11th Advanced ESA Training Course on Land Remote Sensing

Hyperspectral Remote Sensing for Agriculture (and water) - Theoretical Lecture

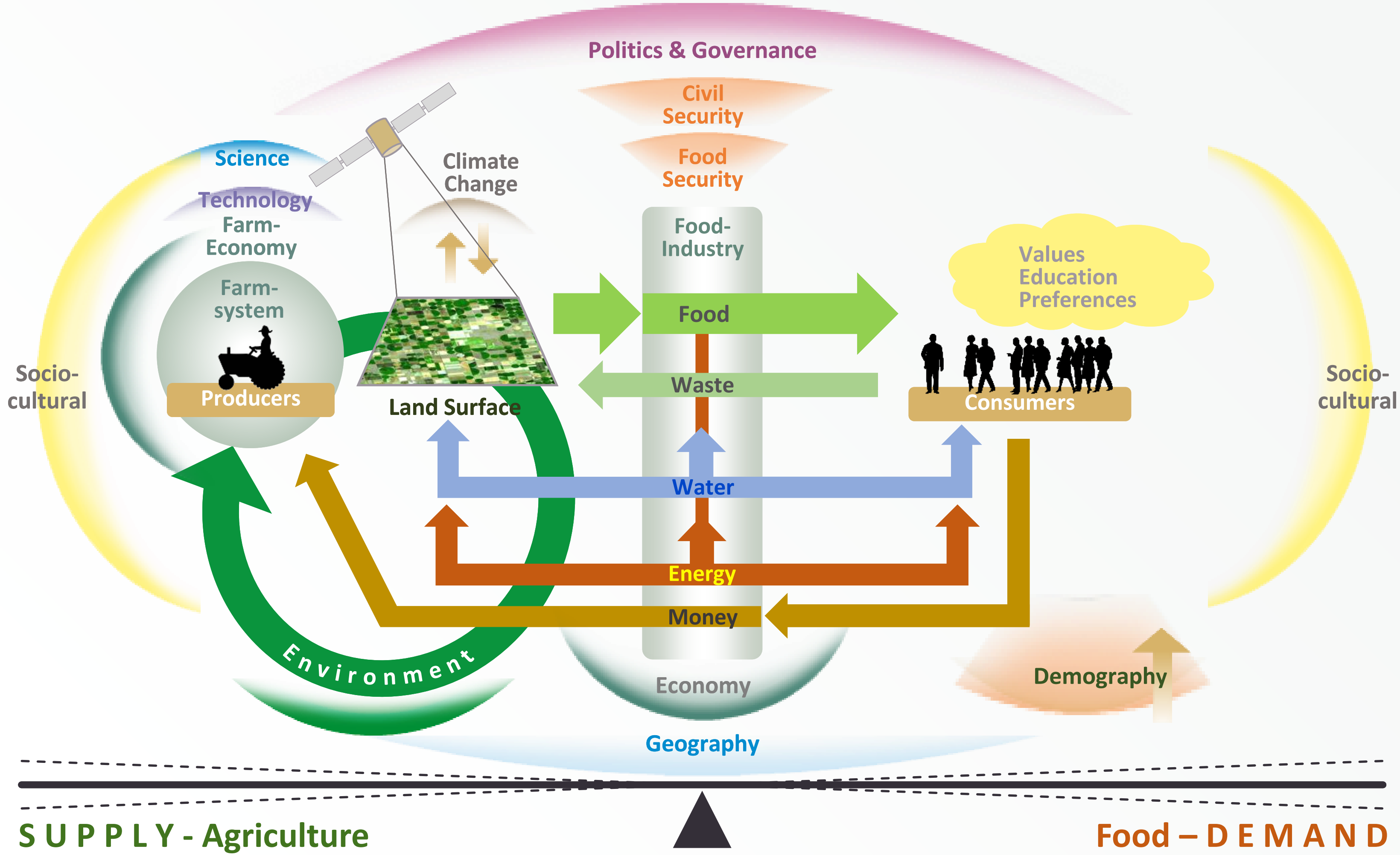
PD Dr. Tobias B. Hank

Dept. of Geography | Faculty of Geosciences
Ludwig-Maximilians-Universität Munich (Germany)

EnMAP

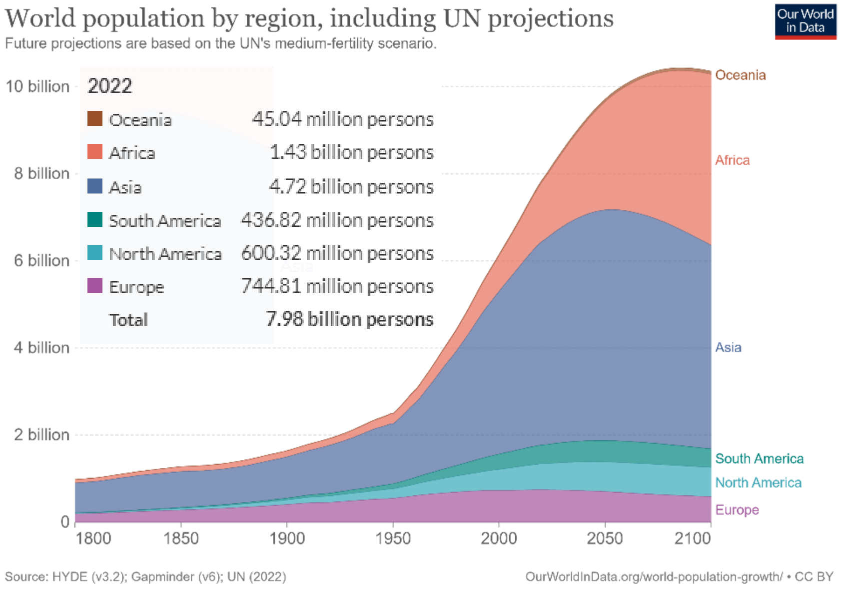
Session: Day 5 – Hyperspectral Data for Agriculture and Water
Place: Institute of Advanced Studies Kőszeg (iASK), Hungary
Time: Friday | November 25th | 08:30-10:00 UTC+1

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN


Politics & Governance
Civil Security
Food Security
Science
Technology
Farm-Economy
Climate Change
Food-Industry
Values
Education
Preferences
Farm-system
Producers
Land Surface
Food
Waste
Consumers
Socio-cultural
Socio-cultural
Water
Energy
Money
Environment
Economy
Demography
Geography
S U P P L Y - Agriculture
Food – D E M A N D

Challenges of food production & sustainability

“Humankind is facing an unprecedented challenge to produce enough food for the coming decades due to population growth and increase in the average demand per capita, changes in climate conditions, and limitations in arable land area, as well as pressure on the water and resources.“ (Ninomiya et al., 2019)



One strategy to meet this increasing food demand has been converting more land into arable land…

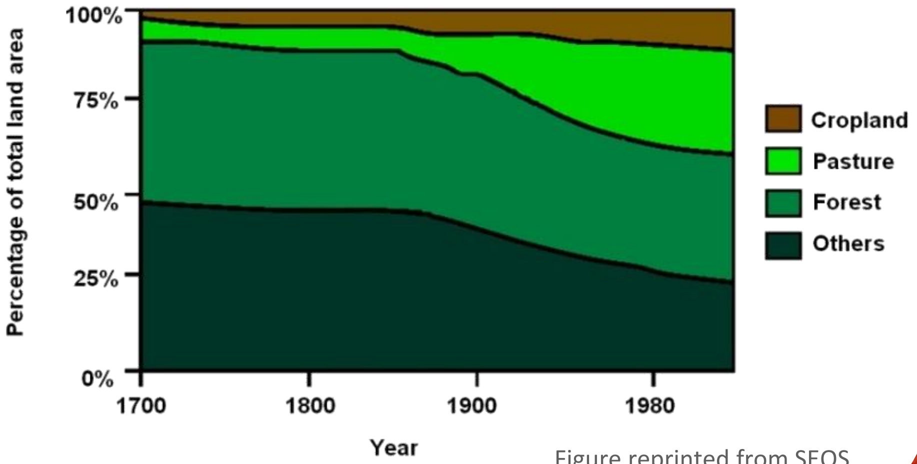
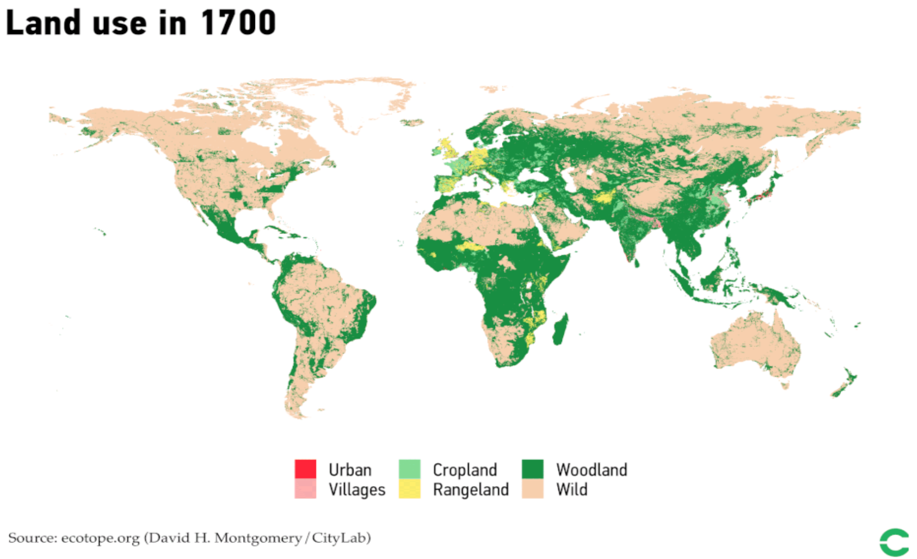
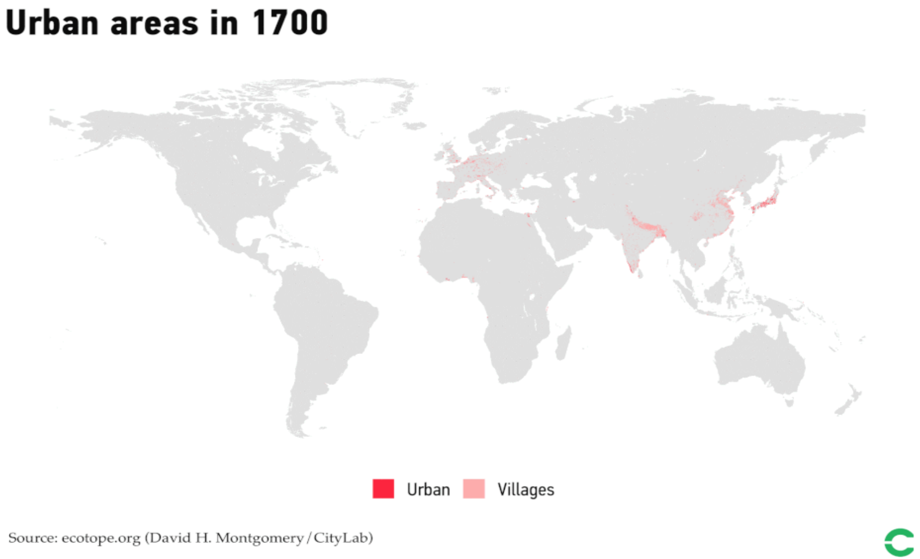


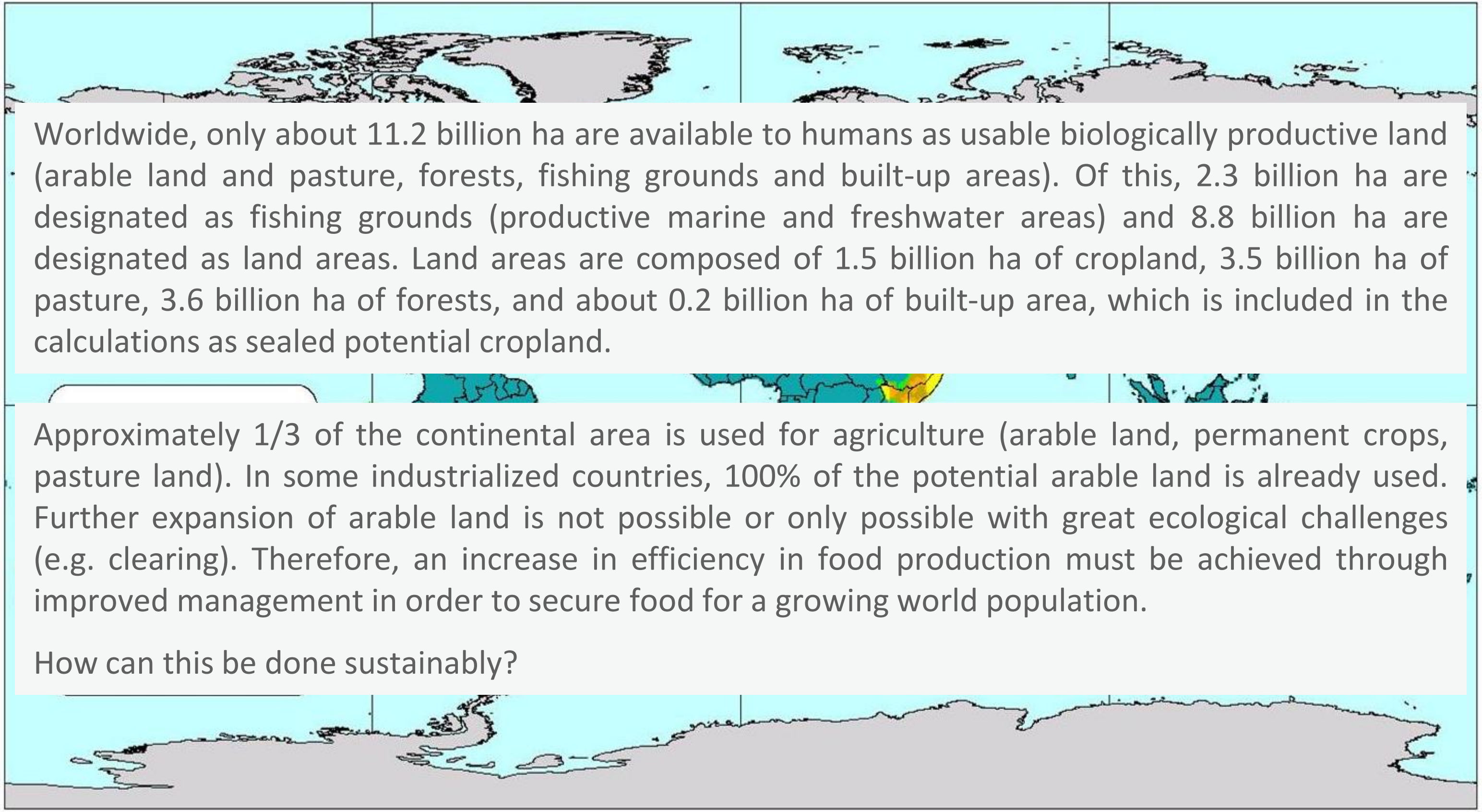
Figure reprinted from SEOS

Challenges of food production & sustainability





Challenges of food production & sustainability

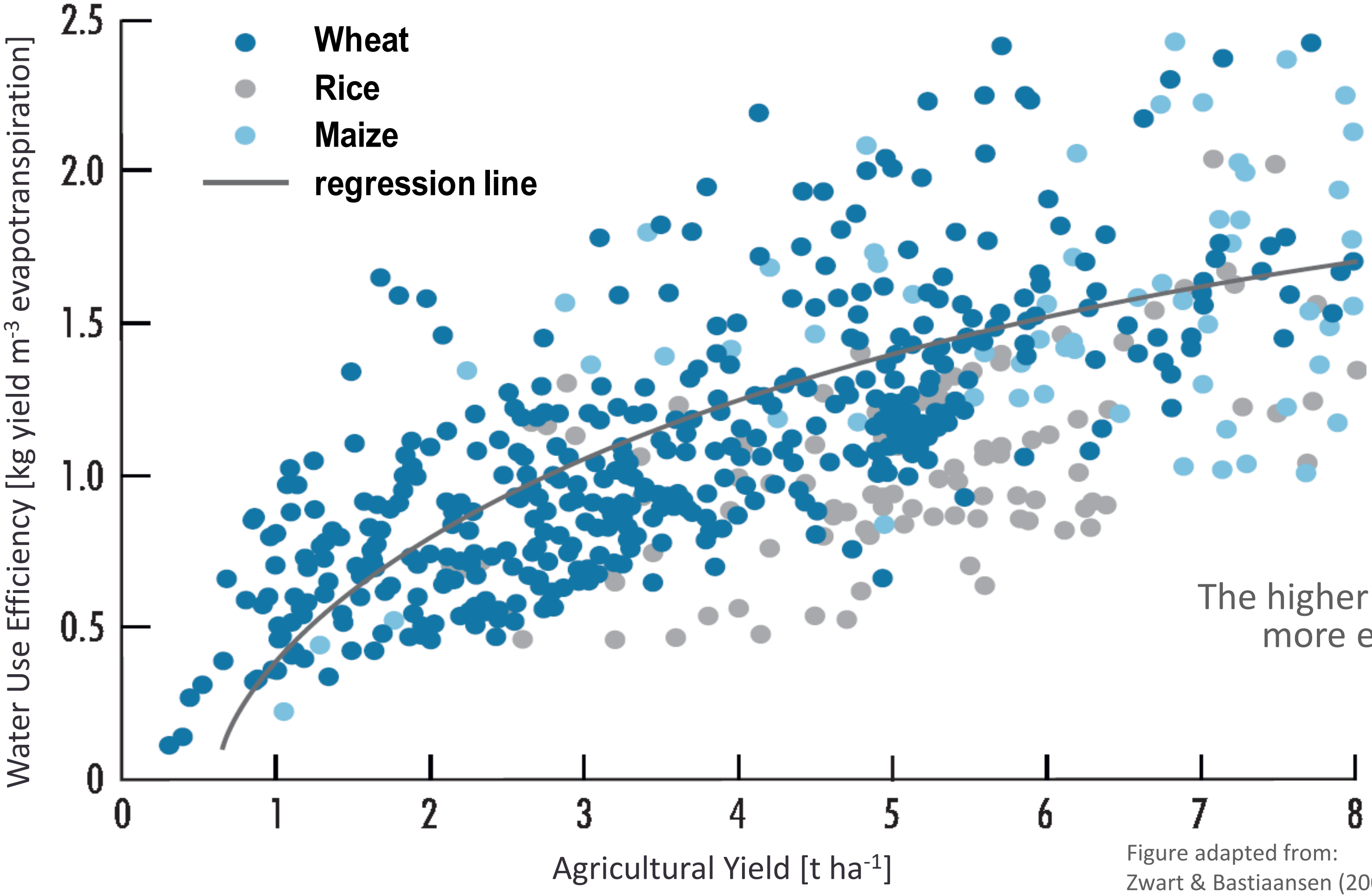


Worldwide, only about 11.2 billion ha are available to humans as usable biologically productive land (arable land and pasture, forests, fishing grounds and built-up areas). Of this, 2.3 billion ha are designated as fishing grounds (productive marine and freshwater areas) and 8.8 billion ha are designated as land areas. Land areas are composed of 1.5 billion ha of cropland, 3.5 billion ha of pasture, 3.6 billion ha of forests, and about 0.2 billion ha of built-up area, which is included in the calculations as sealed potential cropland.

Approximately 1/3 of the continental area is used for agriculture (arable land, permanent crops, pasture land). In some industrialized countries, 100% of the potential arable land is already used. Further expansion of arable land is not possible or only possible with great ecological challenges (e.g. clearing). Therefore, an increase in efficiency in food production must be achieved through improved management in order to secure food for a growing world population.

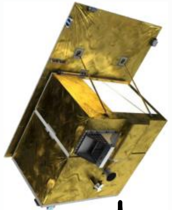
How can this be done sustainably?

Arid regions according to FAO



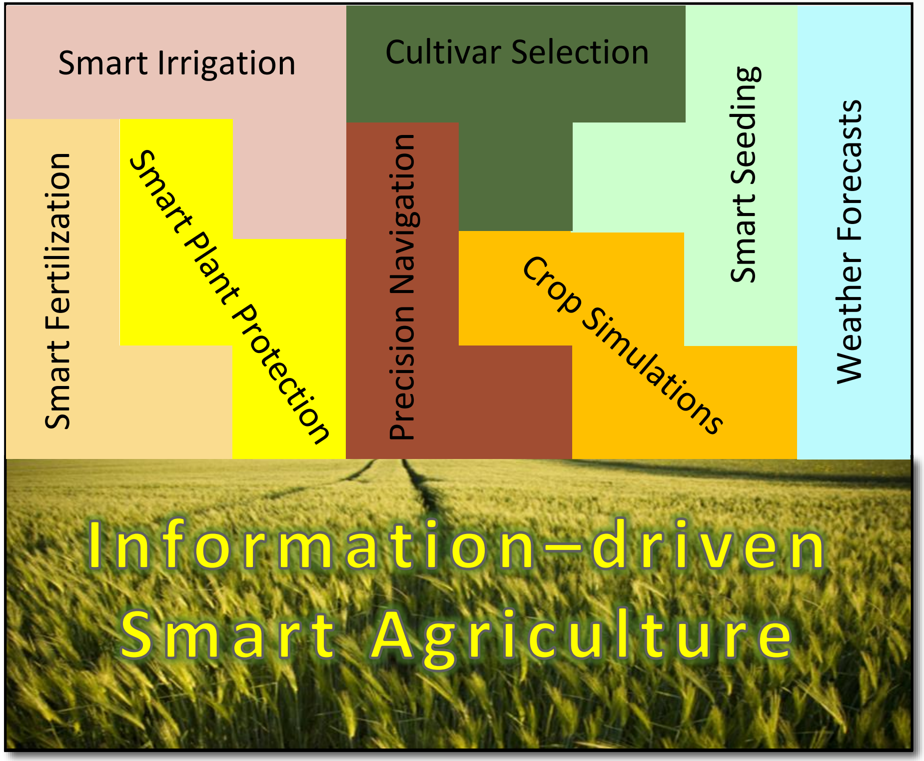
The higher the yield level, the more efficient production becomes!

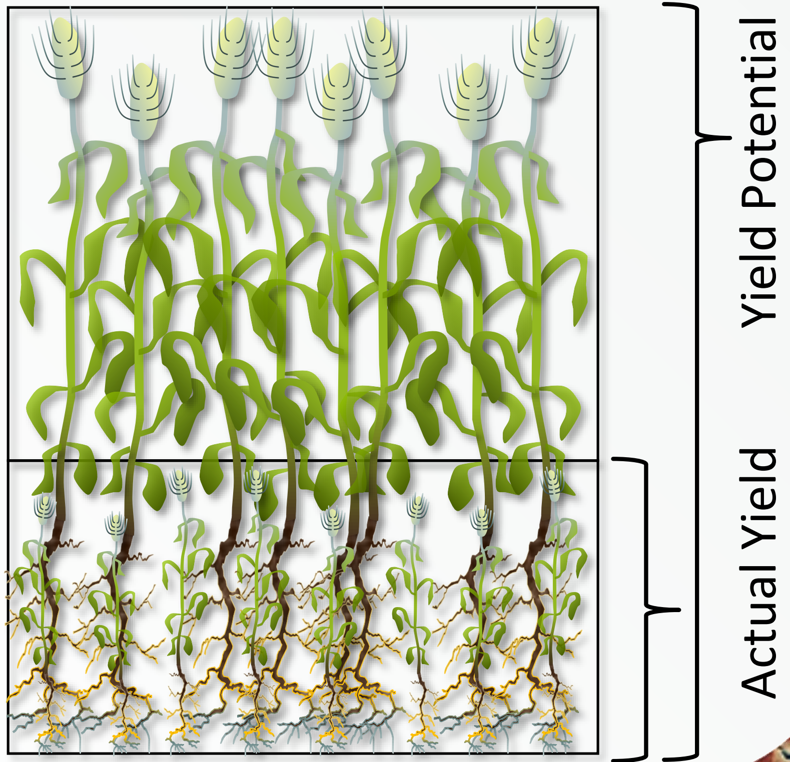
Figure adapted from:
Zwart & Bastiaansen (2004)



All with central contributions from satellite technology!

Results from the project „EO^{4}Food: Earth Observation Needs and Opportunities to Support Sustainable Agriculture and Development", funded by ESA



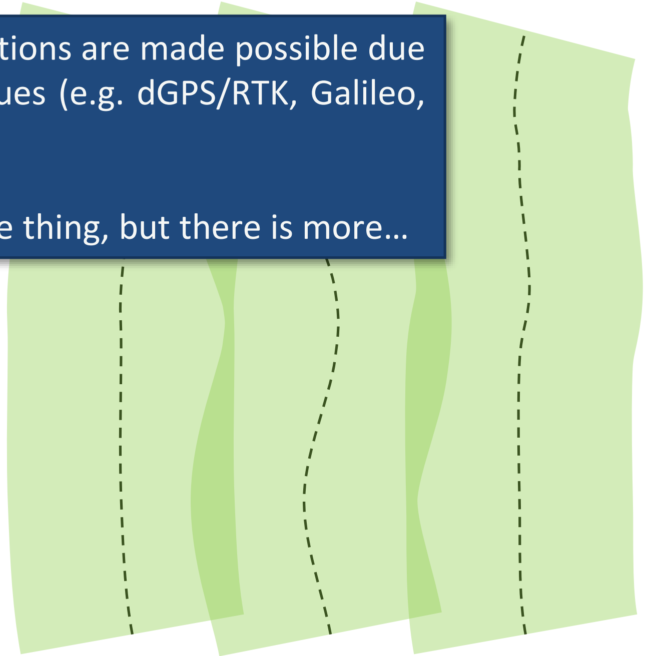


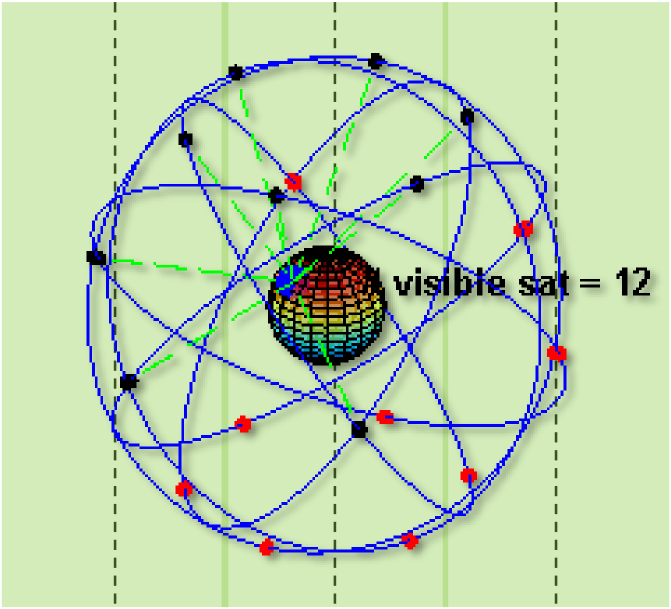
Precision or Smart Farming

Auto Steering

By minimizing the overlap of the lanes, higher efficiency is achieved!

Manual Steering vs. Auto-Steering



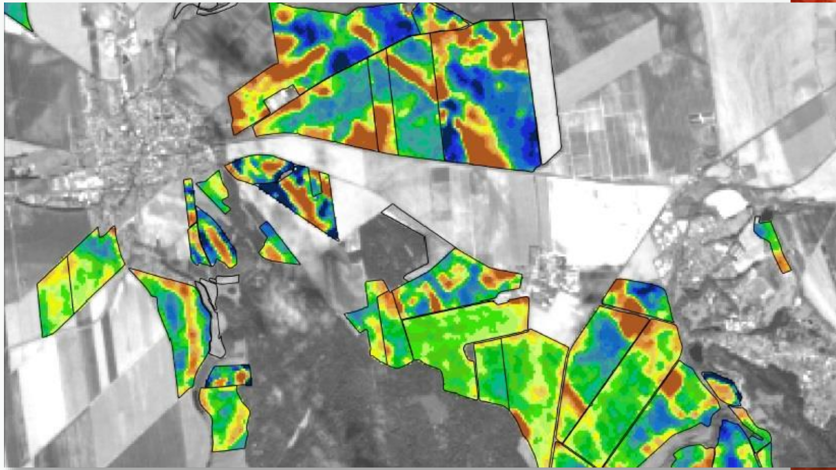


Auto-Steering-Applications are made possible due to navigation techniques (e.g. dGPS/RTK, Galileo, GLONASS, BeiDou).

Navigation is a satellite thing, but there is more…

Precision or Smart Farming
Site Specific Measures

Image credit: VISTA GmbH



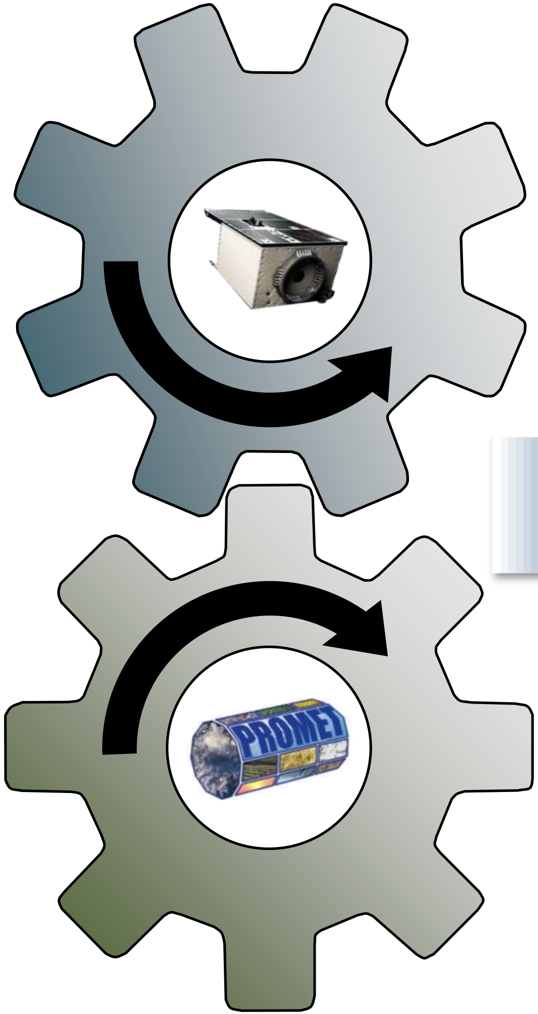
Examples for site-specific management (with the help of earth observation):

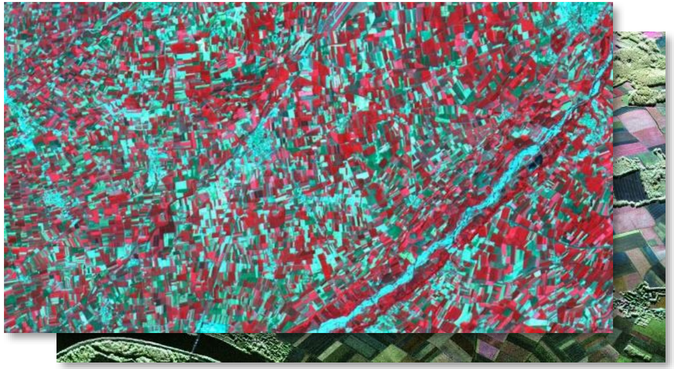
- Distribution of seeds according to yield potential.
- Distribution of fertilizer applications according to current growth activity.
- Distribution of plant protection agents according to current biomass distribution.
- Targeted positioning of soil samples:

Detection of long-term persistent growth patterns

5 samples, 5 soil types

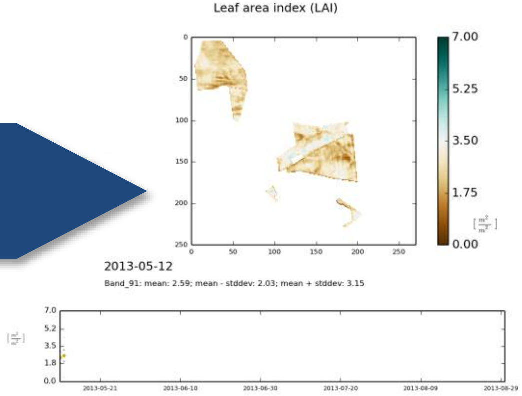
6 samples, 4 soil types



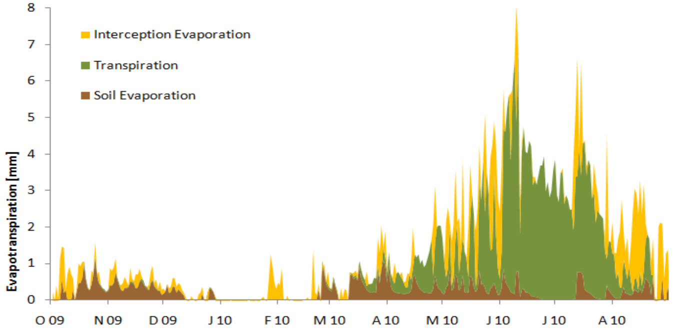


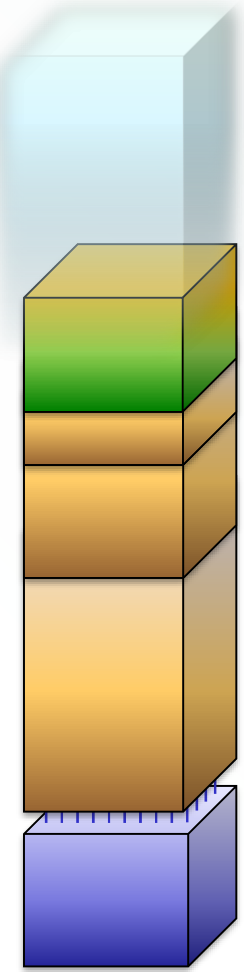
Spatially and temporally dynamic information products

Remote Sensing:
Spatial Dynamics



Process Modelling:
Temporal Dynamics





Atmosphere (wind speed and direction, radiation, temperature, CO_2, relative humidity, cloud cover at 2m above ground)

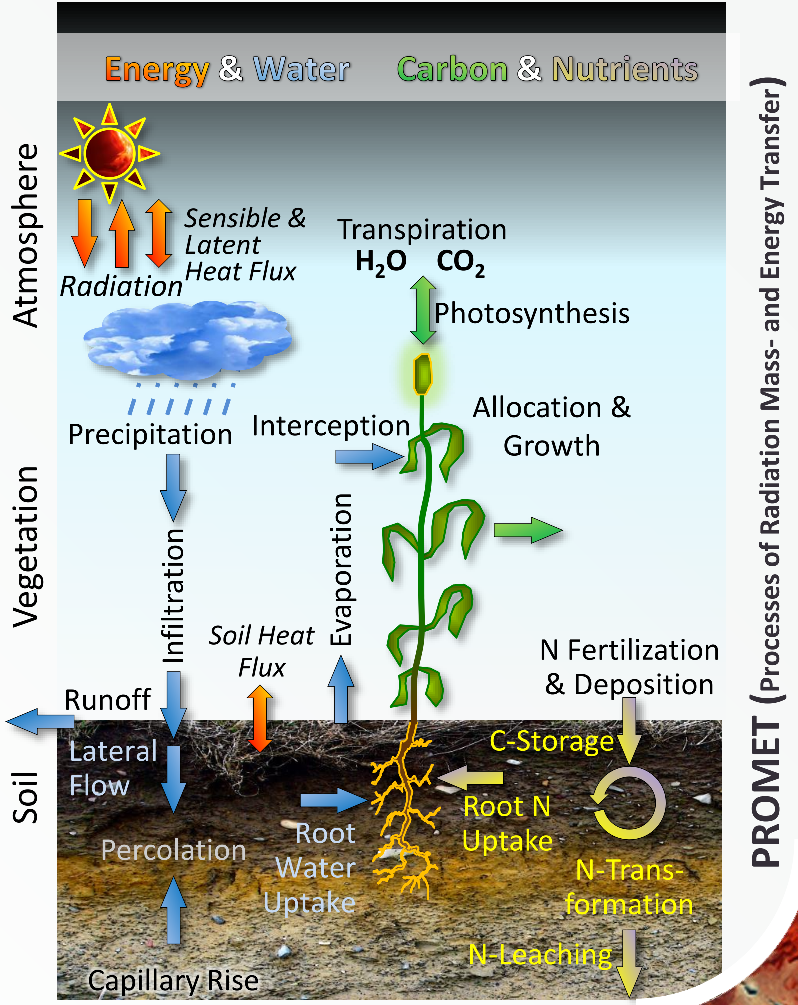
Canopy (fractionized biomass, LAI, height, phenology, root depth/density at -2-20m)

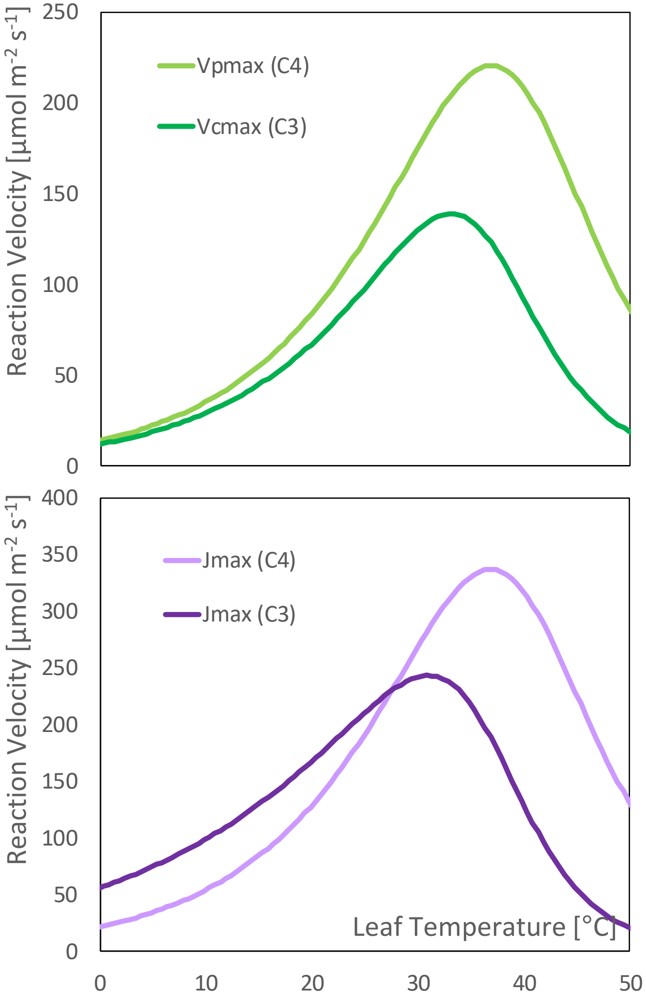
Soil layer 1 (in 0-5cm depth)
Soil layer 2 (in 5-20cm depth)

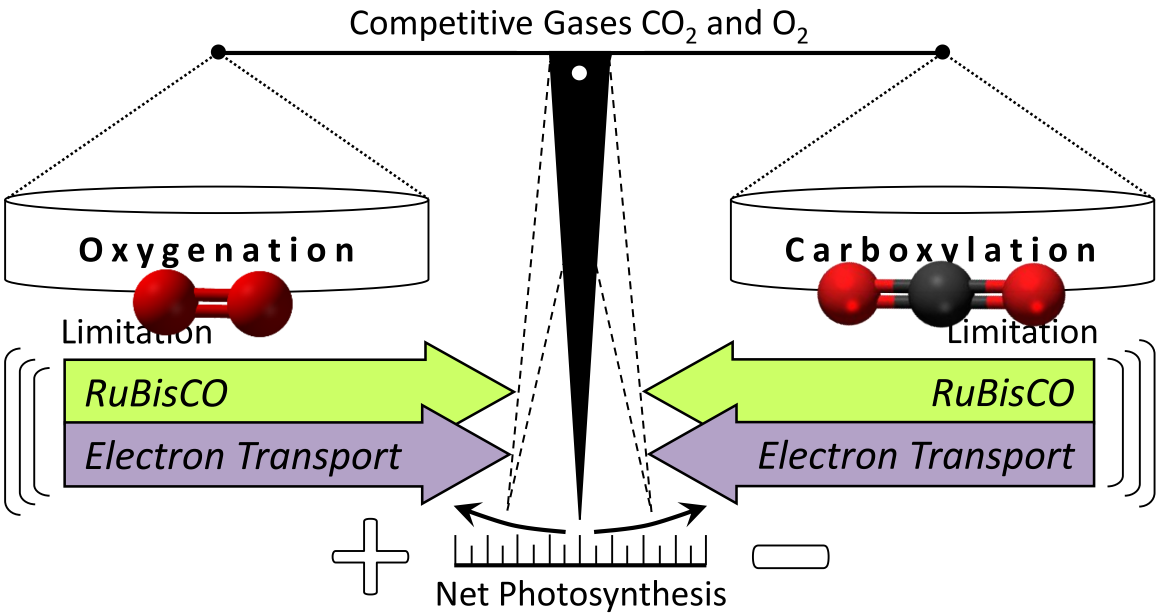
Soil layer 3 (in 20-60cm depth)

Soil layer 4 (in 60-200cm depth)

Groundwater body (capillary rise, variable depth)







\+ Stomatal conductance (drought stress, CO_2-scenarios…)
\+ Phenological model (varieties, yield formation…)
\+ Carbon allocation (canopy structure, radiative transfer…)
\+ Management (sowing, fertilizing, mowing, harvesting…)
\+ etc., etc., etc.

Hank, T. (2008): A Biophysically Based Coupled Model Approach for the Assessment of Canopy Processes Under Climate Change Conditions.

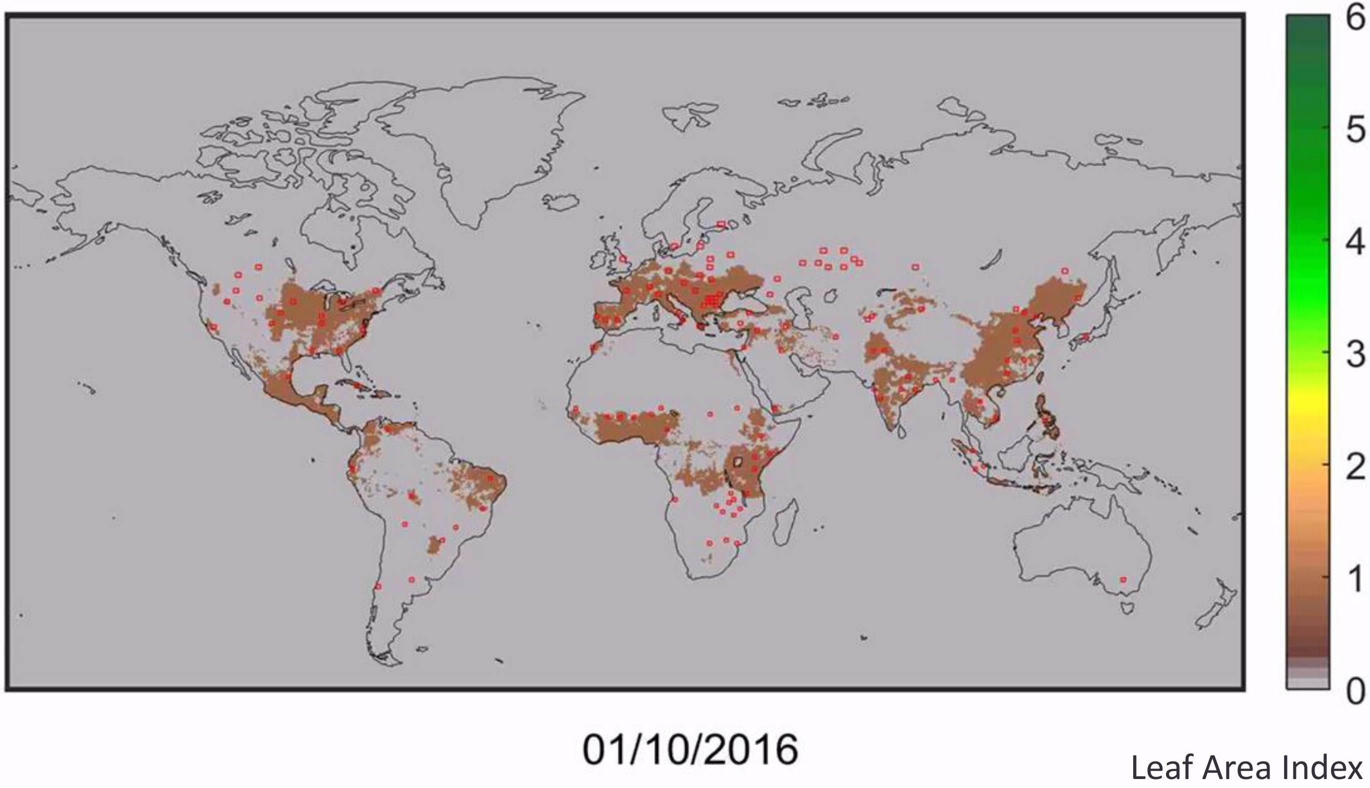
= **Maize**

= standard

= **rainfed**

= 30 arcsec
~1km

= 1 h



Results from the project „ViWA – Virtual Water Values", funded by BMBF

Bundesministerium
für Bildung
und Forschung

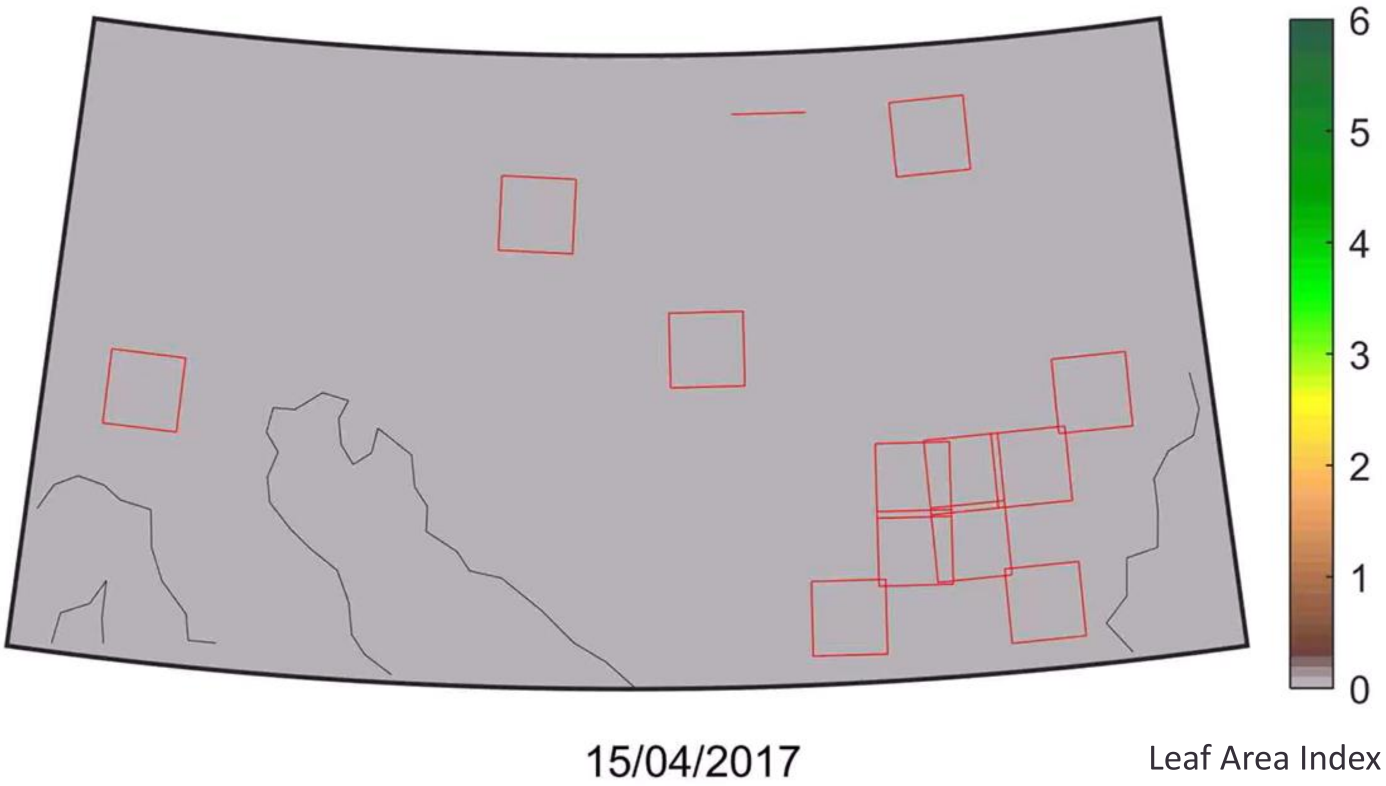
= **Maize**

= standard

= **rainfed**

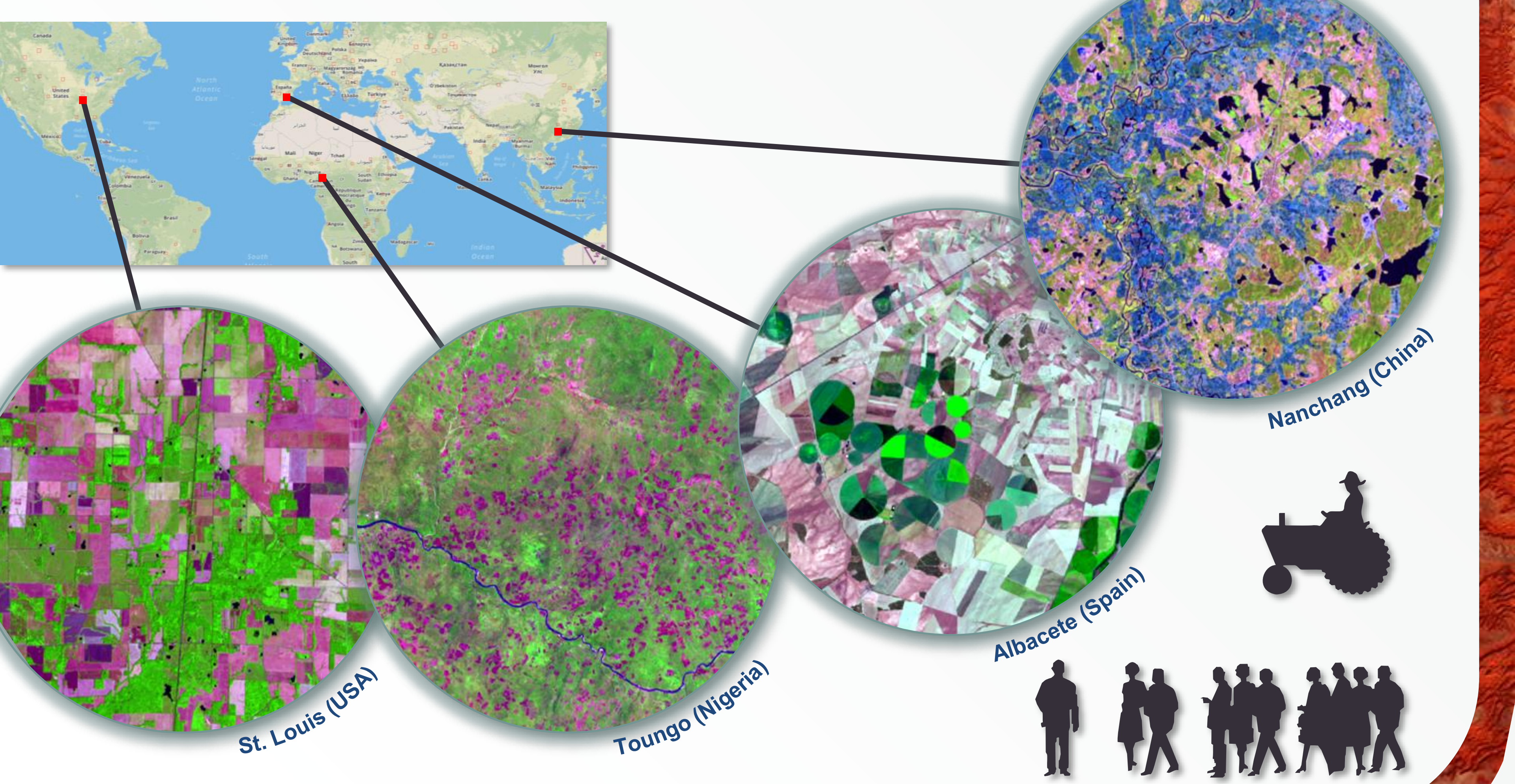
= 30 arcsec
~1km

= 1 h



Results from the project „ViWA – Virtual Water Values“, funded by BMBF

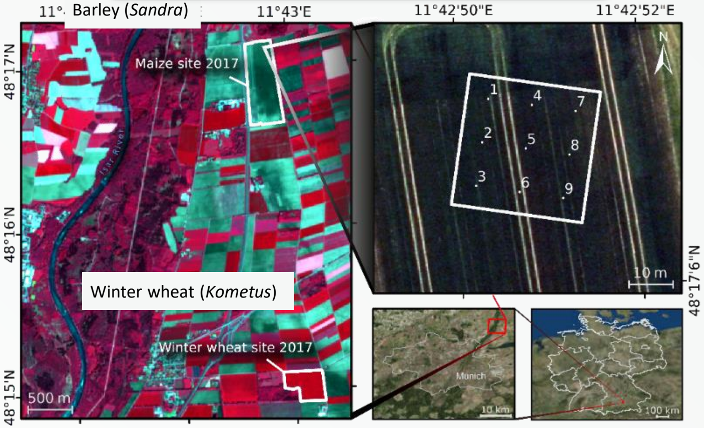
Bundesministerium für Bildung und Forschung


St. Louis (USA)
Toungo (Nigeria)
Albacete (Spain)
Nanchang (China)

Field Campaigns (Test sites MNI & Irlbach)

In-situ data collection of:

- Leaf chlorophyll content (C_{ab})
- Leaf area index (LAI)
- Biomass samples (organ-specific)
  - Leaf water content (EWT),
  - Leaf mass per area (LMA),
  - Nitrogen (N) & Carbon (C) content
- Spectral data (ASD)



The EnMAP Mission and its Scientific Preparation Program

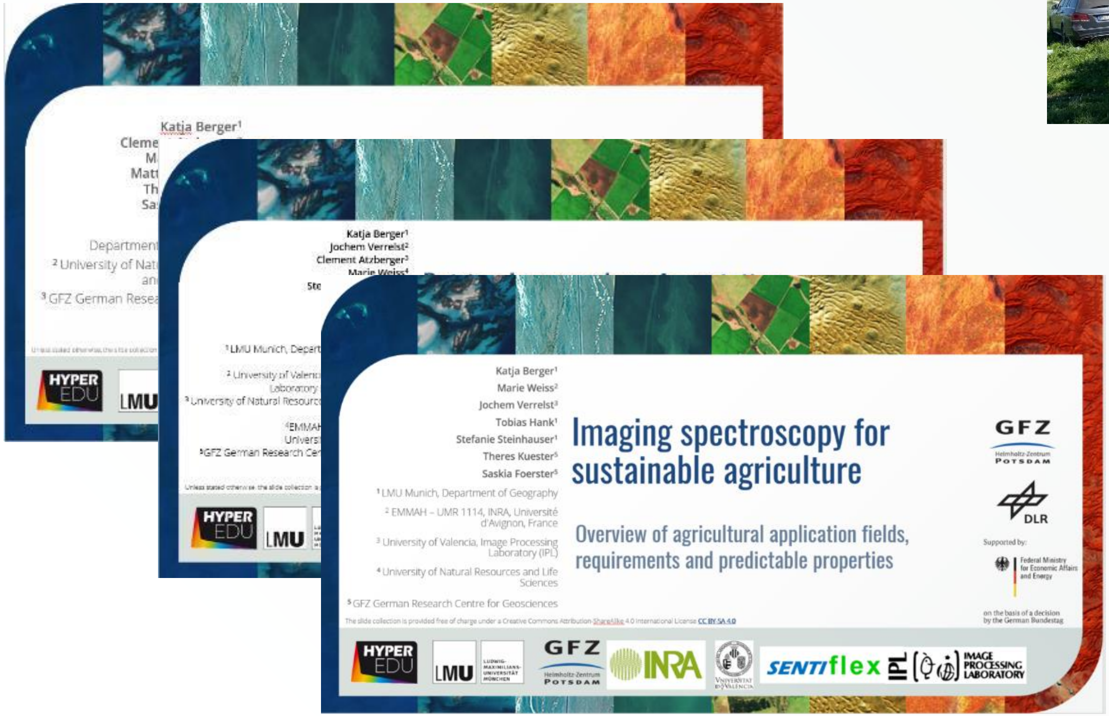
The Mission
- 420 - 2450 nm
- 30 m spatial resolution
- 30 km swath width
- 30° off-nadir pointing
- Launch: April 1st 2022

Science Projects
- Forests & Ecosystems
- Geology & Soils
- Coastal & Inland Waters
- Ecosystem Transitions
- Agriculture & Vegetation

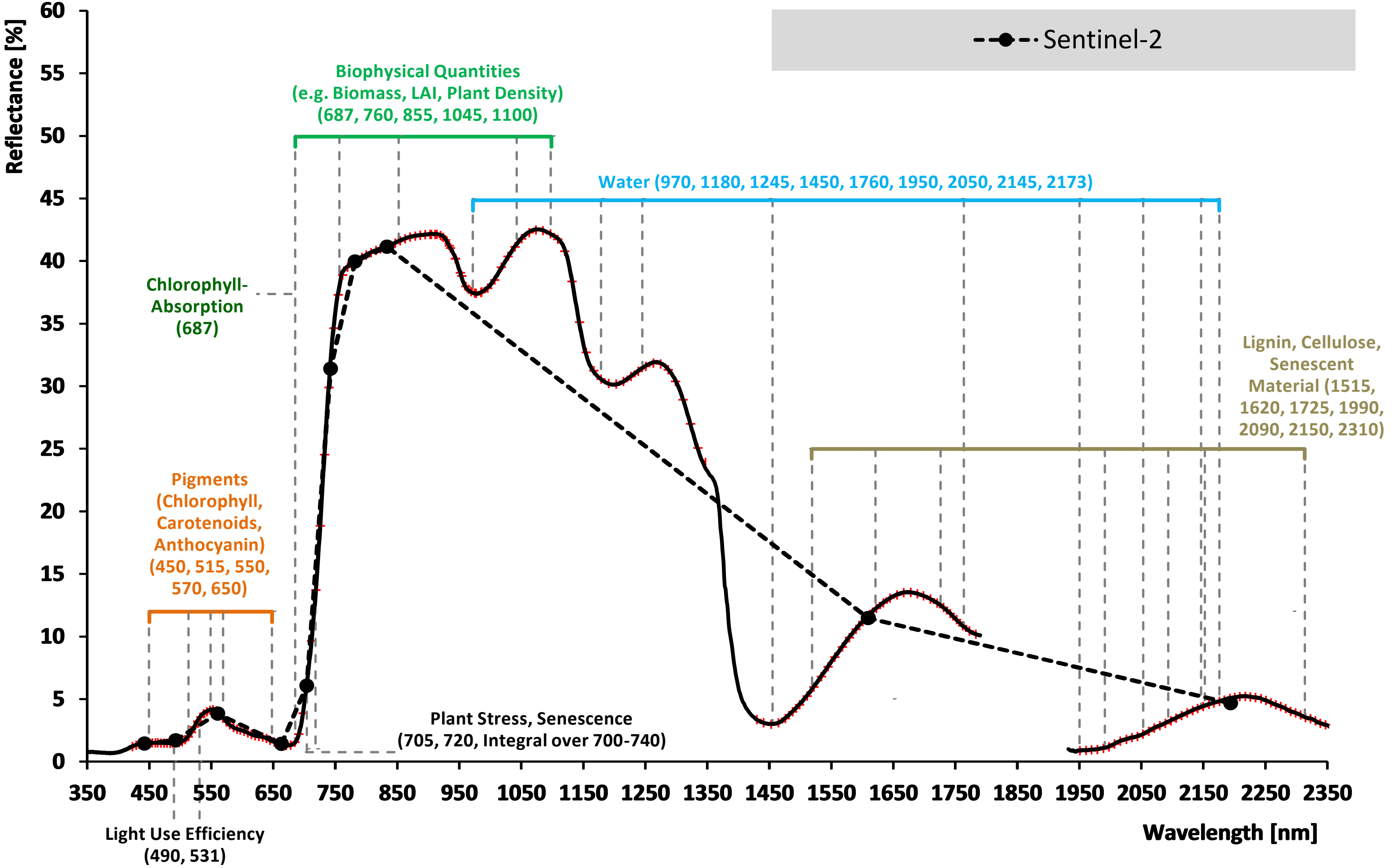
Edu-Program
- HyperEDU Sld Collections
- HyperEDU Workshops
- HyperEDU Tutorials
- HyperEDU MOOCs
- EO College

HyperEDU

Workshops & Summer Schools (SensEco, EnMAP-Box, EARSeL etc.)
EO College (Tutorials, Units, Slide Collections, Video Production)



https://eo-college.org/courses/beyond-the-visible/


Sentinel-2
Reflectance [%]
60
55
50
45
40
35
30
25
20
15
10
5
0
350
450
550
650
750
850
950
1050
1150
1250
1350
1450
1550
1650
1750
1850
1950
2050
2150
2250
2350
Wavelength [nm]
Biophysical Quantities
(e.g. Biomass, LAI, Plant Density)
(687, 760, 855, 1045, 1100)
Water (970, 1180, 1245, 1450, 1760, 1950, 2050, 2145, 2173)
Chlorophyll-
Absorption
(687)
Lignin, Cellulose,
Senescent
Material (1515,
1620, 1725, 1990,
2090, 2150, 2310)
Pigments
(Chlorophyll,
Carotenoids,
Anthocyanin)
(450, 515, 550,
570, 650)
Plant Stress, Senescence
(705, 720, Integral over 700-740)
Light Use Efficiency
(490, 531)

And what about the soils beneath the vegetation canopy?

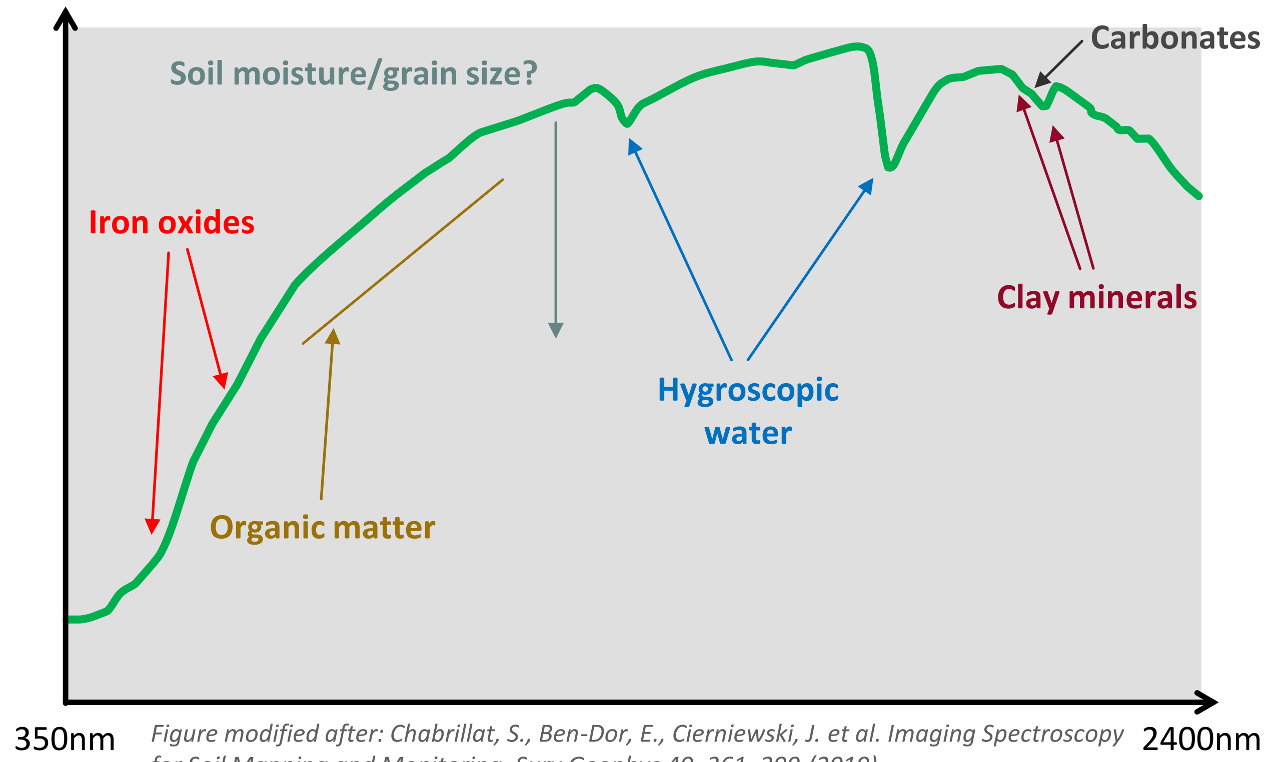
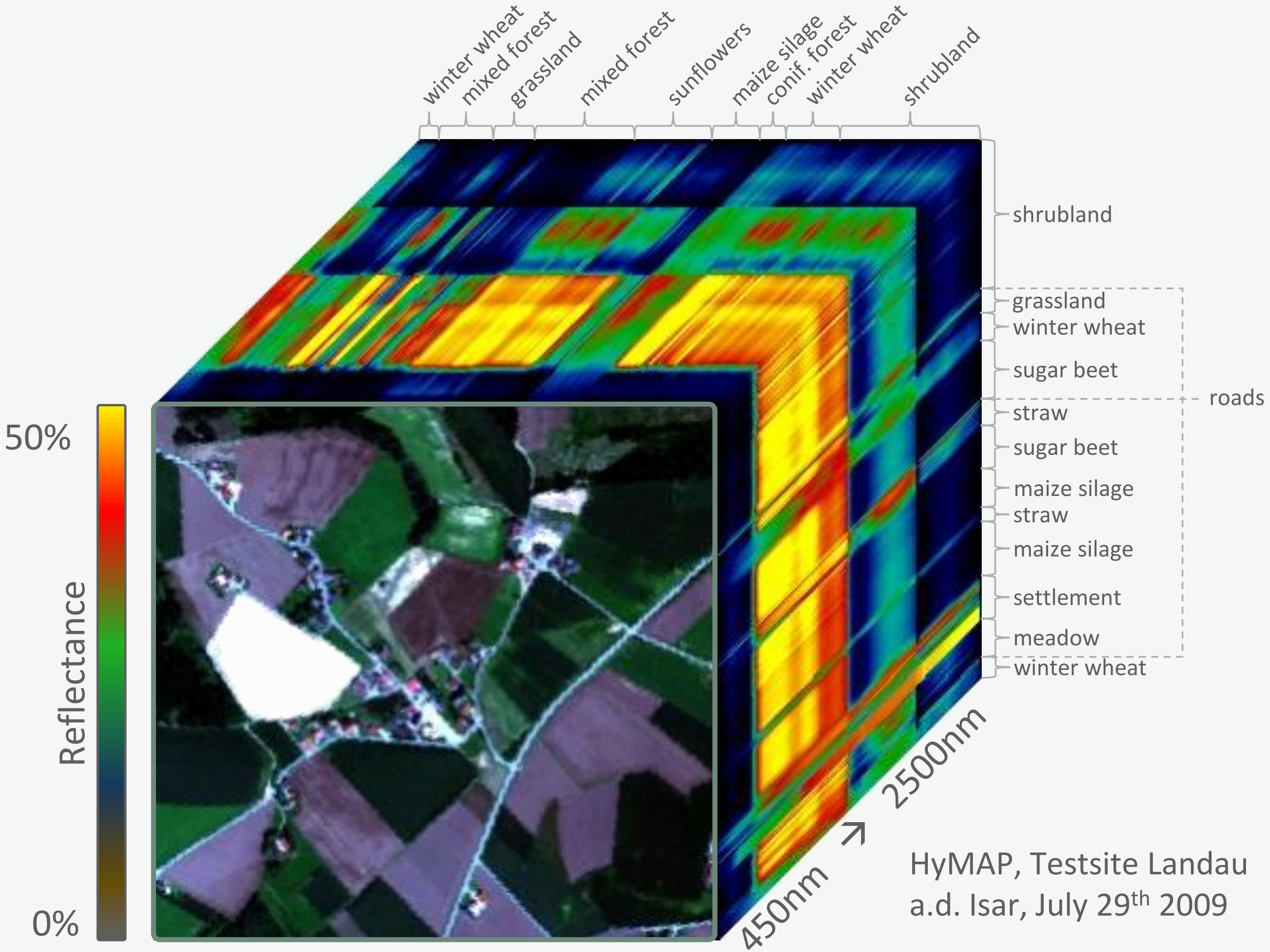


Figure modified after: Chabrillat, S., Ben-Dor, E., Cierniewski, J. et al. Imaging Spectroscopy for Soil Mapping and Monitoring. Surv Geophys 40, 361–399 (2019)

Hyperspectral Data Cube



HyMAP, Testsite Landau a.d. Isar, July 29th 2009

Classification of Biophysical and Biochemical variables

The concept of hyperspectral measurements can be traced to laboratory procedures (quantification is the main goal!). When quantifying our environment with the help of Hyperspectral data, we distinguish between variables that are directly involved in radiative transfer mechanisms and those which are secondarily involved:

Vegetation traits & radiative transfer

Directly drive radiative transfer:
Primary variables

Leaf level traits

- Pigment content (chlorophyll, carotenoid, & anthocyanin contents)
- Leaf mesophyll structure
- Leaf water

Canopy level traits

- LAI (green LAI, brown LAI)
- Leaf inclination
- Green cover fraction
- fAPAR
- Albedo

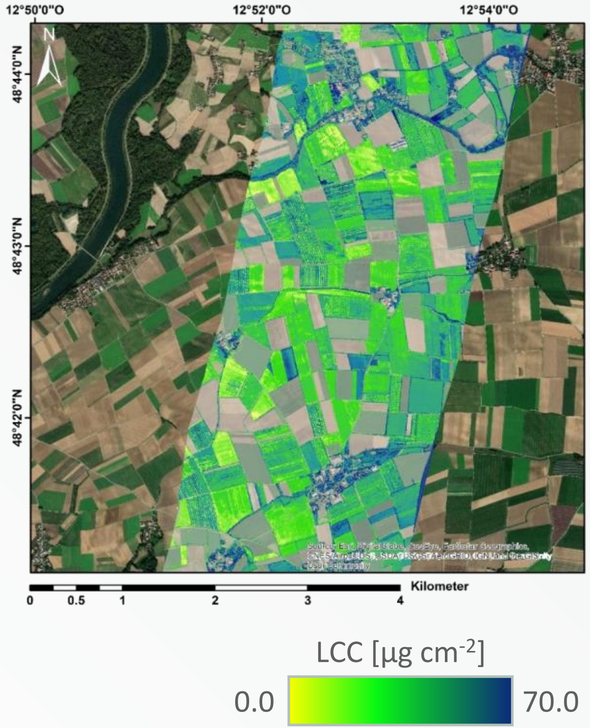
Indirectly drive radiative transfer:
Secondary variables

- Evapotranspiration
- Nitrogen content
- Phenology
- Crop yield
- Water stress
- Irrigation needs
- Crop coefficient

Biophysical and Biochemical Variables
Leaf Level Traits

- Leaf level traits describe the biochemical and morphological properties of leaves, including pigments (chlorophyll a + b, carotenoids, anthocyanins), nitrogen, phosphorus, leaf mass per area, leaf water content, carbon and nonstructural carbohydrates (sugars, starches).
- These traits are mainly involved in photosynthetic processes and carbon uptake.
- Leaf structural compounds include cellulose, fiber, lignin and hemicellulose.
- Further traits are defensive compounds (phenols, condensed tannins), macronutrients with multiple functions (e.g. , Ca, B, Fe, K, Mg, S) and metabolic traits (see Table 12 in Cawse-Nicholson et al. 2021).
- Typically, leaf traits are given in area-based ($\mu g\ cm^{-2}$) or mass-based units (% or $mg\ g^{-1}$).

Image credit: Courtesy of Martin Danner

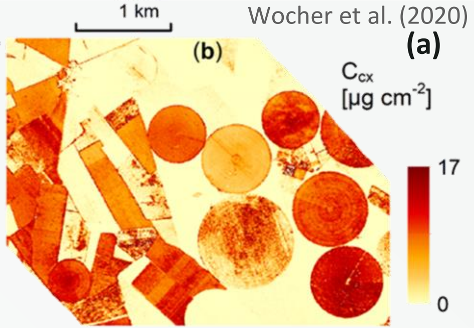


Biophysical and Biochemical Variables

Leaf chlorophyll content (LCC)

- *"In agricultural systems, the accurate spatial mapping of leaf chlorophyll content is important for monitoring vegetation health and plant stress, which can be used to guide fertilizer application in order to optimise crop yield and reduce excessive nutrient loss."* (Croft and Chen 2017)
- Chlorophyll molecules allow the conversion of absorbed solar irradiance into stored chemical energy, through harvesting light energy and supplying electrons to the electron transport chain, leading to the production of NADPH for the reactions of the Calvin–Benson Cycle.
- The amount of solar radiation absorbed by a leaf is largely a function of foliar concentration of photosynthetic pigments. Hence, low leaf chlorophyll content (LCC) limits the photosynthetic capacity and reduces primary productivity of the crops (plants).
- LCC is usually quantified in units of µg chlorophyll per cm^2 (leaf area), or µmol m^{-2} or µg g^{-1} .
- *In situ* measurements of LCC are usually performed non-destructively via handheld devices, e.g. the Konica Minolta SPAD-502Plus:

Biophysical and Biochemical Variables
Leaf carotenoid content (C_{xc})



Wocher et al. (2020)

- *"Carotenoid pigments provide fruits and flowers with distinctive red, orange and yellow colours as well as a number of aromas, which make them commercially important in agriculture, food, health and the cosmetic industries."* (Cuttriss et al. 2011)
- Plants contain a number of different types of carotenoids (C_{xc}), which fall into the subgroups: carotenes or xanthophylls. The most common carotenoid pigments present in leaves are represented by one carotene pigment (b-carotene) and five xanthophylls (lutein, zeaxanthin, violaxanthin, antheraxanthin, and neoxanthin) (Croft & Chen 2017).
- Carotenoids and xanthophylls play an important role in photoprotection, accessory light harvesting and energy transfer (Gitelson et al. 2002; Kong et al. 2017).
- Carotenoids are present in variable proportions during the differentiation and ageing of leaves, but abiotic stress can inhibit carotenoid production (Hank et al. 2019).
- Amount of carotenoids is commonly expressed in different units, e.g., as mass per unit surface area (µg cm^{-2}), or as mass per unit fresh leaf weight (mg g^{-1}).

Biophysical and Biochemical Variables
Leaf anthocyanin content (C_{ant})

- *"Anthocyanins are actively produced as a result of environmental stresses (e.g., low or high temperatures), during senescence and following budburst, before the photosystems are fully developed."* (Gamon & Surfus 1999; Croft & Chen 2017)
- Hence, they can be of interest for precision agriculture, since anthocyanins are typically present when plants suffer from environmental stresses such as drought, freezing, air pollution or nutrient deficiency (Lee & Gould 2002; Springob et al. 2003).
- Anthocyanins are the most common class of flavonoids, i.e. the most widespread red pigments (Hank et al., 2019).
- They are responsible for the orange to red, or purple to blue coloration in the tissue depending on the molecule, temperature and pH value, as it can be found, for instance, in blueberry, raspberry, black rice or black soybean (Tanaka et al. 2008).

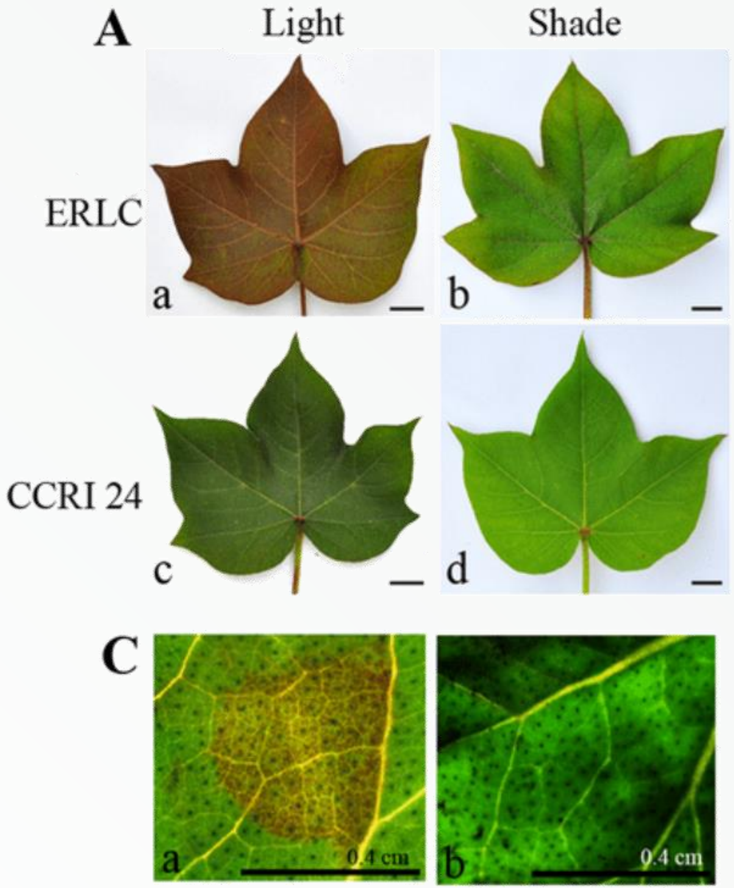


Figure reprinted from Gao et. al. (2013)

Biophysical and Biochemical Variables

Leaf Pigments

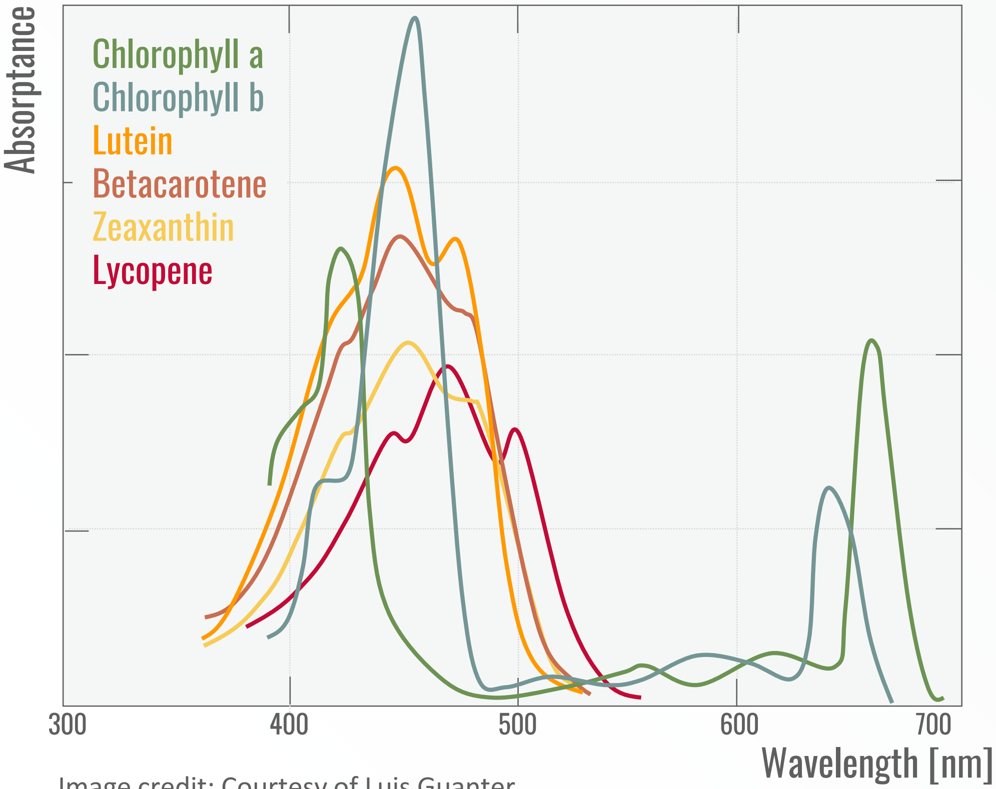
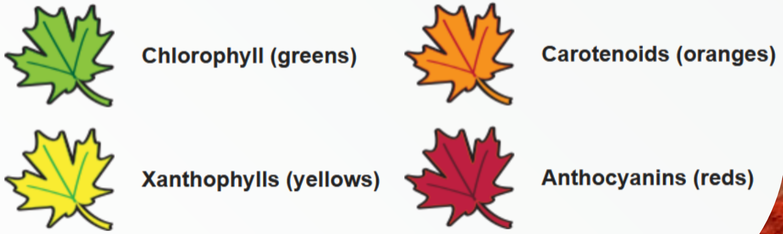


Image credit: Courtesy of Luis Guanter

- Light absorption by pigments in the chloroplast produces a unique absorption pattern in the visible spectrum, with higher absorption in the blue and red wavelengths than in the green wavelengths.
- Photosynthetic pigments, primarily chlorophylls and carotenoids (e.g. lutein, betacarotene, zeaxanthin, lycopene) strongly absorb light.
- Other non-photosynthetic pigments also absorb in this wavelength region, such as anthocyanins (large diverse group of flavonoids creating leaf, flowers and fruit color; Ustin & Jacquemoud 2020).

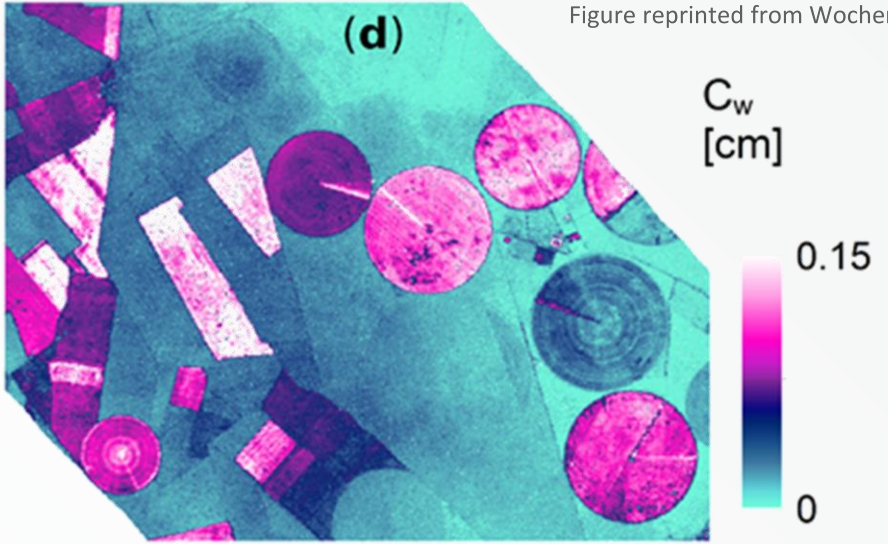


Biophysical and Biochemical Variables

Leaf water content (C_w / EWT)

- *"One aspect of detecting stress in plants from hyperspectral data that has received considerable attention is the measurement of leaf water content."* (Murphy et al. 2019)
- Leaf water content (C_w) or equivalent water thickness (EWT) describes the thickness of a theoretical layer of water (in cm), which absorbs radiation according to the Lambert–Beer law (Nobel 2009).
- Hence, EWT corresponds to the volume of water that is stored within the cells of living vegetation (Hank et al. 2019).

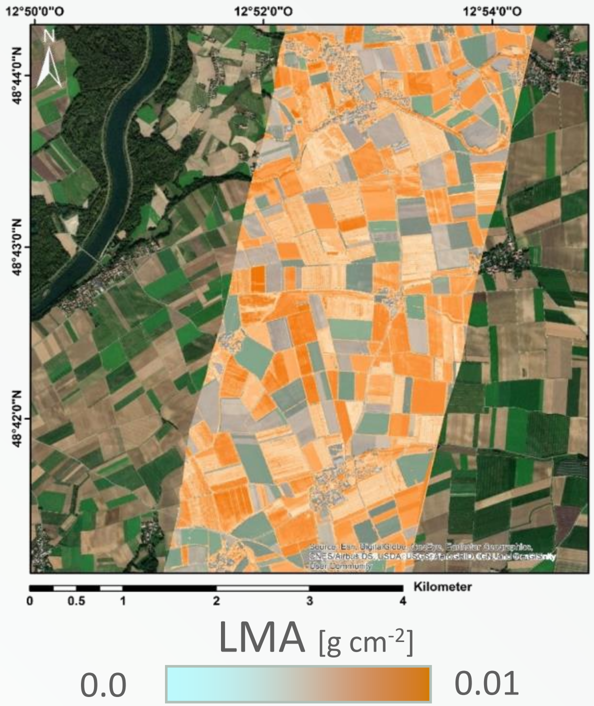
Figure reprinted from Wocher et al. (2020)



- For a remote sensor with a defined field of view, it is difficult to decouple the contributions of leaf water content and LAI. Thus, the total canopy water content per unit ground area (CWC, g m^{-2}), rather than leaf EWT is usually "observed" or retrieved (Clevers et al. 2010).
- CWC is a measure for the moisture state of a canopy, which is of interest for practical farming for the detection of plant water stress.

Biophysical and Biochemical Variables
Leaf mass per area (LMA)

Image credit: Courtesy of Martin Danner



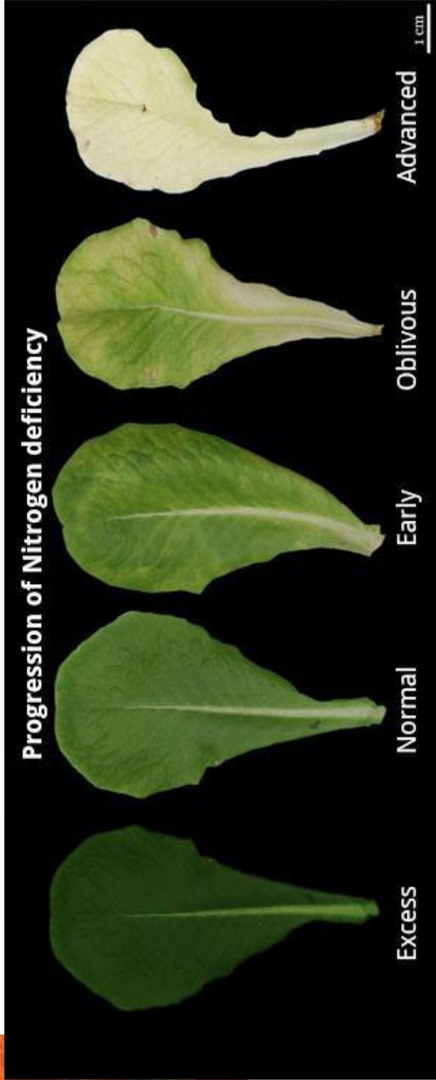
- *"LMA is an essential indicator of plant functioning, including photosynthetic and respiratory rates, chemical composition or resistance to herbivory (de la Riva et al. 2016). The importance of LMA for farming compared to the other variables is therefore rather indirect but nonetheless important, in particular regarding the relationship of LMA to photosynthesis–nitrogen relationships (Poorter & Evans 1998)."*
- Leaf Mass per Area (LMA) denotes the relation of leaf mass to leaf area in a unit of kg dry matter per m^2 or g per cm^2 leaf area.
- LMA is a fundamental leaf functional trait playing a key role in ecosystem modelling (Asner et al. 2011).
- LMA is a measure of the leaf composition: the first leaves developed by a plant at the beginning of its individual growth cycle usually are rather lightweight, so that the area available for the interception of solar radiation expands rapidly during early growth phases. During later development stages, plants tend to invest more biomass into the structural stability of the leaves, causing the LMA to increase over the course of a growing period (Hank et al. 2019).

Biophysical and Biochemical Variables

Leaf protein (C_p) / nitrogen content (N)

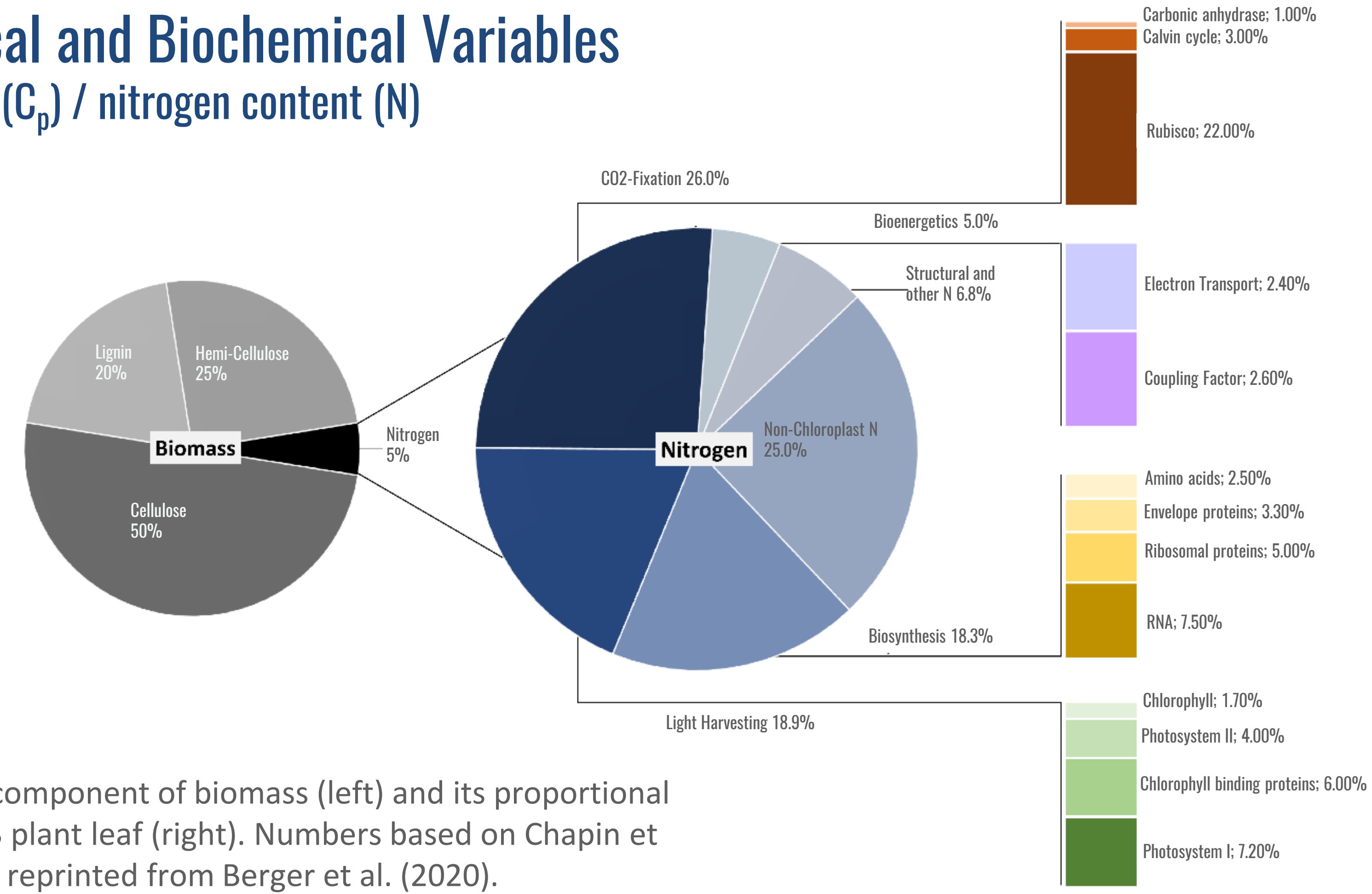
- *„Nitrogen availability enables rapid and early crop growth, increases protein content of crops, facilitates the uptake and utilization of other nutrients such as potassium and phosphorous, improves fruit quality, and controls overall growth of plants [..] Analyzing the N amount in soil and crops and the application of N fertilizer in the event of deficits are essential to improve crop production….“* (Yousfi et al. 2019)
- N is taken up by the roots from soil in the form of $NH4^{+}$ and $NO3^{-}$, and it is a rather small component of leaf dry weight, ranging from 0.3% to 6.4% (Wright et al. 2004).
- A large amount of N is invested in proteins (and chlorophylls) within the leaf cells, with the proteins being the major N containing biochemical constituent of plants (Kokaly et al. 2009).
- Vegetation growth is not a static but a dynamic process of constant nitrogen turnover (Kattge 2002).
- Early in a growing season N is bound in vegetative tissues. During the reproductive phase, N is moved or reallocated from the vegetative organs (leaves) to reproductive structures, such as seeds, ears or fruits (Ohyama 2010).

Image credit: James Hogan



Biophysical and Biochemical Variables

Leaf protein (C_p) / nitrogen content (N)

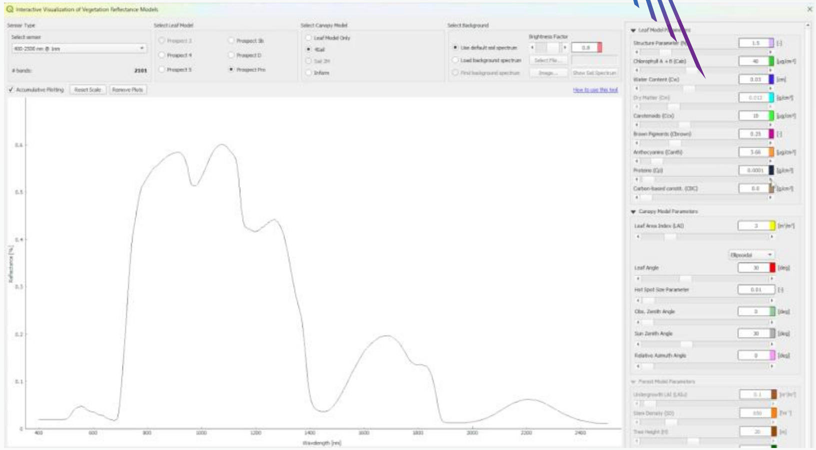


Nitrogen as key component of biomass (left) and its proportional allocation in a C3 plant leaf (right). Numbers based on Chapin et al. (1987). Figure reprinted from Berger et al. (2020).

Biophysical and Biochemical Variables

Carbon-based leaf constituents (CBC)

- *"Lignin from crop residues plays an important role in the soil organic carbon cycling, as it constitutes a recalcitrant carbon pool affecting nutrient mineralization and carbon sequestration. Its function in plants also includes the defence against abiotic and biotic stresses, especially pathogens and insects"* (Frei et al. 2013).
- Carbon-based constituents (CBC) include cellulose, lignin, hemicellulose, sugars and starch. These abundant molecules produced by terrestrial photosynthesis are the main components of non-photosynthetic vegetation (NPV).
- Each constituent of CBC has a specific carbon content (Ma et al. 2018).
- *"Together with the carbohydrate polymers cellulose and hemicellulose, lignin forms the largest portion of "lignocellulosic" plant materials. Thus, lignin accounts for a substantial portion of the total organic carbon in the biosphere, surpassed only by cellulose."* (Frei et al. 2013)



Biophysical and Biochemical Variables
Sun-induced fluorescence (SIF)

- **Concept:** During photosynthesis, plants absorb sunlight in the 400–700 nm spectral range. A small fraction of the energy absorbed is re-emitted at longer wavelengths (650-800 nm) as a faint signal known as sun-induced chlorophyll fluorescence (SIF).
- The strong link between SIF and photosynthesis opens possibilities of inferring gross primary productivity (GPP) (= gross uptake of atmospheric carbon dioxide (CO_2)) through SIF (Schlau-Cohen & Berry 2015; Mohammed et al. 2019).
- Note that the very coarse spatial resolutions of current satellite SIF sensors (40 km GOME-2), and also those in the near future, e.g. FLEX with 300 m, is very limited for agricultural applications, except at regional/global level.

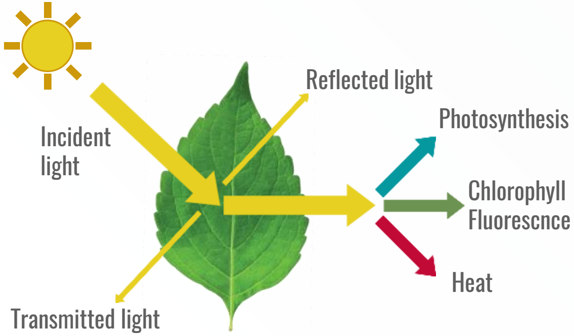


Image reprinted from Sang-O (2017)

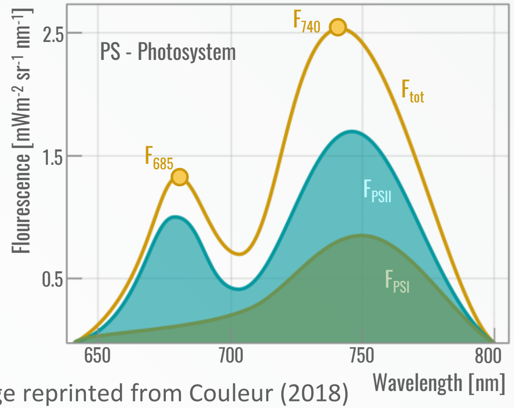


Image reprinted from Couleur (2018)

Biophysical and Biochemical Variables
Canopy Level Traits

Canopy level variables or traits mainly describe the structural properties of a vegetation stand, characterized through:

- Morphology of plants and phytoelements
- Phenology of individual plants
- Vitality of individual plants
- Arrangement and density of the plants
- Composition of plant species (natural vegetation vs. cultivated vegetation)
- Geometry and reflectivity of soil background

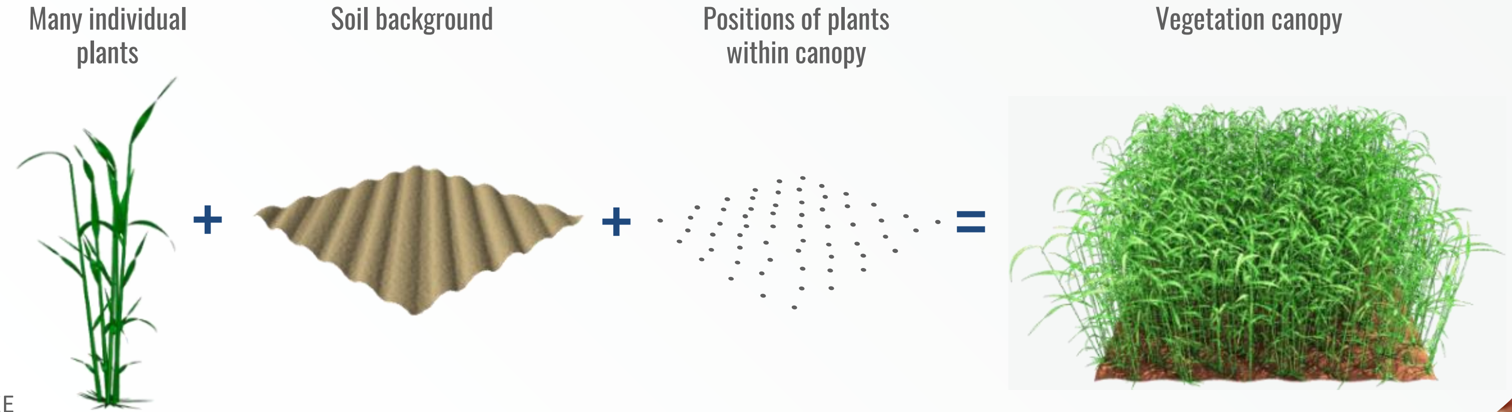
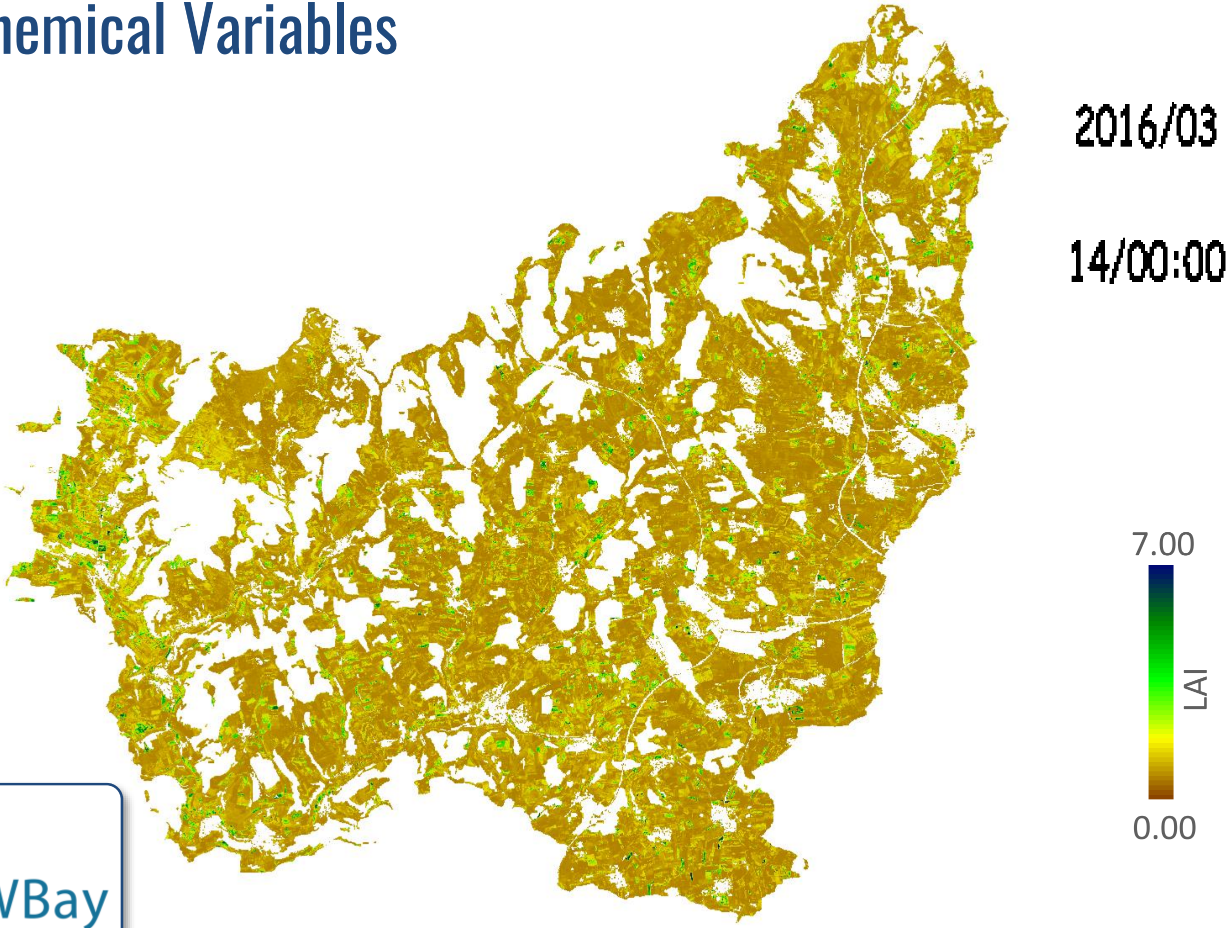


Figure reprinted from Kuester et al. (2014) with permission from IEEE

Biophysical and Biochemical Variables
Leaf Area Index (LAI)

- LAI is the biophysical vegetation trait that attracted most interest in optical remote sensing studies related to agriculture. *"Many applications, including crop growth and yield monitoring, require accurate long-term time series of leaf area index (LAI) at high spatiotemporal resolution with a quantification of the associated uncertainties"* (Yin et al. 2019)
- In general, LAI is defined as half the total leaf area per unit horizontal ground area (Jonckheere et al. 2004), though different LAI definitions exist:
  - **Plant Area Index (PAI)**, accounting for non-green plant elements during the measurements. Note that most indirect methods used to estimate LAI from upward looking canopy transmittance corresponds to PAI rather than LAI!
  - **Green Area Index (GAI)**: accounts for the functioning of the aboveground parts of the plants (crops), which are photosynthetically active during a significant fraction of the growth cycle (Boegh et al. 2002; Duveiller et al. 2012). Very important variable for agriculture & nitrogen content (Verrelst et al. 2014; Amin et al. 2021).
    - **True GAI**: half the developed area of green elements per unit horizontal ground area (destructive measurements).
    - **Apparent GAI (effective LAI, or PAI)**: the value retrieved from remote sensing observations that depends on the (turbid medium) assumptions associated to the estimation algorithm. Effective LAI only considers random positions of leaves, and is referred to the value retrieved from green fraction (gap fraction) measurements based on turbid medium assumption (DHP, LAI2200) (Jonckheere et al. 2004; Richter et al. 2009).

Biophysical and Biochemical Variables
Leaf Area Index (LAI)



Results from the project "VieWBay – Virtual Water Space Bavaria", funded by the Bavarian Ministry of Environment

VieWBay

Biophysical and Biochemical Variables
Leaf inclination angle distribution (LAD)

- Leaf inclination angle distribution (LAD) or average leaf inclination angle (ALIA) is an important characteristic of vegetation canopy structure affecting light interception within the canopy.
- Information of ALIA can be also used as an indicator of water-stress: significant correlations were found between inclination angle and leaf water content in leaflets and petioles of crops (Nagasuga et al. 2013).
- Leaf orientation with respect to the position of the sun is a key factor in determining the amount of light intercepted by a leaf, and also affects the fraction of incident sunlight that penetrates the canopy to lower layers of leaves (Huemmrich 2013).
- Orientation of a leaf is described by its azimuth and inclination angles (in °).

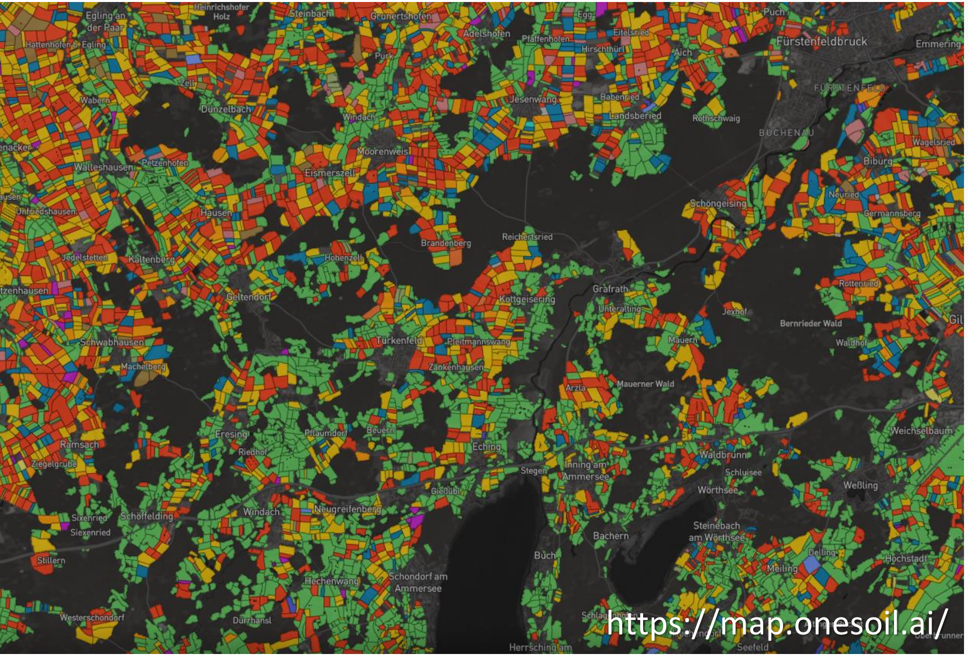
Winter Wheat at noon (12:00)

Maize at noon (12:00)

Maize hourly data…

Biophysical and Biochemical Variables
Crop Mapping

For the retrieval of many of the leaf level and canopy level biophysical and biochemical variables, knowledge about the crop type is crucial as a-priori information.



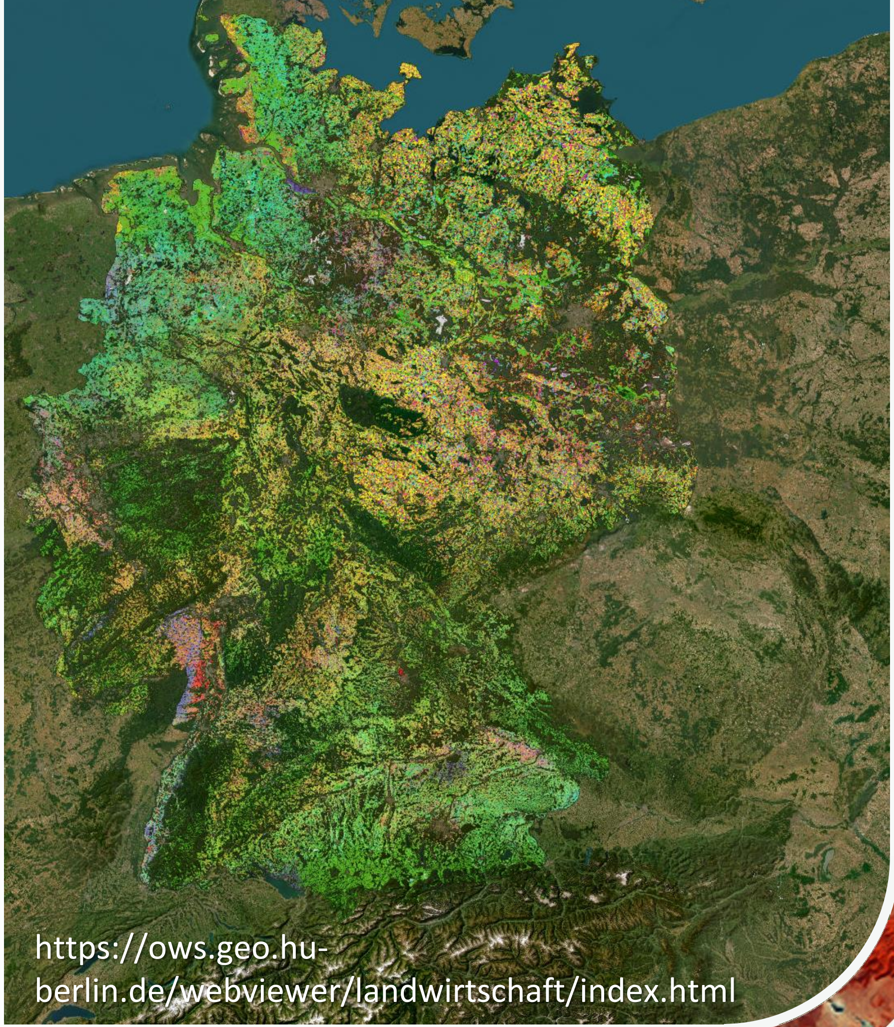
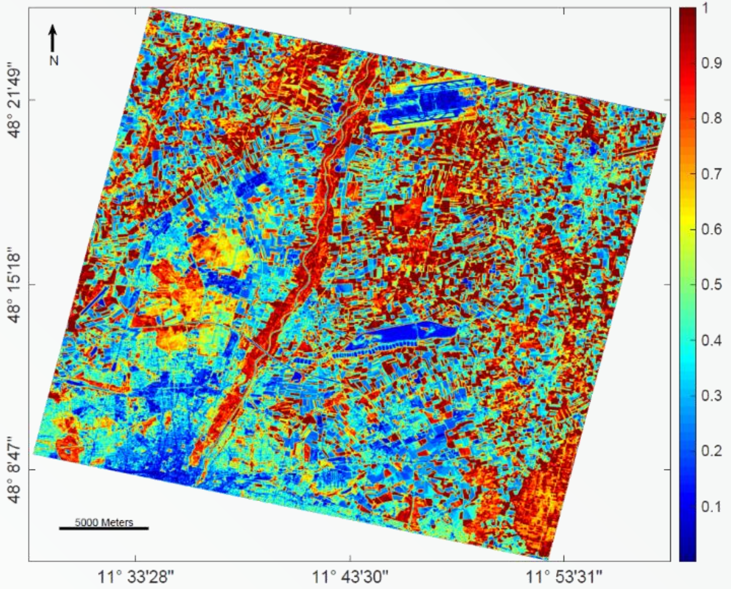


Image credit: Jochem Verrelst / ARTMO

Biophysical and Biochemical Variables

Fractional vegetation cover (fCOVER or FVC)

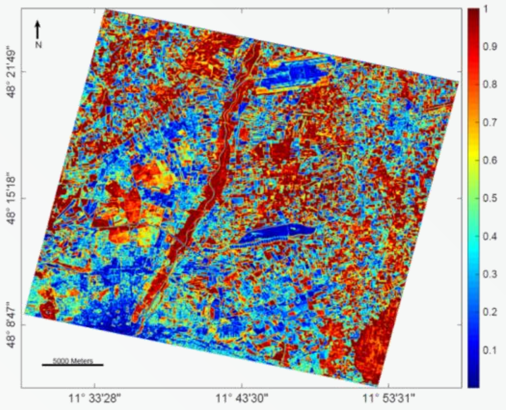


- Fractional vegetation cover (fCOVER or FVC) is useful for various applications in the field of agriculture - ranging from irrigation (e.g., Calera et al. 2001) and crop residues management (e.g., Daughtry et al. 2005) to yield estimations (e.g. Castaldi et al. 2015).
- Green fCOVER is an important biophysical variable describing the Earth's surface. A wide overview is given by Liang & Wang (2020), chapter 12:
- "Fractional vegetation cover is generally defined as the ratio of the vertical projection area of above-ground vegetation organs on the ground to the total vegetation area."
- fCOVER in [%], the fraction of the green vegetation in the nadir direction, is used to separate vegetation and soil in energy balance processes, including temperature and evapotranspiration (Li et al. 2015).
- fCOVER is also known as green ground cover (GGC%) (Zillmann et al. 2015).

Image credit: Jochem Verrelst / ARTMO

Biophysical and Biochemical Variables

Fraction of absorbed photosynthetically active radiation

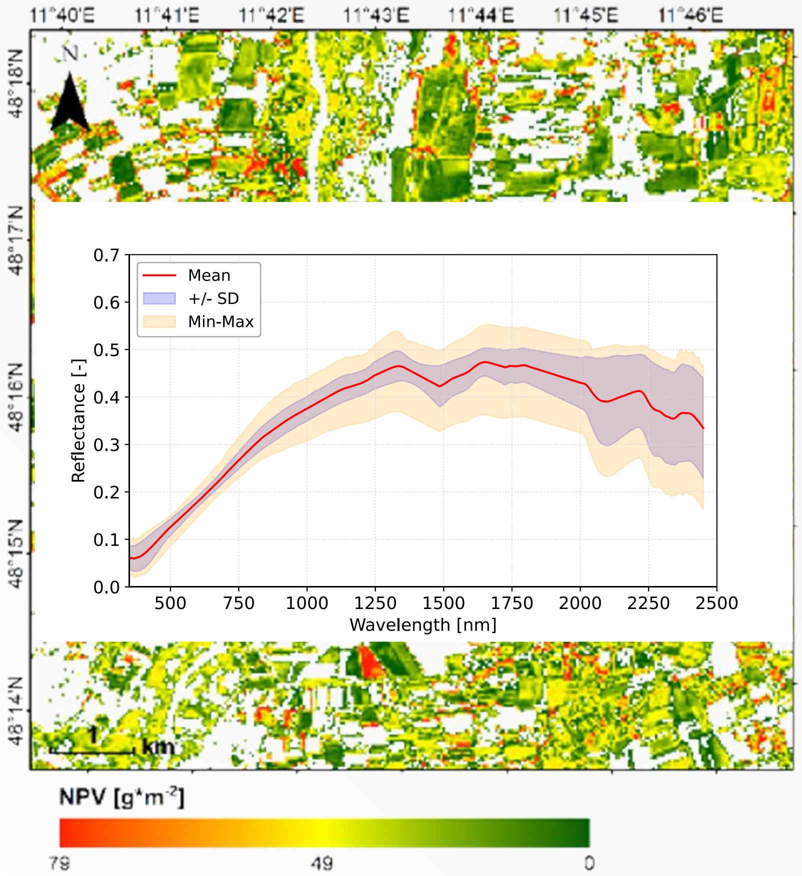


- Remote sensing time-series of fraction of absorbed photosynthetically active radiation (fAPAR) have been *"confirmed to be a reliable tool for regional crop yield forecasting with a strong potential to contribute effectively to operational systems such as those currently running at continental/global level (GIEWS, NASS, FAS, CropWatch or MCYFS)"* (López-Lozano et al. 2015)
- Photosynthetically active radiation (PAR): corresponds to the incoming solar radiation in the spectral range of 400–700 nm.
- Absorbed photosynthetically active radiation (APAR): corresponds to the amount of PAR absorbed by the plant for photosynthesis (Gallo and Daughtry 1986).
- Fraction of absorbed photosynthetically active radiation (fAPAR): is the proportion of PAR absorbed by the plant, expressed as fraction.
- fAPAR is an important biophysical variable in models assessing the primary productivity of vegetation and, more generally, in carbon cycle models between the terrestrial boundary layer and the atmosphere (Vina and Gitelson 2005, Rahmann et al. 2014).
- According to Cawse-Nicholson et al. (2021) fAPAR is not strictly a trait. Though, it was listed alongside others due to its direct relation to primary productivity (Zhang et al. 2012).

Biophysical and Biochemical Variables
Non-photosynthetic vegetation (NPV)

- NPV refer to those plant parts that cannot perform photosynthesis, such as plant litter, crop residues, senescing foliage, branches and stems (Li & Guo 201, Hank et al. 2019).
- For agricultural applications, NPV biomass (e.g. in g/m²) or crop residue cover (%) is particularly interesting since it indicates (seasonal) drought events or other severe injuries.
- Crop residue (CR) cover on the soil surface or a protective mulch:
  - significantly reduces erosion (through wind and water), nutrient loss, evaporation, and soil temperature;
  - reduces soil compaction due to agricultural machinery (Pepe et al., 2020);
  - enhances soil organic C through improvement of the soil structure;
  - crop residues may contain significant amounts of nitrogen and carbon, which enter the soil through ploughing.

Figure reprinted from Berger et. al. (2021)



Biophysical and Biochemical Variable Retrieval

Retrieval Methods

How to translate the complex spectral information into semantic information (which we can understand)?

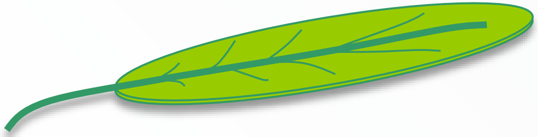
Spectral Information

Regression Models

Parametric

Linear regressions

Semantic Information



Biophysical and Biochemical Variable Retrieval

Parametric Regression Methods

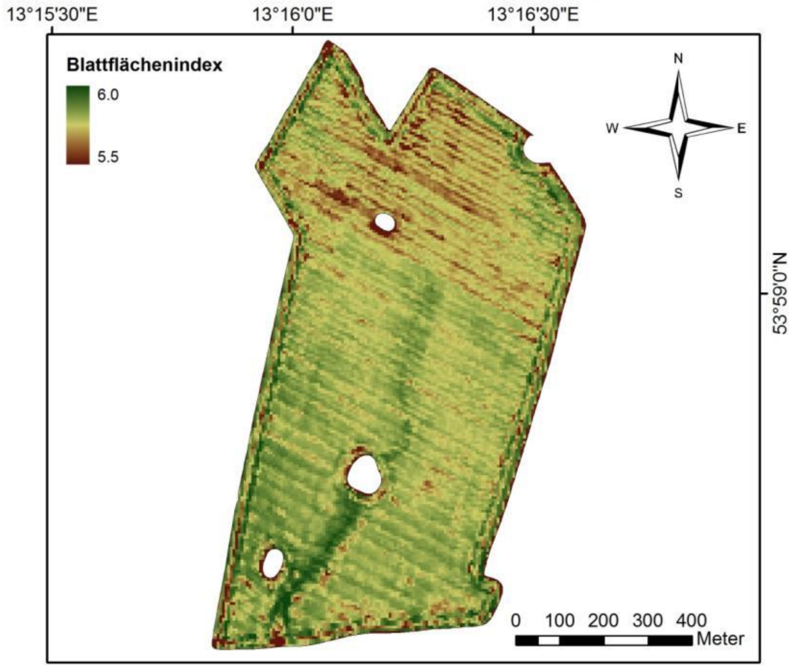
Spectral Information

Dimensionality Reduction

Index

Empirical relationship

Variable



Biophysical and Biochemical Variable Retrieval
Parametric Regression Methods

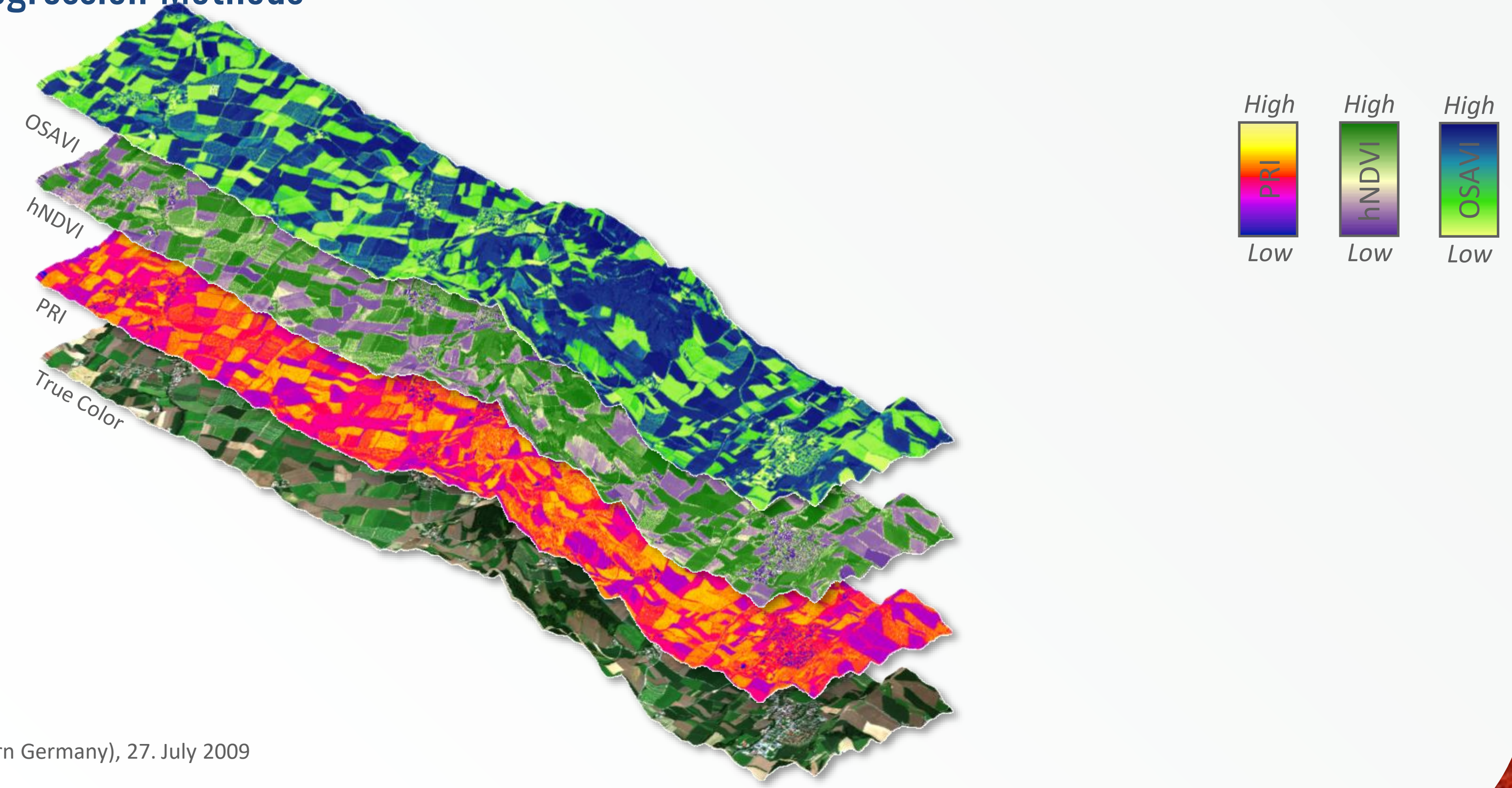
A variety of different indices exist with the help of which the data space can be reduced when focusing on a specific land surface variable.

Widely used examples (in the field of agriculture) are:

OSAVI	Optimized Soil Adjusted Vegetation Index	$OSAVI = \frac{(R_{800nm} - R_{670nm})}{(R_{800nm} + R_{670nm} + 0.16)}$	Rondeaux et al. (1996)
hNDVI	Hyperspectral Normalized Difference Vegetation Index	$hNDVI = \frac{R_{774nm} - R_{677nm}}{R_{774nm} + R_{677nm}}$	z.B. Zarco-Tejada (1999)
PRI_3	Photochemical Reflectance Index (Variation 3)	$PRI_3 = \frac{(R_{570nm} - R_{539nm})}{(R_{570nm} + R_{539nm})}$	Gamon et al. (1992)

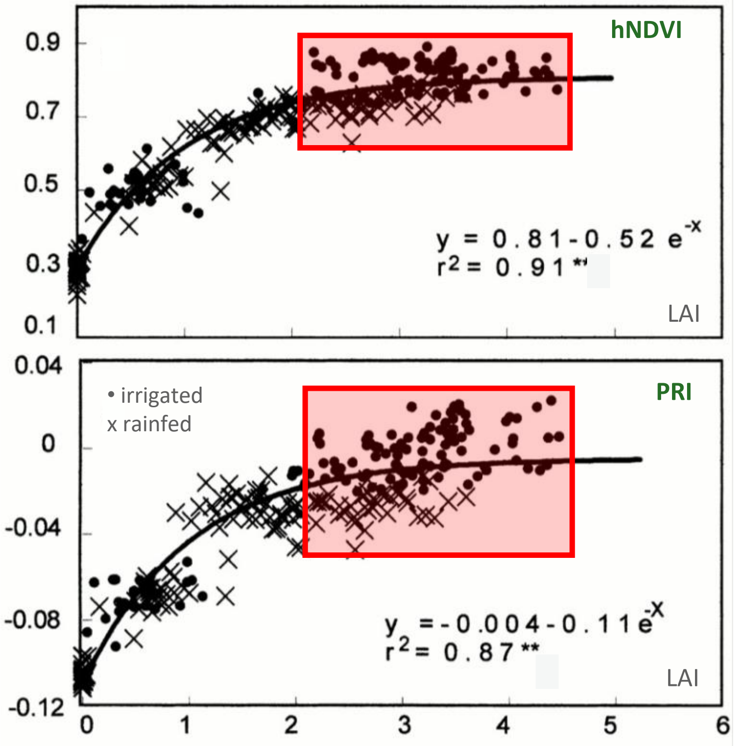
Biophysical and Biochemical Variable Retrieval

Parametric Regression Methods



Example:
HyMap, Neusling (Southern Germany), 27. July 2009

Biophysical and Biochemical Variable Retrieval
Parametric Regression Methods



Example Leaf Area Index:

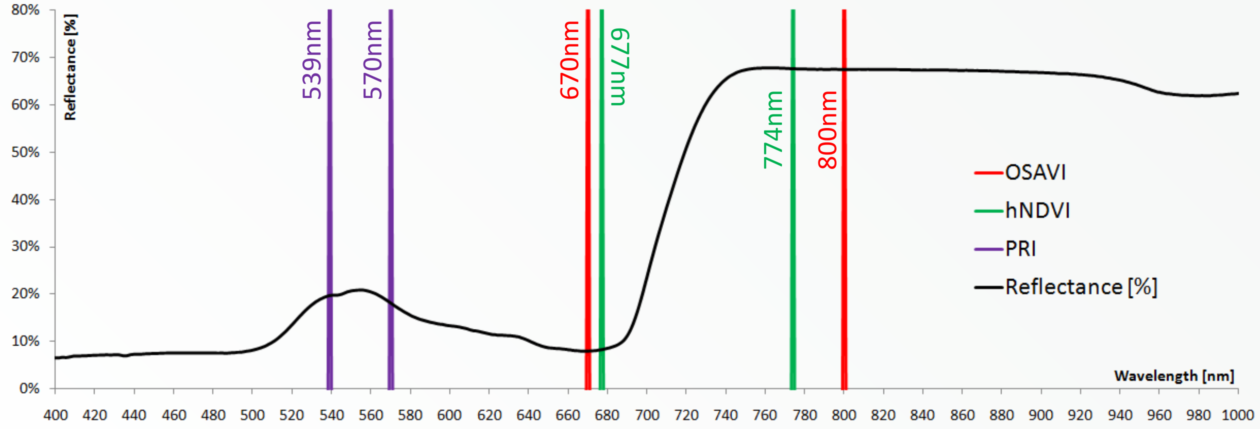
hNDVI and PRI both show a high correlation with the LAI.

From LAI values > 2.0, however, both indices also show strong saturation tendencies...

Figure modified after:
APARICIO et al. (2000): Spectral vegetation indices as nondestructive tools for determining durum wheat yield. *Agron. J.*, 92, pp. 83–91.

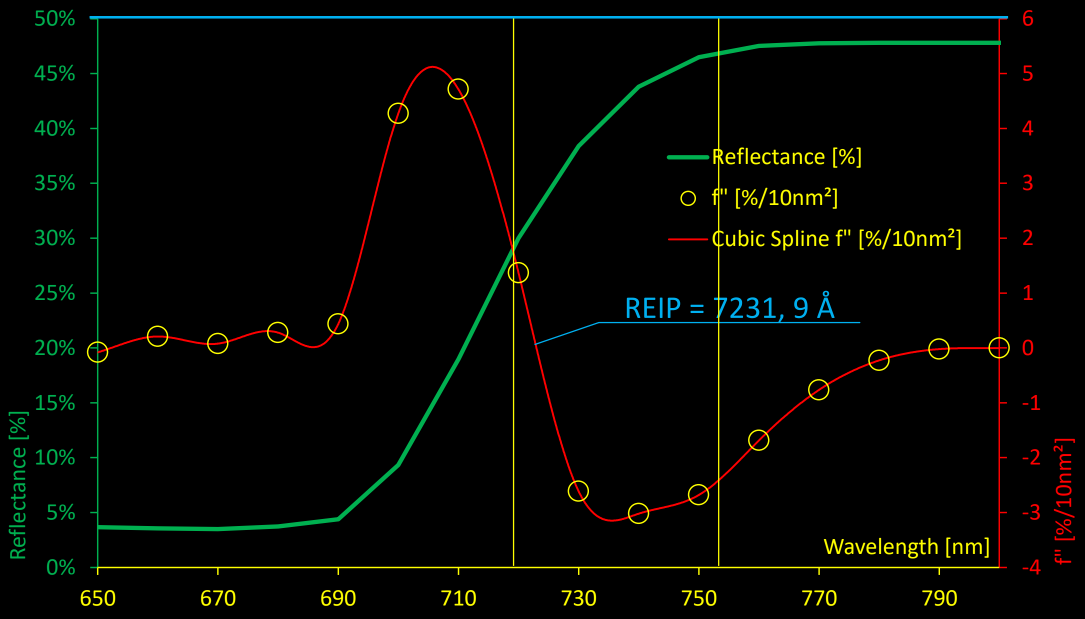
Biophysical and Biochemical Variable Retrieval

Parametric Regression Methods



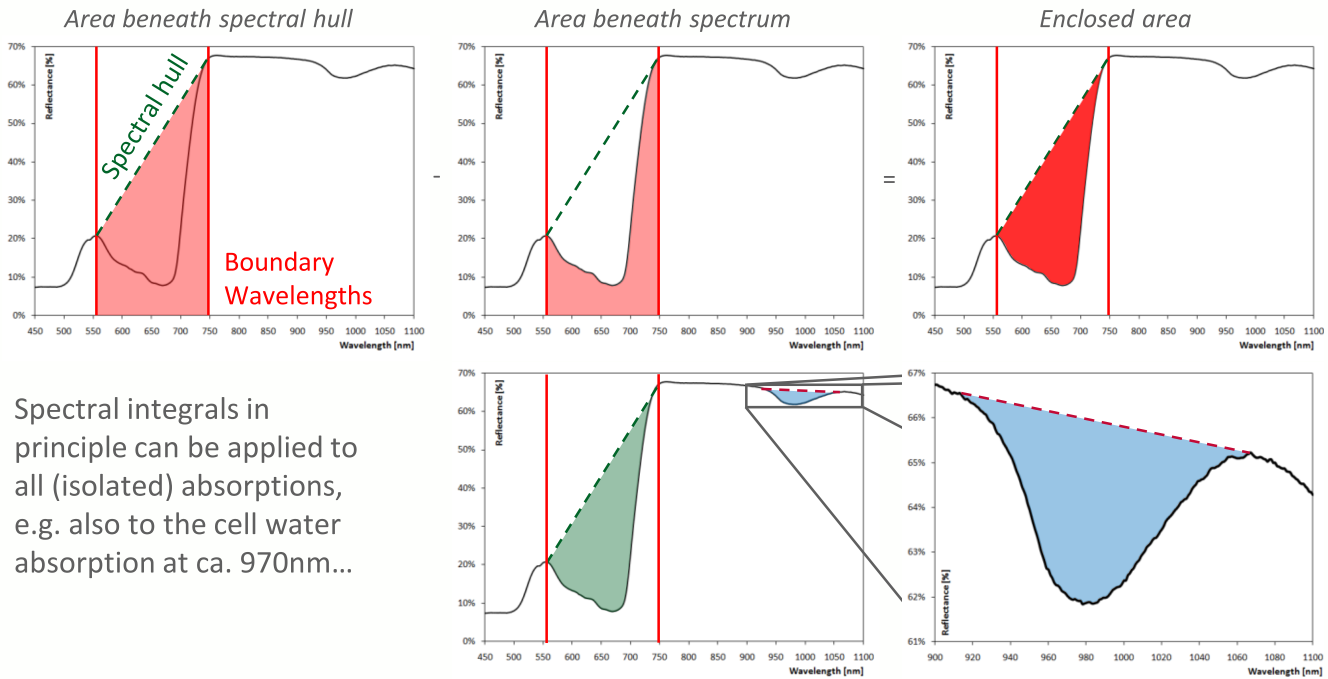
- Hyperspectral vegetation indices make use of the advantage of hyperspectral data by focusing on very narrow bands (low FWHM bandwidth).
- Thus, potentially improved estimation results can be obtained compared to multispectral indices.
- The position of the channels used depends on the variable being investigated.
- However, hyperspectral datasets are characterized not only by narrow channels, but especially by a continuous coverage of the reflectance spectrum...

Biophysical and Biochemical Variable Retrieval
Parametric Regression Methods



Biophysical and Biochemical Variable Retrieval

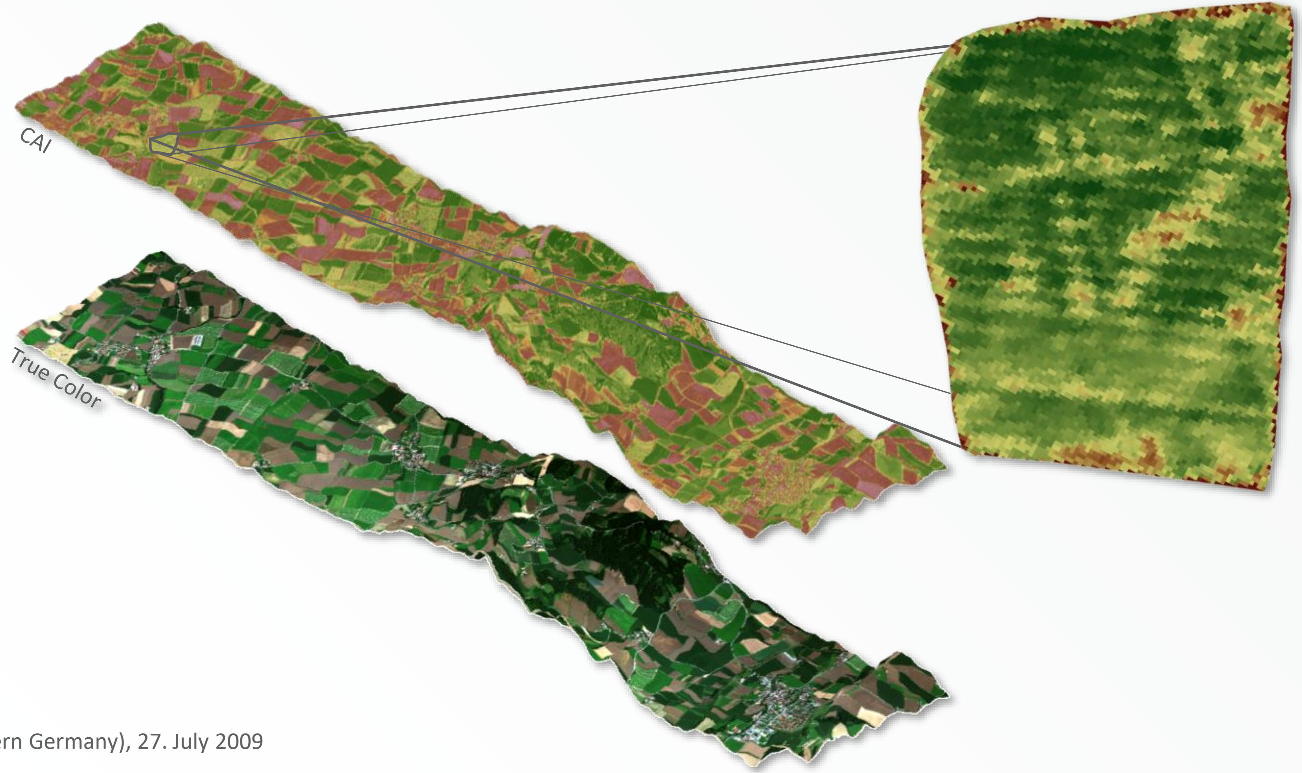
Parametric Regression Methods



Spectral integrals in principle can be applied to all (isolated) absorptions, e.g. also to the cell water absorption at ca. 970nm…

Biophysical and Biochemical Variable Retrieval

Parametric Regression Methods

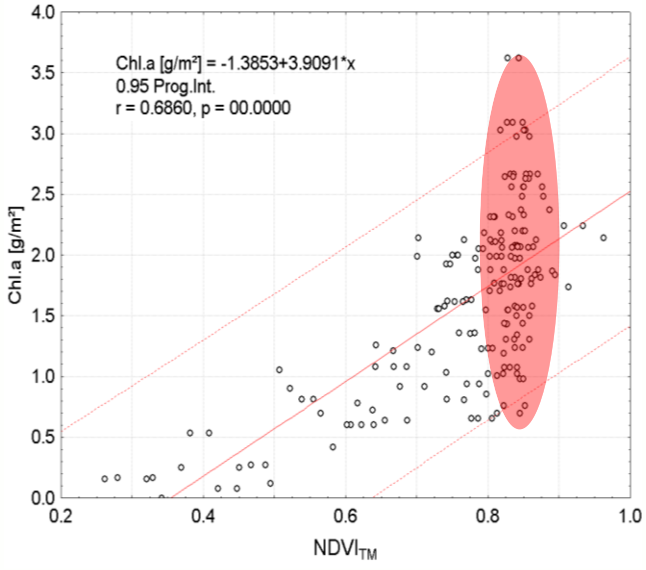


Example:

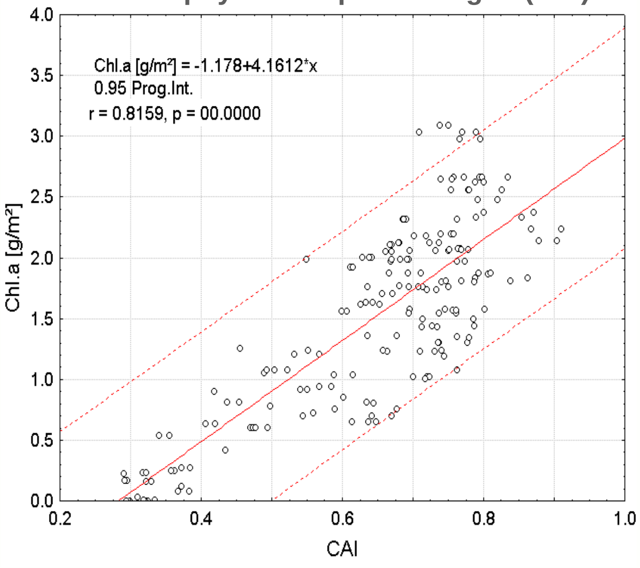
HyMap, Neusling (Southern Germany), 27. July 2009

Biophysical and Biochemical Variable Retrieval
Parametric Regression Methods

Conventional Index: NDVI



Hyperspectral Index: Chlorophyll-Absorption-Integral (CAI)



OPPELT, N. (2004): Hyperspectral Monitoring of Physiological Parameters of Wheat during a Vegetation Period Using AVIS Data. *Int. J. Rem. Sens.*, 25 (1), pp. 145-160.

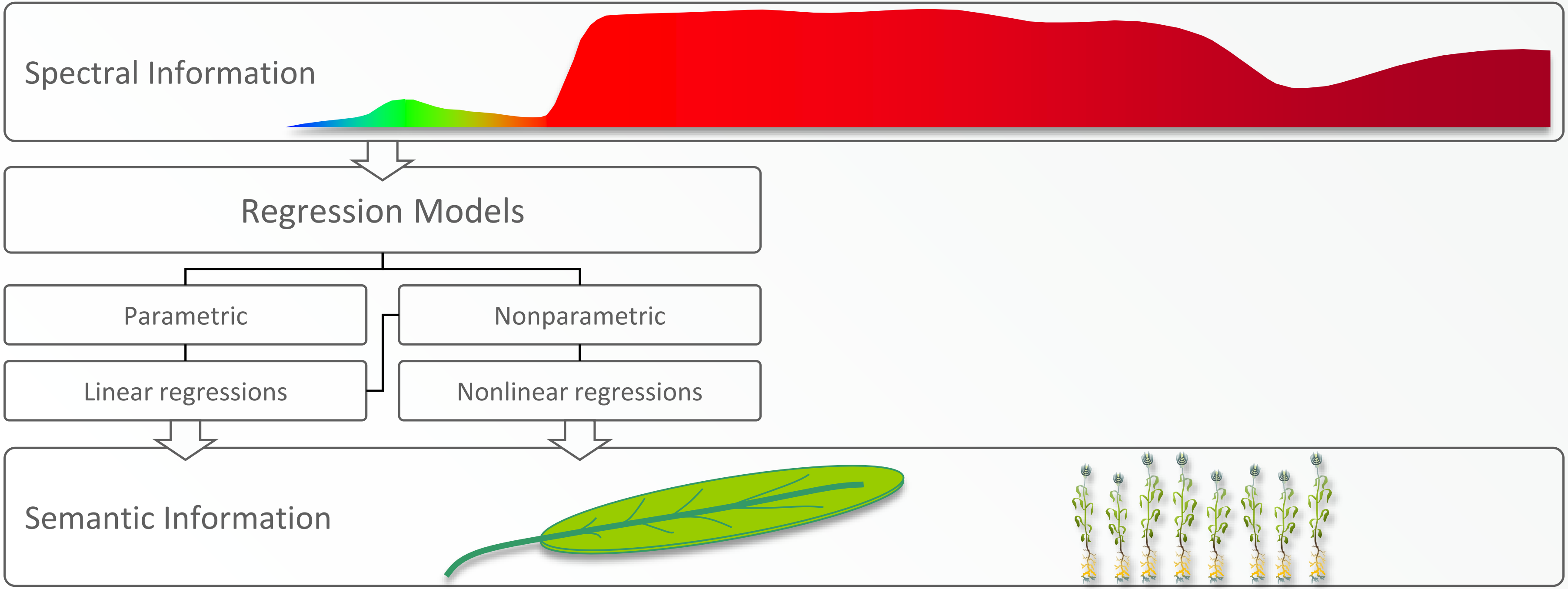
Biophysical and Biochemical Variable Retrieval

Parametric Regression Methods

Strengths	Limitations
▪ Simple and comprehensive regression models ▪ Little knowledge of user required ▪ Easy implementation ▪ Computationally inexpensive: fast in model establishment and mapping	▪ Makes only poor use of the available information within the spectral observation; at most a spectral subset is used. Therefore, they tend to be more noise-sensitive as compared to full-spectrum methods ▪ Parametric regression puts boundary conditions at the level of chosen bands, formulations and regression function ▪ Statistical function accounts for only one variable at a time ▪ Limited transferability to different measurement conditions, sensor & site characteristics, crop and soil type ▪ No uncertainty estimates are provided. Hence the quality of the output maps remains unknown.

Biophysical and Biochemical Variable Retrieval
Retrieval Methods

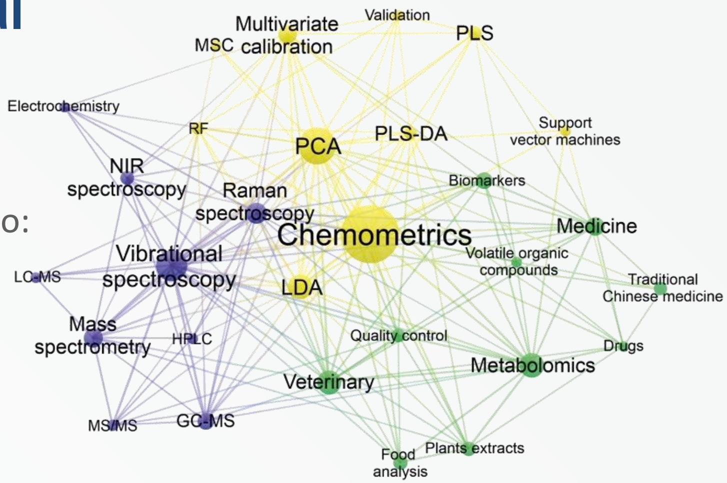
How to translate the complex spectral information into semantic information (which we can understand)?



Biophysical and Biochemical Variable Retrieval

Nonparametric Regression Methods

Figure reprinted from Peris-Díaz & Krezel (2020)



The group of nonparametric regressions can be further subdivided into:

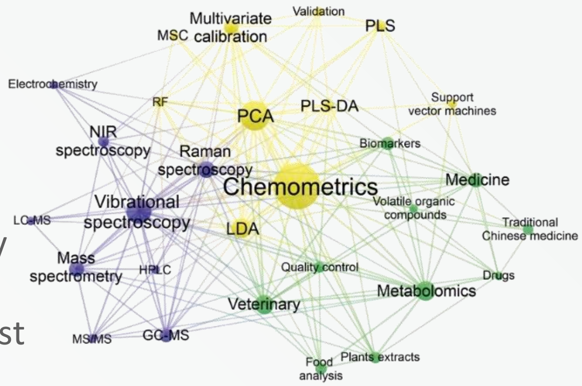
- Linear nonparametric regressions or „chemometrics“ (Lavine and Workman 2013)
- Nonlinear nonparametric regressions, also known as machine learning (ML)

Nonparametric methods directly define regression functions according to information from remotely sensed data. Hence, in contrast to parametric regression methods, a non-explicit choice is to be made on spectral band relationships, transformation(s) or fitting functions (Verrelst et al. 2015).

While in parametric models a finite number of parameters exist, nonparametric models have a (potentially) infinite number of parameters. Hence, nonparametric models exhibit a growing complexity of the model with number of training data.

Figure reprinted from Peris-Díaz & Krezel (2020)

Biophysical and Biochemical Variable Retrieval
Nonparametric Regression Methods



Linear nonparametric regression algorithms, also known as chemometrics, apply linear transformations and have been shown to perform well for retrieving vegetation traits from imaging spectroscopy data, among others, due to their fast performance. These methods also have been used quite frequently since they became standard methods in image processing software packages. Often, dimensionality reduction step is involved: PCR and PLSR are intrinsically based on this principle.

Chemometric methods include, e.g.:

- Stepwise multiple linear regression (SMLR), e.g. Atzberger et al.
- Principal component regression (PCR), e.g. Liu et al. (2017)
- Partial least squares regression (PLSR), e.g. Wold (2001)
- Ridge (regulated) regression (RR) e.g. Geladi and Kowalski (1986)
- Least Absolute Shrinkage and Selection Operator (LASSO) e.g. Tibshirani (1996)

„PLSR is attractive for its straightforward implementation and interpretation, but requires bootstrapping-based methods to estimate and map prediction uncertainties“ (Wang et al. 2019)

Biophysical and Biochemical Variable Retrieval
Nonparametric Regression Methods

- Machine Learning is a subfield of Artificial Intelligence (AI).
- *"The goal of machine learning is to develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest."* (Murphy, 2012).
- ML is closely related to statistics and data mining, but differs slightly in terms of its emphasis and terminology.
- There is a wide variety of ML models for regression and function approximation.
- ML regression algorithms are powerful candidates for the estimation of biochemical & biophysical vegetation traits from imaging spectroscopy data due to their ability to perform adaptive, nonlinear data fitting.
- Note that the term "machine learning" is sometimes also (wrongly…) used for linear regression or chemometric methods.

Biophysical and Biochemical Variable Retrieval
Nonparametric Regression Methods

- Machine learning is based on (often non-parametric and non-linear) regressions.
- So, in principle, it is initially nothing else than working with vegetation indices, but several spectral channels are used at the same time (possibly all of them, whereby a specific so-called band selection/dimensionality reduction, often improves the result, depending on the ML algorithm used).
- Furthermore, with the help of these advanced statistics, it is quasi possible to create several empirical models at the same time (multivariate statistics).
- Thus, multiple input data are linked with multiple target data via statistical models.
- Main difference to simple parametric regression approaches is that in ML methods learning is involved! Algorithms can be based on linear regression, but weights or coefficients of features are iteratively updated with new data until the optimal model is found.
- A large (or a distinctively variable) amount of training data is used for this purpose. Training data could be e.g. spectral signatures and associated biophysical variables.

Biophysical and Biochemical Variable Retrieval
Nonparametric Regression Methods

Families of commonly used ML algorithms include Artificial neural networks, Decision trees, and Kernel-based regressions.

Examples of machine learners commonly used in remote sensing are:

- **SVR: Support Vector Regression** (the values that are positioned in an artificially augmented higher dimensional feature space along a so-called hyperplane define the model).
- **RFR: Random Forest Regression** (ensemble algorithm based on multiple decision trees).
- **ANN: Artificial Neural Networks** (generic term for all algorithms that learn in multiple layers of linked neurons, often e.g. MultiLayer Perceptron Network Regression).
- **GPR: Gaussian Processes Regression** (provide a probabilistic approach for learning generic regression problems with kernels and thus additionally provide confidence intervals → uncertainty assessment).

Biophysical and Biochemical Variable Retrieval
Nonparametric Regression Methods

In addition to feeding training data, all of these algorithms still can/need to be parameterized. The parameters are different for different learners. For SVR, for example, the user can optimize the model using two “tuning screws":

- Kernel Coefficient "Gamma" or γ describes the "effect" of the training data:
    - High gamma value: the algorithm follows the training data very closely, more complex relationships can be mapped.
    - Low gamma value: The model remains more "agile" and not so rigidly bound to the training data, danger of so-called overfitting is reduced.

- C parameter defines the distance between support vectors and hyperplane:
    - High C-value: the result is very accurate, but it may not be possible to convert all input values into results.
    - Low C-value: The probability that many different input values can be converted into results increases, the accuracy of the estimate can decrease.

Biophysical and Biochemical Variable Retrieval

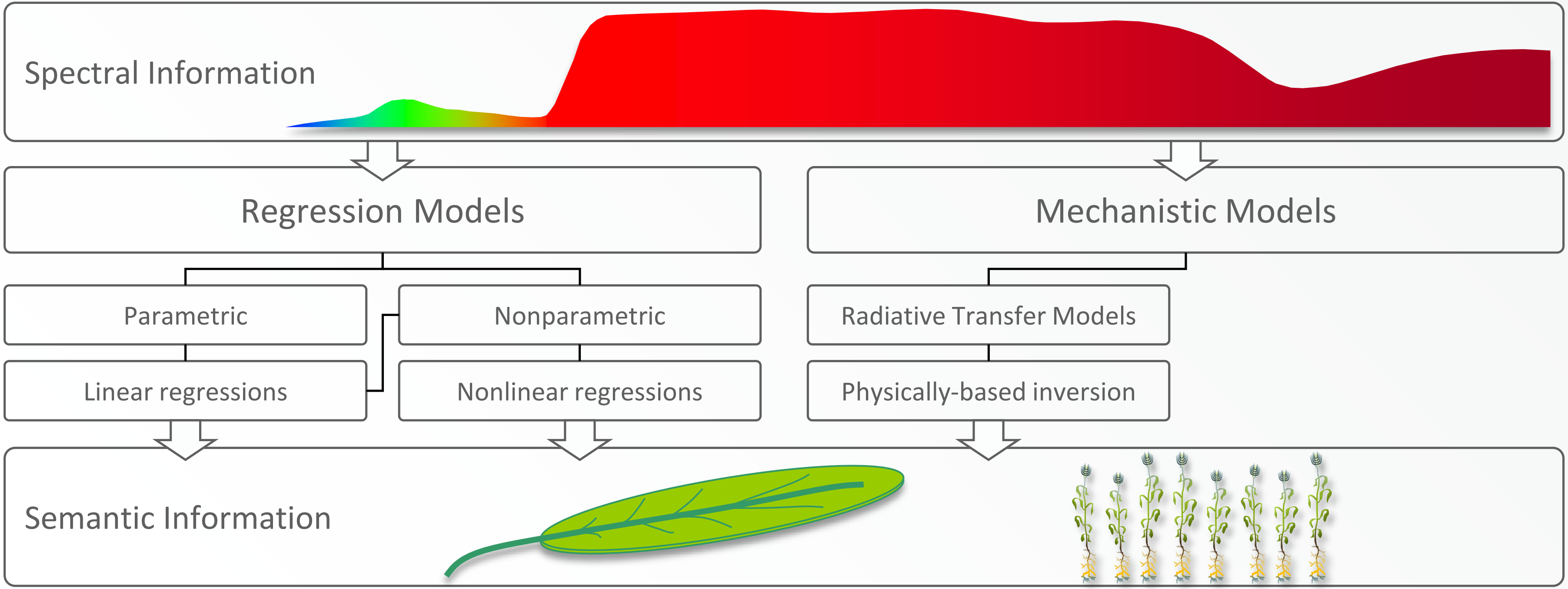
Nonparametric Regression Methods

Strengths	Limitations
▪ Full-spectrum methods: making use of the complete spectral information. ▪ Advanced, adaptive (non-linear) models are built. ▪ Methodologically, accurate and robust performance is enabled. ▪ Some ML regressions cope well with datasets showing redundancy and high noise levels. ▪ Once trained, imagery can be processed time-efficiently. ▪ Some of the nonparametric methods (e.g. ANNs, decision trees) can be trained with a high number of samples (typically >1,000,000). ▪ Some ML methods provide insight in model development (e.g. GPR: relevant bands; decision trees: model structure). ▪ Some statistical models can provide multiple-outputs (e.g. PLRS, ANN, SVR, GPR and KRR) ▪ Some ML methods provide uncertainty intervals (e.g. GPR)	▪ Training can be computationally expensive. ▪ Hypercomplex models can be generated. Their generic potential is limited and hence they do not generalize well, based on the training data, also known as problem of overfitting. ▪ Models' performance on new unseen data depends on the appropriate design of the training data set: representativeness! ▪ Some regression algorithms are difficult (or even impossible) to train with a high number of samples. ▪ Expert knowledge is required, e.g. for tuning. However, toolboxes exist automating some of the steps in this sub-process. ▪ Some regression algorithms elicit instability when applied with datasets statistically deviating from the datasets used for training: underfitting.

Biophysical and Biochemical Variable Retrieval

Retrieval Methods

How to translate the complex spectral information into semantic information (which we can understand)?

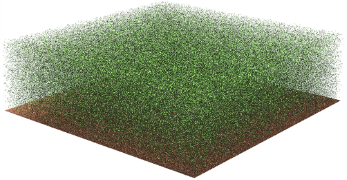


Biophysical and Biochemical Variable Retrieval
Radiative Transfer Models

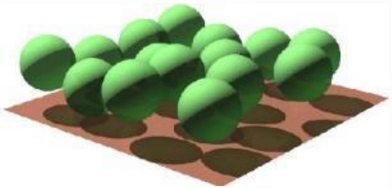
Mechanistic principles of spectral and spatial scaling effects can be analysed by means of physically-based radiative transfer models (RTM). RTMs describe the interaction of photons with biophysical and biochemical plant properties by means of physical laws.

RTMs are widely used in remote sensing science for retrieval (inversion) but also for the development of parametric regressions, generation of training data for nonparametric regressions and beyond that for designing new Earth Observation missions.

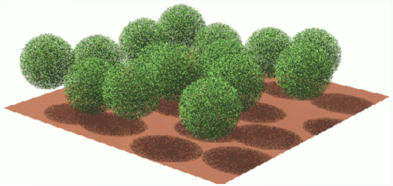
Exemplary groups of RTMs:



Turbid medium*
(e.g. SAIL, Verhoef 1984)



Geometric optics*
(e.g. Chen & Leblanc 1997)



Hybrid model (GO + TM)*
(e.g. DART, INFORM)

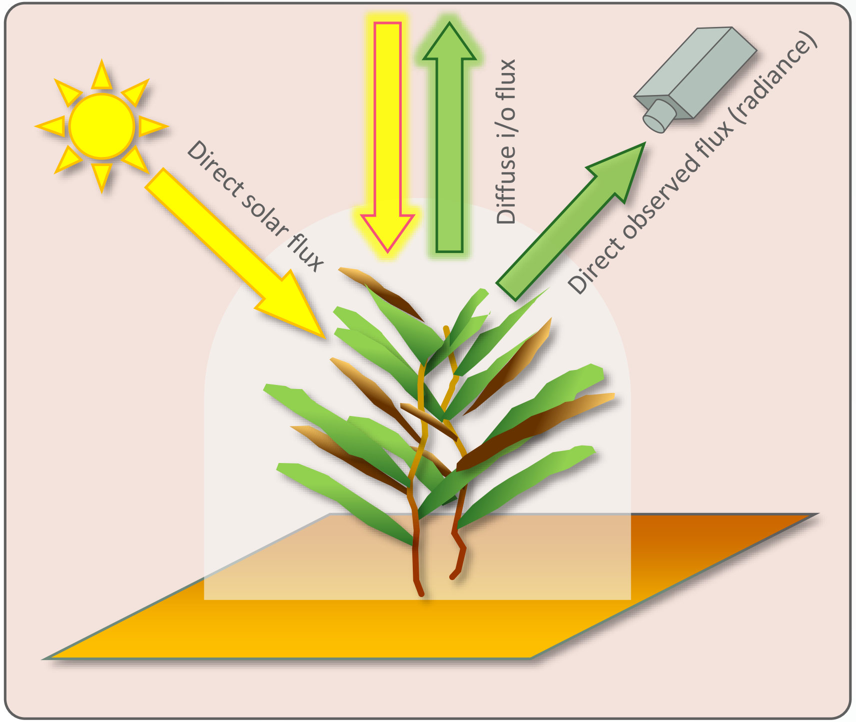
Computer graphic model
(e.g. drat, Raytran, FLIGHT)

* Figures from http://rami-benchmark.jrc.ec.europa.eu/HTML/RAMI3/RAMI3.php

Detailed descriptions & further classifications of RTMs can be found in Verrelst et al. (2019a) and Malenovský et al (2019).

Biophysical and Biochemical Variable Retrieval

Radiative Transfer Models



Input parameters to 4SAIL2:

LAI – Leaf Area Index
Average Leaf slope Parameter a
LIDF bimodality parameter b
Hot spot parameter q
Fraction brown leaf area fB
Layer dissociation factor D
Soil BRDF parameters (b, c, B0, h)
Soil moisture
Crown coverage

structural

Outputs from PROSPECT:

Fraction diffuse sky irradiance
Dry soil reflectance

spectral

Solar zenith angle
Viewing zenith angle
Relative azimuth angle

observational

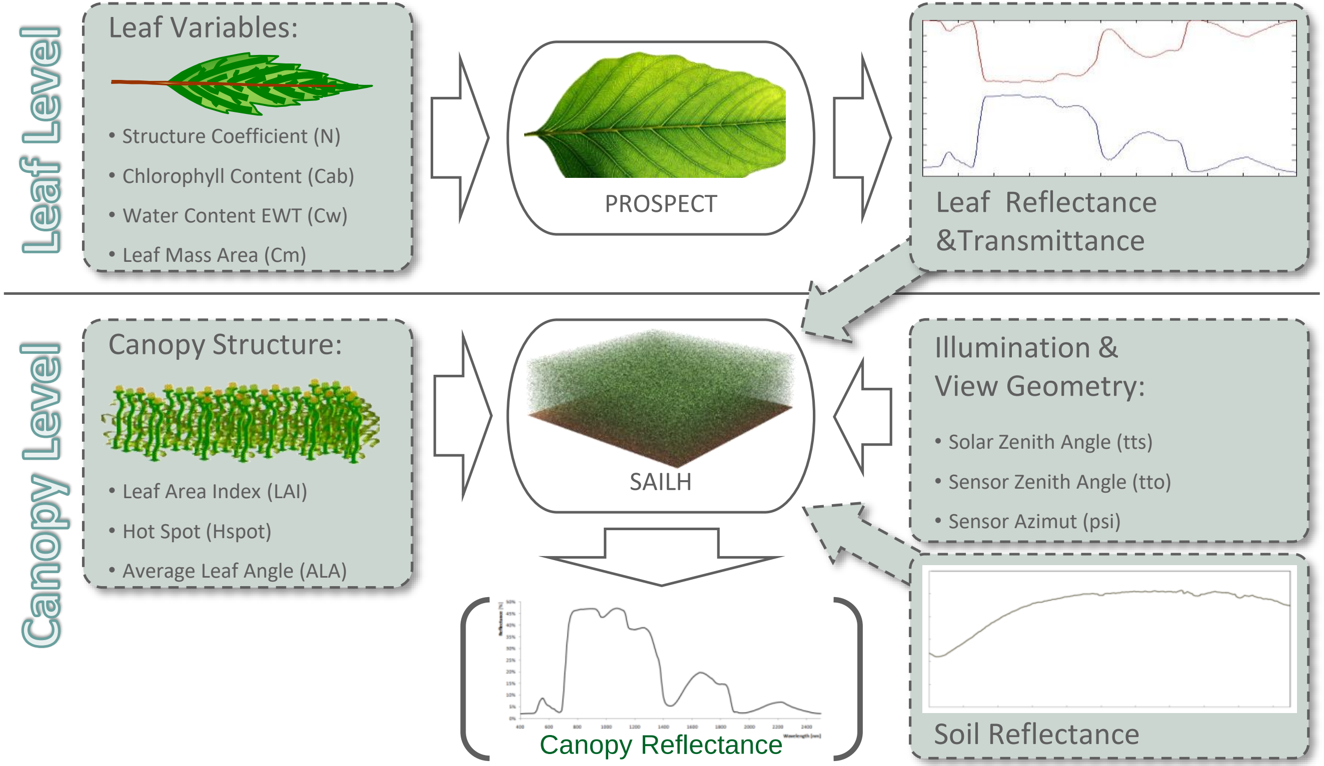
Input parameters to PROSPECT:

Leaf Chlorophyll
Leaf Water
Leaf Dry Matter
Leaf Mesophyll Structure N

Leaf

Biophysical and Biochemical Variable Retrieval

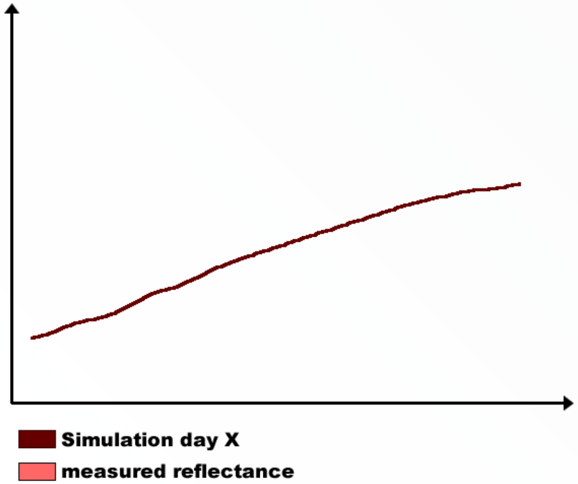
Radiative Transfer Models



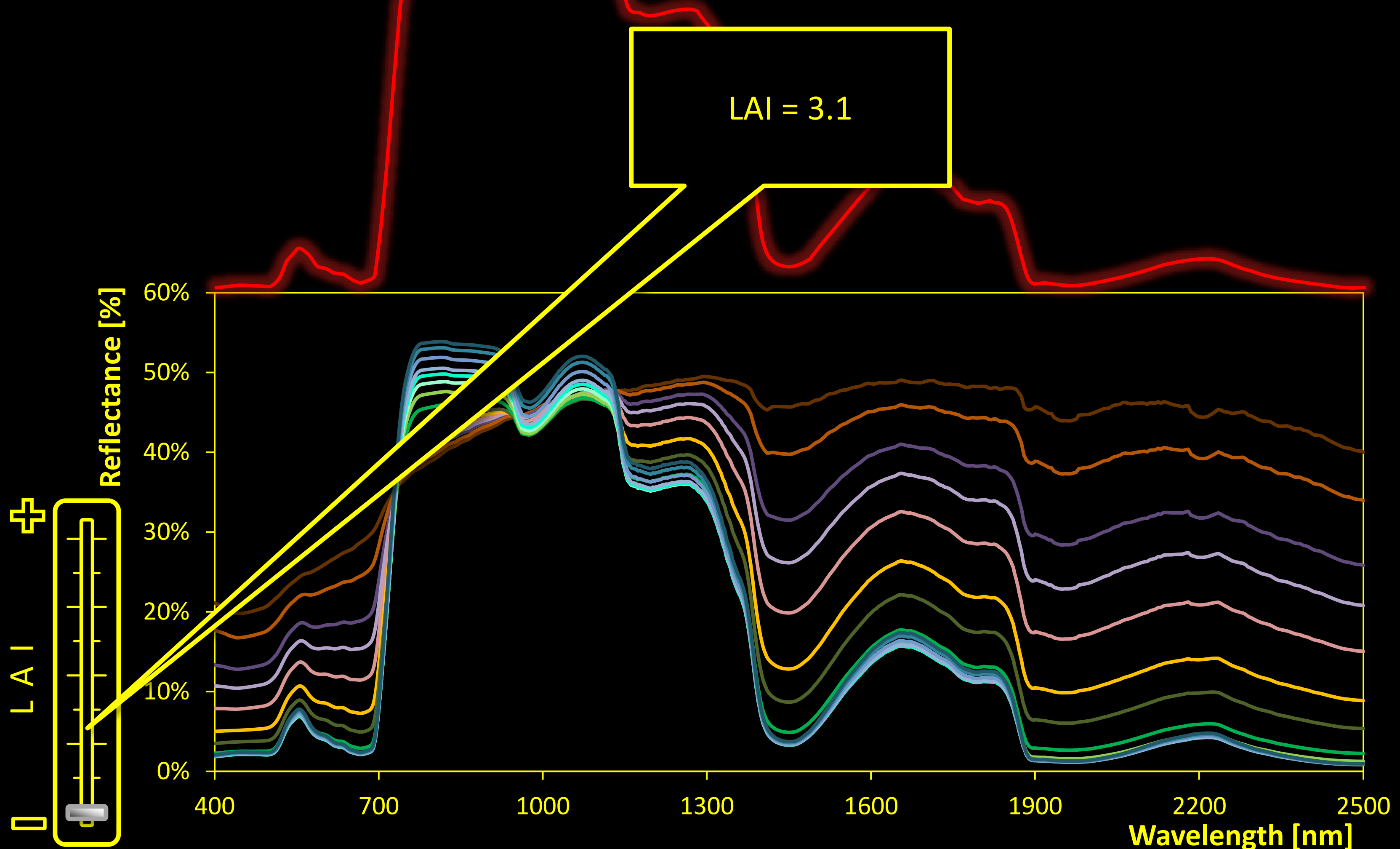
Biophysical and Biochemical Variable Retrieval

Radiative Transfer Model Inversion

Data assimilation

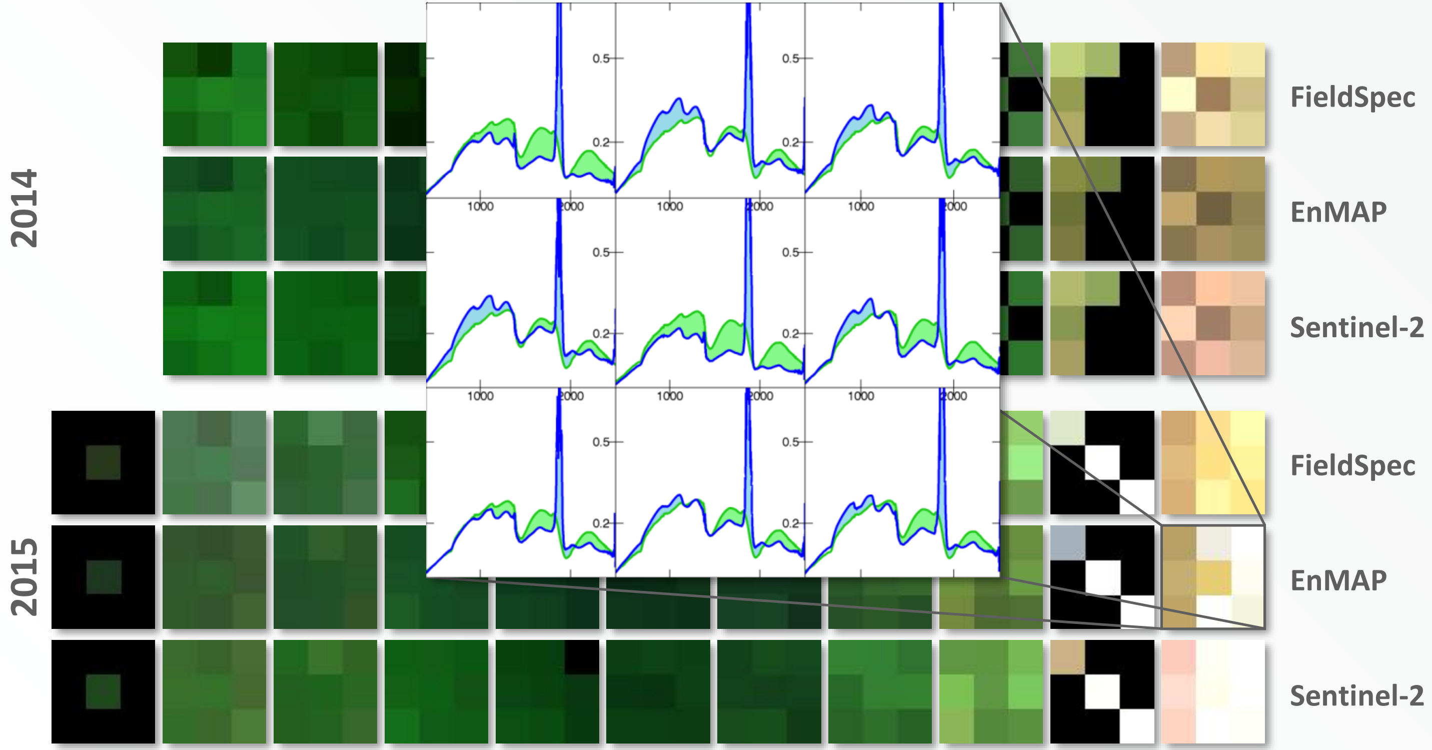


Animation credit:


LAI = 3.1
Reflectance [%]
60%
50%
40%
30%
20%
10%
0%
400
700
1000
1300
1600
1900
2200
2500
Wavelength [nm]
L A I

Biophysical and Biochemical Variable Retrieval

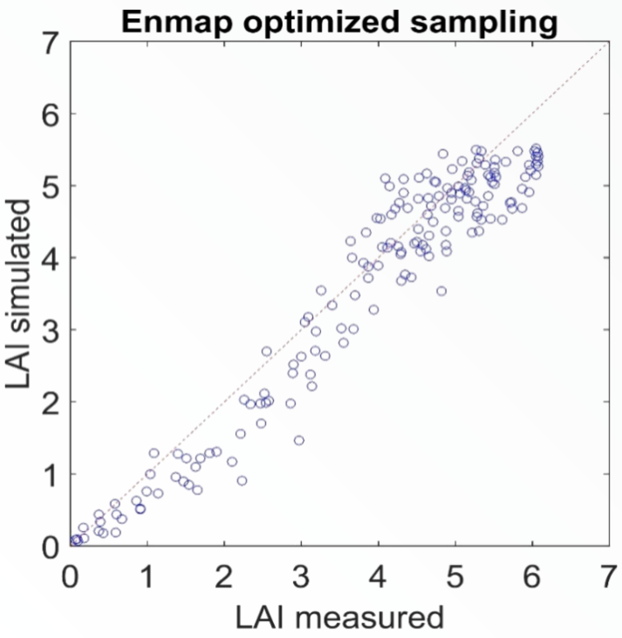
Radiative Transfer Model Inversion



Biophysical and Biochemical Variable Retrieval

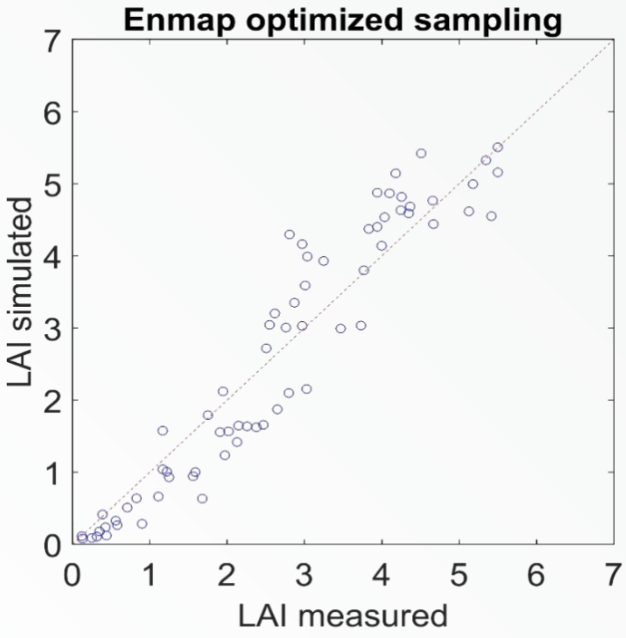
Radiative Transfer Model Inversion

Maize



N bands =42
RMSE = 0.54
R^2 = 0.93

Soybean



N bands =64
RMSE = 0.55
R^2 = 0.91

Biophysical and Biochemical Variable Retrieval

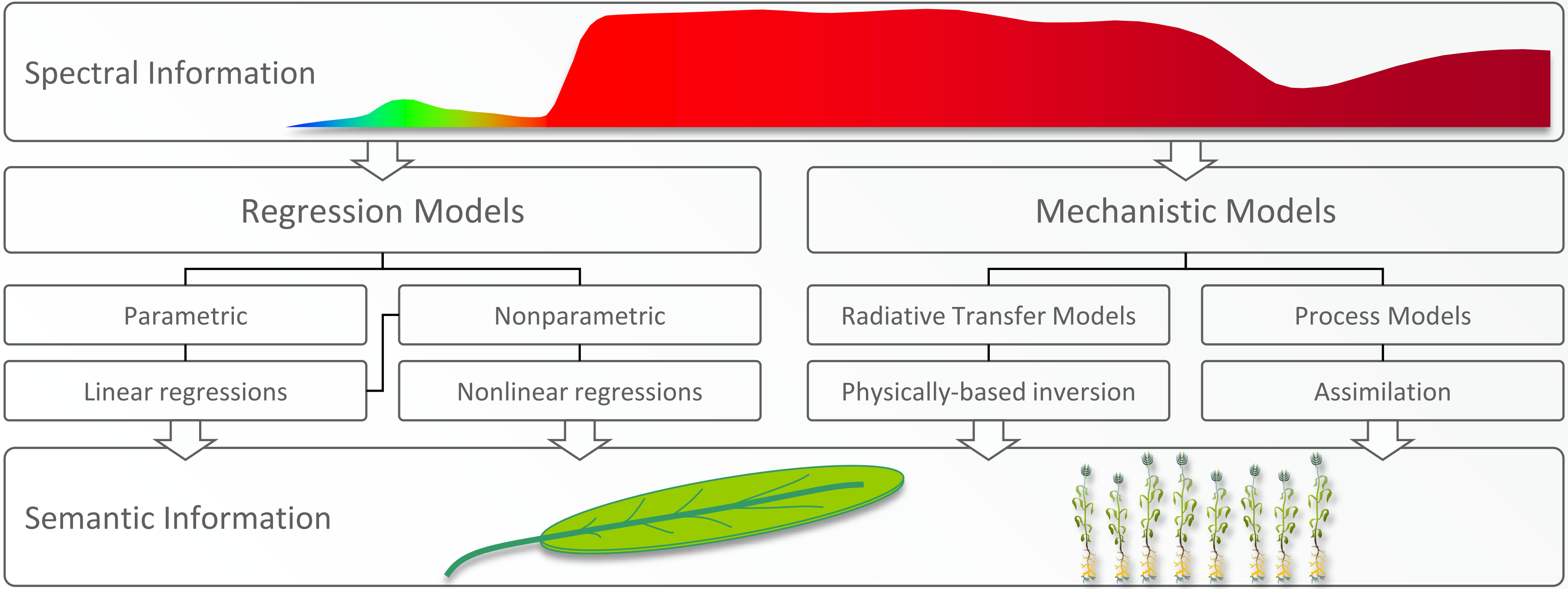
Radiative Transfer Model Inversion

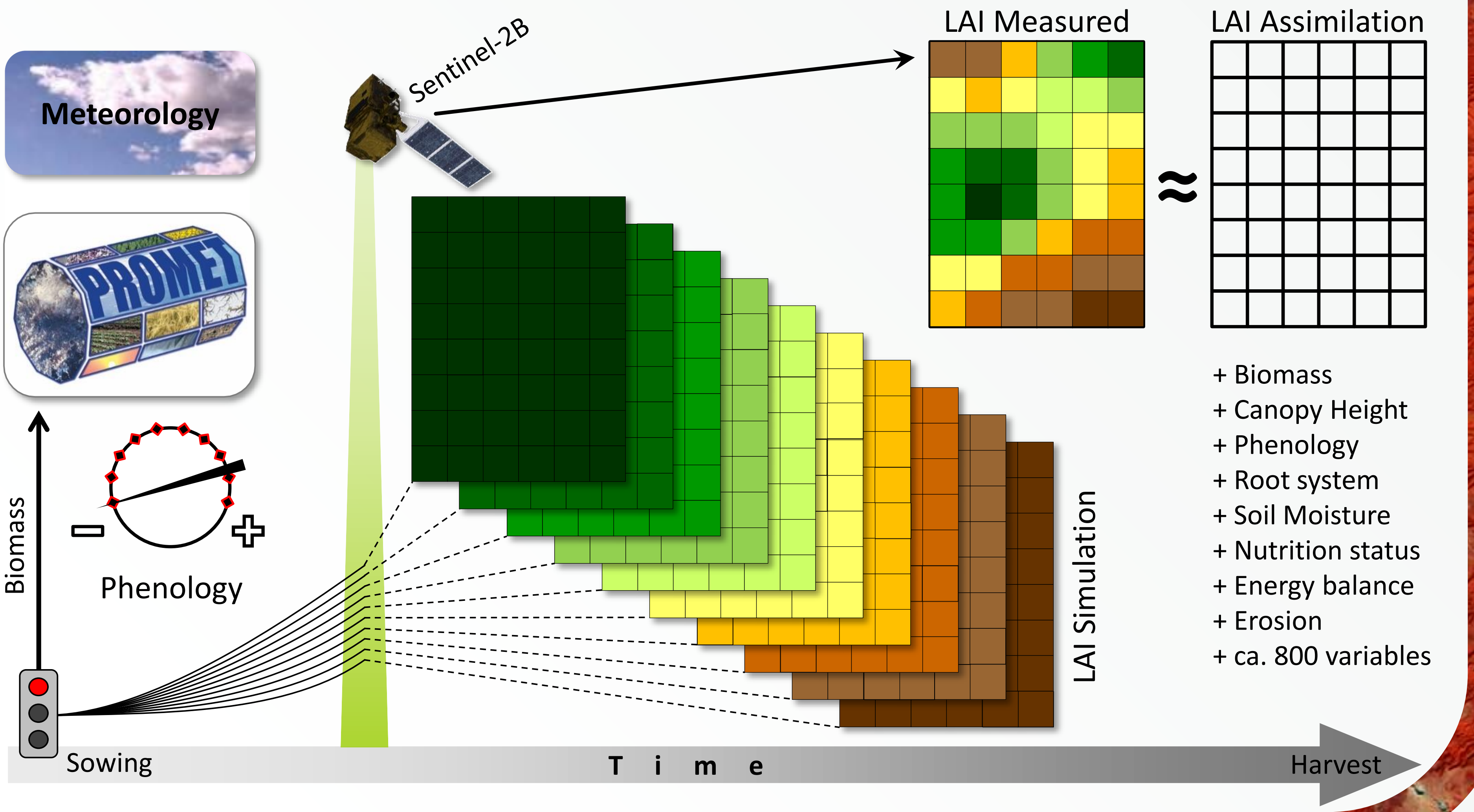
Strengths	Limitations
▪ Full-spectrum methods: making use of the complete spectral information. ▪ Spectral signature is linked to physical processes → we can learn from the models! ▪ Potentially transferable in space and time	▪ Compromise between complexity/accuracy and "invertibility" ▪ Computationally expensive ▪ Equifinality problems ▪ Parameterization (constraining) requires a-priori information

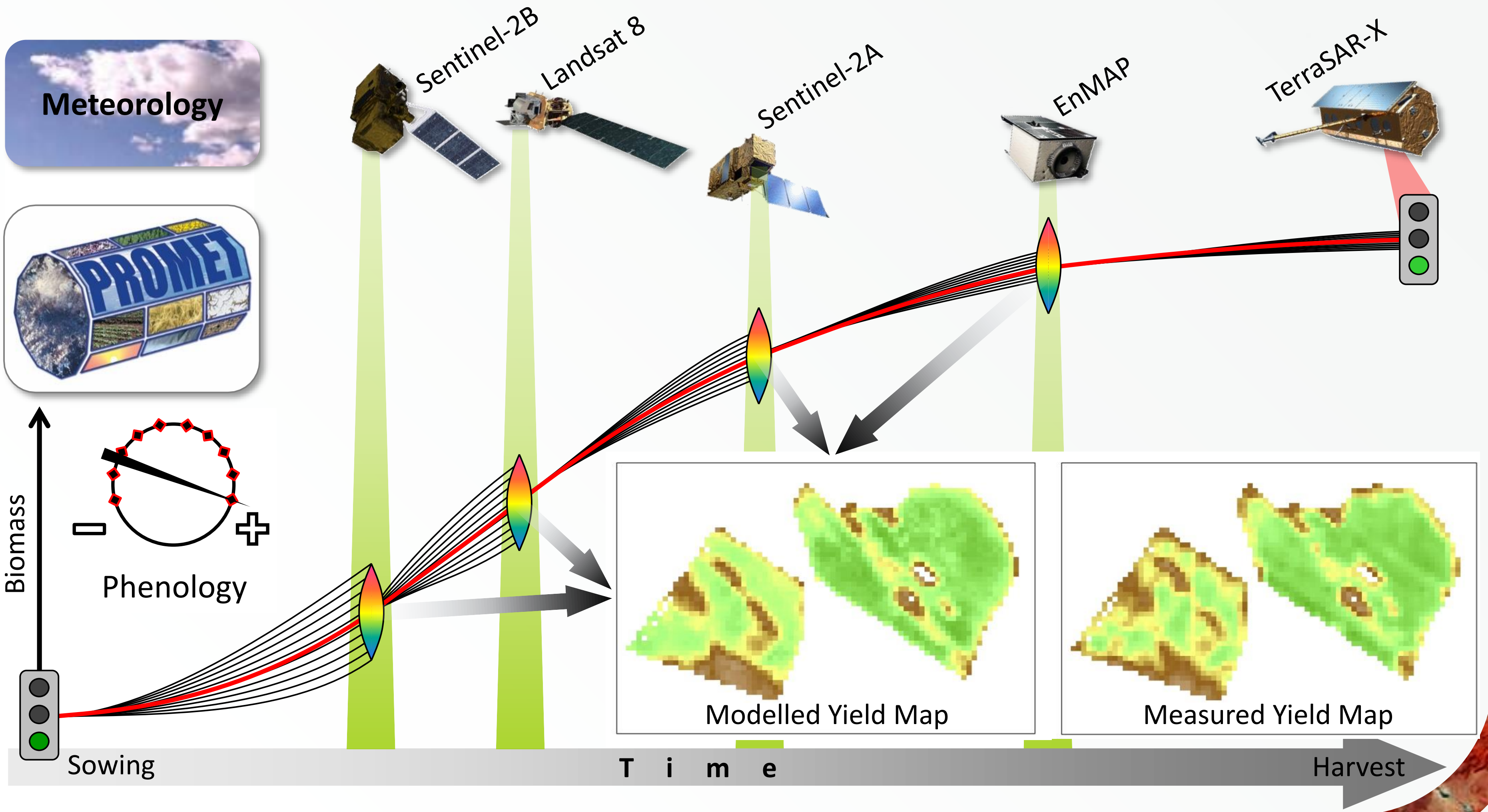
Biophysical and Biochemical Variable Retrieval

Retrieval Methods

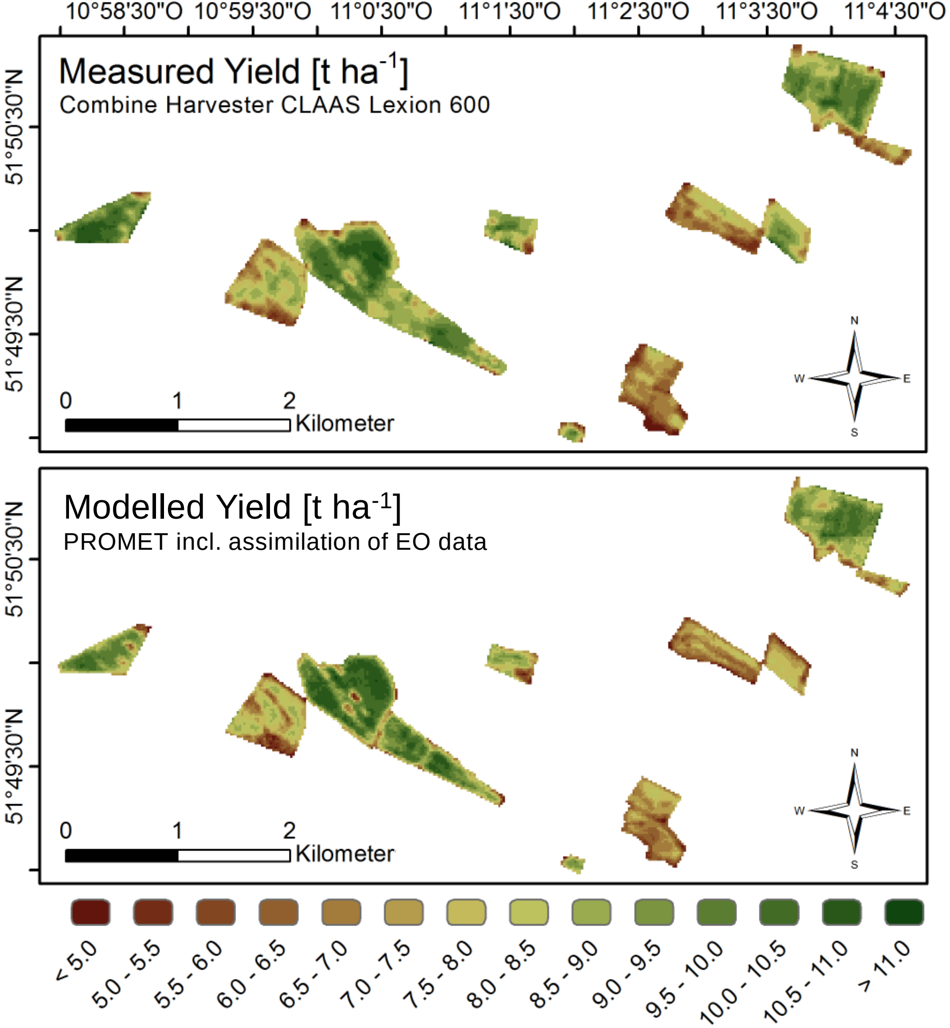
How to translate the complex spectral information into semantic information (which we can understand)?

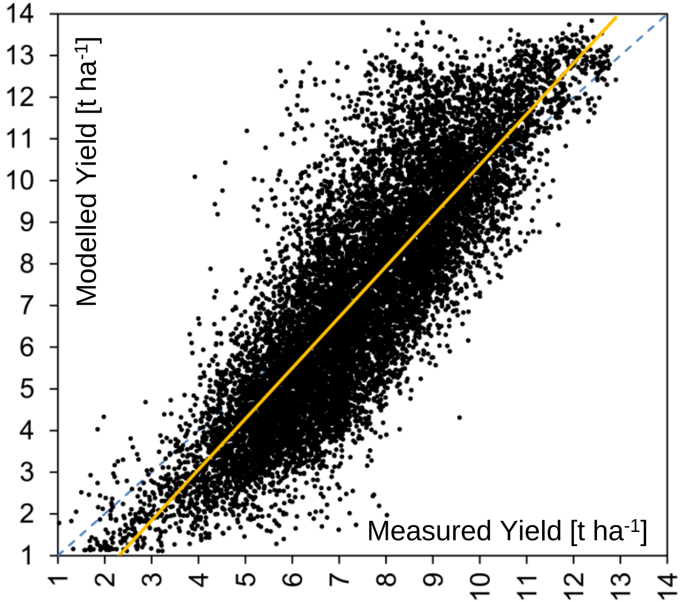



Meteorology
PROMET
Sentinel-2B
LAI Measured
LAI Assimilation
≈
Biomass
Phenology
LAI Simulation
+ Biomass
+ Canopy Height
+ Phenology
+ Root system
+ Soil Moisture
+ Nutrition status
+ Energy balance
+ Erosion
+ ca. 800 variables
Sowing
T i m e
Harvest



Meteorology
PROMET
Sentinel-2B
Landsat 8
Sentinel-2A
EnMAP
TerraSAR-X
Biomass
Phenology
Modelled Yield Map
Measured Yield Map
Sowing
T i m e
Harvest





A	=	525.4 [ha]
R^2	=	0.70 [-]
RMSE	=	1.59 [t ha^{-1}]
Ø Meas. / Mod.	=	7.42 / 7.23 [t ha^{-1}]

Results from the project "TalkingFields – The Use of Satellite Data to optimize Agricultural Production", funded by ESA IAP

Biophysical and Biochemical Variable Retrieval

Physically-based vs. statistical Methods

With their *deductive capabilities*, physical models differ from regression approaches: the combination of logical (mechanistic) principles enables extrapolation to predictions about behaviours not present in earlier sampled field data. In contrast, regression or learning algorithms identify patterns happening in the data and build relationships without asking for a physical basis (=*inductive capability*), see Baker et al. (2018).

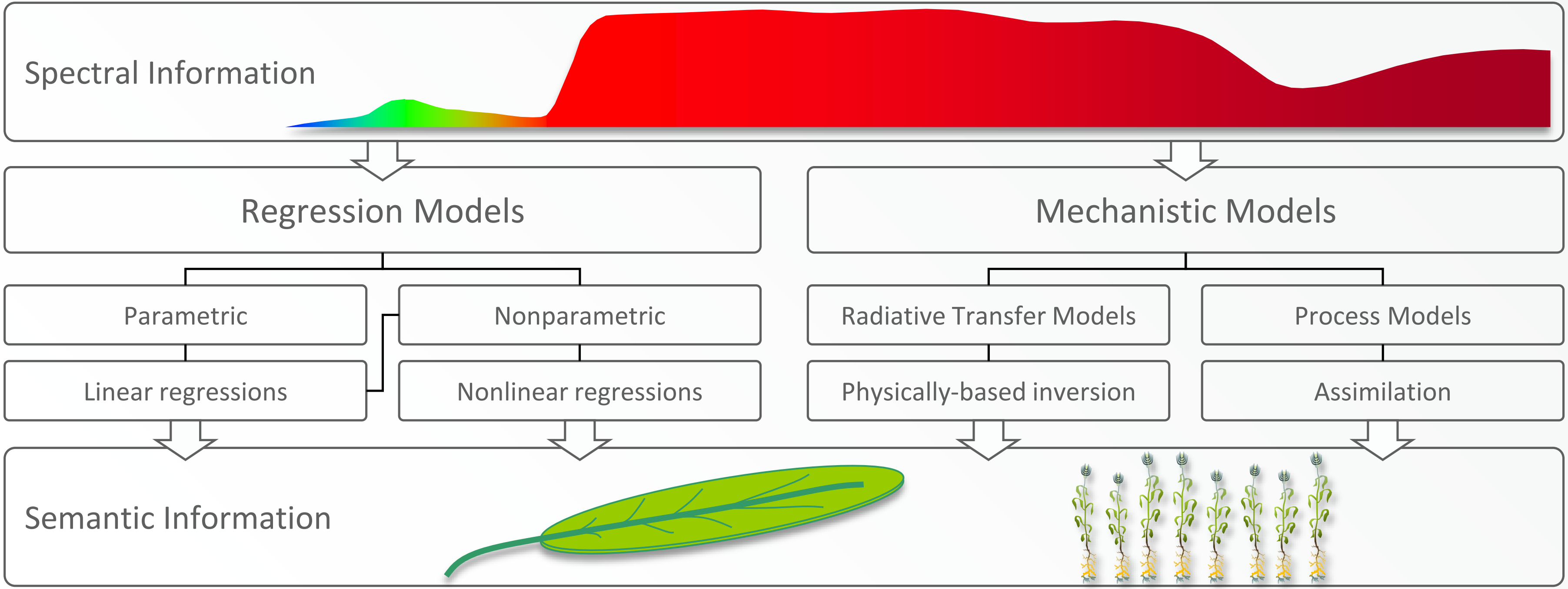
Though physical models are capable of describing causality between inputs and outputs by means of physical laws, they tend to oversimplify the reality. Instead, ML algorithms establish nonlinear relationships between any kind of data without incorporation of physical rules (Camps-Valls et al. 2018). However, one needs to know the physical basis that is underneath the data. This helps for designing the training dataset and selecting the inputs.

So, should we replace physically-based models with machine learning algorithms? No, we need **both**:

- The spatial and temporal prediction ability of machine learning should be, at least, consistent with the patterns observed in physical models." (Reichstein et al. 2019)
- The two approaches are complementary: mechanistic models could be used to provide physical constraints and domain knowledge to machine learning algorithms (Reichstein et al. 2019; Weiss et al. 2020).

Biophysical and Biochemical Variable Retrieval
Retrieval Methods

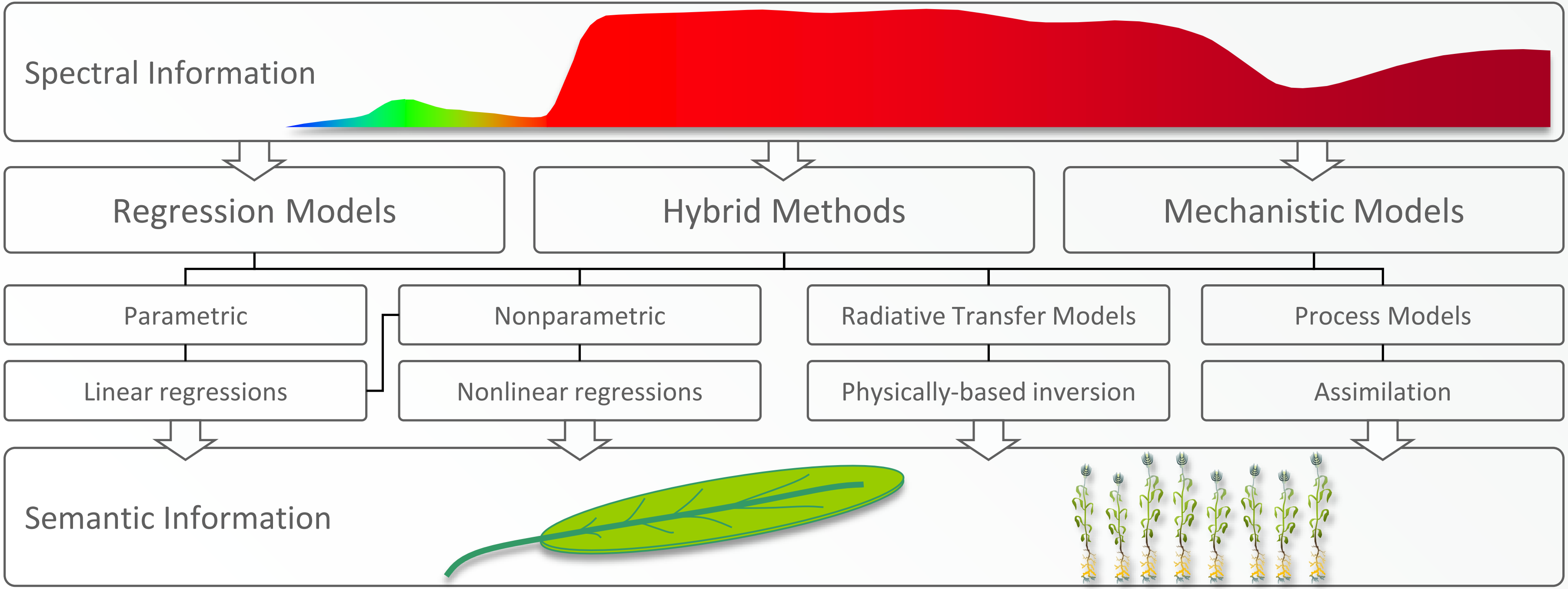
How to translate the complex spectral information into semantic information (which we can understand)?



Biophysical and Biochemical Variable Retrieval

Retrieval Methods

How to translate the complex spectral information into semantic information (which we can understand)?



Biophysical and Biochemical Variable Retrieval
Hybrid Methods

Hybrid techniques denominate a combination of at least two methods in synergic use to obtain the objective (retrieval) more efficiently. For vegetation properties mapping from Earth observation data, hybrid methods are often referred to the combination of machine learning methods and RTMs: thus combining flexibility and scalability of ML while respecting the physics encoded in the RTM (Camps-Valls et al. 2018).

How about a concrete example for hybrid mapping?

WOCHER, M., BERGER, K., VERRELST, J. & HANK, T. (2022): Retrieval of carbon content and biomass from hyperspectral imagery over cultivated areas. *ISPRS Journal of Photogrammetry and Remote Sensing,* Volume 193, pp. 104-114. https://doi.org/10.1016/j.isprsjprs.2022.09.003

Biophysical and Biochemical Variable Retrieval
Hybrid Methods – Carbon Mapping Example

- Agriculture (together with forestry and land use change) is contributing approx. 20% of the global GHG emissions.
- Long-Term storage of carbon in agriculturally used soils may play a role as carbon sink.
- “Carbon Farming” is part of the EU Green Deal.

Wouldn’t it be nice, if we could measure the amount of carbon stored in agricultural crops (and within their residues) with a spectrometer?

Europe → Southern Germany → Bavaria → Near Straubing → Center Coordinates: 48.81° N | 12.73° E

Test Site Irlbach: Location

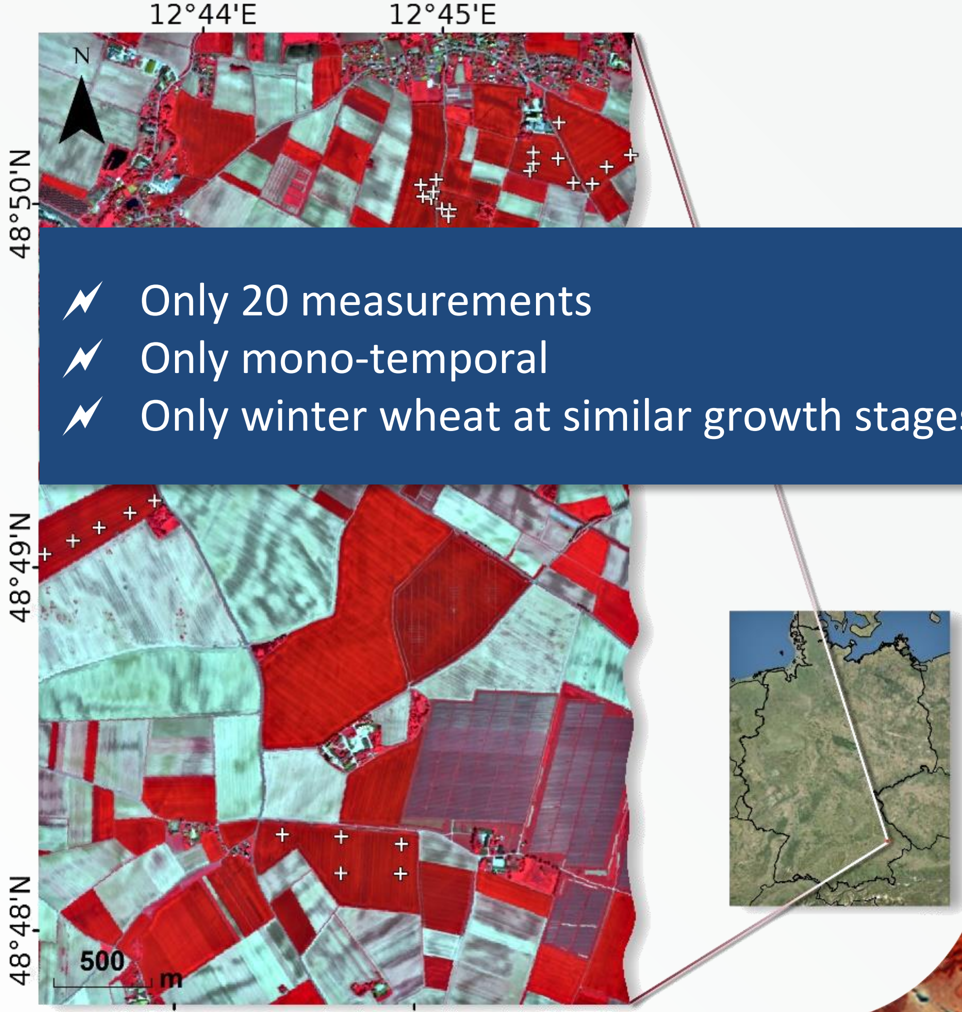
Irlbach site | May 30th 2021

Airborne data:

HyperSense CHIME preparation campaign
AVIRIS-NG airborne acquisitions
(377-2501 nm @425 bands)
→ resampled to EnMAP (233 bands)

In situ data:

- Dry & fresh biomass (AGB_{dry}, AGB_{fresh})
- Carbon content (C_{area})
- Nitrogen content (N_{area})
- Leaf area index (LAI)



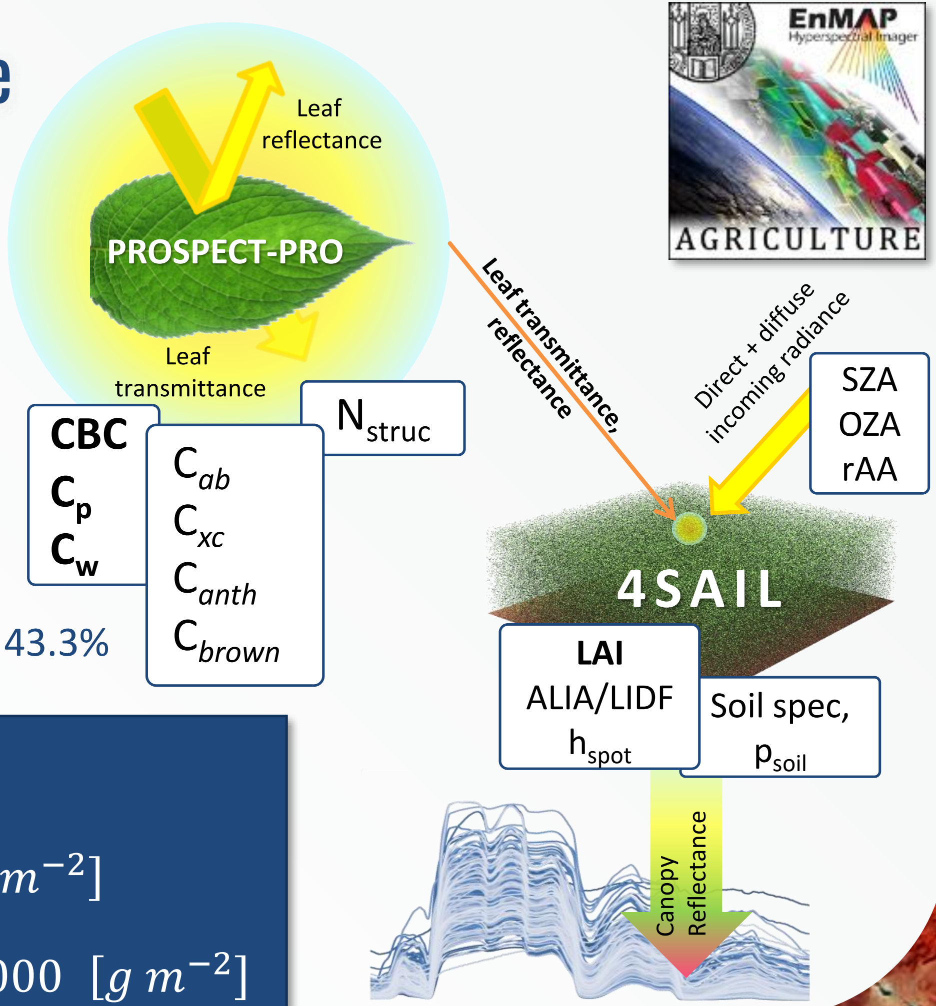
- Only 20 measurements
- Only mono-temporal
- Only winter wheat at similar growth stages

PROSAIL-PRO training database

- 3000 members
- Parameters varied over wide value range

Carbon based constituents (CBC) include sugars, starch, lignin, cellulose and hemicellulose → **CBC ≠ C**

Conversion factor needed!
→ Mean C-concentration (leaves|stalks|fruits) = 43.3%

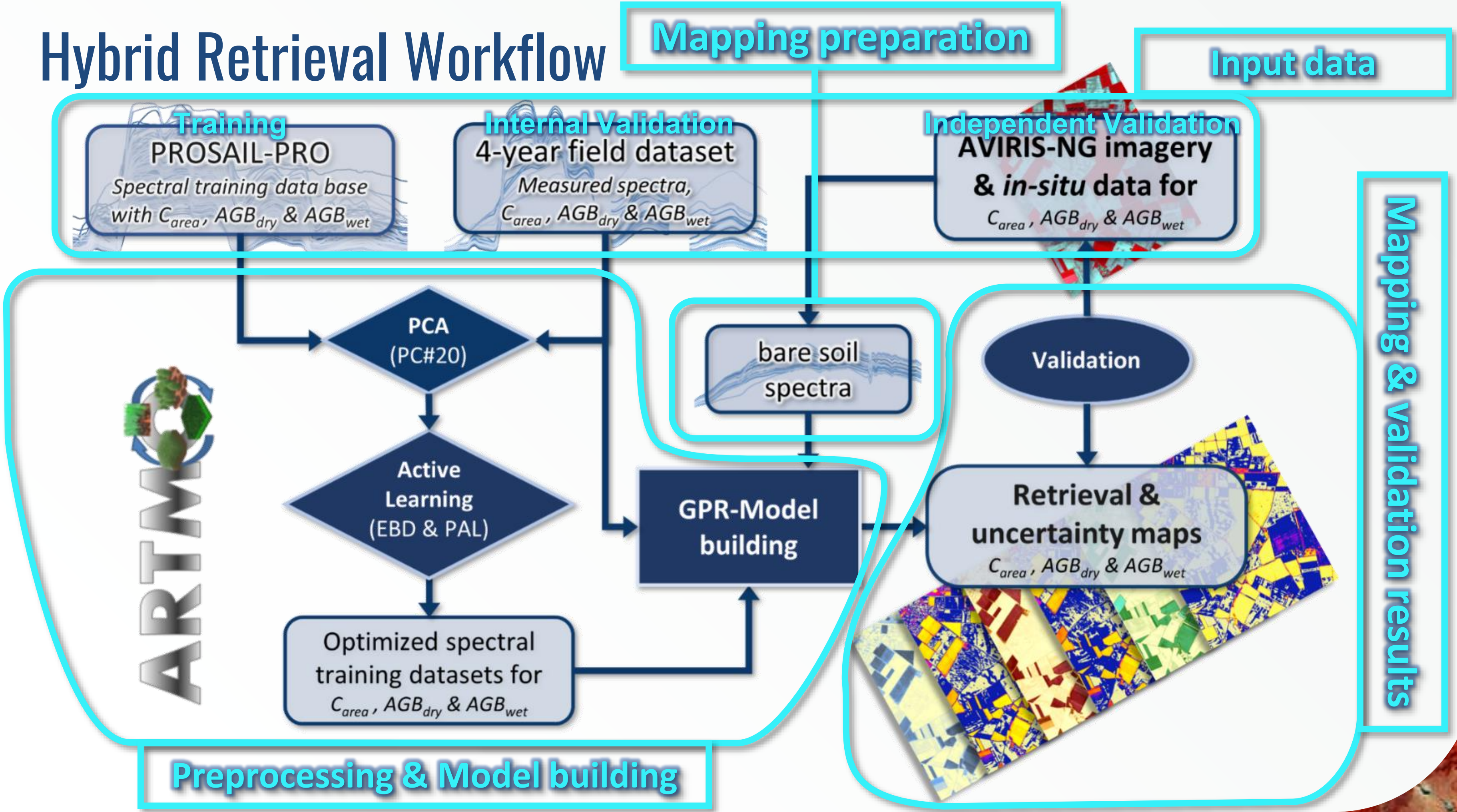


$$C_{area} = \frac{CBC \times LAI \times 10{,}000}{2.31}\ [g\ m^{-2}]$$

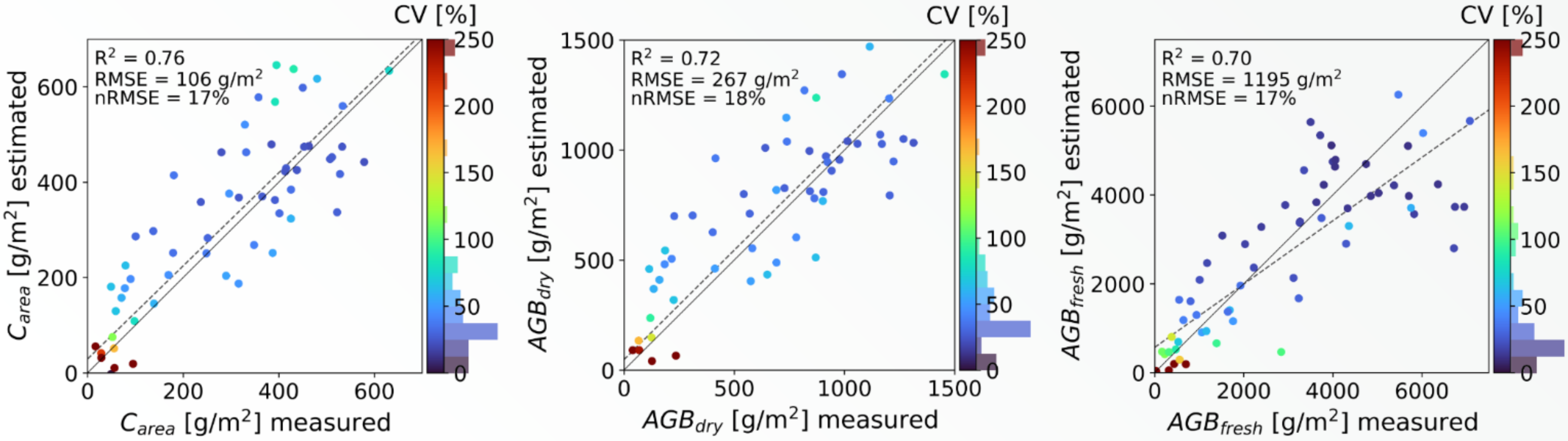
$$AGB_{dry} = \left(C_p + CBC\right) \times LAI \times 10{,}000\ \ [g\ m^{-2}]$$

$$AGB_{fresh} = \left(C_p + CBC + C_w\right) \times LAI \times 10{,}000\ \ [g\ m^{-2}]$$

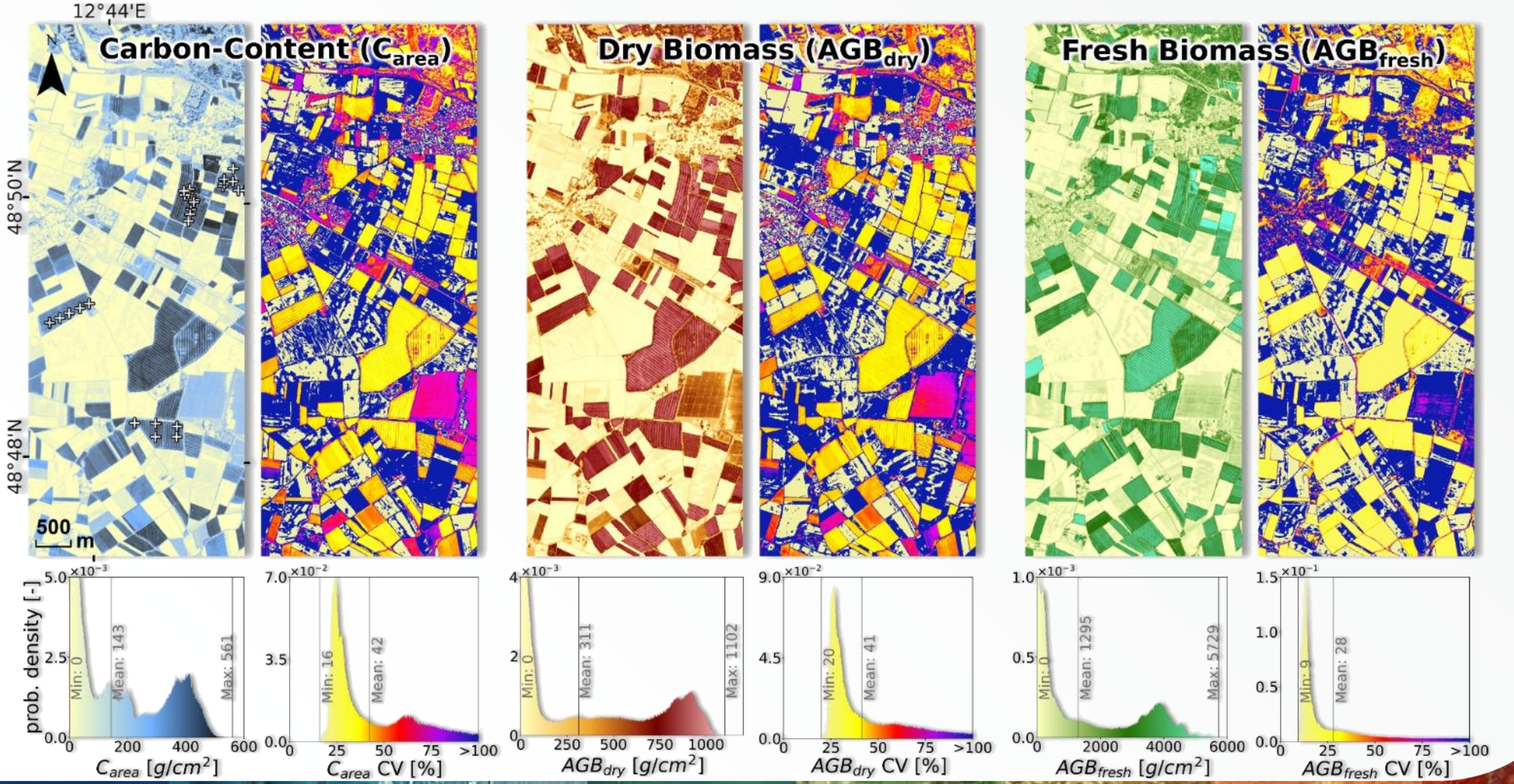
Hybrid Retrieval Workflow



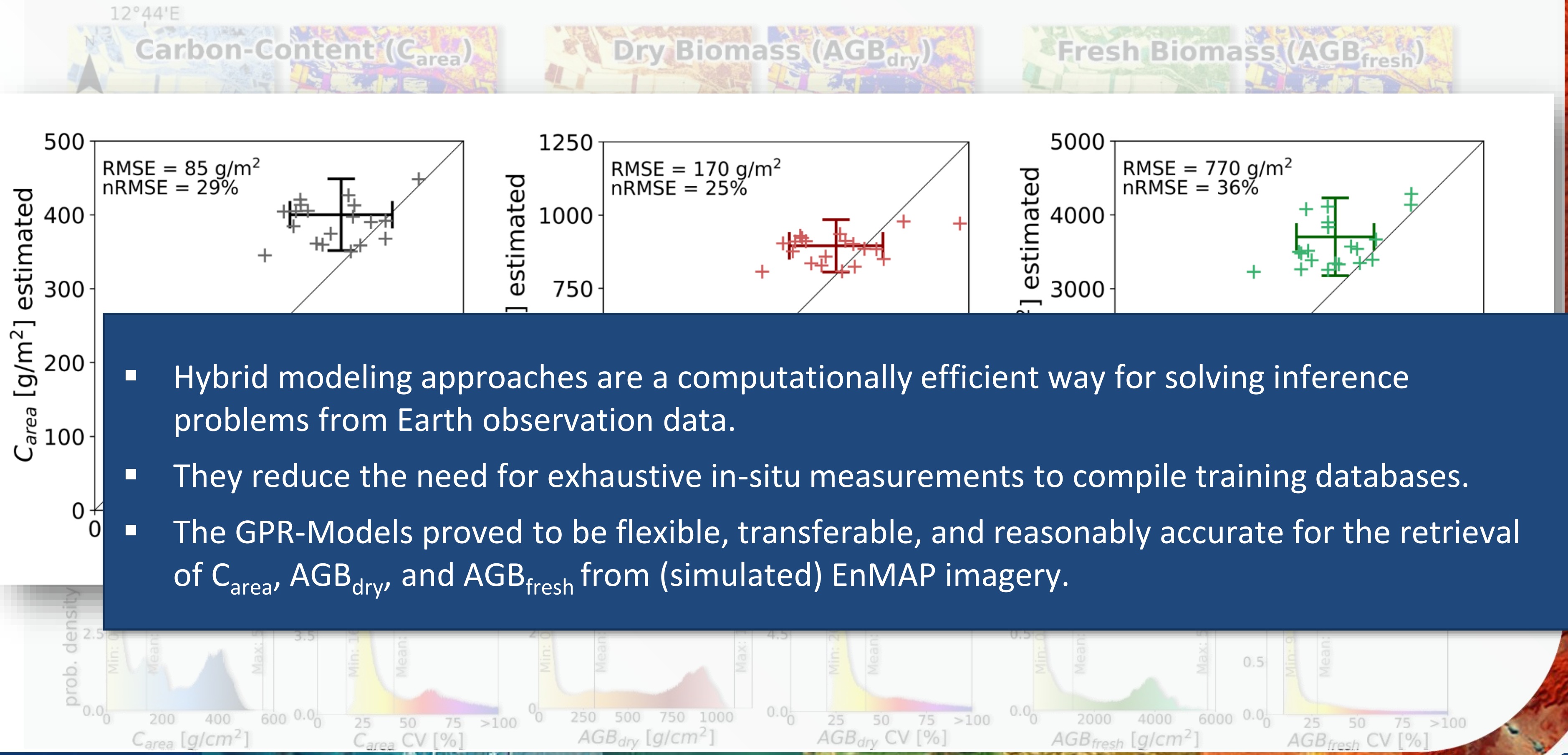
Internal Validation of the GPR Model



Spatial Mapping with the GPR Model

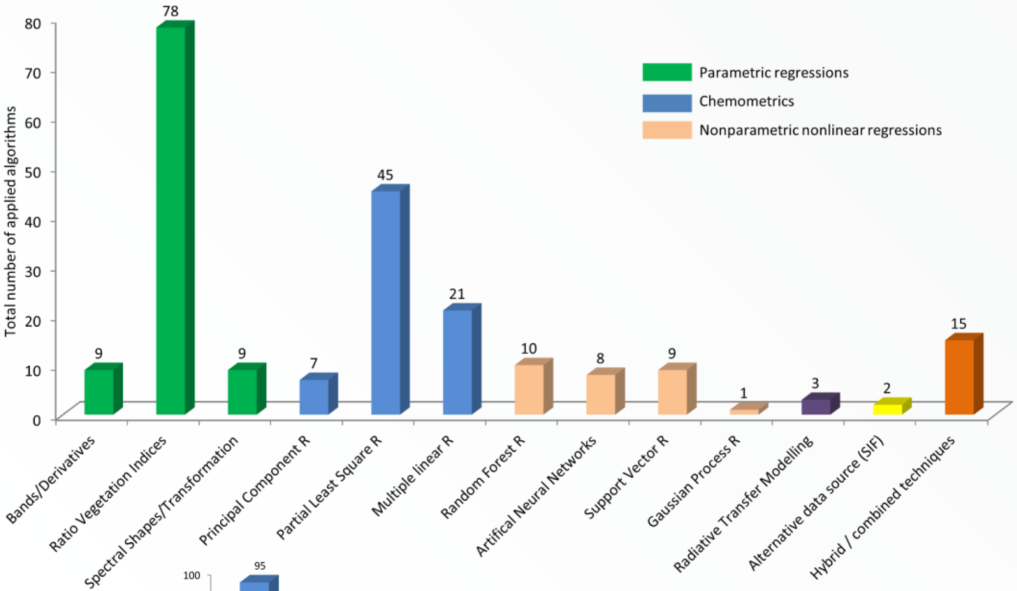


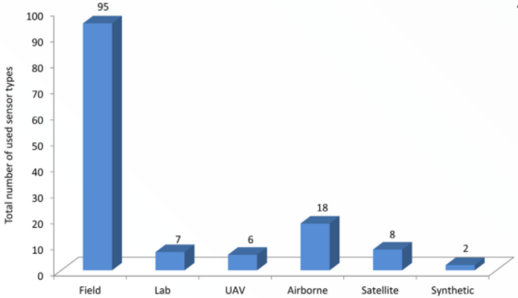
Independent Validation of the GPR Model Results

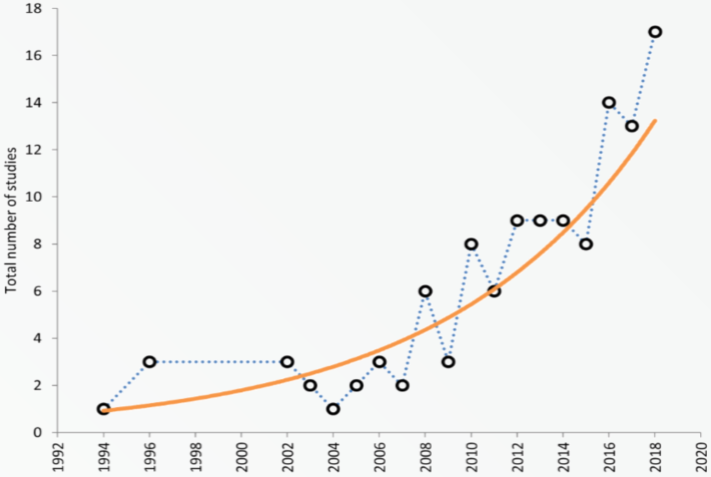


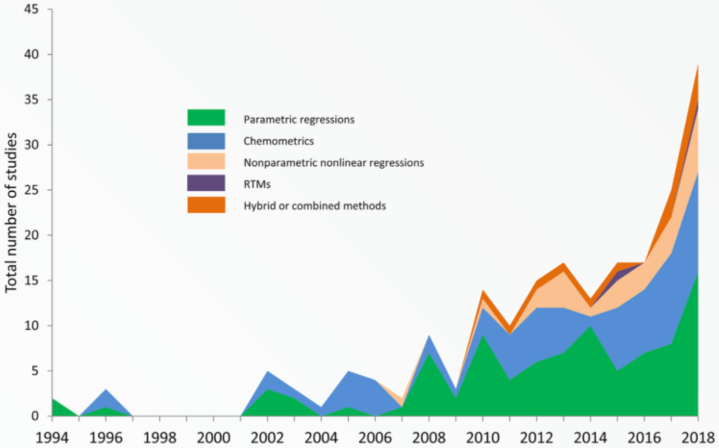
- Hybrid modeling approaches are a computationally efficient way for solving inference problems from Earth observation data.
- They reduce the need for exhaustive in-situ measurements to compile training databases.
- The GPR-Models proved to be flexible, transferable, and reasonably accurate for the retrieval of C_{area}, AGB_{dry}, and AGB_{fresh} from (simulated) EnMAP imagery.

Literature Survey of Applied Methods









Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., & Hank, T. (2020). Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sensing of Environment, 242, 111758

Next-generation spaceborne hyperspectral sensors

Mission (organiz., country)	Spectral range (SSD, no. of bands)	Spatial resolution (swath)	Repeat interval (days)	Launch	Purpose	Reference
DESIS (DLR, Germany)	400-1000 nm (2.55 nm , 235 bands)	30 m (30 km)	3 (63 TOD)	29.06.2018	Scientific precursor	Krutz et al. (2019)
PRISMA (ASI, Italy)	400-2500 nm (6-12 nm, 240 bands)	30 m (30 km)	29 (7, repeat roll m.)	22.03.2019	Technology demonstrator	Loizzo et al. (2019)
HISUI (METI, Japan)	400-2500 nm (10-12 nm, 185 bands)	20 m (cross-track) x 30 m (along-track) (20 km)	3 (63 TOD)	05.12.2019	Operational	Matsunaga et al. (2017)
EnMAP (DLR, Germany)	400-2500 nm (6.5-10 nm, 242 bands)	30 m (30 km)	27 (4 off-nadir)	01.04.2022	Scientific precursor	Guanter et al. (2015)
SHALOM (Italian-Israeli)	400-2500 nm (10 nm, 275 bands)	10 m (30 km)	?	2022	Operational/ commercial	Feingersh & Ben Dor (2015)
CHIME (ESA)	400-2500 nm (225 bands)	20-30 m (290 km?)	10-12.5	2025-2030	Copernicus high-priority mission candidate	Nieke & Rast (2018), Ustin & Middleton (2021)
SBG (NASA, U.S)	VSWIR: 380-2500 nm (10 nm, 210 bands)	30-45 m (150 km)	≤ 16	2026/2027	Operational	NASA (2018); Cawse-Nicholson et al. (2021)
FLEX/FLORIS (ESA)	500-780 nm (434 bands)	300 m (150 km)	27	planned for 2024	Scientific precursor	Coppo et al. (2017)

Conclusions

- The questions of practical agriculture are of a **quantitative** nature.
  → Hyperspectral remote sensing targets variables that come **with physical units**.

- Crop development is a **temporally highly dynamic** process.
  → We are looking forward to the **new missions** to learn how to make the best possible use of **time-series** from imaging spectrometers.

- Models (physical as well as statistical ones) are **digital twins** of the real world.
  → We need to bring **everything that needs mapping** into the digital world.

- There still are strong **deviations** between modeled and measured reflectances.
  → We need to get back into the lab and into the field and **improve the models**.

- At the same time don‘t let us forget **all the data** that we have collected during the last 40 years.
  → We need to share and harmonize this data to improve **transferability** of statistical models.

Thank You for Your Attention!

tobias.hank@lmu.de

www.geographie.uni-muenchen.de

www.enmap.org

EnMAP

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN