3rd International Workshop on Knowledge Discovery from Sensor Data (SensorKDD'09), Paris, France

28.06.2009

KDD-09 PARIS • June 28th - July 1st 2009 The 15th ACM SIGKDD Conference On Knowledge Discovery and Data Mining

Handling Outliers and Concept Drift in Online Mass Flow Prediction in CFB Boilers

Jorn Bakker (TU/e)
Mykola Pechenizkiy (TU/e)
Indre Žliobaite (TU/e)
Andriy Ivannikov (JYU)
Tommi Kärkkäinen (JYU)

NWO Netherlands Organisation for Scientific Research

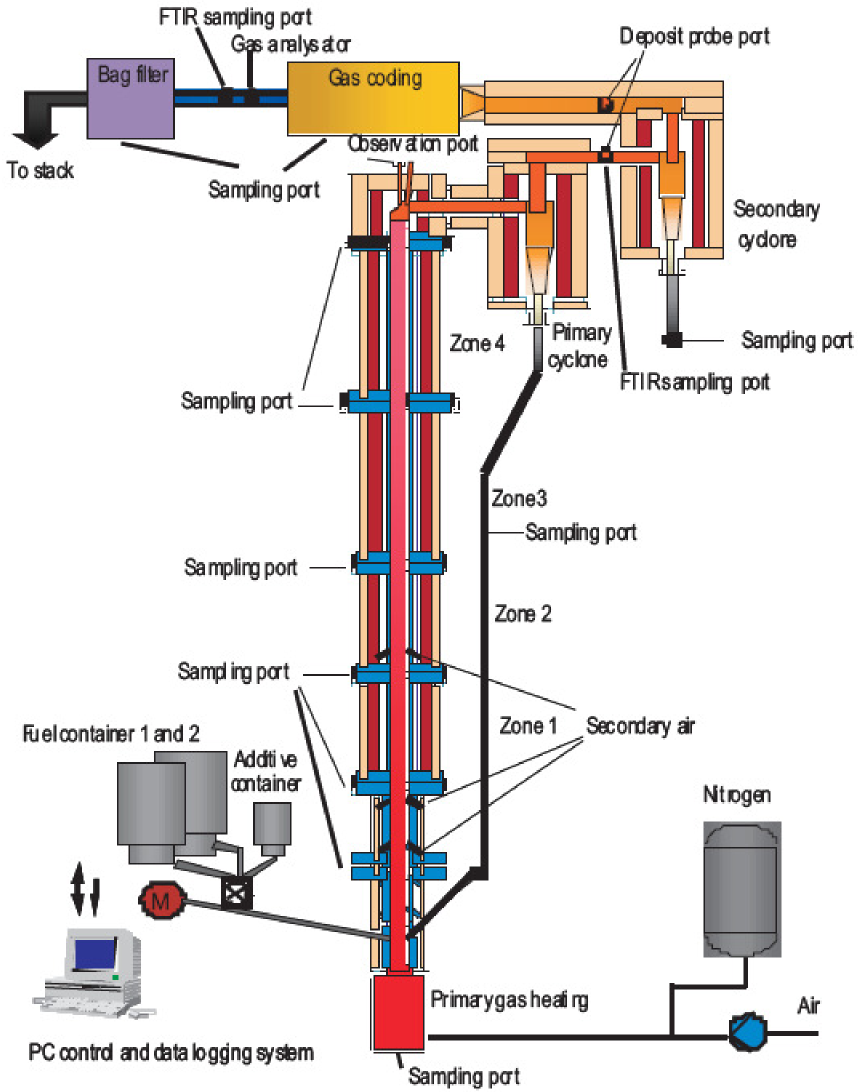
Dynergia Tekes

VTT

TU/e Technische Universiteit Eindhoven University of Technology

Where innovation starts

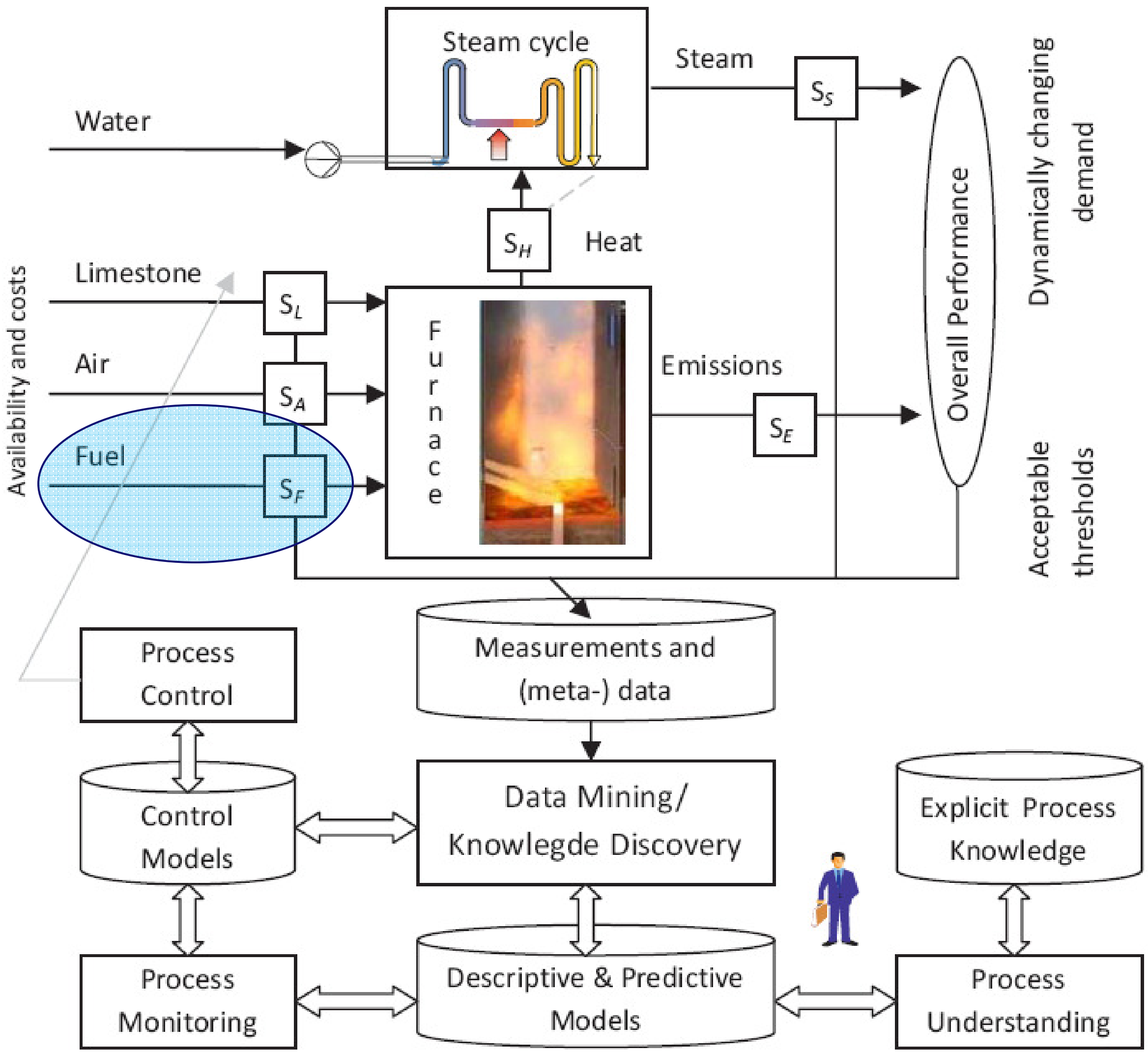
Circulating Fluidized Beds

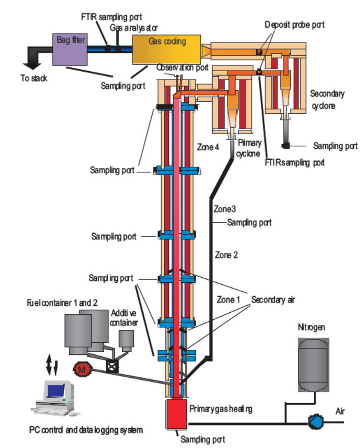


The laboratory scale CFB-reactor.

- The height of the riser of the boiler is **8 m** and
- the inner diameter **167 mm**.
- several separately controlled electrically heated and water/air cooled zones
- several ports for gas and solid material sampling are located in the freeboard area.

SensorKDD: CFB Boiler Optimization

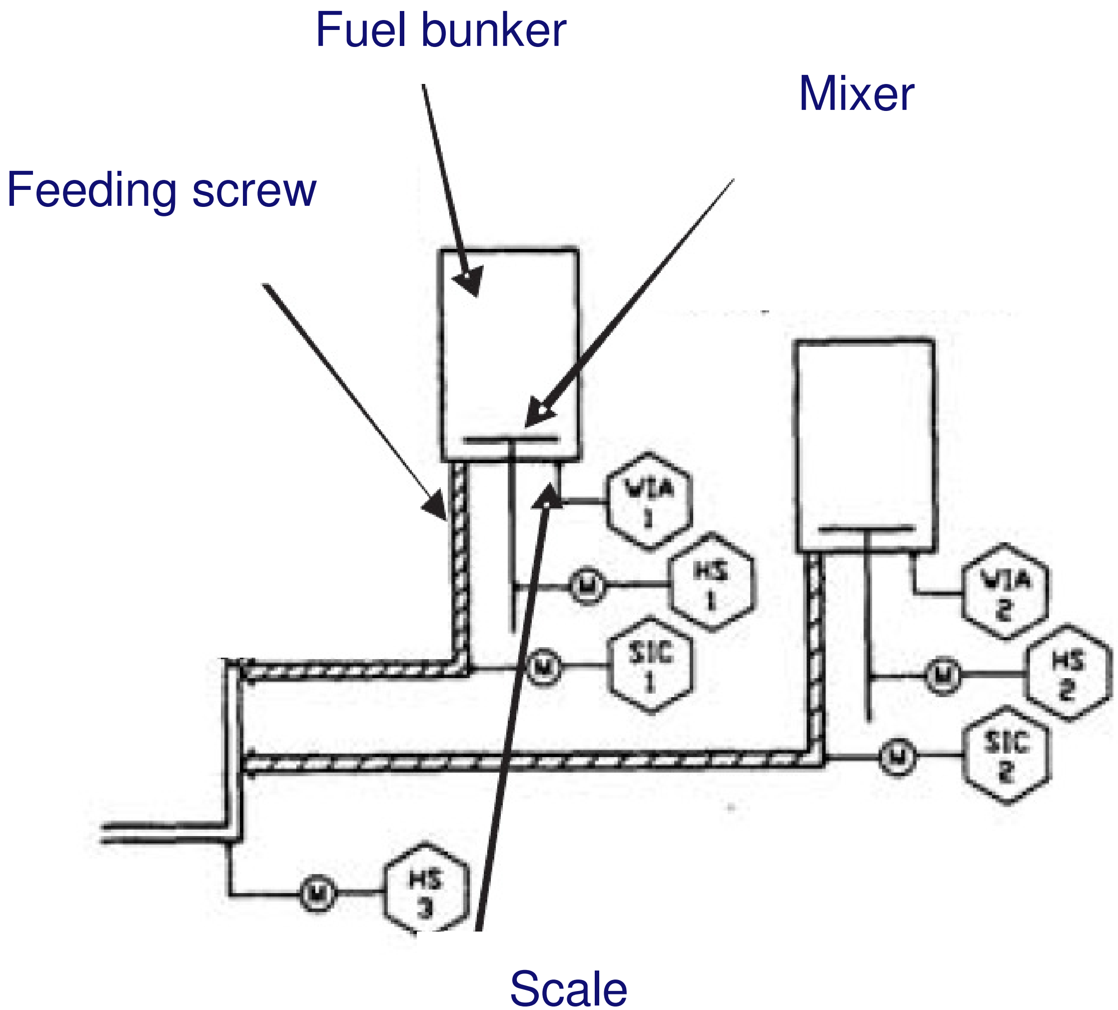




Outline

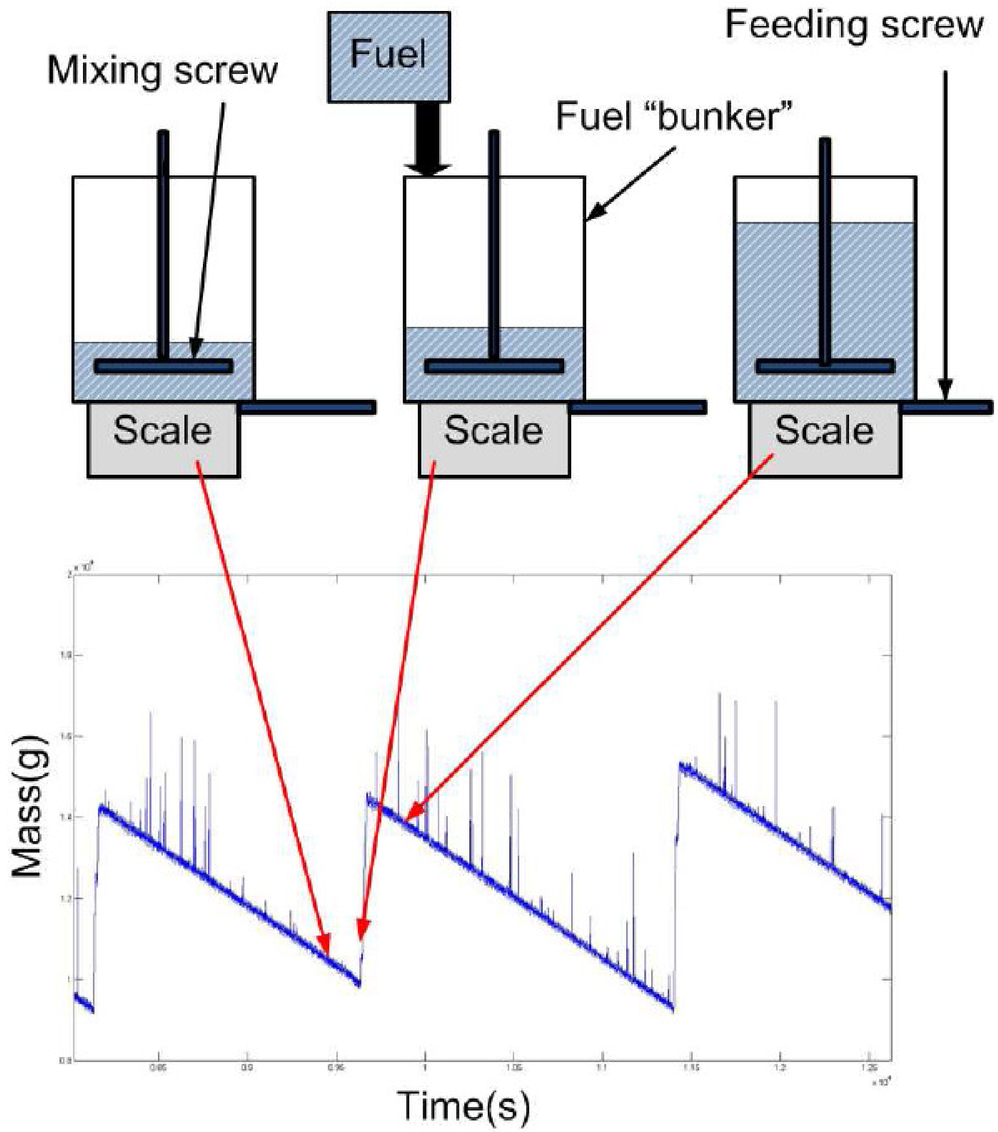
- Research Problem
  - Control of CFB processes
  - Mass flow prediction
- Our Approach
  - Learning the predictor
  - Explicit change detection methods
- Experimental Results
  - Predictor performance
  - Performance of change detection methods
- Conclusion & Further work

A fuel feeding system of the CFB-reactor



Online Mass flow prediction

- fluctuation in the signal of the scales ⇒ no reliable online data can be obtained from the mass flow of fuel to the boiler
- predict the *actual* mass flow based on the sensor measurements

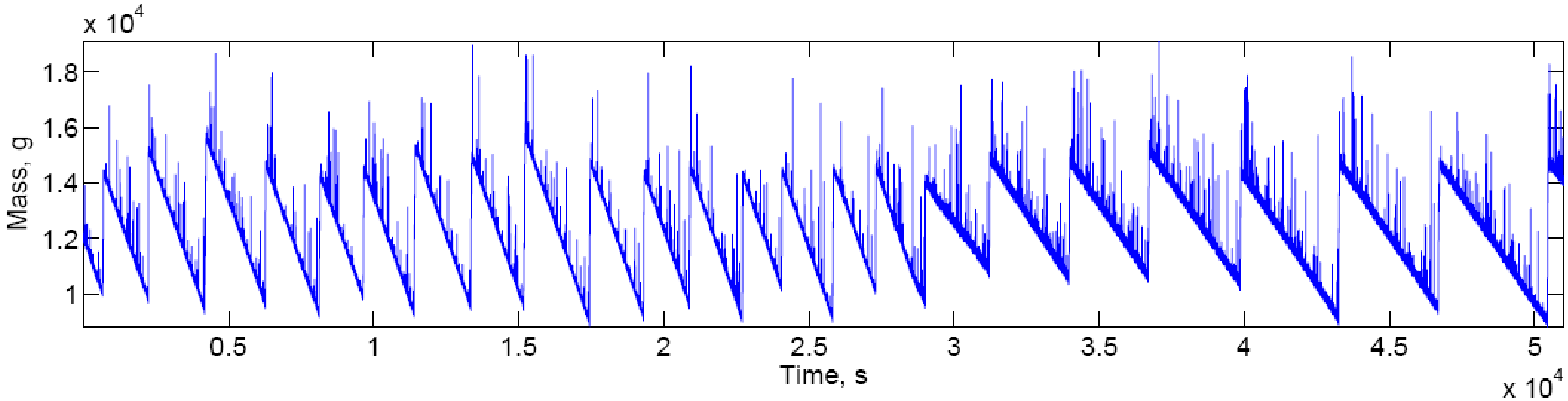


CFB Operational Changes: Concept Drift

- Explicit concept drift detection (context change in the CFB operational settings) and
- Learning of a new model each time after a drift was detected.
- 2 types of change in the mass flow of the boiler;
  - (1) sudden changes - transitions from the process of adding fuel to the process of fuel consumption and vice versa, and
  - (2) gradual changes typically caused by a changing mixture of different fuel types and changes in the speed of fuel consumption of replenishment (the latter type of change can be detected by using the rotation speed signal from the screw).

Measurements of the fuel mass in the tank

An example of mass flow signal collected during a typical experimentation with the CFB boiler

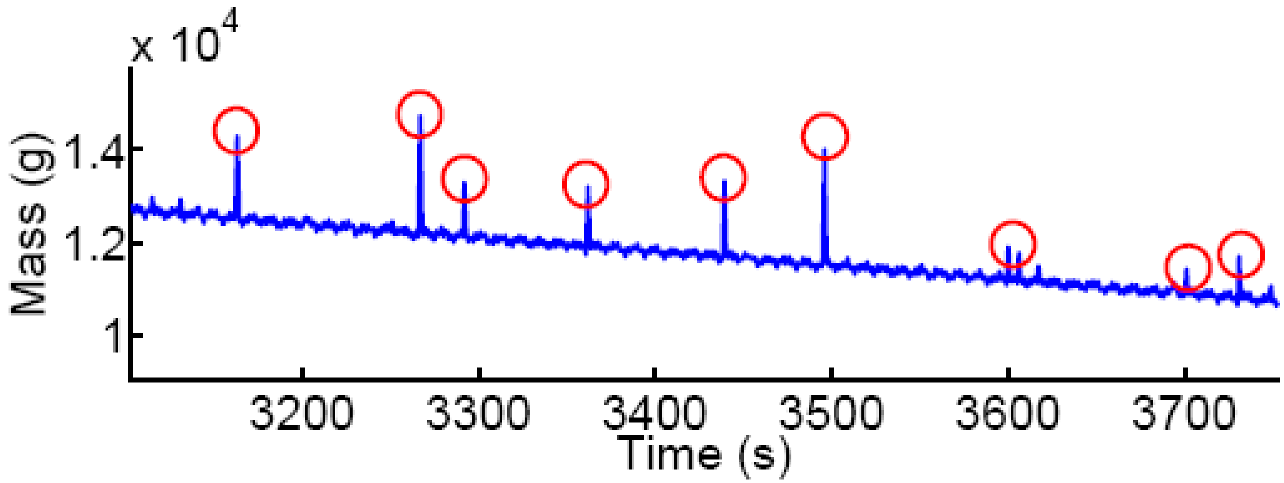


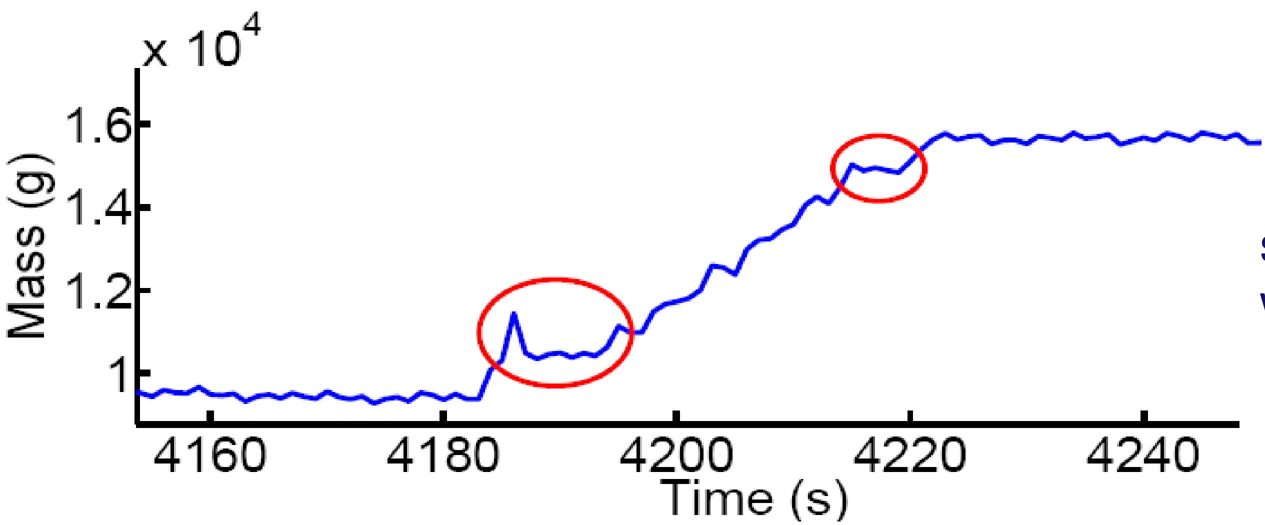
The sources of noise in the measurements are:

- mixing and feeding screws
- the occasional jamming of the fuel particle in the screw.

Why isn't it trivial?

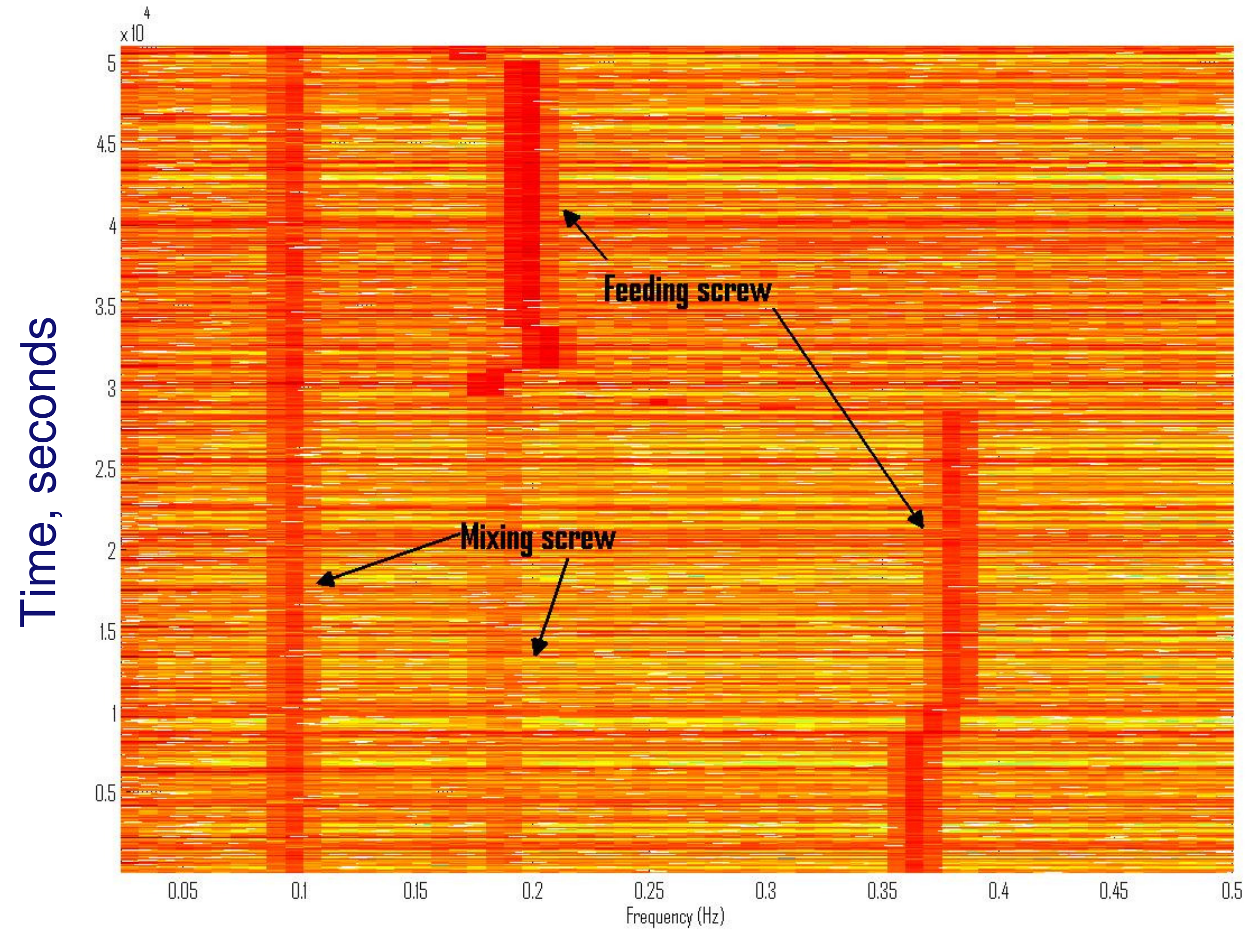
asymmetric nature
of the outliers





short consumption periods
within feeding stages

Influence of the screws on the mass signal



- Spectrogram computed with the Short-Time Fourier Transform.
- The color denotes the power spectral density (PSD).

Learning a predictor: domain knowledge

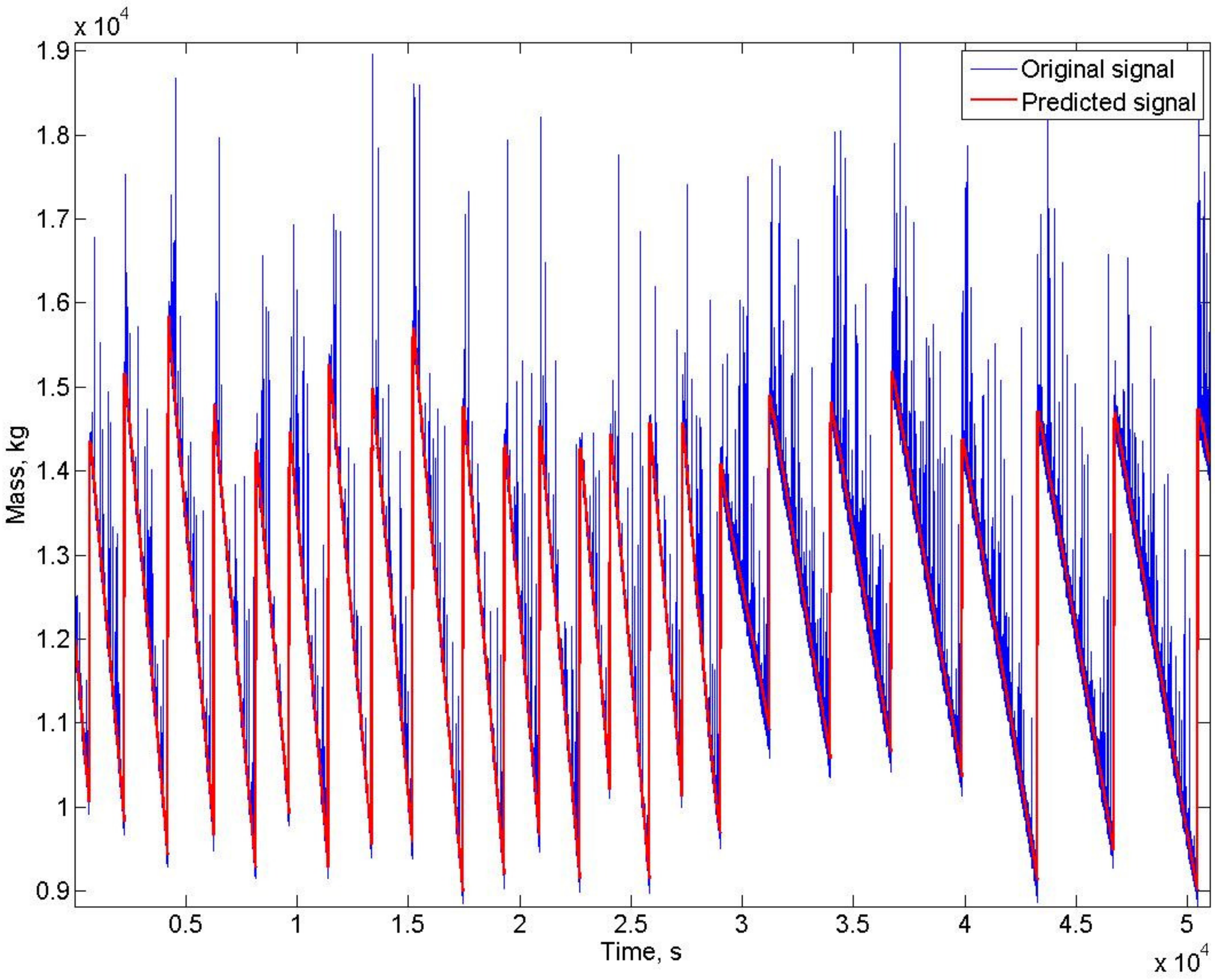
- measured signal at time *t* $y_t = m_t + \Sigma_t.$

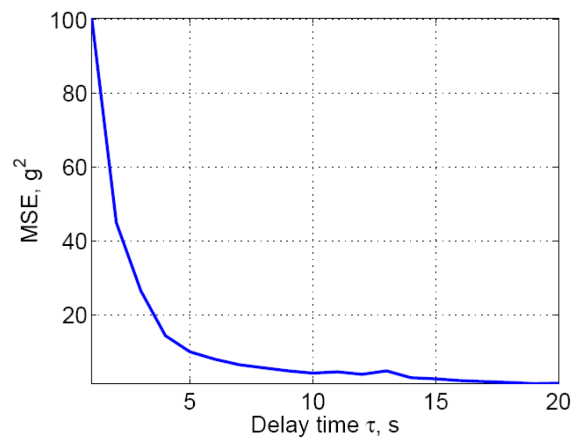
$$y_t = \underbrace{\frac{a \cdot t^2}{2} + v_0 \cdot t + m_0}_{m_t} + \underbrace{A \cdot \sin(\omega_{feed} \cdot t + \alpha_{feed}) + B \cdot \sin(\omega_{mix} \cdot t + \alpha_{mix}) + e(t)}_{\Sigma_t}$$

- *A* and *B*, ω_{feed} and ω_{mix}, α_{feed} and α_{mix} are amplitude, frequency and phase of the fluctuations caused by feeding and mixing screws, respectively;
- e_t - the random peaked high amplitude noise caused by the jamming of the fuel particle at time *t.*
- here we assume t_0 was the time of switch in the feeding/consumption stages.

$$\hat{y}_t = \frac{a \cdot t^2}{2} + v_0 \cdot t + m_0 + E(t)$$

If the state changes were known, prediction *seems* to be accurate





Approaches for changes detection

Identify noisy and change points

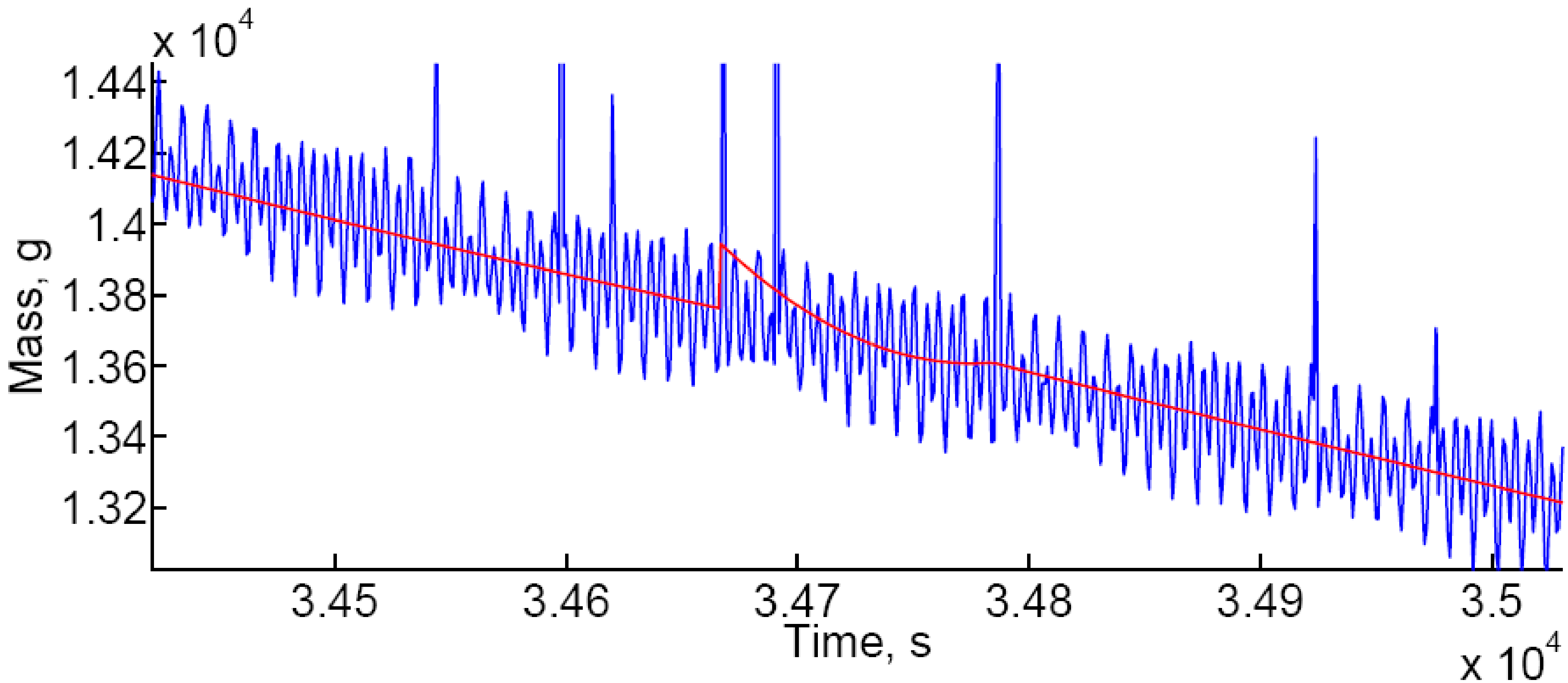
- windowing approaches;
- assumption: when the underlying process in the boiler changes, the current model for the prediction of the data will perform worse;
- state change detection = *several* outliers in a row;
- 1) detection based on comparing the prediction performance obtained on different subsets of data:
  - parametric approach: compares MSE and SD on two windows;
  - non-parametric approach: Mann–Whitney U test;
- 2) detection based on the statistical analysis of the raw data:
  - ADWIN method (Bifet&Gavalda, SDM'07): looks for statistically significant differences between the means of raw data.

Accuracy of detecting sudden changes

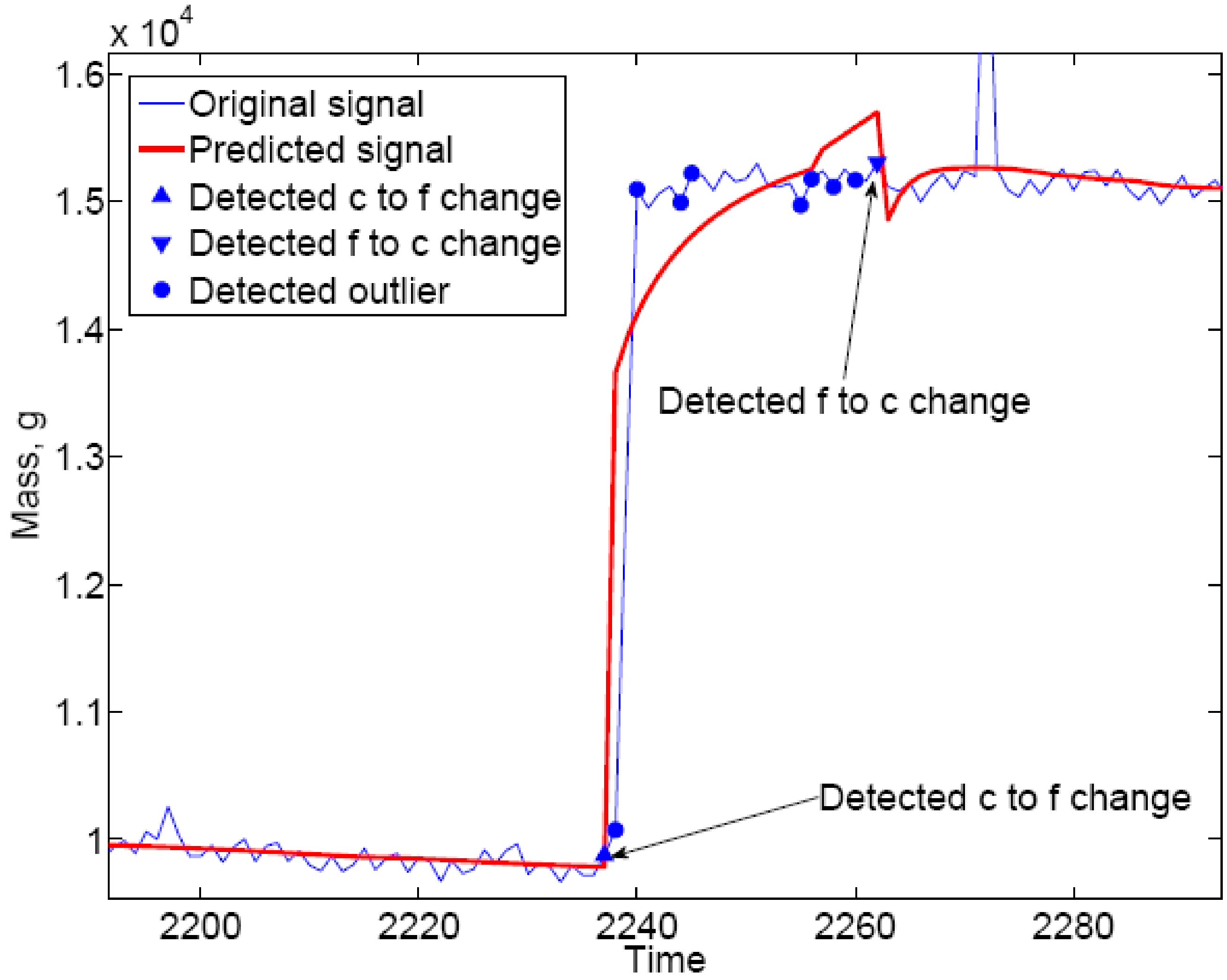
	Parametric test							
ϕ	P	N	κ	P	N	o	P	N
T	19	50898	T	4	50883	T	799	49088
F	74	5	F	89	20	F	892	217
	Nonparametric test							
ϕ	P	N	κ	P	N	o	P	N
T	24	50952	T	8	50942	T	816	49434
F	20	0	F	20	16	F	546	200
	ADWIN method							
ϕ	P	N	κ	P	N	o	P	N
T	24	50960	T	0	50972	T	-	-
F	12	0	F	0	24	F	-	-

φ feeding processes,
κ consumption processes,
o detecting outliers

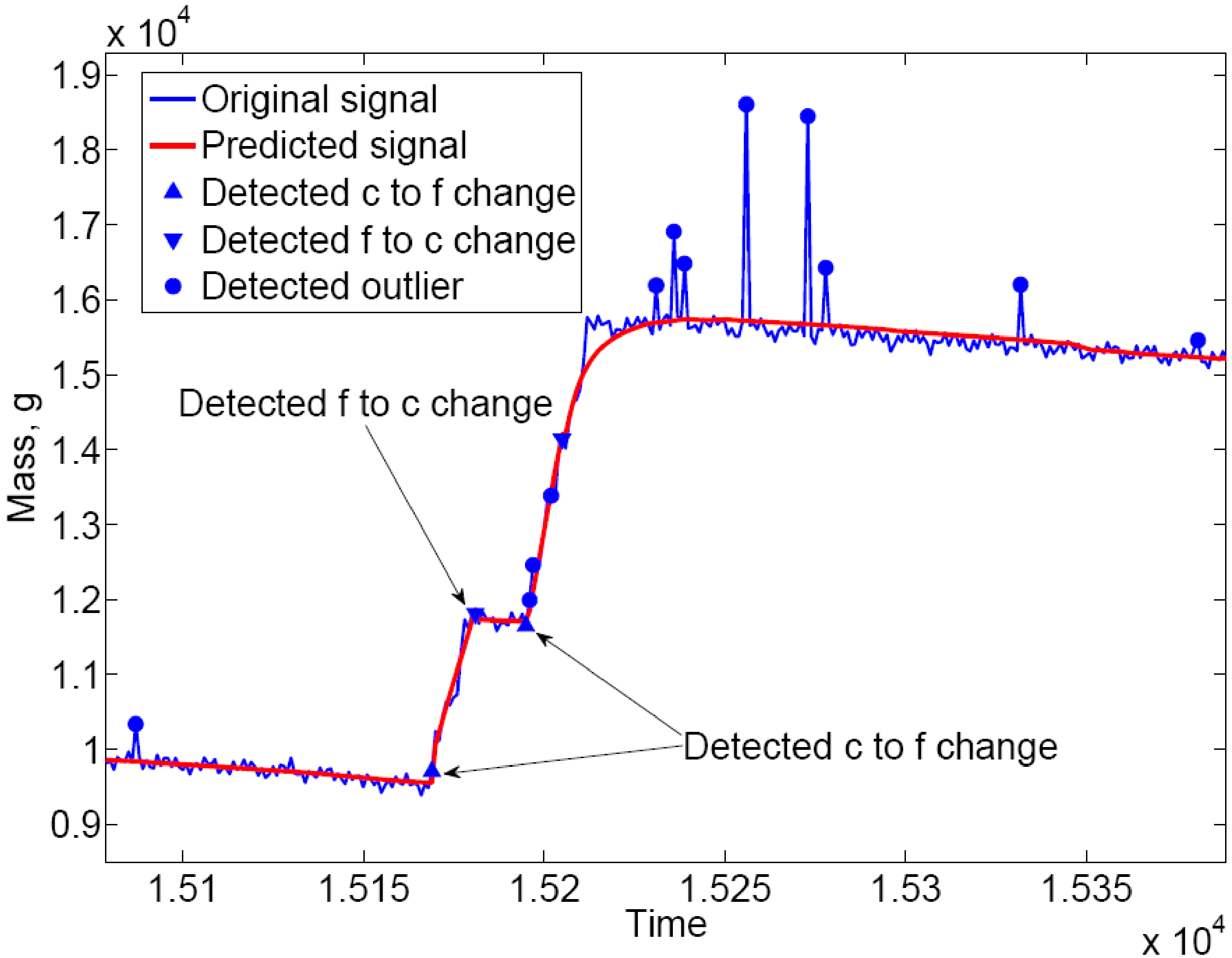
Impact of misclassifying an outlier as a change



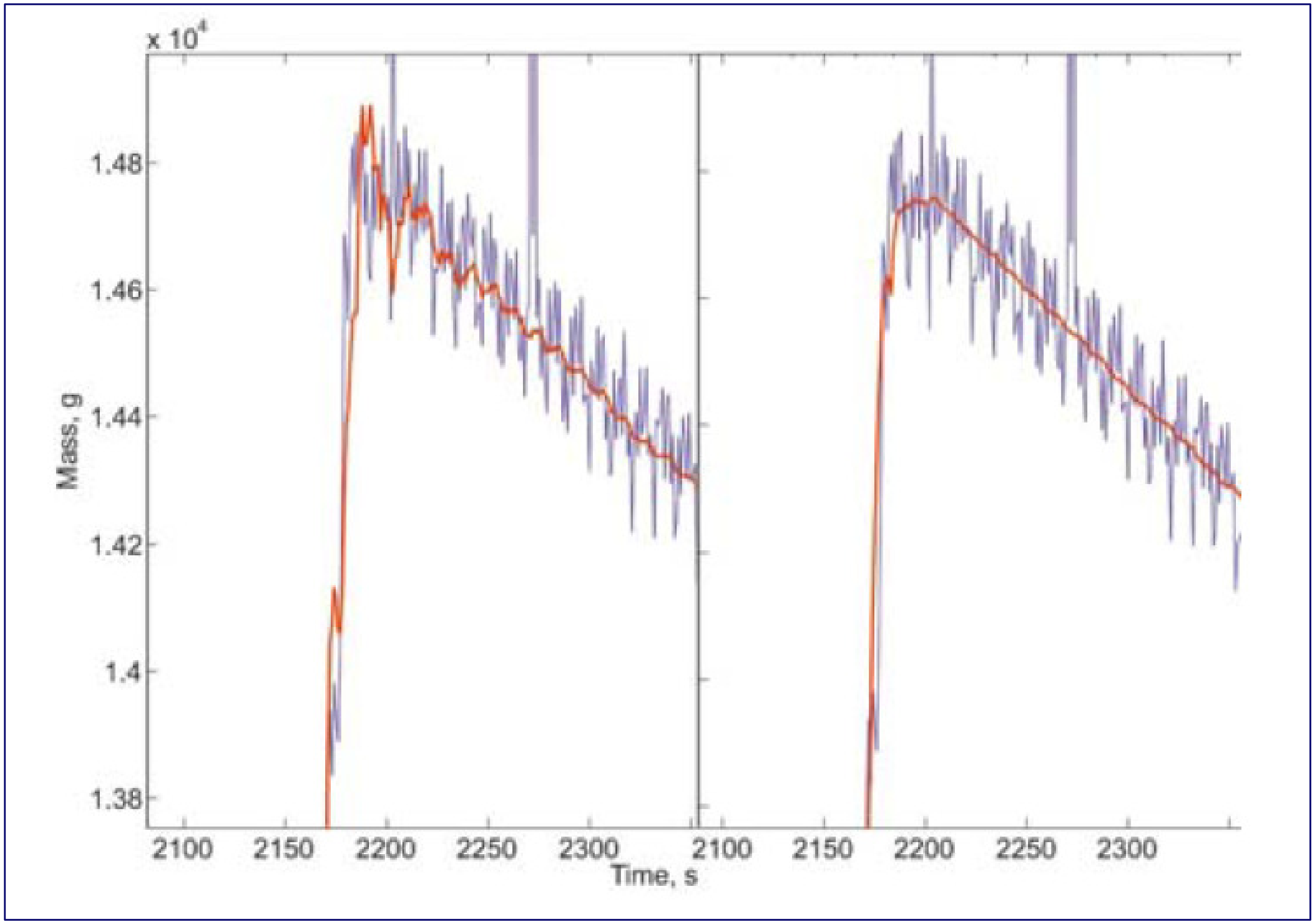
Late detection of the feeding-to-consumption transition



Detecting plateau inside feeding state



Effect of increased delay time



Zooming to the transition point with zero (left) and 20 (right) samples delay

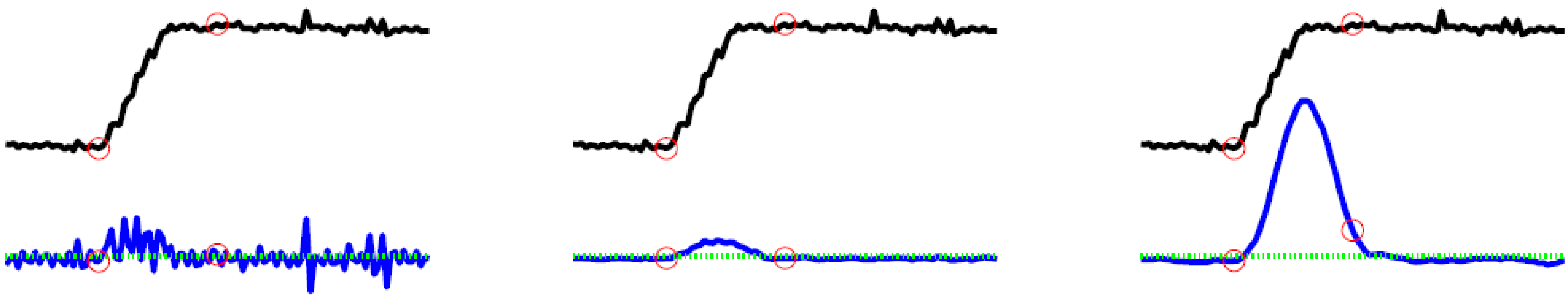
Further work

follow our recently accepted DS’09 paper

- (1) better handling noise and abrupt concept drift
  - Heuristic approach; makes domain experts even happier
- (2) focus more on accuracy of mass-flow prediction
  - not just qualitative, but also quantitative
- (3) constructing the ground truth
  - In offline settings,
    - identify all the noise and change points in the series with the best method available;
    - check the fitting by visual inspection, go back to the previous step if needed;
    - double check with the domain expert.
  - Use this approximation to the ground trutl evaluation of the online approaches

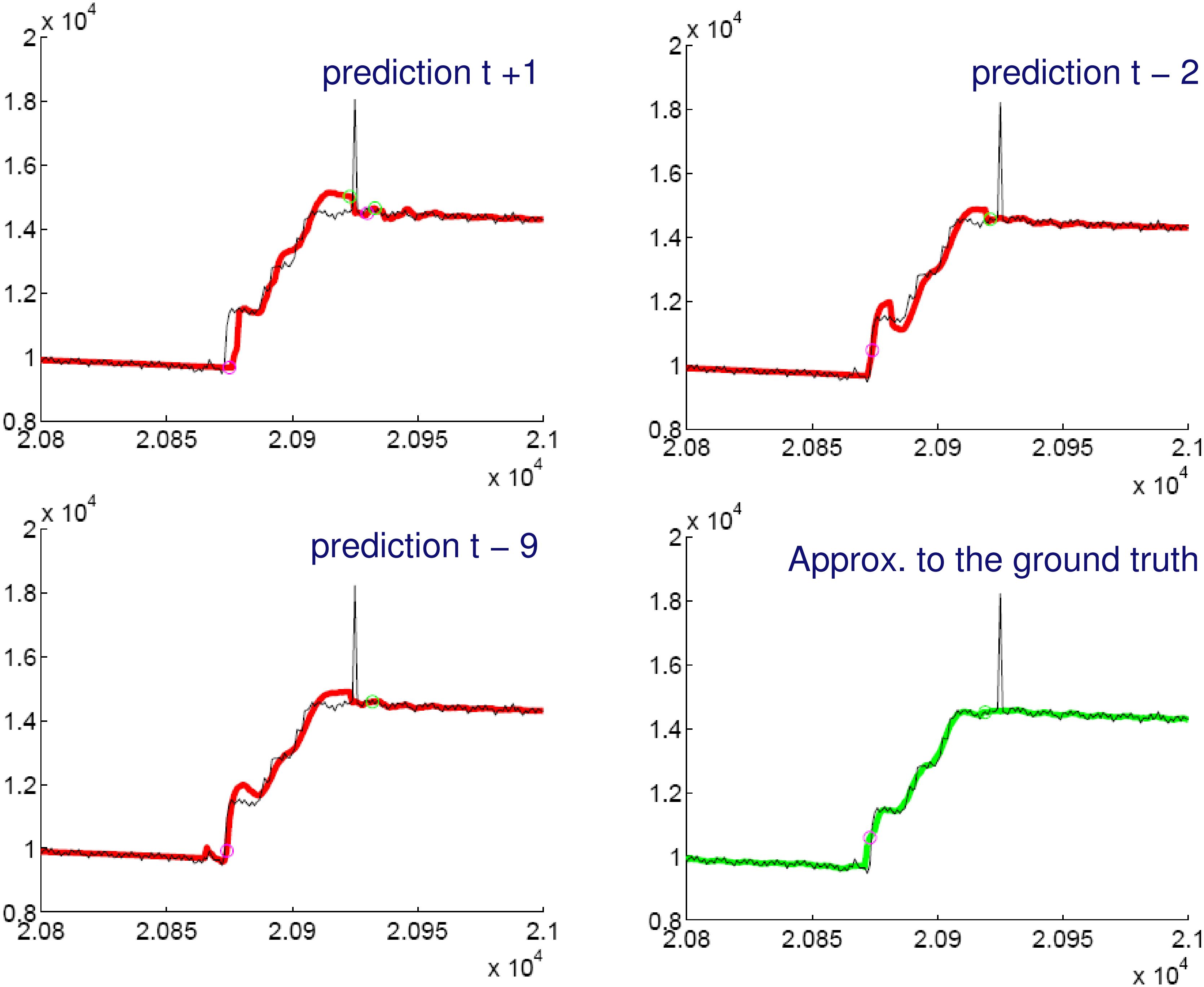
Change detection using signal differences and moving averages

$$d_t^{(L)} = x_t - x_{t-L}$$



1st order difference; no MA   1st order difference with MA   10th order diff. with MA

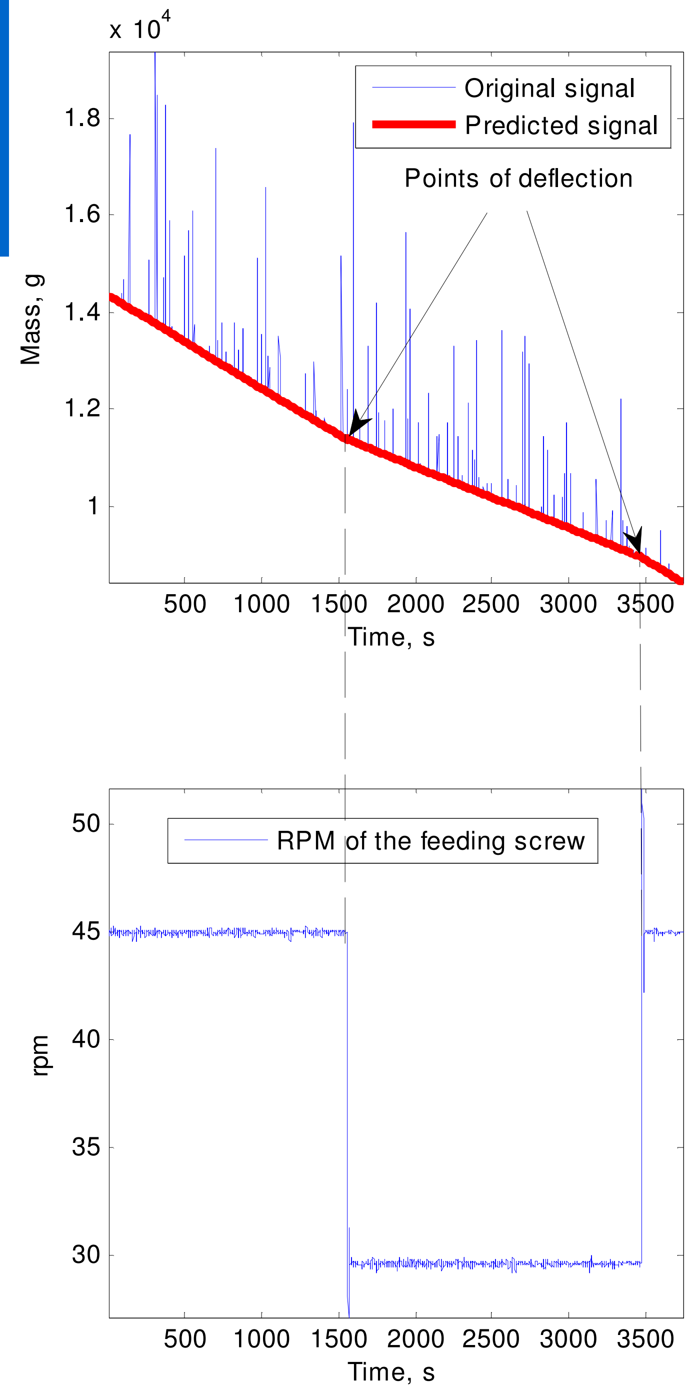
TU/e Technische Universiteit Eindhoven University of Technology


prediction t +1
prediction t – 2
prediction t – 9
Approx. to the ground truth
x 10^4
2
1.8
1.6
1.4
1.2
1
0.8
2.08
2.085
2.09
2.095
2.1

Further work

Gradual concept drift:

- use of domain knowledge and data from related sensors vs. blind detection
- changes in rpm of the feeding screw can be used for detecting the gradual changes in mass flow signal



Conclusion

- when the change points are correctly identified the predictions are reasonably accurate (delay of 5 sec)
  - can be used as a reliable indicator for the CBF control system
- Identification of change point is more accurate with the nonparametric approach than with parametric
  - It is much slower in terms of computation, but applicable
- ADWIN method is very precise in consumption-to-replenishment change detection, but the lag of detection was unacceptably large for the operational settings
  - still can be used as a reference for the “ground truth”
- Identification of the replenishment-to-consumption change points is much more difficult

Future Work

- studying in more detail the transition period at the end of the fuel feeding stage,
- considering an effect of fuel feeding speed, and the effect of using different mixtures of fuels,
- exploring the potential of back-tracking mechanisms for improved detection of state changes,
- ensemble learning ideas, and
- external validation of our online mass flow prediction approach being implemented as part of the control system of the pilot CFB boiler in operational settings with different conditions.

Acknowledgments

- Tekes
  - TEKES (Finnish Funding Agency for Technology and Innovations) DYNERGIA project
- NWO Netherlands Organisation for Scientific Research
  - NWO (The Netherlands Organization for Scientific Research) HaCDAIS project.
- VTT
  - domain experts Timo Leino and Mikko Jegoroff from VTT, Technical Research Centre of Finland.
- University of Jyväskylä

TU/e Technische Universiteit Eindhoven University of Technology

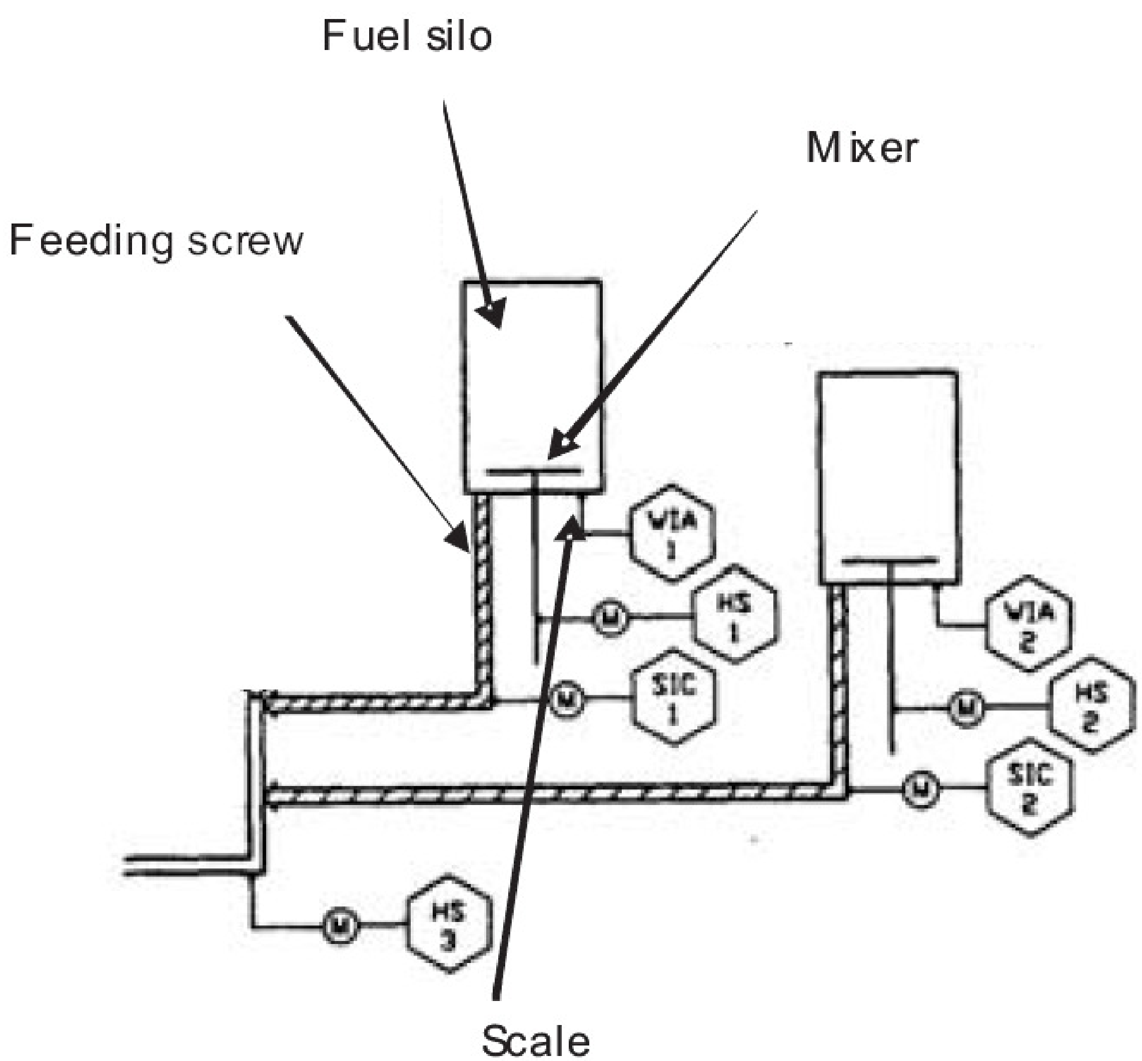
Thank you!

- Questions
- Suggestions
- Collaboration

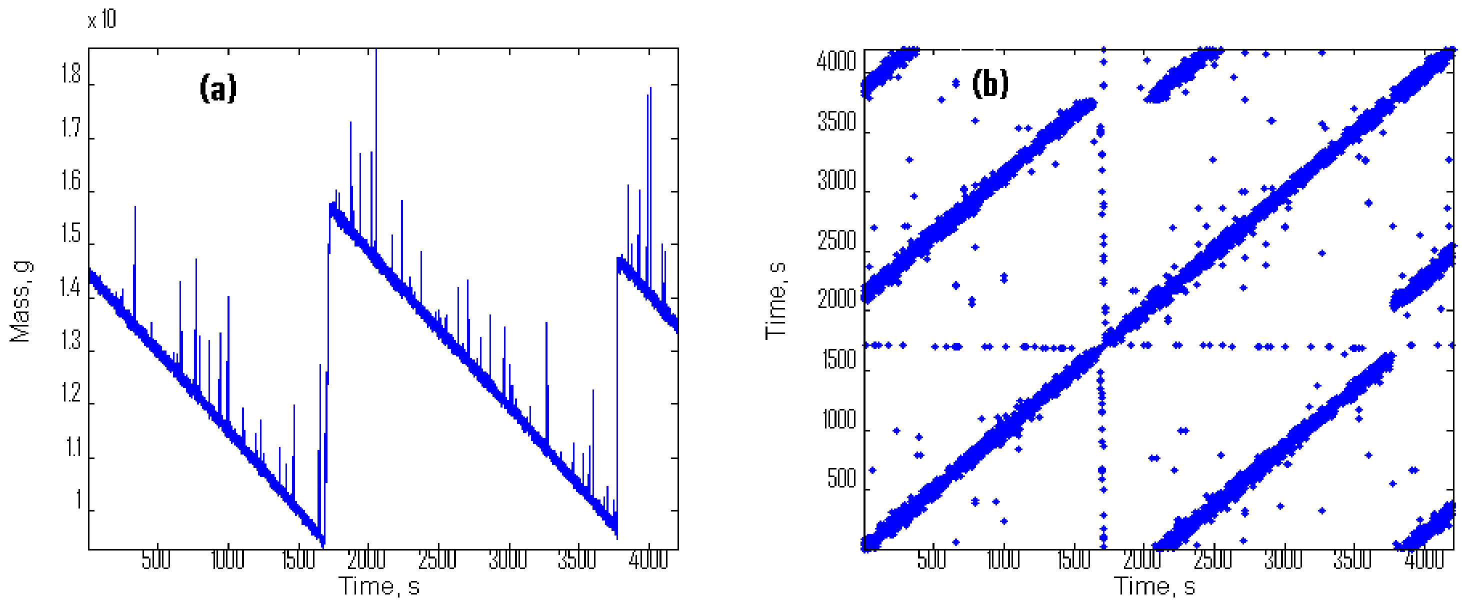
- all welcome

Extra slides

A fuel feeding system of the CFB-reactor

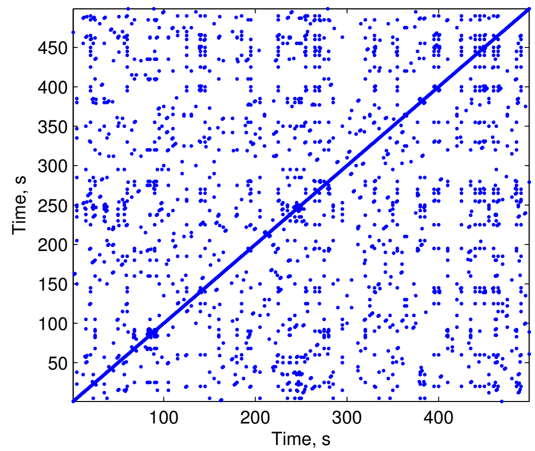


Recurrence plot of the mass flow signal

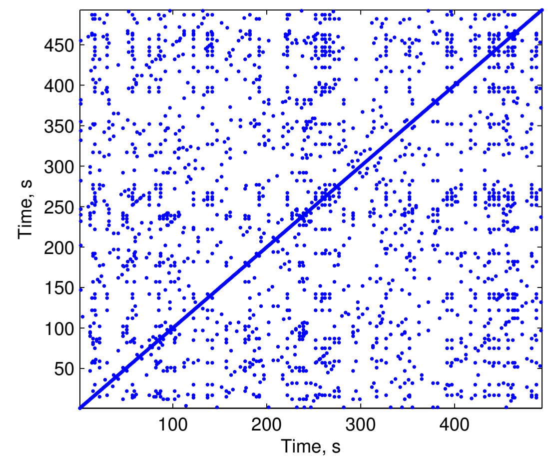


Phase space reconstruction

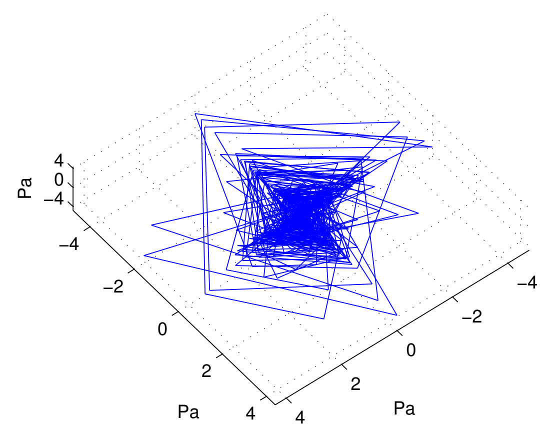
- The techniques for system reconstruction show systematic patterns in the pressure fluctuation signals
- In the figures to the right in the bottom the trajectories of the system's motion are reconstructed in 3D from a time series: (a) delay time equal to 0.3 sec and (b) delay time equal to 1.2 sec. The respective recurrence plots are also shown.

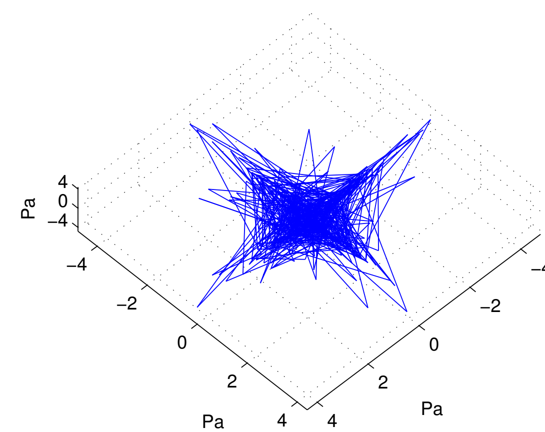


(a)

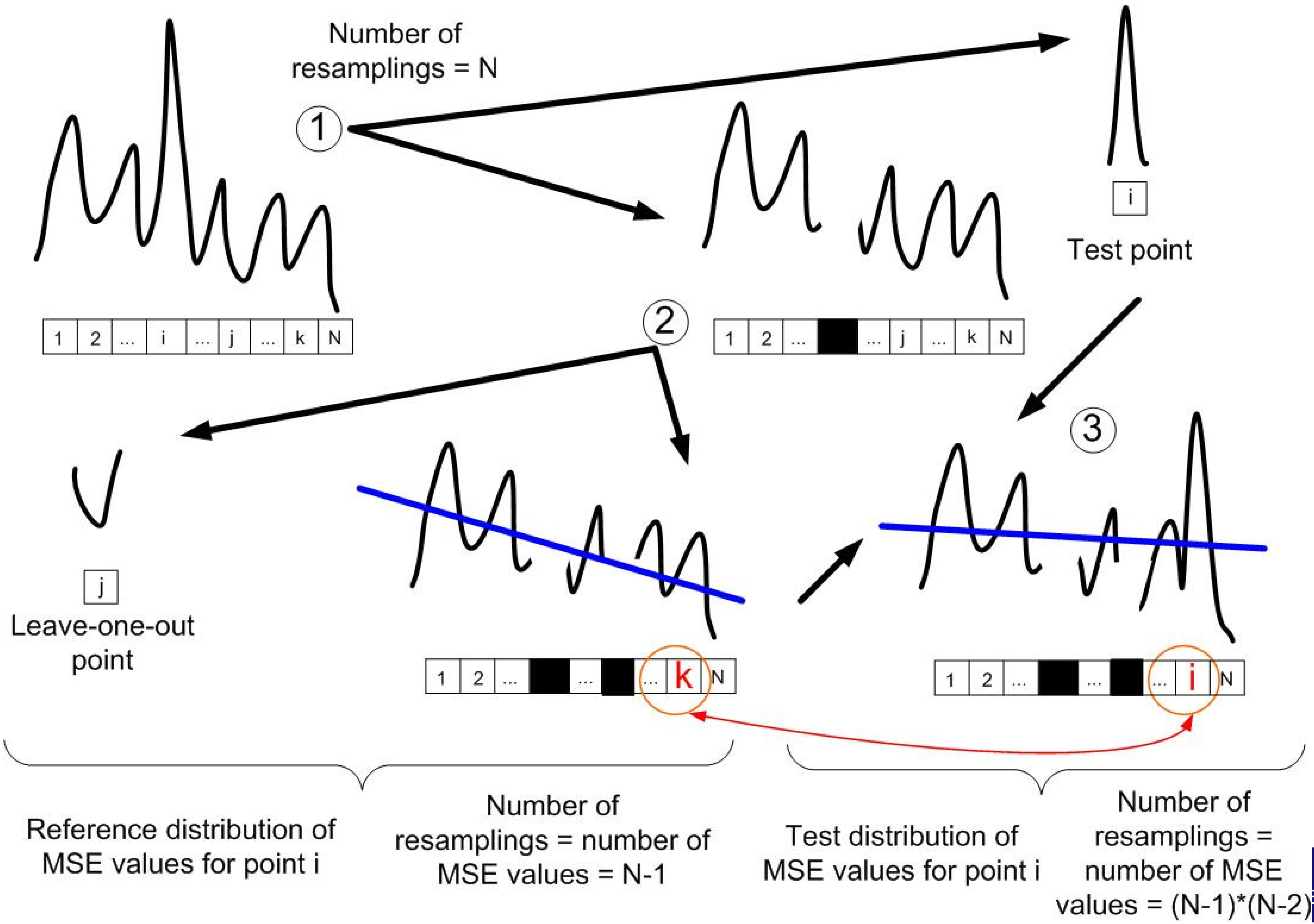


(b)





Noise and change detection with windowing approaches



Early Detection of the Feeding-to-Consumption Transition