Analyst

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Check for updates

Cite this: *Analyst*, 2025, **150**, 3567

Do experimental projection methods outcompete retention time prediction models in non-target screening? A case study on LC/HRMS interlaboratory comparison data†

Louise Malm [a] and Anneli Kruve *[a,b]

Retention time (RT) is essential in evaluating the likelihood of candidate structures in nontarget screening (NTS) with liquid chromatography high resolution mass spectrometry (LC/HRMS). Approaches for estimating the RTs of candidate structures can broadly be divided into projection and prediction methods. The first approach takes advantage of public databases of RTs measured on similar chromatographic systems (CS_{source}) and projects these to the chromatographic system applied in the NTS (CS_{NTS}) based on a small set of commonly analyzed chemicals. The second approach leverages machine learning (ML) model(s) trained on publicly available retention time data measured on one or more chromatographic systems ($CS_{training}$). Nevertheless, the CS_{source} and $CS_{training}$ might differ substantially from CS_{NTS}. Therefore, it is of interest to evaluate the generalizability of projection models and prediction models in CSs routinely applied in NTS. Here we take advantage of the recent NORMAN interlaboratory comparison where 41 known calibration chemicals and 45 suspects were analyzed to evaluate both the projection and prediction approaches on 37 CSs. The accuracy of both approaches was directly linked to the similarity of the CS, and the pH of the mobile phase and the column chemistry were found to be most impactful. Furthermore, for cases where CS_{source} and CS_{NTS} differ substantially but $CS_{training}$ and CS_{NTS} are similar, prediction models often performed on par with the projection models. These findings highlight the need to account for the mobile phase and column chemistry in ML model training and select the prediction model for RT.

Received 19th March 2025,
Accepted 1st July 2025

DOI: 10.1039/d5an00323g

rsc.li/analyst

1. Introduction

Nontarget screening (NTS) with liquid chromatography high resolution mass spectrometry (LC/HRMS) is widely used for screening of chemicals in various samples. For example, NTS is used to pinpoint chemicals of concern in environmental and food samples and also to identify yet unknown metabolites or pharmaceutically active natural products.[1] The aim of these methods is to both detect and identify suspected or yet unknown chemicals and the methods used to identify the detected chemicals leverage tandem mass spectra (MS^2).[2] Albeit information rich, MS^2 spectra rarely indicate only one structure, and a combination of different strategies is needed to reduce the list of likely candidate structures.[3,4] Here empirical analytical information and metadata about the sample can prove useful in prioritizing the correct candidate structure.

Regarding empirical analytical information, retention time (RT),[5–8] collision cross-section (CCS),[9–11] ionization mode, and adduct type[12,13] have proven useful for discriminating candidate structures. Nevertheless, we[14] and others[4,15] have previously observed that RT has high potential in distinguishing between the candidate structures in terms of physical separation during analysis as well as evaluating candidate structures with a prediction model. This arises from the orthogonality of the chromatographic separation and HRMS data and indicates the high importance of incorporating RT in candidate structure prioritization workflows in NTS.

Comparing the RT of the detected LC/HRMS feature and the candidate structure can aid in prioritization by indicating less likely candidate structures, *e.g.*, the candidate structures for which the predicted RT differs from the experimentally observed RT more than the uncertainty limits. The RT of the candidate structures can be obtained by (1) projection of experimental reference retention times and (2) prediction of retention times based on the structure with machine learning (ML). The projection methods leverage RT databases[16,17] of

[a]*Department of Materials and Environmental Chemistry, Stockholm University, 11418 Stockholm, Sweden. E-mail: anneli.kruve@su.se*
[b]*Department of Environmental Science, Stockholm University, 11418 Stockholm, Sweden*

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d5an00323g

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

known structures measured on a different chromatographic system (CS_{source})[5] and is applicable to chemicals already experimentally studied. Nevertheless, the CS_{source} and chromatographic system used in NTS (CS_{NTS}) may differ due to equipment (dead volume, flow rate, and column length) as well as separation mechanisms (column chemistry and mobile phase). A small set of ten to 50 chemicals measured on both CSs can be used to fit a generalized additive model (GAM) or similar model between RTs from the source and target CS, which can account for some of these differences. Later, the same model is applied to project the RT of the chemicals measured with the CS_{source} to the CS_{NTS}.

In the initialization and validation of a well-known projection approach, PredRet,[5] it was observed that the number of chemicals commonly analyzed by both CSs and the similarity of the retention mechanism (reversed phase *vs.* HILIC) impact the projection accuracy. Alternatively, retention time indices (RTIs)[6,18,19] have been suggested to account for the effect of column length, flow rate, and dead volume on RT. Furthermore, retention time order (RTO) is known to be more stable than RT or RTI if the same physical parameters of the CS are changed. Nevertheless, RTO is likely to change for nominally equal CSs for close eluting/co-eluting chemicals due to the variations in the peak shape or with different batches of LC columns[20] as well as changing mobile phase composition when acids or bases are analysed.[21] A recent comparison of RTOs of datasets collected into RepoRT[16] revealed the largest variations for CSs with different columns and mobile phases. In spite of these considerations, an extensive unbiased overview of the applicability of projection methods across different CSs and laboratories is lacking.

The need for experimentally determined RTs can be overcome by ML models[6,8,9,22–32] predicting RT, RTI, or RTO. Despite obvious advantages, the prediction approach is presumed to underperform projection approaches. Such approaches require a relatively large and representative[33] dataset for model training and may require a close match of the CS used for collecting training data ($CS_{training}$) and the CS_{NTS} or conversion of RT to RTIs with a predetermined calibration mix.[6] To overcome this limitation, a complementing prediction approach with projection[5] was suggested by Bouwmeester *et al.*[34] and similar approaches have been successfully used thereafter.[24,35] These approaches project the predicted RT or RTI to CS_{NTS} based on predicted RTs of experimentally measured standards even if commonly measured chemicals are lacking. Nevertheless, a recent comparison by Kwon *et al.*[36] suggested a large performance variability and generally low performance of the combination of ML and projection methods. Though the exact reason remained unclear, it can be hypothesised that chemical space overlap and CS similarity affect the applicability. Similar to the study by Kwon *et al.*[36] benchmarking different RT, RTI, or RTO prediction methods on public datasets has been widely discussed;[28] however, the contribution of variability in chemical space and CSs remains hard to evaluate. In part, this is due to the sparse overlap in the chemical space considered in different studies leading to a set of data where both the chemical space and CSs often vary simultaneously.

Here we leverage the recent NORMAN interlaboratory comparison[37] to evaluate the performance of the RTI projection and prediction approaches across different CSs commonly used in NTS for environmental analysis, while keeping the chemical space constant. In particular, we are interested in answering the following questions: (1) how different are the CSs in terms of agreement in RT, RTI, and RTO; (2) how do CSs affect the accuracy of the projection methods and how do this confine the best practices in NTS; and (3) how accurate are the prediction models in comparison with projection models. To answer these questions, firstly, we used the calibration chemicals to establish a GAM for the projection of RTIs from the CS_{source} to the CS_{NTS} and evaluate these on the suspected chemicals. Secondly, we trained an ML model to predict the retention time from the structure of the suspected chemicals and evaluated its performance based on the root mean square errors (RMSEs). Lastly, we compared the projection and prediction approaches and contextualized the findings based on the similarity of the CS_{source} and $CS_{training}$ to the CS_{NTS}.

2. Materials and methods

2.1 Data

The retention time data for 38 CSs were available from a recent NTS interlaboratory comparison with the NORMAN network.[37] Two sets of chemicals, 41 calibrants and 45 suspects were analyzed with each CS and the number of detected chemicals ranged from 12 to 39 and 17 to 40, respectively. The experimental details for all CSs were published previously by Malm *et al.*[37] In summary, the shortest chromatographic program was 15 min long, leading to RTs of 0.80 min to 9.80 min and the longest program was 54 min, leading to RTs of 11.70 to 36.95 min. All but one CS leveraged reversed phase chromatography (C_{18}, C_8, and C_6-phenyl and biphenyl) with columns from all major vendors; all CSs used an acidic water phase containing only an acid or an acid with the respective ammonium salt as the mobile phase additive.

2.2 Generalized additive model for retention time index projection

To account for the fact that the set of detected calibration chemicals differed from one chromatographic system (CS) to another, the RTIs were calculated for each combination of CSs, so that only the calibration chemicals detected with both CSs were used. In order to allow a straightforward comparison of the performance of the projection and prediction models, independent of the length of the chromatographic program, column dimensions, *etc.*, all retention times were converted to retention time indices as described previously by Aalizadeh *et al.*:[6]

$$\mathrm{RTI} = \frac{\mathrm{RT}_x - \mathrm{RT}_{\min}}{\mathrm{RT}_{\max} - \mathrm{RT}_{\min}} \times 1000, \quad (1)$$

where $\mathrm{RT}_{\min}$ and $\mathrm{RT}_{\max}$ are the minimum and maximum retention times observed for the calibrants and RT_x is the

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

retention time of the chemical of interest. This assures that the RTIs of calibrants for each comparison ranged from 0 to 1000. Scaling RT values between 1 and 1000 has been previously suggested to "have large RTI units between compounds that elute differently and compare the error more realistically" and "define the elution segment for the calibrants".[6]

RTIs enabled accounting for differences in the flow rate and column length and thereby unifying the range of RTI values; however, the RTIs were sensitive to nonlinear gradient programs, that is, plateaus and changing gradient speeds. To account for such variations in the CSs, the RTI projection with a generalized additive model (GAM) was implemented. That is, the *gam* function from the *mgcv* package in R with smooth function "s" was used and the dimensions k for the smooth term was set to 6 for all projections. The GAMs were fitted on the calibration chemicals between the CS_{NTS} and CS_{source}, and applied to predict the RTIs in the CS_{NTS} for the suspects based on the "known" RTIs in the CS_{source}. Only suspects eluting in the calibration range were considered (RTI range 0 to 1000). Notably, the GAM minimizes the residual in the direction of the y-axis (like most models), leading to unsymmetrical projection accuracies and therefore all-to-all CS combinations were considered.

2.3 Retention time prediction models

A retention time prediction model was trained leveraging the RT data of environmentally relevant chemicals from Aalizadeh *et al.*[6] measured with a 15.5 min long chromatographic program using a 5 mM ammonium acetate based buffer as the water phase and methanol as an organic modifier. Prior to modelling, the calibrants and suspects were removed from the dataset to avoid overfitting to these chemicals. For all remaining chemicals in both datasets the 2D and 3D Mordred descriptors were calculated from the mol representation after standardization with RDKit. As part of data cleaning, all descriptors yielding missing values for more than 15 chemicals were discarded, followed by removing all chemicals with any left missing values. Thereafter, the dataset contained retention times for 1795 chemicals and 1328 Mordred descriptors. Prior to modelling, the descriptors with near zero variance and descriptors with high correlation coefficients were removed with *nearZeroVar* (default settings) and *findCorrelated* (cutoff 0.7) functions from the *caret* package in R (version 4.3.2). An *xgbTree* model was trained to predict the RT. The hyperparameters were optimized with five times repeated two-fold cross-validation implemented through the *caret* package in R.

2.4 Statistical analysis

The similarity of RTIs across CSs was analyzed with hierarchical clustering combined with a heatmap. The performance of both the projection and prediction models was evaluated based on the RMSEs of the predicted RTI *vs.* experimental RTI. The statistical comparison of the RMSE values was performed using the F-test ($p < 0.05$, no correction for multiple comparisons was done). Data, code, and additional visualizations are available in the ESI.†

3. Results and discussion

3.1 Comparability of the RT, RTI, and RTO across CSs

The same calibration and suspect compound mix was analyzed with all CSs; however, the number of detected calibrants ranged from 12 to 38 and suspects from 17 to 40. The earliest eluting suspect had RT between 0.8 and 7.3 min, depending on the CS. Simultaneously, the RT of the last eluting suspect varied between 8.1 min and 32.4 min. These differences arise from the substantially different lengths of gradient programs used (15 to 54 min), including segments of isocratic elution implemented at different parts of the elution program.

The similarity in the retention time profiles of CSs was evaluated based on the correlation of the RTs with Pearson correlation coefficients and RTOs with Spearman correlation coefficients. Generally, the agreement in RTs for two CSs (Pearson correlation) followed the same order as that based on RTOs (Spearman correlation), with a vast majority of the CS pairs showing $R^2 > 0.8$. This is expected as both the correlation of RT and RTO values have their advantages and disadvantages. RTOs are less affected by segments of isocratic elution, while RTs are less affected by small differences in the retention time for close eluting or overlapping chemicals. Due to the good agreement of the Pearson and Spearman correlation coefficients (ESI 1†), only the RTO agreement (Spearman correlation) is discussed below.

The Spearman correlation coefficient of RTO values ranged from 0.59 to 0.999 (ESI 1†). Independent of the type of correlation coefficient, the highest similarity was observed for the two pairs of CSs where both CSs were implemented in the same lab, by making use of either a different LC/HRMS instrument (DS_QDF and DS_VQL) or the HRMS mode only (DS_GSB and DS_TDF). The latter is expected to yield different RTs only due to the random variability of the analysis and data processing.

The highest correlation coefficients for CSs implemented in different laboratories (DS_HT and DS_QBD) exceeded a Spearman ρ value of 0.999. These CSs used the same LC column (Agilent ZORBAX Eclipse Plus C18) and mobile phase (water and acetonitrile with 0.1% formic acid). Nevertheless, the CSs differed in LC instrumentation, gradient program, column temperature, injection volume, *etc*. This indicates that the retention order may be well preserved, given the same column chemistry and mobile phase, even if other parameters vary.

The highest Spearman ranked correlation coefficient for CSs (DS_BQW and DS_MT) using different columns from different vendors yielded only a marginally lower Spearman ρ value of 0.998; however, both CSs made use of a very similar mobile phase containing ammonium formate in both mobile phase components. This indicates a good agreement in RTOs despite having different CSs. However, some pairs of CSs leveraging the same retention mechanism yielded Spearman ρ values below 0.9.

Further analysis with hierarchical clustering revealed several clusters of CSs with similar elution profiles that could

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

be associated with the CS parameters. Firstly, one of the CSs showed substantially different RTOs in comparison with all other CSs. A closer investigation revealed that this was the only CS that combined RP and HILIC columns for separation (DS_DX), and the significant differences in the RTI order clearly indicate the orthogonality of this system. Due to the significant differences in the retention mechanism of RP and HILIC, only RP based CSs will be considered further in evaluating the projection or prediction method as both approaches assume the same retention mechanism.

Among the CS using RP-LC, a cluster of five CSs, namely DS_JL, DS_QJS, DS_AWW, DS_JBQ, and DS_JSG, stands out. Furthermore, RTOs from DS_QQT showed a high correlation with the RTOs from the above CSs (Fig. 1). All these CSs made use of ammonium salt as a mobile phase additive either alone or with formic acid. As the majority of the CSs used in this interlaboratory comparison made use of 0.1% formic acid as a water phase additive, the CSs in this cluster use a higher pH of the water phase compared to the majority of the CSs. Comparison of the RTOs from these CSs with RTOs from CSs using formic acid as a mobile phase additive revealed that higher RTOs were exclusively observed for chemicals with basic functional groups (pK_a values between 3.6 and 8.1) with ammonium formate as a mobile phase additive. Similarly, lower retention time order values were observed for acidic chemicals, suggesting the importance of the acid–base equilibrium of the analytes in the RTO. In addition to the above-mentioned CSs, three additional systems leveraged ammonium salts as mobile phase additives, namely DS_STF, DS_GJT and DS_QSB. The RTOs from DS_STF and DS_GJT showed lower similarity to all other CSs, possibly due to the use of a different stationary phase. DS_STF used a biphenyl column, while DS_GJT used a C_8 column. On the other hand, DS_QSB clustered together with CSs facilitating formic acid as a water phase additive.

In two of the clusters of CSs (clusters of DS_JWW, DS_AW, DS_JDF, DS_QDF, DS_VQL, DS_QQG, DS_ QV; and cluster of DS_BIJ, DS_TL, DS_QSB, DS_HT, DS_QBD, DS_DBS, and DS_TSF) leveraged exclusively 0.1% formic acid as the water phase additive. Several CSs in the first of these clusters used Agilent Eclipse Plus or Poroshell columns, while in the second Phenomenex Kinetex EVO columns were used in several CSs. This is likely to indicate that in addition to the mobile phase the stationary phase chemistry has an impact on the agreement of the RTOs even if nominally all CSs considered here applied RP separation.

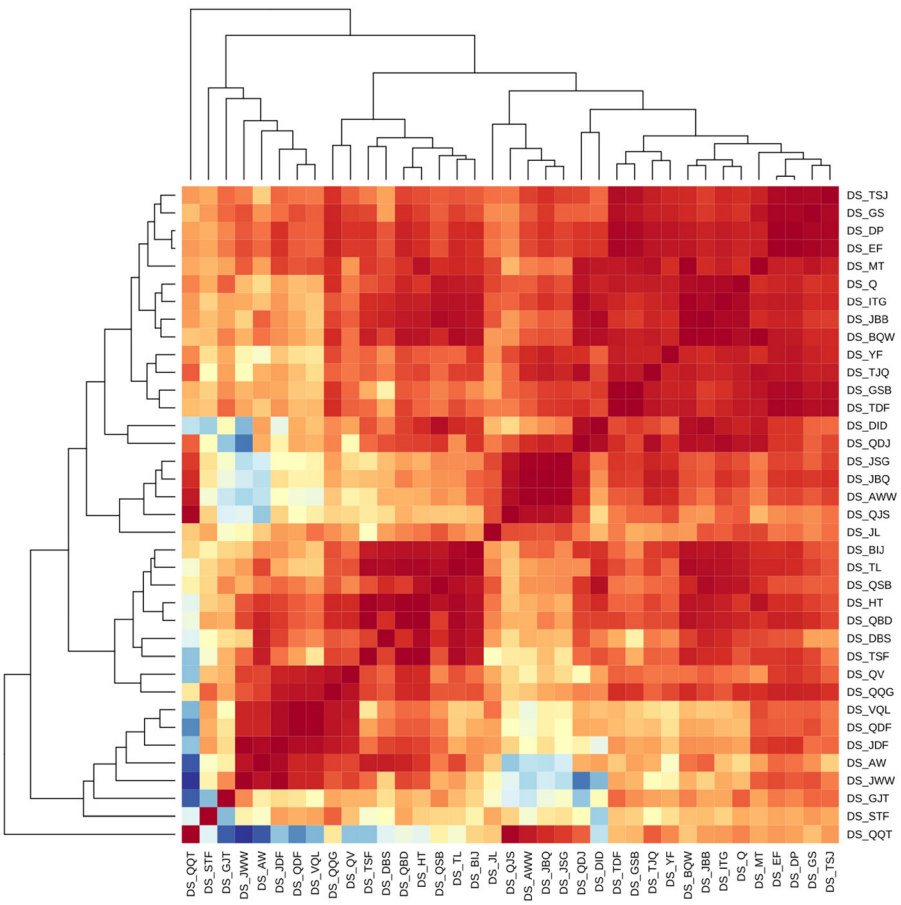

Fig. 1 The heatmap of the CSs based on the Spearman correlation coefficient of the RTOs. Red indicates high and blue indicates low Spearman correlation coefficients.

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Importantly, the similarity in RTOs is impacted by the chemicals commonly detected by the two CSs. For example, if one of the LC/HRMS methods has not detected some of the acidic or basic compounds that are likely to change RTOs, such an effect will be left undetected here. Nevertheless, the cluster analysis demonstrates that the RTOs largely depend on the similarity of the CSs and the impact on the performance of projection and prediction methods needs close scrutiny.

3.2 Peak spacing

Ideal chromatographic systems would allow the separation of all components in the mixture. In NTS, this is virtually impossible due to the high complexity of the environmental samples as well as the limited time available for analyses. For a limited and known number of chemicals, ideal peak spacing would follow a uniform distribution where chemicals elute across the whole elution program and peaks are equally spaced. Under such conditions the peak overlap as well as the resulting matrix effect would be minimized. In the case of NTS, the peak spacing also affects the purity of MS^2 spectra, especially for data independent acquisition. Last but not least, the distribution can affect RT based prioritization, as candidate structures predicted to elute in regions of high peak density are harder to distinguish from each other.

In order to evaluate the peak spacing of CSs used here, RTs were normalized relative to the length of the elution program. The peak spacing was analyzed based on the standard deviation and cumulative distribution of the normalized RTs. It was observed that in some CSs the chemicals eluted almost equally spaced across the whole elution program (Fig. 2, green line). The highest standard deviation of the normalized and non-normalized RTs was observed for CS DS_TSJ, which made use of a relatively long elution program and the duration of the gradient segment exceeded 50% of the total elution program (run time).

Simultaneously, with some CSs a narrow range of normalized RTs was observed. For example, for CS DS_QV (Fig. 2, red line) most of the chemicals yielded RTI values between 200 and 600. This CS has long isocratic elution segments at the beginning and end of the elution program while the gradient segment is relatively short in comparison with the length of the elution program (<40%). The absolute peak spacing based on the distribution of the normalized RTs showed that CS DS_QV yielded one of the lowest RT standard deviations. The distribution of the normalized RTs is therefore also affected by the polarity of the chemicals included in the study. For example, it is fair to believe that incorporating more (very) low and high polarity chemicals would benefit this CS due to the isocratic elution segments. Nevertheless, the chemicals used in this study are chosen as representatives of the chemical contaminants in environmental water samples.

Furthermore, for CS DS_TSF (Fig. 2, yellow line) and many others, the majority of the normalized RTs remained below 400, hinting at a very tight spacing of the peaks. Simultaneously, for CS DS_MT the detailed analysis revealed that the narrow range of normalized RTs is also associated with the low number of detected chemicals. Namely, most of the late eluting chemicals incorporated in this study were undetected with the CS DS_MT, which impacted the analysis of the normalized RTs. The same was true for other CSs yielding the lowest maximum normalized RTs. Nevertheless, no direct association between the standard deviation of the normalized RTs and the number of detected chemicals was observed.

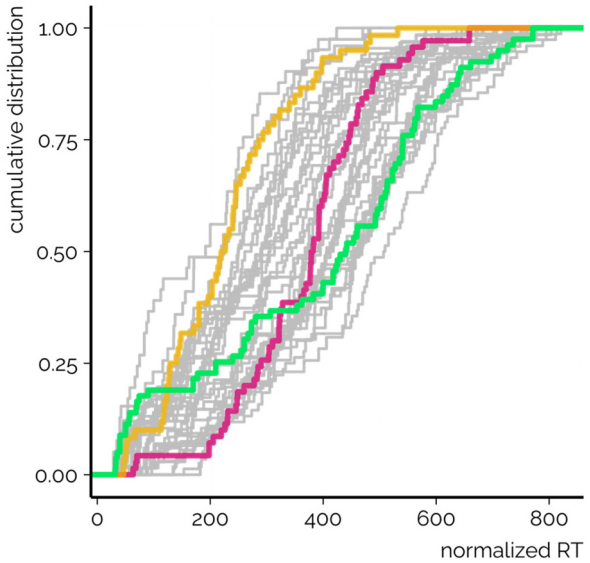

Fig. 2 The cumulative distribution of the normalized RTs for all CSs. In green, the CS with the highest standard deviation of the normalized RTs is shown. In contrast, two CSs yielding relatively narrower peak spacing are highlighted in yellow and red. Both calibrants and suspects are considered.

As a result, the analysis of peak spacing, retention time distribution and the peak detectability suggests that longer elution programs with proportionally larger segments of linear gradients provide a more uniform peak distribution. Importantly, this coincides with the insights from Anderson *et al.*,[38] who found that longer gradient programs increase the number of chemical features detected with LC/HRMS based NTS. It can be simultaneously hypothesized that such elution programs also improve the identification of the detected chemicals based on the RTs due to the expected narrower relative uncertainty.

3.3 Projection of RTI

The RMSE of the RTIs of the suspects with the projection approach was 78.1 RTI units over all labs, that is, 7.8% of the total RTI range for the calibrants (Table S1, ESI 2, ESI/results†). The mean absolute deviation (MAD) was 46.5 RTI units and 95% of the absolute deviations were less than 175.2 RTI units. The projection approach facilitating the GAM clearly improved the transferability of the RTIs across CSs. The RMSE of the RTIs of the suspects prior to the GAM was 150.4 RTI units across all combinations of the CSs and in 92.3% of the CS combinations a lower RMSE point estimate was observed after applying the GAM. In 57.7% of the cases the improve-

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

ment was statistically significant. In 7.7% of the CS combinations (n = 102) the RMSE point estimate was higher after projection; however, in only four of the cases the difference in RMSE was statistically significant according to the F-test. These findings highlight the need for an RT/RTI projection approach even if commonly analyzed calibrants are used for obtaining the RTIs.

The necessity for the GAM for RTI projection results from the fact that RTI calculation, as proposed earlier,[6] makes use of the first and last eluting calibration chemicals and is unaffected by the rest of the calibrants and their RT profiles. Thus, the RTI values are a linear projection of the RTs and overlook any nonlinearity arising from the different gradient profiles, including isocratic segments, used by different CSs. For example, the two CSs exemplified in Fig. 3 used BEH C18 columns and methanol as an organic modifier; however, they differed in the flow rate and organic modifier gradient as well as the position and length of the isocratic elution segment. As a result, the retention order of the chemicals and Spearman correlation of the RTIs from the two CSs are high. Nevertheless, the RTI values are nonlinearly associated, leading to an RMSE of 257.2 RTI units without the GAM. Here the GAM enables an effective projection of the RTIs from one system to another, and the RMSE drops to 55.0 RTI units.

We furthermore observed that the projection of the GAM was very sensitive to the overlap in the calibration chemicals detected by the two CSs. Most of all, the projections were unreliable for suspects outside of the calibration range (lower or higher RTs of suspects than those of calibrants), and all such suspects were omitted in the above analysis. As a result, the projection results are unavailable for 9.8% of the detected suspects across all combinations of CSs. In this sense, it is important to ensure that the calibrants elute across the eluent program.

Lastly, we were also interested in the projection accuracy of the calibrants that can be used to assess the prediction accuracy of the suspects. We therefore compared the RMSE values of the calibrants and suspects for each combination of the CSs. Ideally, the RMSE values would be statistically insignificantly different. Here we observed a weak correlation of the RMSE values of the calibrants and suspects (Pearson's squared correlation coefficient of 0.30). In 63 cases, the RMSE of the suspects was statistically significantly lower than that of the calibrants with a maximum difference of 4.7×. The RMSE was statistically significantly higher for suspects with 535 of the 1332 combinations of CSs and the maximum relative increase in the RMSE was over two orders of magnitude. As a consequence, the RMSE values of the calibrants are insufficient in evaluating the projection accuracy for the suspects for a large fraction of the combinations of CSs. As a result, we suggest evaluating the projection accuracy with a separate set of quality control chemicals that are representative of the chemicals expected to be detected in the samples but are not used for fitting the GAM. This will allow for an independent RMSE evaluation, which can be further used as the uncertainty of the projection approach. For example, given the chemical similarity between the quality controlled chemicals and suspected chemicals, ~95% of the "true" RTs of the suspected chemicals are expected to be within ±2 RMSE from the RT obtained with the GAM.

3.4 Prediction models of RTI

In the case of NTS of complex mixtures, the RTs or RTIs of suspects may be unavailable in public libraries due to, *e.g.*, the lack of analytical standards. This has led to the prediction of the RT(I)s with different ML models.[6,32,39] The models are often evaluated on chemicals measured on the same CS, which therefore leads to potentially optimistic performance evaluations. The

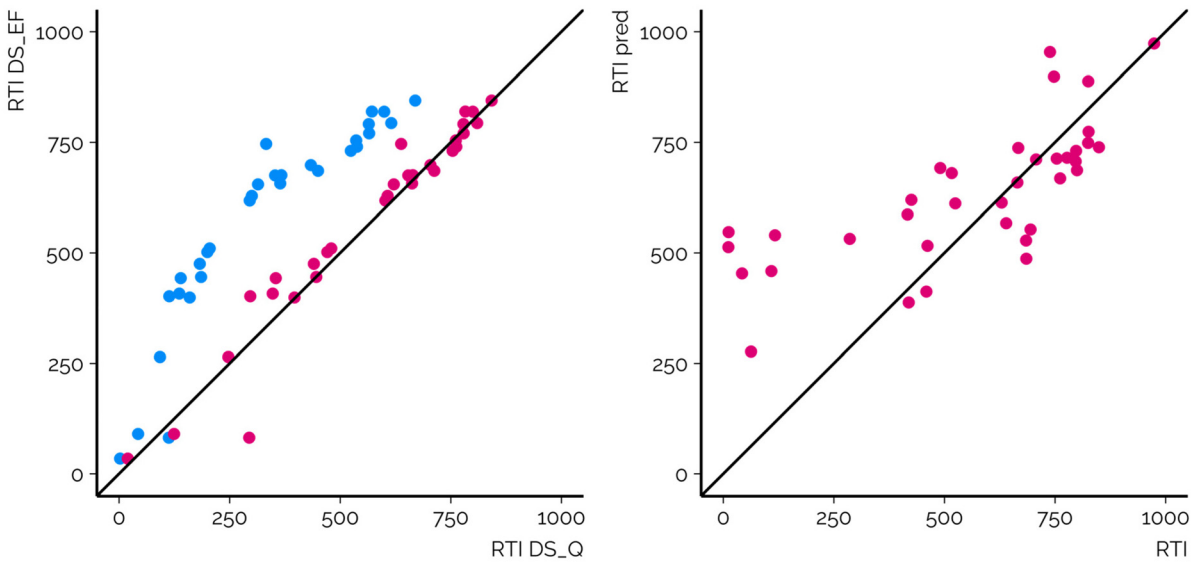

Fig. 3 The comparison of (left) the projection of RTI values for suspects analyzed with two different CSs before (blue, RMSE of 317.4 RTI units) and after (red, RMSE of 53.4 RTI units) application of the GAM (DS_Q → DS_EF) and (right) prediction of the RTI values with the ML model for DS_EF (RMSE of 167.7 RTI units).

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

dataset from the interlaboratory comparison provides a great opportunity to evaluate the consistency of the model performance on CSs used in NTS. For this reason, we trained an RT prediction model based on the cleaned and standardized RT data previously published by Aalizadeh *et al.*[6] The retraining of the model was needed, to provide unbiased evaluation so that the training set did not contain any of the calibration chemicals or suspects. The calibration chemicals were used to project the predicted RTs to the RTI scale of the respective CSs.[6]

The RMSE of suspects across all CSs was 149.0 RTI units, while the overall MAD was 150.3 RTI units and 95% of the chemicals yielded RTI values below 289.8 units across all CSs (Table S1†). The R^2 values showed the same trends as RMSE values; generally CSs with lower RMSE values showed higher R^2 values of RTI predicted and RTI measured. The lowest R^2 value was 0.12 (DS_GJT) and the highest value was 0.82 (DS_AWW). The performance of the prediction models varied starkly depending on the CS (Fig. 4, ESI 3, ESI/results/NORMAN_RTI_prediction_results_summary.csv†): the RMSE values for the individual CSs ranged from 86.8 to 219.7 RTI units. It must be recognized that the RMSE values of suspects are influenced by (1) the suitability of the training data for the CSs and chemical space in question; (2) the accuracy of the model; and (3) the accuracy of the projection of the predicted RTs to the RTIs of the CS. The RMSE values of the calibrants are primarily impacted by the former two factors and less by the projection accuracy as the experimental RTIs are directly used for fitting the projection model.

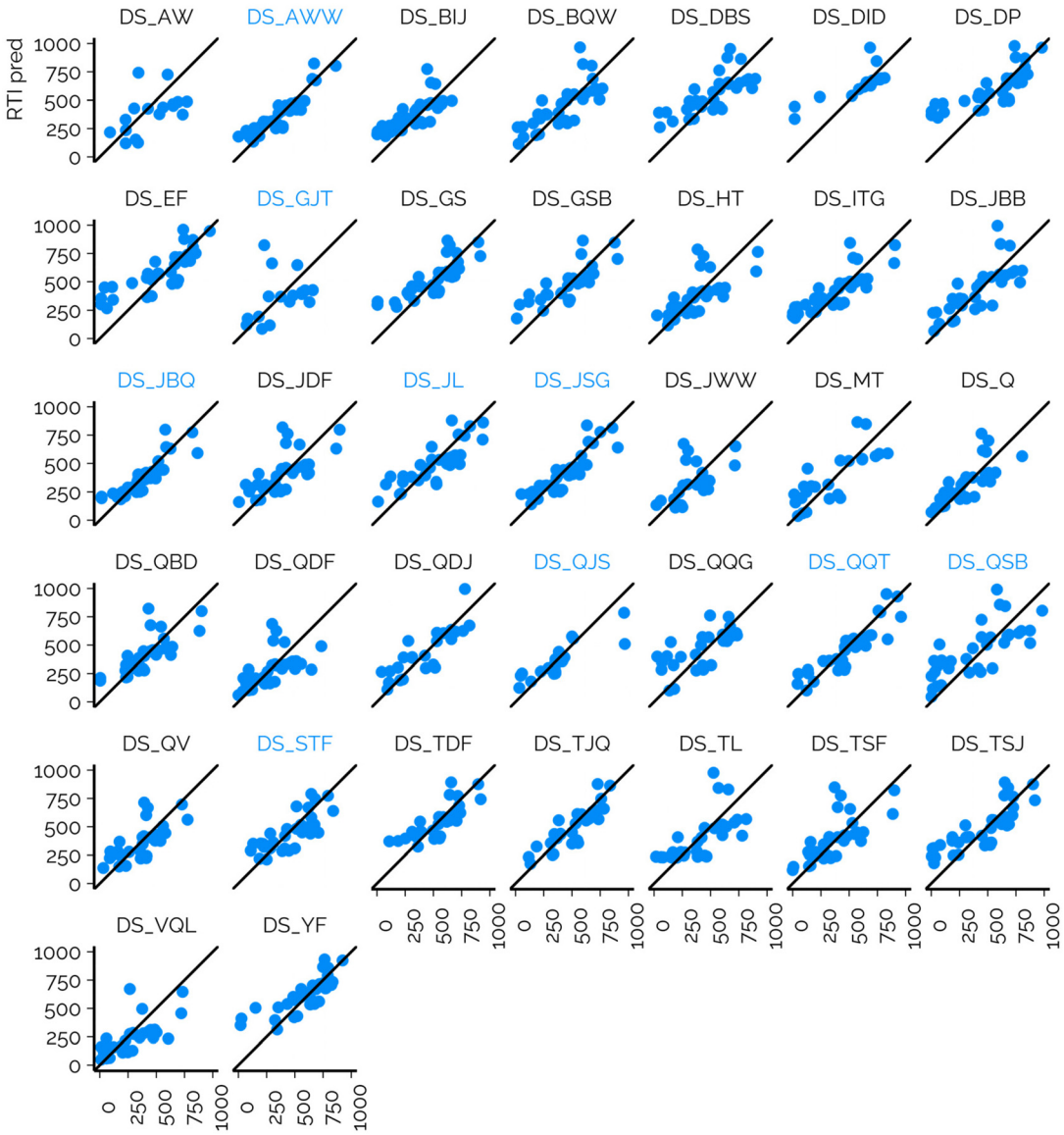

Fig. 4 The comparison of the predicted RTI values with ML *vs.* the experimentally observed RTI values of suspects. DS_AWW, DS_GJT, DS_JBQ, DS_JL, DS_JSG, DS_QJS, DS_QQT, DS_QSB, and DS_STF (marked in blue) all leveraged the ammonium-based buffer in the mobile phase, while the other leveraged the mobile phase with formic acid as an additive.

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM. This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

We observed that the performance in terms of RMSE was directly correlated with the similarity of the CS_{NTS} (CS used for measuring the suspects) and $CS_{training}$ (used for collecting the training data). All four CS_{NTS} on which the trained models showed the lowest RMSEs for the suspects used an ammonium formate based mobile phase. Similarly, the $CS_{training}$ used a mobile phase containing ammonium formate; however, different columns and mobile phase flow rates were used. Furthermore, for these CSs relatively low RMSEs of the calibrants were observed, also indicating a successful projection of the predicted RTs to experimental RTIs. This is in accordance with the RTO correlations observed above, where a good agreement of the RTOs for ammonium salt containing mobile phases was observed independent of the column. As a result, it is expected that predicted RTIs can be successfully used on CSs using similar mobile phases.

Generally, a weak agreement in RMSE values of the calibrants and suspects was observed, where CSs yielding relatively low values for one set also yielded low values for the other and *vice versa* (Pearson's squared correlation coefficient of 0.47). Nevertheless, in a handful of cases the RMSE values of the calibrants were low, while relatively high RMSE values were observed for suspects. No concrete reasons for the substantially higher RMSEs of suspects could be pinpointed either in terms of chemicals detected or the settings of CSs. Nevertheless, the high RMSEs of the calibrants were usually indicative of high RMSEs of the suspects and could be considered in the interpretation of the results.

Recent findings by Kwon *et al.*[36] demonstrated that RT models directly trained on a small dataset might outperform ML models combined with projection models; however, the reasons remained unclear. Inspired by these findings, we trained separate RTI prediction models for each CS using calibration chemicals only (ESI 4†). The RMSE values ranged from 131.5 to 351.3 RTI units and the performance of the models clearly correlated with the number of training instances, where the correlation coefficient between the RMSE and number of training datapoints was −0.63. The model trained on the calibration data from the same CS never outperformed the ML model trained on a larger dataset followed by projection. On the other hand, the ML models trained on external large datasets outperformed the models trained on the calibrant data from the same CS in 11 cases (F-test, $p < 0.05$). In the remaining cases, the performance of the two models was statistically insignificantly different. This indicates that given sufficient overlap in the chemical space, ML models trained on larger datasets followed by the projection of the predicted RTIs may be preferred to in-house trained models if only limited training data are available.

All in all, the application of ML predicted retention times in the data from 37 CSs indicates that the reliability of the predictions is closely linked to the similarity in the CSs used for collecting the training data and the CS used in NTS. This highlights the need for using CSs in NTS that are similar to community standards if the applicability of ML models for structural elucidation or other tasks is desired.

3.5 Comparison of the projection and prediction models

It can be hypothesized that prediction models combined with projection models are less accurate than the projection of experimental RT/RTI values, due to error propagation. Indeed, the RMSE values for the prediction model (larger dataset followed by projection, sections 2.3 and 3.4) are mostly higher than those observed for the projection model (sections 2.2 and 3.3); however, specific combinations of CSs may yield higher RMSE values for the projection method than those for the prediction method (ESI 5†). In total 348 (26.1%) of 1332 combinations of CSs yielded statistically insignificantly different RMSE values with projection and prediction methods, considering only chemicals for which the RTI could be predicted with both. For 32 (2.4%) of the combinations of CSs, the prediction models outperformed projection models, and for the remaining 71.5% of the cases, the projection models yielded statistically significantly lower RMSE values. The comparison is impacted by the accuracy of both the prediction and projection methods.

The prediction models outcompeted the projection models in seven CSs (DS_AWW, DS_DID, DS_JBQ, DS_JSG, DS_QJS, DS_QQT, and DS_YF). In these cases, with the exception of two CSs (DS_DID and DS_YF), the CS_{NTS} mobile phase is similar to the mobile phase of $CS_{training}$ (ammonium salt based buffer solution); therefore, the ML model for RTI prediction has been trained on data with a chromatographically similar mechanism. These five CSs also showed the lowest (top6) RMSE values for the prediction approach.

Furthermore, the comparison of projection and prediction methods is also impacted by the CS_{source} used in the projection approach. Here, most cases where the prediction method outcompeted the projection method are characterized by mismatching the mobile phase of the CS_{source} and CS_{NTS} and therefore high $RMSE_{projection}$ values of the suspects (between 103.7 and 207.7 RTI units). Nevertheless, for the CS_{source} and CS_{NTS} using a similar mobile phase, the projection method outperformed the prediction approach even if the $CS_{training}$ closely matched the CS used in NTS.

This indicates that in the case where the CS used in NTS closely matches the one with experimentally available RTI values, the projection method should be preferred. However, if experimental data with similar CSs are lacking, and the ML model used in the prediction approach is trained on data from a similar CS, the prediction approach could yield similar or higher prediction accuracy than the projection approach from a less similar CS. Moreover, the prediction approach is applicable to all chemicals within the scope of the ML, even the ones that lack previously measured chromatographic data.

Conclusions

A recent interlaboratory study using the NORMAN network has provided a unique opportunity to evaluate the comparability of RT, RTI, and RTOs across 37 CSs used in laboratories in NTS with RP LC/HRMS on a routine basis. Due to an overlapping

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

set of calibration chemicals and suspects analyzed with all CSs, the impact of CSs could be directly evaluated. Expectedly, the similarity of CSs was found to significantly impact the RTIs and RTOs, where the most influential factors were the additive type and/or pH of the water phase. Additionally, the clustering of RTOs from similar column chemistry was observed.

Furthermore, two commonly applied approaches for evaluating RT(I)s for candidate structures, the projection and prediction approaches, could be compared. As expected from the RTO correlations, the performance of both approaches was affected by the similarity in the mobile phase composition. Thus, projection methods generally outperformed prediction models; however, prediction models were found to occasionally outperform the projection method if the prediction model was trained on similar CSs. Furthermore, the prediction models trained on larger external data should be preferred over small in-house data in spite of mismatching CSs. As a result, we suggest using projection methods when the CS_{source} closely matches with the CS_{NTS} and in other cases, prediction models trained on large representative datasets should be preferred due to their accuracy and application scope.

Author contributions

L. M.: data curation, investigation, validation, and writing – review & editing. A. K.: conceptualization, formal analysis, funding acquisition, writing – original draft, and writing – review & editing.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the ESI.†

Acknowledgements

Open access funding was provided by Stockholm University. The study has been kindly supported by VR grant no. 2021-03917 and has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement No. 101124488, LearningStructurE: Machine Learning and Mass Spectrometry for Structural Elucidation of Novel Toxic Chemicals. We are thankful for the NORMAN Network, JPA 2020 "CWG-NTS-CT NTS semi-quantification", for organizing the interlaboratory comparison and participants in the original study, namely, Jaanus Liigand, Reza Aalizadeh, Nikiforos Alygizakis, Kelsey Ng, Emil Egede Frøkjær, Mulatu Yohannes Nanusha, Martin Hansen, Merle Plassmann, Stefan Bieber, Thomas Letzel, Lydia Balest, Pier Paolo Abis, Michele Mazzetti, Barbara Kasprzyk-Hordern, Nicola Ceolotto, Sangeeta Kumari, Stephan Hann, Sven Kochmann, Teresa Steininger-Mairinger, Coralie Soulier, Giuseppe Mascolo, Sapia Murgolo, Manuel Garcia-Vara, Miren López de Alda, Juliane Hollender, Katarzyna Arturi, Gianluca Coppola, Massimo Peruzzo, Hanna Joerss, Carla van der Neut-Marchand, Eelco N Pieke, Pablo Gago-Ferrero, Ruben Gil-Solsona, Viktória Licul-Kucera, Claudio Roscioli, Sara Valsecchi, Austeja Luckute, Jan H Christensen, Selina Tisler, Dennis Vughs, Nienke Meekel, Begoña Talavera Andújar, Dagny Aurich, Emma L Schymanski, Gianfranco Frigerio, André Macherius, Uwe Kunkel, Tobias Bader, Pawel Rostkowski, Hans Gundersen, Belinda Valdecanas, W Clay Davis, Bastian Schulze, Sarit Kaserzon, Martijn Pijnappels, Mar Esperanza, Aurélie Fildier, Emmanuelle Vulliet, Laure Wiest, Adrian Covaci, Alicia Macan Schönleben, Lidia Belova, Alberto Celma, Lubertus Bijlsma, Emilie Caupos, Emmanuelle Mebold, Julien Le Roux, Eugenie Troia, Eva de Rijke, Rick Helmus, Gaëla Leroy, Niels Haelewyck, David Chrastina, Milan Verwoert, and Nikolaos S Thomaidis.

References

1 J. Hollender, E. L. Schymanski, L. Ahrens, N. Alygizakis, F. Béen, L. Bijlsma, A. M. Brunner, A. Celma, A. Fildier, Q. Fu, P. Gago-Ferrero, R. Gil-Solsona, P. Haglund, M. Hansen, S. Kaserzon, A. Kruve, M. Lamoree, C. Margoum, J. Meijer, S. Merel, C. Rauert, P. Rostkowski, S. Samanipour, B. Schulze, T. Schulze, R. R. Singh, J. Slobodnik, T. Steininger-Mairinger, N. S. Thomaidis, A. Togola, K. Vorkamp, E. Vulliet, L. Zhu and M. Krauss, NORMAN Guidance on Suspect and Non-Target Screening in Environmental Monitoring, *Environ. Sci. Eur.*, 2023, **35**(1), 75, DOI: 10.1186/s12302-023-00779-4.

2 E. L. Schymanski, J. Jeon, R. Gulde, K. Fenner, M. Ruff, H. P. Singer and J. Hollender, Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence, *Environ. Sci. Technol.*, 2014, **48**(4), 2097–2098, DOI: 10.1021/es5002105.

3 H. Hupatz, I. Rahu, W.-C. Wang, P. Peets, E. H. Palm and A. Kruve, Critical Review on in Silico Methods for Structural Annotation of Chemicals Detected with LC/HRMS Non-Targeted Screening, *Anal. Bioanal. Chem.*, 2024, **417**, 473–493, DOI: 10.1007/s00216-024-05471-x.

4 T. O. Metz, C. H. Chang, V. Gautam, A. Anjum, S. Tian, F. Wang, S. M. Colby, J. R. Nunez, M. R. Blumer, A. S. Edison, O. Fiehn, D. P. Jones, S. Li, E. T. Morgan, G. J. Patti, D. H. Ross, M. R. Shapiro, A. J. Williams and D. S. Wishart, Introducing "Identification Probability" for Automated and Transferable Assessment of Metabolite Identification Confidence in Metabolomics and Related Studies, *Anal. Chem.*, 2025, **97**(1), 1–11, DOI: 10.1021/acs.analchem.4c04060.

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

5 J. Stanstrup, S. Neumann and U. Vrhovšek, PredRet: Prediction of Retention Time by Direct Mapping between Multiple Chromatographic Systems, *Anal. Chem.*, 2015, **87**(18), 9421–9428, DOI: 10.1021/acs.analchem.5b02287.

6 R. Aalizadeh, N. A. Alygizakis, E. L. Schymanski, M. Krauss, T. Schulze, M. Ibáñez, A. D. McEachran, A. Chao, A. J. Williams, P. Gago-Ferrero, A. Covaci, C. Moschet, T. M. Young, J. Hollender, J. Slobodnik and N. S. Thomaidis, Development and Application of Liquid Chromatographic Retention Time Indices in HRMS-Based Suspect and Nontarget Screening, *Anal. Chem.*, 2021, **93**(33), 11601–11611, DOI: 10.1021/acs.analchem.1c02348.

7 F. Kretschmer, E.-M. Harrieder, M. Witting and S. Böcker, Times Are Changing but Order Matters: Transferable Prediction of Small Molecule Liquid Chromatography Retention Times, December 23, 2024. DOI: 10.26434/chemrxiv-2024-wd5j8.

8 X. Domingo-Almenara, C. Guijas, E. Billings, J. R. Montenegro-Burke, W. Uritboonthai, A. E. Aisporna, E. Chen, H. P. Benton and G. Siuzdak, The METLIN Small Molecule Dataset for Machine Learning-Based Retention Time Prediction, *Nat. Commun.*, 2019, **10**(1), 5811, DOI: 10.1038/s41467-019-13680-7.

9 A. Celma, R. Bade, J. V. Sancho, F. Hernandez, M. Humphries and L. Bijlsma, Prediction of Retention Time and Collision Cross Section (CCS_{H+}, CCS_{H-}, and CCS_{Na+}) of Emerging Contaminants Using Multiple Adaptive Regression Splines, *J. Chem. Inf. Model.*, 2022, **62**(22), 5425–5434, DOI: 10.1021/acs.jcim.2c00847.

10 Z. Zhou, J. Tu, X. Xiong, X. Shen and Z.-J. Zhu, LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision To Support Ion Mobility–Mass Spectrometry-Based Lipidomics, *Anal. Chem.*, 2017, **89**(17), 9559–9566, DOI: 10.1021/acs.analchem.7b02625.

11 D. H. Ross, J. H. Cho and L. Xu, Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections, *Anal. Chem.*, 2020, **92**(6), 4548–4557, DOI: 10.1021/acs.analchem.9b05772.

12 C. D. Broeckling, A. Ganna, M. Layer, K. Brown, B. Sutton, E. Ingelsson, G. Peers and J. E. Prenni, Enabling Efficient and Confident Annotation of LC-MS Metabolomics Data through MS1 Spectrum and Time Prediction, *Anal. Chem.*, 2016, **88**(18), 9226–9234, DOI: 10.1021/acs.analchem.6b02479.

13 R. Costalunga, S. Tshepelevitsh, H. Sepman, M. Kull and A. Kruve, Sodium Adduct Formation with Graph-Based Machine Learning Can Aid Structural Elucidation in Non-Targeted LC/ESI/HRMS, *Anal. Chim. Acta*, 2021, 339402, DOI: 10.1016/j.aca.2021.339402.

14 M. Akhlaqi, W.-C. Wang, C. Möckel and A. Kruve, Complementary Methods for Structural Assignment of Isomeric Candidate Structures in Non-Target Liquid Chromatography Ion Mobility High-Resolution Mass Spectrometric Analysis, *Anal. Bioanal. Chem.*, 2023, **415**(21), 5247–5259, DOI: 10.1007/s00216-023-04852-y.

15 L. Bijlsma, M. H. G. Berntssen and S. Merel, A Refined Nontarget Workflow for the Investigation of Metabolites through the Prioritization by in Silico Prediction Tools, *Anal. Chem.*, 2019, **91**(9), 6321–6328, DOI: 10.1021/acs.analchem.9b01218.

16 F. Kretschmer, E.-M. Harrieder, M. A. Hoffmann, S. Böcker and M. Witting, RepoRT: A Comprehensive Repository for Small Molecule Retention Times, *Nat. Methods*, 2024, **21**, 153–155, DOI: 10.1038/s41592-023-02143-z.

17 Y. Zhang, F. Liu, X. Q. Li, Y. Gao, K. C. Li and Q. H. Zhang, Retention Time Dataset for Heterogeneous Molecules in Reversed–Phase Liquid Chromatography, *Sci. Data*, 2024, **11**(1), 946, DOI: 10.1038/s41597-024-03780-5.

18 J. Boelrijk, D. Van Herwerden, B. Ensing, P. Forré and S. Samanipour, Predicting RP-LC Retention Indices of Structurally Unknown Chemicals from Mass Spectrometry Data, *J. Cheminf.*, 2023, **15**(1), 28, DOI: 10.1186/s13321-023-00699-8.

19 R. Stoffel, M. A. Quilliam, N. Hardt, A. Fridstrom and M. Witting, N-Alkylpyridinium Sulfonates for Retention Time Indexing in Reversed-Phase-Liquid Chromatography-Mass Spectrometry–Based Metabolomics, *Anal. Bioanal. Chem.*, 2022, **414**(25), 7387–7398, DOI: 10.1007/s00216-021-03828-0.

20 P. Yang, T. McCabe and M. Pursch, Practical Comparison of LC Columns Packed with Different Superficially Porous Particles for the Separation of Small Molecules and Medium Size Natural Products, *J. Sep. Sci.*, 2011, **34**(21), 2975–2982, DOI: 10.1002/jssc.201100530.

21 M. Rosés, Determination of the pH of Binary Mobile Phases for Reversed-Phase Liquid Chromatography, *J. Chromatogr. A*, 2004, **1037**(1–2), 283–298, DOI: 10.1016/j.chroma.2003.12.063.

22 H. L. Ngan, V. Turkina, D. Van Herwerden, H. Yan, Z. Cai and S. Samanipour, Machine Learning for Enhanced Identification in RPLC/HRMS Non-Targeted Workflows, January 29, 2025. DOI: 10.26434/chemrxiv-2024-mdl4q-v2.

23 C. B. Mollerup, M. Mardal, P. W. Dalsgaard, K. Linnet and L. P. Barron, Prediction of Collision Cross Section and Retention Time for Broad Scope Screening in Gradient Reversed-Phase Liquid Chromatography-Ion Mobility-High Resolution Accurate Mass Spectrometry, *J. Chromatogr. A*, 2018, **1542**, 82–88, DOI: 10.1016/j.chroma.2018.02.025.

24 Y. Zhang, F. Liu, X. Q. Li, Y. Gao, K. C. Li and Q. H. Zhang, Generic and Accurate Prediction of Retention Times in Liquid Chromatography by Post–Projection Calibration, *Commun. Chem.*, 2024, **7**(1), 54, DOI: 10.1038/s42004-024-01135-0.

25 A. Kajtazi, M. Kajtazi, M. F. Santos Barbetta, E. Bandini, H. Eghbali and F. Lynen, Prediction of Retention Indices in LC-HRMS for Enhanced Structural Identification of Organic Micropollutants in Water: Selectivity-Based Filtration, *Anal. Chem.*, 2025, **97**(1), 65–74, DOI: 10.1021/acs.analchem.4c01784.

26 Y. Wen, M. Talebi, R. I. J. Amos, R. Szucs, J. W. Dolan, C. A. Pohl and P. R. Haddad, Retention Prediction in Reversed Phase High Performance Liquid Chromatography Using Quantitative Structure-Retention Relationships

 This journal is © The Royal Society of Chemistry 2025

Open Access Article. Published on 08 July 2025. Downloaded on 1/22/2026 7:50:35 AM.
This article is licensed under a Creative Commons Attribution 3.0 Unported Licence.

Applied to the Hydrophobic Subtraction Model, *J. Chromatogr. A*, 2018, **1541**, 1–11, DOI: 10.1016/j.chroma.2018.01.053.

27 S. C. Rutan, T. Kempen, T. Dahlseid, Z. Kruger, B. Pirok, J. G. Shackman, Y. Zhou, Q. Wang and D. R. Stoll, Improved Hydrophobic Subtraction Model of Reversed-Phase Liquid Chromatography Selectivity Based on a Large Dataset with a Focus on Isomer Selectivity, *J. Chromatogr. A*, 2024, **1731**, 465127, DOI: 10.1016/j.chroma.2024.465127.

28 Y. Liu, A. C. Yoshizawa, Y. Ling and S. Okuda, Insights into Predicting Small Molecule Retention Times in Liquid Chromatography Using Deep Learning, *J. Cheminf.*, 2024, **16**(1), 113, DOI: 10.1186/s13321-024-00905-1.

29 E. Bach, E. L. Schymanski and J. Rousu, Joint Structural Annotation of Small Molecules Using Liquid Chromatography Retention Order and Tandem Mass Spectrometry Data, *Nat. Mach. Intell.*, 2022, **4**(12), 1224–1237, DOI: 10.1038/s42256-022-00577-2.

30 Q. Yang, H. Ji, X. Fan, Z. Zhang and H. Lu, Retention Time Prediction in Hydrophilic Interaction Liquid Chromatography with Graph Neural Network and Transfer Learning, *J. Chromatogr. A*, 2021, **1656**, 462536, DOI: 10.1016/j.chroma.2021.462536.

31 R. Ju, X. Liu, F. Zheng, X. Lu, G. Xu and X. Lin, Deep Neural Network Pretrained by Weighted Autoencoders and Transfer Learning for Retention Time Prediction of Small Molecules, *Anal. Chem.*, 2021, **93**(47), 15651–15658, DOI: 10.1021/acs.analchem.1c03250.

32 P. Bonini, T. Kind, H. Tsugawa, D. K. Barupal and O. Fiehn, Retip: Retention Time Prediction for Compound Annotation in Untargeted Metabolomics, *Anal. Chem.*, 2020, **92**(11), 7515–7522, DOI: 10.1021/acs.analchem.9b05765.

33 R. Szucs, R. Brown, C. Brunelli, J. Hradski and M. Masár, Impact of Structural Similarity on the Accuracy of Retention Time Prediction, *J. Chromatogr. A*, 2023, **1707**, 464317, DOI: 10.1016/j.chroma.2023.464317.

34 R. Bouwmeester, L. Martens and S. Degroeve, Generalized Calibration Across Liquid Chromatography Setups for Generic Prediction of Small-Molecule Retention Times, *Anal. Chem.*, 2020, **92**(9), 6571–6578, DOI: 10.1021/acs.analchem.0c00233.

35 A. Souihi, M. P. Mohai, E. Palm, L. Malm and A. Kruve, MultiConditionRT: Predicting Liquid Chromatography Retention Time for Emerging Contaminants for a Wide Range of Eluent Compositions and Stationary Phases, *J. Chromatogr. A*, 2022, 462867, DOI: 10.1016/j.chroma.2022.462867.

36 Y. Kwon, H. Kwon, J. Han, M. Kang, J.-Y. Kim, D. Shin, Y.-S. Choi and S. Kang, Retention Time Prediction through Learning from a Small Training Data Set with a Pretrained Graph Neural Network, *Anal. Chem.*, 2023, **95**(47), 17273–17283, DOI: 10.1021/acs.analchem.3c03177.

37 L. Malm, J. Liigand, R. Aalizadeh, N. Alygizakis, K. Ng, E. E. Frøkjær, M. Y. Nanusha, M. Hansen, M. Plassmann, S. Bieber, T. Letzel, L. Balest, P. P. Abis, M. Mazzetti, B. Kasprzyk-Hordern, N. Ceolotto, S. Kumari, S. Hann, S. Kochmann, T. Steininger-Mairinger, C. Soulier, G. Mascolo, S. Murgolo, M. Garcia-Vara, M. López De Alda, J. Hollender, K. Arturi, G. Coppola, M. Peruzzo, H. Joerss, C. Van Der Neut-Marchand, E. N. Pieke, P. Gago-Ferrero, R. Gil-Solsona, V. Licul-Kucera, C. Roscioli, S. Valsecchi, A. Luckute, J. H. Christensen, S. Tisler, D. Vughs, N. Meekel, B. Talavera Andújar, D. Aurich, E. L. Schymanski, G. Frigerio, A. Macherius, U. Kunkel, T. Bader, P. Rostkowski, H. Gundersen, B. Valdecanas, W. C. Davis, B. Schulze, S. Kaserzon, M. Pijnappels, M. Esperanza, A. Fildier, E. Vulliet, L. Wiest, A. Covaci, A. Macan Schönleben, L. Belova, A. Celma, L. Bijlsma, E. Caupos, E. Mebold, J. Le Roux, E. Troia, E. De Rijke, R. Helmus, G. Leroy, N. Haelewyck, D. Chrastina, M. Verwoert, N. S. Thomaidis and A. Kruve, Quantification Approaches in Non-Target LC/ESI/HRMS Analysis: An Interlaboratory Comparison, *Anal. Chem.*, 2024, **96**, 16215–16226, DOI: 10.1021/acs.analchem.4c02902.

38 B. G. Anderson, A. Raskind, H. Habra, R. T. Kennedy and C. R. Evans, Modifying Chromatography Conditions for Improved Unknown Feature Identification in Untargeted Metabolomics, *Anal. Chem.*, 2021, **93**(48), 15840–15849, DOI: 10.1021/acs.analchem.1c02149.

39 J. Xue, B. Wang, H. Ji and W. Li, RT-Transformer: Retention Time Prediction for Metabolite Annotation to Assist in Metabolite Identification, *Bioinformatics*, 2024, **40**(3), btae084, DOI: 10.1093/bioinformatics/btae084.