

PCA Based Data Reconciliation in Soft Sensor Development – Application for Melt Flow Index Estimation

Barbara Farsang^{*a}, Imre Balogh^b, Sándor Németh^a, Zoltán Székvölgyi^b,
János Abonyi^a

^aUniversity of Pannonia, Department of Process Engineering, Veszprém, P.O. Box 158, H-8201, Hungary

^bTisza Chemical Group Public Limited Company, Tiszaújváros, P.O. Box 20, H-3581, Hungary

farsangb@fmt.uni-pannon.hu

Melt flow index (MFI) is a very important property of thermoplastic polymers. Laboratory measurements follow standard methods (ASTM D1238 or ISO 1133) to determine MFI and give accurate values, but these measurements are available only in 2-4 hour sampling intervals. Using soft sensors real time estimation of MFI is available for process control and monitoring. When detailed knowledge about the process is not available, data-driven soft sensors can be applied. In this case historical process data are used to build statistical models to determine the relationship between inputs and outputs. Since these methods are based on measurements, the performance of soft sensor depends on the quality of data. Measurements are always affected by errors so pre-processing of data should be necessary. The measurement noise and process variable can be correlated with each other so one opportunity to improve measurement accuracy is using multivariate statistical methods (PCA, PLS). Statistical methods can be improved when phenomenological knowledge is taken into account (e.g. balance equations). The aim of the presented research is to propose a methodology to support the data-driven development of process monitoring systems. We developed a method which improves the effectiveness of data based MFI soft sensor. This method includes advantages of a priori knowledge based models and data-driven multivariate statistical process monitoring tools. As a case study we developed a soft-sensor to estimate MFI of the products of industrial polypropylene reactor at TVK Plc. The proposed method is able to detect undesirable operation states and it can be used for fault detection.

1. Introduction

Advanced process control and monitoring algorithms are based on process variables which are not always measurable or they are measured off-line. Hence, for the effective application of these tools there is a need for estimation algorithms that are based on the model of the process. In the last few years several methods are developed to estimate unmeasured variables. Soft sensor is a mathematical model which can estimate unmeasured, but important variables from other easily measured variables using computational models. The advantage of these inferential measurements compared to the laboratory measurement is that estimation happens in real time so information is continuously available to improve process control and optimization and increase process safety. Kadlec published a very comprehensive review of the applications of soft-sensors in process industry (Kadlec, 2009).

In polymerization industry more important variables are immeasurable in real time, e.g. conversion rate in a reactor. Teran and Machado built a neural network based soft sensor which can estimate the conversion rate online from temperature and pressure of polystyrene reactor, percent of blowing agent, residual monomer and the average molecular weight (Teran, 2011). Melt flow index (MFI) is an important variable to determine the product quality in the industrial ethylene or propylene polymerization process. It gives information about flow of molten thermoplastic polymers; the high MFI means low viscosity and low molecular weight. It is defined as the mass of polymer, in grams, flowing in ten minutes through a capillary of a specific diameter and length by a pressure applied via prescribed alternative gravimetric weights for alternative prescribed temperatures. So

the exact MFI value is determined by an accurate flow measurement, which can be easily measured in the laboratory. Since laboratory measurements are periodically, continuous information is not available for monitoring system. But MFI involves a lot of information about the product quality. So online monitoring of MFI is recommended to operators can get information about changes in process.

In the last few years several algorithms are developed to estimate MFI. Feil and co-workers developed a semi-mechanistic (neural network) model to estimate MFI of polyethylene (Feil, 2004). Sharmin used PLS algorithm to build a soft sensor to predict melt flow index (Sharmin, 2006). Zhang and Liu made a detailed summary about history of neural network based soft sensors and they improved this method. They used an adaptive fuzzy neural network to estimate MFI of polypropylene products (Zhang, 2013). New research area is least squares support vector machine based soft sensors which are applied in chemical industries (Cheng, 2015).

This theme is significant not only in the academic world. The practical significance of the theme is shown in case of a new project in VIKKK Research Center at the University of Pannonia supported by the largest Hungarian polymer production company (TVK Plc.). Our task is to develop a software sensor, which continuously calculates the melt flow index from other online measured process variables. The laboratory measurements can be used to validate the soft sensor and in further check the reliability of model. We developed a hybrid model to estimate melt flow index from other easily measured variables. This technique uses empirical parameters and a priori information based model.

The aim of this article to show how data based models can improve the effectiveness of hybrid models to estimate MFI of polypropylene. In Section 2 we introduce the Principal Component Analysis (PCA) based methods and we show how probabilistic PCA can support the estimation algorithm. In Section 3 we introduce the principle of polypropylene process in case of Spheripol technology and results are summarized. Some conclusion is drawn in Section 4.

2. Role of Principal Component Analysis in soft sensor development

PCA maps the data points into a lower dimensional space, which is useful in the analysis and visualization of the correlated high-dimensional data. Any PCA based method shall be applied only if the examined process is linear. Unfortunately, many processes are not linear thus PCA should not be used. In addition, original PCA method assumed that time dependency does not exist between data points. This drawback can be solved when PCA is dynamized:

$$Y(k+1) = A_1Y(k) + \dots + A_mY(k-t_m) + B_1U(k) + \dots + B_nU(k-t_n) + C \quad (1)$$

where $A_1 \dots A_m$ and $B_1 \dots B_n$ are coefficient matrix, t_m and t_n show time dependence $U(k)$ is the k^{th} sample of input, $Y(k)$ is the k^{th} sample of output.

In case of standard PCA method variables can be formed in the following way: $X=[Y,U]$. Ku suggested that X data matrix should be formed by considering the process dynamics at every sample points (Ku, 1995). Generally speaking, every sample points were completed with the points they can be depended on, i.e. the past n values (n -order model):

$$X = \begin{bmatrix} Y(k) & U(k) & \dots & Y(k-n) & U(k-n) \\ Y(k+1) & U(k+1) & \dots & Y(k+1-n) & U(k+1-n) \\ \vdots & \vdots & & \vdots & \vdots \\ Y(m) & U(m) & \dots & Y(k+m-n) & U(k+m-n) \end{bmatrix} \quad (2)$$

Tipping and Bishop (1999) developed a method called Probabilistic Principal Component Analysis (PPCA). The probability model offers the potential to extend the scope of conventional PCA method. One of the most important advantages of PPCA models is that it allows their combination into mixture of models (Abonyi, 2005).

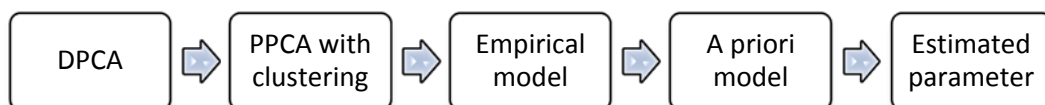


Figure 1: Proposed approach to develop a soft sensor

PPCA is based on an isotropic error model. It seeks to relate a p-dimensional observation vector y to a corresponding k-dimensional vector of latent (or unobserved) variable x . The relationship is

$$y^T = W x^T + \mu + \varepsilon \quad (1)$$

where y is the row vector of observed variable, x is the row vector of latent variables, vector μ permits the model to have a nonzero mean and ε is the isotropic error term. Standard principal component analysis-where the residual variance is zero- is the limiting case of PPCA. PPCA assumes that the values are missing at random through the data set. This means that whether a data value is missing or does not depend on the latent variable, the observed data values are given. There is no closed-form analytical solution for W , so their estimates are determined by iterative maximization of the corresponding loglikelihood using an expectation-maximization (EM) algorithm.

Figure 1 shows our proposed approach to development a soft sensor to predict melt flow index. Firstly credibility test is made using dynamic PCA method. Then the whole operating range is separated to smaller groups called clusters if it is necessary (reason can be the magnitude difference). Then in each cluster projected data are used to identify the coefficients of empirical model to estimate the currently MFI value of produced polymer in case of current operating parameters. Then these values are used in a priori model. In the next chapter we introduce this approach based on a case study.

3. Application in Melt Flow Index estimation

Adequacy of proposed approach is applied in development of a data driven soft sensor which estimates the melt flow index (MFI) real time. In the first subsection technology is described then the results of analysis are introduced.

3.1 Description of the process

Tisza Chemical Group Plc. (TVK) is the largest petrochemical company of Hungary where polymer raw materials (ethylene, propylene, butylenes, etc.) are produced by steam cracking of naphtha or gasoline. In PP-3 plant polypropylene is produced by Spheripol process. This process consists of the following operation steps: catalyst activation, pre-polymerization, polymerization (loop reactor), separation, polymerization (gas phase), steaming, drying. Typical process parameters are summarized in Table 1. As shown in Figure 2, two loop reactors are in TVK plant. The final MFI value is mainly influenced in the two loop reactor so our task was to develop a soft sensor which can estimate online the MFI value from other measured parameters.

In Spheripol process MFI is controlled by hydrogen concentration in reactors, but that it is not the one important variable. So firstly we isolated the more important variables that have got a large effect to the MFI values. We found that hydrogen concentration, ethylene mass flow and triethyl-aluminium (TEAL):donor ratio are the main variables which we have to take into account. Then we separate these measured data into 4 groups because one model cannot be described exactly the whole process due to the magnitude differences; the basic of grouping was the magnitude of MFI value and the type of catalyst. Then we built empirical polynomial model (3) to estimate the produced MFI in reactors in case of all intervals. Since we had got parallel measurements we can validate these models. We had to build a filtering algorithm to the soft sensor due to that empirical models are based on noisy data so the estimation cannot be exact.

Table 1: Typical operating parameters of Spheripol process (Cseryik, 2010)

Process step	Temperature (°C)	Pressure (bar)
Catalyst activation	10	40
Pre-polymerization	20	35
Polymerization (loop reactor)	70	34
High pressure separation	90	18
Polymerization (gas phase)	75-80	10-14
Steaming	105	1.2
Drying	90	1.1

$$MFI_p = f(H_{2,conct}, Et_{mass\ flow}, Cocat, Donor) \quad (3)$$

In order to we can describe product changes exactly in time we have to analyse the dynamic behaviour of system. We take into account the produced MFI in each reactor, the effect of blending, input and output MFI of reactors. The dynamic model can be written as:

$$\frac{dMFI_1^a}{dt} = \frac{Pr_1}{M_{pp1}} \cdot MFI_{p,1}^a - \frac{Pr_1}{M_{pp1}} \cdot MFI_1^a \quad (4)$$

$$\frac{dMFI_2^a}{dt} = \frac{Pr_1}{M_{pp2}} \cdot MFI_1^a + \frac{Pr_2}{M_{pp2}} \cdot MFI_{p,2}^a - \frac{Pr_1 + Pr_2}{M_{pp2}} \cdot MFI_2^a \quad (5)$$

where Pr is the productivity in loop reactors (it is calculated from production rate and conversion), MFI_p is the resulting MFI in reactors (it is calculated based on empirical model), M_{pp} is the mass of polymer in the reactor at any given moment, a is an identified parameter (value: -0.282).

In other words, results of empirical model are used in a *priori* information based model. Since measured, erroneous data are used for identification of parameter of empirical model, the results of soft sensor can be inaccurate and filtering algorithm is needed. So we aimed to improve the effectiveness of soft sensor. The first step is the elimination of the measurement noise from the basic process values measurements. We used different type of PCA method: original PCA method and dynamic PCA. Then we used probabilistic PCA and clustering method to check the adequacy of the grouping. In the next subsection we introduce the results of analysis.

3.2 Results and discussion

Measured process data are available from PP-3 plant. We have got data from 1 y interval: the sampling interval in case of continuous variables is 1 min, while laboratory measurements are in every 2nd – 3rd hours. Numerous polymer grade are produced in this plant. We can analyse steady state and dynamic state of technology. To increase the acceptability of measurements we used PCA and first order dynamic PCA. We have got 22 measured variable, but we find the most of the variance of system can be described with the first 8 principal component. So we projected the 22 dimensional data to an 8 dimensional space. Figure 3 shows the measured and projected PCA values. Since they are industrial process data, we scaled them into a 0-100 interval. Illustrated data basis contains data from steady state and dynamic state.

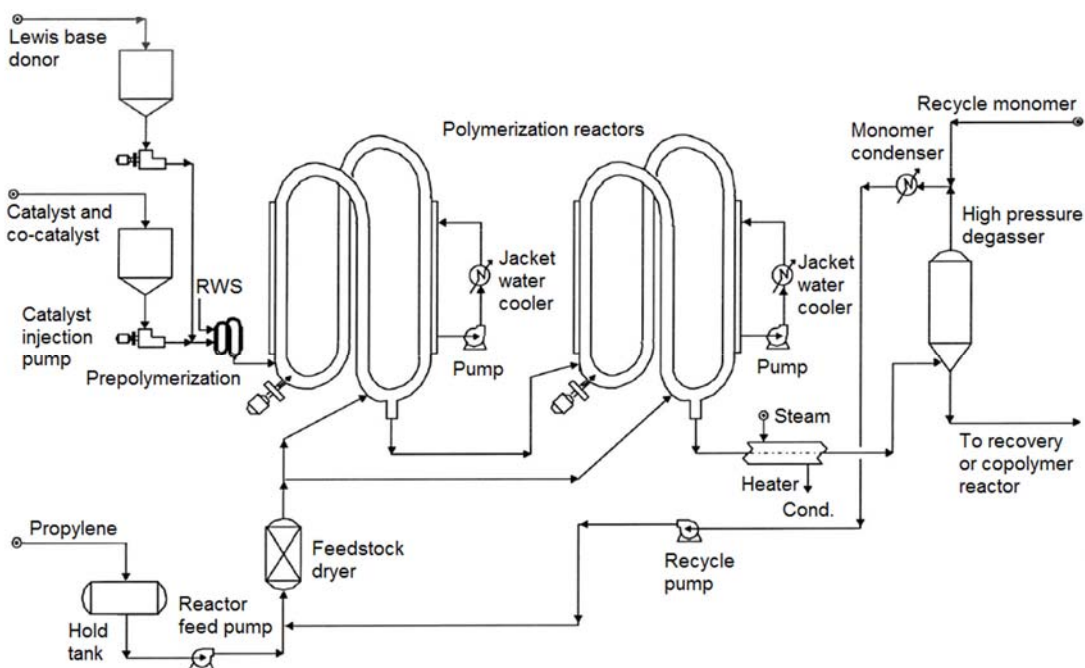


Figure 2: Flowsheet diagram of Spheripol technology with loop reactors to process polypropylene (Cserynik, 2010)

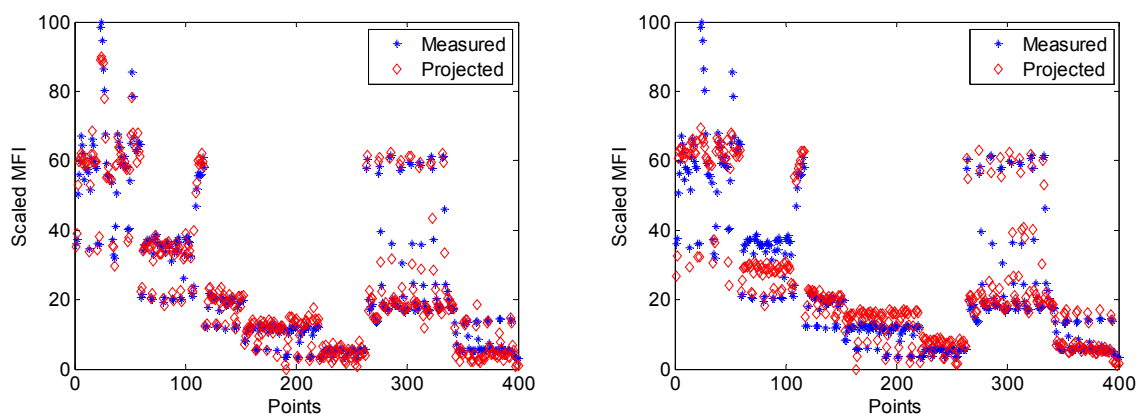


Figure 3: Results of PCA (A) and dynamic PCA (B)

If we use only the original PCA, the accuracy of measured data can be increased, but we do not take into account the time dependency. If we analyse data from dynamic state and apply the dynamic PCA we can see that some of data are outliers. When we built the measured data based soft sensor the estimation of empirical model was difficult in case of large MFI group because these values are higher than others so the estimated data are far from the measured data. If outliers are skipped, the identification of model parameters was more successful and we can increase the efficiency of soft sensor.

In Figure 4 A we can see the organised measured MFI values. We can separate two groups easily: small and large MFI range, because they are obviously differing from other groups. In case of middle MFI range the separation is not unequivocal so we separate the products based on catalyst type. We used probabilistic PCA and clustering algorithm to see which groups are recommended based on main variables (principal components). The result of analysis is visible in Figure 4 B. It shows what the probability is that given point belongs to a group. We can see, that the small and large MFI range form clearly separate groups. In the middle range can be separated into 2 clusters, but in this case the basis of grouping is the MFI range. Then we compare the result of the measured data based and the projected data based soft sensor. We get a more reliable and accurate soft sensor to estimate the MFI value in time, because the mean square error is decreased more than 10 %.

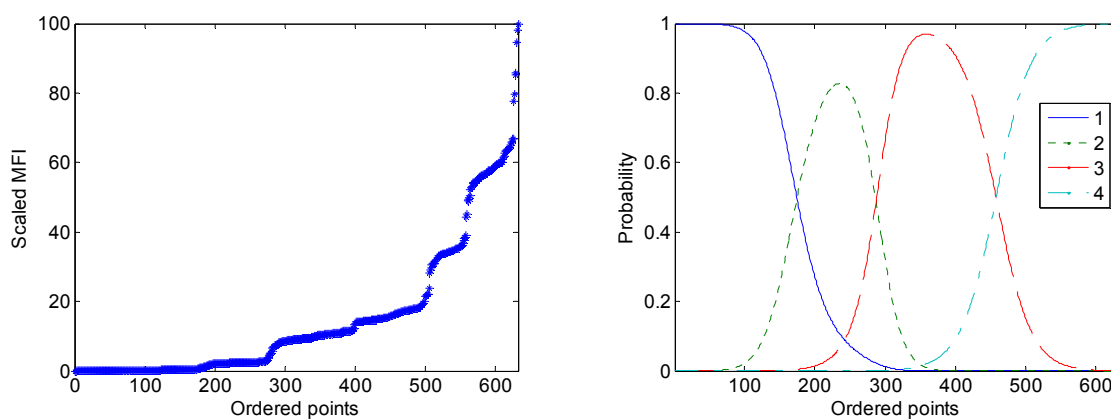


Figure 4: A) Scaled measured MFI values and groups about magnitude of MFI. B) Results of probabilistic principal component analysis and clustering algorithm.

4. Conclusions

In chemical processes some important variables cannot be measured or online measurements are not available. Some properties can be measured in laboratory but data are available periodically. Soft sensors can be used to estimate these values from other easily measured values. Our aim was to develop a soft sensor which can estimate melt flow index of polypropylene product in real time. Firstly we built a measured data based soft sensor. We developed a hybrid model, because empirical and a priori information are used for estimation.

Since measured data are erroneous, the measurement noise can influence the effectiveness of hybrid model. So before the hybrid model data should be pre-processed. We compared the original PCA and dynamic PCA methods. Steady state PCA can increase the accuracy of measurements, but it does not take into account the time-dependency. So dynamic principal components analysis was used and this technique can isolate the outliers or gross errors.

In polymer plant more products are produced which have got different melt flow indexes. Since magnitude difference exists between them, we should grouping data, because one model cannot describe the whole process. Firstly we separated products based on type of catalyst and MFI values. Then we used probabilistic principal component analysis with clustering algorithm to check the rightness of groups. We found that clusters are similar to our groups. Then we re-identified the empirical parameters based on projected data in case of all clusters and we experienced that the projected data based soft sensor results a more reliable MFI values.

Acknowledgements

This research of Barbara Farsang and Janos Abonyi was realized in the frames of TÁMOP 4.2.4. A/2-11-1-2012-0001 „National Excellence Program – Elaborating and operating an inland student and researcher personal support system convergence program” The project was subsidized by the European Union and co-financed by the European Social Fund. This research was supported by the European Union and financed by the European Social Fund in the frame of the TÁMOP 4.2.2.A/11/1-KONV-2012-0071 project.

References

- Abonyi J., Feil B., Nemeth S., Arva P., 2005, Modified Gath–Geva clustering for fuzzy segmentation of multivariate time-series, *Fuzzy Sets and Systems*, 149, 39–56 DOI: 10.1016/j.fss.2004.07.008.
- Cheng Z., Liu X., 2015, Optimal online soft sensor for product quality monitoring in propylene polymerization process, *Neurocomputing*, 149, 1216–1224 DOI: 10.1016/j.neucom.2014.09.006.
- Csernyik I. PP technology, <www.tvk.hu/repository/615517.pdf>, accessed 05.12.2014
- Feil B., Abonyi J., Pach P., Nemeth S., Arva P., Nemeth M., Nagy G., 2004, Semi-mechanistic Models for State-Estimation – Soft Sensor for Polymer Melt Index Prediction, *Artificial Intelligence and Soft Computing*, 3070, 1111-1117.
- Kadlec P., Gabrys B., Strandt S., 2009, Data-driven soft sensors in the process industry. *Computers and Chemical Engineering*, 33(4), 795-814. DOI: 10.1016/j.compchemeng.2008.12.012.
- Ku W., Storer R.H., Georgakis C., 1995. Disturbance Detection and Isolation by Dynamic Principal Components Analysis. *Chemometrics and Intelligent Laboratory Systems*. 30, 179-196. DOI: 10.1016/0169-7439(95)00076-3.
- Sharmin R., Sundararaj U., Shah S., Griend L.V., Sun Y.J., 2006, Inferential sensors for estimation of polymer quality parameters: Industrial application of a PLS-based soft sensor for a LDPE plant, *Chemical Engineering Science*, 61(19), 6372-6384, DOI: 10.1016/j.ces.2006.05.046.
- Teran R.A.C., Machado R.A.F., 2011, Soft-sensor based on artificial neuronal network for the prediction of physico-chemical variables in suspension polymerization reactions, *Chemical Engineering Transactions*, 24, 529-534, DOI: 10.3303/CET1124089.
- Tipping M.E., Bishop C.M., 1999, Mixtures of probabilistic principal component analyzers, *Neural Comput.*, 11(2), 443–482.
- Zhang M., Liu X., 2013, A soft sensor based on adaptive fuzzy neural network and support vector regression for industrial melt index prediction, *Chemometrics and Intelligent Laboratory Systems*, 126, 83–90, DOI: 10.1016/j.chemolab.2013.04.018.