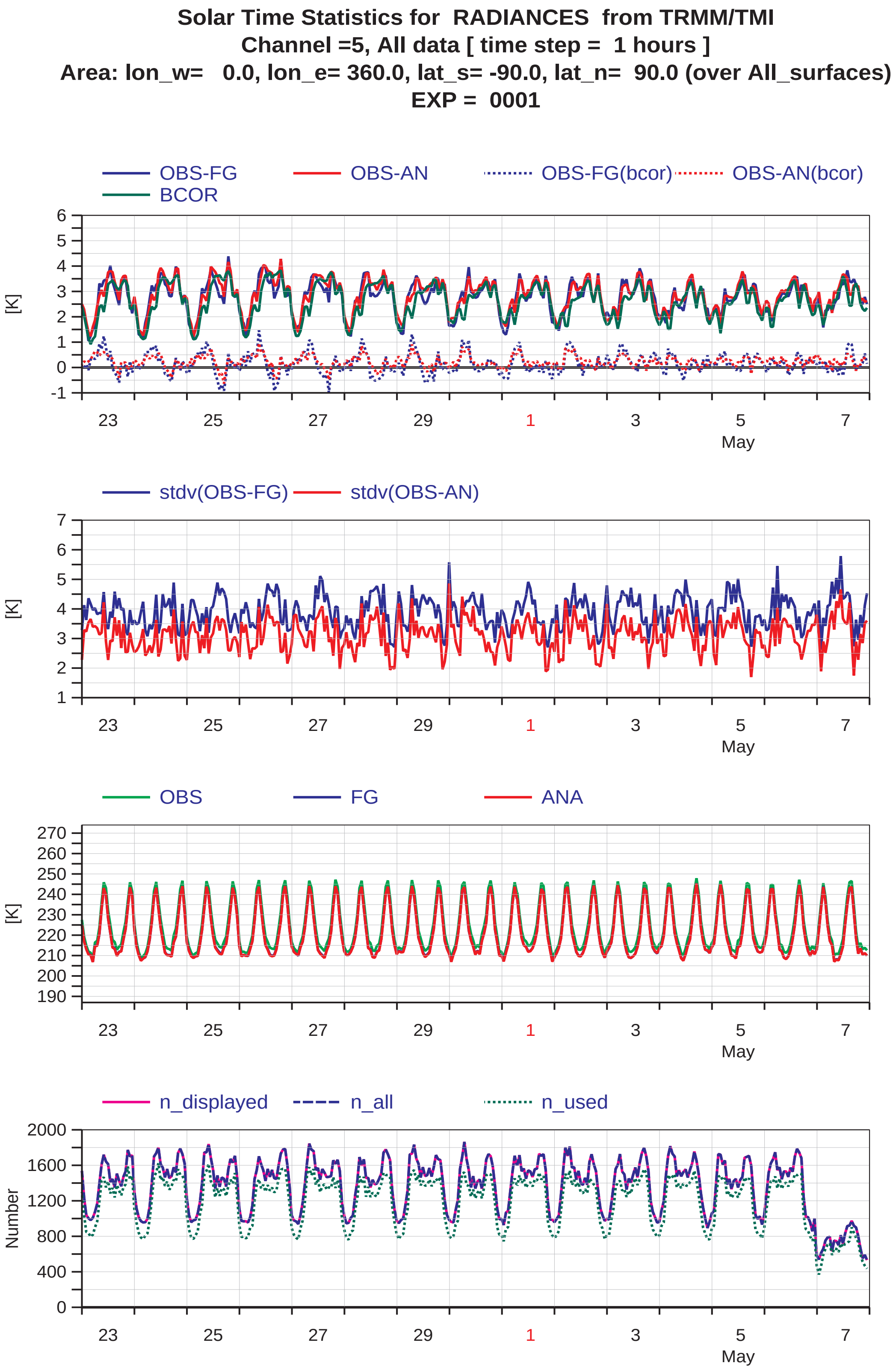

Figure 23: Solar timeseries of area averages for TRMM/TMI.

Statistics for RADIANCES from METOP-A/AMSUA (Global) vs AQUA/AMSUA (Global)
Channel =5, All data [time step = 6 hours]
Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0 (over All_surfaces)
EXP = 0053
[Second data set with symbols]

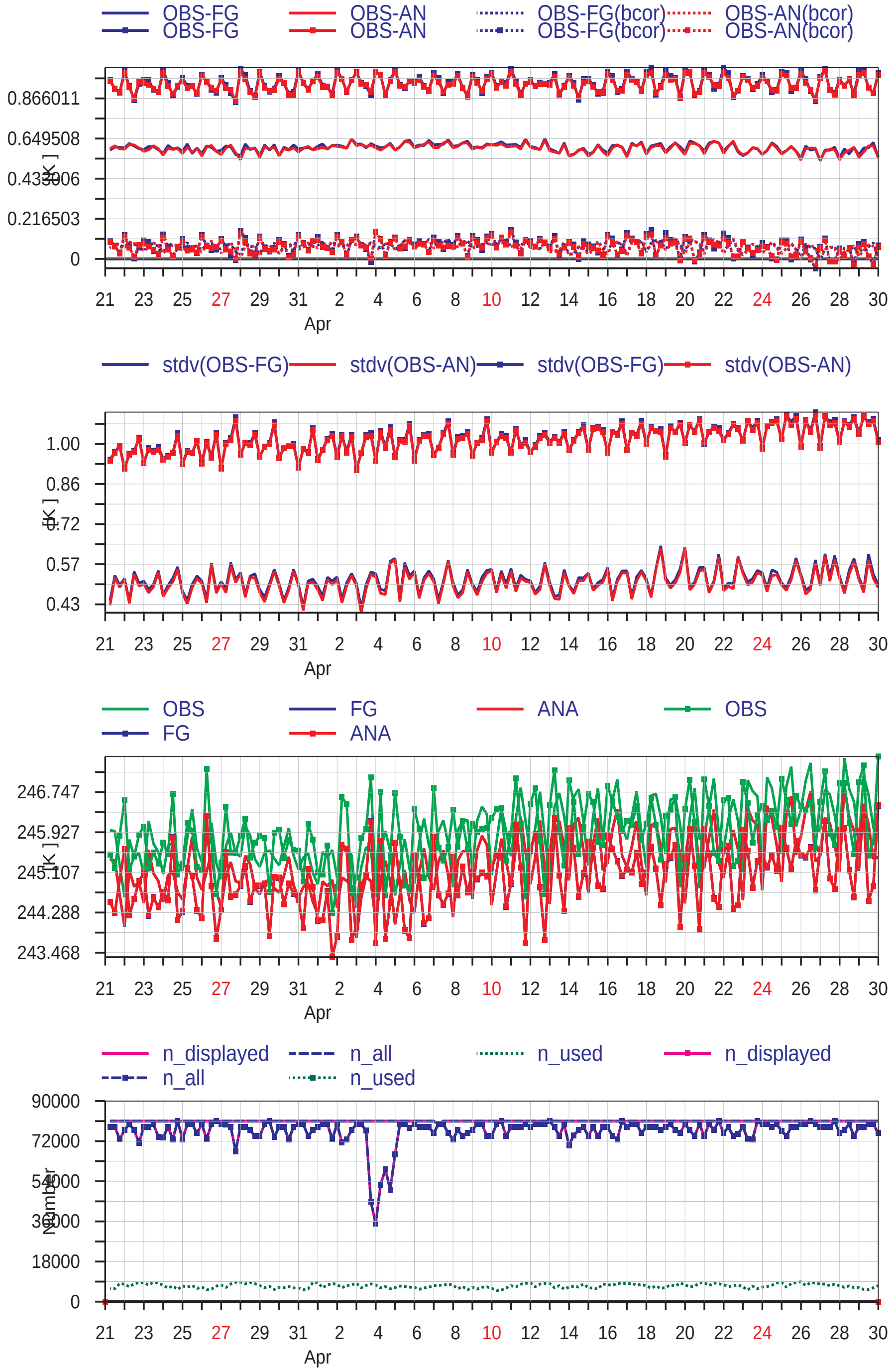

Figure 24: Area averages timeseries of channel 5 radiances from AQUA/AMSU-A and METOP-A AMSU-A

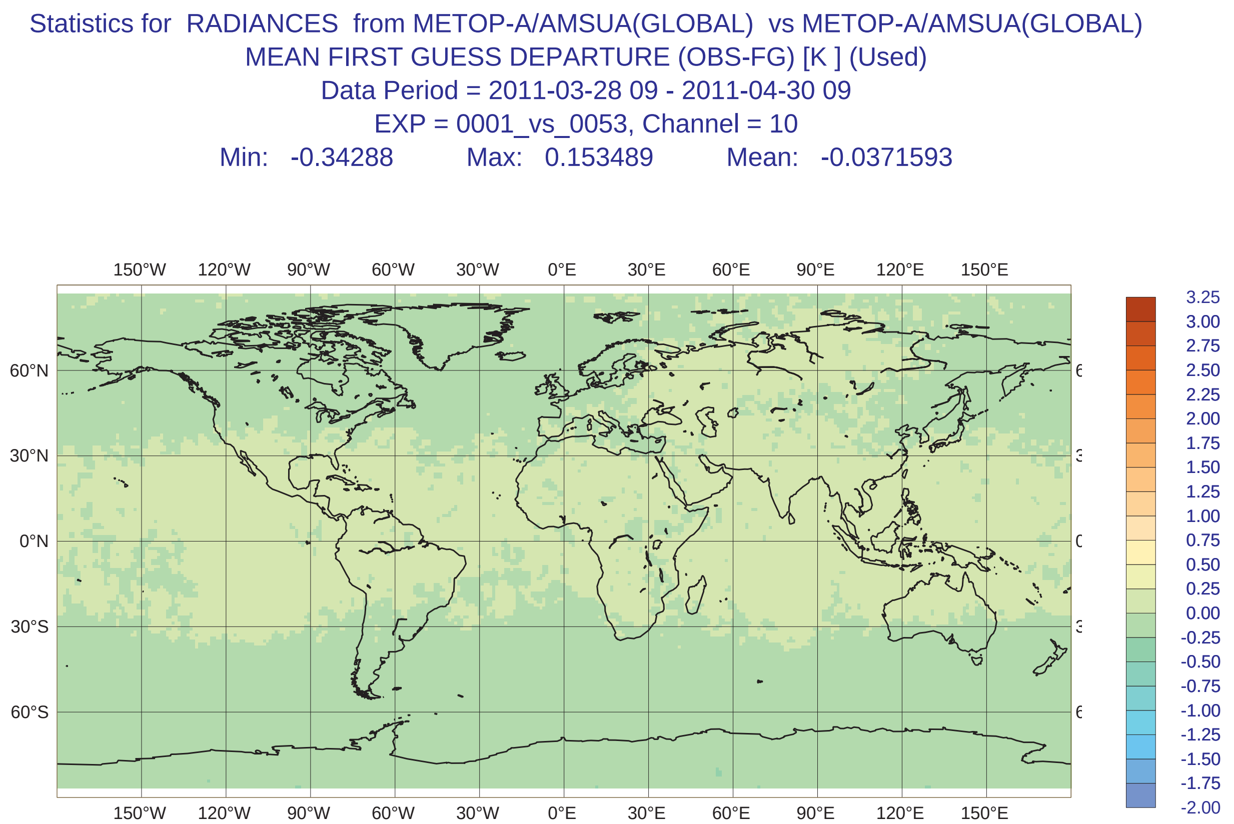

Figure 25: Geo map of differences of FG departures for channel 10 radiances from METOP-A/AMSUA between two experiments

7 Quick summary of using the standalone OBSTAT

This section provides a quick summary on how to use the OBSTAT for statistics calculation and plotting:

- Statistics calculation:
  - Get a copy of the script dobstat from ”/home/mo/obstat/OBSTAT”
  - Modify informations related to the period and experiment identification
  - Provide an updated version of STATDEF if the default version does not contain all user needs. The default version is cycle dependent and it is located on ”/home/rd/rdx/data/$IFSCYCLE/plot
  - Change if necessay the OBSFORMAT, ODBFILETYPE and datalist
  - For more customisation the user is reffered to 3.1
  - Run the script dobstat
- Classical statistics plotting
  - Get a copy of the script dobstat_quick from ”/home/mo/obstat/OBSTAT”
  - Modify informations related to the period and experiment identification
  - For more customisation the user is reffered to 3.1
  - Run the script dobstat_plot
- Gridded statistics plotting

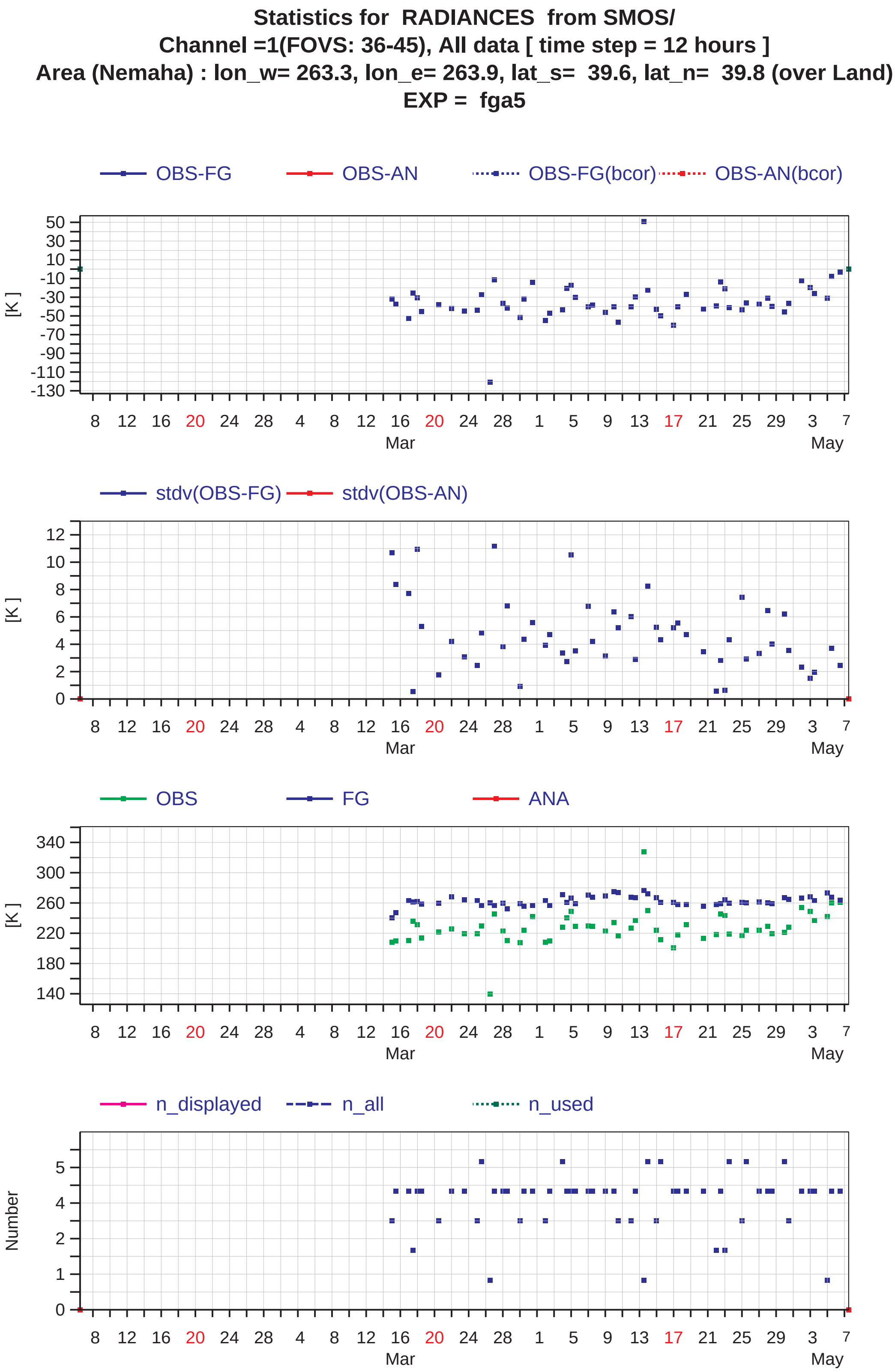

Figure 26: Nemaha time series statistics for SMOS radiances: XX polarisation angle 36

- Get a copy of the script obstat_plot from ”/home/mo/obstat/OBSTAT”
- Modify informations related to the period and experiment identification
- Retrieve the option files produced by OBSTAT during the statistics calculation step. The files are stored in $GRIBDIR (see 5.2 for details)
- Update the script obstat_plot with information about the name of the plotting option to use and where it is.
- For more customisation the user is reffered to 5.2
- Run the script obstat_plot

8 Suggestions for future developments

Below some planned OBSTAT development actions:

- Archive OBSTAT statistics in MARS
- Improve and Extend the Metview 4 support to OBSTAT
- Develop more tools to manipulate the gridded statistics outside the framework of plotting package.
- Optimize the plotting package
- Allow the computation of collocation statistics
- Add the support of other encoding formats of statistics (e.g NetCDF). This is a low priority action

Users are welcome to express their requests for new features. This is considered as an important source of ideas to improve the system.

9 Statistics interpretation guide

Observation departures are extremely valuable in that they provide an independent information to validate the model fields in the background and in the analysis. Statistics of background departures are essentially short-term forecast scores from the previous analysis. Statistics of analysis departures indicate to which extent the model state was unable to fit the data because of constraints in the background, the other observing systems and (in 4D-Var) in the model equations. To interpret the plots correctly, one has to be aware of a number of issues, as suggested by the following list of facts:

- smaller background departures and/or more data used are usually the signs of a well-behaving assimilation.
- variations of statistical quantities are only meaningful if the amount of data used is large enough, and if the variation can be reproduced on several independent samples.
- The decrease from background to analysis departure is an indication of the amount of observed information that was actually taken into account in the analysis.
- the background is such a short-term forecast that its quality is usually comparable to the observation and representativeness errors. Thus large background departures can be caused by poor-quality data.

- because of the probabilistic nature of the analysis procedure, the analysis fit to the data should not be too small (only slightly smaller than the observation error). There have been cases of analysis improvement by fitting *less* the data.
- there are non-negligible biases in most modelled and observed parameters, sometimes with a complex structure. Some observation types (satellite data) are debiased against the model, and different observing systems may exhibit contradicting biases.
- an effect of the quality control procedure is to reject data that is too different from the background. Inclusion or rejection of 'borderline' data may have a large effect on the statistics. Unfortunately there is no way to ensure that the same set of data is used to compare two analyses done at the same time with a different system.
- variations in statistics on the same parameter from one observing network to another may simply be due to a different geographical coverage, rather than to an intrinsic change in the quality of data or model.
- the analysis rms and biases should always be significantly smaller than the background ones. otherwise there is a flaw in the assimilation system, such as competing observing systems with different biases.
- when comparing two assimilations, an increase in background departures may not be a bad sign if more data is used. One should then look for changes in other data types which are not affected by changes in rejection thresholds.
- histograms that have large 'tails' indicate flaws in the quality control system, provided the population is large enough.

10 Acknowledgements

This documentation is partially based on the version written by Lars Isaksen. The `OBSTAT` software was originally developed by Francois Bouttier at ECMWF. It has been improved by Carla Cardinali, Lars Isaksen, Sami Saarinen, Ulf Andrae and Meteo-france staff. More recently `OBSTAT` has been significantly upgraded by Mohamed DAHOUI to support the production and plotting of gridded statistics. Claude Gibert has developed the software to present web plots of time series on ECMWF's web sites.

A OBSTAT GRIB2 tables

Grib Code	Short name	Long Name
1	Active	Active data
2	All	All data
3	Non_Active	Not Active data
4	Best_Active	Best active wind
5	Used Used	data
6	VarQC_Rej	VarQC rejected data
7	Blacklisted	Blacklisted data
8	Failed	Failed data
9	Passed_FgCheck	Data that passed FG check
10	Non_Rejected	All non rejected data
11	VarBC_Passive	VarBC passive channels
12	Failed_FG_Non_Black	Data failed FG check but not blacklisted
13	Failed_FG_VarQC_Rej	Data failed FG check and VARQC rejected
14-19	Reserved	Reserved for additional standard IFS flags
20	QI_LE_20	AMVs with QI ≤ 20
21	QI_LE_66	AMVs with $20 <$ QI ≤ 65
22	QI_GE_65	AMVs with QI > 65
23	QI_GE_80	AMVs with QI > 80
24	QI_GE_90	AMVs with QI > 90
25-29	Reserved	for additional AMVs flags
30	Clear_LE_70%WV_80%IR	CSR data with clear fraction < 70 % (WV) and < 80 % (IR)
31	Clear_GE_70%WV_80%IR	CSR data with clear fraction ≥ 70 % (WV) and ≥ 80 % (IR)
32	Clear_100%	CSR data completely clear (according to IR window channel)
33	Clear_GE_40%WV	CSR data with clear fraction ≥ 40 % (WV)
34	Clear_GE_70%WV	CSR data with clear fraction ≥ 70 % (WV)
35	Clear_100%WV	CSR data completely clear (according to WV channel)
36-39	Reserved	resserved for additional CSR flags
40	Clear	Clear
41	Used_Clear	Used clear data
42	Used_Cloudy_Rainy	Used cloudy and rainy data
43	All_Cloudy_Rainy	All cloudy and rainy data
44	Used_ObsCld_FGClr	Used Obs cloudy and FG clear
45	Used_ObsClr_FGCld	Used Obs clear and FG cloudy
44-49	Reserved	Reserved for additional radiances flags
50	Good_ozone	Good ozone data
51	Daytime	Day time data
52	Nighttime	Night time data
53-69	Reserved	Reserved for additional ozone, trace gases and Aerosol flags
70-79	Reserved	Reserved for GPSRO flags
80-89	Reserved	Reserved for scatterometer flags
90-199	Reserved	Reserved

Table 1: Data selection criteria codes

Grib Code	Short name	Long Name
0	Normal_delivery	Normal delivery
1	EARS	EARS
2	PAC-RARS	PAC-RARS
3	DB_MODIS	DB MODIS winds
4-255	Reserved	Reserved

Table 2: List of data streams codes

Grib Code	Short name	Long Name
1	Land	Land
2	Sea	Sea
3	Sea-ice	Sea-ice
4	All_surfaces	All surface types combined
5-255	Reserved	Reserved

Table 3: List of land sea Mask types

Grib Code	Short name	Long Name
1	count	data count
2	obs	Average of observed values
3	obs_stdv	Standard deviation of observed values
4	fgdep	Average of first guess departure
5	fgdep_stdv	Standard deviation of first guess departure
6	andep	Average of analysis departure
7	andep_stdv	Standard deviation of analysis departure
8	obs_error	Average of observation standard error
9	obs_error_stdv	Standard deviation of observation standard error
10	bkg_error	Average of background standard error
11	bkg_error_stdv	Standard deviation of background standard error
12	lr_andep1	Average of low resolution analysis departure update 1
13	lr_andep1_stdv	Standard deviation of low resolution analysis departure update 1
14	hr_fgdep2	Average of high resolution background departure update 2
15	hr_fgdep2_stdv	Standard deviation of high resolution background departure update 2
16	lr_andep2	Average of low resolution analysis departure update 2
17	lr_andep2_stdv	Standard deviation of low resolution analysis departure update 2
18	bcor	Average of Bias correction
19	bcor_stdv S	tandard deviation of bias correction
20	vbcor	average of Variational bias correction
21	vbcor_stdv	Standard deviation of variational bias correction
22	fgdep_nbcor	Average of background departure without bias correction
23	fgdep_nbcor_stdv	Standard deviation of background departure without bias correction
24	windspeed	Average of wind speed
25	windspeed_stdv	Standard deviation of wind speed
26	norm_andep	Average of normalised analysis fit
27	norm_andep_stdv	Standard deviation of normalised analysis fit
28	norm_fgdep	Average of normalised background fit
29	norm_fgdep_stdv	Standard deviation of normalised background fit
30	fso	Average of forecast sensitivity to observations
31	fso_stdv	stdv of forecast sensitivity to observations
32	norm_obs	Average of normalised observation
33	norm_obs_stdv	stdv of normalised observation
34	anso	Average of analyse sensitivity to observations
35	anso_stdv	stdv of analyse sensitivity to observations
40	fcdep1	Average of forecast departure 1
41	fcdep1_stdv	stdv of forecast departure 1
42	fcdep1	Average of forecast departure 2
43	fcdep1_stdv	stdv of forecast departure 2
44	norm_fcdep1	Average of normalised forecast fit 1
45	norm_fcdep1_stdv	stdv of normalised forecast fit 1
46	norm_fcdep2	Average of normalised forecast fit 2
47	norm_fcdep2_stdv	stdv of normalised forecast fit 2
48-100	Reserved	Reserved

Table 4: List of Observation diagnostics

Grib Code	Short name	Long Name
1	ERS-1	ERS 1
2	ERS-2	ERS 2
3	METOP-1	METOP-1
4	METOP-A	METOP-2
41	CHAMP	CHAMP
42	TERRA-SAR-X	TERRA-SAR-X
46	SMOS	SMOS
54	METEOSAT-7	METEOSAT 7
55	METEOSAT-8	METEOSAT 8
56	METEOSAT-9	METEOSAT 9
57	METEOSAT-10	METEOSAT 10
58	METEOSAT-1	METEOSAT 1
59	METEOSAT-2	METEOSAT 2
60	ENVISAT	ENVISAT
70	METEOSAT-11	METEOSAT 11
140	GOSAT	GOSAT
171	MTSAT-1R	MTSAT-1R
172	MTSAT-2	MTSAT-2
205	NOAA-14	NOAA 14
206	NOAA-15	NOAA 15
207	NOAA-16	NOAA 16
208	NOAA-17	NOAA 17
209	NOAA-18	NOAA 18
223	NOAA-19	NOAA 19
224	NPP	NPP
246	DMSP-13	DMSP 13
247	DMSP-14	DMSP 14
248	DMSP-15	DMSP 15
249	DMSP-16	DMSP 16
253	GOES-9	GOES 9
254	GOES-10	GOES 10
255	GEOS-11	GOES 11
256	GEOS-12	GOES 12
257	GEOS-13	GOES 13
258	GEOS-14	GOES 14
259	GEOS-15	GOES 15
260	JASON-1	JASON-1
261	JASON-2	JASON-2
281	QUIKSCAT	QUIKSCAT
282	TRMM	TRMM
283	CORIOLIS	CORIOLIS
285	DMSP17	DMSP 17
286	DMSP18	DMSP 18
500	FY-1C	FY-1C
501	FY-1D	FY-1D
510	FY-2	FY-2

Table 5: List of meteorological satellites codes: Part 1

Grib Code	Short name	Long Name
512	FY-2B	FY-2B
513	FY-2C	FY-2C
514	FY-2D	FY-2D
515	FY-2E	FY-2E
520	FY-3A	FY-3A
521	FY-3B	FY-3B
722	GRACE-A	GRACE-A
740	COSMIC-1	COSMIC-1
741	COSMIC-2	COSMIC-2
742	COSMIC-3	COSMIC-3
743	COSMIC-4	COSMIC-4
744	COSMIC-5	COSMIC-5
745	COSMIC-6	COSMIC-6
783	TERRA	TERRA
784	AQUA	AQUA
785	AURA	AURA
786	C-NOFS	C-NOFS
820	SAC-C	SAC-C

Table 6: List of meteorological satellites codes: Part 2

Grib Code	Short name	Long Name
0	HIRS	HIRS
3	AMSUA	AMSUA
4	AMSUB	AMSUB
6	SSM/I	SSM/I
9	TMI	TMI
10	SSMI/S	SSMI/S
11	AIRS	AIRS
15	MHS	MHS
16	IASI	IASI
17	AMSRE	AMSR-E
19	ATMS	ATMS
20	MVIRI	MVIRI
21	SEVIRI	SEVIRI
22	GOES	GOES Imager
24	MTSAT-1R	MTSAT-1R imager
27	CrIS	CrIS
30	WINDSAT	WINDSAT
40	MWTS	MWTS
41	MWHS	MWHS
42	IRAS	IRAS
43	MWRI	MWRI
102	GPSRO	GPSRO
103	GPSRO	GPSRO
172	GOMOS	GOMOS
174	MERIS	MERIS
175	SCIAMACHY	SCIAMACHY
202	GRAS	GRAS
207	SEVIRI_O3	SEVIRI O3
220	GOME-2	GOME-2
387	MLS	MLS
394	OMI	OMI
516	TANSO	TANSO
624	SBUV-2	SBUV-2
2000	AMV_WV_CLOUDY	AMV WV cloudy
2001	AMV_IR	AMV IR
2002	AMV_VIS	AMV VIS
2003	AMV_WVMIX	AMV WVMIX
2005	AMV_WV_Clear	AMV Water Vapor clear
2100	AMV_WV_6.2_cloudy	AMV WV 6.2 cloudy
2101	AMV_IR_ch1	AMV IR ch1
2102	AMV_VIS_ch1	AMV VIS ch1
2105	AMV_WV_6.2_clear	AMV WV_6.2 clear
2200	AMV_WV_7.3_cloudy	AMV WV 7.3 cloudy
2201	AMV_IR_ch2	AMV IR ch2
2202	AMV_VIS-2	AMV VIS-2
2205	AMV_WV_7.3_clear	AMV WV 7.3 clear
2300	AMV_WV_cloudy_ch3	AMV WV cloudy ch
2301	AMV_IR-10	AMV IR-10
2305	AMV_WV_clear_Ch3	AMV WV clear Ch3
2350	QUIKSCAT	QUIKSCAT
2150	SCAT	SCAT
2190	ASCAT	ASCAT

Table 7: List of meteorological satellite instruments codes

Grib Code	Short name	Long Name
1	TIROS-N	TIROS-N
2	NOAA-6/HIRS	NOAA-6/HIRS
3	NOAA-7/HIRS	NOAA-7/HIRS
4	NOAA-8/HIRS	NOAA-8/HIRS
5	NOAA-9/HIRS	NOAA-9/HIRS
6	NOAA-10/HIRS	NOAA-10/HIRS
7	NOAA-11/HIRS	NOAA-11/HIRS
8	NOAA-12/HIRS	NOAA-12/HIRS
9	NOAA-14/HIRS	NOAA-14/HIRS
10	NOAA-15/HIRS	NOAA-15/HIRS
11	NOAA-16/HIRS	NOAA-16/HIRS
12	NOAA-17/HIRS	NOAA-17/HIRS
13	NOAA-18/HIRS	NOAA-18/HIRS
14	NOAA-19/HIRS	NOAA-19/HIRS
15	METOP-A/HIRS	METOP-A/HIRS
1001	NOAA-15/AMSUA	NOAA-15/AMSUA
1002	NOAA-16/AMSUA	NOAA-16/AMSUA
1003	NOAA-17/AMSUA	NOAA-17/AMSUA
1004	NOAA-18/AMSUA	NOAA-18/AMSUA
1005	NOAA-19/AMSUA	NOAA-19/AMSUA
1006	NOAA-19/AMSUA	NOAA-19/AMSUA
1007	METOP-A/AMSUA	METOP-A/AMSUA
1008	AQUA/AMSUA	AQUA/AMSUA
2001	NOAA-15/AMSUB	NOAA-15/AMSUB
2002	NOAA-16/AMSUB	NOAA-16/AMSUB
2003	NOAA-17/AMSUB	NOAA-17/AMSUB
2004	NOAA-18/AMSUB	NOAA-18/AMSUB
2005	NOAA-18/AMSUB	NOAA-18/AMSUB
3001	NOAA-19/MHS	NOAA-19/MHS
3002	METOP-A/MHS	METOP-A/MHS
4001	GOES-5/IMAGER	GOES-5/IMAGER
4002	GOES-8/IMAGER	GOES-8/IMAGER
4003	GOES-9/IMAGER	GOES-9/IMAGER
4004	GOES-10/IMAGER	GOES-10/IMAGER
4005	GOES-11/IMAGER	GOES-11/IMAGER
4006	GOES-12/IMAGER	GOES-12/IMAGER
4007	METEOSAT-7/MVIRI	METEOSAT-7/MVIRI
4008	METEOSAT-8/SEVIRI	METEOSAT-8/SEVIRI
4009	METEOSAT-9/SEVIRI	METEOSAT-9/SEVIRI
4010	MTSAT-1R/IMAGER	MTSAT-1R/IMAGER
5001	ERS-2/GOME	ERS-2/GOME
5002	METEOSAT-8/SEVIRI	METEOSAT-8/SEVIRI
5003	METEOSAT-9/SEVIRI	METEOSAT-9/SEVIRI
5004	AURA/MLS	AURA/MLS
5005	AURA/OMI	AURA/OMI

Table 8: List of Reporttypes codes: Part 1

Grib Code	Short name	Long Name
5006	NOAA-9/SBUV	NOAA-9/SBUV
5007	NOAA-11/SBUV	NOAA-11/SBUV
5008	NOAA-14/SBUV	NOAA-14/SBUV
5009	NOAA-16/SBUV	NOAA-16/SBUV
5010	NOAA-17/SBUV	NOAA-17/SBUV
5011	NOAA-18/SBUV	NOAA-18/SBUV
5012	NOAA-19/SBUV	NOAA-19/SBUV
5013	METOP-A/GOME-2	METOP-A/GOME-2
5014	ENVISAT/SCIAMACHY	ENVISAT/SCIAMACHY
5015	ENVISAT/GOMOS	ENVISAT/GOMOS
5016	ENVISAT/MIPAS	ENVISAT/MIPAS
5017	Metror-3/TOMS	Metror-3/TOMS
5018	Nimbus-7/TOMS	Nimbus-7/TOMS
6001	ENVISAT/GOMOS	ENVISAT/GOMOS
6002	ENVISAT/MERIS	ENVISAT/MERIS
7001	METOP-A/GRAS	METOP-A/GRAS
7002	CHAMP	CHAMP
7003	GRACE-A	GRACE-A
7004	COSMIC-1	COSMIC-1
7005	COSMIC-2	COSMIC-2
7006	COSMIC-3	COSMIC-3
7007	COSMIC-4	COSMIC-4
7008	COSMIC-5	COSMIC-5
7009	COSMIC-6	COSMIC-6
8001	METEOSAT-2/AMV	METEOSAT-2/AMV
8002	METEOSAT-3/AMV	METEOSAT-3/AMV
8003	METEOSAT-4/AMV	METEOSAT-4/AMV
8014	METEOSAT-5/AMV	METEOSAT-5/AMV
8005	METEOSAT-6/AMV	METEOSAT-6/AMV
8006	METEOSAT-7/AMV	METEOSAT-7/AMV
8007	METEOSAT-8/AMV	METEOSAT-8/AMV
8008	METEOSAT-9/AMV	METEOSAT-9/AMV
8009	GMS-5/AMV	GMS-5/AMV
8010	MTSAT-1R/AMV	MTSAT-1R/AMV
8011	GOES-9/WV	GOES-9/WV
8012	GOES-10/AMV	GOES-10/AMV
8013	GOES-11/AMV	GOES-11/AMV
8014	GOES-12/AMV	GOES-12/AMV
8015	NOAA-15/AVHRR	NOAA-15/AVHRR
8016	NOAA-16/AVHRR	NOAA-16/AVHRR
8017	NOAA-17/AVHRR	NOAA-17/AVHRR
8018	NOAA-18/AVHRR	NOAA-18/AVHRR
8019	NOAA-19/AVHRR	NOAA-19/AVHRR
8020	TERRA/MODIS	TERRA/MODIS
8021	AQUA/MODIS	AQUA/MODIS
8022	FY-2C/IR	FY-2C/IR

Table 9: List of Reporttypes codes: Part 2

Grib Code	Short name	Long Name
9001	ERS/SCATT	ERS/SCATT
9002	ERS/SCATT	ERS/SCATT
9003	ERS-2/SCATT	ERS-2/SCATT
9004	QuickSCAT/SeaWind	QuickSCAT/SeaWind
9005	METOP-A/ASCAT	METOP-A/ASCAT
10001	DSMP-7/SSMI	DSMP-7/SSMI
10002	DSMP-8/SSMI	DSMP-8/SSMI
10003	DSMP-9/SSMI	DSMP-9/SSMI
10004	DSMP-10/SSMI	DSMP-10/SSMI
10005	DSMP-11/SSMI	DSMP-11/SSMI
10006	DSMP-13/SSMI	DSMP-13/SSMI
10007	DSMP-14/SSMI	DSMP-14/SSMI
10008	DSMP-15/SSMI	DSMP-15/SSMI
10009	DSMP-8/SSMI	DSMP-8/SSMI
10010	DSMP-9/SSMI	DSMP-9/SSMI
10011	DSMP-10/SSMI	DSMP-10/SSMI
10012	DSMP-11/SSMI	DSMP-11/SSMI
10013	DSMP-13/SSMI	DSMP-13/SSMI
10014	DSMP-14/SSMI	DSMP-14/SSMI
10015	DSMP-15/SSMI	DSMP-15/SSMI
11001	METOP-A/IASI	METOP-A/IASI
12001	AQUA/AIRS	AQUA/AIRS
13001	DMSP-16/SSMIS	DMSP-16/SSMIS
14001	TRMM/TMI	TRMM/TMI
15001	AQUA/AMSRE	AQUA/AMSRE
16001	Automatic-Land	Automatic-Land
16002	Manual-Land	Manual-Land
16003	Abbreviated-SYNOP	Abbreviated-SYNOP
16004	METAR	METAR
16005	DRIBU	DRIBU
16006	Automatic-SHIP	Automatic-SHIP
16007	Reduced-SHIP	Reduced-SHIP
16008	SHIP	SHIP
16009	Abbreviated-SHIP	Abbreviated-SHIP
16010	DRIBU-BATHY	DRIBU-BATHY
16011	DRIBU-TESAC	DRIBU-TESAC
16012	Ground-Based-GPS	Ground-Based-GPS
16013	Land-PILOT	Land-PILOT
16014	PILOT-SHIP	PILOT-SHIP
16015	American-WindProfilers	American-WindProfilers
16016	American-WindProfilers	American-WindProfilers
16017	European-WindProfilers	European-WindProfilers
16018	Japanese-WindProfilers	Japanese-WindProfilers
16019	TEMP-SHIP	TEMP-SHIP
16020	DROP-Sonde	DROP-Sonde
16021	Mobile-TEMP	Mobile-TEMP

Table 10: List of Reporttypes codes: Part 3

Grib Code	Short name	Long Name
16022	Land-TEMP	Land-TEMP
16023	ROCOB-TEMP	ROCOB-TEMP
16024	SHIP-ROCOB	SHIP-ROCOB
16025	European-WindProfilers	European-WindProfilers
16026	AIREP	AIREP
16027	CODAR	CODAR
16028	COLBA	COLBA
16029	AMDAR	AMDAR
16030	ACARS	ACARS
16031	PAOB	PAOB
16032	PAOB	PAOB
16033	SATOB_Temperature	SATOB_Temperature
16034	SATOB_Wind	SATOB_Wind
16035	SATOB_Temperature	SATOB_Temperature
16036	SATOB_Temperature	SATOB_Temperature
16037	SATEM_500km	SATEM_500km
16038	SATEM_500km	SATEM_500km
16039	SATEM_500km	SATEM_500km
16040	SATEM_500km	SATEM_500km
16041	SATEM_250km	SATEM_250km
16042	SATEM_250km	SATEM_250km
16043	SATEM_250km	SATEM_250km
16044	SATEM_250km	SATEM_250km
17001	Automatic_Land	Automatic_Land
17002	Manual_Land	Manual_Land
17003	Abbreviated_SYNOP	Abbreviated_SYNOP
17004	METAR	METAR
17005	DRIBU	DRIBU
17006	Automatic_SHIP	Automatic_SHIP
17007	Reduced_SHIP	Reduced_SHIP
17008	SHIP	SHIP
17009	Abbreviated-SHIP	Abbreviated-SHIP
17010	DRIBU-BATHY	DRIBU-BATHY
17011	DRIBU-TESAC	DRIBU-TESAC
17012	Ground-Based_GPS	Ground-Based_GPS
17013	Land-PILOT	Land-PILOT
17014	PILOT-SHIP	PILOT-SHIP
17015	American-Wind	American-Wind
17016	American-Wind	American-Wind
17017	European-Wind	European-Wind
17018	Japanese-Wind	Japanese-Wind
17019	TEMP-SHIP	TEMP-SHIP
17020	DROP-Sonde	DROP-Sonde
17021	Mobile-TEMP	Mobile-TEMP
17022	Land-TEMP	Land-TEMP
17023	ROCOB-TEMP	ROCOB-TEMP
17024	SHIP-ROCOB	SHIP-ROCOB

Table 11: List of Reporttypes codes: Part 4