ESTIMATING NET PHOTOSYNTHESIS OF VEGETATION FROM SOLAR-INDUCED CHLOROPHYLL FLUORESCENCE – A MODELLING STUDY

J. Verrelst*, C. v.d. Tol, F. Magnani, J.P. Rivera, G. Mohammed and J. Moreno

Image Processing Laboratory (IPL), University of Valencia, Spain

*Jochem Verrelst
Image Processing Laboratory
University of Valencia, Spain
jochem.verrelst@uv.es

1. Introduction

In preparation for **ESA's Earth Explorer 8 candidate mission FLEX**, a Photosynthesis Study (PS) has been completed that **aimed to quantitatively link sun-induced fluorescence (SIF) to photosynthesis** based on model and experimental data. One of the objectives of the PS was **to develop a prototype inversion algorithm to retrieve photosynthesis from simulated SIF observations. The SCOPE model** has been selected as baseline model, because it has the ability to simulate the effects of irradiance, vegetation structure and physiology on SIF and photosynthesis.

In this study, the targeted flux is **"Net photosynthesis of the canopy" (NPC)**, which is important for carbon cycle and climate change research. In order to enable estimation of NPC from SIF data, a regression analysis been pursued. This approach enables the use of simulated SIF data in retrieval of NPC for a multitude of theoretical canopy configurations. Because SCOPE is a complex model that consists of over 30 input variables, a first step is to identify the key variables that drive canopy-leaving SIF. Therefore, we had the following **objectives**:

- To apply **a global sensitivity analysis (GSA)** that quantifies the relative importance of SCOPE input variables to SIF
- To assess **the predictive power of SIF wavelengths to estimate NPC**, i.e.:
  (1) linear regression analysis between **individual SIF bands and NPC outputs**.
  (2) Linear and nonlinear regression analysis between **combined SIF bands and NPC outputs**.

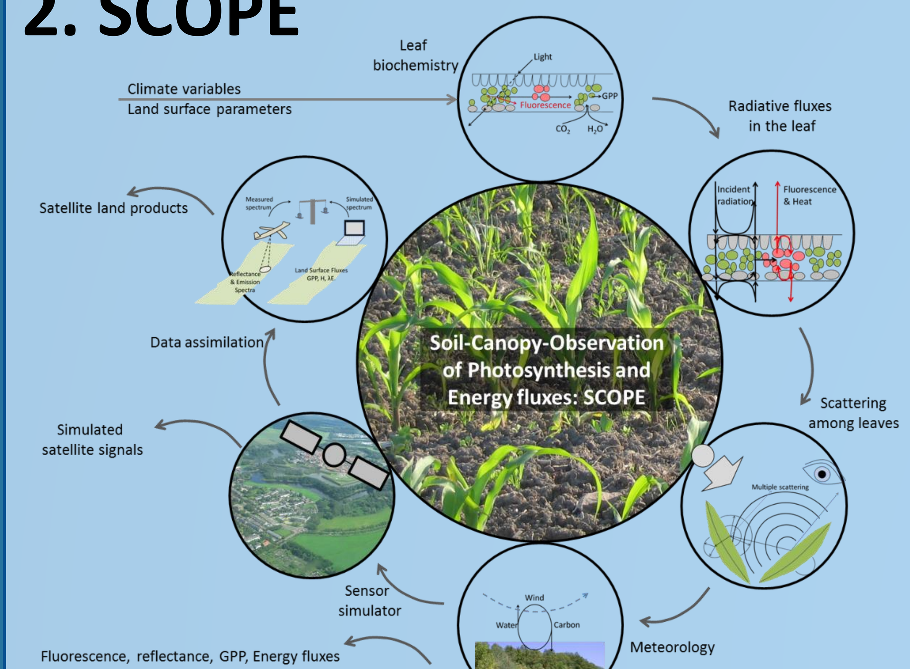
6. Conclusions

A **SCOPE modelling study** was conducted to examine how successfully canopy-leaving SIF can estimate net photosynthesis of the canopy (NPC). Based on identified key variables multiple canopy configurations were simulated. Regression analyses between SIF retrievals and NPC values led to the following general findings:

1. **The most sensitive SIF bands to NPC were located around the first (i.e. red)** emission peak for heterogeneous canopy configurations.
2. **Combining two SIF retrieval bands (e.g., O_2-B and O_2-A) led to stronger correlations than using only one SIF band.**
3. **Using the O_2-B and O_2-A bands produced similar or superior performances than using the two emission peaks**, while using the peak ratio produced poorer relationships than when both bands were individually entered into the regression model.
4. Even stronger correlations were achieved **using four main SIF retrieval bands (Hα, O_2-B, water vapour, O_2-A).**
5. **Nonlinear regression produced stronger relationships than did linear approaches.**

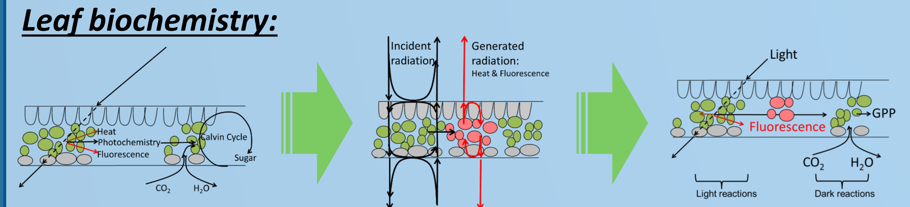
It is recommended to sample the SIF signal in at least the O_2-B and O_2-A bands in order to enable robust quantification of canopy photosynthetic activity.

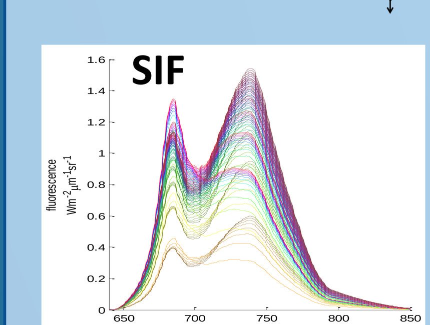
2. SCOPE

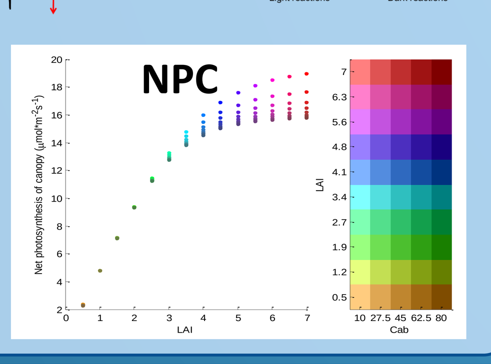


SCOPE is a vertical (1-D) integrated radiative transfer and energy balance SVAT model, with, amongst others, **sun-induced chlorophyll fluorescence (SIF) and net photosynthesis of the canopy (NPC)** as outputs.

Leaf biochemistry:







3. Global sensitivity analysis (GSA)

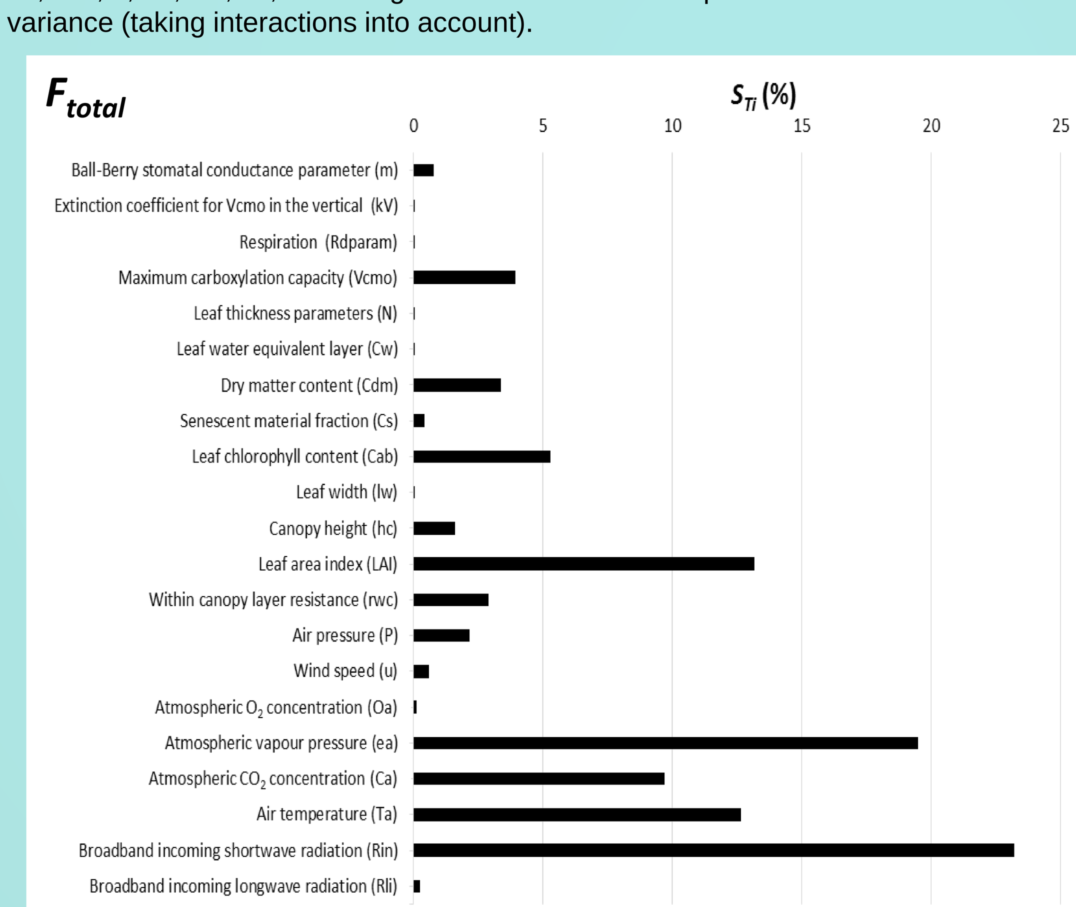
In variance-based GSA methods the output variance is decomposed to the sum of contributions of each individual input parameter and the interactions (coupling terms) between different parameters.

Based on the pioneering work of Sobol the variance-based sensitivity measures are represented as follows: $1 = \sum_i S_i + \sum_i \sum_{j>i} S_{ij} + \dots + S_{12\dots k}$

In this equation, S_i, S_{ij},…,$S_{12,\dots,k}$ are the so-called ***Sobol's global sensitivity indices***.

- The ***total effect sensitivity index S_{Ti}*** measures the whole effect of the variable *Xi*, i.e. the first order effect as well as its coupling terms with the other input variables.

The S_{Ti} over SCOPE's output F_{total} (Integrated SIF from 640 to 850 nm) was calculated to identify the driving input variables. The driving variables were: Vcmo, Cdm, Cab, LAI, hc, rwc, P, ea, Ca, Ta, Rin. Altogether these variables explained 97.5% of the total variance (taking interactions into account).



4. Experimental setup

12 Canopy configurations simulated with increasing heterogeneity. Variables were uniform randomly sampled between model min and max.

Index	Ranging variables	Justification	# Simulations
Biochemistry			
1	V_{cmo}	Vcmo is the main biochemical driver of photosynthesis. Hence, this is the theoretical baseline when SIF is not influenced by any other variable.	2000
2	Biochemistry	All biochemical variables (V_{cmo}, m, Rdparam, kV). Represents the most heterogeneous situation at the biochemical scale.	2000
Biochemistry, leaf			
3	V_{cmo}, Cab	Driving biochemical and leaf variables.	2000
4	V_{cmo}, leaf	Driving biochemical variable and all leaf variables (N, Cw, Cdm, Cs, Cab).	2000
5	Biochemistry, leaf	All biochemical and leaf variables. Represents the most heterogeneous situation at biochemical and leaf scales (V_{cmo}, m, Rdparam, kV, N, Cw, Cdm, Cs, Cab).	2000
Biochemistry, leaf , canopy			
6	Cab, LAI	Driving leaf and canopy variables.	2000
7	V_{cmo}, LAI	Driving biochemical variable (V_{cmo}) with driving canopy variable (LAI)	2000
8	V_{cmo}, canopy	Driving biochemical variable (V_{cmo}) with all varying canopy variables (LAI, lw, hc).	2000
9	V_{cmo}, N, Cw, Cdm, Cs, Cab, LAI, lw, hc (spherical LIDF)	Driving biochemical variable (Vcmo) with all leaf and all canopy (N, Cw, Cdm, Cs, Cab, LAI, lw, hc).	2000
10	Biochemistry, leaf, canopy	Al biochemical, leaf and canopy variables (V_{cmo}, m, Rdparam, kV, N, Cw, Cdm, Cs, Cab, LAI, lw, hc). Represents the most heterogeneous situation at the canopy scale	2000
All biochemistry, leaf , canopy, geometry, micrometeorology			
11	Key SCOPE variables driving SIF	V_{cmo}, Cdm, Cab, LAI, hc, rwc, P, ea, Ca, Ta, Rin. These variables and their interactions explain 97.5% of the variability in F_{total}.	2000
12	All SCOPE variables	All SCOPE variables (V_{cmo}, m, Rdparam, kV, N, Cw, Cdm, Cs, Cab, LAI, lw, hc, VZA, RAA, SZA, rwc, rb, P, u, Oa, ea, Ca, Ta, Rin, Rli). Represents the most heterogeneous configuration.	2000

Regression analysis: **Linear regression (LR) & nonlinear Gaussian processes regression (GPR)**

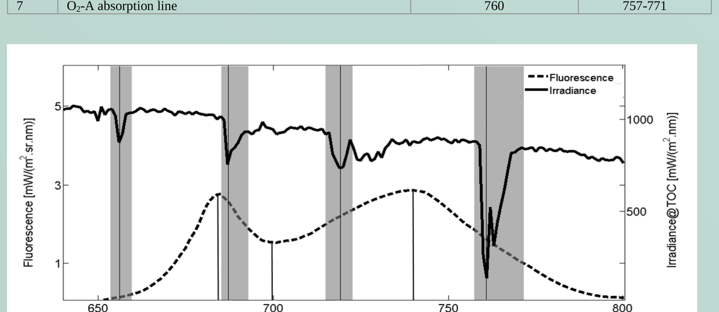
- 50 % used for training
- **50 % used for validation. R^2 and RMSE calculated.**

5. Results: Single & combined band analysis

Both the most important SIF retrieval bands and each single SIF band (1 nm) and combined bands were analyzed on their predictive power to estimate NPC using linear regression.

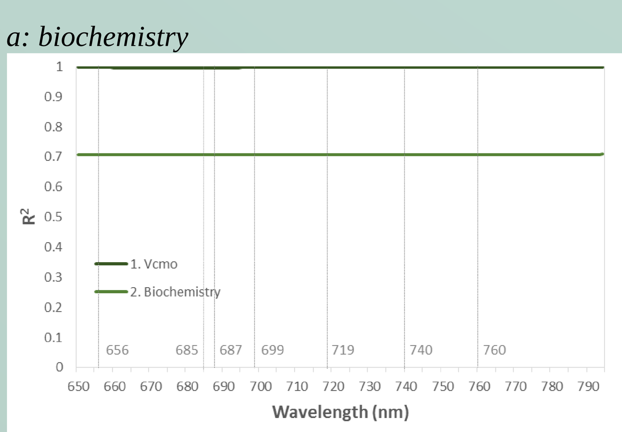
Most important SIF retrieval bands:

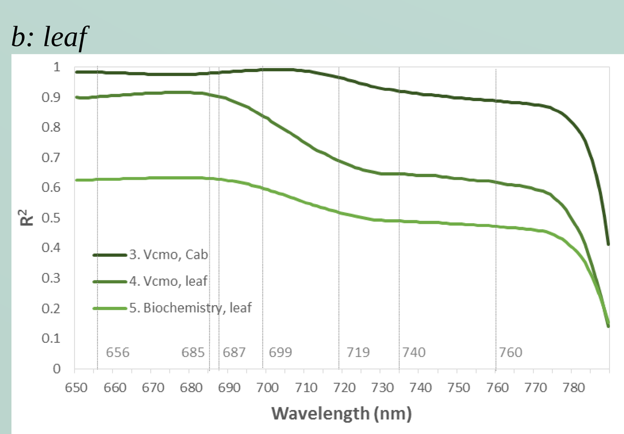
Index	Element	Central Wavelength (nm)	Spectral range (nm) absorption lines
1	Hα absorption line	656	653-662
2	Red peak (attributed to SIF emission of Photosystem II)	685	
3	O_2-B absorption line	687	683-692
4	Mid-valley between red and NIR peaks	699	
5	Water vapor absorption line	719	714-722
6	Near-infrared peak (attributed to SIF emission of Photosystem I and to a lesser extent PSII)	740	
7	O_2-A absorption line	760	757-771

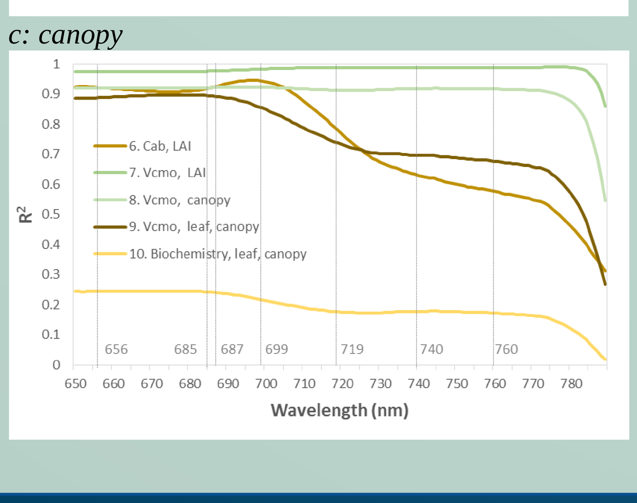


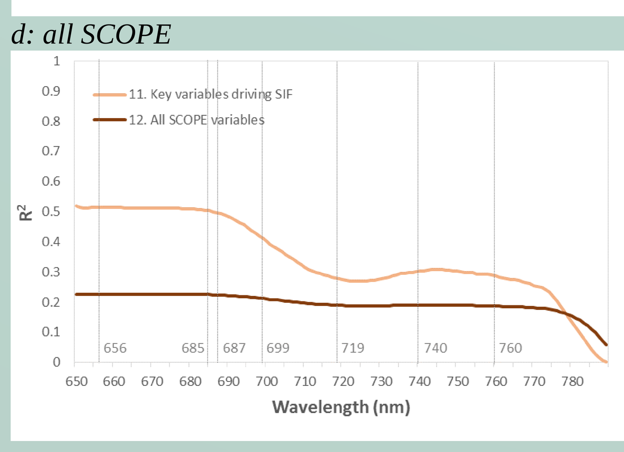
	Ranging SCOPE variables	Hα (656 nm) R²	Hα RMSE	Red peak (685 nm) R²	Red peak RMSE	O2-B (687 nm) R²	O2-B RMSE	Mid-valley (699 nm) R²	Mid-valley RMSE	Water vapor (719 nm) R²	Water vapor RMSE	NIR peak (740nm) R²	NIR peak RMSE	O2-A (760 nm) R²	O2-A RMSE	Best wavelength (nm)	R²
1	V_{cmo}	0.9970	0.7484	0.9966	0.7948	0.9966	0.7890	0.9971	0.7361	0.9975	0.6760	0.9977	0.6491	**0.9978**	0.6426	**790**	**0.9978**
2	biochemistry	0.7070	8.5148	0.7065	8.5221	0.7066	8.5207	0.7072	8.5118	0.7079	8.5031	**0.7082**	8.4986	**0.7082**	8.4975	**790**	**0.7087**
3	V_{cmo}, Cab	0.9830	1.7376	0.9801	1.8811	0.9819	1.7934	**0.9911**	1.2605	0.9652	2.4903	0.9110	3.9809	0.8883	4.4594	**703**	**0.9922**
4	V_{cmo}, leaf	0.9026	4.3075	**0.9092**	4.1596	0.9040	4.2746	0.8371	5.5674	0.6887	7.6947	0.6415	8.2583	0.6175	8.5301	**676**	**0.9159**
5	biochemistry, leaf	0.6275	9.9114	**0.6309**	9.8653	0.6288	9.8938	0.5980	10.2940	0.5178	11.2759	0.4863	11.6382	0.4720	11.8004	**677**	**0.6337**
6	Cab x LAI	0.9208	1.4306	0.9197	1.4411	0.9257	1.3864	**0.9438**	1.2047	0.7789	2.3754	0.6297	3.0730	0.5772	3.2838	**696**	**0.9459**
7	V_{cmo}, LAI	0.9744	2.3695	0.9760	2.2963	0.9766	2.2664	0.9829	1.9356	0.9880	1.6217	0.9869	1.6966	**0.9875**	1.6581	**777**	**0.9895**
8	V_{cmo}, canopy	0.9199	3.4569	0.9211	3.4316	0.9215	3.4221	**0.9227**	3.4001	0.9132	3.6086	0.9179	3.5075	0.9166	3.5353	**696**	**0.9232**
9	V_{cmo}, leaf, canopy	0.8879	4.2744	**0.8947**	4.1411	0.8925	4.1835	0.8540	4.8753	0.7377	6.5325	0.6985	7.0273	0.6768	7.2514	**678**	**0.8974**
10	Biochemistry, leaf, canopy	**0.2453**	29.3064	0.2429	29.3539	0.2411	29.3867	0.2168	29.8472	0.1773	30.5901	0.1782	30.5714	0.1727	30.6767	**650**	**0.2462**
11	Key variables driving SIF	**0.5153**	15.2639	0.5030	15.4557	0.4973	15.5454	0.4109	16.8229	0.2785	18.6249	0.3020	18.3188	0.2866	18.5130	**650**	**0.5190**
12	All SCOPE variables	**0.2260**	39.4620	0.2249	39.4886	0.2241	39.5090	0.2120	39.8133	0.1896	40.3767	0.1902	40.3663	0.1869	40.4500	**650**	**0.2263**

Overall, the red peak, O_2-B, and Hα line show similar predictive strength. The NIR peak and O_2-A are also similar in performance. **In most instances the red peak or O_2-B band are better predictors than the NIR peak or O_2-A.** In realistic canopy scenarios (i.e. with ranging variables at scales of biochemistry, leaf and canopy; scenario 11) the best performing wavelength is situated on the slope before the first peak.









Combined bands into linear and nonlinear regression:

Index	Combined wavelengths	Wavelengths (nm)
1	O_2-B and O_2-A absorption lines	687, 760
2	Hα, O_2-B, O_2-A and water vapor absorption lines	656, 687, 719, 760
3	Two SIF emission peaks	685, 740
4	Peak ratio	685/740
5	Two SIF emission peaks and mid-valley	685, 699, 740
6	F_{total}	Integrated SIF (from 640 to 850nm)
7	F_{all}	All individual SIF wavelengths (from 650 to 790)

Validation results combined bands linear regression

	Ranging SCOPE variables	O_2-B, O_2-A: 687, 760 nm R²	RMSE	Hα, O_2-B, water vapor, O_2-A: 656, 687, 719, 760 nm R²	RMSE	Two peaks: 685, 740 nm R²	RMSE	Peak ratio: 685/740 R²	RMSE	Two peaks and valley: 685, 699, 740 nm R²	RMSE	F_{total}: Integrated SIF (from 640 to 850nm) R²	RMSE	F_{all}: All individual SIF wavelengths (from 650 to 790) R²	RMSE
1	V_{cmo}	1	0.0040	1	0.0053	1	0.0038	0.9900	1.4283	1	0.0039	1	0.0089	**1**	0.0028
2	biochemistry	0.7863	7.0925	0.7901	6.9813	0.7332	7.8987	0.7181	8.2289	0.7389	8.1721	0.7268	35.8424	**0.7996**	6.7456
3	V_{cmo}, Cab	0.9996	0.2598	0.9998	0.1948	0.9996	0.2548	0.6668	7.6360	0.9987	0.4841	0.9905	1.3127	**1**	0.0811
4	V_{cmo}, leaf	0.9695	2.3618	0.9883	1.4783	0.9739	2.1710	0.6382	8.0869	0.9830	1.7833	0.6934	7.5117	**0.9899**	1.3645
5	biochemistry, leaf	0.6797	9.2628	0.6640	9.2910	0.6833	9.0908	0.5779	10.5707	0.6777	9.1064	0.5459	10.6994	**0.7633**	7.9247
6	Cab x LAI	**1**	0.0101	0.9919	0.4640	**1**	0.0066	0.4044	3.8472	0.9879	0.5552	0.9589	1.7122	**1**	0.0040
7	V_{cmo}, LAI	0.9967	0.8520	0.9988	0.5203	0.9993	0.3837	0.8926	4.8546	0.9791	2.1539	0.9945	1.1048	**1**	0.0070
8	V_{cmo}, canopy	0.9456	2.9284	0.9431	2.9300	0.9447	2.9073	0.8274	5.2367	0.9397	3.0608	0.9416	3.0289	**0.9945**	0.9428
9	V_{cmo}, leaf, canopy	0.9164	3.7109	0.9174	3.7290	0.9177	3.7213	0.5272	8.9396	0.8982	4.1313	0.7250	6.7224	**0.9401**	3.1440
10	Biochemistry, leaf, canopy	0.3180	27.7851	0.3573	27.6869	0.3392	30.4935	0.1938	33.3582	0.3280	27.6240	0.1935	31.0465	**0.3819**	28.0068
11	Key variables driving SIF	0.5881	12.1416	**0.6610**	10.9384	0.6000	11.2757	0.3184	15.4214	0.5374	13.0736	0.4053	14.7770	0.6411	11.7772
12	All SCOPE variables	0.3131	37.2426	0.2676	41.4844	**0.3204**	35.4161	0.1350	43.8470	0.2870	38.1309	0.2292	39.3470	0.2873	42.9957

- **Combining the O_2-A and O_2-B bands or the red and NIR peaks produced stronger relationships with NPC than were obtained when the single O_2-A band or the NIR peak was used.**
- **The combination of O_2-B and O_2-A bands produced similar results as combining the two peaks.** Hence, these two combinations could be considered essentially equivalent from this analysis.
- In all cases, the peak ratio (F_{685}/F_{740}) produced considerably poorer correlations than using the two bands individually. Combining the mid-valley with the two peaks produced only marginal improvements over the combined peaks.
- **Also small improvements were obtained when combining SIF retrievals at the four absorption lines (Hα, O_2-B, water vapor, O_2-A).** The F_{total} (integrated SIF) generally did not yield a predictive advantage and in several instances produced weaker correlations than other features.
- Conversely, **further improvements are achieved for most of the scenarios when including all individual wavelengths into the regression analysis**, but gains in explained NPC variance are modest.

Validation results combined bands nonlinear Gaussian processes regression (GPR)

	Ranging SCOPE variables	O_2-B, O_2-A: 687, 760 nm R²	RMSE	Hα, O_2-B, water vapor, O_2-A: 656, 687, 719, 760 nm R²	RMSE	Two peaks: 685, 740 nm R²	RMSE	Peak ratio: 685/740 R²	RMSE	Two peaks and valley: 685, 699, 740 nm R²	RMSE	F_{total}: Integrated SIF (from 640 to 850nm) R²	RMSE	F_{all}: All individual SIF wavelengths (from 650 to 790) R²	RMSE
1	V_{cmo}	0.9965	0.8255	0.9970	0.7412	0.9965	0.7869	0.8686	4.9370	0.9966	0.7831	0.9983	0.5665	**1**	0.0843
2	biochemistry	**0.7234**	8.0229	0.7111	8.1849	0.7195	8.1690	0.5874	9.8065	0.7141	8.3908	0.7201	8.2419	0.7011	8.4288
3	V_{cmo}, Cab	0.9965	0.7792	0.9982	0.5657	0.9963	0.8123	0.4721	9.6652	0.9966	0.7816	0.9900	1.3493	**0.9991**	0.4056
4	Vcmo, leaf	0.9333	3.5056	0.9462	3.0807	0.9305	3.6105	0.3062	11.1958	0.9573	2.8784	0.6467	8.0626	**0.9783**	2.0221
5	biochemistry, leaf	0.6356	9.7534	0.6205	10.2080	0.6337	9.8866	0.4056	12.3210	0.6423	9.3908	0.5176	10.9974	**0.6828**	8.9835
6	Cab x LAI	0.9529	1.1083	0.9354	1.2673	0.9547	1.0751	0.1955	4.5583	0.9434	1.2164	0.9426	2.0248	**0.9728**	0.8098
7	V_{cmo}, LAI	0.9907	1.4723	0.9853	1.7952	0.9898	1.5035	0.7240	7.7957	0.9770	2.2615	0.9923	1.3046	**0.9968**	0.8415
8	V_{cmo}, canopy	0.9315	3.2390	0.9333	3.2096	0.9321	3.2583	0.6404	7.3998	0.9244	3.4306	0.9105	3.7510	**0.9490**	2.8608
9	V_{cmo}, leaf, canopy	0.8950	4.1334	0.9023	4.0438	0.9077	3.9263	0.3495	10.2994	0.8982	4.1313	0.7157	6.8358	**0.9169**	3.7433
10	Biochemistry, leaf, canopy	0.2356	31.8478	0.2388	30.8678	0.2348	30.8049	0.1152	33.2900	0.2144	30.2922	0.1292	32.3184	**0.2805**	30.4726
11	Key variables driving SIF	0.4581	13.7917	0.5068	13.6618	0.5342	12.5735	0.2484	15.8487	0.5407	12.6540	0.3078	16.1429	**0.5693**	12.4038
12	All SCOPE variables	0.2234	41.6271	0.2112	45.0060	**0.2446**	40.2953	0.1260	41.4454	0.2091	41.6138	0.2278	35.7368	0.2275	40.8113

- **The nonlinear GPR produced stronger relationships with NPC in the majority of cases** as opposed to linear regression, although improvements were generally modest.
- Considering the best two-band combinations (i.e. the two peaks or the O_2-A and O_2-B), R^2 values were higher in scenarios 2, 4-6, and 8-12.
- **The strongest improvements were under conditions of increasing canopy and environmental heterogeneity.** Again, including all individual wavelengths into the regression analysis led to strongest relationships for the majority of the scenarios, although improvements as opposed to using the SIF absorption bands were generally modest.
- **From a pragmatic perspective, by using an adaptive, nonlinear regression method and retrieving SIF in the two deepest absorption lines could be sufficient to derive NPC with sufficient accuracy.**