Flanders
State of
the Art

Supporting scientific research @ INBO

Master of Statistical Data Analysis Day, Coupure Campus, UGent

Dimitri Brosens, Peter Desmet, Daniel Du Seuil, Ivy Jansen, Els Lommelen, Jo Loos, Tanja Milotic, Thierry Onkelinx, Paul Quataert, Frederic Piesschaert, Hans Van Calster, Toon Van Daele, Floris Vanderhaeghe, Gert Van Spaendonck, Stijn Van Hoey, Pieter Verschelde, Toon Westra, Carine Wils

Research Institute for Nature and Forest (INBO)

www.INBO.be

INBO

www.INBO.be

INBO

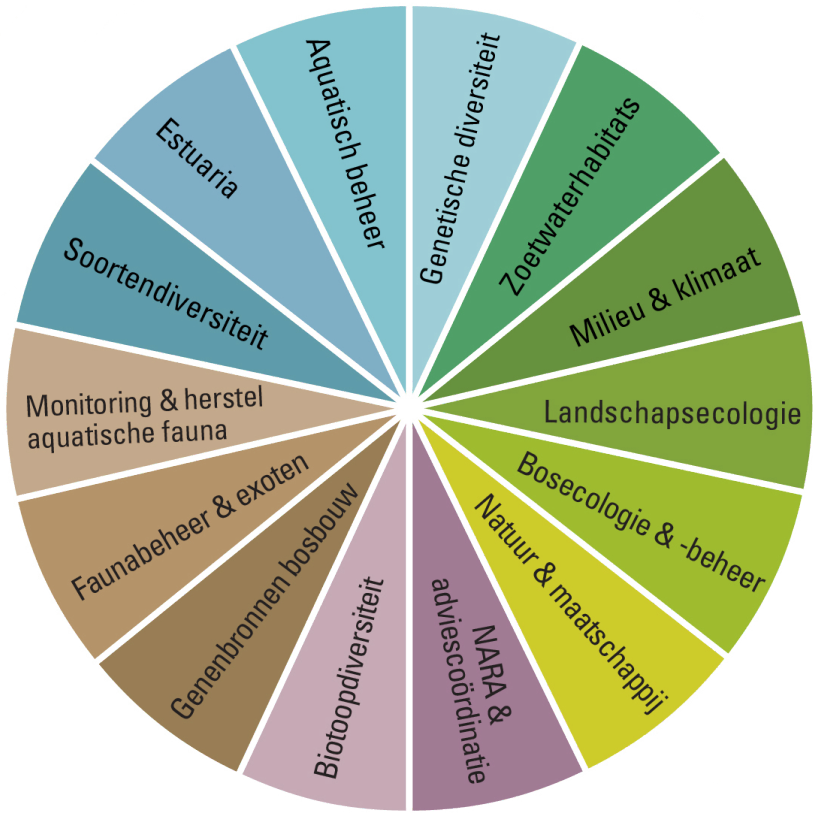
- Instituut voor Natuur- en Bosonderzoek / Research Institute for Nature and Forest
- Flemish research and knowledge centre for nature and its sustainable management and use
- conducts research and supplies knowledge to policy makers and stakeholders
- supports
  - agencies of Flemish government related with countryside or nature within urban areas
  - organisations involved in nature conservation, forestry, agriculture, hunting and fishery

Flanders
State of the Art

Data collection and provision

- INBO collects and provides data for international reporting
- Participates in (inter)national research networks
  - LTER: the Long Term Ecological Research network
  - LifeWatch: e-Science and Technology European Infrastructure for Biodiversity and Ecosystem Research
  - ALTER-Net: A Long-Term Biodiversity, Ecosystem and Awareness Research Network
  - …

Organigram

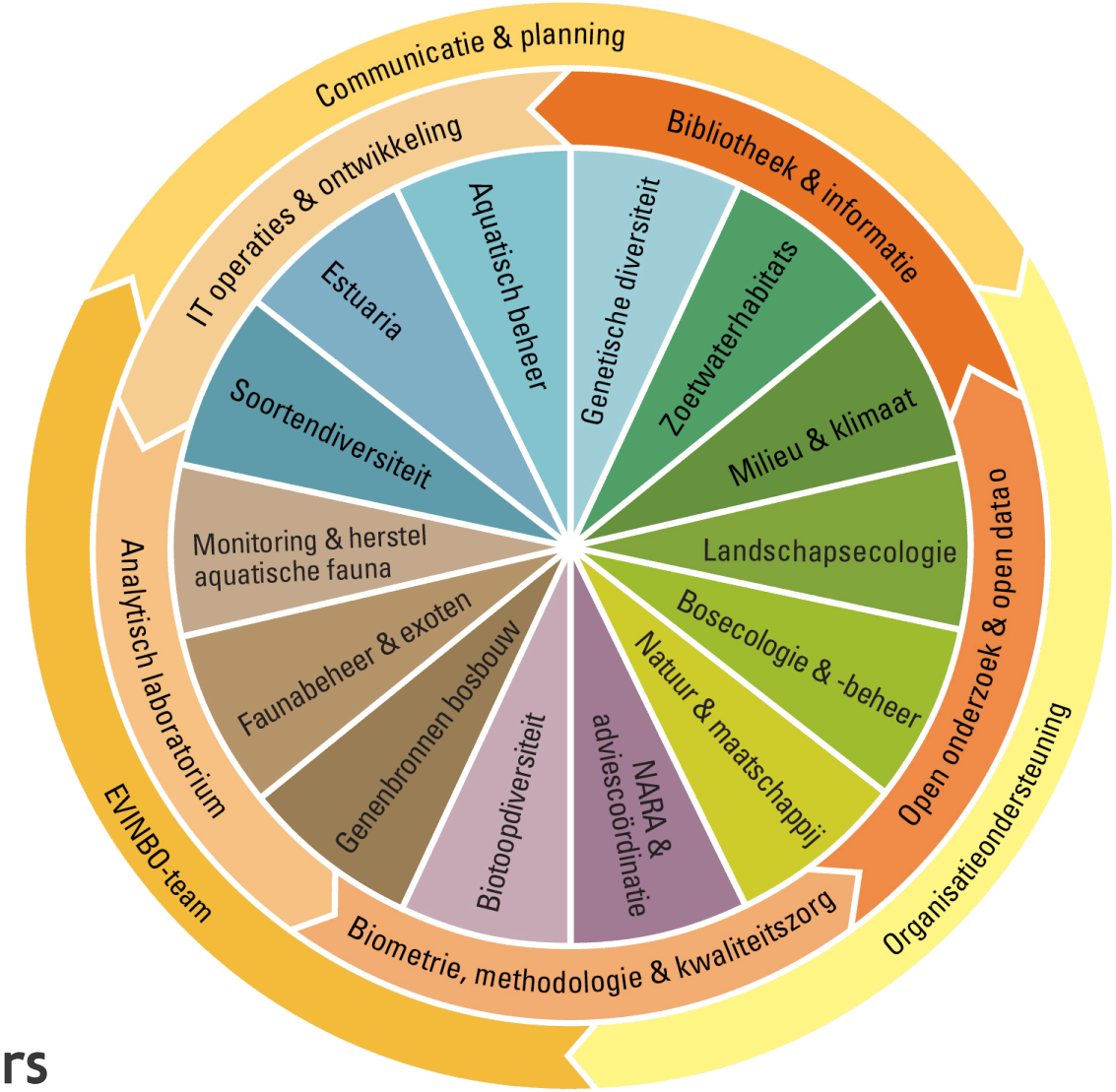


Flanders
State of the Art

Organigram

Flanders
State of the Art

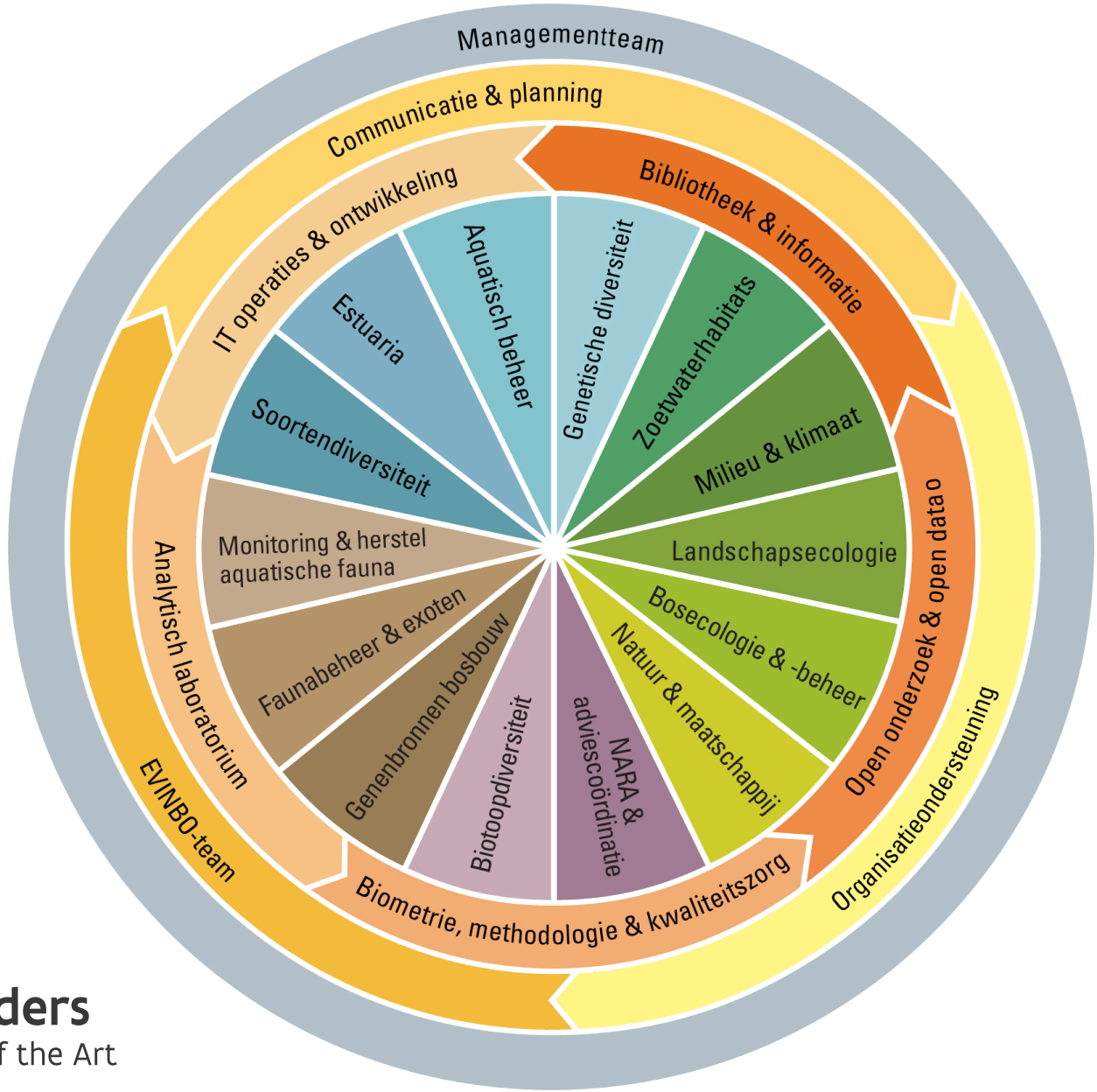
Organigram



Flanders
State of the Art

www.INBO.be

Organigram



Flanders
State of the Art

www.INBO.be

Staff

- 262 staff
  - 191 scientific staff (scientists + technicians)
  - 40 support for scientific staff
  - 22 administrative staff
  - 9 management
- 5 specialised teams dedicated to support the scientific staff
  1. Biometry, Methodology & Quality assurance (8)
  2. Open data & open research (8)
  3. Laboratory (10)
  4. IT operations & development (6)
  5. Library (8)

Flanders
State of the Art

Biometry, Methodology & Quality assurance

www.INBO.be

A scientific corporate identity and quality assurance

- involvement from design phase up to implementation phase
- creating, evaluating and adjusting procedures
  - harmonising protocols for data collection
  - validation of data through statistical modelling
- improving reproducibility
  - scripted analysis (R)
  - dynamic documents (RMarkdown)
  - version control (git)
- gradual change!
  - focus on early adopters and early majority
  - stimulation and support is vital
  - new procedures should lower the burden in the long run
  - allow for a long transition period before enforcing change

Creating a methodological platform

- implementing methodologies
  - standardise methods and analyses with focus on long term monitoring and common experiments
  - testing new methods and analysis techniques
  - developing new methods and analysis techniques
- improving self-reliance of scientists
  - internal statistics courses ranging from “stats 101” to generalised linear mixed models and multivariate analysis
  - offer scientists tools to help themselves and each other: websites, webinars, books, …
- maintaining tools
  - software, manuals, documentation, books, …
  - templates for basic analyses
  - corporate identity styles for graphics (ggplot2) and dynamic documents (markdown, LaTeX, HTML)

Leading specific projects

Some projects are led by the team, in collaboration with scientists. These projects typically:

- have more statistical or technological challenges
- strive towards achieving higher quality and standardization levels at INBO
- have a longer term

Examples:

- design and implementation of monitoring networks to support and evaluate Natura 2000 objectives
- design and implementation of monitoring of nature management plans
- elaboration of population modelling methods

Support and services

- teaching
  - statistical courses (sampling theory, GLMM, multivariate data analysis, …)
  - R and RStudio
- consultancy
  - help desk: ad hoc questions on statistical techniques and R
  - sounding board: critical reflection on research goals, experimental design, statistical analysis, …
- expertise assignments
  - experimental design: spatio-temporal resolution, sample size calculation, financial impact, …
  - complex analyses requiring a full-fledged statistician
- quality control
  - scientists can ask for a review of their code

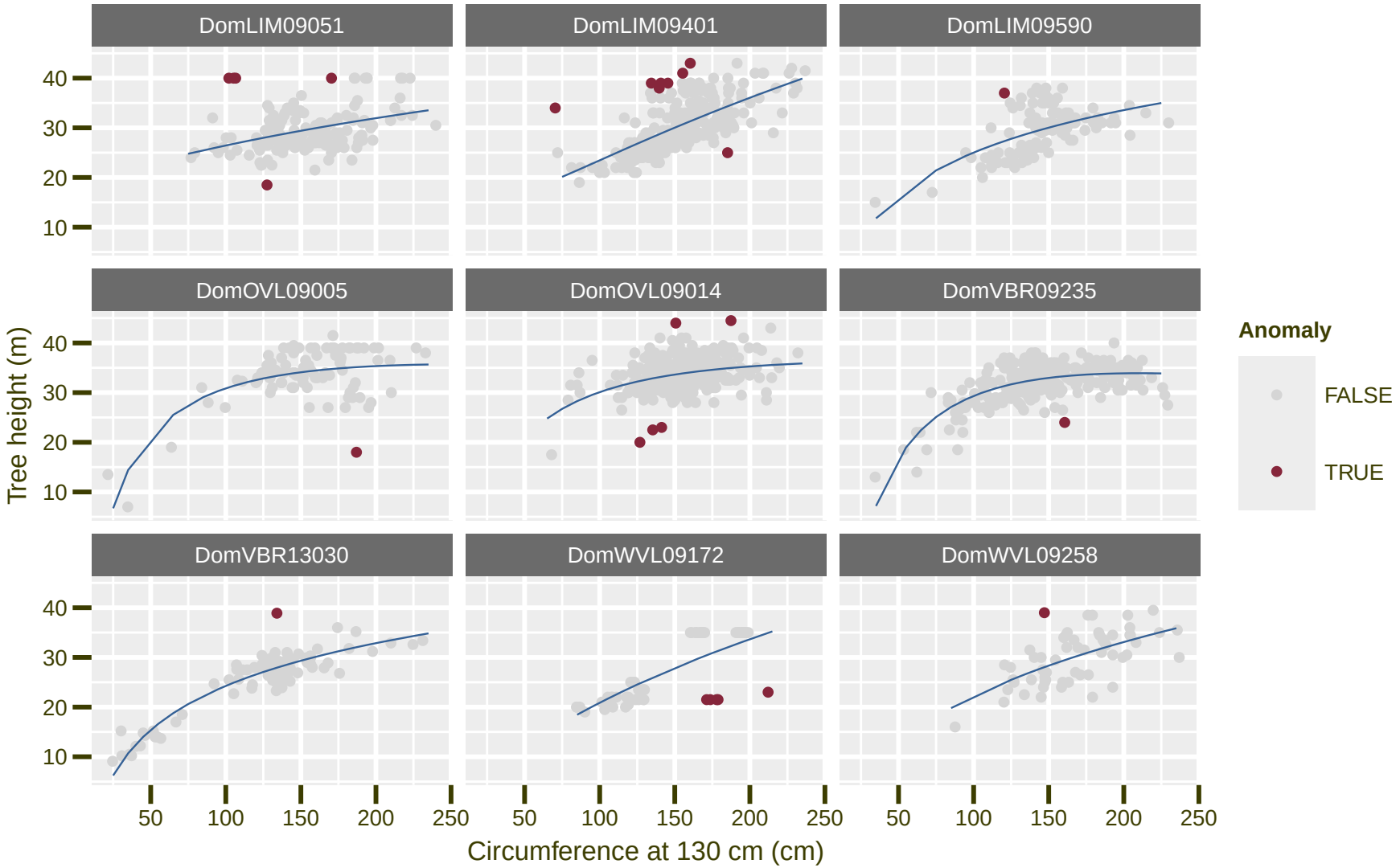
Flanders
State of the Art

Examples

Validation of data through statistical modelling

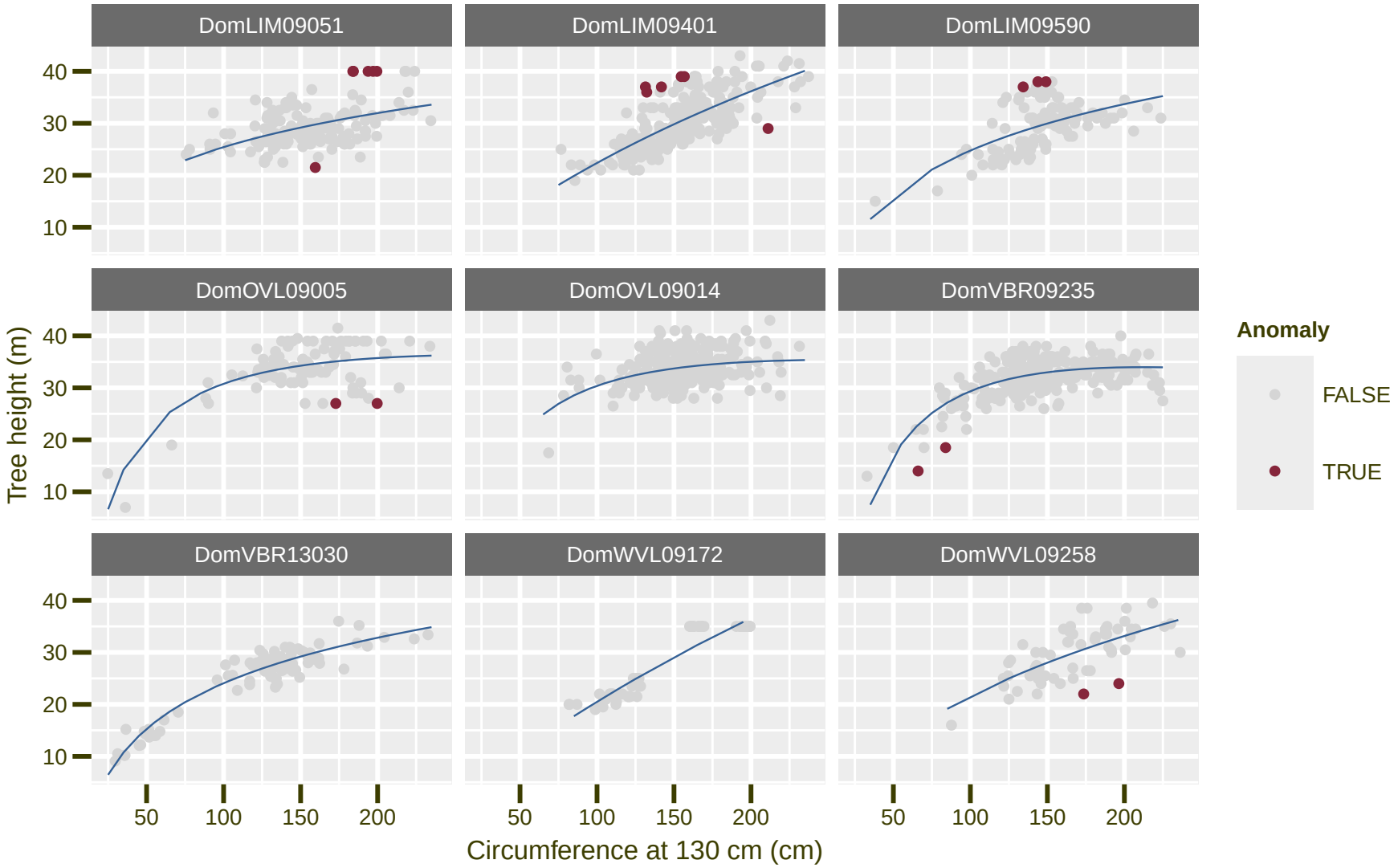
- inspecting each data point in a large dataset is impractical
- observed value is the result of an (unknown) data generating process
- try to model the data generating process
- look for extreme values given the model
  - strong residuals
  - strong (hyper)parameters
  - “anomalies”: *n* strongest values
- thoroughly inspect all observations linked to an anomaly
  - something wrong with the observation?
  - something wrong with the model?
- remove irrelevant observations, fix typos, adjust model
- repeat until
  - model makes sense
  - all anomalies are relevant and correct observations

Example: height-diameter curves for trees



Flanders
State of the Art

Example: updated height-diameter curves for trees



Flanders
State of the Art

Visual estimates of plant cover

- vertical projection of stems and leafs onto ground
- fraction of projected area per plant species
  - does not sum to 1
- estimated in classes with different binwidth
- ordinal data but often treated as continuous data
- survey quadrant size ranges from 1 m² to 256 m²

Flanders
State of the Art

Alternative: integral occurrence probability

- measure the shortest distance r_{ij} from sampling point i to nearest plant part of species j
  - $r_{ij} = 0$ when species j covers the sampling point
  - $r_{ij} = r_{max}$ when species j is not found within r_{max} around the sampling point
- requires several sampling points per location
- measured continuous data instead of visually estimated ordinal data
- $r^2/r^2_{max} \sim BI(p_0, p_1, \mu, \phi)$
  - p_0: cover estimated by point sampling
  - $1 - p_1$: presence within a circle of r_{max}
- integral occurrence probability
  - $1 - \mathrm{E}(r^2/r^2_{max}) = 1 - (p_1 + (1 - p_0 - p_1)\mu)$
  - $\frac{\sum_{i=1}^{n}(1-r_i^2/r^2_{max})}{n}$