Supporting Information

Comparison of chromatographic stationary phases using Bayesian-based multilevel modeling

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

Department of Biopharmaceutics and Pharmacodynamics, Medical University of Gdańsk, Gen. J. Hallera 107, 80-416 Gdańsk, Poland

Table of content:

Table S1. Experimental design

Table S2. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters

Figure S1. pH measurements.

Figure S2. Raw data. Lines connect measurements obtained for a particular analyte.

Figure S3. Functional groups identified by Checkmol. Figures show the number of analytes having at least one functional group of a given type.

Figure S4. Raw data for a set of 6 analytes. .

Figure S5. Individual parameters (neutral form of analyte)

Figure S6. Individual parameters (dissociated forms of analytes)

Figure S7. Individual parameters (pKas)

Figure S8. Population Predictions for a set of 6 analytes.

Figure S9. Limited Data Predictions for a set of 6 analytes.

Figure S10. Individual Predictions for a set of 6 analytes.

Table S1. Experimental design.

|  |  |  |  |
| --- | --- | --- | --- |
| **Gradient time, min** | **Mobile phase pH** | **Temp. [°C]** | **Org. mod.** |
| 270 | 2.5 5.8 10.5 | 25 | MeOH |
| 90 | 2.5-10.5 | 25 | MeOH |
| 30 | 2.5-10.5 | 25 | MeOH |
| 270 | 2.5 5.8 10.5 | 25 | ACN |
| 90 | 2.5-10.5 | 25 | ACN |
| 30 | 2.5-10.5 | 25 | ACN |
| 270 | 2.5 5.8 10.5 | 35 | MeOH |
| 90 | 2.5-10.5 | 35 | MeOH |
| 30 | 2.5-10.5 | 35 | MeOH |
| 270 | 2.5 5.8 10.5 | 35 | ACN |
| 90 | 2.5-10.5 | 35 | ACN |
| 30 | 2.5-10.5 | 35 | ACN |

Table S2. Summary of the MCMC simulations of the marginal posterior distributions of population-level model parameters.

Mean denotes sample mean, MCSE denotes Monte Carlo Standard Error, StdDev denotes sample standard deviation, 5%, 50%, 95% denote corresponding quantiles, N\_Eff denotes effective sample size, R\_Hat denotes a measure of chain equilibrium, must be within 0.05 of 1.0.



Figure S1. pH measurements. Lines provide model predictions.

Figure S2. Raw data. Lines connect measurements obtained for a particular analyte.

Figure S3. Functional groups identified by Checkmol. Figures show the number of analytes having at least one functional group of a given type.

Figure S4. Raw data for 6 selected analytes.

Figure S4. Raw data for 6 selected analytes. Continued

Figure S4. Raw data for 6 selected analytes. Continued

Figure S4. Raw data for 6 selected analytes. Continued

Figure S4. Raw data for 6 selected analytes. Continued

Figure S4. Raw data for 6 selected analytes. Continued

Figure S5. Scatter plots between individual chromatographic parameters and their relationship to log P. Diagonal subplots present histograms.

Figure S6. Scatter plots between individual chromatographic parameters describing the effect of dissociation (*dlogkw*, *dS1m*, *dS1a* were normalized by the absolute charge). Diagonal subplots present histograms. Colors corresponds to different charge state of analyte form (red – anions, blue – cations, yellow – zwitterions).

Figure S7. Scatter plots between individual chromatographic parameters (pKas) and their relationship to literature pKa. Diagonal subplots present histograms. Red color denotes acidic group, blue color basic group.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters. Continued.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters. Continued.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters. Continued.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters. Continued.

Figure S8. Population predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given only population-level parameters. Continued.

Figure S9. Limited Data Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient.

Figure S9. Limited Data Predictions. Figures S6. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient. Continued

Figure S9. Limited Data Predictions. Figures S6. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient. Continued

Figure S9. Limited Data Predictions. Figures S6. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient. Continued

Figure S9. Limited Data Predictions. Figures S6. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient. Continued

Figure S9. Limited Data Predictions. Figures S6. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to the future observations given three preliminary experiments conducted in MeOH, at 25oC for pH = 2.5, 5.8, and 10.5 and 30 min gradient. Continued

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.

Figure S10. Individual Predictions. Predictions represented as posterior median (line) and 5th-95th percentiles (dotted lines) for a 6 exemplary analytes. Observed retention factors are shown as dots. Predictions corresponding to future observations given the population-level parameters and all the retention data measured for a particular analyte.