Modelling the Distribution of Nitrophenols in a Liquid-Liquid System Representative of an Industrial Nitration Process

Ana Leonor C. V. Lopes[a,b,c], Alejandro F. G. Ribeiro[a], Marco S. Reis[b], Dulce C. M. Silva[a], Inês Portugal[c], Cristina M. S. G. Baptista*[b]

[a]CUF - Químicos Industriais S.A., Quinta da Indústria, R. Amoníaco Português nº10, Beduído, 3860-680 Estarreja, Portugal
[b]CIEPQPF - Department of Chemical Engineering, Faculty of Sciences and Technology, University of Coimbra, R. Sílvio Lima, Polo II, 3030-790 Coimbra, Portugal
[c]Department of Chemistry and CICECO, University of Aveiro, Campus Universitário de Santiago, 3810 - 193 Aveiro, Portugal
cristina@eq.uc.pt

The industrial production of mononitrobenzene (MNB) is currently achieved with high yields by liquid phase nitration of benzene (Bz). However, the environmental footprint of this process is highly dependent on the minimization of reaction by-products, such as nitrophenols, whose elimination and disposal procedures are subjected to very strict regulations. The mechanisms behind their formation are still not completely understood and would benefit from detailed information about their distribution in the reaction mixture. This work reports predictive models for the distribution ratio (D) of the by-products 2,4-dinitrophenol (DNP) and 2,4,6-trinitrophenol (TNP) in a bi-phasic organic-aqueous system representative of industrial nitrators.

Equilibrium concentrations of DNP and TNP were measured at 30 ºC for different initial conditions, namely: composition of the aqueous (H_2SO_4 + water) and organic (Bz + MNB) phases, their weight ratio, and the concentration of nitrophenols. Multivariate linear regression (MLR) models were built with a sub-set of the experimental data and confirmed by external validation, exhibiting high fit quality (R^2 and $R^2_{\text{Adj}} > 0.900$) and prediction capability ($R^2_{\text{Pred,DNP}} = 0.936$; $R^2_{\text{Pred,TNP}} = 0.962$). The most significant predictor was the equilibrium concentration of sulphuric acid in the aqueous phase, followed by benzene concentration in the organic phase. The distribution ratios of DNP and TNP decrease with the increase of these concentrations. A higher affinity of DNP with the organic phase was confirmed by $D_{\text{DNP}} \approx 2D_{\text{TNP}}$.

1. Introduction

Mononitrobenzene (MNB) is produced industrially by nitration of benzene with a mixture of sulphuric and nitric acids (mixed acid). The formation of undesired by-products, such as 2,4-dinitrophenol and 2,4,6-trinitrophenol, reduces the selectivity, yield and economy of the process, and it compromises final product specifications. The toxicity of nitrophenols (NP) affects the environmental footprint of the process. Moreover, when MNB is used for aniline production the presence of NP reduces the catalyst lifetime. Consequently, industrial processes include expensive purification stages and elimination procedures complying with strict environmental regulations.

An enormous effort has been made to understand and mitigate NP formation being of note the works of Hanson et al. (1976), Burns and Ramshaw (1999) and Quadros et al. (2004). Nonetheless, detailed mechanims of formation and mass transfer have not been established yet leading to the use of statistical models to account for their presence when modelling benzene nitration (Nogueira et al., 2013). Moreover, some important properties of NP are also lacking, such as their equilibrium distribution in the two liquid immiscible phases characteristic of the nitration system. The correlation of Zaldivar et al. (1995), that describes the equilibrium distribution of benzene as a function of temperature and aqueous phase composition, has been considered appropriate for NP (Dummann, 2009) but its validity has not been confirmed.

Equilibrium distribution of a solute in dilute liquid-liquid (L-L) systems is described by experimental parameters such as distribution constant, K_D in Eq(1), and distribution ratio, D in Eq(2) (Berthod and Carda-Broch, 2004). The former is specific for the cases of solute partition in one chemical form (e.g. molecular, ionic, elemental) being defined, in Eq(1), as the ratio of molar concentrations of the solute i in the organic and in the aqueous phases (C_i^O and C_i^A, respectively).

$$K_D = \frac{C_i^O}{C_i^A} \tag{1}$$

On the other hand, D is specific for solutes which may be present in different chemical forms being defined as the ratio of total molar concentration of all forms of solute i between the two phases. For instance, NP are ionisable compounds ($\text{AH} \rightleftharpoons \text{A}^- + \text{H}^+$) that can be present in molecular or ionic forms and, consequently, D is defined by Eq(2).

$$D = \frac{C_{\text{AH}}^O + C_{\text{A}^-}^O}{C_{\text{AH}}^A + C_{\text{A}^-}^A} \tag{2}$$

The curve representing the influence of pH on D exhibits an inverse sigmoidal profile, with the molecular form (AH) being more soluble in the organic phase, at low pH, and the ionic form (A^-) being more soluble in the aqueous phase, at high pH. Nevertheless, at low pH, when ionization is minimized, $D \cong K_D$ (Berthod and Carda-Broch, 2004). Azevedo (2015) reported that ionization of DNP and TNP in a MNB-water system was minimized at pH < 0.5. In the present work the equilibrium distribution of NP in a bi-phasic organic-aqueous system representative of industrial nitrators (but in the absence of reaction, i.e. without nitric acid) will be described by the distribution ratio, D_j, where j represents DNP or TNP.

2. Statistical-oriented approach

Simple and reliable prediction models can be developed by a statistical-oriented approach integrating three main steps: 1) identification of the main variables (screening); 2) model development based on a training data set and fit analysis; and, 3) model validation with a new testing data set. Both training and testing data sets can be obtained using a statistical design of experiments (DOE) approach, namely by a factorial design to assess the individual and interaction effects of the predictor variables (factors) with a minimum number of experiments. If the predictor variables, x, are uncorrelated the method of least squares for multivariate linear regression (MLR) can be implemented over the training data set (Montgomery et al., 2012) in order to build models of the response variable, y, by Eq(3):

$$y = \beta_0 + \sum_{g=1}^{z} \beta_g x_g + \sum_{g=1}^{z-1} \sum_{h=1+g}^{z} \beta_{gh} x_g x_h + \varepsilon \tag{3}$$

where β_0 designates the intercept, β_g are the partial regression coefficients for the z input variables terms, β_{gh} represent the coefficients of the 2nd order interaction terms and ε is the error of the model. According to the MLR assumptions the error is identically distributed with a normal distribution of zero mean and constant variance σ^2 (homoscedasticity).

Analysis of fit encompasses an evaluation of the significance of the model, the MLR residual assumptions, and the fit quality of the prediction models developed ($\hat{y}$). Significance can be assessed by hypothesis testing (Montgomery et al., 2012) applied to the whole model (ANOVA for regression) or to the individual terms. Statistical significance of the whole model is easily assessed by comparing the ANOVA p-value with the adopted significance level α ($\alpha = 0.05$): if the p-value $< \alpha$, the model is deemed significant. The same approach is followed for the individual hypothesis tests performed on the variables' β coefficients. The MLR residual assumptions are validated by analysis of residuals ($e_c = y_c - \hat{y}_c$, $c = 1,2,\ldots,N$) of the N observations, which are expected to be independent and identically distributed with zero mean and constant variance. Homoscedasticity can be checked using a plot of the residuals, e_c, versus the estimated values, $\hat{y}_c$ (Montgomery et al., 2012). Normality can be evaluated by a QQ-plot or a Shapiro-Wilk test applied to e_c (SAS Institute, 2016). Detection and influence of outliers in the model is analysed by the studentized residuals (defined as $r_c = e_c/\text{SE}(e_c)$ where $\text{SE}(e_c)$ is the standard error of the residuals) and Cook's distance, respectively (Montgomery et al., 2012). A large Cook's distance meaning a significant specific weight of that observation in the model. Fit quality is characterized by the coefficient of determination, R^2, and by the adjusted R^2, R_{Adj}^2, with ideal values being close to the maximum value of 1. These coefficients measure how much of the variance of the response is explained by the regressors in the MLR model, with R_{Adj}^2 being

adjusted to penalize overfitting (because it accounts for the number of estimated coefficients, β). By definition R^2_{Adj} is always smaller than R^2, but for a good fit it is desirable that $R^2_{\text{Adj}} \cong R^2$. For model comparison it is important to evaluate the root mean squared error (RMSE), defined by Eq(4), which is an estimate of the standard error of the model ($\sqrt{\hat{\sigma}^2}$) (Montgomery et al., 2012).

$$\text{RMSE} = \sqrt{\frac{\sum_{c=1}^N (y_c - \hat{y}_c)}{N - p}} \tag{4}$$

An external validation methodology is preferred to evaluate the prediction capability of the fitted MLR models (Esbensen and Geladi, 2010). Ths requires a new test data set, to which the model is applied to calculate the coefficient of determination of prediction, R^2_{Pred}, as well as the root mean squared error of prediction ($\text{RMSE}_{\text{Pred}}$), defined by Eq(5),

$$\text{RMSE}_{\text{Pred}} = \sqrt{\frac{\sum_{t=1}^{N_t} (y_t - \hat{y}_t)}{N_t}} \tag{5}$$

where N_t is the number of representative and independent experiments used to assess the model.

In this work, the development and analysis of MLR models for prediction of D_j in L-L system was carried out with the software JMP® PRO version 12.1.0 (SAS Institute Inc.).

3. Material and methods

In the first stage of this study, experimental data were collected to model the distribution of DNP and TNP between organic and aqueous phases mimicking the reaction mixture of the nitration process. Benzene (99.94 wt %, Gadiv Petrochemical Industries Ltd) and mononitrobenzene (99.995 wt %, CUF – Químicos Industriais S.A.) were mixed to prepare organic phases with 10, 30 and 50 wt % Bz. Precise amounts (± 0.0001 g) of 2,4-dinitrophenol (97 wt %, Aldrich) and 2,4,6-trinitrophenol (99 wt %, BDH) were added to these organic mixtures and dissolved using ultrasound to ensure homogeneity. Sulphuric acid (95-97 wt %, Merck) was diluted with high purity water (Barnstead E-pure purification system) to obtain solutions in the range 50-75 wt %. Sulphuric acid content was confirmed by titration following the procedure described in Quadros et al. (2004).

Different weight ratios of aqueous/organic phases, in the range 2-8, were mixed according to the shake flask method (Berthod and Carda-Broch, 2004), adapted for the current study with shorter agitation ($\approx$ 2 min) and decantation times ($\approx$ 60 min) as suggested by Azevedo (2015). The tests were carried out at 30 °C (± 0.1 °C) using a F25-ED thermostatic bath from Julabo. Phase separation was achieved by decantation followed by centrifugation (1 min at 1250 rpm) in a Rotanta 460 centrifuge from Hettich. Samples of equilibrium aqueous phases were prepared by collection of aqueous phase aliquots (ca. 0.5 g) and dilution with water (10 times). These samples were analysed by HPLC to measure the equilibrium weight fractions of DNP and TNP following the procedure described in Costa et al. (2013). The measurements were carried out in triplicate using an Elite LaChrom HPLC from VWR Hitachi equipped with a LiChroCART® 125-4 column (5 µm, 125 × 4 mm) and a LiChroCART® 4-4 guard column (5 µm, 4 × 4 mm), both from Merck. The HPLC was calibrated for a NP weight fraction range of 1-5×10⁻⁶ in acidic medium (diluted samples with 1 wt % sulphuric acid) with $R^2_{\text{DNP}} = 0.9998$ and $R^2_{\text{TNP}} = 0.9991$. The HPLC system was washed several times between runs. The software EZChrom Elite version 3.1.7 from Agilent was used to calculate peak areas. A repeatability and reproducibility study was performed confirming the capability and adequacy of the measurement system for this purpose. The weight fractions of NP in the organic phase were calculated by mass balance assuming an immiscible L-L system.

4. Results and discussion

A statistical-oriented approach was applied to develop MLR models for the prediction of D_j ($\hat{D}_j$) in L-L system at equilibrium. The main variables were identified in a previous work (unpublished results) being the equilibrium composition of the phases, expressed in terms of molar concentration of sulphuric acid in the aqueous phase ($C^A_{\text{SA},e}$), and of benzene ($C^O_{\text{Bz},e}$) and nitrophenol j ($C^O_{j,e}$) both in the organic phase. For each nitrophenol model, equilibrium temperature and the presence of the other nitrophenol were found to be irrelevant.

The models for $\hat{D}_j$ were developed using the training data set and validated on new independent testing data set. The equilibrium conditions of the data sets were defined by DOE, applied to the following experimental domain: $6.946 < C^A_{\text{SA},e} < 12.42$; $1.471 < C^O_{\text{Bz},e} < 6.442$; $0.414 < C^O_{\text{DNP},e} \times 10^2 < 1.24$; $2.85 < C^O_{\text{TNP},e} \times 10^3 < 9.87$;

and molar ratio $4.0412 < n_e^A/n_e^O < 30.262$. The training set consisted of 34 duplicate experiments given by 17 conditions of a two-level full factorial design with $C_{SA,e}^A$, $C_{Bz,e}^O$, and $C_{j,e}^O$ as factors and one central point. The testing set was composed of 9 experiments resulting from a two-level fractional factorial for the same factors and one central point.

The $\widehat{D}_j$ models developed for DNP and TNP are in Eq(6) and Eq(7), respectively.

$$\widehat{D}_{DNP} = 175 - 11.7C_{SA,e}^A - 3.45C_{Bz,e}^O + 0.852\left(C_{SA,e}^A - 9.682\right)\left(C_{Bz,e}^O - 3.968\right) \tag{6}$$

$$\widehat{D}_{TNP} = 73.7 - 4.60C_{SA,e}^A - 2.08C_{Bz,e}^O + 701C_{TNP,e}^O + 0.369\left(C_{SA,e}^A - 9.682\right)\left(C_{Bz,e}^O - 3.968\right) \tag{7}$$

The models and their respective regressor terms are statistically significant as indicated by the low p-value ($< \alpha = 0.05$) of the corresponding test statistics (Table 1). From these results one can conclude that composition of both phases has a great influence on D_j, with higher concentrations of H_2SO_4 in the aqueous phase and of Bz in the organic phase leading to lower values of D_j. Interactions terms between the two variables were also identified for both models. The concentration of TNP may have a slight influence on D_{TNP} (p-value $= 0.0554 \cong \alpha$) and yet the concentration of DNP does not influence D_{DNP} (p-value > 0.1). The analysis of the MLR coefficients (Table 1) according to the magnitude of test statistic t, showed that the order of significance of each regressor variable was: $C_{SA,e}^A > C_{Bz,e}^O$ ($> C_{TNP,e}^O$ in the case of TNP). The triangular sensitivity indicators in Figure 1 represent the partial derivative of the prediction trace function of the independent variables at those conditions. Their height helps to visualize the higher influence of $C_{SA,e}^A$ on D_j, followed by $C_{Bz,e}^O$, and an increase in these concentrations leads to lower D_j values. Furthermore, Figure 1 also shows that the equilibrium concentration of TNP in the organic phase can have a minor statistical importance in the explanation of D_{TNP}.

Table 1: Parameters, fit and prediction capabilities analysis results of the models in Eq(6) and in Eq(7).

| Model | Coefficients | | | | | RMSE | ANOVA | Fit Quality | | Prediction capability | |
	$\hat{\beta}$	$\hat{\sigma}_\beta$	Test statistic (t)	p-value	95 % CI		p-value	R^2	R_{Adj}^2	R_{Pred}^2	$RMSE_{Pred}$
$\widehat{D}_{DNP}$	175	6.71	26.2	< 0.0001	[162, 189]	9.27	< 0.0001	0.931	0.924	0.936	7.47
	-11.7	0.615	-19.0	< 0.0001	[-13.0, -10.4]						
	-3.45	0.662	-5.21	< 0.0001	[-4.80, -2.10]						
	0.852	0.249	3.43	0.0018	[0.344, 1.36]						
$\widehat{D}_{TNP}$	73.7	4.37	16.9	< 0.0001	[64.8, 82.6]	4.38	< 0.0001	0.919	0.908	0.962	2.74
	-4.60	0.296	-15.6	< 0.0001	[-5.21, -4.00]						
	-2.08	0.322	-6.45	< 0.0001	[-2.73, -1.42]						
	701	351	2.00	0.0554	[-17.4, 1419]						
	0.369	0.117	3.14	0.0038	[0.129, 0.609]						
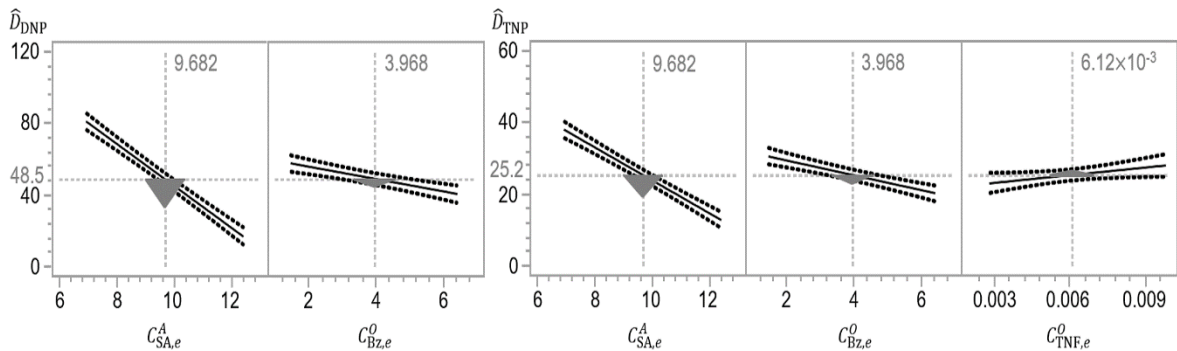
Figure 1: Plot of the sensibility indicator (▼) for the mean point of the experimental range (--) in the prediction traces of $\widehat{D}_j$ (—) with 95 % confidence interval (---).

The high values of R^2 and R_{Adj}^2 (> 0.900) proved good fit quality of the MLR models, slightly better for DNP. Prediction capability of both models was high, mainly for TNP ($R_{Pred}^2 \approx 0.96$), and the RMSE and $RMSE_{Pred}$ values were also satisfactory. The quality of the MLR fittings is illustrated in Figure 2, which

presents the experimental versus predicted values of D_j located at or near the bisector line, within the 95 % confidence interval (CI).
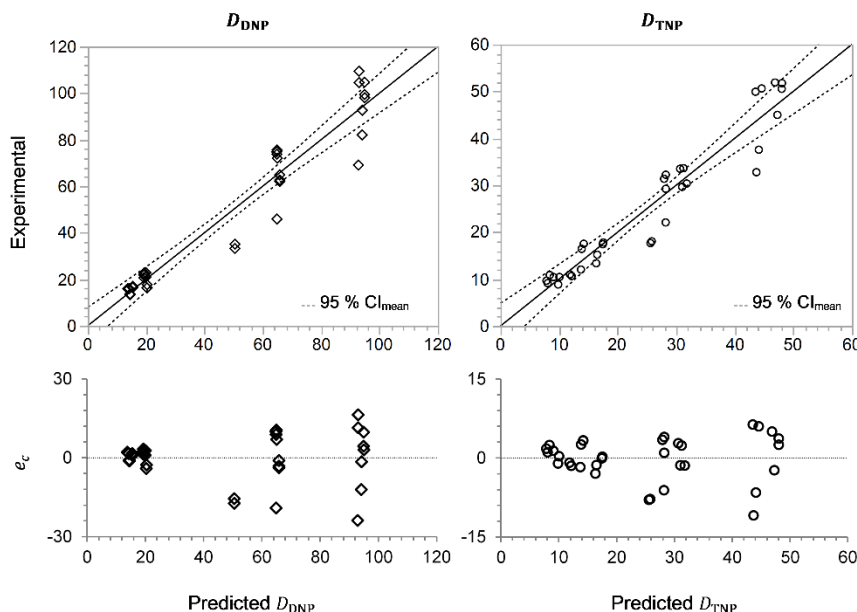

Figure 2: Experimental D_j and residuals (e_c) versus predicted values for the training set.

In Figure 2 it can be seen that the values of D_{DNP} are about twice the values of D_{TNP}. Before this work, D_{DNP} and D_{TNP} were obtained with the correlation of Zaldivar et al. (1995) assuming $D_j \approx K_{D,j}$ (Dummann, 2009). When comparing those predicted values with these experimental results the relative errors are 77-99 % and 89-100 % for DNP and TNP, respectively.

The residuals (e_c) plots presented in Figure 2, reveal a heteroscedastic behaviour (funnel pattern), which induced a non-normal distribution on e_c for both models. Therefore, the assumptions of conventional MLR models, namely homoscedasticity and normal distribution of residuals were not fully validated. Consequently, weighted least squares (WLS) models were assessed, which means the squared residuals are weighted by the variances of residuals ($1/\sigma_{e_c}^2$), so that an observation with a small variance has a higher weight than one with a large variance (Montgomery et al., 2012). The $\widehat{D}_j$ models obtained by WLS are represented by Eq(8) and Eq(9).

$$\widehat{D}_{\mathrm{DNP,WLS}} = 174 - 11.5 C_{\mathrm{SA},e}^{A} - 3.40 C_{\mathrm{Bz},e}^{O} + 0.840\left(C_{\mathrm{SA},e}^{A} - 9.682\right)\left(C_{\mathrm{Bz},e}^{O} - 3.968\right) \tag{8}$$

$$\widehat{D}_{\mathrm{TNP,WLS}} = 76.1 - 4.75 C_{\mathrm{SA},e}^{A} - 1.95 C_{\mathrm{Bz},e}^{O} + 405 C_{\mathrm{TNP},e}^{O} + 0.443\left(C_{\mathrm{SA},e}^{A} - 9.682\right)\left(C_{\mathrm{Bz},e}^{O} - 3.968\right) \tag{9}$$

The coefficients of the regressor terms of the predictive WLS models for D_j are within the 95 % CI of the correspondent models, shown in Table 1. The only hypothesis test with a p-value higher than α, even though very close, was obtained for coefficient of $C_{\mathrm{TNP},e}^{O}$ (p-value = 0.0804). Prediction capability of WLS and conventional MLR models for the training and test set was evaluated, with the values of R_{Pred}^2 and $\mathrm{RMSE}_{\mathrm{Pred}}$ being identical for D_{DNP} and slightly better for conventional MLR, in the case of D_{TNP}. These results validate the D_j MLR models expressed by Eq(6) and Eq(7) for the characterization of NP equilibrium distribution in L-L systems representative of Bz nitration.

5. Conclusions

Distribution ratio of DNP and TNP in a L-L system mimicking the industrial conditions for benzene nitration were experimentally assessed and successfully modelled following a statistical-oriented approach. Experimental values for D_{DNP} (between 10 and 120) were nearly twice the values of D_{TNP} (between 5 and 55) corroborating the higher solubility of DNP in the organic phase. These values were also very distinct from those calculated for Bz distribution at similar conditions. Predictive MLR models based on equilibrium data were developed for D_j exhibiting good fit and prediction capabilities (R^2, R_{Adj}^2 and $R_{\mathrm{Pred}}^2 > 0.900$). Overall the

models show that D_j is strongly influenced by the phase composition, with a higher concentration of sulphuric acid in the aqueous phase or of benzene in the organic phase leading to lower values of D_j in the L-L system. This work provides useful information about the distribution of nitrophenols in benzene nitration L-L systems and presents simple and reliable statistical models for D_j which can be applied in more comprehensive kinetic models.

Acknowledgments

Notation

Bz	Benzene	R^2	Coefficient of determination	g, h	Regressor indexes
C	Molar concentration, mol/dm^3	RMSE	Root mean squared error	i	Generic specie
CI	Confidence interval	SE	Standard error	j	DNP or TNP
D	Distribution ratio	t	Test statistic	O	Organic
DOE	Design of experiments	TNP	2,4,6-Trinitrophenol	Pred	Prediction
DNP	2,4-Dinitrophenol	WLS	Weighted least squares	t	Observation index of
e	Residuals	x	Regressor		test set
K_D	Distribution constant	y	Response		
MLR	Multivariate linear regression	z	Number of regressor terms	*Greek letters and symbols*	
MNB	Mononitrobenzene			α	Significance level
N	Number of observations	*Subscript and superscript*		β	Coefficient regressor
n	Number of moles, mol	A	Aqueous	ε	Error
NP	Nitrophenols	Adj	Adjusted	σ^2	Variance
p	Number of parameters β	c	Observation index of training set	^	Estimate
r	Studentized residuals	e	Equilibrium condition		

Reference

Azevedo J.C., 2015, Distribuição de nitrofenóis no sistema mononitrobenzeno-água, MSc. Thesis, Faculty of Sciences and Technology, University of Coimbra, Portugal (in Portuguese).

Berthod A., Carda-Broch S., 2004, Determination of liquid-liquid partition coefficients by separation methods, J. Chromatogr. A, 1037, 3-14, DOI: 10.1016/j.chroma.2004.01.001

Burns J.R., Ramshaw C., 1999, Development of a microreactor for chemical production, Chem. Eng. Res. Des., 77, 206-211, DOI: 10.1205/026387699526106

Costa T.J.G., Nogueira A.G., Silva D.C.M., Ribeiro A.F.G., Baptista C.M.S.G., 2013, Chemistry, Process Design, and Safety for the Nitration Industry, 49-60, vol. 1155: ACS Symposium Series, Ed. Guggenheim T.L., American Chemical Society, Washinghton, D.C., U.S.A., DOI: 10.1021/bk-2013-1155.ch004

Dummann G., 2009, Intensivierung des Wärme- und Stofftransports bei der Nitrierung von Benzol im Kapillarmikroreaktor, Ph.D. Thesis, Faculty of Biochemical and Chemical Engineering, Technical University of Dortmund, Germany (in German).

Esbensen K. H., Geladi P., 2010, Principles of Proper Validation: use and abuse of re-sampling for validation, J. Chemometrics, 24, 168-187, DOI:10.1002/cem.1310

Hanson C., Kaghazchi T., Pratt M.W.T., 1976, Industrial and Laboratory Nitrations, 132-155, vol. 22: ACS Symposium Series, Eds. Albright L.F., Hanson C., American Chemical Society, Washinghton, D.C., U.S.A., DOI: 10.1021/bk-1976-0022.ch008

Montgomery D.C., Peck E.A., Vining G.G., 2012, Introduction to linear regression analysis, 5th ed, John Wiley & Sons, Inc., Hoboken, U.S.A.

Nogueira A.G, Silva D.C.M., Reis M.S., Baptista C.M.S.G., 2013, Prediction of the by-products formation in the adiabatic industrial benzene nitration process, Chemical Engineering Transactions, 32, 1249-1254. DOI: 10.3303/CET1332209

Quadros P.A., Castro J.A.A.M., Baptista C.M.S.G., 2004, Nitrophenols reduction in the benzene adiabatic nitration process, Ind. Eng. Chem. Res., 43, 4438-4445, DOI: 10.1021/ie034263o

SAS Institute, 2016, Fit distributions <www.jmp.com/support/help/Fit_Distributions.shtml> accessed 15.07.2016

Zaldivar J.M., Molga E., Alós M.A., Hernández H., Westerterp K.R., 1995, Aromatic nitrations by mixed acid, Slow liquid-liquid reaction regime, Chem. Eng. Process., 34,543-559, DOI:10.1016/0255-2701(95)04111-7