ISPRS Journal of Photogrammetry and Remote Sensing 227 (2025) 234–250

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Mitigating NDVI saturation in imagery of dense and healthy vegetation

Zezhong Tian [a], Jiahao Fan [a], Tong Yu [a], Natalia de Leon [b], Shawn M. Kaeppler [b], Zhou Zhang [a,*]

[a] Department of Biological Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA
[b] Department of Plant and Agroecosystem Sciences, University of Wisconsin-Madison, Madison, WI 53706, USA

ARTICLE INFO

Keywords:
Mitigating saturation
Remote sensing
NDVI
Satellites and UAV
PROSAIL

ABSTRACT

The Normalized Difference Vegetation Index (NDVI) is a widely used tool for assessing vegetation in remote sensing. However, its non-linear response to increasing vegetation vigor, especially in dense and healthy canopies, often leads to inaccurate estimations of vegetation status — a phenomenon known as NDVI saturation. This study investigates the underlying causes of NDVI saturation and proposes a two-stage saturation mechanism. In contrast to the first-stage optical saturation caused by biophysical constraints, we demonstrate that second-stage mathematical saturation can be mitigated through functional optimization. To address this issue, we introduce a new vegetation index (VI), Saturation Mitigated NDVI (NDVIsm), which modifies the NDVI structure by integrating an anti-saturation module. This modification enhances the index's sensitivity to changes in vegetation vigor dynamics by amplifying variation among high NDVI values, thereby mitigating saturation effects. The performance of NDVIsm was validated across multiple remote sensing platforms and land cover types. NDVIsm eliminated saturation (defined as areas where more than 80% of pixels fall within less than 20% of the index range) and exhibited a more dispersed distribution with a pronounced right tail in the histogram. The effectiveness of NDVIsm in mitigating saturation was further confirmed by improved Pearson correlations (r) with canopy structure (0.3010 increase in Leaf Area Index (LAI) and 0.2452 increase in Leaf Structure Parameters (N)) and vegetation vigor (0.5575 increase in Chlorophyll Content (Cab)), as well as enhanced performance in machine learning (ML)-based yield prediction models when combined with non-optical and spatial features. Overall, NDVIsm demonstrates improved sensitivity in detecting subtle variations in dense and healthy vegetation and offers a promising solution to the saturation problem in environmental remote sensing.

1. Introduction

A typical and detectable feature of vegetation is the pronounced rise in reflectance from the red (R) to near-infrared (NIR) regions (Thenkabail et al., 2000), resulting from strong light absorption by chlorophyll in the R region and increased light reflection by the cellular structure in the NIR region (Knipling, 1970). Leveraging this feature, the Normalized Difference Vegetation Index (NDVI) was developed and has since become one of the most extensively used vegetation indices (VIs) in environmental remote sensing (Huang et al., 2021). It has been widely applied in monitoring and analyzing the healthy statues of vegetation (Farbo et al., 2024; Meng et al., 2013; van Leeuwen et al., 2006). However, as vegetation thrives and develops, typically with Leaf Area Index (LAI) greater than 3 (Carlson and Ripley, 1997; Mutanga et al., 2023), NDVI becomes less sensitive to the ratio between NIR and R reflectance, leading to inaccurate assessments in healthy and dense vegetation (Huete et al., 1997), which is known as the saturation challenge. With the expanding application in environmental remote sensing, the limitations of NDVI, particularly its saturation issue, has become increasingly apparent (Liu et al., 2012). This saturation issue restricts precise monitoring and further research in areas under healthy and dense vegetation, such as rainforests, farmlands and grasslands. NDVI saturation remains a persistent challenge in environmental remote sensing, attracting ongoing research and attention (Beamish et al., 2020; Huete, 2012; Khanal et al., 2020).

In recent decades, considerable efforts have been devoted to mitigating NDVI saturation through the improvement of VIs, the integration of multi-source data, and the application of machine learning (ML) techniques. The improvement of VIs, either by modifying their formulations or by introducing additional spectral bands, includes examples

* Corresponding author.
E-mail address: zzhang347@wisc.edu (Z. Zhang).

https://doi.org/10.1016/j.isprsjprs.2025.06.013
Received 8 December 2024; Received in revised form 22 May 2025; Accepted 12 June 2025
Available online 18 June 2025

such as the Enhanced Vegetation Index (EVI), Soil-Adjusted Vegetation Index (SAVI), and saturation-adjusted NDVI aim to mitigate NDVI saturation by accounting for soil background (Huete et al., 2002), atmospheric interference (Huete, 1988), or by empirically adjusting the index formulation to better capture vegetation signals under dense canopies (Gu et al., 2013; Liu et al., 2012). The integration of multi-source remote sensing data, such as synthetic aperture radar (SAR), shortwave infrared (SWIR), and LiDAR, has shown potential in reducing saturation effects in dense vegetation canopies by incorporating complementary structural or biochemical information (David et al., 2022; Kumar et al., 2022; Wei et al., 2012). ML methods can mitigate NDVI saturation by extracting latent features that are not directly accessible through conventional indices (Eugenio et al., 2020; Kussul et al., 2017; Wolanin et al., 2019). By leveraging nonlinear relationships and high-dimensional data representations, ML models enhance the detection of vegetation traits such as yield, biomass, and leaf nitrogen content that are often masked under saturation conditions (Johnson et al., 2016; Maimaitijiang et al., 2020; Safonova et al., 2021). Despite notable advances, existing methods remain constrained by high computational demands, dependence on specific sensor types, and limited adaptability across diverse vegetation conditions (Kganyago et al., 2024; Ma et al., 2024; Mutanga et al., 2023). Crucially, the fundamental mechanisms underlying NDVI saturation are still not fully elucidated, highlighting the need for a more comprehensive understanding that includes biophysical constraints and limitations inherent in the NDVI's mathematical formulation.

Therefore, we identify a two-stage saturation mechanism in NDVI. First, optical saturation, driven by biophysical constraints such as increased LAI and Leaf Structure Parameters (N), leads to both NIR and R reflectance becoming insensitive to further increases in vegetation vigor (as indicated by Chlorophyll Content, Cab) (Ciganda et al., 2008; Jiang et al., 2006; Nijland et al., 2014); Second, mathematical saturation, caused by NDVI's nonlinear formulation (Barbosa et al., 2006), further compresses spectral signals even when NIR/R still varies. While previous studies have acknowledged the mathematical limitations of NDVI formulation (Barbosa et al., 2006; Liu et al., 2012; van Leeuwen et al., 2006), they have not independently evaluated its role in exacerbating saturation beyond optical constraints. In this study, derivative analysis of NDVI with respect to the NIR/R reveals that its formulation further compresses spectral responses to biophysical changes in LAI and Cab. Specifically, the first derivative decreases monotonically and the second derivative remains negative, confirming the global concavity of the function. This inherent curvature intensifies signal compression in dense and healthy vegetation, independently of the optical saturation, and can be mitigated through mathematical optimization. To overcome this mathematical saturation, we propose the Saturation Mitigated NDVI (NDVIsm), a mathematically optimized index that modifies the curvature of NDVI in the NIR/R domain. By transforming the NDVI into a concave function with enhanced high-end sensitivity, NDVIsm selectively amplifies spectral differences among dense and healthy vegetation, offering a structural solution to NDVI's saturation problem.

To evaluate the effectiveness in mitigating NDVI saturation, we compared NDVIsm with commonly used VIs using multi-source satellite imagery covering rainforests, farmlands, grasslands at different spatio-temporal resolutions. The saturation criterion was defined firstly and applied alongside quantitative metrics to assess the histogram distribution patterns of VIs and their saturation resistance. Furthermore, Unmanned Aerial Vehicle (UAV)-based hyperspectral imagery, combined with a radiative transfer model (PROSPECT5 and 4-SAIL) and a look-up table (LUT) approach (Berger et al., 2018), was employed to investigate the improved relationships between NDVIsm and key canopy biophysical traits. The applied potential of NDVIsm in ML modeling was also assessed through a yield prediction task. In the discussion, commonly used VIs are found to remain susceptible to saturation effects. The broader applicability of NDVIsm was examined by exploring its potential variants and identifying its limitations. In addition, the relationship between NDVIsm and county-level maize yield across four central U.S. states was visually presented to highlight its potential for large-scale agricultural monitoring. In summary, this study provides new methodological framework for addressing NDVI saturation in dense and healthy vegetation and underscores the potential of saturation-resistant indices to enhance the accuracy of remote sensing-based vegetation assessment.

2. Materials and methods

2.1. Study areas

The study areas vary depending on the data collection platforms—satellite and UAV. For the satellite platform, regions with typical dense vegetation, including rainforests, farmlands, and grasslands, were selected across North and South America in 2023. Each land cover type was represented by three satellite images sourced from Sentinel-2, Landsat-8, and MODIS-09GA with their location shown in Fig. 1. For the UAV platform, a hyperspectral image captured on Aug 23, 2022, was selected from a vigorously growing and uniformly planted maize field at the Arlington Agricultural Research Station, WI, U.S. The field was planted on May 23, 2022, and harvested on November 28, 2022. It consisted of 1,200 plots, encompassing over 500 maize hybrids. Each plot consisted of two rows measuring 6.50 meters (m) in length, with 0.19 m plant spacing and a 0.75 m gap between plots. Water and fertilization management were consistent across all plots. These areas represent diverse vegetation conditions and structural complexities, providing a basis for evaluating the performance and generalizability of NDVIsm.

2.2. Specifications of remote sensing platforms and sensors

The multiple remote sensing platforms utilized in this study varied in detectable bands and spatial–temporal resolution. The key parameters and characteristics are detailed in Table1. The diversity of the remote sensing data provides a solid foundation for validating the stability of NDVIsm in addressing saturation issue.

2.3. Mitigating NDVI saturation solutions

2.3.1. Two-stage NDVI saturation mechanism

PROSAIL is a radiative transfer model that combines leaf optical properties (PROSPECT) with canopy structure (SAIL) and is widely used to simulate vegetation reflectance and retrieve biophysical parameters (Duan et al., 2014; Guo et al., 2023; Kayad et al., 2022). This study investigates the NDVI saturation phenomenon through PROSAIL (PROSPECT5 + 4-SAIL) to explore the relationship between spectral reflectance saturation and key biophysical parameters. The PROSAIL was parameterized with Cab ranging from 20 to 80 μg/cm^2, N from 1.0 to 2.0, and LAI from 0.1 to 8.0. Other parameters were fixed as follows: carotenoid = 10 μg/cm^2, brown pigment = 0.2, water thickness = 0.03 g/cm^2, dry matter = 0.008 g/cm^2, hot spot = 0.3, solar zenith angle = 27°, view zenith angle = 0°, relative azimuth angle = 0°, soil reflectance = 0.15, soil moisture = 0, leaf inclination parameter A = 50, and leaf inclination parameter B = 0. The analysis reveals a two-stage saturation mechanism:

(1) Optical Saturation Stage

As vegetation develops, the increasingly complex canopy structure—indicated by higher N and LAI—leads to near-complete absorption of NIR photons and strong scattering of red radiation (Bannari et al., 1995; Jackson and Huete, 1991; Kooistra and Clevers, 2016). As a result, both NIR and R reflectance enter a saturation phase: NIR reflectance is always insensitive to increasing Cab, while the sensitivity of R reflectance to Cab declines progressively, see Fig. 2(a-d). Given that NDVI is defined as

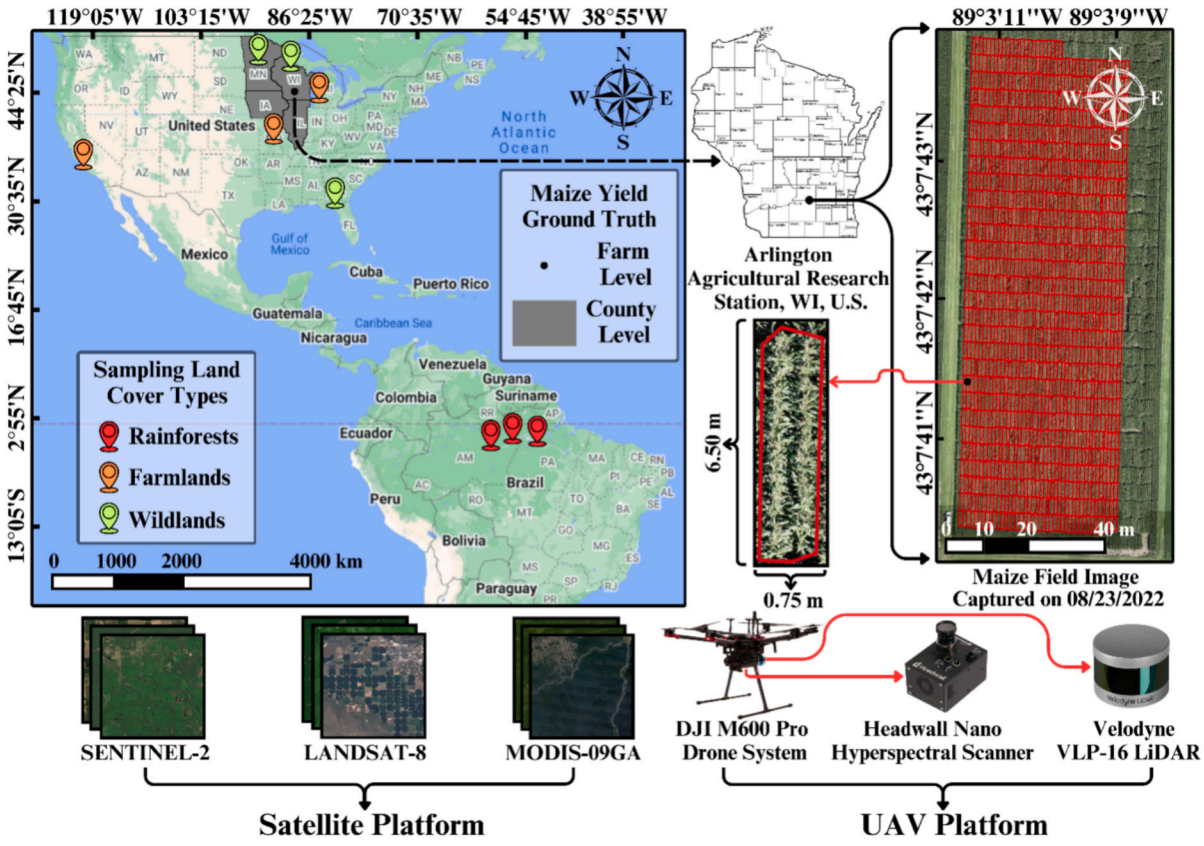

Fig. 1. Descriptions of study locations and data sources.

Table 1
Specifications of remote sensing platforms and sensors.

Platform	Model	SpectralRange (nm)	Numberof Bands	SpatialResolution	RevisitTime
UAV	Headwall Nano Hyperspec-Camera	400–1000	274	3.5 cm	–
DJI M600 Pro (60 m altitude)	VLP-16 LiDAR	–	–	100–400 points/m^2	–
Satellite	Landsat-8	450–885 (Blue-NIR)	4	30 m	16 days
	Sentinel-2	490–865 (Blue-NIR)	8	20 m	5 days
	Modis-09GA	459–876 (Blue-NIR)	4	500 m	Daily

a nonlinear function of NIR/R, saturation in NIR and R inevitably leads to saturation in NIR/R, and thus in NDVI. This results in reduced responsiveness of NDVI to vegetation vigor (indicated by Cab).

(2) Mathematical Saturation Stage

Beyond optical limitations, the inherently non-linearity of NDVI formulation further compresses spectral signals (Montandon and Small, 2008; Zhu and Liu, 2015), see Fig. 2(h). Bivariate derivative analysis (∂NDVI/∂LAI∂N) demonstrates a progressive intensification of saturation, with NDVI and NIR/R considered saturated when the gradient with respect to both LAI and N falls below 0.1, see Fig. 2(e-g):

- Initial Saturation Phase (NIR/Red): 142 points (at Cab = 20), 149points (at Cab = 45), 152 points (at Cab = 80) saturated
- Saturation Exacerbated Phase (NDVI): 190 points (at Cab = 20), 192 points (at Cab = 45), 193 points (at Cab = 80) saturated

When LAI exceeds 3.3 and N is greater than 1.6 (Fig. 2(h)), even subtle spectral variations in NIR/R are compressed into a narrow NDVI range. This compression amplifies the saturation effect and further reduces NDVI's sensitivity to vegetation vigor under dense canopy conditions (when Cab > 40).

Optical saturation, constrained by inherent vegetation properties and radiative transfer limits, is difficult to overcome (Mutanga et al., 2023). However, by constructing a mathematically optimized index guided by differential sensitivity analysis (dNDVI/d(NIR/Red)), the saturation issue of NDVI can be effectively mitigated. This decoupled two-stage mechanism offers a novel pathway for developing anti-saturation VIs by separating physical constraints from mathematical flexibility.

2.3.2. *From derivative to concavity: NDVI mathematical saturation analysis*

The nonlinear mathematical structure of NDVI imposes an additional layer of signal compression beyond the physical saturation of NIR/R. This impact can be evaluated through the derivatives:

$$\frac{d(\mathrm{NDVI})}{d(\mathrm{NIR/R})} = \frac{2}{(\mathrm{NIR/R}+1)^2} \; ; \; \frac{d^2(\mathrm{NDVI})}{d(\mathrm{NIR/R})^2} = -\frac{4}{(\mathrm{NIR/R}+1)^3} \tag{2-1}$$

The first derivative decreases monotonically as NIR/R increases, indicating a continuous loss of sensitivity in dense vegetation areas. The second order derivative remains strictly negative, confirming that NDVI is a globally concave function of NIR/R. This concavity not only reduces responsiveness but also intensifies the compression of spectral differ-

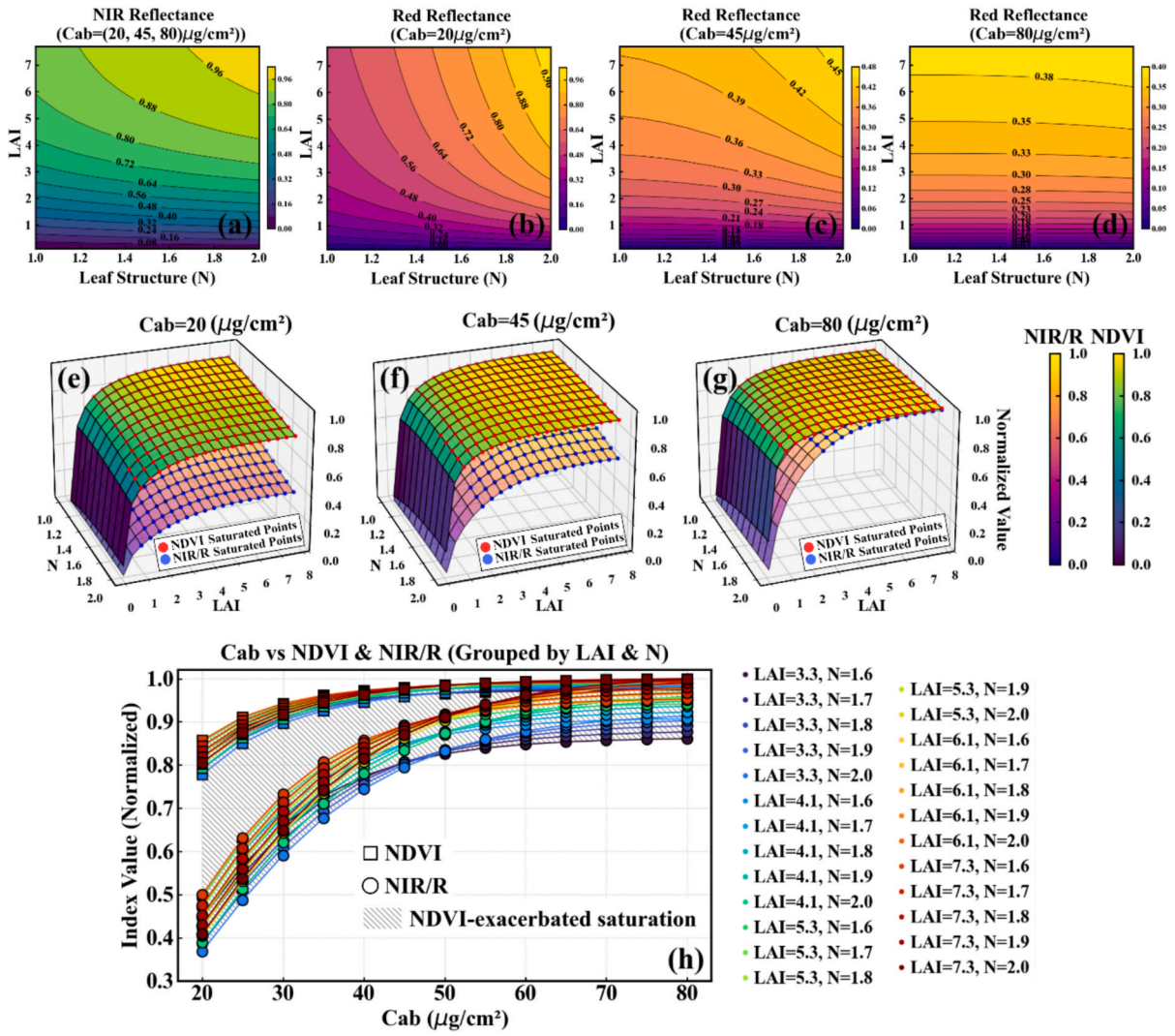

Fig. 2. Two-stage NDVI saturation mechanism: nonlinear formulation amplifies NIR/R saturation.

ences as NIR/R increases

Therefore, redesigning the functional curvature of NDVI—specifically its concavity—with respect to NIR/R is essential for enhancing sensitivity in saturation-prone scenarios. In the 2D graph with NIR/R on the horizontal axis, modifying the concavity of NDVI can yield alternative VI curves that better capture variation in vigorous vegetation. As illustrated in Fig. 3(c), the resulting VI curve can take one of three forms: convex, linear, or concave.

1. Convex VIs

PROSAIL simulations indicate that healthy vegetation typically exhibits R reflectance (3%–10%) approximately 0.67–0.83 times that of the blue reflectance (Schmidt and Skidmore, 2003). Based on this, convex VIs, such as NDVI, EVI, SAVI, can be plotted against the NIR/R axis. As shown in Fig. 3(a), (b), convex VIs exhibit high growth rates in low NIR/R regions. This characteristic enhances sensitivity to low-to-medium vegetation vigor, effectively distinguishing bare soil from

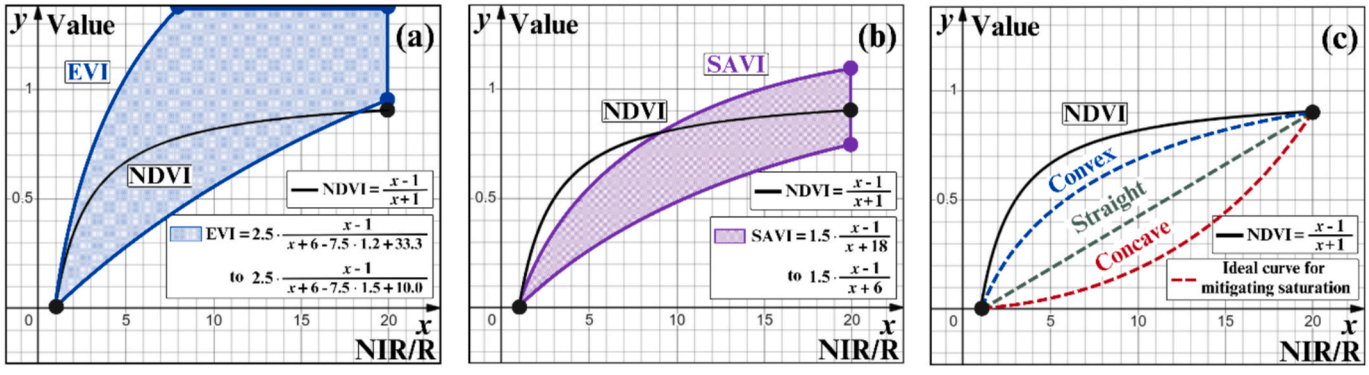

Fig. 3. Convexity and Concavity of VIs in the 2D graph with NIR/R on the horizontal axis.

sparse vegetation. While parameter optimization (EVI: L = 1, C_1 = 6, C_2 = 7.5; SAVI: L = 0.5) can delay saturation, these convex curves ultimately saturate in high NIR/R regions.

2. Linear VIs

Linear VIs define uniform gradients from bare soil to healthy vegetation. While conceptually simple, their constant sensitivity across the entire range of vegetation vigor limits their practical utility. In reality, the relationship between NIR/R and vegetation vigor is highly nonlinear (Xue and Su, 2017) and influenced by multiple interacting factors such as soil properties, moisture availability, and climatic conditions (Yang et al., 2020). As a result, linear indices often fail to capture these complex dynamics, particularly in dense vegetation where more nuanced sensitivity is required (Gnyp et al., 2014).

3. Concave VIs

Concave VIs, as illustrated in Fig. 3(c), exhibit a dual-phase response: they suppress index variation in low NIR/R regions (bare soil and stressed vegetation), while enhancing sensitivity in high NIR/R regions (dense and healthy vegetation) (Kooistra and Clevers, 2016). This asymmetric response improves discrimination among high vigor vegetation while maintaining robustness against background noise in low vigor areas. Consequently, concave formulations offer a theoretically superior approach to mitigating saturation effects.

2.3.3. Saturation mitigated NDVI

Based on the above analysis, the NDVI saturation issue can be addressed by modifying its formulation to construct a concave curve. The new vegetation index should (1) retain NDVI's sensitivity to vegetation vigor and (2) enhance its responsiveness in the high NIR/R domain by altering its curvature.

NDVI can be rewritten as a function of the NIR/R, yielding:

$$\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{R}}{\mathrm{NIR} + \mathrm{R}} \xrightarrow{x=\frac{\mathrm{NIR}}{\mathrm{R}}} \frac{x-1}{x+1} \tag{2-2}$$

As NDVI is convex, achieving concavity requires introducing a concave rate-modification function. Here, we adopt the exponential function for its concave and increasing properties, yielding the following index:

$$\mathrm{VI}_{new} = \frac{x-1}{x+1} \cdot a \cdot b^{x}, x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{2-3}$$

The concavity of this new index can be determined using the second derivative:

$$\mathrm{VI}''_{new} = \frac{4ab^{x}}{(x+1)^2}[\frac{\ln(b)^2(x^2-1)}{4} + \ln b - \frac{1}{x+1}], x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{2-4}$$

When $a > 0$ and $b > e^{1/2}$, the second derivative is constantly greater than 0, and the index is concave over the horizontal range $(1, \infty)$. When $a > 0$ and b decreases below $e^{1/2}$, this concave range narrows slightly. For example, when $a > 0$ and $b = 1.5$, the index will be concave over the horizontal range (1.3033, ∞), which still allows for distinguishing differences at high horizontal axis.

Particularly, when the scaling factor $a = 0.01$, we define the anti-saturation index NDVIsm as:

$$\mathrm{NDVIsm} = 0.01 \cdot \frac{x-1}{x+1} \cdot b^{x}, x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{2-5}$$

Given any NDVI dataset, the parameter b can be determined from its maximum value, which corresponds to the second intersection of NDVI and NDVIsm in Fig. 3(c), leading to the following equation:

$$0.01 \cdot \frac{x_{\max}-1}{x_{\max}+1} \cdot b^{x_{\max}} - \mathrm{NDVI}_{\max} = 0 \tag{2-6}$$

Given the maximum NDVI value ($\mathrm{NDVI}_{\max}$), the corresponding $x_{\max}$ can be derived as:

$$\mathrm{NDVI}_{\max} = \frac{x_{\max}-1}{x_{\max}+1} \rightarrow x_{\max} = \frac{2}{1-\mathrm{NDVI}_{\max}} - 1 \tag{2-7}$$

Accordingly, the parameter b can be determined using the following equation.

$$0.01 \cdot \mathrm{NDVI}_{\max} \cdot b^{x_{\max}} - \mathrm{NDVI}_{\max} = 0 \rightarrow b = 100^{\frac{1}{\left(\frac{2}{1-\mathrm{NDVImax}}-1\right)}} \tag{2-8}$$

Finally, for any NDVI dataset with a known $\mathrm{NDVI}_{\max}$, the anti-saturation NDVIsm can be computed as:

$$\mathrm{NDVIsm} = 0.01 \cdot \mathrm{NDVI} \cdot 100^{\frac{(1+\mathrm{NDVI})\cdot(1-\mathrm{NDVImax})}{(1-\mathrm{NDVI})\cdot(1+\mathrm{NDVImax})}} \tag{2-9}$$

2.4. Comparative saturation mitigated vegetation indices

Comparative analyses are conducted to assess whether NDVIsm mitigates saturation more effectively than widely used VIs. The study compares the distributional patterns of NDVIsm with those of widely used VIs across multiple satellite platforms. Alongside NDVI, we include the EVI, SAVI, Green Normalized Difference Vegetation Index (GNDVI) (Vina and Gitelson, 2011), Normalized Difference Infrared Index (NDII) (Joiner et al., 2018), Normalized Difference Red Edge Index (NDRE) (Xie and Yang, 2020) for comparison with NDVIsm. Definitions and formulations of these indices are summarized in Table 2.

Other approaches to mitigating saturation—such as multi-source data fusion and ML—are not included in these comparative analyses, as they primarily enhance vegetation assessment through additional data rather than structural improvements to NDVI itself (Mutanga et al., 2023). This study focuses on optimizing the mathematical formulation of NDVI. Notably, such structural enhancements to NDVI can also benefit data fusion and ML methods by providing a more reliable and sensitive input VI. The potential improvement in vegetation assessment offered by the proposed anti-saturation index, when combined with non-optical data sources and ML, will be further evaluated in the Results section.

2.5. Evaluation metrics

The change in functional concavity has led to a distinct distribution pattern for NDVIsm compared to other VIs. To quantitatively assess whether this revised distribution mitigates saturation effects, Table 3 presents the evaluation metrics. These metrics jointly characterize the distributional pattern of NDVIsm and its effectiveness in saturation mitigation.

Table 2
Comparative saturation mitigated VIs.

Index	Bands	Formula	Anti-saturation Performance
EVI	NIR, R, B	$2.5 \times \frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+6\mathrm{R}-7.5\mathrm{B}+1}$	Delays saturation, stays responsive up to LAI = 3 (Hinojo-Hinojo and Goulden, 2020)
SAVI	NIR, R	$(1+0.5) \times \frac{\mathrm{NIR}-\mathrm{R}}{\mathrm{NIR}+\mathrm{R}+0.5}$	Reduces soil noise, tunable for high LAI (Zhen et al., 2021)
GNDVI	NIR, G	$\frac{\mathrm{NIR}-\mathrm{G}}{\mathrm{NIR}+\mathrm{G}}$	Less saturation, better in high chlorophyll (Petrović et al., 2025)
NDII	NIR, SWIR	$\frac{\mathrm{NIR}-\mathrm{SWIR}}{\mathrm{NIR}+\mathrm{SWIR}}$	Effective in dense, moist canopies (Zhou et al., 2022)
NDRE	NIR, RE	$\frac{\mathrm{NIR}-\mathrm{RE}}{\mathrm{NIR}+\mathrm{RE}}$	Remains sensitive at higher LAI or nutrient levels (Rehman et al., 2022)

*Blue (B), Green (G), Red (R), Red Edge (RE), Near-Infrared (NIR), Short-Wave Infrared (SWIR).

Table 3
Metrics for evaluating VI distribution pattern, saturation, and model performance.

Metrics	Formula	Interpretation
Coefficient of Variation (CV)	$\frac{\delta}{\mu}=\frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i-\frac{1}{n}\sum_{i=1}^{n}x_i\right)^2}}{\frac{1}{n}\sum_{i=1}^{n}x_i}$	Dispersion of VI values. (Bendel et al., 1989)
Skewness(Skew)	$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\delta}\right)^3$	Asymmetry of VI values. (Bendel et al., 1989)
Saturation Range (SR)	Saturated if $Q_{1.0}-Q_{0.2}\le 0.2\cdot(\max(y)-\min(y))$	Upper 80 % of values span ≤ 20 % of total range.(Defined Firstly)
Coefficient of Determination (R^2)	$1-\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}$	Model-explained variance. (Chicco et al., 2021)
Root Mean Square Error (RMSE)	$\sqrt{\frac{\sum_{i=1}^{n}(y_i-\hat{y}_i)^2}{n}}$	Mean prediction error. (Chicco et al., 2021)
Correlation Coefficient (r)	$r=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2(y_i-\bar{y})^2}}$	Linear correlation between variables. (Benesty et al., 2009)

* n is the number of samples, x_i and y_i are the i th observed values of variables of x and y respectively; $\bar{x}_i$ and $\bar{y}_i$ are their respective means; $\hat{y}_i$ is the predicted value of y_i.

3. Results

This section evaluated the effectiveness of the proposed NDVIsm for mitigating saturation in healthy and dense vegetation across different remote sensing platforms. For satellite platforms, the focus is on comparing the distributional pattern and saturated range of NDVIsm and NDVI across various land cover types. On the UAV platform, the PROSAIL and LUT approach are employed to validate the improved sensitivity of NDVIsm to canopy structure (LAI, N) and vegetation vigor (Cab). Additionally, NDVIsm is incorporated as an input feature in yield prediction models, demonstrating its practical potential when combined with non-optical data in ML frameworks.

3.1. Mitigating NDVI saturation on satellite platforms

Sentinel-2, Landsat-8, and MODIS-09GA are widely used for terrestrial observations, including environmental monitoring, agricultural assessment, and forest studies. In this study, rainforests, farmlands, and grasslands were selected from these satellite images to test the effectiveness of NDVIsm in mitigating saturation in healthy and dense vegetation, exploring its applicability across various land types and observational platforms.

3.1.1. Mitigating NDVI saturation on Sentinel-2 imagery

The Amazon River Basin is characterized by its lush rainforests cover, where NDVI values in the satellite images of this region consistently

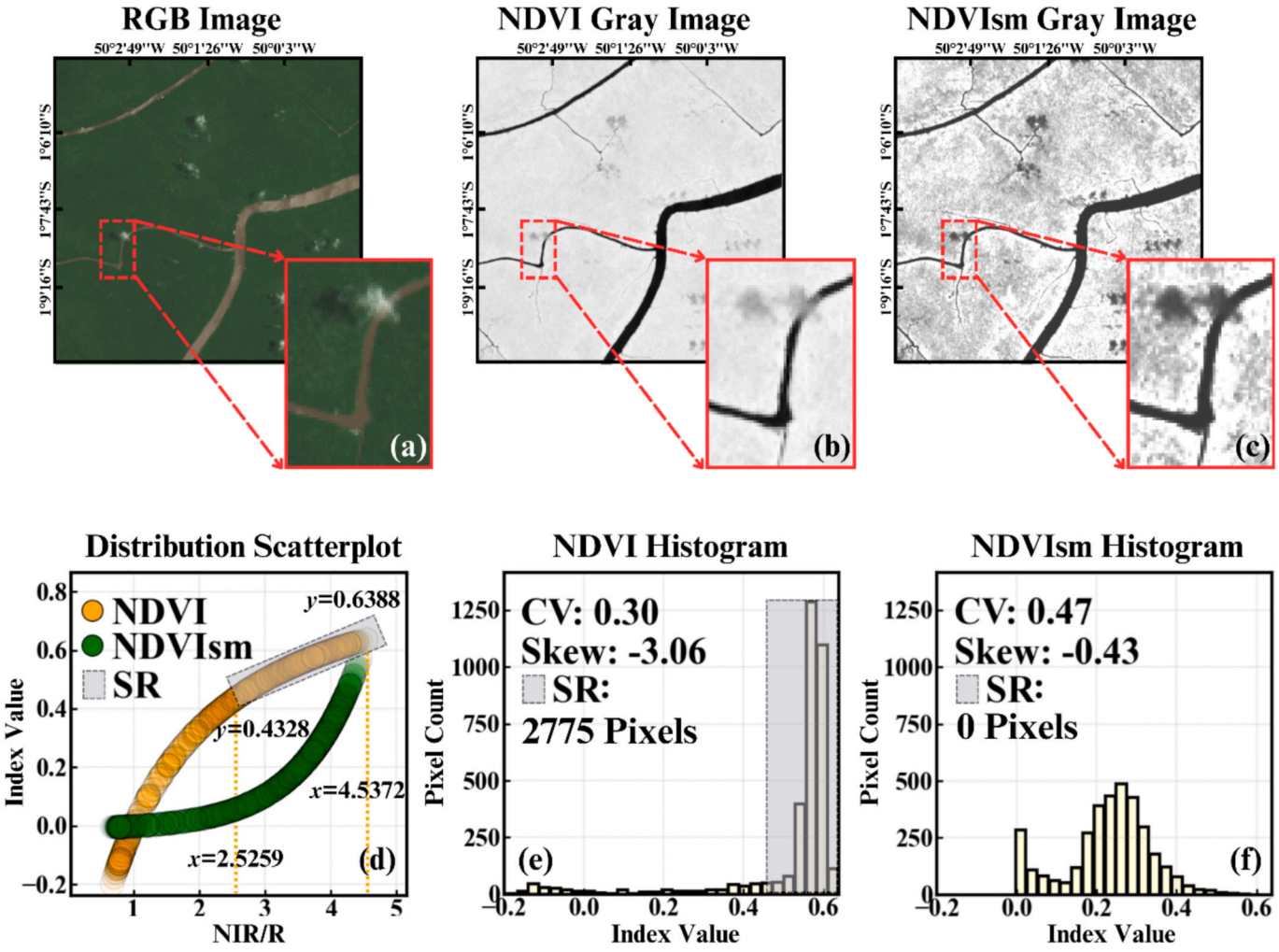

Fig. 4. Comparison of NDVI and NDVIsm in the rainforests of Amazon River Basin.

show high levels. This uniformity in NDVI makes it challenging to discern differences in the trees. Fig. 4(a) presented a typical satellite image of the Amazon rainforests, captured on 2023-09-20. The image has dimensions of 400 by 400 pixels with a spatial resolution of 20 m. The NDVI and NDVIsm for this image are calculated and displayed as grayscale images in Fig. 4(b) and Fig. 4(c). To enable visual comparison, the grayscale images were brightness-normalized by aligning the median of NDVIsm with that of NDVI.

Taking the red areas in Fig. 4(b) and Fig. 4(c) as example, the NDVIsm image shows more pronounced differences in pixel grayscale levels compared to the NDVI image. This is evident from the distribution histograms of the two VIs, as shown in Fig. 4(e) and Fig. 4(f). Most of the pixels in the NDVI image are concentrated with values greater than 0.5, whereas the NDVIsm values are more widely distributed. This is supported by the CV and Skew of all pixel values, which are 0.47 and −0.43 for NDVIsm, respectively—both higher than those for NDVI, at 0.30 and −3.06. As illustrated in Fig. 4(d), within the NDVI image, over 80 % of the pixel values fall within a SR (from 0.4328 to 0.6388), accounting for less 20 % of the total value span—indicating that 2775 pixels are saturated. In contrast, no such saturation is observed in NDVIsm.

The Central Valley of California, U.S., is a fertile plain renowned for its agricultural diversity. The area is known for cultivating a variety of crops, including fruits, vegetables, and nuts like almonds and walnuts. Fig. 5(a) showed a typical satellite remote sensing image of farmlands in the region, captured on 2023-04-14, with a size of 400 by 400 pixels and a spatial resolution of 20 m. The NDVI and NDVIsm indices for this image were calculated and displayed as grayscale images in Fig. 5(b) and Fig. 5(c).

Taking the red areas in Fig. 5(b) and Fig. 5(c) as example, the NDVIsm image exhibits more apparent differences in pixel grayscale levels. As shown in the histograms in Fig. 5(e) and Fig. 5(f), most of the NDVI values are concentrated within the 0.6–0.7 range, while NDVIsm values are more broadly distributed. This distribution pattern is evaluated by the CV and Skew: NDVIsm has a CV of 0.53 and a skewness of 0.10, both higher than those of NDVI, at 0.11 and –2.67, respectively. As shown in Fig. 5(d), more than 80 % of the pixel values in the NDVI image are confined to a SR (0.5867 to 0.7134), which covers less 20 % of the entire value span—suggesting that 2832 pixels are affected by saturation. By contrast, such saturation is not observed in the NDVIsm image.

Northern Minnesota's land cover is characterized by rolling hills, numerous lakes, and extensive forests. The dominant vegetation includes coniferous trees like pine, fir, and spruce, as well as deciduous trees such as oak and maple. Fig. 6(a) was such a typical satellite remote sensing image of wildlands, captured on 2023-05-26. This image, with a size of 400 by 400 pixels and a spatial resolution of 20 m, is processed to calculate the NDVI and NDVIsm, which are shown as grayscale images in Fig. 6(b) and Fig. 6(c).

Taking the red areas in Fig. 6(b) and Fig. 6(c) as example, more obvious differences in gray levels are similarly found in NDVIsm. In Fig. 6(e), most of the NDVI values clustered around 0.6, whereas in Fig. 6(f), the distribution of NDVIsm is more spread out. The CV for NDVIsm is 0.38, significantly higher than that of NDVI at 0.07. The Skew of the NDVIsm is 0.53, compared to NDVI at −2.64. In Fig. 6(d), it also can be seen that over 80 % of the pixel values in the NDVI image are clustered

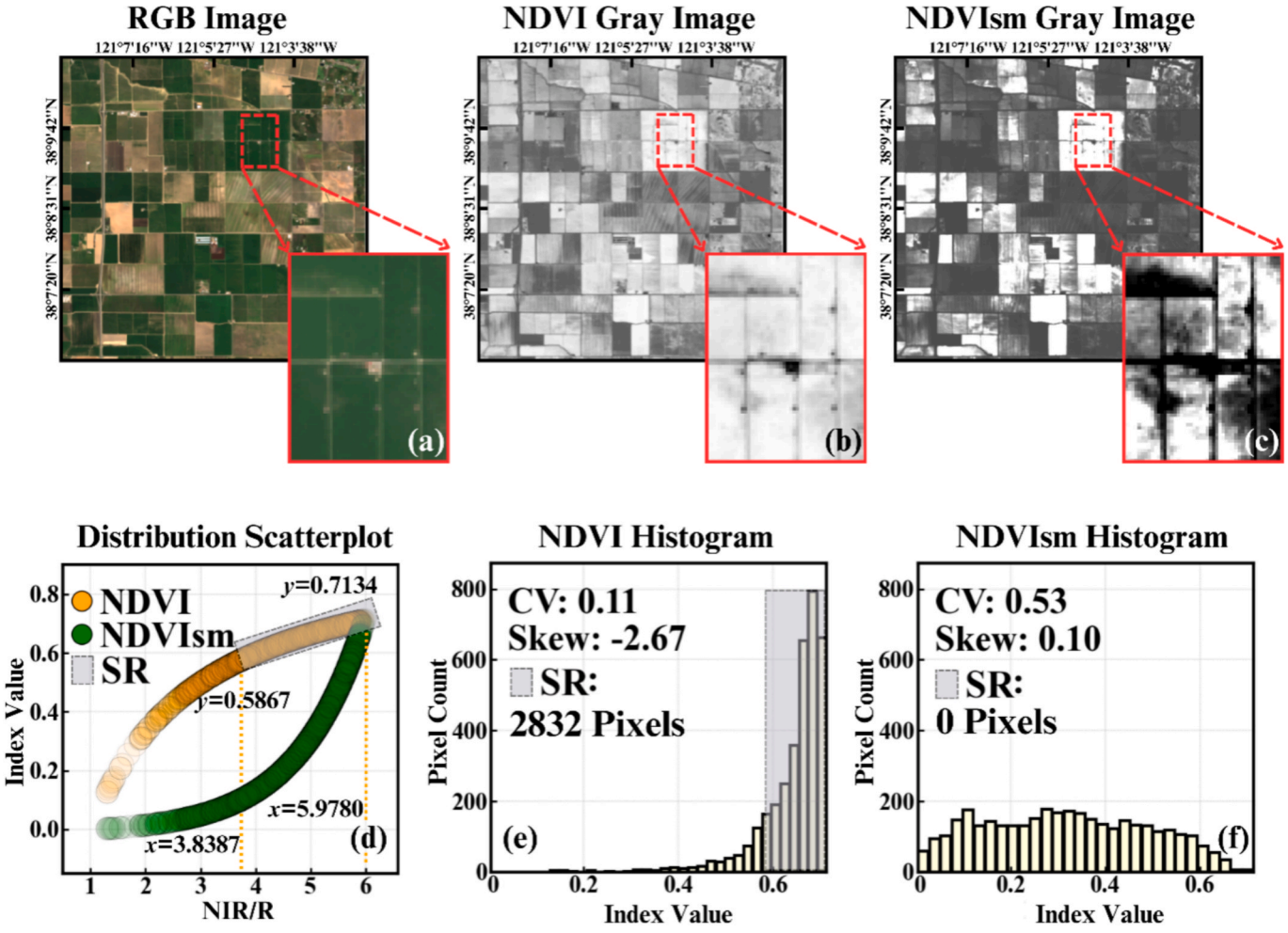

Fig. 5. Comparison of NDVI and NDVIsm in the farmlands of Central Valley of California.

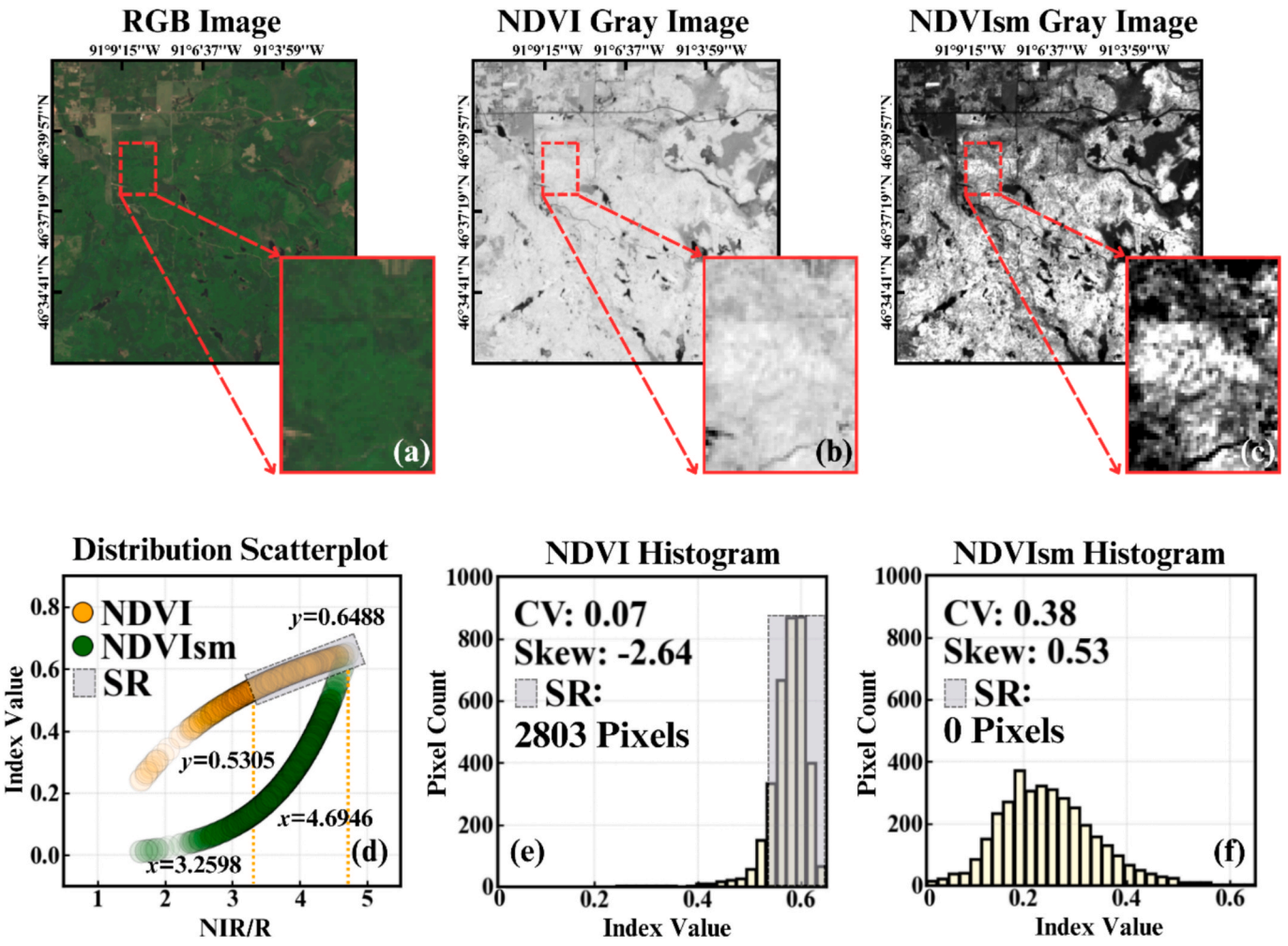

Fig. 6. Comparison of NDVI and NDVIsm in the grasslands of Northern Minnesota.

within a SR (0.5305 to 0.6488), which accounts for less 20 % of the full value range—indicating saturation in 2803 pixels. In comparison, the NDVIsm image exhibits no such saturation.

According to the results from Sentinel-2 imagery for rainforests, farmlands, and wildlands, it is found that compared to NDVI, NDVIsm exhibits a higher CV, indicating a more dispersed data distribution. This dispersion effectively stretches the concentrations of high values found in NDVI, making NDVIsm more adept at differentiating of healthy and dense vegetation. Moreover, NDVIsm demonstrates a greater Skew, indicating a more apparent right-tailed distribution. High Skew suggests that pixels with similar and high index values are effectively separated, significantly mitigating the saturation often encountered in NDVI. In particular, for regions like the Amazon rainforests, negative NDVI values and low positive NDVI values are compressed into a very small range in NDVIsm. These negative and lower positive values typically indicate non-vegetation or unhealthy vegetation. The ability of NDVIsm to compress these values significantly aids in reducing background noise dominated by non-vegetation and unhealthy vegetation. NDVIsm can compress the index values of water, soil, and unhealthy vegetation at lower levels, and widen the disparity in index values among healthy and dense vegetation. In summary, in Sentinel-2 imagery, the improved NDVIsm provides a powerful tool for detecting variation within healthy and dense vegetation, thereby helping to address saturation.

3.1.2. Comparisons of mitigating NDVI saturation on three satellite platforms

In addition to Sentinel-2 imagery, data from Landsat-8 and MODIS-09GA were also utilized to evaluate the effectiveness of NDVIsm in mitigating saturation. In regions of healthy and dense vegetation—such as rainforests, farmlands, and grasslands—the distribution patterns of NDVI and NDVIsm from all three satellite platforms were still evaluated using the CV, Skew and SR, as summarized in Table 4.

As demonstrated in Table 4, NDVI suffers from saturation in all cases, with over 80 % of pixel values are compressed into less 20 % of the total value range. This clustering of values indicates limited sensitivity in dense vegetation areas. In contrast, NDVIsm avoids such saturation by redistributing pixel values across a broader range. Specifically, NDVIsm consistently exhibits higher CV and Skew, indicating a more dispersed and right-skewed distribution. This adjustment not only expands the variation among high-index pixels but also compresses low-index regions that typically associated with non-vegetated surfaces. These distributional improvements make NDVIsm a robust tool for identifying subtle variations in dense and healthy vegetation. Its performance has been consistently validated across Sentinel-2, Landsat-8, and MODIS-09GA imagery, despite differences in spatial–temporal resolution and spectral center wavelengths. This cross-platform stability can be further explained by the two-stage saturation mechanism, in which optical saturation is affected by sensor-specific factors, whereas NDVIsm mitigates the mathematically induced saturation through a reformulated structure that remains independent of these imaging differences. This robustness of NDVIsm highlights its broad applicability in mitigating NDVI saturation across multi-source remote sensing analysis.

Table 4
Comparison of NDVI and NDVIsm distribution metrics across satellite imagery.

Model	Site Info.	VI	Bands (nm)	Resolutions	CV	Skew	SR
Sentinel-2	RainForests	NDVI	665 ± 30	20 m	0.30	−3.06	0.43–0.64
	(1°8′43″S, 50°2′56″W)	NDVIsm	(R)	16 days	0.47	−0.42	–
	Farmlands	NDVI	842 ± 115 (NIR)		0.11	−2.67	0.59–0.71
	(38°9′31″N, 121°4′27″W)	NDVIsm			0.53	0.10	–
	Grasslands	NDVI			0.07	−2.64	0.53–0.65
	(46°38′19″N, 91°8′37″W)	NDVIsm			0.38	0.53	–
Landsat-8	RainForests	NDVI	655 ± 30	30 m	0.14	−4.45	0.31–0.53
	(0°4′34″N, 64°12′16″W)	NDVIsm	(R)	5 days	0.31	−0.63	–
	Farmlands	NDVI	865 ± 20 (NIR)		0.20	−1.87	0.43–0.57
	(40°3′27″N, 103°3′51″W)	NDVIsm			0.51	−0.72	–
	Grasslands	NDVI			0.32	−2.26	0.35–0.56
	(45°0′16″N, 88°1′17″W)	NDVIsm			0.60	0.12	–
Modis-09GA	RainForests	NDVI	645 ± 50	500 mDaily	0.13	−4.82	0.71–0.91
	(3°37′23″S, 54°14′25″W)	NDVIsm	(R)		0.50	0.01	–
	North Florida	NDVI	856 ± 35 (NIR)		0.08	−4.49	0.67–0.86
	(35°4′30″N, 93°51′27″W)	NDVIsm			0.59	1.86	–
	Great Lakes	NDVI			0.14	−3.91	0.69–0.92
	(39°37′19″N, 92°42′33″W)	NDVIsm			0.92	3.17	–

3.2. Mitigating NDVI saturation on UAV platforms

NDVIsm exhibits a more dispersed distribution than NDVI, as indicated by its consistently higher CV and Skew. Although saturation is no longer observed in NDVIsm images, the precise impact of NDVIsm on mitigating saturation requires further investigation. A compelling way to evaluate its effectiveness is to examine whether NDVIsm can significantly enhance correlations with vegetation biophysical parameters, such as LAI, Cab and N, which are proved to exhibit saturation when using NDVI in dense canopies. We further assess whether NDVIsm maintains its stability and can reveal biophysical responses more effectively in saturation-prone conditions. In addition, this section explores the potential of NDVIsm to improve vegetation assessments, such as yield prediction, thereby demonstrating its broader applicability in dense vegetation monitoring.

3.2.1. Distributional differences between NDVIsm and NDVI

Fig. 7(a) displays a UAV remote sensing image of a maize field on Arlington Agricultural Research Station, Madison, WI, U.S. The image was captured by a Headwall Nano Hyperspectrometer mounted on a DJI M600 Pro drone at 60 m altitude on 08/23/2022, with a ground sampling distance of 3.50 cm.

Consistent with the satellite platforms, similar conclusions can be drawn from the UAV platform. In Fig. 7(b), NDVI displays extensive saturated regions due to the high and concentrated index values for healthy and dense maize. However, NDVIsm displays obvious differences in gray scales, making it easier to visually distinguish areas with unhealthy and non-vegetation that represented by dark pixels. The reason is still that NDVI and NDVIsm have different distributional characteristics. Different CV and Skew shown on histograms Fig. 7(e) and Fig. 7(f), the values of NDVI are clustered and mostly above 0.8, while the values of NDVIsm are more dispersed and accompanied by a clear right-tailed distribution. The saturation issue was not observed in NDVIsm.

3.2.2. Correlation improvement with LAI, Cab and N

For the area highlighted in Fig. 7, pixels with NIR reflectance < 0.3 were excluded to remove most soil backgrounds. The PROSAIL was parameterized with LAI (0.1–8.0), Cab (20–80 $\mu g/cm^2$), N (1.0–2.0), carotenoids (2–20 $\mu g/cm^2$), brown pigment (0–0.3), leaf water thickness (0.03 g/cm^2), dry matter (0.005–0.012 g/cm^2), hot spot (0.01–0.5), solar zenith angle (20–35°), view zenith angle (0°), relative azimuth angle (45°), soil reflectance (0.1–0.3), soil moisture (0–0.2), and leaf inclination parameters $A = 50$ and $B = 0$. A total of 100 k spectrum were generated using Latin Hypercube Sampling (Loh, 1996). For each pixel, the most similar simulated spectrum was identified by minimizing RMSE with the observed hyperspectral spectrum (400–1000 nm, 274 bands), and the corresponding LAI, Cab, and N values were retrieved. This process yielded 31,208 vegetation pixels, each associated with NDVI, NDVIsm, and retrieved LAI, Cab, and N values. Fig. 8 shows their relationships.

From the perspective of canopy structure (LAI and N) and vegetation vigor (Cab), NDVI remains affected by saturation in such dense and healthy vegetation. This is reflected by the limited variation in NDVI values, which tend to remain around 0.8 regardless of changes in LAI, Cab, or N. As a result, saturation means NDVI fails to capture subtle differences in vegetation and cause inaccurate assessment. In contrast, NDVIsm demonstrates improved sensitivity to increases in LAI, Cab and N, with correlation r of 0.9280, 0.8083 and 0.3242 respectively—both markedly higher than those of NDVI (0.6720, 0.2508 and 0.0790). This result demonstrates the physical significance of optimizing the concavity of the VI's curve, which enhances sensitivity to canopy structure and vegetation vigor—an effect that is typically obscured by NDVI saturation. As a result, NDVIsm proves more effective in distinguishing subtle variations in dense and healthy vegetation.

3.2.3. Improvement in yield prediction using NDVIsm

Reducing saturation is merely a means; whether the improved VI can play a significant role in vegetation assessment—such as yield prediction—is a key question that deserves attention. The effectiveness of the anti-saturation index NDVIsm for yield prediction under dense vegetation conditions was evaluated from three perspectives:

1. Enhanced correlation between NDVIsm and maize phenotypic traits

Key field traits were collected in this study, including grain yield, plant height, ear height, anthesis and silking dates, stand count, and grain moisture. Yield was determined by harvesting, drying, and weighing ears (kg/ha). Plant and ear heights were measured from the soil to the top of the plant and the ear node, respectively. Flowering time was recorded as days to 50 % anthesis or silking. Stand count referred to the number of plants per plot, while moisture was estimated by oven-drying samples at 60 °C for seven days.

2. Improved feature importance of NDVIsm in random forest models

To quantify the contribution of NDVIsm to yield prediction, a random forest (RF) (Breiman, 2001) model was trained with ten repeated runs. Each model was tested with an increasing number of decision trees from 100 to 200, no depth limit, a minimum of five samples per split and per leaf node, up to six features per split, and bootstrap sampling. Feature importance was quantified based on the average reduction in Gini impurity (Disha and Waheed, 2022) across 10 repeated training runs and visualized using histograms.

3. Yield prediction improvement using NDVIsm with random forest

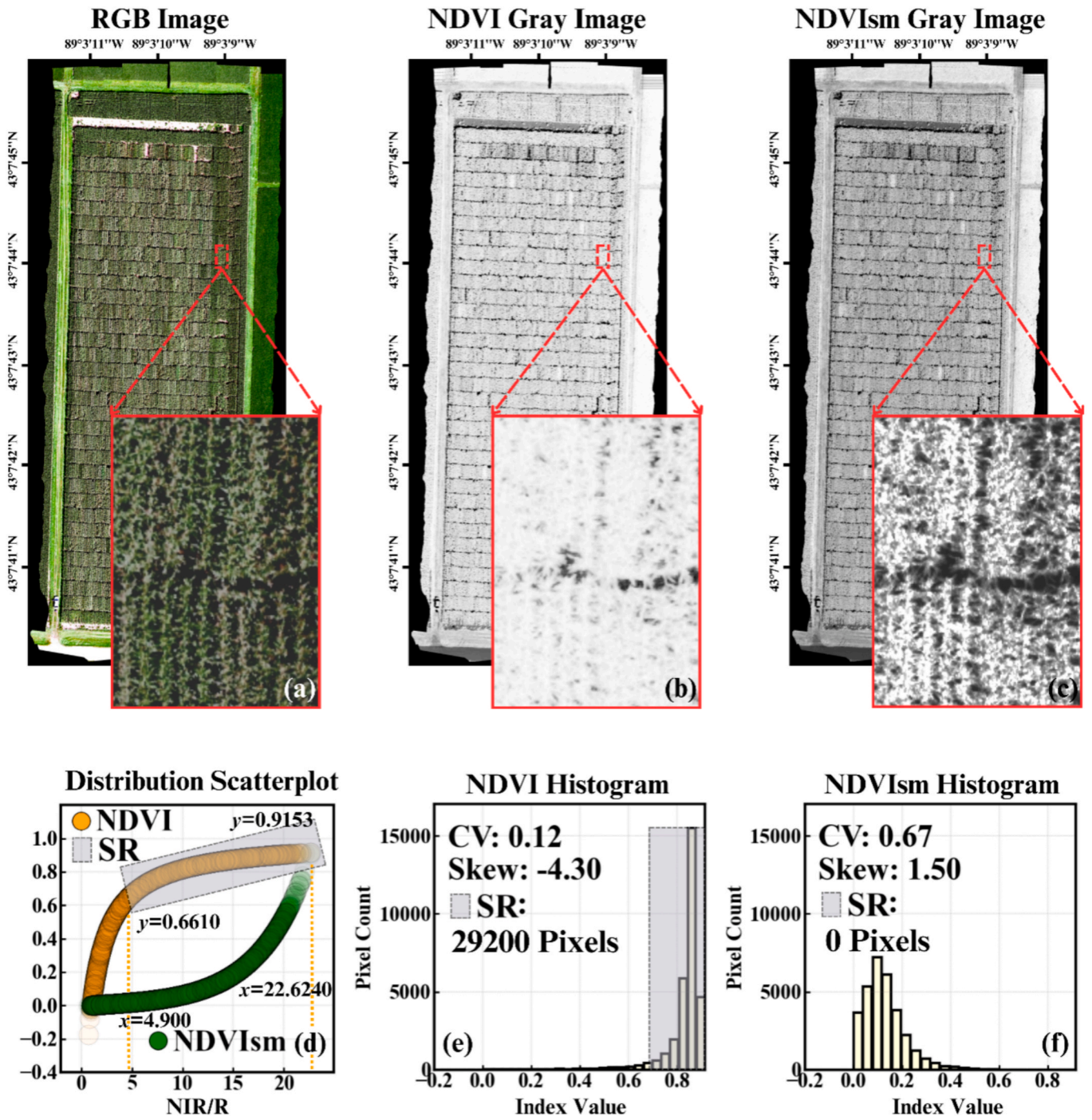

Fig. 7. Comparison of NDVI and NDVIsm in the maize field of Arlington Agricultural Station.

models

The RF models incorporated one VI (NDVI or NDVIsm) along with three common features: plant height, green coverage, and textural metrics. Plant height was derived from LiDAR-based crop height models (Zhang et al., 2016); green coverage was calculated from RGB images in HSV color space (Yang et al., 2015); and textural features (contrast, homogeneity, entropy) were extracted using GLCM methods (Haralick et al., 1973). To minimize variability, only genotypes with $\geq$ 3 plot replicates were retained from the original 1200 plots. After averaging across replicates, 284 representative samples were obtained, with 54 used for testing and 230 for training. Yield prediction accuracy was evaluated using R^2 and RMSE, as illustrated in Fig. 9.

As shown in Fig. 9, the NDVIsm heat map Fig. 9(b) exhibits generally darker red tones than NDVI Fig. 9(a), indicating stronger correlations with phenotypic traits. Among the 35 values, 32 increased under NDVIsm. On average, NDVIsm improves correlation by 2.55 % for grain yield, 5.02 % for plant height, 9.37 % for ear height, 3.98 % for anthesis, 19.40 % for silking, and 26.19 % for moisture. Stand count increased by 0.0864, likely due to enhanced texture visibility in grayscale images, see Fig. 7(c). Percentage improvement in Stand Count was not reported due to the small baseline values, which can lead to exaggerated percentage changes.

Fig. 9(c) and Fig. 9(d) present side-by-side histograms of feature importance derived from ten runs of RF training. NDVIsm achieve a higher average importance (0.4147) compared to NDVI (0.3709), indicating its stronger contribution to yield prediction. Considering that the

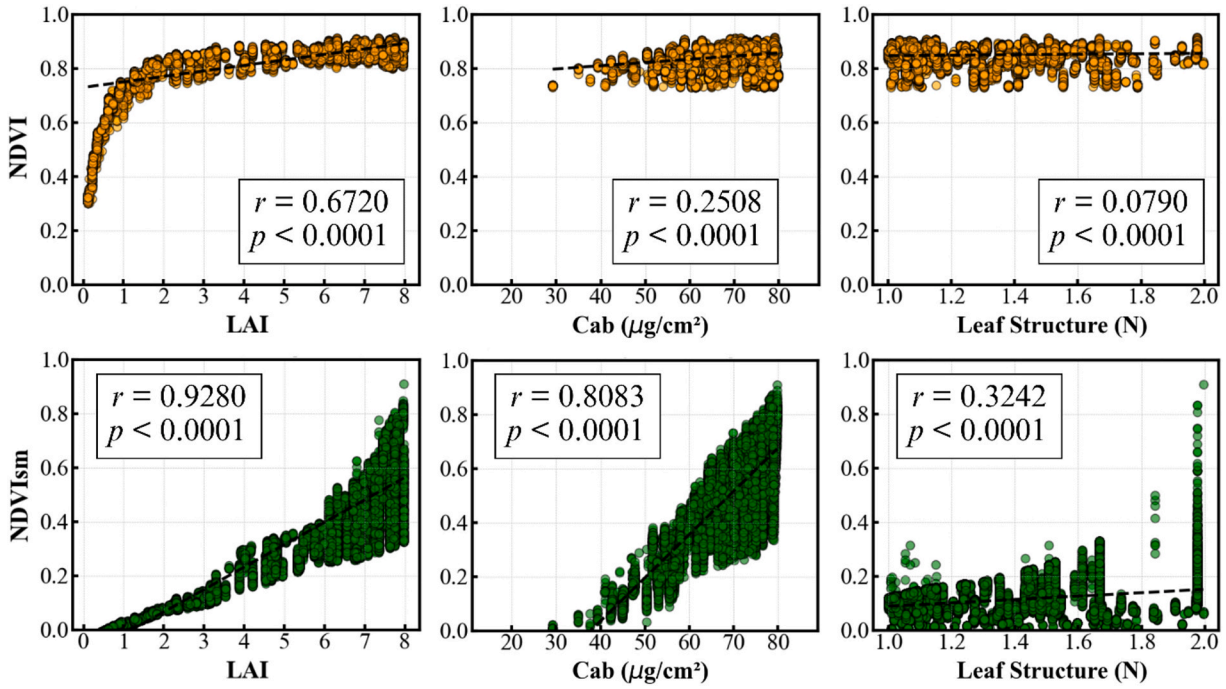

Fig. 8. The relationship between NDVI/NDVIsm with LAI, Cab and N.

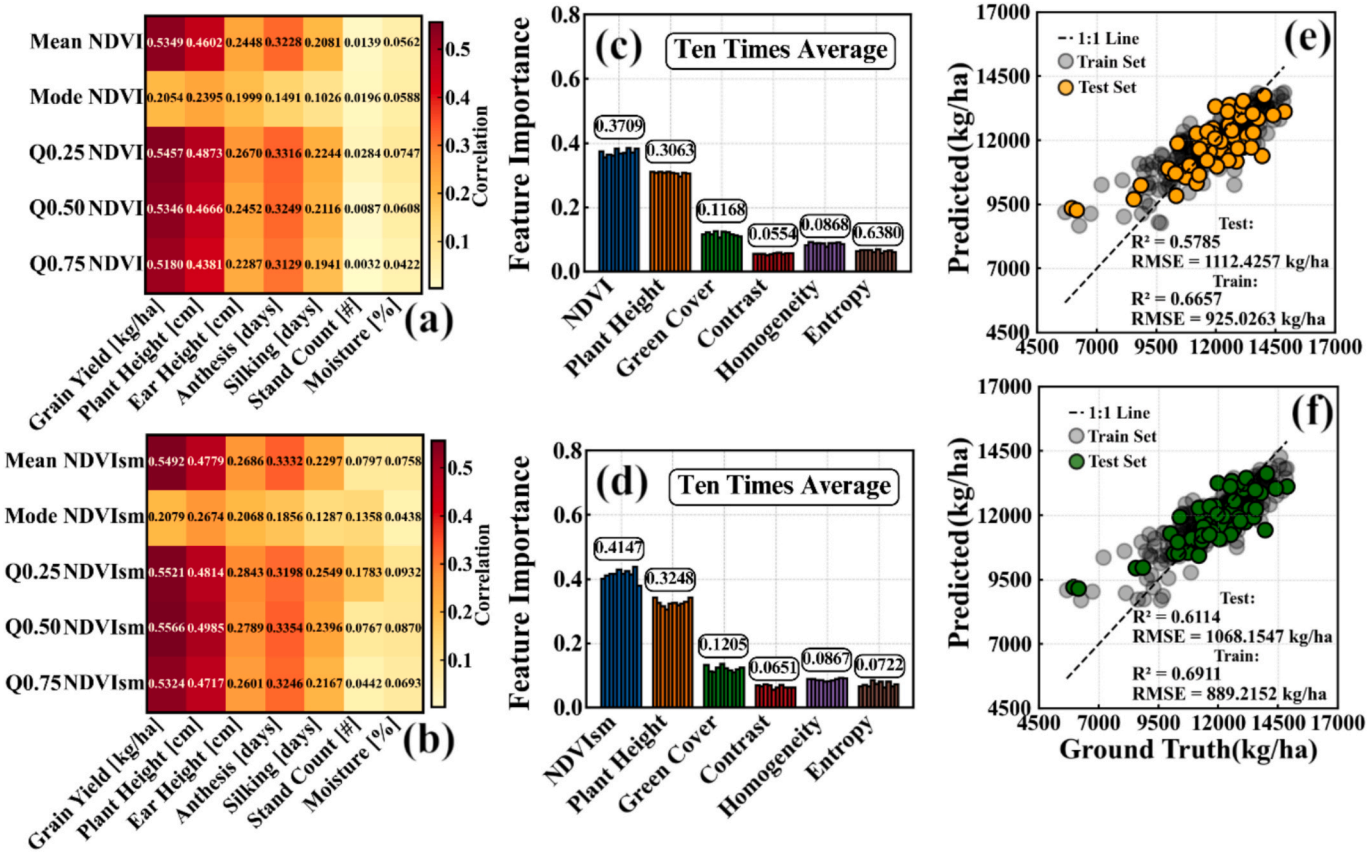

Fig. 9. Results of improved performance in maize yield prediction using NDVIsm: Correlation between phenotypic traits and NDVI (a) or NDVIsm (b). Feature importance in RF models using NDVI (c) or NDVIsm (d). Yield prediction performance using NDVI (e) or NDVIsm (f).

number of decision trees is a key factor influencing RF model performance (Biau and Scornet, 2016), ten tests were conducted with varying tree counts (increased by 10 per test). The results showed that in nine out of ten runs, NDVIsm consistently exhibits greater importance than NDVI, and also surpassed plant height (non-optical data), green coverage, and textural features (spatial data). These findings suggest

that NDVIsm is more reliably recognized by the model as a valuable predictor and shows greater potential for accurate and stable yield estimation.

Fig. 9(e) and Fig. 9(f) are the R^2 and RMSE scores of yield prediction models. The RF model using NDVIsm achieves the highest R^2 (0.6911) and lowest RMSE (889.2152 kg/ha) in training set, outperforming that of NDVI. While in test set, NDVIsm still has higher R^2 (0.6144) and lower RMSE (1068.1547 kg/ha) than NDVI. Although the improvements in R2 and RMSE are moderate, this is expected given that yield prediction is a complex process influenced by various factors such as soil properties, water availability, and genetic traits. Nevertheless, Fig. 9 still demonstrates NDVIsm's consistent advantage throughout the entire yield modeling process. These findings support the use of NDVIsm as a more effective index than NDVI in dense biomass environments.

4. Discussion

4.1. Comparisons of different vegetation indices

With advancements in sensor technologies and analytical methods, many new VIs have been developed to address the saturation issues of NDVI. In this study, widely used VIs—EVI, SAVI, GNDVI, NDII, and NDRE—were compared against the newly proposed NDVIsm, using Sentinel-2 imagery across three dense vegetation land cover types: rainforests, farmlands, and grasslands. Fig. 10 presents the grayscale images and histograms of these VIs respectively.

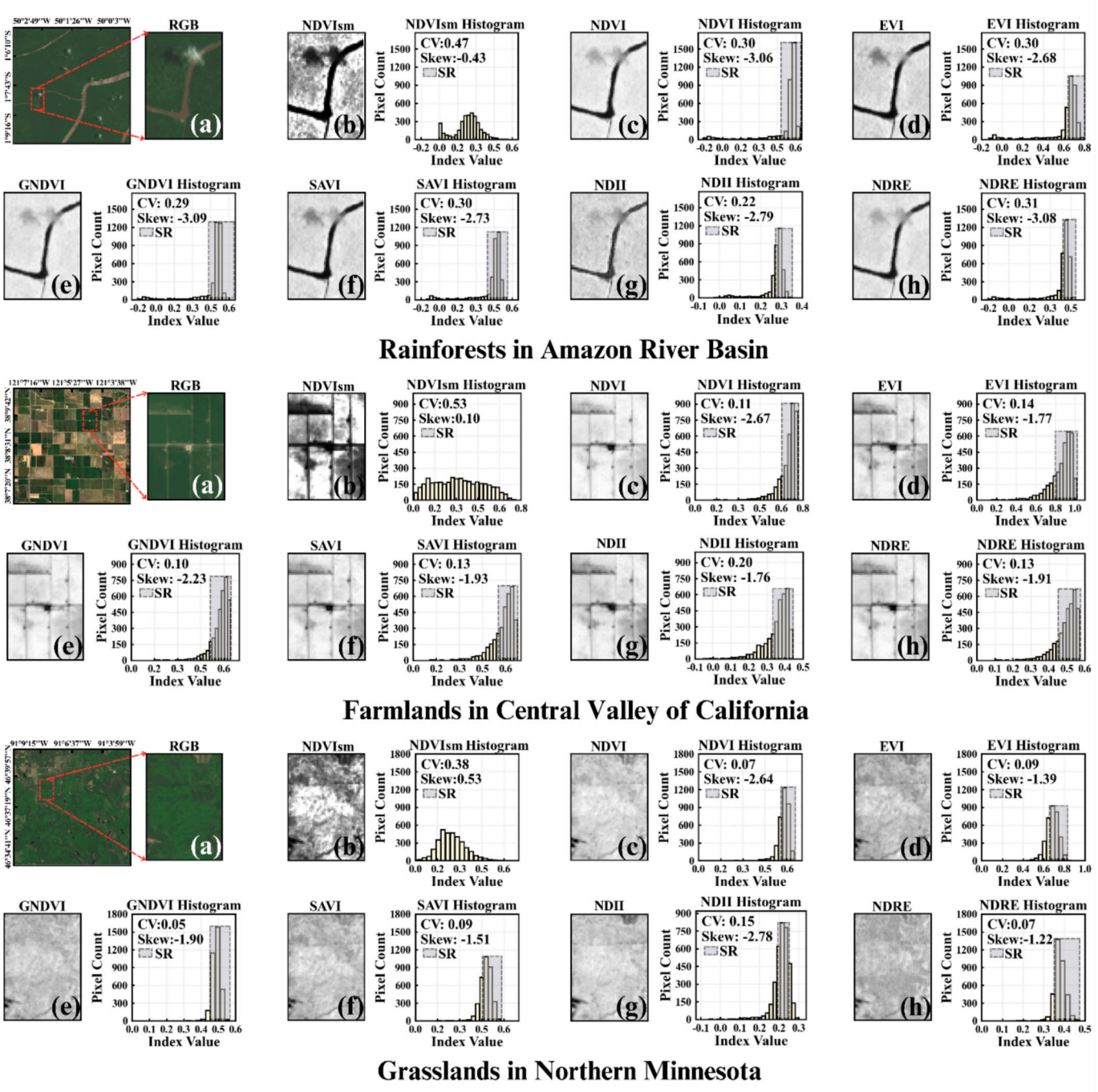

Fig. 10. Comparisons between different VIs in rainforests, farmland and grasslands.

In all three cases, NDVIsm consistently exhibits the most pronounced grayscale variation and the most dispersed histograms, with no observable SR in its histograms. In contrast, the histograms of other VIs remain clustered and exhibit clear SR, despite incorporating additional wavebands or more complex formulations. Specifically, in the rainforests, NDVIsm shows the highest CV at 0.47 and most pronounced right-skewed distribution (Skew = -0.43), whereas other VIs remain concentrated with CVs around 0.30. In farmlands, NDVIsm still maintains the highest CV at 0.53 and shows a positive Skew at 0.10, clearly distinguishing high-index values that other VIs failed to resolve. In grasslands, NDVIsm again achieves the highest CV (0.38) and Skew (0.53), standing out as the only VI with a distinct right-tailed distribution.

The analysis highlights that traditional VIs—though improved in spectral or formula complexity—do not significantly widen the data distribution, as reflected by limited improvement in CV. While they may slightly improve the Skew of the data distribution, the overall trend still shows a pronounced left-skewed histogram and gets in saturation. This suggests that these VIs, much like NDVI, continue to be affected by saturation. In contrast, NDVIsm consistently achieves broader dispersion and right-skewed histograms across diverse vegetation conditions. These distributional characteristics enhance numerical differentiation at higher index values, making NDVIsm particularly effective for mitigating saturation in healthy and dense vegetation.

4.2. Other NDVIsm variants and NDVIsm disadvantages

In the 2D graph with NIR/R on the horizontal axis, the exponential function with a rapid growth rate was selected as the rate modification function to reshape the NDVI curve into a concave form and mitigate saturation. Alternatively, a power function can also be employed for this purpose, and it is defined as follows:

$$\mathrm{VI}_{new} = \frac{x-1}{x+1}\cdot a\cdot x^{c}, x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{4-1}$$

The concavity of this new index can be determined using the second derivative:

$$\mathrm{VI}''_{new} = \frac{a(cx^{c+1}(c-1)(x-1)(x+1)^2 + 4cx^{c+2}(x+1) - 4x^{c+3})}{x^3(x+1)^3}, x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{4-2}$$

When $c > 1, x > 1, a > 0$, the numerator of the second order derivative is constantly greater than 0, because:

$$4cx^{c+2}(x+1) - 4x^{c+3} > 4cx^{c+3} - 4x^{c+3} > 0 \tag{4-3}$$

Therefore, under this condition of $c > 1$, $a > 0$, the new index will be concave.

Moreover, if concave functions are not suitable for regions with unhealthy or sparse vegetation, alternative functions can be employed to shape the new indices into convex or approximately linear forms.

New convex and approximately linear VIs are defined as shown in Equations (4–4) and (4–5):

$$\mathrm{VI}_{new} = \frac{x-1}{x+1}\cdot a\cdot \log_{10}(x^{d}), x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{4-4}$$

$$\mathrm{VI}_{new} = \frac{x-1}{x+1}\cdot a\cdot (x-1), x \in (1, \frac{\mathrm{NIR}_n}{\mathrm{R}_n}) \tag{4-5}$$

In the 2D graph with NIR/R on the horizontal axis, various functions can be used to modify the increase rate of NDVI, leading to NDVIsm variants with different distributional properties. As shown in the Fig. 11, some examples are provided, though these do not represent the full variants. By altering parameters and mathematical definitions, additional variants can be generated that may be better suited for other studies involving different vegetation types or canopy densities.

As shown in Fig. 11, logarithmic, linear, power, and exponential functions were applied to modify the rate of NDVI increase, producing multiple NDVIsm variants with mitigated saturation in the imagery of the Amazon River Basin.

It can be observed that as the concavity of the rate modification function increases, pixel-level differences and textural details in the grayscale images become more pronounced. Correspondingly, the distribution of NDVIsm exhibits higher CV and Skew with increasing concavity. Among these, the exponential modification yields the highest CV (0.38) and a Skew of −0.59. As shown in Fig. 12(f), this enhancement makes areas of dense vegetation—indicated by brighter pixels—more visually distinguishable, revealing details that are not apparent in the original NDVI images. These results reinforce that the exponential-function-based NDVIsm variant offers the most effective saturation mitigation.

Although NDVIsm enhances sensitivity to subtle variations in dense vegetation, its exponential nature leads to premature saturation in sparse vegetation areas, concentrating most index values in the lower range. As shown in Fig. 13, a sparse vegetation region was selected from Landsat-8 imagery over the Mississippi River Basin on August 30, 2023.

In this selected sparse vegetation region, the dominant land cover has shifted from dense vegetation to sparse or non-vegetated areas. The higher CV and Skew indicate that low index values are further concentrated, with most NDVIsm pixels falling within the narrow range of 0.0202–0.0319, see Fig. 13(d). Additionally, since NDVIsm is derived from the maximum NDVI, large differences in vegetation vigor cause it to respond more strongly to highest vigor vegetation (upper-right area in Fig. 13(c)) while compressing other vigor vegetation. This results in minimal value distinction between low-vigor and non-vegetated areas.

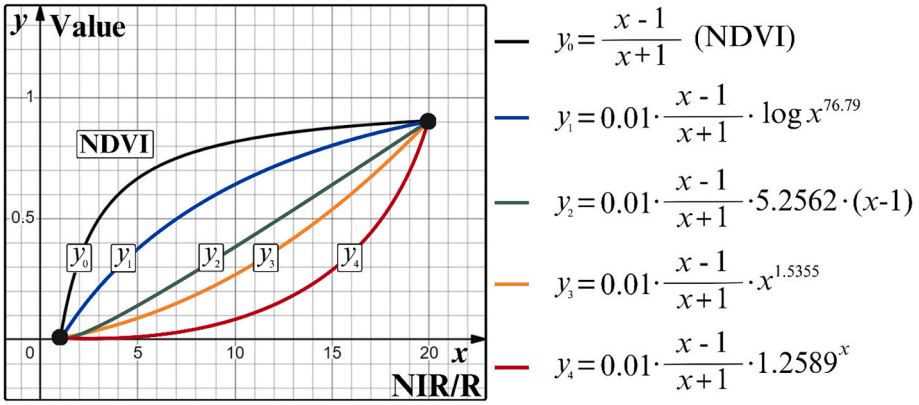

Fig. 11. Some NDVIsm variants in the 2D graph with NIR/R on the horizontal axis.

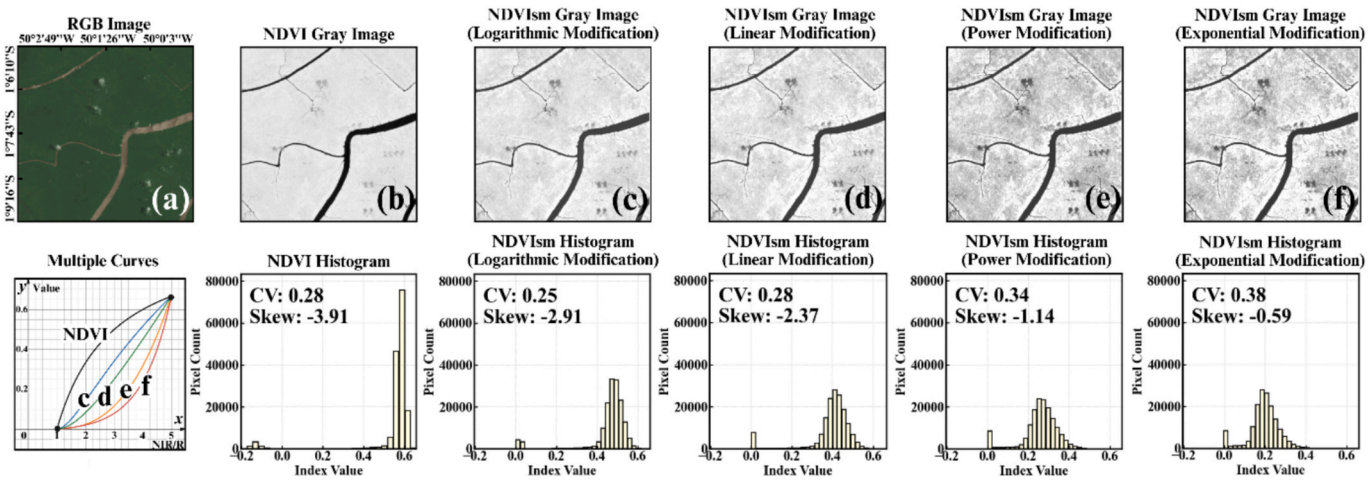

Fig. 12. Comparisons of distributional characteristics of multiple NDVIsm variants.

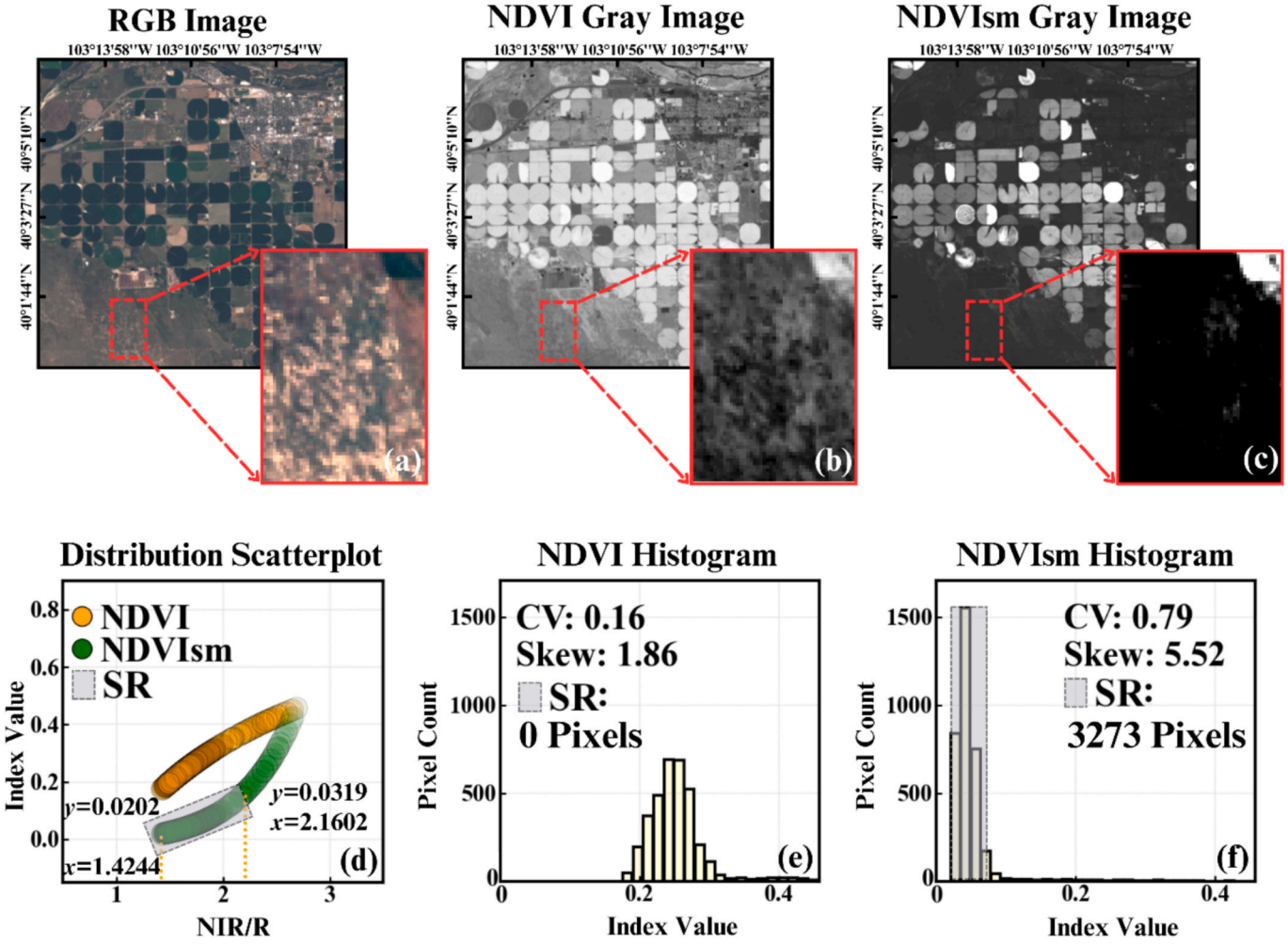

Fig. 13. Limited application of NDVIsm in sparse vegetation area.

Applying NDVIsm in such regions results in forward-saturation. This unintended effect originates from the exponential transformation used in NDVIsm, which, although effective at mitigating saturation in dense vegetation, inadvertently suppresses index variability in sparse vegetation. It is important to note that NDVIsm is specifically developed to address saturation in dense vegetation and has the following limitations:

1.Not suitable for sparse vegetation (typically when LAI $\leq$ 3).

2.Most suitable for uniformly vigorous vegetation (where large-scale saturation occurs).

4.3. NDVIsm reflects county-level maize yield variability

The mechanism by which improvements in CV and Skew help mitigate saturation and enhance yield prediction accuracy still requires further investigation. To explore this, Landsat-8 satellite imagery and county-level maize yield data were used. Yield data for 2022 were obtained from the USDA Quick Stats Survey for Illinois, Iowa, Wisconsin, and Minnesota. Landsat-8 imagery (LANDSAT/LC08/C02/T1_L2) with less than 20 % cloud cover from July 15 to August 15, 2022, was retrieved from Google Earth Engine. CDL masks were applied to isolate maize fields, and the digital numbers were converted to reflectance using a scaling factor of 0.0000275 and an offset of −0.15. Pixels in the NIR with reflectance below 0.3 were masked out, and the same mask was applied to the R. As shown in Fig. 14(a–d), the average NIR/R ratio for each county was calculated, and yield values were normalized relative to the maximum NDVI, determined based on the maximum NIR/R value in the dataset. The resulting scatter plots visualize the relationship between yield and NDVIsm variants, enabling direct comparison of how index formulation affects yield sensitivity.

As shown in Fig. 14(a–d), county-level maize yield increased with the NIR/R ratio, particularly in such dense and uniform farming conditions, exhibiting a clear exponential trend. Among all NDVIsm variants, the exponential form (red line) most accurately captures this relationship, outperforming the linear, logarithmic, and power forms (dashed lines). When these variants are directly applied as regression functions, the exponential NDVIsm consistently has the highest prediction accuracy. Notably, its R^2 values reach 0.4923 in Illinois and 0.4079 in Minnesota, and remain superior to NDVI even in Iowa (0.2644) and Wisconsin (0.1678). Fig. 14(e–h) further illustrates that this exponential variant produces the lowest absolute prediction errors—visualized as lighter areas on the heatmaps—and exhibit the strongest correlation with survey yields ($r = 0.6783$).

These results highlight the distinctive advantage of exponential NDVIsm: it aligns closely with the actual yield distribution in NIR/R space and effectively captures the nonlinear crop response. Notably, even without any model training, NDVIsm alone is able to explain a substantial portion of yield variation (e.g., $R^2 = 0.4923$ in Illinois). This suggests that, when combined with additional data sources and more advanced machine learning models, further improvements in prediction accuracy can be reasonably expected. In contrast, NDVI suffers from saturation and fails to reflect the physiological yield dynamics, potentially introducing noise into prediction models. The concave shape of the exponential NDVIsm curve mirrors the yield-NIR/R trajectory, making it a more robust and interpretable indicator for yield prediction in densely maize regions.

5. Conclusion

To address the NDVI saturation issue common encountered in environmental remote sensing with healthy and dense vegetation, this study thoroughly analyzed its underlying causes and proposed a two-stage saturation mechanism. In the mathematical saturation stage, an exponential function was introduced as a growth rate adjustment to modify the increase rate of NDVI, leading to the development of anti-saturation VI, NDVIsm. Multiple validations across remote sensing platforms, comparative VIs and ML models confirmed the stability and effectiveness of NDVIsm in mitigating saturation effects. The key conclusions are as follows:

1. The NDVIsm enhanced by the exponential function has a higher CV and Skew, which means it has a more diffuse and more pronounced right-skewed distribution. This property helps the high values in the NDVI affected by saturation (Other VIs like EVI and SAVI are also proved to be saturated) to be significantly pulled apart, making index differences in the NDVIsm more apparent in areas with healthy and dense vegetation. This distributional pattern of NDVIsm is crucial for overcoming the saturation.
2. The anti-saturation distributional pattern of NDVIsm were verified across multiple land cover types (rainforests, farmlands and grasslands) under multiple satellite platforms (Sentinel-2, Landsat-8 and Modis-09GA). Experiments based on the UAV platform further demonstrated that NDVIsm exhibited stronger correlations with

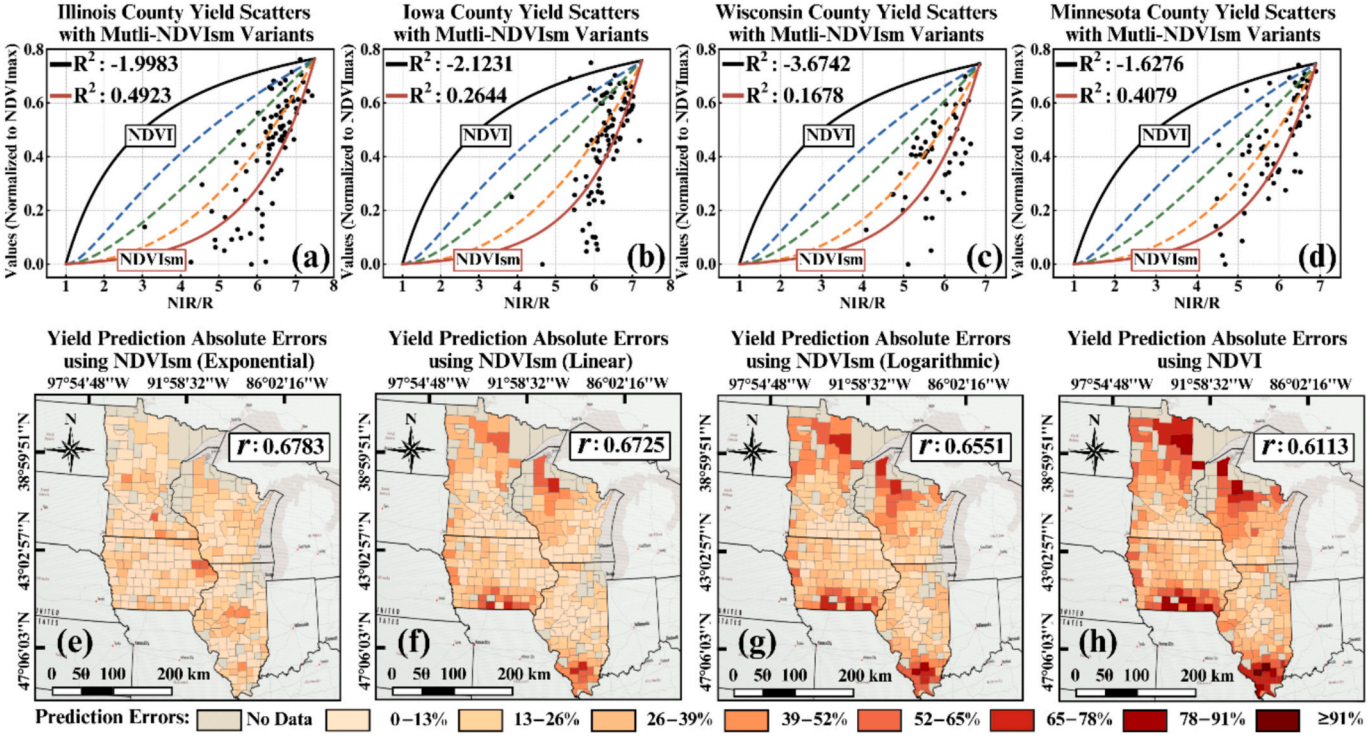

Fig. 14. The relationship between the distributional characteristics of NDVIsm with county level maize yield.

canopy structural traits (LAI, N) and vegetation vigor (Cab), and contributed more effectively to yield prediction models when integrated with non-optical data and spatial distribution features. These improvements highlight the effectiveness of NDVIsm in mitigating saturation under dense vegetation conditions.

3. NDVIsm is not universally applicable for all vegetation assessments, as it was originally designed to address index saturation caused by uniformly vigorous vegetation. NDVIsm offers greater flexibility through its potential for generating customized variants. Designing appropriate variants that align with the characteristics of specific target species is essential for improving assessment accuracy. In this study, the exponential variant of NDVIsm more effectively captured the variation trends of maize yield at the county level, whereas NDVI failed to do so. This also highlights the misuse of NDVI in yield assessments under dense vegetation conditions, where its contribution becomes limited due to saturation effects. Overall, we proposed strategies for generating feasible NDVIsm variants based on NDVI, offering new insights for mitigating saturation problem in related research.

Funding

This work was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture, Agriculture and Food Research Initiative Foundational Program (Award No. 2022-67021-36469); and Wisconsin Alumni Research Foundation.

CRediT authorship contribution statement

Zezhong Tian: Writing – original draft, Visualization, Validation, Software, Methodology, Data curation, Conceptualization. **Jiahao Fan:** Writing – review & editing, Data curation, Conceptualization. **Tong Yu:** Writing – review & editing, Methodology, Formal analysis. **Natalia de Leon:** Writing – review & editing, Data curation. **Shawn M. Kaeppler:** Writing – review & editing, Data curation. **Zhou Zhang:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Bannari, A., Morin, D., Bonn, F., Huete, A.R., 1995. A review of vegetation indices. Remote Sens. Rev. 13, 95–120. https://doi.org/10.1080/02757259509532298.

Barbosa, H.A., Huete, A.R., Baethgen, W.E., 2006. A 20-year study of NDVI variability over the northeast region of Brazil. J. Arid Environ. 67, 288–307. https://doi.org/10.1016/j.jaridenv.2006.02.022.

Beamish, A., Raynolds, M.K., Epstein, H., Frost, G.V., Macander, M.J., Bergstedt, H., Bartsch, A., Kruse, S., Miles, V., Tanis, C.M., Heim, B., Fuchs, M., Chabrillat, S., Shevtsova, I., Verdonen, M., Wagner, J., 2020. Recent trends and remaining challenges for optical remote sensing of Arctic tundra vegetation: a review and outlook. Remote Sens. Environ. 246, 111872. https://doi.org/10.1016/j.rse.2020.111872.

Bendel, R.B., Higgins, S.S., Teberg, J.E., Pyke, D.A., 1989. Comparison of skewness coefficient, coefficient of variation, and gini coefficient as inequality measures within populations. Oecologia 78, 394–400. https://doi.org/10.1007/BF00379115.

Benesty, J., Chen, J., Huang, Y., Cohen, I., 2009. Pearson Correlation Coefficient. pp. 1–4. doi: 10.1007/978-3-642-00296-0_5.

Berger, K., Atzberger, C., Danner, M., D'Urso, G., Mauser, W., Vuolo, F., Hank, T., 2018. Evaluation of the PROSAIL model capabilities for future hyperspectral model environments: a review study. Remote Sens. (Basel) 10, 85. https://doi.org/10.3390/rs10010085.

Biau, G., Scornet, E., 2016. A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s11749-016-0481-7.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324.

Carlson, T.N., Ripley, D.A., 1997. On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens. Environ. 62, 241–252. https://doi.org/10.1016/S0034-4257(97)00104-1.

Chicco, D., Warrens, M.J., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7, e623.

Ciganda, V., Gitelson, A., Schepers, J., 2008. Vertical profile and temporal variation of chlorophyll in maize canopy: quantitative "crop vigor" indicator by means of reflectance-based techniques. Agron. J. 100, 1409–1417. https://doi.org/10.2134/agronj2007.0322.

David, R.M., Rosser, N.J., Donoghue, D.N.M., 2022. Improving above ground biomass estimates of southern Africa dryland forests by combining Sentinel-1 SAR and Sentinel-2 multispectral imagery. Remote Sens. Environ. 282, 113232. https://doi.org/10.1016/j.rse.2022.113232.

Disha, R.A., Waheed, S., 2022. Performance analysis of machine learning models for intrusion detection system using gini impurity-based weighted random forest (GIWRF) feature selection technique. Cybersecurity 5, 1. https://doi.org/10.1186/s42400-021-00103-8.

Duan, S.-B., Li, Z.-L., Wu, H., Tang, B.-H., Ma, L., Zhao, E., Li, C., 2014. Inversion of the PROSAIL model to estimate leaf area index of maize, potato, and sunflower fields from unmanned aerial vehicle hyperspectral data. Int. J. Appl. Earth Obs. Geoinf. 26, 12–20. https://doi.org/10.1016/j.jag.2013.05.007.

Eugenio, F.C., Grohs, M., Venancio, L.P., Schuh, M., Bottega, E.L., Ruoso, R., Schons, C., Mallmann, C.L., Badin, T.L., Fernandes, P., 2020. Estimation of soybean yield from machine learning techniques and multispectral RPAS imagery. Remote Sens. Appl. 20, 100397. https://doi.org/10.1016/j.rsase.2020.100397.

Farbo, A., Sarvia, F., De Petris, S., Basile, V., Borgogno-Mondino, E., 2024. Forecasting corn NDVI through AI-based approaches using sentinel 2 image time series. ISPRS J. Photogramm. Remote Sens. 211, 244–261. https://doi.org/10.1016/j.isprsjprs.2024.04.011.

Gnyp, M.L., Miao, Y., Yuan, F., Ustin, S.L., Yu, K., Yao, Y., Huang, S., Bareth, G., 2014. Hyperspectral canopy sensing of paddy rice aboveground biomass at different growth stages. Field Crops Res. 155, 42–55. https://doi.org/10.1016/j.fcr.2013.09.023.

Gu, Y., Wylie, B.K., Howard, D.M., Phuyal, K.P., Ji, L., 2013. NDVI saturation adjustment: a new approach for improving cropland performance estimates in the greater Platte River basin, USA. Ecol. Ind. 30, 1–6. https://doi.org/10.1016/j.ecolind.2013.01.041.

Guo, A., Ye, H., Huang, W., Qian, B., Wang, J., Lan, Y., Wang, S., 2023. Inversion of maize leaf area index from UAV hyperspectral and multispectral imagery. Comput. Electron. Agric. 212, 108020. https://doi.org/10.1016/j.compag.2023.108020.

Haralick, R.M., Shanmugam, K., Dinstein, I., 1973. Textural features for image classification. IEEE Trans. Syst. Man. Cybern. SMC-3 610–621. https://doi.org/10.1109/TSMC.1973.4309314.

Hinojo-Hinojo, C., Goulden, M.L., 2020. Plant traits help explain the tight relationship between vegetation indices and gross primary production. Remote Sens. (Basel) 12, 1405. https://doi.org/10.3390/rs12091405.

Huang, S., Tang, L., Hupy, J.P., Wang, Y., Shao, G., 2021. A commentary review on the use of normalized difference vegetation index (NDVI) in the era of popular remote sensing. J. Res. (Harbin) 32, 1–6. https://doi.org/10.1007/s11676-020-01155-1.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83, 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2.

Huete, A.R., 2012. Vegetation indices, remote sensing and forest monitoring. Geogr. Compass 6, 513–532. https://doi.org/10.1111/j.1749-8198.2012.00507.x.

Huete, A.R., 1988. A soil-adjusted vegetation index (SAVI). Remote Sens. Environ. 25, 295–309. https://doi.org/10.1016/0034-4257(88)90106-X.

Huete, A.R., HuiQing Liu, van Leeuwen, W.J.D., 1997. The use of vegetation indices in forested regions: issues of linearity and saturation. In: IGARSS'97. 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings. Remote Sensing - A Scientific Vision for Sustainable Development. IEEE, pp. 1966–1968. doi: 10.1109/IGARSS.1997.609169.

Jackson, R.D., Huete, A.R., 1991. Interpreting vegetation indices. Prev. Vet. Med. 11, 185–200. https://doi.org/10.1016/S0167-5877(05)80004-2.

Jiang, Z., Huete, A.R., Chen, J., Chen, Y., Li, J., Yan, G., Zhang, X., 2006. Analysis of NDVI and scaled difference vegetation index retrievals of vegetation fraction. Remote Sens. Environ. 101, 366–378. https://doi.org/10.1016/j.rse.2006.01.003.

Johnson, M.D., Hsieh, W.W., Cannon, A.J., Davidson, A., Bédard, F., 2016. Crop yield forecasting on the Canadian prairies by remotely sensed vegetation indices and machine learning methods. Agric. For. Meteorol. 218–219, 74–84. https://doi.org/10.1016/j.agrformet.2015.11.003.

Joiner, J., Yoshida, Y., Anderson, M., Holmes, T., Hain, C., Reichle, R., Koster, R., Middleton, E., Zeng, F.-W., 2018. Global relationships among traditional reflectance vegetation indices (NDVI and NDII), evapotranspiration (ET), and soil moisture variability on weekly timescales. Remote Sens. Environ. 219, 339–352. https://doi.org/10.1016/j.rse.2018.10.020.

Kayad, A., Rodrigues, F.A., Naranjo, S., Sozzi, M., Pirotti, F., Marinello, F., Schulthess, U., Defourny, P., Gerard, B., Weiss, M., 2022. Radiative transfer model inversion using high-resolution hyperspectral airborne imagery – retrieving maize LAI to access biomass and grain yield. Field Crops Res 282, 108449. https://doi.org/10.1016/j.fcr.2022.108449.

Kganyago, M., Adjorlolo, C., Mhangara, P., Tsoeleng, L., 2024. Optical remote sensing of crop biophysical and biochemical parameters: an overview of advances in sensor technologies and machine learning algorithms for precision agriculture. Comput. Electron. Agric. 218, 108730. https://doi.org/10.1016/j.compag.2024.108730.

Khanal, S., KC, K., Fulton, J.P., Shearer, S., Ozkan, E., 2020. Remote Sensing in agriculture—accomplishments, limitations, and opportunities. Remote Sens (Basel) 12, 3783. doi: 10.3390/rs12223783.

Knipling, E.B., 1970. Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sens. Environ. 1, 155–159. https://doi.org/10.1016/S0034-4257(70)80021-9.

Kooistra, L., Clevers, J.G.P.W., 2016. Estimating potato leaf chlorophyll content using ratio vegetation indices. Remote Sens. Lett. 7, 611–620. https://doi.org/10.1080/2150704X.2016.1171925.

Kumar, S., Arya, S., Jain, K., 2022. A SWIR-based vegetation index for change detection in land cover using multi-temporal landsat satellite dataset. Int. J. Inf. Technol. 14, 2035–2048. https://doi.org/10.1007/s41870-021-00797-6.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Liu, F., Qin, Q., Zhan, Z., 2012. A novel dynamic stretching solution to eliminate saturation effect in NDVI and its application in drought monitoring. Chin. Geogr. Sci. 22, 683–694. https://doi.org/10.1007/s11769-012-0574-5.

Loh, W.-L., 1996. On latin hypercube sampling. Ann. Stat. 24. https://doi.org/10.1214/aos/1069362310.

Ma, Y., Chen, S., Ermon, S., Lobell, D.B., 2024. Transfer learning in environmental remote sensing. Remote Sens. Environ. 301, 113924. https://doi.org/10.1016/j.rse.2023.113924.

Maimaitijiang, M., Sagan, V., Sidike, P., Daloye, A.M., Erkbol, H., Fritschi, F.B., 2020. Crop monitoring using satellite/UAV data fusion and machine learning. Remote Sens. (Basel) 12, 1357. https://doi.org/10.3390/rs12091357.

Meng, J., Du, X., Wu, B., 2013. Generation of high spatial and temporal resolution NDVI and its application in crop biomass estimation. Int. J. Digit. Earth 6, 203–218. https://doi.org/10.1080/17538947.2011.623189.

Montandon, L., Small, E., 2008. The impact of soil reflectance on the quantification of the green vegetation fraction from NDVI. Remote Sens. Environ. 112, 1835–1845. https://doi.org/10.1016/j.rse.2007.09.007.

Mutanga, O., Masenyama, A., Sibanda, M., 2023. Spectral saturation in the remote sensing of high-density vegetation traits: a systematic review of progress, challenges, and prospects. ISPRS J. Photogramm. Remote Sens. 198, 297–309. https://doi.org/10.1016/j.isprsjprs.2023.03.010.

Nijland, W., de Jong, R., de Jong, S.M., Wulder, M.A., Bater, C.W., Coops, N.C., 2014. Monitoring plant condition and phenology using infrared sensitive consumer grade digital cameras. Agric. For. Meteorol. 184, 98–106. https://doi.org/10.1016/j.agrformet.2013.09.007.

Petrović, I., Vučajnk, F., Spanic, V., 2025. Detection of Fusarium head blight in wheat using NDVI from multispectral UAS measurements and its correlation with DON contamination. Agri Eng. 7, 37. https://doi.org/10.3390/agriengineering7020037.

Rehman, T.H., Lundy, M.E., Linquist, B.A., 2022. Comparative sensitivity of vegetation indices measured via proximal and aerial sensors for assessing N status and predicting grain yield in rice cropping systems. Remote Sens. (Basel) 14, 2770. https://doi.org/10.3390/rs14122770.

Safonova, A., Guirado, E., Maglinets, Y., Alcaraz-Segura, D., Tabik, S., 2021. Olive tree biovolume from UAV multi-resolution image segmentation with mask R-CNN. Sensors 21, 1617. https://doi.org/10.3390/s21051617.

Schmidt, K.S., Skidmore, A.K., 2003. Spectral discrimination of vegetation types in a coastal wetland. Remote Sens. Environ. 85, 92–108. https://doi.org/10.1016/S0034-4257(02)00196-7.

Thenkabail, P.S., Smith, R.B., De Pauw, E., 2000. Hyperspectral vegetation indices and their relationships with agricultural crop characteristics. Remote Sens. Environ. 71, 158–182. https://doi.org/10.1016/S0034-4257(99)00067-X.

van Leeuwen, W.J.D., Orr, B.J., Marsh, S.E., Herrmann, S.M., 2006. Multi-sensor NDVI data continuity: uncertainties and implications for vegetation monitoring applications. Remote Sens. Environ. 100, 67–81. https://doi.org/10.1016/j.rse.2005.10.002.

Vina, A., Gitelson, A.A., 2011. Sensitivity to foliar anthocyanin content of vegetation indices using green reflectance. IEEE Geosci. Remote Sens. Lett. 8, 464–468. https://doi.org/10.1109/LGRS.2010.2086430.

Wei, G., Shalei, S., Bo, Z., Shuo, S., Faquan, L., Xuewu, C., 2012. Multi-wavelength canopy LiDAR for remote sensing of vegetation: design and system performance. ISPRS J. Photogramm. Remote Sens. 69, 1–9. https://doi.org/10.1016/j.isprsjprs.2012.02.001.

Wolanin, A., Camps-Valls, G., Gómez-Chova, L., Mateo-García, G., van der Tol, C., Zhang, Y., Guanter, L., 2019. Estimating crop primary productivity with Sentinel-2 and landsat 8 using machine learning methods trained with radiative transfer simulations. Remote Sens. Environ. 225, 441–457. https://doi.org/10.1016/j.rse.2019.03.002.

Xie, C., Yang, C., 2020. A review on plant high-throughput phenotyping traits using UAV-based sensors. Comput. Electron. Agric. 178, 105731. https://doi.org/10.1016/j.compag.2020.105731.

Xue, J., Su, B., 2017. Significant remote sensing vegetation indices: a review of developments and applications. J. Sens. 2017, 1–17. https://doi.org/10.1155/2017/1353691.

Yang, T., Siddique, K.H.M., Liu, K., 2020. Cropping systems in agriculture and their impact on soil health – a review. Glob Ecol. Conserv. 23, e01118. https://doi.org/10.1016/j.gecco.2020.e01118.

Yang, W., Wang, S., Zhao, X., Zhang, J., Feng, J., 2015. Greenness identification based on HSV decision tree. Inform. Process. Agricult. 2, 149–160. https://doi.org/10.1016/j.inpa.2015.07.003.

Zhang, W., Qi, J., Wan, P., Wang, H., Xie, D., Wang, X., Yan, G., 2016. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. Remote Sens. (Basel) 8, 501. https://doi.org/10.3390/rs8060501.

Zhen, Z., Chen, S., Yin, T., Chavanon, E., Lauret, N., Guilleux, J., Henke, M., Qin, W., Cao, L., Li, J., Lu, P., Gastellu-Etchegorry, J.-P., 2021. Using the negative soil adjustment factor of soil adjusted vegetation index (SAVI) to resist saturation effects and estimate leaf area index (LAI) in dense vegetation areas. Sensors 21, 2115. https://doi.org/10.3390/s21062115.

Zhou, H., Zhou, G., Song, X., He, Q., 2022. Dynamic characteristics of canopy and vegetation water content during an entire maize growing season in relation to spectral-based indices. Remote Sens. (Basel) 14, 584. https://doi.org/10.3390/rs14030584.

Zhu, X., Liu, D., 2015. Improving forest aboveground biomass estimation using seasonal Landsat NDVI time-series. ISPRS J. Photogramm. Remote Sens. 102, 222–231. https://doi.org/10.1016/j.isprsjprs.2014.08.014.