

Kartierung von Energiepflanzen und Ableitung des Biomethanpotentials

Workshop des DGPF Arbeitskreises
„Auswertung von Fernerkundungsdaten“
Fernerkundungsbasierte Technologien und Lösungen für
Anwendungen in der Landwirtschaft

Björn Waske

Institut für Geodäsie und Geoinformation, Universität Bonn

EnMAP – BMP

„Methoden zur fernerkundlichen Quantifizierung des Biomethanpotentials landwirtschaftlicher Kulturflächen unter spezieller Berücksichtigung der EnMAP-Mission“

- Thomas Udelhoven (Universität Trier, Fernerkundung und Geoinformatik)
- Björn Waske (Universität Bonn, Institut für Geodäsie und Geoinformation)
- Holger Lilienthal (Julius Kühn-Institut, Institut für Pflanzenbau und Bodenkunde, Braunschweig)
- Philippe Delfosse (Centre de Recherche Public – Gabriel Lippmann, Luxembourg)
- Christoph Emmerling (Universität Trier, Bodenkunde)

Ziele:

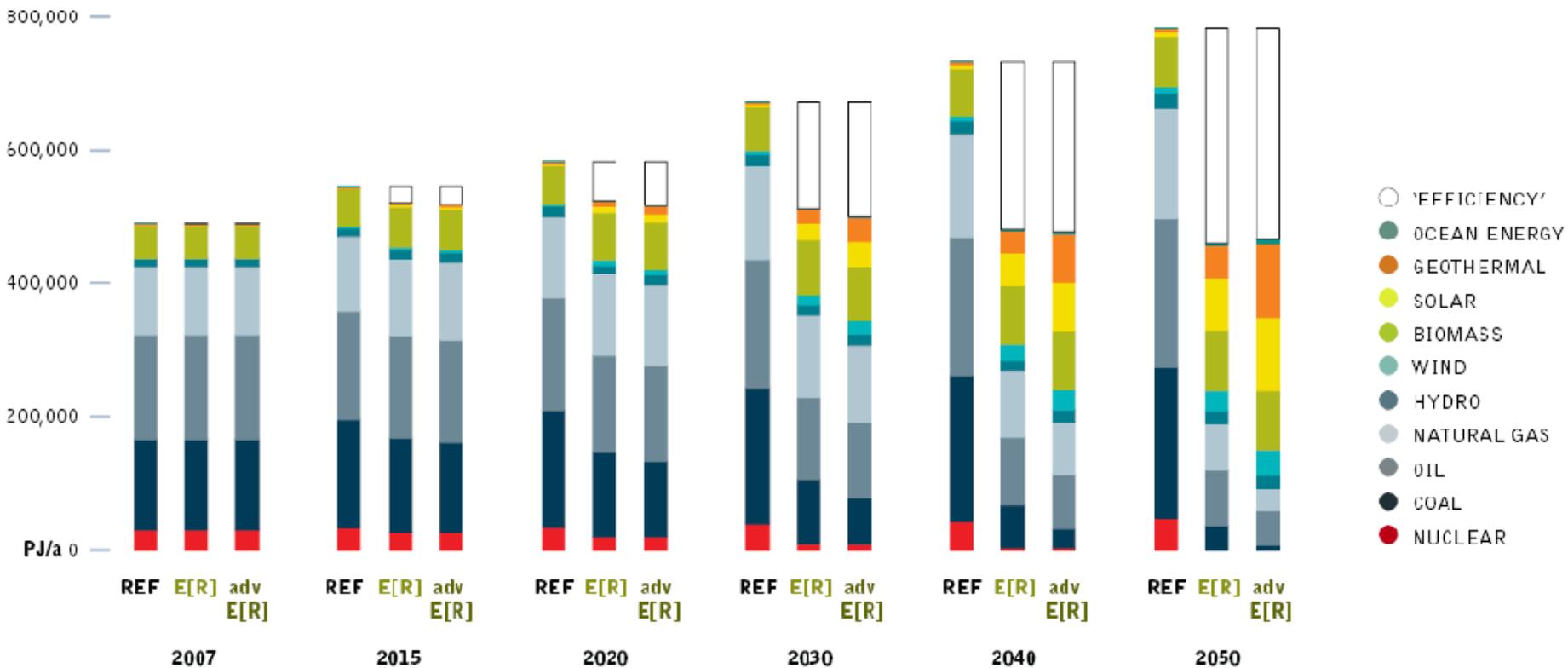
- Entwicklung von Methoden zur
 - (i) Kartierung von Energiepflanzen,
 - (ii) Ableitung des Biomethanpotenzials (BMP) und der Biomasse
 - Integration der Parametern in Pflanzenwachstumsmodelle
- EnMAP-Mission und Integration von SAR-Daten

3

Potentielle Nutzer:

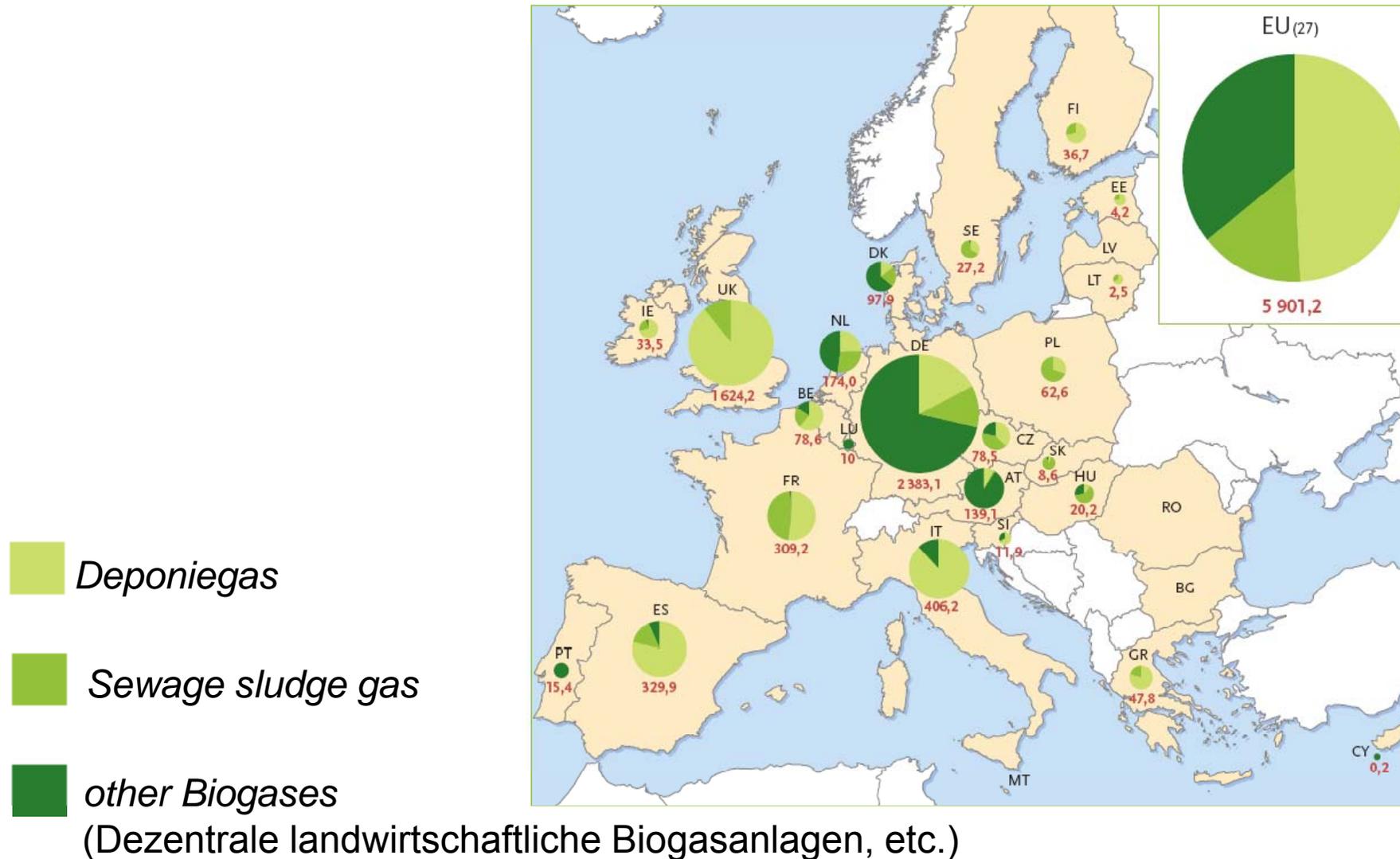
- Öffentliche Einrichtungen
- Landwirte
- Betreiber von Biogasanlagen

- Ursachen für die Förderung erneuerbarer Ressourcen:
 - Abhängigkeit von Energieimporten verringern
 - Treibhausgase reduzieren
 - neue ökonomische Möglichkeiten



Globale Entwicklung des Primärenergiebedarfs unter drei Szenarien
(Greenpeace International, European Renewable Energy Council, 2010)

Biogas Primärenergieproduktion in Europa 2007



Biogas 2008: 49

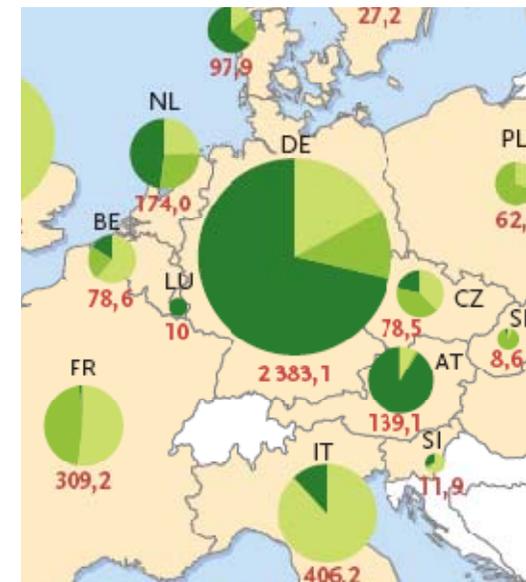
Beispiel: Deutschland

- 2.4 Mtoe durch Biogas 2007 (größter Produzent in der EU)
- 71.2 % davon in kleinen Biogasanlagen landwirtschaftlicher Betriebe erzeugt
- 98 % dieser Anlagen nutzen (u.a.) Energiepflanzen
- wichtigste Pflanze: Mais
- Maispreis in Deutschland:
18 €/Tonne (2006) → 33 €/Tonne (2007)

Kritische Aspekte zu Energiepflanzen

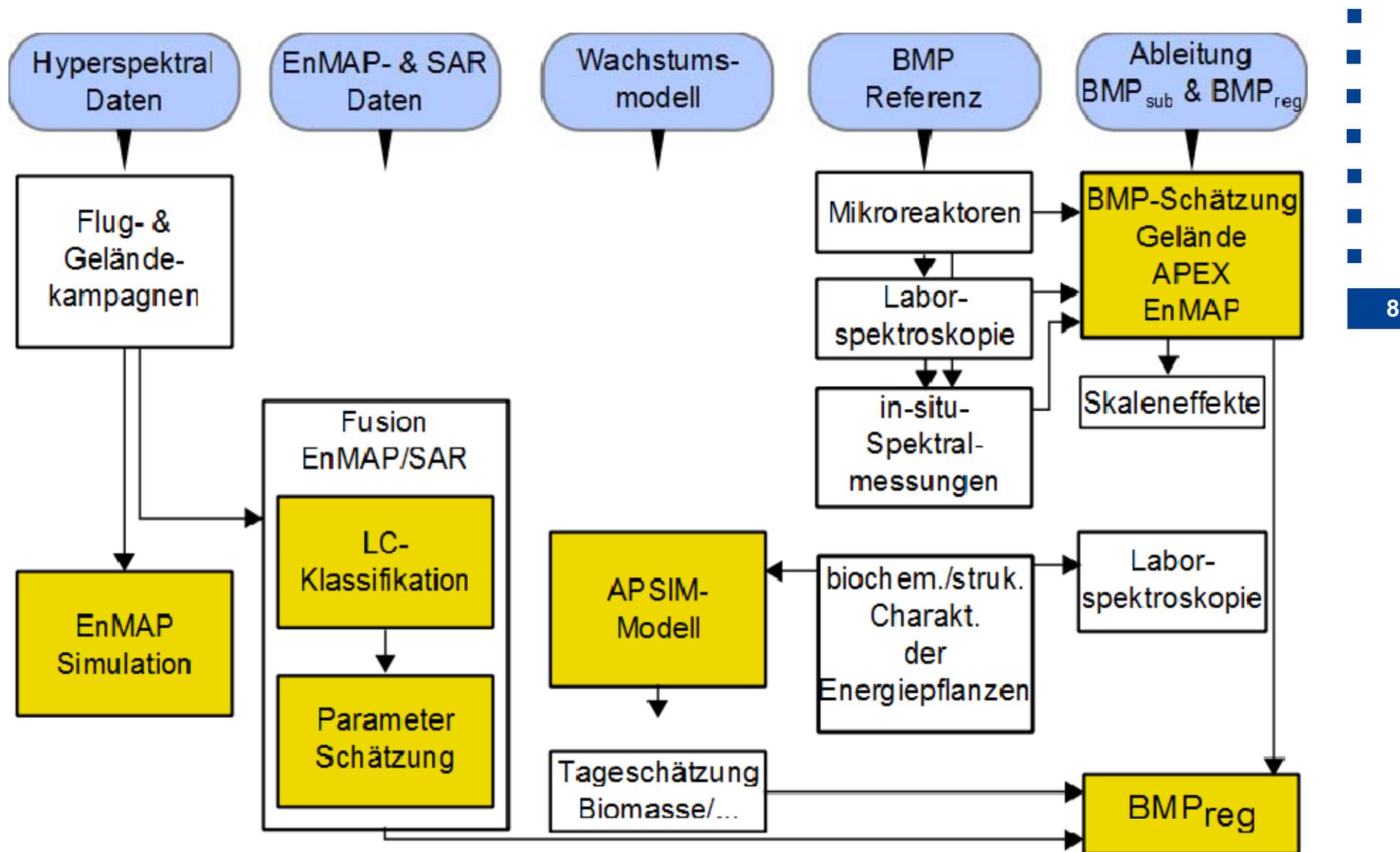
- Artenverarmung (Monokultur Mais)
- CO₂-Verbrauch des Anbaus vs. Energieertrag?
- steigende Nahrungsmittelkosten

Biogas Barometer, 2008

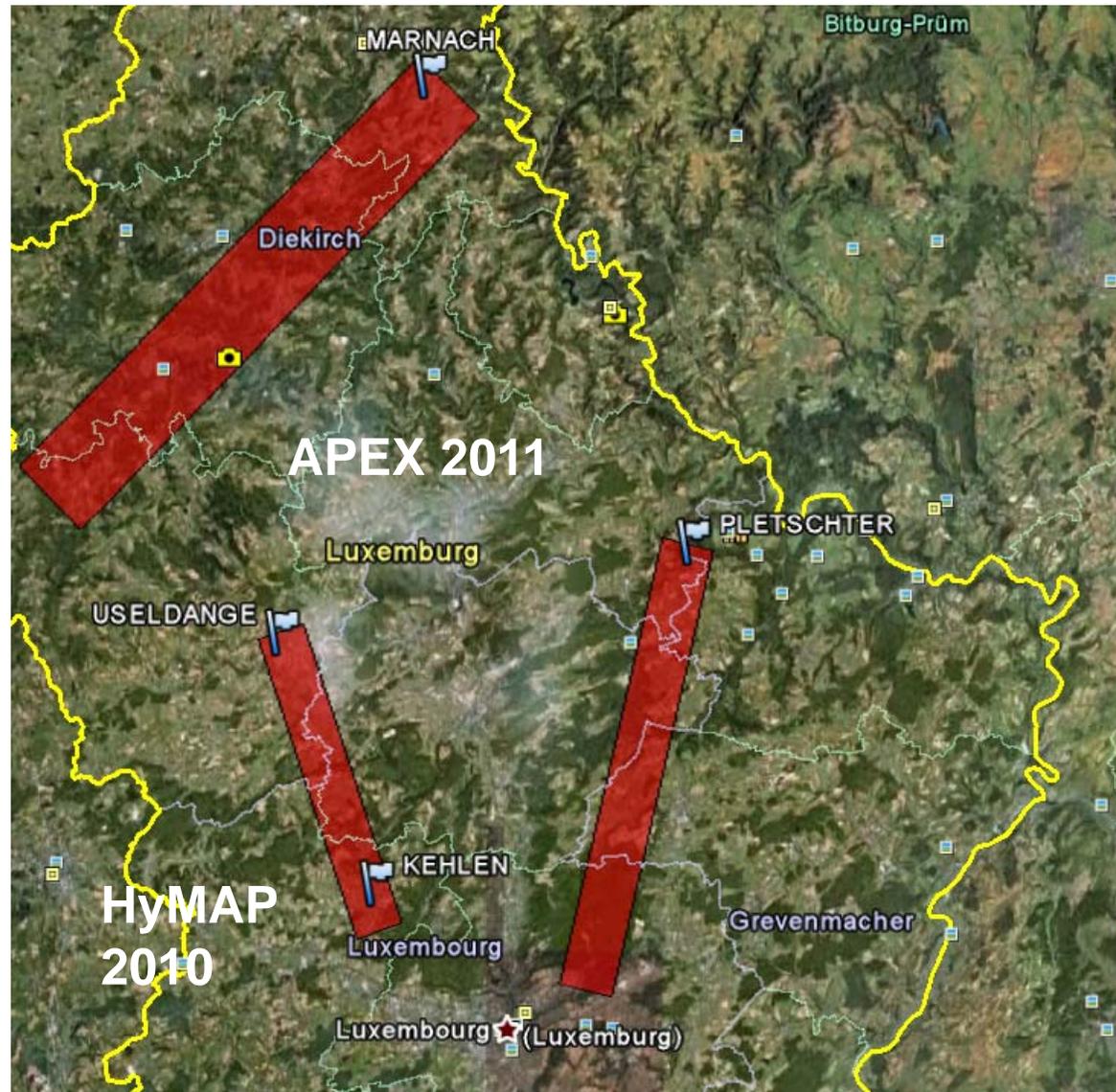


- Nutzen des räumlich differenzierten Biomethanpotentials (BMP) für den landwirtschaftlichen Betrieb:
 - Optimierung der Biogasanlage
 - Vermeidung von Übersäuerung
 - Kostensenkung
 - Gewinnmaximierung
- aber die Bestimmung des BMPs ist aufwendig, und benötigt in der Regel mehr als 30 Tage
- im Rahmen des FNR-BIONIR Projektes hat sich gezeigt das das BMP mittels spektroskopischer Methoden zu bestimmen ist
- erste Versuche sind vielversprechend (Mais: $r^2=0.95$, Sorghum: $r^2=0.99$, Sonnenblumen: $r^2=0.98$)

Übersicht Arbeitsplan

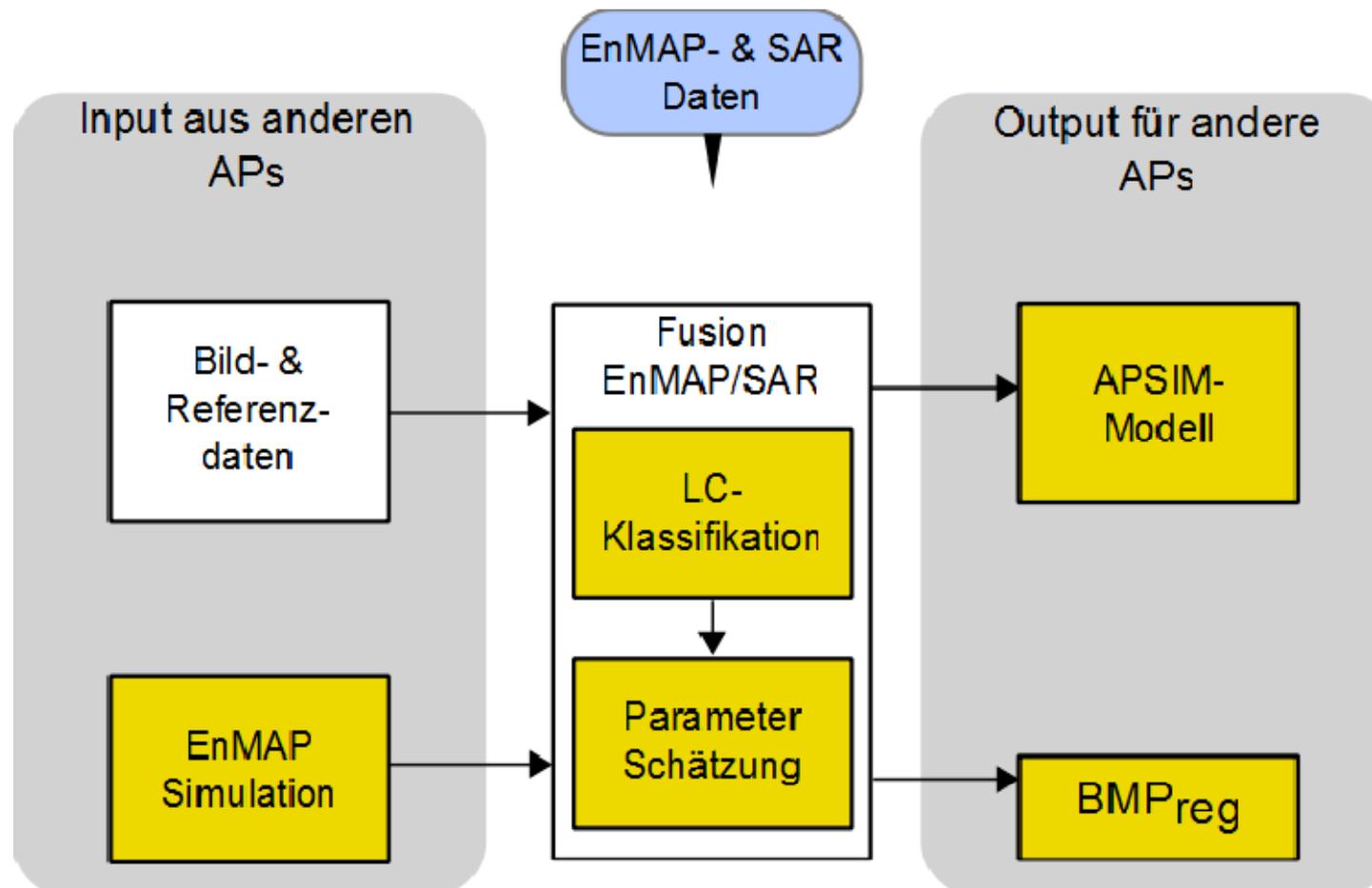


Untersuchungsgebiet



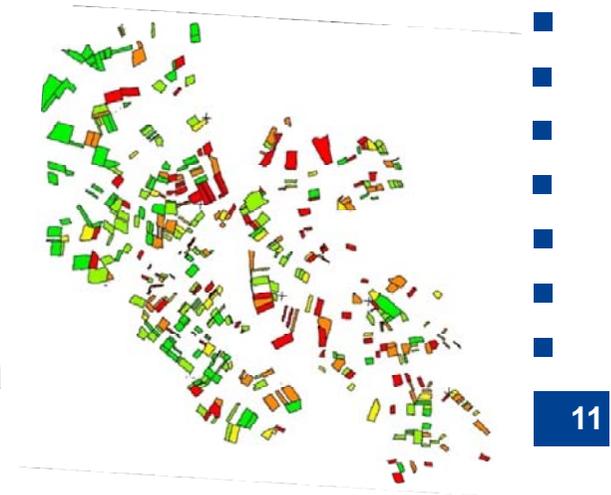
-
-
-
-
-
-
-

II. Fusion EnMAP & SAR



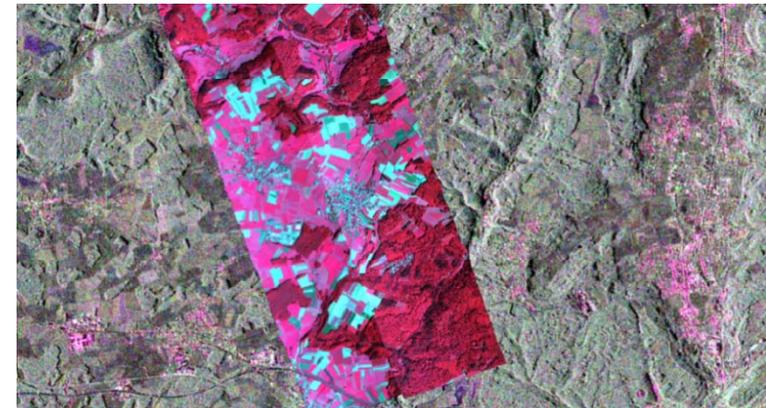
-
-
-
-
-
-
-
-

- Ableitung des regionalen BMP benötigt
 - Erfassung der relevanten landwirtschaftlichen Flächen (d.h., Mais)
 - Informationen bzgl. der räumliche Verteilung von Pflanzenstrukturparametern (Biomasse, LAI, etc.)
- Methoden
 - Klassifikation (Erfassung der Maisflächen)
 - Regression (Ableitung von Pflanzenstrukturparametern, z.B. LAI und Biomasse)
 - Datenfusion
- Daten
 - Hyperspektrale & SAR-Daten



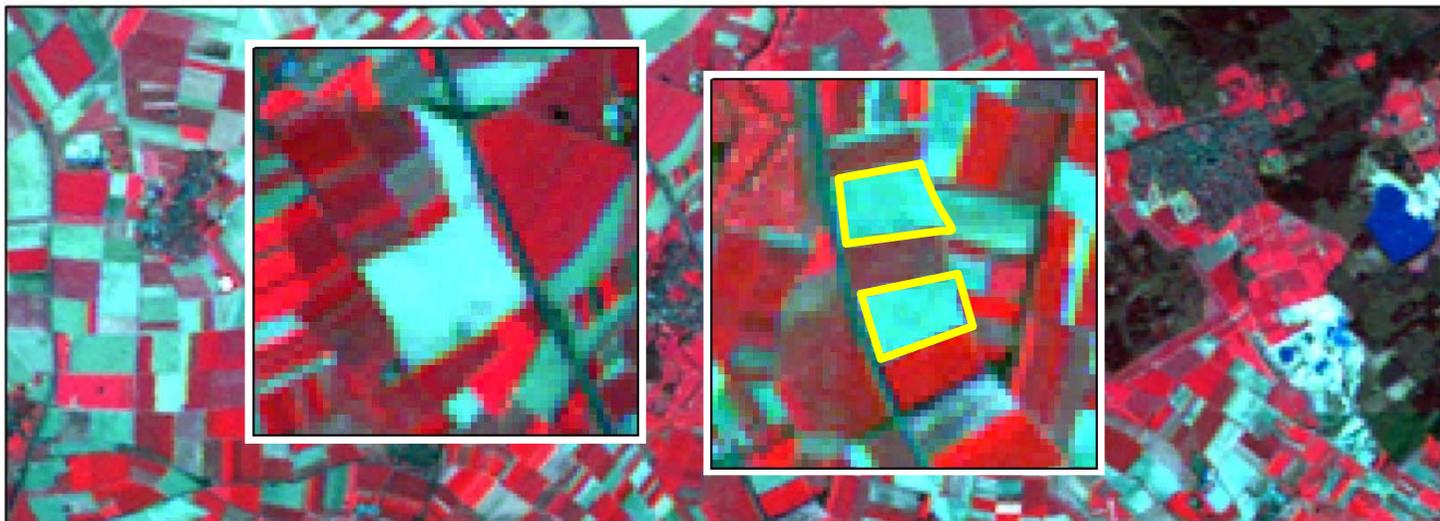
Weshalb Datenfusion?

- zunehmende Verfügbarkeit von multisensoralen FE-Daten (z.B. TerraSAR-X, RapidEye, EnMAP, Sentinel)
 - komplementäre Informationen multisensoraler Daten
 - höhere Genauigkeiten
 - erfordert aber komplexere Modelle
 - Untersuchungen zu gemeinsamer Nutzen von
 - multispektralen & SAR-Daten ✓
 - hyperspektralen & SAR-Daten –
- SAR + (simulierte) EnMAP Daten

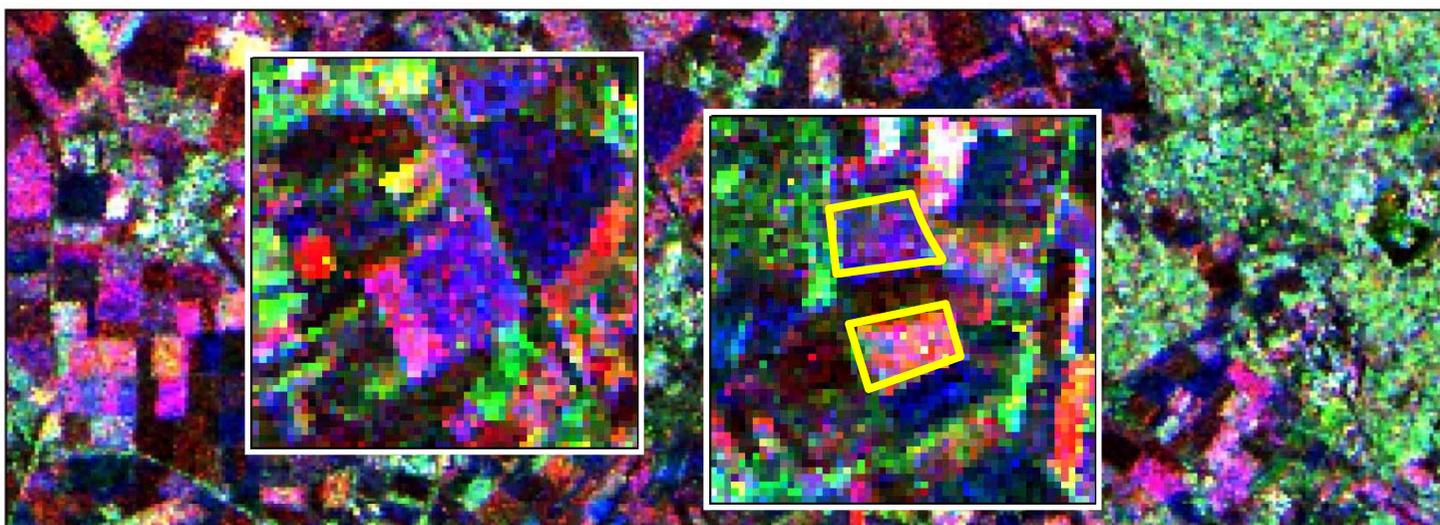


II. Fusion EnMAP & SAR

Landsat TM 5



ERS-2 Komposit



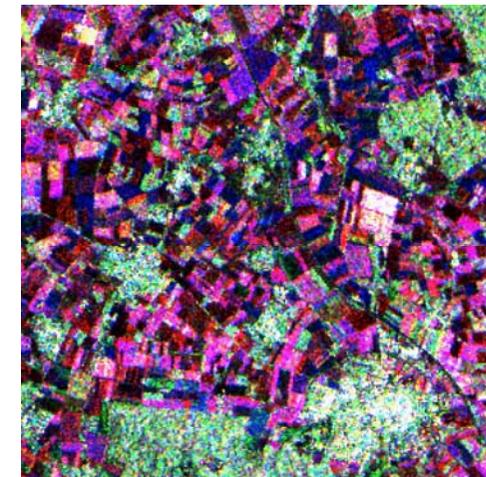
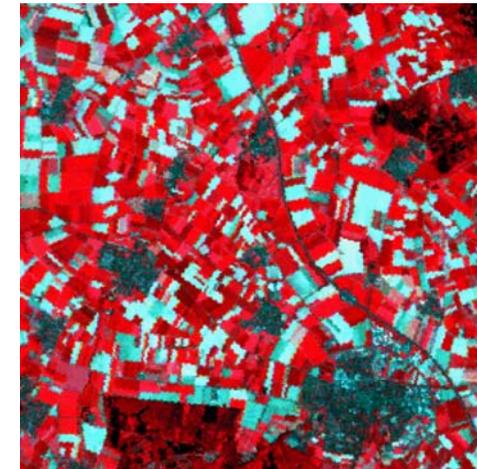
Hypespektrale und multisensorale Datensätze:

Statistische Verfahren (z.B., *Maximum Likelihood Klassifikation*) sind weit verbreitet, aber:

- benötigen „relativ“ große Trainingsdatensätze (MLK min. $d+1$, besser 10d, ideal 100d)
- „Hugh-Phänomen“ bei hoch-dimensionalen Datensätzen, z.B. Hyperspektraldaten (oder Zeitreihen)
- multitemporale und multisensorale Datensätze: passendes statistisches Modell häufig nicht bekannt; bei einer zu geringen Anzahl an Trainingsdaten kann es zudem nicht zuverlässig geschätzt werden
- multitemporal, multisensorale und hyperspektrale Daten können redundante und irrelevante Informationen

Fusion multispektral + SAR

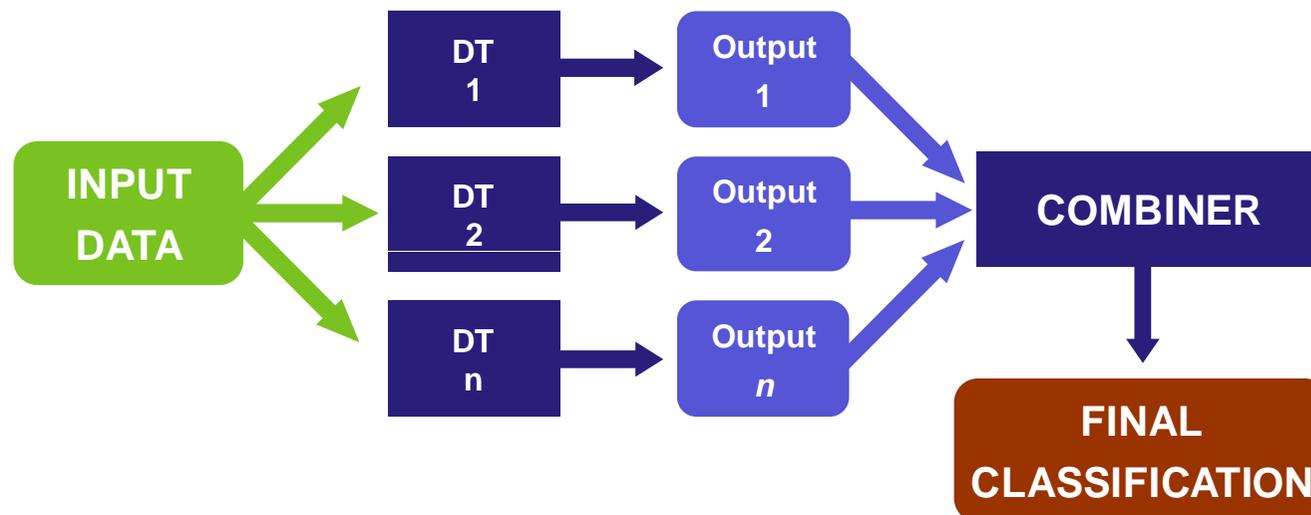
- Klassifikationsverfahren:
 - Maximum Likelihood Klassifikator (MLK)
 - Decision Tree (DT)
 - Support Vector Machines (SVM)
 - Random Forests (RF)
- Datengrundlage:
 - landwirtsch. Region nahe Bonn
 - (1) multitemporale SAR Daten, (2) Landsat 5 TM, (3) SAR+TM
- Validierung:
 - alle Ergebnisse wurden mittels der gleichen unabhängigen Testdaten validiert



Random Forests (Breiman, 2001)

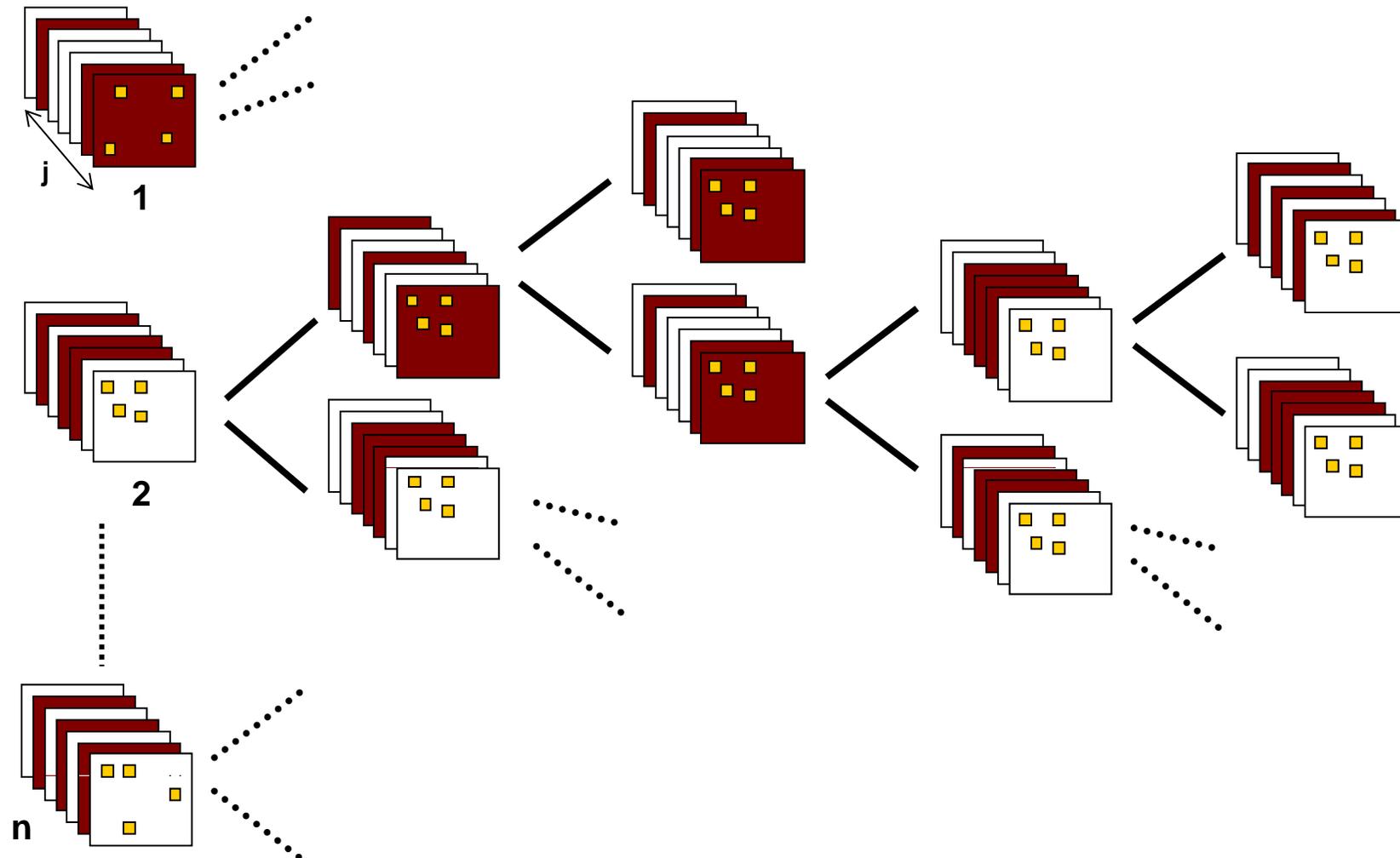
Classifier ensemble bzw. Multiple Classifier System

- Kombination von verschiedenen Klassifikationsmethoden bzw. unterschiedlichen Varianten der selben Methode
- Annahme: jeder Klassifikator generiert Fehler, die nicht von der Mehrheit innerhalb des Ensembles wiederholt werden
- durch eine Kombination der unterschiedlichen Ergebnisse wird die Genauigkeit verbessert



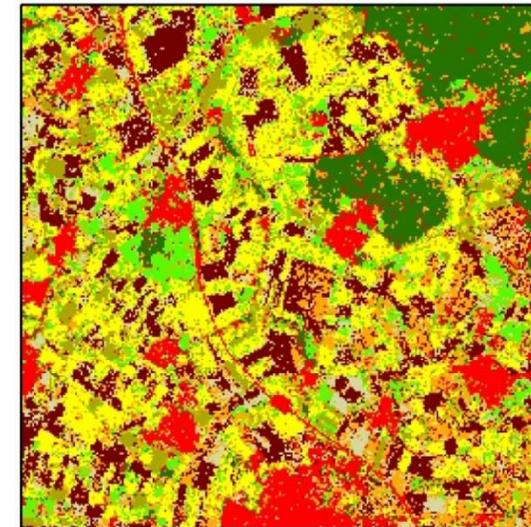
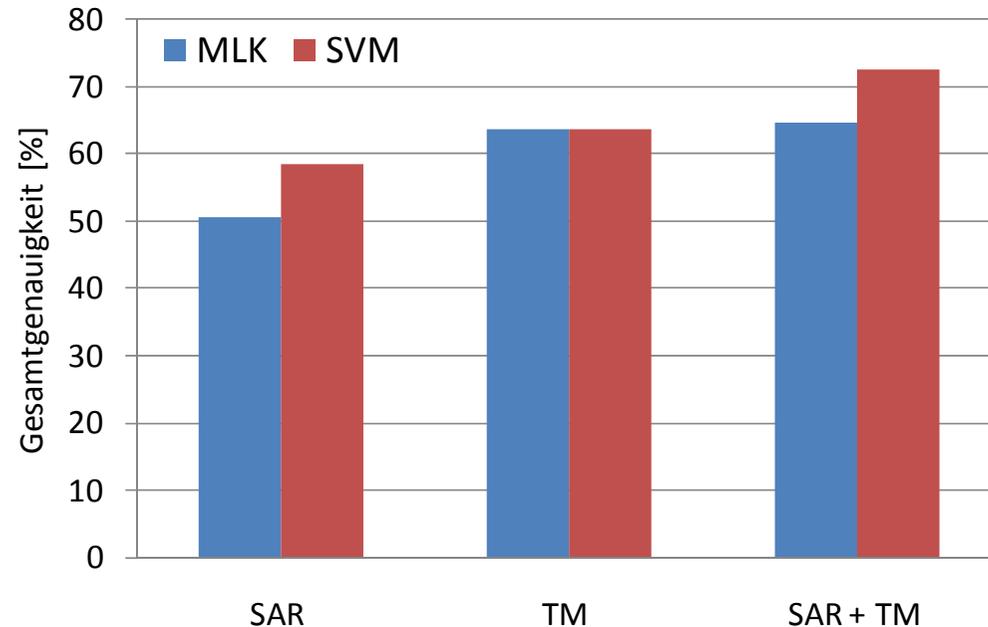
Random Forests:

Decision Tree + „*bagging*“ und „*random feature selection*“

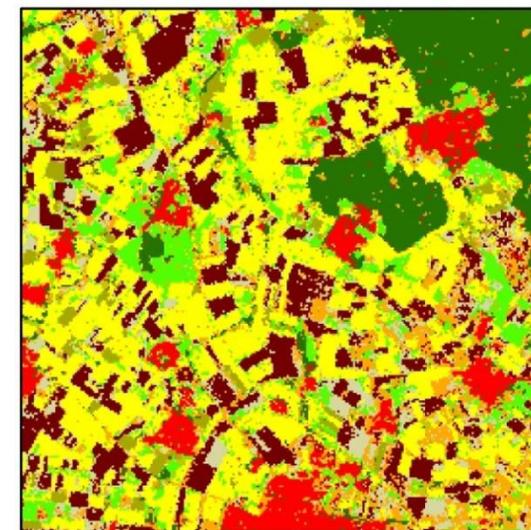


Fusion multispektral + SAR

- Gesamtgenauigkeit (SVM) wird mittels multisensoraler Daten um 14% bzw. 9% verbessert
- SVM und RF zeigen ähnliche Performance
- Ergebnis kann weiter durch *Decision Fusion* verbessert werden



MLC

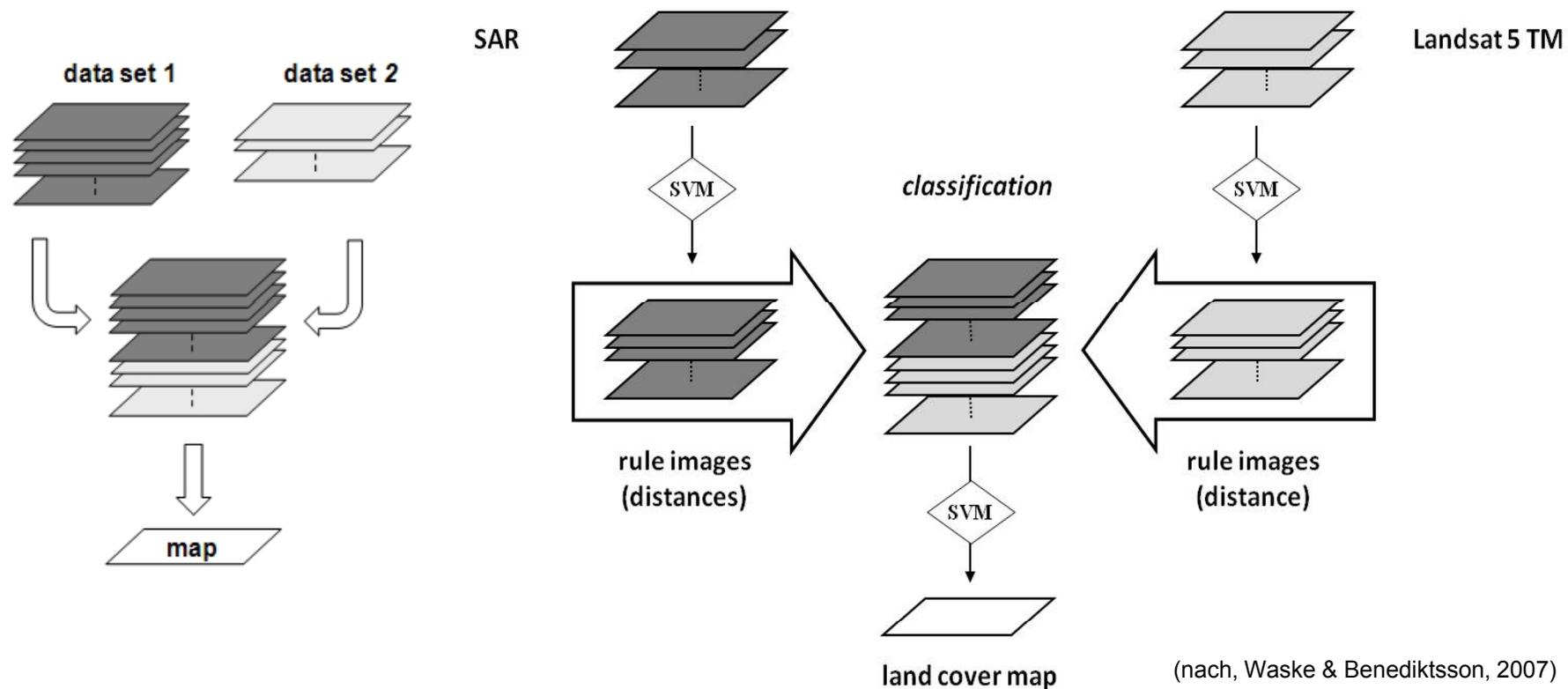


SVM



Fusion multispektral + SAR

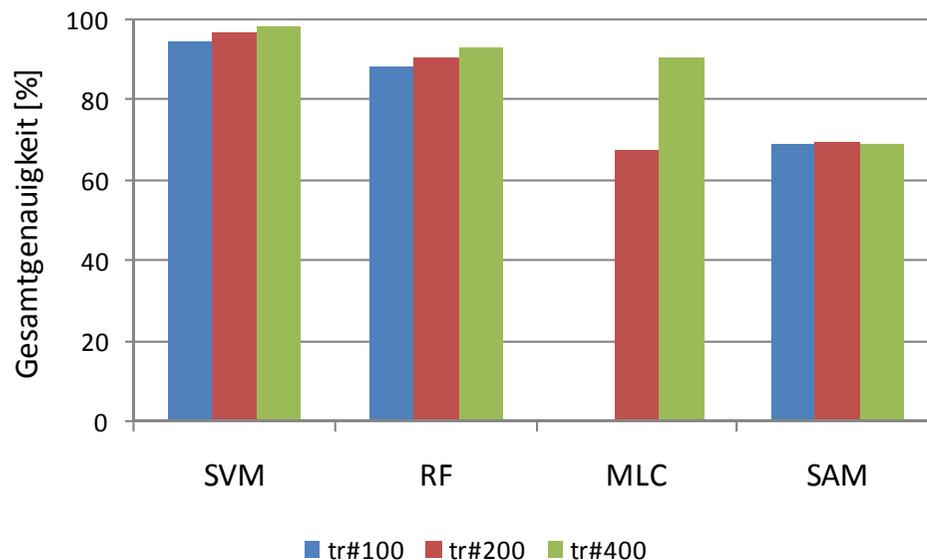
- *Data Fusion*: Fusion auf Datenebene (vor der Klassifikation)
- *Decision Fusion*: Fusion nach der Klassifikation



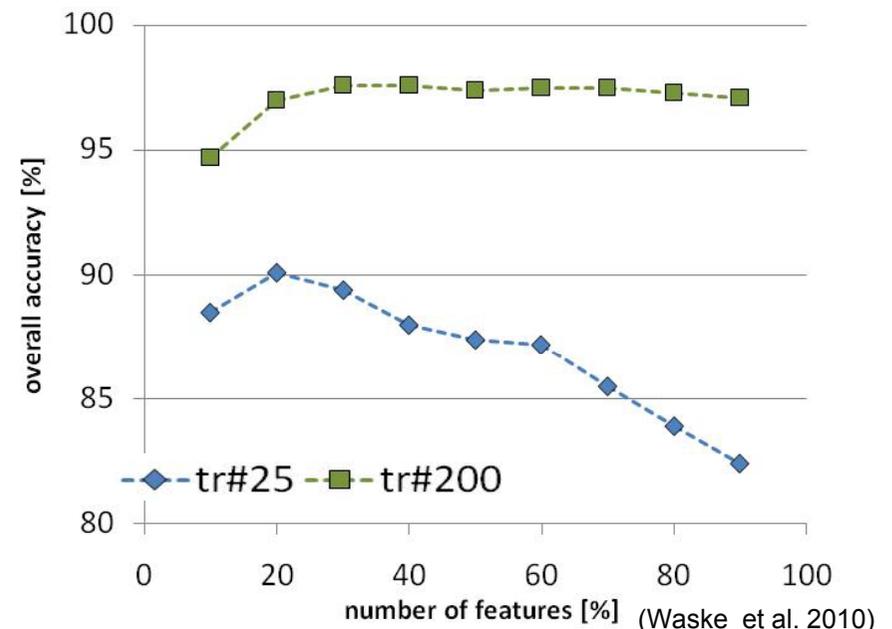
Klassifikation von hyperspektral Daten

- Verfahren wie SVM und RF erzielen häufig höhere Genauigkeiten als Standardverfahren (z.B. MLK und SAM)
- aber: auch Verfahren wie SVM sind vom Hugh-Phänomen betroffen (auch bei Regressionsproblemen)
- mögliche Ansätze: Ensemble-basierte SVM Verfahren und Datenreduktion

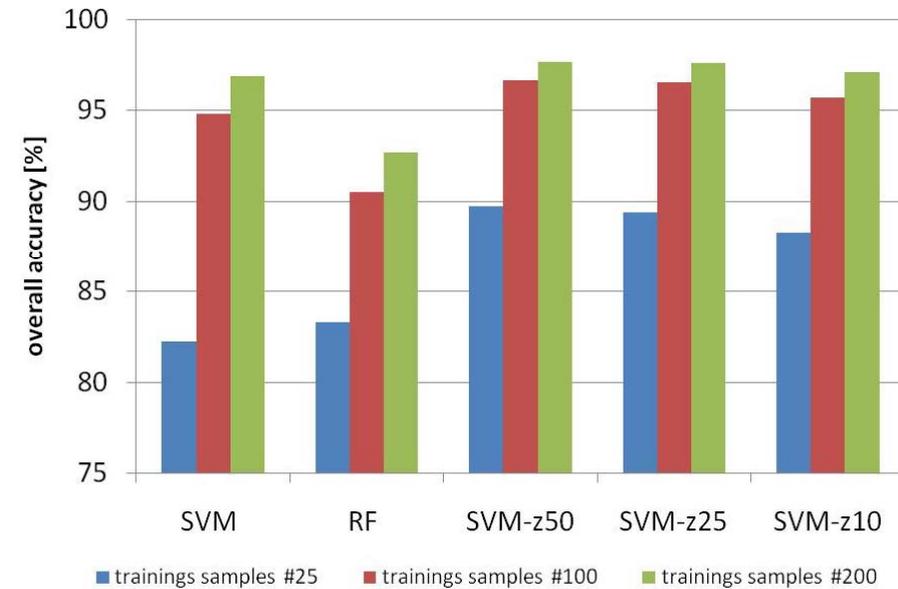
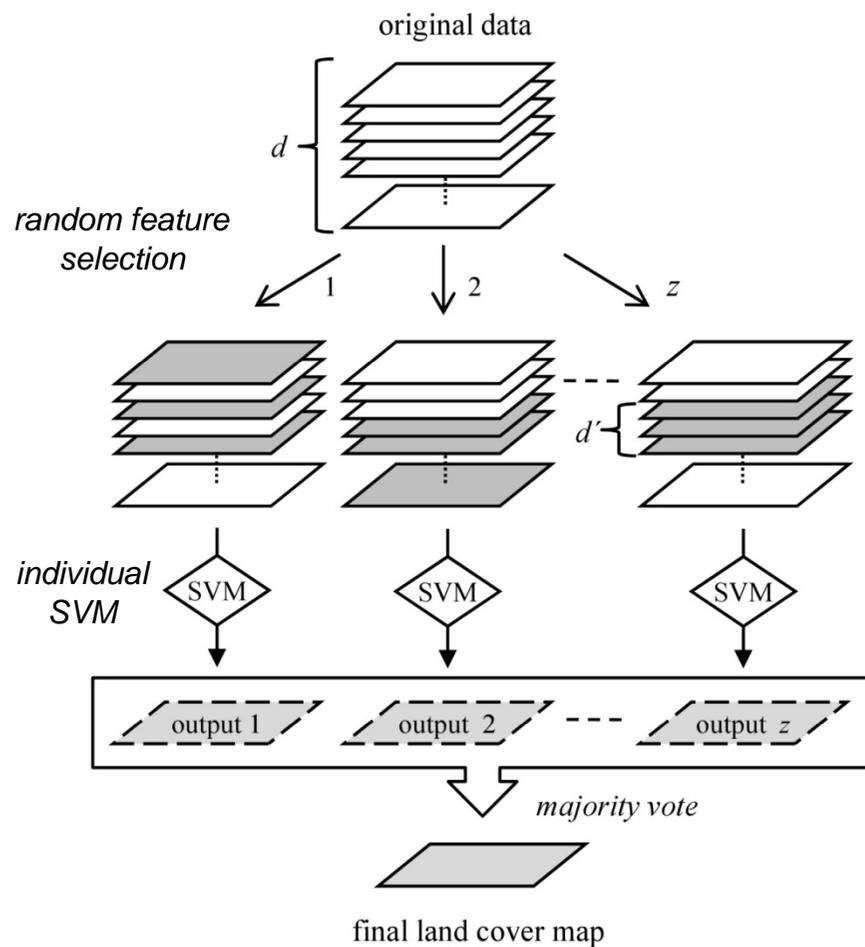
20



(nach, Waske et al. 2009)



Klassifikation von hyperspektral Daten mittels SVM-Ensembles



Klassifikation von multisensoralen und hyperspektralen Fernerkundungsdaten

- diese Datensätze können Ergebnisse verbessern, bedingen aber adäquate Verfahren
- aktuelle Verfahren aus dem Bereich des maschinellen Lernens (z.B. SVM und Random Forests) sind klassischen Methoden häufig überlegen
- aber auch die Performance von Verfahren, wie z.B. SVM, können durch Dimension von hyperspektral Daten beeinflusst werden
- Ansätze können weiter verbessert werden: z.B., *SVM-Ensembles* und *Decision Fusion*
- diese Konzepte lassen sich (theoretisch) auf Regressionsprobleme (z.B. zur Ableitung der Biomasse) übertragen

Wir sind ausschließlich an einer Kartierung von Energiepflanzen interessiert (Mais, d.h. eine Klasse):

- Anforderungen von konventionellen Klassifikationsmethoden an unsere Trainingsdaten:
 - Vollständigkeit (Trainingsdaten für **alle** Klassen)
 - Ausreichend hohe Anzahl von Samples pro Klasse
 - Repräsentativität
- Kosten, Zeit und Expertenwissen
- insbesondere für hyperspektrale Daten:
 - hoch-dimensionaler Merkmalsraum (Hugh-Phänomen), bedingt hoher Anzahl an Trainingsdaten
 - hohe Sensitivität für intraklassen Variationen

Anforderungen an die Trainingsdaten reduzieren?

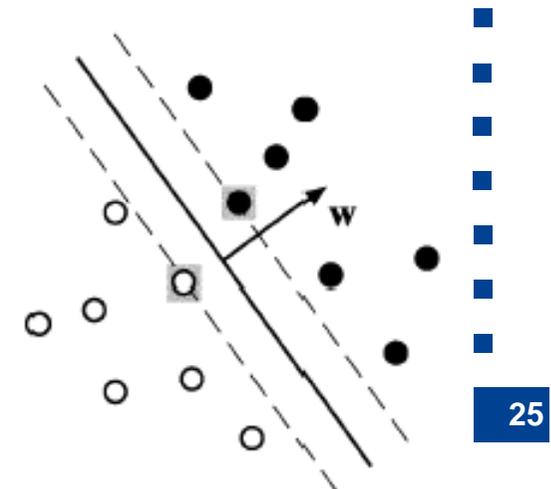
- Binäre Klassifikatoren (BK):
trennen die relevante Klasse (hier: Mais) vom „Rest“
- 1-Klassen-Klassifikatoren (1KK):
klassifizieren ausschließlich die relevante Klasse, die restlichen Klassen werden als Ausreißer behandelt

Vorteile im Vergleich zu konventionellen Ansätzen:

- Zeit- und Kostenersparnisse bei der Erfassung von Trainingsdaten
- Klassifikator wird nur an die *relevante* Klasse angepasst; d.h., nur die Genauigkeit der relevanten Klasse wird optimiert und nicht die Genauigkeit *aller* Klassen

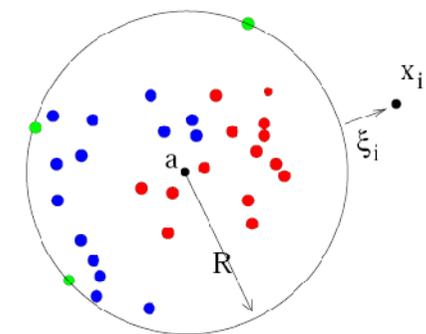
- **Support Vektor-Ansätze:**

- Kernel-basierende Methoden
- robust gegen Overfitting
- Robustheit gegen geringe Anzahl hoch dimensionaler Samples
- BK: Support Vector Machines (SVM)
- 1KK: Support Vector Data Description (SVDD)



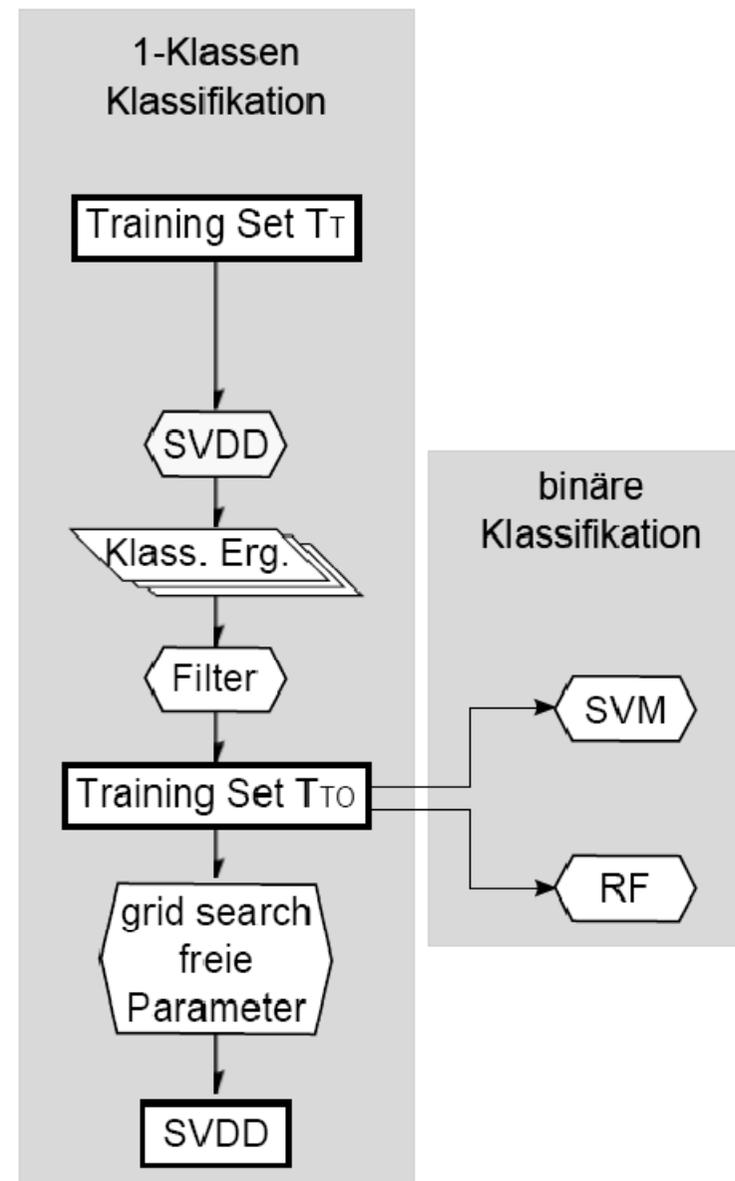
- **Ensemble basierte Verfahren:**

- wenige zu definierende Parameter, daher sehr benutzerfreundlich
- robust gegen Overfitting
- 1KK: z.B. Random Forests (Klasse gegen Rest)



Tax 2001

- Support Vector Machines (und auch SVDD) benötigen Definition von Parametern (C und γ bzw. *rejection rate*)
- Parameter werden in der Regel mittels Trainingsdaten und *grid search* bestimmt
- ist ohne die „Restklasse“ nicht (optimal) durchführbar
- Generierung von *Outliers*, durch Nachbarschaftsinformationen
 - 1KK mit Outlier-Samples
 - binäre Klassifikatoren (SVM, RF)



- Kartierung von Energiepflanzen (Mais) und die Ableitung des BMP, mittels einer Fusion von (simulierten) EnMAP Daten und SAR Daten
 - Laborversuche: signifikanter Zusammenhang zwischen spektral Messungen und BMP
- Nutzung multisensoraler Daten:
 - multisensorale Verfahren (multispektral+SAR) sind mono-sensoralen Ansätzen häufig überlegen
 - „hyperspektral+SAR“? - Potential bisher kaum untersucht
 - EnMAP / Sentinel-1, Radarsat-2, TerraSAR-2

- hyperspektrale und multisensorale Datensätze bedingen komplexere Verfahren:
 - Methoden wie z.B. SVM und Random Forests erscheinen geeignet und können weiter optimiert werden (z.B. *decision fusion* und SVM-Ensembles)
 - diese Verfahren könne auch für Regressionsproblem genutzt werden
 - One-Class Classifier erscheinen interessant, deren Potential muss aber genauer untersucht werden
- Nutzung weitere Sensoren, z.B.:
RapidEye, Sentinel-1, Sentinel-2



Vielen Dank für Ihre Aufmerksamkeit