



On EO-observations for training ML-based weather and climate models

EUMETSAT supporting data-driven ML weather forecasting

R. Tervo¹, J. Schulz¹, V. John¹, R. Huckle¹, A. Lattanzio¹, M. Doutriaux-Boucher¹, A. Heidinger², K. Knapp², J. Mathews²
¹EUMETSAT, Germany - roope.tervo@eumetsat.int, ²NOAA, US

Machine learning models are at best as good as their training data

Training data must be

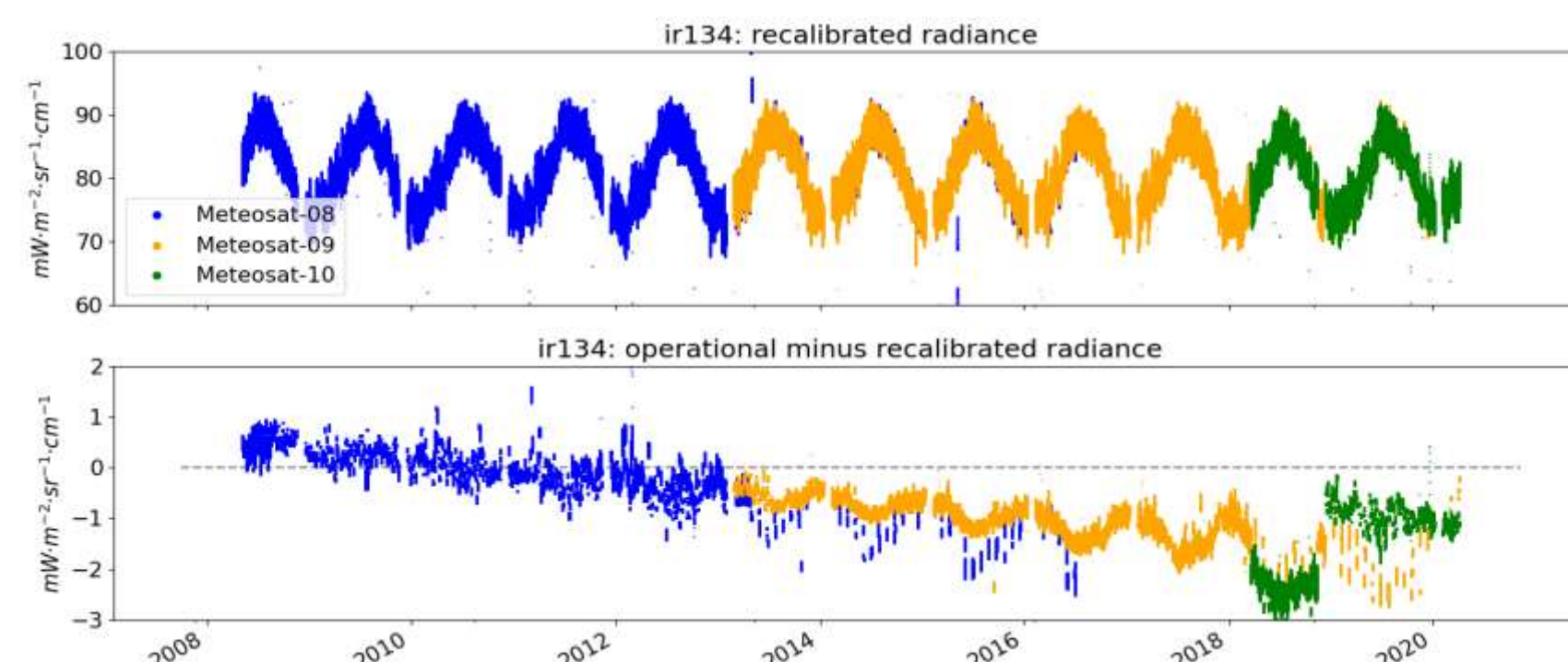
- 1) good quality
- 2) long-enough to contain sufficient natural variability
- 3) consistent and respective to the data used in inference phase

CLIMATE DATA RECORDS of LONG OBSERVATION TIMESERIES

Climate data records (CDR) provide consistent long timeseries (>40 years for satellites) with known quality/uncertainty

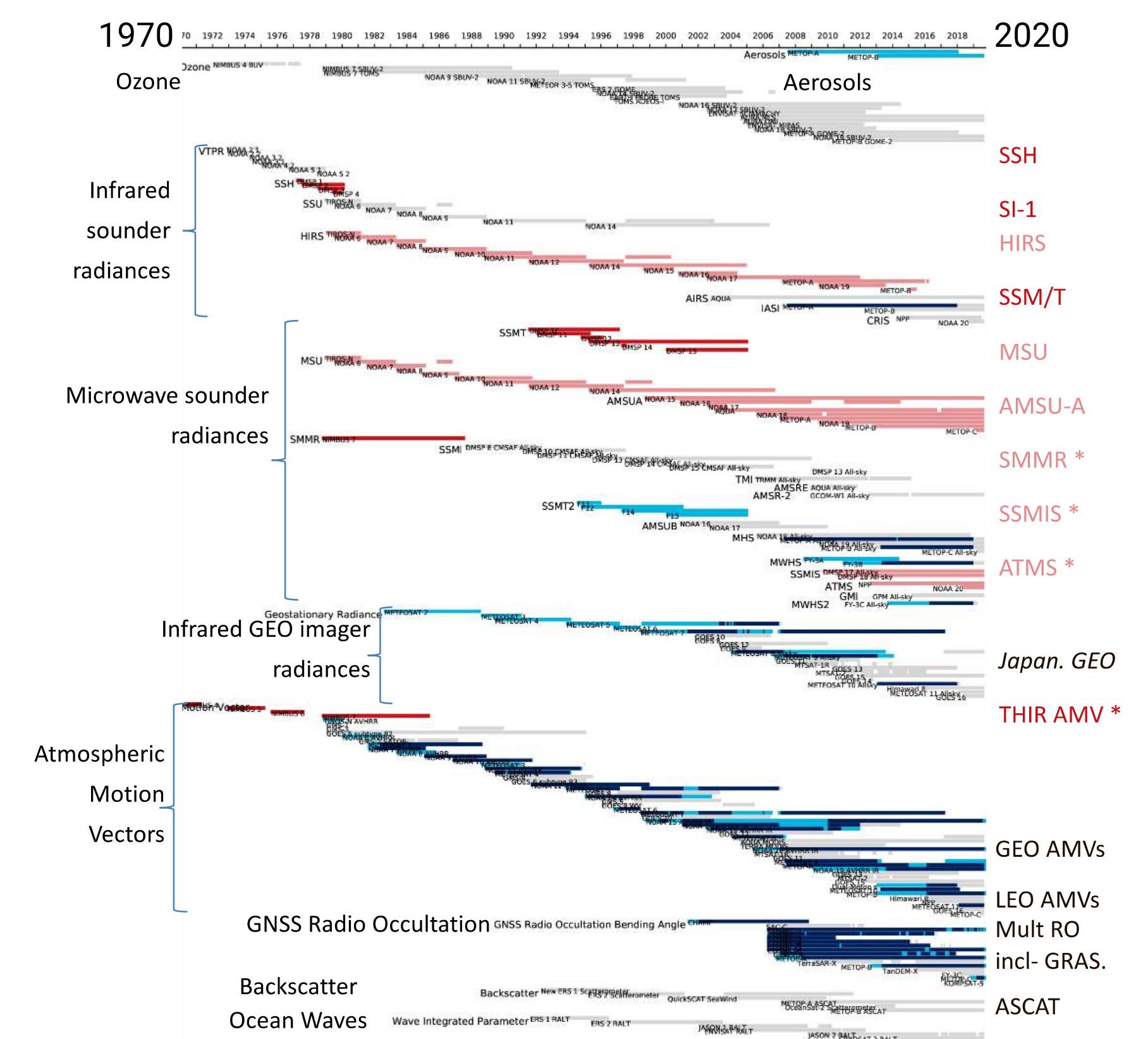


Consult EUMETSAT product navigator for more information



Time series of recalibrated radiance of 13.4 μm SEVIRI channel of MSG satellites 8-10 in Rapid Scan mode (data available every 5 minutes instead of 15 minutes of the normal scanning mode). The upper plot shows recalibrated radiances, the lower plot shows operational minus recalibrated radiances. Recalibration removed spurious jumps and drifts in the operational data.

EUMETSAT contribution to the ERA5



OPERA-SEVIRI FUSION DATASET FOR NOWCASTING

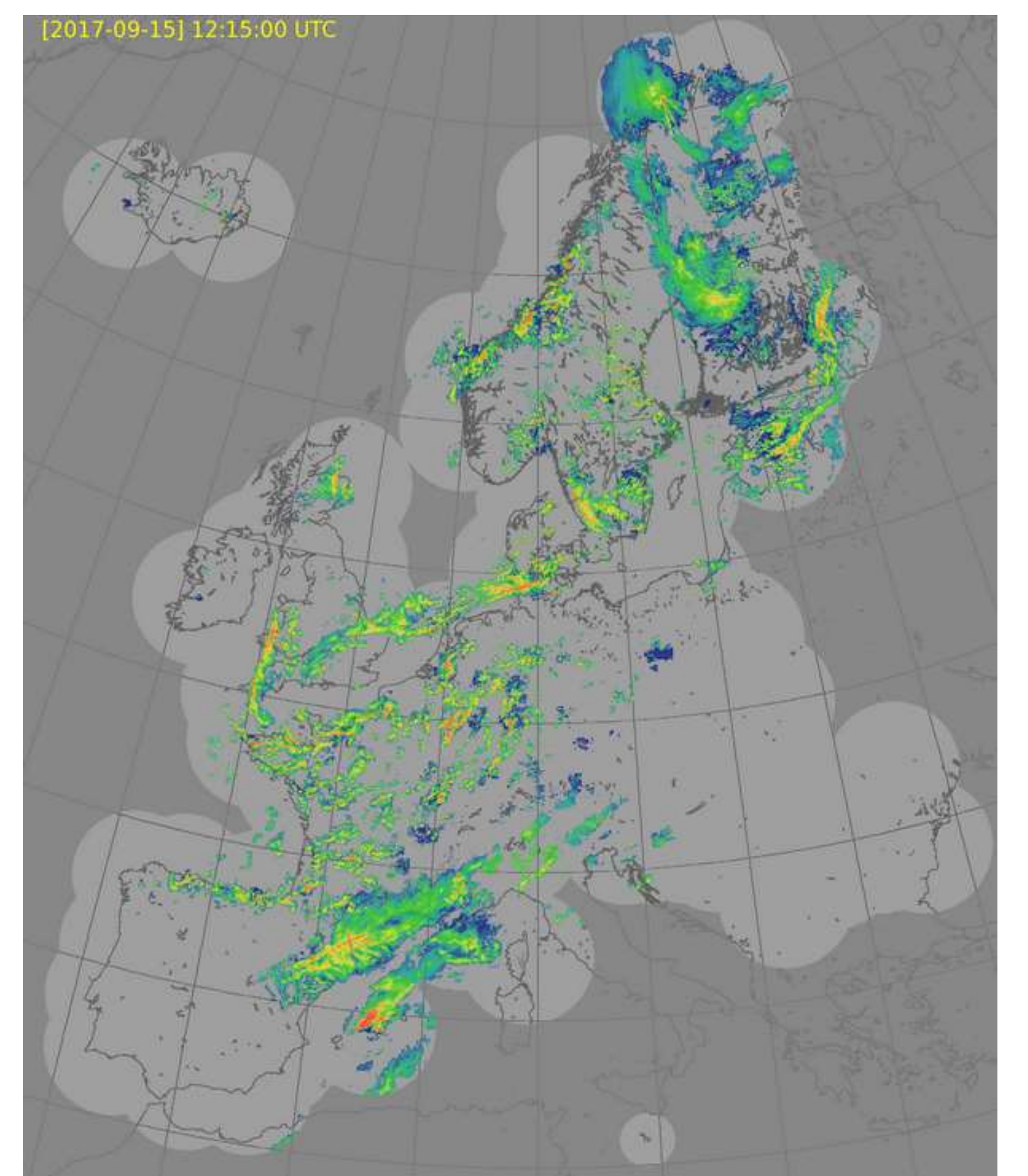
EUMETSAT and AEMET are producing combined dataset of SEVIRI Rapid Scan and OPERA weather radar data for training ML-based nowcasting models and other applications

- Time range: 2018-2023
- Time step 5 minutes time step for SEVIRI and 15 minutes for OPERA
- SEVIRI rapid scan, all channels
- OPERA rain rate (rr), 1 hour precipitation (rr1h), and reflectivity (dbz)

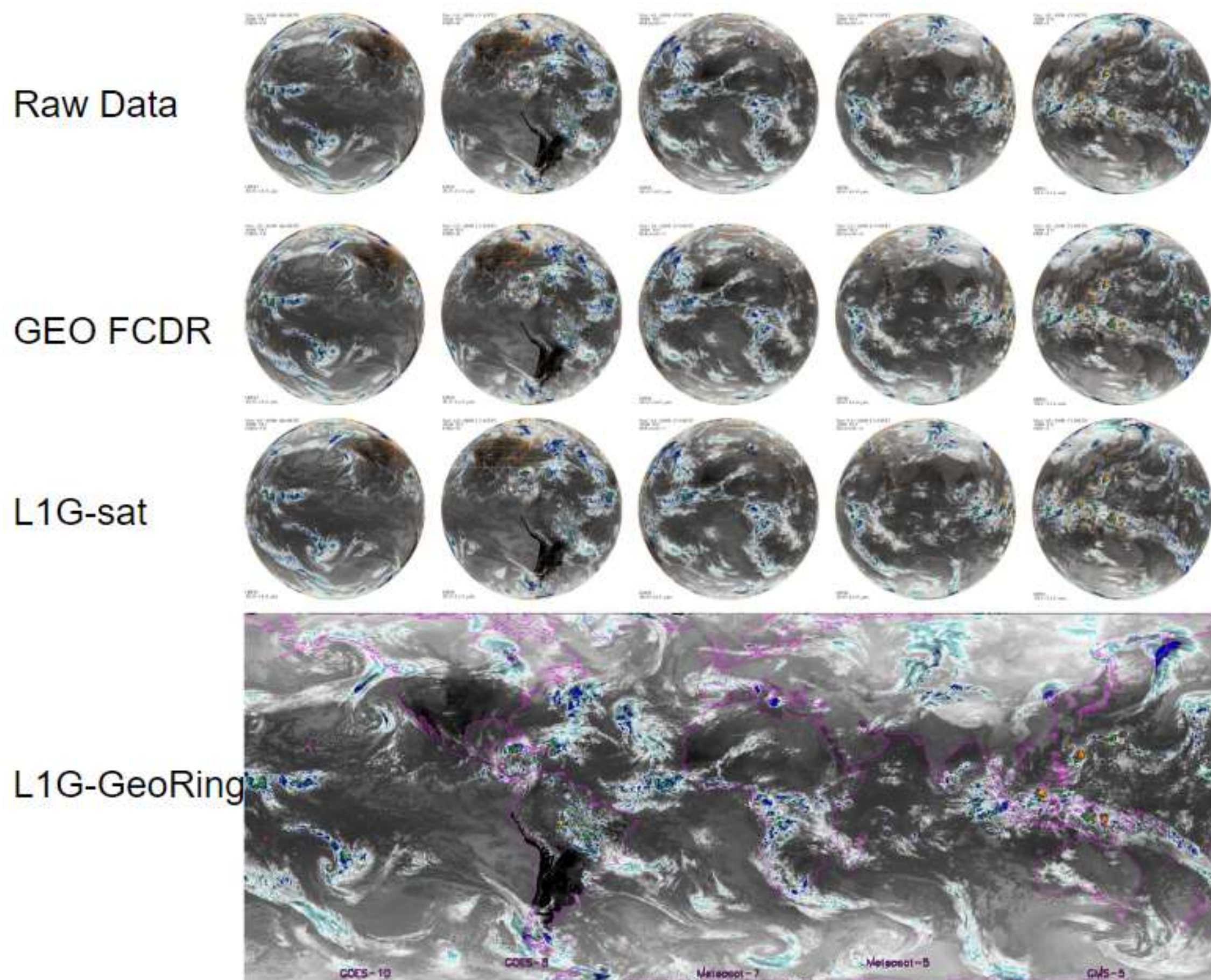
The data will be provided in the same grid

- SEVIRI data will be reprojected to the OPERA grid (2x2km Lambert Equal Area)

Provided first via European Weather Cloud (EWC) data buckets. Possibly later also via other data services



GEO-RING – 50 YEARS' WORTH OF GLOBAL FCDR



The raw data are available from EUMETSAT, NOAA, and JMA archives, each having their own format, quality information, calibration, navigation, etc. In general, IR channels are calibrated using onboard calibration targets while visible channels are calibrated using vicarious calibration based on pseudo invariant desert calibration sites.

GEO FCDR (Geostationary Fundamental Climate Data Record): the raw data is processed and quality controlled. The data are recalibrated using reference measurements. Calibration information for each channel is provided to convert raw counts to radiance data as well as reflectance (for visible channels) and brightness temperatures (for infrared channels)

L1G-sat data are remapped to a fixed grid. The brightness temperatures and reflectances will be stored for infrared and visible channels (respectively) having applied the calibration from the GEO FCDR.

L1G-GeoRing product is the combination of the individual L1G-sat files. The merged product will have the calibration applied along with the spectral band adjustment factors (SBAs) applied to maximize uniformity. It will come with 30-minute temporal and approximately 0.05° spatial sampling.

