A Novel Self-Supervised Sun-Induced Fluorescence Retrieval Using Simulated HyPlant and DESIS Data

Jim Buffat[1], Miguel Pato[2], Kevin Alonso[3], Stefan Auer[2], Emiliano Carmona[2]
Stefan Maier[2], Rupert Müller[2], Patrick Rademske[1], Uwe Rascher[1] and Hanno Scharr[4]

[1] Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[2] German Aerospace Center (DLR), Earth Observation Center, Remote Sensing Technology Institute, Oberpfaffenhofen, Germany
[3] RHEA Group c/o European Space Agency (ESA), Largo Galileo Galilei, Frascati 00044, Italy
[4] Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany

JÜLICH Forschungszentrum

HELMHOLTZ AI | ARTIFICIAL INTELLIGENCE COOPERATION UNIT

DLR

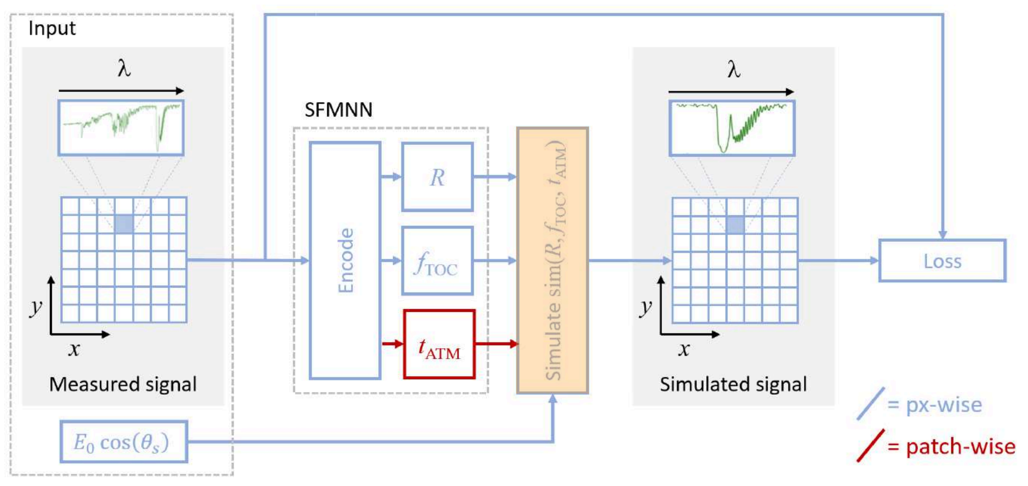
Deep Learning based SIF retrieval

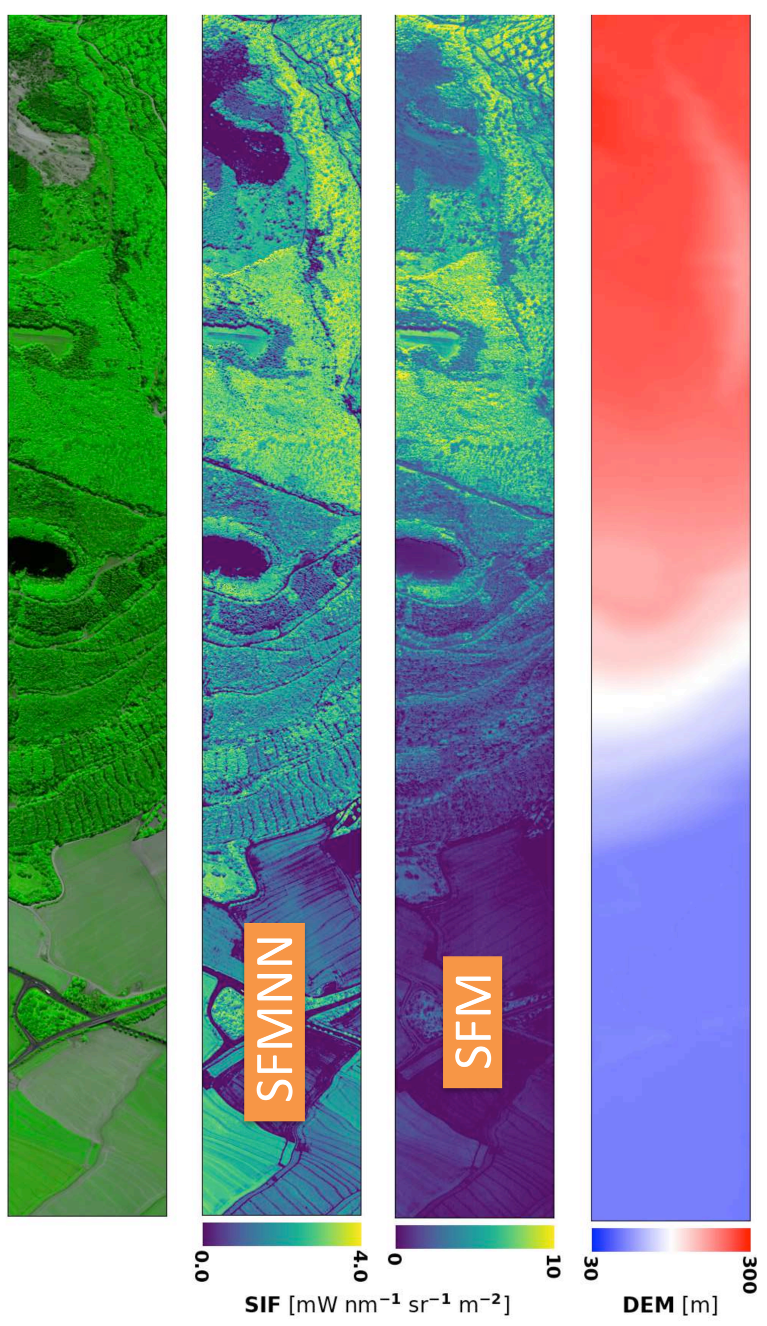
Operationally efficient **retrieval of sun-induced fluorescence (SIF)** from remote sensing data requires **exact atmospheric correction at low computational cost**. Incomplete knowledge of the atmospheric state and surface conditions requires the **formulation of the SIF retrieval as a parameter optimization.** In the present contribution we investigate the use of a neural network to perform this optimization step.

We show-case the possibility to **tightly integrate a neural network with the domain knowledge of radiative transfer codes** simulating observations of the airborne **HyPlant** instrument and the ISS-based **DESIS** spectrometer in a spectral window around the **O_2-A oxygen absorption band** (740-780 nm).

Learning directly from HyPlant acquisitions

- Fully **self-supervised** network training
- **Architectural constraints**: Pixel / patchwise prediction
- **Physically motivated** loss
- Four-stream atmosphere simulation



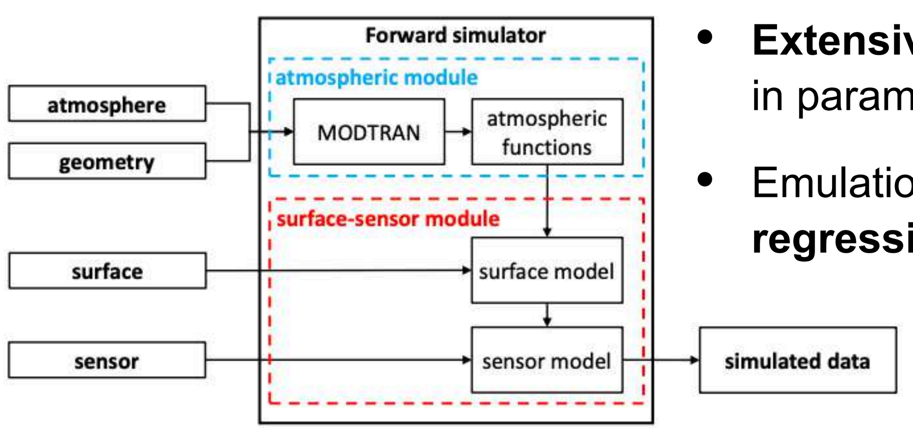


Data Set		r	MAE	N
SEL 2018 (600m)	SFM	0.89	0.80 ± 0.10	11
	SFMNN	**0.96**	**0.68 ± 0.08**	11
	iFLD	0.80	**0.67 ± 0.08**	10
WST 2019 (1500m)	SFM	-0.36*	0.46 ± 0.05	22
	SFMNN	**0.62**	**0.19 ± 0.03**	22
	iFLD	-0.59	4.81 ± 0.09	22
CKA 2020 (350m)	SFM	**0.90**	0.36 ± 0.04	37
	SFMNN	0.87	**0.31 ± 0.04**	37
	iFLD	0.55	**0.28 ± 0.05**	36
CKA 2020 (600m)	SFM	**0.83**	0.42 ± 0.05	23
	SFMNN	**0.83**	**0.24 ± 0.05**	23
	iFLD	0.52	**0.39 ± 0.08**	23

Table 1: FLoX derived SIF measurements compared to SFMNN, SFM and iFLD SIF predictions from HyPlant acquisitions (max. 6 min time difference). Pearson correlation r maked with * have $p > 0.05$. Mean absolute errors (MAE) are given in mW nm^{-1} sr^{-1} m^{-2}.

- Spectral Fitting Neural Network (SFMNN) **outperforms state-of-the-art SFM** (Cogliati et al. 2019)
- Addresses topographic variation by **locally fitting the atmosphere**

Data set creation and emulator



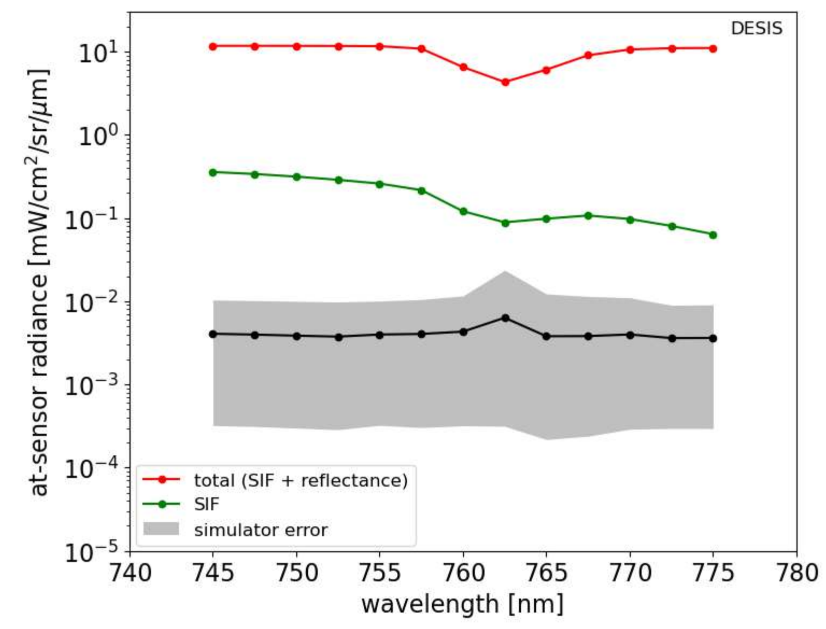
- **Extensive sampling** in parameter space
- Emulation: basic **regression problem**

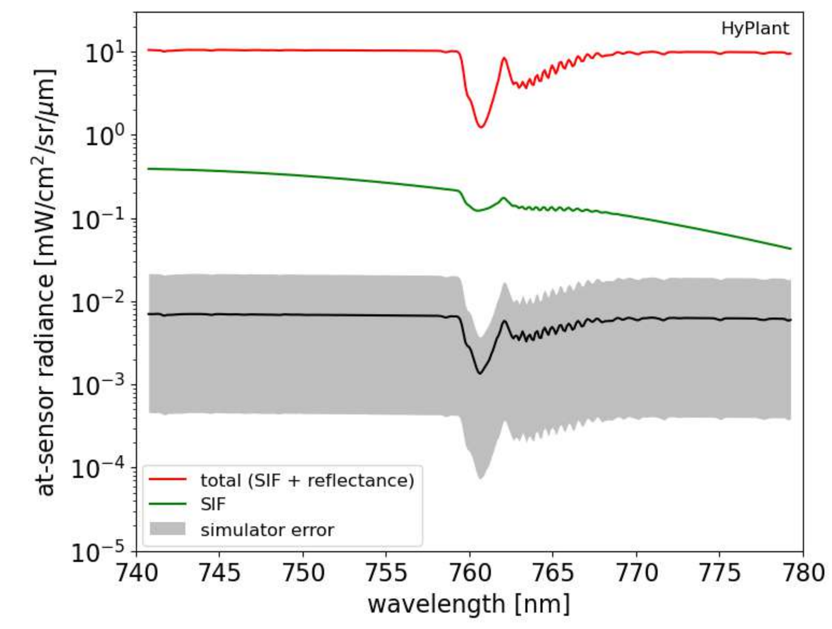
Specification	Databases	
	DESIS	HyPlant
Input dimensions	11	13
Output dimensions	13	349
Number of samples	1.2×10^7	1.5×10^7
Data size [GB]	5.6	64.7

Parameter		HyPlant DB
Atmosphere	model	mid-latitude summer
	H_2O [cm]	0.3–3.0
	O_3 [DU]	332
	AOT_{550} []	0.05–0.40
	aerosol model	rural
	g []	[−1, +1]
Geometry	TA [°]	0–20
	SZA [°]	20–55
	RAA [°]	0–180
	h_{gnd} [m]	0–300
	h_{sen} [km]	0.659–0.691 agl 1.543–1.598 agl
Surface	ρ_{740} []	0.05–0.60
	$d\rho/d\lambda$ [nm^{-1}]	0–0.008
	F_{737}/F_0	0–8
Sensor	δ_λ [nm]	[−0.080, +0.023]
	δ_{FWHM} [nm]	[−0.040, +0.040]

13 parameters

Polynomial 4th degree as emulator is sufficient

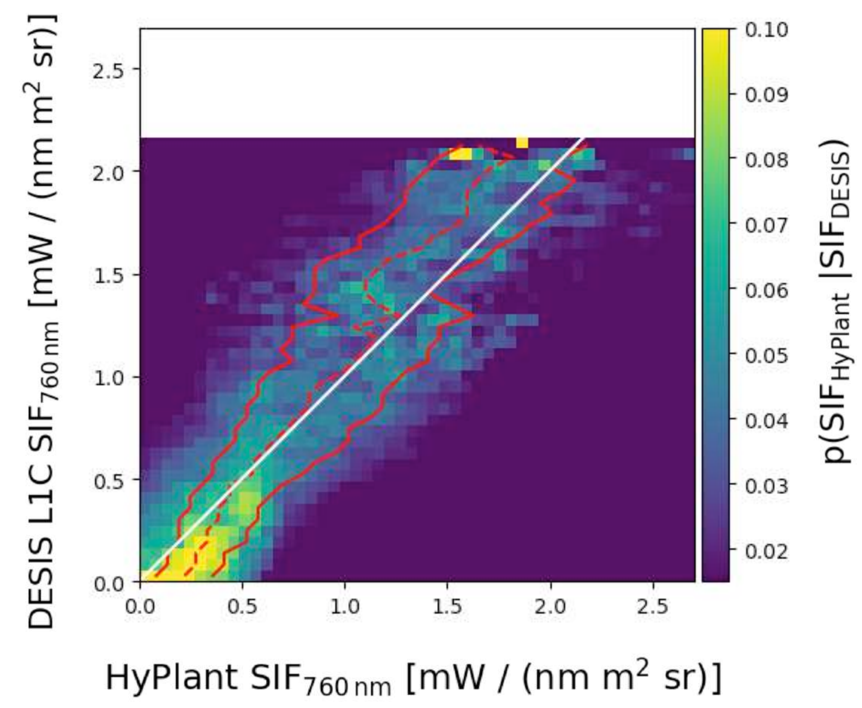




Supervised training on simulated DESIS spectra

- **Calibration** necessary due to L1C processing
- **Multiple matching acquisitions** allow a preliminary validation
- **2023 campaign data** will allow a thorough validation

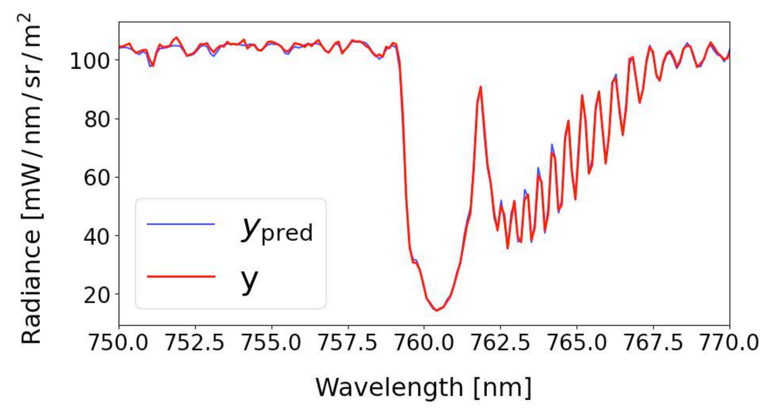
- **Simulation trained network** can be applied to **DESIS** acquisitions
- **No reflectance correlation** due to data generation setup
- Comparison with **quasi-simultaneous HyPlant SFM** product $(\Delta t = 1\mathrm{h})$



Integration of self-supervised approach and emulation

- Supervised training highlights the **simulation data quality** (small domain gap)
- Pixel-wise training limits **decomposition capacity of network**
- **Integrate** SFMNN approach in two steps:
  1. Replace SFM-type simulation by emulator
  2. Semi-supervised training by inclusion of labels

preliminary results



Data Set		r^{pear}	MAE	MAE[calib]	N
			mW nm^{-1}	m^{-2} sr^{-1}	
CKA-2020 (600m)	SFM	**0.85**	**0.43 ± 0.05**	0.17 ± 0.02	18
	SFMNN	0.78	0.90 ± 0.03	0.18 ± 0.04	18
	iFLD	0.53	0.41 ± 0.07	0.24 ± 0.01	18
SEL-2018 (600m)	SFM	0.91	0.53 ± 0.07	0.11 ± 0.00	12
	SFMNN	**0.93**	**0.40 ± 0.03**	0.11 ± 0.00	12
	iFLD	0.82	0.61 ± 0.09	0.18 ± 0.00	12

Conclusions

We reach state-of-the-art SIF prediction performance on HyPlant acquisitions with a deep learning based, self-supervised approach.

A high quality simulation data set could be generated allowing the supervised training of a well performing DESIS SIF predictor.

A tight integration of the emulator with the principles of self-supervised approach derived earlier is subject of further work.

The authors gratefully acknowledge the computing time granted through JARA on the supercomputer JURECA[1] at Forschungszentrum Jülich.

[1] Jülich Supercomputing Centre. (2021). JURECA: Data Centric and Booster Modules implementing the Modular Supercomputing Architecture at Jülich Supercomputing Centre Journal of large-scale research facilities, 7, A182. http://dx.doi.org/10.17815/jlsrf-7-182

j.buffat@fz-juelich.de