#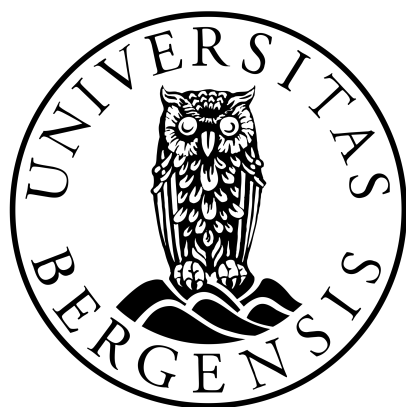 Chemometric Approaches for the Validation and Optimization of a LC-MS/MS Method for the Analysis of Phenanthrene Metabolites in Atlantic Haddock Embryo

Mathias Brevik

Department of Chemistry
University of Bergen

A thesis submitted for the degree of
Master of Science
April 2023

Chemometric Approaches for the Validation and Optimization of a LC-MS/MS Method for the Analysis of Phenanthrene Metabolites in Atlantic Haddock Embryo

Mathias Brevik

Department of Chemistry

University of Bergen

Supervisors

Prof. Svein Are Mjøs, PhD

Department of Chemistry, University of Bergen

Dr. Carey Donald, PhD

Department of Marine Toxicology, Institute of Marine Research

A thesis submitted for the degree of

Master of Science

April 2023

Acknowledgements

Thank you!

Mathias Brevik

Bergen, April 2023

Abstract

The release of oil and petroleum products into the environment has long been a cause for concern due to its toxic effects on marine species. In particular, the Atlantic haddock has shown enhanced sensitivity towards the effects of crude oil. Much of the crude oil toxicity has been attributed to metabolites of polycyclic aromatic hydrocarbons (PAHs). While many PAHs undergo bioactivation, including phenanthrene, recent research has revealed that PAHs in isolation are unable to act with the same potency as PAHs in crude oil mixtures. This suggests that the contribution of PAHs in crude oil toxicity is more complex than what was initially believed.

This thesis aims to validate and apply a quantitative liquid chromatography tandem mass spectrometry method developed for the analysis of phenanthrene metabolites in Atlantic haddock embryos. Attention is also given towards measures for method optimization. The work performed throughout this thesis emphasises the use of rigorous statistical and chemometric approaches to obtain reliable and justifiable results.

Method validation was conducted by spiking non-exposed haddock eggs with standard solutions. The validation parameters included a response function, linearity, trueness, precision, accuracy, limit of detection, limit of quantitation and selectivity. The results demonstrated that the method had good quantification abilities for most of its analytes within the validated range.

The validated method was used to analyse samples from two exposure studies. In the first study, results showed that haddock embryos exposed to crude oil mixtures possess an enhanced rate of metabolite formation compared to PAH-only containing fractions. The second study revealed relative abundances of metabolites formed in haddock embryos after phenanthrene exposure. Quality control samples showed that the method provides accurate results within its validated range. However, the sample application study also revealed that the validated method failed to capture the relevant concentration range. Thus, while trends were visible, no quantification could be performed.

To optimize the method, experimental designs were employed on the electrospray ionization source. Fractional factorial designs were used to screen for relevant parameters. Screening revealed that ion source voltages could be improved. After voltage adjustments, a response surface was generated. The surface revealed that the method was seemingly still unoptimized, and that method response could be further enhanced by changing the gas temperature and flow in tandem. The optimized method was found to provide a considerably improved method response when compared to the initial results.

Both the method validation and sample application revealed that the method works well within its validated range, suggesting that the statistical approaches were fit for their purpose. Additionally, after method optimization, instrumental response was significantly improved. Therefore, although some further optimization may be necessary and a lower concentration level must be validated, the work performed in this thesis demonstrated that the proposed method has good potential for the analysis of phenanthrene metabolites in haddock embryos.

Sammendrag

Olje og andre petroleumsprodukter har lenge vært kjent for å forårsake skadelig effekter i det marine miljøet. Atlantisk hyse har vist seg å være spesielt sensitiv for giftigheten til råolje. Effektene til råolje har i stor grad blitt tilskrevet metabolittene produsert av polysykliske aromatiske hydrokarboner (PAHer). Det er godt kjent at mange PAHer, som fenantren, blir bioaktivert. Likevel har nylig forskning vist at PAHer i isolasjon ikke kan oppnå den samme giftigheten som PAHer funnet i råolje. Dette viser til at bidragene fra PAHer i råolje ikke er forstått så godt som opprinnelig antatt.

Målet med denne oppgaven er å validere og anvende en kvantitativ høypresisjonsvæskekromatografi tandem massespektrometrisk metode for analysen av fenantrenmetabolitter i egg fra Atlantisk hyse. Videre blir oppmerksomhet også rettet mot tiltak for metodeoptimalisering. Arbeidet i denne oppgaven vektlegger anvendelsen av rigorøse statistiske og kjemometriske metoder for å anskaffe pålitelige og forsvarlige resultater.

Metodevalideringen ble utført ved å tilsette standardløsninger til ikke-eksponerte hyseegg. Valideringen undersøkte en kalibreringsfunksjon, linearitet, riktighet, presisjon, nøyaktighet, deteksjonsgrenser, kvantiseringsgrenser og selektivitet. Resultatene viste at metoden har gode kvantifiseringsegenskaper innad dens validerte konsentrasjonsrekkevidde.

Den validerte metoden ble også anvendt på resultater fra to eksponeringsforsøk. I det første forsøket ble det funnet at hyseegg utsatt for råolje har en vesentlig høyere metabolittdannelse kontra de som kun er utsatt for PAHer. Det andre forsøket illustrerte den relative graden av metabolittdannelser i hyseegg utsatt for fenantren. Kontrollprøver viste at metoden gir nøyaktige resultater innenfor det validerte området. Likevel ble det funnet at prøvene fra eksponeringsforsøkene lå under de validerte konsentrasjonsområdene. Dermed, til tross for at trender kunne bli observert, var ikke metoden i stand til å kvantifisere eksponeringsprøvene.

For å optimalisere metoden ble det anvendt eksperimentelle design på ionekilden i massespektrometeret. Fraksjonelle faktorielle design ble anvendt for å søke etter relevante parametere. Resultatene viste at spenningene i ionekilden kunne endres for å øke signalstyrken. Etter optimalisering av spenningene ble det konstruert en responsflate. Modellen viste at metoden kunne bli videre optimalisert ved å endre gasstemperatur og gasstrøm i tandem. Den optimaliserte metoden hadde en betydelig økt signalstyrke sammenlignet med den opprinnelige metoden.

Resultatene fra metodevalideringen og prøveanvendelsene viste at metoden virker godt innenfor sitt validerte område. Dette viser at de anvendte statistiske teknikkene har vært egnet til sine formål. I tillegg, etter videre optimalisering, ble det funnet at metodens signalstyrke var betydelig forbedret. Dermed, til tross for at videre optimalisering sannsynligvis er nødvendig og at et lavere konsentrasjonsnivå må bli validert, har det utførte arbeidet vist at metoden har godt potensial til å kunne anvendes for analysen av fenantrenmetabolitter i hyseegg.

Contents

Abbreviations

AcN	Acetonitrile
AhR	Aryl Hydrocarbon Receptor
ANOVA	Analysis of Variance
AP	Accuracy Profile
CCD	Central Composite Design
CF	Correction Factor
CI	Confidence Interval
CV	Capillary Voltage
DEE	Diethyl Ether
ESI	Electrospray Ionization
FFD	Fractional Factorial Design
FOM	First Order Model
GF	Gas Flow
GT	Gas Temperature
GUM	Guide to the Expression of Uncertainty in Measurement
IP	Intermediate Precision
LC	Liquid Chromatography
LC-MS	Liquid Chromatography - Mass Spectrometry
LC-MS/MS	Liquid Chromatography - Tandem Mass Spectrometry
LoD	Limit of Detection
LoQ	Limit of Quantification
m/z	Mass-to-Charge Ratio
MeOH	Methanol

MS	Mass Spectrometry
MU	Measurement Uncertainty
NP	Nebulizer Pressure
NV	Nozzle Voltage
OHPHN	Hydroxyphenanthrene
OLS	Ordinary Least Squares
PAH	Polycyclic Aromatic Hydrocarbon
PI	Prediction Interval
r	Repeatability
RSD	Relative Standard Deviation
SPE	Solid-Phase Extraction
WLS	Weighted Least Squares

Symbols

α	False Positive Rate, Significance level, Between-Day Effect and Axial Point
β	False Negative Rate and Regression Parameter
σ	Population Standard Deviation
σ^2	Population Variance
μ	Population Mean
s	Sample Standard Deviation
s^2	Sample Variance
$\bar{x}$	Sample Mean
ν	Degrees of Freedom
n	The Total Number of Observations
I	Number of Levels and Number of Parameters
J	Number of Repeated Measurements per Level
b	Regression Parameter Estimate
Y	Dependent Variable
X	Random Variable and Independent Variable
$s_{y/x}$	Residual Standard Deviation
w	Regression Weight
r^2	Coefficient of Determination
r^2_{adj}	Adjusted Coefficient of Determination
SS_E	Error Sums of Squares
SS_M	Model Sums of Squares
SS_T	Total Sums of Squares
MS_E	Within-Level Variance
MS_M	Between-Level Variance
$u(r)$	Repeatability Standard Deviation
$u(b)$	Between-day Standard Deviation

$u(IP)$	Intermediate Precision Standard Deviation
u_c	Standard Combined Uncertainty
U	Expanded Measurement Uncertainty
p	p-value
$\hat{\mu}$	Spike Concentration
$\hat{\delta}$	Absolute bias
$\hat{\delta}_\%$	Relative bias
$R_\%$	Recovery
a	Sensitivity Coefficient
c_p	Correction Factor for Proportional Bias
c_b	Correction Factor for Constant Bias

1 Introduction

1.1 Background

Recent decades have seen some of the largest maritime oil spills to date, as well as their disastrous outcomes [1, 2]. Oil spills can have profound consequences on society in terms of economic and social impacts. Nonetheless, some of the largest ramifications have been observed in the environment [3–6]. In particular, sea-based oil spills, such as the Deepwater Horizon and Exxon Valdez, have shown particularly deleterious effects [1, 7, 8]. As oil situates itself on the water surface, ocean currents allow the dispersal of oil slicks, potentially allowing its area of effect to extend up to several hundred nautical miles [9, 10]. The fatal potential of crude oil has been well established; in the wake of previous spills, large numbers of seabirds, fish, shellfish and other marine organisms have succumbed to injuries related to oil exposure [1, 2, 7, 11]. Furthermore, the oil discharge is not solely reserved for accidental oil spills. Anthropogenic activities also introduce oil into the marine environment by means of urban wastewater, discharge from oil and gas operations, as well as natural seeps [12–14]. Understanding the toxicity of crude oil and other petroleum products is crucial due to their significant impact on marine life. Knowing the effects of oil releases is also necessary to develop effective responses to future spills and to minimize the impact of human-caused petroleum releases [15–18].

Much of the toxicity mediated by crude oil has been attributed to *polycyclic aromatic hydrocarbons* (PAHs) [19–21]. PAHs are natural constituents of the crude oil mixture, with the total PAH content being around 2% of the total oil mass [22]. PAHs native to the crude oil mixture are referred to as *petrogenic PAHs*. Although, in the environment, PAHs of *pyrogenic* orgin are also prevalent; these PAHs are by-products of incomplete combustion reactions and have been investigated in relation to tobacco, occupational exposure to exhaust fumes, road paving and more [14, 23, 24]. Toxicity has been demonstrated for many different PAH congeners, but it is perhaps benzo[*a*]pyrene that remains the most infamous example [1]. This PAH, *inter alia* formed in tobacco smoke, undergoes biotransformation *in vivo*. The resulting metabolite, a diol epoxide, is highly reactive and reacts with cellular macromolecules, such as DNA, ultimately initiating the onset of carcinogenesis [25–27]. The production of damaging metabolites during metabolism, so-called *bioactivation*, is a well known phenomenon. However, the effects of PAHs in marine organisms still remain under scrutiny. As the crude oil mixture is incredibly complex, attributing crude oil toxicity solely based on PAH content is likely an oversimplification [28–30]. Much of modern toxicology is concerned with *mixture effects*; the toxicity of a particular toxicant is often dependent on the presence of other chemicals in the mixture. Given this perspective, it is indeed plausible that the petrogenic PAHs may not be the main culprit for oil toxicity, and that their role is more intricate than previously believed [19, 30].

Recent research suggests previous theories of crude oil toxicity are incomplete. More focus is now on separating out the toxicities of PAHs from the overall crude oil mixture. [30]. A toxicological profile for a chemical can often be derived from its metabolic fingerprint [25]. In *metabolomics*, substrates, intermediates and other products formed during metabolism, are characterised. As these chemicals are closely related with cellular processes, metabolite characterization can

provide an overview of the physiological *status quo* of an organism [25, 31, 32]. Metabolomics requires advanced analytical methods in order to identify and quantitate the metabolites of interest [25, 32]. Nonetheless, while metabolomics is a well established area of research and extends into the analysis of PAHs, the metabolomic profiling of marine life exposed to PAHs remains scarce [33]. While several analytical methods have indeed been developed using liquid chromatography coupled with fluorescence detection or gas chromatography coupled with mass spectrometric detectors, these techniques have some considerable disadvantages. Most notably, the methods are often limited to semi-quantitative analysis, poor limits of detection and a limited range of metabolite polarities [33–38]. One proposed remedy to these problems is the development of analytical procedures using liquid chromatography coupled with tandem mass spectrometric detection (LC-MS/MS) [33, 34]. However, while the potential of LC-MS/MS has been acknowledged, literature describing its application for the task at hand are limited.

1.2   Polycyclic Aromatic Hydrocarbons

PAHs are a group of organic pollutants found in crude oil and as by-products of incomplete combustion reactions (Figure 1) [39–41]. These lipophilic compounds are characterised by their composition of two or more aromatic rings, giving rise to a planar parent structure, sometimes with incorporated heteroatoms, such as nitrogen, oxygen or sulphur [41]. For PAHs of petrogenic orgins, these parent structures often contain additional alkyl chains [41, 42].
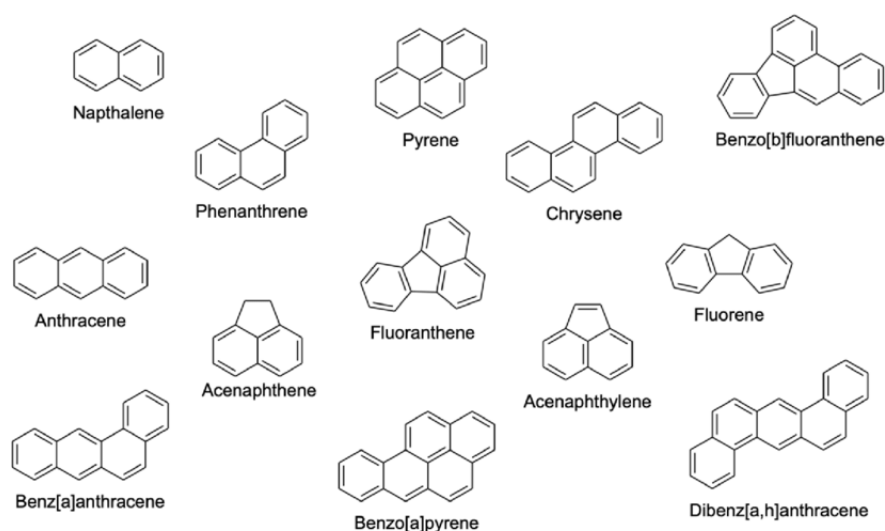
Figure 1: Selected polycyclic aromatic hydrocarbon structures [43].

PAHs pose a considerable issue in the marine environment due to their bioaccumulation in marine organisms [39, 44, 45]. The process of bioaccumulation occurs for all aquatic organisms, with the extent of the PAH uptake being governed by the level of exposure. Generally, the petrogenic PAHs possess both the greatest bioavailibility and toxicity [39]. The additional alkyl chains of the petrogenic PAHs render them more lipophilic than their pyrogenic counterparts. As a consequence, they tend to accumulate within the lipophilic environments of the sediment. Here, great quantities of PAHs can be found in pore water or adsorbed to surfaces [39, 45, 46].

This amplifies PAHs exposure for fish close to the sediment. Therefore, bottom-dwelling fishes, or *demersal fishes*, such as the haddock and cod, tend to experience the most considerable extent of PAH contamination. This is both due to the ingestion of contaminated sediment particles and the process of biomagnification, *i.e.* the up-concentration of toxicants via the food chain [39, 44, 45].

While most aquatic organisms are able to metabolise PAHs, the efficiency of this process is dependent on the metabolic system. Fish possess a particularly efficient metabolism, with 99% of the PAHs being metabolised within 24 hours of the uptake [39]. However, PAH toxicity is not a result of the initial parent compound, but rather the metabolites it produces *in vivo*. Therefore, paradoxically, the metabolic efficiency of fishes render them especially susceptible to the toxicity mediated by PAHs. The toxic effects can be seen as damage in the fish metabolism, tissue and physiological processes [47, 48]. This also renders fish exposed to PAHs potentially dangerous for human consumption [39, 49]. Thus, an understanding of PAH biomagnification and sediment distribution is of high priority due to its potential effects on both marine and human health [49–51].

1.3   Oil Toxicity in Atlantic Haddock

1.3.1   Atlantic Haddock

The Atlantic haddock (*Melanogrammus aeglefinus*) (Figure 2) is a dermersal gadoid with populations distributed across the Barents sea and both sides of North Atlantic. In the Eastern North Atlantic, the species is found from the coast of Northern Spain to Svalbard; in the Western North Atlantic, it stretches from the northeastern coast of the United States to Newfoundland [52].

Figure 2: Atlantic Haddock (*Melanogrammus aeglefinus*) [53].

The fish is found from 10 to 450 meters of depth and often at temperatures between 4 and 10ºC. Spawning usually occurs at a depth between 50 and 150 meters. The time of spawning depends on the localisation of the population; in the Northwestern Atlantic, spawning occurs between January and July, while in the Northeastern Atlantic, it occurs between February and June. The fecundity of the females is largely dependent on its size. For fish around 25 cm, the expected egg count is circa 55 000 which increases to 1 841 000 for corresponding specimen around 91 cm [52].

The Atlantic haddock is an important species for the Northern Atlantic fisheries, particularly in the Northeastern region. One of the most important spawning grounds is the Lofoten-Vesterålen area. This area contains fish populations, such as haddock and cod (*Gadus morhua*), that are of great interest for both economic and ecological reasons [46, 52]. In 2022, Norwegian fishing vessels had a total catch of 88 195 tons of haddock, yielding a catch-value of 1.6 billion NOK [54].

1.3.2   Oil Toxicity in Early Life Stage Atlantic Haddock

The Lofoten-Vesterålen area is not only of interest due to the spawning of economically important fish species; the waters outside these archipelagos are also of economic interest due to the possibilities of oil exploration. However, any area hosting oil exploration and production is associated with an enhanced risk for oil pollution [2]. This necessitates the development of risk assessments that characterises the toxicity experienced by proximate biota. In particular, much attention has recently been given to the toxic effects of crude oil in the Atlantic haddock [2, 30, 46, 55].

While Lofoten-Vesterålen is also a spawning ground for other fish species, such as Atlantic cod, the Atlantic haddock has been found to be especially vulnerable to crude oil toxicity during early life stages [2, 55]. As both species belong to the *Gadidae* family of marine fishes, the differences in toxicities is an interesting phenomenon. Recent research have proposed an explanation to these differences. As oil is introduced to the sea, it can be partitioned into micro-scale droplets. These droplets may be formed as a consequence of chemical dispersants used during oil clean-up procedures, or alternatively via wave energies. Sørhus et.al [55] have shown that, for the haddock embryos, its outermost membrane possesses adhesive properties which attracts oil micro-droplets. Ultimately, this facilitates the transport of crude oil constituents into the embryo, increasing toxicity. This hydrophobic membrane is termed the chorion, and its adhesive properties has been reported to be greater than those observed in similar species, such as the cod.

In relation to early life stage fish, *phenanthrene*, alongside its other tricyclic congeners (Figure 1), have received significant attention due to the cardiotoxic behaviour of their metabolites. While these PAHs in isolation do not have the same toxic potency as those in crude oil, it is important to understand the toxicity of the individual components in order to predict and understand mixture effects. Therefore, characterizing and describing the metabolites of isolated PAHs is vital in understanding the true mechanisms of crude oil toxicity [19, 30].

1.4   Objectives

The aim of this thesis is to apply rigorous statistical and chemometric approaches to provide a validated LC-MS/MS method for the analysis of phenanthrene metabolites in Atlantic haddock embryo. Additional emphasis is put on the ability of the method to provide statistically rigorous and readily applicable characteristics for future results.

The main objectives of the thesis are:

- To validate an analytical method for the analysis of phenanthrene metabolites in Atlantic haddock embryo.

- To demonstrate the validated method by its application to samples of Atlantic haddock embryos exposed to PAH mixtures.

- To optimize the method if the validation or application show that the method is non-satisfactory.

This work is a part of the EGGTOX project: "Unraveling the mechanistic effects of crude oil toxicity during early life stages of cold-water marine teleosts". This project was funded by the Research Council of Norway (project number 267820) and was conducted at the Institute of Marine Research. The results of this work will further develop the techniques for phenanthrene analysis, ultimately enhancing the knowledge of crude oil toxicity.

2 Theory

2.1 The Toxicology of Polycyclic Aromatic Hydrocarbons

Toxicology is an interdisciplinary scientific field, combining different branches of science such as biology, pharmacology, medicine and chemistry, to study the adverse effects of exogenous substances in living organisms. Due to the ever-growing scale of chemical production and unique chemicals, the science of toxicology is essential in modern society to ensure the health of both human and non-human species. Consequently, the quantitative and qualitative analyses of toxicants in living systems have become a research area of great interest for analytical chemists [25, 56].

2.1.1 Toxicokinetics of PAHs in Fish

Toxicokinetics studies the fate of an exogenous compound as it enters the body of an organism. Toxicokinetics is divided into four processes, namely absorption, distribution, metabolism and excretion [25, 57].

2.1.1.1 Absorption

In fish, PAHs may enter the body via all epithelial cells in contact with water and contaminated food. Uptake varies greatly across the different species, but well-perfused organs with large surface areas, such as the respiratory system and gastrointestinal tracts, generally provide the largest rate of absorption. The passage of the hydrocarbons into epithelial cells is facilitated by simple passive diffusion [58, 59].

Uptake is largely governed by the compound lipophilicity, commonly described by the logarithmic octanol-water partitioning coefficient (log K_{OW}) [57, 60]. Many of the PAHs containing two to six rings, have log K_{OW} values around 3-6 [60]. These compounds are sufficiently soluble in both water and fat and may thus enter the fish via the gills [59, 61]. PAH lipophilicity increases alongside the number of rings and eventually, around log $K_{OW} > 6$, absorption via gill epithelia becomes limited. For these substances, gastrointestinal uptake is the most important [59].

2.1.1.2 Distribution

Following absorption in epithelial cells, the PAHs are subsequently distributed to target organs throughout the fish. The distribution process is mediated by the circulatory system. The aqueous blood readily dissolves and transports hydrophilic xenobiotics. Hydrophobic compounds such as PAHs, however, navigate the circulatory system by binding themselves to macromolecules such as lipids or, more commonly, plasma proteins [57, 59].

The complex formed between the PAH and blood protein possesses a high molecular weight. This increase in mass makes it difficult for the xenobiotic to migrate out of the capillary. Thus, the release of PAHs occurs when the blood protein binds to a new molecule with a higher binding affinity [59, 62]. The final distribution is dependent on several factors including physicochemical properties, blood flow rate and rate of tissue migration [57]. The lipid composition of body tissue heavily influences the distribution of xenobiotics. In particular, the fatty tissue of the liver serves as the primary site of accumulation for most lipophilic xenobiotics [57, 59, 63].

2.1.1.3   Metabolism

After initial absorption, PAHs in the liver tissue can undergo *biotransformation*. The purpose of this process is to enhance water solubility of lipophilic PAHs, facilitating their elimination from the body [25]. The pathways of biotransformation are generally divided into two main phases: phase I and phase II metabolism [57, 59].

Phase I metabolism predominantly consists of oxidation reactions but also include reduction and hydrolysis reactions. The oxidation reactions are largely mediated by the *cytochrome P450* (CYP450) family of enzymes [25, 59]. The CYP enzymes show great plasticity and catalytic versatility, possessing the ability to create several metabolites from the same parent compound [25]. Most CYP enzymes are located in the endoplasmic reticulum of the hepatocytes. Here, the enzymes receive substrates which they transform into hydroxylated or other oxygenated metabolites [25, 64]. The oxygenated species from the CYP enzymes are often further metabolised by *epoxide hydrolase* enzymes, forming dihydrodiols [25].

Phase II metabolism are *conjugative reactions*, meaning they enlarge molecules via biosynthetic pathways mediated by relevant enzymes [25, 59]. The process of *glucoronidation* is among the most common conjugative processes. In this reaction, glucuronic acid is transferred from uridine diphosphate-glucuronic acid onto some exogenous substrate. This process is facilitated by uridine 5'-diphospho-glucuronosyltransferase (UGT). The UGT enzymes, similarly to the CYP enzymes, are also strongly expressed in the endoplasmic reticulum and competes with the CYPs for the lipophilic substrates. The UGTs can also further metabolise products from CYP oxidations [25, 64].

An overview of some of the processes from phase I and II metabolism is provided in Figure 3.
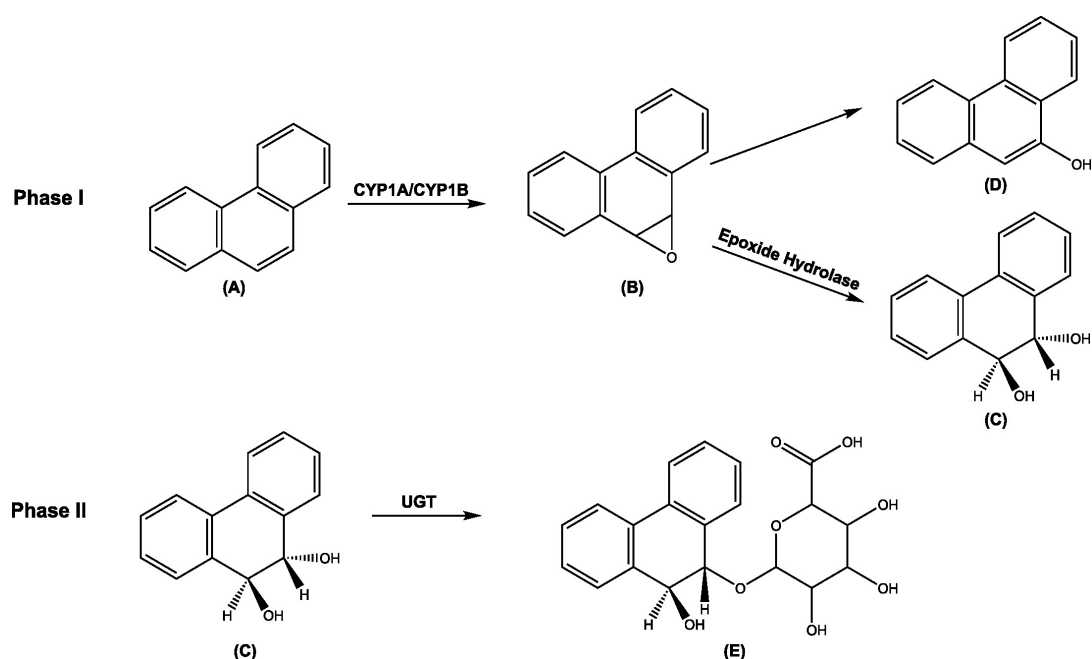
Figure 3: Selected biotransformation pathways for phenanthrene **(A)**. In phase I, the parent compound is oxidized by CYP enzymes to form an epoxide **(B)**. The resulting epoxide can either be converted into an alcohol **(D)** or a dihydrodiol **(C)** via epoxide hydrolase. In phase II metabolism, compounds from phase I can be conjugated with glucuronide via UGT enzymes **(E)** [65] (modified from source).

2.1.1.4   Excretion

After their polarity has been enhanced by biotransformation, the metabolites are subsequently excreted from the organism. There are various pathways for excretion [25, 57, 64]. In fish, the main routes of excretion are via the liver, kidneys and gills [66–68]. However, alternative routes, such as the eggs or scales, are also important [59]. The route of excretion is largely dependent on log K_{OW}. PAHs and their metabolites generally possess large values of log K_{OW} [59, 60], ensuring poor water solubility. As a consequence, the rather lipophilic metabolites are unable to enter the aqueous extracellular environments, preventing excretion via gills or kidneys. Therefore, for PAHs, the main route of excretion occurs via the liver [59, 68].

Excretion via the liver does not require the PAH metabolites to leave the cell [25, 57]. Rather, the metabolites are transported to *bile canaliculi*, which are tubules found in the center of adjacent hepatocytes [57]. The transport from the hepatocyte into the tubule is often termed *phase III* of metabolism [57]. The *ATP-binding cassettes* (ABC) transporters and *solute carrier proteins* (SLC) are the most important families of proteins involved in the transport of metabolites into bile. The ABC family is energy-dependent and utilises ATP to facilitate the transport of xenobiotics. SLC transporters, however, are energy independent and use differences in electrochemical potentials to expel metabolites from the hepatocytes [25, 57]. The bile canaliculi empties the metabolites and other bile constituents into the gallbladder. Here, they are temporarily stored until the organisms finally excrete bile via fecal matter [57, 69].

2.1.2   Aryl Hydrocarbon Receptor and PAH Mediated Toxicity

Phase I, II and III of metabolism are regulated by *xenosensors*, a group of ligand-activated receptors which initiate adaptive responses in organisms [25]. Generally, once bound to a ligand, the xenosensors form transcriptional complexes that enter the nucleus and transcribes xenobiotic-response elements. These response elements often code for enzymes or other proteins that initiate some biological process of interest [25, 57]. One of the most important xenosensors in toxicology, is the *aryl hydrocarbon receptor* (AhR) [25, 64]. This xenosensor is of particular interest for planar molecules, such as the PAHs, and other crude oil constituents [30, 70, 71]. Once bound to its substrate, the AhR forms a transcriptional complex by binding to the aryl nuclear transferase and other ancillary proteins. As it enters the nucleus, the complex binds to specific sequences in the DNA and enhances transcription, producing a range of relevant response proteins. This process of increasing DNA transcription and protein production is called *induction* [25].

Enzyme induction is an adaptive response of the biological system that improves the rate of xenobiotic detoxification and elimination [25]. For PAHs, the transcriptional complex with AhR induces the CYP1A and CYP1B enzymes, enhancing the organism's oxidative processes [57]. However, for several PAHs, the CYP1A/CYP1B enzymes actually produce metabolites that are more toxic than their parent congeners. One example is benzo[*a*]pyrene, which undergoes repeated metabolism via CYP1A/CYP1B and epoxide hydrolase enzymes, ultimately producing a reactive dihydrodiol epoxide. This metabolite possesses strong electrophilic character and thus a strong corresponding affinity for the nucleophilic amino acids of DNA [71, 72]. As the metabolite binds to DNA, errors in DNA transcriptions are introduced. Depending on the location and magnitude of errors, these DNA adducts can potentially initiate the onset of cancer [25].

2.2   Analytical Techniques and Instrumentation

2.2.1   Chromatography

2.2.1.1   Separation Theory and Liquid Chromatography

In analytical chemistry, analytes of interest are often found within complex matrices and must thus be separated from the surrounding interfering compounds by a separation technique. One such technique is *chromatography*. Here, separation is achieved by a partitioning-equilibrium; analytes are distributed between two phases, the *mobile* and the *stationary phases*. In the chromatographic process, a sample is dissolved into the mobile phase, which moves through a column containing the fixed stationary phase. The mobile phase is commonly in the form of a liquid or gas, whereas the stationary phase usually consists of small particles or a liquid [73, 74].

Chromatographic separation is dependent on the chemistry of the sample components, mobile phase and stationary phase. A large number of separation modes can be applied to achieve adequate separation. One of the most commonly applied separation modes is *reverse-phase chromatography*. In this separation mode, the mobile phase is more polar than the stationary phase. For reverse-phase systems, due to their similar polarities, apolar compounds will be absorbed by the stationary phase and be more *retained* than less polar counterparts. This is in contrast to *normal-phase chromatography*, in which the stationary phase is more polar than the mobile phase [73, 74].

The extent to which an analyte is retained is commonly described by the *retention factor*, k:

$$k = \frac{t_r - t_m}{t_m} \tag{1}$$

Where t_m is the time required for an *un*retained compound to *elute* (leave) the chromatographic system, and t_r is the elution time for the analyte.

The quality of a separation can be summarized via the chromatographic *resolution*, R_s:

$$R_s = \frac{t_{r,b} - t_{r,a}}{\frac{1}{2}\left(w_b + w_a\right)} = 0.589 \cdot \frac{t_{r,b} - t_{r,a}}{\frac{1}{2}\left(w_{b,h} + w_{a,h}\right)} \tag{2}$$

Where $t_{r,a}$ and $t_{r,b}$ denote the retention times for two adjacent peaks, a and b, with $t_b > t_a$. Moreover, w denotes the corresponding peak widths measured at the baseline, and w_h is the width of the peak measured at half of its height. Commonly, resolution is determined sufficient if *baseline resolution* is achieved, this occurs when $R_s \geq 1.5$ (Figure 4) [73].
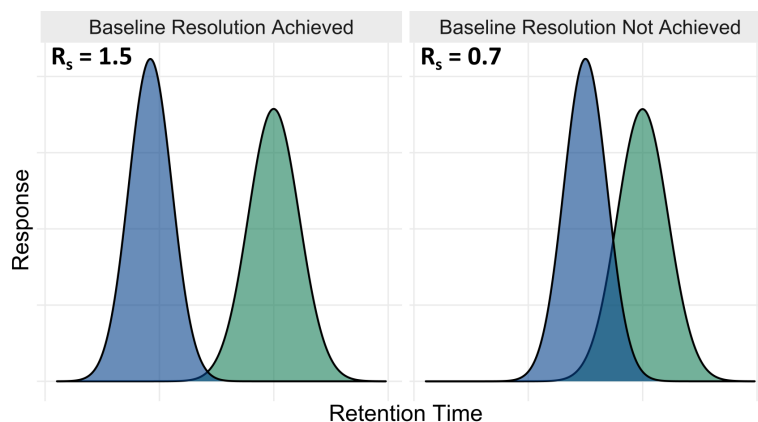
Figure 4: Resolution of chromatographic peaks.

The resolution between two adjacent peaks is determined by three unique variables: the *plate number*, N, the *selectivity*, α and the *retention*, k_b:

$$R_s = \left(\frac{\sqrt{N}}{4}\right)\left(\frac{\alpha - 1}{\alpha}\right)\left(\frac{k_b}{1 + k_b}\right) \tag{3}$$

The value of N is largely dependent on the column in which chromatographic separation takes place. In *liquid chromatography* (*LC*), a liquid mobile phase is forced through a narrow column containing a stationary phase consisting of small particles. It can be shown that N is inversely proportional with the particle diameter (d_p):

$$N \propto \frac{1}{d_p} \tag{4}$$

Therefore, a smaller particle size will always provide better resolution. However, columns with small particles come at the cost of high pressures. The backpressure (P) required to force a liquid through the column is related to particle size via

$$P \propto \frac{1}{d_p^2} \tag{5}$$

Thus, despite an improved chromatographic resolution, decreasing the particle size is not always feasible due to the required backpressure.

The retention, k_b, describes how efficiently the analytes are retained by the stationary phase. This component can be manipulated by changing the mobile phase, stationary phase and temperature.

Finally, the selectivity, α describes the relative retention between the adjacent peaks:

$$\alpha = \frac{k_b}{k_a} \tag{6}$$

This component is affected by the same parameters as k_b. Selectivity, and thus resolution, can be enhanced by choosing mobile and stationary phases which maximizes the difference between k_a and k_b.

Equation (3) is only valid if the mobile phase composition is constant. This is often not the case (see below). Nevertheless, the three variables in the equation (N, α, and k) are always crucial factors that need to be optimized in order to achieve good separations [73, 74].

2.2.1.2   Instrumentation in Liquid Chromatography

In LC, a liquid mobile phase is forced through a column packed with the stationary phase. As the stationary phase consists of small particles, the mobile phase requires high pressures to pass through the column. Instruments which are able to use columns with particles smaller than 2 µm are typically referred to as *ultra-high performance liquid chromatography* instruments. Conversely, instruments which operate with particle sizes above 2 µm are referred to as *high performance liquid chromatography* instruments. Regardless of particle size, a typical LC can roughly be divided into the solvent delivery module, injection valve, column and detector (Figure 5) [73, 74].
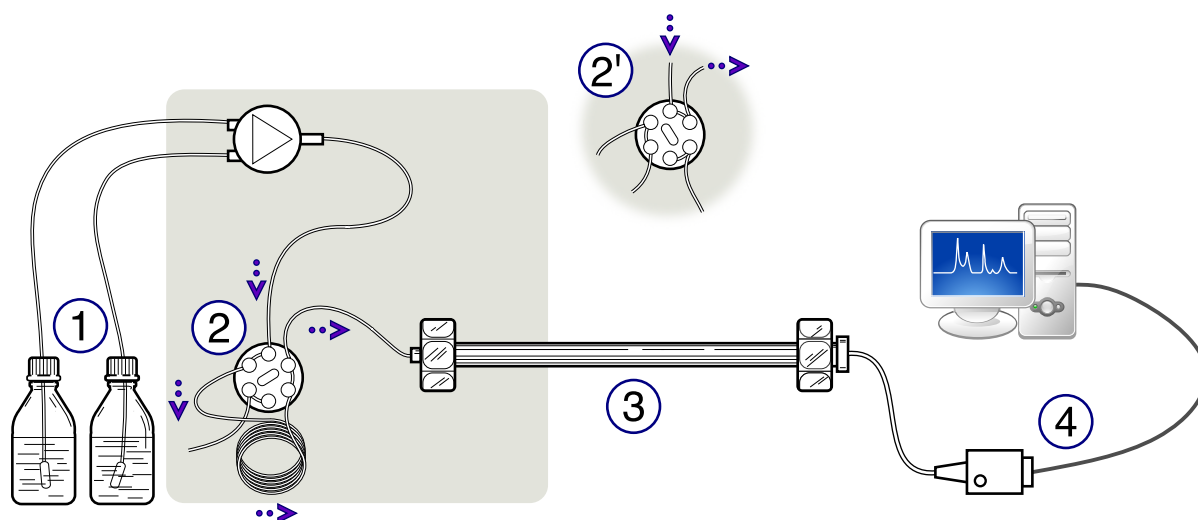
Figure 5: Schematic of an LC instrument. The figure shows the flow path of solvents and analytes through the system. The solvent delivery module (**1**) uses a pump to deliver the mobile phase to an injection valve (**2**), where samples are introduced into the flow stream. The solvent and sample are then pushed through the analytical column (**3**) and ultimately reach the detector (**4**), which produces interpretable signals. Point (**2′**) denotes the injection valve in its loading position; the sample loop is detached from the high-pressure system to allow for sample injection [75] (modified from source).

The main components of the solvent delivery module is the solvent reservoir and pump. The *solvent reservoir* is comprised of two solvents; an *aqueous* and an *organic phase*. Respectively, these are often denoted as solvent A and B. The *pump* is responsible for transporting the solvents from the reservoirs to the column at a smooth and consistent flow rate. Solvents A and B are commonly mixed. If the mixing ratio between A and B is constant during the analysis, the LC program is said to run in *isocratic* mode. If the ratio between A and B changes throughout the

program, the LC runs in *gradient* mode. Most LC programs use gradient elution as these are often able to obtain enhanced separation and peak shape compared to isocratic elution [73, 74].

The sample enters the instrument via the *injection valve*, which can be adjusted into two positions. In the first position, sample solution is injected into the injection port, where it enters a sample loop separated from the column and high-pressured solvent. In the second position, the sample loop becomes a part of the high-pressured system and is pushed into the column via the solvent. The analytes then interact with the column and mobile phase before reaching the detector. There exists a wide array of detection techniques for modern LC equipment. The most powerful detector is the mass spectrometer [73].

2.2.2   Mass Spectrometry

2.2.2.1   Instrumentation in Mass Spectrometry

Mass spectrometry (MS) is an analytical technique which generates and separates ions by their *mass-to-charge ratio* (m/z). The MS is comprised of four distinct units, the sample inlet, ion source, mass analyser and detector [73, 76].

The *sample inlet* is the component responsible for introducing the sample to the MS instrument. Often, in so-called *hyphenated techniques*, the inlet is connected to some separatory technique. The most common of which are gas and liquid chromatography, in which case the techniques are abbreviated GC-MS and LC-MS [76].

After introduction, the sample enters the *ion source*. In this chamber, analyte molecules are ionized via a selected ionization technique. For systems coupled with an LC interface, the most popular technique is *electrospray ionization* (ESI) [73, 76]. This technique nebulizes the sample solution via the application of a strong electric field, creating an aerosol of charged droplets. These droplets are then gradually evaporated by the application of a drying gas (typically N_2). As the surface area shrinks, charge density increases, ultimately tearing the droplets apart due to electrostatic repulsion. This produces new droplets that repeat the process. Each iteration reduces droplet size until the charge finally attaches to the analyte. The final charged analyte is often formed by the removal of a hydrogen ion (H^+). Alternatively, analytes can also obtain charge by bonding to ions, such as a hydrogen, sodium (Na^+) or acetate (CH_3COO^-). If analytes gain charge via analyte bonding, the charged complex is called an *adduct*. Moreover, if the ESI produces ions of negative charge, it is said to operate in *negative mode*. Similarly, if the ESI provides positive ions, it is run in *positive mode* [77].

After ionization, the analyte molecules are subsequently separated via the *mass analyser*. A large number of mass analysers are available, all of which utilize electric and/or magnetic fields to manipulate ion trajectories. One of the most popular mass analysers is the triple quadrupole (*QqQ*) [73, 76]. This mass analyser uses three linear quadrupoles in sequence, each of which consists of four (or more) parallel pairs of metal rods. Each of these pairs possess a joint electrical connection which is used to apply an alternating (AC) and a direct current (DC). The first quadrupole in the sequence (Q_1) separates ions based on their m/z. This is done by adjusting

the AC and DC voltages such that only a particular *m/z* has a stable trajectory and may enter the next quadrupole. The second quadrupole (Q_2) does not apply a DC to its rods, only the AC. This allows all ions from Q_1 to freely pass. Moreover, Q_2 is also filled with a collision gas (*e.g.* N_2), fragmenting the larger ions from Q_1, so-called *parent ions* (or precursor ions), into smaller *daughter ions* (product ions). The fragments then enter the third quadrupole (Q_3) where ions are selected analogously to Q_1 [76, 78].

The ions passing through the mass analyser reach the *detector*. Commonly, an electron multiplier or a related detection technique is applied. As they enter the detector, ions collide with electron multiplier plates, each hit releases electrons. This produces a cascade effect; each released electron is free to collide with a new multiplier plate, quickly enhancing the electric potential. When the electrons reach the end of the detector, they hit a detector plate which converts the detected current into interpretable signals. The final signals are typically presented as a *mass spectrum*. Here, the signal intensity (often normalized to the tallest peak) is plotted alongside the *y*-axis and the *m/z* along the *x*-axis [73, 76].

2.2.2.2   Matrix Effects in ESI

LC-MS is particularly sensitive to *matrix effects*, especially when ESI is applied as the ion source. Matrix effects arise when the analyte ionization is altered by coeluting molecules. *Ion suppression* occurs when the matrix components decrease the ionization efficiency of the analyte; reducing the instrumental signal. This can happen by various mechanisms such as the formation of multiple adducts, the binding of analytes onto matrix components, or the formation of ion clusters that interfere with the ionization process [79, 80].

Alternatively, the ionization efficiency can be increased by *ion enhancement*. This can happen due to the suppression of competing reactions, the formation of adducts with the analyte, or the alteration of the ionization process itself [79, 80].

2.2.3   Solid-Phase Extraction

Solid-phase extraction (SPE) is a widely used sample preparation technique in which samples are concentrated and purified for downstream analysis. The principles of solid-phase extraction are largely similar to those of LC; a sample mixture is separated into smaller components on the basis of a partitioning-equilibrium between the liquid mobile phase and some stationary phase [73].

SPE separates components by pulling a sample solution through disposable cartridges containing a sorbent. The eluate is collected in a vial. If a cartridge has affinities for the impurities, then the eluate contains the analytes of interest. In this case, the cartridge is discarded. If the analytes of interest are retained on the stationary phase, the initial eluate contains only impurities and is discarded. The analytes can then be eluted into a new vial by the addition of an appropriate solvent [73, 81].

2.3   Concepts in Statistics and Chemometrics

The following sections uses specific statistical and mathematical notation:

- $X \sim N(\mu, \sigma)$ is read "*The random variable X is drawn from a normal distribution with mean μ and standard deviation σ*".

- $Pr(X)$ denotes the probability of event X occurring.

- In general, capital letters (*e.g. T* and *Z*) denote random variables whereas lowercase letters (*e.g. t* or *z*) are realized values of a particular random variable (*i.e.* the value of the random variable after it has been drawn).

- Greek letters (*e.g.* μ and β) imply *population parameters* whereas lowercase latin letters (*e.g.* $\bar{x}$ and *b*) imply *sample estimates*.

- Lowercase letters in bold refer to the use of vectors, whereas uppercase letters in bold refer to matrices.

- All vectors are assumed to be column vectors.

- The -1 and T superscripts denote the matrix inverse and vector/matrix transpose, respectively.

2.3.1   The Normal Distribution and the *z*-test

Measurements in the physical sciences are inherently random [73, 82–84]. Consequently, all measurements may be thought of as a random variable, *X*, following a particular probability distribution. The most important of which being the *normal distribution* [73, 82]. This distribution is characterized by two parameters, the population mean (μ) and standard deviation (σ). The former parameter describes the distribution center and the latter, the dispersion (Figure 6). Moreover, the standard deviation is also often presented in its squared form (σ^2) in which case it is called the *variance* [82, 85].
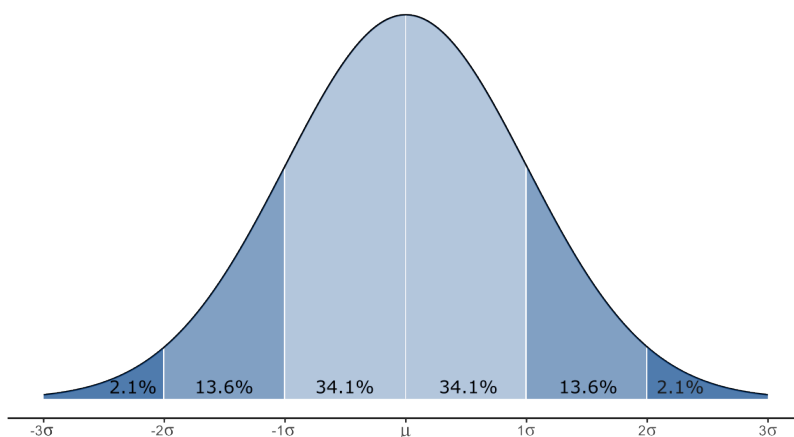
Figure 6: The Normal Distribution.

Having generated a sample, it is often of interest to compare the collected data against one or more parameters of some suspected probability distribution [73, 85]. For the normal distribution, the parameter of interest is often μ [86].

Consider a sample consisting of n instances of the random variable $X \sim N(\mu, \sigma)$. Furthermore, let σ be known and μ unknown. An estimate of μ is then given by $\bar{x}$:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{7}$$

The standard deviation of $\bar{X}$ is called the *standard error of the mean*, $\sigma_{\bar{X}}$, and is given by

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \tag{8}$$

Let the suspected distribution have a mean of μ_0. To investigate if the sample mean ($\bar{x}$) agrees with the suspected distribution, two contradictory hypotheses are proposed, the null (H_0) and alternative (H_a) hypotheses. In two-tailed tests, the hypotheses are expressed as

$$H_0 : \mu = \mu_0$$

$$H_a : \mu \neq \mu_0$$

Whereas in one-tailed tests, the hypotheses may be written as

$$H_0 : \mu \leq \mu_0 \qquad\qquad H_0 : \mu \geq \mu_0$$
$$\text{or}$$
$$H_a : \mu > \mu_0 \qquad\qquad H_a : \mu < \mu_0$$

The hypotheses are evaluated by the z-statistic:

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{X}}} = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} \tag{9}$$

If H_0 is correct, then $Z \sim N(0, 1)$. Therefore, if z differs considerably from 0, there is evidence suggesting that H_0 is false and may be rejected in favour of H_a. Nonetheless, as Z is a random variable, it is possible that some extreme deviation from 0 occurred by chance. Therefore, to evaluate if there is some difference between $\bar{x}$ and μ, the probability of Z must be further investigated [85].

The p-value (p) is the probability of observing some value of Z at least as extreme as its realized instance found from the sample data (z), assuming H_0 is true [85, 86]. That is,

$$p = Pr(|Z| \geq |z| \,|\, H_0) \tag{10}$$

If the p-value falls below a predefined *significance level*, α, the results are deemed *statistically significant*; H_0 is then rejected in favour of H_a.[1] Given some z, its p-value can readily be obtained using tabulated values. The relation between α and the types of tests is shown in Figure 7. Moreover, it is also possible to determine significance by *quantiles*. The z_γ^* quantile denotes the point in the normal distribution at which a proportion γ of all measurements are expected to fall below (or equivalently, a proportion $1 - \gamma$ falls above). Therefore, if z is more extreme than $z_{1-\alpha}^*$, the results of a one-sided test may be deemed significant. Equivalently, if z is more extreme than $z_{1-\alpha/2}^*$, the results of a two-sided test are significant [85, 86].

The value of α also represents the risk of mistakenly rejecting a true H_0. This is called a *false positive* or *type I error*. A *false negative* or *type II error* is the probability, denoted β, of mistakenly failing to reject H_0 [86].
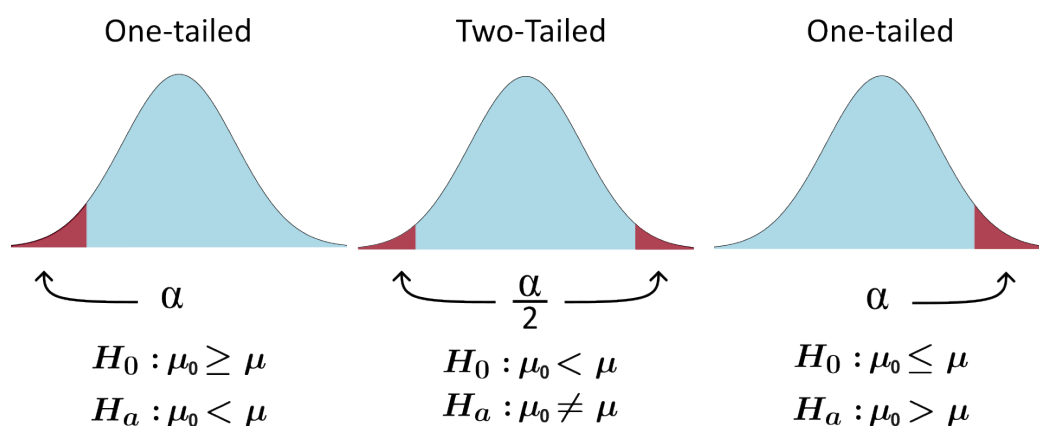
Figure 7: The different pairs of hypotheses for z- and t-tests. The red-shaded region denotes the rejection region, *i.e.* the values of $\bar{x}$ which provide a p-value lower than α. For the one-tailed tests, significance only occurs on one end of the curve. For instance, for the one tailed test on the left-hand side of the figure, only extreme results that are lower than μ may be deemed significant; extreme results greater than μ cannot warrant significance. The opposite holds true for the corresponding right-hand test. For the two-tailed test, as it accounts for both sides of the distribution, there are two rejection regions, each possessing an area of $\alpha/2$ which together sum to α.

2.3.2   Student's t-Test

As shown in the previous section, statistical inference via the z-test requires the knowledge of σ. However, since σ is seldom known, it must be estimated from the sample data. This estimate, the *sample standard deviation* (s), is given by

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} \tag{11}$$

Where the denominator, $n - 1$, is known as the *degrees of freedom*, ν. The value of ν represents the number of independent observations in a sample that are available for estimating a statistical parameter.

---

[1]NB: While a low p-value provides evidence suggesting that H_0 is false, a high p-value does not imply that H_0 is true. Per definition (10), the p-value is already conditioned on H_0. Therefore it may *not* be used to infer that H_0 is true.

The *Student's t-test* is a rephrasing of the *z*-test, accounting for an unknown σ. The associated test statistic, t_ν, is obtained by replacing σ with its estimate in (9):

$$t_\nu = \frac{\bar{x} - \mu_0}{s / \sqrt{n}} \tag{12}$$

The random variable T_ν follows the *t*-distribution. This distribution, characterized by ν, is similar to the normal distribution. However, the *t*-distribution has fatter tails,[2] which makes it more probable to obtain extreme results. The tails shrink as ν increases. In essence, this allows the *t*-distribution to account for the uncertainty in s arising due to a low sample size [85].

As was the case with z, the significance of t_ν may either be determined via *p*-values or quantiles. Moreover, as ν increases, the values of $t^*_{\gamma,\nu}$ converge towards those of z^*_γ [85, 86].

2.3.3   Least Squares Regression Analysis

2.3.3.1   Ordinary Least Squares

Much of modern analytical chemistry depends on instrumental analyses. This type of analysis often requires the usage of a *response function* relating standard solutions of known concentrations to corresponding instrumental response. This allows quantification of unknowns by relating their response to an associated concentration via interpolation [85, 86]. In this case, the instrumental response is a *dependent variable* as its value is determined by the concentration, which is the corresponding *independent variable*.

For many calibration purposes, a dependent variable (Y) is related to an independent variable (X) via

$$y = \beta_0 + \beta_1 X + \varepsilon \tag{13}$$

Where β_0 and β_1 are coefficients describing the relationship between the X and Y. The final term, the model *error*, is a random variable representing experimental noise, $\varepsilon \sim N(0, \sigma_\varepsilon)$ [73, 85]. When the independent variable(s) only has a power of one, the model is often called a *first-order model* (FOM). Moreover, Equation (13) can be generalized to account for any number of independent variables [87, 88]. Thus, for a total of n data points and I independent variables:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{14}$$

Where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{1,1} & \cdots & x_{1,I} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & \cdots & x_{n,I} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_I \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

---

[2]Formally, it is said that the t-distribution has greater *kurtosis* than the normal distribution

Equation (14) is commonly solved by *ordinary least squares* (OLS) regression. In OLS, a regression model is generated such that the sum of the squared differences between the experimental (y) and predicted ($\hat{y}$) values are as low as possible (Figure 8) [85, 88]. This sum, the *error sums of squares* (SS_E), is defined as

$$SS_E = \sum_{k=1}^{n}(y_k - \hat{y}_k)^2 = (y - Xb)^T(y - Xb) \tag{15}$$

Solving for the minimum of Equation (15) yields estimates (b) of the true coefficients (β) [85, 88]:

$$b = [X^TX]^{-1}X^Ty \tag{16}$$
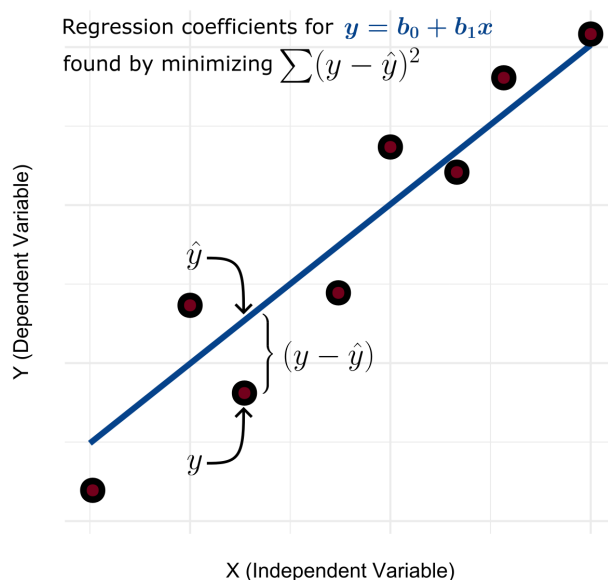
Figure 8: The method of least squares for $y = b_0 + b_1x$.

OLS regression assumes *homoscedasticity*. That is, the variance in y, *i.e.* the experimental noise, as determined by σ_ε, is independent of X. An estimate of σ_ε is given by the residual standard deviation, $s_{y/x}$ [85]:

$$s_{y/x} = \sqrt{\frac{SS_E}{n - I}} \tag{17}$$

2.3.3.2   Weighted Least Squares

If the assumption of homoscedasticity is violated, the data is *heteroscedastic*. In this case σ_ε is dependent on X; changes in the independent variables alters the variance in the dependent variable [85, 89]. To remedy the problem of heteroscedasticity, *weighted least squares* (WLS) regression is applied. WLS builds upon the same equations as OLS, but with the incorporation of a *weighting term*, w. In most cases of heteroscedasticity, the variance in y is assumed to increase alongside X. Consequently, smaller values of X carry the most information (due to less

uncertainty) and are given more *weight* during the estimation of the parameters, β [89]. WLS seeks to minimize [89, 90]

$$SS_{E,w} = \sum_{k=1}^{n} w_k(y_k - \hat{y}_k)^2 = w(y - Xb)^T(y - Xb) \tag{18}$$

The parameter estimates are then given by [87, 89]

$$b = [X^T w X]^{-1} X w y \tag{19}$$

2.3.3.3 The Coefficient of Determination

It is often of interest to investigate the goodness-of-fit between the regression model and the data. One such descriptor is the *coefficient of determination*, r^2. In OLS, r^2 denotes the fraction of data variance explained by the regression model. Thus, an r^2 close to one indicates a good model fit [85, 88]. The statistic is estimated by

$$r^2 = \frac{SS_M}{SS_T} = 1 - \frac{SS_E}{SS_T} \tag{20}$$

Where SS_T denotes the *total sums of squares* and describes the total variance of the dependent variable. Furthermore, SS_M, the *model sums of squares* quantifies the variance explained by the regression model. These may be calculated via

$$SS_T = \sum_{k=1}^{n} (y_k - \bar{y})^2 = (y - \bar{y})^T(y - \bar{y}) \tag{21}$$

and

$$SS_M = \sum_{k=1}^{n} (\hat{y}_k - \bar{y})^2 = (\hat{y} - \bar{y})^T(\hat{y} - \bar{y}) \tag{22}$$

The three types of sums of squares are related in accordance with [85]

$$SS_T = SS_E + SS_M \tag{23}$$

However, the r^2 can easily be inflated by additional predictors. Therefore, for models with a larger number of independent variables, the r^2 is commonly modified to provide the *adjusted coefficient of determination*, r^2_{adj}:

$$r^2_{adj} = 1 - \frac{SS_e}{SS_T} \frac{n-1}{n-(I+1)} \tag{24}$$

The value of r^2_{adj} is penalised for each additional regressor. Thus, a value of r^2_{adj} much lower than r^2 suggests that the model is overfitted [85].

For WLS, there is no universal method for calculating r^2 or r^2_{adj}. Consequently, different statistical software and approaches can give different estimates [91].

2.3.3.4   The General Linear Hypothesis

The error sums of squares can also used for the *general linear hypothesis*. In this test, a proposed *full* regression model, $Y = \beta_0 + \sum_{i=1}^{I} \beta_i + \varepsilon$, is compared to a *reduced* counterpart which contains only the intercept, $Y = \beta_0 + \varepsilon$. The hypotheses become

$$H_0 : \beta_1 = ... = \beta_I = 0$$

$$H_a : \beta_i \neq 0, \text{for at least one } i$$

If H_0 is true, the additional regressors do not improve the model. Thus, the residual variances for both models should be equal. However, if H_0 is false, then at least one of the regressors in the full model improves model fit. The test statistic becomes

$$F_{\nu_1,\nu_2} = \frac{SS_E(R) - SS_E(F)}{\nu_R - \nu_F} \div \frac{SS_E(F)}{\nu_F} \tag{25}$$

Where $SS_E(R)$ and ν_R is the error sums of squares and its associated degrees of freedom for the reduced model; $SS_E(F)$ and ν_F denote the corresponding parameters for the full model.

The statistic shown in Equation (25) follows the *F-distribution*. Similarly to the *t*-statistic, the *F*-statistic is also dependent on degrees of freedom. One is associated with the variance in its numerator, $\nu_1 = \nu_R - \nu_F$, and one with the variance in its denominator, $\nu_2 = \nu_F$ [92, 93].

2.3.4   One-Way Analysis of Variance

The *analysis of variance* (ANOVA) is a set of statistical techniques in which observed variances are partitioned into individual sources of dispersion. Most commonly, the ANOVA is applied as a generalization of the *t*-test in which the differences between three or more group means are assessed by their variances [85, 89].

In the ANOVA, samples are characterized by unique *levels* of some independent variable, termed a *factor*. For instance, if the experimental factor is exposure duration, the levels could correspond to different exposure times. The variance within each level is then calculated and compared to the variance between the levels (Figure 9) [85].
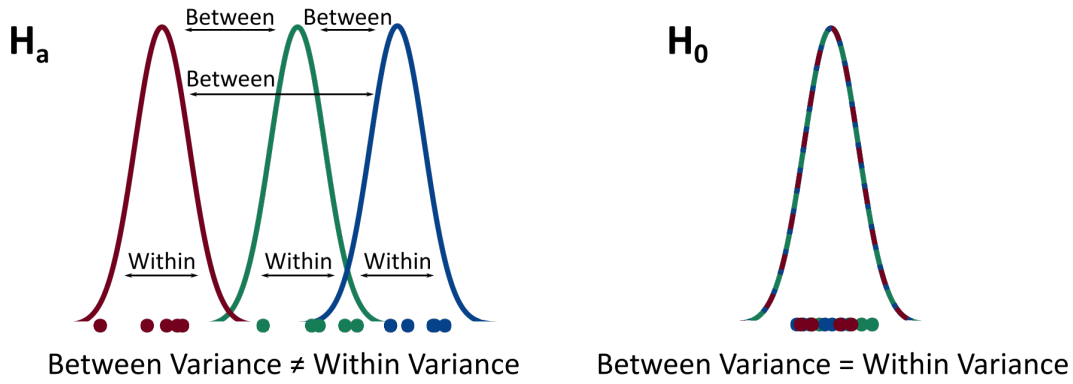
Figure 9: The ANOVA hypotheses. For H_a, at least one of the levels possess a unique μ_i. In this figure, the three different levels (as distinguished by their colors) all possess characteristic instances of μ_i. This means that between-level variance differ from that of the within-level variance. For H_0, all levels are drawn from the same distribution, meaning they all share the same μ_i. In this case, the between variance equals the within variance.

In the simplest case of ANOVA, only one single factor is under study and all factor levels possess the same number of measurements. This is called the balanced *one-way ANOVA*. If the different levels of the factor are drawn from different populations, the variance between the levels will be larger than the variance within (H_a). However, if all levels follow the same population, then the between and within variances are equal (H_0) [85]. The hypotheses may be expressed as

$$H_0 : \ \mu_1 = \mu_2 = ... = \mu_I$$

$$H_a : \ \text{At least two of } \mu_i\text{'s are different}$$

Where μ_i denotes the true mean for some level i. The ratio of the variances can be compared in an *F*-test. The ANOVA test statistic becomes

$$F_{\nu_1, \nu_2} = \frac{\text{Between Variance}}{\text{Within Variance}} = \frac{SS_M/(I-1)}{SS_E/I(J-1)} = \frac{MS_M}{MS_E} \tag{26}$$

Where SS_M is the *model sums of squares*, SS_E is the *error sums of squares*, and J is the number of replicate measurements, which is constant across all I levels. Moreover, MS_M and MS_E are estimates of the between and within variances, which respectively possess $\nu_1 = I - 1$ and $\nu_2 = I(J-1)$ degrees of freedom. The *F*-statistic can be compared to tabulated values to determine whether one or more levels are statistically significant [85, 89].

The ANOVA is often expressed via the equation

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{27}$$

Where μ represents the mean across all I levels and serves as a baseline response, α_i is the effect of level i, *i.e.* its deviation from the baseline, and ε_{ij} is the noise for the jth measurement at level i [85].

The ANOVA is a special case of OLS in which the predictors are qualitative rather than quantitative. Therefore, by defining $\mu_i = \mu + \alpha_i$, Equation (27) can be re-expressed as

$$Y_{ij} = \mu_1 X_{ij1} + \mu_2 X_{ij2} + ... + \mu_I X_{ijI} + \varepsilon_{ij} \tag{28}$$

Using Equation (28), the ANOVA F-test can then be conducted by the general linear hypothesis procedure (Section 2.3.3.4) [92, 94].[3]

2.3.5   Experimental Design

2.3.5.1   Factorial Designs

The *design of experiments* is concerned with the process of planning, conducting and analysing experiments. In analytical chemistry and chemometrics, one of the most popular exploratory designs is the *factorial design* [95]. This design screens for factors that alter some chosen response variable.

Factorial designs evaluates pre-selected factors at a high and low level. Typically, instead of using their numerical values, the high and low factor levels are coded as +1 and −1, respectively. The center point (coded 0) represents the average level of each factor, and the coded values of +1 and −1 represent a distance of one unit from the center point in the positive and negative directions.

If all factors and their interactions are of interests, a *full factorial design* is applied. In this case, if I factors are investigated, a total of 2^I experiments must be conducted. The effects of the factors and their interactions are estimated by the OLS techniques presented previously [95–97]. Consider, a 2^3 full factorial design. The *design matrix* becomes

$$
\begin{array}{cccccccc}
X_0 & X_1 & X_2 & X_3 & X_{12} & X_{13} & X_{23} & X_{123}
\end{array}
$$
$$
\begin{bmatrix}
1 & -1 & -1 & -1 & 1 & 1 & 1 & -1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & 1 & -1 & 1 & -1 & 1 & -1 & -1 \\
1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

---

[3]An example is provided in Appendix A.2

In the matrix above, X_0 refers to the estimation of β_0, X_1 to β_1 and so forth. Moreover, subscripts containing two numbers refer to interaction terms (*e.g.* β_{12} is the regression coefficient of the interaction between factors 1 and 2). Note that all the columns are orthogonal;[4] this allows the model to estimate as many parameters as there are observations [97]. Finding the regression estimates via $b = [X^T X]^{-1} X^T y$ (Section 2.3.3) provides the model

$$y = b_0 + b_1 X_1 + b_2 X_2 + X_3 X_3 + b_{12} X_{12} + b_{13} X_{13} + b_{23} X_{23} + b_{123} X_{123}$$

The magnitude of the regression coefficients and their statistical significance can reveal the next step in the experimentation. If one or more regression coefficients are very large, it is often of interest to perform additional experiments along their direction. If the design is suspected to be close to an optimum, response surface modelling should be performed (Section 2.3.5.2) [96–98].

However, as the number of experiments increases exponentially alongside the number of factors, many full factorial designs require an unrealistic number of experiments. Consider the following matrix

$$
\begin{array}{cccccccc}
X_0 & X_1 & X_2 & X_3 & X_{12} & X_{13} & X_{23} & X_{123} \\
\end{array}
$$
$$
\begin{bmatrix}
1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\
1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
$$

The matrix above presents a *fractional factorial design* (FFD). The FFD attempts to provide the same information as a full factorial design with a *fraction* of the experiments [92]. The FFD possesses several identical, or *confounding*, columns (*e.g.* X_3 and X_{12}). Confounding columns columns in a FFD are given by a *defining relation* [92, 95, 96, 98]. The defining relation for the FFD above is

$$0 = 123$$

Where the italics in the defining relation are used to emphasise that the numbers refer to the *columns* of the design matrix. To reveal the confounding pattern, a particular effect is multiplied on both sides of the defining relation. If an arbitrary column, *p*, is multiplied by itself, it yields the intercept column, *0*. Meaning $p \cdot p = 0$. Moreover, the product between any column, *p*, and the intercept column, is just *p*. That is $p \cdot 0 = p$. Thus, from the defining relation above, it can be shown

$$0 \cdot 3 = 123 \cdot 3 = 120$$

---

[4]That is, the scalar product of any two columns is zero. Meaning, that for two arbitrary vectors *p* and *q*, then $p^T q = 0$

$$3 = 12$$

Meaning X_3 is indeed confounded with X_{12}. The complete confounding pattern becomes

$$1 = 23 \quad 2 = 13 \quad 3 = 12 \quad 0 = 123$$

However, in the final design matrix, identical columns cannot be present. If two or more of its columns are identical, the matrix $X^T X$ is singular (*i.e.* does not have an inverse), meaning the previously introduced OLS techniques are inapplicable. To remedy this problem, all confounding columns must be identified and reduced into one single column [92]. This yields a new FFD design matrix with four columns

$$
\begin{array}{cccc}
X_0 + X_{123} & X_1 + X_{23} & X_2 + X_{13} & X_3 + X_{12} \\
\end{array}
$$
$$
\begin{bmatrix}
1 & -1 & -1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 1 & -1 \\
1 & 1 & 1 & 1
\end{bmatrix}
$$

Using the design matrix above for the regression, the final model is given by

$$y = (b_0 + b_{123}) + (b_1 + b_{23})X_1 + (b_2 + b_{13})X_2 + (b_3 + b_{12})X_3$$

Notice that confounded variables are found as linear combinations. The *sparsity of effects* principle states that in most models, the responses are determined almost exclusively by main effects and lower-order interactions [96]. Thus, by constructing a FFD with a well-chosen defining relation and resolution (see below), the linear combinations can be manipulated such that (suspected) important effects are only confounded with negligible ones [96]. If one or more factors are found to provide negligible effects, the FFD can often be interpreted as a full factorial design (Figure 10) [95].
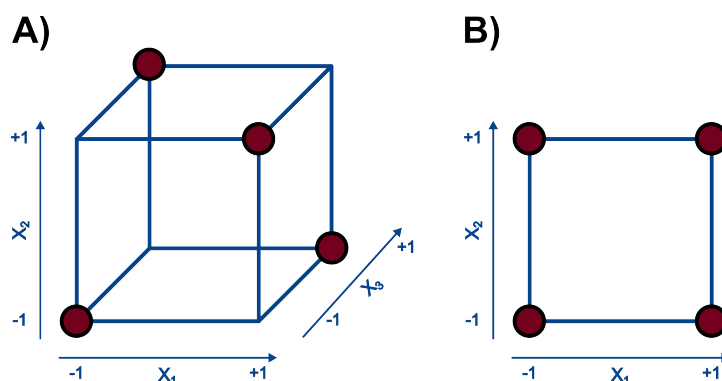
Figure 10: Illustration of the FFD. In **A**, a 2^{3-1} FFD is illustrated. **B** illustrates the same design projected onto the $X_1 X_2$ plane. Therefore, if the effects of X_3 are considered negligible a full factorial design can be obtained in the experimental space spanned by X_1 and X_2. The same logic applies for projecting the design onto the $X_1 X_3$ and $X_2 X_3$ planes.

The features of a FFD can be summarized on the form 2^{I-f}_R. The value 2^{I-f} is the number of experiments conducted in the FFD, where I is the number of factors and f is the fractionation number. If $f = 1$, the design is a *half-fraction* as it uses half the number of experiments required for a full factorial design. Similarly, $f = 2$ is a quarter-fraction and so forth [92]. The R subscript is the *resolution* of the design and is given by the number of non-zero terms in the defining relation, expressed in roman numerals. The following is characteristic for each instance of R [92]:

Table 1: Worst-case confounding found at various design resolutions.

Design Resolution	Worst-Case Confounding
III	Main effects are confounded with two-factor interactions
IV	Main effects are confounded with three-factor interactions and two-factor interactions are confouded with other two-factor interactions
V	Main effects are confounded with four-factor interactions and two-factor interactions are confounded with three-factor interactions

2.3.5.2   Central Composite Design

The purpose of *response surface modelling* is to apply a second-order polynomial to estimate the optimum of some process. Polynomials are applied since they can the locate stationary points, *i.e.* maximums, minimums and saddle points, on a curved response surface (Figure 11) [96, 98].
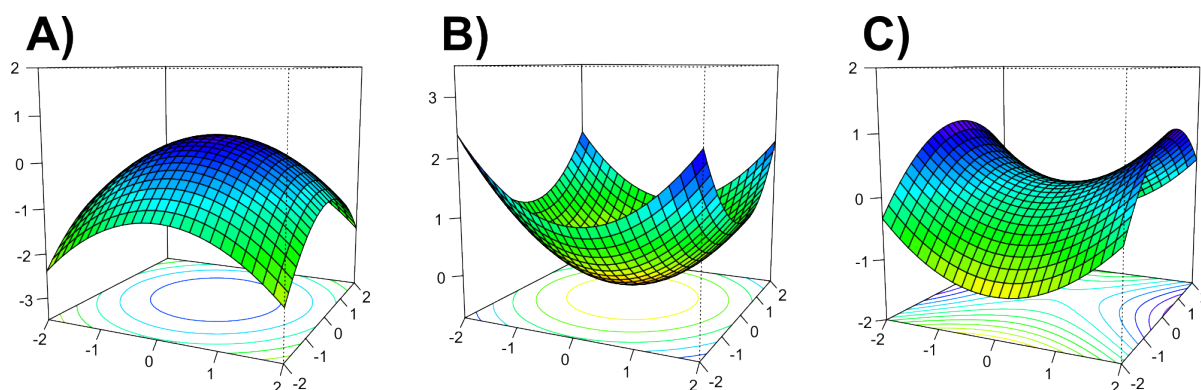
Figure 11: Stationary points are points in a function where the derivative is zero. They can take one of three forms: maximum, minimum, and saddle point. A maximum (**A**) is represented as a peak, while a minimum (**B**) is represented as a trough. These points represent the highest or lowest points in the entire experimental range, respectively. A saddle point (**C**) is a point at which the derivative is zero, but it is neither a maximum nor a minimum. This means that there are points in the experimental range where the response is higher than the stationary point.

A common way to estimate the polynomial model is via *central composite designs* (CCDs) (Figure 12). These designs are either full or fractional factorial designs, expanded with axial points and center points. If an FFD is used as the basis of a CCD, then its resolution must be at least *V* to ensure integrity of the response surface [92].[5] An *axial point* is an experiment where one

---

[5]If the FFD's resolution is *IV* or lower, second-order interactions will be confounded with each other, disallowing any rigorous estimation of the response surface (Table 1).

factor is held at a level α and all remaining factors at their center levels. Similarly to the factorial experiments, axial points are also run at high $(+\alpha)$ and low $(-\alpha)$ levels [92, 96–98].
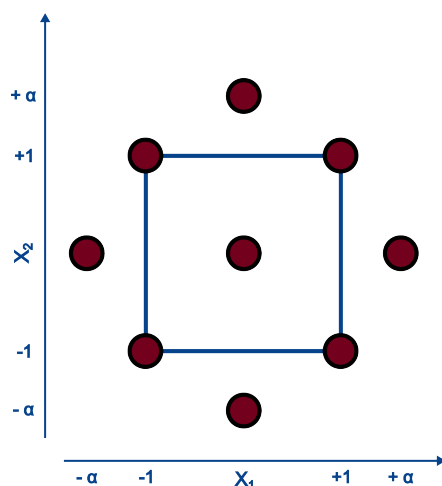
Figure 12: Illustration of the CCD. The square illustrates the underlying factorial design. The axial points, surrounding the factorial design in a circle-wise fashion, allow the estimation of curvature via quadratic terms.

Quadratic effects can be calculated for any $|\alpha| > 1$, yet choosing a value for α can be complex. In general, a good choice of α is provided by

$$\alpha = \left[\frac{2^{k-f} \cdot n_c}{n_\alpha} \right]^{\frac{1}{4}} \tag{29}$$

Where n_α is number of replicates at each axial point and n_c is the number of replicates at each unique experiment in the factorial design. The advantage of estimating α from (29) is that all estimates within the experimental region possess the same uncertainty [91, 98].

2.3.6   Uncertainty Intervals

2.3.6.1   Confidence Intervals

Confidence intervals (CI) are the most commonly applied uncertainty intervals in statistics and are used to estimate population parameters. CIs are defined such that, upon repeated sampling, a proportion $1 - \alpha$ of the constructed intervals will contain the population parameter of interest. Considering the mean of the normal distribution, this may formally be expressed as

$$Pr\left(\bar{X} - t^*_{1-\alpha/2,n-1} \cdot S_{\bar{X}} < \mu < \bar{X} + t^*_{1-\alpha/2,n-1} \cdot S_{\bar{X}} \right) = 1 - \alpha \tag{30}$$

Where $S_{\bar{X}}$ is the sample standard error of the mean. A CI for μ satisfying Equation (30) is given by

$$CI_{1-\alpha} = \bar{x} \pm t^*_{1-\alpha/2,n-1} \frac{s}{\sqrt{n}} \tag{31}$$

The value $1 - \alpha$ is the CIs *confidence level*. Moreover, the CI is closely related to a two-tailed t-test with a significance level, α. In fact, the CI is simply a reformulation of aforementioned test; the CI contains all the values which are not statistically significant from $\bar{x}$ at a level, α [85].

2.3.6.2   Prediction Intervals

A *prediction interval* (PI) concerns the prediction of a future random variable. This is in contrast to the CI, which seeks to estimate a fixed (non-random) population parameter. The PI is defined similarly to the CI; if a random sample is generated and a subsequent PI constructed, in the long run, a proportion $1 - \alpha$ of the PIs will contain the next instance of the future random variable. Denoting the mean of the collected sample data as $\bar{Z}_n$ and the future random variable as Z_{n+1}, this may be written as

$$Pr\left(\bar{Z}_n - t^*_{1-\alpha/2,\nu} \cdot S_p < Z_{n+1} < \bar{Z}_n + t^*_{1-\alpha/2,\nu} \cdot S_p\right) = 1 - \alpha \tag{32}$$

Where S_p is the prediction error of Z_{n+1}.

Consider first a random sample consisting of n draws of the random normal variable, X. It can then be shown [85] that a PI for the future random variable X_{n+1} is given by

$$PI_{1-\alpha/2,X_{n+1}} = \bar{x} \pm t^*_{1-\alpha/2,n-1} \cdot s \sqrt{1 + \frac{1}{n}} \tag{33}$$

Alternatively, consider a univariate FOM where $X = x^*$, a PI for the future instance of Y is then calculated in accordance with [85]

$$PI_{1-\alpha/2,Y_{n+1}} = b_0 + b_1 x^* \pm t^*_{\alpha/2,n-2} \cdot s_{y/x} \sqrt{1 + \frac{1}{n} + \frac{(x^* - \bar{x})^2}{\sum_{k=1}^{n}(x_k - \bar{x})^2}} \tag{34}$$

2.3.7   Statistical Diagnostics

When a regression model is proposed, it is essential to assess whether the model can adequately describe that data or not. To do this, *statistical diagnostics* are applied [85].

2.3.7.1   Residual Plots

The goodness of fit of a regression line can found by various means. A common way to assess model fits is via the *residual plot*. Here, the residuals are calculated and plotted against the independent variable. The distribution of the residuals can then be used to evaluate model fit [85]. Figure 13 illustrates some typical residual plots produced when a univariate FOM has been fitted to different kinds of data.
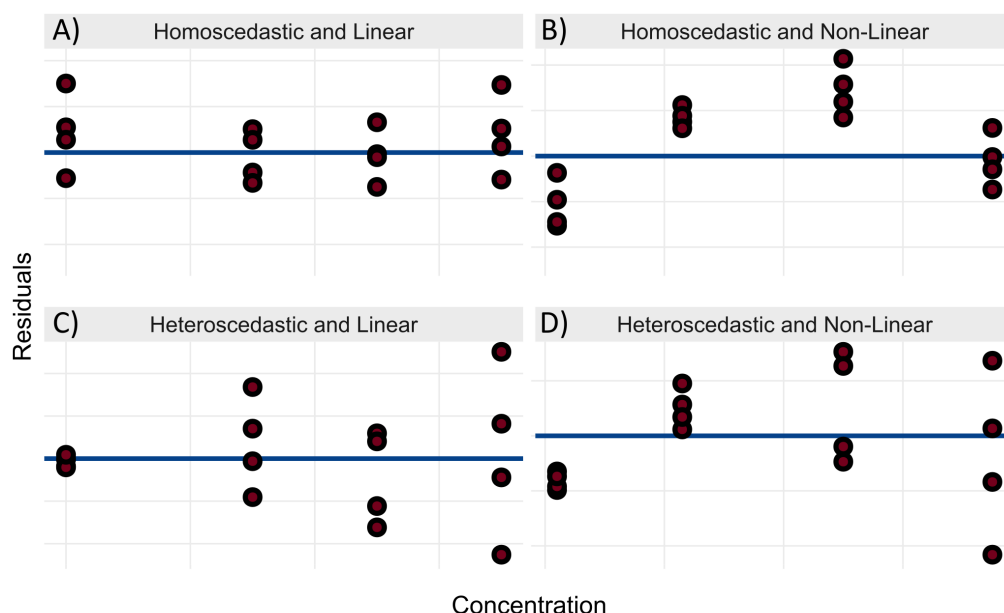
Figure 13: Typical trends in residual plots.

If the data are well fitted by the model, the points in the residual plot should be evenly centered around zero (plot **A**, Figure 13). An incomplete model might be evidenced by shapes in the residuals. For instance, curvature in the residuals (plots **B** and **D**) could suggest that the data is non-linear and that the FOM must be expanded with a quadratic term. Moreover, if the residuals are not scattered evenly, but increase alongside the independent variable (plots **C** and **D**), there is evidence suggesting heteroscedasticity; a weighted regression model is required [85, 89].

2.3.7.2   Mandel's Test for Goodness of Fit

It is also possible to compare method fits by the general linear hypothesis. In chemical literature, when the two models being compared are first order ($Y = \beta_0 + \beta_1 X + \varepsilon$) and second order polynomials ($Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$), the test is often referred to as *Mandel's test* [99]. The relevant hypotheses become

$$H_0 : \ \beta_2 = 0$$

$$H_a : \ \beta_2 \neq 0$$

If H_0 is rejected, there is significant evidence suggesting that a quadratic term is required to adequately describe the data [93]. The test statistic becomes

$$F_{1,n-3} = \left(SS_{E,f} - SS_{E,q} \right) \div \frac{SS_{E,q}}{n-3} \tag{35}$$

Where n is the number of data points used to estimate the models. Moreover, $SS_{E,f}$ and $SS_{E,q}$ denote the error sums of squares for the first-order and quadratic polynomials, respectively. The F-statistic have 1 and $n-3$ degrees of freedom and can be compared with tabulated values to

assess potential significance [93]. If the data is heteroscedastic, the polynomials can be weighted and their weighted error sums of squares be applied [92].

2.3.7.3 Levene's Test for Homoscedasticity

Levene's test is a common statistical test applied for the detection of heteroscedasticity (lack of homoscedasticity). The test utilizes the one-way ANOVA to test for differences in variances between distinct groups. Within each group, the absolute differences between the individual measurements (Y_{ij}) and the group mean $\bar{Y}_i$ are calculated:

$$Z_{ij} = |Y_{ij} - \bar{Y}_i|$$

Using Z_{ij} as the dependent variable, the one-way ANOVA can be carried out as described in Section 2.3.4 with the hypotheses [85]

$$H_0 : \text{ All group variances are equal}$$

$$H_a : \text{ All group variances are not equal}$$

2.4 Method Validation

Method validation is the process in which analytical methods are proven to be fit for their intended use. This is done by examining an analytical procedure via a number of quality parameters [100]. The number of parameters, as well as their definitions and evaluations, can differ greatly between guidelines [101]. The sections below provide an overview over some of the most commonly encountered validation parameters.

2.4.1 Response Function

Modern analyses largely depend on analytical instrumentation where instrumental response is determined by the magnitude of analyte present in the sample. The relationship between instrumental response and analyte quantities is described by the method *response function* [102, 103].

This function may be established by regressing standard solutions against their associated instrumental responses. The most commonly applied regression techniques include OLS and WLS methods [85, 88, 89], meaning the response functions can be summarized by

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon} \tag{14}$$

In the case of WLS, some frequently applied weights include the reciprocal variance of the dependent variable:

$$w_k = \frac{1}{s_k^2} \tag{36}$$

and the reciprocal independent variable, raised to a power τ:

$$w_k = \frac{1}{x_k^\tau} \tag{37}$$

where the value of τ is often chosen to be 1 or 2.

When a response function has been fitted to the calibration solutions, statistical diagnostics must be applied to confirm model adequacy [102].

2.4.2   Precision

The *precision* investigates the random error of the method. The *random error* is the unpredictable error found during replicate measurements. This error is quantified by the estimated method standard deviation [104, 105].

The method precision is heavily influenced by the conditions at which it was evaluated. The two extreme types precision are the repeatability and reproducibility. The *repeatability* (*r*) provides the lowest level of precision. This level studies the random error found by a single analyst using the same equipment over a short time period. The *reproducibility* (*R*) describes the greatest level of method variance and investigates the between-lab effect. Here, the random error of the method is evaluated when replicates are generated by several analysts at different laboratories over a longer period of time. Between these two extremes is the *intermediate precision* (*IP*) which investigates the between-day effect. This type of precision describes the method variation obtained when results are produced by a single laboratory with several analysts over a longer period of time [85, 106].

Precision is often determined through the ANOVA [85, 100, 106, 107]. In Section 2.3.4, it was shown that the one-way ANOVA model can be expressed as

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \tag{27}$$

Let the between-day effect denote the factor under study. Then, the variable Y_{ij} represents the jth measurement result generated by the method on day i. The recorded result depends on the true method result, μ, between-day effects, α_i, and random noise, ε_{ij}.

The repeatability variance, $u^2(r)$, is represented by the variance of the noise, ε_{ij}. In turn, this term is estimated by the within-day variance, MS_E with $\nu_r = I(J-1)$ degrees of freedom. An estimate of $u(r)$ becomes

$$u(r) = \sqrt{MS_E} \tag{38}$$

The IP variance, $u^2(IP)$, considers method dispersion due to both the within-, $u^2(r)$, and between-day variance, $u^2(b)$:

$$u(IP) = \sqrt{u^2(r) + u^2(b)} \tag{39}$$

Where the estimate for $u^2(b)$, with $\nu_b = I - 1$ degrees of freedom, is given by

$$u^2(b) = \frac{MS_M - MS_E}{J} \tag{40}$$

For any precision component, P, its uncertainty can be expressed relative to its estimate ($\bar{x}$) via the *relative standard deviation* (RSD) [100]:

$$RSD_P = \frac{u(P)}{\bar{x}} \cdot 100\% \tag{41}$$

When plotted against concentration, the *RSD* for larger concentration ranges often show a characteristic shape. At low concentrations, instrumental noise may interfere with signals from the test sample, providing high RSDs. However, as the concentration increases, the noise often becomes negligible and the *RSD* stabilizes (Figure 14). In these higher concentration ranges, if the *RSD* is known, it can be used to estimate the precision of any measurement falling within its relevant range [83, 100, 106].
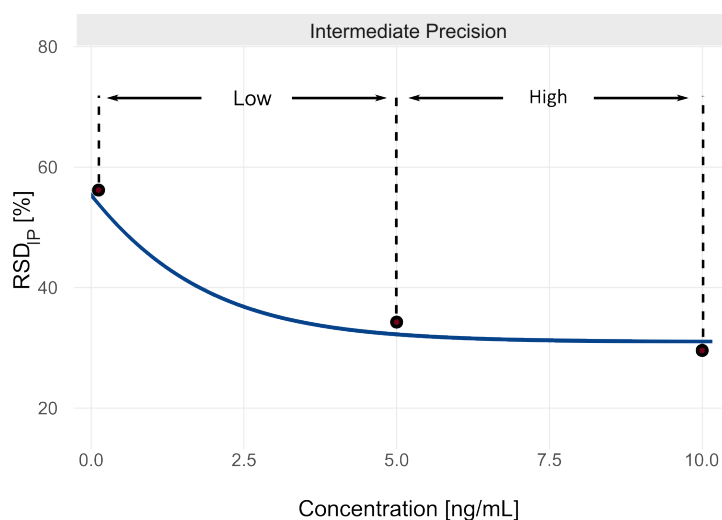
Figure 14: RSD_{IP} as a function of concentration. The blue line denotes the true and unknown IP and each point marks a validated concentration level at which the IP has been estimated. The RSD at the low level rapidly decreases as the concentration increases. For the high level, the RSD is essentially constant. Thus, any estimate falling between 5.0 and 10 ng/mL can be assumed to possess the same precision.

To determine if the random error is satisfactory, precision estimates are often compared to a reference value. Reference values for the IP ($RSD_{IP,h}$) and repeatability ($RSD_{r,h}$) are often provided by *Horwitz equation* [108–110]:

$$RSD_{IP,h} = 2C^{-0.15} \tag{42}$$

$$RSD_{r,h} = \frac{2}{3}RSD_{IP,h} \tag{43}$$

If a method has an acceptable random error, the ratio of its precision components to those estimated by Horwitz should be close to one. This ratio, the *HorRat*, thus indicates precision quality. If the *HorRat* is too large, the precision of the method is likely unsatisfactory [108–110].

2.4.3   Trueness

The *trueness* is the systematic error of a measurement result. It describes the closeness between the mean of an infinite number of method results and some accepted reference value. The distance between the mean result and reference value is described by the *bias*, δ [85].

Ideally, to quantify bias, suitable reference values with representative concentration levels and matrix composition should be used. However, these materials are not always readily obtainable and the bias must thus be quantified by means of spiking experiments [85, 100].

In *spiking experiments*, matrix blanks are added a known amount of analyte, $\hat{\mu}$. The test solutions are subsequently analysed and the average concentration estimate ($\bar{x}$) calculated. Bias estimates may then be provided in both absolute ($\hat{\delta}$) and relative forms ($\hat{\delta}_{\%}$) [100]:

$$\hat{\delta} = \bar{x} - \hat{\mu} \tag{44}$$

$$\hat{\delta}_{\%} = \frac{\bar{x} - \hat{\mu}}{\hat{\mu}} \cdot 100\% \tag{45}$$

Moreover, the recovery, $R_{\%}$, is a measure of how well the estimated concentrations agree with the spiked concentrations, expressed as a percentage [111, 112]:

$$R_{\%} = \frac{\bar{x}}{\hat{\mu}} \cdot 100\% = 100\% - \hat{\delta}_{\%} \tag{46}$$

2.4.4   Accuracy

The *accuracy* provides an overall estimate of the difference between a test result and an accepted reference value, accounting for both random and systematic errors.

2.4.4.1   Measurement Uncertainty

The most common way to express the accuracy is via the *measurement uncertainty* (MU), which expresses accuracy as a CI.

In general, the final measurement estimate, y, can be seen as a function of I other components related by some function f [105]. That is,

$$y = f(x_1, ..., x_I)$$

Using the function f, the MU combines all I uncertainty components, $u(x_i)$, into a single uncertainty estimate via the *law of propagation of uncertainty* [105]. This law expresses that the variance in y, $u_c^2(y)$, is given by

$$u_c^2(y) = \sum_{i=1}^{I} a_i^2 u^2(x_i) \equiv \sum_{i=1}^{I} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) \tag{47}$$

The constants a_i are defined as the partial derivatives of the function f with respect to x_i. These *sensitivity coefficients* expresses how much a change in x_i affects the final uncertainty $u_c(y)$.

For the final uncertainty estimate, $u(y)$, it can be shown [113, 114] that its *effective degrees of freedom* (ν_{eff}) is well approximated by *Satterthwaite's equation*:

$$\nu_{\text{eff}} \approx \frac{u_c^4(y)}{\sum_{i=1}^{I} \left(a_i^4 u^4(x_i)/\nu_i\right)} \tag{48}$$

After estimating u_c and ν_{eff} via the equations above, a $1 - \alpha$ CI for y can be constructed. This is done by multiplying the *standard uncertainty*, u_c, by its $1 - \alpha/2$ quantile, providing the final *expanded measurement uncertainty* (U) [105]:

$$U_{1-\alpha} = y \pm u_c(y) \cdot t^*_{1-\alpha/2, \nu_{\text{eff}}} \tag{49}$$

2.4.4.2 Accuracy Profile

Another way to assess accuracy is to construct a PI, rather than CI. The PI provides an interval where a future measurement is expected to lie with some confidence $1 - \alpha$.

An analytical procedure can be deemed accurate if it can produce errors smaller than some maximum acceptance limit, λ, Therefore, if the method's PI is entirely contained within the acceptance limits, the procedure is expected to provide sufficiently accurate results with $1 - \alpha$ confidence. The PI in this context is referred to as an *accuracy profile* (AP) (Figure 15) [107].
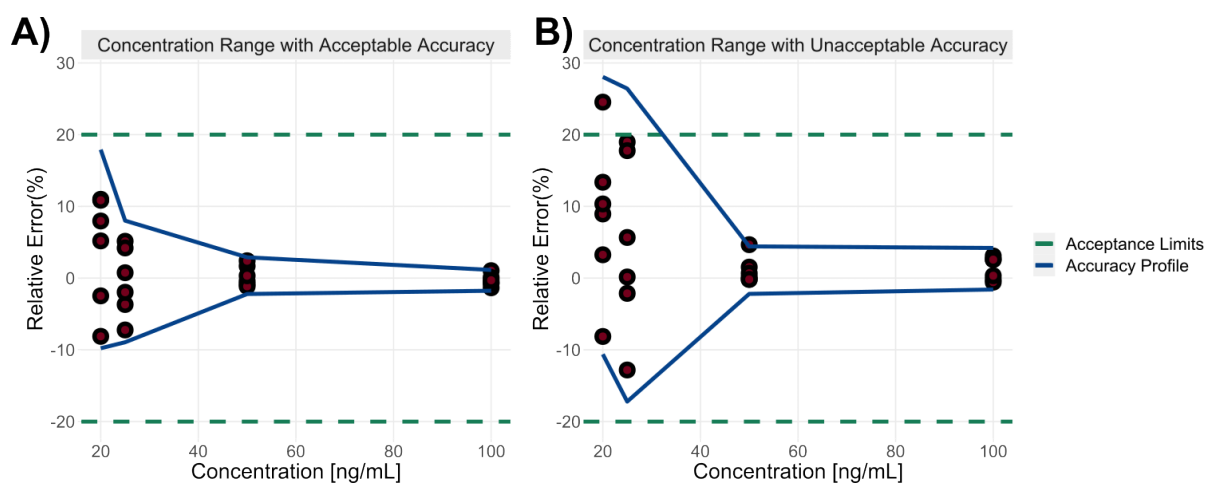
Figure 15: The two possible outcomes for the AP. **A** presents a method with acceptable accuracy; the AP is found entirely between the acceptance limits across the concentration range. With $1 - \alpha$ confidence, future measurements are thus expected to fall within the acceptance limits. **B** presents a method with unacceptable accuracy; the AP of the first two concentration levels are found outside the acceptance limits. This means that, with $1 - \alpha$ confidence, these measurements are likely to fall outside the acceptance limits.

The AP may calculated in absolute form

$$AP_{1-\alpha} = \bar{x} \pm t_{1-\alpha/2, \nu_{\text{eff}}} \cdot u(IP) \sqrt{1 + \frac{1}{n_{eff}}} \tag{50}$$

or alternatively in relative form

$$AP_{1-\alpha,\%} = \hat{\delta} \pm t_{1-\alpha/2, \nu_{\text{eff}}} \cdot RSD_{IP} \sqrt{1 + \frac{1}{n_{eff}}} \tag{51}$$

Where n_{eff} denotes the *effective sample size*

$$n_{eff} = \frac{u^2(IP)}{u^2(\bar{x})} \tag{52}$$

and $u(\bar{x})$ is the uncertainty of $\bar{x}$ with ν_{eff} degrees of freedom [105, 107, 115].

2.4.5 Limit of Detection

The *limit of detection (LoD)* is the lowest concentration of analyte that can readily be distinguished from noise [85, 100].

To distinguish analyte signal from noise, both type I (α) and II (β) errors should be accounted for. Let b_{blk} and s_{blk} denote the mean response and standard deviation for a blank sample, respectively. The instrumental response at the LoD (Y_{LoD}) can then be estimated via

$$Y_{LoD} = b_{blk} + (t^*_{v,\alpha} + t^*_{v,\beta}) \cdot s_{blk} \tag{53}$$

Y_{LoD} can then be converted into an estimate of the LoD via a response function [85, 104].[6]

Alternatively, assuming the response function is a univariate FOM, it is possible to estimate LoDs directly using the regression coefficients:

$$LoD = (t^*_{v,\alpha} + t^*_{v,\beta}) \cdot \frac{s_{y/x}}{b_1} \tag{54}$$

However, since b_1 is a random variable, the LoD proposed by Equation (54) can give too optimistic estimates. To account for the uncertainty in the slope, the formula can be expanded with its (slightly simplified) PI (34) [106, 116, 117]:

$$LoD = \left(t^*_{v,1-\alpha} + t^*_{v,1-\beta}\right) \cdot \frac{s_{y/x}}{b_1} \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i + \bar{x})^2}} \tag{55}$$

While the equations vary in complexity, equations (53) through (55) all build upon the same concepts (Figure 16). In essence, the only difference separating the three equations is how they estimate instrumental noise term.
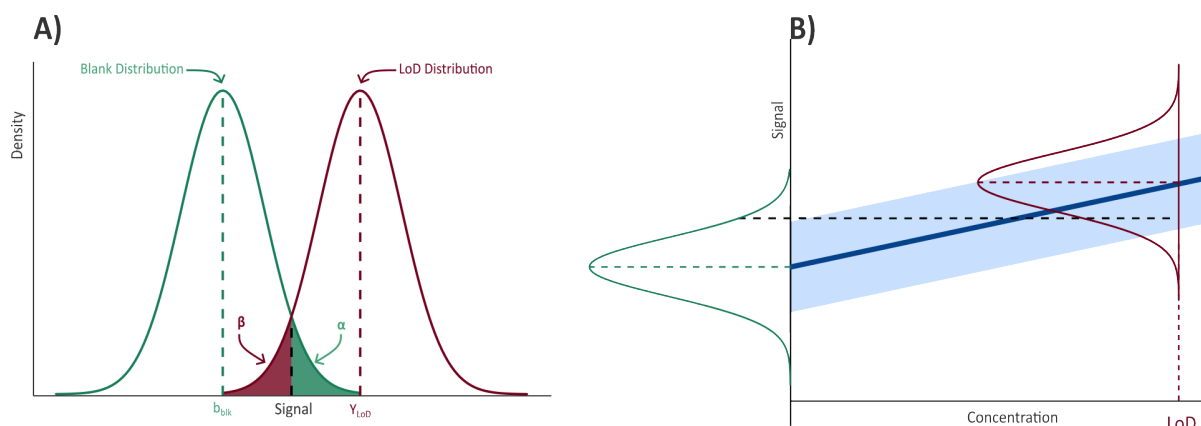
Figure 16: The estimation of the LoD. **A** shows how the instrumental signal at the LoD is distinguished from the noise; a distance between the mean blank response (b_{blk}) is determined by $t^*_{\alpha,v}$ and $t^*_{\beta,v}$ to provide an acceptable type I (α) and II (β) error rate, as given by the green and red shaded areas respectively. **B** shows how the regression curve (dark blue) relates to the instrumental signal at the LoD and blank levels. The light blue rectangle corresponds to the expanded uncertainty of the noise (as given by multiples of $s_{y/x}$ or the PI) and is assumed constant across the entire concentration range (homoscedastic).

2.4.6   Limit of Quantification

The *limit of quantification* (LoQ) is the lowest concentration at which the analytical method is able to provide results with acceptable accuracy.

---

[6]A commonly observed estimate is $Y_{LoD} = 3s_{blk}$. This estimate is a simplification of (53). If the type I and II errors are both set at the 5% level, $t^*_{v,\alpha} = t^*_{v,\beta} = t^*_{v,0.05}$, both values will approach 1.645 as v increases. Thus, assuming negligible noise, $Y_{LoD} = (1.645 + 1.645) \cdot s_{blk} = 3.29s_{blk} \approx 3s_{blk}$. [85]

The LoQ is often taken to be a multiple of the LoD. Generally,

$$LoQ = 3 \cdot LoD \tag{56}$$

However, this estimate for the LoQ is merely a convention and the choice of the 3 multiple is arbitrary [104].

An alternative approach to estimate the LoQ is via the AP. The LoQ may be set to the lowest concentration at which which the AP is fully contained within the acceptance limit. Again, consider Figure 15 and its four concentration levels. For the AP in **A**, the lowest measured concentration level shows acceptable accuracy, and may thus serve as the LoQ. In **B**, however, the two lower levels provide unacceptable accuracy. In this case, the LoQ may either be taken as the lowest level at which the AP is fully contained within the acceptance limits (in the figure, around 50 ng/mL). Alternatively, the LoQ may set to the concentration where the AP intersects the acceptance limits (around 30 ng/mL) [102, 107].

2.4.7   Linearity

Method *linearity* describes the ability of the analytical procedure to generate concentration estimates that are directly proportional to the true concentration. That is, the linearity investigates if concentration estimates yield a straight line when they are plotted against the true concentrations.

The linearity is a requirement for the evaluation of trueness. If the method is biased, but the bias is constant and/or proportional to the concentration, there will be predictable correction factors that may be applied to the measurement result, ensuring adequate accuracy. However, any potential deviations from a straight line suggests that the method is not able provide accurate measurements [85, 102, 106, 107].

2.4.8   Selectivity

The extent to which a method can detect analytes without being disturbed by other compounds is termed *selectivity* [100].

For chromatographic methods, selectivity can be demonstrated by providing a chromatographic separation with baseline resolution. Alternatively, for LC-MS methods, overlapping peaks can be deemed to possess acceptable selectivity if they have unique MS transitions [100, 118]. Overlapping peaks with no unique transitions can also be deemed adequately selective if they can be quantified with reasonable accuracy [73].

3 Materials and Methods

3.1 Chemicals and Reagents

3.1.1 Standards and Materials

Standards for 1-hydroxyphenanthrene (1-OHPHN), 2-hydroxyphenanthrene (2-OHPHN), 3-hydroxyphenanthrene (3-OHPHN), 4-hydroxyphenanthrene (4-OHPHN), trans-9,10-dihydroxy-9,10-dihydrophenanthrene (9,10-OH-9,10-HPHN) and 4-hydroxyphenanthrene-d_9 (4-OHPHN-d9) were purchased from Chiron (Trondheim, Norway). The 9-hydroxyphenanthrene (9-OHPHN) and bisphenol-d_{16} (BPA-d16) were purchased from Sigma-Aldrich (Saint Louis, MO, USA). Trans-1,2-dihydroxy-1,2-dihydrophenanthrene (1,2-OH-1,2-HPHN) was purchased from MRI Global Carcinogen Repository (Kansas City, MO, USA). The 2000 U β-Glucuronidase from *Helix pomatia*, was purchased from Sigma-Aldrich (Saint Louis, MO, USA). Phree (1 mL) and Strata-X (6 mL) SPE cartridges were purchased from Phenomenex (Torrance, CA, USA).

3.1.2 Sample Preparation

The following chemicals were purchased from Merck KGaA (Darmstadt, Germany): Methanol (MeOH) for gas chromatography of ECD and FID SupraSolv® quality; MeOH for liquid chromatography of LiChrosolv® quality; Acetic acid of Suprapur® quality; Diethyl ether (DEE) of SeccoSolv® quality; ultra pure water was obtained through a Milli-Q® system from the same vendor. DEE was also purchased from Aldrich (Saint Louis, MO, USA).

3.1.3 Chromatography

Acetonitrile (AcN) of LC-MS grade and HiPerSolv CHROMANORM® quality was purchased from VWR (Leuven, Belgium). AcN of CHROMASOLV™quality was bought from Honeywell Riedel-de Haën (Seelze, Germany). Ultra pure water was obtained by a Milli-Q® system from Merck KGaA (Darmstadt, Germany).

3.2 Sample Preparation

Haddock eggs were added to a 2 mL polypropylene vial and homogenized with a hand mixer using cone-shaped tips. A total of 40 μL of 500 ng/mL surrogate standard (4-OHPHN-d9) was added to each vial, along with spike mixes where applicable. The vials were then added 1 mL of MeOH, followed by 30 minutes of sonication and 5 minutes of centrifugation at 8000 rpm. For phospolipid removal, the supernatants were quantitatively transferred to Phree cartridges and eluted under vacuum into pre-labeled 1.5 mL autosampler vials. The solutions were concentrated by evaporation down to 1 mL under N_2 at 40°C. The vials were then vortexed and the extracts split; a 0.5 mL aliquot was kept as a non-hydrolyzed fraction, and the remaining 0.5 mL was quantitatively transferred to a 12 mL glass tube for further hydrolysis.

To hydrolyse the phase II metabolites in the extract, converting them into detectable phase I congeners, the 12 mL tubes were treated with 5 mL of sodium acetate buffer (1 M, pH 5) containing 2000 U β-Glucoronidase. The samples were vortexed and incubated for 60 minutes at 40ºC with a loose lid. After incubation, 300 µL glacial acetic acid were added to the samples to stop the hydrolysis.

The Strata-X cartridges were equilibrated with 4.5 mL of MeOH followed by 4.5 mL water. The solutions in the 12 mL tubes were then quantitatively transferred to the cartridges. Cartridges were washed with 6.8 mL water followed by 4 mL MeOH/water 63/37 (v/v). Afterwards, samples were dried for 30 minutes under vacuum. Samples were subsequently eluted from the cartridges into 12 mL tubes by the addition of 6 mL MeOH/DEE 65/35 (v/v). The 12 mL tubes were then added 50 µL of 1000 ng/mL internal standard (BPA-d16). The sample volumes were then evaporated down to approximately 0.5 mL under N_2 at 40ºC. Finally, samples were quantitatively transferred to autosampler vials and further evaporated to a final volume of 0.5 mL.

3.3 Chemical Analysis

After preparation, samples were analysed using an Agilent Technologies LC-MS system (Santa Clara, CA, USA) consisting of an Agilent 1290 Infinity II LC, which included an Agilent 1290 Infinity Binary Pump, an Agilent 1290 Infinity Autosampler and an Agilent 1290 Infinity Thermostatted Column Compartment. Separation was performed using a 2.1 x 150 mm BEH Shield RP18 reversed-phase column with 1.7 µm particle size from Waters (Milford, MA, USA). The MS system was an Agilent G6460C Triple Quadrupole with an Agilent Jet Stream ESI.

For each run, 5 µL of sample were injected into the LC-MS system. The column compartment was held at a constant temperature of 45ºC. The mobile phase consisted of water (solvent A) and AcN (solvent B). The LC was run in gradient mode: the initial composition (solvent A/solvent B) was 40/60 at a flow of 0.2 mL/min; the initial composition was held for 0.2 min; 15.0 min to 70/30; 0.3 min to 98/2 and 0.35 mL/min; the conditions were held for 2.5 min; 0.1 min to 60/40 and 0.3 mL/min; the conditions were held for 8 min; 0.1 min to initial conditions at 60/40 and 0.2 mL/min. The gradient program is illustrated in Figure 17.
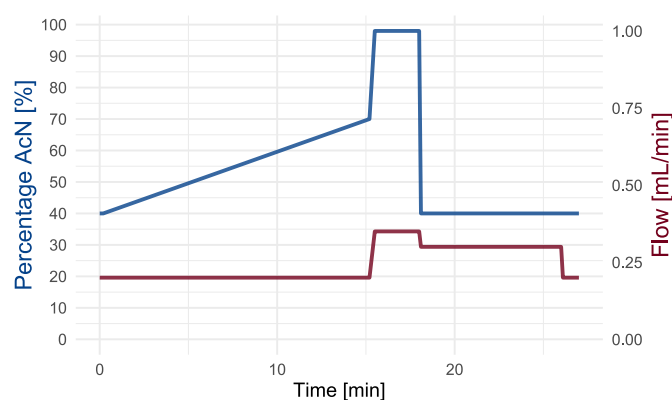
Figure 17: LC gradient program, change in percentage organic composition (blue) and flow rate (red) is shown as a function of time.

The ion source was operated in negative mode with the following conditions: gas temperature and flow at 250ºC and 12 L/min; nebulizer pressure at 25 psi; sheath gas temperature and flow at 300ºC and 11 L/min; the capillary and nozzle voltages at 3500 V and 2000 V.

The MS scan settings [119] are presented in table 2

Table 2: Scan conditions for the MS.

Compound	Precursor [m/z]	Product [m/z]	Fragmentor [V]	Collision Energy [eV]	RT [min]	Cell Accelerator [V]
1-OHPHN	193.1	165.1	145	30	12.6	1
2-OHPHN	193.1	165.1	145	30	11.1	1
3-OHPHN	193.1	165.1	145	30	11.2	1
4-OHPHN	193.1	165.1	145	30	12.9	1
9-OHPHN	193.1	165.1	145	30	12.2	1
4-OHPHN-d9	202.1	174.1	156	33	12.8	1
1,2-OH-1,2-HPHN	211.1	193.0	110	10	4.5	1
9,10-OH-9,10-HPHN	211.1	181.0	110	10	4.5	1
BPA-d16	241.1	223.1	150	27	6.8	2

The relevant compounds for the analysis are presented in Figure 18.
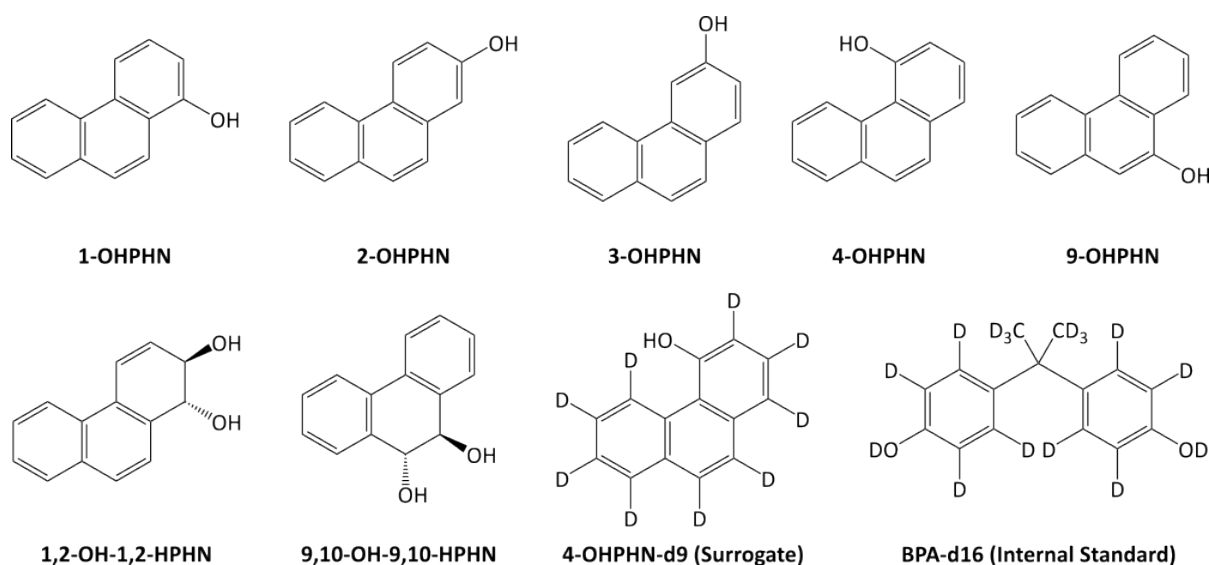
Figure 18: Chemical structures of the analytes, surrogate standard and internal standard.

3.4   Quantification and Data Treatment

To allow for initial quantification, an internal standard, BPA-d16, was added at the end of the sample preparation. The relative responses between the analytes and BPA-d16 were related to concentrations via a response function.

To correct for analyte loss during extraction, a surrogate standard, 4-OHPHN-d9, was added at the beginning of the sample preparation. Initial analyte concentrations were corrected using the following formula:

$$c_{cor} = c_{ncor} \cdot \frac{20 \text{ ng/mL}}{c_{sur}} \tag{57}$$

Here, c_{cor} and c_{ncor} represent the corrected and non-corrected analyte concentrations, respectively, and c_{sur} denotes the estimated concentration of the surrogate standard after extraction. The concentrations in (57) are in units of ng/mL. However, the final units of the method are in ng/g wet weight (ng/g ww), which are calculated using the following formula:

$$c_{ng/g \ ww} = c_{cor} \cdot \frac{1 \text{ mL}}{n_{egg}} \cdot \frac{1 \text{ egg}}{0.00218 \text{ g ww}} \tag{58}$$

Here, $c_{ng/g \ w/w}$ is the final estimate (in ng/g ww), c_{cor} represents the surrogate corrected method result (ng/mL), and n_{egg} denotes the number of haddock eggs in the experiment.

Initial quantification was performed using the MassHunter Quantitative QqQ Software (Santa Clara, CA, USA). Other calculations, including the corrections given by (57) and (58), as well as additional statistical calculations, were conducted in R using the packages ggplot2, dplyr, readxl, writexl, RSM, car, and propagate (Appendix E). The confidence level for all statistical analyses was set to $\alpha = 0.05$. The following codes were used to denote statistical significance: * p < 0.05, ** p < 0.01, *** p < 0.001.

Moreover, two R-packages were developed for the method validation study, which are available for download via GitHub.[7] These packages provide statistical calculations for various validation parameters, as well as a tool to export MassHunter Quantitative data to manageable xlsx files.

3.5   Method Validation

The method validation was conducted in accordance with the Eurachem guidelines [120]. The validation parameters included linearity, trueness, precision, accuracy, LoD, LoQ and selectivity. The response function for the analytical procedure was also evaluated. For all spiking experiments, approximately 20 blank (non-exposed) eggs were used (Appendix D).

The validation study was conducted on the basis of three different sets of experiments, namely the calibration, accuracy and LoD studies. The calibration study encompassed only the response function and the LoD study only the LoDs. The accuracy study encompassed all remaining parameters; these experiments were first used to estimate precision and linearity, with the latter also serving to evaluate trueness. Using the trueness and precision parameters as a basis, the accuracy was determined and further applied to assess the LoQ and selectivity (Figure 19).

---

[7]https://github.com/matbre17/R-Scripts
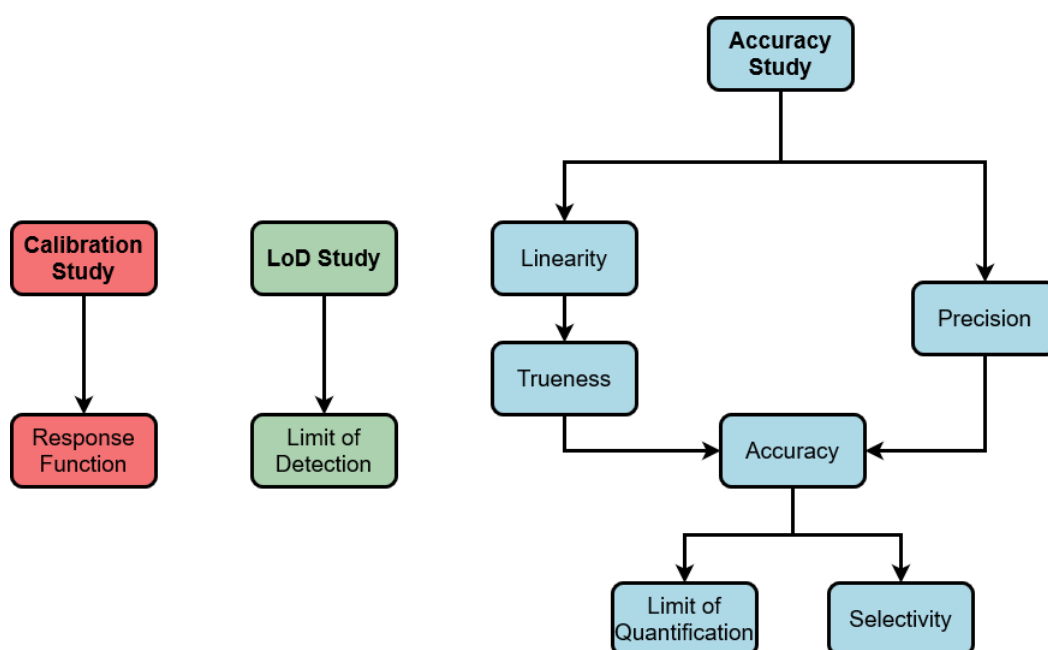
Figure 19: The three sets of validation experiments and their associated validation parameters.

3.5.1 Calibration Study

The calibration study concerned the establishment of the method response function. To do this, seven calibration standards were prepared in blank matrix extracts. The standard concentrations were 0.100, 0.300, 0.500, 1.00, 5.00, 10.0 and 30.0 ng/mL. Each calibration standard were injected and analysed by the LC-MS/MS four times.

The choice of response function was evaluated by plotting instrumental responses against the standard concentrations and consulting relevant statistical diagnostic tools including Levene's test, Mandel's test and residual plots. The selection of weights for the response functions was limited to the ones available in the MassHunter Quantitative Software ($1/x$, $1/x^2$, and $1/s^2$).

3.5.2 LoD Study

The LoDs were estimated by spiking blank eggs with concentrations close to the suspected LoDs and constructing response functions with the extracts (Table 3).

Table 3: Spike concentrations for the LoD study, in units of ng/mL and ng/g ww.

	Concentration				
ng/mL	0.040	0.050	0.060	0.080	0.09
ng/g ww	0.92	1.2	1.3	1.8	2.1

Using the generated response functions, the LoDs were generated in accordance with

$$LoD = \left(t^*_{v,1-\alpha} + t^*_{v,1-\beta}\right) \cdot \frac{s_{y/x}}{b_1} \sqrt{1 + \frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i + \bar{x})^2}} \tag{55}$$

where $\alpha = \beta = 0.05$.

3.5.3 Accuracy Study

All samples in the accuracy study were quantified using a quadratic polynomial with weights $1/x^2$ and four replicates per level.

3.5.3.1 Linearity

Spiking and extraction of blank haddock eggs were performed across four concentration levels (Table 4). The experiments were conducted over a period of five days, with two replicates per level performed each day.

Table 4: Spike Concentrations for the linearity study, in units of ng/mL and ng/g ww.

	Concentration			
ng/mL	0.250	1.00	5.00	10.0
ng/g ww	5.73	22.9	115	229

The estimated concentrations from the extracts were regressed against the spike concentrations using the same weights presented in Section 3.5.1. Statistical diagnostics, including residual analysis, Levene and Mandel's tests, were conducted to evaluate linearity.

3.5.3.2 Trueness

Using the regression models from the linearity study, the trueness was assessed by comparing their slopes and intercepts against 1 and 0, respectively. These differences were assessed by t-tests. Significant differences indicate the presence of considerable bias and warrants the application of bias corrections.

Trueness estimates were also estimated by the spiking experiments and expressed as recoveries

$$R_\% = \frac{\bar{x}}{\hat{\mu}} \cdot 100\% = 100\% - \hat{\delta}_\% \tag{46}$$

and relative bias

$$\hat{\delta}_\% = \frac{\bar{x} - \hat{\mu}}{\hat{\mu}} \cdot 100\% \tag{45}$$

3.5.3.3   Precision

Precision was also assessed by the same experiments used for the linearity and trueness evaluations. Using ANOVA techniques, the repeatability and IP were estimated via

$$u(r) = \sqrt{MS_E} \tag{38}$$

and

$$u(IP) = \sqrt{u^2(r) + u^2(b)} \tag{39}$$

The final precision components were provided in relative forms. Additionally, the laboratory precision estimates were compared to theoretical estimates from Horwitz equation:

$$RSD_{IP,h} = 2C^{-0.15} \tag{42}$$

$$RSD_{r,h} = \frac{2}{3}RSD_{IP,h} \tag{43}$$

3.5.3.4   Accuracy

Accuracy was evaluated by combining the trueness and precision components into a single uncertainty estimate. Accuracy was expressed via the MU and AP.

A general expression to calculate the standard uncertainties was derived via

$$u_c^2(y) = \sum_{i=1}^{I} c_i^2 u^2(x_i) \equiv \sum_{i=1}^{I} \left(\frac{\partial f}{\partial x_i}\right)^2 u^2(x_i) \tag{47}$$

The final accuracy components were provided as standard uncertainties in relative form.

The APs, in relative form, were constructed using the formula

$$AP_{1-\alpha,\%} = \hat{\delta} \pm t_{1-\alpha/2,\nu_{\text{eff}}} \cdot RSD_{IP} \sqrt{1 + \frac{1}{n_{eff}}} \tag{51}$$

Additionally, the estimated APs were compared with theoretical APs generated from Horwitz (42). Accuracy was deemed satisfactory if the AP estimates did not exceed twice those of the theoretical ones.

The degrees of freedom were calculated from Satterthwaite's approximation:

$$\nu_{\text{eff}} \approx \frac{u_c^4(y)}{\sum_{i=1}^{I} \left(c_i^4 u^4(x_i)/\nu_i\right)} \tag{48}$$

3.5.3.5   Limit of Quantification

The LoQ was determined using the APs found during the accuracy evaluation. The LoQ was set to be the lowest concentration at which the APs were contained within the acceptance limits,

3.5.3.6   Selectivity

To determine if the selectivity of the method was sufficient, the accuracy of each validated level was evaluated. Satisfactory accuracy was taken as evidence that method selectivity was adequate. Selectivity was also investigated by calculating the chromatographic resolution of the compounds

$$R_s = 0.589 \cdot \frac{t_{R,B} - t_{R,A}}{\frac{1}{2}\left(w_{B,h} + w_{A,h}\right)} \tag{2}$$

3.6   Method Application

The validated method was applied to 47 different samples obtained from two different exposure studies (Table 5). In the first study (WSF19), oil microdroplets were dispersed into sea water and the resulting water soluble fraction (WSF) removed. The mixture was then fractionated by running the solution through reverse-phase and ion exchange SPE cartridges; separating the components based on polarity and charge, respectively. A total of eight fractions were obtained by the ion exchange columns, whereas four fractions were obtained by the reverse-phase column. In the four reverse-phase fractions, SI0 denotes the least retained fraction and SI4 the most retained. Haddock eggs were exposed to three different mixtures. Namely, the unfractionated WSF, the SI0 fraction and a mixture comprised of all 12 individual fractions (WSF Comb). Exposures were conducted at *low*, *medium* and *high* concentration levels. A publication with the complete details for this study are in preparation (C. Donald, personal communication, 17 April 2023)

In a second study [30], haddock eggs were exposed to individual PAH compounds, including phenanthrene. This study was conducted two times (PD18 and PD19). The exposures were performed using passive dosing. In this technique, a silicon rod is first loaded with some toxicant of interest. Afterwards, by placing the rod in a MeOH solution with eggs, the rods transfer their loaded compounds into the aqueous environment; providing the eggs a continuous exposure of the toxicant. The passive dosing was conducted at five levels; *S0* is the saturated solution, *S1* is the three-fold dilution of S0, *S2* the three-fold dilution of S1 and so forth until S4. In addition, two different control samples were run. In the *Si Control*, eggs were kept in vials with unloaded silicon rods. The *control* consisted of only eggs and passive dosing material that had been loaded in clean MeOH. Additional details on the experiments are provided in Donald *et al*. [30]. While numerous PAHs were included in Donald *et al*., only the samples from the exposures to phenanthrene were investigated in this thesis.

In both studies, three replicates (*A, B* and *C*) were made for each unique combination of exposure and level. However, some samples were lost due to laboratory error, or unavailable due to high mortality following exposure. In both studies, the eggs underwent exposure two days post fertilization. Exposure lasted three days. The samples from the two studies were extracted over four unique days. On each day of extraction, a quality control sample consisting of 20 blank eggs and a 115 ng/g ww spike was prepared. All samples were quantified with a quadratic polynomial with weights $1/x^2$ and four replicates per level.

Table 5: Extracted samples for WSF (WSF19) and passive dosing (PD18 and PD19) exposure studies.

Study	Exposure	Level	Replicate	n_{egg}	Extraction Day
PD18	PHN	S4	A	28	1
PD18	PHN	S4	B	24	2
PD18	PHN	S4	C	20	4
PD18	PHN	S3	C	19	4
PD18	PHN	S2	A	11	1
PD18	PHN	S1	C	20	4
PD18	Control		B	20	2
PD18	Control		C	20	4
PD19	PHN	S2	A	10	1
PD19	PHN	S2	B	10	2
PD19	PHN	S2	C	10	3
PD19	PHN	S1	A	10	1
PD19	PHN	S1	C	10	3
PD19	PHN	S0	B	10	2
PD19	Si Control		A	10	1
PD19	Si Control		B	10	2
PD19	Si Control		C	10	3
PD19	Control		A	10	1
PD19	Control		B	10	2
PD19	Control		C	10	3
WSF19	WSF	low	A	20	4
WSF19	WSF	low	B	20	2
WSF19	WSF	low	C	20	3
WSF19	WSF	med	A	20	4
WSF19	WSF	med	B	20	2
WSF19	WSF	med	C	20	3
WSF19	WSF	high	A	20	4
WSF19	SI0	low	A	20	1
WSF19	SI0	low	B	20	2
WSF19	SI0	low	C	20	3
WSF19	SI0	med	A	20	1
WSF19	SI0	med	B	20	2
WSF19	SI0	med	C	17	3
WSF19	SI0	high	A	20	1
WSF19	SI0	high	B	20	2
WSF19	SI0	high	C	20	3
WSF19	DCM Control		A	20	1
WSF19	DCM Control		B	20	2
WSF19	WSF Comb.	low	A	20	1
WSF19	WSF Comb.	low	B	12	4
WSF19	WSF Comb.	low	C	18	3
WSF19	WSF Comb.	med	A	12	1
WSF19	WSF Comb.	med	B	19	4
WSF19	WSF Comb.	med	C	19	3
WSF19	WSF Comb.	high	A	10	1
WSF19	WSF Comb.	high	B	11	4
WSF19	WSF Comb.	high	C	10	3

3.7   Method Optimization

The ESI source was optimized through a series of experimental designs that investigated five different ion source parameters. The chosen parameters included the gas temperature, gas flow, nebulizer pressure, capillary voltage and nozzle voltage. Each experiment injected 2 µL of the 5.00 ng/mL calibration standard from the response function. The peak area of 4-OHPHN was chosen to be representative of the ESI response. The experimental designs were analysed using the RSM package in R.

4 Results and Discussion

4.1 Estimating the Measurement Uncertainty from Method Validation Data

The following section derives selected equations for the estimation of MUs via validation data. These formulas are derived specifically for the validation work conducted in this thesis and are necessary for the evaluation of accuracy. All calculations are done in accordance with the *Guide to the expression of uncertainty in measurements* (GUM) [105].

4.1.1 Precision of a Method Estimate

Consider a method with negligible bias. If the MU is to be estimated for one single measurement, X, then its $1 - \alpha$ CI can be found from

$$U_{1-\alpha} = x \pm u(IP) \cdot t^*_{1-\alpha/2, \nu_{\text{eff}}}$$

However, the expression above only works for a single value. If the validation data is to be applied to the mean value of several measurements, the calculation of its MU becomes more complex.

Assume the mean value of an analysis is found by preparing J replicates per day across I days. This mean, the *grand mean* ($\bar{\bar{x}}$), can be found by averaging all $I \cdot J$ values. Equivalently, it can also be found by averaging I different means found within each day ($\bar{x}_i$):

$$\bar{\bar{x}} = \frac{\sum_{i=1}^{I} \sum_{j=1}^{J} x_{ij}}{IJ} = \frac{\sum_{i=1}^{I} \bar{x}_i}{I} \tag{59}$$

Consider first the variance of the within-day mean, $u^2(\bar{x})$. The total variance of $\bar{x}$ is a combination of two other variances. Firstly, each day is affected by some between-day variance, $u^2(b)$. In addition, within each day, each measurement is affected by the repeatability variance, $u^2(r)$ (Figure 20).
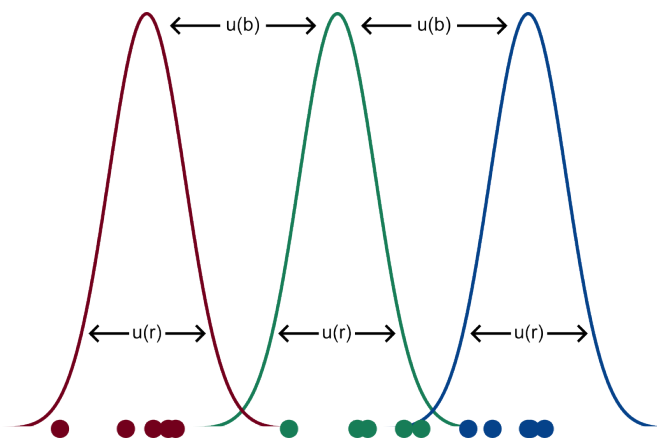
Figure 20: Illustration of the between- and within-day variances. Each colour corresponds to an unique day. Within any day i, the repeatability, $u^2(r)$, scatters measurements evenly about the within-day mean, $\bar{x}_i$. The between-day variance, $u^2(b)$, randomly shifts *all* measurements recorded within a particular day. Therefore, for instance, if a recorded measurement on a particular day is much smaller than usual, then the remaining measurements on that day will likely also be small.

A total of J replicate measurements are made *within* each day. In accordance with the standard error of the mean (8), these replicate measurements will reduce the within-day variance, *i.e.* the repeatability. Thus, the variance of the within-day mean, $u^2(\bar{x})$, becomes

$$u^2(\bar{x}) = u^2(b) + \frac{u^2(r)}{J} \tag{60}$$

Moreover, $\bar{\bar{x}}$ is the mean of I different within-day means. Therefore, again in accordance with the standard error of the mean, its variance, $u^2(\bar{\bar{x}})$, becomes

$$u^2(\bar{\bar{x}}) = \frac{u^2(\bar{x})}{I} = \frac{u^2(b) + u^2(r)/J}{I} = \frac{u^2(b)}{I} + \frac{u^2(r)}{IJ} \tag{61}$$

or as a standard uncertainty

$$u(\bar{\bar{x}}) = \sqrt{\frac{u^2(r)}{IJ} + \frac{u^2(b)}{I}} \tag{62}$$

From equations (61) and (62) it is possible to see that if only one measurement is made ($I = J = 1$), then the uncertainty simplifies to the IP, since

$$u(IP) = \sqrt{u^2(r) + u^2(b)} \tag{39}$$

Nonetheless, whenever replicate measurements are made, the uncertainty of the mean value cannot be expressed solely via $u(IP)$.[8] These conclusions have been noted before [105, 106], however, they are not widespread and are thus emphasised here.

---

[8]Unless $u^2(r) \gg u^2(b)$

4.1.2   Correcting for Bias

It is now necessary to distinguish the two different kinds of bias. Systematic errors can be found as *proportional bias*. That is, for the analytical method, the percentage bias remains consistent for all measurement results across its entire concentration range (*e.g.* results are always 90% of the true value). Alternatively, the method can be be affected by *constant bias*. Here the analytical method possesses some absolute bias which is consistent across its concentration range (*e.g.* all results are off the true value by 0.4 ng/mL). It is possible to neglect absolute bias if the measurement results are too large to be affected. Alternatively, some correction factor (CF) may be applied. For proportional bias, if the magnitude of the systematic error is too large, the bias can only be removed by the application of a CF.

The GUM establishes that all systematic errors present in the measurement result must either be negligible or corrected. Given some mean measurement result, $\bar{x}$, generated from a biased method, the corrected result, y, is provided by

$$y = f(\bar{x}, c_b, c_p) = (\bar{x} + c_b) \cdot c_p \tag{63}$$

Where c_b and c_p are CFs for constant and proportional bias, respectively. These CFs can be estimated from the linearity evaluation. During the investigation of linearity, estimated concentrations are regressed against the true concentrations. The regression coefficients thus provide information about the bias. The CF for the constant bias is estimated by $c_b = -b_0$ and the proportional bias by $c_p = 1/b_1$ [121, 122].

4.1.3   Combining the Uncertainty Components

After having related the random and systematic errors to the final measurement result (63), the MU can be calculated. To combine these uncertainty components, the law of propagation of uncertainty is applied:

$$u_c^2(y) = \sum_{i=1}^{n} a_i^2 u^2(x_i) \equiv \sum_{i=1}^{n} \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) \tag{47}$$

Substituting (63) into (47) yields the sensitivity coefficients, a_i:

$$a_1 = \frac{\partial f}{\partial \bar{x}} = c_p = \frac{y}{\bar{x} + c_b} \tag{64}$$

$$a_2 = \frac{\partial f}{\partial c_b} = c_p = \frac{y}{\bar{x} + c_b} \tag{65}$$

$$a_3 = \frac{\partial f}{\partial c_p} = \bar{x} + c_b = \frac{y}{c_p} \tag{66}$$

It is then possible to write combined uncertainty as

$$u_c^2(y) = \left(\frac{y}{\bar{x} + c_b}\right)^2 u^2(\bar{x}) + \left(\frac{y}{\bar{x} + c_b}\right)^2 u^2(c_b) + \left(\frac{y}{c_p}\right)^2 u^2(c_p)$$

which can be simplified to

$$u_c(y) = y \sqrt{\frac{u^2(\bar{x})}{(\bar{x} + c_b)^2} + \frac{u^2(c_b)}{(\bar{x} + c_b)^2} + \frac{u^2(c_p)}{c_p^2}}$$

On the basis of the identity derived in (61), it is possible to write

$$u_c(y) = y \sqrt{\frac{u^2(r)}{IJ(\bar{x} + c_b)^2} + \frac{u^2(b)}{I(\bar{x} + c_b)^2} + \frac{u^2(c_b)}{(\bar{x} + c_b)^2} + \frac{u^2(c_p)}{c_p^2}}$$

However, the between-day variance is not usually reported in validation studies, so the equation is better expressed as

$$u_c(y) = y \sqrt{\frac{u^2(r)}{IJ(\bar{x} + c_b)^2} + \frac{u^2(IP) - u^2(r)}{I(\bar{x} + c_b)^2} + \frac{u^2(c_b)}{(\bar{x} + c_b)^2} + \frac{u^2(c_p)}{c_p^2}} \tag{67}$$

Satterthwaite's approximation (48) can then be used to estimate the degrees of freedom for the uncertainty estimate:

$$\nu_{\text{eff}} \approx \frac{u_c^4(y)}{\sum_{i=1}^{I} \left(a_i^4 u^4(x_i)/\nu_i\right)} \tag{48}$$

The sensitivity coefficients obtained in (64) through (66) can be substituted into (48) to give a final expression for the degrees of freedom:

$$\nu_{\text{eff}} = u_c^4(y) \cdot \left[\left(\frac{c_p^2 \cdot u^2(r)}{IJ \sqrt{\nu_r}}\right)^2 + \left(\frac{c_p^2 \cdot \{u^2(IP) - u^2(r)\}}{I \sqrt{\nu_b}}\right)^2 + \left(\frac{c_p^2 \cdot u^2(c_b)}{\sqrt{\nu_c}}\right)^2 + \left(\frac{\{\bar{x} + c_b\}^2 \cdot u^2(c_p)}{\sqrt{\nu_c}}\right)^2\right]^{-1} \tag{68}$$

Where ν_r, ν_b and ν_c refers to the degrees of freedom for the repeatability, between-day variance and bias corrections, respectively.

Having estimated the degrees of freedom of the final measurement result, it is possible to create the expanded MU in accordance with[9]

$$U_{1-\alpha} = y \pm u_c(y) \cdot t^*_{1-\alpha/2,\nu_{\text{eff}}} \tag{49}$$

---

[9]Appendix A.3 presents an example that demonstrates how to use the derived equations to calculate MUs.

4.2   Method Validation

To evaluate the performance of the method, a total of seven validation parameters and a response function were investigated. The validation parameters included linearity, trueness, precision, accuracy, LoD, LoQ and selectivity. The validation work was conducted in accordance with the Eurachem guidelines [120].

4.2.1   Response Function

The response functions were chosen to be quadratic polynomials, $y = b_0 + b_1x + b_2x^2$, with weights, $w = 1/x^2$ (Table 6 and Figure 21). To construct and evaluate the response functions, a total of seven standard solutions in matrix extracts were prepared in a range of 0.100 to 30.0 ng/mL. There were four repeated injections per level.

Table 6: Regression coefficients and uncertainties for the response functions.

Compound	b_0	$u(b_0)$	b_1	$u(b_1)$	b_2	$u(b_2)$
1-OHPHN	-1.10e-03	1.13e-03	5.41e-02	5.50e-03	9.03e-04	3.69e-04
2-OHPHN	-4.93e-03	1.47e-03	1.02e-01	7.17e-03	7.79e-04	4.80e-04
3-OHPHN	-7.63e-03	2.51e-03	1.33e-01	1.23e-02	1.09e-03	8.23e-04
4-OHPHN	-3.09e-03	1.87e-03	8.51e-02	9.13e-03	1.27e-03	6.12e-04
9-OHPHN	-3.46e-03	8.32e-04	5.65e-02	4.07e-03	7.94e-04	2.73e-04
4-OHPHN-d9	-3.32e-03	1.19e-03	5.93e-02	5.83e-03	8.36e-04	3.91e-04
1,2-OH-1,2-HPHN	8.75e-02	6.82e-02	1.89e-02	1.39e-02	6.87e-04	4.33e-04
9,10-OH-9,10-HPHN	4.38e-02	5.80e-02	2.77e-02	1.18e-02	5.10e-04	3.68e-04
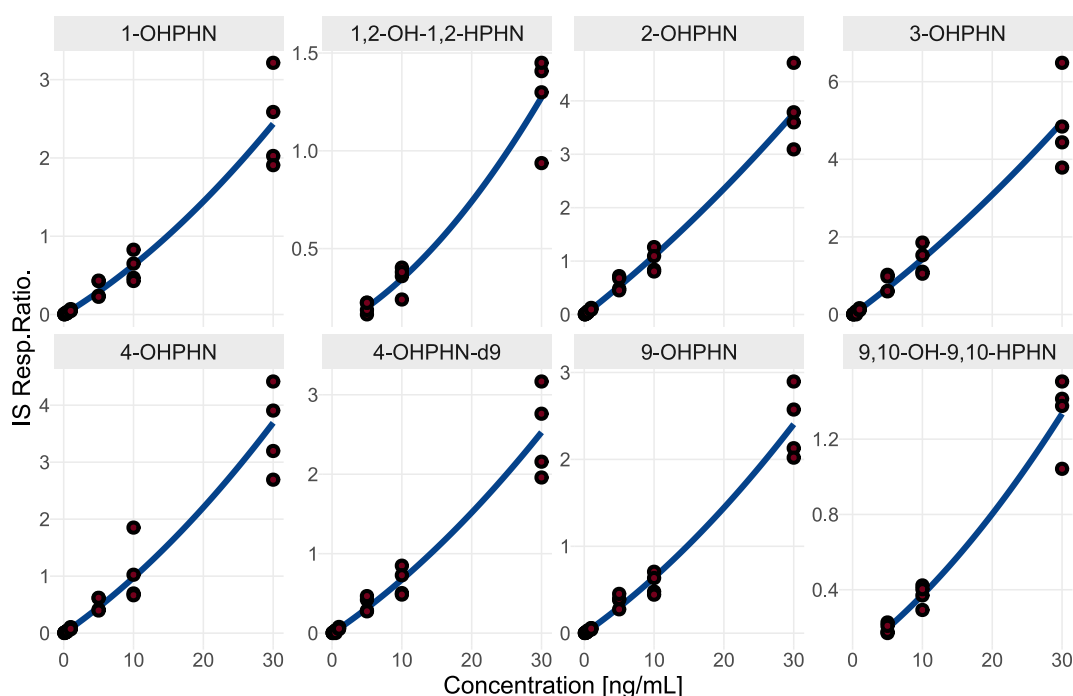
Figure 21: Final response functions, on the form $y = b_0 + b_1x + b_2x^2$ with weights, $w = 1/x^2$.

The monohydroxylated metabolites showed responses across the entire concentration range. However, the dihydrodiols only showed observable responses at the three highest concentration levels (5.00, 10.0 and 30.0 ng/mL) (Table 7). Most validation guidelines recommend a minimum of five to six standards when constructing a response function [120, 123]. Consequently, as the dihydrodiols fail to meet the recommended minimum number of calibration standards, the validation results of these metabolites can only be considered tentative.

The poor response of the dihydrodiols can possibly be attributed to acidic behaviour (distorting retention and detection) or compound reactivity (forming undetected adducts). This is discussed in more detail in Section 4.4.7.

Table 7: Instrumental responses at the concentration levels for the response function study, '✓' denotes presence and '-' denotes absence.

Compound	Concentration [ng/mL]						
	0.100	0.300	0.500	1.00	5.00	100	30.0
1-OHPHN	✓	✓	✓	✓	✓	✓	✓
2-OHPHN	✓	✓	✓	✓	✓	✓	✓
3-OHPHN	✓	✓	✓	✓	✓	✓	✓
4-OHPHN	✓	✓	✓	✓	✓	✓	✓
9-OHPHN	✓	✓	✓	✓	✓	✓	✓
4-OHPHN-d9	✓	✓	✓	✓	✓	✓	✓
1,2-OH-1,2-HPHN	-	-	-	-	✓	✓	✓
9,10-OH-9,10-HPHN	-	-	-	-	✓	✓	✓

After injecting each standard four times, the instrumental responses were plotted as a function of standard concentration. While the plots do not reveal any obvious curvature trends, the variance in instrumental response clearly increases with concentration, demonstrating heteroscedasticity (Figure 22). This can also be assessed more formally via Levene's test. The results from Levene's test, alongside other relevant statistical tests discussed below, are presented at the end of this section (Table 8). For the monohydroxylated metabolites, the suspicion of heteroscedasticity is confirmed by the small p-values (rejecting the null hypothesis of homoscedasticity). For the dihydrodiols, however, homoscedasticity can not rejected by Levene's test as both p-values are greater than 0.05. Nonetheless, it will here be assumed that lack of statistical significance is a consequence of the smaller number of data points, and that the dihydrodiols are also heteroscedastic.
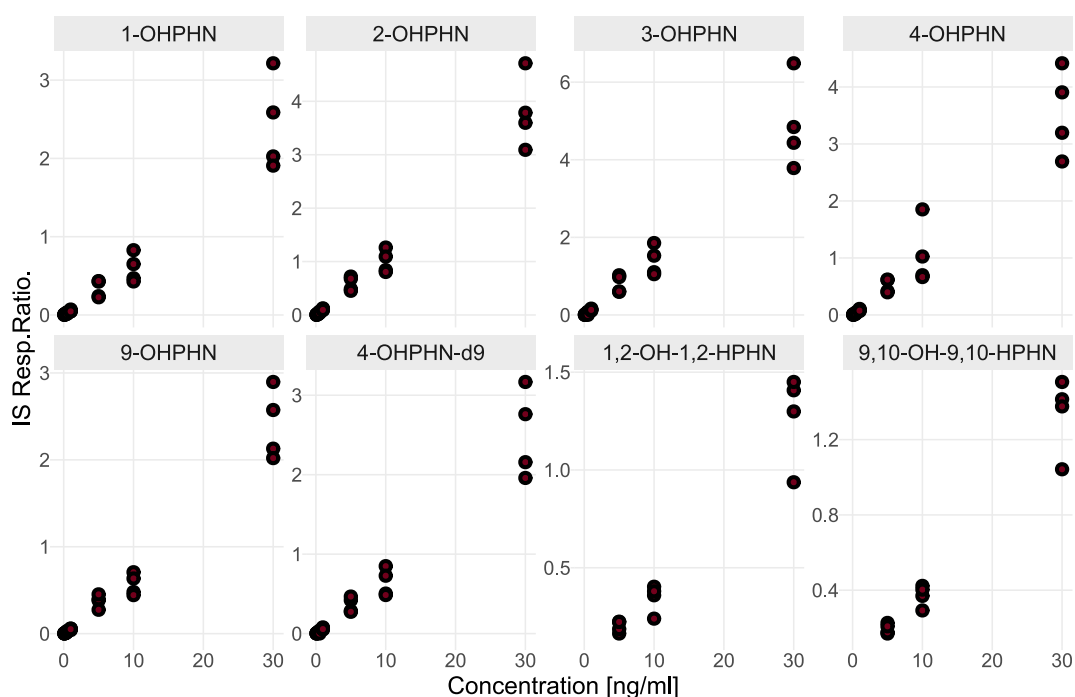
Figure 22: Scatterplots for the response functions. The variance of the instrumental noise increases with concentration, suggesting heteroscedasticity. No curvature trends are immediately obvious in the raw calibration data.

Since Figure 22 does not express any strong curvature, initially, a weighted FOM will be constructed and the residuals inspected. Curvature in the residuals indicates that the response functions require a quadratic term. Generally, the simplest model is the best choice when constructing a response function. However, the heteroscedastic nature of the data warrants the application of weights. The choice of weights will be limited to the options available in the MassHunter Quantification software. This is to minimize the use of additional software. The most common weights in the software include $1/x$, $1/x^2$ and $1/s^2$. Consequently, these are the ones that will be considered here.

For the sake of illustration, the residual plots for the weight $1/s^2$ are shown in Figure 23. The residuals at the lower concentrations are smaller than than those found at the higher ones by several orders of magnitude. This makes it difficult to discover any trends in the residuals. This holds true for the weights $1/x$ and $1/x^2$ as well (Appendix B).
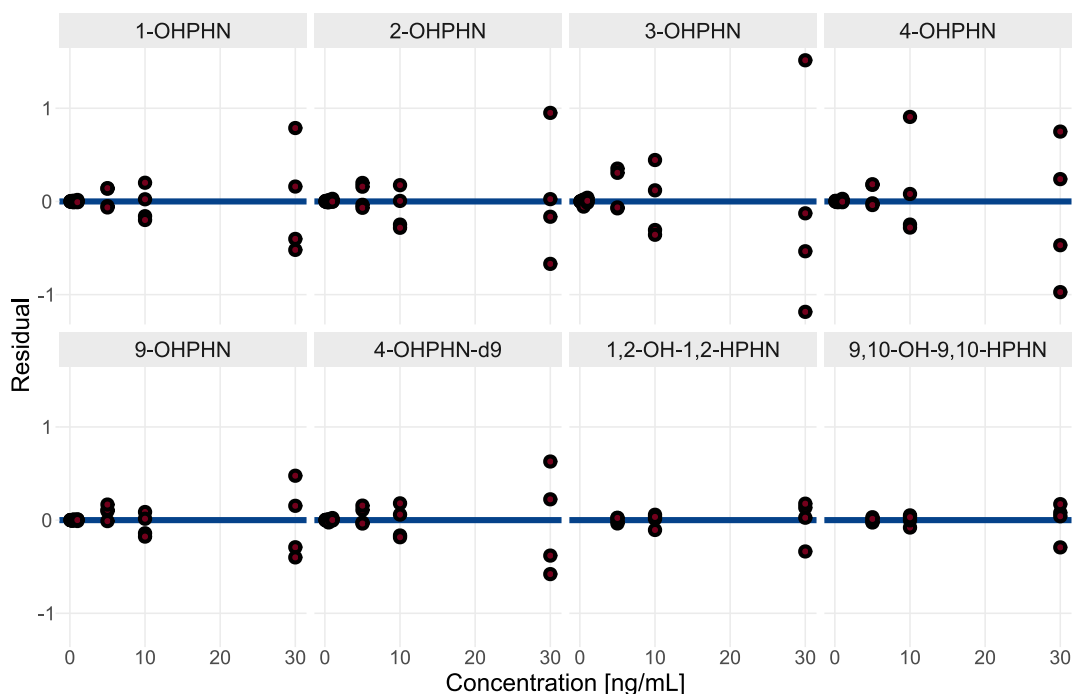
Figure 23: Residual plots for FOMs with weights, $w = 1/s^2$. Due to the large differences between the residuals at the different concentration levels, it is difficult to notice any trends.

To remedy this problem, each residual (e), can be multiplied by the square-root of its associated weight (w) providing the *standardized residuals*, e_w:

$$e_w = e \cdot \sqrt{w} \tag{69}$$

The standardized residual plots for $1/s^2$ show no obvious trends (Figure 24). However, some slight curvature is visible at the three highest concentration levels. This is especially noticeable for 9-OHPHN. This could indicate the necessity of a quadratic term in the response function, but more evidence is necessary to confirm the non-linearity of the data. Consultation of the standardized residual plots associated with $1/x$ and $1/x^2$ (Appendix B) does not yield any additional insight as the distribution of the residuals are essentially identical across the three weights.
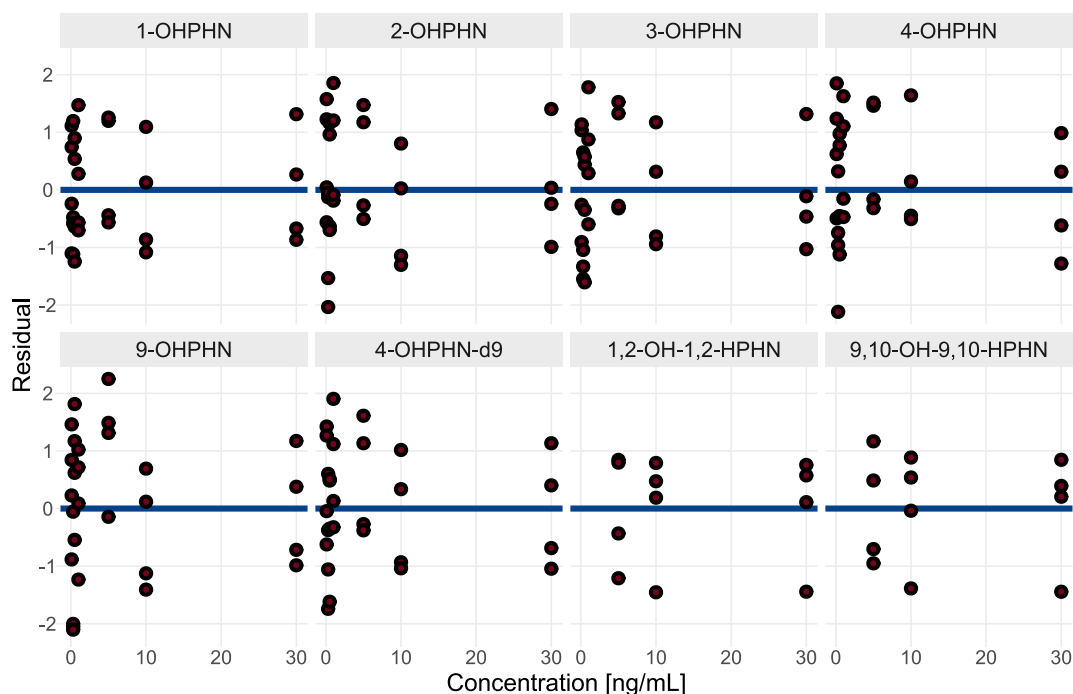
Figure 24: Standardized residual plots for FOMs with weights, $w = 1/s^2$. The application of standardized residuals makes it possible to see trends. The plots reveal slight curvature at the three upper levels.

The final insight in the curvature evaluation is provided by Mandel's test (Table 8). The test provides several significant p-values, suggesting that the null hypothesis of a FOM should be rejected for a quadratic fit. For $1/s^2$ and $1/x^2$, significance is observed for 1-, 4- and 9-OHPHN, as well as 4-OHPHN-d9. For $1/x$, significance is observed for 1-, 2- and 9-OHPHN, in addition to 4-OHPHN-d9. Thus, only 3-OHPHN and the dihydrodiols show no significance for any weight. Nonetheless, for the dihydrodiols, with only three concentration levels, it is difficult to rigorously determine whether curvature is present or not. Additionally, given the other monohydroxylated metabolites have shown some evidence for curvature, it reasonable to assume 3-OHPHN does too, despite non-significance. Therefore, for the sake of simplicity and consistency, all compounds are assumed to be sufficiently modelled by a weighted polynomial.

The final step in establishing a response function is selecting a weighting term. To do this, back calculations of the calibration standards are performed for each weight, and the results are compared. Scatterplots illustrating the percentage error found from the back calculations are provided in Figure 25 at the end of this section. In the plots for $1/s^2$ and $1/x^2$, results are essentially identical; the errors are random, of similar magnitude, and centered about zero. For the lower concentration levels of $1/x$, the errors are more poorly centered than those of the two other weights. The choice of weights can therefore be limited to $1/s^2$ and $1/x^2$. Moreover, as $1/x^2$ does not require any replicate measurements when constructing the response function, this weight is preferable over $1/s^2$.

Therefore, the selected response function for the method will take the form of $y = b_0 + b_1x + b_2x^2$, with weights, $w = 1/x^2$.
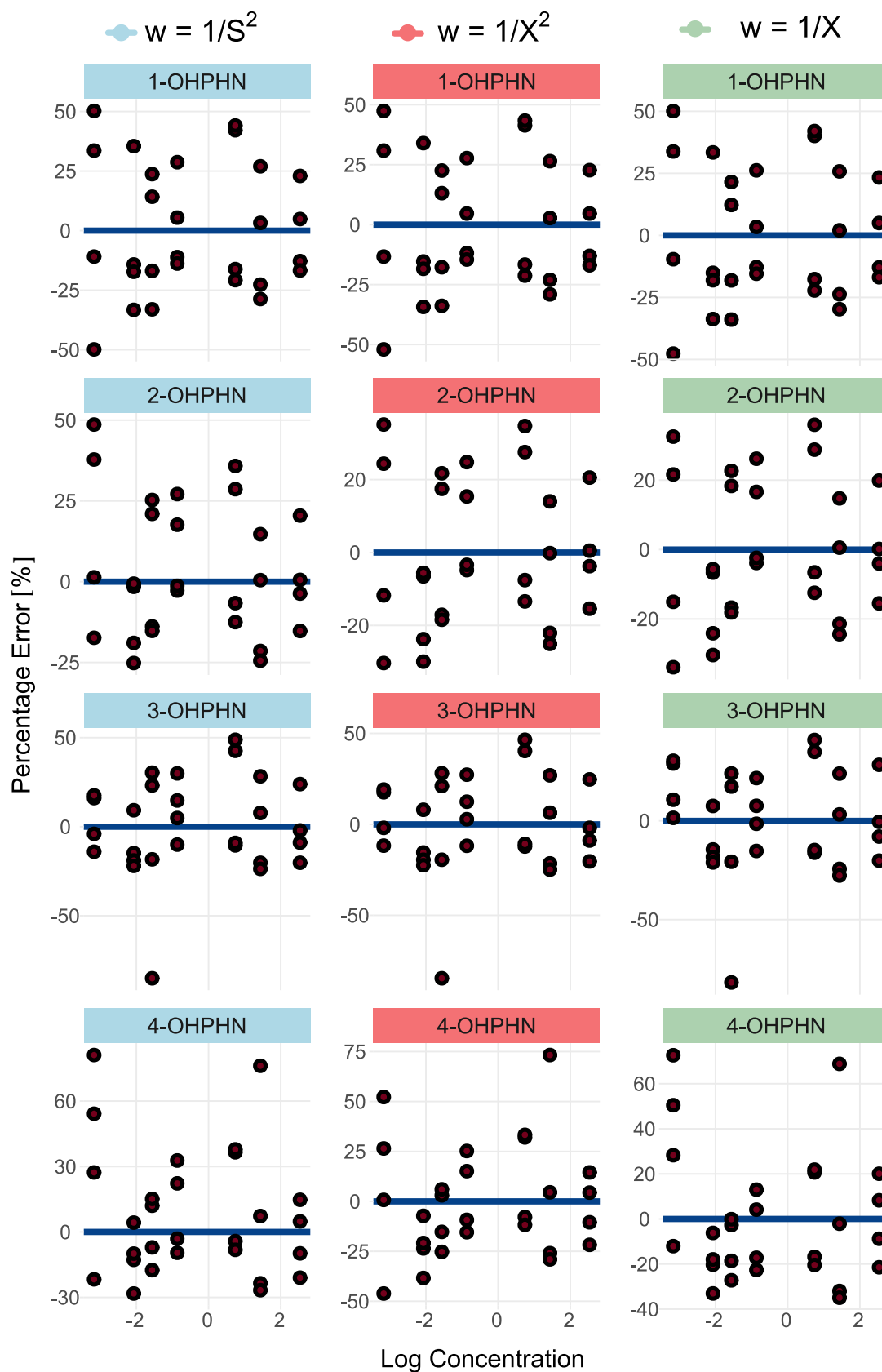
Figure 25: Back calculation errors for quadratic polynomials with various weights. The percentage error between standard concentration and response function estimate is plotted against the standard concentration on a logarithmic scale. The figure provides errors for the weights $1/s^2$ (blue), $1/x^2$ (red) and $1/x$ (green). The error distributions between all three weights are similar.

Figure 25: Back calculation errors for quadratic polynomials with various weights (Continued).

Table 8: Outcomes for Levene and Mandel's tests for the response function evaluation.

Compound	Levene's Test p	Mandel's Test $1/s^2$ p	$1/x$ p	$1/x^2$ p
1-OHPHN	1.47e-05 ***	1.85e-02 *	1.76e-02 *	2.16e-02 *
2-OHPHN	5.92e-03 **	1.09e-01	4.45e-02 *	1.18e-01
3-OHPHN	2.97e-03 **	1.17e-01	3.12e-01	1.96e-01
4-OHPHN	5.85e-05 ***	1.65e-02 *	2.59e-01	4.89e-02 *
9-OHPHN	2.56e-07 ***	6.71e-03 **	8.76e-03 **	7.46e-03 **
4-OHPHN-d9	8.80e-08 ***	2.64e-02 *	3.26e-02 *	4.24e-02 *
1,2-OH-1,2-HPHN	6.29e-02	1.59e-01	2.42e-01	1.47e-01
9,10-OH-9,10-HPHN	5.86e-02	1.98e-01	3.06e-01	1.99e-01

4.2.2 Linearity

Linearity was demonstrated for the monohydroxylated metabolites in the range of 5.73 to 229 ng/g ww with the weights $1/s^2$ (Figure 26). The evaluation of linearity was done via experiments from the accuracy study. Here, blank eggs were spiked at four different concentration levels. Two spikes were performed at each level and this was repeated across five days. These experiments were also used to estimate the trueness, precision, accuracy, LoQ and selectivity.
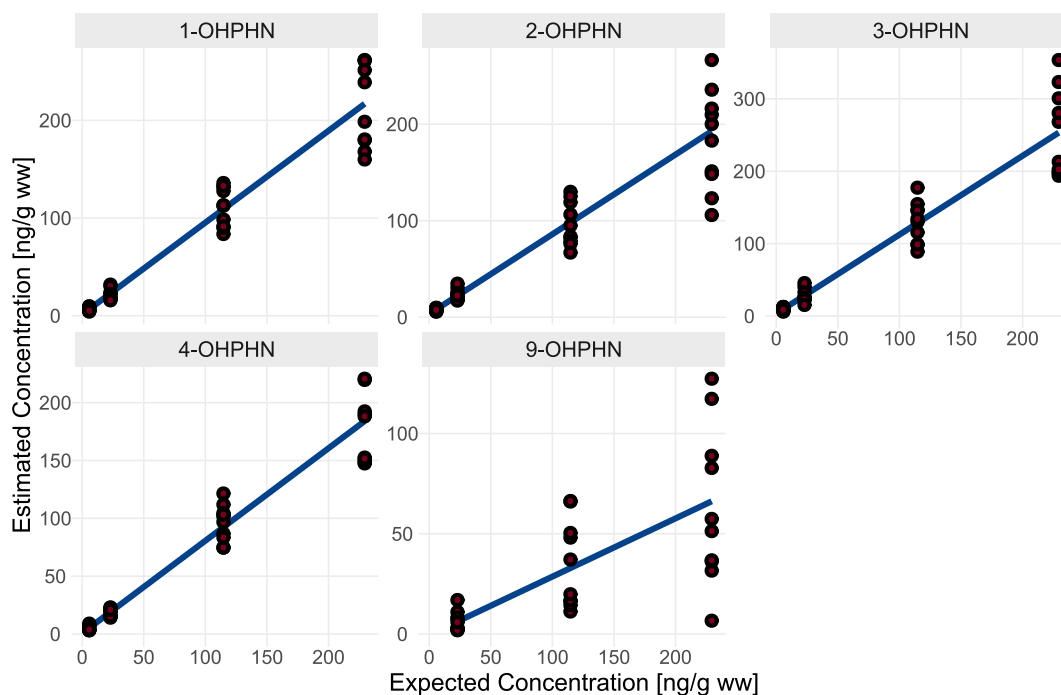
Figure 26: Final linearity graphs, on the form $y = b_0 + b_1 x$ with weights, $w = 1/s^2$.

The compounds 1,2-OH-1,2-HPHN and 9,10-OH-9,10-HPHN only provided instrumental responses at two levels. Moreover, 9-OHPHN could only provide instrumental responses at the

three highest levels. The remaining monohydroxylated phenanthrenes provided instrumental responses at all levels (Table 9). A regression line between two points always forms a straight line; a minimum of three levels is thus required for investigation of potential curvature. Therefore, only the monohydroxylated metabolites will undergo a linearity assessment. By extension, these compounds are also the only ones that will undergo complete method validation.

Table 9: Instrumental responses at the validated levels for the accuracy study, '✓' denotes presence and '-' denotes absence.

Compound	Concentration [ng/g ww]			
	5.73	22.9	115	229
1-OHPHN	✓	✓	✓	✓
2-OHPHN	✓	✓	✓	✓
3-OHPHN	✓	✓	✓	✓
4-OHPHN	✓	✓	✓	✓
9-OHPHN	-	✓	✓	✓
1,2-OH-1,2-HPHN	-	-	✓	✓
9,10-OH-9,10-HPHN	-	-	✓	✓

Consulting the initial scatterplot, the estimated concentrations are seemingly linear when plotted against the expected concentrations. Moreover, as the calibration standards showed heteroscedastic behaviour, it is expected that the validation samples do too. Again, the suspicion of heteroscedasticity is essentially confirmed by visual evaluation; the variance of the estimated concentrations increase drastically alongside the expected concentrations (Figure 27). Nonetheless, visual evaluation of heteroscedasticity is again accompanied by formal statistical testing by means of Levene's test. The outcomes, which are presented at the end of this section (Table 11), reject the null hypothesis of homoscedasticity for all monohydroxylated compounds.
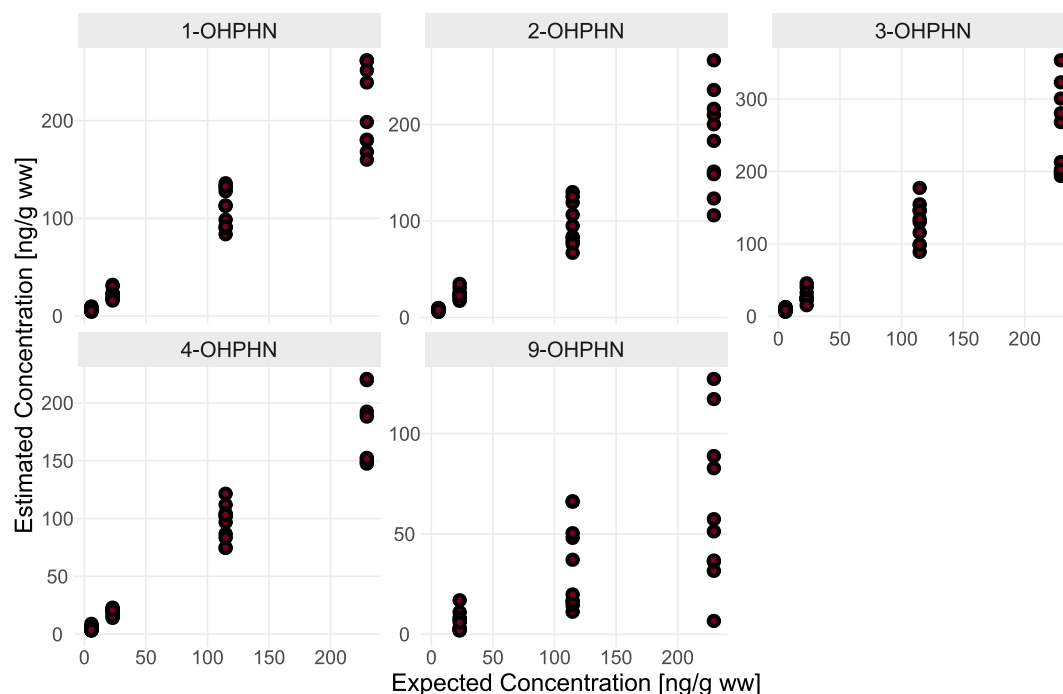
Figure 27: Scatterplots of estimated concentrations against expected concentrations. The concentration estimates as a function of the expected concentrations are seemingly linear in the investigated range. The data are also heteroscedastic, as evidenced by an increasing spread in concentration estimates alongside spike concentrations.

To further investigate the linearity of the validation data, a FOM may be fitted. However, as the data is heteroscedastic, a weighting term must be added to the model. Consider again the three weights $1/x$, $1/x^2$ and $1/s^2$. The choice of weight has a minimal effect on the produced regression estimates (Table 10 and Figure 28). For both b_0 and b_1, the regression coefficients remain approximately constant for all weights. However, in terms of b_0's standard error, $u(b_0)$, the weight $1/x$ possesses much greater uncertainties than the two others. For $u(b_1)$, the differences in errors are negligible.

Table 10: Regression coefficients and uncertainties for the linearity graph candidates.

Compound	$1/x^2$				$1/x$				$1/s^2$			
	b_0	$u(b_0)$	b_1	$u(b_1)$	b_0	$u(b_0)$	b_1	$u(b_1)$	b_0	$u(b_0)$	b_1	$u(b_1)$
1-OHPHN	1.350	0.556	0.932	0.050	1.138	1.304	0.944	0.032	1.232	0.679	0.941	0.038
2-OHPHN	3.243	0.532	0.830	0.048	3.815	1.494	0.797	0.037	3.314	0.590	0.826	0.043
3-OHPHN	3.359	0.772	1.087	0.070	3.284	1.851	1.090	0.046	3.331	0.874	1.090	0.055
4-OHPHN	0.527	0.480	0.803	0.043	0.817	0.972	0.786	0.024	0.659	0.652	0.800	0.029
9-OHPHN	-0.476	1.965	0.292	0.051	-0.146	3.456	0.286	0.043	-0.424	2.058	0.290	0.047
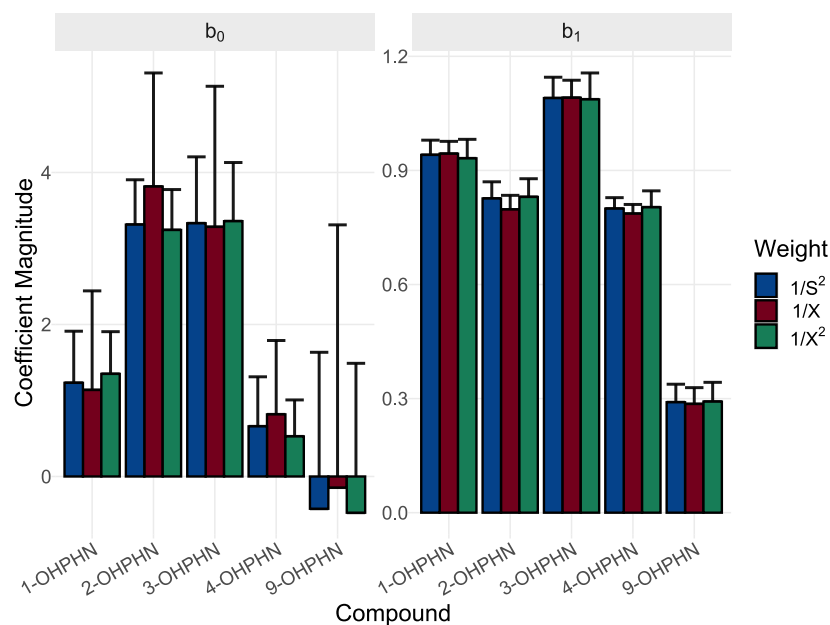
Figure 28: Regression coefficients for linearity graph candidates. Regression coefficients are presented alongside their standard error. The bars in in blue corresponds to regression estimates generated with weights $1/s^2$, plots in red with $1/x$ and green with $1/x^2$. All compounds have essentially identical regression coefficients. The standard uncertainties for the intercept using $1/s^2$ and $1/x^2$ are comparable, while the corresponding errors for $1/x$ are significantly higher. The standard uncertainties in the slope are similar across all weights.

The selection of $1/s^2$ as the weighting term is reasonable as it consistently produces small errors across both b_0 and b_1. This choice is further supported by the residual plots (Figure 29). While all three weights provide residuals with a center about zero, $1/x$ provide standardized residuals which 'fan out' alongside the concentration. Additionally, the residuals seem to be more evenly spread for $1/s^2$ when compared to $1/x^2$. This is especially noticeable for 4-OHPHN, where the lowest level has a much higher spread than the three remaining ones. While neither of these observations violate the WLS assumptions, it is still reasonable to assume that $1/s^2$ likely models the data better as its weighting provides the most even spread of residuals.

Additionally, for all three weights, Mandel's test fails to reject linearity (Table 11). Therefore, on the basis that linearity could not be rejected and the residual plots did not show any curvature, the method can be deemed sufficiently linear. The method linearity is assumed to be adequately modelled by $y = b_0 + b_1 x$ with weights, $w = 1/s^2$.
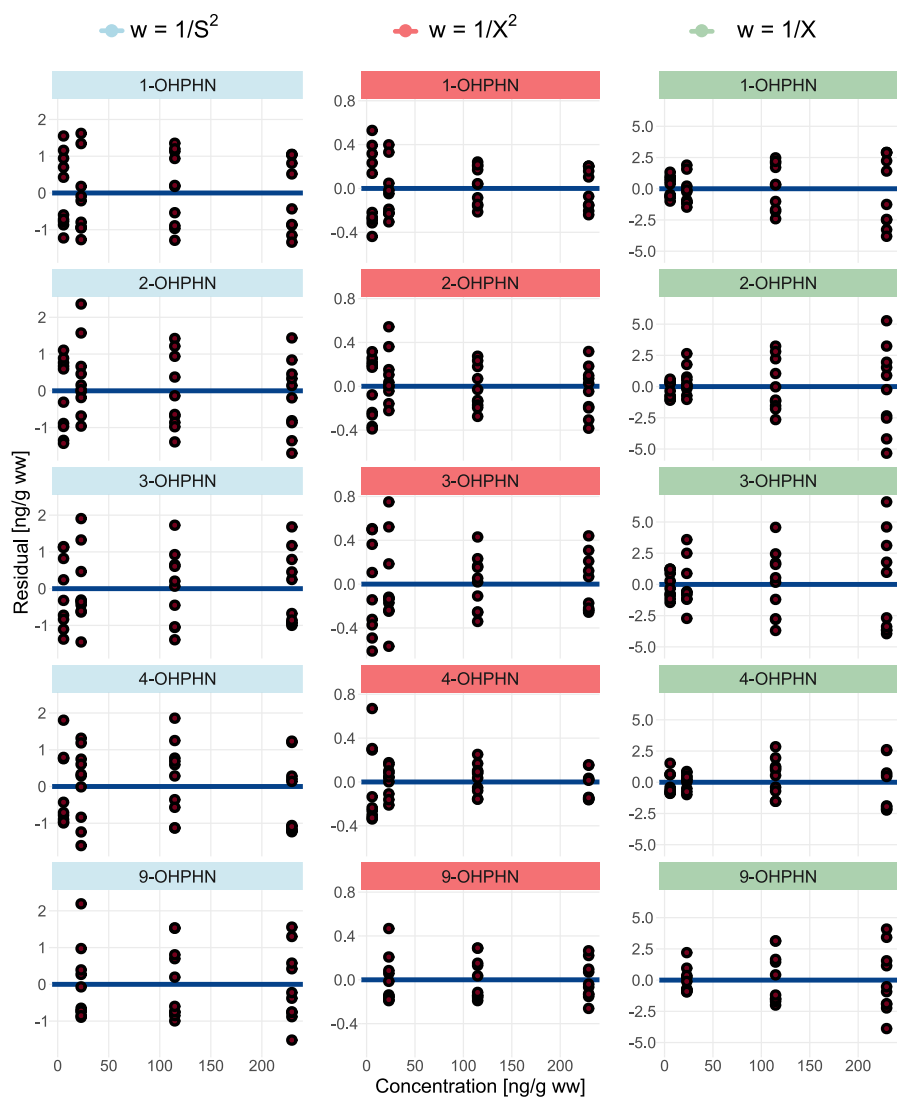
Figure 29: Standardized residuals for linearity graphs for the weights $1/s^2$ (blue), $1/x^2$ (red) and $1/x$ (green). The error distributions between all three weighting terms are similar, although the most even spread is achieved for $1/s^2$.
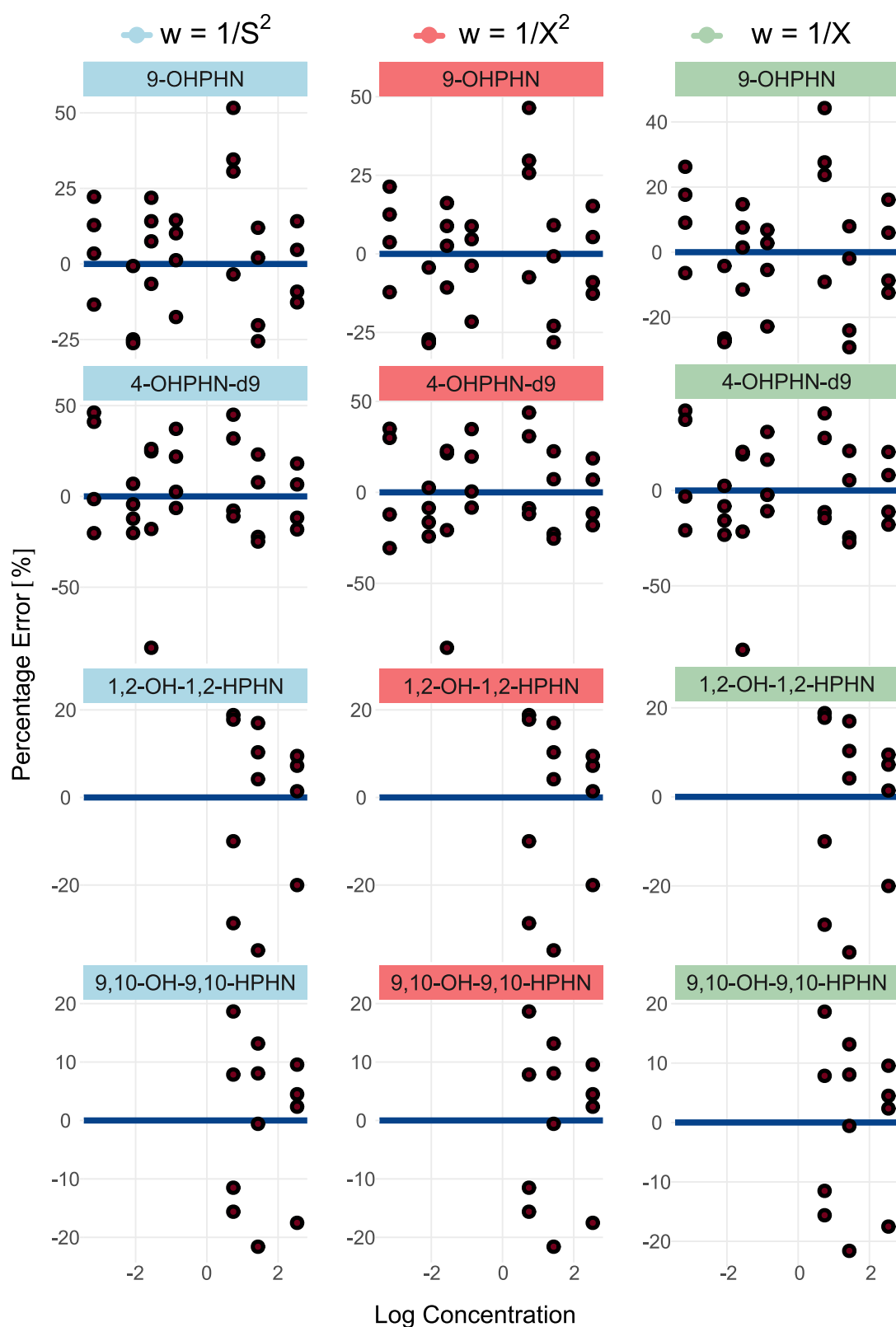
Table 11: Outcomes for Levene and Mandel's tests for the linearity evaluation of the monohydroxylated metabolites.

Compound	Levene's Test	Mandel's Test		
		$1/s^2$	$1/x$	$1/x^2$
	p	p	p	p
1-OHPHN	2.53e-12 ***	8.04e-01	8.94e-01	7.03e-01
2-OHPHN	2.10e-06 ***	2.78e-01	2.84e-01	5.48e-01
3-OHPHN	6.92e-08 ***	9.23e-01	9.60e-01	9.88e-01
4-OHPHN	3.98e-08 ***	5.45e-01	3.31e-01	1.97e-01
9-OHPHN	1.66e-03 **	7.57e-01	7.43e-01	7.22e-01

4.2.3 Trueness

The final recoveries of the method were close to 100%, ranging from 94.7 to 109.5% (Table 12). For the monohydroxylated metabolites, this was demonstrated at all validated levels.

Table 12: Corrected recoveries [%] and biases [%] for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	Recovery [%]	Bias [%]
1-OHPHN	5.73	102.2	2.2
	22.9	94.7	-5.3
	115	102.4	2.4
	229	99.6	-0.4
2-OHPHN	5.73	96.9	-3.1
	22.9	109.5	9.5
	115	98.3	-1.7
	229	95.4	-4.6
3-OHPHN	5.73	100.4	0.4
	22.9	98.4	-1.6
	115	100.6	0.6
	229	99.9	-0.1
4-OHPHN	5.73	97.1	-2.9
	1	101.3	1.3
	5	103.9	3.9
	229	95.7	-4.3
9-OHPHN	22.9	99.3	-0.7
	115	105.3	5.3
	229	96.2	-3.8

In order to achieve final recoveries above, the method needs to apply CFs which account for bias. If the method only accounts for surrogate corrections, the recoveries are much poorer (Table 13 and Figure 30). Most compounds have biases of a similar magnitude at around ±20%, although 9-OHPHN distinguishes itself with a bias around −70%. Nonetheless, some common trends in the bias are observed. For all compounds except 2-OHPHN, the magnitude of the bias changes significantly when moving from 5.73 to 22.9 ng/g ww, after which it stabilizes. This suggests that a constant bias with considerable magnitude is present at 5.73 ng/g ww, but becomes negligible around 22.9 ng/g ww, with proportional bias becoming the primary source of systematic error thereafter. For 2-OHPHN, proportional bias dominates at around 115 ng/g ww.

Table 13: Uncorrected recoveries [%] and biases [%] for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	Recovery [%]	Bias [%]
1-OHPHN	5.73	117.7	17.7
	22.9	94.5	-5.5
	115	97.4	-2.6
	229	94.3	-5.7
2-OHPHN	5.73	137.9	37.9
	22.9	104.9	4.9
	115	84.1	-15.9
	229	80.2	-19.8
3-OHPHN	5.73	167.6	67.6
	22.9	121.8	21.8
	115	112.5	12.5
	229	110.4	10.4
4-OHPHN	5.73	89.1	-10.9
	22.9	83.9	-16.1
	115	83.6	-16.4
	229	76.7	-23.3
9-OHPHN	22.9	30.2	-69.8
	115	27.0	-72.0
	229	27.6	-72.4
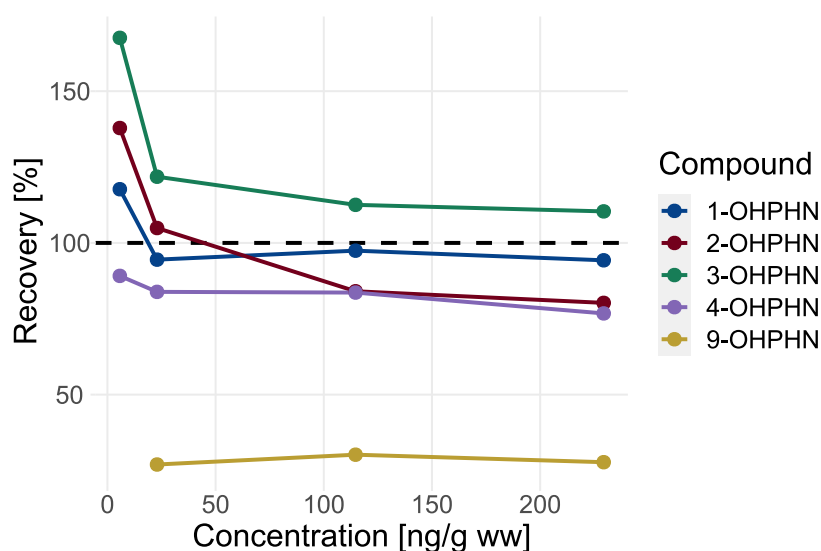
Figure 30: Non-corrected recoveries against spike concentrations. Dashed black line corresponds to 100% recovery. No compound completely aligns with the 100% recovery line, meaning every compound possess some systematic error. The 9-OHPHN recovery is considerably worse than the remaining analytes.

The necessity of bias corrections can be revealed by the regression coefficients from the linearity study (Table 10). If there is constant bias, the intercept will likely differ significantly from zero. Similarly, if there is proportional bias, the slope will differ significantly from unity. The two sets of hypotheses, which evaluate constant and proportional bias, become

$$H_0 : \beta_0 = 0 \qquad\qquad H_0 : \beta_1 = 1$$
$$\text{and}$$
$$H_a : \beta_0 \neq 0 \qquad\qquad H_a : \beta_1 \neq 1$$

For the compounds 1-, 2- and 3-OHPHN, the p-values for the constant bias (p_0) are significant (Table 14). This agrees with Figure 30; all of these compounds possess large jumps in recovery upon moving from 5.73 to 22.9 ng/g ww. Similarly, the p-values for proportional bias (p_1) are significant for 2-, 4- and 9-OHPHN. This again agrees well with the figure, as all of these compounds lie well below 100% recovery.

Table 14: p-values for constant and proportional bias for the monohydroxylated metabolites.

Compound	p_0	p_1
1-OHPHN	3.88e-02*	1.35e-01
2-OHPHN	9.56e-07***	3.17e-04***
3-OHPHN	2.47e-04***	1.09e-01
4-OHPHN	1.59e-01	2.85e-08***
9-OHPHN	5.81e-01	1.66e-17***

All monohydroxylated metabolites have significant bias, but only 2-OHPHN has both constant and proportional bias. As stated by the GUM [105], CFs must be applied to ensure that the final measurement results are free for systematic errors. In order to keep consistency in the applied formulas (Section 4.1), all compounds will be corrected for both constant and proportional bias. The corrected results are found via

$$y = (\bar{x} + c_b) \cdot c_p \tag{63}$$

Where $\bar{x}$ is the mean measurement result, y is the corrected result, moreover c_p and c_b represent the CFs for proportional and constant bias, respectively.

The linearity study can also be used to estimate the relevant CFs (Table 15). Since the slope corresponds to the proportional bias and the intercept to the absolute bias, the CFs become $c_p = 1/b_1$ and $c_b = -b_0$. For c_b, the uncertainty in the CF remains the same as for b_0. However, for c_p, the uncertainty must be rewritten in accordance to the rules of uncertainty propagation.[10]

---

[10]Since $c_p = 1/b_1$, then from (47), $u^2(c_p) = \left(\frac{\partial c_p}{\partial b_1}\right)^2 \cdot u^2(b_1) = \left(\frac{\partial}{\partial b_1}\frac{1}{b_1}\right)^2 \cdot u^2(b_1) = \left(-\frac{1}{b_1^2}\right)^2 \cdot u^2(b_1) = \frac{u^2(b_1)}{b_1^4}$

Table 15: Constant and proportional bias CFs with corresponding uncertainties for the monohydroxylated metabolites.

Compound	c_p	$u(c_p)$	c_b	$u(c_b)$
1-OHPHN	1.062	0.043	-1.232	1.304
2-OHPHN	1.211	0.063	-3.314	1.494
3-OHPHN	0.917	0.046	-3.331	1.851
4-OHPHN	1.250	0.045	-0.659	0.972
9-OHPHN	3.448	0.559	0.424	3.456

In addition to its significance, it is also beneficial to evaluate why bias is present in the method. If the cause is known, future method optimization can possibly reduce the bias or even remove it in its entirety. Firstly, some bias is possibly attributed to the surrogate correction; the recovery of 4-OHPHN-d9 need not perfectly reflect the recoveries of the analytes. However, most of the bias is likely a consequence of matrix effects. The matrix effects are seemingly similar for 1-, 2-, 3- and 4-OHPHN, but differ drastically from 9-OHPHN. Similar results have also been observed in two other methods for the analysis of PAH metabolites by Da Silva *et al.* and Onyemauwa *et al.* [124, 125]. This could suggest that the poor recovery is an intrinsic property of the 9-OHPHN metabolite.

It has been previously reported that the double bond between carbons 9 and 10 in phenanthrene is highly reactive. [126, 127]. This reactivity may be enhanced when phenanthrene is oxidised to 9-OHPHN. This kind of behaviour can be seen in the closely related benzene. When benzene is oxidized to phenol, its reactivity increases considerably. This is because the hydroxy group introduces additional electrons that stabilizes reaction intermediates via resonance [128, 129]. Therefore, if the additional hydroxy group amplifies the reactivity of an already sensitive double bond, the 9-OHPHN metabolite could become very susceptible to reactions with other matrix components in the ESI, forming non-detected adducts. In contrast, the milder reactivities of the 1,2 and 3,4 bonds [130] suggest that the extent of non-detected adducts could be smaller for these metabolites. Therefore, the differences in biases can possibly be attributed to compound reactivity. Although this remains speculative.

4.2.4   Precision

The precision of the method can be reported for both the uncorrected and corrected results, the former of which is required to calculate the final MU (Section 4.1). For the corrected results, the repeatability (RSD_r) was found to vary from 6.1 to 50.6% and the IP (RSD_{IP}) from 18.8 to 76.4%. The uncorrected results displayed a repeatability ($RSD_{n,r}$) ranging from 6.1 to 54.0% and an IP ($RSD_{n,IP}$) from 16.3 to 81.6% (Table 16 and Figure 31).

Table 16: Repeatability [%] and IP [%] for corrected (RSD_r and RSD_{IP}) and non-corrected ($RSD_{n,r}$ and $RSD_{n,IP}$) concentration estimates for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	$RSD_{n,r}$ [%]	$RSD_{n,IP}$ [%]	RSD_r [%]	RSD_{IP} [%]
1-OHPHN	5.73	15.3	31.0	18.7	37.9
	22.9	8.5	27.1	9.0	28.8
	115	8.3	18.5	8.4	18.8
	229	8.5	20.7	8.6	20.8
2-OHPHN	5.73	14.9	20.7	25.7	35.7
	22.9	10.8	23.0	12.5	26.7
	115	13.5	24.2	14.0	25.0
	229	8.5	29.4	8.7	29.9
3-OHPHN	5.73	19.1	27.3	29.2	41.7
	22.9	13.6	33.8	15.4	38.4
	115	13.0	22.9	13.3	23.5
	229	10.1	24.7	10.3	25.1
4-OHPHN	5.73	28.8	41.7	33.1	47.8
	22.9	10.3	16.3	10.7	16.9
	115	6.1	17.2	6.1	17.3
	229	7.4	17.7	7.4	17.8
9-OHPHN	22.9	54.0	81.6	50.6	76.4
	115	35.9	65.4	35.5	64.6
	229	31.2	64.5	31.0	64.1
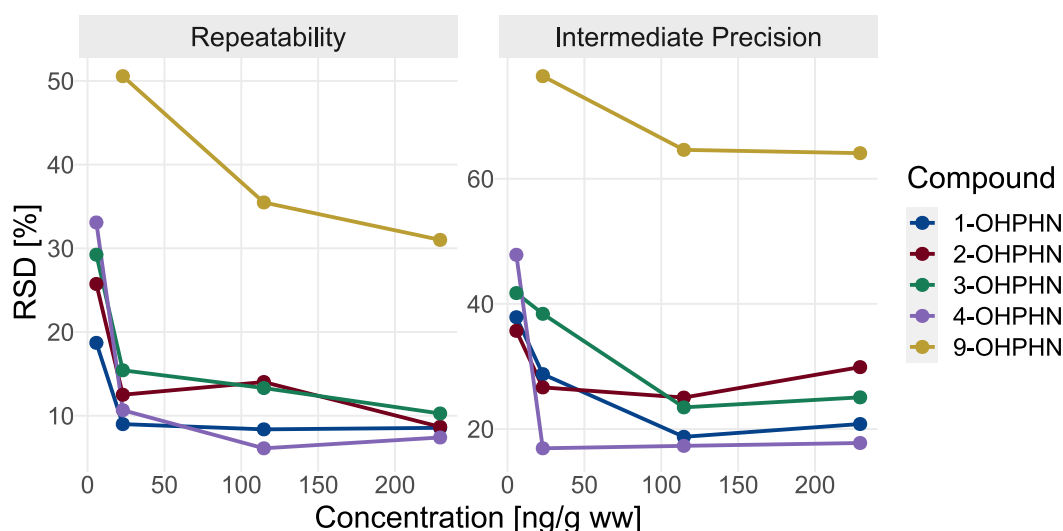
Figure 31: RSD_r and RSD_{IP} for the corrected measurement results. The precision components drastically decrease before stabilising when moving from a low to high concentration. The RSDs of 9-OHPHN are notably higher than those of the other compounds.

The table and figure show an expected pattern, the method has high RSDs at low concentrations, but they reduce and stabilise as the concentration increases. Again, the results from 9-OHPHN stands out; its precision is considerably worse than any other compound. This is true for both the repeatability and IP. This is a consequence of the very poor instrumental response found for

9-OHPHN at all levels. The weak analyte signal renders it very susceptible to baseline noise, ensuring a poor precision.

To further assess the quality of the method precision, the calculated RSDs can be compared to precisions found from Horwitz equation. This is done calculating the *HorRat*; the ratio between the estimated and predicted precisions. In general, values of the *HorRat* should not exceed two [108, 110]. Horwitz predicts IPs of 56, 45, 36 and 32 % for the levels 5.73, 22.9, 115 and 229 ng/g ww, respectively. By comparing these to the RSDs in Table 16, it is found that the *HorRat*s for 1-, 2-, 3- and 4-OHPHN do not exceed one, while the largest *HorRat* for 9-OHPHN is two (Table 17). Although 9-OHPHN meets the minimum criteria, it does so barely. Some further optimization to the method may therefore be reasonable. However, for the time being, the precision of 9-OHPHN is considered satisfactory.

Table 17: HorRat estimates for the repeatability (*HorRat$_r$*) and IP (*HorRat$_{IP}$*) of the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	$HorRat_r$	$HorRat_{IP}$
	5.73	0.50	0.68
1-OHPHN	22.9	0.30	0.64
	115	0.35	0.53
	229	0.40	0.65
	5.73	0.69	0.64
2-OHPHN	22.9	0.41	0.59
	115	0.59	0.70
	229	0.41	0.93
	5.73	0.79	0.75
3-OHPHN	22.9	0.51	0.85
	115	0.56	0.66
	229	0.48	0.78
	5.73	0.89	0.86
4-OHPHN	22.9	0.35	0.37
	115	0.26	0.49
	229	0.35	0.56
	22.9	1.68	1.69
9-OHPHN	115	1.50	1.82
	229	1.45	2.00

4.2.5 Accuracy

The standard combined MUs were calculated by the formulas derived in Section 4.1. The standard uncertainties ranged from 17.7 to 84.1 % (Table 18).

Table 18: Standard uncertainties [%] for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	RSD_u [%]
1-OHPHN	5.73	40.0
	22.9	29.3
	115	19.2
	229	21.2
2-OHPHN	5.73	38.3
	22.9	27.3
	115	25.6
	229	30.4
3-OHPHN	5.73	44.3
	22.9	38.9
	115	24.0
	229	25.6
4-OHPHN	5.73	50.2
	22.9	17.7
	115	17.7
	229	18.1
9-OHPHN	22.9	84.1
	115	66.9
	229	66.2

The standard uncertainty RSD_u is expected to be larger than RSD_{IP}. This is because RSD_u incorporates additional uncertainties from the bias CFs. However, comparisons between the values of RSD_{IP} (Table 16) and RSD_u show that the differences between the MUs and IPs are small. This is important information since, in general, to justify bias corrections, the uncertainties in the CFs cannot bloat the uncertainty budget [131]. This means that the bias corrections in this method are warranted, as their contributions to the uncertainty budget are minimal.

To investigate if the method accuracy is satisfactory, APs of the monohydroxylated metabolites are consulted. The APs show the expected errors of future measurements (Table 19 and Figure 32). Most monohydroxylated metabolites have predicted errors around 90-100% at 5.73 ng/g ww, which quickly decrease and stabilize at around 50% at 22.9 ng/g ww and above. However, 9-OHPHN consistently shows a high error, starting at 180% for 22.9 ng/g ww and decreasing to around 150% at 115 ng/g ww and above. This error is, again, noticeably larger than the errors observed for the other compounds.

Table 19: 95% APs [%] for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	$AP_{95\%}$ [%]
1-OHPHN	5.73	[-92.1, 96.6]
	22.9	[-75.4 , 64.9]
	115	[-42.7 , 47.6]
	229	[-50.6 , 49.9]
2-OHPHN	5.73	[-93.6 , 87.4]
	22.9	[-54.3 , 73.3]
	115	[-61.3 , 57.8]
	229	[-78.1 , 68.9]
3-OHPHN	5.73	[-103 , 104]
	22.9	[-95.5 , 92.4]
	115	[-55.2 , 56.4]
	229	[-60.6 , 60.5]
4-OHPHN	5.73	[-120 , 113.9]
	22.9	[-39.5 , 42.1]
	115	[-38.1 , 45.8]
	229	[-47.2 , 38.5]
9-OHPHN	22.9	[-204 , 202]
	115	[-151 , 162]
	229	[-159 , 152]
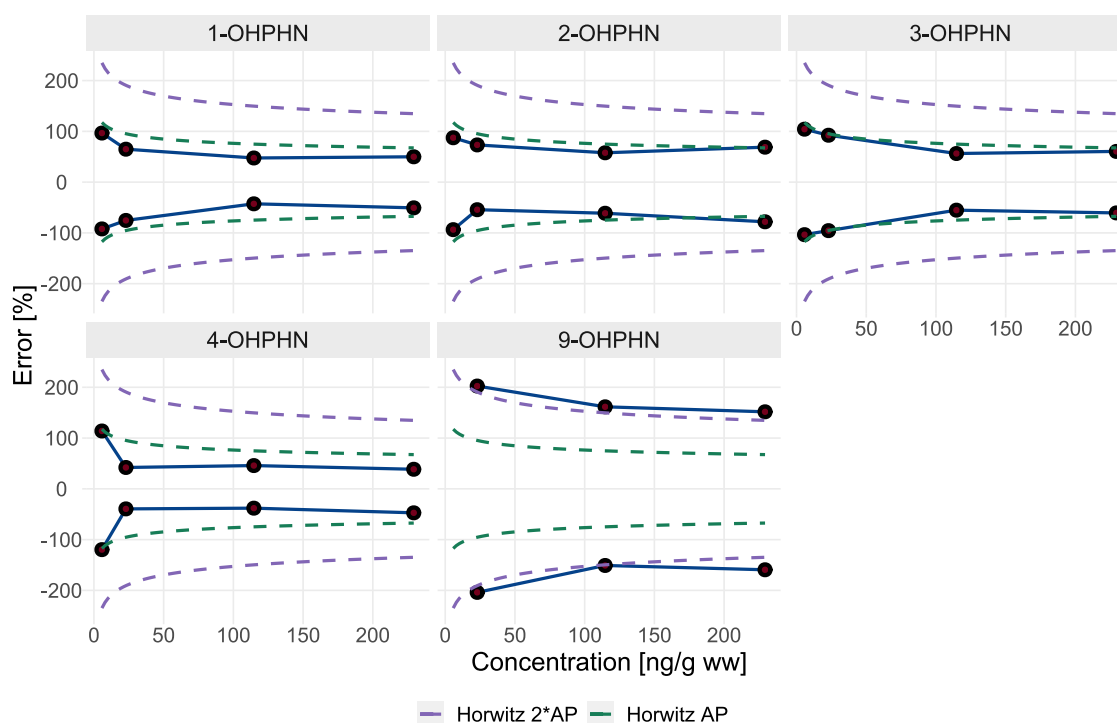
Figure 32: The 95% APs for the monohydroxylated metabolites. The dashed green line (Horwitz AP) corresponds to the theoretical 95% AP generated from Horwitz equation, whereas the dashed purple line (Horwitz 2*AP) is twice the theoretical AP. The purple lines denote the acceptance limits. For all compounds, besides 9-OHPHN, the AP fall below the acceptance limits. For 9-OHPHN, the AP slightly exceeds the limits.

In Figure 32, the APs of the analytes are compared to a theoretical AP generated by the Horwitz equation. Similarly to the precision evaluation, a two-fold difference between the Horwitz and analyte estimates is considered a reasonable acceptance limit. For most compounds, this is achieved; the measured errors are smaller than the acceptance limits (purple line). The errors of 9-OHPHN, however, are very close to these limits and even exceeds them at some points.[11] Despite this, the analyte is considered to be sufficiently accurate at present, as it does not greatly exceed the two-fold theoretical limit. Nonetheless, as previously mentioned, optimization is likely beneficial and should be considered.

4.2.6   Limit of Quantification and Detection

The monohydroxylated metabolites provided satisfactory quantification at all validated levels. Their LoQs can therefore be set as the lowest validated concentration level. For the dihydrodiols, the LoQs are set as the lowest level at which they demonstrated an instrumental response.

The LoDs were estimated by spiking blank eggs close to the suspected LoDs and constructing response functions with the extracts. Spiking was performed in the range of 0.92-2.1 ng/g ww. The estimated LoDs were found in the range of 1.15-2.38 ng/g ww for 1-, 2-, 3- and 4-OHPHN. For the remaining compounds, for the sake of simplicity, the LoDs are taken to be same as the LoQs (Table 20).

Table 20: Limits of Quantification and Detection. For the dihydrodiols and 9-OHPHN, both the LoQs and LoDs are set as the lowest level at which they demonstrated an instrumental response.

Compound	LoD [ng/g ww]	LoQ [ng/g ww]
1-OHPHN	2.75	5.73
2-OHPHN	1.83	5.73
3-OHPHN	1.15	5.73
4-OHPHN	1.38	5.73
9-OHPHN	22.9	
1,2-OH-1,2-HPHN	115	
9,10-OH-9,10-HPHN	115	

The responses at the LoQ levels can be illustrated (Figure 33). For 9-OHPHN, the LoQ signal is close to the baseline, making it very susceptible to noise. This, as discussed earlier, is the reason for the poor accuracy of the metabolite. A similar signal-to-noise ratio is found for 1,2-OH-1,2-HPHN. For the remaining monohydroxylated compounds and 9,10-OH-9,10-HPHN, the peaks are quite tall and readily discernible from the baseline noise.

---

[11]The *HorRat*s for the APs in Table 19 are presented in Appendix B.
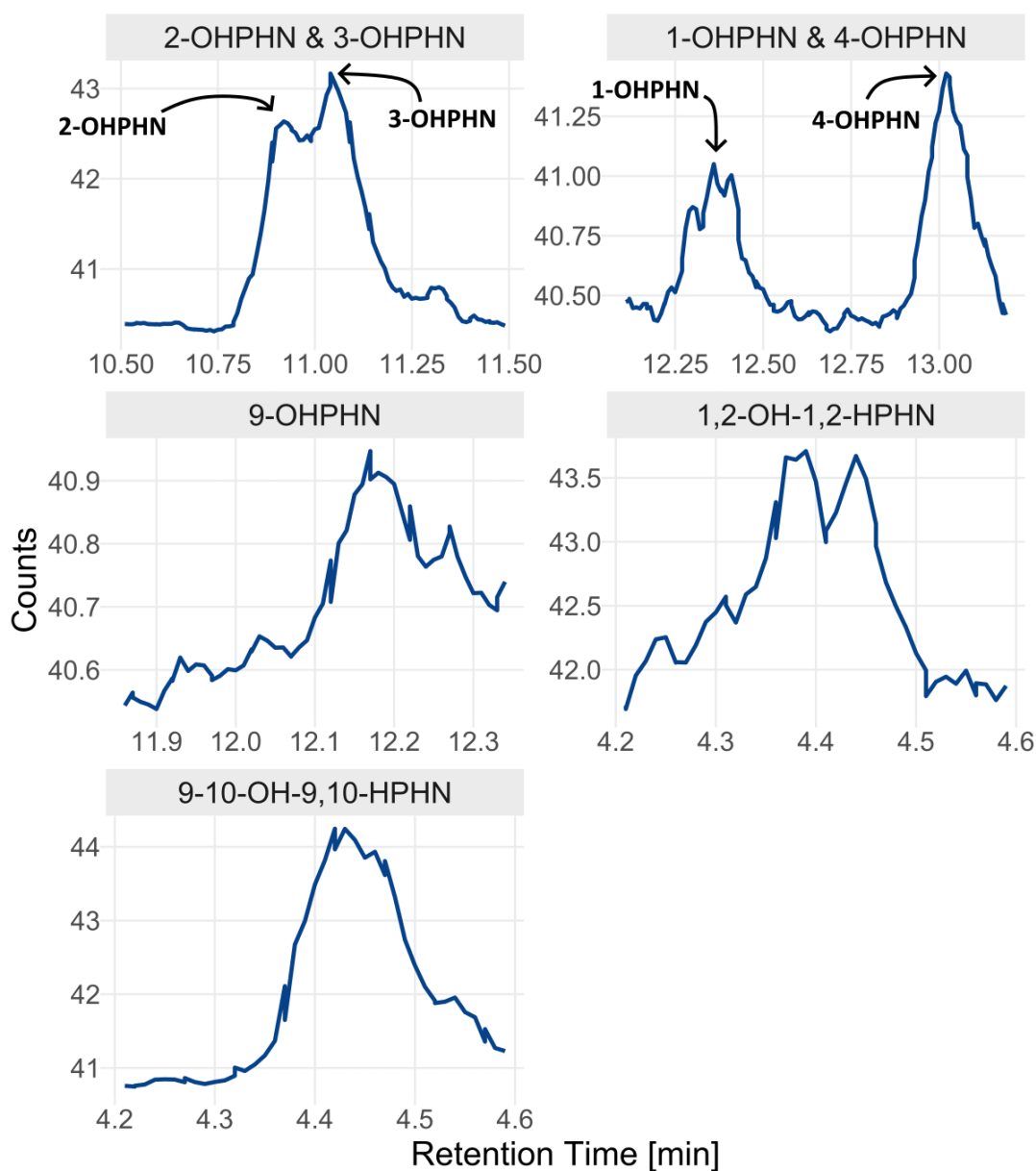
Figure 33: Chromatograms recorded at the LoQ Level. For 1-, 2-, 3- and 4-OHPHN and 1,2-OH-1,2-HPHN the signals at their LoQs are considerably above the baseline noise. For 9-OHPHN and 1,2-OH-1,2-HPHN, the signals are quite poor.

It is also possible to illustrate responses around the LoD. For 9-OHPHN and the dihydrodiols, the LoDs were set to be the same as the LoQs and are thus illustrated in Figure 33. Moreover, while there were no analyses recorded for the remaining monohydroxylated compounds at their exact LoD levels, the concentrations of the four upper standard solutions from the response functions are close. These levels thus serve as good estimates for the LoD signals. As seen in Figure 34, these chromatograms show responses that readily discernible while still being close to the baseline. Therefore, the estimated LoDs are seemingly reasonable.
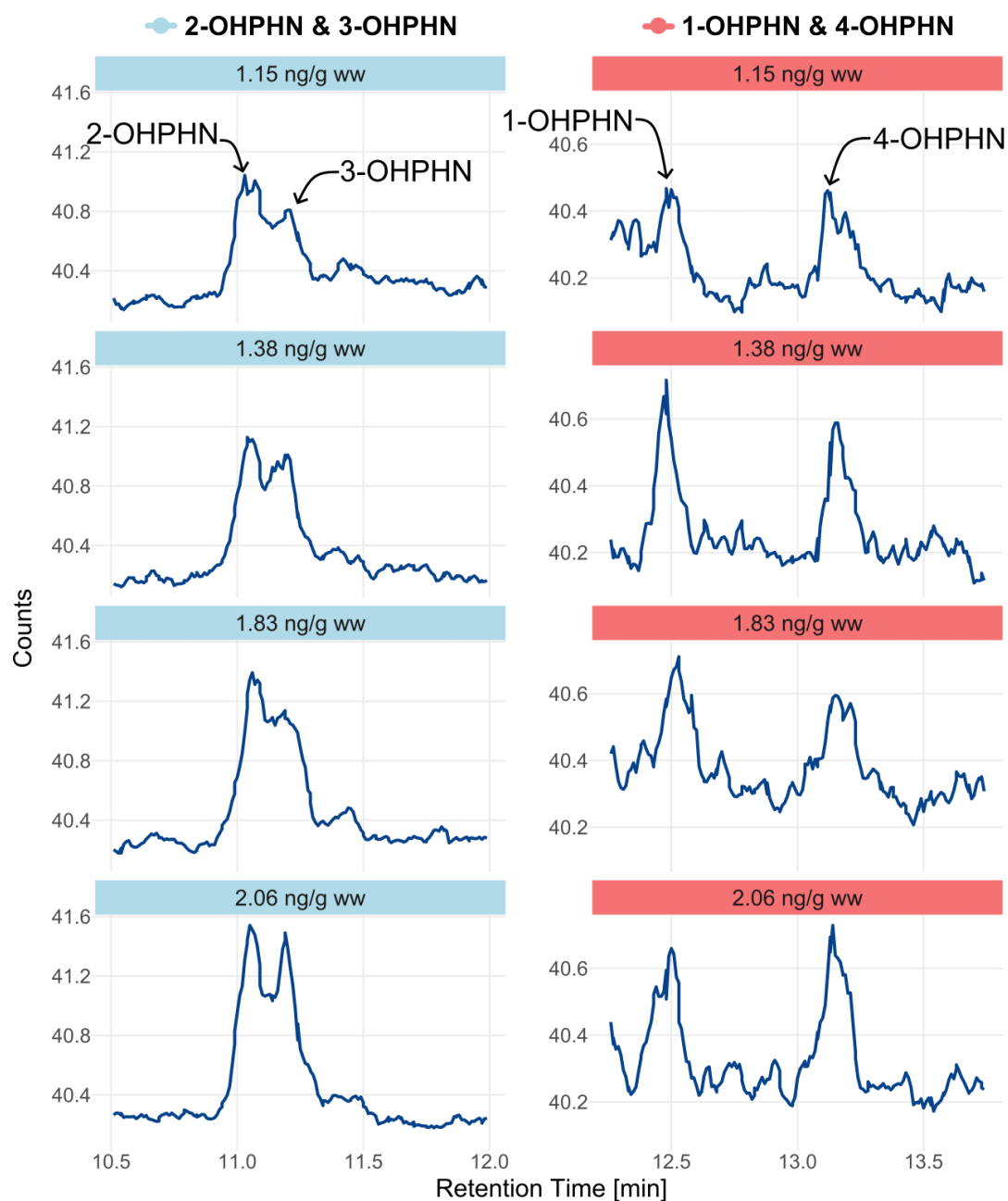
Figure 34: Chromatograms recorded at concentrations close to the LoD Level. For all presented compounds, responses are discernible but close to the baseline.

4.2.7   Selectivity

The selectivity of all compounds were deemed satisfactory. For the monohydroxylated metabolites, there is considerable chromatographic overlap between the two pairs 2-OHPHN & 3-OHPHN and 1-OHPHN & 9-OHPHN (Table 21). Despite this, the compounds are resolved well enough to provide satisfactory accuracy (Section 4.2.5), and may thus be deemed adequately separated. The remaining compounds all have distinct MRMs, providing unique chromatograms and complete selectivity (Figure 35).

Table 21: Chromatographic resolutions for the monohydroxylated metabolites at their validated levels.

Compound	Concentration [ng/g ww]	R_s
1-OHPHN	5.73	3.70
	22.9	1.52
	115	1.13
	229	1.18
2-OHPHN	5.73	0.52
	22.9	0.57
	115	0.52
	229	0.57
3-OHPHN	5.73	0.52
	22.9	0.57
	115	0.52
	115	0.57
4-OHPHN	5.73	3.70
	22.9	2.99
	115	2.99
	229	3.20
9-OHPHN	22.9	1.52
	115	1.13
	229	1.18
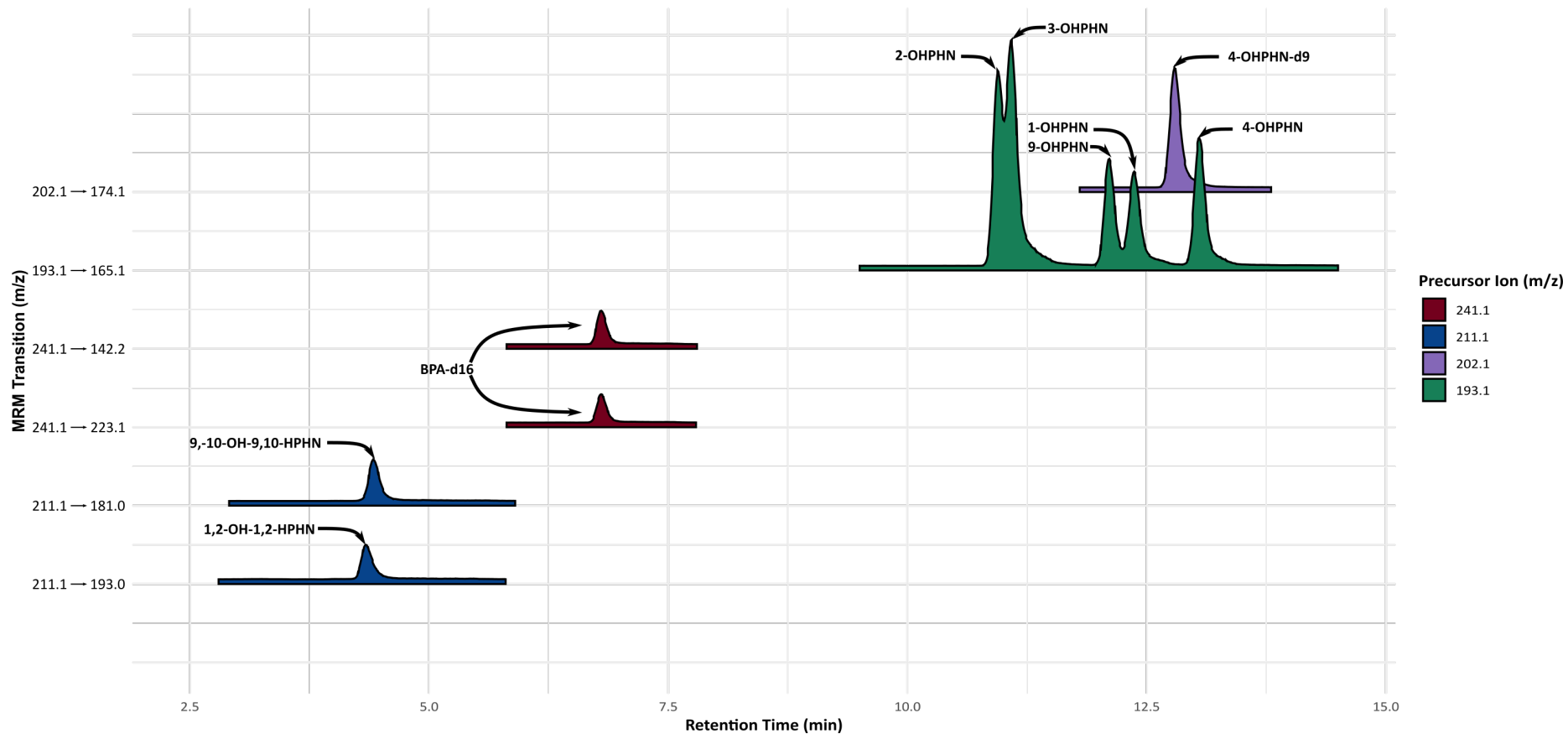
Figure 35: Chromatograms of different analytes and their associated MS transitions at the 30.0 ng/mL calibration standard. While there is considerable chromatographic overlap between the monohydroxylated metabolites, all peaks can be readily discerned. Moreover, every other compound in the method has unique MS transitions and thus have sufficient selectivity.

4.3   Application of the Analytical Method

After the method validation, the method was applied to 47 unique samples from two separate sets of PAH exposure studies. These samples were extracted over the course of four separate days, each day also including a quality control (QC) sample. The QC consisted of 20 blank eggs spiked to a 115 ng/g ww concentration.

In the first study, oil microdroplets were dispersed into sea water and the resulting water soluble fraction (WSF) collected. The WSF was fractionated into 12 subfractions; 8 fractions via ion exchange SPE and 4 fractions via reverse-phase SPE. The haddock eggs were exposed to the WSF, the least retained reverse-phase SPE fraction (SI0) and a mixture comprised of all 12 individual fractions (WSF Comb).

The exposures in WSF and WSF Comb seemingly lead to a higher formation of metabolites when compared to SI0 (Figure 36). These data correspond well with expected results. In the crude oil mixture, there is a range of compounds that induce the oxidising CYP1A/CYP1B enzymes via the AhR xenosensor. This induction ultimately amplifies the production of PAH metabolites. As PAH toxicity is mediated by its metabolites, these findings agree with previous work that proposed the individual PAH fractions are not able to provide the same toxicity as the crude oil mixture [30, 55].
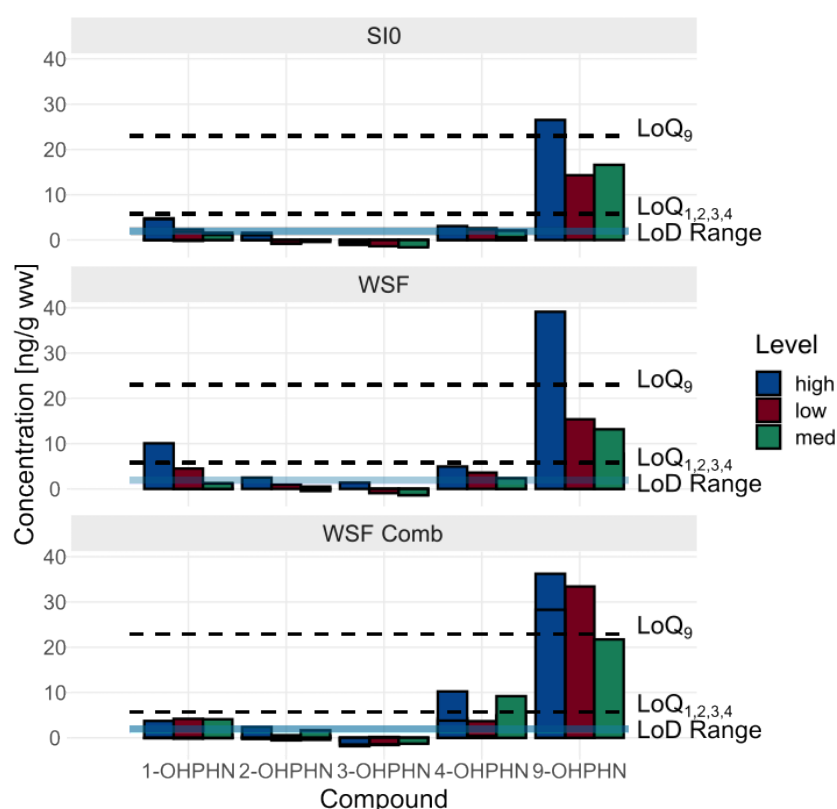
Figure 36: Estimated metabolite concentrations from the WSF Study. *LoD Range* denotes the range of LoDs estimated for the monohydroxylated metabolites, except 9-OHPHN. LoQ_9 denotes the LoQ and LoD for 9-OHPHN. $LoQ_{1,2,3,4}$ denotes the LoQ for all remaining monohydroxylated metabolites. In general, metabolite formation is higher for 9-OHPHN. Additionally, the WSF and WSF Comb exposures are associated with a higher formation of metabolites when compared to SI0.

A second set of exposure experiments were conducted in 2018 (PD18) and 2019 (PD19). In these studies, haddock eggs were exposed to individual PAHs via passive dosing. Exposures were done at five levels, S0-S4, with S0 being the highest and S4 the lowest.

In contrast to the previous results, the data from this study do not show any distinct trends (Figure 37). It is possible to roughly make out that 9-OHPHN is the most abundant metabolite, followed by 4-OHPHN, then 1-OHPHN and 2-OHPHN, and finally 3-OHPHN. However, beyond the relative formation of metabolites, it is difficult to spot any trends. Since the S0 exposure was greatest, it would also be expected to have the highest metabolite levels. The subsequent dilutions (S1, S2, S3, and S4) should thus have progressively lower levels, but this is not reflected by the data.
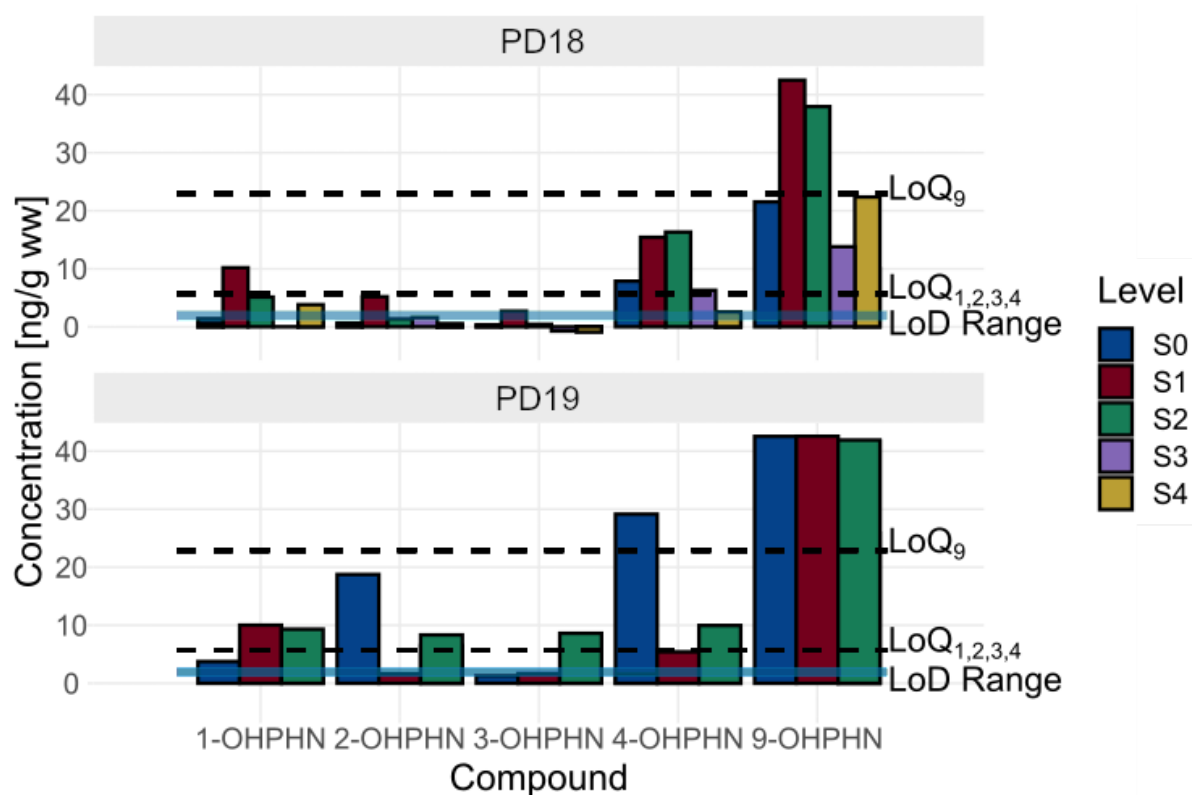
Figure 37: Estimated metabolite concentration from the passive dosing study. *LoD Range* denotes the range of LoDs estimated for the monohydroxylated metabolites, except 9-OHPHN. LoQ_9 denotes the LoQ and LoD for 9-OHPHN. LoQ_{1,2,3,4} denotes the LoQ for all remaining monohydroxylated metabolites. Again, 9-OHPHN is the most abundant metabolite, but additional trends are not prominent.

In addition to the exposure samples, four QC samples were also prepared. Expanded uncertainties, with 95% confidence levels, were constructed for every monohydroxylated metabolite in each sample (Table 22 and Figure 38). Out of the twenty uncertainty intervals, only one did not contain the spike concentration. This is exactly what is expected from a CI with a 95% confidence level.[12] Therefore, the method seemingly works as intended within its validated range.

---

[12]This is because a CI with a 95% confidence level has an error rate of 5% = 0.05 = 1/20.

Table 22: Concentration estimates [ng/g ww] alongside their 95% expanded uncertainties, $U_{95\%}$ [ng/g ww], for 115 ng/g ww spikes at different extraction days for the monohydroxylated metabolites.

Compound	Estimate [ng/g ww]	$U_{95\%}$ [ng/g ww]	Extraction Day
1-OHPHN	87.5	[47, 128]	1
	113	[61, 165]	2
	94.7	[50.9, 138.5]	3
	107	[58, 156]	4
2-OHPHN	84.0	[33.8, 134.2]	1
	100	[40, 160]	2
	84.1	[33.9, 134.3]	3
	85.4	[34.4, 136.4]	4
3-OHPHN	155	[70, 240]	1
	113	[51, 175]	2
	103	[46, 160]	3
	96.6	[43, 150.2]	4
4-OHPHN	117	[65, 169]	1
	91.8	[51.3, 132.3]	2
	84.2	[47, 121.4]	3
	64.1	[35.7, 92.5]	4
9-OHPHN	71.0	[-38 180]	1
	64.0	[-34.2, 162.2]	2
	127	[-70, 324]	3
	154	[-85, 393]	4
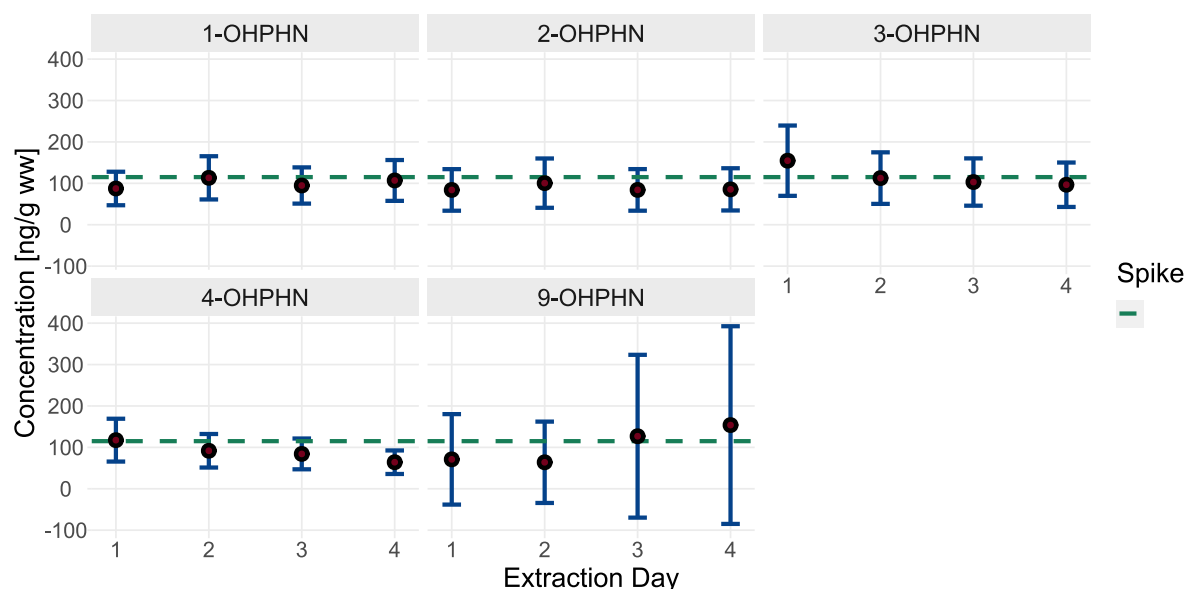
Figure 38: Concentration estimates for QC samples alongside their 95% expanded uncertainties. The red dot corresponds to the estimated concentration and the blue line its expanded uncertainty. The green line illustrates the added spike concentration (115 ng/g ww). The figure shows that 19 out of 20 expanded uncertainties contain the spike value.

Despite the exposure studies providing results that are correspondent with recent research, they must still be interpreted with caution due to their very low concentrations. The experimental

data shown above was initially assumed to be, at least mostly, above the LoQ. However, this is not the case. For 1-, 2-, 3- and 4-OHPHN, all samples generally fall below their LoQ at 5.73 ng/g ww. Similarly, despite its high abundance, 9-OHPHN also falls below its own LoQ at 22.9 ng/g ww for approximately half of the samples.

Therefore, while the method is seemingly accurate, it is unable to quantify samples at relevant concentrations. A lower concentration level must thus be validated. The above results showed that sample concentrations are close to their LoDs (Figure 36 and Figure 37). However, at the LoD levels, the chromatographic shapes and signals are rather poor and will likely have an unsatisfactory accuracy (Figure 33 and Figure 34). Therefore, in its current state, the method is unable to be validated at any lower concentration level. The method should thus be further optimized in order to allow for additional validation (and quantification). Changes could also be made to future experimental designs. Most notably, a larger number of eggs could be collected per sample. If the sample preparation is sufficient, this should also enhance the signal.

4.4   Optimization of the ESI

4.4.1   ESI Parameters

The sample data above shows that many experiments possess concentrations below the LoQ. Therefore, the method should be further optimized in an attempt to provide quantifiable responses for the samples. The ESI was selected as a starting point for the optimization.

The ion source was an Agilent Jet Stream ESI (Figure 39). The *gas temperature* (GT) and *flow* (GF) refers to the temperature and flow of nitrogen drying gas. Thus, these parameters are responsible for desolvation; the reduction of ion droplets into individual ions. Ion droplets are formed at the spray nozzle and are dependent on both the *nozzle voltage* (NV) and *nebulizer pressure* (NP). The former parameter concerns the electric potential at the end of the nozzle capillary. This voltage is responsible for applying charges to clusters of sample molecules. The latter parameter, describes the pressure of the nitrogen gas flow alongside the nozzle capillary. This pressure is intimately related to the aerosol droplet formation. Orthogonal to the nozzle capillary is a second capillary. This heated needle also possesses an electric potential, the purpose of which is to accelerate the generated ions into the mass analyser. The *capillary voltage* (CV) thus determines which ions may enter the quadrupoles [76, 132].

The GT, GF, NP, NV and CV were manipulated in an attempt to enhance the signal. For simplicity, the peak area of 4-OHPHN was chosen to be representative of the ESI response. This metabolite was chosen because of its good instrumental response and baseline resolution. Both of these properties make it easier to monitor changes in the peak area as the ESI is manipulated.

The ESI had constraints that determined the experimental settings. The GT could be adjusted between 0 and 350 °C, with increments of 5; the GF could be adjusted between 0 and 13 L/min, with increments of 0.1 L/min; the NP could be adjusted between 0 and 60 psi, with increments of 0.5 psi; the CV could be adjusted between 0 and 6000 V, with increments of 50 V; the NV could be adjusted between 0 and 2000 V, with increments of 50 V.
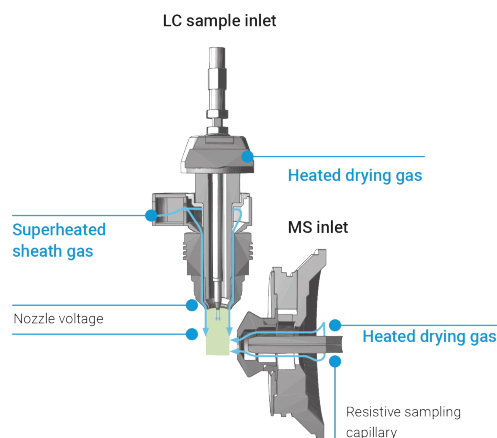
Figure 39: The Agilent Jet Stream ESI (modified from source) [132].

4.4.2   Screening for Significant Factors

To screen for ESI settings with considerable effects on the signal, a two-level 2_V^{5-1} FFD was employed. The defining relation was *0 = 12345*. The GT was evaluated at 250 (low level) and 350 °C (high level); GF at 9 and 13 L/min; NP at 15 and 25 psi; CV at 3000 and 4000 V; NV at 1000 and 2000 V. The initial (unoptimized) method and four replicates of the center point were also run. A complete overview of the design is provided in Table C.1 in Appendix C.

Seemingly, three of the experiments (4, 7 and 14) provide enhanced signals when compared to the initial method (Figure 40). This suggests that there are indeed some parameters which can be changed to enhance method performance.
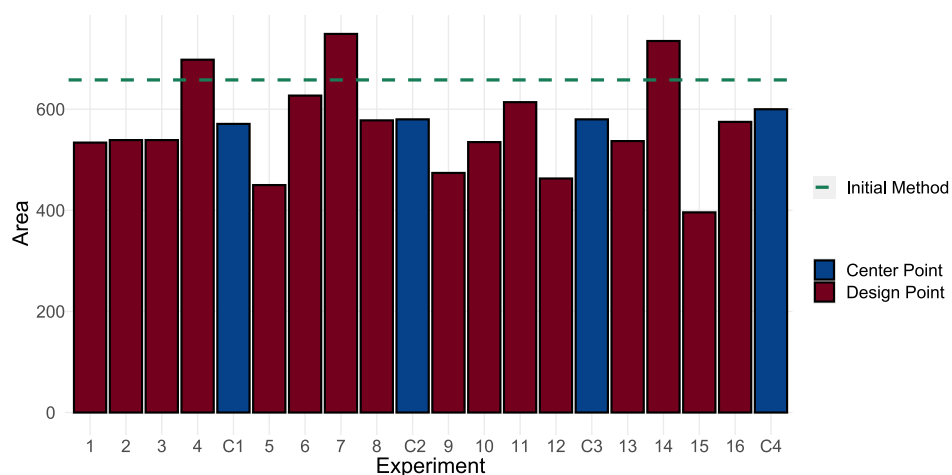
Figure 40: Areas for the initial FFD. The colors denote the different experiments, blue is the center points and red the experiments from the FFD. The dashed green line refers to the recorded area of the initial method. Experiments 4, 7 and 14 provide responses which are better than the initial method.

OLS regression techniques can be used to generate the regression coefficients for the design. Additionally, the center point replicates enable the estimation of instrumental noise, allowing the coefficients to be tested for statistical significance. Inspection of the coefficients reveals that the NV and CV are much more significant than the rest (Table 23). Additionally, the model seems to be well fit as both r^2 and r^2_{adj} are large. Based on these observations, manipulation of the CV and NV likely provides a good starting point for further optimization.

The effects of the voltages can also be revealed via a *half-normal probability plot*. If the investigated factors have no affect on the response, their regression coefficients are simply a result of random noise. Assuming the noise is normally distributed, the regression coefficients of negligible factors should thus form a straight line when plotted against the quantiles of a normal distribution.[13] Any regression coefficient deviating from the line is then likely statistically significant. Indeed, when plotting the regression coefficients from Table 23 (excluding the intercept) against the theoretical quantiles of the normal distribution, the significance of the CV and NV is clearly seen (Figure 41).

Table 23: FOM with two-way interactions for the initial 2^{5-1}_V FFD.

Coefficient	Estimate	Std.Error	p
Intercept	568.70	4.24	1.85e-08***
GT	-21.69	4.74	1.02e-02*
GF	-14.56	4.74	3.72e-02*
NP	1.56	4.74	7.58e-01
CV	-68.56	4.74	1.33e-04***
NV	50.19	4.74	4.50e-04***
GT:GF	-8.19	4.74	1.59e-01
GT:NP	10.19	4.74	9.80e-02
GT:CV	6.06	4.74	2.70e-01
GT:NV	3.81	4.74	4.66e-01
GF:NP	19.06	4.74	1.58e-02*
GF:CV	11.19	4.74	7.76e-02
GF:NV	18.44	4.74	1.77e-02*
NP:CV	-16.69	4.74	2.44e-02*
NP:NV	-1.94	4.74	7.04e-01
CV:NV	-15.06	4.74	3.36e-02*
r^2		0.991	
r^2_{adj}		0.955	

---

[13] In essence, this is the same as the commonly encountered quantile-quantile plots.
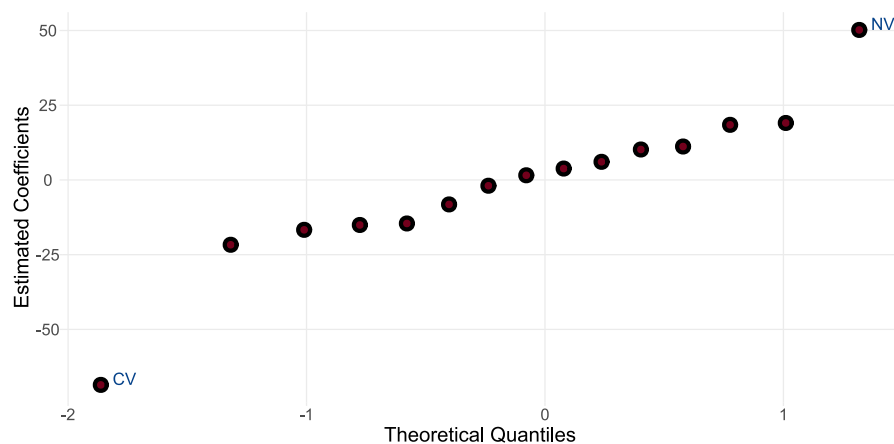
Figure 41: Half-normal probability plot of initial screening regression coefficients. The plot reveals that the CV and NV deviate from the "noise line" and are significant parameters.

4.4.3   Optimization by the Path of Steepest Ascent

The regression coefficients of CV and NV are negative and positive, respectively. This suggests that the area will be enhanced when the CV is *decreased* and the NV *increased*. However, as the NV is already at its maximum capacity (2000 V), only the CV can be altered. This technique, where the change in a parameter is given by the sign and magnitude of its regression coefficient, is referred to as the method of *steepest ascent* [96–98].

In an attempt to enhance the signal, all parameters except the CV were held at their center point values. The starting condition of the CV was 3000 V. From here, using a step size $\Delta CV = -100$ V, the CV was further reduced across ten experiments. When moving from 2900 to 2300 V, there is an improvement in the response. However, when further decreasing the voltage down to 2000 V, the response deteriorates (Figure 42). The area thus peaks at CV = 2300 V. This voltage provides a reasonable starting point for a new design.
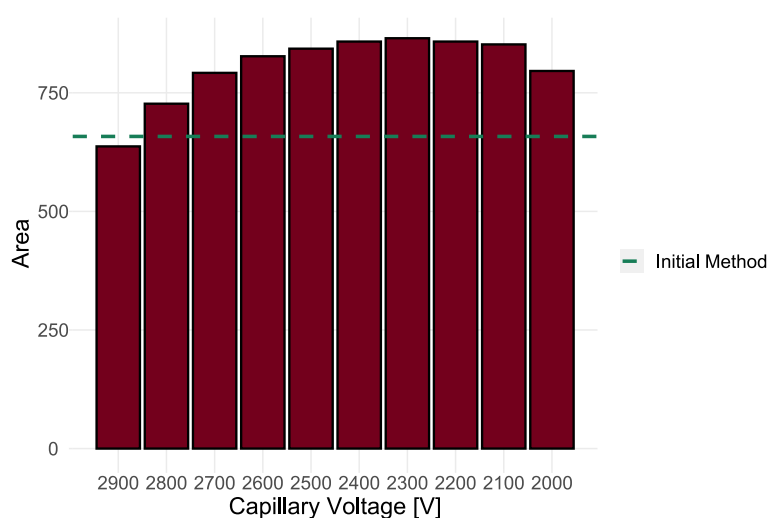
Figure 42: Area as a function of the capillary voltage (CV). The axis decrease from left to right, each bar reduces the voltage by 100 V. The green line represents the area of the initial method. The area peaks at 2300 V.

4.4.4 Follow-Up Screening

A new 2_V^{5-1} FFD was constructed using the CV from the steepest ascent, providing the new levels 1800 and 2800 V. The FFD levels for the other parameters remained unchanged from the previous design (Table C.2). For this design, the defining relation *0 = -12345* was used. Two center point replicates and the initial method were also run.

Noteworthy, all experimental outcomes are now better than the initial method (Figure 43).
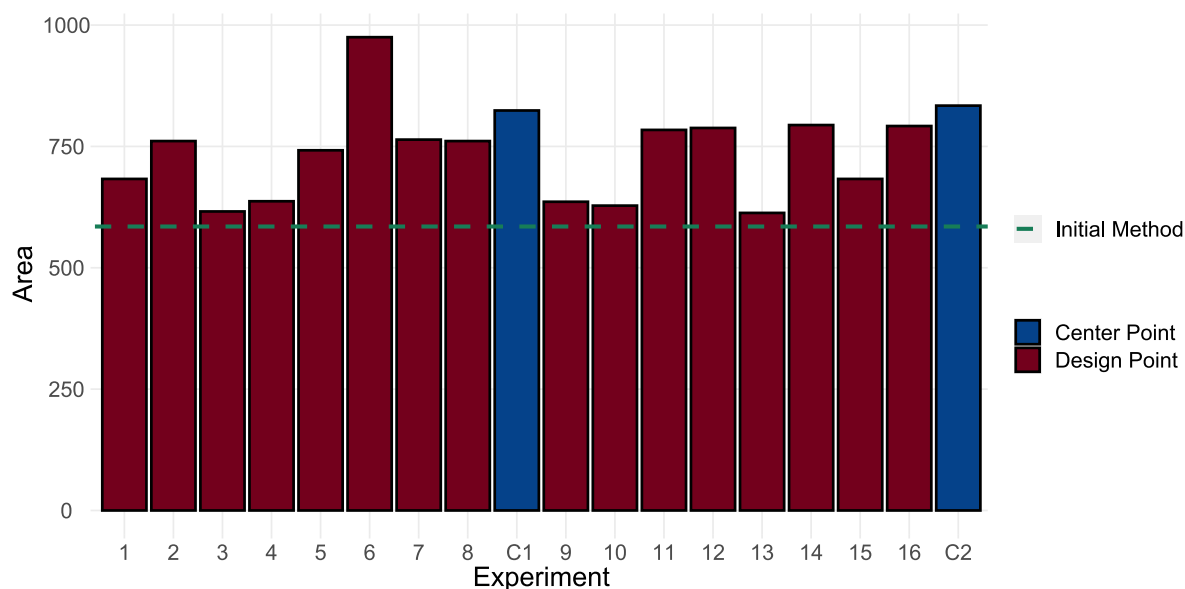
Figure 43: Areas for the follow-up FFD. The colors denote the different experiments, blue is the center points and red the experiments from the FFD. The dashed green line refers to the recorded area of the initial method. In this design, all experiments have a response which is higher than the initial method.

Upon examining the regression model, it can be seen that, other than the intercept, no other regression coefficient is statistically significant (Table 24). Nonetheless, with only two replicates at the center point, the standard errors of the regression estimates only have one single degree of freedom, making false negatives very likely. Thus, to further investigate the significance of the regression coefficients, a half-normal probability plot is constructed (Figure 44). The plot reveals that the GF, GT, NP and NV can be altered to enhance the method response. However, while the model has a seemingly high r^2, its r^2_{adj} is close to zero (Table 24). Due to the poor model fit, the path of steepest ascent will likely not provide any considerable increase in the method response. The poor fit may be a consequence of curvature in the response. Therefore, at this point, a reasonable next step in the optimization process is to estimate the quadratic coefficients and generate a response surface.

Table 24: FOM with two-way interactions for the 2_V^{5-1} follow-up FFD.

Coefficient	Estimate	Std.Error	p
Intercept	739.72	22.35	9.12e-04
GT	-36.44	23.71	0.264
GF	-5.94	23.71	0.836
NP	-9.19	23.71	0.746
CV	4.06	23.71	0.880
NV	-58.31	23.71	0.133
GT:GF	20.06	23.71	0.496
GT:NP	3.31	23.71	0.902
GT:CV	-6.44	23.71	0.811
GT:NV	7.19	23.71	0.790
GF:NP	38.31	23.71	0.257
GF:CV	0.060	23.71	0.998
GF:NV	10.69	23.71	0.696
NP:CV	16.31	23.71	0.562
NP:NV	24.94	23.71	0.403
CV:NV	26.19	23.71	0.384
r^2		0.884	
r^2_{adj}		0.010	
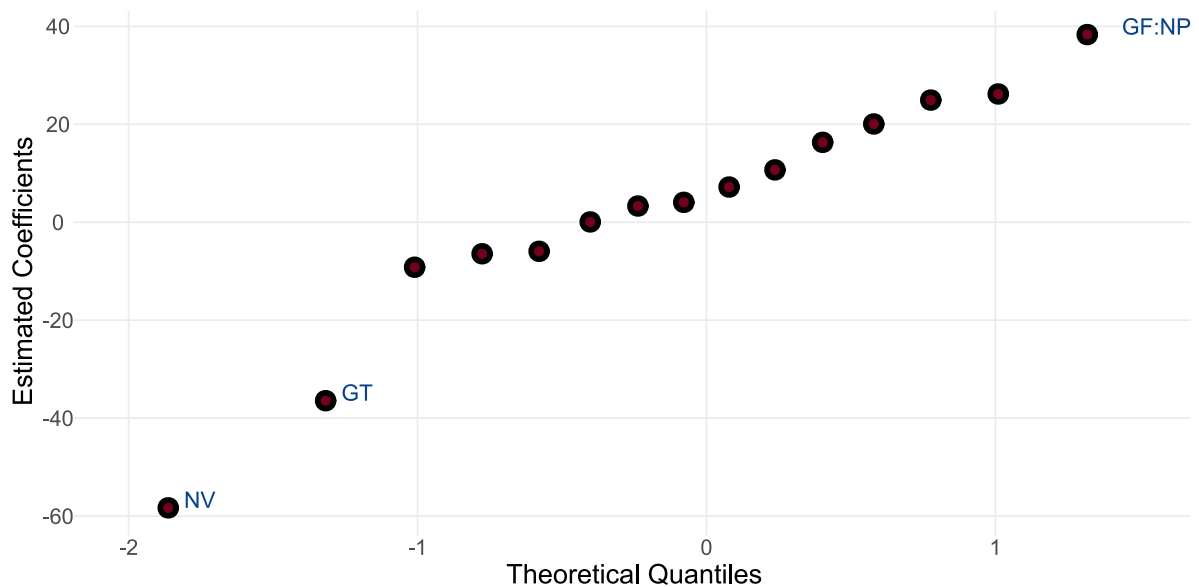
Figure 44: Half-normal probability plot for follow-up screening regression coefficients. The plot suggests that NV, GT and GF:NP are significant.

4.4.5   Response Surface Modelling

4.4.5.1   Generating the Response Surface

To generate the response surface, a third 2_V^{5-1} fractional factorial design was constructed on the basis of experiment 6 from the previous FFD.[14] The defining relation was $0 = 12345$. The GT was evaluated at 225 and 275 °C; GF at 7 and 11 L/min; NP at 18 and 22 psi; CV at 1550 and 2050 V; NV at 750 and 1250 V. Four center point replicates and the initial method were also run (Table C.3). Moreover, axial points were added to this design to create a CCD. Axial points are found by evaluating a factor at two extreme levels, one high ($+\alpha$) and one low ($-\alpha$), while the other factors are held at their center points. From Equation (29), in coded units, the value of α becomes:

$$\alpha = \left[\frac{2^{k-f} \cdot n_c}{n_\alpha} \right]^{\frac{1}{4}} = \left[\frac{2^{5-1} \cdot 1}{1} \right]^{\frac{1}{4}} = 2$$

By converting α from coded units into numerical values, the extreme levels for GT are found at 200 and 300 °C; GF at 5 and 13 L/min; NP at 16 and 24 psi; CV at 1300 and 2300 V; NV at 500 and 1500 V.

In the CCD, the responses are very similar across most of the experiments (Figure 45). This might suggest that the experimental factors were varied across too narrow ranges and could have benefited from larger intervals. Nonetheless, the design can still provide useful information regarding the state of the method and future means of optimization.
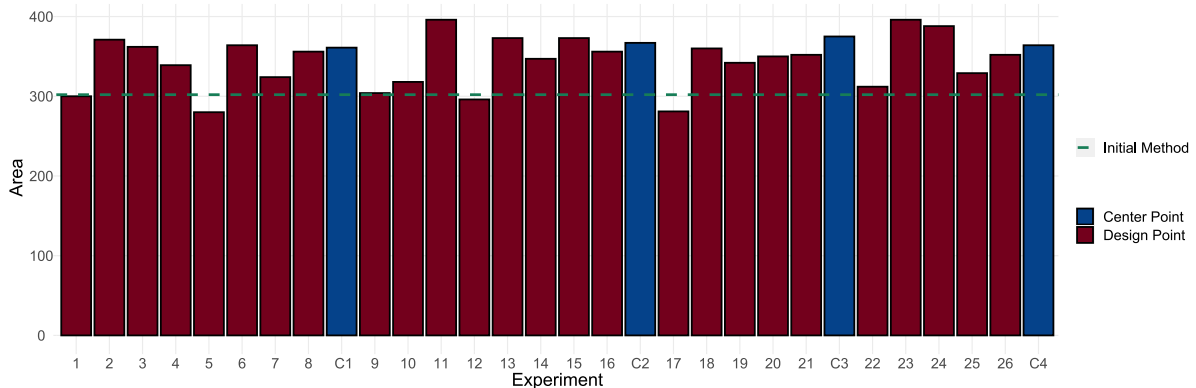
Figure 45: Areas for the CCD. The colors denote the different experiments, blue is the center points and red the experiments from the CCD. The dashed green line refers to the recorded area of the initial method. All experiments produce similar results.

According to the *t*-tests and half-normal probability plot, the statistically significant regression coefficients include GF, GF:CV, NP:NV and CV^2 (Table 25 and Figure 46). Thus, all but one ESI parameters are associated with some significant regression coefficient. Moreover, the r_{adj}^2 is now at 0.759 as opposed to 0.010 found in the last design. These results suggest that the method response is indeed heavily influenced by curvature.

---

[14]The experiment from this design can be found in Table C.2

Table 25: Second order model for a CCD constructed from a 2_V^{5-1} FFD.

Coefficient	Estimate	Std.Error	p
Intercept	366.72	7.56	3.37e-12***
GT	-5.04	3.19	1.49e-01
GF	-21.38	3.19	8.95e-05***
NP	4.13	3.19	2.29e-01
CV	3.88	3.19	2.56e-01
NV	-3.13	3.19	3.54e-01
GT:GF	4.31	3.91	2.99e-01
GT:NP	-1.31	3.91	7.45e-01
GT:CV	-3.06	3.91	4.54e-01
GT:NV	-3.44	3.91	4.02e-01
GF:NP	-2.19	3.91	5.90e-01
GF:CV	12.81	3.91	9.61e-03**
GF:NV	-3.56	3.91	3.86e-01
NP:CV	3.44	3.91	4.02e-01
NP:NV	10.31	3.91	2.71e-02*
CV:NV	-1.94	3.91	6.32e-01
GT^2	1.84	3.03	5.59e-01
GF^2	-1.04	3.03	7.40e-01
NP^2	-5.16	3.03	1.23e-01
CV^2	-17.54	3.03	2.64e-04***
NV^2	-3.66	3.03	2.58e-01
r^2		0.925	
r_{adj}^2		0.759	
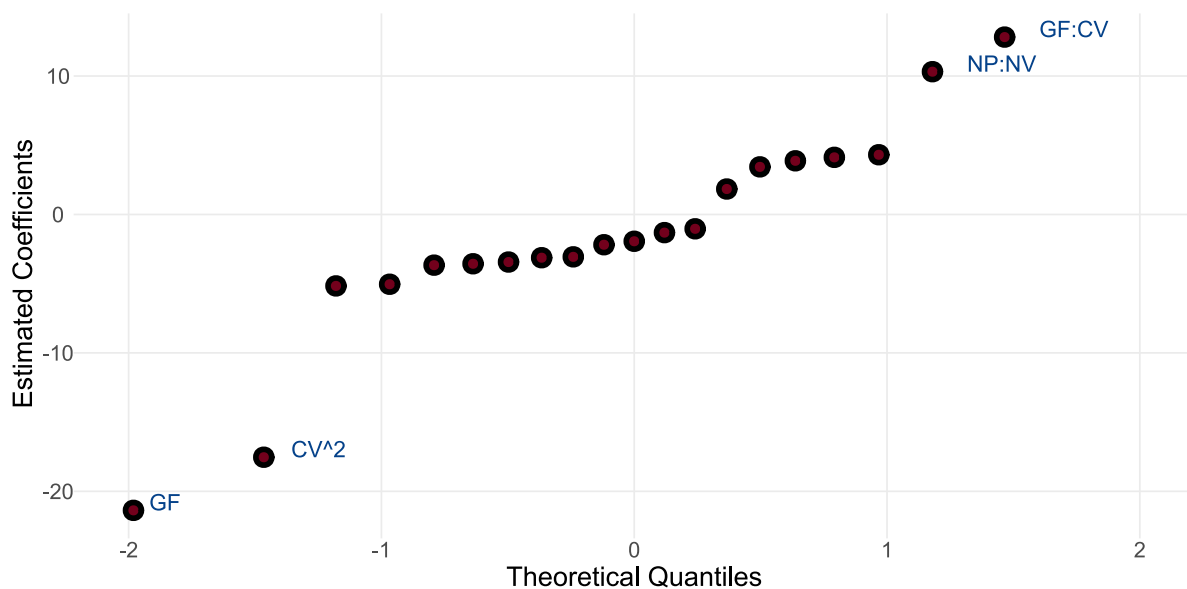
Figure 46: Half-normal probability plot of the response surface regression coefficients. The plot suggests that GF, CV^2, NP:NV and GF:CV are significant.

Using the second order model, it is now possible to visualise the generated response surface. Given that there are five variables, there are $\binom{5}{2} = 10$ unique plots which may be generated (Figure 47).

The contour plots readily illustrate the optimal settings for the NP, CV and NV. These settings are most apparent in plots **H** and **J**, where the CV is compared against the NP and NV, respectively. These plots show a clear peak denoting the maximum response. In both plots, optimal signals are obtained with a CV just above 1800 V, alongside a NP around 19 psi and a NV around 600 V. These settings also yield the highest responses in the remaining plots containing these parameters. Moreover, the plots also reveal that the method response can be increased by reducing GF and GT. This is especially visible in plot **A**, but also reflected in all other plots including the two gas parameters. Thus, an enhanced method can likely be improved by changing these in tandem. However, doing this solely on the basis of the contour plots is not ideal. As the true response surface exceeds three-dimensional space, it is difficult to determine an optimum visually. To ensure future optimization is as precise as can be, a more formal approach must be applied.
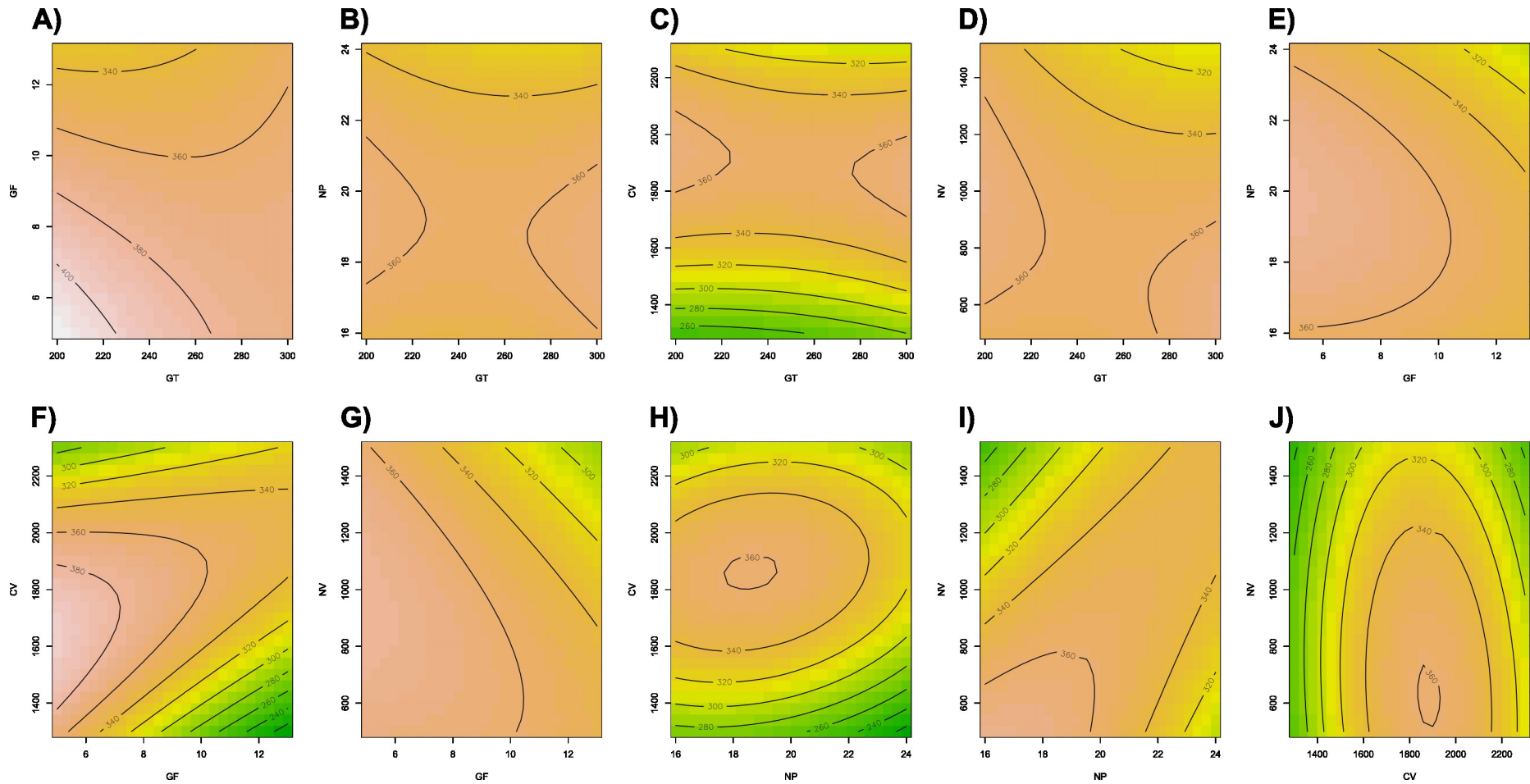
Figure 47: Contour plots generated by the CCD. One contour plot is presented for each combination of the ESI parameters. For any contour plot of two variables, the three other parameters are held at their center point values. The CV is close to an optimum around 1800-1900 V, whereas NV and NP seem to have their highest responses close to 19 psi and 600 V (plots **H** and **J**). The plots including GT and/or GF (*e.g.* plot **A**) show that the parameters are seemingly too large and will likely enhance method response by being lowered.

4.4.5.2   Investigating the Response Surface by Means of Canonical Analysis

Canonical analysis is a formal method to investigate the nature of response surfaces. In essence, the approach transforms the data to a new coordinate system where the origin is given by the stationary point, x_s, and the principal axes given by the response surface. That is, y can be written

$$y = y_s + \sum_{i=1}^{I} \lambda_i w_i^2 \tag{70}$$

Where y_s is the response at the x_s. Moreover, λ_i and w_i denote the ith eigenvalue and eigenvector of the symmetric matrix containing the estimated second-order coefficients; the quadratic terms are found along the diagonal and interaction terms on the off-diagonal. These eigenvectors constitute the principal axes of the transformed system.

Equation (70) describes the nature of the stationary point. More precisely, it describes how the response y changes as one moves away from the stationary point along some axis w_i. Therefore, if one moves along some w_i with an associated negative λ_i, then y will decrease. Equivalently, if λ_i is positive, y will increase when moving along w_i. If *all* λ_is of a system are negative, then the response will decrease when moving away from the stationary point in any direction. This means that the stationary point, x_s, is at an optimum. Alternatively, all positive λ_is means that x_s is a minimum. If the λ_is are of mixed signs, x_s is a saddle point. Moreover, the magnitudes of the eigenvalues describe how much the response changes as one moves along its associated eigenvector. Further mathematical details of this analysis are beyond the scope of this thesis, however the derivation of (70) is provided in Appendix A.1.

On the basis of the paragraphs above, an attempt can now be made to gain a more rigorous understanding of the generated response surface. The eigenvalues and eigenvectors are readily calculated using the RSM package in R. Eigenvectors w_1 through w_5 are provided below, with their corresponding eigenvalues (λ_i) on top:

$$
\begin{array}{ccccc}
4.88 & 0.11 & -0.84 & -9.43 & -20.28 \\[4pt]
w_1 = \begin{bmatrix} 0.64 \\ 0.54 \\ -0.31 \\ 0.10 \\ -0.44 \end{bmatrix} &
w_2 = \begin{bmatrix} 0.73 \\ -0.58 \\ 0.13 \\ -0.27 \\ 0.19 \end{bmatrix} &
w_3 = \begin{bmatrix} -0.22 \\ -0.50 \\ -0.59 \\ -0.20 \\ -0.56 \end{bmatrix} &
w_4 = \begin{bmatrix} -0.02 \\ -0.14 \\ 0.72 \\ 0.13 \\ -0.67 \end{bmatrix} &
w_5 = \begin{bmatrix} 0.10 \\ -0.32 \\ -0.15 \\ 0.93 \\ 0.08 \end{bmatrix}
\end{array}
$$

As the eigenvalues are present with both positive and negative signs, the stationary point of the estimated response surface is a saddle point. This means that the stationary point of the method is not the area of maximum response (Section 2.3.5.2). Nonetheless, the saddle point can still be used to direct the next step in the optimization. The point can be estimated from Equation (70). In the order GT, GF, NP, CV and NV, the proposed stationary point is given by

$$x_s = [-417\ 46\ 3\ 4067\ -1843]^T$$

That is, the canonical analysis estimates that the stationary point of the method is at GT = -417 °C, GF = 46 L/min, NP = 3 psi, CV = 4067 V and NV = -1843 V. This estimate is unrealistic (and inapplicable) as it extends far beyond the possible settings of the ESI. The poor estimate can be attributed to the differences in the magnitude of the eigenvalues. Consider the three eigenvalues $\lambda_2 = 0.11$, $\lambda_3 = -0.84$ and $\lambda_5 = -20.28$. Clearly, λ_2 and λ_3 are negligible when compared to λ_5. As the response shows little sensitivity to the paths defined by w_2 and w_3, it is necessary to travel extreme distances along these paths to reach the predicted stationary point. These extreme distances distort the stationary point estimate. A reasonable approximation is thus to let $\lambda_2 \cong 0$ and $\lambda_3 \cong 0$.

Equating the two negligible eigenvalues to zero provides a new stationary point:

$$x_s = [274\ 10\ 20\ 1905\ 768]^T$$

In units, this point is given by GT = 274 °C, GF = 10 L/min, NP = 20 psi, CV = 1900 V and NV = 768 V. This result agrees well with Figure 47. However, as the stationary point is a saddle point, it is still possible to further optimize the method. Since λ_2 and λ_3 are equated to zero, the remaining eigenvalues are $\lambda_1 = 4.88$, $\lambda_4 = -9.43$ and $\lambda_5 = -20.28$. As λ_1 is positive, moving along w_1 is predicted to improve the response. Optimization by moving along the path of a quadratic function is called optimization by the *canonical path* [96] (in contrast to the path of steepest ascent for a FOM). This path, as given by the eigenvector w_1, is presented below (again in the order GT, GF, NP, CV and NV):

$$w_1 = [0.64\ 0.54\ -0.31\ 0.10\ -0.44]^T$$

As a positive eigenvalue is associated with a minimum, moving along the axis of w_1 in *any* direction will (in theory) increase the response. One way of moving along w_1 is by lowering GT, GF and CV while increasing NP and NV. The RSM package suggests doing this with the following increments: ΔGT = -10 °C, ΔGF = -0.5 L/min, ΔNP = +0.3 psi, ΔCV =-10 V and ΔNV = +25 V. For NP, CV and NV these increments are too small for the ESI to recognize, meaning they are essentially constant when moving along w_1. Thus, the response is expected to improve by reducing GT and GF while keeping the other parameters constant. Notice that this conclusion is congruent with the one made during the visual evaluation of Figure 47. Albeit, by using canonical analysis, more precise details was obtained regarding how optimization should be performed.

4.4.6 Optimization by the Canonical Path

Using the path found in the previous section, step-wise optimization may be attempted once again. For this iteration, the GT and GF were simultaneously adjusted using the step-sizes ΔGT = -10 °C and ΔGF = -0.5 L/min, respectively. The starting conditions for the experiments were taken to be the estimated stationary point. Again, the parameters of the MS can only be adjusted with specific increments. The stationary point was thus approximated by GT = 275 °C, GF = 10 L/min, NP = 20 psi, CV = 1900 V and NV = 750 V.

Starting at GT = 275 °C and GF = 10 L/min, there is an increase in the response until the conditions reach GT = 195 °C and GF = 6.0 L/min. After this, the response deteriorates (Figure 48). Thus, a new optimum for the method is found at GT = 195 °C, GF = 6.0 L/min, = 20 psi, CV = 1900 V and NV = 750 V.
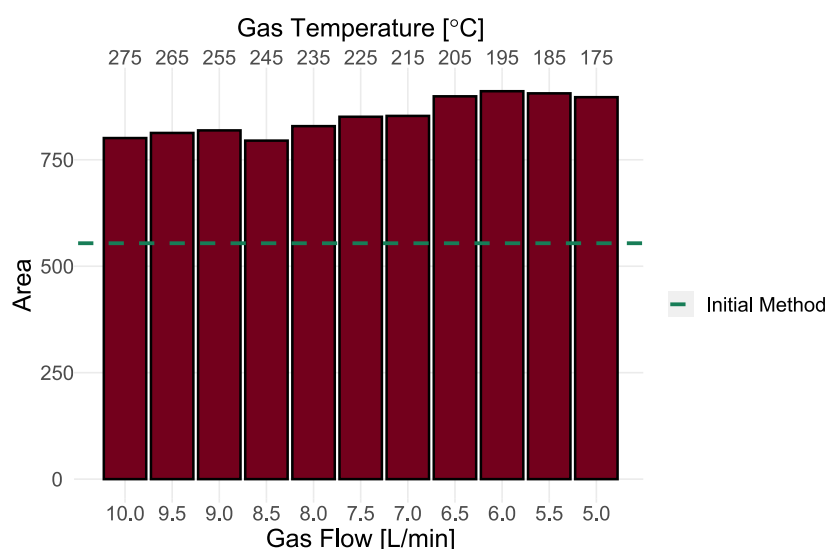
Figure 48: Area as a function of the gas temperature and flow. The axes decrease from left to right, each bar reduces the GT by 10°C (top axis) and GF with 0.5 mL/min (bottom axis). The green line represents the area of the initial method. The area peaks at GT = 195°C and GF = 6.0 L/min.

Due to time constraints the optimization process was terminated at this stage. During the optimization shown in Figure 48, the highest recorded area (found at GT = 195°C and GF = 6.0 L/min) and the area of the initial method were 911 and 554, respectively. This means that the optimization has provided a 64% improvement over the initial method. Indeed, by looking at the analyte chromatograms, it can be seen that the monohydroxylated metabolites have undergone a considerable increase in their response. For the dihydrodiols, however, the response is essentially the same for both the optimized and initial methods (Figure 49).
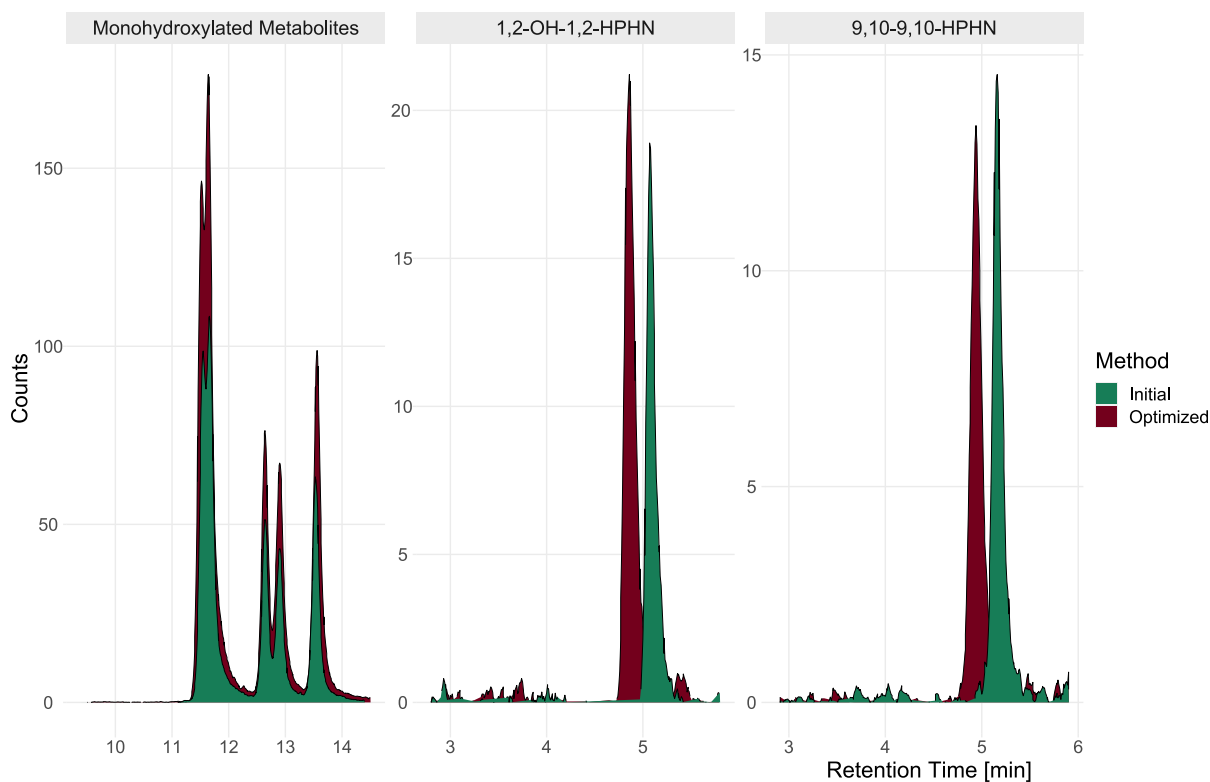
Figure 49: Chromatograms of analytes after optimization. The responses reveal that the optimized method (green) provides a considerably greater signal when compared to the initial method (red) for the monohydroxylated metabolites. For the dihydrodiols, however, the responses have been reduced.

While the chromatographic area has increased, it is the *signal-to-noise ratio* (S/N) which is the most important factor. If the increase in noise is directly proportional to the increase in instrumental response, the LoDs and LoQs will remain constant. Nonetheless, in this case, the signal-to-noise ratio was also found to improve. A comparison between the initial and optimized method reveals that all monohydroxylated metabolites underwent an increase in their signal-to-noise by at least 50% (Table 26). The increase of 2-OHPHN was even found to be at 232%, although this value is likely somewhat influenced by the overlapping signal from 3-OHPHN. These results show that optimization has been successful for the monohydroxylated metabolites. For the dihydrodiols, however, a slight decrease was observed in the signal-to-noise ratio.

Table 26: Signal-to-Noise ratios and percentage differences for analytes before and after optimization.

Compound	Optimized S/N	Initial S/N	Difference [%]
1-OHPHN	6.51	3.72	+75
2-OHPHN	412	124	+232
3-OHPHN	48.9	29.7	+64
4-OHPHN	26.6	17.0	+56
9-OHPHN	20.5	13.7	+50
1,2-OH-1,2-HPHN	11.8	13.7	-14
9,10-OH-9,10-HPHN	10.4	15.0	-31

4.4.7 Future Steps for Optimization

It is likely that the optimum of the ESI is close to the current conditions. The method can reach this optimum by constructing a new response surface at the current settings, but it may not provide significant improvements and could yield diminishing returns. It is therefore worth discussing some other measures which might enhance the method response. Further optimization could be done by changes to MS scan settings or extraction protocol. Additionally, the simplest means to enhance the response could be by simply increasing the injection volume. Nonetheless, at this point, the most considerable increase in instrumental response can likely be found from further changes to the chromatography.

Recall that the (isocratic) chromatographic process can be described by the Purnell equation:

$$R_s = \left(\frac{\sqrt{N}}{4} \right)\left(\frac{\alpha - 1}{\alpha} \right)\left(\frac{k_b}{1 + k_b} \right) \tag{3}$$

Improving the chromatography by increasing the number of theoretical plates, N, is not feasible. The current particle size of the column is already small with its 1.7 µm particle size. However, adjusting the retention, k_b, and selectivity, α, might provide necessary improvements to the method. Both of these parameters influence peak shape. A decrease in peak width can possibly enhance the signal-to-noise ratio of the method. This might provide the method with its necessary LoDs and LoQs.

To change the method retention and selectivity, the analyte chemistry must be understood. In this method, all analytes possess some weakly acidic behaviour due to their hydroxylated nature and ability to undergo resonance. Additionally, for the dihydrodiols, the presence of two hydroxy groups likely provide the dihydrodiols with both an enhanced reactivity and acidity. The enhanced reactivity may reduce instrumental response by forming undetected adducts (possibly akin to the effects of 9-OHPHN discussed in Section 4.2.3). Moreover, the ions produced by deprotonation hinders retention on the non-polar stationary phase, ultimately reducing instrumental response.

There are several alternatives to mitigate the analyte acidity. Firstly, a more protic solvent such as MeOH or the addition of a buffer can likely decrease the deprotonation, enhancing instrumental response. The application of MeOH to acidic analytes is also supported by literature, some even stating this should be the default solvent of choice [133]. Additionally, potential additives can also be used to enhance the response. For instance, adding small volumes of acetic acid can improve the overall signal as its anion produced in the ESI has a strong affinity for protons, thus deprotonating the analytes. However, too much acetic acid will decrease the response, so the concentration of the additive must also be optimized [134]. Additionally, it is not necessary to be restricted to only protic solvents. Another popular solvent is that of tetrahydrofuran (THF). In the case of a LC-MS analysis of fatty acids (also with negative ESI), it was found that a change from MeOH or AcN to THF more than doubled the instrumental response [135]. This can possibly be attributed to the fact that THF has been reported to provide a smaller extent of

adduct formation for selected analytes, when compared to MeOH and AcN [136]. Thus, both THF and MeOH are good candidates for alternate mobile phases.

While a mobile phase may be suitable for chromatographic separation, it need not be compatible with the MS. This includes not only the choice of solvent, but also the time of elution for the analytes. It is possible that the analytes elute at a time where the mobile phase composition suppresses ionization. In this case, there is limited gain in optimizing the MS. Rather, attention should be given towards adjustments of the chromatographic gradient. Additionally, the ESI often yields higher instrumental responses for analytes eluting with a mobile phase of higher organic composition. This has been reported to be a consequence of the faster droplet drying rate in the ESI [137]. Therefore, evaluating different start and ending compositions for the gradient program could also contribute to a better method response.

5 Conclusions

The main objective of this thesis was to validate a quantitative LC-MS/MS method for the analysis of phenanthrene metabolites using rigorous statistical and chemometric approaches. Additionally, the study aimed to demonstrate the method's applicability to real samples and to optimize it if necessary.

A quantitative LC-MS/MS method was successfully validated for the analysis of seven different PAH metabolites, including five monohydroxylated and two dihydrodiol phenanthrene metabolites. Due to poor instrumental response, the dihydrodiols were only evaluated for the response function, LoD, LoQ, and selectivity. For the response function, the standard solutions were found to be sufficiently described by quadratic polynomials weighted with the reciprocal of their squared concentration. Linearity was demonstrated in the range 22.9-229 ng/g ww for 9-OHPHN, and 5.73-229 ng/g ww for the four remaining monohydroxylated metabolites. Trueness investigations revealed considerable bias and that the method concentration estimates required correction factors. After corrections, the method recoveries ranged from 95.7 to 109.5%. Corrected results also demonstrated good precision, with repeatability RSDs ranging from 6.1-50.6% and intermediate precision from 17.3-76.4%. The LoQs were determined to be the lowest validated concentration level, which corresponded to 115 ng/g ww for the dihydrodiols, 22.9 ng/g ww for 9-OHPHN, and 5.73 ng/g ww for the remaining monohydroxylated metabolites. For the four latter metabolites, their estimated LoDs ranged from 1.15 to 2.75 ng/g ww. For the dihydrodiols and 9-OHPHN, the LoD was set equal to the LoQ. Finally, the method demonstrated sufficient selectivity for all compounds.

The validated method was applied to samples obtained from two different exposure studies. QC samples were also run. In the first study, haddock eggs were exposed to the water-soluble fraction of crude oil and a fraction consisting only of PAHs. It was found that there was a greater metabolite production in the crude oil fractions than that of the PAH fractions. This ultimately suggests the presence of mixture effects, which are indeed reported in previous research. In the second study, the method was applied to haddock eggs exposed to isolated PAHs via passive dosing techniques, but this study failed to reveal any trends except that 9-OHPHN appeared to be the most abundant metabolite. The QC samples revealed that the method is able to produce accurate results within its validated range. Nonetheless, the samples from the exposure studies were found to possess concentrations below the LoQ. Optimization was thus attempted.

To optimize the method, chemometric approaches involving experimental designs were employed. By using factorial designs, it was found that the nozzle and capillary voltages in the ESI could be improved. These settings were optimized using the path of steepest ascent. A response surface was subsequently generated. The surface revealed that the current method was at a saddle point. Using canonical analysis, the method was further optimized by adjusting the gas flow and temperature. This ultimately provided a method response that was enhanced by 64% compared to the initial method.

Although the method failed to quantify the investigated samples, it is probable that with further optimization, these issues can be resolved. As a result, the initial validation and application of the method has shown promise for its use in elucidating the role of phenanthrene in crude oil toxicity.

6 Further Work

Further optimization may be necessary for the analytical method. At present, the largest gain in method response can likely be obtained via changes to the chromatography. It is therefore worth considering alternative solvents, such as MeOH and THF, as well as potential additives. Additional signal may also be obtained by changes to the gradient program. Potential adjustments to the injection volume and MS scan settings can also be considered. Moreover, for future experiments, collecting larger samples can also help in generating sufficient analyte signals.

Method validation needs to be conducted anew for the optimized method. Additionally, a lower concentration level needs to be validated to ensure that the method can quantify the analyte levels in the exposure samples. The validation samples extracted during this study can be reused for the validation of the optimized method; additional experimental work is thus only required for the new validation level. The new level should be extracted using the same protocol as the previous levels (*i.e.* two samples per day across five days). This is to keep the precision conditions as similar as possible as the previous samples.

7 References

1. Honda, M. & Suzuki, N. Toxicities of Polycyclic Aromatic Hydrocarbons for Aquatic Animals. *Int J Environ Res Public Health* **17,** 1363. ISSN: 1660-4601,1661-7827 (2020).

2. Sørensen, L. *et al.* Oil droplet fouling and differential toxicokinetics of polycyclic aromatic hydrocarbons in embryos of Atlantic haddock and cod. ISSN: 1932-6203 (2017).

3. Jesus, F. *et al.* A review on polycyclic aromatic hydrocarbons distribution in freshwater ecosystems and their toxicity to benthic fauna. *Sci Total Environ* **820,** 153282–153282. ISSN: 0048-9697 (2022).

4. Hjermann, D. Ø. *et al.* Fish and oil in the Lofoten–Barents Sea system: synoptic review of the effect of oil spills on fish populations. *Marine ecology. Progress series (Halstenbek)* **339,** 283–299. ISSN: 0171-8630 (2007).

5. Duran, R. & Cravo-Laureau, C. Role of environmental factors and microorganisms in determining the fate of polycyclic aromatic hydrocarbons in the marine environment. *FEMS Microbiol Rev* **40,** 814–830. ISSN: 0168-6445,1574-6976 (2016).

6. Chang, S. E., Stone, J., Demes, K. & Piscitelli, M. Consequences of oil spills: a review and framework for informing planning. *Ecology and society* **19,** 1–25. ISSN: 1708-3087 (2014).

7. Beyer, J., Trannum, H. C., Bakke, T., Hodson, P. V. & Collier, T. K. Environmental effects of the Deepwater Horizon oil spill: A review. ISSN: 0025-326X (2016).

8. Wiens, J. A. *Oil in the environment : legacies and lessons of the Exxon Valdez oil spill* Electronic Book. 2013.

9. Brakstad, O. G., Lewis, A. & Beegle-Krause, C. J. A critical review of marine snow in the context of oil spills and oil spill dispersant treatment with focus on the Deepwater Horizon oil spill. *Mar Pollut Bull* **135,** 346–356. ISSN: 0025-326X (2018).

10. Barron, M. G., Vivian, D. N., Heintz, R. A. & Yim, U. H. Long-Term Ecological Impacts from Oil Spills: Comparison of Exxon Valdez , Hebei Spirit , and Deepwater Horizon. *Environ Sci Technol* **54,** 6456–6467. ISSN: 0013-936X (2020).

11. Samiullah, Y. Biological effects of marine oil pollution. *Oil and petrochemical pollution* **2,** 235–264. ISSN: 0143-7127 (1985).

12. Fulton, M. H., Scott, G. I., Fortner, A., Bidleman, T. F. & Ngabe, B. The effects of urbanization on small high salinity estuaries of the southeastern United States. *Archives of Environmental Contamination and Toxicology* **25,** 476–484. ISSN: 1432-0703 (1993).

13. Wang, D., Tian, F., Yang, M., Liu, C. & Li, Y.-F. Application of positive matrix factorization to identify potential sources of PAHs in soil of Dalian, China. *Environ Pollut* **157,** 1559–1564. ISSN: 0269-7491 (2009).

14. Da Silva, D. A. M. & Bícego, M. C. Polycyclic aromatic hydrocarbons and petroleum biomarkers in São Sebastião Channel, Brazil: Assessment of petroleum contamination. *Mar Environ Res* **69,** 277–286. ISSN: 0141-1136 (2010).

15. Liu, X. *et al.* Assessing oil spill risk in the Chinese Bohai Sea: A case study for both ship and platform related oil spills. *Ocean and coastal management* **108,** 140–146. ISSN: 0964-5691 (2015).

16.   Wang, Y., Liu, X., Yu, X. & Zheng, X. Assessing response capabilities for responding to ship-related oil spills in the Chinese Bohai Sea. *International journal of disaster risk reduction* **28,** 251–257. ISSN: 2212-4209 (2018).

17.   Ferguson, A., Solo-Gabriele, H. & Mena, K. Assessment for oil spill chemicals: Current knowledge, data gaps, and uncertainties addressing human physical health risk. *Mar Pollut Bull* **150,** 110746–110746. ISSN: 0025-326X (2020).

18.   Chilvers, B. L., Morgan, K. J. & White, B. J. Sources and reporting of oil spills and impacts on wildlife 1970–2018. *Environ Sci Pollut Res Int* **28,** 754–762. ISSN: 0944-1344 (2021).

19.   Meador, J. P. & Nahrgang, J. Characterizing Crude Oil Toxicity to Early-Life Stage Fish Based On a Complex Mixture: Are We Making Unsupported Assumptions? *Environ. Sci. Technol* **53,** 11080–11092. ISSN: 0013-936X (2019).

20.   Sørhus, E. *et al.* Untangling mechanisms of crude oil toxicity: Linking gene expression, morphology and PAHs at two developmental stages in a cold-water fish. ISSN: 0048-9697 (2020).

21.   Incardona, J. P. Molecular Mechanisms of Crude Oil Developmental Toxicity in Fish. *Arch Environ Contam Toxicol* **73,** 19–32. ISSN: 0090-4341 (2017).

22.   Wu, B., Guo, S. & Wang, J. Assessment of the human health risk of polycyclic aromatic hydrocarbons in soils from areas of crude oil exploitation. *Environ Res* **193,** 110617. ISSN: 0013-9351 (2021).

23.   Lawal, A. T. Polycyclic aromatic hydrocarbons. A review. *Cogent environmental science* **3,** 1339841. ISSN: 2331-1843 (2017).

24.   Abdel-Shafy, H. I. & Mansour, M. S. M. A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation. *Egyptian journal of petroleum* **25,** 107–123. ISSN: 1110-0621 (2016).

25.   Burcham, P. C. *An Introduction to Toxicology* Electronic Book. 2014.

26.   Bukowska, B., Mokra, K. & Michałowicz, J. Benzo[a]pyrene-Environmental Occurrence, Human Exposure, and Mechanisms of Toxicity. *Int J Mol Sci* **23,** 6348. ISSN: 1422-0067,1661-6596 (2022).

27.   Bogdan, D. P. *Interaction of benzo(a)pyrene and nucleic acids in the presence of a mixed-function oxidase system* Thesis (1973).

28.   Carls, M. G. & Meador, J. P. A Perspective on the Toxicity of Petrogenic PAHs to Developing Fish Embryos Related to Environmental Chemistry. *Human and ecological risk assessment* **15,** 1084–1098. ISSN: 1080-7039 (2009).

29.   González-Doncel, M., González, L., Fernández-Torija, C., Navas, J. M. & Tarazona, J. V. *Aquat Toxicol* **87,** 280–288. ISSN: 0166-445X (2008).

30.   Donald, C. E. *et al.* Alkyl-phenanthrenes in early life stage fish: differential toxicity in Atlantic haddock (Melanogrammus aeglefinus) embryos. *Environ Sci Process Impacts* (2023).

31.   Ramirez, T. *et al.* Metabolomics in Toxicology and Preclinical Research. *ALTEX* **30,** 209–225. ISSN: 1868-596X (2013).

32.   Sweedler, J. V. Metabolomics in Analytical Chemistry. *Anal. Chem* **84,** 5833–5833. ISSN: 0003-2700 (2012).

33. Beyer, J., Jonsson, G., Porte, C., Krahn, M. M. & Ariese, F. Analytical methods for determining metabolites of polycyclic aromatic hydrocarbon (PAH) pollutants in fish bile: A review. *Environ Toxicol Pharmacol* **30,** 224–244. ISSN: 1382-6689 (2010).

34. Wang, Y. *et al.* Quantification of urinary mono-hydroxylated metabolites of polycyclic aromatic hydrocarbons by on-line solid phase extraction-high performance liquid chromatography-tandem mass spectrometry. *Anal Bioanal Chem* **409,** 931–937. ISSN: 1618-2642 (2017).

35. Li, Z. *et al.* Measurement of Urinary Monohydroxy Polycyclic Aromatic Hydrocarbons Using Automated Liquid-Liquid Extraction and Gas Chromatography/Isotope Dilution High-Resolution Mass Spectrometry. *Anal. Chem* **78,** 5744–5751. ISSN: 0003-2700 (2006).

36. Campo, L., Rossella, F. & Fustinoni, S. Development of a gas chromatography/mass spectrometry method to quantify several urinary monohydroxy metabolites of polycyclic aromatic hydrocarbons in occupationally exposed subjects. *J Chromatogr B Analyt Technol Biomed Life Sci* **875,** 531–540. ISSN: 1570-0232 (2008).

37. Krahn, M. M., Myers, M. S., Burrows, D. G. & Malins, D. C. Determination of metabolites of xenobiotics in the bile of fish from polluted waterways. *Xenobiotica* **14,** 633–646. ISSN: 0049-8254 (1984).

38. Dissanayake, A. & Galloway, T. S. Evaluation of fixed wavelength fluorescence and synchronous fluorescence spectrophotometry as a biomonitoring tool of environmental contamination. *Mar Environ Res* **58,** 281–285. ISSN: 0141-1136 (2004).

39. Pampanin, D. M. & Sydnes, M. O. *Petrogenic polycyclic aromatic hydrocarbons in the aquatic environment : analysis, synthesis, toxicity and environmental impact* Electronic Book. 2017.

40. Saha, M. *et al.* Sources of sedimentary PAHs in tropical Asian waters: Differentiation between pyrogenic and petrogenic sources by alkyl homolog abundance. *Mar Pollut Bull* **58,** 189–200. ISSN: 0025-326X (2009).

41. Alegbeleye, O. O., Opeolu, B. O. & Jackson, V. A. Polycyclic Aromatic Hydrocarbons: A Critical Review of Environmental Occurrence and Bioremediation. *Environ Manage* **60,** 758–783. ISSN: 0364-152X (2017).

42. Neff, J. M., Stout, S. A. & Gunster, D. G. Ecological Risk Assessment of Polycyclic Aromatic Hydrocarbons in Sediments: Identifying Sources and Ecological Hazard. *Integr Environ Assess Manag* **1,** 22–33. ISSN: 1551-3777 (2005).

43. Manousi, N. & Zachariadis, G. A. Recent Advances in the Extraction of Polycyclic Aromatic Hydrocarbons from Environmental Samples. *Molecules* **25,** 2182. ISSN: 1420-3049 (2020).

44. Arnot, J. A. & Gobas, F. A. P. C. A food web bioaccumulation model for organic chemicals in aquatic ecosystems. *Environmental Toxicology and Chemistry* **23,** 2343–2355. ISSN: 0730-7268 (2004).

45. McGrath, J. A., Joshua, N., Bess, A. S. & Parkerton, T. F. Review of Polycyclic Aromatic Hydrocarbons (PAHs) Sediment Quality Guidelines for the Protection of Benthic Life. *Integr Environ Assess Manag* **15,** 505–518. ISSN: 1551-3777 (2019).

46. Sørensen, L. *et al.* Accumulation and toxicity of monoaromatic petroleum hydrocarbons in early life stages of cod and haddock. *Environ Pollut* **251,** 212–220. ISSN: 0269-7491 (2019).

47.   Gauthier, P. T., Norwood, W. P., Prepas, E. E. & Pyle, G. G. Metal–PAH mixtures in the aquatic environment: A review of co-toxic mechanisms leading to more-than-additive outcomes. *Aquat Toxicol* **154,** 253–269. ISSN: 0166-445X (2014).

48.   Michel, C. *et al.* Seasonal and PAH impact on DNA strand-break levels in gills of transplanted zebra mussels. *Ecotoxicol Environ Saf* **92,** 18–26. ISSN: 0147-6513 (2013).

49.   Meador, J. P., Casillas, E., Sloan, C. A. & Varanasi, U. Comparative bioaccumulation of polycyclic aromatic hydrocarbons from sediment by two infaunal invertebrates. *Marine ecology. Progress series (Halstenbek)* **123,** 107–124. ISSN: 0171-8630 (1995).

50.   Van der Oost, R., Beyer, J. & Vermeulen, N. P. E. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. *Environ Toxicol Pharmacol* **13,** 57–149. ISSN: 1382-6689 (2003).

51.   De Lange, H. J., Peeters, E. T. H. M., Harmsen, J., Maas, H. & De Jonge, J. Seasonal variation of total and biochemically available concentrations of PAHs in a floodplain lake sediment has no effect on the benthic invertebrate community. *Chemosphere* **75,** 319–326. ISSN: 0045-6535 (2009).

52.   Cohen, D. M., Inada, T., Iwamoto, T. & Scialabba, N. *FAO species catalogue : Vol. 10 : Gadiform fishes of the world : (order Gadiformes) : an annotated and illustrated catalogue of cods, hakes, grenadiers and other gadiform fishes known to date* ISBN: 0014-5602 9251028907 (Food and Agriculture Organization of the United Nations, Rome, 1990).

53.   Britannica. *Haddock* Encyclopedia. 2022. Accessed 23/01/23. https : / / www . britannica.com/animal/haddock.

54.   The Norwegian Directorate Of Fisheries, *Fangst og kvoter* Web Page. 2023. Accessed 23/01/23. https://www.fiskeridir.no/Yrkesfiske/Tall-og-analyse/ Fangst-og-kvoter.

55.   Sørhus, E. *et al.* Unexpected interaction with dispersed crude oil droplets drives severe toxicity in Atlantic haddock embryos. *PLoS One* **10,** e0124376. ISSN: 1932-6203 (2015).

56.   Gallo, M. A. *History and Scope of Toxicology* in *Casarett and Doull's Toxicology : The Basic Science of Poisons* (eds Casarett, L. J. & Klaassen, C. D.) 3rd, 1–5 (McGraw-Hill Medical, New York, 2015). ISBN: 9780071847087.

57.   Lehman-McKeeman, L. *Absorption, Distribution, and Excretion of Toxicants* in *Casarett and Doull's Toxicology : The Basic Science of Poisons* (eds Casarett, L. J. & Klaassen, C. D.) 3rd, 61–75 (McGraw-Hill Medical, New York, 2015). ISBN: 9780071847087.

58.   Schanker, L. S. Mechanisms of Drug Absorption and Distribution. *Annual review of pharmacology* **1,** 29–45. ISSN: 0066-4251 (1961).

59.   Tierney, K. B., Kennedy, C. J., Gobas, F., Gledhill, M. & Sekela, M. *Organic Contaminants and Fish* in, 1–52 (United States: Elsevier Science and Technology, United States, 2014). ISBN: 1546-5098 9780123982544,0123982545.

60.   Sahu, S. K. & Pandit, G. G. Estimation of Octanol-Water Partition Coefficients for Polycyclic Aromatic Hydrocarbons Using Reverse-Phase HPLC. *Journal of liquid chromatography and related technologies* **26,** 135–146. ISSN: 1082-6076 (2003).

61. Qiao, P. & Farrell, A. P. Uptake of hydrophobic xenobiotics by fish in water laden with sediments from the fraser river. *Environmental Toxicology and Chemistry* **15,** 1555–1563. ISSN: 0730-7268 (1996).

62. Kennedy, C. J., Gill, K. A. & Walsh, P. J. In-vitro metabolism of benzo(a)pyrene in the blood of the gulf toadfish, Opsanus beta. *Marine environmental research* **31,** 37–53. ISSN: 0141-1136 (1991).

63. Clark, K. E., Gobas, F. A. P. C. & Mackay, D. Model of organic chemical uptake and clearance by fish from food and water. *Environ. Sci. Technol* **24,** 1203–1213. ISSN: 0013-936X (1990).

64. A, P. *et al. Biotransformation of Xenobiotics* in *Casarett and Doull's Toxicology : The Basic Science of Poisons* (eds Casarett, L. J. & Klaassen, C. D.) 3rd, 79–106 (McGraw-Hill Medical, New York, 2015). ISBN: 9780071847087.

65. Seidel, A. *et al.* New Biomarkers of Occupational Exposure to Polycyclic Aromatic Hydrocarbons. *J Toxicol Environ Health A* **71,** 734–745. ISSN: 1528-7394 (2008).

66. Hunn, J. B. & Allen, J. L. Movement of Drugs Across the Gills of Fishes. *Annual review of pharmacology* **14,** 47–54. ISSN: 0066-4251 (1974).

67. Christiansen, J. S., Dalmo, R. A. & Ingebrigtsen, K. Xenobiotic excretion in fish with aglomerular kidneys. *Marine ecology. Progress series (Halstenbek)* **136,** 303–304. ISSN: 0171-8630 (1996).

68. Vuorinen, P. J. *et al.* Use of biliary PAH metabolites as a biomarker of pollution in fish from the Baltic Sea. *Mar Pollut Bull* **53,** 479–487. ISSN: 0025-326X (2006).

69. Jaeschke, H. *Toxic Responses of the Liver* in *Casarett and Doull's Toxicology : The Basic Science of Poisons* (eds Casarett, L. J. & Klaassen, C. D.) 3rd, 195–208 (McGraw-Hill Medical, New York, 2015). ISBN: 9780071847087.

70. Larsson, M., Giesy, J. P. & Engwall, M. AhR-mediated activities of polycyclic aromatic compound (PAC) mixtures are predictable by the concept of concentration addition. *Environ Int* **73,** 94–103. ISSN: 0160-4120,1873-6750 (2014).

71. Vondráček, J. *et al.* Aryl Hydrocarbon Receptor-Activating Polychlorinated Biphenyls and Their Hydroxylated Metabolites Induce Cell Proliferation in Contact-Inhibited Rat Liver Epithelial Cells. *Toxicol. Sci* **83,** 53–63. ISSN: 1096-6080,1096-0929 (2005).

72. Jiang, H. *et al.* Metabolism of Benzo[a]pyrene in Human Bronchoalveolar H358 Cells Using Liquid Chromatography–Mass Spectrometry. *Chem. Res. Toxicol* **20,** 1331–1341. ISSN: 0893-228X (2007).

73. Harris, D. C. & Lucy, C. A. *Quantitative chemical analysis* Tenth edition. ISBN: 9781319324506,1319324509 (Macmillan International Higher Education, New York, NY, 2020).

74. Miller, J. M. *Chromatography : concepts and contrasts* 2nd. ISBN: 0471472077,9780470530252 (Wiley, Hoboken, N.J, 2005).

75. Raja, P. M. V. & Barron, A. R. *High Performance Liquid chromatography* Web Page. 2022. Accessed 24/04/23. https://chem.libretexts.org/@go/page/55862.

76. Gross, J. H. *Mass Spectrometry : A Textbook* Electronic Book. 2017.

77. Banerjee, S. & Mazumdar, S. Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte. *Int J Anal Chem* **2012,** 282574–40. ISSN: 1687-8760 (2012).

78. Savaryn, J. P., Toby, T. K. & Kelleher, N. L. A researcher's guide to mass spectrometry-based proteomics. *Proteomics* **16,** 2435–2443. ISSN: 1615-9853 (2016).

79. Furey, A., Moriarty, M., Bane, V., Kinsella, B. & Lehane, M. Ion suppression; A critical review on causes, evaluation, prevention and applications. *Talanta* **115,** 104–122. ISSN: 0039-9140 (2013).

80. Taylor, P. J. Matrix effects: the Achilles heel of quantitative high-performance liquid chromatography–electrospray–tandem mass spectrometry. *Clin Biochem* **38,** 328–334. ISSN: 0009-9120 (2005).

81. Falaki, F. *Gas Chromatography* in *Sample Preparation Techniques for Gas Chromatography* (ed Kusch, P.) chap. 3 (IntechOpen, Rijeka, 2019).

82. Bureau international des poids et mesures, *Guide to the expression of uncertainty in measurement* ISBN: 9267101889 (International Organization for Standardization, Genève, 1993).

83. Magnusson, B., Näykki, T., Hovind, H., Krysell, M. & Sahlin, E. Handbook for Calculation of Measurement Uncertainty in Environmental Laboratories. *Nordtest TR 537 edition 4* (2017).

84. Miller, J. N. & Miller, J. C. *Statistics and chemometrics for analytical chemistry* in. 6th, 1–15 (Pearson Prentice Hall, Harlow, 2010). ISBN: 9780273730422.

85. Devore, J. L., Berk, K. N., Carlton, M. A. & Devore, J. L. *Modern mathematical statistics with applications* Third edition. ISBN: 9783030551551 (Springer, Cham, 2021).

86. Miller, J. N. & Miller, J. C. *Statistics and chemometrics for analytical chemistry* in. 6th, 37–74 (Pearson Prentice Hall, Harlow, 2010). ISBN: 9780273730422.

87. Lay, D. C., Lay, D. C., Lay, S. R. & McDonald, J. J. *Linear algebra and its applications* in. 6th global edition., 347–410 (Pearson Education, Harlow, Essex, 2021). ISBN: 9781292351216.

88. Dobson, A. J. & Barnett, A. G. *An introduction to generalized linear models* in. 3rd, 89–122 (CRC Press, Boca Raton, FL, 2008). ISBN: 0-367-80784-X,1-58488-951-9.

89. Miller, J. N. & Miller, J. C. *Statistics and chemometrics for analytical chemistry* in. 6th, 120–164 (Pearson Prentice Hall, Harlow, 2010). ISBN: 9780273730422.

90. Dobson, A. J. & Barnett, A. G. *An introduction to generalized linear models* in. 3rd, 1–18 (CRC Press, Boca Raton, FL, 2008). ISBN: 0-367-80784-X,1-58488-951-9.

91. Willett, J. B. & Singer, J. D. Another Cautionary Note about R 2: Its Use in Weighted Least-Squares Regression Analysis. *The American statistician* **42,** 236–238. ISSN: 0003-1305 (1988).

92. Kutner, M., Nachtsheim, C., Neter, J. & Li, W. *Applied Linear Statistical Models* Fifth Edition. ISBN: 978-0073108742 (McGraw-Hill Higher Education, 2005).

93. Mandel, J. *The Statistical Analysis of Experimental Data* (Interscience, New York, 1964).

94. Dobson, A. J. & Barnett, A. *Introduction to Generalized Linear Models, Third Edition* ISBN: 9781584889502,1584889500 (Philadelphia, PA: CRC Press LLC, Philadelphia, PA, 2008).

95. Thelin, B., Lundstedt, T., Seifert, E., Abramo, L. & Calrson, R. *Försöksplanering och optimering för eksperimentalisten* in *Anvendelse av kjemometri innen forskning og industri* (eds Nordtvedt, R., Brakstad, F., Kvalheim, O. & Lundstedt, T.) 81–106 (Tidsskriftforlaget Kjemi, Oslo, 1996). ISBN: 8291294011.

96. Box, G. E. P., Hunter, J. S. & Hunter, W. G. *Statistics for experimenters : design, innovation, and discovery* Second edition. ISBN: 9780471718130 (Wiley-Interscience, Hoboken. N.J, 2005).

97. Montgomery, D. C. *Design and analysis of experiments* 10th. ISBN: 978-1119722106 (Wiley, Hoboken, N.J, 2019).

98. Myers, R. H., Montgomery, D. C. & Anderson-Cook, C. M. *Response surface methodology : process and product optimization using designed experiments* Electronic Book. 2009.

99. Andrade, J. & Gómez Carracedo, M. P. Notes on the use of Mandel's test to check for nonlinearity in laboratory calibrations. *Anal. Methods* **5,** 1145–1149 (2013).

100. Prichard, F. E. & Barwick, V. *Quality Assurance in Analytical Chemistry* in, 73–98 (John Wiley & Sons ; VAM/LGC, Chichester ;,Hoboken, NJ :,Teddington, Middlesex, 2007). ISBN: 1-280-97410-9,9786610974108,0-470-51777-8,0-470-51776-X.

101. Kollipara, S., Bende, G., Agarwal, N., Varshney, B. & Paliwal, J. International Guidelines for Bioanalytical Method Validation: A Comparison and Discussion on Current Scenario. *Chromatographia* **73,** 201–217. ISSN: 0009-5893 (2011).

102. Hubert, P. *et al.* Harmonization of strategies for the validation of quantitative analytical procedures - A SFSTP proposal - Part II. *J Pharm Biomed Anal* **45,** 70–81. ISSN: 0731-7085,1873-264X (2007).

103. Agency, E. *Committee for Medicinal Products for Human Use (EMEA/CHMP/EWP/192217/2009)* (2011).

104. Ellison, S., Farrant, T., Barwick, V. & Chemistry, R. S. o. *Practical Statistics for the Analytical Scientist: A Bench Guide* in, 144–160 (Royal Society of Chemistry, 2009). ISBN: 9780854041312.

105. Bureau international des poids et mesures, & International Organization for Standardization, *Guide to the expression of uncertainty in measurement* Generic. 2008. https://www.bipm.org/documents/20126/2071204/JCGM_100_2008_E.pdf/cb0ef43f-baa5-11cf-3f85-4dcd86f77bd6..

106. Ermer, J. & Nethercote, P. *Method Validation in Pharmaceutical Analysis - A Guide to Best Practice* in. Second edition, 73–182 (Wiley-VCH, Weinheim, Germany, 2004). ISBN: 3-030-55156-3.

107. Hubert, P. *et al.* Harmonization of strategies for the validation of quantitative analytical procedures - A SFSTP proposal - Part III. *J Pharm Biomed Anal* **45,** 82–96. ISSN: 0731-7085,1873-264X (2007).

108. Boix, A., Bellorini, S. & von Holst, C. Validation of an analytical method for the determination of glyceroltriheptanoate (GTH) in processed animal by-products: Results of a collaborative study. *Food Addit Contam Part A Chem Anal Control Expo Risk Assess* **27,** 793–800. ISSN: 1944-0049 (2010).

109. Horwitz, W. & Albert, R. The Horwitz Ratio (HorRat): A Useful Index of Method Performance with Respect to Precision. *Journal of AOAC INTERNATIONAL* **89,** 1095–1109. ISSN: 1060-3271 (2019).

110. Latimer George W., J. *Guidelines for Collaborative Study Procedures to Validate Characteristics of a Method of Analysis* in *Official Methods of Analysis of AOAC INTERNATIONAL 22nd Edition* 0 (Oxford University Press, 2023). ISBN: 9780197610138.

111. González, A. G. & Herrador, M. Á. Accuracy profiles from uncertainty measurements. *Talanta* **70,** 896–901. ISSN: 0039-9140 (2006).

112. González, A. G., Herrador, M. Á. & Asuero, A. G. Intra-laboratory assessment of method accuracy (trueness and precision) by using validation standards. *Talanta* **82,** 1995–1998. ISSN: 0039-9140 (2010).

113. Satterthwaite, F. E. An Approximate Distribution of Estimates of Variance Components. *Biometrics bulletin* **2,** 110–114. ISSN: 0099-4987 (1946).

114. Taylor, B. & Kuyatt, C. Guidelines for Evaluating and Expressing the Uncertainty of NIST Measurement Results. *NIST Technical Note 1297* (1994).

115. Hoffman, D. & Kringle, R. TWO-SIDED TOLERANCE INTERVALS FOR BALANCED AND UNBALANCED RANDOM EFFECTS MODELS. *J Biopharm Stat* **15,** 283–293. ISSN: 1054-3406 (2005).

116. Van Loco, J. *et al.* Calculation of the decision limit and the detection capability for banned substances: The imperfect marriage between the quantitative and the qualitative criteria. *Anal Chim Acta* **586,** 8–12. ISSN: 0003-2670 (2007).

117. Vogelgesang, J. & Hadrich, J. Limits of detection, identification and determination: a statistical approach for practitioners. *Accreditation and quality assurance* **3,** 242–255. ISSN: 0949-1775 (1998).

118. Kruve, A. *et al.* Tutorial review on validation of liquid chromatography–mass spectrometry methods: Part I. *Anal Chim Acta* **870,** 29–44. ISSN: 0003-2670 (2015).

119. Brevik, M. *Initial Steps in the Development of a Quantitative LC-MS/MS Method for PAH Metabolites in Marine Fish* Thesis (2021).

120. Magnusson, B. & Örnemark, U. *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics* Electronic Book. 2014.

121. Hubert, P. *et al.* Harmonization of strategies for the validation of quantitative analytical procedures : A SFSTP proposal Part IV. Examples of application. *J Pharm Biomed Anal* **48,** 760–771. ISSN: 0731-7085,1873-264X (2008).

122. Gustavo Gonzalez, A. & Angeles Herrador, M. A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles. *TrAC, Trends in analytical chemistry (Regular ed.)* **26,** 227–238. ISSN: 0165-9936 (2007).

123. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use, *ICH HARMONISED GUIDELINE VALIDATION OF ANALYTICAL PROCEDURES Q2(R2)* Generic. 2022. https://www.ema.europa.eu/en/documents/scientific-guideline/ich-guideline-q2r2-validation-analytical-procedures-step-2b_en.pdf.

124. Da Silva, D. A. M., Gates, J. B., O'Neill, S. M., West, J. E. & Ylitalo, G. M. Assessing hydroxylated polycyclic aromatic hydrocarbon (OHPAH) metabolites in bile of English sole (Parophrys vetulus) from Puget Sound, WA, USA by liquid chromatography/tandem mass spectrometry (LC-MS/MS). *Sci Total Environ* **865,** 161229. ISSN: 0048-9697 (2023).

125. Onyemauwa, F. *et al.* Using liquid chromatography–tandem mass spectrometry to quantify monohydroxylated metabolites of polycyclic aromatic hydrocarbons in urine. *Journal of Chromatography B* **877,** 1117–1125. ISSN: 1570-0232 (2009).

126. Hunsberger, I. M., Ketcham, R. & Gutowsky, H. S. The Determination of Double-bond Character in Cyclic Systems. II. The 9,10-Bond of Phenanthrene. *J. Am. Chem. Soc* **74,** 4839–4845. ISSN: 0002-7863 (1952).

127. Roberts, J. D. & Caserio, M. C. *Basic principles of organic chemistry* 2nd. ISBN: 0805383298; 978-0805383294 (W.A. Benjamin, New York, 1977).

128. McMurry, J. *Organic chemistry* Ninth edition. ISBN: 9781305080485 (Cengage Learning, Australia, 2016).

129. Vollhardt, K. P. C. & Schore, N. E. *Organic chemistry : structure and function* 7th. ISBN: 9781464120275,1464120277 (W.H. Freeman, New York, 2014).

130. Gore, P. H. Abnormal Substitution Reactions of Anthracene and Phenanthrene. *J. Org. Chem* **22,** 135–138. ISSN: 0022-3263 (1957).

131. Linsinger, T. P. J. Use of recovery and bias information in analytical chemistry and estimation of its uncertainty contribution. *TrAC, Trends in analytical chemistry (Regular ed.)* **27,** 916–923. ISSN: 0165-9936 (2008).

132. Mordehai, A. & FJeldsted, J. *Agilent Jet Stream Thermal Gradient Focusing Technology* Generic. 2009. https://www.agilent.com/cs/library/technicaloverviews/Public/5990-3494en_lo%20CMS.pdf.

133. Huffman, B. A., Poltash, M. L. & Hughey, C. A. Effect of Polar Protic and Polar Aprotic Solvents on Negative-Ion Electrospray Ionization and Chromatographic Separation of Small Acidic Molecules. *Anal. Chem* **84,** 9942–9950. ISSN: 0003-2700 (2012).

134. Wu, Z. *et al.* Favorable Effects of Weak Acids on Negative-Ion Electrospray Ionization Mass Spectrometry. *Anal. Chem* **76,** 839–847. ISSN: 0003-2700 (2004).

135. Sanz Alvarez, I. M. *Analysis of Fatty Acids by High Performance Liquid Chromatography and Electrospray Ionization-Mass Spectrometry* Thesis (2019).

136. Sichilongo, K., Kolanyane, P., Mazimba, O. & Masesane, I. ESI behavior of thiamphenicol and florfenicol in methanol and tetrahydrofuran solutions for LC-MS. *International Journal of Mass Spectrometry* **367** (2014).

137. Liigand, P., Liigand, J., Kaupmees, K. & Kruve, A. 30 Years of research on ESI/MS response: Trends, contradictions and applications. *Analytica Chimica Acta* **1152,** 238117. ISSN: 0003-2670 (2021).

8 Appendices

A Selected Mathematical and Statistical Derivations

A.1 Canonical Analysis

Recall that the response surface is a second degree polynomial estimating some true response. Consequently, as the function is of second order, there can only be a single stationary point in the estimated response function. It is of interest to establish whether this point is a maximum, minimum or a saddle point. This evaluation is simple if there are only two independent variables of interest. However, if there are more three or more independent variables influencing the response surface, other means of evaluating the stationary point must be applied. A formal analysis of the multivariate response surface may be done by *canonical analysis* [96–98].

A second-order model can be expressed as

$$y = b_0 + x^T b + x^T B x \tag{A.1}$$

Where

$$x = \begin{bmatrix} x_1 \\ \vdots \\ x_I \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ \vdots \\ b_I \end{bmatrix}, \quad B = \begin{bmatrix} b_{11} & b_{12}/2 & \cdots & b_{1I}/2 \\ & b_{22} & \vdots & b_{2I}/2 \\ & & \ddots & \vdots \\ sym. & & & b_{II} \end{bmatrix}$$

Here, b is a vector of length I containing the first-order coefficients (main effects) whereas the matrix B is a symmetric $I \times I$ matrix with the pure quadratic terms along the diagonal and mixed quadratic (interactions) along the off-diagonals. To find the stationary point of the response surface, the derivative of (A.1) is set to zero. Akin to the scalar derivatives $\frac{d}{dx}ax = x$ and $\frac{d}{dx}ax^2 = 2ax$, matrix calculus shows that

$$\frac{\partial}{\partial x}x^T b = b \quad \text{and} \quad \frac{\partial}{\partial x}x^T B x = 2Bx$$

Thus the stationary point, x_s, can be found equating the second order model to zero

$$\frac{\partial y}{\partial x} = b + 2Bx_s = 0$$

Meaning

$$x_s = -\frac{1}{2}B^{-1}b \tag{A.2}$$

Substituting (A.2) into (A.1) yields the response at the stationary point, y_s. Due to the B's symmetry $x_s^T = \left(-\frac{1}{2}B^{-1}b\right)^T = -\frac{1}{2}b^T B^{-1}$. Thus

$$y_s = b_0 + x_s^T b + x_s^T B x_s \tag{A.3}$$

111

$$y_s = b_0 + x_s^T b + \left(-\frac{1}{2} b^T B^{-1}\right) B x_s = b_0 + x_s^T b - \frac{1}{2} x_s^T b$$

$$y_s = b_0 + \frac{1}{2} x_s^T b \tag{A.4}$$

However, this equation can re-expressed via eigendecomposition. If P is the normalized eigenvectors of B and Λ is a diagonal matrix with its corresponding eigenvalues, then

$$PBP^T = \Lambda \tag{A.5}$$

Moreover, the model can be centered and undergo a change of basis

$$z = x - x_s$$

$$w = P^T z \tag{A.6}$$

This can be substituted into (A.1) to yield

$$y = b_0 + (z + x_s)^T b + (z + x_s) B (z + x_s)$$

Multiplying and collecting the terms gives

$$y = [b_0 + x_s^T b + x_s^T B x_s] + [z^T B + z^T B z + 2 x_s^T B z] \tag{A.7}$$

Moreover, from (A.2),

$$x_s^T = \frac{1}{2} b^T B^{-1}$$

$$2 x_s^T B = -b^T$$

Multiplying both sides by z

$$2 x_s^T B z = -b^T z = -z^T b$$

Substituting this expression into (A.7) and recalling (A.3) allows the following simplification

$$y = [b_0 + x_s^T b + x_s^T B x_s] + [z^T B z + z^T b - z^T b]$$

$$y = y_s + z^T B z \tag{A.8}$$

Since P consists of normalized eigenvectors, its transpose is equal to its inverse (*i.e.* P is orthogonal). Thus, $z = wP$. Substituting this into (A.8)

$$y = y_s + w^T P^T B P w$$

$$y = y_s + w^T \Lambda w$$

Which can be expressed as

$$y = y_s + \sum_{i=1}^{I} \lambda_i w_i^2 \tag{A.9}$$

Where $\lambda_1, \lambda_2..., \lambda_I$ refers to the eigenvalues of B and the corresponding variables $w_1, w_2, ..., w_I$ are the *canonical variables*.

The magnitude of the eigenvalues reveals how the response changes as one moves away from the center, x_s. Therefore, if all eigenvalues are negative, the response will decrease as one moves away from the center. Therefore, all negative magnitudes means that the center is a maximum. Equivalently, if all eigenvalues are positive, the center is a minimum. If there is a combination of positive and negative signs, the center is a saddle point.

A.2   The Regression Approach to the Analysis of Variance

This section provides a brief example on the usage of OLS techniques for conduction of the ANOVA F-test. In actuality, OLS estimation is actually the most common technique applied for ANOVA estimation in statistical software. Although it is more commonly known as the *corner point parameterization* technique [94].

Assume an analyst investigates effect of various buffer concentrations in a mobile phase. Denote the three buffer concentrations A, B and C. There were two replicates per concentration, yielding responses 0.26 and 0.31 for A; 0.38 and 0.43 for B; and 0.23 and 0.28 for C.

The full regression model for the response becomes

$$Y_{ij} = \mu_A X_{ijA} + \mu_B X_{ijB} + \mu_C X_{ijC} + \varepsilon_{ij}$$

The design matrix, X_F, for the full model is given by

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

The reduced model consists only of an intercept and is represented by

$$Y_{ij} = \mu_0 + \varepsilon_{ij}$$

The design matrix, X_R, become

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Calculating both models it can be found that the full model y_F becomes

$$y_F = 0.285 X_{ij1} + 0.405 X_{ij2} + 0.255 X_{ij3}$$

and the reduced model, y_R

$$y_R = 0.315$$

Using these models, the residuals and the associated residual standard error, be estimated. For the full and reduced model, respectively, the residual standard errors becomes

$$SS_E(F) = 0.00075 \text{ and } SS_E(R) = 0.00579$$

The degrees of freedom for the numerator and denominator are given by $n - I$, where n is the number of data points and I the number of model parameters. As there are 6 data points and 3 parameters in the full model, $SS_E(F)$ have 3 degrees of freedom. Similarly, $SS_E(R)$ has 5. Recalling the Equation (25):

$$F_{v_1, v_2} = \frac{SS_E(R) - SS_E(F)}{v_R - v_F} \div \frac{SS_E(F)}{v_F}$$

The F-statistic can be found via

$$F_{5-3,3} = \frac{0.00579 - 0.00075}{5 - 3} \div \frac{0.00075}{3} = 10.1$$

For the test statistic $F_{2,3} = 10.1$, a corresponding p-value of $p = 0.465$.

The corresponding ANOVA output from the *aov()* function in R, provides $F_{2,3} = 10.14$ and $p = 0.466$. The differences between the two sets of answers are easily attributed to the number of significant figures.

A.3   Calculating Expanded Measurement Uncertainties from Validation Data

This appendix provides an example on how the formulas derived in Section 4.1 can be used to estimate the MU. For illustrative purposes, the measurement result of interest was chosen to be 4-OHPHN from the QC sample extracted on the first day.

After surrogate corrections, the initial 4-OHPHN concentration (x) was found to be 94.598 ng/g ww. This estimate has to be further adjusted by the bias CFs. Thus, from Table 15, it is found that the corrected 4-OHPHN (y) is given by

$$y = (x + c_b)c_p = (94.598 + (-0.659)) \cdot 1.250 = 117.4 \text{ ng/g ww}$$

In this case, the spike concentration was at 115 ng/g ww. Thus, the proposed estimate seems reasonable.

The next step is to find find all uncertainty components. Recall that the uncertainty of the final estimate y is given by

$$u_c(y) = y \sqrt{\frac{u^2(r)}{IJ(\bar{x} + c_b)^2} + \frac{u^2(IP) - u^2(r)}{I(\bar{x} + c_b)^2} + \frac{u^2(c_b)}{(\bar{x} + c_b)^2} + \frac{u^2(c_p)}{c_p^2}} \tag{67}$$

In this case, it is known that the sample concentration is exactly 115 ng/g ww. Therefore, the repeatability and IP at this concentration level can be used. From Table 16, it is found that the RSD repeatability and IP of x are $RSD_{n,r} = 6.1\%$ and $RSD_{n,r} = 17.2\%$. Thus, the estimated precision components of x becomes

$$u(r) = 94.598 \cdot 6.1\% = 5.77 \text{ ng/g ww and } u(IP) = 94.598 \cdot 17.2\% = 16.27 \text{ ng/g ww}$$

Additionally, $u(c_p) = 0.045$ and $u(c_b) = 0.972$. Finally, as there was only replicate, $I = J = 1$. The combined uncertainty becomes

$$u_c(y) = 117.4 \sqrt{\frac{5.77^2}{(94.598 + \{-0.659\})^2} + \frac{16.27^2 - 5.77^2}{(94.598 + \{-0.659\})^2} + \frac{0.972^2}{(94.598 + \{-0.659\})^2} + \frac{0.045^2}{1.250^2}}$$

$$u_c(y) = 20.8 \text{ ng/g ww}$$

The degrees of freedom for $u_c(y)$ is provided by

$$\nu_{\text{eff}} = u_c^4(y) \cdot \left[\left(\frac{c_p^2 \cdot u^2(r)}{IJ \sqrt{\nu_r}} \right)^2 + \left(\frac{c_p^2 \cdot \{u^2(IP) - u^2(r)\}}{I \sqrt{\nu_b}} \right)^2 + \left(\frac{c_p^2 \cdot u^2(c_b)}{\sqrt{\nu_c}} \right)^2 + \left(\frac{\{\bar{x} + c_b\}^2 \cdot u^2(c_p)}{\sqrt{\nu_c}} \right)^2 \right]^{-1} \tag{68}$$

The degrees of freedom for repeatability, between-day variance are denoted ν_r and ν_b respectively. Similarly, the degrees of freedom for the constant and proportional CFs are given by ν_b and ν_p. An accuracy study was conducted over five days with two replicates per day, resulting in $\nu_r = 5$ and $\nu_b = 4$, as shown in Section 2.4.2 The linearity of monohydroxylated metabolites, except 9-OHPHN, was evaluated with 40 measurements (30 for 9-OHPHN). Therefore, the linearity correction factors have $\nu_b = \nu_p = 38$ degrees of freedom (28 for 9-OHPHN).

$$\nu_{\text{eff}} = 20.8^4 \cdot \left[\left(\frac{1.250^2 \cdot 5.77^2}{1\,\sqrt{5(2-1)}} \right)^2 + \left(\frac{1.250^2 \cdot \{16.27^2 - 5.77^2\}}{1\,\sqrt{5-1}} \right)^2 \right.$$
$$\left. + \left(\frac{1.250^2 \cdot 0.972^2}{\sqrt{38}} \right)^2 + \left(\frac{\{94.598 + (-0.659)\}^2 \cdot 0.045^2}{\sqrt{38}} \right)^2 \right]^{-1}$$
$$\nu_{\text{eff}} = 5.6$$

The final uncertainty is given by

$$U_{95\%} = y \pm u_c(y) \cdot t^*_{0.975,\nu_{\text{eff}}}$$

In many cases $t^*_{0.975,\nu_{\text{eff}}}$ is simply taken to be 2. This simplification is justified by the fact that $t^*_{0.975,\nu_{\text{eff}}} \approx z^*_{0.975} = 1.96$ for sufficiently large values of ν_{eff}. While this simplification is very widespread, blind application of this coverage factor can provide considerable errors if ν_{eff} is too small. As a rule of thumb, it has been proposed that the usage of 2 as a coverage factor should be avoided if ν_{eff} is lower than 8 [114].

In this case, since $\nu_{\text{eff}} = 5.6$, the use of the approximation is necessary. It it further found that $t^*_{0.975,5.6} = 2.49$. The expanded MU becomes

$$U_{95\%} = 117 \pm 2.5 \cdot 20.8$$

$$U_{95\%} = 117 \pm 52 \text{ ng/g ww}$$

B   Statistical Outputs
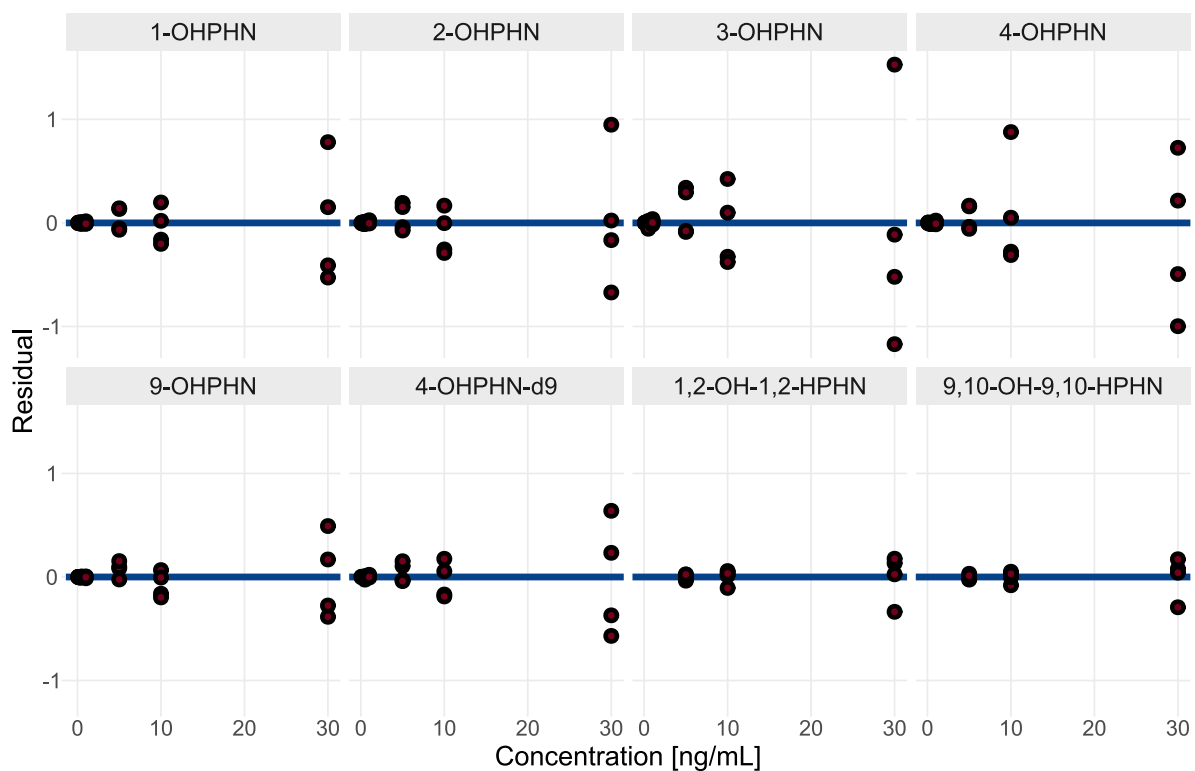
Figure B.1: Non-Standardized residual plots for a FOM with weights, $w = 1/x^2$.
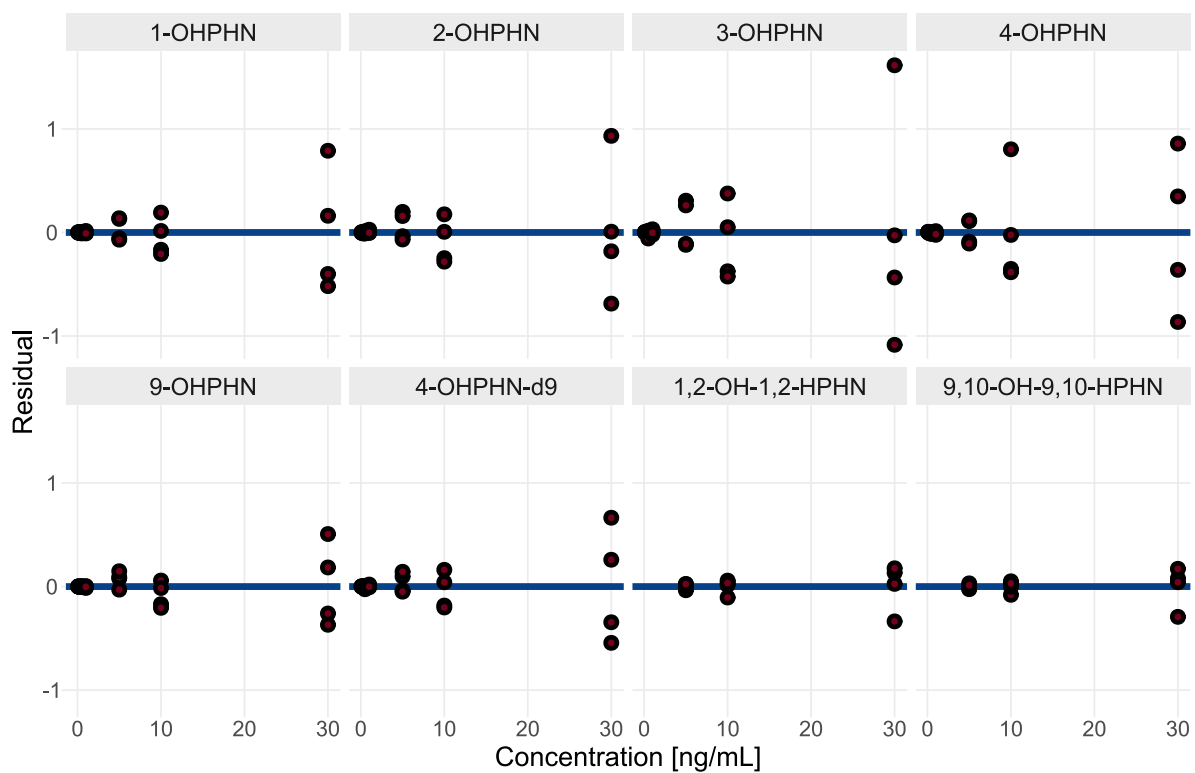
Figure B.2: Non-standardized residual plots for a FOM with weights, $w = 1/x$.
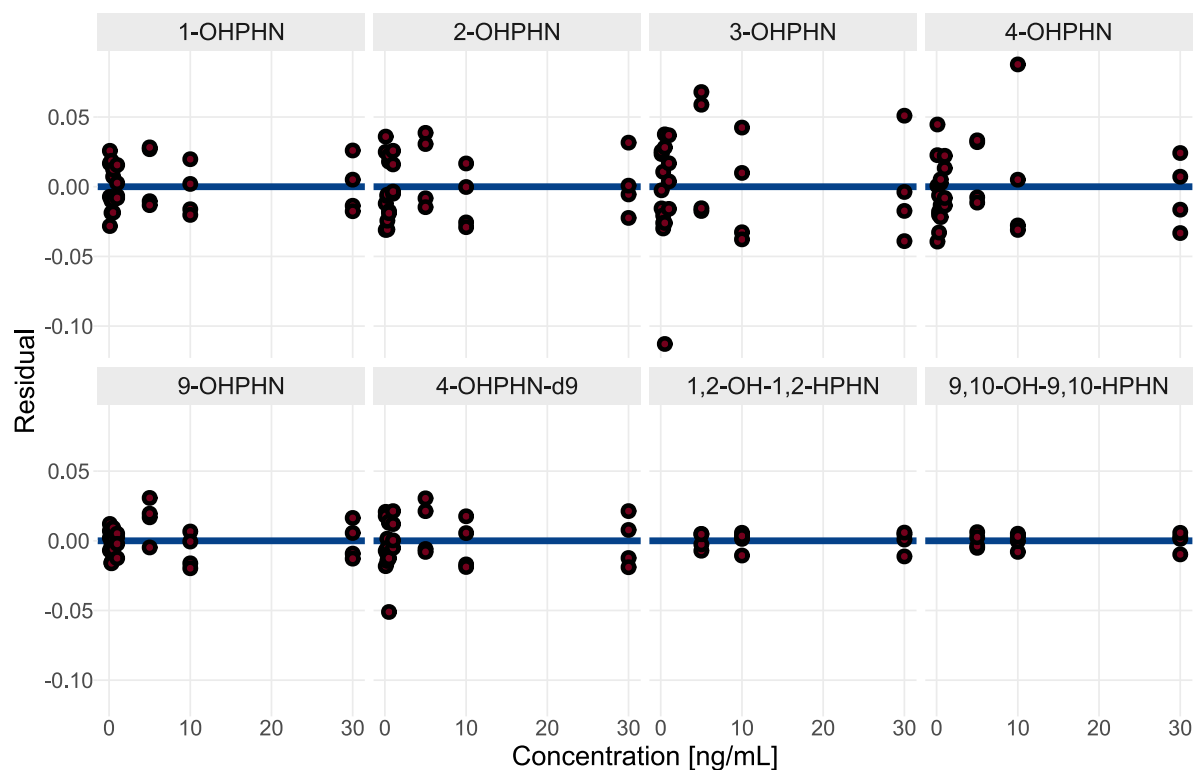
Figure B.3: Standardized residual plots for a FOM with weights, $w = 1/x^2$
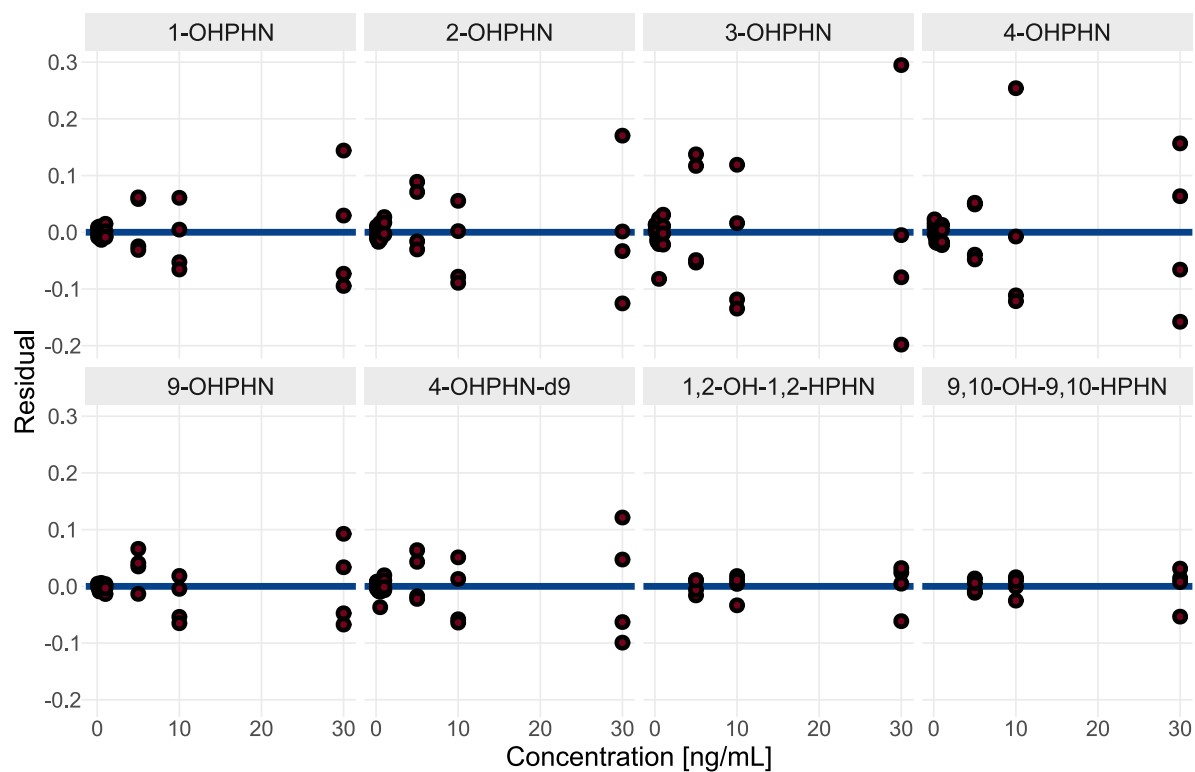
Figure B.4: Standardized residual plots for a FOM with weights, $w = 1/x$

Table B.1: *HorRat* values from APs

Compound	Concentration [ng/g ww]	*HorRat* of upper prediction error	*HorRat* of lower prediction error
1-OHPHN	5.73	-0.78	0.82
	22.9	-0.79	0.68
	115	-0.57	0.64
	229	-0.75	0.74
2-OHPHN	5.73	-0.80	0.74
	22.9	-0.57	0.77
	115	-0.82	0.77
	229	-1.16	1.02
3-OHPHN	5.73	-0.88	0.89
	22.9	-1.00	0.97
	115	-0.74	0.75
	229	-0.90	0.90
4-OHPHN	5.73	-1.02	0.97
	22.9	-0.41	0.44
	115	-0.51	0.61
	229	-0.70	0.57
9-OHPHN	22.9	-2.14	2.12
	115	-2.02	2.16
	229	-2.36	2.25

C   Factorial Designs

This appendix presents the experimental designs as well as the recorded area of 4-OHPHN. All designs are presented in their standard order.[15] In all runs designs, *Experiment 0* denotes the area of the initial method. For all runs, the sheath gas and temperatures were 300ºC and 12 L/min.

Table C.1: Initial 2_V^{5-1} FFD for screening

Experiment	GT	GF	NP	CV	NV	Area
0	250	12	25	3500	2000	658
14	250	9	20	3000	2000	735
8	350	9	20	3000	1000	578
10	250	13	20	3000	1000	535
11	350	13	20	3000	2000	614
C1	300	11	25	3500	1500	571
6	250	9	30	3000	1000	627
4	350	9	30	3000	2000	698
7	250	13	30	3000	2000	749
1	350	13	30	3000	1000	534
C2	300	11	25	3500	1500	580
13	250	9	20	4000	1000	537
2	350	9	20	4000	2000	539
16	250	13	20	4000	2000	575
15	350	13	20	4000	1000	396
C3	300	11	25	3500	1500	580
9	250	9	30	4000	2000	474
5	350	9	30	4000	1000	450
12	250	13	30	4000	1000	463
3	350	13	30	4000	2000	539
C4	300	11	25	3500	1500	600

---

[15]That is, the levels in first column is sorted -1, 1,..., -1, 1, the second column sorted -1, -1, 1, 1,..., -1, -1 , 1, 1, and so forth.

Table C.2: Follow-up 2_V^{5-1} FFD for screening

Experiment	GT	GF	NP	CV	NV	Area
0	250	12	25	3500	2000	585
6	250	9	20	1800	1000	975
10	350	9	20	1800	2000	628
13	250	13	20	1800	2000	613
11	350	13	20	1800	1000	784
9	250	9	30	1800	2000	636
15	350	9	30	1800	1000	683
14	250	13	30	1800	1000	794
1	350	13	30	1800	2000	683
C1	300	11	25	2300	1500	824
2	250	9	20	2800	2000	761
7	350	9	20	2800	1000	764
8	250	13	20	2800	1000	761
3	350	13	20	2800	2000	616
16	250	9	30	2800	1000	792
4	350	9	30	2800	2000	637
12	250	13	30	2800	2000	788
5	350	13	30	2800	1000	742
C2	300	11	25	2300	1500	834

Table C.3: CCD built from a 2_V^{5-1} FFD

Experiment	GT	GF	NP	CV	NV	Area
0	250	1211	25	3500	2000	302
14	225	11	22	2050	750	347
8	275	11	18	2050	750	356
10	225	11	18	2050	1250	318
11	225	7	22	2050	1250	396
6	225	7	22	1550	750	364
4	275	11	22	2050	1250	339
7	275	7	18	2050	1250	324
1	225	11	22	1550	1250	300
C1	250	9	20	1800	1000	361
13	275	7	18	1550	750	373
2	225	7	18	2050	750	371
16	275	7	22	2050	750	356
15	275	7	22	1550	1250	373
9	225	11	18	1550	750	304
5	275	11	18	1550	1250	280
12	275	11	22	1550	750	296
3	225	7	18	1550	1250	362
C2	250	9	20	1800	1000	367
24	200	9	20	1800	1000	388
18	300	9	20	1800	1000	360
23	250	5	20	1800	1000	396
25	250	13	20	1800	1000	329
19	250	9	16	1800	1000	342
C3	250	9	20	1800	1000	375
20	250	9	24	1800	1000	350
22	250	9	20	1300	1000	312
17	250	9	20	2300	1000	281
26	250	9	20	1800	500	352
21	250	9	20	1800	1500	352
C4	250	9	20	1800	1000	364

Table C.4: Optimization by the path of steepest ascent

Experiment	CV [V]	Area
0	3500	634
1	2900	637
2	2800	727
3	2700	792
4	2600	827
5	2500	843
6	2400	858
7	2300	865
8	2200	858
9	2100	852
10	2000	796

Experiments conducted at GT = 250^{oC}, GF = 13 mL/min, NP = 30 psi, NV = 2000 V

Table C.5: Optimization by the canonical path

Experiment	GT [^{oC}]	GF [mL/min]	Area
0	250	12	554
1	275	10	801
2	265	9.5	813
3	255	9	819
4	245	8.5	795
5	235	8	829
6	225	7.5	851
7	215	7	853
8	205	6.5	899
9	195	6	911
10	185	5.5	906
11	175	5	897

Experiments conducted at NP = 20 psi, CV = 1900 V, NV = 750 V

D   Raw Data from from Method Validation and Application Studies

Table D.1: Relative responses for the response functions

Concentration [ng/mL]	ISTD Response							
	1-OHPHN	2-OHPHN	3-OHPHN	4-OHPHN	9-OHPHN	4-OHPHN-d9	1,2-OH,12-HPHN	9,10-OH-9,10-HPHN
0.100	0.0036	0.0078	0.0080	0.0077	0.0047	0.0034	-	-
	0.0069	0.0022	0.0054	0.0055	0.0019	0.0024	-	-
	0.0060	0.0089	0.0082	0.0099	0.0008	0.0029	-	-
	0.0015	0.0041	0.0041	0.0015	0.0044	0.0015	-	-
0.300	0.0127	0.0238	0.0355	0.0207	0.015	0.0128	-	-
	0.0122	0.0241	0.0261	0.0165	0.013	0.0088	-	-
	0.0096	0.0166	0.0245	0.0127	0.0102	0.0089	-	-
	0.0208	0.0185	0.0233	0.0172	0.0116	0.0087	-	-
0.500	0.0298	0.0554	0.0731	0.0411	0.0330	0.0275	-	-
	0.0324	0.0576	0.0778	0.0424	0.0334	0.0296	-	-
	0.0213	0.0376	0.0460	0.0332	0.0203	0.0219	-	-
	0.0169	0.0369	0.0026	0.0289	0.0010	0.0257	-	-
1.00	0.0695	0.1239	0.1631	0.1055	0.0781	0.0565	-	-
	0.0565	0.1141	0.1430	0.0966	0.0688	0.0589	-	-
	0.0473	0.0931	0.1103	0.0698	0.0517	0.0413	-	-
	0.0458	0.0946	0.1300	0.0752	0.0571	0.0516	-	-
5.00	0.427	0.719	1.023	0.614	0.420	0.383	0.164	0.170
	0.433	0.679	0.977	0.621	0.466	0.396	0.187	0.177
	0.240	0.484	0.596	0.416	0.285	0.275	0.224	0.226
	0.226	0.452	0.606	0.397	0.274	0.453	0.223	0.208
10.0	0.828	1.262	1.853	1.852	0.849	0.707	0.240	0.293
	0.650	1.093	1.528	1.025	0.729	0.635	0.359	0.370
	0.469	0.839	1.102	0.697	0.504	0.479	0.403	0.423
	0.428	0.805	1.051	0.665	0.485	0.443	0.380	0.403
30.0	3.22	4.71	6.49	4.42	3.17	2.90	0.94	1.04
	2.03	3.60	4.84	3.20	2.16	2.13	1.41	1.41
	2.59	3.79	3.79	3.91	2.76	2.57	1.30	1.38
	1.91	3.09	4.44	2.69	1.96	2.02	1.45	1.51

Table D.2: Initial concentration estimates [ng/mL] from the accuracy study

Day	Replicate	Egg Count	Concentration [ng/mL]								
			Spike	1-OHPHN	2-OHPHN	3-OHPHN	4-OHPHN	9-OHPHN	4-OHPHN-d9	1,2-OH,12-HPHN	9,10-OH-9,10-HPHN
1	1	17	0.250	0.466	0.323	0.681	0.491	-	25.1	-	-
1	2	18	0.250	0.472	0.489	0.538	0.359	-	24.2	-	-
2	1	16	0.250	0.299	0.291	0.384	0.210	-	14.1	-	-
2	2	16	0.250	0.315	0.358	0.458	0.171	-	18.0	-	-
3	1	20	0.250	0.090	0.165	0.142	0.067	-	8.41	-	-
3	2	20	0.250	0.199	0.261	0.331	0.183	-	12.2	-	-
4	1	17	0.250	0.186	0.243	0.223	0.117	-	16.8	-	-
4	2	18	0.250	0.134	0.185	0.239	0.109	-	14.7	-	-
5	1	17	0.250	0.224	0.268	0.279	0.134	-	18.9	-	-
5	2	19	0.250	0.180	0.262	0.304	0.131	-	15.9	-	-
1	1	18	1.00	1.09	1.180	1.17	0.910	0.776	21.0	-	-
1	2	20	1.00	0.976	1.000	1.10	0.856	0.341	20.7	-	-
2	1	17	1.00	1.25	1.200	1.58	0.902	0.0728	18.0	-	-
2	2	19	1.00	0.806	0.924	1.21	0.601	0.0703	12.2	-	-
3	1	20	1.00	0.381	0.502	0.707	0.435	0.239	9.97	-	-
3	2	20	1.00	0.506	0.560	0.574	0.482	0.185	10.4	-	-
4	1	18	1.00	0.635	0.766	0.904	0.592	0.110	16.5	-	-
4	2	19	1.00	0.673	0.683	0.828	0.517	0.0983	16.8	-	-
5	1	18	1.00	0.649	0.637	0.846	0.565	0.0765	17.0	-	-
5	2	18	1.00	0.456	0.650	0.443	0.604	0.171	13.3	-	-
1	1	20	5.00	5.99	5.92	6.69	5.10	3.01	20.9	3.46	5.54
1	2	19	5.00	6.31	5.53	7.18	5.64	3.08	21.3	3.08	5.67
2	1	16	5.00	3.07	2.59	3.55	2.77	1.31	12.5	1.85	1.88
2	2	19	5.00	4.76	4.67	6.61	3.61	0.418	17.1	4.55	5.21
3	1	20	5.00	4.54	4.27	5.84	4.19	1.49	18.4	1.51	1.37
3	2	20	5.00	6.17	3.84	6.23	4.79	2.34	21.3	1.34	4.32
4	1	17	5.00	3.52	3.52	4.14	3.14	0.687	19.3	2.49	1.64
4	2	20	5.00	3.35	2.68	3.93	2.95	0.489	15.6	0.510	0.738
5	1	17	5.00	3.18	2.69	3.14	2.64	0.589	16.2	0.642	3.19
5	2	17	5.00	3.43	2.51	3.71	3.13	0.744	17.2	0.28	0.804
1	1	17	10.0	12.3	10.3	13.8	11.30	4.26	23.6	5.92	10.20
1	2	17	10.0	12.1	9.70	13.9	10.20	5.42	21.2	6.21	9.99
2	1	17	10.0	11.0	9.92	13.6	8.00	1.33	19.3	9.24	12.80
2	2	16	10.0	10.2	10.4	13.8	7.51	0.259	17.9	8.65	11.70
3	1	20	10.0	8.27	7.62	8.87	6.14	3.70	19.1	2.34	4.34
3	2	20	10.0	8.06	6.93	8.99	6.03	4.08	14.7	3.89	7.14
4	1	20	10.0	6.30	3.97	7.43	5.56	1.36	17.2	3.02	3.29
4	2	16	10.0	5.58	4.30	6.76	5.32	1.79	16.0	2.53	3.38
5	1	20	10.0	7.29	6.13	8.03	6.07	1.49	18.6	2.57	2.79
5	2	17	10.0	7.83	6.44	8.78	6.58	2.49	19.9	1.70	2.35

Table D.3: Initial concentration estimates [ng/mL] from the method application study

Study	Exposure	Level	Replicate	n_{egg}	Day	Concentration [ng/mL]					
						1-OHPHN	2-OHPHN	3-OHPHN	4-OHPHN	9-OHPHN	4-OHPHN-d9
PD18	PHN	S4	A	28	1	0.211	0.182	0.129	0.153	0.179	20.4
PD18	PHN	S2	A	11	1	0.237	0.194	0.173	0.405	0.286	21.9
PD18	PHN	S4	B	24	2	0.301	0.163	0.122	0.149	0.346	20.6
PD18	Control		B	20	2	0.114	0.128	0.120	0.243	0.198	14.7
PD18	PHN	S4	C	20	4	0.254	0.184	0.119	0.168	0.304	23.4
PD18	PHN	S3	C	19	4	0.0980	0.211	0.116	0.272	0.169	20.6
PD18	PHN	S1	C	20	4	0.644	0.406	0.334	0.720	0.641	24.1
PD18	Control		C	20	4	0.101	0.142	0.131	0.144	0.169	17.0
PD19	PHN	S1	A	10	1	0.234	0.195	0.172	0.145	0.289	21.7
PD19	PHN	S2	A	10	1	0.130	0.154	0.114	0.140	0.165	20.4
PD19	Si Control		A	10	1	0.0891	0.154	0.113	0.152	0.169	23.3
PD19	Control		A	10	1	0.0907	0.148	0.114	0.140	0.162	23.9
PD19	PHN	S1	C	10	3	1.680	0.640	1.01	0.802	0.884	108
PD19	PHN	S2	C	10	3	0.351	0.386	0.456	0.208	0.337	25.7
PD19	Si	Control	C	10	3	0.144	0.214	0.201	0.342	0.193	49.2
PD19	Control		C	10	3	0.141	0.257	0.288	0.180	0.203	44.7
PD19	PHN	S0	B	10	2	0.130	0.361	0.132	0.420	0.197	14.8
PD19	PHN	S2	B	10	2	0.185	0.167	0.114	0.143	0.163	12.5
PD19	Si Control		B	10	2	0.169	0.163	0.129	0.153	0.206	17.5
PD19	Control		B	10	2	0.112	0.146	0.115	0.143	0.182	18.7
WSF19	DCM Control		A	20	1	0.0936	0.162	0.113	0.195	0.163	15.3
WSF19	SI0	low	A	20	1	0.196	0.131	0.115	0.146	0.180	20.1
WSF19	SI0	med	A	20	1	0.212	0.174	0.114	0.162	0.268	25.8
WSF19	SI0	high	A	20	1	0.364	0.250	0.117	0.141	0.400	24.1
WSF19	WSF Comb.	low	A	20	1	0.215	0.180	0.164	0.146	0.166	21.1
WSF19	WSF Comb.	med	A	12	1	0.092	0.160	0.112	0.158	0.164	20.1
WSF19	WSF Comb.	high	A	10	1	0.197	0.175	0.114	0.145	0.255	22.5
WSF19	DCM Control		B	20	2	0.188	0.137	0.114	0.145	0.165	24.1
WSF19	SI0	low	B	20	2	0.096	0.134	0.115	0.148	0.165	22.2
WSF19	SI0	med	B	20	2	0.154	0.165	0.120	0.135	0.165	21.5
WSF19	SI0	high	B	20	2	0.124	0.220	0.117	0.196	0.192	24.2
WSF19	WSF	low	B	20	2	0.232	0.182	0.114	0.149	0.286	22.8
WSF19	WSF	med	B	20	2	0.116	0.199	0.114	0.347	0.176	19.2
WSF19	SI0	low	C	20	3	0.112	0.184	0.121	0.234	0.181	18.6
WSF19	SI0	med	C	17	3	0.402	0.290	0.187	0.196	0.272	46.8
WSF19	SI0	high	C	20	3	0.339	0.346	0.224	0.198	0.208	56.0
WSF19	WSF	low	C	20	3	0.640	0.413	0.260	0.287	0.307	43.7
WSF19	WSF	med	C	20	3	0.607	0.227	0.212	0.254	0.376	42.0
WSF19	WSF Comb.	low	C	18	3	0.392	0.357	0.214	0.209	0.192	53.4
WSF19	WSF Comb.	med	C	19	3	0.221	0.333	0.213	0.164	0.360	50.5
WSF19	WSF Comb.	high	C	10	3	0.616	0.316	0.204	0.324	0.217	46.9
WSF19	WSF Comb.	low	B	12	4	0.179	0.281	0.213	0.231	0.239	38.6
WSF19	WSF Comb.	med	B	19	4	0.099	0.170	0.113	0.136	0.178	18.5
WSF19	WSF Comb.	high	B	11	4	0.103	0.177	0.115	0.144	0.173	21.0
WSF19	WSF	med	A	20	4	0.422	0.377	0.221	0.214	0.275	56.6
WSF19	WSF	low	A	20	4	0.094	0.145	0.116	0.197	0.166	21.9
WSF19	WSF	high	A	20	4	0.408	0.185	0.164	0.175	0.377	15.4
QC	-	-	-	20	1	4.39	3.82	9.03	4.97	1.06	24.1
QC	-	-	-	20	2	5.31	4.25	6.21	3.65	0.893	22.6
QC	-	-	-	20	3	4.57	3.68	5.85	3.44	1.84	23.2
QC	-	-	-	20	4	4.84	3.51	5.16	2.47	2.10	21.8

Table D.4: Relative responses for the LoD study

Concentration [ng/g ww]	Egg Count	ISTD Response			
		1-OHPHN	2-OHPHN	3-OHPHN	4-OHPHN
0.92	17	0.042	0.044	0.042	0.036
		0.039	0.042	0.043	0.037
		0.043	0.044	0.039	0.040
		0.046	0.029	0.043	0.039
1.2	20	0.056	0.062	0.045	0.044
		0.061	0.067	0.045	0.066
		0.048	0.056	0.057	0.074
		0.063	0.054	0.068	0.056
1.3	16	0.142	0.045	0.037	0.030
		0.063	0.056	0.057	0.062
		0.075	0.108	0.065	0.037
		0.037	0.037	0.049	0.035
1.8	18	0.069	0.075	0.080	0.083
		0.11	0.092	0.071	0.076
		0.061	0.094	0.071	0.067
		0.073	0.075	0.086	0.061
2.1	19	0.106	0.079	0.092	0.088
		0.080	0.079	0.105	0.11
		0.090	0.090	0.086	0.091
		0.083	0.086	0.082	0.091

E   R-Scripts

E.1   Accuracy Study

```r
require(ggplot2) #Used for Plotting
require(car) #Used for Levene's Test
require(readxl) #Reads in Excel Files
require(dplyr) #Data manipulation
source('AnalyticalMethodValidation.r') #Custom validation package
source('MyTheme.r') #Loads in pre-set plotting themes


#Load in sample data
sample.data <- data.frame(read_xlsx('x2.xlsx', col_names = TRUE)) %>%
   filter(Type %in% 'Sample')%>%  filter(!(Compound %in% c('COOHs'))) %>%
   arrange(Exp..Conc., Compound)




#Remove the blank samples and the dihydrodiols from the linearity evaluation
sample.data <- arrange(sample.data, Compound) %>% filter(!(Type == 'Blank')) %>%
   filter(!(Compound %in% c('9,10-OH-9,10-HPHN', '1,2-OH-1,2-HPHN')))

#Convert spike conc. (Exp..Conc.) to estimated conc (Final.Conc) to ng/g ww units
sample.data$Exp..Conc.<- c(0.25,0.25,1,1,5,5,10,10)/(20*0.00218)
sample.data$Final.Conc. <- as.numeric(sample.data$Final.Conc.)/(20*0.00218)


#Surrogate corrections
FinalCor <-20/as.numeric(sample.data$Final.Conc.[sample.data$Compound == '4-OHPHN-d9'])
sample.data <- filter(sample.data, !(Compound %in% '4-OHPHN-d9'))
sample.data$Final.Conc. <- as.numeric(sample.data$Final.Conc.)*
   as.numeric(FinalCor)/(20*0.00218)




lvls <- unique(sample.data$Exp..Conc.) #Stores the unique concentration levels

I <- 5 #Sample preparation done over five days with two replicates per day
J <- 2

lvl.len <- length(unique(sample.data$Exp..Conc.))

sample.data$Days <- rep(1:I, each = rep(lvl.len*J))

#Remove lowest concentration level for 9-OHPHN as it did not possess a response
sample.data <- sample.data  %>%
   filter(!(Compound %in% c('9-OHPHN') & Exp..Conc. %in% c(0.25)))


#Separate betweeen bias corrected and non-corrected concentrations
non.cor.sample.data <- sample.data
non.cor.sample.data <- (non.cor.sample.data %>% arrange(Compound, Exp..Conc.)) %>%
   filter(!(Compound == '9-OHPHN' & Exp..Conc. == sample.data$Exp..Conc.[1]))


conc <- c(rep(c(0.25,1,5,10), length(unique(sample.data$Compound))-1),
```

```
            c(1,5,10))

comp <- rep(unique(sample.data$Compound),each = 4)[
  (-length(unique(sample.data$Compound))*4)]

#Linearity
#############################################################################

#Obtain linearity parameters
b0 <- c()
nu <- c()
b0u <- c()
b1 <- c()
b1u <- c()
r2 <- c()
w <- c()
recovery <- c()

res <- c()
for (cmp in unique(sample.data$Compound)){

  temp.dat <- non.cor.sample.data[non.cor.sample.data$Compound == cmp,]

  #Finds the variance at each level
  var.dat <-    aggregate(temp.dat$Final.Conc., by = list(temp.dat$Exp..Conc.), FUN = var)[,2]

  Weights <- 1/rep(var.dat, each = I*J)

  temp.dat$Weights <- Weights

  w <- c(w, Weights)

#Fits the linearity graph to using each of the three weights (1/X, 1/X^2 and 1/S^2)
m1 <- lm(Final.Conc. ~ Exp..Conc.,data=temp.dat, weights = 1/Exp..Conc.^1)
m2 <- lm(Final.Conc. ~ Exp..Conc.,data=temp.dat, weights = 1/Exp..Conc.^2)
m3 <- lm(Final.Conc. ~ Exp..Conc.,data=temp.dat, weights = Weights)

  #Collects slope and intercept coefficients and their standard errors
  b0 <- c(b0, summary(m1)$coefficients[1,1],
          summary(m2)$coefficients[1,1],summary(m3)$coefficients[1,1])
  b0u <- c(b0u, summary(m1)$coefficients[1,2],
           summary(m2)$coefficients[1,2],summary(m3)$coefficients[1,2])

  b1 <- c(b1, summary(m1)$coefficients[2,1],
          summary(m2)$coefficients[2,1],summary(m3)$coefficients[2,1])
  b1u <- c(b1u, summary(m1)$coefficients[2,2],
           summary(m2)$coefficients[2,2],summary(m3)$coefficients[2,2])

  #Performs Levene's Test
  print(leveneTest(Final.Conc. ~ as.factor(Exp..Conc.),data = temp.dat))

  res <- c(res, weighted.residuals(m1)) #Collected weighted residuals of m1
  #m1 above is replaced with m2 or m3 for other models


#Calculates recovery
recovery <- c(recovery,aggregate(x=(100*temp.dat$Final.Conc./temp.dat$Exp..Conc.),
                                  by = list(temp.dat$Exp..Conc.),FUN = mean)[,2])
}
```

```
#Perform t-tests on the slope and intercept
#the df of 9-OHPHN must manually be changed to 28 instead of 38 due to one less level
b0.test <- 2*pt(q=(b0-0)/b0u, df=38, lower.tail = FALSE)
b1.test <- 2*pt(q=abs((b1-1)/b1u), df=38, lower.tail = FALSE)


#Linearity residual plot
gg.res <- ggplot(data = data.frame(res = res, x = non.cor.sample.data$Exp..Conc.,
                                   Compound = non.cor.sample.data$Compound),
                 aes(x=x, y=res))+
  ylab('Residual')+xlab('Concentration [ng/g ww]')+my_theme()+theme(axis.text.y = NULL)+
  geom_hline(yintercept = 0, colour =  "#07438B", lwd = 2)+
  geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
facet_wrap(~Compound,ncol = 1)

#Linearity plot
non.cor.sample.data$Weights <- w
ggplot(data = non.cor.sample.data, aes(x = Exp..Conc., y = Final.Conc.))+
  stat_smooth(method = 'lm',se = FALSE, colour = "#07438B",
              lwd = 2, aes(weight = Weights))+
  geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
  facet_wrap(~Compound, scales = 'free_y', nrow = 2)+my_theme()+
  xlab('Expected Concentration [ng/g ww]')+
  ylab('Estimated Concentration [ng/g ww]')+theme(axis.text.y = NULL)



#Plots regression coefficients
my.lab <- as_labeller(c(b_0 = 'b[0]', b_1 = 'b[1]'), default = label_parsed)

 reg.df2 <- data.frame(b = c(b0,b1), u = c(b0u,b1u),
                        compound = rep(unique(sample.data$Compound),each = 3),
                        Weight = c('1/X', '1/X^2', '1/S^2'),
                        Coefficient=rep(c('b_0','b_1'), each = length(b0) ))

ggplot(data = reg.df2, aes(x = compound, y = b, fill = Weight))+
geom_errorbar( aes( ymin=b, ymax=b+u),#width=1,
                colour="black", size=1,alpha=0.9, position =  position_dodge())+

geom_col(position = ('dodge'), color = 'black', lwd = 0.8)+
  facet_wrap(~Coefficient, scales = 'free',
             labeller = my.lab)+my_theme()+
  scale_fill_manual(labels = c('1/X' = expression('1/X ' ),
                               '1/X^2' = expression('1/X'^2),
                               '1/s^2'= expression('1/S'^2)),
                    values = c('#07438B', '#73001d','#177F57')) +
  xlab('Compound')+ ylab('Coefficient Magnitude')+

  theme(axis.text.y = NULL)+guides(x = guide_axis(angle = 30))



#Collects recoveries into a data frame
recovery.df <- data.frame(Recovery = recovery, Lvl = conc ,Compound = comp)

ucf <- c((1/b1)*(b1u/b1)) #Find the uncertainty of the proportional bias CF

#Collects all corrections
corrections <- data.frame(cp = 1/b1, cb = -b0, up = ucf, ub = b0u,
                          Compound =rep(unique(sample.data$Compound),each = 3),
```

```
                              Weight = c('1/X', '1/X^2', '1/s^2')) %>%
                                  filter((Weight %in% '1/s^2'))




#Plot recoveries
ggplot(data = recovery.df, aes(x=Lvl/(20*0.00218), y = Recovery))+
  geom_hline(yintercept = 100, lwd = 1, linetype = 'dashed')+
  geom_line(aes(colour = Compound), lwd = 1)+xlab('Concentration [ng/g ww]')+
  geom_point(aes(colour = Compound ), size = 3)+
  ylab('Recovery [%]')+  scale_fill_manual(values =
  c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  my_theme()+theme(axis.text.y = NULL)




#Trueness
###############################################################################

#Collects corrected recoveries by applying the CFs
recovery.c <- c()
means.c <- c()
means.uc <- c()

corrected.concs <- c()

for (cmp in unique(sample.data$Compound)){

  temp.dat <- non.cor.sample.data[non.cor.sample.data$Compound == cmp,]
  means.uc <- c(means.uc,aggregate(x=(temp.dat$Final.Conc.), by = list(temp.dat$Exp..Conc.),
                              FUN = mean)[,2])


  temp.dat$Final.Conc. <- (temp.dat$Final.Conc. +
     corrections[corrections$Compound == cmp,]$cb)*corrections[
        corrections$Compound == cmp,]$cp
  corrected.concs <- c(corrected.concs, temp.dat$Final.Conc.)
  recovery.c <- c(recovery.c,aggregate(x=(100*temp.dat$Final.Conc./temp.dat$Exp..Conc.),
                              by = list(temp.dat$Exp..Conc.),
                           FUN = mean)[,2])


  means.c <- c(means.c,aggregate(x=(temp.dat$Final.Conc.), by = list(temp.dat$Exp..Conc.),
                              FUN = mean)[,2])
}
recovery.c.df <- data.frame(Recovery = recovery.c,
                              Lvl = conc,
                              Means = means.c,
                              Compound = comp)

#Plots corrected recoveries
ggplot(data = recovery.c.df, aes(x=Lvl/(20*0.00218), y = Recovery))+
  #geom_line(aes(colour = Compound), lwd = 1)+
  xlab('Concentration [ng/mL]')+
  geom_hline(yintercept = 100, lwd = 1, linetype = 'dashed')+
  geom_point(aes(colour = Compound ), size = 3)+
  ylab('Recovery [%]')+  scale_fill_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
```

```
   my_theme()+theme(axis.text.y = NULL)




#Precision
###############################################################################
#Collects precision components for corrected and non-corrected data
r <- c()
btw <- c()
IP <- c()
#nu <- c()

r.n <- c()
btw.n <- c()
IP.n <- c()

for (cmp in unique(sample.data$Compound)){
  temp.dat <- sample.data[sample.data$Compound == cmp,]
  temp.dat$Final.Conc.cor <- (temp.dat$Final.Conc. + corrections[
    corrections$Compound == cmp,]$cb)*
    corrections[corrections$Compound == cmp,]$cp

  for (lvl in unique(temp.dat$Exp..Conc.)){
    temp.dat.lvl <- temp.dat[temp.dat$Exp..Conc. == lvl,]
    precision.data <- PrecisionANOVA(Days='Days',Response = 'Final.Conc.cor',
                                     I = 5, J = 2, data = temp.dat.lvl)

    precision.data.noncor <- PrecisionANOVA(Days='Days',Response = 'Final.Conc.',
                                     I = 5, J = 2, data = temp.dat.lvl)

    r.n <- c(r.n, unlist(precision.data.noncor[2]))
    btw.n <- c(btw.n, unlist(precision.data.noncor[4]))
    IP.n <- c(IP.n, unlist(precision.data.noncor[6]))

    r <- c(r, unlist(precision.data[2]))
    btw <- c(btw, unlist(precision.data[4]))
    IP <- c(IP, unlist(precision.data[6]))
  }
}



precision.df <- data.frame(r = r, IP = IP, btw = btw,# nu = nu,
                           lvl = rep(unique(sample.data$Exp..Conc.),5),
                           r.n = r.n, btw.n = btw.n, IP.n = IP.n,
                           Compound = compound)

precision.df <- precision.df[-(length(precision.df$r)-3),]



#Removes lowest concentration level of 9-OHPHN
precision.df$cb <-rep((corrections$cb),each = 4)[
  (-length(unique(sample.data$Compound))*4)]
precision.df$cp <-    rep((corrections$cp),each = 4)[
  (-length(unique(sample.data$Compound))*4)]
precision.df$ub <-    rep((corrections$ub),each = 4)[
  (-length(unique(sample.data$Compound))*4)]
precision.df$up <-    rep((corrections$up),each = 4)[
  (-length(unique(sample.data$Compound))*4)]
precision.df$means.cor <- means.c
precision.df$means.uncor <- means.uc
```

```
#Find RSDs
precision.df$RSDr <- 100*precision.df$r/precision.df$means.cor
precision.df$RSDIP <- 100*precision.df$IP/precision.df$means.cor
precision.df$RSDr.n <- 100*precision.df$r.n/precision.df$means.uncor
precision.df$RSDIP.n <- 100*precision.df$IP.n/precision.df$means.uncor



precision.df2 <- data.frame(RSD   = c(precision.df$RSDr, precision.df$RSDIP),
                            RSD.n = c(precision.df$RSDr.n, precision.df$RSDIP.n),
                            lvl = precision.df$lvl,
                            means.cor <- c(precision.df$means.cor, precision.df$means.cor),
                            Compound = precision.df$Compound,
                            Type = rep(c('Repeatability', 'Intermediate Precision'),
                                       each = length(precision.df$IP)))
#Plot precision data
ggplot(data=precision.df, aes(x = lvl, y = RSDIP))+
  geom_line(aes(colour = Compound), lwd = 1)+xlab('Concentration [ng/g]')+
  geom_point(aes(colour = Compound), size = 3)+
  ylab('Recovery [%]')+  scale_fill_manual(values =
                         c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  my_theme()+theme(axis.text.y = NULL)



ggplot(data=precision.df, aes(x = lvl, y = RSDr))+
  geom_line(aes(colour = Compound), lwd = 1)+xlab('Concentration [ng/mL]')+
  geom_point(aes(colour = Compound), size = 3)+
  ylab('Recovery [%]')+  scale_fill_manual(values =
                         c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  my_theme()+theme(axis.text.y = NULL)




ggplot(data=precision.df2, aes(x = lvl, y = RSD))+
  geom_line(aes(colour = Compound), lwd = 1)+xlab('Concentration [ng/g ww]')+
  geom_point(aes(colour = Compound), size = 3)+
  ylab('RSD [%]')+  scale_fill_manual(values =
                    c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  my_theme()+theme(axis.text.y = NULL)+facet_wrap(~Type, scales= 'free')




#Accuracy Profiles and Standard MU estimates
############################################################################

#Calculate the HorRat
horratIP <- precision.df$RSDIP/Horwitz(precision.df$lvl*(20*0.00218)/(10^9))
horratr <- precision.df$RSDr/((2/3)*Horwitz(precision.df$lvl*(20*0.00218)/(10^9)))



uAP <-c()
nuAP <- c()
```

```r
muAP <- c()

nu <- c()
mu <- c()
RSDmu <- c()


for (cmp in unique(precision.df$Compound)){
  temp.dat <- precision.df[precision.df$Compound == cmp,]
  for (lvl in unique(temp.dat$lvl)){

    #This loop calculates the combined uncertainty and the effective degrees
    #of freedom
    temp.dat.lvl <- temp.dat[temp.dat$lvl == lvl,]


    u <- c(temp.dat.lvl$r.n, temp.dat.lvl$btw.n, temp.dat.lvl$ub, temp.dat.lvl$up)
    c <- c(temp.dat.lvl$cp, temp.dat.lvl$cp, temp.dat.lvl$cp,
          temp.dat.lvl$means.uncor + temp.dat.lvl$cb)

    df <- c(I*(J-1), I -1, (I*J*length(temp.dat$lvl))-2, (I*J*length(temp.dat$lvl))-2)
    nu <- c(nu, unlist(WelchSatter(ui = u, ci =c ,df = df))[1])
    urss <- sqrt(sum((c*c*u*u)))

    mu <- c(mu, urss)
    RSDmu <- c(RSDmu, 100*urss/temp.dat.lvl$means.cor)


    #Estimates to be usdd in accuracy proiles
    ##########################################################################
    uAP <- c(temp.dat.lvl$r.n/(I*J), temp.dat.lvl$btw.n/I, temp.dat.lvl$ub, temp.dat.lvl$up)
    urssAP <- sqrt(sum((c*c*uAP*uAP)))
    muAP <- c(muAP, urssAP)
    nuAP <- c(nuAP, unlist(WelchSatter(ui = uAP, ci =c ,df = df))[1])
  }
}


#Constructing the accuraacy profile
AP.df <- data.frame(xbar = precision.df$means.cor, uxbar = muAP,
                    ux = mu,
                    neff = mu/muAP, nu = nuAP, Level = precision.df$lvl,
                    Compound = precision.df$Compound)

AP.df$RSD.ux <- 100*AP.df$ux/AP.df$xbar

AP.df$Lower <- (100*(AP.df$xbar-AP.df$Level)/AP.df$Level) -
  100*qt(p = 0.975, df = AP.df$nu)*
          AP.df$ux*sqrt(1+(1/AP.df$neff))/AP.df$xbar

AP.df$Upper <- (100*(AP.df$xbar-AP.df$Level)/AP.df$Level) +
  100*qt(p = 0.975, df = AP.df$nu)*
          AP.df$ux*sqrt(1+(1/AP.df$neff))/AP.df$xbar

AP.df2 <- data.frame(Value= c(AP.df$Lower, AP.df$Upper),
                     Level = AP.df$Level, Type = rep(c('Lower', 'Upper'),
                           each = length(AP.df$xbar)),
                     Compound = AP.df$Compound)
```

```r
#Create a function for plotting the Theoretical Horwitz AP
HorPred <- function(c, I, J, neg = FALSE,k = 1){
  c <- c*20*0.00218

  R <- (2^(1-0.5*log10(c/10^9)))/100 #Estimated reproducibility
  r <- R*2/3  #Estimated repeatability

  R.sq <- (R*c)^2
  r.sq <- (r*c)^2

  b.sq <- R.sq-r.sq

  u.xbar <- sqrt((b.sq/I) +(r.sq/(I*J)))
  neff <- ((R*c)^2)/(u.xbar^2)


  interval <- 1.96*100*R*sqrt(1+(1/neff))*k

  if (neg == TRUE){interval <- -interval}

  return(interval)
  }


#Plotting the final uncertainty components
ggplot(data = AP.df2, aes(x = Level, y = Value))+
  geom_line(aes(colour = Type), lwd = 1)+xlab('Concentration [ng/g ww]')+
  geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
  ylab('Error [%]')+
  stat_function(fun = HorPred, args = list(I=5, J=2, neg = TRUE), lwd = 1,
                aes(color = 'Horwitz AP'), linetype = 'dashed') +
  stat_function(fun = HorPred, args = list(I=5, J=2, neg = FALSE), lwd  =1,
                aes(color = 'Horwitz AP'), linetype = 'dashed')+

  stat_function(fun = HorPred, args = list(I=5, J=2, neg = TRUE,k=2), lwd  =1,
                aes(color ='Horwitz 2*AP'), linetype = 'dashed')+
  stat_function(fun = HorPred, args = list(I=5, J=2, neg = FALSE,k=2), lwd  =1,
                aes(color ='Horwitz 2*AP'), linetype = 'dashed')+
  scale_colour_manual(values =
      c('#8467B7','#177F57','#07438B','#07438B','#07438B','#07438B'))+
  my_theme()+theme(axis.text.y = NULL, legend.position = 'bottom')+ facet_wrap(~Compound)


ggplot(data=AP.df, aes(x = Level, y = RSD.ux))+
  geom_line(aes(colour = Compound), lwd = 1)+xlab('Concentration [ng/mL]')+
  geom_point(aes(colour = Compound ), size = 3)+
  ylab('RSD [%]')+  scale_fill_manual(values =
      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values =
                      c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  my_theme()+theme(axis.text.y = NULL)
```

E.2   Calibration Study

```r
require(ggplot2)
require(car)
require(readxl)
require(dplyr)
source('AnalyticalMethodValidation.r')
source('MyTheme.r')

#Reads in calibration data for statistical evaluation.
my.dat <- data.frame(read_xlsx('270123x2.xlsx'))
cal.dat <- arrange(my.dat)

#Sorter fixes the order of the compounds when using facet_wrap
cal.dat$sorter <- (rep(c(8,7,0,2,3,5,1,6,4),
                    each = length(cal.dat$Type)/length(unique(cal.dat$Compound))))
cal.dat <- cal.dat %>% filter(Type == 'Cal') %>% filter(!(Compound %in% c('COOHs')))
cal.dat$Exp..Conc. <- as.numeric(cal.dat$Exp..Conc.)
cal.dat$ISTD.Resp..Ratio <- as.numeric(cal.dat$ISTD.Resp..Ratio)




#Removes concentration levels which are undetected
cal.dat <- cal.dat %>% arrange(Exp..Conc., Compound) %>% filter(!(Exp..Conc. %in% c(0.1)))#
cal.dat <- rbind(filter(cal.dat, !(Compound %in% c('1,2-OH-1,2-HPHN', '9,10-OH-9,10-HPHN'))),
                 filter(cal.dat, !(Exp..Conc. %in% c(0.1,0.2,0.3,0.5,1)) &
                        (Compound %in% c('1,2-OH-1,2-HPHN', '9,10-OH-9,10-HPHN'))))
cal.dat$Exp..Conc.[cal.dat$Exp..Conc. == 0.2] <- 0.1 #Updates an error in the excel file

#Plots the calibration data
ggplot(data = cal.dat, aes(x = Exp..Conc., y =   ISTD.Resp..Ratio))+
  facet_wrap(~Compound, scales = 'free')+
  geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
  my_theme()+theme(axis.text.y = NULL)+xlab('Concentration [ng/ml]') +
  ylab('IS Resp.Ratio.')+
  facet_wrap(~sorter ,nrow = 2, scales ='free_y',
            labeller = as_labeller( c('1' = '1-OHPHN','2' = '2-OHPHN','3'= '3-OHPHN',
            '4'= '4-OHPHN','5' = '9-OHPHN','6' = '4-OHPHN-d9','7' = '1,2-OH-1,2-HPHN',
            '8' = '9,10-OH-9,10-HPHN')))

#Back calculation and residuals
back.calc <- c()
all.w <- c()
all.r <- c()
cal.dat <- arrange(cal.dat, Compound)
for (cmp in unique(cal.dat$Compound)){
  print(cmp)
  #Define the variances
  v <- rep(aggregate(x = cal.dat[cal.dat$Compound == cmp,]$ISTD.Resp..Ratio ,
                    by = list(as.factor(cal.dat[cal.dat$Compound == cmp,]$Exp..Conc.)),
                    FUN = var)[,2], each = 4)

  #The three different weights are provided below. Uncomment to run the weight
  #of interest

  w <- 1/(cal.dat[cal.dat$Compound == cmp,]$Exp..Conc.)
  #w <- 1/(cal.dat[cal.dat$Compound == cmp,]$Exp..Conc.)^2
  #w<- 1/v


  temp.dat <- cal.dat[cal.dat$Compound == cmp,]
```

```r
temp.dat$Weights <- w

#Perform Levene's test
print(leveneTest(y = cal.dat[cal.dat$Compound == cmp,]$ISTD.Resp..Ratio,
                 group= as.factor(cal.dat[cal.dat$Compound == cmp,]$Exp..Conc.),
                 center = 'mean'))
m1 <- lm(ISTD.Resp..Ratio ~  Exp..Conc. + I(Exp..Conc.^2),
         weights = w,data = cal.dat[cal.dat$Compound == cmp,])
print(summary(m1))

#Perform Mandel's test
print(MandelLoF(data=temp.dat,y = 'ISTD.Resp..Ratio', x = 'Exp..Conc.',
                weights = 'Weights',m1.deg = 1,m2.deg = 2))


all.w <- c(all.w, w)
all.r <- c(all.r, weighted.residuals(m1))

for(resp in cal.dat[cal.dat$Compound == cmp,]$ISTD.Resp..Ratio){
  back.calc <- c(back.calc,   InvPrediction(b0 = m1$coefficients[1],
                                            b1= m1$coefficients[2],
                                            b2 = m1$coefficients[3],
                                            y = as.numeric(resp))[1])}

for(resp in sample.data[sample.data$Compound == cmp,]$ISTD.Resp..Ratio){

  samples.calc <- c(samples.calc,   InvPrediction(b0 = m1$coefficients[1],
                                                  b1=m1$coefficients[2],
                                                  b2 = m1$coefficients[3],
                                                  y = as.numeric(resp))[1])}

}
back.calc <-as.numeric(replace(replace(back.calc, is.na(back.calc), 0),
                               back.calc == 'No real roots',0))




cal.dat$BackCalculation <- 100*(back.calc - cal.dat$Exp..Conc.)/(cal.dat$Exp..Conc.)
cal.dat$Residual <- all.r
cal.dat$Weights <- all.w

#Plots data
ggplot(data = cal.dat, aes(x = (as.numeric((Exp..Conc.))), y =Residual))+
  geom_point()+facet_wrap(~Compound,scales = 'free_y')+my_theme()+
  xlab('Concentration [ng/mL]') + ylab('Residual')+theme(axis.text.y = NULL,
                                      plot.title = element_text(size = 12))+
  geom_hline(yintercept = 0, colour =  "#07438B", lwd = 2)+
  geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
  facet_wrap(~sorter ,nrow = 2 ,#scales ='free_y',
             labeller = as_labeller( c('1' = '1-OHPHN','2' = '2-OHPHN','3'= '3-OHPHN',
             '4'= '4-OHPHN','5' = '9-OHPHN','6' = '4-OHPHN-d9','7' = '1,2-OH-1,2-HPHN',
             '8' = '9,10-OH-9,10-HPHN')))



ggplot(data = cal.dat, aes(x = Exp..Conc., y = ISTD.Resp..Ratio))+
  facet_wrap(~Compound, nrow = 2, scales = 'free_y')+
  geom_smooth(method = 'lm', aes(weight = Weights),formula = y ~ x + I(x^2), lwd = 2,
              se = FALSE, colour = '#07438B')+
```

```
geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
my_theme()+theme(axis.text.y = NULL)+
xlab('Concentration [ng/mL]') + ylab('IS Resp.Ratio.')


ggplot(data=cal.dat, aes(x=log(Exp..Conc.), y = BackCalculation))+
geom_point()+facet_wrap(~Compound,scales = 'free_y', nrow = 2)+my_theme() +
ylab('Percentage Error (%)')+theme(axis.text.y = NULL)+
geom_hline(yintercept =
            0, colour =  '#07438B', lwd = 2)+
geom_point(color = "black",fill = '#73001d' , size = 2, stroke = 2, shape = 21)+
ggtitle((bquote('w = 1/S'^2)))+xlab('Log Concentration')
```

E.3 LoD Study

```r
require(ggplot2)
require(car)
require(readxl)
require(dplyr)
source('AnalyticalMethodValidation.r')
source('MyTheme.r')


exclude <- c('9,10-OH-9,10-HPHN','1,2-OH-1,2-HPHN', 'COOHs', '9-OHPHN', '4-OHPHN-d9')

my.lod <-data.frame(read_xlsx('LoDs.xlsx', col_names  = TRUE))
my.lod$Exp..Conc. <- as.numeric(my.lod$Exp..Conc)
my.lod$Final.Conc. <- as.numeric(my.lod$Final.Conc)
my.lod$Resp. <- as.numeric(my.lod$Resp.)
my.lod$Rel.Resp. <- as.numeric(my.lod$ISTD.Resp..Ratio)


my.lod$Exp..Conc. <- rep(c(0.03,0.04,0.05,0.06,0.08,0.09),4)


my.lod <- my.lod %>% arrange(as.numeric(Exp..Conc.)) %>% arrange(Compound, Name) %>%
  filter(Type == 'Cal')
my.lod <- my.lod %>% filter(Exp..Conc. != '0.03') %>% filter(!(Compound %in% '9-OHPHN'))


res <- c()

for (cmp in unique(samples$Compound)){
  temp.dat <- samples[samples$Compound == cmp,]
  res.t <- weighted.residuals(lm(as.numeric(Rel.Resp.) ~Exp..Conc., weights = Weight,
                                 data = temp.dat[temp.dat$Compound == cmp,]))
  res <- c(res, res.t)
  print(cmp)
  qqnorm(res.t)
  print(LoDCalc(temp.dat[temp.dat$Compound == cmp,]$Exp..Conc.,
                temp.dat[temp.dat$Compound == cmp,]$Rel.Resp.))
  }
```

E.4 Sample Application

```r
require(readxl)
require(dplyr)
require(ggplot2)
require(forcats)
require(propagate)
source('MyTheme.R')


compound <- c('1-OHPHN','2-OHPHN','3-OHPHN','4-OHPHN', '9-OHPHN')

sample.dat <- data.frame(read_xlsx('Samples180323.xlsx'))
QC.dat <- data.frame(read_xlsx('QC.xlsx'))
QC.dat$Concentration <-( QC.dat$Concentration*20/QC.dat[
  QC.dat$Compound == '4-OHPHN-d9',]$Concentration)/(20*0.00218)

#Only investigate monohydroxys
QC.dat <- QC.dat[QC.dat$Compound %in% compound,]


#Surrogate correction
surrogate <- 20/as.numeric(sample.dat$Uncorrected[sample.dat$Compound == '4-OHPHN-d9'])
sample.dat$Sur.Cor <- as.numeric(sample.dat$Uncorrected)*surrogate


sample.dat <- sample.dat %>% arrange(Compound) %>%
  filter(Compound %in% compound) %>% filter(!(Exposure %in% c('QC')))


#Defining Correcetion Factors
##############################################################################
b1 <- c(0.941,
        0.826,
        1.09,
        0.8,
        0.29)


ub1 <- c(0.038,
         0.043,
         0.055,
         0.029,
         0.047)

cp <- 1/b1
up <- cp* sqrt(ub1/b1)^2


cb <- -c(1.232,
         3.314,
         3.331,
         0.659,
         -0.424
)


ub <- c(0.679,
        0.590,
        0.874,
        0.652,
        2.058
```

```
)
precision.n.df <- data.frame(Compound  =
                            rep(c('1-OHPHN','2-OHPHN','3-OHPHN','4-OHPHN', '9-OHPHN')
, each = 4)[(-20)],

Level = rep(c(1,2,3,4), 5)[(-20)],

r.n =    c(15.28951,
           8.483115,
           8.277881,
           8.503384,
           14.949962,
           10.783032,
           13.541355,
           8.512334,
           19.101748,
           13.587494,
           12.963642,
           10.129968,
           28.818541,
           10.295399,
           6.065493,
           7.372994,
           54.034539,
           35.91353,
           31.213395

),



IP.n = c(30.95303,
         27.13528,
         18.54565,
         20.68079,
         20.72817,
         22.98249,
         24.16622,
         29.35427,
         27.25967,
         33.83195,
         22.85462,
         24.7326,
         41.67108,
         16.33926,
         17.1891,
         17.70005,
         81.63241,
         65.40917,
         64.5149),
Levels = c(rep(4, 16), 3,3,3)
)

#Storing compound names
comp <- c()

#Create a data frame with correction factors
val.trueness <- data.frame(Cor.p = cp, Cor.b = cb, Cor.ub = ub, Cor.up = up,
                           Compound = compound)
```

```r
QC.dat <- arrange(QC.dat, Compound)
QC.dat$CorConc <- (QC.dat$Concentration+rep(val.trueness$Cor.b, each = 4))*
  rep(val.trueness$Cor.p, each = 4)


nu <- c()
U <- c()
for (result in QC.dat$CorConc){

temp.dat <- QC.dat[QC.dat$CorConc== result ,]
cmp <- temp.dat$Compound
J <- 2
I <- 5

means.uncor <- temp.dat$Concentration

#Spike was also performed at level 3
lvl.len <- precision.n.df[
  precision.n.df$Compound == cmp & precision.n.df$Level == 3,]$Levels
r <- means.uncor*precision.n.df[
  precision.n.df$Compound == cmp & precision.n.df$Level == 3,]$r.n/100
IP <- means.uncor*precision.n.df[
  precision.n.df$Compound == cmp & precision.n.df$Level == 3,]$IP.n/100
print(r)
btw <- sqrt(IP^2- r^2)


up <- val.trueness[val.trueness$Compound == cmp,]$Cor.up
ub <- val.trueness[val.trueness$Compound == cmp,]$Cor.ub

cp <- val.trueness[val.trueness$Compound == cmp,]$Cor.p
cb <- val.trueness[val.trueness$Compound == cmp,]$Cor.b

#Define uncertainty components
u <- c(r, btw, ub, up)
#Define sensitivity coefficients
c <- c(cp, cp, cp, means.uncor + cb)

#Define degrees of freedom for uncertainty estimates
df <- c(I*(J-1), I -1, I*J*lvl.len -2, I*J*lvl.len -2)
#Find the effective degrees of freedom for the combined uncertainty
nu <- c(nu, unlist(WelchSatter(ui = u,ci =c ,df = df))[1])

#Calculate the standard combined uncertainty
U <- c(U, sqrt(sum(c*c*u*u)))
comp <- c(comp,cmp)
}

#Create the expanded  uncertainty
U <- c(U*qt(p = 0.975,df = nu))

qc <- data.frame(CI = U, Compound = comp, Conc = QC.dat$CorConc, nu = nu)

#Plot the data
ggplot(data = data.frame(Concentration = QC.dat$CorConc,
                         Uncertainty = U,
                         Compound = QC.dat$Compound,
                         Spike = c(1,2,3,4)),
       aes(x = Spike,  y = Concentration))+facet_wrap(~Compound)+
```

```
      geom_errorbar(aes(ymin=Concentration−Uncertainty,
                        ymax=Concentration+Uncertainty), width=.2,
              lwd = 1.25,color = '#07438B'  ,position=position_dodge(.9)) +
   geom_hline(aes(linetype ='Spike',
                yintercept = 115), color = '#177F57', lwd = 1.25)+
   geom_point(color = "black",fill = "#73001d" , size = 2, stroke = 2, shape = 21)+
   ylab('Concentration [ng/g ww]')+xlab('Extraction Day')+my_theme()+
   theme(axis.text.y = NULL)+
   scale_linetype_manual(name = "Spike", values = "dashed", labels = "")


cor.n <- length(sample.dat$Uncorrected)/length(unique(compound))

sample.dat$EggConc <- sample.dat$Sur.Cor/(sample.dat$Eggs*0.00218)

sample.dat$EggConc <- (sample.dat$EggConc+rep(val.trueness$Cor.b,
                       each = cor.n))*rep(val.trueness$Cor.p, each = cor.n)
sample.dat$Level <- fct_relevel(sample.dat$Level, 'low', 'med', 'high')


max.lod  <- 0.12/(20*0.00218)
min.lod <- 0.05/(20*0.00218)

loq1 <- 0.25/(20*0.00218)
loq2 <- 1/(20*0.00218)

sample.dat <- sample.dat[sample.dat$Treatment != 'DCM Control',]


ggplot(data = sample.dat[sample.dat$Exposure == 'PD',],
       aes(x = Compound, y = EggConc,fill = Level, color = Level, order = Level))+
  geom_bar(stat = 'identity', color = 'black', lwd = 0.8, position = 'dodge')+
  facet_wrap(~Treatment,nrow = 2,ncol = 1, strip.position = 'top')+
  my_theme()+theme(axis.text.y =  NULL)+
  geom_rect(aes(ymin = min.lod, ymax = max.lod, xmin = 0.4, xmax = 5.5,
              color = NULL ),fill = "#2C77BF",
           alpha = 0.02)+
  scale_fill_manual(values =c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  scale_colour_manual(values = c('#07438B','#73001d','#177F57','#8467B7','#BBA033'))+
  geom_segment(aes(x= 0.4, xend = 5.5, y = loq1, yend = loq1), color = 'black',
             lwd = 1, linetype = 'dashed')+
  geom_segment(aes(x= 0.4, xend = 5.5, y = loq2, yend = loq2), color = 'black',
             lwd = 1, linetype = 'dashed')+
  scale_x_discrete(expand=c(−1.25, 6.2))+ylab('Concentration [ng/g ww]')+
  geom_text(aes(x =5.86 , y = max.lod, label = 'LoD Range'),size = 5,
           color = 'black',check_overlap = TRUE)+
  geom_text(aes(x =5.8 , y = loq1, label = 'LoQ Low'),size = 5,
           color = 'black',check_overlap = TRUE)+
    geom_text(aes(x =5.8 , y = loq2, label = 'LoQ High'),size = 5,
             color = 'black', check_overlap = TRUE)


ggplot(data = sample.dat[sample.dat$Exposure == 'WSF',],
      aes(x = Compound, y = EggConc,fill = Level, color = Level,
          order = Level))+geom_bar(stat = 'identity', color = 'black',
                                  lwd = 0.8, position = 'dodge')+
```

```r
facet_wrap(~Treatment, nrow = 4, ncol = 1, strip.position = 'top')+
my_theme()+theme(axis.text.y =   NULL)+
geom_rect(aes(ymin = min.lod, ymax = max.lod, xmin = 0.4, xmax = 5.5,
              color = NULL ), fill = "#2C77BF",
         alpha = 0.02)+scale_fill_manual(values = c('#07438B','#73001d','#177F57'))+
scale_colour_manual(values =
                    c('#07438B','#73001d','#177F57'))+
geom_segment(aes(x= 0.4, xend = 5.5, y = loq1, yend = loq1), color = 'black',
             lwd = 1, linetype = 'dashed')+
geom_segment(aes(x= 0.4, xend = 5.5, y = loq2, yend = loq2), color = 'black',
             lwd = 1, linetype = 'dashed')+
scale_x_discrete(expand=c(-1.25, 6.2))+
geom_text(aes(x =5.86 , y = max.lod, label = 'LoD Range'),size = 4,
          color = 'black',check_overlap = TRUE)+
geom_text(aes(x =5.8 , y = loq1, label = 'LoQ Low'),size = 4,
          color = 'black',check_overlap = TRUE)+
geom_text(aes(x =5.8 , y = loq2, label = 'LoQ High'),size = 4,
          color = 'black', check_overlap = TRUE)+
ylab('Concentration [ng/g ww]')
```

E.5 Optimization

```r
require(rsm)
require(readxl)
require(ggplot2)
require(dplyr)
source('MyTheme.r')

#Labels for first-order models with interactions
tw.labs <- c('Intercept', 'GT', 'GF', 'NP', 'CV','NV','GT:GF','GT:NP',
             'GT:CV','GT:NV', 'GF:NP','GF:CV','GF:NV','NP:CV','NP:NV','CV:NV')

#Read data

df <- data.frame(read_xlsx('XlFacDatV2.xlsx'))
area.initial <- df$Area[df$Experiment == 'Initial Method']
#Remove Blank Method it is only relevant for comparisons sake, not anything else
df <- df[df$Experiment != 'Initial Method',]
#Plot initial areas
#####
ggplot(data = df, aes(x = reorder(Experiment, seq(1:length(Area))), y = Area))+
  geom_bar(position = 'dodge',stat = 'identity', aes(fill = Type),
           color = 'black', lwd = 0.8)+
  geom_hline(lwd = 1.2,aes(yintercept = area.initial,
             linetype = 'Initial Method'), colour = '#177F57')+
  theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=1))+
  scale_fill_manual(name='',labels = c('Center Point',
                                'Experiment' = 'Factorial Experiment',
                                'Initial Method'),
                 values = c('#07438B', '#73001d','#177F57'))+my_theme()+
  theme(axis.text.y = NULL)+xlab('Experiment')+scale_linetype_manual(
    values = c('Initial Method' = 'dashed'), name = '', labels = c('Initial Method')
  )

#Seemingly, experiments 4, 7 and 14 provide areas greater than  the initial method
#This can be aassessed via a t-test using center points as an estimate for noise

diff.1 <- data.frame(Experiment = c('E4', 'E7', 'E14'),
  t.stat = c(698-area.initial,749-area.initial,
             735-area.initial)/sd(df[df$Type == 'Center Point','Area']),
  p.val = 1-pt(c(698-area.initial,749-area.initial,
                 735-area.initial)/sd(df[df$Type == 'Center Point','Area'])
          ,length(df[df$Type == 'Center Point','Area'])-1)
)


#Create coded data frame
df.c <- coded.data(df, x1 ~ (GT-300)/50, x2 ~ (GF-11)/2, x3 ~ (NP-25)/5,
                   x4 ~ (CV-3500)/500, x5 ~ (NV-1500)/500)

#Try first order model
m.f <- rsm(Area ~ FO(x1,x2,x3,x4,x5), data = df.c)
summary(m.f)
#Plot Q-Q Plot of coefficients (except intercept)
qqnorm(m.f$coefficients[2:length(m.f$coefficients)])

#First order model with two-way interactions
m.ft <- rsm(Area ~FO(x1,x2,x3,x4,x5) + TWI(x1,x2,x3,x4,x5), data = df.c)
summary(m.ft)


ggplot(
```

```r
    data = data.frame(Label = tw.labs, Coefficients = m.ft$coefficients,
        Quantiles = qqnorm(m.ft$coefficients)$x)[-1,],#Collect data required for qqplot
        aes(x = Quantiles, y = Coefficients, label = Label))+
        geom_point(color = "black",fill = "#73001d" , size = 3, stroke = 2.5,
                    shape = 21)+
        geom_text(aes(label=ifelse(abs(Coefficients)>25,as.character(Label),'')),
                    #Only label the two largest coefficients
            color = '#07438B', hjust = -.5, vjust = 0, size = 5)+
    my_theme()+xlab('Theoretical Quantiles')+ylab('Estimated Coefficients') x
#Only label the two largest coefficients

#Note that many coefficients are statistically significant
#However, merely two coefficients are able to "break away" from the "noise line"
#these are nozzle voltage and capillary voltages, however, as nozzle is already
#maxed out, we may only change the capillary. The negative coefficient
#suggests that we need to reduce this voltage. We follow the path of
#steepest ascent:

sa <- data.frame(read_xlsx('SteepestAscent.xlsx', col_names = TRUE))


ggplot(data = sa[sa$Experiment != 'Initial Method',], aes(x = as.factor(seq(1:10)),
                                                            y = Area))+
    geom_bar(stat = 'identity', fill = '#73001d', color = 'black',
            lwd = 0.8)+geom_hline(lwd = 1.2,aes(yintercept = area.initial,
                        linetype = 'Initial Method'), colour = '#177F57')+
    my_theme()+ theme(axis.text.y = NULL)+xlab('Step')+scale_linetype_manual(
        values = c('Initial Method' = 'dashed'), name = '', labels = c('Initial Method')
    )

#Follow-up experiments:
############################################################################
df.2 <- data.frame(read_xlsx('FacI3v2.xlsx'))
area.initial.2 <- df.2$Area[df.2$Experiment == 'Initial Method']
#Remove Blank Method it is only relevant for comparisons sake, not anything else
df.2 <- df.2[df.2$Experiment != 'Initial Method',]


ggplot(data = df.2, aes(x = reorder(Experiment, seq(1:length(Area))), y = Area))+
    geom_bar(position = 'dodge',stat = 'identity', aes(fill = Type), color = 'black',
            lwd = 0.8)+
    geom_hline(lwd = 1.2,aes(yintercept = area.initial.2,
                            linetype = 'Initial Method'), colour = '#177F57')+
    theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=1))+
    scale_fill_manual(name='',labels = c('Center Point',
                                        'Experiment' = 'Factorial Experiment',
                                        'Initial Method'),
                    values = c('#07438B', '#73001d','#177F57'))+my_theme()+
    theme(axis.text.y = NULL)+xlab('Experiment')+scale_linetype_manual(
        values = c('Initial Method' = 'dashed'), name = '', labels = c('Initial Method')
    )

#Create coded data frame
df.c.2 <- coded.data(df.2, x1 ~ (GT-300)/50, x2 ~ (GF-11)/2, x3 ~ (NP-25)/5,
                    x4 ~ (CP-2300)/500, x5 ~ (NV-1500)/500)

#Try first order model
m.f.2 <- rsm(Area ~ FO(x1,x2,x3,x4,x5), data = df.c.2)
summary(m.f.2)
```

```
m. ft .2 <- rsm( Area ~FO(x1,x2,x3,x4,x5) + TWI(x1,x2,x3,x4,x5), data = df.c.2)
summary(m. ft .2)


ggplot(
   data = data.frame(Label = m. ft .2$newlabs, Coefficients = m. ft .2$coefficients,
                     Quantiles = qqnorm(m. ft .2$coefficients)$x)[-1,],
   #Collect data required for qqplot
   aes(x = Quantiles, y = Coefficients, label = Label))+
   geom_point(color = "black",fill = "#73001d" , size = 3, stroke = 2.5, shape = 21)+
   geom_text(aes(label=ifelse(abs(Coefficients)>25,as.character(Label),'')),
             color = '#07438B', hjust = -.5, vjust = 0, size = 5)+
   my_theme()+xlab('Theoretical Quantiles')+ylab('Estimated Coefficients')


####################################
df.3 <- data.frame(read_xlsx('FinalExperiments.xlsx'))
area. initial .3 <- df.3$Area[df.3$Experiment == 'Initial Method']
#Remove Blank Method it is only relevant for comparisons sake, not anything else
df.3 <- df.3[df.3$Experiment != 'Initial Method',]


ggplot(data = df.3, aes(x = reorder(Experiment, seq(1:length(Area))), y = Area))+
   geom_bar(position = 'dodge',stat = 'identity', aes(fill = Type), color = 'black',
            lwd = 0.8)+
   theme(axis.text.x = element_text(angle = 0, vjust = 0.5, hjust=1))+
   scale_fill_manual(name='',labels = c('Center Point',
                                         'Experiment' = 'Factorial Experiment',
                                         'Initial Method'),
                     values = c('#07438B', '#73001d','#177F57'))+my_theme()+
   theme(axis.text.y = NULL)+xlab('Experiment')+scale_linetype_manual(
      values = c('Initial Method' = 'dashed'), name = '', labels = c('Initial Method'))
   )


#Create coded data frame
df.c.3 <- coded.data(df.3, x1 ~ (GT-250)/25, x2 ~ (GF-9)/2, x3 ~ (NP-20)/2,
                     x4 ~ (CV-1800)/250, x5 ~ (NV-1000)/250)


#Try first order model
m. f .3 <- rsm(Area ~ FO(x1,x2,x3,x4,x5), subset = (Block == '1') ,data = df.c.3)
summary(m. f .3)


m. ft .3 <- rsm(Area ~FO(x1,x2,x3,x4,x5) + TWI(x1,x2,x3,x4,x5),
               subset = (Block == '1'),data = df.c.3)
summary(m. ft .3)


m.so.3 <- rsm(Area ~ SO(x1,x2,x3,x4,x5), data = df.c.3)
summary.rsm(m.so.3, threshold = 1)


ggplot(
   data = data.frame(Label = m.so.3$newlabs, Coefficients = m.so.3$coefficients,
                     Quantiles = qqnorm(m.so.3$coefficients)$x)[-1,],
   aes(x = Quantiles, y = Coefficients, label = Label))+
   geom_point(color = "black",fill = "#73001d" , size = 3, stroke = 2.5, shape = 21)+
   geom_text(aes(label=ifelse(abs(Coefficients)> 10 ,as.character(Label),'')),
             color = '#07438B', hjust = -.5, vjust = 0, size = 5)+
   my_theme()+xlab('Theoretical Quantiles')+ylab('Estimated Coefficients')
```

```
contour(m.so.3, ~x1+x2+x3+x4+x5, image = TRUE)
canonical(m.so.3, threshold = 1)
points(20,1000, pch = 18, cex =1.75)
persp(m.so.3,~x1+x2+x3+x4+x5, contour = 'colors')
canonical.path(m.so.3, threshold = 1)


#Perform new steepest ascent, reduce GT by increments of 10 and GF by 0.5
#For all other components the center point is (ish) at the peak
#Perhaps also increase NP with 0.3


sa2 <- data.frame(read_xlsx('SteepestAscent2.xlsx', col_names = TRUE))
area.initial.5 <- sa2$Area[sa2$Experiment == 'Initial Method']

ggplot(data = sa2[sa2$Experiment != 'Initial Method',],
       aes(x = as.factor(1:length(Area)), y = Area))+
  geom_bar(stat = 'identity',
              fill = '#73001d', color = 'black', lwd = 0.8)+
  geom_hline(lwd = 1.2,aes(yintercept = area.initial.5,
                             linetype = 'Initial Method'), colour = '#177F57')+
  my_theme()+
  theme(axis.text.y = NULL)+xlab('Experiment')+scale_linetype_manual(
     values = c('Initial Method' = 'dashed'), name = '', labels = c('Initial Method'))
```

E.6 Method Validation Package

```r
require(ggplot2)
require(propagate)

#Mandels Test
################################################################################
MandelLoF <- function(x, y, weights = NULL, m1.deg = 1, m2.deg = 2, data){
  x <- data[[x]] #Redefine x to denote the x values of the data frame, df
  y <- data[[y]] #same for y...

  #Find sample size (which equals the length of any arbitrary column of df)
  n <- length(x)

  #Provides equal weight to all points, i.e. normal OLS, if no weights are provided
  if (is.null(weights) == TRUE){w <- rep(c(1), n)}
  else {w <- data[[weights]]} #Assigns weights to w, if weights are given

  #Defines regression models
  lm.m1 <- anova(lm(y ~ poly(x, degree = m1.deg, raw = TRUE), weights = w))
  lm.m2 <- anova(lm(y ~ poly(x, degree = m2.deg, raw = TRUE), weights = w))


  #degree of polynomial plus 1a denotes number of parameters
  m1.df <- n - (m1.deg + 1)
  m2.df <- n - (m2.deg + 1)

  m1.sse <- lm.m1$'Sum Sq'[2] #Extracts residual sums of squares
  m2.sse <- lm.m2$'Sum Sq'[2]

  F.statistic <- ((m1.sse - m2.sse)/(m1.df - m2.df))/(m2.sse/m2.df)

  #Converts F-quantile into a p-value
  p.val <- pf(F.statistic, df1 = (m1.df-m2.df), df2 = m2.df, lower.tail = FALSE)


  #Lines below prints out relevant information

  cat('\nF =', F.statistic, 'with num df =', (m1.df-m2.df), 'and denom df =',
      m2.df, 'degrees of freedom\n')
  cat('p-value =', p.val,'\n\n')

  invisible(list(F.stat = F.statistic, p.value = p.val))
}
################################################################################

#Defines weights which sum to 1
################################################################################
#Returns weights of compounds, sorted in accordance with the levels
VarWeight <- function(y, level, data){
  y <- data[[y]] #Redefine y to denote the x values of the data frame, df
  lvl <- data[[level]] #.'..and levels

  #Find sample size (which equals the length of any arbitrary column of df)
  n <- length(y)

  weights <- c()

  for (l in unique(lvl)){
    replicates <- sum(lvl == l) #Number of replicates at the particular level
```

```
        sr <- 1/var(y[lvl == l]) #Extract data individually for each level

        weights <- c(weights, rep(sr, replicates))
    }

    return(weights/(sum(weights)/n))
}
###############################################################################

#Youden Plot
###############################################################################
Youden <- function(estimates, replicates, day, TrueValue = NULL, data){
    est <- data[[estimates]]

    x.rep <- unique(data[[replicates]])[1]
    y.rep <- unique(data[[replicates]])[2]

    day <- as.factor(data[[day]][data[[replicates]] == x.rep])




    x <- est[data[[replicates]] == x.rep]
    y <- est[data[[replicates]] == y.rep]

    x.max <- max(x)
    y.max <- max(y)

    df <- data.frame(x,y,day, TrueValue)

    return(ggplot(data = df, aes(x = x, y = y))+geom_point(aes(color = day))+
        geom_vline(xintercept = TrueValue)+geom_hline(yintercept = TrueValue)+
        geom_abline(slope = 1, intercept = 0)+xlab('Parallel 1')+ylab('Parallel 2')+
            xlim(c(-x.max,x.max))+ylim(c(-y.max,y.max)))

}
###############################################################################

#ANOVA-based Precision estimates
###############################################################################
#Defining a function which enables the calculation of accuracy components
PrecisionANOVA <- function(Days, Response, I, J, data){

    x.dat <- data[[Days]] #Define the days
    y.dat <- data[[Response]] #Define the response

    #The repeatability variance is the average within-day variance among the I days

    r.ev <- sum(aggregate(x = y.dat, by = list(x.dat), var)$x)/I

    #x.ev is the standard deviation of the grand mean

    x.ev <- var(aggregate(x = y.dat, by = list(x.dat), mean)$x)

    #It can be shown that the between-day variance is the difference between
    #the standard deviation of the means and the standard variance of the repeatability
    #see the thesis

    b.ev <- x.ev - r.ev/J
```

```
    r.es <- sqrt(r.ev) #Repeatability estimate

    if (b.ev < 0){ #If between-day variance is less than zero, is assumed to be absent
      b.ev <- 0
      b.es <- 0

      R.ev <- r.ev
      R.es <- sqrt(R.ev)

#Due to the means of estimation, the degrees of freedom becomes I-1,
#rather than IJ-1, see GUM
      nu.es <- I-1

    }


    else{#If between-day variance is positive
      b.es <- sqrt(b.ev)
      R.ev <- r.ev+b.ev
      R.es <- sqrt(R.ev) #Intermediate precision estimate


      #Intermediate variance estimate
      nu.es <- unlist(WelchSatter(ui=c(sqrt(r.ev), sqrt(b.ev)), df=c(I*(J-1), I-1))[1])

    }



    results <- list(rep.ev=c(r.ev), rep.es=c(r.es), btw.ev=c(b.ev),
                        btw.es=c(b.es), IP.ev=c(R.ev), IP.es=c(R.es),
                        Nu.eff=c(nu.es)) #Provide all results

    return(results)

}
################################################################################

AccuracyProfile <- function(Days, Response, I, J, TrueMean,gamma, data,
                            customcoverage = FALSE){

  aov.dat <- PrecisionANOVA(Days = Days, Response = Response, I = I, J = J, data = data)

  mean.response <- mean(data[[Response]])

  sr <- aov.dat[2] #Extracts repeatability
  sb <- aov.dat[4] #Extracts between-day variance
  nu <- aov.dat[7] #Effective degrees of freedom

  v.ip <- (sr^2)+(sb^2)
  s.ip <- sqrt(v.ip)
  v.xbar <- (sr*sr/(I*J))+(sb*sb/I)

  bias <- mean(Response) - unique(data[[TrueMean]])

  neff <- v.ip/v.xbar

  if (customcoverage == FALSE){Q <- qt(p = (1+gamma)/2, df = nu)}
  if (customcoverage != FALSE){Q <- customcoverage}
```

```
  uct <- s.ip*(1+(1/neff))^0.5

  return(Q*c(-uct,uct)+bias)

}
################################################################################
#Inverse predicition - Calculate concentration from first or second degree polynomial
################################################################################
InvPrediction <- function(b0,b1,b2=0,y=0, negatives = FALSE){
  b0 <- b0 - y #Update coefficient after solving for roots

  roots <- polyroot(z = c(b0,b1,b2)) #Find the roots of the polynomial

  real.roots <- c()

  x1.real <- round(Im(roots[1]), 8) == 0 #Check if complex component is negligble
  x2.real <- round(Im(roots[2]), 8) == 0

  #If there are no real roots, end process
  if(x1.real == FALSE & x2.real == FALSE){final <- 'No real roots'}

  else{ #If there's atleast one real root, proceed on...
    if(x1.real == TRUE){real.roots <- c(real.roots,roots[1])}
    if(x2.real == TRUE){real.roots <- c(real.roots,roots[2])}

    #Only use real non-negative components}
    if (negatives == FALSE){final <- (Re(real.roots)[Re(real.roots) > 0])}
    else{final <- Re(real.roots)} #Return all roots
  }
  return(final)}



################################################################################
#Limit of detetion calculations
################################################################################
LoDCalc <- function(x, y, gamma = 0.95, weighting = FALSE){
  if (weighting == FALSE){w <- rep(c(1), length(x))}
  if (weighting == TRUE){w <- WeightCalc(Values = y, Factor = x)}

  m1 <- summary(lm(y~x, weights = w))

  df <- m1$df[2]
  Q <- qt(df = df,p = gamma)

  res <- m1$sigma

  simple.lod <- (2*Q)*res/(m1$coefficients)[2]
  n <- length(x)
  xbar <- mean(x)
  ssx <- var(x)*(n-1)

  PI <- (1+(1/n)+(xbar*xbar/ssx))^0.5

  PI.lod <- simple.lod*PI

  return(c(simple.lod, PI.lod))
}

################################################################################
#Horwitz Equation
```

```
###############################################################################
Horwitz <- function(c, r = FALSE){

  RSD <- 2^(1-(0.5*log10(c)))

  if (r == TRUE){RSD <- RSD*0.5}
  return(RSD)
  }

###############################################################################
```

E.7   MassHunter Quantification Software to R Package

```r
require(dplyr)
require(writexl)
#Compounds to be used
#######################################################################
compounds <- c("9,10-OH-9,10-HPHN", "1,2-OH-1,2-HPHN","COOHs",
               "2-OHPHN","3-OHPHN", "9-OHPHN",
               "1-OHPHN","4-OHPHN-d9" ,"4-OHPHN")


#Data cleaning
#######################################################################
data <- data.frame(read.csv('MassHunter.csv', header = F)) #load the data
#%>% c(data[1,] == "BPA-d16 (ISTD) Results" )

col.names <- unlist(gsub('Method', '', data[1,])) #Cleaning up row names
col.names <- unlist(gsub('Results', '', col.names)) #to allow for easier
col.names <- unlist(gsub(' ', '', col.names)) #data treatment later on

#locate the column for type of sample
type.index <- match('Type',c(unlist(filter(data[2,]))))
name.index <- match('Name',c(unlist(filter(data[2,]))))




for (i in 1:(length(col.names) -1 )){
  #iterates through the row names to give them all a name
  if (col.names[i+1] == ""){col.names[i+1] <- col.names[i]}

}
data[1,] <- col.names
sample.name <- data[1, type.index]

###############################################
colnames(data) <- NULL
rownames(data) <- NULL

#Using the cleaned data, the dataframe can be made more compact
df.dat <- data.frame()



for (compound in compounds){
  dat.temp <- data[,unlist(data[1,]) == compound]
  dat.temp$Compound <- compound
  dat.temp[2,][length(dat.temp[2,])] <- "Compound"
  colnames(dat.temp) <- dat.temp[2,]
  dat.temp <- dat.temp[-c(1,2),]
  df.dat <- rbind(df.dat, dat.temp)
}

df.dat$Type <- data[,type.index][3:length(data[,type.index])]
df.dat$Name <- data[,name.index][3:length(data[,name.index])]


write_xlsx(df.dat,"Excel.xlsx")
```