

Observations monitoring and related diagnostics

M. Dahoui, S. Abdallah, M. Bonavita, N. Bormann and T. McNally

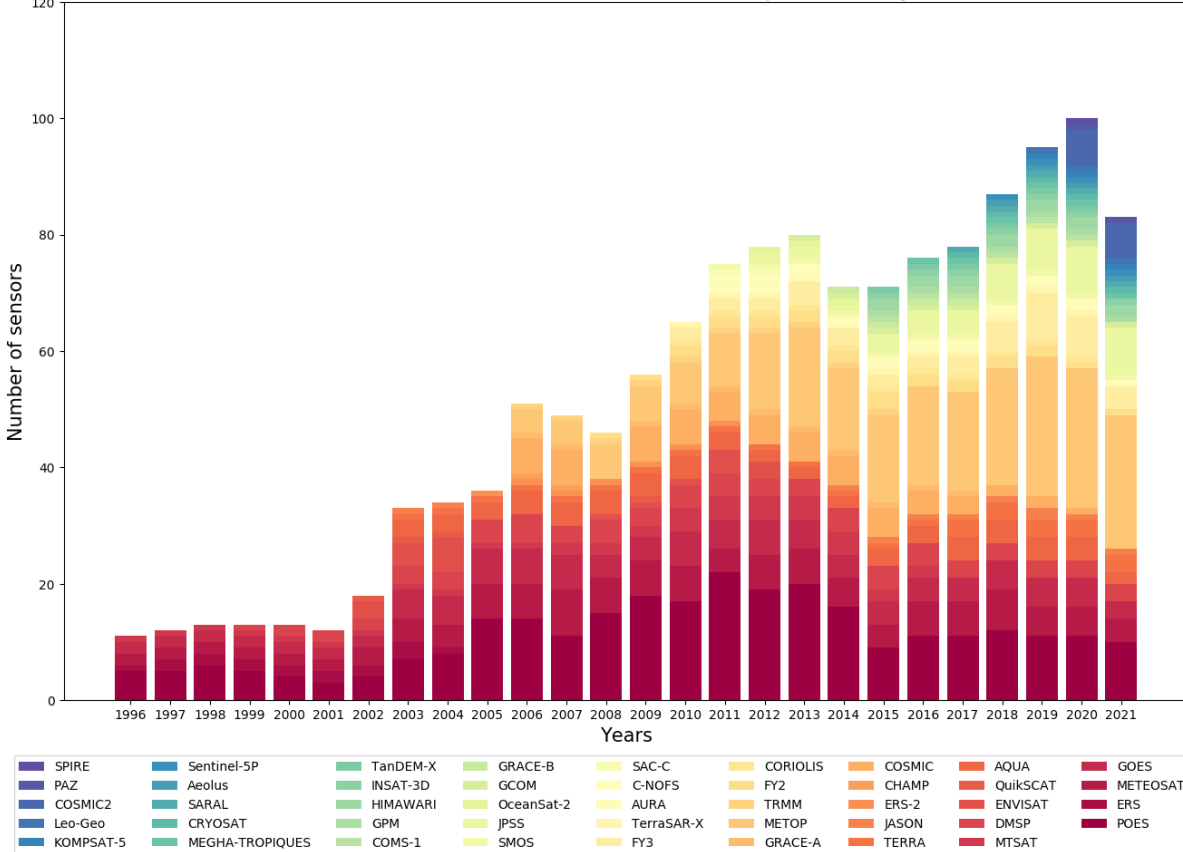
mohamed.dahoui@ecmwf.int

Outline

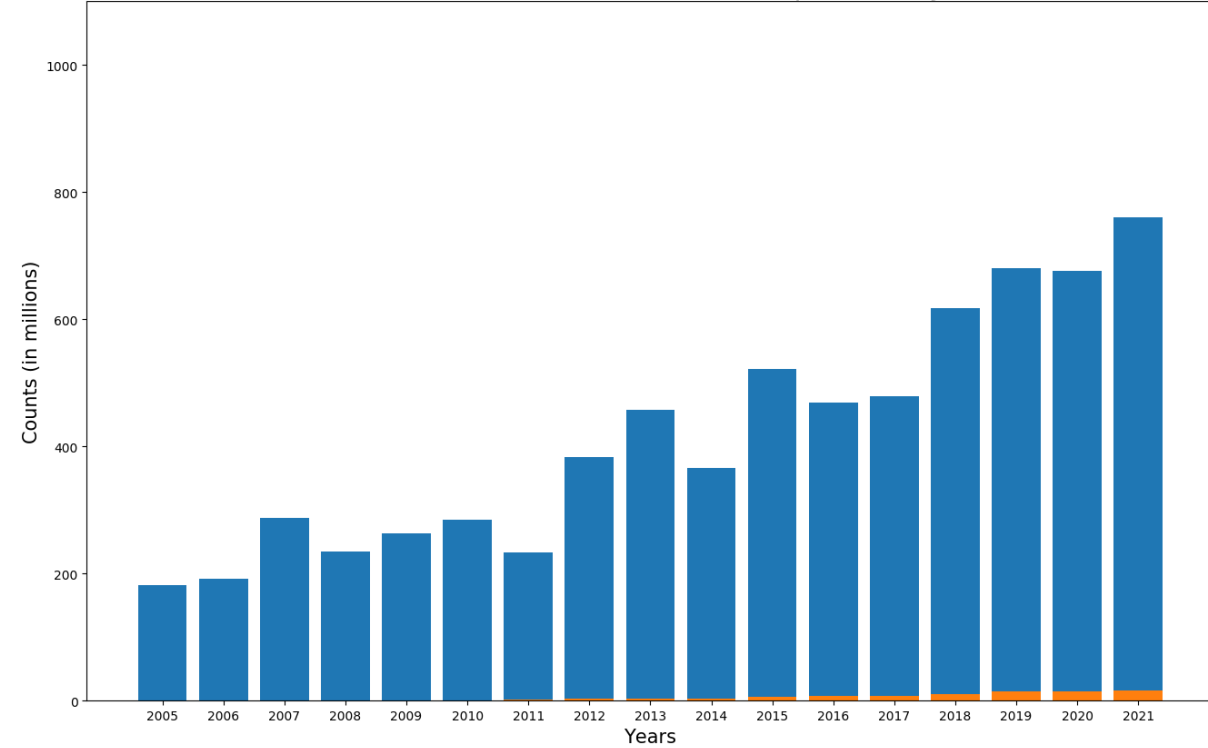
1. **Monitoring of observations: Why and how ?**
2. **Observation monitoring capabilities**
3. **Observation based diagnostics**
4. **Conclusion**

Evolution of data counts and diversity

Number of satellite instruments monitored operationally at ECMWF

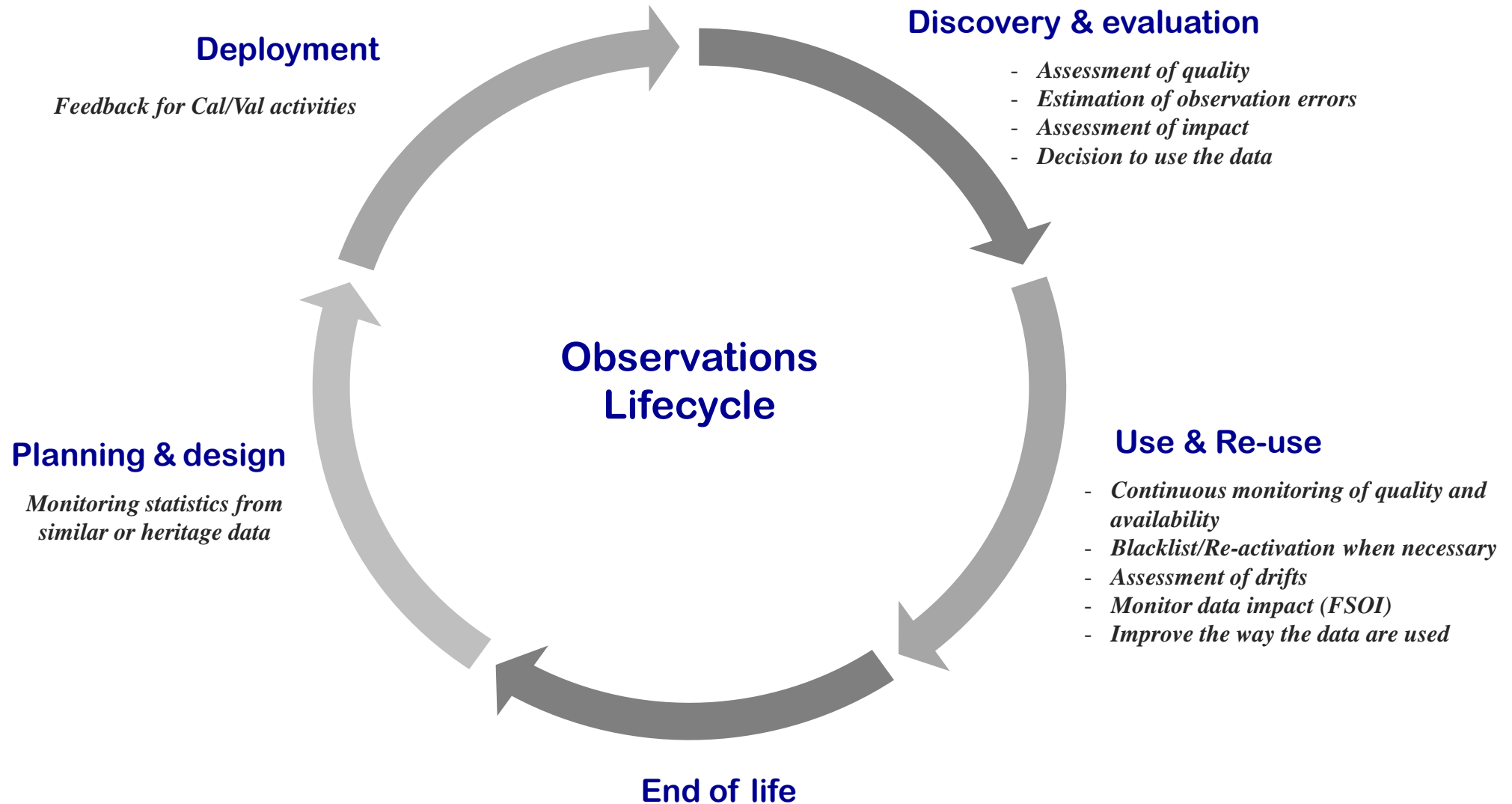


Number of satellite observations monitored operationally at ECMWF



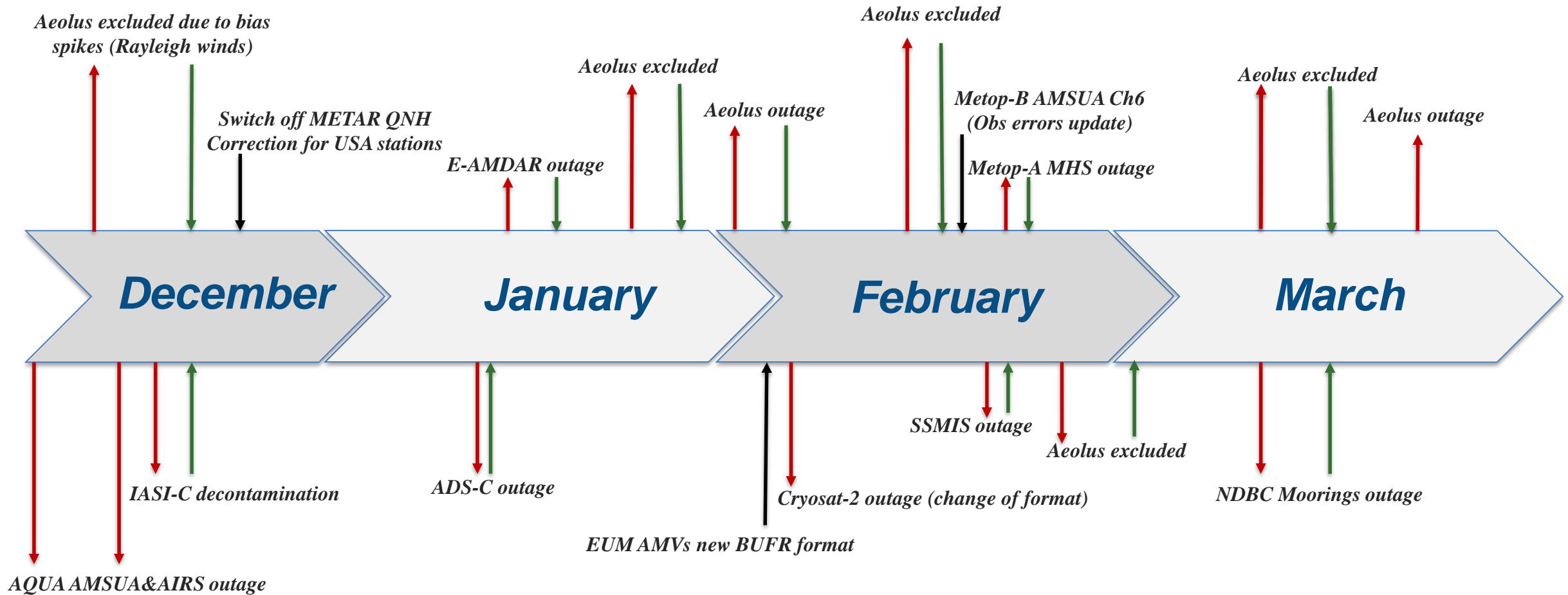
- ~ 800 million pieces of observations received daily and ~60 millions active . A lot of potential
- Data characteristics are subject to variation with possible consequences on the data impact.

Monitoring of observations: Why ?



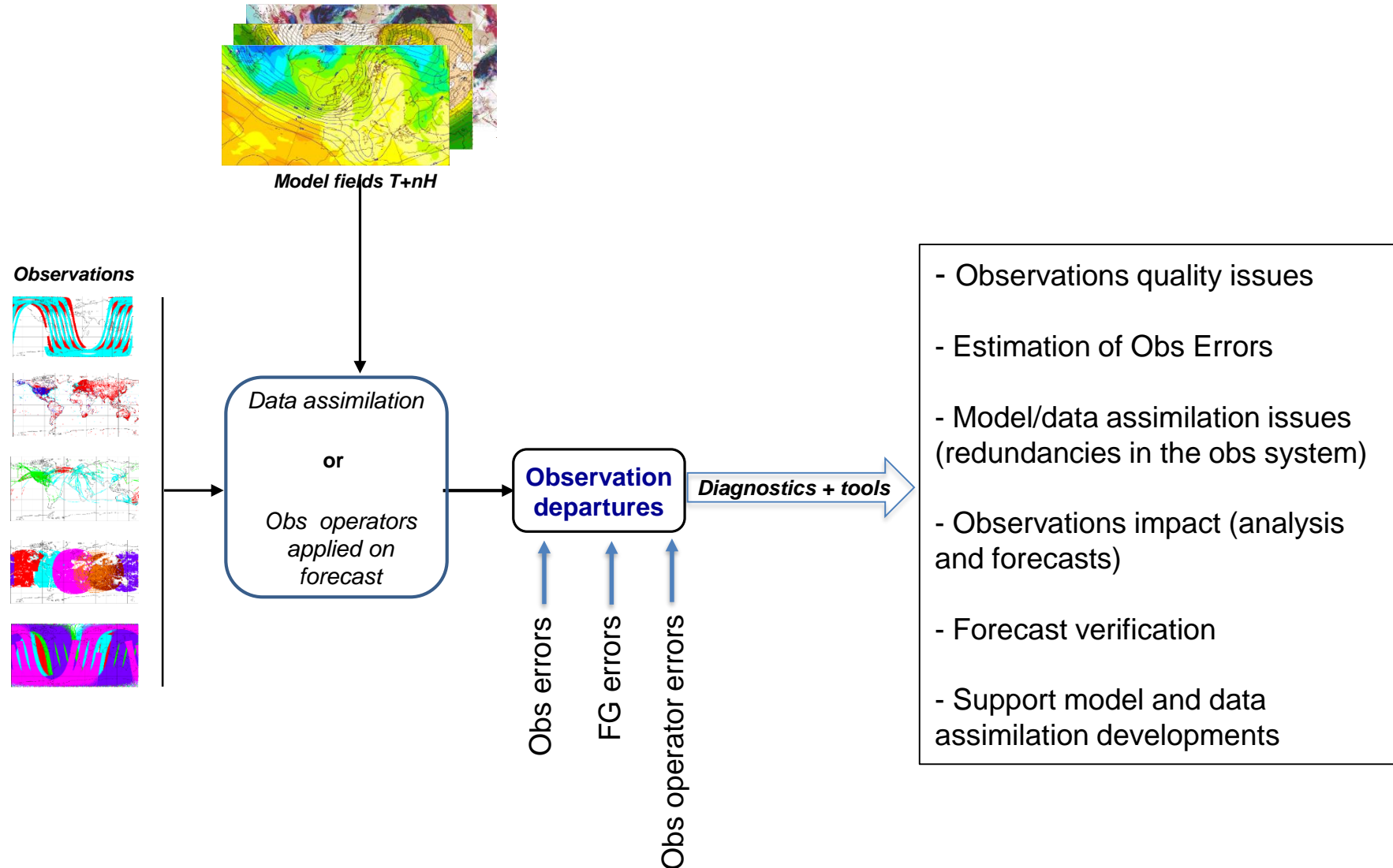
Monitoring of observations: Why ?

Data events (DJFM 2021)



Monthly update of data selections (4D-VAR in-situ and LDAS)

Monitoring of observations: how ?



Observations monitoring: how ?

- Routine production and display of statistics over large data samples
- Statistics are generally computed for observation quantities related to the data assimilation: departures, bias correction, data counts, etc.
- Statistics are produced for various data selection criteria (crucial for data usage monitoring and quality assessment)
- Availability of statistical tools to generate products allowing the investigation of data from various perspectives: time, area, vertical, FOV, etc.
- Availability of tools allowing generic comparison of statistics from different data assimilation setups
- Tools Should be generic and future proof (easy to maintain) to support new instruments

Outline

1. Monitoring of observations: Why and how ?
- 2. Observation monitoring capabilities**
3. Observation based diagnostics
4. Conclusion

Observations monitoring resources

- Operational monitoring suite: repository of generic monitoring statistics for almost all observations (internal and external use)
- Automatic detection of quality/availability issues
- Research monitoring suite: systematic for all experiments and allows inter-comparison (data impact on short ranges)
- Standalone monitoring tools: allow users to explore and investigate other aspects of data usage
- Other sources of information (data providers notifications, external monitoring resources, cross-activities monitoring, etc)

Operational monitoring

<https://www.ecmwf.int/en/forecasts/quality-our-forecasts/monitoring-observing-system>

ECMWF Search

Home About Forecasts Computing Research Learning

Charts | Datasets | **Quality of our forecasts** | About our forecasts | Access to

View published New draft Revisions Access control

Monitoring of the observing system

The purpose of the monitoring is to provide a detailed statistical information on the quality and availability of the different components of the observing system used/monitored by ECMWF. The monitoring results are primarily produced to help improve the usage of observations within the ECMWF data assimilation systems. Most of the products are updated on a daily basis.

In this page

- [Availability](#)
- [Satellite data monitoring](#)
- [Conventional data monitoring](#)
- [Ocean observation monitoring](#)
- [Data automatic checking](#)
- [Monitoring of GUAN stations](#)
- [ECMWF Global Data Monitoring Report Archive](#)

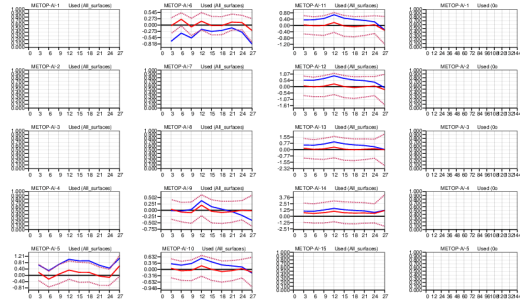
Monitoring

- Almost all earth system observations supported
- ~ 140000 plots updated daily or weekly
- Various types of plots for each data type
- Publicly accessible

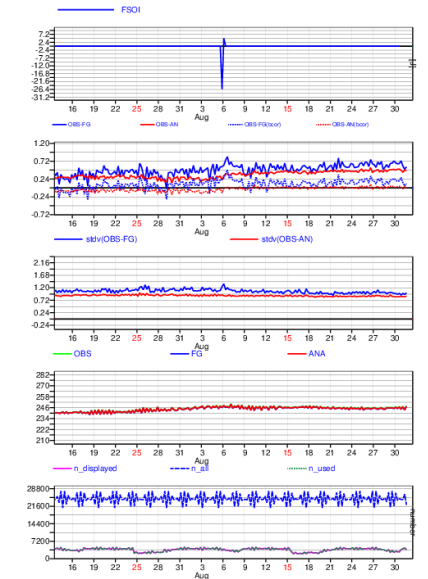
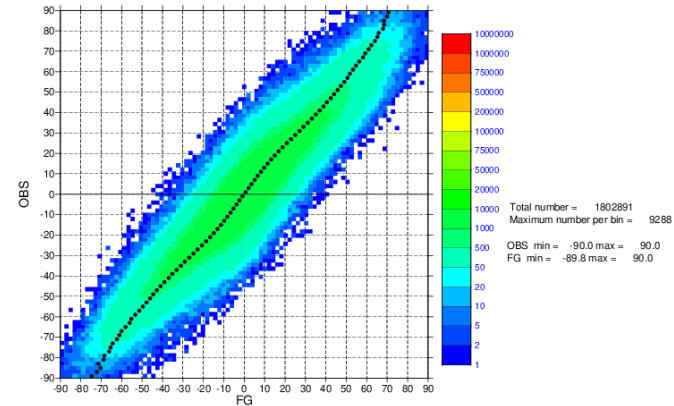
Selected monitoring products

STATISTICS FOR RADIANCES FROM METOP-A/AMSUA (GLOBAL)
 CHANNEL = 14, USED DATA [TIME STEP = 6 HOURS]
 Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0
 EXP = 0001 (LAST TIME WINDOW: -1)

Scan dependent statistics for RADIANCES from METOP-A & METOP-AV
 Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0
 EXP = 29 Jul - 28 Aug 2021
 Departures: blue = uncorrected, red = bias corrected +/- SD (dots) [scan pos in x-axis]

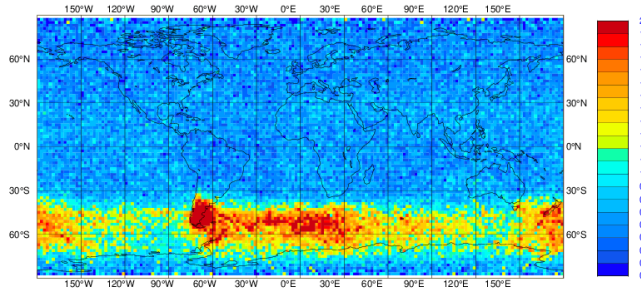


SCATTER PLOT OF FG VERSUS OBS
 AEOLUS HLOS WINDS RAYLEIGH CLEAR
 EXP = 0001 ; PERIOD = 2021080812 - 2021082712
 ALL - GLOBE

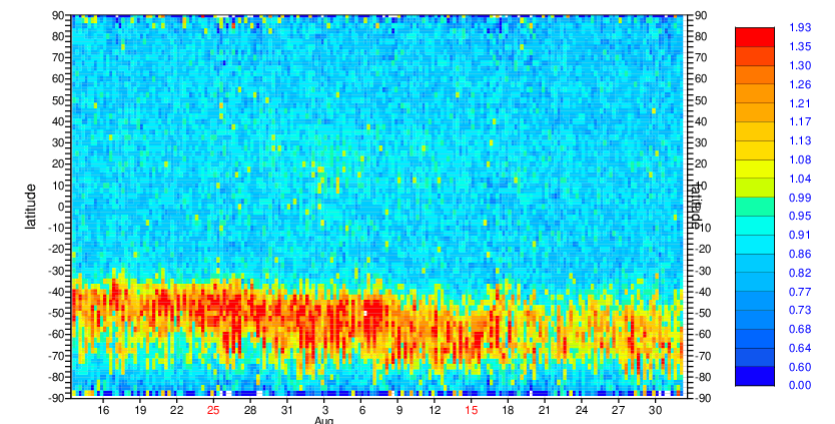


STATISTICS FOR RADIANCES FROM METOP-A
 STDV OF FIRST GUESS DEPARTURE (USED)
 DATA PERIOD = 2021-06-30 21 - 2021-08-27 21
 EXP = CHANNEL = 14

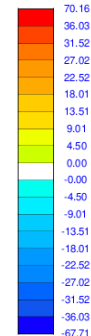
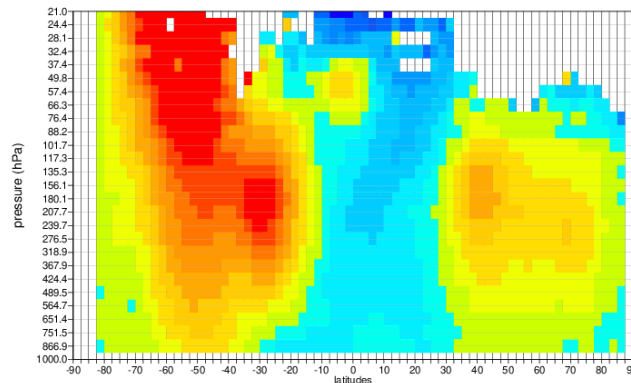
Min: 0.475 Max: 1.960 Mean: 0.930
 GRID: 2.00x 2.00



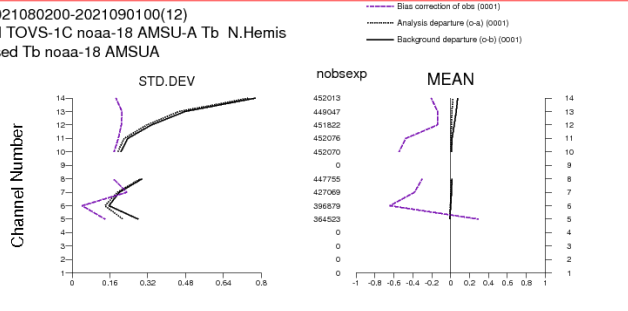
STATISTICS FOR RADIANCES FROM METOP-A/AMSUA
 CHANNEL = 14 [TIME STEP = 6 HOURS]
 STDEV FIRST GUESS DEPARTURE , USED
 EXP = 0001, DATA PERIOD = 2021071309 - 2021090115, AREA = 90S - 90N/ 00 - 360
 Min: 0.000 Max: 1.931 Mean: 0.916



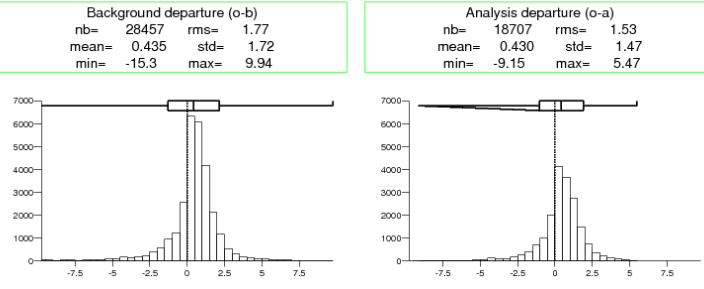
STATISTICS FOR HLOS FROM AEOLUS/RAYLEIGH CLEAR (ASCENDING NODE)
 LEVEL = 21.00 - 1000.00 HPA [TIME STEP = 3 HOURS]
 MEAN OBSERVATION VALUE , ALL (QC FILTERED)
 EXP = 0001, DATA PERIOD = 2021080809 - 2021082809, AREA = 90S - 90N/ 00 - 360
 Min: -67.710 Max: 70.160 Mean: 8.371



2021080200-2021090100(12)
 All TOVS-1C noaa-18 AMSU-A Tb N.Hemis
 used Tb noaa-18 AMSUA



0001 : 2021080300-2021090200(24)
 ocean prof potential temperature (XBTS) N.Hemis
 all POTM



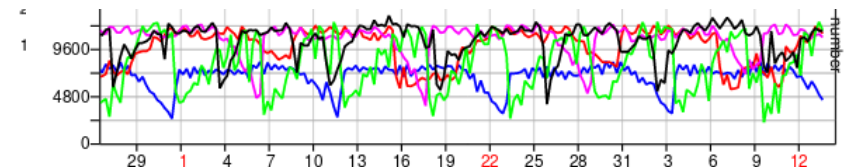
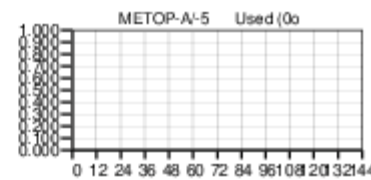
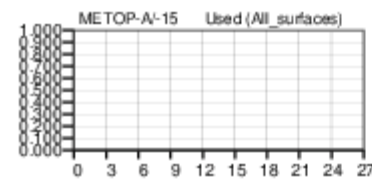
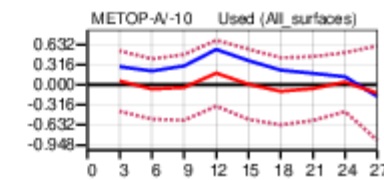
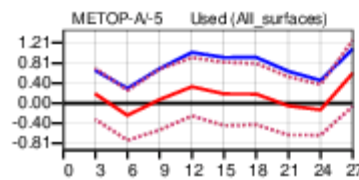
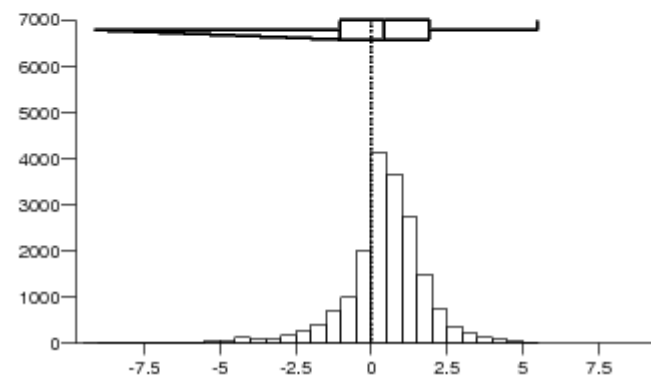
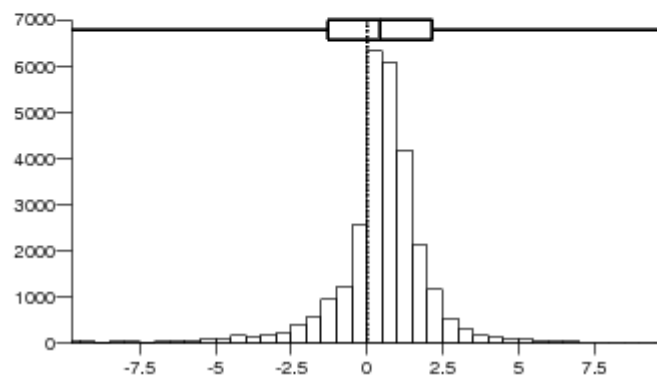
Selected monitoring products

STATISTICS FOR RADIANCES FROM AMSUA (GLOBAL)
 CHANNEL =6, USED 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)
 Scan dependent statistics for RADIANCES from METOP-A/ & METOP-A/
 Area: lon_w= 0.0, lon_e= 360.0, lat_n= -90.0, lat_s= 90.0
 EXP =, 29 Jul - 28 Aug 2021
 Departures: blue = uncorrected, red = bias corrected +/- SD (dots) [scan pos in x-axis]

0001 : 2021080300-2021090200(24)
 ocean prof potential temperature (XBTS) N.Hemis
 all POTM

Background departure (o-b)
 nb= 28457 rms= 1.77
 mean= 0.435 std= 1.72
 min= -15.3 max= 9.94

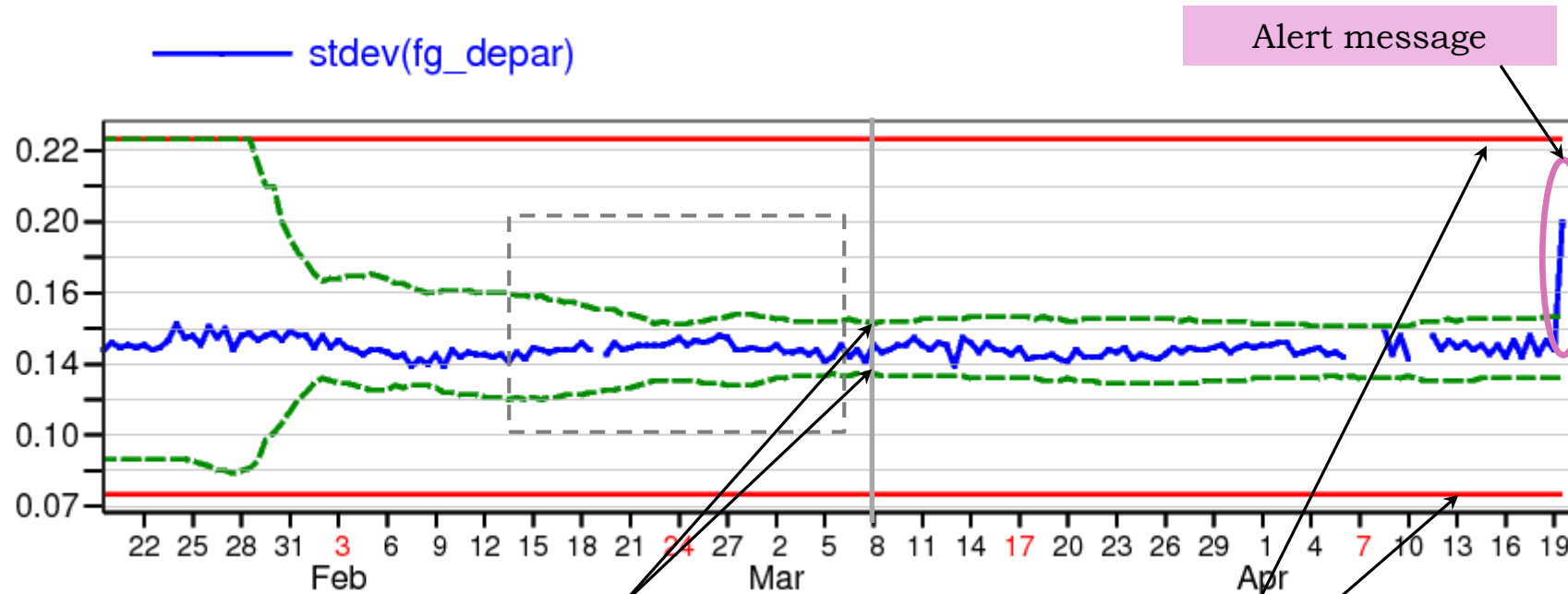
Analysis departure (o-a)
 nb= 18707 rms= 1.53
 mean= 0.430 std= 1.47
 min= -9.15 max= 5.47



Automatic data checking system

- The large amount of active observations makes it difficult to timely detect availability/quality issues
- An automatic data checking system is implemented at ECMWF to continuously monitor satellite and in-situ data (main trigger for corrective actions). Warnings available to internal and selected external users, at user specified levels of detail
- The same system is used to detect improved in-situ data that are currently excluded (timely activation of improved observations)
- Detects data problems but informs also when extreme events are taking place or in case of issues with the model/DA

Automatic data checking system



Soft limits ($5 \pm \text{stdev}$ of statistics to be checked, calculated from past statistics over a recent period ending 2 days earlier and excluding extremes)

Hard limits (fixed)

**Observations with model
feedback info (ODB)**

Current statistics
Selected Obs quantities

Past statistics
Selected Obs quantities

Hard limits
(only for satellites)
Detects slow drifts

Soft limits
computed
dynamically
Detects sudden changes

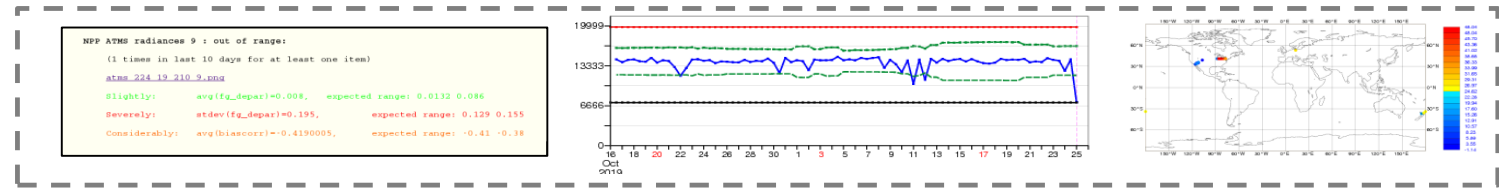
anomaly detection
Thresholds based tests
Static tests
Filters
Flexibility to add other tests

Record of new/missing datasets

Ignore facility

Past warnings

Periodic reporting



Email

Web

Event Data base

Blacklist procedure

Automatic data checking system

Three-component system

All observations

Detects widespread **availability** and **quality** problems over selected **areas**

Based on FG and Analysis departures, bias correction, data counts, etc

In-situ observations

Detects severe **quality** problems affecting **individual stations**

Detect stations persistently missing

Detect new stations

Based on FG departures and the IFS produced PGE

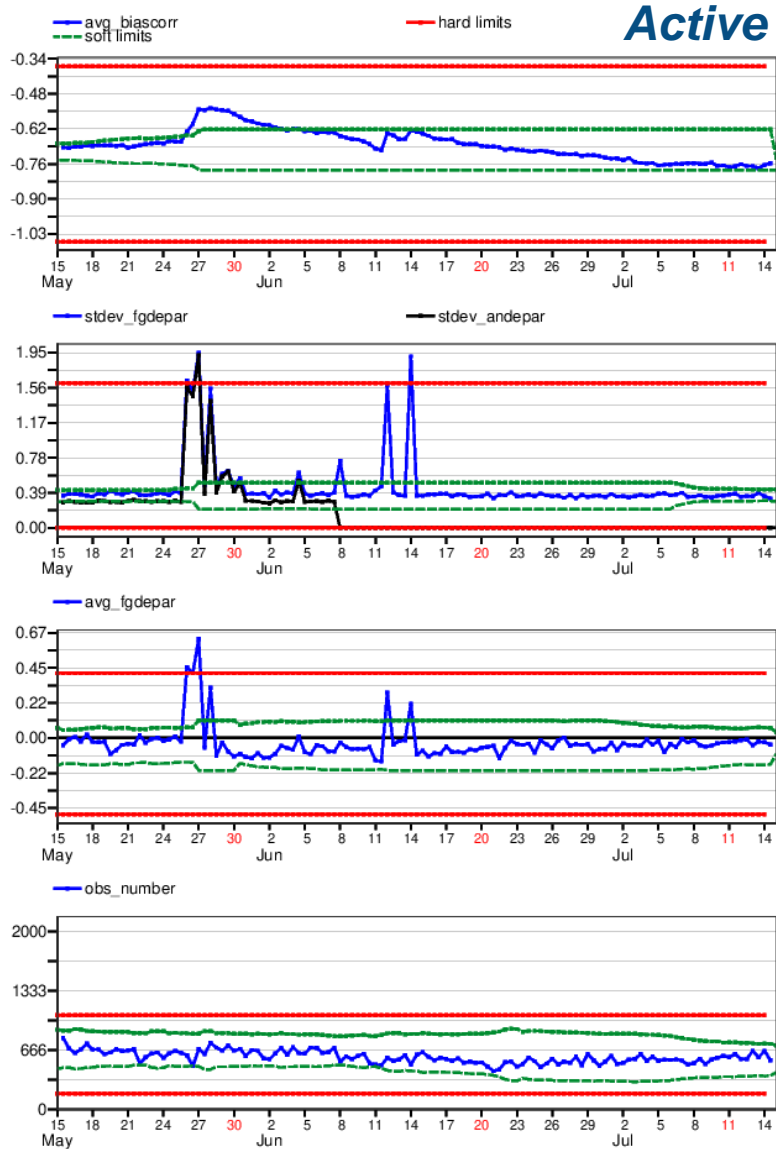
In-situ observations

Detects persistent **quality** improvements of “blacklisted” stations

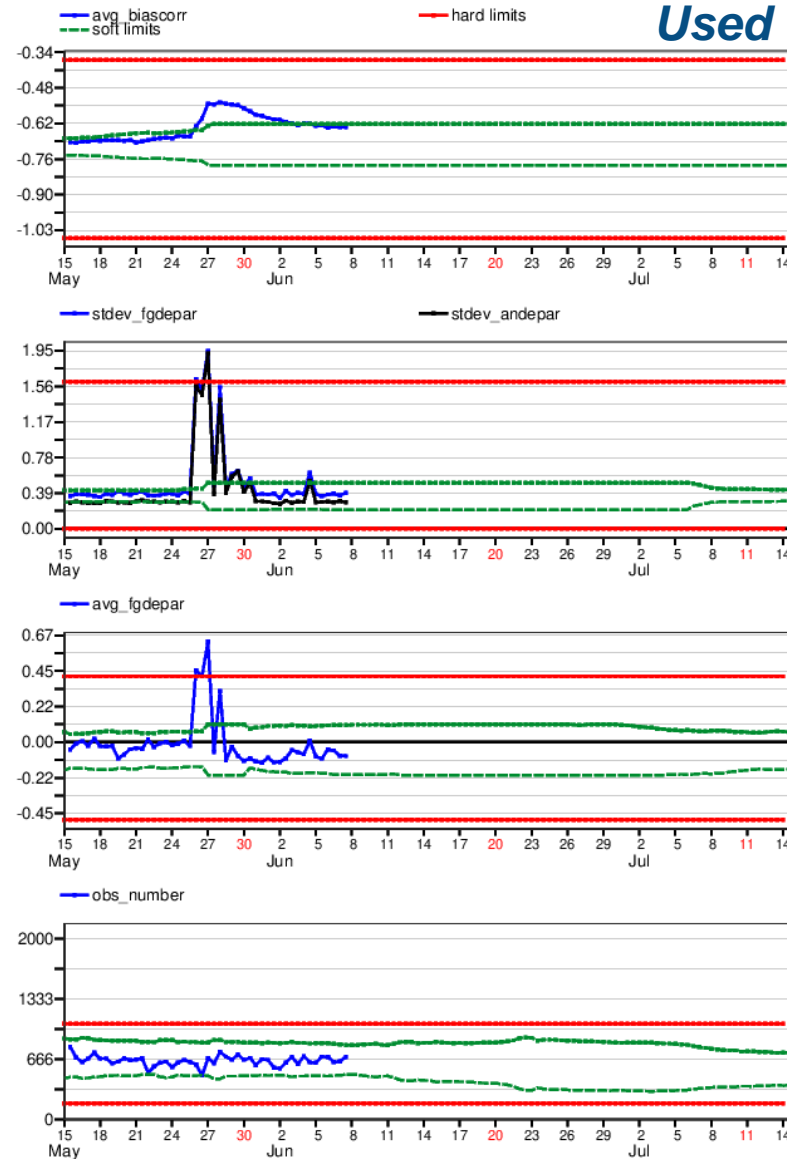
Based on an estimated PGE

Automatic data checking system (examples)

AQUA AIRS Radiances Global channel: 1092
All data, EXP =
12001_119_112_1092/12001_119_112_1092 (used)



AQUA AIRS Radiances Global channel: 1092
All data, EXP =
12001_119_112_1092/12001_119_112_1092 (used)

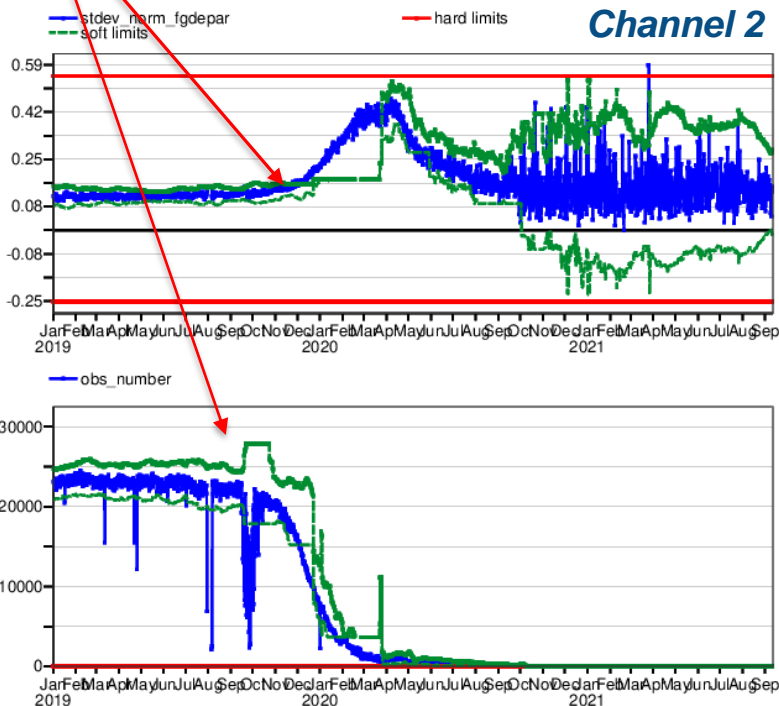


- AIRS channel 1092 erratic behaviour
- Data excluded after few bad cycles

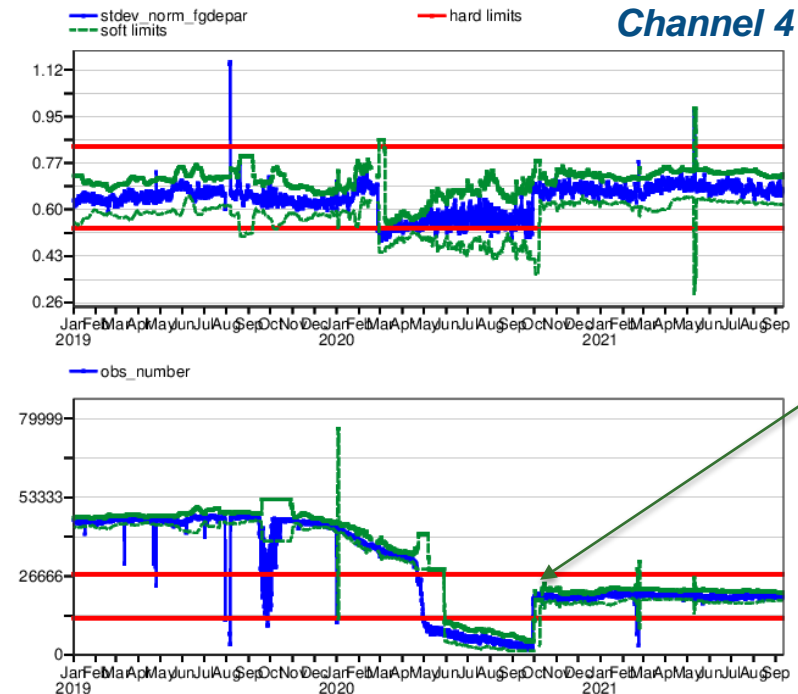
Automatic data checking system (examples)

- METOP-A MHS channel 2 quality degradation resulting in the drop of counts below the processing thresholds and hence the reduction of data. Fixed by data provider
- METOP-A MHS channel 2 continued to behave badly. This had an impact on the use of other channels until we found a way to use the rest of data without checks based on channel 2

METOP-A MHS ALLSKY Radiances All surface type Global channel: 2
All data, EXP =
44002_119_3_112_2/44002_119_3_112_2



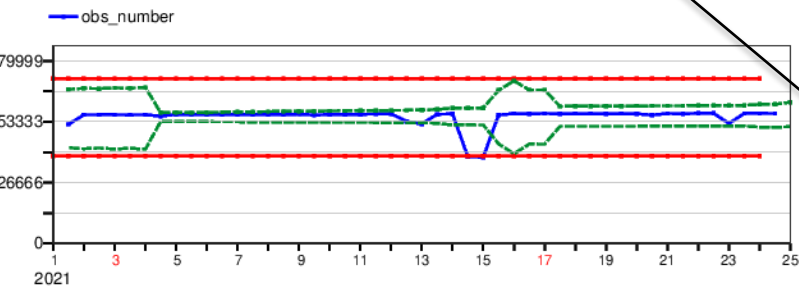
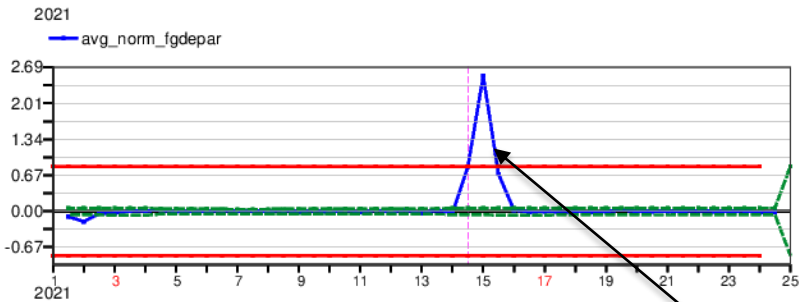
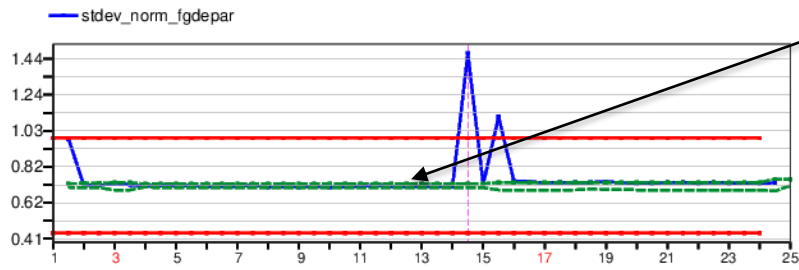
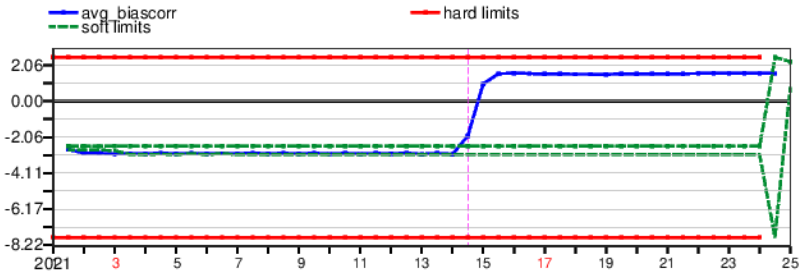
METOP-A MHS ALLSKY Radiances All surface type Global channel: 4
All data, EXP =
44002_119_3_112_4/44002_119_3_112_4



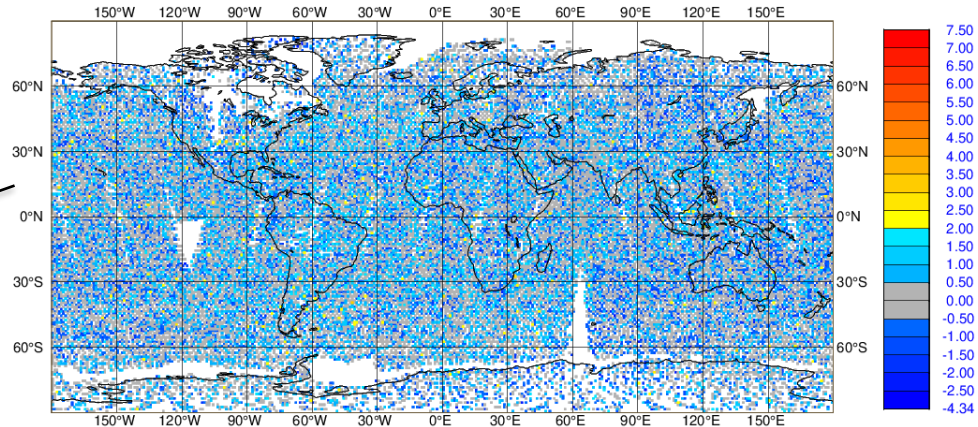
Use of data without
channel 2 based
checks

Automatic data checking system (examples)

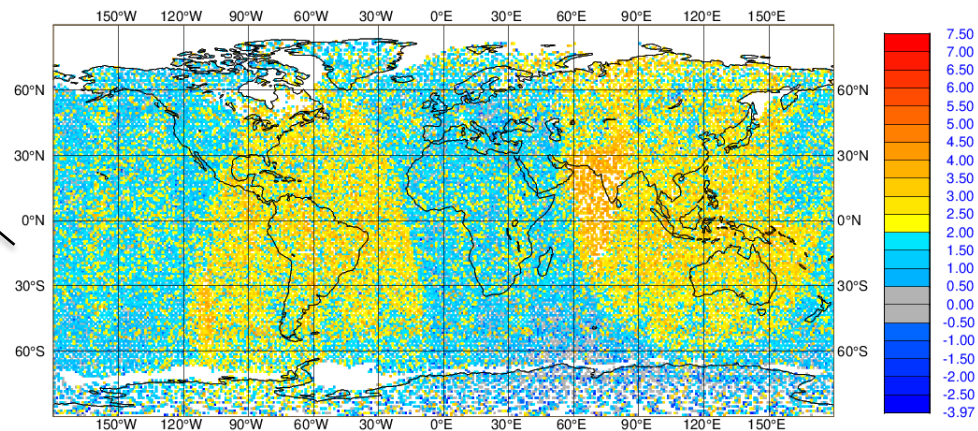
FY-3D MWHS2 ALLSKY Radiances All surface type Global channel: 3
 All data, EXP =
 57002_119_3_112_3 (used)



MWHS2 FROM (REPORTYPES: 57002)
 MEAN FIRST GUESS DEPARTURE (OBS-FG) [] (ACTIVE)
 DATA PERIOD: 2021011321 - 2021011409
 ACTIVE-CHANNEL:3
 Min: -3.838 Max: 3.280 Mean: -0.035
 GRID: 1.00x 1.00



MWHS2 FROM (REPORTYPES: 57002)
 MEAN FIRST GUESS DEPARTURE (OBS-FG) [] (ACTIVE)
 DATA PERIOD: 2021011409 - 2021011521
 ACTIVE-CHANNEL:3
 Min: -3.467 Max: 4.384 Mean: 1.516
 GRID: 1.00x 1.00

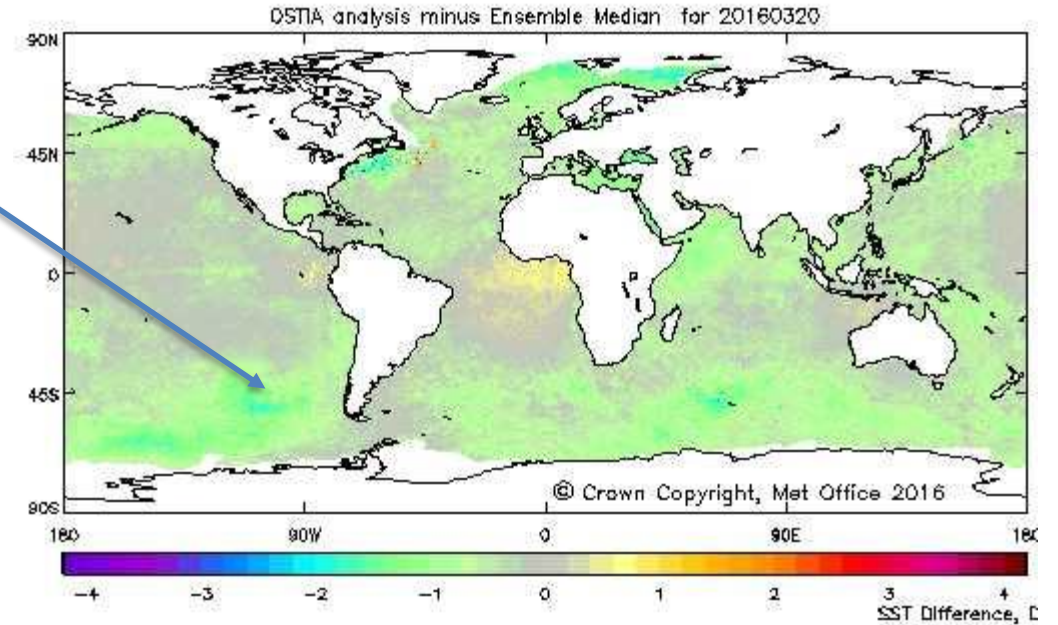
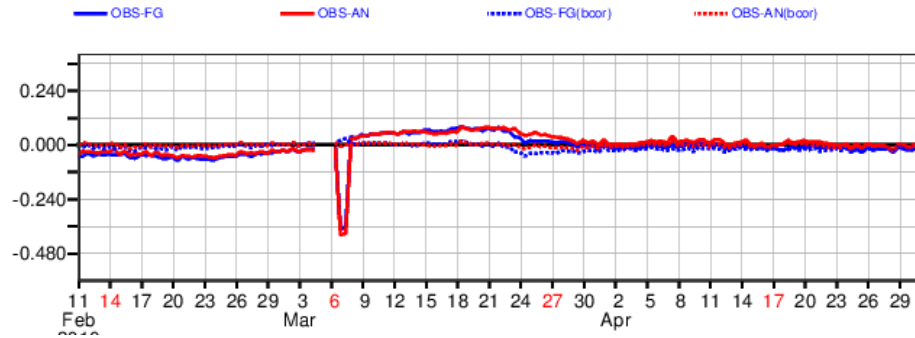


- FY-3D MWHS2 Ch3 sudden degradation
- Quick recovery (no action needed)

Earth system automatic warnings

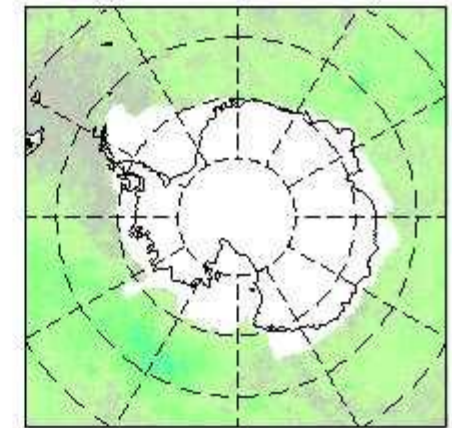
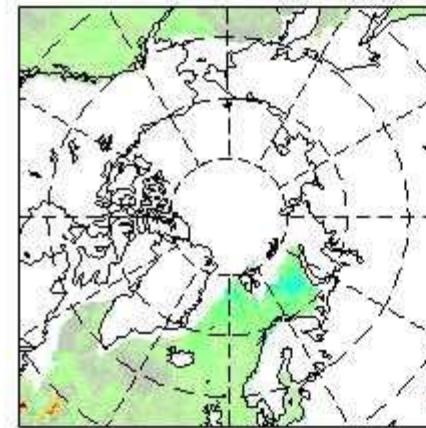
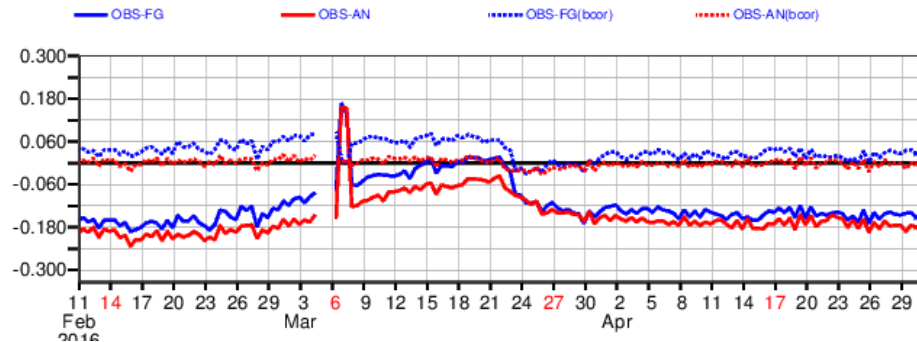
STATISTICS FOR CRIS FROM NPP/CRIS
CHANNEL = 457, ACTIVE DATA [TIME STEP = 12 HOURS]
Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0 (over All_surfaces)
EXP = 0001 (LAST TIME WINDOW: 2016043021)

CRIS Ch 457



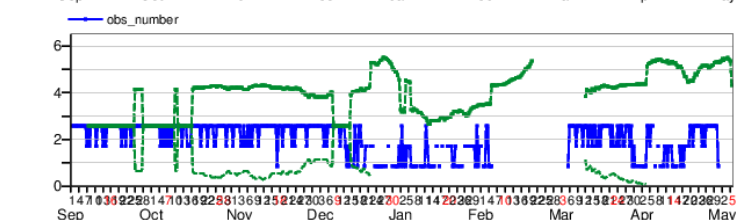
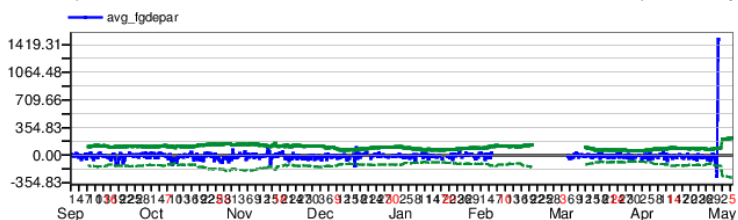
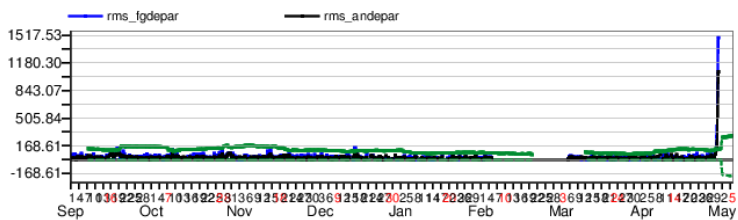
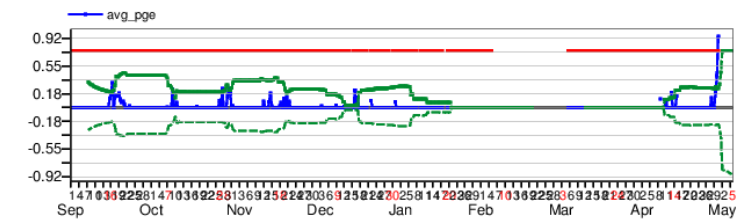
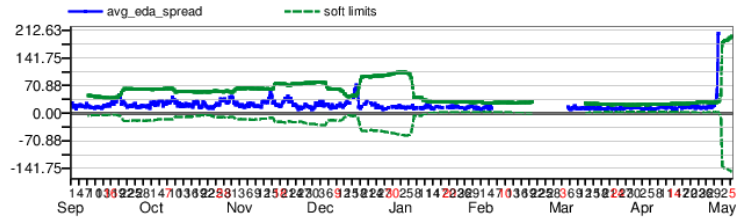
STATISTICS FOR RADIANCES FROM METOP-A/IASI
CHANNEL = 1090, ACTIVE DATA [TIME STEP = 12 HOURS]
Area: lon_w= 0.0, lon_e= 360.0, lat_s= -90.0, lat_n= 90.0 (over All_surfaces)
EXP = 0001 (LAST TIME WINDOW: 2016043021)

IASI Ch 1090

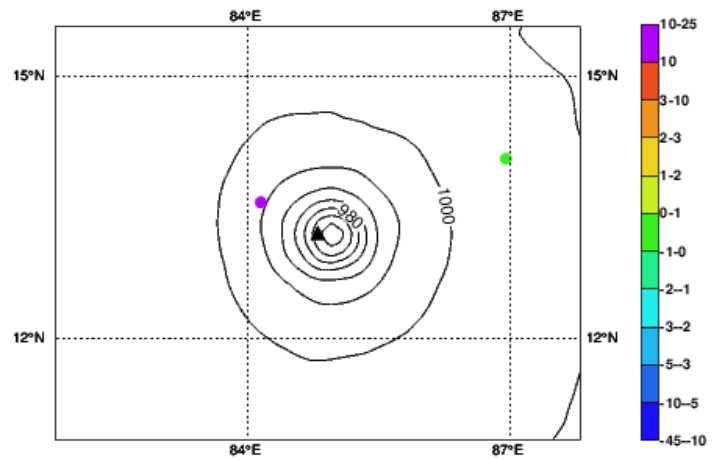


Automatic data checking system (examples)

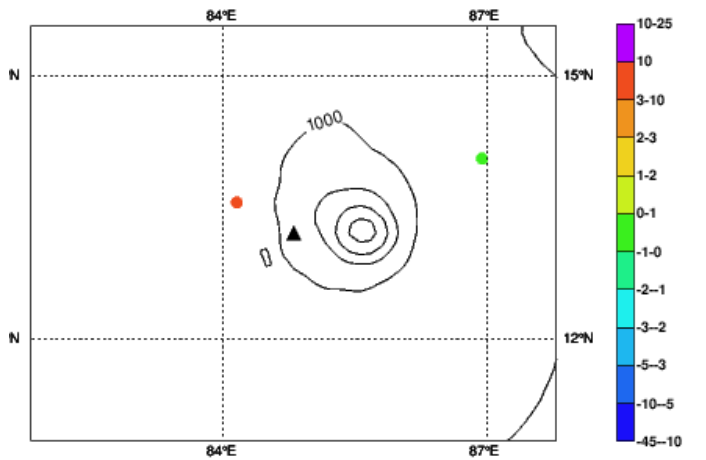
DRIBU Surface_pressure ID 23094
All data, EXP =0001
16005_110_0_23094 (used)



Surface pressure OBS-FG (Surface Surface) hPa [All 9H to 15H]
0001 06h MSLP from 20190430 06 LWDA [FANI(967.475625)]
[contour interval every 5 hPa/ observed position in black triangle (964)]
Mean: 4.33054 StDev: 7.45137 Data Count: 3



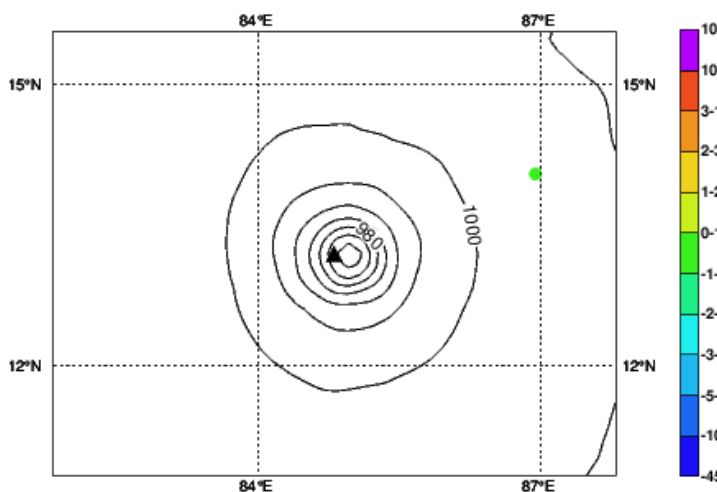
Surface pressure OBS-AN (Surface Surface) hPa [All 9H to 15H]
0001 AN MSLP for 20190430 12 [FANI(983.350625)]
[contour interval every 5 hPa/ observed position in black triangle (964)]
Mean: 230.3904266 StDev: 4.57726 Data Count: 3



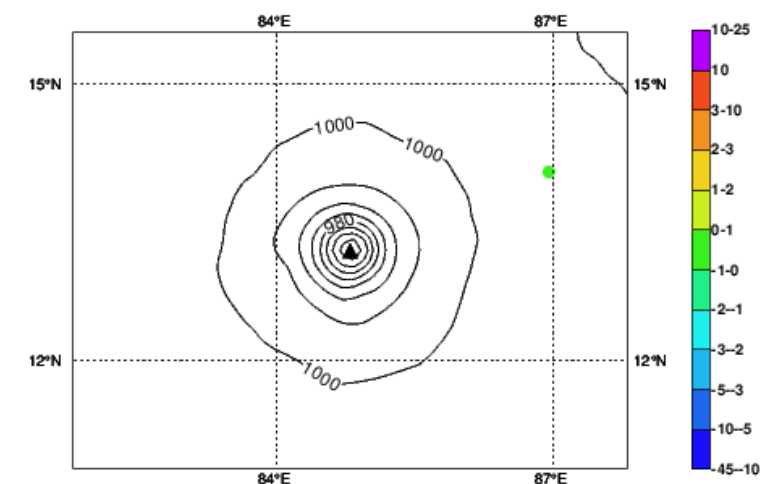
The forecast was also - impacted

Denial of the buoy 23094

Surface pressure OBS-FG (Surface Surface) hPa [All 9H to 15H]
h6py 06h MSLP from 20190430 06 LWDA [FANI(967.475625)]
[contour interval every 5 hPa/ observed position in black triangle (964)]
Mean: -0.93544 StDev: 0.304465 Data Count: 2

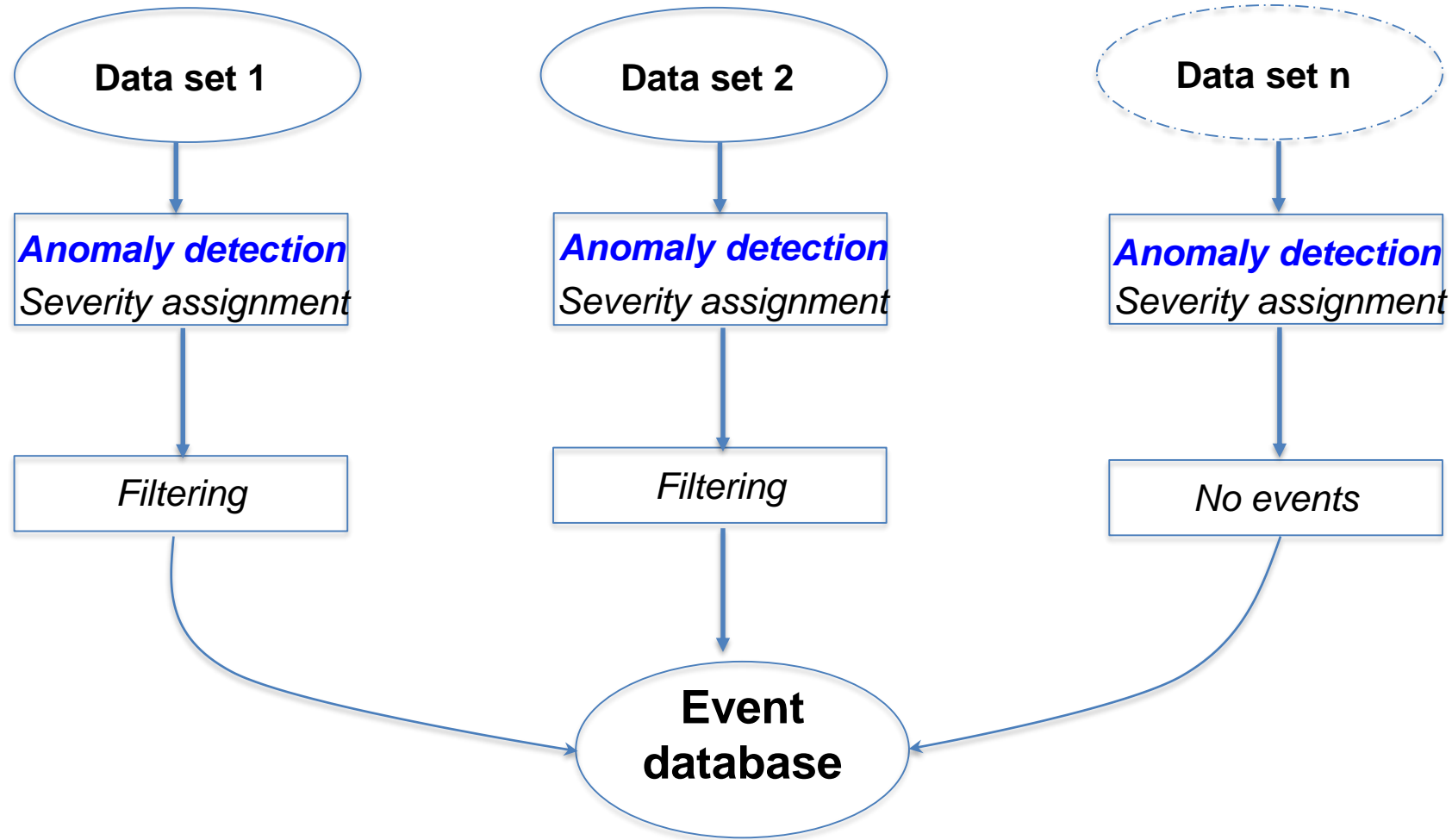


Surface pressure OBS-AN (Surface Surface) hPa [All 9H to 15H]
h6py AN MSLP for 20190430 12 [FANI(961.315625)]
[contour interval every 5 hPa/ observed position in black triangle (964)]
Mean: -79.390976 StDev: 0.205471 Data Count: 2

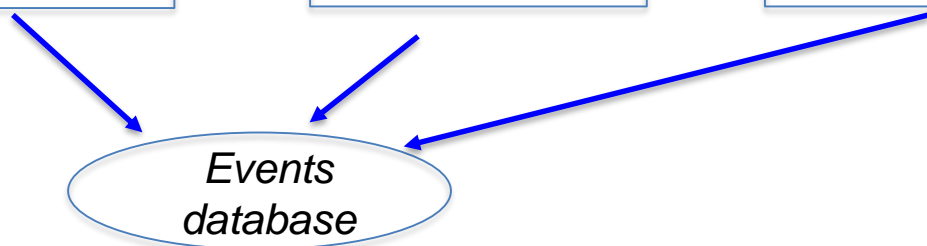
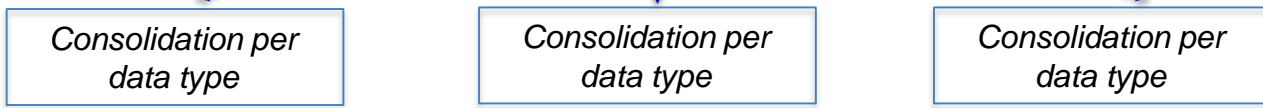
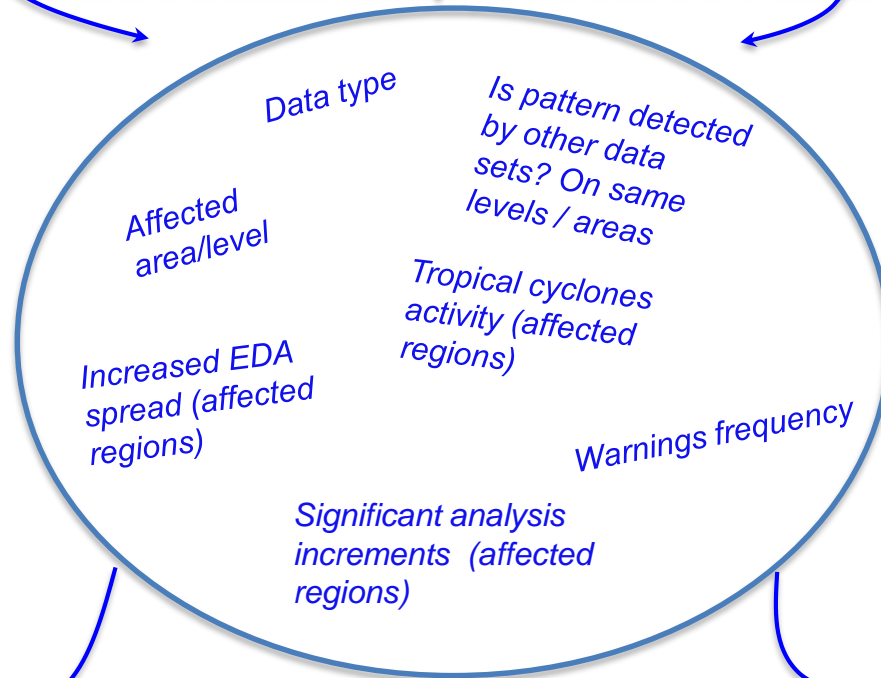
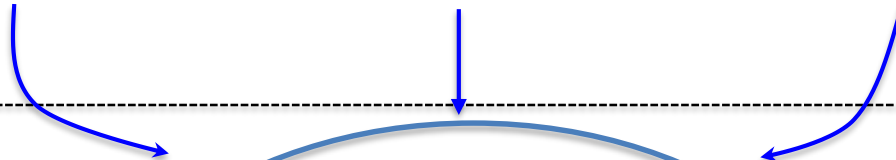
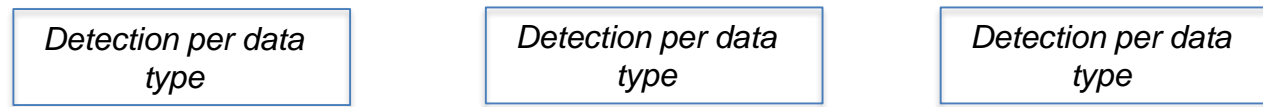


Earth system automatic warnings (ML approach)

Work in progress



- Current system works well but requires time and expert knowledge to distinguish data and model issues. Both important but require different actions
- Although false alarms are not common but many alarms do not require actions.



TensorFlow (LSTM + auto-encoders)

- Data de-seasonalised and normalised
- Models trained (once every quarter) on long data sets (> year)
- Models trained (every cycle) on short data sets (~60 days)
- Previous alarms excluded from training

ML classifier (random forests)

- Training based on labelled warnings populated from the operational alarm system)
- During the labelling process, decisions are made based on the data types, area of interest, etc.
- The classifier will adjust: Severity, likely cause of the problem (model, obs) and the need for action.

Consolidation

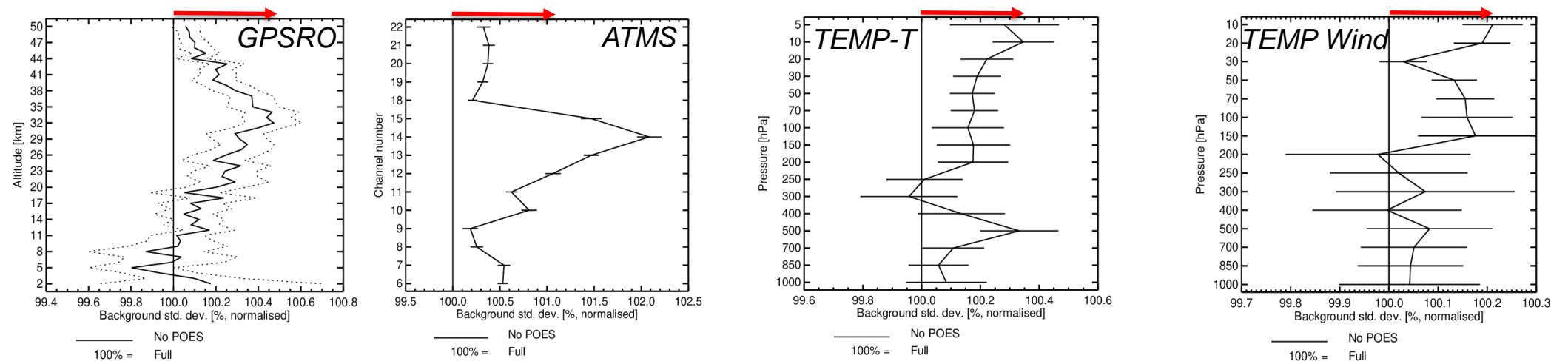
- Warnings are consolidated and processed for each data type



Research monitoring suite

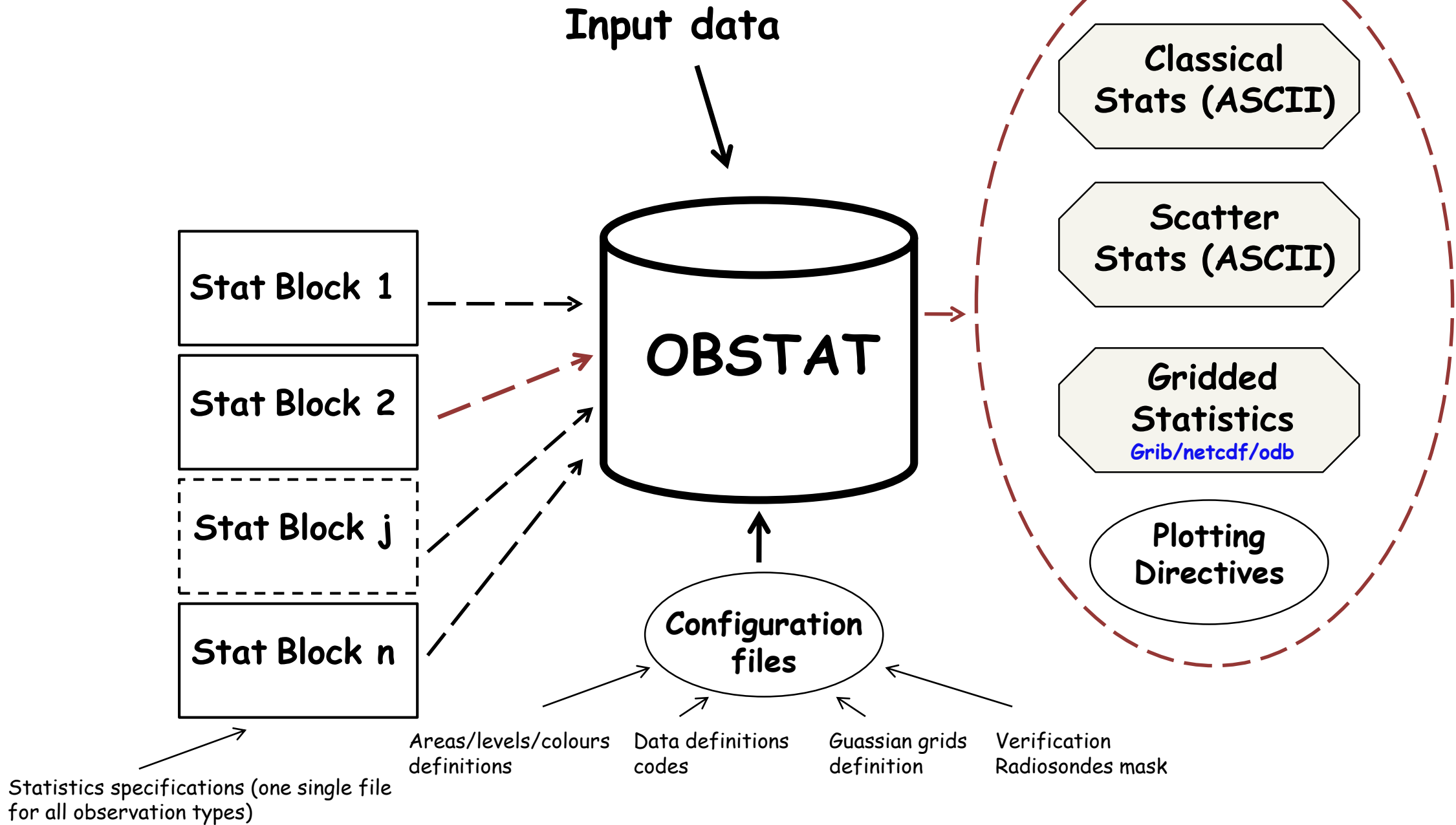
- Generic statistics are computed systematically for all research experiments (used observations only)
- Quick comparison of data usage between experiments: Data counts, First guess and analysis fit to used observations
- Provides idea on data impact (for data denial experiments) and the impact of scientific changes

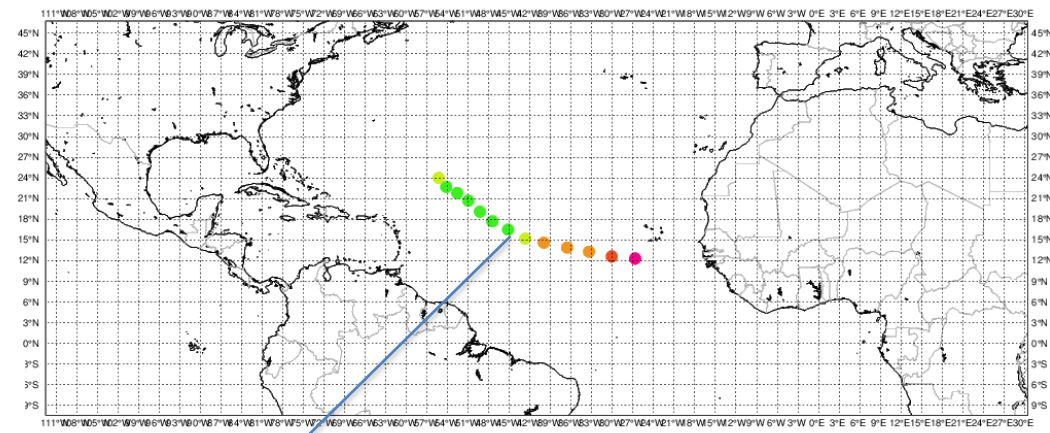
Degradation of short-range forecasts from losing the legacy POES satellites for temperature wind and humidity



Standalone monitoring tools

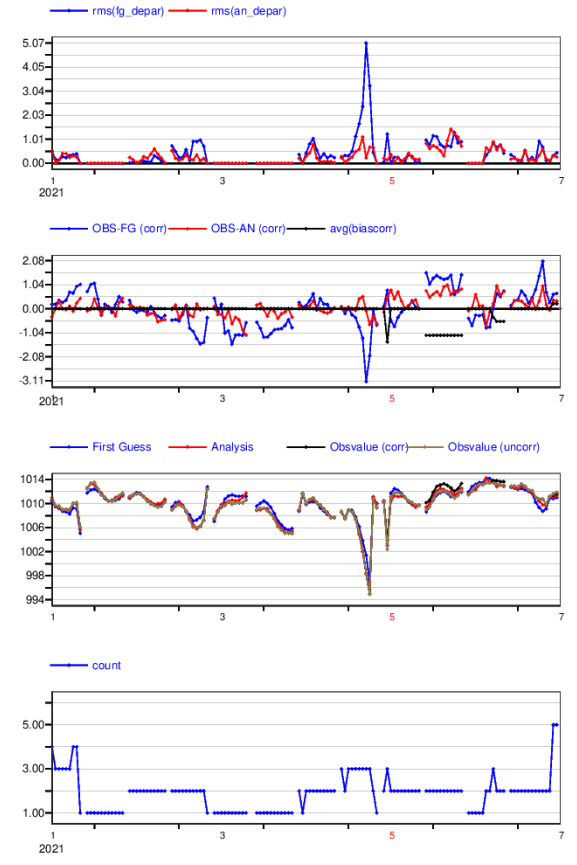
- Designed to allow users to conduct special investigations of data usage (Beyond the default set of statistics)
- Allows users to compute statistics for various **diagnostics** (departures, bias correction, FSOI, etc.)
- Production of statistics from various perspective (time, vertical, latitudes, asc/desc, day/night, FOV, Lagrangian, scatterplots, time series for individual stations, etc.)
- Sufficiently generic to handle all data types
- Used to overlay results from different model versions
- Advanced functions such as co-colocation, Observation matching, etc
- At ECMWF the OBSTAT software is used as a standalone monitoring tool in addition to being the core of the operational and research monitoring suites.





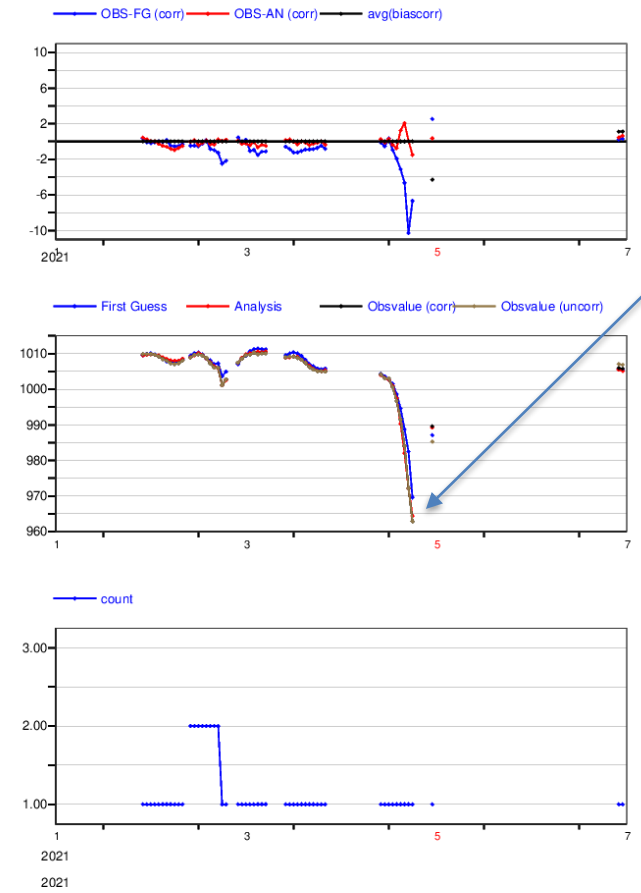
TC LARRY (4x4 degrees)

Statistics for Surface pressure (hPa)
Used data, EXP =0001 [each 1 hours]
Statistics within a rectangular box [4 deg] around LARRY



TC LARRY (2x2 degrees)

Statistics for Surface pressure (hPa)
Used data, EXP =0001 [each 1 hours]
Statistics within a rectangular box [2 deg] around LARRY



Outline

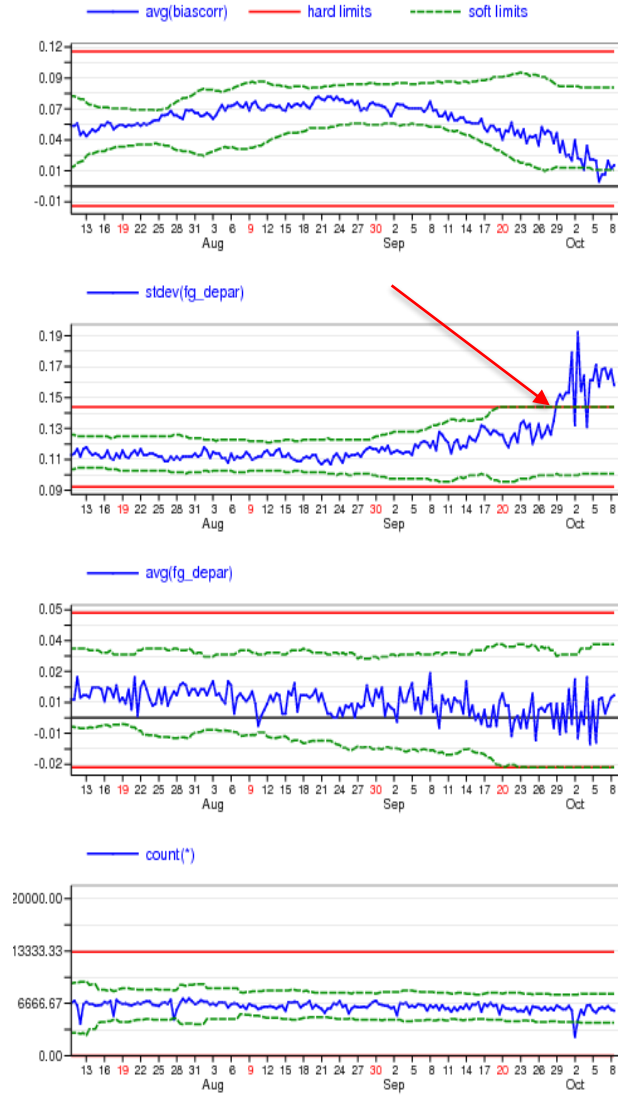
1. Monitoring of observations: Why and how ?
2. Observation monitoring capabilities
- 3. Observation based diagnostics**
4. Conclusion

Observations related diagnostics

- Inference from redundancies in the observing system (supported by collocation when possible)
- Impact on departures from changes to the observing system or how the data are assimilated
- Diagnostics in observation space based on departures
- Diagnostics from the EDA
- Triple-collocation

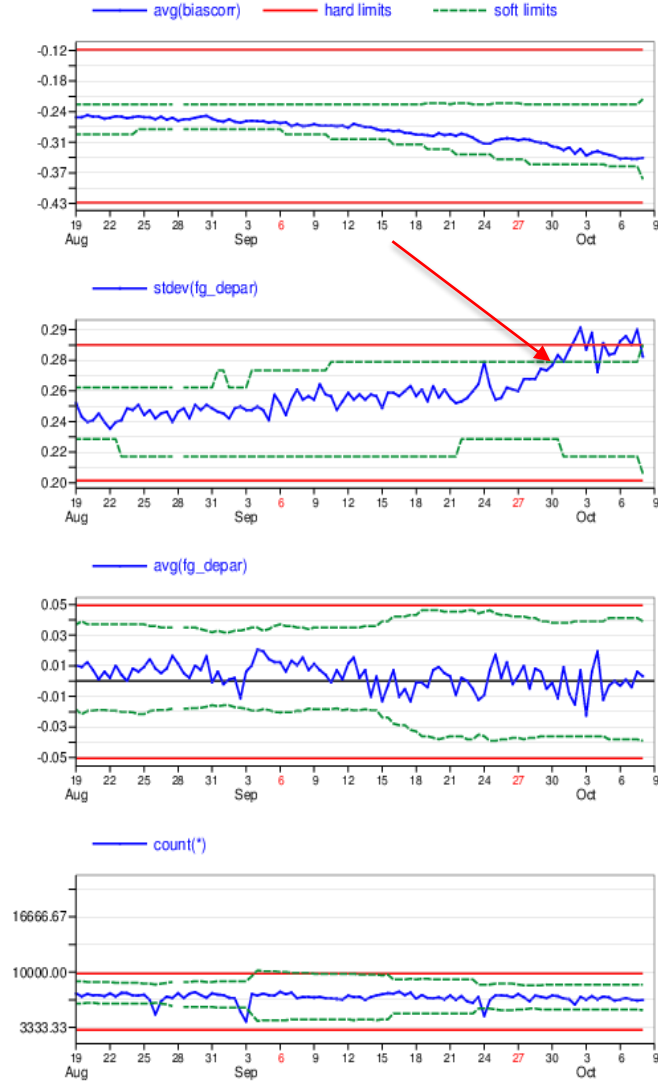
S-NPP CrIS Ch 101

NPP CrIS 101 radiances
Active data, EXP =0001
cris_224_27_101_210



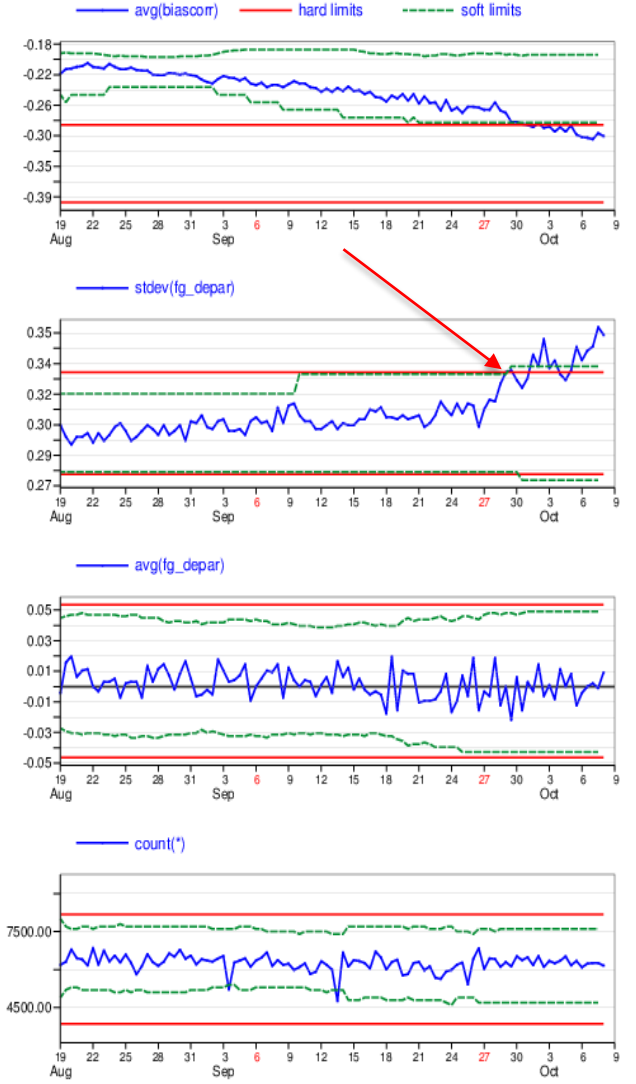
METOP-B IASI Ch 272

METOP-B IASI 272 radiances
Active data, EXP =0001
iasi_3_16_272_210



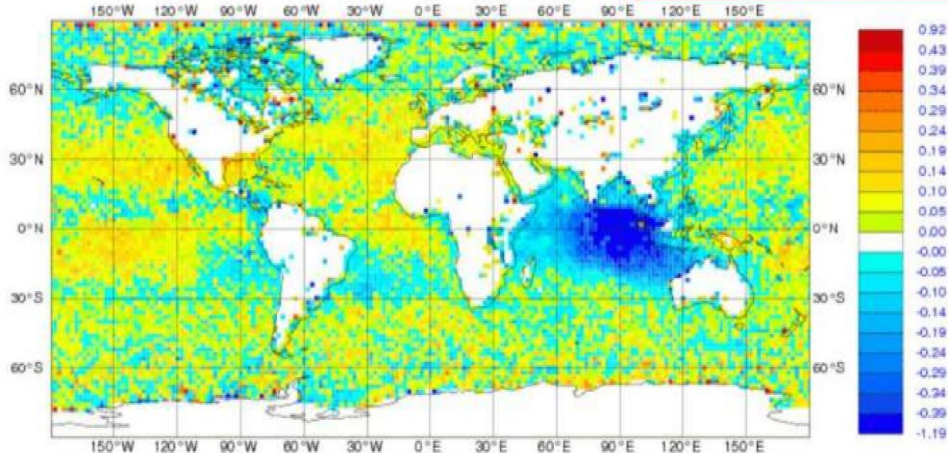
AQUA AIRS Ch 221

AQUA AIRS 221 radiances
Active data, EXP =0001
airs_784_11_221_210

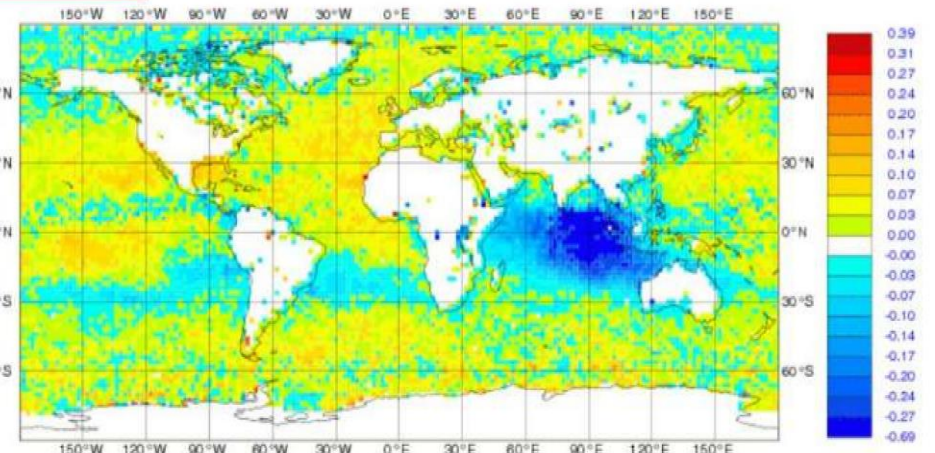


**Gridded monthly mean O-B
October 2015**

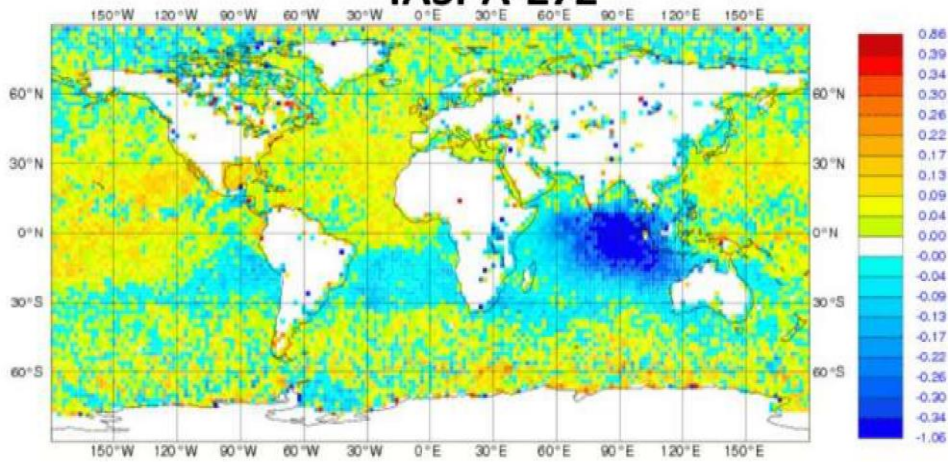
AIRS-221



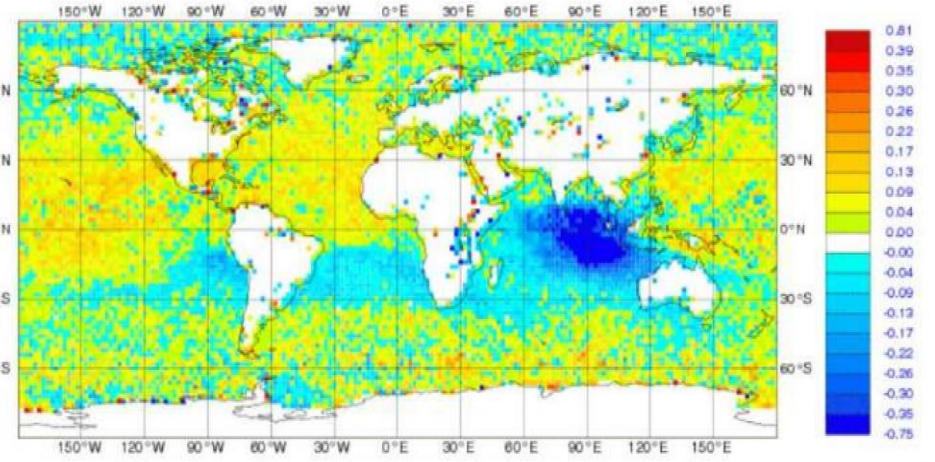
CRIS-101



IASI-A-272

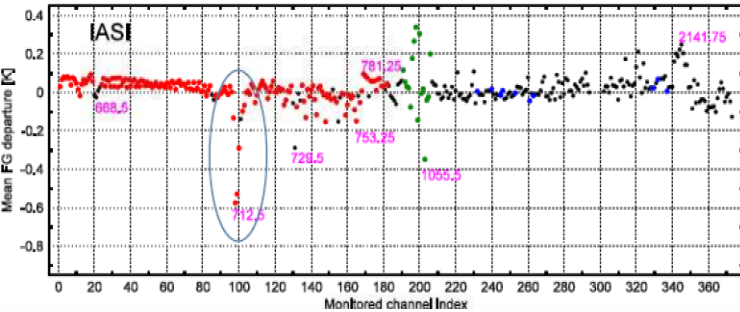
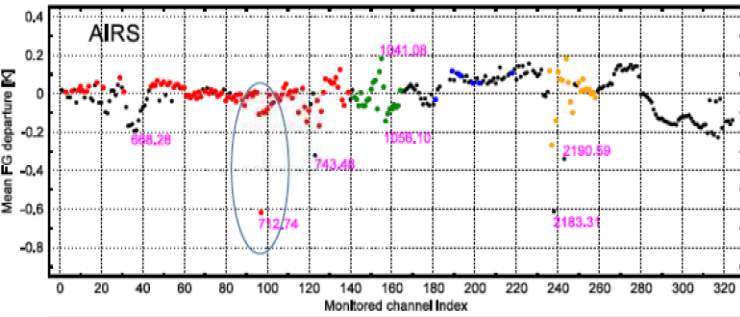
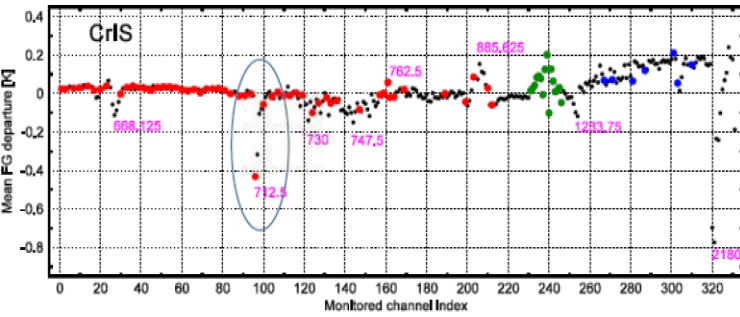


IASI-B-272



M. Matricardi

M. Matricardi



The mean O-B statistics in other IR channels shows that the anomaly can be seen in any IR channel sensing around 712 cm⁻¹ (or 14 micron) from CrIS, AIRS and both IASI instruments (A and B)



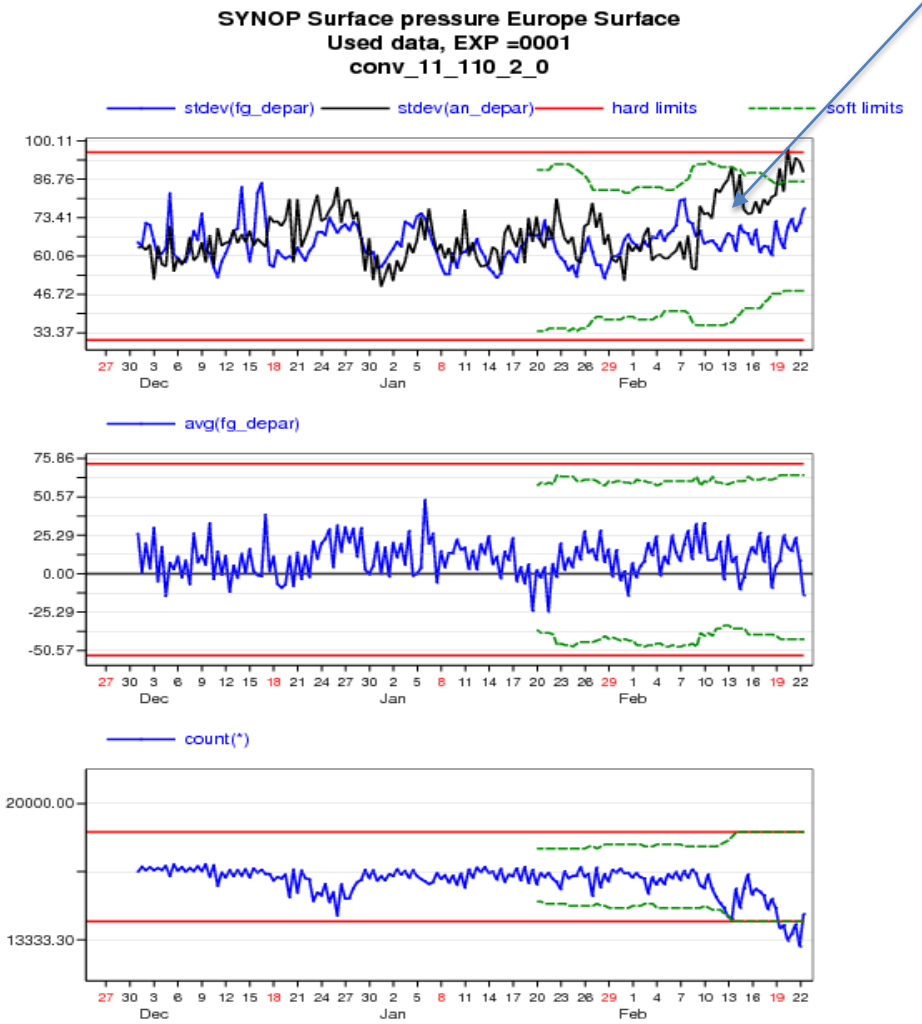
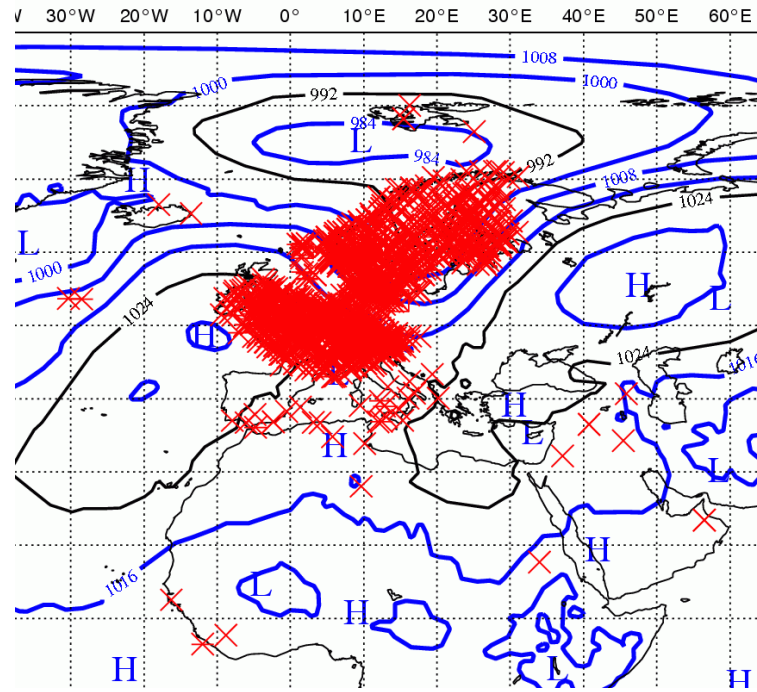
However, there is no evidence of a similar contamination signal in the IR window channels indicating that the signal cannot be attributed to residual clouds or aerosols

Decision to blacklist the affected channels (unwanted temperature increments)

Spectral signature of Hydrogen Cyanide (HCN)

HCN is a known pollutant associated with biomass burning and the alarms coincided with the Indonesian fires

Surface pressure from SYNOP @20120222



SYNOP Surface pressure Europe Surface : out of range:
(19 times in last 10 days for at least one item)
http://wedit.ecmwf.int/products/forecasts/satellite_check//do/get/satcheck/969/59315?showfile=true
Slightly: count(*)=13632, expected range: 14227.5(H) 18606.9(H)
Severely: stdev(fg_depar)=73.073 < stdev(an_depar)=88.41

SYNOP and METAR co-location

Before METAR QNH correction

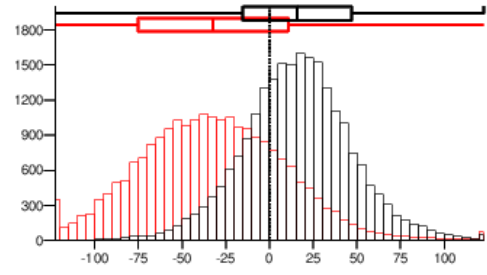
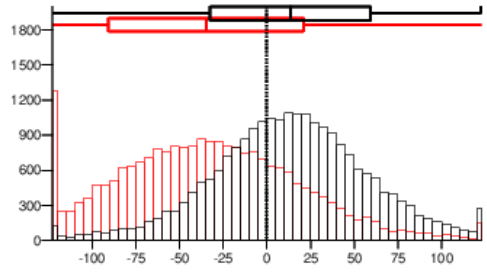
SYNP vs METR : 2020040100-2020041912(12)

ALL P [PA] Europe

used p

Background departure (o-b)
 nb= 22613 (ref= 22620) rms= 47.8 (65.7)
 mean= 13.5 (-34.8) std= 45.9 (55.7)
 min= -229. (-300.) max= 244. (280.)

Analysis departure (o-a)
 nb= 22613 (ref= 22620) rms= 35.0 (53.8)
 mean= 15.9 (-32.3) std= 31.2 (43.0)
 min= -153. (-254.) max= 167. (283.)



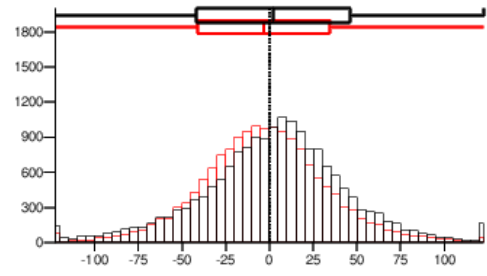
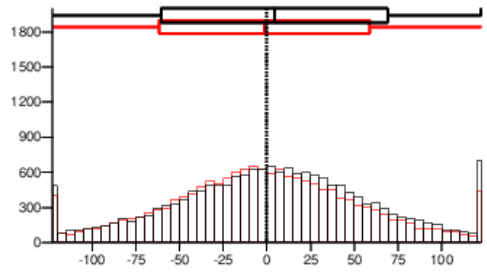
SYNP vs METR : 2020040100-2020041912(12)

ALL P [PA] N.Amer

used p

Background departure (o-b)
 nb= 17405 (ref= 16143) rms= 64.9 (60.3)
 mean= 4.28 (-1.40) std= 64.8 (60.3)
 min= -379. (-409.) max= 373. (321.)

Analysis departure (o-a)
 nb= 17405 (ref= 16143) rms= 44.0 (38.0)
 mean= 2.19 (-3.40) std= 43.9 (37.8)
 min= -215. (-412.) max= 219. (294.)



After METAR QNH correction (Apr 2020)

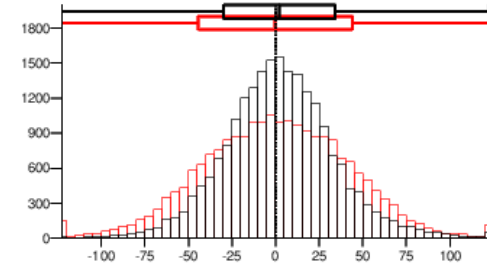
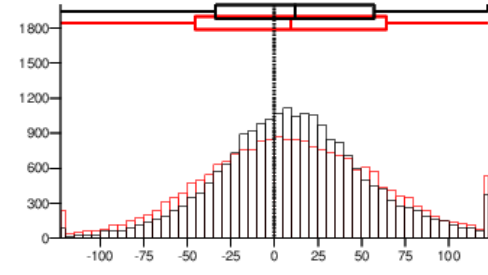
SYNP vs METR : 2020050100-2020051912(12)

ALL P [PA] Europe

used p

Background departure (o-b)
 nb= 21527 (ref= 21534) rms= 47.0 (55.7)
 mean= 11.8 (9.43) std= 45.5 (54.9)
 min= -301. (-347.) max= 249. (295.)

Analysis departure (o-a)
 nb= 21527 (ref= 21534) rms= 32.0 (44.2)
 mean= 2.22 (-0.200) std= 31.9 (44.2)
 min= -180. (-366.) max= 230. (312.)



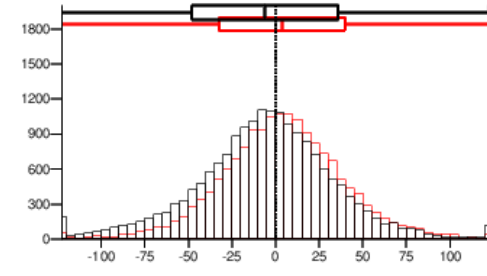
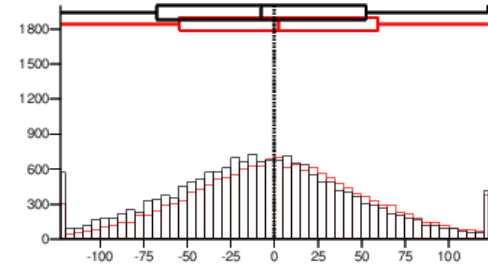
SYNP vs METR : 2020050100-2020051912(12)

ALL P [PA] N.Amer

used p

Background departure (o-b)
 nb= 18161 (ref= 16579) rms= 60.5 (56.9)
 mean= -7.47 (2.48) std= 60.0 (56.8)
 min= -433. (-371.) max= 426. (461.)

Analysis departure (o-a)
 nb= 18161 (ref= 16579) rms= 42.4 (36.3)
 mean= -6.10 (3.64) std= 42.0 (36.1)
 min= -225. (-253.) max= 213. (312.)



Observation statistics and data impact

- Changing the data usage or the assimilation algorithms impact the fit of “used” observations to the background and analysis
- The change of fit is indicative of the (positive/negative/neutral) impact of the change being tested
- The availability of forecast departures (O-F computed similarly to O-B but for QC data and for longer ranges) allows verification against observations

Diagnostics in observation space based on departures

Consistency diagnostics on departures (Desroziers et al. 2005)

$$E[\mathbf{d}_b^o (\mathbf{d}_b^o)^T] = \mathbf{R} + \mathbf{HBH}^T \quad \text{Global consistency check}$$

\mathbf{d}_b^o (O-B) Background departures

$$E[\mathbf{d}_a^o (\mathbf{d}_b^o)^T] = \mathbf{R} \quad \text{Observation errors estimation}$$

\mathbf{d}_a^o (O-A) Analysis departures

$$E[\mathbf{d}_b^a (\mathbf{d}_b^o)^T] = \mathbf{HBH}^T \quad \text{Background errors check}$$

\mathbf{d}_b^a (A-B) Increments

$$E[\mathbf{d}_b^a (\mathbf{d}_a^o)^T] = \mathbf{HAH}^T \quad \text{Analysis errors}$$

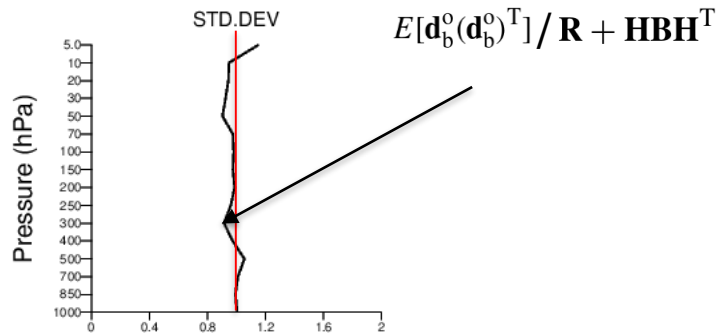
These diagnostics are easily computed by the monitoring system for cases without obs error correlations

$$E[\mathbf{d}_b^0 (\mathbf{d}_b^0)^T] = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$$

- Observation and background errors are assumed uncorrelated
- In a well tuned system this equation verifies
- Otherwise further tuning is needed for R, B or H

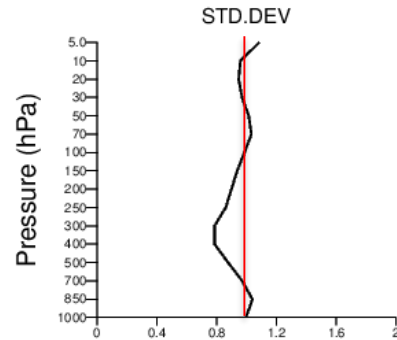
TEMP temperature

2021090100-2021090400(12)
TEMP-T N.Hemis
used T



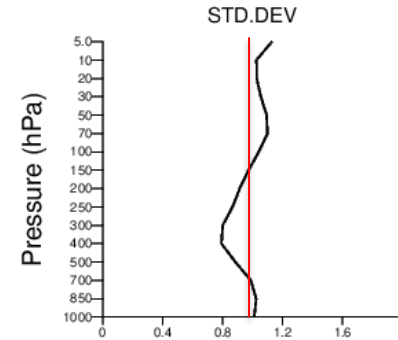
TEMP U-comp

2021090100-2021090400(12)
TEMP-Uwind N.Hemis
used U



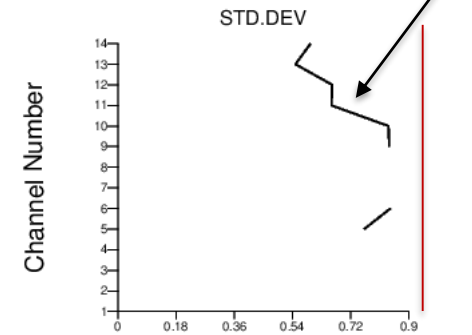
TEMP V-comp

2021090100-2021090400(12)
TEMP-Vwind N.Hemis
used V

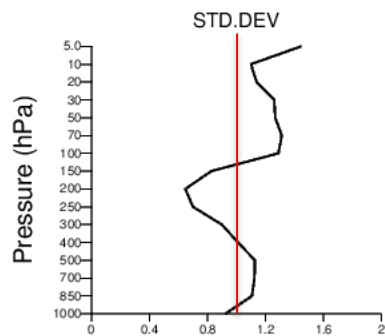


METOP-A AMSUA

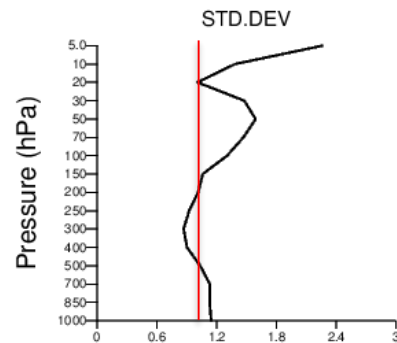
2021090100-2021090400(12)
METOP-A AMSU-A Tb N.Hemis
used Tb METOP-A AMSUA



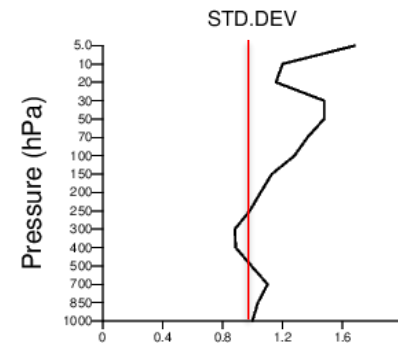
2021090100-2021090400(12)
TEMP-T Tropics
used T



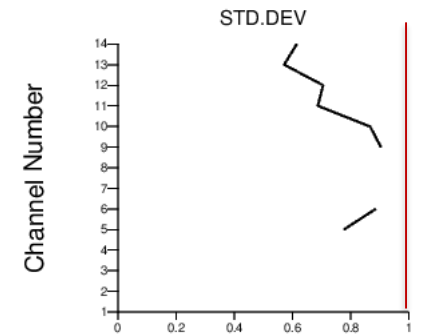
2021090100-2021090400(12)
TEMP-Uwind Tropics
used U



2021090100-2021090400(12)
TEMP-Vwind Tropics
used V



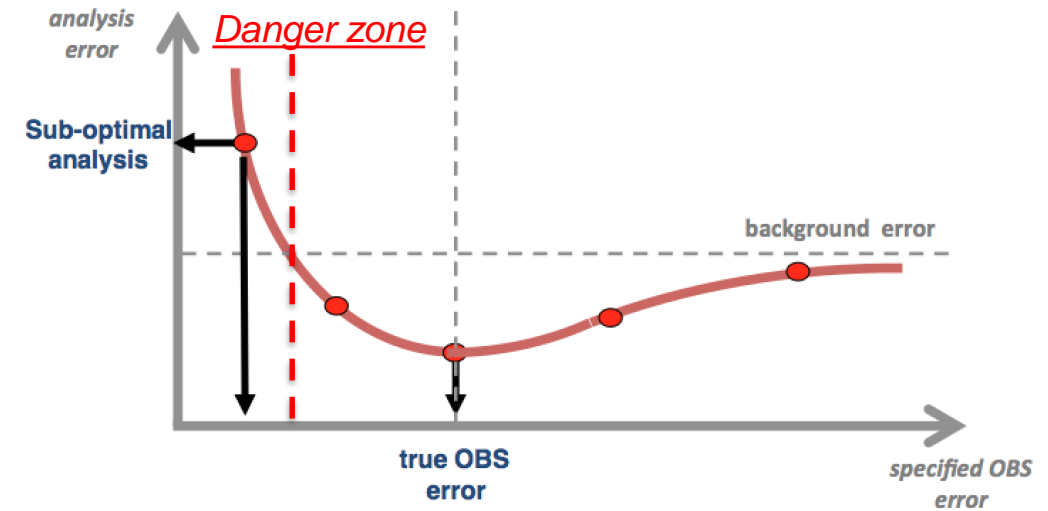
2021090100-2021090400(12)
METOP-A AMSU-A Tb Tropics
used Tb METOP-A AMSUA



Diagnostics in observation space based on departures

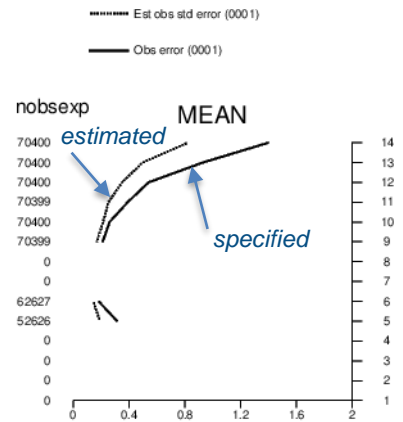
$$E[\mathbf{d}_a^o (\mathbf{d}_b^o)^T] = \mathbf{R}$$

$$E[\mathbf{d}_b^a (\mathbf{d}_b^o)^T] = \mathbf{H}\mathbf{B}\mathbf{H}^T$$

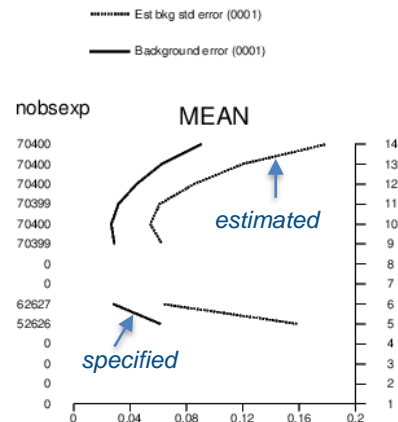


METOP-A/ AMSUA (01-04 Sep 2021)

Estimated and specified Obs errors

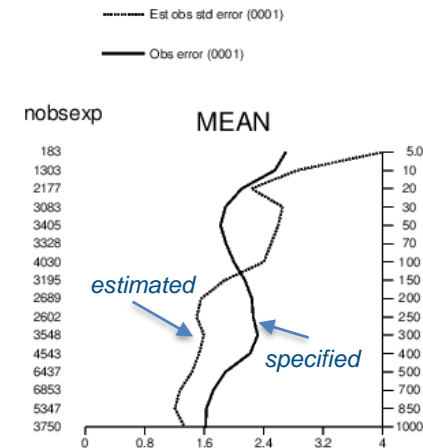


Estimated and specified FG errors

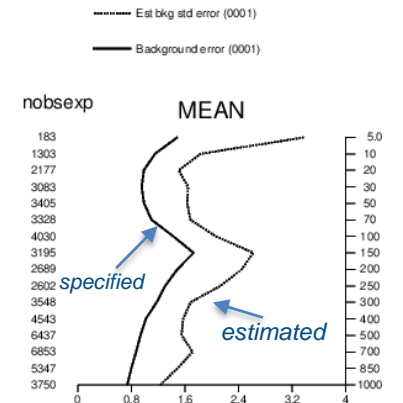


TEMP V-comp Tropics (01-04 Sep 2021)

Estimated and specified Obs errors



Estimated and specified FG errors



*Consistency check for observation errors but can also be used to estimate them.
Works well if B is properly specified*

Diagnostics from the EDA

Monitoring of EDA spread

$$E[\mathbf{d}_b^o (\mathbf{d}_b^o)^T] = \mathbf{R} + \mathbf{H}\mathbf{B}\mathbf{H}^T$$

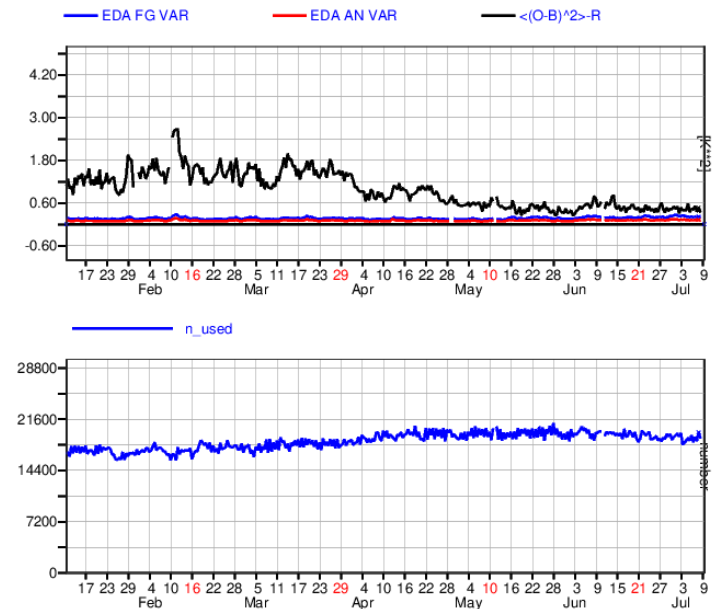
EDA control
(Variance of FG departures)

Computed from EDA members (variance of FG in obs points)

If(\mathbf{R} and $\mathbf{H}\mathbf{B}\mathbf{H}^T$ are correctly specified)

- Check the consistency of EDA spread with what is expected
- Monitor the evolution of EDA spread (including seasonal variations and model upgraded)

Temperature from Radiosondes 0 – 100 hPa North Hemi Extra-tropics



Diagnostics from the EDA

(Cardinali et al. 2004, Liu and Kalnay 2008)

Observations influence on the analysis

$$\mathbf{S}^o = \frac{\partial \hat{\mathbf{y}}^a}{\partial \mathbf{y}^o} = \mathbf{K}^T \mathbf{H}^T = \mathbf{R}^{-1} \mathbf{H} \mathbf{P}^a \mathbf{H}^T = \frac{1}{n-1} \mathbf{R}^{-1} (\mathbf{H} \mathbf{X}^a) (\mathbf{H} \mathbf{X}^a)^T$$

$$S_{jj}^o = \frac{\partial \hat{y}_j^a}{\partial y_j^o} = \left(\frac{1}{n-1} \right) \frac{1}{\sigma_j^2} \sum_{i=1}^n (\mathbf{H} \mathbf{X}^{ai})_j \times (\mathbf{H} \mathbf{X}^{ai})_j$$

Self-sensitivity

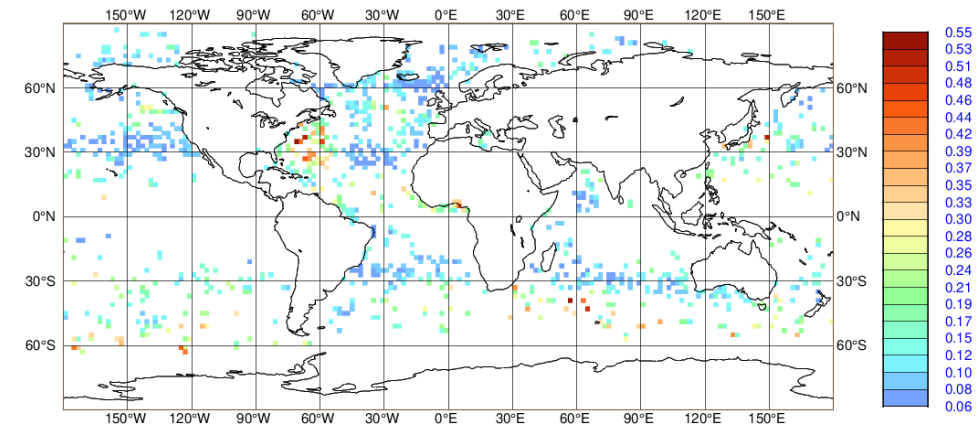
Assumed obs errors

EDA analyses in obs space

$$\mathbf{H} \mathbf{X}^{ai} \cong h(\mathbf{x}^{ai}) - \frac{1}{n} \sum_{i=1}^n h(\mathbf{x}^{ai})$$

- Input information readily available (No computation time)
- The computation assumes no observation error correlations (true for most data types)
- Less expensive way to monitor the impact of observations on analysis (weather related) and has the potential to flag data with gross errors (unusual influence)
- This diagnostics needs to be assessed against other implementations of self-sensitivity (Cardinali 2013)

PRESSURE FROM BUOY
 SELSENS (USED)
 DATA PERIOD = 2021-09-02 09 - 2021-09-03 21
 EXP = 0001, CHANNEL = 1
 Min: 0.044 Max: 0.926 Mean: 0.138
 GRID: 2.00x 2.00



Random Error estimations using Triple-collocation

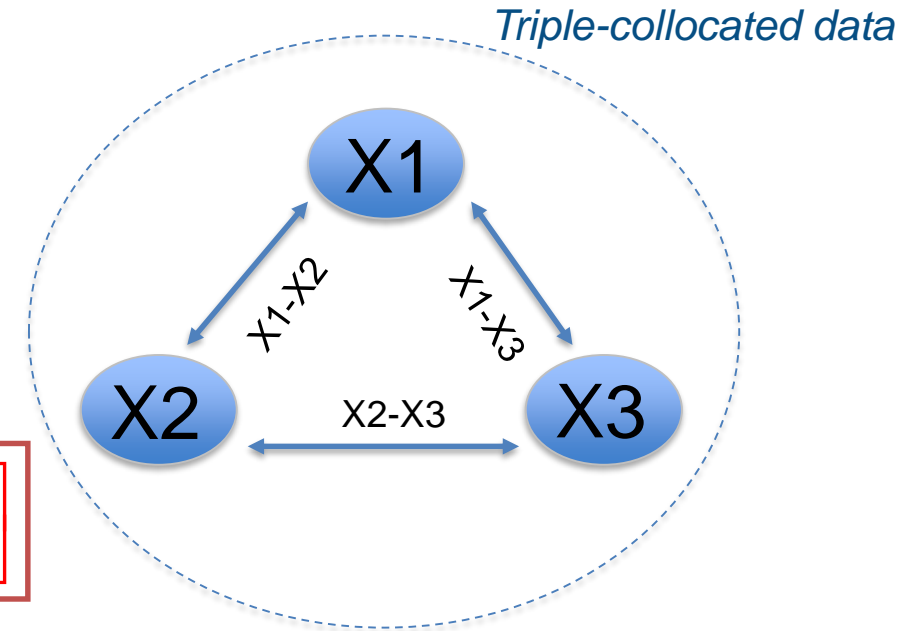
S. Abdalla et al 2017

- X_i (3 or more) should be independent and measure the same parameter at the *same* location and time (triple collocation criteria)
- Each measurement X_i consists of a unknown truth T (calibrated with β_i) and an unknown error e_i ($X_i = \alpha_i + \beta_i T_i + e_i$)
- The unknown error variance can be written as:

$$\langle e_p^2 \rangle = 0.5 \left[\langle (X_p - X_{p1})^2 \rangle + \langle (X_p - X_{p2})^2 \rangle - \langle (X_{p1} - X_{p2})^2 \rangle \right]$$

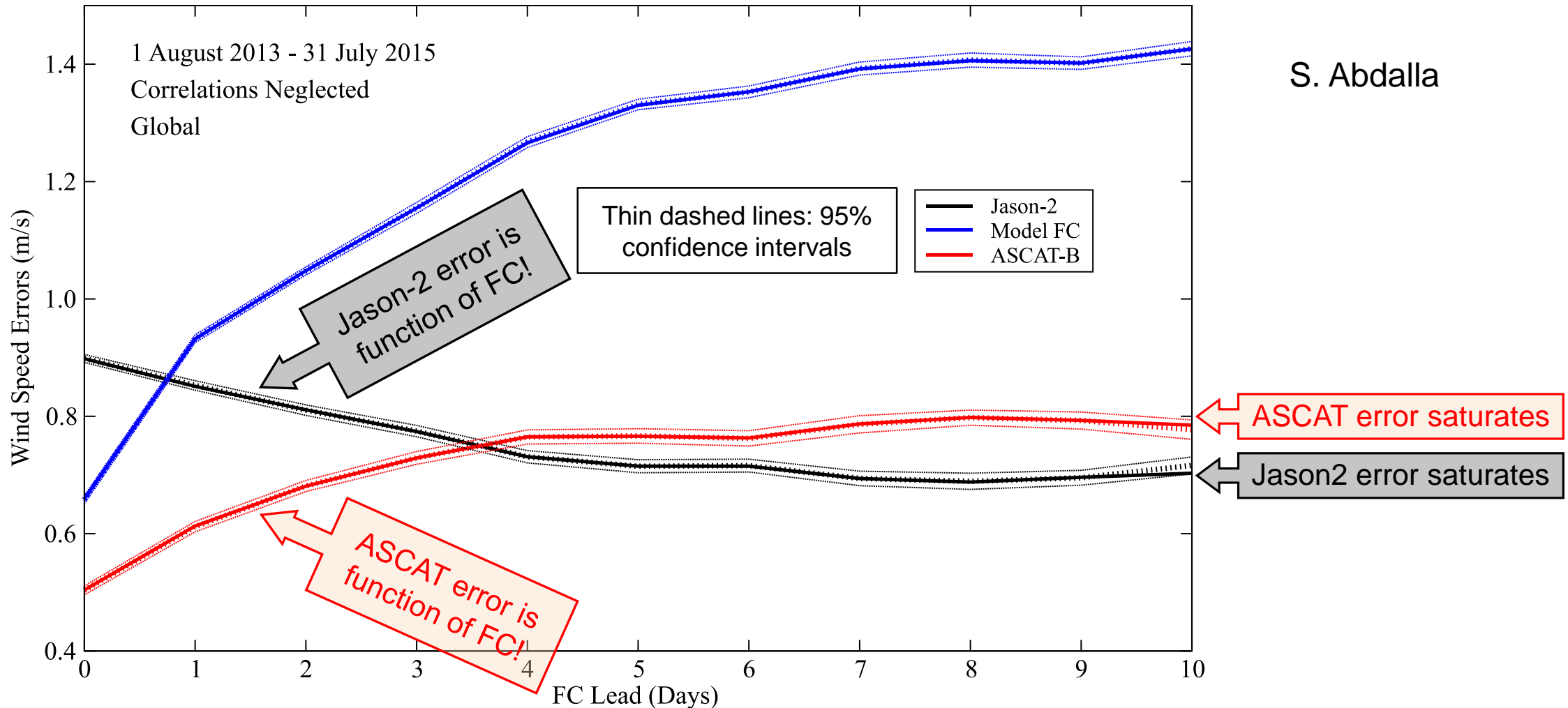
$\langle \dots \rangle$ is the average, $p1$ & $p2$ refer to the other 2 systems.

- This assumes there is **no correlation between the errors** in the triplets. Calibration constants, β_p , are found by iteration.
- The third dataset is usually the model (analysis and forecasts at various ranges). Forecast departures are potentially correlated with **used** datasets but the correlations tend to fade with forecast range



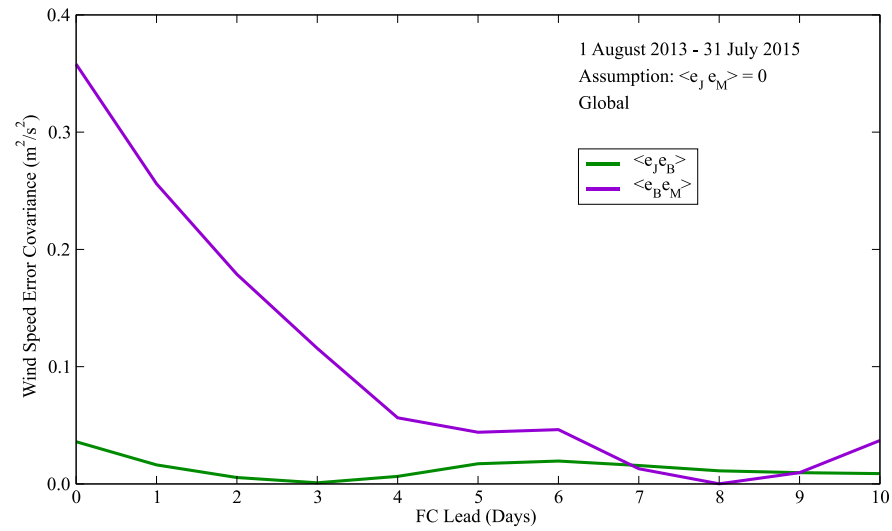
Initially the Errors estimated by ignoring the correlations

S. Abdalla

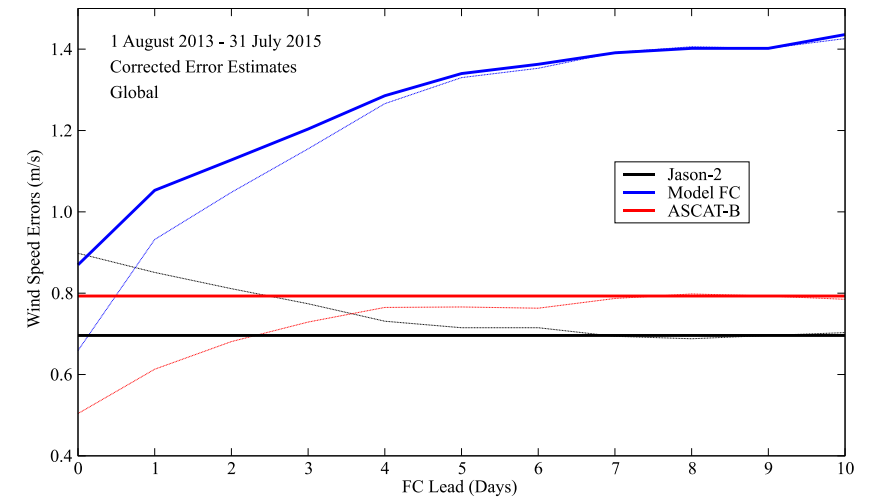


We can estimate the correlations (2 out of 3 only) and correct the error estimates.

Estimation of correlations (covariances)



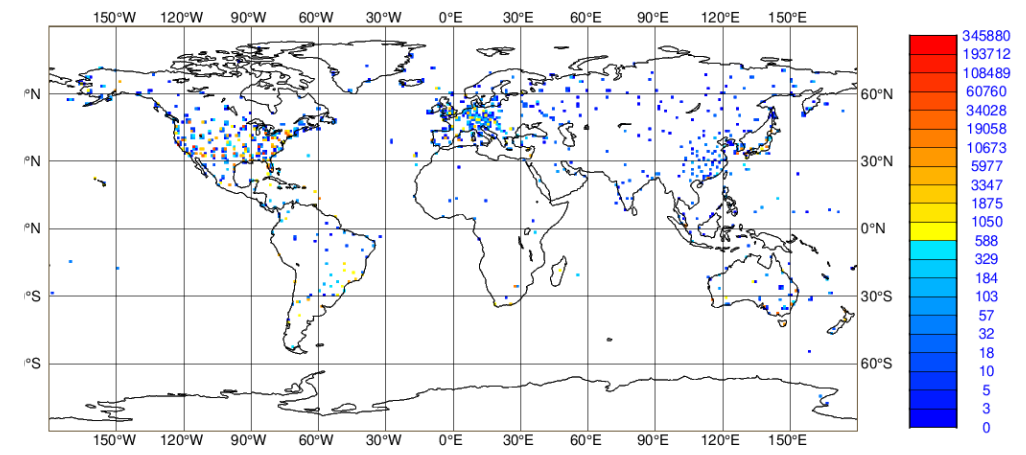
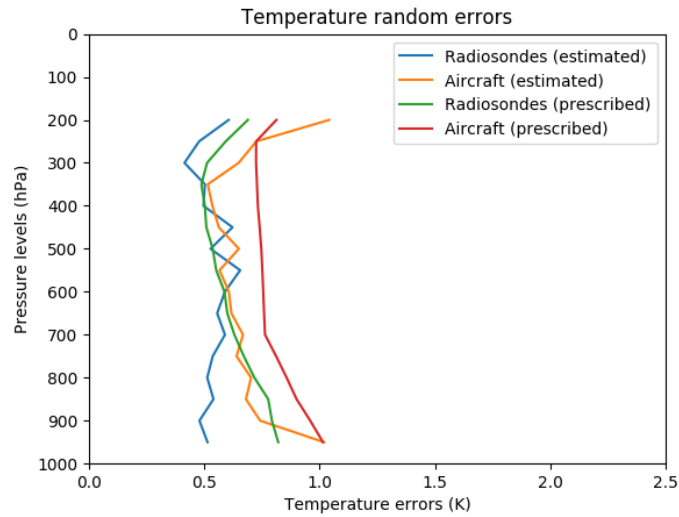
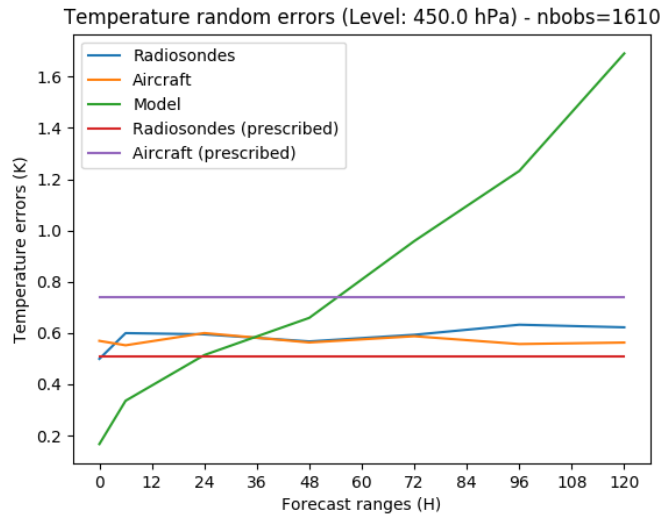
Errors corrected for correlations



Radiosonde/Aircraft/Model (temperature and wind in the troposphere)

Co-located aircraft and radiosondes

TEMP FROM ALL
 NUMBER OF OBSERVATIONS (ALL)
 DATA PERIOD = 2020-05-09 21 - 2020-05-10 09
 EXP = 0001, CHANNEL = 1
 Min: 0 Max: 187350 Mean: 61501.4 Total: 1857756.0
 GRID: 1.00x 1.00



Conclusions

- **Observations monitoring activities are a key component of the data assimilation diagnostics**
- **An integrated monitoring system improves the data usage by assisting operational and research activities**
- **Monitoring activities benefits from a wide range of resources (internal and external). Collaboration and timely sharing of information is vital**
- **Various diagnostics are available to check the consistency of the DA system and to suggest estimates of errors**