Prototyping Preventive Maintenance Tools with R

Erich Neuwirth, Julia Theresa Csar

The R User Conference 2010
National Institute of Standards and Technology (NIST),
Gaithersburg, Maryland, USA

Introduction

- **Machinery is constantly monitored**
  - A lot of data is collected (rotation, temperature)

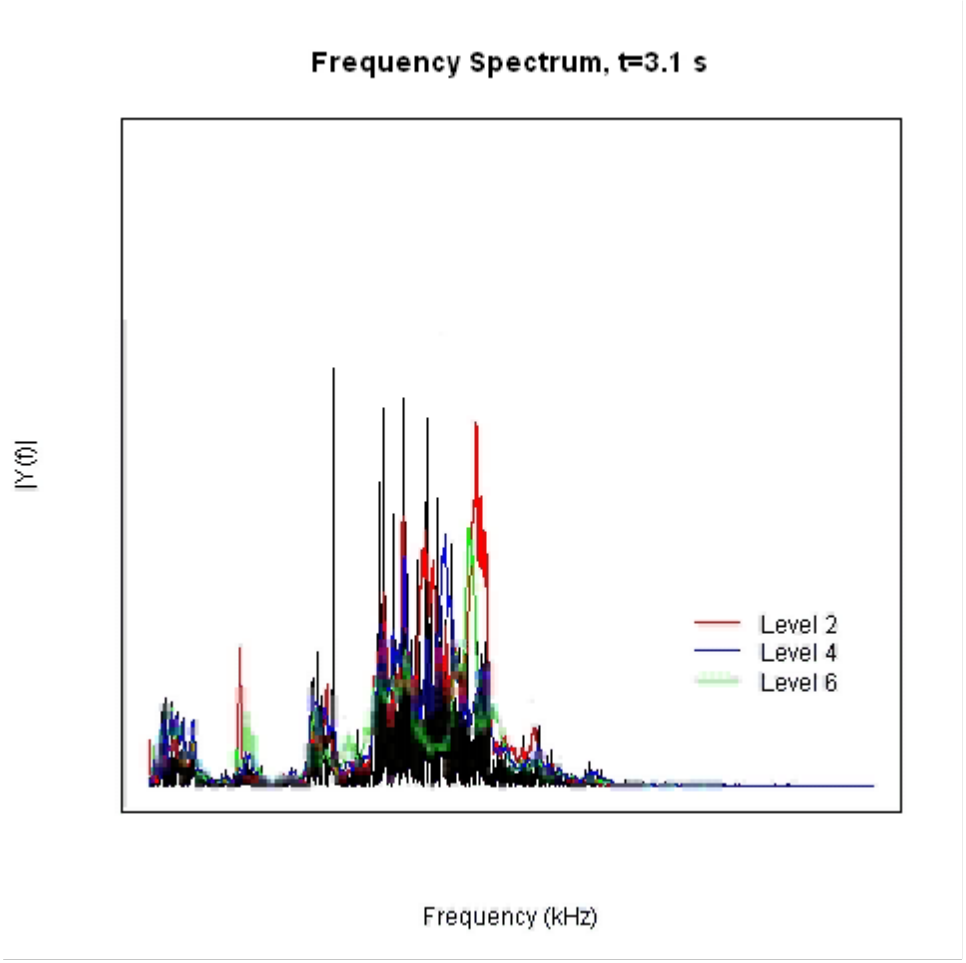
- **Extract a low resource representation for the monitored data**
  - to detect unusual behavior
  - to detect long time development

Example: Coffee Machine

- **Noise of the crushing mill is constantly monitored**

- **The goal is the detection of**
  - Low charging level of coffee beans
  - Level of grinding texture
  - Over long time: erosion

Frequency Spectrum



The 90%-confidence intervals of the crushing levels 2,4 and 6 are shown in the background

Extract multidimensional Representation

- Identify some important frequency intervals
  - Coffemachine: One Interval to identify the crushing level and one interval to recognize low bean charging level
- Calculate RMS over these intervals

  → Multidimensional Points
- Store those points and gain representing data points using the algorithm.
- Update those representation points frequently.
- The number of representation points is kept constant

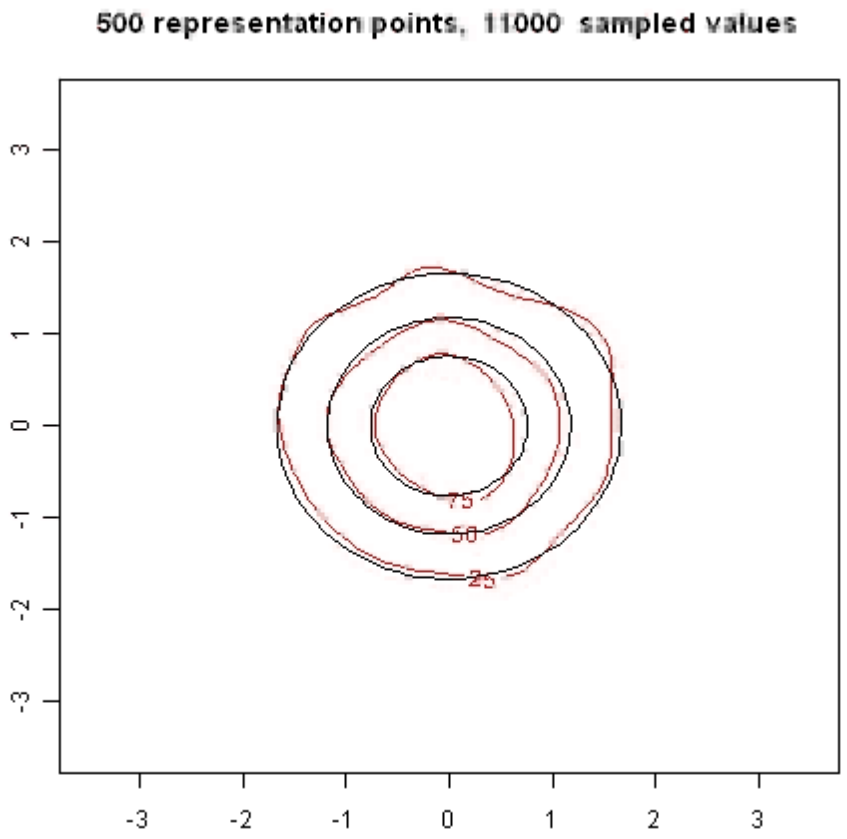
Algorithm

- Based on the algorithm for incremental quantile estimation presented in „Monitoring Networked Applications With Incremental Quantile Estimation” by John M. Chambers et al.
- Generalisation for multidimensional data was reached by using adaptive principal components analysis

Algorithm

- **Parameters to set:**
  - m...Number of Representation Points
  - n...Number of new points used for updating
- **Buffering Datapoints**
- **Starting algorithm after buffer is filled with *n* new points**
- **Updating the representation points using those new points**
- **Reset representation points after some time**

Algorithm

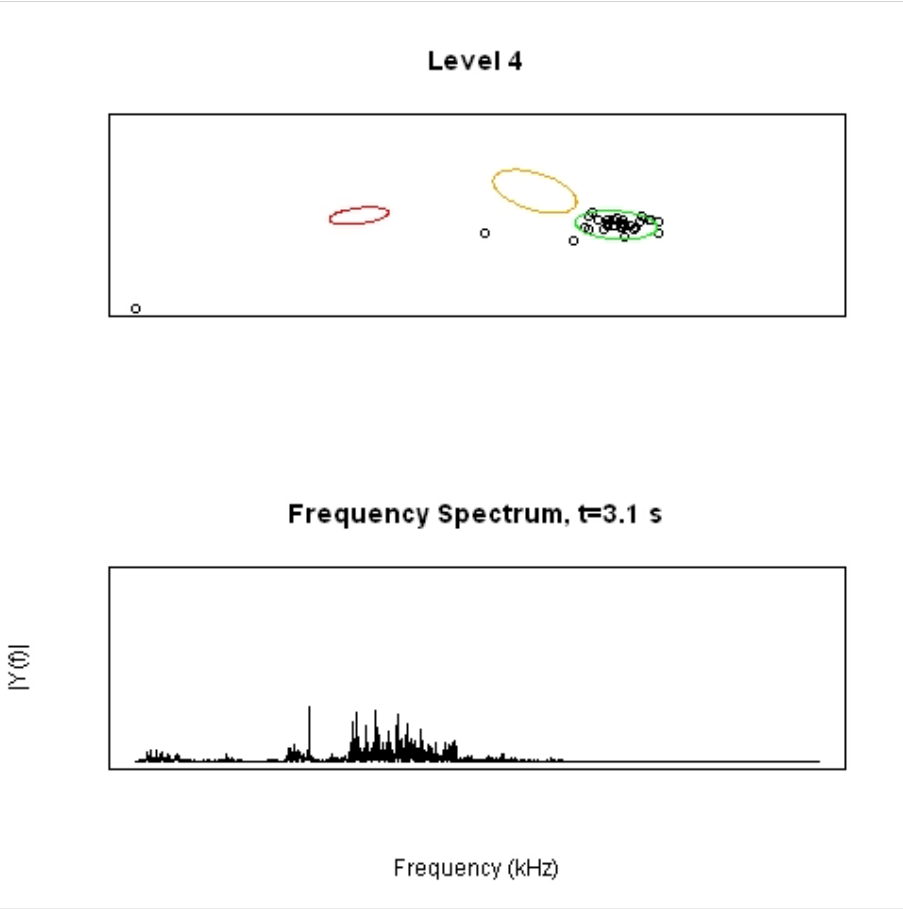


- The *Black Confidence Ellipsoids* are from the distribution used for generating random numbers
- Random numbers were generated using function „mvrnorm“ from R-Package „MASS“
- The *Red Ellipsoids* are derived from the calculated representation points using function „kde“ from R-package „ks“

Two-Dimensional representation of Coffeemachine

- **Identify two frequency intervals which contain information about the status:**
  - Coffee bean charging level
  - Crushing level
- **Use those points to gain the two-dimensional representation**
  - Visualization: confidence ellipsoids

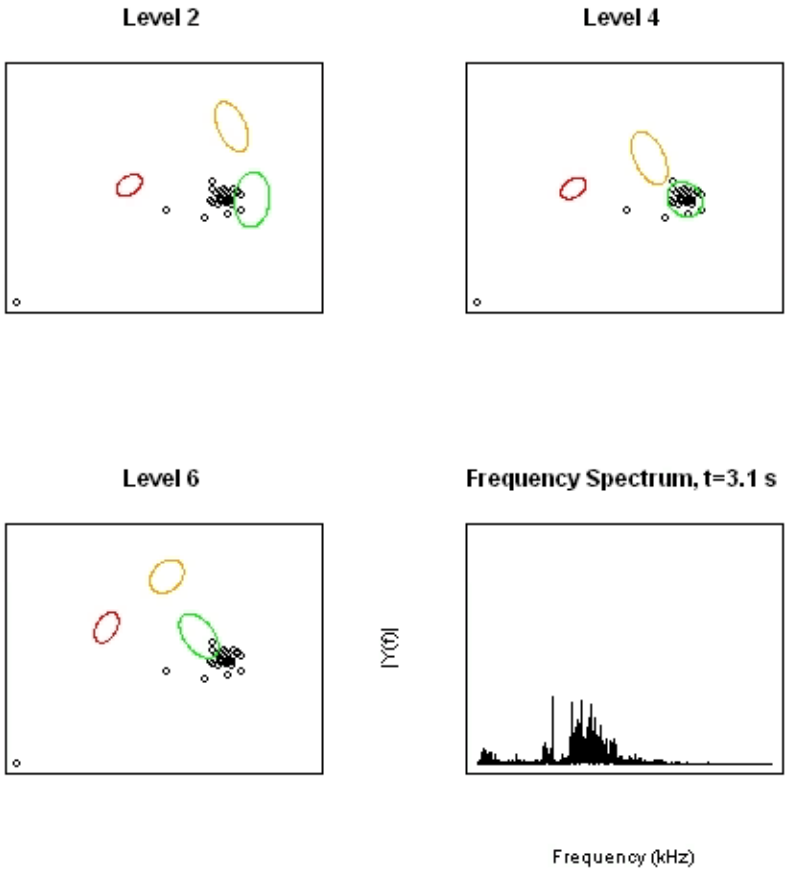
Two-Dimensional representation of Coffeemachine Status



Crushing level 4

- **Green:** OK
- **Orange:** Warning
- **Red:** out of coffee beans

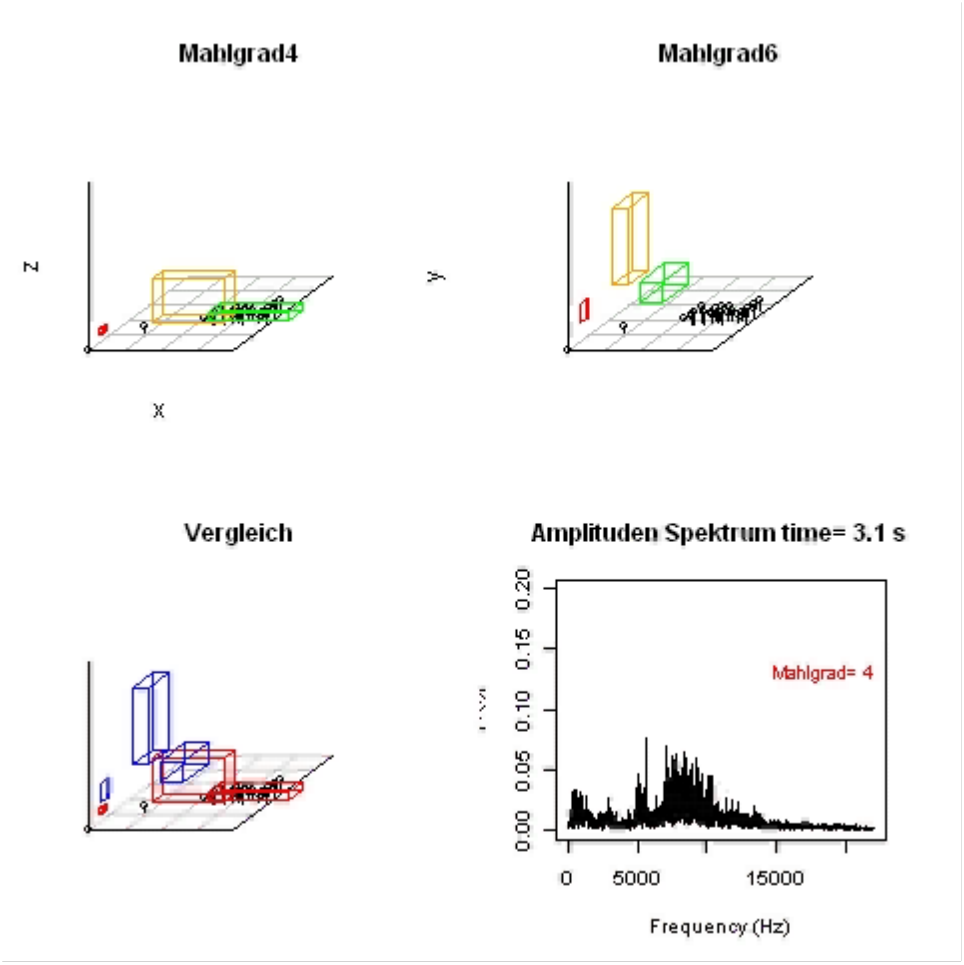
Two-Dimensional representation of Coffee Machine Status



- Confidence Ellipsoids are different at each crushing level

- Green: OK
- Orange: Warning
- Red: out of coffee beans

Three Dimensional



Crushing Levels 4 and 6

Red: Crushing Level 4
Blue: Crushing Level 6

R-Packages Used

- KS: **Kernel smoothing, Tarn Duong**
  - **kde**: Kernel density estimate for 1- to 6-dimensional data.
  - **rmvnorm.mixt**: Multivariate normal mixture distribution
- MASS: **Venables, W. N. & Ripley, B. D. (2002) Modern Applied Statistics with S. Fourth Edition. Springer, New York. ISBN 0-387-95457-0**
  - **mvrnorm**: Simulate from a Multivariate Normal Distribution

References

- John M. Chambers, David A. James, Diane Lambert and Scott Vander Wiel (2006). Monitoring Networked Applications With Incremental Quantile Estimation. *Statistical Science*, 2006, Vol. 21, No. 4, 463-475.