Observation Pre-processor for WRF-Var

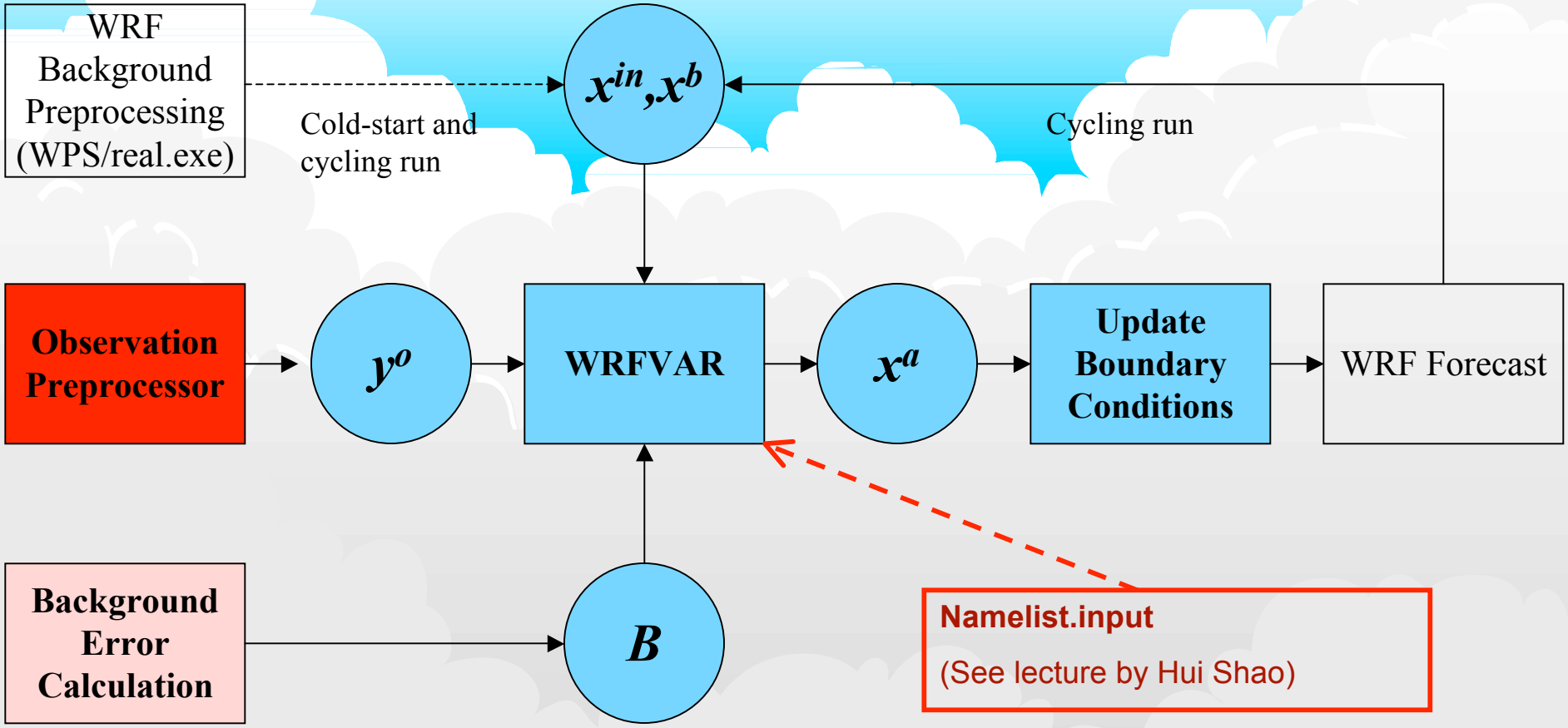
Yong-Run Guo

National Center for Atmospheric Research
P.O.Box 3000, Boulder, CO 80307

Presented in WRF Tutorial, 21 July 2008

Thanks to everyone in NCAR/MMM/WRFDA group, especially to Hui-Chuan Lin, Zhiquan Liu, Hans Huang, and Jim Bresch

WRF-Var in the WRF Modeling System



The observations are one of the important input files for WRF-Var, no observations will be no data assimilation.

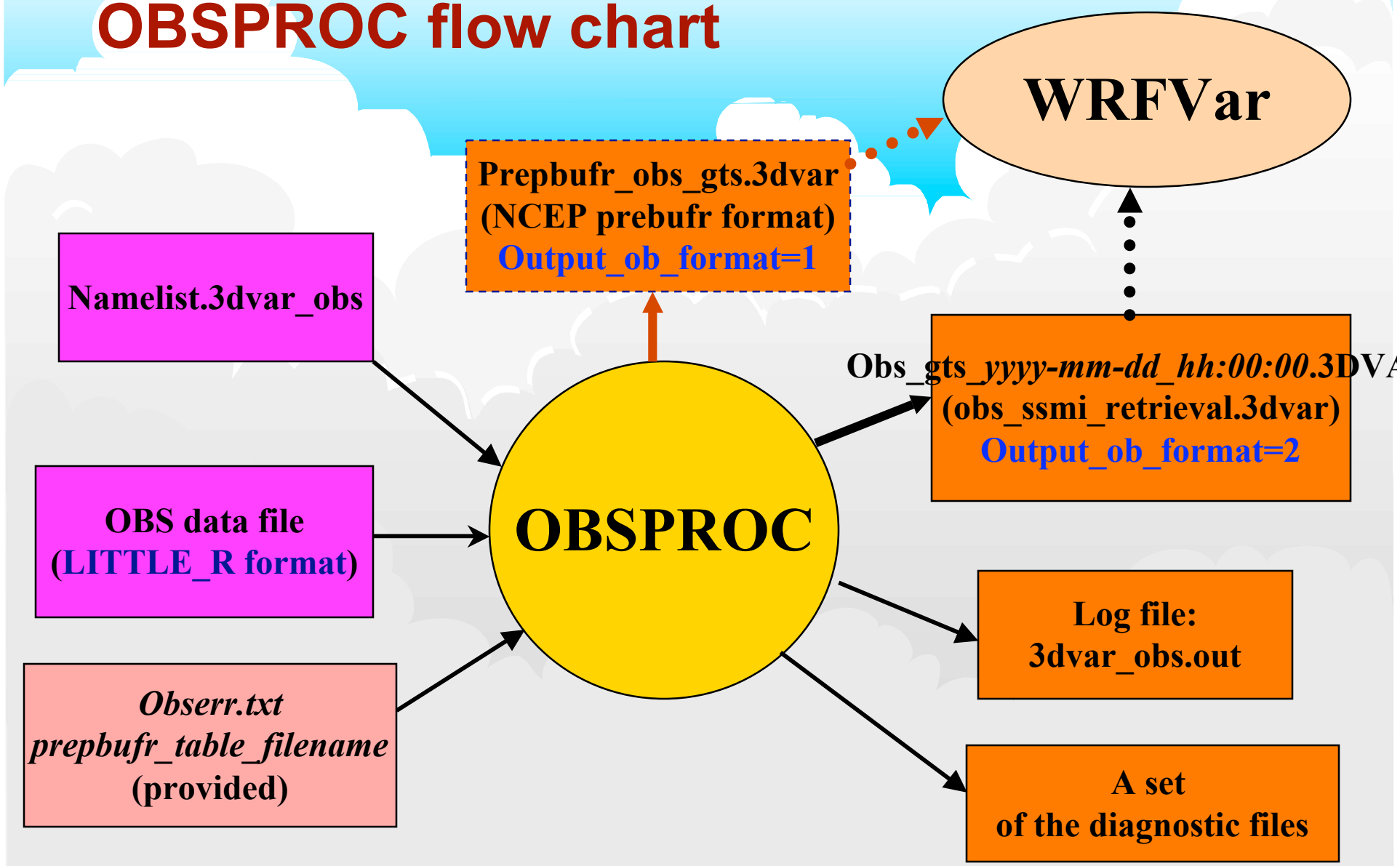
* In certain cases, no obs WRF-Var run is still useful (e.g. using *RIP4 diff* to plot increments).

Why do we need the OBS preprocessor?

To prepare the OBS data file suitable for WRF-Var needs (3DVAR, FGAT, 4DVAR, etc.)

- Screening of the conventional observations (time window, domain, duplication, etc.), and keep the necessary information for WRF-Var assimilation
- Assign the observation errors to each of the observations
- Do the basic quality control (gross check and consistent check)
- Save the OBS data processing time, *one OBS file* can be repeatedly used for multiple times of WRF-Var experiments.

OBSPROC flow chart



Input files to obsproc LITTLE_R format

- The ***input file*** to 3DVAR_OBSPROC is the LITTLE_R format (see below)

  This is a report-based format, so all types of the observation data can be easily '*cat*'ted together to form a monolithic file.

  It is easily to read, edit,…. with an ASCII file.

- Users' duty is just to convert their own observations in any format to the LITTLE_R format. Then that data can be processed by 3DVAR_OBSPROC, and an observation file suitable for assimilation with WRF-Var is created.

- Conventional observation data can also be downloaded from NCAR MSS:

  mss:/BRESCH/RT/DATA/*yyyymm*/obs.*yyyymmddhh*.gz (available starting from 2003040800Z, every 6 hours)

  mss:/RTFCST/ARCHIVE/RAIN/init.*yyyymmddhh*.tar (available starting from 1999012400Z to 2004072612Z, every 12 hours)

Structure of the LITTLE_R format

- *OBS decoded file* in LITTLE_R format containing *Reports*
- *A Report* containing *Records (header, data,…, and ending) and 3 tail integers* (3I7)
- *There are 3 types of Records, each of them contains a number of the fields*
  - The *fields* in the *header record* (Fortran format in parenthesis)
  - The *fields* in the *data record* (Fortran format in parenthesis)
  - The *fields* in the *ending record*

Input OBS (LITTLE_R) file to preprocessor

- *OBS decoded file* in LITTLE_R format containing *Reports*
- *Report* containing *Records (header, data,…, and ending) and 3 tail integers* (3I7)
- *Record* containing *fields*
  - The *fields* in the *header record* (Fortran format in parenthesis)
  - The *fields* in the *data record* (Fortran format in parenthesis)
  - The *fields* in the *ending record*

The details for each of records are described below:

The fields in the header record:

No	Field	No	Filed	No	Field
1	Latitude (f20.5)	2	Longitude (f20.5)	3	ID (a40)
4	Name (a40)	5	Platform (a40)	6	Source (a40)
7	Elevation (f20.5)	8	Num_vld_fld (i10)	9	Num_error (i10)
10	Num_warning (i10)	11	Seq_num (i10)	12	Num_dupd (i10)
13	Is_sound (L10)	14	Bogus (L10)	15	Discard (L10)
16	Valid_time%sut (i10)	17	Valid_time%julian (i10)	18	Valid_time%date_char(a20)
19	Slp%data (f13.5)	20	Slp%qc (i7)	21	Ref_pres%data (f13.5)
22	Ref_pres%qc (i7)	23	Ground_t%data (f13.5)	24	Ground_t%qc (i7)
25	SST%data (f13.5)	26	SST%qc (i7)	27	Psfc%data (f13.5)
28	Psfc%qc (i7)	29	Precip%data (f13.5)	30	Precip%qc (i7)
31	T_max%data (f13.5)	32	T_max%qc (i7)	33	T_min%data (f13.5)
34	T_min%qc (i7)	35	T_min_night%data (f13.5)	36	T_min_night%qc (i7)
37	P_tend03%data (f13.5)	38	P_tend03%qc (i7)	39	P_tend24%data (f13.5)
40	P_tend24%qc (i7)	41	Cloud_cvr%data (f13.5)	42	Cloud_cvr%qc (i7)
43	Celling%data (f13.5)	44	Celling%qc (i7)	45	Pw%data (f13.5)
46	Pw%qc (i7)	47	Tb19v%data (f13.5)	48	Tb19v%qc (i7)
49	Tb19h%data (f13.5)	50	Tb19h%qc (i7)	51	Tb22v%data (f13.5)
52	Tb22v%qc (i7)	53	Tb37v%data (f13.5)	54	Tb37v%qc (i7)
55	Tb37h%data (f13.5)	56	Tb37h%qc (i7)	57	Tb85v%data (f13.5)
58	Tb85v%qc (i7)	59	Tb85h%data (f13.5)	60	Tb85h%qc

The fields in the data record (Fortran format in parenthesis)

No	Field	No	Field
1	Pressure%data (f13.5)	2	Pressure%qc (i7)
3	Height%data 9f13.5)	4	Height%qc (i7)
5	Temperature%data (f13.5)	6	Temperature%qc (i7)
7	Dew_point%data (f13.5)	8	Dew_point%qc (i7)
9	Speed%data (f13.5)	10	Speed%qc (i7)
11	Direction%data (f13.5)	12	Direction%qc (i7)
13	U%data (f13.5)	14	U%qc (i7)
15	V%data (f13.5)	16	V%qc (i7)
17	RH%data (f13.5)	18	RH%qc (i7)
19	Thickness%data (f13.5)	20	Thickness%qc (i7)

The fields in the ending record

No	field	No	field	No	field	No	field
1	-777777.00000	2	0	3	-777777.00000	4	0
5	-888888.00000	6	0	7	-888888.00000	8	0
9	-888888.00000	10	0	11	-888888.00000	12	0
13	-888888.00000	14	0	15	-888888.00000	16	0
17	-888888.00000	18	0	19	-888888.00000	20	0

Note

1.The tail fields are not need to all filled in. For example, if no SSMI Tb (brightness temperature) available, the header record may only have 46 fields.

2.For certain type of observations, the some of the fields in data record are just used as the storage, not the actual data as the field's name. For example, for QuikScat SeaWind, the fields: U%data and V%data are the speed and direction errors, respectively.

3.For certain types of observations, such as GPSREF, etc., the observation data are not the wind, temperature, moisture, etc., so specific arrangements are made with the fields to hold the refractivity, perigee point location, etc.

COSMIC GPS data format

Level record in little_r file:

Press.	Geo height	Temp.	Dew-p	speed	Dir.	u	v	rh	thick
Miss.	height	miss	Refractivity	Impact parameter	Azimuth angle	latitude	longitude	Bending angle	Opt. bending

press	latitude	longitude	height	temp	Refractivity	Azimuth angle	Impact parameter *1.e-3	Bending angle*1.e7
Ref. Atmos	N	E	m	miss	N	Deg.		rad

** The NCEP ADP observation data (upper air from 20 December 1972 to 28 February 2007 and surface data from 10 February 1975 to 28 February 2007) may also be downloaded.
http://www.mmm.ucar.edu/mm5/mm5v3/data/free_data.html

Then use a MM5 utility to convert these ADP data to LITTLE_R format data (Recently we did not try this procedure, and not sure if it is still working).

Output WRFDA/var/obsproc

1. **NCEP prepBUFR format (not fully tested yet!)**

In obsproc namelist.3dvar_obs

&record9

prepbufr_output_filename='*prepbufr_obs_gts.3dvar*'

prepbufr_table_filename='*prepbufr_table_filename*'

output_ob_format=1 (or 3)

In WRFVar namelist.input,

&wrfvar3

ob_format=1,

The prepbufr file is a binary file, and an endian dependent. See

http://www.nco.ncep.noaa.gov/sib/decoders/BUFRLIB/toc/cwordsh/

prepbufr observation data from NCEP ftp site:

You can download the NCEP real-time prepbufr observation data from

ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/

The NCEP GDAS archived prepbufr data can be downloaded from

http://nomads.ncdc.noaa.gov/data/gdas

Or you can download the archived prepbufr observation data from NCAR MSS:

mss:/LIUZ/GDAS/*yyyymm*/*yyyymmddhh*/gdas1.t*hh*z.prepbufr.unblok.nr
(using var/obsproc/lib/cwordsh/cwordsh to add the blocking informationto the BUFR file in the little-endian (Linux/PGI) system) or

mss:/LIUZ/GDAS/*yyyymm*/*yyyymmddhh*/gdas1.t*hh*z.prepbufr.nr
(can be directly used in big-endian machine, such as IBM)

The archived data are available starting from 20060718Z to date, every 6 hours.

2, ASCII format --- Easy to manipulate: read, edit,etc. and endian independent (fully supported)

In obsproc namelist.3dvar_obs

&record9

output_ob_format=2 (or 3)

```
;     Select the obs_gts (ASCII) files used for 3DVAR, FGAT,
        and 4DVAR:
 use_for             = '3DVAR',  ; '3DVAR' obs file, same as
                                 before, default
                                 ; 'FGAT ' obs files for FGAT
                                 ; '4DVAR' obs files for 4DVAR
;      num_slots_past and num_slots_ahead are used ONLY for
FGAT     and 4DVAR:
num_slots_past    = 3, ; the number of time slots before
                         time_analysis
num_slots_ahead   = 3, ; the number of time slots after
                         time_analysis
```

In WRFVar namelist.input,

&wrfvar3

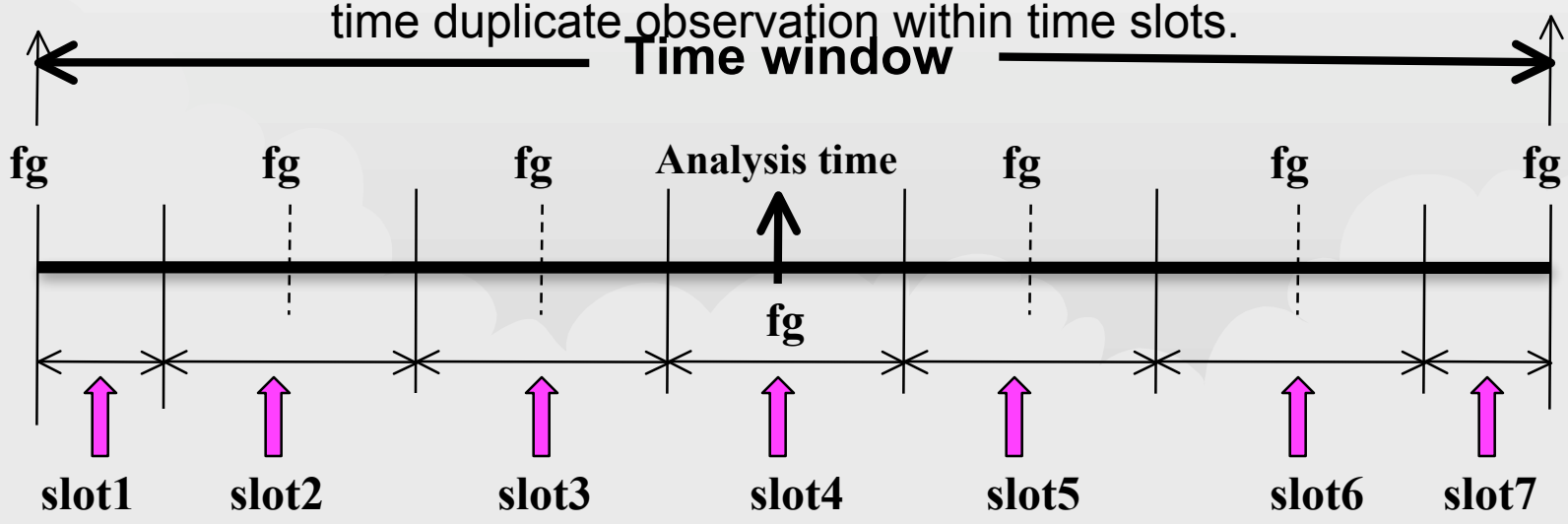
ob_format=2,

3DVAR, FGAT, and 4DVAR have the different requirements:

3DVAR ---- not allowed the time duplicate observation within time window

FGAT ---- multiple time slots within the time window, but not allowed the time duplicate observation within time window
(First Guess at Appropriate Time)

4DVAR ---- multiple time slots within the time window, but not allowed the time duplicate observation within time slots.



Output files for WRFVar

For 3DVAR,

obs_gts_yyyy-mm-dd_hh:00:00.3DVAR

For FGAT,

obs_gts_yyyy-mm-dd_hh:00:00.FGAT

For 4DVAR,

obs_gts_yyyy-mm-dd_hh:00:00.4DVAR

Types of observations to be processed

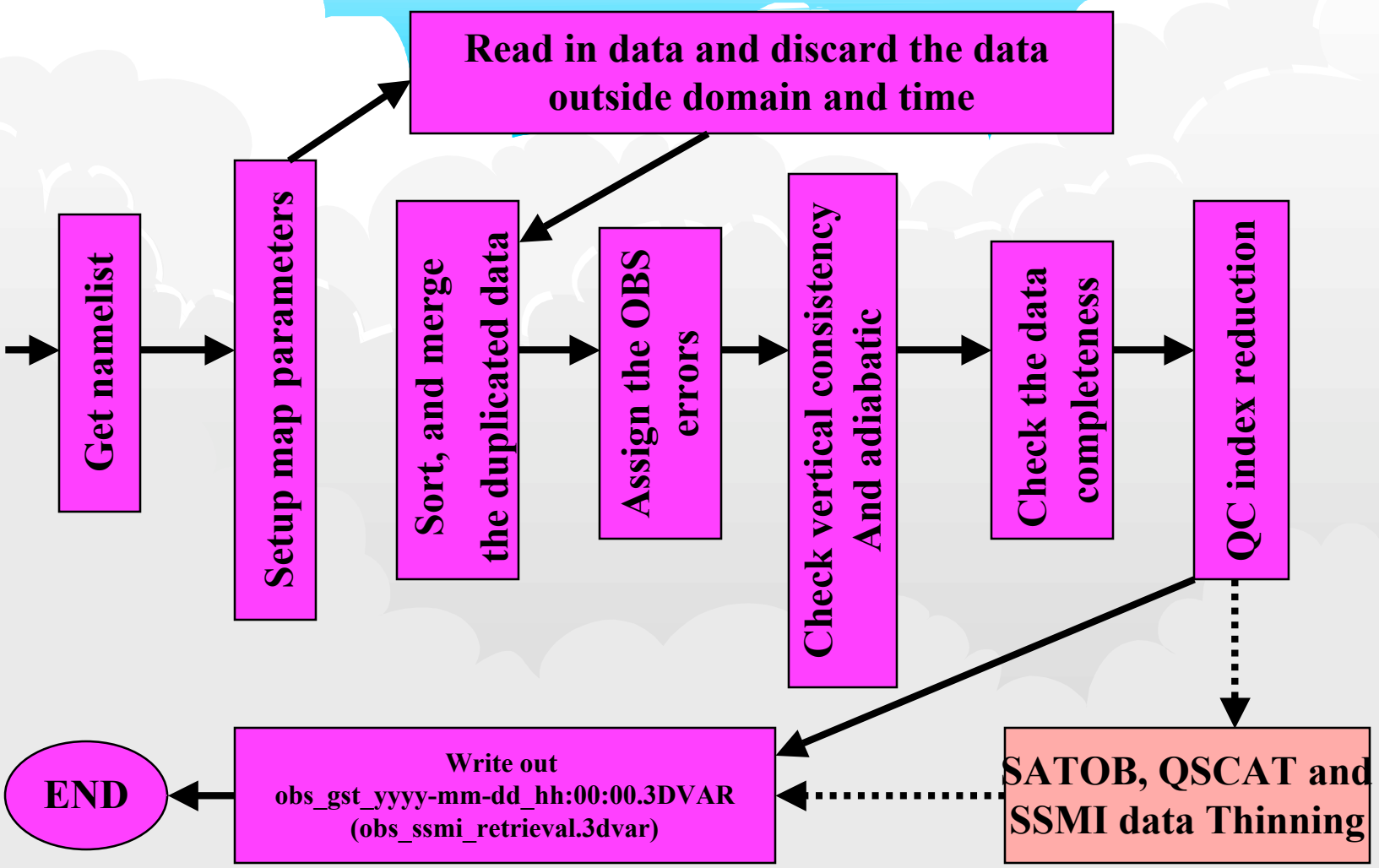
→18 types (SYNOP, SHIPS, METAR, TEMP , AIREP, PILOT , AMDAR, PROFL, SATOB, SATEM, SSMT1, SSMT2, SSMI, GPSPW, GPSZD, GPSRF, QSCAT, BOGUS)

Each type of the observations is identified by its WMO code in WRF-Var. If the standard WMO code is not available to a new data type, user should assign a 3-digit code to that data type.

WMO code for each type of observations

Name	WMO code	WMO code name
SYNOP	12, 14	SYNOP, SYNOP MOBIL
SHIP	13	SHIP
METAR	15, 16	METAR, SPECI
PILOT	32, 33, 34	PILOT, PILOT SHIP, PILOT MOBIL
SOUND	35, 36, 37, 38	TEMP, TEMP SHIP, TEMP DROP, TEMP MOBIL
AMDAR	42	AMDAR
SATEM	86	SATEM
SATOB	88	SATOB
AIREP	96, 97	AIREP
GPSPW	111	GPSPW (Ground-based GPS precipitable water)
GPSZD	114	GPSZD (Ground-based GPS Zenith Total Delay)
GPSRF	116	GPSRF (Space-based GPS Refractivity)
SSMT1	121	SSMT1
SSMT2	122	SSMT2
SSMI	125	SSMI
PROFL	132	WIND PROFILER
BOGUS	135	TCBOU (Typhoon bogus), BOGUS (other bogus)
QSCAT	281	Quik SCAT level-2B SeaWind
OTHER		UNKNOWN

Flow chart of OBS preprocessor



Tasks of the OBS preprocessor: obsproc

1, To perform a time-windowed and, in case of regional application (domain_check_h = .TRUE.), geographically-filtered dump of the ingested observations

Currently, there is *no time-check for observation data in WRF-Var assimilation code*, so to select the observation data within a suitable time-window must be performed in 3DVAR_OBSPROC.

For the regional application with the IPROJ = 1 (Lambert conformal), 2 (Polar Stereographic), or 3 (Mercator), there is a geographic-filtered performed based on the model domain settings. For the global application of WRF-Var, it should set IPROJ = 0 and no geographic-filtered is performed.

Gross check during the data ingestion:

•Ignore the data with the invalid WMO code.

•Any data values in header record > 888887 or < -888887 or pressure%data <= 0.0, etc., will be regarded as missing.

•Elevations for SHIP and BUOY data outside the Great Lakes are always set to zero. If the pressure < 85,000 Pa for SHIP and BUOY, the data are tossed out.

•Gross pressure/height consistent check based on the reference atmosphere defined by namelist variables: base_pres, base_temp, and base_lapse

•If both pressure and height are missing, the whole data are discarded.

•……

Tasks of the OBS preprocessor: 3DVAR_OBSPROC (cont.)

2, Sort and merge the duplicated data

- To retrieve the pressure or height based on the observed information with the hydrostatic assumption
- To remove the duplicate reports of observations: for 3DVAR and FGAT only observations nearest to the analysis time are kept while for 4DVAR, the observations nearest to the central time of each of the slots are kept.
- To re-order (from bottom to top) and merge the data reports with the same platform, time, and location based on the pressure.

Tasks of the OBS preprocessor: 3DVAR_OBSPROC (cont.)

3, To assign the observation errors to the different types of observations

Observations errors

→ NCEP OBS error (Parrish and Derber 1992)

→US Air Force (AFWA) OBS error file

→Directly from the observation reports

4, To perform the quality control (QC) for soundings

- Vertical consistency check: super adiabatic check and wind shear check
- Dry convective adjustment
- To discard the data above the model top ($p<p_{top}$) in the upper-air observations (remove_above_lid = .TRUE.)

Tasks of the OBS preprocessor: 3DVAR_OBSPROC (cont.)

8, To complete thinning with the SATOB, SSMI, and QSCAT data

The data points nearest to the model grid-points will be picked up for assimilation for SATOB, SSMI, and QSCAT.

9, To write out the OBS files in ASCII format as the WRF-Var input

→ GTS data (*obs_gts_yyyy-mm-dd_hh:00:00.3DVAR*): pressure, Wind, height, temperature, dew-point, RH, thickness, etc.

→ SSMI data (*obs_ssmi_retrieval.3dvar*): PW and surface wind speed

10, To plot the distribution for each type of observations

→MAP_plot

Output files

1, *Obs_gts_yyyy-mm-dd_hh:00:00.3DVAR* and *obs_ssmi_retrieval.3dvar*

Header: the information for this OBS file and data format

Data : *header* record and *data* records for each of levels

- These are the OBS input file to WRF-Var program
- obs_ssmi_retrieval.3dvar needed only when SSMI retrieval data available
- These files can be used as input to MAP_plot to obtain the gmeta plot file with NCAR GRAPHICS

2, 3dvar_obs.out ---- a program execution log file

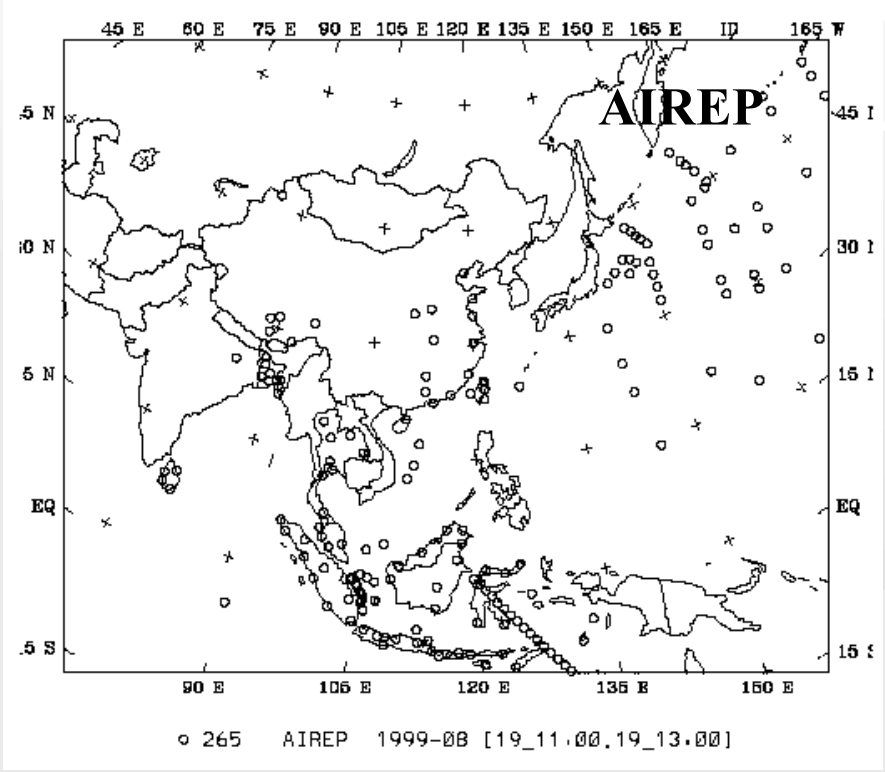
The printing out from the program execution. It can used to monitor the execution and to identify the troubles if any

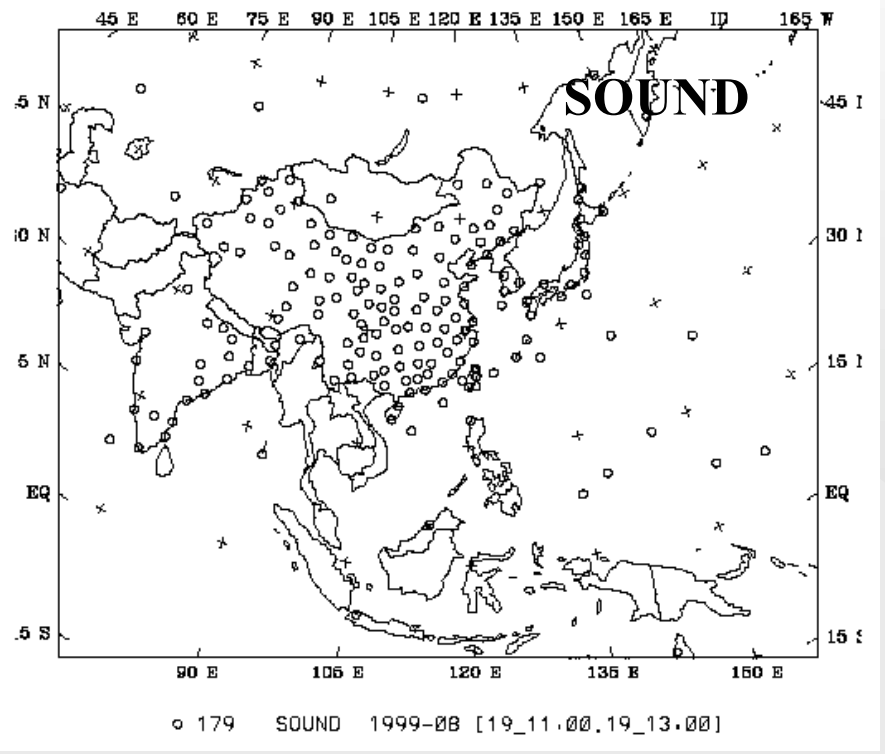
3, Diagnostic files depended on the print switches in namelist

How to plot the OBS distribution?

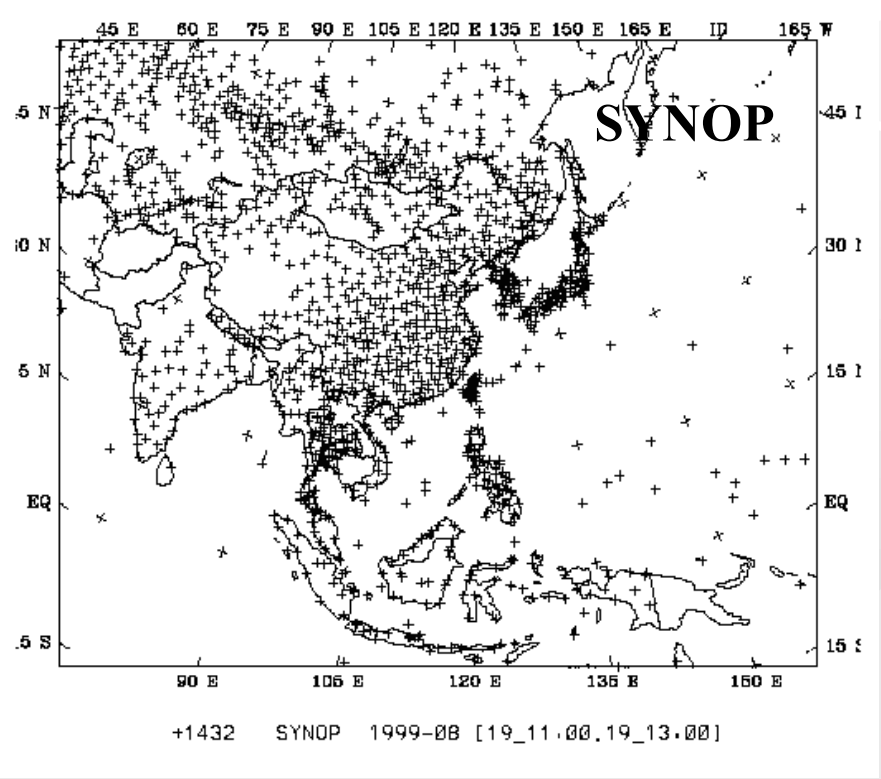
- Go to the directory *var/obsproc/MAP_plot*
- Modify the shell script *Map.csh*
  - » To fill in *TIME_ANALYSIS*, etc., and *OBSDATA* file name
- Run shell script *Map.csh*
  - » You will have a gmeta file: *gmeta.${TIME_ANALYSIS}* to show the the distribution of observations contained in *OBSDATA* file.

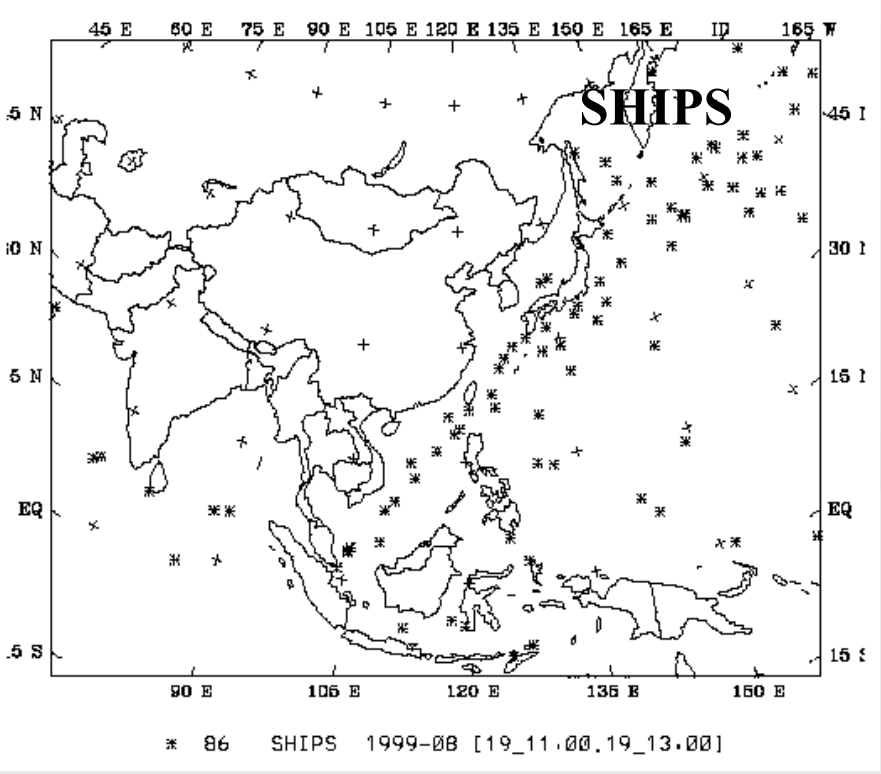
Distribution for each type of observations





Distribution for each type of observations





Distribution for each type of observations