CROP YIELD PREDICTION INDEX

#### SEMINAR REPORT

***Submitted in partial fulfilment of th*e *requirements for the award of the degree of***

## BACHELOR OF TECHNOLOGY

***in***

## ELECTRONICS & COMMUNICATION ENGINEERING

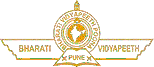
***by***

#### DEEPANSHU SATIJA

**Enrollment No:** 042115028042

***Guided by***

#### Dr. Rubeena Vohra Mentor

****

**DEPARTMENT OF ELECTRONICS & COMMUNICATION ENGINEERING BHARATI VIDYAPEETH’S COLLEGE OF ENGINEERING**

#### (AFFILIATED TO GURU GOBIND SINGH INDRAPRASTHA UNIVERSITY, DELHI)

**NEW DELHI – 110063**

# CANDIDATE’S DECLARATION

It is hereby certified that the work which is being presented in the B. Tech Seminar entitled

**"CROP YIELD PREDICTION INDEX"** in partial fulfilment of the requirements

for the award of the degree of **Bachelor of Technology** and submitted in the

#### Department of

**Electronics & Communication Engineering** of **BHARATI VIDYAPEETH’S COLLEGE**

**OF ENGINEERING, New Delhi (Affiliated to Guru Gobind Singh Indraprastha University, Delhi)** is an authentic record of our work carried out during a period from **Sep 0423 to Dec 0423** under the guidance of **Dr. Rubeena Vohra, Mentor.**

#### DEEPANSHU SATIJA En. No: 04211502820

This is to certify that the above statement made by the candidate is correct to the best of my knowledge. He/She/They are permitted to appear in the External Major Project Examination

#### Dr. Rubeena Vohra Dr. Kirti Gupta

**Mentor Head, ECE**

# ABSTRACT

Crop yield prediction is crucial for ensuring food security and optimizing agricultural practices. This study aims to develop a robust and reliable crop yield prediction index for a specific crop in a particular region. The index will consider various factors influencing crop yield, including historical data, climate data, soil data, remote sensing data, and management practices.

The proposed index will leverage machine learning and artificial intelligence techniques to analyze the complex relationships between these factors and crop yield. The index will be validated and calibrated using historical data to ensure its accuracy and reliability.

These indices will enable farmers to forecast future crop yields with increased accuracy, identify potential risks and areas of concern, and gain valuable insights into the impact of various factors on crop production. This information can then be used to inform decision-making, optimize resource allocation, and ultimately improve crop yield and sustainability. The development of this index has the potential to contribute significantly to food security, sustainable agriculture, and precision farming practices.

In the field of remote sensing applications, scientists have developed vegetation indices (VI) for qualitatively and quantitatively evaluating vegetative covers using spectral measurements. The spectral response of vegetated areas presents a complex mixture of vegetation, soil brightness, environmental effects, shadow, soil color and moisture. Moreover, the VI is affected by spatial‐temporal variations of the atmosphere. Over forty vegetation indices have been developed during the last two decades in order to enhance vegetation response and minimize the effects of the factors described above. This paper summarizes, refers and discusses most of the vegetation indices found in the literature. It presents different existing classifications of indices and proposes to group them in a new classification.

# ACKNOWLEDGEMENT

I, DEEPANSHU SATIJA express my deep gratitude to **Dr. Rubeena Vohra**, **Mentor**, Department of Electronics & Communication Engineering for her valuable guidance and suggestion throughout my project work.

I would like to extend my sincere thanks to the Head of the Department, **Dr. Kirti Gupta** for her time-to-time suggestions to complete my project work.

I am also thankful to Dr. Dharmendra Saini, Principal for providing me the facilities to carry out my project work.

DEEPANSHU SATIJA (En. No: 04211502820)

**TABLE OF CONTENTS**

## CANDIDATE DECLARATION

* **ABSTRACT**

## ACKNOWLEDGEMENT

* **INTRODUCTION**

## REFLECTANCE BAND & VI

* **VEGETATION INDEX**

## NDVI

* + **SAVI**
  + **EVI**
* **CONCLUSION**
* **REFERNCES**

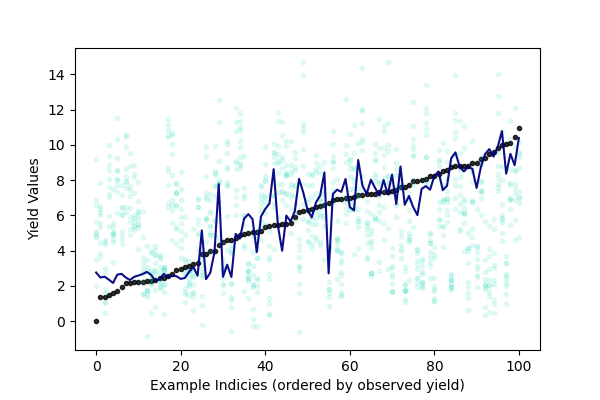
# INTRODUCTION

Crop yield prediction plays a crucial role in modern agriculture, aiding farmers in making informed decisions about their crops. This involves estimating the number of crops expected in a given area, considering various factors like soil type, weather conditions, and crop management practices. In recent years, machine learning (ML) has emerged as a powerful tool for predicting crop yields with increasing accuracy.

The importance of crop yield estimation extends beyond individual farmers, impacting the broader context of precision agriculture. This approach utilizes data and technology to manage crops efficiently, optimizing resource allocation and maximizing yield. Crop yield estimation indexes serve as valuable tools in this regard, generating data-driven maps that guide farmers in making informed decisions about various aspects like planting, irrigation, and fertilization, even at the individual plant level.

By providing valuable insights and enabling data-driven decision-making, crop yield prediction indexes offer numerous benefits:

* **Informed Decision-Making:** Farmers can anticipate crop yield based on real-time data and predictions, allowing for optimal planning and resource allocation.
* **Risk Identification:** Proactive measures can be taken by identifying potential risks like drought, pests, and diseases, mitigating potential losses.
* **Insights into Impact Factors:** The index helps understand how various factors like climate, soil, and management practices influence crop yield, enabling farmers to adopt optimal practices for improved productivity.
* **Data-Driven Agriculture:** Empowers farmers to make informed decisions based on real-time data and predictions, leading to sustainable and efficient agricultural practices.



**REFLECTANCE BANDS**

Vegetation Indices (VIs) are combinations of two or more reflectance bands from satellite images.

Bands in the 1st zone are in the **visible wavelength spectrum** (RGB: red, green and blue) and provide information on leaf pigmentation, which is useful to determine the growing stage of the crop, and a basic indicator of crop health. During photosynthesis, plants rely on the green pigment chlorophyll to convert energy from the sun for fuel.

The 2nd zone **near-infrared (NIR)** wavelengths are a great way to detect healthy growing plants through chlorophyll detection, representing plant productivity more directly.

**Bands in shortwave infrared** 3rd zone provide information on the water content with the plants as well as biochemical components in leaves.

**Spectral Reflectance curve:** Spectral reflectance curve shows the relationship between the electromagnetic spectrum (distribution of the continuum of radiant energies plotted either as a function of wavelength or of frequency) and the associated percent reflectance for any given material. It is plotted in a chart that represents wavelengths on the horizontal axis and percent reflectance on the vertical axis (fig. 1). This curve will visualize the formula of NDVI, NDBI, and NDWI.

#### 

**VEGETATION INDICES**

### **NDVI**

**Introduction: The Meaning of NDVI**

# The Normalized Difference Vegetation Index (NDVI) is a widely-used metric for quantifying the health and density of vegetation on Earth because it is robust, versatile and easy to interpret. It is robust because it has a high correlation with ground truth, irrespective of the type of vegetation in question. It is versatile because it has application in a wide range of fields.

# It is easy to interpret, not only because of its simplicity but because of its near linearity. By construction, NDVI will be a value between -1 and 1. A region with absolutely nothing growing in it will have an NDVI of zero. As you move across either time or space, NDVI will increase in proportion to the quantity and health of the vegetation in the region. It will reach its maximum value of 1.0 if the region is entirely covered with dense, healthy vegetation. NDVI values less than 0 suggest a lack of dry land; a lake or ocean will have an NDVI of -1.

# The conceptual simplicity of NDVI, while attractive, can also be a trap for practitioners: Interpreting NDVI requires a proper theoretical understanding of how it is constructed so that one can know when, where and how it can be deceptive. This article presents NDVI from first principles so as to equip a practitioner with a proper understanding of the index and how to use it.

**Derivation of the NDVI Formula**

NDVI follows from a scientific understanding of how different wavelengths of light are reflected (or not) by plants. There are two bands (red and near-infrared) that are key. First consider red light: if you shine it on a plant, not much of it will be reflected back because chlorophyll, (like any green matter), absorbs it.  Hence, if you measure the amount of red light that is reflected back to you, you are implicitly measuring “greenness”.

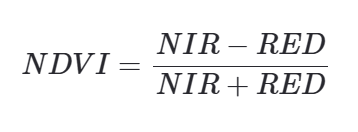
There is another wavelength that is useful: If you shine near-infrared light on healthy plants most of it will be reflected back. This is because the particular way in which plant cells are organized prevents near-infrared light from being absorbed. Thus the amount of near-infrared light reflected back in any region on Earth will vary with the amount of healthy plant cells present. (Whereas the amount of red light reflected back will vary inversely.)

For any particular location, if you measure the reflectance (in lumens per square meter) at the near-infrared (NIR) and red (RED) bands, you can take a simple measurement:



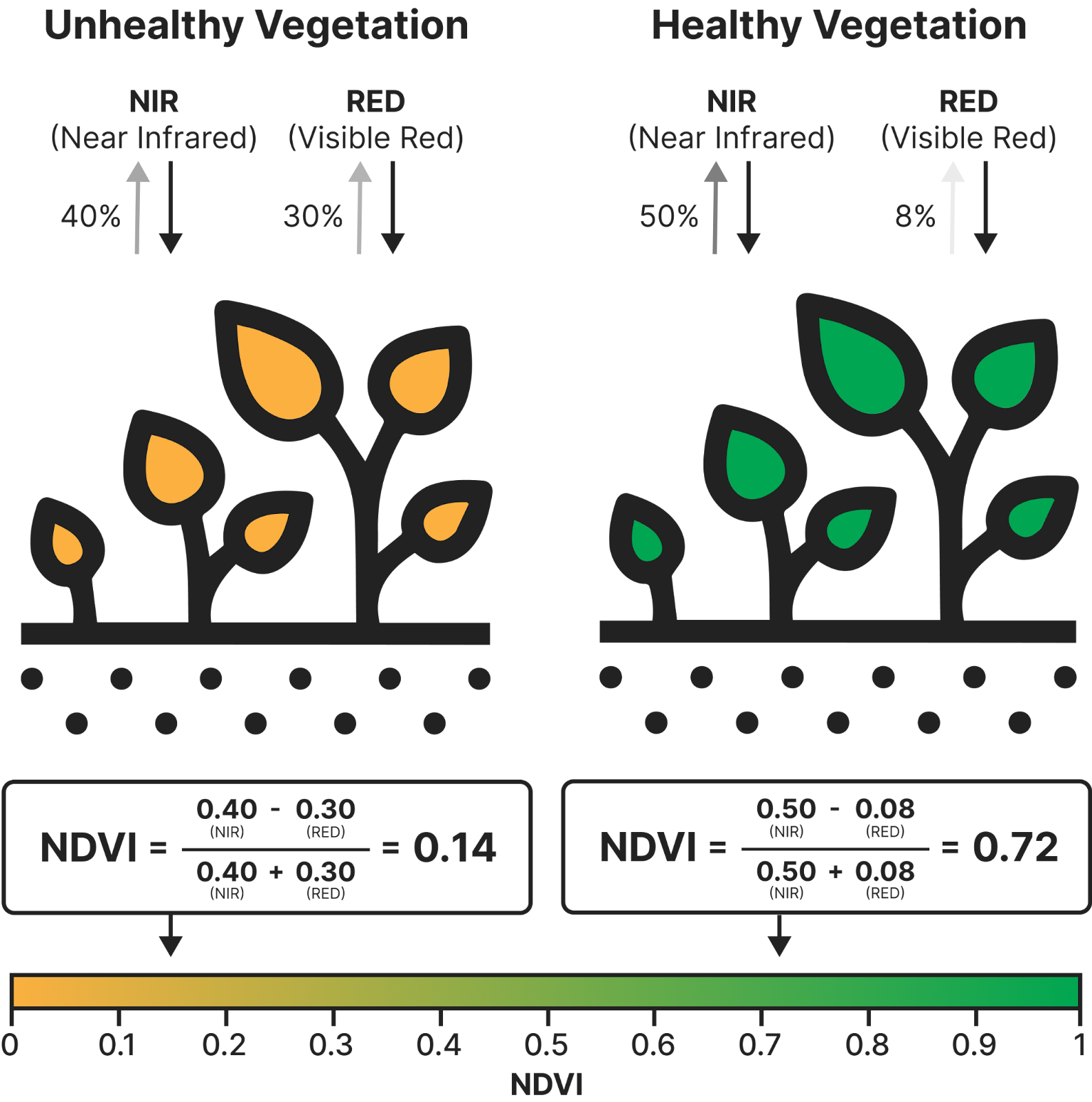
‍This difference will vary proportionately with the quantity and quality of the vegetation present. You can do this using sensor data from a satellite. The only difference being that the sensor output is a number ranging from 0 to 1, where 1 implies 100% of light is reflected back and 0 means none is.

There is one problem however: the measure will also vary with the intensity of the light. (When light intensity is 10% higher, so too is the difference between NIR and RED.) But this is easily corrected by normalizing for the total intensity (NIR+RED), hence the term “normalized difference”:

‍

NDVI is that simple; it is an elegant, unitless measurement that, by construction, will vary between -1 and 1, proportional to the health and density of the vegetation in the area being measured:

‍

‍

**Soil Effects and NDVI**

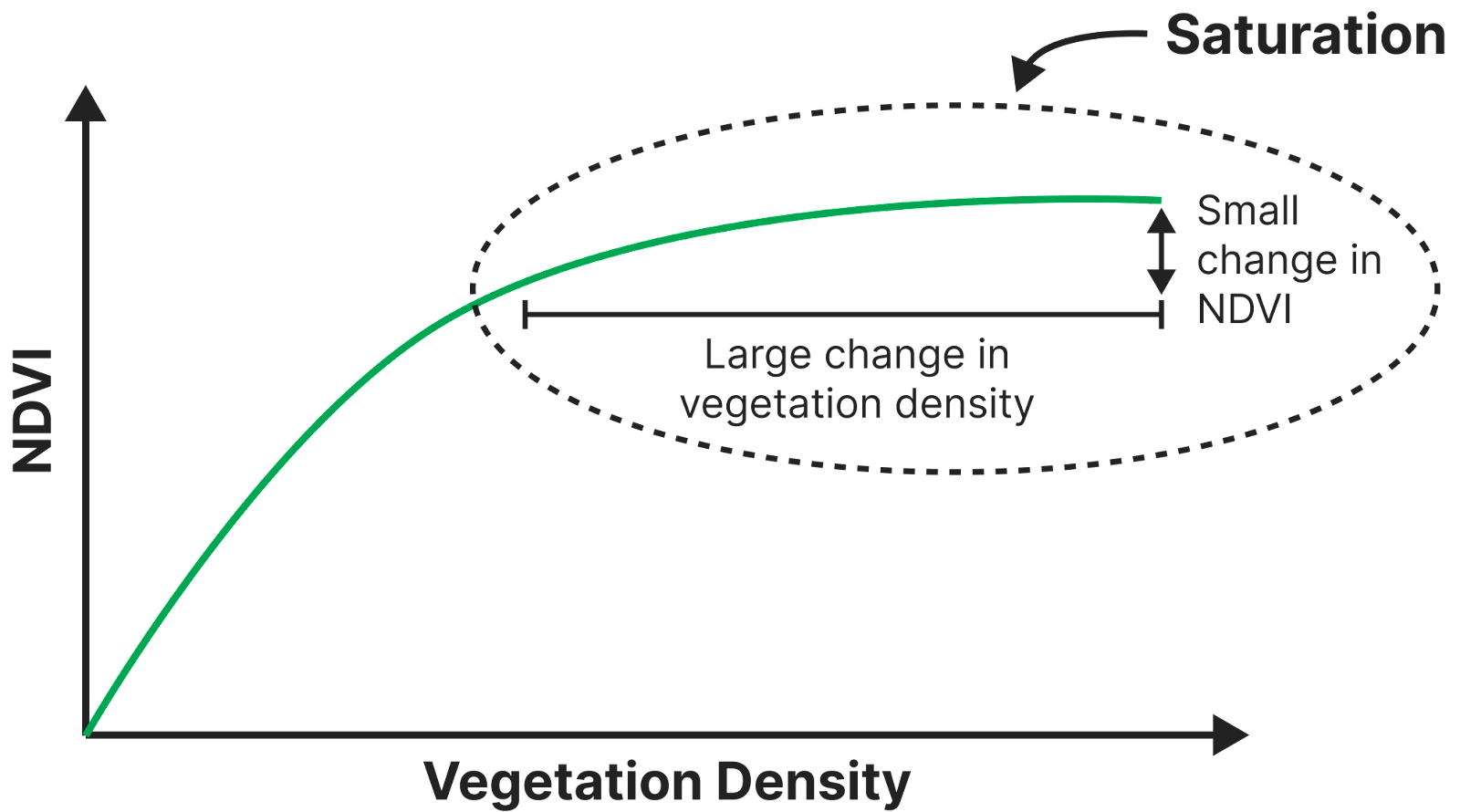
When soil gets wet, it gets darker. When it gets darker, it reflects less red light and less near-infrared light. Recall the denominator of the NDVI formula is the total near-infrared and red light reflected by the area in question. Therefore changing soil color will change the denominator of the NDVI equation leading to an altered NDVI value not because of vegetation changes but because of soil color variation. This is highly undesirable of course.

A practitioner using NDVI to study areas where there is sparse vegetation has to be alert to this effect and control for it. Aware of this limitation, they may In fact decide to adopt a different index, such as the [Soil Adjusted Vegetation Index (SAVI).](https://www.streambatch.io/knowledge/savi-from-first-principles)

**Vegetation Classification and NDVI**

NDVI at or near 1.0 tells you the ground is essentially covered by vegetation, but it doesn’t tell you what exactly is there. Consider an old growth rainforest and a golf course. The former is one of the highest density biomasses on Earth. The latter most certainly is not. If their densities were on a spectrum, these two terrains would be at opposite ends.

Rain forests and golf courses are indistinguishable from NDVI’s vantage point. This illustrates an important limitation of NDVI (and many other indexes too): they cannot tell you much about the nature of the vegetation they detect; they see chlorophyll and plant matter. As such, all vegetation basically looks the same. For this reason, NDVI is completely inappropriate for classification.



# SAVI

**Introduction: The Effect of Soil Color on NDVI**

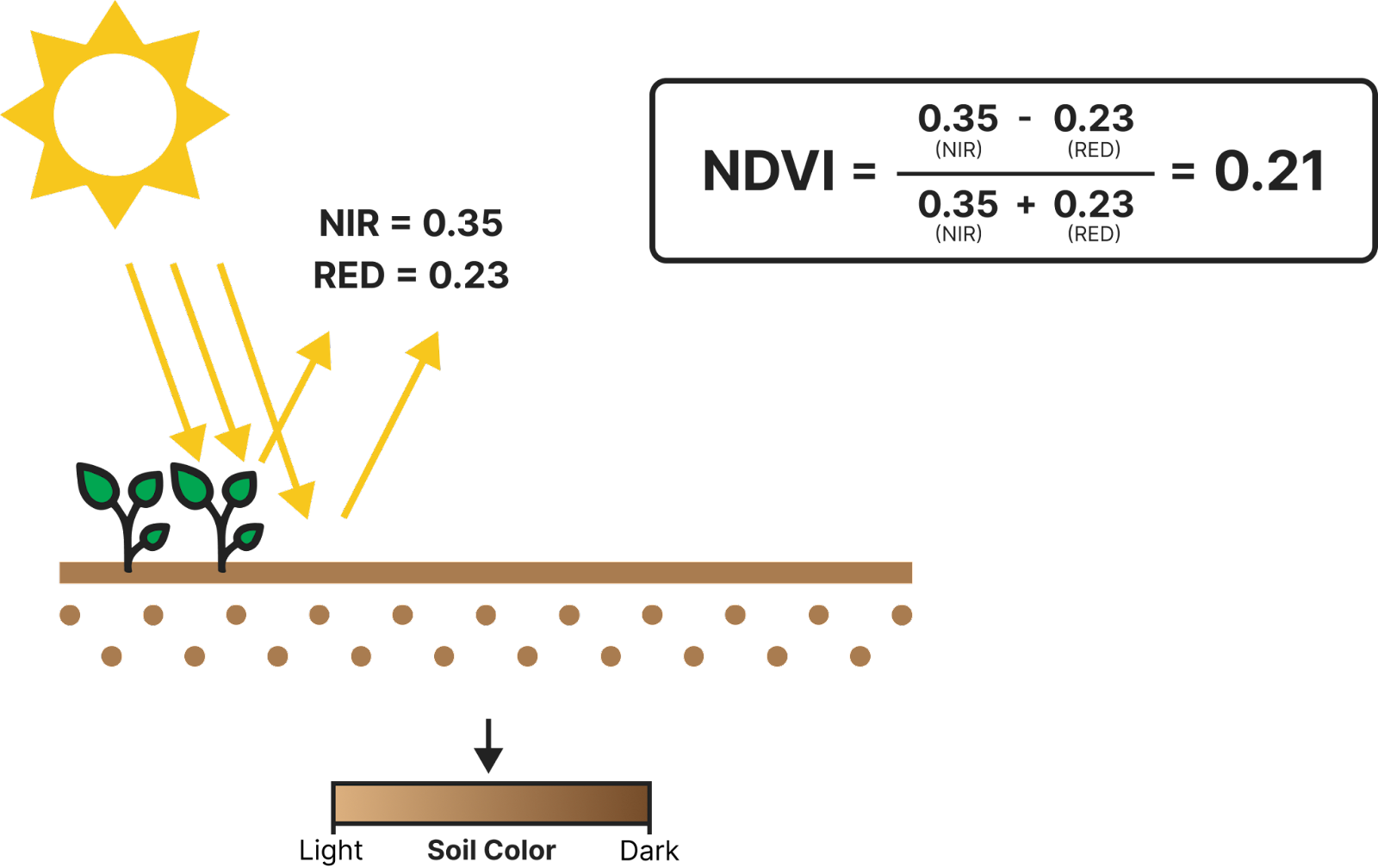
NDVI is a robust and versatile metric for assessing vegetation health on a global scale, but it is not optimal for all ground conditions. When vegetation is sparse, NDVI can fluctuate even if the state of vegetation does not change. This is a consequence of how soil in the area changes brightness depending on how wet or dry it is.

To understand this phenomenon, recall that NDVI follows from the idea that the reflection differential between red light and near infrared light (NIR - RED) will vary proportionally with the amount of vegetation present in the area being observed (See: [NDVI from First Principles).](https://streambatch.io/knowledge/ndvi-from-first-principles)

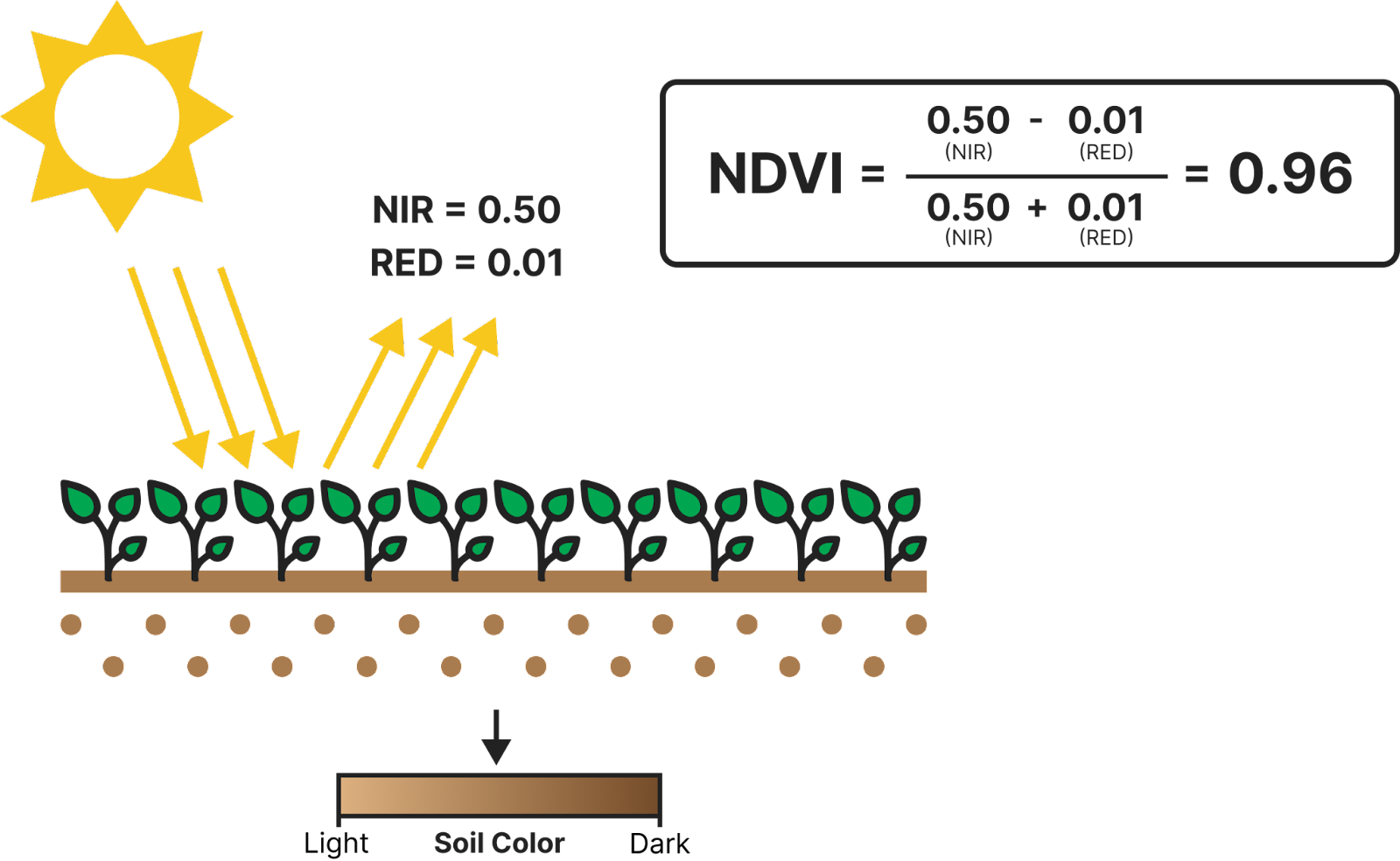
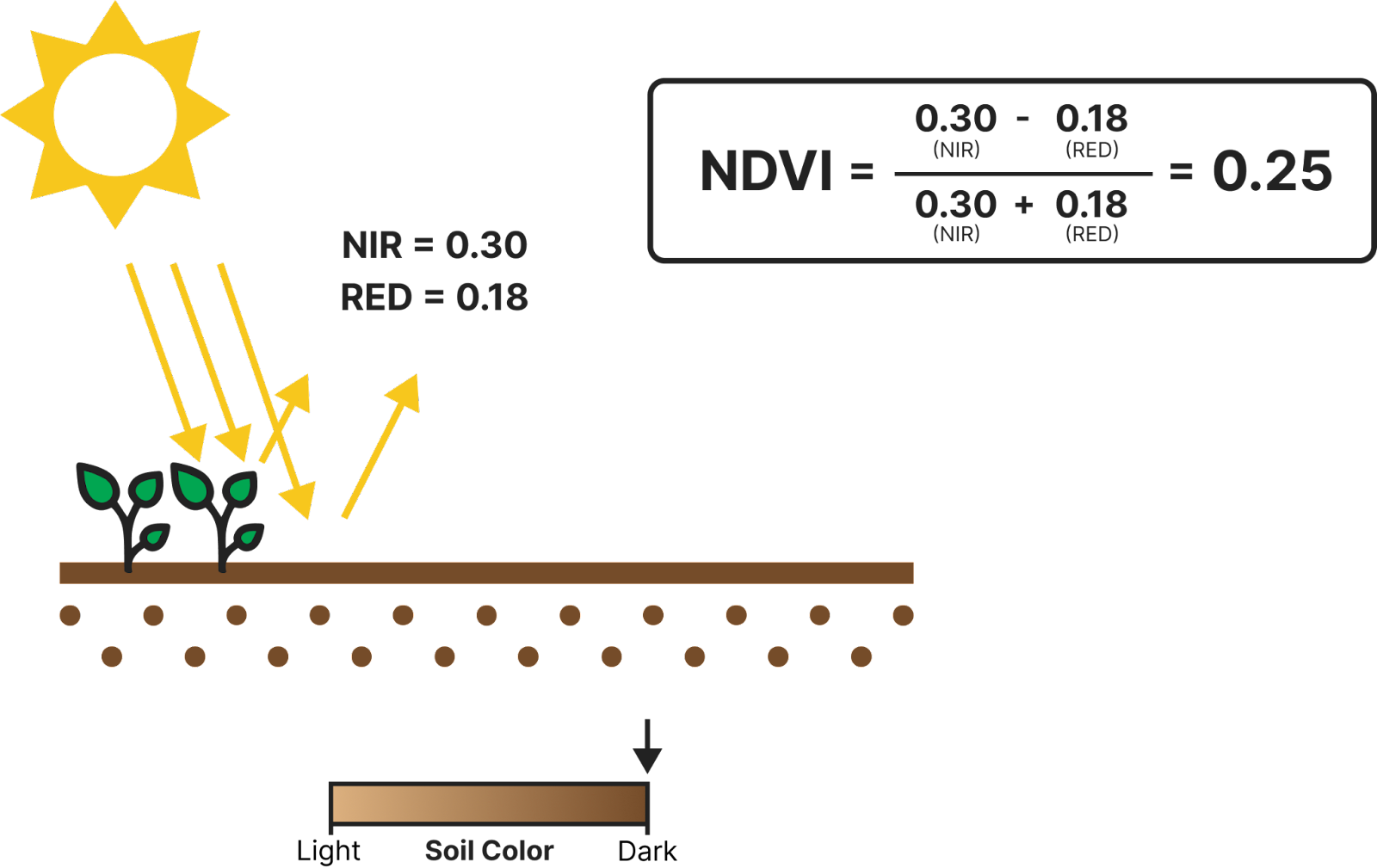
We can create an intuition for how NDVI works and where it breaks down with a stylized example, starting with an empty field. An empty field will reflect near-infrared and red light at nearly equal levels leading to a low NDVI:

# 

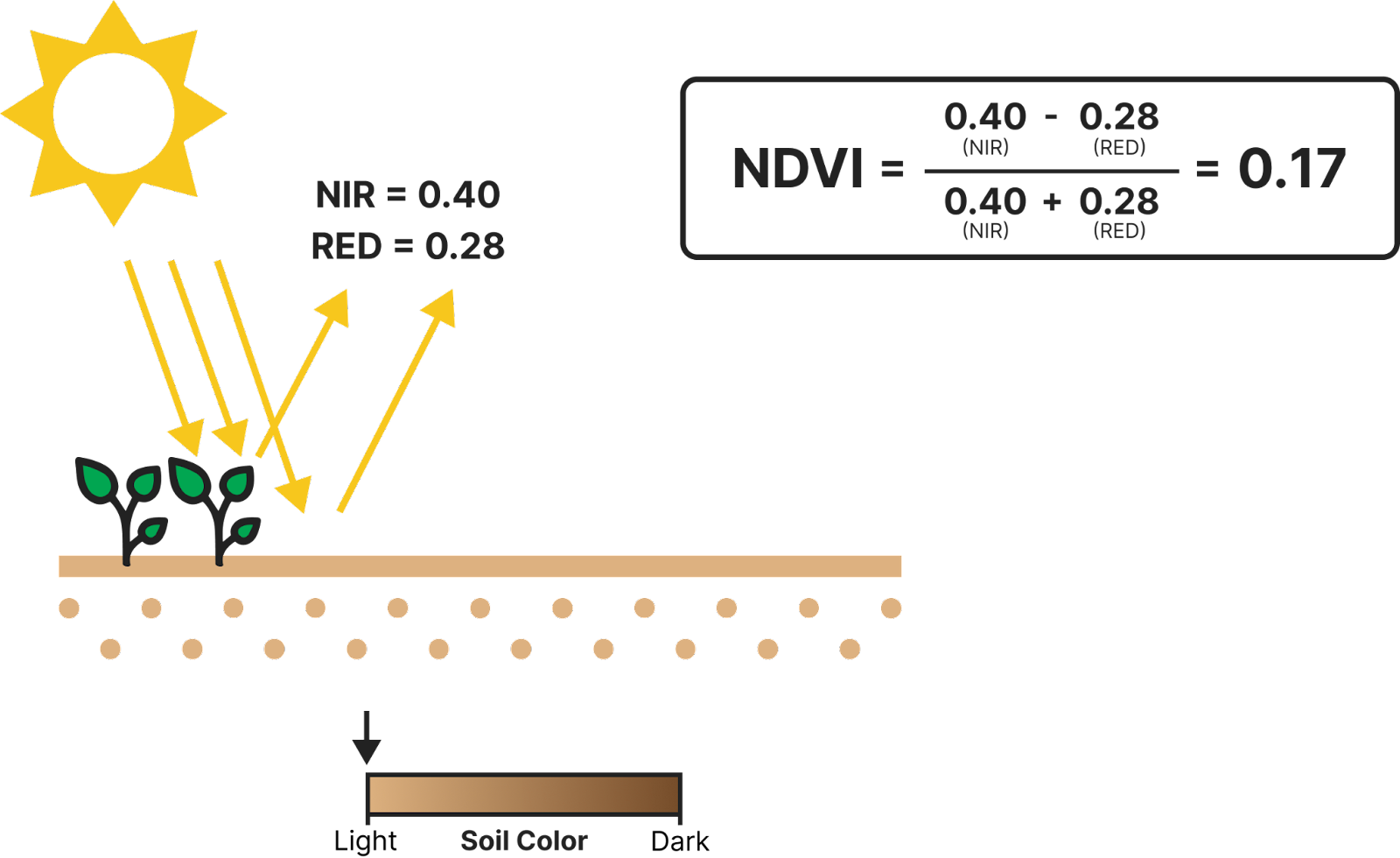
Now suppose we cover 20% of this field with vegetation. NDVI goes up of course:



If we cover the entire field with vegetation, NDVI will approach 1.0:

Now let’s return to our field that is 20% vegetated.  Suppose it rains. The 80% of the field covered in soil is now wet and, as a result, darker. The optical implication is that both near-infrared and red reflection levels will go down, by about the same amount. Note this causes NDVI to increase:

This of course is a problem. The NDVI for this area was 0.21. It rained, and NDVI went up to 0.25. We want a vegetation index to vary only as vegetation changes and not because it rains! The opposite thing happens when soil gets lighter as it becomes more dry:



**Remedies for Soil Effects: SAVI**

## **The change in reflectance from varied soil color causes** NIR **and** RED **to increase or decrease by similar amounts. The changes are similar enough that we can assume they are actually the same without affecting the analysis that follows. We’ll denote the change as epsilon. The NDVI result for changed soil conditions will thus be:**

## **Epsilon will be determined by how much of the field is covered by soil and how dark or light that soil is. If the field is mostly covered in vegetation, there will not be much visible soil at all meaning epsilon will be small relative to (NIR+RED). Ie, NIR+RED. Thus:**

## 

## **The math here simply affirms that if there is not much visible soil, NDVI will not be sensitive to soil color. However, when substantial soil is visible epsilon will be larger and will be significant relative to (**NIR+RED**):**

## 

## **The NDVI value will be materially impacted. What can be done? To start, let’s go back to this equation:**

## 

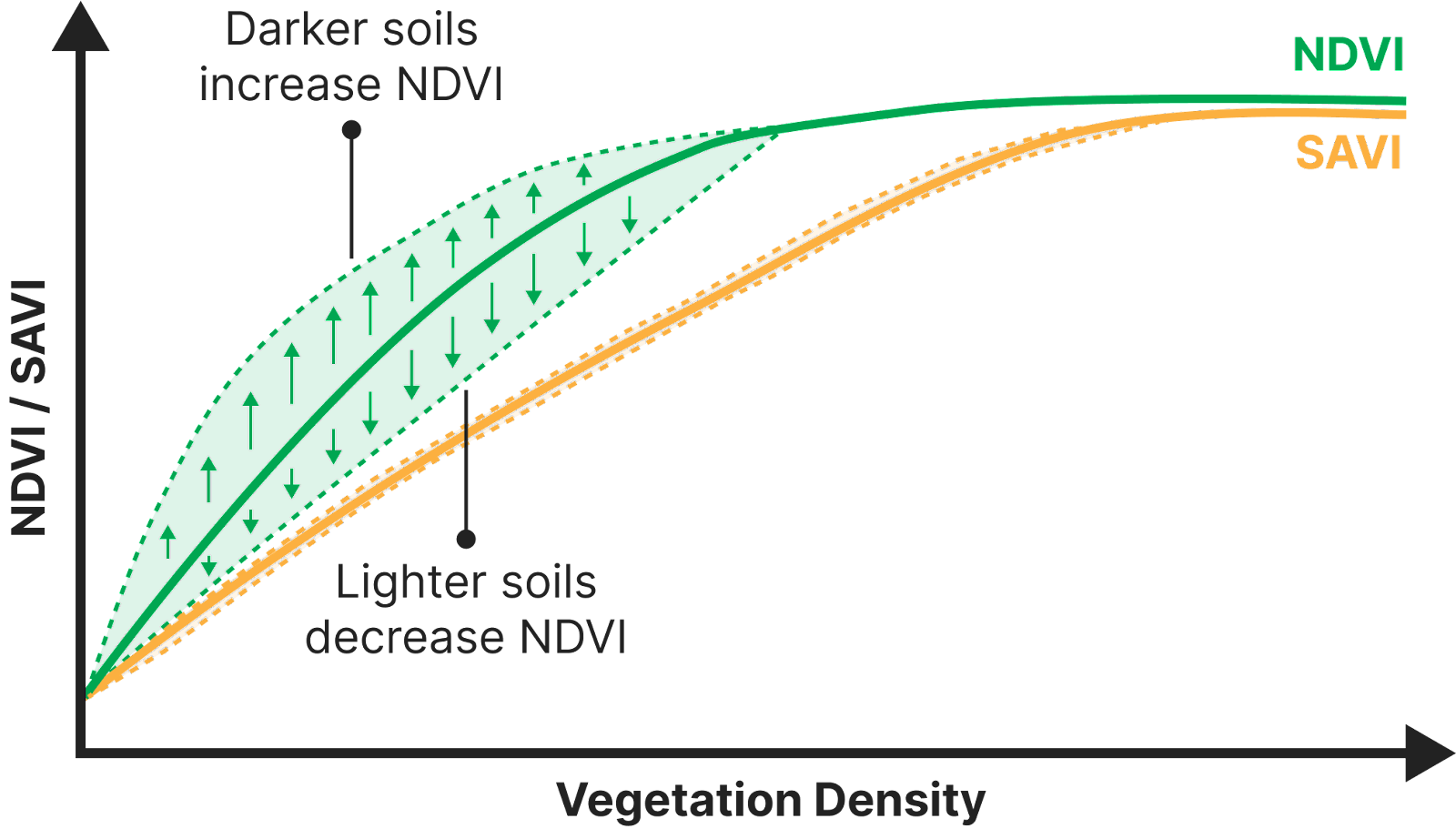
**Notice that epsilon does not affect the numerator of NDVI because it cancels. So, if you simply used** NIR-RED **as your index, you would not have any trouble from varying soil colors! You would have a bigger problem though. NDVI would vary based on just how sunny it is: On a day where total light intensity was 50% of normal because of moderate clouds,** NIR-RED **would also be 50% lower. Hence the need for normalizing by dividing by total intensity (**NIR+RED**).**

There is actually a simple mathematical adjustment you can make to the NDVI formula to reduce its sensitivity to epsilon: add a constant value to the denominator:

‍

## 

### **SAVI vs NDVI**



The figure above illustrates the difference between NDVI and SAVI at different vegetation densities. Notice that:

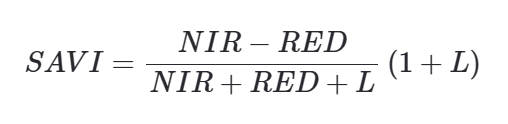
* At high vegetation density, NDVI is stable because minimal soil is visible
* At sparse and moderate vegetation levels, NDVI has substantial variance due to soil color sensitivity. SAVI has only modest variance.

A practitioner will prefer SAVI to NDVI in any situation where significant soil is visible and brightness changes are possible. SAVI is not unequivocally superior however; in mitigating soil brightness effects we have compromised the index’s previous complete insensitivity to total light intensity.

**SAVI and Saturation**

There is a second mathematical property that SAVI gains from the additional L term: the SAVI saturation threshold is higher than NDVI. To get an intuition for why, recall that NDVI saturates because red reflectance (RED) gets close to zero well before vegetation density reaches a maximum.

As RED gets close to zero, NDVI will still increase as NIR does, but only asymptotically. Ever larger increases in NIR have an ever diminishing effect on NDVI; i.e. saturation. But now consider the effect of adding L to the denominator:



‍

SAVI will still reach saturation, but it will happen at a larger NIR. (For intuition on this, set RED = 0 and make L very large relative to NIR. As such, L dominates the denominator. The function is now effectively just NIR/L. It increases linearly with NIR; i.e. there is no convexity and thus no saturation! Generalizing: as L increases from 0, the saturation point increases with it.)

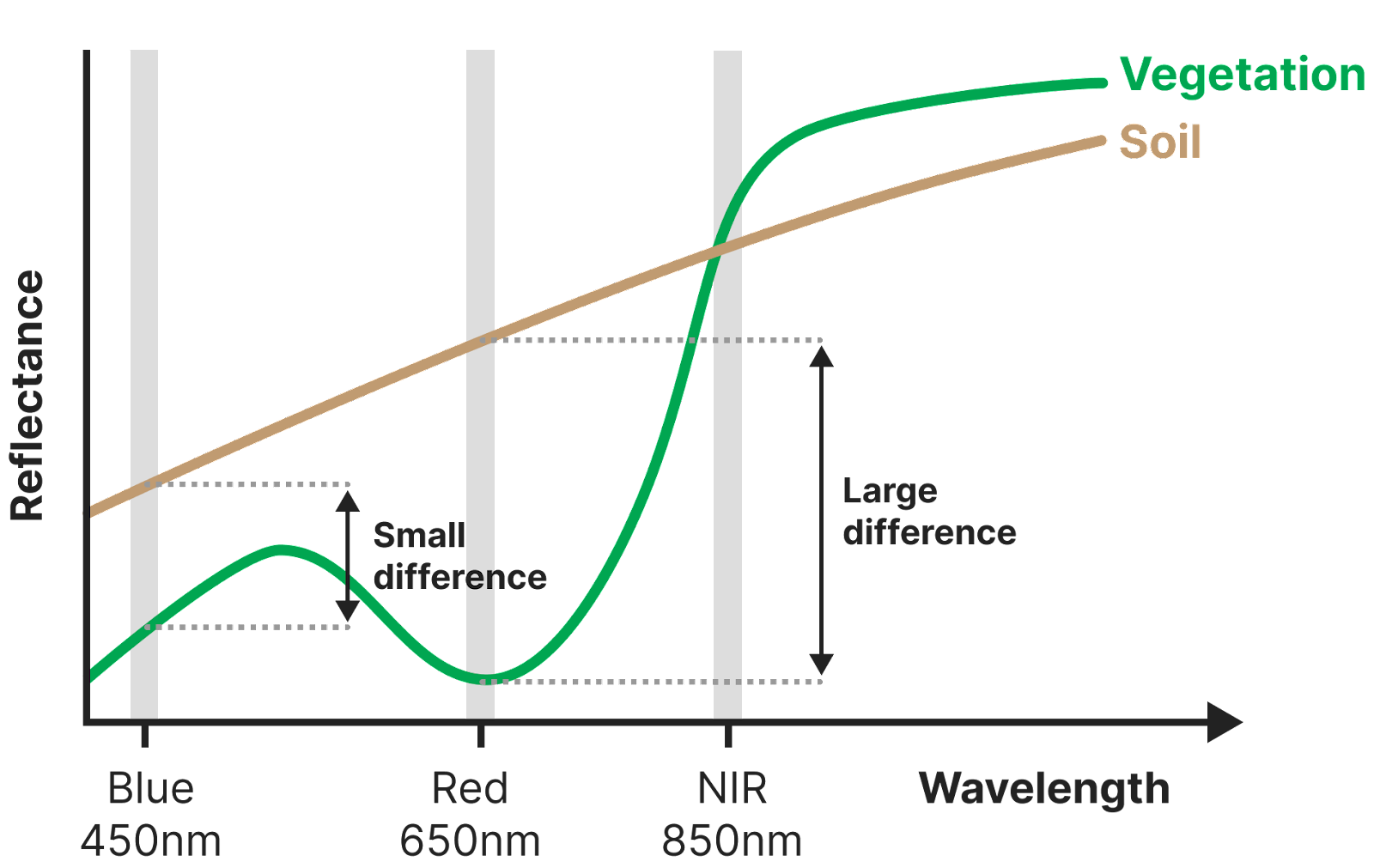
Thus, the additional term in the numerator allows SAVI to register increasing vegetation density after NDVI has already saturated.

## ****EVI****

### **Introduction: ARVI**

### In the 1980s and 1990s, atmospheric correction mechanisms were less developed than today, specifically in the case of aerosols. Aerosols are a problem because some red light, that might otherwise be absorbed by vegetation, gets reflected by aerosol particles and detected by sensors on satellites. This can lead to an overstated measure of red reflectance and hence an underestimate of vegetation levels.

### Blue light however, is far less susceptible to aerosol effects. It tends to pass through aerosols with minimal interference. Therefore, to minimize aerosol effects, one might consider building a vegetation index using blue reflectance instead of red. The problem with that is the reflectance differential between soil and vegetation for blue light is not as significant as the difference for red light; i.e. its harder to tell the difference between vegetation and soil using blue light compared to red light:

However, you can still use blue light to help deal with aerosol effects. The idea is this: when there is zero aerosol interference, there will be some baseline difference between how much red light is reflected and how much blue light is reflected from the Earth’s surface. That difference will of course not be completely insensitive to the relative amounts of soil and vegetation present, but that difference will actually be more sensitive to aerosol interference.

When there is aerosol interference, red reflectance will increase and blue will not. Therefore the difference between red reflectance and blue reflectance will correlate with the amount of aerosol interference.

This leads to the key idea behind the Atmospherically Resistant Vegetation Index (ARVI): The increase in red reflectance due to aerosols will be proportional to the difference between red and blue reflectance. We can use this relationship to correct the value for RED:



## Because RED is overstated due to aerosol effects, we apply a correction by reducing it by an amount proportional to the amount of aerosol interference. γ is determined by a calibration. And replace RED, with REDcorrected, we get:

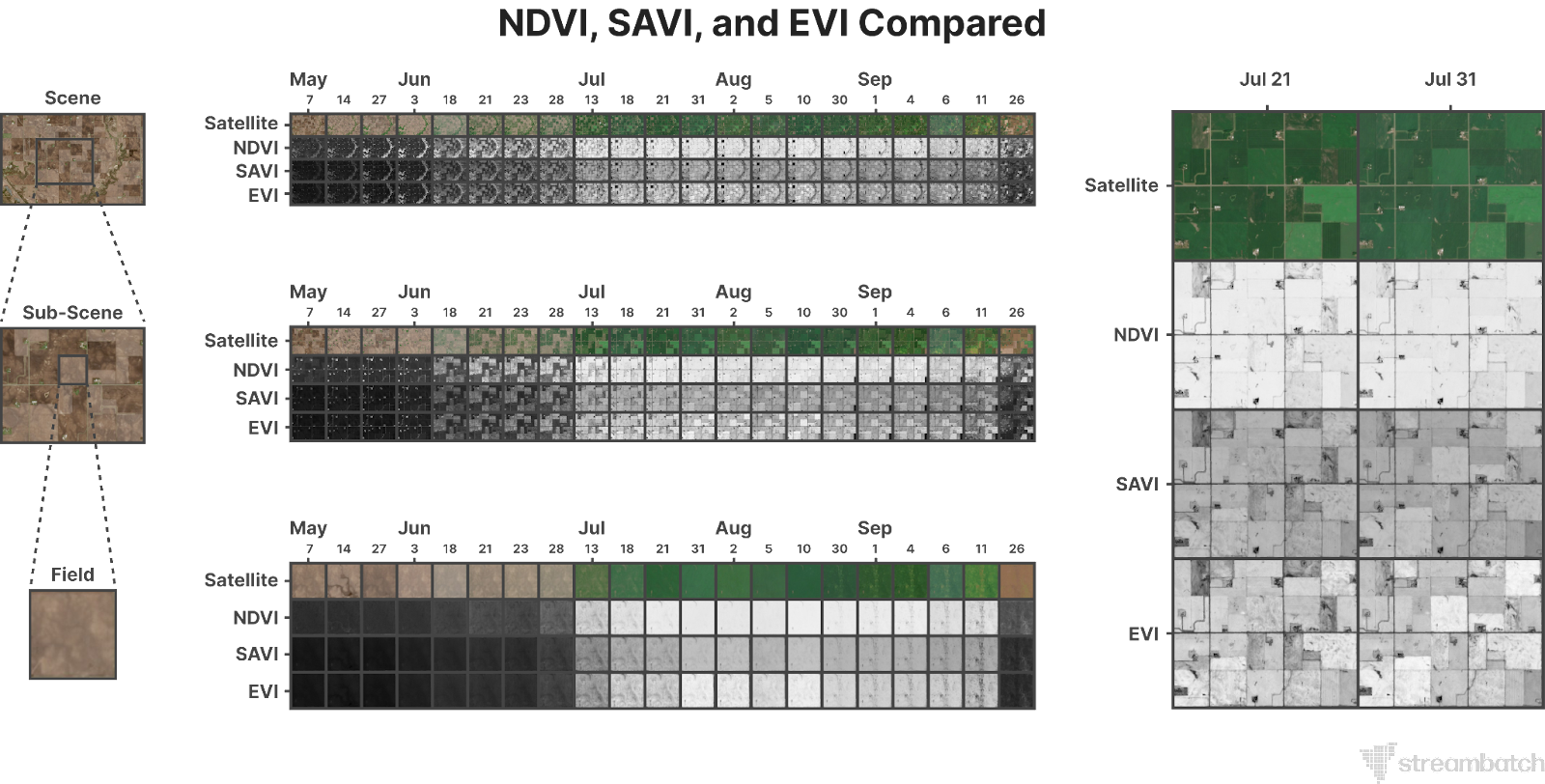
## 

# CONCLUSION

NDVI is easy to understand, derive and utilize. Because it offers a robust intuitive measure of vegetation health, it is readily used industrially for tracking seasonal changes and long term structural trends.

Any time there is soil exposure and especially if its moisture content can change, SAVI is going to be superior to NDVI. The better saturation characteristics of SAVI are highly desirable in any dense vegetation scenario.

EVI leverages the blue band to mitigate aerosol effects. But modern atmospheric correction mechanisms do that anyway making the use of the blue band redundant today. The one valuable innovation from EVI worth keeping, however, is the idea of designing an index with slightly less sensitivity to the red band than NDVI or SAVI.



Looking towards the future, several exciting advancements are expected in crop yield prediction:

* **Integration of New Data Sources:** Incorporation of data from drones, ground sensors, and other sources can further improve the accuracy and granularity of predictions.
* **Machine Learning Advancements:** Utilizing advanced machine learning techniques like deep learning can lead to even more accurate and reliable predictions.
* **Crop-Specific Models:** Customized models tailored to specific crops and regions can provide enhanced accuracy and insights relevant to the specific context.
* **Open Access Data and Tools:** Promoting open access to data and tools can empower farmers and democratize access to advanced crop yield prediction capabilities.

# REFERENCES

1. “CROPUP –A Crop Yield Prediction and Recommendation System with Geographical Data using DNN and XGBoost” Sobhana M1, Smitha Chowdary Ch2, D.N.V.S.L.S. Indira3, Konduru Kranthi Kumar. International Journal on Recent and Innovation Trends in Computing and Communication, 2321-8169 Volume: 10 Issue: 11
2. A Front. Plant Sci., 22 May 2019 Sec. Computational GenomicsVolume 10- 2019 “Crop Yield Prediction Using Deep Neural Networks” (https://doi.org/10.3389/fpls.2019.00621)
3. “Crop yield estimation model for Iowa using remote sensing and surface parameters” Anup K. Prasad, Lim Chai, Ramesh P. Singh, Menas Kafatos. International Journal of Applied Earth Observation and Geoinformation 8 (2006) 26–33.
4. Mateo-Sanchis, A.; Piles, M.; Muñoz-Marí, J.; Adsuara, J.E.; Pérez-Suay, A.; Camps-Valls, “G. Synergistic integration of optical and microwave satellite data for crop yield estimation. Remote Sens. Environ.” (2019), 234, 111460.