Dealing with Uncertainty in Polymer Manufacturing by Using Linear Regression Metrics and Sensitivity Analysis

Aarón David Bojarski[a], Carlos Rodrigo Álvarez[a], Luis Puigjaner[a]

[a]*Universitat Politecnica de Catalunya, Chemical Engineering, ETSEIB, Pavello G-2, Av. Diagonal 647, E-08028 Barcelona, Spain; luis.puigjaner@upc.edu*

Abstract

In the polymers industry, models are heavily parametrizised, and the effect of each parameter on model outputs has not been extensively studied. A wide range of values for most of model's parameters can be found in literature [1], a thorough analysis regarding the model's sensitivity to the parameters' values is needed to find the set of parameters that have the most impact in the output results and consequently deserve an extra effort and care during their estimation. In this work, a global sensibility analysis of a styrene emulsion polymerization reactor model is carried out in order to determine the set of critical parameters.

Keywords: polymer production modelling, sensitivity analysis, regression analysis, uncertainty, Monte Carlo sampling

1. Introduction

In the case of the chemical industry uncertainty sources are found due to the model which includes all kinds of information arising from experimental and pilot plant data (kinetic constants, physical properties and transfer coefficients) and the process including information regarding stream quality, (variations in flow rate, pressure and temperature). Other uncertainty sources are external uncertainty including variability regarding information that is outside the system boundary and finally discrete uncertainty which is related to equipment availability and other random events [2].

During the last decades, polymerization processes have played a central part in petrochemical industry for the production of plastics, rubbers, paints and other many kinds of products. The many possible phases present inside some types of polymerization reactors and transfer phenomena occurring between them make emulsion polymerization systems very difficult to model. In addition to the common difficulties of modelling mass transfer phenomena and equilibrium conditions between several phases, the lack of reliable values for parameters makes the modelling task even more complex.

The big disparity between parameters values used in those models, either, related to reaction kinetics or used within transfer and equilibrium equations, causes serious problems when one attempts to validate the model. Some of these parameters like pre-

exponential constants or activation energies for the different polymerization reactions involved (propagation, termination, chain transfer etc) may differ in more than one order of magnitude in similar conditions [1]. This big disparity, can also be found in other parameters like diffusion coefficients or phase surface tension values [3,4].

Due to this big parameter uncertainty ranges, it is important to estimate the actual effect of each one of these uncertain parameters on the final model outcomes, aiming at determining which are the most influential ones. The knowledge of this set of critical parameters can focus attention on them for more accurately estimation via physical experiments or for further model enhancements. Moreover, it also allows for the modeller to know which parameters are the least influential to model outcomes, to ease the modelling of such model parts, by assuming model simplifications.

A sensibility analysis (SA) [5], is the study of how the variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation and of how this model depends on information fed into it. Various methods have been proposed to make uncertainty operational due to parameter uncertainty, such as the use of analytical uncertainty propagation methods; calculations based on intervals; applied fuzzy logic computations; and stochastic modelling describing parameters as uncertainty distributions [6].

The usage of analytical propagation methods suffers from complexity in algebra that increases rapidly with the complexity of the model, the method produces moments of distributions (mean and variance) making hard to obtain reliable estimates for the tails of the output distribution. It is basically a local approach and will not be accurate if the uncertainties are large, if the model is not smooth or if important covariance terms are omitted. In this respect Saltelli [5] points out that the usage of error propagation methods (derivative methods) provide only of a local glimpse at model factors effect on model outputs. The usage of sampling methods which explore the full space of possible model parameters values is recommended.

2. Problem statement

Several metrics can be calculated based on sampling results, such as standard statistics (mean, standard deviations and confidence intervals) or regression analysis metrics. Regression metrics are based on a linear correlation resulting from input variables (x) and model output results (y). Standardization of input variables and output results is performed by subtracting the mean value and normalizing the standard deviation. SRCs are obtained from a fit to a minimum square difference optimization. A value of any SRC close to zero indicates that the output variable is not correlated to that input variable. The sign of SRC also indicates the relationship between input and output variables.

$$\frac{y_l - \overline{y_l}}{\sigma_l} = \sum_{l=1}^{M} SRC_{lh} \frac{x_h - \overline{x_h}}{\sigma_h} \qquad \forall l = 1:N \tag{1}$$

Dealing with Uncertainty in Polymer Manufacturing by Using Linear Regression Metrics and Sensitivity Analysis

Another commonly used metrics are Partial Correlation Coefficients (PCCs); which are calculated by performing several regressions including or not the variable under study. In this case a PCC tries to show how much each input variable affects the behaviour of the output variables. This can be obtained by performing two separate regressions one where all input variables are used (y_{lf}^{xh}) and one with the subject input variable ignored ($y_{lf}^{\sim xh}$) [5].

$$PCC_{hl}^2 = \frac{\sum_{f=1}^{M}\left(y_{lf} - y_{lf}^{\sim x_h}\right)^2 - \sum_{f=1}^{M}\left(y_{lf} - y_{lf}^{x_h}\right)^2}{\sum_{f=1}^{M}\left(y_{lf} - y_{lf}^{\sim x_h}\right)^2} \qquad \forall l = 1:N; h = 1:M \qquad (2)$$

PCCs have only positive values; the values that are closer to one represent more important variables.

3. Paper approach

In this work we adopt a stochastic sampling approach, which varies input data (model and process parameters) according to given probability distributions. The model is run for a given set of input values realizations and stores its output results. This procedure is repeated until the appropriate uncertainty ranges are obtained for the output variables.

3.1. Methodology and Case Study

An isothermal batch reactor model for styrene emulsion polymerization is considered. The model includes the mass balances for initiator, surfactant, monomer, radical and polymer species. Both, micellar and homogeneous nucleation are considered and the radical flow into particles and micelles and also its desorption from particles are included following a similar approach as in Gao et al. [3]. The model also includes polymer molecular weight calculations by using the moments approach. The model renders a detailed description of the physical phenomena and chemical reactions that take place in the aqueous and polymer phases. Model results have been able to reproduce experimental data provided in the literature [1,3,4].

3.2. Sampling methodology

A Monte Carlo sampling (MCS) methodology was used. Variable input probability distribution functions (pdfs) can be seen in table 1. In order to set these pdfs data from literature was used. Uniform distribution functions were used for parameters for which no information except a range was found, while normal distributions were used in the case of parameter for which more information was available (process operating conditions).

Table 1. Input parameter tested distributions

Variable Name	Unit	Distribu-tion shape	Distribution parameters		Variable Remark
			Mean / Lower Bound	STD / Upper Bound	
A_{kt}	[l/mol·min]	Uniform	3.33E+15	4.50E+15	A
E_{kt}	[kJ/mol·K]	Uniform	7.54E+03	1.02E+04	A
A_{kt}^{w}	[l/mol·min]	Uniform	3.33E+15	4.50E+15	A
E_{kt}^{w}	[kJ/mol·K]	Uniform	7.54E+03	1.02E+04	A
A_{kp}	[l/mol·min]	Uniform	9.64E+10	1.30E+11	A
E_{kp}	[kJ/mol·K]	Uniform	8.84E+03	1.20E+04	A
A_{kp}^{w}	[l/mol·min]	Uniform	9.64E+10	1.30E+11	A
E_{kp}^{w}	[kJ/mol·K]	Uniform	8.84E+03	1.20E+04	A
alfaTerDes	[1]	Uniform	0.00E+00	1.00E+00	A
A_{kf}^{m}	[l/mol·min]	Uniform	4.69E+03	6.34E+03	B
E_{kf}^{m}	[kJ/mol·K]	Uniform	2.86E+04	3.86E+04	B
Chi	[1]	Uniform	4.00E-01	7.00E-01	C
Dw	[dm^2/min]	Uniform	1.20E-07	1.76E-09	D
Dp	[dm^2/min]	Uniform	1.20E-08	1.76E-12	D
mmd	[1]	Uniform	8.50E+11	1.15E+12	D
Jcr	[1]	Uniform	5.00E+00	8.00E+00	D
DMM	[dm2/min]	Uniform	8.72E-12	1.18E-11	D
F	[1]	Uniform	5.00E-01	7.00E-01	E
A_{kd}	[l/mol·min]	Uniform	1.30E+18	1.75E+18	E
E_{kd}	[kJ/mol·K]	Uniform	2.83E+04	3.83E+04	E
r_{mic}	[dm]	Uniform	2.00E-08	5.00E-08	F
Surfactant	[g]	Normal	1.43E+01	2.00E+00	G
Monomer	[g]	Normal	3.00E+02	1.67E+01	G
Initiator	[g]	Normal	7.07E+00	1.00E+00	G
Reactor Temp.	[K]	Normal	3.23E+02	6.67E+00	H
Reactor Vol.	[l]	Normal	7.00E-01	3.33E-02	H

Variables marked as A, represent values associated to the propagation-termination reaction system, B related to inhibition and chain transfer mechanisms, C, D, E and F are related to monomer, radical, initiator and surfactant physical properties respectively. G and H refer to initial charge and reactor working conditions. In the case of kinetic related parameters the superscript refers to the phase where reaction occurs.

Each scenario was created by sampling all variables from their respective distributions, these variables realizations were used to run the model. Output variables were calculated at three different time intervals namely: 0-10, 45-55 and 90-100 minutes. These three time regions were defined due to expected different model behaviour. Random variable value generation and SRC and PCC calculation were performed using Matlab®. The number of scenarios used to compute SRCs and PCCs was gradually

increased until the obtained values did not change appreciably. The total number of simulation runs performed was 5000.

3.3. Results and discussion

The model output results studied were monomer conversion (X), polymer molecular weight (MW) and polydispersion (PD). These variables allow for a global interpretation of model results.

Table 2. Calculated time average SRCs and PCCs

Variable	X		MW		PD	
Metric	SRC	PCC	SRC	PCC	SRC	PCC
A_{kt}	0,003	0,005	-0,012	0,016	0,024	0,026
E_{kt}	-0,011	0,048	0,064	0,082	-0,110	0,115
A_{kt}^{w}	-0,002	0,004	-0,001	0,001	-0,010	0,010
E_{kt}^{w}	-0,015	0,065	0,024	0,031	0,006	0,007
A_{kp}	0,020	0,042	0,020	0,026	-0,005	0,008
E_{kp}	**-0,306**	**0,521**	**-0,215**	**0,268**	0,014	0,017
A_{kp}^{w}	-0,009	0,020	0,005	0,007	0,008	0,008
E_{kp}^{w}	0,000	0,005	0,000	0,003	-0,006	0,006
alfaTerDes	-0,005	0,029	0,021	0,028	0,085	0,089
A_{kf}^{m}	-0,011	0,022	-0,003	0,004	0,002	0,002
E_{kf}^{m}	**0,147**	**0,295**	**0,114**	**0,146**	0,056	0,058
Chi	-0,006	0,017	-0,003	0,004	-0,009	0,010
Dw	0,003	0,011	-0,008	0,010	0,004	0,005
Dp	-0,076	0,159	0,073	0,095	0,011	0,012
mmd	0,009	0,022	0,001	0,001	-0,010	0,011
Jcr	0,000	0,003	0,013	0,017	-0,005	0,005
DMM	-0,002	0,022	-0,004	0,005	-0,021	0,022
F	0,018	0,038	0,003	0,004	0,010	0,011
A_{kd}	0,019	0,039	-0,018	0,023	-0,004	0,004
E_{kd}	**-0,757**	**0,839**	**0,629**	**0,625**	**0,295**	**0,292**
r_{mic}	0,037	0,079	-0,041	0,053	-0,066	0,069
Surfactant	-0,015	0,039	0,025	0,033	-0,010	0,014
Monomer	-0,009	0,029	0,008	0,010	0,017	0,018
Initiator	0,029	0,061	-0,012	0,016	0,001	0,004
Reactor Temp.	**0,217**	**0,414**	**-0,111**	**0,142**	-0,041	0,043
Reactor Vol.	-0,005	0,011	-0,002	0,003	0,017	0,018

The values were calculated for each variable at three different time intervals, the reported value is the arithmetic average (see table 2). It is found that each input parameter affects in different ways to the selected output variables results. In all cases a high value for PCC is also associated to a value significantly different than zero for SRC. The most influencing input variables found were: reactor temperature (*T*) and the

activation energies for the initiator decomposition (E_{kd}), inhibition (E_{kfm}) and polymer phase polymerization (E_{kp}) reactions.

It can be seen from SRCs values that increments in reactor temperature for this system will impact increasing conversion while decreasing polymer molecular weight. The other parameters found to be influential are the activation energies for the polymer phase, while all other reactions are mostly non influential. The only input parameter influencing polydispersion was found to be the activation energy for the initiator decomposition reaction.

Table 3. Calculated SRCs for different time intervals for the four most significant variables

Variable	X			MW			PD		
Time Interval	[0-10]	[45-55]	[90-100]	[0-10]	[45-55]	[90-100]	[0-10]	[45-55]	[90-100]
E_{kp}	*-0.456*	**-0.248**	*-0.214*	**-0.231**	**-0.206**	**-0.207**	0.040	0.005	-0.003
$E_{kf}{}^{m}$	**0.143**	**0.152**	**0.144**	**0.110**	**0.117**	**0.115**	0.052	0.058	0.056
E_{kd}	*-0.666*	**-0.795**	*-0.808*	**0.640**	**0.625**	**0.621**	**0.309**	**0.298**	**0.279**
Reactor Temp.	**0.210**	**0.222**	**0.220**	**-0.105**	**-0.115**	**-0.113**	-0.055	-0.036	-0.031

From table 3, it can be seen that some input variables that show an appreciable change over time in its influence over and output variable is E_{kp} and E_{kd} over X. At the reaction start their SRCs values are different than at reaction end. In the first case E_{kp} influence is higher at reaction start, while lower at the end, the opposite behaviour is found for E_{kd}. For the remaining variables their influences over output variables remain similar over the whole time interval, no changes in the SRCs values are found. A similar trend is found when analysing the PCCs results.

4. Conclusions

It has been shown that the usage of SRCs and PCCs eases the selection procedure for variables that influence the most on model outputs. It also enables the study of how the input parameter behaviour affects model output. The current procedure enables to focus attention only on four parameters for further studying, instead of the original set of 26. The most influencing parameters found were related to kinetic reaction constants.

5. Acknowledgements

6. References

[1] J. Brandrup, E. H. Immergut, E. A. Grulke, Eric A. Grulke, D. Bloch (eds); Polymer Handbook, Wiley (2003).
[2] Pistikopoulos, E., Computers and Chemical Engineering No. 19 S553 (1995).
[3] Gao, J. and A. Penlidis, Prog. Polym. Sci., No. 27 403 (2002).
[4] M. Rajabi-Hamane, S. Engell; Chemical Engineering Science No. 62 5282 (2007).
[5] Saltelli, A., Chan, K., Scott, E. M. Sensitivity Analysis, Ch 6, Wiley , (2002).
[6] Huijbregts, M., International Journal of Life Cycle Assessment, No 3, 5, 273 (1998).