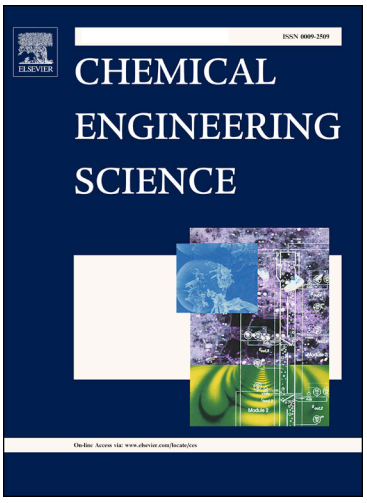

Distribution models for nitrophenols in a liquid-liquid system

A.L.C.V. Lopes, A.F.G. Ribeiro, M.P.S. Reis, D.C.M. Silva, I. Portugal, C.M.S.G. Baptista

PII: S0009-2509(18)30270-7
DOI: https://doi.org/10.1016/j.ces.2018.04.056
Reference: CES 14188

To appear in: *Chemical Engineering Science*

Received Date: 30 March 2017
Revised Date: 22 March 2018
Accepted Date: 25 April 2018

Please cite this article as: A.L.C.V. Lopes, A.F.G. Ribeiro, M.P.S. Reis, D.C.M. Silva, I. Portugal, C.M.S.G. Baptista, Distribution models for nitrophenols in a liquid-liquid system, *Chemical Engineering Science* (2018), doi: https://doi.org/10.1016/j.ces.2018.04.056

Distribution models for nitrophenols in a liquid-liquid system

A.L.C.V. Lopes[a,b,c], A.F.G. Ribeiro[a], M.P.S. Reis[b], D.C.M. Silva[a], I. Portugal[c], C.M.S.G. Baptista[b,*]

[a] *CUF - Químicos Industriais S.A., Quinta da Indústria, Rua do Amoníaco Português nº.10, Beduído, 3860-680 Estarreja, Portugal*
[b] *CIEPQPF - Department of Chemical Engineering, Faculty of Sciences and Technology, University of Coimbra, Rua 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*

Abstract

The formation of nitrophenols by-products is still of major concern for the economics and environmental impact of the industrial process of benzene (Bz) nitration to mononitrobenzene (MNB) with mixed acid (sulphuric and nitric acids). The knowledge of nitrophenol (NP) distribution ratios in the liquid-liquid mixture (D_j, $j = \{NP\}$) is desirable for process optimization and for understanding the reaction mechanisms behind nitrophenols formation.

In this study, a data-driven approach was implemented to provide prediction models for D_j of 2,4-dinitrophenol (DNP) and of 2,4,6-trinitrophenol (TNP) in a biphasic liquid system with a composition representative of the industrial processes. In the first step, screening tests were performed to identify the main variables influencing the experimental equilibrium weight fractions of nitrophenols in the aqueous phase ($w^{\mathrm{A}}_{j,e}$). Subsequently two independent data sets were built for development and external validation of prediction multivariate linear regression (MLR) models, at 30 °C. The fitting results (R^2 and $R^2_{ad} \geq 0.90$) and the prediction results ($R^2_{pred,\mathrm{DNP}} = 0.931$, $R^2_{pred,\mathrm{TNP}} = 0.908$) confirmed the quality of the $w^{\mathrm{A}}_{j,e}$

*Corresponding author: cristina@eq.uc.pt
Email addresses: alopes@eq.uc.pt (A.L.C.V. Lopes), alejandro.ribeiro@cuf-qi.pt (A.F.G. Ribeiro), marco@eq.uc.pt (M.P.S. Reis), dulce.silva@cuf-qi.pt (D.C.M. Silva), inesport@ua.pt (I. Portugal)

models. Statistical significant predictive MLR models were also developed for D_j (which is related with $w_{j,e}^{\mathrm{A}}$), at 30 °C, with DNP evidencing a higher affinity for the organic phase (i.e. $D_{\mathrm{DNP}} \approx 2D_{\mathrm{TNP}}$).

Keywords: benzene nitration, distribution ratio, multivariate linear regression, nitrophenols, predictive models

2017 MSC: 00-01, 99-00.

1. Introduction

The nitration of benzene (Bz) to mononitrobenzene (MNB) in a mixture of aqueous sulphuric and nitric acids (mixed acid), is an heterogeneous liquid-liquid (L-L) reaction that is accompanied by the formation in a reduced concentration (i.e. ppm) of undesirable and toxic nitrophenols by-products. In spite of the great attention received, the production of the nitrophenols is not yet fully understood (Burns and Ramshaw, 2002; Dummann et al., 2003; Hanson et al., 1976; Quadros et al., 2005a). A mechanistic model for nitrophenols formation, namely 2,4-dinitrophenol (DNP) and 2,4,6-trinitrophenol (TNP), has not been established yet and data-driven approaches have been used to model the benzene nitration (Nogueira et al., 2013; Portugal et al., 2009; Quadros et al., 2005b).

Another aspect is the presently unknown distribution of each NP between the two liquid phases of the system, which is important for modelling the system and better comprehension of reaction mechanism. Available studies for the distribution of the mixture components include: the partition of Bz between organic and aqueous phases characterized as a function of equilibrium temperature and mole fractions in the mixed acid (Zaldivar et al., 1995); the distributions of sulphuric and nitric acids aqueous phases in equilibrium with a MNB organic phase, expressed in terms of weight ratio, at 298 K (Suresh et al., 2009). Equilibrium data can also be found in the study of Schiefferle et al. (1976) for all species in the reaction system except nitrophenols, and in the works of Rahaman et al. (2007, 2010) for MNB in concentrated sulphuric acid. On the other hand, experimental values for NP distributions between

L-L phases are available for octanol-water, cyclohexane-water, Bz-water and MNB-water systems (Abraham et al., 2000; Berthod and Carda-Broch, 2004; Azevedo, 2015). The effects of different temperatures, pH, and solute concentrations were studied on the distribution of each NP in MNB-water systems by Azevedo (2015). None of the above works studied the distribution of nitrophenols in an acid phase.

1.1. Fundamentals of liquid-liquid equilibrium

The equilibrium distribution of a solute in a closed L-L system at a given temperature (T) and pressure (P) is characterized by a certain concentration of the solute in both phases. When solutes are present at very low concentrations and the phases are mostly immiscible, it is usually considered that they do not interfere in the phases solubility (Leo et al., 1971; Sangster, 1989). In these conditions, the closed biphasic system reaches a minimum Gibbs energy in which the phase-specific contribution is given by,

$$G_i^f = \sum_i x_i^f \mu_i^f \tag{1}$$

with

$$\mu_i^f = \mu_i^o + RT \ln \left(\gamma_i^f x_i^f\right) \tag{2}$$

where x_i^f, μ_i^f and γ_i^f are the molar fraction, chemical potential and activity coefficient of the pure compound i in phase f, respectively; μ_i^o is the chemical potential of i in the standard state; and R is the universal constant (Elliott and Lira, 1999). Under equilibrium conditions, the chemical potentials in each phase, I and II, are equal, $\mu_i^I = \mu_i^{II}$, and according to Eq. (2):

$$\gamma_i^I x_i^I = \gamma_i^{II} x_i^{II} \tag{3}$$

In the case of dilute solutions, one can assume ideal behaviour ($\gamma_i^f \simeq 1$) and the molar fraction can be expressed as the product of the molar concentration (C_i^f) by the molar volume (V_m^f), leading to $x_i^f \simeq C_i^f V_m^f$ (Chiou et al., 1982).

The distribution of solutes in L-L systems at specific experimental conditions can be characterized by the thermodynamic distribution constant (K_x) and, in the case of dilute solutions, by the distribution constant (K_D) defined by (Berthod and Carda-Broch, 2004; Leo et al., 1971; Sangster, 1989):

$$K_x = \frac{x_i^I}{x_i^{II}} \tag{4}$$

$$K_D = \frac{C_i^I}{C_i^{II}} \tag{5}$$

For dilute solutions, when the molar volumes of phases I e II are identical K_x and K_D become similar ($K_x \simeq K_D$). The distribution constant (K_D) characterizes the partition of a solute in one chemical form (e.g. molecular, ionic, elemental) between the two phases at equilibrium. However, when the solute ionizes, precipitates, reacts or participates in complexation reactions with other species in the system, the distribution ratio (D) should be adopted instead (Berthod and Carda-Broch, 2004). This parameter is defined as the ratio between the sum of all concentrations of the chemical forms in which the solute may be present, in each phase (Berthod et al., 1999; Berthod and Carda-Broch, 2004; Berthod and Mekaoui, 2011; Ingram et al., 2011). Since the nitrophenols in this study are ionizable acids, the distribution ratio depends on the ionization extent in the aqueous phase ($AH \rightleftharpoons A^- + H^+$) and is given by,

$$D_{AH} = \frac{C_{AH}^O + C_{A^-}^O}{C_{AH}^A + C_{A^-}^A} \tag{6}$$

where C_{AH}^f and $C_{A^-}^f$ are the molar concentrations of solute in its molecular and

ionized forms in phase f, respectively. Rearranging Eq. (6) one obtain,

$$D_{\mathrm{AH}} = \frac{K_{D,\mathrm{AH}} + K_{D,\mathrm{A^-}}(K_{\mathrm{AH}}/C^{\mathrm{A}}_{\mathrm{H^+}})}{1 + K_{\mathrm{AH}}/C^{\mathrm{A}}_{\mathrm{H^+}}} \tag{7}$$

where K_{AH} is the equilibrium ionization constant ($K_{\mathrm{AH}} = (C^{\mathrm{A}}_{\mathrm{H^+}} C^{\mathrm{A}}_{\mathrm{A^-}})/C^{\mathrm{A}}_{\mathrm{AH}}$) and $K_{D,\mathrm{AH}}$ and $K_{D,\mathrm{A^-}}$ are the distribution constants of the molecular and ionic forms of the solute (AH and $\mathrm{A^-}$, respectively).

The dependence of D_{AH} with pH usually exhibits a decreasing sigmoidal profile (Berthod et al., 1999) being higher at low pH, because the molecular form of the acid solute is mostly located in the organic phase. However, when the pH rises, D_{AH} decreases due to the enhanced solubility of the anionic form in the aqueous phase. Moreover, from Eq. (7), it can be seen that $D \simeq K_D$ when $C^{\mathrm{A}}_{\mathrm{H^+}}$ is much higher than K_{AH} (Berthod et al., 1999). In the particular case of DNP, ionization can be minimized by operating at pH = 2.6 (Abraham et al., 2000), while for TNP this is achieved for pH < 0.5 (Azevedo, 2015).

Given the lack of detailed fundamental knowledge about reaction system, mechanistic models cannot be developed, and the characterization of the distribution of DNP and TNP in the L-L system representative of an industrial nitration process should be accomplished by developing prediction models based on experimental data.

1.2. Predictive Models Development

A systematic data-driven approach schematically represented in Fig. 1 was followed in order to obtain simple and reliable predictive models for the distribution ratio of each NP, D_j. The main steps of this approach include the identification of the main variables by conducting a set of experiments planned according to a statistical design of experiments (DOE) strategy, the development of the models following the MLR methodology, and evaluation of the fit and predictive capabilities of the MLR models.

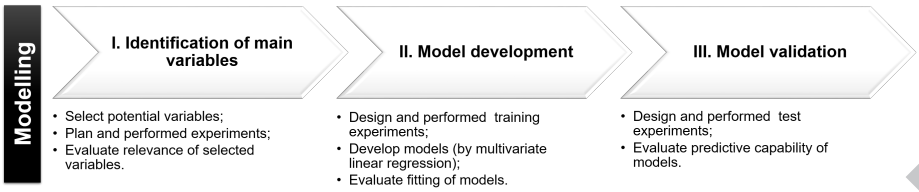

Fig. 1. Systematic approach for modelling the distribution of each NP in the L-L system.

The systematic approach for developing data-driven predictive models is strongly dependent of the availability or reliable and informative experimental data. In this context, DOE is a very efficient statistical methodology for planning the experiments for building and validation of predictive models (training and test sets, respectively) (Montgomery, 2001). The individual effects and the interactions between the independent variables (factors) can be estimated and evaluated with minimum effort (number of experiments) by application of a suitable experimental design.

Once experimental data becomes available, the conventional MLR homocedastic framework, based on the least squares method can be adopted, as long as the predictors are not highly correlated (a situation that can be avoided by selecting an appropriate DOE approach; for instance, the fractional factorial designs, used in this work, are orthogonal designs and therefore the factors do not present any mutual correlation) (Montgomery, 2012). The generic structure of the MLR models that is usually considered in DOE applications, is the following (quadratic model):

$$y_c = \hat{\beta}_0 + \sum_{g=1}^{z} \hat{\beta}_g x_{g,c} + \sum_{g=1}^{z} \hat{\beta}_{gg} x_{g,c}^2 + \sum_{g=1}^{z-1} \sum_{h=g+1}^{z} \hat{\beta}_{gh} x_{g,c} x_{h,c} + \epsilon \tag{8}$$

The error term, ϵ, is considered to have homocedastic variance, being estimate by the root mean squared error (RMSE) defined by the square root of the

variance, $\hat{\sigma}^2$,

$$\hat{\sigma}^2 = \frac{\sum_{c=1}^{N} (y_c - \hat{y}_c)^2}{N - p} \tag{9}$$

where y_c and $\hat{y}_c$ are the experimental and predicted values of the response variable for observation c, respectively; $\hat{\beta}$ designates the estimates of the coefficients of the MLR terms ($\hat{\beta}_0$ is the intercept, $\hat{\beta}_g$ are the partial regression coefficients, $\hat{\beta}_{gg}$ and $\hat{\beta}_{gh}$ with $g \neq 0$ are the coefficients of the second order interaction and quadratic terms, respectively); $x_{g,c}$ and $x_{h,c}$ are the g^{th} and h^{th} regressors for observation c; and z, p, and N are the number of regressors, of β parameters and of total observations, respectively.

When, upon inspection of the residuals, one verifies that the hypothesis of homocedasticity of the error term is not verified, the estimation machinery of the MLR model needs to be modified accordingly. In this case, instead of the usual least squares approach, it becomes necessary to apply a weighted least squares (WLS) scheme, that takes into consideration the heterocedastic (i.e., non-homogeneous) nature of the errors. In this method each response is weighted by a factor inversely proportional to the variance of its residual ($e_c = y_c - \hat{y}_c$), keeping unchanged the structure of the MLR model. Therefore, more importance is given to observations with lower variances than to those with higher variances (smaller weight). This consists in determining the estimates of the β parameters by minimizing the following WLS problem,

$$\sum_{c=1}^{N} W_c \left(y_c - \beta_0 - \sum_{g=1}^{z} \beta_g x_{g,c} + \sum_{g=1}^{z} \beta_{gg} x_{g,c}^2 + \sum_{g=1}^{z-1} \sum_{h=g+1}^{z} \beta_{gh} x_{g,c} x_{h,c} \right)^2 \tag{10}$$

with $W_c = 1/\sigma_{e_c}^2$ and where the variance was estimated from the residuals obtained in a preliminary least squares regression model fitted to the data. These residuals were grouped forming clusters and the variance of each cluster

was computed and used to calculate the weights of each observation, W_c.

The statistical significance of the conventional MLR models is usually assessed following standard procedures based on hypothesis testing. Analysis of variance (ANOVA) is used to evaluate the significance of the model as a whole with the null hypothesis, H_0: $\beta_g = \beta_{gh} = 0$ with $g \neq 0$, and the alternative hypothesis, H_1: at least one of the β's $\neq 0$ (with the exception of β_0, that is not considered in this test). The statistical significance is given by the rejection criterion of H_0, where the value of the test statistic F_0 is greater than the *Ficher* or F distribution value for a certain significance level α (normally $\alpha = 0,05$): $F_0 > F_{\alpha,z,N-p}$. The significance of each regression term is evaluated by another type of statistical hypothesis test which checks if the corresponding coefficient: is equal to zero, $\beta = 0(H_0)$, or not, $\beta \neq 0(H_1)$, for a certain significance level, α (Montgomery, 2012). The rejection of H_0 in favor of H_1 indicates that the regression term is significant and should be considered in the model developed to explain the response. This is obtained when the absolute value of the test statistic, t, is greater than the value of the t student distribution: $|t| > t_{\alpha/2,N-p}$. The outcome of the statistical tests can also be based on the p-value approach, with the rejection of H_0 occurring whenever the p-value $< \alpha$.

In addition, it is important to validate the conventional MLR assumptions namely, that observation errors are independent and identically distributed with a normal distribution around the zero mean and with constant variance σ^2 (homocedasticity). For this purpose the model residuals, e_c, which are estimates of the observation errors, must be analysed (Chatterjee and Hadi, 2006; Montgomery, 2001). Normality can be assessed by a QQ-plot or a Shapiro-Wilk test to the residual sample (SAS Institute, 2016). The studentized residuals, r_c, defined in Eq. (11) for observation c, can also be

analysed for detection of outlier observations (Montgomery, 2012).

$$r_c = \frac{e_c}{\mathrm{SE}(e_c)} \tag{11}$$

where $\mathrm{SE}(e_c)$ is the standard error of the raw residual for observation c (SAS Institute, 2016). When this normalized residuals sample follows a Normal distribution then it is expected that about 95% of its values fall within ± 2, and for r_c values higher than ± 3 the observations are considered potential outliers (Montgomery, 2001). Cook's distance measure ($d_c = r_c/p$) is used to assess the influence of an outlier observation in the model, which is indicated by a value of d_c higher than 1 (Montgomery, 2012). Homoscedasticity is observed by a random and an homogeneous dispersion (with no trend) of the e_c values plot against $\hat{y}_c$ values (Montgomery, 2012).

The fitting quality of the conventional MLR models to the training set data is assessed by the coefficient of determination, R^2, and the adjusted R^2, R^2_{ad}, defined by:

$$R^2 = \frac{\sum_{c=1}^{N} (\hat{y}_c - \bar{y}_c)^2}{\sum_{c=1}^{N} (y_c - \bar{y}_c)^2} \tag{12}$$

$$R^2_{ad} = \frac{\dfrac{\sum_{c=1}^{N} (y_c - \hat{y}_c)^2}{N-p}}{\dfrac{\sum_{c=1}^{N} (y_c - \bar{y}_c)^2}{N-1}} \tag{13}$$

where $\bar{y}_c$ is the mean of the N observations of the response variable. From Eqs. (12) and (13), it can be seen that $R^2 > R^2_{ad}$ with values close to the maximum value 1 indicating a good fit. The final validation of MLR models should be done using a test. This test set should be representative of the experimental conditions under analysis and independent from the training set (external validation). This validation approach is preferred to other

methodologies (Esbensen and Geladi, 2010) and should be followed whenever possible. In external validation, similar performance metrics to the fit quality are used to evaluate the predictive capability of the N_t experiments of the test set namely, the R^2 of prediction, R^2_{pred},

$$R^2_{pred} = \frac{\sum_{t=1}^{N_\mathrm{t}} (\hat{y}_t - \bar{y}_t)^2}{\sum_{t=1}^{N_\mathrm{t}} (y_t - \bar{y}_t)^2} \tag{14}$$

and the root mean squared error of prediction (RMSE_{pred}),

$$\mathrm{RMSE}_{pred} = \sqrt{\frac{\sum_{t=1}^{N_\mathrm{t}} (\hat{y}_t - y_t)^2}{N_\mathrm{t}}} \tag{15}$$

also called root average squared error (RASE) (SAS Institute, 2016). Another error metric useful for model comparison is the mean absolute error of prediction (MAE_{pred}) given in Eq. (16), also designated as average absolute error (AAE) (SAS Institute, 2016).

$$\mathrm{MAE}_{pred} = \frac{\sum_{t=1}^{N_\mathrm{t}} |\hat{y}_t - y_t|}{N_\mathrm{t}} \tag{16}$$

The aim of the present study is to build simple, robust and representative prediction models for the distribution ratios of DNP and TNP, D_j, under operating conditions as similar as possible to those prevailing in industrial nitration processes, but in the absence of reaction (i.e. with no nitric acid). The statistical software used in this work was JMP® Pro version 12.1.0 of SAS Institute Inc. .

2. Material and methods

During the experiments, the organic phase contained nitration grade benzene (99.994 wt%) from Gadiv Petrochemical Industries Ltd,

mononitrobenzene (99.995 wt%) from CUF - Químicos Industriais S.A., 2,4-dinitrophenol (97 wt%) from Sigma-Aldrich, and 2,4,6-trinitrophenol (99 wt%) from BDH. Other chemicals used in this work were: sulphuric acid (95–97 wt%), sodium hydroxide solution (Titripur® 1 M) and potassium dihydrogen phosphate (≥99.5 wt%) all from Merck, and acetonitrile (99.9 wt%) acquired from VWR Chemicals. The water used in this study was obtained from a Barnstead E-pure purification system.

Three different organic phases (Bz and MNB) were prepared, containing 10, 30 and 50 wt% of Bz. Precise amounts of DNP and TNP were weighted (±0.0001 g), added to the organic phase and placed in an ultrasound bath at room temperature for about 10 min, in order to ensure complete dissolution of nitrophenols. The aqueous phase consisted of five solutions of sulphuric acid, in the range of 50–75 wt%, prepared by weighting and dilution with water. The exact compositions were measured by titration with NaOH (1 M) in a 907 Titrando from Metrohm, equipped with a combined pH electrode iEcotrode Plus® as described in the studies of Santos (2005); Nogueira (2015).

The experimental procedure for mixing and reaching phase equilibrium followed the shake flask methodology (Berthod and Carda-Broch, 2004; Sangster, 1989) using the manual agitation (≈ 2 min) and decantation times (≈ 60 min) adapted and proposed by Azevedo (2015). Different volumes of each phase were employed (initial weight ratio of phases in the range of $2 - 8$) and equilibrium temperature was reached by immersing the flasks in a F25-ED thermostatic bath from Julabo (±0.1 °C) for about 5 min. After mixing, complete separation of phases was achieved by decantation followed by centrifugation in a Rotanta 460 centrifuge from Hettich for 1 min at 1250 rpm.

The equilibrium weight fractions of nitrophenols in the aqueous phase, $w_{j,e}^{\mathrm{A}}$ were measured by HPLC as described by Costa et al. (2013); Azevedo (2015). Briefly, aliquots of the aqueous phase (ca. 0.5 g) were collected and diluted 10 times with water. The diluted samples (10 µL) were injected in triplicate in an Elite LaChrom HPLC from VWR Hitachi equipped with a LiChroCART® 125-4 column (5 µm, 125 mm × 4 mm) and a LiChroCART® 4-4 guard column

(5 µm; 4 mm × 4 mm), both from Merck. Calibration was done in acidic medium (1 wt% of sulphuric acid) for the NP weight fractions in the range 1–5 $\times 10^{-6}$ with $R^2_{\mathrm{DNP}} = 0.9998$ and a $R^2_{\mathrm{TNP}} = 0.9991$. The software EZChrom Elite version 3.1.7 from Agilent was used for peak area calculations. Capability and adequacy of the measurement system for nitrophenols in the aqueous phase was confirmed by a gauge reproducibility and repeatability (R&R) study. Weight fractions of each NP in the organic phase ($w^{\mathrm{O}}_{j,e}$) were calculated by mass balances assuming an immiscible L-L system.

3. Results and Discussion

This work consists of three steps (Fig. 1) based on experimental data. The relevant process variables were identified in the first step. The second and third steps take advantage of this information to built and validate prediction MLR models.

3.1. Screening of relevant variables

This first step of the systematic data-driven approach (Fig. 1) was divided in four stages, in order to identify the most influential variables on the equilibrium weight fractions of nitrophenols in aqueous phase ($w^{\mathrm{A}}_{j,e}$). The variables assessed in each stage were:

- Stage I – equilibrium temperature (T);
- Stage II – initial weight fraction of each NP in the organic phase ($w^{\mathrm{O}}_{j,0}$);
- Stage III – initial composition of organic and aqueous phases (expressed in terms of weight fraction of Bz ($w^{\mathrm{O}}_{\mathrm{Bz},0}$) and sulphuric acid ($w^{\mathrm{A}}_{\mathrm{AS},0}$));
- Stage IV – initial weight ratio of the aqueous and organic phases ($W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$).

3.1.1. Stage I – equilibrium temperature

The influence of temperature in the distribution of nitrophenols was assessed at six temperatures in the range 20–60 °C, with $w^{\mathrm{A}}_{\mathrm{AS},0} = 0.65$, $w^{\mathrm{O}}_{\mathrm{Bz},0} = 0.50$, $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 = 2$ and $w^{\mathrm{A}}_{j,0} = 0.001$ for each NP. The experiments

were carried out in triplicate. The equilibrium weight fractions of DNP and TNP in the aqueous phase, $w^{\mathrm{A}}_{j,e}$, are plotted as a function of temperature in Fig. 2. The experimental results reveal a slight reduction in $w^{\mathrm{A}}_{\mathrm{TNP},e}$ with increasing equilibrium temperature while, on the other hand, $w^{\mathrm{A}}_{\mathrm{DNP},e}$ did not vary significantly. Therefore for TNP the temperature increase could lead to a smooth increased of its affinity for organic phase. It may also be observed that $w^{\mathrm{A}}_{\mathrm{TNP},e}$ is about twice $w^{\mathrm{A}}_{\mathrm{DNP},e}$. This difference may be explained by the stronger acidity of TNP in comparison with DNP ($\mathrm{p}K_{\mathrm{DNP}} = 4.0$ and $\mathrm{p}K_{\mathrm{TNP}} = 0.4$ (Carey, 2006)). In fact, the presence of strong electron-withdrawing nitro groups bonded to the aromatic ring of NP increases the stability of the phenoxide ion by resonance.

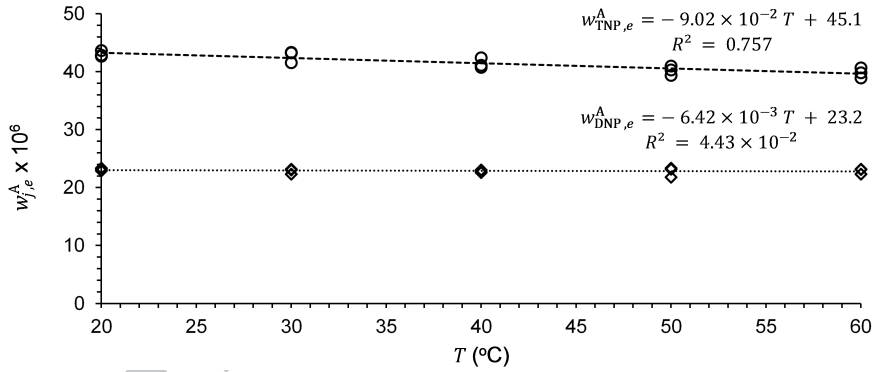

Fig. 2. Influence of temperature in the equilibrium composition of the aqueous phase (◇, DNP; and ○, TNP).

Accordingly to these results, the effect found for temperature was weak and the experiments in the next stages were all performed at 30 °C.

3.1.2. Stage II – Initial weight fraction of each NP in the organic phase

The effect of the initial composition of each NP in the organic phase on their equilibrium distributions between phases was investigated based on a three-level complete factorial DOE, with $w^{\mathrm{A}}_{j,0}$ factors in the range 0.25–2 $\times 10^{-3}$. The experiments were performed in triplicate at the following conditions: $w^{\mathrm{A}}_{\mathrm{AS},0} = 0.65$, $w^{\mathrm{O}}_{\mathrm{Bz},0} = 0.5$ and $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 = 2$. In this stage a three-level factorial was

implemented for better modelling of the expected curvature, as reported by Azevedo (2015). However, in this work, non-linear dependencies were not found as can be seen in Fig. 3, and therefore, in section 3.2 ahead a two level factorial will be adopted.

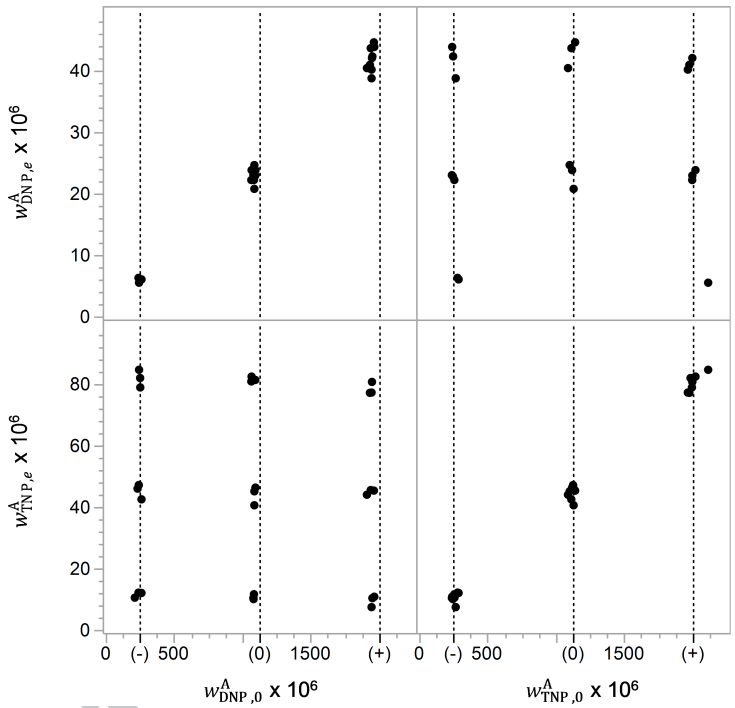

Fig. 3. Scatter plot of the equilibrium fractions in the aqueous phase in function of initial fractions in the organic phase data for DNP and TNP in the low (–), intermediate (0) and high (+) levels of the DOE.

Fig. 3 shows the linear associations between $w^{\mathrm{A}}_{j,e}$ and $w^{\mathrm{A}}_{j,0}$ with no effect of one NP in the equilibrium weight fraction of the other. Analysis of the regressors' coefficients of the MLR models (Table S.1 in Supplementary Material) revealed that these associations were positive. The quality of fits was also found to be good (R^2 and $R^2_{ad} > 0.98$).

These data were also used to relate the equilibrium weight fractions of each NP in aqueous and organic phases. Fig. 4 shows that DNP has an higher affinity for the organic phase than TNP as $w^{\mathrm{O}}_{\mathrm{DNP},e}/w^{\mathrm{A}}_{\mathrm{DNP},e} \approx 2w^{\mathrm{O}}_{\mathrm{TNP},e}/w^{\mathrm{A}}_{\mathrm{TNP},e}$.

Therefore, for the same $w^{\mathrm{O}}_{j,0}$, more TNP had been transferred once equilibrium was reached.

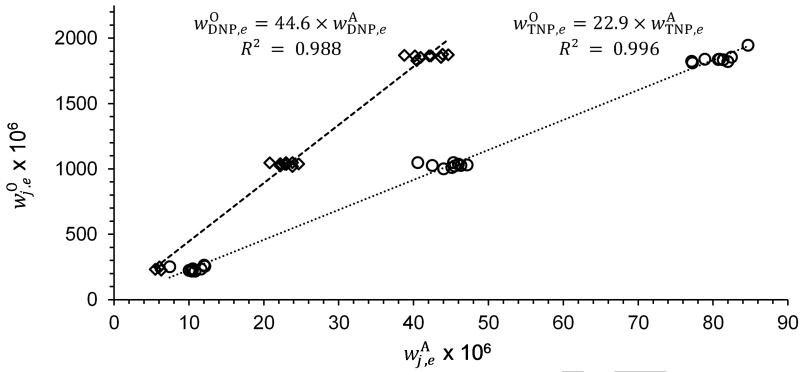

Fig. 4. Equilibrium distribution of each NP between the two phases (◇, DNP; and ○, TNP).

3.1.3. Stage III – Initial composition of organic and aqueous phases

The dependence of the $w^{\mathrm{A}}_{j,e}$ on the initial composition of the two phases was examined for $0.1 \leq w^{\mathrm{O}}_{\mathrm{Bz},0} \leq 0.5$ and $0.5 \leq w^{\mathrm{A}}_{\mathrm{AS},0} \leq 0.75$, which are within the range of industrial adiabatic nitrators operating conditions. The experiments were carried out in duplicate and designed according to a three-level complete factorial with $w^{\mathrm{O}}_{\mathrm{Bz},0}$ and $w^{\mathrm{A}}_{\mathrm{AS},0}$ as factors, keeping constant the initial weight fractions of nitrophenols, $w^{\mathrm{O}}_{j,0} = 1 \times 10^{-3}$, and the initial weight ratio of phases, $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 = 2$. A positive association of the $w^{\mathrm{A}}_{j,e}$ with the phases composition was observed, in particular for $w^{\mathrm{A}}_{\mathrm{AS},0}$ (Fig. S.1 in Supplementary Material). The statistical significance of the MLR model terms for the variables $w^{\mathrm{O}}_{\mathrm{Bz},0}$ and $w^{\mathrm{A}}_{\mathrm{AS},0}$ was confirmed and the model also presented good fitting ability, as can be seen from the high values of R^2 and R^2_{ad} (> 0.895) - see Table S.2 in Supplementary Material. The magnitude of the test statistic t for the coefficients of $w^{\mathrm{A}}_{\mathrm{AS},0}$ and $w^{\mathrm{O}}_{\mathrm{Bz},0}$ proved that the former is the most influential among the two.

The positive coefficient of $w^{\mathrm{A}}_{\mathrm{AS},0}$ indicates that the affinity of the nitrophenols with the aqueous phase increased when this phase was richer in sulphuric acid. One possible explanation is that higher values of $w^{\mathrm{A}}_{\mathrm{AS},0}$ lead to an increase of the electronic charge which, on the other hand, may favour the electrostatic forces

between the nitrophenols and the ionic species. Another interpretation is based on the solubility increment of Bz and MNB for higher sulphuric acid content (Schiefferle et al., 1976; Suresh et al., 2009). Since the nitrophenols are mostly dissolved in the organic phase in acidic conditions, they might be transferred to the aqueous phase in a higher extent due to the higher solubilities of Bz and MNB.

On the other hand, the positive coefficient of $w^{\mathrm{O}}_{\mathrm{Bz},0}$ demonstrates that the affinity of nitrophenols for organic phase was reduced by the higher content in Bz leading to an increase of $w^{\mathrm{A}}_{j,e}$. This result may be explained by the fact that nitrophenols have an higher affinity with MNB rather than with Bz.

The MLR model for $w^{\mathrm{A}}_{\mathrm{DNP},e}$ exhibited an additional and significant quadratic term with the variable $w^{\mathrm{A}}_{\mathrm{AS},0}$ (see Eq. (8)) indicating that the relation between these two variables is non linear (Table S.2). The importance of this effect will be re-evaluated again in section 3.1.4.

3.1.4. Stage IV – Initial weight ratio of the aqueous and organic phases

This last stage aimed at inferring about the effect of aqueous and organic weight ratio ($W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$) on the equilibrium compositions, $w^{\mathrm{A}}_{j,e}$. The experiments were based on a two-level complete factorial DOE for $w^{\mathrm{A}}_{\mathrm{AS},0}$, $w^{\mathrm{O}}_{\mathrm{Bz},0}$, $w^{\mathrm{O}}_{j,0}$ and $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$ factors, with $0.5 \leq w^{\mathrm{A}}_{\mathrm{AS},0} \leq 0.75$, $0.1 \leq w^{\mathrm{O}}_{\mathrm{Bz},0} \leq 0.5$, $1 \leq w^{\mathrm{O}}_{j,0} \times 10^3 \leq 2$, and $2 \leq W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 \leq 8$.

The experimental data were also well fitted by MLR models (R^2 and $R^2_{ad} > 0.98$) and the coefficients of the model terms (Table S.3 in Supplementary Material) confirmed the associations previously mentioned in stages II and III. The slightly negative relation between $w^{\mathrm{A}}_{j,e}$ and the ratio $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$ (illustrated in Fig. S.2. in Supplementary Material) is consistent with the higher affinity of nitrophenols with the organic phase under acidic conditions.

The order of relevance (from high to low importance) of the main variables on $w^{\mathrm{A}}_{j,e}$, ranked by the magnitude of the respective test statistic (t), was: $w^{\mathrm{A}}_{\mathrm{AS},0} > w^{\mathrm{O}}_{j,0} > W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 > w^{\mathrm{O}}_{\mathrm{Bz},0}$ (Table S.3). The statistical importance of an aliased

quadratic term was verified for both nitrophenols. In stage III this was valid only for DNP and identified in the $w^{\mathrm{A}}_{\mathrm{AS},0}$ term. On the other hand, for TNP the quadratic term with higher t value, although not significant, was the $w^{\mathrm{A}}_{\mathrm{AS},0}$ term. Moreover, these MLR models showed that interaction terms (Eq. (8)) between $w^{\mathrm{A}}_{\mathrm{AS},0}$ and $w^{\mathrm{O}}_{\mathrm{Bz},0}$, and between $w^{\mathrm{A}}_{\mathrm{AS},0}$, $w^{\mathrm{O}}_{\mathrm{Bz},0}$ and $w^{\mathrm{O}}_{j,0}$ have also significant contributions (Table S.3).

3.2. Equilibrium prediction models for DNP and TNP

Prediction models for the distribution ratio of DNP and TNP in L-L systems at equilibrium conditions were developed and validated (steps II and III of the systematic data-driven approach depicted in Fig. 1) with two independent data sets, namely training and test experiments. The experiments were carried out in the same range of initial conditions at 30 °C: $0.5 \leq w^{\mathrm{A}}_{\mathrm{AS},0} \leq 0.75$, $0.1 \leq w^{\mathrm{O}}_{\mathrm{Bz},0} \leq 0.5$, $250 \leq w^{\mathrm{O}}_{j,0} \times 10^6 \leq 2000$, $2 \leq W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 \leq 8$. The training set was designed with a two-level complete factorial for $w^{\mathrm{A}}_{\mathrm{AS},0}$, $w^{\mathrm{O}}_{\mathrm{Bz},0}$, $w^{\mathrm{O}}_{j,0}$ and $W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$ factors with one central point leading to 17 experimental conditions that were run in duplicate. The test set comprised 9 experiments planned by a two-level fractional factorial for the same factors, with one central point.

3.2.1. Multivariate linear regression models for DNP

The predictive MLR model for the equilibrium weight fraction $w^{\mathrm{A}}_{\mathrm{DNP},e}$, expressed in Eq. (17), was developed with the training data set using the main variables identified in section 3.1.

$$\begin{aligned}\hat{w}^{\mathrm{A}}_{\mathrm{DNP},e} = {} & -8.44 \times 10^{-5} + 1.36 \times 10^{-4} w^{\mathrm{A}}_{\mathrm{AS},0} + 1.13 \times 10^{-5} w^{\mathrm{O}}_{\mathrm{Bz},0} \\ & + 1.96 \times 10^{-2} w^{\mathrm{O}}_{\mathrm{DNP},0} - 5.03 \times 10^{-7} W^{\mathrm{A}}_0/W^{\mathrm{O}}_0 \\ & + 9.25 \times 10^{-2} (w^{\mathrm{A}}_{\mathrm{AS},0} - 0.6239)(w^{\mathrm{O}}_{\mathrm{DNP},0} - 1.51 \times 10^{-3})\end{aligned} \tag{17}$$

The positive associations between $w^{\mathrm{A}}_{\mathrm{DNP},e}$ and the initial compositions of both phases and of DNP ($w^{\mathrm{A}}_{\mathrm{AS},0}$, $w^{\mathrm{O}}_{\mathrm{Bz},0}$,$w^{\mathrm{O}}_{\mathrm{DNP},0}$) were previously mentioned in

section 3.1.4, as well as the slightly negative relation with $W_0^{\mathrm{A}}/W_0^{\mathrm{O}}$ and an interaction term between the main variables, see Eq. (17).

The distribution ratio of DNP, D_{DNP}, in the L-L system was modelled taking into consideration Eq. (6), for initial conditions expressed in molar concentrations (C), $6.946 \leq C^{\mathrm{A}}_{\mathrm{AS},0} \leq 12.42$, $1.471 \leq C^{\mathrm{O}}_{\mathrm{Bz},0} \leq 6.442$, $0.529 \leq C^{\mathrm{O}}_{\mathrm{DNP},0} \times 10^2 \leq 1.32$, $0.430 \leq C^{\mathrm{O}}_{\mathrm{TNP},0} \times 10^2 \leq 1.05$, and the ratio between phases expressed as molar ratio (n), $4.0403 \leq n_0^{\mathrm{A}}/n_0^{\mathrm{O}} \leq 30.255$. The prediction MLR model for $\hat{D}_{\mathrm{DNP}}$ given by Eq. (18) and illustrated in Fig. 5, shows that an aqueous phase with a higher sulphuric acid concentration, or an organic phase richer in Bz, lead to lower values of D_{DNP}, being $C^{\mathrm{A}}_{\mathrm{AS},0}$ the more influential factor. The MLR model for $\hat{D}_{\mathrm{DNP}}$ also exhibited a centred interaction term for $C^{\mathrm{A}}_{\mathrm{AS},0}$ and $C^{\mathrm{O}}_{\mathrm{Bz},0}$.

$$\begin{aligned}\hat{D}_{\mathrm{DNP}} = 175 - 11.7C^{\mathrm{A}}_{\mathrm{AS},0} - 3.45C^{\mathrm{O}}_{\mathrm{Bz},0} \\ + 0.852(C^{\mathrm{A}}_{\mathrm{AS},0} - 9.682)(C^{\mathrm{O}}_{\mathrm{Bz},0} - 3.968)\end{aligned} \tag{18}$$

The statistical significance of the two prediction MLR models (ANOVA in Table S.4 of Supplementary Material) and of their terms (coefficients analysis in Table S.5 of Supplementary Material) was further confirmed by analyses of their quality of fits. The significance of the main variables followed the order $w^{\mathrm{A}}_{\mathrm{AS},0} > w^{\mathrm{O}}_{\mathrm{DNP},0} > w^{\mathrm{O}}_{\mathrm{Bz},0} > W_0^{\mathrm{A}}/W_0^{\mathrm{O}}$ and $C^{\mathrm{A}}_{\mathrm{AS},0} > C^{\mathrm{O}}_{\mathrm{Bz},0}$ depending on the prediction model (Table S.5). These results demonstrate that the content of sulphuric acid in the aqueous phase is the most influential variable for both $w^{\mathrm{A}}_{\mathrm{DNP},e}$ and D_{DNP}.

The prediction capability of the MLR models for DNP, in Eqs. (17) and (18), was evaluated using the test data set. The results for RMSE, quality of fit metrics R^2 and R^2_{ad} (Eqs. (12) and (13)), and prediction capability metrics R^2_{pred}, RMSE_{pred} and MAE_{pred} (Eqs. (15) and (16)) are presented in the Table 1. The high values of the coefficients R^2, R^2_{ad} and R^2_{pred} (near or higher than 0.90) prove the high fitting and prediction capabilities of these models. In residual

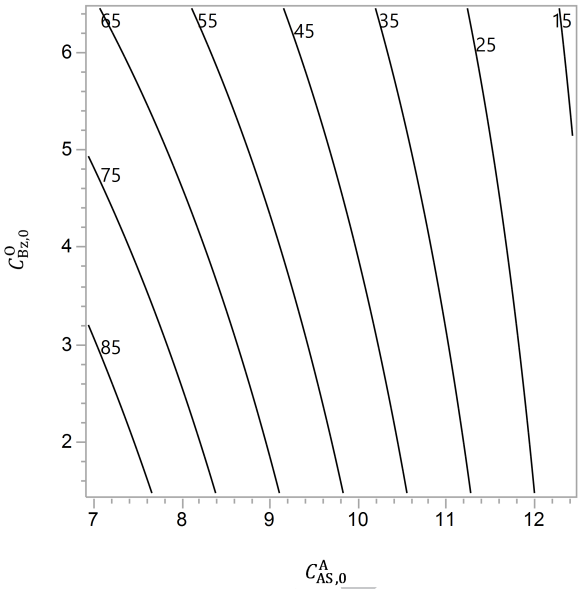

Fig. 5. Counter plot of the predicted D_{DNP} MLR model in Eq. (18).

analysis, the plots of r_c as a function of predicted $w^{\mathrm{A}}_{\mathrm{TNP},e}$ and D_{TNP} in Fig. 6 do not show outliers (values of $r_c$are within ± 3). As well influential observations were not identified (d_c values in Figs. S.3 and S.4 of Supplementary Material).

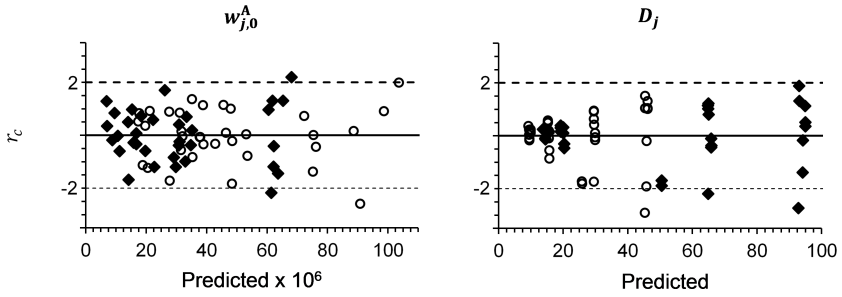

Fig. 6. Studentized residuals in function of the model values of $w^{\mathrm{A}}_{j,e}$ and D_j for training set (♦, DNP; ○, TNP).

The residuals e_c for predicted values of $\hat{w}^{\mathrm{A}}_{\mathrm{DNP},e}$ exhibit an aleatory distribution and were homoscedastic (Fig. S.5 of Supplementary Material), which behaviour was reflected on r_c sample as observed in Fig. 6. The

normality of the e_c distribution was confirmed by a QQ-plot, with the majority of the points falling along the bisection line, and by the Shapiro-Wilk test, where the p-value was greater than α, validating the MLR assumptions (Fig. S.3). On the other hand, the e_c for $\hat{D}_{\mathrm{DNP}}$ presented an heteroscedastic behaviour (funnel pattern in Fig. S.5) and, as consequence, did not followed a normal distribution (Fig. S.4). Therefore, for the later model, the conventional MLR assumptions were not fully validated. This was attributed to the calculation procedure of D_{DNP}, where a small variation on the equilibrium concentration of NP in the aqueous phase ($C^{\mathrm{A}}_{j,e}$) on the lower range, introduces a great difference in D_{DNP} values. This effect is attenuated as $C^{\mathrm{A}}_{j,e}$ increases.

To deal with the heteroscedasticity of the residuals of $\hat{D}_{\mathrm{DNP}}$ the WLS methodology was applied. The corresponding WLS model for DNP is presented in Eq. (19):

$$\begin{aligned}\hat{D}_{\mathrm{DNP,WLS}} = 174 - 11.5C^{\mathrm{A}}_{\mathrm{AS},0} - 3.40C^{\mathrm{O}}_{\mathrm{Bz},0} \\ + 0.840(C^{\mathrm{A}}_{\mathrm{AS},0} - 9.682)(C^{\mathrm{O}}_{\mathrm{Bz},0} - 3.968)\end{aligned} \tag{19}$$

All parameters of the $\hat{D}_{\mathrm{DNP,WLS}}$ model fall within the confidence interval at 95% of the $\hat{D}_{\mathrm{DNP}}$ MLR model and the prediction results for training and test sets values (R^2_{pred}) were identical, with the WLS model being only slightly worse. Hence, the use of the conventional MLR model - Eq. (18) was validated for prediction of D_{DNP} in L-L systems similar to those industrially implemented for Bz nitration.

3.2.2. Multivariate linear regression models for TNP

The same procedure described in section 3.2.1 was adopted for the development, analysis and validation of the prediction MLR models for $w^{\mathrm{A}}_{\mathrm{TNP},e}$ and D_{TNP} in the L-L system.

The MLR model for $w^{\mathrm{A}}_{\mathrm{TNP},e}$ - Eq. (20), exhibits similar associations between the main variables and the response variables as those previously observed in

the prediction models of DNP (Eqs. (17) and (18)). The coefficients in Eq. (20) confirm that higher values of $w^{\mathrm{A}}_{\mathrm{AS},0}$ or of $w^{\mathrm{O}}_{\mathrm{Bz},0}$ lead to higher values of $w^{\mathrm{A}}_{\mathrm{TNP},e}$, and enable the identification of a larger number of interactions.

$$\begin{aligned}\hat{w^{\mathrm{A}}_{\mathrm{TNP},e}} = &-8.98\times10^{-5} + 1.42\times10^{-4}w^{\mathrm{A}}_{\mathrm{AS},0} + 3.03\times10^{-5}w^{\mathrm{O}}_{\mathrm{Bz},0}\\ &+ 2.89\times10^{-2}w^{\mathrm{O}}_{\mathrm{TNP},0} - 1.02\times10^{-6}W^{\mathrm{A}}_{0}/W^{\mathrm{O}}_{0}\\ &+ 9.91\times10^{-5}(w^{\mathrm{A}}_{\mathrm{AS},0} - 0.6239)(w^{\mathrm{O}}_{\mathrm{Bz},0} - 0.2992)\\ &+ 0.105(w^{\mathrm{A}}_{\mathrm{AS},0} - 0.6239)(w^{\mathrm{O}}_{\mathrm{TNP},0} - 1.52\times10^{-3})\\ &+ 2.15\times10^{-2}(w^{\mathrm{O}}_{\mathrm{Bz},0} - 0.2992)(w^{\mathrm{O}}_{\mathrm{TNP},0} - 1.52\times10^{-3})\\ &- 3.25\times10^{-6}(w^{\mathrm{O}}_{\mathrm{Bz},0} - 0.2992)(W^{\mathrm{A}}_{0}/W^{\mathrm{O}}_{0} - 4.9060)\end{aligned} \tag{20}$$

In the MLR model for prediction of D_{TNP} expressed by Eq. (21), it is also evident that higher $C^{\mathrm{A}}_{\mathrm{AS},0}$ or $C^{\mathrm{O}}_{\mathrm{Bz},0}$ (i.e. $w^{\mathrm{A}}_{\mathrm{AS},0}$ or $w^{\mathrm{O}}_{\mathrm{Bz},0}$) lead to lower values of $\hat{D}_{\mathrm{TNP}}$ as shown in Fig. 7. The interaction term between $C^{\mathrm{A}}_{\mathrm{AS},0}$ and $C^{\mathrm{O}}_{\mathrm{Bz},0}$ has also an important contribution.

$$\begin{aligned}\hat{D}_{\mathrm{TNP}} = &\ 79.7 - 4.71C^{\mathrm{A}}_{\mathrm{AS},0} - 2.23C^{\mathrm{O}}_{\mathrm{Bz},0}\\ &+ 0.370(C^{\mathrm{A}}_{\mathrm{AS},0} - 9.682)(C^{\mathrm{O}}_{\mathrm{Bz},0} - 3.968)\end{aligned} \tag{21}$$

The statistical significance of both $w^{\mathrm{A}}_{\mathrm{TNP},e}$ and D_{TNP} prediction models and of their terms was confirmed (results in Tables S.4 and S.5). The significance of the variables follows the same order as previously reported for the DNP models, i.e. $w^{\mathrm{A}}_{\mathrm{AS},0} > w^{\mathrm{O}}_{\mathrm{TNP},0} > w^{\mathrm{O}}_{\mathrm{Bz},0} > W^{\mathrm{A}}_{0}/W^{\mathrm{O}}_{0}$ for $\hat{w}^{\mathrm{A}}_{\mathrm{TNP},e}$ model and $C^{\mathrm{A}}_{\mathrm{AS},0} > C^{\mathrm{O}}_{\mathrm{Bz},0}$ for $\hat{D}_{\mathrm{TNP}}$ model.

In residuals analysis, no outliers and influential observations were identified (see Fig. 6 and Figs. S.3 and S.4, respectively). The e_c of $\hat{w}^{\mathrm{A}}_{\mathrm{TNP},e}$ were homocedastic and exhibit a Normal distribution (see Fig. S.5). In comparison to $\hat{D}_{\mathrm{DNP}}$, the e_c of $\hat{D}_{\mathrm{TNP}}$ presented the same heteroscedastic pattern following a nonnormal distribution (Fig. S.5).

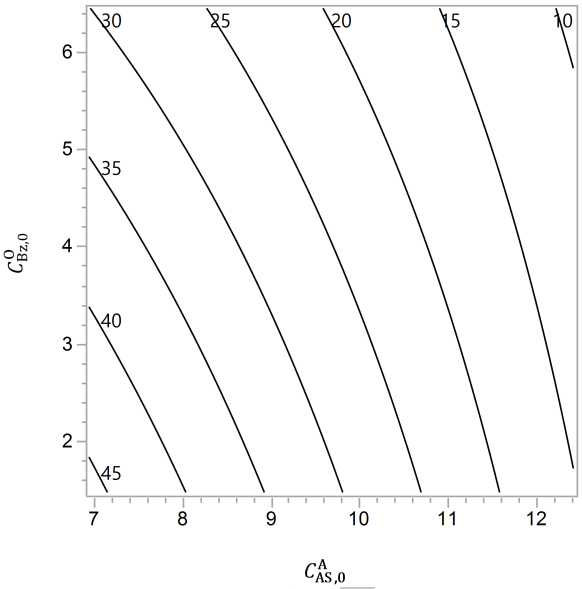

Fig. 7. Counter plot of the predicted D_{TNP} MLR model in Eq. (21).

The fitting and prediction capabilities of the MLR models of $w^{\mathrm{A}}_{\mathrm{TNP},e}$ and D_{TNP} are very satisfactory as indicated by the high values of the coefficients R^2, R^2_{ad} and R^2_{pred} ($>$0.880) presented in Table 1. Furthermore, the linear relationship between the experimental and predicted values illustrated in Fig. 8 not only for TNP but also for DNP models corroborates the high quality of the fit results.

Fig. 8 also shows that the experimental values of $w^{\mathrm{A}}_{\mathrm{TNP},e}$ are higher than the values of $w^{\mathrm{A}}_{\mathrm{DNP},e}$ and, consequently, the values of D_{DNP} are almost twice the values of D_{TNP}. Recalling that $D_j \approx K_{D,j}$, when nitrophenols ionization is minimized (i.e. under acidic conditions - see section 1.1) the D_j values from this work were compared with the K_D values calculated for Bz with the correlation of Zaldivar et al. (1995). The experimental D_j values differed significantly from the values calculated for $K_{D,\mathrm{Bz}}$ for the same equilibrium conditions, with relative deviations between 77–99 % for DNP and 89–100 % for TNP.

Table 1
Fitting and prediction capabilities of the prediction models of $w_{j,e}^{\mathrm{A}}$ and D_j in the L-L system.

Model	RMSE	Quality of fit		Prediction capability		
		R^2	R_{ad}^2	R_{pred}^2	RMSE_{pred}	MAE_{pred}
$\hat{w}_{\mathrm{DNP},e}^{\mathrm{A}}$	2.95×10^{-6}	0.982	0.979	0.924	6.16×10^{-6}	4.70×10^{-6}
$\hat{w}_{\mathrm{TNP},e}^{\mathrm{A}}$	4.22×10^{-6}	0.978	0.971	0.884	9.92×10^{-6}	6.22×10^{-6}
$\hat{D}_{\mathrm{DNP}}$	9.27	0.931	0.924	0.936	7.47	5.24
$\hat{D}_{\mathrm{TNP}}$	4.59	0.908	0.899	0.917	4.06	3.08

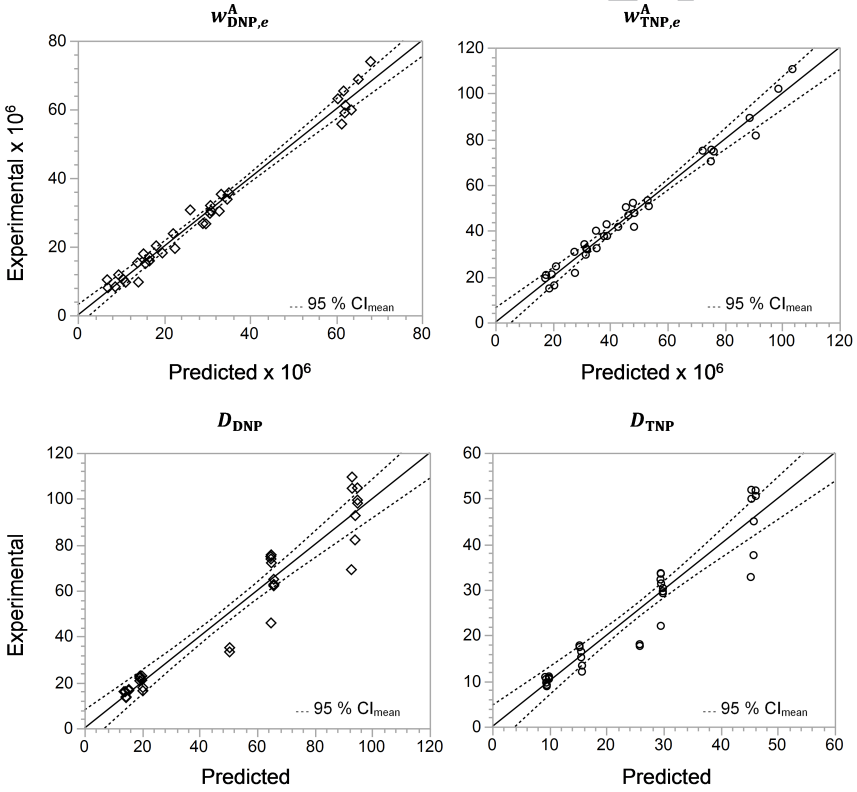

Fig. 8. Experimental versus model values for training set (◇, DNP; ○, TNP).

4. Conclusions

The distribution of DNP and TNP in a L-L system similar to the industrial production of MNB was characterized through relationships between the distribution ratio parameter, D_j, and the most influential operating variables in a range of interest for the full scale process. The values for D_{DNP} (between 10 and 120) were about two times larger than the values for D_{TNP} (between 5 and 55). Moreover, these values of D_j differed in one to two orders of magnitude when compared to the distribution of Bz ($K_{D,\mathrm{Bz}}$) calculated with the Zaldivar et al. (1995) correlation for identical experimental conditions.

The systematic data-driven approach adopted in this work proved to be very efficient in getting simple, representative and reliable prediction models. Predictive models were developed for the parameter D_j and for the equilibrium weight fraction of each NP in the aqueous phase ($w^{\mathrm{A}}_{j,e}$) with high quality of fit (R^2 and $R^2_{ad} > 0.90$) and prediction capability ($0.88 < R^2_{pred} < 0.93$ for the $w^{\mathrm{A}}_{j,e}$ and $0.91 < R^2_{pred} < 0.94$ for the D_j MLR models). The $w^{\mathrm{A}}_{j,e}$ MLR model was significantly affected by the composition of both phases (expressed by $w^{\mathrm{A}}_{\mathrm{AS},0}$ and $w^{\mathrm{O}}_{\mathrm{Bz},0}$), the weight fraction of the NP in the organic phase ($w^{\mathrm{O}}_{j,0}$), and the weight ratio of phases ($W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$) in the following order of importance: $w^{\mathrm{A}}_{\mathrm{AS},0} > w^{\mathrm{O}}_{j,0} > w^{\mathrm{O}}_{\mathrm{Bz},0} > W^{\mathrm{A}}_0/W^{\mathrm{O}}_0$. The effect of temperature was found to be negligible and the hypotheses of interaction between the two nitrophenols was discarded, because during the development of the models, in the screening step, the respective terms were not statistically significant. In the D_j MLR model, the variables found to be statistically significant were the aqueous phase composition ($C^{\mathrm{A}}_{\mathrm{AS},0}$) followed by the organic phase composition ($C^{\mathrm{O}}_{\mathrm{Bz},0}$).

The predictive models developed showed that an aqueous phase richer in sulphuric acid and higher concentrations of Bz in the organic phase (in depleting of MNB) provide higher $w^{\mathrm{A}}_{j,e}$ and, consequently, lower values of D_j.

This study also indicates that an industrial process at similar L-L system composition, with each NP tested in a concentration lower than 5000 ppm, will have at least 95% of the NP in the organic phase.

Nomenclature

Abbreviations

AAE	Average absolute error
ANOVA	Analysis of variance
Bz	Benzene
CI	Confidence interval
DNP	2,4-Dinitrophenol
DOE	Design of experiments
L-L	Liquid-liquid
MLR	Multivariate linear regression
MNB	Mononitrobenzene
MAE_{pred}	Mean absolute error
MSE	Mean squared error
NP	Nitrophenol
RASE	Root average squared error
RMSE	Root mean squared error
R&R	Reproducibility and repeatability
SA	Sulphuric acid
SE	Standard error
TNP	2,4,6-Trinitrophenol
WLS	Weighted least squares

Symbols

C	Molar concentration, $\mathrm{mol\,dm^{-3}}$
D	Distribution ratio
d_c	Cook's distance measure
e	Residual
F_0, t	Test statistics
F_α	Fisher distribution

G	Gibbs energy, $\mathrm{J\,mol^{-1}}$
K_{AH}	Equilibrium ionization constant
$K_{D,i}$	Distribution constant for species i
K_x	Thermodynamic distribution constant
N	Number of observations
n	Number of moles, mol
p	Number of parameters β
P	Pressure, Pa
R	Universal constant, $\mathrm{J\,mol^{-1}\,K^{-1}}$
r	Studentized residual
R^2	Coefficient of determination
t	Observation index of test set
t_α	t Student distribution
T	Temperature, °C
V_m	Molar volume, $\mathrm{dm^3\,mol^{-1}}$
w	Weight fraction
$W^{\mathrm{A}}/W^{\mathrm{O}}$	Initial weight ratio of aqueous phase for organic phase
x	Molar fraction
x_g, x_h	Regressors
y	Response
$\bar{y}$	Mean response
z	Number of regressor terms
$\wedge$	Estimate

Superscript

A	Aqueous
f	Phase
o	Standard state
O	Organic

Subscript

A^-	Acid in ionized form
ad	Adjusted
AH	Acid in molecular form
c	Observation index
e	Equilibrium condition
i	Generic species
j	DNP or TNP
pred	Prediction
t	Test set
0	Initial condition

Greek symbols

α	Significance level
β	Coefficient regressor
γ	Activity coefficient
μ	Chemical potential
σ^2	Variance

Acknowledgements

This work was supported by *Fundação para a Ciência e Tecnologia* (FCT) and *CUF - Químicos Industriais S.A.* [Ph.D. Grant SFRH/BDE/52071/2013].

Appendix A. Supplementary Material

Supplementary data associated with this article can be found, in the online version, at URL.

References

Abraham, M.H., Du, C.M., Platts, J.A., 2000. Lipophilicity of the Nitrophenols. The Journal of Organic Chemistry 65, 7114–7118. URL: http://dx.doi.org/10.1021/jo000840w, doi:10.1021/jo000840w.

Azevedo, J.C., 2015. Distribuição de nitrofenóis no sistema mononitrobenzeno-água. MSc. Thesis. University of Coimbra. Department of Chemical Engineering, Faculty of Sciences and Technology.

Berthod, A., Carda-Broch, S., 2004. Determination of liquid–liquid partition coefficients by separation methods. Journal of Chromatography A 1037, 3–14. URL: http://dx.doi.org/10.1016/j.chroma.2004.01.001, doi:10.1016/j.chroma.2004.01.001.

Berthod, A., Carda-Broch, S., Garcia-Alvarez-Cue, M.C., 1999. Hydrophobicity of Ionizable Compounds. A Theoretical Study and Measurements of Diuretic Octanol–Water Partition Coefficients by Countercurrent Chromatography. Analytical Chemistry 71, 879–888. URL: http://dx.doi.org/10.1021/ac9810563, doi:10.1021/ac9810563.

Berthod, A., Mekaoui, N., 2011. Distribution ratio, distribution constant and partition coefficient. Countercurrent chromatography retention of benzoic acid. Journal of Chromatography A 1218, 6024–6030. URL: http://dx.doi.org/10.1016/j.chroma.2010.12.027, doi:10.1016/j.chroma.2010.12.027.

Burns, J.R., Ramshaw, C., 2002. A Microreactor for the Nitration of Benzene and Toluene. Chemical Engineering Communications 189, 1611–1628. URL: http://dx.doi.org/10.1080/00986440214585, doi:10.1080/00986440214585.

Carey, F.A., 2006. Organic Chemistry. 6th ed., McGraw-Hill, New York, U.S.A.

Chatterjee, S., Hadi, A.S., 2006. Regression Analysis by Example. Wiley Series in Probability and Statistics. 4th ed., John Wiley & Sons, Inc., Hoboken, New Jersey, U.S.A.

Chiou, C.T., Schmedding, D.W., Manes, M., 1982. Partitioning of organic compounds in octanol-water systems. Environmental Science & Technology 16, 4–10. URL: http://dx.doi.org/10.1021/es00095a005, doi:10.1021/es00095a005.

Costa, T.J.G., Nogueira, A.G., Silva, D.C.M., Ribeiro, A.F.G., Baptista, C.M.S.G., 2013. Nitrophenolic By-Products Quantification in the Continuous Benzene Nitration Process, in: Chemistry, Process Design, and Safety for the Nitration Industry. guggenheim, thomas l. ed.. American Chemical Society, Washington, D.C., U.S.A.. volume 1155 of *ACS Symposium Series*, pp. 49–60. URL: http://pubs.acs.org/doi/abs/10.1021/bk-2013-1155.ch004.

Dummann, G., Quittmann, U., Gröschel, L., Agar, D.W., Wörz, O., Morgenschweis, K., 2003. The capillary-microreactor: a new reactor concept for the intensification of heat and mass transfer in liquid–liquid reactions. Catalysis Today 79-80, 4339–439. URL: http://dx.doi.org/10.1016/S0920-5861(03)00056-7, doi:10.1016/S0920-5861(03)00056-7.

Elliott, J.R., Lira, C.T., 1999. Introductory Chemical Engineering Thermodynamics. Prentice Hall International Series in the Physical and Chemical Engineering Sciences. 1st ed., Prentice Hall, Upper Saddle River, New Jersey, U.S.A.

Esbensen, K.H., Geladi, P., 2010. Principles of Proper Validation: use and abuse of re-sampling for validation. Journal of Chemometrics 24, 168–187. URL: http://dx.doi.org/10.1002/cem.1310, doi:10.1002/cem.1310.

Hanson, C., Kaghazchi, T., Pratt, M.W.T., 1976. Side Reactions During Aromatic Nitration, in: Albright, L.F., Hanson, C. (Eds.), Industrial and Laboratory Nitrations. American Chemical Society, Washington, D.C., U.S.A.

volume 22 of *ACS Symposium Series*, pp. 132–155. URL: http://pubs.acs.org/doi/abs/10.1021/bk-1976-0022.ch008.

Ingram, T., Richter, U., Mehling, T., Smirnova, I., 2011. Modelling of pH dependent n-octanol/water partition coefficients of ionizable pharmaceuticals. Fluid Phase Equilibria 305, 197–203. URL: http://dx.doi.org/10.1016/j.fluid.2011.04.006, doi:10.1016/j.fluid.2011.04.006.

Leo, A., Hansch, C., Elkins, D., 1971. Partition coefficients and their uses. Chemical Reviews 71, 525–616. URL: http://dx.doi.org/10.1021/cr60274a001, doi:10.1021/cr60274a001.

Montgomery, D.C., 2001. Design and analysis of experiments. 5th ed., John Wiley & Sons, Inc., New York, U.S.A.

Montgomery, D.C., 2012. Introduction to linear regression analysis. Wiley Series in Probability and Statistics. 5th ed., John Wiley & Sons, Inc., Hoboken, New Jersey, U.S.A.

Nogueira, A.G., 2015. Optimização da Nitração de Aromáticos. Ph.D. Thesis. University of Coimbra. Department of Chemical Engineering, Faculty of Sciences and Technology. URL: http://hdl.handle.net/10316/26346.

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. URL: http://www.aidic.it/cet/13/32/209.pdf, doi:10.3303/CET1332209.

Portugal, P.A.G., Reis, M.S., Baptista, C.M.S.G., 2009. Extending model prediction ability for the formation of nitrophenols in benzene nitration. Chemical Engineering Transactions 17, 117–122. URL: http://dx.doi.org/10.3303/CET0917020, doi:10.3303/CET0917020.

Quadros, P.A., Oliveira, N.M.C., Baptista, C.M.S.G., 2005a. Continuous adiabatic industrial benzene nitration with mixed acid at a pilot plant scale.

Chemical Engineering Journal 108, 1–11. URL: http://dx.doi.org/10.1016/j.cej.2004.12.022, doi:10.1016/j.cej.2004.12.022.

Quadros, P.A., Reis, M.S., Baptista, C.M.S.G., 2005b. Different Modeling Approaches for a Heterogeneous Liquid–Liquid Reaction Process. Industrial & Engineering Chemistry Research 44, 9414–9421. URL: http://dx.doi.org/10.1021/ie050205m, doi:10.1021/ie050205m.

Rahaman, M., Mandal, B.P., Ghosh, P., 2007. Nitration of nitrobenzene at high-concentrations of sulfuric acid. AIChE Journal 53, 2476–2480. URL: http://dx.doi.org/10.1002/aic.11222, doi:10.1002/aic.11222.

Rahaman, M., Mandal, B.P., Ghosh, P., 2010. Nitration of nitrobenzene at high-concentrations of sulfuric acid: Mass transfer and kinetic aspects. AIChE Journal 56, 737–748. URL: http://dx.doi.org/10.1002/aic.11989, doi:10.1002/aic.11989.

Sangster, J., 1989. Octanol–Water Partition Coefficients of Simple Organic Compounds. Journal of Physical and Chemical Reference Data 18, 1111–1229. URL: http://dx.doi.org/10.1063/1.555833, doi:10.1063/1.555833.

Santos, P.A.Q.d.O., 2005. Nitração de Compostos Aromáticos: Transferência de Massa e Reação Química. Ph.D. Thesis. University of Coimbra. Department of Chemical Engineering, Faculty of Sciences and Technology. URL: http://hdl.handle.net/10316/2022.

SAS Institute, I., 2016. Fit Distributions. URL: http://www.jmp.com/support/help/Fit_Distributions.shtml.

Schiefferle, D.F., Hanson, C., Albright, L.F., 1976. Heterogeneous Nitration of Benzene, in: Hanson, C., Albright, L.F. (Eds.), Industrial and Laboratory Nitrations. American Chemical Society, Washington, D.C., U.S.A. volume 22 of *ACS Symposium Series*, pp. 176–189. URL: http://pubs.acs.org/doi/abs/10.1021/bk-1976-0022.ch011.

Suresh, K.R., Ghosh, P., Banerjee, T., 2009. Liquid–Liquid Equilibria of Nitrobenzene–Inorganic Acid Systems at 298.15 K. Journal of Chemical & Engineering Data 54, 1302–1307. URL: http://dx.doi.org/10.1021/je800878b, doi:10.1021/je800878b.

Zaldivar, J.M., Molga, E., Alós, M.A., Hernández, H., Westerterp, K.R., 1995. Aromatic nitrations by mixed acid. Slow liquid-liquid reaction regime. Chemical Engineering and Processing: Process Intensification 34, 543–559. URL: http://dx.doi.org/10.1016/0255-2701(95)04111-7, doi:10.1016/0255-2701(95)04111-7.

Highlights

- Experimental data for liquid-liquid equilibrium distribution of nitrophenols (DNP and TNP);
- Statistical predictive models for distribution ratio based on operating conditions;
- Distribution ratio of each nitrophenol between the phases differed ($D_{DNP} \approx 2D_{TNP}$).