- CHAIRS: Rossella Arcucci, Marcin Chrust
- PARTICIPANTS: 16 in person, 26 online
- ORGANISATION TYPE: Academia, Industry, NWP centres, Research Centres

CURRENT ML APPLICATIONS IN THE THEMATIC AREA:

a) Fully AI based analysis update
   a) Using the neural network to do analysis update and adversarial method to bypass Gaussian assumptions
   b) Emulate data assimilation based on an existing conventional DA system (train on background and analysis increments OmB)
   c) Recurrent Neural Networks as replacement of 4DVar (4DVar-Net) – joint learning of variational data assimilation models and solvers
   d) Extended Elman Network for Bayesian data assimilation (estimate the posterior covariance matrices)

CURRENT ML APPLICATIONS IN THE THEMATIC AREA:

a) Hybrid ML/DA (analysis)
   a) Learning full model with a DA step followed by ML step
   b) Model hybridization: combining physics based model with a statistical model
      a) Offline training of neural network based model error
      b) Joint estimation of model error/parameters and state
   c) Covariance matrix estimation with ML techniques
      a) Observation Error Covariance matrix estimation – use recurrent neural networks with/without parametric estimation
      b) Background Error Covariance matrix estimation – use generative neural networks (adversarial, variational)
      c) Estimation of normalization factors of correlation operators using ML techniques
   d) Estimation of model parametrizations
   e) Making a mapping from state space to control/latent space using encoders (for model reduction, to work with Gaussian error statistics)
   f) Estimation of hyper parameters in DA,
      a) e.g. optimal localization matrix
   g) Learning mapping between observation and state space

CURRENT ML APPLICATIONS IN THE THEMATIC AREA:

a) AI for prediction/correction
   a) Developing emulators of the dynamics with the purpose of using them in ensemble forecasts, error covariance statistics estimation
   b) Emulation of tangent linear and adjoint models
   c) Emulation of radiative transfer models, or observation operators more generally
   d) State dependent model error correction applied throughout forecast
   e) Using ML for pre-processing of observations (identification of outliers, quality control, observation information augmentation, reconstruction of missing observations)
   f) Applying ML for reconstruction of past climate

What Data Assimilation can bring in to Machine Learning world

a) Ensemble Kalman Smoother as replacement for the backpropagation to address exploding/vanishing gradient problems
b) Account for uncertainties in the ML loss function/advocate Bayesian framework for ML
c) Use of error estimation for characterizing training data/model error; e.g. use ENKF/iterative ENKF to learn parameters of neural networks
d) ML would benefit from DA tools and methods to deal with sparse and noisy observations

LIMITATIONS, CHALLENGES AND OPPORTUNITIES:

a) Training ML model on noisy/biased observations result in imperfect/biased models
b) Learning full model is challenging for high dimensional systems
c) Propagation of uncertainty in case of neural networks is very challenging due to requiring very large ensembles
d) Handling of extreme events requires dedicated ML techniques
e) Communication between ML practitioners and NWP community is gaining momentum and should be encouraged
f) When generating ML based observation operators we may introduce correlations between background and observation errors
g) Accounting for physical constraints and multivariate relationships by ML techniques is a challenge

LIMITATIONS, CHALLENGES AND OPPORTUNITIES:

a) Hybrid ML/DA techniques, such as joint estimation of neural network based parameterizations face similar challenges to those of parameter estimation
b) Encourage operational NWP centres to account for ML requirements when producing reanalysis data sets
c) Ability to combine all observational information with physical constraints to produce forecasts
d) Background error statistics can be difficult to transfer from state space to the latent space
e) Can the ML revolution sweep away traditional NWP?

3. ADVANTAGES (DISADVANTAGES?) OF ML TECHNIQUES FROM TRADITIONAL STATISTICAL METHODS:
   a) Speeding up the Data Assimilation process important from operational applications perspective
   b) Low cost generation of ensembles for non-linear DA approaches
   c) We typically rely on Gaussian assumptions in DA, while ML can deal with non-Gaussianity and non-linearity
   d) Non-linear ML based analysis update vs typically linear analysis update
   e) Use ML to transform problems into Gaussian space
   f) Build ML based observation operators, in particular for observations which we do not know the mapping from the state space

4. FUTURE DIRECTIONS

a) Towards unified ML-DA framework

b) Multi-modal ML to merge different sources of information (e.g. satellite radiance observations with imagers)

c) Use ML techniques to make use of new sources of observations

d) Use ML methods for observation pre-processing, e.g. more intelligent thinning, data selection to maximize the information content, estimate observation statistics

e) AI embedded satellite instruments geared towards useful NWP information

f) Develop ML techniques exploiting information embedded in an ensemble for computation of an analysis update

g) ML techniques bridging observations directly with products (e.g. from satellite observations to day 2 precipitation forecasts)