Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/7/9/eabc7447/DC1

Supplementary Materials for

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, Steven W. Running

*Corresponding author. Email: gustau.camps@uv.es

Published 26 February 2021, *Sci. Adv.* **7**, eabc7447 (2021)
DOI: 10.1126/sciadv.abc7447

This PDF file includes:

S1 Generalizing Vegetation Indices with Kernels

The new family of nonlinear vegetation indices is based on kernel methods (*38, 24*), a machine learning methodology to derive nonlinear algorithms from linear ones while still resorting to linear algebra operations. We first review the main theoretical properties of feature maps and kernel functions. Then we exemplify the framework of kernel-based vegetation indices and illustrate it with the particular case of the NDVI.

S1.1 Feature maps and kernel functions

Deriving nonlinear (kernel) indices requires the definition of a feature mapping $\phi(\cdot)$ to a Hilbert space $\mathcal{H}$ endorsed with the kernel reproducing property.

Definition S1.1 Reproducing kernel Hilbert space (RKHS). *Given a Hilbert space $\mathcal{H}$ with functions over [d], i.e. $f : \mathbb{R}^d \to \mathbb{R}$, the function $k(\cdot,\cdot) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is called reproducing kernel of $\mathcal{H}$ if $k(x,\cdot) \in \mathcal{H}$, and $\mathcal{H}$ is a RKHS.*

Property S1.1 Properties of Hilbert spaces. *A Hilbert space $\mathcal{H}$ is a space endorsed with an inner product. Let $\mathcal{H}$ be a vector space over . A function $\langle\cdot,\cdot\rangle_{\mathcal{H}} : \mathcal{H} \times \mathcal{H} \mapsto$ is said to be an inner product on $\mathcal{H}$ if: (1) $\langle\alpha_1 f_1 + \alpha_2 f_2, g\rangle_{\mathcal{H}} = \alpha_1\langle f_1, g\rangle_{\mathcal{H}} + \alpha_2\langle f_2, g\rangle_{\mathcal{H}}$; (2) $\langle f, g\rangle_{\mathcal{H}} = \langle g, f\rangle_{\mathcal{H}}$; and (3) $\langle f, f\rangle_{\mathcal{H}} \geq 0$, and $\langle f, f\rangle_{\mathcal{H}} = 0$ iff $f = 0$.*

Property S1.2 Reproducing property. *If $\forall x \in \mathbb{R}^d$ and $\forall f \in \mathcal{H}$ then $f(x) = \langle f, k(x,\cdot)\rangle$ and the product $\langle k(\cdot,x), k(\cdot,z)\rangle_{\mathcal{H}} = k(x,z)$. This is the reproducing property of the kernel. A function f can thus be represented as a linear function defined by an inner product in the vector space $\mathcal{H}$.*

S1.2 An illustrative example: NDVI

The normalized difference vegetation index is defined as NDVI $= \frac{n-r}{n+r}$, where n and r are the reflectances in the NIR and the red bands, respectively. This is a difference-ratio operation: the difference in the numerator can be cast as the 'physical' component, while the sum in the denominator is a 'normalization' factor. For the formulation of the kernel NDVI let us treat the two components separately. Given scalars $n, r \in \mathbb{R}$, $d = 1$, let us define a feature map $\phi \mapsto \phi(n) \in \mathcal{H}$ with an associated reproducing kernel $k(n,\cdot) = \langle\phi(n),\cdot\rangle_{\mathcal{H}}$, likewise for r. Now let us define two feature maps that work on the joint (n, r) feature vector:

$$\psi((n,r)) := \phi(n) - \phi(r) \in \mathcal{H} \quad \textit{and} \quad \varphi((n,r)) := \phi(n) + \phi(r) \in \mathcal{H},$$

with associated physical and normalization kernels:

$$\begin{aligned} m((n,r),(n,r)) &= \langle\psi((n,r)), \psi((n,r))\rangle_{\mathcal{H}} = k(n,n) + k(r,r) - k(n,r) - k(r,n) \\ \ell((n,r),(n,r)) &= \langle\varphi((n,r)), \varphi((n,r))\rangle_{\mathcal{H}} = k(n,n) + k(r,r) + k(n,r) + k(r,n). \end{aligned}$$

We can estimate the kernel NDVI transformation for (n, r) simply as:

$$\text{kNDVI} = \frac{m((n,r),(n,r))}{\ell((n,r),(n,r))} = \frac{k(n,k) - k(n,r)}{k(n,n) + k(n,r)}.$$

Property S1.3 *All kernels in kNDVI are positive definite. By construction ϕ leads to a positive definite kernel k. The difference between feature maps in ψ might not lead to a valid kernel because the third property of kernel functions in S1.1 could be violated because m could be negative in principle. The kernel is however symmetric since $m((n,r),(n,r)) = m((r,n),(r,n))$, and positive by construction, since $\langle \phi(n) - \phi(r), \phi(n) - \phi(r) \rangle_{\mathcal{H}} = \|\phi(n) - \phi(r)\|^2_{\mathcal{H}} \geq 0$. Actually, for the particular case of the RBF kernel function, we have $k(x,x) = 1$, and therefore $m((n,r),(n,r)) = 2(1-k(n,r)) \geq 0$ by construction since $0 \leq k(n,r) \leq 1$. Also note that $m((n,r),(n,r)) = 0$ iff $n = r$ so that $k(n,r) = 1$. Following similar arguments, the summation feature map φ also leads trivially to a positive definite kernel ℓ and $\ell((n,r),(n,r)) = 2(1 + k(n,r)) \geq 0$. In conclusion, all defined feature maps ϕ, ψ and φ need to lead to positive definite kernels k, m and ℓ respectively, and the multiplication (ratio) of kernels is a valid kernel too thus the kNDVI is a valid kernel.*

S1.3 The choice of the kernel function

The core of any kernel method in general, and the kNDVI in particular, is the appropriate definition of the kernel function, $k(a,b)$. Popular examples of valid reproducing kernels are the linear kernel, $k(a,b) = ab$, the polynomial $k(a,b) = (ab+1)^p$, $p \in \mathbb{Z}^+$, and the radial basis function (RBF) kernel, $k(a,b) = \exp(-\frac{1}{2\sigma^2}(a-b)^2)$, $\sigma \in \mathbb{R}^+$.

Property S1.4 NDVI is equivalent to kNDVI with a linear kernel function. *In the linear kernel, the associated RKHS is the space $\mathbb{R}$, and kNDVI trivially reduces to the standard linear NDVI:*

$$\textit{kNDVI} \overset{\textit{lin}}{:=} \frac{n\,n - n\,r}{n\,n + n\,r} = \frac{n-r}{n+r} = \textit{NDVI}.$$

Property S1.5 *Higher moments kernels. In polynomial kernels of degree p, kNDVI effectively only accounts for moments up to order p:*

$$\textit{kNDVI} \overset{\textit{poly}}{:=} \frac{(n\,n)^p - (n\,r)^p}{(n\,n)^p + (n\,r)^p} = \frac{n^p - r^p}{n^p + r^p}.$$

For the Gaussian kernel, the RKHS is of infinite dimension and kNDVI measures higher order spectral dependencies between the reflectances in the NIR and the red channels. In addition, note that for RBF kernel above, self-similarity $k(a,a) = 1$, and thus the kNDVI measure simply reduces to

$$\textit{kNDVI} \overset{\textit{RBF}}{:=} \frac{1 - k(n,r)}{1 + k(n,r)} = \tanh\left(\left(\frac{n-r}{2\sigma}\right)^2\right).$$

S1.4 Prescription and interpretation of the kernel parameter

In kernel methods, setting the kernel parameters is critical and has an important impact in the solution (*25*). We used in all our experiments the RBF and set the lengthscale parameter σ to the average value between NIR and red, $\sigma = 0.5(n + r)$. This prescription of σ is a reasonable choice; note that σ should reflect the notion of similarity between input data (in our case, NIR and red reflectances). It is customary in the kernel methods literature to fix it to the average distance among objects (here the reflectances in NIR and red channels). This choice can be also interpreted as a rough estimation of the pixel's albedo, see Fig. S1: higher σ are automatically selected for bare soils.

Interestingly, by virtue of this approximation, the simplified kNDVI is a convenient double non-linear transformation of NDVI as it reduces to kNDVI $= \tanh(\text{NDVI}^2)$. First NDVI is squared, and then the result is squashed with a sigmoid function. On the one hand, the squared NDVI has been proposed in (*11*) as a proxy of fAPAR times LUE, and hence very useful to estimate GPP. On the other hand, the $\tanh$ function allows to improve sensitivity at high values, such as in managed croplands, and reduce the well-known bias of NDVI at low values, where photosynthetic activity is low or non-existent.

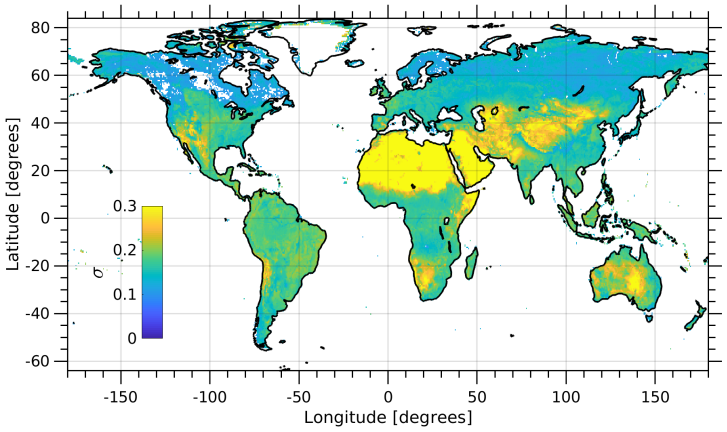

Figure S1: Distribution of the kernel parameter σ computed as the average of NIR and red, $\sigma = 0.5(n + r)$ over the 506 MODIS images (2007-2017) used in the SIF experiment.

The parameter σ directly affects the nonlinearity and may have a strong impact on the index performance. In our experiments, however, we used the mean heuristic that worked very well. Actually, optimizing σ per biome or climatic region to approximate GPP, LAI pr SIF did not improve the results much over the proposed heuristic (results not shown). The reason is that the index is already pixel-adapted. The different σ value per pixel endorses the kNDVI with a high degree of adaptation to dynamic ranges, thus resolving challenging cases of arid, dry, densely and sparsely vegetated regions. The prescribed parameter stretches the predictions to account for high dynamic ranges (e.g for GPP estimation in Fig. 3 and for LAI in Fig. S8), while reducing bias and saturation problems. This behaviour is explained by looking at the sensitivity of the index to NDVI, see S2 and Fig. S3. The suggested $\sigma = 0.5(n + r)$ actually leads to virtually no sensitivity to sparsely vegetated regions (low NDVI, avoiding the bias problem), that varies roughly linearly with NDVI for mixed-pixels (moderate NDVI) and that decreases for highly vegetated regions (high NDVI, reducing the NDVI saturation problem). This, in turn, has a positive effect in terms of error propagation, see S2.5.

S1.5 Generalizing standard vegetation indices with kernel methods

The kernel methodology can be readily applied to any vegetation index available in the literature, provided that it can be expressed as a function of dot products between spectral channels. Table S1 shows some illustrative examples of kernelized indices.

Firstly, one has then to select an appropriate kernel function k (e.g. linear, polynomial, or RBF). We recommend the Gaussian kernel -RBF kernel- function because it captures all higher-order relations between the spectral channels involved, it only contains one hyperparameter to choose, and generally gives good results in many applications. Secondly, one has to choose, or optimize, the kernel parameter(s). This can be very challenging and problem dependent. While for the kNDVI the prescription of setting the σ parameter as the average between NIR and red reflectances worked very well, this can be troublesome in other 'kernelized' indices because of the nature of relations between the involved channels.

Table S1: Examples of vegetation indices and their kernel versions.

Indices	**Example**	**VI**	**Kernelized VI**
Ratio	GI (*39*)	$\frac{R_1}{R_2}$	$\frac{k(R_1,R_1)}{k(R_1,R_2)}$
Percentage	IPVI (*40*)	$\frac{R_1}{R_1+R_2}$	$\frac{k(R_1,R_1)}{k(R_1,R_1)+k(R_1,R_2)}$
2-bands	NDVI (*3*)	$\frac{R_1-R_2}{R_1+R_2}$	$\frac{k(R_1,R_1)-k(R_1,R_2)}{k(R_1,R_1)+k(R_1,R_2)}$
3-bands	EVI (*9*)	$\frac{G(R_1-R_2)}{R_1+C_1R_2-C_2R_3+L}$	$\frac{G(k(R_1,R_1)-k(R_1,R_2))}{k(R_1,R_1)+C_1k(R_1,R_2)-C_2k(R_1,R_3)+k(R_1,L)}$
3-bands	VARI (*41*)	$\frac{R_1-R_2}{R_1+R_2-R_3}$	$\frac{k(R_1,R_1)-k(R_1,R_2)}{k(R_1,R_1)+k(R_1,R_2)-k(R_1,R_3)}$
Area	NAOC (*42*)	$1-\frac{\sum_{\lambda=\lambda_1}^{\lambda_2}R_\lambda}{R_2(\lambda_2-\lambda_1)}$	$1-\frac{\sum_{\lambda=\lambda_1}^{\lambda_2}k(R_2,R_\lambda)}{k(R_2,R_2)(\lambda_2-\lambda_1)}$

S2 Mathematical properties of kNDVI

We give some mathematical properties of the kernel NDVI that ensure its generality: the kNDVI generalizes NDVI and NIRv, it captures all (infinite) higher-order moments of the NIR and red band relations when the RBF kernel function is used, the kNDVI adapts to sparsely-vs-densely vegetated areas by means of the kernel parameter, and the index propagates less uncertainty in the spectral bands.

Property S2.1 A kernel vegetation index generalizes its original vegetation index counterpart. *The kernel version of an index reduces to the standard counterpart when linear kernels are used. As an example, using the linear kernels* $k(n, r) = n\, r$ *and* $k(n, n) = n\, n$ *into Eq.* (1)*, it is easy to show that the kNDVI reduces to the standard NDVI.*

Property S2.2 The NIR$_v$ in (*20*) is a particular case of kNDVI. *The NIR$_v$ index proposed in (20) departs from the standard NDVI and assumes that pixel reflectance* x *is composed of a portion* δ *of vegetation and* $1 - \delta$ *of soil, i.e.* $x = \delta x^v + (1 - \delta)x^s$ *for every wavelength* λ*. Then, by assuming that the soil component remains roughly constant across the spectrum,* $n^s \approx r^s$*, and that for the vegetation component the NIR reflectance is typically much higher than the red reflectance,* $n^v \gg r^v$*, one can show that NIR$_v$* $= \delta n^v \approx$ *NDVI* $\times\, n$*. Now, it is easy to show that there exist a* σ *parameter in the proposed kNDVI that yields the same result as NIR$_v$. Essentially, using an RBF kernel in the kNDVI and isolating* σ *from the equation*

$$\textit{kNDVI} = \frac{1 - k(n, r)}{1 + k(n, r)} = \textit{NDVI} \times n,$$

it is easy to show that using

$$\sigma = \sqrt{\frac{n - r}{2\sqrt{\textit{atanh}(\textit{NDVI}\; n))}}}$$

returns NIR$_v$, and therefore demostrating how n$_v$ is a particular case of kNDVI.

Property S2.3 Any kernelized vegetation index with a Gaussian kernel exploits all relations between the considered spectral bands. *We show that replacing a dot product with a kernel function, in particular the Gaussian RBF kernel function, allows us to account for all higher-order moments of similarity between the involved spectral bands. Let us assume the kernel* $k(a, b) = \langle \phi(a), \phi(b) \rangle = \exp(-\gamma(a - b)^2)$*, where for simplicity we define* $\gamma = 1/(2\sigma^2) > 0$*. Then, the explicit feature map* ϕ *is infinite dimensional, and can be expressed as*

$$\phi(a) = \exp(-\gamma a^2) \left[1, \sqrt{\frac{2\gamma}{1!}} a, \sqrt{\frac{(2\gamma)^2}{2!}} a^2, \sqrt{\frac{(2\gamma)^3}{3!}} a^3, \dots \right]^\top.$$

Note that the kernel $k(n, r) = \langle \phi(n), \phi(r) \rangle_{\mathcal{H}}$ *is thus a dot product between infinite-dimensional expansions of both* n *and* r*, and thus the kernel summarizes the all higher order differences between the NIR and red reflectance bands as* $k(n, r) = \sum_{t=0}^{\infty} (-1)^t \gamma^t (n - r)^{2t}/t!$

Figure S2 compares the correlation between SIF and different indices (NDVI, NIRv and kNDVI with polynomial and RBF kernels). Using a polynomial kernel for kNDVI with $p = 1$ recovers the solution of NDVI, while as p increases, higher order relations between the red and NIR bands are captured. In the limit, using the RBF kernel exploits all higher order relations and shows the best correlation, improving results over NIRv.

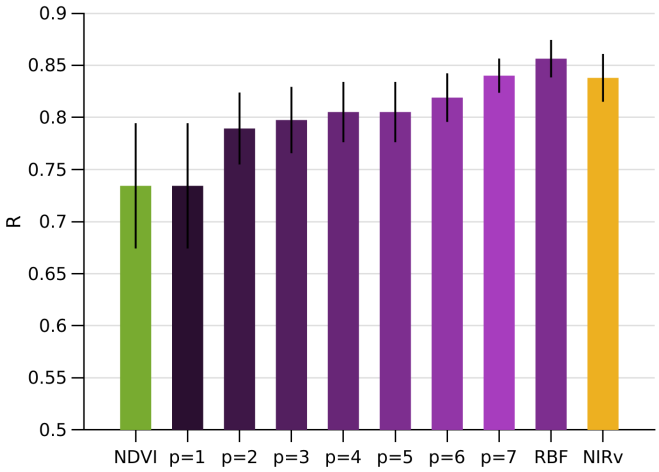

Figure S2: Correlation coefficient R (average and standard deviation) between SIF and the considered indices: NDVI, kNDVI with polynomial and RBF kernels, and NIRv.

Property S2.4 Sensitivity maps of the index. *The derivative of the kNDVI with respect its linear counterpart NDVI, can be easily computed from the complete expression of kNDVI,*

$$\mathrm{kNDVI} = \tanh\left(\left(\frac{n-r}{2\sigma}\right)^2\right) = \tanh\left(\left(\frac{\mathrm{NDVI}}{2\tau}\right)^2\right), \tag{4}$$

where for convenience we used a lengthscale parameter σ that scales linearly with the average of NIR and red reflectances $\sigma = \tau(n + r)$. The derivative can be readily obtained:

$$\frac{d\,\mathrm{kNDVI}}{d\,\mathrm{NDVI}} = \frac{1}{2\tau^2}(1 - \mathrm{kNDVI}^2)\,\mathrm{NDVI}\,.$$

Note that with our recommended value $\tau = 0.5$, the index largely simplifies, $\mathrm{kNDVI} = \tanh(\mathrm{NDVI}^2)$ *and the derivative becomes* $\frac{d\,\mathrm{kNDVI}}{d\,\mathrm{NDVI}} = 2(1 - \mathrm{kNDVI}^2)\,\mathrm{NDVI}\,.$

The value, and thus the sensitivity, of the new index strongly depends of the selected σ (through τ) parameter, see Fig. S3. The higher the σ (or τ) value, the lower the derivative and hence more sensitive to densely vegetated regions. On the contrary, the lower the σ (or τ) value, the more sensitive will be the kernel index to sparsely vegetated regions. The selection of σ has an impact on the

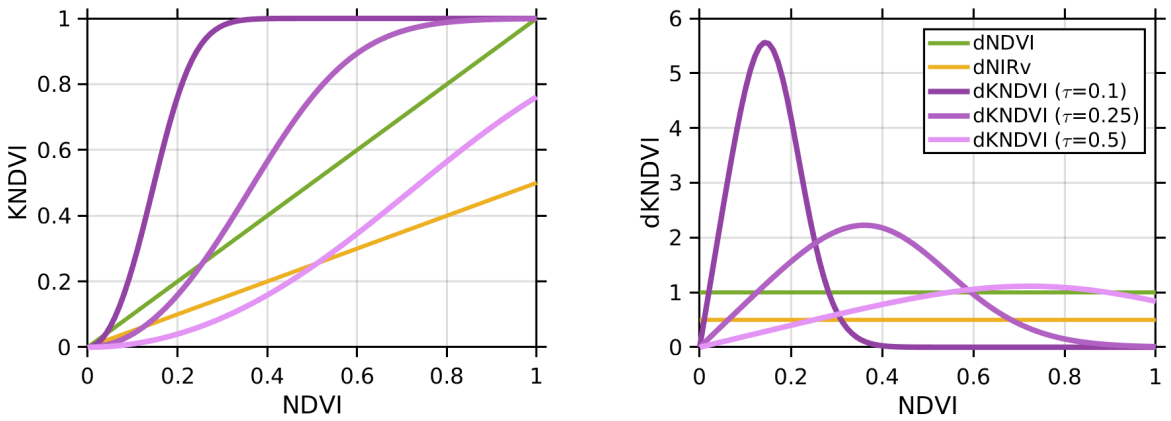

Figure S3: Derivative (sensitivity) of kNDVI and NIRv with respect NDVI (right) and dependence of the indices with NDVI (left) for different values of $\tau \propto \sigma$ (we assumed an arbitrary value of NIR reflectance of 0.5 for the NIRv illustration purposes).

desaturation effect of the index. Lower values of τ would increase the sensitivity to soils and sparsely vegetated pixels. A $\tau = 0.25$ would lead to Gaussian-like sensitivity around NDVI=0.4 but would emphasize too much the lower values and would not reduce the saturation of NDVI at high values. The suggested $\tau = 0.5$, on the contrary, leads to virtually no sensitivity to sparsely vegetated regions (low NDVI, avoiding the bias problem), then varies roughly linearly with NDVI for mixed-pixels (moderate NDVI) and then decreases for highly vegetated regions (high NDVI, reducing the NDVI saturation problem). Note that, unlike NIRv whose sensitivity increases linearly with NIR, the kNDVI with $\sigma = 0.5(n + r)$ copes with the saturation problem with a nonlinear function. In principle, one could optimize the τ value per biome of climatic region to increase the sensitivity or reduce the bias. In our experiments, however, $\tau = 0.5$ showed a good compromise between accuracy and simplicity.

Property S2.5 Error propagation. *Let us compare the indices in terms of uncertainty propagation in the spectral bands. Given the transformation* $\mathrm{kNDVI} = \tanh(((n - r)/(2\sigma))^2)$*, and independent distortions in each channel with standard deviations* σ_n *and* σ_r*, one can calculate the first order linear approximation of the error propagation by using the variance formula:*

$$\sigma^2(\mathrm{kNDVI}) = \left(\frac{d\,\mathrm{kNDVI}}{dn}\right)^2 \sigma_n^2 + \left(\frac{d\,\mathrm{kNDVI}}{dr}\right)^2 \sigma_r^2,$$

where the derivatives of kNDVI with respect the reflectances of the NIR and the red bands are:

$$\frac{d\,\mathrm{kNDVI}}{dn} = \frac{(n-r)}{2\sigma^2} sech^2\left(\left(\frac{n-r}{2\sigma}\right)^2\right) \quad \textit{and} \quad \frac{d\,\mathrm{kNDVI}}{dr} = -\frac{(n-r)}{2\sigma^2} sech^2\left(\left(\frac{n-r}{2\sigma}\right)^2\right).$$

The error propagation for the NDVI involve

$$\frac{d\,\mathrm{NDVI}}{dn} = \frac{2r^2}{(n+r)^2} \quad \textit{and} \quad \frac{d\,\mathrm{NDVI}}{dr} = -\frac{2n^2}{(n+r)^2}$$

and for the NIRv involve

$$\frac{d\,\mathrm{NIRv}}{dn} = \frac{(n^2 + 2nr - r^2)}{(n+r)^2} \quad \textit{and} \quad \frac{d\,\mathrm{NIRv}}{dr} = -\frac{2n^2}{(n+r)^2}.$$

See a comparison between the three indices in Fig. S4. Results suggest that the kNDVI propagates a lower amount of error than the rest of the indices, especially resistant to increased noise variance, which may result in more robust estimates.

(a) $\sigma_n = \sigma_r = 0.01$ (b) $\sigma_n = \sigma_r = 0.05$ (c) $\sigma_n = \sigma_r = 0.10$

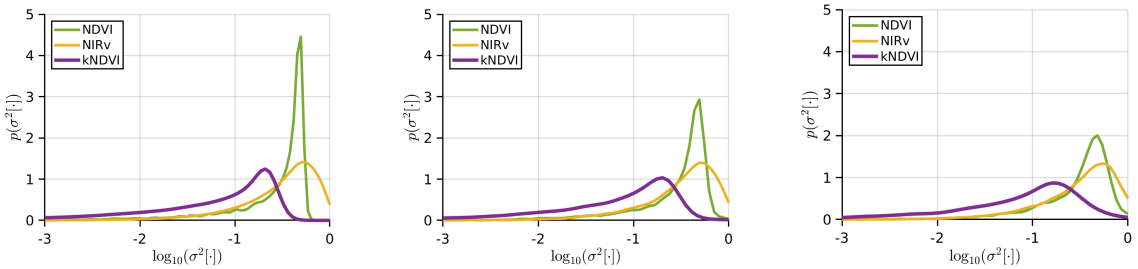

Figure S4: Density p of the propagated errors by all indices (σ^2[NDVI], σ^2[NIRv] and σ^2[kNDVI]) over a uniform grid of 10^4 combinations of NIR and red reflectance values, and fixing $\sigma_n = \sigma_r$ to 0.01 (a), 0.05 (b) and 0.1 (c) standard deviation of additive white Gaussian noise (distortion/error level) in each channel.

S3 IGBP groups

Table S2: IGBP classification.

Class	IGBP	Acronym
0	Water	WAT
1	Evergreen Needleleaf Forest	ENF
2	Evergreen Broadleaf Forest	EBF
3	Deciduous Needleleaf Forest	DNF
4	Deciduous Broadleaf Forest	DBF
5	Mixed Forest	MF
6	Close Shrublands	CSH
7	Open Shrublands	OSH
8	Woody savannas	WSA
9	Savannas	SAV
10	Grasslands	GRA
11	Permanent wetlands	WET
12	Croplands	CRO
13	Urban and built-up	URB
14	Cropland/Natural vegetation mosaic	CVM
15	Snow and ice	SNO
16	Barren or sparsely vegetated	BSV

Table S3: The considered IGBP classes and their grouping in our study.

Class	Name	Acronym	IGBP classes cf. S2 merged
C1	Needleleaf Forest	NF	1+3
C2	Evergreen Broadleaf Forest	EBF	2
C3	Deciduous Broadleaf Forest	DBF	4
C4	Mixed forest	MF	5
C5	Shrublands	SH	6+7
C6	Savannas	SAV	8+9
C7	Herbaceous	GRA	10
C8	Cultivated	CRO	12

S4 Correlation with remotely-sensed Leaf Area Index (LAI)

LAI is a key biophysical parameter for both Earth vegetation modelling and monitoring. Many studies have reported nonlinear empirical relations between NDVI and LAI. However, it is acknowledged that this relation varies temporally according to the phenological development of plants and trees, as well as with the changing environmental conditions (*43*). The correlation of kNDVI with LAI, also compared to both NDVI and NIRv, is presented here.

S4.1 LAI data and surface reflectances

The MCD43A4 and MCD15A3H MODIS v006 products were used as reflectance data and LAI estimates, respectively. Both satellite products are provided at 500 m spatial resolution and generated combining data from Terra and Aqua spacecrafts. They are disseminated from the Land Processes Distributed Active Archive Center (LP DAAC) also available at Google Earth Engine (GEE). MCD43A4 offers a daily global Bidirectional Reflectance Distribution Function (BRDF) product from a Nadir view in seven MODIS land bands (red, near infrared, blue, green, short wave infrared-1, short wave infrared-2, and middle wave infrared). The MCD43A2 MODIS product, which contains the quality information for the corresponding MCD43A4 product, was also used for avoiding low-quality BRDF estimates. The MCD15A3H collection 6 product provides LAI estimates every 4 days, and uses for the retrieval a look-up-table (LUT) approach simulated from a 3-D radiative transfer model. The product also provides with a quality flag information of the LAI estimates.

S4.2 Processing

We used GEE for processing the MODIS products' time series over 445 global biome-representative sites from July 4, 2002 to March 14, 2017. The selected sites belong to the BEnchmark Land Multisite ANalysis and Intercomparison of Products dataset (BELMANIP) (*44*). It was built using 420 sites from existing experimental networks (FLUXNET, AERONET, VALERI, BigFoot, etc) completed with selected sites from the GLC2000 land cover map. The updated one, BELMANIP2.1 dataset complements BELMANIP by adding 25 sites corresponding to bare soil areas (deserts) and tropical forests (Figure S5). Site selection was performed by keeping the same proportion of biome types within the selected sites as within the 10^o-width latitudinal bands. Attention was paid so that the sites were homogeneous over a $10\times$ 10 km^2 area, almost flat, and with a minimum proportion of urban area and permanent water bodies.

Since the used MODIS products differ in temporal frequency of production, only coincident dates among them were selected. The MCD43A4 was used to compute the indices after filtering non-valid pixels. This was carried out excluding clouds, cloud shadows, snow, as well as poor-quality BRDF parameter retrievals according to the pixel-based quality flag provided by the MCD43A2 MODIS product, which is also available in GEE. In addition, only LAI estimates provided by the MCD15A3H main algorithm were used, and intentionally filtered out estimates from the back-up algorithm as they

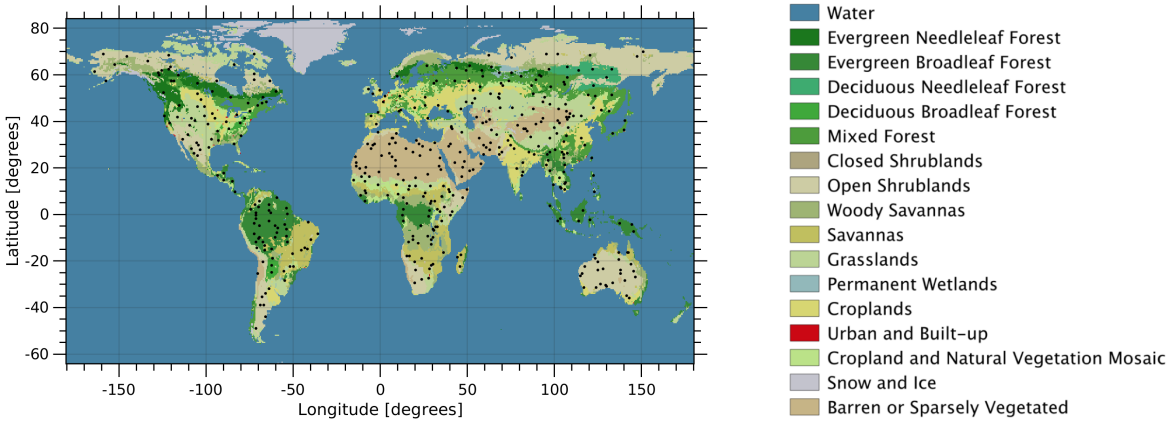

Figure S5: Location of the BELMANIP2.1 sites and associated biomes.

internally use NDVI and related biophysical parameters. Hence, we did not use MODIS-derived LAI estimates that can be affected by NDVI to avoid biased results and conclusions. This yielded 60,078 observations. LAI correlations with kNDVI, NIRv, and NDVI, were computed using these observations in the temporal domain. Lastly, the correlations are also reported per global biomes.

S4.3 Results

We evaluated our proposed kNDVI as a proxy for LAI over a large dataset of MODIS LAI estimates. Results indicate that kNDVI (R=0.81) correlates better with the MODIS LAI product than NDVI (R =0.74) and NIRv (R=0.76), see details in Table S4. These results are observed over all biomes and conditions (Fig. S6).

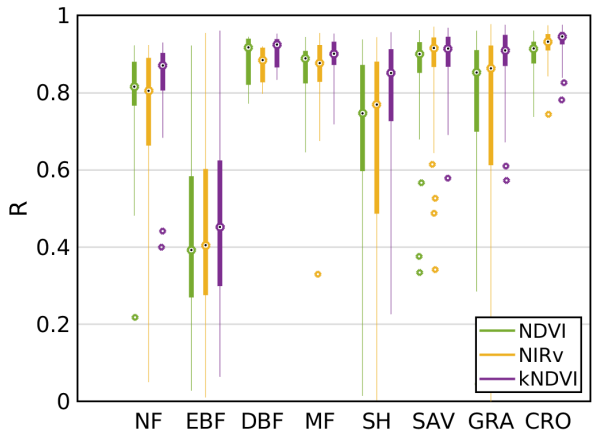

Figure S6: Boxplots of the averaged correlations between LAI and NDVI, NIRv and, kNDVI per biome type.

Table S4: Correlation coefficient between the three vegetation indices (NDVI, NIRv, kNDVI) and LAI per biome. Darker green indicates higher correlation values.

		LAI		
	Biome	**NDVI**	**NIRv**	**kNDVI**
C1	**NF**	0.75	0.79	**0.82**
C2	**EBF**	0.41	0.43	**0.45**
C3	**DBF**	0.87	0.88	**0.91**
C4	**MF**	0.85	0.86	**0.89**
C5	**SH**	0.63	0.67	**0.79**
C6	**SAV**	0.86	0.87	**0.89**
C7	**GRA**	0.74	0.79	**0.89**
C8	**CRO**	0.90	0.93	**0.94**
	ALL	0.74	0.76	**0.81**

Assessment per biome type reveals kNDVI as the most correlated index wih LAI (see Fig. S6). In general the correlations are high, except over EBF the correlation is clearly lower. This can be due to the fact that the MODIS LAI retrieval rate of the main algorithm is very low in the case of EBF caused by reflectance saturation (*45*). In addition, the distribution of correlations reveals that kNDVI outperforms both NDVI and NIRv (Fig. S7).

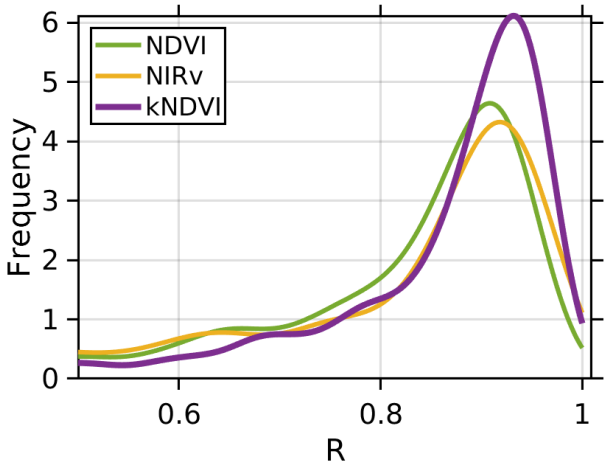

Figure S7: Estimated density of the correlation coefficient between the indices and LAI.

We show the temporal evolution of the considered indices and LAI over both cultivated and herbaceous areas, see Fig. S8. The time series reveal that kNDVI follows similarly the LAI temporal behaviour whereas NDVI performance is worse mainly in sparse vegetation periods. The index actually

adapts better to phenological cycles, and is more sensitive to low vegetation too (see Fig. S8). The kNDVI values are close to zero when no (or sparse) vegetation is present, whereas NDVI systematically retrieves values around 0.2. This highlights the normalization power of kNDVI in very early phenological stages that present high brightness variability in the underlying soil background.

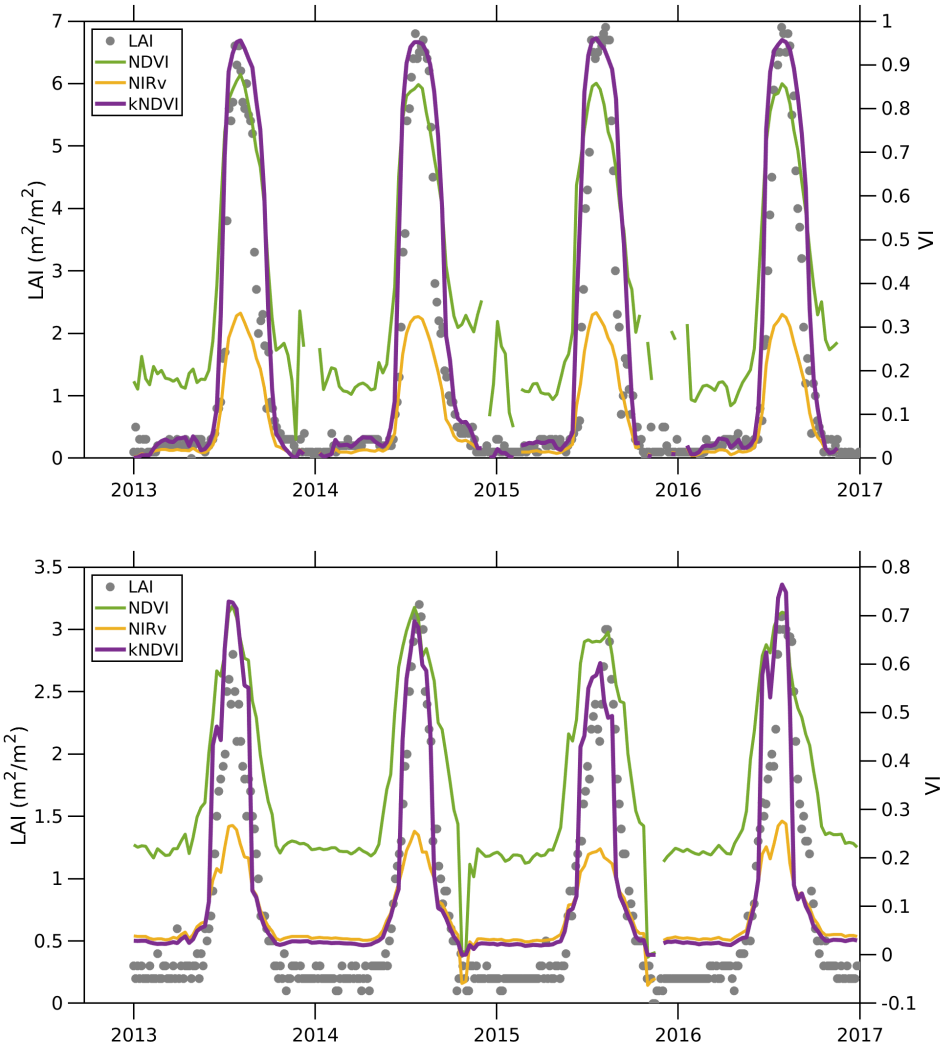

Figure S8: Time series over a cultivated area (top) and an herbaceous area (bottom) in the BELMANIP2.1 collection during the period 2013-2016.

S5 Additional analysis of GPP results

S5.1 Quantification of tower-level correlations per biome type

The per biome type assessment reveals that kNDVI generally outperforms the rest of VIs to predict GPP estimates over 4 of 7 considered biomes types (see Fig. S9). Correlations are moderate to high in all biomes, except for the EBF biome type where none of the considered VI performs adequately. This can be attributed to reflectance saturation issues (*45*)

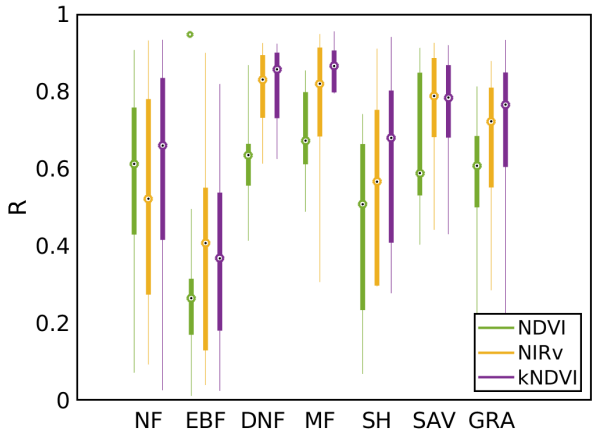

Figure S9: Boxplots of correlations between GPP and NDVI, NIRv and, kNDVI per biome type.

S5.2 On the linearization effect of normalizing GPP with radiation

Here we compare the effect of normalizing the GPP by PAR on the indices performance. Table S5 shows the results of association between the different indices (NDVI, NIRv and kNDVI) on both situations (GPP and GPP/PAR). In the comparison we used different measures of association (that is, statistical dependence) both linear and nonlinear; Pearson's correlation coefficient R; Spearman's correlation coefficient, RS (*26*); Mutual information, MI (*27*); and Distance Correlation, DC (*28*). Such analysis is imperative to have a clear view of the impact of the normalization on the indices.

S5.2.1 On the linearization via normalization versus the implicit linearization via kernels.

Results suggest that normalizing GPP by PAR has an obvious linearization effect since differences between indices are smaller independently of the dependence measure used (note that while tempting, one should not compare the scores obtained in the normalized versus the unnormalized case as a nonlinear transformation is applied and they cast different problems now). It is also observed that such normalization affects NIRv the most, which yields virtually no numerical difference with NDVI.

Table S5: Average results obtained with different measures of dependence between the indices and GPP or GPP/PAR.

	GPP			GPP/PAR		
	NDVI	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**
R	0.59	0.68	**0.68**	0.58	0.58	**0.60**
RS	0.62	**0.66**	0.63	0.58	0.56	**0.58**
MI	0.31	0.46	**0.47**	0.33	0.34	**0.36**
DC	0.64	0.71	**0.71**	0.65	0.65	**0.66**

A noticeable gain is still obtained with the proposed kNDVI. After all, kernel methods in general, and the kNDVI in particular, implement the original operation –the NDVI– in a feature space where NIR and red have been mapped to. The kNDVI is a linear operation in that space, which is nonlinear in the original (bands) input space. The simplicity and elegance of the framework allows us to accomplish the ever-sought linearization transformation implicitly. This means that no *ad hoc* parametric transformations are needed, just the kernel trick (*25, 46*). But, this also implies that virtually no gain over other indices will be obtained when the relation between the bands and the parameter of interest is linear, such as for instance after PAR normalization or when working (averaging) over larger spatial or temporal scales (see S6.3). Our results showed that the kNDVI improved results in all cases but, as expected, the gain was moderate when the domain was previously linearized.

S5.2.2 On the linear versus nonlinear regime

GPP is routinely estimated from satellite data with the light use efficiency (LUE) model (*47,48,49,50*). It is a simple model which consists of the product of the photosynthetically active radiation (PAR), the fraction of PAR absorbed by the vegetation (fAPAR), and an energy conversion efficiency factor or LUE. Within the LUE modelling logic, the fAPAR (often calculated as a linear function of the NDVI) is in charge of capturing the dynamics in photosynthetic biomass (green leaves, green stems, and shoots), while the LUE and PAR variables provide the relationship between GPP and light. However, the LUE model assumes a linear relationship between the GPP and the absorbed PAR, which is valid in a broad range of biomes and environmental conditions but breaks at high temporal resolutions (daily variation) due to nonlinear asymptotic light saturation effects, which is not the case of the present study. This seems to be the reason why the weekly GPP/PAR appears not to be greatly benefited by using higher-order (nonlinear) approaches like the kNDVI.

S6 Additional analysis of SIF results

S6.1 Spatial correlations

An alternative study with SIF was done computing the spatial correlation and averaging results through time. The overall average correlation over the 506 images (16-daily, 0.5^o), see Table S6, shows outstanding results of kNDVI ($R = 0.84$) over NDVI ($R = 0.69$), and improves performance over NIRv ($R = 0.81$). The kNDVI excels in characterizing all vegetation types (gains in R of +21.7% over NDVI and +3.7 over NIRv). Interestingly, in needle-leaf forests, kNDVI largely improves NDVI (gain of +18.5%) but performs slightly worse than NIRv (-6.7%). Accuracy of the kNDVI ($R = 0.82$) is also higher than NDVI ($R = 0.64$) or NIRv ($R = 0.80$) at different latitudes, yet far more noticeable in higher latitudes ($\geq 30^o$). This matches results when disaggregated by climatic zones (Köppen regions): the index achieves averaged improvements in correlation above +35% with regard to the NDVI and around +3% over NIRv in cold regions.

Table S6: Spatial correlation coefficients between the vegetation indices and SIF per biome. Greener colors indicate higher correlations.

		Spatial SIF		
	Biome	**NDVI**	**NIRv**	**kNDVI**
C1	**NF**	0.63	**0.80**	0.74
C2	**EBF**	0.65	0.79	**0.81**
C3	**DBF**	0.49	**0.53**	0.51
C4	**MF**	0.59	0.75	**0.77**
C5	**SH**	0.65	0.82	**0.84**
C6	**SAV**	0.77	0.83	**0.84**
C7	**GRA**	0.72	0.82	**0.84**
C8	**CRO**	0.71	0.83	**0.85**
	ALL	0.69	0.81	**0.84**

S6.2 Monthly and seasonal correlations

Figure S10 shows the obtained correlations between the indices and SIF for the whole period 2007-2018 grouped by month and season. kNDVI and NIRv perform similarly in all cases and much better than NDVI. A noticeable gain is observed during the SON months and Fall season.

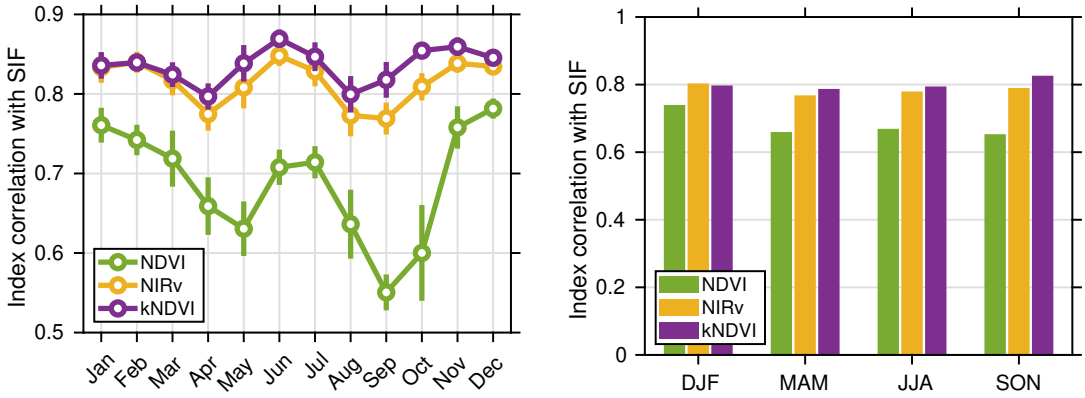

Figure S10: Monthly correlations between the index and SIF for all considered biomes and all 10 years of data (left) and analysis per season (right).

S6.3 Impact of spatial and temporal scales

We analyze here the correlation between SIF and the indices at different temporal (biweekly, monthly and bimonthly) and spatial (0.5, 1, 2) scales, see Fig. S11. Results confirm that kNDVI is more competitive at finer temporal resolutions with a noticeable advantage over NDVI (+15%) and NIRv (+4%), but the gain over NIRv disappears at bimonthly scales. A broader spatial aggregation tends to improve results of all indices and kNDVI outperforms the others independently of the spatial scale.

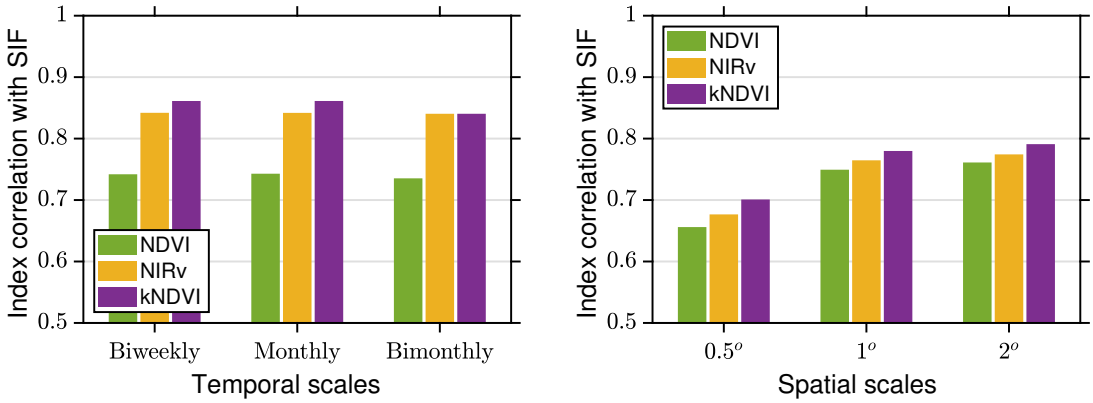

Figure S11: Average global correlation between the indices and SIF at different temporal (biweekly, monthly and bimonthly) and spatial (0.5, 1, 2) scales.

S6.4 On the linearization effect of normalizing SIF with radiation

Here we study the impact of considering SIF normalized by radiation instead of the raw SIF (to create an expression of "SIF efficiency"). In our study we approximated PAR with the cos(SZA) and studied the effect of such normalization, PAR/cos(SZA), on the results. The idea behind this is to 'discount' the associations due to seasonality. In the unnormalized case, the nonlinear similarity measures (Spearman, mutual information, and distance correlation) agree with Pearson's correlation, and are favourable to kNDVI, see Table S7. When SIF is normalized, all measures still indicate that the proposed kNDVI aligns better, yet results are deemed similar to NIRv. Note that kernel methods in general, and kNDVI in particular, solve a linear problem in a nonlinearly transformed space. Since the main effect of dividing SIF by the cos(SZA) is to linearize the problem, getting rid of the strong nonlinear seasonal cycle that dominates the distribution, a significant improvement over NIRv is not expected. Yet, still for all measures the index tends to generalize (improve) both indices. These results are also observed per biome (see Table S8), climate zone (see Table S9) and latitude (see Table S10).

Table S7: Average results over time obtained with different measures of dependence between the indices and SIF or SIF/cos(SZA) as a proxy to PAR normalization. Greener colors indicate higher values of linear and nonlinear association.

	SIF			SIF/cos(SZA)		
	NDVI	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**
R	0.54	0.63	**0.65**	0.55	0.58	**0.59**
RS	0.53	0.62	**0.65**	0.55	0.58	**0.59**
MI	0.33	0.43	**0.45**	0.34	0.37	**0.38**
DC	0.59	0.65	**0.67**	0.60	0.62	**0.63**

Table S8: Averaged temporal correlation between the indices and SIF/cos(SZA) per biome. Greener colors indicate better linear and nonlinear association values.

		R			RS			MI			DC		
	Biome	**NDVI**	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**
C1	**NF**	0.65	0.69	**0.70**	0.67	0.69	**0.70**	0.39	0.44	**0.44**	0.69	0.71	**0.71**
C2	**EBF**	0.65	0.69	**0.69**	0.67	**0.70**	**0.70**	0.41	0.46	**0.46**	0.69	**0.72**	**0.72**
C3	**DBF**	**0.46**	0.45	0.40	**0.47**	0.45	0.39	**0.21**	**0.21**	0.19	**0.49**	0.47	0.44
C4	**MF**	0.60	0.64	**0.65**	0.61	**0.64**	**0.64**	0.38	0.42	**0.43**	0.64	0.66	**0.66**
C5	**SH**	0.63	0.66	**0.67**	0.63	**0.66**	**0.66**	0.39	0.43	**0.43**	0.67	0.69	**0.69**
C6	**SAV**	**0.57**	0.56	0.55	**0.57**	0.56	0.54	**0.34**	**0.34**	**0.34**	**0.60**	0.59	0.59
C7	**GRA**	0.50	0.55	**0.58**	0.50	0.55	**0.58**	0.29	0.33	**0.34**	0.55	0.58	**0.60**
C8	**CRO**	0.54	0.59	**0.61**	0.55	0.59	**0.60**	0.30	0.34	**0.35**	0.58	0.61	**0.62**
	ALL	0.55	0.58	**0.59**	0.55	0.58	**0.59**	0.34	**0.37**	**0.37**	0.60	**0.62**	**0.62**

Table S9: Correlation between the indices and SIF (left) or SIF/cos(SZA) (right) per climate zone. Greener colors indicate higher correlations.

	SIF			SIF/cos(SZA)		
Climatic zone	**NDVI**	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**
A- Tropical	0.59	0.68	**0.69**	0.61	**0.65**	**0.65**
B- Arid	0.47	0.55	**0.61**	0.49	0.51	**0.56**
C- Temperate	0.61	0.72	**0.74**	0.61	0.63	**0.65**
D- Cold	0.51	0.59	**0.66**	0.53	0.55	**0.62**
E- Polar	0.66	0.74	**0.77**	0.70	**0.72**	**0.72**

Table S10: Temporal correlation between the indices SIF (left) or SIF/cos(SZA) (right) per latitude. Greener colors indicate higher correlations.

			SIF		SIF/cos(SZA)		
Latitude range		**NDVI**	**NIRv**	**kNDVI**	**NDVI**	**NIRv**	**kNDVI**
-60	-30	0.43	0.54	**0.59**	**0.53**	**0.53**	0.49
-30	0	0.50	0.57	**0.58**	**0.52**	**0.52**	**0.52**
0	30	0.65	0.73	**0.75**	0.66	0.69	**0.70**
30	60	0.53	0.61	**0.67**	0.53	0.58	**0.62**
60	90	0.64	0.72	**0.73**	0.65	0.68	**0.69**

S7 Dependence between the index and in-situ Chlorophyll content, LAI and FVC

We quantitatively assess the performance of kNDVI in real *in situ* measurements of chlorophyll content (Chl-a), leaf-area index (LAI) and fractional vegetation cover (FVC). For this purpose, we will use the SPARC dataset (*51, 52*). The SPectra bARrax Campaign (SPARC) field dataset encompasses different crop types, growing phases, canopy geometries and soil conditions. The SPARC-2003 campaign took place from 12 to 14 July in Barrax, La Mancha, Spain (coordinates 30°3'N, 28°6'W, 700 m altitude). Bio-geophysical parameters have been measured within a total of 108 Elementary Sampling Units (ESUs) for different crop types (garlic, alfalfa, onion, sunflower, corn, potato, sugar beet, vineyard and wheat). An ESU refers to a plot, which is sized compatible with pixel dimensions of about 20 m $\times$ 20 m. In the analysis no differentiation between crops was made.

Table S11: Linear and nonlinear dependence measures between the vegetation indices and the biophysical parameter.

		NDVI	NIRv	kNDVI
R	LAI	0.71	0.72	**0.78**
	fAPAR	0.75	0.83	**0.86**
	FVC	0.85	0.86	**0.88**
RS	LAI	0.45	**0.52**	0.50
	fAPAR	0.70	0.83	**0.85**
	FVC	0.63	0.79	**0.81**
MI	LAI	0.33	0.37	**0.47**
	fAPAR	0.41	0.59	**0.67**
	FVC	0.63	0.66	**0.73**
DC	LAI	0.69	0.72	**0.74**
	fAPAR	0.79	0.84	**0.85**
	FVC	0.82	0.83	**0.86**
MSE	LAI	9.44	9.95	**9.00**
	fAPAR	1.00	0.83	**0.77**
	FVC	0.17	0.16	**0.15**

The data used in this study were obtained in two terrestrial campaigns in Barrax, Spain. The test area has a rectangular form and an extent of 5 km $\times$ 10 km, and is characterized by a flat morphology and large, uniform land-use units. The region consists of approximately 65% dry land and 35% irrigated land. Several instruments were used to measure the variables: a calibrated CCM-200 Chlorophyll Content Meter for Chl-a, the LiCor LAI-2000 for LAI, and hemispherical photographs taken with a digital camera with a fish-eye lens for FVC. Simultaneously we used satellite images from the CHRIS sensor. CHRIS measures over the visible/ near-infrared spectra from 400 to 1050 nm. For this study, we used CHRIS data in Mode 1 (62 bands, full spectral information) for the four campaign days, where *in situ* measurements of surface properties were measured in conjunction with the satellite overpass. The images were geometrically and atmospherically corrected. Three sets of 135 measurements were collected in total. Results are shown in Table S11, where again kNDVI is a better proxy of the different *in situ* measurements of biophysical parameters, independently of the adopted measure: higher values of Pearson's correlation R; Spearman's correlation, RS; Mutual information, MI; and Distance Correlation, DC; and lower values of MSE of a linear fit indicate better performance.

S8 Crop yield estimation

Accurate and timely crop yield estimation is currently one of the major challenges in agricultural research and of paramount interest to governments, public administrations, and farm managers (*53, 54, 55*). Earth observation (EO) data has opened new ways for efficient agricultural mapping, crop monitoring and assessment, as it allows deriving spatially explicit and temporally resolved maps of production and yield (*56, 57*). Most studies on the use EO data for crop estimation are centered on visible and infrared sensors. Actually, optical vegetation indices are easy to compute and useful to monitor the quantity, quality and behavior of the vegetation representing the intra-annual vegetation dynamics (*58, 59, 60*). Among the most widely used VIs, the NDVI has been extensively and successfully used in agricultural mapping and monitoring, as well as in many crop yield studies (*61, 62, 63, 64, 65*).

We used five years of Multi-angle Imaging SpectroRadiometer (MISR) data over the state of Kansas (US) in the "corn belt" and derived weekly averaged time series of NDVI, NIRV and kNDVI at county scale. A total of 79 time series with co-located yield were used for 13 counties. The goal is to estimate the crop yield of both corn and wheat from the time series. The target yield comes from the U.S. Department of Agriculture (USDA) records. To evaluate the indices, we developed an extremely simple crop yield estimation model: the index time series were used as a feature vector to fit a linear regression model. We then computed the correlation coefficient between the estimated and the surveyed USDA crop yield. We also measured the maximum correlation obtained between each index and the yield, as a measure of estimation power. Results are given in Table S12. In both approaches, the kNDVI improves results over the other indices.

Table S12: Correlation coefficient between the estimated and the surveyed crop yield in two settings: (left) using the year time series in a multivariate linear regression (MLR); and (right) maximum correlation between the weekly observation and the yield.

	MLR	**R_{max} (week)**
Corn		
NDVI	0.5591	0.1960 (23)
NIRv	0.5967	0.2446 (29)
kNDVI	0.6157	0.2775 (29)
Wheat		
NDVI	0.7001	0.1591 (39)
NIRv	0.7195	0.3134 (39)
kNDVI	0.7530	0.3598 (39)

The RMSE (bushels/acre) of each model can be translated into actual production (in bushels) by normalizing over the acres planted. Information obtained from USDA.gov. Results are shown in Table S13, and reveal that the lower error obtained by using kNDVI in the linear prediction model generally translates into lower production estimates (around 330'000 bushels/year of corn and 400'000 bushels/year of wheat) compared to the standard NDVI.

Table S13: Translation of RMSE (bushels/acre) into bushels for the particular example of using a linear regression for yield estimation over Kansas.

Corn	NDVI	NIRv	kNDVI
RMSE (bushels/acre)	15.9352	16.123	15.8321
RMSE (bushels)	52108104	52722210	51770967
Diff relative to NDVI (bushels)	-	614106	-337137
Wheat	NDVI	NIRv	kNDVI
RMSE (bushels/acre)	8.3266	8.7451	8.2861
RMSE (bushels)	83682330	87888255	83275305
Diff relative to NDVI (bushels)	-	4205925	-407025

S9 Change detection

We show results of applying vegetation indices in the detection of changes in multispectral Sentinel-2b images. Two scenes are considered: natural floods caused by cyclone Debbie in Australia 2017, and consequences of wildfires in a mountainous area of California (USA), see Fig. S12. Following the standard change vector analysis (CVA) procedure, we used the absolute difference of the vegetation indices between the pre- and post-event dates as the anomaly detector.

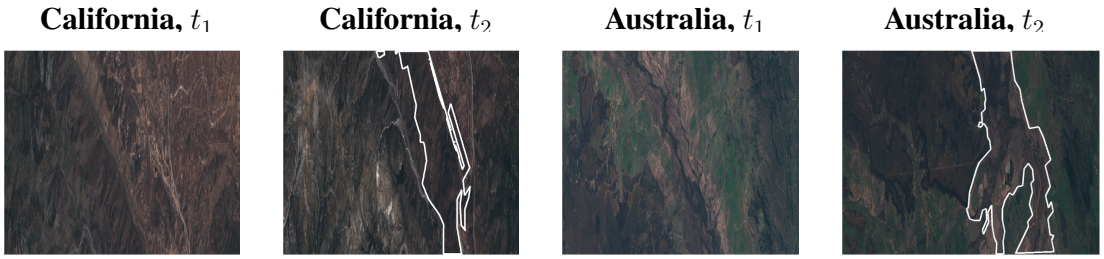

Figure S12: RGB composite S2-b pre- and post-event images of California wildfires (left) and Australia floods (right). The changed area boundary is highlighted in white, and used for computing the ROC and AUC. **Credits:** Images are freely available from ESA Copernicus Hub.

Figure S13 shows the Receiver Operating Curves (ROCs) of the indices, the area under the curves (AUC), and the change detection maps. It can be noted that the kNDVI achieves an improved detection performance over NDVI and NIRv, especially noticeable in the false positive rate regimes. This can be confirmed in the detection images, where kNDVI provides sharper detection maps.

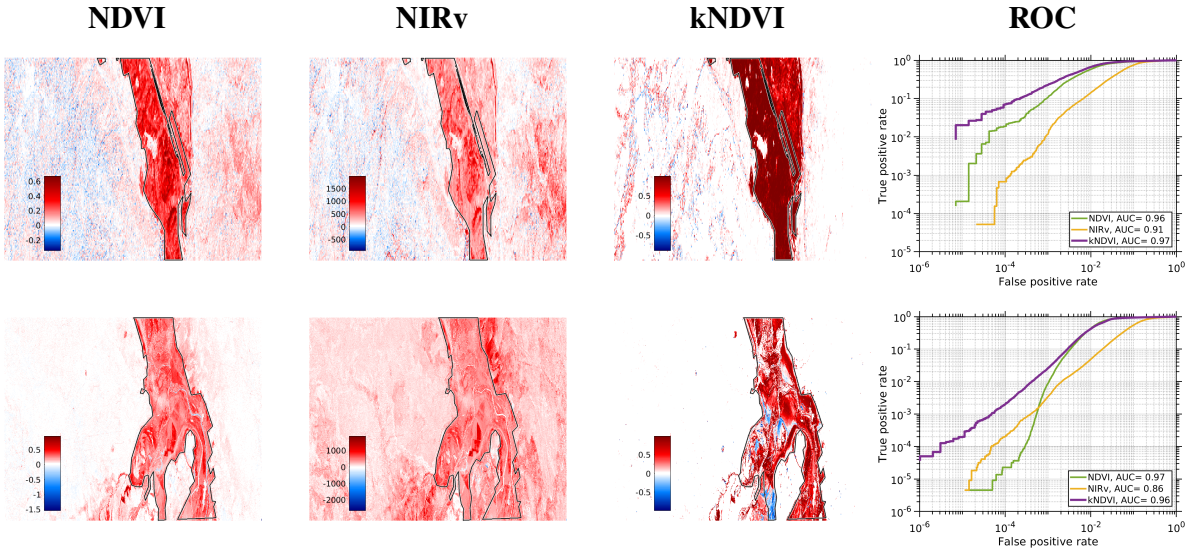

Figure S13: Difference maps by each index for California wildfires (top row) and Australia floods (bottom row) scenes. The difference is taken as the change indicator to compute the ROC and AUC (right plots).

S10 Source code implementation

The kernel-based vegetation indices can be easily programmed and applied. Here we give implementations in standard programming languages: MATLAB, R, Python, Julia, IDL and Google Earth engine (GEE) code. In all cases, and for illustration purposes, we used the standard RBF kernel function in the kNDVI.

S10.1 MATLAB

Given the NIR and RED values for a particular pixel in scalar MATLAB variables `xn` and `xr`, the kNDVI is computed as:

```
sigma = 1.0;
knr = exp(-(xn-xr)^2/(2*sigma^2));
kndvi = (1-knr)/(1+knr);
```

Listing 1: MATLAB code snippet for the kNDVI index

which can be easily computed for a whole image using right array divisions on bands.

The kernel parameter σ was fixed to 1 for illustration purposes. In our experiments we used a common heuristic in machine learning that fixes σ to the mean distance between the involved objects in the kernel similarity measure, in our case the NIR and red bands. Optimization of σ, e.g. per biome or climatic region, is also possible. However, this simple heuristic performed very well in our experiments.

S10.2 R

```
sigma <- 1
knr <- exp(-(xn-xr)^2/(2*sigma^2))
kndvi <- (1-knr) / (1+knr)
```

Listing 2: R code snippet for the kNDVI index

S10.3 Python

```
import numpy as np
sigma = 1.0
knr = np.exp(-(xn-xr)**2/(2*sigma**2))
kndvi = (1-knr) / (1+knr)
```

Listing 3: Python code snippet for the kNDVI index

S10.4 Julia

```
sigma = 1.0
knr = exp(-(xn-xr)^2 / (2*sigma^2))
```

```
kndvi = (1-knr) / (1+knr)
```

Listing 4: Julia code snippet for the kNDVI index

S10.5 IDL

Similarly to the MATLAB code, given an image loaded in IDL environment and the NIR and RED bands assigned to `nir` and `red` variables, the kNDVI is computed as:

```
sigma = 1
k = exp(-( nir*1.0-red )^2/(2*sigma^2))
kNDVI = (1-k)/(1+k)
```

Listing 5: IDL code snippet for the kNDVI index

S10.6 Google Earth Engine (GEE)

Given the NIR and red bands identified by the variables `nir` and `red`, the kNDVI is computed using a map function defined as:

```
var addKNDVI = function(image) {
  // Compute D2 a rename it to d2
  var D2 = nir.subtract(red).pow(2)
              .select([0],['d2']);

  // Gamma, defined as 1/sigma^2
  var gamma = ee.Number(4e6).multiply(-2.0);

  // Compute kernel (k) and KNDVI
  var k = D2.divide(gamma).exp();
  var kndvi = ee.Image.constant(1)
      .subtract(k).divide(
         ee.Image.constant(1).add(k))
           .select([0],['knd']);

  return image.addBands(kndvi);
}
```

Listing 6: GEE/JavaScript code snippet for the kNDVI index

We provide a simple demo in the following GEE link that computes and compares time series of NDVI, NIRv and kNDVI vegetation indices in selected areas of interest.

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).