EPFL

Image processing for Earth Observation

Information extraction

1. Spectral indices

Devis TUIA

EPFL, fall semester 2024

École polytechnique fédérale de Lausanne

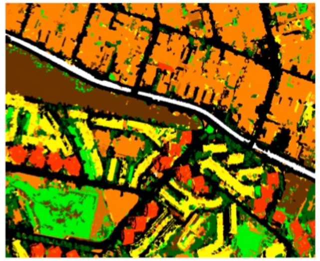
Why information extraction?

- We want the computer lo learn a pixel → product mapping

$$f(x_i) = y_i$$

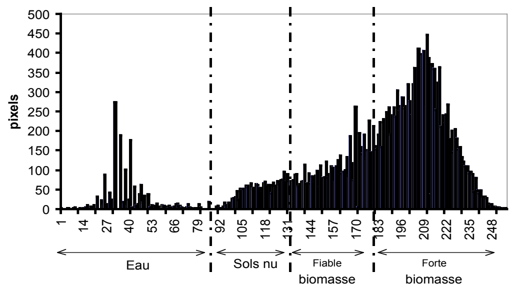
VHR image

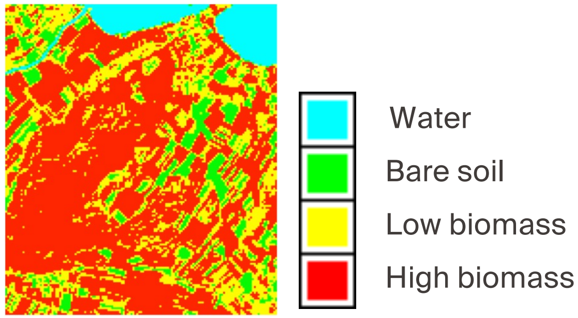
→ Model →



Classified image

- With only the pixel value itself (x_i = band value at pixel i), it is difficult!





Why information extraction?

- This is because a computer only sees numerical values
- It knows nothing of the context of your problem

For you

100	134	79	191	78	3	34	12	132	62
23	121	78	156	100	21	22	134	156	100
29	34	90	121	121	231	134	99	189	86
12	127	95	232	23	230	212	90	198	67
11	134	78	132	100	233	245	78	212	34
190	156	121	145	77	230	7	56	200	230
189	234	231	111	78	34	8	3	255	100
40	45	67	89	245	234	222	21	245	1

For the computer

OR

100	100	100
23	24	101
29	151	123
12	165	111

⋮

100	76	212
1	24	156

Why information extraction?

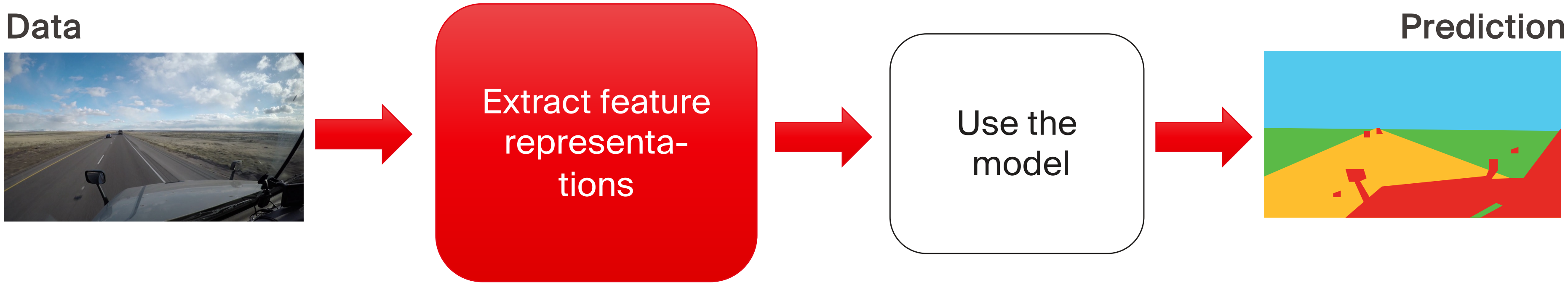
- Fortunately, we know already quite some a-priori about the problem
  - Images are multi-band, some bands are good at specific questions
  - Images are spatial, looking at groups of pixels can help recognition

- We can use this information to help the model
- The information will be extracted as feature representations

- (we will see the models later on in the course, don't worry)

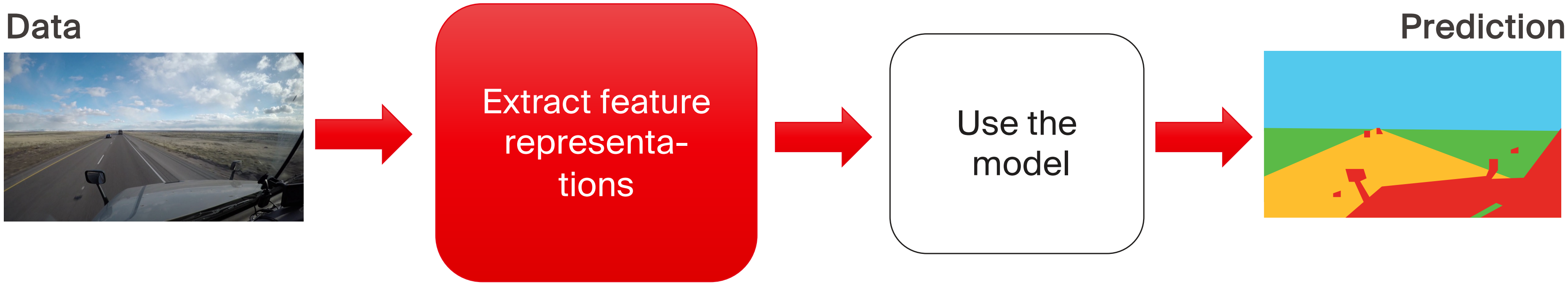
What we will study these two weeks

- Features: new variables issued from the data that are more expressive to solve the problem



What we will study these two weeks

- Features are specific to the type of data / problem
  - Vegetation → features related to reflectivity of vegetation → vegetation indices from NIR bands
  - Urban → features relative to the shape of objects → spatial context in visible bands
  - Clouds → features relative to thermal reflectivity → TIR bands



In the next two weeks

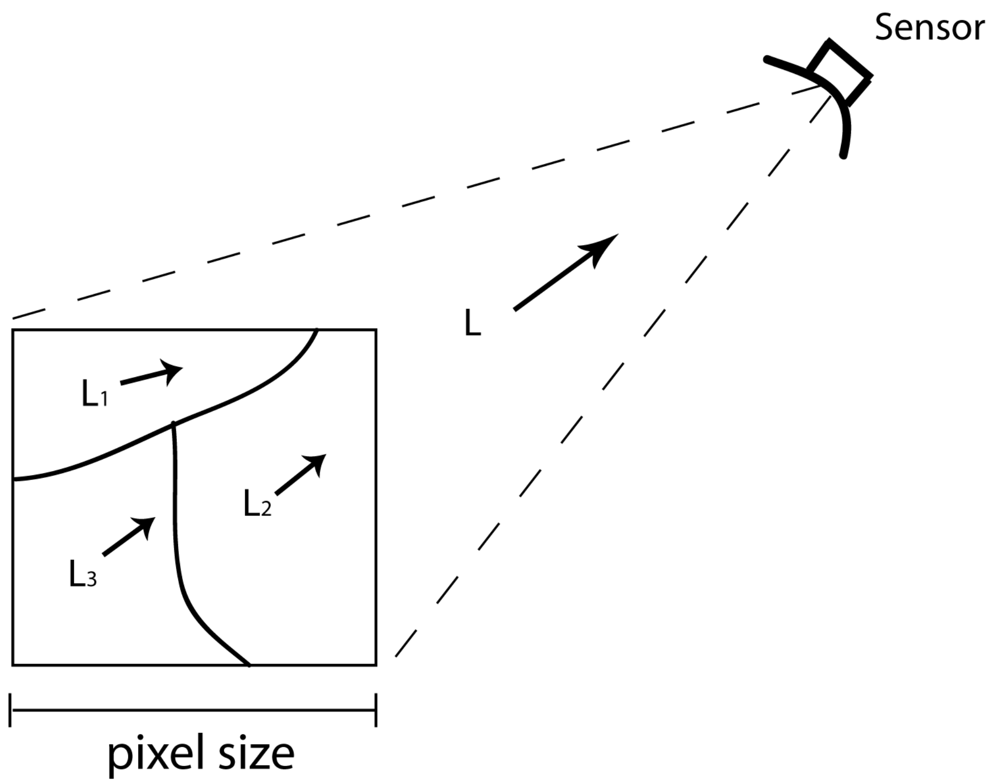
- We will study a number of information (feature) extraction techniques
  - spectral indices: enhance spectral relations between the bands of a pixel
  - spatial indices: extract information about spatial relationships
- We will also discuss how to deal with the increase in number of variables and see some data reduction techniques

Images, pixels and histograms

So that we all speak the same language

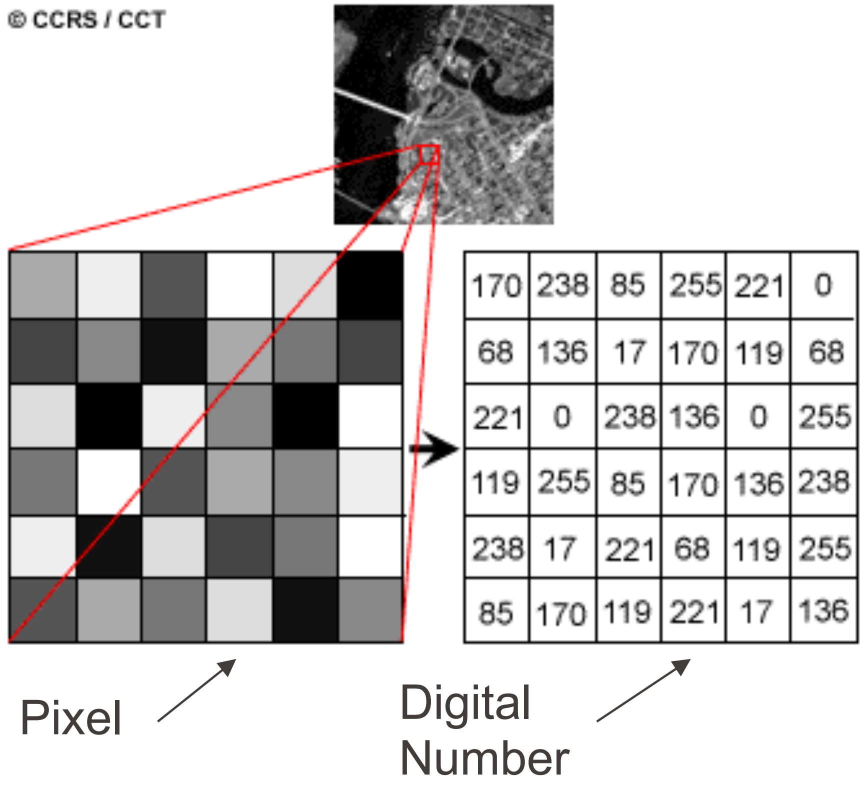
Pixels

- A pixel is the smallest surface unit in an image (a cell of the matrix)
- Pixel = picture element
- A pixel is the integration of the reflectance of all the materials



Pixels

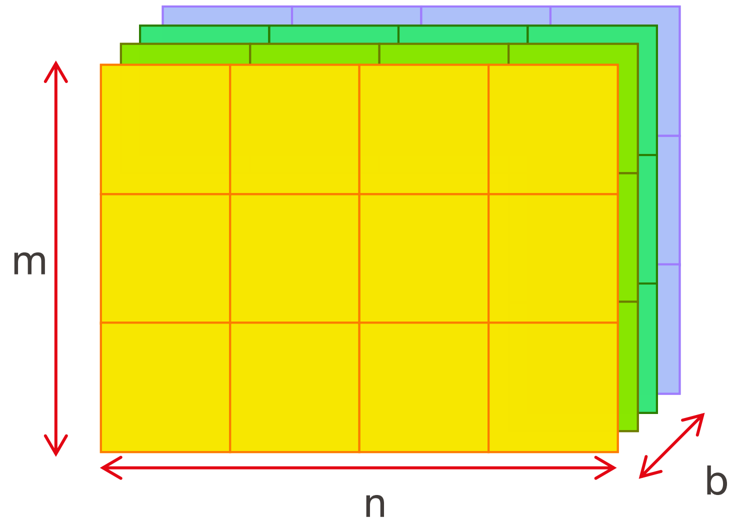
- A pixel is the integration of the reflectance of all the materials
- It receives an integer value, called Digital Number (DN)
- The maximal value depends on the radiometric resolution
  - 8-bits: 256 (range=[0, 255])
  - 12-bits: 4096 (range=[0, 4095])
  - 16-bits: 65536 (range=[0, 65535])



170	238	85	255	221	0
68	136	17	170	119	68
221	0	238	136	0	255
119	255	85	170	136	238
238	17	221	68	119	255
85	170	119	221	17	136

Images and pixels

- An image can be summarized as a 3 dimensional matrix
  - m lines
  - *n* columns
  - *b* bands



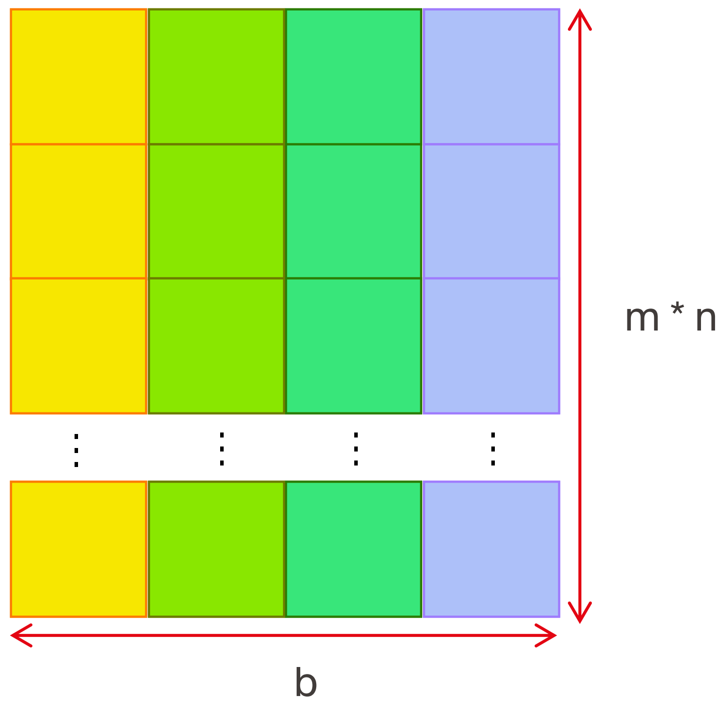
- As a convention, we'll refer to a pixel as $x_{ij}^{(k)}$, with

$$i \in 1,\cdots,m$$
$$j \in 1,\cdots,n$$
$$k \in 1,\cdots,b$$

Images and pixels

- An alternative visualisation is to convert it to a two-dimensional matrix
  - (m * n) lines
  - *b* columns

- Like this, the image cannot be visualized
(it will be useful when performing classification)



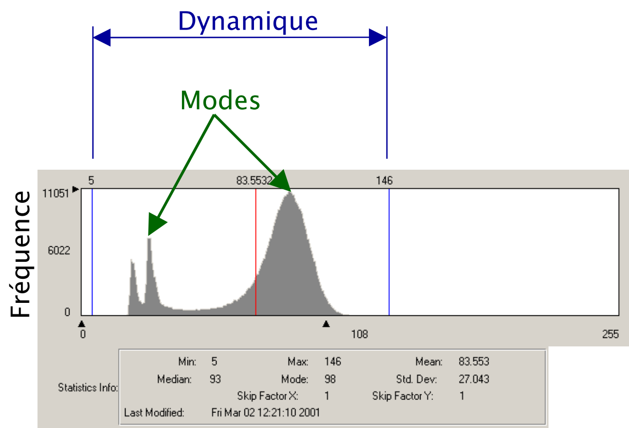
EPFL

Image bands

D. Tuia. ECEO

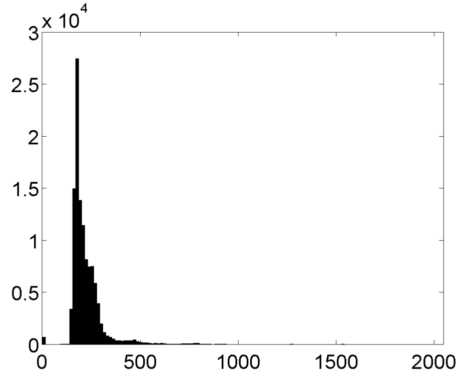
- A band is a suite of numerical values (DN)

- It can be approached as a discrete distribution with
  - Mean
  - Mode
  - Median
  - Standard deviation
  - ...



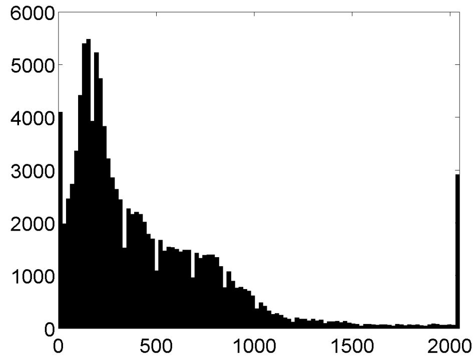
Histogram enhancements

- Sometimes, information is not well visible in the image, as it is
  - Blurry
  - Low illumination
  - Sensor failures
  - Hazy
  - ...



RGB composition

Histograms of band 1 (blue)



- One can correct for these effects by performing histogram modifications
- Not in this course

  (but in remote sensing, ENV-341)

Spectral indices

How to extract relevant across-bands information about one pixel

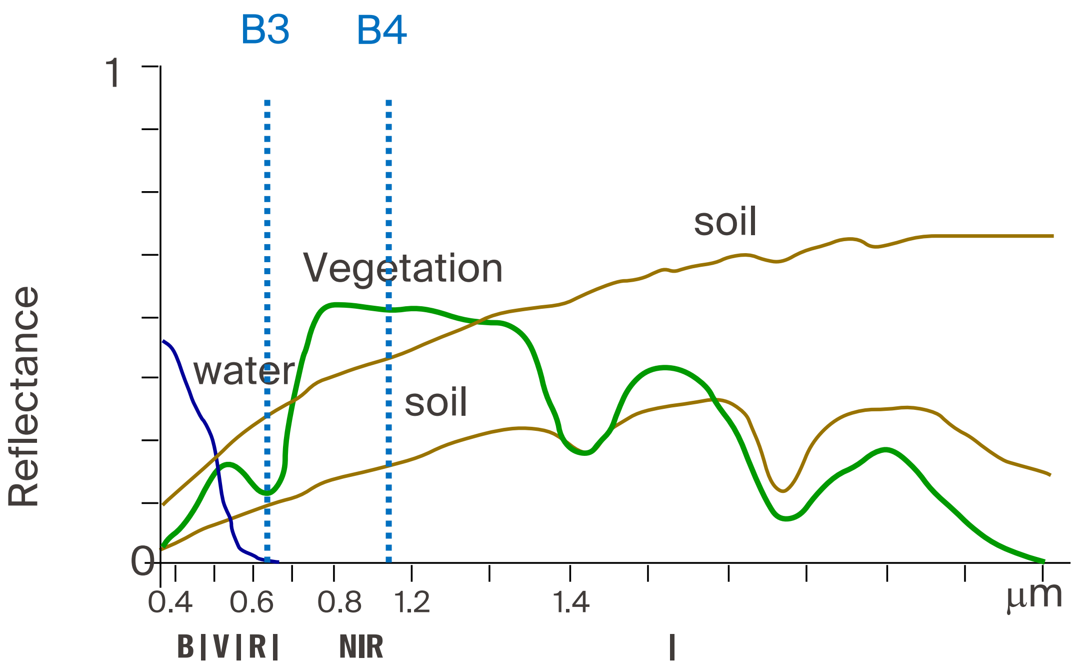
Spectral indices

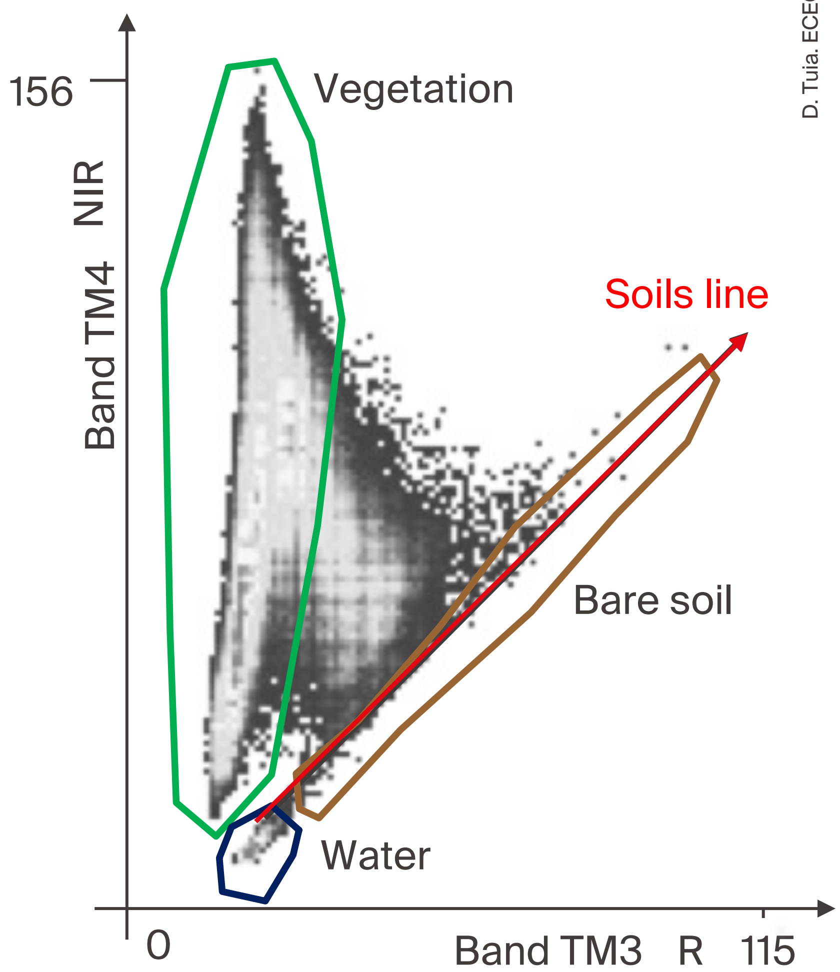
- With spectral indices we extract one character of interest about one pixel
- It is usually a complementing information with respect to the bands
- It usually relates to the final problem under study

- Basic spectral indices are often based on band ratios
- We take the same pixel in two bands and divide their values

$$SI = \frac{x_i^{(j)}}{x_i^{(k)}}$$

Physical bases on VIs

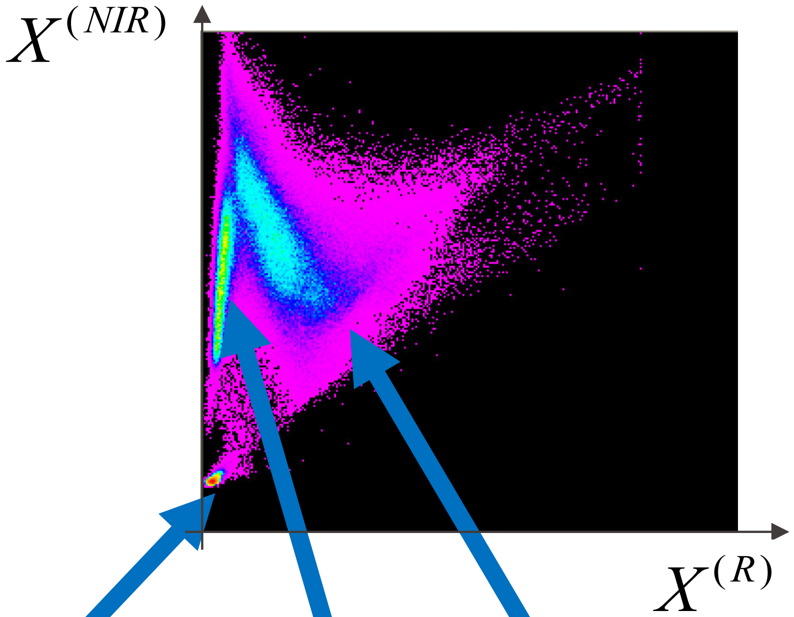


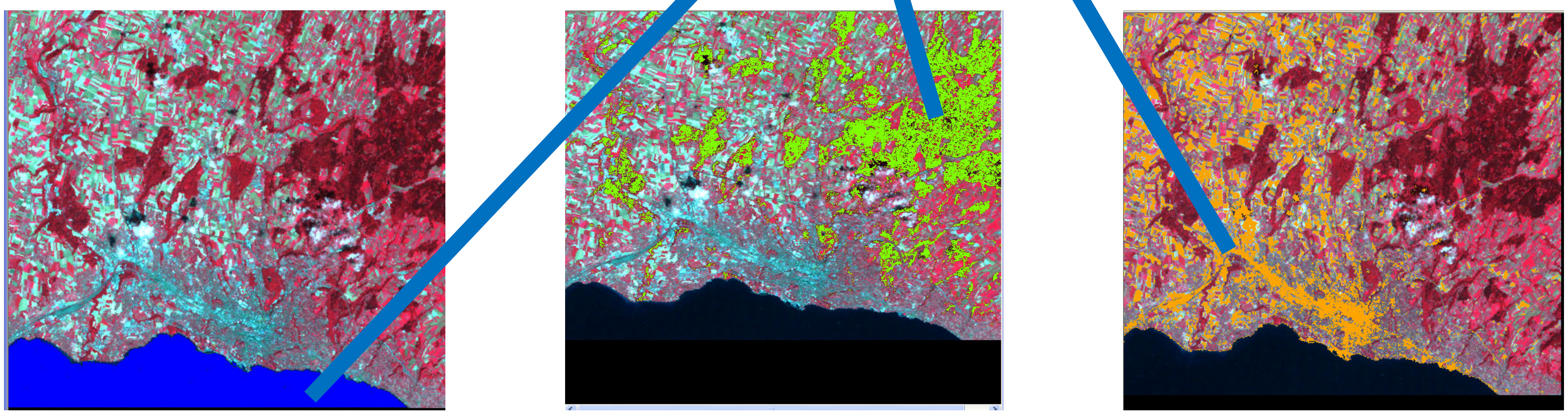


- Red and NIR bands have a strong discriminative power for vegetation
- Soils have an increasing signature w.r.t λ

Spectrogram of Lausanne

- Vegetation is on the leftside (low R, high NIR)
- The soils agglomerate around the diagonal
- Water is near the origin (low reflectance in R & NIR)

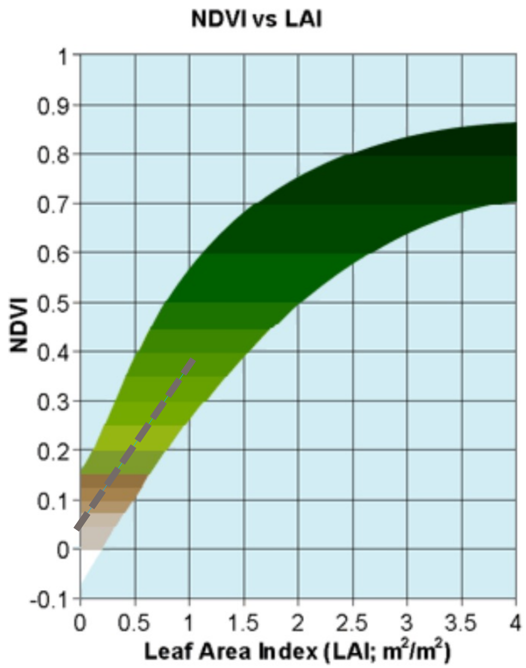




Vegetation indices

- We would like the index calculated by field measurements!
- L.A.I (Leaf area index)
  - Total area of leaves contained in a given surface for a same height (Saint, 1979) => leaf area / total ground area
  - Need to be measured on the field:

- VIs only approximate the L.A.I
  - To have the L.A.I map, need to go through regression and/or physical models (Stagakis et al., 2010; Tuia et al., 2011)

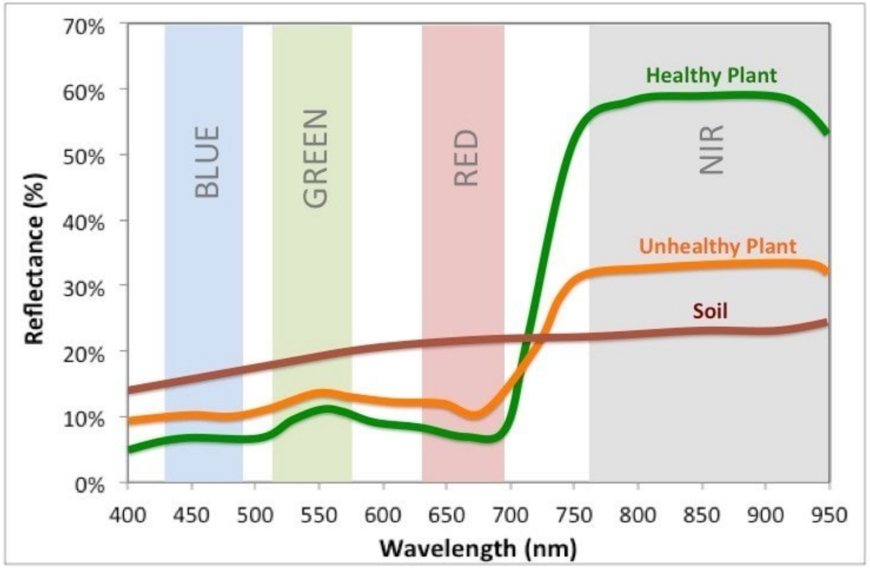


Source: http://web.pdx.edu/~nauna/resources/8-2012_lecture1-vegetationindicies.pdf

Vegetation indices: NDVI

- Many of such indices are about vegetation
- They mostly exploit the red and near infrared bands
- Vegetation is highly reflective in the NIR, other surfaces are not
- This led to the Normalized Difference Vegetation Index (NDVI)
- Theoretically between [-1;1],
- Values are more often between [-0.5;0.5]

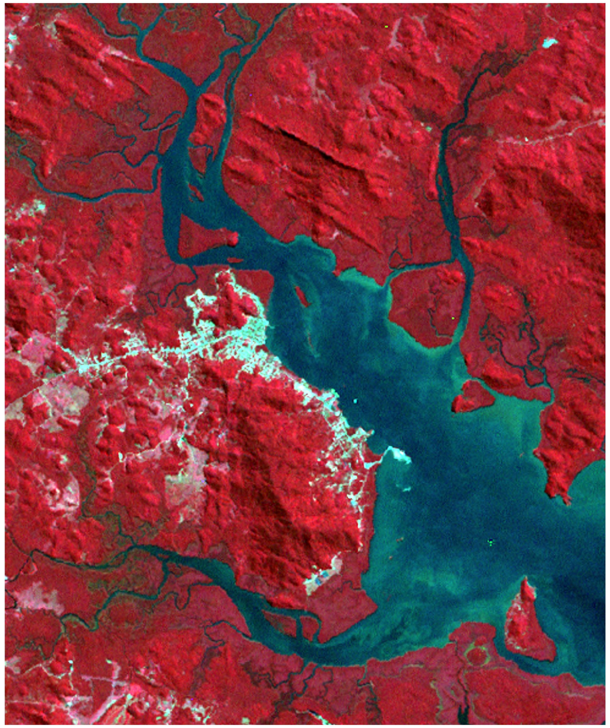
$$NDVI = \frac{x_i^{(NIR)} - x_i^{(R)}}{x_i^{(NIR)} + x_i^{(R)}}$$



This is why, for a remote sensing person, vegetation is red

(a) combinaison 321

RGB



(b) combinaison 432

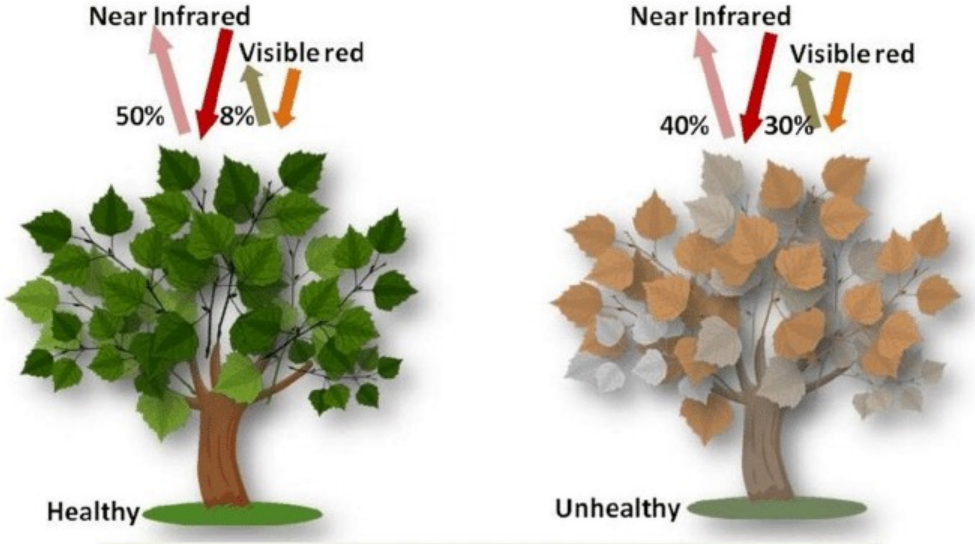
(NIR)RG

https://e-cours.univ-paris1.fr/modules/uved/envcal/html/compositions-colorees/3-differentes-composition-coloree/2-1-vraie-fausse-couleur%20.html

Vegetation indices: NDVI

- NDVI is also used for plants health

- Healthy plants show higher NDVI values than unhealthy
  - When a plant is unhealthy, it replaces chlorophyll with carotenoids.
  - With no chl, plants start reflecting more in the red wavelenghts, so NDVI decreases
  - Carotenoids absorb light in the blue as much as chl



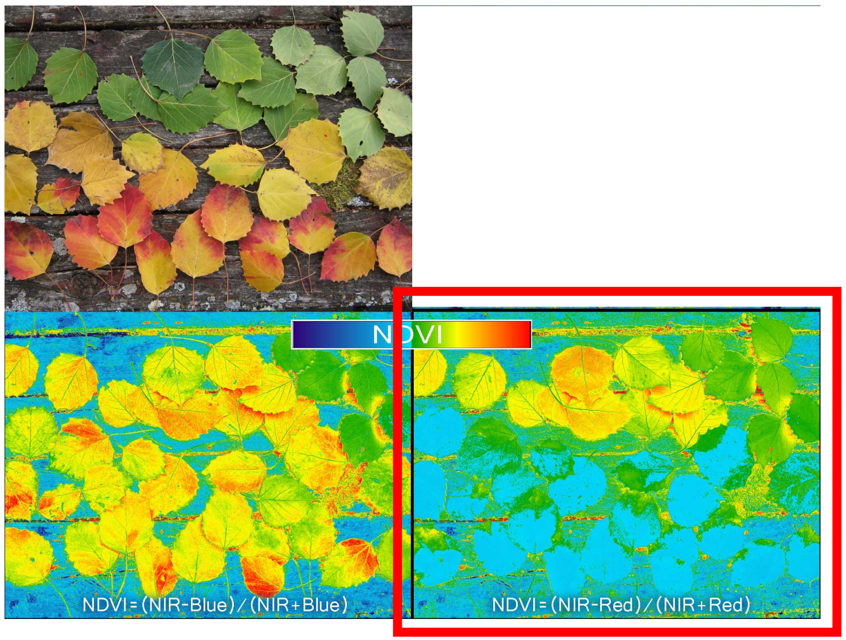
$$\frac{0.50-0.08}{0.50+0.08}=0.72 \qquad \frac{0.4-0.3}{0.4+0.3}=0.14$$

Vegetation indices

D. Tuia. ECEO

- NDVI is also used for plants health

- Healthy plants show higher NDVI values than unhealthy
  - When a plant is unhealthy, it replaces chlorophyll with carotenoids.
  - With no chl, plants start reflecting more in the red wavelenghts, so NDVI decreases
  - Carotenoids absorb light in the blue as much as chl

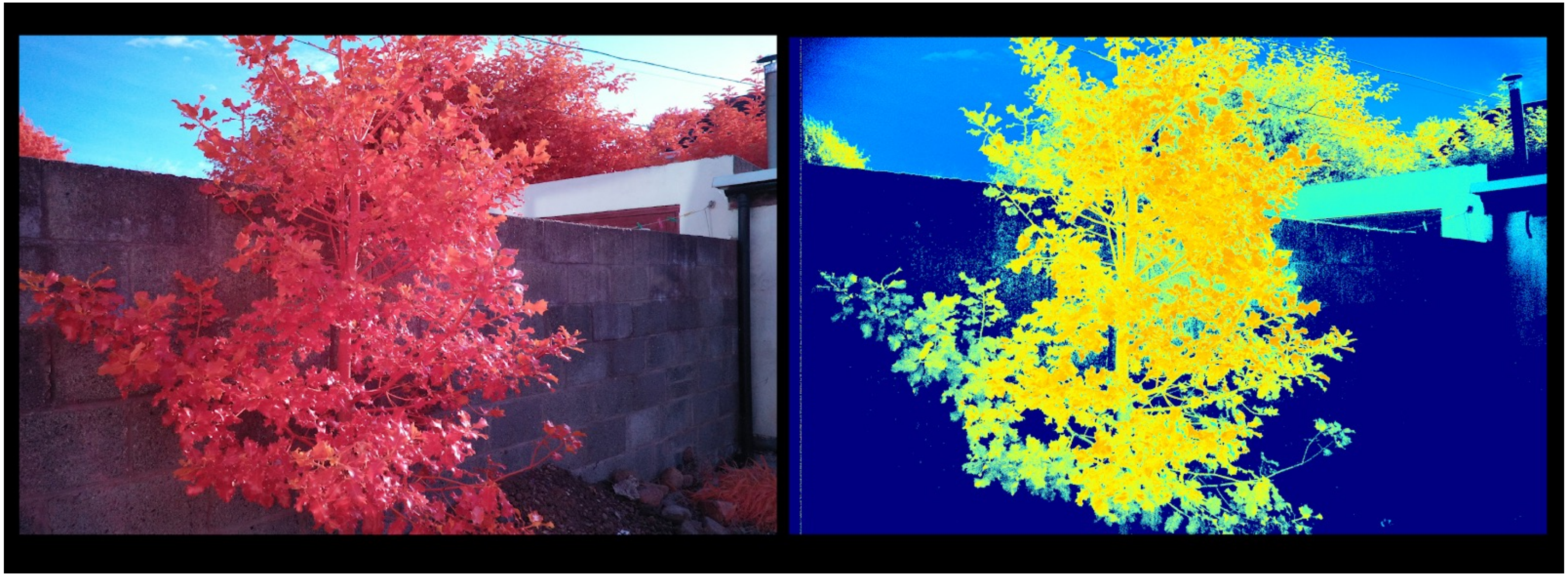


https://publiclab.org/notes/cfastie/10-11-2013/oaktober

Back to NDVI

(NIR)RG

NDVI



http://muonray.blogspot.com/2019/07/ndvi-vegetation-mapping-project-with.html

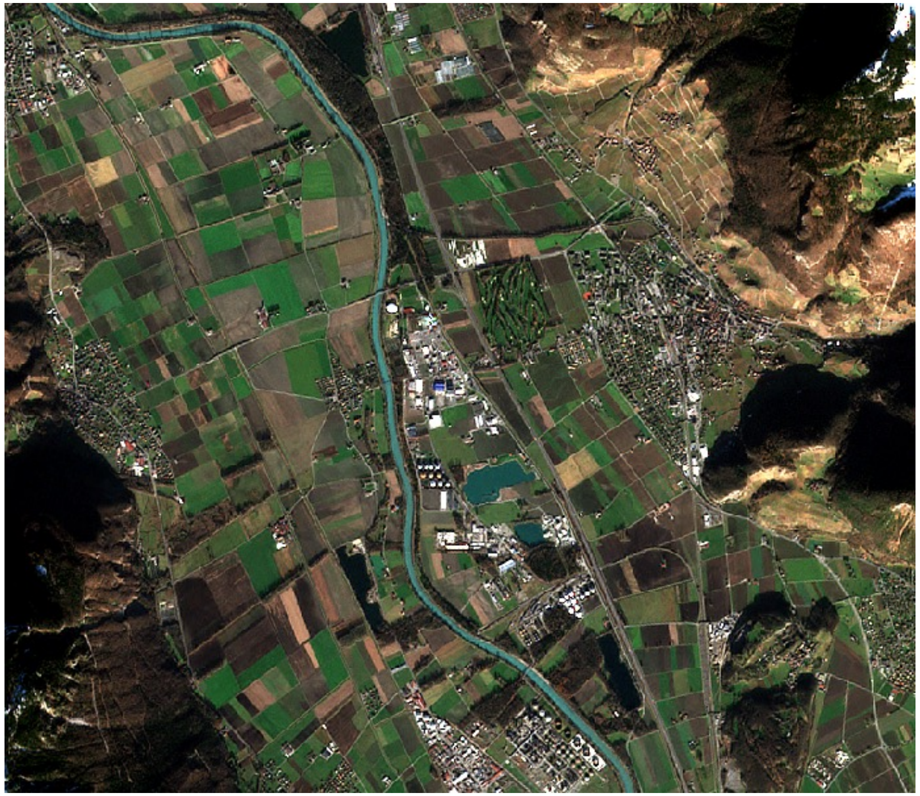
Vegetation indices: NDVI

http://www.isafarmnet.com/Replicated_Strip_Trials/Imagery/ST2013IA215_NDVI.jpg

Vegetation indices: NDVI

D. Tuia. ECEO

Sentinel-2 image (7/12/2015): RGB

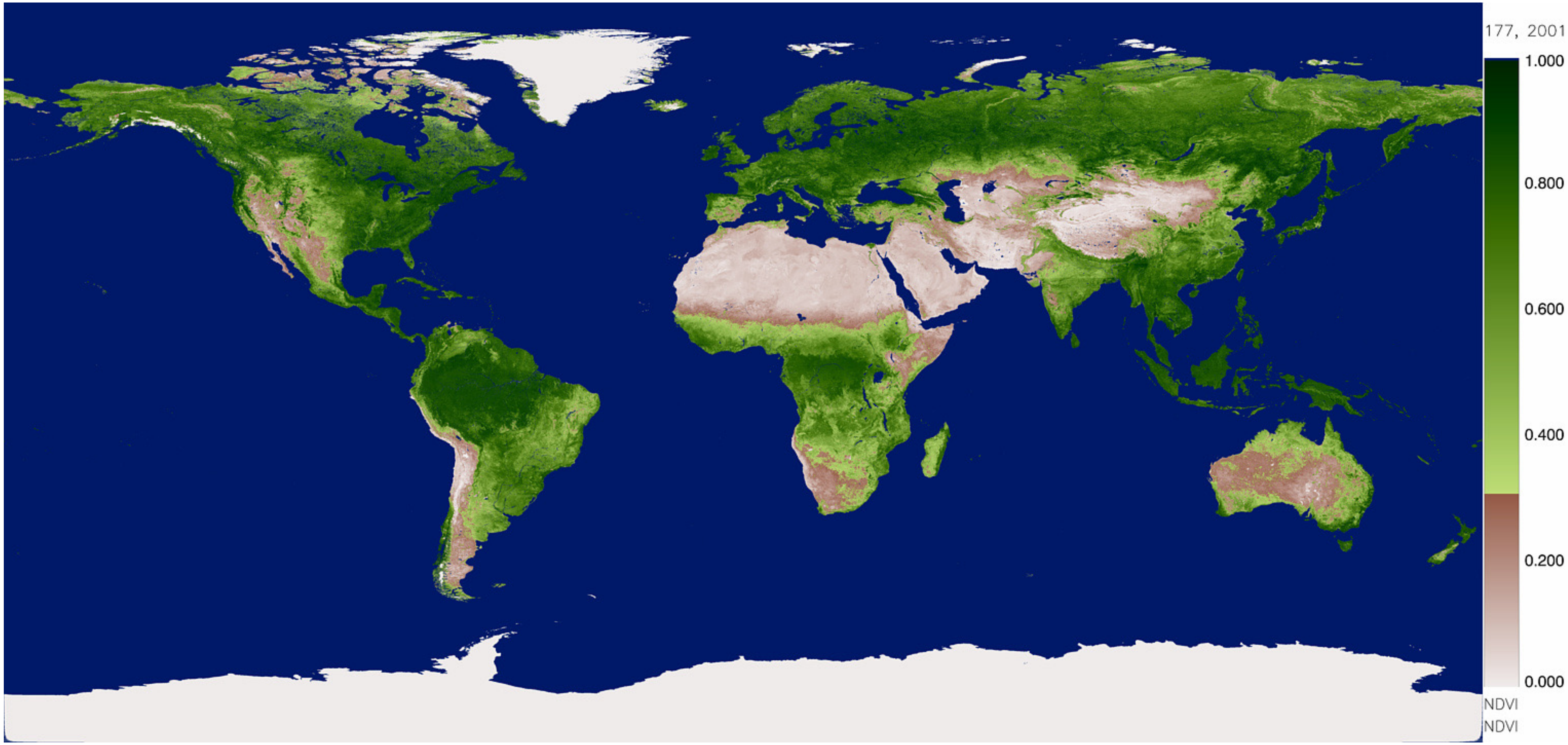


NDVI values, no stretch: -1 1

EPFL

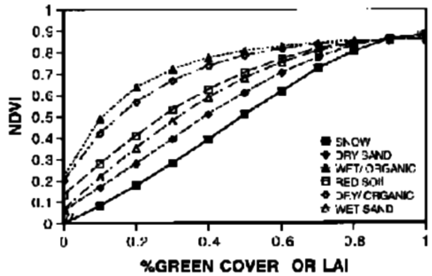
Vegetation indices: NDVI

D. Tuia. ECEO



Corrected vegetation indices: SAVI

- The Soil-Adjusted Vegetation Index (Huete, 1988) corrects NDVI by soil type
- It reduces contribution of bare soil



- L is the soil brightness correction factor

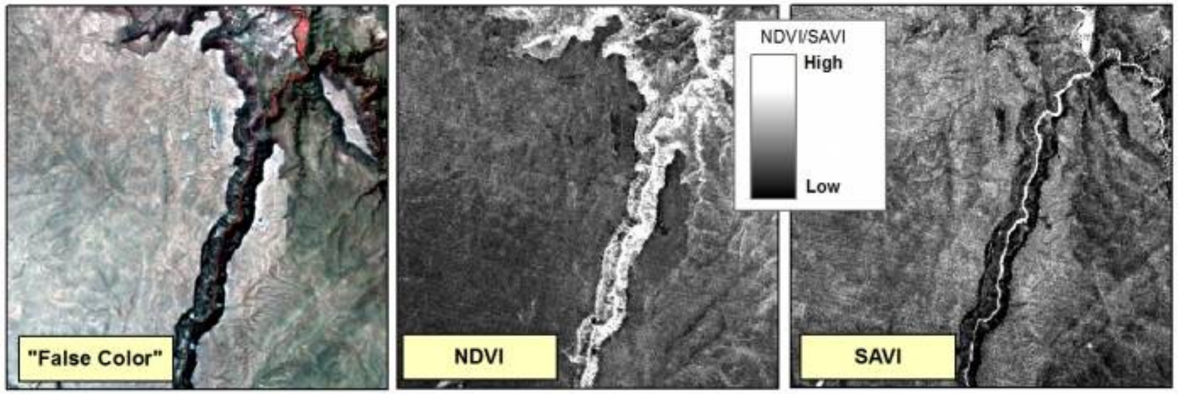
$$SAVI = \frac{x^{(NIR)} - x^{(R)}}{x^{(NIR)} + x^{(R)} + L}(1+L)$$

$$L = \begin{cases} 1 & \text{if low density vegetation} \\ 0.5 & \text{if mid-dense vegetation} \\ 0.25 & \text{if dense vegetation} \end{cases}$$

Difficult to estimate…

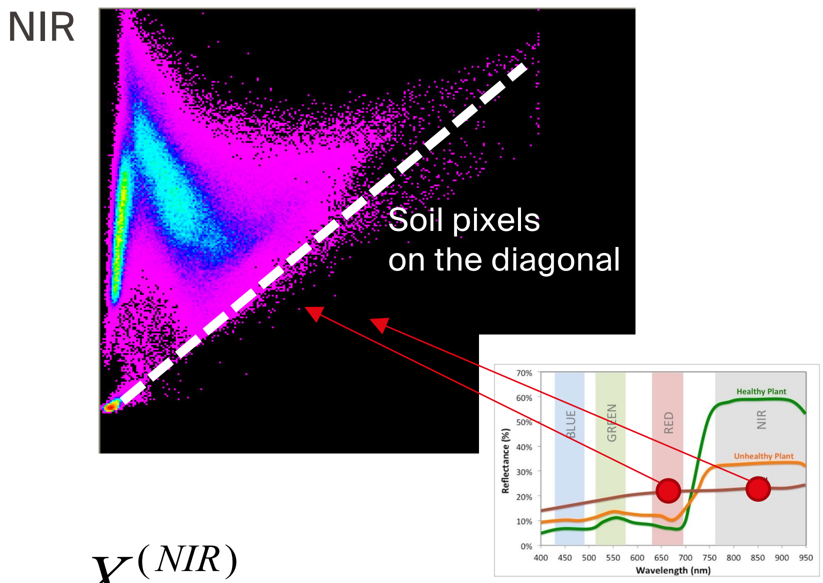
Corrected vegetation indices: SAVI

- In the center of the image, the canyon has too strong values in the NDVI, while having a low vegetation cover

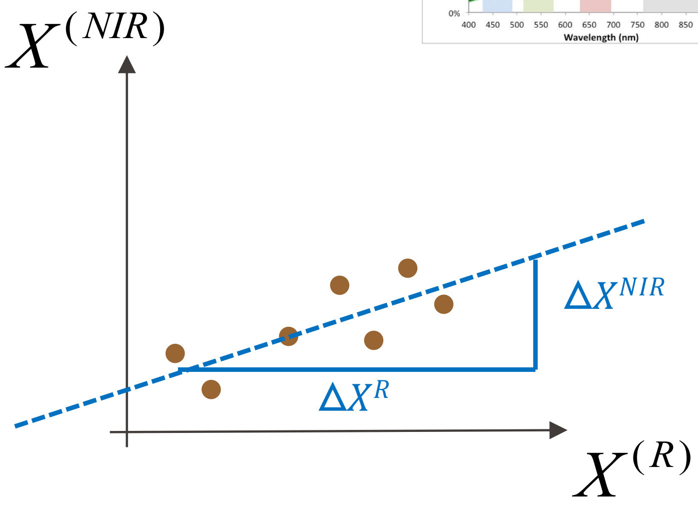


http://wiki.landscapetoolbox.org/doku.php/remote_sensing_methods:soil-adjusted_vegetation_index

Corrected vegetation indices: the soil line



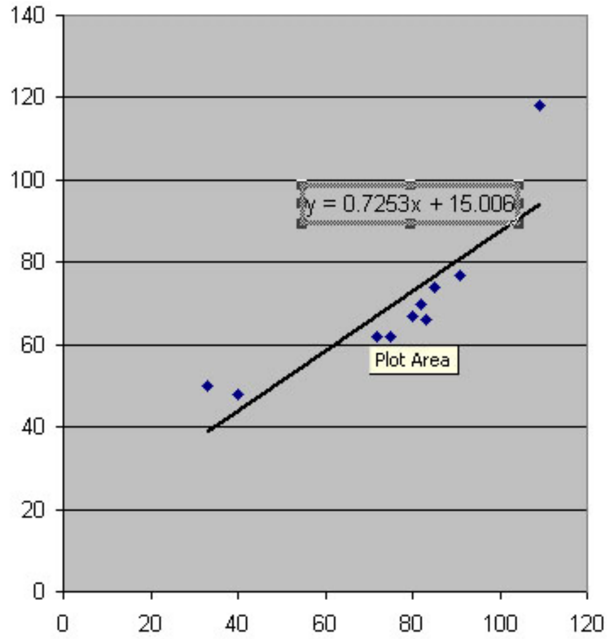
- If we were working with reflectance, the soil line would be unique.
- Since there are atmospheric effects, its position changes for each acquisition



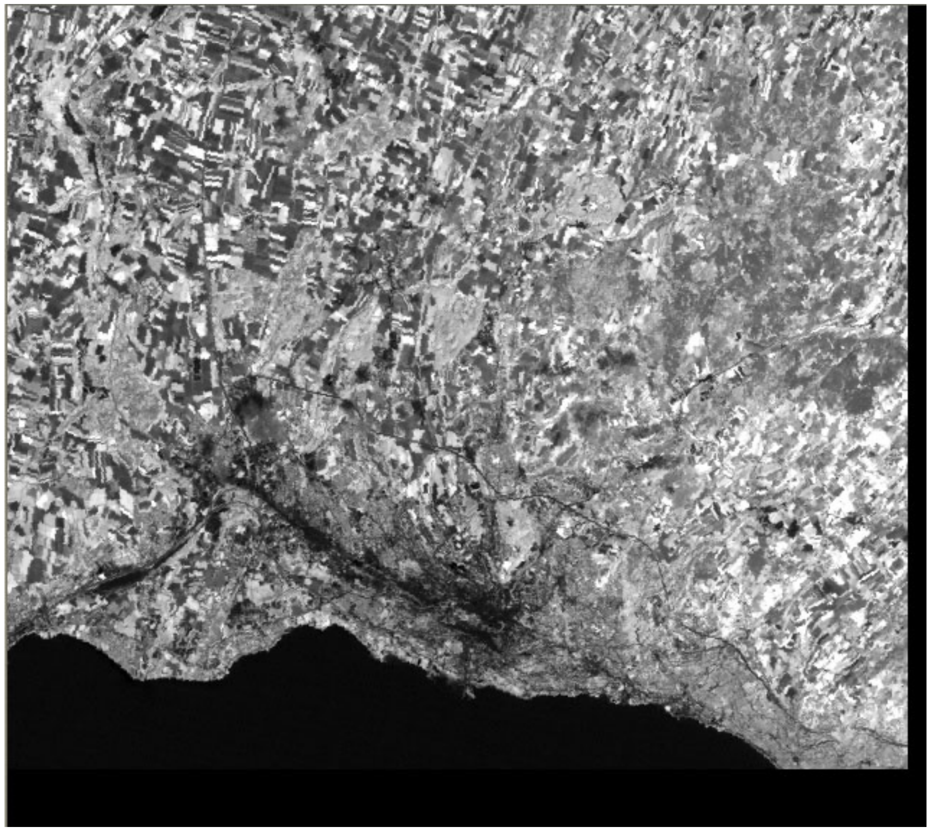
- Must be estimated by linear regression
  - Use some reference soil samples
  - Infer regression parameters *a* and *b*

$$X^{(NIR)} = aX^{(R)} + b$$

Corrected vegetation indices: the soil line



$$a = 0.7253$$



Corrected vegetation indices: MSAVI

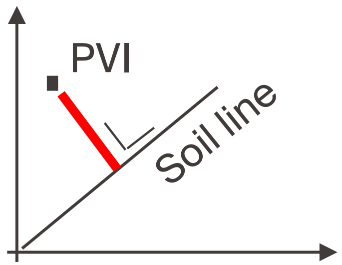
- Modified Soil-Adjusted Vegetation Index (Chehbouni, 1994)
- The L "soil brightness" correction factor is self-adjusted
- L is defined by the NIR and R bands and *a* the slope of the soil line

$$MSAVI = \frac{X^{(NIR)} - X^{(R)}}{X^{(NIR)} + X^{(R)} + L}(1 + L)$$

$$L = 1 - \frac{2 \cdot a \cdot (X^{(NIR)} - X^{(R)}) \cdot (X^{(NIR)} - a \cdot X^{(R)})}{X^{(NIR)} + X^{(R)}}$$

EPFL

Corrected vegetation indices: PVI



- Perpendicular Vegetation Index (Richardson et al., 1977)

$$PVI = \frac{X^{(NIR)} \overbrace{- a}^{\text{Slope}} \cdot X^{(R)} \overbrace{- b}^{\text{Offset}}}{(\sqrt{a^2 + 1})}$$

- Exploits directly the soil line
- The further a pixel is from the soil line, the more densely vegetated

$$\begin{cases} > 0 & \text{for vegetated pixels} \\ = 0 & \text{for bare soil} \\ \leq 0 & \text{for water with low Chl - a} \end{cases}$$

- Its quality depends on the soil line adjustment
(*ax+b* is only an approximation!)

One index to rule them all?

- NDVI is by far the most used index
- In the vegetation indices world there are A LOT of indices…
- E.g. indices based on red – NIR bands (and equivalents for Sentinel 2)

Vegetation index	Formulation	S-2 bands used	Original author
NDVI	(NIR – R)/(NIR + R)	(B7 – B4)/(B7 + B4)	Rouse et al. (1973)
NDI45	(NIR – R)/(NIR + R)	(B5 – B4)/(B5 + B4)	Delegido et al. (2011b)
MTCI	(NIR – RE)/(RE – R)	(B6 – B5)/(B5 – B4)	Dash and Curran (2004)
MCARI	[(RE – R) – 0.2(RE – G)] * (RE – R)	[(B5 – B4) – 0.2(B5 – B3)] * (B5 – B4)	Daughtry et al. (2000)
GNDVI	(NIR – G)/(NIR + G)	(B7 – B3)/(B7 + B3)	Gitelson et al. (1996)
$PSSR_a$	NIR/R	B7/B4	Blackburn (1998)
S2REP	705 + 35 * ((((NIR + R)/2) – RE1)/(RE2 – RE1))	705 + 35 * ((((B7 + B4)/2) – B5)/(B6 – B5))	
IRECI	(NIR – R)/(RE1/RE2)	(B7 – B4)/(B5/B6)	

One index to rule them all?

- NDVI is by far the most used index
- In the vegetation indices world there are A LOT of indices…
- With hyperspectral data we can access to continuous 10nm bands. Then we can design indices for specific vegetation properties:
  - Leaf pigment specific (Gitelson et al 2001, Zarco-Tejada et al. 2005)
  - Vegetation stress (Zarco-Tejada et al. 2003, Merzlyak et al., 1999)
  - Chlorophyll absorption (Haboudane et al., 2002)

Spectral indices for other properties

- Normalized Difference Water index:
  - NIR & SWIR bands (Landsat 8 bands 5 & 6)
  - Plant water content (liquid & in spongy mesophyll), relates to plant water stress
  - Modified Normalized Difference Water Index: G instead of NIR (better on open water)
  - Normalized Multi-band Drought Index: SWIR is replaced by (SWIR1 – SWIR2)

$$NDWI = \frac{X^{(NIR)} - X^{(SWIR)}}{X^{(NIR)} + X^{(SWIR)}}$$

- Normalized Difference Built-up Index: very similar
  - Built-up Index: NDWI – NDVI
  - Applications: Watershed run-off predictions, land use planning

- Normalized Burn Ratio Index: same bands
  - Normalized Burn Ratio Thermal adds the Thermal band (10.4-12.5 μm)

in °Kelvins

$$NBRT = \frac{X^{(NIR)} - X^{(SWIR)} \cdot \frac{X^{(TIR)}}{1000}}{X^{(NIR)} + X^{(SWIR)} \cdot \frac{X^{(TIR)}}{1000}}$$

Spectral indices for other properties

- Clay Minerals Ratio: SWIR1 / SWIR2
  - SWIR 1: 1.55-1.75 µm SWIR2: 2.08-2.35 µm (Landsat 8 bands 6 & 7)
- Ferrous Minerals Ratio: SWIR1 / NIR
  - Landsat 8 bands 6 & 5

http://www.harrisgeospatial.com/docs/BackgroundGeologyIndices.html

- Normalized Difference Snow Index:
  - Discriminates well snow from clouds
  - Snow absorbs light in the SWIR range

$$NDSI = \frac{X^{(SWIR)} - X^{(R)}}{X^{(SWIR)} + X^{(R)}}$$

RGB composition

Cloud mask

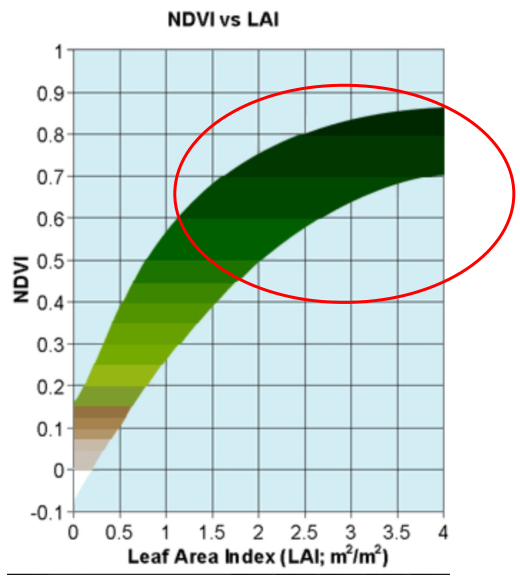
RGB composition

Cloud mask

Cloud-Screening Algorithm for ENVISAT/MERIS Multispectral Images, Gomez-Chova et al. 2007

Final thoughts

- Spectral indices are important if you study vegetation
- There are many of those! NDVI remains the most used, but it has its limitations
- And NDVI is NOT biomass. It saturates, remember!



- Further reading

https://advances.sciencemag.org/content/7/9/eabc7447

RESEARCH ARTICLE | ENVIRONMENTAL STUDIES

A unified vegetation index for quantifying the terrestrial biosphere

Gustau Camps-Valls[1,*], Manuel Campos-Taberner[2], Álvaro Moreno-Martínez[1,3], Sophia Walther[4], Grégory D...
+ See all authors and affiliations

Science Advances 26 Feb 2021:
Vol. 7, no. 9, eabc7447
DOI: 10.1126/sciadv.abc7447