

# Toward the General Mechanistic Model of Liquid Chromatographic Retention

Agnieszka Kamedulska, Łukasz Kubik, Julia Jacyna, Wiktoria Struck-Lewicka, Michał J. Markuszewski, and Paweł Wiczling\*



Cite This: *Anal. Chem.* 2022, 94, 11070–11080



Read Online

ACCESS |



Metrics & More

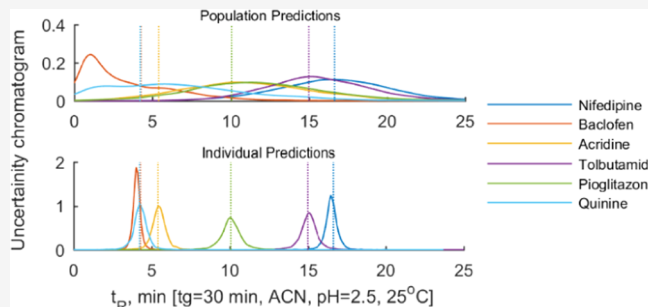


Article Recommendations



Supporting Information

**ABSTRACT:** Large datasets of chromatographic retention times are relatively easy to collect. This statement is particularly true when mixtures of compounds are analyzed under a series of gradient conditions using chromatographic techniques coupled with mass spectrometry detection. Such datasets carry much information about chromatographic retention that, if extracted, can provide useful predictive information. In this work, we proposed a mechanistic model that jointly explains the relationship between pH, organic modifier type, temperature, gradient duration, and analyte retention based on liquid chromatography retention data collected for 187 small molecules. The model was built utilizing a Bayesian multilevel framework. The model assumes (i) a deterministic Neue equation that describes the relationship between retention time and analyte-specific and instrument-specific parameters, (ii) the relationship between analyte-specific descriptors ( $\log P$ ,  $pK_a$ , and functional groups) and analyte-specific chromatographic parameters, and (iii) stochastic components of between-analyte and residual variability. The model utilizes prior knowledge about model parameters to regularize predictions which is important as there is ample information about the retention behavior of analytes in various stationary phases in the literature. The usefulness of the proposed model in providing interpretable summaries of complex data and in decision making is discussed.



Analysis of multicomponent mixtures using liquid chromatography coupled to mass spectrometry (LC/MS) yields large datasets of retention times. Although such datasets are relatively easy to collect, they are usually heterogeneous, messy, and can contain missing records or even systematic errors, e.g., due to a lack of precise control over the experiments, instruments, and data collection process. Data analysis often requires preprocessing in the form of data cleaning and filtering. Additionally, relatively complex analysis is required to extract useful information, e.g., due to the presence of analytes with different retention characteristics and various sources of variation in the data. Either empirical (statistical) models or mechanistic (generative) models can be used to describe such datasets. Empirical models are usually built based on convenience without connections to the available chromatographic theory. On the other hand, mechanistic models describe how the observed data could have arisen from the principles and fundamentals of liquid chromatography. Such models are more appropriate for extrapolations. Regardless of the approach, the accurate prediction of the retention time in liquid chromatography is required for rapid column screening, computer-assisted method development and method transfer, and unambiguous compound identification by LC/MS analyses.<sup>1</sup>

Regression models are usually employed to predict retention time based on a set of scouting (preliminary) experiments and/or various predictors, such as the molecular structure or physicochemical properties of compounds.<sup>2</sup> There are many methods for fitting a model. However, for more complex problems involving large datasets, supervised machine learning algorithms, such as different variants of partial least squares regression (PLS),<sup>3,4</sup> support vector regression (SVR),<sup>5</sup> Bayesian ridge regression (BRR), least absolute shrinkage and selection operator regression (LASSO),<sup>6</sup> adaptive boosting (AB), gradient boosting (GB), random forest (RF),<sup>7</sup> and artificial neural network (ANN), are often employed.<sup>1,8,9</sup> Most of these methods are based on classical statistics. However, Bayesian methods are gaining an increasing amount of popularity, as they allow us to take into account previously acquired knowledge about the process and to quantify the uncertainty of model parameters and predictions.<sup>10–14</sup>

Received: May 10, 2022

Accepted: July 14, 2022

Published: July 29, 2022



Bayesian methods can also be helpful in a search for a desired separation under uncertainty.<sup>15,16</sup>

In this work, we aimed to propose a general mechanistic model based on LC/MS data collected for 84 controlled experiments using a mixture of 300 small-molecule analytes. The gradient experiments differed with respect to pH, type of organic modifier, gradient duration, and temperature. The data were described using the Bayesian multilevel framework.<sup>12–14</sup> Multilevel (hierarchical) models are statistical models that contain two types of parameters: population-level parameters and individual-level parameters. Population parameters are the same for each analyte belonging to a certain population (set) of analytes. In contrast, individual-level parameters differ for each analyte.<sup>17</sup> The multilevel model usually assumes that the same deterministic equation describes the relationship between retention time and analyte-specific and instrument-specific parameters, the relationship between analyte-specific descriptors (log *P*, *pK*<sub>a</sub>, and functional groups) and chromatographically specific parameters of the analyte, and various stochastic components. The benefit of such a model is that it considers the natural nesting and various heterogeneities of the data. The Bayesian method of modeling also allows the use of prior knowledge about the analyzed phenomenon and allows the building of complex models that include all of the known and relevant facts related to the particular problem one aims to solve. Such models also quantify the uncertainty of model parameters and predictions, which is particularly useful for solving problems with limited amounts of experimental data.

This paper is organized as follows: In the following section, we describe the experimental design, data acquisition, and data filtering steps. Next, we present the model using standard statistical notation. Subsequently, we present the results of the inferences based on the model. Finally, we illustrate the usefulness of the model for summarizing complex data and in making predictions given access to different types of preliminary data. We close with discussion and conclusions.

## EXPERIMENTAL SECTION

**Data.** The data were collected by performing 84 different liquid chromatography experiments using a mixture of 300 analytes. The experiments differed with respect to gradient duration (30, 90, and 270 min), pH of the mobile phase (from 2.5 to 10.5), type of organic modifier (methanol (MeOH) or acetonitrile (ACN)), and column temperature (25 and 35 °C). The detailed conditions are listed in Table S1.

The liquid chromatography experiments were carried out using an Agilent Technologies 1260 Infinity system (Agilent Technologies, Waldbronn, Germany) composed of a binary pump, a membrane degasser, an autosampler, a column thermostat, and a 6224 time-of-flight (TOF) mass spectrometer with a dual electrospray ionization source (Dual ESI) in positive polarity, using an XBridge-C18 (Waters Ltd., Milford, MA, 3 mm × 50 mm, 2.5 μm) column. A drying gas (nitrogen) flow was set to 11 L/min at 350 °C. The nebulizer pressure was set to 50 psi at a capillary voltage of 4000 V. The skimmer, octopole voltage, and fragmentor voltage were set to 65, 750, and 84 V, respectively. The analyses were carried out in scan mode using a 50–1200 *m/z* range. The reference substances (with masses of 121.050873 and 922.009798) were delivered to the electrospray ionization (ESI) source and monitored to control the mass measurement accuracy. The extra column volume and system dwell volume (*V*<sub>d</sub>) equaled 0.020 and 1.05 mL, respectively. The column hold-up volume (*V*<sub>0</sub>) was 0.266

mL; the flow rate (*F*) was 0.5 mL/min; and the injection volume was 2 μL. After the end of each gradient, the final content of organic modifier was held for an additional 6 min. A short, 3.5 min, postrun program was performed to maintain the initial conditions of the system and to equilibrate the chromatographic column.

Approximately 5 μmol of each reference substance was weighed and diluted in MeOH to obtain approximately 15 μmol/mL concentrations of each analyte. Undissolved substances were treated with NaOH or formic acid solutions until complete dissolution was achieved. Subsequently, all diluted substances were mixed (using volumes corresponding to 0.3 μmol of each substance) into one sample. The resulting sample was diluted with MeOH 1:499 (v/v) (1 nmol/mL).

Ammonium bicarbonate, ammonium acetate, and ammonium formate were selected as buffers to control the pH of the mobile phase during chromatographic separation. The buffers were prepared at a concentration of 10 mM. The pH of the buffers (nominal aqueous pH) was adjusted to the desired pH (ammonium formate: 2.5, 3.3, 4.1, 8.9, and 9.7; ammonium acetate: 4.9 and 5.8; and ammonium bicarbonate: 6.8 and 10.5) by an appropriate addition of formic acid, acetic acid, and ammonia, respectively. The prepared buffers were filtered using 0.45 μm nylon filters and degassed.

The pH was measured at 25 and 35 °C using an S220 pH meter (Mettler Toledo, Greifensee, Switzerland) with an InLab Routine Pro ISM electrode after mixing an organic modifier with the buffer solution. The pH meter electrode was calibrated using a standard aqueous standard. The relationship between pH and the content of organic modifier for various combinations of organic modifier and buffer was experimentally determined prior to the chromatographic analysis. In this setting, pH, and consequently, *pK*<sub>a</sub> values correspond to an absolute pH scale.<sup>18</sup> The obtained data were then described using quadratic equations for each nominal pH, temperature, and organic modifier (36 equations in total)

$$\text{pH}_{m,b,t,j} = \text{pHo}_{m,b,t} + \alpha 1_{m,b,t} \cdot \varphi_j + \alpha 2_{m,b,t} \cdot \varphi_j^2$$

where *m* denotes the type of organic modifier (1 for MeOH, 2 for ACN), *b* represents nominal pH (*b* = 1...9, corresponding to nominal pH of 2.5, 3.3, 4.1, 4.9, 5.8, 6.8, 8.9, 9.7, and 10.5, respectively), *t* indicates temperature (1–25 and 2–35 °C), *j* denotes levels of organic modifier content, and *pHo*<sub>*m,b,t*</sub>, *α*1<sub>*m,b,t*</sub>, and *α*2<sub>*m,b,t*</sub> are regression coefficients specific for a given condition.

The pH measurements and predictions are given in Figure S1. The estimated values of *pHo*, *α*1, and *α*2 for each chromatographic condition were added to the dataset.

**Data Extraction Procedure and Data Filtering Step.** The MassHunter Profinder B.08.00 (Agilent Technologies, Waldbronn, Germany) was selected to find all of the matches per formula using “Batch Targeted Feature Extraction” (containing 300 predefined masses for each analyte included in the mixture). A match tolerance of ±20.00 ppm was set for the identification of compounds, and possible ionization adducts of pseudomolecular ions (H<sup>+</sup>, NH<sub>4</sub><sup>+</sup>) were taken into account. The resulting data were exported as a detailed CSV. All 84 files were then merged and combined with the experimental design data and analyte-specific information.

The data for analysis were restricted to analytes that had “Identification Scores” higher than 95%, that were present on at least 42 chromatograms, and that had less than 2

dissociation steps in a pH range from 2 to 11. This process led to the final dataset with 187 analytes (out of the initial 300 analytes).

The molecular structure of the analytes was converted from SMILES format to MDL mol format using OpenBabel.<sup>19</sup> The input molecules were then analyzed for the presence of approximately 204 functional groups and structural elements using Checkmol (version 0.5b N. Haider, University of Vienna, 2003–2018).<sup>20</sup> Functional groups that were not present on any analyte and functional groups merging other simpler functional groups were excluded from the analysis. The lipophilicity ( $\log P$ ), dissociation constant ( $\text{pK}_{\text{a}}^{\text{lit}}$ ), and predicted error of  $\text{pK}_{\text{a}}$  ( $\text{pK}_{\text{a}}^{\text{lit}} - \text{pK}_{\text{a}}^{\text{pred}}$ ) were calculated using the ACD/Labs program<sup>21</sup> based on the structures of analytes generated from SMILES strings. Only  $\text{pK}_{\text{a}}^{\text{lit}}$  from 2 to 11 were considered. The  $\log P$  value of the analytes ranged from −4.49 to 7.81, and the molecular mass ranged from 120 to 915. There were 62 and 126 compounds with at least one acidic and basic group. 21 (11.2%) analytes were neutral, 111 (59.4%) analytes were monoprotic and 55 (29.4%) analytes were diprotic in the considered range of pH values. 60 unique functional groups were identified and utilized during the model-building process. The raw data applied in this work are shown in the Supporting Information and graphically presented in Figure S2. The functional groups and their frequency of occurrence are characterized in Figure S3.

**Structural Model.** A standard chromatographic model was employed in this work.<sup>22,23</sup> The following function describes the relationship between the isocratic retention factor and pH for an analyte with  $R$  dissociation steps and  $R + 1$  forms<sup>24</sup>

$$k_{i,m,b,t,j} = \frac{k_{1,i,m,b,t,j} + \sum_{r=1}^R k_{r+1,i,m,b,t,j} 10^{r \cdot \text{pH}_{m,b,t,j} - \sum_{r=1}^R \text{pK}_{a,r,i,m,j}}}{1 + \sum_{r=1}^R 10^{r \cdot \text{pH}_{m,b,t,j} - \sum_{r=1}^R \text{pK}_{a,r,i,m,j}}} \quad (1)$$

where  $r$  represents the dissociation step,  $i$  denotes the analyte,  $m$  indicates the organic modifier,  $j$  represents the organic modifier content,  $b$  denotes the pH, and  $t$  indicates the temperature. Thus,  $\text{pK}_{a,r,i,m,j}$  denotes the  $r$ th dissociation constant of the  $i$ th analyte for the  $m$ th organic modifier and  $j$ th organic modifier content,  $k_{r,i,m,b,t,j}$  represents the retention factor of a particular form of the  $i$ th analyte in a given chromatographic condition and  $k_{i,m,b,t,j}$  represents the isocratic retention factors in a given chromatographic condition. Furthermore, it was assumed that  $k$  depends on the organic modifier content, pH, and temperature according to the following equation

$$\log k_{r,i,m,b,t,j} = \log kw_{r,i} - \frac{S1_{r,i,m} \cdot (1 + S2_m) \cdot \varphi_j}{1 + S2_m \cdot \varphi_j} + d \log kT_i \cdot \frac{T_t - 25}{10} + |\text{charge}A_{r,i}| \cdot \text{apHA} \cdot (\text{pH}_{m,b,t,j} - 7) + |\text{charge}B_{r,i}| \cdot \text{apHB} \cdot (\text{pH}_{m,b,t,j} - 7) \quad (2)$$

where  $\log kw_{r,i}$  represents the logarithm of retention factors extrapolated to 0% of organic modifier content at 25 °C for mobile phase at pH = 7 for the neutral and dissociated forms of the analyte;  $S1_{r,i,m}$  and  $S2_m$  denote slopes in the Neue equation;

$d \log kT_i$  denotes the change in  $\log kw$  due to the increase in temperature by 10 °C, a pH denotes pH effects for cations and anions (common for all analytes);  $\text{charge}A_{r,i}$  and  $\text{charge}B_{r,i}$  denote a charge state of an analyte ( $\text{charge}A_{r,i} = \{0, -1, -2, \dots\}$  for anions, and  $\text{charge}B_{r,i} = \{0, 1, 2, \dots\}$  for cations); and  $|l|$  denotes absolute value. In this parametrization of the Neue equation, the  $S1$  parameter reflects the difference between the logarithm of retention factors corresponding to water (0% organic modifier content) and MeOH or ACN (100% organic modifier content) as eluents.

Furthermore, a linear relationship between  $\text{pK}_{\text{a}}$  values and organic modifier content was assumed

$$\text{pK}_{a,r,i,m,j} = \text{pK}_{a,r,i} w_{r,i} + \alpha_{r,i,m} \cdot \varphi_j \quad (3)$$

where  $\text{pK}_{a,r,i,m,j}$  denotes dissociation constants of an analyte in given chromatographic conditions,  $\text{pK}_{a,r,i} w_{r,i}$  denotes aqueous  $\text{pK}_{\text{a}}$ , and  $\alpha_{r,i,m}$  denotes the slope due to changes in the organic modifier. The linear relationship is generally valid for  $\varphi_j < 0.8$ .

**Measurement-Error Model.** The observed retention factors ( $t_{\text{Robs},z}$ ) were modeled using the following model

$$t_{\text{Robs},z} \sim \text{student}_t(\nu, t_{\text{R},z}, \sigma_{[z]}) \quad (4)$$

where  $z$  denotes the  $z$ th measurement and  $\text{student}_t$  denotes the Student's  $t$ -distribution with the mean given by the predicted retention time  $t_{\text{R},z}$ , scale  $\sigma_i$  (analyte-specific), and normality parameter  $\nu$ . The retention time  $t_{\text{R},z}$  under an organic modifier gradient was calculated utilizing the well-known integral equation

$$\int_0^{t_{\text{R},z} - t_0 - t_e} \frac{dt}{t_0 \cdot k_i(t)} = 1 \quad (5)$$

where  $t_0$  denotes column hold-up (dead) time,  $t_e$  denotes extra column time, and  $k_i(t)$  denotes the instantaneous isocratic retention factor corresponding to the mobile phase composition at time  $t$  at the column inlet for a particular measurement. The numerical solution of this integral equation was carried out using the method of steps with 4 and 10 steps for methanol and acetonitrile gradients using the method proposed by Nikitas et al.

**Analyte-Level Model.** The  $\log kw_{r,i}$  and  $S1_{r,i,m}$  parameters for each analyte form were calculated based on  $\log kw$  and  $S1$  of the neutral form of an analyte and the difference in  $\log kw$  or  $S1$  values between the neutral form of an analyte and the ionized form of an analyte. The  $S1$  parameter was separately estimated for MeOH ( $m = 1$ ) and ACN ( $m = 2$ )

$$\log kw_{r,i} = \log kwN_i + |\text{charge}A_{r,i}| \cdot d \log kwA_{r,i} + |\text{charge}B_{r,i}| \cdot d \log kwB_{r,i} \quad (6)$$

$$S1_{r,i,m=1} = S1mN_i + |\text{charge}A_{r,i}| \cdot dS1mA_{r,i} + |\text{charge}B_{r,i}| \cdot dS1mB_{r,i} \quad (7)$$

$$S1_{r,i,m=2} = S1aN_i + |\text{charge}A_{r,i}| \cdot dS1aA_{r,i} + |\text{charge}B_{r,i}| \cdot dS1aB_{r,i} \quad (8)$$

Furthermore, the  $\alpha$  parameters were assumed to be different for acids and bases and were separately estimated for MeOH ( $m = 1$ ) and ACN ( $m = 2$ )

$$\alpha_{r,i,m=1} = \alpha mA_{r,i} \cdot \text{group}A_{r,i} + \alpha mB_{r,i} \cdot \text{group}B_{r,i} \quad (9)$$

$$\alpha_{r,i,m=2} = \alpha a A_{r,i} \cdot \text{group} A_{r,i} + \alpha b B_{r,i} \cdot \text{group} B_{r,i} \quad (10)$$

where  $\text{group} A_{r,i}$  and  $\text{group} B_{r,i}$  denote the type of dissociating group ( $\text{group} A_{r,i} = 1$  if acidic and 0 otherwise;  $\text{group} B_{r,i} = 1$  if basic and 0 otherwise).

The second-level part of the model describes the relationship between analyte-specific parameters and predictors. The parameters for the neutral form of an analyte were assumed to be correlated and related to  $\log P$  and functional groups

$$\begin{bmatrix} \log kwN_i \\ S1mN_i \\ S1aN_i \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \theta_{\log kwN} + \beta_{\log kwN} \cdot (\log P_i - 2.2) + \pi_{\log kwN} \cdot X \\ \theta_{S1mN} + \beta_{S1mN} \cdot (\log P_i - 2.2) + \pi_{S1mN} \cdot X \\ \theta_{S1aN} + \beta_{S1aN} \cdot (\log P_i - 2.2) + \pi_{S1aN} \cdot X \end{bmatrix}, \Omega \right) \quad (11)$$

where MVN denotes the multivariate normal distribution;  $\theta_{\log kwN}$ ,  $\theta_{S1mN}$ , and  $\theta_{S1aN}$  are the mean values of individual chromatographic parameters that correspond to a typical analyte with  $\log P = 2.2$ , with no functional groups at 25 °C;  $\beta_{\log kw}$ ,  $\beta_{S1m}$ , and  $\beta_{S1a}$  are regression coefficients between the individual chromatographic parameters and the  $\log P_i$  values; and  $\pi$  is an effect of each functional group on chromatographic parameters with separate values for  $\log kwN$ ,  $S1mN$ , and  $S1aN$ .  $\pi$  represents the difference in chromatographic parameters due to the presence of a functional group, assuming all else being equal.  $X$  is a matrix of size  $187 \times 60$  that decodes the number of functional groups present on each analyte. The lack of a particular functional group was denoted as 0, and the presence of a functional group was denoted as  $n$ , with  $n$  denoting the number of functional groups of the same type present on each analyte and  $\Omega$  denoting a variance-covariance matrix. To ease the specification of the prior distribution,  $\Omega$  was decomposed into a vector of scales (vector  $\omega = [\omega_{\log kwN}, \omega_{S1mN}, \omega_{S1aN}]$ ) and a correlation matrix ( $3 \times 3$  matrix  $\rho_1$ ) based on the formula

$$\Omega = \text{diag}(\omega) \cdot \rho_1 \cdot \text{diag}(\omega) \quad (12)$$

The difference in retention between the ionized form of an analyte and the neutral form of an analyte was separately estimated for acids and bases

$$\begin{aligned} d \log kw A_{r,i} &\sim N(\theta_{d \log kw A}, \kappa_{d \log kw}), \quad d \log kw B_{r,i} \\ &\sim N(\theta_{d \log kw B}, \kappa_{d \log kw}) \end{aligned} \quad (13)$$

$$\begin{aligned} dS1m A_{r,i} &\sim N(\theta_{dS1m A}, \kappa_{dS1m}), \quad dS1m B_{r,i} \\ &\sim N(\theta_{dS1m B}, \kappa_{dS1m}) \end{aligned} \quad (14)$$

$$dS1a A_{r,i} \sim N(\theta_{dS1a A}, \kappa_{dS1a}), \quad dS1a B_{r,i} \sim N(\theta_{dS1a B}, \kappa_{dS1a}) \quad (15)$$

where  $\theta_{d \log kw}$  and  $\theta_{dS1}$  denote the mean, and  $\kappa_{d \log kw}$  and  $\kappa_{dS1}$  represent the standard deviation for acids (A) and bases (B) in MeOH (a) or ACN (m).

The effect of temperature was assumed to differ for each analyte and to follow a normal distribution

$$d \log kT_i \sim N(\theta_{d \log kT}, \omega_{d \log kT}) \quad (16)$$

The  $pK_a w_{r,i}$  in water was assumed to be equal to the aqueous literature values  $pK_{a \text{lit},r,i}$  assuming a reported measurement error  $pK_{a \text{literror},r,i}$

$$pK_a w_{r,i} \sim N(pK_{a \text{lit},r,i}, pK_{a \text{literror},r,i}) \quad (17)$$

Furthermore, the  $\alpha$  values for MeOH and ACN were assumed to be correlated for acids and bases

$$\begin{aligned} \begin{bmatrix} \alpha m A_{r,i} \\ \alpha a A_{r,i} \end{bmatrix} &\sim \text{MVN} \left( \begin{bmatrix} \theta_{\alpha mA} \\ \theta_{\alpha aA} \end{bmatrix}, \text{Tau} \right), \quad \begin{bmatrix} \alpha m B_{r,i} \\ \alpha a B_{r,i} \end{bmatrix} \\ &\sim \text{MVN} \left( \begin{bmatrix} \theta_{\alpha mB} \\ \theta_{\alpha aB} \end{bmatrix}, \text{Tau} \right) \end{aligned} \quad (18)$$

where  $\theta_{\alpha mA}$ ,  $\theta_{\alpha aA}$ ,  $\theta_{\alpha mB}$ , and  $\theta_{\alpha aB}$  denote the mean  $\alpha$  for acids (A) and bases (B) in MeOH (a) or ACN (m).  $\text{Tau}$  was also decomposed into a vector of scales vector  $\tau = [\tau_{\alpha m}, \tau_{\alpha a}]$  and a correlation matrix ( $2 \times 2$  matrix  $\rho_2$ ) based on the following formula

$$\text{Tau} = \text{diag}(\tau) \cdot \rho_2 \cdot \text{diag}(\tau) \quad (19)$$

**Priors.** The Bayesian model requires specification of priors that provide a likely range of model parameters expected before the data are observed. Priors also provide the appropriate scales for a given analysis and introduce regularization into the analysis. In this work, the prior information was selected to be weakly informative and in agreement with known facts about analyte retention and gradient chromatography.

The retention factor of the typical neutral form of an analyte ( $\theta_{\log kwN}$ ) was assumed to equal  $2.2 \pm 2$ , where 2.2 and 2 correspond to the mean and standard deviation of  $\log P_i$  values. The typical S1 values in MeOH and ACN ( $\theta_{S1mN}$  and  $\theta_{S1aN}$ ) were assumed to be approximately 4 and 5, respectively, with a standard deviation of 1.<sup>25</sup>

$$\theta_{\log kwN} \sim N(2.2, 2), \quad \theta_{S1mN} \sim N(4, 1), \quad \theta_{S1aN} \sim N(5, 1) \quad (20)$$

The slope ( $\beta$ ) was assumed to be nearly 1 ( $\pm 0.125$ ) for the  $\log kwN$  vs  $\log P$  relationship and to be 0.5 ( $\pm 0.5$ ) for the S1 vs  $\log P$  relationships

$$\beta_{\log kwN} \sim N(1, 0.125), \quad \beta_{S1mN}, \beta_{S1aN} \sim N(0.5, 0.5) \quad (21)$$

The regression parameters that describe the effects of substituents were given the priors that assume small effects

$$\pi_{\log kwN,1:60}, \pi_{S1mN,1:60}, \pi_{S1aN,1:60} \sim N(0, \sigma_\pi) \quad (22)$$

$$\sigma_{\pi \log kwN}, \sigma_{\pi, S1mN}, \sigma_{\pi, S1aN} \sim N_+(0, 0.1) \quad (23)$$

where  $\sigma_\pi$  is a standard deviation of the individual  $\pi_{1:60}$  values for a particular parameter.

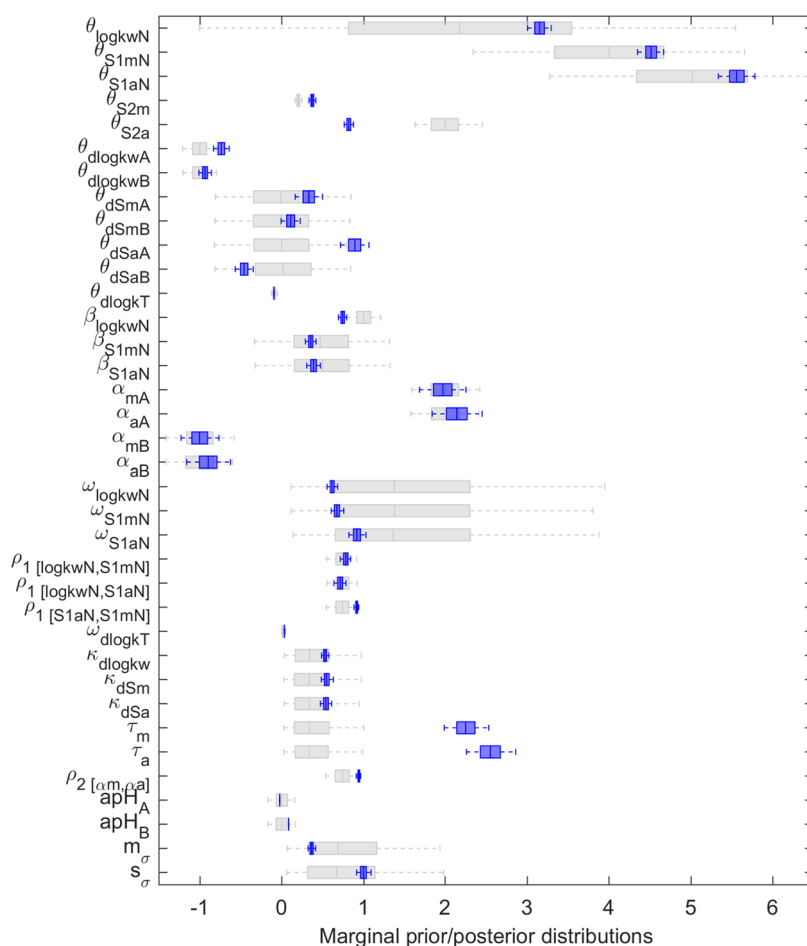
The value of  $d \log kw$ , which was assumed to be nearly  $-1$  ( $\pm 0.125$ ), corresponds to a typical ratio of the retention factors of the neutral and acidic/basic forms on the order of 10.<sup>26</sup> The difference in S1 values between the dissociated forms of acids and bases and the neutral form of an analyte was assumed to be similar on average to that in water

$$\begin{aligned} \theta_{d \log kw A}, \theta_{d \log kw B} &\sim N(-1, 0.125), \quad \theta_{dS1m A}, \theta_{dS1m B}, \\ \theta_{dS1a A}, \theta_{dS1a B} &\sim N(0, 0.5) \end{aligned} \quad (24)$$

The priors for parameters that describe the effect of pH on retention for cations and anions were selected to avoid such a relationship

$$\text{pH} \sim N(0, 0.1) \quad (25)$$





**Figure 1.** Graphical display of the marginal posterior (blue) and prior (gray) distributions for the population-level parameters. The exact values are given in Table S2.

The S2 parameters in the Neue equation were assumed to be similar for all analytes and analyte forms but were assumed to differ for each organic modifier. S2 was assumed to be positive and nearly 0.2 for MeOH and 2 for ACN.<sup>27</sup>

$$\theta_{S2m} \sim \text{lognormal}(\ln(0.2), 0.125), \theta_{S2a} \sim \text{lognormal}(\ln(2), 0.125) \quad (26)$$

According to the literature, acids (such as carboxylic acids and phenol) show modest increases in  $pK_a$  (1–2 pK units) in solvent mixtures containing up to 60–70% MeOH and ACN. Basic compounds (such as amines and anilines) display a universal decrease in  $pK_a$  (~1 pK unit) up to approximately 80% organic solvent.<sup>28</sup> Based on these results, it was assumed that  $pK_a$  increases with MeOH/ACN content for acids with a typical slope of 2.0 ( $\pm 0.125$ ) and that  $pK_a$  decreases for bases with a typical slope of  $-1$  ( $\pm 0.125$ ).

$$\theta_{\alpha mA}, \theta_{\alpha aA} \sim N(2, 0.125), \theta_{\alpha mB}, \theta_{\alpha aB} \sim N(-1, 0.125) \quad (27)$$

Between-analyte variabilities were given weekly informative priors of the following form

$$\omega \sim N_+(0, 2), \kappa \sim N_+(0, 0.5), \tau \sim N_+(0, 0.5) \quad (28)$$

where  $N_+$  denotes the half-normal distribution. The correlation matrix was given the following prior:

$$p(\rho_1) \propto \text{LKJ}(3) \cdot \prod_u N(0.75, 0.125) \quad (29)$$

$$p(\rho_2) \propto \text{LKJ}(2) \cdot \prod_u N(0.75, 0.125) \quad (30)$$

where  $\rho_1$  and  $\rho_2$  have a joint prior consisting of a uniform LKJ prior (Lewandowski et al. distribution<sup>29</sup>) on the matrix and a normal on its elements.<sup>30</sup> The symbol  $\propto$  means “is proportional to”. LKJ(3) and LKJ(2) ensure that the densities are uniform over correlation matrices of order 3 and order 2, respectively, and  $u$  denotes the unique lower triangular elements of  $\rho_1$  and  $\rho_2$ . High correlations can be expected for  $\log kwN_i$ ,  $S1mN_i$ , and  $S1aN_i$ , and similarly, between  $\alpha m_{i,j}$  and  $\alpha a_{i,j}$ . Here, we assumed that all of the correlation coefficients are high and positive.

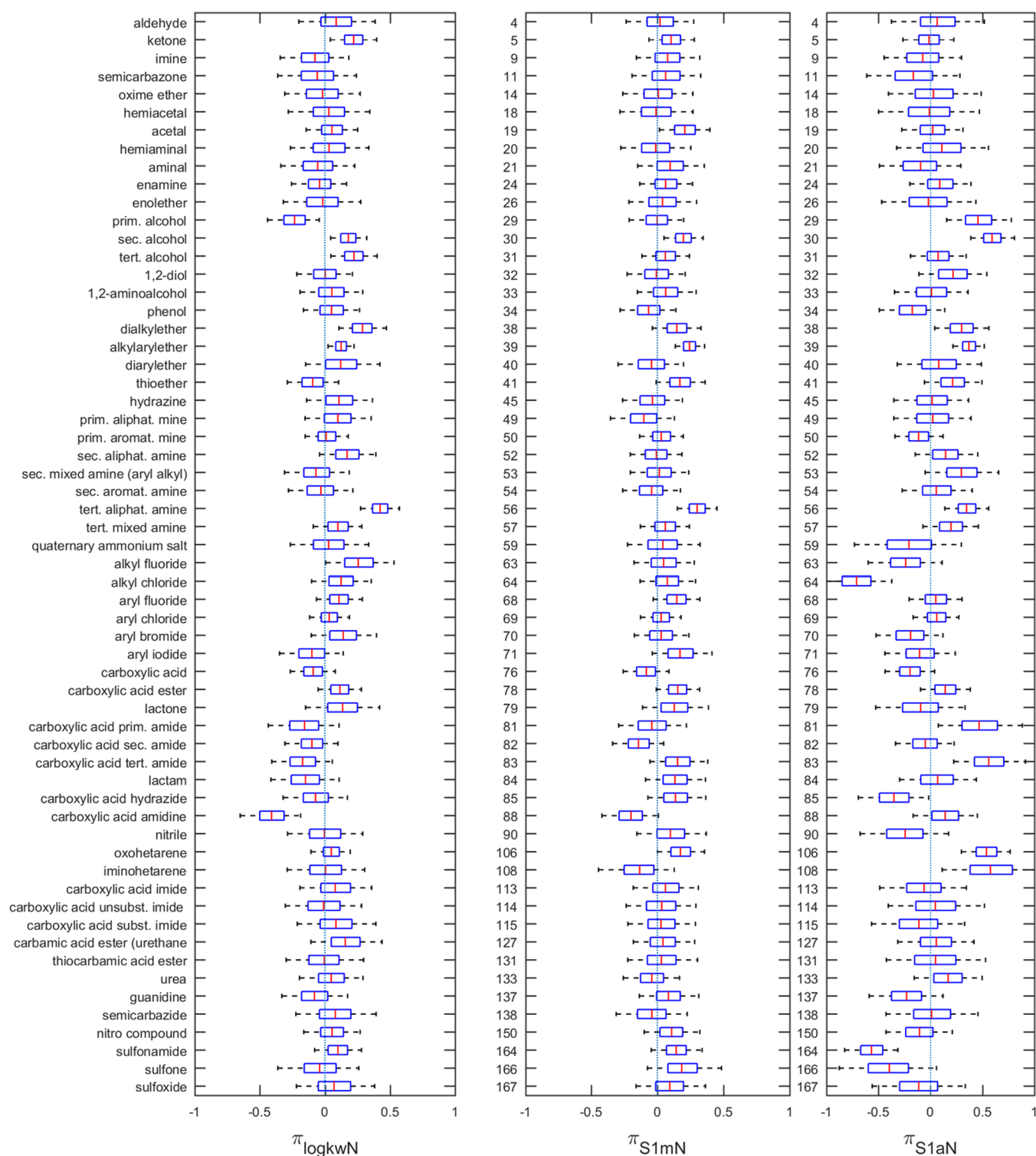
Priors for the effects of temperature on the retention factor assume a 1–3% decrease in retention factor per unit increase in temperature.<sup>31</sup> This finding corresponds to the following priors

$$\theta_{d \log kT} \sim N(-0.087, 0.022), \omega_{d \log kT} \sim N_+(0.022) \quad (31)$$

Priors for residual variability ( $\sigma_i$ ,  $\nu$ ) equal

$$\sigma_i \sim \text{lognormal}(\ln(m_\sigma), s_\sigma) \quad (32)$$

$$m_\sigma \sim N_+(0, 1), s_\sigma \sim N_+(0, 1), \nu = 3 \quad (33)$$

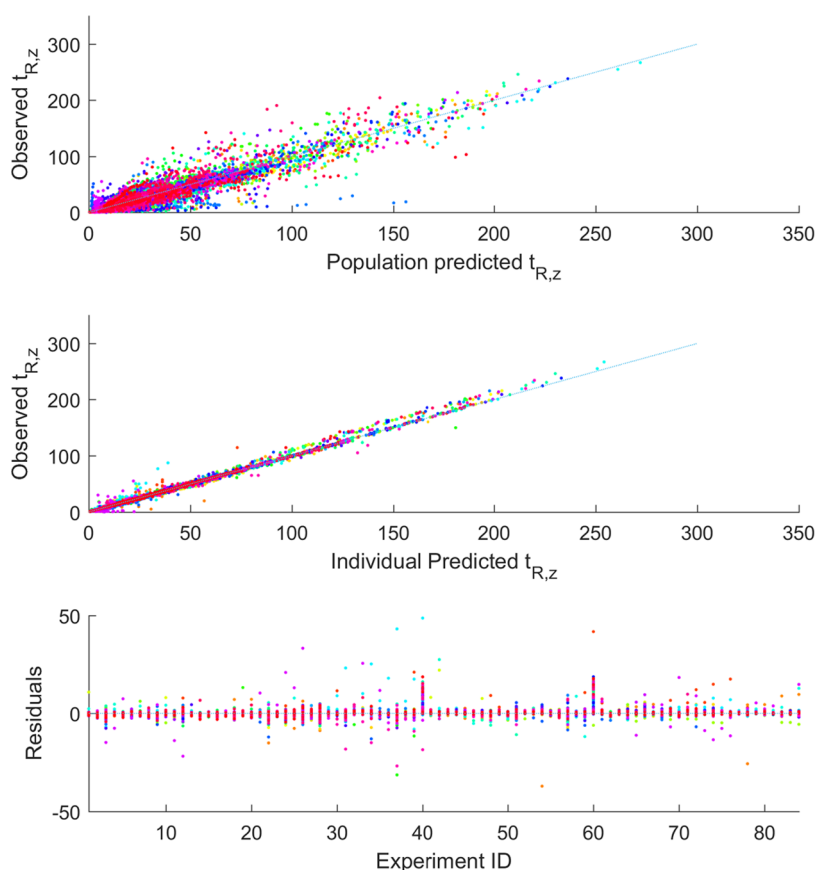


**Figure 2.** Graphical display of the marginal posterior distributions for the effects of each functional group on  $\log k_{wN}$ ,  $S1mN$ , and  $S1aN$ .

Under this model,  $\sigma_i$  is lognormally distributed with a typical value of approximately 0.5. Large between-analyte variabilities of  $\sigma_i$  values were assumed. The normality parameter was assumed to equal 3 due to the large number of outlying measurements in the data.

**Bayesian Inference. Technical.** Multilevel modeling was performed in Stan/CmdStan<sup>32</sup> software linked with MATLAB R2017a<sup>33</sup> using MATLAB-Stan 2.15.<sup>34</sup> For the inference and simulation calculations, we used the following values of the

Stan parameters: number of iterations = 1000, warmup = 1000, and number of Markov chains = 4. The reduce\_sum function, which was selected to accelerate the calculations, works by parallelizing the execution of a single Stan chain across multiple cores. Convergence diagnostics were checked using Gelman–Rubin statistics and trace plots. No divergence was reported in the model. The MATLAB code, data, and Stan code used to analyze the data are publicly available from



**Figure 3.** Goodness-of-fit plots. The observed vs the mean population-predicted retention factors (i.e., a posteriori means of predictive distributions corresponding to the future observations of a new analyte), the observed vs the mean individual-predicted retention times (i.e., a posteriori mean of a predictive distribution conditioned on the observed data from the same analyte), and the residuals vs experiment ID.

GitHub (<https://github.com/wiczling/lcms>). The raw data are also available through a repository.<sup>35</sup>

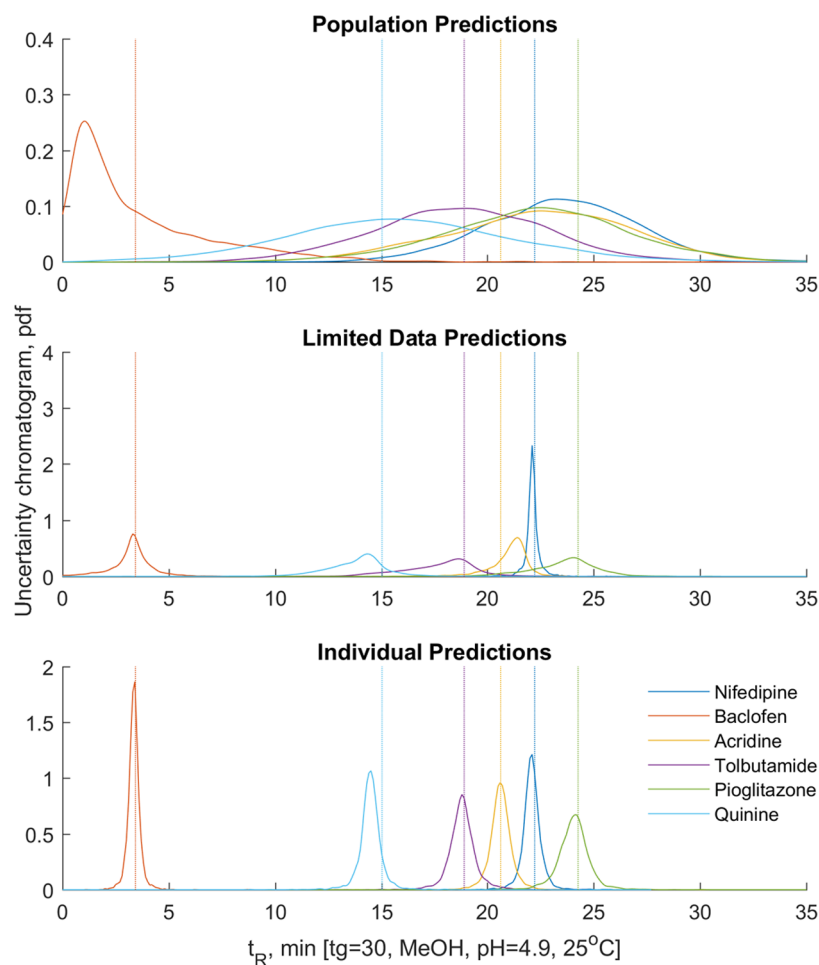
**Predictions Using a Limited Set of Experiments.** To illustrate the usefulness of the proposed model, we selected six analytes with different acidic/basic properties (within the range of considered pH values): acridine (monoprotic acid), baclofen (zwitterion: acidic and basic group), nifedipine (neutral), pioglitazone (zwitterion: basic and acidic group), quinine (diprotic: 2 basic groups), and tolbutamide (monoprotic base). The experimental data for these analyses are presented in Figure S4. Three types of predictions are shown: (i) individual predictions that correspond to future observations given access to all of the experimental data collected for these analytes (84 experiments), (ii) population predictions that correspond to future observations when no experimental data are available, and (iii) limited data predictions that correspond to future predictions given access to the limited experimental data collected for these analytes (three experiments collected for pHs of 2.5, 5.8, 10.5, and 30 min MeOH gradient at 25 °C). The predictions are summarized as uncertainty chromatograms (posterior distribution of retention times expected for a given set of chromatographed analytes under given conditions).<sup>12</sup> The uncertainty chromatogram visualizes the uncertainty for the locations of the maximum of each peak on a given chromatogram. Any area under the uncertainty chromatogram for a particular analyte can be probabilistically interpreted as a fraction of analytes (similar with respect to predictors and gathered data) that are expected

to have a retention time within the range that the area was calculated.

## RESULTS AND DISCUSSION

The available data were described using a single mechanistic model. This model was built using simple blocks/components, known fundamentals of gradient chromatography, and prior knowledge available in the literature. This approach allowed us to build a fairly realistic model using simple and interpretable parameters (such as  $\log k_w$ ,  $S$ , and  $pK_a$ ). However, we are aware that the results of this work are sensitive to the choice of priors. For this reason, we explicitly described our choice and encouraged readers to criticize it and modify it according to a different state of knowledge about the problem.

Figure 1 (Table S1) shows a summary of the marginal posterior distributions for population-level parameters compared with an assumed prior knowledge. The effects of functional groups are shown in Figure 2. These population-level parameters summarize the behavior of a typical analyte and contain the information required to predict retention for a new analyte. The developed model generated chromatographic parameters that generally show agreement with the literature knowledge and assumed priors. It is expected, as many heuristics about chromatographic parameters are available in the literature (e.g.,  $S_1$  is nearly 4 for methanol, or unit increase in the temperature decrease retention factor by 1–3%). The most surprising finding (difference between prior/posterior distribution) was observed for the  $dS1a$ ,  $S_2$ , and  $\tau$  parameters. The  $dS1a$  parameters describe the retention of anions/cations



**Figure 4.** Uncertainty chromatograms displaying the predictions for six selected analytes using different preliminary information. Each peak represents the range of analyte retention factors compatible with prior and preliminary data. Predictions were based on three experiments conducted at pH values of 2.5, 5.8, and 10.5 for a 30 min MeOH gradient at 25 °C. Colors correspond to different analytes that are identified in the bottom subplot. Vertical lines represent actual measurements.

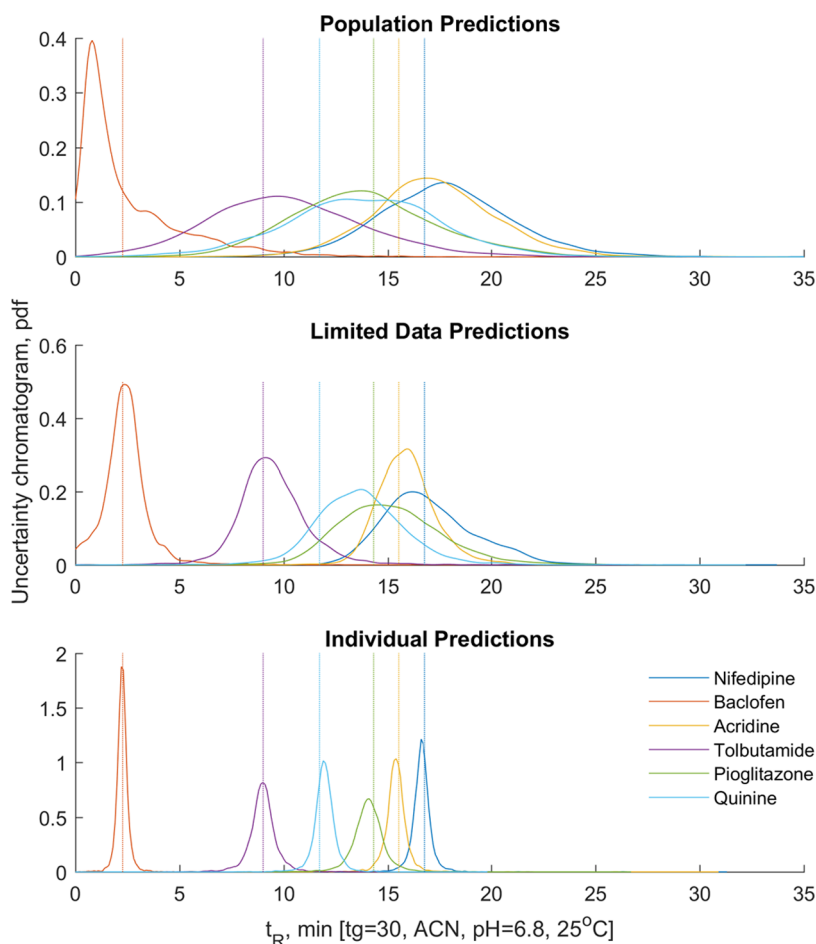
in the ACN-rich mobile phase. It was observed that  $\theta_{dS1aA} = 0.89$  (0.72 – 1.07) and that  $\theta_{dS1aB} = -0.46$  (–0.57 – –0.35). These parameters have different signs, suggesting different retention characteristics of anions and cations in ACN-rich mobile phases. This difference is less evident for MeOH, as  $\theta_{dS1mA} = 0.33$  (0.17 – 0.50) and  $\theta_{dS1mB} = 0.11$  (–0.01 – 0.23). Between-analyte variability for  $\alpha$  is large ( $\tau_m = 2.26$  (1.99 – 2.53) and  $\tau_a = 2.56$  (2.26 – 2.86)). To some degree, it might be a consequence of misidentification of certain analytes and consequential error in the value of predictors ( $pK_a$ ,  $\log P$ , charges, and groups). The  $S_2$  parameters are higher for MeOH and lower in ACN than expected. It was also evident that the stationary phase changes its properties with the pH of the mobile phase (or buffer type). In this work, this change was quantitated by determining the slope of the relationship between  $\log k_w$  and pH. The slope is negative for anions  $-0.02$  (–0.03 – –0.02) and positive for cations (0.09 (0.08 – 0.09)). This effect is likely caused by a combination of various mechanisms related to the pH of the mobile phase, such as the presence of surface silanol groups and the formation of ion pairs with buffer components.<sup>26,36</sup>

Figures S5–S7 visualize the distribution of analyte-specific (individual) chromatographic parameters and their relationship to predictors ( $\log P$  and  $pK_a$ ). There are strong correlations between individual parameters corresponding to

neutral forms of analytes: 0.78 (0.71 – 84), 0.71 (0.64 – 0.78), and 0.92 (0.88 – 0.94) for  $\log kwN-S1mN$ ,  $\log kwN-S1aN$ , and  $SmN-S1aN$  correlations, respectively. Additionally, the  $\alpha$  values for MeOH and ACN are highly correlated (the correlation is 0.94 (0.91 – 0.96)). A high correlation implies mutual information between the variables. Simply, one can gain knowledge about one parameter by knowing the value of another parameter (e.g., the knowledge about  $\log kwN_i$  narrows the range of possible  $S1mN_i$  values for a particular analyte). This finding is confirmed in practice, as one scouting gradient run (assuming MeOH or ACN content as a single design variable) usually provides much information about retention. This result is a consequence of a high correlation between  $\log kw$  and  $S_1$ , which implies that one “effective” parameter drives retention.

It would be valuable to identify chromatographic parameters that are approximately independent of the stationary phase. Such parameters (if precisely and accurately identified) can serve as a prior for data predictions for other stationary phases (chromatographic columns). The parameters meeting this criterion are  $pK_a$  and  $\alpha$  (they reflect acid–base properties of compounds in the solvent). Additionally,  $S_1$  and  $dS_1$  seem to describe the properties of the mobile phase.  $S_1$  represents the difference in retention between  $\log kw$  and  $\log km$  (or  $\log ka$ ). This parameter reflects the total free energy of transfer of the





**Figure 5.** Uncertainty chromatograms displaying the predictions for six selected analytes using different preliminary information. Each peak represents the range of analyte retention factors compatible with prior and preliminary data. Predictions were based on three experiments conducted at pH values of 2.5, 5.8, and 10.5 for a 30 min MeOH gradient at 25 °C. Colors correspond to different analytes that are identified in the bottom subplot. Vertical lines represent actual measurements.

compound from water to MeOH or ACN.  $dS1$  corresponds to a change in  $S1$  due to dissociation. In contrast, the parameters that strongly depend on stationary phase characteristics are  $\log k_{wN}$ ,  $d\log k_{wN}$ ,  $\pi_{\log k_{wN}}$ , and  $\alpha pH$ . The question remains whether they can be accurately estimated based on the experimental design used in this work.

Various interactions occurring in a chromatographic column are usually characterized by a carefully designed set of experiments using probe analytes. This process allows for a detailed physical description of the chromatographic systems.<sup>37</sup> One can also characterize these interactions using retention time data collected for a relatively large and heterogeneous group of compounds chromatographed in a broad range of conditions. Nevertheless, population and individual parameters obtained this way should be treated as “macro” parameters (parameters that describe general behaviors of analytes in the column). As pointed out by Gritti and Guiochon,<sup>37</sup> such “macro” parameters (e.g.,  $\log k$ ) mask the physical reality of various interactions in the chromatographic column, as they lump the consequences of different effects in one parameter. We agree with this statement. Nevertheless, we suggest that these parameters can be interpreted (under the assumed model) and used for predictions.

The model predictions are well calibrated with the data, as shown in Figure 3. The considerable number of outlying

measurements was handled using robust residual error. Figure 3 also summarizes the calibration and sharpness of predictions<sup>38</sup> expected after cross-validation. Specifically, individual predictions approximate leave-one-measurement-out cross-validation, and population predictions correspond to leave-one-analyte-out cross-validation. The population-level parameters are typically insensitive to the lack of a single observation (individual predictions) or all observations for a particular analyte (population predictions). In scenarios with limited preliminary data, one can expect the calibration and sharpness to fall between those observed for population predictions and those observed for individual predictions. The prediction for 6 selected analytes is illustrated in Figures S8–S10. As expected, the individual predictions are highly accurate because they are based on population-level parameters, predictors, and all of the observed retention time measurements. In contrast, the population predictions are rather imprecise, as predictions are based on population-level parameters and predictors. Access to a set of preliminary experiments reduces uncertainty in predictions. This decrease depends on the choice of experiments and analyte characteristics. The comparison of population, individual, and limited data predictions for two chromatographic conditions are shown in Figures 4 and 5. Three preliminary experiments conducted in MeOH allowed prediction of retention in MeOH

and ACN and for other pH values. In all cases, the propagated uncertainty can be applied for decision making, e.g., to determine whether achieving the desired separation is feasible. In our opinion, the quantification of uncertainty is virtually missing in method development. However, it is a crucial element for the more realistic use of chromatographic models in practice, especially in problems involving limited preliminary data. In this case, all of the decisions regarding further analytical steps are made under uncertainty.

The multilevel models can also be beneficial in supporting MS identification of unknown compounds by LC/MS analysis. The probability that a particular peak corresponds to a given analyte can be refined by adding information from the observed retention time. The Bayesian approach seems suitable for these types of problems, as for this problem, predictions need to be made under uncertainty. Since the prediction accuracy of current models (that are based on analyte structure) is rather limited, one can also expect that a limited amount of predictive information from retention time measurement will be added to the probability of correct identification/annotation. The methods using Bayesian algorithms for peak detection and compound screening in LC/MS databases are already available in the literature and can be combined with the proposed model.<sup>39,40</sup>

In this work, we proposed a mechanistic model to describe the retention data of 187 small molecules obtained for a wide range of chromatographic conditions. We are fully aware that the proposed model can be improved by adding several omitted complexities (e.g., effects of temperature or the effects of ionic strength on certain parameters). We also acknowledge the complexity and very time-consuming nature of the approach. Nevertheless, even in the current form, the model provides a step forward in the process of finding a general mechanistic model that is applicable to describe chromatographic retention of a heterogeneous group of analytes. We plan to apply the model to different columns to better understand the column variations in the model parameters. We also suggest that the model has value in improving the accuracy and precision of parameter estimation (due to prior information in subsequent analysis) and in providing the necessary input to identify better experimental designs.

## CONCLUSIONS

This work provides a pilot study that aims to demonstrate the application of a Bayesian multilevel model to describe the retention time data collected for 187 analytes for a wide range of chromatographic conditions. The analysis characterizes the chromatographic retention of neutral, acidic, and basic analytes. The model is interpretable and provides a compact summary of complex data. The model can be used to predict retention uncertainty based on various numbers of preliminary experiments, and as such, can be useful for decision making under uncertainty. The model also provides prior information for subsequent analysis.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.2c02034>.

Experimental design (Table S1); summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters (Table S2);

pH measurements (Figure S1); raw data (Figure S2); functional groups identified by Checkmol (Figure S3); raw data for a set of six analytes (Figure S4); individual parameters (Figures S5–S7); population predictions (Figure S8); limited data predictions (Figure S9); and individual predictions (Figure S10) (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Paweł Wiczling – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland; [orcid.org/0000-0002-2878-3161](https://orcid.org/0000-0002-2878-3161); Email: [wiczling@gumed.edu.pl](mailto:wiczling@gumed.edu.pl)

### Authors

Agnieszka Kamedulska – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland

Lukasz Kubik – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland

Julia Jacyna – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland

Wiktoria Struck-Lewicka – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland

Michał J. Markuszewski – Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, 80-416 Gdańsk, Poland

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.analchem.2c02034>

### Author Contributions

L.K., J.J., and W.S.L. collected the experimental data; L.K. and A.K. prepared the data for analysis; A.K. and P.W. analyzed the data; P.W. and A.K. wrote the paper with input from all authors; P.W. conceived of the presented idea, designed the study, and supervised the project; and M.M. helped supervise the project.

### Funding

This project was supported by the National Science Centre, Poland (grant 2015/18/E/ST4/00449). A.K. was also supported by the project POWR.03.02.00-00-I035/16-00 cofinanced by the European Union through the European Social Fund under the Operational Programme Knowledge Education Development 2014–2020.

### Notes

The authors declare no competing financial interest.

## REFERENCES

- (1) Gritti, F. *Anal. Chem.* **2021**, 93, 5653–5664.
- (2) Bouwmester, R.; Martens, L.; Degroove, S. *Anal. Chem.* **2019**, 91, 3694–3703.
- (3) Muteki, K.; Morgado, J. E.; Reid, G. L.; Wang, J.; Xue, G.; Riley, F. W.; Harwood, J. W.; Fortin, D. T.; Miller, I. J. *Ind. Eng. Chem. Res.* **2013**, 52, 12269–12284.
- (4) Talebi, M.; Schuster, G.; Shellie, R. A.; Szucs, R.; Haddad, P. R. *J. Chromatogr. A* **2015**, 1424, 69–76.
- (5) Golmohammadi, H.; Dashtbozorgi, Z.; Heyden, Y. V. *Chromatographia* **2014**, 78, 7–19.

- (6) Daghir-Wojtkowiak, E.; Wiczling, P.; Bocian, S.; Kubik, Ł.; Kośliński, P.; Buszewski, B.; Kaliszan, R.; Markuszewski, M. *J. Chromatogr. A* **2015**, *1403*, 54–62.
- (7) Hancock, T.; Put, R.; Coomans, D.; Heyden, Y. V.; Everingham, Y. *Chemom. Intell. Lab. Syst.* **2005**, *76*, 185–196.
- (8) Haddad, P. R.; Taraji, M.; Szűcs, R. *Anal. Chem.* **2021**, *93*, 228–256.
- (9) Liebal, U. W.; Phan, A. N. T.; Sudhakar, M.; Raman, K.; Blank, L. M. *Metabolites* **2020**, *10*, No. 243.
- (10) Briskot, T.; Stücker, F.; Wittkopp, F.; Williams, C.; Yang, J.; Konrad, S.; Doninger, K.; Griesbach, J.; Bennecke, M.; Hepbildikler, S.; Hubbuch, J. *J. Chromatogr. A* **2019**, *1587*, 101–110.
- (11) Yamamoto, Y.; Yajima, T.; Kawajiri, Y. *Chem. Eng. Res. Des.* **2021**, *175*, 223–237.
- (12) Wiczling, P.; Kamedulska, A.; Kubik, Ł. *Anal. Chem.* **2021**, *93*, 6961–6971.
- (13) Wiczling, P.; Kubik, Ł.; Kaliszan, R. *Anal. Chem.* **2015**, *87*, 7241–7249.
- (14) Wiczling, P. *Anal. Bioanal. Chem.* **2018**, *410*, 3905–3915.
- (15) He, Q.-L.; Zhao, L. *Sep. Purif. Technol.* **2020**, *246*, No. 116856.
- (16) Wiczling, P. *Sep. Sci. plus* **2018**, *1*, 63–75.
- (17) Gelman, A.; Vehtari, A.; Hill, J., *Regression and Other Stories*; Cambridge University Press: Cambridge, 2020.
- (18) Rosés, M. *J. Chromatogr. A* **2004**, *1037*, 283–298.
- (19) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. *J. Cheminf.* **2011**, *3*, No. 33.
- (20) Haider, N. *Molecules* **2010**, *15*, 5079–5092.
- (21) ACD/Labs, *Release 12.0*; Advanced Chemistry Development Inc.: Toronto, ON, Canada, 2011.
- (22) Nikitas, P.; Pappa-Louisi, A. *J. Chromatogr. A* **2009**, *1216*, 1737–1755.
- (23) Nikitas, P.; Pappa-Louisi, A. *J. Chromatogr. A* **2002**, *971*, 47–60.
- (24) Jano, I.; Hardcastle, J. E.; Zhao, K.; Vermillion-Salsbury, R. *J. Chromatogr. A* **1997**, *762*, 63–72.
- (25) Téllez, A.; Rosés, M.; Bosch, E. *Anal. Chem.* **2009**, *81*, 9135–9145.
- (26) Neue, U. D.; Phoebe, C. H.; Tran, K.; Cheng, Y. F.; Lu, Z. *J. Chromatogr. A* **2001**, *925*, 49–67.
- (27) Pappa-Louisi, A.; Nikitas, P.; Balkatzopoulou, P.; Malliakas, C. *J. Chromatogr. A* **2004**, *1033*, 29–41.
- (28) Cox, B. G. *Org. Process Res. Dev.* **2015**, *19*, 1800–1808.
- (29) Lewandowski, D.; Kurowicka, D.; Joe, H. *J. Multivar. Anal.* **2009**, *100*, 1989–2001.
- (30) Martin, S. R. *Informative Priors for Correlation Matrices: An Easy Approach*, 2021, <http://srmart.in/informative-priors-for-correlation-matrices-an-easy-approach/>.
- (31) Snyder, L. R.; Kirkland, J. J.; Glajch, J. L., *Practical HPLC Method Development*; Wiley: New York, 1997.
- (32) Stan Development Team, *Stan Modeling Language Users Guide and Reference Manual*, VERSION. <https://mc-stan.org/mc-stan.org/users/citations/> (accessed Jan 18, 2022).
- (33) MATLAB, 9.2.0.556344 (R2017a); The MathWorks Inc.: Natick, MA, 2018.
- (34) MatlabStan, <https://github.com/brian-lau/MatlabStan> (accessed Jan 24, 2022).
- (35) Kubik, Ł.; Jacyna, J.; Struck-Lewicka, W.; Markuszewski, M.; Wiczling, P. *LC-TOF-MS Data Collected for 300 Small Molecules XBridge-C18 Column*. 2022.
- (36) Méndez, A.; Bosch, E.; Rosés, M.; Neue, U. D. *J. Chromatogr. A* **2003**, *986*, 33–44.
- (37) Gritti, F.; Guiochon, G. *Anal. Chem.* **2005**, *77*, 4257–4272.
- (38) Gneiting, T.; Balabdaoui, F.; Raftery, A. E. *J. R. Stat. Soc.: Ser. B (Stat. Methodol.)* **2007**, *69*, 243–268.
- (39) Woldegebriel, M.; Gonsalves, J.; van Asten, A.; Vivó-Truyols, G. *Anal. Chem.* **2016**, *88*, 2421–2430.
- (40) Woldegebriel, M.; Vivó-Truyols, G. *Anal. Chem.* **2015**, *87*, 7345–7355.

## Recommended by ACS

### Introducing Multifactorial Peak Crossover in Analytical and Preparative Chromatography via Computer-Assisted Modeling

Imad A. Haidar Ahmad, Erik L. Regalado, *et al.*

JULY 30, 2020

ANALYTICAL CHEMISTRY

READ 

### Achiral–Chiral Two-Dimensional Liquid Chromatography Platform to Support Automated High-Throughput Experimentation in the Field of Drug Development

Alexandre Goyon, Kelly Zhang, *et al.*

NOVEMBER 03, 2020

ANALYTICAL CHEMISTRY

READ 

### Comparison of Cosine, Modified Cosine, and Neutral Loss Based Spectrum Alignment For Discovery of Structurally Related Molecules

Wout Bittremieux, Pieter C. Dorrestein, *et al.*

AUGUST 12, 2022

JOURNAL OF THE AMERICAN SOCIETY FOR MASS SPECTROMETRY

READ 

### Microscale Purification with Direct Charged Aerosol Detector Quantitation Using Selective Online One- or Two-Dimensional Liquid Chromatography

Kanaka Hettiarachchi, Thomas J. Greshock, *et al.*

JUNE 03, 2022

ANALYTICAL CHEMISTRY

READ 

Get More Suggestions >