

Screening and Minimization

Hirlam-B Training Week on HARMONIE system
Norrköping, Wednesday, 17 September, 2014

Martin Ridal



Screening – last chance to be accepted

The main emphasis on rejecting crude observation errors and on removing redundant information. Select observations suitable for data assimilation.

Can the observation be accepted and used by the model?

- Control the quality of the data
- Remove wrong data or data we already have

scr/Screening

The SCREENING executable MASTERODB is launched from Screening through:

```
$MPPEXEC $BINDIR/$MODEL -c002 -maladin -vmeteo -  
eMIN1 -t001 -ft0 -asli -procs ${NPROC} || exit
```

- c: configuration (screening = 002)**
- v: version of the code (always “meteo” for ARPEGE/ALADIN/HARMONIE)**
- m: aladin for LAM**
- e: experiment name (MIN1 for instance)**
- t: time-step length (“1.” for screening)**
- f: duration of the integration (“t0” or “h0” for screening)**
- a: dynamical scheme (does not matter for screening, Eulerian = “eul” or semi-Lagrangian = “sli”)**
- procs: needs number of processors \${NPROC}**

scr/Screening

Needed Input

- x ECMA ODB (observation, sigma-o)
- x First guess (ICMSHMIN1INIT, ICMSHMIN1IMIN, ICMRFGMIN10000, ELSCFGMIN1ALBC000, ELSCFGMIN1ALBC)
- x Costants, statistics
- x Namelists

Output

CCMA ODB and UPDATED ECMA ODB

scr/Screening needed input: Namelists

```
NAMEL=namelist_screening
```

```
#--- prepare namelist
```

```
sed -e "s/NBPROC/${NPROC}/g" \
    -e "s/NPROCX/${NPROCX}/g" \
    -e "s/NPROCY/${NPROCY}/g" \
    -e "s/NBZONVAR_EW=NBZONVAR_EW/NBZONVAR_EW=${NBZONVAR_EW}/g" \
    -e "s/LVARBC=LVARBC/LVARBC=${LVARBC}/g" \
    ${HM_LIB}/nam/$NAMEL > fort.4
```

Change in *nam/harmonie_namelists.pm* %screening -> mods in namelist_screening

scr/Screening needed input: Namelists

namelist_ screening

(NAMSCC and NAMOBS
contains tunable parameters)

&NAMCTO

LSCREEN=.TRUE.,

LOBS=.TRUE.,

LSIMOB=.FALSE.,

LOBSC1=.TRUE.,

/

...

&NAMSCC

RFIND_AIREP=25000.,

RFIND_RAD1C(0)=80000.,

RFIND_RAD1C(1)=80000.,

RFIND_RAD1C(15)=80000.,

RFIND_RAD1C(16)=80000.,

RFIND_RAD1C(2)=80000.,

RFIND_RAD1C(3)=80000.,

RFIND_RAD1C(4)=80000.,

RFIND_SATAM=45000.,

RFIND_SATOB=45000.,

RMIND_RAD1C(0)=60000.,

RMIND_RAD1C(1)=60000.,

RMIND_RAD1C(15)=40000.,

RMIND_RAD1C(16)=60000.,

RMIND_RAD1C(2)=60000.,

RMIND_RAD1C(3)=60000.,

RMIND_RAD1C(4)=40000.,

RMIND_SATAM=25000.,

RMIND_SATOB=25000.,

/

&NAMOBS

LAMSUB_FULL=.TRUE.,

LCACHMT=.TRUE.,

LCAPACH=.TRUE.,

LHRRH2=.TRUE.,

LMKCMARPL(2,9)=.TRUE.,

LRSTBIAS_ODB=.FALSE.,

LSCATT_NEUTRAL=.TRUE.,

LSLRW10=.TRUE.,

LSOE=.TRUE.,

NOBSHOR=201,

/

&NAMJO

SIGMAO_COEF=0.9,

/

scr/Screening needed input: Constants, statistics

```
cp $WRK/VARBC.cycle ./VARBC.cycle
```

```
ln -s ${HM_LIB}/const/bias_corr/correl.dat correl.dat
```

```
ln -s ${HM_LIB}/const/bias_corr/sigmab.dat sigmab.dat
```

```
ln -s ${HM_LIB}/const/bias_corr/rszcoef_fmt rszcoef_fmt
```

```
ln -s ${HM_LIB}/const/bias_corr/errgrib0scr errgrib
```

```
ln -s ${HM_LIB}/const/bias_corr/cstlim_noaa.dat cstlim_noaa.dat
```

```
ln -s ${HM_LIB}/const/bias_corr/var.amv* .
```

```
#--- RADIOSONDE BIAS CORRECTIONfor tabfile in table1 table2 table3;do
```

```
  cp ${HM_LIB}/const/bias_corr/$tabfile .
```

```
done
```

SCREENING of observations



!**** SUBROUTINE DECIS - CONTROLS IFS OBS. SCREENING DECISIONS

! HEIKKI JARVINEN ECMWF 15/MAR/1995

! DRASKO VASILJEVIC ECMWF 21/10/1999

!* PURPOSE

! CONTROLS IFS OBS. SCREENING DECISIONS

! DECIS IS CALLED BY SUBROUTINE SCREEN

!* EXTERNALS

! GETDB- GET DB DATA

! PUTDB- PUT DB DATA

! PRECH - PRELIMINARY CHECKS: RDB FLAGS, REPORTING PRACTICE ETC.

! BLACK - BLACKLISTING

! FIRST - FIRST GUESS CHECK

! VERO - VERTICAL CONSISTENCY OF MULTILEVEL REPORTS

! MISCE - A MISCELLANEOUS COLLECTION OF INDEPENDENT CHECKS

! GLOBA - MAKE GLOBAL TIME-LOCATION ARRAYS FOR SELECTED PLATFORMS

! DUPLI - CHECK FOR DUPLICATED AUREP REPORTS

! REDUN - ALL REDUNDANCY CHECKS

! THINN - THINNING OF DATA

1. → Preliminary check of observations

→ *Check of completeness of the reports*

→ *Check if station altitude is present*

→ *Check of the reporting practice for SYNOP & TEMP
mass observations [Ps & Φ]*

2. → Blacklisting

A blacklist is applied to discard observations of known poor quality and/or that cannot be properly handled by the data assimilation. A selection of variables for assimilation is done using the data selection part of the blacklist file and the information hard-coded in Arpege/Aladin (orographic rejection limit, land-sea rejection...).

Decisions based on the blacklist are feedback to the CMA

2. → Blacklisting

src/bla/mf_blacklist.b

```
if (OBSTYP = satob) then
! Geographical blacklist by satellites
  if (STATID) in (" 171") and (abs(LAT)>50. or (LON>-170. and LON<90.)) then fail(CONSTANT); endif;
  if (STATID) in (" 172") and (abs(LAT)>50. or (LON>-165. and LON<95.)) then fail(CONSTANT); endif;
endif; ! satob

if (OBSTYP = radar) then
! remove reflectivity
  if ( VARIAB = refl ) then
    if (statid = ' 07569') and (elevation < 2.6) then fail(CONSTANT); endif;
    if (statid = ' 07745') and (elevation < 2.8) then fail(CONSTANT); endif;
    if (statid = ' 07629') and (elevation < 2.6) then fail(CONSTANT); endif;
    if (statid = ' 07572') then fail(CONSTANT); endif;
  endif;

  if (statid = ' 02606') then fail(CONSTANT); endif; ! Remove the radar Ängelholm until it is upgraded

! To run radar in passive mode
  fail (EXPERIMENTAL);
endif; ! Radar
```

3. → Background quality control

- Flags are assigned to observations by checking:

$$\frac{y - H(x_b)}{\sigma_b} \Big)^2 > \text{flaglimits} * \lambda$$

$$\text{where } 1 + \frac{\sigma_o^2}{\sigma_b^2} \equiv \lambda$$

1: probably correct

2: probably incorrect `flaglimits(i=1,3)` `DEFRUN (RBGQC)`

3: incorrect

from file `errgrib` and interpolated - σ^b

from ECMA- σ^o

`bator_init_mod.F90` (+ tuning through `SIGMAO_COEF` in `batnam`)

4. → Vertical consistency of multilevel reports

→ *The duplicated levels, in multi-level reports, are removed from the reports*

→ *If 4 consecutive layers are found to be of suspicious quality then these layers are rejected*

5. → Removal of duplicated reports

→ *In case of co-located airep reports of the same observation types (time, position), some or all of the content of one of the reports is rejected*

6. → Redundancy check

→ *It is performed for the active reports that are co-located and originate from the same station*

→ *For land **Synop**, the report closest to the centre of the screening time window with most active data is retained*

→ *The ship **Synop** are redundant if the moving platforms are within a circle of 1 degree radius*

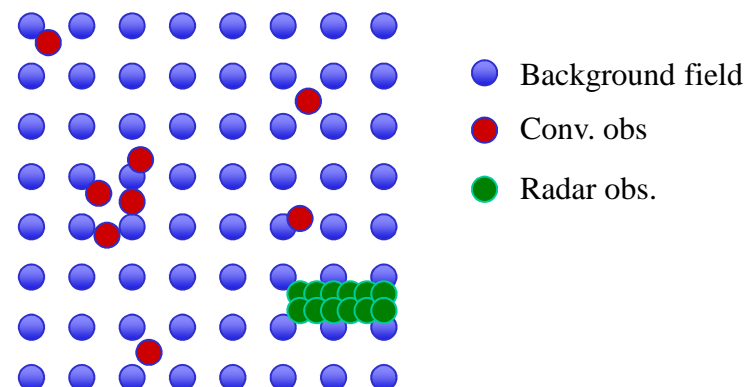
(arp/obs_preproc/sufglim.F90: RSHIDIS = 111000._JPRB)

→ ***Temp** and **Pilot** from same station are considered at the same time in the redundancy check*

→ *A **Synop** mass observation is redundant if there are any **Temp** geopotential height observations (made in the same time and the same station) that are no more than 50hPa above the **Synop** mass observation*

7. → Thinning

- High resolution data needs to be reduced
 - Reduce correlated errors
 - Reduce the amount of data
 - The model cannot make use of the data
- Thinning procedure
 - Decide on a minimum distance between observations
 - Decide which observation to keep (clear or cloudy, land or sea...)
- Alternative: Super observations
 - Averaging over an area
 - Needs to be done clever (e.g. not mixing land and sea)



NAMSCC

&NAMSCC

RFIND_AIREP=25000.,
RFIND_RAD1C(0)=80000.,
RFIND_RAD1C(1)=80000.,
RFIND_RAD1C(15)=80000.,
RFIND_RAD1C(16)=80000.,
RFIND_RAD1C(2)=80000.,
RFIND_RAD1C(3)=80000.,
RFIND_RAD1C(4)=80000.,
RFIND_SATAM=45000.,
RFIND_SATOB=45000.,
RMIND_RAD1C(0)=60000.,
RMIND_RAD1C(1)=60000.,
RMIND_RAD1C(15)=40000.,
RMIND_RAD1C(16)=60000.,
RMIND_RAD1C(2)=60000.,
RMIND_RAD1C(3)=60000.,
RMIND_RAD1C(4)=40000.,
RMIND_SATAM=25000.,
RMIND_SATOB=25000.,

/

Screening output

IN LOGFILE:

Valuable summary about screening decisions can be found in HM_Date_YYYYMMDDHH.html:

- Look for “**SCREENING STATISTICS**” to get:

- STATUS summary**

- EVENT summary**

- Number of variables, departures and missing departures**

- Diagnostic JO-table**

- + CCMA ODB and updated ECMA ODB**

Screening output

STATUS SUMMARY OF REPORTS:

OB.TYP	REPORTS	ACTIVE	PASSIVE	REJECTED	BLACKLISTED
1	2028	807	0	1221	0
2	22	22	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	9	9	0	0	0
6	0	0	0	0	0
7	1705	60	0	1645	976
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	0	0	0	0	0
12	0	0	0	0	0
13	37632	0	0	37632	37632
14	0	0	0	0	0
15	0	0	0	0	0
<hr/>					
TOT	41396	898	0	40498	38608

Screening output

EVENT SUMMARY OF REPORTS:

OB.TYP:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	29	0	0	0	0	0	1614	0	0	0	0	0	37632	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1192	0	0	0	0	0	31	0	0	0	0	0	0	0	0

1=NO DATA IN THE REPORT

2=ALL DATA REJECTED

3=BAD REPORTING PRACTICE

4=REJECTED DUE TO RDB FLAG

5=ACTIVATED DUE TO RDB FLAG

6=ACTIVATED BY WHITELIST

7=HORIZONTAL POSITION OUT OF RANGE

8=VERTICAL POSITION OUT OF RANGE

9=TIME OUT OF RANGE

10=REDUNDANT REPORT

11=REPORT OVER LAND

12=REPORT OVER SEA

13=MISSING STATION ALTITUDE

14=MODEL SUR. TOO FAR FROM STAT. ALT.

15=REPORT REJECTED THROUGH THE NAMELIST

16=FAILED QUALITY CONTROL

Screening output

EVENT SUMMARY OF DATA:

- | | |
|---|--|
| 1=MISSING VERTICAL COORDINATE | 18=DUPLICATED DATUM/LEVEL |
| 2=MISSING OBSERVED VALUE | 19=TOO MANY SURFACE DATA/LEVELS |
| 3=MISSING FIRST GUESS VALUE | 20=MULTI LEVEL CHECK |
| 4=REJECTED DUE TO RDB FLAG | 21=LEVEL SELECTION |
| 5=ACTIVATED DUE TO RDB FLAG | 22=VERTICAL CONSISTENCY CHECK |
| 6=ACTIVATED BY WHITELIST | 23=VERTICAL COORDINATE CHANGED FROM Z TO P |
| 7=BAD REPORTING PRACTICE | 24=DATUM REJECTED THROUGH THE NAMELIST |
| 8=VERTICAL POSITION OUT OF RANGE | 25=COMBINED FLAGGING |
| 9=REFERENCE LEVEL POSITION OUT OF RANGE | 26=DATUM REJECTED DUE TO REJECTED REPORT |
| 10=TOO BIG FIRST GUESS DEPARTURE | 27=VARIATIONAL QC PERFORMED |
| 11=TOO BIG DEPARTURE IN ASSIMILATION | 28=OBS ERR INCREASED |
| 12=TOO BIG OBSERVATION ERROR | 29=CLOUD CONTAMINATION |
| 13=REDUNDANT DATUM | 30=RAIN CONTAMINATION |
| 14=REDUNDANT LEVEL | 31=AEROSOL CONTAMINATION |
| 15=REPORT OVER LAND | |
| 16=REPORT OVER SEA | |

Minimization

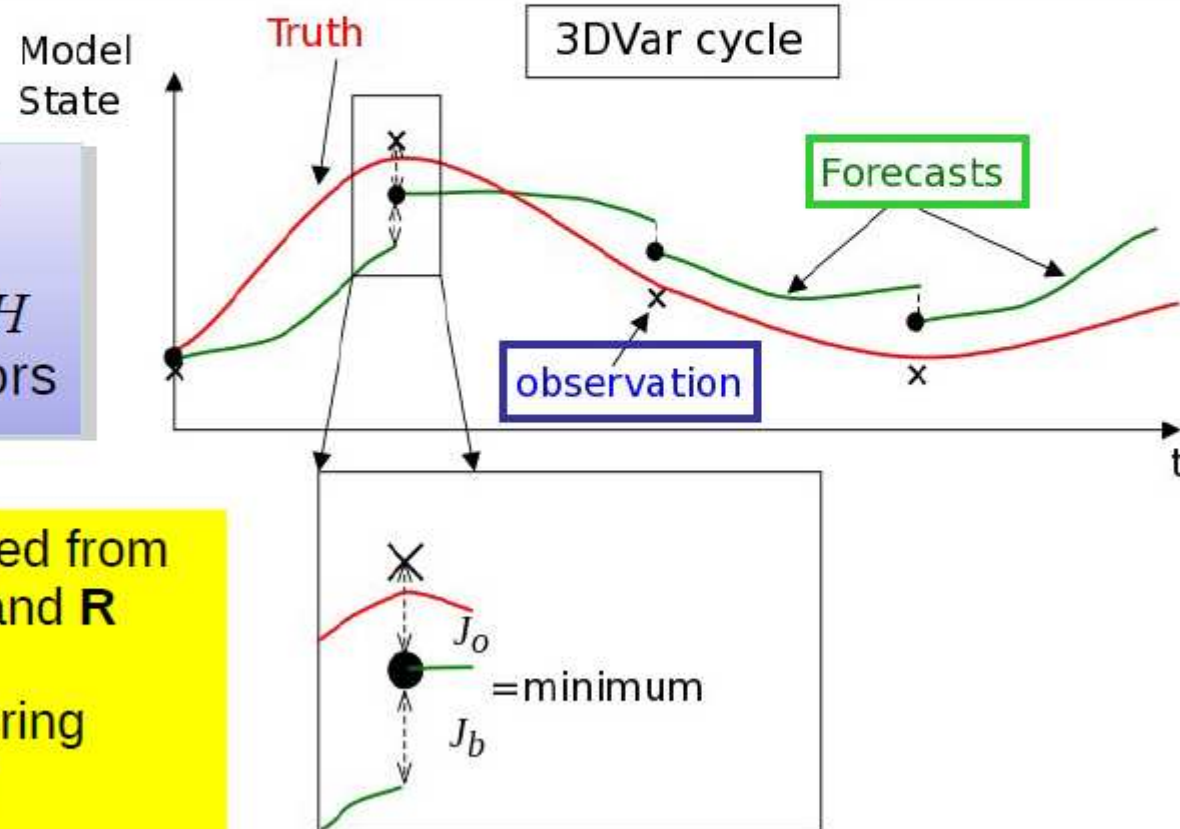
Find the best estimate of the true atmospheric state by minimizing the cost function

Data assimilation - 3D-Var

$$J(\mathbf{x}) = J_b + J_o = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{y}^o - H[\mathbf{x}])^T \mathbf{R}^{-1}(\mathbf{y}^o - H[\mathbf{x}])$$

- **B**, **R** background and observations error covariances matrices, H NL observation operators

The analysis \mathbf{x}^a is deduced from $\text{grad}(J)=0$, supposing **B** and **R** uncorrelated, unbiased innovations and considering linearized versions of H



$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H[\mathbf{x}])^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}])$$

Cost function and its gradient

$$\nabla J(\mathbf{x}) = 2\mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) - 2H^T \mathbf{R}^{-1} (\mathbf{y} - H[\mathbf{x}])$$

We need:

- **Observations** (y), **background or first-guess** (x_b), observation operator (H), **background** (B) and **observation** (R) error statistics;
- Minimizer - M1QN3 (Gilbert and Lemarechal, INRIA) ;
- System to do/realize what we need - **Harmonie**.

scr/Minim

The minimization executable MASTERODB is launched from Minim through:

```
$MPPEXEC $BINDIR/$MODEL -c131 -maladin -vmeteo -eMIN1 -t001  
-ft0 -asli -procs ${NPROC} || exit
```

Needed Input

- x CCMA ODB (observations, departures, sigma-o)
- x First guess (ICMSHMIN1INIT, ICMSHMIN1IMIN, ICMRFBMIN10000, ELSCFMIN1ALBC000, ELSCFMIN1ALBC)
- x Costants, statistics
- x Namelists

Output

Updated CCMA ODB and analysis (MXMIN1999+0000)

scr/Minim needed input: Constants, statistics

```
gunzip -c ${HM_LIB}/const/jb_data/$f_JBCV.gz > stabal96.cv  
gunzip -c ${HM_LIB}/const/jb_data/$f_JBBAL.gz > stabal96.bal  
${HM_LIB}/const/bias_corr/rszcoef_fmt  
${HM_LIB}/const/sat_const/rt_coef_atovs_newpred_ieee.dat  
${HM_LIB}/const/sat_const/rmtberr_airs.dat rmtberr_airs.dat  
${HM_LIB}/const/sat_const/var.sat.noaa_chanspec.11  
${HM_LIB}/const/sat_const/var.sat.noaa_rmtberr.11  
${HM_LIB}/const/bias_corr/cstlim_noaa.dat  
${HM_LIB}/const/bias_corr/$stabfile  
VARBC.cycle
```

scr/Minim needed input: Namelists

```
sed -e "s/NBPROC/${NPROC}/g" \  
-e "s/NPROCX/${NPROCX}/g" \  
-e "s/NPROCY/${NPROCY}/g" \  
-e "s/NREDNMC/${REDNMC}/" \  
-e "s/NBZONVAR_EW=NBZONVAR_EW/NBZONVAR_EW=${NBZONVAR_EW}/g" \  
-e "s/LVARBC=LVARBC/LVARBC=${LVARBC}/" \  
{HM_LIB}/nam/namelist_minimization > fort.4
```

Change in harmonie_namelists.pm %minimization -> mods in namelist_minimization

SMHI

0=to be used, -1=not to be used

Active or passive?

- Active: Use in the minimisation, i.e. contributing to the analysis
- Passive: Treated as active observations but not used in the minimisation
 - BATOR, screening, monitoring...

This is chosen in the namelist NAMJO:

```
&NAMJO
```

```
...
```

```
NOTVAR(1,1)= 0, 0,-1,-1,-1,-1,-1, 0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
```

```
NOTVAR(1,2)= 0, 0,-1,-1,-1,-1, 0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,
```

```
...
```

```
/END
```

where NOTVAR(1,1) = SYNOP, NOTVAR(1,2) = AIREP

and from "yomcosjo.F90" one can see NVAR_U = 1; NVAR_U10= 2; NVAR_H = 5; NVAR_T = 7; NVAR_Z=8;

Radar reflectivity = 26

Doppler wind = 44

NOTVAR(1 1)= 0 0 -1 -1 -1 -1 -1 0 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1 -1

NOTVAR(1,2)= 0, 0,-1,-1,-1,-1, 0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,

NOTVAR(1,3)= 0, 0,-1,

NOTVAR(1,4)= 0, 0,-1,-1,-1,-1,-1, 0,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,

NOTVAR(1,5)= 0, 0,-1,-1,-1,-1, 0, 0,-1,-1,-1,-1,-1,-1,-1,-1,-1, 0,-1,-1,-1,-1,-1,-1

NOTVAR(1,6)= 0, 0,-1,

[illegible]

NOTVAR(1.8)=-1,-1

[illegible][illegible][illegible]
$$1, -1, -1, -1, -1, -1, -1, -1, -1, 0,$$

/END

/END

Minimization output

IN LOGFILE:

Open HM_Date_YYYYMMDDHH.html in editor:

Look for 'Diagnostic JO-table (JOT) MINIMISATION'

try

grep -i 'grepcost' HM_Date_YYYYMMDDHH.html

+ MXMIN1999+0000, updated VARBC.cycle, updated CCMA (an_depar)

$fg_depar = obs - \text{guess which is bias corrected}$

By the way

$biascorr = \text{bias to be corrected}$

$fg_depar + biascorr = obs - \text{guess before bias correction}$