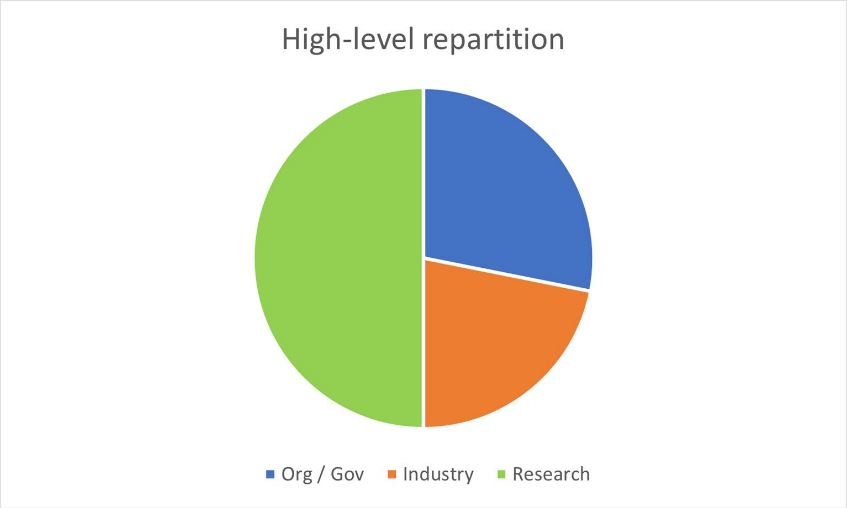
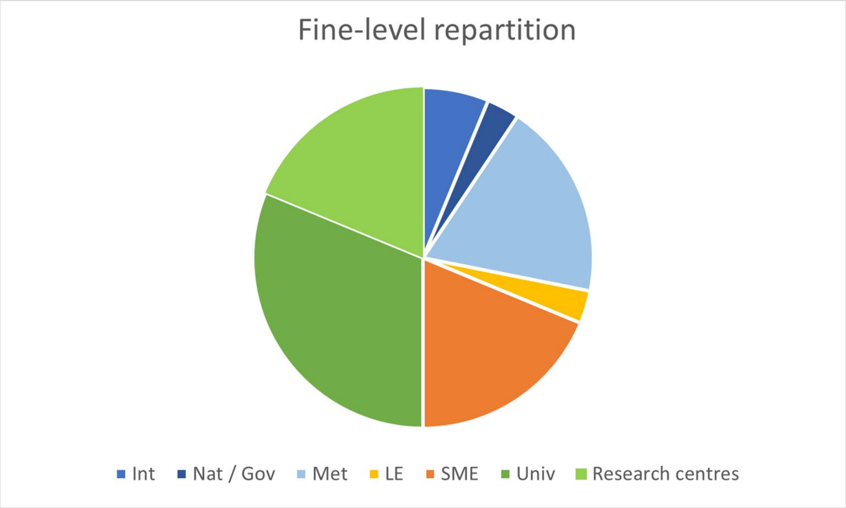
CHAIRS: Bertrand Le Saux (ESA), Claudia Vitolo (ESA) and Patrick Laloyaux (ECMWF)

ATTENDEES/AFFILIATIONS

- 32 active participants (registered: 22 in presence / 30 online)
- From Organisation (int / nat / gov), Industry (large and SMEs), Research (universities, research centres)





CURRENT ML APPLICATIONS IN THE THEMATIC AREA

a) Weather & Climate:
- Nowcasting precipitation (DL)
- Extreme event detection and forecasting e.g. tropical cyclones
- Weather regimes and atmospheric patterns (considering teleconnections) on S2S time scale
- Multivariate dependencies for climate variables, predicting climate risks
- Post-processing to correct systematic errors of NWP models, e.g. for probabilistic weather forecasting
- Combining statistical ensemble models with AI for gridded nowcasting → subseasonal predictions (so far temp+precip) (unet, SAMOS, imbalance regression)
- downscaling+correcting subseasonal predictions from coarse to fine scale (~1 km)
- Incorporation of information theoretic quantities to analyze the relationships between climate variables. Transfer Entropy, providing cause and effect relations

b) Environmental applications … actionable information using EO data
- Energy sector (Solar energy applications, energy distribution and demand, etc
- Hydrosphere applications: Lake ice models - lake water quality products - downscaling Antarctic surface melt estimations
- Early warnings, environmental predictions, wildfire predictions, predictions of river flows
- Public health
- Crop yield forecasting
- Modelling of near-shore bathymetry

c) Industrial applications
- Aviation sector, meteo events, during flight planning
- (re)insurance market
- Road temperature and state (wet, ice, dry) prediction based on road maintenance data, NWP, +webcam
- Forecasters exposed to new (AI) products, testbeds

d) Emerging fields:
- (network) federated learning
- Security applications
- Collecting requirements for satellite tech and from policy DGs to prototype applications (develop hybrid models, quantify uncertainties, xAI)
- ML for Quantum Computing, onboard applications, Digital Twins, Data Fusion, Assimilation, …

LIMITATION

- Cannot use cloud computing for security applications
- Time consuming
- Pay for data/services

CHALLENGES

- Lack of trust, black box (XAI is very difficult for complex models, including GANs)
- Lack of understanding, Fear of being replaced by robots
- Lack of common language (for industry and people that are not necessarily domain experts) - Improving communication (technical presentation) - the common language is fundamental when you work across disciplines and among people with different level of expertise.
- Data access (esp. from industries), data quality, and missing values imputation
- Reproducibility/comparability - Need for more benchmark datasets
- For new AI products (for forecasters) it is important to provide good baseline skills (testbeds), Forecasters need to adjust to new products.
- Operationalisation
  - Maintenance
  - Computer resources? (outdated CUDA libs which only permit old tensorflow/pytorch libraries could be an issue)
  - Local data availability
- Same project underway in different centers / research base
- Generators - these models are very difficult to understand and explain
- Transparency in the AI pipeline / dataflow

OPPORTUNITIES

- XAI, Hybrid modelling
- Continuous learning, online learning, open-set learning
- Self-supervised learning: pre-training of large models, sharing pre-trained weights with the community, boosting advancements & innovation
- Forecasting surrogates to improve f. skills
- Community building
  - Human in the loop, better/more dialogue between academics and industry (for rapidly building operational prototypes)
  - Training, hands-on activities, Make presentation less technical (less jargon
  - Interdisciplinary education between CS, atmospheric sciences, statistics etc for the next generation of scientists
- Software / data initiatives
  - Open source, Open repositories
- Hardware Initiatives
  - Cloud computers becoming more accessible
  - Public owned and operated with data subscriptions (sovereignty)

ADVANTAGES (DISADVANTAGES?) OF ML TECHNIQUES FROM TRADITIONAL STATISTICAL METHODS

1. TRUST - EXPLAINABILITY
   - Backpropagation and looking at activation to explain results
   - Explainability depends on the complexity of the model, in general more traditional models feel easier to explain
   - ML can (sometimes) learn the physics, when not possible explainability tools are important
   - How do you validate models? Missing unit test in ML pipelines. Audience reporting on: observing system experiments (running models with and without a subset of data, to understand the how that data influences the result); define a single error metric (de-assimilate observation). Central certification standards.
2. MODEL PERFORMANCE - are ML techniques better that traditional stat models?
   - It depends on the task, for some tasks (computer visions, processing large images from satellites, need to learn physical processes, need to combine heterogeneous data sizes/types/formats) ML is often better
   - Recent papers claim to predict better than ECMWF model (using ERA5)
   - For physically based models we have a good understanding of what works, for ML models we need continuous improvement
3. REPRODUCIBILITY - TRANSPARENCY
   - More and more papers with code
   - Open source commonly accepted, open development less (varying degrees of opensess especially in industry)
   - EU committed to make application code open (e.g. code.europa.eu). There is still lots of work to do to implement that across projects, departments and so on
   - Stochasticity of ML models make it more difficult to reproduce
4. COMPUTE MEANS (inc. accessibility)
   - Compute constraints, GPUs available through the NVIDIA grant
   - Is it more expensive to run in-cloud or on premise? It depends on how much you use GPUs. Possibility to purchase GPUs utilised in-cloud?
   - Time consuming, crashed program …solution : run on server or online
   - High computational requirements are needed when scaling up applications, initial tests could be done on smaller scale on local machines. Latest hardware is powerful for most basic ML tasks
5. ETHICS & PRIVACY
   - For some governmental organisations, open data is acceptable - source code should be closed?
   - Big companies are moving towards open source - new business models are taking over
6. OPERATIONALISATION
   - ML models are fairly new, maybe will become operational in 5 to 10 years

FUTURE DIRECTIONS

a) IMPACT ON SOCIETY

- Low cost replacement for complex NWP, or satellite obs replaced by low cost sensors
- Fast simulations, obtaining quick actionable information

b) IMPACT ON CLIMATE - AI vs GREEN COMPUTING?

- ML could help adopt more sustainable/renewable solutions for the energy sector
- ML could improve early warnings
- Energy used for simulations - we need to find a way to use computers more efficiently

c) TECHNICAL ASPECTS (PROMISING SUB-FIELDS OF ML)

- Transformers, capabilities and accurate decision making. Currently used in energy sector. Seem to be state-of-the-art, however we need to acknowledge some techniques loose popularity quickly. One nice feature of transformers is that you can plot the attention matrix at inference time. So the model can (pretty literally) show you what it's focusing on. Which could help explain what the model is "thinking".
- Explainable AI - more technique coming up - still developing and difficult to validate explanations. Less hesitancy than before to adopt AI.

d) FURTHER NEEDS IN EO EARTH SCIENCES TO BE INVESTIGATED?

- AI could be used more for predicting extreme events
- Need for more general usage → Using IoT in Digital Twins, ontologies, comparing HRes output of DTs to observation, ML to describe relationship amongst Earth system processes.

e) OPPORTUNITIES (FUNDING-GRANTS) FOR ACCESSING COMPUTING RESOURCES?

- Funding sources for maintenance of algorithms for downstream applications (updating after major releases of IFS, change of resolutions).
- Training on how optimally use GPUs/ HPCs

f) Impact on pushing scientific boundaries for process understanding in Earth science ?