Integrating Context-based Recommendations with Deep NN Image Classification for Plant Identification Tasks

Hans Christian Wittich, David Boho, Patrick Mäder

ICEI 2018

25.09.2018

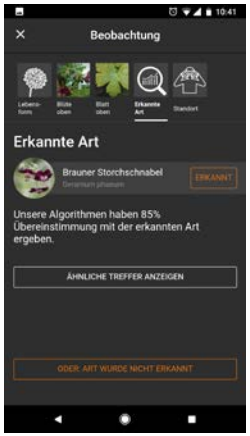
TECHNISCHE UNIVERSITÄT ILMENAU

Software Engineering for Safety-Critical Systems Group
Technische Universität Ilmenau

Plant Identification

- Flora Incognita Project
- Free app for iOS/Android devices
- Identifying species from their photos interactively

- Machine Learning: Deep CNN (NASNet)
- 2770 Classes (wild flowering plants in Germany)



Contextual Recommender

- Identification of large number of plant taxa is hard
- Recommender: Shortlist of likely candidate taxa (“mobile field guide“)

- Context: Information on circumstances an observation is made under
- Easily available metadata on mobile devices:
  - Geographic position
  - Current date and time

- Benefit for users:
  - Validity check when plants found in atypical locations

Factors Influencing Plant Observability

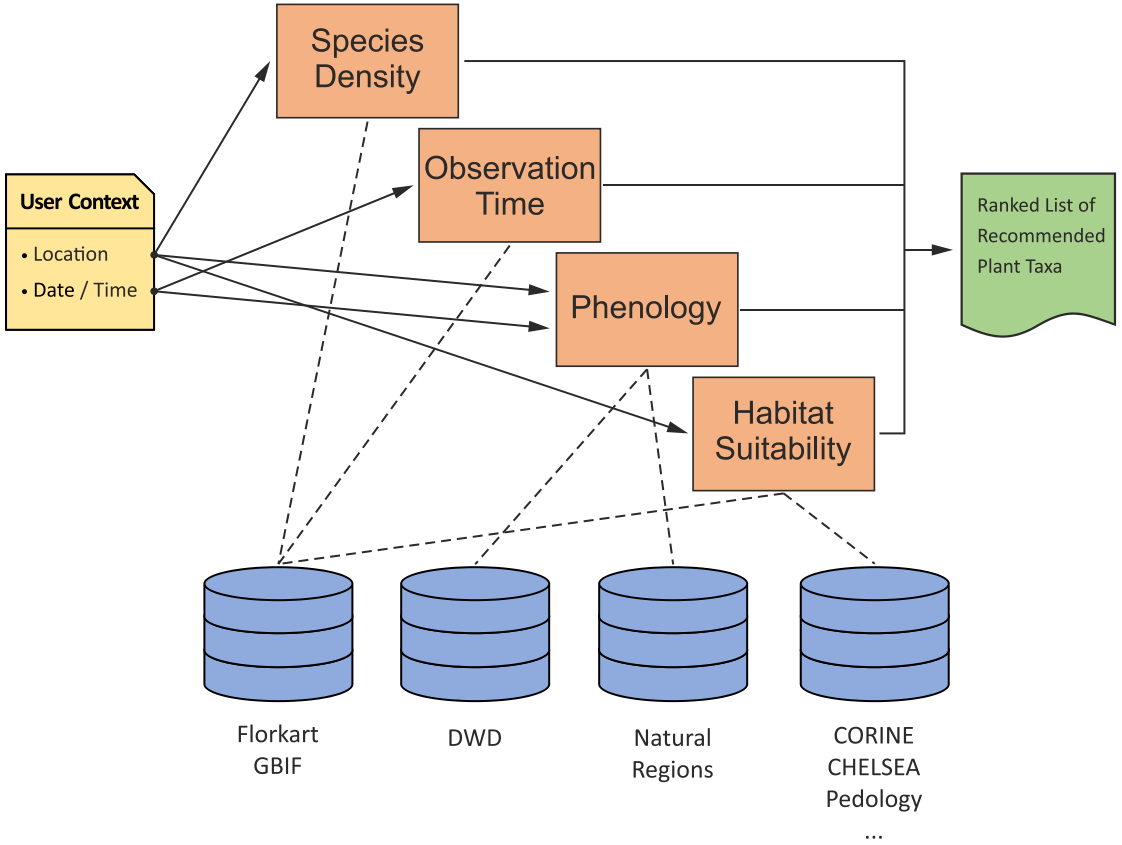
1. a) Geographic range of plant
   - Plant known to occur somewhere in area (large scale)

   b) Occurrence of plant individual
   - Likely to occur near location of known observation
   - Likely to be observed around same time of year

2. Phenology
   - Known plant-specific flowering periods
   - Most plants best identified when flowering

3. Presence of suitable habitat/environment
   - Plant-specific ecological conditions present -> plant present (potentially)

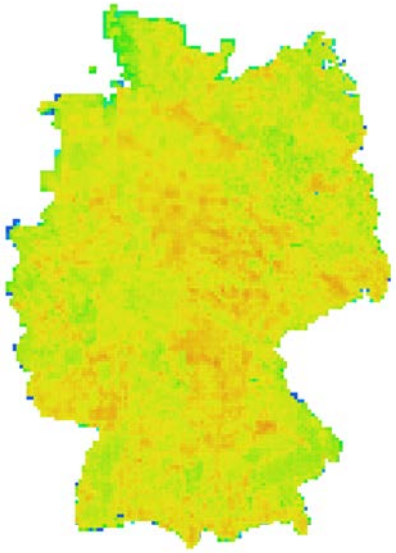
Recommender Components



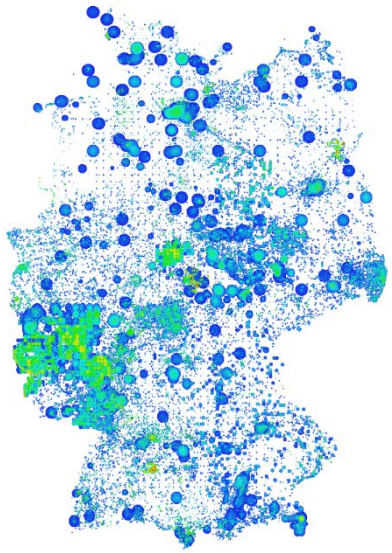
Biogeography: Expert Knowledge

- Plant distribution grid maps
  - Large scale, low resolution
  - Binary presence/absence
  - Comprehensive (all of Germany)
- Individually observed occurrences
  - Local scale, high accuracy
  - Observation times
  - Presence only
  - Irregularly sampled

BfN FloraWeb
Bundesamt für Naturschutz

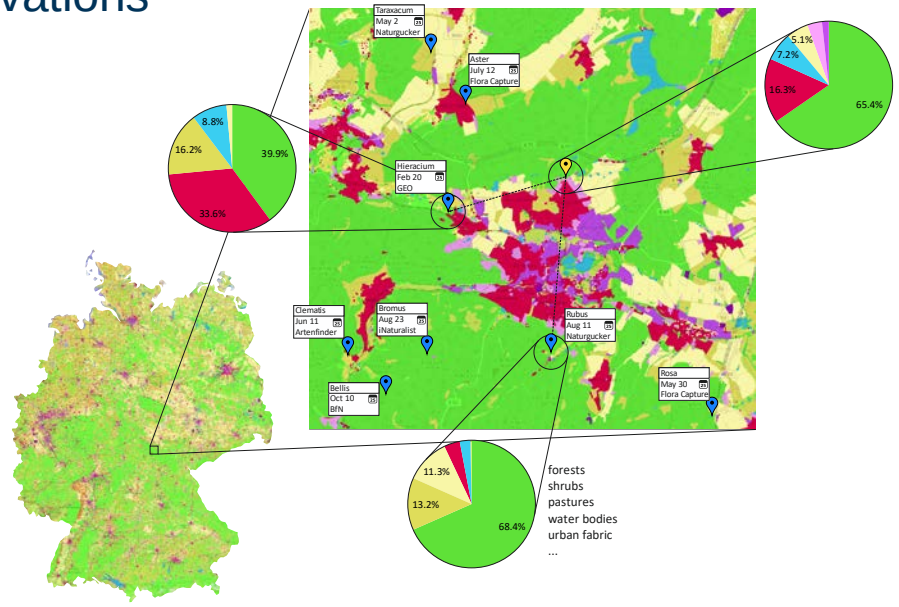


GBIF
Global Biodiversity Information Facility



Biogeography: Contribution

- Similarity to known plant observations
  - Geographical distance
  - Land cover similarity
  - Time difference

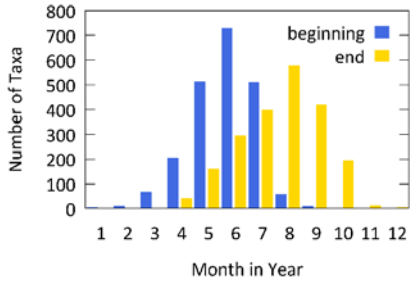


	Average Recall	Top-20 Recall	Median Rank	Avg List Length
Species Density	95%	13%	156	1551
Observation Time	50%	20%	37	240

Phenology: Expert Knowledge

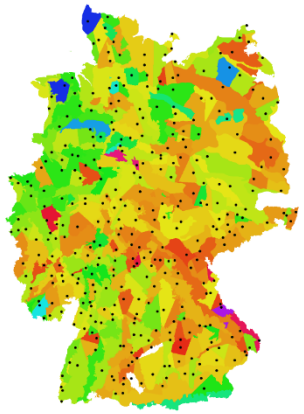
Static flowering periods

- Known for most taxa
- Coarse-grained (per month)
- Independent of location



Dynamic flowering periods

- Groups of taxa sharing similar phenological characteristics
- Ten phenological seasons, related to onset of specific phenophase
- Location-specific begin of seasons

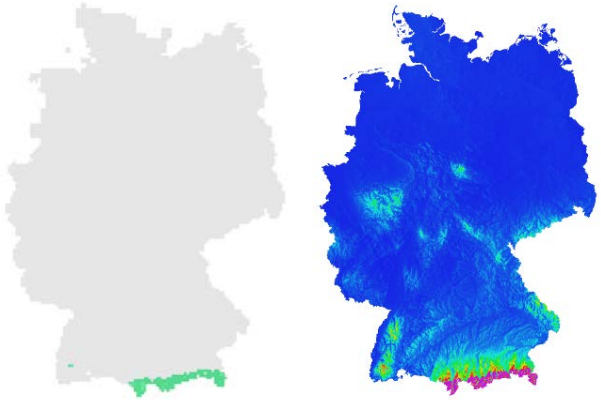


	Average Recall	Top-20 Recall	Median Rank	Avg List Length
Phenology	70%	3%	466	1117

Habitat Suitability

- Machine-learning approach: maximum entropy density estimation[1]
  - Relating taxon occurrences to environmental predictors
  - Model for each of 2770 taxa

- Predictor variables: environmental geodata
  - Climate
  - Height
  - Land cover
  - Soil
  - Geomorphology



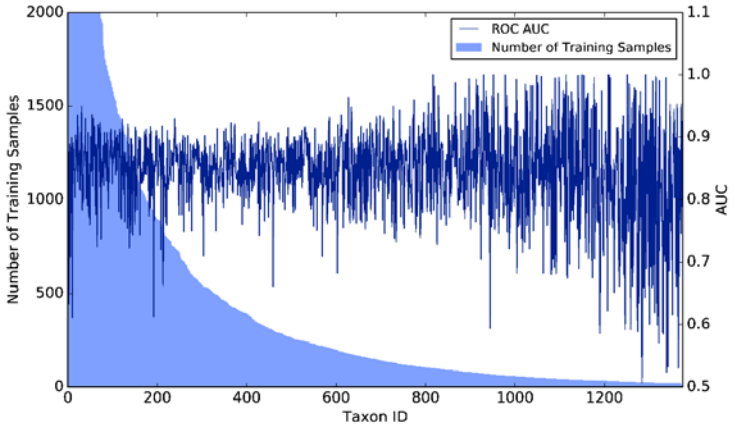
Soldanella alpina

- Training data acquisition:
  - Known taxon presence at observation locations (GBIF)
  - Randomly chosen absence locations (Florkart)

[1] Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling.

Habitat Suitability: Contribution

- Model performance dependent on
  - Training sample selection
  - Background geodata selection
  - Number of training samples

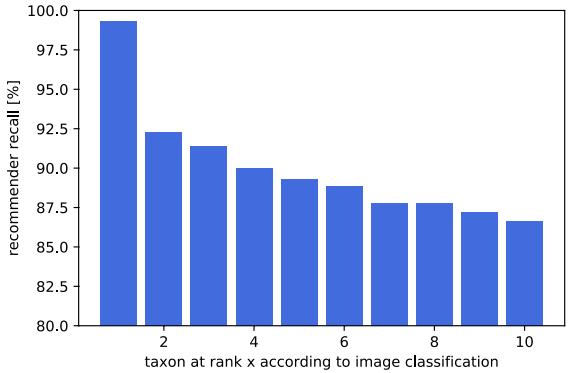


	Average Recall	Top-20 Recall	Median Rank	Avg List Length
Habitat Suitability	93%	7%	247	1379

Results

	Average Recall	Top-20 Recall	Median Rank	Avg List Length
Compound Recommender	90%	25%	70	860

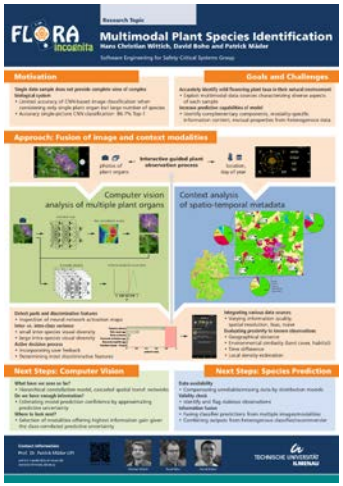
- Feedback for image classification
  - Taxa predicted from photos (top-1) have 99% recall in recommended list
  - 92% or less for following ranks
- Flora Incognita app
  - “Sanity check“ for image recognition
  - Flag observations as dubious/uncertain and inform user
- Prospects
  - Integration into image classification pipeline
  - Needs more comprehensive data basis -> more densely sampled map of Germany



Wrap-up

Thanks for listening!

Questions?



Poster Session 17:00