Supported by:

Federal Ministry for the Environment, Nature Conservation and Nuclear Safety

Federal Agency for Nature Conservation

EOCap 4Africa

EOCap4Africa

5a Remote Sensing Data: Preprocessing

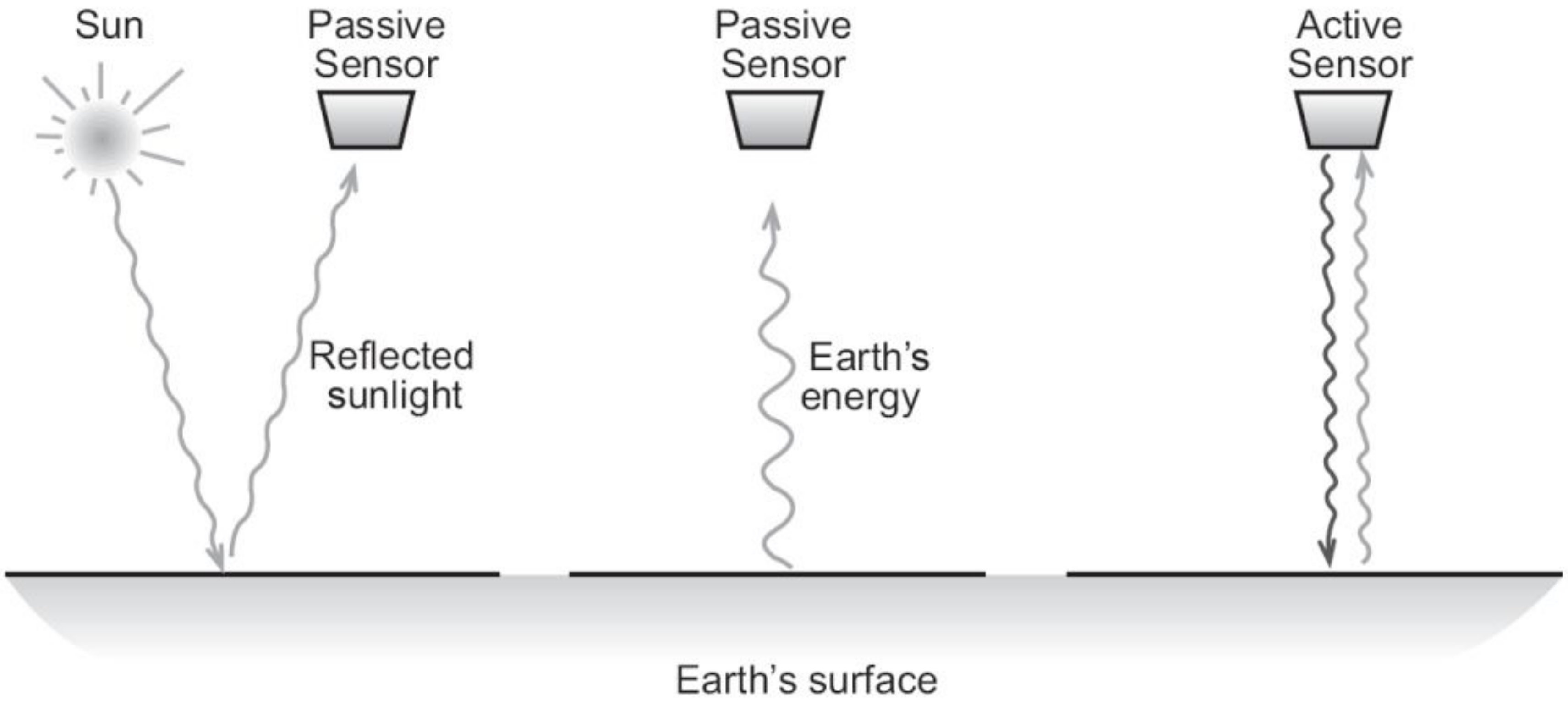
Julius-Maximilians-Universität Würzburg · DLR · Martin-Luther-Universität Halle-Wittenberg · University of Botswana · INES Ruhengeri Institute of Applied Sciences

EARTH OBSERVATION RESEARCH

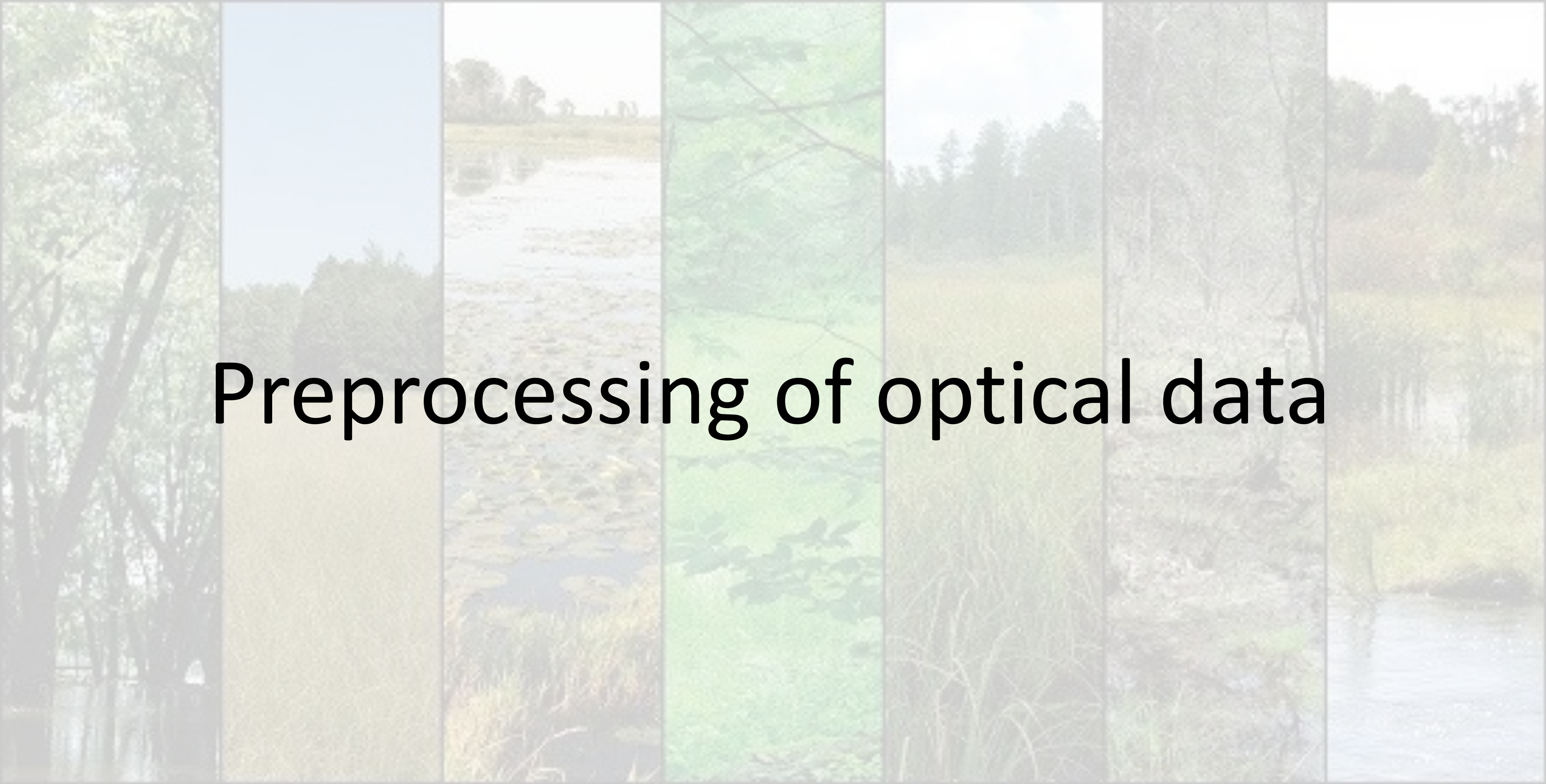
Recap

Active vs. passive remote sensing

- Active sensor: emits EM energy and detects the energy returning from the object or surface
- **Passive sensors: measure solar or terrestrial energy**



A remote sensor measures reflected or emitted energy. An active sensor has its own source of energy.



Preprocessing of optical data

The purpose of preprocessing satellite imagery

- Raw satellite imagery inherits systematic error of various kinds
- These systematic errors depend on the acquisition technique
- These ‘unwanted’ effects can be of radiometric or spatial nature, meaning that the value that was recorded, does not relate to the actual ground observation or that the pixel we look at is distorted in some way
- Using this as a basis for any kind of scientific analysis propagates errors, which can lead to misclassifications and wrong conclusions
- Consequently, solely completely calibrated data allows to detect and precisely quantify trends of changes in land cover

Factors influencing the RS image acquisition

- Sensor characteristics (radiometry)
- Weather, particular cloud cover
- Earth surface (geometry)
- Atmosphere
- Acquisition method: satellite or airborne
- Others

However, RS images should be comparable:

  - In time (e.g., monitoring)
  - Between sensors (e.g., MODIS and Landsat TM)

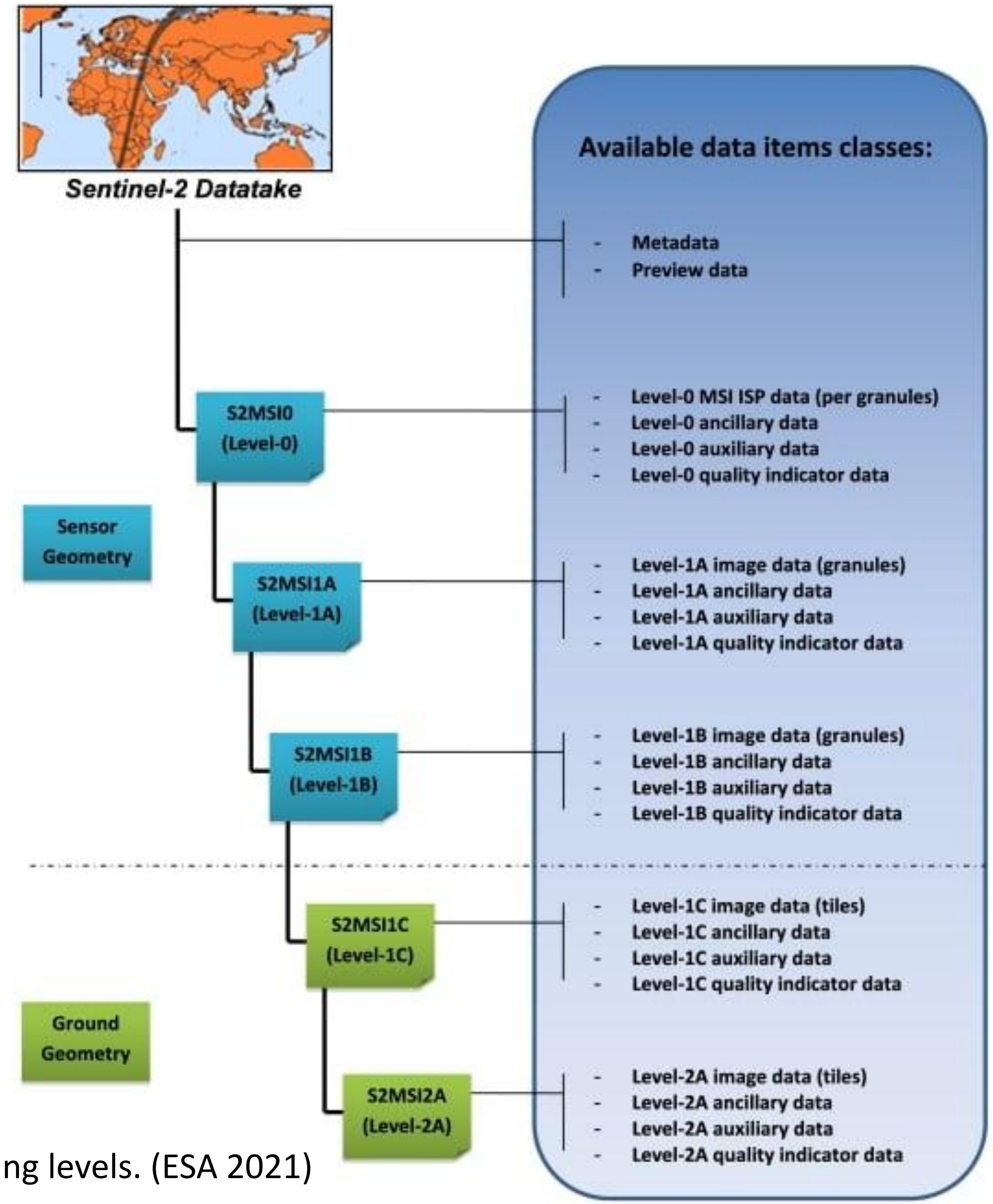
Preprocessing of optical data

Processing Levels

- With optical data sets, we can divide the data into what is called ‘processing levels’
- This term describes how many different calibration and registration algorithms were performed on a certain data set
- On the next slide, you can see the breakdown of the structure of Sentinel-2 data that is created by the European Space Agency (ESA). The structure is exemplary for other commonly used optical satellites

Preprocessing of optical data

Processing Levels



Sentinel-2 Processing levels. (ESA 2021)

Preprocessing of optical data

Processing Levels – General Systematic:

- Level 0: Raw data without any calibration or correction (not of interest for us)
- Level 1A: Raw data bundled with radiometric and geometric calibration and correction coefficients as well as geolocation information, where radiometric correction deal with internal anomalies in detector sensitivity or detector noise and geometric corrections deal with image distortions, e.g., due to the optics of the sensor or the rotation of the Earth
- Level 1B: Radiometric and geometric calibration and corrections from Level 1A have been applied to the raw data. Each pixel now corresponds to an explicit location on Earth – it is georeferenced. This is the minimum processing level you should strive for

Preprocessing of optical data

Processing Levels – General Systematic:

- Level 2: Geophysical parameters, such as reflectance at the Earth's surface, land surface temperature or soil moisture, have been completed. The calculation of these parameters requires the application of various transformations and corrections, including, e.g., corrections for atmospheric scattering or surface properties

- Level 3: Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.

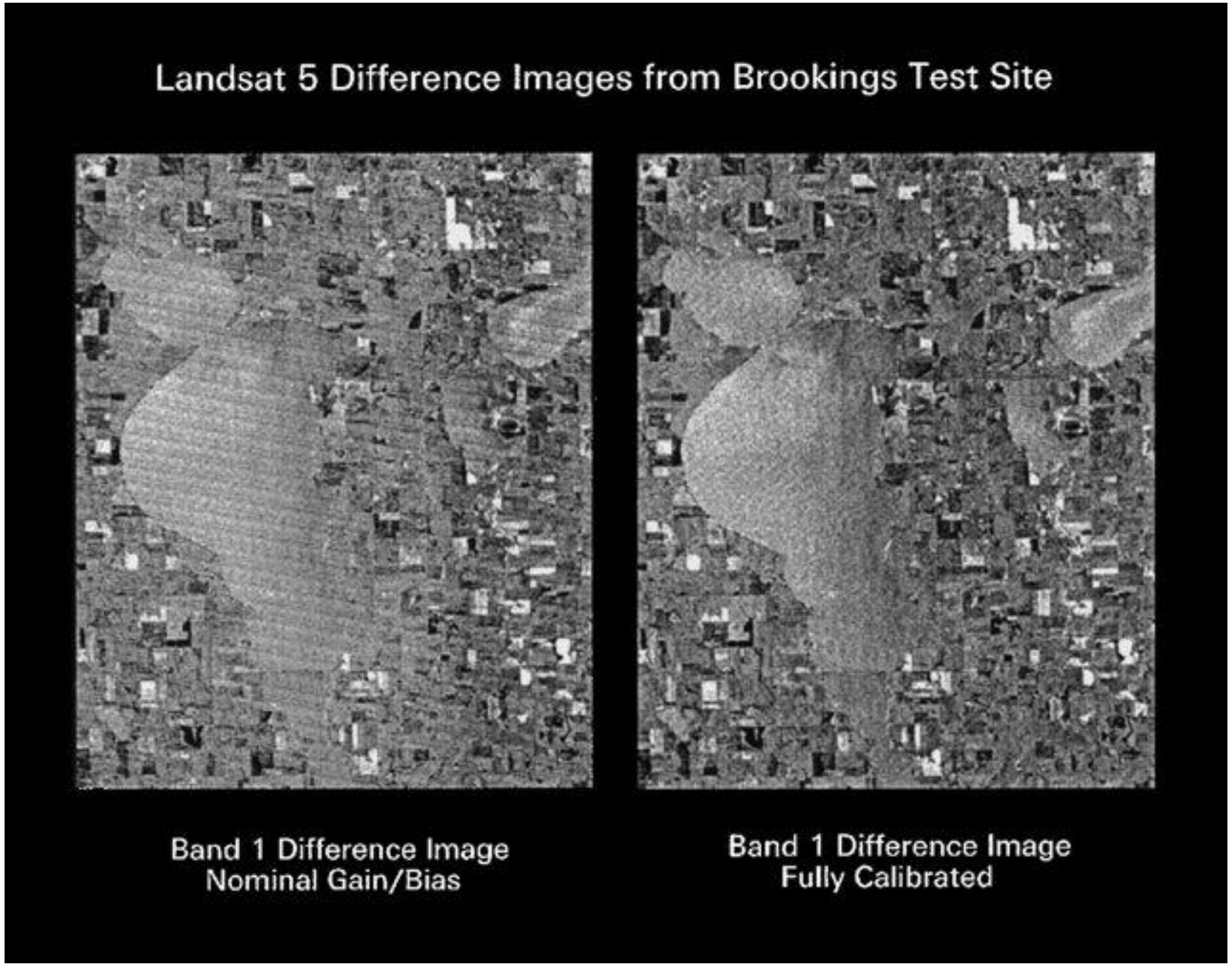
- Level 3A: L3A data are generally periodic summaries (weekly, 10-days, monthly) of L2 products

- Level 4: Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements)

Striping

- Due to non-identical detector response
    - Detector characteristics
    - Change with time / rise of temperature
    - failure
- Various methods (sometimes used in combination)
    - Look up tables (radiometric response measurements at different brightness levels)
    - Onboard calibration
    - Histogram matching (gain and offset) – line pattern

Striping



Comparison between Landsat 5 TM Band 1 difference images with different levels of calibrations and corrections applied. Bright areas in the lakes represent the point of no change. Most values are between - 50 (dark, Band 1 drop between 1997 and 1998 images) and + 20 digital numbers (DN) (bright, Band 1 gain between 1997 and 1998).
(Vogelmann et al. 2001)

(Partially) missing lines

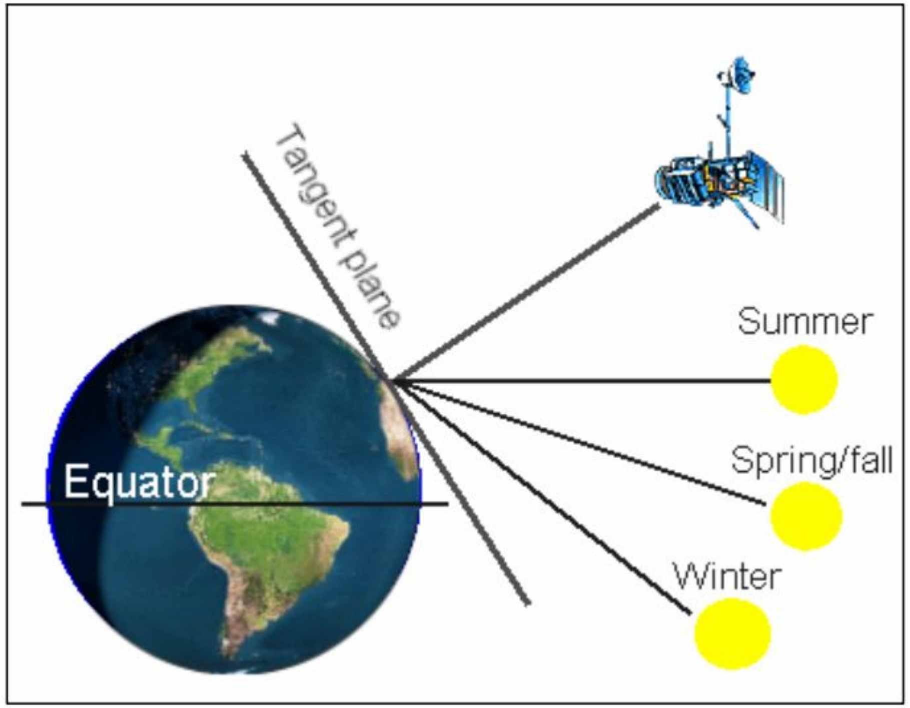
- Errors in
  - Sampling or scanning equipment
  - Transmission or recording of image data
  - Reproduction of the media containing the data

- Two methods
  - Interpolations using data from adjacent scan lines
  - Interpolation data at the same scan line from different spectral bands

Band gaps in Landsat 7
(NASA, 2025)

Solar angle correction

- Position of the sun relative to the earth changes depending on time of the day and the day of the year
- In the northern hemisphere the solar elevation angle is smaller in winter than in summer



Sun elevation difference per season
(adapted from Lillesand and Kiefer)

Solar angle correction

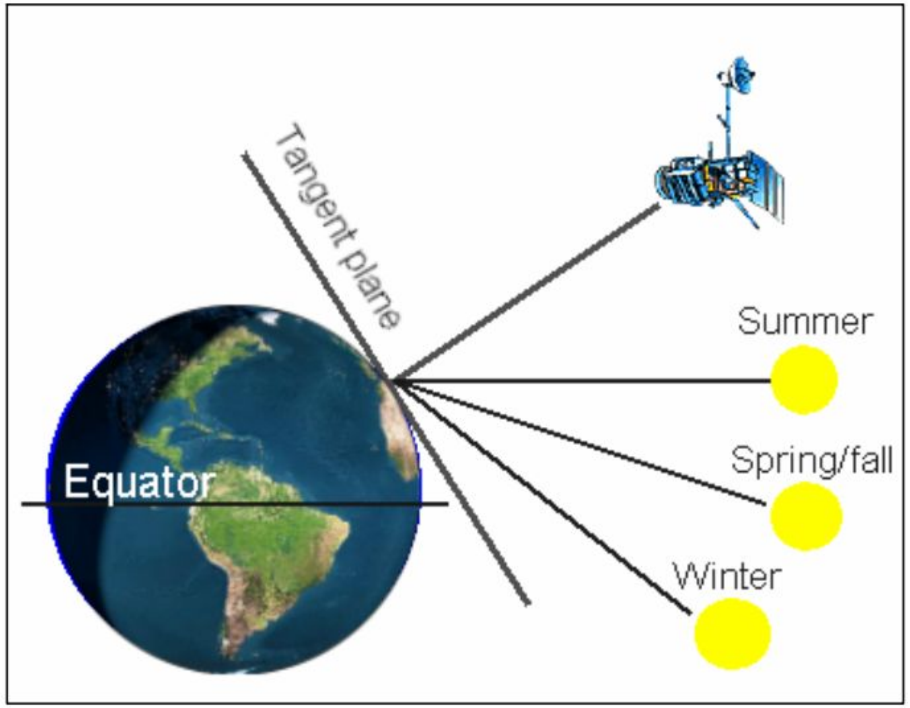
- An absolute correction involves dividing the DN-value in the image data by the sine of the solar elevation angle
- Size of the angle is given in the header of the image data

$$DN_{corr} = \frac{DN}{\sin \alpha}$$

Solar irradiation correction

Scene illumination

- Position of sun
  - Sun elevation (sun-angle)
  - Sun- earth distance
- Correction elevation
  - Normalization
  - Division of each pixel value by the sine of solar elevation angle for particular time and location per spectral band
- Correction distance
  - Sun irradiance decreases with square of distance
  - normalization

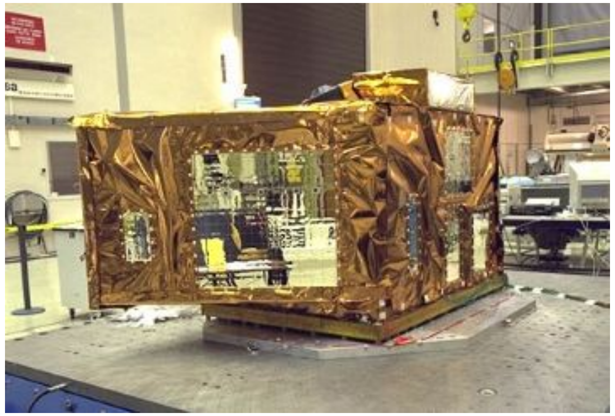


Sun elevation difference per season
(adapted from Lillesand and Kiefer)

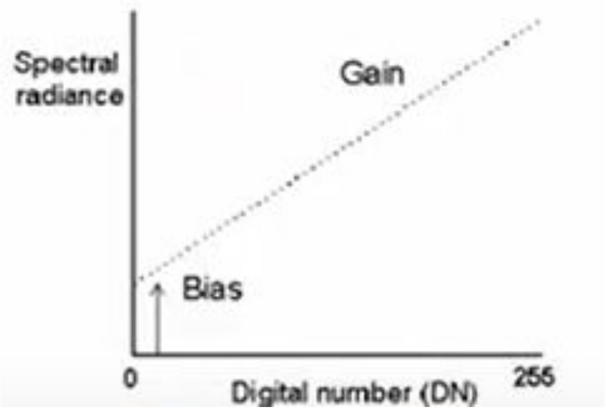
Radiometric calibration

Sensor calibration before launch:

- From DN to physical units
- Pre-launch calibration
- Correction factors for all channels:
  - Bias (A_0)
  - Gain (A_1)
- $L = A_0 + A_1 * DN$
  - L = spectral radiance
  - DN = digital number



MERIS pre-launch calibration in lab. (ESA 2025)



Response function for a TM band

Sensor calibration

- Necessary to generate absolute data on physical properties
  - Reflectance
  - Temperature
  - Emissivity
  - Backscatter
- Values provided by data provider / agency

Value conversion

- The conversion from DN to radiance and reflectance is an important step in the preprocessing
- In case of Sentinel-2 and Landsat data, we are usually interested in the bottom-of-atmosphere (BOA) reflectance, which is, in contrast to the top-of-atmosphere (TOA) reflectance, corrected for atmospheric impacts and can thus directly be interpreted
- However, the initially available data sets are being converted from DN to TOA format
- For the conversion, the data is rescaled using sensor specific equations containing information about sun-related angles and other parameters

Value conversion

- An example for such a conversion, the expression used to correct Landsat 8 OLI data is given below:

$$\rho\lambda = \frac{M_\rho Q_{cal} + A_\rho}{\cos\theta_{SZ}}$$

$\rho\lambda$ = TOA reflectance

M_ρ = multiplicative rescaling factor

Q_{cal} = DN (quantized & calibrated)

A_ρ = additive rescaling factor

θ_{SZ} = local solar zenith angle, $\theta_{SZ} = 90° - \theta_{SE}$ = local sun elevation angle

- By using these equations, we can compute TOA values which are subject to strong illumination geometry difference-related effects and therefore should not be favored over BOA reflectances

Terrain effects

- Cause differential solar illumination
  - Some slopes receive more sunlight than others

- Magnitude of reflected radiance reaching the sensor
  - Topographic slope and aspect
  - Bidirectional reflectance distribution function (BRDF)

Terrain effects

- **Minnaert correction**
  - First order correction for terrain illumination effects

$$L_n = L * \cos(e)^{k-1} * \cos(i)^k$$

L_n - normalized radiance

L - measured radiance

e - slope (derived from DEM)

i - incidence angle of solar radiation

k - Minnaert constant (estimated for each image)

Terrain effects

- **Shaded relief model (SRM)**
  - Requires digital elevation model
  - Generated with constant albedo (brightness dependent solely on topographic effects)
  - Ratio of image and SRM yields spectral radiance of ground cover (noisy)

  - Alternative:

$$DN_{corr} = m * (DN - SRM_{DN}) + a$$

Atmospheric correction – why?

- Physical relation of radiance to surface property
  - Atmospheric component needs to be removed
- Multispectral data for visual analysis
  - Scattering increases inversely with wavelength
- Image ratios
  - Leads to biased estimate
- Time difference between image acquisition and ground truth measurements

Atmosphere and radiation

- Relationship between radiance received at the sensor (above atmosphere) and radiance leaving the ground

$$L_s = H * \rho * T + L_p$$

L_s - at sensor radiance

H - total downwelling of target

ρ - reflectance of target

T - atmospheric transmittance

L_p - atmospheric path radiance (wavelength dependent)

Atmospheric correction methods

- Image-based methods
    - Histogram minimum method
    - Regression method

- Radiative transfer models

- Empirical line method

Atmospheric correction methods

Histogram minimum method

$$L_s = H \cdot \rho \cdot T + L_p$$

- Histograms of pixel values in all bands
- Pixel values of low reflectance areas near zero
  - Exposure of dark colored rocks
  - Deep shadows
  - Clear water
- Lowest pixel values in visible and near-infrared are approximation to atmospheric path radiance
- Minimum values subtracted from image

Atmospheric correction methods

Regression method

$$L_s = H \cdot \rho \cdot T + L_p$$

- Applicable to dark pixel areas
- Near-infrared pixel values are plotted against values in other bands
- Least square line fit using standard regression methods
- Resulting offset is approximation for atmospheric path radiance
- Offset subtracted from image

Atmospheric correction methods

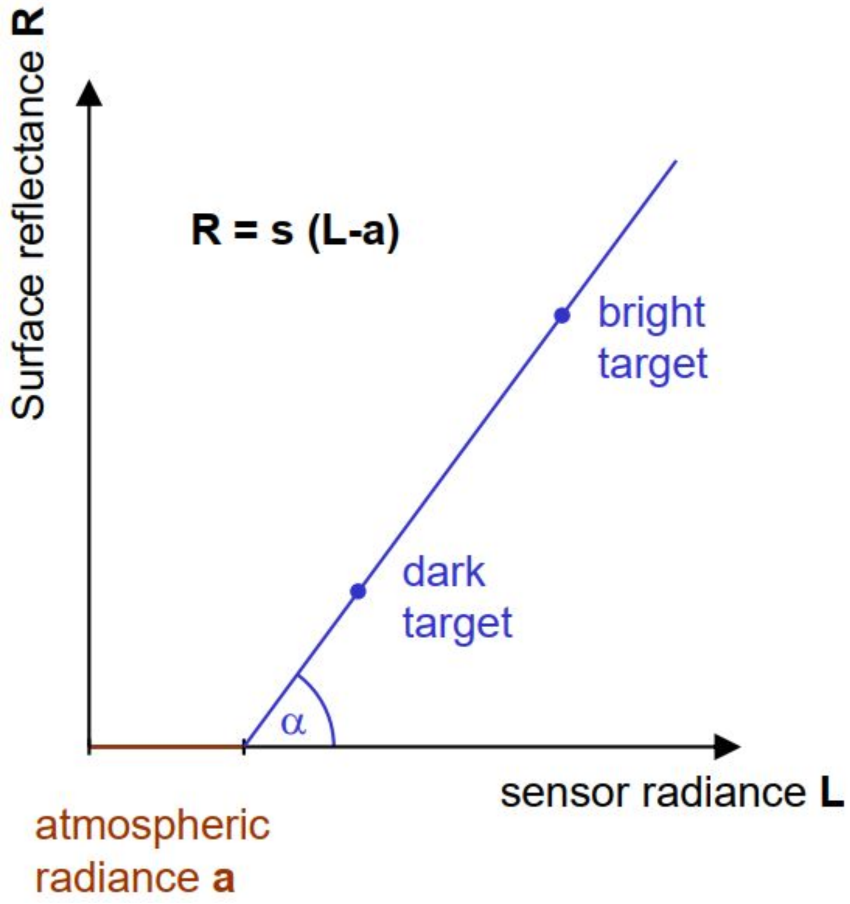
Radiative transfer models

- Limited by the need to supply data about atmospheric conditions at time of acquisition
- Mostly used with “standard atmospheres”
- Available numerical models
    - LOWTRAN 7
    - MODTRAN 4
    - ATREM
    - ATCOR
    - 6S (**S**econd **S**imulation of the **S**atellite **S**ignal in the **S**olar **S**pectrum)

Atmospheric correction methods

Empirical line method

- Selection of one dark and one bright target
- Ground reflectance measurement
  - Field radiometer
- Sensor radiance computed from image
- Slope s = cos (α) and intercept a of line joining two targets



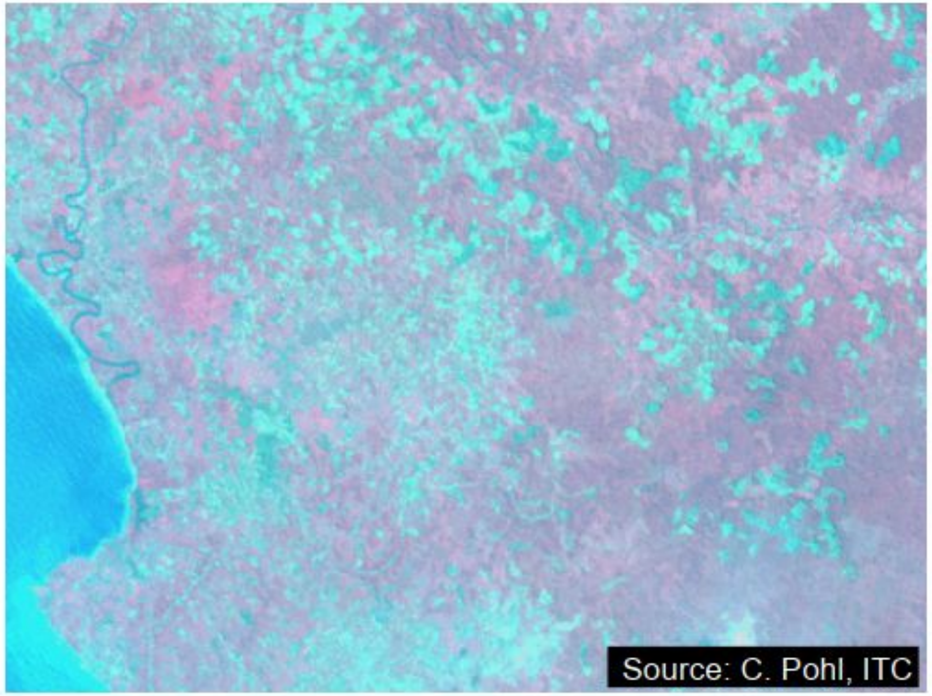
Atmospheric correction methods

Haze

- Due to Rayleigh scattering
  - Particles size responsible for effects smaller than the radiation‘s wavelength (e.g., oxygen and nitrogen)
- Haze has an additive effect resulting in higher DN values
- Decreases the general contrast of the image
- Effect is wavelength dependent
  - More pronounced in shorter wavelengths and negligible in the NIR

Atmospheric correction methods

Haze



Hazy

Corrected

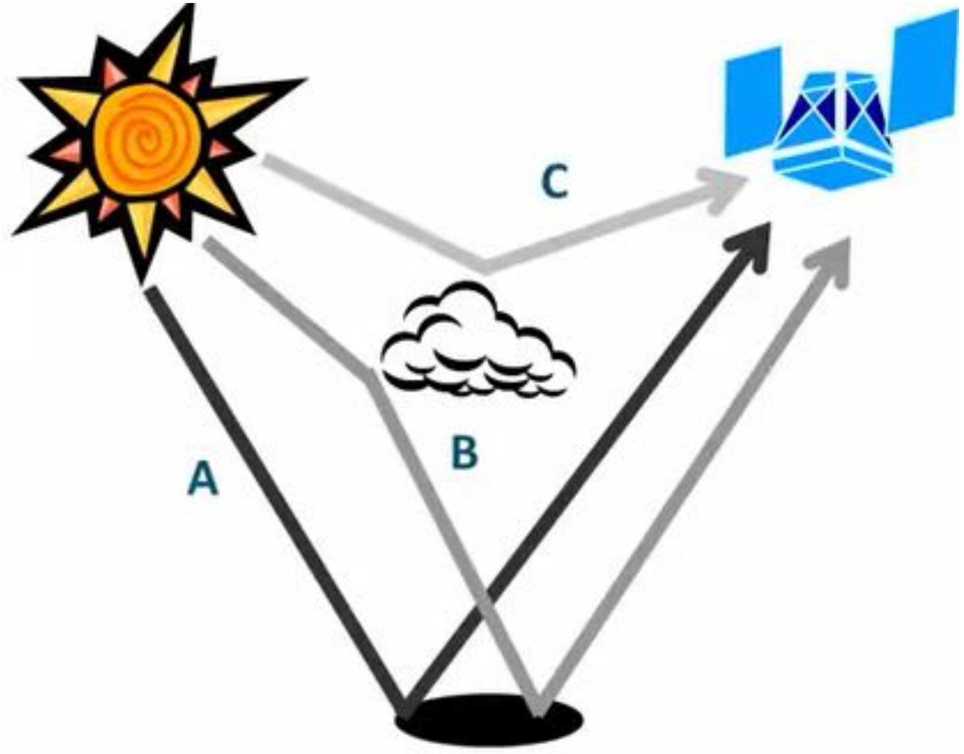
Influence of Cloud Cover

- Cloud cover is a common problem in optical remote sensing images due to its contamination of the reflectance from the Earth's surface
- Cloud removal has become a necessary preprocessing step for most applications
- Cloud cover may render an optical remote sensing image useless

ASTER satellite,
north of the Netherlands

Influence of Cloud Cover

- Pathway A: direct reflected sunlight
- Pathway B: skylight
- Pathway C: air light
- Atmospheric influence in two directions (up and down)

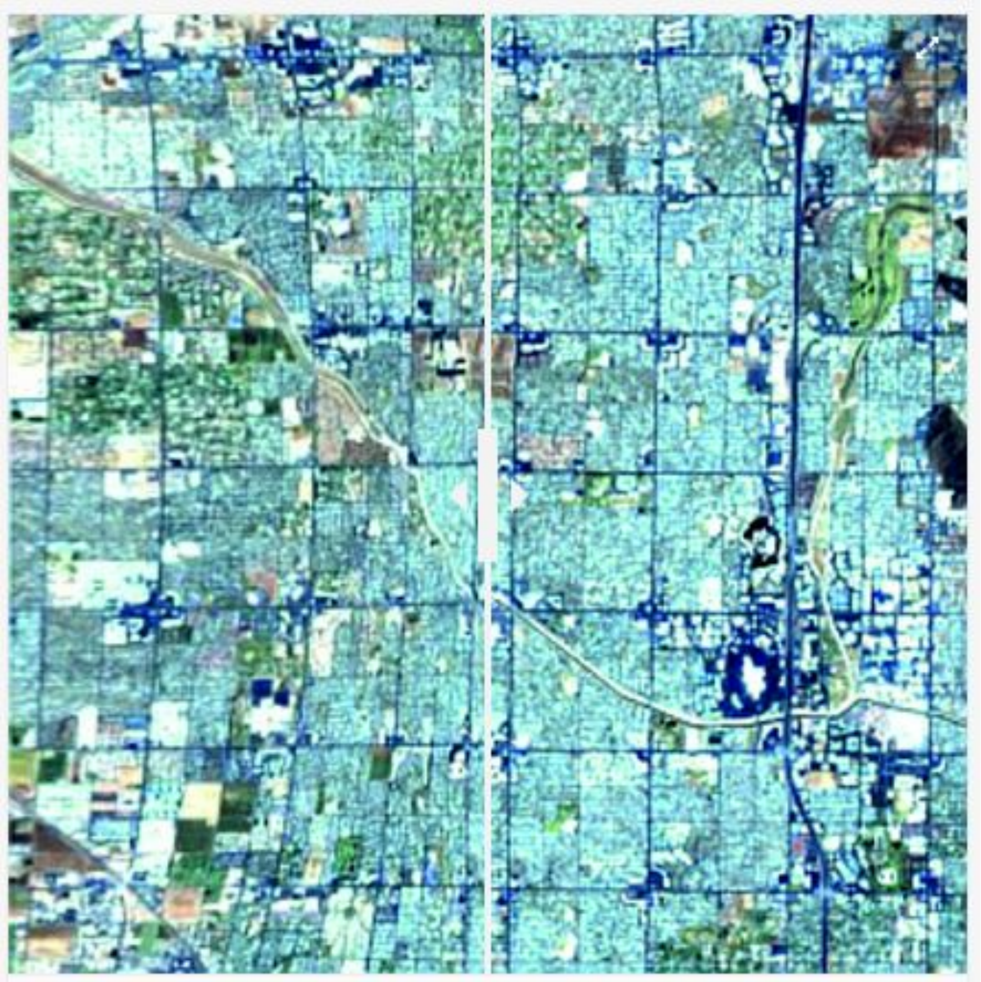


Generalized overview of pathways from sun to remote sensing sensor

Histogram stretching

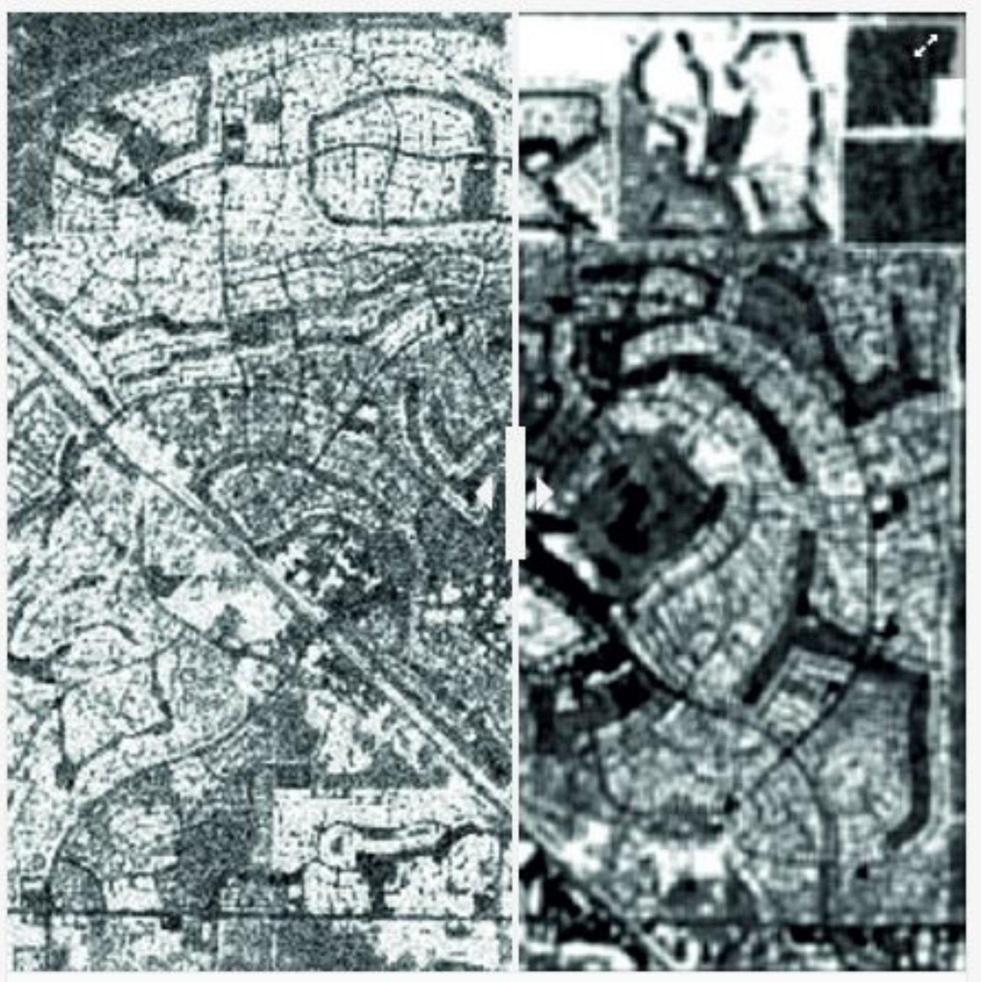
Stretching of digital numbers (rescaling over the range of the greyscale histogram) to allow better visibility and pixel distinction. (EOcollege, 2025)

Remove spatial distortion



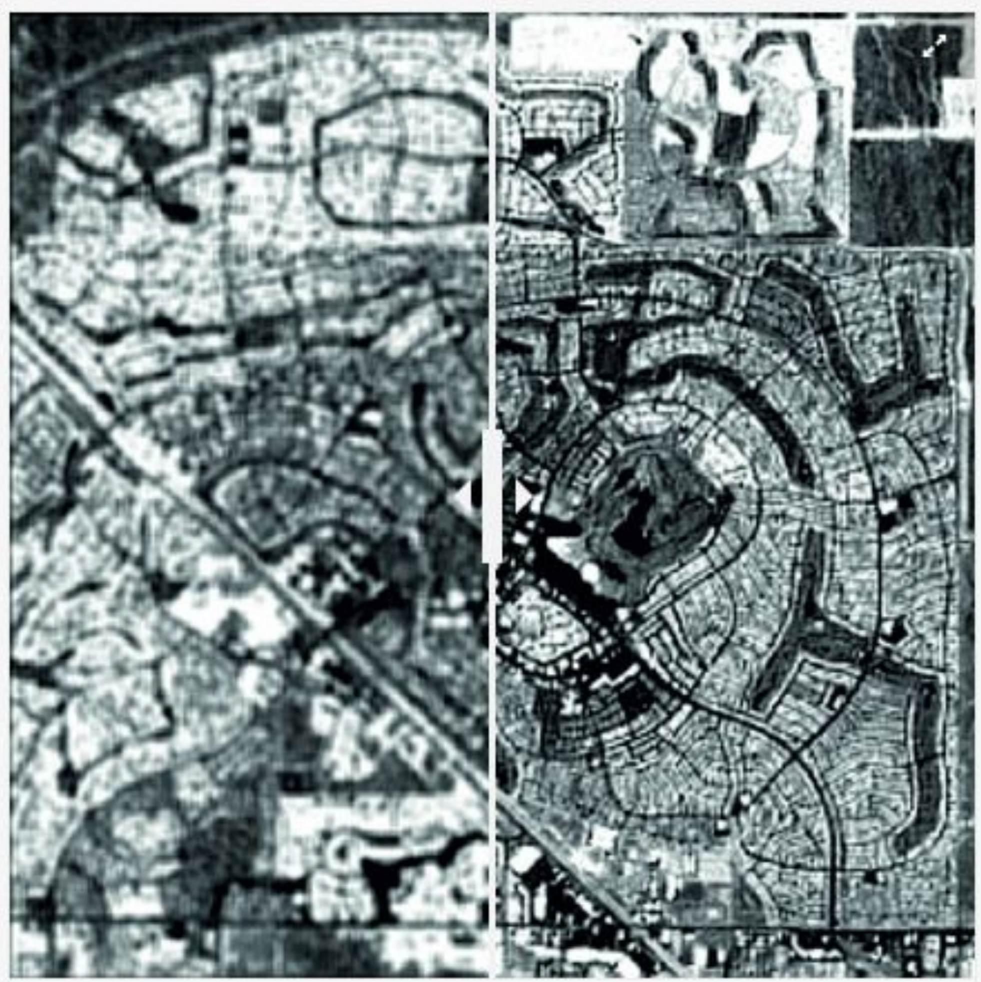
Geometric correction to remove spatial distortions. This process corrects the discrepancy between an image coordinate and the actual position on the Earth's surface. Exemplary, this can be done using ground control points (GCPs) from known 'correct' images.
(EOcollege, 2025)

Low pass filtering



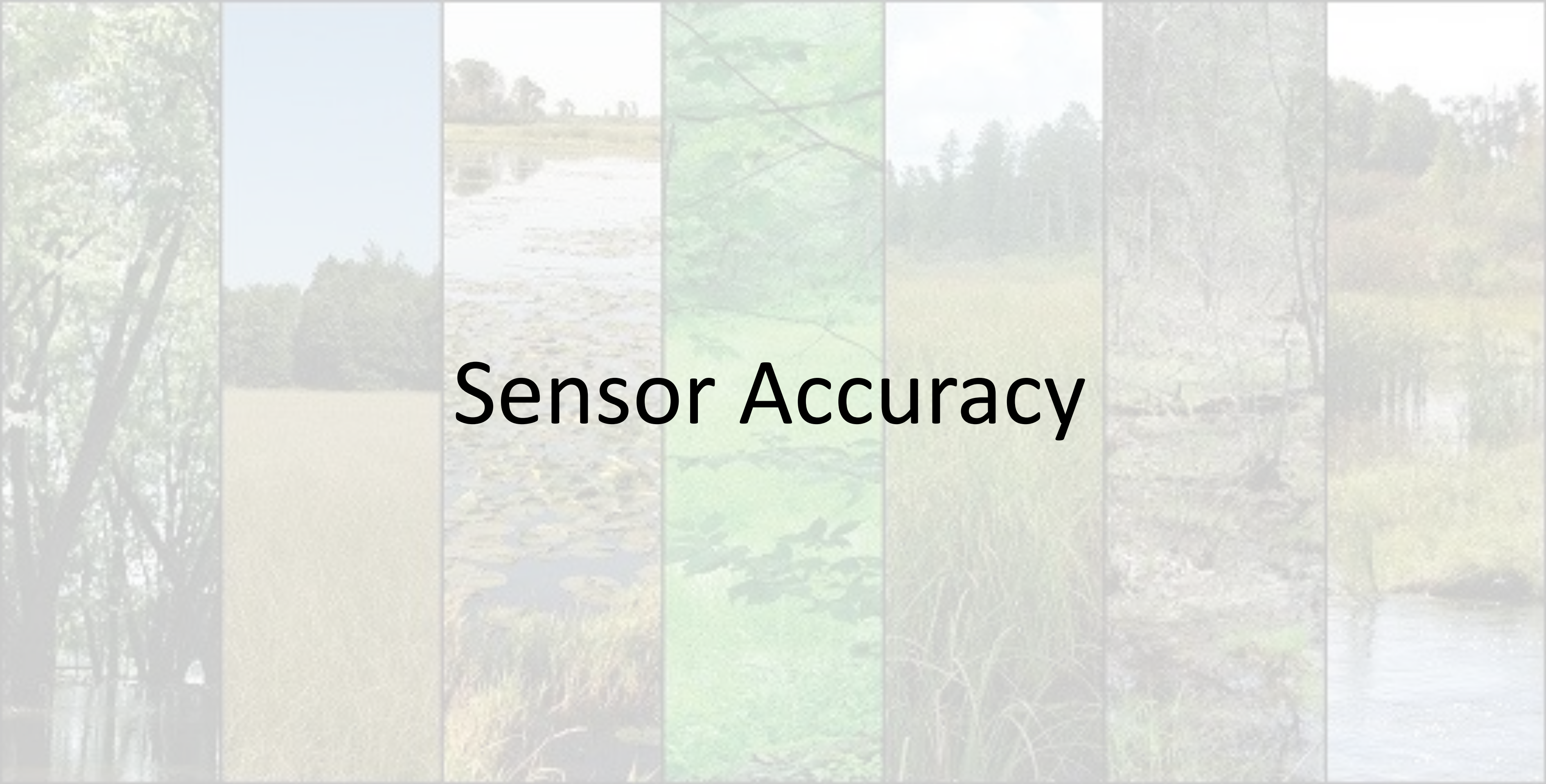
Low pass filtering to smoothen satellite imagery using moving window (e.g., 3 x 3 pixels), which are used to calculate local averages for kernel pixels.
(EOcollege, 2025)

High pass filtering



High pass filtering to sharpen edges and increasing contrast. They enhance spatial frequencies / components (e.g., edges). Subtracts the low pass filter from the original image.
(EOcollege, 2025)

Sensor Accuracy



Sensor Accuracy

General Accuracy Assessment

Principles of Validation

- Remote Sensing is an indirect technique to measure or quantify the studied objects Errors during the estimation process are inevitable.
- Error assessment or validation is typically performed regarding an *independent measurement* of the target quantity – e.g., the measurement was taken with another instrument or by applying another, more robust method.
- Usually, these validation measurements are also expected to have a higher quality than the RS derived estimates G*round truth* is used to talk about these measurements.

Sensor Accuracy

General Accuracy Assessment

Accuracy and precision

- Relate to quality

Accuracy

- Refers to distance of a measure (or prediction) to the true value
- Can be referred to as trueness
- An accurate measurement produces results that are close to the true value

Sensor Accuracy

General Accuracy Assessment

Accuracy and precision

- Relate to quality

Precision

- Refers to repeatability or reproducibility of a measurement
- Refers to the spread that occurs if the measurement (or prediction) is repeated under the same conditions
- A precise measurement produces results that are close to each other, no matter their distance from the true value

Sensor Accuracy

General Accuracy Assessment

Quality metrics

- For validation purposes, many quality indicators or metrics exist
   meant to make the error assessment quantifiable and objective
- They are used in scientific publications and reports to allow comparing different RS products with each other.
- The error metric that should be used to assess the quality of a RS product depends first on its type: interval and ratio scale variables like reflectance or Leaf Area Index require different metrics than nominal and ordinal scale variables like land cover.
- Error assessment for nominal and ordinal scale variables, important for land cover, are discussed in the classification topic.

Sensor Accuracy

General Accuracy Assessment

Quality metrics – R-squared

- R-squared is a measure of how well a linear regression model fits the data. It can be integrated as the proportion of variance of the outcome Y explained by the linear regression model.
- It is a number between 0 and 1 ($0 \leq R^2 \leq 1$). The closer its values is to 1, the more variability the model explains. And $R^2 = 0$ means that the model cannot explain any variability in the outcome Y.

Sensor Accuracy

General Accuracy Assessment

Quality metrics – R

On the other hand, the **correlation coefficient r** is a measure that quantifies the strength of the linear relationship between 2 variables.

r is a number between -1 and 1 ($-1 \leq r \leq 1$):

- **A value of r close to -1**: means that there is negative correlation between the variables (when one increases the other decreases and vice versa)
- **A value of r close to 0**: indicates that the 2 variables are not correlated (no linear relationship exists between them)
- **A value of r close to 1**: indicates a positive linear relationship between the 2 variables (when one increases, the other does)

Sensor Accuracy

General Accuracy Assessment

Quality metrics – Root Mean Square Error (RMSE)

- RMSE measures the accuracy of an RS product, i.e., the distance of RS predictions from the corresponding ground truth measurements. It is defined as:

$$RMSE = \sqrt{\frac{\sum(\hat{x}_i - x_i)^2}{n}}$$

- It should be used when comparing metrics that should by definition be the same quantities, e.g., if Leaf Area Index (LAI in m^2/m^2) is derived from RS data, it can be compared to ground truth LAI data with the RMSE.

- has the same physical unit as the RS product and the ground truth. It should be noted that the RMSE weights large errors heavier than smaller errors because of the square. Also, positive (overestimation) and negative (underestimation) errors have the same impact on RMSE.

Sensor Accuracy

General Accuracy Assessment

Quality metrics – Mean Absolute Error (MAE)

- MAE measures the average magnitude of the errors in a set of predictions
- Is not considering their directions
- It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight

$$\mathrm{MAE} = \frac{1}{n}\sum_{j=1}^{n} |y_j - \hat{y}_j|$$

Sensor Accuracy

Surface Reflectance

- Surface reflectance is a basic product in remote sensing and starting point for many applications
- Still, there are several error sources contributing to the “measured” surface reflectance
  - Sensor calibration
  - Atmospheric correction

Supported by:

Federal Ministry for the Environment, Nature Conservation and Nuclear Safety

Federal Agency for Nature Conservation

EOCap 4Africa

Thank you for your attention!

Dr. Insa Otte

insa.otte@uni-wuerzburg.de