ENVIRONMENTAL STUDIES

A unified vegetation index for quantifying the terrestrial biosphere

Gustau Camps-Valls[1]*, Manuel Campos-Taberner[2], Álvaro Moreno-Martínez[1,3], Sophia Walther[4], Grégory Duveiller[5], Alessandro Cescatti[5], Miguel D. Mahecha[6,7,8], Jordi Muñoz-Marí[1], Francisco Javier García-Haro[2], Luis Guanter[9], Martin Jung[4], John A. Gamon[10,11], Markus Reichstein[4], Steven W. Running[3]

Empirical vegetation indices derived from spectral reflectance data are widely used in remote sensing of the biosphere, as they represent robust proxies for canopy structure, leaf pigment content, and, subsequently, plant photosynthetic potential. Here, we generalize the broad family of commonly used vegetation indices by exploiting all higher-order relations between the spectral channels involved. This results in a higher sensitivity to vegetation biophysical and physiological parameters. The presented nonlinear generalization of the celebrated normalized difference vegetation index (NDVI) consistently improves accuracy in monitoring key parameters, such as leaf area index, gross primary productivity, and sun-induced chlorophyll fluorescence. Results suggest that the statistical approach maximally exploits the spectral information and addresses long-standing problems in satellite Earth Observation of the terrestrial biosphere. The nonlinear NDVI will allow more accurate measures of terrestrial carbon source/sink dynamics and potentials for stabilizing atmospheric CO_2 and mitigating global climate change.

INTRODUCTION

Quantifying vegetation cover, biochemistry, structure, and functioning from space is key to study and understand global change, biodiversity, and agriculture. In practice, remote sensing has relied vastly on the use (and abuse) of vegetation indices (VIs) derived from spectral reflectance owing to their generally decent performance. VIs are parametric transformations of a few spectral bands designed to maximizing their sensitivity to particular biophysical phenomena (e.g., greenness, water content, or photosynthetic activity) while minimizing their sensitivity to factors such as soil properties, solar illumination, atmospheric conditions, and sensor viewing geometry. A plethora of narrow-band indices has been proposed in the literature (*1*). Indices are designed for specific applications and conditions, and their parameters are fixed empirically.

The most widely used VI in Earth observation is undoubtedly the normalized difference vegetation index (NDVI) (*2*, *3*). This index exploits the fact that green healthy vegetation shows contrasting behavior in how it reflects red and near-infrared (NIR) radiation. The more chlorophyll there is in a canopy, the more visible light (including the red) can potentially be absorbed to drive photosynthesis, and thus the higher the absorbed energy that can potentially be consumed in carbon fixation. On the other hand, as more living plant biomass is present, the vegetation will scatter and reflect more NIR radiation, which is unusable for photosynthesis. By calculating the difference between bands measuring red and NIR reflectances, NDVI accentuates the particular signature of green vegetation while attenuating undesired influences from nonvegetative elements. NDVI, and other similar indices, have proven effective in assessing chlorophyll content (*4*, *5*), being a good proxy of vegetation density parameters, like the leaf area index (LAI) and the fractional vegetation cover (FVC) (*6–8*), as well as the fraction of absorbed photosynthetically active radiation (fAPAR). The success of NDVI relies on its ease of use and its availability over long observational records expanding more than three decades, notably thanks to the Advanced Very High Resolution Radiometer (AVHRR), Landsat optical sensors (Multi Spectral Scanner, Thematic Mapper, Enhanced Thematic Mapper, Operational Land Imager), and the Moderate Resolution Imaging Spectroradiometer (MODIS).

However, NDVI has two major limitations. First, the relationship between NDVI and green biomass is nonlinear and saturates. Some indices such as the enhanced vegetation index (EVI) (*9*) have tried to compensate for this using information from other bands, but the saturation problem remains. Other approaches have tried to improve NDVI heuristically to obtain a good proxy of both fAPAR and light-use efficiency, and hence suggested it for gross primary productivity (GPP) estimation (*10*). Actually, some authors have proposed $NDVI^2$ (*11*) and other arbitrary exponentiations (*12*) to cope with the nonlinear issue. The second issue is that VIs, by construction, react to the presence of green leaves, but not to photosynthesis per se. GPP can thus decline without any leaf abscission (i.e., a reduction of LAI) or reduction in chlorophyll. A relatively new way to estimate GPP variability from satellite measurements to retrieve sun-induced chlorophyll fluorescence (SIF) (*13*). However, the relationship between canopy GPP and SIF retrieved from space is still not fully understood (*14*), and more importantly, this technique is still only available with an overly coarse spatial resolution and a very shallow temporal archive (*15*, *16*).

Using radiative transfer models, Sellers *et al.* (*17–19*) noted early on that NIR reflectance is a better proxy for fAPAR than NDVI. The problem is then to disentangle the fraction of the NIR that is reflected from the vegetation from the remaining fraction of NIR reflected

[1]Image Processing Laboratory, Universitat de València, 46980, Paterna, Spain. [2]Environmental Remote Sensing group (UV-ERS), Universitat de València, 46100, Burjassot, Spain. [3]Numerical Terradynamic Simulation Group (NTSG), University of Montana, Missoula, MT, USA. [4]Max Planck Institute for Biogeochemistry, 07745 Jena, Germany. [5]European Commission Joint Research Centre, Ispra, Italy. [6]Remote Sensing Centre for Earth System Research, Leipzig University, Talstr. 35, 04103 Leipzig, Germany. [7]Helmholtz Centre for Environmental Research - UFZ, Permoserstraße 15, 04318 Leipzig, Germany. [8]German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Puschstr. 4, 04103 Leipzig, Germany. [9]Universitat Politècnica de València, 46022 València, Spain. [10]University of Alberta, Edmonton, Alberta, Canada. [11]University of Nebraska–Lincoln, Lincoln NE, USA.
*Corresponding author. Email: gustau.camps@uv.es

from nonvegetated elements within a mixed pixel. To address this issue, Badgley *et al.* (*20*) proposed considering NDVI as a proxy for vegetation coverage instead of a proxy for fAPAR, and thus multiply NDVI times NIR to calculate a new index, NIRv, which shows high correlations with SIF and GPP at specific temporal scales. Despite its wide reception in the community, NIRv also raises some intriguing questions. For example, given that fAPAR is estimated by both of its components (NIR and NDVI), how does this affect the interpretation of the index? Also, as NIRv linearly scales with the NIR reflectance, how does it deal with saturation? Last, NIRv still uses the same bands as NDVI, but it is not clear how the adopted approximations and assumptions affect NIRv or whether it exploits all available information in these spectral bands.

This paper introduces a methodology to generalize the broad family of VIs based on differences and ratios of spectral bands. Unlike previous approaches to improve indices based on principled (*10*, *20*, *21*) or heuristic parametric transformations (*11*, *12*, *22*, *23*), here we adopt a machine learning standpoint using the theory of kernel methods, which has been widely used to derive nonlinear algorithms from linear ones, while still resorting to linear algebra operations (*24*, *25*). Kernel methods map the involved spectral bands using a nonlinear feature map to a higher dimensional space where the index is defined. The calculation can be expressed in terms of the spectral channels by the definition of a kernel (similarity) function, so one does not need to define the feature map explicitly. The main property of kernel methods is that of linearizing the problem, which is what most of the indices seek either heuristically or based on first principles. Also, by using a particular kernel function, we have guarantees that all higher-order relations between the spectral channels are accounted for, not just the first-order ones. For example, when using differences between NIR and the red bands, the kernel function summarizes all monomials of the differences too, i.e., {NIR-red, (NIR-red)^2, (NIR-red)^3, …} in a single scalar. Although kernel methods can, in principle, be applied to any VI (see section S1.5 and table S1), the framework is illustrated here to generalize NDVI, largely because of the long history and wide utility of this index, most notably to perform global and long-term studies. We specifically define the NDVI in Hilbert spaces and adopt the radial basis function (RBF) reproducing kernel, $k(\text{NIR}, \text{red}) = \exp\left(-(\text{NIR} - \text{red})^2/(2\sigma^2)\right)$, where the σ parameter controls the notion of distance between the NIR and red bands. The presented kernel NDVI (kNDVI) reduces to compute

$$\text{kNDVI} = \tanh\left(\left(\frac{\text{NIR} - \text{red}}{2\sigma}\right)^2\right)$$

where σ is a length-scale parameter to be specified in each particular application and represents the sensitivity of the index to sparsely/densely vegetated regions. A reasonable choice is taking the average value σ = 0.5(NIR + red) (see sections S1 and S2 for mathematical and ecophysiological justifications), which leads to a simplified operational index version expressed as kNDVI = tanh (NDVI^2). The selection of the kernel function and prescription of its parameter allows the kNDVI to perform an automatic and pixel-wise adaptive stretching and guarantees that all moments of the relations between the NIR and red channels are taken into account. This also allows kNDVI to cope with saturation effects, complex phenological cycles, and seasonal variations, to deal with the mixed-pixel problem (*20*), and to propagate lower uncertainty than other indices (section S2.5). It can be shown that kNDVI actually generalizes NDVI and NIRv

theoretically (see sections S1 and S2 and Properties S2.1 and S2.2), which ensures an improved performance. Last, the presented methodology, and the kNDVI in particular, are easy to implement and use in practice (section S10), which is of paramount relevance in operational studies.

RESULTS AND DISCUSSION

We show that kNDVI exhibits consistently stronger correlations than NDVI and NIRv in key independent products [GPP at flux tower estimates and SIF from Global Ozone Monitoring Experiment–2 (GOME-2)]. In general, the proposed index performs better than NDVI and NIRv in all applications, biomes, and climatic zones. The kNDVI is more resistant to saturation, bias, and complex phenological cycles and shows enhanced robustness to noise and stability across spatial and temporal scales (sections S6.2 and S6.3). Additional results for approximating MODIS LAI (section S4), correlation with other related parameters (like fAPAR and FVC) acquired in situ (section S7), crop yield estimation (section S8), and kNDVI's use for image change detection (section S9) further confirm the validity of the approach. All these properties and performance are achieved without adopting any specific assumption, just exploiting all higher order statistical relations between the involved reflectances.

Accurate proxy to GPP

We evaluated and compared the performance of kNDVI with NDVI and NIRv as a GPP proxy using flux tower GPP estimates from the FLUXNET database (section S5). The proposed kNDVI provides correlations with GPP similar to or better than the other indices over all considered biomes and across all the 169 flux tower sites (Table 1). The weakest relationships are observed for evergreen broad-leaved forests, which can be expected because of the stronger saturation effect in such ecosystem (similarly clear when using the index for LAI estimation, see section S4). The kNDVI excels in each biome individually, confirming its adaptive nature, and globally

Table 1. Temporal correlation coefficient between the VIs and the parameters GPP and SIF per biome. Only vegetation biomes are considered and classes in IGBP were grouped as indicated in parentheses: C1 = NF=Needle-leaf Forest (1 + 3), C2 = EBF = Evergreen Broadleaf Forest (2), C3 = DBF=Decidious Broadleaf Forest (4), C4 = MF = Mixed forest (5), C5 = SH=Shrublands (6 + 7), C6 = SAV=Savannas (8 + 9), C7 = GRA = Herbaceous (10), C8 = CRO=Cultivated (12). Best results per biome indicated in bold and darker green indicates higher correlation.

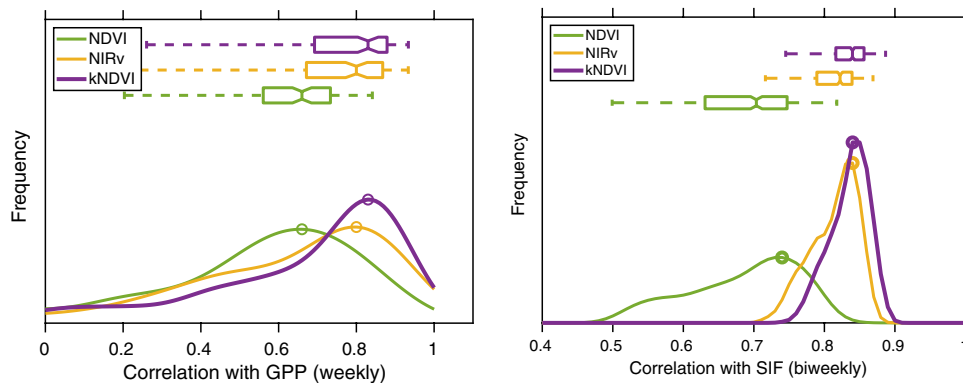|	Biome	GPP			SIF		
		NDVI	NIRv	kNDVI	NDVI	NIRv	kNDVI
C1	NF	0.59	**0.65**	0.62	0.66	0.73	**0.74**
C2	EBF	0.37	**0.45**	**0.45**	0.65	0.72	**0.73**
C3	DBF	0.61	0.79	**0.82**	0.41	0.48	**0.51**
C4	MF	0.69	0.84	**0.86**	0.61	0.68	**0.70**
C5	SH	0.57	0.68	**0.72**	0.63	0.69	**0.71**
C6	SAV	0.63	**0.74**	**0.74**	0.56	0.62	**0.63**
C7	GRA	0.61	0.71	**0.72**	0.48	0.59	**0.66**
C8	CRO	0.51	**0.58**	**0.58**	0.53	0.62	**0.67**
	ALL	0.59	**0.68**	**0.68**	0.54	0.62	**0.65**

Fig. 1. Correlation between the indices and parameters. Histogram of the correlation coefficient between the **VIs** and the parameters: for GPP (left) correlation computed over 169 FLUXNET sites, and for SIF (right) averaged over all 506 global images.
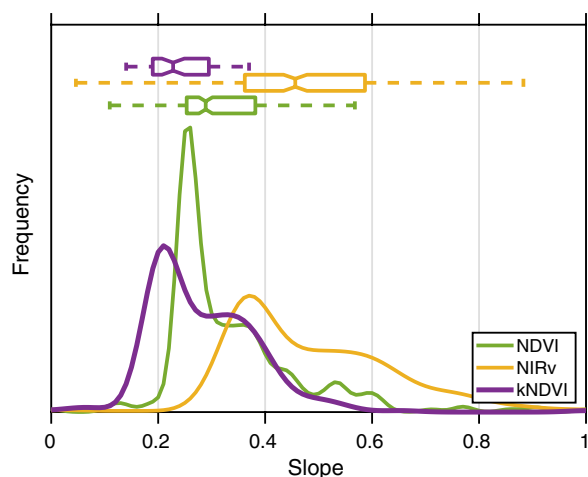

Fig. 2. Goodness of fit between the indices and GPP. Distribution of slopes of site-level linear regressions (normalized between 0 and 1) between the indices and biweekly GPP from 169 FLUXNET sites.

shows a clear gain (Fig. 1). Although photosynthesis is driven by the amount of vegetation photosynthetic mass within a pixel, solar irradiation and environmental constraints also play a critical role. The latter is not accounted for by the spectral information provided by NIR and red bands. This explains why all indices present lower correlation with GPP and SIF than with LAI for all biomes (Table 1, cf. section S4). The correlation is, however, higher for the kNDVI in almost all cases. Alternative measures of nonlinear association between GPP and the indices, such as Spearman's correlation (*26*), mutual information (*27*), and distance correlation (*28*), yielded similar results and conclusions (see sections S5 and S6), thus confirming the good capabilities of kNDVI to implicitly linearize the problem.

We studied the robustness of the indices across sites. Figure 2 shows the density and boxplots of the slopes (scaled between 0 and 1) for all 169 flux tower sites. The NIRv index shows a mean closer to 0.5, but the spread is higher than for the kNDVI. Both NDVI and NIRv show very wide whiskers (and hence pathological behaviors and high sensitivity to outliers), while kNDVI shows higher robustness and stability across sites. A simple analysis over all the towers shows that kNDVI outperformed in 84 of the towers (50%), NIRv in 59 (35%), and NDVI in 26 (15%). The kNDVI gains are more

noticeable in deciduous and evergreen forests, which confirms the good adaptation to varying photosynthetic phenology of different biomes, primarily forests. This is confirmed when looking at the seasonal patterns of stand photosynthesis for some illustrative sites in Fig. 3, expressed as monthly GPP. For example, the CA-TP4 (Ontario–Turkey Point 1939 Plantation White Pine site) is a region dominated by densely covered woody vegetation and displays green foliage all year round. Unlike NDVI that shows relatively too much and too little sensitivity, respectively, to seasonally changing GPP, the kNDVI follows much better the temporal shape and captures the higher and lower GPP values too. This might be due to the subtle pigment shifts that are largely invisible to NDVI, but may be more detectable by kNDVI, as it was recently shown with NIRv (*29*). For grasslands, like the CH-Oe1 (Oensingen, Switzerland), neither NDVI nor NIRv can disentangle the phenological cycle of the vegetation from the background noise, while the kNDVI returns acceptable results with larger dynamic range. Here, the tree and shrub cover is less than 10% and a permanent mixture of water and herbaceous or woody vegetation is observed, inducing a strong mixed-pixel problem aggravated by complex topography. The IT-Ro1 (Roccarespampani-1 near Viterbo site) is a deciduous broad-leaved forest consisting of broadleaf tree communities with a clear annual cycle of long leaf-on and leaf-off periods, which are followed faithfully by the kNDVI index. NIRv and kNDVI reveal very similar characteristics. An interesting case is that of closed shrublands. The mixed shrub foliage in the Kennedy Space Center site CSH US-KS2, which can be either evergreen or deciduous, is efficiently handled by kNDVI ($R = 0.72$) over NIRv ($R = 0.68$) and NDVI ($R = 0.57$). Here, unlike NIRv, the proposed kNDVI does not over- and underestimate GPP. Overall, we observed that the kNDVI closely tracked the seasonal dynamics of photosynthesis, presenting a better agreement with GPP. This is achieved by adaptively stretching the dynamic range to better capture time-series extremes (e.g., sparsely and densely vegetated, as well as cold and dry regions). The proposed kNDVI seems to largely correct for "background effects" (important in sparse vegetation or snow) and saturation and may be more sensitive to subtle greenness shifts (e.g., evergreens) based on pigments rather than structure per se.

Closer monitoring of photosynthetic activity of ecosystems

Recent studies have linked SIF and VIs, such as NDVI and NIRv (*20*), as a pragmatic alternative to more sophisticated machine learning
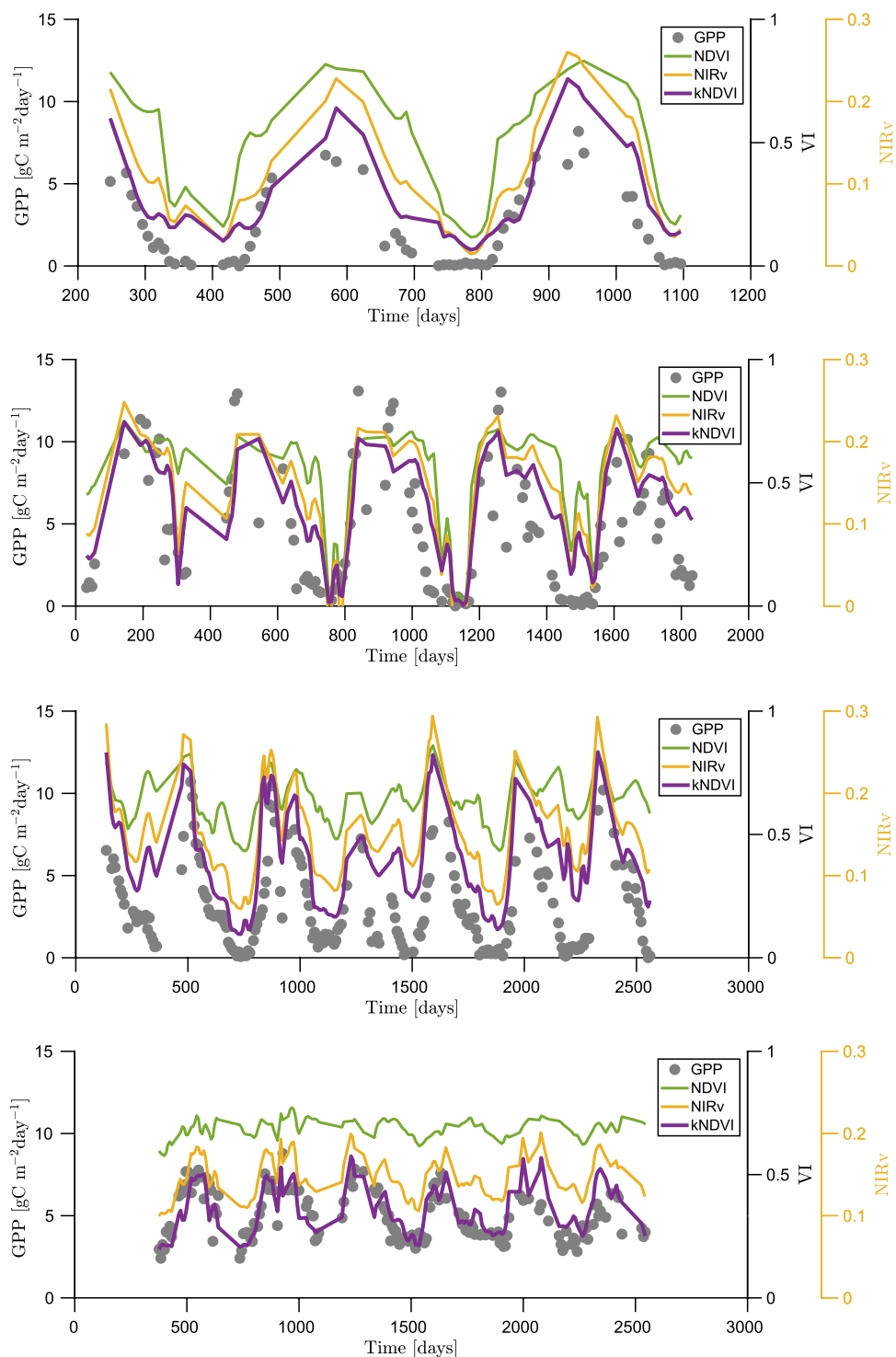
Fig. 3. Monitoring GPP at tower level. Illustrative results over four flux towers covering evergreen needle-leaved forests (CA-TP4), grasslands (CH-Oe1), deciduous broadleaf forest (IT-Ro1), and closed shrublands (US-KS2).

approaches (*30*). We here evaluate the kNDVI computed from the MODIS reflectance bands to approximate globally gridded GOME-2 SIF at 16-day temporal resolution. Despite the fact that GOME-2 can measure both SIF and the NIR and red bands simultaneously, we intentionally estimated all indices independently from coincident

MODIS data (see processing details in section S6). We computed the correlation between time series. The kNDVI outperforms the other indices in general (Fig. 1) and in all biomes individually (Table 1), especially in DBF, GRA, and CRO: 5 to 11% gain in correlation over NIRv and 20 to 35% over NDVI.
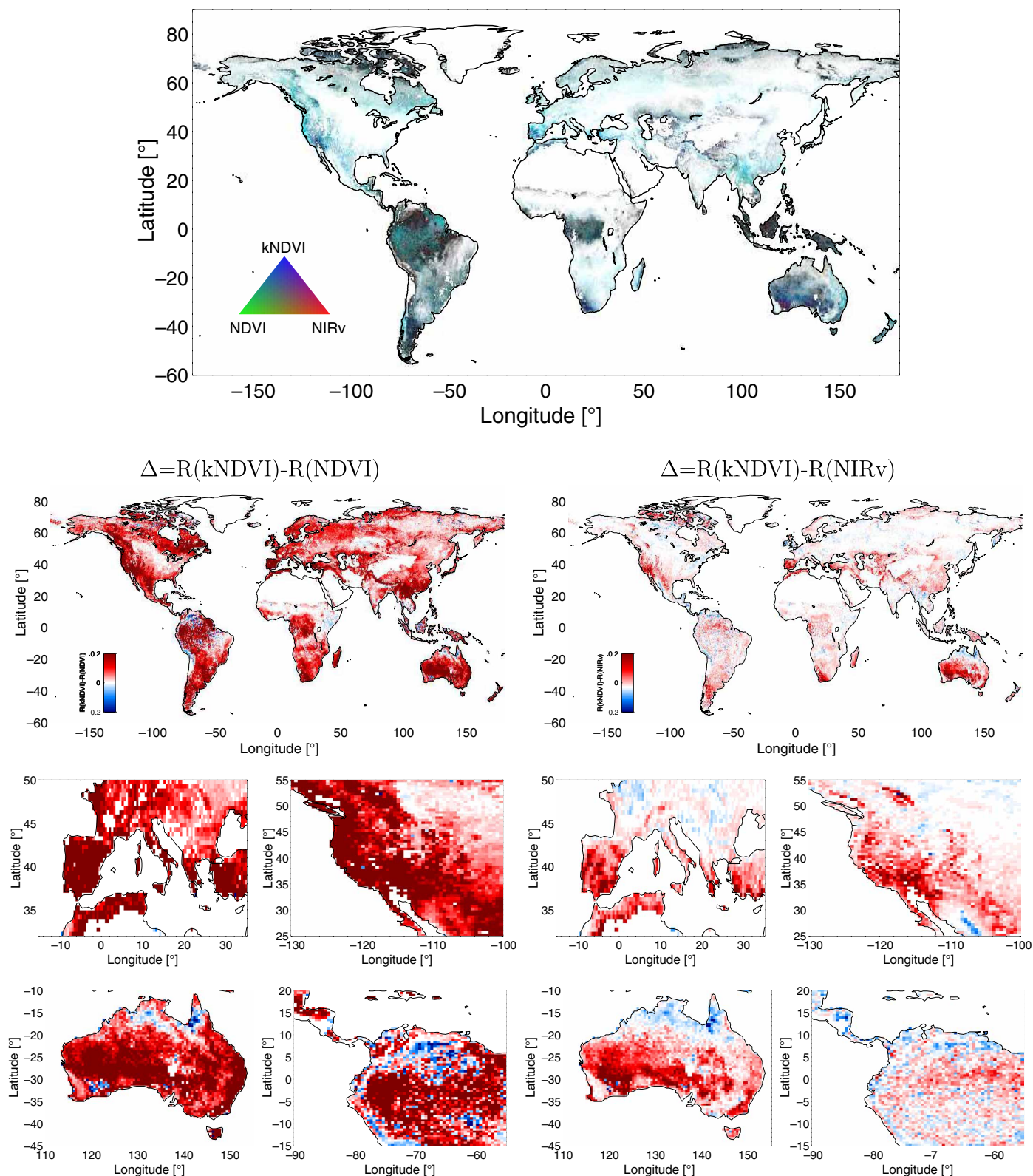
Fig. 4. Temporal correlation between indices and SIF globally. Top: Color composite of indices-to-SIF correlation, (R, G, B) = (NIRv, NDVI, kNDVI). Bluish means kNDVI outperforms the rest, which generally happens [in 91.32% of the pixels over NDVI (left) and 69.69% of the cases over NIRv (right)] and particularly in the extreme (low and high) vegetation covers or in cold and dry regions. Bottom: Differences of correlation-with-SIF between the proposed index kNDVI and NDVI (left) and NIRv (right), both globally and for extreme regions. Red colors indicate a higher correlation for kNDVI, and blue indicates a lower correlation for kNDVI (relative to the other indices).
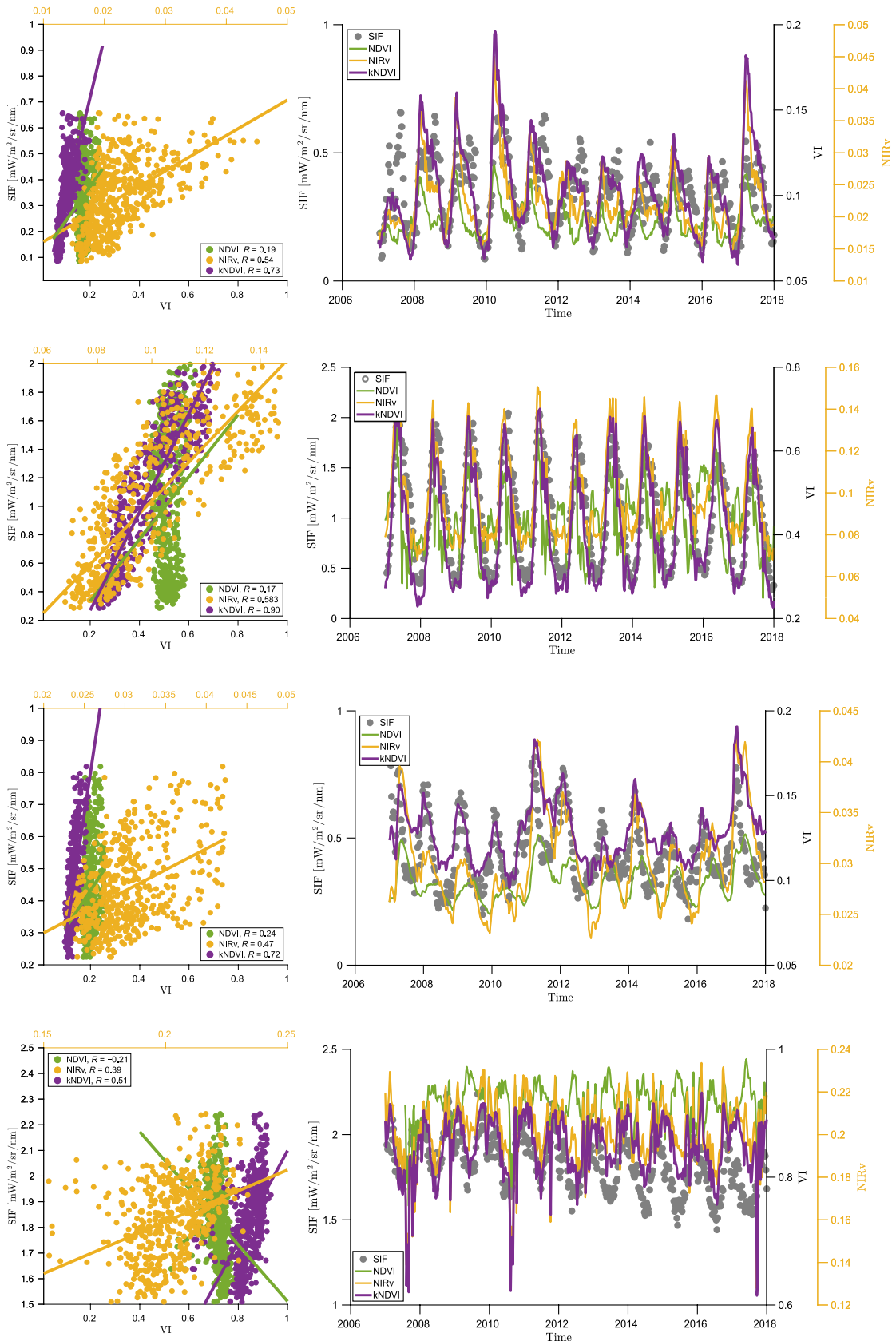
Fig. 5. Temporal analysis over selected study areas. Scatterplots of the different indices versus SIF (left), and the average time series over the study areas (right). Axes limits were optimized to improve visualization of all indices.
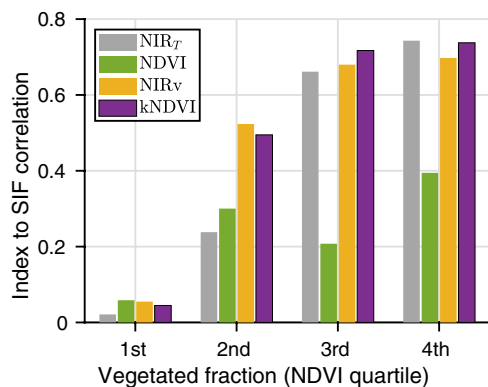
Fig. 6. Correlation with vegetated fraction. Correlation coefficient between the indices and SIF increases with vegetated fraction (computed from NDVI percentiles). We include the total NIR, NIR_T, as a reference. The lower bounds of the NDVI quartiles are as follows: 0, 0.25, 0.50, and 0.75.

To confirm the robustness to capture extreme SIF values, we studied the spatial maps of temporal correlation coefficients. kNDVI dominates in all regions (Fig. 4, top) and correlates better with SIF in 69.69% of the pixels compared to NIRv and in 91.32% of cases compared to NDVI (see Fig. 4, bottom). Results suggest that the kNDVI clearly outperforms the other indices in densely vegetated tropical (e.g., Amazonia and Indonesia) and arid regions (e.g., Australia and Mediterranean). As for the case of GPP, other measures of correlation yielded identical conclusions (table S7). Further analysis confirmed the dominant performance of kNDVI in all latitudes, especially in higher and lower ones (table S10), as well as in all climatic zones, especially in the arid and cold regions (table S9).

The study areas in Fig. 4 showed the biggest differences between the kNDVI and NDVI and NIRv, and are further scrutinized in Fig. 5. The kNDVI provides improved fit scores in all cases, larger excursions in general, and more resistance to noise and saturation. The higher accuracy by kNDVI (e.g., in California, +19% in R over NIRv) comes mainly from the better behavior in the presence of sharp phenological cycles. In the Iberian peninsula, kNDVI and NIRv perform similarly in quantitative terms, but the proposed kNDVI appears less affected by high-frequency components and covers the whole dynamic range nicely. In Australia, the favorable numerical gain in R (+25%) and the much lower scatter highlight that kNDVI better approximates SIF and closely follows the cycles (especially in March-April-May periods). Despite the big challenges in the Amazon for SIF estimation with GOME-2, the kNDVI can be a more convenient choice compared to other indices, as it deals better with noise and background effects (e.g., soil, standing water, or snow). All in all, the proposed kNDVI seems better qualified to cope with noise, saturation, and complex phenologies.

Similar conclusions were obtained when we studied spatial correlations through time: The proposed index achieves noticeable improvements over NDVI and NIRv, especially between August and November, thus improving autumn phenology owing to its adaptive stretching (see fig. S10 in section S6). The kNDVI is more competitive at finer temporal resolutions (native biweekly) with a noticeable advantage over NDVI (+15%) and NIRv (+4%), but the gain over NIRv disappears at bimonthly scales, since the temporal aggregation induces a "more linear" problem. Likewise, a broader spatial aggregation (from 0.5 up to 2) yielded improved results of all indices, but kNDVI still outperformed the others independently of the spatial scale (section S6 and fig. S11).

We lastly studied the capabilities of kNDVI to deal with the mixed-pixel problem (Fig. 6). Both kNDVI and NIRv scale with the total NIR, NIR_T, unlike NDVI that clearly saturates. The kNDVI strongly correlates with SIF over highly vegetated pixels, but the correlation decreases with lower vegetated fractions (Fig. 6). The difference between kNDVI and NDVI stands out, and kNDVI is slightly higher correlated with SIF than NIRv, thus suggesting that the index can reliably isolate the proportion of reflectance attributable to vegetation as well. These properties emerge directly from the NIR-red relations since no assumption is made in designing the index. Accounting for all NIR-red relations allows us to optimally disentangle the mixed-pixel problem efficiently, especially in the densely vegetated areas (e.g., LAI and GPP phenology of crops in section S4 and Fig. 3).

The study of natural and agricultural systems should greatly benefit from the kNDVI proposed here because of its solid theoretical foundation combined with its ease of calculation and application. The high correlation with GPP and SIF across all biomes, especially in grasslands, croplands, and mixed forests as well as in arid regions, suggests that the index can efficiently cope with both the saturation and the mixed-pixel problems encountered with traditional indices. The proposed kNDVI explains a large fraction of the variance of GPP at flux tower level, showed good robustness capabilities to noise and saturation, and enhanced stability across space. The kNDVI also highly correlates with SIF derived from an independent sensor, paving the way toward improving our quantification and understanding of photosynthesis at the global scale. Its application and usefulness goes beyond vegetation monitoring and embraces change and extreme detection, phenological and greening studies, upscaling parameters, and all applications where VIs in general and NDVI in particular have previously demonstrated their utility. Our results demonstrate that an agnostic statistical approach is sufficient to explain most of the observed signal.

The kernel methods framework allowed us to generalize all VIs, but we focused on the NDVI case only. Kernel methods, in general, and the kNDVI, in particular, implement the original operation (e.g., NDVI) in a high-dimensional feature space where spectral bands are mapped to. The solution of kNDVI is thus a nonlinear version of NDVI. The framework allows us to accomplish the ever-sought linearization operation implicitly. This means that no ad hoc parametric transformations are needed, just the kernel operation. This also implies that virtually no gain should be obtained over other indices when the relation between the bands and the parameter of interest is linear, such as for instance when an appropriate PAR normalization is applied (see sections S5.2 and S6.4) or whenever one averages over larger spatial or temporal scales (see section S6.3). Our results, however, suggested that the kNDVI instantiation improved results in all problems, even when the domain was previously linearized. This makes the index a very powerful and practical default choice. We anticipate a wide use and development of the proposed index in particular, and of the family of nonlinear VIs in general, to derive informative indicators for operational Earth monitoring and the quantification of the terrestrial biosphere vital signs.

MATERIALS AND METHODS
Datasets and processing
GPP and FLUXNET data
The GPP data were obtained from FLUXNET, which is a collection of sites from multiple regional networks (*31*). This network provides

a compilation of in situ observations to measure the exchanges of carbon dioxide, water vapor, and energy between the biosphere and atmosphere (*32*). To calculate the GPP, the carbon dioxide flux, i.e., net ecosystem exchange, is measured by means of the eddy covariance method. This flux is further partitioned into ecosystem respiration and GPP [gC m^{-2} day^{-1}] using the daytime (*33*) or nighttime (*34*) partitioning methods. For our analyses, we used GPP estimates from the freely available Tier 1 dataset that were obtained with the daytime partitioning method. Of all available sites (212), we selected a subset of 169 sites corresponding to natural vegetation having less than 50% of missing remotely sensed data due to cloud contamination. In addition, we only considered sites where we had more than 4 months of available flux data.

SIF data from GOME-2

We generated GOME-2 0.5 fluorescence at 740 nm and reflectance at 670 and 780 nm from level 2 data obtained from measurements of the GOME-2 sensor flying onboard MetOp-A. The retrieval algorithm of SIF [mW/m^2/sr/nm] proposed in (*35*) uses the filling-in of Fraunhofer lines caused by the plants' chlorophyll fluorescence. Data were gridded to 16 day and 0.5° resolutions from the individual soundings and cover 11 years (2007–2017). No spatial smoothing or temporal averaging was performed before computing or averaging results. High sun zenith angle (SZA) observations (SZA > 70°) were removed from the analysis as well as cloudy scenes with a cloud fraction over 50% and observations taken between 2 p.m. and 8 a.m. local time. The illumination corrected SIF/cos(SZA) was considered, cf. section S6.

MODIS BRDF-corrected reflectances

MODIS reflectance data were derived from the MCD43A4.006 bidirectional reflectance distribution function (BRDF)-Adjusted Reflectance 16-Day L3 Global 500m product (*36*). They are disseminated from the Land Processes Distributed Active Archive Center (LP DAAC) also available at Google Earth Engine (GEE). The MCD43A2 MODIS product, which contains ancillary quality information for the corresponding MCD43A4 product, was also used for avoiding low-quality BRDF estimates. We computed the indices and conducted the analysis at 16-day temporal and 500-m spatial scales over the 11 years of SIF data.

Analysis

General rationale

In all our experiments, we used reflectance values from MODIS, yet radiances or digital counts could also be used. The flux tower GPP estimates in our experiments come from the site-level data in (*31*). The SIF product comes from GOME-2, so the product is fully independent of MODIS reflectances. GPP and SIF correlations are computed in the time domain, while for SIF, we additionally compute correlations in space and then average results over time (results shown in section S6).

In all cases, we compute correlations between indices (NDVI, NIRv, and kNDVI) and the considered product only in meaningful vegetation classes: Needleleaf Forest, Evergreen Broadleaf Forest, Decidious Broadleaf Forest, Mixed forest, Shrublands, Savannas, Herbaceous, and Cultivated. These resulted from a meaningful grouping of International Geosphere-Biosphere Programme (IGBP) classes (see section S3). Analysis of the SIF results also considered aggregated climatic zones (Tropical, Arid, Temperate, Cold, and Polar), monthly means, and latitude averages (see section S6).

kNDVI calculation

The kNDVI index is defined as

$$\text{kNDVI} = \frac{k(n,n) - k(n,r)}{k(n,n) + k(n,r)} \tag{1}$$

where $n, r \in \mathbb{R}$ refer to the reflectances in the NIR and red channels, respectively, and the kernel function k measures the similarity between these two bands. We used in all cases the RBF kernel, $k(a, b) = \exp\left(-(a-b)^2/(2\sigma^2)\right)$, where the σ parameter controls the notion of distance between the NIR and red bands. This kernel function induces an important simplification

$$\text{kNDVI} := \frac{1 - k(n,r)}{1 + k(n,r)} = \tanh\left(\left(\frac{n-r}{2\sigma}\right)^2\right) \tag{2}$$

Other kernel functions are possible, but the RBF kernel is the most widely used one because of its theoretical and practical advantages (see sections S1 and S2) (*24, 25*). We calculated the kNDVI fixing the length-scale parameter σ equal to the mean distance between the NIR and red bands, $\sigma = 0.5(n + r)$, which is a standard heuristic in the kernel methods literature, makes the index adaptive to each pixel, and worked very well in practice. Note that this simplification further reduces the index to

$$\text{kNDVI} = \tanh(\text{NDVI}^2) \tag{3}$$

Further optimization of σ per biome was done, but results did not improve substantially (results not shown).

Reproducibility: Open-source software and data

All calculations, visualization, and analyses were performed using the MATLAB programming language. We stored and processed netCDF files and tabular data. The kNDVI can be easily coded and applied. We give implementations in five standard programming languages (MATLAB, R, Python, Julia, and IDL) and in the GEE in section S10.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at http://advances.sciencemag.org/cgi/content/full/7/9/eabc7447/DC1

REFERENCES AND NOTES

1. J. Xue, B. Su, Significant remote sensing vegetation indices: A review of developments and applications. *J. Sens.* **2017**, 1–17 (2017).
2. J. Rouse Jr, R. Haas, J. Schell, D. Deering, Monitoring vegetation systems in the great plains with ERTS (NASA Special Publication, 1974).
3. C. Tucker, Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **8**, 127–150 (1979).
4. R. B. Myneni, F. G. Hall, P. J. Sellers, A. L. Marshak, The interpretation of spectral vegetation indexes. *IEEE Trans. Geosc. Rem. Sens.* **33**, 481–486 (1995).
5. D. Haboudane, J. R. Miller, E. Pattey, P. J. Zarco-Tejada, I. B. Strachan, Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture. *Remote Sens. Environ.* **90**, 337–352 (2004).
6. G. Le Maire, C. Francois, K. Soudani, D. Berveiller, J. Y. Pontailler, N. Breda, H. Genet, H. Davi, E. Dufrane, Calibration and validation of hyperspectral indices for the estimation of broadleaved forest leaf chlorophyll content, leaf mass per area, leaf area index and leaf canopy biomass. *Remote Sens. Environ.* **112**, 3846–3864 (2008).
7. D. Haboudane, N. Tremblay, J. Miller, P. Vigneault, Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **46**, 423–437 (2008).
8. J. Berni, P. Zarco-Tejada, L. Suárez, E. Fereres, Thermal and narrowband multispectral remote sensing for vegetation monitoring from an unmanned aerial vehicle. *IEEE Trans. Geosci. Remote Sens.* **47**, 722–738 (2009).

9. A. Huete, K. Didan, T. Miura, E. Rodriguez, X. Gao, L. Ferreira, Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **83**, 195–213 (2002).

10. J. Joiner, Y. Yoshida, Y. Zhang, G. Duveiller, M. Jung, A. Lyapustin, Y. Wang, C. J. Tucker, Estimation of terrestrial global gross primary production (GPP) with satellite data-driven models and eddy covariance flux data. *Remote Sens.* **10**, 1346 (2018).

11. S. Wang, L. Zhang, C. Huang, N. Qiao, An NDVI-based vegetation phenology is improved to be more consistent with photosynthesis dynamics through applying a light use efficiency model over boreal high-latitude forests. *Remote Sens.* **9**, 695 (2017).

12. W. Wu, The generalized difference vegetation index (GDVI) for dryland characterization. *Remote Sens.* **6**, 1211–1233 (2014).

13. Y. Ryu, J. A. Berry, D. D. Baldocchi, What is global photosynthesis? History, uncertainties and opportunities. *Remote Sens. Environ.* **223**, 95–114 (2019).

14. A. Porcar-Castell, E. Tyystjärvi, J. Atherton, C. Van der Tol, J. Flexas, E. E. Pfündel, J. Moreno, C. Frankenberg, J. A. Berry, Linking chlorophyll a fluorescence to photosynthesis for remote sensing applications: Mechanisms and challenges. *J. Exp. Bot.* **65**, 4065–4095 (2014).

15. G. Duveiller, F. Filipponi, S. Walther, P. Köhler, C. Frankenberg, L. Guanter, A. Cescatti, A spatially downscaled sun-induced fluorescence global product for enhanced monitoring of vegetation productivity. *Earth Syst. Sci. Data* **12**, 1101–1116 (2020).

16. J. Wen, P. Köhler, G. Duveiller, N. Parazoo, T. Magney, G. Hooker, L. Yu, C. Chang, Y. Sun, A framework for harmonizing multiple satellite instruments to generate a long-term global high spatial-resolution solar-induced chlorophyll fluorescence (SIF). *Remote Sens. Environ.* **239**, 111644 (2020).

17. P. J. Sellers, Canopy reflectance, photosynthesis and transpiration. *Int. J. Remote Sens.* **6**, 1335–1372 (1985).

18. P. Sellers, Canopy reflectance, photosynthesis, and transpiration, II. The role of biophysics in the linearity of their interdependence. *Remote Sens. Environ.* **21**, 143–183 (1987).

19. P. Sellers, J. Berry, G. Collatz, C. Field, F. Hall, Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. *Remote Sens. Environ.* **42**, 187–216 (1992).

20. G. Badgley, C. Field, J. Berry, Canopy near-infrared reflectance and terrestrial photosynthesis. *Sci. Adv.* **3**, e1602244 (2017).

21. J. A. Gamon, K. F. Huemmrich, C. Y. S. Wong, I. Ensminger, S. Garrity, D. Y. Hollinger, A. Noormets, J. Peñuelas, A remotely sensed pigment index reveals photosynthetic phenology in evergreen conifers. *Proc. Natl. Acad. Sci. U.S.A* **113**, 13087–13092 (2016).

22. J. Clevers, The derivation of a simplified reflectance model for the estimation of leaf area index. *Remote Sens. Environ.* **25**, 53–69 (1988).

23. P. Gong, R. Pu, G. S. Biging, M. R. Larrieu, Estimation of forest leaf area index using vegetation indices derived from hyperion hyperspectral data. *IEEE Trans. Geosci. Remote Sens.* **41**, 1355–1362 (2003).

24. G. Camps-Valls, L. Bruzzone, *Kernel Methods for Remote Sensing Data Analysis* (Wiley & Sons, UK, 2009).

25. J. Rojo-Álvarez, M. Martínez-Ramón, J. Muñoz-Marí, G. Camps-Valls, *Digital Signal Processing with Kernel Methods* (Wiley & Sons, UK, 2018).

26. M. Hollander, D. A. Wolfe, E. Chicken, *Nonparametric Statistical Methods* (John Wiley & Sons, 2013), vol. 751.

27. T. M. Cover, J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)* (Wiley-Interscience, USA, 2006).

28. G. J. Székely, M. L. Rizzo, N. K. Bakirov, Measuring and testing dependence by correlation of distances. *Annals Stat.* **35**, 2769–2794 (2007).

29. C. Y. Wong, P. D'Odorico, M. A. Arain, I. Ensminger, Tracking the phenology of photosynthesis using carotenoid-sensitive and near-infrared reflectance vegetation indices in a temperate evergreen and mixed deciduous forest. *New Phytol.* **226**, 1682–1695 (2020).

30. P. Gentine, S. Alemohammad, Reconstructed solar-induced fluorescence: A machine learning vegetation product based on MODIS surface reflectance to reproduce GOME-2 solar-induced fluorescence. *Geophys. Res. Lett.* **45**, 3136–3146 (2018).

31. G. Tramontana, M. Jung, G. Camps-Valls, K. Ichii, B. Raduly, M. Reichstein, C. R. Schwalm, M. A. Arain, A. Cescatti, G. Kiely, L. Merbold, P. Serrano-Ortiz, S. Sickert, S. Wolf, D. Papale, Predicting carbon dioxide and energy fluxes across global fluxnet sites with regression algorithms. *Biogeosciences* **13**, 4291–4313 (2016).

32. D. D. Baldocchi, Assessing the eddy covariance technique for evaluating carbon dioxide exchange rates of ecosystems: Past, present and future. *Glob. Chang. Biol.* **9**, 479–492 (2003).

33. G. Lasslop, M. Reichstein, D. Papale, A. Richardson, A. Arneth, A. Barr, P. Stoy, G. Wohlfahrt, Separation of net ecosystem exchange into assimilation and respiration using a light response curve approach: Critical issues and global evaluation. *Glob. Chang. Biol.* **16**, 187–208 (2010).

34. M. Reichstein, E. Falge, D. Baldocchi, D. Papale, M. Aubinet, P. Berbigier, C. Bernhofer, N. Buchmann, T. Gilmanov, A. Granier, T. Grünwald, K. Havránková, H. Ilvesniemi,

D. Janous, A. Knohl, T. Laurila, A. Lohila, D. Loustau, G. Matteucci, T. Meyers, F. Miglietta, J.-M. Ourcival, J. Pumpanen, S. Rambal, E. Rotenberg, M. Sanz, J. Tenhunen, G. Seufert, F. Vaccari, T. Vesala, D. Yakir, R. Valentini, On the separation of net ecosystem exchange into assimilation and ecosystem respiration: Review and improved algorithm. *Glob. Chang. Biol.* **11**, 1424–1439 (2005).

35. P. Köhler, L. Guanter, J. Joiner, A linear method for the retrieval of sun-induced chlorophyll fluorescence from GOME-2 and SCIAMACHY data. *Atmos. Meas. Tech.* **8**, 2589–2608 (2015).

36. C. B. Schaaf, F. Gao, A. H. Strahler, W. Lucht, X. Li, T. Tsang, N. C. Strugnell, X. Zhang, Y. Jin, J.-P. Muller, P. Lewis, M. Barnsley, P. Hobson, M. Disney, G. Roberts, M. Dunderdale, C. Doll, R. P. d'Entremont, B. Hu, S. Liang, J. L. Privette, D. Roy, First operational brdf, albedo nadir reflectance products from modis. *Remote Sens. Environ.* **83**, 135–148 (2002).

37. J. Shawe-Taylor, N. Cristianini, *Kernel Methods for Pattern Analysis* (Cambridge Univ. Press, 2004).

38. P. Zarco-Tejada, A. Berjn, R. López-Lozano, J. Miller, P. Martín, V. Cachorro, M. González, A. de Frutos, Assessing vineyard condition with hyperspectral indices: Leaf and canopy reflectance simulation in a row-structured discontinuous canopy. *Remote Sens. Environ.* **99**, 271–287 (2005).

39. R. E. Crippen, Calculating the vegetation index faster. *Remote Sens. Environ.* **34**, 71–73 (1990).

40. A. A. Gitelson, Y. J. Kaufman, R. Stark, D. Rundquist, Novel algorithms for remote estimation of vegetation fraction. *Remote Sens. Environ.* **80**, 76–87 (2002).

41. J. Delegido, L. Alonso, G. González, J. Moreno, Estimating chlorophyll content of crops from hyperspectral data using a normalized area over reflectance curve (NAOC). *Int. J. Appl. Earth Observ. Geoinform.* **12**, 165–174 (2010).

42. Q. Wang, S. Adiku, J. Tenhunen, A. Granier, On the relationship of NDVI with leaf area index in a deciduous forest site. *Remote Sens. Environ.* **94**, 244–255 (2005).

43. F. Baret, J. T. Morissette, R. A. Fernandes, J. L. Champeaux, R. B. Myneni, J. Chen, S. Plummer, M. Weiss, C. Bacour, S. Garrigues, J. E. Nickeso, Evaluation of the representativeness of networks of sites for the global validation and intercomparison of land biophysical products: Proposition of the CEOS-BELMANIP. *IEEE Trans. Geosci. Remote Sens.* **44**, 1794–1803 (2006).

44. K. Yan, T. Park, G. Yan, C. Chen, B. Yang, Z. Liu, R. Nemani, Y. Knyazikhin, R. Myneni, Evaluation of MODIS LAI/FPAR product collection 6. Part 1: Consistency and improvements. *Remote Sens.* **8**, 359 (2016).

45. B. Schölkopf, A. Smola, *Learning with Kernels—Support Vector Machines, Regularization, Optimization and Beyond* (MIT Press Series, 2002).

46. J. L. Monteith, in *Symposia of the Society for Experimental Biology* (Cambridge Univ. Press (CUP), 1965), vol. 19, pp. 205–234.

47. W. K. Smith, S. C. Reed, C. C. Cleveland, A. P. Ballantyne, W. R. Anderegg, W. R. Wieder, Y. Y. Liu, S. W. Running, Large divergence of satellite and Earth system model estimates of global terrestrial CO_2 fertilization. *Nat. Clim. Chang.* **6**, 306–310 (2016).

48. S. W. Running, R. R. Nemani, F. A. Heinsch, M. Zhao, M. Reeves, H. Hashimoto, A continuous satellite-derived measure of global terrestrial primary production. *Bioscience* **54**, 547–560 (2004).

49. J. Delegido, J. Verrelst, L. Alonso, J. Moreno, Evaluation of Sentinel-2 red-edge bands for empirical estimation of green LAI and chlorophyll content. *Sensors* **11**, 7063–7081 (2011).

50. J. Delegido, J. Verrelst, C. Meza, J. Rivera, L. Alonso, J. Moreno, A red-edge spectral index for remote sensing estimation of green LAI over agroecosystems. *Eur. J. Agron.* **46**, 42–52 (2013).

51. S. B. Idso, R. D. Jackson, R. J. Reginato, Remote-sensing of crop yields. *Science* **196**, 19–25 (1977).

52. B. Marinković, J. Crnobarac, S. Brdar, B. Antić, G. Jaćimović, V. Crnojevi, Data mining approach for predictive modeling of agricultural yield data, Paper presented at the *Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment (BioSense09), Novi Sad, Serbia* (2009), pp. 1–5.

53. S. Fritz, L. See, J. C. L. Bayas, F. Waldner, D. Jacques, I. Becker-Reshef, A. Whitcraft, B. Baruth, R. Bonifacio, J. Crutchfield, F. Rembold, O. Rojas, A. Schucknecht, M. V. der Velde, J. Verdin, B. Wu, N. Yan, L. You, S. Gilliams, S. Mcher, R. Tetrault, I. Moorthy, I. McCallum, A comparison of global agricultural monitoring systems and current gaps. *Agr. Syst.* **168**, 258–272 (2019).

54. R. Fieuzal, C. M. Sicre, F. Baup, Estimation of corn yield using multi-temporal optical and radar satellite data and artificial neural networks. *Int. J. Appl. Earth Observ. Geoinform.* **57**, 14–23 (2017).

55. M. Weiss, F. Jacob, G. Duveiller, Remote sensing for agricultural applications: A meta-review. *Remote Sens. Environ.* **236**, 111402 (2020).

56. X. Zhang, M. A. Friedl, C. B. Schaaf, A. H. Strahler, J. C. Hodges, F. Gao, B. C. Reed, A. Huete, Monitoring vegetation phenology using MODIS. *Remote Sens. Environ.* **84**, 471–475 (2003).

57. M. O. Jones, L. A. Jones, J. S. Kimball, K. C. McDonald, Satellite passive microwave remote sensing for monitoring global land surface phenology. *Remote Sens. Environ.* **115**, 1102–1114 (2011).

58. P. Zhang, B. Anderson, B. Tan, M. Barlow, R. Myneni, *Geoscience and Remote Sensing Symposium (IGARSS), 2010 IEEE International* (IEEE, 2010), pp. 1815–1818.

59. N. A. Quarmby, M. Milnes, T. L. Hindle, N. Silleos, The use of multi-temporal NDVI measurements from AVHRR data for crop yield estimation and prediction. *Int. J. Remote Sens.* **14**, 199–210 (1993).

60. M. Mkhabela, P. Bullock, S. Raj, S. Wang, Y. Yang, Crop yield forecasting on the Canadian Prairies using MODIS NDVI data. *Agric. For. Meteorol.* **151**, 385–393 (2011).

61. D. K. Bolton, M. A. Friedl, Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* **173**, 74–84 (2013).

62. Y. Chen, D. Lu, E. Moran, M. Batistella, L. V. Dutra, I. D. Sanches, R. F. B. da Silva, J. Huang, A. J. B. Luiz, M. A. F. de Oliveira, Mapping croplands, cropping patterns, and crop types using MODIS time-series data. *Int. J. Appl. Earth Observ. Geoinform.* **69**, 133–147 (2018).

63. A. Kern, Z. Barcza, H. Marjanović, T. Árendás, N. Fodor, P. Bónis, P. Bognár, J. Lichtenberger, Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agric. For. Meteorol.* **260-261**, 300–320 (2018).

Citation: G. Camps-Valls, M. Campos-Taberner, Á. Moreno-Martínez, S. Walther, G. Duveiller, A. Cescatti, M. D. Mahecha, J. Muñoz-Marí, F. J. García-Haro, L. Guanter, M. Jung, J. A. Gamon, M. Reichstein, S. W. Running, A unified vegetation index for quantifying the terrestrial biosphere. *Sci. Adv.* **7**, eabc7447 (2021).

Science Advances

A unified vegetation index for quantifying the terrestrial biosphere

Gustau Camps-Valls, Manuel Campos-Taberner, Álvaro Moreno-Martínez, Sophia Walther, Grégory Duveiller, Alessandro Cescatti, Miguel D. Mahecha, Jordi Muñoz-Marí, Francisco Javier García-Haro, Luis Guanter, Martin Jung, John A. Gamon, Markus Reichstein and Steven W. Running

ARTICLE TOOLS	http://advances.sciencemag.org/content/7/9/eabc7447
SUPPLEMENTARY MATERIALS	http://advances.sciencemag.org/content/suppl/2021/02/22/7.9.eabc7447.DC1
REFERENCES	This article cites 53 articles, 3 of which you can access for free http://advances.sciencemag.org/content/7/9/eabc7447#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service