Development of an open reference data base for tree species classification from the German national forest inventory

Magdon, P.[1] Freudenberg, M.[2], Pietras-Couffignal, K.[3] and Schnell, S.[3]

[1]University of Applied Sciences and Arts (HAWK) Göttingen, Germany
[2]Forest Inventory and Remote Sensing, Georg-August University, Göttingen, Germany
[3]Thünen Institute of Forest Ecosystems, Eberswalde, Germany

Introduction

Information on past and current tree species distribution is a core variable for forest management

- Tree species classification is among the most studied topics in forest remote sensing
- Review by Fassnacht et al. (2016):
  - More large area studies with sound sample/experimental designs are needed
  - More hypothesis driven studies are needed
  - Exchange of reference data should be fostered

Applications within the context of the NFI
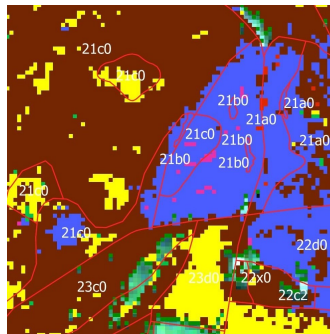
- Species-specific estimates within small-area estimation framework
- Augmentation of forest disturbance statistics
- Risk evaluation of current forest stands
- Facilitate model-assisted estimates of forest area by tree species (with focus on more rare species if possible)
- Provide wall-to-wall maps as a public service: forest owners, scientific community, forest industry, etc.

Introduction

- Earth observation programs provide access to standardized timeseries with spatial resolutions suitable for forest management
- Machine learning algorithms allow for classification of complex non-linear timeseries data

- Forest inventory programs systematically collect information on tree species in the field
- National Forest Inventory (NFI) program provides the largest field dataset on tree species in Germany

Limitation of reference data

Quality, amount and access to reference data is described as one of the major limitations by many authors.

Why are plot coordinates not public?

- Permanent sample plots with a long time series since 1987
- Representative and unbiased results only if the surroundings of the sample plots are treated the same way as the remaining forest
- This cannot be guaranteed if coordinates are made public or sample trees have permanent and visible markers
- Past investments (about 80 million Euro) and validity of the NFI at risk
- Common understanding to keep exact locations secret in forest inventory related literature and most NFIs across the globe
- Option to establish the link between field and remote sensing data by Thünen Institute
- Other solutions are in preparation (reference data base, web service, etc.)

# Objectives & Research Questions

Project Objectives:

- Development of a public national reference data base for tree species classifications based on the German National Forest Inventory (NFI) and Copernicus satellite data
- Circumvent the publication of exact geographical location

Research Questions:

- How to utilize the NFI data to source reference data for remote sensing tree species classification?
- How to create Sentinel-2 timeseries as spatial, temporal and spectral fingerprints of tree species?
- How to structure and publish an open flexible reference data set?

Sample design of the German NFI

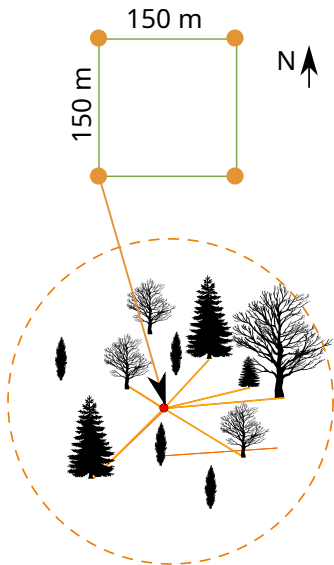Figure 1: NFI sampling grid, 2012
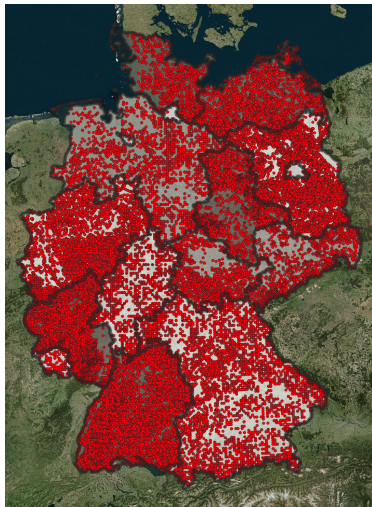
- Probability design with systematic grid of varying density
- Four subplots at each NFI plot
- Regular revisiting interval: 10 years
- Coordinated by the Thünen Institute and implemented by the federal states

Year	Plots	Subplots	Trees
2012	25,382	68,201	522,958
2017	5,768	15,740	119,311
2021	*29,000*	*81,000*	*1 mil.*

- Angle count sampling with $k = 4$
- Average maximum distance from plot centre to tree ~ 13 m
- ~ 9 trees per subplot
- Observed variables:
  - Distance & angle from center
  - Tree species
  - Diameter at breast height (DBH)
  - Tree height (subset)
  - ... and many more

Representative vs. unbalanced

- Abundance of tree species in the sample as they occur in the population of trees
- Results of the NFI are representative for the population: only four species represent 75 % of all individuals
- However, the data set is unbalanced with respect to tree species: Implications for training, model building, etc.
- Training data should be balanced: model-based paradigm
- Data for validation should be representative: design-based, probabilistic
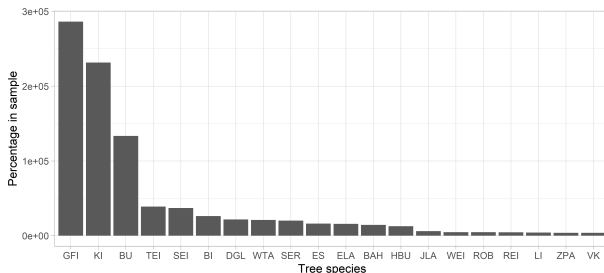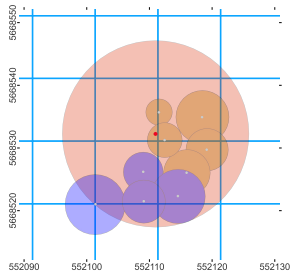
Design considerations for the reference data set

Challenges of the NFI plot design:

- No fixed plot area with angle count sampling
- Precise subplot coordinates must not be released

Option 1: Restrict to single/two species plots:

- Blickensdörfer et al., 2022, under review
- Welle et al., 2022

Approach

Utilize the single tree information to provide the fractional cover and times series for the S2 pixels

Remote Sensing Data

Sentinel-2 Satellite Images:

- National data cube 2017–2021 with all S2A/B images $< 90\%$ cloud cover
- Image processing to L2A timeseries with FORCE v3.6 (Frantz, 2019)
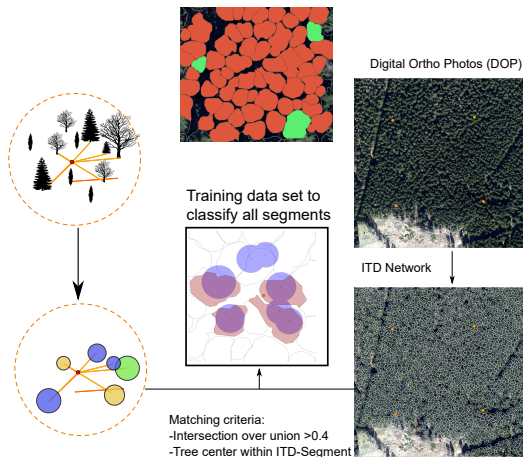- Data provided via CODE-DE platform

Digital Orthophotos (DOP):

- National coverage 2017–2021 of aerial images with 20cm spatial resolution
- RGB+IR images
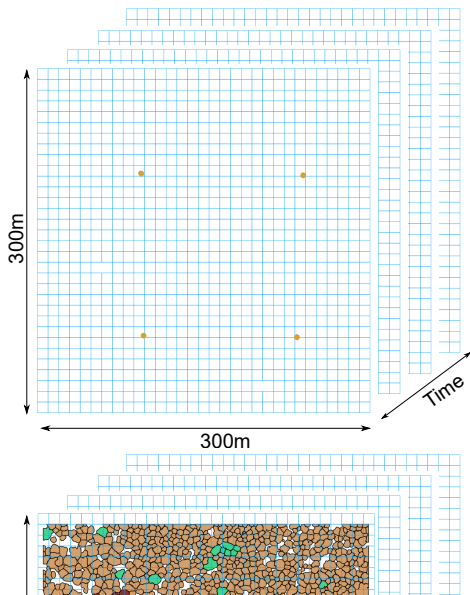- Federal Agency of Cartography and Geodesy (BKG)

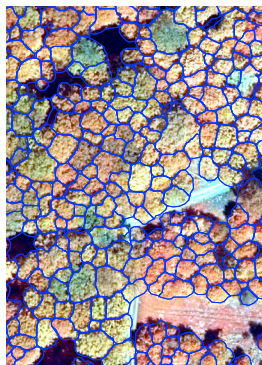# Compilation of the reference data: Step 1

Workflow:

1. Single tree positions
2. Modelled crown projection area
3. U-Net based Individual Tree Detection (ITD) (Freudenberg et al., 2022)
4. Matching of crown areas and ITD segments
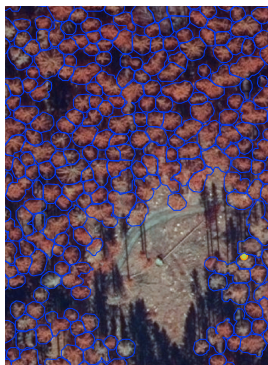5. Species classification of ITD segments (Nölke et al., 2020)

Digital Ortho Photos (DOP)

Training data set to classify all segments

ITD Network

Matching criteria:
-Intersection over union >0.4
-Tree center within ITD-Segment
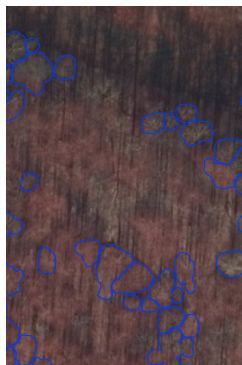
First Results: Tree crown segmentation

PlotID: 18608   PlotID: 46804   PlotID: 19362

- Segmentation done for 300x300m buffer of 27200 NFI plots
- Accuracy 65%; Precision 77%; Recall 81%
- Only 50.6% of the NFI-plots from 2017 are covered by leaf-on DOPs (DOY > 130)

- For 11.5% of the field measured trees an ITD segment could be assigned based on the matching criteria (Position and IOU)

- Implementing enhanced pattern matching methods is planned
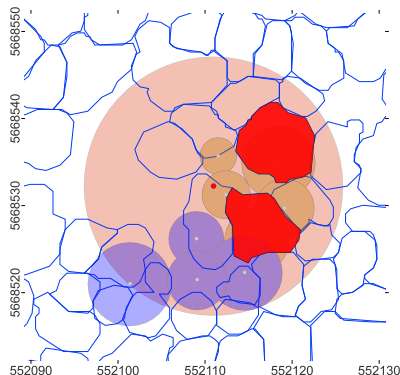
First Results: Extraction of the time series

- The data base will provide the fractional cover of the different species for each S2 pixel in the 300x300m support area
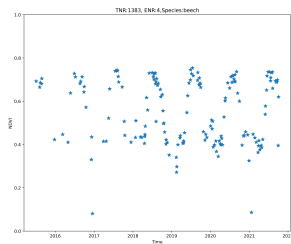- On average 152 observations per pixel are extracted

Figure 2: Sentinel L2A NDVI Time series of a single tree

Data Access

The reference data will be provided via a public web service which allows customized filtering and visualization in the temporal, spatial and spectral domain.

Discussion & Outlook

- The German NFI was not designed to collect reference data
- Applying the single tree approach allows for compiling a quite universal reference data set
- Linking field measured trees to ITD segments remains challenging and advanced pattern matching algorithms will be tested
- Measuring the exact plot positions with differential GNSS receivers should become a standard in forest inventories!

We are looking forward to the NFI data from 2022, which will enlarge the data base drastically

Acknowledgments

Thank you for your attention!

treespecies.de

Paul.Magdon@hawk.de

 @MagdonPaul

References

References:

Blickensdörfer, L., Katja, O., Langer, N., Pflugmacher, D., Kleinschmit, B., and Hoster, P. (2022). Large-area tree species mapping across environmental gradients: Combining sentinel-2 and sentinel-1 time series with german national forest inventory data. under review.

Fassnacht, F. E., Latifi, H., Stereńczak, K., Modzelewska, A., Lefsky, M., Waser, L. T., Straub, C., and Ghosh, A. (2016). Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*, 186:64–87.

Frantz, D. (2019). FORCE—Landsat + Sentinel-2 Analysis Ready Data and Beyond. *Remote Sensing*, 11(9):1124.

Freudenberg, M., Magdon, P., and Nölke, N. (2022). Individual tree crown delineation in high-resolution remote sensing images based on u-net. *Neural Computing and Applications*.

Nölke, N., Freudenberg, M., Kleinn, C., Fuchs, H., and Magdon, P. (2020). Baumartenerkennung mithilfe von künstlicher Intelligenz. *AFZ - Der Wald*, 15.

Welle, T., Aschenbrenner, L., Kuonath, K., Kirmaier, S., and Franke, J. (2022). Mapping dominant tree species of german forests. *Remote Sensing*, 14(14):3330.

Example GNSS Coordinates

TNR 19764

◆ Field measured trees (2017)
● Plot centre (GNSS 2021)

DOP (21.04.2019)
- Band 1 (Red)
- Band 2 (Green)
- Band 3 (Blue)

0   2,5   5   7,5   10 m

Pure Plots

Table 1: Subplots with $\geq 90\%$ basal area of one species

Species	Subplots
Spruce	10695
Pine	12693
Beech	5066
Oak	2154