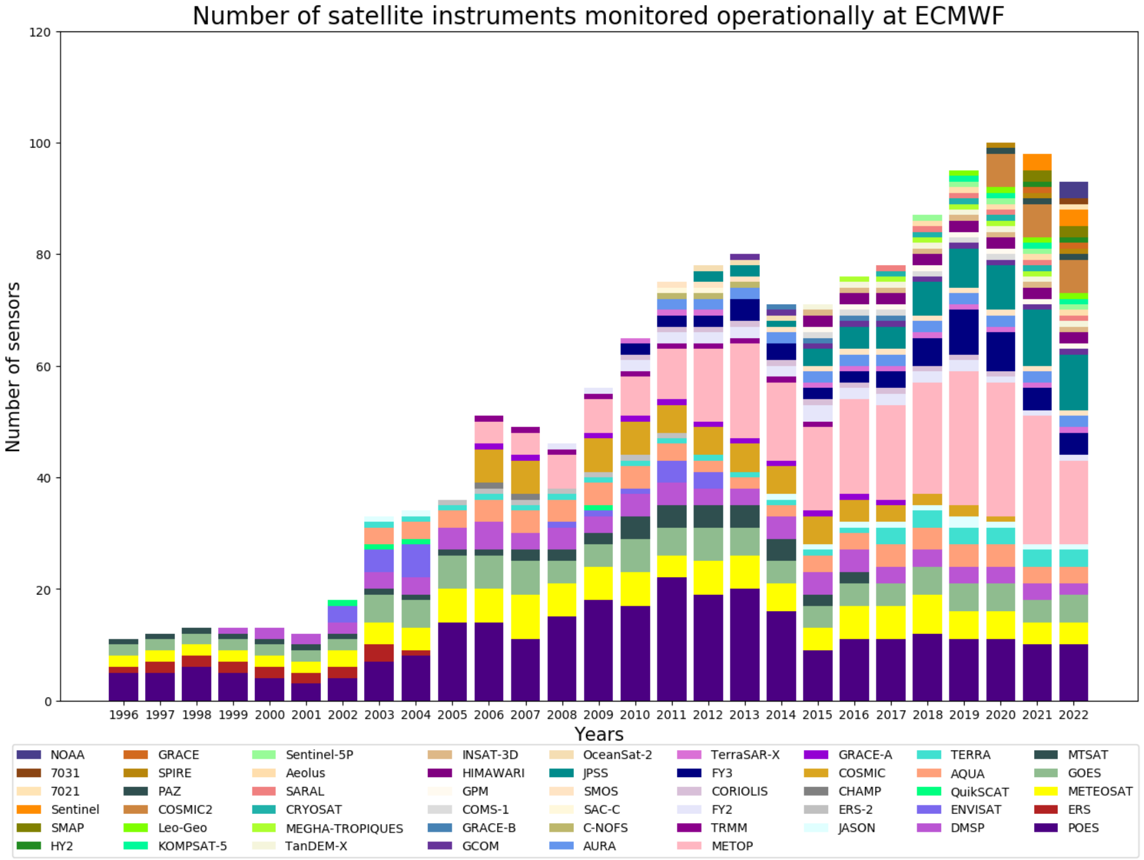
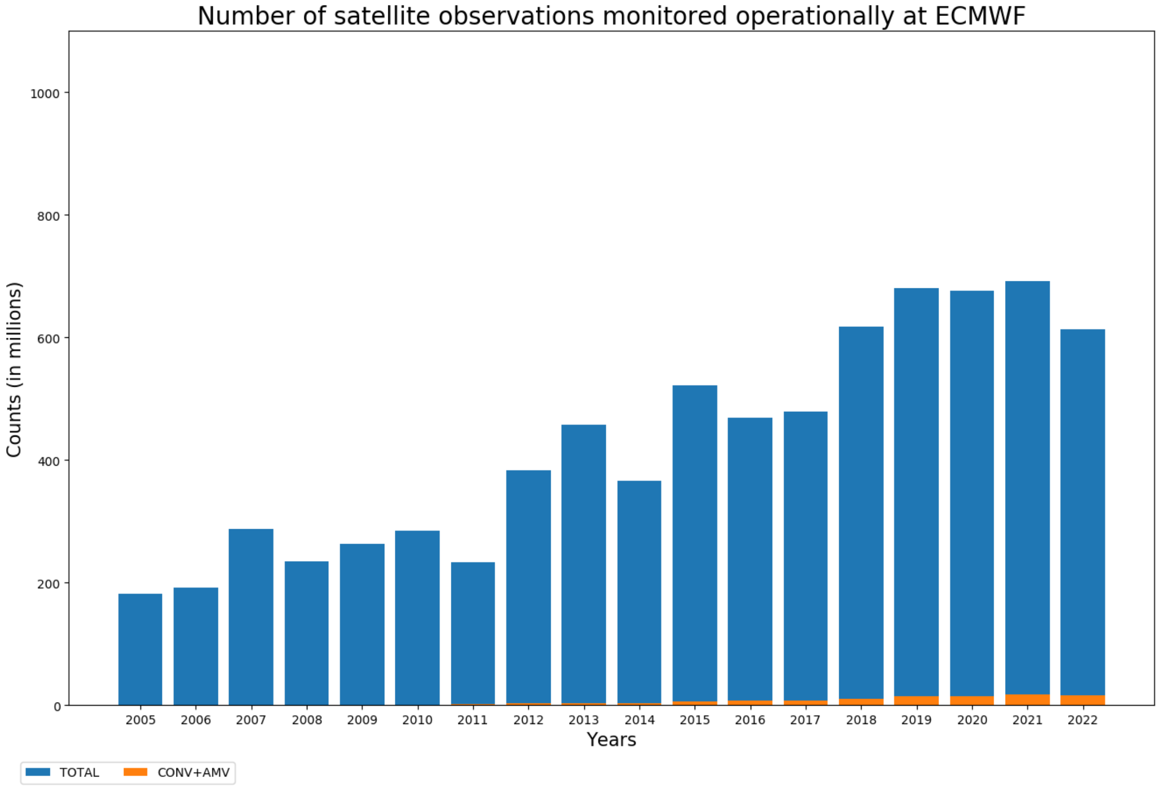
Use of Machine learning for the detection and classification of observation anomalies

M. Dahoui

mohamed.dahoui@ecmwf.int

Evolution of data counts and diversity

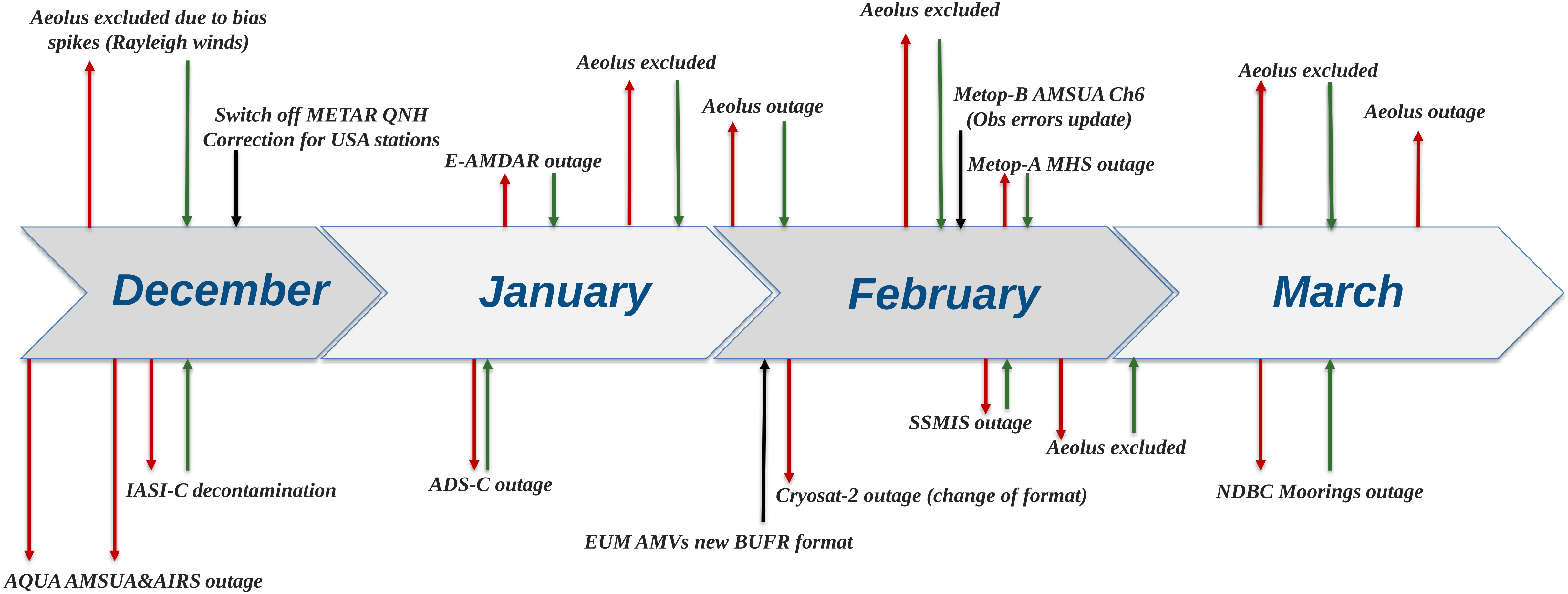




- ~ 700 million pieces of observations received daily and ~60 millions active. Data characteristics are subject to variation with possible consequences on the data impact.

Monitoring of observations: Why ?

Data events (DJFM 2021)

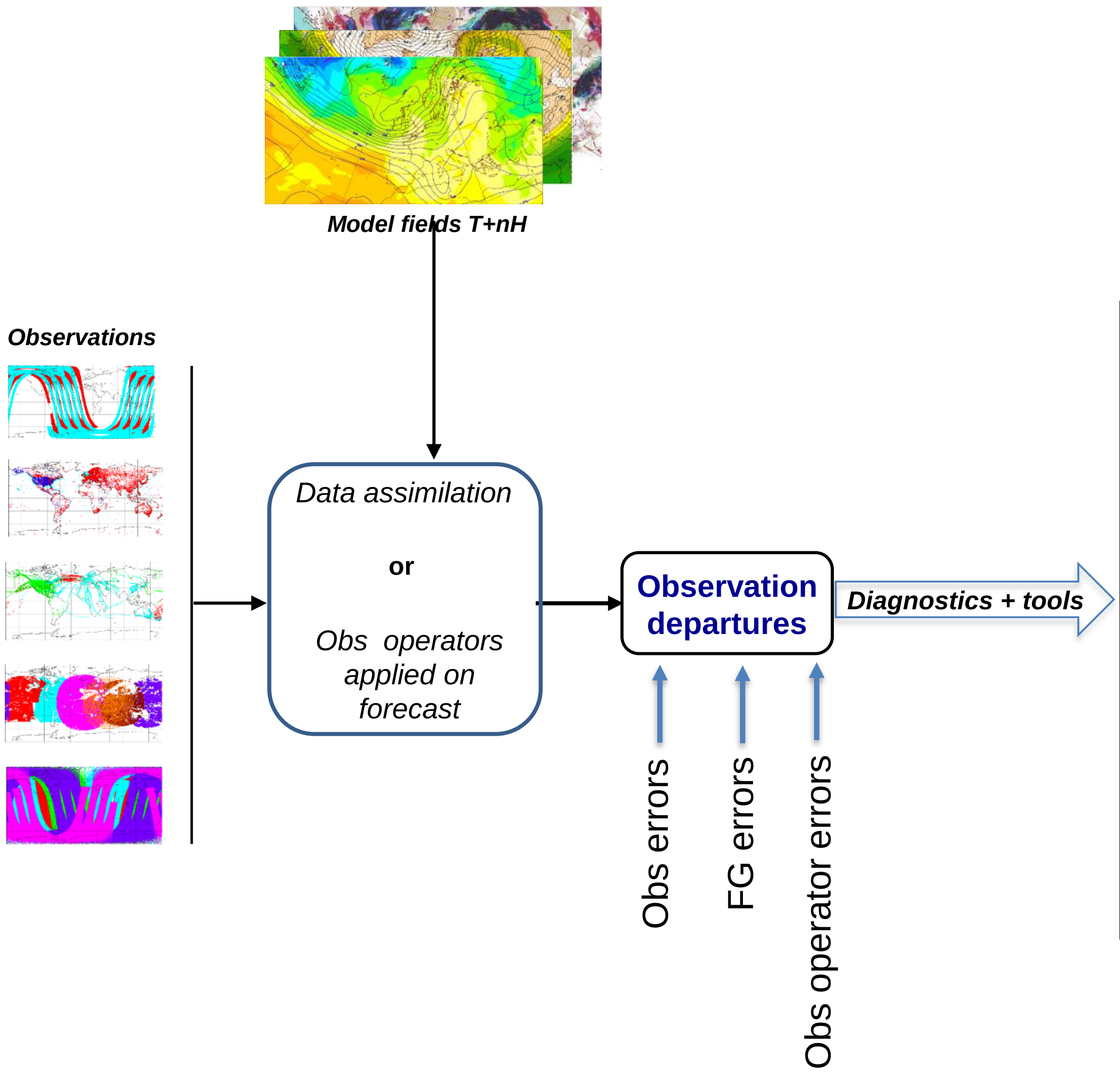


Monthly update of data selections (4D-VAR in-situ and LDAS)

ECMWF

https://confluence.ecmwf.int/display/FCST/Observations+data+events

Monitoring of observations



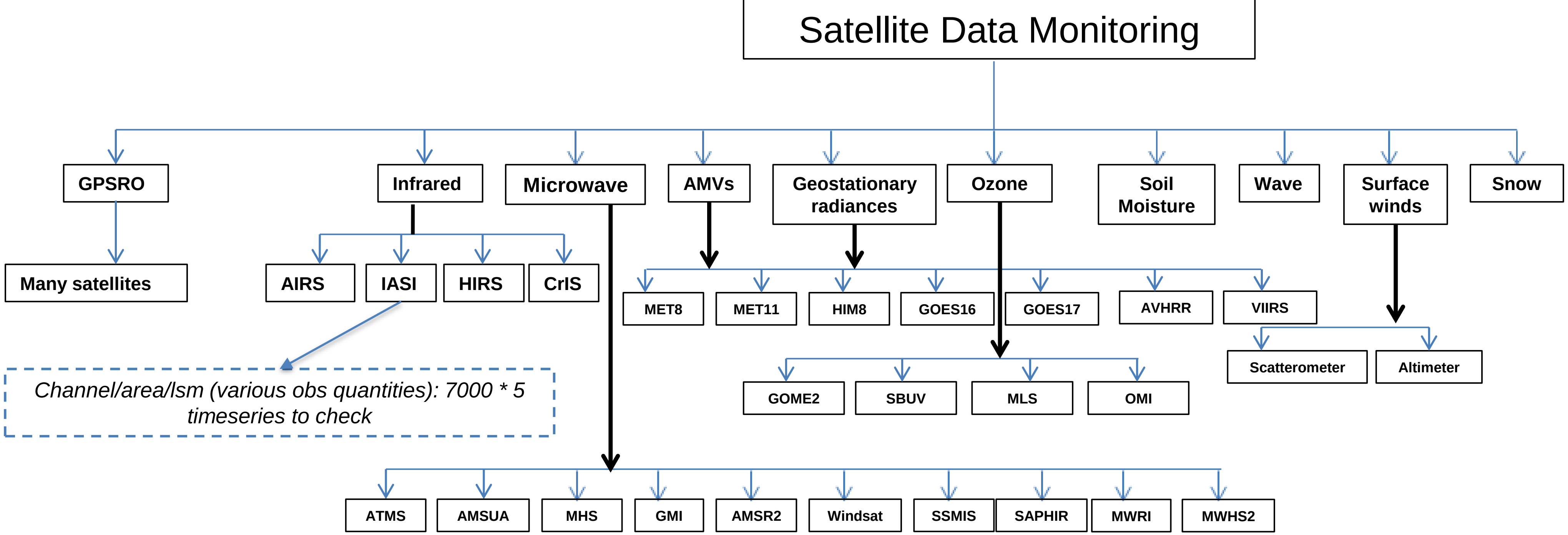
- Observations quality issues
- Estimation of Obs Errors
- Model/data assimilation issues (redundancies in the obs system)
- Observations impact (analysis and forecasts)
- Forecast verification
- Support model and data assimilation developments

Automatic data checking system

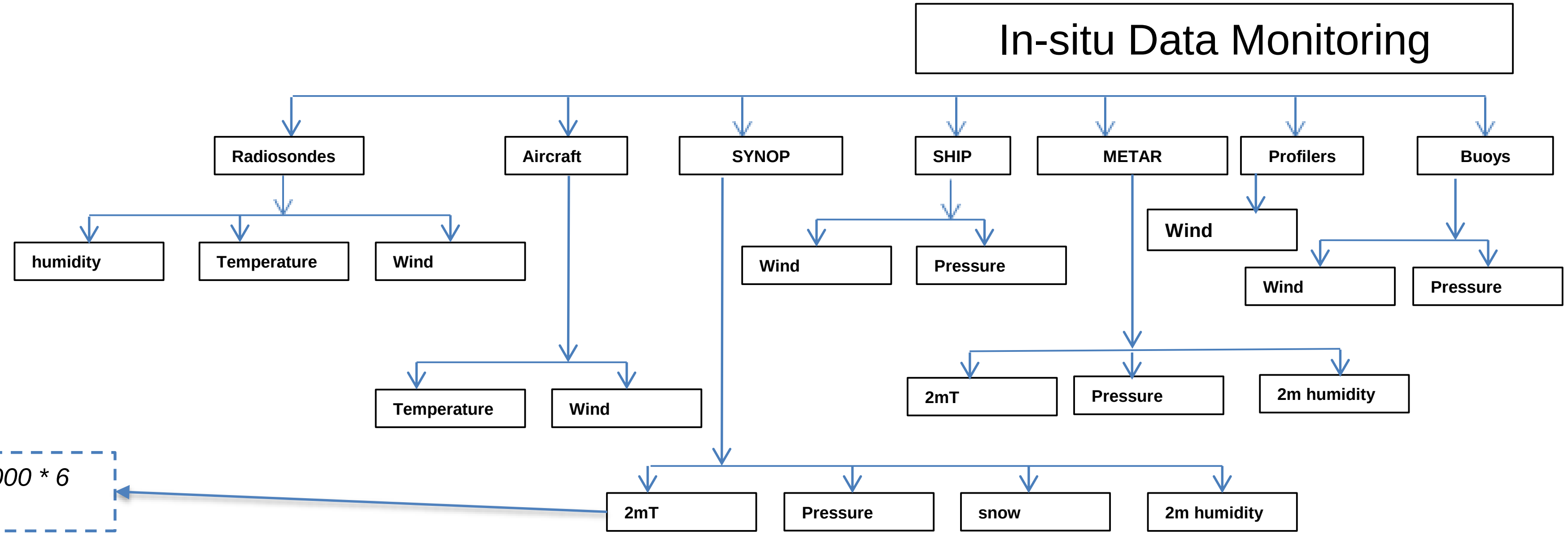
- The large amount of active observations makes it difficult to timely detect availability/quality issues

- An automatic data checking system is implemented at ECMWF to continuously monitors satellite and in-situ data (main trigger for corrective actions). Warnings available to internal and selected external users, at user specified level of detail

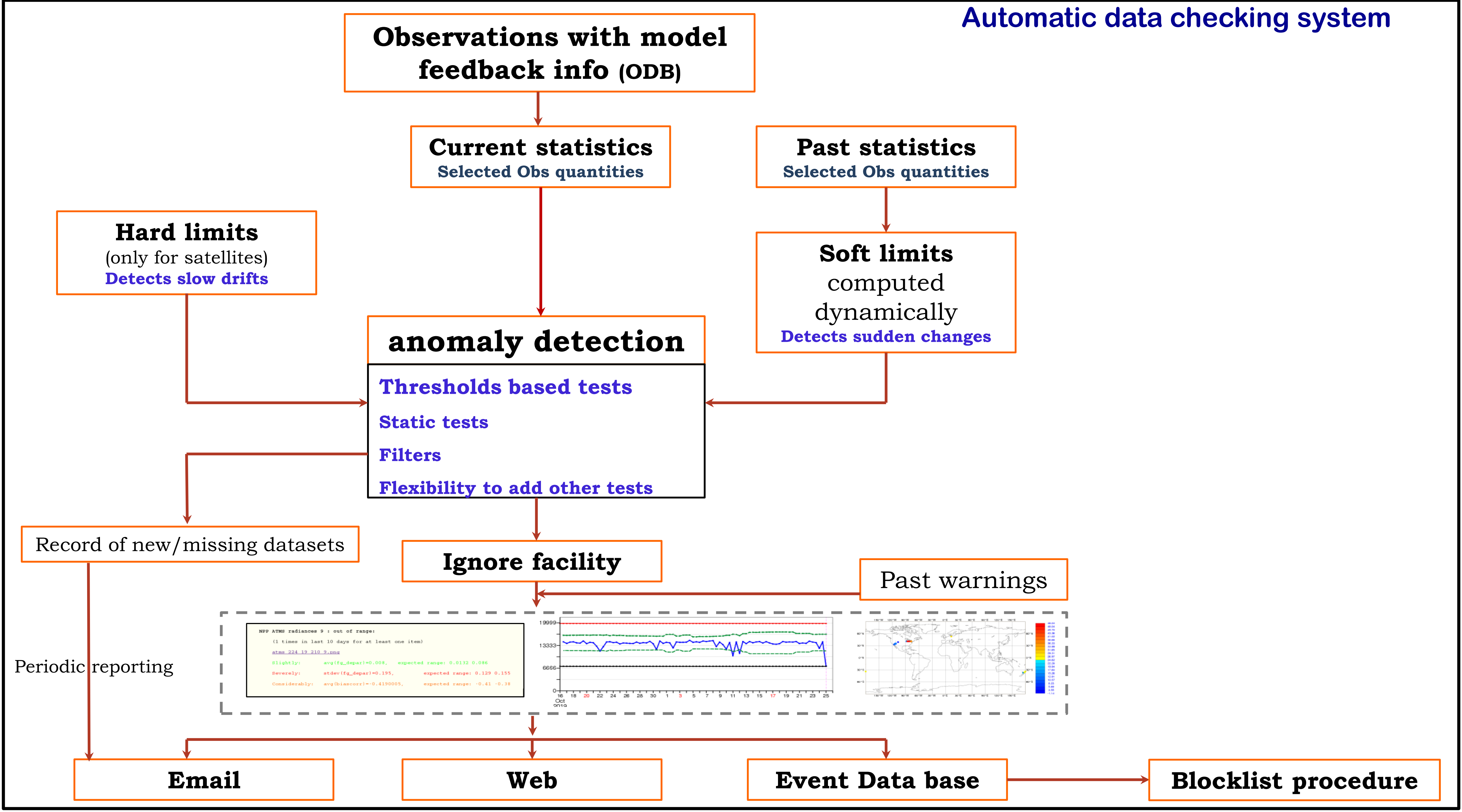
- The same system is used to detect improved in-situ data that are currently excluded (timely activation of improved observations



Detection needs to be well tuned to avoid reporting large number of irrelevant events



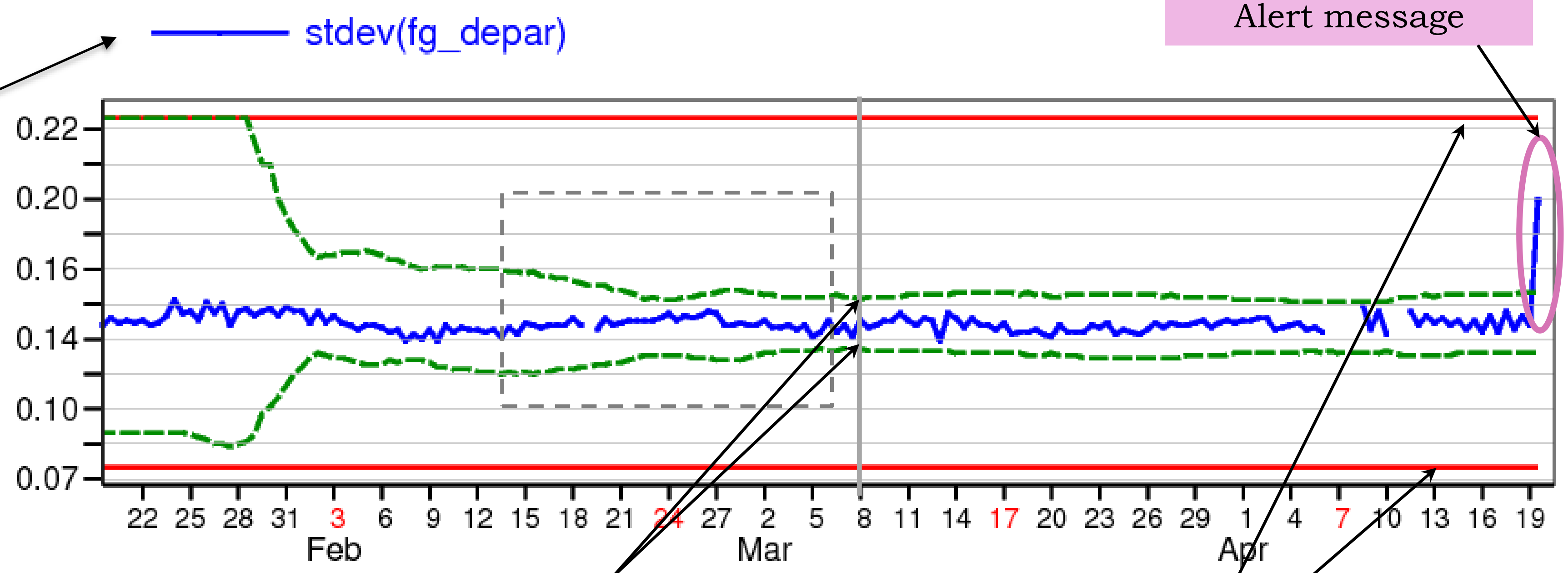
*Statid/layer/area/varno (various obs quantities): 13000 * 6 timeseries to check*


Automatic data checking system
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
Blocklist procedure

Automatic data checking system

Observation quantities of interest: data counts, stdev O-FG, bias correction, etc

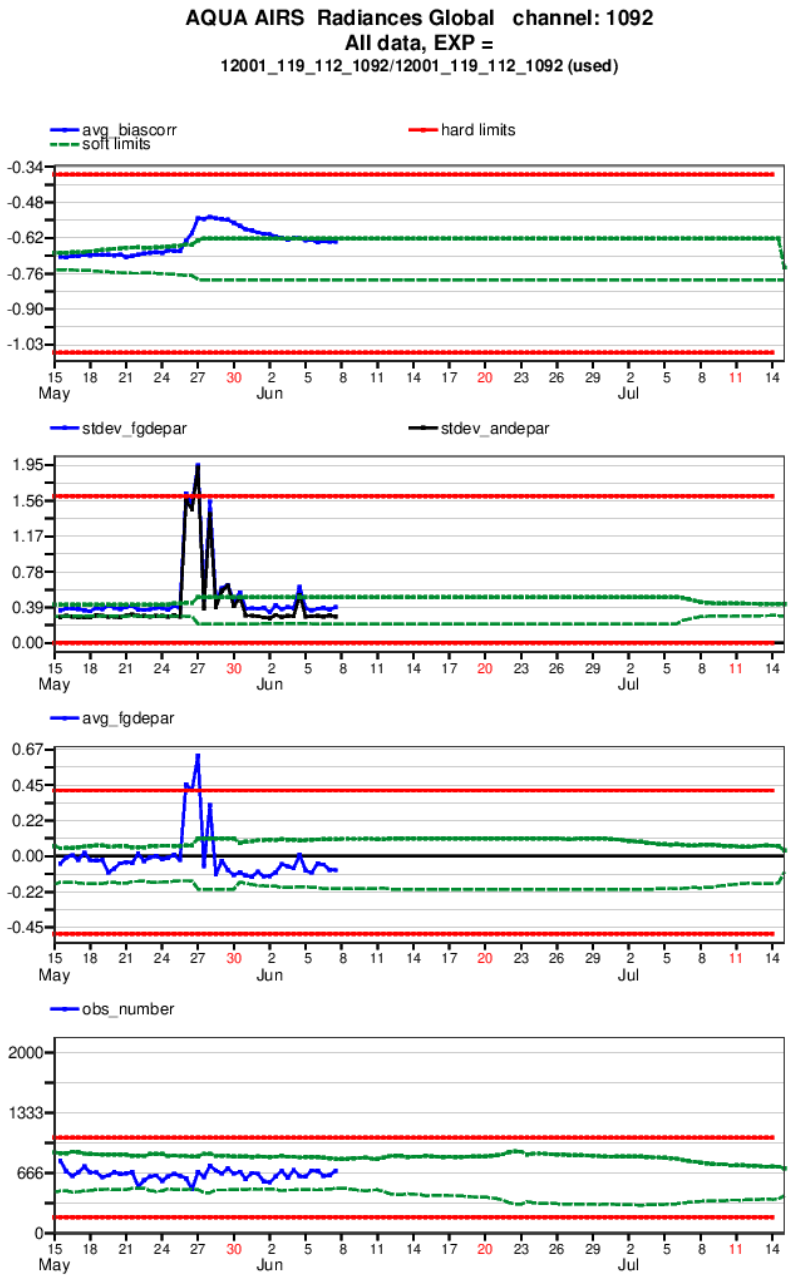
Alert message



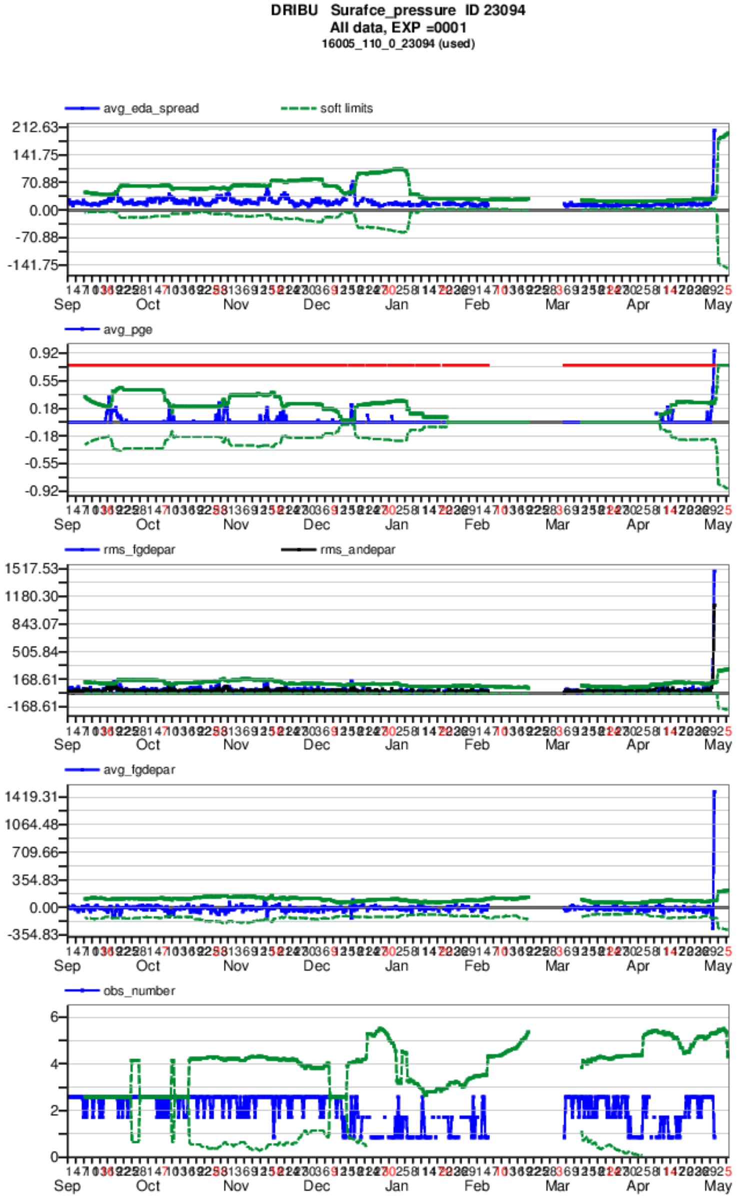
Soft limits: Detects sudden changes. Updated dynamically: (5±stdev of statistics. calculated from past statistics over a recent period ending 2 days earlier and excluding extremes)

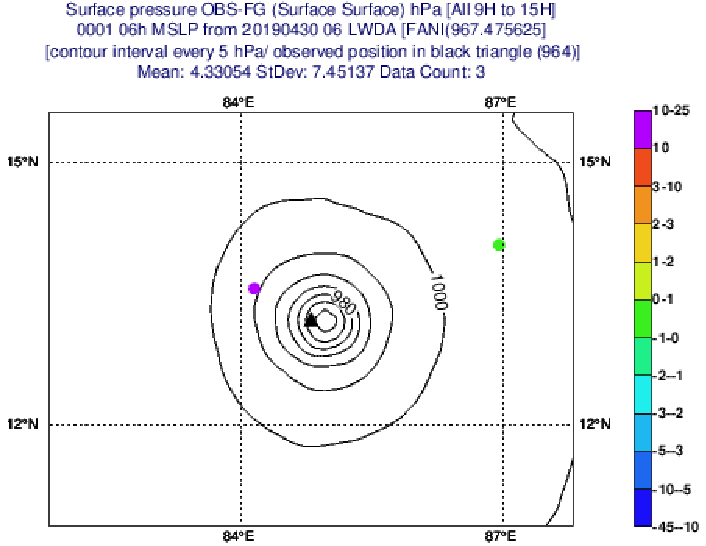
Hard limits: adjusted manually based on long time series. Detects drifts

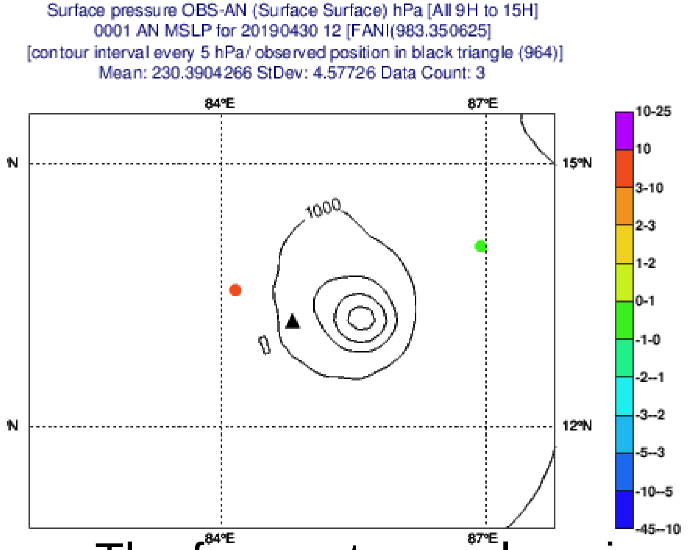
Automatic data checking system (examples)



Automatic data checking system (examples)

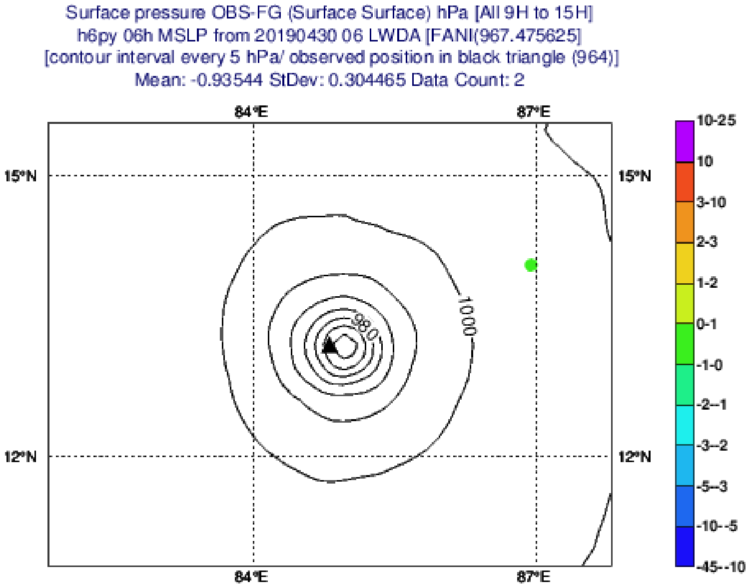


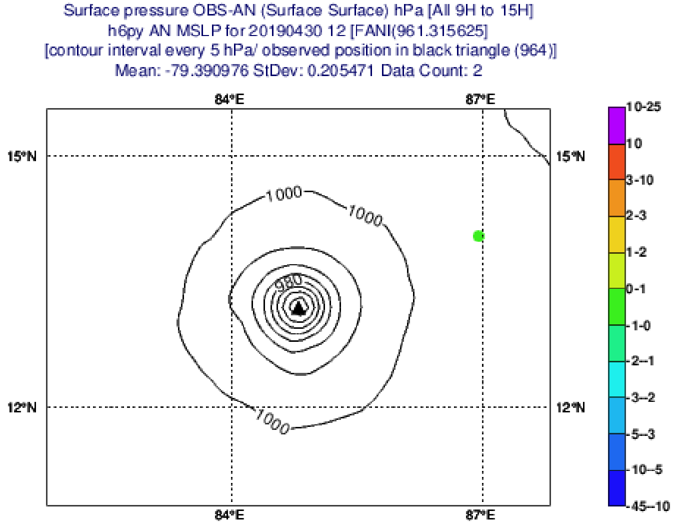




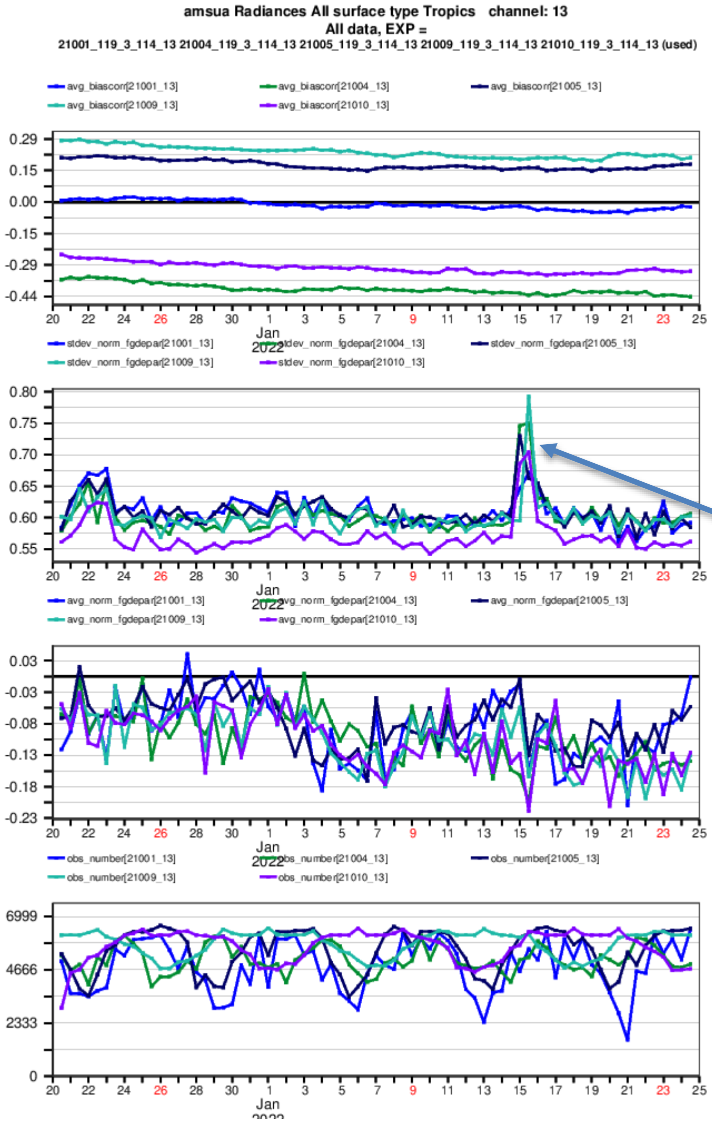
The forecast was also - impacted

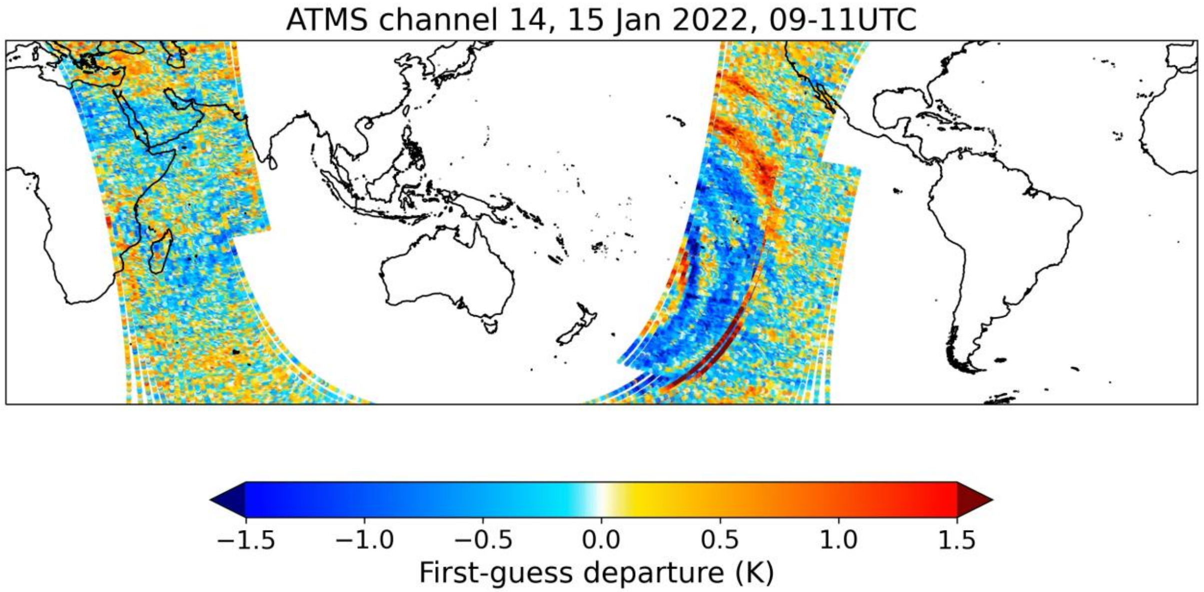
Denial of the buoy 23094



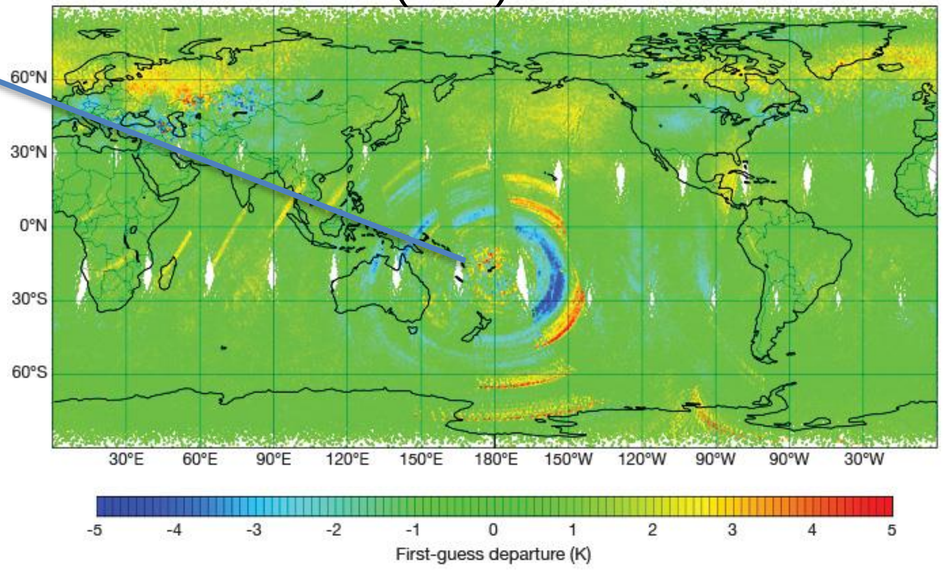


AMSUA Ch13 Tropics



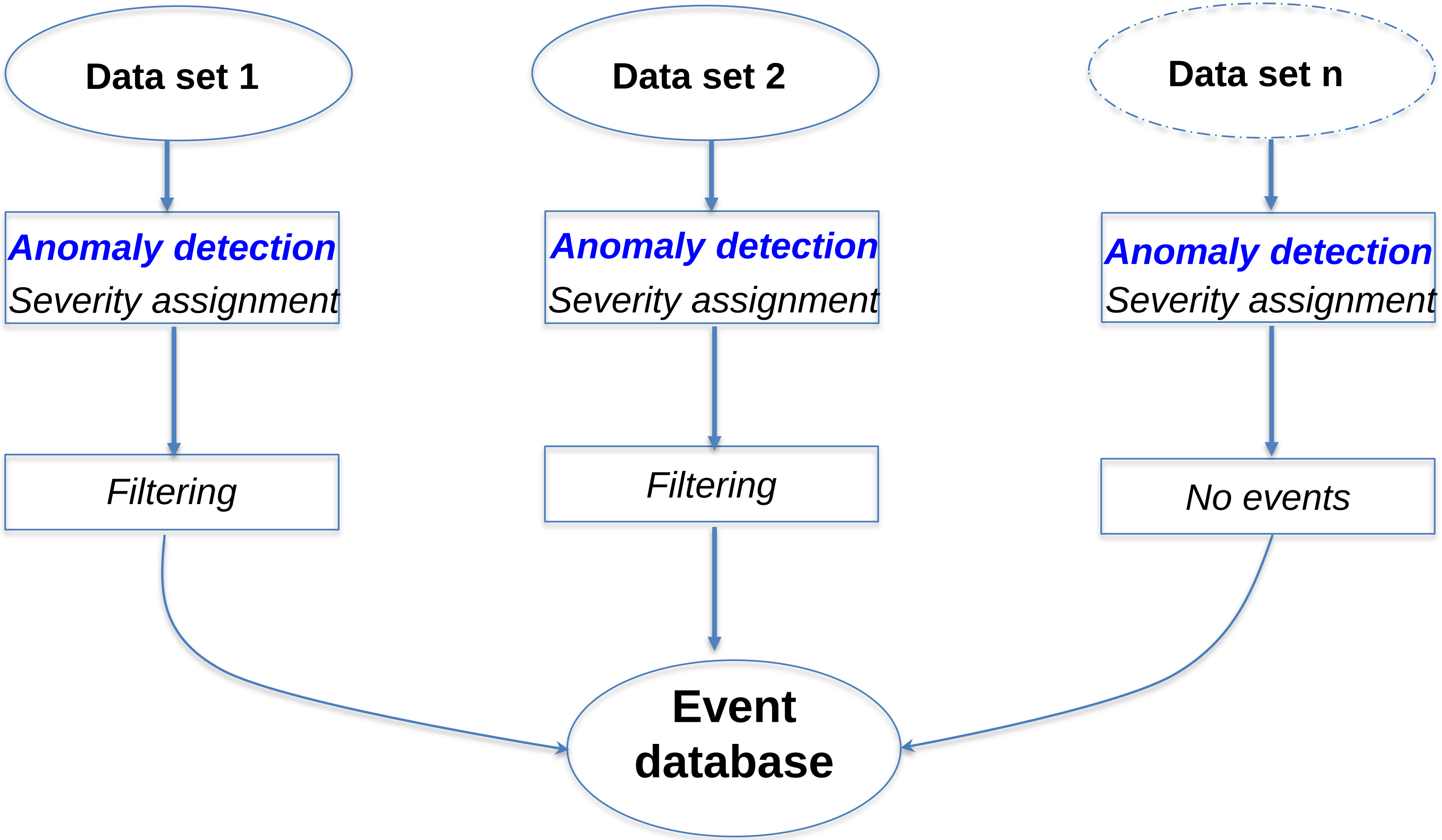


IASI Ch 92 (O-B) – 15/02/2022



https://www.ecmwf.int/en/about/media-centre/news/2022/hunga-tonga-eruption-seen-ecmwf

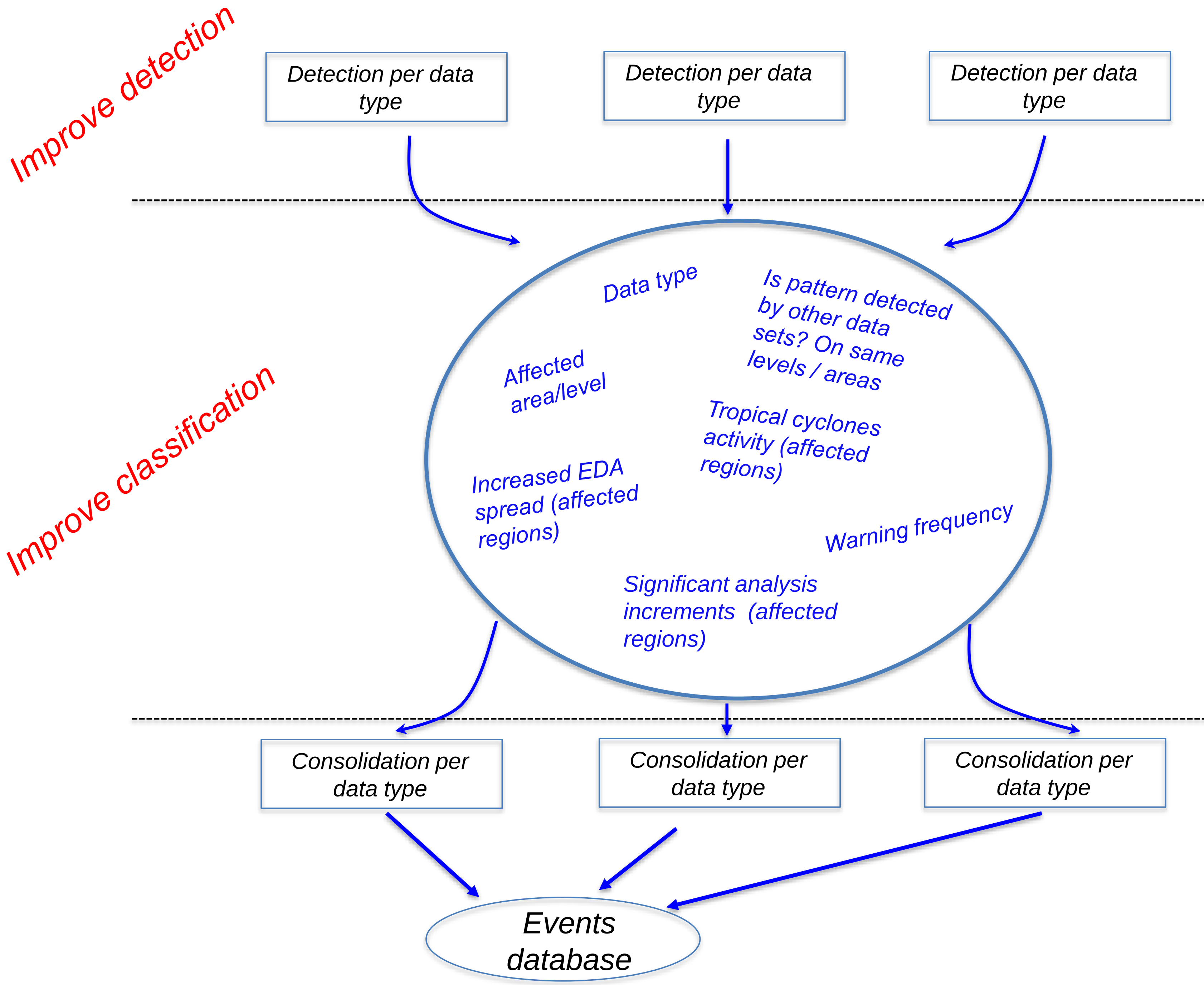
Earth system automatic warnings (ML approach)



- 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 (Statistical severity of events does not necessarily reflect an impactful situation)

Earth system automatic warnings (ML approach)

- Unsupervised machine learning to improve anomaly detection
- Supervised machine learning to improve classification and severity assignment.



TensorFlow (LSTM autoencoder)

- Data de-seasonalised and normalised
- Models trained 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)

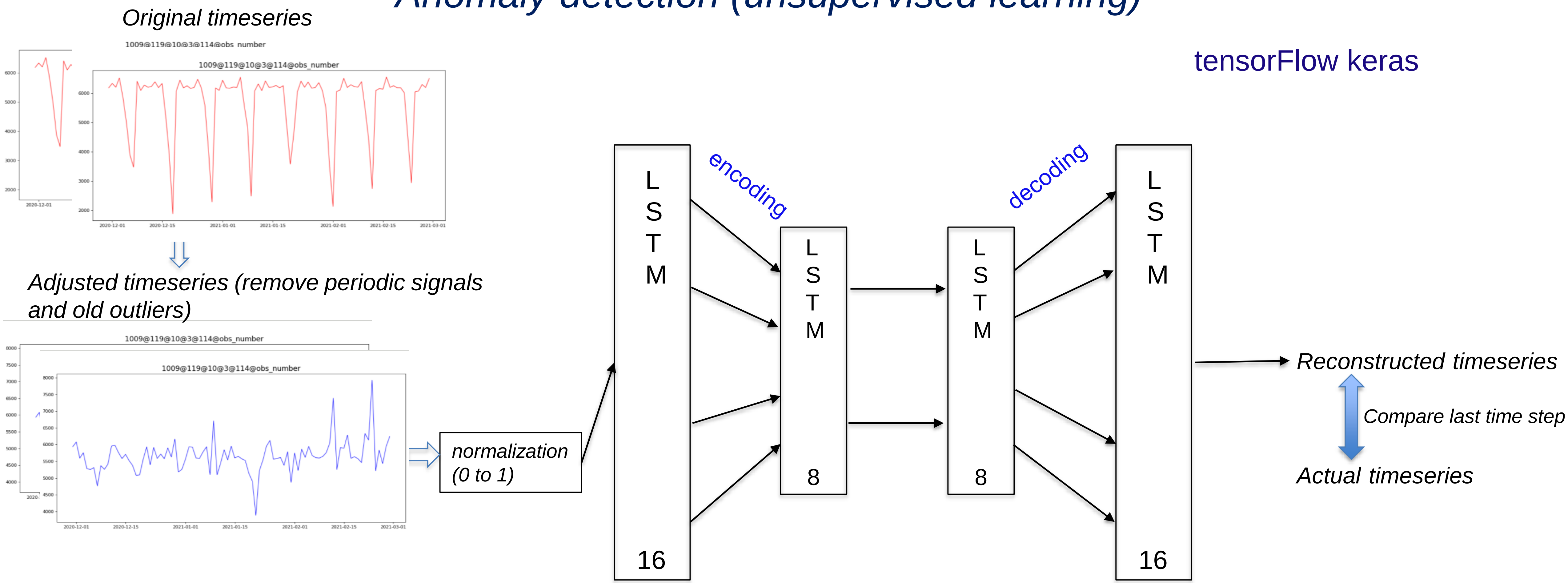
- During the labelling process, decisions are made based on the data types, area of interest, etc.

- 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

Anomaly detection (unsupervised learning)



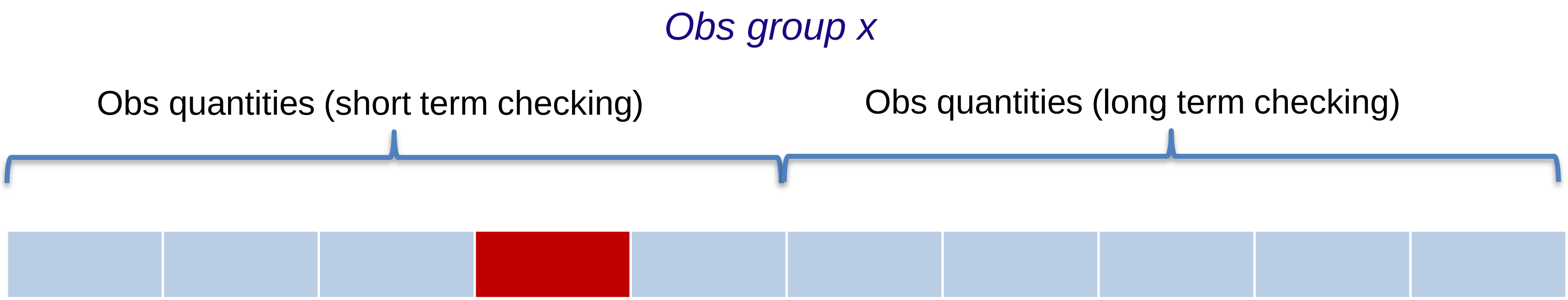
- Training performed every cycle. Number of epochs 1000 with early stopping option activated
- Thresholds are estimated from the loss distribution (from training dataset)
- The trained model is applied to the last time step of timeseries. Reconstructed values are compared to actual ones. Data are flagged in case of threshold exceedance
- A Severity level is assigned based on how far the stats from thresholds (again based on loss distribution)

Anomaly detection (unsupervised learning)

- Detection applied to short time series (~2 months):
  - Periodicity adjustment applied to data counts only
  - Flags sudden changes of statistics

- For satellite data detection applied to long time series (~12 months when available):
  - Flags drift of statistics
  - Periodicity adjustment applied to data bias correction and stdev(O-B)
  - When an anomaly is detected an additional test (Theil-Sen estimates) to confirm trends and exclude cases with changing slopes

- All flagged cases are recorded (bitfields) in the data files and excluded from subsequent training (positive feedback)

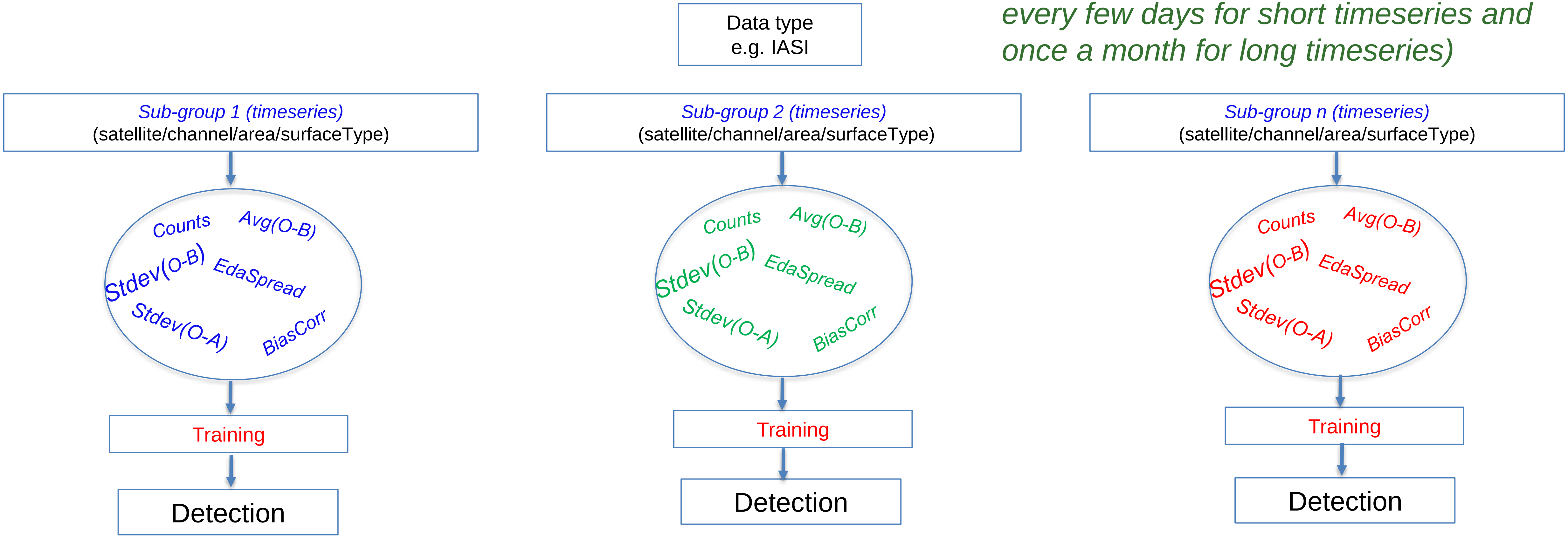


Anomaly detection (unsupervised learning)

- All relevant events detected during the testing period

- The very large amounts of data (~ 7000 data groups just for IASI) require an efficient data structure to overcome processing time. Training is fast (~1/2 seconds) but collectively very time consuming if the training is performed for each data group independently.

Training can be done periodically (e.g. every few days for short timeseries and once a month for long timeseries)



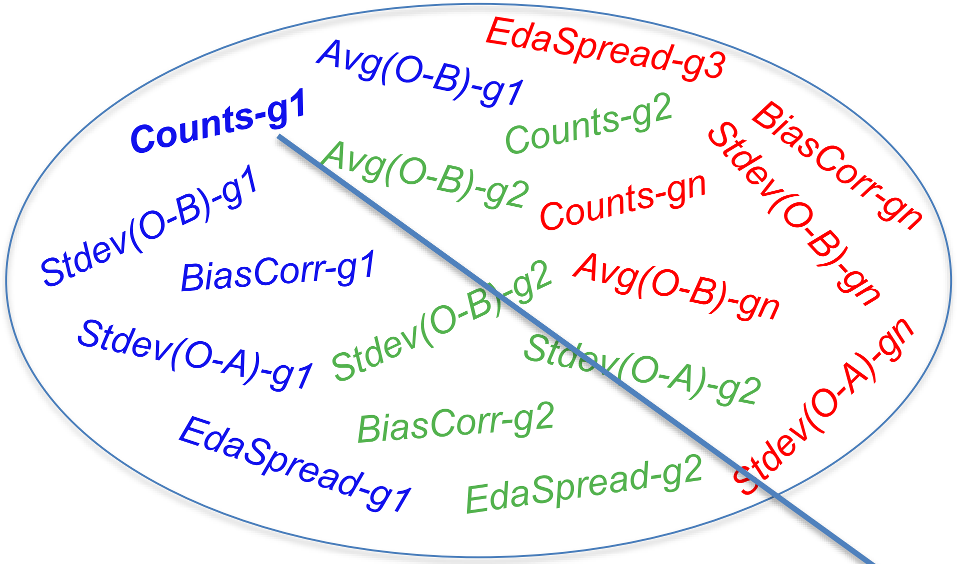
Data type
e.g. IASI

Sub-group 1
(satellite/channel/area/surfaceType)

Sub-group 2
(satellite/channel/area/surfaceType)

Sub-group n
(satellite/channel/area/surfaceType)

Combined features (distributed by processor)

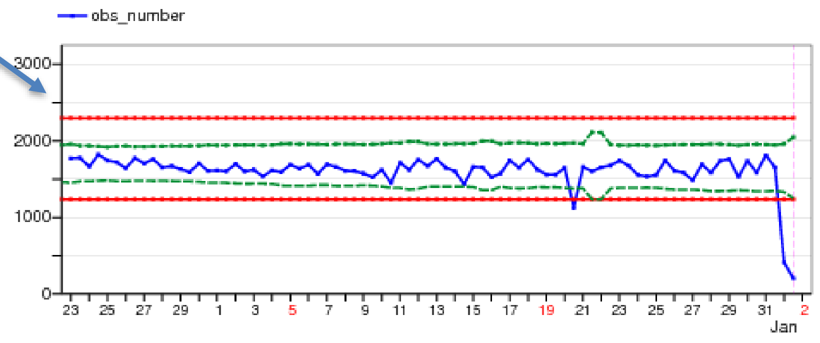


Training best done every cycle to avoid issues with inconsistent bucket content (very dynamic observing system)

- Reasonably fast
- Allow detection of anomalies affecting individual groups: no compromise on details
- Combining datasets require a good treatment of missing timeslots

Training

Detection



Improve classification of warnings (supervised learning)

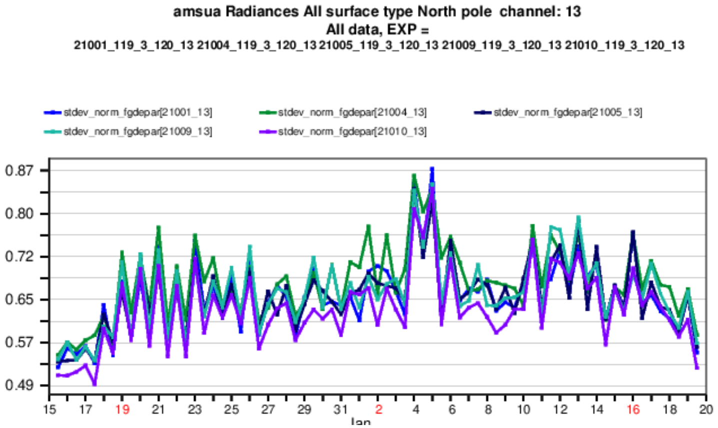
- Improve the system to distinguish data and DA/model issues.

- Improved severity assignment of warnings (Severe, considerable, slight, false alarm)

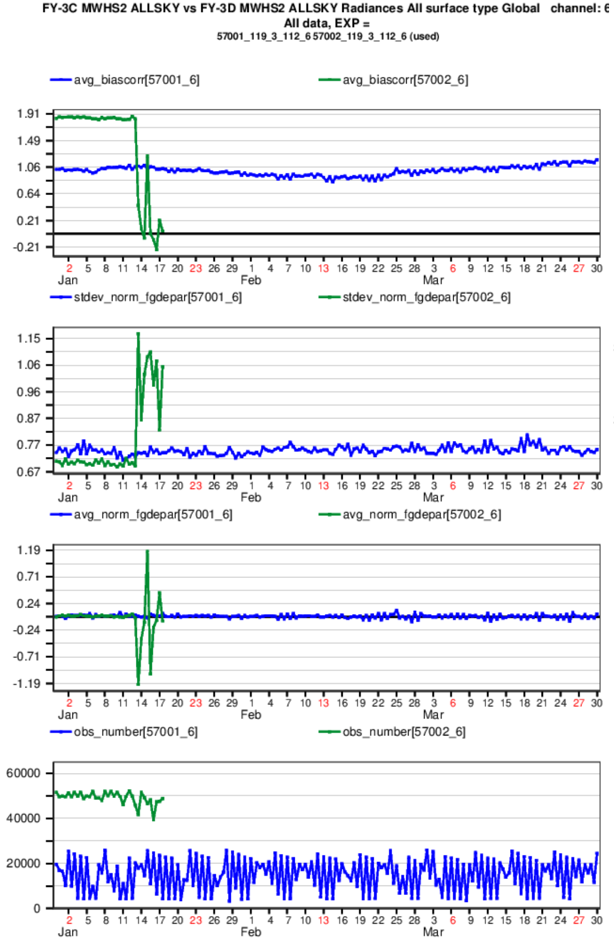
- Suggest if an action is needed

MWHS2 Ch 6 Global

AMSUA Ch 13 N-Pole

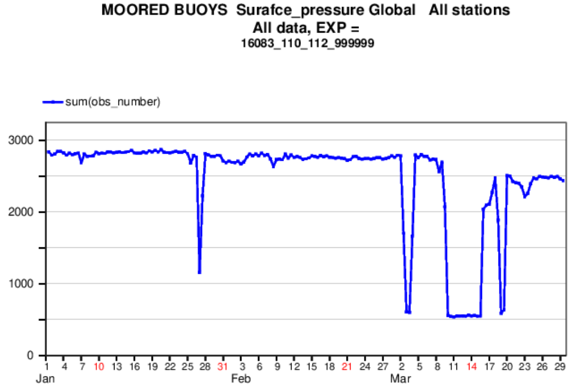


- Warnings affected many satellites (not a data issue)
- Severity: Severe
- Cause: Atmospheric variability
- No need for action

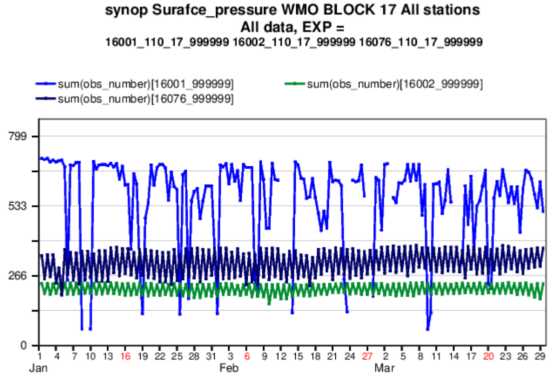


- Warnings affected only one satellite Severity: Severe
- Cause: Data issue
- Action needed (stop using the data)

Moored buoys



- Moored buoys reduced (important dataset)
- Severity: Severe
- Cause: Data outage
- Action needed (contact data provider)



- Automatic SYNOP data intermittent over Turkey (Other datasets present)
- Severity: Severe
- Cause: Data outage
- Action needed (contact data provider)

All preliminary warnings

- Data category (satellite, conv, ocean, ostia)
- location or area
- Type of event (missing, out of range)
- Observation quantity
- Frequency of similar incidents
- Severity of deviation
- Other similar datasets present in the area and how many affected ?
- EDA spread increased ?
- Severe weather events in the area
- All levels/channels affected ?
- How many areas affected ?
- parameters affected ?
- Data counts
- Usage status (satellite data)

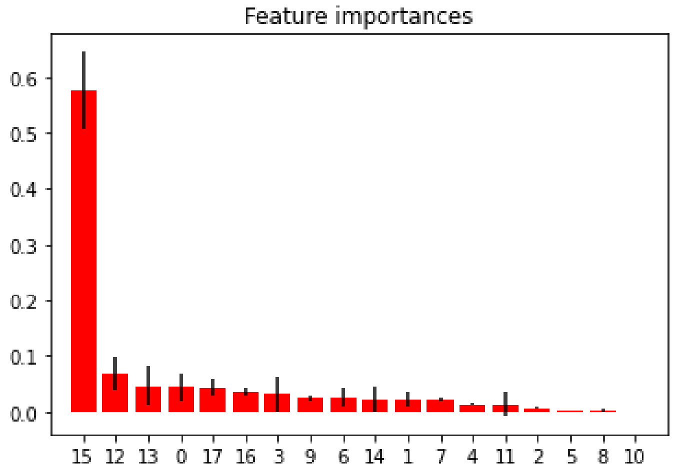
Random forest classifier

- False alarm (yes/no)
- Slight (yes/no)
- Considerable (yes/no)
- Severe (yes/no)
- Cause (data/other)
- Action required (data/other)

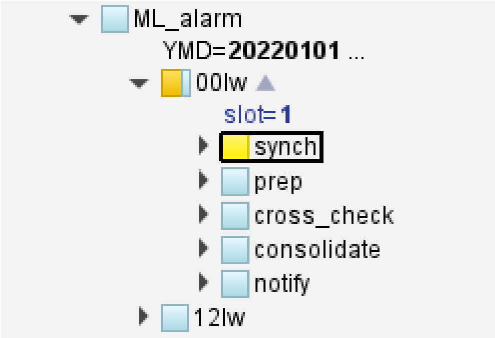
– Training dataset populated from *labelled* warnings (from the event database). Dataset is balanced before training
– The labelling process allows classification based on the context of the warnings
– The labelling process was largely semi-automated but will be further improved.
– Continuous labelling is an option (cases to be added occasionally when relevant)

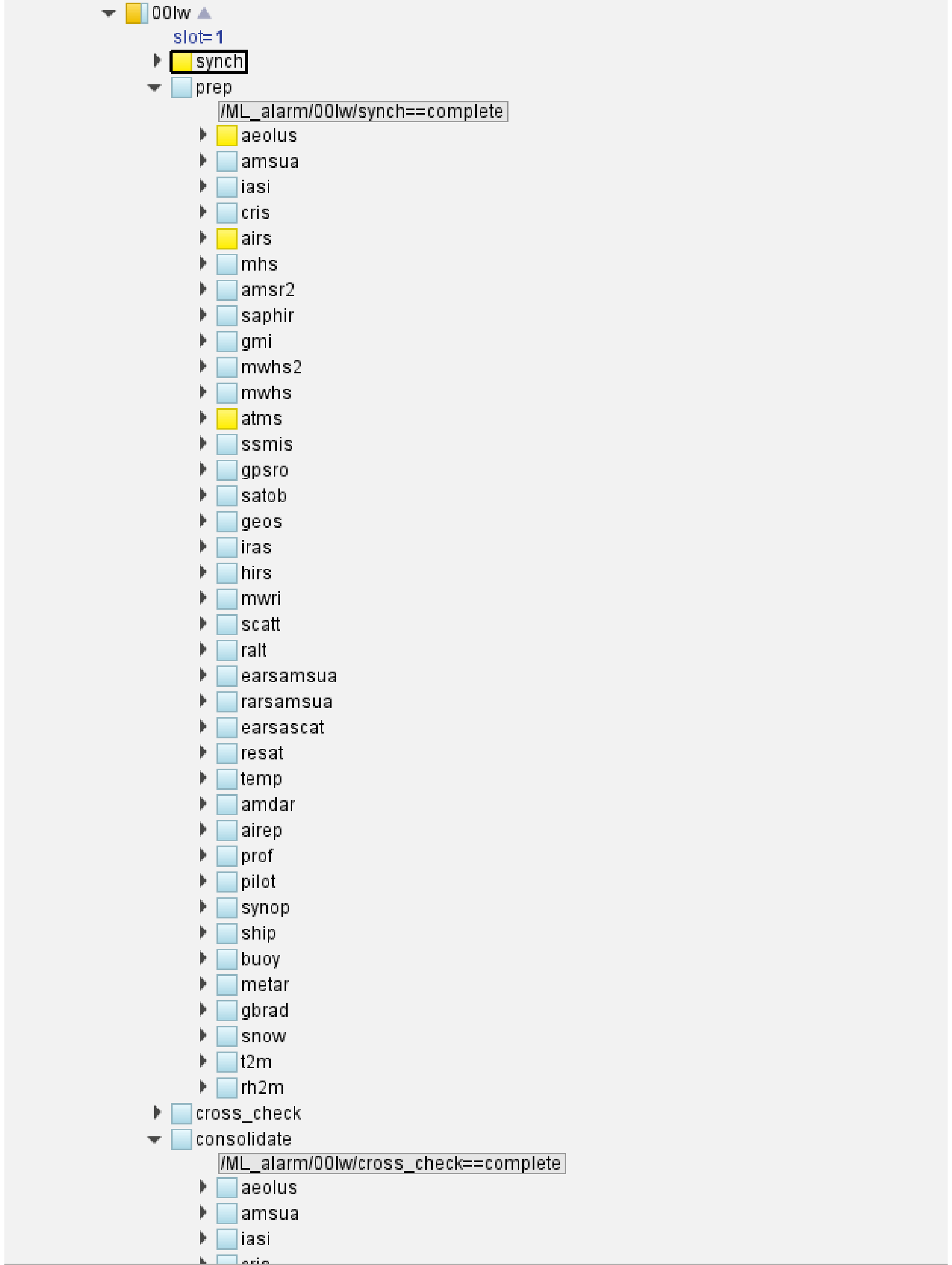
Feature ranking (target feature: cause)

```
Feature ranking:
1. feature 15 nbr_weather_events (0.576049)
2. feature 12 nbr_similar_reportypes_walarms (0.068808)
3. feature 13 NBR_similar_wmoblock (0.045815)
4. feature 0 Reportype (0.044922)
5. feature 17 ratio2 (0.042841)
6. feature 16 ratio1 (0.036956)
7. feature 3 Area (0.030702)
8. feature 9 NbrPast (0.025465)
9. feature 6 varno (0.024970)
10. feature 14 Nbr_similar_near (0.023117)
11. feature 1 Vertco1 (0.021050)
12. feature 7 AlarmGroup (0.020834)
13. feature 4 Obs_quantity (0.014144)
14. feature 11 Nbr_areas (0.013520)
15. feature 2 vertco_type (0.006806)
16. feature 5 Type_event (0.002447)
17. feature 8 usage (0.001555)
18. feature 10 eda (0.000000)
```



System implementation





Conclusions

- Automatic data checking is very important for an operational data assimilation system
- ML is being tested to improve detection and classification of warnings. Promising behaviour
- Further optimisation of the implementation is expected to bring more benefits (features detection, rare events detection and improved diagnostics)