Group #3
AI Contribution to Climate Data and Modelling

Science challenges & feasibility of potential AI solutions

Connecting Atmospheric CCI ECVs to Ground Phenomena

Cloud Masking for CCI ECVs

Seed Questions

1. Where is AI already used in the area of connecting atmosphere and ground measurements, and for cloud masking? What are lessons learnt and what are the unresolved challenges?
2. Atmospheric ECVs and ground measurements are very different in data type, spatial and temporal representivity etc. How can AI facilitate an optimal connection of them?
3. Machine learning requires training data. Cloud masking is a typical application for machine learning. What are the requirements on training data and by what means can we collect them? Can AI support the validation of cloud masking?

Never forget, where the information comes from …

Where is AI already used in the area of connecting atmosphere and ground measurements, and for cloud masking? What are lessons learnt and what are the unresolved challenges?

- Speeding up RTM,
  - has been used in Inversion schemes as replacement for slow RTMs, in particular under cloudy conditions
  - The question was raised, if there is a chance where AI can help to improve RTM in complex situations , e.g. 3D parametrizations, aerosol, cloud (although it is important that the physics are not violated)
  - For RTM replacement AI more a kind of fast/compressed LUT replacement,
- Cloud masking: Many great examples where ML enables seamless integration of spectral, spatial and temporal state vectors to detect clouds. Actually, pattern recognition is a key competence of ML
- Example was given, of connecting PM (air quality) point observations with satellite observations. These examples were not very convincing, since they reflect local statistics rather than information content of measurements. Great amount of care is needed …
- AI can’t be better than training data …

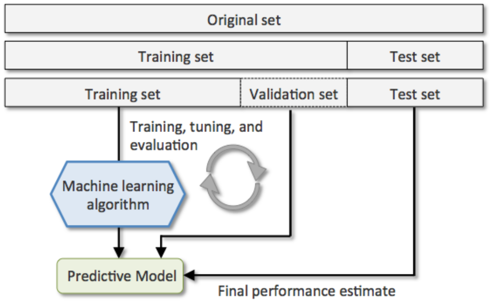
Atmospheric ECVs and ground measurements are very different in data type, spatial and temporal representivity etc. How can AI facilitate an optimal connection of them?

- AI has the capacity to be used a 'transfer' function between different spatial, spectral and temporal samplings. As for RTM's, great care must be taken, that AI is not used as a cosmetic gap filler, which invents information, where is none. (As always) the quality of training and test data is crucial
- Examples were given, for the quality control of cloud CCI data using MWR (liquid water path information), to detect erroneous data in a high speed and efficiency, further to discriminate high altitude / snow as false positive cloud detections – in this case this was a faster, less costly option to reprocessing all of the data.
- AI can be a very good tool (fast!) to define valid ranges of measurements or physical parameters in high dimensional space. Self organizing maps or other kernel density estimators could be helpful.

Machine learning requires training data. Cloud masking is a typical application for machine learning. What are the requirements on training data and by what means can we collect them? Can AI support the validation of cloud masking?

- Requirements on training data and test data: **Read the text books!** Training data must be representative spectrally, temporally and spatially – often ML algorithms do not transfer well between sensors.
- Clarity in naming: ‘training’ <-> ‘validation’ <-> ‘test’ datasets
- Great care must be taken to avoid (hidden) biases.
- Training data can come from other cloud detection algorithms, expert classification, citizen science. Biases in any training data have to be investigated and quantified by independent means.

Data warping (geometric spectral …)



Source: http://www.cs.nthu.edu.tw/~shwu/courses/ml/labs/08_CV_Ensembling/fig-holdout.png