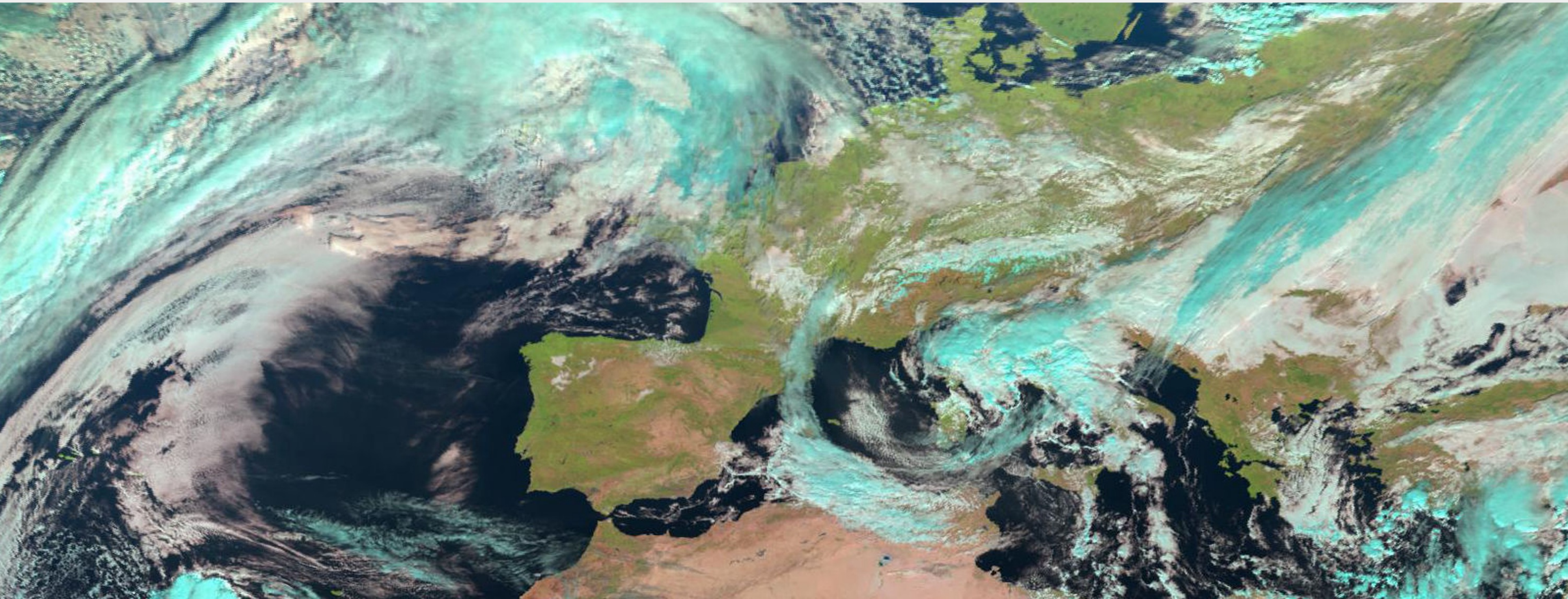
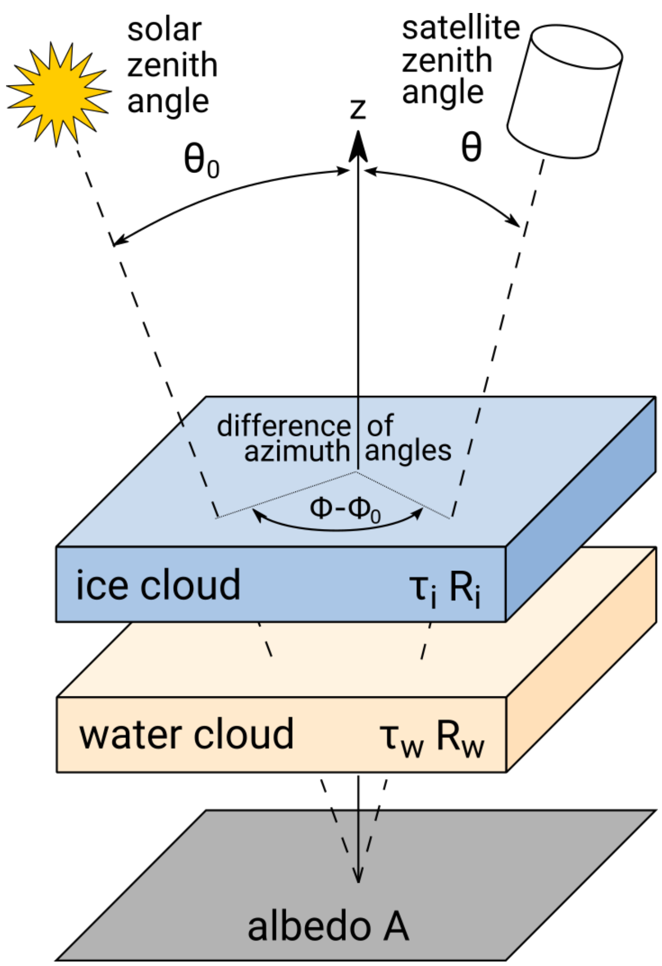
Neural network-based methods for simulating cloud- and aerosol-affected solar satellite channels

Leonhard Scheck[1,2], Florian Baur[1,2], Olaf Stiller[2], Christina Stumpf[2], Christina Köpken-Watts[2]
1) Hans-Ertel-Center for Weather Research / LMU Munich, Germany
2) Deutscher Wetterdienst, Offenbach, Germany



MFASIS for cloud-affected visible channels (LUT-based)

Method for Fast Satellite Image Synthesis



The "conventional" (no machine learning) method that has been used for model evaluation + assimilation experiments.

- **Simplify vertical cloud structure:** Complex structure can be replaced by two homogeneous clouds with same optical depth without changing reflectance significantly
  → only 4 parameters (optical depth, particle size)
  + 3 angles, albedo → **8 parameters per column**
- **Compute 8-dimensional reflectance look-up table** (LUT) with discrete ordinate method (DOM) for all parameter combinations → **8GB,** use **lossy compression** → **21MB** = O(CPU cache)
- Determine parameters from profile, interpolate in LUT

fast (O(μsec/column)), mean reflectance error < 0.01
Implem. in RTTOV 12.2 by DWD in collab. with MetOffice

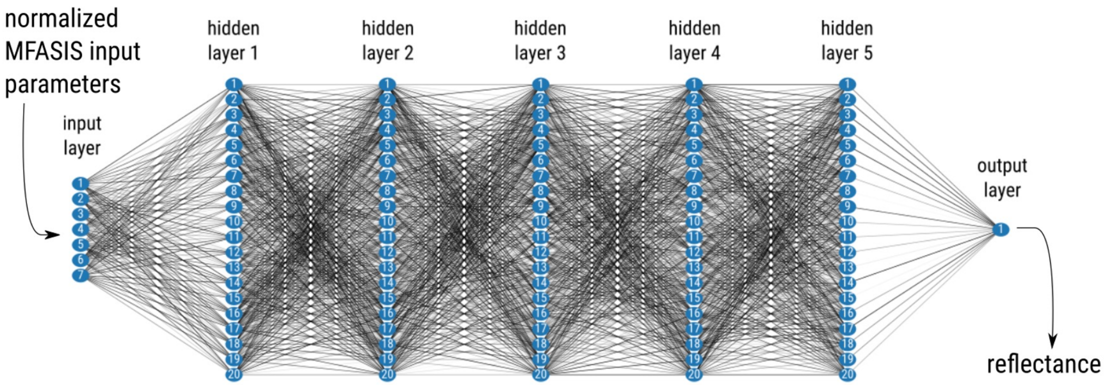
- Simple corrections for mixed-phase clouds and weakly water vapor sensitive channels (0.8μm SEVIRI)
- Preliminary correction for 1.6μm channels

Could we replace the LUT by a neural network (NN)?

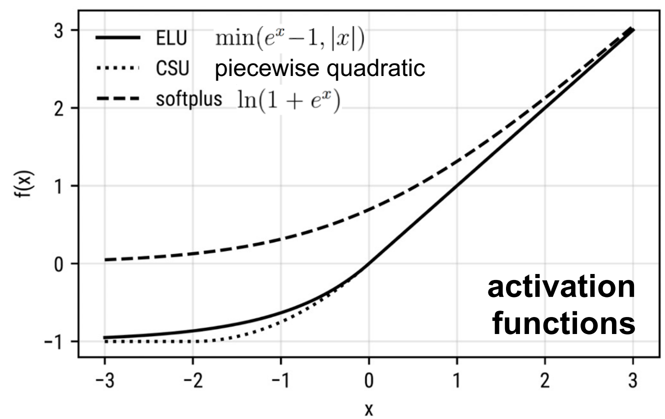
Motivation: Absorbing channels (water vapor, trace gases, clouds) and aerosols (many different species) require additional input variables → LUT size would explode…

Approach: Keep idealized profile strategy (low number of input parameters) but use relatively small (= fast) feed-forward neural network (several 1000 params.) instead of LUT

First goal: Replace LUT by NN for the visible 0.6µm channel (no additional inputs)



$$s_{l+1,n} = f(b_{l,n} + \sum s_{l,m} * w_{l,n,m})$$



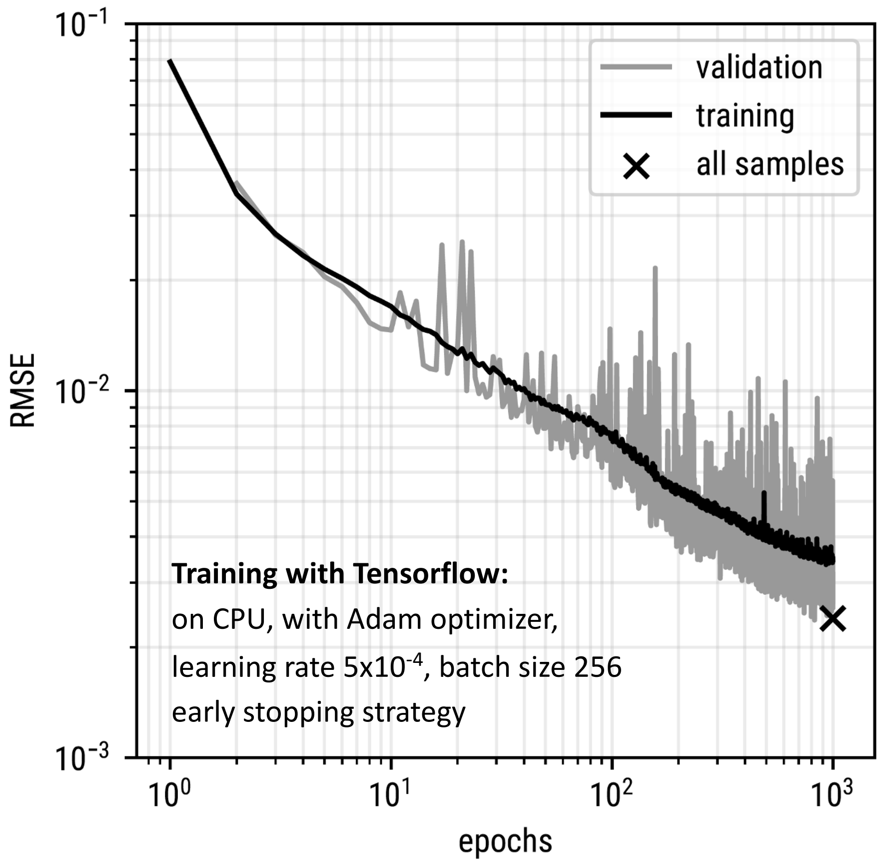
NN structure: best results for 4 – 8 hidden layers (“deep”), CSU activation function

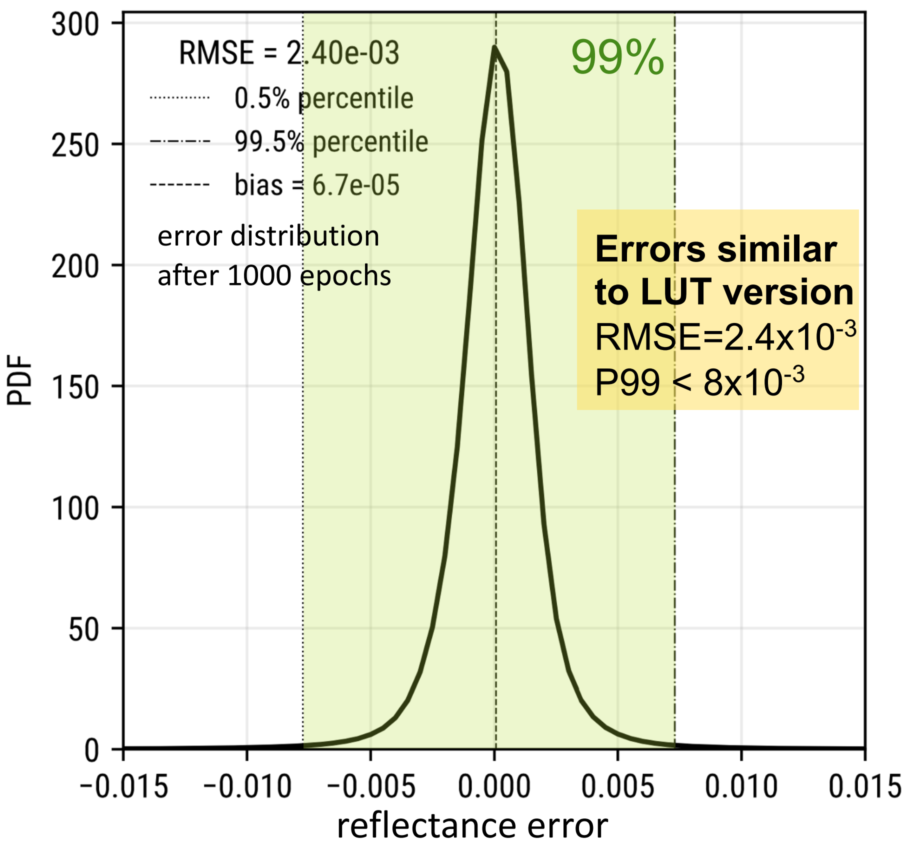
Training data: Synthetic (random numbers for input params., reflectance computed with DOM)
→ produce as much data as is required, cover full parameter space with constant density

Training process: Tensorflow standard methods (Adam optimizer, early stopping strategy)

Error evolution during training for an example

3000 parameters, 6 hidden layers, 23 nodes/layer, 3.4x10^6 samples, trained for 13h





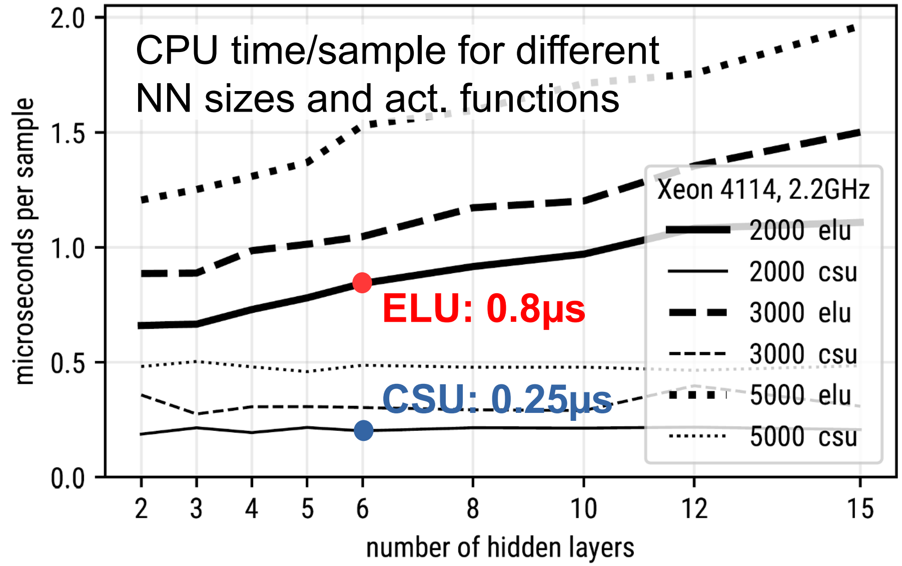
DOM-generated training data can be several 100 times smaller than DOM-generated LUT

Performance

- Development of **Fortran inference code optimized for small NNs** (<100 nodes/layer) (vectorized, much faster than Tensorflow)
- Using a **activation function without exp()** (CSU, piecewise linear/quadratic)
  → inference 3-4 times faster for small NNs

→ Final version for SEVIRI 0.6µm channel 11 x faster than MFASIS-LUT, similar errors

(and MFASIS-LUT is ~200 x faster than DOM)

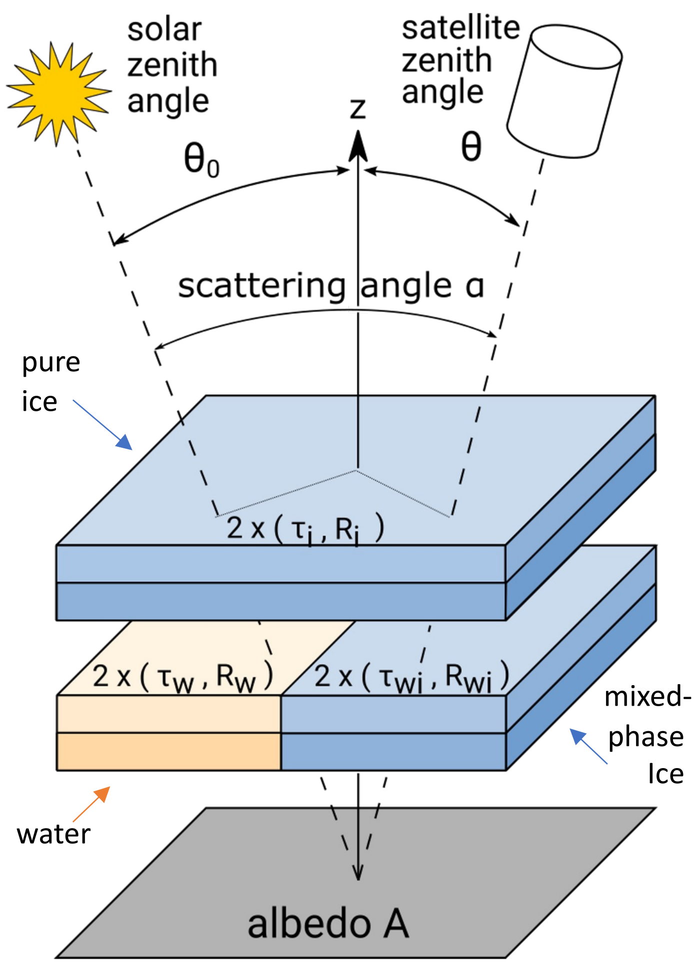


Adjoint / tangent linear codes

- Adjoint (AD) + tangent linear (TL) versions of the nonlinear NN inference code (NL) are required for variational and hybrid DA methods. → AD+TL implemented for Fortran code
- Advantage of neural networks: **AD/TL codes easy to derive, do not have to be modified** when training data or network structure is changed.

For more details see Scheck, L., 2021: *A neural network based forward operator for visible satellite images and its adjoint,* Journal of Quantitative Spectroscopy and Radiative Transfer, DOI:10.1016/j.jqsrt.2021.107841

Additional input parameters for the 1.6μm channel



Interesting for DA & model eval.: 1.6μm can distinguish water from ice, is sensitive to particle size

- Stronger sensitivity to effective radius profiles
  → use **two-layer clouds** to provide information on vertical effective radius gradients
- (Dark) ice in mixed-phase clouds is often below water
  → add a two-layer **mixed-phase ice cloud** in the same location as the water cloud
- Weak absorption by CO2, CH4
  → use **surface pressure and cloud top pressure** as input parameters to quantify influences
- Weak absorption by water vapor
  → use **integrated water vapor** as input parameter

→ **In total 16 input parameters** -- feasible with NNs, 16-dimensional LUT would have been very problematic.

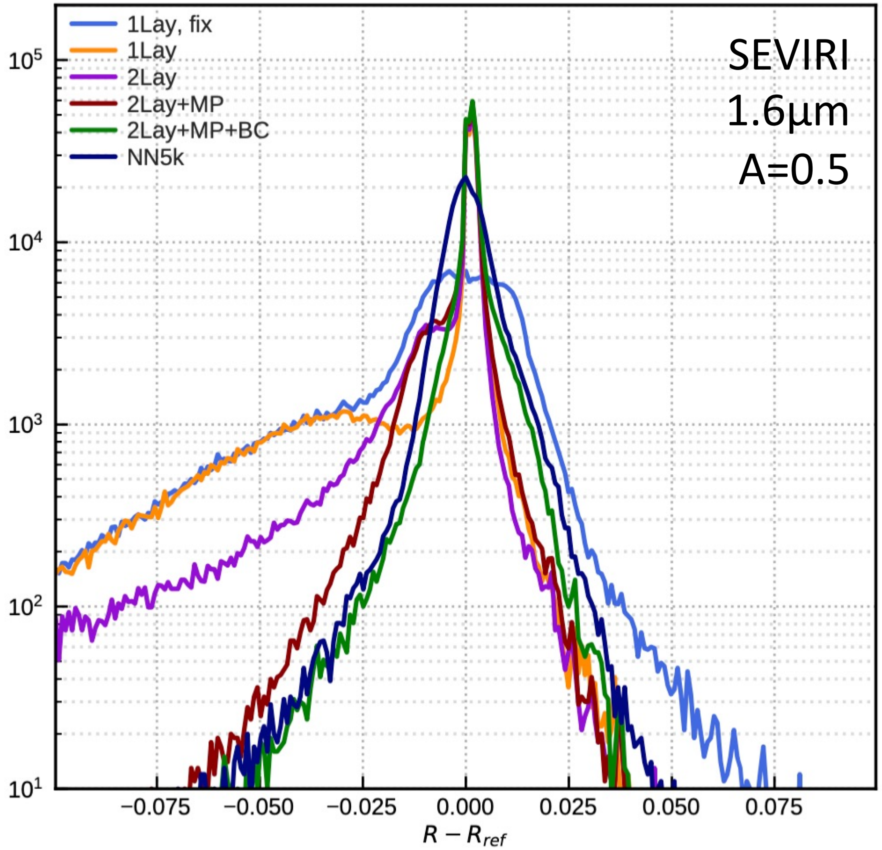
NN learns more complex function → 2.5 times larger NN and 4 times more training data required than for 0.6μm.

Profile simplification and network training errors

Evaluation with IFS profile collection available from NWP SAF

Simplification error: Reflectance error caused by replacing full profiles by idealised profiles (both reflectances computed with DOM reference RT method), $R_{full} - R_{ideal}$

MFASIS-NN error: Simplification error + error caused by imperfect network training



Simplification error for idealised profiles with

(1) One-layer clouds at fixed height (7 param.)

(2) + surface & cloud top pressure, integrated WV (10 param.)

(3) + two-layer clouds (12 param.)

(4) + mixed-phase ice in water cloud (all 16 param.)

(5) + bias correction
→ MAE 0.003, 99th percentile 0.024

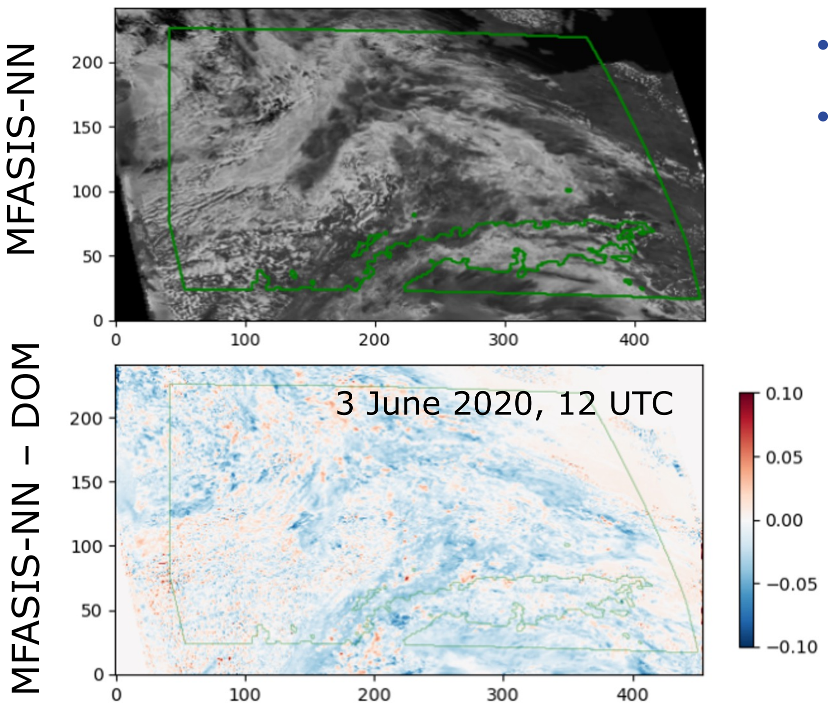
more input parameters

→ all new input parameters reduce errors

MFASIS-NN (training data based on (5)):
MAE 0.004, 99th percentilde 0.027

→ additional training error is relatively small (for a well-trained 5000 parameter NN)

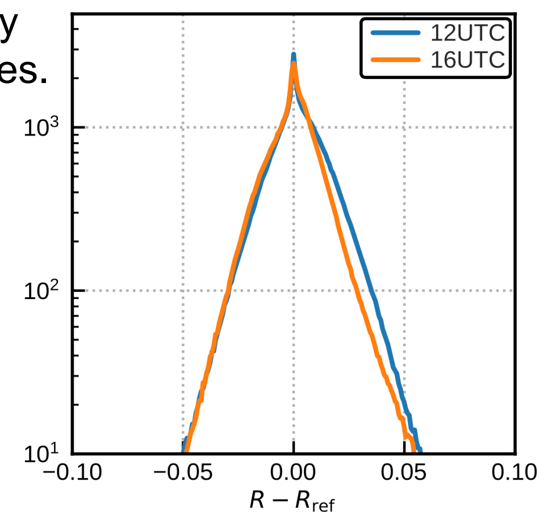
MFASIS-NN SEVIRI 1.6µm results for ICON-D2 forecasts



- Evaluation w. independent, regional model (30 days)
- For IFS profiles parameterizations for effective radii were used, but here **2-moment microphysics scheme** provides prognostic information on radii

	Mean absolute error	99th percentile
5000 IFS profiles	0.010	0.035
ICON-D2 (12UTC)	0.011	0.046
ICON-D2 (16UTC)	0.013	0.056

Statistics are only slightly worse than for IFS profiles.

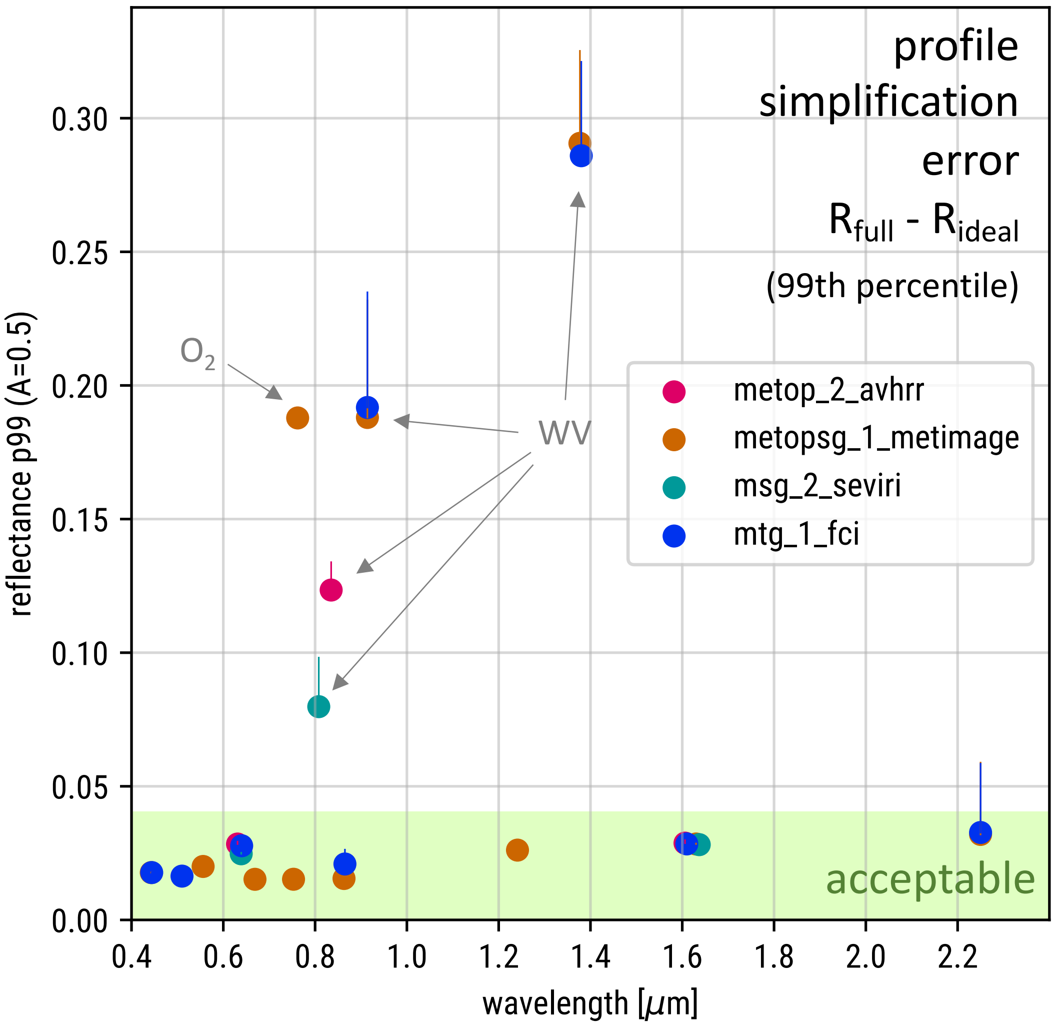


Summary: 1.6µm works -- errors are now similar to 0.6µm errors

For more details see Baur F. et al., 2023: *A neural network-based method for generating synthetic 1.6µm near-infrared satellite images* (submitted to AMT)

Except for WV input variable and vectorization everything is **in RTTOV 13.2.**

Other instruments and channels

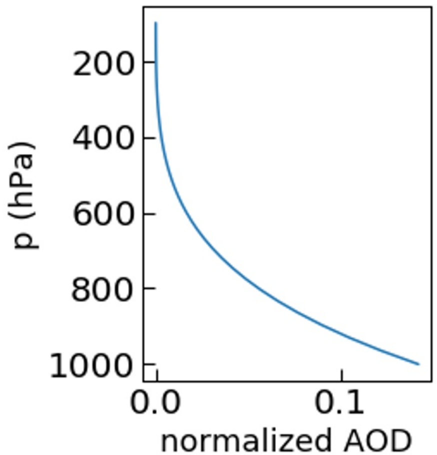


- **Examples:** Solar imager channels on MSG, MTG, MetOp, MetOp-SG
- **λ ≤ 0.7μm channels work well** (enough information on Rayleigh scattering in pressure input variables)
- **Weakly WV-affected channels also work well** (2.2μm, 1.6μm, 1.2μm and most 0.8μm channels)
- **Errors are still too high for stronger WV sensitivity**: 1.3μm, 0.9μm, broad 0.8μm channels (MSG, MetOp)
  → will be solved by additional WV input parameters
- Channels affected by other absorption lines/bands will also need addtitional input parameters
  (e.g. 0.76μm, oxygen A band)

First results for aerosols (preliminary)

A prototype based on CAMS aerosols

- Can we generate reflectances for arbitrary combinations of many aerosols species with one NN and still have sufficiently small errors?
- Same strategy as for clouds: **Replace complex aerosol profile by simplified version** with same AOD, scale height, surface presure and relative humidity



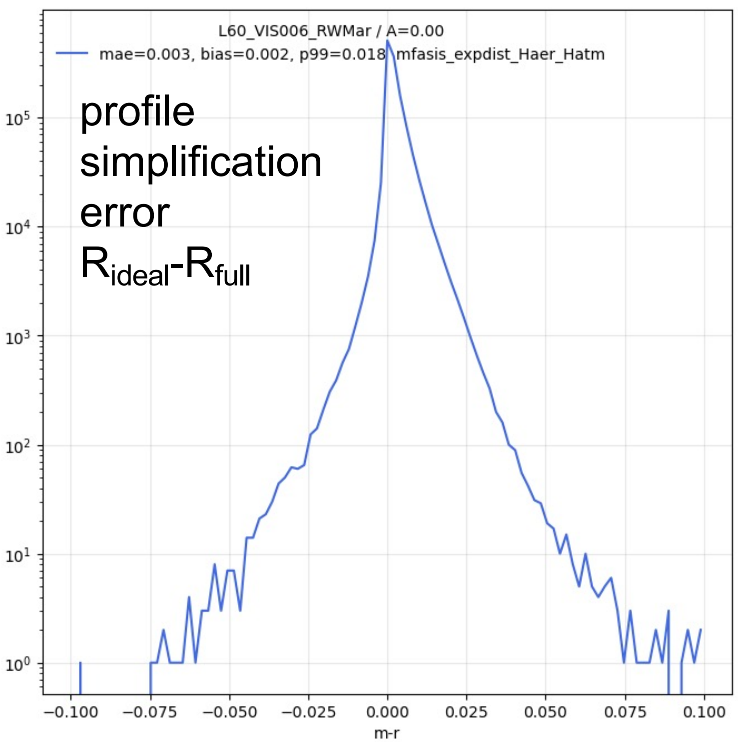
$$\beta_{ext} \propto \left(\frac{p}{p_0}\right)^{H_{atm}/H_{aer}}$$

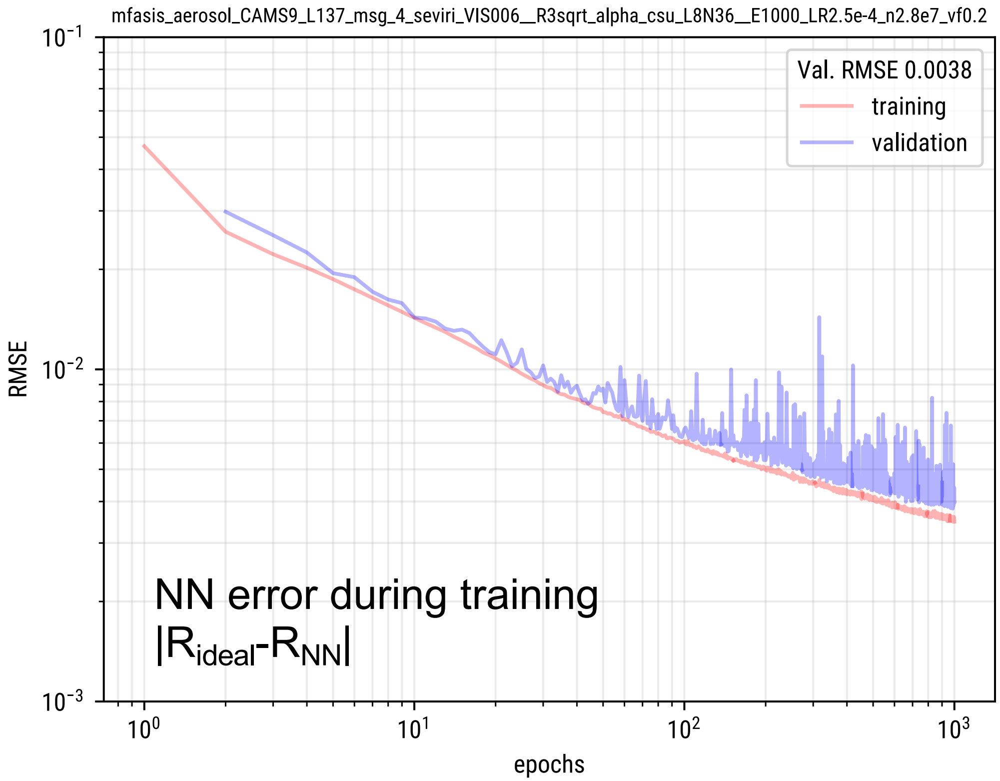
p_0 … Surface pressure
H_{atm} … Scale-height atmosphere
H_{aer} … Scale-height aerosol
β_{ext} … extinction coefficient

- First tests: Assume power law profile for extinction
- Prototype with **23 input variables** (incl. AOD + scale height for 9 CAMS species), NNs with 8 layers, 36 nodes/layer (a bit larger than for clouds, still fast -- <1µs/sample)

First results for aerosols (preliminary)

Evaluation with MACC-60L data set (5 x 4000 IFS profiles optimized for different species):
Profile simplification errors are similar to cloud case (mean abs. reflectance error 0.003)





NN with 10^4 parameters can be trained to error levels similar to cloud case → **promising…**

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN MIM EUMETSAT NWP SAF NUMERICAL WEATHER PREDICTION Hans-Ertel-Zentrum für Wetterforschung Deutscher Wetterdienst Wetter und Klima aus einer Hand DWD

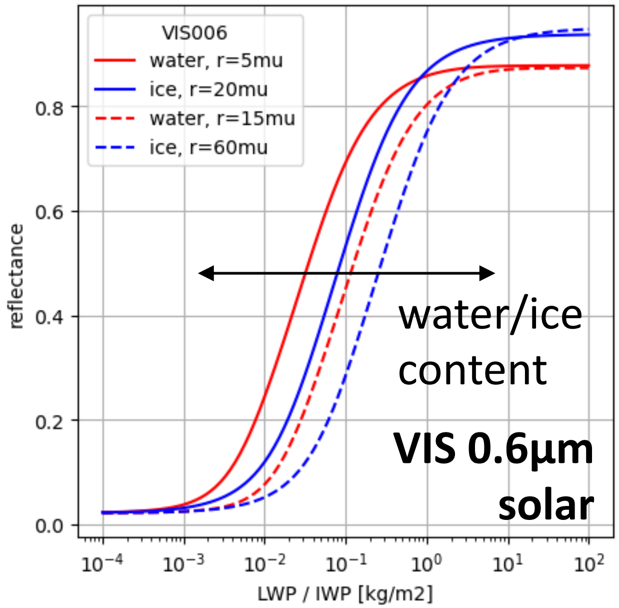
Summary

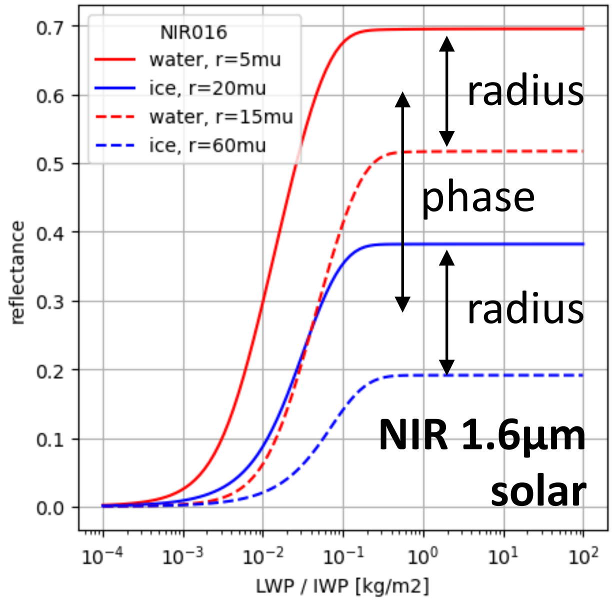
- Replacing a compressed look-up table by a **neural network** in a forward operator for solar satellite channels **has significant advantages** in terms of speed, memory consumption, number of input parameters and time required to generate training data
- **MFASIS-NN for clouds** with additional input parameters **yields good results** for the 1.6µm channels and many other channels. The new developments have been implemented in RTTOV 13.2 (except for WV input variable and code vectorization)
- A prototype of **MFASIS-NN for CAMS aerosols looks promising**
- Research at HErZ/LMU and DWD will continue. In particular, we plan to
  - finish first aerosol version, support further channels (in particular WV sensitive)
  - use feature extraction capabilities of neural networks for more complex channels
  - include 3D radiative transfer effects in neural network

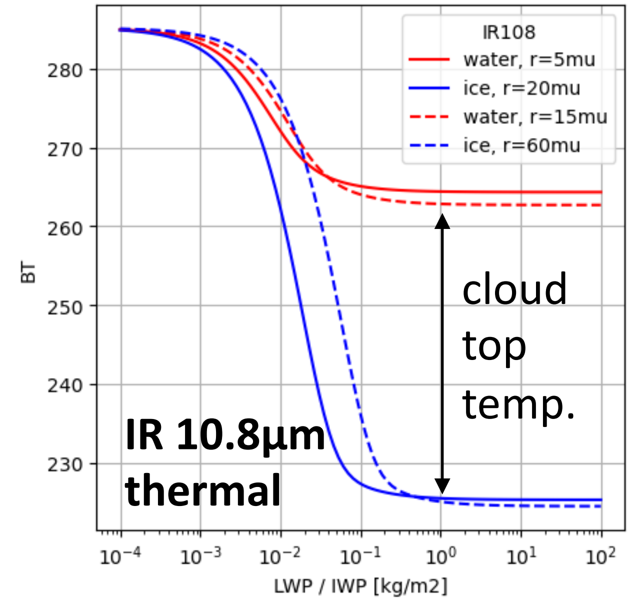
Publications

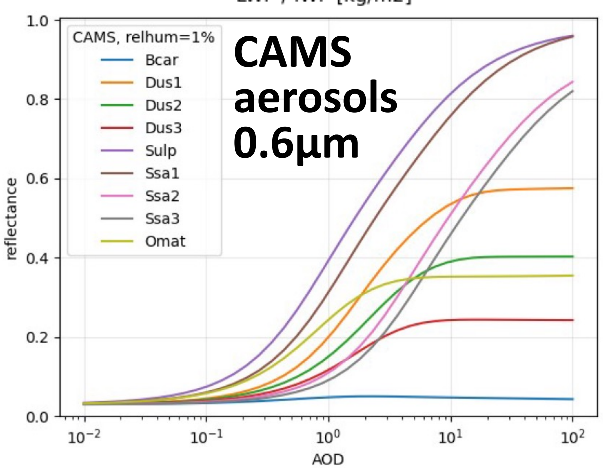
- Baur F. et al., 2023: A neural network-based method for generating synthetic 1.6µm near-infrared satellite images (submitted to AMT)
- Scheck, L., 2021: *A neural network based forward operator for visible satellite images and its adjoint,* Journal of Quantitative Spectroscopy and Radiative Transfer, DOI:10.1016/j.jqsrt.2021.107841
- Scheck, Weissmann, Mayer (2018): *Efficient methods to account for cloud top inclination and cloud overlap in synthetic visible satellite images*, JTECH, Vol. 35, Issue: 3, p. 665-685.
- Scheck, Frerebeau, Buras-Schnell, Mayer (2016): *A fast radiative transfer method for the simulation of visible satellite imagery*, Journal of Quantitative Spectroscopy and Radiative Transfer, 175, p. 54-67.

Why are we interested in solar channels?









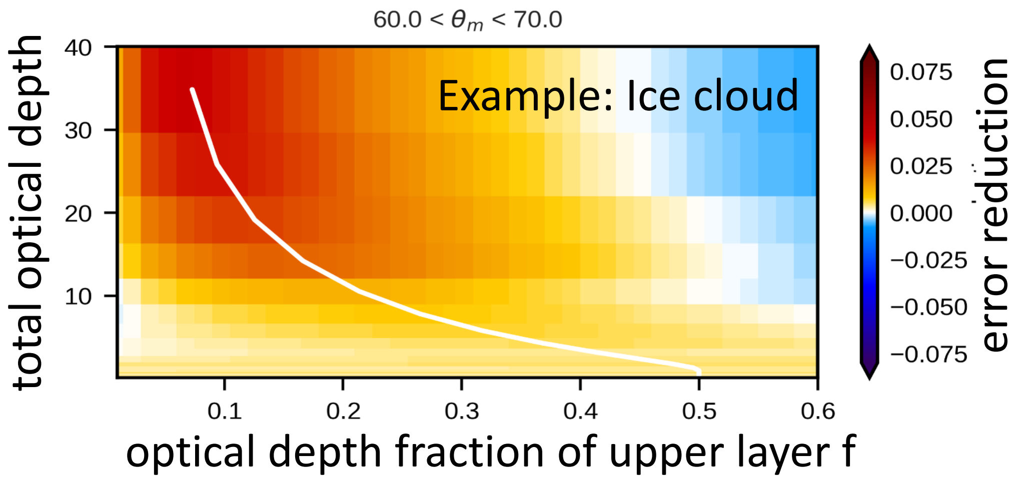
- High-res. cloud information (also for low clouds, complementary to thermal channels) for improving forecasts (including radiation)
- Aerosols (optical properties differ significantly)
- Channels sensitive to water vapor, O2, …

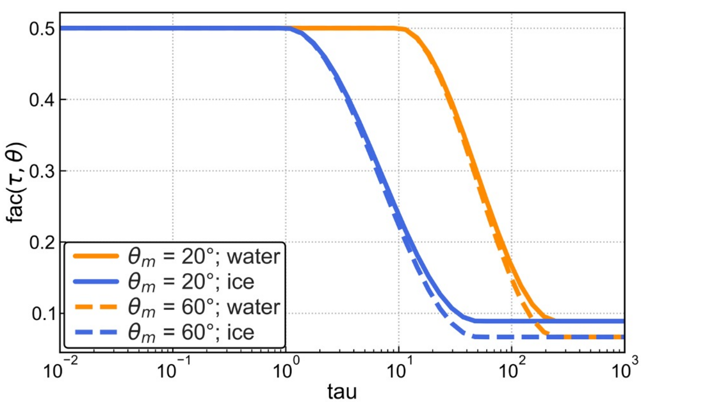
Challenges: Multiple scattering, 3D effects

→ Standard radiative transfer methods too slow for DA

How to split clouds in two layers

- 50% + 50% of optical depth for thin clouds (same probability to be scattered)
- Thicker clouds: e.g. upper 10% + rest, because deeper layers contribute less to reflectance and it is more important to resolve effective radius gradients near the top of the cloud
- Aim: Maximize error reduction, compared to one-layer clouds
- Error reduction computed for collection of 5000 IFS profiles sampled from one year of global IFS short-term forecasts (maximizing diversity in cloud profiles). 64 random angle combinations were used for each profile. Effective radii computed with parameterizations.
  → Parameterization for near-optimal depth fraction of upper layer f

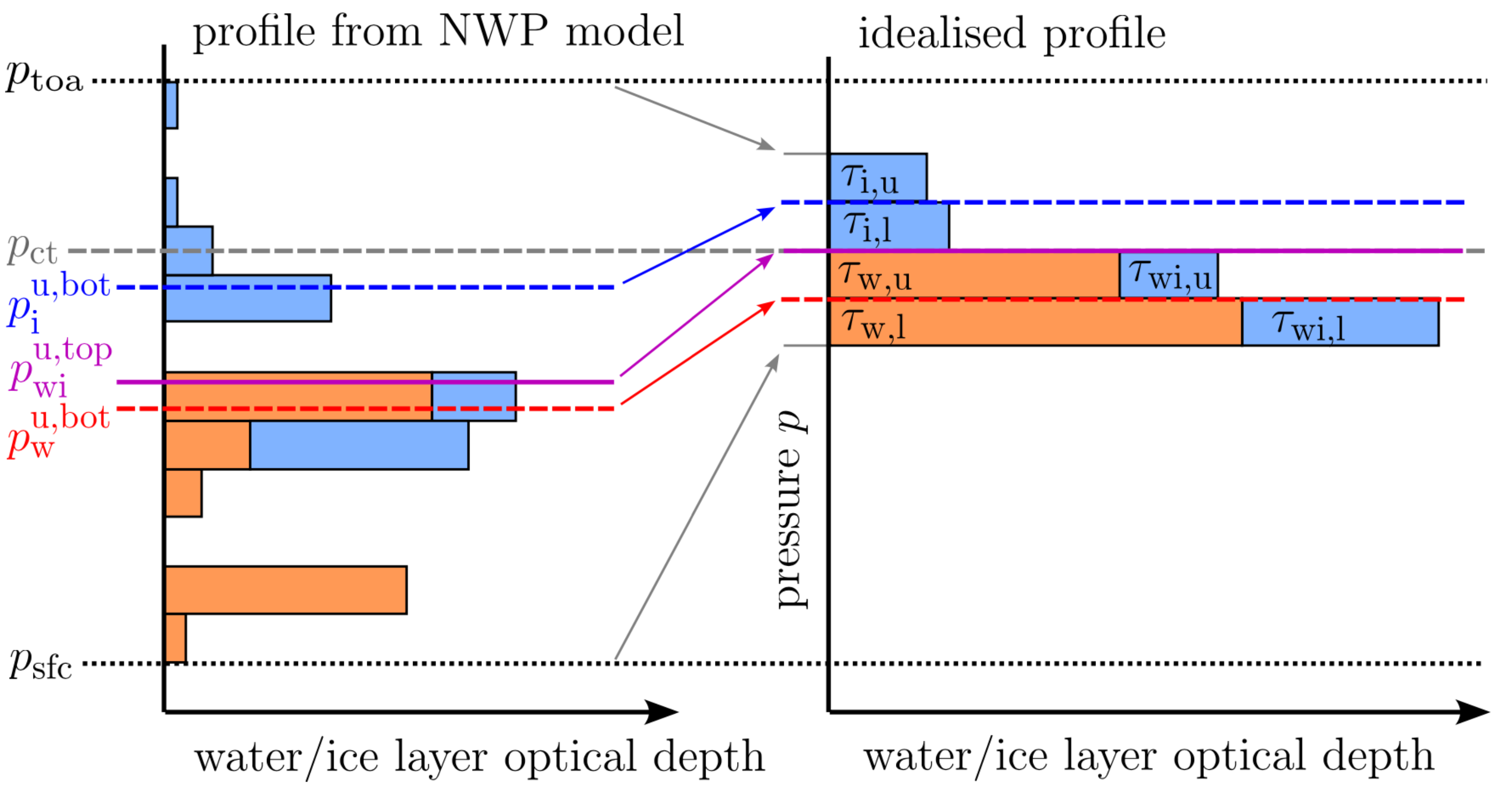




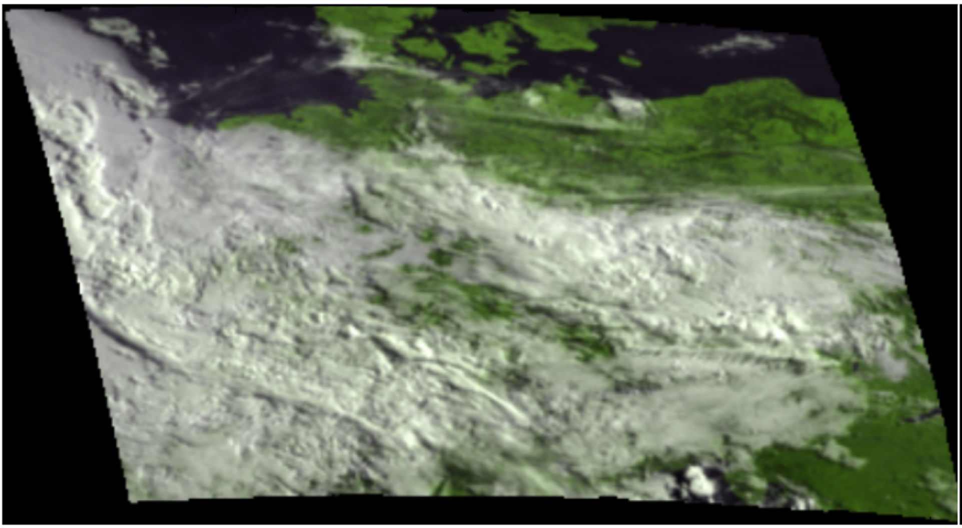
Parameterization of the fraction as function of optical depth, zenith angle

How to compute parameters from NWP profiles

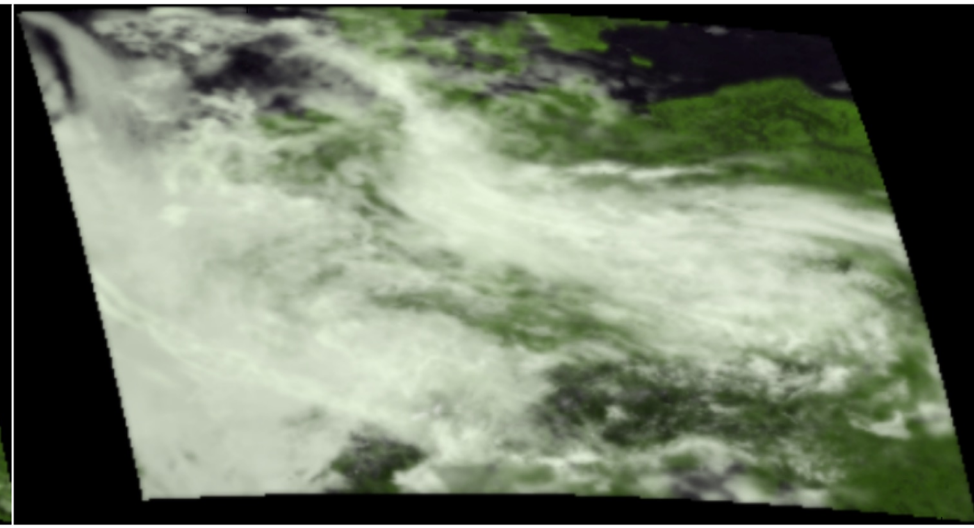
- Cloud top height is where cloud optical depth exceeds 1
- Boundary between water-or-mixed-phase / pure ice cloud determined by water cloud optical depth > 0.5 threshold
- Boundaries between lower / upper cloud layer for water/ice determined by parameterization



Taking cloud top inclination into account (3D RT effect)



SEVIRI 0.6mu+0.8mu, 3 June 2016, 6UTC



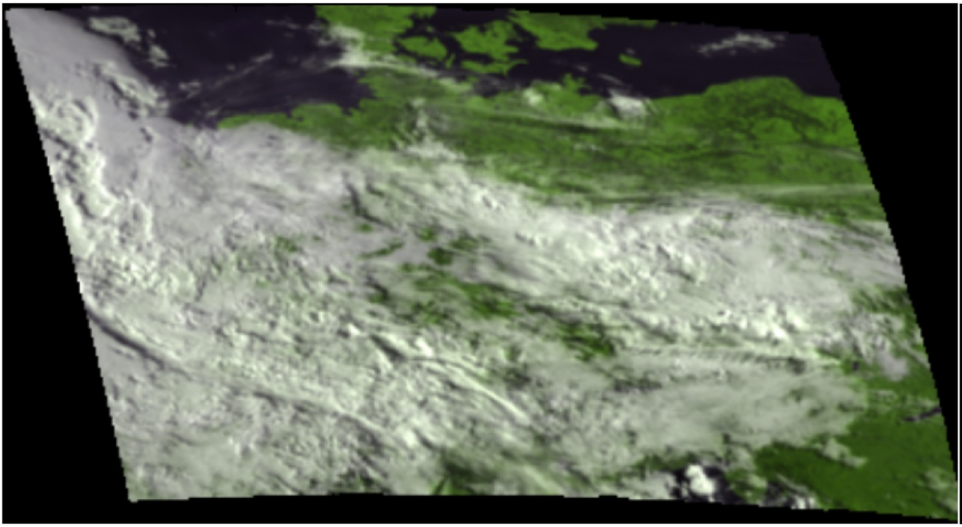
3h COSMO fcst **without 3D correction**

Cloud top tilted away from/towards sun → reflectance lower / higher

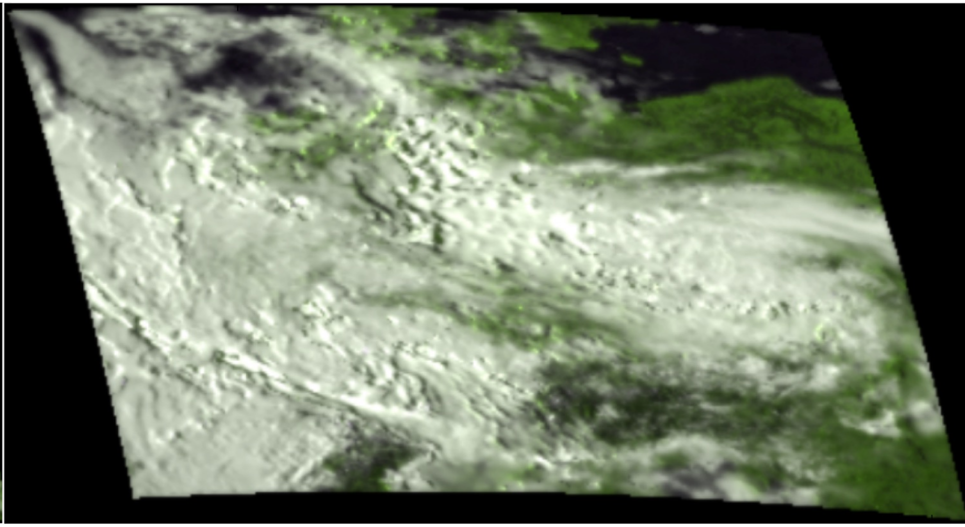
Fast approximation: Find optical depth 1 surface, determine inclination angles
Compute 1D RT solution in rotated frame of reference (in which inclination is zero), transform reflectance back to non-rotated frame

Cloud top inclination correction → Reduced errors, increased information content
Much more cloud structure is visible, in particular for larger SZAs

Taking cloud top inclination into account (3D RT effect)



SEVIRI 0.6mu+0.8mu, 3 June 2016, 6UTC



3h COSMO fcst **with 3D correction**

Cloud top tilted away from/towards sun → reflectance lower / higher

Fast approximation: Find optical depth 1 surface, determine inclination angles
Compute 1D RT solution in rotated frame of reference (in which inclination is zero), transform reflectance back to non-rotated frame

Cloud top inclination correction → Reduced errors, increased information content
Much more cloud structure is visible, in particular for larger SZAs