LMU

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

MIM

5th ECMWF-ESA Machine Learning Workshop

Hans-Ertel-Zentrum für Wetterforschung

Using Distributional Regression Networks to Retrieve Cloud Properties from Solar Satellite Channels for Data Assimilation

Stefano Franzoni[1], Christopher Bülte[2], Leonhard Scheck[1,3], Christian Keil[1], George Craig[1]

(1) Hans-Ertel-Zentrum für Wetterforschung / LMU München, (2) Mathematics Department / LMU München, (3) Deutscher Wetterdienst

State of the Art and Motivation

Solar channels (λ<4μm) observed by satellites commonly include Visible (VIS) and Near-Infrared (NIR) regions, and contain a wealth of information on clouds and aerosols. For this reason, they could be very useful in Data Assimilation (DA), in particular to improve cloud and radiation forecasts.

Current status: only one visible channel currently assimilated operationally (0.6μm SEVIRI, ICON-D2, DWD)

Now many solar channels are available, e.g. from the Flexible Combined Imager (FCI) instrument on Meteosat Third Generation (MTG):

- Most of them are already supported by a forward operator, e.g. in Method for FAst Satellite Image Simulation (MFASIS) available in the RTTOV package v14
- MTG channels available in MFASIS:
  - 4 channels in the visible (VIS): 0.4, 0.5, 0.6, 0.8μm
  - 2 channels in the near-infrared (NIR): 1.6, 2.2μm

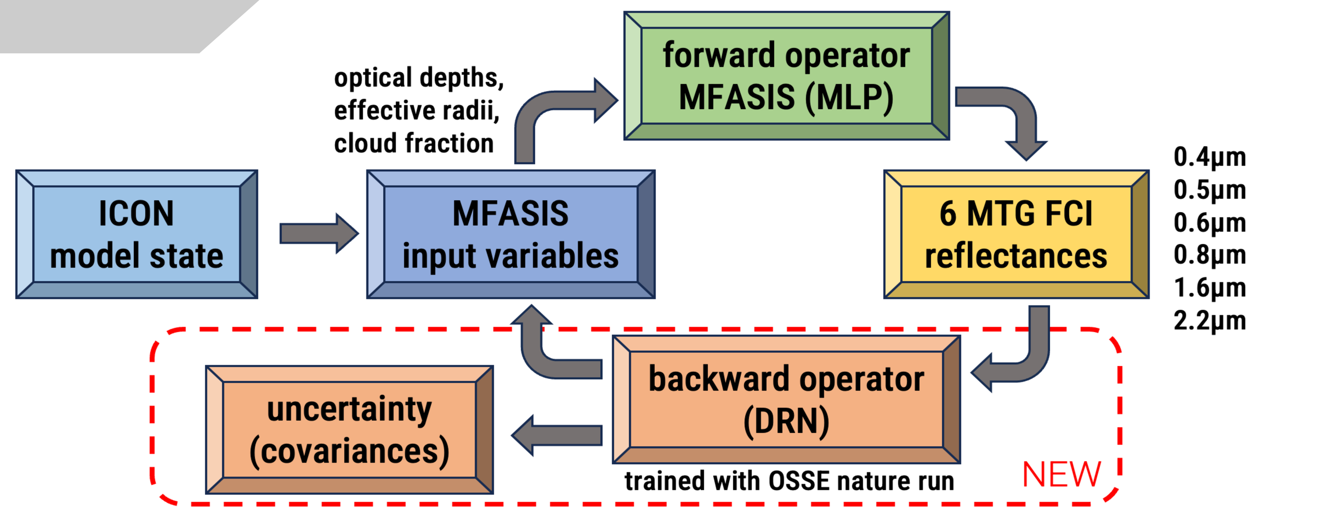
Issues with MFASIS and solar channels:
- non-linearity
- ambiguity: same reflectance from multiple states
- solar channels show very high correlations

Consequences on DA:
- suboptimal analysis
- assimilation of multiple channels may not result in a strong improvement, but it would require considerable effort

Questions:
1. What is the ***joint information content*** of the solar MTG channels, independent of the DA system's limitations?
2. How to ***best exploit it with DA***?

Strategy

Objectives:
1. ***To assess the joint information content*** (independent of the DA system limitations): useful to provide insights into the potential observation impact and to design new instruments
2. Demonstrate that a ***probabilistic retrieval with situation-dependent uncertainties can be designed, which avoids typical problems for DA*** (e.g., need for prior information)



Solution: Backward Operator (BO) with Distributional Regression Networks (DRNs)

	Typical regression methods	Distributional Regression Networks (DRNs)
Loss	RMSE (or similar, e.g., L^p norms)	***proper scoring rule*** (e.g., CRPS, negative log-likelihood, ***energy score***)
Learned features	The mean $E[X_{out} \mid X_{in}]$	(a parametrization of) the ***full predictive conditional distribution*** $P[X_{out} \mid X_{in}]$

Modelling assumption:
$P[X_{out} \mid X_{in}] = N[\mu(X_{in}), \Sigma(X_{in})]$

- Multivariate Gaussian with mean μ and covariance matrix Σ
- ***Covariance matrix provides uncertainties and correlation structure***

Which variables to retrieve?
Simplified set of MFASIS variables: allows for physics-based characterisation of the properties of the column while being more closely related to model state. The size of the set must be comparable to the number of input channels used.

Advantages of the Backward Operator:
- ***No need for prior information***, as the input-output relationship is learned during training based on model climatology
- ***Instrumental error can be accounted for*** by injecting noise on inputs with a prescribed error model
- The ***consistency between BO and the NWP*** model is guaranteed if the forward operator and the NWP model are consistent (which should be the case)
- BO outputs a ***situation-dependent covariance matrix which could be used in the assimilation of BO retrievals*** (along with an estimate of the representativeness error)
- ***Assimilating retrievals could be beneficial:*** lower correlations compared to reflectances in a direct DA approach, and closer relationship to model state variables, ***potentially mitigating both ambiguity and nonlinearity errors***.

Forward Operator

MFASIS (Method for FAst Satellite Image Simulation, Baur et al., available in RTTOV) is a Neural Network (NN) based **forward operator (H) for reflectances**:

- NN computing reflectances from features (MFASIS input variables, see below a subset of them)
- Trained using synthetic data generated with the (much slower) Discrete Ordinates Method (DOM) solver
- MTG channels now available in RTTOV v14
- **Ambiguous and highly non-linear relationship x↔H(x)**

Main model state variables:
- QC (liquid water fraction)
- QI (ice fraction)
- CLC (cloud cover)

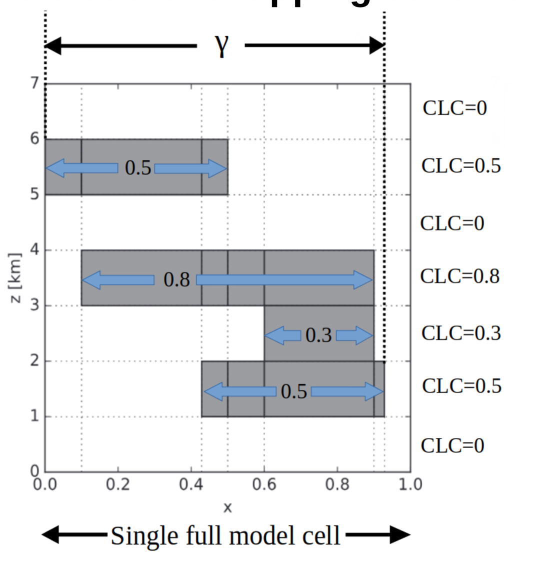
↓ **MFASIS variables**

- τ_w, τ_i (optical thicknesses), $\tau=\tau_w+\tau_i$
- r_{effw}, r_{effi} (effective radii)

Refl = (1-γ)Refl$_{clear}$+γRefl$_{cloudy}$

Using the NN in MFASIS

Column cloud fraction γ from the cloud overlapping scheme



Adapted from Scheck et al. 2018

To Do Next ...

- ***Assimilating retrievals*** from the Backward Operator in a OSSE, using the output covariances as R-matrix and quantifying the representativeness component (Desroziers)
- ***Comparing analysis quality with direct assimilation of reflectances*** using OSSE
- Assessing the ***effect of an imperfect model*** in the OSSE
- Testing the assimilation of retrievals with ***real observations***

Sensitivity to Input Channels

More input channels →

	OnlyNIRs	NIRs-VIS4	NIRs-VIS5	NIRs-VIS6	NIRs-VIS8	NIRs-VIS4,5	NIRs-VIS4,5,6	All
$log(\tau)$	0.544	0.439	0.439	0.441	0.440	0.416	0.405	0.396
γ	0.126	0.120	0.117	0.118	0.117	0.105	0.102	0.100
f_{ice}	0.120	0.112	0.111	0.111	0.105	0.099	0.097	0.093
r_{effw} [μm]	1.372	1.109	1.105	1.116	1.135	1.068	1.021	0.993
r_{effi} [μm]	7.510	7.088	7.012	7.056	6.856	6.754	6.640	6.490

No VIS channel is more informative than the others

Despite the high correlation, more VIS channel use improves performance

- **Above:** RMSE of the output variables. Comparison between a set of models trained using different input reflectances+albedos. The cell shading is row-wise normalised, and proportional to the cell value.

Results

Data and Settings:
- Training dataset based on 4-weeks long ICON-D2 Nature run (covering August-September)
- Energy score as loss function
- Noise added on input reflectances and albedos (Noise):
  - σ_{Refl}=Rad_{Refl}+3%Refl , where Rad is the radiometric noise
  - σ_A=min(0.04, A/3)
- Additional version without noise (No-Noise) for comparison

Backward Operator (BO)

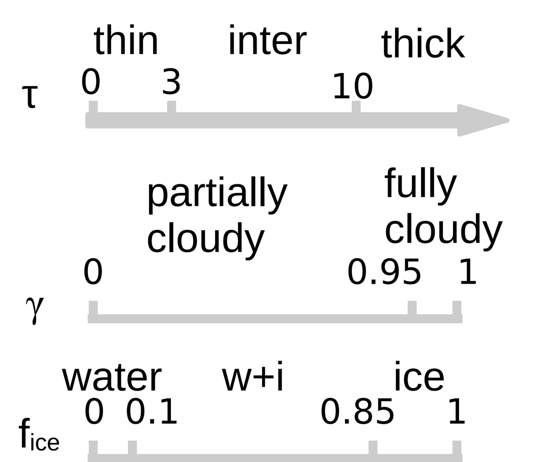
Input:
- 6 MTG reflectances
- 6 albedos
- Zenith and scattering angles

MLP
8 layers,
256 nodes per layer

Output:
- Mean: log(τ), γ, f_{ice}, r_{effw}, r_{effi}
- Full uncertainty covariance matrix

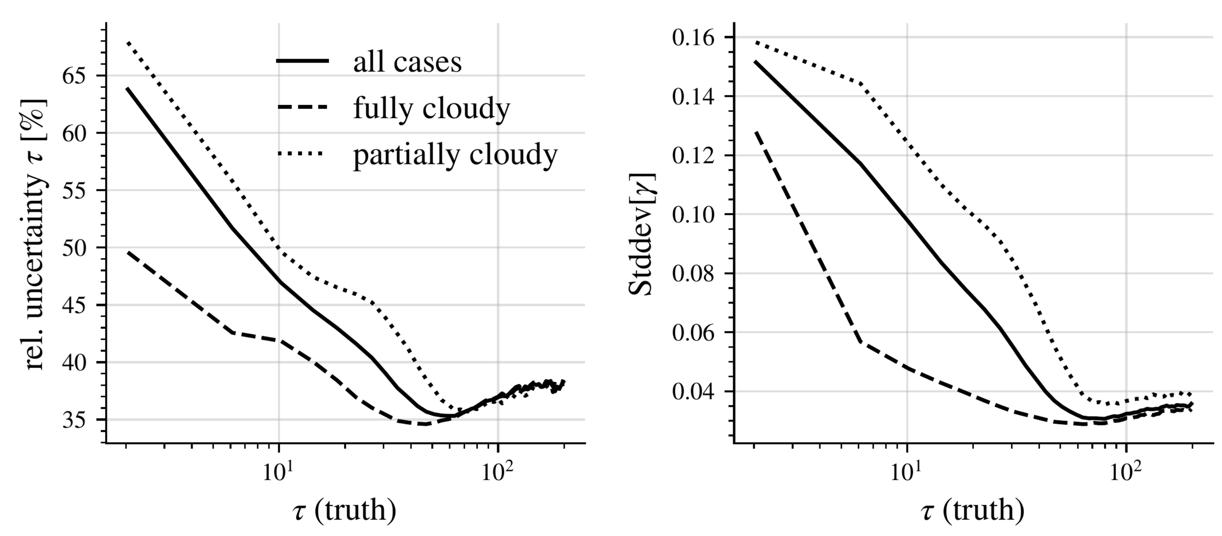
Information Content: Situation Analysis

Definition of Situations



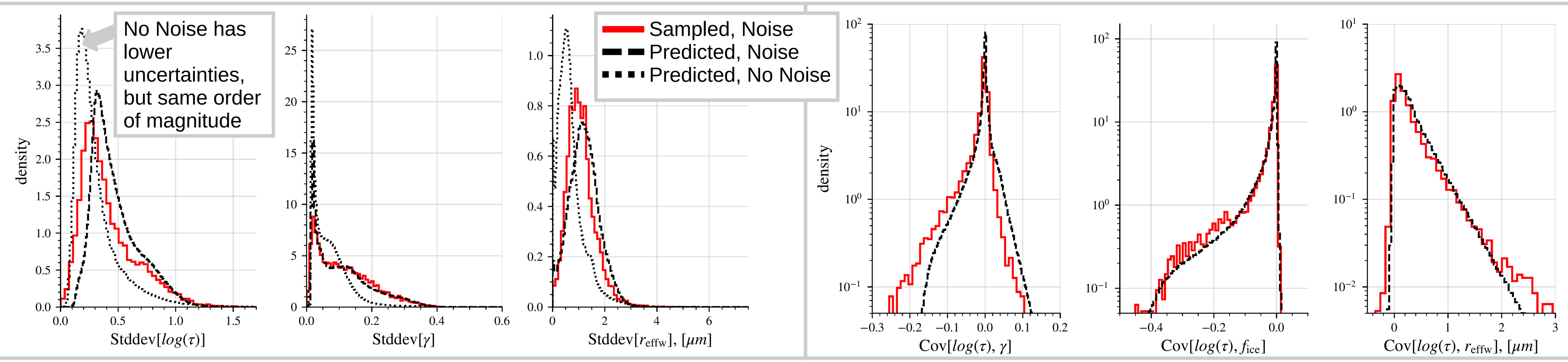
cloudiness	all	fully cloudy					partially cloudy				
thickness	**all**	**thick**		**inter**	**thin**		**thick**	**inter**	**thin**		
phase	**all**	**water**	**w+i**	**w+i**	**w+i**	**ice**	**water**	**water**	**water**	**w+i**	**ice**
rel.err. τ [%]	52.640	37.983	35.102	41.773	55.464	43.748	45.920	58.004	76.952	70.556	48.176
$log(\tau)$	0.475	0.364	0.338	0.392	0.495	0.398	0.428	0.521	0.660	0.611	0.435
γ	0.111	0.041	0.022	0.042	0.127	0.155	0.098	0.151	0.129	0.191	0.206
f_{ice}	0.083	0.030	0.102	0.160	0.219	0.173	0.020	0.036	0.067	0.220	0.146
r_{effw} [μm]	1.213	1.332	1.301	1.330	1.175	1.153	1.481	1.356	0.814	0.988	0.994
r_{effi} [μm]	5.146	5.322	2.399	3.127	4.099	3.027	8.095	6.887	4.043	5.062	2.992

- **Top-right:** predicted uncertainties for the subset of most frequent situations defined above (mean within bin is shown). The cell shading is row-wise normalised, and proportional to the cell value.
- **Right:** predicted relative uncertainty (%) of τ and predicted uncertainty of γ conditioned on the ground truth of τ, distinguishing fully cloudy and partially cloudy cases (mean within bin is shown).



Key Results:
- uncertainties vary significantly
- τ and γ best constrained in thick, fully cloudy cases
- f_{ice} uncertainty tends to be higher in w+i and ice clouds
- effective radii uncertainties tend to be lower in thin clouds (also in thin-partially cloudy cases)

Predicted Covariances: Effect of Noisy Inputs and Calibration



- **Top-Left:** predicted uncertainties for a selection of output variables. Comparison between BO with noise (black dashed), without noise (black dotted) and sampled uncertainties with noise (red).
- **Top-Right:** predicted covariances for a selection of pairs of output variables. Comparison between BO with noise (black dashed) and sampled uncertainties with noise (red).
- **Right:** table of the 1-σ coverage (BO with noise)

Key Results:
- Independent ground truth for the predicted covariances based on sampled conditional distributions from the training dataset
- Good overlap between predicted and sampled (both Noise)
- The input noise is not the main source of uncertainty on the output variables
- Coverage is close to the nominal value (≈0.68)

	1-σ Coverage
$log(\tau)$	0.694
γ	0.721
f_{ice}	0.761
r_{effw}	0.720
r_{effi}	0.745