Estimating near-infrared reflectance of vegetation from hyperspectral data

Yelu Zeng, Dalei Hao, Grayson Badgley, Alexander Damm, Uwe Rascher, Youngryel Ryu, Jennifer E. Johnson, Vera Krieger, Shengbiao Wu, Han Qiu, et al.

Estimating near-infrared reflectance of vegetation from hyperspectral data

Yelu Zeng [a,b], Dalei Hao [c,*], Grayson Badgley [b,d], Alexander Damm [e,f], Uwe Rascher [g], Youngryel Ryu [h], Jennifer Johnson [b], Vera Krieger [g], Shengbiao Wu [i], Han Qiu [a], Yaling Liu [j], Joseph A Berry [b], Min Chen [a,k,*]

[a] Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Drive, Madison, WI 53706, USA

[b] Department of Global Ecology, Carnegie Institution for Science, Stanford, CA 94305, USA

[c] Atmospheric Sciences and Global Change Division, Pacific Northwest National Laboratory, Richland, WA, USA

[d] Black Rock Forest, Cornwall, NY 12518, USA

[e] Department of Geography, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[f] Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland

[g] Institute of Bio- and Geosciences, IBG-2: Plant Sciences, Forschungszentrum Jülich GmbH, Leo-Brandt-Str., 52425 Jülich, Germany

[h] Department of Landscape Architecture and Rural Systems Engineering, Seoul National University, South Korea

[i] Laboratoire des Sciences du Climat et de l'Environnement, CEA-CNRS-UVSQ, UMR8212, Gif-sur-Yvette, France

[j] Kettle Reinsurance, 83 Norwood Ave, Kensington, CA 94707

[k] Nelson Institute Center for Climatic Research, University of Wisconsin-Madison, 1225 W. Dayton St. Madison, WI 53706 USA

* *Corresponding authors*: Min Chen (min.chen@wisc.edu) and Dalei Hao (dalei.hao@pnnl.gov)

Abstract: Disentangling the individual contributions from vegetation and soil in measured canopy reflectance is a grand challenge to the remote sensing and ecophysiology communities. Since Solar Induced chlorophyll Fluorescence (SIF) is uniquely emitted from vegetation, it can be used to evaluate how well reflectance-based vegetation indices (VIs) can separate the vegetation and soil components. Due to the residual soil background contributions, Near-infrared (NIR) reflectance of vegetation (NIRv) and Difference Vegetation index (DVI) present offsets when compared to SIF (i.e., the value of NIRv or DVI is non-zero when SIF is zero). In this study, we proposed a simple framework for estimating the true NIR reflectance of vegetation from Hyperspectral measurements (NIRvH) with minimal soil impacts. NIRvH takes advantage of the spectral shape variations in the red-edge region to minimize the soil effects. We evaluated the capability of NIRvH, NIRv and DVI in isolating the true NIR reflectance of vegetation using the data from both the model-based simulations and Hyperspectral Plant imaging spectrometer (HyPlant) measurements. Benchmarked by simultaneously measured SIF, NIRvH has the smallest offset (0~0.037), as compared to an intermediate offset of 0.047~0.062 from NIRv, and the largest offset of 0.089~0.112 from DVI. The magnitude of the offset can vary with different soil reflectance spectra across spatio-temporal scales, which may lead to bias in the downstream NIRv-based photosynthesis estimates. NIRvH and SIF measurements from the same sensor platform avoided complications due to different geometry, footprint and time of observation across sensors when studying the radiative transfer of reflected photons and SIF. In addition, NIRvH was primarily determined by canopy structure rather than chlorophyll content and soil brightness. Our work showcases that NIRvH is promising for retrieving canopy structure

parameters such as leaf area index and leaf inclination angle, and for estimating fluorescence yield with current and forthcoming hyperspectral satellite measurements.

Keywords:

Solar-Induced chlorophyll Fluorescence (SIF), hyperspectral remote sensing, soil contamination, near-infrared reflectance of vegetation (NIRv), singular value decomposition (SVD), red edge

1.Introduction

Solar-induced chlorophyll Fluorescence (SIF) has been increasingly used to estimate the terrestrial gross primary productivity (GPP) (Frankenberg et al., 2011; Guanter et al., 2014; Ryu et al., 2019), because it originates from vegetation photosynthetic activity and is minimally influenced by soil background (Wang et al., 2019; Zeng et al., 2019). Despite these advantages, applications of SIF for GPP estimations are limited by a number of factors, including the relatively coarse spatial resolution, low revisit frequency, low signal-noise ratio, and short history of measurements (since early 2000s) compared to the optical remote sensing (1972~) (Guanter et al., 2015). These limitations have been partly compensated with the exploitation of the newest satellites/sensors, e.g., the TROPOspheric Monitoring Instrument (TROPOMI) and Orbiting Carbon Observatory-2 (OCO-2), making a step change particularly in the spatial resolution. In the near future, the Fluorescence Explorer (FLEX) mission by the European Space Agency (Drusch et al., 2017) will provide SIF measurements at an unprecedented spatial resolution (300 m). Other limitations mentioned above will remain, including the short time series of SIF and high retrieval uncertainties.

Meanwhile, substantial knowledge on the relationship between SIF and traditional optical remote sensing has been accumulated over the recent years and can be beneficial to

compensate limitations inherent to SIF. Badgley et al. (2017), for example, proposed the near-infrared (NIR) reflectance of vegetation (NIRv) as the product of normalized difference vegetation index (NDVI) and NIR reflectance (NIRv=NIR·NDVI), which has been demonstrated as a prominent indicator of SIF under low stress conditions (Turner et al., 2020). This suggests that NIRv can be considered as the "potential SIF" (Zeng et al., 2019) and can be either used as a complement of SIF to separate the physiological and structural components (Wang et al., 2020), or as a proxy of SIF when SIF measurement is unavailable (Wu et al., 2020; Peng et al., 2020). Since NIRv is entirely based on optical surface reflectance which is much easier to acquire, it addresses many of the aforementioned limitations of SIF. Further, the availability of a "potential SIF" proxy could open new opportunities to derive SIF yield (Φ_F), which is a higher-level metric with even more direct linkage with vegetation photosynthesis. To date, NIRv has been used in GPP estimates, view-angle correction of SIF, Φ_F and crop yield retrievals (Badgley et al., 2019; Hao et al., 2021a, 2021b; Peng et al., 2020; Wang et al., 2020).

However, a common limitation in traditional optical vegetation remote sensing is that observations are typically a mixture of solar radiation reflected by vegetation and the soil background. How to remove the soil contribution from observed remote sensing signals has been a challenge for optical remote sensing (Asner et al., 2002; Yang et al., 2019). As demonstrated by Badgley et al. (2017), NIRv can significantly reduce a major portion of the soil impact in the total NIR reflectance of the pixel. Zeng et al. (2019) explored the underlying physical mechanism of NIRv and found that NIRv was a good approximation of the NIR reflectance only contributed by the vegetation with a black soil background (NIR_{BS}) and shared a similar canopy radiative transfer process as SIF (Zeng et al., 2019; Dechant et al., 2020). However, there remain residual impacts of changing soil background on NIRv (i.e., slightly violating the black-soil assumption of NIR_{BS}), which will yield uncertainties that can

propagate in the downstream NIRv applications. In fact, the varying offset (i.e., the value of a vegetation index (VI) when SIF is zero) in the NIRv-SIF relationship due to different levels of soil contribution may complicate the universal application of NIRv as the "potential SIF" proxy, and thus its downstream applications such as estimating GPP or Φ_F across spatiotemporal scales.

The current formula of NIRv as NIR·NDVI can still be partially impacted by the soil background. In this formula, NDVI serves as an adjusting factor of NIR. Ideally NDVI should range from zero (pure soil) to one (dense canopy), but in reality the soil NDVI is larger than zero, and NDVI for dense canopy is always smaller than one. Yang et al. (2020) proposed the Fluorescence Correction Vegetation Index (FCVI), which requires the NIR reflectance and the broadband visible reflectance covering the range of photosynthetically active radiation (PAR, 400~700 nm). The Difference Vegetation Index (DVI=NIR-Red; Richardson et al., 1977) was used as a proxy for FCVI (Joiner et al., 2020). DVI can also partially reduces the soil contribution due to the contrast spectra of soil and vegetation in the red-NIR region. However, DVI can still be affected by the soil background. Because soil reflectance generally increases from visible to NIR wavelengths (i.e., NIR reflectance > red reflectance), the soil DVI (=NIR-Red) is typically larger than zero and can increase with the slope of the soil visible-NIR spectrum. It must be noted that the DVI is sensitive to the underlying soil spectral shape. This sensitivity can be compensated using a combination of an *a-priori* acquired soil spectrum and the weighted DVI (WDVI) that employs weighted red and NIR spectral bands to account for varying soil spectral shapes (Richardson et al., 1977). NDVI can be rewritten as DVI/(NIR+Red). Considering the stronger variation of the denominator than the numerator, darker soil background results in higher NDVI values, while brighter soil background leads to lower NDVI values (Qi et al., 1994; Huete et al., 2002). The increase of canopy reflectance with the wavelength is contributed by the increase of both the

vegetation and soil reflectance contributions, and this effect is obvious at the red edge (e.g., 675~800 nm). With the support of current and forthcoming hyperspectral missions, e.g., the TROPOMI covering 675~775 nm (Guanter et al., 2015), and the FLEX covering 500~780 nm (Drusch et al., 2016), it is expected to be promising to employ the red edge to better reduce the soil contribution than NIRv and DVI. For multi-spectral sensors, it is difficult to detect subtle shifts in the location and variation of the red edge reflectance driven by the chlorophyll content (Ollinger et al., 2011).

Implementing an approach that allows NIRv to be less dependent on changing soil background will open new possibilities to mitigate current limitations in SIF and enable new applications. In this study, we explored features in leaf and soil spectral shapes at the red edge to define a strategy for separating vegetation and soil contributions from measured canopy reflectance. We implemented this strategy as a new derivative of the NIR reflectance of vegetation applicable to hyperspectral data (NIRvH) with substantially reduced sensitivity to soil background, and thoroughly evaluated the proposed NIRvH against the existing NIRv and DVI approaches.

2.Theoretical foundation

The total bidirectional reflectance factor BRF_T at wavelength λ of a vegetated surface can be described as a sum of three components (Zeng et al., 2019):

$$BRF_T(\lambda)=BRF_{BS}(\lambda)+BRF_S(\lambda)+BRF_M(\lambda) \quad (1)$$

where BRF_{BS} is the contribution of photons at wavelength λ reflected back by only the vegetation component over a black soil background (i.e., no soil reflection), BRF_S is the single scattering contribution of photons at wavelength λ reflected back by the soil alone, and BRF_M is the contribution of photons by multiple scattering between vegetation and soil. The

goal of this study is thus to reduce the soil contribution from the total reflectance (BRF_T). Whether the BRF_M needs to be removed or not depends on the specific application case, and will be discussed in Section 5.2. Based on the spectral invariants theory which considers the photon interactions with leaves and soil within the canopy in a successive order of scattering (Knyazikhin et al., 2013; Zeng et al., 2018), we can write $BRF_{BS} = \Sigma\Sigma a_{i,j} \cdot \rho(\lambda)^i \cdot \tau(\lambda)^j$, $BRF_S = b \cdot R_S(\lambda)$, and $BRF_M = \Sigma\Sigma\Sigma c_{i,j,k} \cdot \rho(\lambda)^i \cdot \tau(\lambda)^j \cdot Rs(\lambda)^k$, where $\rho(\lambda)$ and $\tau(\lambda)$ are the leaf reflectance and transmittance at wavelength λ, respectively, $R_S(\lambda)$ is the soil reflectance at wavelength λ, b is the bidirectional gap fraction, while $a_{i,j}$ and $c_{i,j,k}$ are the wavelength-independent coefficients that are within the range of 0~1 and are only determined by the canopy structure and the multiple scattering orders i, j and k. To achieve a smaller offset (the value of VI when SIF or GPP is zero) in the SIF-VI (or GPP-VI) relationship, we need to reduce the soil contribution (BRF_S and part of BRF_M) from the total reflectance (BRF_T). The distinct features of the leaf and soil spectrum at the red edge are the key to reduce the soil contribution in BRF. Below we will describe the features of the leaf spectrum and soil spectrum at the red edge in Section 2.1 and 2.2, respectively. Then we will describe two approaches for better approximating the true NIR reflectance of vegetation with hyperspectral data in Section 2.3.

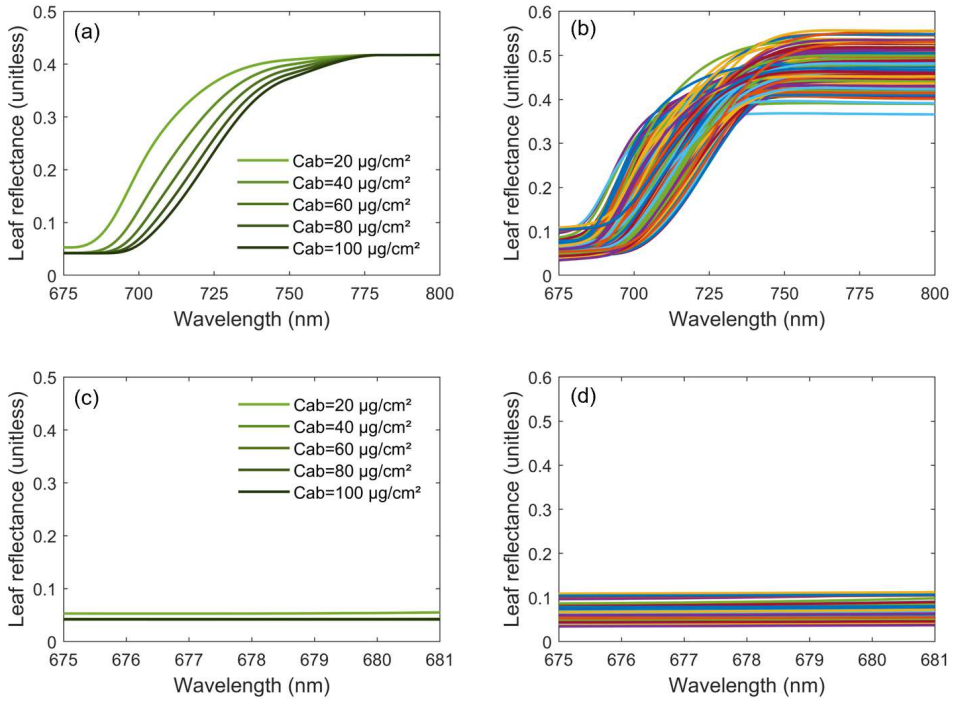
2.1 Leaf spectra exhibit strong absorption around 678 nm and a flat plateau in NIR

The first step to decouple the soil and vegetation contributions in total reflectance is to identify the spectral regions where the leaf spectrum is invariant while the soil spectrum is changing. There is a strong absorption feature in the leaf spectrum in the red spectral domain (675~681 nm, Fig. 1c, d), as shown in the simulations with the PROSPECT-D model and measurements in the ANGERS Leaf Optical Properties Database (Fig. 1a, b) (Feret et al.,

2008; Feret et al., 2017). This feature originates from a strong absorption by chlorophyll around 675~681 nm peaking at 678 nm (Fig. S2a).

Secondly, a relatively flat plateau of the leaf spectrum within the NIR range of 778~800 nm (Figs. 1 and S1) is expected if brown pigments are not taken into consideration. The variation of the chlorophyll content only shifts the starting and ending points of the red edge, but does not influence the reflectance in the NIR bands (Fig. 1). This is because in this wavelength region, there is little absorption of most leaf biochemical constituents, except the dry matter and brown pigments (Fig. S2). Although the brown pigments quantified by senescent material fraction can lead to positive slope in this wavelength region (Fig. S1b), flat plateaus are observed in the ANGERS Database (Feret et al., 2008), as shown in Fig. 1b. Overall, we find little change of the leaf spectrum in the NIR band in our modelling experiments due to the stable absorption of leaf biochemical constituents. The brown pigment-induced uncertainties will be discussed in Section 3.1.

Due to these features, the red band (675~681 nm, Fig. 1c, d) and NIR band (778~800 nm) can be used as candidate spectral regions for unmixing vegetation and soil contributions to the canopy reflectance. In practice, the selection of spectral region depends on two data features: (1) the spectral resolution, i.e., the narrow red band (675~681 nm) must have high spectral resolution (<2 nm); (2) the spectral coverage, i.e., the sensor must cover either the red region (675~681 nm) or the NIR region (778~800 nm).



192

193 **Fig. 1** Leaf reflectance spectra simulated by the PROSPECT-D model (a, c) and measured by
194 INRA (National Institute for Agricultural Research) from the ANGERS Leaf Optical
195 Properties Database (Feret et al., 2008) available at
196 http://opticleaf.ipgp.fr/index.php?page=database (b, d). In the PROSPECT-D simulations, the
197 chlorophyll content ranged from 20 µg cm^{-2} to 100 µg cm^{-2} at a step of 20 µg cm^{-2}. The
198 carotenoid content was set to 20 µg cm^{-2}, the dry matter content was 0.012 g cm^{-2}, the
199 equivalent water thickness was 0.009 cm and the leaf structure parameter *N* was 1.4. The
200 spectral range of 675~800 nm in (a) and (b) represents the red edge which we are interested
201 in for the NIRvH estimation. The colors in (b) and (d) only represent different leaf spectra.
202 The spectral range of 675~681 nm in (c) and (d) represents the narrow region where the leaf
203 spectra is relatively stable in the red band.

204

205 **2.2 Soil spectra exhibit a steady continuous change at the red edge**

206 The shape of the soil reflectance spectra is typically stable with a smooth increase around
207 the red edge from 675 nm to 800 nm (Fig. 2). Because the leaf spectrum is generally flat in
208 red band (675~681 nm) and NIR band (778~800 nm) (Fig. 1c,d), the increase of the canopy
209 total reflectance in the two spectral regions (Fig. 3) is primarily due to the increase of the soil

reflectance. When LAI increases from 0.5 to 5 m^2 m^{-2}, the contribution from soil single scattering (BRF_S) to the canopy total reflectance (BRF_T) may decrease, and thus the canopy spectrum becomes flatter, especially in the NIR bands. When LAI is large (e.g., 5 m^2 m^{-2}), BRF_S may be minimal and neglectable, and thus BRF_T behaves similarly to the leaf spectrum and becomes flat in the red (675~681 nm) and NIR (778~800 nm) regions. Such features offer an opportunity to extract the profile of BRF_S with measurements in the two spectral regions.

In particular, we find that the red-band BRF_T at the chlorophyll absorption peak (678 nm) is almost solely contributed by BRF_S. For different LAIs, the vegetation contribution (BRF_T - BRF_S) at 678 nm does not exceed 0.02 (Fig. 3), which is even smaller than the leaf reflectance at 678 nm (generally <0.1 and mostly <0.05, Fig. 1d) considering the fractional vegetation cover, multiple scattering, and strong reabsorption. Therefore, BRF_T at 678 nm provides useful information of soil reflectance. Once we know the soil spectrum profile or its slope, this information can be extrapolated to other wavelengths and used to estimate the soil contribution to BRF_T.

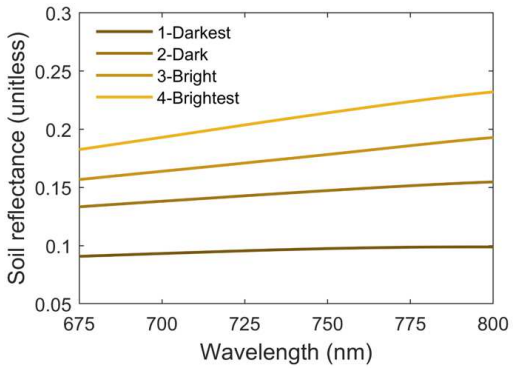


Fig. 2 Four soil spectrum samples with different brightness from the soil database in the SCOPE model (van der Tol et al., 2009). The spectral range of 675~800 nm in (a) and (b) represents the red edge which we are interested in for the NIRvH estimation.

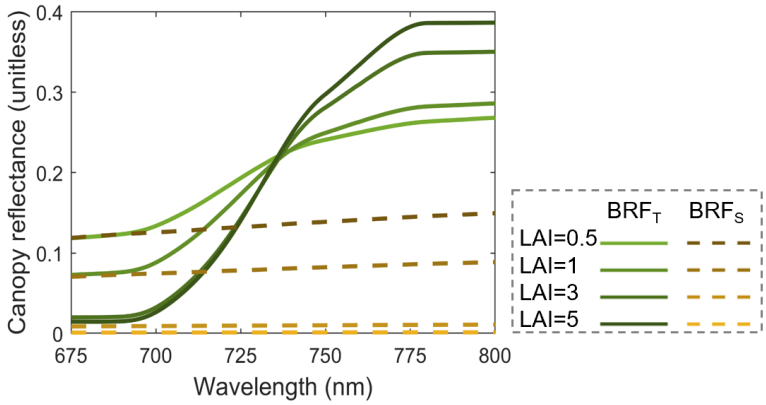


Fig. 3 Canopy-scale total reflectance (BRF_T) with increasing LAI and the same soil background (dried soil, 4-Brightest in Fig. 2), simulated by the SCOPE model. The soil single scattering contribution (BRF_S) is calculated by the product of the bidirectional gap fraction times the soil reflectance. The spectral range of 675~800 nm represents the red edge which we are interested in for the NIRvH estimation.

2.3 Two approaches to develop the hyperspectral NIRv (NIRvH)

Based on the analysis above, we developed two simple and practical approaches (i.e., NIRvH1 and NIRvH2) for the NIRvH index without using any additional canopy structure parameters. In principle, BRF_T can be written in another way as the sum of vegetation and soil contributions (BRF_{Veg} and BRF_{Soil}, respectively):

$$BRF_T(\lambda) = BRF_{Veg}(\lambda)+BRF_{Soil}(\lambda) \quad (2)$$

NIRvH1 is a generalized approach that uses a singular value decomposition (SVD) method to estimate BRF_{Soil} and a logistic function to fit the BRF from vegetation (BRF_{Veg}) at the red edge (675~800 nm). Compared to Eq. (1), BRF_{Soil} is the sum of BRF_S and a part of BRF_M, while BRF_{Veg} includes BRF_{BS} and the remaining portion of BRF_M in BRF_T.

BRF_{Soil} at the red edge can be estimated as:

$$BRF_{Soil}(\lambda)=\Sigma\square_i \cdot PC_i(\lambda) \quad (3)$$

where PC_i are the principal components extracted from the neighbouring pure soil pixels in the same image or from the soil spectral database, and $\square_i$ are the fitted coefficients. The shape of soil spectrum at the red edge within 675~800 nm is typically as simple as a linear line (Fig. 2). The first one or two *PC*s usually can explain more than 95% of the soil spectrum variance, and the complexity of the soil spectrum shape determines the number of *PC*s needed for the fitting. With the SVD method, we do not need to assume the shape of the soil spectrum is linear at the red edge.

BRF_{Veg} at the red edge can be approximated as a logistic function:

$$BRF_{Veg}(\lambda)=a/(1+b\cdot\exp(-t\cdot(\lambda-\lambda_0))) \quad (4)$$

where a, b and t are the fitted coefficients, and λ_0 is 675 nm where the canopy spectrum begins to increase.

Combining Eqs. X and Y gives:

$$BRF_T(\lambda)= BRF_{Veg}(\lambda)+ BRF_{Soil}(\lambda)=\Sigma\square_i\cdot PC_i(\lambda)+a/(1+b\cdot\exp(-t\cdot(\lambda-\lambda_0))) \quad (5)$$

Therefore, we can use the canopy reflectance to fit Eq. 5 to determine the coefficients ($\square_i$, a, b, and t), and thus directly estimate NIRvH1 using Eq. 4 and the fitted *a, b, t* and λ in the NIR band (e.g., 778 nm).

Considering the delicate atmospheric correction of high spectral resolution data, using the "atmospheric windows" is a more pragmatic strategy to avoid these errors. Therefore, we used the wavelengths at the “atmospheric windows” instead of the full spectral range in our NIRvH1 approach.

NIRvH2 was developed by assuming linear soil spectra at the red edge (Wang et al., 2017; Yang et al., 2019). Although this assumption is not universally accurate, it is reasonable for the most common cases. With this assumption, NIRvH2 can be written as

$$NIRvH2= BRF_T(NIR)-BRF_T(R)-k\cdot(\lambda_{NIR}-\lambda_R) \quad (6)$$

where $BRF_T(R)$ and $BRF_T(NIR)$ are the reflectance in the red and NIR bands, respectively (Fig. 4). According to the typical shapes of soil and vegetation spectrum shapes, λ_R is set to be 678 nm, while λ_{NIR} could be within 778~800 nm for the best performance. BRF_T(678 nm) is contributed mostly by BRF_S(678 nm) (the product of the bidirectional gap fraction '*b*' and the soil reflectance R_S(678 nm)) in Eq. 1, and thus by the fitting with the baseline reflectance BRF_T(678 nm) in Fig. 4, we do not need to know either '*b*' or R_S(678 nm), separately. The slope parameter, *k*, can be estimated by linear fitting several BRF_T against wavelengths in either one of the two leaf spectral-invariant regions: red band (675~681 nm) or NIR band (778~800 nm) as discussed in Section 2.1, depending on the data availability. Compared to NIRvH1, NIRvH2 is simpler and does not require the neighbouring pure soil pixels or a soil spectral database to extract the *PCs*. Hyperspectral reflectance is not necessary for NIRvH2 because the slope *k* can be derived from a few multispectral bands in the two spectral regions. However, the performance of NIRvH2 depends on linearity of the soil spectrum at the red edge, and thus has a narrower range of applications than NIRvH1.

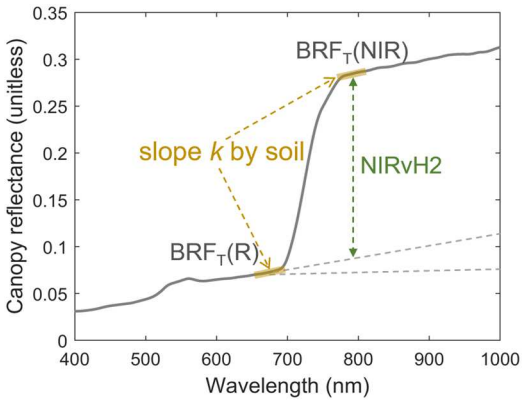


Fig. 4 A diagram demonstrating the calculation of the NIRvH2 index from the canopy total reflectance by the baseline reflectance $BRF_T(R)$ at 678 nm, and by the slope *k* (Eq. 6) from either the red region (675~681 nm) or the NIR region (778~800 nm).

3. Materials and methods

3.1 Model-based evaluation with the SCOPE model

In order to test the performance of NIRvH1 and NIRvH2, the Soil Canopy Observation, Photochemistry and Energy (SCOPE) model version 1.70 (van der Tol et al., 2009) was used to simulate canopy reflectance and SIF with various combinations of canopy structure, leaf property, sun-sensor geometry and soil spectrum (Table 1). Four soil spectra with different brightness levels from the soil database in the SCOPE model (Fig. 2) were used in the simulations. All the other parameters were kept as the default values in SCOPE v1.70. In total, 20,736 different combinations were generated. The broadband incoming shortwave radiation (400-2500 nm) was set at 600 W m^{-2}. The simulated reflectance covered 400~1000 nm with a sampling interval of 1 nm, and simulated SIF at the O_2-A band (760 nm) was used for the analysis. NIR_{BS} was obtained by replacing the soil with a black-body background in the model setup. We chose the central wavelengths of the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor at the red band (648 nm) and NIR band (858 nm) (Vermote et al., 1997) to calculate NIRv and DVI.

In order to evaluate the practical performance of NIRvH1 and NIRvH2, we chose to use the spectral range of 675~775 nm to be identical to the spectral coverage of existing hyperspectral satellite sensors, e.g., TROPOMI. The four soil spectra in Fig. 3 were used in the SVD method for the NIRvH1 calculation. For calculating NIRvH2, we set λ_{NIR}=775 nm. The slope parameter k in Eq. 6 was estimated using the simulated red band (675~681 nm) canopy reflectance to be consistent with TROPOMI.

We compared four simulated VIs (NIRvH1, NIRvH2, DVI and NIRv) to evaluate their performance in reducing the soil impacts. To highlight the conditions of sparse vegetation, we first examined the results where SIF was no more than 1 W m^{-2} um^{-1} sr^{-1} and all the VIs were no more than 0.3. Both SIF and NIR_{BS}, which had little soil influence, were used as the benchmark to evaluate the soil impacts on the four VIs. The offsets (i.e., the value of a

vegetation index when SIF (or NIR_{BS}) is zero) in the VI-SIF (or VI-NIR_{BS}) relationship, were used to indicate the impact of soil brightness on VIs. For example, if the linear fitting line between a VI and SIF or NIR_{BS} goes through the origin point of coordinates (0, 0), the offset is zero and suggests promising performance of the VI in reducing the soil influence. In addition, the relative offset (RO) was also calculated as the ratio between the offset and the mean of the VI. Similar analyses were also conducted with the model simulations where NIRvH1 and NIRvH2 were larger than 0.3 for the evaluation over dense canopies.

As suggested in several earlier studies (Zeng et al., 2019; Yang et al., 2020; Hao et al., 2021a), pure vegetation reflectance has similar sun-sensory geometry effects as SIF and thus can be used to correct the anisotropy of SIF. Therefore, a VI that can well approximate pure vegetation reflectance should demonstrate good performance on correcting the anisotropy of SIF observations at different sun-sensor geometries. With the setting of varied sun-sensor geometry in our SCOPE simulations, we evaluated the four VIs for their ability of correcting the anisotropy of SIF in two approaches. One approach was to calculate the total emitted SIF by all leaves (SIF_{Total}) from top-of-canopy SIF (SIF_{TOC}) with different sun-sensor geometries: $SIF_{Total} = SIF_{TOC}/fesc$, and compared it with the true value provided by model simulation. SIF_{Total} can be estimated from SIF_{TOC} obtained under a sun-sensor geometry by dividing SIF_{TOC} by a photon escape ratio $fesc = VI/(fPAR \cdot w)$, where VI is one of the four abovementioned VIs under the same sun-sensor geometry of SIF_{TOC}, fPAR is the fractional of absorbed photosynthetically active radiation (PAR) from SCOPE in this study, and w is the leaf single scattering albedo at 760 nm (Zeng et al., 2019; Dechant et al., 2020). The second approach was to normalize SIF_{TOC} to the nadir viewing direction (SIF_{Nadir}), with the corresponding VIs at the same angle of SIF_{TOC} (by VI_{TOC}) and nadir VIs (VI_{Nadir}) (Zeng et al., 2019; Hao et al., 2021a, 2021b): $SIF_{Nadir}=SIF_{TOC}*VI_{Nadir}/VI_{TOC}$. The calculated SIF_{Nadir} can be compared with the true value provided by model simulations to evaluate the anisotropy of

VI in comparison to SIF. The first approach may have systematic biases due to different behaviours of SIF and VI over the bare soil, because VI may not be equal to zero when SIF is zero. The second approach highly depends on the anisotropy distribution of SIF and VI, while the soil impacts on VIs could vary with the sun-sensor geometry with different fractions of sunlit/shaded soil in view.

Table 1. List of variables and their ranges used in SCOPE v1.70 simulations of canopy reflectance and fluorescence (Zeng et al., 2019), resulted in 20,736 combinations. Default values were used for all the other variables.

	Variable	Values
Canopy Structure	Leaf Area Index (LAI, m^2 m^{-2})	[0.5, 1, 3, 5]
	Leaf Angle Distribution	Spherical, Erectophile, Planophile
Leaf Property	Leaf Chlorophyll Content (Cab, μg cm^{-2})	[40, 60, 80]
	Dry Matter Content (Cm, g cm^{-2})	[0, 0.025, 0.05]
	Brown Pigments (Senescent material fraction, Cs)	[0, 0.3, 0.6, 0.9]
Sun-sensor Geometry	Solar Zenith Angle	[30°, 50°]
	View Zenith Angle	[0°, 30°, 50°]
	Relative Azimuth Angle	[0°, 180°]
Soil Background	Soil Spectra	Four soil spectra

3.2 HyPlant airborne dataset

We also used hyperspectral and SIF measurements from an airborne remote sensing experiment with the HyPlant sensor to evaluate NIRvH1 and NIRvH2. The HyPlant sensor

(Specim, Finland) is an airborne imaging spectrometer for vegetation monitoring with two sensors operating in push-broom mode (Rascher et al., 2015; Siegmann et al., 2019). The fluorescence imager (FLUO) module of HyPlant acquires radiance at high spectral resolution (0.25 nm) in the spectral region of the two oxygen absorption bands (670~780 nm), enabling simultaneous acquisition of SIF and reflectance with consistent geometry. The DUAL module of HyPlant provides imageries covering the spectral range of 380~2500 nm with the spectral resolution at about 1.7 nm at the red edge. In this study, four HyPlant airborne imageries were acquired with the flight altitude of 600 m above the ground surface, including three regions of agricultural fields (Region A~C), and one region of complex land surfaces (Region D). The imageries were acquired on August 23rd, 2012 (Region A: 50°52'37.52"N/ 6°26'41.89"E, sugar beet), June 30th, 2015 (Region B: 50°52'33.42"N/ 6°26'10.83"E, potato, winter wheat and sugar beet), July 19th, 2016 (Region C: 50°52'43.95"N/ 6°26'11.79"E, sugar beet, potato and maize), and August 23rd, 2012 (Region D: 50°54'39.10"N/ 6°26'41.89"E, forest with woody components, water body and urban areas), respectively (Fig. S3).

The improved Fraunhofer Line Discrimination (iFLD) module was adopted to retrieve SIF at the O_2–A band (760 nm) from the observed radiance (Damm et al., 2011). The HyPlant imageries were atmospherically corrected using the MODTRAN radiative transfer model to obtain the relevant atmospheric transfer functions and enable the calculation of surface reflectance (Berk et al., 2005). The spatial resolution for reflectance and SIF was originally at 1 m but was aggregated to 5 m to reduce the noise and compensate the impact of a wide spatial-spectral point spread function. For NIRvH1, we chose the same wavelength of TROPOMI covering 675~775m. Again, only the wavelengths at the atmospheric windows within 675~775 nm of the HyPlant FLUO surface reflectance data were used for the NIRvH1 (cf. Section 2.3 for a justification). The soil pixels for the SVD method in NIRvH1 were

selected by a criterion which has a threshold value of 0.2 for NDVI at the same HyPlant imagery. For NIRvH2, we set λ_{NIR}=775 nm and extracted *k* in Eq. 6 with linear fitting of the total reflectance within the range of 675~681 nm or 778~800 nm. Because NIR_{BS} cannot be directly acquired by the airborne data, we only used far-red SIF at 760 nm as the reference. NIRv and DVI were also included in the HyPlant analysis to compare with NIRvH1 and NIRvH2. Reflectance at 648 nm (red) and 858 nm (NIR), which were identical to the central wavelengths of MODIS, were used to calculate NIRv and DVI.

4. Results

4.1 Evaluation by SCOPE simulations

All the four VIs demonstrated strong linear relationships with SIF and NIR_{BS}, and the coefficient of determination (R^2) was no less than 0.79 (Fig. 5) by the SCOPE simulations at different combinations of canopy structure, leaf property, sun-sensor geometry and soil optical properties in Table 1. NIRvH1 and NIRvH2 stood out and showed the strongest and similar correlations with SIF and NIR_{BS}, with the highest R^2 and lowest RMSE, as well as the smallest offset and RO. In contrast to the small offsets of NIRvH1 and NIRvH2, DVI and NIRv had variable and larger offsets in the SIF-VI or NIR_{BS}-VI relationships in response to the four soil spectra with different levels of brightness in the SCOPE simulations. ROs of DVI and NIRv were also higher (Fig. 5). In the SIF-VI relationship, ROs were 17.7% and 10.4% for DVI and NIRv, while they were 4.2% and 2.1% for NIRvH1 and NIRvH2, respectively. Similarly, in the NIR_{BS}-VI relationship, ROs were 19.5% and 11.6% for DVI and NIRv, and were 4.8% and 4.6% for NIRvH1 and NIRvH2, respectively.

Larger offsets were found for DVI and NIRv (Figs. 5~6) with brighter soil backgrounds, which result in steeper red-edge slopes in the TOC reflectance spectra (Fig. 2). The largest

offset over the brightest soil in the SIF-VI relationship for DVI and NIRv could be as large as 0.049 and 0.028, respectively, while the largest offset for NIRvH1 or NIRvH2 was generally less than 0.013 (Fig. 6a). The largest offset over the brightest soil in the NIR_{BS}-VI relationship for NIRvH1 and NIRvH2 was 0.013 and 0.008, and was 0.029 and 0.050 for NIRv and DVI, respectively (Fig. 6b). Over dense canopies when all the four VIs were larger than 0.3, NIRvH1 and NIRvH2 still performed better than DVI and NIRv in approximating NIR_{BS}, with a slightly higher R^2 and lower RMSE when compared to NIR_{BS} (Fig. S4).

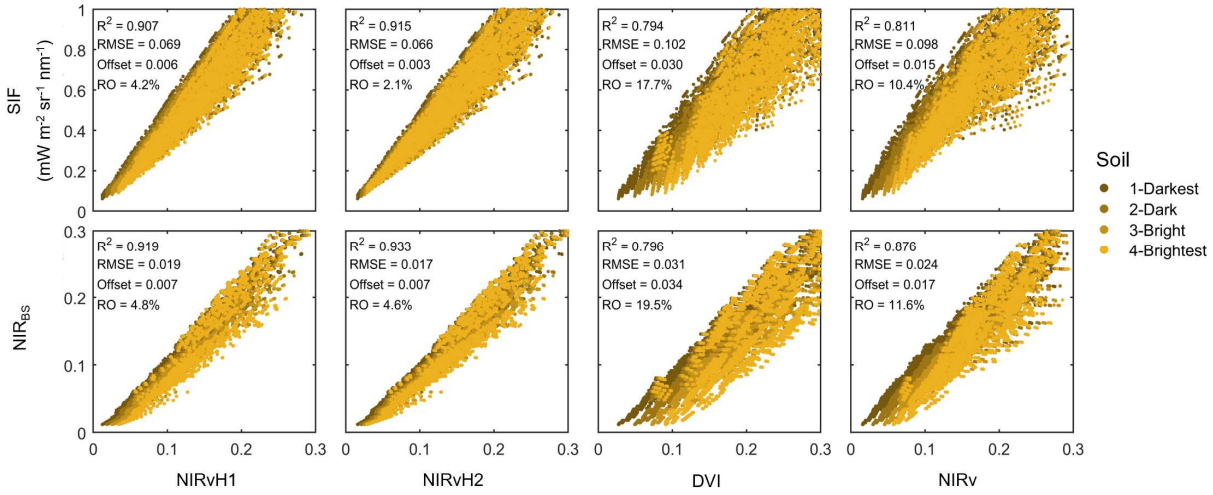


Fig. 5 Scatter plots between SIF at 760 nm or NIR_{BS} and different VIs by SCOPE simulations at different combinations of canopy structure, leaf property, sun-sensor geometry and soil spectrum in Table 1. To highlight the condition of sparse vegetation, we first examined the results where SIF was no more than 1 W m^{-2} um^{-1} sr^{-1} and all the four VIs were no more than 0.3. Note there were four soil spectra in the simulation, which led to four different offsets for the SIF-VI or NIR_{BS}-VI relationships for DVI and NIRv as in Fig. 6. The relative offset (RO) was estimated by the ratio of the offset and the mean of the corresponding VI.

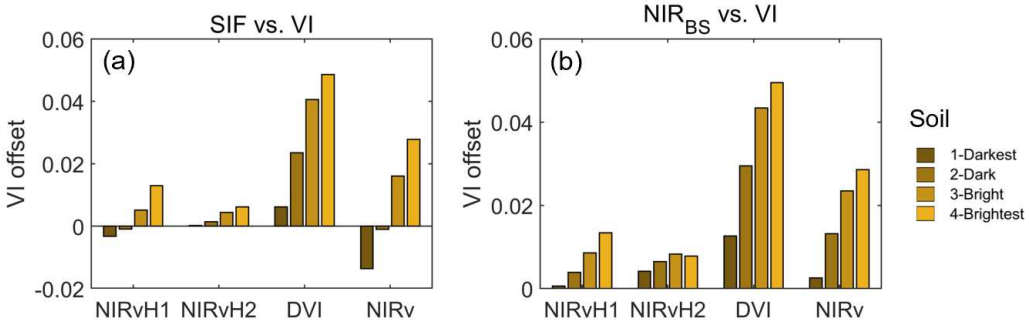


Fig. 6 Offsets in the relationship between SIF at 760 nm or NIR_{BS} with different VIs (Fig. 4) over four different soil brightness by SCOPE simulations. The simulations were conducted with different combinations of canopy structure, leaf property, sun-sensor geometry and soil reflectance spectra in Table 1.

All the four VIs demonstrated good performances in the SIF anisotropy correcting experiments with SCOPE, i.e., the R^2s between the corrected and simulated data were no lower than 0.78 (Fig. 7). NIRvH1 and NIRvH2 had higher R^2s and lower RMSEs than DVI and NIRv for both SIF_{Total} and SIF_{Nadir}. For sparse vegetation canopies (e.g., LAI=0.5 m^2 m^{-2}), using DVI and NIRv may result in underestimated SIF_{Total} by over 30% and 10%, respectively (Fig. 8). This is due to the overestimations of *fesc* by DVI and NIRv, caused by the residual soil effects. SIF_{Nadir} using DVI and NIRv were overestimated by about 20% and 10% at maximum, again due to the residual soil contributions that led to different anisotropic distributions between SIF and DVI or NIRv. The residual soil contributions to DVI or NIRv changed with the sun-sensor geometry, caused by different fractions of sunlit/shaded soil in view. In contrast, we found the relative error was less than 10% for both SIF_{Total} and SIF_{Nadir} over all the sun-sensor geometries if NIRvH1 or NIRvH2 were used.

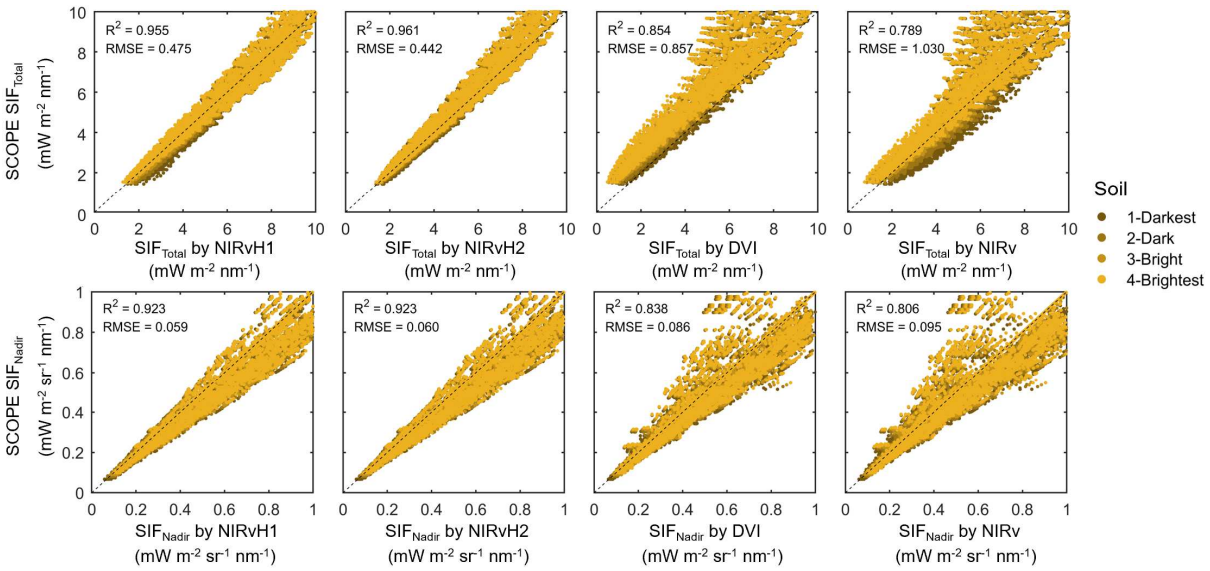


Fig. 7 Scatter plots between SIF_{Total} or SIF_{Nadir} at 760 nm normalized by different VIs and SCOPE references at different combinations of canopy structure, leaf property, sun-sensor geometry and soil spectrum in Table 1.

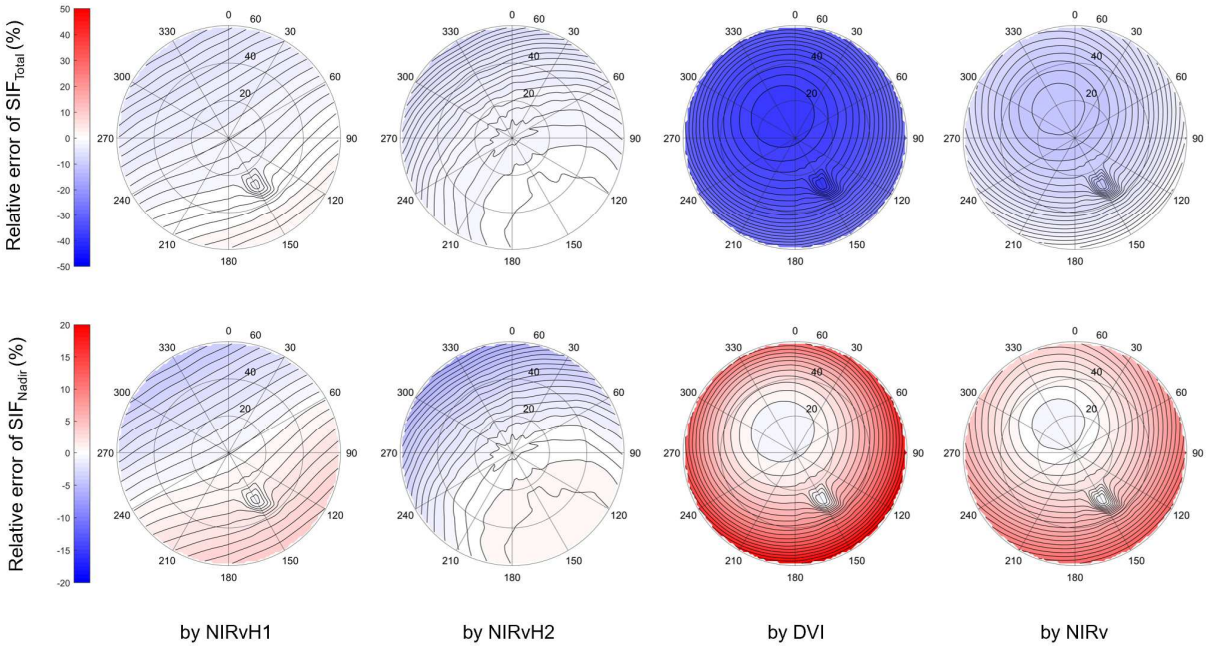


Fig. 8 Relative error of SIF_{Total} or SIF_{Nadir} normalized by different VIs over sparse canopies when compared to SCOPE references at different viewing geometries, with the solar zenith angle at 30°, the leaf area index at 0.5 m^2 m^{-2}, the same soil background (dried soil, 4-Brightest in Fig. 2), spherical leaf angle distribution, the chlorophyll content at 60 μg cm^{-2}, the carotenoid content at 20 μg cm^{-2}, the dry matter content at 0.025 g cm^{-2}, the equivalent water thickness at 0.009 cm, the senescent material fraction at 0.3, and the leaf structure parameter N at 1.4. The contour lines divide each plot into equal intervals, and the color from blue to red represents the relative error from negative to positive.

4.2 Evaluation by airborne HyPlant dataset

The performance of the NIRvH1 calculation on the spectral fitting by the SVD method and the logistic function at three spectra from the HyPlant imagery was displayed in Fig. 9. The displayed three spectra of soil or total reflectance were sampled at 20%, 50%, and 80% percentiles of the data (Fig. 9). The increase of HyPlant-measured soil reflectance was not strictly linear at the red edge (675~775 nm). The first *PC* by the SVD method explained 98%

of the soil spectrum variance for the pure soil pixels at region A, and only the first *PC* was thus used in the fitting. The results show that the SVD-reconstructed soil reflectance agreed well with the HyPlant-measured soil reflectance (Fig. 9a). After the removal of the soil contribution by the SVD method, the remaining HyPlant vegetation reflectance of the three pixels agreed well with the shape of the leaf spectrum, and could be well fitted by a logistic function (Fig. 9b). Overall, the HyPlant total reflectance of the three pixels with different soil brightness as shown in the red band (Fig. 9c), can be well fitted by the SVD method and the logistic function in the NIRvH1 calculation.

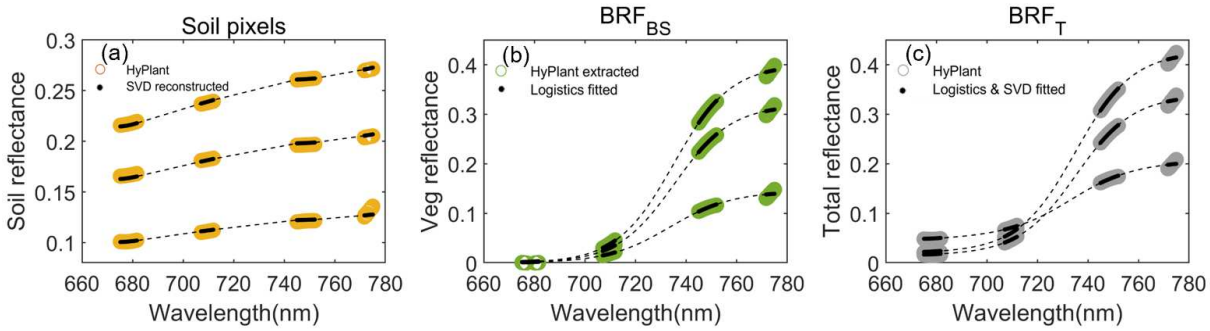


Fig. 9 The performance of the fitting by the singular value decomposition (SVD) method and the logistic function in the NIRvH1 calculation at three spectra from the HyPlant fluorescence imager at region A in Fig. S3. (a) The HyPlant measured reflectance of three soil pixels representing the 20%, 50%, and 80% percentiles of the soil reflectance values, and the corresponding SVD reconstructed reflectance. (b) The extracted HyPlant vegetation reflectance (BRF_{BS}) after the removal of the soil contribution by the SVD approach at three mixed pixels representing the 20%, 50%, and 80% percentiles of the total reflectance values, and the corresponding logistic function-fitted vegetation reflectance. (c) The HyPlant observed total reflectance (BRF_T) at the three mixed pixels, and the corresponding fitted total reflectance by the SVD method and the logistic function in the NIRvH1 calculation. Note only the wavelengths in the atmospheric windows with high transmittance around 680 nm, 710 nm, 750 nm and 775 nm were used.

Scatter plots in Fig. 10 show the comparison between HyPlant-measured SIF and different Vis. In the three agricultural field regions (region A~C), all the four VIs showed strong linear

484 correlations with SIF (R^2 > 0.73). Positive offsets were found for all the four VIs in the SIF-
485 VI relationship over sparse vegetation canopies. DVI and NIRv had higher ROs at the levels
486 of 26.9%~37.1% and 16.3%~23.7%, respectively. In contrast, ROs for NIRvH1 and NIRvH2
487 were only 0.1%~10.5% and 3.2%~12.8%, respectively. NIRvH1 and NIRvH2 had the
488 smallest offsets within 0~0.028 and 0.008~0.037, respectively. NIRvH1 performed slightly
489 better than NIRvH2 because the shape of soil reflectance spectra was not strictly linear (Fig.
490 10a). NIRv had an intermediate offset within 0.047~0.062, and DVI had the largest offsets
491 ranged between 0.089 and 0.112, two to three folds higher than NIRvH1 and NIRvH2. In the
492 region D with more complex land surfaces, NIRvH1 and NIRvH2 kept showing better
493 performance than DVI and NIRv, with higher R^2, lower RMSE, offset and ROs. ROs for
494 NIRvH1 and NIRvH2 were only 4.1% and 0.4%, respectively, while for DVI and NIRv were
495 as large as 20.7% and 11.9%, respectively. Overall, Fig. 10 suggests that the traditional DVI
496 and NIRv are not promising at removing soil impacts with non-zero offset in their linear
497 regression line with SIF, likely due to the increasing shape of soil reflectance at the red edge.
498 By contrast, the soil effect was effectively reduced with the proposed NIRvH1 and NIRvH2.

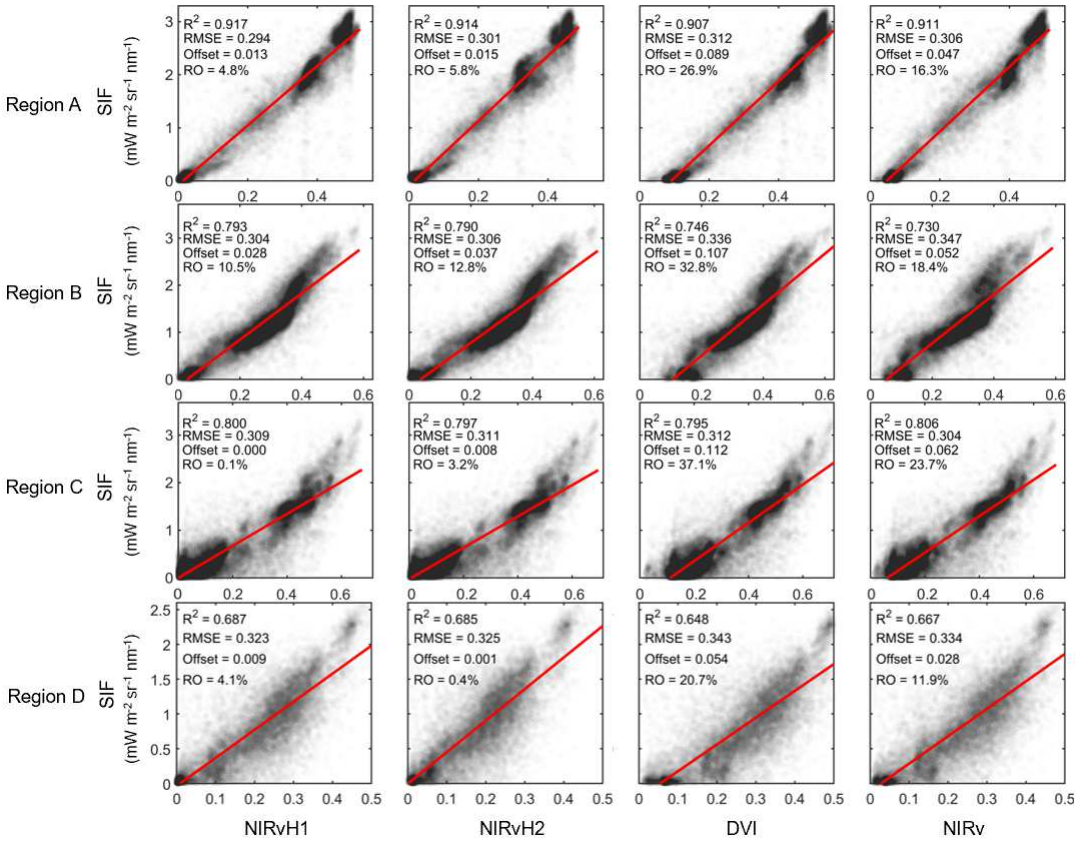


Fig. 10 Scatter plots of SIF at 760 nm and different VIs for HyPlant imageries at four study regions on August 23rd, 2012, June 30th, 2015, July 19th, 2016 and August 23rd, 2012, respectively. Regions A~C were agricultural fields, while region D was over complex land surfaces composed of forest, water body and urban areas. The relative offset (RO) was estimated by the ratio of the offset to the mean value of the corresponding VI in the SIF-VI relationship.

5. Discussion

5.1 Towards better separation of soil and vegetation contributions in canopy reflectance by NIRvH

At present, NIRv and DVI are considered to be effective approaches for minimizing soil impacts and are widely used in measuring vegetation growing status. This may be of particular concern when these VIs are used to reduce structural effects inherent to SIF retrievals (Zeng et al., 2019; Yang et al., 2020; Joiner et al., 2020). Here we found that they

514 may have varying positive offsets when compared to SIF and BRF_{BS}, i.e., positive values at
515 bare soil condition. This is mainly caused by the increasing soil reflectance at the red edge
516 (Figs. 5 and 9), and thus soil still partly contributes to the difference between NIR and red
517 reflectance, which was not fully accounted in their mathematical formulas. Furthermore,
518 different sensor settings may introduce an additional level of complexity due to sampling red
519 and NIR reflectance at different spectral wavelengths (e.g., MODIS at 648 nm and 858 nm,
520 TROPOMI at 675 nm and 775 nm for red and NIR central wavelength, respectively) which
521 are associated with different soil reflectance.

522 Sellers et al (1992) emphasized the importance of disentangling the individual
523 contributions from vegetation and soil in the red and NIR, and that this was particularly
524 challenging with multispectral data especially over sparse vegetation canopies. We made full
525 use of the shape of the soil and leaf spectrum at the red edge covered by hyperspectral data,
526 and developed the NIRvH1 and NIRvH2. Both indices show strong linear correlations with
527 SIF and BRF_{BS} (Figs. 5 and 10), and can be directly used for the angular correction of
528 directional observed SIF with good performance (Figs. 7 and 8). These two indices are
529 promising for improving recent studies with NIRv, especially over sparse vegetation canopies
530 at agricultural, forest and urban ecosystems (Fig. 10). While the goal of reducing the soil
531 contribution in BRF can also be achieved with a physically-based canopy radiative transfer
532 model, this requires retrieval of soil and vegetation structural parameters (e.g., leaf area index
533 and leaf angle distribution) and thus introduces additional complexity, and may become
534 challenging over heterogeneous land surfaces which violate the assumptions of the model. By
535 contrast, NIRvH1 and NIRvH2 completely avoid such additional complexity by taking
536 advantage of the features of soil and vegetation spectra, and they show better performance
537 than DVI and NIRv over sparse canopies (Fig. 5). NIRvH1 and NIRvH2 also performed
538 better than DVI and NIRv over dense canopies when the soil has minimal impacts (Fig. S4),

539 possibly due to that DVI (=NIR-Red) and NIRv (=NIR·NDVI) were not equal to NIR with
540 the additional term of Red (≠0) and NDVI (≠1) in DVI and NIRv, respectively.

541 Table 2 summarizes the major characteristics of the new (NIRvH1 and NIRVH2) and
542 traditional vegetation indices (DVI and NIRv) discussed in this study. NIRvH1 has the
543 advantage that no assumption on the spectral shape of background soil is required and thus is
544 more flexible. However, NIRvH1 needs the spectral information of the nearby pure soil
545 pixels or at least a soil spectral database to drive the SVD approach. In addition,
546 hyperspectral data are required which also limits the applicability of NIRvH1 in practice.
547 NIRvH2 is simple to use and does not need any other additional information such as the
548 nearby pure soil pixels or a soil spectral database. NIRvH2 also does not necessarily require
549 hyperspectral data, as long as the slope k in Eq. 6 can be derived from a few red and NIR
550 bands by multispectral sensors based on the assumption of a linearly increasing soil
551 reflectance at the red edge. This assumption may not always be true but is reasonable for
552 most soil types (Wang et al. 2017; Yang et al. 2019). In addition, both NIRvH1 and NIRvH2
553 need to use reflectance measurement at 678 nm, the central wavelength of the chlorophyll
554 absorption peak (Fig. S2a) which is the least sensitive to chlorophyll content and has the local
555 minimal reflectance in the red band (Fig. 1). However, the central wavelength of the red band
556 of many existing multispectral sensors is away from 678 nm (e.g., it is 648 nm for MODIS)
557 in red. Cautions must be taken that NIRvH1 and NIRvH2 calculated with a red band
558 departing from 678 nm may not have the best performance on removing soil impacts because
559 of red band's sensitivity to the chlorophyll content (Fig. 1).

560 In addition to NIRv and DVI, some other vegetation indices have been proposed to
561 reduce the soil impacts, such as the soil-adjusted vegetation index (SAVI) (Huete et al., 1988),
562 enhanced vegetation index (EVI) (Huete et al., 2002), plant phenology index (PPI) (Jin and
563 Eklundh, 2014), normalized difference phenology index (NDPI) (Wang et al., 2017) and

normalized difference greenness index (NDGI) (Yang et al., 2019). They are widely used in the time-series analysis of phenology, but it is difficult to link their absolute magnitudes with the photon escape ratio (*fesc*) of SIF, and thus they cannot be directly used to calculate the total emitted SIF by directional observed SIF. Thus we mainly focus on the improvement of NIRv and DVI with hyperspectral dataset as a proxy of "potential SIF".

Table 2 Major characteristics of the four vegetation indices: NIRvH1, NIRvH2, DVI and NIRv on soil impacts as candidates for NIR_{BS}.

Remote sensing indices	Reduce the soil effect?	Influenced by the increase of soil reflectance?	Require hyperspectral data?	Other characteristics
NIRvH1	Yes	No	Yes	Requires soil pixel or a soil spectral database for SVD; Narrow-band at 678 nm required
NIRvH2	Yes	No	Not necessarily	No additional information needed; Simple; Narrow-band at 678 nm required
DVI	Partially	Yes	No	Maximum offset on the SIF-VI relationship
NIRv	Partially	Yes	No	Medium offset on the SIF-VI relationship

5.2 Uncertainties and requirements in NIRvH estimations

Multiple scattering between the vegetation and soil, i.e., BRF_M in Eq. 1, also contributes to the canopy total reflectance in both leaf spectral-invariant regions at 675~681 nm and 778~800 nm. In this study, BRF_M has been partially included in NIRvH (NIRvH1 and NIRvH2) by the curve fitting, enabling a better correlation between NIRvH and SIF or NIR_{BS} than the widely-used NIRv (Fig. 5). However, the importance of including the multiple scattering contributions in NIRvH depends on the specific application scenarios. For example, if the research goal is to use NIRvH for estimating GPP, to normalize SIF and retrieve Φ_F, or

to correct SIF for view-angle effects, BRF_M should be included in NIRvH, since soil reflected photons can be reabsorbed by the canopy, and then excite fluorescence and trigger photosynthesis. When applying NIRvH for retrieving canopy structural parameters (e.g., LAI, leaf inclination angle and clumping index), BRF_M should be excluded in the calculation of the NIRvH, because all the soil contributions, no matter from single or multiple scattering, belong to the noise signal for estimating canopy structural parameters.

Leaf reflectance at 678 nm is small but has a non-zero positive value (usually <0.1 and mostly <0.05, Fig. 1). This is because non-photosynthetic biochemical constituents within the leaves, such as the dry matter and brown pigments, always scatter a small proportion of incident radiation at 678 nm. NIRvH2 not only reduces the soil contribution from the canopy total reflectance, but also reduces the scattering by non-chlorophyll leaf constituents. In fact, all scattering contributions by non-chlorophyll components can be reduced from BRF_T(NIR) by NIRvH. In principle, NIRvH is robust and flexible for various application scenarios. As long as the background spectra of non-chlorophyll components (e.g., soil, litter, impervious surfaces and water body) can be identified in nearby pixels (for NIRvH1), or the slope of the background spectra varies minimally in the range of 675~800 nm (for NIRvH2), NIRvH is applicable over either natural or urban ecosystems (Fig. S3). Note that NIRvH can only separate the NIR contribution by chlorophyll and non-chlorophyll components, while cannot distinguish the chlorophyll contribution from leaves, woody parts (e.g., branches, stems and bark) and understory (e.g., moss and lichen). This allows NIRvH and SIF have similar radiative transfer process and results in the strong correlation between NIRvH with SIF (Figs. 5 and 10), although differences and uncertainties could come from the contribution of brown pigments that cause a slight positive slope in the NIR region of 778~800 nm and violates the assumption of flat canopy NIR reflectance (Fig. S1).

We recommended using two leaf spectral-invariant regions in the red and NIR bands for calculating NIRvH. In fact, the blue band at 450~480 nm is another spectral-invariant region that is not sensitive to canopy structure variations and potentially useful for calculating NIRvH. However, it may be challenging to use the blue band to calculate NIRvH, particularly NIRvH2, for two reasons. First, it might not be reasonable to assume soil reflectance linearly increases over a large spectral region from blue to red edge, and thus does not meet the prerequisite of the NIRvH2 algorithm. Second, atmospheric correction at the blue band is more complicated due to aerosol scattering and remaining larger artefacts than the red and NIR bands (Vermote et al., 1997). However, the NIRvH algorithm can still be promising to reduce the soil impacts of other vegetation indices at the canopy scale, such as the Photochemical Reflectance Index (PRI) (Gamon et al., 1992) and Chlorophyll/Carotenoid Index (CCI) (Gamon et al., 2016), which have typically been applied with airborne or ground-based measurements with limited atmospheric scattering impacts on the blue band.

5.3 Potential applications of the hyperspectral-based NIRvH

NIRvH could be useful to decouple the soil and vegetation contribution in mixed pixels for current and forthcoming hyperspectral missions. TROPOMI covers 675~775 nm (Guanter et al., 2015) but not the largest chlorophyll absorption wavelength at 778 nm. However, applying both of our NIRvH approaches within the TROPOMI's spectral coverage range could still perform well, as indicated by SCOPE experiments (Figs. 6 and 9). The forthcoming FLEX mission covers a wider spectral range of 500~780 nm (Drusch et al., 2016). The new hyperspectral instrument German Aerospace Center (DLR) Earth Sensing Imaging Spectrometer (DESIS) onboard the International Space Station (ISS) covers the wavelength of 400~1000 nm with the spectral resolution of 2.5 nm (Krutz et al., 2019), which

will be promising for further validating our NIRvH approaches especially by including the spectral-invariant NIR region of 778~800 nm. However, similar to other VIs for reducing soil impacts such as NDGI (Yang et al. 2019), NIRvH needs inputs from at least two red or NIR wavelengths/bands. This suggests that NIRvH designed for hyperspectral sensors is not applicable for most existing multispectral sensors/satellites such as MODIS and Landsat-8. Nevertheless, the NIRvH2 framework may be applicable with Sentinel-2 measurements which have 2 bands in NIR regions to partially reduce the soil impacts, although both Sentinel-2 NIR bands are beyond 800 nm and could introduce additional uncertainties by extrapolating Eq. 6.

The reduced sensitivity of the NIRvH for soil background compared to the NIRv and its strong linear correlation to SIF enables it to approximate "potential SIF". This may open many important applications, including the normalization of SIF and obtaining canopy-scale SIF yield (Φ_F). Zeng et al (2019) demonstrated that SIF and NIRv radiance (Baldocchi et al., 2020; Wu et al., 2020) were related through their joint dependence on canopy structure and sun-sensor geometry. The residual between the two measurements (NIRv or NIRvH radiance and SIF) should be directly related to Φ_F (Wang et al., 2020). Variations in Φ_F relate to the differential partitioning of absorbed PAR between photochemical and non-photochemical processes at the photosystem level. While the difference in performance between NIRvH and NIRv might be minor, small uncertainties could propagate directly into larger errors in Φ_F. When combined with high-frequency SIF and eddy covariance measurements, NIRvH can be used to infer plant responses to environmental changes and stresses such as drought and thus improve our mechanistic understanding of the underlying physiological and ecological processes. Furthermore, since SIF measurements are usually sparse and coarse, NIRvH, as the "potential SIF", is a good candidate as the reference for estimating SIF at a higher spatio-temporal resolution. This would also allow us to look back in time and possibly extend SIF

time series to the past when suited hyperspectral data (e.g., Hyperion) or applicable multi-spectral data were available.

6. Conclusions

Our study demonstrates that the widely used NIRv and DVI have a considerable dependence on the soil background, and that this translates into uncertainties in downstream applications such as estimates of GPP or the retrieval of Φ_F. We showed that a rigorous exploitation of spectral invariant regions in the red and NIR enables the design of new VIs (NIRvH1 and NIRvH2) that show substantially reduced sensitivity for soil background contributions and enables decoupling the combined contributions from vegetation and soil. We recommend using hyperspectral data for the NIRvH retrieval since such data enable a more robust usage of spectral invariant wavelength regions to estimate the spectral shape of soil background reflectance and to compensate for this contribution. The newly designed NIRvH can be considered as a robust proxy of "potential SIF" and enables various applications to contribute to ecological research. We suggest evaluating the potential of NIRvH to directly approximate GPP, to normalize SIF retrievals for structural effects and illumination effects, and to retrieval Φ_F. The NIRvH also holds potential to fill gaps in sparse spatio-temporal sampling of actual SIF and could even be exploited to extend available SIF time series to the past. These fields of applications are hypothetical, and we recommend further detailed experiments to exploit these possibly highly interesting application fields.

Acknowledgement

This research was supported by the National Aeronautics and Space Administration (NASA) through Remote Sensing Theory and Arctic Boreal Vulnerability Experiment (ABoVE) grants 80NSSC21K0568 and NNH18ZDA001N granted to Min Chen. Y. R. was supported

678 by National Research Foundation of Korea (NRF-2019R1A2C2084626). Airborne
679 acquisition and data analysis were financed by the European Space Agency (ESA) in the
680 frame of the HyFLEX campaign (ESA contract No. 4000107143/12/NL/FF/If) and the
681 Photoproxy campaign (ESA contract No. 4000125731/19/NL/LF). This work was partially
682 funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under
683 Germany's Excellence Strategy – EXC 2070 – 390732324.

684

685 References

686 Alonso, L., Gómez-Chova, L., Vila-Francés, J., Amorós-López, J., Guanter, L., Calpe, J.,
687 Moreno, J. 2008. Improved Fraunhofer Line Discrimination method for vegetation
688 fluorescence quantification. IEEE Geoscience and Remote Sensing Letters, 5, 620-624.
689 Asner, G. P., Heidebrecht, K. B. 2002. Spectral unmixing of vegetation, soil and dry carbon
690 cover in arid regions: comparing multispectral and hyperspectral observations.
691 International Journal of Remote Sensing, 23(19), 3939-3958.
692 Badgley, G., Field, C.B., Berry, J.A. 2017. Canopy near-infrared reflectance and terrestrial
693 photosynthesis. Science Advances, 3, e1602244.
694 Badgley, G., Anderegg, L. D., Berry, J. A., Field, C. B. 2019. Terrestrial gross primary
695 production: Using NIRv to scale from site to globe. Global change biology, 25(11),
696 3731-3740.
697 Baldocchi, D. D., Ryu, Y., Dechant, B., Eichelmann, E., Hemes, K., Ma, S., ... Verfaillie, J.
698 2020. Outgoing Near Infrared Radiation from Vegetation Scales with Canopy
699 Photosynthesis Across a Spectrum of Function, Structure, Physiological Capacity and
700 Weather. Journal of Geophysical Research: Biogeosciences, e2019JG005534.
701 Berk, A., Anderson, G. P., Acharya, P. K., Bernstein, L. S., Muratov, L., Lee, J., ...
702 Lockwood, R. B. 2005, June. MODTRAN 5: a reformulated atmospheric band model

703 with auxiliary species and practical multiple scattering options: update. In Algorithms
704 and technologies for multispectral, hyperspectral, and ultraspectral imagery XI (Vol.
705 5806, pp. 662-667). International Society for Optics and Photonics.
706 Damm, A., Erler, A., Hillen, W., Meroni, M., Schaepman, M. E., Verhoef, W., Rascher, U.
707 2011. Modeling the impact of spectral sensor configurations on the FLD retrieval
708 accuracy of sun-induced chlorophyll fluorescence. Remote Sensing of Environment,
709 115(8), 1882-1892.
710 Dechant, B., Ryu, Y., Badgley, G., Zeng, Y., Berry, J. A., Zhang, Y., ... Li, J. 2020. Canopy
711 structure explains the relationship between photosynthesis and sun-induced chlorophyll
712 fluorescence in crops. Remote Sensing of Environment, 241, 111733.
713 Drusch, M., Moreno, J., Del Bello, U., Franco, R., Goulas, Y., Huth, A., ... Nedbal, L. 2016.
714 The fluorescence explorer mission concept—ESA's earth explorer 8. IEEE Transactions
715 on Geoscience and Remote Sensing, 55(3), 1273-1284.
716 Féret, J.-B., Gitelson, A., Noble, S., Jacquemoud, S. 2017. PROSPECT-D: towards modeling
717 leaf optical properties through a complete lifecycle. Remote Sensing of Environment,
718 193, 204-215
719 Feret, J.-B., François, C., Asner, G.P., Gitelson, A.A., Martin, R.E., Bidel, L.P., Ustin, S.L.,
720 Le Maire, G., Jacquemoud, S. 2008. PROSPECT-4 and 5: Advances in the leaf optical
721 properties model separating photosynthetic pigments. Remote Sensing of Environment,
722 112, 3030-3043
723 Gamon, J. A., Penuelas, J., Field, C. B. 1992. A narrow-waveband spectral index that tracks
724 diurnal changes in photosynthetic efficiency. Remote Sensing of environment, 41(1),
725 35-44.
726 Gamon, J. A., Huemmrich, K. F., Wong, C. Y., Ensminger, I., Garrity, S., Hollinger, D. Y., ...
727 Peñuelas, J. 2016. A remotely sensed pigment index reveals photosynthetic phenology in

evergreen conifers. Proceedings of the National Academy of Sciences, 113(46), 13087-13092.

Guanter, L., Aben, I., Tol, P., Krijger, J., Hollstein, A., Köhler, P., Damm, A., Joiner, J., Frankenberg, C., Landgraf, J. 2015. Potential of the TROPOspheric Monitoring Instrument (TROPOMI) onboard the Sentinel-5 Precursor for the monitoring of terrestrial chlorophyll fluorescence. Atmospheric Measurement Techniques, 8, 1337-1352.

Hao, D., Zeng, Y., Qiu, H., Biriukova, K., Celesti, M., Migliavacca, M., Rossini, M., Asrar, G., Chen, M. 2021a. Practical approaches for normalizing directional solar-induced fluorescence to a standard viewing geometry. Remote Sensing of Environment, 255, 112171.

Hao, D., Asrar, G.R., Zeng, Y., Yang, X., Li, X., Xiao, J., Guan, K., Wen, J., Xiao, Q., Berry, J.A., & Chen, M. 2021b. Potential of hotspot solar-induced chlorophyll fluorescence for better tracking terrestrial photosynthesis. Global Change Biology, 27, 2144-2158

Huete, A. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment. Remote Sensing of Environment, 25, 295-309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment, 83, 195-213.

Jin, H., Eklundh, L. 2014. A physically based vegetation index for improved monitoring of plant phenology. Remote Sensing of Environment, 152, 512-525.

Joiner, J., Yoshida, Y., Köehler, P., Campbell, P., Frankenberg, C., van der Tol, C., ... Sun, Y. 2020. Systematic Orbital Geometry-Dependent Variations in Satellite Solar-Induced Fluorescence (SIF) Retrievals. Remote Sensing, 12(15), 2346.

753 Köhler, P., Guanter, L., Joiner, J. 2015. A linear method for the retrieval of sun-induced
754 chlorophyll fluorescence from GOME-2 and SCIAMACHY data. Atmospheric
755 Measurement Techniques, 8(6), 2589-2608.
756 Knyazikhin, Y., Schull, M. A., Stenberg, P., Mõttus, M., Rautiainen, M., Yang, Y., ... Disney,
757 M. I. 2013. Hyperspectral remote sensing of foliar nitrogen content. Proceedings of the
758 National Academy of Sciences, 110(3), E185-E192.
759 Krutz, D., Müller, R., Knodt, U., Günther, B., Walter, I., Sebastian, I., ... Venus, H. 2019. The
760 instrument design of the DLR earth sensing imaging spectrometer (DESIS). Sensors,
761 19(7), 1622.
762 Liu, X., Guanter, L., Liu, L., Damm, A., Malenovský, Z., Rascher, U., ... Gastellu-Etchegorry,
763 J. P. 2019. Downscaling of solar-induced chlorophyll fluorescence from canopy level to
764 photosystem level using a random forest model. Remote sensing of environment, 231,
765 110772.
766 Ollinger, S. V. 2011. Sources of variability in canopy reflectance and the convergent
767 properties of plants. New Phytologist, 189(2), 375-394.
768 Qi, J., Chehbouni, A., Huete, A.R., Kerr, Y.H., Sorooshian, S., 1994. A modified soil
769 adjusted vegetation index. Remote Sens. Environ. 480 (2), 119–126.
770 Rascher, U., Alonso, L., Burkart, A., Cilia, C., Cogliati, S., Colombo, R., Damm, A., Drusch,
771 M., Guanter, L., Hanus, J. 2015. Sun-induced fluorescence–a new probe of
772 photosynthesis: First maps from the imaging spectrometer HyPlant. Global Change
773 Biology, 21, 4673-4684.
774 Peng, B., Guan, K., Zhou, W., Jiang, C., Frankenberg, C., Sun, Y., ... Köhler, P. 2020.
775 Assessing the benefit of satellite-based Solar-Induced Chlorophyll Fluorescence in crop
776 yield prediction. International Journal of Applied Earth Observation and Geoinformation,
777 90, 102126.

778 Richardson, A.J., Wiegand, C. 1977. Distinguishing vegetation from soil background
779 information. Photogrammetric Engineering and Remote Sensing, 43, 1541-1552.
780 Ryu, Y., Berry, J.A., Baldocchi, D.D. 2019. What is global photosynthesis? History,
781 uncertainties and opportunities. Remote Sensing of Environment, 223, 95-114.
782 Sellers, P. J., Berry, J. A., Collatz, G. J., Field, C. B., Hall, F. G. 1992. Canopy reflectance,
783 photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a
784 new canopy integration scheme. Remote sensing of environment, 42(3), 187-216.
785 Siegmann, B., Alonso, L., Celesti, M., Cogliati, S., Colombo, R., Damm, A., ... Kraska, T.
786 2019. The high-performance airborne imaging spectrometer HyPlant—From raw images
787 to top-of-canopy reflectance and fluorescence products: Introduction of an automatized
788 processing chain. Remote Sensing, 11(23), 2760.
789 Turner, A. J., Köhler, P., Magney, T. S., Frankenberg, C., Fung, I., Cohen, R. C. 2020. A
790 double peak in the seasonality of California's photosynthesis as observed from space.
791 Biogeosciences, 17(2), 405-422.
792 van der Tol, C., Berry, J., Campbell, P., Rascher, U. 2014. Models of fluorescence and
793 photosynthesis for interpreting measurements of solar-induced chlorophyll fluorescence.
794 Journal of Geophysical Research: Biogeosciences, 119, 2312-2327.
795 van der Tol, C., Verhoef, W., Timmermans, J., Verhoef, A., Su, Z. 2009. An integrated model
796 of soil-canopy spectral radiances, photosynthesis, fluorescence, temperature and energy
797 balance. Biogeosciences, 6, 3109-3129.
798 Vermote, E. F., El Saleous, N., Justice, C. O., Kaufman, Y. J., Privette, J. L., Remer, L., ...
799 Tanre, D. 1997. Atmospheric correction of visible to middle-infrared EOS-MODIS data
800 over land surfaces: Background, operational algorithm and validation. Journal of
801 Geophysical Research: Atmospheres, 102(D14), 17131-17141.

802 Wang, C., Chen, J., Wu, J., Tang, Y., Shi, P., Black, T.A., Zhu, K. 2017. A snow-free
803 vegetation index for improved monitoring of vegetation spring green-up date in
804 deciduous ecosystems. Remote Sensing of Environment, 196, 1-12.
805 Wang, C., Guan, K., Peng, B., Chen, M., Jiang, C., Zeng, Y., ... Frankenberg, C. 2020.
806 Satellite footprint data from OCO-2 and TROPOMI reveal significant spatio-temporal
807 and inter-vegetation type variabilities of solar-induced fluorescence yield in the US
808 Midwest. Remote Sensing of Environment, 241, 111728.
809 Wang, S., Ju, W., Peñuelas, J., Cescatti, A., Zhou, Y., Fu, Y., ... Zhang, Y. 2019. Urban–
810 rural gradients reveal joint control of elevated CO2 and temperature on extended
811 photosynthetic seasons. Nature Ecology Evolution, 3(7), 1076-1085.
812 Wu, G., Guan, K., Jiang, C., Peng, B., Kimm, H., Chen, M., ... Moore, C. E. 2020. Radiance-
813 based NIRv as a proxy for GPP of corn and soybean. Environmental Research Letters,
814 15(3), 034009.
815 Yang, W., Kobayashi, H., Wang, C., Shen, M., Chen, J., Matsushita, B., Tang, Y., Kim, Y.,
816 Bret-Harte, M.S., Zona, D. 2019. A semi-analytical snow-free vegetation index for
817 improving estimation of plant phenology in tundra and grassland ecosystems. Remote
818 Sensing of Environment, 228, 31-44.
819 Yang, P., van der Tol, C., Campbell, P. K., Middleton, E. M. 2020. Fluorescence Correction
820 Vegetation Index (FCVI): A physically based reflectance index to separate physiological
821 and non-physiological information in far-red sun-induced chlorophyll fluorescence.
822 Remote sensing of environment, 240, 111676.
823 Zeng, Y., Xu, B., Yin, G., Wu, S., Hu, G., Yan, K., ... Li, J. 2018. Spectral invariant provides
824 a practical modeling approach for future biophysical variable estimations. Remote
825 Sensing, 10(10), 1508.

826 Zeng, Y., Badgley, G., Dechant, B., Ryu, Y., Chen, M., Berry, J. A. 2019. A practical
827 approach for estimating the escape ratio of near-infrared solar-induced chlorophyll
828 fluorescence. Remote Sensing of Environment, 232, 111209.
829 Zeng, Y., Badgley, G., Chen, M., Li, J., Anderegg, L. D., Kornfeld, A., ... Berry, J. A. 2020.
830 A radiative transfer model for solar induced fluorescence using spectral invariants
831 theory. Remote Sensing of Environment, 240, 111678.
832 Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C., Gao, F., ... Huete, A.
833 2003. Monitoring vegetation phenology using MODIS. Remote sensing of environment,
834 84(3), 471-475.