ALGODAN Algorithmic Data Analysis

Parsimonious models in monitoring the environment and wireless sensor networks

Mika Sulkava, Mikko Korpela, Janne Toivola
Aalto University School of Science
Department of Information and Computer Science

28.10.2011

University of Helsinki

Aalto University

Research group

- Group leader:
  - Jaakko Hollmén
- Post-doc:
  - Mika Sulkava
- Graduate students:
  - Mikko Korpela
  - Janne Toivola
  - Prem Raj Adhikari
- Visitors 2010-2011: Miguel Angel Prada, Serafín Alonso Castro, Antonio Morán, and Andrés Sanz García

ALGODAN Algorithmic Data Analysis

Methodological goal: parsimony

- Learning parsimonious models from large and high-dimensional noisy data sets
- Parsimonious = compact, sparse, economical
- *Mission: model representation* should be as simple as possible or minimal
- Computationally intensive algorithms
- Good basis for interpretation

Current research projects

- Environmental Informatics for Analyzing the Role of Forests in the Global Carbon Cycle (2009-2011)
- Environmental proxy selection in temperature reconstruction
- Intelligent Structural Health Monitoring System (2008-2011)
- Gene selection in time-series gene expression data

Recent activity

- Organization of the 14th International Conference on Discovery Science (DS 2011)
- Local organization of DS 2011 and the 22nd International Conference on Algorithmic Learning Theory (ALT 2011), 5-7.10.2011

ALGODAN Algorithmic Data Analysis

Examples of current research

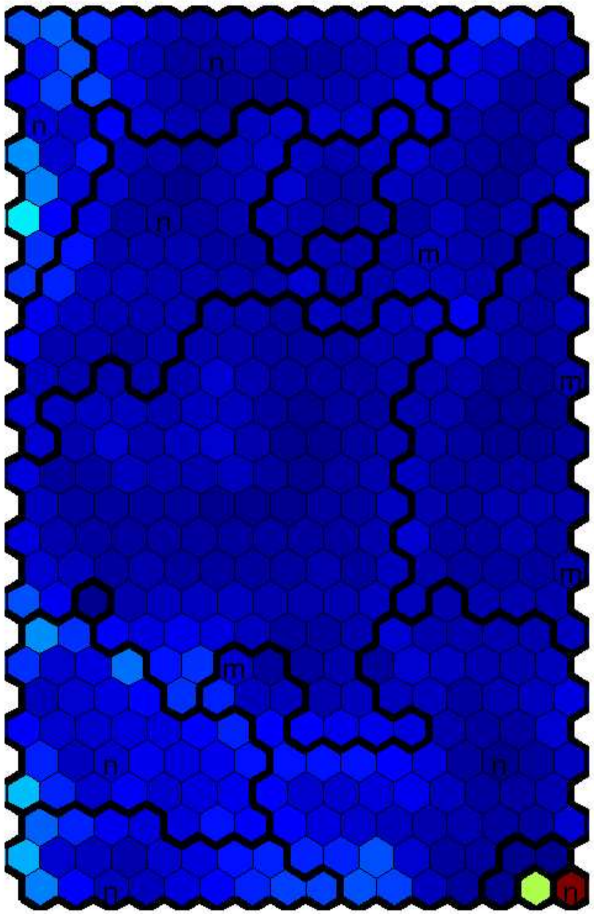
- Design of ecosystem monitoring networks (Mika Sulkava)
- Environmental proxy selection problems in temperature reconstruction (Mikko Korpela)
- Collaborative filtering for coordinated monitoring in wireless sensor networks (Janne Toivola)

Monitoring network design

- Understand the complex variability of the global flow of carbon and the climate, the growth of forests and the interconnections between them
- Design a representative network of measurement towers
- Carbon exchange between ecosystems and atmosphere
- Collaboration: Laboratoire des Sciences du Climat et de l'Environnement, France, University of Tuscia, Italy

ALGODAN Algorithmic Data Analysis

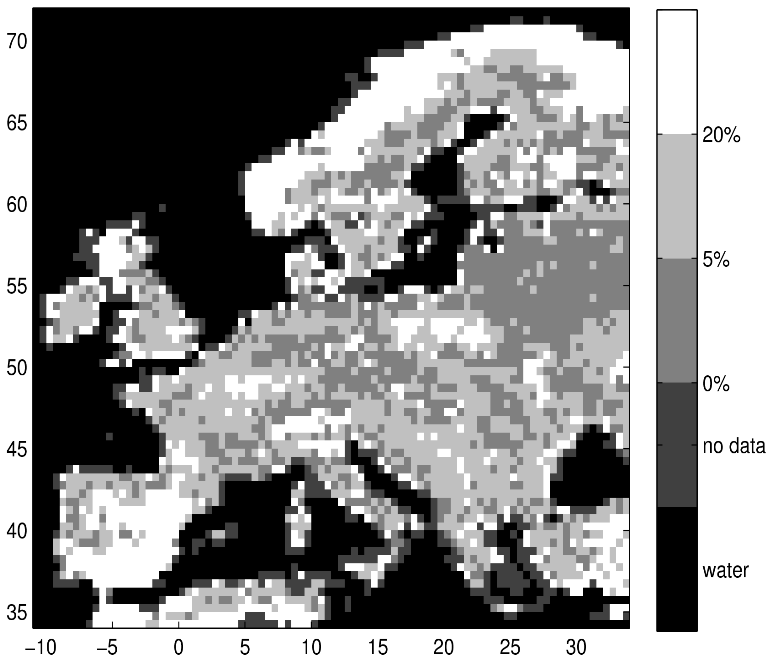
Monitoring network design



- CO_2 exchange of ecosystems is measured in towers around the world
- How many towers are needed for good representation?
- Where should new towers be located?
- Multidimensional climatic space
- Stratify study area using SOM and clustering

ALGODAN Algorithmic Data Analysis

Monitoring network design



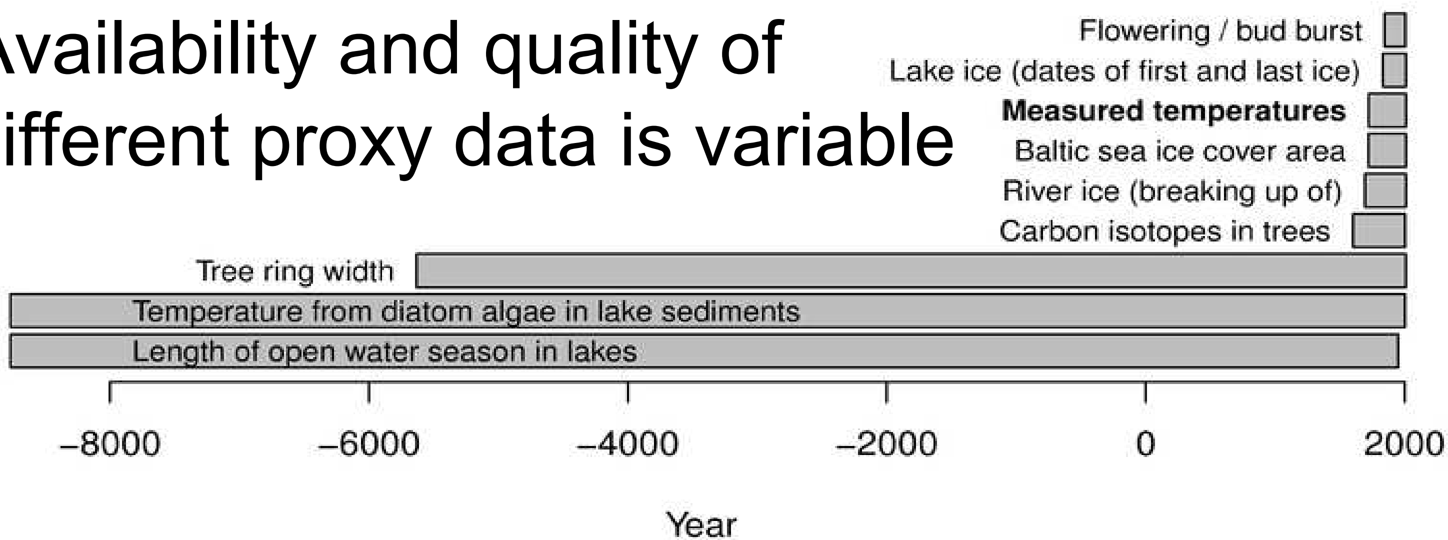
Upscaling error of a tower network

- Select locations of towers based on clustering
- Use an ecosystem model to estimate carbon exchange
- Upscale carbon exchange to the whole study area using the selected tower locations
- Use upscaling results to compare different networks
- Current research: importance of extreme climatic conditions

ALGODAN Algorithmic Data Analysis

Environmental proxy selection problems in temperature reconstruction

- Direct temperature measurements (partly) available from the past few hundred years
- Proxies must be used for estimating longer temperature records
- Availability and quality of different proxy data is variable



ALGODAN Algorithmic Data Analysis

Comparison of published temperature records

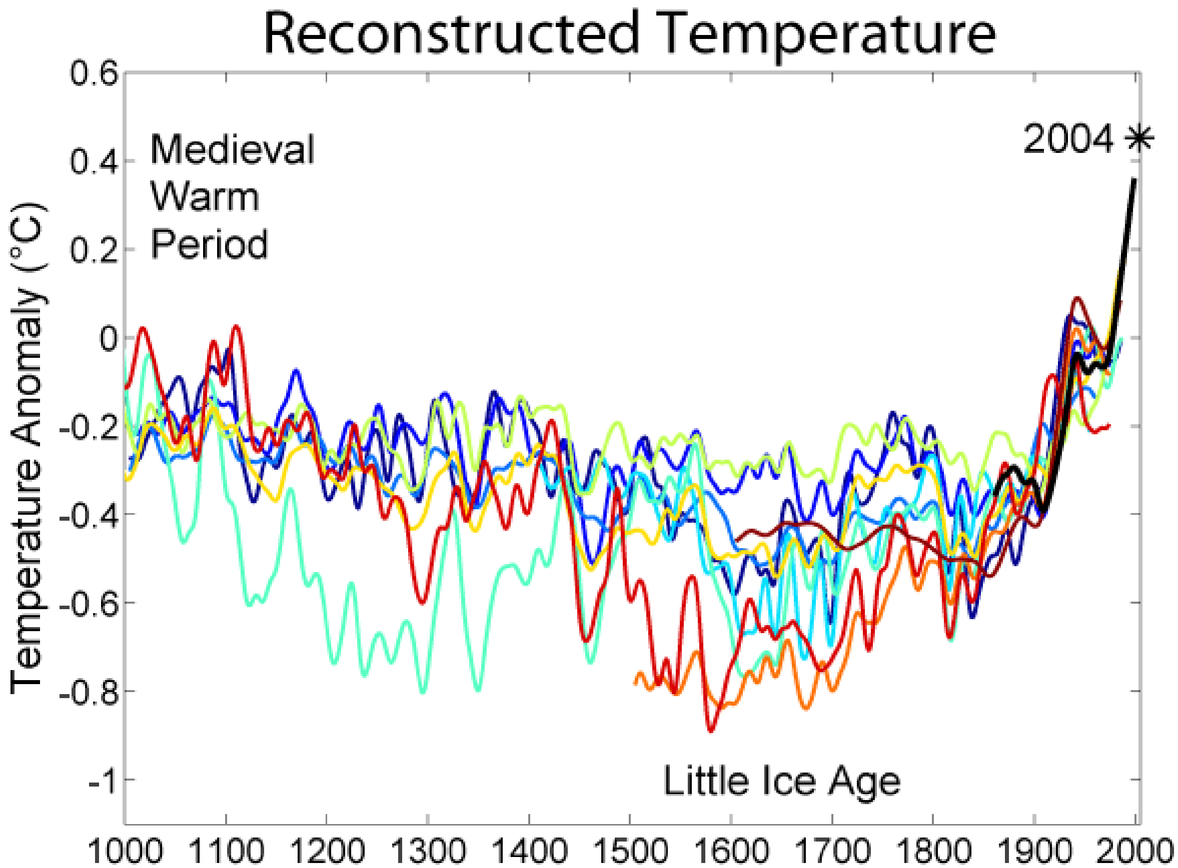
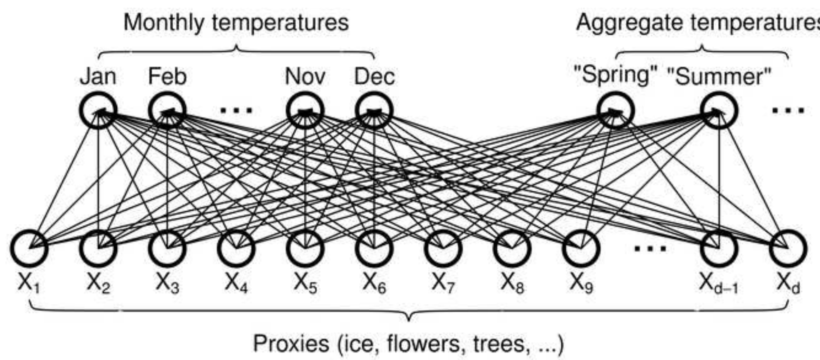


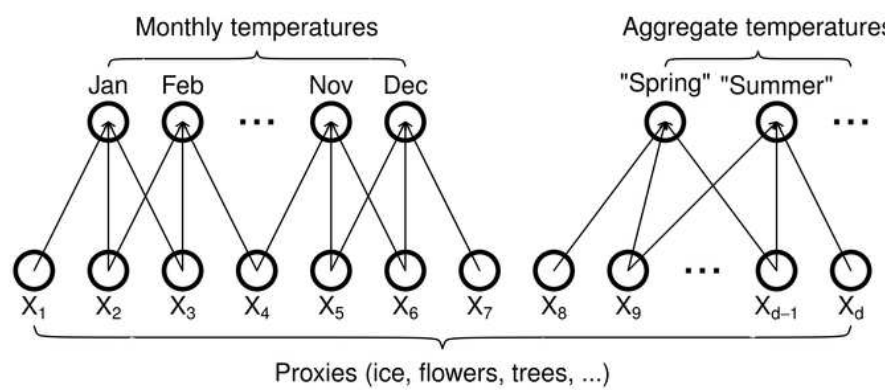
Image created by Robert A. Rohde / Global Warming Art
http://www.globalwarmingart.com/wiki/File:1000_Year_Temperature_Comparison_png

ALGODAN Algorithmic Data Analysis

Variable selection problem

- Full model is not plausible
- Find relevant proxies
  - Dependent on month / season
- Good temperature reconstruction possible for all seasons?

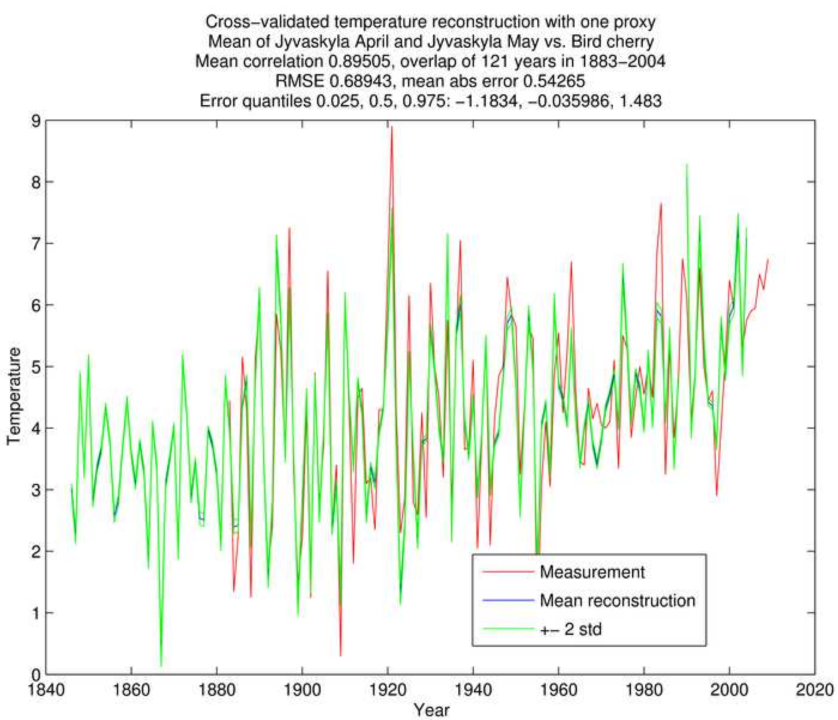


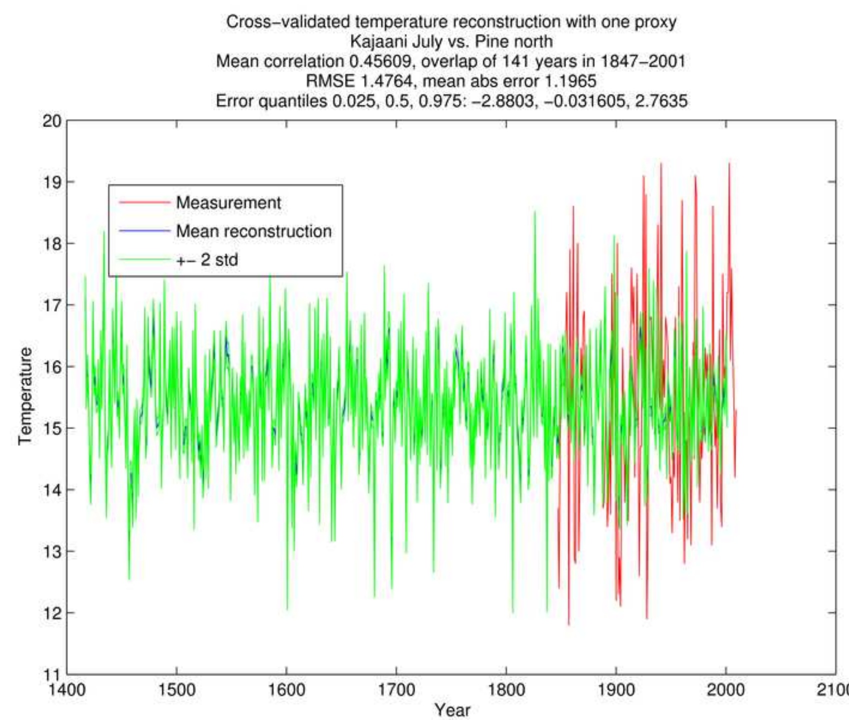


ALGODAN Algorithmic Data Analysis

Temperature reconstruction experiments (early results)

- With models using one proxy, there are large differences in the quality of the reconstructions

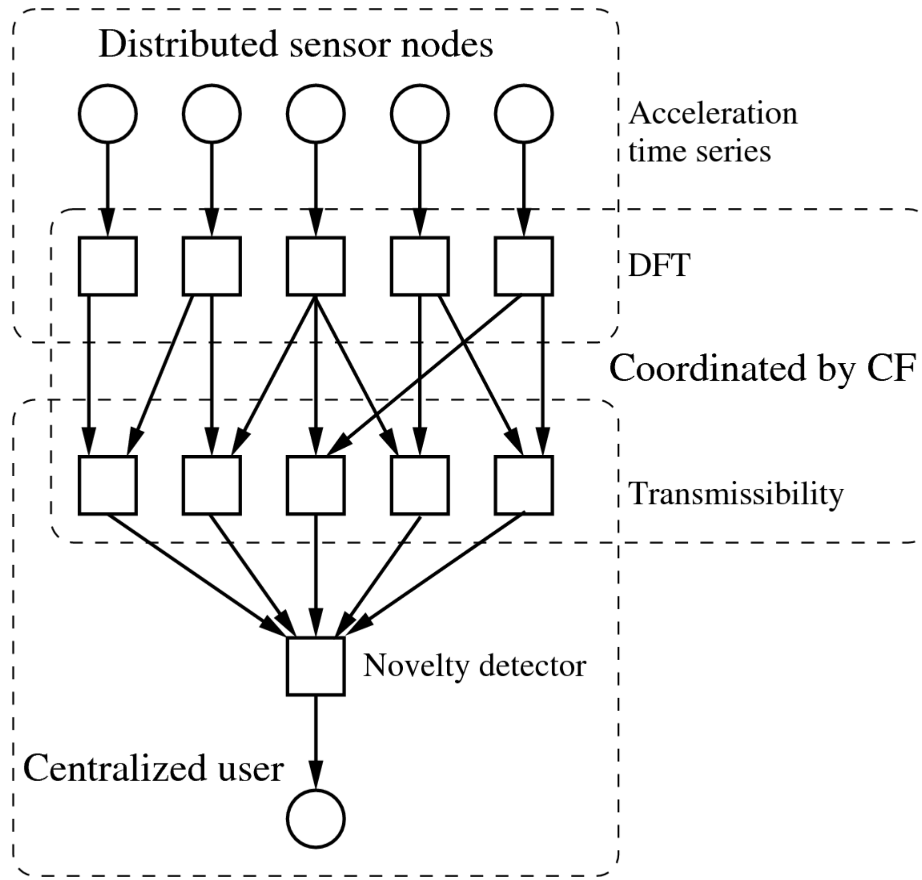




ALGODAN Algorithmic Data Analysis

Collaborative filtering in WSN

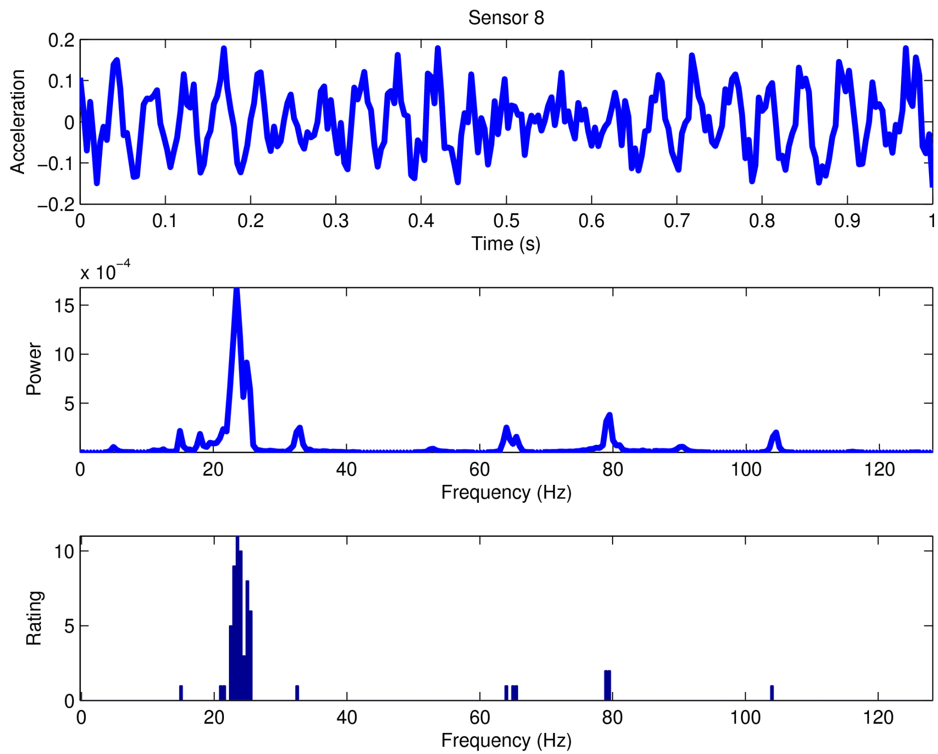
- Set of sensors
  - Not detectors…
- Local features
  - Max. D / node
- Combined features
  - Depend on distributed nodes
- Novelty detection
  - Wrapper-based feature selection not possible



ALGODAN Algorithmic Data Analysis

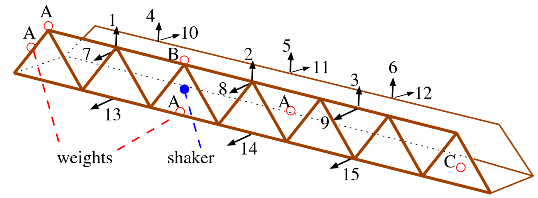
SHM: vibration power spectrum ratings

- Accelerometers
- Local features
  - Power spectrum: X[k]
- Local ratings
  - Accumulated votes for D “best” bins
- Combined features
  - Transmissibility: $X^1[k] / X^2[k]$
- Collaborative filtering
  - Select best features

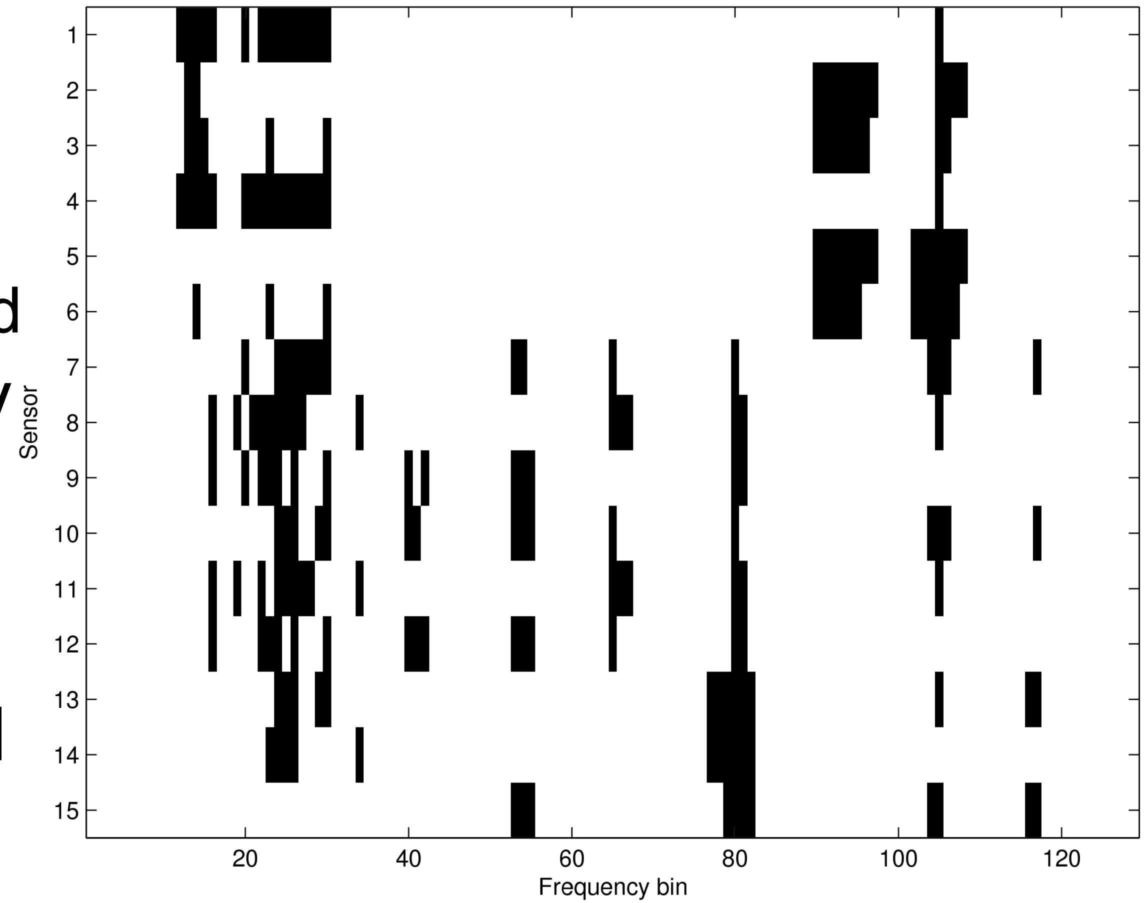


ALGODAN
Algorithmic Data Analysis

SHM: filtering example



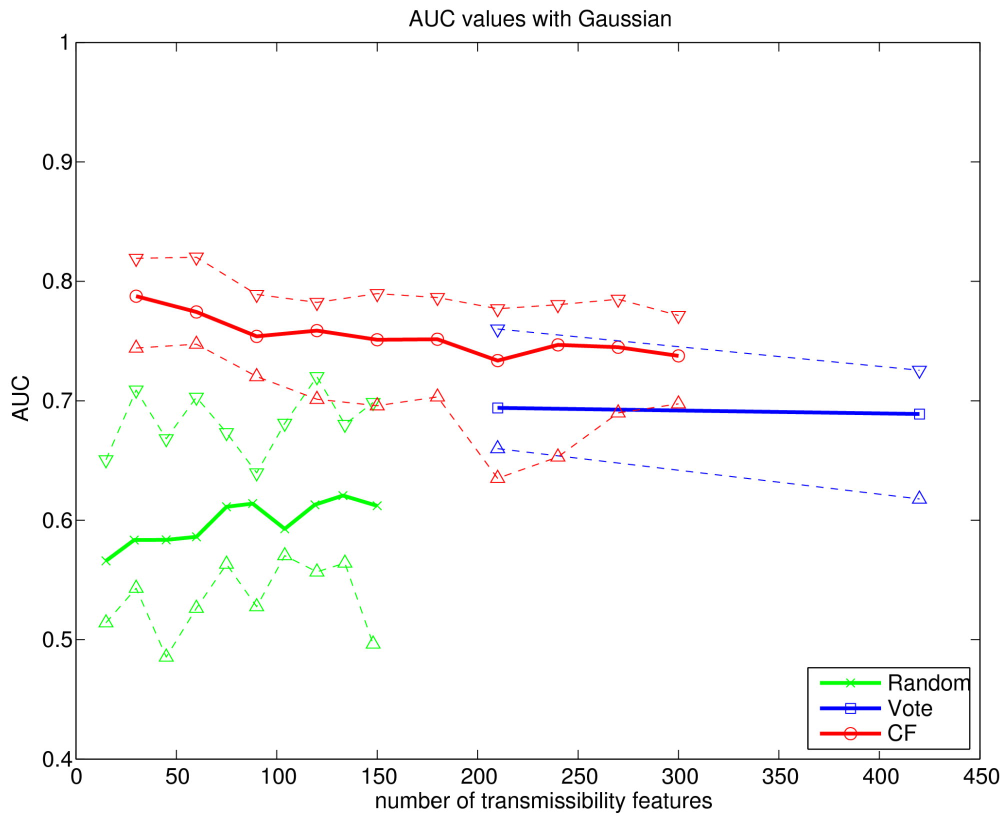
- 15 sensors
- 128 features
  - Max. 20 / node to be monitored
- Transmissibility
  - Pairs of black bars
- Sensors both collaborate and specialize



ALGODAN Algorithmic Data Analysis

SHM: detection accuracy (AUC)

- Area under ROC curve
  - 10 iterations
  - min, med, max
- D = 2...20
- Collaborative filtering
- Majority vote
- Random selection



ALGODAN Algorithmic Data Analysis