Automatic checking of observations at ECMWF

Mohamed Dahoui and Niels Bormann

ECMWF, Shinfield Park, Reading, UK

Introduction

Observations are essential for numerical weather prediction (NWP) systems. They are used by the data assimilation system to produce the best estimate of the initial conditions. The quality of the analysis and the subsequent forecasts depend, amongst other factors, on the quality and availability of observations. A reliable monitoring system is therefore needed for early detection of observational data issues that can potentially degrade the quality of the analysis. The steady increase in data volumes makes it difficult to rely solely on manual checking procedures.

Many years ago, ECMWF started to use an automatic data checking system to trigger warnings if there are sudden changes in the quality or availability of the data actively assimilated by the ECMWF data assimilation system (4DVAR). The checking system was initially limited to satellite observations, but it was later extended to cover in-situ measurements. More recently, the system has been complemented with an automatic checking of persistent quality improvements of in-situ data that are blacklisted ((not used because of previously poor or unknown data quality) in the IFS allowing a quick re-activation of observations. A test version of the system is being used to detect sudden significant changes affecting the SST and sea-ice fields from OSTIA.
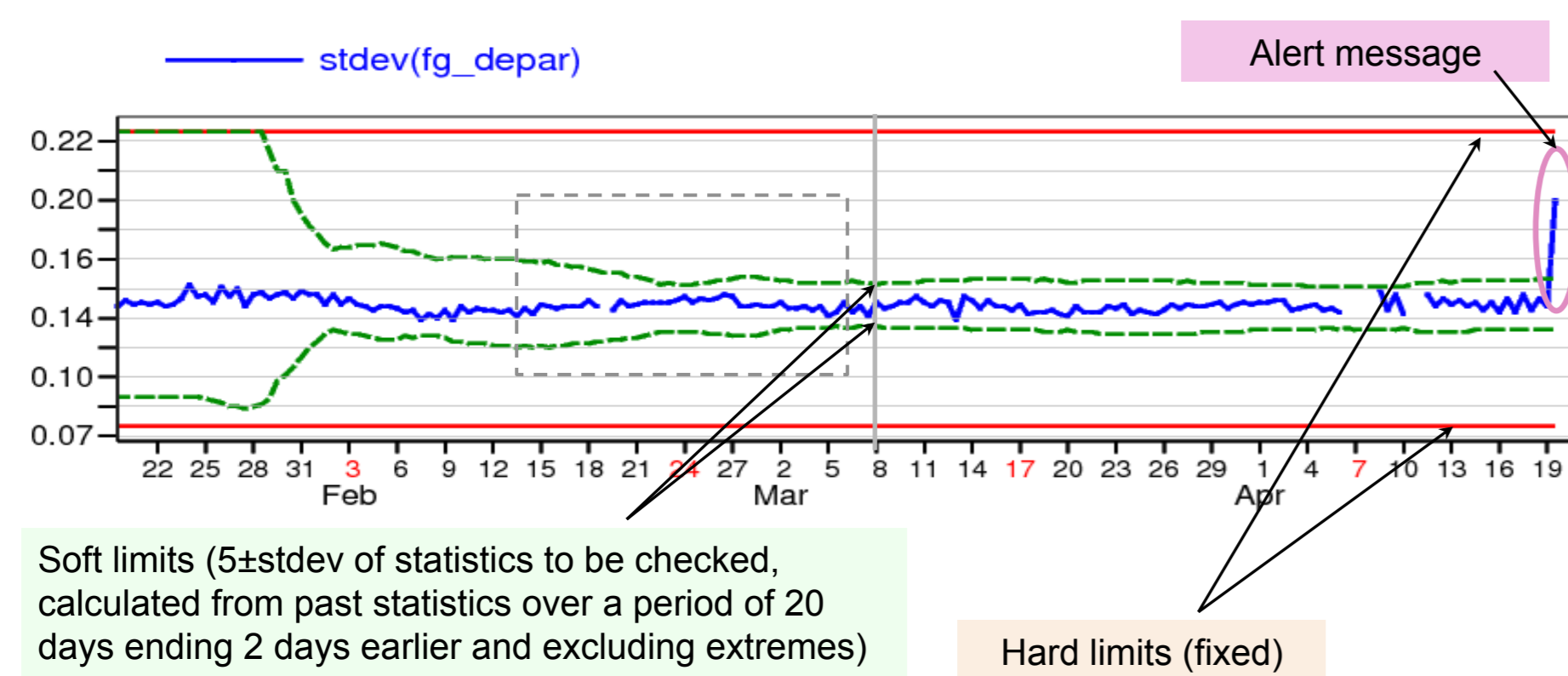
Soft limits (5±stdev of statistics to be checked, calculated from past statistics over a period of 20 days ending 2 days earlier and excluding extremes)

Hard limits (fixed)

Figure 2. Preparation of checking limits

Application for satellite data

For satellite data, we check separately global statistics from each satellite, instrument, channel, parameter and statistical quantities. The system offers the flexibility to add new statistical quantities. The current setup is summarized in Table 1.

Table 1. Current setup for the automatic checking of satellite data

Data types	Statistical quantities
Radiances	• Data counts • Average of innovations • Standard deviation of innovations • Bias correction
All sky microwave radiances	• Data counts • Average of normalized innovations • Standard deviation of normalized innovations • Bias correction
AMVs	• Data counts • Average of innovations • Standard deviation of innovations • Average pressure (pressure of the assigned height) • Standard deviation of pressure
GPS radio occultation	• Data counts • Average of normalized innovations • Standard deviation of normalized innovations
Scatterometer (wind speed)	• Data counts • Average of innovations for the best ambiguity • Standard deviation of innovations for the best ambiguity
Ozone	• Data counts • Average of innovations • Standard deviation of innovations • Bias correction

Application for conventional observations

Unlike satellite observations, in-situ measurement issues are typically specific to each individual station. However, on some occasions widespread issues might be caused by data routing problems, significant weather events, model errors or data assimilation issues. To cover both aspects, the automatic data checking system has been extended to perform two kinds of automatic check.

Individual stations: For each assimilation cycle, the system checks the quality of available stations based on the 4DVAR estimated Probability of Gross Error (PGE), the mean and root-mean-square of innovations and bias correction. The automatic system calculates the percentage of in-situ reports with PGE above a pre-defined threshold (currently 0.75). It then triggers a warning if this percentage is high or if there are significant changes in the other statistical quantities.

Main data types over a number of pre-defined geographical areas and WMO blocks: This component of the system detects widespread issues. It follows the same method as applied for satellite data but with an additional test comparing the standard deviation of innovations and analysis departures.

Application to SST and sea-ice from OSTIA

ECMWF SST and Sea-ice analysis is mainly driven by the OSTIA fields from the UK Met Office. Despite its importance for the different components of forecasting system, limited quality control checks are currently applied to OSTIA SST and Sea-ice fields. An automatic checking is being tested to allow early detection of sudden and significant changes over pre-selected geographical areas. The system monitors the Root Mean Square of day to day differences (computed using the differences of values for each grid point within the area) and the day to day differences of the area mean. With these two measures, the system will detect unrealistic day to day variations.

Conclusion

The automatic data checking is becoming a key component of data monitoring activities at ECMWF. It reached a level of maturity allowing the detection of almost all severe data quality issues. Work is on-going to further improve the reliability of the system and explore its potential for improving the daily monitoring of the operational data assimilation and forecasting system at ECMWF. The future technical reorganization of the data pre-processing and assimilation (continuous observation processing environment) will allow **early** automatic detection of quality and availability issues well before the data assimilation takes place.

Approach

For each set of observations, selected statistical quantities (e.g. number of observations, bias correction, mean bias- corrected innovations (i.e. observation minus background) and analysis departures) are checked, for each assimilation cycle, against an expected range. An appropriate alert message is generated if statistics are outside the specified limits. A severity level ('slightly', 'considerably', 'severely' or 'severely persistent') is then assigned to each message depending on how far statistics are from the expected values. The automatic checking uses two kinds of ranges: soft and hard limits (see Figure 2).
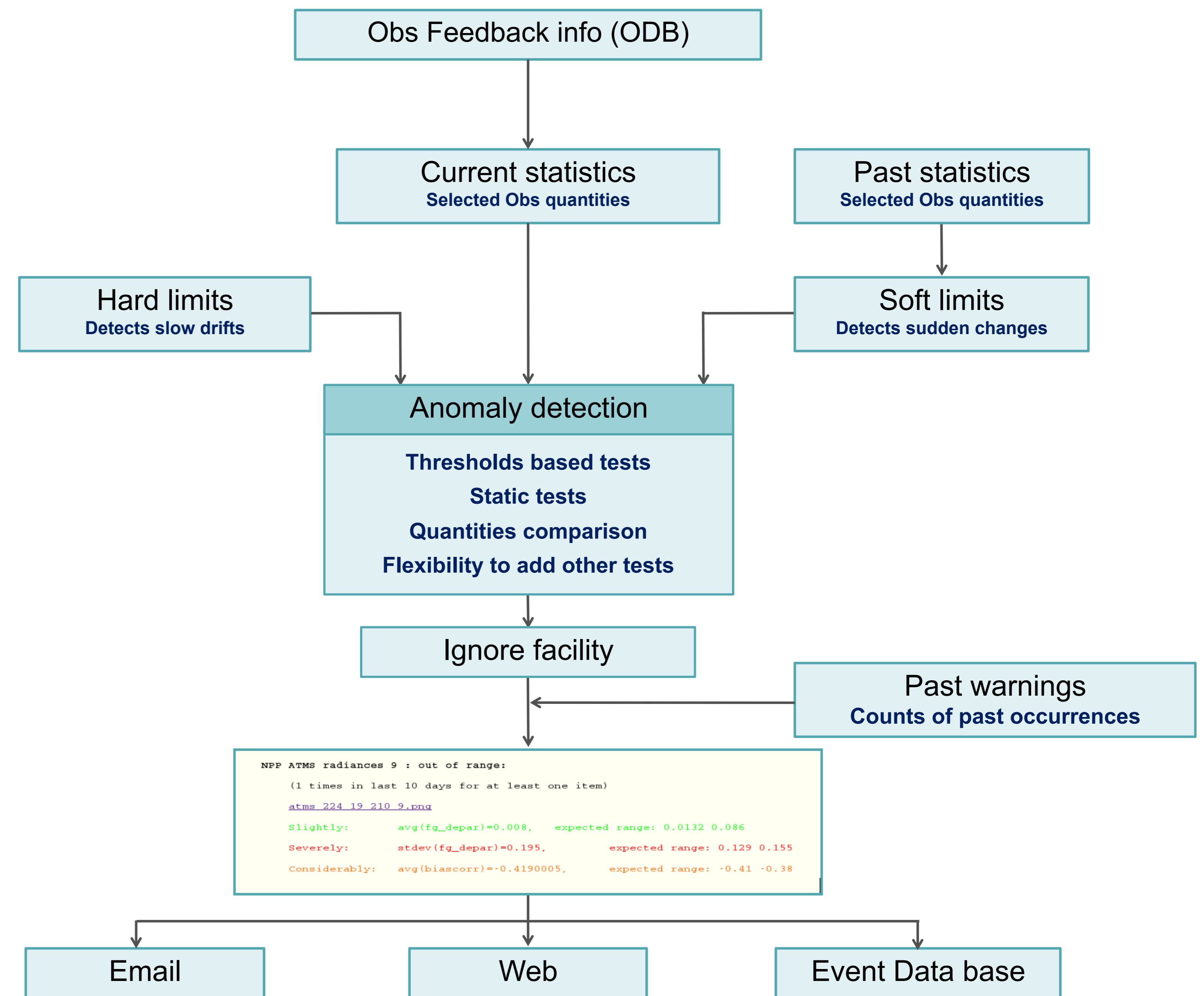
Figure 1. Automatic data checking working flow diagram

Lunar intrusion event

On 19 April 2013, the automatic checking system triggered severe warnings for several NPP/ATMS channels . The alerts were caused by a sudden sharp increase in the innovations standard deviation (see the example in Figures 2 and 3 for ATMS channel 9 first guess departures covering 09 to 21 UTC on 19 April 2013) due to the moon intrusion into the ATMS field of view used for the calibration of the instrument.
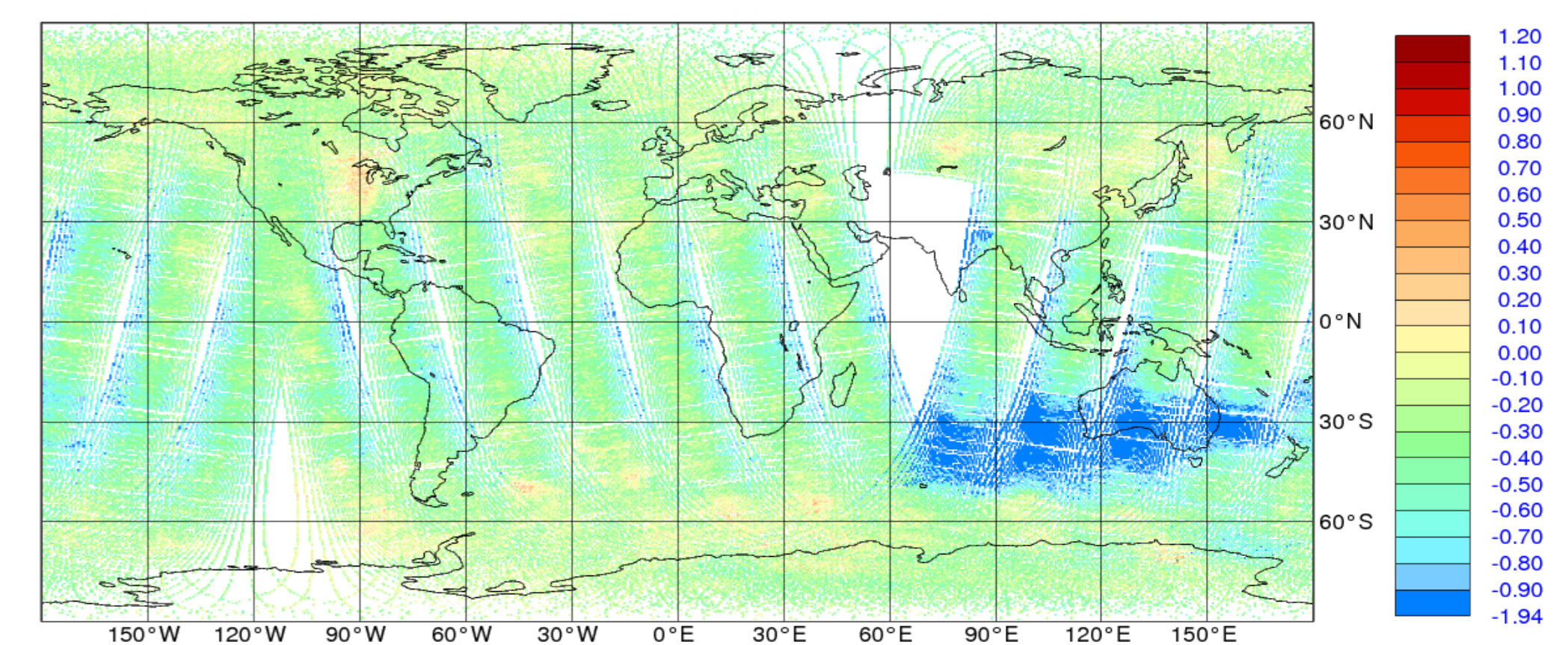
Figure 3. First guess departures for brightness temperatures (in K) from ATMS Channel 9 on 19/04/2013 (09 to 21 UTC)

Potential for model performance monitoring

The automatic data checking system was primarily designed to detect data-related issues. However, since the quality assessment is mainly based on innovations, there is a potential to detect situations where the forecast model or the data assimilation itself has weaknesses. This can be the case in the following situations.

- The system triggers alerts for independent satellites/ instruments that provide similar observations of the atmosphere. Large departures associated with independent data suggest that the forecast model is not able to fully capture the phenomena being observed. As an example, on many occasions the system triggered alerts associated with microwave and infrared satellite upper-stratospheric channels (see Figure 4). These alerts were caused by sudden stratospheric warming episodes not well captured by the model

- Widespread warnings affecting in-situ measurements (apart from availability issues) are likely to be related to the forecast model or the data assimilation

- Individual warnings occurring in the vicinity of dynamic atmospheric situations need to be investigated. They can potentially point to model or data assimilation issues (e.g. a weather system moving too slowly in the model).
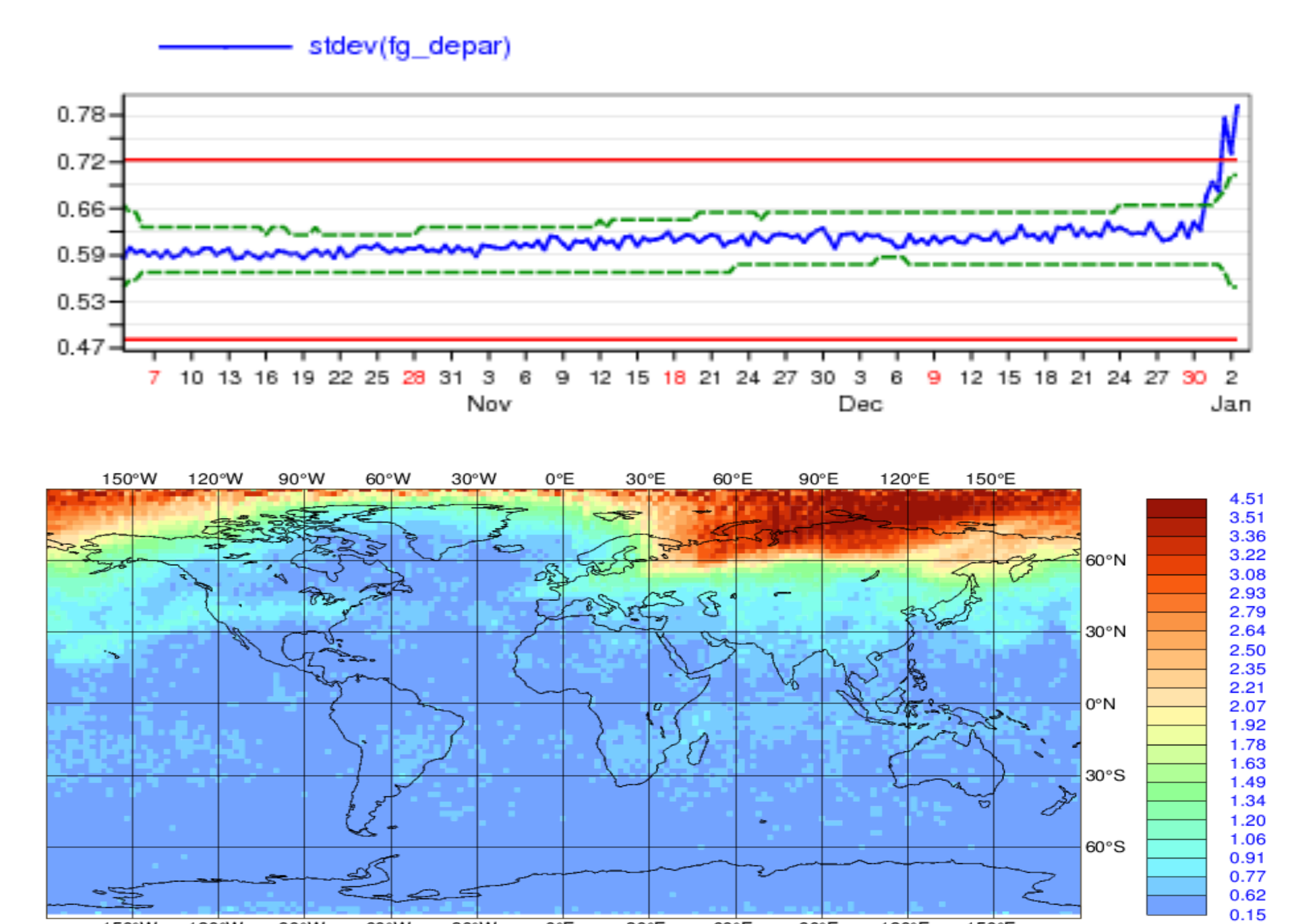
Figure 4. Standard deviation of first guess departure for brightness temperatures (in K) from METOP-A/AMSUA Channel 13 (27/12/2012 to 06/01/2013). Top panel shows long time series.

Example of a warning triggered by abnormal OSTIA sea-ice field over the North Hemisphere on 01/07/2015. Both monitored quantities were affected by a sudden strong change (see Figure 5) caused mainly by unrealistic sea-ice over Hudson and Baffin bays (see Figure 6). The ECMWF analysis rejected the corrupted fields and persisted the sea-ice from the previous day.
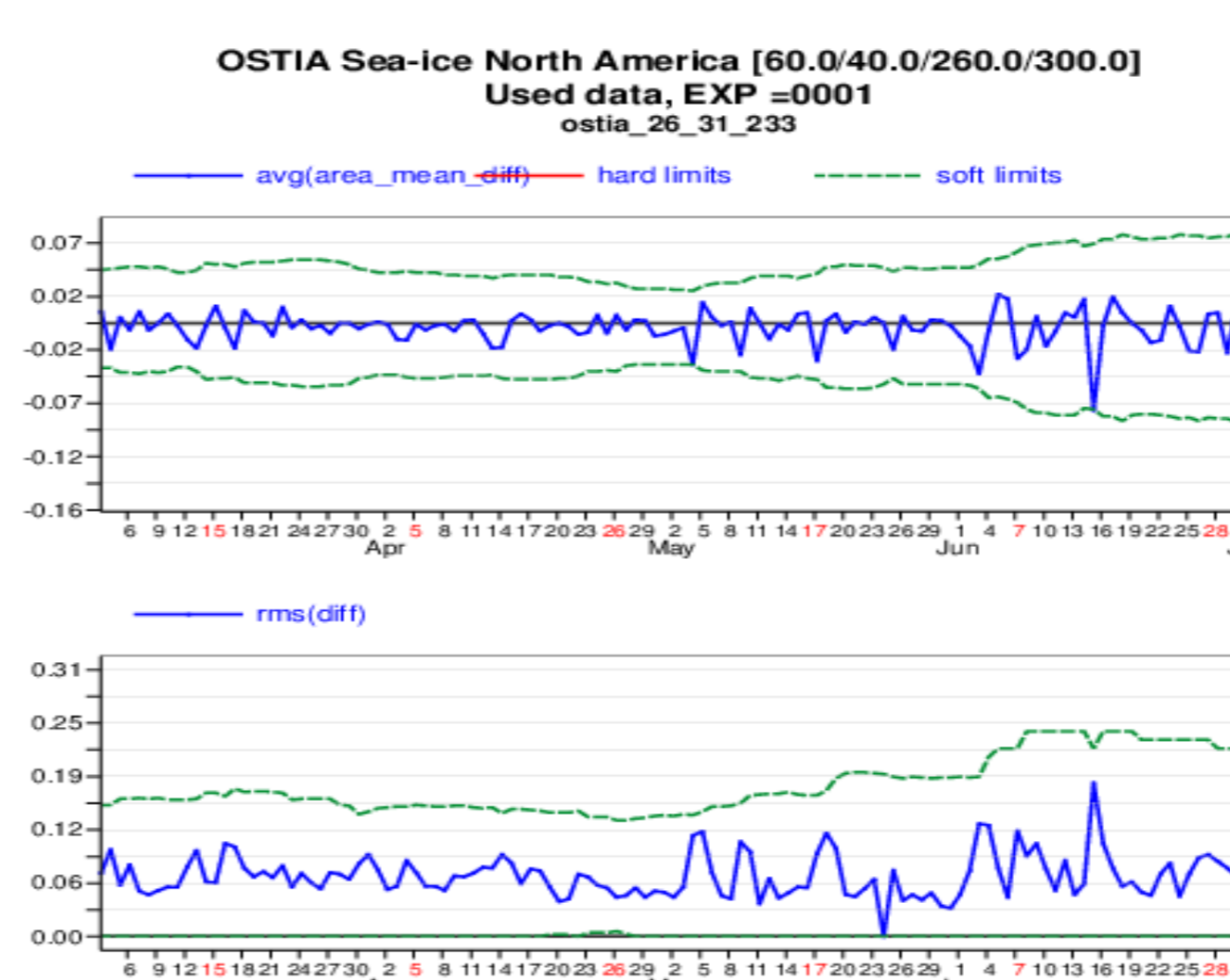
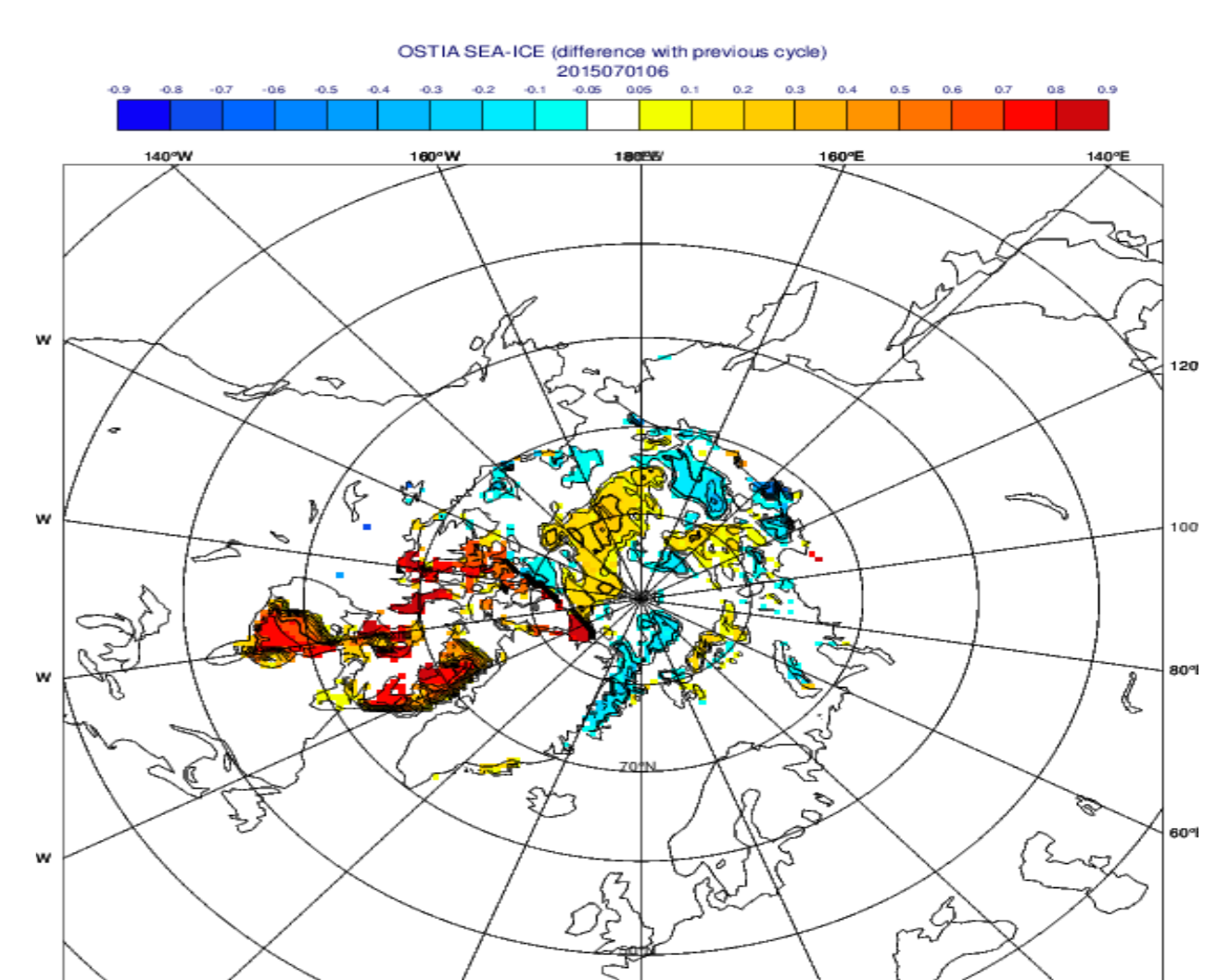**Figure 5.** Time series of day to day variability of sea-ice mean and RMS over North America.

Figure 6. Difference of sea-ice between 01/07/2015 and 30/6/2015