



AI-ML next steps: proposed use cases for CGMS

WG II

Topics and actions to be discussed at the plenary include:

- Data curation activities to build best practices for AI readiness, considering also standards for data processing and formats.
- Cooperation across international working groups on ML applications.
- Specific collaborative projects under CGMS.

Next Steps

- Focusing on few use cases:
 - Constructing robust data sets for training of ML prediction systems, especially at regional scale
 - Supporting ML Nowcasting development
- What is needed
 - Select satellite parameters; regridding functions; assess the impact on ML systems; Working with ML experts to build lesson learnt on standards
- Who has responsibilities
 - WG I for standards; WG II for selection, quality and methods; WG IV for ML operations
- CGMS internal coordinator role across WGs: identify a person
- Future extension to Space Weather use case

Example for Supporting ML Nowcasting development

- ML Nowcasting systems
 - Nowcasting models enable the forecasting of extreme precipitation, making them highly relevant
 - ML nowcasting systems work first with regrided radar data as a training base, then they use satellite data to build a proxy system
 - There is a possibility to exploit high-resolution (in space and time) hyperspectral radiosoundings and lightning data
- What is needed
 - High quality precipitation satellite data as for verification
 - Select satellite parameters (high resolution space-time) from Geostationary platform, but new microwave constellation could provide similar temporal resolution in the future
 - Regrid on a regular grid for targeted areas the selected periods
 - Assess the skill of ML nowcasting models
 - Build lesson learnt on standards
 - Investigate the relevance of other parameters (hyperspectral infrared sounder, lightning)
- Who has responsibilities
 - WGii to identify areas, select satellite parameters
 - Required additional effort on data curation (regridding, QC) – who and how?
 - WGi infer the standards on data processing
 - WGiv ML operations analysis

Thanks

Specific Example for constructing robust data sets for ML training

- ML regional weather prediction
 - There is a need for high-resolution variables that enable ML models to learn about the vertical structure of the Planetary Boundary Layer during the training phase
 - The experience from the AI-FS stretched system suggests a limited number of years, around 3-4 years, as a good initial training time window for fine-tuning global ML for a regional area
- What is needed
 - Identify target areas
 - Select satellite parameters (high resolution) targeting lower atmosphere and convective related variables
 - Regrid on a regular grid for the selected periods with a reasonable time frequency
 - Assess the impact on ML fine-tuning for existing models
 - Working with ML experts to build lesson learnt on standards
- Who has responsibilities
 - WGii to identify areas, select satellite parameters
 - Required additional effort on data curation (regridding, QC) – who and how?
 - WGi infer the standards on data processing