DOCTORAL THESIS

Analytical Quality by Design in Environmental Analysis

Jelena Jurjeva

TALLINNA TEHNIKAÜLIKOOL
TALLINN UNIVERSITY OF TECHNOLOGY
TALLINN 2025

TALLINN UNIVERSITY OF TECHNOLOGY
DOCTORAL THESIS
65/2025

Analytical Quality by Design in Environmental Analysis

JELENA JURJEVA

TAL TECH PRESS

TALLINN UNIVERSITY OF TECHNOLOGY
School of Engineering
Department of Energy Technology
This dissertation was accepted for the defence of the degree 12/08/2025

Supervisor: Dr. Mihkel Koel
School of Engineering
Tallinn University of Technology
Tallinn, Estonia

Opponents: Prof. Miguel de la Guardia Cirugeda
Department of Analytical Chemistry
University of Valencia
Valencia, Spain

Dr. Riin Rebane
Institute of Chemistry
University of Tartu
Tartu, Estonia

Defence of the thesis: 09/09/2025, Tallinn

Declaration:
Hereby I declare that this doctoral thesis, my original investigation and achievement, submitted for the doctoral degree at Tallinn University of Technology has not been submitted for doctoral or equivalent academic degree.

Jelena Jurjeva

---
signature

European Union
European Regional
Development Fund

Investing
in your future

ISSN 2585-6898 (publication)
ISBN 978-9916-80-366-0 (publication)
ISSN 2585-6901 (PDF)
ISBN 978-9916-80-367-7 (PDF)
DOI https://doi.org/10.23658/taltech.65/2025
Printed by Koopia Niini & Rauam

Jurjeva, J. (2025). *Analytical Quality by Design in Environmental Analysis* [TalTech Press]. https://doi.org/10.23658/taltech.65/2025

TALLINNA TEHNIKAÜLIKOOL
DOKTORITÖÖ
65/2025

Analüütilise disainiruumi kontseptsiooni rakendamine saasteainete pinnases määramise meetodi väljatöötamisel

JELENA JURJEVA

TAL TECH
KIRJASTUS

Contents

List of Publications

The list of author's publications, on the basis of which the thesis has been prepared:

I **Jurjeva, J**.; Koel, M. (2019). The chemometric approach to identification of residual oil contamination at former primitive asphalt pavement plants. Oil Shale, 36(3), 410-430. https://doi.org/10.3176/oil.2019.3.04

II Koel, M.; **Jurjeva, J**. (2022). Implementing greening into design in analytical chemistry. Talanta Open 6, 100136. https://doi.org/10.1016/j.talo.2022.100136

III **Jurjeva, J**.; Koel, M. (2025). Application of the concept of design space in method development for the determination of contaminants in soil. Proceedings of the Estonian Academy of Sciences, 74(1), 71-81. https://doi.org/10.3176/proc.2025.1.07

Author's other publications

Jurjeva, J.; Koel, M. (2022). Using Analytical Quality by Design principles in the analysis of petroleum compounds in the soil. Current Topics in Analytical Chemistry 14, 45-59.

Author's Contribution to the Publications

Contribution to the papers in this thesis are:

I The author performed all experiments, analysed the data, and wrote the manuscript.

II The author was designated as a co-author in this paper, who performed experiments and contributed to writing and editing the manuscript.

III The author performed all experiments, analysed the data, and wrote the manuscript.

Introduction

The Analytical Quality by Design (AQbD) approach contributes to achieving an improved method performance to ensure the analytical procedure is clearly defined, reliable, and suitable for its intended purpose. For development of the analytical process related to quality by design, it is important to follow Guideline Q8 (R2) on Quality by Design from the International Conference on Harmonization (ICH), which is defined as "a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management" (ICH Harmonized Tripartite Guidance Q8(R2), 2009). The process of developing an analytical method in a QbD environment is containing several steps - starting with setting of Analytical Target Profile (ATP), which specifies what will be measured, in which matrix, and over what concentration range, following the selection of a proper method; accompanied with definition of necessary robustness of analytical procedure and design of experiments setting the attributes in the development process and in frame of that the selection of parameters for optimal method to be developed (Tome et al., 2019).

In this work, the AQbD approach was chosen for analytical method development (**Study I**). Analytical target in this case is oil spill into the soil and compounds that are spilled, namely 16 parent polycyclic aromatic hydrocarbons (PAHs) and their alkylated analogs, 7 polychlorinated biphenyls (PCBs), and four aliphatic fractions from octane to pentatriacontane. Method to use is gas-chromatography mass spectrometry (GC-MS) where the design of the experiment (DoE) was applied, in order to choose the optimal chromatographic conditions. Subsequently, the set-up of the Design Space (DS) for method development was investigated, including the selection of a suitable GC column.

The method was validated as per the guidelines established by the Nordtest Guide and the International Council for Harmonisation (ICH) (**Study II**). The developed method was validated by the validation parameters (linearity, the limit of detection (LoD), and the limit of quantification (LoQ), within laboratory precision and the measurement uncertainty) evaluated. The Z scores were calculated when using the results of the proficiency test for 16 PAHs and 7 PCBs.

A combination of chemometric and analytical methods was used to identify the sources of PAHs, which could be attributed to the soil pollution in the former primitive asphalt pavement plants (APPs) in Estonia (**Study III**). The identification and classification of oil spills were performed using chemometric techniques, such as the principal component analysis (PCA) and the clustering analysis (CA).

Statistical methods (Pearson's product-moment correlation (PPMC) and the Shapiro-Wilk test) were employed to compare different analytical methods for analysing soil samples polluted with PAHs (**Study IV**). The Analytical Greenness Metric (AGREE) was employed to assess the greenness of the methods considering the twelve principles of green analytical chemistry (GAC) (**Study V**).

Thus, this study developed a robust, rapid, simple, sensitive, and selective GC method for PAHs, PCBs, and aliphatic and aromatic compounds. The reduction of the different pretreatment steps on a single sample and reduced amount of the consumables makes this method greener. The validation of developed method demonstrated that it could be used for controlled testing of the soil and the sediment.

Abbreviations

Acy	acenaphthylene
ANOVA	analysis of Variance
Ant	anthracene
APP	asphalt pavement plants
AQbD	Analytical Quality by Design
B(a)Ant	benz(a)anthracene
B(a)P	benzo(a)pyrene
B(b)F	benzo(b)fluoranthene
B(g,h,i)P	benzo(g,h,i)perylene
B(k)F	benzo(k)fluoranthene
Chr	chrysene
CMAs	critical method attributes
CRM	certified reference material
DiAnt	dibenz(a,h)anthracene
DoE	Design of Experiments
DS	design space
FDA	Food and Drug Administration
Fla	fluoranthene
Flu	fluorene
FMEA	failure mode and effects analysis
GC-FID	gas chromatography-flame ionisation detection
GC-MS	gas chromatography-mass spectrometry
HFO	high fuel oil
HMW	high molecular weight
HPLC	high-performance liquid chromatography
ICH	International Council for Harmonisation
Ind	indeno(1,2,3-cd)pyrene
ISTD	internal standards
LFO	light fuel oil
LMW	low molecular weight
ME	matrix effect
MU	measurement uncertainty
Nap	naphthalene
PAHs	polycyclic aromatic hydrocarbons
PCA	principal component analysis
PCBs	polychlorinated biphenyls
Phe	phenanthrene

Pyr	pyrene
QRM	quality risk management
TPH	total petroleum hydrocarbon
SIM	selected ion monitoring
UCMs	unresolved complex mixtures

1 Literature overview

1.1 Analytical method development

Analytical method development and validation are continuous and inter-dependent tasks associated with research and development, quality control (QC), and quality assurance (Doltade et al., 2019). All developed analytical methods need to be validated to ensure that each measurement of the analyte content in the sample during routine analysis is close to the true values. International guidelines have been established for the validation of analytical methods by organisations such as the Food and Drug Administration (FDA), European Medicines Agency (EMA), International Union of Pure and Applied Chemistry (IUPAC), and Association of Official Analytical Chemists (AOAC) (Moosavi and Ghassabian, 2017). The primary purpose of method development is to obtain an analytical procedure fit for the intended purpose, and which can measure the attribute(s) of the material with the required specificity/selectivity, accuracy, and precision over the reportable range (ICH harmonised guideline Q14, 2023).

ISO 17025 lays out requirements for the operation of laboratories to deliver consistent and high-quality results that fulfil the expectations of the client. According to this standard, a method should be validated when it is necessary to demonstrate that its performance characteristics are adequate for use for a particular purpose (United nations industrial development organization, 2009). The laboratory should decide which performance characteristics require investigation and what level of detail the investigation must entail. The most important performance characteristics that are typically included in a validation study are the selectivity, working range and linearity, limit of detection/limit of quantification (LoD/LoQ), precision, accuracy, ruggedness, and measurement uncertainty (MU) (Eurachem Guide, 2014). Changes dependent on the analyst, new equipment or equipment part (new column, new detector), or a new reagent batch are subject to control variation and validation parameters (Hibbert, 2007).

Chromatographic methods such as gas chromatography-flame ionisation detection (GC-FID), gas chromatography-mass spectrometry (GC-MS), and high-performance liquid chromatography (HPLC) are typically used for identification of oil compounds in soil. GC-MS has emerged as an effective and sensitive method for tracking the presence of oil pollutants in soil. The obtained mass spectrum can be used to resolve compounds that are spectrally similar yet chemically independent. However, different factors interfere with the accurate and efficient identification of each compound in the dataset. These factors include baseline drifts between samples, shifts in the timing of the elution peak of individual compounds, co-elution of compounds contaminating specific mass spectra, and chemical and electronic background noise that masks the presence of low-concentration analytes (Murphy et al., 2012). In addition, when the co-eluted molecules possess the same mass and same fragmentation pattern (e.g., alkylated polycyclic aromatic hydrocarbons, PAHs), MS provides incorrect identification. It is common practice to employ single ion monitoring (SIM) or single ion extraction (SIE) from full scan data to clean and extract complex GC-MS spectra. SIM can help eliminate background noise and impurities with other mass fragments (Zhao et al., 2014; Christensen and Tomasi, 2007).

1.2 Analytical Quality by Design

Analytical Quality by Design (AQbD) is employed in the development and optimisation of analytical methods. Figure 1 shows a flowchart of the process of analytical method development in an AQbD environment.

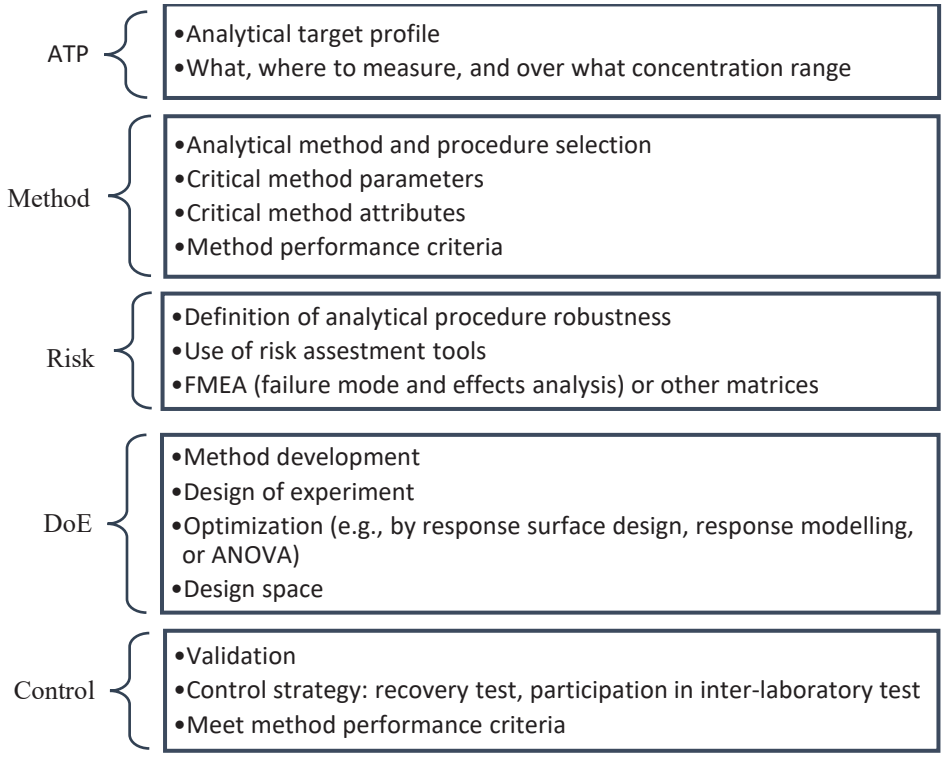

Figure 1. Flowchart depicting the process of analytical method development in an AQbD environment.

The AQbD approach commences with defining the goal of the proposed method, also termed as the analytical target profile (ATP), the critical method parameters (CMPs), whose influence will be evaluated, and the critical method attributes (CMAs), or the variables to be monitored that will confirm the reproducibility and quality of the method. Risk assessment is related to sample preparation for further analyte determination, focusing on data analysis. Subsequently, further potential variables such as noise variables can be identified, which can be evaluated by measurement system analysis and instrumental parameters, and assessed by design of experiment (DoE) strategies. The experiments commences with the scouting phase, which aims to define the parameters that are critical to the process. To guide the choice of experiments and assist in their efficient execution, DoE is an important component of the AQbD approach (Ellwanger, 2020; Araújo, 2021). From the DoE results and their interpretation, the Design Space (DS, the region where all input parameters can vary without altering the quality of the product) of the analytical method is obtained. An analytical procedure

control strategy is then designed to ensure that the analytical procedure is suitable for the intended purpose during routine use throughout its lifecycle (ICH Harmonized Tripartite Guidance Q8(R2), 2009).

1.3 Analytical target profile

Establishing the ATP is the first step in the development of an analytical method by the AQbD approach. The ATP drives the choice of analytical technology, sets the criteria that define what will be measured, in which matrix, and over what concentration range, and defines the required performance criteria of the method. The first stage in the life cycle of an analytical procedure includes knowledge gathering, establishing the ATP, understanding the effects of various process parameters on the procedure performance, optimising them, and defining the initial control strategy. The ATP is independent of any specific analytical procedure; hence, only after defining the ATP can the analyst choose the analytical technique. Once a technique is selected, the ATP serves as a foundation to derive the analytical procedure attributes and performance criteria for validating the procedure. The ATP facilitates ongoing monitoring and continuous improvement of the analytical procedure. The ATP is maintained over the lifecycle and can also be used as a basis for lifecycle management to ensure that the existing, revised, or new analytical procedure remains fit for the intended purpose (Rozet et al., 2013; Szoleczky at al., 2024; ICH harmonised guideline Q14, 2023).

In environmental analysis, the establishment of an ATP in any environment (air, water, or soil) inherently depends on the threshold values of the contaminants established by relevant authorities. Compliance control requires the use of reliable and reproducible methods of sampling, sample pre-treatment prior to analysis, and analytical measurements to produce results that are legally valid. The ATP is similar to the environmental parameter profile, which prospectively summarises the requirements associated with the quality attributes of the measurement that the analytical procedure needs to meet (Jurjeva and Koel, 2025; Jackson et al., 2019).

1.4 Critical method parameters and critical method attributes

CMPs are essential elements in method development that significantly impact the quality of analysis. CMPs can include the conditions of extraction in sample preparation, GC-MS instrument parameters, and gradient program in chromatography, all of which have a significant impact on the performance of the analytical procedure. The influence of each CMP on the performance characteristics such as the sensitivity, selectivity, precision, accuracy, linearity, LoQ, and LoD should be assessed during method development. To this end, the required method development experiments and their results are evaluated to establish the acceptable criteria for each critical parameter. The subsequent step is to optimise the analytical procedure conditions to achieve the best compromise among the sensitivity, selectivity, and other performance criteria. CMAs can include the run time of the analysis, resolution of the critical pair, precision of the analytical method, and lower limit of quantification of the analytical method (Jurjeva and Koel, 2025; Beg et al., 2015).

1.5 Risk assessment management

Risk communication should be employed to support the continuous improvement of the performance of the analytical procedure throughout its lifecycle. The AQbD approach requires a good understanding and control of the analytical method based on quality risk management (QRM) (ICH harmonised guideline Q9(R1), 2023; Koel and Jurjeva, 2022).

The use of QRM is encouraged in the development of a robust analytical procedure to mitigate the risks of poor performance and the reporting of incorrect results. A risk assessment is typically performed early in the development cycle of the analytical procedure, and this assessment is updated with the increasing availability of information. The analytical method can be structured in a flowchart that highlights the main steps of the procedure, from sample preparation and instrumental settings to final data analysis. This helps determine the critical attributes related to product quality as well as the analytical method parameters that should be examined during method development. Different risk assessment matrices are used to identify, evaluate, and manage risks. Failure mode and effects analysis (FMEA) helps reveal potential failures in processes to prevent them from occurring, or mitigates their impact by identifying their potential location of occurrence and impact. The risk assessment tools described in ICH Q9 can be used for the following purposes:

- Identify analytical procedure parameters and their potential impact on the performance. Visualisation can be achieved with the help of a fishbone or Ishikawa diagram (Figure 2), which identifies potential factors and distributes factor-related risks into categories associated with instrumentation, materials, methods, measurements, laboratory environments (temperature, relative humidity, and light), and human factors.
- Assess the potential impact of analytical procedure parameters on the performance of the procedure.
- Identify and prioritise analytical procedure parameters to be investigated experimentally.
- Obtain information on the need and extent of ongoing monitoring as part of a risk review (ICH harmonised guideline Q14, 2023; Kardosa et al., 2021).

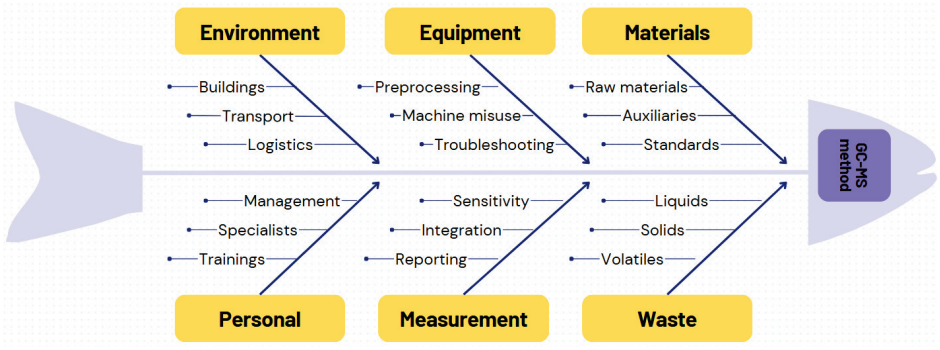

Figure 2. Fishbone diagram for risk assessment in an AQbD environment.

1.6 Design of Experiments

DoE is a statistical technique to screen factors, plan, conduct, analyse, and optimise the conditions of the system to be used as well as interpret experimental data (Araújo et al., 2021). The experiments commence in the scouting phase, which aims to define the parameters that are critical to the process. DoE is an important component of the AQbD environment to guide the choice of experiments and execute them more efficiently. After statistical analysis, the DoE results can help determine the relationships among factors affecting a process and the output of such a process (Ellwanger et al., 2020). Full factorial designs incorporate every possible combination of factors at the designated levels, and there are L^k combinations of k factors at L levels. Full factorial design is employed in industrial settings to identify the most active factors that affect responses or processes. However, the high cost of performing a full factorial is the reason why fractional factorial designs are used (Alomair et al., 2020). Fractional factorial designs are a specific subset of a full design. Other experimental designs include the Plackett–Burman, central composite, Box–Behnken, Doehlert, D-optimal, G-optimal, and mixture designs (Hibbert, 2012).

1.7 Design space in the development of analytical procedures

From the DoE results and their interpretation, the DS of the analytical method is obtained. Working within the DS is not considered a change. The DS allows for responsive action when “process drifts” are observed. These drifts could be caused by changes in reagents or materials for analysis, like the presence of different types of soil or water. These factors cannot be actively controlled to maintain the environmental parameters at the intended quality level. Additionally, in analytical procedures, it is possible to move within the DS without the need to initiate a regulatory post-approval change process (Hakemeyer et al., 2016; Ng and Rajagopalan, 2009). In the framework of method development, the DS can be considered as a zone of theoretical robustness, as no drastic changes in the levels of CMAs are typically observed. Hence, to define an analytical DS, an appropriately selected number of factors (here, CMPs) that affect the analytical technique under development should be studied simultaneously. Generally, the CMPs are obtained from a risk analysis and prioritisation strategy (Rozet et al., 2013).

1.8 Method validation

Validation of an analytical procedure involves a series of laboratory studies to confirm that the analytical procedure is suitable for its intended purpose. It aims to strike a balance among costs, risks, and technical possibilities. For validation, the laboratory should factor customer and regulatory requirements, available tools, satisfactory laboratory design, stable environmental conditions, and suitable QC procedures. Moreover, the laboratory should identify and evaluate relevant performance characteristics and verify them against the analytical requirements. The effects of changes in other parameters, such as the analyst, instrument, or laboratory environment, should also be verified. Validation is completed with a conclusion and statement about whether or not the analytical requirement is met. If not met, further method development is necessary (Eurachem Guide, 2014).

Statistical analysis of the data obtained during method validation should be performed to demonstrate the validity of the analytical method. There are many statistical parameters and tests that are essential for summarising data and deriving objective judgements on differences between sets of data (significance testing). The most important parameters are the mean, standard deviation, relative standard deviation, confidence intervals, regression, and correlation. Other statistical tools used in method validation include comparative studies employing Student's t-test, Fisher's test, analysis of variation (ANOVA), and assessment of outliers. These calculations are characteristically performed using statistical software packages such as SPSS, R, Excel Stat, and Minitab. Analysts should be able to evaluate the precision, bias, linear range, LoD, LoQ, and MU using statistical methods (Eurachem Guide, 2014; Belouafa et al., 2017).

1.9 Analytical procedure control strategy

The control strategy for an analytical procedure comprises a set of controls derived from the current understanding of the analytical procedure, including development data, risk assessments, robustness, and prior knowledge. This control strategy should be defined before validation and should be confirmed after the validation is complete (ICH harmonised guideline Q2(R2), 2023). The control strategy for an analytical procedure includes analytical procedure parameters that require control. The analytical procedure should describe the steps necessary to perform each analytical test, related to the sample, reference materials and reagents, sample and control preparations, use of the apparatus, generation of the calibration curve, number of replicates, and use of the formulae for calculating the reportable results. Conducting an inter-laboratory proficiency test is a part of the control strategy. Another methodology that can be implemented as a control strategy for analytical methods is to use statistical control charts such as Shewhart X and R charts. Thus, out-of-control methods can be detected efficiently, and corrective actions can be realised by tracking the daily performance of analytical methods on such charts (Rozet et al., 2013).

1.10 Soil pollution and health

The AQbD approach used in pharmacy can also be applied in environmental analysis, specifically, soil analysis (Jurjeva and Koel, 2022). Soil is an important component of the earth’s biosphere, playing a critical role not only in the production of food and fibre, but also in the maintenance of local, regional, and global environmental quality (Dorana and Zeiss, 2000). Healthy soils are critical for supporting human health. Depending on the chemicals involved, soil pollutants can affect various organs, as well as immune, reproductive, nervous, and cardiovascular systems (European Environment Agency). Soils become polluted in various ways, including through seepage from a landfill, discharge of industrial waste into soil, percolation of contaminated water into soil, rupture of underground storage tanks, excess application of pesticides, herbicides or fertiliser, and solid waste seepage (Ashraf et al., 2013).

Currently, approximately 80% of land is contaminated/polluted by products of a petroleum origin (e.g., hydrocarbons or solvents), which are used as energy sources in the oil industry, as well as chemicals. Various pollutants affect soil and subsoil, such as

fuel and oil products, hydrocarbon residues, crude oil, and other products from industrial operations (e.g., saturated and unsaturated aliphatic hydrocarbons, and monocyclic and polycyclic aromatics) (Marinescu et al., 2010). The most commonly analysed components from soil for the identification of oil pollution are PAHs, volatile organic compounds (benzene, toluene, ethylbenzene, and three xylene isomers), polychlorinated biphenyls (PCBs), heavy metals, total petroleum hydrocarbons (TPHs), and the hydrocarbon oil index (HOI). The major source of PAHs is the incomplete combustion of organic material such as coal, oil, and wood. In soil or sediments, PAHs tend to adsorb strongly onto suspended particulate matter (Jefimova et al., 2014; Retnam et al., 2013; Lau et al., 2010). The presence of organic pollutants such as PAHs in the soil used for crop production is a major concern in modern agriculture. Owing to the toxic nature as well as mutagenic, cancerogenic, and teratogenic activities of some compounds from PAH groups, their occurrence in soil has numerous disastrous consequences, such as economic losses (decreased yield), hazards to human health, and toxic effects on flora and fauna. Human health is at risk as these pollutants can migrate to plant tissues, thereby contaminating food or feed. The source of PAH contamination in soil can be pyrogenic (i.e., caused or produced by combustion or by the application of heat) or petrogenic (i.e., related to the origin or formation of rocks) (Włóka and Placek, 2015; Pongpiachan et al., 2018; Dudhagara et al., 2016). The USA EPA has listed 16 PAH compounds as priority pollutants, including naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, chrysene, benz(a)anthracene, benzo(b)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene, benzo(g,h,i)perylene, dibenz(a,h)anthracene, and indeno(1,2,3-cd)pyrene (Kumar et al., 2011).

The HOI is defined as the total amount of all hydrocarbons extracted from a sample by a non-polar solvent (e.g., hexane) and then eluted between n-decane and n-tetracontane on an apolar capillary GC column. Mineral oils are combinations of paraffins, napthenes, and aromatic oils. Mineral oils are petroleum-based hydrocarbons, which can be found in diesel, kerosene, home heating oils, heavy fuel oils, transformer oil, and lubricants. These hydrocarbons can contaminate water or be consumed by organisms that can enter the human food chain (Koch et al., 2005; Konečný et al., 2003; Agilent Technologies).

PCBs comprise a class of 209 individual compounds. In the past, they were used in several applications, including as hydraulic fluids and dielectric fluids in transformers, and as additives in paints, paper, sealants, and plastics. PCBs have been widely used as additives in thermal oils or transformer oils owing to their non-flammability, chemical and thermal stability, high boiling point, and low electrical conductivity. Exposure to transformer oils containing PCBs is known to be associated with numerous health issues, including cancer, neurological disorders, and adverse effects on the immune system, reproductive system, and endocrine system as endocrine disruptors. PCB concentrations are often reported as the sum of seven congeners (PCB7, IUPAC No. 28, 52, 101, 118, 138, 153, 180) or as the 'total PCB' (Faraday et al., 2024; Kalbe et al., 2019).

There are studies in the literature on the development of methods for the quantification and identification of PAHs, (Yamada et al., 2009; Volk and Gratzfeld-Huesgen, 2011) PCBs (Krzemień-Konieczka and Buszewski, 2015), and aliphatic and aromatic fractions (Sadler and Connell, 2003) in soil, along with method validations.

Different factors such as the type of sample, concentration of target compounds, required sensitivity, and available instrumentation influence the choice of the method. Further, the methods differ in terms of sample preparation steps (e.g., sample clean up, filtration, concentration, evaporation) and extraction type (e.g., Soxhlet extraction, sonication or ultrasonic treatment, mechanical agitation, accelerated solvent extraction (ASE), solid phase microextraction (SPM)). The selection of a suitable extraction method is challenging as the efficiency of extracting PAHs from soils and sediments is influenced by several factors such as the soil moisture, PAH content in samples, and soil texture (Lau et al., 2010). The extraction solution and its volume are important in determining the greenness of the method. Typically, acetone, hexane, or dichloromethane is used as the extraction solvent. Dichloromethane can extract many compounds; however, some countries have restrictions on working with dichloromethane owing to its potential health hazards (Swedish Chemicals Agency). The analytical instrumentation methods for determining contaminants from soil include HPLC with UV/fluorescence detectors and GC-MS. The most commonly used methods for PAH analysis in soils are chromatographic techniques with MS detection, including HPLC-MS and GC-MS. In most cases, GC-MS is the preferred instrumentation method, as it can be used for most parameters and required LoQs (Włóka and Placek, 2015; Adeniji et al., 2018).

1.11 Chemometric approach in oil identification

The chemometric approach with chromatographic methods can help identify and classify soil based on the type of contamination. It is used to classify and recognise similarities and differences in data, correct signals obtained from the analytical instrument, and control the statistical process (Mali et al., 2017; Singh et al., 2013). In PCA, data analysis can be performed in the matrix form (rectangular table) consisting of *n* rows and *m* columns. Each row corresponds to an object, for instance, a sample name or number, while each column corresponds to a particular feature of the object, such as the physical or chemical properties of the sample. PCA decomposes the matrix into the product of scores matrix and transposed loading matrix plus the residual matrix. The loading plot shows the importance of different variables responsible for clustering in the score plot. The score plot provides information about the relationships between each object, showing groups and outliers (Panchuk et al., 2018; Lubes and Goodarzi, 2017; Gad et al., 2013).

Before PCA, the data are standardised and the Min-Max scaled and autoscaled methods are used. In the Min-Max scaled (normalisation) method for every feature, the minimum value of that feature is transformed into 0, the maximum value is transformed into 1, and every other value is transformed into a decimal between 0 and 1. The Min-Max scaled is typically calculated via the following equation:

$$Xnorm = \frac{X - Xmin}{Xmax - Xmin}$$

To avoid the problem of different scales, the data are often centred, and all values are divided by the standard deviation for each variable. The equation for auto-scaling is expressed as:

$$Xaut = \frac{X - Xavg}{Xstd}$$

(Raschka, 2014; Risvik, 2007).

Cluster analysis (CA) or clustering is a statistical technique employed to sort a set of observations (individuals) into different groups called clusters. The first step of CA is to determine the similarity between objects (e.g., correlation coefficients between samples, the Euclidean distance, squared Euclidian distance, and Manhattan distance). The next step is to link the objects. Two of the most similar objects (e.g., with the highest correlation coefficient) form a group. The main task is to recalculate the numerical similarity values between the new group and the remaining objects. The next step continues to group the data until all objects have joined one large group. The results of hierarchical clustering are usually presented in a dendrogram, which visualises grouping among samples in a two-dimensional space (Nováka et al., 2017; Brereton, 2003).

The Jaccard similarity index combines the two most similar sites to each other. It can be expressed as follows:

$$Sj = \frac{a}{a + b + c}$$

where coefficient *a* is the number of species in sample A and in sample B; *b* is the number of species in sample B, but not in sample A; and *c* is the number of species in sample A, but not in sample B (Oksanen, 2011; Romesburg, 2004). To quantify the distance between the two clusters, single, complete, and average linkages are used.

1. *Single linkage clustering*

As in all CAs of an agglomerative type, the single linkage begins with a matrix of similarity (or dissimilarity) coefficients. First, the most similar pair(s) of samples, or the first clusters, are identified. Next, the second most similar pair(s) of samples, or those with the highest similarity between a sample and the first cluster (whichever is greater), are identified. However, a disadvantage of single linkage clustering is its tendency to produce long clusters.

2. *Complete linkage clustering*

Compared to single linkage clustering, complete linkage clustering often tends to cause opposite extremes by producing very tight compact clusters. This method calculates the similarity measures after new groups are formed, and the two groups with the highest similarity are always joined first.

3. *Average linkage clustering*

This technique avoids the extremes introduced by single and complete linkage clustering. To compute the average similarity between a sample and an existing cluster, the type of average should be precisely defined using unweighted or weighted techniques. In the unweighted technique, the new similarity (*Sab*) is calculated by the following equation. If group A has *Na* objects and group B has *Nb* objects, then:

$$Sab = \frac{NaSa + NbSb}{Na + Nb}$$

The weighted mean value is calculated by:

$$Sab = \frac{Sa + Sb}{2}$$

(Brereton, 2003; Jurjeva and Koel, 2019).

There are studies describing the use of a combination of analytical methods and chemometrics for the identification and grouping of samples. Retnam et al. used a combination of CA and discriminant analysis (DA), and were able to group and discriminate aquaculture sites according to contamination levels. The combination of CA and PCA was effective for identifying PAH sources in aquaculture surface sediments (Retnam et al., 2013). Mali et al. demonstrated that the chemometric approach is advantageous for assessing and modelling contamination patterns of highly polluted areas, and hence, could contribute to effectively monitoring their quality. They analysed the distribution of contaminants using two complementary chemometric techniques, PCA/CA and factorial ANOVA (Mali et al., 2017). Miki et al. employed CA using Ward's method with squared Euclidean distances to identify the sources of parent and alkylated PAHs in sediments with the R language (Miki et al., 2014).

1.12 Green chemistry approach

Since analytical method development also involves the use of toxic and hazardous solvents and the generation of a large amount of waste, it is important to calculate and understand their impact on the environment, health, and general safety. Green analytical chemistry (GAC) has twelve basic principles (Table 1). These basic principles include replacement with greener solvents, miniaturisation of analytical devices, reduced analysis time, and reduced amount of waste generated. Analytical methods should be developed and evaluated by considering these principles. GAC aims to eliminate or reduce the use of hazardous chemicals in analytical processes to improve their environmental and health friendliness without compromising the method performance (Saroj et al., 2017; Yabréa et al., 2020).

Table 1. Twelve principles of GAC (Pena-Pereira et al., 2020).

Number	Principle
1.	Direct analytical techniques should be applied to avoid sample treatment.
2.	The sample size and number of samples should be minimal.
3.	In situ measurements should be performed.
4.	Analytical processes and operations should be integrated to save energy and reduce the use of reagents.
5.	Automated and miniaturised methods should be selected.
6.	Derivatisation should be avoided.

Number	Principle
7.	The generation of a large volume of analytical waste should be avoided and proper analytical waste management measures should be implemented.
8.	Multianalyte or multiparameter methods are preferred versus methods using one analyte at a time.
9.	The use of energy should be minimised.
10.	Reagents obtained from renewable sources should be preferred.
11.	Toxic reagents should be eliminated or replaced.
12.	The safety of the operator should be improved.

GAC encourages the reduced usage of toxic chemicals/reagents, while promoting the use of energy-efficient equipment and generating minimal waste. Recent trends in analytical method development have focused on the miniaturisation of sample preparation devices, development of solventless or solvent-minimised extraction techniques, and utilisation of less toxic solvents. The twelve principles of GAC serve as basic guidelines for improving greenness in an analytical procedure. However, in many scenarios, some additional steps are unavoidable. Therefore, it is important to evaluate the greenness of analytical procedures to assess and, if possible, reduce their impact on the environment and operators (Sajid and Płotka-Wasylka, 2022).

Several metrics have been developed for evaluating the greenness of analytical procedures. Among these, the Analytical Eco-Scale, Green Analytical Procedure Index, Blue Applicability Grade Index (BAGI), and Analytical Greenness (AGREE) metric, are important tools for assessing the greenness of analytical procedures. All these metrics consider different aspects of the analytical procedure to provide the green index of the procedure (Manousi et al., 2023). The AGREE metric system assesses the greenness of analytical procedures based on the SIGNIFICANCE principles. The assessment criteria are obtained from the twelve principles of GAC and are transformed into a unified 0–1 scale. The final score is calculated based on the SIGNIFICANCE principles. The result is a pictogram indicating the final score, performance of the analytical procedure in each criterion, and weights assigned by the user (Pena-Pereira, 2020).

2 Aim of work

The specific aims of this thesis are:

I. Applying the AQbD concept to target compounds (PAHs, PCBs, and aliphatic and aromatic fractions) with the following steps: obtaining the ATP, identifying the CMPs, carrying out a performance assessment, establishing acceptable ranges, and optimising conditions.

II. Validating the robust, rapid, simple, sensitive, and selective GC method for PAHs, PCBs, and aliphatic and aromatic compounds, along with monitoring and control.

III. Using the analytical and chemometric tools for soil contamination identification. PCA and CA are employed to combine the samples into groups. The binary ratios of the PAHs are used to distinguish the potential sources of PAHs in the environmental samples.

IV. Comparing different PAH methods using statistical tools.

V. Evaluating the greenness of the method.

3 Materials and Methods

3.1 Sampling, extraction, and sample preparation

All soil and sediment samples used during study were collected in plastic Rilsan bags or in pre-washed glass jars. The samples were stored in a dry, dark, and cool place (refrigerator) and analysed within one week (calculated from their arrival time in the laboratory). Before weighing the samples, they were homogenised by removing rocks, stones, wooden pieces, and other debris, and shaken or mixed thoroughly. The samples were not dried prior to the analysis. The water content was determined according to ISO 11465:1993 (ISO 11465:1993) and the analysis results were calculated on a dry matter basis.

Paper I describes the method suitable for determining the HOI, C_{10}-C_{40}, and PAH content in soil and sediment samples. First, the samples were extracted (and shaken) with acetone and the extraction solution for 16 h. After extraction, ultrapure water was added, which bound the acetone and separated the extraction solution with the analytes (sample extract). The PAHs were analysed from a layer of the sample extract, which was dried over anhydrous Na_2SO_4. Another part of the extract was purified with a Florisil column, using which the HOI was determined. The HOI in C_{10}-C_{40} sample solutions was determined using GC-FID and the PAHs were analysed with GC-MS in the SIM mode.

Paper III describes the method suitable for determining aliphatic and aromatic fractions of hydrocarbons (C_8-C_{35}), PAHs, and PCBs in soil and sediment samples. The method is suitable for the determination of hydrocarbons in diesel, light fuel oils, lubricant oils, and motor oils. Gasoline (e.g., 95E) could not be determined with this method as it consists of very volatile hydrocarbons that are mostly $<C_8H_{22}$. Fractioning was performed based on mass ions with GC-MS, and not by sample pre-treatment. The aliphatic fraction of hydrocarbons C_8-C_{35} includes all aliphatic hydrocarbons from octane to pentatriacontane without a benzene ring structure in the boiling range of 126-490 °C. The aromatic fraction of hydrocarbons C_8-C_{35} includes all aromatic hydrocarbons from C_8 to C_{35} containing a benzene ring. First, the samples were extracted (shaken) with acetone containing internal standards (ISTDs) for 40 min. Subsequently, hexane and sodium pyrophosphate solution were added, and the samples were shaken for approximately 10 min. Two layers appeared at this point: a hexane mixture in the upper layer with the analytes and a water/acetone mixture in the lower layer. From the sample extract, the aliphatic and aromatic fractions of hydrocarbons, PAHs, and PCBs were analysed using GC-MS.

3.2 Target analytes

The target analytes in this research were 16 EPA PAHs (naphthalene, acenaphthylene, acenaphthene, fluorene, phenanthrene, anthracene, fluoranthene, pyrene, benz(a)anthracene, chrysene, benzo(b)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene, indeno(1,2,3,c,d)pyrene, dibenz(a,h)anthracene, benzo(g,h,i)perylene), the alkylated PAHs (methylnaphthalenes, dimethylnaphthalenes, trimethylnaphthalenes, 1,4,6,7-tetramethylnaphthalene, 1,2,5,6-tetramethylnaphthalene, 1,2-methylanthracene/methylphenanthrene,

methylpyrenes/fluoranthenes, methylchrysenes/benz(a)anthracenes), seven PCBs (PCB28, PCB52, PCB101, PCB118, PCB153, PCB138, PCB180), four aliphatic fractions (aliphatic >C_8-C_{10}, aliphatic >C_{10}-C_{12}, aliphatic >C_{12}-C_{16}, aliphatics >C_{16}-C_{35}), and biphenyls. These compounds were analysed by GC-MS in the SIM mode using two or three ions for each compound for identification (Table 2). The HOI was determined using GC-FID.

Table 2. Ions used for identification.

Compound	**Ions**
o-Terphenyl ISTD	*215 (230)*
Aliphatic C8-C10	71 (85)
Aliphatic C10-C12	71 (85)
Aliphatic C12-C16	71 (85)
Aliphatic C16-C35	71 (85)
Methylnaphthalenes	142 (115, 141)
Biphenyl	154 (153, 152)
Dimethylnaphthalenes	156 (141)
Trimethylnaphthalenes	170 (171)
1,4,6,7-Tetramethylnaphthalene	169 (184)
1,4,6,7-Tetramethylnaphthalene	169 (184)
Methylanthracene	192 (193, 165)
Methylpyrenes/methylfluoranthene	216 (215)
Methylchrysene/methylbenz(a)anthracenes	242 (241)
Naphthalene D8 ISTD	*136 (108)*
Naphthalene	128 (102)
Acenaphthylene D10 ISTD	*162 (164)*
Acenaphthylene	152 (153)
Acenaphthene	153 (152)
Fluorene	166 (165)
Phenanthrene D10 ISTD	*188 (184)*
Phenanthrene	178 (176)
Anthracene D10 ISTD	*188 (184)*
Anthracene	178 (176)
Pyrene D10 ISTD	*212 (208)*
Fluoranthene	202 (200)
Pyrene	202 (200)
Benzo(a)anthracene D12 ISTD	*240 (236)*
Benzo(a)anthracene	228 (226)
Chrysene	228 (226)
Benzo(a)pyrene D12 ISTD	*264 (260)*
Benzo(b)fluoranthene	252 (126)
Benzo(k)fluoranthene	252 (126)
Benzo(a) pyrene	252 (126)
Dibenzo(ah)anthracene D14 ISTD	*292 (291)*

Compound	Ions
Indeno(123cd)pyrene	276 (274)
Perylene-D12	*264 (260)*
Dibenzo(ah)anthracene	278 (276)
Benzo(g h i)perylene	276 (274)
PCB No.: 28	258 (256)
PCB No.: 52	292 (290)
PCB No.: 101	326 (324)
PCB No.: 118	326 (324)
PCB No.: 153	360 (362)
PCB No.: 138	361 (362)
PCB No.: 180	396 (394)
Isopropylbenzene	105 (120)
Propylbenzene	91 (120)
3-4-2_ethyltoluene 1,3,5-trimethylbenzene	105 (120)
1,2,4-trimethylbenzene	105 (120)
1,2,3-trimethylbenzene	105 (120)
1,4-1,2_diethylbenzene	119 (134)
1,2,4,5-1,2,3,5_tetramethylbenzene	119 (134)

3.3 Materials and reagents

The stock solutions used in this work for preparing the calibration standards and ISTDs were purchased from AccuStandard (New Haven, USA), Sigma-Aldrich/Merck (Darmstadt, Germany), Dr. Ehrenstorfer GmbH Reference Materials (Augsburg, Germany), and Chiron Petroleum Reference Standards (Trondheim, Norway). Hexane and acetone were purchased from Honeywell International Inc., Charlotte, USA. Table 3 shows all standards used in **Paper II** with the respective suppliers.

Table 3. Standards and solutions used for method development.

Purpose	Substance	Supplier
Aliphatic hydrocarbons: C_8-C_{35} standards		
Calibration standard	Calibration window defining the hydrocarbon standard (C_8-C_{40} 1000 mg/L of each compound)	AccuStandard
Internal standard	o-Terphenyl, 99%	Agilent
Reference standard	Hydrocarbon window defining standard in chloroform (C_8 - C_{40} 500 mg/L of each compound)	AccuStandard

Purpose	Substance	Supplier
Certified reference material	TPH Sand 1 (CRM372-100G)	Sigma-Aldrich
RT checking solution	Hydrocarbon window defining standard in chloroform (C_8 - C_{40} 500 mg/L of each compound)	AccuStandard
Aromatic hydrocarbons: C_8-C_{35} standards		
Calibration standard	Custom mix, multi-standard alkylated PAHs and aromatics in toluene (100 mg/L of each compound)	Chiron
Internal standard	Naphthalene D-8 for aromatic fractions C_8-C_{16}. Pyrene D-12 for aromatic fractions C_{16}-C_{35}	Dr. Ehrenstorfer
Certified reference material	TPH Sand 1 (CRM372-100G)	Sigma-Aldrich
PAHs		
Calibration standards	PAH mix-9 100 ng/µL in acetonitrile	Dr. Ehrenstorfer
Internal standard	Separate isotope-labelled solid stock standards: Naphthalene D8, Pyrene D10, Perylene D-12	Dr. Ehrenstorfer, Supelco
Reference standard	PAH-Mix 63 (1000 mg/L in toluene)	Dr. Ehrenstorfer
Certified reference material	Soil CRM BAM ERM-U013b	LGC Standards
PCBs		
Calibration standard	PCB Mix-3 (10 ng/µL in iso-octane)	Dr. Ehrenstorfer

Purpose	Substance	Supplier
Internal standards	Pyrene D10	Dr. Ehrenstorfer
Reference standard	PCB Mix-20 (10 ng/µL in iso-octane)	Dr. Ehrenstorfer
Certified reference material	SETOC sample	WEPAL

3.4 Instrumentation and operating parameters

The HOI analysis described in **Paper I** was performed on an Agilent 7890B instrument equipped with a flame ionisation detector and an Agilent 7693 autosampler. The column used for these analyses was an Agilent J&W GC Column DB-1 of dimensions 15 m × 320 µm × 0.10 µm. The target EPA PAH compounds were analysed on an Agilent 6890N GC system equipped with a mass-selective detector (Agilent Technologies 5973), together with an Agilent 7683 series autosampler. The column used for this analysis had dimensions of 20 m × 180 µm × 0.18 µm (ZB-5MS). For both instruments, all samples were injected in the splitless injection mode with helium as the carrier gas.

In **Paper III**, Agilent GC-MS systems 7890B GC and 5977B MS were used for method development. An RXI-XLB column was used, possessing dimensions of 30 m × 250 µm × 0.25 µm. GC-MS was carried out under the following conditions: pulsed pressure spitless injection; column flow, 1.8 mL min^{-1}; oven temperature program: initial temperature, 40 °C, hold time, 3 min; 40 °C min^{-1} to 160 °C; 25 °C min^{-1} to 315 °C; hold time, 3.5 min, 40 °C min^{-1} to 340 °C; hold time, 10 min; interface temperature, 320 °C.

3.5 Three-level Design of Experiment (Study I)

This study captured the projection properties of fractional factorial designs with three levels, namely, the low, intermediate, and high levels. The three-level design was implemented to investigate the optimal levels for the initial oven temperature, final oven temperature, and inlet temperature of the gas chromatograph. DoE was performed using the Custom Design Platform in JMP software. The chromatographic peak areas and resolution between the peaks were compared for different instrument parameters.

3.6 Method validation, statistical methods, and control strategy (Study II)

The within-laboratory validation was performed according to Eurachem, the GUM guide, the Nordtest guide and ICH guideline (Eurachem Guide, 2014; GUM JCGM 100:2008, 2008; Nordtest project, 2012; ICH harmonised guideline Q2(R2), 2023). The optimised method was ratified in terms of the linearity, LoQ, LoD, precision, accuracy, and MU. The effect of the matrix was also evaluated. Statistical tools, namely the lack-of-fit test, T-test statistics and null hypothesis, and one-way ANOVA supported the validation. For

the control strategy, the performance characteristics were evaluated through inter-laboratory ring tests.

Linearity and working range

The linearity was assessed by a visual inspection of the plot, supported by statistical results (the lack-of-fit test) and residuals plot from the linear regression. For each analyte, these calculations were performed according to linear calibration models. The residuals were normally distributed around zero in the linear range. A calibration curve was generated for each analyte and used to calculate the concentration of the analyte in unknown samples. To determine the linearity of the calibration curve, a set of calibration standards at concentrations of 10, 50, 500, 2000, 5000, 10000, and 20000 µg/L was prepared for determining the PAHs, and at concentrations of 0.3, 3, 15, 60, 150, and 600 µg/L for determining the PCBs. The linearity of the aliphatic compounds was determined in the ranges of 5000 to 50000 µg/L, and for alkylated PAHs (aromatic compounds), in the ranges of 200 to 5000 µg/L for the individual compounds. All measurements were performed in triplicate to confirm the homoscedasticity and heteroscedasticity of the data.

Determination of LoD and LoQ

The LoD and LoQ were determined separately for each matrix (soil and sediment) using the following equations (Eq. (1, 2)):

$$LOD = Xavg+3*STDEV \quad (1)$$

$$LOQ = Xavg+10*STDEV \quad (2)$$

where *Xavg* is the average concentration of the replicates and *STDEV* is the standard deviation of 10 replicate samples.

Accuracy, precision, and measurement uncertainty estimation

Two approaches were used during the validation for determining the bias and MU, analysis of the certified reference materials (CRMs) and recovery tests (standard addition to a sample), as suggested by the Nordtest guide. The minimum number of replicates used in the calculation of the precision and bias was 10 for each matrix. The recovery experiments were performed by comparing the analytical results for the extracted samples at three concentration levels (low, medium, and high). To calculate the expanded MU, the coverage factor k = 2 was used, providing a level of confidence of 95%.

Matrix effect

The effect of the matrix was studied by analysing the zero matrix. For this, clean soil without the analyte was spiked with the calibration solutions at three levels of concentration (near the LoQ, intermediate, and upper levels of the calibration curve). The matrix effect was calculated as follows (Eq. (3):

$$ME\% = \frac{SLsp}{SLcal} * 100 \quad (3)$$

where *SLcal* (the sensitivity of the method) is the slope of the calibration curve and *SLsp* is the slope of the calibration curve with the matrix at the same concentration levels. An *ME%* value of 100% indicates no effect, a value less than 100% indicates ionisation suppression, and an *ME%* over 100% indicates ionisation enhancement due to the co-elution of sample compounds (Zhou et al., 2017). Figure 3 shows representative standard calibration curves for the aliphatic fractions of C_{10}-C_{12}, with and without the spiked matrix.

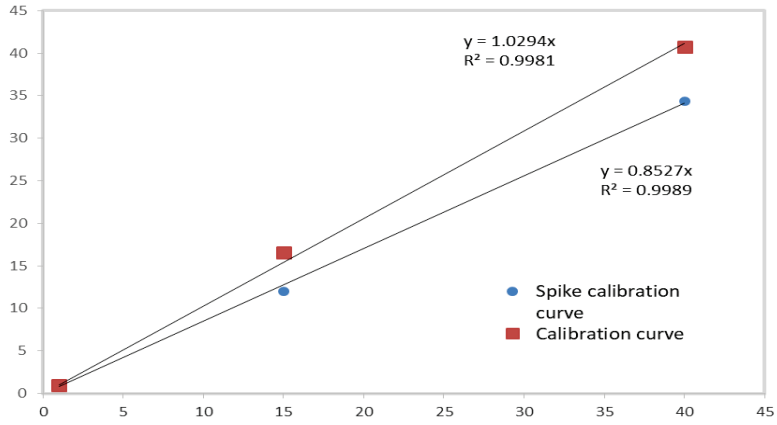

Figure 3. Standard and spiked matrix calibration curves for aliphatic fractions of C_{10}-C_{12}.

Lack-of-fit test

A linear regression model was expected to be obtained for all the compounds in this work. The relationship between x (the concentration of the analyte) and y (the instrument response) was a straight line defined by $y = b_1x + b_0$, where b_0 is the y-intercept and b_1 is the slope of the line. The three replicates of the five expected levels of concentration values were measured by GC-MS for all the compounds included in the method.

The linear calibration model was validated by the lack-of-fit test in conjunction with a residual plot. The F-statistic was used to test the null hypothesis that the linear model was correct when used. The F value was obtained and the probability (P) associated with the F statistic was calculated using the Excel 2010 function =F.INV.RT (0.05 probability; DoF (degree of freedom) (numerator, n-2); DoF (denominator, n*(p-1)) (Moosavi and Ghassabian, 2017; Thompson et al., 2002).

T-test statistics and null hypothesis

T-test statistics and the null hypothesis showed that the candidate method (the new validated method described in **Paper III**) provided PAH results that were equivalent to an existing method (the old laboratory method described in **Paper I**). The t-test was used to evaluate whether there was a statistically significant difference between the mean value for a series of determinations and the accepted reference value (Cherry et al., 2012). The sample used for the t-test was a real soil sample, analysed three times each with the old and new validated methods. For the t-test, 16 EPA PAH differences (replicate average results) were calculated.

The t-test is expressed in Eq. (4) as the mean value of the differences, *Xavg dif*, divided by the mean value of the deviation of the differences, *STDEV dif*, multiplied by the square root of the resulting pairs.

$$t = \frac{Xavg\ dif}{STDEV\ dif} * \sqrt{n} \quad (4)$$

One-way ANOVA

One-way ANOVA was employed to confirm the reproducibility of the methods. Four different chemists performed the analyses on different days. The F, critical F, and P values allowed for direct conclusions to be drawn on whether the variations between the results obtained by the different chemists were significantly greater than the variation in the results obtained by one chemist.

Z-Score in chemical proficiency testing

The final step during the validation of the analytical method was to evaluate the performance characteristics through inter-laboratory ring tests, namely, proficiency tests, in order to demonstrate that the method was fit for its intended purpose. The results of the laboratory tests were assessed by converting them to a Z score using the following equation (Eq. (5)):

$$Z = \frac{Xi - Xa}{\sigma} \quad (5)$$

where *Xi* is the result of laboratory *i*, *X*a is the assigned reference value, and σ is the assigned standard deviation for the proficiency test. The Z-scores were typically interpreted as Z ≤ 2 (acceptable), outside the range of ±2 (questionable), and outside the range of ±3 (actionable) (Hibbert, 2007; Analytical Methods Committee, 2016). The reported value (laboratory result) was compared with the assigned value (the value used to assess the proficiency), and the Z scores for the 16 PAHs and 7 PCBs were calculated by the organiser of the proficiency test (Merck).

3.7 Oil spill fingerprinting (Study III)

On a dry matter basis, soil pollution from former asphalt pavement plants (APPs) was identified using the binary ratios of the PAH isomers, such as Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(g,h,i)P, and B(a)Ant/(B(a)Ant + Chr), as well as the ratios of the four-to-six-ring parent PAHs to the sum of the two-and -three-ring parent PAHs (HMW/LMW). For PCA, a matrix consisting of 54 samples and 7 binary ratios was used. The Min-Max scaled and the autoscaled data were subjected to PCA. In this study, a Microsoft Excel Macro using the NIPALS (nonlinear iterative partial least squares) algorithm was applied to determine the eigenvectors of the most important principal components.

CA was performed for clustering the samples. The data consisted of 7 PAH isomers (Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(g,h,i)P, B(a)Ant/(B(a)Ant + Chr), and (HMW/LMW) from 18 Jänesselja APP samples. The distance was used as the similarity/dissimilarity measure between the samples. Before computing the distances, the PAH binary ratios were first normalised by the maximum-minimum range. Three kinds of linkages were selected (single, average, and complete)

to determine which among them was most suitable for the PAH ratios. All linkage methods showed almost the same groups in CA. The average distance of the linkage methods better reflected the relationships between the binary ratios of the PAHs and the oil types in Jänesselja APP.

To define the similarity between the samples, R software was employed. R is a programming language for statistical analysis, graphics representation, and reporting (Zuur et al., 2009).

3.8 Comparison of different methods and statistical analysis (Study IV)

The PPMC and Shapiro–Wilk test were applied to compare the different methods. PPMC is a measure of the linear relationship between two ratio variables and can have a value between -1 and 1. PPMC is advantageous in that it is a simple means to assess the association between two variables in terms of whether they share variance (covary) and if the relationship is positive or negative, as well as to assess the degree to which they correlate. The disadvantage of using PPMC is that it cannot identify relationships that are not linear, and may show a correlation of zero when the correlation has a relationship other than a linear one (Chee, 2013).

The Shapiro-Wilk test is a hypothesis test applied to a sample with a null hypothesis that the sample has been generated from a normal distribution. If the p-value is less than the chosen alpha level, then the null hypothesis is rejected and the sample is considered to not have been generated from a normal distribution. If the sample size is sufficiently large, this test may detect even trivial departures from the null hypothesis. Additional investigation of the effect size is typically advisable. A Q-Q (quantile-quantile) plot is used to verify if a dataset follows a particular theoretical distribution. It compares the quantiles of the observed data to the quantiles of the other distribution (Builtin; Datacamp).

3.9 Analytical GREEnness Metric (Study V)

The determination of PAH compounds in soil or sediment is a complicated analytical procedure, requiring significant energy and material inputs; hence, it is important to ascertain the environmental friendliness of this method. For this purpose, AGREE metric analysis following the twelve principles of GAC was applied in **Paper III**. The assessment was performed using software with an automatically generated clock-like graph and an assessment report. The performance of the procedure according to each of the twelve principles is reflected with a red-yellow-green colour scale, while the weight of each principle is reflected with the width of its corresponding segment. The final assessment result is the product of the assessment results for each principle. The overall score is shown in the middle of the pictogram. A summary of transformations applied to every principle is presented in graphical form in Figure 4. The maximum score of 1 represents a methodology that is fully compliant with the twelve principles of GAC (Pena-Pereira et al., 2020; Sajid and Płotka-Wasylka, 2022).

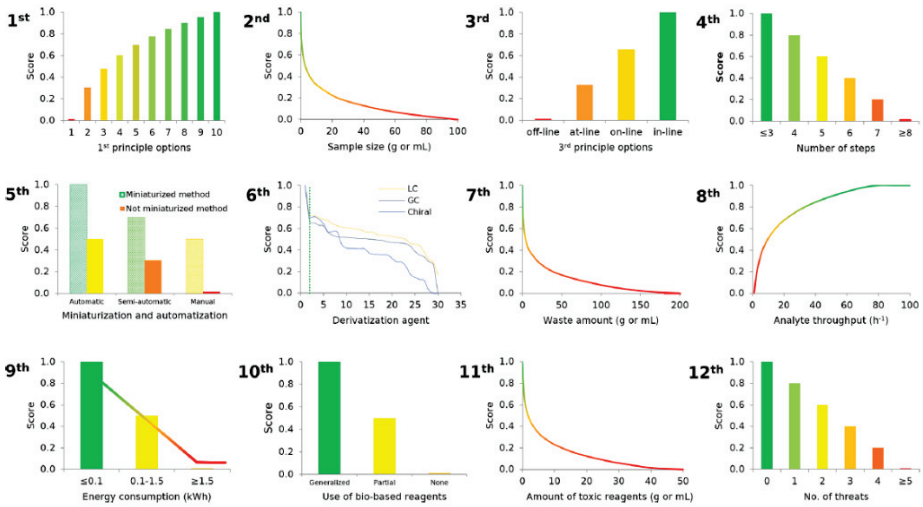

Figure 4. Graphical representation of the functions applied to convert the variables to scores in the 0–1 scale (Pena-Pereira et al., 2020).

4 Results and discussion

4.1 Analytical target profile, method optimisation, and design of experiment (Study I)

The AQbD concept was applied to develop a modified analytical method for compounds from oil spills, namely PAHs, PCBs, and aliphatic and aromatic fractions. Establishing the ATP is an important step in the development of a precise and accurate method to determine oil compounds in soil and sediment using GC-MS. The ATP supports uncertainty estimation throughout the range of analyte concentrations and achieves the necessary limits of quantification. Here, the peak areas on the chromatograms and the resolution between the peaks were used as CMAs.

The analytical target in this case is oil spills into soil and the compounds that are spilled, namely 16 parent PAHs and their alkylated analogues, 7 PCBs, 4 aliphatic fractions from octane to pentatriacontane, and different aromatic compounds.

Experiments were conducted using different sample weights, extraction volumes, times, and speeds, as well as two types of GC columns (Rxi-XLB, low-polarity phase, proprietary and Rxi-17Sil, mid-polarity phase, Crossbond), to evaluate the impact of these parameters on the performance characteristics (sensitivity, selectivity, precision, accuracy, linearity, and LoQ). The columns were selected based on the manufacturer's recommendation (RESTEC) as they offer excellent inertness and selectivity for active environmental compounds such as PCB congeners and PAHs. However, there was no information available regarding their selectivity for aliphatic compounds. Preliminary experiments were conducted using the chromatographic conditions suggested by the column manufacturer for the PAH and PCB analyses. A mixture of the target compounds was injected into the GC-MS instrument. The Rxi-17Sil column enabled the best separation of co-eluted PAHs (e.g., benzo(b)fluoranthene, benzo(k)fluoranthene, and benzo(j)fluoranthene), while the Rxi-XLB column exhibited the highest responses for aliphatic compounds. The latter column was chosen for further validation as the separation between benzo(b)fluoranthene, benzo(k)fluoranthene, and benzo(j)fluoranthene was not crucial in this study.

An initial screening study of the analytical procedure was carried out by examining the oven and inlet temperatures, flow rate, and temperature gradient, which are parameters known to significantly affect peak retention and separation in GC-MS analysis. The preliminary experiments achieved the target LoQ, reproducibility, and MU (Table 4).

Table 4. Target LoQ, reproducibility, and MU.

Compound	**LoQ, mg kg^{-1}**	**Reproducibility, %**	**Measurement uncertainty, %**
Individual PAHs	< 0.030	< 20	< 30
Ali > C8–C10	< 5	< 20	< 30
Ali > C10–C12	< 5	< 20	< 30
Ali > C12–C16	< 5	< 20	< 30

Compound	LoQ, mg kg^{-1}	Reproducibility, %	Measurement uncertainty, %
Ali > C16–C35	< 10	< 20	< 30
Aromatic > C8–C10	< 3	< 20	< 30
Aromatic > C10–C16	< 1	< 20	< 30
Aromatic > C16–C35	< 1	< 20	< 30
Methyl-pyr/fluorant	< 0.5	< 20	< 30
Methyl-chry/benz(a)a	< 0.5	< 20	< 30
Individual PCBs	< 0.002	< 20	< 30

The results of the initial screening study and primary risk assessment revealed that three factors, namely the initial and final oven temperatures and the inlet temperature, considerably affected the retention time and separation of peaks. The DoE used in this work consisted of three factors, each at three levels. It can be expressed as a $3 \times 3 \times 3 = 3^3$ design. Table 5 displays the design matrix, incorporating the selected factors during the screening studies, along with their respective low (0), intermediate (1), and high (2) levels.

Table 5. 3^3 Design.

Factor B (final oven temperature)	Factor C (final inlet temperature)	Factor A (initial oven temperature)		
		0	1	2
0	0	000	100	200
0	1	001	101	201
0	2	002	102	202
1	0	010	110	210
1	1	011	111	211
1	2	012	112	212
2	0	020	120	220
2	1	021	121	221
2	2	022	122	222
Selected Parameter		**Levels of factor studied**		
		Low (0)	**Intermediate (1)**	**High (2)**
Initial oven temperature		40	50	60
Final oven temperature		300	320	340
Final inlet temperature		300	310	320

For this model, 21 experimental runs were conducted in total. A standard concentration of 25 mg/L for each of the aliphatic compounds was used for all the runs. They were analysed for critical analytical attributes, namely, the C_{35}/C_{10} ratios (response of the C_{35} peak to the response of the C_{10} peak). The response of the

experiments was the ratio between the areas of the C_{35} and C_{10} peaks, and the aim was to find a ratio of 1.

The actual and predicted plots in Figure 5 provide a visual assessment of the model fit, reflecting the variations due to random effects. For a good fit, the points should be close to the fitted line (diagonal line).

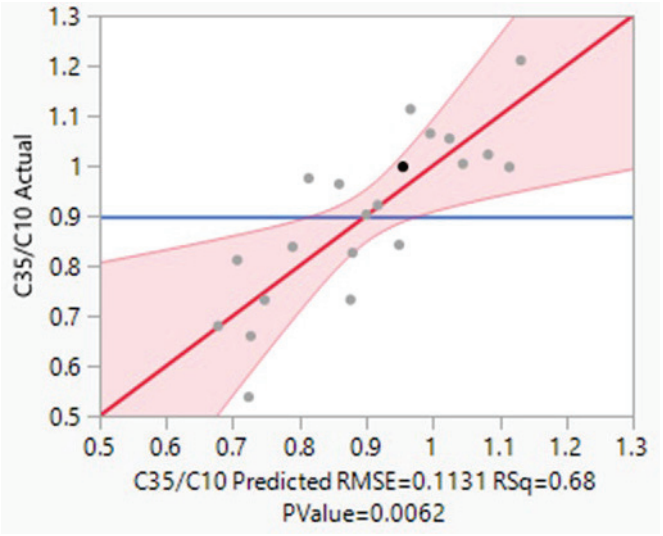

Figure 5. Actual and predicted plots.

Interaction effects were also incorporated in the model. The influence of second-order interactions between the initial and final temperatures of the oven, the oven and final inlet temperatures, and the initial oven and final inlet temperatures was evaluated. The last two interactions (Table 6) were not statistically significant (p-value > 0.05) and hence were removed from the model.

Table 6. Summary of model effects.

Source	P Value
Final inlet temperature	0.0022
Initial oven temperature*Final oven temperature	0.0066
Initial oven temperature	0.0133
Final oven temperature	0.0487
Final oven temperature*Final inlet temperature	0.1164
Initial oven temperature*Final inlet temperature	0.2933

Figure 6 shows the influence of the temperatures on the change in the C_{35}/C_{10} ratios. There is no difference in the ratios when the final temperature of the oven is changed from 300 °C to 340 °C. However, when the final inlet temperature is increased from 300 °C to 320 °C, the ratios of C_{35}/C_{10} change significantly (the slope of the design curve increases).

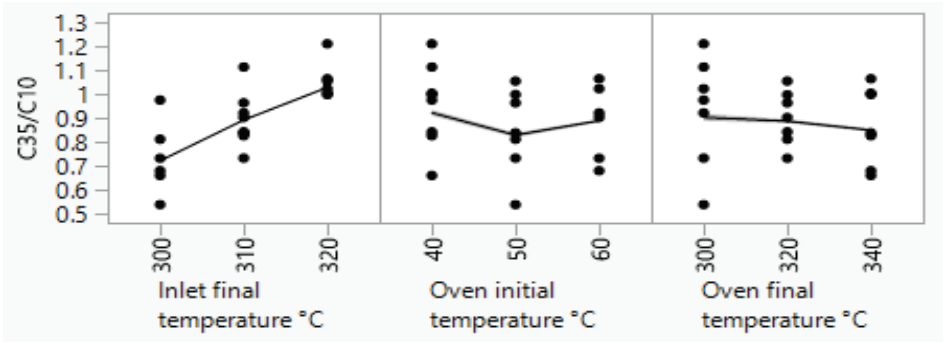

Figure 6. Influence of temperatures on the change in the ratios of C_{35}/C_{10}.

The optimal chromatographic conditions were chosen based on the ratios of C_{35}/C_{10} that were close to 1. The optimised chromatographic solution was observed at an initial oven temperature of 40 °C, with a final oven temperature of 340 °C and a final inlet temperature of 320 °C (Figure 7).

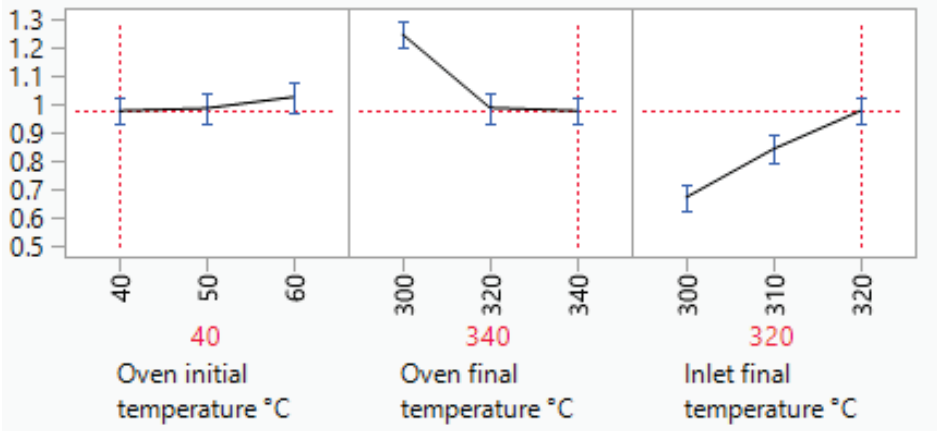

Figure 7. Optimal chromatographic conditions.

4.2 Validation of the analytical method and control strategy (Study II)

The following parameters were evaluated for each component during validation: linearity, LoQ, LoD, accuracy, precision, and MU. A linear calibration curve was confirmed in the ranges of 50-5000 µg/L for all the parent PAHs, 400-5000 µg/L for the alkylated PAH compounds, 15-600 µg/L for the PCBs, and 10000-50000 µg/L for the aliphatic compounds, with a coefficient of determination $R^2 \geq 0.998$. The expected concentrations and obtained concentrations did not differ by more than 20%. The lowest calibration point was equal to the LoQ. At this level, each analyte was determined with acceptable uncertainty.

Table 7 shows the reproducibility, accuracy, MU, LoQ, and recovery results. The results fall within the expected laboratory uncertainty (30%). The reproducibility of the measurements of the PCBs, PAHs, and aliphatic and aromatic fractions is lower than

20%. The difference between the expected and obtained concentrations is below 20%. All the calculated LoQs are lower than those required by Swedish legislation.

The within-laboratory reproducibility, accuracy, and uncertainty were validated for the two matrixes (sediment and soil). For the sediment matrix, the values of the parameters were higher than they were for the soil matrix. The higher values obtained in the analyses of the oil compounds in the sediment met expectations, as this matrix had more complicated mineral structures than the soil.

For the recovery test, blank matrix samples were spiked with known concentrations of the analyte in the range of the calibration curve (three levels: LoQ, intermediate, and high concentration). At each level, the samples were prepared in triplicate. The criterion of recovery was selected according to the calculated MU, implying that the recovery must fall between 80% and 120% in order to be acceptable (2/3rd of the calculated MU 30%). In the present work, all the compounds exhibited a recovery within the allowed range. Table 7 shows the recoveries in soil at the intermediate concentration level.

Table 7. Calculated reproducibility, accuracy, uncertainty, LoQ, and recovery.

Compound	Reproducibility, %	Accuracy, %	Uncertainty, %	LoQ, mg kg^{-1}	Recovery %
Naphthalene	2	11	21	0.015	93
Acenaphthylene	5	10	19	0.014	90
Acenaphthene	4	12	23	0.016	92
Fluorene	4	5	12	0.018	99
Phenanthrene	5	6	10	0.014	102
Anthracene	7	8	16	0.016	99
Fluoranthene	3	8	17	0.014	97
Pyrene	2	10	20	0.015	94
Benz(a)anthracene	6	7	14	0.019	97
Chrysene	5	8	16	0.017	96
Benzo(b)fluoranthene	4	8	15	0.019	110
Benzo(k)fluoranthene	8	8	17	0.020	99
Benzo(a)pyrene	2	11	22	0.018	112
Indeno(1,2,3,c,d)pyrene	6	7	14	0.014	105
Dibenz(a,h)anthracene	4	6	13	0.016	102
Benzo(g,h,i)perylene	7	9	18	0.016	103
Ali > C8–C10	10	12	24	5	88
Ali > C10–C12	9	9	17	5	101
Ali > C12–C16	8	8	19	5	105
Ali > C16–C35	10	10	19	10	99
Aromatics C8–C10	7	9	18	0.9	100
Aromatics C10–C16	3	3	19	0.9	109

Compound	Reproducibility, %	Accuracy, %	Uncertainty, %	LoQ, mg kg^{-1}	Recovery %
Methyl-pyr/fluorant	10	13	26	0.2	91
Methyl-chry/benz(a)a	12	14	27	0.1	92
Aromatics C16–C35	10	12	24	0.9	92
PCB28	6	9	17	0.001	106
PCB52	9	10	20	0.001	104
PCB101	9	12	24	0.002	92
PCB118	5	5	10	0.002	99
PCB153	6	13	26	0.002	89
PCB138	4	4	9	0.002	102
PCB180	3	11	22	0.002	89

The two PAH laboratory methods (described in **Paper I** and **Paper III**) were compared using the t-test statistic. The mean value of the differences between the 16 PAHs was 0.0041 and the deviation was 0.8629. The t-test provided a reading of t = 0.02. Compared to the table value of 2.13 with a freedom level of 15 (16 PAHs -1) and 95% probability, there were no differences in the results between the two methods. A reduction in the extraction time of a single sample positively influenced production in the laboratory.

One-way ANOVA was employed to compare the results obtained by four different chemists, and determine whether the variation in the results obtained by the different chemists was significantly greater than the variation in the results obtained by one chemist. Table 8 shows representative ANOVA results. There are no differences between the results obtained by the chemists in the analyses of the PAHs and PCBs (F crit > F).

Table 8. F statistical, P-value, and F critical calculations using one-way ANOVA.

Source of variation	SS	df	MS	F	P-value	F crit
Between groups	1.762	3	0.587	0.963	0.436	3.287
Within groups	9.151	15	0.610			
Total	10.912	18				

The results of the proficiency test (inter-laboratory comparison, Table 9) for the 16 PAHs and 7 PCBs were satisfactory. The Z-scores for the 14 PAHs and 7 PCBs were in the range of -1.5 to 1.7 (acceptable), and that of 2 PAHs, chrysene and benzo(b)fluoranthene, was 2.8, which is questionable. The separation between benzo(b)fluoranthene and benzo(k)fluoranthene was insufficient. Benzo(j)fluoranthene, which was not part of the EPA PAHs, was eluted between the two analytes; this could lead to higher concentrations than expected. For known interferences in the determination of chrysene, triphenylene was much more resonance-stable than its isomers. In cases where interferences are unavoidable, the sum of the individual compounds can be reported, for instance, triphenylene/chrysene and benzo(b,j,k) fluoranthene.

Table 9. Z-scores of 16 PAHs and 7 PCBs.

Compound	Z-score	Result
Naphthalene	1,7	acceptable
Acenaphthylene	-0,4	acceptable
Acenaphthene	-0,4	acceptable
Fluorene	-0,5	acceptable
Phenanthrene	1,3	acceptable
Anthracene	0,3	acceptable
Fluoranthene	0,6	acceptable
Pyrene	0,0	acceptable
Benzo(a)anthracene	-0,2	acceptable
Chrysene	2,8	questionable
Benzo(b)fluoranthene	2,8	questionable
Benzo(k)fluoranthene	1,2	acceptable
Benzo(a)pyrene	0,5	acceptable
Indeno(1,2,3-cd)pyrene	1,7	acceptable
Dibenz(a,h)anthracene	-0,5	acceptable
Benzo(g,h,i)perylene	1,1	acceptable
2,2',5,5'-Tetrachlorobiphenyl (PCB52)	-1,5	acceptable
2,2',4,5,5'-Pentachlorobiphenyl (PCB101)	-0,7	acceptable
2,3',4,4',5-Pentachlorobiphenyl (PCB118)	-1,5	acceptable
2,2',3,4,4',5'-Hexachlorobiphenyl (PCB138)	-0,7	acceptable
2,2',4,4',5,5'-Hexachlorobiphenyl (PCB153)	-1,0	acceptable
2,2',3,4,4',5,5'-Heptachlorobiphenyl (PCB180)	-0,2	acceptable
2,4,4'-Trichlorobiphenyl (PCB 28)	-1,3	acceptable

4.3 Chemometrics approach for identifying residual oil contamination from former primitive asphalt pavement plants (Study III)

Contaminated soil from two Jänesselja and Maadevahe APPs was analysed to identify the oil type. GC-FID chromatograms were acquired for an initial screening of the oils. The chromatograms of the soil sample were visually compared with those of the control oils, and the levels of oil weathering were estimated. Figures 8-13 show the control oils used in this study for oil spill identification: diesel (Figure 8), light fuel oil (LFO) (Figure 9) shale oil grade C (Figure 10), used motor oil (Figure 11), new motor oil (Figure 12), and high fuel oil (HFO) (Figure 13). Retention times between 3.5 min and 6 min defined the HOI from decane (C_{10}) to heneicosane (C_{21}), and those between 6 min and 8.5 min defined the fractions C_{21}-C_{40}. The chromatogram in Figure 8 shows that the common

diesel components include alkanes from C_{10} to C_{24}, while LFO (Figure 9) includes components from C_{10} to C_{32}. Broader unresolved complex mixtures (UCMs) are observed in the motor oils (Figure 11 and Figure 12), but no UCMs are identified in diesel (Figure 8) or LFO (Figure 9).

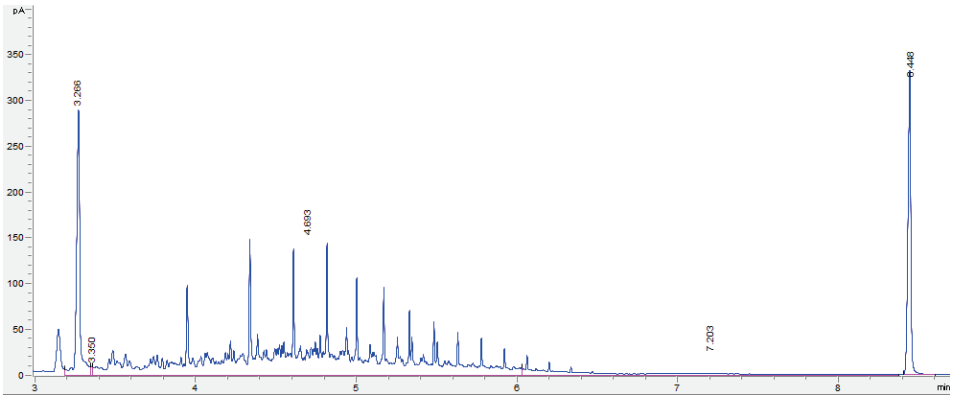

Figure 8. GC-FID chromatograms of diesel.

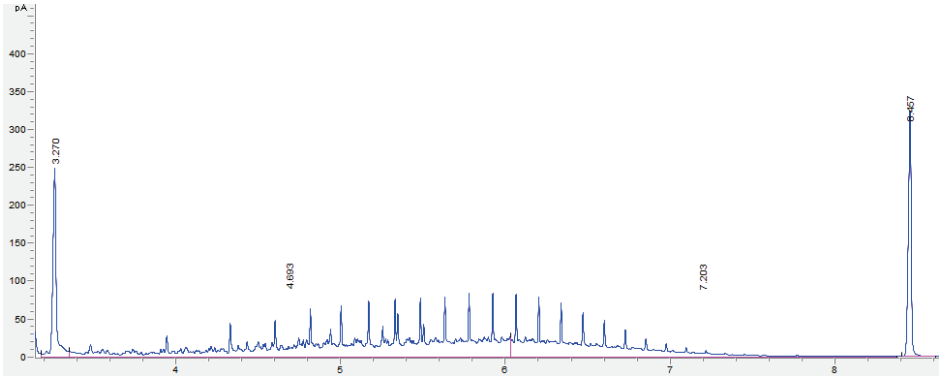

Figure 9. GC-FID chromatograms of LFO.

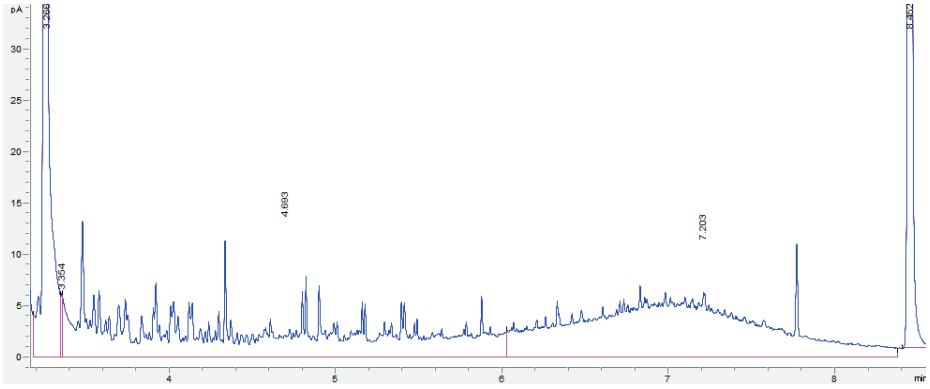

Figure 10. GC-FID chromatograms of shale oil grade C.

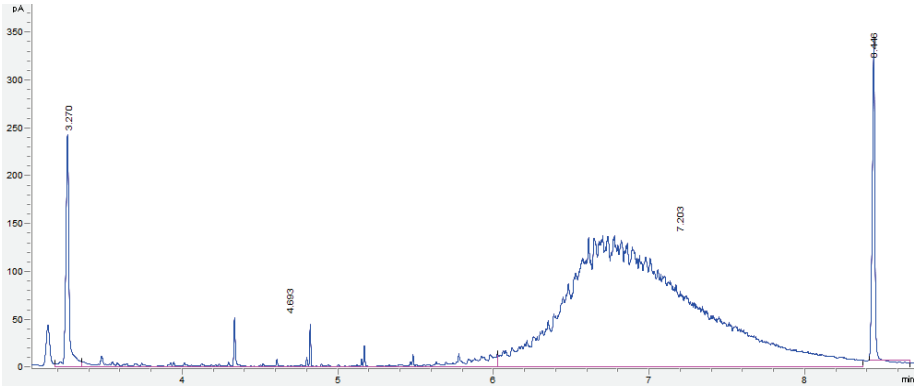

Figure 11. GC-FID chromatograms of used motor oil.

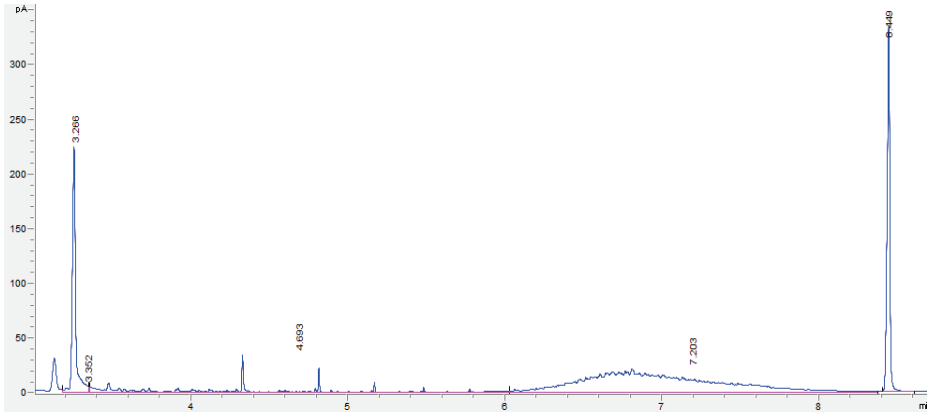

Figure 12. GC-FID chromatograms of new motor oil.

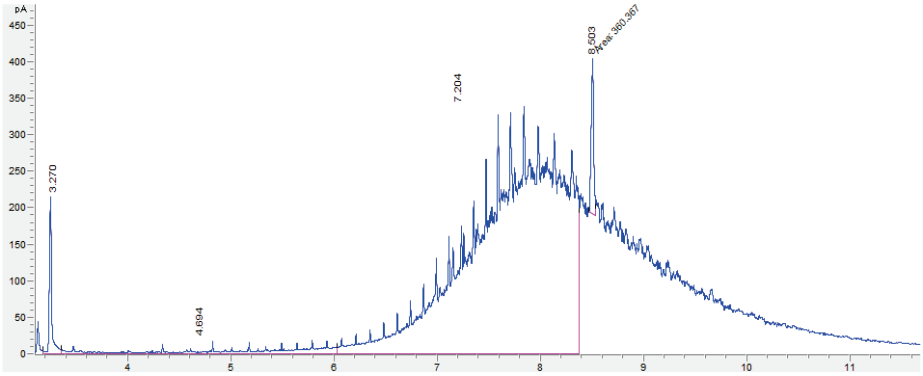

Figure 13. GC-FID chromatograms of HFO.

The seven diagnostic ratios (Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(g,h,i)P, B(a)Ant/(B(a)Ant + Chr), and HMW/LMW), together with PCA, were used in the identification of oil sources in the Jänesselja and Maadevahe APP soil samples. The same diagnostic ratios of the control oils, like diesel, LFO, used motor oil, shale oil, and HFO, were also calculated and compared with those of the soil samples.

The average ratio of Ind/(Ind + B(g,h,i)P) for all 54 samples was 0.48. One sample had a value > 0.5, which may be attributed to biomass combustion. The other samples had values between 0.42 and 0.5, indicating petroleum combustion as the possible source of the PAHs. The variations in the PAH sources, as indicated by these indices, could be due to the different sampling sites, which have different hydrological and spatial variations. The Ant/(Ant+Phe) for the Jänesselja samples was in the range of 0.18-0.84 and for the Maadevahe samples, it was in the range of 0.15-0.33 (Table 10). Ant/(Ant+Phe) < 0.1 indicates that the PAHs in the soil are from petroleum contamination sources, whereas Ant/(Ant+Phe) > 0.1 may imply that the PAHs are from wood and coal combustion sources.

Table 10. Ranges of the binary ratios for the Jänesselja and Maadevahe APPs.

Binary ratios	Ant/ Phe	Fla/ Pyr	Ant/ Ant+Ph	Fla/ Fla+Pyr	Ind/ Ind+B(g,h,i)P	B(a)Ant/ B(a)Ant+Chr	HMW/ LMW
Jänesselja APP	0.21-5.19	0.63-0.95	0.18-0.84	0.38-0.49	0.42-0.71	0.51-0.65	0.02-2.63
Maadevahe APP	0.18-0.50	0.70-0.87	0.15-0.33	0.41-0.46	0.44-0.50	0.38-0.62	0.06-0.70

Most of the samples from the APPs exhibited Fla/(Fla+Pyr)<0.5, B(a)Ant/(B(a)Ant + Chr > 0.4, and Ind/(Ind + B(g,h,i)P > 0.35. These indices from the APPs indicate that the sources of contamination in the APPs were oil shale and petroleum combustion.

The score plot of the first and second PCs (Figure 14) shows 54 samples from the APPs, including diesel, LFO, used motor oil, shale oil, and HFO. The first two PCs account for 70.69 % of the total components in the complex mixtures of the oil samples. PC1 accounts for 45.46 % of the total oil spill compounds, while PC2 accounts for 25.23 % of the total variance. The score plot reveals that most of the

samples are similar to the used motor oils. This study used motor oil in the form of a mixture of different used motor oils from diesel and petroleum engines. Diesel, shale oil, HFO, and LFO clasters were far from the analysed oils in the soil samples. Sample 2 from the Jänesselja APP was detected as an outlier, since it lay far apart from the other soil samples and control oils in the score plot. The Ant/Phe ratio was 5, while for all other samples, Ant/Phe < 0.66. The source of anthracene in this sample was unknown.

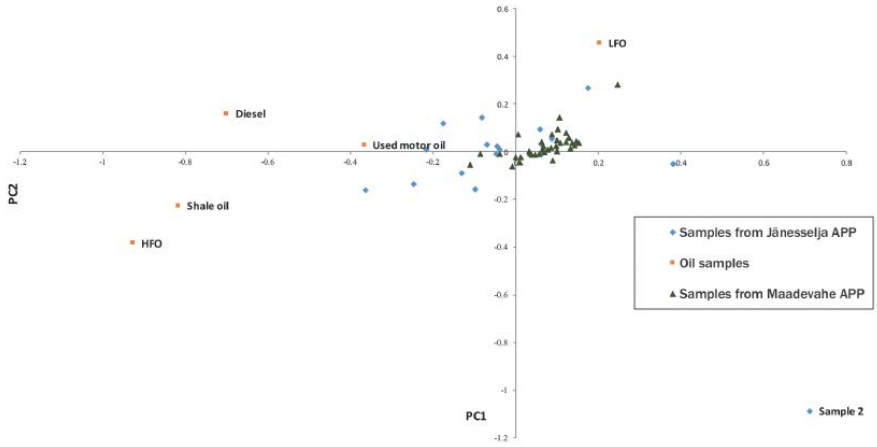

Figure 14. Score plot of the binary ratios of the samples from the Jänesselja APP, Maadevahe APP, LFO, diesel, shale oil, used motor oil, and HFO.

The 16 relative PAHs (Figure 15) were divided into 5 groups: 2 benzene rings, 3-rings, 4-rings, 5-and 6- rings, and acenaphthylene. In this study, the behaviour of acenaphthylene in the GC system differed from that of other 3-ring PAHs. In some cases, this compound decomposed in the GC inlet, rendering the analysis of acenaphthylene more difficult than that of other PAHs. This issue can be overcome using a programmable temperature vaporisation (PTV) inlet and specific ISTD. The other individual PAHs with similar chemical characteristics were clustered into the same group. For example, Cluster 1 was composed of Nap (2-ring benzene), Cluster 2 had 3-ring PAHs, Cluster 3 consisted of 4-ring PAHs (Fla, Pyr, Chr, B(a)Ant). B(b)F, B(k)F, B(a)P, B(g,h,i)P, DiAnt, and Ind), and Cluster 4 consisted of high-molecular-weight 5- and 6-ring PAHs.

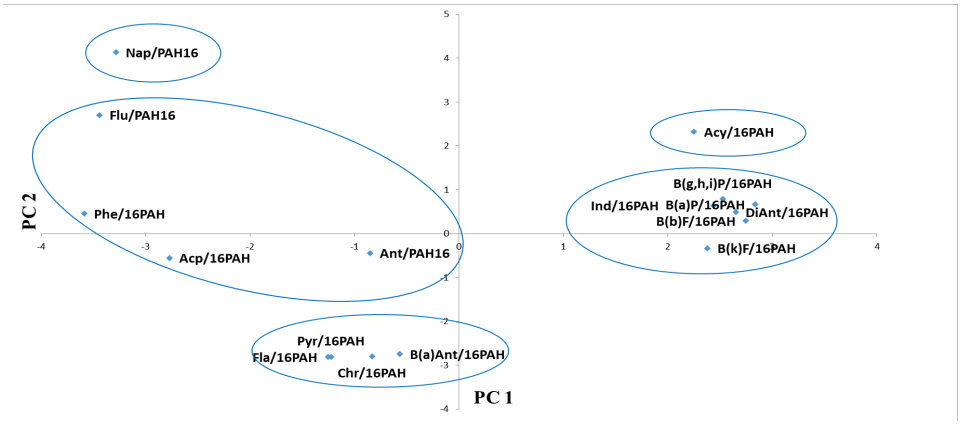

Figure 15. Groups of relative ratios of PAHs (Jänesselja APP).

Based on the clusters obtained from the CA of the samples from the Jänesselja APP (Figure 16) and using the chromatograms, the sampling sites in this study were grouped according to the oil mixture type. Cluster 1 included Sample 2 (the outlier). In Cluster 2, Sample 13 was polluted mostly with LFO. Sample 4 was polluted with a mixture of shale oil and diesel. Cluster 4 represented the samples that were mostly polluted with HFO (probably mazut). Cluster 5 represented samples polluted with a shale oil/LFO mixture. Some sample pairs, namely, 1-10, 7-14, 8-16, and 3-15, were located next to each other on the dendrogram. The clusters that were generated were highly convincing, as the samples in the groups had similar PAH binary ratios and similar sampling backgrounds.

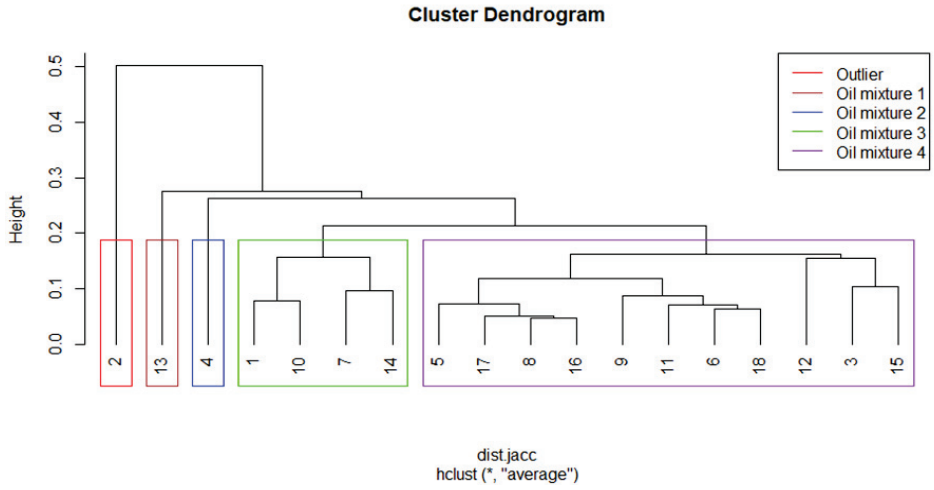

Figure 16. Dendrogram of the groups of the samples from Jänesselja APP.

4.4 Comparison of different methods and statistical analysis between the parameters (Study IV)

The advantages and disadvantages of the 10 different GC-MS, GC-FID, and HPLC methods, including the one developed in the current study, were compared (Table 11) in the analysis of PAHs. The comparison considered the cost of one sample, energy consumption, time of extraction, the time spent by the analyst for sample preparation, and the amount of waste generated, all of which were crucial in the choice of the suitable method. Additionally, the LoQ, repeatability/reproducibility, and recovery rate of the methods were compared.

Different methods used samples weighing from 1 to 60 g. The sample weight is usually selected considering the type of solid material and the LoQs to be achieved. The weight of 20 g for the validated method was chosen to obtain a representative sample with a sufficient LoQ. In the validated method, hexane and acetone were used as extraction solvents, as they enable the extraction of more polar compounds. The modified method in this study first used acetone alone, followed by the addition of hexane. It was confirmed that using the solvents in this specific order resulted in a higher extraction efficiency than that obtained using the solvents individually.

Most methods used ISTDs or surrogate standards. The number of standards ranged from 3 to 11. Methods 2, 4, 7, and 9 applied external calibration. The correct ISTDs can compensate for the low extraction efficiency and reduce the matrix effect. In this study, four ISTDs were used: one for aliphatic compounds, one for high PAHs and PCBs, one for intermediate PAHs, one for light PAHs.

Method 7 employed solid-phase microextraction (SPME) with a needle-trap device (NTD), which is a green and solvent-free sample preparation method. However, the packing of sorbents into the NTD by this method was a time-consuming process with low repeatability (Dalvand and Ghiasvand, 2019). Methods 2 and 5 used Soxhlet extraction, which required 24 and 6 h, respectively, to complete a single extraction. Besides being time-consuming, this method used a large amount of solvent, which is disadvantageous from the cost and environmental perspectives. In Method 5, a dichloromethane extraction solution volume of 120 mL was used. The accelerated solvent extraction employed in Method 3 was faster, consumed less solvent, and could be easily automated; however, the equipment for accelerated solvent extraction is expensive. In addition, heat-sensitive compounds may still be affected during accelerated solvent extraction (Han et al., 2019). Similar to Method 10, the present study employed mechanical extraction. Mechanical extraction can be used for different soil types and is easy to implement in a laboratory setting. Mechanical agitation can accelerate the extraction process by facilitating the movement of the solvent through the soil matrix. However, continuous mechanical agitation requires energy, which may increase operational costs.

Table 11. Methods of PAH analysis.

Method	Reference	Equipment	Extraction type	Number of stages	~Time, min	Weight, G	Extracts	Volume, mL	ISTDs	Greenness score	LoQ, mg kg^{-1}	Reproducibility, %	Recovery, %
1 (current study)	N/A	GC-MS	Mechanical	5	60	20	2	30	3	0.42	0.014 –0.020)	2–8	88–112
2	Huang et al., 2013	HPLC-FLD-DAD	Soxhlet	9	1460	2	4	~72	0	0.33	0.001–0.0095	< 5	29.8–96
3	Han et al., 2019	GC-MS	ASE	11	40	5	3	~80	11	0.32	0.0001–0.00017	1.3–18.4	80–100
4	Kumar et al., 2014	HPLC UV-DAD	Sonication	12	190	20	4	~237	0	0.25	0.00005–0.00171	3.34–14.68	82– 106
5	Environment Agency, 2003	GC-MS	Soxhlet	9	400	10	2	~185	6	0.22	< 1	1-33	N/A
6	Yamada et al., 2009	GC-MS	Sonication	12	150	10	3	~200	6	0.23	0.00136–0.01186	2–20	68–108
7	Dalvand and Ghiasvand, 2019	NTD-GC-FID	SPM	2*	40	2	0	0	0	0.68	0.000001–0.00001	7.3–13.2	92.53–110.75
8	Dong et al., 2012	GC-MS	Sonication	12	100	1	2	15	5	0.33	0.0006–0.0054	1.1–13.7	87–128
9	Volk and Gratzfeld-Huesgen, 2011	HPLC FLD-DAD	Sonication	7	40	5	3	~50	0	0.34	0.00741–0.45258	< 14	84–87
10	Reflab metode 4:2008, 2008	GC-MS	Mechanical	5	730	60	2	40	3	0.45	0.005	< 15	90

**The preparation of the NTD prior to extraction is not incorporated in the preparation stage.*

Statistical analysis between the parameters used for comparing different methods

The PPMC coefficients (Table 12) between the parameters (the weight (*Weight*) of the sample (in grams), number of solvents (*Extracts*), number of method preparation steps (*Stages*), number of ISTDs, volume of solvents (*Volume*), extraction time in min (*Time*), minimum and maximum LoQ values obtained using these methods (*LoQ.min* and *LoQ.max*), average repeatability (*Repeatability*) of the method in %, and average recovery *(Recovery)* of the control samples) were calculated using the R statistic to determine the strength of the relationship between the parameters.

Table 12. PPMC results.

	Weight	Extracts	Stages	ISTDs	Volume	Time	LoQ min	LoQ max	Repeatability	Recovery
Weight	**1.000**	0.274	-0.112	0.045	0.195	0.496	0.071	-0.184	0.536	-0.166
Extracts	0.274	**1.000**	0.291	0.320	-0.071	0.118	-0.252	0.136	-0.114	-0.034
Stages	-0.112	0.291	**1.000**	0.030	0.647	-0.248	0.310	-0.047	-0.289	-0.360
ISTDs	0.045	0.320	0.030	**1.000**	0.186	0.625	0.242	-0.393	0.483	-0.068
Volume	0.195	-0.071	0.647	0.186	**1.000**	0.126	0.840	-0.040	0.112	-0.786
Time	0.496	0.118	-0.248	0.625	0.126	**1.000**	0.215	-0.211	0.456	-0.164
LoQ min	0.071	-0.252	0.310	0.242	0.840	0.215	**1.000**	0.103	0.119	-0.654
LoQ max	-0.184	0.136	-0.047	-0.393	-0.040	-0.211	0.103	**1.000**	-0.804	-0.394
Repeatability	0.536	-0.114	-0.289	0.483	0.112	0.456	0.119	-0.804	**1.000**	0.167
Recovery	-0.166	-0.034	-0.360	-0.068	-0.786	-0.164	-0.654	-0.394	0.167	**1.000**

The Shapiro-Wilk test was conducted to calculate the p values and to test the null-hypothesis. High degrees of positive correlation (from 0.5 to 1) were obtained for *Weight* and *Repeatability* at 0.536; *Stages* and *Volume* at 0.647; *ISTDs* and *Time* at 0.625; and *Volume* and *LoQ min* at 0.840. High degrees of negative correlation (from -0.5 to -1) were obtained for *Volume* and *Recovery* at -0.786 and for *Recovery* and *LoQ min* at -0.654. No correlation (less than ±0.100) was obtained for *Weight* and *ISTDs* at 0.045; *Weight* and *LoQ min* at 0.071; number of *Extracts* and *Volume* at -0.071; number of *Extracts* and *Recovery* at - 0.034, and number of *ISTDs* and *Recovery* at - 0.068.

Figure 17 shows the linear relationship between the two sets of data and the plot with a normal distribution error: *Volume* vs. *Recovery* (r = -0.786 and p-value = 0.971*)*, *Volume* vs. *LoQ min* (r = 0.840 and p-value = 0.098), *Stages* vs. *ISTDs* (r = 0.030 and p-value = 0.767*)*. An increase in the volume of the extract decreases the recovery of the PAHs, while a decrease in the volume of the extract results in a lower LoQ. Figure 17 (*Stages vs. ISTDs*) is a representative example where the correlation between the parameters is absent.

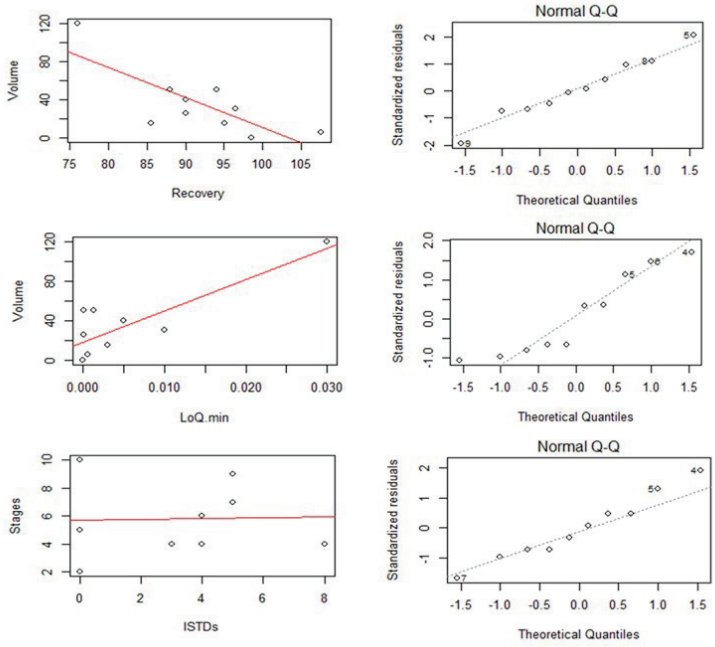

Figure 17. Linear relationship between the two sets of data and the plot with a normal distribution error.

4.5 Greenness assessment of the method (Study V)

The column “Greenness score” in Table 11 presents the calculated AGREE scores for all methods under consideration. AGREE metric analysis was chosen as it is a comprehensive, flexible, and straightforward assessment approach that provides an easily interpretable and informative result. The weight for all GAC principles was set to

two, as all principles are equally important. The low AGREE scores in all methods can be attributed to the use of traditional extraction methods. Soxhlet extraction and sonication exhibit an explicitly worse score (less than 0.35), while mechanical procedures (our method and Method 10) have similar scores (0.42 and 0.45, respectively). One exception is Method 7, where the use of SPME for analyte extraction yields a higher score.

Table 13 shows the results of the AGREE analysis for the current method. The procedure involved an external sample treatment with a reduced number of steps (principle (p) 1), and the use of 20 g of the soil sample (p 2). The measurement was conducted off-line (p 3) and the procedure involved five distinct steps (p 4). The procedure was not automated, but incorporated a miniaturised sample preparation technique (p 5). No derivatisation agents were used in the analysis (p 6). The analytical waste contained 20 g of the sample, 20 mL of acetone, 10 mL of hexane, and 1 g of Na_2SO_3 (p 7). In all, 43 analytes were determined in a single run, and the sample throughput was 2 samples per hour (p 8). GC-MS was found to be the most energy-intensive analytical technique (p 9). Some of the reagents (acetone) may have originated from bio-based sources (p 10). The procedure required 10 mL of toxic solvents (hexane) (p 11), while PAHs were persistent and toxic to aquatic life (p 12).

Table 13. Results of AGREE analysis.

No	Criterion	Score	Amounts / Notes
1	Sample treatment	0.3	External sample treatment with a reduced number of steps
2	Sample amount	0.22	20 g of the soil sample
3	Device positioning	0.0	Off-line
4	Sample preparation steps	0.6	Five
5	Automation & miniaturisation	0.5	The procedure was not automated but incorporated miniaturised sample preparation technique
6	Derivatisation	1.0	No derivatisation agents
7	Waste generation	0.17	20 g of sample, 20 mL of acetone, 10 mL of hexane, and 1 g of Na_2SO_3
8	Analysis throughput	1.0	43 analytes in a single run, 2 samples per hour
9	Energy use	0.0	GC-MS is the most energy-intensive analytical technique
10	Reagent source	0.5	Acetone may originate from bio-based sources
11	Toxicity	0.2	10 mL of toxic solvents (hexane)
12	Operator hazard	0.6	PAHs are persistent and toxic compounds

5 Conclusions

This work aimed to develop and apply a suitable analytical method for the precise, robust, and selective analysis of oil compounds in soil, as well as to investigate the use of different statistical/chemometric techniques in the method development phase and identify the oil type in contaminated soil.

The main conclusions can be summarised as follows:

I. The AQbD approach in environmental analysis was used for the simultaneous identification and quantification of 16 PAHs, 7 PCBs, alkylated PAHs, and aliphatic compounds in the soil. The ATP, CMPs, and CMAs were identified. The acceptable ranges were established. DoE enabled the determination of optimal GC conditions for the new method to be developed.

II. The robust, rapid, simple, sensitive, and selective GC method was validated. The reproducibility, accuracy, MU, and recovery of the method were highly acceptable, and the LoQ satisfied regulatory requirements.

III. Two chemometric techniques (CA and PCA) and two analytical methods (GC-MS and GC-FID) were applied to better understand the distribution of PAHs in the soils from former APPs, as well as to identify the possible sources of the old oils. The combination of analytical and chemometric methods in oil spill fingerprinting helped in interpreting the data based on the ratios of the PAHs, while determining the possible sources of the old oils, thus contributing to considerable time and cost savings in oil classification.

IV. Comparison showed that there was no single “best” method for measuring all types of soil-contaminating PAH compounds. A comparison of 10 methods revealed the newly developed method to be most suitable for the intended purpose in every aspect, including for rapid analysis. Including all steps of sample preparation, the analysis time was approximately 1 h. The PPMC coefficients between the parameters were calculated with the use of R statistic. Volume vs. Recovery showed a negative correlation, Volume vs. LoQ min showed a positive correlation, while Stages vs. ISTDs showed no correlation between the parameters.

V. The greenness of the method was evaluated using AGREE metric analysis. The greenness score of the present method generally surpassed that of the other methods. The present method involved five steps of sample preparation without cleanup and evaporation, making it greener than the other methods.

References

Adeniji, A.O.; Okoh, O.O.; Okoh, A.I. (2018). *Analytical Methods for Polycyclic Aromatic Hydrocarbons and their Global Trend of Distribution in Water and Sediment: A Review.* In book: Recent Insights in Petroleum Science and Engineering, DOI: 10.5772/intechopen.71163.

Agilent Technologies. *Measuring Hydrocarbon Oil Index according to ISO 9377-2 (DIN H53). Environmental Application. Gas Chromatography.* Available online: https://www.agilent.com/library/applications/app_note.pdf

Alomair, M.; Georgiou, S.; Stylianou, S. (2020). *Three-level designs: Evaluation and comparison for screening purposes.* Wiley Online Library, 90-107. https://doi.org/10.1002/qre.2722.

Analytical Methods Committee. (2016). *z-Scores and other scores in chemical proficiency testing—their meanings, and some common misconceptions*. Anal. Methods 28. https://doi.org/10.1039/c6ay90078j.

Araújo, A. S.; Andrade, D. F.; Babos, D. V.; Castro, J. P.; Garcia, J. A.; Sperança, M. A.; Gamela, R.R.; Machado, R. C.; Costa, V. C.; Guedes, W. N.; Pereira-Filho, E. R.; Pereira, F. M. V. (2021). *Key Information Related to Quality by Design (QbD) Applications in Analytical Methods Development*. Brazilian Journal of Analytical Chemistry, 8(30), 14-28. doi: 10.30744/brjac.2179-3425.RV-27-2020.

Ashraf, M.A.; Maah, M.J.; Yusoff, I. (2013). *Soil Contamination, Risk Assessment and Remediation. Environmental Risk Assessment of Soil Contamination*. DOI: 10.5772/57287. Available online: https://books.google.ee/books?hl=en&lr=&id=EBCQDwAAQBAJ&oi=fnd&pg=PA3&dq=soil+contamination+Exposure+Pathways&ots=gBrISrB4E5&sig=nQwH_f9X_MEYyH6eoRjxQHWil0Q&redir_esc=y#v=onepage&q=soil%20contamination%20Exposure%20Pathways&f=false

Beg, S.; Sharma, G.; Katare, O.P.; Lohan, S.; Singh, B. (2015). *Development and Validation of a Stability-Indicating Liquid Chromatographic Method for Estimating Olmesartan Medoxomil Using Quality by Design*. Journal of Chromatographic Science, 53, 1048–1059. doi:10.1093/chromsci/bmu165.

Belouafa, S.; Habti, F.; Benhar, S.; Belafkih, B.; Tayane, S.; Hamdouch, S.; Bennamara, A.; Abourriche, A. (2017). *Statistical tools and approaches to validate analytical methods: methodology and practical examples.* Int. J. Metrol. Qual. Eng., 8, 9. https://doi.org/10.1051/ijmqe/2016030.

Brereton, R. G. (2003). *Chemometrics. Data Analysis for the Laboratory and Chemical plant*. Wiley, Chichester.

Builtin. *An Introduction to the Shapiro-Wilk Test for Normality.* Available online: https://builtin.com/data-science/shapiro-wilk-test

Chee, J. (2013). *Pearson's Product-Moment Correlation: Sample Analysis.* Available online: https://www.researchgate.net/publication/262011045

Cherry, S. R.; Sorenson, J. A.; Phelps, M. E. (2012). *Nuclear Counting Statistics in: Physics in Nuclear Medicine*. Fourth Edition, 125-140.

Christensen, J.H.; Tomasi, G. (2007). *Practical aspects of chemometrics for oil spill fingerprinting.* Journal of Chromatography A, 1169, 1–22. https://doi.org/10.1016/j.chroma.2007.08.077.

Dalvand, K.; Ghiasvand, A. (2019). *Simultaneous analysis of PAHs and BTEX in soil by a needle trap device coupled with GC-FID and using response surface methodology involving Box-Behnken design.* Anal. Chim. Acta., 1083, 119-129. https://doi.org/10.1016/j.aca.2019.07.063.

Datacamp. *The Q-Q Plot: What It Means and How to Interpret It*. Available online: https://www.datacamp.com/tutorial/qq-plot

Doltade, M.; Saudagar, R. (2019). *The Analytical Method Development and Validation: A Review.* Journal of Drug Delivery and Therapeutics, 9(3), 563-570, doi:10.22270/jddt.v9i3.2774.

Dong, C-D.; Chen, C-F.; Chen, C-W. (2012). *Determination of Polycyclic Aromatic Hydrocarbons in Industrial Harbor Sediments by GC-MS*. Int J Environ Res Public Health, 9(6), 2175-2188. https://doi.org/10.3390/ijerph9062175.

Dorana, J. W.; Zeiss, M.R. (2000). *Soil health and sustainability: managing the biotic component of soil quality*. Applied Soil Ecology, 15, 1, 3-11.

Dudhagara, D.R.; Rajpara, R.K.; Bhatt, J.K.; Gosai, H.B.; Sachaniya, B.K.; Dave, B.P. (2016). *Distribution, sources and ecological risk assessment of PAHs in historically contaminated surface sediments at Bhavnagar coast, Gujarat, India.* Environmental Pollution, 213, 338-346. https://doi.org/10.1016/j.envpol.2016.02.030.

Ellwanger, J.B.; Wingert, N.R.; Volpato, N.M.; Garcia, C.V.; Schapoval, E.E. S.; Steppe, M. (2020). *Analytical Quality by Design Approach for a Stability-Indicating. Method to Determine Apixaban and Its Related Impurities.* Chromatographia 83(R2). DOI:10.1007/s10337-019-03815-9.

Environment Agency. (2003). *The determination of polycyclic aromatic hydrocarbons in soil by dichloromethane extraction using gas chromatography with mass spectrometric detection.* Available online: https://assets.publishing.service.gov.uk/media/5be998cde5274a082fc7f203/pahinsoil.pdf

Eurachem Guide. (2014). *The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics.* Second Edition. https://www.eurachem.org/images/stories/Guides/pdf/MV_guide_2nd_ed_EN.pdf

European Environment Agency. *Soil pollution and health*. Available online: https://www.eea.europa.eu/publications/zero-pollution/health/soil-pollution

Faraday, E.T.; Ifeanyi, U.J.; Rofiat, A.F. (2024). *Determination of Polychlorinated Biphenyls (PCBS) using Gas Chromatography Electron Capture Detector in Dust from Schools in Okitipupa, Ondo State, Nigeria.* Dutse Journal of Pure and Applied Sciences, 10, 3c. 10.4314/dujopas.v10i3c.2.

Gad, H. A.; El-Ahmady, S. H.; Abou-Shoer, M. I.; Al-Azizi, M.M. (2013). *Application of Chemometrics in Authentication of Herbal Medicines: A Review.* Phytochem. Anal., 24, 1–24. DOI: 10.1002/pca.2378.

GUM JCGM 100:2008. (2008). *Evaluation of measurement data — Guide to the expression of uncertainty in measurement*, first edition, 2008.

Hakemeyer, C.; McKnight, N.; St. John, R.; Meier, S.; Trexler-Schmidt, M.; Kelley, B.; Zettl, F.; Puskeiler, R.; Kleinjans, A.; Lim, F.; Wurth, C. (2016). *Process characterization and Design Space definition.* Biologicals, 44, 306-318. https://doi: 10.1016/j.biologicals.2016.06.004.

Han, M.; Kong, J.; Yuan, J.; He, H.; Hu, J.; Yang, S.; Li, S.; Zhang, L.; Sun, C. (2019). *Method development for simultaneous analyses of polycyclic aromatic hydrocarbons and their nitro-, oxy-, hydroxy- derivatives in sediments.* Talanta, 205, 120128. doi:10.1016/j.talanta.2019.120128.

Hibbert, D. B. *Experimental design in chromatography: A tutorial review*. (2012). J. Chromatogr. B, 910, 2– 13. DOI: 10.1016/j.jchromb.2012.01.020.

Hibbert, D.B. (2007). *Quality assurance for the analytical chemistry laboratory*. Oxford University Press.

Huang, Y.; Wei, J.; Song, J.; Chen, M.; Luo, Y. (2013). *Determination of low levels of polycyclic aromatic hydrocarbons in soil by high performance liquid chromatography with tandem fluorescence and diode-array detectors.* Chemosphere, 92 (8), 1010-1016. https://doi.org/10.1016/j.chemosphere.2013.03.035.

ICH harmonised guideline. (2023). *Analytical procedure development Q14*. https://database.ich.org/sites/default/files/ICH_Q14_Guideline_2023_1116_1.pdf

ICH harmonised guideline. (2023). *Validation of analytical procedures Q2(R2).* https://database.ich.org/sites/default/files/ICH_Q2%28R2%29_Guideline_2023_1130.pdf

ICH Harmonized Tripartite Guidance. (2009). *Pharmaceutical Development Q8(R2).* https://database.ich.org/sites/default/files/Q8%28R2%29%20Guideline.pdf.

ICH harmonised guideline. (2023). *Quality Risk Management Q9(R1).* https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_Step4_2025_0115.pdf

ISO 11465:1993. (1993). *Soil quality — Determination of dry matter and water content on a mass basis — Gravimetric method*

Jackson, P.; Borman, P.; Campa, C., Chatfield, M.; Godfrey, M.; Hamilton, P.; Hoyer, W.; Norelli, F.; Orr R.; Schofield, T. (2019). *Using the Analytical Target Profile to Drive the Analytical Method Lifecycle.* Anal. Chem., 91, 2577−2585. https://doi: 10.1021/acs.analchem.8b04596.

Jefimova, J.; Irha, N.; Reinik, J.; Kirsoa, U.; Steinnes, E. (2014). *Leaching of polycyclic aromatic hydrocarbons from oil shale processing waste deposit: A long-term field study.* Science of the Total Environment, 481, 605–610. DOI: 10.1016/j.scitotenv.2014.02.105.

Jurjeva, J.; Koel, M. (2019). *The chemometric approach to identification of residual oil contamination at former primitive asphalt pavement plants*. Oil Shale, 36(3), 410–430. doi: https://doi.org/10.3176/oil.2019.3.04.

Jurjeva, J.; Koel, M. (2022). *Using analytical quality by design principles in the analysis of petroleum compounds in the soil.* Current Topics in Analytical Chemistry, 14, 45 – 59.

Jurjeva, J.; Koel, M. (2024). *Application of Concepts of Design Space in Method Development for Determination of Contaminants in Soil.* Proceedings of the Estonian Academy of Sciences, 74, 1, 71–81. DOI:10.2139/ssrn.4789276.

K. Ng; N. Rajagopalan. (2009). *Application of quality by design and risk assessment principles for the development of formulation design space*. Quality by Design for Biopharmaceuticals, Wiley, Hoboken, NJ, 161–174.

Kalbe, U.; Lehnik-Habrink, P.; Bandow, N.; Sauer, A. (2019). *Validation of European horizontal methods for the analysis of PAH, PCB, and dioxins in sludge, treated biowaste and soil.* Environ. Sci. Eur., 31 (29) https://doi:10.1186/s12302-019-0211-3.

Kardosa, P.; Lahutaa, P.; Hudakova, M. (2021). *Risk Assessment Using the FMEA method in the Organization of Running Events.* Transportation Research Procedia, 55, 1538–1546. https://doi.org/10.1016/j.trpro.2021.07.143.

Koch, M.; Liebich, A.; Win, T.; Nehls, I. (2005). *Certified Reference Materials for the determination of mineral oil hydrocarbons in water, soil and waste.* Forschungsbericht 272, Berlin.

Koel, M.; Jurjeva, J. (2022). *Implementing greening into design in analytical chemistry.* Talanta Open, 6, 100136. https://doi.org/10.1016/j.talo.2022.100136.

Konečný, F.; Boháček, Z.; Müller, P.; Kovářová, M.; Sedláčková, I. (2003). *Contamination of soils and groundwater by petroleum hydrocarbons and volatile organic compounds – Case study: ELSLAV BRNO.* Bulletin of Geosciences, 78(3), 225–239.

Krzemień-Konieczka, I.; Buszewski, B. (2015). *Determining Polychlorinated Biphenyls in Soil Using Accelerated Solvent Extraction (ASE).* Pol. J. Environ. Stud., 24(5), 2029–2033. https://doi.org/10.15244/pjoes/41589.

Kumar, A.; Ashok, M.; Sawhney R. (2011). *Crude oil PAH constitution, degradation pathway and associated bioremediation microflora: an overview.* International journal of environmental sciences, 1, 7.

Kumar, B.; Verma, V. K.; Gaur, R.; Kumar, S.; Sharma, C. S.; Akolkar, A. B. (2014). *Validation of HPLC method for determination of priority polycyclic aromatic hydrocarbons (PAHS) in wastewater and sediments.* Adv. Appl. Sci. Res., 5 (1), 201-209.

Lau, E.V.; Gan, S.; Ng, H.K. (2010). *Extraction Techniques for Polycyclic Aromatic Hydrocarbons in Soils.* International Journal of Analytical Chemistry, 398381. https://doi.org/10.1155/2010/398381.

Lubes, G.; Goodarzi, M. (2017). *Spectroscopic method are still used for screening in some oil spill investigations, but capillary gas chromatography (GC) clearly has proven most effective alternative.* Chem. Rev., 117, 6399−6422.

Mali, M.; Dell'Anna, M.M.; Notarnicola, M.; Damiani, L.; Mastrorilli, P. (2017). *Combining chemometric tools for assessing hazard sources and factors acting simultaneously in contaminated areas. Case study: "Mar Piccolo" Taranto (South Italy).* Chemosphere, 184, 784-794.

Manousi, N.; Wojnowski, W.; Płotka-Wasylka, J.; Samanidou, V. (2023). *Blue applicability grade index (BAGI) and software: a new tool for the evaluation of method practicality.* Green Chem., 25, 7598–7604. https://doi.org/10.1039/d3gc02347h.

Marinescu, M.; Toti, M.; Tanase, V.; Carabulea, V. Plopeanu,G.; Calciu, I. (2010). *An assessment of the effects of crude oil pollution on soil properties.* Annals. Food Science and Technology.

Merck. Available online: https://www.sigmaaldrich.com/analytical-chromatography/proficiency-testing.html

Miki, S.; Uno, S.; Ito, K.; Koyama, J.; Tanaka, H. (2014). *Distributions of polycyclic aromatic hydrocarbons and alkylated polycyclic aromatic hydrocarbons in Osaka Bay, Japan.* Marine Pollution Bulletin, 85, 2, 30, 558-565. https://doi.org/10.1016/j.marpolbul.2014.04.004.

Moosavi, S.M.; Ghassabian, S. (2017). *Linearity of Calibration Curves for Analytical Methods: A Review of Criteria for Assessment of Method Reliability.* In book: Calibration and Validation of Analytical Methods - A Sampling of Current Approaches. http://dx.doi.org/10.5772/intechopen.72932.

Murphy, K.R.; Wenig, P.; Gavin Parcsi, G.; Thomas Skov, T.; Stuetz, R.M. (2012). *Characterizing odorous emissions using new software for identifying peaks in chemometric models of gas chromatography–mass spectrometry datasets.* Chemometrics and Intelligent Laboratory Systems, 118, 41-50, doi:10.1016/j.chemolab.2012.07.006.

Nordtest project 1589-02. (2012). *Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories.* Version 3.1.

Nováka, M.; Palya, D.; Bodai, Z.; Nyiri, Z.; Magyar, N.; Kovács, J.; Eke, Z. (2017). *Combined cluster and discriminant analysis: An efficient chemometric approach in diesel fuel characterization.* Forensic Science International, 270, Pages 61-69.

Oksanen, J. R package version. (2011). *Multivariate Analysis of Ecological Communities in R: vegan tutorial*.

Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D. (2018). *Application of chemometric methods to XRF-data - A tutorial review.* Analytica Chimica Acta, 1-14.

Pena-Pereira, F.; Wojnowski, W.; Tobiszewski, M. (2020). *AGREE—Analytical GREEnness Metric Approach and Software.* Analytical Chemistry, 92, 14

Pongpiachan, S.; Hattayanone, M.; Tipmanee, D.; Suttinun, O.; Khumsup, C.; Kittikoon, I.; Hirunyatrakul, P. (2018). *Chemical characterization of polycyclic aromatic hydrocarbons (PAHs) in 2013 Rayong oil spill-affected coastal areas of Thailand.* Environmental Pollution, 233, 992-1002. https://doi.org/10.1016/j.envpol.2017.09.096.

Raschka, S. (2014). About *Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms.* Available online: https://sebastianraschka.com/Articles/2014_about_feature_scaling.html.

Reflab metode 4:2008. Bestemmelse af kulbrinter, BTEX og PAH i jord ved gaskromatografi. Available online: http://cdnmedia.eurofins.com/Microsites/media/1148/metode_4_2008_2_udg.pdf, 2008

Retnam, A.; Zakaria, M.P.; Juahir, H.; Aris, A. Z.; Zali, M. A.; Kasim, M. F. (2013). *Chemometric techniques in distribution, characterisation and source apportionment of polycyclic aromatic hydrocarbons (PAHs) in aquaculture sediments in Malaysia.* Marine Pollution Bulletin, 69(1-2), 55–66. DOI:10.1016/j.marpolbul.2013.01.009.

Risvik, H. (2007). *Principal Component Analysis (PCA) & NIPALS algorithm.* Available online: https://folk.uio.no/henninri/pca_module/pca_nipals.pdf.

Romesburg, H.C. (2004.) *Cluster Analysis for Researchers.* Lulu Press.

Rozet, E.; Lebrun, P.; Debrus, B.; Boulanger, B.; Hubert, P. (2013). *Design Spaces for analytical methods.* Trends in Analytical Chemistry 42. https://doi.org/10.1016/j.trac.2012.09.007.

Sadler, R.; Connell, D. (2003). *Analytical Methods for the Determination of Total Petroleum Hydrocarbons in Soil.* Proceedings of the Fifth National Workshop on the Assessment of Site Contamination, 133-150. https://www.nepc.gov.au/sites/default/files/2022-09/asc-wkshoppaper-09-tphs-sadler-petrol-hydro-soil-200301.pdf.

Sajid, M.; Płotka-Wasylka, J. (2022). *Green analytical chemistry metrics: A review.* Talanta, 238, 2, 123046https://doi.org/10.1016/j.talanta.2021.123046

Saroj, S.; Jairaj, V.; Rathod, R. (2017). *Applying green analytical chemistry for development and validation of RP-HPLC stability indicating assay method for estimation of fenoverine in bulk and dosage form using quality by design approach.* Journal of Liquid Chromatography & Related Technologies, 40, 7, 340-352. https://doi.org/10.1080/10826076.2017.1304414

Singh, I.; Juneja, P.; Kaur, B.; Kumar, P. (2013). *Pharmaceutical Applications of Chemometric Techniques.* Hindawi Publishing Corporation ISRN Analytical Chemistry, 795178.

Swedish Chemicals Agency. Swedish rules on chlorinated solvents https://www.kemi.se/en/rules-and-regulations/rules-applicable-in-sweden-only/certain-swedish-restrictions-and-bans/chlorinated-solvents.

Szoleczky, R.; Kovács, A.; Berkó, S.; Budai-Szűcs, M. (2024). *An Analytical Target Profile for the Development of an In Vitro Release Test Method and Apparatus Selection in the Case of Semisolid Topical Formulations.* Pharmaceutics, 16(3), 313. https://doi.org/10.3390/pharmaceutics16030313.

Thompson, M.; S. Ellison, L. R.; Wood, R. (2002). *Harmonized guidelines for single laboratory validation of methods of analysis (IUPAC Technical Report).* Pure Appl. Chem., 74 (5), 835–855.

Tome, T.; Žigart, N.; Časar, Z.; Obreza, A. (2019). *Development and Optimization of Liquid Chromatography Analytical Methods by Using AQbD Principles: Overview and Recent Advances*. Org. Process Res. Dev. 23, 1784−1802

United nations industrial development organization. (2009). *Complying with ISO 17025. A practical guidebook for meeting the requirements of laboratory accreditation schemes based on ISO 17025:2005 or equivalent national standards.* https://www.unido.org/sites/default/files/2010-08/Complying_with_ISO_17025_A_practical_guidebook_0.pdf

Volk, S.; Gratzfeld-Huesgen, A. (2011). *Analysis of PAHs in soil according to EPA 8310 method with UV and fluorescence detection. Agilent Application Solution.* Available online: https://www.agilent.com/cs/library/applications/5990-8414EN.pdf.

Włóka, D.; Placek, A. (2015). *Study of the effectiveness of the new method for the rapid PAHs analysis in soil.* Acta Innovations, 17, 5-11.

Yabréa, M.; Fereya, L.; Soméb, T.I.; Sivadierc, G.; Gaudina, K. (2020). *Development of a green HPLC method for the analysis of artesunateand amodiaquine impurities using Quality by Design.* Journal of Pharmaceutical and Biomedical Analysis, 190, 113507. DOI: 10.1016/j.jpba.2020.113507

Yamada, T.M.; Souza, D.A.; Morais, C.R.; Mozeto, A.A. (2009). *Validation of a Method for the Analysis of PAHs in Bulk Lake Sediments Using GC–MS*. J. Chromatogr. Sci., 47 (9),794-799. https://doi:10.1093/chromsci/47.9.794.

Zhao, Y.; Hong, B.; Fan, Y.; Wen, M.; Han, X. (2014). *Accurate analysis of polycyclic aromatic hydrocarbons (PAHs) and alkylated PAHs homologs in crude oil for improving the gas chromatography/mass spectrometry performance.* Ecotoxicology and Environmental Safety, 100, 2014, 242-250. https://doi.org/10.1016/j.ecoenv.2013.10.018.

Zhou, W.; Yang, S.; Wang, P. G. (2017). *Matrix effects and application of matrix effect factor.* Bioanalysis 9, 23. https://doi.org/10.4155/bio-2017-0214.

Zuur, A.; Ieno, E.; Meesters; E. A. (2009). *Beginner's Guide to R. Springer Publishing Company.*

Acknowledgements

This work was performed at School of Engineering, Department of Energy Technology, Tallinn University of Technology and at Eurofins Environment Testing Estonia OÜ. Part of this thesis was carried out at The J. Molner Company OÜ.

I would like to thank the Department of Energy Technology and all its member's staff for all the considerate guidance, especially my supervisor **Dr. Mihkel Koel**, without whom I would not have been able to complete this research, and without whom I would not have made it through my PhD! Mihkel continuously provided encouragement and was always willing and enthusiastic to assist in any way he could throughout the research project. Mihkel provided great ideas regarding analysis, helped in writing the research papers. I am profoundly grateful to him for his sage advice, rigorous academic guidance, and the confidence he instilled in me. His expertise in the field of analytical chemistry greatly enriched my work. I am extremely grateful for our friendly chats and his personal support in my academic and business endeavours. I am deeply grateful to **Prof. Mihkel Kaljurand** for revising my work and to my opponents **Dr. Riin Rebane** and **Prof. Miguel de la Guardia Cirugeda.**

It has been an amazing experience working in two different research groups across two different companies during my PhD study. Thanks to everyone from the Eurofins Environment Testing Estonia OÜ I've worked with and from whom I've learned so much about gas chromatography, method validation and environmental analysis. Thank you, The J. Molner Company OÜ from whom I am continuously learning about generic drug development and solving analytical chemistry challenges.

I would like to express my deepest gratitude to my family for their unwavering support throughout this journey. To my spouse Pavel, to my children Konstantin, Vladislav and Diana, who believes in me and have kept my spirits and motivation high during this process. I am also thankful for my friends, who provided both distractions when needed and encouragement when it seemed impossible to continue.

I would like to acknowledge the financial support from the Estonian Research Council for supporting this work. Last but not least, I would like to thank late **Hans Luik** who was my supervisor and gave me an opportunity of starting my PhD journey.

Abstract

Analytical Quality by Design in environmental analysis

Analytical Quality by Design (AQbD) is a holistic method and development approach that encompasses the entire analytical procedure, from risk assessment to lifecycle management. The enhanced AQbD approach reduces the time and effort necessary to develop reliable analytical methods, enables flexible change control through the method operable design region, and lowers out-of-specification results. In this work, the AQbD approach was chosen for analytical method development. This approach aided in the simultaneous identification and quantification of polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), alkylated PAHs, and aliphatic compounds in soil and sediment by gas chromatography-mass spectrometry (GC-MS). Additionally, the mechanical extraction method was employed for sample preparation, resulting in the development of a modified method with superior analytical parameters. Subsequently, the set-up of the Design Space (DS) for method development was investigated, including the selection of a suitable GC column. The obtained experimental design allowed the determination of the optimal GC conditions to develop a new method that yields adequate resolution and recovery. The method development process involved modifying the sample preparation procedure to make it simpler and faster. Decreasing the number of extraction steps contributed to reducing the staff required and the consumable costs for extraction. Thus, a method to analyse PAHs, PCBs, and aliphatic and aromatic compounds in a single run with reliable accuracy, precision, linearity, and recovery rates was developed. This new method was validated in the laboratory, and an inter-laboratory comparison was also carried out. With this new validated method and chemometric tools (principal component analysis (PCA) and cluster analysis (CA)), the source of soil contamination could be identified. Statistical methods (Pearson's product-moment correlation (PPMC) and the Shapiro-Wilk test) were employed to compare different analytical methods for analysing soil samples polluted with PAHs. The Analytical Greenness Metric (AGREE) was employed to assess the greenness of the methods considering the twelve principles of green analytical chemistry (GAC). Thus, this study developed a robust, rapid, simple, sensitive, and selective GC method for PAHs, PCBs, and aliphatic and aromatic compounds. The method was validated as per the guidelines established by the Nordtest Guide and the International Council for Harmonisation (ICH).

Lühikokkuvõte

Analüütilise disainiruumi kontseptsiooni rakendamine saasteainete pinnases määramise meetodi väljatöötamisel

Analüütiline disainiruumi kontseptsioon (AQbD) on terviklik meetodite arendamise lähenemisviis, mis hõlmab kogu analüütilist protseduuri alates riskihindamisest kuni meetodi elutsükli juhtimiseni. Täiustatud AQbD lähenemine vähendab usaldusväärsete analüütiliste meetodite väljatöötamiseks vajalikku aega ja töömahtu, võimaldab paindlikku muudatuste kontrolli läbi meetodi toimiva disainipiirkonna ning vähendab väljaspool spetsifikatsioone (out-of-specification) tulemuste esinemist.

Käesolevas töös valiti analüütilise meetodi arendamiseks AQbD lähenemine. See võimaldas samaaegselt tuvastada ja kvantifitseerida polütsüklilisi aromaatseid süsivesinikke (PAH-id), polüklooritud bifenüüle (PCB-id), alküülitud polütsüklilisi aromaatseid süsivesinikke ja alifaatseid ühendeid pinnases ja setetes, kasutades gaaskromatograafiat-massispektromeetriaga (GC-MS).

Proovide ettevalmistamiseks kasutati mehaanilist ekstraheerimismeetodit, mille tulemusel töötati välja muudetud meetod paremate analüütiliste parameetritega. Edasi uuriti meetodi arendamiseks vajaliku disainiruumi, sealhulgas sobiva GC kolonni valikut. Eksperimentaalne disain võimaldas määrata optimaalseid GC tingimuseid uue meetodi väljatöötamiseks, mis tagas piisava lahutusvõime ja saagise.

Meetodi arendamise käigus lihtsustati ja kiirendati proovide ettevalmistamise protseduuri. Ekstraheerimisetappide arvu vähendamine aitas kaasa tööjõukulu ja tarvikukulude vähenemisele. Seega töötati välja meetod, mis võimaldab määrata PAH-id, PCB-id ning alifaatseid ja aromaatseid ühendeid ühe mõõtmistsükli jooksul usaldusväärse täpsuse, lineaarsuse ja saagisega.

Uus meetod valideeriti laboris ning viidi läbi ka laboritevaheline võrdluskatse. Valideeritud meetodi ja keemomeetriliste tööriistade (peakomponentide analüüsi ja klasteranalüüsi) abil oli võimalik määrata pinnase saastumise allikas. Statistilisi meetodeid, nagu Pearsoni korrelatsioonikordaja ja Shapiro-Wilki test, kasutati erinevate analüütiliste meetodite võrdlemiseks PAH-idega saastunud pinnaseproovide analüüsimisel. Lisaks kasutati Analüütilise Rohelisuse Maatriksi (AGREE) meetodite keskkonnasõbralikkuse hindamiseks, tuginedes rohelise analüütilise keemia kaheteistkümnele põhimõttele.

Seega töötati selles uuringus välja usaldusväärne, kiire, lihtne, tundlik ja selektiivne GC meetod PAH-ide, PCB-de ning alifaatsete ja aromaatsete ühendite määramiseks. Meetod valideeriti vastavalt Nordtesti juhistele ja rahvusvahelise nõuete ühtlustamise nõukogu (ICH) juhistele.

Appendix 1

Publication I

Jurjeva, J., Koel, M. (2019). The chemometric approach to identification of residual oil contamination at former primitive asphalt pavement plants. Oil Shale, 36(3), 410-430. https://doi.org/10.3176/oil.2019.3.04

Oil Shale, 2019, Vol. 36, No. 3, pp. 410–430
doi: https://doi.org/10.3176/oil.2019.3.04
ISSN 0208-189X

The chemometric approach to identification of residual oil contamination at former primitive asphalt pavement plants

Jelena Jurjeva[(a)]*, Mihkel Koel

[(a)] Department of Materials and Environmental Technology, Tallinn University of Technology, Ehitajate tee 5, 19086 Tallinn, Estonia; Eurofins Environment Testing Estonia OÜ, Paavli 5–3, 10412 Tallinn, Estonia
[(b)] Department of Chemistry and Biotechnology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

Abstract. *This study investigated polycyclic aromatic hydrocarbons (PAHs) and hydrocarbon oil index (HOI) pollution in the soil on the territories of two former primitive asphalt pavement plants (APPs) in Estonia. The standard quantitative methods for the chemical characterisation of the oils consisted of an initial screening, by using a gas chromatography-flame ionization detector (GC-FID), and, for a more detailed analysis, of gas chromatography-mass spectrometry (GC-MS). A combination of chemometric and analytical methods was used to identify the sources of PAHs, which could be attributed to the soil pollution at the plants. The identification and classification of oil spills were performed using chemometric techniques, such as the principal component analysis (PCA) and the clustering analysis (CA), which is based on Jaccard similarity. The application of the chemometric techniques enabled the clustering and discrimination of polluted soils into four groups, according to oil type. Several different methods of CA, such as single, complete and average linkages, were tested and the results were compared.*

Keywords: *residual pollution, oil spills, chemometrics, principal component analysis, cluster analysis.*

1. Introduction

Residual pollution is still a crucial environmental problem in Estonia and its clean-up has been too slow. Any assessment of pollution includes a subsequent and constant monitoring of carefully selected parameters, which give information about the contamination size and risks to the groundwater, surface water, air and soil. According to the Estonian Ministry of the Environment, the most serious past contamination was detected in the soil at former primitive asphalt pavement plants (APPs). The main sources of contamination were old oils, which were spilt into the soil. At the moment, there are approximately 30

* Corresponding author: e-mail *jelenajurjeva@eurofins.com*

unclean asphalt pavement plants on the territory of Estonia [1–3].

For the identification of oil pollution the most commonly analysed compounds found in the soil are polycyclic aromatic hydrocarbons (PAHs), volatile organic compounds (benzene, toluene, ethylbenzene and three xylene isomers), polychlorinated biphenyls (PCBs), heavy metals and hydrocarbon oil index (HOI). The U.S. Environmental Protection Agency (EPA) has reported 16 PAH compounds (Fig. 1) as being priority pollutants, including naphthalene (Nap), acenaphthylene (Acy), acenaphthene (Acp), fluorene (Flu), phenanthrene (Phe), anthracene (Ant), fluoranthene (Fla), pyrene (Pyr), chrysene (Chr), benz(*a*)anthracene (B(*a*)Ant), benzo(*b*)fluoranthene (B(*b*)F), benzo(*k*)fluoranthene (B(*k*)F), benzo(*a*)pyrene (B(*a*)P), benzo(*g*,*h*,*i*)perylene (B(*g*,*h*,*i*)P), dibenz(*a*,*h*)anthracene (DiAnt) and indeno(*1*,*2*,*3-cd*)pyrene (Ind) [4, 5].

PAHs are important to be determined because of their possible toxic, carcinogenic and mutagenic properties. These compounds are released into the atmosphere during the incomplete combustion of organic materials (e.g. coal, oil, petrol, wood) and they are then precipitated onto the soil. In the soil, or in sediments, they tend to adsorb tightly onto suspended particulate matter

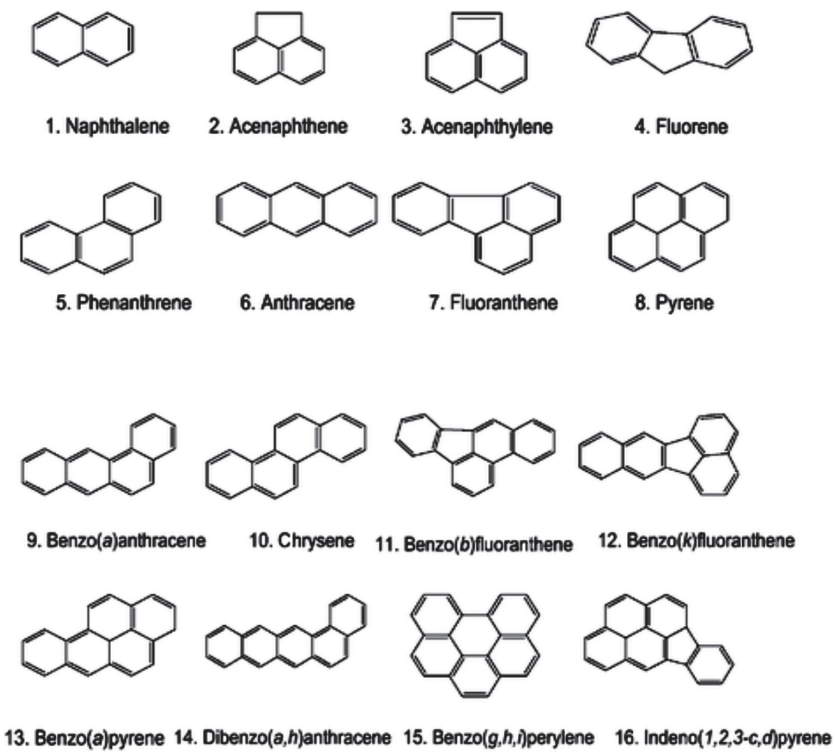

Fig. 1. The chemical structure of 16 EPA-reported PAHs [6].

[7–9]. The source of PAHs contamination in the soil can be pyrogenic, being caused or produced by combustion of heat fuels, or petrogenic, relating to the origin or formation of rocks. PAH diagnostic (binary) ratios can be applied to tracing out the sources of contamination in polluted areas [10, 11].

The hydrocarbon oil index is defined as the total amount of hydrocarbons which can be extracted from the sample by a non-polar solvent and then eluted between *n*-decane ($C_{10}H_{24}$) and *n*-tetracontane ($C_{40}H_{82}$) on an apolar capillary gas chromatographic (GC) column. The analysis of HOI is performed using a gas chromatography-flame ionization detector (GC-FID). The general term "mineral oil" comprises petroleum products, with a complex mixture of hydrocarbons, which can be found in diesel, kerosene, home heating oils, heavy fuel oils, transformer oil and lubricants. Due to the widespread use of mineral oils, these petroleum hydrocarbons are the most common organic contaminants in the soil and sediments, especially at former industrial and military sites. The said hydrocarbons can contaminate water or be consumed by organisms that can enter the human food chain [4, 12–14].

For oil hydrocarbon fingerprinting, the pollution source identification is based on chromatographic methods, which are the most effective for this purpose [15]. Gas chromatography-flame ionization detection (GC-FID) is a technique which is used for an initial screening. Gas chromatography-mass spectrometry (GC-MS) gives information about the content of oil and it is able to separate compounds in complex mixtures [15, 16].

The analytical methods generate a large amount of data, which is sometimes difficult to interpret. The multivariate statistical methods, like the principal component analysis (PCA) and the cluster analysis (CA), give a better resolution and a more adept separation quality of the samples. The chemometric methods extract the hidden information, to individualise and classify the samples into groups. The chemometric approach, together with the chromatographic methods, can help to identify and classify the soil, based on the type of contamination. Various statistical and numerical software programs (SPSS, R, MATLAB, Minitab and Excel Stat) are used to simplify these processes [17–19].

The principal component analysis decomposes the matrix into products of the scores matrix, the transposed loadings matrix plus the residuals matrix. This reduction of the data allows presenting the initial data in new coordinates or principal components (PCs). The newly generated PCs explain most of the information from the dataset. The loadings plot shows the importance of the different variables that are responsible for the clustering in the scores plot. The scores plot provides information about the relationships between individual objects, showing the groups, outliers, etc. [16, 20, 21].

The cluster analysis is used for the grouping of samples, according to the type of similarity. Its main task is to recalculate the numerical values of similarity between the new group and the rest of the objects. The next step consists in the further grouping of the data until all the objects have been

merged into one large group. The output of the hierarchical cluster analysis (HCA) is a dendrogram which visualises the grouping of samples in a two-dimensional space [16, 22, 23].

The study by Mali et al. [19] demonstrated that the chemometric approach (PCA/CA and factorial analysis of variance (ANOVA)) was advantageous for assessing and modelling the contamination patterns of highly polluted areas, and thus, it could contribute to the effective monitoring of their quality. In the study by Miki et al. [24], CA was used to identify the sources of the parent and alkylated PAHs in the sediments. The contaminated sites were categorised on the basis of PAHs composition, in order to find their primary sources within the site groups.

In spite of the numerous studies that have focused on the analysis of PAHs, there are only a few investigations on their distribution and contamination identification on the territories of past plants. In this work, analytical and chemometric tools were used to identify soil contamination at two former primitive asphalt pavement plants, Jänesselja asfaltbetoonitehas in Pärnu County, Southeast Estonia and Maadevahe asfaltbetoonitehas on Saaremaa Island, West Estonia. First of all, GC-FID chromatograms were recorded for an initial screening, in order to determine soil types and estimate the extent of weathering. For a detailed fingerprinting, the content and distribution of PAHs were determined by GC-MS. The binary ratios of PAHs were calculated. The samples from both APPs were compared and, using oil standards (diesel, light fuel oil (LFO), used motor oil, shale oil, heavy fuel oil (HFO)) were classified. PCA and CA were used to cluster the samples. The binary ratios of PAHs were used to distinguish their potential sources in the environmental samples.

2. Material and methods

2.1. Soil samples, reagents and equipment

The chemical fingerprints of 20 spilt oil-containing soil samples from Jänesselja APP and 36 samples from Maadevahe APP were analysed. The samples were collected from various locations and from different depths. Prior to the analyses, the samples were registered and equipped with numbers, and stored at 4 °C. Hexane was used as a solvent for extraction of oils and PAHs. In this study, for PAHs calculation, the specific internal standards (ISTDs) Naphthalene D8, Acenaphthene D10, Phenanthrene D10, Anthracene D10, Pyrene D10, Benz(*a*)anthracene D12, Benzo(*a*)pyrene D12 and Dibenz(*ah*)-anthracene D14 were chosen. The mixtures of the above-mentioned standards were prepared and spiked for all the samples, together with the blank sample and control samples. For the HOI internal standard, *n*-tetracontane with a final concentration of 20 mg/l in an extraction solution (hexane) was used. Analyses for HOI were performed on an Agilent 7890B equipped with a flame ionization detector (FID), and an Agilent 7693 autosampler. The column used

for the analyses was an Agilent J&W GC Column DB-1 (15 m × 320 μm × 0.10 μm). The analyses of the target EPA PAH compounds were performed on an Agilent 6890N GC system equipped with an Agilent Technologies 5973 mass selective detector, and an Agilent 7683 Series Autosampler. The column used was the Zebron 20 m × 180 μm × 0.18 μm ZB-5MS. In case of both instruments, all the samples were injected in a splitless injection mode, with helium as the carrier gas. Further, the PAH quantitation was performed in the selected ion monitoring (SIM) mode. The validations were established based on six calibration points for PAHs and on eight calibration points for HOI, with the correlation coefficient (R^2) greater than 0.995 for each component.

2.2. Samples preparation

The samples were prepared according to a modified method, ISO 16703:2011 [25]. Approximately 10 g of each soil sample was weighed into a tube. The PAH mix internal standards were spiked into the soil samples from APPs, the blank sample and control samples. 10 ml of acetone, 5 ml of the extraction solution with the HOI internal standard and 5 g of NaCl were then added to the sample tube. The samples were shaken for 16 hours. 1 ml of each extract was used for PAHs analysis by GC-MS and approximately 4 ml of the rest of the solutions was eluted through a Florisil column for the HOI analysis by GC-FID. The polar substances were removed by a clean-up with Florisil and the non-polar compounds were eluted through the column with the extraction solutions. One blank sample was processed together with the soil samples. The limit of quantification (LOQ) of each PAH was 0.003 mg/kg and 20 mg/kg for HOI. The identification of the compounds was based on the retention times of the calibration standards. The quantification of PAHs was performed using the internal standard quantification method, by comparing the area of the quantification ion to that of the corresponding deuterated quantification standard.

2.3. Principal component analysis

The initial data matrix consisted of 54 samples (rows) and concentrations of 16 individual PAHs (columns). Two samples were excluded from the chemometric analysis because the concentrations of most of the PAHs contained were below LOQ. The PAH concentrations were calculated as follows:

$$C_{\mathrm{PAH}} = \frac{(C - C_0) \times V}{M}, \tag{1}$$

where C_{PAH} is the PAH concentration in the soil sample, mg/kg (on a dry matter basis); C is the PAH concentration obtained from the calibration curve, mg/l; C_0 is the PAH concentration in the blank sample obtained from the calibration curve, mg/l; V is the extraction solution volume, l; and M is the weight of the

soil sample, kg (on a dry matter basis).

To identify soil pollution on a dry matter basis, the binary ratios of PAH isomers, such as Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(*g*,*h*,*i*)P) and B(*a*)Ant/(B(*a*)Ant + Chr), and the ratios of the four-to-six-ring parent PAHs to the sum of the two- and three-ring parent PAHs (HMW/LMW) were calculated. In PCA, a matrix consisting of 54 samples and 7 PAH binary ratios was used.

Prior to PCA, the data was standardised using two methods, Min-Max scaling and autoscaling. In the Min-Max scaling (normalisation), the data was scaled to a fixed range, usually to 0 to 1, and this was typically calculated via the following equation:

$$X\text{norm} = \frac{X - X\text{min}}{X\text{max} - X\text{min}}, \tag{2}$$

where *X*norm stands for the the normalised *X* value, *X*min is the minimum *X* value and *X*max is the maximum *X* value.

In order to avoid the problem of incompatibility between different scales, the data was often centred and all the values were divided by the standard deviation for each variable. The autoscaling was performed via the following equation:

$$X\text{aut} = \frac{X - X\text{avg}}{X\text{std}}, \tag{3}$$

where *X*aut means the *X* value (recalculated value) after using the autoscaling method, *X*avg stands for the average *X* value and *X*std denotes the standardised *X* value.

The Min-Max scaled and autoscaled data were then subjected to PCA. In this study, the Microsoft Excel Macro using the nonlinear iterative partial least squares (NIPALS) algorithm was applied to finding the eigenvectors of the most important principal components [26–28].

2.4. Cluster analysis

The hierarchical clustering methods grouped the objects according to the similarity between them. Based on the Jaccard similarity index, two sites that were most similar to each other were classified into one group. The Jaccard similarity index was calculated by the following equation:

$$Sj = \frac{c}{a + b - c}, \tag{4}$$

where *a* is the number of PAH species in sample A, *b* is the number of PAH species in sample B and *c* is the number of similar PAH species in sample A and sample B [29, 30].

To quantify the distance between the two clusters, single, complete and average linkages were used.

2.4.1. Single linkage clustering

As in all agglomerative cluster analyses, the single linkage began with a matrix of similarity (or dissimilarity) coefficients. Firstly, the most similar pair(s) of the samples, or the first clusters, was(were) found. The second most similar pair(s) of the samples, or the highest similarity between a sample and the first cluster, whichever was greater, was(were) then found. A disadvantage of the single linkage clustering was that it tended to produce long clusters.

2.4.2. Complete linkage clustering

Unlike the single linkage clustering, the complete linkage clustering was often inclined to generate opposite extremes by producing highly compact clusters. The method calculated the similarity measures after the new groups were formed and two groups with the highest similarity were always merged first.

2.4.3. Average linkage clustering

The average linkage clustering, differently from the single and complete linkage clusterings, produced no extremes. In order to compute the average similarity between a sample and the existing cluster, the type of "average" had to be precisely defined by using the unweighted or weighted technique. If group A consisted of *Na* objects and group B of *Nb* objects, then in the unweighted technique, the new similarity (*Sab*) was calculated by the following equation:

$$Sab = \frac{NaSa+NbSb}{Na+Nb}, \quad (5)$$

where *Na* is the number of objects in group A and *Nb* is the number of objects in group B.

The weighted mean value was calculated as follows [23, 30]:

$$Sab = \frac{Sa+Sb}{2}. \quad (6)$$

For defining the similarity between the samples, R software was used. R is a programming language for statistical analysis, graphical representation and reporting [31].

3. Results and discussion

3.1. Initial screening

In this study, GC-FID was used for an initial screening of oils. The respective chromatograms of soil samples were compared with those of control oils and the levels of oils weathering were estimated. Figure 2 shows the chromatograms of control oils used in the study for identification of the following oil spills: diesel, LFO, shale oil grade C, used motor oil, fresh motor

oil and HFO. The retention times between 3.5 and 6 min were attributable to HOI from decane (C_{10}) to heneicosane (C_{21}), and the retention times between 6 and 8.5 min were assignable to fractions C_{21}–C_{40}. From Figure 2a it can be seen that the major components in diesel were alkanes C_{10}–C_{24}, while LFO (Fig. 2b) contained mostly alkanes C_{10}–C_{32}. Motor oils (Figs. 2d and 2e) were found to consist mainly of unresolved complex mixtures (UCMs), while no UCMs were identified in diesel (Fig. 2a) or LFO (Fig. 2b). Figure 3 depicts the GC-FID chromatograms of soil samples from Jänesselja and Maadevahe APPs. All the chromatograms indicate the presence of the mixture of different oils. This study found no samples that would have been polluted with one type of oil which would have been similar to control oils. The samples from Jänesselja APP reveal different patterns of chromatograms, depending on the source background (Figs. 3b, 3c, 3d). These chromatograms display pollution with different oils, like diesel, LFO, lubricating oil, HFO, waste oil and shale oil. Hydrocarbons are represented by a wide range of species, from *n*-decane

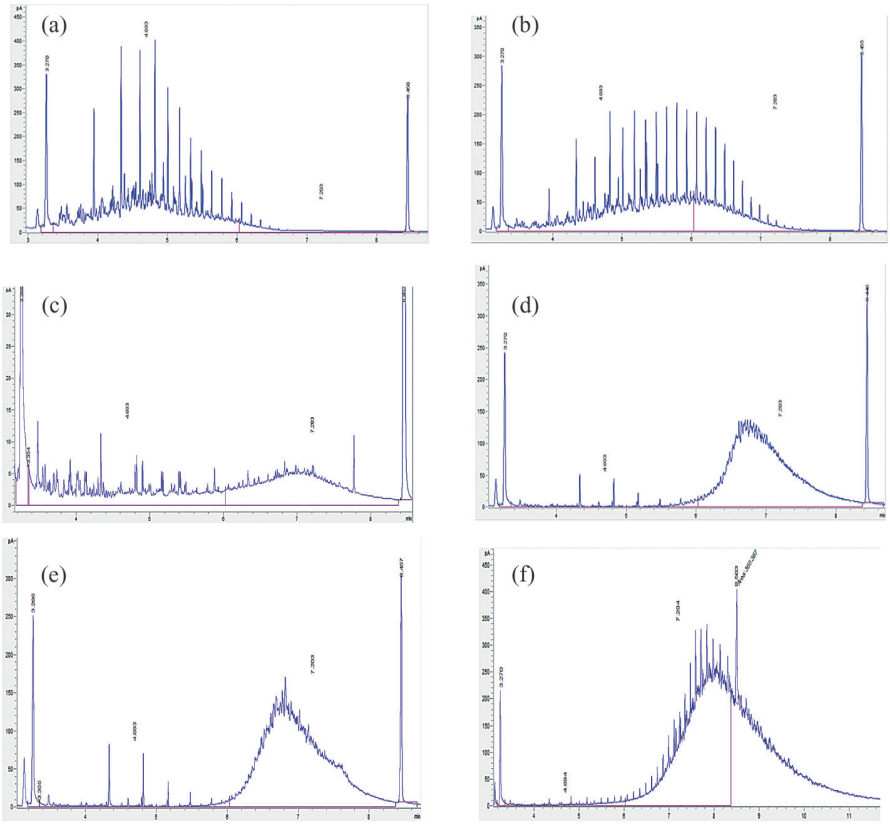

Fig. 2. GC-FID chromatograms of control oils: (a) diesel, (b) LFO, (c) shale oil grade C, (d) used motor oil, (e) fresh motor oil, (f) HFO.

to high molecular weight paraffins (> C_{40}). The GC-FID chromatograms of selected oil samples from Maadevahe APP (Fig. 3a) exhibit a similar pattern of chromatograms (hydrocarbons C_{10}–C_{40}), which suggests soil pollution sourced from one type of oil or a mixture of oils. Based on this information, mainly shale oil was used by Maadevahe APP [32].

The soil samples from Maadevahe APP (Fig. 3a) and sample 16 from Jänesselja APP (Fig. 3b) were probably polluted with a mixture of diesel, sample 14 (Fig. 3c) was polluted with a mixture of LFO, waste oil and HFO, and sample 4 (Fig. 3d), with a mixture of LFO and shale oil. When petroleum products are released into the environment, they tend to weather by evaporation, water solubilisation and oxidation. Oxygen from the air and biological organisms transform the petroleum products, increasing the degradation of *n*-alkanes. After *n*-alkanes have been removed, the remaining constituents appear as a hump on the GC-FID chromatograms, with a few discernible peaks. The chromatograms of oil samples from both APPs show much lower *n*-alkane concentrations (small peaks) and higher phytane concentrations in them, in addition to UCMs, than in fresh oils. The loss of *n*-alkanes and increase in UCMs are indicative of oil weathering processes [33, 34].

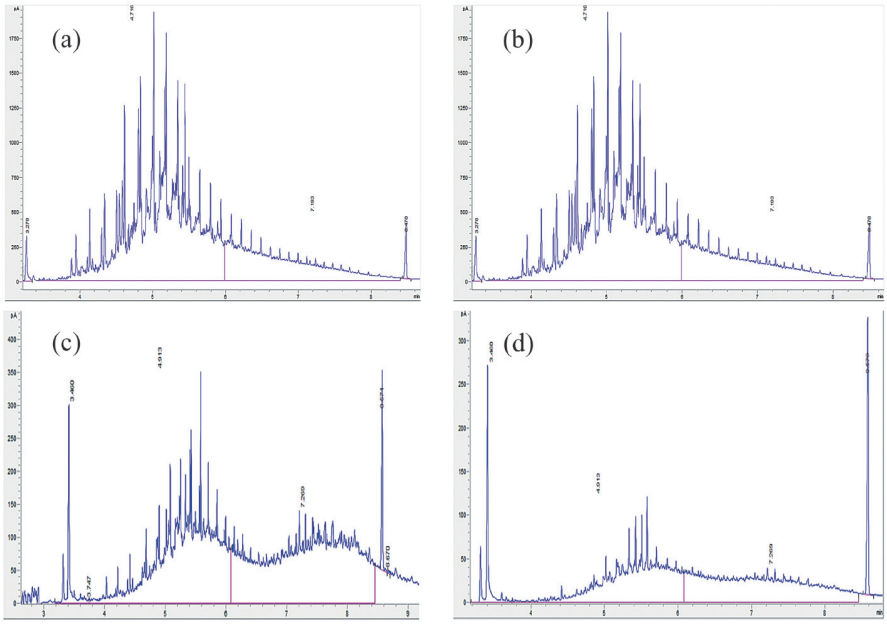

Fig. 3. GC-FID chromatograms of samples from: (a) Maadevahe APP, (b)–(d) Jänesselja APP.

3.2. Determination of PAHs in soils

PAHs in soil were identified using a modified method, ISO 18287:2006 [35]. For the identification of the compounds, one appropriate analyte retention time was chosen and two or three different fragmentation ions (quantifier and qualifier) were used (Table 1). The qualifier to quantifier ion area ratio, m/z, is unique and should not differ more than ± 20% from the same ratio in calibration standards.

The most sensitive ions were used to calculate the contents of individual PAHs by using the corresponding calibration curve. The analyte concentrations were calculated from the ratio of the samples to the internal standard peak area.

Table 1. Ions used for the qualification of PAHs and ISTDs used for the quantification of the final concentrations of PAHs

PAH	Mass fragment, m/z	Internal standard (ISTD) and m/z
Naphthalene	128, 127	Naphthalene D8; 136,137
Acenaphthylene	152, 151	Acenaphthene D10; 162, 164
Acenaphthene	154, 152	Acenaphthene D10; 162, 164
Fluorene	166, 165	Acenaphthene D10; 162, 164
Phenanthrene	178, 176	Phenanthrene D10; 188,184
Anthracene	178, 176	Anthracene D10; 188, 184
Fluoranthene	202, 200	Pyrene D10; 212, 210
Pyrene	202, 200	Pyrene D10; 212, 210
Benz(*a*)anthracene	228, 226	Benz(*a*)anthracene; 240, 236
Chrysene	228, 226	Benz(*a*)anthracene; 240, 236
Benzo(*b*)fluoranthene	252, 250	Benzo(*a*)pyrene D12; 264, 260
Benzo(*k*)fluoranthene	252, 250	Benzo(*a*)pyrene D12; 264, 260
Benzo(*a*)pyrene	252, 250	Benzo(*a*)pyrene D12; 264, 260
Dibenz(*ah*)anthracene	278, 277, 276	Dibenz(*ah*)anthracene D14;292, 291
Benzo(*g*,*h*,*i*)perylene	276, 277	Dibenz(*ah*)anthracene D14;292, 291
Indeno(*1*,*2*,*3-cd*)pyrene	276, 277	Dibenz(*ah*)anthracene D14;292, 291

3.3. Detailed fingerprinting

3.3.1. Soil samples contamination identification using PAH binary ratios and PCA

Seven PAH diagnostic ratios, namely Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(*g*,*h*,*i*)P), B(*a*)Ant/(B(*a*)Ant + Chr) and HMW/LMW, as well as PCA were used to identify the sources of oil in the soil samples from Jänesselja and Maadevahe APPs. The same PAH diagnostic ratios were also calculated for control oils, like diesel, LFO, used motor oil, shale oil and HFO, and compared with those for the soil samples. Figure 4 shows the average PAH binary ratios for the samples from both APPs, with standard deviations. All the above ratios of samples from Jänesselja APP had higher standard deviations, which indicated that the samples differed from each other to a greater degree than did those from Maadevahe APP. The high dispersion of Ant/Phe and HMW/LMW ratios just indicates a wider distribution in the grouping of the samples in further PCA, but this does not influence the grouping itself.

The average binary ratio of Ind/(Ind + B(*g*,*h*,*i*)P) for all the 54 samples was 0.48. One sample had a value > 0.5, which might be due to biomass combustion. The other samples had values between 0.42 and 0.5, indicating petroleum combustion as the possible source of PAHs. The variation in PAH sources revealed by these indices could be due to the different sampling sites, with different hydrological and spatial conditions. The Ant/(Ant + Phe) for

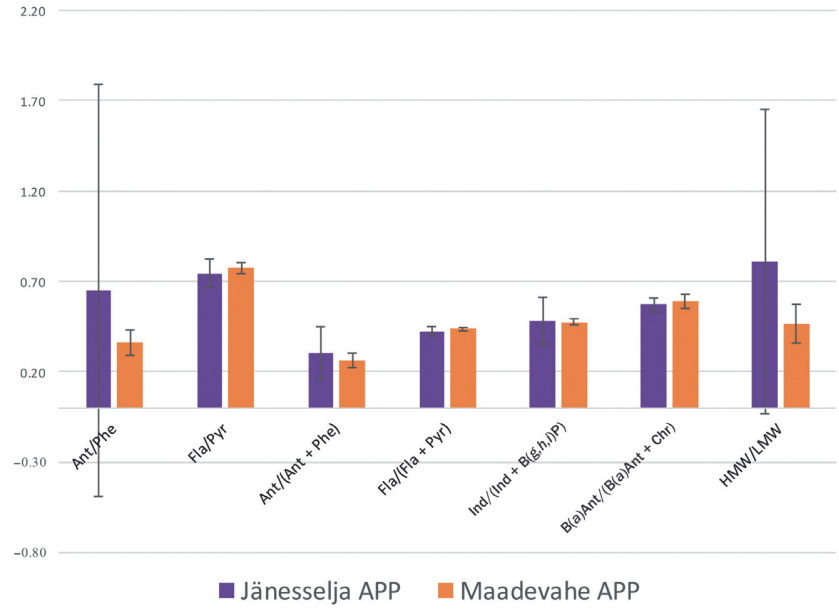

Fig. 4. The average PAH binary ratios for the samples from Jänesselja and Maadevahe APPs, with standard deviations.

Table 2. PAH binary ratios for samples from Jänesselja and Maadevahe APPs

PAH binary ratio / Sample	Ant/Phe	Fla/Pyr	Ant/(Ant + Phe)	Fla/(Fla + Pyr)	Ind/(Ind + B(*g*,*h*,*i*)P)	B(*a*)Ant/(B(*a*)Ant + Chr)	HMW/LMW
Jänesselja APP	0.21–5.19	0.63–0.95	0.18–0.84	0.38–0.49	0.42–0.71	0.51–0.65	0.02–2.63
Maadevahe APP	0.18–0.50	0.70–0.87	0.15–0.33	0.41–0.46	0.44–0.50	0.38–0.62	0.06–0.70

the samples from Jänesselja APP was in the range of 0.18–0.84 and for those from Maadevahe APP, between 0.15 and 0.33 (Table 2). Generally, Ant/(Ant + Phe) < 0.1 indicates that PAHs in the soil sourced from petroleum combustion, whereas Ant/(Ant + Phe) > 0.1 might mean contamination from wood and coal combustion [10].

PAHs in the soils with Fla/(Fla + Pyr) < 0.4, B(*a*)Ant/(B(*a*)Ant +Chr) < 0.2 and Ind/(Ind + B(*g*,*h*,*i*)P) < 0.2 were mainly from petroleum contamination. PAHs with ratios of 0.4 < Fla/(Fla + Pyr) < 0.5, 0.2 < B(*a*)Ant/(B(*a*)Ant + Chr) < 0.4 and 0.2 < Ind/(Ind + B(*g*,*h*,*i*)P) < 0.35 originated predominantly from the combustion of petroleum. PAHs with Fla/(Fla + Pyr) > 0.5, B(*a*)Ant/(B(*a*)Ant + Chr) > 0.4 and Ind/(Ind + B(*g*,*h*,*i*)P) > 0.35 came chiefly from the combustion of coal, shale oil and biomass. Most soil samples from APPs had Fla/(Fla+Pyr) < 0.5, B(*a*)Ant/(B(*a*)Ant + Chr > 0.4 and Ind/(Ind + B(*g*,*h*,*i*)P) > 0.35. The above values suggested that contamination at both APPs originated from the combustion of oil shale and petroleum [10].

The scores plot in Figure 5 shows the PAH binary ratios of 54 samples from APPs, which contained diesel, LFO, used motor oil, shale oil and HFO as the first and second principal components (PC1 and PC2 respectively). The first two PCs comprised 70.69% of the total components of the complex mixture of oil samples. The first principal components were responsible for 45.46% of the total oil spill compounds, while the second principal components were responsible for 25.23% of the total variance. The scores plot reveals that the soil samples mostly contained used motor oil. In this study, the researchers used for analysis fresh motor oil in the mixture of different used motor oils, from diesel to engine petroleum. Diesel, shale oil, HFO and LFO were not identified in the soil samples. The chemical composition of weathered oils differed from that of fresh oils. The most noticeable changes in PAHs composition in the weathered soils were the depletion of naphthalenes, degradation of alkylated PAHs and increase of chrysenes [33, 36]. Being situated distant from the other

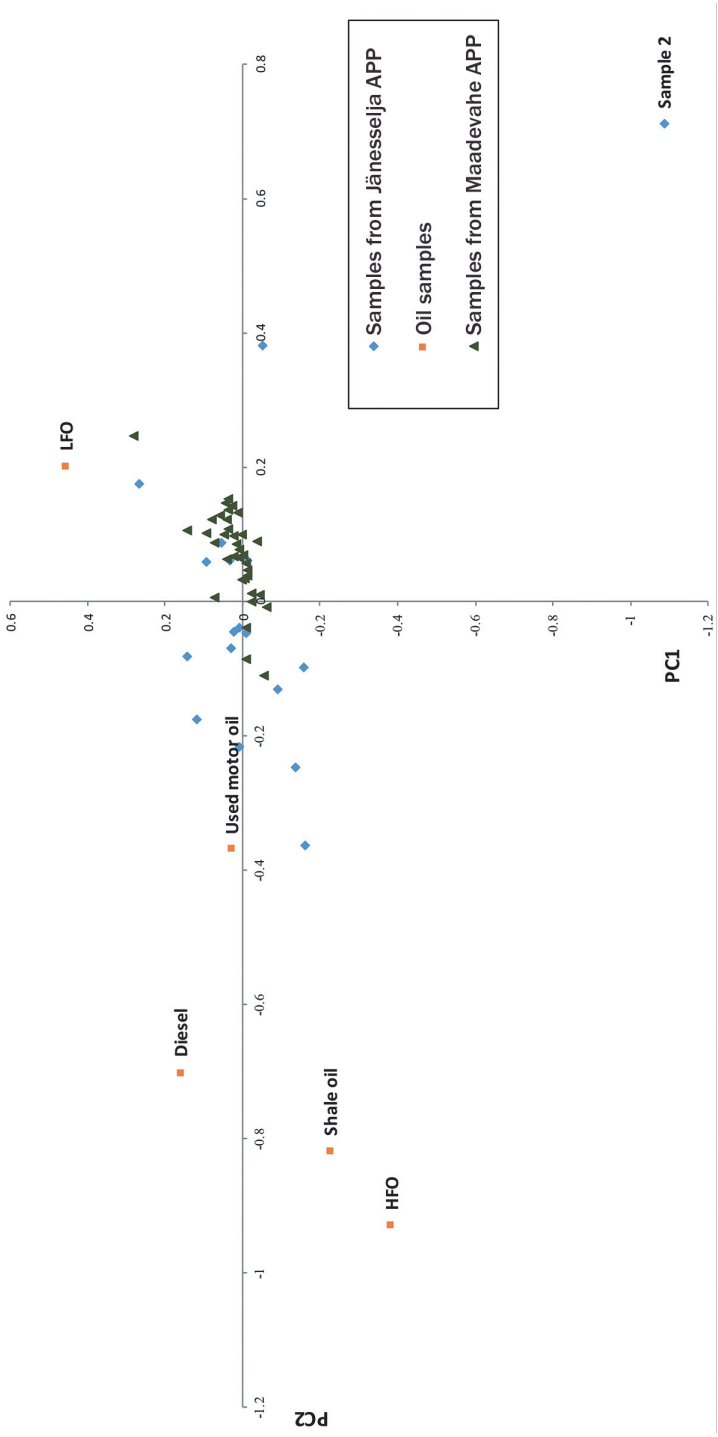

Fig. 5. The scores plot of PAH binary ratios of the samples from Jänesselja and Maadevahe APPs and oil samples.

soil samples and control oils in Figure 2, sample 2 from Jänesselja APP was identified as an outlier. The Ant/Phe ratio of the sample was 5, while for all the other samples it was < 0.66. The source of anthracene in sample 2 could not be identified.

3.3.2. The classification of soil samples from Jänesselja APP

To group the soil samples from Jänesselja APP, another PCA approach was applied, leaving sample 2 out as an outlier from further analysis (Fig. 6).

Sample 4 was classified into an individual cluster. Unlike all the other samples, this sample had a high Fla/Pyr ratio. With Ind/(Ind + B(*g*,*h*,*i*)P) > 0.5, sample 13 was distinguished from the other samples whose respective ratio was considerably lower. This suggested contamination of the sample with PAHs from combustion of biomass. The HMW/LMW of sample 15 was less than 0.1, in samples 10 and 1 more than 2 and in samples 14 and 7, approximately 1. HMW/LMW ≥ 1.0 was indicative of a pyrogenic source and HMW/ LMW < 1.0 suggested a petrogenic source. Most samples from Jänesselja and Maadevahe APPs had similar PAH binary ratios and could therefore be classified into one group. Contamination in the soil samples from both plants originated most likely from shale oil combustion. Samples 4 and 13 could be clearly differentiated from the other samples by PAH binary ratios. Samples 1, 3, 7, 10, 12 and 15 could also be distinguished from the other samples by PAH binary ratios, but this difference was not as great as in case of samples 3 and 14. Based on PCA, several samples were found to be similar in PAH binary ratios. Sample pairs 5 and 6, 9 and 11, and 16 and 17 were established to be similar in composition, based on origin, sampling time and location. However, in the case of sample pairs 3 and 15, 1 and 10, and 14 and 7, the similarity could not be traced, based on the information available. The oil samples that were closely located in the scores plot (central cluster) had a similar chemical composition, based on PAH binary ratios (Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(*g*,*h*,*i*)P, B(*a*)Ant/(B(*a*)Ant + Chr), and HMW/LMW). Conversely, the oil samples that were located far apart in the scores plot differed in chemical composition, and this dissimilarity increased as the distance between the samples increased. Figure 7 shows the scores plot and the loadings plot of the samples. The grouping of the samples was mainly based on the ratio HMW/LMW. In grouping sample 2, mainly Ant/Phe and Ant/(Ant + Phe) were taken into account, while sample 4 was grouped chiefly on the basis of Fla/Pyr and Fla/(Fla + Pyr), and sample 13, on the basis of Ind/(Ind + B(*g*,*h*,*i*)P).

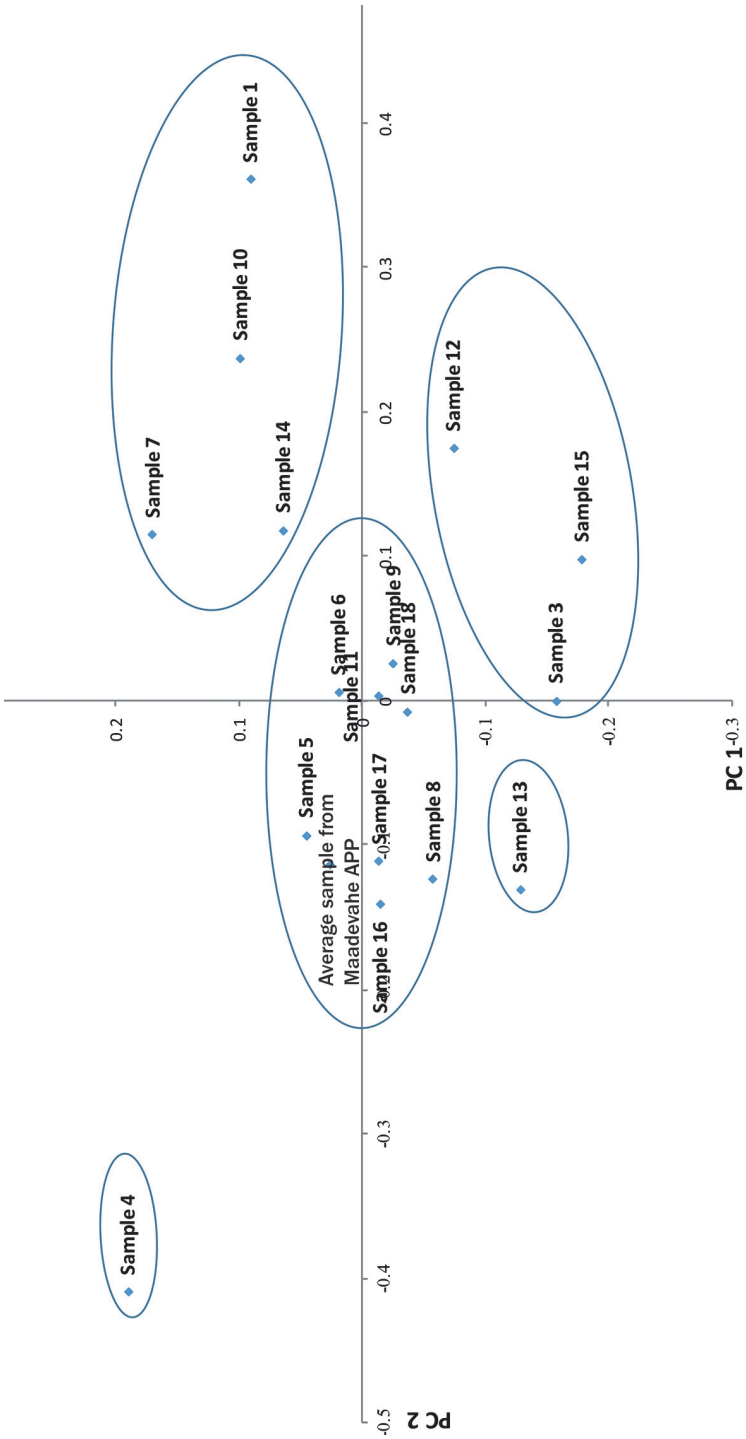

Fig. 6. The grouping of samples from Jänesselja APP on the basis of PAH binary ratios.

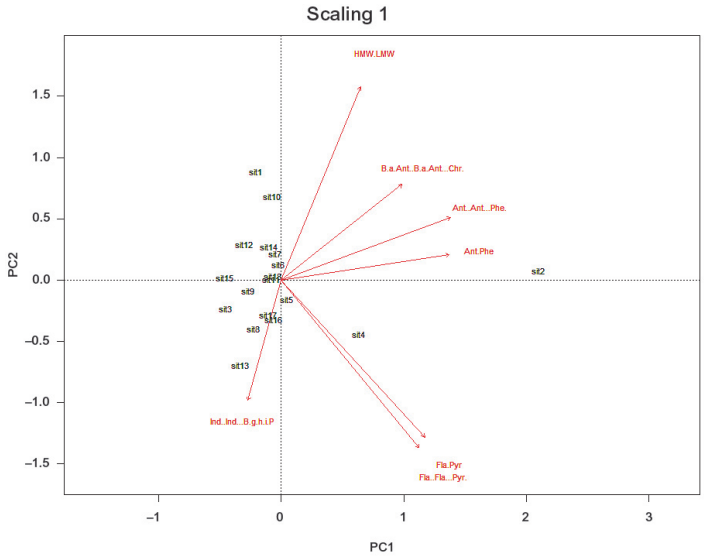

Fig. 7. The scores plot and the loadings plot of samples.

3.3.3. Cluster analysis

Another chemometric technique, cluster analysis, was used for clustering the samples. The data consisted of binary ratios of seven PAH isomers (Ant/Phe, Fla/Pyr, Ant/(Ant + Phe), Fla/(Fla + Pyr), Ind/(Ind + B(*g*,*h*,*i*)P), B(*a*)Ant/(B(*a*)Ant + Chr), and HMW/LMW) contained in 18 samples from Jänesselja APP. Distance was used as a measure of similarity/dissimilarity between the samples. Before the distances were computed, the PAH binary ratios were first normalised by the maximum-minimum range. Three kinds of linkage methods (single, average, complete) were tested in order to find out the most suitable linkage method for determination of these ratios. All the linkage methods merged the samples almost into the same groups, but the average distance between the groups better reflected the relationship between PAH binary ratios and oil types.

Figure 8 depicts the output of the average linkage method using the Jaccard distance as a similarity measure. From the figure it can be seen that the sampling sites fall into two major groups. The dendrogram obtained from HCA revealed that sampling sites with similar PAH concentrations were clustered into the same group. Based on the clusters obtained from CA using chromatogram patterns, as described in Section 3.1, the sampling sites in this study were grouped according to oil mixture type. Cluster 1 included sample 2 (the outlier). Sample 13 in cluster 2 was polluted mostly with LFO and sample 4 in cluster 3 with a mixture of shale oil and diesel. Cluster 4 consisted of samples which were primarily polluted with HFO (probably mazut), while cluster 5 comprised samples containing the mixture of shale oil and LFO.

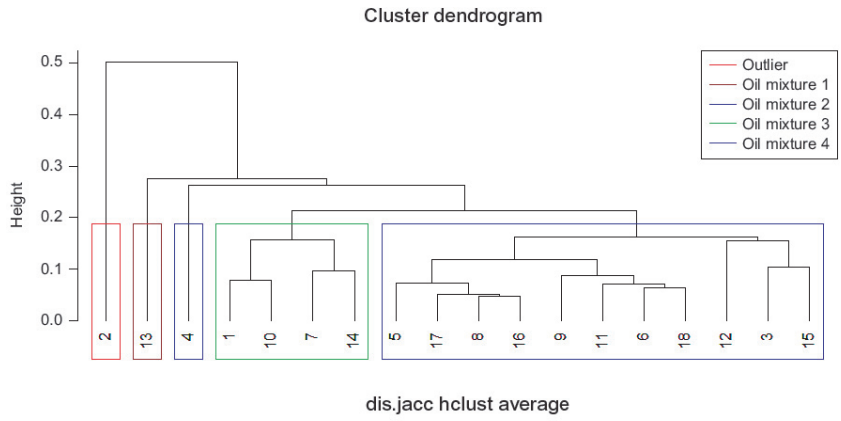

Fig. 8. The dendrogram of the groups of samples from Jänesselja APP.

Some sample pairs, namely 1 and 10, 7 and 14, 8 and 16, and 3 and 15, were located next to each other on the dendrogram (Fig. 8) and in the scores plot (Fig. 7). Oil mixtures 3 and 4 on the dendrogram included the same samples as in the scores plot (Fig. 6). PCA and CA classified the samples into the same groups. The clusters generated were highly reliable as the samples in the groups had similar PAH binary ratios and similar sampling backgrounds.

4. Conclusions

The application of two chemometric techniques, cluster analysis and principal component analysis, and two analytical methods, gas chromatography-mass spectrometry and gas chromatography-flame ionization detection, were used to determine the distribution of polycyclic aromatic hydrocarbons in the soils on the territories of past asphalt pavement plants, as well as to identify possible sources of old oils. The initial oil screening revealed that the soil samples from Jänesselja APP were contaminated with different mixtures of oils, such as diesel, light fuel oil, heavy fuel oil, lubricating oil, waste oil and shale oil. The GC-FID chromatograms of the selected oil samples from Maadevahe APP showed a similar pattern, which suggested that the soil at this plant was polluted with one type of oil, namely shale oil. The composition and physical properties of spilt oils in the soil samples changed during the weathering processes (evaporation, water solubilisation and oxidation), and differed from those of fresh oils.

The binary ratios of 16 PAHs revealed that the contamination in the samples from Maadevahe and Jänesselja APPs originated from petroleum and shale oil combustion. Based on the results of PCA and CA, oils or oil mixtures, with

which the soil samples from Jänesselja APP were mostly polluted, could be divided into four groups: oil mixture 1 (mostly LFO), oil mixture 2 (shale oil and diesel), oil mixture 3 (HFO, waste oils and diesel), oil mixture 4 (shale oil, LFO and diesel). As a result of degradation and migration of PAHs in soil, it was difficult to identify the sources of mixed oils and the relative contribution of each possible source to pollution according to PAH binary ratios only. It can be concluded that using the combination of analytical and chemometric methods in oil spill fingerprinting considerably contributed to interpreting the data, based on PAH binary ratios, while in determining possible sources of old oils, it saved a great amount of time and avoided the high costs of oils classification. PCA and CA can be useful tools to examine various relationships among different samples. These methods are able to identify similar sample groups or sample pairs in the scores plots or on the dendrograms.

Acknowledgements

The authors would like to thank Sander Sannik of Eurofins Environment Testing Estonia OÜ for his help in writing the manuscript, Mati Salu of Maves AS for assisting in sampling at Jänesselja APP, and the Estonian Ministry of Education and Research for financial support (Grant No. IUT 33-20).

REFERENCES

1. Eesti Päevaleht. *In the road construction there occurred a 6680-tonne mazut spill.* Report. http://epl.delfi.ee/news/eesti/tee-ehitusele-sattus-ette-6680-tonnine-masuudireostus?id=51193631. Last visited September 2018 (in Estonian).
2. AS Maves. *The control and monitoring of hazardous residual contamination*. Report. Tallinn, 2004 (in Estonian).
3. Mas, S., de Juan, A., Tauler, R., Olivieri, A. C., Escandar, G. M. Application of chemometric methods to environmental analysis of organic pollutants: A review. *Talanta*, 2010, **80**(3), 1052–1067.
4. Ramirez, M. I., Arevalo, A. P., Sotomayor, S., Bailon-Moscoso, N. Contamination by oil crude extraction – Refinement and their effects on human health. *Environ. Pollut.*, 2017, **231**(1), 415–425.
5. Eesti keskkonnauuringute keskus. *The development of the methodology for reducing residual pollution at former military and industrial objects, Phase I*. Report. 2013 (in Estonian).
6. Bruzzoniti, M. C., Fungi, M., Sarzanini, C. Determination of EPA's priority pollutant polycyclic aromatic hydrocarbons in drinking waters by solid phase extraction-HPLC. *Anal. Methods*, 2010, **2**, 739–745.
7. Jefimova, J., Irha, N., Reinik, J., Kirso, U., Steinnes, E. Leaching of poly-

cyclic aromatic hydrocarbons from oil shale processing waste deposit: A long-term field study. *Sci. Total Environ.*, 2014, **481**, 605–610.
8. Retnam, A., Zakaria, M. P., Juahir, H., Aris, A. Z., Zali, M. A., Kasim, M. F. Chemometric techniques in distribution, characterisation and source apportionment of polycyclic aromatic hydrocarbons (PAHs) in aquaculture sediments in Malaysia. *Mar. Pollut. Bull.*, 2013, **69**(1–2), 55–66.
9. Lau, E. V., Gan, S., Ng, H. K. Extraction Techniques for Polycyclic Aromatic Hydrocarbons in Soils. *Int. J. Anal. Chem.*, **2010**, Article ID 398381, 2010.
10. Pongpiachan, S., Hattayanone, M., Tipmanee, D., Suttinun, O., Khumsup, C., Kittikoon, I., Hirunyatrakul, P. Chemical characterization of polycyclic aromatic hydrocarbons (PAHs) in 2013 Rayong oil spill-affected coastal areas of Thailand. *Environ. Pollut.*, 2018, **233**, 992–1002.
11. Dudhagara, D. R., Rajpara, R. K., Bhatt, J. K., Gosai, H. B., Sachaniya, B. K., Dave, B. P. Distribution, sources and ecological risk assessment of PAHs in historically contaminated surface sediments at Bhavnagar coast, Gujarat, India. *Environ. Pollut.*, 2016, **213**, 338–346.
12. Koch, M., Liebich, A., Win, T., Nehls, I. Certified Reference Materials for the determination of mineral oil hydrocarbons in water, soil and waste. *Forschungsbericht 272*, Berlin, 2005.
13. Konečný, F., Boháček, Z., Müller, P., Kovářová, M., Sedláčková, I. Contamination of soils and groundwater by petroleum hydrocarbons and volatile organic compounds – Case study: ELSLAV BRNO. *Bull. Geosci.*, 2003, **78**(3), 225–239.
14. Wüst, B. *Measuring Hydrocarbon Oil Index according to ISO 9377-2 (DIN H53). Environmental Application. Ga Chromatography*. https://www.agilent.com/cs/library/applications/app_note.pdf. Last visited October 2018.
15. Christensen, J. H., Tomasi, G. Practical aspects of chemometrics for oil spill fingerprinting. *J. Chromatogr. A*, 2007, **1169**(1–2), 1–22.
16. Lubes, G., Goodarzi, M. Analysis of volatile compounds by advanced analytical techniques and multivariate chemometrics. *Chem. Rev.*, 2017, **117**(9), 6399−6422.
17. Hopke, P. K. Chemometrics applied to environmental systems. *Chemom. Intell. Lab. Syst.*, 2015, **149**(B), 205–214.
18. Singh, I., Juneja, P., Kaur, B., Kumar, P. Pharmaceutical applications of chemometric techniques. *ISRN Anal. Chem.*, **2013**, Article ID 795178, 2013.
19. Mali, M., Dell'Anna, M. M., Notarnicola, M., Damiani, L., Mastrorilli, P. Combining chemometric tools for assessing hazard sources and factors acting simultaneously in contaminated areas. Case study: "Mar Piccolo" Taranto (South Italy). *Chemosphere*, 2017, **184**, 784–794.
20. Panchuk, V., Yaroshenko, I., Legin, A., Semenov, V., Kirsanov, D. Application of chemometric methods to XRF-data - A tutorial review. *Anal.*

Chim. Acta, 2018, **1040**, 19–32.
21. Gad, H. A., El-Ahmady, S. H., Abou-Shoer, M. I., Al-Azizi, M. M. Application of chemometrics in authentication of herbal medicines: a review. *Phytochem. Analysis*, 2013, **24**(1), 1–24.
22. Novák, M., Palya, D., Bodai, Z., Nyiri, Z., Magyar, N., Kovács, J., Eke, Z. Combined cluster and discriminant analysis: An efficient chemometric approach in diesel fuel characterization. *Forensic Sci. Int.*, 2017, **270**, 61–69.
23. Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*. Wiley, Chichester, 2003.
24. Miki, S., Uno, S., Ito, K., Koyama, J., Tanaka, H. Distributions of polycyclic aromatic hydrocarbons and alkylated polycyclic aromatic hydrocarbons in Osaka Bay, Japan. *Mar. Pollut. Bull.*, 2014, **85**(2), 558–565.
25. ISO 16703:2011. *Soil quality - Determination of content of hydrocarbon in the range C10 to C40 by gas chromatography*, 2011.
26. Hupp, A. M., Marshall, L. J., Campbell, D. I., Smith, R. W., McGuffin, V. L. Chemometric analysis of diesel fuel for forensic and environmental applications. *Anal. Chim. Acta*, 2008, **606**(2), 159–171.
27. Raschka, S. *About Feature Scaling and Normalization – and the effect of standardization for machine learning algorithms*. https://sebastianraschka.com/Articles/2014_about_feature_scaling.html. Last visited November 2018.
28. Risvik, H. *Principal Component Analysis (PCA) & NIPALS algorithm*, 2007. https://folk.uio.no/henninri/pca_module/pca_nipals.pdf. Last visited November 2018.
29. Oksanen, J. *R package version. Multivariate Analysis of Ecological Communities in R: vegan tutorial*, 2011.
30. Romesburg, H. C. *Cluster Analysis for Researchers*. Lulu Press, 2004.
31. Zuur, A., Ieno, E., Meesters, E. *A Beginner's Guide to R*. Springer Publishing Company, 2009.
32. AS Maves. http://www.maves.ee/Projektid/2003/OJRK_pingerida.htm. Last visited January 2019 (in Estonian).
33. Barakat, A. O., Mostafa, A. R., Qian, Y., Kennicutt II, M. C. Application of petroleum hydrocarbon chemical fingerprinting in oil spill investigations – Gulf of Suez, Egypt. *Spill Sci. Technol. B.*, 2002, **7**(5–6), 229–239.
34. Zemo, D. A., Bruya, J. E., Graf, T. E. The application of petroleum hydrocarbon fingerprint characterization in site investigation and remediation. *Ground Water Monit. R.*, 1995, **15**(2), 147–156.
35. ISO 18287:2006. *Soil quality - Determination of polycyclic aromatic hydrocarbons (PAH) - Gas chromatographic method with mass spectrometric detection (GC-MS)*, 2006.
36. Malmquist, L. M., Olsen, R. R., Hansen, A. B., Andersen, O.,

Christensen, J. H. Assessment of oil weathering by gas chromatography – mass spectrometry, time warping and principal component analysis. *J. Chromatogr. A*, 2007, **1164**(1–2), 262–270.

Presented by A. Siirde
Received April 10, 2019

Appendix 2

Publication II

Koel, M., Jurjeva, J. (2022). Implementing greening into design in analytical chemistry. Talanta Open 6, 100136. https://doi.org/10.1016/j.talo.2022.100136

Talanta Open 6 (2022) 100136

Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open

Implementing greening into design in analytical chemistry

Jelena Jurjeva [a,b], Mihkel Koel [b,*]

[a] Eurofins Environment Testing Estonia OÜ, Paavli 5-3, 10412 Tallinn, Estonia
[b] Department of Chemistry and Biotechnology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

ARTICLE INFO

Keywords:
Quality-by-design
Quality control
Green analytical chemistry
Design of experiments

ABSTRACT

From the pharmaceutical industry is known the approach Quality-by-Design (QbD) and this has great influence on analytical methods and procedures it employs. There is even proposed "analytical quality-by-design" to emphasise the role of analytical chemistry in quality control system of this industry.

However, every branch of chemical industry in principle has the quality system where big part is related to chemical analysis and respective quality control system. By time more green chemistry principles like minimising the generation of waste and reducing the use of materials and energy as well as including analytical measurements in large-scale quality-control processes, are integrated into production of chemicals. The integration of green chemistry principles in analytical chemistry will bring these principles also into QbD of particular industry. In present work the review of examples is given on introduction of green analytical methods into QbD.

Introduction

Green Chemistry has been defined "as the utilisation of techniques and methodologies that reduce or eliminate the use or generation of feedstocks, products, by-products, solvents, reagents, etc. that are hazardous to human health or the environment" [1]. In the green approach of chemistry, the main attention is directed towards the design of a material or chemical process: four of twelve principles [2] are associated with design - Designing Less Hazardous Chemical Synthesis (3); Designing Safer Chemicals and Products (4); Designing for energy efficiency (6) and Designing for Degradation (10). The same is in engineering where the most of principles for greening also are related to design [3].

Research in analytical chemistry is directed towards optimising the analytical process to provide the required information in a manner that is inherently safe, non-toxic and environmentally benign, and with the least possible consumption of materials and energy and generation of waste [4]. Now the question is about the design and implementation of large-scale analytical procedures in industry where green chemistry principles should be considered, especially with regard to quality control. Here the analytical chemistry itself is able provide the tools for determining the greenness of a chemical product or technology. Important advocate of green analytical chemistry Prof. J.Namiesnik proposed the principles of green analytical chemistry (SIGNIFICANCE) [5], and these cover important aspects what should considered also in analytical chemistry for process technology.

The aim of present work is to discuss how the quality by design and green analytical chemistry can have common points and review what is done into area Quality-by-Design on introduction of green chemistry printciples into analytical side of design. Also some new perspectives in this area are proposed.

Quality by design and analytical chemistry

In industry analytical chemistry is mainly regarded as a measurement tool for chemists and chemical engineers for monitoring chemical technological processes and with this being main instrument of quality control of technology and products. Process analytical technology (PAT) includes the process monitoring and control aspects of analytical chemistry, in which it is mainly used for problem solving and for determining the composition of the products. The aim is to develop and keep running technological processes that yield high-quality products. This approach for that is called Quality by Design (QbD). The pharmaceutical industry is at the forefront of the implementation of QbD into the development of technological processes; which in turn has a direct impact on the analytical methods and the process analytical technologies it employs [6]. However, quality is not well-defined term in dictionaries and even Wikipedia give wide scale for interpretation. The

Abbreviations: QbD, Quality-by-design; QC, quality control; GAC, green analytical chemistry; DoE, design of experiments.

* Corresponding author.
E-mail address: mihkel.koel@taltech.ee (M. Koel).

https://doi.org/10.1016/j.talo.2022.100136
Received 15 June 2022; Received in revised form 8 August 2022; Accepted 10 August 2022
Available online 13 August 2022

most important aspect of quality related to products and processes is defined as being suitable for its intended use while satisfying customer expectations. In PAT a clear understanding of the process parameters and material attributes and how they affect the quality of the final product is required. This involves also monitoring of all the manufacturing steps in order to follow the evolution of quality throughout the entire process [7]. In literature the QbD approach combined with Risk Analysis (RA) and Design of Experiments (DoE) is leading to the development of robust chemical processes with proper analytical monitoring and control that ensure product quality [8].

Considering the sustainable production, where green aspects should be integrated in every step, it can be presented also as fishbone diagram (Fig. 1):

This diagram illustrates the necessary inputs which are related to product quality and further to sustainability of process. Here all inputs have more or less strong relation with green chemistry – especially Materials, Energy and Waste and should be considered in process/product development in sense of fit-of-purpose (high quality).

Nowadays, the QbD is becoming the main target in pharmaceutical development and with this the side related to analytical chemistry is intensively studied. According to Web of Science more than 100 papers in a year are published with keywords "Analytical Quality by Design". It seems that the quality-by-design is becoming an essential part of modern approach in process development in pharmacy and chemistry. For that in the first place the assessing of the performance profile of the desired product define the quality profile of the related product, and from that will be identified critical quality attributes for the process and analytical side of the process control.

Analytical targets

The customers' expectations from analytical chemistry is from one side to obtain with sufficient accuracy and precision the chemical information - both qualitative and quantitative - about the object under the study, and from other side to quantity of the relevant elements and molecules in order to solve problems related to the purity of compounds or the safety of products and processes. The data produced by a measurement process have to enable a user of the data to make technically correct decisions for a stated purpose [9].

A huge range of analytical methods, procedures and instruments are available, and those that are best suited to the purpose of the analysis must be identified and used. In this stage of development of the analytical process related to quality-by-design, it is useful to follow Guideline Q8 (R2) on Quality by Design from the International Conference on Harmonization (ICH), which is defined as "a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management" [10]. It stresses that the method of analysis and the performance characteristics of the analytical process have to be chosen to fulfil the specific objectives of the study.

The following the Guideline means a change of paradigm from quality-by-testing of the end product or batches released to the development of an efficient process of production with the complete understanding of the process with the possibility to influence of operational parameters and the same time improving measurement reliability and leaving the room for regulatory flexibility. This definition is in accord with the role of analytical chemistry proposed by the principles of Green Chemistry - perform real-time analysis to prevent pollution. Green chemistry is usually not mentioned in quality-control systems clearly, but dealing with waste and toxicity of chemicals in use leads directly to strongly follow the related green chemistry principles [11].

Limits set by quality-control systems, including definition of the problem, a method of obtaining representative samples, specification of the required level of accuracy and acceptable uncertainty (often based on current legislation), a method of data treatment, and communication of results, they provide a framework for the analytical targets of the technological process in the first place. Within this framework, product quality is ascertained by measuring and analysing the degree of impurities, level of toxicity, amount of waste, etc.

The principles of Green Chemistry, which include limitations on the use of solvents, the toxicity of the chemicals and the generation of waste, additionally to customer specifications and regulatory requirements could be used as the basis and framework on setting analytical targets of the chemical process. In principle, any analytical procedure conforming to the target would be acceptable. However, environmentally friendliness and using of safe and non-toxic materials, also economy of energy use in analytical process should be considered.

In the process of design, it means to perform the first step of optimisation of analytical approach. There are several assessment methods that enable the selection of the most efficient method of analysis in terms of greenness [12–15]. The number of parameters required to profile the greenness of an analytical method can be debated, but it is certain that the availability of this type of profiling must facilitate the selection of a method that performs best in terms of environmental impact.

The first attempt to quantify the green profile of analytical methods was proposed by Keith et al. [13]. They used four criteria - Persistent/Bioaccumulative/Toxic (PBT), Hazardous, Corrosive, and Waste – for chemicals and processes. The criteria on energy used in the process of analysis by instruments and for heating/cooling was added later. This approach is somehow straightforward and can be easily combined into analytical design space for proper selection of analytical method

There is another method - "Eco-Scale", which has become popular on comparison of different methods. This metrics was developed by Prof. Namiesnik et al., where each possible hazard in analytical procedure (amount of chemicals and waste generated, energy used, also physical

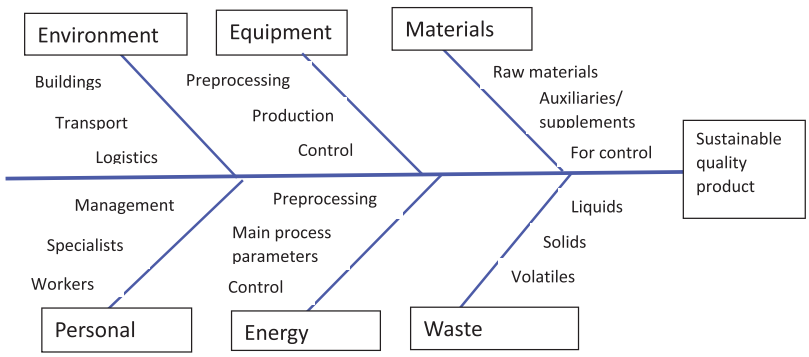

Fig. 1. Fishbone diagram for production.

occupational effect) is assigned with penalty points [12].

The most popular metric used in literature is "AGREE" developed by the group of M.Tobiszewski based on previously mentioned 12 principles of green analytical chemistry [15], what is recently modified considering sample preparation [16]. It is little more complicated system of scores and weights considering each principle, which results final score in 0 – 1 scale for particular analytical method. The result is a pictogram indicating the final score, performance of the analytical procedure in each criterion, and weights assigned by the user. This way input variables that usually have of different nature - binary, discrete, or continuous functions are transformed to give comparable and easy to interpret numerical results. There are the level of analysis automation and miniaturization considered, also the location of analytical devise (laboratory or field) is taken account. Freely available software makes the assessment procedure straightforward. Proposed approach for assessment of analytical methods is well complex and can provide useful information for selecting of analytical methods for process analytical technology.

The process of developing an analytical method in a QbD environment is summarised in the following flow chart (Fig. 2) (adapted from J. N. Sangshetti et al. [17] and from T.Tome et al., [18]) All of the steps in the flow chart, especially those related to the analytical target profile, selection of technique, and risk assessment, include opportunities to apply the principles of Green Chemistry and Green Engineering.

The adoption of QbD principles in analytical measurement and related design of methods starts with the concept of an "Analytical Target Profile (ATP)", which is analogous to the Quality Target Product Profile described in ICH Guideline Q8 [19,20]. This set of criteria defines what will be measured, in which matrix, over what range of concentration, and the required performance of the method, which comprises factors such as accuracy, precision, selectivity, sensitivity, linearity, robustness; also, some more general parameters like analysis time, throughput capacity, and the total cost of analysis. In the list of general parameters could be included restrictions from green approach, like use of environmentally benign solvent and non-toxic chemicals, limits on use of energy, *etc*. The ATP describes the requirements for the analysis process to ensure that the resulting data is suited to its intended purpose. These requirements are established prior to developing the method, and are directly linked to the performance of the method. The QbD approach, in which the principles of Green Chemistry are integrated into analytical method used, supports the development of robust and environmentally friendly process analytical technologies.

A review by T. Tome et al. [18] with references more than 70 studies on the use of QbD in the development of analytical procedures indicates that QbD is becoming to spread in analytical chemistry, accompanying even with term "Analytical QbD". An important consideration in the development of process analytical systems is that QbD enables greater regulatory flexibility across the system lifecycle, because of wider performance criteria related to the method, not strict instrument settings [21].

There are not many publications on implementing the principles of Green Chemistry in QbD. However, those that exist, show that applying the concepts of Green Chemistry and QbD to analytical procedures results some environmentally benignity together with the robustness of the particular HPLC procedure [22].

In the literature on analytical QbD the main trend of implementing green element in design involves replacing the solvent - commonly used in HPLC acetonitrile with ethanol [23–25], and similarly in UHPLC [26, 27]. Ethanol is chosen due to its lower toxicity and availability from renewable sources. However, the use of ethanol gives rise to different critical factors in the separation process and brings up new method parameters need to optimised during the design of the experiment and

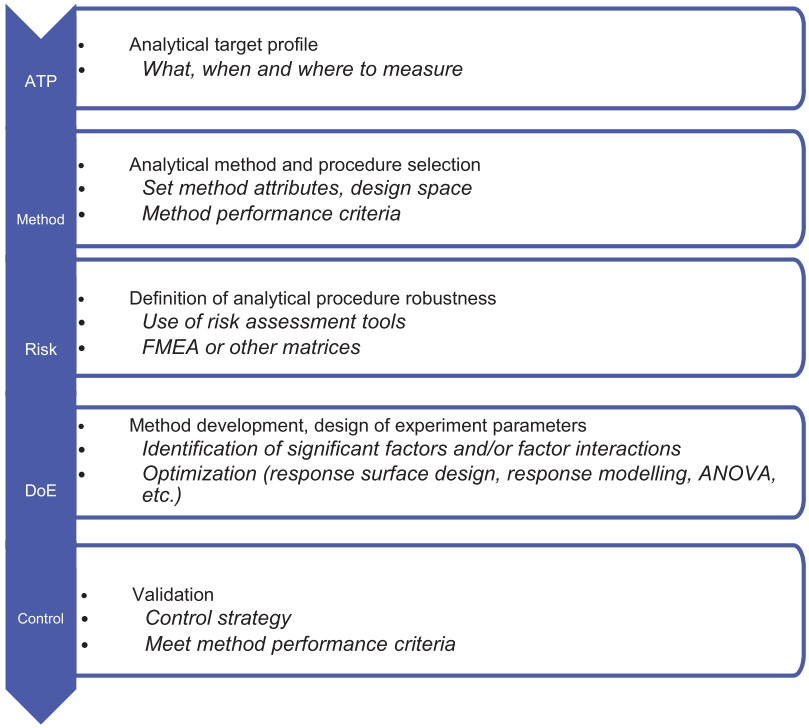

Fig. 2. A flow chart of the process of analytical method development in a QbD environment.

optimisation of the analytical procedure.

The three main aspects of an analytical procedure to consider in terms of greenness are: sample pre-treatment, efficiency of analytical devices, and waste reduction together with the use of a proper waste treatment methodology. Selecting the proper solvent is usually the first choice in greening an analytical procedure. For this kind of selection there is existing proper selection guide [28] what is well valid for analysis methods also. Similar selection guide is developed for acids and bases [29] and this is valuable in assessing the supplement compounds used in analysis methods. These guide systems are based on environmental, safety and health issues and problems with the occurrence of hazardous by-products or disposal problems. The similar approach was used by Y.Gaber [30] in development of the assessment tool for liquid chromatography.

Analytical design space and design of experiments

The establishment of the frame for developing of an analytical method for certain chemical process is defined in the ICH Q8 document as the analytical Design Space (DS): "the multidimensional combination and interaction of input variables (e.g., material attributes) and process parameters that have been demonstrated to provide assurance of quality'' [15].

The most important determinant of the quality of an analytical procedure is measurement uncertainty, as this ascertains whether the analytical procedure will provide the data with necessary quality. The DS is establishing certain framework inside of which ensure a proper analytical procedure which provides reliable results to be a base for making correct decisions. There is a range of possibilities in the selection of a procedure and its development for particular types of analysis: for example, in chromatographic separation, the variability due to HPLC instrument configuration, column selection, nature of mobile phase, injection volumes, and so forth is controlled by the selection of method; whereas the other variables (pH, column temperature, flow rate, and percentage ratio of the mobile phase) can be assigned through robustness testing using the Design-of-Experiment (DoE) approach [31]. As the analytical procedure is part of process design in PAT, the design space gives a zone of robustness in which drastic changes of the required performance criteria of the method should not occur.

The analytical method can be presented in a flow chart that highlights the main steps of the procedure from sample preparation and instrumental settings to final data analysis in order to determine the critical attributes related to product quality as well as the analytical method parameters that should be examined during the method development. This is the first level of risk analysis, what is considered important step in QbD. Here the Ishikawa diagram [32] can facilitate the selection of variables, such as instrumentation, materials, methods, chemicals and reagents, measurements, human factors, and environmental issues (e.g., laboratory temperature, relative humidity, and light), that can have an impact on a particular analytical target (Fig. 3), also have a potential effect on the performance of the method and the total quality of the analytical result.

Most of the conditions and parameters are usually fixed by preliminary experiments and/or prior knowledge, but some are typically evaluated by a DoE during method development. In this step is important to consider that all activities build the total environmental impact of the analytical method and materials inputs become a waste.

With regard to the inclusion of green measures into the analytical process, it is the designer/developer decision in order to achieve the required performance of the method to make changes on the list of factors. Therefore, a number of carefully selected factors (e.g., gradient time in chromatography, concentration of reagents in sample preparation, etc.) that affect the analytical method must be considered simultaneously during the development of a DS [34]. It is advantageous to employ mathematical methods for planning analytical procedures and integrating communications technology into measurement processes and data handling – chemometrics - in order to ensure that the analysis is safe and less resource-intensive in terms of chemicals and solvents [35, 36].

There are some examples where green analytical chemistry metrices are used to find out greener analytical procedures. The Analytical Eco-Scale metric tool [12] was popular method to assess resulting method with including the greener features into the method: A. A. Habib *et al.* developed a fast micellar HPLC method in which greenness was increased by using small amounts of an organic solvent containing a surfactant) above its critical micelle concentration [37]. Since several constraints related to analytical procedures have been identified due to the influence of the solvent, studies have been undertaken to develop new HPLC methods, such as enantiomeric purity testing that incorporates green features [38]. This Analytical Eco-Scale tool was applied to confirm that methods developed were far superior to the official one in terms of greenness. The Eco-Scale score was also used for optimisation of analytical procedure based liquid chromatography with screening solvents and selecting reagents and modifying instrumentation in order to achieve at a greener procedure [39].

Sample preparation is a vital and influential part of an analytical procedure, also from a greenness perspective, and there are studies in which advantageous microextraction methods are combined with method optimisation. In a study by Sumithra *et al.* the sample solution is prepared using dispersive liquid–liquid extraction (DLLME). Favourable green aspects of DLLME include the use of microlitre amounts of organic solvent, production of low levels of waste, and minimisation of effluent

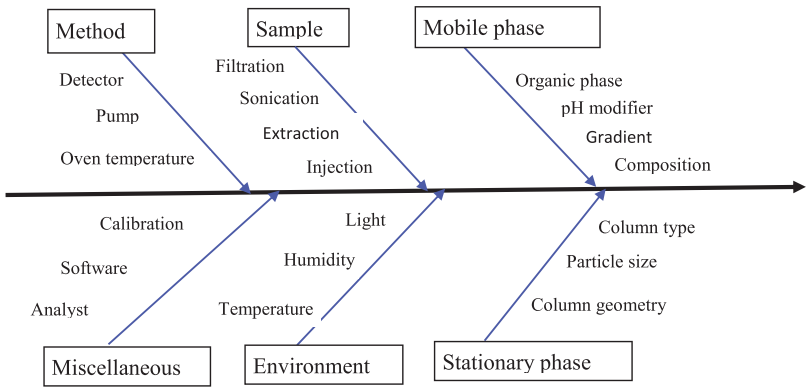

Fig. 3. Ishikawa diagram for reviewing the effect of possible factors for the LC method based on preliminary trials and previous knowledge [adapted from 33].

treatment [40].

An important objective of QbD is to achieve regulatory flexibility in terms of the level of robustness of analytical procedures for production control [41]. Simpler and more robust methods with sufficient specificity and accuracy are usually considered greener than others because they generate less or no waste and consume minimal if any energy. Good examples are paper-based assays for quantitative colorimetric determination of phytochemicals in plant extracts [42] and smartphone-based spectrometry [43,44]. Less research has been devoted to this topic and no examples of integrating this kind of methods into QbD. However, it has significant potential – not for replacing high-performance laboratory instruments but for obtaining high-quality analytical information onsite or at the point-of-care where simple and robust methods often yield enough information to make the decision and solve the problems [45].

Despite of number of fixed parameters of selected analytical method there are parameters where for selection some experimental work is needed before final decision in design. Big part in the literature on analysis process development is dedicated to Design of Experiment (DoE). A common difficulty at the DoE stage is the definition of ranges employed for the process parameters included in the experimental matrix. These ranges must suit the purpose of the experiment, whether screening, optimization or assessment of robustness, and fulfil the objectives of the study. One of the keys to the success of DoE is to perform a convincing but modest number of preliminary experiments that can inform the design and produce meaningful results. Using different procedures for method optimization (like Box-Behnken experimental plan [46]) and designing the experiment for quantitative robustness (like Plackett-Burman[47]) will result in a method that is best suited to its intended purpose.

One example of using DoE in analytical QbD is presented by Santos Moreira et al. for optimization of HPLC chromatography conditions [48]. A Placket-Burman experimental design was adopted to assess the most relevant analytical conditions (type of buffer used in mobile phase, buffer concentration, buffer pH, organic solvent in mobile phase, additive concentration in mobile phase, column temperature, mobile phase flow rate, injection volume of test solutions, time of mobile phase first gradient elution, time of mobile phase second gradient elution, proportion of organic solvent in mobile phase during gradient elution, time of isocratic elution) that affect the chromatographic responses (peak asymmetry, number of theoretical plates, peak purity, capacity factor for the first eluted peak, resolution between different impurities, peak asymmetry for impurity, signal-to-noise for impurity, and retention time of impurity peak). The 20 experiments were conducted. The optimized analytical conditions were obtained and the method operable design space (DS) was defined.

Another very thorough example where statistical evaluation of analysis parameters helped to finalize chromatographic conditions in the method development for drug purity and results a robust method with proven method operable region in a short time is presented by Pawar *et al.* [49]. There the analysis of variance (ANOVA) and response surface graphs from DoE were used for evaluation of the critical attributes and robustness of the procedure.

Risk assessment

The next crucial step in QbD approach is the risk assessment what is especially important in drug production and related to safety of patients. The accepted definition of risk is the combination of the probability of the occurrence of harm and the severity of that harm. When in Green Chemistry main attention is given to the hazards from chemicals and related risks, where risk is defined as the likelihood that the toxic properties of a substance will be produced and persons can fall under their actual conditions of exposure [50], then in the case of QbD, risk is related to the quality of the product – usually related to purity of and the functioning of the production process as controlled and assessed by analytical chemistry procedures. The Quality Risk Management (QRM) process is described in detail in ICH Guideline Q9 [11] and is comprised of assessment, control and review of risk. The risk assessment identifies the critical method variables - the parameters that have an impact on the performance of the analytical method - the risks associated with variability such as sample preparation, instrument configuration, and environmental conditions, as well as factors that will affect the environmental impact of the analytical procedure. From the other side it can be said that with this will be defined an analytical procedure robustness for particular method. The risk assessment serves as a guide to understanding the linkage between of the analytical parameters and materials used in analytical measurement to quality of final product. In principle for analytical methods used in process control, there is very difficult to draw solid line between risk assessment and developing the design space (DS). The variables are assessed quantitatively for analytical measurement and qualitatively from the influence on measurement quality. It must also provide a high degree of confidence that the analytical method will provide results that meet the specified requirements under the conditions of the process, and as the method progresses through its lifecycle. LCA seems to be the most comprehensive tool to assess the application of analytical reagents or analytical procedures, as well to get a holistic view on the environmental impact of the process [20].

There are several methods of risk assessment and risk factor ranking, such as Failure Mode Effect Analysis (FMEA) and the Prioritisation Matrix (PM), that may be employed throughout various stages of the development of an analytical method. There are robust methods that are insensitive to small variations in some parameters, as well as delicate or rugged methods that can't deal with variations or noise likely to be encountered during use. The high noise risk factors (e.g., analysts, equipment, days of analysis, batches of reagent or materials, and sample batches) must be evaluated experimentally to determine acceptable ranges and to study their impact on the repeatability and reproducibility of the method, including the sample preparation step and data analysis [51]. Factors that directly affect the quality and safety of the product must first be identified, and their possible effect on method development studied. For example, impurities affecting the quality of the product are directly related to compound impurity control and to the development of the method of analysis. Again, here important is the feedback in development of the analytical method what is closely connected with the design of the experiment, risk assessment and its procedures, and must ensure that the methods are "fit-for-purpose" [52].

Further developments

Introducing green limits into analytical target profile in the QbD approach is related to risk assessment, which facilitates the understanding of variables affecting the performance and the fit-for-purpose of the analytical procedures. With this also the feedback for design of analytical method is obtained. New and green conditions to target profile lead to careful design of experiment to optimise analytical processes not only of process parameters, but also easily can be added the consumption of materials and energy, generation of waste. This will lead to inherent safety, and environmental benignity of methods in order to provide the required analytical information with the necessary sensitivity, accuracy and precision.

Chemometric and statistical tools as well as other mathematical methods play a crucial role in ensuring the quality of the final product, especially in statistical design of experiments, also in multivariate statistics for data analysis. Data processing is not one of the principles of Green Chemistry, but because it prevents human error and saves time, it can be considered as part of an environmentally benign laboratory system. Here should be noted that often by using mathematical procedures instead of chemical processing it is possible to extract useful information from raw measurement data [53,54] with no aid of chemicals and respectively chemical waste.

PAT is now being redefined by miniaturisation, automatic sampling systems, high-throughput measurement processes, and increased

sophistication in data modelling and processing, in which critical process parameters are determined onsite and in real time. This control can be included into the analytical process and assessed during the design. Once these parameters are identified and monitored, process models can be changed accordingly; the presence of actuators and regulators substantially improves control over the process.

At present, analytical chemists everywhere are under pressure to increase throughput and shorten response times to produce data for decision making. Automation in particular is closely related to the elimination of defects and human error in laboratory processes, which is an important part of quality assurance.

Combining automation, sensors and networking onto one platform reflects the definition of an Internet of Things (IoT): "a dynamic global network infrastructure with self-configuring capabilities based on standard and interoperable communication protocols where physical and virtual things have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into the information network" [55]. Development of these types of smart systems of collecting, communicating information and controlling for production processes certainly can be included into QbD.

Conclusions

Quality by Design (QbD) is a systematic approach to method development that utilises predefined objectives and emphasises understanding and control of chemical processes and products. As seen, there are already papers where QbD principles are complemented with green chemistry principles allowing the development of environmentally friendlier analytical procedures where analytical methods are optimised within determined measurement uncertainty with specified robustness parameters and this way satisfying customer expectations on quality. In quality design stage the formation of analytical target profile it is very useful of analysing analytical methods available according to metrics of green analytical methods. This enables the selection of analytical procedure what is safer and reducing its environmental impact. Greening analytical chemistry methods does not only mean reducing the use of solvents and harmful chemicals resulting decrease of amount of waste with regard to a particular chemical measurement procedure, but also using proper instruments with necessary performance that is suited for the purpose. The economy obtained from the selection of instruments in a given situation not only conserves materials and energy but also reduces health and safety hazards for the analyst and for the environment.

Contemporary Design-of-Experiment methods provide limited number of additional experiments used to determine acceptable ranges of method variables and their impact on the repeatability and reproducibility of the measurement procedure (including sample preparation and data analysis) is remarkable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] P.T. Anastas, Green chemistry and the role of analytical methodology development, Crit. Rev. Anal. Chem. 29 (3) (1999) 167–175, https://doi.org/10.1080/10408349891199356.

[2] P.T. Anastas, J.C. Warner, Green Chemistry: Theory and Practice, Oxford University Press, New York, 1998.

[3] P.T. Anastas, J.B. Zimmerman, Design through the 12 principles of green engineering, Env. Sci. Tech. 37 (2003) 94A–101A, https://doi.org/10.1021/es032373g.

[4] M. Koel, Do we need Green analytical chemistry? Green Chem 18 (2016) 923–931, https://doi.org/10.1039/c5gc02156a.

[5] A. Gałuszka, Zd. Migaszewski, J. Namiesnik, The 12 principles of green analytical chemistry and the SIGNIFICANCE mnemonic of green analytical practices, TrAC - Trends Anal. Chem. 50 (2013) 78–84. doi.org/10.1016/j.trac.2013.04.010.

[6] QbD Considerations for Analytical Methods – FDA Perspective – IFPAC annual meeting 2013. http://qbdworks.com/wp-content/uploads/2014/06/AnalyticalQbDIFPAC.pdf.

[7] J. Henriques, et al., On demand for new process analytical technologies applied to injectable drug products, Eur. J.Pharma. Sci. 137 (2019), 105023. doi.org/10.1016/j.ejps.2019.104975.

[8] A.A. Zlota, Fit-for-purpose Quality by Design (QbD) implementation, part 1: risk analysis and statistical Design of Experiments (DoE), Chimica Oggi - Chem. Today 38 (2020) 36–38.

[9] M. Thompson, M.H. Ramsey, Quality concepts and practices applied to sampling—an exploratory study, Analyst 120 (1995) 261–270, https://doi.org/10.1039/AN9952000261.

[10] International Conference on Harmonization (ICH), Tripartite guidelines,' ICH Q8 (R2): pharmaceutical development', 2009, London.

[11] S. Tang, R. Bourne, R. Smith, M. Poliakoff, The 24 principles of Green Engineering and Green Chemistry: "Improvements productively, Green Chem 10 (2008) 268–269, https://doi.org/10.1039/B719469M.

[12] A. Gałuszka, Z.M. Migaszewski, P. Konieczka, J. Namieśnik, Analytical Eco-Scale for assessing the greenness of analytical procedures, TrAC–Trend. Anal. Chem. 37 (2012) 61–72. doi.org/10.1016/j.trac.2012.03.013.

[13] L.H. Keith, L.U. Gron, J.L. Young, Green analytical methodologies, Chem. Rev. 107 (2007) 2695–2708, https://doi.org/10.1021/cr068359e.

[14] J. Płotka-Wasylka, A new t ool for the evaluation of the analytical procedure: green analytical procedure index, Talanta 181 (2018) 204–209, https://doi.org/10.1016/j.talanta.2018.01.013.

[15] F. Pena-Pereira, W. Wojnowski, M. Tobiszewski, AGREE—analytical GREEnness metric approach and software, Anal. Chem. 92 (2020) 10076–10082, https://doi.org/10.1021/acs.analchem.0c01887.

[16] W. Wojnowski, M. Tobiszewski, F. Pena-Pereira, E. Psillakis, AGREEprep - Analytical greenness metric for sample preparation, TrAC - Trends Anal. Chem. 149 (2022), 116553, https://doi.org/10.1016/j.trac.2022.116553.

[17] J.N. Sangshetti, M. Deshpande, et al., Quality by design approach: regulatory need, Arabian J. Chem. 10 (2017) S3412–S3425, https://doi.org/10.1016/j.arabjc.2014.01.025.

[18] T. Tome, N. Žigart, Z. Časar, A. Obreza, Development and optimization of liquid chromatography analytical methods by using AQbD principles: overview and recent advances, Org. Process Res. Dev. 23 (2019) 1784–1802, https://doi.org/10.1021/acs.oprd.9b00238.

[19] Analytical target profile: structure and Application throughout the analytical lifecycle, Pharmacopeial. Forum 42 (5) (2016).

[20] M.K. Parr, A.H. Schmidt, Review: life cycle management of analytical methods, J. Pharm. Biomed. Anal. 147 (2018) 506–517, https://doi.org/10.1016/j.jpba.2017.06.020.

[21] P. Jackson, P. Borman, et al., Using the analytical target profile to drive the analytical method lifecycle, Anal. Chem. 91 (4) (2019) 2577–2585, https://doi.org/10.1021/acs.analchem.8b04596.

[22] S. Saroj, et al., Green analytical chemistry and quality by design: a combined approach towards robust and sustainable modern analysis, Curr. Anal. Chem. 14 (2018) 367–381, https://doi.org/10.2174/1573411013666170615140836.

[23] E.M. Abdel-Moety, M.R. Rezk, M. Wadie, M.A. Tantawy, A combined approach of green chemistry and Quality-by-Design for sustainable and robust analysis of two newly introduced pharmaceutical formulations treating benign prostate hyperplasia, Microchem. J. 160 (2021), 105711, https://doi.org/10.1016/j.microc.2020.105711.

[24] M. Yabre, et al., Development of a green HPLC method for the analysis of artesunate and amodiaquine impurities using Quality by Design, J. Pharm. Biomed. Anal. 190 (2020), 113507, https://doi.org/10.1016/j.jpba.2020.113507.

[25] L. Ferey, et al., UHPLC method for multiproduct pharmaceutical analysis by Quality-by-Design, J. Pharm. Biomed. Anal. 148 (2018) 361–368, https://doi.org/10.1016/j.jpba.2017.10.014.

[26] J. Saxena, et al., Development and validation of an environmentally benign and robust stability indicating assay method for lenalidomide: comprehensive degradation kinetics study and application of synergistic approach involving green analytical chemistry and quality by design methodology, Indian J. Pharm. Edu.Res. 53 (1) (2019) 133–143, https://doi.org/10.5530/ijper.53.1.18.

[27] T. Sharma, et al., Implementation of analytical quality-by-design and green analytical chemistry approaches for the development of robust and ecofriendly UHPLC analytical method for quantification of chrysin, Separation Science Plus 3 (9) (2020) 384–398, https://doi.org/10.1002/sscp.202000028.

[28] R.K. Henderson, C. Jimenez-Gonzalez, D.J. Constable, et al., Expanding GSK's solvent selection guide–embedding sustainability into solvent selection starting at medicinal chemistry, Green Chem. 13 (2011) 854–862, https://doi.org/10.1039/C0GC00918K.

[29] R.K. Henderson, A.P. Hill, A.M. Redman, H.F. Sneddon, Development of GSK's acid and base selection guides, Green Chem 17 (2015) 945–949, https://doi.org/10.1039/C4GC01481B.

[30] Y. Gaber, U. Törnvall, M.A. Kumar, M.Ali Amine, R. Hatti-Kaula, HPLC-EAT (Environmental Assessment Tool): a tool for profiling safety, health and environmental impacts of liquid chromatography methods, Green Chem 13 (2011) 2021–2025, https://doi.org/10.1039/C0GC00667J.

[31] A. Bouabidi, et al., Usefulness of capability indices in the framework of analytical methods validation, Anal. Chim. Acta 714 (2012) 47–56, https://doi.org/10.1016/j.aca.2011.11.019.

[32] What is Total Quality Control? in: K. Ishikawa (Ed.), The Japanese Way Prentice-Hall, Englewood Cliffs, NJ, USA, 1985, pp. 63–64.

[33] X.-T. Sun, Q.-J. Tan, S.-X. Wang, J.-F. Shan, J.-L. Jiang, Systematic, computer-assisted development of high performance liquid chromatography for multicomponent analysis, Anal. Methods 7 (2015) 5428–5435, https://doi.org/10.1039/C5AY01000D.
[34] E. Rozet, P. Lebrun, B. Debrus, B. Boulanger, P. Hubert, Design Spaces for analytical methods, TrAC, Trend. Anal. Chem. 42 (2013) 157–167, https://doi.org/10.1016/j.trac.2012.09.007.
[35] M. Bystrzanowska, M. Tobiszewski, Chemometrics for selection, prediction, and classification of sustainable solutions for green chemistry—a review, Symmetry (Basel) 12 (2020) 2055, https://doi.org/10.3390/sym12122055.
[36] M. Tobiszewski, Metrics for green analytical chemistry, Anal. Methods 8 (2016) 2993–2999, https://doi.org/10.1039/C6AY00478D.
[37] A.A. Habib, S.F. Hammad, S.M. Megahed, A.H. Kamal, Innovative quality by design approach for development of green micellar HPLC method for simultaneous determination of atorvastatin and amlodipine, Chromatographia 83 (10) (2020) 1221–1231, https://doi.org/10.1007/s10337-020-03937-5.
[38] M. Mitrovic, et al., Analytical quality by design development of an ecologically acceptable enantioselective HPLC method for timolol maleate enantiomeric purity testing on ovomucoid chiral stationary phase, J. Pharm. Biomed. Anal. 180 (2020) 13034, https://doi.org/10.1016/j.jpba.2019.113034.
[39] S.S. Panda, S. Dutta, R.K.V. Varaha Bera, S. Jammula, Analytical eco-scale and quality by design-oriented liquid chromatography method for simultaneous quantification of metoprolol succinate, telmisartan, and cilnidipine in their fixed-dose combination, Separat. Sci. Plus: 4 (2021) 128–136, https://doi.org/10.1002/sscp.202000115.
[40] M. Sumithra, V. Ravichandiran, Green approaches in sample preparation for the estimation of atenolol, aspirin, atorvastatin calcium and losartan potassium in bulk and its pharmaceutical dosage form prior to chromatographic technique, Eurasian J. Anal. Chem. 13 (5) (2018) em62, https://doi.org/10.29333/ejac/97223.
[41] S.S. Panda, R.K.V. Varaha Bera, S.S. Pradhan, Development and validation of a quality by design enabled robust LC method for estimation of daptomycin in pharmaceutical dosage form, Separat. Sci. Plus 2 (2019) 397–404, https://doi.org/10.1002/sscp.201900062.
[42] M. Vaher, et al., Automatic spot preparation and image processing of paper microzone-based assays for analysis of bioactive com pounds in plant extracts, Food Chem 143 (2014) 465–471, https://doi.org/10.1016/j.foodchem.2013.08.007.
[43] J.I. Hong, B.Y. Chang, Development of the smartphone-based colorimetry for multi-analyte sensing arrays, Lab. Chip 14 (2014) 1725–1732, https://doi.org/10.1039/C3LC51451J.
[44] A.R. Campos, C.M. Knutson, T.R. Knutson, A.R. Mozzetti, C.L. Haynes, R.L. Penn, Quantifying gold nanoparticle concentration in a dietary supplement using smartphone colorimetry and google applications, J. Chem. Educ. 93 (2016) 318–321, https://doi.org/10.1021/acs.jchemed.5b00385.
[45] A.W. Martinez, S.T. Phillips, G.M. Whitesides, Three-dimensional microfluidic devices fabricated in layered paper and tape, PNAS 105 (50) (2008) 19606–19611, https://doi.org/10.1073/pnas.0810903105.
[46] G. Box, D. Behnken, Some new three level designs for the study of quantitative variables, Technometrics 2 (1960) 455–475, https://doi.org/10.2307/1266454.
[47] R.L. Plackett, J.P. Burman, The Design of Optimum Multifactorial Experiments, Biometrika 33 (4) (1946) 305–325, https://doi.org/10.1093/biomet/33.4.305.
[48] C. dos Santos Moreira, F. R.Lourenço, Development and optimization of a stability-indicating chromatographic method for verapamil hydrochloride and its impurities in tablets using an analytical quality by design (AQbD) approach, Microchem. J. 154 (2020), 104610, https://doi.org/10.1016/j.microc.2020.104610.
[49] A. Pawar, N. Pandita, Statistically designed, targeted profile uplc method development for assay and purity of haloperidol in haloperidol drug substance and haloperidol 1 mg tablets, Chromatographia 83 (2020) 725–737, https://doi.org/10.1007/s10337-020-03889-w.
[50] CJ van Leeuwen, JLM Hermens (Eds.), Risk Assessment of Chemicals: An Introduction, 1st ed., Springer, 2012.
[51] E. Rozet, P. Lebrun, B. Debrus, B. Boulanger, P. Hubert, Design spaces for analytical methods, TrAC - Trend. Anal. Chem. 42 (2013) 157–167, https://doi.org/10.1016/j.trac.2012.09.007.
[52] Validation of analytical methods –to be fit for the purpose, in: Bernd W. Wenclawiak, M. Koch, E. Hadjicostas (Eds.), Quality Assurance in Analytical Chemistry: Training and Teaching, Springer, 2010, pp. 215–245.
[53] M.B. Anzardi, J.A. Arancibia, A.C. Olivieri, Processing multi-way chromatographic data for analytical calibration, classification and discrimination: a successful marriage between separation science and chemometrics, TRAC-Trend. Anal. Chem. 134 (2021) 11612, https://doi.org/10.1016/j.trac.2020.116128.
[54.] J. Chapman, et al., Combining chemometrics and sensors: toward new applications in monitoring and environmental analysis, Chem. Rev. 120 (13) (2020) 6048–6069, https://doi.org/10.1021/acs.chemrev.9b00616.
[55] 38 CERP-IoT, Vision and challenges for realising the Internet of Things, in: H. Sundmaeker, P. Guillemin, P. Friess, S. Woelfflé (Eds.), Vision and challenges for realising the Internet of Things, European Commission: Brussels (2010).

Appendix 3

Publication III

Jurjeva, J., Koel, M. (2025). Application of the concept of design space in method development for the determination of contaminants in soil. Proceedings of the Estonian Academy of Sciences, 74(1), 71–81. https://doi.org/10.3176/proc.2025.1.07

Proceedings of the Estonian Academy of Sciences
2025, **74**, 1, 71–81

https://doi.org/10.3176/proc.2025.1.07

www.eap.ee/proceedings
Estonian Academy Publishers

CHEMICAL ANALYSIS IN CONTAMINATED SOIL

RESEARCH ARTICLE

Received 9 September 2024
Accepted 2 December 2024
Available online 17 February 2025

Keywords:
environmental chemistry, method validation, gas chromatography, experimental design, analytical quality by design, design space

Corresponding authors:
Jelena Jurjeva
jurjeva.jelena@gmail.com
Mihkel Koel
mihkel.koel@taltech.ee

Citation:
Jurjeva, J. and Koel, M. 2025. Application of the concept of design space in method development for the determination of contaminants in soil. *Proceedings of the Estonian Academy of Sciences*, **74**(1), 71–81. https://doi.org/10.3176/proc.2025.1.07

Application of the concept of design space in method development for the determination of contaminants in soil

Jelena Jurjeva[a] and Mihkel Koel[b]

[a] The J. Molner Company OÜ, Akadeemia tee 21–5, 12618 Tallinn, Estonia

[b] Department of Energy Technology, Tallinn University of Technology, Akadeemia tee 15, 12618 Tallinn, Estonia

ABSTRACT

The analysis of polycyclic aromatic hydrocarbons and polychlorinated biphenyls is an important topic in environmental analysis, and several analysis procedures have already been proposed. Following the approach of analytical quality by design (AQbD) contributes to the development of simultaneous identification and quantification of polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs), alkylated PAHs and aliphatic compounds in soil by the gas chromatography-mass spectrometry (GC-MS) analytical technique, using the mechanical extraction method for sample preparation, which results in the modified method with superior analytical parameters. In the study, the set-up of design space (DS) for method development was introduced, including the selection of an appropriate gas chromatography (GC) column. Subsequently, design of experiment enabled finding the optimal GC conditions for the new method that allowed achieving adequate resolution and recovery. Method development involved changes in sample preparation, making it simpler and faster. Decreasing the number of different extraction steps enabled reducing staff numbers and consumable costs. As a result, methods of analysis of PAHs, PCBs and aliphatic and aromatic compounds in a single run were developed, offering reliable accuracy, precision, linearity and recovery rates that fully comply with Swedish regulations. This new method and its introduction to laboratory practice was validated, and the greenness of the method was evaluated.

1. Introduction

Analytical quality by design (AQbD) – the term mostly used in the pharmaceutical industry – aims to integrate analysis and chemical state control development, resulting in more reliable analytical procedures that need less effort for performance verification and post-approval changes because of a better-defined design space (DS) – the region where all the input parameters can vary without altering the quality of the product [1].

According to the International Conference on Harmonisation (ICH) of Technical Requirements for Registration of Pharmaceuticals for Human Use [2], the setting of this space starts from the specifications of a product or the quality target product profile (QTPP), which forms the basis for the design and development of the product. It summarizes all the terms of processes, product properties, product quality criteria as well as safety issues both for the production process and the product itself through the use of which the product can have the desired quality. The next step is directly related to the identification of critical quality attributes (CQAs) of the product, defining the appropriate limits for the product properties or distribution to ensure the desired product quality [2]. Here the direct relationship with quality risk management comes into play to identify which material attributes and process parameters impact CQAs [2,3]. This approach was originally designed for the production process, but similar principles can be applied to the environmental analysis of soil, water or air, where the quality of analysis is an important characteristic describing the state of the environment and controlling the quality of remediation.

DS is important in relation to quality management because working within DS is not considered a change. DS allows to react when the ‘process drifts’ are observed that could be caused, for example, by changes in analysed materials, such as different types of soil or water that cannot be actively controlled in order to keep the quality of environmental parameters at the intended level. Such changes within an approved

DS do not require the authorities' notification or preapproval [4], while the movement out of DS does and should normally initiate a post-approval regulatory change process [5].

1.1. Design space in the development of analytical procedures

There are different approaches and methods for defining DS and its relationship with the product – in the present case, the environmental parameters [6]. There are no official guidelines for establishing these kinds of parameters, and different methods can be used for this purpose, for example, literature search and relevant statistics [7].

Developers of analytical measurements are often asked the following questions: what methods are available for the measurements? What are the characteristics of the samples? What analytical performance criteria are requested by regulatory agencies? What is the method-provided performance? What laboratory resources are available? These questions and the related answers enable the developers to define the exact concept of the method and the limits for analytical procedures, which in principle form the DS for a particular analytical process.

The environmental quality and level of monitoring define and constitute the basis for the control and limits of the environmental parameters in DS, and the procedure development and control are aligned with a similar DS for the analytical parameters. The analytical target profile (ATP) is similar to the environmental parameters profile, which prospectively summarizes the requirements associated with the quality attributes of the measurements that need to be met by an analytical procedure. The general criteria defined in ATP relate to the maximum uncertainty associated with the reportable result that gives an acceptable confidence in the quality of the decision. Other important performance characteristics for each ATP are specificity, accuracy and precision, taking into account the expected range of the analyte [8]. One must keep in mind that ATP does not refer to a specific analytical technique or an operative mode; rather it establishes the criteria required for the technique as well as the measurement parameters. During the method search phase, prior knowledge is gathered, and, if necessary, preliminary experiments are carried out. In separation procedures, which are common in environmental analysis, stationary phases and solvents in chromatography, operative mode/pseudo-stationary phases in liquid chromatography as well as sample preparation are quickly evaluated or selected based on the literature, univariate experiments and/or the laboratory's experience, with the aim of approaching ATP. Regarding sample preparation, the choice of target concentration values as well as the preparation of reference solutions should be considered, paying attention to high performance and efficiency. For example, in environmental sample preparation, one needs to carefully consider the selection of analyte(s), pay due attention to the properties of the extraction solvent as well as carry out the sampling of the environmental probe in a representative way.

In environmental analysis, the setting of an ATP depends on the threshold values of contaminants established by relevant authorities for each case (air, water, soil). Compliance control requires using reliable and reproducible methods for sampling, sample pre-treatment prior to analysis and analytical measurements to produce results that are valid for legal purposes. This stresses the importance of the methods of validation, which is especially relevant in the development of reliable analytical procedures [9,10]. Quality parameters such as accuracy and robustness are therefore evaluated only at the end of the method development process. Notably, method robustness, defined by the ICH guideline Q2(R1) as 'a measure of the capacity of the method to remain unaffected by small but deliberate variations in method parameters' [11], is a key point in assessing the quality of an analytical procedure.

AQbD involves good understanding and control of an analytical method, based on quality risk management [12]. Risk assessment can be carried out using several tools. In the first place, for visualisation it is helpful to use a fishbone or Ishikawa diagram, which identifies potential factors and distributes factor-related risks into categories associated with instrumentation, materials, methods, measurements, laboratory environment and human factors [13]. In addition to the above, the recognition of significant factors for DS and their optimization can be made using statistical analysis and methods of design of experiment (DoE) [14].

In this study, the concept of DS was applied to developing a modified analytical method for compounds from oil-spills, namely polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and aliphatic and aromatic fractions, which are known as accumulating contaminants. PAHs are formed mainly during the incomplete combustion of organic materials. PCBs can be released into the soil by leakages from landfills with PCB-containing products and disposed industrial waste [9]. ATP was related to the development of a precise and accurate method for determining oil compounds in soil and sediment using the gas chromatograph-mass spectrometry (GC-MS) technique in order to support uncertainty estimation throughout the whole range of analyte concentrations and achieve the necessary limits of quantification (LoQ). In this case, the peak areas on chromatograms and the resolution between the peaks were used as critical methods attributes (CMAs).

An analytical target in this case is the oil spill into the soil and the compounds that are spilled, namely 16 parent PAHs and their alkylated analogues, seven PCBs, four aliphatic fractions from octane to pentatriacontane and different aromatic compounds. The concept of DS applied to the target compounds in this study consists of the following steps: identification of critical method parameters, performance assessment, establishment of acceptable ranges, optimization of conditions, method validation, monitoring and control.

1.2. Legal limits of design space for oil products in the environment (Scandinavian examples)

An increasing number of countries have established threshold values to monitor and evaluate the content of contaminants in soil, water and air. Compliance control requires using reliable, reproducible and validated methods for sampling,

sample pre-treatment prior to analysis and analytical measurements to produce results valid for legal purposes [10]. This way, DS is referring to a multidimensional combination of method parameters and conditions that ensures method robustness, reliability and performance [15]. The use of chemometrics in environmental studies is a powerful tool that distinguishes three main areas of interest: quantitative chemical analysis, monitoring for environmental quality assessment and prediction of toxicological effects [16]. It is particularly important in the development of analytical methods for complex analyses, such as the determination of PAHs, PCBs and aliphatic and aromatic compounds in a single run.

The compounds used for this study are emerging contaminants, which were chosen due to their potential risks and increasing occurrence in soil and sediment. In Scandinavian countries, great attention is paid to the pollution of soil and its remediation [17]. In Norway, a set of guidelines has been developed by the Norwegian Environment Agency [18,19]. The Environmental Damage Fund, managed by the Ministry of the Environment, is responsible for compensating for the costs of remediating oil-contaminated soil and groundwater when the party that caused the accident is not known [20]. The Soil Contamination Act, which entered into force on 1 January 2000 in Denmark, has two main objectives: one is groundwater protection and the other is prevention of health problems that may arise from the contaminated areas [21,22]. The Swedish Environmental Code harmonizes the general rules and principles in environmental legislation [23]. The Swedish Environmental Protection Agency has established guidelines and standards for acceptable levels of contaminants in soil (Table 1) [24,25].

Several methods exist in Sweden for analysing oil compounds in soil. First, there is a method devised by the Swedish Environmental Protection Agency, which was then further developed under the so-called SPIMFAB project (Svenska Petroleum Institutet Miljösaneringsfond AB) [26] for analysing aliphatic fractions in soil. Also, the instruction for carrying out the analysis of the aromatic fractions >C16–C35 [27] is available. In addition, a method for analysing PAHs that meets regulatory requirements of contaminants in soil can be used. Table 1 presents the limits of contaminants in the living and industrial zones for PCB-7 (sum of PCB 28, PCB 52, PCB 101, PCB 118, PCB 138, PCB 153, PCB 180), PAH-L (sum of light PAHs: naphthalene, acenaphthene, acenaphthylene), PAH-M (sum of medium PAHs: fluorene, phenanthrene, anthracene, fluoranthene, pyrene), PAH-H (sum of heavy PAHs: benz(a)anthracene, chrysene, benzo(b)fluoranthene, benzo(k)fluoranthene, benzo(a)pyrene, indeno(1,2,3-cd)pyrene, dibenz(a,h)anthracene, benzo(g,h,i)perylene) as well as for different aromatic and aliphatic fractions. Table 2 lists the limits of contaminants in the living zone for aliphatic fractions and PAHs in Denmark. In that country, all the methods of analysis developed must comply with the existing soil quality criteria for oil and/or petrol products in soil and must be carried out according to the methods prescribed by the Danish Environmental Protection Agency [22].

Table 1. Concentration limits for contaminated land in Sweden [25]

Compound	Limit in the living zone, mg kg^{-1}	Limit in the industrial zone, mg kg^{-1}
PCB-7	0.008	0.2
PAH-L	3	15
PAH-M	3	20
PAH-H	1	10
Aliphatic > C8–C10	20	120
Aliphatic > C10–C12	100	500
Aliphatic > C12–C16	100	500
Aliphatic > C16–C35	100	1000
Aromatic > C10–C16	10	50
Aromatic > C10–C16	3	15
Aromatic > C16–C35	10	30

Table 2. Soil quality criteria given by the Danish Environmental Protection Agency for sensitive lands [21,22,28]

Compound	Soil quality criteria for sensitive land, mg kg^{-1}
C6–C10	25
> C10–C15	35
> C15–C20	55
> C20–C40	150
C6–C40	150
PAHs, total*	1.5
Benzo(a)pyrene	0.1
Dibenz(a,h)anthracene	0.1
PCBs, total	0.02

* defined as the sum of individual components: fluoranthene, benzo(b,k)fluoranthene, benzo(a)pyrene, dibenz(a,h)anthracene and indeno(1,2,3-cd)pyrene

The values for the hydrocarbon fractions given in the Danish regulation differ from those presented in the Swedish regulation. The difference may be mainly due to the fact that the hydrocarbon values in Denmark were determined by gas chromatography flame ionization detector (GC-FID) and in Sweden, by GC-MS. For the PAHs analysis, both countries used GC-MS.

2. Steps for the new method development

2.1. Analysis scheme

In this study, the legal aspect of DS for oil compounds in soil is based on the guidelines devised by the Swedish Environmental Protection Agency. Also, the analytical performance criteria requested were defined by this agency. In soil analysis, the number of the samples to be analysed is usually relatively high, and the analysis method has to be applicable to different types of soil.

In the present study, all samples were analysed for aliphatic and aromatic fractions as well as for PAHs, and approximately 80% of samples for PCBs. It means that the combined method was highly preferable for analysing all compounds in one run.

The concentration limits of contaminants set in regulations and the expected detection limits require defining the critical method parameters for method development. The flow chart (Fig. 1) shows the steps of the concept of DS applied to the target compounds in this study.

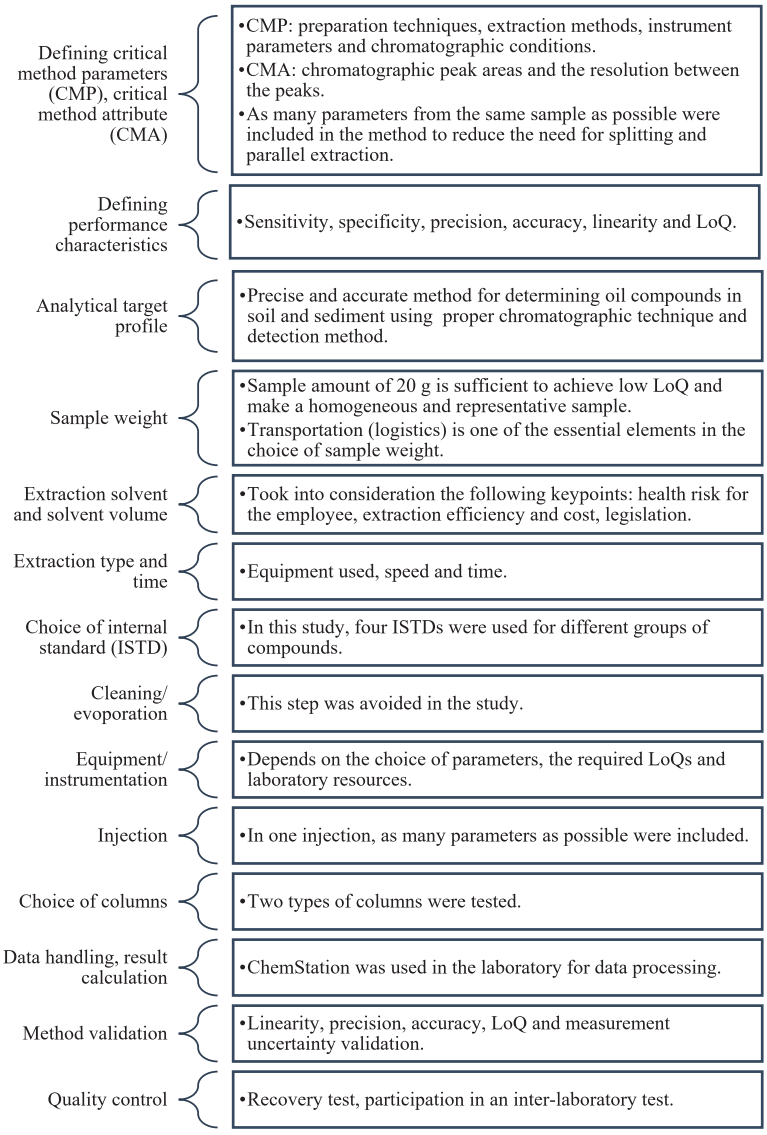

Fig. 1. Setting DS for oil products in soil.

2.2. Critical method parameters

There are several critical method parameters to choose from, such as the conditions of extraction in sample preparation, GC-MS instrument parameters and the chromatographic conditions in the analysis of contaminants. All these conditions and parameters have a significant impact on the performance of the analysis. The next step is to assess the influence of each critical method parameter on the performance characteristics, such as sensitivity, selectivity, precision, accuracy, linearity and LoQ. For that, the method development experiments and, respectively, the results were evaluated to establish the acceptable criteria for each critical parameter. The further step was to optimize the analytical procedure conditions to achieve the best compromise between sensitivity, selectivity and other performance criteria. Here the peak areas on chromatograms and the resolution between the peaks were CMAs because these are directly related to the strict threshold values of the content of contaminants in the soil established by the authorities [25]. After this kind of optimization, the method was validated using standard procedures. To ensure that the method remains robust and reliable over time, the monitoring and control of critical parameters was performed by a recovery test and participation in an interlaboratory test.

3. Method optimization

3.1. Preliminary studies

The GC-MS technique was chosen for determining oil compounds in soil and sediment because of its ability to separate compounds in complex mixtures and support accuracy and precision within the whole range of analyte concentrations required by official regulations [25,26].

The method for the PAHs analysis was based on the former laboratory method [29] with GC-MS instrumentation used. The analyses of the target compounds were performed on an Agilent 7890B GC system that was equipped with a mass-selective detector 5977B MS. Experiments were conducted using different soil types (loamy, clayey, sandy), different weights, extraction volumes, time and speed as well as two types of GC columns (Rxi-XLB (low-polarity proprietary phase) and Rxi-17Sil (midpolarity crossbond phase)) to evaluate the impact of the above parameters on performance characteristics (sensitivity, selectivity, precision, accuracy, linearity and LoQ). Both columns, according to the recommendation of the producer (RESTEK), exhibit excellent inertness and selectivity for active environmental compounds, such as PCB congeners and PAHs. However, for aliphatic compounds, no information about selectivity was available. Preliminary experiments were conducted using the chromatographic conditions recommended by the column producer for the PAH and PCB analyses. A mixture of target compounds diluted in n-hexane was injected into GC-MS. The standards used were: calibration/window defining hydrocarbon standard (C8–C40 1000 mg/L of each compound) from AccuStandard, New Haven, USA; custom mix, multistandard alkylated PAHs and aromatics in toluene (100 mg/L of each compound) from Chiron, Trondheim, Norway; PAH mix-9 100 ng/µL in acetonitrile from Dr. Ehrenstorfer GmbH, Augsburg, Germany; PCB mix-3 (10 ng/µL in iso-octane) from Dr. Ehrenstorfer GmbH. Both columns showed sufficient resolution for PAHs and PCBs; however, Rxi-XLB column was selected as the analytical column because it produced the best peak shape for aliphatic compounds.

An initial screening study of the analytical procedure was carried out examining the oven and inlet temperatures, flow rate and temperature gradient – the parameters known to have a significant impact on peak retention and separation in GC-MS analysis. The LoQ, reproducibility and measurement uncertainty experiments were conducted and calculated according to the Nordtest guide [30]. Table 3 shows results from the preliminary study.

3.2. Selection of chromatographic parameters

Two kinds of problems need to be solved within the framework of DS by experimental design. The first is to discover which factors may significantly affect the response of an experiment, and the second is to find factor values that optimize the response [31]. The results of the initial screening and primary risk assessment revealed that the three factors – the initial and final oven temperatures as well as the inlet temperature – greatly affected the retention time and separation of peaks.

The three-level DoE was performed by using JMP software [32] to investigate the optimal levels for the initial oven temperature, the final oven temperature and the inlet temperature. The chromatographic peak areas and the resolution between the peaks were compared under different

Table 3. Target LoQ, reproducibility and measurement uncertainty

Compound	LoQ, mg kg^{-1}	Reproducibility, %	Measurement uncertainty, %
Individual PAHs	< 0.030	< 20	< 30
Aliphatics > C8–C10	< 5	< 20	< 30
Aliphatics > C10–C12	< 5	< 20	< 30
Aliphatics > C12–C16	< 5	< 20	< 30
Aliphatics > C16–C35	< 10	< 20	< 30
Aromatics > C8–C10	< 3	< 20	< 30
Aromatics > C10–C16	< 1	< 20	< 30
Aromatics > C16–C35	< 1	< 20	< 30
Methyl-pyrene/fluoranthene	< 0.5	< 20	< 30
Methyl-chrysene/benz(a)anthracene	< 0.5	< 20	< 30
Individual PCBs	< 0.002	< 20	< 30

instrument parameters. The response of the experiments was the ratio of the peak areas of C35 and C10. The aim was to find whether this ratio was equal to one. When the areas of C10 and C35 are similar, the chromatogram looks symmetrical and the high-mass discrimination effect relatively weak [33]. A total of 21 experiments were conducted using the mixture of aliphatic compounds C10–C35 in the concentration of 25 mg L^{-1} for each compound, and the area ratio of C35/C10 was calculated. Based on these results, it was demonstrated that the required peak area ratio equal to one would be achieved under any conditions within the ranges of 35–40 °C for the initial oven temperature, 335–340 °C for the final oven temperature and 320–325 °C for the inlet temperature. The separation between critical pairs (e.g. benz(a)anthracene and chrysene) and the chromatographic peak areas of PAHs, PCBs and aromatic compounds were evaluated under these conditions, and, furthermore, method validation was conducted.

3.3. Method validation

3.3.1. Extraction and sample preparation

The extraction of PAHs, aliphatics, aromatics and PCBs was based on a modified method of ISO 16703:2011 [34]. Aliquots of 20 g from the soil sample were first extracted in acetone, and then hexane and sodium pyrophosphate (22g L^{-1} $Na_4O_7P_2$) were added. The samples were not dried prior to the analysis. The water content was determined according to ISO 11465:1993 and the analysis results calculated on the dry matter basis [35]. The samples were centrifuged and the upper layer dried with Na_2SO_3 to remove residual water. The dried extract was transferred into the GC vial for the GC-MS analysis.

3.3.2. Instrumental parameters

The GC-MS analysis was carried out under the following conditions: pulsed pressure splitless injection; column flow 1.8 mL min^{-1}; oven temperature programme: initial temperature 40 °C (hold time 3 min); 40 °C min^{-1} to 160 °C; 25 °C min^{-1} to 315 °C (hold time 3.5 min); 40 °C min^{-1} to 340 °C (hold time 9.75 min); interface temperature 320 °C. The analyses were performed in a selected ion monitoring (SIM) mode using two or three ions for each compound for identification [25,29,36].

3.3.3. Results

Table 4 shows that all the results fall within the expected laboratory uncertainty (30%). The reproducibility of the measurements of PCBs, PAHs and aliphatic and aromatic fractions was lower than 20%. The difference between the expected concentrations and those obtained was below 20%. The recovery of PAH compounds ranged from 88 to 112%. For all parent PAHs, a linear calibration curve in the range of 50–5000 ug L^{-1} was obtained; for alkylated PAH compounds it was 0.4–5 mg L^{-1}, for PCBs it was 15–600 ug L^{-1}, and for aliphatic compounds with a coefficient of determination ($R^2 \geq 0.998$) it was 10–50 mg L^{-1}. LoQ, reproducibility, uncertainty and accuracy were assessed using certified reference materials (CRMs): soil CRM BAM-U013b from LGC standards, Teddington, Middlesex, UK; TPH sand 1 (CRM372-100G) from Sigma-Aldrich/Merck, Darmstadt, Germany; SETOC sample from WEPAL, Bennekom, The Netherlands. Two sources of uncertainty were estimated: the uncertainty component from the certified or nominal value and the uncertainty component for bias. Finally, the calculated standard uncertainty was multiplied by a 'coverage factor', k = 2, to provide an expanded uncertainty [30].

In the concentration range investigated, the accuracy, precision and uncertainty achieved were satisfactory. All the calculated LoQs were lower than those required by the Swedish legislation. The only limitation of the method was its failure to analyse highly polluted samples. In that case, the samples needed to be diluted and the instrument more frequently maintained.

4. Analytical limits for design space for oil products in the environment

In the literature, one can find studies which deal with the development of the methods of quantification and identification of PAHs [37,38], PCBs [39] and aliphatic and aromatic fractions [40] in soil and also provide the validation of these methods. The methods differ in terms of sample preparation (sample cleanup, filtration, concentration, evaporation), extraction type (Soxhlet extraction, sonication or ultrasonic treatment, mechanical agitation, accelerated solvent extraction (ASE), solid-phase microextraction (SPM)) and analysis/detection (GC-MS, GC-FID, high-performance liquid chromatography (HPLC)). It is outlined that various factors such as sample type, target compound concentration, required sensitivity and available instrumentation influence method choice. Each method has its advantages and disadvantages that the laboratories must consider in setting DS for improving the methods.

In this study, the pros and cons of 10 different GC-MS, GC-FID and HPLC methods, including the one developed in the current study, were compared (Table 5). The cost of extraction techniques, the approximate time spent for the preparation of one sample (including the extraction time) and the greenness of the method were also evaluated. In addition, LoQ, repeatability/reproducibility and recovery rate of the methods were compared.

This comparison of the methods reported in the literature disclosed the absence of a combined method for the simultaneous analysis of PAHs, PCBs and aliphatic and aromatic compounds. The present comparison had to be made between the methods of analysis of PAHs only.

4.1. Analysis techniques used

The most commonly used methods for the PAHs analysis in soils are chromatographic techniques with mass-spectrometry detector, including HPLC and GC-MS. In most cases, GC-MS is the preferred instrument [41]. It is applicable to the majority of parameters and required LoQs. GC-MS was also used in our method due to the availability in the laboratory.

Table 4. The calculated reproducibility, accuracy, uncertainty and LoQ

Compound	Reproducibility, %	Accuracy, %	Uncertainty, %	LoQ, mg kg^{-1}
Naphthalene	2	11	21	0.015
Acenaphthylene	5	10	19	0.014
Acenaphthene	4	12	23	0.016
Fluorene	4	5	12	0.018
Phenanthrene	5	6	10	0.014
Anthracene	7	8	16	0.016
Fluoranthene	3	8	17	0.014
Pyrene	2	10	20	0.015
Benz(a)anthracene	6	7	14	0.019
Chrysene	5	8	16	0.017
Benzo(b)fluoranthene	4	8	15	0.019
Benzo(k)fluoranthene	8	8	17	0.020
Benzo(a)pyrene	2	11	22	0.018
Indeno(1,2,3-cd)pyrene	6	7	14	0.014
Dibenz(a,h)anthracene	4	6	13	0.016
Benzo(g,h,i)perylene	7	9	18	0.016
Aliphatics > C8–C10	10	12	24	5
Aliphatics > C10–C12	9	9	17	5
Aliphatics > C12–C16	8	8	19	5
Aliphatics > C16–C35	10	10	19	10
Aromatics C8–C10	7	9	18	0.9
Aromatics C10–C16	3	3	19	0.9
Methyl-pyrene/fluoranthene	10	13	26	0.2
Methyl-chrysene/benz(a)anthracene	12	14	27	0.1
Aromatics C16–C35	10	12	24	0.9
PCB28	6	9	17	0.001
PCB52	9	10	20	0.001
PCB101	9	12	24	0.002
PCB118	5	5	10	0.002
PCB153	6	13	26	0.002
PCB138	4	4	9	0.002
PCB180	3	11	22	0.002

4.2. Sample size and extraction solvents

Samples of 1 to 60 g were used in different methods. Sample weight is usually selected by the type of solid material and the LoQs to be achieved. The weight of 20 g for a validated method was chosen to get a representative sample with a sufficient LoQ.

The extraction solution volume in the methods under comparison was in the range of 0–150 mL. This is important in determining the greenness of the method. Usually, the extraction solvent used is acetone, hexane or dichloromethane. Dichloromethane can extract many compounds; however, some countries avoid working with dichloromethane as it may cause health issues [42]. The advantage of using acetone and hexane is that it enables more polar compounds to be extracted. The modified method in this study first used acetone alone and then hexane was added. It was confirmed that using solvents in this specific order allows for higher extraction efficiency compared to using solvents individually.

4.3. Calibration standards

Internal standards (ISTDs) or surrogate standards were used in most methods. The number of standards ranged from 3 to 11. External calibration was applied in Methods 2, 4, 7 and 9. The right internal standards can compensate for the low extraction efficiency and reduce the matrix effect. The researchers of the current study used four ISTDs: one for aliphatic compounds, one for high PAHs and PCBs, one for middle PAHs and one for light PAHs.

4.4. Extraction method

Different extraction methods are used. The selection of the extraction method is challenging because the efficiency of the

Table 5. Methods of the PAHs analysis

Method	Reference	Equipment	Extraction type	Number of stages	Time, min	Greenness score	LoQ, mg kg^{-1}	Reproducibility, %	Recovery, %
1 (current study)	N/A	GC-MS	Mechanical	5	60	0.42	0.014–0.020	2–8	88–112
2	42	HPLC-FLD-DAD	Soxhlet	9	1460	0.33	0.001–0.0095	< 5	29.8–96
3	43	GC-MS	ASE	11	40	0.32	0.0001–0.00017	1.3–18.4	80–100
4	44	HPLC-UV-DAD	Sonication	12	190	0.25	0.00005–0.00171	3.34–14.68	82–106
5	45	GC-MS	Soxhlet	9	400	0.22	< 1	1–33	N/A
6	37	GC-MS	Sonication	12	150	0.23	0.00136–0.01186	2–20	68–108
7	46	NTD-GC-FID	SPM	2	40	0.68	0.000001–0.00001	7.3–13.2	92.53–110.75
8	47	GC-MS	Sonication	12	100	0.33	0.0006–0.0054	1.1–13.7	87–128
9	38	HPLC-FLD-DAD	Sonication	7	40	0.34	0.00741–0.45258	< 14	84–87
10	48	GC-MS	Mechanical	5	730	0.45	0.005	< 15	90

extraction of PAHs from soils and sediments is influenced by several factors, such as soil moisture, PAHs content in samples and the texture of soils [43]. The solid-phase microextraction (SPME) with a needle-trap device (NTD) used in Method 7 is a green, solvent-free preparation of the sample; however, packing of sorbents into NTD by this method is a time-consuming process with low repeatability [46]. Soxhlet extraction allows for prolonged and continuous extraction, which can give higher yields compared to other methods [44]. Methods 2 and 5 using Soxhlet extraction require 24 and 6 hours, respectively, to complete a single extraction. Besides being time-consuming, this method uses a large amount of solvent, which is a concern in terms of cost and environmental impact. In Method 5, the extraction solution volume of 120 mL of dichloromethane was used. The accelerated solvent extraction (Method 3) was faster, consumed less solvent and could be easily automated; however, the equipment for accelerated solvent extraction is expensive. In addition, heat-sensitive compounds may still be affected during accelerated solvent extraction [43]. Similarly to Method 10, the present study used mechanical extraction. This extraction is adaptable to different soil types and is easy to implement in a laboratory setting. Mechanical agitation can accelerate the extraction process by facilitating the movement of the solvent through the soil matrix. However, continuous mechanical agitation requires energy, which may increase operational costs.

4.5. Cleanup steps and sample concentration

Different matrices require the application of specific extraction and cleanup steps [45]. As they are mostly related to the maintenance of the instrument, it should be considered whether cleaning is necessary. Some methods include all the steps of preparation, which makes a method complex and time-consuming (Methods 4, 6, 8). In our method, the cleanup stage was found to be unnecessary. Evaporation of extract is usually needed to obtain low LoQ. Almost all methods (2, 3, 4, 5, 6, 8, 9) decreased the amount of extract via rotary evaporation. Method 2 showed the lightest PAHs to be quite sensitive to the evaporation step due to the distinct volatility loss during the evaporation process [42]. The LoQ in our method without evaporation was less than 0.020 mg kg^{-1}. Comparing the threshold values in terms of the content of contaminants in soil established by the Swedish authority, the modified method was sensitive enough. Methods 2, 3, 4, 6, 7, 8 (Table 5) have lower LoQs compared to the current method. However, the cleanup and sample concentration steps make them complicated and time-consuming (with the exception of Method 7 where the number of stages is two). An extra purification step of the samples before the GC analysis could be skipped in the current study, making this part simpler and faster.

5. Assessment of method greenness

The determination of PAH compounds in soil or sediment is a complex analytical procedure that requires significant energy and material inputs. It is also important to assess the environmental friendliness of the methods used. For the latter purpose, the AGREE metric analysis following the 12 principles of green analytical chemistry (GAC) was applied in the present study. The maximum score equal to one represents a methodology that is fully compliant with the 12 principles of GAC [46]. The assessment was performed using the software with an automatically generated graph and an assessment report [47]. Recently, some complementary green assessment tools have become available with the software that can estimate the applicability and functionality of an analytical method [48].

In Table 5, column 'Greenness score', the calculated AGREE scores for all the methods under consideration are presented. Using old classical extraction techniques is the reason for the low AGREE analysis scores in all the methods. The Soxhlet extraction and sonication procedures are shown to have explicitly worse scores (less than 0.35), while the mechanical procedures (our method and Method 10) have similar scores (0.42 and 0.45, respectively). One exception is Method 7, where using SPM for the extraction of analytes gives a higher score.

Figure 2 shows the results of the AGREE analysis for our method. The procedure involves an external sample treatment with a reduced number of steps (principle (p) 1) and a soil sample of 20 g (p 2). The measurement is off-line (p 3) and the procedure involves five distinct steps (p 4). The procedure is not automated but involves a miniaturized sample preparation technique (p 5). No derivatization agents are used in the analysis (p 6). Analytical wastes contain 20 g of sample, 20 mL of acetone, 10 mL of hexane and 1 g of Na_2SO_3 (p 7). 43 analytes are determined in a single run, and the sample throughput is two samples per hour (p 8). GC-MS is the most energy-demanding analytical technique (p 9). Some of the reagents (acetone) may originate from bio-based sources (p 10). The procedure requires 10 mL of toxic solvents (hexane) (p 11), while PAHs are persistent and toxic to aquatic life (p 12).

The comparison shows that there is no single 'best' method for measuring all types of soil-contaminating PAH compounds. The comparison took into account the cost per sample, energy consumption, time of extraction, the time the analyst spent on sample preparation and the amount of waste, which was crucial in the choice of the appropriate method.

The comparison of the 10 methods revealed the authors' method to be the most suitable for the intended purpose in every aspect, being also fit for rapid analysis. With all steps of sample preparation included, the time of analysis was approximately one hour. Precision, recovery and peak resolution of the method were highly acceptable and the LoQ complied with the regulatory requirements. The method involves five steps of sample preparation without cleanup and evaporation, making it greener. The greenness score of the present method generally surpasses that of the other methods.

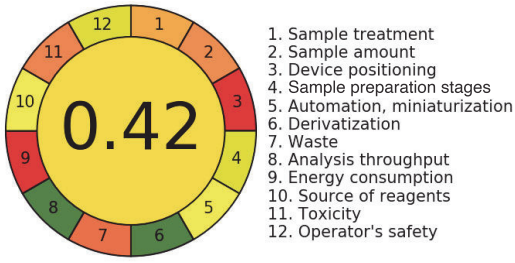

Fig. 2. Results of the AGREE analysis.

6. Conclusions

Design space as an important step in method development for environmental analysis involves a careful study of limits set by authorities as well as the possibilities of existing analytical methods. This provides a strong foundation for the design of experiment, which has helped to identify the optimal instrumental conditions for developing the new method. The simultaneous identification and quantification of 16 polycyclic aromatic hydrocarbons, seven polychlorinated biphenyls, alkylated PAHs and aliphatic compounds in soil involved determining their quality by a design approach in environmental analysis. For sample preparation, the GC-MS analytical technique with the mechanical extraction method was employed. Using a proper GC column with the optimized oven programme enabled achieving sufficient resolution and recovery. Sample preparation was included as part of method development through which an extra purification step of the samples before the GC analysis could be skipped, making this part simpler and faster. By modifying the extraction procedure, it was possible to analyse multiple analytes of PAHs, PCBs and aliphatic and aromatic compounds routinely in a single run with results that comply to Swedish regulations. Decreasing the number of extraction steps in the analysis of a single sample will reduce staff numbers as well as consumable costs. In addition, it reduces the complexity of laboratory organization by diminishing the number of different processes to be followed. Nowadays, it has become almost mandatory to assess the greenness of the developed method and to provide numerical measures for that. Another mandatory part of developing a new method and introducing it to laboratory practice is its validation, which includes an interlaboratory comparison.

The present experiment demonstrated that there is room for method development with improved technical and analytical parameters, which can be used for controlled testing of soil and sediment for PAHs, PCBs and aliphatic and aromatic compounds with reliable accuracy, precision, linearity and recovery rates in the investigated concentration range.

Data availability statement

All data are available in the article.

Acknowledgements

We thank the Estonian Research Council (grant No. PRG1784) for supporting this work. The publication costs of this article were covered by the Estonian Academy of Sciences.

References

1. ICH. *ICH Harmonized Tripartite Guideline. Pharmaceutical Development Q8(R2) 2009.* https://database.ich.org/sites/default/files/Q8%28R2%29%20Guideline.pdf (accessed 2024-11-24).
2. ICH. *ICH Harmonized Tripartite Guideline. Pharmaceutical Quality System Q10 2008.* https://database.ich.org/sites/default/files/Q10%20Guideline.pdf (accessed 2024-11-24).
3. ICH. *ICH Harmonized Guideline. Quality Risk Management Q9(R1) 2023.* https://database.ich.org/sites/default/files/ICH_Q9%28R1%29_Guideline_Step4_2023_0126_0.pdf (accessed 2024-11-24).
4. Hakemeyer, C., McKnight, N., St. John, R., Meier, S., Trexler-Schmidt, M., Kelley, B. et al. Process characterization and design space definition. *Biologicals*, 2016, **44**(5), 306–318. https://doi.org/10.1016/j.biologicals.2016.06.004
5. Ng, K. and Rajagopalan, N. Application of quality by design and risk assessment principles for the development of formulation design space. In *Quality by Design for Biopharmaceuticals: Principles and Case Studies* (Rathore, A. S. and Mhatre, R., eds). Wiley, Hoboken, NJ, 2008, 161–174. https://doi.org/10.1002/9780470466315.ch9
6. Manzon, D., Claeys-Bruno, M., Declomesnil, S., Carité, C. and Sergent, M. Quality by design: comparison of design space construction methods in the case of design of experiments. *Chemometr. Intell. Lab. Sys.*, 2020, **200**, 104002. https://doi.org/10.1016/j.chemolab.2020.104002
7. Cocco, P., Ayaz-Shah, A., Messenger, M. P., West, R. M. and Shinkins, B. Target product profiles for medical tests: a systematic review of current methods. *BMC Med.*, 2020, **119**(18). https://doi.org/10.1186/s12916-020-01582-1
8. Jackson, P., Borman, P., Campa, C., Chatfield, M., Godfrey, M., Hamilton, P. et al. Using the analytical target profile to drive the analytical method lifecycle. *Anal. Chem.*, 2019, **91**(4), 2577–2585. https://doi.org/10.1021/acs.analchem.8b04596
9. Kalbe, U., Lehnik-Habrink, P., Bandow, N. and Sauer, A. Validation of European horizontal methods for the analysis of PAH, PCB and dioxins in sludge, treated biowaste and soil. *Environ. Sci. Eur.*, 2019, **31**(29). https://doi.org/10.1186/s12302-019-0211-3
10. Eurachem. *The Fitness for Purpose of Analytical Methods. A Laboratory Guide to Method Validation and Related Topics. 2nd ed., 2014.* https://www.eurachem.org/images/stories/Guides/pdf/MV_guide_2nd_ed_EN.pdf (accessed 2024-11-24).
11. ICH. *ICH Harmonized Tripartite Guideline. Validation of Analytical Procedures: Text and Methodology Q2(R1) 1994.* https://database.ich.org/sites/default/files/Q2%28R1%29%20Guideline.pdf (accessed 2024-11-24).
12. Deidda, R., Orlandini, S., Hubert, P. and Hubert, C. Risk-based approach for method development in pharmaceutical quality control context: a critical review. *J. Pharm. Biomed. Anal.*, 2018, **161**, 110–121. https://doi.org/10.1016/j.jpba.2018.07.050
13. Ishikawa, K. *Introduction to Quality Control*. Taylor & Francis, 1990.
14. Fukuda, I. M., Pinto, C. F. F., dos Santos Moreira, C., Saviano, A. M. and Lourenço, F. R. Design of experiments (DoE) applied to pharmaceutical and analytical quality by design (QbD). *Braz. J. Pharm. Sci.*, 2018, **54**(special), e01006. https://doi.org/10.1590/s2175-97902018000001006
15. von Stosch, M., Schenkendorf, R., Geldhof, G., Varsakelis, C., Mariti, M., Dessoy, S. et al. Working within the design space: do our static process characterization methods suffice? *Pharmaceutics*, 2020, **12**(6), 562. https://doi.org/10.3390/pharmaceutics12060562
16. Mas, S., de Juan, A., Tauler, R., Olivieri, A. C. and Escandar, G. M. Application of chemometric methods to environmental analysis of organic pollutants: a review. *Talanta*, 2010, **80**(3), 1052–1067. https://doi.org/10.1016/j.talanta.2009.09.044
17. European Environment Agency. *Down to Earth: Soil Degradation and Sustainable Development in Europe 2000.* https://www.eea.europa.eu/publications/Environmental_issue_series_16/file (accessed 2024-11-24).
18. Lehoux, A. P., Petersen, K., Leppänen, M. T., Snowball, I. and Olsen, M. Status of contaminated marine sediments in four Nordic countries: assessments, regulations, and remediation approaches, *J. Soils Sediments*, 2020, **20**, 2619–2629. https://doi.org/10.1007/s11368-020-02594-3
19. Norwegian Pollution Control Authority. *Guidelines for the Risk Assessment of Contaminated Sites*. https://www.miljodirektoratet.no/globalassets/publikasjoner/klifsft/publikasjoner/andre/1691/ta1691.pdf (accessed 2024-11-24).
20. Ministry of the Environment. *Environmental Damage Fund*. https://ym.fi/en/environmental-damage-fund (accessed 2024-11-24).
21. Danish Environmental Protection Agency. *Environmental Factors and Health. The Danish Experience 2001.* https://www2.mst.dk/udgiv/publications/2001/87-7944-519-5/pdf/87-7944-518-7.pdf (accessed 2024-11-24).
22. Grøn, C., Borling, P., Andersen, L., Cohr, K.-H., Hansen, J. B., Oberender, A. et al. *Olie i jord – forslag til analysemetode og justering af jordkvalitetskriterier, samt grundlag for afskæringskriterier. Miljøprojekt Nr. 1225 2008. Teknologiudviklingsprogrammet for jord- og grundvandsforurening* (*Oil in soil – analysis method proposal and adjustment for soil quality criteria, as well as basis for cutoff criteria. Environmental project No. 1225 2008. Technology development program for soil and groundwater pollution*). https://www2.mst.dk/udgiv/publikationer/2008/978-87-7052-725-5/pdf/978-87-7052-726-2.pdf (accessed 2024-11-24).
23. Nordic Council of Ministers. *Nordic Environmental Permitting Processes.* https://www.researchgate.net/publication/372090099_Title_Nordic_Environmental_Permitting_Processes (accessed
24. Svenska Petroleum Institutet. *Efterbehandling av förorenade bensinstationer och dieselanläggningar 2010* (*Remediation of contaminated gas stations and diesel facilities*). https://drivkraftsverige.se/wp-content/uploads/2023/02/SPBI-rekebh-fororenade-bensinst-dieselanluppdaterad20120129.pdf (accessed 2024-11-24).
25. Naturvårdsverket. *Riktvärden för förorenad mark. Modellbeskrivning och vägledning* (*Guideline values for contaminated soil. Model description and guidance 2009*). https://www.naturvardsverket.se/globalassets/media/publikationer-pdf/5900/978-91-620-5976-7.pdf (accessed 2024-11-24).
26. Eliaeson, K., Kaj, L., Rosenqvist, L., Allard, A-S. and Graae, L. *Påverkan från naturligt organiskt material i GC-MS analyser 2018* (*The influence of natural organic material in GC-MS analyses*). https://www.ivl.se/download/18.694ca0617a1de98f473864/1628417134444/FULLTEXT01.pdf (accessed 2024-11-24).
27. Olivares, L. and Lindén, P. *Instruktion för analys av fraktionen Aromater >C16-C35 2010* (*Instruction for analysis of the aromatic fraction >C16–C35*). https://www.naturvardsverket.se/4a437e/globalassets/vagledning/fororenade-omraden/riktvarden/sis-tk-535-n-012-sis-instruktion-tunga-aromater-c16-c35.pdf (accessed 2024-11-24).
28. Gomes, H. I., Ottosen, L. M., Ribeiro, A. B. and Dias-Ferreira, C. Treatment of a suspension of PCB contaminated soil using iron nanoparticles and electric current. *J. Environ. Manage.*, 2015, **151**, 550–555. https://doi.org/10.1016/j.jenvman.2015.01.015

29. Jurjeva, J. and Koel, M. The chemometric approach to identification of residual oil contamination at former primitive asphalt pavement plants. *Oil Shale*, 2019, **36**(3), 410–430. https://doi.org/10.3176/oil.2019.3.04
30. Magnusson, B., Näykki, T., Hovind, H. and Krysell, M. *Handbook for calculation of measurement uncertainty in environmental laboratories* (*NT TR 537, ed. 3.1) 2012.* https://www.nordtest.info/wp/2012/10/29/handbook-for-calculation-of-measurement-uncertainty-in-environmental-laboratories-nt-tr-537-edition-3-1/ (accessed 2024-11-24).
31. Hibbert, D. B. Experimental design in chromatography: a tutorial review. *J. Chromatogr. B*, 2012, **910**, 2–13. https://doi.org/10.1016/j.jchromb.2012.01.020
32. JMP. https://www.jmp.com/en_us/home.html (accessed 2024-11-24).
33. Wang, X., Cheng, Y., Dong, M., Zhao, L., Chen, L., Feng, T. et al. Research on correcting high-mass discrimination effect of the quadrupole mass spectrometer with dodecane. 2023. *SSRN*, 2023, early access. https://ssrn.com/abstract=4501270 (accessed 2024-11-24).
34. ISO 16703:2011. *Soil quality – Determination of content of hydrocarbon in the range C10 to C40 by gas chromatography.* https://www.evs.ee/en/evs-en-iso-16703-2011 (accessed 2024-11-24).
35. ISO 11465:1993. *Soil quality – Determination of dry matter and water content on a mass basis – Gravimetric method.* https://www.iso.org/standard/20886.html (accessed 2024-11-24).
36. Robbat, A., Jr. and Wilton, N. M. A new spectral deconvolution – selected ion monitoring method for the analysis of alkylated polycyclic aromatic hydrocarbons in complex mixtures. *Talanta*, 2014, **125**, 114–124. https://doi.org/10.1016/j.talanta.2014.02.068
37. Yamada, T. M., Souza, D. A., Morais, C. R. and Mozeto, A. A. Validation of a method for the analysis of PAHs in bulk lake sediments using GC–MS. *J. Chromatogr. Sci.*, 2009, **47**(9), 794–799. https://doi.org/10.1093/chromsci/47.9.794
38. Volk, S. and Gratzfeld-Huesgen, A. *Agilent application solution. Analysis of PAHs in soil according to EPA 8310 method with UV and fluorescence detection 2011.* https://www.agilent.com/cs/library/applications/5990-8414EN.pdf (accessed 2024-11-24).
39. Krzemień-Konieczka, I. and Buszewski, B. Determining polychlorinated biphenyls in soil using accelerated solvent extraction (ASE). *Pol. J. Environ. Stud.*, 2015, **24**(5), 2029–2033. https://doi.org/10.15244/pjoes/41589
40. Sadler, R. and Connell, D. Analytical methods for the determination of total petroleum hydrocarbons in soil. In *Proceedings of the Fifth National Workshop on the Assessment of Site Contamination* (Langley, A., Gilbey, M. and Kennedy, B., eds), 2003, 133–150. https://www.nepc.gov.au/sites/default/files/2022-09/asc-wkshoppaper-09-tphs-sadler-petrol-hydro-soil-200301.pdf (accessed 2024-11-24).
41. Włóka, D. and Placek, A. Study of the effectiveness of the new method for the rapid PAHs analysis in soil. *Acta Innov.*, 2015, **17**, 5–11.
42. Swedish Chemicals Agency. *Swedish rules on chlorinated solvents.* https://www.kemi.se/en/rules-and-regulations/rules-applicable-in-sweden-only/certain-swedish-restrictions-and-bans/chlorinated-solvents (accessed 2024-11-24).
43. Lau, E. V., Gan, S. and Ng, H. K. Extraction techniques for polycyclic aromatic hydrocarbons in soils. *Int. J. Anal. Chem.*, 2010, **2010**, 398381. https://doi.org/10.1155/2010/398381
44. Zhou, W., Zhang, H., Deng, C., Chen, Y., Liao, J., Chen, Z. et al. Solvent-assisted vacuum desorption coupled with gas chromatography-tandem mass spectrometry for rapid determination of polycyclic aromatic hydrocarbons in soil samples. *J. Chromatogr. A*, 2019, **1604**, 460473. https://doi.org/10.1016/j.chroma.2019.460473
45. Song, Y. F., Jing, X., Fleischmann, S. and Wilke, B.-M. Comparative study of extraction methods for the determination of PAHs from contaminated soils and sediments. *Chemosphere*, 2002, **48**(9), 993–1001. https://doi.org/10.1016/S0045-6535(02)00180-7
46. Sajid, M. and Płotka-Wasylka, J. Green analytical chemistry metrics: a review. *Talanta*, 2022, **238**(P2), 123046. https://doi.org/10.1016/j.talanta.2021.123046
47. Pena-Pereira, F., Wojnowski, W. and Tobiszewski, M. AGREE–analytical GREEnness metric approach and software. *Anal. Chem.*, 2020, **92**(14), 10076–10082. https://doi.org/10.1021/acs.analchem.0c01887
48. Manousi, N., Wojnowski, W., Płotka-Wasylka, J. and Samanidou, V. Blue applicability grade index (BAGI) and software: a new tool for the evaluation of method practicality. *Green Chem.*, 2023, **25**, 7598–7604. https://doi.org/10.1039/d3gc02347h

Disainiruumi kontseptsiooni rakendamine saasteainete pinnases määramise meetodi väljatöötamisel

Jelena Jurjeva ja Mihkel Koel

Disainiruum (DS) määrab protsessi sisendite (nt materjalid ja protsessi parameetrid) ja kriitiliste kvaliteediatribuutide vahelise seose. Kvaliteedijuhtimise seisukohast on see kontseptsioon väga oluline, sest DSis töötamist ei peeta muudatuseks. Disainiruumi rakendamine on vajalik samm keskkonnaanalüüsi meetodite väljatöötamisel, et arvestada saasteainete piirnormidega ja toetada olemasolevate analüüsimeetodite arendamist. Selles töös rakendati disainiruumi kontseptsiooni naftareostusest pärinevate ühendite analüüsimeetodi väljatöötamiseks, keskendudes polütsüklilistele aromaatsetele süsivesinikele, polüklooritud bifenüülidele ning erinevatele alifaatsetele ja aromaatsetele fraktsioonidele. Sihtühenditele rakendatud disainiruumi kontseptsioon koosneb järgmistest etappidest: meetodi kriitiliste parameetrite tuvastamine, tulemuslikkuse hindamine, vastuvõetavate tööpiiride määramine, tingimuste optimeerimine, meetodi valideerimine, seire ja kontroll. Meetodi arendus hõlmas ka muudatusi proovide ettevalmistuses, mis lihtsustas ja kiirendas vastavat osa mõõtmisprotsessis. Lisaks võimaldas disainiruum tuvastada olulisi tegureid ning optimeerida neid statistilise analüüsi ja katsete planeerimise meetodite abil.

Curriculum vitae

Personal data

Name:	Jelena Jurjeva
Date of birth:	16.05.1986
Place of birth:	Narva, Estonia
Citizenship:	Estonian

Contact data

E-mail:	jurjeva.jelena@gmail.com

Education

2014–…	Tallinn University of Technology, School of Engineering, Department of Energy Technology, PhD
2010–2012	Tallinn University of Technology, Faculty of Chemical and Materials Technology, Department of Environmental and Chemical Technology, MSC
2006–2010	Tallinn University of Technology, Faculty of Chemical and Materials Technology, Department of Environmental and Chemical Technology, BSC
1993–2005	Narva Soldino Gymnasium, High school

Language competence

Russian	Native
English	Fluent
Estonian	Fluent

Professional employment

2023–…	The J. Molner Company OÜ, Senior Research Scientist
2019–2023	Eurofins Environment Testing Estonia OÜ, The Head of Gas Chromatography Department
2015–2019	Eurofins Environment Testing Estonia OÜ, Laboratory Analyst
2012–2015	Tallinn University of Technology, Faculty of Chemical and Materials Technology, Department of Polymer Materials, Laboratory of Oil Shale and Renewables Research, engineer
2012 (7 months)	Tallinn Wastewater Treatment Plant, Laboratory Technician

Elulookirjeldus

Isikuandmed

Nimi:	Jelena Jurjeva
Sünniaeg:	16.05.1986
Sünnikoht:	Narva, Eesti
Kodakondsus:	Eesti

Kontaktandmed

E-post:	jurjeva.jelena@gmail.com

Hariduskäik

2014–...	Tallinna Tehnikaülikool, Inseneriteaduskond, Energiatehnoloogia instituut, PhD
2010–2012	Tallinna Tehnikaülikool, Keemiatehnika instituut, MSC
2006–2010	Tallinna Tehnikaülikool, Keemiatehnika instituut, BSC
1993–2005	Narva Soldino Gümnaasium, keskharidus

Keelteoskus

Vene keel	emakeel
Inglise keel	kõrgtase
Eesti keel	kõrgtase

Teenistuskäik

2023–…	The J. Molner Company OÜ, vanemteadur
2019–2023	Eurofins Environment Testing Estonia OÜ, GC osakonna juhataja
2015–2019	Eurofins Environment Testing Estonia OÜ, Laborianalüütik
2012–2015	Tallinna Tehnikaülikool, Polümeermaterjalide instituut, Põlevkivi ja taastuvkütuste teaduslaboratoorium, insener
2012 (7 kuud)	Paljassaare reoveepuhastusjaam, Laborant

ISSN 2585-6901 (PDF)
ISBN 978-9916-80-367-7 (PDF)