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

J. Bakker, M. Pechenizkiy, I. Žliobaitė
Dept. of CS, TU Eindhoven
P.O. Box 513, NL-5600MB
the Netherlands
{j.bakker,m.pechenizkiy,i.zliobaite}@tue.nl

A. Ivannikov, T. Kärkkäinen
Dept. of MIT, U. Jyväskylä
P.O. Box 35, FIN-40014
Finland
aivanni@cc.jyu.fi,tka@mit.jyu.fi

ABSTRACT

In this paper we consider an application of data mining technology to the analysis of time series data from a pilot circulating fluidized bed (CFB) reactor. We focus on the problem of the online mass prediction in CFB boilers. We present a framework based on switching regression models depending on perceived changes in the data. We analyze three alternatives for change detection. Additionally, a noise canceling and a state determination and windowing mechanisms are used for improving the robustness of online prediction. We validate our ideas on real data collected from the pilot CFB boiler.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning; H.2.8 [**Database Management**]: Database applications—*Data mining*

General Terms

Experimentation, Performance, Reliability

1. INTRODUCTION

Continuous and growing increase of fluctuations in electricity consumption brings new challenges for the control systems of boilers. Conventional power generation will face high demands to ensure the security of energy supply because of increasing share of renewable energy sources like wind and solar power in power production. This can lead to frequent load changes which call for novel control concepts in order to minimize emissions and to sustain high efficiency during load changes.

From combustion point of view the main challenges for the existing boilers are caused by a wider fuel selection, increasing share of low quality and bio fuels, and co-combustion. In steady operation, combustion is affected by the disturbances in the feed-rate of the fuel and by the incomplete mixing of the fuel in the bed, which may cause changes in the burning rate, oxygen level and increase CO emissions. This

is especially important, when considering the new biomass based fuels, which have increasingly been used to replace coal. These new biofuels are often rather inhomogeneous, which can cause instabilities in the feeding. These fuels are usually also very reactive. Biomass fuels have much higher reactivity compared to coals and the knowledge of the factors affecting the combustion dynamics is important for optimum control. The knowledge of the dynamics of combustion is also important for optimizing load changes [11].

Data-mining approaches can be used to develop better understanding of underlying processes in CFB boiler, or learn a model for optimizing its efficiency [9]. For instance, the selection of process variables to optimize combustion efficiency has been studied in [12]. One typical problem, which we address in this work, is an online prediction of the system parameters. These are required for the control optimization but cannot be measured reliably in a direct way. Mass flow is one of such parameters. Currently, it is calculated offline using relatively long time period averages. Recently in [5] a method for learning regression models on mass flow prediction in CFB boilers was introduced. That approach assumed known points in time where a state transition would occur (e.g. changing from replenishment to consume state).

In this paper we propose a framework for online mass flow prediction that is based on learning and switching regression models, where the switching is regulated by an online change detection mechanism.

The paper is organized as follows. Section 2 presents a wider context of sensor data mining for developing better understanding and control of CFB reactors, and particularly the problem of obtaining mass flow signal from the CFB pilot boiler relating this problem to learning under concept drift. In Section 3 we focus on two major subtasks, learning a predictor and identifying the change points, for which we explore three alternative change detection methods. The results of the experimental study with real data collected from the pilot CFB boiler are discussed in Section 4. We conclude with a summary and discussion of the further work in Section 5.

2. RESEARCH PROBLEM

In this section we will first describe the wider context of our research and then will focus on the particular problem of online mass flow prediction.
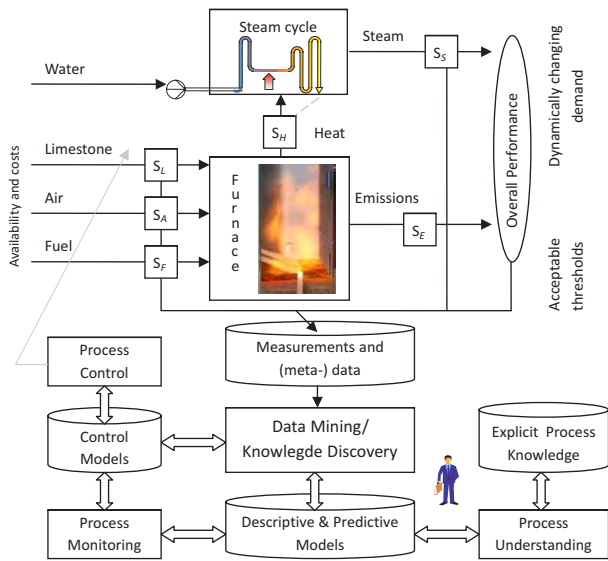
Figure 1: A simplified view of a CFB boiler operation with the data mining approach.

2.1 Control of CFB processes

The supercritical CFB combustion utilizes coal, biofuels, and multifuels in relatively clean, efficient and sustainable way. However, it needs advanced automation and control systems because of its physical characteristics (relatively small steam volume and absence of a steam drum).

The requirements for control systems are tight, since fuel, air and water mass flows are directly proportional to the power output of the boiler. That is especially relevant in CFB operation where huge amount of solid material exist in the furnace [11]. For large CFB boilers understanding of the process and the process conditions affecting heat transfer, flow dynamics, carbon burnout, hydraulic flows is critical in addition to the mechanical design.

A simplified schematics of a CFB boiler operation is presented in Figure 1. Fuel or a mixture of fuels, air and limestone are the controlled inputs to the furnace. The fuel is used for heat production. The air is added to enhance the combustion process. Limestone reduces the emission of sulfur dioxides (SO_2).

The produced heat is used to convert water into steam, which is the final output. The systems performance is monitored using the sensors S_F, S_A, S_L, S_H, S_S and S_E. The measurements are collected in database repository together with meta-data describing process conditions and used for offline and online analysis. That is where DM techniques come into play. Predictive and descriptive models that can be further utilized to facilitate process monitoring, process understanding, and process control.

2.2 Mass flow prediction

The combustion and emission performance of different types of solid fuels and their mixtures are studied at VTT's 50 kW CFB pilot boiler (see Appendix for a schematic view).

Fuel can be fed into the reactor through two separate feeding lines (Figure 2). There is a fuel screw feeder in each line at the bottom of the silo and a mixer, which prevents arching of the fuel in the silo. The fuel silos are mounted on top of the scales. The scales measure a mass flow rate of the solid fuels as a decrease in weight in time. The signal is fluctuating with constant screw feeder rotational speed. It depends on the quality of the fuel changes (e.g. moisture content, particle size), the fuel grades in the silo. Particles might jam in between the screw rotor and the stator causing a peak in the mass signal. fuel addition causes a step change in the mass signal. The mass flow decreases in between fillings in line with the level of fuel in the tank.
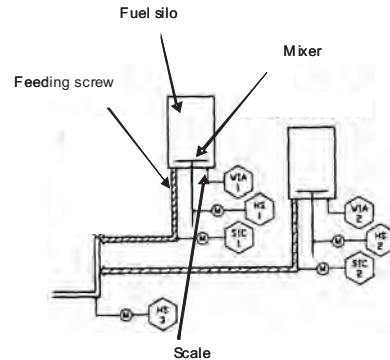

Figure 2: A fuel feeding system of the CFB-reactor.

Due to the fluctuation in the signal of the scales no reliable online data can be obtained from the mass flow of fuel to the boiler. The measurement system cannot be easily improved and therefore the mass flow is calculated offline using longer time period averages. The ultimate goal of our study is to predict the *actual* mass flow based on the sensor measurements, that would allow to improve the control system of the pilot CFB boiler.

2.3 Online mass flow prediction as learning under concept drift

In case of CFB boiler, like in many other dynamic environments, the data flows continuously and the target concept, i.e. the mass flow in this case could change over time due to the different operational processes (like fuel reloading) or changes in these processes themselves (like change of consumption speed). Therefore, learning algorithms must recognize an abrupt change mass flow (and be able to distinguish it from the outliers) and adjust a model accordingly.

In data mining and machine learning this problem is generally known as concept drift, that is the changes in the (hidden) context inducing more or less radical changes in the target concept [13].

Roughly, adaptation methods can be divided into two categories, blind methods that adapt a learner at regular intervals (e.g. with a sliding window approach) without considering whether changes have really occurred, and informed methods, that only modify learning approach after a change was detected [3]. The later group of methods must be used

14

in conjunction with a drift detection mechanism which could be implemented in several ways such as estimating statistical difference between *old* and *new* sample, or observing the deterioration of the predictor performance with more recent data point.

In this work we employ the idea of explicit concept drift detection (that corresponds to the context change in the CFB operational settings) and learning of a new model each time after a drift was detected.

There are two important types of change in the mass flow of the boiler; (1) 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). In the context of this paper we will only focus on the sudden changes.

We limit the scope of this study to identification of sudden changes, leaving the discussion of handing gradual changes as a future work.

3. OUR APPROACH

In this section we first discuss some peculiarities and domain knowledge of the problem and explain how we learn a predictor for the case when the change points are assumed to be known, and then discuss three approaches for the explicit detection of change points.

3.1 Learning the predictor

The data were recorded with 1 Hz sampling rate (an example of mass flow signal collected during a typical experimentation with the CFB boiler is presented in Figure 3). In each test of the experiment the type of fuel and/or the rotation speed of the feeding and mixing screws can be varied.

The three major sources of noise in the measurements are mixing and feeding screws and the occasional jamming of the fuel particle between the stator and rotor in the screw. The rotation of the screws causes vibrations to the system that are reflected by the scales as high frequency fluctuations around the true mass value. In Figure 4 the evolution of the frequency content of the measurements is shown by means of the short-time Fourier transform [8], from which the influence of the screws is evident. Namely, the rotating parts induce oscillations to the measurements of the same frequency as the rotation frequency. The frequency content due to the screws is identified from the figure as contrasting vertical curves.

The jamming of the fuel particle causes an abnormally large upward peak to the measurements that can be seen from Figure 3. The speed of the mass change in the tank at a given time depends not only on the rotation speed of the feeding screw and the speed of the replenishment of the fuel in the tank, but also on the amount of the fuel in the tank. The more fuel is in the tank the more fuel gets in the screw, since the weight of the fuel at the higher levels of the tank compresses (increases the density) the fuel in the lower levels and in the screw. The size and grade of fuel also have an effect on the compression rate of the fuel. Therefore, we assume that the mass flow signal has a nonzero second derivative.

The measured signal at time t can be modeled as the sum of the true mass (m_t) and measurement noise (Σ_t):

$$y_t = m_t + \Sigma_t. \qquad (1)$$

In a light of the discussed reasoning from the domain perspective, true mass at time t can be described as:

$$m_t = \frac{a \cdot (t - t_0)^2}{2} + v_0 \cdot (t - t_0) + m_0, \qquad (2)$$

where a is acceleration of the mass change, v_0 stands for the speed of the mass change at time t_0, m_0 is the initial mass at time t_0.

The noise component at time t can be described as:

$$\begin{aligned} \Sigma_t &= A \cdot \sin(\omega_{feed} \cdot (t - t_0) + \alpha_{feed}) + \\ &+ B \cdot \sin(\omega_{mix} \cdot (t - t_0) + \alpha_{mix}) + e_t, \end{aligned} \qquad (3)$$

where 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 denotes the random peaked high amplitude noise caused by the jamming of the fuel particle at time t.

One solution to predict the true mass flow signal is to use stochastic gradient descent [2] to fit the model (1) without e_t term to the measured data with the high amplitude peaks skipped from the learning process. This is closely related to fitting the model (2) to the same data in the mean-least-squares sense as the noise fluctuations are symmetric relatively to the true mass signal. Alternatively, a linear regression approach with respect to the second order polynomial can offer a better local stability and faster convergence. As the accurate mass flow measurements are required on-line by a control system the choice of the linear regression method seems more reasonable. The chosen prediction model is as follows:

$$\hat{y}_t = \frac{a \cdot (t - t_0)^2}{2} + v_0 \cdot (t - t_0) + m_0 + \epsilon_t, \qquad (4)$$

where ϵ_t summarizes all the noise components and is treated as random.

To learn a regressor, the Vandermonde matrix [4] $\mathbf{V}$, which elements $v_{i,j}$ are the powers of independent variable x, can be used. In our case the independent variable is time $x_i = t_{i-1} - t_0$, $i = 1, \ldots, T$, where T denotes the number of the time samples. If the linear regression is done for a polynomial of order n ($p^n(x) = p_n x^n + p_{n-1} x^{n-1} + \ldots + p_1 x + p_0$), $\mathbf{V}$ is computed from the observed time series of the independent variable as follows:

$$v_{i,j} = x_i^{n-j+1}, \ i = 1, \ldots, T, \ j = 1, \ldots, n+1, \qquad (5)$$

where i and j run over all time samples and powers, respectively. Provided with $\mathbf{V}$ the problem of polynomial interpolation is solved by solving the system of linear equations
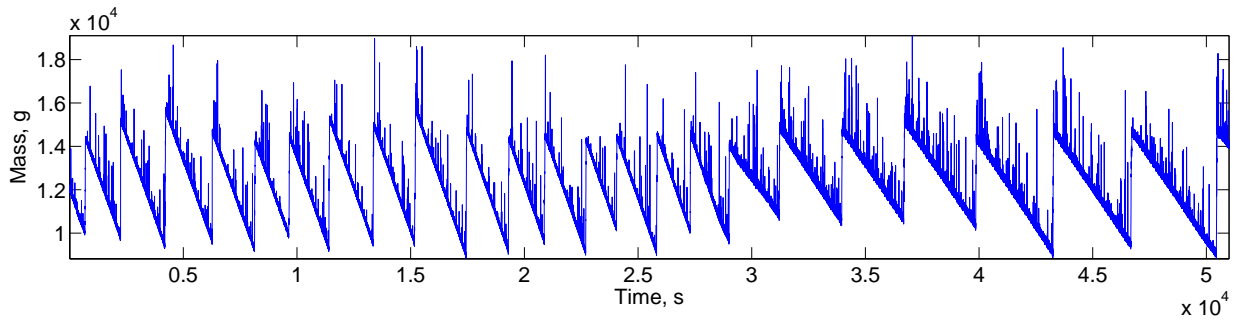
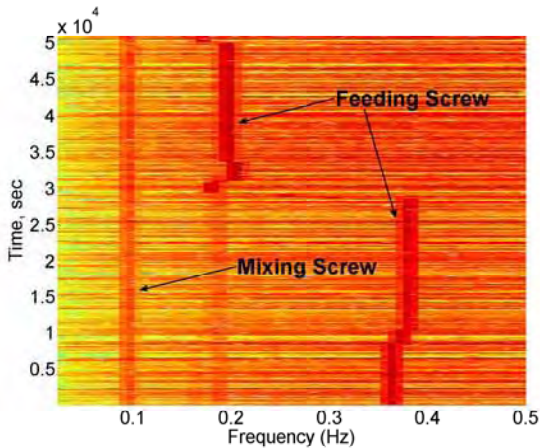Figure 3: Measurements of the fuel mass in the tank.

Figure 4: Spectrogram of the fuel mass measurements computed with Short-Time Fourier Transform. The color denotes the power spectral density.

$\mathbf{Vp} \cong \mathbf{y}$ with respect to $\mathbf{p}$ in the least square sense:

$$\widehat{\mathbf{p}} = \operatorname{argmin}_{\mathbf{p}} \sum_{i=1}^{T} (\sum_{j=1}^{n+1} V_{i,j} p_{n-j+1} - y_i)^2 \qquad (6)$$

Here, $\mathbf{p} = [p_n\ p_{n-1}\ \ldots\ p_1\ p_0]^{\mathrm{T}}$ denotes the vector of the coefficients of the polynomial, and $\mathbf{y} = [y(x_1)\ y(x_2)\ \ldots y(x_T)]^{\mathrm{T}} = [y_1\ y_2\ \ldots y_T]^{\mathrm{T}}$ is the time series of the dependent variable that is indication of the scales. Provided that the $n+1$ columns of the matrix $\mathbf{V}$ are linearly independent, this minimization problem has a unique solution given by solving the normal equation [6]:

$$(\mathbf{V}^{\mathrm{T}}\mathbf{V})\widehat{\mathbf{p}} = \mathbf{V}^{\mathrm{T}}\mathbf{y}. \qquad (7)$$

For the modeling of the mass flow we have chosen second order polynomial, where $\widehat{p}_0 = \widehat{m}_0$, $\widehat{p}_1 = \widehat{v}_0$, $\widehat{p}_2 = \widehat{a}$. We distinguish the two types of the periods in the experiment: the consumption (fuel is only consumed) and the consumption with fuel replenishment.

When one period of the CFB operating changes to another (i.e. a new portion of fuel is being added) the process of mass flow prediction starts over again, as the model of the mass flow changes. Thus, the most problematic unstable regions are the transitions intervals, when the parameters of the model change their values.

When a period of the session starts the samples of measurements start to accumulate in the buffer. The data in the buffer are used to fix the parameters of the mass flow prediction model. Only the samples that do not contain high amplitude peak (due to jamming) are placed to the buffer. The buffer is emptied after each change point (that assumed to be known in this setting) and starts to accumulate new data.

The first measurement within the experiment is taken as the prediction of the mass flow signal at that time $\widehat{m}_1 = y_1$ and as the first point that is placed to the buffer. In contrast, the first prediction of the mass within a following period is taken as the last prediction of the mass from the previous period. In addition, the last predicted point from the previous period $\widehat{m}_{i_c}$ is placed to the buffer as the first point for the new period, where i_c denotes the number of the sample when the change of the periods occurs. When a new sample arrives the parameters of the model are estimated based on the points that are in the buffer independently of whether the current point was placed to the buffer or not. The current prediction is computed respectively based on the current model. Depending on the number of data points T in the buffer different models are used:

- if $T = 1$, then $\widehat{m}_i = \widehat{m}_{i-1}$;

- if $1 < T \le 4$, then $\widehat{p}_2 = 0$ and $\widehat{p}_0$ and $\widehat{p}_1$ of the first order polynomial are estimated from the available data, and $\widehat{m}_i = \widehat{p}^1(x_i)$, where $\widehat{p}^1(x_i)$ is the current prediction by the first order polynomial.

- if $T > 4$, then the second order polynomial model is fitted to the data, and $\widehat{m}_i = \widehat{p}^2(x_i)$.

In practice the operational settings often allow a delay between the data arrival and the evaluation of the signal of interest at this time sample. This means that the estimate of the signal at a given time sample is obtained based on the data that is accumulated also during the future times. This allows the more accurate estimates of the signal to be computed. Note that in our case this will have an effect of

16

increased accuracy mainly for the beginning of a new experiment, when the amount of the data in the buffer is small yet.

3.2 Explicit change detection methods

In this section we consider three different approaches to detect change points and thus facilitate modeling the transitions from one state of the system to another. All three are based on measuring statistical data in a moving window. The first approach is nonparametric and is based on the Mann-Whitney U test. The second approach is based on a parametric test on the performance of the local models. The third approach is based on the statistical analysis of the raw data.

3.2.1 Nonparametric change detection

Let's assume that when the underlying process in the boiler changes, the current model for the prediction of the data will perform worse. The idea behind nonparametric change test is to compare the prediction performances obtained on different subsets of the data. For the classification of whether an observed point is outlier (peak due to a fuel particle jamming) or not, the statistical nonparametric Mann-Whitney U test [7] that compares the two distributions of MSE values can be used. Namely, if a point is to be tested for being an outlier with respect to a window of previous points or, in general, with respect to a set of surrounding points, the following procedure is applied to construct the two samples of MSE values:

- First, a leave-one-out cross validation (LOOCV) is done on a reference set of $N-1$ points, a line is fitted to each of the possible subsets of size $N-2$, and MSE of the fit is computed for fitted $N-2$ points only. Thus, at this stage we obtained a sample of $N-1$ MSE values that characterizes the set of reference points. We call the distribution of these MSE values a reference distribution for the tested point.

- Then, the test distribution composed of $(N-1)(N-2)$ MSE values is computed by replacing each point in each subset of size $N-2$ from previous stage for a test point, computing a linear fit and MSE value for the $N-2$ points used to compute the fit. Therefore, test distribution contains MSE values for a linear fit, when a test point is used. This way we ensure that the fittings and MSE computations are always done using $N-2$ points, but the sizes of reference and test samples are different.

The above considerations are demonstrated in Figure 5 in the second and third steps. In fact, the first step is just a LOOCV, when each point in the initial window of size N is made a test point, and the rest form reference set.

Note that in Figure 5 a less general case is shown compared to a discussion. The illustration considers a set of neighboring points, although in general the points in the set can be drawn from distanced time locations depending on the pursued goal. By changing the size and spread of the reference set, one controls how the decision made using the Mann-Whitney U statistics is tailored/diversified over time.
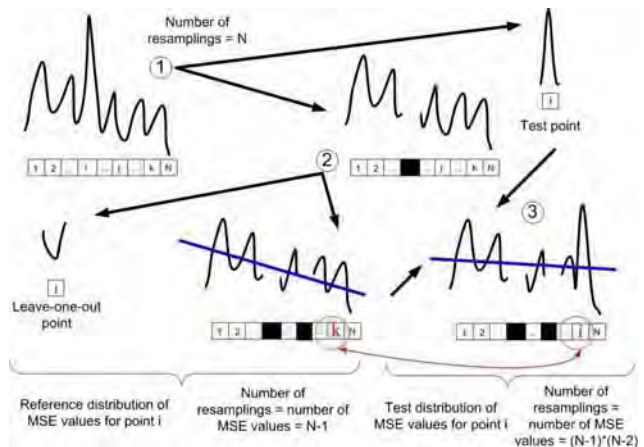
Figure 5: The resampling procedure for constructing two distributions to be compared by Mann-Whitney U test.

In other words, varying the window size adjusts the context for decision making to be more local, global, or balanced.

The Mann-Whitney U test applied to the reference and test distributions for a given confidence level α shows whether MSE of the fitted points including test point is systematically different from the case when the test point is excluded. If such tendency is detected by the test, the test point is classified as outlier.

Based on the process of identification of the outlier, we define a criterion for state change detection as the fact of observing several outliers in a row ($o2c$).

In our experiments we used $\alpha = 10^{-9}$ (confidence level of Mann-Whitney U test); $d = 10$ points as delay after which the prediction for the point appeared d points prior to the current point must be output, $r = 16$ points as the maximal size of the reference window (composed of the most recent precedents to the currently considered point) against which the newly arrived point is compared using outlier test, and $r_a = max\{d, r\}$ is the actual reference window size; $f = 150$ points as the maximal size of the window that is used to compute the model parameters to predict the current output, and $f_a = max\{d, f\}$ is the actual size of the fit window; finally $o2c = 5$ is a (currently predetermined) number of detected outliers in a row that is a sufficient condition to assume the state change.

The method operates as follows. Until the first $d+1$ points are accumulated from the start of the new state nothing is output. In the following all considerations are assumed to be referenced to a start of the current state. For simplicity we assume that we start in the fuel consumption state.

When first d points are accumulated, an outlier test is made for each of the d accumulated points. All found outlier points are substituted with an interpolation (by fitting a line to the non-outlier points). After this, when a new point arrives, it is tested for outlier with respect to r_a previous points. If it is

17

classified as outlier, it is replaced by an extrapolation from the r_a previous points. Then the point 'delayed' by d points from the current one is predicted by the linear fit to the latest sequence of outlier-corrected f_a points. If $o2c$ outliers in a row are noticed, the state change (pointing $o2c$ points back in time) is alarmed and the detected 'false outliers' (that actually belong to the change) are restored (from the backup buffer). The process continues then from the beginning of the newly detected phase.

3.2.2 Parametric change detection

The nonparametric approach is relatively time consuming, thus we look for a more efficient change detectors.

When the underlying process in the boiler changes, the current model for the prediction of the data will perform worse. This behavior can be detected by keeping track of two statistical properties of the performance [3]. The first is the error rate that signifies the probability of miss classifying the actual value y_t of the signal. The second is the standard deviation of the error rate. In [3] it is assumed that the classification task can be modeled by a binomial distribution. Since the mass flow is continuous, we have to assume that given a large enough sample or window, the binomial distribution is close to a normal distribution with the same mean and standard deviation. This is used in an online test to see whether the signal has changed.

Since this task is in continuous space, we use the Mean-Squared-Error (MSE) as metric for change detection instead of the error rate. For each time step the window is moved and for point x_t all local reference MSEs are calculated with LOOCV.

We assume that the model prediction performance is stable if the local found reference MSE (E_t) satisfies

$$E_t + S_t < E_{min} + \alpha S_{min}, \qquad (8)$$

where E_{min} is the minimal found reference MSE, S_{min} the minimal found reference standard deviation, and α a parameter that reflects the level of confidence. If there is a large enough deviation in the signal the algorithm will report a warning. This level is determined by the condition:

$$E_{min} + \alpha S_{min} < E_t + S_t < E_{min} + \beta S_{min}, \qquad (9)$$

where β is the upper confidence bound signifying a change. Since it might occur that there is a local change in the signal (an outlier) it is not possible to go the change state immediately. If $E_t + S_t > E_{min} + \beta S_{min}$ the algorithm reports a change and it will switch to the initial state of the other process. When this happens E_{min} and S_{min} are reset to the minimal found values in the new regime.

This procedure puts a strict lower bound on the window size. Since we have to assume that the normal distribution is representative for the distribution of reference MSEs, the window size should be at least 30 consecutive points. In principle, a small window size is preferable when trying to detect rapid changes. But a smaller window size will result in a higher variation in the local models. In our experiments we used $\alpha = 2$, $\beta = 3$ (lower and upper bounds for the confidence interval), and the window size of 30 points.

For each 30 accumulated points, the LOOCV is used. For each $N - 1$ local model the MSE is calculated and it's standard deviation. The transition conditions are checked and when an outlier is detected it is ignored by the global fit. If, after this, the algorithm detects a change, the boundaries for the transition criteria are reset and the global fit is relearnt for the future points.

3.2.3 Change detection using raw data

To reduce the degrees of freedom introduced by MSE based change detectors, we employ a change detection method based on raw data. We choose ADWIN, which was presented in [1]. It is based on the differences between the means of raw data. The method was originally designed for univariate sequential data. The method works as follows: given a sequence of signals it checks whether there are statistically significant differences between the means of each possible split of the sequence. If statistically significant difference is found, the oldest portion of the data backwards from the detected point is dropped and the splitting procedure is repeated until there are no significant differences in any possible split of the sequence. More formally, suppose m_1 and m_2 are the means of the two subsequences as a result of a split. Then the criterion for a change detection is $|m_1 - m_2| > \epsilon_{cut}$, where

$$\epsilon_{cut} = \sqrt{\frac{1}{2m} \log \frac{4n}{\delta}}, \qquad (10)$$

$$m = \frac{1}{\frac{1}{n_1} + \frac{1}{n_2}}, \qquad (11)$$

here n is total size of the sequence, while n_1 and n_2 are sizes of the subsequences respectively. Note that $n = n_1 + n_2$. $\delta \in (0, 1)$ is a hyperparameter of the model. In our experiments we used $\delta = 0.3$, $n = 200$.

4. EXPERIMENTAL RESULTS

In this Section we first present the performance results under the assumption of known change points and then demonstrate the accuracy of the considered change detection techniques.

4.1 Predictor performance under the assumption of known change points

As the main points of interest are the points, where the change of the period occurs, for this analysis we took an interval of the experiment containing both types of the transition points (see Figure 6).

The computations of mass flow were done for the delay time varying from 0 to 20 samples/seconds with step 1.

It takes about $100 - 150$ seconds from the beginning of each consumption period for the model of the mass signal stabilizes. Since the replenishment period lasts during shorter times ($3 - 40$ seconds), the model may exhibit instability even at the end of the period. However, overall the prediction of the mass during the replenishment period is satisfactory, because changes of the mass are rather steep and prominent against the background of the noise. For comparison, the effect of increased delay time is shown in Figure 7.

18

Table 1: Accuracy of detecting sudden changes in feeding (ϕ) and consumption (κ) processes, and of detecting outliers (o).

				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	-	-

Figure 7: Zooming to the transition point with zero (left) and 20 (right) samples delay.

To indicate the rate of convergence of the model with respect to the delay time we computed the mean-square error between the true mass prediction for each pair of the consecutive delays (see Figure 8):
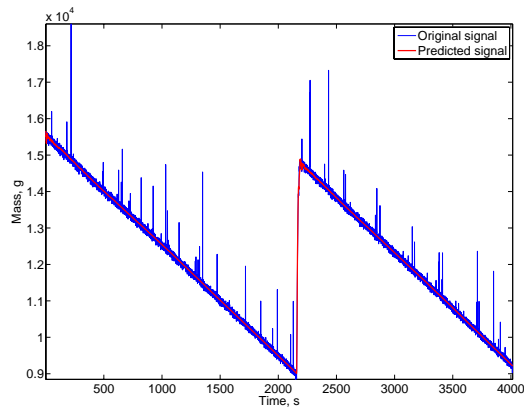
$$MSE(\tau) = \frac{1}{T} \sum_{i=1}^{T} (\widehat{m}_i^{\tau} - \widehat{m}_i^{\tau-1})^2. \qquad (12)$$

It can be clearly seen from the figure that for the small delays ($1 - 4$ samples) the prediction accuracy improves dramatically, and for the larger delays the improvement slows down.
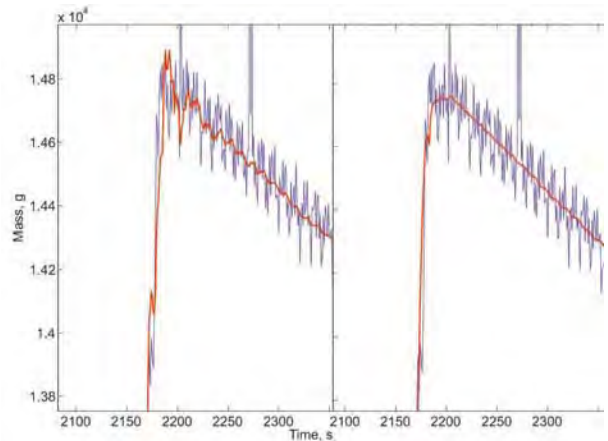
4.2 Performance of change detection methods

The results are summarized in Table 1. For each method we present confusion matrixes of detecting sudden changes in feeding (ϕ) and consumption (κ) processes, and of detecting outliers (o).
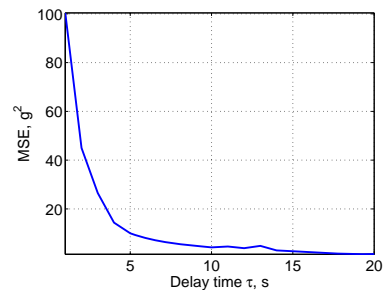
The parametric test succeeds in identifying 19 out of 24 changes from feeding state to consumption state. However, it finds only 4 out of 24 transitions the other way

Figure 6: Online mass flow prediction with the switching regression models and zero delay time.

Figure 8: Mean-square error between the mass signal predictions for each pair of the successive delays.

around. The worst part of the results is the number of falsely identified changes(74 for feed-to-consumption and 89 for consumption-to-feed). At these points the algorithm will signal a change where in fact there is none. This results in a very poor fit as demonstrated with an example in Figure 9.

The number of unidentified outliers is also important, since it will greatly affect the performance of the model if this number is high. For the parametric test this number is 217 out of 1016. This is roughly 20% of the total number of outliers. This number is too high for the algorithm to be
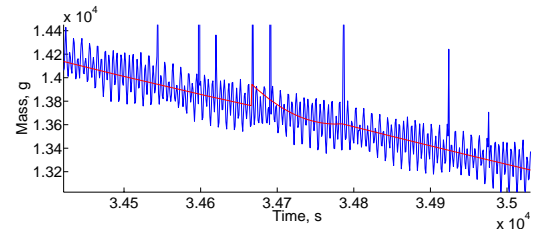

Figure 9: An example of impact of misclassifying an outlier as a change (red line).

19

effective. The relatively bad results might be caused by the fact that the assumption that the local MSEs are normally distributed is not valid.

The Mann-Whitney U-test does not assume a normal distribution and its accuracy is noticeably better. It enables finding all of the consumption-to-feed transitions. However, this test also incorrectly identifies 20 (false) change points for this transition. The other type of change is recognized 8 times out of 24 and the number of false classifications is also high. The number of outliers detected is comparable to that of the parametric test.

The most intolerable and harmful for prediction detection errors are related to early (see Figure 10) and late (see Figure 11) determination of the feeding-to-consumption transition. If the early or late detection happens, not only the predictions of the transition period itself deteriorate, but also a large portion of the following or prior to the detection event consumption interval lacks accurate interpolation, because points from feeding are also used to compute fit for consumption. Short perturbations of signal during the feeding state may lead to early false detection of the transition, because the outliers may be misinterpreted as indication of the change (see Figure 10).

At the same time the recognition of the fact that feeding continues may be problematic, since no points of appropriate number and properties are observed in the following window. The current implementation does not take into account prior signal once state change is detected. This has an effect of resetting the context. This way only one level of verification if the change is true is used: observing relatively large numbers of outliers in a row. But additional level of verification can be advised by continuing back-tracking, while new points arrive. In other words, we may wish to check if the situation is really what it appears to be also with respect to the following part of the signal, and whether the change was real. The late determination of the feeding-to-consumption transition may be a consequence of a too short feeding phase that is perceived by the method as outlier (see Figure 11).
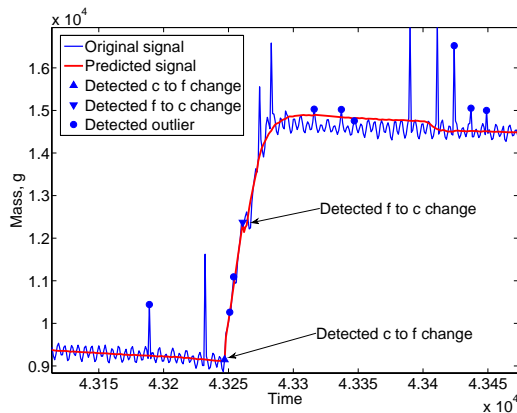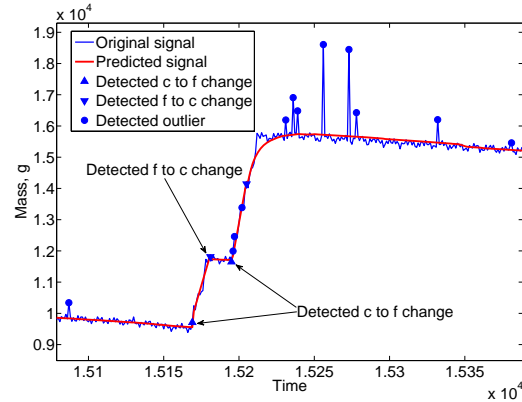

Figure 10: An example of early detection of the feeding-to-consumption transition.

The reasons for delayed transition detection considered above
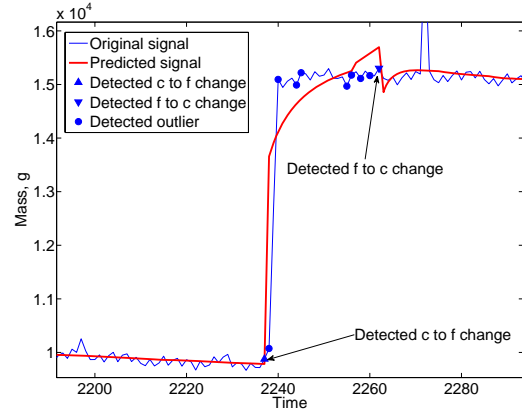
Figure 11: An example of late detection of the feeding-to-consumption transition.

Figure 12: Detecting plateau inside feeding state.

are well understood. The errors happen due to a locally tuned criteria for outlier and change detection. However, the more sophisticated windowing schemes can be used. For example, an adaptive window size approach can be employed (however, this requires further study of the data patterns observed in the data). Besides, a simple back-tracking mechanism can be used to verify the already made decision with respect to subsequently arrived points.

It is worth noticing that not always outlier behavior can be unambiguously distinguished from the true behavior of the signal. For example, in Figure 12 the plateau inside the feeding phase was successfully identified. In this case it is clear that feeding was stopped for a while and than continued again. However, if such plateau is small it is generally unknown if this is a false measurement or consequence of real process of mass flow. In our computations, the small plateaus inside feeding phase sometimes can be ignored by a detection procedure or lead to an early detection of the transition to consumption with or without following detection of continuing feeding.

The ADWIN method required on average a lag of 40.7

20

seconds to detect the start of the feeding stage. Given the fact that the average length of this stage (in our data) is 26.0 seconds the detection in most cases occurred after it was already finished. The shortage of the feed stage explains the fact that there were no detections of the consumption stage using this method.

From the careful visual inspection it was possible to conclude that the method detected all the feed stages and made no false alarms (no outliers were mistreated for the change). The method seems to be slow in reaction but reliable and can be of a particular use for verification of the changes, since there are no "ground truth" available.

Note that although the change is detected with a lag, but the actual signaled change point is quite close to the original (on average 7.5 seconds away from the "true" change points). The numbers in Table 1 are based on the detected change points, not the detection time.

Computational complexity of the considered change detection methods. Assuming that at time t we have a history of t datapoints available, but we delimit the back search to N past data points. Both parametric and nonparametric change detection algorithms employed use LOOCV procedure, which requires multiple scans to build local prediction models to compute MSEs. The parametric model would require $\sim N^2$ iterations to output the detection decision at time t, but once the records for MSE are kept, the subsequent detections would require only $\sim N$ iterations per time point. Nonparametric change detection involves comparison between the two distributions, thus it requires $\sim N^3$ iterations per time point. Change detection using raw data requires N splits to test for a change at time t.

The methods were implemented in MATLAB and the performance time of all the methods is reasonable for online operation. We have a series consisting of ~ 50000 entries, which correspond to seconds. That gives a restriction for implementation to be reasonable for online operation if all the experiment is completed in less than 14 hours. The factual performance in our implementation was $3 - 4$ times more quick.

5. CONCLUSIONS AND FURTHER WORK

Prediction of mass flow in CFB boilers in online settings is an important and challenging problem having clear connections to the problem of learning under the presence of sudden and gradual concept drifts.

In this paper we presented a regression learning framework for fuel mass prediction. Change and outlier detection is the key component in this framework, it regulates switching of the predictors. We studied three alternatives of the explicit detectors.

Our results demonstrated that (1) when the change points are correctly identified our predictors perform reasonably well and a delay of less than 5 seconds allows to predict the mass flow accurately enough to be used as a reliable indicator for the CBF control system, (2) when the change points need to be determined online without any additional input, the nonparametric approach shows much better performance

than parametric one in terms of accuracy; however it is much slower and therefore requires more computational power to guarantee in time prediction, (3)The ADWIN method was the very precise in consumption-to-replenishment change detection, while the lag of detection was quite large and therefore this method can be used only as a reference for the "ground truth", but is not applicable for the operational settings of the CFB boiler.

The directions of our further work in mining CFB sensor data include (1) studying in more detail the transition period at the end of the fuel feeding stage, (2) considering an effect of fuel feeding speed, and the effect of using different mixtures of fuels, (3) exploring the potential of adaptive context window size approach and back-tracking mechanisms for improved detection of state changes, and that is not least important (4) 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.

We anticipate that local reliability estimation of mass flow prediction [10] may be helpful for the domain experts. Furthermore, our experimental study also suggests that learning from multiple experiments and better utilizing additional relevant information available from various sensors (e.g., rotation speed of the feeding screw) should bring a further improvement of the reliability of online mass flow prediction.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] A. Bifet and R. Gavaldà. Learning from time-changing data with adaptive windowing. In *SDM*. SIAM, 2007.

[2] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification (2nd Edition)*. Wiley-Interscience, 2001.

[3] J. Gama and G. Castillo. Learning with local drift detection. In X. Li, O. R. Zaïane, and Z. Li, editors, *ADMA*, volume 4093 of *Lecture Notes in Computer Science*, pages 42–55. Springer, 2006.

[4] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, 1991.

[5] A. Ivannikov, M. Pechenizkiy, J. Bakker, T. Leino, M. Jegoroff, T. Karkkainen, and S. Ayramo. Online mass flow prediction in cbf boilers. In *Proc. 9th Industrial Conference on Data Mining (ICDMŠ09)*. Springer-Verlag, 2009.

[6] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems*. Prentice-Hall, Englewood Cliffs, NJ, 1974.

[7] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Annals of Mathematical Statistics*, 18:50–60, 1947.

[8] S. Nawab and T. Quatieri. Short time fourier

transform. In J. Lim and A. Oppenheim, editors, *Advanced topics in signal processing*, pages 289–337. Prentice Hall, 1988.

[9] M. Pechenizkiy, A. Tourunen, T. Kärkkäinen, A. Ivannikov, and H. Nevalainen. Towards better understanding of circulating fluidized bed boilers: a data mining approach. In *Proceedings ECML/PKDD Workshop on Practical Data Mining*, pages 80–83, 2006.

[10] P. P. Rodrigues, J. Gama, and Z. Bosnic. Online reliability estimates for individual predictions in data streams. In *ICDM Workshops*, pages 36–45. IEEE Computer Society, 2008.

[11] J. Saastamoinen. Modelling of dynamics of combustion of biomass in fluidized beds. *Thermal Science*, 8(2):107–126, 2004.

[12] Z. Song and A. Kusiak. Constraint-based control of boiler efficiency: A data-mining approach. *IEEE Trans. Industrial Informatics*, 3(1):73–83, 2007.

[13] G. Widmer and M. Kubat. Learning in the presence of concept drift and hidden contexts. *Mach. Learn.*, 23(1):69–101, 1996.

Appendix. The laboratory scale CFB-reactor. The height of the riser of the boiler is 8 m and the inner diameter 167 mm. The reactor is equipped with several separately controlled electrically heated and water/air cooled zones in order to control the process conditions, for example, oxygen level, temperature and load almost independently. Several ports for gas and solid material sampling are located in the freeboard area.

22