

Supporting Weather and Climate Application Development with ML-Friendly Earth Observation Data

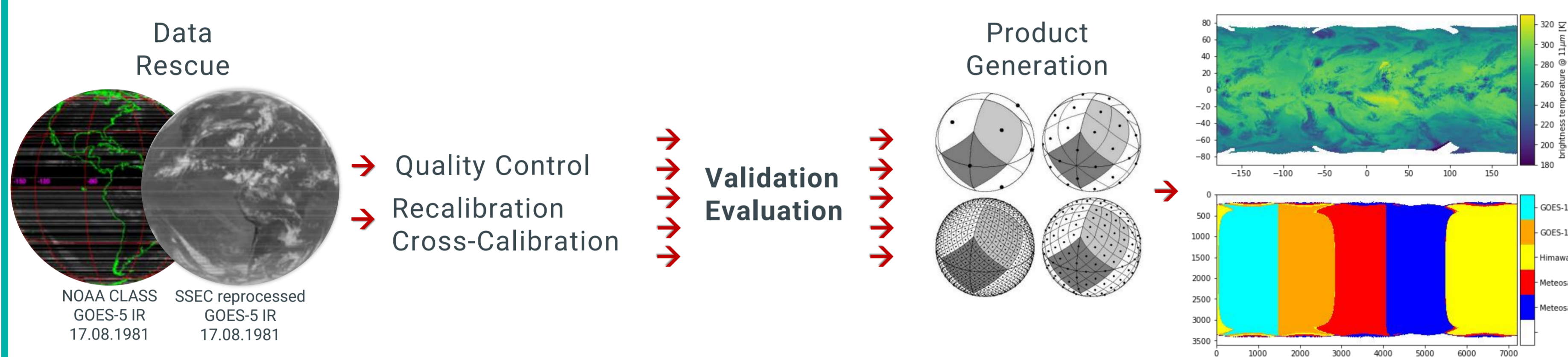
Author: Roope Tervo
Co: Lauren Biermann, Anna-Lena Erdmann, Andrew Heiding, Viju John, Paolo Ruti and Jörg Schulz

INTRODUCTION

The adoption and operational use of Machine Learning (ML) for weather and climate are often hindered by the availability of consistent, high-quality training data and the required data preparation. EUMETSAT are addressing this challenge through the generation and provision of two high-value datasets: the GEO-Ring climate data records (CDRs) and the OPERA-SEVIRI-NORDLIS fused data cube.

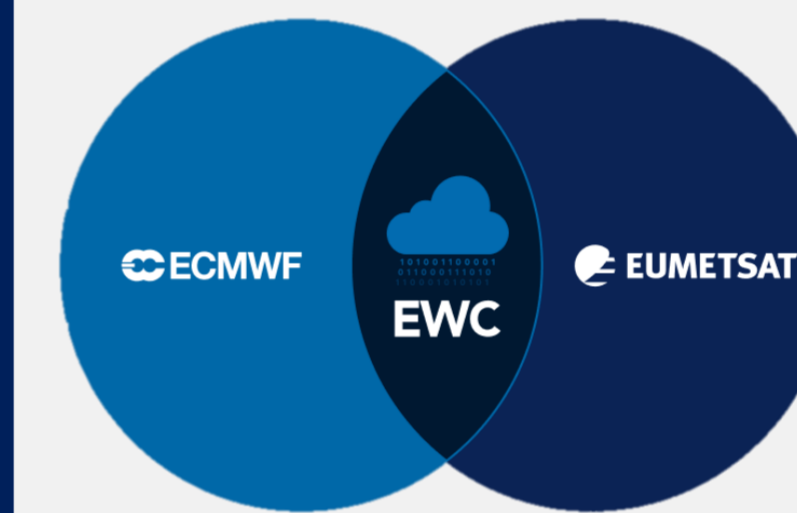
EUMETSAT CDRs

- EUMETSAT builds highest quality CDRs from past and current satellite data using level-1 radiance or level-2 reflectance data.
- State-of-art data processors are applied to ensure data from different sources are harmonized to form a coherent series.
- The steps in establishing these operational CDR's involve a very strict **research-to-operations process** as shown in figure 1 below:



EUROPEAN WEATHER CLOUD (EWC)

- Community cloud for EUMETSAT and ECMWF Member and Cooperating States for their official duties and R&D.
- Users can deploy computing resources close to the EUMETSAT and ECMWF data holdings.
- Available via Member State Computing representatives or EUMETSAT R&D calls and ECMWF Special Projects.



<https://europeanweather.cloud>



GEO-Ring (Level-1g)

EUMETSAT and NOAA are working to produce a gridded GEO-Ring CDR from global geostationary satellites through consistent quality control and recalibration of Level-1 radiances. A static test dataset is available covering 2019 – 2024.

- Data scope: ABI (GOES), AHI (Himawari) and SEVIRI (Meteosat).
- 0.05° grid sampling on a HEALPix Grid with a temporal resolution of 30 minutes.
- S3 access from within the EWC, available to internal projects only.

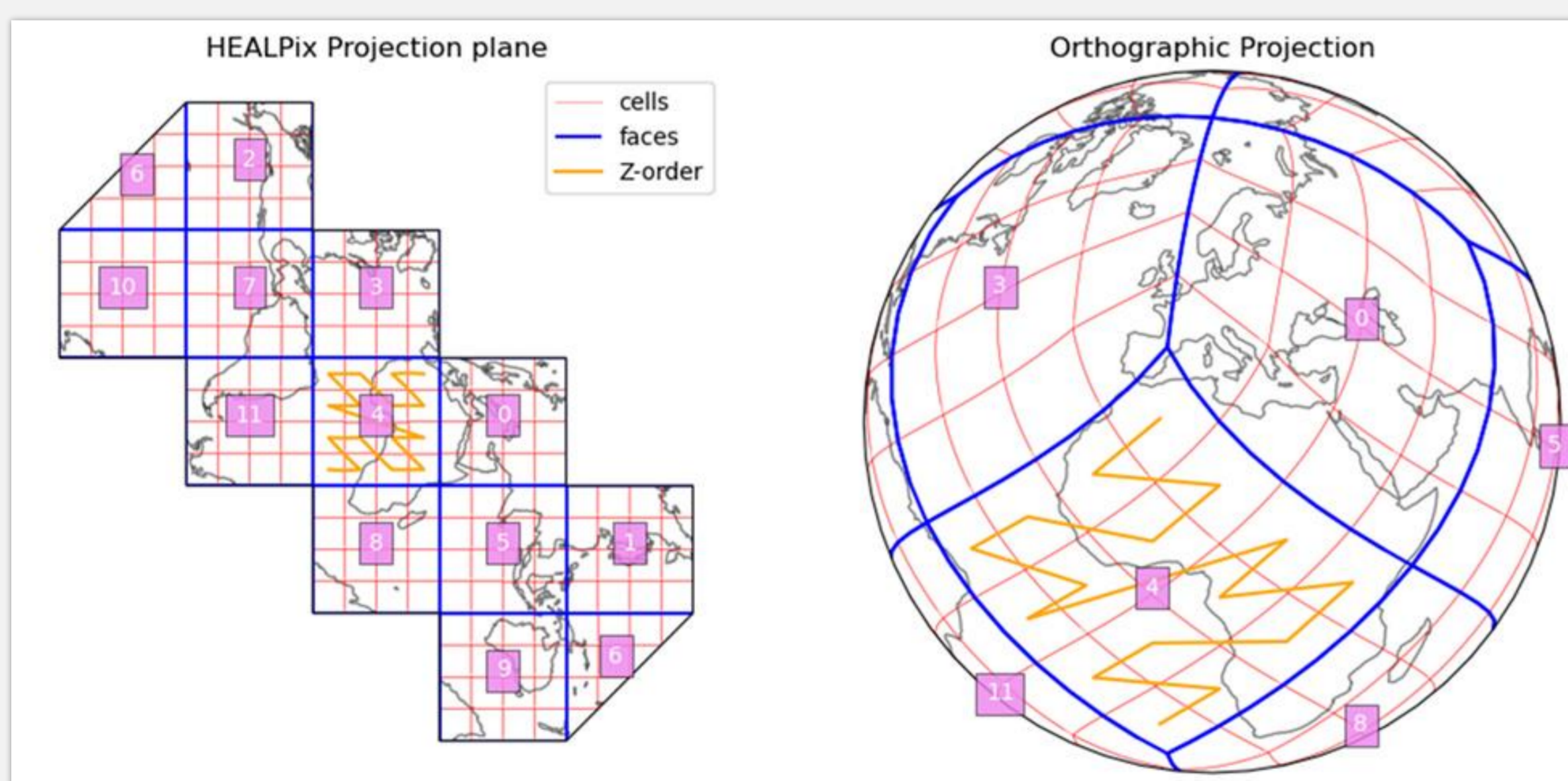
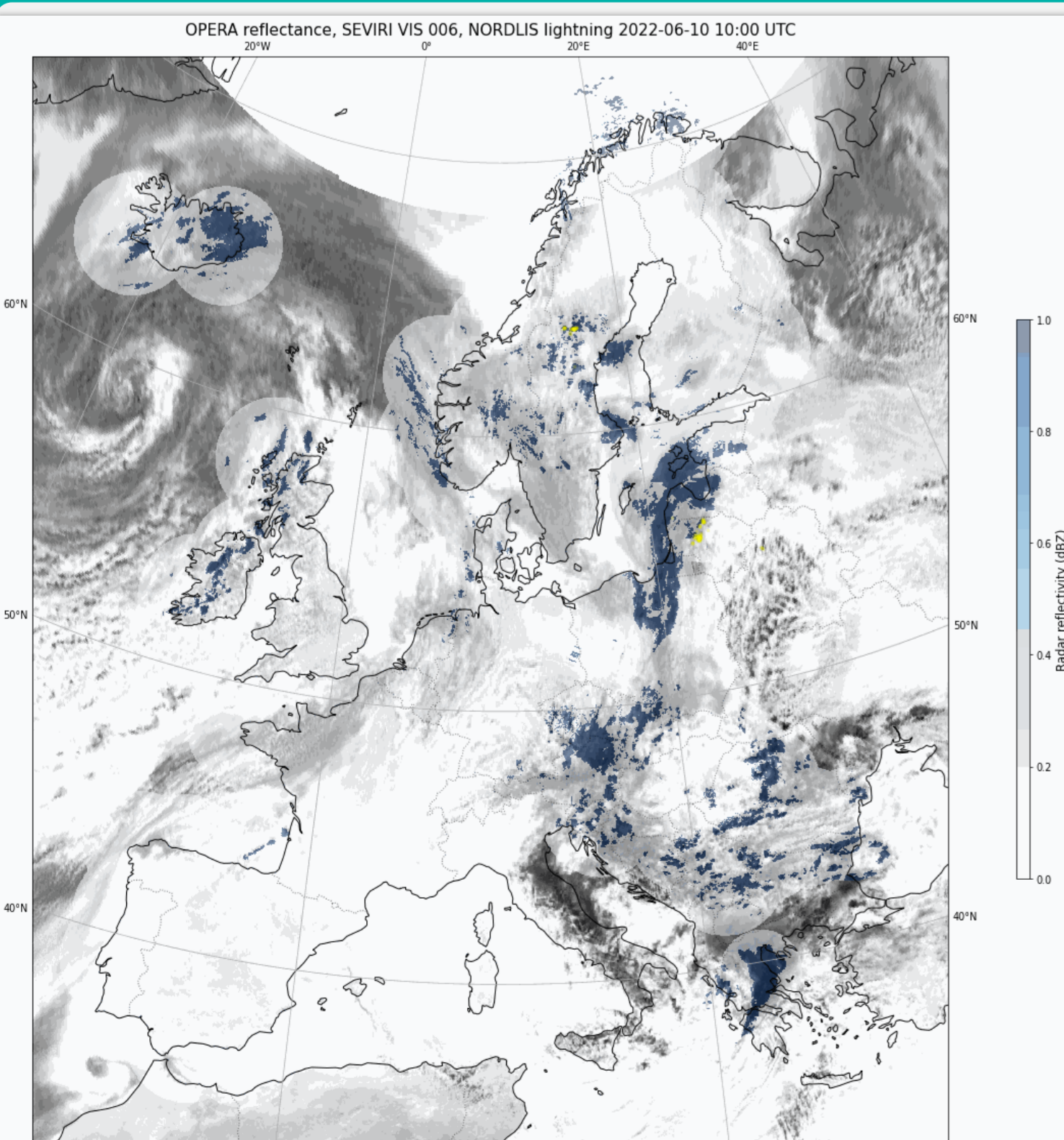
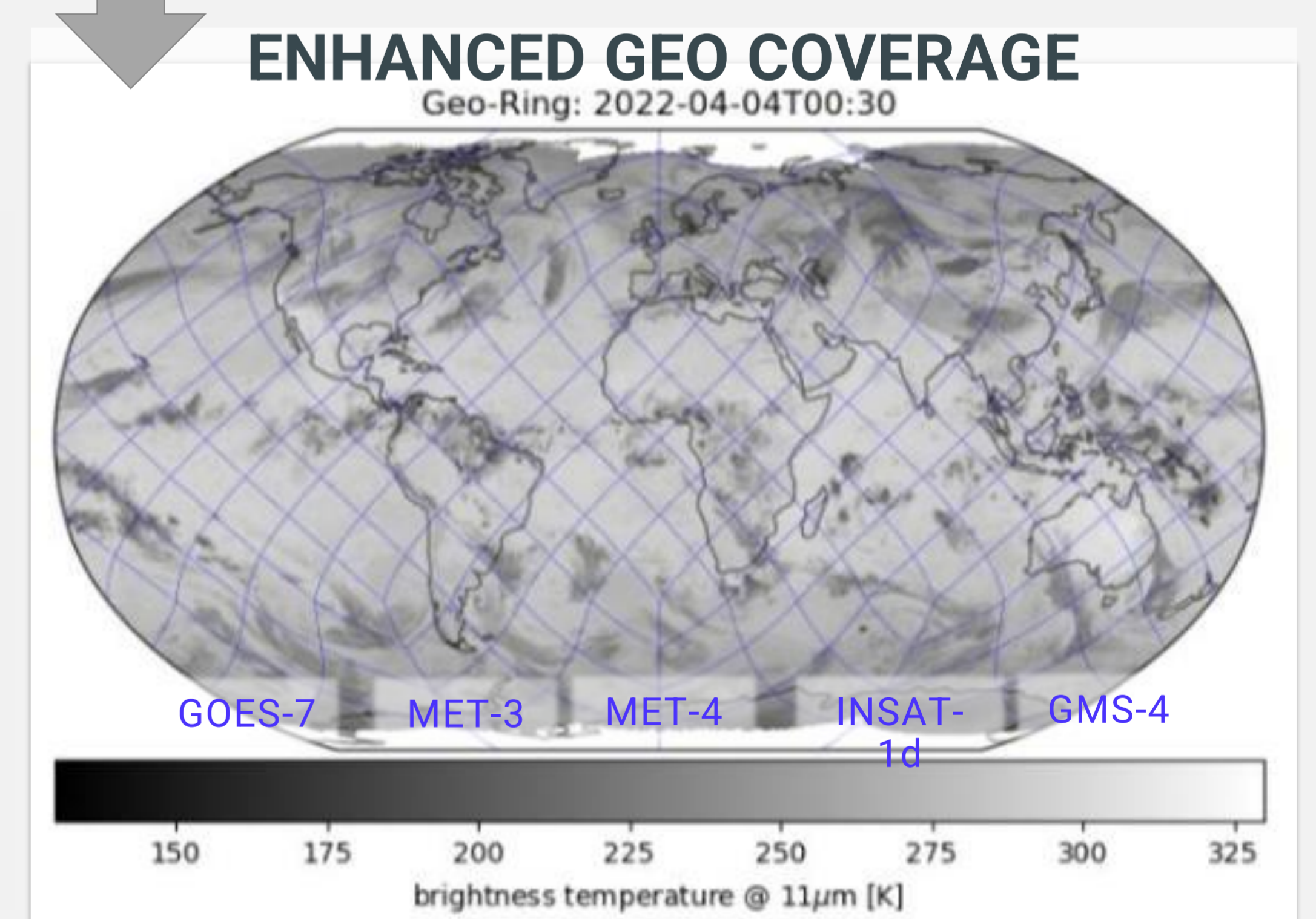
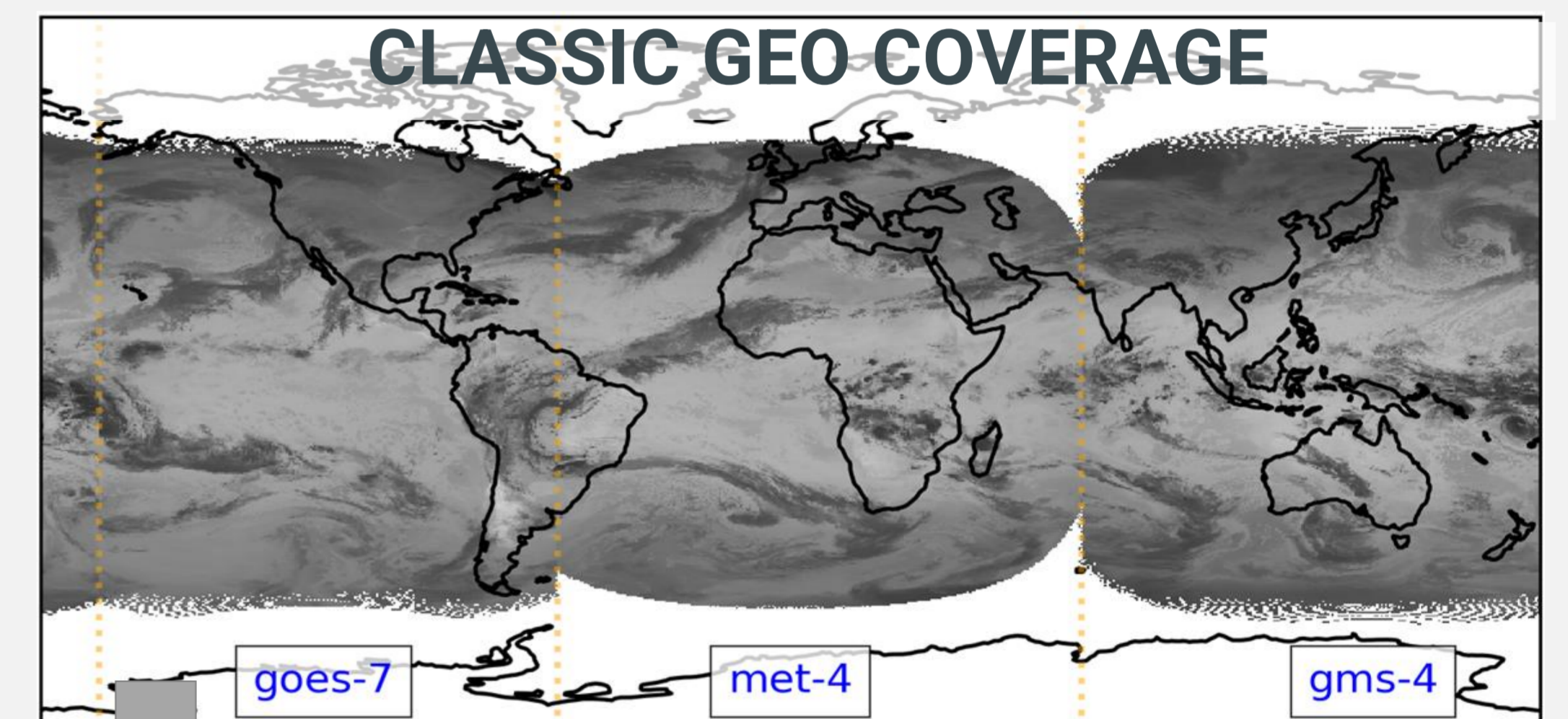


Figure 2 (Left) illustration HEALPix Projection Plane.

Figure 3 (Right) showing improvement in coverage in new gridded GEO-Ring dataset.



OPERA-SEVIRI-NORDLIS fused data cube

Combined data cube of OPERA, SEVIRI, and NORDLIS covering 2018-2023 provides solid foundation for ML-based nowcasting models and other applications.

Available formats:

- NetCDF: one S3 bucket per dataset, nc files for each time step – by year, month, day.
- Zarr conversion in progress: one zarr cube for all datasets, one group for each dataset.

Data access: S3 access from within the EWC, available to internal projects only.

Temporal resolution: 5 minutes (SEVIRI, NORDLIS), 15 minutes (OPERA).

Data scope: **OPERA:** rain rate (rr), 1 hour precipitation (rr1h), and reflectivity (dbz); **SEVIRI:** all channels; **NORDLIS:** flash count.

Contact person:

Dr Roope Tervo – roope.tervo@eumetsat.int