

Modul 1: Einführung, Problemerkfassung und Daten

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)
www.ada-oeffentliche-verwaltung.de



BERD
@NFDI



Bayerisches Staatsministerium
für Digitales



Über uns



Prof. Dr. Frauke Kreuter



Dr. Malte Schierholz



Dr. Marcel Neunhoeffer



Dr. Heidi Seibold



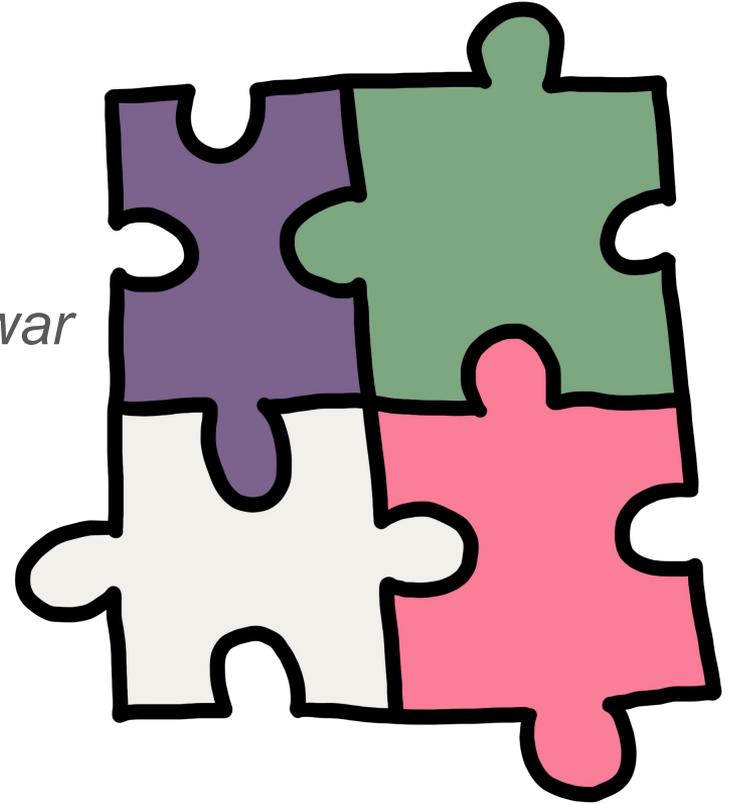
Felix Henninger



Leonhard Kestel

Über euch/Sie

- Name
- Fachlicher Hintergrund (kurz!)
- Grüner Zettel: *Wenn das passiert, war es ein guter Workshop*
- Roter Zettel: *Das soll hier auf gar keinen Fall passieren*



Übrig gebliebene Zettel



Die drei Workshop-Tage

1. **Tag: Gemeinsame Probleme verstehen**
2. Tag: Best-Practice: Zielvorstellungen entwickeln
3. Tag: Infrastruktur (Fokus: Einzelbaumerkennung)

Einführung und Vorstellung	13:00 - 13:30
Problemstellung klären	13:30 - 13:50
Pause	13:50 - 14:05
Fortsetzung: Problemstellung klären	14:05 - 14:55
Pause	14:55 - 15:10
Gruppenarbeit: Datenpipelines und Arbeitsabläufe	15:10 - 16:20
Wrap-Up	16:20 - 16:30

1. Was ist das Ziel?
2. Wo liegen die Daten?
3. Was darf mit den Daten gemacht werden (derzeit)?
4. Warum kann das, was man gerne machen würde, nicht gemacht werden?
5. Was wäre der Wunsch für die Zukunft?

Problemstellung klären

Stadt Bamberg

Was ist das Ziel?

❖ **BaKIM Forschungsprojekt**

- Smart City Bamberg, Lehrstuhl für Kognitive Systeme, Forstamt Bamberg
- Open-Source Tool-Box für Förster*innen
- Fokus auf hochauflösenden Drohnendaten
- Nutzung von statistischen, ML und DL Methoden

❖ **Ziel**

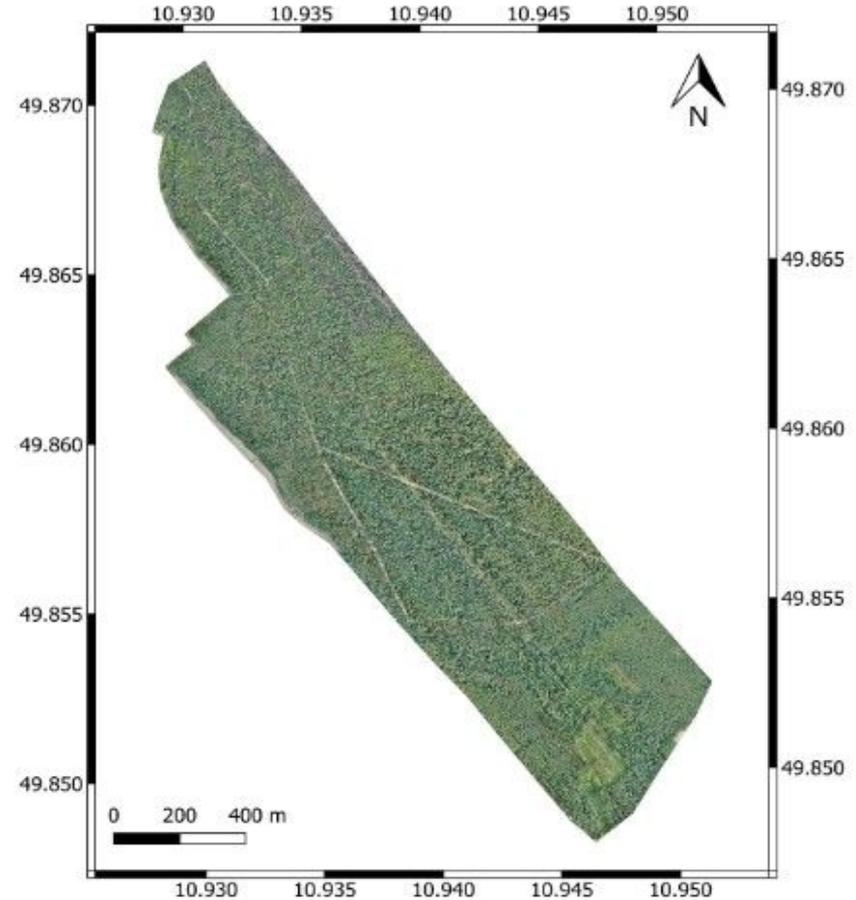
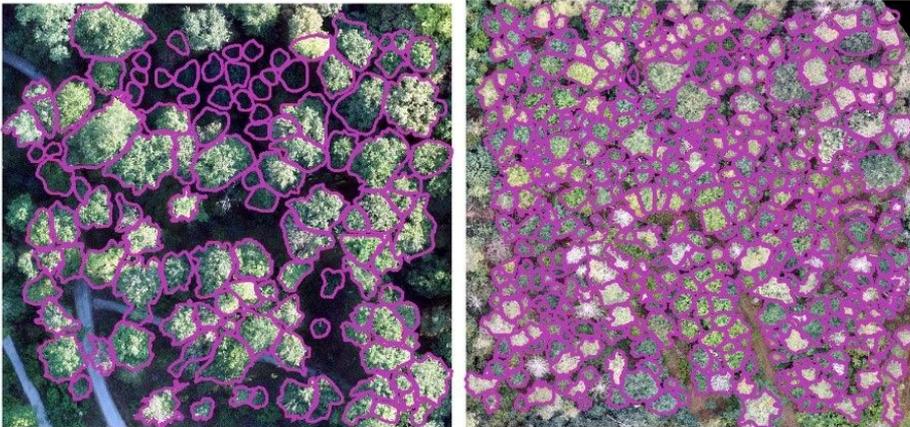
- Bessere Informationslage über Baumbestände schaffen
- Analyse der Bestandsentwicklung
- Zielgerichteterer Einsatz von Ressourcen (Geld und Arbeitskraft)
- Erhaltung und Umstrukturierung der Baumbestände

❖ **Aufgaben**

- Einzelbaumerkennung
- Baumhöhenableitung
- Baumartenerkennung
- Vitalitätsschätzung
- Mistelerkennung

Wo liegen die Daten?

- ❖ **Datenzugang**
 - Öffentliche Daten
 - Gesichert auf dem Lehrstuhlserver
 - [BAMFORESTS](#)
 - <https://doi.org/10.3390/rs16111935>
- ❖ **Datenarten**
 - RGB-Daten (GSD: ~1.5cm)
 - (Höhenprofil (DSM) (GSD: ~1.5cm))
 - (Multispektraldaten (GSD: ~5,3cm))



Was darf mit den Daten gemacht werden (derzeit)?

Regelung nach Open-Source-Lizenzvereinbarung

Warum kann das, was man gerne machen würde, nicht gemacht werden?

❖ Aufgaben

- Einzelbaumerkennung
- *Baumhöhenableitung*
- Baumartenerkennung
- Vitalitätsschätzung
- *Mistelerkennung*

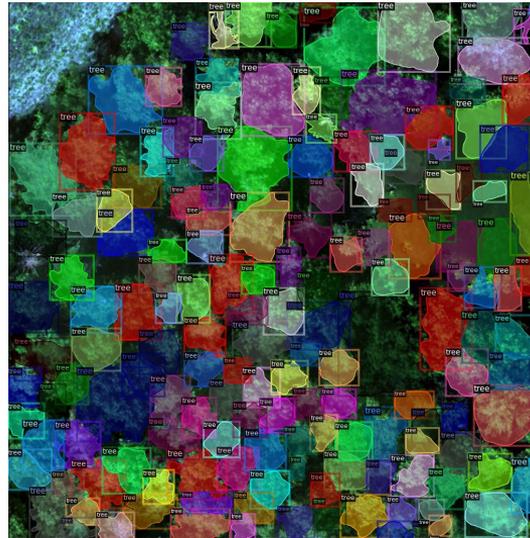
❖ Probleme

- Multispektraldatenqualität
- Unterscheidung von gesunden und kranken Bäumen
- Trennung von dichten Laubbäumen (~70% akkurat)

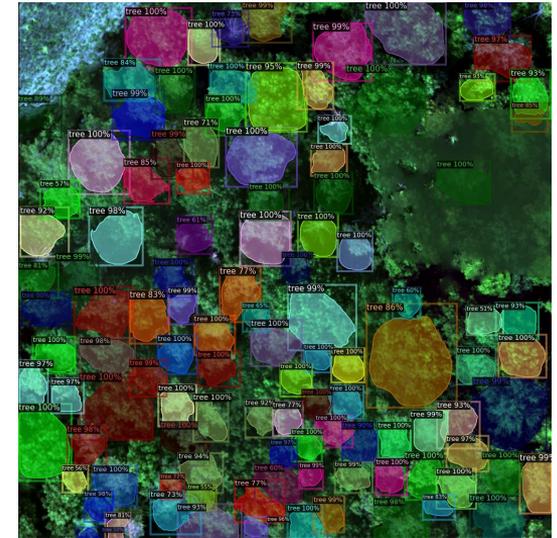


Was wäre der Wunsch für die Zukunft?

- ❖ Bessere Performanz der Ansätze
- ❖ Schnellere/vereinfachte Datenerhebung
- ❖ Mehr öffentlich verfügbare Trainingsdatensätze
- ❖ Robustere Modelle
- ❖ Effiziente gute Pipelines und SaaS



Annotation

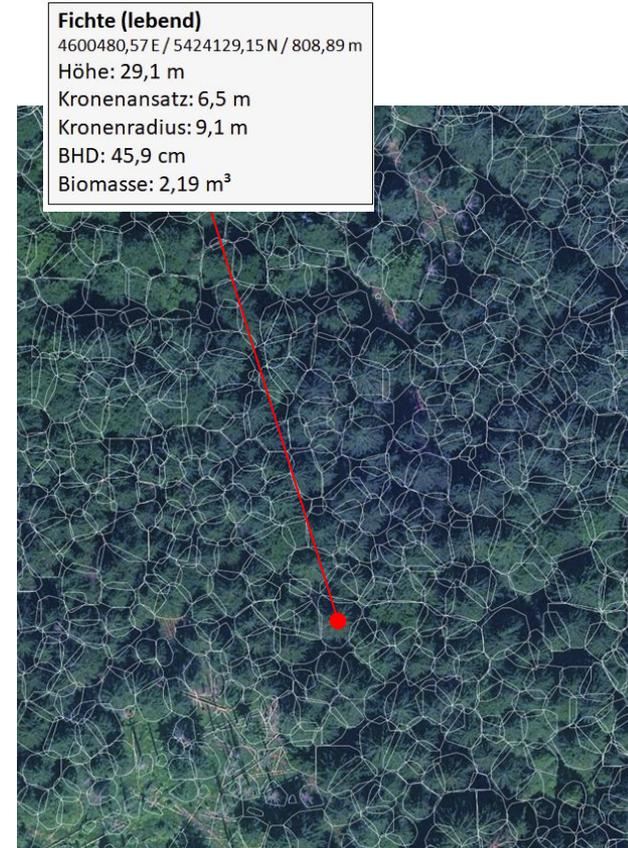


Vorhersage

Nationalpark Bayerischer Wald

Was ist das Ziel?

- ❖ **Flächendeckende, wiederkehrende Waldinventur:**
 - Kartierung aller Bäume inkl. Art, Zustand, Höhe, Kronenparameter
 - Verschneidung mit Bodendaten aus Dauerbeobachtungsflächen und Inventuren
 - Ableitung flächendeckender Parameter zum Waldzustand (Baumartenverteilung, Totholzkartierung, Biomasse, Waldzustand)
 - Evaluierung der langzeitlichen Waldentwicklung im Nationalpark
- ❖ **Bereitstellung von Basisdaten für weitere Aufgabenbereiche des NP:**
 - Waldmanagement (Borkenkäferbekämpfung)
 - Habitatkartierung für verschiedene Tier/Pflanzenarten
 - Evaluierung von Naturschutzpraktiken



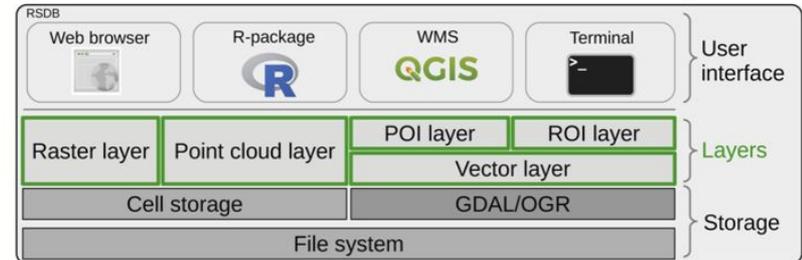
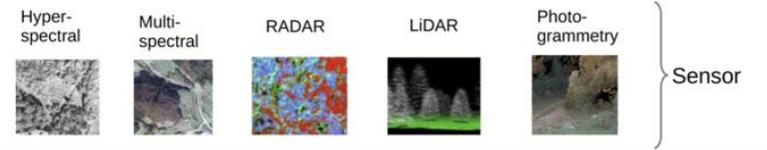
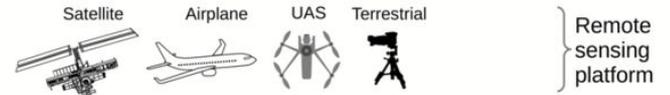
Wo liegen die Daten?

- ❖ **Alle Daten liegen File-basiert auf Servern des RZ Süd**

- Problem: Zugriffsgeschwindigkeit
- Lokale Kopien auf Festplatten zum arbeiten

- ❖ **LiDAR-Punktwolken zusätzlich in spezieller Fernerkundungsdatenbank abgespeichert**

<input type="checkbox"/> 32804_5445.laz	02.11.2023 21:16	LAZ-Datei	14.144 KB
<input type="checkbox"/> 32804_5446.laz	02.11.2023 21:16	LAZ-Datei	145.096 KB
<input type="checkbox"/> 32804_5447.laz	02.11.2023 21:17	LAZ-Datei	468.574 KB
<input type="checkbox"/> 32804_5448.laz	02.11.2023 21:17	LAZ-Datei	499.762 KB
<input type="checkbox"/> 32804_5449.laz	02.11.2023 21:17	LAZ-Datei	41.459 KB
<input type="checkbox"/> 32805_5444.laz	02.11.2023 21:17	LAZ-Datei	187.374 KB
<input type="checkbox"/> 32805_5445.laz	02.11.2023 21:18	LAZ-Datei	882.507 KB
<input type="checkbox"/> 32805_5446.laz	02.11.2023 21:19	LAZ-Datei	1.035.968 KB
<input type="checkbox"/> 32805_5447.laz	02.11.2023 21:20	LAZ-Datei	1.152.871 KB
<input type="checkbox"/> 32805_5448.laz	02.11.2023 21:21	LAZ-Datei	1.061.153 KB
<input type="checkbox"/> 32805_5449.laz	02.11.2023 21:22	LAZ-Datei	744.042 KB
<input type="checkbox"/> 32805_5450.laz	02.11.2023 21:22	LAZ-Datei	53.447 KB
<input type="checkbox"/> 32806_5443.laz	02.11.2023 21:22	LAZ-Datei	157.585 KB
<input type="checkbox"/> 32806_5444.laz	02.11.2023 21:23	LAZ-Datei	832.962 KB
<input type="checkbox"/> 32806_5445.laz	02.11.2023 21:24	LAZ-Datei	1.168.811 KB
<input type="checkbox"/> 32806_5446.laz	02.11.2023 21:25	LAZ-Datei	1.109.322 KB
<input type="checkbox"/> 32806_5447.laz	02.11.2023 21:26	LAZ-Datei	1.128.275 KB
<input type="checkbox"/> 32806_5448.laz	02.11.2023 21:27	LAZ-Datei	1.371.602 KB
<input type="checkbox"/> 32806_5449.laz	02.11.2023 21:28	LAZ-Datei	755.957 KB
<input type="checkbox"/> 32806_5450.laz	02.11.2023 21:28	LAZ-Datei	577.504 KB
<input type="checkbox"/> 32806_5451.laz	02.11.2023 21:28	LAZ-Datei	15.195 KB
<input type="checkbox"/> 32807_5442.laz	02.11.2023 21:28	LAZ-Datei	134.753 KB
<input type="checkbox"/> 32807_5443.laz	02.11.2023 21:28	LAZ-Datei	726.475 KB



Was darf mit den Daten gemacht werden (derzeit)?

- ❖ Interne Verwendung frei
- ❖ Externe Verwendung mit Nutzungsvereinbarung möglich

Warum kann das, was man gerne machen würde, nicht gemacht werden?

❖ **Einschränkungen bei IT-Infrastruktur**

- Rechenkapazitäten
- Speicherkapazitäten und Datenzugriff (kein zentraler Speicherort für alle Daten mit Updates)
- Aktuell kein nutzbares Tool zur Einzelbaumerkennung

❖ **Einschränkungen bei der Algorithmik**

- Trennung von Einzelbäumen in dichten Laubbaumbeständen
- Beschränkung auf die in den Bilddaten sichtbaren Bäume (keine Bäume im Unterstand)
- Baumartenerkennung: aktuell nur Differenzierung von Laub- und Nadelbäumen

❖ **Einschränkungen bei den Daten**

- Qualität der True-Orthophotos nicht gewährleistet
- Homogenisierung von Datensätzen über verschiedene Zeitschritte

Was wäre der Wunsch für die Zukunft?

- ❖ **Entwicklung einer Pipeline zur Einzelbaumdetektion im Nationalpark**
 - Entwicklung von zuverlässigen Algorithmen
 - Entwicklung eines anwendbaren Tools
- ❖ **Einrichtung von zentralen Rechen- und Speicherkapazitäten in der Cloud**
 - Effektiver, Schneller Zugriff (evtl. Auch von Anwendern im Feld)
 - Datendokumentation
 - Rohdaten (True-Orthophotos, Laserscanning Punktwolken)
 - Manuell gelabelte Daten und Ergebnisse der Einzelbaumerkennung

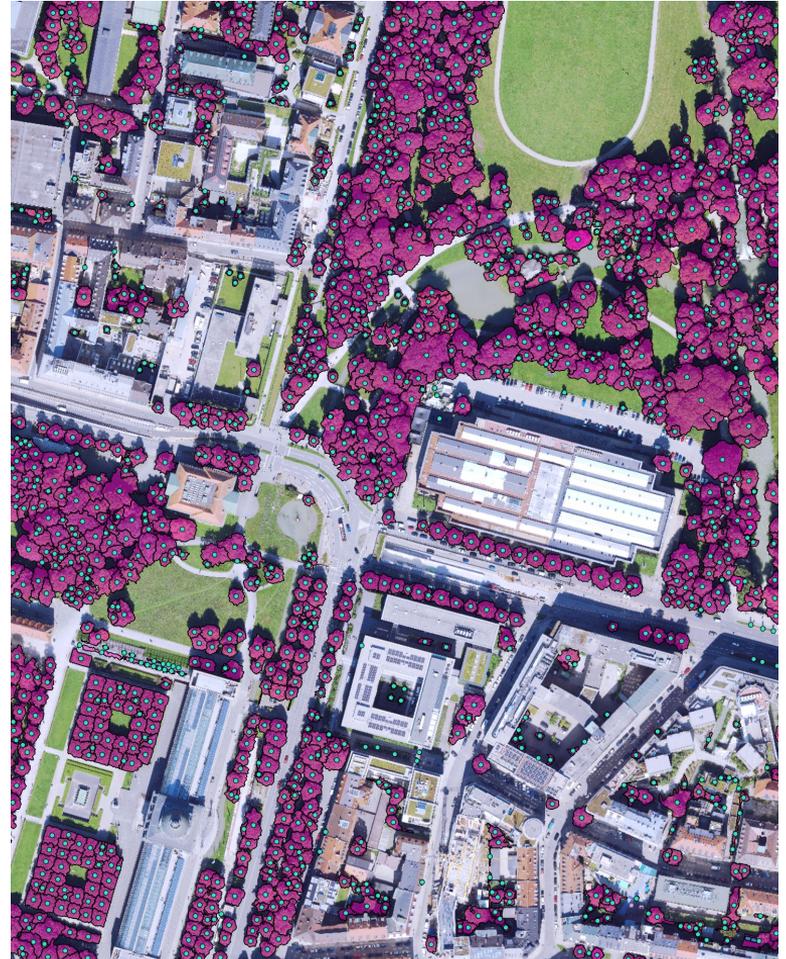
Pause: 13:50-14:05



Stadt München

Was ist das Ziel?

- ❖ **Vollständiges “Fernerkundungs-Baumkataster” für München**
 - Stand heute
 - Erstmals auf Luftbilddaten 2021 (RGBI und LiDAR) Erfassung aller Einzelbäume:
 - Punkte
 - Kronen (kreisförmig)
 - Umrisse
 - **Ziel:** Fortführung alle zwei Jahre
 - Hybrider Ansatz mithilfe von Bodenperspektive aus Mobile-Mapping-Daten (zusätzliche Informationen wie Stammfußpunkt oder Stammdurchmesser)



Wo liegen die Daten?

- ❖ Sicherung auf unseren Servern
- ❖ Bereitstellung intern über Geodatenpool der Stadt München (in Arbeit)
- ❖ Visualisierung über Portal for ArcGIS

Was darf mit den Daten gemacht werden (derzeit)?

- ❖ interne Nutzung frei
- ❖ externe Verwendung nur mit Nutzungsvereinbarung

Warum kann das, was man gerne machen würde, nicht gemacht werden?

- ❖ Deep Learning
- ❖ Baumarten (evtl. Projekt auf Schrägbildern und KI)
- ❖ Vitalität der Bäume
- ❖ Verbesserung der Ergebnisse

Probleme:

- ❖ keine Multispektraldaten
- ❖ städtische IT-Infrastruktur
- ❖ Qualität in Waldstrukturen

Was wäre der Wunsch für die Zukunft?

- ❖ gemeinsame Pipeline zur Einzelbaumdetektion im urbanen Raum
- ❖ agilere Zusammenarbeit mit städtischer IT bzgl. Software und Programmiermöglichkeiten
- ❖ mehr öffentlich zugängliche Trainingsdaten

LDBV

Was ist das Ziel?

- Baumdetektion für die Visualisierung in eigenen Kartenprodukten in 2D und 3D
- momentan existieren dazu 2 Ansätze:
 - durch Maskierung (produktiv)
 - durch KI (Testphase)



Wo liegen die Daten?

- auf eigenen Servern

Was darf mit den Daten gemacht werden (derzeit)?

- Daten vorerst nur für interne Nutzung vorgesehen

Warum kann das, was man gerne machen würde, nicht gemacht werden?

- Maskierungsansatz:
 - grobe Fehler in Bereichen, in denen die Masken nicht genau genug sind (Stromleitungen, Brücken, Gebäude)
 - Probleme bei Laubbäumen mit großen Kronen
 - als Ergebnis erhält man nur Baumhöhen und -standorte, keine Klassen (Laub/Nadel) und keinen Kronendurchmesser
- KI-Ansatz:
 - Probleme bei dichten Laubwäldern, einzelne Bäume zu unterscheiden
 - keine oder nur wenige Ground Truth Daten vorhanden
 - schlechte Übertragbarkeit bei großen radiometrischen Unterschieden der Eingangsdaten

Was wäre der Wunsch für die Zukunft?

- robusterer KI - Ansatz hinsichtlich Radiometrie

Sonstige

Was ist das Ziel?

Wo liegen die Daten?

Was darf mit den Daten gemacht werden (derzeit)?

Warum kann das, was man gerne machen würde, nicht gemacht werden?

Was wäre der Wunsch für die Zukunft?

Pause: 14:55-15:10



Datenpipelines und Arbeitsabläufe

Teamarbeit 1

1. Vorstellungsrunde: Was sind die Expertisen der Team-Mitglieder?
 - a. Wer in der Gruppe hat bereits Hands-on mit Fernerkundungs-daten gearbeitet?
2. Das Team wählt ein Projekt aus und bespricht das Projekt im Detail, was dort durchgeführt wurde
 - a. Welche Daten (genau) sind der Input?
 - b. Welche Verarbeitungsschritte werden durchgeführt?
 - c. Was ist der Output (genau)?
3. Die Ergebnisse sollen zur weiteren Nutzung (graphisch) festgehalten werden

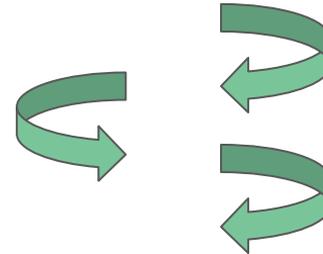
Datenarbeit = Teamarbeit

Expertenwissen einbinden!

Beispiel

Input:

1. Laserscanner (Lidar)
2. Digitales Oberflächenmodell



Output:

Position von Baumkronen
Erkannte Bäume (Position, Art, Höhe, ...)

Die Teams

Gruppe 1
Singer

Felix

Isabell R.

Marco H.

Johannes N.

Gruppe 2
Nightingale

Marcel

Jakob E.

Thomas K.

Peter K.

Gruppe 3
R. 456

Heidi

Michael D.

Sebastian D.

Manuela E.

Theresa M.

Gruppe 4
R. 454

Leo

Fabian F.

Johannes W.

Lukas L.

Die Teams

Gruppe 1
Singer

Felix

Fabian F.

Johannes N.

Leo

Gruppe 2
Nightingale

Marcel

Jakob E.

Thomas K.

Theresa M.

Gruppe 3
R. 456

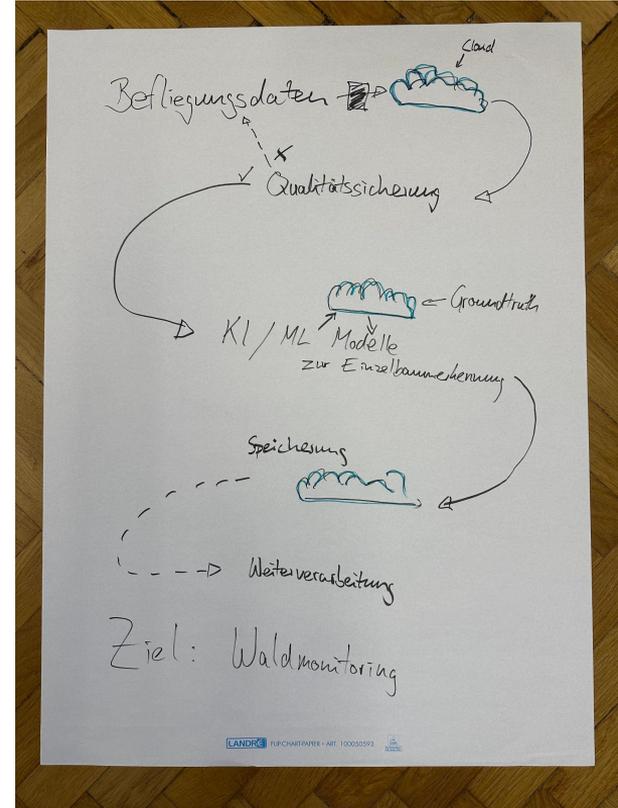
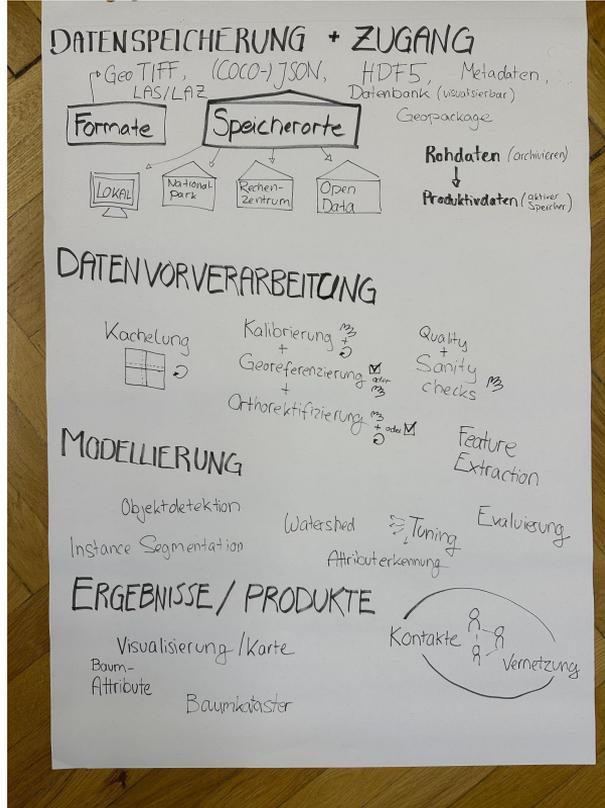
Heidi

Michael D.

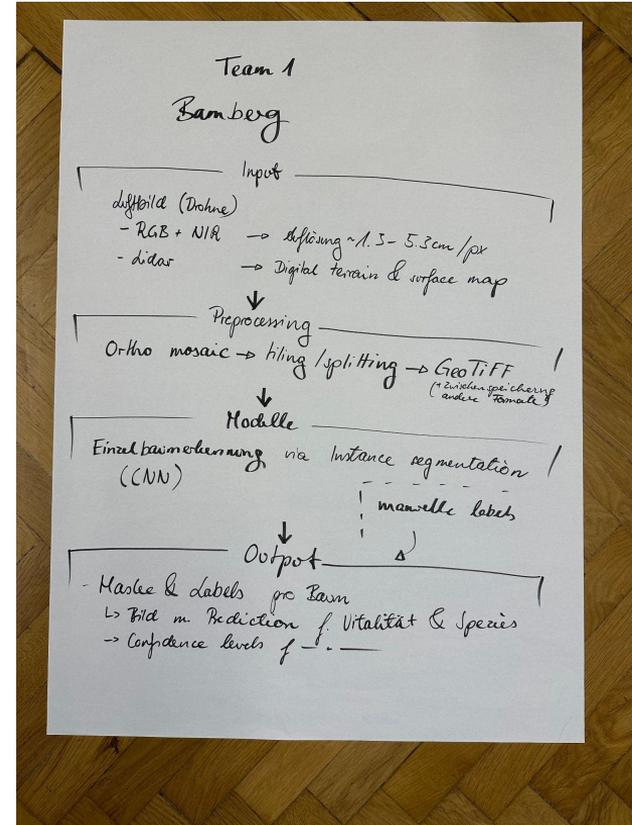
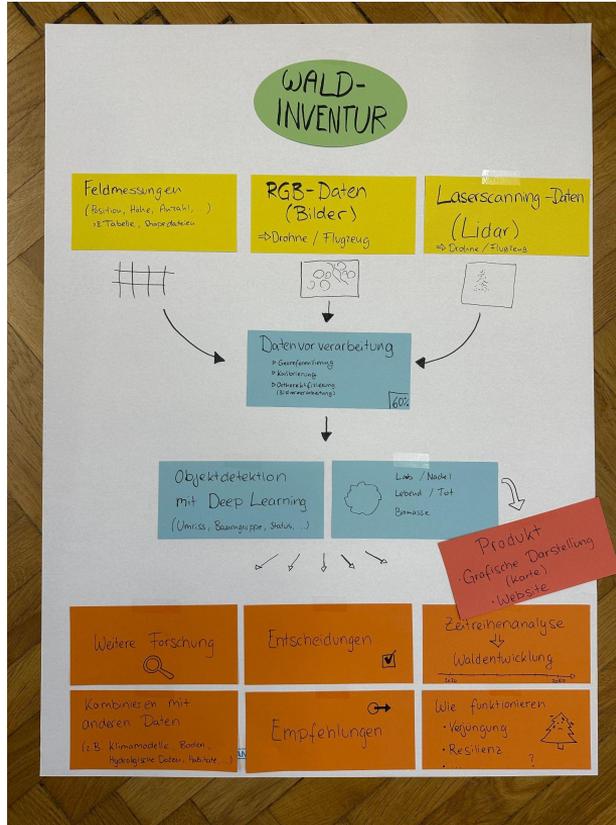
Sebastian D.

Manuela E.

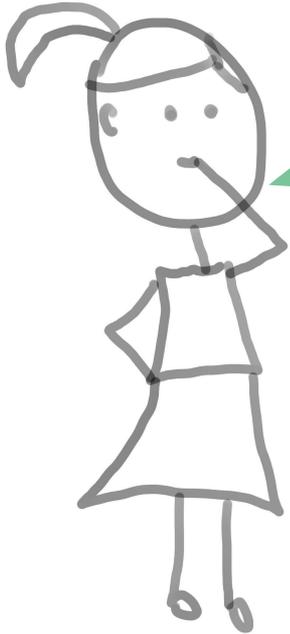
Vorstellung der Ergebnisse



Vorstellung der Ergebnisse

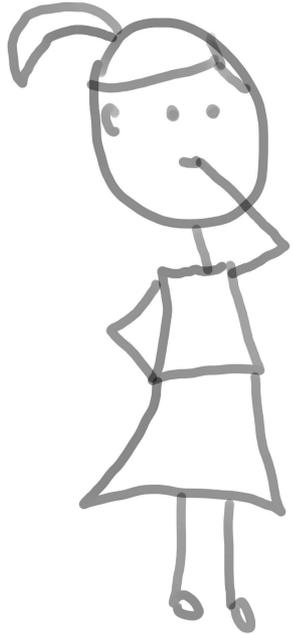


Gemeinsames Ziel festlegen



Wie können wir **gemeinsam...**

- *effektiver Einzelbaumerkennung durchführen und nutzen,*
- *Infrastruktur nutzen (Cloud?),*
- *Daten, Modelle und Code austauschen / gemeinsam nutzen?*
- *Datentypen überdenken / standardisieren, harmonisieren in die Vergangenheit, Robustheit gegenüber z.B. Auflösung*
- *Weitere Partner identifizieren und einbeziehen -> Skaleneffekte (Bayern-weit?)*



Was nehmen Sie von heute mit?

Welche Fragen haben Sie?

Ausblick: Morgen

Übersicht: Was ist die Cloud?	09:35 - 10:00
ADA Bayern und die Cloud	10:00 - 10:30
Pause	10:30 - 10:45
Recap: Arbeitsabläufe	10:45 - 10:55
Funktionsweisen von Algorithmen	10:55 - 12:00
Mittagspause	12:00 - 12:50
Herausforderungen an die IT-Infrastruktur	12:50 - 13:25
Brainstorming: Mögliche Anwendungen & Datenprodukte (inkl. Pause)	13:25 - 14:45
Pause	14:45 - 14:55
Wünsche an die Infrastruktur	14:55 - 15:20
Wrap-Up und optional Parkspaziergang	ab 15:20