LINEAR SPECTRAL UNMIXING FOR LARGE SPACEBORNE HYPERSPECTRAL DATASETS - CHALLENGES AND SOLUTIONS FOR THE AUTOMATED DLR FCOVER CHAIN

Martin Bachmann[1], David Marshall[1], Kevin Kühl[1], Frederic Schwarzenbacher[2] and Sarah Asam[1]

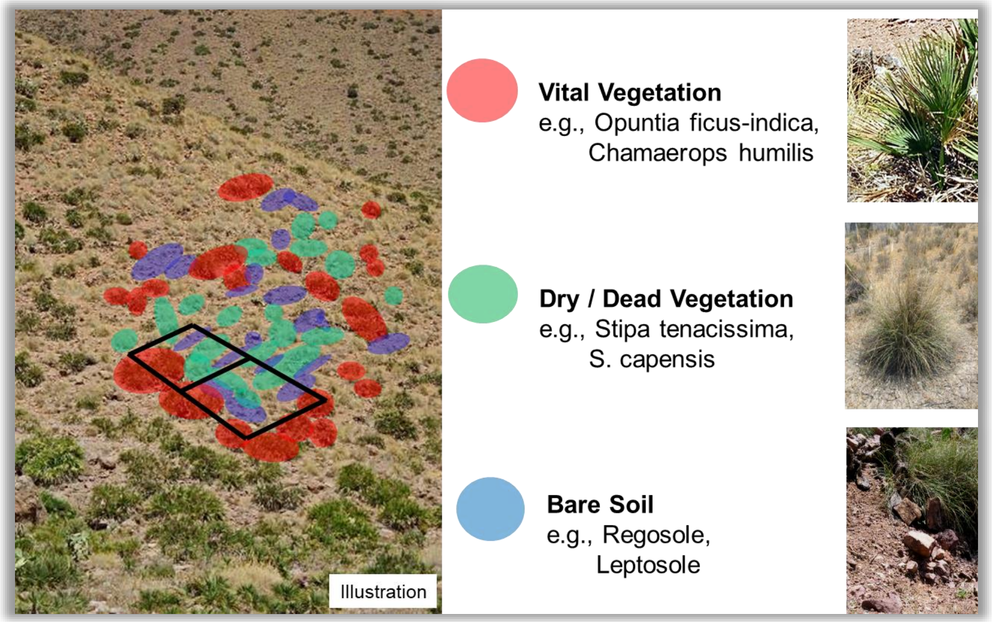
[1] **DLR-EOC Earth Observation Center**

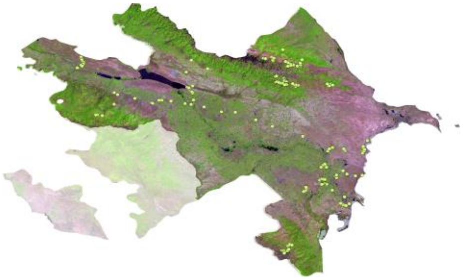
[2] **Topodat®, Koenig, Schwarzenbacher & Will GbR**

DLR

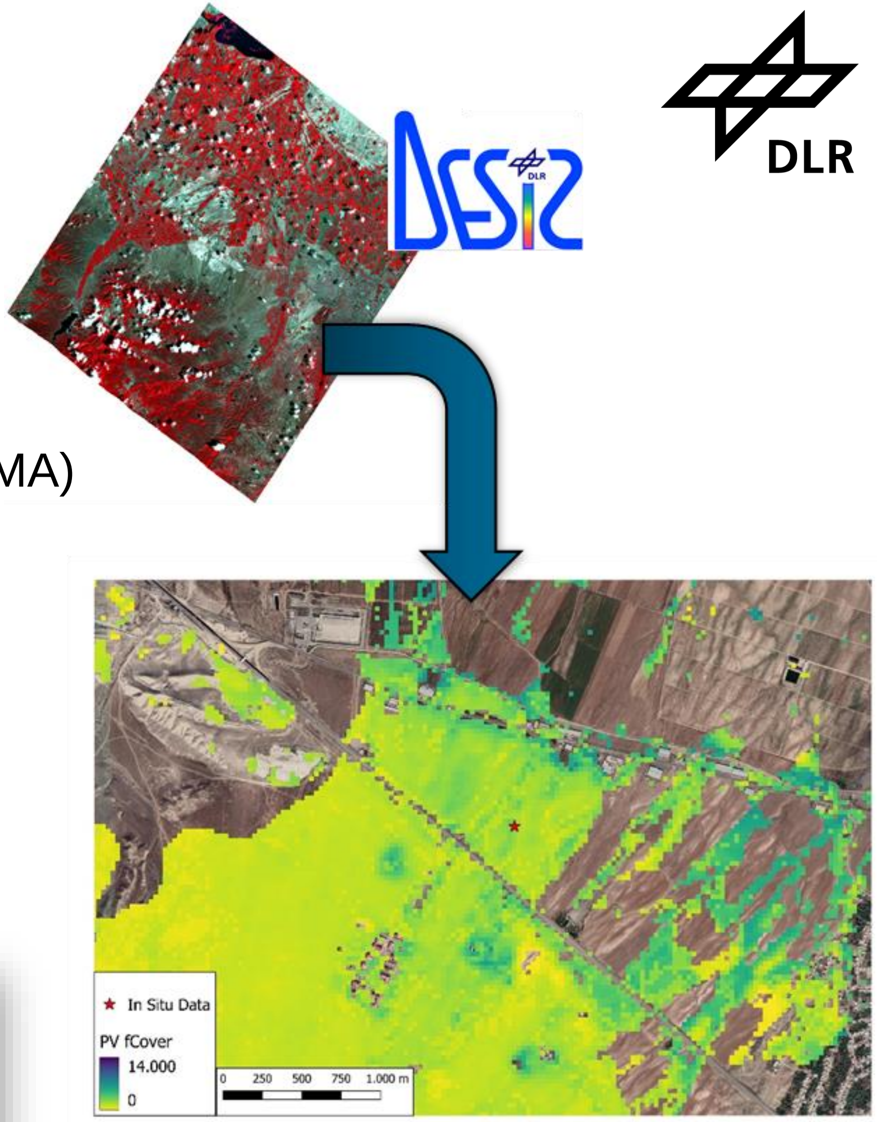
Rationale

- Linear spectral unmixing can be highly accurate
  - If all End-Members (EM) are known
  - If spectral variability of EMs is included in mixture model (MESMA)



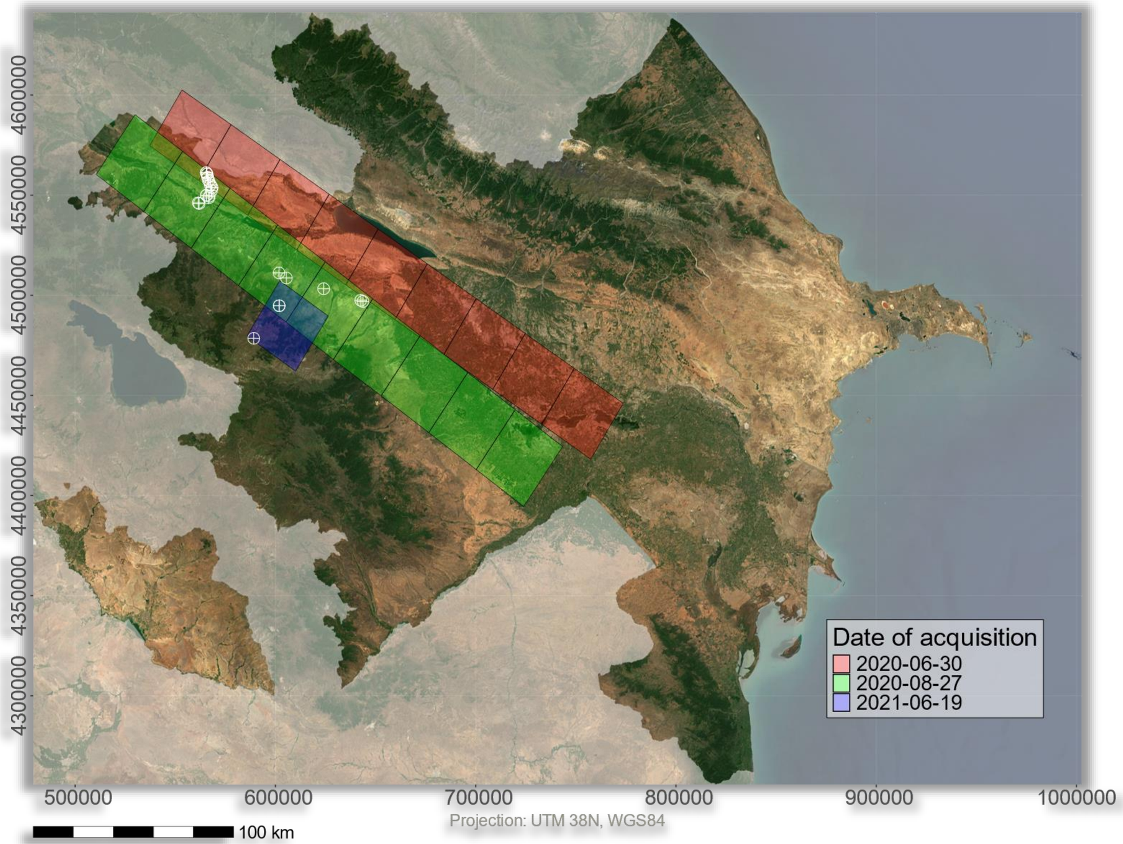


Azerbaijan,
Degradation of grasslands



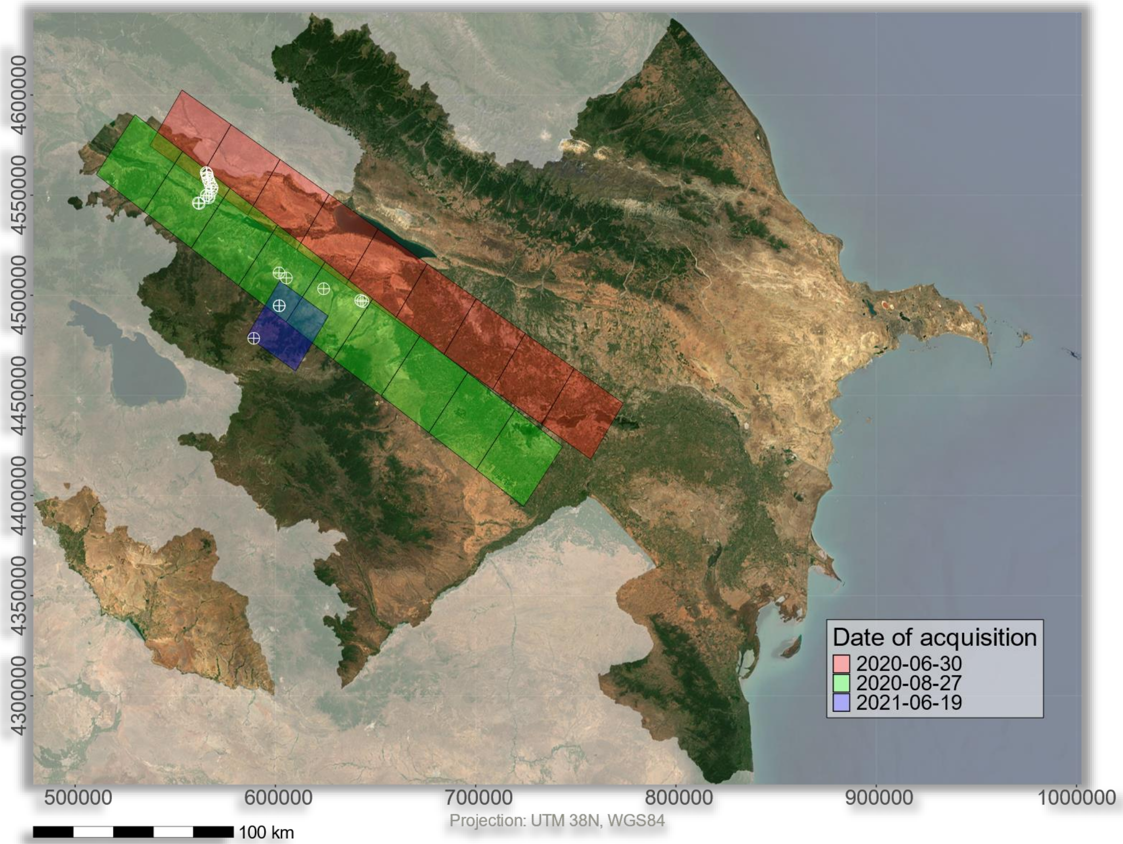
Rationale

- Linear spectral unmixing is highly accurate
  - If all End-Members (EM) are known
  - If spectral variability of EMs is included in mixture model (MESMA)
- But in real-life:
  - Not all EM are known
  - If strictly tile-based: EM change between tiles
  - Some EM are already mixtures
  - Some combinations of EM result in ill-conditioned mixing model
  - View angle effects generally limit accuracy



Rationale

- Linear spectral unmixing can be highly accurate
  - If all End-Members (EM) are known
  - If spectral variability of EMs is included in mixture model (MESMA)
- But in real-life:
  - Not all EM are known
  - If strictly tile-based: EM change between tiles
  - Some EM are already mixtures
  - Some combinations of EM result in ill-conditioned mixing model
  - View angle effects generally limit accuracy
- Resulting in:
  - Reduced overall accuracy
  - Variable accuracy over scene
  - Bordering effects between tiles



Background – Linear Spectral Unmixing

- Linear Spectral Mixture Model

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$Ax = b$$

Where

a_{mn}: reflectance of EM n in band m
b_m: measured reflectance in band m
x_m: abundance for EM n

A: m*n EM-matrix
x: abundance vector for n EM
b: measured spectrum in m bands

Overdetermined problem, solving by Least-Squares approximation, e.g. $x = A^+ b$ where $A^+ = (A^T A)^{-1} A^T$ plus constraints (sum-to-one, non-neg.)

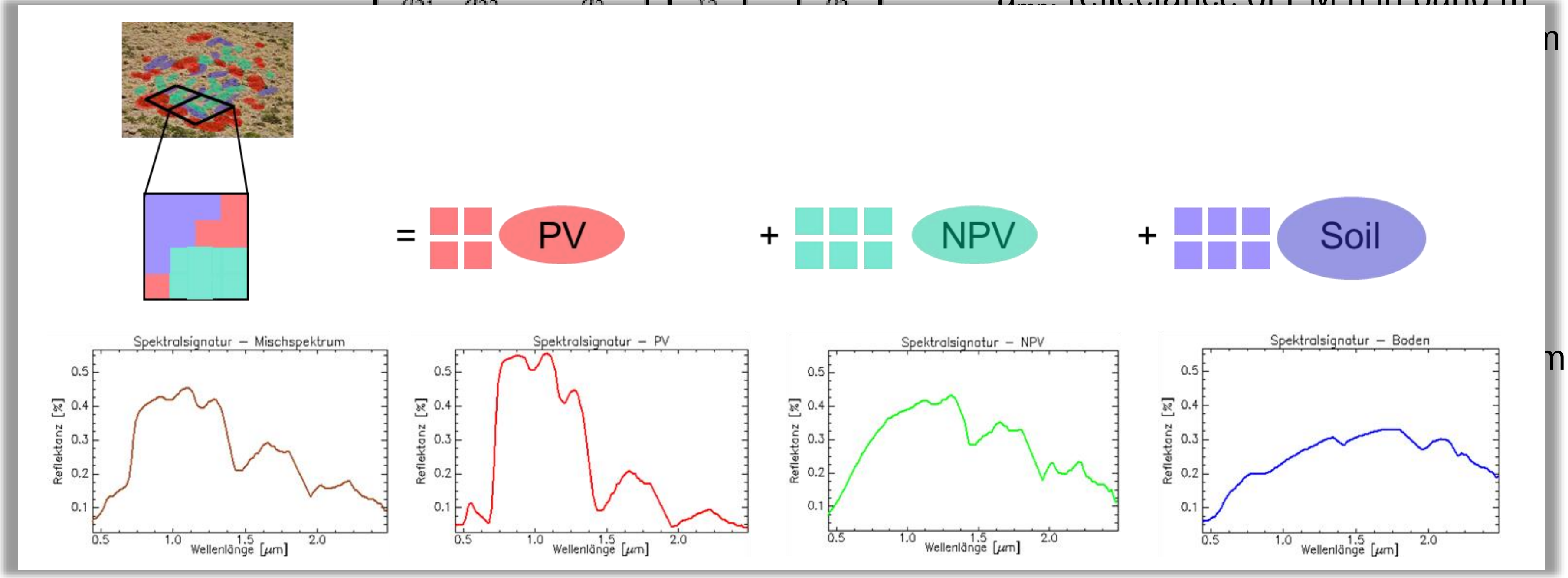
Background – Linear Spectral Unmixing

- Linear Spectral Mixture Model

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix} \begin{pmatrix} x_1 \end{pmatrix} \quad \begin{pmatrix} b_1 \end{pmatrix}$$

Where
a_{mn}: reflectance of EM n in band m
n

m-to-one, non-neg.)



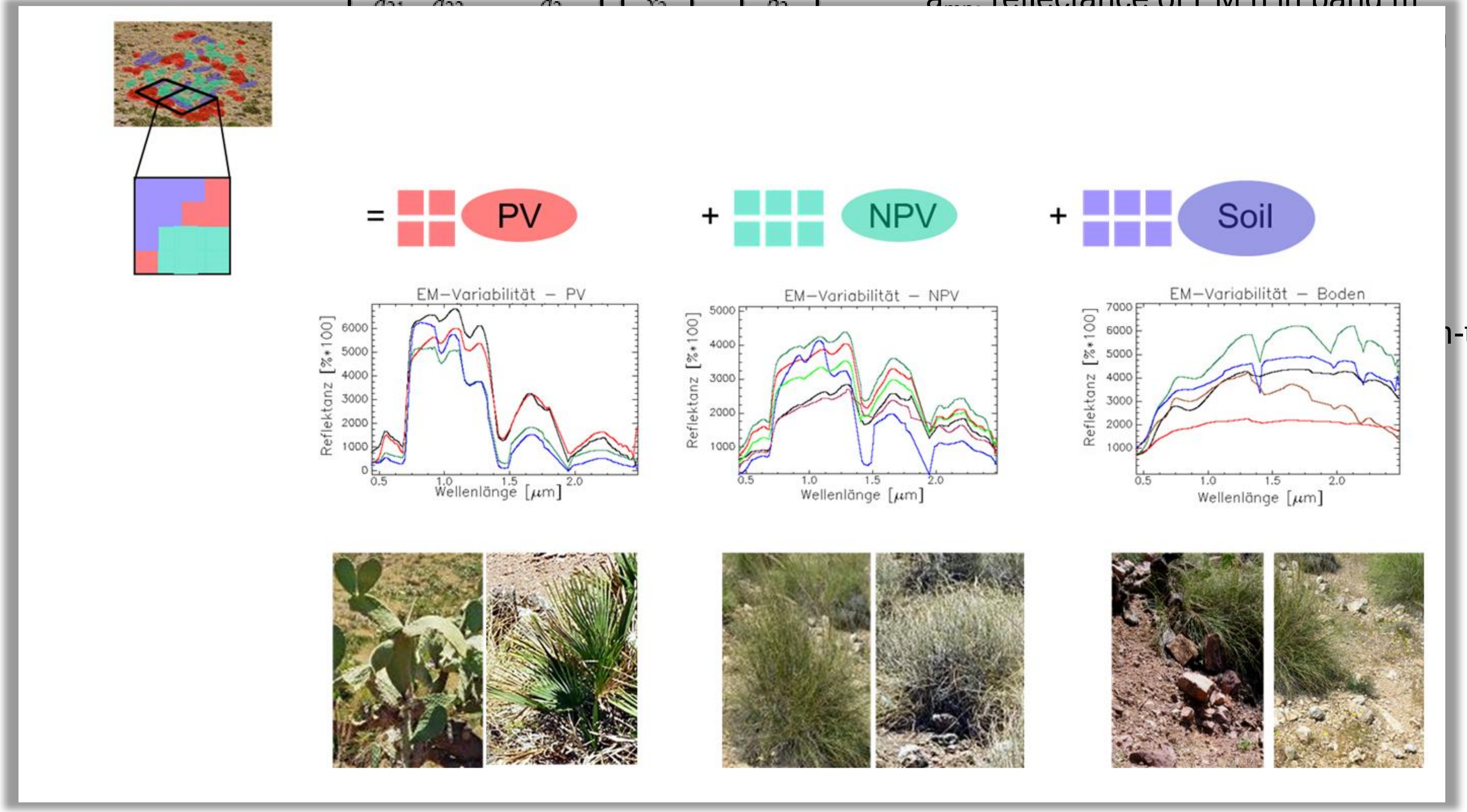
Background – Linear Spectral Unmixing

- Linear Spectral Mixture Model

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix} \begin{pmatrix} x_1 \end{pmatrix} \quad \begin{pmatrix} b_1 \end{pmatrix}$$

Where

a_{mn}: reflectance of EM n in band m



n-to-one, non-neg.)

Background – Linear Spectral Unmixing

- Linear Spectral Mixture Model

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}$$

$$Ax = b$$

Where
a_{mn}: reflectance of EM n in band m
b_m: measured reflectance in band m
x_m: abundance for EM n

A: m*n EM-matrix
x: abundance vector for n EM
b: measured spectrum in m bands

Overdetermined problem, solving by Least-Squares approximation, e.g. $x = A^+ b$ where $A^+ = (A^T A)^{-1} A^T$ plus constraints (sum-to-one, non-neg.)

↗ **M**ultiple **E**ndmember **S**pectral **M**ixture **A**nalysis

Particular **EMs used** to model a pixel and **number of EMs** (i.e., **matrix A**) varies on a **per-pixel** basis
Usually model with smallest RMS error selected

➡ More materials and spectral variability of EM included

Background – Accuracy of Image EMs

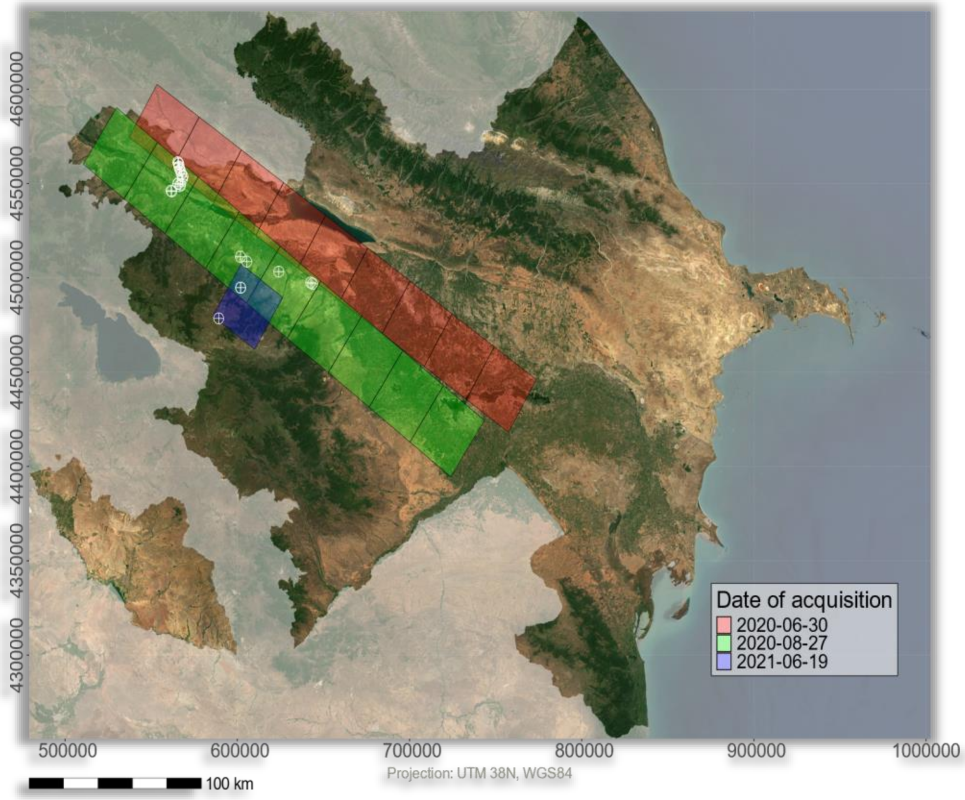
- **Test case based on synthetic scenes**
  (pure field spectra from 3 campaigns in Spain, Namibia **plus:** MEDSPEC, JHU SpecLib, ECOSTRESS)
  - SMACC (Sequential max. angle convex cone) on pre-segmented image
  - Additional EMs included after 1st unmixing iteration
  - ~70% of all EM could be automatically retrieved
    But: incorrect spectra also detected!
- Subsequent automated EM classification in 3 classes (PV – NPV – Soil)
  - In simulations:
    ~93% correctly classified EM
- Consequences:
  30% of EM missing
  ~1 in 10 EM is incorrect
- **Usually an incorrect and incomplete mixture model**

Challenges & Solutions (I)

- Now – what can we do to improve the situation?
  - Take advantage of EM-derivation over multiple tiles: SSEE (Rogge et al.)
  - Exclude linear dependencies in mixture model
  - Select mixture models which are "reasonable"
  - Identify mixture models which are likely to be wrong
  - Address general limitations - view angle effects
  - Use an appropriate shade component
  - Improve solving algorithms
  - …

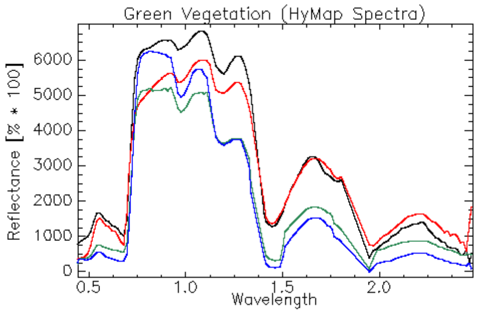
  … and explicitly include a per-pixel reliability score

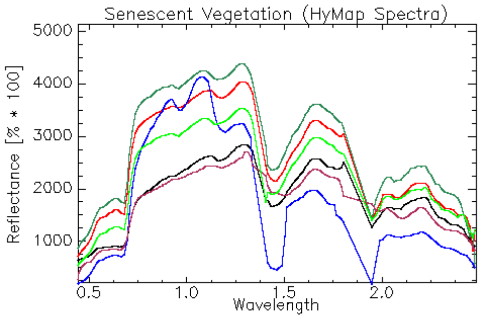
- Presenting one approach: **automated fCover unmixing**

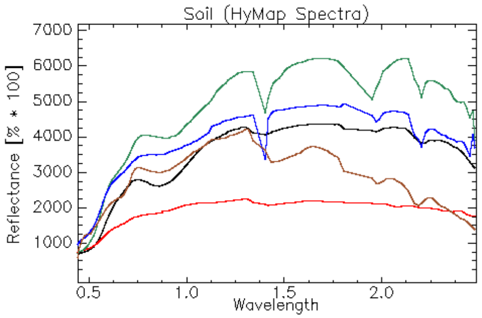


Addressing Numerical Problems

- Matrix inversion may result in numerical problems!
  - **Linear dependencies** between spectrally similar classes







  - Higher number of EM – higher probability of linear dependencies
  - Result: **ill-conditioned problem**

Thus: check condition number κ of EM-Matrix A:

$\kappa = ||A||\ ||A^{-1}||$ where || denotes the euklidian L2-norn

➚ **Exclude EM combinations** which result in ill-conditioned problem

For the mentioned SpecLibs: **>10% of all mixing models**

(esp. combination of Dry Veg. – Soil)

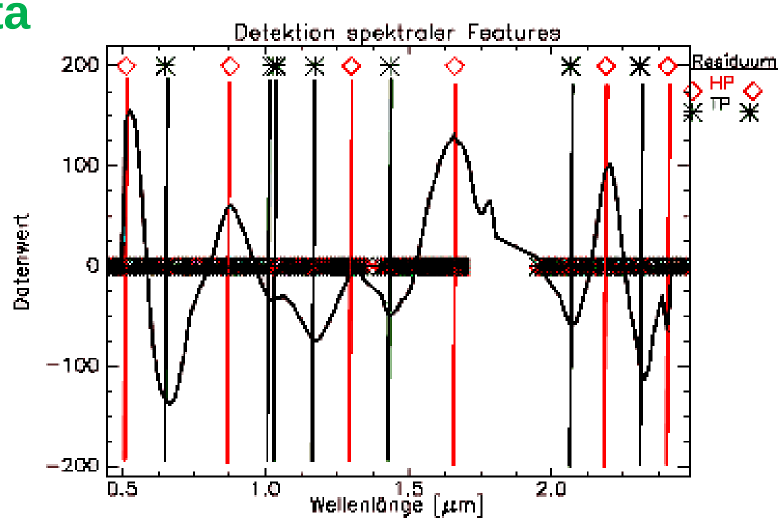
Selecting Reasonable Mixture Models

Model selection in standard MESMA: **RMS Error** (i.e., goodness of fit)

Some improvements in µMESMA:

- Integrate information from **spatial dimension**
  Soil type -and thus soil EM- unlikely to change between pixels.
  Thus: **check**, if unmixing error significantly increases when
  using **dominant soil EM in spatial neighborhood**

- Take advantage of **spectroscopic data**
  - **Automated residual analysis**
  - Check residual spectra for **diagnostic absorption features**
  - Identify & parameterize these features



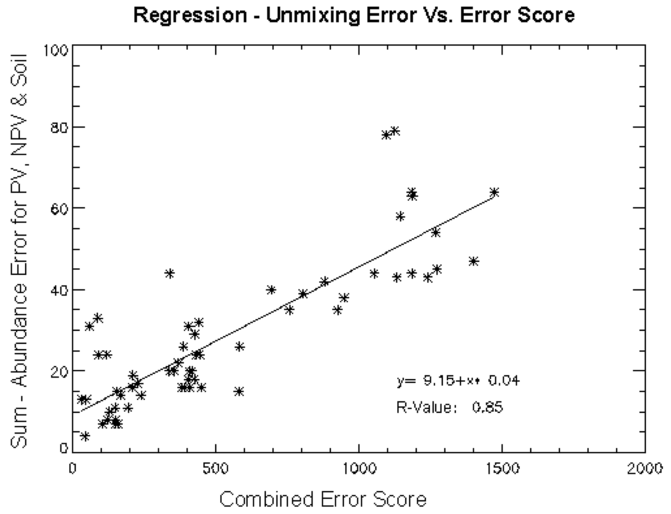
➡ Increased stability against incorrect EM

Identify Error-Prone Pixels

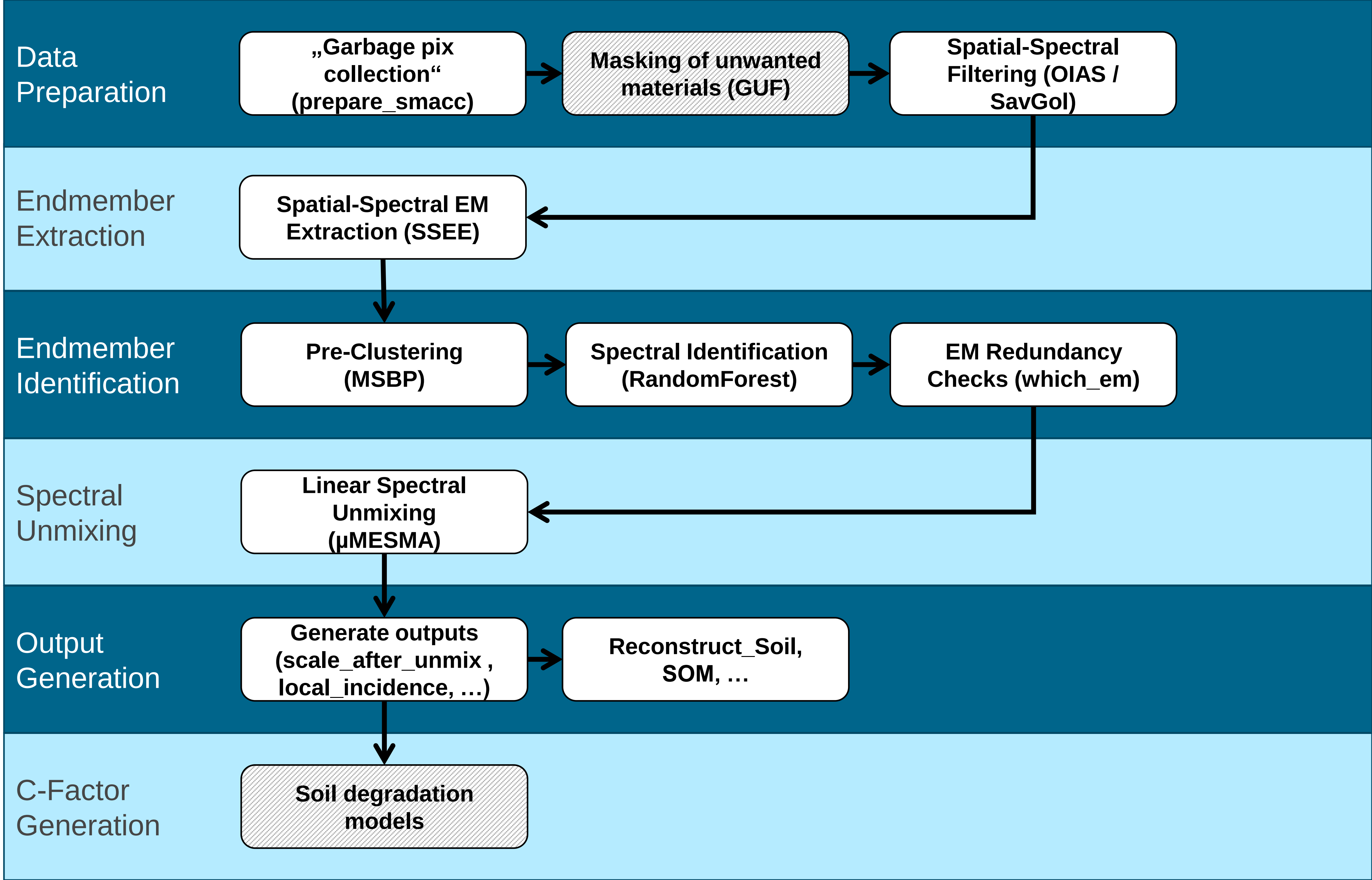
- Linear Spectral Unmixing already offers measure for **“goodness of fit”,** i.e. the model RMSE
- **Improved** detection of pixels which are likely error-prone based on:
  - **Residual** analysis **&** weighted model **RMSE**
  - Agreement with **empirical regression models** between cover % and band indices
  - Critical local **incidence angle**
- L2 data quality flags (from pre-processing)

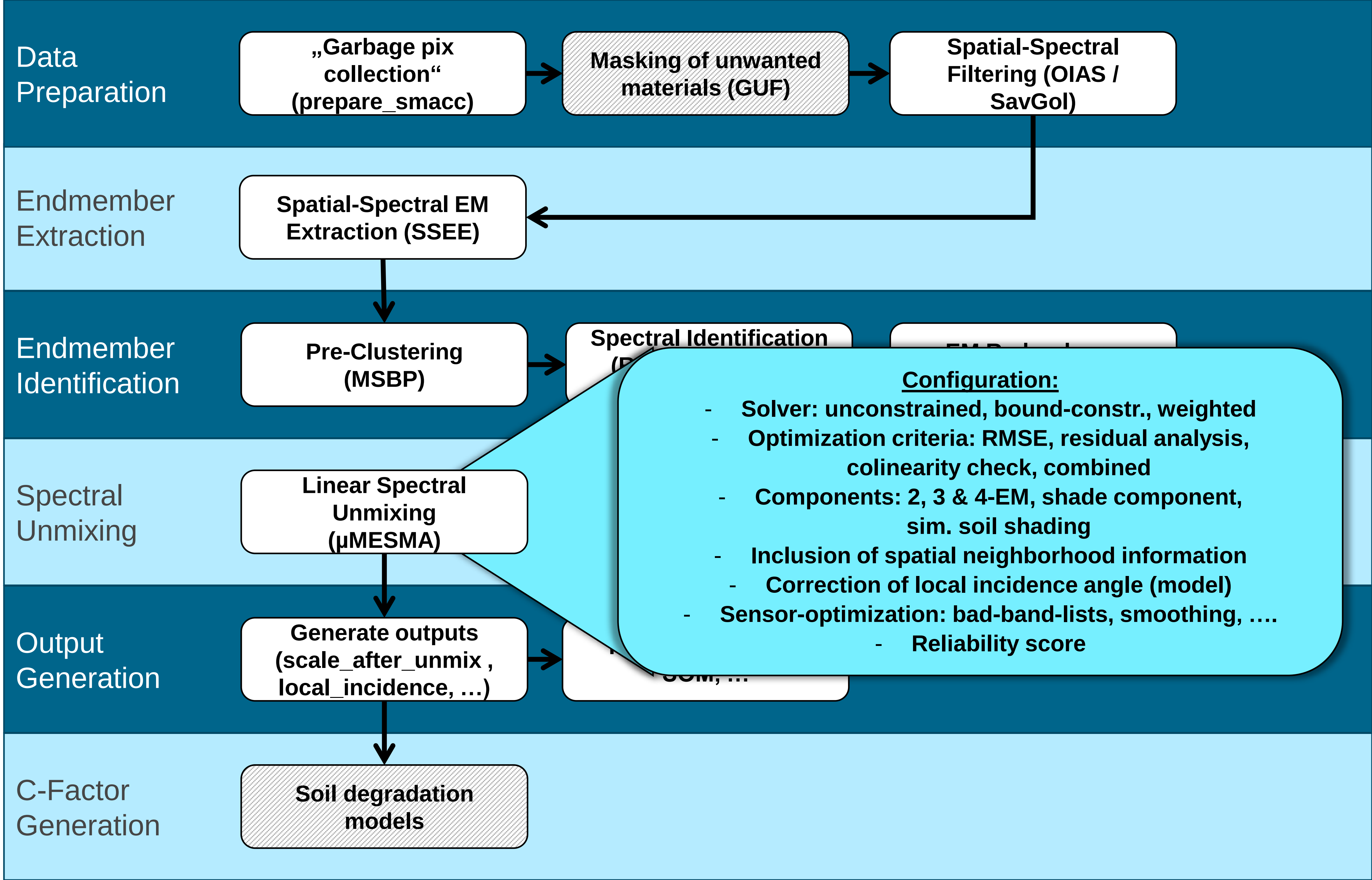
➡ **Baseline for a Reliability Score**

R^2 ~0.73,
n= 61 tests, 61.000 models



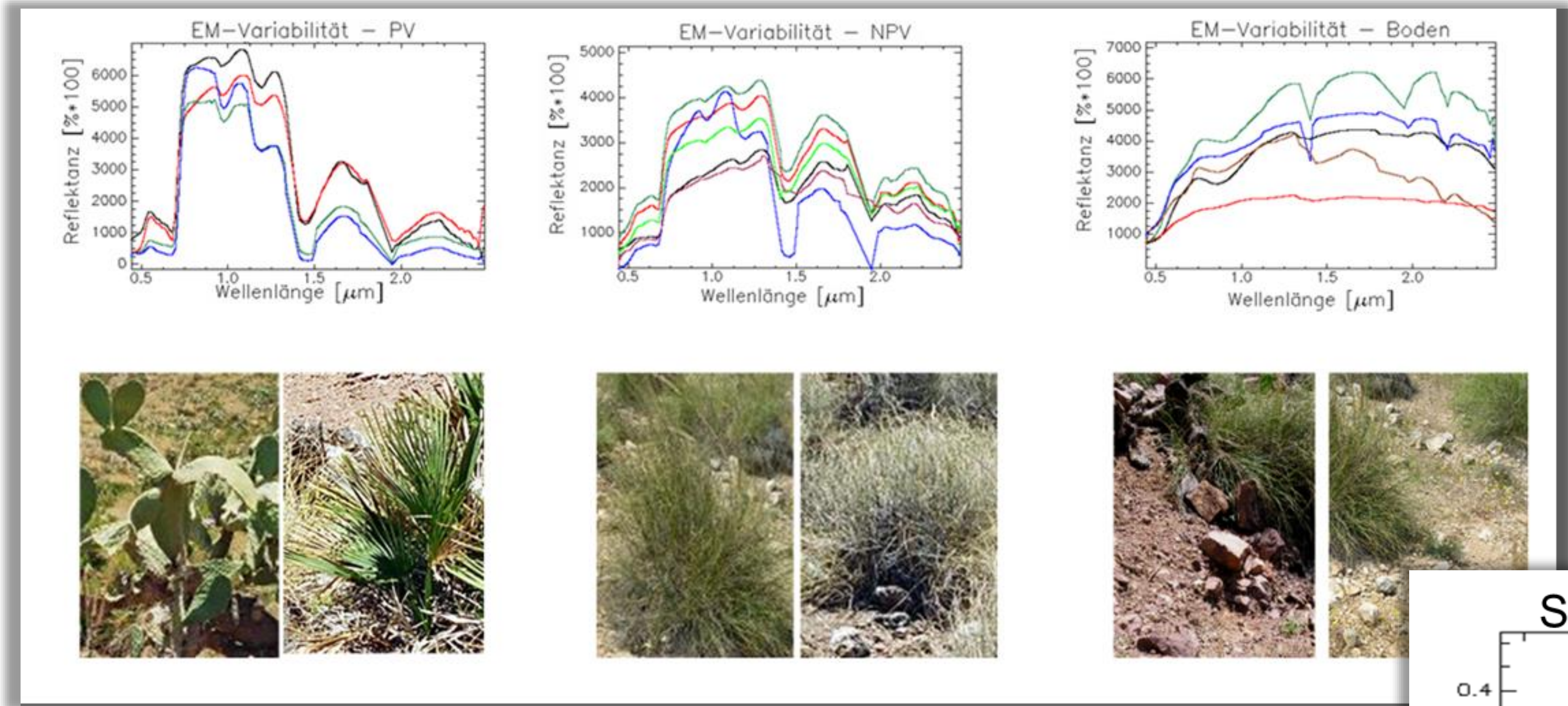
Data Preparation	**Mask erroneous pixels, mask unwanted materials, filter image**
Endmember Extraction	**Detect and extract potentially pure pixel spectra**
Endmember Identification	**Cluster and identify these spectra, group into PV, NPV, SOIL, and remove redundant spectra**
Spectral Unmixing	**Calculate ground cover fractions („abundances“) for PV, NPV, SOIL**
Output Generation	**Generate additional unmixing-related products (classification by EM maps, reliability, statistics, …)**
C-Factor Generation	**„Translate“ abundance images into C-Factors**

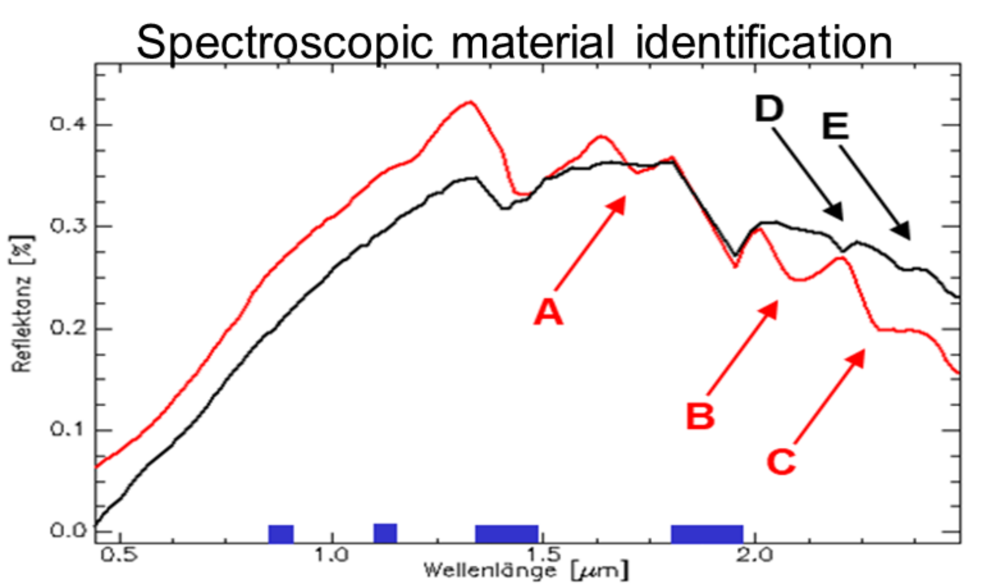

Data Preparation
„Garbage pix collection“ (prepare_smacc)
Masking of unwanted materials (GUF)
Spatial-Spectral Filtering (OIAS / SavGol)
Endmember Extraction
Spatial-Spectral EM Extraction (SSEE)
Endmember Identification
Pre-Clustering (MSBP)
Spectral Identification (RandomForest)
EM Redundancy Checks (which_em)
Spectral Unmixing
Linear Spectral Unmixing (µMESMA)
Output Generation
Generate outputs (scale_after_unmix , local_incidence, …)
Reconstruct_Soil, SOM, …
C-Factor Generation
Soil degradation models



Data Preparation
„Garbage pix collection“ (prepare_smacc)
Masking of unwanted materials (GUF)
Spatial-Spectral Filtering (OIAS / SavGol)
Endmember Extraction
Spatial-Spectral EM Extraction (SSEE)
Endmember Identification
Pre-Clustering (MSBP)
Spectral Identification
Spectral Unmixing
Linear Spectral Unmixing (µMESMA)
Output Generation
Generate outputs (scale_after_unmix , local_incidence, …)
C-Factor Generation
Soil degradation models
Configuration:
- Solver: unconstrained, bound-constr., weighted
- Optimization criteria: RMSE, residual analysis, colinearity check, combined
- Components: 2, 3 & 4-EM, shade component, sim. soil shading
- Inclusion of spatial neighborhood information
- Correction of local incidence angle (model)
- Sensor-optimization: bad-band-lists, smoothing, ….
- Reliability score

… but wait:
can DESIS separate between NPV & soil ?





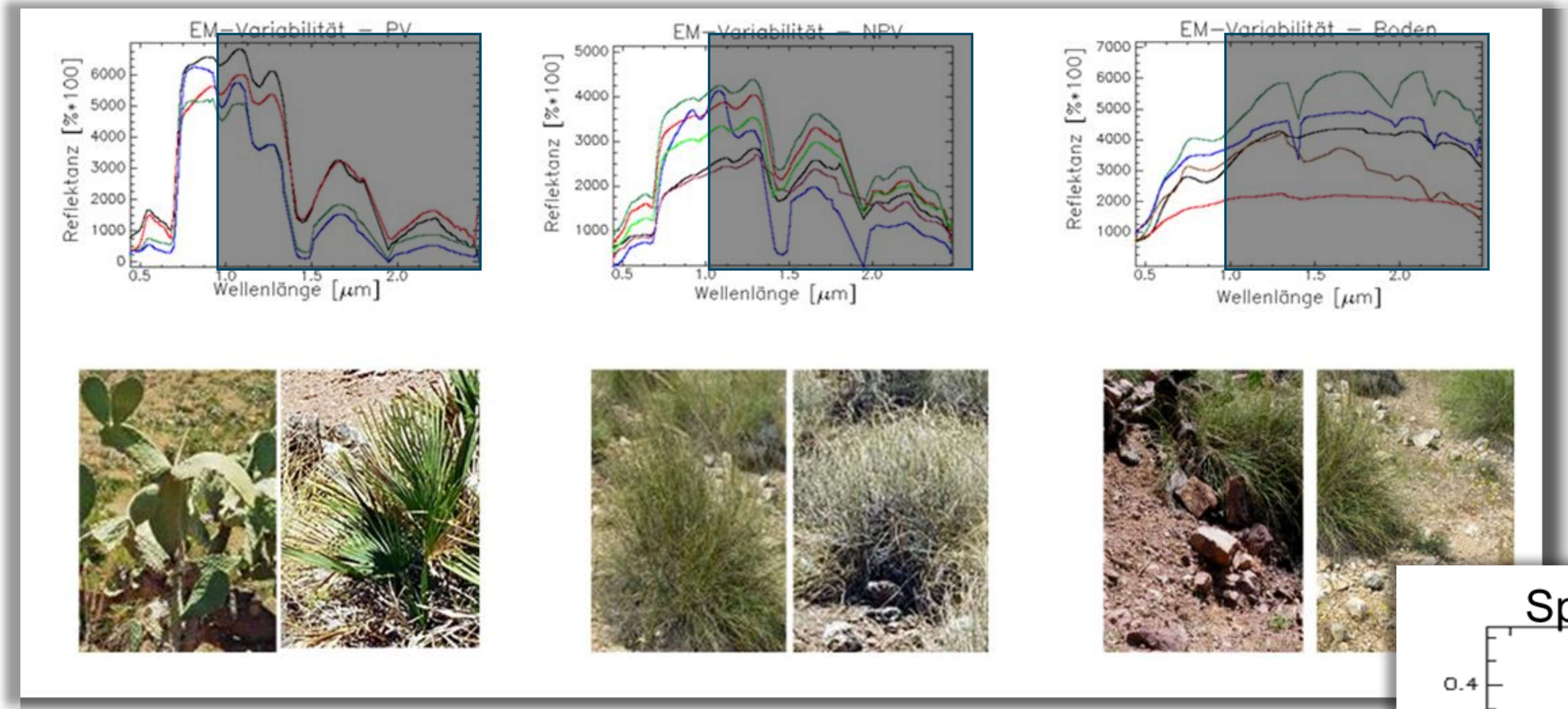
Absorption features of
A: Xylan & Cellulose
B: Lignin & Cellulose
C: Cellulose

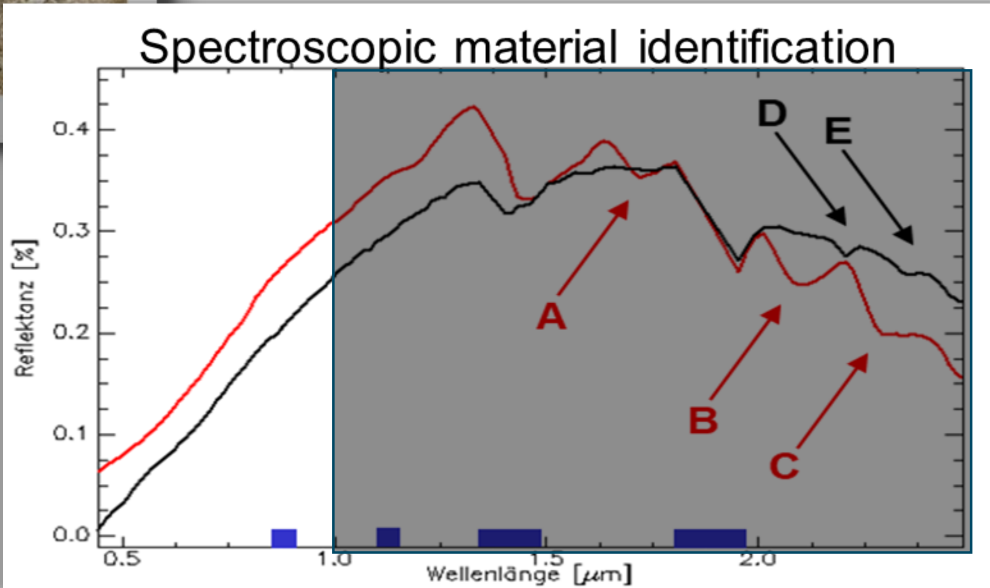
Vegetation without Chlorophyll => NPV

D: Clay
E: Carbonates

=> Soil

… but wait:
can DESIS separate between NPV & soil ?





Absorption features of

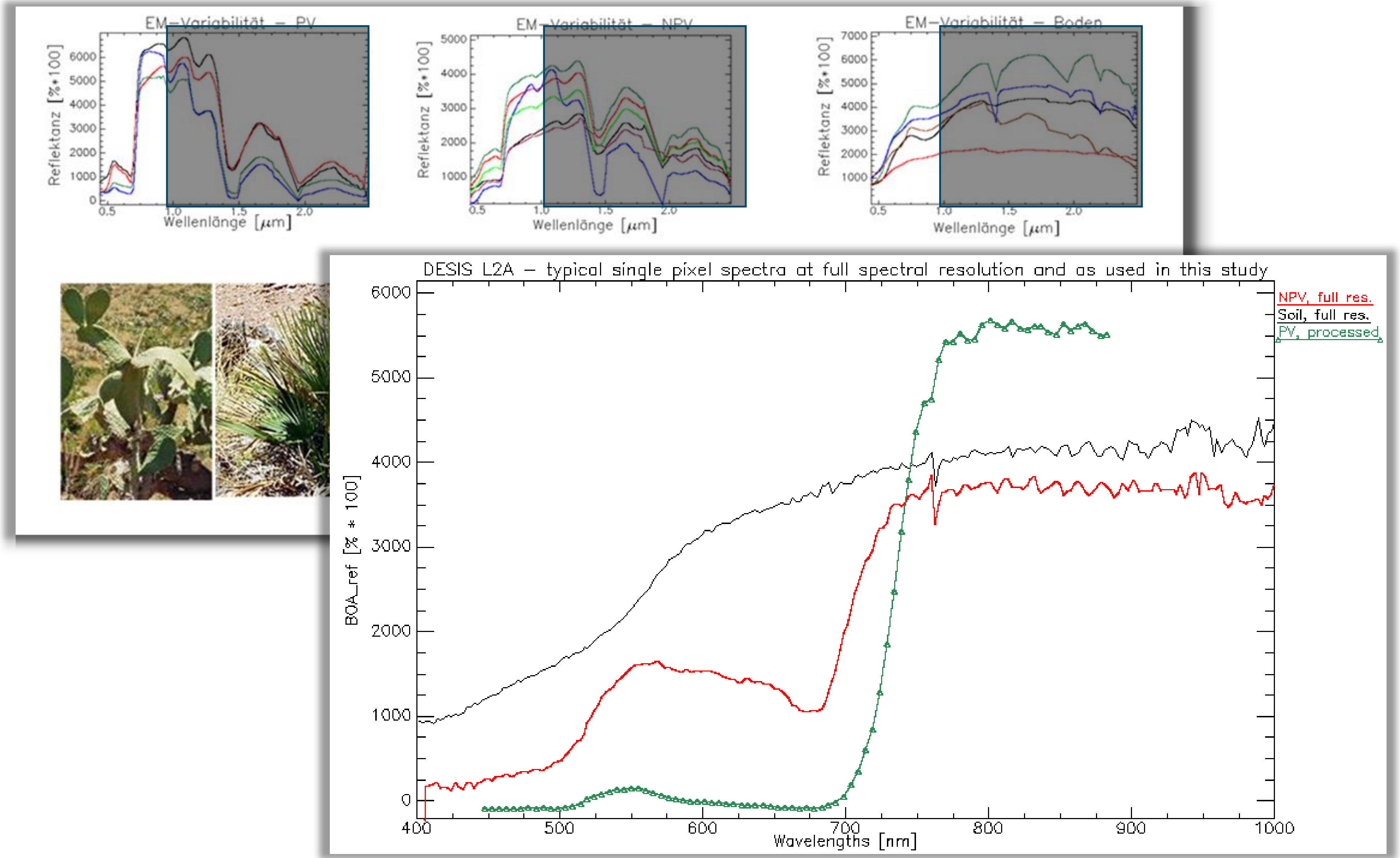
- A: Xylan & Cellulose
- B: Lignin & Cellulose
- C: Cellulose

} Vegetation without Chlorophyll => NPV

- D: Clay
- E: Carbonates

} => Soil

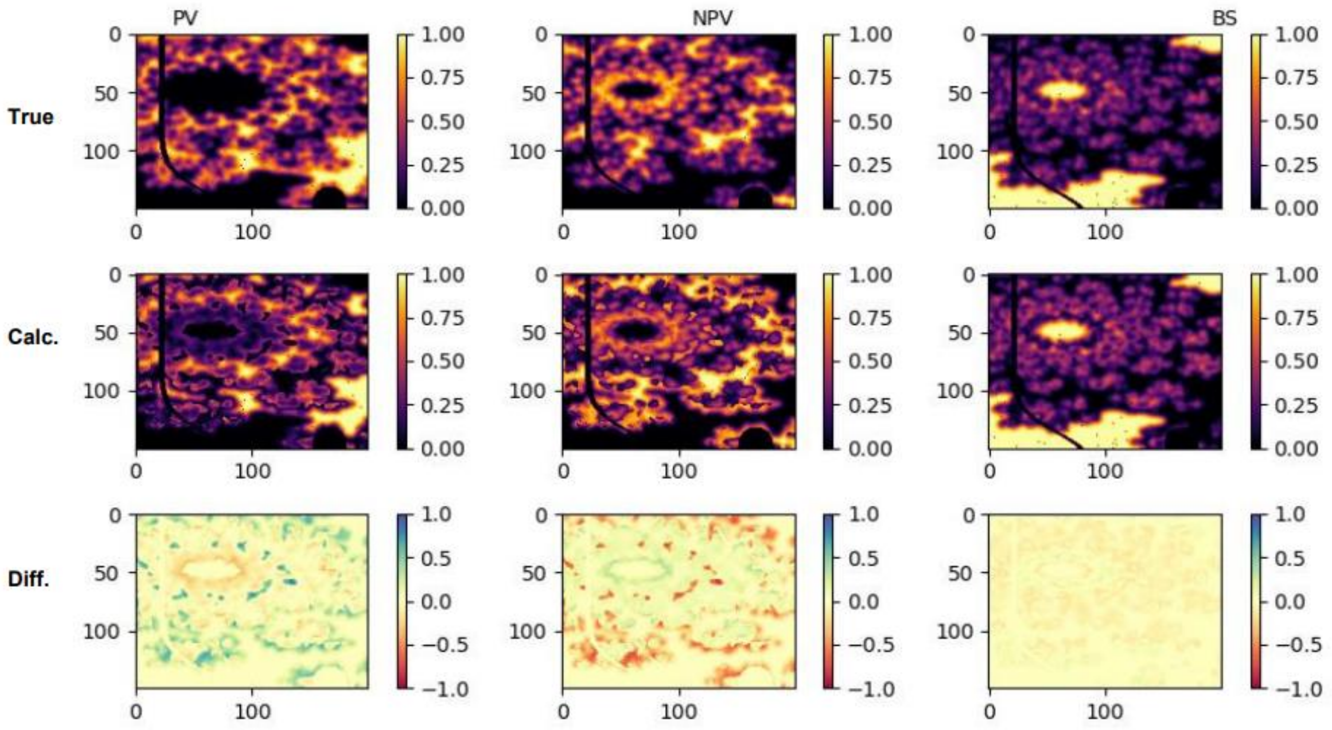
… but wait:
can DESIS separate between NPV & soil ?



DESIS
single pixel
spectra
of
PV (pre-proc),
NPV &
soil

fCover for DESIS data – simulations

- Validation using simulated fractal scenes

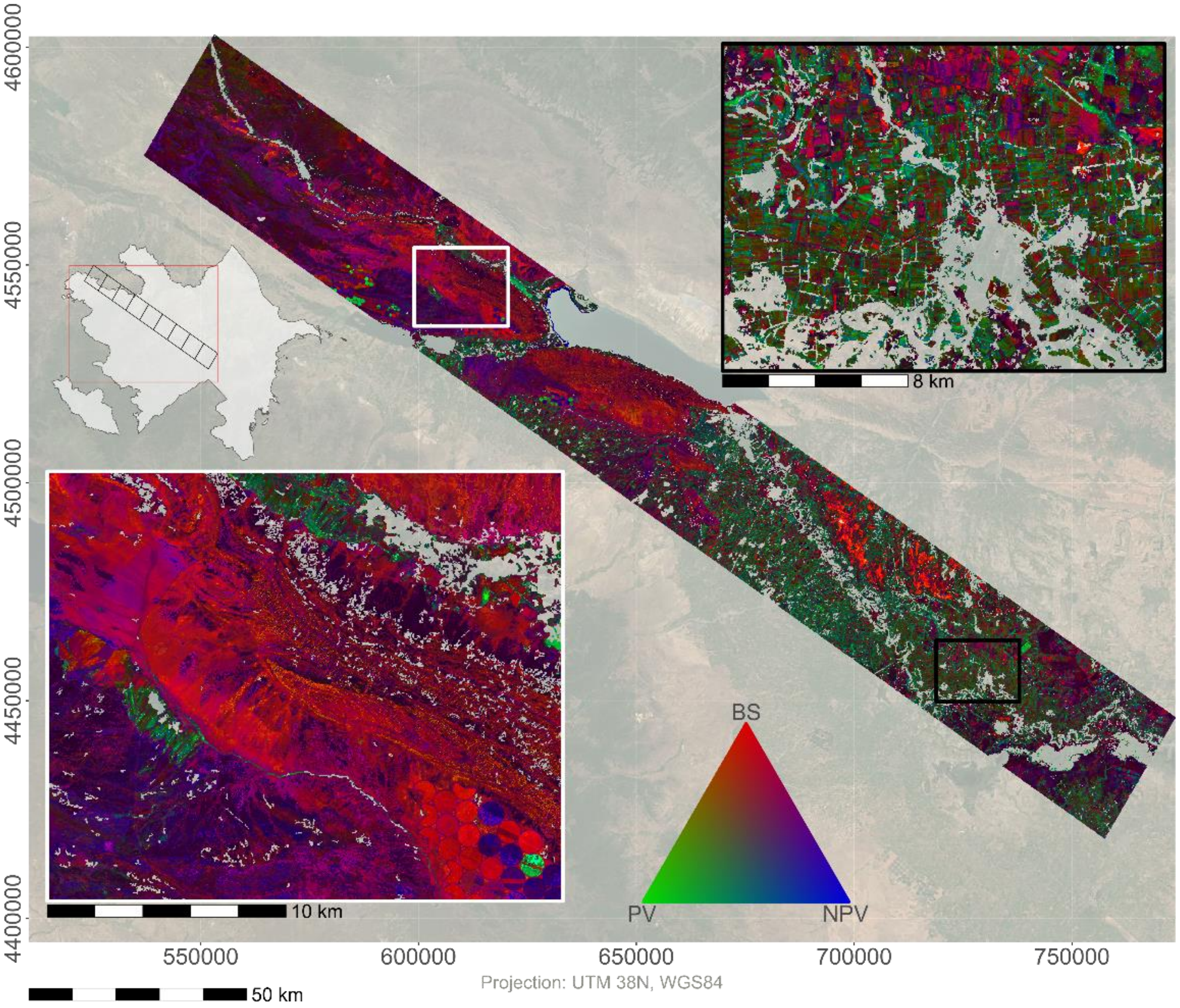


		RMSE
EnMAP	All	8%
	PV	7%
	NPV	10%
	BS	6%
DESIS	All	14%
	PV	8%
	NPV	15%
	BS	10%

An automated operational processor for the determination of fractional vegetation cover from DESIS observations

D. Marshall[1*], M. Bachmann[1], M. Habermeyer[1], U. Heiden[2], S. Holzwarth[1], T. Schmid[3]

fCover for DESIS – Azerbaijan Results



Note:

- Field measurements: 296 samples from August & Oct. 2018 (different years!)
- Living biomass, pot. vegetation & other parameters sampled, so MESMA classes are combined

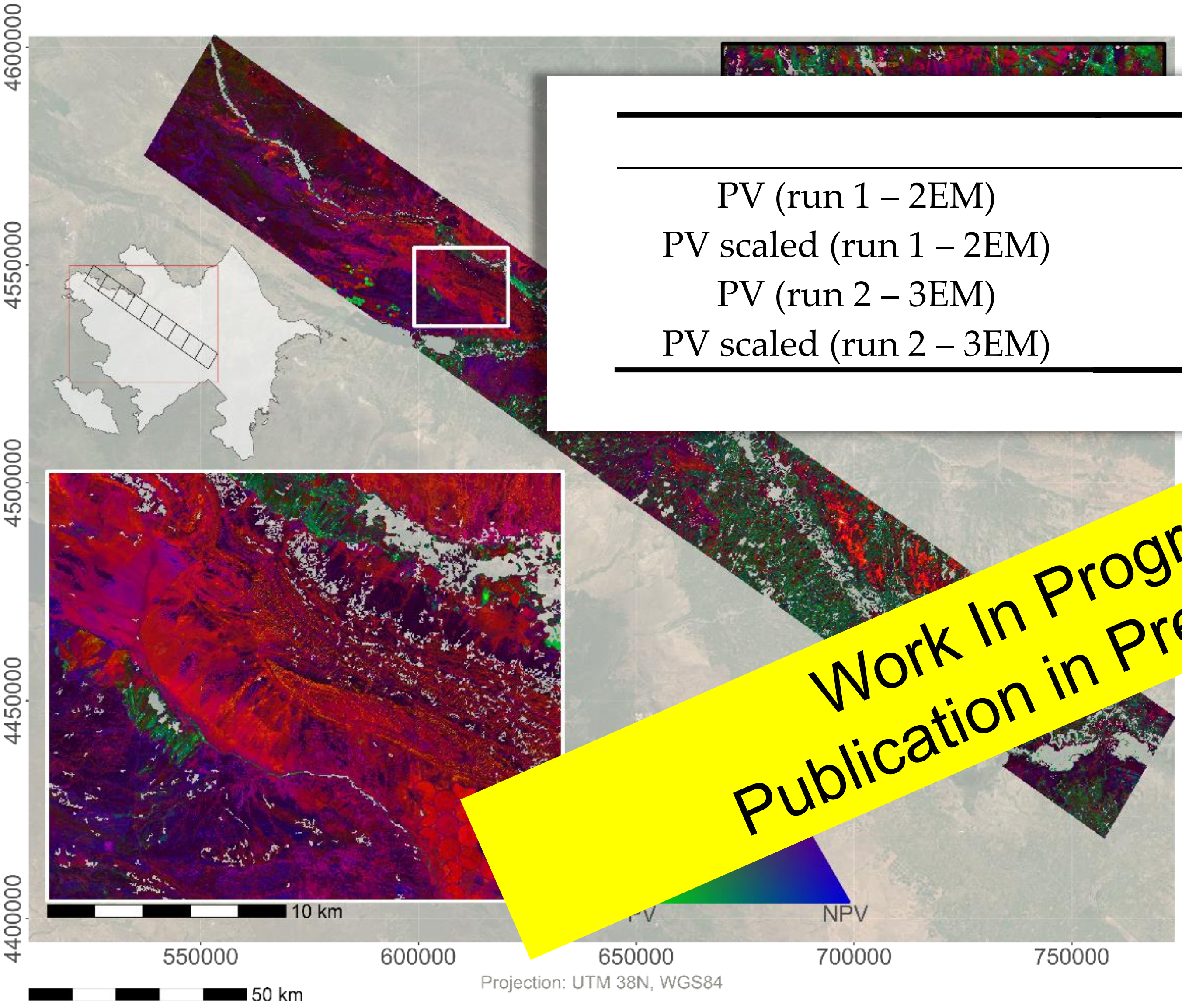
sensors
MDPI

Article

Detection of Grassland Degradation in Azerbaijan by Combining Fractional Cover Estimates Based on DESIS Data with Multi-Decadal Landsat NDVI Time Serie

Sarah Asam [1,*], Frederic Schwarzenbacher [2], David Marshall [1], and Martin Bachmann [1]

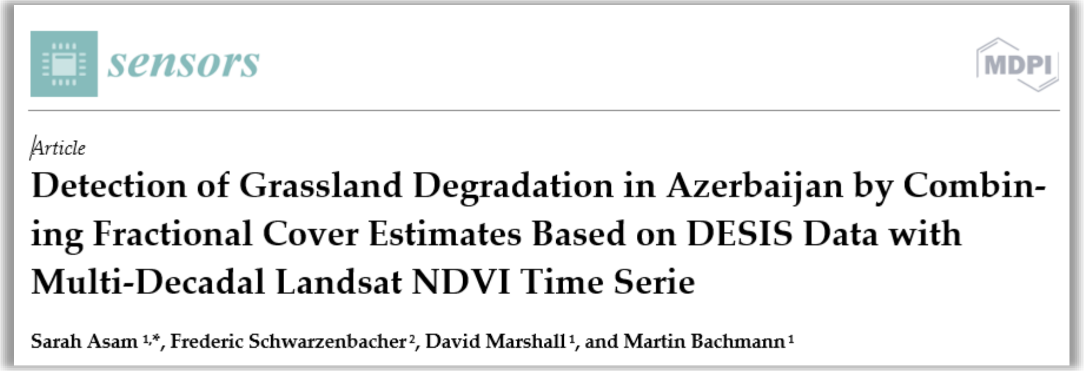
fCover for DESIS – Azerbaijan Results



	MAE	SD of absolute errors
PV (run 1 – 2EM)	8.13	10.97
PV scaled (run 1 – 2EM)	8.52	7.63
PV (run 2 – 3EM)	11.74	11.80
PV scaled (run 2 – 3EM)	13.88	14.32

Work In Progress
Publication in Preparation

...surements: 296 samples from ...gust & Oct. 2018 (different years!)
Living biomass, pot. vegetation & other parameters sampled,
so MESMA classes are combined


sensors

MDPI

Article

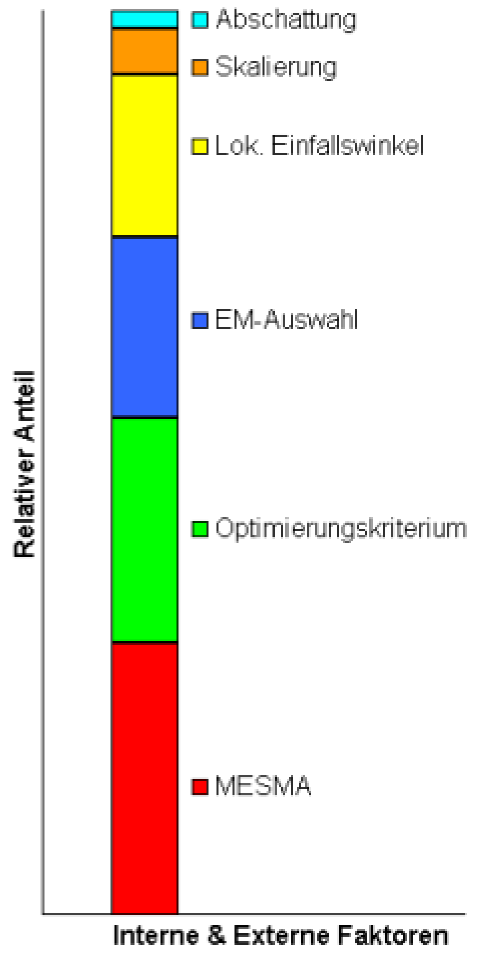
Detection of Grassland Degradation in Azerbaijan by Combining Fractional Cover Estimates Based on DESIS Data with Multi-Decadal Landsat NDVI Time Serie

Sarah Asam [1,*], Frederic Schwarzenbacher [2], David Marshall [1], and Martin Bachmann [1]

Summary – challenges and solutions

- EM variability and inherent between-class EM similarity
  - MESMA & check for ill-conditioned mixture models
- EM derivation in large datasets
  - SSEE spawning multiple tiles
  - EM labeling using RF classifiers trained on field SpecLibs
    - … but it‘s never 100% accurate !
- Identification of likely incorrect abundances
  - Include per-pixel reliability model and residual analysis
- Soil & NPV separation with DESIS
  - Indeed some potential, but better combine after unmixing to PV-cover
  - EnMAP, PRISMA and EMIT cover the SWIR with diagnostic NPV & soil features

- EM labeling, reduction of EM set <=> computation times
  - Work in progress…


Abschattung
Skalierung
Lok. Einfallswinkel
EM-Auswahl
Optimierungskriterium
MESMA
Relativer Anteil
Interne & Externe Faktoren

Land Degradation – Sustainable Development Goal (SDG)

SDG decided by the UN General Assembly in September 2015, came into effect January 2016

Land Degradation – SDG 15.3

DLR

SUSTAINABLE DEVELOPMENT GOALS KNOWLEDGE PLATFORM

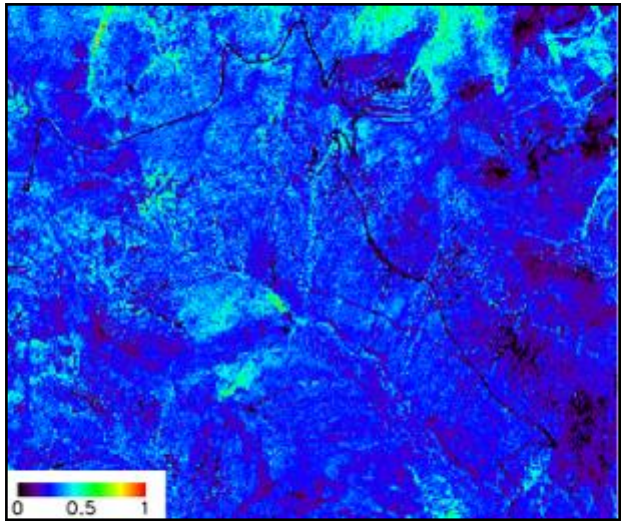
HOME SDGS HLPF STATES SIDS UN SYSTEM STAKEHOLDERS TOPICS PARTNERSHIPS RESOURCES ABOUT

TARGETS		INDICATORS	
15.1	By 2020, ensure the conservation, restoration and sustainable use of terrestrial and inland freshwater ecosystems and their services, in particular forests, wetlands, mountains and drylands, in line with obligations under international agreements	15.1.1	Forest area as a proportion of total land area
		15.1.2	Proportion of important sites for terrestrial and freshwater biodiversity that are covered by protected areas, by ecosystem type
15.2	By 2020, promote the implementation of sustainable management of all types of forests, halt deforestation, restore degraded forests and substantially increase afforestation and reforestation globally	15.2.1	Progress towards sustainable forest management
15.3	By 2030, combat desertification, restore degraded land and soil, including land affected by desertification, drought and floods, and strive to achieve a land degradation-neutral world	15.3.1	Proportion of land that is degraded over total land area
15.4	By 2030, ensure the conservation of mountain ecosystems, including their biodiversity, in order to enhance their capacity to provide benefits that are essential for sustainable development	15.4.1	Coverage by protected areas of important sites for mountain biodiversity
		15.4.2	Mountain Green Cover Index

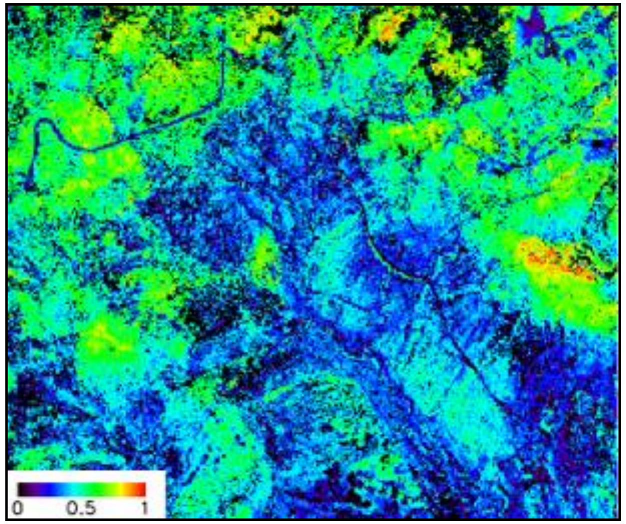
15 LIFE ON LAND

- fCover – Products & applications

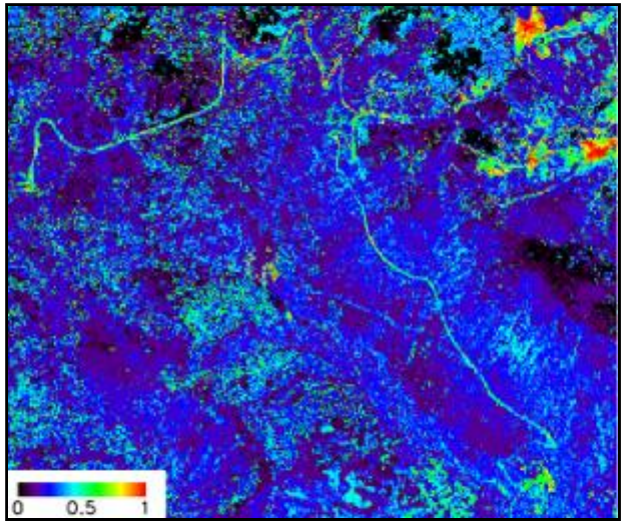
Products & Applications: Cabo de Gata (PhD Martin)



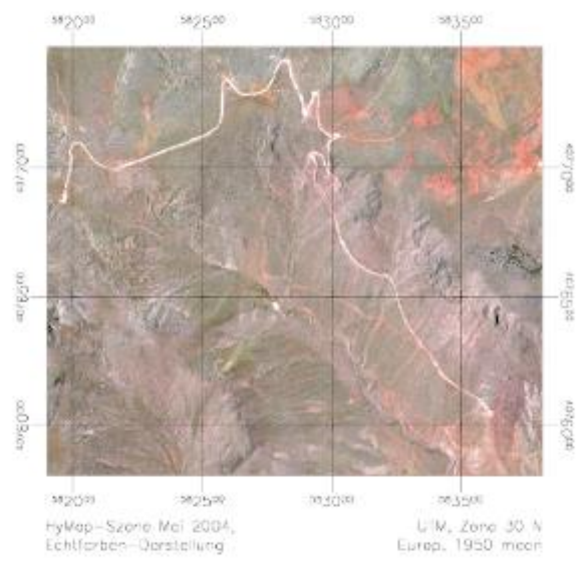
Bedeckungsgrad – PV

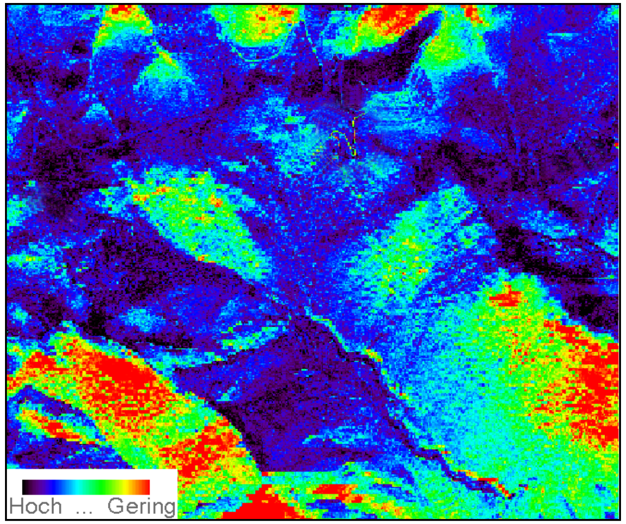


Bedeckungsgrad – NPV



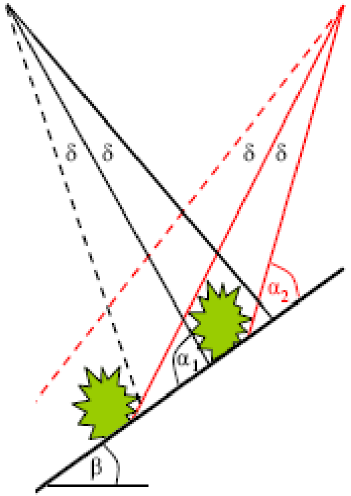
Bedeckungsgrad – Boden



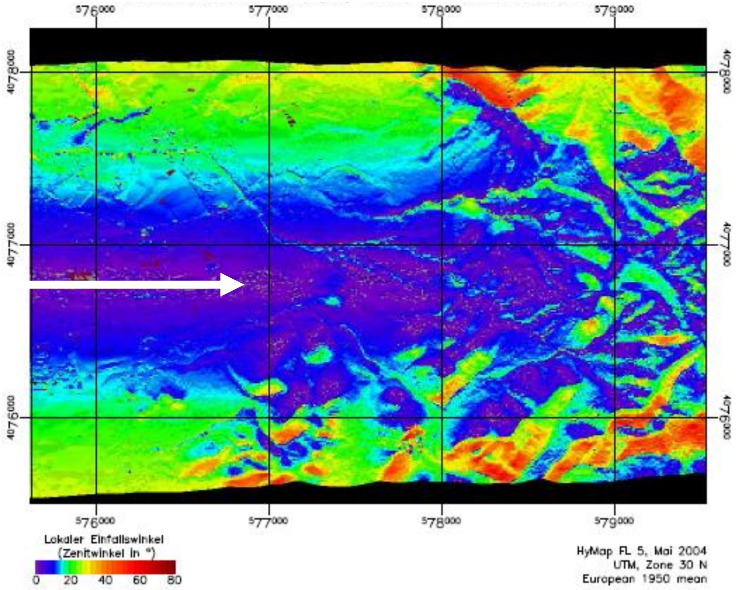


Bewertung der Güte

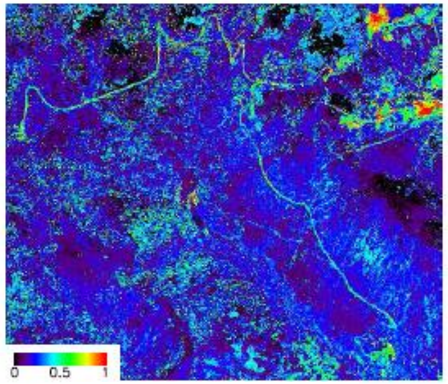
Products & Applications: Cabo de Gata (PhD Martin)



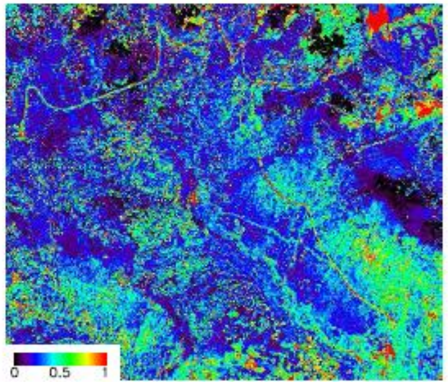
Lokaler Einfallswinkel – Übergang zu bewegtem Relief



Anteil an offenliegendem Boden



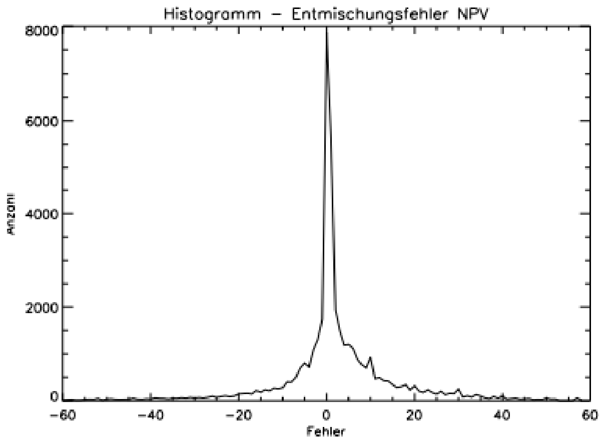
… ohne Korrektur



… mit empirischer Korrektur

Products & Applications: Cabo de Gata (PhD Martin)

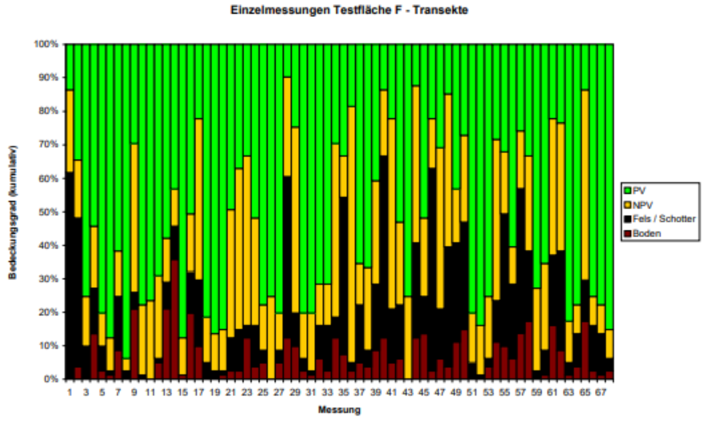
- Validierung:
  - Nutzung spektraler Bibliotheken zur Simulation von ~80.000 Mischmodellen
    Feldmessungen aus CdG, Calanas, Namibia, MEDSPEC (JRC), USGS & Johns Hopkins U.

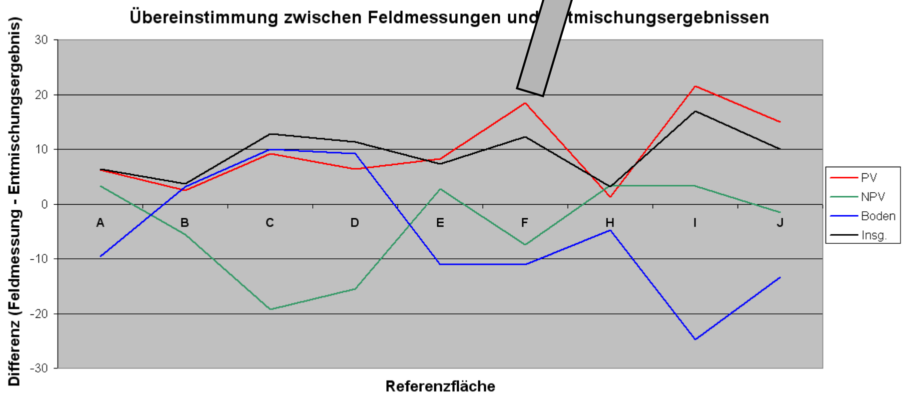


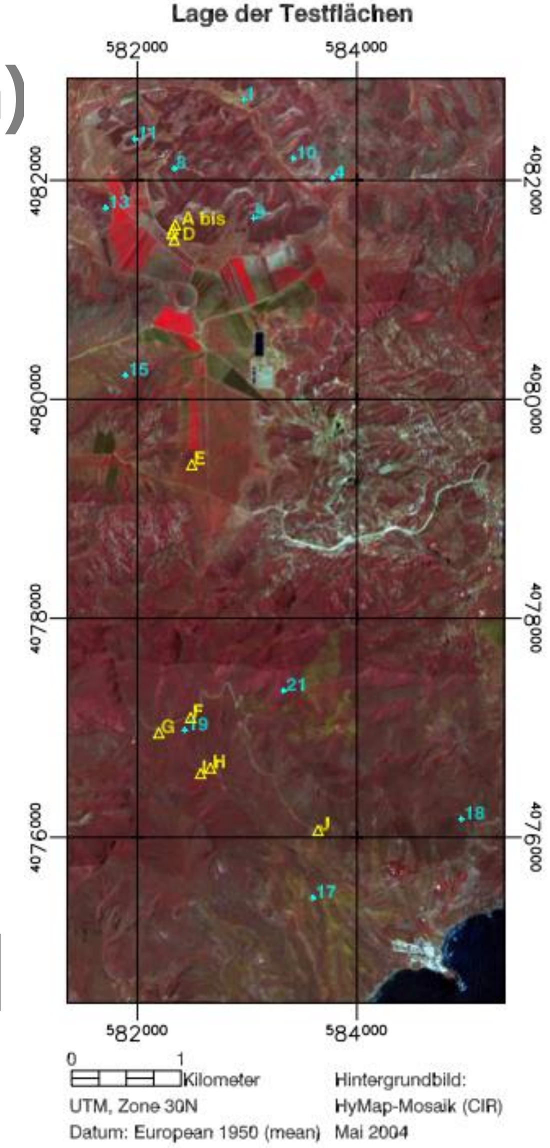
	Mittlerer Fehler*	Davon EM-Fehler**	R^2	Anteil in Prozent – Modelle mit Fehler unter 3%	Anteil in Prozent – Modelle mit Fehler über 10%
PV	5,2	3,2 (62%)	0,88	63	16
NPV	7,9	3,9 (49%)	0,73	52	23
Boden	6,9	4,0 (58%)	0,80	56	25

Products & Applications: Cabo de Gata (PhD Martin)

- Validierung:
  - Feldmessungen an 9 vorstratifiz. Tesfläche aus je 40-80 Einzelmessungen



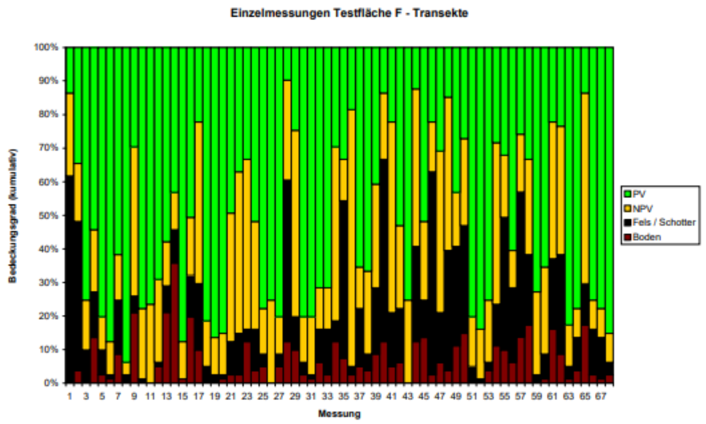


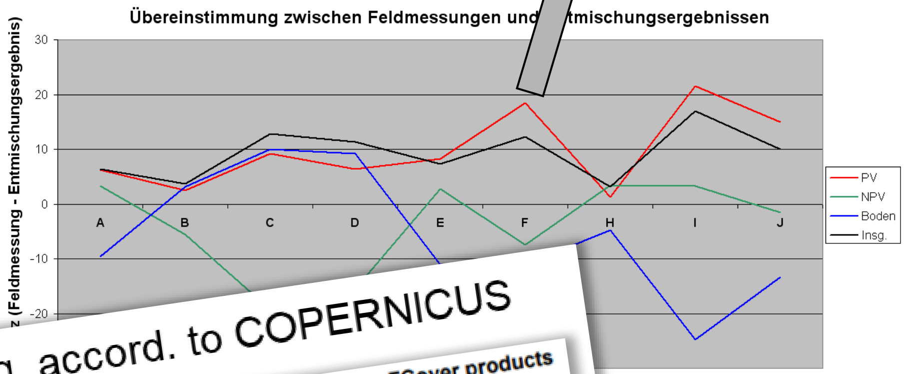


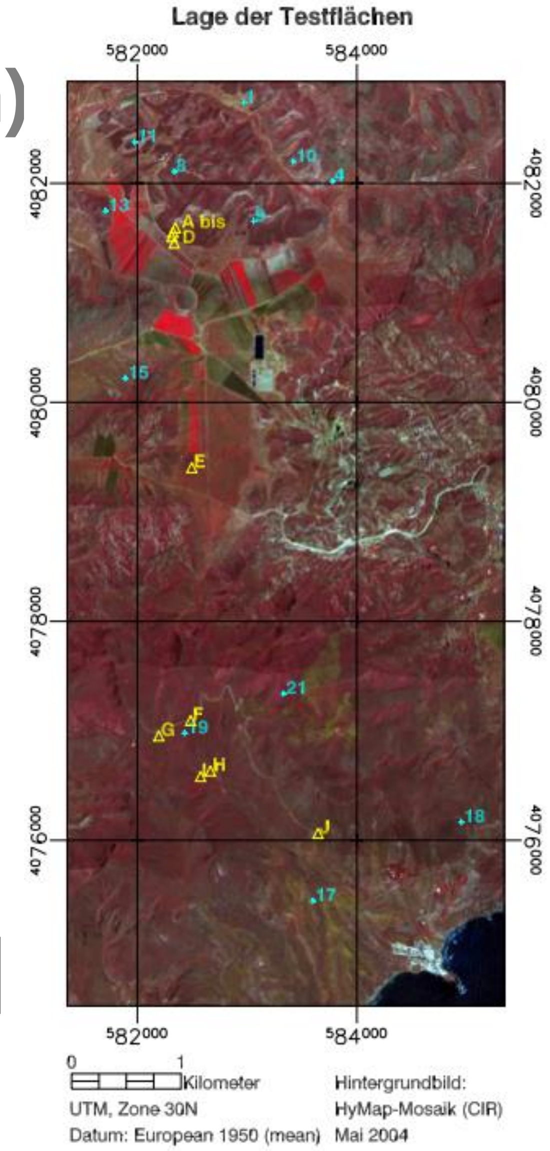
	Klasse	Mittlerer Fehler*	Stdev **	Anteil in Prozent Modelle mit Fehler ≤ 5%	≥ 10%
µMESMA, lok. Einfallsw. korrigiert	PV	8,5	5,4	40	40
	NPV	10,4	13,1	50	30
	Boden	15,1	9,3	20	70
	Insg.	**11,3**	**6,8**	**37**	**47**
	Insg. RW ***	**9,3**			

Products & Applications: Cabo de Gata (PhD Martin)

- Validierung:
  - Feldmessungen an 9 vorstratifiz. Tesflächen aus je 40-80 Einzelmessungen







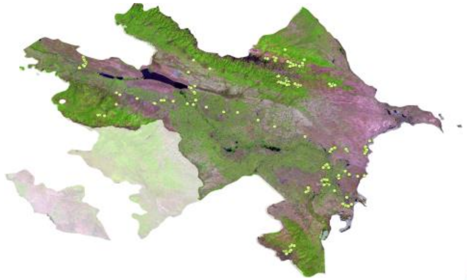
	Klasse	Mittlerer Fehler*	Stdev **	Anteil in Prozent Modelle mit Fehler $\leq 5\%$	$\geq 10\%$
µMESMA, lok. Einfallsw. korrigiert	PV	8,5	5,4	40	40
	NPV	10,4	13,1	50	30
	Boden	15,1	9,3	20	70
	Insg.	**11,3**	**6,8**	**37**	**47**
	Insg. RW ***	**9,3**			

Uncertainty req. accord. to COPERNICUS

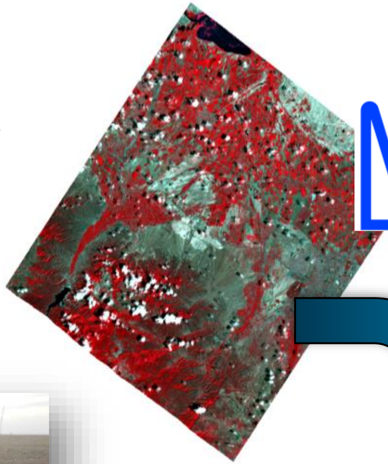
Table 2: Copernicus Global Land product requirements for FAPAR, FCover products

	Optimal	Target	Threshold
FAPAR FCover	5%	10%	20%

Products & Applications: Aserbaidschan (w. Sarah & David)



Aserbaidschan,
Degradation of grasslands



EnMAP fCover
Prozessor