Reanalysis Blacklists at ECMWF

Paul Poli

General concepts

- For a reanalysis produced in 2 stages:
  - "Historical part"
    - Only specify an 'exclusion' part
    - With time-dependent entries
  - "Near-real-time continuation"
    - 'exclusion' part changes with evolutions of the observing system (time-varying entries extended to allow re-run)
    - 'Monthly' part received from ECMWF Operations, reflecting evolutions in the in situ surface and upper-air observing networks

ERA-Interim blacklist

- Historical run uses information:
  - From prior reanalyses: FGGE, ERA-15, ERA-40, JRA-25!
  - From Operations for 2002-2008
  - From any other “accrued knowledge”
  - Rarely traceable and verifiable!
- NRT Continuation since 2009:
  - Differs substantially from ECMWF OPS
    - Usually less conservative…
    - More precise with exact dates of instrument turning ‘bad’…
    - …Benefit of hindsight!
  - Misses on all the new instruments in the observing system that could not be supported by ERA-Interim

Blacklist pseudo-code

- Another brainchild of the ODB inventor (Sami Saarinen)
- Turns human-readable pseudo-code into C
- C gets compiled into a library
- Library is dynamically linked
- 4DVAR can use the latest available blacklist without need to re-compile

Blacklist sample

```
elif SENSOR = iras then      ! IRAS data selection (FY-3A)

   if vert_co = TOVS_CHA and PRESS in (5,6,7,8,9,10,11,12,13,14,15)
      and (MODORO > 1500)                                    then fail(CONSTANT); endif;

   if (LSMASK = land)  then
    if vert_co = TOVS_CHA and PRESS notin (1,2,3,12)      then fail(CONSTANT); endif;
   endif;

   if vert_co = TOVS_CHA and PRESS in (16,17,18,19,20)
                                                          then fail(CONSTANT); endif;
   if vert_co = TOVS_CHA and PRESS notin (4,5,6,7,11,12,14,15)
                                                          then fail(EXPERIMENTAL); endif;

   if FOV in (1,2,3,54,55,56)                             then fail(CONSTANT); endif;

   if  SATELLITE_IDENTIFIER in (520) then ! FY-3A erroneous data at lat/lon = 0/0
     if( abs(LAT) < 0.1 and abs(LON) < 0.1 )              then fail(CONSTANT); endif;
   endif;
```

Blacklist decisions are based on the following attributes

- analysis date, analysis time, observation date, observation time, mean solar local time
- observation type, observation code type, observation report type, station identifier, sonde type
- generating centre, generating sub-centre, data stream, data source, collection identifier, retrieval source
- satellite identifier, satellite sensor type, satellite instrument identifier, retrieval type, product type,
- latitude, longitude, station altitude
- satellite zenith angle, field of view number, solar zenith angle, solar elevation, radar elevation
- geophysical variable, surface pressure reporting practice, type of vertical coordinate, vertical coordinate, reference level vertical coordinate
- observed value, (for radiosonde only:) observation temperature at same level
- Quality of retrieval, EUMETSAT Quality Indicator with forecast dependence, EUMETSAT Quality Indicator without forecast dependence, CIMSS Quality Indicator Recursive Filter Flag
- (for radiances) cloud cover, cloud top pressure, cloud high/middle/low amount, % of clear pixel
- first guess departure, observation error, first-guess error, window or alternative channel departure, model land-sea mask, model orography, model sea-ice fraction, model surface pressure, model surface type indicator, model surface temperature, model 2 metre temperature, model top pressure,
- surface transmittance (for AIRS)

More than 60 parameters! We could rationalize some of that into: obs. identification, obs. metadata (position, time, orientation, geometry, quality flags…)**, and model (feedback)-added value**

“Major instrument records”

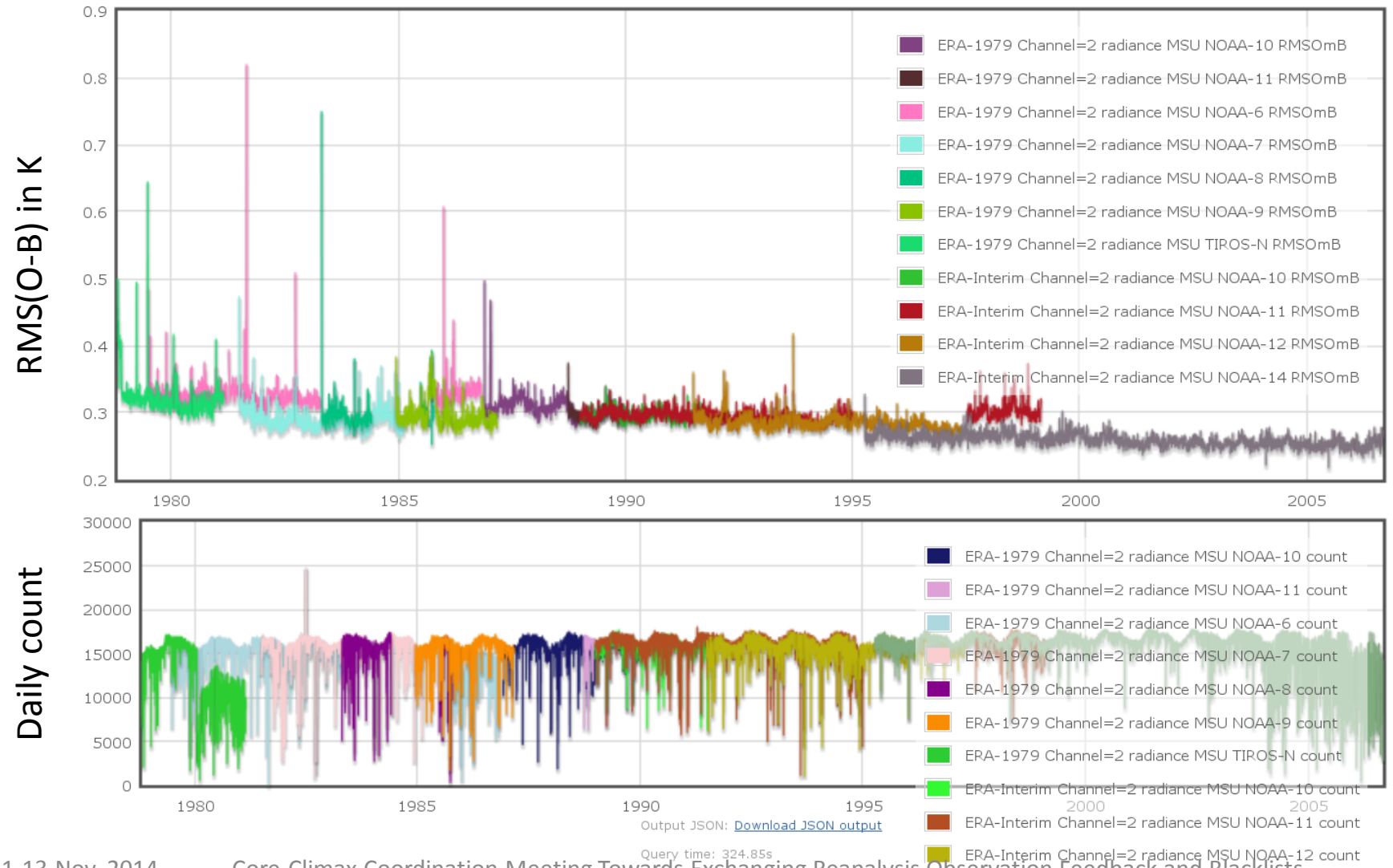
- SSU
- HIRS
- MSU
- AMSU-A
- AMSU-B & MHS
- SSM/I

MSU

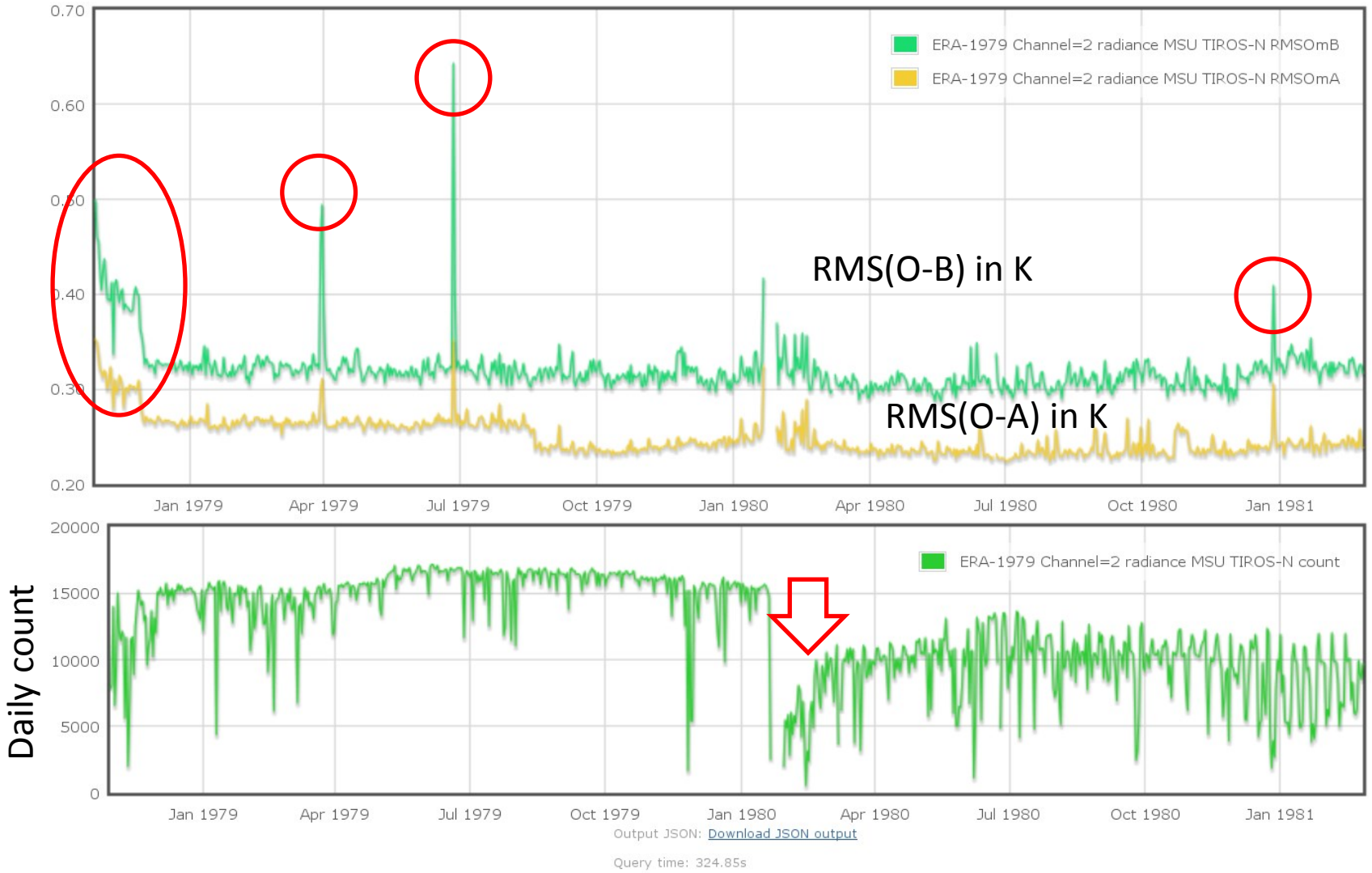
General:

- Reject first and last FOVs (1,11)
- Reject channel 1
- Reject channel 2 if model orography>1500m
- Reject channel 2…
  - … if |O-B| for channel 1 > 3K over land
  - … or if |O-B| for channel 1 > 5K over other areas

MSU, All satellites, Channel 2



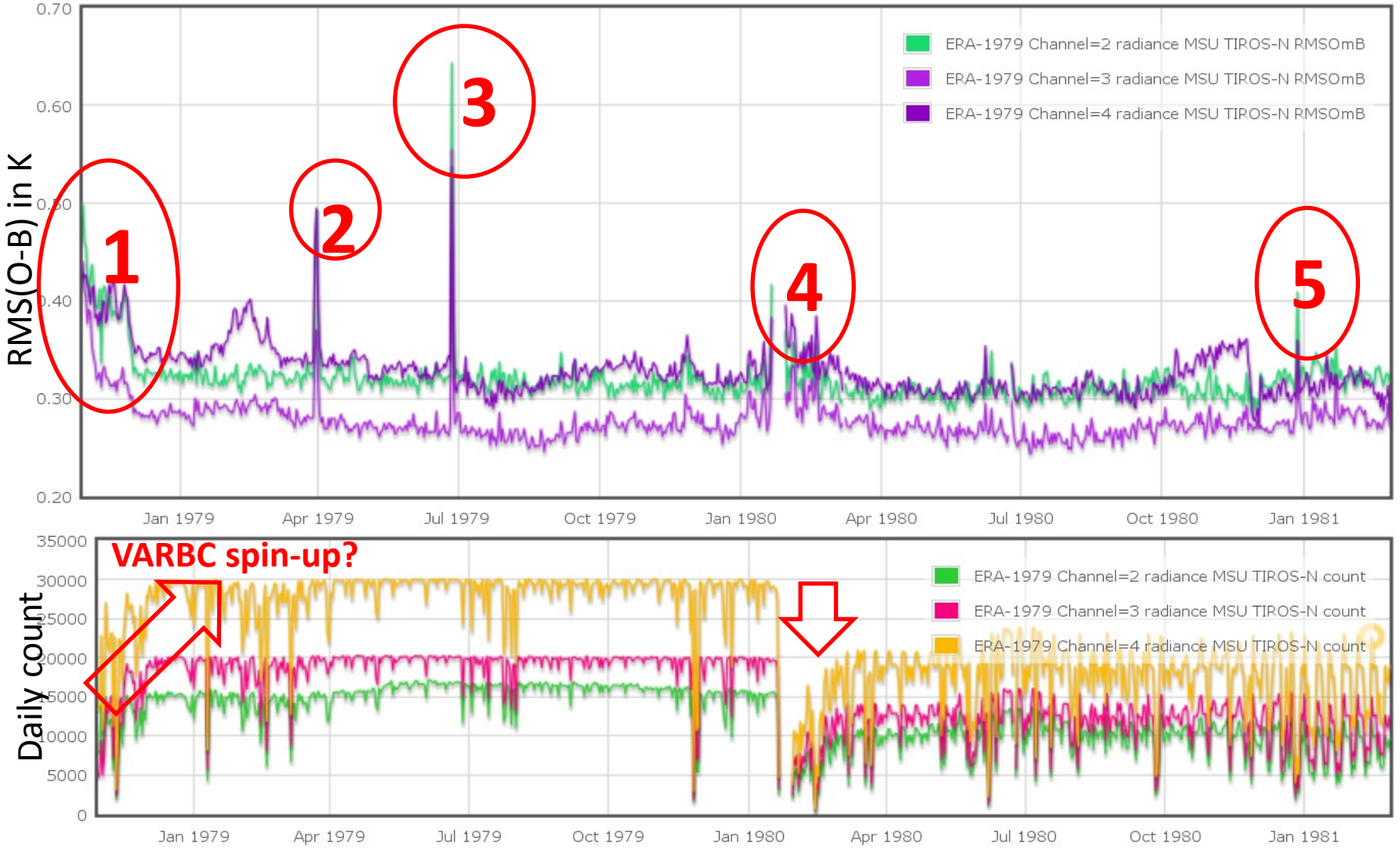
MSU, TIROS-N, Channel 2



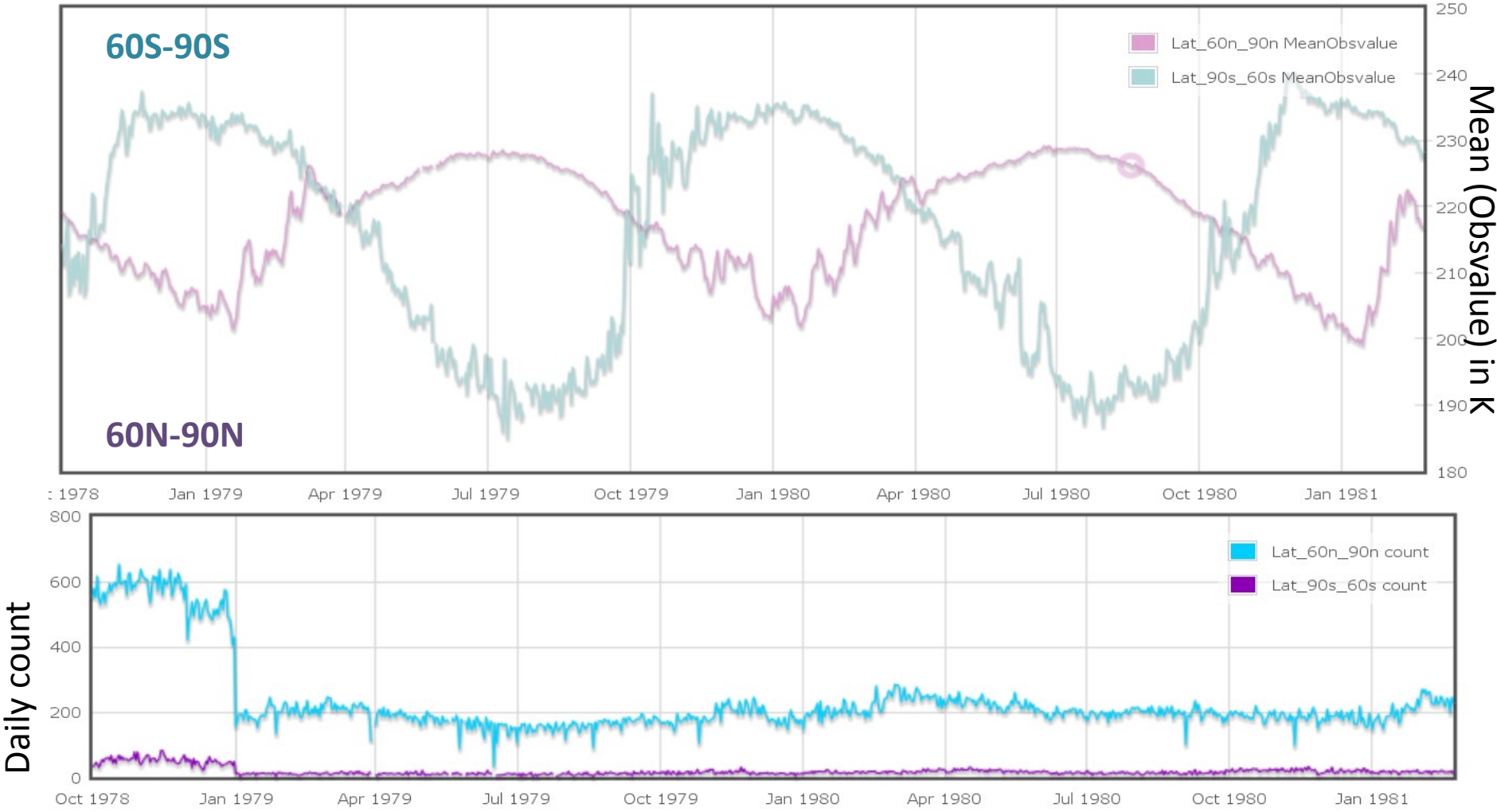
MSU

- TIROS-N
  - Expected usage dates 19781201 – 19810227
  - MSU changes behaviour on 19790626, should probably reject after that, but ignored that in ERA-Interim (and proved right?)
  - TIROS-N bad Earth-location (all instruments)
    - (DATE = 19790713 and TIME >= 150000 and TIME < 210000)
    - (DATE = 19790724 and TIME >= 090000 and TIME < 210000)
    - (DATE = 19790729 and TIME >= 090000 and TIME < 150000)
    - (DATE = 19790801 and TIME >= 090000 and TIME < 150000)
    - (DATE = 19800306 and TIME >= 210000)
    - (DATE = 19800307)
    - (DATE = 19800308 and TIME < 210000)
    - (DATE = 19800822 and TIME >= 210000)
    - (DATE = 19800823 and TIME < 090000)
    - (DATE = 19800924 and TIME >= 090000 and TIME < 210000)
    - (DATE = 19800924 and TIME >= 210000)
    - (DATE = 19800925 and TIME < 030000)

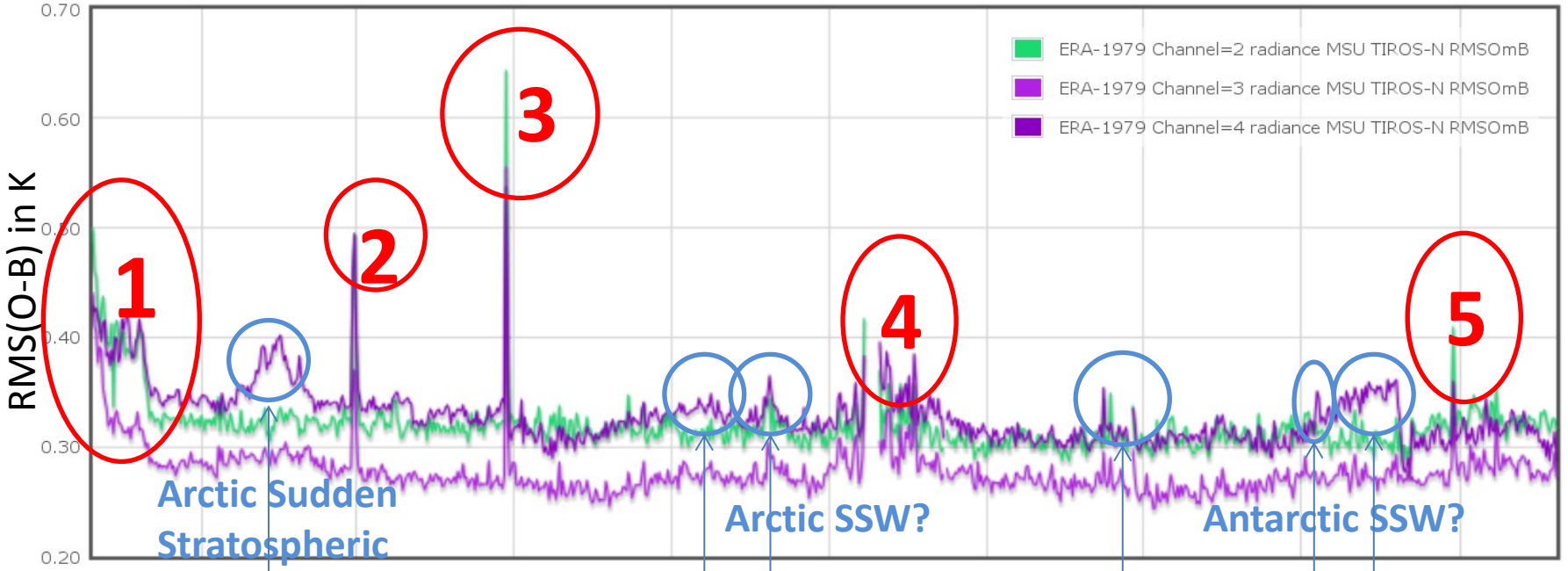
MSU, TIROS-N, all channels

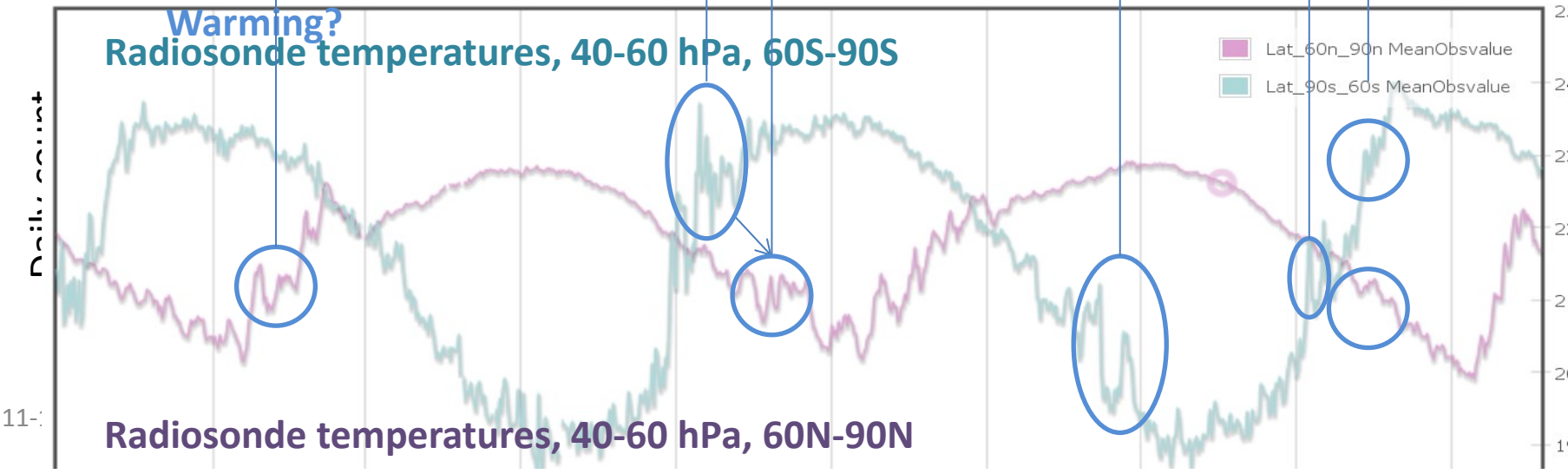


Land radiosonde temperatures, 40-60hPa



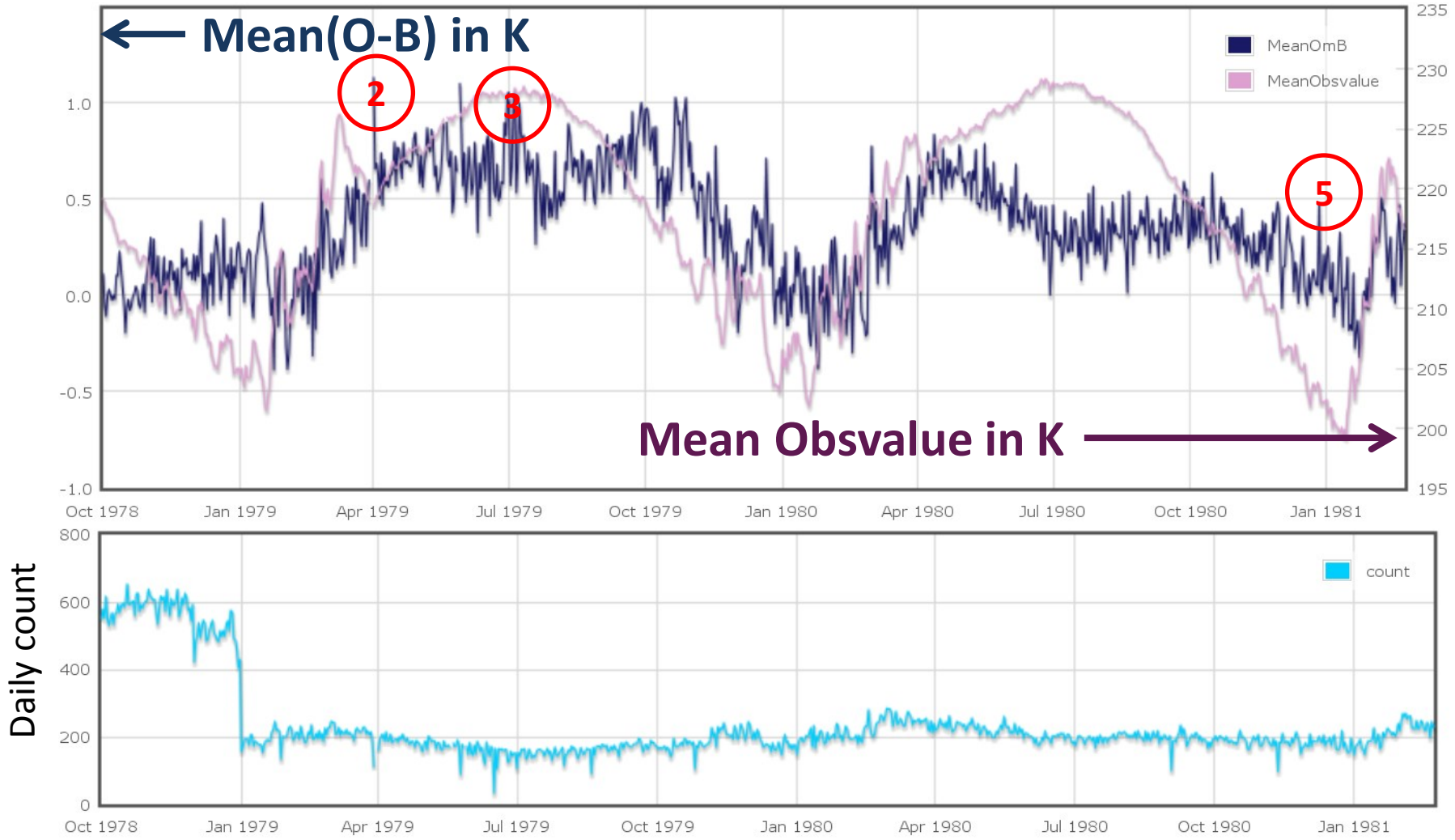
MSU, TIROS-N, all channels





RS temperatures Land 60-90N

Spikes in RMS(O-B) in MSU channel 4 cause spike in RMS(O-B) (possibly next cycle) for RS?



Even for this simple time-series

- Takes time to look through and make a consolidated (best-practice?) suggestion
- Not straightforward to conclude about 'bad' data based solely on the RMS O-B of a given data type:
  - 'Problem' could be in real events missed by the model
    - Can be assessed by looking at other observations
  - 'Problem' in the observation data passes on to the background, and then to other observation data departures

SSM/I

General:

- Blacklist all channels over land
- Avoid inland seas and lakes:
  - Great Lakes: Lat 40-50N, Lon 75W-95W
  - Caspian Sea and Aral Sea: Lat 35-50N, Lon 45E-65E
- Only CM-SAF reprocessed data up to Dec 2008, no other SSM/I data
  - Question whether to use F-13 GTS data from 2009

SSM/I

- F-8
  - Expected date range 19870708 – 19911219
  - Allow for one-month spin-up : passive until 19870808
  - Bad data on 19890107
  - 85 GHz channels defective (large departures, corroborated by FCDR PUM), so reject channels 6 & 7 after 19880401
- F-10
  - Expected date range 19910106 – 19971115
  - Allow for one-month spin-up: passive until 19910206
- F-11
  - Expected date range 19911231 – 20000101
  - Allow for one-month spin-up: passive until 19920131
- F-13
  - Expected date range 19950502 – 20090101
  - Allow for one-month spin-up: passive until 19950602
  - Potential bad data caused by solar panel manoeuvre: 20090517 - 20090527
  - 85H (channel 7) deteriorated after 20090401
- F-14
  - Expected date range 19970506 – 20080824
  - Allow for one-month spin-up : passive until 19970606
- F-15
  - Expected date range 20000227 – 20060801
  - Allow for one-month spin-up: passive until 20000327
  - (22 GHz (channel 3) suffers from RFI after 20060810)

Conclusions

- Difficult to put the blacklist information into simple text
- Prone to making mistakes, too
- However, it should be feasible to parse the pseudo-code through a converter to extract only the date-related information, for examle, and then automatically get a list of blacklisted dates

Blacklist questions

- How could we keep track of all the reasons why we blacklist data?
  - Keeping plots supporting the decision would be useful – e.g. a reprocessed dataset may come along later, that fixes the problem – we need to be able to compare with about the same metrics, to check that
- List of flat 'events' (Blacklist/de-blacklist) could also be considered
  - Could be easy to compare, if we agreed on the general principle and granularity
  - The main driver would then be the time dimension – everything else coming on top

MSU

- NOAA-6
  - Hibernation period 19830417- 19851101
  - Second life ends 19861101
  - Noisy, unknown cause, on 19790906 between 9UTC and 15UTC
  - NOAA-6 bad Earth-location (all instruments)
    - (DATE = 19800302 and TIME >= 030000)
    - (DATE = 19800303)
    - (DATE = 19800304 and TIME < 030000)
    - (DATE = 19801104 and TIME >= 030000)
    - (DATE = 19801105 and TIME < 030000)
    - (DATE = 19851231 and TIME >= 210000)
    - (DATE = 19860101)
    - (DATE = 19860102 and TIME < 210000)

MSU

- NOAA-7
  - End of life 19850218
  - Noisy, unknown cause:
    - (DATE = 19840721 and 150000>TIME >= 090000)
    - (DATE = 19850207 and TIME >= 210000)
    - (DATE = 19850208 and TIME < 090000)
  - NOAA-6 bad Earth-location (all instruments)
    - (DATE = 19840301 and TIME >= 210000)
    - (DATE = 19840302 and TIME < 090000)