Towards a machine learning retrieval of solar-induced fluorescence from DESIS data

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

13th EARSeL Workshop on Imaging Spectroscopy
València, 17.04.2024

[1] German Aerospace Center (DLR), Remote Sensing Technology Institute, Oberpfaffenhofen, Germany
[2] Forschungszentrum Jülich GmbH, Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Jülich, Germany
[3] RHEA Group c/o European Space Agency (ESA), Frascati, Italy
[4] Forschungszentrum Jülich GmbH, Institute of Advanced Simulations, IAS-8: Data Analytics and Machine Learning, Jülich, Germany
* Miguel.FigueiredoVazPato@dlr.de

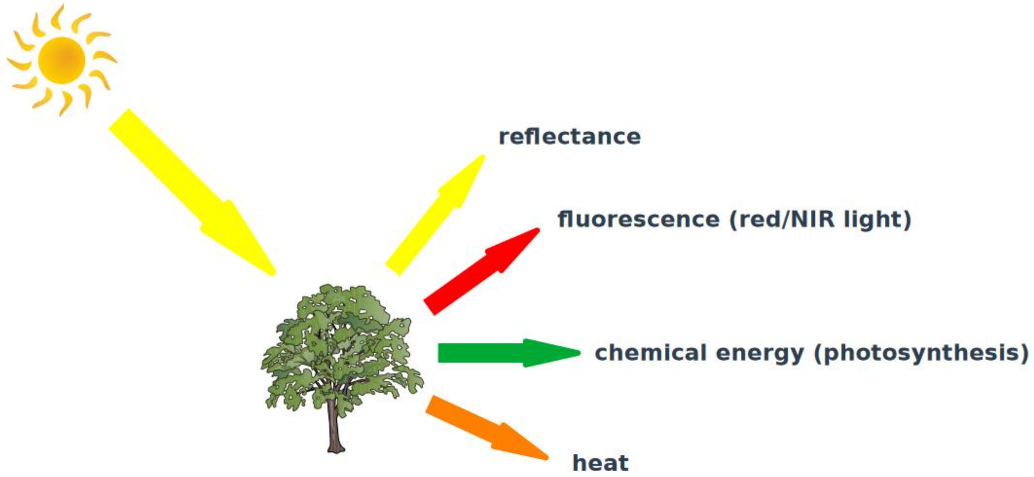
Motivation

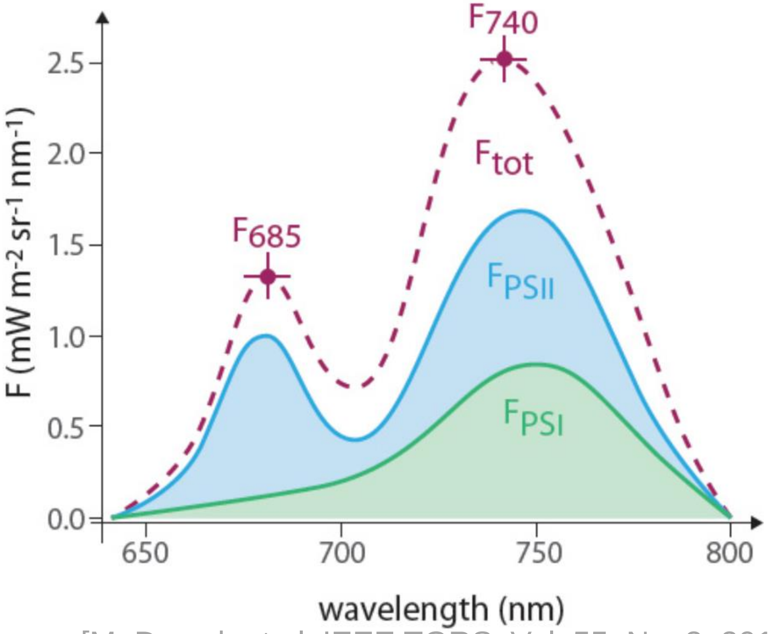
What is solar-induced fluorescence (SIF)?

- During photosynthesis chlorophyll emits fluorescence light at red and near infrared wavelengths.
- This light output is an indicator of photosynthesis efficiency and plant stress.

Remote sensing of SIF is challenging:

- SIF is very much smaller than reflectance signal.
- Fraunhofer lines or absorption features typically used.
- Atmospheric effects need to be corrected for.
- Very strict requirements on detector: high spectral resolution, spectral stability, reasonable SNR, good radiometric accuracy.



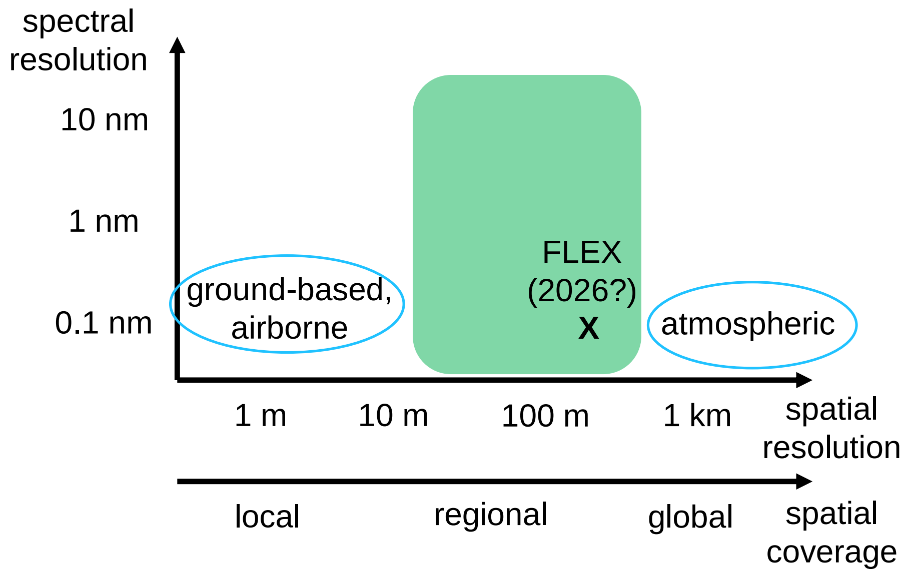


[M. Drusch et al, IEEE TGRS, Vol. 55, No. 3, 2017]

Motivation

Current status of SIF measurements:

- High spectral resolution instruments (ground-based, airborne, space-based) provide either high spatial resolution or large spatial coverage.
- Moderate spectral resolution instruments fill in the gap.



Motivation

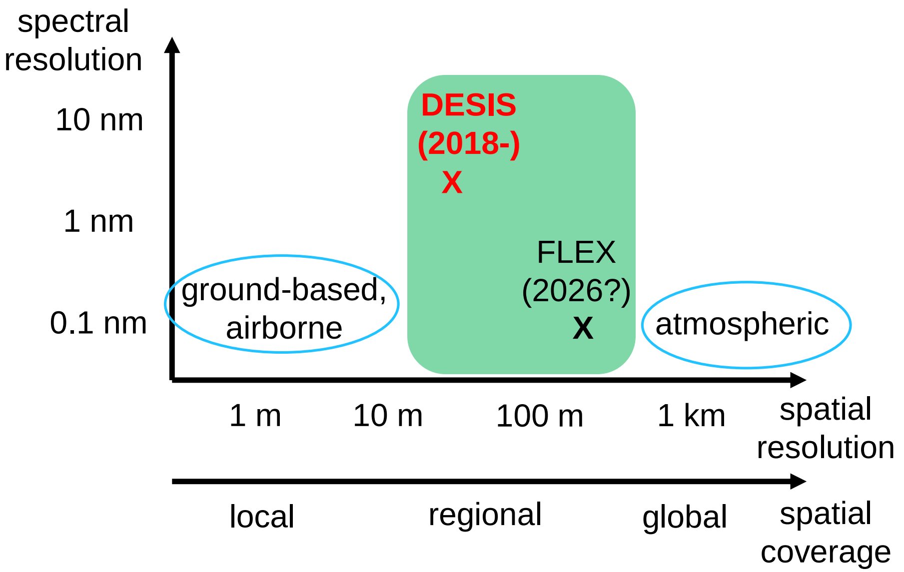
Current status of SIF measurements:

- High spectral resolution instruments (ground-based, airborne, space-based) provide either high spatial resolution <u>or</u> large spatial coverage.
- Moderate spectral resolution instruments fill in the gap.

DESIS advantages for SIF:

- Regional coverage around the globe possible.
- Different hours of day for same site.
- Large archive of data available.

Goal: Machine learning SIF retrieval for DESIS



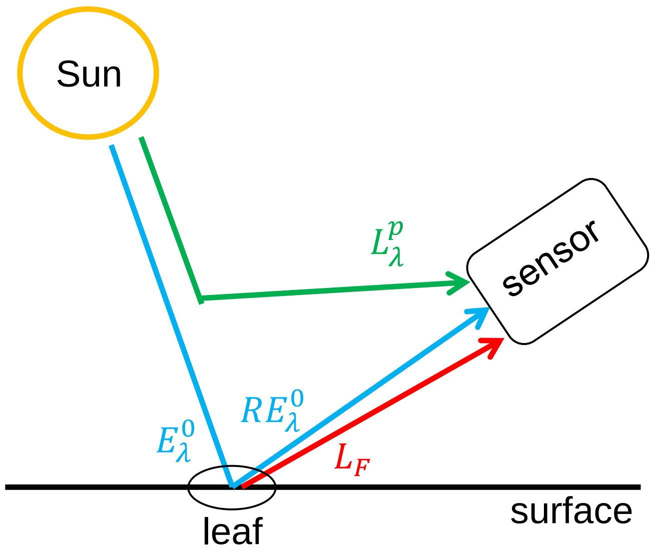
DESIS specification	
Spectral range	420 – 1000 nm
Number of spectral bands	235
Spectral sampling distance	2.5 nm
Spectral full width at half maximum	3.5 nm
Spectral accuracy	0.5 nm
Signal-to-noise ratio	>150
Orbit type, altitude and inclination	ISS, 400 km, 51.64°
Local time and revisit time	variable
Ground sampling distance	30 m
Product size	30 km x 30 km

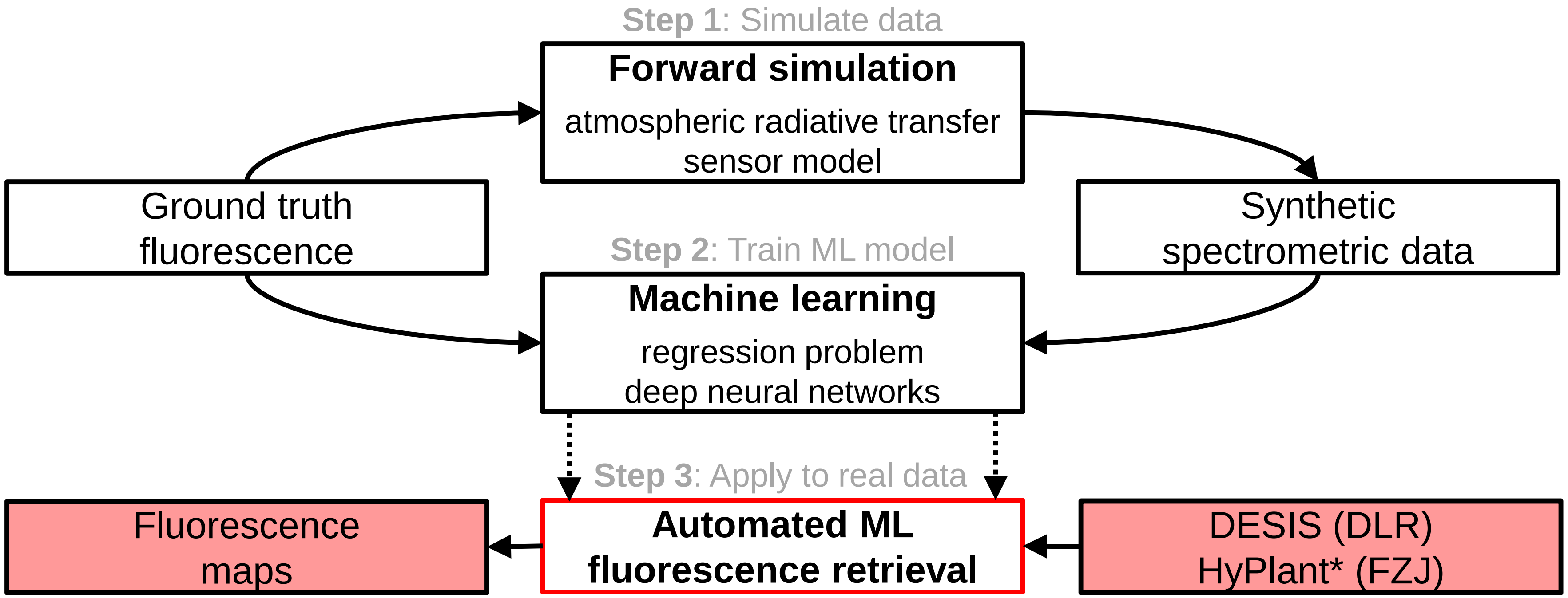
Strategy

fluorescence

$$L_\lambda = \left(\frac{RE^0_\lambda}{\pi} + L_F \right) T_\lambda + L^p_\lambda$$

data reflectance path radiance





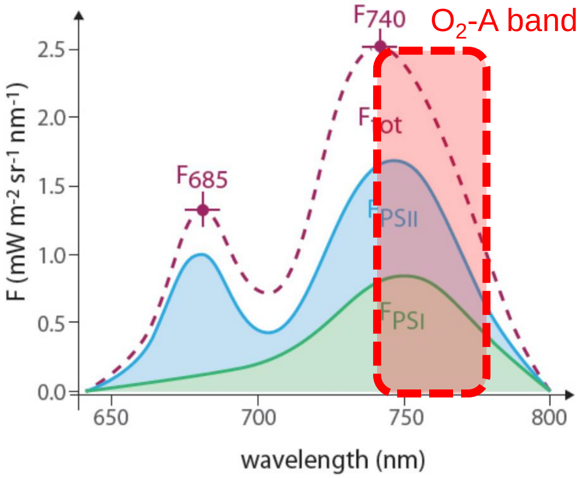
* For HyPlant results, see talk by Jim Buffat on Wed 08:30–10:00 (SIF session).

Step 1: Simulate data

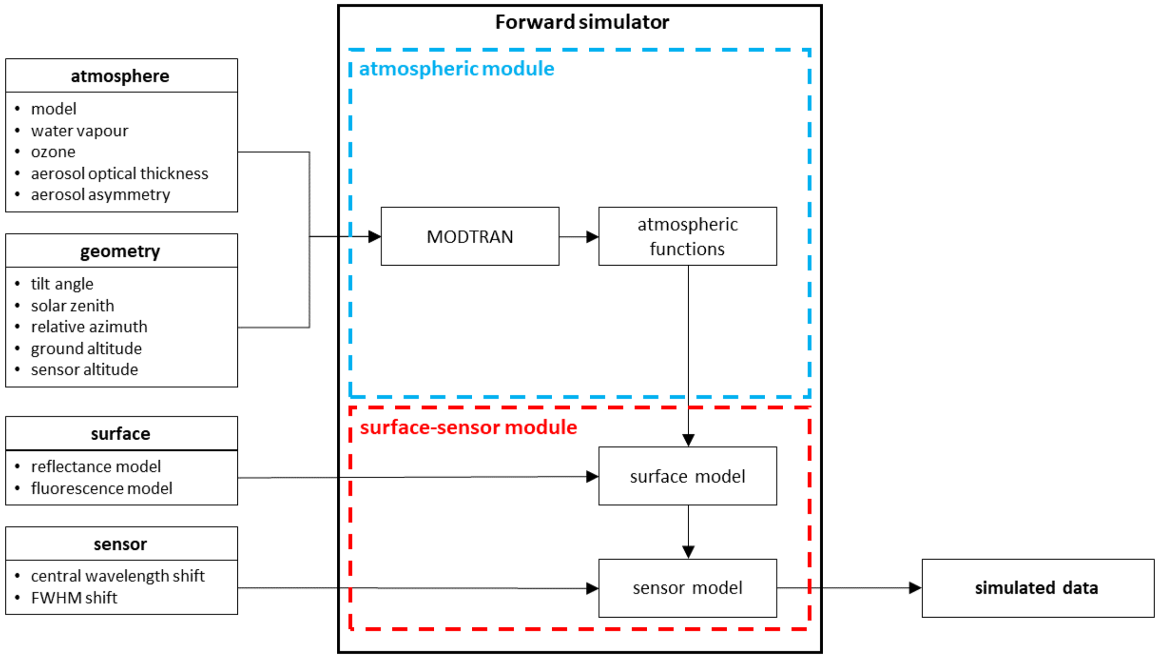
Simulation setup:

- At-sensor radiances around O_2-A band: 740–780 nm.
- Atmosphere+geometry: radiative transfer with MODTRAN6.
- Surface: reflectance and fluorescence parametric models.
- Sensor: based on expert DESIS and HyPlant knowledge.

Note: Other specialized simulation codes exist, but we opted to design a dedicated tool for our needs.



[M. Drusch et al, IEEE TGRS, Vol. 55, No. 3, 2017]

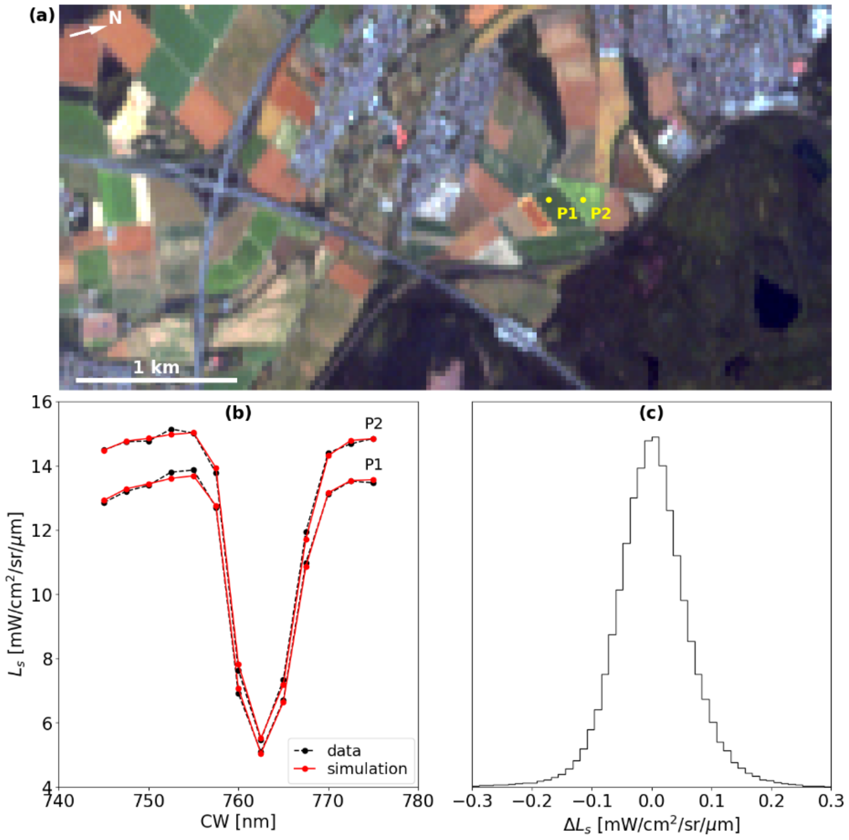


Step 1: Simulate data

Simulated datasets:

- Key parameters and ranges set after sensitivity analysis (see backup slides).
- Hierarchical complexity of datasets.
- Sampling: uniform grid, random, Halton.
- Early comparison to real data.

DB	Parameter		
ATM	Atmosphere	model	mls, trop
		H_2O [cm]	0.3–5.0
		O_3 [DU]	332
		AOT_{550} []	0.02–0.30
		aerosol model	rural
		g []	–
	Geometry	TA [°]	0–25
		SZA [°]	0–55
		RAA [°]	0–180
		h_{gnd} [m]	0–600
		h_{sen} [km]	100
SENSOR	Surface	ρ_{740} []	0.05–0.60
		s [nm^{-1}]	0–0.012
		e []	0–1
		F_{737}/F_0	0–8
	Sensor	δ_{CW} [nm]	[−1.75, +1.25]
		δ_{FWHM} [nm]	[−0.30, +0.30]



Outcome: Highly realistic simulated DESIS data in O_2-A band

Step 1: Simulate data

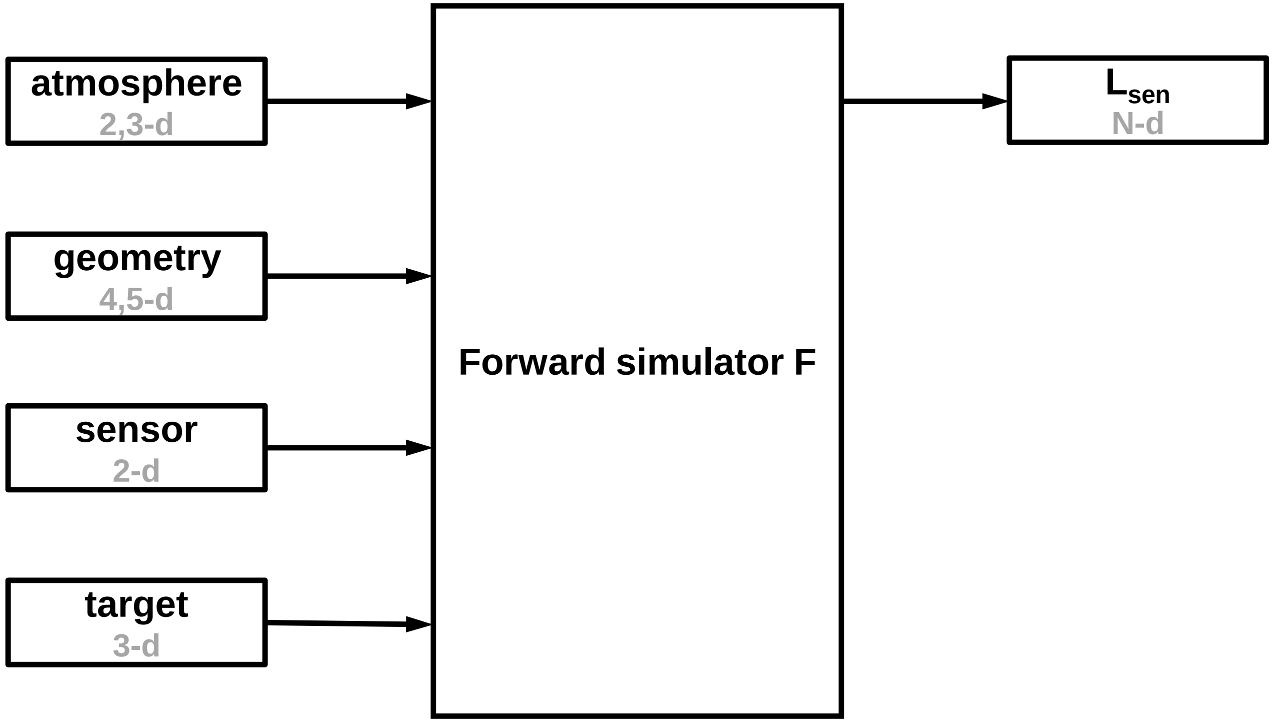
[Pato et al, IGARSS 2023]

Fast machine learning simulator:

- Aim: generalize the slow physics-based forward simulator with a fast ML-based model.
- Framework:

  $x = [\mathrm{atm}, \mathrm{geo}, \mathrm{sen}, \mathrm{tar}] \qquad L_{\mathrm{sen}} = F(x)$

  learn forward simulator $\hat{F} \approx F$
- Input: DESIS/HyPlant simulated data.
- Output: trained ML forward simulator.



N=11 (DESIS)
N~300 (HyPlant)

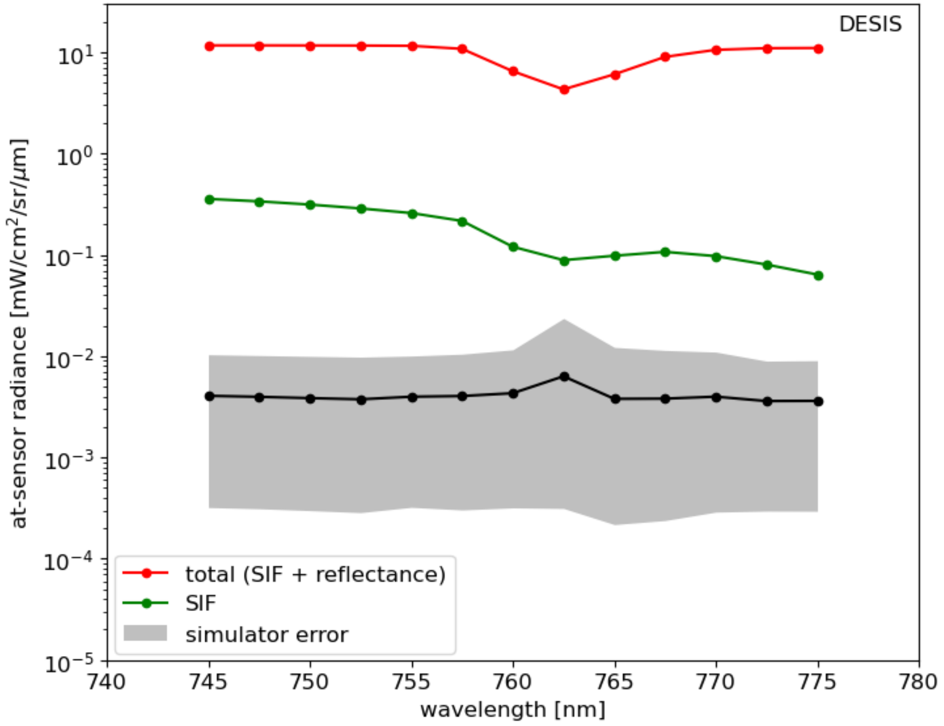
Step 1: Simulate data

[Pato et al, IGARSS 2023]

Fast machine learning simulator:

- Simple ML models are adequate to emulate the full-fledged simulation in the case considered.
- Polynomials of 4th degree are both fast and accurate.
- Speed: 10^7 faster than the simulation.
- Accuracy: 10 times below SIF signal.
- There is room for improvement.

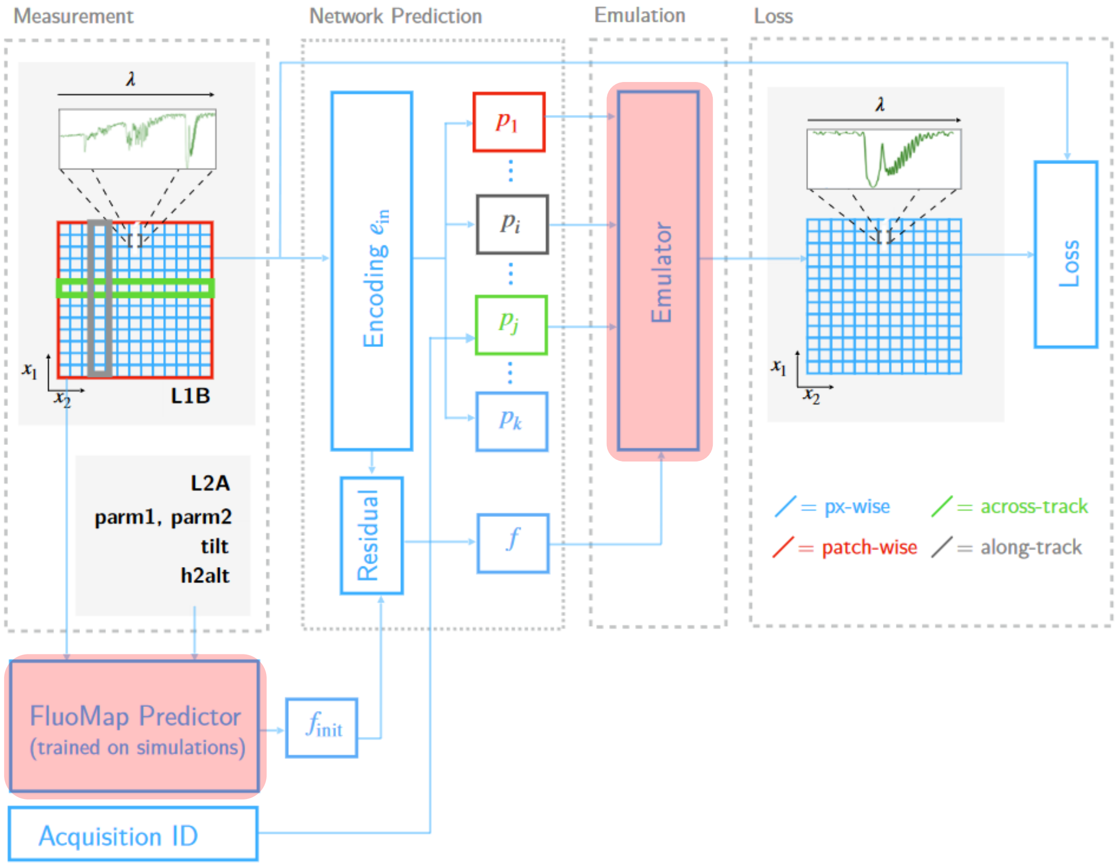
Performance parameter	DESIS			
	OLS	P2	P4	NN
Test set MAE [mW/cm^2/sr/μm]	0.65	0.13	0.0041	0.017
Total training time	1.6 s	14 s	1.4 min	1.7 h
Prediction time per sample [μs]	0.04	0.9	11	31



Step 2: Train ML model

DESIS SIF ML model:

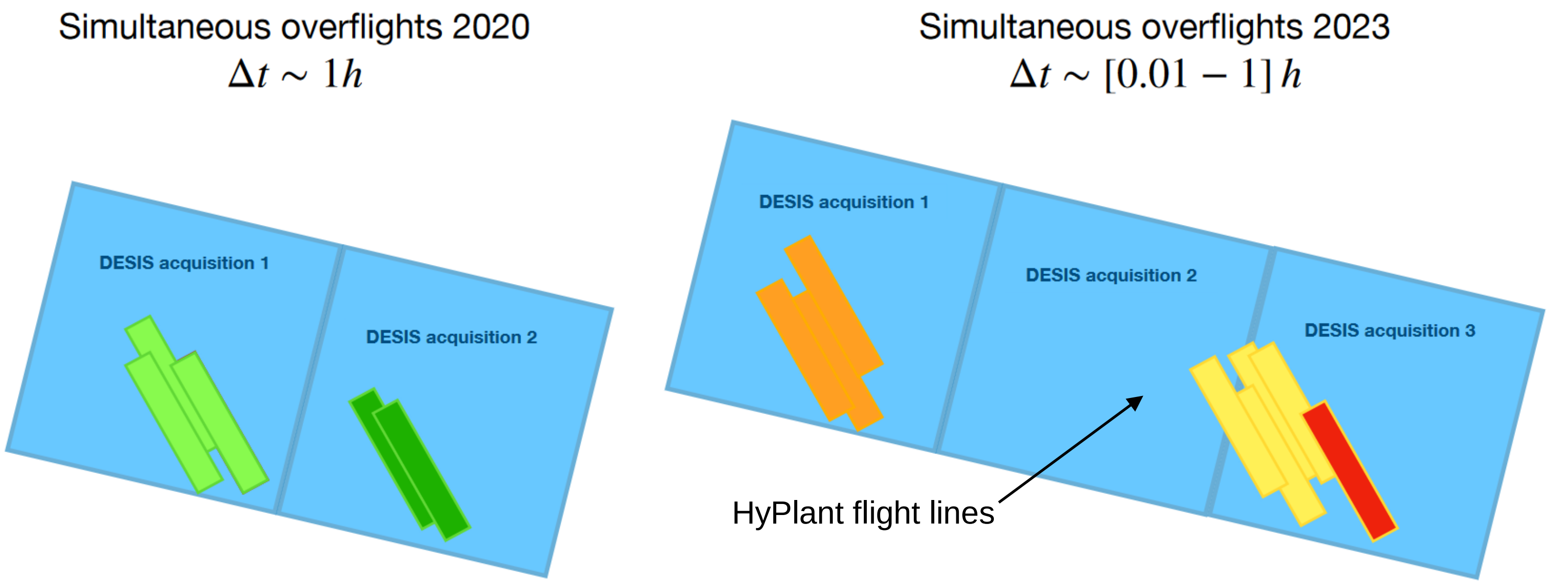
- Idea: self-supervised scheme initialised by supervised predictor.
- Inputs from simulation: simulated datasets, fast ML simulator (see red blocks).
- Differentiated treatment of parameters:
  - Pixel: reflectance, fluorescence
  - Patch: atmosphere
  - Across-track: sensor
- Encoder: multi-layer perceptrons with residual links.
- Loss: least squares, O_2A boost, physiological constraint.

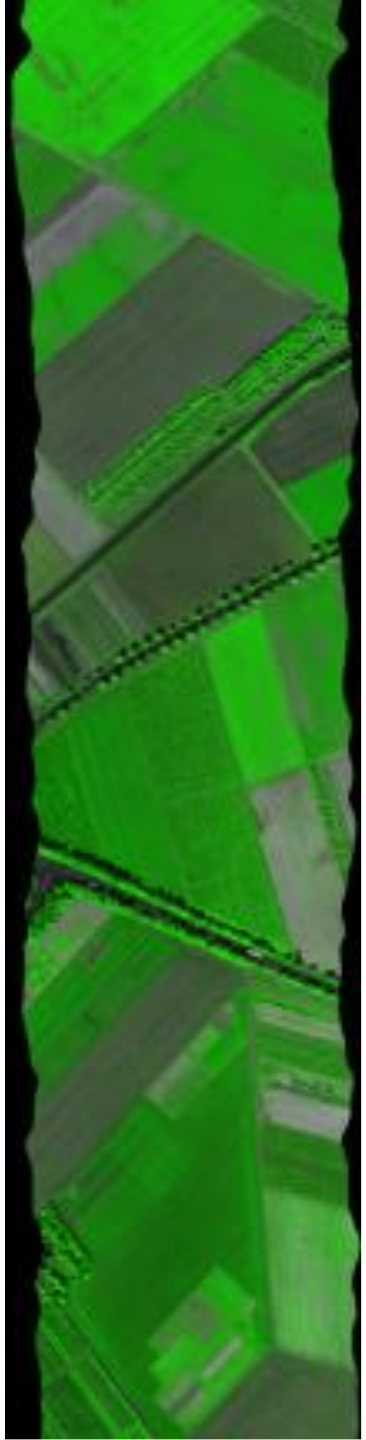


Step 3: Apply to real data

Dataset:

- SW of Cologne, Germany in Jun 2020 and Jul 2023.
- Coincident DESIS/HyPlant overflights (Δt: minutes to 1 hour).
- HyPlant SIF estimates obtained with spectral fitting method (SFM) serve as validation for our DESIS SIF predictions.

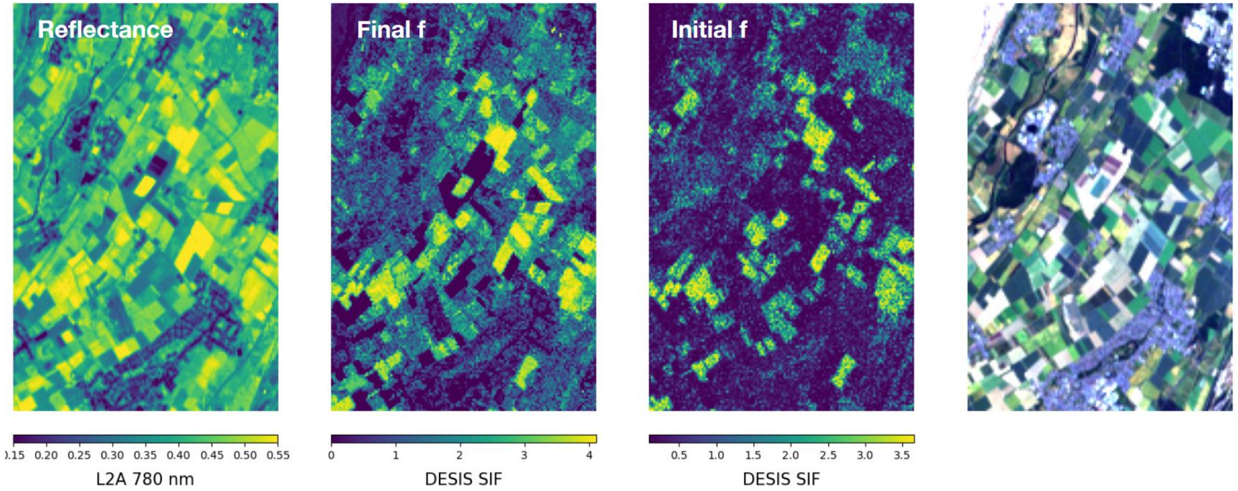


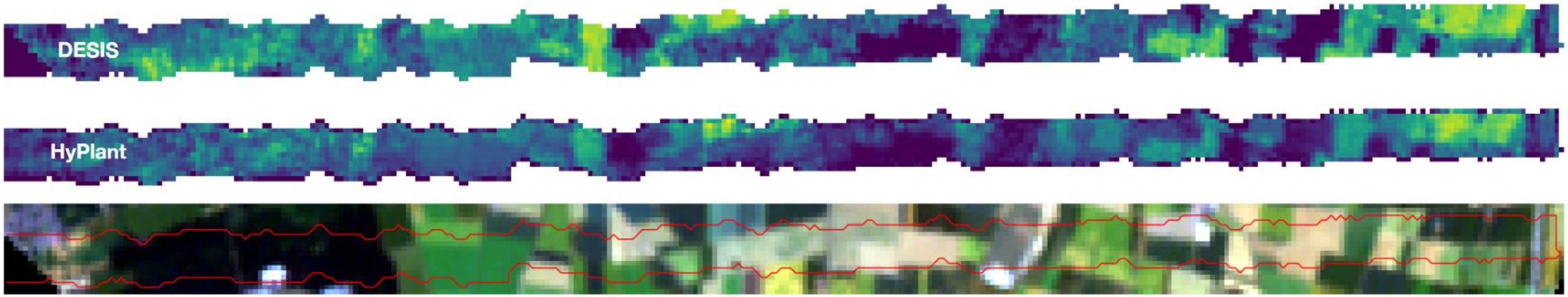


Step 3: Apply to real data

Executive summary:

- Final SIF prediction is much improved over initial (supervised) prediction.
- Results are qualitatively plausible and correlate well with vegetation fields.
- Model is not learning a correlation between reflectance and fluorescence, but fluorescence.
- Spatial patterns of HyPlant SFM SIF reproduced.

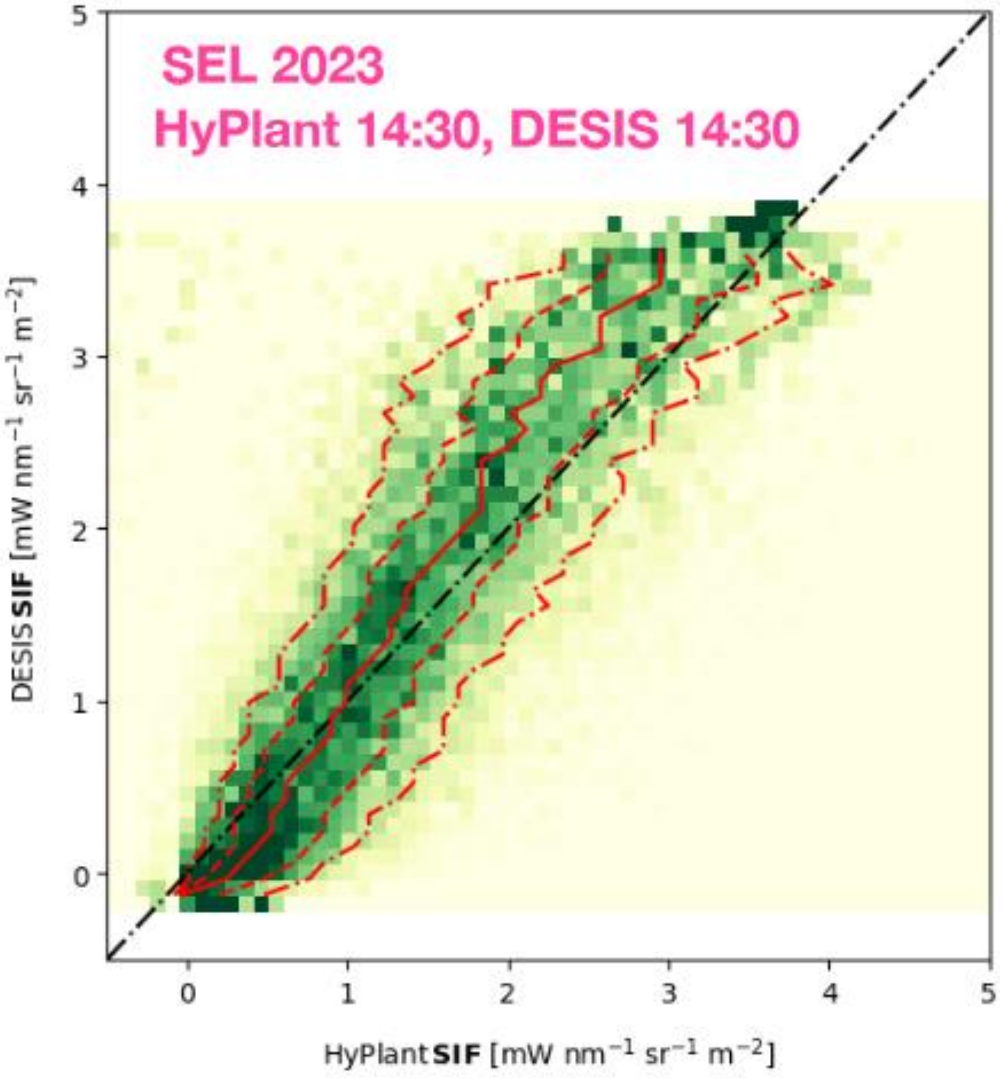




Step 3: Apply to real data

Executive summary:

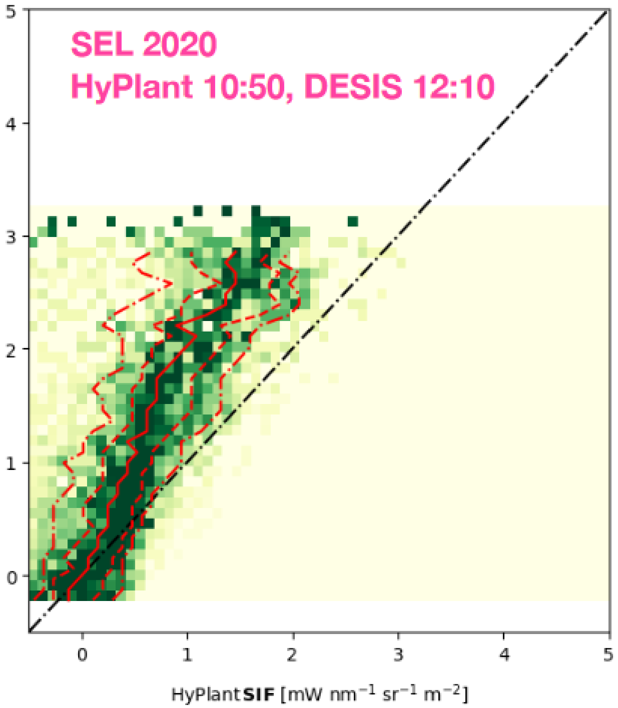
- DESIS SIF predictions match very well HyPlant SFM SIF estimates for a coincident acquisition.
- There appears to be a slight tendency for our model to overpredict SIF wrt HyPlant SFM SIF.
- **Note: HyPlant SFM provides a completely independent check of our results (different method, detector and spatial resolution).**

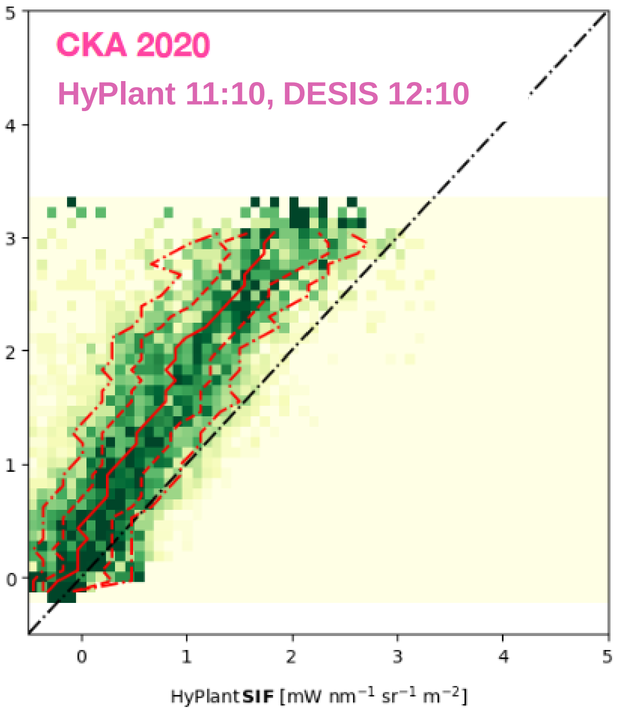


Step 3: Apply to real data

Executive summary:

- DESIS SIF predictions match very well HyPlant SFM SIF estimates for a coincident acquisition.
- There appears to be a slight tendency for our model to overpredict SIF wrt HyPlant SFM SIF.
- **Note: HyPlant SFM provides a completely independent check of our results (different method, detector and spatial resolution).**
- DESIS SIF larger than HyPlant SFM SIF later in the morning (as expected physiologically).
- Comparison with in-situ data is needed for actual validation.
- **It is possible to retrieve SIF with DESIS.**

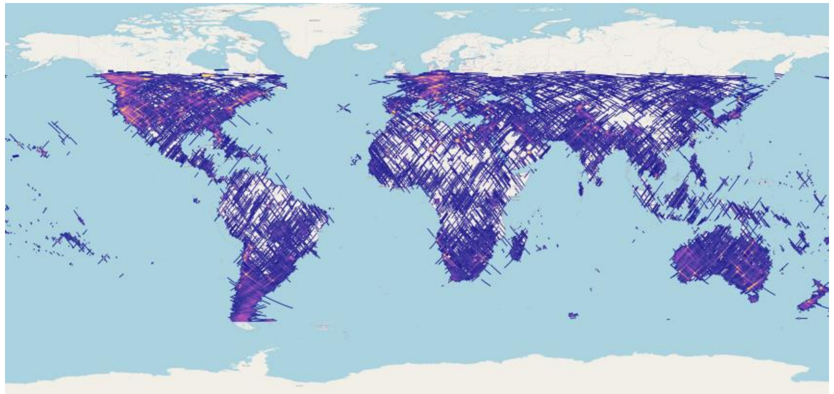




Conclusion

- ML-based SIF retrieval for DESIS developed with the help of careful simulations.
- Results are very encouraging and suggest DESIS can be used for SIF retrieval.
- Proposed model is work in progress and needs proper validation with in-situ data.
- If validated, the proposed model can be used to:
  - derive SIF from large DESIS archive; and
  - help in the validation of upcoming FLEX mission.
- Our approach is generic and can be applied to other sensors.

DESIS archive



[FLEX, earth.esa.int]

Acknowledgements

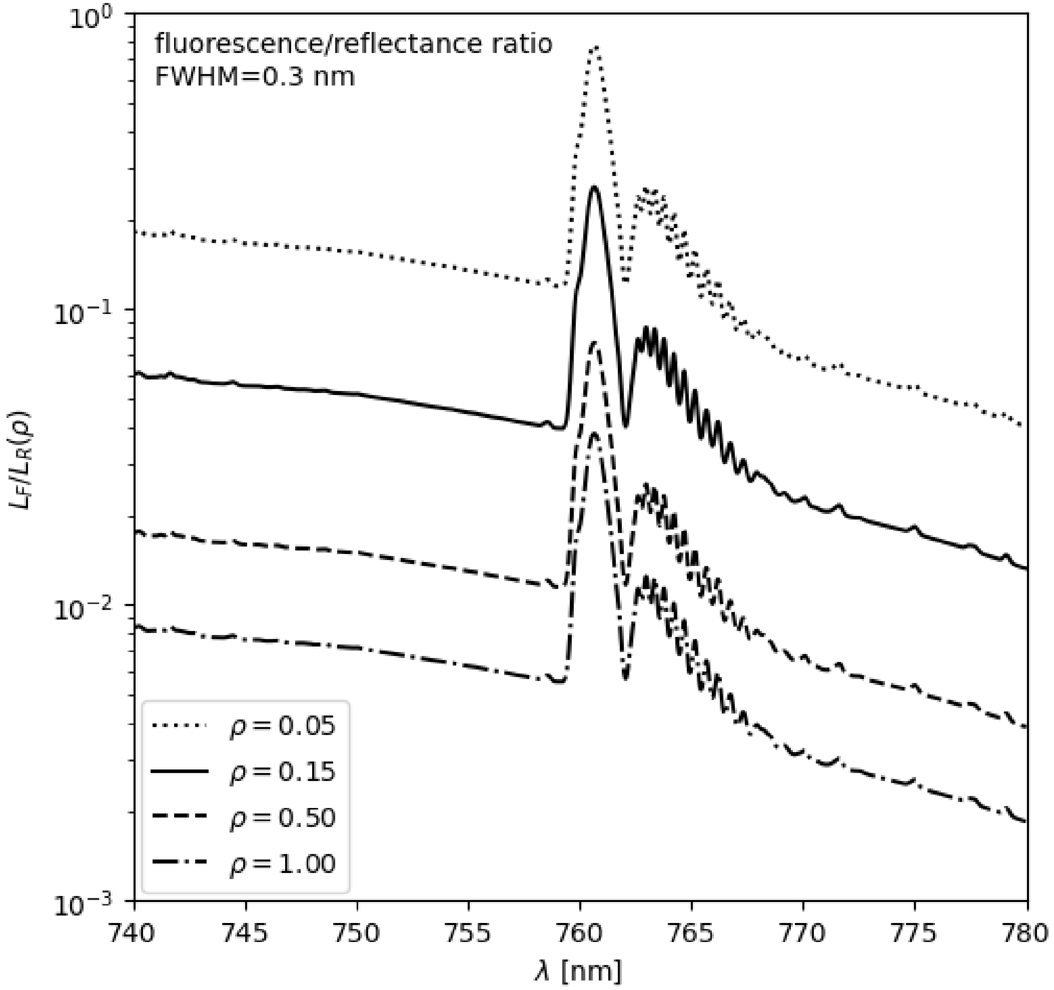
This work is part of the project “FluoMap” funded by the Helmholtz Initiative and Networking Fund, Helmholtz AI, Deutsches Zentrum für Luft- und Raumfahrt (DLR) and Forschungszentrum Jülich GmbH (FZJ). The authors gratefully acknowledge the computing time granted by the JARA Vergabegremium and provided on the JARA Partition part of the supercom- puter JURECA [1] at Forschungszentrum Jülich.

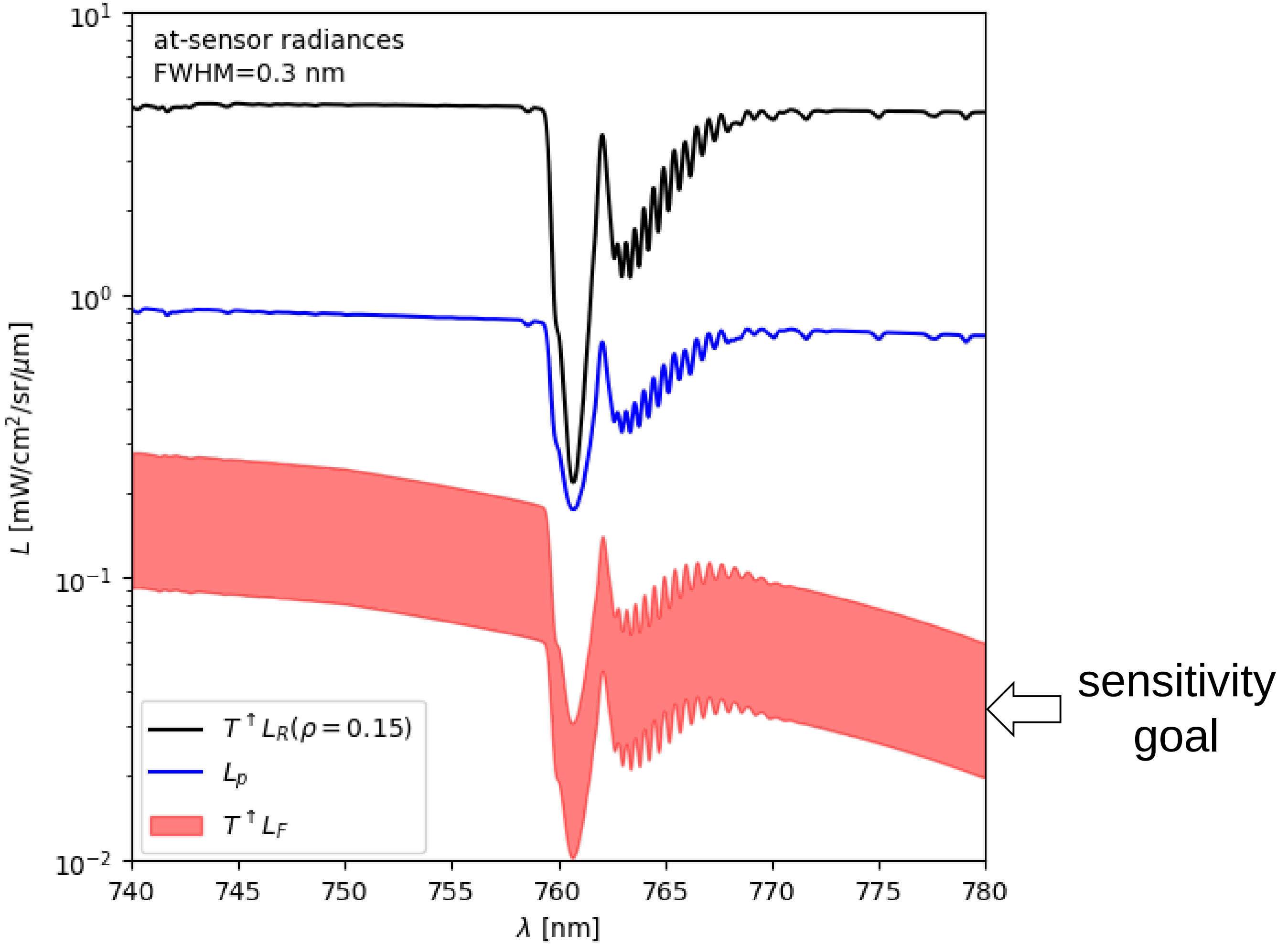
[1] Jülich Supercomputing Centre, JURECA: Data centric and booster modules implementing the modular supercomputing architecture at Jülich supercomputing centre, Journal of large-scale research facilities 7 (A182) (2021). https://doi.org/10.17815/jlsrf-7-182

BACKUP SLIDES

Sensitivity analysis

RT: LBL, 0.1 cm^{-1} / 100, DISORT (8S)
Atm: mid-lat summer, rural, 23 km vis
Geo: nadir, 30 deg Sun, h=0 km

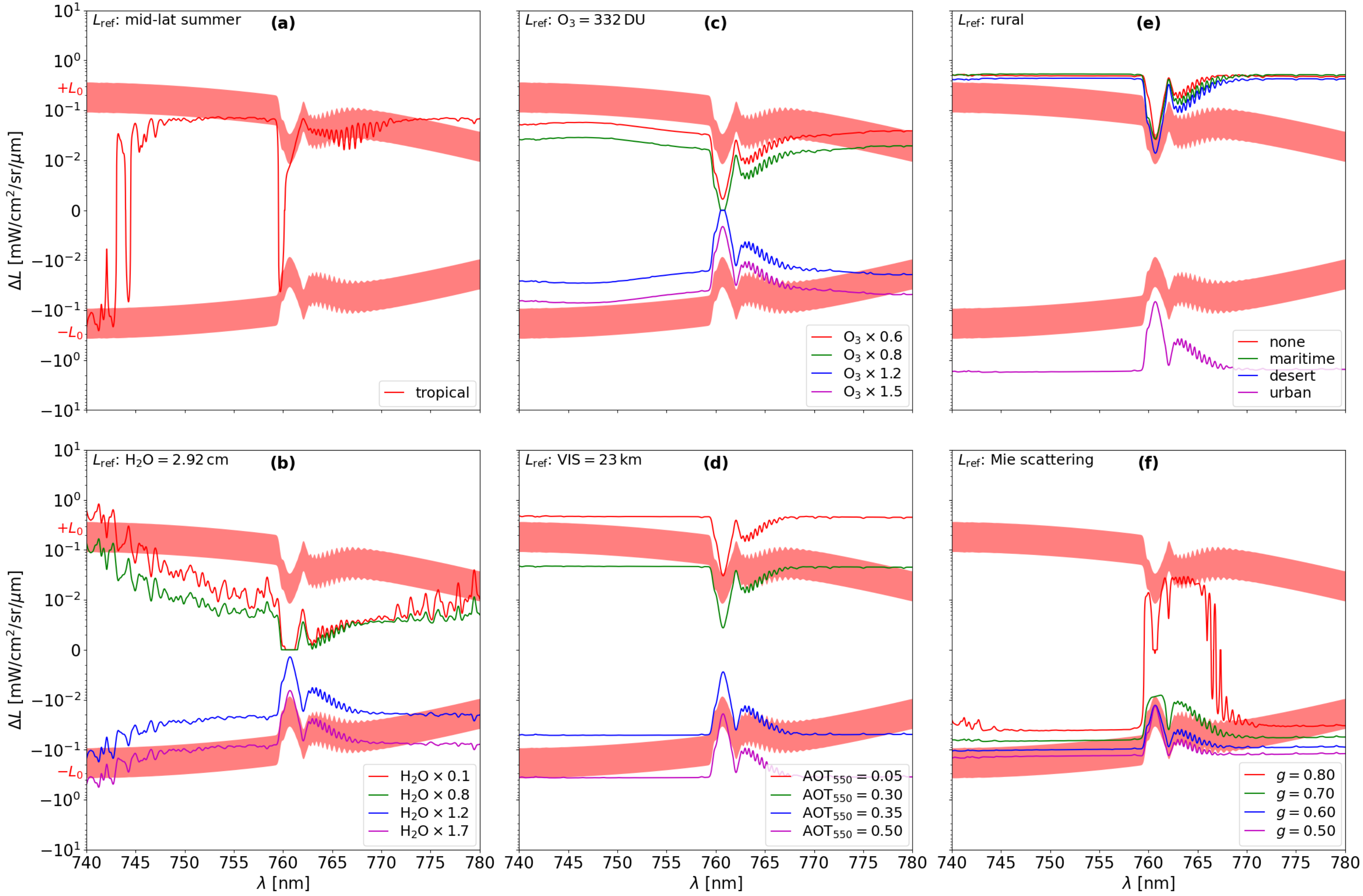




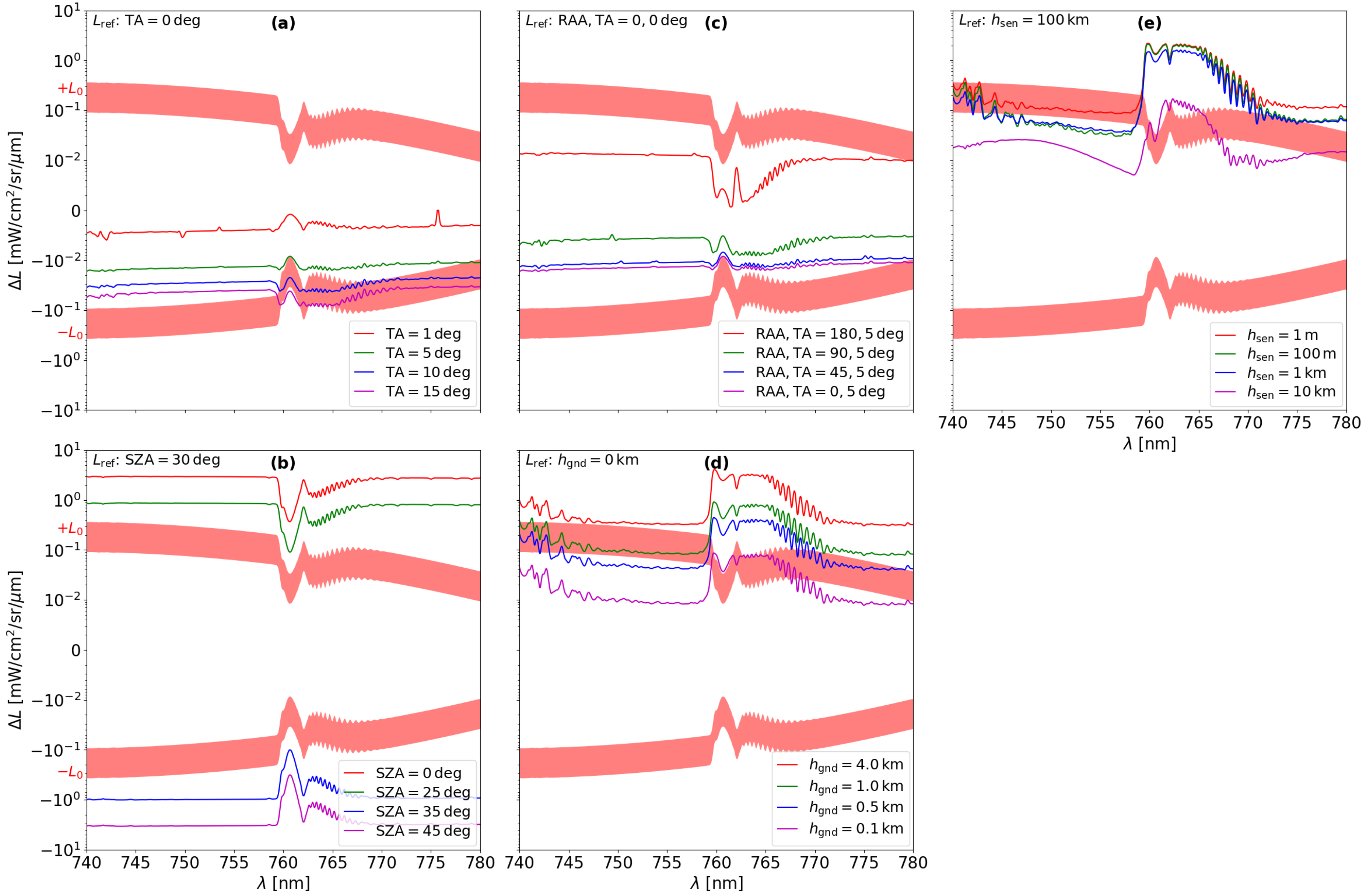
Sensitivity analysis: radiative transfer

Case	Model	Resolution	Multiple scattering	Run time
00	correlated-k (fast)	1.0 cm^{-1} / –	Isaacs scaled (8S)	00:02
A1	line-by-line	0.1 cm^{-1} / 100	DISORT (8S)	09:13
B2	correlated-k (slow)	0.1 cm^{-1} / –	DISORT (8S)	01:40
B3	correlated-k (fast)	0.1 cm^{-1} / –	DISORT (8S)	00:59
B4	band model	0.1 cm^{-1} / –	DISORT (8S)	00:09
C2	line-by-line	0.1 cm^{-1} / 50	DISORT (8S)	04:34
C3	line-by-line	0.1 cm^{-1} / 20	DISORT (8S)	01:55
C4*	line-by-line	0.1 cm^{-1} / 10	DISORT (8S)	00:57
C5	line-by-line	0.1 cm^{-1} / 5	DISORT (8S)	00:30
C6	line-by-line	0.1 cm^{-1} / 3	DISORT (8S)	00:20
C7	correlated-k (slow)	1.0 cm^{-1} / –	DISORT (8S)	00:20
C8	correlated-k (slow)	5.0 cm^{-1} / –	DISORT (8S)	00:05
D2	line-by-line	0.1 cm^{-1} / 100	Isaacs scaled (8S)	failed
D3	correlated-k (slow)	0.1 cm^{-1} / –	Isaacs scaled (8S)	00:07
D4	correlated-k (fast)	0.1 cm^{-1} / –	Isaacs scaled (8S)	00:06
D5	line-by-line	0.1 cm^{-1} / 100	None	00:05

Sensitivity analysis: atmosphere



Sensitivity analysis: geometry



Sensitivity analysis: uncertainties