Developing tools for non-target analysis and digital archiving of organic urban water pollutants

Cathrin Veenaas
To my parents and my sister

The good thing about science is that it’s true whether or not you believe in it.
– Neil deGrasse Tyson
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table of Contents</td>
<td>V</td>
</tr>
<tr>
<td>Abstract</td>
<td>VIII</td>
</tr>
<tr>
<td>Sammanfattning på svenska</td>
<td>X</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>XI</td>
</tr>
<tr>
<td>Publications</td>
<td>XIII</td>
</tr>
<tr>
<td>Author’s contributions</td>
<td>XIV</td>
</tr>
<tr>
<td>General Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Sewage Sludge</td>
<td>2</td>
</tr>
<tr>
<td><strong>PART I</strong></td>
<td>5</td>
</tr>
<tr>
<td>Data generation</td>
<td>5</td>
</tr>
<tr>
<td>Introduction to environmental analyses</td>
<td>6</td>
</tr>
<tr>
<td>Additional analytical details</td>
<td>8</td>
</tr>
<tr>
<td>Comprehensive two-dimensional gas chromatography</td>
<td>8</td>
</tr>
<tr>
<td>Mass spectrometry</td>
<td>8</td>
</tr>
<tr>
<td>Tandem mass spectrometry</td>
<td>8</td>
</tr>
<tr>
<td><strong>Method development for gas chromatography analysis</strong></td>
<td>10</td>
</tr>
<tr>
<td>Sample extraction and clean-up methods</td>
<td>10</td>
</tr>
<tr>
<td>Method development experiments</td>
<td>12</td>
</tr>
<tr>
<td>Method development results</td>
<td>13</td>
</tr>
<tr>
<td>Comparison of extraction methods</td>
<td>13</td>
</tr>
<tr>
<td>Sulfur removal efficiencies</td>
<td>14</td>
</tr>
<tr>
<td>Evaluation of evaporation techniques</td>
<td>15</td>
</tr>
<tr>
<td>Method validation</td>
<td>15</td>
</tr>
<tr>
<td>Method application</td>
<td>16</td>
</tr>
<tr>
<td>Conclusions and improvements</td>
<td>19</td>
</tr>
<tr>
<td><strong>Method development for liquid chromatography analysis</strong></td>
<td>20</td>
</tr>
<tr>
<td>Sample extraction for LC-MS analysis</td>
<td>20</td>
</tr>
<tr>
<td>Method development experiments</td>
<td>21</td>
</tr>
<tr>
<td>Method development and validation results</td>
<td>21</td>
</tr>
<tr>
<td>Method application: Non-target screening approach</td>
<td>24</td>
</tr>
<tr>
<td>Conclusions and improvements</td>
<td>25</td>
</tr>
<tr>
<td><strong>Part I summary</strong></td>
<td>26</td>
</tr>
<tr>
<td><strong>PART II</strong></td>
<td>27</td>
</tr>
<tr>
<td>Data evaluation</td>
<td>27</td>
</tr>
<tr>
<td>Time trend analysis</td>
<td>28</td>
</tr>
<tr>
<td>Statistical analysis for time trends</td>
<td>28</td>
</tr>
<tr>
<td>Time trend analysis methodology</td>
<td>30</td>
</tr>
<tr>
<td>Data-reduction strategies</td>
<td>31</td>
</tr>
<tr>
<td>Time trends detected in the GC data</td>
<td>33</td>
</tr>
<tr>
<td>Time trends detected in the LC data</td>
<td>34</td>
</tr>
<tr>
<td>Title</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Time trends of selected compounds</td>
<td>35</td>
</tr>
<tr>
<td>Comparison of parametric and non-parametric trend tests</td>
<td>39</td>
</tr>
<tr>
<td>Conclusions and improvements</td>
<td>40</td>
</tr>
<tr>
<td><strong>A new retention index system for GC×GC using polyethylene glycols</strong></td>
<td>42</td>
</tr>
<tr>
<td>Development of retention indices and GC×GC</td>
<td>42</td>
</tr>
<tr>
<td>PEG-2 retention index development and tests</td>
<td>43</td>
</tr>
<tr>
<td>Theory and calculations</td>
<td>44</td>
</tr>
<tr>
<td>Robustness and application range of the new system</td>
<td>47</td>
</tr>
<tr>
<td>Conclusions and improvements</td>
<td>49</td>
</tr>
<tr>
<td><strong>Retention time prediction</strong></td>
<td>50</td>
</tr>
<tr>
<td>Retention-time prediction approaches</td>
<td>50</td>
</tr>
<tr>
<td>Quantitative structure-retention relationships</td>
<td>51</td>
</tr>
<tr>
<td>Model types</td>
<td>51</td>
</tr>
<tr>
<td>Partial Least Squares</td>
<td>51</td>
</tr>
<tr>
<td>“Federation of local models”</td>
<td>52</td>
</tr>
<tr>
<td>Modelling workflow</td>
<td>53</td>
</tr>
<tr>
<td>Retention time prediction in GC×GC</td>
<td>54</td>
</tr>
<tr>
<td>Data generation and modelling procedure</td>
<td>54</td>
</tr>
<tr>
<td>“Quality assurance”</td>
<td>55</td>
</tr>
<tr>
<td>PLS model results</td>
<td>56</td>
</tr>
<tr>
<td>Number of PLS components</td>
<td>56</td>
</tr>
<tr>
<td>Influence of descriptors and transformation</td>
<td>56</td>
</tr>
<tr>
<td>Local models</td>
<td>58</td>
</tr>
<tr>
<td>“Federation of local models” results (ChromGenius)</td>
<td>59</td>
</tr>
<tr>
<td>Validation of models</td>
<td>61</td>
</tr>
<tr>
<td>Applying the retention prediction models</td>
<td>64</td>
</tr>
<tr>
<td>Conclusions and improvements</td>
<td>64</td>
</tr>
<tr>
<td><strong>Part II summary</strong></td>
<td>65</td>
</tr>
<tr>
<td><strong>PART III</strong></td>
<td>67</td>
</tr>
<tr>
<td><strong>Data Archiving</strong></td>
<td>67</td>
</tr>
<tr>
<td>Digital archiving</td>
<td>68</td>
</tr>
<tr>
<td>What are digital archives?</td>
<td>68</td>
</tr>
<tr>
<td>What data are needed?</td>
<td>69</td>
</tr>
<tr>
<td>Sample collection</td>
<td>69</td>
</tr>
<tr>
<td>Sample treatment</td>
<td>70</td>
</tr>
<tr>
<td>Sample analysis</td>
<td>70</td>
</tr>
<tr>
<td>Selection and standardization of chromatographic techniques</td>
<td>70</td>
</tr>
<tr>
<td>Selection of ionization techniques for GC-MS and LC-MS</td>
<td>71</td>
</tr>
<tr>
<td>Suitable mass analyzer</td>
<td>72</td>
</tr>
<tr>
<td>Changes in the analytical procedure and workflow</td>
<td>73</td>
</tr>
<tr>
<td>Data storage</td>
<td>74</td>
</tr>
<tr>
<td>Data size</td>
<td>74</td>
</tr>
<tr>
<td>Data evaluation software</td>
<td>76</td>
</tr>
</tbody>
</table>
What advantages do digital archives have? 76
What can digital archives be used for? 77
What are the current options? 77

**Final words** 81

**Acknowledgements** 82

**Bibliography** 84

**Appendix** 95

Non-target screening and time trend analysis of sewage sludge via liquid chromatography – ion mobility spectrometry – mass spectrometry 95

*Experimental part*

*Materials* 95

*Method development* 95

*Sample treatment for time trend study* 97

*Sample analysis* 98

*Data analysis for time trend samples* 99
Abstract

This thesis describes efforts to develop robust methods for the creation and use of digital archives of environmental samples, and proposes guidelines based on the results. Digital archives are repositories that store environmental samples digitally. Traditionally, samples are stored physically in environmental specimen banks over long time periods. However, this has several drawbacks, for example degradation effects and limited accessibility.

During the course of my PhD project I developed methods that allow the comprehensive analysis of sewage sludge samples. Sewage sludge is a complex matrix that contains many commercial chemicals. In addition, sewage treatment plants form a link between the human society that generates the sewage and the environment, making sewage sludge a very interesting matrix to analyze. The developed methods enable analysis and subsequent identification of compounds of all sizes and with diverse chemical characteristics. I further explain how unknown compounds can be identified (non-target screening) using mass spectral analysis and several other approaches (e.g. retention indices).

The thesis is divided into three parts. In the first part, Data Generation, I describe the development of sample preparation methods for analyzing sewage sludge with gas chromatography (GC) and liquid chromatography (LC) coupled to high resolution mass spectrometry (HRMS). For the GC approach, two methods involving use of different extraction techniques, solvents, and matrix reduction techniques are presented while for the LC approach different extraction techniques are compared. The methods have been developed to enable the generation of data suitable for digital archiving.

In the second part of the thesis, Data Evaluation, I present ways to find and identify compounds of interest. Firstly, time trend analyses provide a way to prioritize pollutants, for example by focusing on pollutants that are increasing with time. Thousands of compounds with significant time trends were detected and several hundred of them were tentatively identified. Compounds with strong increasing trends included, for example, UV-filters from sunscreens. Secondly, a new retention index system for comprehensive two-dimensional chromatography (GC×GC) is introduced to characterize compounds in terms of their retention times in the second dimension. The new retention index system is based on co-injection of polyethylene glycols and was validated for various compounds of diverse classes. Thirdly, I tested different ways to predict GC×GC retention times or indices. Those methods include a multivariate prediction (PLS) approach using molecular descriptors, which proved to be the best approach, and use of commercially available software. The last part of my thesis, Data Archiving, discusses requirements to create digital archives and how they can be used. Here I present the current...
state and options for archiving data files, and give recommendations for each step, from sample collection, through instrumental analysis to storage of the final data.
Sammanfattning på svenska

I denna avhandling beskrivs innovativa metoder för att skapa och använda digitala arkiv för miljörelaterade prover, såsom biologisk vävnad, sediment och rötslam. Digitala arkiv skiljer sig från traditionella miljöprovbanker genom att resultat från analys av miljöprover fryses digitalt, istället för att fysiska prover placeras i frys. För att testa detta nya koncept utvecklades nya metoder för omfattande kemisk analys av slam från avloppsreningsverk. Avloppsslam är spännande för att det kan ge en integrerad bild av vilka kemikalier som används i samhället. Det används också för gödsling av åkermark vilket kan leda till exponering av olika organismer, inklusive människa.


Analys av tidstrender användes för att prioritera bland detekterade föröreningar, till exempel för att finna föröreningar som ökar i halt med tiden. Tusentals föröreningar med statistiskt säker-ställda tidstrener upptäcktes och flera hundra av dem kunde ges en preliminär identitet. Föroreningar med starkt ökande trender inkluderade exempelvis kemikalier med UV-blockerande egenskaper och UV-blockerande egenskaper som används i solskyddsmedel.

Slutligen presenteras nuvarande status och utsikter för framtida användning av digitala arkiv. Lämpliga rutiner för digital arkivering diskuteras och den ges rekommendationer för varje steg, från insamling av prover, genom instrumentanalys till lagring av slutdata. Förhoppningen är att digitala arkiv framöver helt eller delvis kan ersätta miljöprovbanker och därmed undvika problem såsom begränsad tillgång till material, nedbrytning eller kontamination under lagring.
List of Abbreviations

2D  Two-dimensional
3D  Three-dimensional
A   Number of PLS components
ANDI-MS  Analytical data interchange format for mass spectrometry
ANOVA  Analysis of variances
APCI  Atmospheric pressure chemical ionization
APPI  Atmospheric pressure photoionization
ASE  Accelerated solvent extraction
ASTM  American Society for Testing and Materials
CCF  Central composite face-centered (design)
CI  Chemical ionization
CSV  Comma-separated values
DBMS  Database-management system
DCM  Dichloromethane
DCT  Data collection template
DDT  Dichlorodiphenyltrichloroethane
DROP met  Data resources of plant metabolomics
d-SPE  Dispersive solid phase extraction
EDA  Exploratory data analysis
EG  Ethylene glycol
EI  Electron ionization
ESB  Environmental specimen bank
ESI  Electrospray ionization
FAME  Fatty acid methyl ester
FTP  File transfer protocol
GC  Gas chromatography
GC×GC  Comprehensive two-dimensional gas chromatography
GPC  Gel permeation chromatography
GUI  Graphical user interface
HPLC  High-performance liquid chromatography
HRT  High resolution time-of-flight mass spectrometer
$I$  Second dimension retention index
IMS  Ion-mobility spectrometry
IS  Internal standard
$K_{OW}$  Octanol-water partition coefficient
K-S test  Kolmogorov-Smirnov’s goodness-of-fit test
LC  Liquid chromatography
LOD  Limit of detection
LOQ  Limit of quantification
LRI  Linear retention index
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-MBC</td>
<td>3-(4-Methylbenzylidene)-camphor</td>
</tr>
<tr>
<td>MICRE</td>
<td>Matrix induced chromatographic response enhancement</td>
</tr>
<tr>
<td>MLR</td>
<td>Multiple linear regression</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>MS/MS</td>
<td>Tandem mass spectrometry</td>
</tr>
<tr>
<td>MSRI</td>
<td>Mass spectral and retention time index</td>
</tr>
<tr>
<td>NIST</td>
<td>National Institute of Standards and Technology</td>
</tr>
<tr>
<td>NMR</td>
<td>Nuclear magnetic resonance spectroscopy</td>
</tr>
<tr>
<td>NPE</td>
<td>Nonylphenol ethoxylate</td>
</tr>
<tr>
<td>OP</td>
<td>Organophosphate</td>
</tr>
<tr>
<td>PAC</td>
<td>Polycyclic aromatic compound</td>
</tr>
<tr>
<td>PAH</td>
<td>Polycyclic aromatic hydrocarbon</td>
</tr>
<tr>
<td>PBDE</td>
<td>Polybrominated diphenyl ether</td>
</tr>
<tr>
<td>PCA</td>
<td>Principle component analysis</td>
</tr>
<tr>
<td>PCB</td>
<td>Polychlorinated biphenyl</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PEG–2I</td>
<td>GC×GC second-dimension retention index</td>
</tr>
<tr>
<td>PLE</td>
<td>Pressurized liquid extraction</td>
</tr>
<tr>
<td>PLS</td>
<td>Partial least squares</td>
</tr>
<tr>
<td>PPCP</td>
<td>Pharmaceutical and personal care product</td>
</tr>
<tr>
<td>PTV</td>
<td>Programmable temperature vaporizer</td>
</tr>
<tr>
<td>QuEChERS</td>
<td>Quick easy cheap effective rugged and safe</td>
</tr>
<tr>
<td>QSPPR</td>
<td>Quantitative structure–property relationships</td>
</tr>
<tr>
<td>QSRR</td>
<td>Quantitative structure–retention relationship</td>
</tr>
<tr>
<td>RAR</td>
<td>Roshal archive</td>
</tr>
<tr>
<td>RI</td>
<td>Retention index</td>
</tr>
<tr>
<td>RMSE&lt;sub&gt;CV&lt;/sub&gt;</td>
<td>Root mean square error of cross validation</td>
</tr>
<tr>
<td>RS</td>
<td>Recovery standard</td>
</tr>
<tr>
<td>SIM</td>
<td>Selected ion monitoring</td>
</tr>
<tr>
<td>SPLE</td>
<td>Selective pressurized liquid extraction</td>
</tr>
<tr>
<td>SQL</td>
<td>Structured query language</td>
</tr>
<tr>
<td>SRM</td>
<td>Selected reaction monitoring</td>
</tr>
<tr>
<td>STP</td>
<td>Sewage treatment plant</td>
</tr>
<tr>
<td>( t_R )</td>
<td>First-dimension retention time</td>
</tr>
<tr>
<td>( t_R' )</td>
<td>Second-dimension retention time</td>
</tr>
<tr>
<td>TBA</td>
<td>Tetra butyl ammonium hydrogen sulfate</td>
</tr>
<tr>
<td>USE</td>
<td>Ultrasound assisted extraction</td>
</tr>
<tr>
<td>UV</td>
<td>Ultraviolet</td>
</tr>
<tr>
<td>XML</td>
<td>Extensible markup language</td>
</tr>
</tbody>
</table>
Publications

This thesis is based on the following appended papers, which are referred to in the text by the corresponding Roman numerals.


Published papers and figures are reproduced with permission of the respective publisher (Springer Nature and Elsevier).
Author’s contributions

**Paper I**
I was responsible for sampling, lab work, sample analysis, and data interpretation such as identification of the compounds. I contributed to the planning of the experiments and wrote the paper.

**Paper II**
I carried out the lab work and sample analysis. I was responsible for the data treatment prior to statistical analysis. I identified the compounds and evaluated the time-trend analysis results and wrote the manuscript.

**Paper III**
I was involved in the planning of the experiments and was responsible for the analysis. I was partly responsible for the data evaluation, optimization of calculations and writing the paper.

**Paper IV**
I did the planning of the study, modelling and interpretation of the results with support from my co-authors. I was in charge of the instrumental analysis and writing the manuscript.
General Introduction

“There was once a town in the heart of America where all life seemed to live in harmony with its surroundings. The town lay in the midst of a checkerboard of prosperous farms, with fields of grain and hillsides of orchards where, in spring, white clouds of bloom drifted above the green fields. In autumn, oak and maple and birch set up a blaze of color that flamed and flickered across a backdrop of pines. Then foxes barked in the hills and deer silently crossed the fields, half hidden in the mists of the fall mornings. [...] Then a strange blight crept over the area and everything began to change. Some evil spell had settled on the community: mysterious maladies swept the flocks of chickens; the cattle and sheep sickened and died. Everywhere was a shadow of death. [...] There was a strange stillness. The birds, for example — where had they gone? [...] It was a spring without voices. [...] No witchcraft, no enemy action had silenced the rebirth of new life in this stricken world. The people had done it themselves.” [1]

By describing this fictitious small town in her book *Silent Spring*, Rachel Carson started a movement. She increased awareness of what pollution can do to nature not only among scientists but also among the broad public. In the introduction to an updated version of *Silent Spring* Linda Lear states that Carson’s *Silent Spring* started “a national debate on the use of chemical pesticides, the responsibility of science, and the limits of technological progress” and consequently helped ban domestic production of DDT in the USA [1]. Furthermore, Lear writes that “Carson’s writing initiated a transformation in the relationship between humans and the natural world and stirred awakening of public environmental consciousness.” [1].

Rachel Carson dedicated her book to Albert Schweitzer who said “Man has lost the capacity to foresee and to forestall. He will end by destroying the earth.” Nowadays, most people are aware that chemicals, like pesticides, can have negative effects on the environment, but I still think Schweitzer’s comment is highly relevant. Numerous chemicals are being produced, used, and emitted to the environment although we do not know what environmental effects they will have. Moreover, some chemicals have known adverse effects, but they can still be produced and used because policy making and legislative changes require time and effort. To help policy-makers formulate regulations, scientists have studied (and are studying) effects of numerous chemicals and environmental levels of pollutants. Many chemical compounds are known to have bad effects on the environment and are already banned, for example DDT and polychlorinated biphenyls (PCBs) [2]. However, large numbers of compounds are not regulated or even known. Thus, my doctoral work,
reported in this thesis, was intended to help efforts to detect unknown chemicals that enter the environment and may adversely affect it.

Unknown and new, emerging compounds greatly outnumber the known and regulated compounds (like the mass of an iceberg; most of which is hidden under water and is not easily detectable). In navigation, technologies such as radar and sonar are used to detect icebergs. Similarly, in environmental analysis we use modern technologies such as gas chromatography (GC), liquid chromatography (LC) and high resolution (HR) mass spectrometry (MS) to identify pollutants in environmental samples. In “target screening” approaches, analysts focus on known compounds that are expected to be in the samples (the visible tip of the metaphorical iceberg), but “non-target screening” enables capture of new or unexpected compounds (some of the previously hidden ice), as shown later in this thesis. However, before continuing with the introduction, I would like to inform the reader that the thesis itself is merely intended to provide an informative overview of the work. The details of the work, such as quality assurance or exact measures of chemicals, are not explained in the thesis. If you would like to repeat the described experiments, or merely scan the details, please refer to the appended papers and manuscripts.

**Sewage Sludge**

Every day more than 100,000 chemicals are currently used globally [3]. Large amounts of these chemicals, including pollutants, are being disposed of through the sewer system, and consequently enter sewage treatment plants (STPs; Figure 1). The purpose of STPs is to reduce contamination of the sewage they receive sufficiently for its safe release to the environment by removing nutrients, some metals and organic chemicals. Hence, large amounts of the nutrients, metals and organic chemicals end up in the sewage sludge. Due to the nutrients contained in the sludge, it can be used for applications such as fertilization of agricultural fields. During my PhD project I have focused on analysis of sewage sludge since it is a matrix that includes many of the chemicals used commercially and is a link between the human societies and the surrounding environment, as illustrated in Figure 1.

Between 2005 and 2014, 54% of the sewage sludge in Europe was used in land applications, 10% was deposited in landfills, and 25% was disposed of in other ways [4]. Data suggest similar disposal routes in North America: 61% in land applications, 17% in landfilling and 22% subjected to incineration [5]. Interestingly, use of sewage sludge in agriculture is increasing in various countries, including Sweden, where annual amounts used agriculturally increased from 19,800 t to 59,500 t between 2004 to 2015 [6].

In order to safely dispose of, or reuse, sewage sludge on agricultural fields, European Union directive 86/278/EEC requires member states to monitor
heavy metal concentrations in sludge and soil regularly when sewage sludge is used as fertilizer [7]. Similarly, the U.S. government has defined maximum loadings of heavy metal pollutants for sewage sludge when used as fertilizer on agricultural fields [8]. For EU member states national requirements also apply, and may be much stricter (e.g. in Denmark, Finland, Sweden, and The Netherlands) [9]. In Sweden, for example, there are additional limits for concentrations of PCBs, nonylphenol ethoxylates (NPEs), polycyclic aromatic hydrocarbons (PAHs), and toluene when sludge is applied as fertilizer [9].

![Figure 1 Schematic overview of the wastewater flow (left) and processes in a sewage treatment plant (STP; right) [10].](image)

Nevertheless, sewage sludge might pose a risk when applied to arable land as in all cases only target compounds are monitored and plants can take up diverse pollutants from the surrounding soil [11–13]. Thus, a major goal of this doctoral project was to develop methodology enabling comprehensive screening of sewage sludge, thereby permitting the detection and monitoring of unknown organic contaminants present in sewage sludge. The development of sample extraction and clean-up methods is described in the first part of the thesis (Data generation). Evaluations of methods to identify (e.g. retention indices) and prioritize (time trend analysis) identified pollutants are also presented and discussed. The results of these studies are presented in part two (Data evaluation). Finally, Part III – Data archiving – addresses important factors to consider when creating suitable data for digitally archiving environmental samples, and uses of such data.
PART I
Data generation
Introduction to environmental analyses

Before explaining in more detail the techniques used to generate data, I present here a basic introduction to the analysis of environmental samples and meaning of non-target screening.

The analysis of environmental samples always starts with sampling, i.e. collection of samples. For the first part of my PhD project sewage sludge was collected from an STP in Umeå (Sweden). In analytical contexts, such a bulk medium is also called a matrix. Other environmental matrices include water, soil and air. After sampling, compounds of interest must be extracted from the matrix, and, for example, transferred to an appropriate solvent. Various techniques can be used for this, some of which are covered in this thesis. Since not only compounds of interest but also other, unwanted, compounds (for example lipids, i.e. fats and oils) are transferred to the extraction solvent, further clean-up of the samples is sometimes needed to remove some unwanted substances and thus enable better analysis.

After the samples are extracted and cleaned, they are ready for analysis. Various types of instrumentation can be used for this. Since the extracts are usually complex mixtures (containing many compounds of interest and other, unwanted compounds), the samples are first separated using a chromatographic technique, either GC or LC. In GC, a gas is used as a mobile phase that passes through a thin capillary column while in LC a liquid solvent is used that passes through an LC column. After the chromatography, the separated sample constituents (analytes) are directly introduced into a detection system where signals are recorded. Those can be graphically displayed as “peaks” in chromatograms, specifying the analytes’ retention times and areas proportional to the amount of analyte eluting from the system. If, in addition, masses of the compounds causing the signals are recorded as well as the relative amount, spectra are obtained showing masses versus signal intensities (mass spectrum). The recorded signals can be subsequently transformed to corresponding amounts or concentrations through appropriate calibration.

Good chromatographic separation is crucial to obtain “clean” spectra, showing clearly distinct peaks with no overlap, for unequivocal identification of compounds. However, in reality compounds often co-elute, i.e. have nearly identical retention times on the chromatographic column. Thus, their peaks substantially overlap. Several variables influence the separation of compounds in GC and LC. In both cases the type of stationary phase in the column is an important factor. To increase the separation, two columns that separate compounds by exploiting different properties can be used sequentially. This process is called comprehensive two-dimensional chromatography (GC×GC or LC×LC) if all analytes eluting from the first
column enter the second column in small defined portions. In GC×GC a
standard column set [14] is a non-polar first dimension column, which
separates the compounds according to boiling point, coupled to a (semi-)polar
secondary column, which separates the analytes according to polarity
(roughly). Using GC×GC greatly increases peak capacities, and more peaks
can be identified, a feature that I exploited throughout my studies.

After GC or LC separation the analytes are ionized, i.e. charged. Various
techniques are available for this. Typically, the ionization occurs under
vacuum in GC, but at atmospheric pressure in LC. However, nowadays
analytes can also be ionized at atmospheric pressure in GC (APGC). In my
doctoral work I used the most common ionization techniques for both GC
(electron ionization; EI) and LC (electrospray ionization; ESI). In EI, the
analytes are bombarded with fast electrons. In addition to charging analytes,
this can also fragment them, i.e. create smaller, sometimes characteristic,
pieces. In ESI the compounds are ionized in solution. A current is applied in
the source that supports formation of small charged droplets from the mobile
phase. Charges within the droplets are transferred to the analytes while the
droplets disperse due to a heated gas flow. This is a rather mild (“soft”) procedure that does not generally cause fragmentation of analytes.

After the analytes are ionized or fragmented they are introduced into a
mass analyzer. There are many types of mass analyzers. The type I used
most often is high-resolution time-of-flight (TOF) MS. The charged analytes
and/or analyte fragments (ions) are pushed from the ion source into the
so-called flight tube and the time they spend traveling through the tube is
measured by the instrument. While low weight compounds travel quickly
through the vacuum in the tube, high molecular weight compounds are
slower. The principle is similar to macromolecular phenomena. Imagine
kicking a light tennis ball and a heavy football with the same force across a
football field. The tennis ball will be faster than the football and reach the end
of the field sooner. Similarly, small ions travel faster and reach the end of the
flight tube sooner than large ions. So, when the ions reach the end of the flight
tube, their flight times (which can be converted to masses) are recorded, and
their numbers are counted in the detector.

Having briefly described the basic principles of analyzing environmental
samples, I will explain potential uses. Generally, there are three types of data
analysis: target, suspect and non-target screening. Target screening focuses
on quantifying specific compounds in samples (e.g. PCBs, PAHs or
pharmaceuticals like paracetamol). To enable this, the pure, dissolved,
compounds are injected into the GC-MS or LC-MS system to acquire reference
spectra and chromatographic retention times. These data can be used to find
and quantify the compounds in samples. A suspect screening usually starts
with the compilation of possible analytes that may be present the samples.
Molecular weights, chemical formulae, spectra and retention times of these
analytes are collected, if possible, and then used to detect the compounds in the samples. In some cases, the pure analytes (reference standards) may be bought and subjected to the same analytical procedures to confirm the results. In contrast, during non-target screening no assumptions about possible analytes are made. Spectra and retention times are evaluated by comparing them to data in libraries (information such as spectra and relative retention times stored in databases), using computational prediction tools or by manual interpretation. Again, reference standard may be used for final confirmation of the suspected candidate structures. Throughout my thesis I provide increasing information about each of these steps and the available tools.

**Additional analytical details**

**Comprehensive two-dimensional gas chromatography**

In GC×GC small defined portions are introduced from the first chromatographic column to the second column in defined time intervals. The size of those portions is typically around one third of the first-dimension peak volume. Hence, the length of the interval is, theoretically, around a third of a chromatographic peak width. This value is called the modulation period, and typically ranges from 3 to 10 seconds. During this time, all analytes that enter the second column pass through it before the next portion enters the column. The second-dimension peaks originating from the same first dimension peak are called slices (see Figure 2, upper right corner). The slices are often stacked next to each other and used to create a two-dimensional (2D) or three-dimensional (3D) chromatogram (bottom part of Figure 2), with first dimension retention time on the horizontal axis, second dimension retention time on the vertical axis, and signal height represented by color (2D) or peak height (3D). For each time point in the chromatogram, one spectrum is recorded that shows which masses were obtained after the ionization including their corresponding intensities.

**Mass spectrometry**

Different mass analyzers have different properties and, hence, different advantages. For example, there are both high- and low-resolution TOFMS instruments. Increasing the effective length of the flight tube, i.e. flight path, increases the measurements’ resolution, and ability to measure ions’ exact masses. In a TOFMS instrument, all ions reach the detector, but other mass analyzers, quadrupole MS instruments, for example, can be used to sort out (select) certain ions.

**Tandem mass spectrometry**

As I previously mentioned, EI generates characteristic fragments of analytes that can be compared, for example, to entries in a mass spectral library. In atmospheric pressure ionization (e.g. ESI) generally only a few
fragments are generated. In most cases the molecular ion can be seen. In addition, impurities present in the mobile phase sometimes form complexes with the analytes (so-called adducts), which can then also be seen.

Figure 2 GC×GC chromatograms: the translation form 2D to 3D chromatograms.

To obtain characteristic fragments with soft ionization techniques, e.g. ESI, tandem mass spectrometry (MS/MS) was developed. Here, the ion source is connected to the first mass analyzer, which is then connected to a collision cell. In this cell the ions are fragmented into smaller pieces, for example using collision induced dissociation (CID), and subsequently transferred to the second mass analyzer and, finally the detector. Such setups can be operated in various “modes”. Table 1 shows the modes that are possible when using a quadrupole – time of flight (QTOF) mass spectrometer as it was used in my doctoral work.

Table 1 Experiments and data that can be obtained in MS and MS/MS mode using Quadrupole – Time of flight mass spectrometry.

<table>
<thead>
<tr>
<th>Scan mode</th>
<th>Quadrupole</th>
<th>CID*</th>
<th>TOF</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS: Full-spectrum</td>
<td>Transmitting</td>
<td>off</td>
<td>Full MS-spectra</td>
</tr>
<tr>
<td>MS/MS: Target MS/MS</td>
<td>Selected m/z</td>
<td>off/on</td>
<td>Full MS-spectra and MS/MS spectra</td>
</tr>
<tr>
<td>MS/MS: Data Dependent acquisition (auto MS/MS)</td>
<td>Selected m/z</td>
<td>off/on</td>
<td>Full MS spectra and MS/MS spectra</td>
</tr>
<tr>
<td>MS/MS: Data independent acquisition (all-ion MS/MS)</td>
<td>Transmitting</td>
<td>off/on</td>
<td>Full MS spectra at zero and high collision energy</td>
</tr>
</tbody>
</table>

*Off: no additional energy applied, on: additional energy applied (e.g., 10, 20 or 40 V), off/on: first a full spectrum without collision and then with collision is created
Method development for gas chromatography analysis

Sample extraction and clean-up methods

Many published studies have described analyses of sewage sludge, but none before my first study included non-target screening of compounds in sewage sludge. Apart from classic extraction techniques such as Soxhlet and ultrasound-assisted extraction, pressurized liquid extraction (PLE, also called accelerated solvent extraction, ASE) has been the most frequently used extraction technique [15]. In PLE, a sample is placed in a stainless-steel cell. The cell is then heated to a certain temperature and solvent is introduced under pressure. After a short extraction time, the solvent is pushed out of the cell and collected. In a next cycle new solvent is introduced and the process is repeated.

PLE has several advantages over the traditional extraction methods, including lower solvent consumption, shorter extraction times, higher extraction rates, and greater possibilities to extract polar as well as non-polar compounds [15,16]. Hence, PLE seems to be a very good method for the comprehensive extraction of sewage sludge. An additional advantage of PLE is the possibility of in-cell clean-up using, for example, Florisil, silica gel or alumina [17], or combinations of such materials [18]. Moreover, by decreasing the amount of co-extracted interfering matrix compounds from sewage sludge through adsorption of matrix compounds to the clean-up agent, the extracts can ideally be analyzed directly after the extraction [18]. This method, called selective PLE (SPLÉ), has been widely used, for example for extracting soil and sediment [19–21], food and feed samples [22], and sewage sludge [18,23].

For extractions of sewage sludge using SPLÉ in some cases no further steps besides filtering and/or derivatization have been applied [18,23]. However, in some studies further clean-up has been applied prior to GC analysis [24,25], to reduce interferences and, hence, improve limits of detection (LODs) [26]. Nevertheless, in most cases, PLE has been followed (without in cell clean-up) by further clean-up steps before GC analysis [26]. In non-target screening, a non-destructive clean-up technique, such as gel permeation chromatography (GPC), is preferred, to retain as many compounds of interest as possible.

One of the aims of the studies underlying this thesis was to develop a robust approach for comprehensive non-target screening of sewage sludge. To cover all compound classes, several methods must be applied. For this purpose, the utility and validity of PLE and SPLÉ with silica, in conjunction with several solvents and solvent mixtures, were evaluated in terms of extraction efficiency and amounts of co-extracted matrix.
Figure 3 gives an overview of the three methods I propose to use to cover (almost) the whole chemical space of analytes. Since the PLE method involves size exclusion chromatography, it will remove large analytes from the extracts together with unwanted matrix compounds. Hence, the PLE method will allow analysis of small and medium-sized compounds. The SPLE method proposed here uses silica, a polar sorbent, in on-line clean-up to remove polar matrix compounds. Naturally, polar analytes will also be removed from the extracts. Hence, the SPLE method will extract analytes of low to medium polarity. To complement those methods, an LC method (described in the next chapter) was developed for the analysis of larger and more polar compounds. The only compounds that are not covered by any of the proposed extraction methods and analytical techniques are very large and non-polar compounds (e.g., plastic polymers).

**Figure 3** Method overview [Paper I].

Figure 3 provides a simplified overview of the three methods’ coverage of analytes, and of course overlaps between their coverage may occur, i.e. a compound may be detected by two or even all three methods.

The experiments and results from the GC method development phase of the project are summarized in this chapter and presented in detail in Paper I.
Method development experiments

The aim of this first part of my PhD studies, and the study reported in Paper I, was to develop a sample treatment method that allows, with subsequent GC analysis, the detection of virtually all GC-amenable compounds. In practice, there are several limitations, imposed for example by the columns chosen (for instance, very large compounds with high boiling points cannot be detected using a 30-m long column) or injector temperatures (e.g., high temperatures can degrade analytes).

As described above, PLE and SPLE with silica were used to extract the sewage sludge samples. In the SPLE approach the samples are cleaned within the cells. A layer of silica is placed in the bottom of the PLE extraction cell followed by the sample (at a 5:1 silica to sample ratio). When the solvent is introduced from the top, it first passes through the sample, then (with extracted substances) through the layer of silica. This removes polar matrix compounds and leaves non-polar compounds in the extract. The PLE extracts are then cleaned-up by size exclusion chromatography (GPC). This removes large matrix compounds. Hence, only small and medium sized analytes remain in the extract.

The factor I varied during the method development experiments was the extraction solvent. Two solvents, both alone and in two ratios, were tested: dichloromethane (DCM), \( n \)-hexane/DCM (1:1), \( n \)-hexane/DCM (8:2) and \( n \)-hexane. The results were evaluated in terms of amounts of co-extracted matrix, and recoveries of a large number of target analytes. For the first of these variables, the samples were extracted, the solvent was evaporated, and the solid residue was weighed. For the latter, the sludge was spiked with a portion of a 76-component contaminant mixture (8270 MegaMix), the sample was extracted, and the recoveries were determined using GC-HRMS analysis.

Since sewage sludge can contain a high amount of elemental sulfur, which disturbs GC-MS analysis, sulfur must be removed before analysis. For this, I tested two materials that are reportedly good for sulfur removal and compared their effects on analyte recovery: granular copper activated with concentrated hydrochloric acid, and a tetra butyl ammonium (TBA) sulfite reagent mixture. Portions of solvent were spiked with the 8270 MegaMix and some additional pesticides at known concentrations, and subjected to both sulfur removal procedures. Recoveries of all analytes were subsequently determined.

Finally, I compared two systems (Turbo-Vap and Rota-Vap) for solvent evaporation. In the Turbo-Vap system, a stream of nitrogen creates dispersion within the extract and, with warming by a water bath, facilitates solvent evaporation. The Rota-Vap uses a combination of vacuum and a water bath to evaporate the solvent, rotating the flask containing the extract to form a thin solvent film on the flask’s walls. To compare the two techniques, again
solvent was spiked with a portion of the 8270 MegaMix, the solvent was evaporated to ~1 mL and the recoveries were compared.

To evaluate the final method’s utility, and validate it, I extracted unspiked sewage sludge and sewage sludge spiked with 99 analytes (including the MegaMix, PCBs, organophosphates (OPs), fragrances, pesticides, and others) using the PLE and SPLE methods with n-hexane/DCM (8:2). The PLE samples were cleaned by GPC, and finally both the PLE- and SPLE-generated extracts were treated with the TBA-sulfite reagent for sulfur removal. Generally, sulfur can be removed with GPC. However, this would mean excluding some other small analytes or large PAHs, and thus was not used in this study. All samples were analyzed using GC×GC-HRMS to determine the recoveries. As an additional application test, a non-target screening of the unspiked sewage sludge samples was performed. The identification of the unknown compounds detected is explained in the results section.

**Method development results**

*Comparison of extraction methods*

As outlined above, PLE and SPLE (using n-hexane and DCM, in both pure forms and mixtures) were tested and evaluated. Initially, use of pure DCM was excluded as it led to very high matrix co-extraction (8.1% and 5.0% with PLE and SPLE extraction, respectively). Recovery ranges for all compounds, and amounts of co-extracted matrix, obtained in all the experiments with the other solvents, are shown in **Figure 4**. Box plots of results of the PLE experiments (without silica) show a clear difference between results obtained with pure n-hexane and a mixture of n-hexane and DCM as the extraction solvent. This indicates that n-hexane alone recovers the selected compounds less strongly than a mixture with DCM. The differences in recovery obtained with the 8:2 and 1:1 mixtures of n-hexane and DCM are less pronounced. However, there was a difference in amounts of co-extracted matrix obtained with these solvents. Although this difference was not statistically significant, it was decided that the 8:2 mixture is more suitable than the 1:1 mixture, affording good recoveries and relatively low amounts of co-extracted matrix.

For the SPLE recoveries the relations are less clear. However, higher percentages of matrix were co-extracted with n-hexane/DCM (1:1), with no compensatory improvement in recoveries. Thus, this combination of extraction technique and solvent mixture was regarded as unsuitable. Moreover, as there was only a small difference in extraction efficiency between pure n-hexane and n-hexane/DCM (8:2), an additional extraction using n-hexane/DCM (95:5) was performed to test if this would improve results. Since the 95:5 mixture yielded lower recovery values (median and 90-percentile) than the 8:2 mixture, n-hexane/DCM (8:2) was chosen as the final solvent.
**Figure 4** Comparison of amounts of co-extracted matrix (green) and recoveries obtained (box plots) with PLE and SPLE in conjunction with indicated solvents: medians, 10- and 90-percentiles, minimum and maximum recoveries and mean amounts of co-extracted matrix with standard deviations.

**Sulfur removal efficiencies**

The pre-studies included comparison of the suitability of treatments with copper and the TBA-sulfite reagent for sulfur removal. As shown in Figure 5, total numbers of compounds extracted using the two procedures were similar, but they were slightly higher following the copper treatment. There were also differences in recoveries, for example, the copper treatment resulted in slightly higher median recoveries for both the MegaMix and the pesticide mix.

Furthermore, the copper treatment resulted in recoveries with a lower overall spread than treatment with the TBA-sulfite reagent, manifested in narrower boxes in Figure 5. In addition, the copper treatment is much easier to apply, and less time- and labor-intensive. For these reasons, the copper treatment was preferred over the TBA-sulfite reagent method.

US EPA method 3660B [27] states that two materials, copper powder and TBA-sulfite reagent, can be used for sulfur removal. Furthermore, it states that use of copper powder negatively affects recovery of certain pesticides, for example, reducing recoveries of heptachlor to 5%, and completely removing malathion, ethion, and diazinon. However, in this study recoveries for those four compounds were 82-92%. The reason for the differences could be in the copper used. I used copper granulate while US EPA method 3660B suggests use of fine copper powder.
Figure 5 Comparison of effects of removing sulfur from samples using the copper and TBA-sulfite reagent treatment: median, 10- and 90-percentile, minimum and maximum recoveries, and total numbers of compounds recovered following each treatment.

This part of the method development was carried out after the method validation (see below) took place, to further improve the method. Since the TBA-sulfite reagent method has been presented as a soft and non-invasive method [27,28], it was used in the method validation experiments. However, I recommend use of copper granulate in future applications.

**Evaluation of evaporation techniques**

For the comparison of the two evaporation systems the recoveries of all compounds after evaporation with the Rota-Vap and Turbo-Vap systems were compared. For almost all compounds the Rota-Vap yielded better results. The only compound for which the Turbo-Vap provided better results was bis(2-ethylhexyl) adipate. However, standard errors indicated that the difference was not significant. The Rota-Vap system was especially advantageous for compounds with low boiling points. Hence, it was used to reduce solvent volumes in all further experiments.

**Method validation**

Acceptable recoveries of both non-polar and moderately polar compounds included in the method validation process were obtained with both of the presented PLE and SPLE methods. Recoveries of these compounds ranged
from 64 to 136%. However, recoveries were sometimes lower for the more polar compounds, for example OPs and chlorophenols. Those compounds would better be analyzed using the LC method presented below.

A few other compounds, for example bisphenol A and some other phenols, showed high recoveries and standard deviations. I believe this was due to an effect called matrix-induced chromatographic response enhancement (MICRE) [29,30]. It involves matrix compounds in the sample binding to active sites in the chromatographic equipment, thereby reducing their availability for analytes to bind to. In calibration solutions prepared for quantification no such matrix compounds are present. Hence, active sites are free to bind the analytes. This could reduce amounts of some analytes reaching the MS and, thus, reduce signals for those analytes in calibration runs. To prevent MICRE, it has been reported that analyte protectants can be added to the sample and/or standards immediately before the analysis to block the active sites in the GC [31–33].

Since matching internal standards are not usually available during a non-target screening, exact quantification is not possible, and in most cases analyses are semi-quantitative. For this procedure, areas of peaks of analyzed standards that are most similar to tentatively identified compounds are compared to those of the tentatively identified compounds. This enables rough estimates of concentrations of the tentatively identified compounds.

**Method application**

To test the proposed methods, a non-target screening of sewage sludge was performed. The compounds were identified in a tiered approach with five steps:

1. The detected compounds were classified into groups (phthalates, long chain amides, long chain ketones, alkanes and alkenes, and fatty acids) using the instrument software. A combination of characteristic elution regions in the GC×GC chromatograms and characteristic EI fragments was used to identify compounds belonging to those classes.

2. The NIST library was then used to tentatively identify compounds with high spectral similarity match (>750) or high probability score (>7000) in combination with linear retention indices (LRIs), where possible. If no LRI information was available, a simple comparison of boiling points of possible candidate compounds and closely eluting alkanes was performed. This was done to rule out compounds that clearly did not match the apparent boiling point range.

3. In the next step, chlorine and bromine filters were applied. Here I exploited the high mass accuracy (and high resolution) of the MS instrumentation used. Only compounds with characteristic losses from the molecular ion, e.g. losses of \( \text{Cl}_2 \), \( \text{Br}_2 \), HCl or HBr were displayed.
4. The final step targeted halogenated compounds that were not detected by the instrumental software’s peak-picking, but might still be important. The mass defects of chlorine and bromine were used to identify chlorinated and brominated compounds. As a reference for normalization the nominal isotope spacing divided by the exact isotope spacing was used. Thus, compounds containing chlorine or bromine, will show up as lines of horizontally aligned peaks in a mass defect plot.

Numbers of compounds that could be identified in each step using both the PLE and SPLE methods are shown in Table 2. In total, more compounds were detected using the PLE method. However, this does not necessarily mean that the PLE method is better, it may simply mean that the methods have different strengths and weaknesses (as shown later). Furthermore, Table 2 shows that the higher the tier the less compounds are identified, partly because when compounds are identified they are removed from the list of compounds that have not yet been identified. Thus, a compound can only appear in one tier, and will be counted in the first tier where it appears. For both the PLE and SPLE methods, the two largest groups of compounds that were detected are alkanes/alkenes and alkylated PAHs. This is also reflected in Figure 6.

Furthermore, Figure 6 shows the characteristics of compounds extracted using the PLE and SPLE methods. While the PLE method allowed extraction of more polar compounds (right hand side of the PLE contour plot), the SPLE method allowed extraction of a few additional large compounds (upper side of the SPLE contour plot) that were not extracted using the PLE method. Figure 6 also shows that sets of compounds extracted using the two methods substantially overlapped, but both methods extracted some compounds that the other missed. For example, while the SPLE method yielded more alkyl PAHs around the elution time of heneicosane (C21), or more alkanes, the PLE method allowed extraction of more phthalates.

<table>
<thead>
<tr>
<th>Tier</th>
<th>Technique</th>
<th>PLE</th>
<th>SPLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Classification</td>
<td>231</td>
<td>187</td>
</tr>
<tr>
<td>2</td>
<td>NIST Similarity</td>
<td>267</td>
<td>174</td>
</tr>
<tr>
<td>3</td>
<td>NIST Probability</td>
<td>37</td>
<td>11</td>
</tr>
<tr>
<td>4</td>
<td>Chlorine/bromine filters</td>
<td>11</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>Mass defect</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Sum</td>
<td>552</td>
<td>379</td>
</tr>
<tr>
<td></td>
<td>Tentatively identified</td>
<td>321</td>
<td>192</td>
</tr>
</tbody>
</table>
**Figure 6** Numbers of compounds of indicated groups with increasing size towards the top and increasing polarity towards the right (to improve visibility the lowest category includes zero, one and two compounds while the highest category includes 11 up to 20 compounds).

Besides these groups, I identified other compounds, including alkylphenols and plant extractives, several OPs, pharmaceuticals and personal care
products (PPCPs), stabilizers, antioxidants, UV screens, and other halogenated and process chemicals in steps 2 and 3 of the tiered approach. I detected 12 of these compounds using both the PLE and SPLE methods. Furthermore, the PLE and SPLE methods respectively detected 49 and 6 compounds that the other method missed in the tier 2 and 3 groups. A complete list of detected compounds is presented in Paper I.

The final group consisted of chlorinated or brominated compounds identified in steps 4 and 5. Seven of these compounds were identified using both methods (five PCBs, one DDT metabolite and triphenylchloromethane, which is used in organic synthesis and has not previously been reported in sludge). The PLE method detected a further 10 compounds: an additional PCB (PCB 151), three PAH derivatives, a commercial disinfectant (dichloroxylenol), two pharmaceutical impurities of sertraline (4-(3,4-dichlorophenyl) tetralone isomers), which have not previously been reported in sewage sludge, two compounds used in tire production, and p-(6-chloro-4-phenyl-2-quinolyl) aniline.

Several of the tentatively identified compounds had not been previously detected in sewage sludge. This was probably partly because this was the first non-target screening of sewage sludge, and partly due to the increases in peak capacity and separation afforded by use of GC×GC.

**Conclusions and improvements**

Although more compounds were detected using the PLE method, the acquired data show that the two methods are complementary. As mentioned before, the sets of analytes extracted by the two methods substantially overlap, and are not as easily separable as Figure 3 might suggest, but both methods extract compounds that the other misses.

When I reflect on the overall method development, I realize that effects of other factors in addition to the extraction solvent could have been examined. One aspect that I could have addressed is the type of sorbent used in the SPLE approach. As mentioned before, Florisil and alumina can be used in SPLE [17], and other (vendor-specific) sorbents might also be available. Moreover, comparison of alternative extraction methods (as applied in the LC method development, see below) might give an improvement. However, the advantage of in-cell clean-up would be lost if some other extraction methods, for example ultrasound-assisted extraction, were applied. Then, either off-line silica gel clean-up would be required or only the PLE method (before GPC) would need to be replaced. In both cases, effects of many factors would need to be studied.
Method development for liquid chromatography analysis

Sample extraction for LC-MS analysis

In the introduction to the GC method development I mentioned some of the advantages of “new” sample extraction techniques compared to classical techniques, for example lower time and solvent consumption. Besides PLE, many other methods have these advantages. In this chapter I discuss four methods (a classic extraction technique and three novel techniques) for extracting sewage sludge with subsequent LC-MS analysis to perform a non-target screening.

The first method I tested, ultrasound-assisted extraction (USE), is considered a classical extraction technique [15] and has been used many times before. A sample and solvent are mixed and sonicated for a certain time, then the solvent is usually transferred to a suitable container and the procedure is repeated with fresh solvent. Examples of USE applications include extraction of PPCPs [34–36], pesticides [37], linear alkylbenzene sulfonates (LAS) and PAHs [38] from sewage sludge.

The second method is called QuEChERS (Quick Easy Cheap Effective Rugged and Safe) and was first described in 2003 for the extraction of pesticides from fruits and vegetables [39]. The original QuEChERS method consists of an extraction step and a clean-up step using dispersive solid phase extraction (d-SPE). As the name suggests, it is easy and consumes low amounts of time and solvent. A QuEChERS extraction consists of three simple steps: add solvent and shake, add salts and shake, then centrifuge and transfer the supernatant. Apart from the extraction of pesticides from food samples, the QuEChERS method has been previously applied to extract various groups of compounds from sludge, for example pharmaceuticals and hormones [40], PPCPs [41], and pesticides [42]. In my studies, I tested the QuEChERS extraction without d-SPE clean-up for the extraction of sewage sludge.

The third method I tested involved use of a BeadBeater to homogenize samples by adding suitable beads (for example, zirconium beads) and solvent then rapidly vibrating the resulting mixtures [43]. The method uses very low amounts of solvent, between 1 and 1.5 mL per extraction cycle, and is fast and easy to apply. Originally this method was mainly applied to tissue samples [43,44] but a BeadBeater has been used for sewage sludge extractions in several other studies [45,46].

The last method that was tested was the PLE. As described in the GC method development section, it provides the advantages of being solvent and
time efficient and has been proven a good method for sample extraction when followed by GC-MS analysis (Paper I).

The experiments performed for the LC method development are described in the Appendix.

**Method development experiments**

The **aim** of this part of the study was to develop a reliable method for extracting samples before analysis with LC-MS. I examined effects of several factors during this study. Firstly, I compared the BeadBeater, PLE, QuEChERS, and USE extraction methods in terms of extraction efficiency, matrix effects and reliability/repeatability. After that, the method that proved to be best was further improved.

Sewage sludge samples were spiked with a range of internal standards of different compound groups (e.g. OPs, phthalates, pesticides, pharmaceuticals) and then **extracted** with each method using acetonitrile. All samples were analyzed with LC-MS using electrospray ionization (ESI) in positive and negative modes. Recoveries of all internal standards, and the matrix effect (slope of the calibration curve obtained with matrix dilution divided by the slope of the calibration curve with solvent dilution) were then determined for each of the compounds. The extraction efficiency was obtained by dividing the recovery by the matrix effect value.

In addition to the extraction efficiency, I tested effects of several other factors to improve the final extraction method (the BeadBeater method). More specifically, I compared BeadBeater extraction efficiencies at neutral, basic and acidic **pH**, and examined effects of **hydrating** dry samples (freeze-dried sludge) before extraction.

**Method development and validation results**

The main factor I studied was the type of extraction method. Results obtained with all four tested methods, including an overview of the extraction efficiencies (i.e. recoveries with matrix effects taken into account) are compared in Figure 7. The recoveries were divided by matrix effects calculated using a matching or closely eluting (LC retention time) internal standard. A matrix effect of 100% (or 1) indicates no effect at all, while a value below 100% indicates ion suppression and a value above 100% ion enhancement.

The first observations I would like to highlight are the very high recoveries provided by the **PLE** method. The recovery of one compound (trimethylphosphate) was excluded from the graph of PLE recoveries to enhance visibility, since it was much higher than recoveries of other compounds. The median extraction efficiency for the extraction with PLE
exceeded 150\%, and the 90-percentile exceeded 400\%. This, and the large overall spread of recoveries, indicates rather low reliability. High ion suppression effects – on average a 43\% matrix effect with a minimum of 1\%, i.e. 99\% ion suppression – caused the overall high extraction efficiency values, since the recoveries were divided by matrix effects to obtain the extraction efficiency. Matrix effects for the other three methods were much weaker. Visual examination of the PLE extract (background of Figure 7) showed that a large amount of visible matrix was extracted. The extracts were clearly darker than those obtained using the other methods, and solid residues were attached to the walls of the sample container. For these reasons, I regarded PLE as unsuitable in this case. The use of PLE in combination with a clean-up technique, as applied in the GC methods for example, could be considered. However, I decided to include as few steps during the sample treatment as possible.

**Figure 7** Results of the comparison of indicated extraction methods: the horizontal double bar indicates a change in axis spacing, and a picture of the extracts is shown in the background. Median, 10- and 90-percentiles, minimum and maximum recoveries are shown, and total numbers of suspects (from suspect screening) identified following each treatment.

Of the remaining methods, the **BeadBeater** method yielded results closest to 100\%. Although the median extraction efficiency was only slightly above 50\% it still provided higher median efficiency than the other remaining methods. The 90-percentile was at 114\%, while corresponding values for the other methods, **QuEChERS** and **USE**, only reach 47\% and 71\%, respectively. Additionally, the matrix effects were reasonable for the BeadBeater method:
78\% (i.e. 22\% ion suppression) on average, while the 10- and 90-percentiles were 53\% and 97\%, respectively. Moreover, in a suspect screening using the Agilent Water library and an in-house library, the most compounds were identified following BeadBeater extraction (green dots in Figure 7). Consequently, I chose it as the best extraction method. I also paid particular attention to compounds for which the two GC methods provided low recoveries, but LC-MS methodology provided better recoveries. Four of those compounds were included in the LC method development experiments (three OPs and one phthalate), and all of them showed improved recoveries using the BeadBeater method, with values ranging from 77 to 107\%.

In addition to the extraction method, I examined effects of the pH of the sample and solvent mixture on the BeadBeater extraction. The results, presented in Figure 8, show that the recovery values could be improved by adding additional extraction steps with a different pH after the neutral extraction. Both the median and 90-percentile shown in the graphs increased when using either an acidic or basic solvent after extraction with a neutral solvent. However, the acidic extraction yielded greater increases in recoveries. The basic solvent provided better recoveries than the acidic solvent for just four compounds: two OPs and two pharmaceuticals (fluconazole and tramadol). To minimize the number of steps, I decided to only perform the acidic extraction after the neutral extraction. To further reduce the amount of extracts for analysis, the neutral and acidic extracts were combined, after neutralizing the acidic extract by adding sodium hydrogen carbonate.

![Figure 8 Results of comparisons of effects of pH and hydration on extraction efficiency: median, 10- and 90-percentiles, minimum and maximum recoveries.](image-url)
In further efforts to improve the extraction and, hence, recovery values for the analytes of interest we studied effects of **hydrating** samples prior to extraction. The hypothesis behind this experiment was that adding water to the samples weakens interactions between analytes and the sample matrix [47]. Accordingly, soaking samples with water prior to extraction increased the median recovery value by 15%. Moreover, the overall spread was reduced (see inter-percentile ranges in **Figure 8**). Thus, hydration could improve recoveries of compounds with previously low extraction efficiencies.

### Method application: Non-target screening approach

The identification of unknown compounds via LC-MS is considered more complex than identification via GC-MS. In GC-MS analysis, spectra generated with EI are comparable across all instruments, and several extensive commercial libraries (for example, the NIST library) that facilitate the identification of unknown compounds have been compiled.

In contrast, since LC-MS usually involves soft ionization techniques (e.g., ESI), often only a molecular ion (with various adducts) can be detected and no characteristic spectrum is obtained. Using LC-MS/MS, however, allows the generation of fragment ions, but no standard collision energy has been defined yet, and big differences between different instrument designs and vendors exist. Consequently, only a few commercial libraries are available, and most are MS instrument- and vendor-specific. Generally, they include only a few thousand compounds, while GC-MS libraries are considerably larger.

Nevertheless, the existing libraries can be used for suspect screening. In addition to these MS/MS libraries, simple lists of suspected analytes and the corresponding formulae as well as retention information (if available), can be used to perform a suspect screening. However, to ensure that a compound is correctly identified, a reference standard should be used for confirmation.

In 2014, an article describing several levels of confidence for the identification of unknown compounds via LC-MS analysis was published [48]. To reach level 1, the highest confidence level, confirmation using a reference standard is necessary. Level 2 confidence is obtained by determining, at one of two sub-levels, a “probable structure”. Using MS/MS libraries to perform a suspect screening would result in level 2a confidence. Level 2b confidence is reached by excluding all, but one, possible structures. This can be done using fragment information obtained from MS/MS spectra (for example, by using in-silico fragmentation tools) or information about the precursor ion (depending on the scan mode). At level 3, a tentative candidate structure determined from a formula (using accurate mass measurements of the molecular ion), MS/MS data, and retention time information, is obtained. At level 4, MS/MS data are unavailable and only a molecular formula is assigned, whereas, at level 5, only the accurate mass could be determined.
The identification of unknown compounds involves three steps, as explained by several previous authors, e.g., [49]:

1. **Target analysis** is performed using reference standards for comparison. A limited number of compounds is sought, using mass and retention time information obtained from the reference standard.

2. **Suspect screening** is also aimed at identifying a limited number of compounds. As previously stated, MS/MS libraries (i.e., suspect lists) can be used for identifying compounds. Furthermore, reference standards are initially unavailable, so retention times must be predicted and then matched in addition to the MS/MS spectra.

3. **Non-target screening** is aimed at identifying as many compounds as possible. In this step, molecular formulae are generated for compounds that were not identified through target or suspect screening. Databases can be subsequently searched for possible structures. These structures are then compared with the unknown compounds using, for example, retention-time prediction or MS/MS prediction tools (e.g., MetFrag [50]). To identify truly unknown compounds, i.e. compounds that are not included in any databases, tools such as MOLGEN-MS can be used to derive structures directly from spectra [51,52].

Compounds that are tentatively identified through a suspect screening are sometimes deemed incorrect when, for example, reference standards are used for the final confirmation. In these cases, the molecular formulae of the candidates can still be maintained and used in the next step. The formulae for the compounds are assumed to be correct, since the exact mass and isotopic patterns match, and only the structures are wrongly assigned. The next step would then be similar to non-target screening.

**Conclusions and improvements**

To summarize, the BeadBeater method appears to be a good method for extracting sewage sludge. In addition, recovery of target analytes can be improved by applying additional extractions with acidified solvent and including a hydration step. Compounds exhibiting relatively low recoveries with both GC methods exhibit better recoveries when BeadBeater extraction is employed.

Effects of several factors were considered during method development, but (as always) there is scope for further improvements. The optimal type of extraction solvent was not considered in this study. Compared with those used for GC-MS, more polar solvents are usually used for extracting samples for LC-MS analysis. Widely used solvents include methanol and acetonitrile. In some cases, the solvent properties can be changed by mixing the solvents with water. In this study, it was decided to use acetonitrile to reduce amounts of co-extracted lipids and other potentially interfering substances.
Part I summary

In this part I have shown, hopefully convincingly, that several analytical methods are needed for comprehensive non-target screening and how unknown compounds are, or can be, identified. The three tested methods (PLE and SPLE with subsequent GC-MS analysis and BeadBeater extraction followed by LC-MS analysis), allow characterization or even identification of the many unknown compounds that are represented by the bottom part of the iceberg that is hidden from our view. In the following part, Data Evaluation, I describe several additional methods for identifying unknown compounds in GC×GC analysis and prioritizing among the compounds present in environmental samples.
PART II
Data evaluation
Time trend analysis

In this part of the thesis, I describe a time-trend study we performed on sewage sludge collected from Stockholm once every year between 2005 and 2015. Time-trend studies are performed to determine if, for example, concentrations of certain compounds in the samples are increasing or decreasing over time. These are referred to as monotonic trends. A decreasing trend may indicate that a certain regulation or law applied to ban or limit use of a compound has been successful. However, increasing trends may reveal where laws are missing, and attention of both authorities and scientists is required. Increasing trends could reflect an increased production and use in society. Time-trend analyses can also show if a compound has increased up to a certain point and started decreasing again (or vice versa). These more complex patterns are called non-monotonic trends. For all trend analysis, statistical tests can be applied to determine if a trend is significant.

Statistical analysis for time trends

Time-trend analysis is a widely used technique for exploring patterns in data. Temporal trends are important indicators and helpful tools for authorities in identifying chemicals that pose potential threats due to increasing concentrations [53]. They can also indicate the effectiveness of adopted measures. Sewage sludge contains chemicals from urban stormwater and wastewater inputs, so it serves as a useful “index” of environmental pollution trends associated with catchment areas [54].

Generally, trend tests can be divided into two categories: parametric and non-parametric. Parametric tests provide more information than non-parametric tests about the type of trends (for example, linear) displayed in a dataset. The data must also meet certain criteria, such as being normally distributed. Simple parametric measures include standard deviations and mean values. The non-parametric alternatives for these are interquartile ranges and medians.

Changes in levels of compounds in the environment over time, i.e. time trends, are often exponential (rather than linear), so in a first step, data are often logarithmically transformed [55–57]. A simple linear regression analysis can then be applied to detect relevant log-linear trends. Linear regression analysis is well-established and widely applied. As described in detail by various authors, for example [58], a straight line through the data is adjusted in such a way that the sum of the distances of each point from the line (“residuals”) is minimized. A null hypothesis significance test (for example, analysis of variance; ANOVA) can then be applied to check if the slope of the
regression line differs significantly from zero, indicating a significant trend is present [59].

Two of the most often used parametric tests are the t-test and ANOVA. For example, a paired t-test and ANOVA have been used to determine significant differences with time in average concentrations of DDTs and PCBs in guillemot eggs collected from 1969 to 1989 [60]. Eggs laid both early and late during a year were sampled. ANOVA was used to detect trends in parts of the dataset (e.g., data obtained from samples collected early in the year), whereas paired t-tests were used to detect differences between samples collected early and late in each year. Generally, ANOVA is used when several averages are compared (e.g., one average from each year) while a t-test is used when only two averages are compared (e.g., average concentrations of a compound in eggs laid early and late in a year).

Other common types of trend tests include non-parametric Spearman or Kendall Tau rank correlation analyses [53], which only consider relative distances between the points, or rather rankings of these distances. Therefore, these tests can only qualitatively describe monotonic trends (i.e., trends that either increase or decrease over time) rather than the nature of a trend. Non-parametric tests only identify whether a trend is significant, whereas parametric tests also identify the type of trend (e.g., log-linear), so parametric tests are often used. However, as previously mentioned, parametric tests require a normal distribution of the data. In addition, similar variances of the data are required when several groups are compared. Fulfillment of these criteria requires larger datasets than those required for non-parametric tests. If the datasets are too small, two options may be considered: choose a non-parametric test to be on the safe side, or assume that the data are normally distributed [61]. In any case, if a test for normality is performed on a sufficiently large dataset and it reveals that the data are non-normally distributed, a non-parametric test should be chosen. Non-parametric tests are also suitable for datasets containing outliers or extreme values that are retained even after the data are transformed. Those values may influence the statistics associated with a parametric test, but have less effect on the statistics associated with non-parametric tests, which are less sensitive to outliers [62].

As already mentioned, parametric tests require normal distribution of the data and equal variances. In trend testing, as performed in this work, the residuals (i.e., distances of the points from the regression line) must be normally distributed. The normality of the distribution can be determined via the Kolmogorov-Smirnov goodness-of-fit test (K-S test). The K-S test compares two distributions, i.e., the normal distribution and the distribution of data in the dataset of interest, and checks whether or not they significantly differ [60]. The equality of variances in a dataset can be tested using Bartlett’s test (parametric) or Levene’s test (non-parametric)[63].
An issue that must be considered when dealing with statistics like these is that **errors** occur. As previously noted, “Given that we are working with probabilities and not certainties [sic], it is possible that you will reject a null-hypothesis that is true” [64]. An error of this type, i.e. a false positive, is a type I error. Our trend testing was based on the assumption that the slopes did not differ significantly from zero, i.e., no increasing or decreasing trends were present. In some cases, we may have falsely concluded that a trend was present, a scenario that is suitably described by the term false positive. The probability of such an error occurring depends on the significance level applied during the testing. A significance level of 5 or 1% is typically used. These levels indicate that 5% or 1%, respectively, of the positively identified trends are actually insignificant [64]. These errors can only be reduced by decreasing the significance level and must be considered during data analysis.

Results of the time-trend analysis of the GC data are presented in **Paper II**. Corresponding analysis of the LC data will be published in a separate paper. Details about the treatment and analysis of samples for LC-MS experiments can be found in the Appendix.

**Time trend analysis methodology**

The **aim** of this analysis was to detect monotonic and non-monotonic time trends in sewage sludge samples collected over 10 years using the proposed GC and LC methods introduced in **Part I**.

Sewage sludge **samples** were obtained from the Swedish Museum of Natural History's environmental specimen bank (ESB). These samples had been collected each autumn from a sewage treatment plant in Stockholm (Henriksdal) from 2005 to 2015, freeze-dried, and stored in a freezer.

The samples were extracted and analyzed using the three methods described in **Part I (Data generation)**:

1. PLE followed by GPC, sulfur removal, and GC×GC-HRMS
2. SPLE with silica followed by sulfur removal, and GC×GC-HRMS
3. BeadBeater extraction with neutral and acidic solvents followed by LC-IMS-MS/MS\(^1\) in positive ESI (ESI+) and negative ESI (ESI-) modes

These methods are referred to hereafter as the PLE/GC, SPLE/GC and BeadBeater/LC/ESI+ and BeadBeater/LC/ESI- methods, respectively. The data were first aligned and then reduced in a stepwise manner, by removing compounds that were:

1. detected with signals less than three times stronger than blank values
2. detected in less than two out of three replicates

\(^1\) Liquid chromatography – ion-mobility spectrometry – tandem mass spectrometry
detected in samples from fewer than four of the 10 covered years.

We investigated two types of trends: **log-linear** and **non-monotonic**. Log-linear trends were investigated by linear regression analysis of the logarithmic data. Only trends that were significant at a level of $\alpha=0.05$ were considered. Slopes of the obtained curves represent the yearly increases or decreases of compounds’ concentrations in percent. Non-monotonic trends were investigated using a 3-point running average smoothing function (smoother).

In addition, **Mann-Kendall** tests were applied to the LC data. These tests reveal differences in results obtained from log-linear regression and the non-parametric alternative, and were subsequently applied to detect monotonic trends. In environmental analysis, a normal distribution of the residuals is often assumed, so parametric tests are used. However, LC-MS data are less reproducible than GC-MS data. Furthermore, GC data are normalized using a volumetric standard, whereas normalization of LC data was not possible. Hence, the residuals for LC data may be non-normally distributed.

Suspected **extreme values** were detected using the distances between the measured and smoothed values. The limit was calculated using the average standard deviation for all points in a curve. A value was considered extreme if it deviated more than three standard deviations ($\pm 99\%$) from the smoother function. Extreme values were always reported, and never excluded from the dataset. However, particular attention was paid to time series that included an extreme value by manually checking the quantification.

**Data-reduction strategies**

The stepwise procedure for reducing the amount of data is shown in **Figure 9**. After the peak picking step, the amount of data was reduced stepwise, as described in the methodology section. The first step of this procedure was alignment, followed by normalization (for GC data) and removal of data that were lower than three times the blank levels. The number of peaks was reduced mainly in the second step, retaining peaks detected in at least two replicates and thus eliminating all randomly picked-up “peaks”. In the third and final step, compounds detected in samples from less than four of the 10 covered years were eliminated. The first, second and third steps respectively yielded: 8, 87 and 75% reductions of the PLE/GC dataset; 2, 92, and 70% reductions of the SPLE/GC dataset; 13, 43 and 12% of the BeadBeater/LC/ESI+ dataset; and 0.5, 56 and 68% of the BeadBeater/LC/ESI- dataset. Finally, only 3, 2, 43 and 14%, of the data in the respective datasets were subjected to trend analysis.
Figure 9 Workflow for the time-trend analysis of sewage sludge via GC and LC. PLE, SPLE, LC ESI+, and LC ESI- refer to the PLE/GC, SPLE/GC, BeadBeater/LC/ESI+ and BeadBeater/LC/ESI- methods, respectively.

We performed a log-linear regression and smoother analysis to detect monotonic and non-monotonic trends, respectively, in the remaining data. Almost identical numbers of compounds displaying increasing and decreasing log-linear trends were detected, but a slightly lower number of non-monotonic trends was detected in the PLE/GC dataset (Table 3). Significantly more compounds displaying decreasing than increasing trends were detected in the SPLE/GC dataset. The number of non-monotonic trends was equal to the number of decreasing trends.

Table 3 Numbers of detected compounds in sewage sludge displaying log-linear (increasing and decreasing) and non-monotonic (smoother) trends

<table>
<thead>
<tr>
<th>Detection method</th>
<th>Increasing trend</th>
<th>Decreasing trend</th>
<th>Smoother</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PLE/GC</td>
<td>51</td>
<td>55</td>
<td>43</td>
</tr>
<tr>
<td>SPLE/GC</td>
<td>12</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>LC</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BeadBeater/LC/ESI+</td>
<td>939</td>
<td>295</td>
<td>786</td>
</tr>
<tr>
<td>BeadBeater/LC/ESI-</td>
<td>832</td>
<td>229</td>
<td>598</td>
</tr>
</tbody>
</table>

However, the possibility that a compound may simultaneously exhibit a significant log-linear trend and a significant non-monotonic trend must be
taken into account. Eight and 32 compounds, respectively, exhibiting both monotonic and non-monotonic trends were detected in the PLE/GC and SPLE/GC datasets.

In the BeadBeater/LC/ESI+ and BeadBeater/LC/ESI- datasets, 199 and 115 compounds simultaneously exhibited a significant log-linear and a significant non-monotonic trend. Similar ratios of increasing to decreasing trends, with considerably more increasing trends, were found in both datasets. Moreover, a large number of non-monotonic trends was detected in both datasets.

**Time trends detected in the GC data**

Now we will further examine the types of compounds displaying trends detected using the log-linear regression and smoother analysis in the PLE/GC and SPLE/GC datasets. As shown in Figure 10, many hydrocarbons displaying trends (pale green) were detected in the SPLE/GC dataset, especially with the smoother function. Hydrocarbons (many biogenic) comprised more than half of the compounds exhibiting a non-monotonic trend, and several exhibited decreasing log-linear trends. In addition, PAHs (technically also hydrocarbons) constitute the second-largest group of compounds with a trend detected in the SPLE/GC dataset. Most compounds in this group displayed decreasing trends during the study period. These trends are consistent with the log-linear trends (i.e., decreasing between 2005 and 2015) detected in the PLE/GC dataset, where PAHs rank among the largest groups. PAHs are generated mainly by vehicular emissions. However, the number of cars using fossil fuels and the use of cars older than 15 years has increased (up to 2016) [65], so the decrease in PAH concentrations is puzzling. It may have resulted from technological advances. Moreover, the overall fuel consumption of vehicles may have decreased, and the performance of catalysts may have improved, leading to a decline in PAH emissions.

Another large group of compounds that characteristically displayed both log-linear and non-monotonic trends in the PLE/GC dataset are classified as “other compounds” and will be discussed later. The group of flavor and fragrances and (other) natural substances exhibiting log-linear trends is rather small, but these compounds were disproportionately abundant in the set of compounds with trends detected by the smoother function. Moreover, most of these compounds exhibited increasing (log-linear) trends. This may have resulted from an increase in the use of natural substances in personal care products and other articles of daily use, as reported previously [66].

The last two groups, alkylbenzenes and aldehydes and ketones, constitute a rather small number of the detected compounds with significant trends. Alkylbenzene emissions in urban areas are associated with traffic [67]. Hence,
the primarily decreasing trends for alkylbenzenes concur with the decreasing trends of PAHs.

**Figure 10** Percentages of detected compounds displaying increasing/decreasing log-linear and non-monotonic trends in the PLE/GC and SPLE/GC datasets.

**Time trends detected in the LC data**

The final evaluation of the BeadBeater/LC datasets is incomplete. So far, we have used two approaches to identify unknown compounds. Compounds displaying significant log-linear or non-monotonic trends in the ESI+ dataset have been tentatively identified using MS/MS library data. In addition, both
the ESI+ and ESI- datasets have been searched, using exact masses from compounds that exhibited significant changes in either the PLE/GC or SPLE/GC datasets. The few compounds with trends tentatively identified in the LC dataset are discussed in the following section.

**Time trends of selected compounds**

I have considered many compounds (see Table 4), but several of them remain unidentified, as discussed in the following paragraphs.

The compound exhibiting by far the most significant change between 2005 and 2015 is an unidentified compound (referred to as Unknown 11). This compound has been detected in the SPLE/GC dataset. Unknown 11 first appeared, at low concentrations, in samples from 2009, but its concentration increased considerably until 2013, without the occurrence of extreme values. Such an increase (yearly increase: 150%) is unusually high and may have resulted from evaporative losses or degradation of a parent compound during storage. The EI-spectrum of Unknown 11 indicates a low molecular weight (M+ at m/z 136) and, thus, evaporation during storage cannot be excluded.

*Table 4* Results of the time trend analysis performed on GC and LC data showing yearly increase or decrease of log-linear trends, significance (below 1% and 5% for two or one asterisk, respectively) of non-monotonic trends (smoother), and extreme values (extr. values). For compounds identified in several datasets, results obtained using each of the methods are separated by a slash.

<table>
<thead>
<tr>
<th>Compound</th>
<th>GC: PLE</th>
<th>GC: SPLE</th>
<th>LC: ESI+</th>
<th>LC: ESI-</th>
<th>Log-linear</th>
<th>Smoother</th>
<th>Extr. value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Natural substances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,6-Xanthine</td>
<td>2a</td>
<td>22%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Guanine</td>
<td>2a</td>
<td>46%</td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Thymine</td>
<td>2a</td>
<td>26%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Flavor and Fragrances</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-(1,3,4,4a,5,6,7-Hexahydro-2,5,5-trimethyl-2H-2,4a-ethanonaphthalen-8-yl) ethanone</td>
<td>x</td>
<td>4</td>
<td>-13%/-6%</td>
<td>2005/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Similar to Tonalid</td>
<td>x</td>
<td>4</td>
<td>4%/-</td>
<td>-*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galaxolide isomer</td>
<td>x</td>
<td>4</td>
<td>6%/</td>
<td>-*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galaxolide isomer</td>
<td>x</td>
<td>4</td>
<td>4%/-</td>
<td>-*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Galaxolide isomer</td>
<td>x</td>
<td>4</td>
<td>4%/</td>
<td>-*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-Methyl-5-hepten-2-one</td>
<td>x</td>
<td>4</td>
<td>-24%</td>
<td>*/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plasticizers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tri(2-butoxyethyl) phosphate (TBEP)</td>
<td>x</td>
<td>1</td>
<td>-13%/-7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Triphenyl thiophosphate</td>
<td>4</td>
<td>7%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diethylphthalate</td>
<td>x</td>
<td>4%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown Phthalate 1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Unknown Phthalate 2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>Bis(2-ethylhexyl) phthalate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-Hexadecylnaphthalene</td>
<td>x</td>
<td>8%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Compound</td>
<td>GC: PLE</td>
<td>GC: SPLE</td>
<td>LC: ESI+</td>
<td>Log-Linear</td>
<td>Smoother</td>
<td>Extr. value</td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
<td>------------</td>
<td>----------</td>
<td>-------------</td>
<td></td>
</tr>
<tr>
<td><strong>PPCPs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-(3,4-Dichlorophenyl) tetralone</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8%</td>
<td></td>
</tr>
<tr>
<td>Homosalate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41%</td>
<td></td>
</tr>
<tr>
<td>Octocrylene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>19%</td>
<td></td>
</tr>
<tr>
<td>Clozapine</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5%</td>
<td></td>
</tr>
<tr>
<td>Atazanavir</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>2-Lauryl-p-cresol</td>
<td>x</td>
<td>4</td>
<td>19%/−</td>
<td>-/*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-(4-Methylbenzylidene)-camphor</td>
<td>2b</td>
<td></td>
<td>-17%</td>
<td></td>
<td></td>
<td>2005</td>
<td></td>
</tr>
<tr>
<td>Triclosan</td>
<td>x</td>
<td>4</td>
<td>-18%/−34%</td>
<td>-/*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7-Pentylbicyclo[4.1.0]heptane</td>
<td>x</td>
<td>x</td>
<td>-9%/−</td>
<td>-/*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Technical/Industrial chemicals</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2,5-Dichloroaniline</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30%</td>
<td></td>
</tr>
<tr>
<td>4-tert- Octylphenol</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>* 2013</td>
<td></td>
</tr>
<tr>
<td>Tetraethylene glycol (PEG-4)</td>
<td>2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>Hexaethylene glycol (PEG-6)</td>
<td>2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>**</td>
<td></td>
</tr>
<tr>
<td>4-Octyl-(4-octylphenyl)-benzenamine</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-6%</td>
<td></td>
</tr>
<tr>
<td>12-Hydroxystearic acid (Lexiol G21)</td>
<td>2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>Nonylphenol isomer 1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10%</td>
<td></td>
</tr>
<tr>
<td>Nonylphenol isomer 2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-18% *</td>
<td></td>
</tr>
<tr>
<td>Nonylphenol isomer 3</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-8%</td>
<td></td>
</tr>
<tr>
<td>Nonylphenol isomer 4</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10%</td>
<td></td>
</tr>
<tr>
<td>1-Methyl-3-nonylindane or 1-Methyl-4-octyl-1,2,3,4-tetrahydronaphthalene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td><strong>Process chemical</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetradecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14%</td>
<td></td>
</tr>
<tr>
<td>Tetradecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Pentadecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10% *</td>
<td></td>
</tr>
<tr>
<td>Pentadecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Pentadecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6% *</td>
<td></td>
</tr>
<tr>
<td>Pentadecyl phenyl ester carbonate</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>C1-Carbazole</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
</tr>
<tr>
<td><strong>Other compounds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14,14-Bis(2-methylene cyclopropyl)-13,15-dioxo-14-siladispiro[5.0.5.3]-pentadecane</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2% *</td>
<td></td>
</tr>
<tr>
<td>1-(3-Fluoropropyl)-4-(hexyloxy)-benzene</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-10%</td>
<td></td>
</tr>
<tr>
<td>2-(2,3-Dihydro-1H-inden-1-yl)-1-(4-{{[(E)-2-phenylvinyl]sulfonyl}-1-piperazinyl}ethanone</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-3%</td>
<td></td>
</tr>
<tr>
<td>2,6-Di-tert-butyl-4-ethylidene-cyclohexa-2,5-en-1-one</td>
<td>2b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-17% 2005</td>
<td></td>
</tr>
<tr>
<td><strong>Unidentified compounds</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 1</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15% *</td>
<td></td>
</tr>
<tr>
<td>Unknown 2</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
</tr>
<tr>
<td>Unknown 3</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-9%</td>
<td></td>
</tr>
<tr>
<td>Unknown 4</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7%</td>
<td></td>
</tr>
<tr>
<td>Unknown 5</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-17%</td>
<td></td>
</tr>
<tr>
<td>Unknown 6</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>7% 2009</td>
<td></td>
</tr>
<tr>
<td>Unknown 7</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21%</td>
<td></td>
</tr>
<tr>
<td>Compound</td>
<td>GC: PLE</td>
<td>GC: SPLE</td>
<td>LC: ESI+</td>
<td>Log-Linear</td>
<td>Smoother Extrinsic value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>---------</td>
<td>----------</td>
<td>----------</td>
<td>------------</td>
<td>--------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 8</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>24%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 9</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 10</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-8%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 11</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>150% *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 12</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 13</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-6%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 14</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-11%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 15</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-10% *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 16</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-6% **</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 17 (C28H48O2)</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td>-4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 18 (C13H18O3)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>17%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 19 (C10H36O3)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 20 (C20H26O5)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>40%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 21 (C5H4N4O)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 22 (C24H36O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 23 (C25H38O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>* 2010</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 24 (C12H18O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>13%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 25 (C24H31N4O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>38% *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 26 (C11H18N2O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>16%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 27 (C10H20N2O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 28 (C16H35NO)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>-20% *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 29 (C11H9NO3)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown 30 (C10H13N5O4)</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>11%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* These compounds are included in the representation in Figure 10 as well.

a GC: PLE, GC: SPLE, LC: ESI+, and LC: ESI- refer to the PLE/GC, SPLE/GC, BeadBeater/LC/ESI+ and BeadBeater/LC/ESI- methods, respectively.

b For GC trends compounds that were identified with a trend are marked with an “x”, for LC the confidence level (page 24) for identification is given.

Significant changes (yearly increases of 41% and 19%, respectively) were also observed for the UV-filters homosalate and octocrylene, which are used in sun screens. Interestingly, the number of sun-hours and the intensity of sunlight in Stockholm remained approximately constant during the covered period [68,69], so the use of UV-filters should have remained steady. However, the Swedish Medical Products Agency reported that the use of octocrylene and homosalate as UV-filters in cosmetic products increased between 2012 and 2016 [70,71], which may explain the increasing concentrations of these products in sewage sludge. Another UV-filter, 3-(4-methylbenzylidene)-camphor (4-MBC), with a trend was also detected, but its concentration decreased (by 18% per year) during the study period. However, the overall decrease in 4-MBC only accounts for part of the increase in homosalate and octocrylene. Data from the Swedish Medical Products Agency indicate that use of 4-MBC remained unchanged between 2012 and 2016 [70,71]. Notably, 4-MBC, homosalate, and octocrylene were detected in sewage sludge from the same STP in Stockholm in samples from 2009 and 2014 [72]. Those data concur with our results, which show that homosalate
and octocrylene levels increased from 2009 to 2014, whereas the 4-MBC levels decreased.

The group of PPCPs also includes pharmaceuticals or pharmaceutical impurities, for example, clozapine (an antipsychotic agent) and atazanavir (used to treat and/or prevent HIV and AIDS). Both compounds showed moderately strong trends. Clozapine exhibited a yearly increase of 5% and atazanavir a non-monotonic trend (Figure 11). Atazanavir prescriptions in Stockholm increased from 2006 to 2011 and decreased thereafter [73], which is consistent with the trends observed in sewage sludge (Figure 11). Similarly, clozapine prescriptions increased slowly, but constantly, between 2006 and 2015 [73]. Both compounds have been previously detected in wastewater from Stockholm [74], and sewage sludge (see [75] and Paper I). Likewise, a yearly increase of 8% was observed for 4-(3,4-dichlorophenyl) tetralone (Figure 11 and Paper I), which had not previously been reported in environmental samples. This compound is considered a potential impurity of sertraline [76], a pharmaceutical used as an anti-depressant. However, sertraline was not detected in our samples, possibly due to its metabolization, concentrations lower than its detection limits, or partitioning to the water phase. The increase in 4-(3,4-dichlorophenyl)tetralone levels can, nevertheless, be attributed to an increase in the number of sertraline prescriptions in Stockholm between 2006 and 2015 [73].

![Figure 11](image-url)  
*Figure 11 Time trends of Atazanavir, 4-(3,4-dichlorophenyl) tetralone and 4-tert-octylphenol (extreme value indicated by the red circle).*

I also considered disinfectants and compounds related to dyes and pigments. The concentration of triclosan (a disinfectant) decreased
significantly during the study period, consistent with previously reported trends for triclosan in sewage sludge collected across Sweden [77].

Extreme values in the first and last year, 2005 and 2015, respectively, may have influenced the significance and degree of log-linear trends. However, extreme values among the other years may indicate outliers and change points, as suggested by significant non-monotonic trends. Such a trend was observed for 4-tert-octylphenol, which increased until 2013 and then decreased again, as shown in Figure 11. Octylphenol is an intermediate in the production of phenolic resins, which are used to produce rubber, inks, and surfactants (octylphenol ethoxylates). In 2011, octylphenol was added to the list of compounds regulated by REACH, the EU legislation for chemicals, and classified as a Substance of Very High Concern. Hence, companies searching for replacements (in anticipation of its prohibition) might explain the decrease in octylphenol use after 2013. The use of octylphenol and octylphenol ethoxylates in Sweden decreased from 2005 to 2014 and is inconsistent with the trend we observed [78]. Moreover, 4-tert-octylphenol was previously detected in the study presented in Paper I.

**Comparison of parametric and non-parametric trend tests**

As I explained in the Introduction, non-parametric tests are employed for cases involving non-normal distributions of data. Here, I compare monotonic trends revealed by parametric (log-linear regression) and non-parametric (Mann-Kendall test) tests. This comparison focuses on the LC datasets, which are considered less stable than the GC data, as discussed above.

**Table 5** Number of significant trends in the BeadBeater/LC/ESI+ (ESI+) and BeadBeater/LC/ESI- (ESI-) datasets revealed by log-linear regression analysis (LR) and Mann-Kendall (MK) tests.

<table>
<thead>
<tr>
<th></th>
<th>Total numbers of compounds</th>
<th>Total numbers of significant trends</th>
<th>Numbers of significant trends only identified by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with significant trends</td>
<td>revealed by</td>
<td>LR</td>
</tr>
<tr>
<td>ESI+</td>
<td>1372</td>
<td>1234</td>
<td>248</td>
</tr>
<tr>
<td>ESI-</td>
<td>1262</td>
<td>1061</td>
<td>346</td>
</tr>
</tbody>
</table>

In Table 5, I compare numbers of compounds with trends revealed by LR and MK tests in the two datasets. As the table shows, these analyses revealed similar overall numbers of compounds with significant trends. However, despite the similarity in numbers, some compounds with significant trends were only detected by the LR tests, while others were only detected by MK tests. In the ESI+ and ESI- datasets, 138 and 201 compounds, respectively,
showed significant trends according to MK tests, but not LR tests. This may have resulted from non-normal distribution of residuals, as LR is a parametric test, and thus requires a normal distribution of the residuals. The differences in LR and MK results may also be partly due to effects of the extreme values in the middle of the timeline on the tests (for example, see the right-hand side of Figure 12). Moreover, trends that are significant according to LR-tests, but not MK tests, are probably influenced by extreme values in the first or last year (for example, see left-hand side of Figure 12). Hence, evaluating compounds for which the two tests yield different results is essential for ensuring reliable results. Compounds for which both tests yield similar results are considered “safe” within the given uncertainties.

**Figure 12** Trends that were significant according to log-linear regression (LR) analysis, but insignificant according to the Mann-Kendall (MK) tests (left side) and vice versa (right side). The regression line in the right graph was obtained after exclusion of the extreme value (indicated by red circle).

**Conclusions and improvements**

In this section, I have presented a method for ranking compounds that comprise the large group of unknown compounds that lie hidden under water in the bottom of the metaphorical iceberg. I have also presented time-trend data that reveal increasing as well as decreasing concentrations of various compounds. Legislation appears to be lacking for compounds characterized by increasing trends (for example, PPCP chemicals). Therefore, these compounds require further attention from scientists and policy-makers.
We have obtained results from both GC methods (PLE and SPLE) and the LC methods (BeadBeating followed by LC in ESI+ and ESI- modes). Unfortunately, many compounds, detected predominantly via the LC approach, remain unidentified. The identification of those compounds and a good means of handling the large amount of data will be considered in future work.
A new retention index system for GC×GC using polyethylene glycols

As mentioned in previous sections of the thesis, chromatographic retention time information can be used in the process of identifying compounds. Linear retention indices (LRIs) are widely used in GC analyses during compound identification. These indices are based on retention times of reference compounds (alkanes of different sizes) that have different elution times [79]. If, for example, \(n\)-decane, \(n\)-undecane, and \(n\)-dodecane elute after 7, 10, and 13.5 min, respectively, retention times of other compounds (for example, limonene) can be defined in relation to these times, thereby yielding retention indices (RIs). By definition, \(n\)-decane and \(n\)-undecane have LRIs of 1000 and 1100, respectively. If limonene would elute midway between these compounds (after 8.5 min), it would have an LRI of 1050. Indices, such as LRIs, are already used in GC analysis. However, no RIs are widely used as yet for the second-dimension retention time in GC×GC analyses. Therefore, in this section, I present a new method of calculating such indices.

Development of retention indices and GC×GC

For many years, RIs have been used in GC analyses in the course of identifying unknown compounds. The relationship between the retention time of an analyte and retention times of \(n\)-alkanes, formulated by Ervin Kováts in 1958 [79], provides foundations for the use of RIs in isothermal one-dimensional GC. In 1963, van den Dool and Kratz [80] presented the first LRIs for temperature-programmed GC (see also Figure 13). In subsequent work, Lee et al. [81] used PAHs, rather than \(n\)-alkanes, as retention markers for indexing polycyclic aromatic compounds (PACs).

Two-dimensional GC (i.e., GC×GC) was first developed by Phillips and Liu in 1991 [82]. This technique increases the peak capacity of chromatograms and allows the separation of complex mixtures [83]. The abovementioned RIs are applicable to first-dimension elution in GC×GC, but a separate system is required for the second dimension. Several attempts have been made to develop such a system. In the first attempt, Beens et al. [83] included the creation of so-called isovolatility curves through the continuous injection of reference compounds (in this case, \(n\)-alkanes). This procedure, which allowed 2D mapping of the analytes in GC×GC, was applied in several cases using \(n\)-alkanes [84,85], fatty acid methyl esters (FAMEs) [86] or PAHs [87] as reference compounds. However, this procedure suffered from drawbacks, such as complexity, difficulty of application, validity within only a given range of boiling points [83,88], and low precision [89]. Although several improvements have been proposed [86,88,90], application of this method
remains difficult. More theoretical approaches (than previous methods) have been proposed by Arey et al. [91] and Dorman et al. [92] who calculated RIs based on stationary – gas phase equilibrium partition coefficients and thermodynamic RIs, respectively.

[Image: Milestones in development of RIs in GC and GC×GC analyses.]

Figure 13 Milestones in development of RIs in GC and GC×GC analyses.

A new, easily applied and robust RI system for GC×GC analysis is presented in the following text. In this system, polyethylene glycols (PEGs) and \( n \)-alkanes are used as reference compounds. The resulting 2D RIs (\( 2I \)), PEG-\( ^{2}I \) values, can be obtained using a standard column configuration (an apolar × polar column set [14]) by applying three simple steps. The work described in the following text was published in Paper III.

**PEG-\( ^{2}I \) retention index development and tests**

The aim of this study, as described in Paper III, was to develop and test a new RI system for the second dimension in GC×GC.

To calculate the indices, I used a range of PEGs, \( n \)-alkanes, and the 8270 MegaMix (76 compounds). The PEGs and \( n \)-alkanes are needed for the calculations while the MegaMix compounds were used to test the indices. All experiments were performed using an apolar × polar column set consisting of a 5% phenyl column and a 50% phenyl column for the first and second dimension separations, respectively.

The robustness of the new RI system was evaluated in two ways. First, effects of several method parameters (carrier gas flow, oven ramping rate, and secondary oven offset) on the calculated PEG-\( ^{2}I \) values were examined. In addition, results obtained with three second-dimension columns (with 0.1, 0.18, and 0.25 mm internal diameters) were compared. The method
parameters were varied in accordance with a central composite face-centered (CCF; Figure 14) experimental design, with the center point repeated five times to determine the repeatability. The experiments performed with columns of different dimensions were conducted with settings based on the center point of the CCF design.

![Diagram of the central composite face-centered design and the experiments that were performed](image)

**Figure 14** Illustration of the central composite face-centered design and the experiments that were performed (red, middle of the cube; green, middle of each face). Each dot represents an experimental setup.

**Theory and calculations**

Analysis of the GC×GC data revealed that PEGs were spread almost equidistantly across the 2D space (see, for example, **Figure 15**). This results from the incremental increase in both the first- and second-dimension retention times associated with a PEG oligomer unit (CH₂CH₂O). In our system, a value of 10 is assigned to an oligomer unit. The PEGs considered in this study have between two and 10 oligomer units (PEG-2 and PEG-10, assigned PEG⁻²⁻I values of 50 and 130, respectively. In addition, ethylene glycol (EG) is assigned a PEG⁻²⁻I value of 20 (see **Table 6**).

The reference point of the PEG⁻²⁻I system is the alkane band, i.e., all n-alkanes have (by definition) a PEG⁻²⁻I of zero. Thus, retention times of the other compounds must all be “shifted” with respect to the second-dimension retention times of n-alkanes, as follows:
The first equation determines the second-dimension “position” of the alkane band at the first-dimension elution time of the analyte, as illustrated for azobenzene in Figure 15. The “excess retention time” (\(2t_{R,E}\)), i.e., the temporal distance of an analyte from the alkane band in the second dimension, is calculated using the second equation. Furthermore, the PEG–\(2I\) is determined from \(2t_{R,E}\) using a linear regression function for the PEGs between PEG–\(2I\) and \(2t_{R,E}\). This function can then be used to determine the PEG–\(2I\) value of an analyte from the corresponding \(2t_{R,E}\). Although the calculation may seem rather complicated, I would like to show that this is not the case by describing an example.

**Step 1:** The \(2t_{R,alkane}\) and \(2t_{R,E}\) for our model analyte, azobenzene are calculated using the two equations above. The second-dimension retention time (\(2t_R\)) of the alkane eluting before the analyte (in the first dimension), i.e., \(2t_{R,alkane}\), is required for this calculation. The alkane eluting before azobenzene is hexadecane (\(2t_R: 2.07\), Table 6). The next term includes the first- and second-dimension retention times (\(t_R\) and \(2t_R\)) for the alkanes eluting before and after azobenzene (\(\alpha_1\) and \(\alpha_2\); hexadecane and heptadecane, respectively).
In addition, the $t_R$ of the analyte is required (1712). Inserting all these values into the equation yields:

$$2 \ t_{R,alkane}(azobenzene) = 2.07 + \frac{2.11 - 2.07}{1808 - 1676} \times (1712 - 1676) = 2.08$$

**Step 2:** The excess retention time $t_{R,E}$ is calculated using the $t_R$ of azobenzene (4.53) and the calculated retention time of azobenzene projected onto the alkane band (2.08):

$$2 \ t_{R,E} = 4.53 - 2.08 = 2.45$$

**Step 3:** As explained above, a linear relationship between $t_{R,E}$ and PEG-$2I$ values of the PEGs is required to calculate PEG-$2I$ values for the analytes. Figure 16 shows an example based on data presented in Table 6. The PEG-$2I$ value of azobenzene can be calculated from the slope of the line:

$$PEG^{-2I}(azobenzene) = slope \times 2 \ t_{R,E} = 29.33 \times 2.45 = 71.9$$

Once calculated for an entire dataset, the slope can then be used to calculate PEG-$2I$ values of all the analytes.

**Table 6** Retention time data required for calculating the PEG-$2I$ value for azobenzene (indicated with three question marks in the table).

<table>
<thead>
<tr>
<th>Compound</th>
<th>$t_R$ (s)</th>
<th>$t_R$ (s)</th>
<th>$t_{R,E}$</th>
<th>PEG-$2I$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Azobenzene</td>
<td>1712</td>
<td>4.53</td>
<td>2.45</td>
<td>???</td>
</tr>
<tr>
<td>Hexadecane</td>
<td>1676</td>
<td>2.07</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Heptadecane</td>
<td>1808</td>
<td>2.11</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Ethylene glycol</td>
<td>230</td>
<td>2.38</td>
<td>0.89</td>
<td>20</td>
</tr>
<tr>
<td>Diethylene glycol</td>
<td>614</td>
<td>3.81</td>
<td>2.00</td>
<td>50</td>
</tr>
<tr>
<td>Triethylene glycol</td>
<td>1082</td>
<td>4.40</td>
<td>2.43</td>
<td>60</td>
</tr>
<tr>
<td>Tetraethylene glycol</td>
<td>1514</td>
<td>4.81</td>
<td>2.76</td>
<td>70</td>
</tr>
<tr>
<td>Pentaethylene glycol</td>
<td>1898</td>
<td>5.06</td>
<td>2.95</td>
<td>80</td>
</tr>
<tr>
<td>Hexaethylene glycol</td>
<td>2234</td>
<td>5.32</td>
<td>3.14</td>
<td>90</td>
</tr>
<tr>
<td>Heptaethylene glycol</td>
<td>2534</td>
<td>5.51</td>
<td>3.26</td>
<td>100</td>
</tr>
<tr>
<td>Octaethylene glycol</td>
<td>2792</td>
<td>5.84</td>
<td>3.53</td>
<td>110</td>
</tr>
<tr>
<td>Nonaeethylene glycol</td>
<td>3032</td>
<td>6.20</td>
<td>3.80</td>
<td>120</td>
</tr>
<tr>
<td>Decaethylene glycol</td>
<td>3248</td>
<td>6.62</td>
<td>4.13</td>
<td>130</td>
</tr>
</tbody>
</table>
A key question I needed to answer is: How robust is the system we developed? As previously mentioned, I calculated PEG-$2I$ values associated with different conditions and secondary columns of different dimensions. The robustness was then evaluated by first determining the repeatability, by examining results for the central experimental point, which was repeated five times. Values of 0.7% and 1.8% were obtained for the average and maximum relative standard deviations, respectively, indicative of very good repeatability.

I then evaluated results of the experiments conducted in accordance with the CCF design (see Figure 14, and Figure 17 for the corresponding results). In the latter figure, the relative standard deviation for each compound is shown in relation to the corresponding PEG-$2I$ and LRI values. The LRI presented here is calculated via the method proposed by van den Dool and Kratz [80]:

$$LRI = 100 \times \left( n_{\alpha 1} + \frac{t_{R,analyte} - t_{R,\alpha 1}}{t_{R,\alpha 2} - t_{R,\alpha 1}} \right)$$

Here: $t_{R,analyte}$ is retention time of the analyte; $t_{R,\alpha 1}$ and $t_{R,\alpha 2}$ are retention times of the $n$-alkanes eluting before and after the analyte, respectively; and $n_{\alpha 1}$ is the number of carbon atoms in $n$-alkane $\alpha 1$.

As Figure 17 shows, the deviations among runs with different settings were largest for compounds that eluted late in the second dimension. Moderate RSD values were obtained for compounds that eluted late in the first dimension, but not the second dimension. The large variations among compounds that eluted late in the second dimension can be explained as

**Figure 16** Calibration graph for determining PEG-$2I$ values from corrected $t_{R,E}$ values.

**Robustness and application range of the new system**
follows. By definition, PEG-10 has a PEG-\(2I\) of 130. Large deviations were observed for late-eluting compounds with PEG-\(2I\) values >130, hence, PEG-\(2I\) values are extrapolated for those compounds. For practical purposes, confidence intervals provide better indications than RSD values of the exactitude of the RI system. Average 95% confidence interval values of 1.3 PEG-\(2I\) units and 0.8 PEG-\(2I\) units were obtained with and without the extrapolated compounds, respectively. This corresponds to a 95% probability that the errors are \(\leq 0.8\) PEG-\(2I\) units (on average), indicating that our proposed RI system is robust toward changes in instrument settings.

**Figure 17** Deviations (RSD) for all compounds in relation to their first- and second-dimension elution times (LRI and PEG-\(2I\) values, respectively).

The experiments described above were conducted using a secondary column with an inner diameter of 0.18 mm. When using a column with an inner diameter of 0.25 mm in the second dimension, calculated PEG-\(2I\) values deviated on average by 1.7% from those obtained using the 0.18 mm column, both with and without the extrapolated compounds.
Similarly, use of a secondary column with a smaller inner diameter (0.1 mm) resulted in average deviations of 2.7 and 1.9% from those obtained using the 0.18 mm column, with and without extrapolated compounds, respectively. A constant flow mode was used in all experiments, so the instrument software constantly adjusted the head pressure at the inlet to maintain the set flow. Calculating pressures within the equipment is complex, especially when two columns with different diameters are coupled. In addition, viscosity effects in the transfer line can influence the precision of these calculations. Due to the very high temperatures in the transfer line, the mobile-phase gas in the line will be more viscous than in other regions, resulting in additional back pressure. This back pressure will increase with reductions in column diameter, thus variations will be larger for narrow columns than for wider columns. However, other factors may have also contributed to these column-related variations.

Conclusions and improvements

Overall, the proposed new RI system seems robust against changes in GC settings and moderate changes in secondary-column diameter. In the corresponding paper (Paper III), we present suggestions for improving the system, which suffers from the following drawbacks: limited range of PEGs, difficulty in analyzing the PEGs (which are very polar), and high volatility of EG (leading to premature EG elution from non-polar columns). These drawbacks can be overcome by increasing the number of PEGs considered, thereby avoiding extrapolation to the PEG-\(2I\) values and improving the exactitude of the system. Alternatively, the PEGs can be replaced with glyme compounds, which are structurally similar to PEGs: the hydrogen atom of the alcohol group at both ends of the PEGs is simply replaced by methyl groups. Hence, glymes have lower polarity and are easier to analyze than PEGs, especially when using aging columns (see Table 6 in Paper III for PEG-\(2I\) values of glymes). Another option is to replace EG with 1-octanol. The LRI of EG is 699. When solvents with high boiling points (e.g., toluene) are used, EG will elute in the solvent peak, and hence be impossible to analyze. However, with an LRI of 1070 and a PEG-\(2I\) of 17, 1-octanol is more easily analyzed than EG (assigned PEG-\(2I\): 20) and is a good substitute for it.
Retention time prediction

Different compounds have different properties and therefore differing retention times that, combined with the corresponding RIs, can be used in the identification of unknown compounds by excluding possible structures from a list of candidates. However, the theoretical retention time or RI of a candidate structure is sometimes unknown, so comparison with our measured values is prevented. Therefore, different computational models (*in-silico* tools) for predicting the retention times and RIs are presented in this last section of Part II. The retention times and RIs of almost 900 compounds have been measured and other properties (for example, the structure and molecular weight) of the compounds have been generated. The computational models then determine relationships (if any) between these attributes and the measured retention times and RIs. These relationships (equations) can be subsequently used to calculate the theoretical retention times and RIs of new compounds.

Development of a tool for predicting the retention times and RIs in both GC×GC dimensions was the goal of the study presented in Paper IV. In non-target screening, the acquired mass spectra and retention times are used to identify compounds. The prediction of retention times or RIs in GC×GC provides an additional qualification point (in addition to that provided by GC), which helps in the identification of compounds and selection of probable structures from numerous possibilities. Using information from two retention characteristics will therefore help to reduce numbers of false positives obtained in the identification of unknown compounds.

Retention-time prediction approaches

Retention-time prediction is a powerful tool that is being used in proteomics [93,94], metabolomics [95] and analyses of organic pollutants using GC-MS or LC-MS techniques [95,96]. It can be applied to compounds with varying molecular weights, polarities, and boiling points [96]. Numerous approaches for predicting retention-times have already been applied. For example, Creek et al. [95] and Garkani-Nejad [97] used quantitative structure–retention relationships (QSRRs) and quantitative structure–property relationships (QSPR), respectively, to derive analytes’ retention times or RIs from their structures. The main difference between QSRRs and QSPRs is that in QSRRs the retention is defined as output, whereas in QSPRs (generally) “properties” are modeled. However, one modelled property may be retention in a chromatographic column, so QSRRs comprise a special type of QSPR. In both studies, so-called molecular descriptors (where the structure and properties of a molecule are denoted by numbers) were used to predict
analytes’ retention times from their structures. Simple examples of these descriptors include the number of double bonds, numbers of chlorine or bromine atoms, and logarithms of the octanol-water partition coefficient (logKow or logP).

**Quantitative structure-retention relationships**

QSRRs have been used to identify compounds of specific classes (for example, peptides and steroids [98,99]) via retention-time prediction. The QSRR method involves use of molecular descriptors that are directly derived from the structure of each compound. Applying this method can help to identify the structural factors that play important roles in the retention characteristics of compounds [100,101]. For example, QSRRs of reversed phase liquid chromatography parameters are relatively simple compared to those associated with other techniques, because the main influencing factors (size and polarity of the analytes) have been extensively studied [98]. In GC, for instance, use of a column with a non-polar stationary phase yields an almost linear relation between compounds’ retention times and boiling points [102].

Essentially, QSRRs are statistical models, i.e. equations, to link the physico-chemical and structural properties of analytes to their retention behavior, using data for compounds with properties similar to those of the analytes [100,103]. Multivariate methods, such as multiple linear regression, artificial neural networks and support vector machines, are widely used for constructing the models [101].

**Model types**

*Partial Least Squares*

The first approach I used included Partial Least Squares Projections to Latent Structures or (in short) Partial Least Squares (PLS), which is a regression analysis method that relates many X-variables (e.g., molecular descriptors) to one or more Y-variables (responses, e.g., retention time). The development of PLS started in the 1960’s and 70’s and was driven by Herman Wold [104,105]. This method is aimed at extracting latent variables that account for maximum variability while adjusting those in direction of the response [106]. The latent variables are also called components. The number of significant components must be determined during model optimization, and the model fit improves with increasing number of components. However, beyond a certain number of components, further increases only result in “over-fitting” by modelling noise. An over-fitted model may be quite suitable for the set used to build it, but will have poor predictive power [107]. The predictive power and fit of a model may be evaluated by cross-validation. Here, the dataset is randomly divided into fractions, for example seven
sub-sets. Each sub-set is excluded once from the model-building process and then predicted using the model. The deviation of the observed values from the predicted values ($y_{\text{obs}}$ and $y_{\text{pred}}$, respectively) is then calculated and used as a measure of predictive power. This process is repeated iteratively for each sub-set, thus the root-mean-square error of cross validation ($\text{RMSE}_{\text{CV}}$) can be calculated. The predictive power can also be evaluated by calculating the root-mean-square error of prediction, as follows:

$$\text{RMSEP} = \sqrt{\frac{1}{N-1} \sum (y_{\text{obs}} - y_{\text{pred}})^2}$$

Here: $N$ is the number of data points and $y_{\text{obs}}$ and $y_{\text{pred}}$ are the experimental (observed) and predicted values (in this case retention times and RIs), respectively. The $y_{\text{obs}}$ and $y_{\text{pred}}$ for the RMSEP are obtained by predicting compounds from a separate dataset (test set) that was excluded from the model building.

“Federation of local models”

The other approach I used included the ChromGenius software package (ACD/Labs). In this software package, some physico-chemical properties (e.g., boiling point or logP) are calculated for each compound within the training set. Data for these compounds comprise the so-called knowledge base. When new compounds are imported for the prediction, the software applies an algorithm to search the knowledge base for compounds that are similar (in terms of both structure and physico-chemical properties) to each new compound. The search is performed by generating similarity coefficients, i.e., numerical measures of similarity between two compounds. These coefficients can be calculated in several ways [108]. The so-called dice coefficient describing the similarity of two compounds (A and B) is calculated as follows:

$$S_{A,B} = \frac{2 \sum_{j=1}^{n} x_{jA} x_{jB}}{\left(\sum_{j=1}^{n} (x_{jA})^2 + \sum_{j=1}^{n} (x_{jB})^2\right)}$$

The $x$ vectors include $n$ physico-chemical properties that the calculations are based upon. Another metric that is often used for chemical-similarity searching is the Euclidean distance. The Euclidean distance between two compounds, A and B, is calculated as:

$$D_{A,B} = \sqrt{\sum_{j=1}^{n} (x_{jA} - x_{jB})^2}$$

As in the dice coefficient equation, the $x$ vector includes the physico-chemical properties. The compounds from the database (knowledge
base) that are closest (i.e., for which $S_{A,B}$ is highest or $D_{A,B}$ is smallest) are then used to predict, via multiple linear regression (MLR), the retention time of the new compound [109]. Since local models are calculated using structurally similar compounds, this approach is referred to as the “Federation of local models” [110,111]. The ChromGenius software package, has been previously used to predict GC and LC retention times [109,112]. In addition to the basic ChromGenius model, an add-on (Absolv) is available that includes Abraham solvation parameters [113] in the modeling.

**Modelling workflow**

Most modeling approaches involve three basic steps that require three sets of data: training of the model, using a training set, optimization of the model, using a test set (optimization of model parameters to maximize predictive power), and validation of the model, by determining the error of prediction of an external validation set. In some cases, especially when insufficient data points are available, the dataset is only divided into two sub-sets: a training set and a test set. The training set is used to build the model and the test set is used to optimize and validate the model.

To obtain a training set, test set and external validation set, I divided the whole dataset systematically into three groups. The training set, external validation set, and test set contained the highest number, second-highest number, and fewest compounds, respectively. Before dividing the data, I applied principal component analysis (PCA) [105,114] using all the data points and molecular descriptors. PCA identifies patterns in the data (here the molecular descriptors) by finding a line (or plane or hyperplane) that best approximates the data. The number of dimensions is thereby reduced, and the data are projected onto the new coordinate system.

The scores (“principle component values”) and loadings assigned by the PCA represent the projections of observations, i.e. the compounds, and the projections of variables, i.e. the descriptors, respectively. Similar compounds will have similar scores. To divide the data into three sets, the scores for each compound are sorted in ascending order, i.e., based on increasing value of the first component. Every fifth, eight, and ninth value is then assigned to the training set, test set, and external validation set, respectively. The remaining data are then sorted with respect to the second-component scores and the procedure is repeated. These steps are performed for all components. However, other methods can be used to divide the data sets. In any case, after dividing the dataset into two or three sub-groups, PCA (unsupervised) and PLS (supervised) score plots should be generated. These plots reveal the goodness of the data division. For example, inadequate division is achieved if some parts lie outside the test set. In contrast, good division is achieved if the test set covers the same space as the training set.
Retention time prediction in GC×GC

Before the work presented in Paper III, widely applicable RIs for the second dimension of GC×GC were lacking. The approach presented in this chapter is based on the RIs developed in Paper III and includes the prediction of retention times and RIs for the first and second dimension of GC×GC. Some previous attempts have been made to predict GC×GC separations. Notably, Beens et al. [83] and Vendeuvre et al. [89] used LRIs, alone and in combination with vapor pressures, to predict GC×GC chromatograms, while Seeley and Seeley [115] created so-called retention diagrams showing the relative retention of analytes in GC×GC, and used one-dimensional retention times to predict 2D-chromatograms. These approaches are helpful when selecting stationary-phase combinations for a given set of target analytes. However, in temperature-programmed GC the first-dimension separation depends primarily on the analytes vapor pressure, and the second-dimension retention depends on both analyte-stationary phase interactions and the elution temperature of the analyte, which is related to its LRI [115]. Thus, predicting analytes’ behavior is more complex in the second-dimension separation than in the first-dimension separation.

Data generation and modelling procedure

The aim of this study was to create four models: one each for retention times in the first and second dimensions, one for the first-dimension LRIs, and one for second-dimension PEG−2I values. Furthermore, I wanted to compare the predictive power associated with the retention times and respective indices to determine which (retention time or RI) models yielded the best results. This was achieved by subjecting 859 compounds to GC×GC-MS analysis with an apolar×(semi-)polar column set (the most commonly used column set [14]). Diverse types of compounds—e.g., pesticides, OPs, FAMEs, PAHs, dioxins and furans, bisphenols, polybrominated diphenyl ethers (PBDEs), and all 209 PCBs—were included in the analysis. These compounds were divided into three sets: a training set, a test set, and an external validation set. Molecular descriptors of each compound were then calculated using Molecular Operating Environment (MOE) software (104 2D physico-chemical descriptors) [116] and the Percepta module of PhysChem Suite 2014 (ACD/Labs) [117] (29 descriptors). To compensate for the size dependency of some properties (e.g., lipophilicity; logKOW), each descriptor was subsequently normalized to the weight of the compound. Each descriptor was then manually checked to determine if (compared with the original data) transformation of the data would improve the linear relationship between descriptors and responses (i.e., retention times and RIs). Consequently, twenty and three transformed descriptors were added to the first- and second-dimension models, respectively.
We used two software packages to model the retention times and RIs: SIMCA [118] and ChromGenius [119]. The calculated descriptors and determined retention times as well as RIs for the compounds from the training and test sets were imported to SIMCA, and four separate PLS models were created for retention times and RIs. In addition, the data were scaled to unit variance and mean-centered. Effects of various factors on the model quality, i.e., the predictive power and prediction error, were then determined. These factors included the:

1. Number of PLS components (A)
2. Stepwise increase in the number of descriptors
3. Automatic transformation of variables
4. Removal of all molecular descriptors with high uncertainties
5. Removal of all molecular descriptors of low importance
6. Stepwise removal of descriptors with high uncertainties and low importance
7. Development of separate models for different groups of compounds

In ChromGenius I varied parameters that influence the search for similar compounds, which the models would be based upon, building-block parameters (e.g. physico-chemical properties) of the models, and numbers of compounds used to derive the equation for the MLR. The following parameters were investigated for each model:

1. Two sets of similarity and distance coefficients (Dice coefficients and Euclidian distances) for the similarity search
2. Number of similar compounds used for the predictions
3. Exclusion of Abraham parameters
4. Exclusion of all parameters, except the Abraham parameters
5. Number of compounds per parameter included in the equation for the retention-time calculation

During the prediction of each retention time (or RI) associated with each parameter, a certain number of compounds from the sub-set of the knowledge base is used to estimate the parameter for the compound of interest. The number of compounds is determined in step 5.

Results obtained using the PLS and ChromGenius package are compared. In addition, the average of each set of results is calculated and compared with the experimental values to see if this leads to any improvement.

“Quality assurance”

I also developed simple linear models using only a single basic descriptor for each dimension. The first-dimension separation in GC×GC is (generally) based on differences in boiling points when an apolar column is used in the first dimension. Therefore, the boiling point was taken as the basic descriptor.
for the first dimension. The separation on a polar second-dimension stationary phase is driven by differences in the compounds’ polarity. However, the temperature at which an analyte enters the second column is determined by the first-dimension separation. GC×GC systems are generally tuned such that the boiling-point dependence in the second dimension is canceled. Unfortunately, many parameters describing the polarity (for example, logK_{OW}) are size dependent. Therefore, the logK_{OW} values and logK_{OW} values normalized to the weight of a compound were separately considered. The ability of these simple models to predict the retention times and RIs was compared with that of the previously developed models. If such models exhibit good predictive power, the PLS and ChromGenius models would be considered overly complex, and one could use the simple one-parameter relationships instead.

**PLS model results**

**Number of PLS components**

Figure 18 shows the dependence of the predictive power (cumulative Q²), cross-validation error (RMSE_{CV}), and prediction error (RMSE_\text{P}) of the model constructed from the test set on the number of components. The Q² and RMSE_{CV} exhibit the expected behavior. These values improve up to a certain point (here seven components for each model) and deteriorate thereafter, due to overfitting.

The continuously decreasing RMSE_\text{P} values are unexpected, because over-fitted models are expected to have relatively high prediction errors. However, for the models developed in this work, the prediction error continuously decreased with increasing numbers of components. This suggests that a model with 30 or 40 components (depending on the retention time or RI modeled) would yield the best result. However, all four models only predict one variable. In theory, if the variables and response are linearly related one component should be sufficient for modeling the relationship. However, Figure 18 shows that use of a single component will yield rather high errors. The maximum Q² and minimum RMSE_{CV} were obtained with seven components, so a seven-component model was used in further analyses. The optimal number of components was assessed after each step, as outlined above.

**Influence of descriptors and transformation**

In addition to the number of components included in each model, effects of several other factors were assessed. The first of these was increasing the number of descriptors. The original models solely included MOE descriptors. To improve the predictive power of the models, additional descriptors from the Percepta module of ACD/Labs were included. The descriptors that could
be transformed to resemble a linear relationship were subsequently identified, all descriptors were divided by the analytes’ molecular weight, and the resulting values were included as additional descriptors.

Table 7 shows the resulting RMSE_P for each model. The models with the best predictive power for the first dimension included all MOE, Percepta, and manually transformed descriptors. Using all MOE, Percepta and manually transformed descriptors, as well as all descriptors normalized to molecular weight, yielded the best models for the second dimension. Separation in the second dimension is driven by differences in polarity and polarizability. The polarizability is size dependent, so results are improved by including descriptors normalized to molecular weight. Auto-transforming all variables, i.e., letting the SIMCA software choose which parameters should be
transformed, as well as removing descriptors with high uncertainties and/or low importance, yielded no improvement.

Table 7 Lowest and highest values of each response and RMSEP following each developmental step of each PLS (SIMCA) model. Numbers of PLS components are shown in parentheses.

<table>
<thead>
<tr>
<th>Steps for model improvement</th>
<th>( t_R (s) )</th>
<th>LRI</th>
<th>2( t_R (s) )</th>
<th>PEG–I</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lowest measured value</td>
<td>270</td>
<td>808</td>
<td>1.68</td>
<td>0</td>
</tr>
<tr>
<td>Highest measured value</td>
<td>3325</td>
<td>3413</td>
<td>6.62</td>
<td>215.1</td>
</tr>
<tr>
<td>MOE descriptors only</td>
<td>142 (6)</td>
<td>116 (8)</td>
<td>0.37 (8)</td>
<td>15.3 (8)</td>
</tr>
<tr>
<td>MOE and Percepta descriptors</td>
<td>118 (7)</td>
<td>104 (6)</td>
<td>0.33 (6)</td>
<td>13.2 (7)</td>
</tr>
<tr>
<td>MOE, Percepta, and manually transformed descriptors</td>
<td>112 (6)</td>
<td>101 (7)</td>
<td>0.33 (6)</td>
<td>13.3 (7)</td>
</tr>
<tr>
<td>MOE, Percepta, manually transformed, and normalized-to-weight descriptors</td>
<td>119 (7)</td>
<td>106 (7)</td>
<td>0.30 (7)</td>
<td>12.5 (8)</td>
</tr>
<tr>
<td>Autotransform all variables</td>
<td>114 (9)</td>
<td>107 (8)</td>
<td>0.32 (7)</td>
<td>20.2 (8)</td>
</tr>
<tr>
<td>Remove all descriptors with high uncertainties</td>
<td>123 (6)</td>
<td>114 (6)</td>
<td>0.32 (7)</td>
<td>14.4 (7)</td>
</tr>
<tr>
<td>Removal all descriptors with low importance</td>
<td>132 (6)</td>
<td>108 (8)</td>
<td>0.30 (7)</td>
<td>14.2 (8)</td>
</tr>
<tr>
<td>Remove descriptors with high uncertainties and of low importance stepwise</td>
<td>145 (4)</td>
<td>-</td>
<td>0.41 (2)</td>
<td>-</td>
</tr>
</tbody>
</table>

Local models
To further improve the predictive power of the four models, I divided the compounds into four groups and developed local models for each group. The first group consisted of fluorinated compounds. These compounds were generally predicted with relatively low accuracy, and thus were treated as a separate group. The other groupings were based on a PCA score plot. Compounds that exhibit similar behavior (with regard to their molecular descriptors) will cluster in these plots. Selections based on this clustering generated three groups, which were the same for the first- and second-dimension models. The second group consisted of chlorinated and brominated compounds. The third group consisted of compounds with long carbon chains and associated compounds (e.g., alkanes, PEGs, glymes, FAMEs, long chain alcohols, ketones, and aldehydes). The fourth and final group contained all remaining compounds. Fluorinated compounds comprise the first group. Therefore, a compound that (for example) contains a fluorine and a chlorine was still only assigned to group one.
To compare the local model results with the aforementioned results, I predicted the four groups using the best of the models described above, and calculated the RMSE$_P$ for each group (Table 8). Compared with the global PLS models, the local models yielded better results for group 2 (chlorinated and brominated compounds) and group 3 (long chain compounds). The results for group 2 improved only modestly, but those of group 3 improved significantly with the use of local models. Global models are (generally) less time-consuming than local models, so the global model was used for group 2. Thus, at the end, a local model was only used for the compounds in group 3.

Table 8 RMSE$_P$ values obtained using local models and the best global model for comparison (from Table 7) using PLS. The higher of two numbers from a direct comparison is shown in red font.

<table>
<thead>
<tr>
<th></th>
<th>Group 1 (F)</th>
<th></th>
<th>Group 2 (Cl/Br)</th>
<th></th>
<th>Group 3 (Long chain)</th>
<th></th>
<th>Group 4 (Other compounds)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Global model</td>
<td>Local model</td>
<td>Global model</td>
<td>Local model</td>
<td>Global model</td>
<td>Local model</td>
<td>Global model</td>
<td>Local model</td>
</tr>
<tr>
<td>$t_R$ (s)</td>
<td>179</td>
<td>174</td>
<td>129</td>
<td>125</td>
<td>64</td>
<td>14</td>
<td>127</td>
<td>129</td>
</tr>
<tr>
<td>LRI</td>
<td>168</td>
<td>174</td>
<td>110</td>
<td>93</td>
<td>49</td>
<td>5</td>
<td>111</td>
<td>117</td>
</tr>
<tr>
<td>$t_R$ (s)</td>
<td>0.46</td>
<td>0.43</td>
<td>0.32</td>
<td>0.28</td>
<td>0.21</td>
<td>0.02</td>
<td>0.29</td>
<td>0.30</td>
</tr>
<tr>
<td>PEG-$t_I$</td>
<td>20.7</td>
<td>18.9</td>
<td>13.4</td>
<td>12.4</td>
<td>7.6</td>
<td>0.9</td>
<td>12.3</td>
<td>13.8</td>
</tr>
</tbody>
</table>

“Federation of local models” results (ChromGenius)

Optimization of the ChromGenius models was also assessed, by calculating the RMSE$_P$ for each model (Table 9). In the first step of the model evaluation, factors that influence the search for molecules (within the knowledge base) that are similar to the compounds of interest were assessed. The first factor was the function used to calculate the similarity (or distance) of two compounds. The dice coefficient and Euclidian distance seemed to be most suitable (i.e., yielded the best results) for the first-dimension and second-dimension models, respectively. For the second-dimension models, similar Euclidian distances were obtained even when the number of compounds used to create the sub-group from the knowledge base was varied. However, in both cases ($t_R$ and PEG-$t_I$), increasing the number of compounds improved results. Slightly different results were obtained for the first dimension. In accordance with the second-dimension results, 25 compounds yielded the best result for the $t_R$ model, but 20 compounds yielded the best results for the LRI model. The differences in RMSE$_P$ were also larger (93 vs. 83) for the LRI model when comparing the number of compounds used from the knowledge base.
The physico-chemical property parameters used in the similarity calculations and the retention time or RI prediction equation were then considered. All parameters were included in the previous step. However, the Abraham parameters are included in a separate add-on in the ChromGenius software. As shown in Table 9, effects of including these parameters depended on the model. Using only Abraham parameters or omitting them had very little effect on the performance of the LRI and \( t_R \) models. In contrast, the performance of the \( t_R \) and PEG-1 model showed a difference in step two. However, compared with the predictions obtained when only Abraham parameters were used, better prediction errors (RMSE\(_P\)) were obtained from all models (except the \( t_R \) model) when all parameters were included.

Table 9 Lowest and highest values of each response and RMSE\(_P\) following each developmental step of each model constructed using ChromGenius. The black boxes and the values shown in green font indicate results of the final model and best model following the first step, respectively.

<table>
<thead>
<tr>
<th>Step</th>
<th>Lowest measured value</th>
<th>( t_R ) (s)</th>
<th>LRI</th>
<th>( t_R ) (s)</th>
<th>PEG-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Lowest measured value</td>
<td>270</td>
<td>808</td>
<td>1.68</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>Highest measured value</td>
<td>3325</td>
<td>3413</td>
<td>6.62</td>
<td>215.1</td>
</tr>
<tr>
<td></td>
<td>Model settings</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dice coefficient (25 compounds)</td>
<td>159</td>
<td>93</td>
<td>0.28</td>
<td>14.6</td>
</tr>
<tr>
<td></td>
<td>Dice coefficient (20 compounds)</td>
<td>160</td>
<td>84</td>
<td>0.32</td>
<td>17.2</td>
</tr>
<tr>
<td></td>
<td>Euclidian distance (25 compounds)</td>
<td>196</td>
<td>105</td>
<td>0.28</td>
<td>14.5</td>
</tr>
<tr>
<td></td>
<td>Euclidian distance (20 compounds)</td>
<td>204</td>
<td>98</td>
<td>0.29</td>
<td>15.5</td>
</tr>
<tr>
<td>2</td>
<td>Best model setting, no Abraham parameters</td>
<td>176</td>
<td>155</td>
<td>0.34</td>
<td>17.5</td>
</tr>
<tr>
<td></td>
<td>Best model setting, Abraham parameters only</td>
<td>158</td>
<td>153</td>
<td>0.34</td>
<td>22.6</td>
</tr>
<tr>
<td>3</td>
<td>Reduction of the number of molecules used per parameter</td>
<td>160</td>
<td>135</td>
<td>0.26</td>
<td>18.1</td>
</tr>
</tbody>
</table>

In the last step, parameters affecting the final retention time or RI calculations were considered. Four molecules per parameter (the default value) were used in the final equation so far. According to the reference manual for the ChromGenius software, four compounds should be used for large datasets containing more than 1000 compounds and three for smaller datasets. Therefore, the default value was reduced to three to determine if such a reduction would yield an improvement. Improved results were obtained for the \( t_R \) model, but worse results (than the original results) for the other models. The results of the final and best models are enclosed in the bold boxes shown in Table 9. In addition, the best results from the first step are
shown in green font, as these values were used in the subsequent modeling steps. The calculated RMSE\textsubscript{P} values are similar for some models (e.g. 159 s vs. 158 s for the $^{1}\tau\text{R}$ model). Models could then be chosen in such a manner that the $^{1}\tau\text{R}$ model and LRI model and the $^{2}\tau\text{R}$ model and PEG\textsuperscript{-2}I model, respectively, are calculated using the same parameters. For example, the full 20 component Dice model (step 1) predicts both first dimension data well.

**Validation of models**

In the PLS modeling, I predicted compounds from group 3 (long chain compounds) using the corresponding local model, and all other compounds using the best global model. Overall, the first-dimension models exhibited better precision (upper part in **Figure 19**) than their second-dimension counterparts, as expected. Values of 5\% (RMSE\textsubscript{P}: 109 s) and 4\% (RMSE\textsubscript{P}: 95) were obtained for the average relative deviation of the predicted values from the experimental values of the retention time ($^{1}\tau\text{R}$) and index (LRI), respectively. The corresponding 90-percentiles were 10\% and 8\%, respectively. For the second dimension, average deviation values of 5 and 12\% were obtained for the retention time and index, respectively, and 90-percentiles of 12 and 25\%, respectively.

In the final models for the first dimension, the boiling point (as expected) contributes significantly to the model results. The surface tension and index of refraction also have large contributions, and make the largest contributions to the second-dimension models. The index of refraction is linked to the polarizability of a compound [120], a driving force for separation on the secondary column, which explains its relatively large contribution.

The settings used for the final ChromGenius models are indicated in green font in **Table 9**. For the $^{1}\tau\text{R}$ model, only the Abraham parameters were used, whereas for the other models all parameters were included. Three compounds per parameter were used to calculate the predicted retention times (or RIs) for the $^{2}\tau\text{R}$ model, and four compounds for the other models. Similar to findings with the PLS models, the ChromGenius predictions fitted the first-dimension models better than the second-dimension models (**Figure 20**). Average relative deviation values of 6\% (RMSE\textsubscript{P}: 142 s) and 3\% (RMSE\textsubscript{P}: 89), with corresponding 90-percentiles of 12\% and 6\%, were obtained for predictions of retention times ($^{1}\tau\text{R}$) and indices (LRIs), respectively. The relative deviations are comparable to those of the PLS models described above, although the RMSE\textsubscript{P} values differ. The $^{1}\tau\text{R}$ model has a larger error using ChromGenius (109 s for PLS vs. 142 s for ChromGenius) while the LRI model has a smaller error using ChromGenius (**Table 10**). The average relative deviations between the predicted and measured values for the $^{2}\tau\text{R}$ and PEG\textsuperscript{-2}I models (5\% and 12\%, respectively; 90-percentiles, 10\% and 24\%) do not significantly differ from those obtained for the PLS models.
Figure 19 Predicted vs. experimental values for the first- and second-dimension retention time and index PLS models, respectively.

In addition, similarly to Dossin et al. [109], I calculated average predicted retention times and indices from the PLS and ChromGenius models. This reduced prediction errors for all four models (Table 10), but not sufficiently to justify the additional work.

Generally, for all models, the relative differences between the predicted and measured retention times and indices were largest for early-eluting compounds and fluorinated compounds (e.g., fipronil, an insecticide). For early eluting compounds, small deviations will have a strong impact on relative deviations, resulting in relatively large errors. The large deviations for the fluorinated compounds may have resulted from the inclusion of just a few of these compounds in the model-building process.

The index models were compared with corresponding retention time models by dividing the final RMSE$_p$ (from Table 10) by the range of values (from Table 9). This yielded the same relative errors in all except one case. The only RI model that performed better than the corresponding retention-time model was the first-dimension ChromGenius LRI model. In addition, the final models were compared to previously published studies. Consequently, the here developed models were equally good or better than previous modelling approaches (see Paper IV).
Results of the quality assurance tests (Table 10) show that the models using the best single descriptor do not give a lower error. In fact, the models described above are considerably better than a simple basic model.

**Table 10** RMSEP (obtained using the external validation set) for the final PLS and ChromGenius models and the average of both models in comparison to results from simple "one-descriptor-models" (for quality assurance).

<table>
<thead>
<tr>
<th></th>
<th>$t_R$ (s)</th>
<th>LRI</th>
<th>$t_R$ (s)</th>
<th>PEG-$^2$I</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>109</td>
<td>95</td>
<td>0.27</td>
<td>11.3</td>
</tr>
<tr>
<td>ChromGenius</td>
<td>142</td>
<td>89</td>
<td>0.29</td>
<td>11.9</td>
</tr>
<tr>
<td>Average of PLS and ChromGenius prediction</td>
<td>103</td>
<td>72</td>
<td>0.22</td>
<td>9.6</td>
</tr>
<tr>
<td>Boiling point</td>
<td>200</td>
<td>163</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>log$K_{OW}$</td>
<td>-</td>
<td>-</td>
<td>0.72</td>
<td>30.6</td>
</tr>
<tr>
<td>log$K_{OW}$/Weight</td>
<td>-</td>
<td>-</td>
<td>0.71</td>
<td>29.2</td>
</tr>
</tbody>
</table>

**Figure 20** Predicted vs. experimental values for the first- and second-dimension retention-time and index models constructed using ChromGenius.
Applying the retention prediction models

The final models can now be used for non-target screening purposes. A detected peak may correspond to several possible structures that match the MS spectrum. If so, retention times or RIs can be predicted for each of these structures and compared with the measured values of the unknown compound. The deviations between the predicted (first- or second-dimension) and measured values for a possible candidate may lie outside the given error range. In such cases the candidate can be excluded, thereby reducing the list of probable structures.

A possible risk when predicting new compounds is that they could lie outside the model domain, which is difficult to assess. An option would be to create a PCA score plot using the molecular descriptors for the compounds used to build the model and the compounds that will be predicted. This will give an overview of the degree of similarity between the predicted compounds and the compounds that define the model domain. Relatively large errors are expected for compounds lying outside the model domain.

Conclusions and improvements

In this section, I have presented robust models for predicting first- and second-dimension retention times in GC×GC as well as LRI and PEG-2I values. Overall, compared with the other models, the PLS method yields models with low prediction errors, relative to those of the other models, and only the LRI models are improved by use of ChromGenius. Therefore, the use of PLS models is recommended for retention-time and RI predictions. However, PLS is complex and some knowledge of the software and PLS (in general) is required.

To improve the predictive power of the PLS models, all descriptors characterized by high uncertainties or low importance are removed and a new model excluding those descriptors is created. Another option would be to individually remove the descriptors to determine their individual influence on the model. In addition, the local models can be improved in the same manner as the other PLS models. The number of descriptors is large, and each will have to be applied to four models. Therefore, a very high number of models must be created, rendering the task very time-consuming. To obtain “ideal” models, individual removal of all descriptors characterized by high uncertainties or low importance, to determine their influence, and further improvements of the local models are recommended.
Part II summary

In this second part of my thesis (Data evaluation) I have presented ways of analyzing and evaluating the data, starting by considering time trends. Time trends have been identified for several compounds (for example, UV-filters). Similarly, spatial trends could be used. In addition, I have presented a new RI system (PEG-\(\alpha I\)) for evaluating unknown compounds in the second GC×GC dimension. This new system works well for all compounds lying within the retention range of the applied PEGs when an apolar × polar column set is used. PEG-\(\alpha I\), like the LRI, will help to identify unknown analytes when GC×GC is used. I have also presented a method for predicting (via PLS) LRI values, PEG-\(\alpha I\) values, as well as first- and second-dimension retention times. These predictions can be used to reduce the number of possible structures for unknown compounds in non-target screening approaches. In the following part of the thesis, Data Archiving, the concept and application of digital archiving is described.
PART III
Data Archiving
Digital archiving

A review of notes from meetings at the start of my PhD project to remind myself of the project’s original main goals and how well we “stayed on track” showed that it was quite well planned. The ultimate goal was to create digital archives. We have not yet reached this goal, but I researched and developed several ways to create data to put in digital archives, and evaluate data from digital archives, as described in Part I (Data generation) and Part II (Data evaluation). Thus, the overall course of my PhD was consistent with the objectives, but of course there were some deviations. Notably, the time devoted to the creation of protocols for data storage and archiving was much shorter than initially planned. This was probably due to a combination of factors: projects change because of unforeseen developments, people and knowledge change, and Murphy’s Law (whatever can go wrong will go wrong) should not be forgotten. Nevertheless, in this last chapter of my thesis, I try to provide guidelines for data storage and emphasize important considerations when data are stored (in digital archives) rather than samples. This chapter is also intended to provide a summary of the thesis and an overview of future plans and needs for further work.

What are digital archives?

In previous parts of the thesis I briefly mentioned some aspects of digital archiving and the nature of digital archives, or rather associated challenges. Here I explain in more detail what digital archives are.

Digital archives are data repositories. My research area concerns analysis of organic pollutants in environmental matrices. As previously explained, this analysis is typically performed using a chromatographic technique (GC, LC, GC×GC, or other combinations), which separates the sample components, and an MS detection system. Such analysis yields three (or more)-dimensional data, i.e., chromatograms linked to mass spectra that show the intensities of ion species. The data size increases with the resolution of the analytical system. Thus, GC×GC and high-resolution MS generate very large datasets. We would like to store large amounts of data in digital archives, i.e. repositories, for use by other researchers. In addition to the raw data, processed data files could be uploaded in the form of peak tables with accurate masses or formulae and intensity information for already obtained results. Such data repositories could be either open so that everyone can upload information or restricted so that only a few users can upload data. However, with open repositories, quality assurance is difficult. How can we ensure that the uploaded data are of good quality? Furthermore, the data source must be
verified and use of the data requires sufficient metadata, which begs the following question.

**What data are needed?**

The first question one might ask when dealing with digital archiving is: **What** kind of data do we need? Several other questions also arise, including: How much data? What metadata (if any) should be included? What types of samples should be covered? How should we treat samples? What analytical aspects should be considered? What instrumentation do we need?

The purpose of digital archives is the long-term storage of data over many years. To enable long-term storage and still ensure good comparability, a harmonized approach for reporting data and metadata must be developed and appropriate quality assurance measures must be taken. The key word in this sentence is *harmonization*. Providers of the archive must agree on a common format for data submission that will allow present and future (in 10 or 100 years) data analysis, comparison, and evaluation. When common ground is set for these values, a digital archive can be successfully created.

**Sample collection**

The first step in establishing an archive is sampling, i.e., sample collection. The samples should be representative of the sampled matrix, and large fluctuations in the sample composition must be taken into consideration. A representative fraction of the sample matrix should be collected either by taking sufficiently large amounts and then homogenizing the samples or by repeated sampling and pooling the final samples before homogenization.

Samples of the focal matrix in my doctoral studies, sewage sludge, were collected after anaerobic digestion (three weeks residence in the digester). Hence, the sludge samples obtained were rather homogeneous and, thus, representative. However, other matrices may be relatively inhomogeneous. For example, the incoming water (influent) has significant inter-day variability, so grab samples of this water will have low representativeness. Therefore, collecting several samples on different days may be a good strategy. These samples can then be pooled and extracted. Other matrices (for example, soil) exhibit lower inter-day variability than influent water. Changes in these matrices occur rather slowly and collecting a sample once (time-wise) may be sufficient. However, soil samples are typically spatially inhomogeneous, with contaminants occurring in very high concentrations in some regions and much lower concentrations in other regions. This drawback may be overcome by collecting and then mixing (for homogenization) several samples from the same area. Other sampling techniques (for example, passive sampling of water or air) are slow processes, so the inhomogeneity of the samples is reduced. These “time-integrative” sampling approaches are good methods for
obtaining average concentrations over certain periods of time. The composition of air samples may also vary significantly with the wind direction and/or type and magnitude of local emission sources. Hence, reporting how and when samples were taken is essential for users of data archives.

In addition, if yearly sampling is performed the samples should be taken at the same time each year. This stratified sampling approach will help to reduce the variability between samples. Additional metadata related to the sampling should be noted, as these will be helpful for subsequent data-evaluations. In relation to STPs, this information includes, for example, the amount of incoming water, amount of sludge produced, or number of people connected to the STP. Other examples of important metadata include information on flood or spill events. Obtaining such metadata may be difficult, especially when many years (or decades) have passed.

**Sample treatment**

Sample treatment is considered in the first part of the thesis, using the extraction of sewage sludge as an example. The sample treatment may have a significant effect on the type of compounds that can be obtained. For example, when GPC is used (as in my doctoral studies), the fact that GPC is a size-exclusion method must be taken into account. In fact, GPC may exclude some of the large analytes, and measurement of these analytes will require a complementary method. This approach yields the methods depicted in **Figure 3** (page 11). Therefore, storing both the data and detailed information on the sample-treatment techniques is essential.

**Sample analysis**

As already discussed, samples are usually analyzed using a chromatographic technique, such as GC or LC, coupled with MS. Generally, samples analyzed using GC are small or medium-sized compounds with low or medium polarity. Compounds that can be analyzed with LC are often more polar than those analyzed with GC, and have various sizes. However, in both GC-MS and LC-MS, the results vary significantly with the ionization technique, as discussed in the following sections.

**Selection and standardization of chromatographic techniques**

Peak capacity is an important characteristic of any chromatographic technique. A rather long chromatographic column and runtime (one hour or more) provides high peak capacity, which is desirable. It is also desirable to detect and separate as many compounds as possible during a run. The peak capacity can be increased through 2D chromatographic techniques. In my PhD project, the separation and peak capacity were improved via GC×GC. The first- and second-dimension columns separate the compounds based on...
different properties or separation mechanisms, thereby improving the separation.

In addition, the co-injection of reference compounds must be considered to ensure that the data are comparable. These could be carbon-13 ($^{13}\text{C}$)-labeled or deuterated internal standards that are spiked to the samples, or additional compounds that are separately analyzed (e.g., alkanes or PEGs in GC and GC×GC) to determine retention times. Such co-injection will allow correction of retention-time shifts for qualitative analysis. Moreover, as discussed in Part II of the thesis (pages 42-49), the alkanes and PEGs can then be used to calculate RIs. The spiked reference compounds can also be used for the relativization of signals in quantitative analysis. Here, a constant (or at least known) concentration of the injected reference compounds is required. Furthermore, the reference compounds should cover the entire chromatographic range from early- to late-eluting compounds and, for (semi-)quantitative purposes, be the same over the years considered. Of course, additional reference compounds can always be included.

Selection of ionization techniques for GC-MS and LC-MS

Electron ionization (EI) at a standard electron energy of 70 eV is, by far, the most commonly used ionization technique for GC-MS. Spectra recorded at this energy are reproducible and can be compared across instruments and with entries in libraries. However, sometimes a molecular ion is lacking, thereby hindering identification of unknown compounds. Molecular ions may be obtained during GC-MS analysis by using complementary techniques, such as chemical ionization (CI). In CI, ions are formed from a reagent gas (for example, methane) and then react with the analytes. This leads to formation of molecular or quasimolecular ions (e.g., protonated molecular ions) that can be detected. Molecular ions may also be obtained via low-energy EI, also referred to as soft EI. In an attempt to identify some of the unknown GC compounds from the time-trend analysis described in Part II (page 28-41) and Paper II, samples were analyzed using low-energy EI (at 15 eV). The acquired data allowed tentative identification of an additional 19 compounds and exclusion of one false positive. Based on these results, a combination of “normal” EI (at 70 eV) and soft EI or CI is recommended for obtaining molecular ions and characteristic fragmentation spectra via GC-MS analysis.

In LC-MS, the most common ionization techniques (in descending order of use) are ESI, atmospheric pressure chemical ionization (APCI), and atmospheric pressure photoionization (APPI). These techniques can all be used in positive and negative modes to target different types of analytes. Some

---

2 This information is based on the hits obtained through Google Scholar searches. ESI yields roughly 335,000 hits, APCI yields roughly 44,900 hits and APPI yields roughly 7,460 hits when searching for the unabbreviated term in quotes (as at December 6, 2017).
studies have reported that the matrix effects in APCI are weaker than those associated with ESI. However, ESI covers more compounds and is, hence, more universal than APCI. Therefore, ESI remains the generally preferred ionization technique for LC-MS, as shown (for example) for pesticides, hormones, and pharmaceuticals [121–123]. In contrast to ESI and APCI, APPI can ionize non-polar analytes [124]. However, these analytes can generally also be analyzed via GC-MS. Therefore, if only one ionization technique can be used, I recommend use of ESI for sample analysis with LC-MS techniques aimed at digital archiving. Ionization with ESI should then be performed in positive and negative modes. In positive mode, positive ions are formed through addition of a hydrogen cation or other positively charged species (e.g., sodium ion), while in negative mode negative ions are formed via the abstraction of a hydrogen nucleus or addition of a negatively charged species. Compounds with high proton affinities (e.g., amines) are easily ionized in positive ESI mode, whereas compounds such as phenols are better ionized in the negative mode, where oxygen loses a hydrogen nucleus.

In cases where two ionization techniques for LC-MS analysis can be used, I recommend a combination of APCI and ESI. For some sample matrices and compounds, APCI might be more suitable than ESI for ionization of the compounds. Analytes in sample matrices containing many neutral compounds that are difficult to ionize in solution (e.g., nitro compounds) may be easier to ionize using APCI.

Suitable mass analyzer

Various types of mass analyzers are commercially available. A high-resolution mass analyzer that provides accurate mass information should be used to maximize the number of identification points, thereby facilitating identification of unknown compounds [125]. The two most widely used types of high resolution mass analyzers are the time-of-flight mass spectrometer (TOFMS) and orbitrap, but the former is used more often than the latter. The TOFMS is well-established, has a large mass range, and is cheaper than the orbitrap [126]. However, the orbitrap has significantly higher resolution than the TOFMS. In addition, it enables MS\textsuperscript{n} experiments, and thus the isolation and fragmentation of specific ions. This procedure can, in theory, be repeated many times. Similar experiments (MS\textsuperscript{2}, i.e., MS/MS) can be conducted by coupling a quadrupole mass analyzer to a TOFMS (QTOF).

The principles of MS/MS and the mentioned operational modes are explained in Part I (page 8 and 9). Characteristic fragment spectra, similar to EI spectra from GC-MS, can be generated by using MS/MS following LC. As previously explained, there are several possible operational modes. The two options for non-target purposes (if the instruments support these modes) are auto-MS/MS and all-ion-MS/MS. Auto-MS/MS is a data-dependent
acquisition method where a pre-scan is performed and the most abundant ion(s) is/are chosen for fragmentation and subsequent analysis. In practice, the number of ions that can be selected is limited, for two reasons. Firstly, since only one precursor ion at a time can pass through the first mass analyzer, the reference masses used for the internal mass calibration must be recorded separately, and, hence, the accuracy of the mass measurements decreases from scan to scan within an auto-MS/MS cycle. And secondly, the available time-frame for an auto-MS/MS cycle is limited by the peak width, and has to be short enough to not miss important information.

In all-ion MS/MS mode, all ions are fragmented in the collision cell and the first mass analyzers is in transmission (RF only) mode. This data-independent mode was used in the time-trend study presented in Part II. The spectra were compared with entries in MS/MS libraries for tentative identification of compounds present in the libraries. In addition, for confirmation, target MS/MS (SRM mode) spectra were recorded for compounds with significant trends.

Both, all-ion MS/MS and auto-MS/MS, can be performed at different collision energies. The advantage of auto-MS/MS is that a product ion can be directly linked to a precursor ion. The disadvantage is that ions (or analytes) with low signal intensities have low likelihoods of triggering MS/MS fragmentation, so their identification is difficult.

One of the main aims of digital archiving is to ensure high comparability of samples and maximize the coverage of analytes, so ideally fragmentation spectra are collected for all (or most) analytes and similar spectra are obtained across samples. This is prevented if some of the data are missing, owing to different precursors being chosen for fragmentation in auto-MS/MS. Hence, I recommend use of all-ion MS/MS at different collision energies if a specific fragmentation mode must be selected. If time and funding allow, auto-MS/MS could also be performed to create a complementary dataset.

Effects associated with the main disadvantage of all-ion MS/MS, i.e., chimeric spectra, can be reduced by cleaning-up the mass spectra. This can be done through peak deconvolution, which is quite routine in GC-EI-MS analyses. The peak shapes of several fragments or precursors and fragments can be compared, and ions with matching peak shapes and apexes can be combined into a single spectrum. Methods and software that can deconvolute and link LC-ESI-MS precursor and product ions generated in all-ion MS/MS, for example, MS-DIAL (Mass Spectrometry – Data Independent AnaLysis software) [127], have been reported in previous studies.

**Changes in the analytical procedure and workflow**

Data that are stored in digital archives are collected over many years. However, changes in the sampling or analytical procedures may occur, possibly due to improvements in these procedures, the workflow or analytical
techniques [53]. As previously mentioned, providing maximum information about the samples and analyses (metadata) is essential for data archiving. The information and metadata will be helpful for data interpretation. Changing the analytical instrument directly affects the resulting measurements and can lead to time series artifacts [53]. Hence, when changing to a newer analytical technique or different sampling procedure, simultaneous processing of samples using the old and new methods is recommended. This will ensure comparability, before data from the new method are used or stored in a digital repository. Furthermore, a harmonized Data Collection Template (DCT) can be developed to ensure that sufficient information is always provided when data are uploaded. Such efforts are being made by the Network of reference laboratories, research centers and related organizations for monitoring of emerging environmental substances (NORMAN), in connection with the development of a digital archive – the Digital Sample Freezing Platform (DSFP) [128].

In addition, the previously mentioned reference compounds (page 71) will help to ensure good comparability between new samples and old datasets.

**Data storage**

Ideally, open data formats should be used to ensure that data stored in digital archives can be accessed by everyone. One of the most common open formats is mzML, which was developed by merging the previously used formats mzXML and mzData [129], through joint efforts of various organizations (including instrument vendors), and is based on Extensible Markup Language (XML). Another XML based format, Analytical Data Interchange Format for Mass Spectrometry (ANDI-MS), is based on netCDF [130] and was developed by the American Society for Testing and Materials (ASTM) [131]. The netCDF or ANDI-MS file format is often used for GC-MS data. The most common open source or vendor-independent file format for LC-MS is mzML. This format is, for example, supported by the previously mentioned deconvolution software MS-DIAL [127] (page 73) and used in the DSFP of NORMAN.

**Data size**

The increasing size of datasets is a common present-day problem. For example, up to 1.5 GB³ and 5 GB³ of data are contained in the Leco GC×GC high-resolution MS files (runtime: ~1 h) and Agilent LC-IM-QTOF files (runtime: ~1 h), respectively, obtained in my doctoral work. The size increases when the data are changed from an instrument-specific format to an open format. For example, the size of a 1.3-GB Leco GC×GC high-resolution MS file in vendor format (SMP file), increases to 3.5 GB, 7.9 GB, and 8.2 GB when

---

³ Using the vendor file format, respectively.
converted to mzML, ANDI-MS, and CSV formats, respectively. This increase will multiply when several files are considered, leading to extremely large archive databases. To my knowledge, the size of the file format can be reduced in four ways: storing data processing results (peak lists) rather than the raw data, recording MS spectra in centroid mode (peak apex) instead of profile mode, applying a cutoff to the spectra, and compressing the files for storage.

The first issue to consider is whether the original raw (i.e., unprocessed) data should be stored, or data retained after application of a peak selection process. The advantage of only storing the results of peak picking is that the data size is reduced. However, peaks overlooked by the automatic peak-picking algorithm will be undetected and the corresponding information will be lost. Therefore, storage of the raw files (for retaining all the information) and a harmonized set of peak-picked data is recommended.

Another question that needs to be answered is whether data should be collected and stored in centroid or profile mode. Centroid mode means that only the m/z at the apex of each detected ion (rather than all individual data points) is stored. This preserves most of the information while reducing file sizes, and is therefore recommended.

As previously stated, the file size may be reduced by applying a cutoff to the MS spectra. This cutoff should be linked to the electronic noise level, which can be determined by examining the spectra for regions where few (if any) peaks are located, and then checking high-mass fragments. I suggest a cutoff at five times the electronic noise. All fragments below this level will be removed from the spectra and the file size will be reduced significantly [132].

Compressing files by creating ZIP or Roshal Archive (RAR) archives also reduces file sizes. One of the best software tools for data compression, 7zip, is freely available [133]. For example, a 7.9-GB ANDI-MS file can be reduced to 1.3-GB and 0.99-GB files using 7zip’s normal and ultra-compression rates, respectively. Moreover, an 8.2-GB CSV file and a 3.5-GB mzML file can be reduced to 1.2 GB and 0.97 GB, respectively, using the normal compression rate in 7zip. Files in the SMP format associated with LECO instruments are already compressed, so additional compression provides no significant reduction in file size. However, an undesirable effect of data compression is that it increases the time required to create and open files. The data must then be extracted before comparisons can be made.

Furthermore, many instrument vendors include an archive option in their software so that the data can be immediately compressed. However, one must consider the format. The compressed files, which are exported from the instrument software, may be in a specific format that can only be opened by the instrument software.

As previously discussed, the mzML formatting produces smaller files than ANDI-MS (netCDF) or CSV and is therefore preferred to these formats. GC-MS and LC-MS files can be stored as mzML files and processing tools such
as MS-DIAL are compatible with mzML. Hence, I recommend storage of data in mzML format when possible. If mzML is unavailable, another open format (e.g., ANDI-MS) should be used, thereby allowing everyone to access the data.

**Data evaluation software**

When storing raw data in a vendor-independent open-source format (e.g., mzML) a factor that needs to be considered is the evaluation of those data. In addition to open source data formats, open source software packages for processing and visualizing of data exist (e.g., MZmine 2 [134], XCMS and XCMS² [135,136], and Skyline [137]). These open source software packages are vendor independent, freely available for everyone, and include advanced tools for processing, data evaluation and data visualization. Consequently, they often offer advantages that data processing in digital archives alone cannot provide. In addition, these software formats usually do not change quickly and, in general, compatibility between different versions is ensured. Another advantage is that these initiatives are often community driven and, hence, often offer good user support [135,137,138].

**What advantages do digital archives have?**

As collected samples can be stored in environmental specimen banks (ESBs), two other questions that should be addressed are why do we need digital archives, and why is digital archiving a better option than storing physical samples in ESBs? The most obvious advantage is that degradation or contamination effects are prevented. When stored for long times, samples may become exposed to, for example, air and some compounds might be transformed, as previously mentioned in the time-trend analysis in Part II (page 35). Evaporation, transformation or degradation will lead to reductions in concentrations of some compounds in long-stored samples or increases in concentrations of others. Therefore, the nature of the samples will change. In contrast, data stored in digital archives remain constant.

Another advantage is the general accessibility. For example, data from digital archives can be downloaded and used repeatedly. In contrast, samples from ESBs will eventually be used up and they can only be used in a limited number of studies. If those samples are then only used to perform target analysis, valuable information will be lost. In addition, the data can generally be more easily and (more importantly) rapidly accessed than data associated with physically stored samples. Whether a certain compound occurred in a different matrix or at a different time point can simply be determined by analyzing the data files in the archive. This is much less time-consuming than applying for samples from an ESB, then extracting and analyzing them. Moreover, this accessibility is transnational. Using digital archives, data from
different countries can be compared more easily than when ESBs are used, and common features and differences can be evaluated.

The main potential drawback is that the development of information technology is very rapid and ensuring that files created today can be accessed and used decades or even centuries later may be difficult. Nowadays, software versions, file formats, and computer media rapidly become obsolete. Hence, an open format that will remain largely unaltered in the future is essential for sustained compatibility of different versions.

**What can digital archives be used for?**

How can we use the archived data and why do we want to create digital archives? Archives enable retrospective analysis. For example, imagine that analysis of a recently collected sample of sewage sludge (or any other matrix) reveals a new compound that has not been previously detected. Retrieving data files for samples from previous years (if available) and extracting the necessary information from digital archives will enable researchers to find out if the compound was not previously present in the sampled matrix, or was simply unidentified or undetected due to the use of automated processes.

The archived data can also be used for time-trend analysis, as presented in **Part II** of the thesis. The data can be aligned using the retention markers previously suggested (page 71) and then normalized through the labeled reference standards that were added to compensate for changes in instrument sensitivity. This will allow comparison of data collected during different years, thereby enabling subsequent time-trend analysis.

As previously mentioned, digital archives would also allow comparison of data from different countries (if the archives are open for this purpose) or different matrices. For example, if a compound is found in sewage sludge, the compound’s origin, such as the wastewater, may be of interest. Similarly, sediment or water from a lake can be subsequently compared with fish samples obtained from that same lake, other lakes or nearby rivers. The archived data can be used in several ways and many research questions can be addressed. The accessibility of data (easy, quick, and transnational if allowed by the archive), is a major advantage of digital archives.

**What are the current options?**

In this section, the current options for digital archives are examined and the question of how digital archives can be created is addressed. Examples where raw data have been stored and made available publicly are presented. Furthermore, the methods of storing data in digital archives are highlighted.

Let us first consider the means of storing data in an archive. Databases containing the metadata and MS data form the core of a digital archive. The
MS data can be stored as either raw data files (vendor or open format) or as processed files, i.e., peak lists. Processed data files can be inserted and searched directly, while raw data files stored in the databases may need translation. An additional application or script will then be needed to visualize the data. Therefore, a decision has to be made whether non-modified raw data files, modified or transformed data files should be stored. Modifications may include peak-picking or normalization of signal heights and retention times [128,139].

A database-management system (DBMS) can be used for creating and using databases [140]. Many common DBMSs are written using a Structured Query Language (SQL), such as MySQL, MariaDB or PostgreSQL. PostgreSQL has additional advanced features and is often used for business applications, whereas MySQL is used by many private users. MariaDB, which was developed by the developer of MySQL, is considered the current open-source alternative [141].

A general example of the structure and functionality of a digital archive is shown in Figure 21. The raw data are uploaded along with the metadata using, for example, a File Transfer Protocol (FTP) or web client (i.e., browser). The data can then be saved directly (e.g., in mzML format) or processed, for example, via peak picking and creation of a peak list [128,139]. The data are then stored in the database and can be accessed through an interface (e.g., a web client, or desktop graphical user interface, GUI).

I next present some examples of digital archives and databases, starting with the Data Resources Of Plant Metabolomics (DROP met) database, a repository of data obtained in the PRIMe project [142]. This database contains raw data obtained from GC-MS, LC-MS and LC-MS/MS analyses of plants in metabolomics studies, which are located on a file server. Most data in this database are converted to the netCDF format and compressed to a zip archive. Some files are accompanied by an additional file that includes metadata. The database includes references to 25 datasets. Each of these contain up to several hundred chromatograms that can be downloaded without special access rights.

Another example, the MetabolomeExpress [143], is also linked to metabolomics data and was developed by Adam J. Carroll of the Australian Research Council Centre of Excellence in Plant Energy Biology. The database contains GC-MS data that are stored and organized through an FTP repository combined with an SQL-based database that stores metadata and response statistics for metabolites [139]. In addition, a data processing algorithm is embedded in the database. When a user uploads raw data through an FTP client (e.g., FileZilla), a retention index file and a file containing metadata must also be uploaded. The imported files are then treated by a peak detection algorithm and the data are processed using a Mass Spectral and Retention time Index (MSRI) library matching algorithm [139,144].
Another example, MetaboLights, is an open-access repository for raw experimental data and metadata related to metabolomics studies [145,146]. The database contains information from 315 studies, on almost 25,000 compounds [147]. The submitted raw data are GC-MS, LC-MS or nuclear magnetic resonance spectroscopy (NMR) data that are organized in an SQL database [146]. Search fields allow searches of metadata for easy identification of datasets or compounds. Unprocessed raw data files can be submitted in any format to the MetaboLights database. However, the recommended open source format for submission is mzML [146].

The use of digital archives is well-established in the field of metabolomics, but less established in other fields. However, the NORMAN network has established a database called EMPODAT [148]. This is the largest database (containing information and metadata) worldwide for emerging substances in the environment, but contaminant concentrations (rather than raw data files) are stored. Thus, it is a database rather than digital archive. Nevertheless, the mass spectra of unknown compounds can be stored in MassBank [149], a spectral database supported by the network. MassBank hosts more than 47,000 spectra of more than 15,000 compounds. One goal of the NORMAN
network is to promote transnational use of the NORMAN format for the collection of data on emerging substances [150].

The NORMAN database is widely used, with many contributors (who upload data) from different countries. Currently, it consists merely of a database for storing sample and compound information, without the possibility of uploading data files. However, it warrants inclusion in a discussion of current digital archiving efforts, as a digital archiving option is being developed for the network. Although still in the test phase, this option (the Digital Sample Freezing Platform, DSFP) [128] will be accompanied by a DCT for non-target screening. In its current state, data in mzML format and the DCTs are inserted into the DSFP database. In addition to the mzML files, and as part of the DCT, a peak picking operation is performed and is followed by a retention-time normalization. The database will be fully searchable for individual compounds by specifying a mass of interest and a retention index. The compounds can then be searched in all chromatograms across all countries. In addition, interactive maps will be available. These can be used, for example, to show the detection of compounds across Europe, thereby allowing the determination of sources. In the current test phase, only LC-MS data are included, but the inclusion of GC-MS data is planned.

An implementation as presented in the DROP met database, MetabolomeExpress or the DSFP would allow the “actual” digital archiving of samples and might be a good alternative to ESBs. This would improve both the sample and data quality by avoiding effects of prolonged sample storage and allowing worldwide use of the data. In addition, MetabolomeExpress and the DSFP have a raw-data viewer embedded as an online function and have the potential to enable rapid and efficient data analysis. The NORMAN network has many contributors and is supported by many instrument vendors. Hence, their solution – the Digital Sample Freezing Platform – is a promising tool for digital archiving linked to non-target screening.
Final words

During my PhD project I developed two gas chromatography (GC)-based methods (Paper I) and one liquid chromatography (LC)-based method (future manuscript) for the untargeted analysis of organic contaminants in sewage sludge. The combination of these methods enables coverage of a large part of the chemical domain of analyzable compounds. I used, and will use, those methods for non-target screening of sewage sludge samples, which were collected from 2005 to 2015. Time trends of compounds in these samples were analyzed and several compounds, especially anthropogenic contaminants (e.g., UV-filters from sunscreens), with significant time trends were detected (Paper II and future manuscript). To help identify unknown compounds in complex samples, a newly developed retention-index system for comprehensive two-dimensional gas chromatography (GC×GC; Paper III) was developed. This accurately (within certain limits) determines the retention times of compounds in the second dimension of GC×GC (e.g., apolar × polar stationary phase combination). To increase the use of this index and to help further characterize the compounds, retention index prediction models were developed for GC×GC along with retention-time prediction models for the same data (Paper IV). These models are beneficial for compound characterization as the former can be applied to datasets obtained using different instruments and/or in different studies and the latter can be directly applied to a dataset without calculation of the RIs. In addition, a general method for generating data that can then be used for storage in digital archives was presented. The most important aspect here is harmonization. The data that are stored must be easily understood (e.g., supported by sufficient metadata), easily accessible (e.g., saved in an open file format), and comparable over a large time span.

The tasks associated with future work have been summarized in the conclusions of each chapter. In addition to each of the listed tasks, a time-trend analysis of the LC data must be performed. These data will be further analyzed via non-target screening aimed at identifying the remaining compounds that exhibit significant time trends. The acquired results will be subsequently summarized in a manuscript.

Finally, I hope that this work provided some insight into the detection of unknown compounds and helped to identify those that need further attention, just as sonars and radars help to reveal the hidden parts of icebergs and identify the threats they pose. I hope that I was able to show good methods for evaluating large amounts of data as well as the means and need for creating digital archives.
Acknowledgements

First and foremost, I would like to thank my supervisor, Peter, for giving me the chance to work on this interesting topic, which in turn made me learn to love non-target screening. I am thankful for all the things he taught me, for the many productive meetings and helpful comments throughout the past four and a half years. I would also like to thank my assistant supervisors, Per and Christine, for their valuable input to my work, my manuscripts and my thesis. In particular, I would like to thank Christine for teaching me how to use LC and evaluate the data, and Per for lending me his ears when I needed help and advice.

Moreover, I am grateful to my co-authors and other helpers who worked on the manuscripts. Thanks go to Anders Bignert for teaching me about time trend analysis and answering my many questions regarding statistics, and to Anna Linusson Jonsson and David Andersson for help with MOE, SIMCA and multivariate modelling in general. In addition, I would like to thank Anders, Anna and David for their comments and suggestions for my manuscripts and thesis.

I am grateful to everyone from Environmental Chemistry, and particularly those in my group (including Peter and Christine as well as Kristin and Lisa and former members Matyas and Ivan), who have helped me during the past four and a half years. I am grateful for their support of my work and during meetings, but also for their company during lunches and fikas.

In particular, I would like to thank Darya for her help in the lab on my first paper and for her support in general, Joao for his explanations on databases, Sandra for her help in crises, reading my thesis and her explanations about matrix effects, and Mirva for lending me Silent Spring. My gratitude extends to Lisa and Madeleine who formed the committee for my annual evaluation: I have them to thank for making sure that my PhD stayed on track and that I fulfilled all the goals.

A big “thank you” goes to Jana for her long-lasting friendship, for interesting trips to different places, for her help in creating work schedules, for fashion advices and for being one of the few people to have actually read my complete thesis.

I would also like to acknowledge my friends in Germany for the years of friendship from kindergarten age (Steff) and from high school (Anne, Nico, Steffi, Annika, Kris, Sarah, Miriam and Gaby) which have lasted until now even though I have missed so many parties and gatherings (in German for Nico: Ich danke meinen Freunden in Deutschland für die jahrelange Freundschaft vom Kindergartenalter (Steff) und OS (Anne, Nico, Steffi, Annika, Kris, Sarah, Miriam und Gaby) an bis jetzt, obwohl ich jede Menge Partys und Treffen verpasst habe.)
My gratitude goes to the Casparis (Carmen, Ralf, Vincent, Jost and Judith) for giving me two “extra brothers” and an “extra sister” and for coming here to Umeå to celebrate with me.

A very important and big thank you goes to Alexander. Thank you for your support of my thesis, my work and everything else I have done. Thank you for always being there for me and never complaining when I work too much.

I would also like to thank Alexander’s family for welcoming me into their family and making it so easy to feel comfortable with them.

Finally, I would like to thank my parents and my sister Melanie for their support throughout my life. I especially want to thank my sister for being such a good friend and going with me on vacation to places I picked. Lastly, I would like to apologize, not only to my parents and sister, but also all my aunts, uncles, cousins, grandparents and friends who have complained that I moved away so far: I am sorry that I moved abroad and don’t work at Emsland Group. But I am very grateful that you still supported me and that some of you even visited me here in Umeå.
Bibliography

[10] A part of the image was made by http://www.freepik.com retrieved from https://www.flaticon.com/.


[88] S. Bieri, P.J. Marriott, Generating multiple independent retention index data in dual-secondary column comprehensive two-dimensional


[116] Chemical Computing Group Inc., The Molecular Operating Environment (MOE), (2016) 1010 Sherbrooke St. West, Suite #910, Montreal, QC.


[119] Advanced Chemistry Development UK Ltd (ACD/Labs), ChromGenius.


G. Nowakowski, Open source relational databases and their capabilities in constructing a web-based system designed to support the functioning of a health clinic, Czas. Tech. Autom. 1–AC (2013).


A.J. Carroll, MetabolomeExpress.


Appendix

Non-target screening and time trend analysis of sewage sludge via liquid chromatography – ion mobility spectrometry – mass spectrometry

Experimental part

In the following, the experiments and analysis for the method development and time trend analysis in connection with LC-MS analysis are described.

Materials

Methanol (≥ 99.9% purity), sodium chloride (NaCl) and sodium sulfate (Na₂SO₄) were obtained from Merck KGaA (Darmstadt, Germany). Acetonitrile (ACN, 99.99% purity) was purchased from Fisher Scientific (Loughborough, UK). Formic acid was acquired from Sigma Aldrich (St. Louis, MO, USA). Sodium hydrogen carbonate was obtained from VWR (Leuven, Belgium). Magnesium sulfate (MgSO₄) was purchased from Scharlab SL (Sentmenat, Barcelona, Spain). Triethylamine (> 99%) was obtained from Merck-Schuchardt (Hohenbrunn, Germany). Glass fiber filters (GFFs) with a diameter of 27 mm were acquired from Dionex (Sunnyvale, CA, USA). Zirconia/Silica Beads (2.3 mm diameter) were acquired from BioSpec Products Inc. (Bartlesville, OK, USA). An internal standard (IS) mixture containing 19 deuterated organophosphates, 12 pharmaceuticals (10 deuterated, one deuterated and carbon-13 labeled, and one carbon-13 and nitrogen 15-labeled), nine deuterated pesticides, five deuterated phthalates and nine additional deuterated compounds (triclosan, bisphenol A, octocrylene, benzophenone, oxybenzone, benzothiazole, n-butylbenzenesulfonamide, tetramethyldecynediol, α-tocopheryl acetate) was used to spike samples. The standard solution for matrix effect determination contained four deuterated organophosphates, six pharmaceuticals (five deuterated and one carbon-13 and nitrogen 15-labeled), six deuterated pesticides, three deuterated phthalates, and six other deuterated compounds (triclosan, octocrylene, benzophenone, oxybenzone, n-butylbenzenesulfonamide, tetramethyldecynediol).

Method development

Sewage sludge was obtained from the sewage treatment plant in Umeå. The sludge was freeze-dried and stored in the freezer until use.

For the sample extraction prior to LC-MS analysis four different methods were tested, those being BeadBeater extraction, pressurized liquid extraction (PLE), QuEChERS and ultrasound assisted extraction (USE). Three times (triplicates) 1 g freeze-dried sewage sludge was spiked with 50 µL of the IS mixture, extracted with ACN using the four different methods and the
extraction efficiency as well as matrix effects were determined to compare all four methods. In addition, for each method one blank was prepared and treated the same way.

In the BeadBeater extraction the sludge was divided between three 2 mL BeadBeater vials, resulting in nine vials (three vials for each replicate). To each vial five zirconia/silica beads and 1.5 mL ACN were added. The samples were extracted in a Mini-BeadBeater (BioSpec Products) for 4 minutes at 4000 rpm, then centrifuged for 10 minutes at 13500 rpm and the supernatant was transferred. Fresh solvent (1 mL ACN) was added followed by extraction and centrifugation. Once more, the supernatant was transferred and the extracts from all three vials belonging to one sample (but not the replicates) were combined.

For the PLE, the samples were first homogenized with mortar and pestle and mixed with pre-baked (550°C) sodium sulfate (approximately 3:2, w/w). The extraction was carried out on a Dionex™ ASE™ 350 system using 22 mL stainless steel extraction cells. The samples were extracted with in total approximately 50 mL ACN under the following conditions: 120°C, 5 min static extraction, 3 extraction cycles, 100% flush volume, and 60 sec nitrogen purge. A GFF was placed in the bottom of each cell followed by the homogenized, spiked sludge and glass beads to fill the cell. The cells, GFF and glass beads were pre-cleaned in the PLE using acetone under the following conditions: 100°C, 1 min static extraction, 3 extraction cycles, 100% flush volume and 60 sec nitrogen purge. The blank consisted of pre-cleaned sand.

The QuEChERS (Quick, Easy, Cheap, Effective, Rugged and Safe) method as described by Anastassiades et al. [9] was slightly modified. The sludge was shaken with 10 mL ACN. To each tube 4 g MgSO4 and 1 g NaCl were added. The sample was shaken, centrifuged (10 minutes at 2000 rpm) and the supernatant was transferred.

Finally, for the USE sludge and 10 mL ACN were mixed in a centrifuge tube and sonicated for 10 minutes. The extracts were centrifuged (10 minutes at 2000 rpm) and the supernatant was transferred. The same procedure was repeated two more times with fresh solvent (each 5 mL ACN) and all supernatants were collected and combined.

Prior to analysis, the volume of all extracts was reduced to a few mL using a rotary evaporator, and to a final volume of 1 mL using a stream of nitrogen. All samples were analyzed in MS mode in positive and negative ESI. Additionally, all samples were analyzed in all-ion MS/MS mode in positive ESI for comparison with in house libraries.

To determine the matrix effect for each extraction method a standard dilution was prepared in solvent and in matrix (for each method separately). A four-point curve was produced by adding 0, 2, 4 or 8 µL of standard solution to 100, 98, 96 and 92 µL of extract. For each compound in each matrix the slope of the curve given by the amount of standard solution added and the
area was calculated and divided by the slope of the same curve in solvent. The resulting value was the matrix effect. A value lower than one (i.e. lower than 100%) indicates matrix suppression while a value above one (i.e. above 100%) indicates matrix enhancement. The recovery calculated from the calibration curve (in solvent) was divided by the matrix effect value to obtain the extraction efficiency.

In addition to the extraction methods different extraction conditions were tested and compared. These conditions include the pH and water content of the samples. For the former, four samples of each 0.25 g freeze-dried sewage sludge were spiked with the IS mixture, the solvent was evaporated, and 0.5 g milli-Q water was added. The sludge was extracted using the BeadBeater method as described above. Each sample was extracted first using three times 1 mL ACN. Afterwards, two samples were extracted two times with 1 mL ACN with 0.1M formic acid (acidic pH) and the other two samples were extracted two times with 1 mL ACN with 5% triethylamine (basic pH). The neutral, acidic and basic fractions were analyzed by LC-MS (ESI+ and ESI-) each separately. To evaluate the effect of hydration on the extraction efficiency, the internal standard recoveries from the neutral extraction of the acidic/basic extraction test (hydrated) were compared with the recoveries obtained from the BeadBeater extraction during the method development (no hydration step). Furthermore, a non-target feature extraction was performed to compare the influence of the hydration on the detection of non-target compounds.

**Sample treatment for time trend study**

Sewage sludge samples were obtained from the environmental specimen bank (ESB) at the Swedish Museum of Natural History. The samples were collected in 2005 and from 2007 to 2015 each year in autumn at a sewage treatment plant in Stockholm (Henriksdal), freeze-dried, and stored in the ESB at -25°C.

Prior to analysis the samples were thawed at room temperature and then homogenized with mortar and pestle. For each year three replicates were prepared consisting of each 0.25 g freeze-dried sewage sludge. The extraction was performed in 2 mL BeadBeater vials. Each sample was spiked with the IS mixture, mixed and the solvent was evaporated. Afterwards, 0.5 mL milli-Q water was added, the extract was mixed and let to soak for approximately 15 minutes. Before the extraction, five zirconia/silica beads and 1 mL ACN were added to each vial. The samples were extracted in a Mini-BeadBeater for 4 minutes at 4000 rpm followed by centrifugation for 10 minutes at 13500 rpm. The supernatant was transferred, and the samples were extracted two more times with 1 mL ACN and afterwards two times with 1 mL ACN with 0.1M formic acid. The extracts from the acidic extraction were combined and mixed with approximately 0.2 g sodium hydrogen carbonate to neutralize the
extracts. Afterwards, the neutralized acidic fraction was filtered through sodium sulfate and combined with the neutral extract.

Prior to adding sodium hydrogen carbonate to the extracts possible degradation or transformation effects were studied by comparing a neutralized acidic extract and a non-neutralized acidic extract analyzed by LC-MS. After performing a feature extraction on both chromatograms, the results were compared but no big differences were visible.

As a final step, the volume of all samples was reduced to 1 mL using a rotary evaporator and a stream of nitrogen for final adjustment.

**Sample analysis**

The analysis was carried out on an Agilent 1290 Infinity LC coupled to a 6560 ion-mobility Q-TOF (Quadrupole – time of flight; Agilent Technologies, Santa Clara, CA, USA). A volume of 4 µL was injected and a C18 column (2.1 mm × 150 mm, Phenomenex) with 2.6 µm particles was used for separation with a C18 pre-column. The column temperature was 40°C. Mobile phase flow was 0.5 mL/min and solvents were methanol with 0.1% formic acid and milli-Q water with 0.1% formic acid and methanol and milli-Q water for the positive and negative ionization, respectively. The gradient started at 10% methanol and increased to 95% in 50 minutes. The final composition was held for 10 minutes. Reference compounds for mass accuracy measurements were delivered with a flow of 0.8 mL/min using a split of 1:100.

Electrospray ionization (ESI) was used as interface to the MS. In positive ESI the settings were as follows: the source temperature and sheath gas temperature were 225 and 350°C, respectively. The drying gas and sheath gas flow were 13 and 12 L/min during the initial method development and 5 and later, after further optimization, 12 L/min for the time trend samples, respectively. The nebulizer was operated at 50 psig (method development) and 30 psig (time trend samples), while VCap and Nozzle voltage were 3500 V and 1000 V for method development and 4500 and 500 V for the time trend samples, respectively.

In negative ESI mode the source was operated at 225°C for the method development experiments and 150°C for the time trend samples. The sheath gas temperature was 350°C in both cases. Drying gas and sheath gas flow were 13 and 12 L/min, respectively. The Nebulizer was operated at 50 psig and VCap and Nozzle voltage were 3500 V and 1000 V for method development experiments and 4400 V and 1400 V for time trend samples, respectively.

The instrument was initially operated in IMS-MS mode for the time trend samples (for positive as well as negative ESI) using an m/z range of 50-1700 and a frame rate of 1 frame/s. The maximum drift time, fill time and trap release time were 45 ms, 40 ms and 150 µs, respectively. In addition, IMS-MS/MS (all-ion MS/MS mode) data were collected in positive ESI. In these experiments, the frame rate was changed to 2 frames/s, while all other
parameters were the same. The collision energy was defined in relation to the drift time, since larger molecules require a higher energy to fragment and smaller molecules would fragment too much and not yield characteristic fragments using a high energy. The collision energy was ramped (in relation to the drift time) as follows: from 5 to 40V (0-20 ms) and then from 40 to 90V (20-45 ms). The data acquisition speed in the TOF was 4 spectra/s and 3 spectra/s for MS and MS/MS experiments, respectively. All MS/MS experiments without IMS measurements were performed using following collision energies: 10, 20 and 40V, while the other parameters were the same as before. As stated above, method development samples were analyzed in MS mode in positive and negative ESI and, in addition, using all-ion MS/MS in positive ESI. The time trend data acquisition experiments are explained below.

Data analysis for time trend samples

In the data processing two approaches were carried out. The first approach aimed at identifying all time trends whereas the second approach first identified suspects within the complete data set and afterwards detected time trends within the pool of suspects. While the first approach was carried out using data collected in positive and negative ESI, the second approach was, so far, only performed on data recorded in positive ESI mode.

For the first approach the data processing was done using the “Find feature” function of the Agilent MassHunter IM-MS Browser (version B.08.00) and the recursive workflow of the Agilent MassHunter Profinder (version B.08.00). Due to large data files the recursive search had to be done batchwise, i.e. for each replicate batch separately. The charge state was limited to maximum two charges whereas the peak height was set to a minimum of 1500 counts in both the IM-MS Browser and Profinder processing. The mass window and retention time window were ±(15 ppm + 2mDa) and ±0.3 min, respectively. The data were aligned using the Agilent Mass Profiler Professional (MPP) software. Subsequently, the aligned data were stepwise reduced by removing peaks with intensities lower than three times the blank values, removing peaks that appear in less than two out of three replicates, and removing peaks in samples from fewer than four of the 10 covered years. The final remaining peaks were subjected to a time trend analysis. The exact masses of compounds that had a significant trend were compared to the compounds previously detected using GC-MS to tentatively identify some compounds.

In the second approach data were processed using the “Find feature” function of the Agilent MassHunter IM-MS Browser (version B.08.00). The charge state was limited to maximum two charges whereas the peak height was set to a minimum of 1500 counts. Afterwards, the data were aligned using the Agilent Mass Profiler software (version B.08.00). Maximum charge state,
minimum peak height, mass window and retention time window were the same as above (two, 1500 counts, ±(15 ppm + 2mDa), and ±0.3 min, respectively). Next, a suspect screening was performed. Detected features were compared to four different libraries (three Agilent libraries: Water, Extractable Leachable PCDL, Forensic Toxicology (ForTox) and one in-house curated library) using retention time and m/z information and possible formula were obtained. All features/compounds that had a potential match were subjected to a data reduction and time trend analysis as described above. In addition, one replicate of each year was analyzed using IMS-MS/MS in all-ion MS/MS mode to obtained MS/MS spectra and a suspect screening using the same four libraries was performed. Compounds that showed a significant trend in the time trend analysis performed on the IMS-MS experiment data were searched in the IMS-MS/MS dataset to get, in addition to retention time and m/z match, a confirmation through MS/MS spectra. In the final step, the identification of compounds was confirmed by performing a targeted MS/MS experiment using one replicate per year and reference standards were used, where possible.

In both approaches the retention time information were used in an extra step to exclude hits that do not match based on a simple linear regression between logP values and retention times obtained from 350 previously analyzed compounds.