Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation

RESEARCH REPORT

Kamil Dedecius*, Ladislav Jirsa

Impact of forgetting on models of rolling mills

No. 2283 September 22, 2010

ÚTIA AV ČR, P.O.Box 18, 182 08 Prague, Czech Republic
Tel: +420 286892337, Fax: +420 266052068, Url: http://www.utia.cas.cz,
E-mail: *dedecius@utia.cas.cz

1 Introduction

This report deals with modelling of the cold sheet rolling process. It describes several models proposed for this purpose, one of them being based directly on related physical principle, one being a physical approximation, using the rolling force and two models are so called 'blackbox models', i.e., models with user-selected structure.

As one can expect that the models' parameters vary in time, we estimate them with exponential and partial forgetting; estimation without forgetting is used for comparison. While the first forgetting method is known as time weighted least squares (TWLS) [3] or as flattening of the posterior probability density function (pdf) [7], and is well established in the theory of estimation of time-varying parameters, hence the most popular, the other method is a recently developed approach to the issue [1]. The partial forgetting tries to solve the main drawbacks of the exponential forgetting, i.e., use of the same forgetting factor for all parameters and the liability to covariance blow-up, a situation, when the gain of the estimation algorithm grows without bounds for non-persistently exciting signals [6]. The drawback of the latter method consists in its higher complexity, imposing certain needs on computing device.

None of the forgetting methods used a predefined alternative pdf. If such a pdf were at hand, the estimation would (logically) be significantly improved. However, the purpose of this report is to demonstrate the pros and cons of both the methods for modelling of the rolling process.

2 Bayesian modelling with forgetting

We employ the Gaussian regressive model in the form

$$f(y_t|\boldsymbol{\theta}, \boldsymbol{\psi}) \sim \mathcal{N}(\boldsymbol{\psi}_t' \boldsymbol{\theta}_t, r) \tag{1}$$

where f denotes a probability density function (pdf) of the argument, $\boldsymbol{\theta} \in \mathbb{R}^n$ is a vector of regression coefficients and $\boldsymbol{\psi} \in \mathbb{R}^n$ is the regression vector. y_t denotes the scalar value of interest, corrupted by a Gaussian white noise with zero mean value and constant variance $r \in \mathbb{R}^+$. This model coincides with the recursive least squares with

$$y_t = \boldsymbol{\psi}_t' \boldsymbol{\theta}_t + e_t, \qquad t = 1, 2, \ldots \tag{2}$$

where $e_t \sim \mathcal{N}(0, r)$.

The estimation of parameters θ has two steps:

- Data update – new data are incorporated into the parameter pdf.
- Time update – the unknown transition from $\theta_{t-1} \to \theta_t$ is reflected.

Both these steps are given in literature, e.g. [7, 4]. The time update step in our case has the form of exponential forgetting [7] or partial forgetting [1]. The latter is attached in Appendix.

3 Data transformations

The available data represent various physical variables with different scales. There are two potentially 'dangerous' situations:

- Different measuring units – the data are mix of speed, length and other variables.
- Different scale units – there are different scales for length (millimeters, meters etc).

Both of these reasons can potentially lead to modelling problems, especially when algorithms working with Euclidean norms are used. As the Euclidean distance is computed as a sum of variable differences, its result greatly depends on the ranges of the variables. Therefore, we demonstrate the use of data normalization to $\mathcal{N}(0,1)$.

Proposition 1 (Normalization to $\mathcal{N}(0,1)$) *Let the random variable $X \in \mathbb{R}$ have finite first and second order moments, namely mean value μ and variance σ^2. Then,*

$$Z = \frac{X-\mu}{\sigma} \tag{3}$$

has normalized normal distribution $\mathcal{N}(0,1)$.

Proof omitted.

Since in our case are the moments unknown, we use their sample variants, i.e., sample mean $\bar{X}$ and sample variance S^2. Instead of the random variable X we work on its n realizations, i.e.

$$z_i = \frac{x_i - \bar{X}}{S}, \qquad i = 1, \ldots, n, \tag{4}$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2.$$

For online estimation purposes, a need of recursive evaluation of these moments arises. A suitable solution is to calculate them on a finite sliding window. A potential risk consists in situations when a failure occurs – faulty data may significantly corrupt the regular data when used for normalization, as the moments can diverge to nonsense values. This may especially occur, if the normalized values are constant (then $\sigma^2 = 0$ leads to division by zero) or if the measured invalid value is significantly different to the reasonable values (outliers).

A short analysis of normalization impact is given in Section 5.1 on page 14.

4 Models used

There are multiple ways how to model the output y, i.e., the output thickness deviation of the rolled sheet, e.g., [2]. Some of them are 'recycled' below, while others are new proposals. Let's introduce the following notation for the related physical variables, used further in this reading:

H_1, H_2	input, output thickness – absolute
h_{1nom}, h_{2nom}	input, output thickness – nominal
h_1, h_2	input, output thickness – deviation
y	output thickness deviation
F	total rolling force
v_1, v_2	input, output speed
v_r	speed ratio v_1/v_2
z	uncompensated rolling gap

where

$$H_i = h_{inom} + h_i, \qquad i = 1, 2 \tag{5}$$

Upon (2), we can build the following models, characterized by their regression vectors:

- The mass-flow based model

$$\psi_t = [h_{1nom} v_r, h_1 v_r, 1]'. \tag{6}$$

- Gaugemeter-based model, evaluated as polynomials of order m

$$\psi_t = [F^1, \ldots, F^m, z, 1]', \qquad m = 2, 3, \ldots \tag{7}$$

- Black-box model 1

$$\psi_t = [h_1, z, 1]', \tag{8}$$

- Black-box model 2

$$\psi_t = [h_1, z, v_r, 1]' \tag{9}$$

Like in many industrial applications, the sheet rolling process is typical for delayed data. This means, that when some of the measurements are being gathered, these measurements are delayed to the others due to the traffic delay. Here, the traffic delay is connected with the output thickness deviation, measured by a distant device. The distance corresponds usually to either 19 or even 120 time steps. The only straightforward solution is to use such data in place of the current one.

4.1 Analysis of the mass-flow model

The mass-flow model is specific for its underlaying physical principle. If we neglect the change of the strip width and other variables, the following equation holds:

$$\frac{v_1}{v_2} = \frac{H_2}{H_1}, \tag{10}$$

i.e., the ratio of speeds is equal to the reciprocal fraction of absolute thicknesses. Rewriting (10) yields

$$v_r = \frac{v_1}{v_2} = \frac{h_{2nom} + h_2}{h_{1nom} + h_1} \tag{11}$$

from which follows the relation for the output thickness deviation

$$y = h_2 = h_{1nom} v_r + h_1 v_r - h_{2nom}.$$

Considering the last term as an absolute term yields a model with the regression vector (6).

Let us now focus on the validity of this model, using standard data batch file *Gi-200810211323_206-10-B_3.mat*. To avoid working with initial unstable data, measured at the beginning of the rolling process, the first 219 samples were dropped. The length of the batch was 1000 samples.

First, we can compare the right- and left-hand sides of (10). Apparently, the speed ratio v_r has significantly more noisy character than the ratio of absolute thicknesses. Some selected

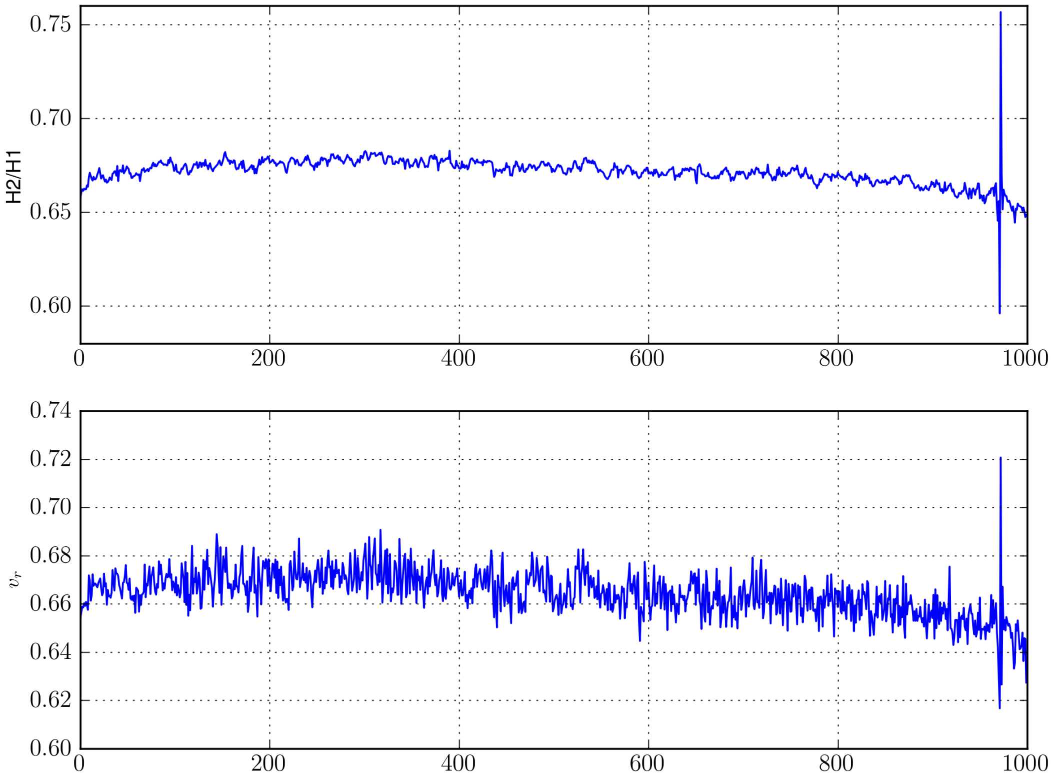


Figure 1: Evolution of H_2/H_1 and $v_r = v_1/v_2$, respectively.

statistics of series v_r and H_2/H_1 are given in Table 1. The Spearman correlation coefficient is 0.707, which indicates certain, but not really tight correlation between the two data.

Figure 2 shows two boxplots for H_2/H_1 and v_r, respectively. The median of the latter is lower and its inter-quartile range (IQR) is much higher than the ones of the former. H_2/H_1 contains a lot of outliers, i.e. data with absolute values more than 1.5IQR from median, which indicates the need of adaptive tracking. However, one would expect the two boxplots to be almost the same, they are not. This undermines the arguments justifying the mass-flow model as well.

The scatter plot 3 visualises the dependence between H_2/H_1 and v_r; the outliers were filtered out. The blue line denotes the perfect correlation between the two variables. On the other side, if the markers were concentrated on a horizontal line, there would be no (linear) correlation at all. In the empirical case, the data are scattered around a cluster with centroid above the blue line and their correlation is (visually) not very high; the red line depicts the linear trend ($0.5362v_r + 0.3150$). Let us remind that the correlation coefficient is 0.707.

Let us now focus on a subset of the data batch. It contains 400 measurements of h_1, h_2, v_1, v_2, z and F. This sample starts with $t_0 = 400$, which is identical to the 180th sample in the previous batch. The course of selected variables is depicted in Fig. 4. It is worth to mention that the correlation coefficient of this batch was even 0.419! It is necessary to pose the following questions:

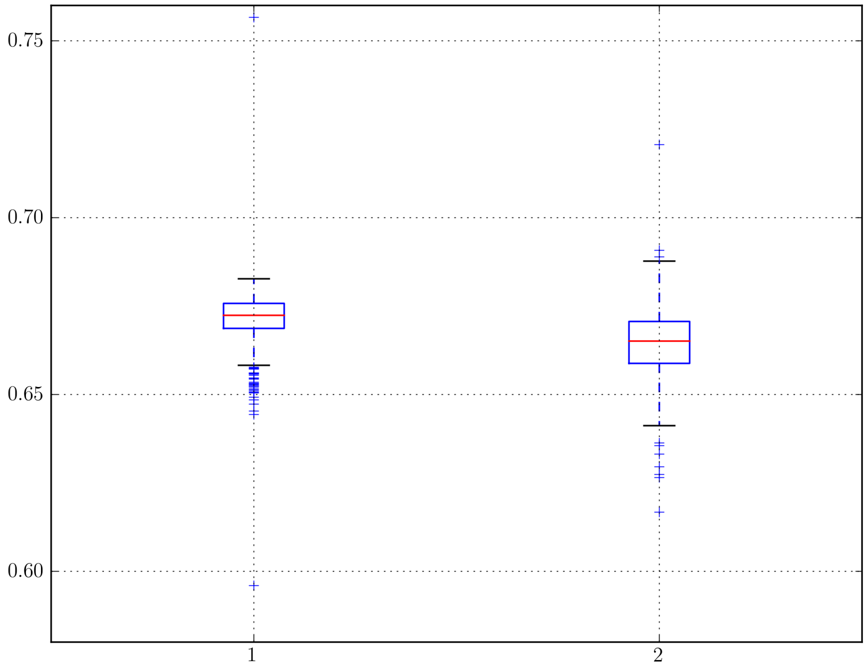


Figure 2: Boxplots of H_2/H_1 (1) and $v_r = v_1/v_2$ (2), respectively. The blue box extends from the lower to the upper quartile, hence it contains 50% of data samples. The red line is the median. The whiskers show the range of the data, their length is 1.5IQR (inter-quartile range). + denote outlier data w.r.t. quartiles.

1. The input thickness deviation h_1 course resembles periodic character, which is visually not similar to evolution of any other variable. Is this caused by a systematic error, e.g., the measuring device? If yes, can it be eliminated or detected?

2. The input speed v_1 has very noisy character. What is the cause? The measuring device? What are the (dis)continuity properties of its measurements?

3. The output speed v_2 measurement is polluted with the noise even more and the amplitude is (significantly) higher. Is the reason the higher speed of the strip?

4. The rolling gap z is measured indirectly, based on the piston/plunger position. What is the reliability and precision of such measurements? What may cause the change of the measured gap size?

5. The rolling force F can be measured either directly or indirectly. What is the reliability and precision of such measurement?

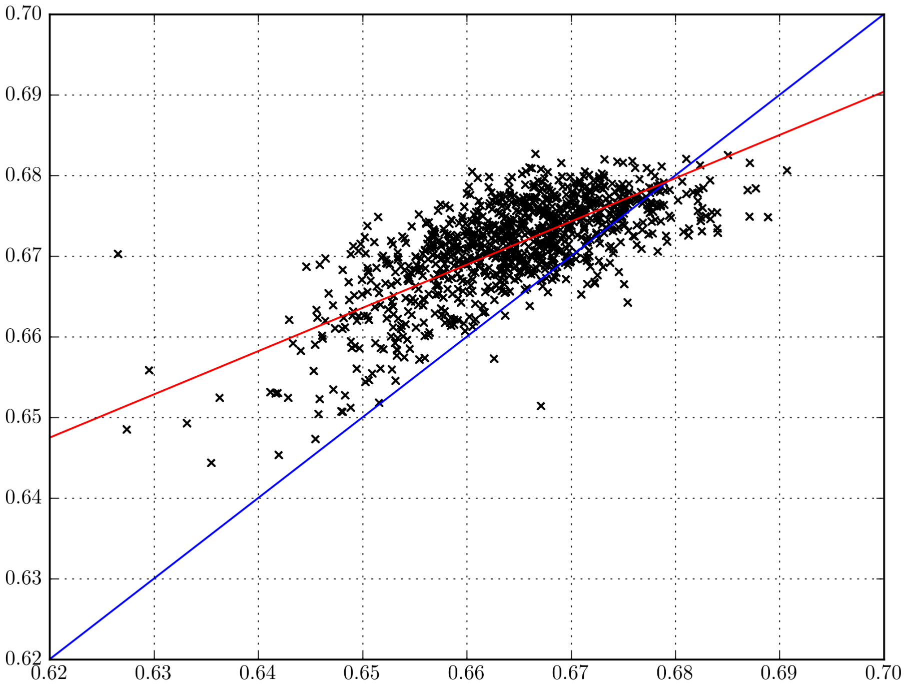


Figure 3: Scatterplot $H_2/H_1 \times v_r$ with line of theoretical dependence (blue) and real linear regression line (red).

Statistics	v_r	H_2/H_1
Minimum	0.6167	0.5960
Maximum	0.7207	0.7566
Range	0.1040	0.1606
Mean	0.6646	0.6714
Median	0.6651	0.6724
Standard deviation	0.0092	0.0070
Variance	8.492e-05	4.884e-05
Correlation coefficient (Spearman)	**0.707**	

Table 1: Descriptive statistics of data.

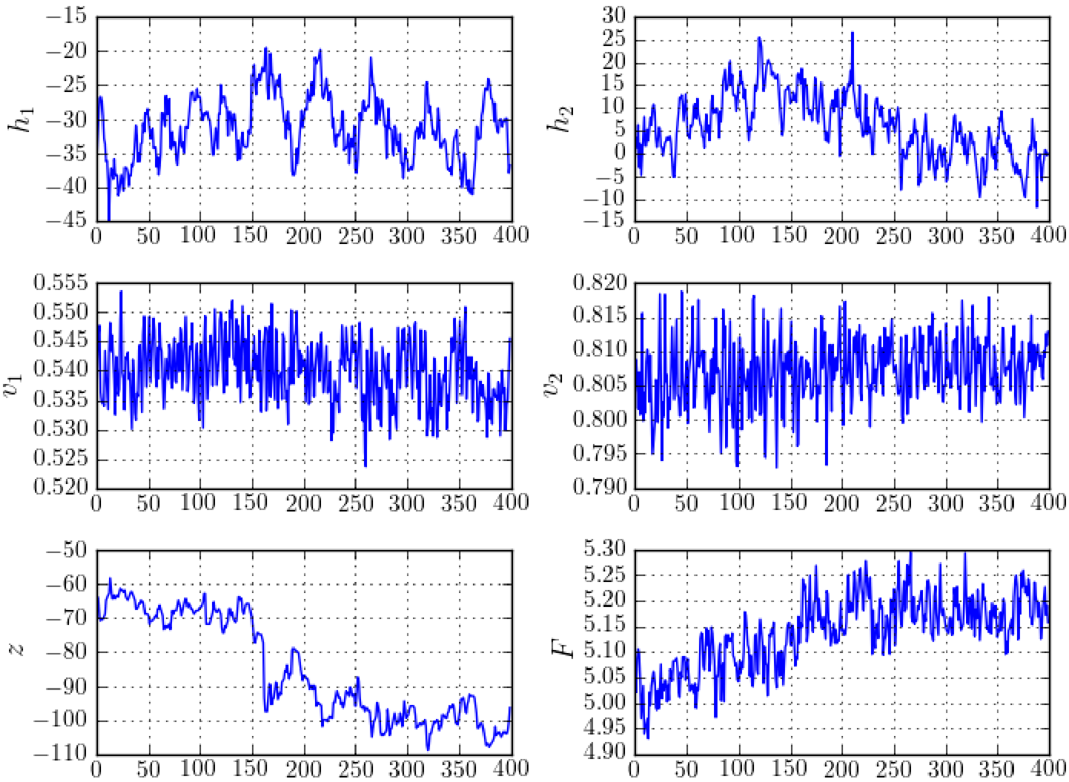


Figure 4: Course of h_1, h_2, v_1, v_2, z and F

4.1.1 Discussion

The mass-flow equation (10) was studied with empirical data. Some selected properties of the data were visualised using a 'classical' plot and boxplots. Selected descriptive statistics were given in the table. The idea of closely deterministic relation between both sides of the equation was corrupted to some degree, in particular due to:

1. Spearman correlation coefficient $r_S = 0.707$.
2. The very noisy character of v_r, at least in comparison to H_2/H_1.
3. Difference in some descriptive statistics.
4. The scatterplot visually declines full causality between v_r and H_2/H_1.

Possible causes:

- Plastic deformation of the rolled material structure.
- Deformations in the rolling mill.
- Change of the (not measured) width of the rolled material.
- Sensors' inaccuracy.
- Other causes.

5 Analysis of the gaugemeter model

The gaugemeter model (7) models the output thickness deviation h_2 with a regression vector containing 1 for absolute term, z and a series of powers of F^n, $n = 1, 2, \ldots$. Theoretically, the output thickness (deviation) should be well correlated with the rolling force, however, the linear correlation coefficient was, for standard data batch used in the previous chapter, namely *Gi-200810211323_206-10-B_3.mat*, equal to 0.323. This indicates very low linear correlation between these two variables, anyway, it does not say anything about a non-linear relation. Let us compare 2nd, 3rd and 4th order gaugemeter-based models without forgetting.

First, let us work on non-normalized data. In the figures 5 and 6 are depicted 2nd and 4th order-models characteristics. Some selected characteristics of prediction errors are given in the Table 3. Apparently, the increasing order of the model does not lead to significant improvement.

Normalization of regressors leads to the balancing of the powers of F. This is evident from the Fig. 8. In the Fig. 7 are depicted courses for the 3rd order gaugemeter model. The difference between normalized F and F^2 are shown in Fig. 9.

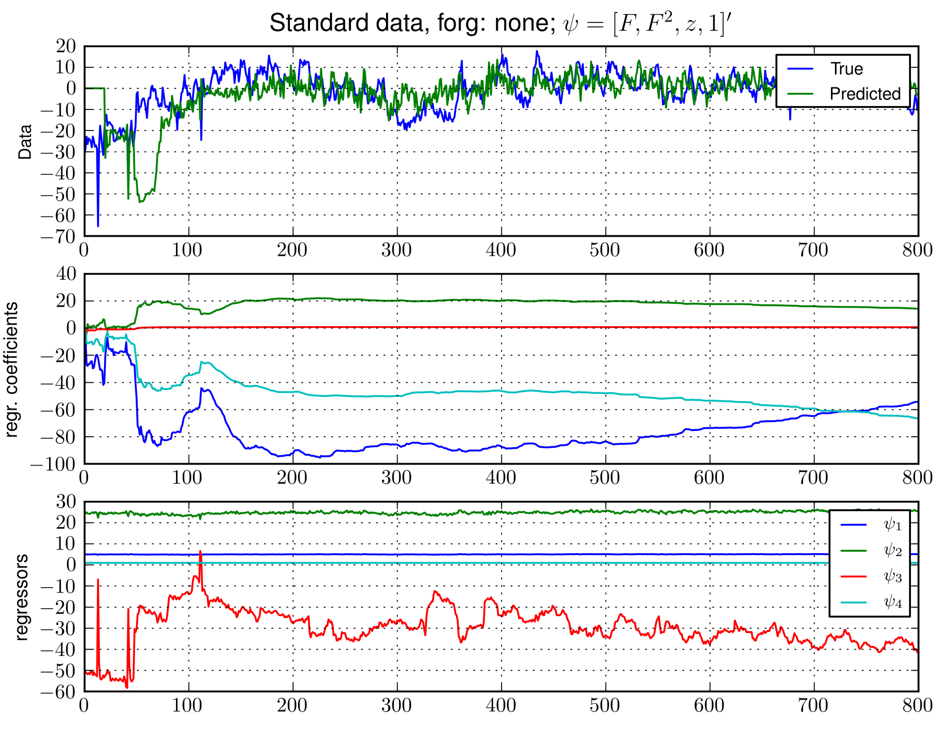


Figure 5: Gaugemeter model – 2nd order

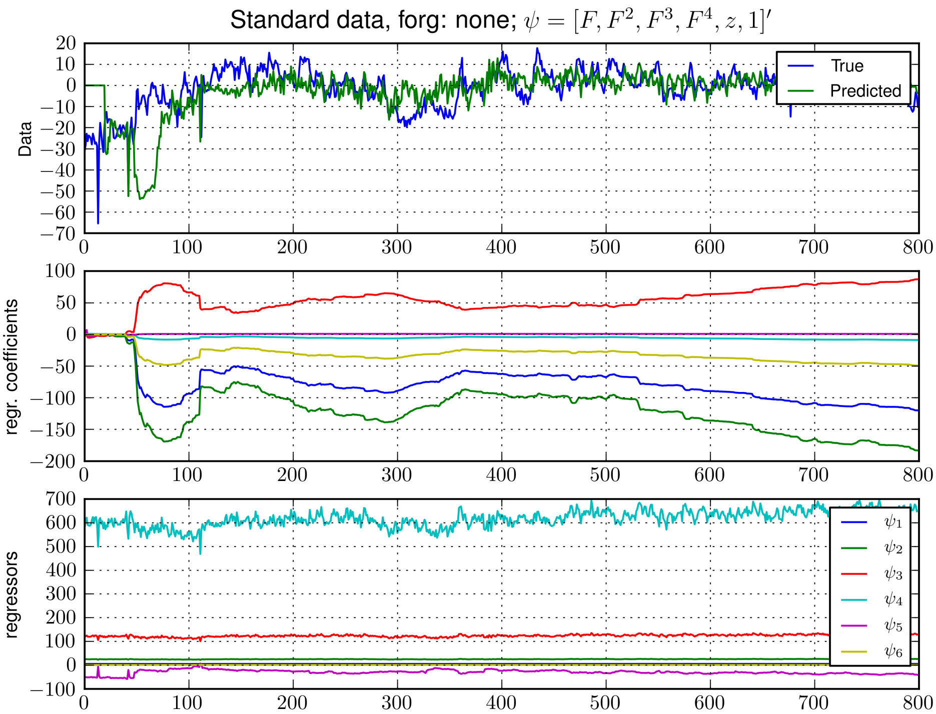


Figure 6: Gaugemeter model – 4th order

Statistics	2nd order	3rd order	4th order
Maximum	48.969	48.810	48.727
Minimum negative	-65.434	-65.434	-65.434
Mean	0.928	0.934	1.273
Median	0.566	0.465	0.921
Standard deviation	10.970	10.985	10.845
Variance	120.344	120.681	117.605
RMSE	121.206	121.554	119.226

Table 2: Gaugemeter models – descriptive statistics of prediction errors for non-normalized regressors.

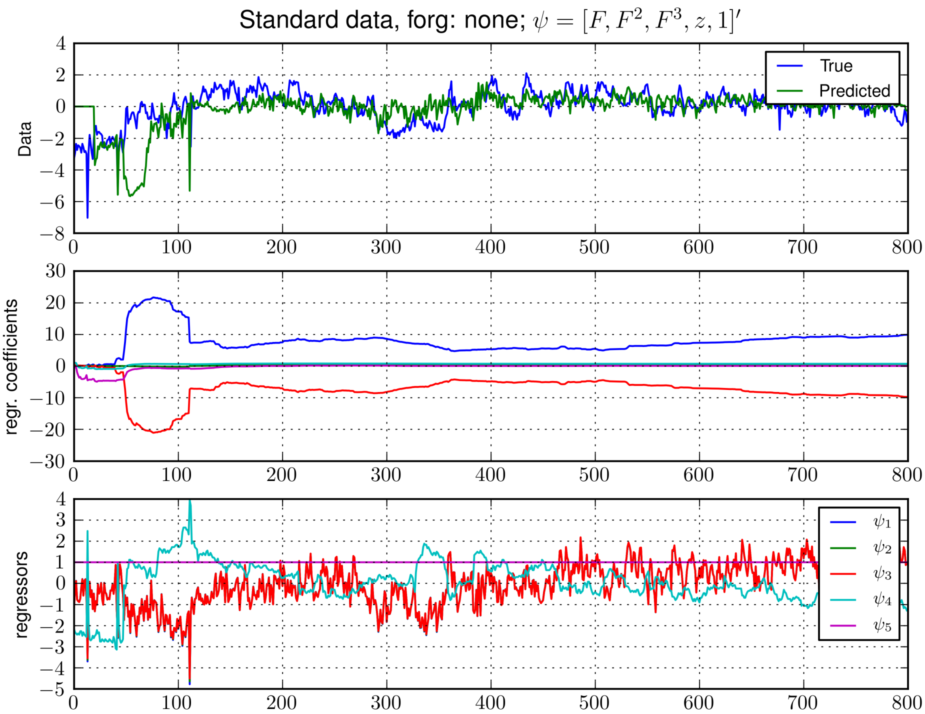


Figure 7: Gaugemeter model – 3rd order – normalized data

Statistics	2nd order	3rd order	4th order
Maximum	5.275	5.522	7.065
Minimum negative	-7.034	-7.034	-7.034
Mean	0.145	0.156	0.157
Median	0.103	0.115	0.122
Standard deviation	1.175	1.177	1.179
Variance	1.380	1.386	1.391
RMSE	1.402	1.410	1.415

Table 3: Gaugemeter models – descriptive statistics of prediction errors for non-normalized regressors.

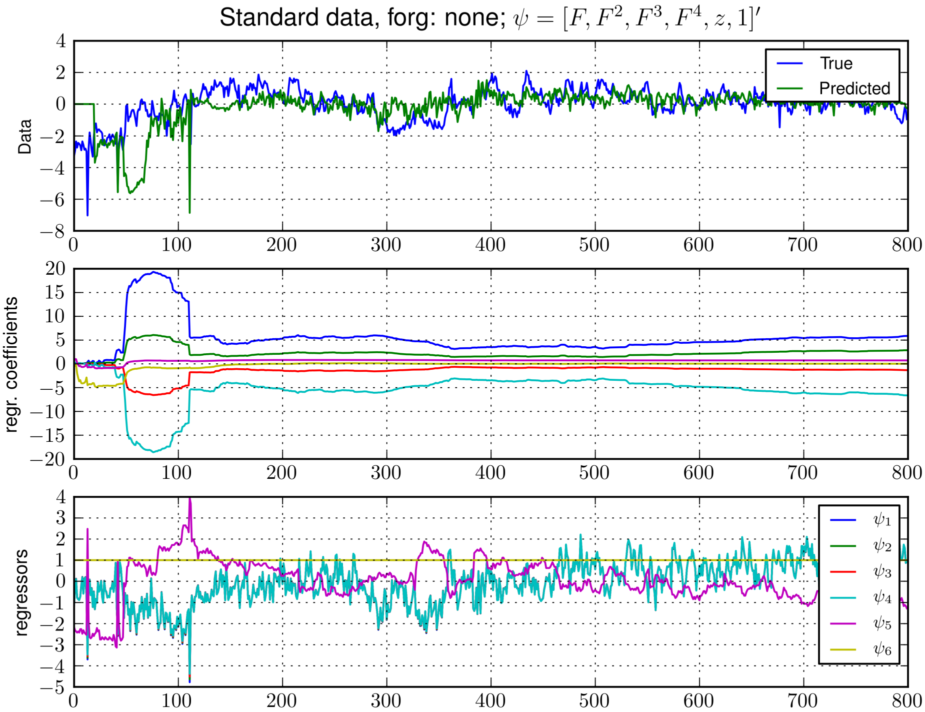


Figure 8: Gaugemeter model – 4th order – normalized data

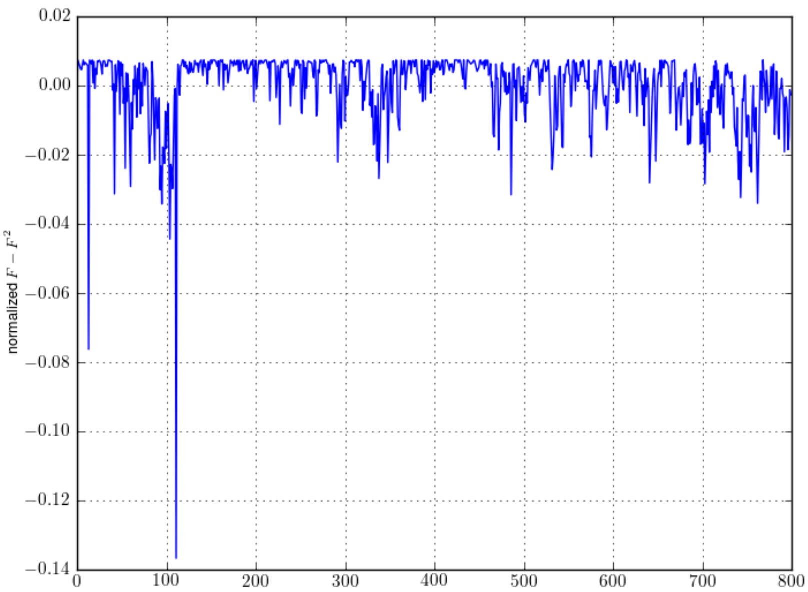


Figure 9: Difference between normalized F and F^2

5.1 Brief analysis of normalization impact

In this section we bring a short analysis of impact of data normalization to normal normalized distribution $\mathcal{N}(0,1)$. This procedure was described in Section 3 by Proposition 1.

Let us deal with the mass-flow model (6), whose physical validity was studied in Section 4.1. Under rather weak assumption, that the regression coefficients should lay in interval [0, 1], and should be close to 1 (or, at least, should not be close to 0), we can compare the modelling process for both normalized and non-normalized data by means of coefficients evolution.

We chose file *Gi-200810211323_206-10-B_3.mat*, containing stable data. The window started with $t_0 = 1250$ and was 800 samples long. The traffic delay for h_2 was 19 time steps, the modelling started with non-informative prior with $V_0 = diag(0.1, 0.01, 0.01, 0.01)$ and $\nu_0 = 7$, the estimation used exponential forgetting with factor 0.99.

The results are depicted in Figs. 10 and 11 for non-normalized and normalized data, respectively. From the figures follows, that the former case suffers from anti-correlation of θ_1 and θ_3, i.e., between $h_{1nom}v_r$ and the absolute term. This indicates suspicion of model overparametrization. The second regression coefficient θ_2 remains close to 0 and its impact on modelling is therefore very low. On the other side, the normalization of the data (regressors) led to release of the correlation.

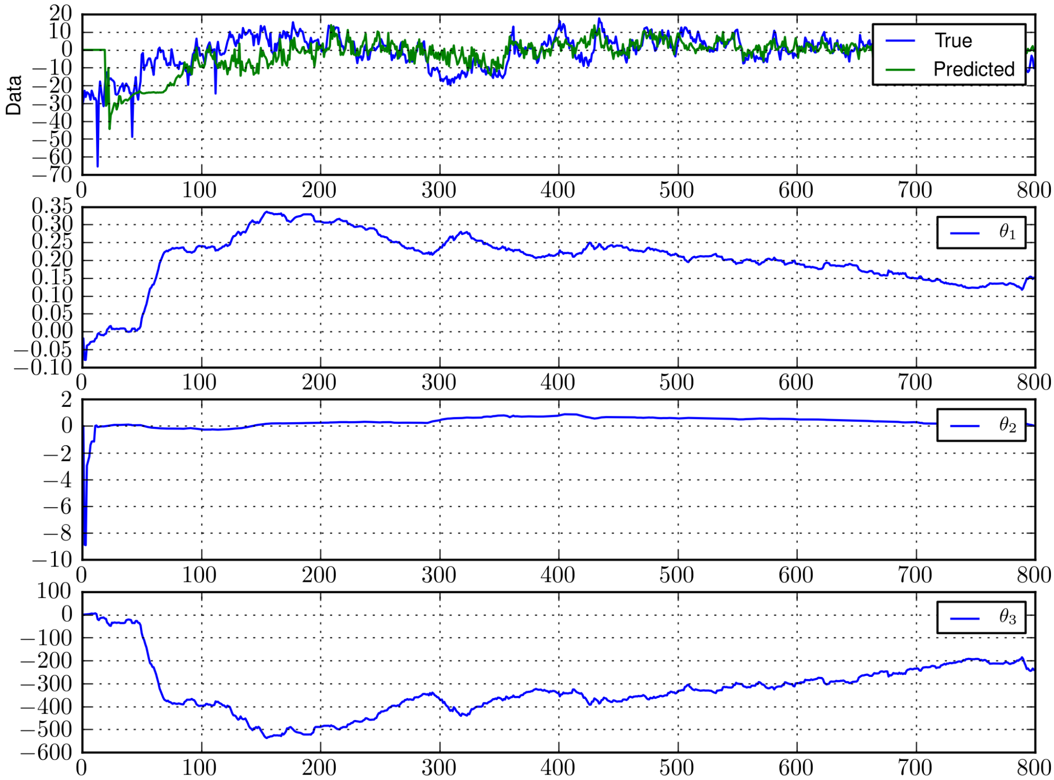


Figure 10: Course of regression coefficients estimation for non-normalized data.

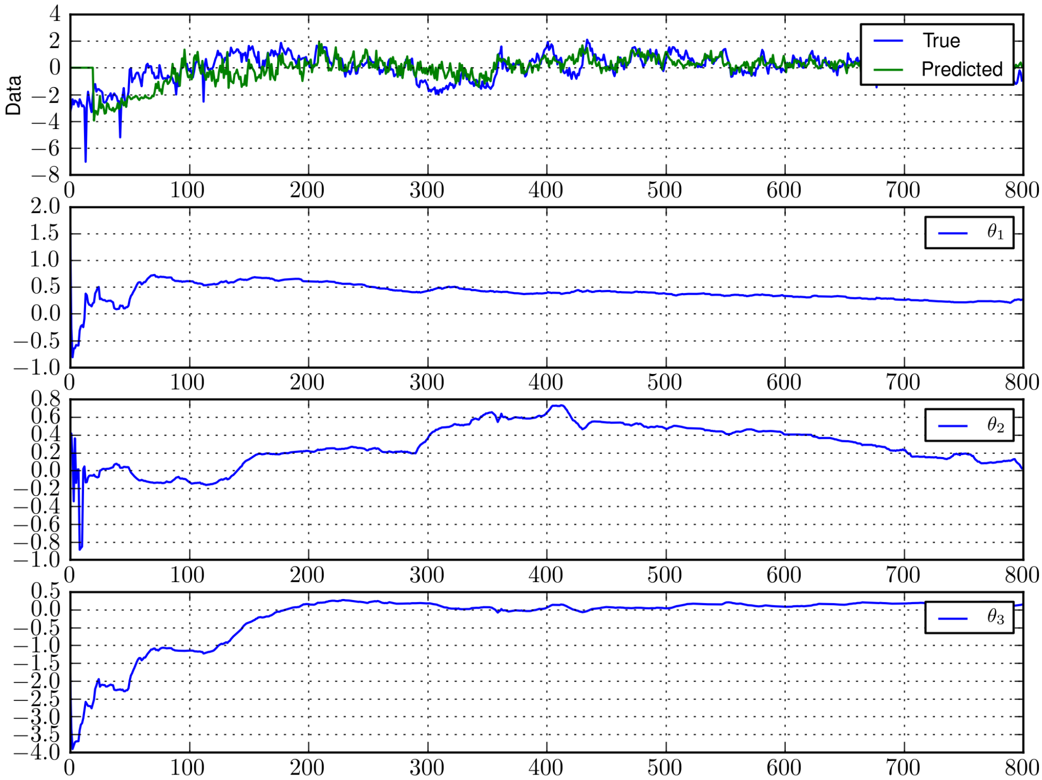


Figure 11: Course of regression coefficients estimation for normalized data.

6 Standard data modelling

This and the following chapters bring analyses of models mentioned in the former parts of the report. This chapter deals with the full set of models defined and for each of them, it describes the impact of various estimation techniques on data prediction. In contrast to the following chapters, each case is accompanied by a relevant figure here. The following chapters depict only the no-forgetting cases.

The standard data batch consisted of data from file *Gi-200810211323_206-10-B_3.mat*. To avoid the initial stabilization, the modelling started from 1250th sample, 800 samples were used. They were modelled with the following four models:

- Mass-flow model (6), which was widely discussed in Section 4.1, where its limited theoretical validity was studied.
- Blackbox model 1 (8).
- Blackbox model 2 (9).
- 2nd order gaugemeter model (7).

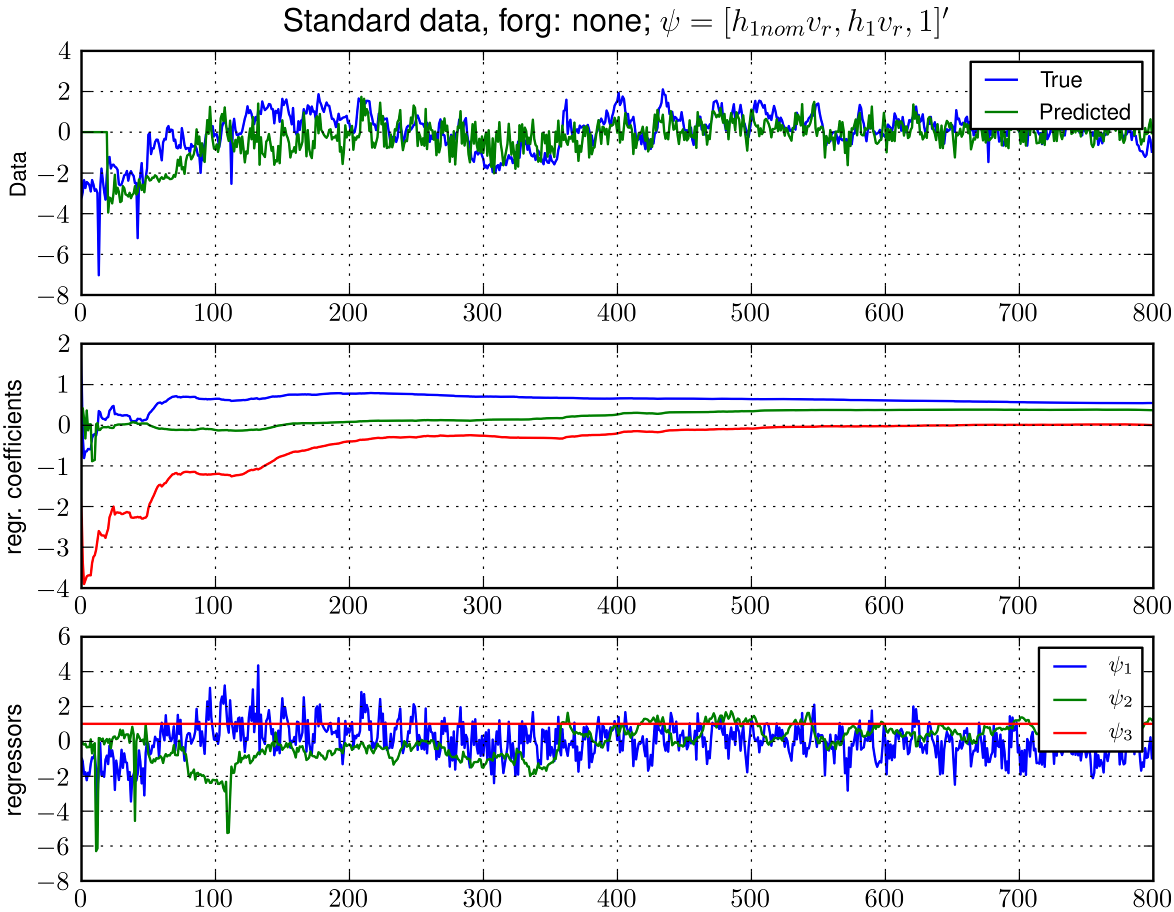
The parameter estimation was evaluated:

- Without forgetting, i.e., no time update of parameters pdf was done.
- With exponential forgetting. The forgetting factor was 0.99.
- With partial forgetting. The hypotheses weights were evaluated automatically during the estimation. The factors were:
  - 0.95 for hypotheses' weights,
  - 0.99 for construction of alternative pdf f_A for hypothesis H_1
  - 0.9 for construction of alternative pdf for f_A for hypothesis H_2

The results are summarized below each model.

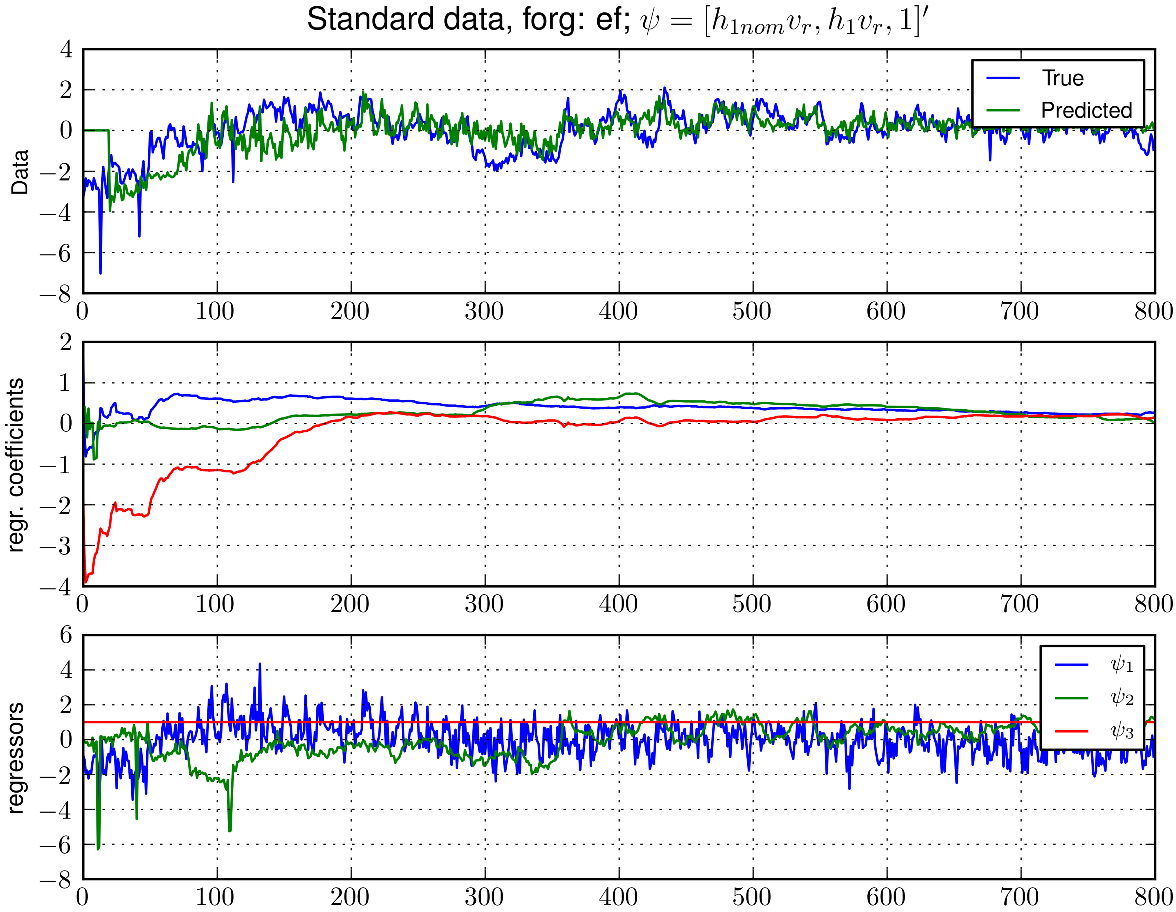
6.1 Mass-flow model

6.1.1 No forgetting



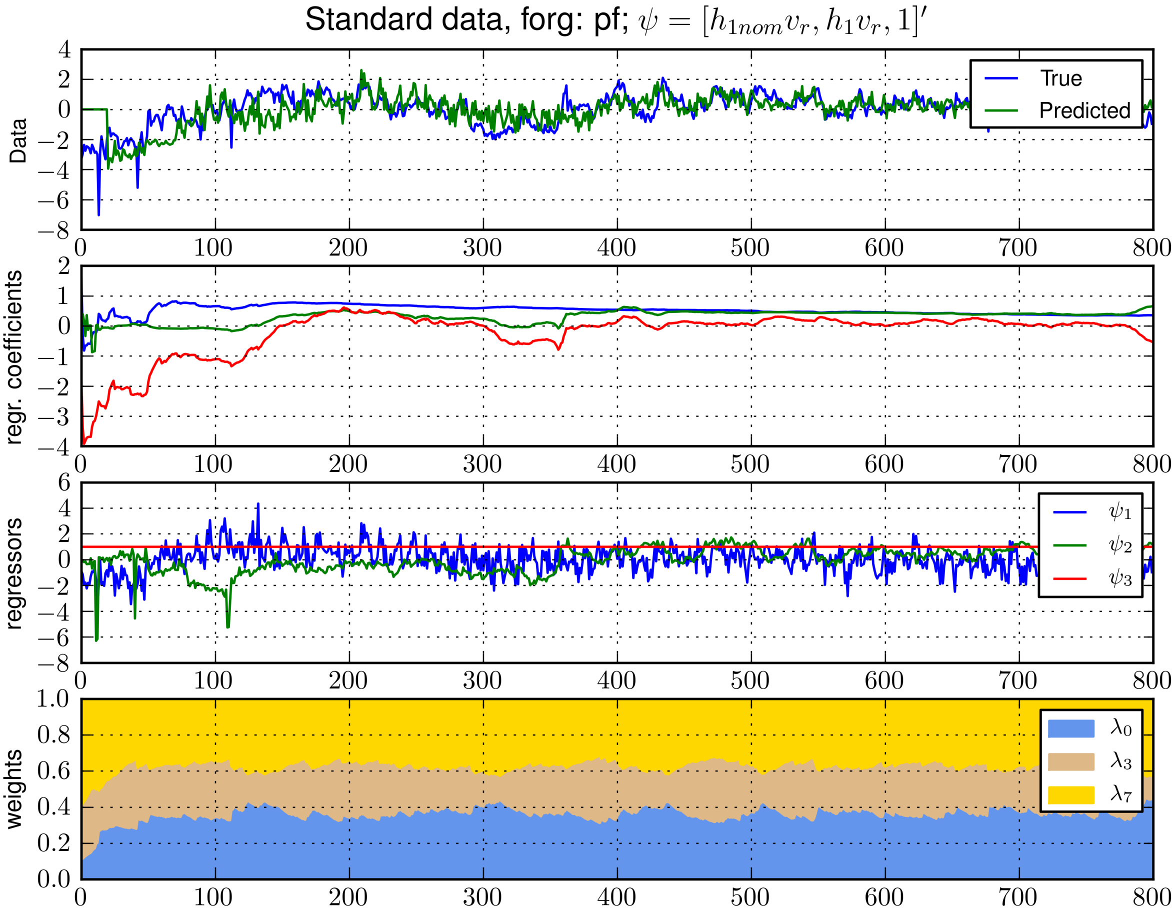
Error statistic	Value
Mean	0.254
Median	0.273
Variance	0.956
Standard deviation	0.978
RMSE	1.021

6.1.2 Exponential forgetting



Error statistic	Value
Mean	0.028
Median	0.024
Variance	0.892
Standard deviation	0.945
RMSE	0.893

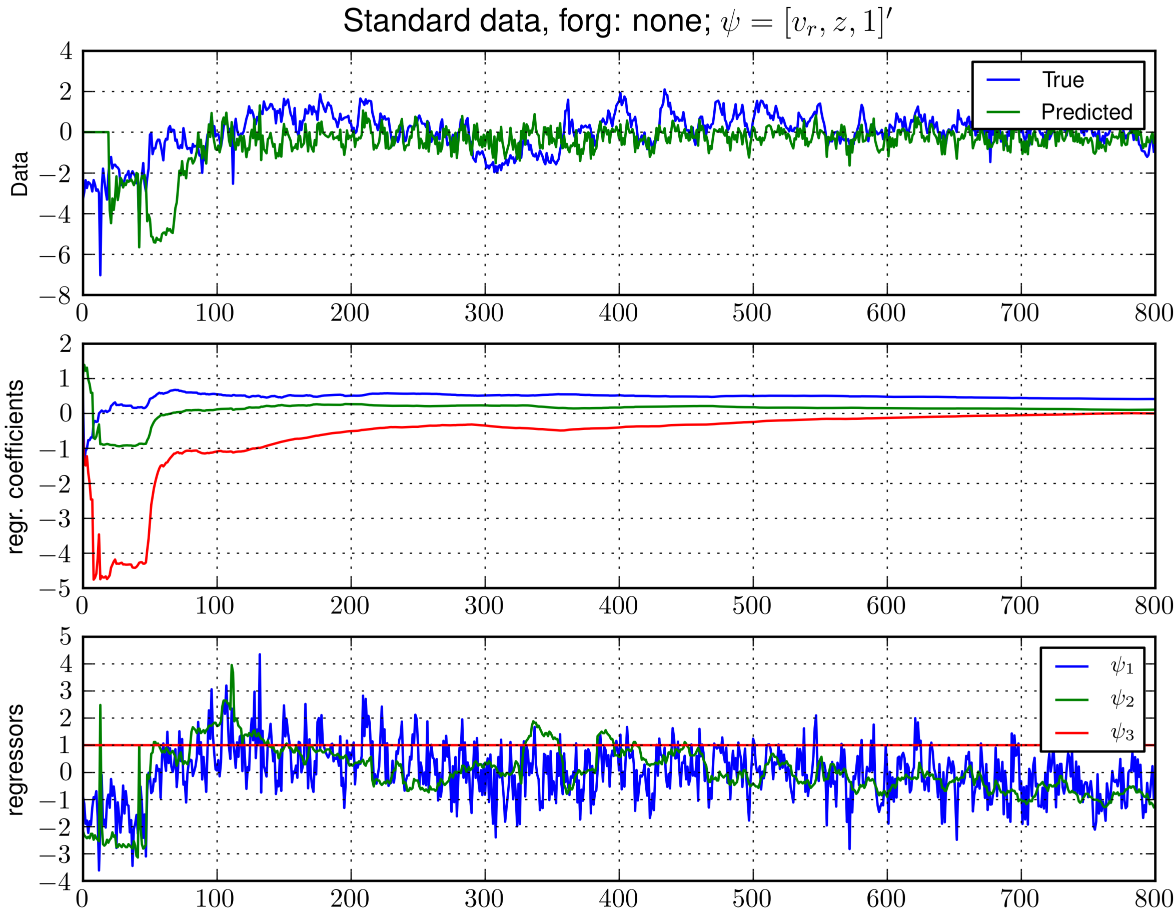
6.1.3 Partial forgetting



Error statistic	Value
Mean	0.007
Median	0.003
Variance	0.911
Standard deviation	0.955
RMSE	0.911

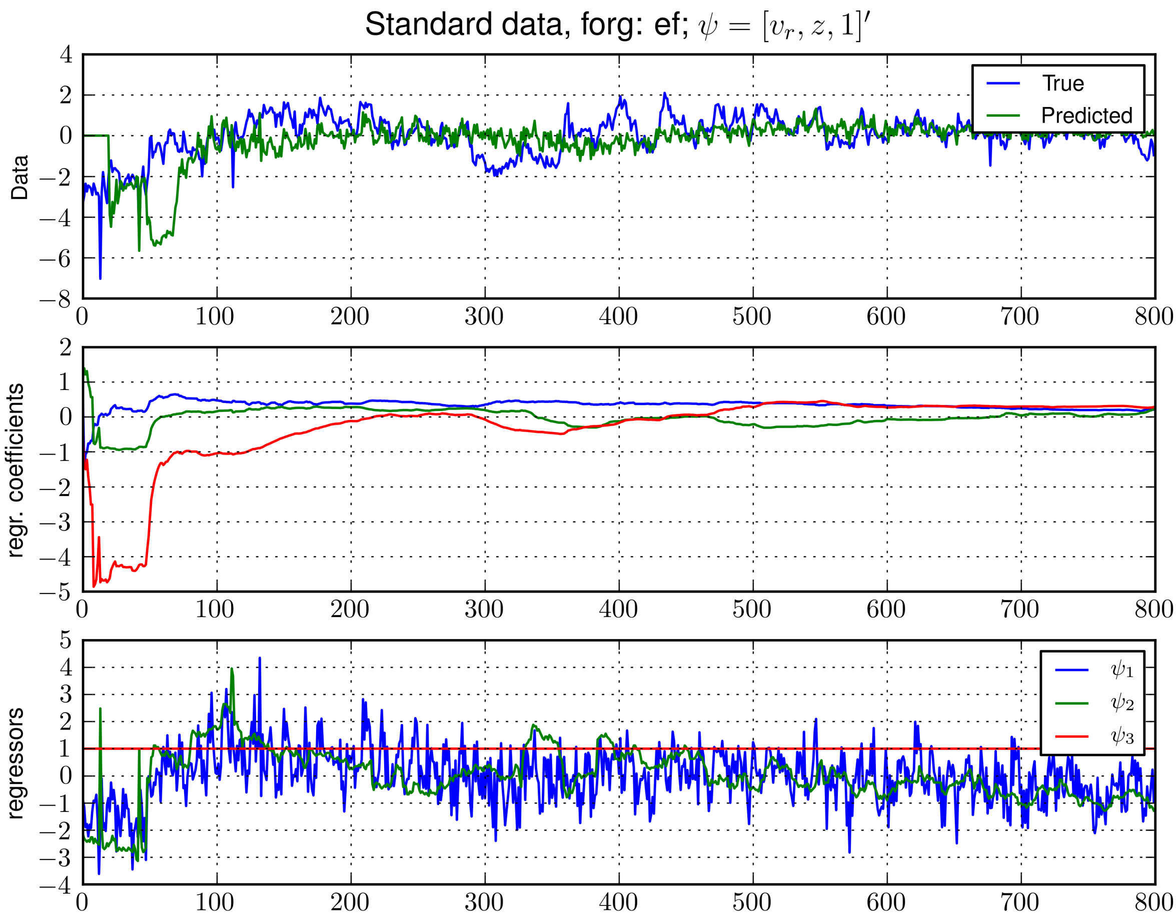
6.2 Blackbox 1 model

6.2.1 No forgetting



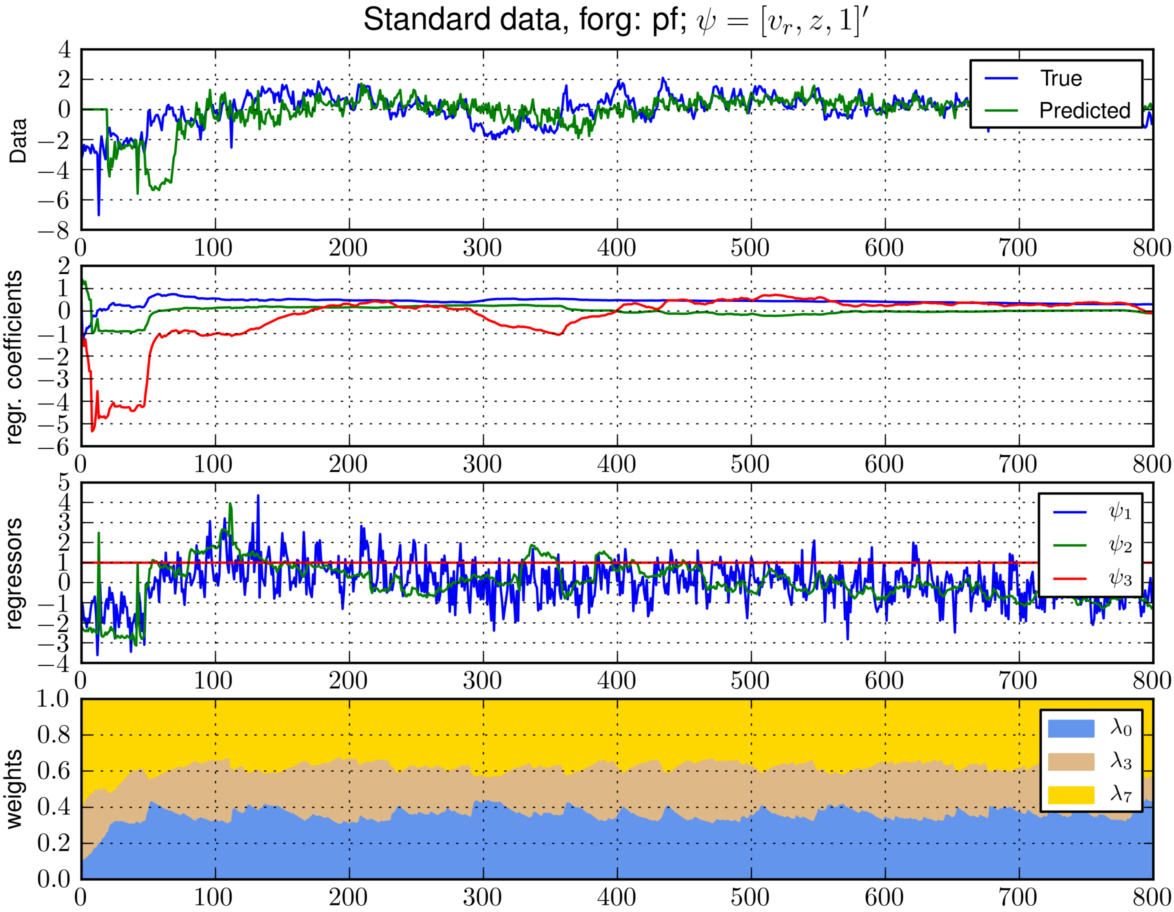
Error statistic	**Value**
Mean	0.521
Median	0.528
Variance	1.33
Standard deviation	1.15
RMSE	1.603

6.2.2 Exponential forgetting



Error statistic	Value
Mean	0.219
Median	0.152
Variance	1.38
Standard deviation	1.18
RMSE	1.431

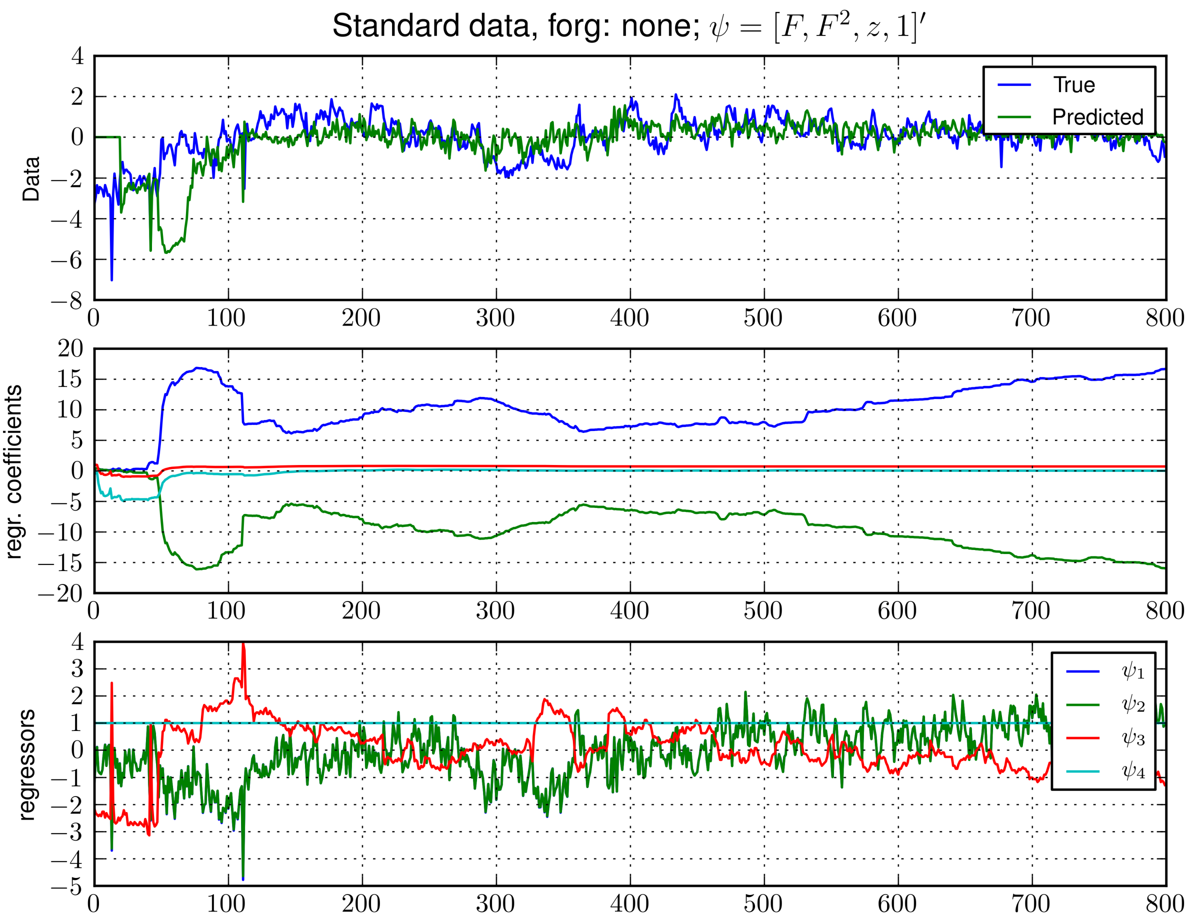
6.2.3 Partial forgetting



Error statistic	Value
Mean	0.122
Median	0.058
Variance	1.34
Standard deviation	1.16
RMSE	1.360

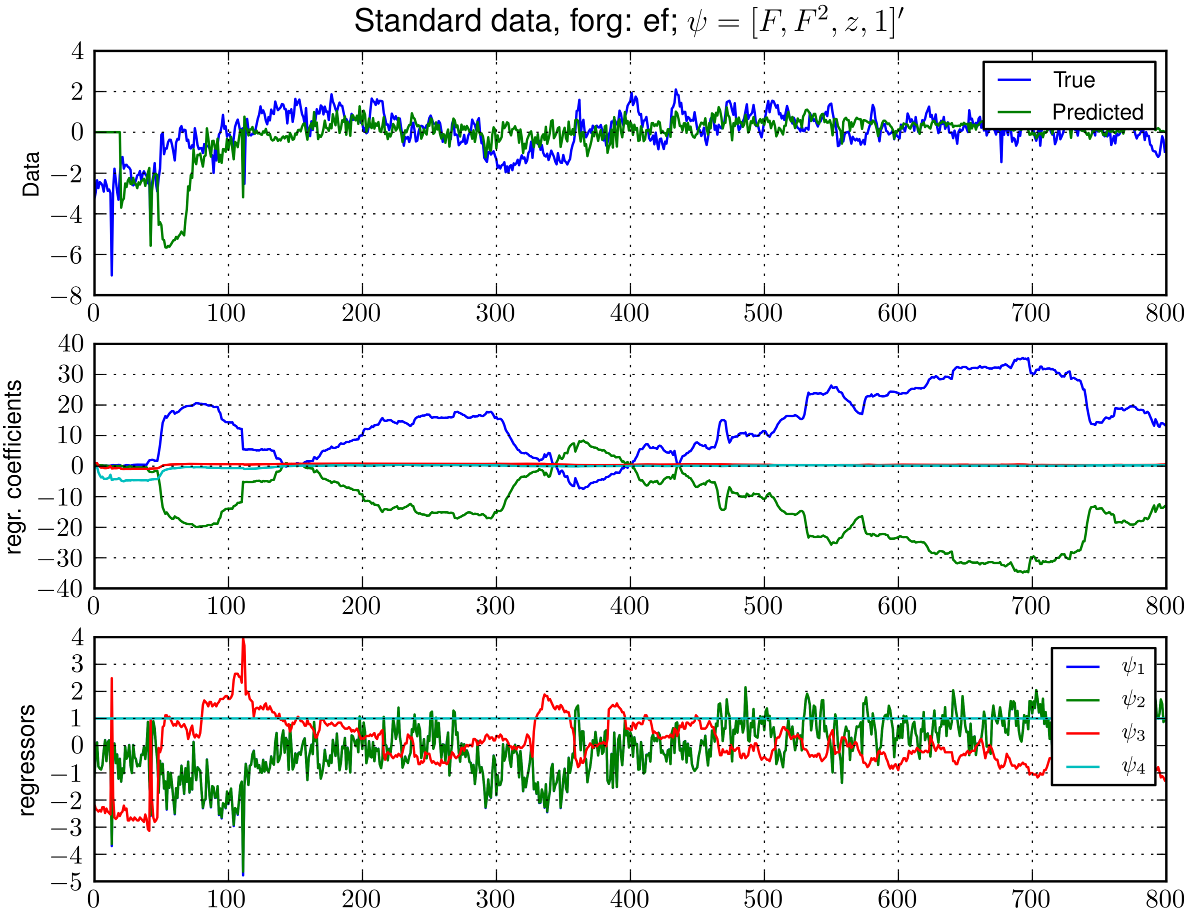
6.3 2nd order gaugemeter model

6.3.1 No forgetting



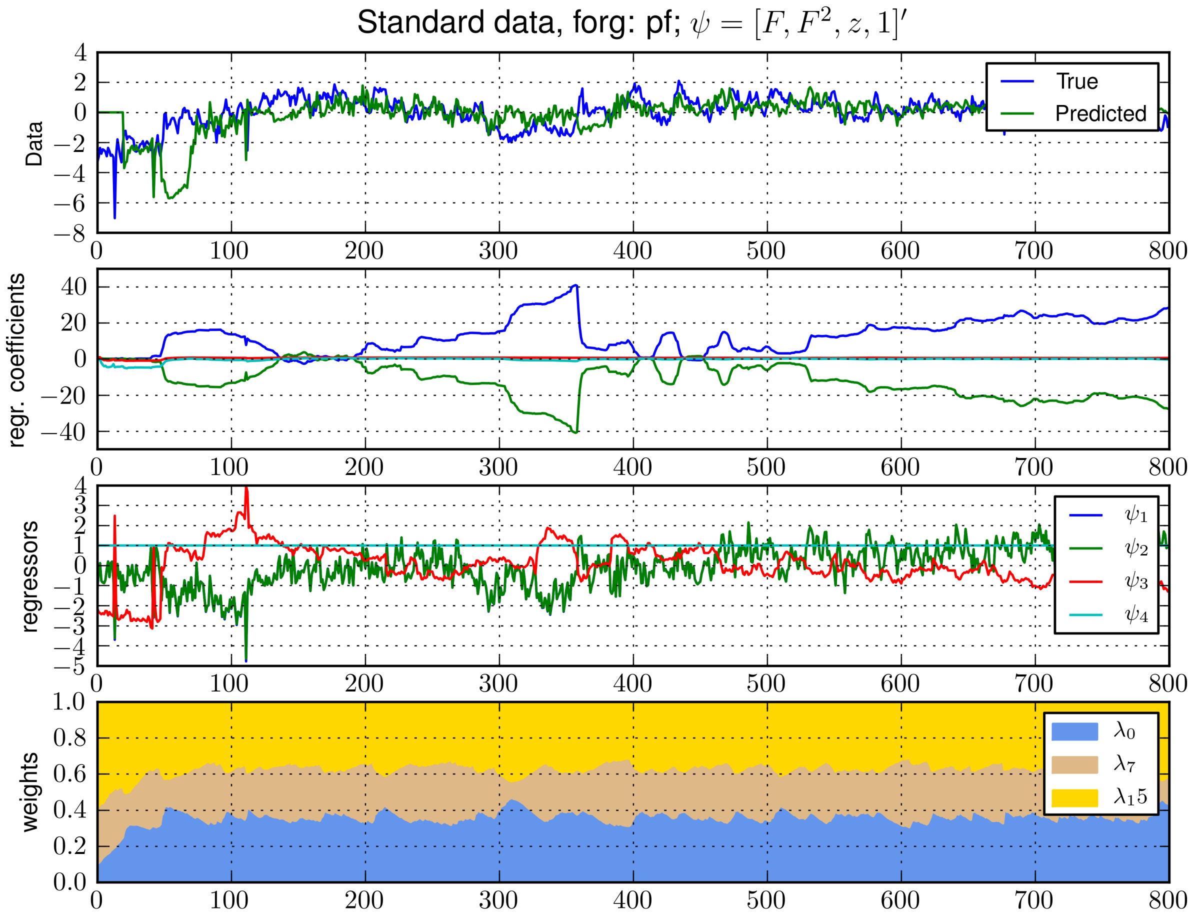
Error statistic	Value
Mean	0.145
Median	0.103
Variance	1.38
Standard deviation	1.18
RMSE	1.402

6.3.2 Exponential forgetting



Error statistic	Value
Mean	0.122
Median	0.045
Variance	1.37
Standard deviation	1.17
RMSE	1.384

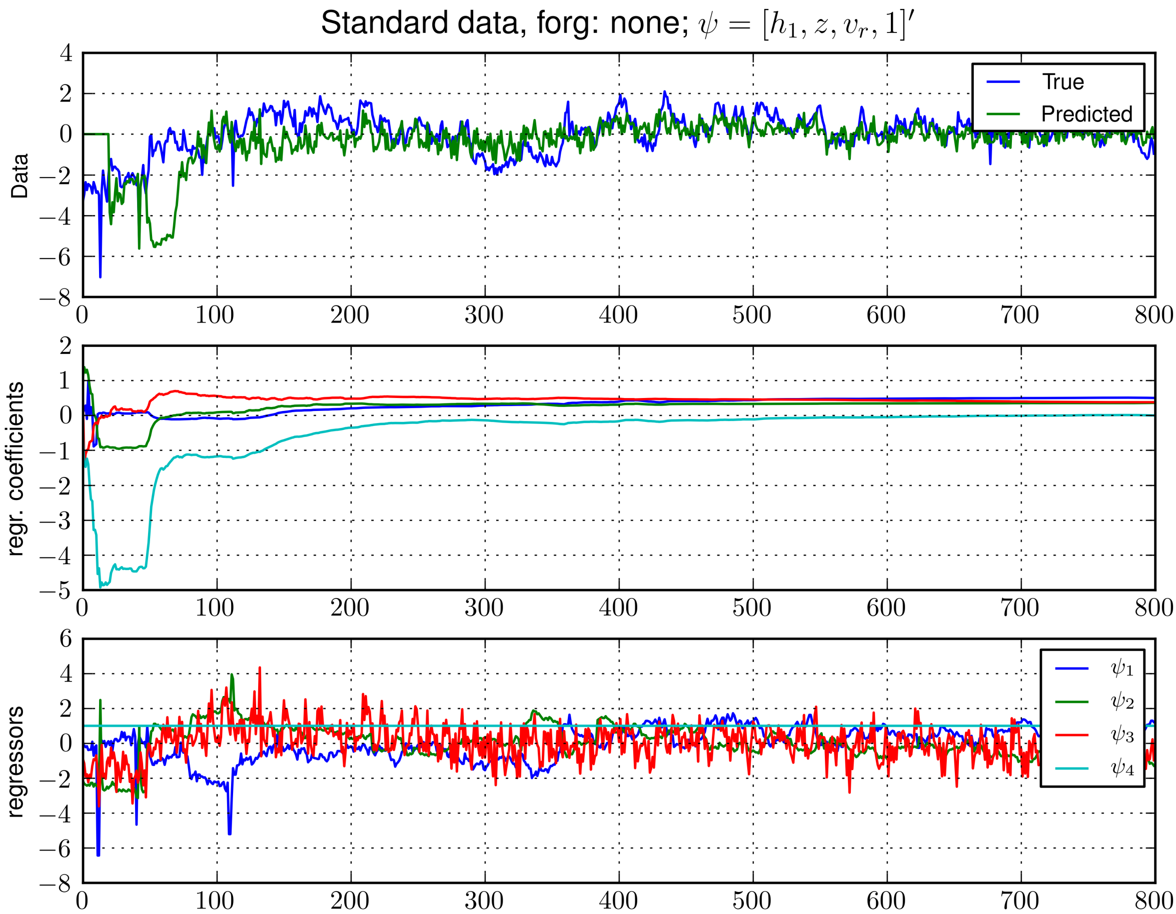
6.3.3 Partial forgetting



Error statistic	Value
Mean	0.128
Median	0.051
Variance	1.4
Standard deviation	1.18
RMSE	1.420

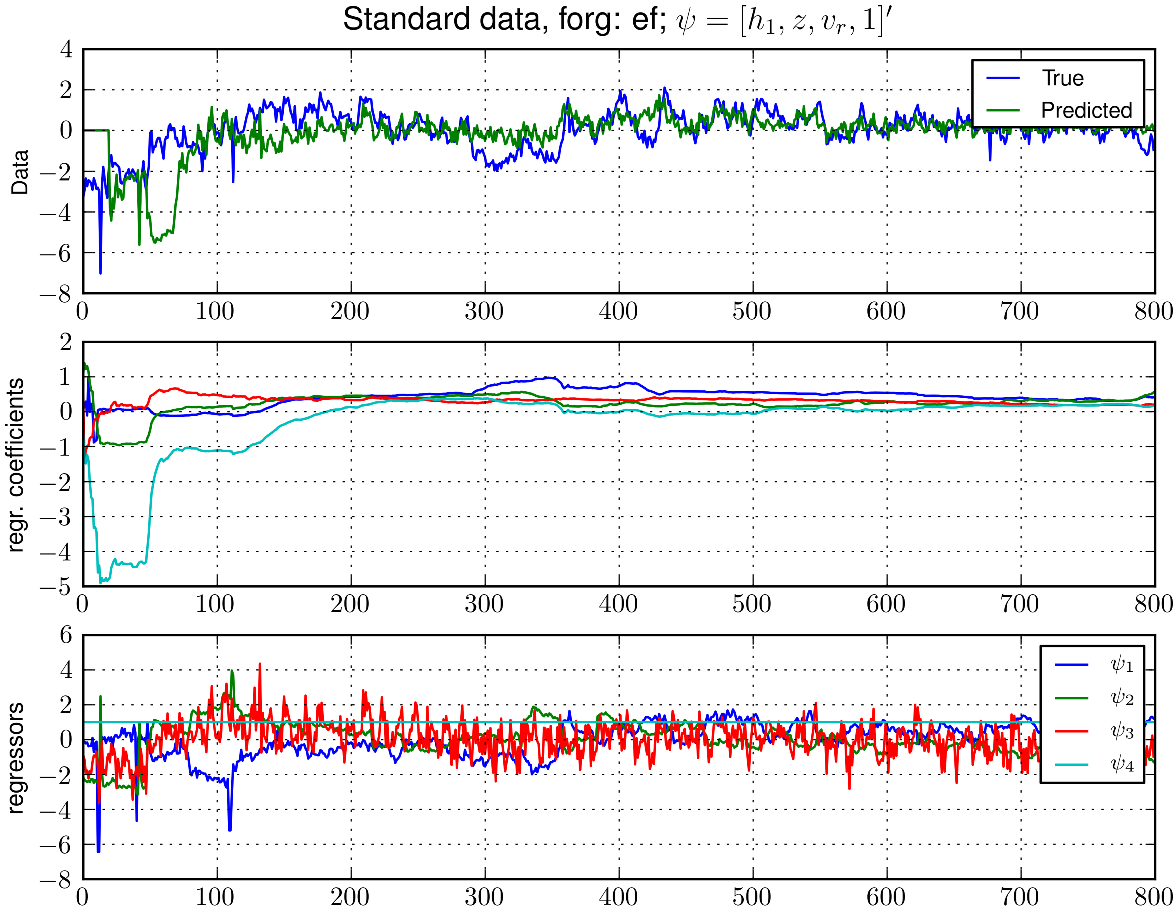
6.4 Blackbox 2 model

6.4.1 No forgetting



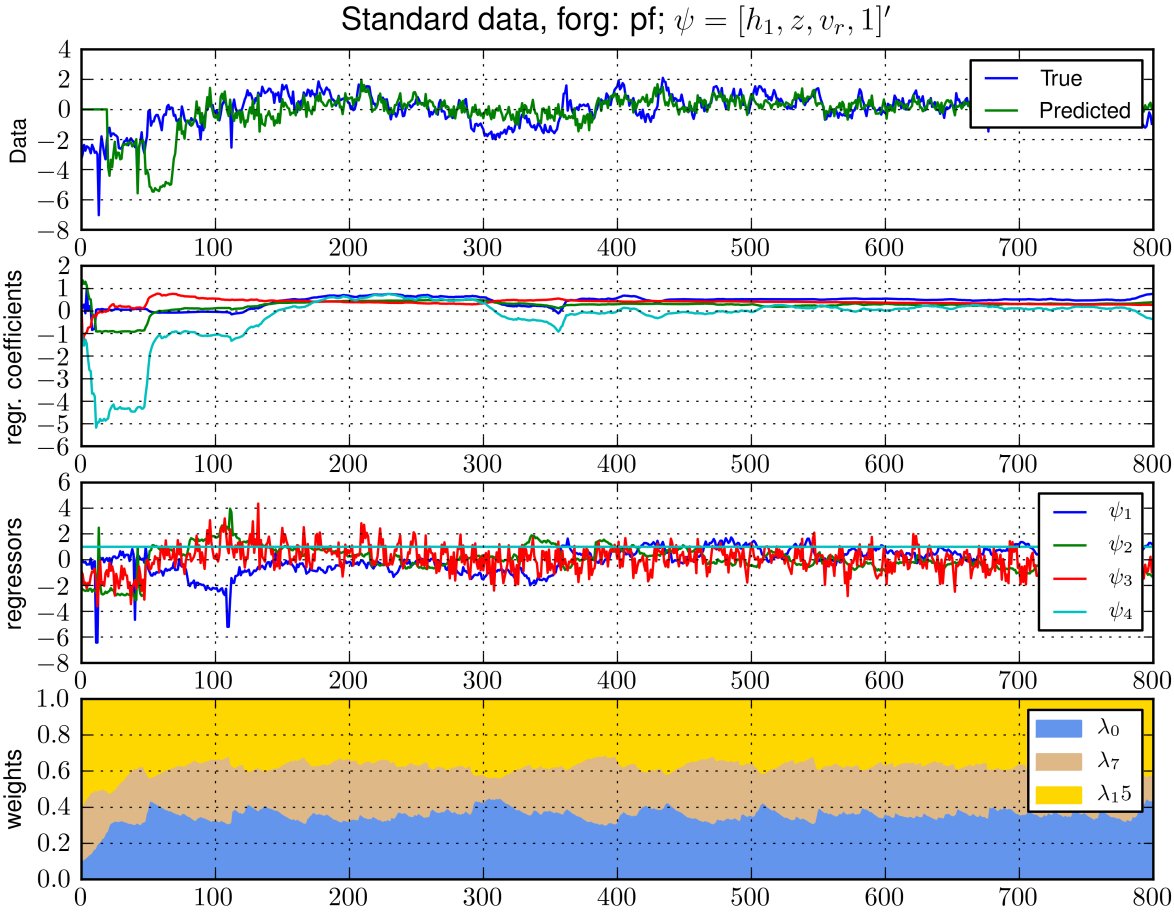
Error statistic	**Value**
Mean	0.313
Median	0.252
Variance	1.34
Standard deviation	1.16
RMSE	1.438

6.4.2 Exponential forgetting



Error statistic	Value
Mean	0.113
Median	0.064
Variance	1.34
Standard deviation	1.16
RMSE	1.354

6.4.3 Partial forgetting



Error statistic	Value
Mean	0.085
Median	0.012
Variance	1.31
Standard deviation	1.14
RMSE	1.315

7 Periodic signal modelling

The periodic signal data batch consisted of data from file *Ki-200703010856_7632171-081-001s_2-example_data_fault_2.mat*. To demonstrate the modelling of interesting data batch, it started from 2000th sample and 800 samples were used. The datafile did not contain the rolling force measurements necessary for the gaugemeter model (7), which restricted the class of available models to the following three members:

- Mass-flow model (6), which was widely discussed in Section 4.1, where its limited theoretical validity was studied.
- Blackbox model 1 (8).
- Blackbox model 2 (9).

The parameter estimation was evaluated:

- Without forgetting, i.e., no time update of parameters pdf was done.
- With exponential forgetting. The forgetting factor was 0.99.
- With partial forgetting. The hypotheses weights were evaluated automatically during the estimation. The factors were:
  - 0.95 for hypotheses' weights,
  - 0.99 for construction of alternative pdf f_A for hypothesis H_1
  - 0.9 for construction of alternative pdf for f_A for hypothesis H_2

Apparently, the periodicity is related to both v_r and z.

7.1 Mass-flow model

7.1.1 No forgetting

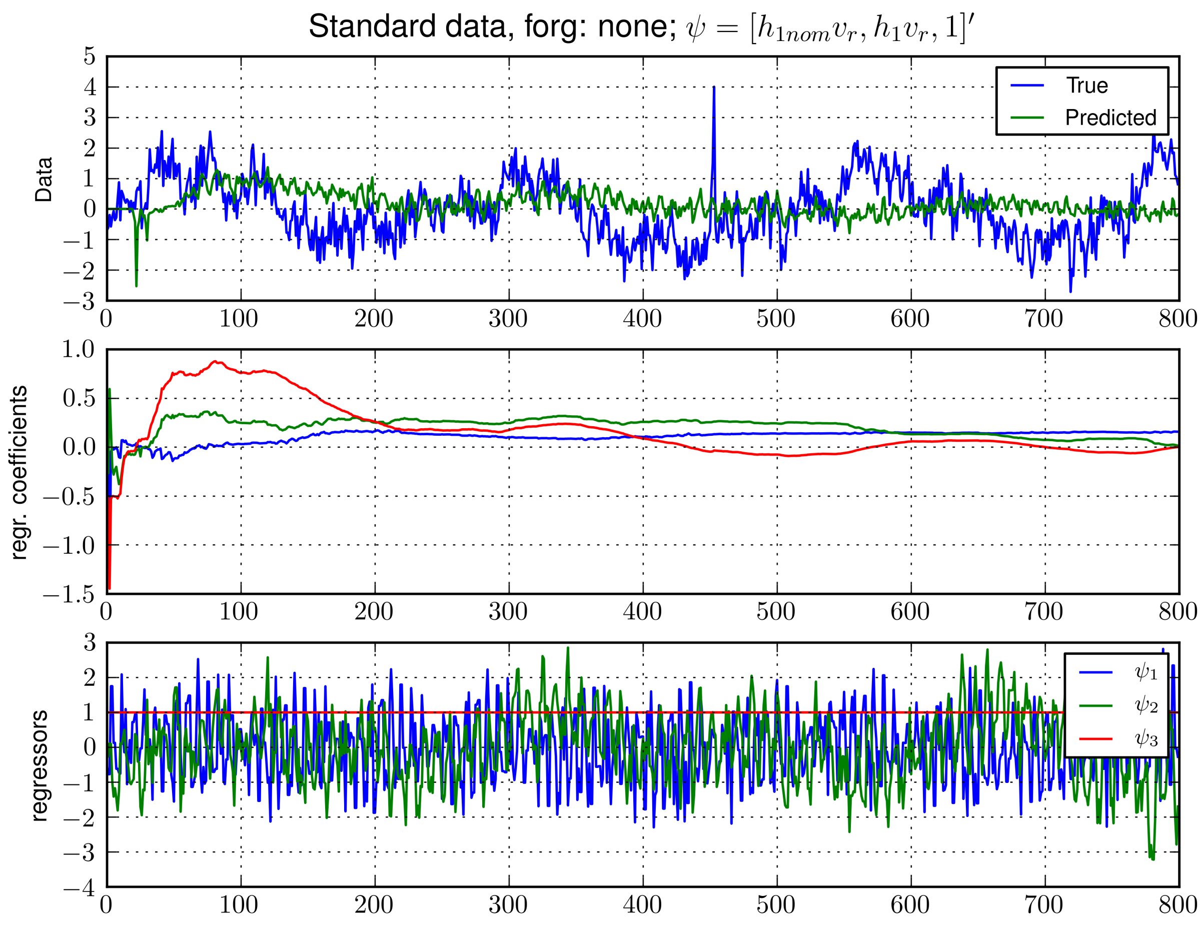
Error statistic	**Value**
Mean	-0.174
Median	-0.265
Variance	1.09
Standard deviation	1.04
RMSE	1.116

7.1.2 Exponential forgetting

Error statistic	**Value**
Mean	-0.031
Median	-0.095
Variance	1.11
Standard deviation	1.05
RMSE	1.106

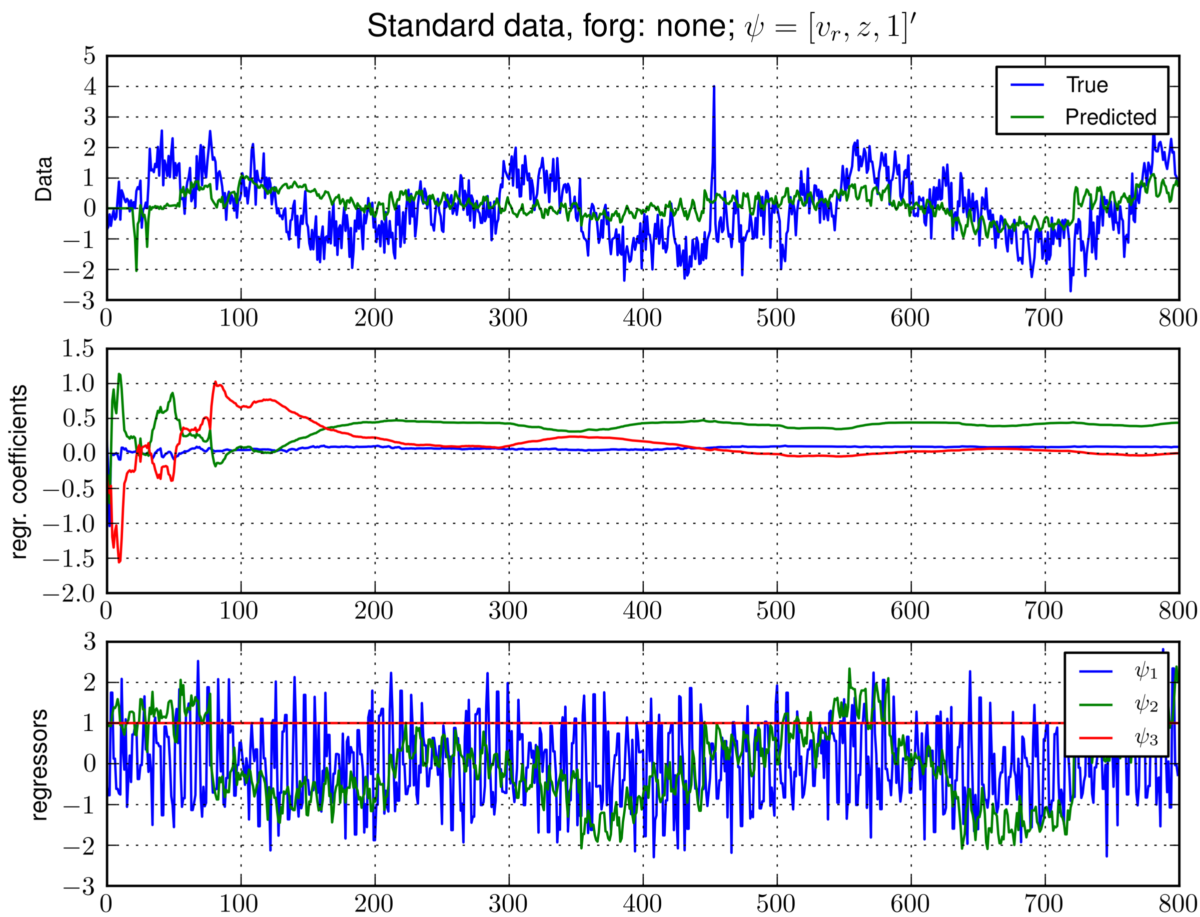
7.1.3 Partial forgetting

Error statistic	Value
Mean	0.033
Median	-0.001
Variance	1.01
Standard deviation	1
RMSE	1.011



7.2 Blackbox 1 model

7.2.1 No forgetting



Error statistic	Value
Mean	-0.137
Median	-0.146
Variance	0.905
Standard deviation	0.951
RMSE	0.924

7.2.2 Exponential forgetting

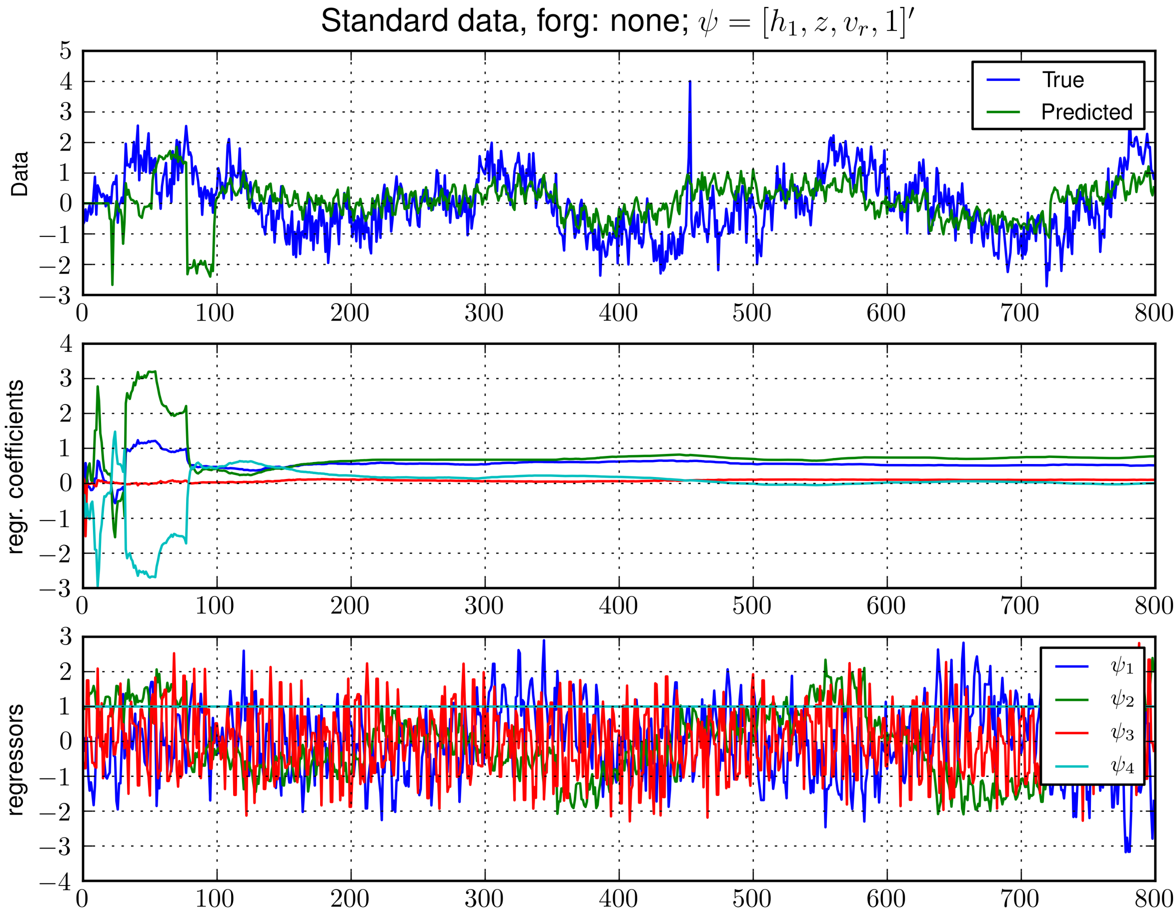
Error statistic	**Value**
Mean	-0.042
Median	-0.084
Variance	1.01
Standard deviation	1
RMSE	1.012

7.2.3 Partial forgetting

Error statistic	**Value**
Mean	0.037
Median	0.016
Variance	0.92
Standard deviation	0.959
RMSE	0.922

7.3 Blackbox 2 model

7.3.1 No forgetting



Error statistic	Value
Mean	-0.064
Median	-0.135
Variance	0.96
Standard deviation	0.98
RMSE	0.965

7.3.2 Exponential forgetting

Error statistic	**Value**
Mean	0.036
Median	-0.023
Variance	0.988
Standard deviation	0.994
RMSE	0.989

7.3.3 Partial forgetting

Error statistic	**Value**
Mean	0.077
Median	-0.023
Variance	0.882
Standard deviation	0.939
RMSE	0.888

8 Periodic signal modelling II

Another periodic signal data batch consisted of data from file *Ki-200903020808_0000000-000-0000f_2_2-example_data_lesser-fault_2.mat*. The data of interest started again from 2000th sample and 800 samples were used. Since the datafile did not contain measurements of the rolling force, the gaugemeter model was not available. Hence, the data were modelled with the three models:

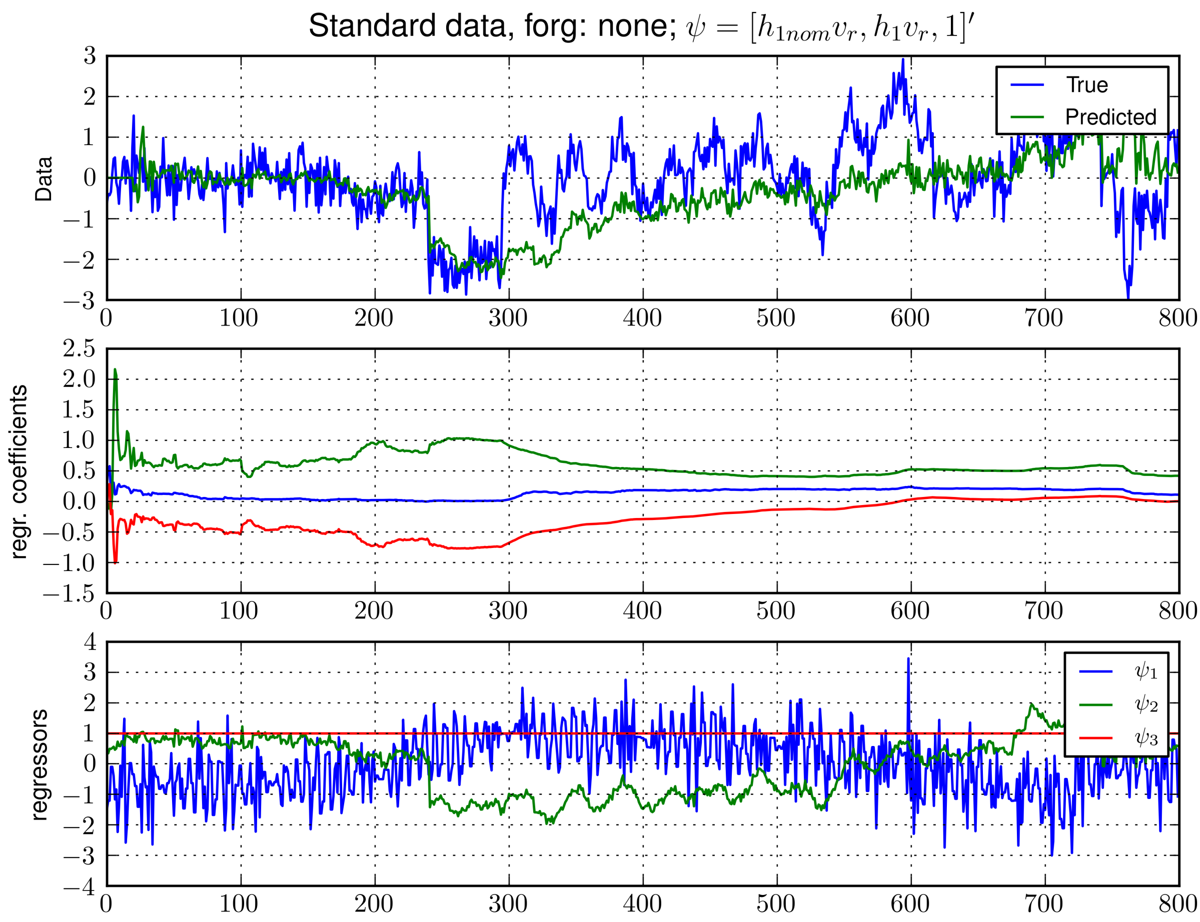
- Mass-flow model (6), which was widely discussed in Section 4.1, where its limited theoretical validity was studied.
- Blackbox model 1 (8).
- Blackbox model 2 (9).

The parameter estimation was evaluated:

- Without forgetting, i.e., no time update of parameters pdf was done.
- With exponential forgetting. The forgetting factor was 0.99.
- With partial forgetting. The hypotheses weights were evaluated automatically during the estimation. The factors were:
  - 0.95 for hypotheses' weights,
  - 0.99 for construction of alternative pdf f_A for hypothesis H_1
  - 0.9 for construction of alternative pdf for f_A for hypothesis H_2

8.1 Mass-flow model

8.1.1 No forgetting



Error statistic	Value
Mean	0.336
Median	0.276
Variance	0.899
Standard deviation	0.948
RMSE	1.013

8.1.2 Exponential forgetting

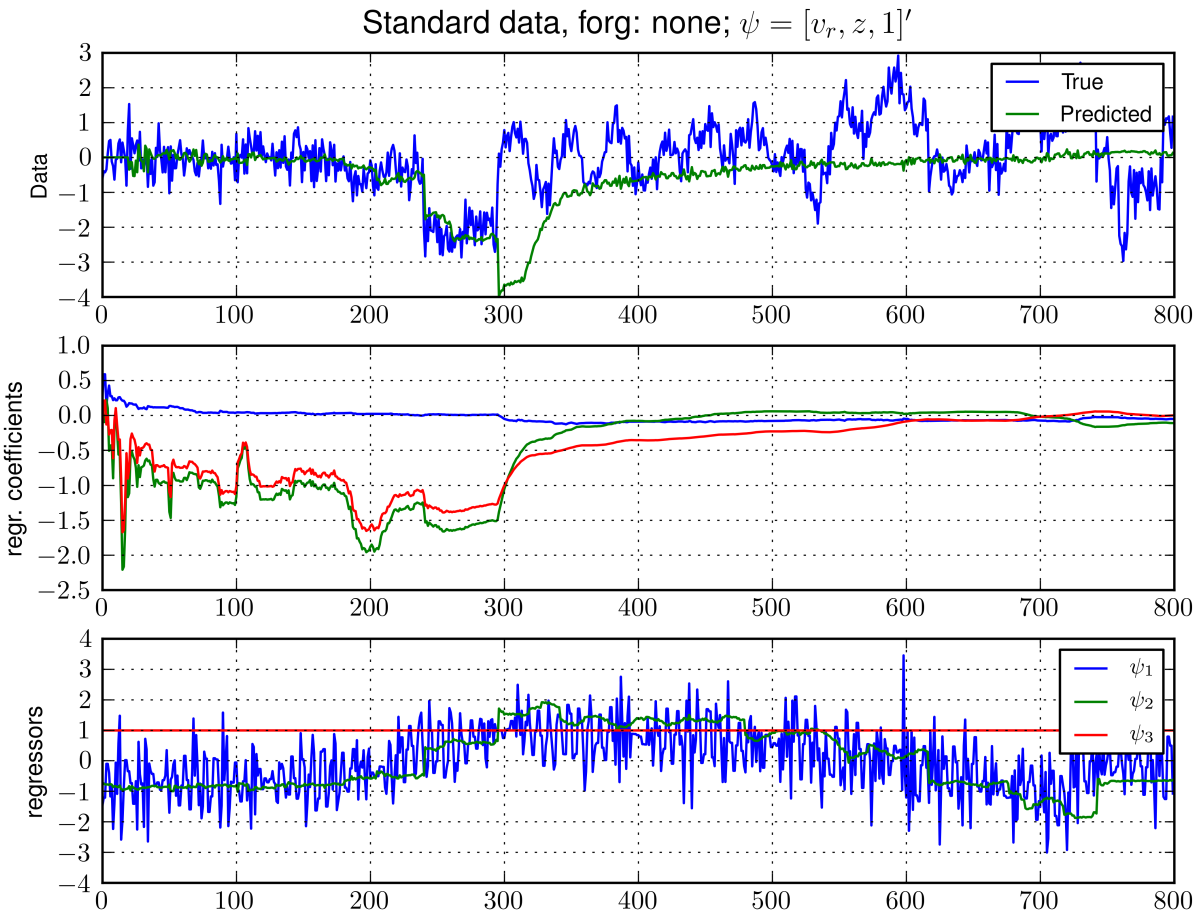
Error statistic	Value
Mean	0.042
Median	0.096
Variance	0.918
Standard deviation	0.958
RMSE	0.920

8.1.3 Partial forgetting

Error statistic	Value
Mean	-0.002
Median	0.030
Variance	0.844
Standard deviation	0.919
RMSE	0.844

8.2 Blackbox 1 model

8.2.1 No forgetting



Error statistic	Value
Mean	0.486
Median	0.324
Variance	1.14
Standard deviation	1.07
RMSE	1.381

8.2.2 Exponential forgetting

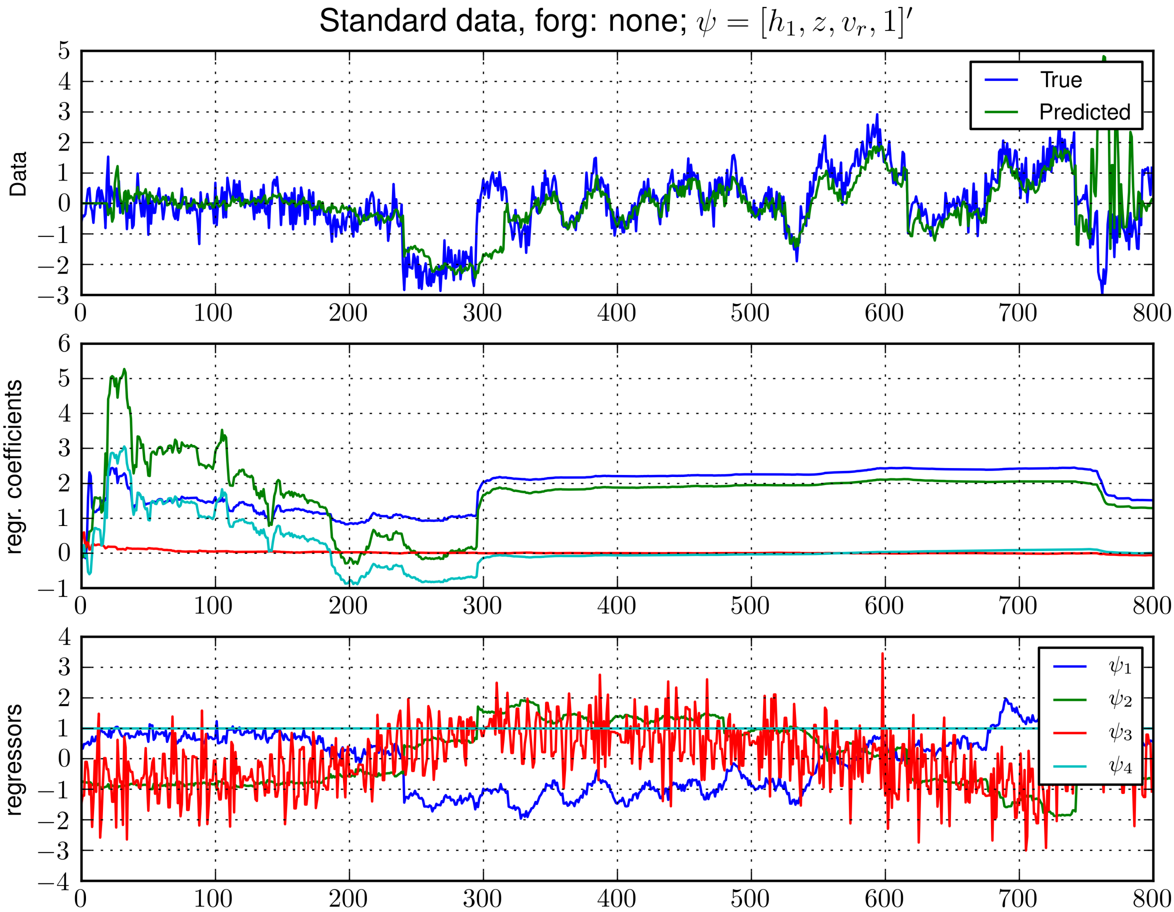
Error statistic	**Value**
Mean	0.218
Median	0.181
Variance	1.35
Standard deviation	1.16
RMSE	1.398

8.2.3 Partial forgetting

Error statistic	**Value**
Mean	0.107
Median	0.100
Variance	1.37
Standard deviation	1.17
RMSE	1.383

8.3 Blackbox 2 model

8.3.1 No forgetting



Error statistic	Value
Mean	0.045
Median	0.105
Variance	0.905
Standard deviation	0.952
RMSE	0.907

8.3.2 Exponential forgetting

Error statistic	**Value**
Mean	-0.078
Median	0.001
Variance	0.796
Standard deviation	0.892
RMSE	0.802

8.3.3 Partial forgetting

Error statistic	**Value**
Mean	-0.071
Median	0.002
Variance	0.85
Standard deviation	0.922
RMSE	0.855

9 Isolated outliers in the signal

A data batch containing isolated outliers consisted of data from file *Vt-200909211137-example_data_fault_3.mat*. The data of interest started from 23300th sample and 800 samples were used. The modelling with this datafile was restricted to just two models – the datafile did not contain measurements necessary for both the mass-flow model and the blackbox 2 model.

- Blackbox model 1 (8).
- Gaugemeter model (7).

The parameter estimation was evaluated:

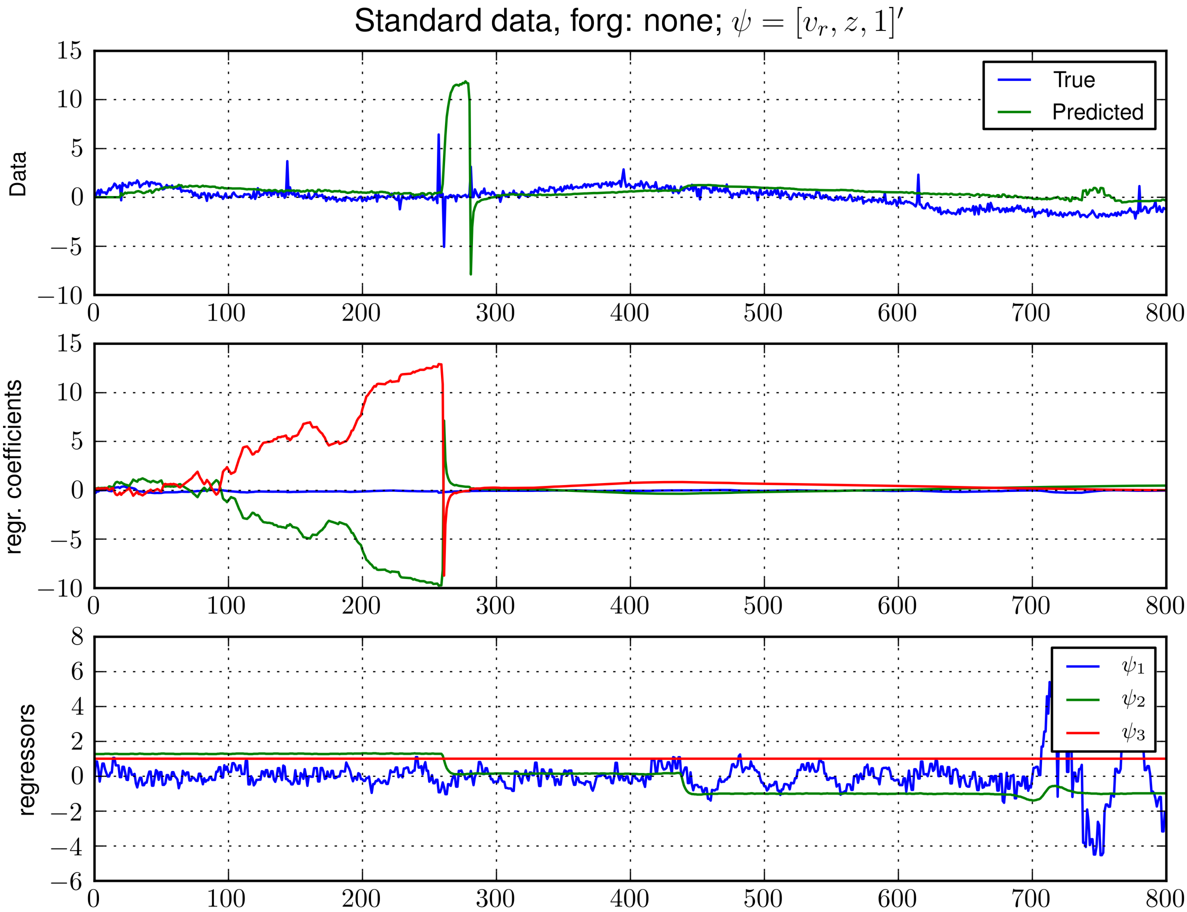
- Without forgetting, i.e., no time update of parameters pdf was done.
- With exponential forgetting. The forgetting factor was 0.99.
- With partial forgetting. The hypotheses weights were evaluated automatically during the estimation. The factors were:
  - 0.95 for hypotheses' weights,
  - 0.99 for construction of alternative pdf f_A for hypothesis H_1
  - 0.9 for construction of alternative pdf for f_A for hypothesis H_2

The modelling results are not very satisfactory here. Apparently, the models used were very sensitive to the outliers, which caused significant immediate deterioration of modelling quality. However, the stabilization is still very fast. This example shows, that it would be reasonable to consider some kind of outlier filtration.

If we focus on the particular models, we may conclude, that the difficulties are probably connected with the regressor z as well. Its constant setting suppresses its use in modelling, while its forgetting may potentially cause problems with numerical stability.

9.1 Blackbox 1 model

9.1.1 No forgetting



Error statistic	Value
Mean	-0.713
Median	-0.587
Variance	3.48
Standard deviation	1.87
RMSE	3.989

9.1.2 Exponential forgetting

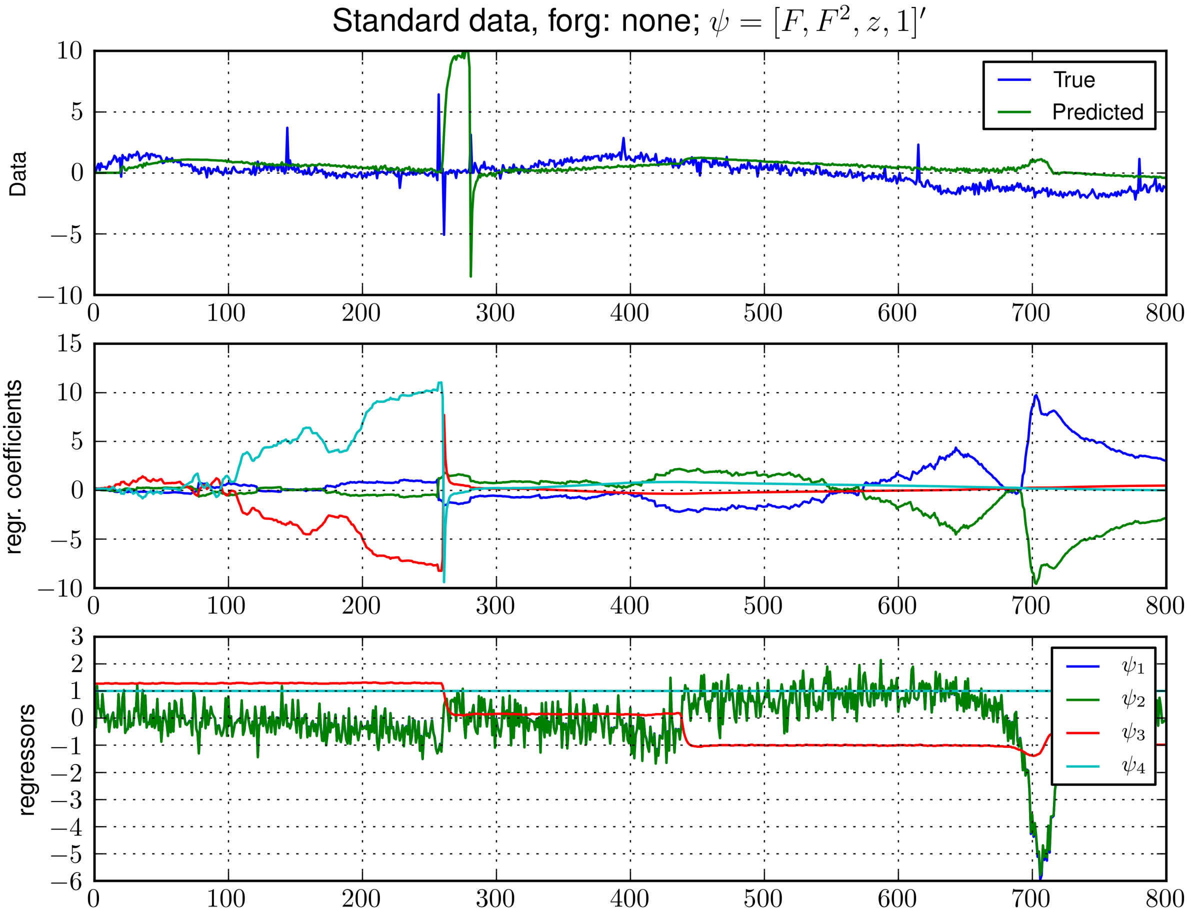
Error statistic	**Value**
Mean	-0.546
Median	-0.448
Variance	3.42
Standard deviation	1.85
RMSE	3.716

9.1.3 Partial forgetting

Error statistic	**Value**
Mean	-0.683
Median	-0.370
Variance	10.8
Standard deviation	3.28
RMSE	11.223

9.2 2nd order gaugemeter model

9.2.1 No forgetting



Error statistic	**Value**
Mean	-0.659
Median	-0.556
Variance	2.74
Standard deviation	1.65
RMSE	3.170

9.2.2 Exponential forgetting

Error statistic	**Value**
Mean	-0.513
Median	-0.441
Variance	2.95
Standard deviation	1.72
RMSE	3.210

9.2.3 Partial forgetting

Error statistic	**Value**
Mean	-0.280
Median	-0.411
Variance	18.2
Standard deviation	4.26
RMSE	18.266

10 Outliers (dirty strip)

10.1 Brief analysis

The course of thickness deviation h_2 due to dirty strip is depicted in Fig. 12. The difficult part starts around the instant $t = 1500$ and continues until around 2500. In comparison to isolated outliers, here the situation gets much worse. The term *outlier* is rather abused – the improper data form a significant signal component, which cannot be filtered out by the basic means.

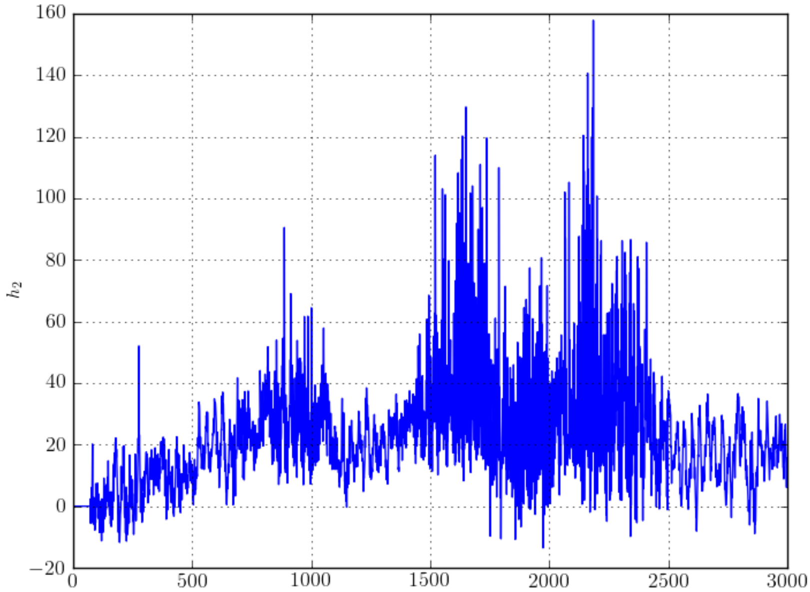


Figure 12: Course of h_2 (dirty strip).

To evaluate a brief analysis, let us focus on the signal power spectral properties. The Figure 13 shows signal power spectra for various data files, namely:

(A, B) – *Vi-200909171440_0_2-example_data_fault_4.mat* – the dirty strip data, i.e., a case relevant in this section.

(C, D) – *Vi-201003201554_2011507_3-example_data_fault_4-free.mat* – example of a file free of fault of the same type.

(E, F) – *Gi-200810211323_206-10-B_3.mat* – the regular and correct data.

(G, H) – *Vt-200909211137-example_data_fault_3.mat* – the impact of isolated outliers on thickness deviation.

(I, J) – *Vt-201003220733-example_data_fault_3-free.mat* – example of a file free of fault of the same type.

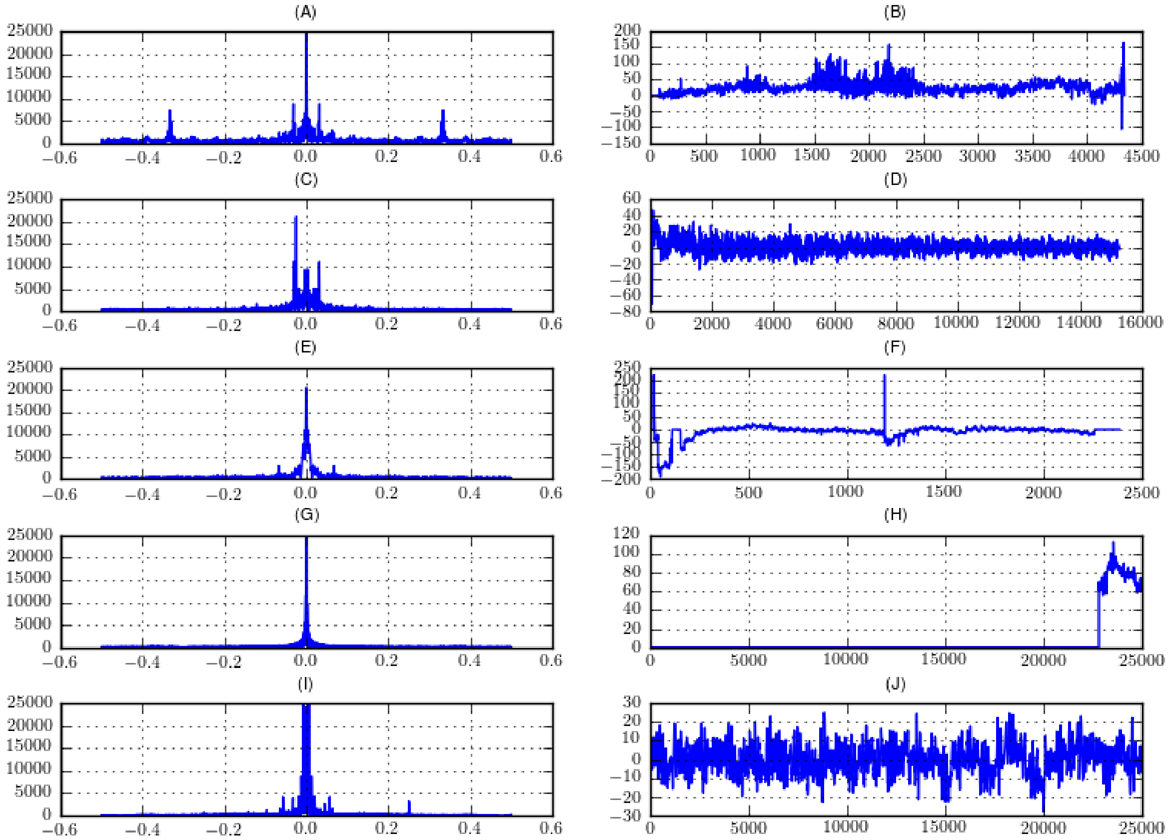


Figure 13: Power spectral densities (left) and the course of data (right).

The lengths of data are not the same for each data file, which can potentially affect the low-frequency components of the signal. The vertical axes are limited by 25 000, which can truncate some 0-peaks, however, this does not limit the information content of the figure. Notice the high-frequency band around ± 0.35 in Fig. (A), which do not have their counterparts in other files. The shape indicates, that there exist frequencies, typical for this type of fault, which considerably influence the signal and which will almost surely cause deterioration of the modelling quality. It would be advisable to either pre-process the data concerned by a suitable filter (e.g., low-pass one) or to react by other means.

11 Alternative blackbox models

In this section, we discuss the searching for alternative blackbox models. This class of models, introduced in the preliminary sections, are characteristic with their ignorance of explicit physical character of the rolling process. They are constructed ad-hoc, using potentially suitable combinations of measured variables as regression vector elements. It can happen, that a chosen combination models well the regressand (output thickness deviation), on the other hand, the opposite situation can occur as well. Therefore, a thorough analysis of suitability of a chosen combination is strongly advisable.

11.1 Settings

For our purpose, we finally carried out a brute-force analysis of all possible combinations of measured variables, used as elements of tested regression vectors. We used the same standard data file as in some previous sections, namely *Gi-200810211323_206-10-B_3.mat*. The data window started with 1250th measurement and was 800 data long. The traffic delay was 19 time steps. The suitable variables, stored in the data file, were:

h_1 – input thickness deviation

H_{1nom} – nominal input thickness

H_{2nom} – nominal output thickness

v_1 – input strip speed

v_2 – output strip speed

v_r – ratio v_1/v_2

z – uncompensated rolling gap

F – rolling force

I – main drive current

The total count of the tested cases (for $n = 9$ variables) was

$$\binom{n}{1} + \binom{n}{2} + \ldots + \binom{n}{n} = 511. \tag{12}$$

This means, that the shortest regression vector used one measured variable (of nine possible), while the longest one used all nine. Additionally, all tested vectors were extended by the absolute term. The models used three types of estimation procedure:

1. no forgetting
2. exponential forgetting with factor 0.99
3. partial forgetting with online weights determination and alternative pdf derived by exponentiation of the posterior by 0.99 for hypothesis 'all parameters vary' and 0.9 for hypothesis 'absolute term varies'. The online determined weights were forgotten with factor 0.95.

The initial statistics of the models were $\nu_0 = 10$ and extended information matrix V_0 with $[0.1, 0.01, \ldots, 0.01]$ on diagonal. These settings, i.e., regression vectors, initial statistics and estimation methods, were used both for normalized data and raw data. This led to $511 \times 3 \times 2 = 3066$ tested cases, which were consequently compared using their log-likelihood.

11.2 Computing environment

The number of tested cases, in combination with the growing number of numerical operations in partial forgetting when the length of regression vector increased, represented a computationally challenging task. To shorten the time necessary for the evaluation, we decided to use the python programming language with the numerical and scientific libraries 'numpy' and 'scipy' and the module 'parallel-python' for cluster computation, the modelling itself was carried out with the 'pybamo' module. The computing cluster consisted of three computers with 2, 4 and 8 Intel processor cores, i.e., 14 cores totally. The results (log-likelihoods) were stored in a sqlite database, which simplified their processing and significantly reduced the risk of their corruption.

11.3 Results and proposals

The following two tables – Tab. 4 and 5 contain top-30 results ordered by log-likelihoods of models with exponential (EF) and partial (PF) forgetting; apparently, the underlaying physical principle (cf. mass-flow model) plays still important role and models regarding evolution of strip thickness and speed represent the best members of a class of suitable models. These results may be taken into account as a helpful source for selection of the most suitable regression models. However, it is advisable to avoid choosing models with very similar structure, because a dropout of one measured variable could discard a lot (if not all) of them.

A convenient way to selection of suitable models consists in employing of two principles – Occam's window and Occam's razor [5]. The former one rejects those models, that predict far less well than the best model. Formally, models not belonging to

$$\mathcal{A} = \left\{ M_k; \frac{\max \mathsf{probability}(M_l|d(t))}{\mathsf{probability}(M_k|d(t))} \leq \mathcal{C} \right\},$$

where $\mathcal{C}$ is chosen by the user, should be excluded from the class of used models. The Occam's razor excludes those models, whose structure is more complex than that one of the simpler models, but which receive less support. Formally

$$\mathcal{B} = \left\{ M_k; M_l \subset M_k, \frac{\mathsf{probability}(M_l|d(t))}{\mathsf{probability}(M_k|d(t))} > 1 \right\}$$

should be excluded.

combination	none norm.	none raw	EF norm	EF raw	PF norm	PF raw
$(h_{2nom}, h_{1nom}, v_1)$	-525.201	-2326.421	-27.371	-256.934	-21.752	—
$(h_{2nom}, h_{1nom}, v_1, v_2, v_r)$	-504.018	-2308.037	-28.722	-252.371	-25.662	—
$(h_{2nom}, h_{1nom}, v_1, v_2)$	-502.818	-2316.467	-29.652	-255.952	-25.145	—
$(h_1, h_{2nom}, h_{1nom}, z, v_1)$	-482.05	-2286.65	-29.797	-263.998	-32.973	—
$(h_{2nom}, h_{1nom}, z, v_1)$	-530.539	-2334.07	-29.841	-261.615	-27.721	-113.297
$(h_1, h_{2nom}, h_{1nom}, v_1)$	-493.522	-2297.375	-30.934	-262.926	-29.027	—
$(h_1, h_{2nom}, h_{1nom}, z, v_1, v_2, v_r)$	-439.547	-2257.35	-30.978	-259.29	-30.979	—
$(h_{2nom}, h_{1nom}, v_1, v_r)$	-504.602	-2307.469	-30.991	-255.539	-26.486	—
$(h_{2nom}, h_{1nom}, v_2, v_r)$	-504.963	-2316.702	-31.079	-255.961	-26.601	—
$(h_{2nom}, h_{1nom}, z, v_1, v_2, v_r)$	-503.912	-2313.164	-31.103	-256.965	-31.268	—
$(h_{2nom}, h_{1nom}, v_1, I)$	-522.125	-2327.428	-31.132	-263.78	-25.662	—
(h_{2nom}, v_1)	-525.201	-2325.829	-31.391	-260.362	-23.407	-105.656
(h_{1nom}, v_1)	-525.201	-2326.238	-31.391	-260.771	-23.407	-105.827
$(h_{2nom}, h_{1nom}, F, v_1)$	-477.113	-2276.062	-31.449	-258.345	-30.861	—
$(h_1, h_{2nom}, h_{1nom}, z, v_1, v_2)$	-443.208	-2268.33	-31.929	-262.867	-35.854	—
$(h_{2nom}, h_{1nom}, z, v_1, v_2)$	-503.107	-2323.056	-32.034	-260.544	-30.771	—
$(h_1, h_{2nom}, h_{1nom}, v_1, v_2, v_r)$	-441.218	-2260.022	-32.297	-258.374	-32.339	—
$(h_{2nom}, h_{1nom}, v_1, v_2, v_r, I)$	-508.59	-2314.765	-32.507	-259.241	-29.146	—
$(h_{2nom}, v_1, v_2, v_r)$	-504.018	-2307.445	-32.743	-255.8	-27.285	—
$(h_{1nom}, v_1, v_2, v_r)$	-504.018	-2307.854	-32.743	-256.208	-27.285	—
$(h_{2nom}, h_{1nom}, F, v_1, v_2, v_r)$	-420.576	-2233.177	-32.779	-253.762	-34.3	—
$(h_{2nom}, h_{1nom}, z, v_1, I)$	-521.622	-2330.79	-33.014	-267.872	-31.573	—
$(h_1, h_{2nom}, h_{1nom}, z, v_1, I)$	-461.156	-2271.818	-33.073	-270.358	-37.618	—
$(h_1, h_{2nom}, h_{1nom}, v_1, v_2)$	-444.646	-2274.325	-33.227	-261.955	-31.96	—
$(h_1, h_{2nom}, h_{1nom}, z, v_1, v_r)$	-445.912	-2258.016	-33.267	-262.457	-37.202	—
$(h_{2nom}, h_{1nom}, z, F, v_1)$	-438.709	-2237.87	-33.349	-262.456	-35.455	—
$(h_1, h_{2nom}, h_{1nom}, z, v_2, v_r)$	-446.473	-2267.469	-33.361	-262.88	-37.315	—
$(h_{2nom}, h_{1nom}, z, v_1, v_r)$	-505.049	-2312.707	-33.372	-260.133	-32.112	—
$(h_{2nom}, h_{1nom}, v_1, v_2, I)$	-507.234	-2322.18	-33.438	-262.822	-28.638	—
$(h_{2nom}, h_{1nom}, z, v_2, v_r)$	-505.481	-2323.225	-33.462	-260.552	-32.228	—

Table 4: 30 best regression vectors (without abs. term), ordered by EF (normalized regressors)

combination	none norm.	none raw	EF norm	EF raw	PF norm	PF raw
$(h_{2nom}, h_{1nom}, v_1)$	-525.201	-2326.421	-27.371	-256.934	-21.752	—
(h_{2nom}, h_{1nom})	-707.756	-2508.706	-37.383	-270.522	-22.064	—
(h_{2nom}, v_1)	-525.201	-2325.829	-31.391	-260.362	-23.407	-105.656
(h_{1nom}, v_1)	-525.201	-2326.238	-31.391	-260.771	-23.407	-105.827
$(h_{2nom},)$	-707.756	-2508.115	-41.403	-273.95	-23.77	—
$(h_{1nom},)$	-707.756	-2508.523	-41.403	-274.359	-23.77	—
$(h_{2nom}, h_{1nom}, v_r)$	-612.808	-2433.684	-33.593	-261.718	-24.299	—
$(h_{2nom}, h_{1nom}, v_2)$	-569.09	-2368.225	-39.134	-269.01	-24.832	—
$(v_1,)$	-525.201	-2320.126	-35.411	-257.014	-25.083	-105.866
$(h_{2nom}, h_{1nom}, v_1, v_2)$	-502.818	-2316.467	-29.652	-255.952	-25.145	—
$(h_{2nom}, h_{1nom}, v_1, I)$	-522.125	-2327.428	-31.132	-263.78	-25.662	—
$(h_{2nom}, h_{1nom}, v_1, v_2, v_r)$	-504.018	-2308.037	-28.722	-252.371	-25.662	—
(h_{2nom}, v_r)	-612.808	-2433.092	-37.613	-265.146	-25.991	—
(h_{1nom}, v_r)	-612.808	-2433.501	-37.613	-265.555	-25.991	—
(h_{2nom}, h_{1nom}, I)	-694.04	-2498.072	-41.401	-277.623	-26.041	—
$(h_{2nom}, h_{1nom}, v_1, v_r)$	-504.602	-2307.469	-30.991	-255.539	-26.486	—
(h_{2nom}, v_2)	-569.09	-2367.634	-43.154	-272.438	-26.519	—
(h_{1nom}, v_2)	-569.09	-2368.042	-43.154	-272.847	-26.519	—
$(h_{2nom}, h_{1nom}, v_2, v_r)$	-504.963	-2316.702	-31.079	-255.961	-26.601	—
(h_{2nom}, v_1, v_2)	-502.818	-2315.876	-33.672	-259.38	-26.794	—
(h_{1nom}, v_1, v_2)	-502.818	-2316.285	-33.672	-259.789	-26.794	—
(h_{2nom}, h_{1nom}, z)	-670.226	-2473.385	-35.989	-271.342	-27.068	—
$(h_{2nom}, v_1, v_2, v_r)$	-504.018	-2307.445	-32.743	-255.8	-27.285	—
$(h_{1nom}, v_1, v_2, v_r)$	-504.018	-2307.854	-32.743	-256.208	-27.285	—
(h_{2nom}, v_1, I)	-522.125	-2326.836	-35.152	-267.208	-27.316	—
(h_{1nom}, v_1, I)	-522.125	-2327.245	-35.152	-267.617	-27.316	—
$(h_{2nom}, h_{1nom}, z, v_1, v_2, v_r, I)$	-503.649	-2315.534	-34.328	-263.274	-27.603	—
$(v_r,)$	-612.808	-2433.213	-41.634	-261.794	-27.702	-106.262
$(h_{2nom}, h_{1nom}, z, v_1)$	-530.539	-2334.07	-29.841	-261.615	-27.721	-113.297

Table 5: 30 best regression vectors (without abs. term), ordered by PF (normalized regressors)

12 Conclusion

This report demonstrated the use of various models defined by their structures and estimators. The impact of the forgetting techniques is shown in the related figures. The situations was slightly complicated by the fact that the regressors are not available for each model. Therefore, in some cases, only a subset of defined models was used.

In the first few sections, the analysis of the impact of data normalization was studied. It has been shown, that the normalization to $\mathcal{N}(0, 1)$ can lead to significantly improved estimation. The validity of the mass-flow model was analyzed then. It revealed its limitation, indicated by a low correlation between the left and right-hand side of the continuity equation. This situation deserves further analysis, related to the physical laws.

The main part of the report contains comparison of the models equipped with various forgetting methods. The figures are accompanied by the statistics of the prediction errors. Based on the situation, it seems reasonable to use the partial forgetting for certain cases.

The last part discusses the difficulties of dirty strip thickness deviation modelling and briefly brings the possible computational-intensive search for the best blackbox models.

References

[1] Dedecius, K. et al. (2010). *Parameter Tracking with Partial Forgetting Method.* Submitted to IFAC Automatica.

[2] Ettler, P. and Andrýsek, J. (2007). *Mixing models to improve gauge prediction for cold rolling mills.* In Preprints of the 12th IFAC Symposium on Automation in Mining, Mineral and Metal Processing. Québec City: Université Laval.

[3] Jazwinski, A.H. (1970). *Stochastic processes and filtering theory.* New York: Academic Press.

[4] Kárný, M. et al. (2005). *Optimized Bayesian Dynamic Advising*, Springer.

[5] Madigan, D. and Raftery, A.E. (1994) *Model selection and accounting for model uncertainty in graphical models using Occam's window.* Journal of the American Statistical Association, vol. 89, no. 428, pp. 1535–1546.

[6] Milek, J.T. (1995) *Stabilized Adaptive Forgetting in Recursive Parameter Estimation.* vdf Hochschulverlag AG.

[7] Peterka, V. (1981). *Bayesian Approach to System Identification*, in *Trends and Progress in System Identification*, P. Ekhoff, Ed., pp. 239–304. Pergamon Press, Oxford.