Data-Driven Modeling for Industrial Plants Based on Non-Stationary and Sparsely Sampled Data Streams

Changrui Xie,a Yujie Hu,a Lingyu Zhu,b Han Gong,c Hongyang Li,c Xi Chen a

a State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou, Zhejiang, 310027, China

bCollege of Chemical Engineering, Zhejiang University of Technology, Hangzhou, Zhejiang, 310014, China

cZhejiang Amino-Chem Co., Ltd, Shaoxing, Zhejiang, 312369, China

[*xi\_chen@zju.edu.cn*](mailto:xi_chen@zju.edu.cn)

Abstract

In this paper, a robust auto-regressive exogenous regression model and the associated parameter inference algorithm is developed. Student’s- distribution is adopted to accounted for the noise in dealing with outliers. The parameter inference is conducted through the Streaming Variational Bayes (SVB) approach based on sparsely sampled data streams. To address the non-stationarity in data streams, power priors are established for each streaming batch by means of the exponential weighting mechanism. The uncertainty of parameter estimates is accounted for by formulating the problem under a full Bayesian framework. Furthermore, an analytical posterior predictive distribution is approximately derived, enabling the model to provide not only a point estimate but also the associated predictive uncertainty. The effectiveness of the proposed model and the adaptation algorithm is validated through an industrial distillation process.

**Keywords**: Streaming variational inference, power priors, Student’s-t distribution, non-stationary data streams, sparse sampling.

* 1. Introduction

In industrial processes, quality variables play a crucial role in process monitoring and production optimization. Traditional offline measurement has substantial delays. Data-driven modeling techniques offer an alternative to soft sensing by leveraging easily measurable process variables to model quality variables.1 Industrial plants continuously generate data streams during production, presenting a challenge in adapting data-driven models to these evolving real-time streams. And the variations in operation conditions can introduce shifts in data distribution, making global models unsuitable for real-time applications. Several adaptive mechanisms, including the moving, window approach, the recursive approach and the Just-In-Time learning approach, have been used to online update the data-driven model.2-4 Nevertheless, sparse sampling of quality variables and the associated problem of data scarcity pose a considerable challenge to model adaptation. In such circumstances, probabilistic approaches can remedy this issue by accounting for the uncertainty in limited data, yielding more generalizable prediction.5 They not only provide a point estimate for the quality variable, but also give the associated predictive uncertainty through the posterior predictive distribution. Streaming Variational Bayes (SVB) has been proposed for parameter inference in probabilistic model base on streaming data.6 It can naturally serve as a Bayesian recursive updating approach for probabilistic models. However, SVB’s assumption of data interchangeability across different batches makes it unsuitable for handling non-stationary streams. The power prior approach mitigates the limitations of SVB through an exponential forgetting mechanism.7 Additionally, the presence of outliers in industrial data streams may cause a considerable impact on the model adaptation. A prevalent method to deal with this problem is using Student -noise to account for the effect of outliers.8 Student’s- distribution has a longer tail for explaining the noise and an adjustable parameter named Degree of Freedom (DOF), making it robust to outliers. In this work, we develop a robust probabilistic model for industrial plants by using Student’s-distribution, which is initially identified through variational inference (VI), and continually updated through SVB with power prior approach. Therefore, our model can adapt to variations in the operation condition by using non-stationary and sparsely sampled data streams. Its robustness helps avoid the adverse effects of outliers on model adaptation. The processes of model identification and adaptation are conducted within a full Bayesian framework.

* 1. Robust Regression Model with SVB and Power Priors
     1. Model Development

Due to the data scarcity caused by the sparse sampling of quality variables, a simple Bayesian linear regression model is preferred to build the mapping from to :

|  |  |
| --- | --- |
|  |  |

where denotes the input variables, denotes the regression weights, corresponds to the sampling instance, and is the noise of the model. In this study, Student’s- distribution is employed to account for the contaminating noise:

|  |  |
| --- | --- |
|  |  |

where denotes the precision (inverse of variance) and denotes the degrees of freedom. is the Gamma function. Student's- distribution can be decomposed into an infinite mixture of scaled Gaussian distributions. Therefore, the likelihood function of the regression model can be yielded as

|  |  |
| --- | --- |
|  |  |

where denotes the Gaussian distribution with mean and variance , denotes the Gamma distribution, and is the intermediate latent variable. A hierarchical Normal-Gamma prior is assigned to the weights and precision :

|  |  |
| --- | --- |
|  |  |

where is the precision matrix parameterizing the Normal distribution of . We assume to be a diagonal matrix such that , where the vector is unknown. This sparse configuration is known as automatic relevance determination (ARD). Conjugate priors are also chosen for and the DOF :

|  |  |
| --- | --- |
|  |  |

Therefore, the joint probability over all random variables, including , the local latent variable and the global variables , can be described hierarchically as

|  |  |
| --- | --- |
| . |  |

* + 1. Parameter Inference

In this study, mean-field variational inference is employed for parameter inference. Thus, the variational posterior can be approximately factored as

|  |  |
| --- | --- |
|  |  |

Through the variational inference, the variational posterior can be yielded as

|  |  |
| --- | --- |
|  |  |

where the hyperparameters in (8) can be calculated by

|  |  |
| --- | --- |
|  |  |
|  |  |

As for ARD parameter , the update equation for each is given as

|  |  |
| --- | --- |
|  |  |

Similarly, the variational posterior distribution of the latent variable can also be yielded. The variational posterior distribution of the DOF is approximated by utilizing the Stirling's Series.9 The variational posterior distributions of and are summarized as:

|  |  |
| --- | --- |
|  |  |

where

|  |  |  |
| --- | --- | --- |
|  |  | |
|  | |  |

The expectation operations are denoted by in the aforementioned equations. These equations are iteratively applied to search for the optimal variational posterior distributions over parameters . After each iteration, the lower bound is evaluated as an indicator of training convergence:

|  |  |
| --- | --- |
|  |  |

* + 1. Online Update with SVB and Power Priors

SVB with power prior approach serves as an adaptation mechanism for the regression model, making it capable of accommodating the non-stationarity in streaming data. Upon receiving a new batch , the power priors are initially computed and subsequently used for parameter inference within the VI framework. The size of one batch, denoted as , may be extremely small due to sparse sampling. The power priors for the batch is constructed by the exponentially weighted combination of uninformative priors and variational posteriors obtained in the batch. Given the predefined forgetting factor , the power priors for , and for the batch can be obtained as

|  |  |
| --- | --- |
|  |  |
|  |  |

where and ( is in ) are parameters of the posteriors for the batch; and are parameters of the uninformative priors; and are parameters of the desired power priors for the batch. Using the power priors, the update equations for and can be obtained as

|  |  |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |

The update equations for the latent variable remain the same as (13). The convergence of the training procedure can also be monitored by the .

* + 1. Posterior Predictive Distribution

Given a new, unseen sample , the proposed model provides a posterior predictive distribution for unknown as:

|  |  |  |
| --- | --- | --- |
|  | |  |
|  | |  |
|  | |  |
|  | |  | |
|  | |  | |

where denotes the training set. Note that in (22), we substitute the latent variable with its (Maximum-A-Posterior) estimate in order to derive an analytical solution. Ultimately, the mean and variance of the prediction are given by

|  |  |
| --- | --- |
|  |  |

When an unseen sample is received, the posterior predictive distribution for can be calculated according to (24) using the latest posteriors of parameters.

* 1. Industrial Application

In this work, a distillation column is employed to evaluate the proposed model. A dataset spanning 25 days was collected from the plant and subsequently partitioned into training, testing, and validation sets in a 2:2:1 ratio by time. We aim to predict the yield concentration at the top of column, named . The process variables, involving eight

Table 1. Numerical metrics of the predictions given by five models.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Methods | LS | RLS | SVB-PP | Robust SVB | Robust SVB-PP |
|  | 191.24 | 155.88 | 97.00 | 207.03 | 91.63 |
|  | 156.25 | 119.43 | 75.59 | 145.18 | 69.13 |

temperatures at different trays ( for ) and one inlet flowrate , are recorded at one-minute intervals and used as input variables. The target variable, i.e., concentration, is measured every five minutes. In this work, a lag of 5min is set for the input variables. Therefore, the regression model is constructed as

|  |  |
| --- | --- |
| . |  |

Bayesian neural networks can deal with more complex, non-linear relationships, and model uncertainty by assigning probability distributions to the weights and biases. In this work, we prefer to choose a simple linear regression to model the mapping from the process variables to the target variables, due to data scarcity in streaming batches arising from sparse sampling. The optimal value of the forgetting factor is selected as 0.98 after a grid search between 0.95 and 1. Our proposed robust linear regression model with SVB and power priors (Robust SVB-PP) is compared with other four models including a non-adaptive linear model (LS), the recursive least square (RLS) model, the Robust linear regression model with SVB (Robust SVB) and the linear regression model with SVB and power priors (SVB-PP). To evaluate the efficacy of the model adaptation under sparse sampling, only one sample is used for every 4h from the testing set, serving as a streaming batch for online model updates. Table 1 lists the and of the five models on the testing set. Detailed results as well as predictive uncertainties are presented in Fig. 1, where the sparsely sampled streaming data used for model updates are marked by the star symbols, and three outliers are also noted in the figure.

It is obvious that the robust SVB-PP, SVB-PP and RLS model can address the non-stationarity in the data streams, compared to the LS and Robust SVB model. However, the RLS model and the SVB-PP model are significantly misled by the three outliers, resulting in wrong model updates and terrible predictions. Sparse sampling worsens the problem as it may cost a long time to correct the model back. One potential approach to address this issue is resorting to outlier detection algorithms to identify and exclude outliers before updating the model. However, it is often laboursome to perform outlier detection online. Moreover, the old criterion for outlier elimination, obtained from past batches, may be inappropriate for the new batch due to variations in the industrial platform. And data scarcity may hinder the application of most outlier detection algorithms on a new streaming batch. Instead, our proposed Robust SVB-PP model is inherently robust to the outliers and can respond with high predictive uncertainties. Thereafter, once a normal sample is sampled for model updates, these uncertainties will return to the regular level. These excellent characteristics make our model more robust and informative in industrial applications.

* 1. Conclusion

In this paper, a Streaming Variational Bayes (SVB) with power prior approach was developed for parameter inference in a robust linear regression model. Non-stationary and sparsely sampled data streams were used for online model adaptation within a full Bayesian framework. The effectiveness of the proposed model and the associated adaptation algorithm was validated through an industrial application. Results show our model can adapt to the variations in operation conditions by using sparsely sampled data, and accomplish higher and more reliable predictive accuracy. Student’s- distribution

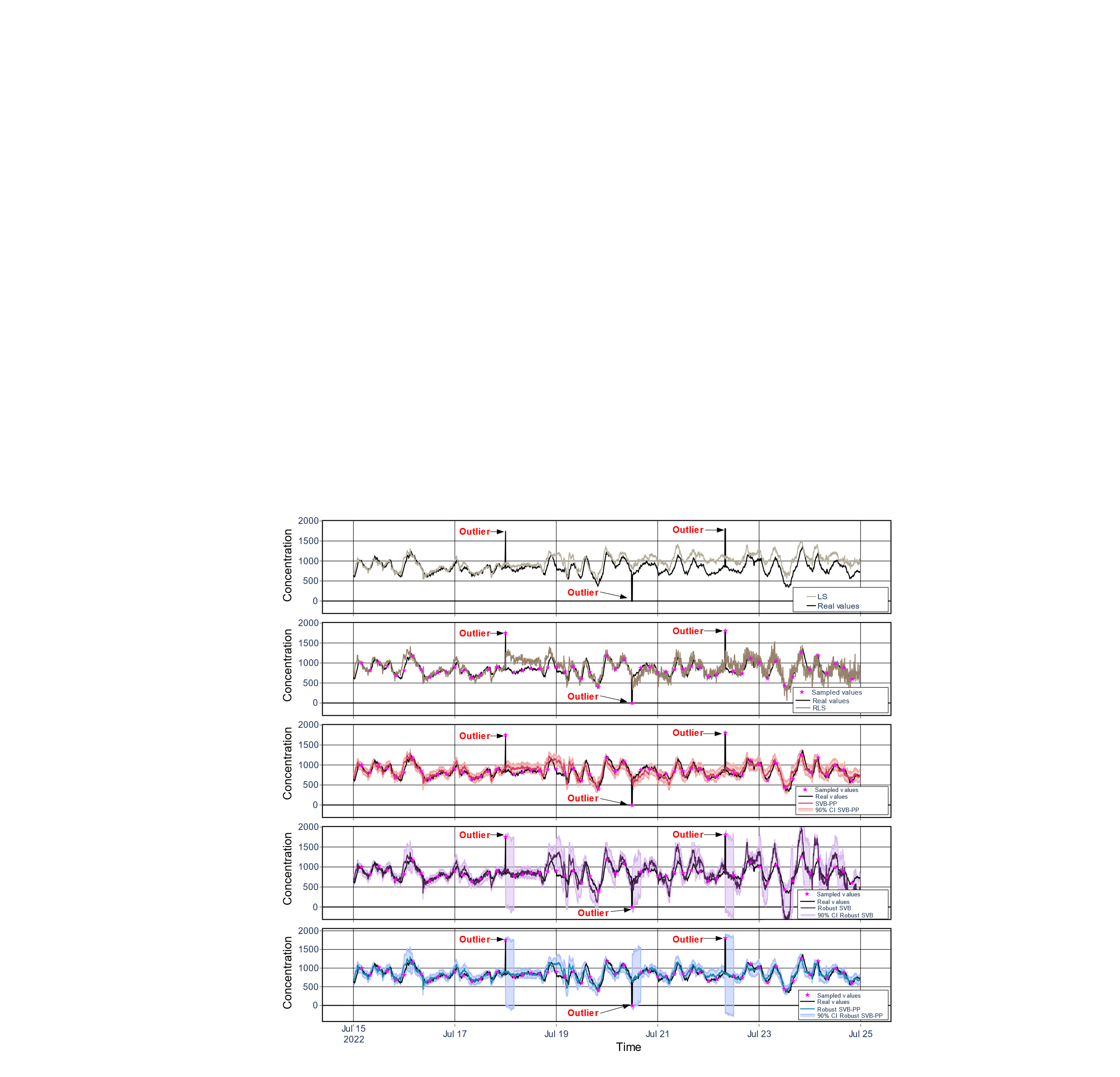


Figure 1. Prediction results given by the five models.

makes it robust to potential outliers in streaming batches. Predictive uncertainties given by the model provide more information on the reliability of the regression model.

Reference

P. Kadlec, B. Gabrys, and S. Strandt, 2009, Data-driven soft sensors in the process industry, *Comput. Chem. Eng.*, vol. 33, no. 4, pp. 795-814.

J. Liu, D.S. Chen, and J.F. Shen, 2010, Development of self-validating soft sensors using fast moving window partial least squares, *Ind. Eng. Chem. Res*, vol. 49, no. 22, pp. 11530-11546.

S.J. Qin, 1998, Recursive PLS algorithms for adaptive data modeling, *Comput. Chem. Eng.*, vol. 22, no. 4-5, pp. 503514.

A. Saptoro, 2014, State of the art in the development of adaptive soft sensors based on just-in-time models, *Procedia Chem.*, vol. 9, pp. 226-234.

C.M. Bishop and N.M. Nasrabadi, 2006, *Pattern Recognition and Machine Learning*, New York.

T. Broderick, N. Boyd, A. Wibisono, A.C. Wilson, and M.I. Jordan, 2013, Streaming variational bayes.

Masegosa A, Nielsen TD, Langseth H, Ramos-López D, Salmerón A, Madsen AL, 2017, Bayesian models of data streams with hierarchical power priors.

M. Svensén and C.M. Bishop, 2005, Robust Bayesian mixture modelling, Neurocomputing, pp. 235-252.

C. Impens, 2003, Stirling’s series made easy. *Am. Math. Mon.*, vol. 110, no. 8, pp. 730-735.