



UNIVERSITÀ DI PARMA

UNIVERSITÀ DEGLI STUDI DI PARMA

Dipartimento di Scienze Chimiche, della Vita e della Sostenibilità Ambientale

Dottorato di Ricerca in Scienza e Tecnologia dei Materiali

Ciclo XXXV

Innovative materials for miniaturized sample preparation techniques, environmental and food packaging applications

Coordinatore:

Prof. Enrico Dalcanale

Tutore:

Prof.ssa Federica Bianchi

Dottorando: Fabio Fornari

Anni Accademici: 2019/2020 – 2021/2022

This page was intentionally left blank.

Innovative materials for miniaturized sample preparation techniques, environmental and food packaging applications

PhD Thesis

by

Fabio Fornari

Dipartimento di Scienze Chimiche, della Vita e della Sostenibilità Ambientale

Università degli Studi di Parma

Title

Innovative materials for miniaturized sample preparation techniques, environmental and food packaging applications

Submission date

January 29th, 2023

Final revisions

March 21th, 2023

Defense date

May 19th, 2023

Supervisor

Associate Professor Federica Bianchi

Department of Chemistry, Life Sciences and Environmental Sustainability,
University of Parma, Italy

Opponents

Full Professor Elisabetta Venuti

Department of Industrial Chemistry "Toso Montanari", University of Bologna, Italy

Associate Professor Alessandro Airoidi

Department of Aerospace Science and Technology, Polytechnic University of Milan,
Italy

Associate Professor Elisa Robotti

Department of Science, Technology and Innovation, University of Eastern Piedmont
"Amedeo Avogadro", Italy

Fonts: Cambria for text, Arial for pictures

PhD Thesis 2023 © Fabio Fornari

Table of Contents

Table of Contents.....	i
Preface	xi
List of publications	xiii
Chapter 1 Introduction.....	1
1.1. General context.....	1
1.2. Outline.....	3
References	6
Chapter 2 Methodologies.....	9
2.1. Chemometric background.....	9
2.1.1. Pattern recognition.....	11
2.1.1.1. Unsupervised Pattern Recognition.....	11
2.1.1.2. Supervised Pattern Recognition	13
2.1.2. Design of Experiments	26
2.1.2.1. Screening Designs.....	30
2.1.2.2. Response Surface Designs	34
2.1.2.3. Multicriteria decision making.....	43
2.2. Validation of analytical methods.....	46
2.2.1. Limits of Detection and Limits of Quantitation	47
2.2.2. Linearity, significance, and validity.....	48
2.2.3. Precision.....	50
2.2.4. Trueness.....	52
2.2.5. Enrichment capability	54
Note of the author	54

References.....	54
Chapter 3 Evaluating the potential of carbon nanotubes for the solid-phase microextraction of personal care products from water samples.....	59
3.1. Introduction	59
3.1.1. Emerging contaminants.....	59
3.1.2. Personal care products.....	60
3.1.3. Solid-phase microextraction.....	63
3.1.4. Carbon nanotubes.....	65
3.2. Materials and methods.....	68
3.2.1. Chemicals and materials.....	68
3.2.2. Preparation and characterization of the fibers	68
3.2.2.1. Fiber fabrication	68
3.2.2.2. Fiber characterization.....	69
3.2.3. Optimization of the extraction procedure	69
3.2.3.1. Preliminary evaluation of the coating performance	69
3.2.3.2. Optimization.....	70
3.2.4. Operating procedure and instrumental conditions.....	70
3.2.4.1. Sample preparation.....	70
3.2.4.2. GC-MS.....	71
3.2.5. Method validation.....	72
3.2.6. Analysis of real samples.....	73
3.2.7. Software.....	73
3.3. Results and discussion	74
3.3.1. Characterization of the CNTs-based coatings	74
3.3.2. Optimization of the SPME procedure	76

3.3.2.1. Preliminary evaluation of the coating performance.....	76
3.3.2.2. Optimization.....	77
3.3.3. Method validation.....	80
3.3.4. Analysis of real samples	82
3.4. Conclusions.....	83
Note of the author	85
References	85

Chapter 4 | Development of a magnetic Metal-Organic Framework composite for the dispersive micro solid-phase extraction of polycyclic aromatic hydrocarbons from water samples

4.1. Introduction	91
4.1.1. Polycyclic aromatic hydrocarbons	91
4.1.2. Magnetic dispersive micro solid-phase extraction.....	93
4.1.3. Metal-Organic Frameworks.....	94
4.1.3.1. MOFs as sorbents for sample preparation.....	95
4.1.3.2. Fabrication of magnetic MOF composites for MD- μ SPE	95
4.1.3.3. PUM198	97
4.2. Materials and methods	98
4.2.1. Chemicals.....	98
4.2.2. Preparation and characterization of the sorbent.....	99
4.2.2.1. Synthetic procedures	99
4.2.2.2. Characterization	100
4.2.3. Optimization of the extraction procedure	100
4.2.3.1. Preliminary evaluation of the elution solvent.....	101
4.2.3.2. Screening for the important factors.....	101
4.2.3.3. Optimization of the critical factors.....	101

4.2.4. Operating procedure and instrumental conditions.....	102
4.2.4.1. Sample preparation.....	102
4.2.4.2. GC-MS.....	103
4.2.5. Method validation.....	104
4.2.6. Analysis of real samples.....	105
4.2.7. Software.....	105
4.3. Results and discussion.....	106
4.3.1. Characterization of the magnetic composite.....	106
4.3.2. Optimization of the MD- μ SPE procedure.....	110
4.3.2.1. Preliminary evaluation of the elution solvent.....	110
4.2.3.2. Screening for the important factors.....	111
4.2.3.3. Optimization of the critical factors.....	113
4.3.3. Method validation.....	116
4.3.4. Analysis of real samples.....	119
4.4. Conclusions.....	119
Note of the author.....	121
References.....	121
Chapter 5 Exploring the potential of aquatic plants and carbon-based materials in detoxifying polluted sediments.....	125
5.1. Introduction.....	125
5.1.1. Eutrophication of water bodies and bioremediation.....	125
5.1.1.1. Bioremediation.....	126
5.1.1.2. Carbon-based materials in remediation processes.....	127
5.1.1.3. Ecophysiology of submerged macrophytes: an outline.....	128
5.1.2. Mantua: the city shaped by water.....	129

5.1.2.1. Rise of the industries.....	130
5.1.2.2. The Mantuan contaminated site	131
5.1.3. Aromatic organic compounds in the Mantuan SNI.....	133
5.2. Materials and methods	135
5.2.1. Chemicals and materials.....	135
5.2.2. Optimization of the extraction procedure	135
5.2.3. Operating procedure and instrumental conditions	136
5.2.3.1. Sample preparation	136
5.2.3.2. GC-MS.....	136
5.2.4. Method validation.....	137
5.2.5. Evaluation of bioremediation capability	138
5.2.5.1. Plants collection and hydrochar production	138
5.2.5.2. Sediment collection and processing	138
5.2.5.3. Experimental setup	139
5.2.5.4. Data analysis.....	139
5.2.6. Software	139
5.3. Results and discussion	140
5.3.1. Optimization of the SPME procedure.....	140
5.3.2. Method validation.....	143
5.3.3. Evaluation of bioremediation capability	145
5.3.3.1. Exploratory data analysis	147
5.4. Conclusions.....	150
Note of the author	151
References	151

Chapter 6 | Cocrystals for agriculture and food packaging applications: a chemometric approach for their discovery 157

6.1. Introduction 157

 6.1.1. Cocrystallization as a crystal engineering strategy 157

 6.1.2. Essential oils and Generally Recognized As Safe substances 159

 6.1.3. Computational approaches to cocrystal discovery 162

 6.1.3.1. Quantitative Structure-Property Relationship 163

 6.1.4. Cocrystal-based active food packaging 164

6.2. Materials and methods 165

 6.2.1. Data collection 165

 6.2.1.1. Class assignation 165

 6.2.1.2. Molecular modelling 166

 6.2.2. Data cleaning and validation strategy 170

 6.2.3. Unsupervised modelling 172

 6.2.4. Supervised modelling 172

 6.2.5. Software 172

6.3. Results and discussion 173

 6.3.1. Data cleaning 173

 6.3.2. Unsupervised modelling 174

 6.3.3. Supervised modelling 176

 6.3.3.1. Model diagnostics 178

 6.3.3.2. Analysis of the pseudo-regression coefficients 180

 6.3.4. Validation 181

6.4. Conclusions 183

Note of the author 183

References	183
Appendix.....	189
Chapter 7 Novel strategies to produce secondary packaging material from cosmetic waste	197
7.1. Introduction	197
7.1.1. Chemistry of hair dyes.....	197
7.1.1.1. Environmental and health concerns.....	201
7.1.2. Biology of <i>Fungi</i>	202
7.1.2.1. Growth factors.....	204
7.1.2.2. The potential of laccases in bioremediation	205
7.1.3. Mycelium-based materials	207
7.2. Materials and methods	208
7.2.1. Chemicals and materials.....	208
7.2.2. Culture preparation	209
7.2.3. Optimization of the growth factors.....	209
7.2.4. Optimization of the QuEChERS procedure	212
7.2.5. Operating procedure and instrumental conditions	213
7.2.5.1. Sample preparation	213
7.2.5.2. UV-vis spectroscopy	213
7.2.6. Preliminary evaluation of the degradative capability.....	214
7.2.7. Software	214
7.3. Results and discussion	215
7.3.1. Preliminary experiments	215
7.3.2. Optimization of the growth factors.....	216
7.3.2.1. Germination time	217

7.3.2.2. Shape factor	219
7.3.2.3. Amount of hair dye	221
7.3.2.4. Multicriteria optimization.....	221
7.3.3. Optimization of the QuEChERS procedure	223
7.3.4. Evaluation of the degradative capability	226
7.4. Conclusions.....	228
Note of the author	228
References.....	229
Chapter 8 Remarks and perspectives	233

"I saw my life branching out before me like the green fig tree in the story. From the tip of every branch, like a fat purple fig, a wonderful future beckoned and winked. [...] I saw myself sitting in the crotch of this fig tree, starving to death, just because I couldn't make up my mind which of the figs I would choose. I wanted each and every one of them, but choosing one meant losing all the rest, and, as I sat there, unable to decide, the figs began to wrinkle and go black, and, one by one, they plopped to the ground at my feet."

_ Sylvia Plath, *The Bell Jar*

Preface

This dissertation has been submitted to the PhD Program of Materials Science and Technology (Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma) to fulfill the requirements to obtain the PhD degree.

I did not know what I was signing up for when I have decided to enroll in this PhD program. I remember telling my parents in Summer 2019 if I could give it a go and they said, "If that's what you want, then go for it". Then, I applied.

I have experienced many ups and downs during the whole path, a roller-coaster of joy and desperation. From the excitement of seeing the very first paper with my name on it in a scientific journal, to the subtle but persistent feeling of inadequacy and not being good enough.

At the beginning of 2020 I had just finished attending a course about Chemometrics in Modena and, suddenly, everyone was stuck inside their house. Outside there was a new infectious disease. A global pandemic was something I did not expect to stumble upon during my PhD path.

So, there I was, sitting home in my room forced to take a break from the lab, from colleagues and from what I was used to. I was lucky enough to have a huge dataset to analyze and, besides studying and baking -occasionally-, Chemometrics was what kept me mentally active for the vast majority of the time I spent home. You, the reader, will find the results in **Chapter 6**, my favorite, because, despite the unfortunate circumstances, I was happy to carry out part of this research in my hometown, close to my parents and my brother.

Fabio Fornari

Montichiari, January 2023

List of publications

Chemometric-assisted cocrystallization: supervised pattern recognition for predicting the formation of new functional cocrystals

F. Fornari, F. Montisci, F. Bianchi, M. Cocchi, C. Carraro, F. Cavaliere, P. Cozzini, F. Peccati, P. P. Mazzeo, N. Riboni, M. Careri, A. Bacchi
Chemometrics and Intelligent Laboratory Systems, 226 (2022) 104580

Dispensing Essential Oil Components through Cocrystallization: Sustainable and Smart Materials for Food Preservation and Agricultural Applications

F. Montisci, P. P. Mazzeo, C. Carraro, M. Prencipe, P. Pelagatti, **F. Fornari**, F. Bianchi, M. Careri, A. Bacchi
ACS Sustainable Chemistry and Engineering, 10 (2022) 8388–8399

Metal-organic framework-based magnetic dispersive micro-solid-phase extraction for the gas chromatography–mass spectrometry determination of polycyclic aromatic compounds in water samples

F. Fornari, F. Bianchi, N. Riboni, F. Casoli, A. Bacchi, P. P. Mazzeo, P. Pelagatti, M. Careri
Journal of Chromatography A, 1671 (2022) 463010

Development of novel cocrystal-based active food packaging material by a Quality by Design approach

F. Bianchi, **F. Fornari**, N. Riboni, C. Spadini, C. S. Cabassi, M. Iannarelli, C. Carraro, P. P. Mazzeo, A. Bacchi, S. Orlandini, S. Furlanetto, M. Careri
Food Chemistry, 347 (2021) 129051

A simple and efficient Solid-Phase Microextraction – Gas Chromatography – Mass Spectrometry method for the determination of fragrance materials at ultra-trace levels in water samples using multi-walled carbon nanotubes as innovative coating

N. Riboni, **F. Fornari**, F. Bianchi, M. Careri
Talanta, 224 (2021) 121891

Deciphering the Supramolecular Organization of Multiple Guests Inside a Microporous MOF to Understand their Release Profile

D. Balestri, P. P. Mazzeo, R. Perrone, **F. Fornari**, F. Bianchi, M. Careri, A. Bacchi, P. Pelagatti
Angewandte Chemie, 133 (2021) 10282–10290

A zinc mixed-ligand microporous metal-organic framework as solid-phase microextraction coating for priority polycyclic aromatic hydrocarbons from water samples

F. Bianchi, A. Pankajakshan, **F. Fornari**, S. Mandal, P. Pelagatti, A. Bacchi, P. P. Mazzeo, M. Careri
Microchemical Journal, 154 (2020) 104646

Recent Advances in In Vivo SPME Sampling

N. Riboni, **F. Fornari**, F. Bianchi, M. Careri

Separations, 7 (2020) 6

This page was intentionally left blank.

Chapter 1 | Introduction

1.1. General context

In Europe approximately 2.5 billion tons/year of waste are producedⁱ, deriving from households, agriculture, livestock, and industrial activities. Moreover, it has been estimated that, in 2011, 20% of the manufactured food went to waste, for a total of nearly 130 thousand tonsⁱⁱ. The recent increase in worldwide populationⁱⁱⁱ produced an acceleration of economic development, intensifying manufacturing activities to satisfy the worldwide demand^{iv}, thus requiring a great effort to food system and environmental policies in keeping pace.

Economic processes still based on a linear approach constitute a serious threat to the environment, sustenance of food system, and biodiversity^v. This is not only for the amount of waste that is produced, but also for the emission into the environment of substances and biological entities that alter the equilibrium in ecosystems and pose risks to human health. Up till now only a reduced portion of contaminants are monitored by competent authorities, whereas a lot of emerging contaminants are not under control due to the lack of knowledge and absence of specific regulations [1,2].

The European Union has enacted a series of directives and regulations that aim at improving the waste management policies and regulating certain pollutants for a more sustainable development. Examples are represented by the Directive EU 2018/851^{vi} for waste management, the Regulation EU 2019/1021^{vii} for limiting the emission of persistent organic pollutants (POPs), and the NORMAN project^{viii} (started in 2005) to gain knowledge towards emerging contaminants.

In December 2019, the European Commission announced the so-called European Green Deal^{ix}, comprising a series of policies aimed at meeting the climate objectives of the next future. The main figures of merit include the reduction of European

ⁱ <https://www.europarl.europa.eu/news/en/headlines/society/20180328ST000751/eu-waste-management-infographic-with-facts-and-figures> (accessed 07/01/2023)

ⁱⁱ https://food.ec.europa.eu/system/files/2021-04/fw_lib_stud_rep_pol_ec-know-cen_bioeconomy_2021.pdf (accessed 07/01/2023)

ⁱⁱⁱ <https://unric.org/en/8-billion-people-10-facts-on-the-worlds-population/> (accessed 07/01/2023)

^{iv} <https://www.un.org/en/dayof8billion> (accessed 07/01/2023)

^v <https://www.un.org/development/desa/pd/content/global-population-growth> (accessed 07/01/2023)

^{vi} <https://eur-lex.europa.eu/eli/dir/2018/851/oj> (accessed 07/01/2023)

^{vii} <http://data.europa.eu/eli/reg/2019/1021/oj> (accessed 08/01/2023)

^{viii} <https://www.norman-network.net/> (accessed 08/01/2023)

^{ix} <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1576150542719&uri=COM%3A2019%3A640%3AFIN> (accessed 08/01/2023)

greenhouse gasses emission by 55% before 2030 (with respect to 1990 levels) and the achievement of the climate-neutrality by 2050 to create a toxic-free environment. Along with the regulation of certain pollutants, as among which the POP regulation previously mentioned, it is fundamental to re-design economic processes to make them more circular^x.

Within this frame of reference, the advancements in the field of Materials Science might help in reaching the climate requirements set by the European Green Deal. This Thesis, in the framework of the PhD Program in Materials Science and Technology (Department of Chemistry, Life Sciences and Environmental Sustainability; University of Parma) aims at exploring the potential of novel materials for applications in the field of Analytical Chemistry, and on materials devoted to packaging, and environmental applications.

The development of analytical methodologies able to keep pace with very low contamination thresholds and fast-changing regulations is of vital importance, not only to comply to the regulating authorities, but to enable people in making decision with confidence [3]. Additionally, the advent of Green Analytical Chemistry [4] has posed new challenges in the development of analytical methods focusing on their impact on the environment. In this context, novel sorbent materials and devices for miniaturized sample treatment offer a plethora of possibilities for the development of methods with enhanced selectivity, sensitivity, and lower detection limits, reducing both the amount of sample required for the analyses the use of organic solvents [5,6].

Along with the introduction of technologies with a reduced climate footprint, it is fundamental to implement strategies able to remediate contaminated ecosystems to effectively depollute the environment. In this context, bioremediation provides a sustainable and cost-effective approach for the removal of pollutants from contaminated ecosystems [7]. Along with the metabolic activity of living organisms, the use of carbon-based materials could simultaneously aid the removal of certain pollutants [8,9], promoting the survival of microbial communities, to increase their bioremediating effect [10].

Finally, the reduction of waste or its reuse as raw material for novel applications are fundamental to decrease the overall climate footprint [11,12]. Within this frame of reference, the manufacturing of active food packaging could help in preventing food waste [13,14]. The impact of active food packaging materials could increase even more dramatically if the active ingredients exploiting antibacterial and

^x<https://www.europarl.europa.eu/news/en/headlines/economy/20151201STO05603/circular-economy-definition-importance-and-benefits> (accessed 30/12/2022)

antioxidant activity can be derived from natural and renewable sources, like in the case of essential oils and their active components [15,16], substituting conventional food preservatives. However, efforts have to be made in making such substances more appealing for technological applications [17]. On a closing note, the fabrication of biocomposites [18,19] to be used as building materials or secondary packaging materials could offer a valuable solution for the valorization of waste in the context of circular economy. This process could become even more interesting when the feedstock is represented by toxic wastes, integrating bioremediation and the production of safe biocomposite materials in a single step.

1.2. Outline

This Thesis is organized in 8 **Chapters**, each one exploring a different topic.

Chapter 1 | This **Chapter** provides the general context revolving around the present Thesis, along with an overview of the topics presented in each **Chapter**.

Chapter 2 | This **Chapter** gives a general background about the methodologies utilized in the following **Chapters**. A brief introduction to Chemometrics is provided, addressing its two main branches, i.e., Pattern Recognition and Design of Experiments. The most common Pattern Recognition techniques are discussed, both for unsupervised and supervised modelling. As for Design of Experiments, the process and the reasoning behind an effective experimental planning are outlined, with a focus on few of the most utilized experimental plans for process and mixture factors. The last Section is dedicated to the importance of validation of analytical methods, with an overview on the main quality parameters involved in quantitative analysis.

Chapter 3 | This **Chapter** deals with the noticeable issue of emerging contaminants. The study aims at exploring the potential of carbon nanotubes (CNTs) as novel coating materials for the solid-phase microextraction (SPME) of a pool of emerging contaminants commonly found into surface water samples. The investigated compounds comprised fragrances, UV filters, antioxidants, and plant protection products. The SPME procedure was optimized in terms of selection of the most

suitable CNTs for the extraction of the investigated analytes together with the effect of extraction time and temperature. Method validation proved the reliability of the developed method for the determination of the investigated compounds in surface water samples, collected in various water bodies in Northern Italy.

Chapter 4 | As for the previous one, also this **Chapter** deals with the evaluation of the potential of new materials in sample preparation. The rational design of a magnetic composite based on the Metal-Organic Framework PUM198 is presented. The developed material was used as a sorbent for the magnetic dispersive micro solid-phase extraction (MD- μ SPE) of ubiquitous contaminants, i.e., polycyclic aromatic hydrocarbons, from water samples. The effect of different factors on extraction procedure was studied, firstly with a screening experimental design, and then by optimizing the most critical factors, paying great attention to the principles of Green Analytical Chemistry. Only 5 mL of sample were processed and only 50 μ L of organic solvent were used *per* extraction. The validated MD- μ SPE-GC-MS method was successfully utilized for the determination of the investigated compounds in underground water samples, collected in a contaminated site in the province of Parma.

Chapter 5 | This **Chapter** is dedicated to the MACHY project, a multidisciplinary project carried out at the University of Parma in collaboration with the University of Trento. The aim is the development of a clean and cost-effective bioremediation technology based on the combination of aquatic plants and carbon-based materials derived from hydrothermal carbonization of biomass harvested from an eutrophicated site. The site involved in this study is the Site of National Interest of Mantua (Northern Italy), whose notoriously complex hydrographic network received heavy contamination from the intense industrial activities that started in mid '40s. What is presented in this **Chapter** are the optimization and validation of a SPME-GC-MS method to quantify a pool of volatile aromatic compounds that have been found in past analytical records of groundwater on the Mantuan contaminated site. The method

was, then, applied to assess remediation capabilities of the proposed technology by analyzing the pore water of microcosms set up in the laboratory.

Chapter 6 | The antimicrobial and antioxidant activity of essential oils (EOs) and their active components are well documented in the scientific literature. Being Generally Recognized As Safe substances, EOs and their active components have unparalleled potential in substituting conventional antioxidants and pesticides, with application ranging from the formulation of active food packaging to agriculture. Unfortunately, their industrial applicability suffers from technological limitations, mainly related to their physicochemical properties. As a matter of fact, most of these compounds are volatile liquids or low-melting solids and lacks thermal and light stability. In this context, cocrystallization can offer a solution for the stabilization of these high valuable chemicals into the solid state by pairing them with suitable partner molecules, enwidening their range of applications. A Section of this **Chapter** reports the results of the successful application of cocrystallized EOs active compounds for the manufacturing of active packaging material to extend the shelf life of foodstuff. Unfortunately, the proper selection of the molecular partner still represents the bottleneck of cocrystal engineering. This **Chapter** focusses on the application of supervised modelling in linking the molecular properties of the two partner molecules and the outcome of cocrystallization trials, by following the general idea of Quantitative Structure-Property Relationship.

Chapter 7 | The study reported in this **Chapter** was carried out in close collaboration with Davines S.p.A. (Parma, Italy). The aim is the fabrication of a mycelium-based biocomposite to be utilized as secondary packaging material starting from production wastes. In this case, the starting materials were cardboard scraps and permanent hair dyes, the latter being classified as a source of emerging contaminants. The core idea is to take advantage of exoenzymes produced by *Fungi*, i.e., laccases, that have been reported being active towards the degradation of organic dyes, and the filamentous structure produced during fungal growth,

to integrate the bioremediation of toxic wastes and the production of a biocomposite. What is presented in this **Chapter** is the optimization of the growth medium *via* Mixture Design, in terms of inoculum size, amount of cardboard scraps and hair dye. Finally, a preliminary evaluation of the degradation capability was carried out.

Chapter 8 | The final **Chapter** presents the future perspectives of the results obtained in the framework of this Thesis.

References

- [1] A. Grobelak, A. Kowalska, Emerging environmental contaminants—current status, challenges, and technological solutions, *Emerging Contaminants in the Environment: Challenges and Sustainable Practices*. (2022) 39–53. <https://doi.org/10.1016/B978-0-323-85160-2.00010-X>.
- [2] S. Dey, F. Bano, A. Malik, Pharmaceuticals and personal care product (PPCP) contamination—a global discharge inventory, *Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology Emerging Contaminants and Micro Pollutants*. (2019) 1–26. <https://doi.org/10.1016/B978-0-12-816189-0.00001-9>.
- [3] M. Valcárcel, G.D. Christian, R. Lucena, Teaching social responsibility in analytical chemistry, *Anal Chem*. 85 (2013) 6152–6161. https://doi.org/10.1021/AC400323M/ASSET/IMAGES/LARGE/AC-2013-00323M_0007.JPEG.
- [4] M. de la Guardia, S. Garrigues, Past, Present and Future of Green Analytical Chemistry, in: M. de la Guardia, S. Garrigues (Eds.), *Challenges in Green Analytical Chemistry*, 2nd ed., The Royal Society of Chemistry, 2020: pp. 1–18. <https://doi.org/10.1039/9781788016148-00001>.
- [5] N. Riboni, F. Fornari, F. Bianchi, M. Careri, A simple and efficient Solid-Phase Microextraction – Gas Chromatography – Mass Spectrometry method for the determination of fragrance materials at ultra-trace levels in water samples using multi-walled carbon nanotubes as innovative coating, *Talanta*. 224 (2021) 121891. <https://doi.org/10.1016/J.TALANTA.2020.121891>.
- [6] F. Fornari, F. Bianchi, N. Riboni, F. Casoli, A. Bacchi, P.P. Mazzeo, P. Pelagatti, M. Careri, Metal-organic framework-based magnetic dispersive micro-solid-phase extraction for the gas chromatography–mass spectrometry determination of polycyclic aromatic compounds in water samples, *J Chromatogr A*. 1671 (2022) 463010. <https://doi.org/10.1016/J.CHROMA.2022.463010>.
- [7] P. Singh, V.K. Singh, R. Singh, A. Borthakur, S. Madhav, A. Ahamad, A. Kumar, D.B. Pal, D. Tiwary, P.K. Mishra, Bioremediation: a sustainable approach for management of environmental contaminants, in: P. Singh, A. Kumar, A. Borthakur (Eds.), *Abatement of Environmental Pollutants: Trends and Strategies*, Elsevier, Amsterdam, Netherlands, 2020: pp. 1–23. <https://doi.org/10.1016/B978-0-12-818095-2.00001-1>.
- [8] I. Budiman, D. Hermawan, F. Febrianto, G. Pari, Subyakto, Char properties and pollutant adsorption capability of oil palm shell using hydrothermal process, *Biomass Convers Biorefin*. 9 (2019) 681–688. <https://doi.org/10.1007/S13399-019-00394-5/TABLES/5>.
- [9] C.W. Purnomo, D. Castello, L. Fiori, Granular Activated Carbon from Grape Seeds Hydrothermal Char, *Applied Sciences*. 8 (2018) 331. <https://doi.org/10.3390/APP8030331>.
- [10] A. Sinharoy, B. Jyoti Deka, L.E. Amabilis-Sosa, E.I. Valenzuela, J.A. Quezada-Renteria, A.M. Pat-Espadas, Biochar-Assisted Bioengineered Strategies for Metal Removal: Mechanisms, Key Considerations, and Perspectives for the Treatment of Solid and Liquid Matrixes, *Sustainability*. 14 (2022) 17049. <https://doi.org/10.3390/SU142417049>.

- [11] H. El-Ramady, E.C. Brevik, Y. Bayoumi, T.A. Shalaby, M.E. El-Mahrouk, N. Taha, H. Elbasiouny, F. Elbehiry, M. Amer, N. Abdalla, J. Prokisch, S. Solberg, W. Ling, An Overview of Agro-Waste Management in Light of the Water-Energy-Waste Nexus, *Sustainability*. 14 (2022) 15717. <https://doi.org/10.3390/SU142315717>.
- [12] C. v. Berenguer, C. Andrade, J.A.M. Pereira, R. Perestrelo, J.S. Câmara, Current Challenges in the Sustainable Valorisation of Agri-Food Wastes: A Review, *Processes*. 11 (2022) 20. <https://doi.org/10.3390/PR11010020>.
- [13] S. Yildirim, B. Röcker, M.K. Pettersen, J. Nilsen-Nygaard, Z. Ayhan, R. Rutkaite, T. Radusin, P. Suminska, B. Marcos, V. Coma, Active Packaging Applications for Food, *Compr Rev Food Sci Food Saf*. 17 (2018) 165–199. <https://doi.org/10.1111/1541-4337.12322>.
- [14] F. Bianchi, F. Fornari, N. Riboni, C. Spadini, C.S. Cabassi, M. Iannarelli, C. Carraro, P.P. Mazzeo, A. Bacchi, S. Orlandini, S. Furlanetto, M. Careri, Development of novel cocrystal-based active food packaging by a Quality by Design approach, *Food Chem*. 347 (2021) 129051. <https://doi.org/10.1016/j.foodchem.2021.129051>.
- [15] F. Montisci, P.P. Mazzeo, C. Carraro, M. Prencipe, P. Pelagatti, F. Fornari, F. Bianchi, M. Careri, A. Bacchi, Dispensing Essential Oil Components through Cocrystallization: Sustainable and Smart Materials for Food Preservation and Agricultural Applications, *ACS Sustain Chem Eng*. 10 (2022) 8388–8399. https://doi.org/10.1021/ACSSUSCHEMENG.2C01257/SUPPL_FILE/SC2C01257_SI_001.CIF.
- [16] P.P. Mazzeo, C. Carraro, A. Monica, D. Capucci, P. Pelagatti, F. Bianchi, S. Agazzi, M. Careri, A. Raio, M. Carta, F. Menicucci, M. Belli, M. Michelozzi, A. Bacchi, Designing a Palette of Cocrystals Based on Essential Oil Constituents for Agricultural Applications, *ACS Sustain Chem Eng*. 7 (2019) 17929–17940. <https://doi.org/10.1021/acssuschemeng.9b04576>.
- [17] F. Fornari, F. Montisci, F. Bianchi, M. Cocchi, C. Carraro, F. Cavaliere, P. Cozzini, F. Peccati, P.P. Mazzeo, N. Riboni, M. Careri, A. Bacchi, Chemometric-assisted cocrystallization: supervised pattern recognition for predicting the formation of new functional cocrystals, *Chemometrics and Intelligent Laboratory Systems*. 226 (2022) 104580. <https://doi.org/10.1016/j.CHEMOLAB.2022.104580>.
- [18] S. Yamanaka, R. Kikuchi, Complex of fibers and fungi and a process for preparation thereof, 5.074.959, 1991.
- [19] C. Girometta, A.M. Picco, R.M. Baiguera, D. Dondi, S. Babbini, M. Cartabia, M. Pellegrini, E. Savino, Physico-Mechanical and Thermodynamic Properties of Mycelium-Based Biocomposites: A Review, *Sustainability*. 11 (2019) 281. <https://doi.org/10.3390/SU11010281>.

Chapter 2 | Methodologies

2.1. Chemometric background

Since the Digital Revolution took place there has been an increasing imbalance in the technical capability of producing complex data and the technologies allowing for the extraction of relevant information from them [1,2]. Chemistry was also involved in this revolution. In fact, the advances in fields such as molecular modelling and analytical sciences have made chemists able to easily access a huge amount of complex data, but not so easily able to extract pertinent information from them and to interpret it. This last sentence summarizes the underlying reason that, in the early '70s, brought Bruce R. Kowalski [3] and Svante Wold [4], considered the fathers of chemometrics, to move the first steps in this discipline.

The International Chemometrics Society defines chemometrics as *“the chemical discipline that uses mathematical and statistical methods to design or select optimal measurement procedures and experiments and to provide maximum chemical information by analyzing chemical data.”*

Few considerations can be drawn from the aforementioned paragraph:

- There are two main branches of chemometrics: one devoted to the extraction of relevant information from data, and one devoted to experimental planning and to the identification of optimal operating procedures. The first one goes under the name of Pattern Recognition, whereas the second one is known as Design of Experiments.
- The nature of the data involved in chemical problems is almost always multivariate. This means that the experimental observations are rarely limited to a single variable, but rather they encompass multiple variables measured together. Given their nature, multivariate data are difficult to visualize, as the human brain can perceive no more than three dimensions. Therefore, the analysis of these data requires techniques that belongs to the field of multivariate statistics [5], able to simultaneously account for the effect of the single variables, the correlation between them, and their interactions (Figure 2.1). This aspect is of paramount importance, as only a small fraction of information can be explained in univariate terms, i.e., taking into consideration the effects of single variables, one at a time.
- Data and information are not synonyms (Figure 2.1). Data is what has been collected, and it comprises:

- **Information:** it represents the fraction of variability within the data that is structured. It involves, e.g., correlation between variables, the presence of clusters of observations that share similar characteristics, and anomalies. It should be noted that not all the information present in the data is useful to improve the knowledge about the problem of interest. Therefore, information can be split into useful information and non-pertinent information. In addition, good data are demanded to extract useful information, meaning that the data must be relevant to the problem of interest and that the information of interest should not be covered by the always present random variation.
- **Noise:** it represents the fraction of variability in the data due to random factors that cannot be always controlled. It should be noted that the noise is always present, it can be reduced, but not eliminated.

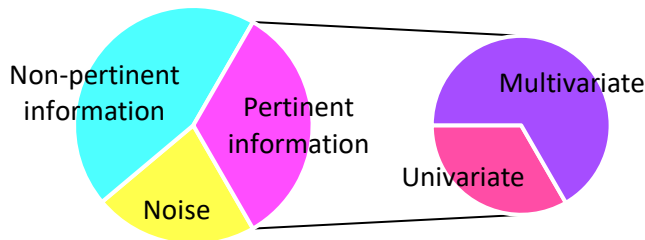


Figure 2.1. Schematic representation of data components. The proportions are arbitrary.

The goodness of the results of data analysis primarily depends on how the data has been collected and only in a minor part on the data analysis itself. Therefore, good sampling practices and a proper experimental design are mandatory to exert control on the random variability and to unbiasedly collect data that are representative of the problem of interest.

Multivariate data are organized in matrices, with I objects (e.g., samples, individuals, observations, experiments) on the rows and J variables on the columns (Figure 2.2). The measured or calculated value of the j^{th} variable on the i^{th} object is reported at the intersection of each row and column.

Rows and columns can be enriched with additional pieces of information, also called attributes. Attributes on the rows can be categorical, such as class indexes, and/or quantitative, such as batch number or an additional property that has been

measured. Attributes on the columns typically are block indexes that define the nature of the variable and/or quantities associated to the column indexes themselves (e.g., wavelengths, m/z ratios, retention times).

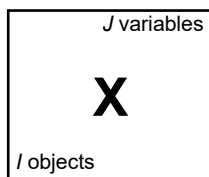


Figure 2.2. Graphical representation of a generic multivariate dataset X . The i objects are on the rows, and the j variables are on the columns.

2.1.1. Pattern recognition

Pattern recognition comprises chemometric techniques devoted to the extraction of pertinent information from chemical data.

Depending on the goal of the data analysis, two main groups of pattern recognition techniques can be identified:

- **Unsupervised:** these techniques allow to visualize the data structure, highlighting the presence of correlations, trends, and clusters without the need of any additional information. In other words, the goal of unsupervised modelling is to explore the data.
- **Supervised:** these techniques allow to find a functional relationship that links one or more attributes of interest associated to the objects with the input variables. The functional relationship that has been found can be used to predict the attributes of interest of future data.

2.1.1.1. Unsupervised Pattern Recognition

Exploratory Data Analysis (EDA) can be considered as the very first step to be performed after the data has been collected [6]. EDA allows to evaluate the quality of the collected data, as well as to assess the sources of information contained in the data itself [7]. Unsupervised Pattern Recognition techniques are particularly suitable for EDA as they aim at highlighting the information within the data without the need of any *a priori* hypothesis or knowledge about the system of interest [6,7].

Principal Component Analysis (PCA) is arguably the unsupervised pattern recognition technique of choice for EDA. PCA [7–9] is an orthogonal transformation that compresses high-dimensional data into a low-dimensional space defined by a set of new variables called Principal Components (PCs).

PCs are derived by the linear combination of the original variables and identify the *directions of maximum variance*. PCs are:

- **Hierarchical:** meaning that the first PC will capture the major source of variation within the data, and each following PC will account for the maximum variance not captured by previous PCs.
- **Orthogonal:** meaning that each following PC must be orthogonal with respect to all the previous PCs. Consequently, the variance captured by each following PC will be uncorrelated with the variance captured by the previous PCs.

The core idea of PCA is summarized in the so-called decomposition equation (Equation 2.1) and can be visualized according to Figure 2.3:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \hat{\mathbf{X}} + \mathbf{E}$$

Equation 2.1

According to Equation 2.1 and Figure 2.3, a model with F PCs decomposes the matrix \mathbf{X} in a matrix $\hat{\mathbf{X}}$ that contains the modelled part of \mathbf{X} and in a matrix \mathbf{E} , also known as residual matrix, that accounts for the unmodelled part of \mathbf{X} . In a good PCA model, all the structured variability in \mathbf{X} should be found in $\hat{\mathbf{X}}$, whereas non-structured variability, i.e., the noise, should remain in the \mathbf{E} matrix.

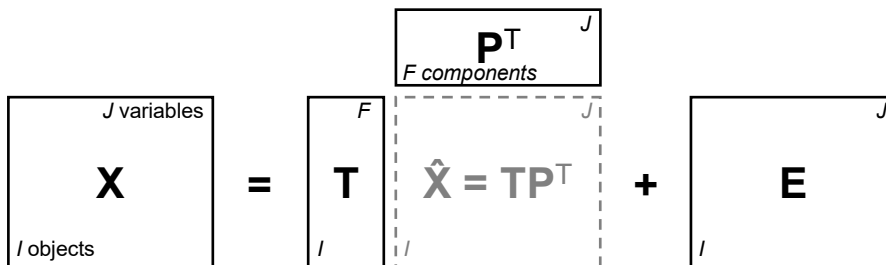


Figure 2.3. Schematic representation of a PCA model.

For any given data matrix \mathbf{X} , the maximum number of PC that can be calculated is the minimum between $I - 1$ and J .

The matrix \mathbf{P} is obtained by the diagonalization of the covariance matrix \mathbf{S} (i.e., a square matrix $J \times J$ whose elements are the covariances between couple of variables of the data matrix \mathbf{X}), along with the diagonal matrix $\mathbf{\Lambda}$ according to Equation 2.2:

$$\mathbf{S} = \mathbf{PAP}^T$$

Equation 2.2

\mathbf{P} contains the eigenvectors, also known as *loadings*. Each eigenvector \mathbf{p} contains the weight of each original variable on a given PC. $\mathbf{\Lambda}$ contains the eigenvalues that are linked to the fraction of variance explained by each PC.

The matrix \mathbf{T} , known as the *score* matrix, is deduced from the original data matrix \mathbf{X} and the loading matrix \mathbf{P} , according to Equation 2.3:

$$\mathbf{T} = \mathbf{XP}$$

Equation 2.3

In other words, each vector \mathbf{t} contains the coordinates of each object of \mathbf{X} on a given PC in the low-dimensional space.

The data modelled by PCA can be visualized by plotting the loadings and the scores in bi- and/or tridimensional scatter plots:

- The score plot provides information about the structure of the data, highlighting, for instance, the presence of clusters or patterns within the data.
- The loading plot provides information about the correlation between the original variables and describes how the original variables relate to a given PC.

The abovementioned derivation of a PCA model goes under the name of Singular Value Decomposition (SVD) [9], but other algorithms such as Non-linear Iterative Partial Least-Squares (NIPALS) have been proposed [10].

2.1.1.2. Supervised Pattern Recognition

While the core idea of Unsupervised Pattern Recognition is to explore the main sources of variability within the data independently by the presence of any attribute, Supervised Pattern Recognition aims at identifying the portion of variability within the data that is related to an attribute (or attributes) of interest, often called *response* (commonly denoted as y) [11–13]. Quantitative and qualitative (or categorical) responses need different supervised techniques for their modelling. *Regression* is suitable for modelling quantitative responses, whereas *classification* is suitable for modelling categorical responses [14].

Classification can be operated in two different ways depending on the purpose of the analysis [12]:

- **Discriminant modelling:** given K modelled classes, a new sample will be mandatorily classified in one, and one only, of those K classes. Linear Discriminant Analysis (LDA) is an example of discriminant method.

- **Class modelling:** given K modelled classes, a new sample will be classified in one class, multiple classes, or none of those K classes. When $K = 1$, the model distinguishes the samples that belongs to the modelled class from everything else. This approach is called one-class classification and it is often used in food authenticity problems. Soft Independent Modelling of Class Analogies (SIMCA) is an example of class modelling method.

The result of a supervised analysis is a model that links the responses with the original data. The model allows the prediction of the attribute of interest of future data.

After representative data has been collected, the steps of supervised modelling are:

1. **Model development:** this phase is dedicated to the selection of the proper model to fit the collected data and to its calculation. This step is also called *training* or *calibration*, and the data set used to build the model is usually called *training set* or *calibration set*. Model development also includes the selection of the proper pretreatments to apply to the data (*preprocessing*) and to tune certain parameters of the model (e.g., number of latent variables, hyperparameters). In this phase, the fraction of variance explained by the model can be assessed. This is expressed with the so-called determination coefficient R^2 [15] (a real number between 0 and 1). The closer R^2 is to 1, more the variance the model is able to explain. Given SS_{TOT} the total variation in the dataset and SS_{RES} the amount of this variation the model is not able to account for (i.e., the residual), R^2 is expressed as in Equation 2.4:

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}}$$

Equation 2.4

2. **Validation:** once the model has been calculated, its predictive performance must be assessed before its application to future data. This step is called *validation* and consists in predicting the response of data included in the so-called *test set* or *validation set* [12-14,16]. The data belonging to this dataset i) are data for whom the response is available, ii) must not be included in the calibration set, and iii) are treated as they were future data. So, the basic idea of validation is to predict the response of the test set and to compare it to the actual values. The parameters that define the goodness of the model in prediction are different depending on the response to be predicted:

- Quantitative: with regression problems the estimate of the predictive capability is expressed in terms of *root mean square error (RMSE)*. This indicates the average deviation that can be expected on the predicted values \hat{y} from their actual values y , according to Equation 2.4.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^I (y_i - \hat{y}_i)^2} = \sqrt{\frac{PRESS}{n}}$$

Equation 2.5

Where n is the number of samples contained in the test set and/or not included in the calibration set. The lower the *RMSE*, the better. *PRESS* stands for sum of squares in prediction, and it can be utilized to estimate an indicator similar to R^2 , i.e., Q^2 (Equation 2.6). Q^2 is a real number lower than 1 and it indicates the predictive capability of the model.

$$Q^2 = 1 - \frac{PRESS}{SS_{TOT}}$$

Equation 2.6

- Categorical: in classification problems the predictive capability of the model is evaluated by means of a table called *confusion matrix*. The confusion matrix summarizes the performance by taking into consideration the actual class of a sample and the class it has been assigned to in prediction (Table 2.1).

Table 2.1. Example of confusion matrix for two classes.

True class	Predicted class	
	P ^a	N ^b
P	TP ^c	FN ^d
N	FP ^e	TN ^f

^a P: positives. ^b N: negatives. ^c TP: true positives. ^d FN: false negatives. ^e FP: false positives. ^f TN: true negatives.

Several parameters can be estimated from the confusion matrix to assess the goodness of the model. *Sensitivity (Sn)* indicates the ability of the model to correctly identify the samples that belongs to a certain class (Equation 2.7a). *Specificity (Sp)* denotes the

ability of the model to correctly identify the samples that do not belong to a certain class (Equation 2.7b).

$$Sn = \frac{TP}{TP + FN}$$

Equation 2.7a

$$Sp = \frac{TN}{FP + TN}$$

Equation 2.7b

The abovementioned parameters refer to one specific class. A global indicator is represented by the *Non-Error Rate (NER)*, calculated as the mean of the Sn values for the K classes included in the model, according to Equation 2.8:

$$NER = \frac{1}{K} \sum_{k=1}^K Sn_k$$

Equation 2.8

The higher the *NER*, the better.

Different validation strategies can be adopted:

- External test set: the test set is constituted of observations that are truly independent from the data included in the calibration set. The use of an external test set allows for a robust estimation of the predictive capability, as important factors of variability are taken into consideration.
- Single-evaluation test set: this approach involves the fractionation of the whole dataset into two subsets of data at the beginning of the data analysis. One of these two subset will be used as a calibration set and the other one as a test set: the latter must not be utilized to test the model until the end of the data analysis. During the splitting of the whole dataset, attention must be paid to replicated measurements: for example, if all observations were conducted in duplicates, the duplicate observation must be kept together either in the calibration set or the test set. In addition, the samples assigned to the test set must be representative of the dataset as a whole and numerous enough to give a reliable

estimate of the predictive capability. The samples to be included in the test set can be selected mainly in two ways:

- Randomly: a randomly selected test set gives a more realistic estimate of the predictive capability of the model, as the distribution of the true response values of future data is not known unless they had been measured.
 - According to a sampling scheme: sampling algorithms have been proposed for this purpose. For instance, the Kennard-Stone algorithm [17] calculates the distance between all the samples and assigns the two that are the farthest apart to the calibration set, then the procedure is iterated until the number of samples to be assigned to the test set is reached. Another example is represented by the DUPLEX algorithm [18] that, similarly to the previous one, assigns the farthest apart samples in the calibration set for odd iterations and in the test set for even iterations until the number of samples to be assigned to the test set is reached.
- Cross-validation: also known as *internal validation*, cross-validation is an iterative procedure that estimates the predictive capability of the model involving only the samples included in the calibration set. This strategy is often used in the development stage to optimize certain parameters of supervised models (e.g., data preprocessing, number of latent variables to be retained, hyperparameters tuning) and to prevent *overfitting*. This phenomenon typically occurs when the data in the calibration set are fitted with a model that is too complex: this results in a high fraction of explained variance, but in a poor predictive capability towards future data. The basic idea of cross-validation is to select a number $C \leq I$ of *cancellation groups*, which is also equal to the number of iterations that will be carried out. At each iteration, I/C samples will be assigned to an *evaluation set* and then predicted with a model calculated with the remaining data. When $C < I$, more samples at a time are left out and the procedure is called *leave-more out* cross-validation, whereas when $C = I$, only one sample at

a time is left out and the procedure is called *leave-one out* cross-validation (sometimes improperly referred to as full cross-validation). The latter typically produces overly optimistic estimates of the predictive capability of the model, especially when there are a lot of samples –since the data used to calculate the model at each iteration differs from the others only for one sample– whereas with smaller datasets it might be a necessity. Different cancellation schemes can be adopted with leave-more out cross validation (Figure 2.4). As stated earlier for the single-evaluation test set strategy, also in cross-validation replicates must be left-out together and no systematicities must occur at each iteration (e.g., leaving-out all the samples belonging to the same class in one iteration in classification problems).

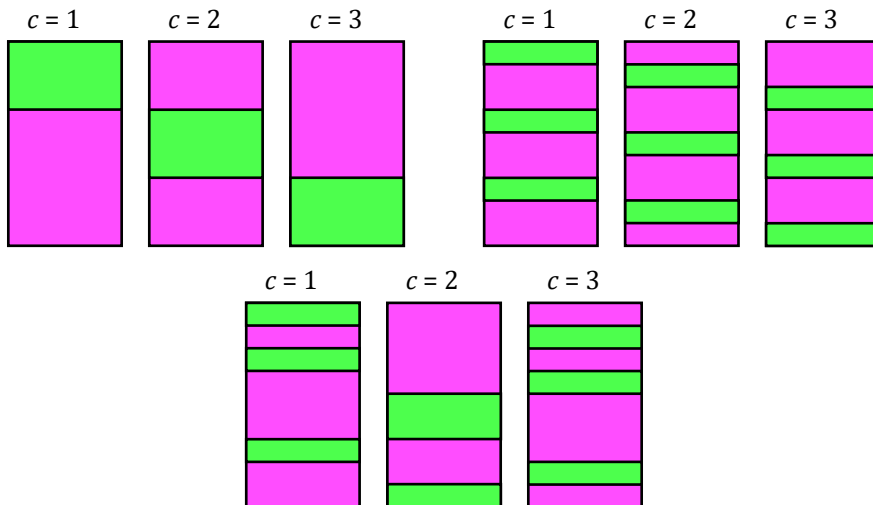


Figure 2.4. Examples of cancellation schemes that can be adopted in leave-more-out cross-validation with $C = 3$ cancellation groups. Each column represents an iteration. The portion of matrix depicted in magenta is utilized to calculate the model whereas the part depicted in green is sent to the evaluation set at each iteration. Top left: contiguous blocks; top right: Venetian blinds; bottom: random subsets.

3. **Application:** once the model has been developed and validated, it is ready to be applied to future samples. The quality of the predictions can be checked by measuring the response of a certain number of future samples.
4. **Updating:** the quality of the prediction might not remain the same over time. This issue can be addressed by including a representative set of future samples for whom the response has been measured in the calibration set. This allows for the inclusion of variability that the old data

might have not been able to explain. Thereafter, the model must be recalculated and revalidated.

Ordinary Least Squares (OLS), also known as Multivariate Linear Regression (MLR), can be considered the simplest regression technique [19,20]. OLS can be applied in a variety of instances including calibration problems, the estimation of response surfaces in Design of Experiments, and in Quantitative Structure-Property and Structure-Activity Relationships (QSPR and QSAR, respectively).

The basic idea of OLS is to find a relationship that captures the variance in the *calibration matrix* \mathbf{X} able to explain the variations observed in the *response matrix* \mathbf{Y} . The general form of an OLS model is summarized in Equation 2.9 and graphically represented in Figure 2.5:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E} = \hat{\mathbf{Y}} + \mathbf{E}$$

Equation 2.9

The variance captured by the model is found in the matrix $\hat{\mathbf{Y}}$, whereas the fraction of variance that the model does not explain is accounted in the residual matrix \mathbf{E} . The matrix \mathbf{B} (Equation 2.10) contains the estimate of the *regression coefficients* that links the responses R to the variables J . In most cases, the sign and the magnitude of a regression coefficient b_{jr} are related to the effect that the j^{th} variable in \mathbf{X} has on the r^{th} response in \mathbf{Y} . An example of exception occurs when interpreting the OLS model obtained in Mixture Design: this aspect will be discussed in a later Section.

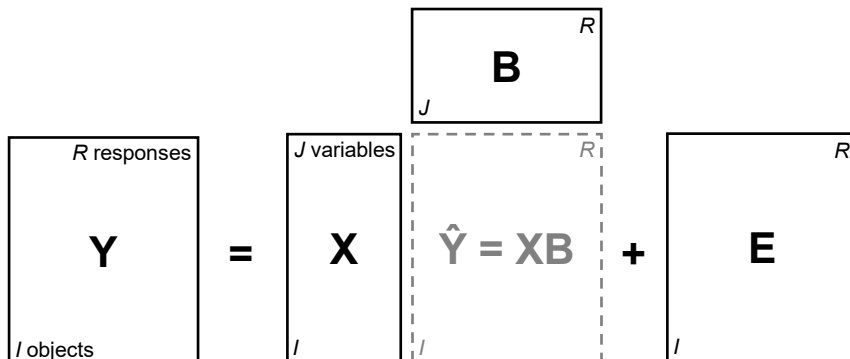


Figure 2.5. Schematic representation of an OLS model.

The estimated regression coefficients in \mathbf{B} assume the values that allow the *minimization of the sum of squared residuals* contained in the matrix \mathbf{E} : this is where the Ordinary *Least Squares* method gets its name from. The minimization of the

residuals is translated in seeking \mathbf{B} so that the *maximum correlation* between \mathbf{X} and \mathbf{Y} is reached.

$$\mathbf{B} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$$

Equation 2.10

The 95% confidence interval of the true regression coefficient β_j associated to the r^{th} response is calculated according to Equation 2.11:

$$\beta_j = b_j \pm t_{\left(\frac{0.05}{2}, \nu\right)} \sqrt{h_{jj}MS_{\text{RES}}}$$

Equation 2.11

Where MS_{RES} is the residual variance of the model and t is the Student's t with $I - J$ degrees of freedom, i.e., the degrees of freedom associated to the residual variance. The term h_{jj} is the element on is the element on the j^{th} row and the j^{th} column of the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$. The calculation of the confidence interval is useful for assessing the significance of a coefficient: if the interval contains the value 0, it means that the considered coefficient is not statistically significant from 0 and, therefore, its associated j variable has a small effect on the r^{th} response.

Once the model has been calibrated and validated, it can be utilized for making predictions. The r^{th} predicted response \hat{y}_{new} of a new object \mathbf{x}_{new} , with the same variables utilized for the calibration of the model, is derived from Equation 2.12 by using the r^{th} set of regression coefficients \mathbf{b} :

$$\hat{y}_{\text{new}} = \mathbf{x}_{\text{new}}\mathbf{b}$$

Equation 2.12

The 95% confidence interval related to the true value y'_{new} is expressed according to Equation 2.13:

$$y'_{\text{new}} = \hat{y}_{\text{new}} \pm t_{\left(\frac{0.05}{2}, \nu\right)} \sqrt{\mathbf{x}_{\text{new}}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_{\text{new}}^T MS_{\text{RES}}}$$

Equation 2.13

Where MS_{RES} is the residual variance of the model and t is the Student's t with $I - J$ degrees of freedom, i.e., the degrees of freedom associated to the residual variance. The term $\mathbf{x}_{\text{new}}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_{\text{new}}^T$ is the *leverage* of \mathbf{x}_{new} . The leverage denotes the contribution of the error in estimating \hat{y}_{new} and its value depends only on the distribution of the objects in \mathbf{X} . The leverage can be calculated also for each sample in \mathbf{X} , and its value can be utilized to assess whether that sample is influential.

In Equations 2.11 and 2.13 the residual variance can be substituted with the pure experimental variance (MS_{PE}) that can be estimated by performing enough replicated measurements, and the degrees of freedom associated to the Student's t needs to be changed accordingly to $I - n_{obs}$, where n_{obs} is the number of independent observations not including the replicates.

The underlying hypotheses of OLS includes that the residuals:

- Should be normally distributed with mean equals to zero,
- Should have the same variance,
- Should be independent, meaning that the value of each individual residual should not depend on any other nor on any external factor, i.e., the value assumed by any variable in \mathbf{X} .

In OLS the variables contained in \mathbf{X} are *assumed as independent*. This aspect can be deduced from Equation 2.10. Since the matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ needs to be computed for the calculation of the coefficients contained in \mathbf{B} , the determinant of the matrix $\mathbf{X}^T\mathbf{X}$ must be non-zero. This means that the variables contained in \mathbf{X} must be orthogonal. It is important to note that a determinant close to zero still allows for the computation of the coefficients, but the obtained model will be unstable, and it will not perform well in prediction as the leverage and the error on the regression coefficient will be inflated. This phenomenon is known as *multicollinearity*. Within this frame of reference, attention must be paid when the model is interpreted, because the regression coefficients can be interpreted one at a time only if the variables are perfectly orthogonal. Another crucial aspect to note is that the computation of a model with finite variance requires the calibration matrix \mathbf{X} to account for *more samples than variables* ($I > J$). On a closing note, when the model is calculated, the correlation between the responses in the \mathbf{Y} matrix is not taken into consideration. This means that R responses yields R models that are *independent* on each other.

The multicollinearity issue can be resolved by forcing the variables contained in \mathbf{X} to be orthogonal. This can be done, for instance, by performing a PCA on the \mathbf{X} matrix and to use the score matrix \mathbf{T} as the calibration matrix. This supervised technique is known as Principal Component Regression (PCR) [21,22] and has two main advantages with respect to OLS:

- Since the F variables of the matrix \mathbf{T} are the PCs, they are orthogonal by definition, so there is no risk of multicollinearity.

- The decomposition of the \mathbf{X} matrix in PCs allows to retain only the major sources of variability within the data.

On the other hand, there is no guarantee that the variability retained by the decomposition in PCs is correlated with the responses in \mathbf{Y} and, again, the correlation between the responses in the \mathbf{Y} matrix is not taken into consideration.

Partial Least Squares (PLS) is, like PCR, another projection-based approach [10,23–25]. PLS seeks a compromise between: i) solving the multicollinearity issue, ii) finding the portion of variability in \mathbf{X} that correlates with \mathbf{Y} , iii) taking into consideration the correlation between the responses in \mathbf{Y} and between the variables in \mathbf{X} , and iv) explaining the maximum variance in \mathbf{X} and \mathbf{Y} .

The general form of a PLS model looks the same as in OLS (Equation 2.7, Figure 2.5), but there are substantial differences in the way the relationship is found between the \mathbf{X} and the \mathbf{Y} matrices.

PLS projects the \mathbf{X} and \mathbf{Y} matrices onto a space characterized by a lower dimensionality. This reduced space is constituted by a new set of variables called Latent Variables (LVs) that are *shared* between the two matrices \mathbf{X} and \mathbf{Y} .

As it happens in PCA, both matrices are decomposed according to Equation 2.14:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

Equation 2.14a

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}$$

Equation 2.14b

As mentioned earlier in Section 2.1.1.1., \mathbf{T} and \mathbf{P} are the score and loading matrices, respectively, that results from the decomposition of the calibration matrix \mathbf{X} . Similarly, \mathbf{U} and \mathbf{Q} are the scores and the loadings, respectively, that results from the decomposition of the response matrix \mathbf{Y} . \mathbf{E} and \mathbf{F} are the residuals matrices. Equation 2.14 has both something in common with the decomposition equation that was mentioned earlier for PCA (Equation 2.1) and something different:

- The scores \mathbf{T} are under an orthogonality constraint as it happens for PCA, so that each LV will capture a portion of variability that is not correlated with the one captured by other LVs. Note that this does not necessarily hold true for the scores \mathbf{U} .
- LVs seek the directions that *maximizes the covariance* between the two score matrices \mathbf{T} and \mathbf{U} . This criterion is a combination between the one of OLS, i.e., maximum correlation, and the one of PCA, i.e., maximum variance.

In fact, in PLS, the relationship between \mathbf{X} and \mathbf{Y} is postulated assuming the existence of a linear dependence of the scores \mathbf{U} on the scores \mathbf{T} . This is called *inner relationship*, since the connection of \mathbf{X} and \mathbf{Y} takes place in the space defined by the LVs. For the f^{th} component, the inner relationship between the score matrices is expressed by Equation 2.15:

$$\mathbf{u}_f = \mathbf{t}_f d_{ff} + \mathbf{h}_f \Rightarrow \mathbf{U} = \mathbf{T}\mathbf{D} + \mathbf{H}$$

Equation 2.15

Where \mathbf{D} is a diagonal matrix that contains the inner regression coefficients d_{ff} and \mathbf{H} is the residual matrix.

The directions of maximum covariance are found by rotating the \mathbf{X} and \mathbf{Y} matrices by means of two new matrices \mathbf{W} and \mathbf{C} , respectively, that contains the *weights*. The relationships that include the weights are similar to the ones expressed in Equation 2.14, and are reported in Equation 2.16:

$$\mathbf{X} = \mathbf{T}\mathbf{W}^T$$

Equation 2.16a

$$\mathbf{Y} = \mathbf{U}\mathbf{C}^T$$

Equation 2.16b

The determination of the PLS weights \mathbf{W} and \mathbf{C} , and the inner regression coefficients \mathbf{D} is carried out one component at a time through an iterative procedure that can be done *via* different algorithms, such as the NIPALS and the SIMPLS algorithms [10,26]. In NIPALS [10], the first step is devoted to the calculation of the scores on the f^{th} component \mathbf{t}_f and \mathbf{u}_f and the respective weights \mathbf{w}_f and \mathbf{c}_f by means of an iterative procedure summarized in Figure 2.6.

Then, the \mathbf{p}_f and \mathbf{q}_f (i.e., the loadings of \mathbf{X} and \mathbf{Y} , respectively) can be derived by regressing \mathbf{X} and \mathbf{Y} onto their respective score vectors \mathbf{t}_f and \mathbf{u}_f (Equation 2.17).

$$\mathbf{p}_f = \mathbf{X}^T \mathbf{t}_f (\mathbf{t}_f^T \mathbf{t}_f)^{-1}$$

Equation 2.17a

$$\mathbf{q}_f = \mathbf{Y}^T \mathbf{u}_f (\mathbf{u}_f^T \mathbf{u}_f)^{-1}$$

Equation 2.17b

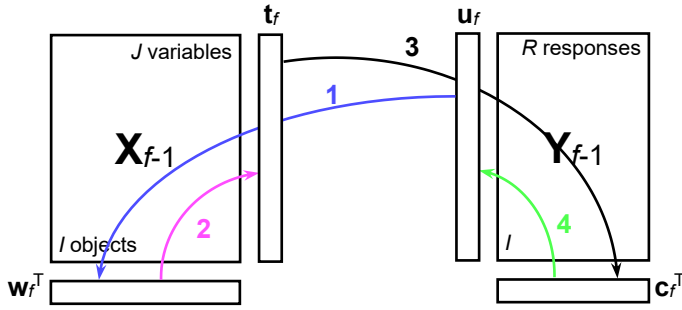


Figure 2.6. Steps of the f^{th} iteration of the NIPALS algorithm. **1.** The scores \mathbf{u}_f are initialized by selecting a random column of \mathbf{Y} , and \mathbf{X} is regressed on \mathbf{u}_f to calculate the weights $\mathbf{w}_f = \mathbf{X}^T \mathbf{u}_f (\mathbf{u}_f^T \mathbf{u}_f)^{-1}$. The weights \mathbf{w}_f are then normalized. **2.** \mathbf{X} is regressed on \mathbf{w}_f to calculate the scores $\mathbf{t}_f = \mathbf{X}^T \mathbf{w}_f (\mathbf{w}_f^T \mathbf{w}_f)^{-1}$. **3.** \mathbf{Y} is regressed on \mathbf{t}_f to calculate the weights $\mathbf{c}_f = \mathbf{Y}^T \mathbf{t}_f (\mathbf{t}_f^T \mathbf{t}_f)^{-1}$. The weights \mathbf{c}_f are then normalized. **4.** \mathbf{Y} is regressed on \mathbf{c}_f to calculate the updated scores $\mathbf{u}_f^* = \mathbf{Y}^T \mathbf{c}_f (\mathbf{c}_f^T \mathbf{c}_f)^{-1}$. If $\mathbf{u}_f^* \approx \mathbf{u}_f$ convergence is reached, deflation is operated, and the algorithm proceeds in the calculation of the next LV. Otherwise \mathbf{u}_f^* will be used as the starting loading vector and another iteration begins.

Additionally, the inner regression coefficient d_{ff} can be derived by regressing the scores \mathbf{u}_f on \mathbf{t}_f (Equation 2.18):

$$d_{ff} = \mathbf{u}_f^T \mathbf{t}_f (\mathbf{t}_f^T \mathbf{t}_f)^{-1} \quad \text{Equation 2.18}$$

The very last step is *deflation*. Deflation removes from the \mathbf{X} and \mathbf{Y} matrices the variance and covariance, respectively, already captured by the previous LV (Equation 2.19). After deflation, the matrices \mathbf{X} and \mathbf{Y} will be used for the calculation of the $f^{\text{th}} + 1$ LV.

$$\mathbf{X}_{f+1} = \mathbf{X} - \mathbf{t}_f \mathbf{p}_f^T \quad \text{Equation 2.19a}$$

$$\mathbf{Y}_{f+1} = \mathbf{Y} - \mathbf{u}_f \mathbf{c}_f^T = \mathbf{Y} - d_{ff} \mathbf{t}_f \mathbf{c}_f^T \quad \text{Equation 2.19b}$$

By isolating \mathbf{U} from Equation 2.14b and introducing it in Equation 2.15, the matrix \mathbf{Y} can be expressed first as a function of the scores \mathbf{T} (Equation 2.16). Then, by comparing Equation 2.16 with Equation 2.15b it can be noted that $\mathbf{C}^T = \mathbf{DQ}^T$. Therefore, Equation 2.20 can be obtained:

$$\mathbf{Y} = \mathbf{TDQ}^T + \mathbf{HQ}^T + \mathbf{F} = \mathbf{TC}^T + \mathbf{F}' \quad \text{Equation 2.20}$$

Equation 2.20 expresses, basically, the OLS regression of \mathbf{Y} onto the reduced orthogonal space \mathbf{T} , and $\mathbf{F}' = \mathbf{H}\mathbf{Q}^T + \mathbf{F}$ accounts for the residuals deriving both from the regression of \mathbf{Y} on \mathbf{T} and from the regression of \mathbf{U} on \mathbf{T} . For the sake of clarity, \mathbf{Y} can be expressed as a function of \mathbf{X} (Equation 2.21). For doing so, Equation 2.14a can be multiplied on the right by the weights \mathbf{W} so that the scores can be expressed as $\mathbf{T} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}$, and then introduced in Equation 2.14:

$$\mathbf{Y} = \mathbf{X}\mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T + \mathbf{F}'$$

Equation 2.21

Equation 2.15 can be rearranged by expressing the *pseudo-regression coefficients* \mathbf{B}_{PLS} (Equation 2.22):

$$\mathbf{B}_{\text{PLS}} = \mathbf{W}(\mathbf{P}^T\mathbf{W})^{-1}\mathbf{C}^T$$

Equation 2.22

Then, the general form of a PLS model is expressed as in Equation 2.23:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}_{\text{PLS}} + \mathbf{F}' = \hat{\mathbf{Y}} + \mathbf{F}'$$

Equation 2.23

Where the matrix $\hat{\mathbf{Y}}$ denotes the fraction of total variance in the data that was captured by the model. As it is for OLS (Equation 2.12), a new object \mathbf{x}_{new} can be projected according to Equation 2.17 to obtain an estimate of the predicted response $\hat{\mathbf{y}}_{\text{new}}$.

As it happens for PCA and OLS, respectively, the inspection of the distribution of the scores, loadings, and weights, and the sign and the magnitude of the pseudo-regression coefficients \mathbf{B}_{PLS} are fundamental to retrieve useful information.

- **Scores:** as in PCA, the scores \mathbf{T} and \mathbf{U} can be plotted in scatter plots to visualize the samples in a low-dimensional space. In addition, \mathbf{U} can be plotted component-wise against \mathbf{T} to observe the inner relationships.
- **Leverage:** as in OLS, the leverage shows the influence of a sample in the model. It is calculated similarly as it is for OLS, but it involves the scores.
- **Loadings and weights:** the scores \mathbf{P} and \mathbf{Q} can give indications about the correlation between the variables in the \mathbf{X} and \mathbf{Y} spaces, respectively. The weights \mathbf{W} denotes which variables in \mathbf{X} correlates with the responses in \mathbf{Y} .
- **Pseudo-regression coefficients:** \mathbf{B}_{PLS} denotes how a given variable j in the \mathbf{X} space contributes to the determination of a response r in the \mathbf{Y} space.

The importance of each j variable in \mathbf{X} in explaining a specific r response in \mathbf{Y} can be expressed in terms of Variable Importance on Projection (VIP score) [12,27,28]. For a given variable j and in relation of a response r , the VIP score is the squared root of the average variance of \mathbf{Y} captured by each component SSY_f weighted by the PLS weight w_{jf} that variable has on each component (Equation 2.24):

$$VIP_j = \sqrt{\frac{J}{SSY_{TOT}F} \sum_{f=1}^F w_{jf}^2 SSY_f}$$

Equation 2.24

Where J is the number of variables in \mathbf{X} , F is the number of LVs that were retained, and SSY_{TOT} is the total variance of \mathbf{Y} captured by the model. Since the sum of the squared VIP scores for all the variables is equal to the number of variables J , the proposed threshold above which a variable can be considered important is set to 1. The latter consideration is particularly useful in variable selection problems.

Contrary to OLS, in PLS the nature of \mathbf{Y} can be both quantitative and categorical. Therefore, PLS can be utilized both in regression problems (PLS-R) [23,25] and in classification (PLS-DA) [12,29]. In classification, classes in \mathbf{Y} are coded with dummy variables. Equation 2.25 that provides an example of a classification problem with three classes:

$$\mathbf{Y} = \begin{bmatrix} \text{class A} \\ \text{class B} \\ \text{class C} \\ \text{class A} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Equation 2.25

A new sample \mathbf{x}_{new} is assigned to the class that shows the highest predicted response $\hat{\mathbf{y}}_{new}$, since the predicted values are continuous and not dummy-coded. By following the aforementioned example in Equation 2.25, a sample whom predicted response vector is $\hat{\mathbf{y}}_{new} = [1.15 \ -0.02 \ 0.36]$ will be assigned to class A. Nevertheless, the classification rule may vary depending on the application. As it was mentioned earlier in this Section, the classification operated by PLS-DA is discriminant.

2.1.2. Design of Experiments

The other major branch of Chemometrics is Design of Experiments (DoE). The terminology *Design of Experiments* was firstly reported in a book published in mid '30 by Ronald A. Fisher [30], widely considered the father of statistics. In "*The Design of Experiments*", Fisher elucidates the foundations behind strategies to

extract the maximum possible information from a number of experiments. This is achieved by selecting among the possible experimental conditions the most informative ones. These aspects summarize the core idea and the goal of DoE.

An experimentation is characterized by [31,32]:

- **Variables:** also referred to as *factors* within the DoE terminology, represent all the sources of variation within an experimentation. Factors can be categorial (e.g., the solvent in which a chemical reaction is carried out), discrete (e.g., power settings expressed as arbitrary units), and continuous (e.g., temperature). The values that a factor can assume are called *levels* and the range of values in which the experimenter decides to study a given factor is called *experimental domain*. Factors can be:
 - Under control: if their value can be set at a desired level.
 - Out of control: if their value changes deliberately. Examples of factors that are almost always out of control include the operator, aging, and the environmental conditions (i.e., room temperature, ambient pressure, and humidity).
- **Responses:** represent the properties of interest that can be measured as the *outcome* of an experimentation. As for the factors, the nature of the responses can be categorial and quantitative, either discrete or continuous. For the sake of brevity, only the strategies for the analysis of quantitative responses will be reported in this Section.

The goal of an experimentation is to find how factors relate to the responses the experimenter is interested in. DoE methodologies allow the experimenter to study *simultaneously* the effects the factors of interest and their interactions have on a response within the experimental domain in a single set of *planned experiments*. The set of experiments involves that all the factors taken into consideration are varied together in a reasonable way. After the data has been collected, an empirical model can be produced, permitting i) the investigation on how the factors influence the response, and, in some instances, ii) the prediction of the response in any point within the experimental domain with a known precision, even for experimental conditions for which the experiment was not performed.

This general idea behind DoE is in contraposition with the one behind the so-called One-Factor-At-a-Time (OFAT) approach. The latter is still, unfortunately, widely utilized for the study and optimization of processes. According to the OFAT strategy, the experimenter will keep all the factors fixed at one level except one that

will be varied at various levels. Once the best value of the response has been reached, the factor will be set and another one will be varied, and so on until all the factors have been varied once. It might seem that the OFAT approach is simpler and allows the experimenter a better control of the experimental conditions than DoE, but it has severe limitations [32,33]:

- The fact that only one factor at a time is changed assumes the complete independence between factors. In other words, the interactions between factors will be completely overlooked.
- The total number of experiments to be carried out cannot be forecasted *a priori*.
- Since OFAT allows to know the value of the response only in the experimental conditions for which the experiment has been performed, the *true* optimal conditions can be missed in the exploration of the experimental domain.

The steps involved when aiming to solve a problem with a set of experiments according to DoE can be summarized as follows [33]:

1. **Definition of the problem:** this step is dedicated to the understanding of the system object of the experimentation, to the definition of the resources, and to the definition of the problem that want to be solved with a set of experiments. This step is the most important one, because if it is carried out in a lousy way, the results of the experimentation will not be in the end as informative as they could.
2. **Definition of the factors and the responses:** in this step the experimenter must select which factors to study. The selection of the factors study should reflect to the amount of prior knowledge about the problem. In this context, three main objectives can be identified:
 - Exploration: exploration reflects a situation in which there is almost no prior knowledge about a given system. Therefore, the number of factors that are taken into consideration can be very large. In addition, in exploration it should be questioned how wide the experimental domain associated to each factor should be.
 - Screening: screening is midway, and its aim is the identification of the most important factors that are worth further investigation. In addition, in screening important hints about whether the

experimental domain that was selected is useful or it has to be adjusted can be retrieved, as well as information about interactions between factors.

- **Optimization:** for optimization a high level of knowledge is demanded. The factors to be studied are almost always known and they are very limited in terms of numerosity. The objective in optimization is the identification of the direction in which the response assumes the desired value.

For what concerns the response, it must be decided what properties that relate to the problem of interest are measurable.

3. **Selection of the design and experimental planning:** depending on the objective, the amount of prior information that is available, the presence of constraints, how many experiments can be afforded, and the most appropriate experimental plan have to be chosen. During this step, attention should be paid towards factors that are out of control, as they can introduce trends within the observation. Two possible solutions are represented by *randomization* and *blocking*: the first strategy consists in randomizing the order in which the experiments will be performed, whereas the second one involves the execution of all the experimental conditions with respect to a factor that cannot be controlled.
4. **Execution of the experiments:** after experimental planning, the experiments are executed according to the established order.
5. **Analysis of the results:** this step is dedicated to the extraction of information from data.

As it was stated earlier, it is not guaranteed that the problem of interest can be solved with a single set of experiments. Therefore, after the analysis of the results, the experiments can be re-planned to take into consideration also the amount of information that was gained with the first experimentation.

In DoE, the *design matrix* contains the standardized values of the investigated factors, and it will be used as the **X** matrix, whereas the responses will be stored into a **Y** matrix. In the case of quantitative responses, the data analysis process typically involves OLS, but in the case of multiple responses also PLS-R can be utilized.

2.1.2.1. Screening Designs

Screening designs are particularly useful both in explorative and screening studies, as they allow to evaluate the effect of many factors with a parsimonious number of experimental runs.

Plackett-Burman Designs^{xi} (PBDs) are exceptionally useful both in exploration and in screening, as they allow the evaluation of the importance of a high number of factors with a limited number of experimental runs [31–34]. In this design, the first-order interaction terms are aliased with the main factors and no main factors are aliased with each other (Resolution III): this means that PBD allows only the evaluation of the effect the main factors have on the response.

The general form of an OLS model for a PBD for k factors is reported in Equation 2.26:

$$y = b_0 + \sum_{i=1}^k b_i X_i + e$$

Equation 2.26

Where y is the response, b_0 is the intercept and b_i is the regression coefficient associated to the factor X_i , and e represents the residual.

In a PBD each factor is studied at two levels (-1 and +1) being i) the extremes of the experimental domain for continuous and discrete factors or ii) the two conditions of a dichotomous categorial factor. The N experiments are planned according to an Hadamard’s matrix, which is a square matrix of order N , where N is a multiple of 4. Table 2.1 reports an example of Hadamard’s matrix for $N = 8$.

Table 2.1. Hadamard’s matrix for $N = 8$. The first column contains only ones and it is utilized for the estimation of the intercept b_0 .

Standard order	Intercept	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	1	1	1	1	1	1	1	1
2	1	-1	1	-1	1	-1	1	-1
3	1	1	-1	-1	1	1	-1	-1
4	1	-1	-1	1	1	-1	-1	1
5	1	1	1	1	-1	-1	-1	-1
6	1	-1	1	-1	-1	1	-1	1
7	1	1	-1	-1	-1	-1	1	1
8	1	-1	-1	1	-1	1	1	-1

^{xi} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri335.htm> (accessed 12/10/2022)

As it can be seen from the example, an experimental plan accounting for $N = 8$ experiments allows the evaluation of the effect of maximum $k = 7$ factors. This means that, given k the number of factors to be investigated, the minimum number of experiments N to be carried out is going to be the minimum multiple of 4 that is greater than k . Therefore, N increases somewhat slowly as k increases.

If k is lower than $N - 1$ the design defined as *unsaturated*. In this case, the remaining $N - k - 1$ columns can be utilized to evaluate the coefficient of *dummy variables*, i.e., redundant factors whose coefficient can be used to estimate the pure experimental variance. The utilization of dummy variables allows to capture aspects of the normal variability the process under study is subjected, e.g., operator, lot number of a starting material, etc. The experimental variance associated to the dummy variables MS_d is expressed in Equation 2.27:

$$MS_d = \frac{N}{N - k - 1} \sum_{i=1}^{N-k-1} d_i^2$$

Equation 2.27

The 95% confidence interval for a given coefficient can be derived from Equation 2.11 by using the variance estimated as in Equation 2.27. The associated degrees of freedom are equal to the number of dummy variables. In a PBD, the term h_{ij} in Equation 2.11 is equal to $1/N$.

If k is greater than $N - 1$, the design defined as *saturated*, and no dummy variables can be fitted into the design matrix. In this instance Fang Dong [35] proposed a procedure to evaluate the experimental variance (Equation 2.28) based on the median absolute value of the coefficients of the factors that were investigated:

$$MS_D = \frac{1}{m} \sum_{i=1}^m b_i^2$$

Equation 2.28

Where the elements of the sum are the regression coefficients that do not satisfy Equation 2.29 and m is the number of them:

$$|b_i| > \frac{15}{4} \text{Me}\{|\mathbf{b}|\}$$

Equation 2.29

Where $\text{Me}\{|\mathbf{b}|\}$ represents the median of the absolute values of the regression coefficients of the investigated factors. The 95% confidence interval is expressed according to Equation 2.30. Note that a Bonferroni correction is necessary.

$$\beta_i = b_i \pm t_{\left(\frac{0.05}{2m}; m\right)} \sqrt{MS_D}$$

Equation 2.30

Full Factorial Designs^{xii} (FFDs) are another option for screening [31–34]. FFD offers an advantage with respect to PBD, as in FFD none of the main factors nor interactions are aliased (“Infinite” Resolution): this means that with an FFD the main factors and their higher-order interactions can be determined. The general form of an OLS model for a FFD for $k = 3$ factors is reported in Equation 2.31:

$$y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3 + e$$

Equation 2.31

In two-levels FFD, each factor is studied on two levels, similarly to PBD. Unlike PBD, the number of experiments N increases more rapidly as k increases, as the number of experiments to be performed is expressed as in Equation 2.32:

$$N = 2^k$$

Equation 2.32

Table 2.2 reports an example of a design matrix for $k = 3$ planned according to a FFD. Figure 2.7 shows the disposition of the experiments for a 2^3 FFD in the multivariate space.

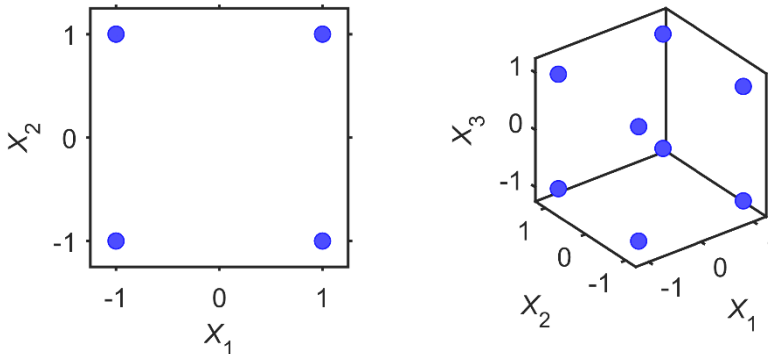


Figure 2.7. Location of the experiments of a two-levels FFD in the multivariate space for $k = 2$ factors (left) and for $k = 3$ factors (right).

^{xii} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3331.htm> (accessed 12/10/2022)

Table 2.2. Design matrix for $k = 3$. The first column contains only ones and it is utilized for the estimation of the intercept b_0 . Note that the columns related to the interaction terms can be obtained by multiplying the elements of the columns of the main factors involved in the interaction.

Standard order	Intercept	X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	$X_1 X_2 X_3$
1	1	-1	-1	-1	1	1	1	-1
2	1	1	-1	-1	-1	-1	1	1
3	1	-1	1	-1	-1	1	-1	1
4	1	1	1	-1	1	-1	-1	-1
5	1	-1	-1	1	1	-1	-1	1
6	1	1	-1	1	-1	1	-1	-1
7	1	-1	1	1	-1	-1	1	-1
8	1	1	1	1	1	1	1	1

The significance of the coefficients associated to the main factors and the interactions can be assessed by taking into consideration the experimental variance, which can be estimated in two ways:

- **By replicating all the experiments:** in this case all the experiments are carried out in independent replicates. This strategy allows a more reliable estimate of the experimental variance, as the associated degrees of freedom are higher, but it is expensive as it needs all the experimental conditions to be replicated. The experimental variance is calculated as the average variance weighted by the degrees of freedom ν for each of the experimental conditions (Equation 2.33):

$$MS_{PE} = \frac{1}{\sum_{i=1}^N \nu_i} \sum_{i=1}^N \nu_i s_i^2$$

Equation 2.33

- **By replicating a single experimental condition:** in this case only a single experimental condition is replicated. This condition is additional to the experiments accounted for the experimental plan, and it is typically performed in the center of the experimental domain, at the level 0 of all the investigated factors (for dichotomous categorical factors, one of the two conditions can be selected to perform these additional experiments). This strategy assumes that the estimated variance in the center of the experimental domain is representative of the variance that can be expected all over it. In terms of reliability, it is less reliable than the previous strategy in terms of degrees of freedom, but it is more advantageous in

terms of experiments that have to be performed, as the number of replicated experiments n_0 is typically lower than the number of replicated experiments that need to be carried out if all the experimental conditions were replicated. An additional advantage of this strategy is that it allows to estimate whether the response y behaves linearly or if there is a significant curvature.

In both cases, the 95% confidence interval for the coefficients is expressed according to Equation 2.10.

As the number of experiments to be performed in FFD increases rapidly as the number of factors k increases, based on the assumption that high-order interactions become less and less likely to be significant, less experiments have to be planned. This is the core idea of Fractional Factorial Designs^{xiii} (FrFDs), which are totally similar to FFDs.

The number of experiments to be performed is (Equation 2.34):

$$N = 2^{k-p}$$

Equation 2.34

Where p is a reduction factor. The price to pay to perform less experiments is Resolution: this means that depending on k and p it will or will not be possible to retrieve information about all the factors and interactions that have been investigated. There are optimum combination of k and p that still permits a good Resolution.

2.1.2.2. Response Surface Designs

The designs discussed so far are suitable in the initial stages of the understanding of a process. When the number of factors has been narrowed down to the very critical ones and the experimental domains were properly chosen the prior knowledge about the problem can be considered sufficient to proceed with the optimization stage. Response surface designs are suitable for optimization and allow a deeper understanding of the problem. These designs allow the estimation of the so-called *response surface*^{xiv}.

The OLS model involved in this typology of designs is of a higher order than the ones involved for screening. Typically, a quadratic model (Equation 2.35) is sufficient to describe most of the processes under study, but cubic models are also possible. Similarly to what was stated earlier, high-order interactions are typically

^{xiii} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3344.htm> (accessed 12/10/2022)

^{xiv} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3336.htm> (accessed 14/10/2022)

neglected under the assumption they become less and less likely to be significant as the order of interaction increases.

$$y = b_0 + \sum_{i=1}^k b_i X_i + \sum_{i \neq j}^{C_{k,2}} b_{ij} X_i X_j + \sum_{i=1}^k b_{ii} X_i^2 + e$$

Equation 2.35

In this case, categorical and discrete factors are not so easy to handle, and experiments should be planned according to computer-aided designs, and D-optimal design [36] is an example of them. This strategy is valuable also in the presence of constraints that results in the experimental domain having an irregular shape (i.e., incompatibility between certain combinations of factors).

In order to detect and resolve quadratic terms, each factor must be studied at least on three levels (-1, 0, and +1). Even if in FFD with additional experiments in the center point the factors are technically studied on three levels, quadratic terms are aliased with each other. A possible solution is represented by three-levels FFDs (Figure 2.8), but they have severe limitations:

- The number of experiments increases way more rapidly than two-levels FFDs as k increases (Equation 2.36):

$$N = 3^k$$

Equation 2.36

- They lack a favorable property called *rotability*^{xv}. A design is rotatable if the variance associated to the prediction \hat{y} depends only on the distance between the desired point and the center point of the experimental domain. Rotability depends on the distribution of the experiments in the multivariate space.

Box-Wilson Central Composite Design^{xvi} (CCD) is arguably the most common response surface design [31–34]. It is called *composite* because it is formed by the combination of a two-levels FFD, center points, and a so-called *star-design* (Figure 2.9). The experimental conditions involved in the star-design are known as *star points* or *axial points* and their purpose is to estimate the squared terms in the response function. The distance between the center points and the axial points is denoted as α , whose absolute value is greater or equal than 1. Depending on the

^{xv} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri336.htm> (accessed 14/10/2022)

^{xvi} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3361.htm> (accessed 15/10/2022)

value of α , different properties can be achieved, and different typologies of CCD can be obtained (Figure 2.10).

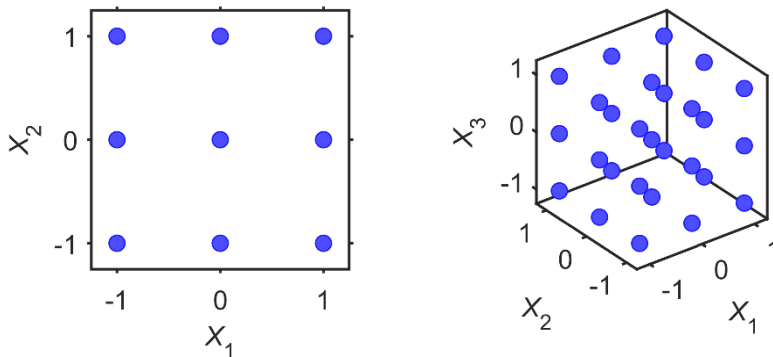


Figure 2.8. Location of the experiments of a three-levels FFD in the multivariate space for $k = 2$ factors (left) and for $k = 3$ factors (right).

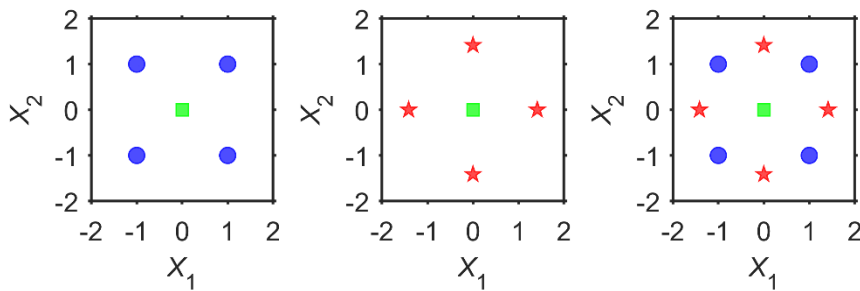


Figure 2.9. Location of the experiments in the multivariate space for a CCD with $k = 2$ factors. Left: two-levels FFD with an additional center point; middle: star-design with an additional center point; right: CCD. Blue dots: factorial points; red stars: axial points; green squares: center points.

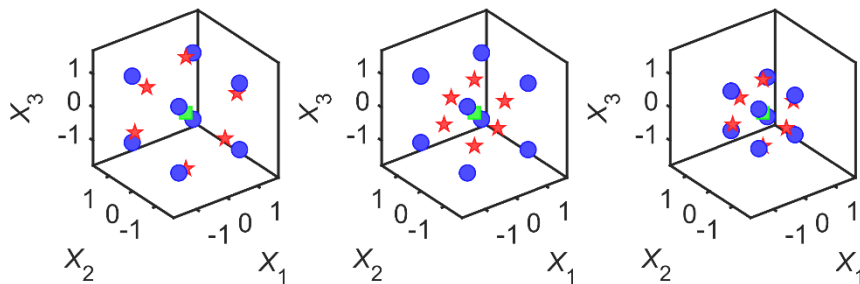


Figure 2.10. Location of the experiments in the multivariate space for the three typologies of CCD with $k = 3$ factors. Left: circumscribed; middle: faced; right: inscribed. Blue dots: factorial points; red stars: axial points; green squares: center points.

- **Circumscribed (CCC):** the experiments are arranged in k -dimensional spherical symmetry. This means that both factorial and axial points are at the same distance from the center point. The distance between the center points and the axial points α is related to the number of factors k according to Equation 2.37:

$$\alpha = \sqrt[4]{2^k}$$

Equation 2.37

This design is rotatable and the quality of the prediction is high all over the experimental domain, but each factor requires to be studied on five levels $(-\alpha, -1, 0, +1, +\alpha)$. Therefore, in the step dedicated to the selection of the experimental domain, it must be evaluated in advance whether the conditions $\pm\alpha$ (which are outside the experimental domain) are feasible or not.

- **Faced (CCF):** in this design the distance between the axial points and the center point is $\alpha = 1$. This means that factorial points and axial points are at a different distance from the center point. An advantage with respect to the CCC, the faced design requires the factors to be studied only on three levels $(-1, 0, +1)$ still guaranteeing a good prediction quality all over the experimental domain. Nevertheless, as in three-levels FFD, this design lacks rotatability and, moreover, the precision to estimate quadratic terms is poorer than the CCC. Note that for $k = 2$, the arrangement of the experiments for a CCF is the same as a three-levels FFD.
- **Inscribed (CCI):** as in the CCC, also in CCI the experiments are arranged in k -dimensional spherical symmetry. This means that both factorial and axial points are at the same distance from the center point and, as in CCF, $\alpha = 1$. It has, therefore, the same properties of a CCC, but the portion of the experimental domain in which the quality of the prediction is high is much more limited. It requires each factor to be studied on five levels as well.

The number of experimental runs N of a CCD increases less rapidly than in a three-levels FFD (Equation 2.38):

$$N = 2^k + 2k + n_0$$

Equation 2.38

n_0 is the number of center points that should be sufficient to have a good estimate of the pure experimental variance. n_0 can be also selected to give the design

properties such as orthogonality and nearly uniform variance on the prediction [34,37].

As an example, the experimental plan for a CCF for $k = 3$ factors and $n_0 = 1$ is reported in Table 2.3.

Table 2.3. Design matrix for $k = 3$ for a CCF. The first column contains only ones and it is utilized for the estimation of the intercept b_0 . Note that the columns related to the interaction terms can be obtained by multiplying the elements of the columns of the main factors involved in the interaction.

Standard order	Intercept	X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	X_1^2	X_2^2	X_3^2
1	1	-1	-1	-1	1	1	1	1	1	1
2	1	1	-1	-1	-1	-1	1	1	1	1
3	1	-1	1	-1	-1	1	-1	1	1	1
4	1	1	1	-1	1	-1	-1	1	1	1
5	1	-1	-1	1	1	-1	-1	1	1	1
6	1	1	-1	1	-1	1	-1	1	1	1
7	1	-1	1	1	-1	-1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1
9	1	-1	0	0	0	0	0	1	0	0
10	1	1	0	0	0	0	0	1	0	0
11	1	0	-1	0	0	0	0	0	1	0
12	1	0	1	0	0	0	0	0	1	0
13	1	0	0	-1	0	0	0	0	0	1
14	1	0	0	1	0	0	0	0	0	1
15	1	0	0	0	0	0	0	0	0	0

Box-Behnken Design^{xvii} (BBD) is another experimental design suitable for the estimation of the response surface [31–34]. This experimental plan can be utilized only if the number of main factors to be investigated is $k \geq 3$.

BBD designs are rotatable and all the experiments are located at the same distance from the center points and, similarly to a CCF, they require each factor to be studied at three levels (-1, 0, +1). In a BBD the experiments are located differently with respect to a CCD (Figure 2.11).

The number of experimental runs N of a BBD is derived from Equation 2.39:

$$N = 2k(k - 1) + n_0$$

Equation 2.39

^{xvii} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3362.htm> (accessed 16/10/2022)

As an example, the experimental plan for a BBD for $k = 3$ factors and $n_0 = 1$ is reported in Table 2.4.

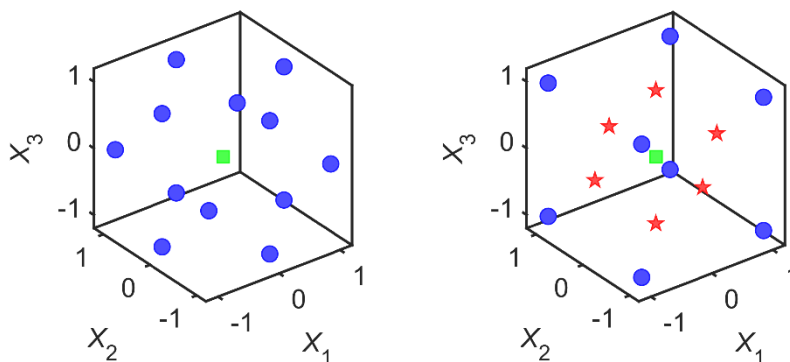


Figure 2.11. Location of the experiments in the multivariate space for a BBD (left) and a CCF (right) with $k = 3$ factors. Blue dots: experiments for the BBD and factorial points for the CCF; red stars: axial points; green squares: center points.

Table 2.4. Design matrix for $k = 3$ for a BBD. The first column contains only ones and it is utilized for the estimation of the intercept b_0 . Note that the columns related to the interaction terms can be obtained by multiplying the elements of the columns of the main factors involved in the interaction.

Standard order	Intercept	X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	X_1^2	X_2^2	X_3^2
1	1	-1	-1	0	1	0	0	1	1	0
2	1	1	-1	0	-1	0	0	1	1	0
3	1	-1	1	0	-1	0	0	1	1	0
4	1	1	1	0	1	0	0	1	1	0
5	1	-1	0	-1	0	1	0	1	0	1
6	1	1	0	-1	0	-1	0	1	0	1
7	1	-1	0	1	0	-1	0	1	0	1
8	1	1	0	1	0	1	0	1	0	1
9	1	0	-1	-1	0	0	1	0	1	1
10	1	0	1	-1	0	0	-1	0	1	1
11	1	0	-1	1	0	0	-1	0	1	1
12	1	0	1	1	0	0	1	0	1	1
13	1	0	0	0	0	0	0	0	0	0

Compared to CCD, BBD are more efficient as they require less experiments for the same number of factors k . On the contrary, due to the so-called *missing corners*

there are regions of the experimental domain in which the quality of the prediction is poor^{xviii}.

This feature of the BBDs can be taken advantage of, because in this experimental plan, all the experimental conditions that involve the combination of factors at the extremes of the experimental domain are not tested. Difficulties can be encountered when working with this design, as it cannot always be divided in orthogonal blocks when factors out of control have to be taken into consideration.

Mixture Designs are another class of experimental designs suitable to study formulation problems [31–34,38]. All the experimental plans discussed so far are suitable for studying the so-called *process factors*, whose values can be changed within the experimental domain without affecting the value of the other factors. On the contrary, *mixture factors* (e.g., the components of a formulation, the ingredients of a recipe) add up to each other to give 100%: this means that if the value of one factor is changed, the remaining ones must be changed accordingly to sum up to 100%.

This fact has two main consequences:

- Since the total must be 100%, the problem is constrained, meaning that the mixture factors are not independent on each other.
- The problem is not about finding the optimal values between parameters, as it is for process factors, but it is about finding the optimal ratios between mixture factors.

A typical experimental plan for $k = 3$ mixture factors can be visualized as an equilateral triangle (Figure 2.12) in which the vertexes correspond to the three pure components. For $k = 4$ mixture factors, the experimental planned can be visualized as a tetrahedron and the degree of complexity and dimensions increases as k increases.

The experimental plan has no treatment in which all the values for the mixture factors are set at 0%, meaning that all the components of the mixture are absent and, therefore, there is no mixture and no response to be measured. Hence, the OLS model for such an experimental plan has no intercept.

^{xviii} <https://www.itl.nist.gov/div898/handbook/pri/section3/pri3363.htm> (accessed 16/10/2022)

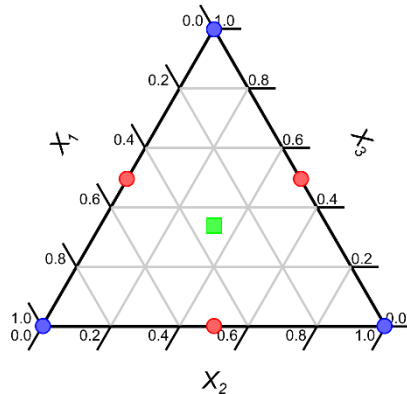


Figure 2.12. Example of the arrangement of the experiments in the multivariate space for a mixture design with $k = 3$ factors. The vertexes (blue dots) represent experimental conditions in which only one of the pure components is utilized. The mid-point of the sides (red dots) represents 50:50 mixtures between the two components that share that side. The center of the experimental domain (green square) represents a condition in which the three components are 1/3 each.

Equation 2.40 shows the OLS model for a mixture design with $k = 3$ factors that can be postulated if the experiments are planned as in Figure 2.12.

$$y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3 + e$$

Equation 2.40

In particular:

- Experiments at the vertexes allow for the estimation of the coefficients related to the main factors.
- Experiments at the mid-points allow for the estimation of the coefficients related to the first-order interactions.
- The experiment at the center point allows for the estimation of the coefficients related to the second-order interaction. This experiment can be replicated for estimating the pure experimental variance instead of replicating all the experimental conditions.

The design matrix that refers to the abovementioned example (Figure 2.12 and Equation 2.33) is reported in Table 2.5. If squared or cubic terms are relevant, then the experiments are planned in a different way as more levels are required to estimate such coefficients.

Table 2.5. Design matrix for $k = 3$. Note that there is no column for the intercept and that for each experimental run, the sum of the fractions for the three main factors is 1.

Standard order	X_1	X_2	X_3	$X_1 X_2$	$X_1 X_3$	$X_2 X_3$	$X_1 X_2 X_3$
1	1.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	1.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	1.00	0.00	0.00	0.00	0.00
4	0.50	0.50	0.00	0.25	0.00	0.00	0.00
5	0.50	0.00	0.50	0.00	0.25	0.00	0.00
6	0.00	0.50	0.50	0.00	0.00	0.25	0.00
7	0.33	0.33	0.33	0.11	0.11	0.11	0.04

An important aspect to note is that in mixture design, the magnitude of the regression coefficients does not reflect the effect that a given factor has on the response as it is generally for OLS models. The magnitude of the effects in mixture design can be deduced only by inspecting the response surface.

It is much more frequent in formulation problems than in process problems to have constraints, e.g., a cake is not a cake anymore if it is made up of eggs only. In this context two main situations can be identified^{xix}:

- Pseudo-component domain:** in this case, the constrained experimental domain can be considered a sized-down version of the original experimental domain (i.e., a smaller equilateral triangle in the case of $k = 3$ factors; Figure 2.13). The vertexes of the constrained experimental domain are called *pseudo-components* that are none other than mixtures with a well-defined composition of the original mixture factors to be investigated [31–33,38]. This situation is particularly advantageous both from the analytical and operational point of view, as: i) the experiments can be planned in the usual way and the coefficients that will result from the analysis are related to the pseudo-components, and ii) the experimenter can prepare bigger batches of the pseudo-components and utilize them to execute the experiments that has been planned.
- Irregular domain:** in this case, the constrained experimental domain is of an irregular shape. As discussed earlier for experimental plans suitable for process factors, computer-aided designs as the D-optimal design [36] are valuable strategies to identify the most relevant combinations of factors to be tested based on the model that has been postulated.

^{xix} <https://www.itl.nist.gov/div898/handbook/pri/section5/pri544.htm> (accessed 18/10/2022)

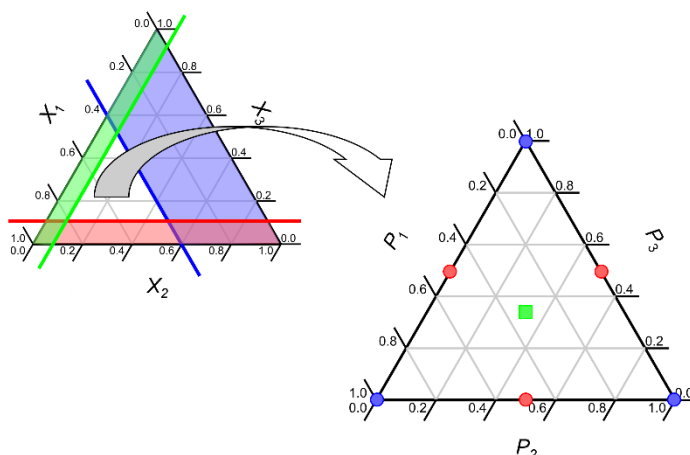


Figure 2.13. Example of constrained experimental domain for $k = 3$ factors in which pseudo-components (P_1 , P_2 , and P_3) can be identified. Their composition corresponds to the proportions of the main factors X_1 , X_2 , and X_3 at the vertexes of the not shaded triangle, i.e., the portion of the experimental domain whose combinations of the main factors are allowed due to the constraints. The experiments in the pseudo-components domain can be planned depending on the model complexity, in this example they have been planned to estimate the coefficients for main factors, first-, and second-order interactions.

There are instances in which mixture and process factors must be taken into consideration together^{xx}. In this case, under the assumption that mixture factors and process factors are independent (i.e., they do not interact), the experiments are almost always planned according to D-optimal design [36].

2.1.2.3. Multicriteria decision making

When the problem under study accounts for more than one response, a strategy to find the set of process and/or mixture factors that guarantees the best performance simultaneously for all the investigated responses must be available [34]. In the case of multiple responses, one way is to utilize PLS instead of OLS in the modelling stage to build a model that accounts for all the responses. In this case, optimization can be carried out graphically by checking both how the experiments (samples) are distributed in the score space and which variables correlate with the responses (weights).

Another solution is the one proposed by George Derringer and his coworker Ronald Suich in 1980 [34,39]. This is known as the Derringer's method of the desirability functions and involves the transformation of each of the R responses in new functions, called *desirability functions* that are then merged into a *global desirability function*.

^{xx} <https://www.itl.nist.gov/div898/handbook/pri/section5/pri545.htm> (accessed 20/10/2022)

The desirability function d_i is derived from the predicted value $\hat{y}_i(\mathbf{x})$ of the OLS model for the i^{th} response and can assume values between 0 and 1:

- If $d_i = 0$ it means that \hat{y}_i assumes an undesirable value for that setting of factors \mathbf{x} .
- If $d_i = 1$ it means that \hat{y}_i assumes a desirable value for that setting of factors \mathbf{x} .

Desirability functions are defined depending on the objective: this means that the shape of the function will be different whether the response have to be maximized (Equation 2.41a), minimized (Equation 2.41b) or it has to reach a target value (Equation 2.41c).

$$d_i = \begin{cases} 0 & \text{if } \hat{y}_i(\mathbf{x}) < L_i \\ \left(\frac{\hat{y}_i(\mathbf{x}) - L_i}{U_i - L_i}\right)^s & \text{if } L_i \leq \hat{y}_i(\mathbf{x}) \leq U_i \\ 1 & \text{if } \hat{y}_i(\mathbf{x}) > U_i \end{cases}$$

Equation 2.41a

$$d_i = \begin{cases} 1 & \text{if } \hat{y}_i(\mathbf{x}) < L_i \\ \left(\frac{U_i - \hat{y}_i(\mathbf{x})}{U_i - L_i}\right)^s & \text{if } L_i \leq \hat{y}_i(\mathbf{x}) \leq U_i \\ 0 & \text{if } \hat{y}_i(\mathbf{x}) > U_i \end{cases}$$

Equation 2.41b

$$d_i = \begin{cases} 0 & \text{if } \hat{y}_i(\mathbf{x}) < L_i \\ \left(\frac{\hat{y}_i(\mathbf{x}) - L_i}{T_i - L_i}\right)^s & \text{if } L_i \leq \hat{y}_i(\mathbf{x}) < T_i \\ 1 & \text{if } \hat{y}_i(\mathbf{x}) = T_i \\ \left(\frac{U_i - \hat{y}_i(\mathbf{x})}{U_i - T_i}\right)^s & \text{if } T_i < \hat{y}_i(\mathbf{x}) \leq U_i \\ 0 & \text{if } \hat{y}_i(\mathbf{x}) > U_i \end{cases}$$

Equation 2.41c

Where L_i , U_i , and T_i are the lower, upper, and target value, respectively, for a given response to be optimized. In other words, they represent the optimization boundaries between which the desirability value d_i varies depending on the predicted response $\hat{y}_i(\mathbf{x})$. It is interesting to note that desirability functions can be computed also for the factors, and not only for the responses. The exponent s (a positive real number) denotes a shape factor that can be set depending on how strict the optimization criteria should be:

- $s < 1$: the desirability function assumes values close to 1 for a wider range of $\hat{y}_i(\mathbf{x})$. This is related to a less strict optimization criterion.
- $s = 1$: the desirability value varies linearly depending on $\hat{y}_i(\mathbf{x})$.
- $s > 1$: the desirability function assumes values close to 1 only if $\hat{y}_i(\mathbf{x})$ is very close to the optimal value. This is related to a stricter optimization criterion.

As already stated, once the desirability functions have been defined for all the responses R , they can be merged into a global desirability function (D , Equation 2.42) which represents the weighted geometric mean for the single desirability values. The geometric mean ensures that experimental conditions \mathbf{x} that produces an undesirable value ($d_i \approx 0$) for a response \hat{y}_i will make $D \approx 0$ as well.

$$D = \sqrt[\sum_{i=1}^R r_i]{\prod_{i=1}^R (d_i(\hat{y}_i(\mathbf{x})))^{r_i}}$$

Equation 2.42

Where r_i (a positive integer) denotes the weight of the i^{th} desirability value with respect to the others. By default, all the desirability values will have the same importance ($r_i = 1$), but their relative importance can be set according to the optimization criteria.

It has to be noted that $D \neq 0$ means that there is at least one set of factors \mathbf{x} that satisfies the optimization criteria for all the responses that have been taken into consideration. A global desirability value close to 1 means that there is a high level of agreement between the single desirability values and, therefore, the responses behave in a similar way depending on the factors. The best compromise is represented by the set of factors \mathbf{x} that maximizes D .

With respect to the validity of the identified conditions there are two alternatives:

- The identified conditions can be validated *a posteriori* by carrying out enough replicated experiments and by comparing the experimental results with the ones predicted by the OLS models.
- The identified conditions can be considered validated *a priori* if the OLS models are not affected by a significant lack-of-fit. The estimation of lack-of-fit requires that replicated experiments have been already carried out in advance.

2.2. Validation of analytical methods

Diagnoses, water and food safety and quality, and the entity of the sentence for the detention of drugs of abuse are just a few of the aspects of modern society that strictly depends on the results of an analytical measurement [40]. This means that, if a decision has to be made based on the result of the analysis, the result itself must be reliable enough to allow the decision to be made with confidence. Therefore, the aim of an analytical determination goes far beyond the mere need of determining an analyte in a given matrix, and it might have strong socio-economic implications [41].

This is the paradigm of *fit for purpose*, and validation is a process that allows an analytical laboratory to claim that the results produced with a given method are reliable for a given purpose. Many definitions of validation have been provided. To mention one, ISO 9000^{xxi} defines it as “*confirmation, through the provision of objective evidence, that the requirements for a specific intended use or application have been fulfilled*”. The same norm also defines verification as the process of “*confirmation, through provision of objective evidence, that specified requirements have been fulfilled*”. Although the two definitions are very similar, there are substantial differences between validation and verification:

- Validation comes before, and it is intended to assess whether a method can deliver results reliable enough depending on the application.
- Verification comes after, and its purpose is the assessment of the capability of a laboratory to produce results reliable enough with a method that has been already validated.

According to ISO/IEC 12025^{xxii} there are instances in which validation is not necessary, for example for standardized methods. In the case of standard methods utilized outside the intended scope or modification of standard methods, or after the development of a new method, validation is necessary. There is no one prescribed way to validate a method: in fact, many guidelines are available to support validation studies depending on the field of interest. In general, validation involves the evaluation of certain quality parameters of the method. The ones to be evaluated may depend on the application, and surely depends on whether the method is intended for qualitative or quantitative analysis.

The following Sections will be devoted to few of the quality parameters to be evaluated in quantitative methods of analysis. For the sake of brevity this and the

^{xxi} <https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v1:en> (accessed 20/10/2022)

^{xxii} <https://www.iso.org/obp/ui/#iso:std:iso-iec:17025:ed-3:v1:en> (accessed 20/10/2022)

following Sections refers to the EURACHEM guidelines [40], as they were followed for the validation of methods presented in the following **Chapters**.

2.2.1. Limits of Detection and Limits of Quantitation

Limit of Detection (LOD) and Limit of Quantitation (LOQ) can be defined as the lowest concentration of analyte that can be detected and quantified, respectively, at a specified confidence level [40,42].

LODs and LOQs can be ascribed both to a specific instrument that has been used for a measurement and to the whole analytical method [42]. The *method LOD and LOQ* reflect the detection and quantitation capabilities of the whole analytical method, whereas *instrumental LOD and LOQ* are related to the capability of a given instrument in detecting or quantifying a given analyte.

The evaluation of LODs and LOQs requires an initial estimate of the standard deviation at near-zero concentration level (s_0). Few aspects have to be considered:

- For spectroscopic techniques s_0 can be evaluated on a blank matrix, whereas for separative techniques fortification is usually required, as in the absence of the analyte no peak will be produced.
- The evaluation should be representative of the conditions in which the analyses will be performed and should cover the scope of the method. This means that if method LODs and LOQs are under evaluation, the blanks should be submitted to the whole analytical procedure. In addition, if more matrices are taken into consideration, the standard deviation should be evaluated separately for each matrix. Furthermore, if blank correction or averaging will be performed, LODs and LOQs should reflect that.
- The evaluation of s_0 is typically done under repeatability conditions. A more reliable estimate can be carried out by evaluating s_0 under intermediate precision conditions (Section 2.2.3). In addition, enough independent determinations m should be carried out. Typically, during validation, $m = 10$ replicated measurements are considered sufficient for estimating s_0 .

After s_0 has been estimated, it can be utilized to estimate the LOD [40] according to Equation 2.43 (a 95% confidence level is used):

$$\text{LOD} = y_0 + 2t_{(0.05;\nu)} \frac{s_0}{\sqrt{n}}$$

Equation 2.43

Where y_0 is the average signal of several procedural blanks that have been measured (note that for spectroscopic techniques this number is the same as m as no fortification is required), n is the number of independent determinations *per* sample that will be averaged when reporting the results^{xxiii}, and t is the one-sided Student's t with $m - 1$ degrees of freedom. The factor $2t$, often approximated at 3, allows to claim that, if a signal greater than the LOD has been measured, the analyte has been detected with a 95% of confidence and a 95% of statistical power.

The LOQ is estimated in a similar way, but with a different factor, usually set to 10 (Equation 2.44):

$$\text{LOQ} = y_0 + 10 \frac{s_0}{\sqrt{n}}$$

Equation 2.44

LODs and LOQs calculated as above are expressed in terms of signal. The conversion in the concentration domain can be done through the calibration functions.

2.2.2. Linearity, significance, and validity

According to the International vocabulary of metrology^{xxiv} (i.e., the ISO/IEC Guide 99:2007), the *working range* is the range of concentrations in which the method itself can provide results with a suitable uncertainty. The lower boundary of the working range is represented by the LOQ, whereas the upper boundary is represented by a threshold above which variation in the concentration of the analyte does not produce any significant variations in the analytical signal. The definition is slightly different depending on whether the *instrument working range* or the *method working range* are taken into consideration: the first one refers to the capability of the instrument of producing a signal proportional to the concentration in a standard, whereas the second one refers to the capability of the method of producing a signal proportional to the concentration in a sample. Therefore, the method working range takes into consideration also any sample preparation step before the analysis.

The definition of the concentration interval to be explored during validation depends on the range of expected concentration of the analytes in the test samples. The EURACHEM guidelines [40] suggests evaluating this range by expanding the range of expected concentrations by 10% or 20% in each direction. Then,

^{xxiii} Note that if blank correction is necessary, the number of blanks that have been measured to make the correction must also be taken into consideration.

^{xxiv} <https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:99:ed-1:v2:en> (accessed 20/10/2022)

concentration levels (typically $k = 6-10$) are chosen so that they are going to be equally spaced across the range of interest and are measured in replicate (typically duplicate or triplicate). If these concentration levels refer to the concentration of a standard, then the working range under evaluation will be the instrumental one, whereas if matrix matched calibration is adopted (i.e., submitting to the whole analytical process fortified blank matrix), then the method working range will be evaluated.

Once the signals have been measured, the nature of the relationship between the signal and the concentration has to be established. This can be carried out by evaluating the regression statistics.

- **Linearity:** typically, the nature of the relationship between the signal and the concentration is either linear or quadratic, and high-order relationship are generally neglected. The concentration until which the relationship is linear, is called Limit of Linearity (LOL). Note that the linear range is always narrower than the working range. The nature of the relationship can be tested, for instance, with the Mandel's fitting test [43]. This inferential test is based on the calculation of the Fisher's F statistic (Equation 2.45), and it compares the difference between the variance that is not explained by a linear and quadratic model with the residual variance of a quadratic one. In other words: if the quadratic model does not account for more variability than the linear one, then a linear model is sufficient for describing the relationship at a given confidence level.

$$F = \frac{\Delta SS_{RES}}{MS_{RES}^Q} = \frac{SS_{RES}^L - SS_{RES}^Q}{MS_{RES}^Q}$$

Equation 2.45

Where SS_{RES}^L and SS_{RES}^Q are the sum of squared residuals for the linear model and for the quadratic model, respectively, and MS_{RES}^Q is the mean of squared residuals for the quadratic model. It must be noted that the R^2 is not an indication of linearity, as it just provides an indication about the fraction of variance that the model is able to explain [15]. R^2 always increases as the model become more and more complex. These consideration holds true also in the case of multivariate techniques.

- **Significance of the model:** the significance of the model is still based on Fisher's F statistic, and it provides information regarding whether there is a relationship between the signal and the concentration at a given confidence level. The F -test compares the variance captured by the

regression with the one that is not retained, i.e., the residual (Equation 2.46):

$$F = \frac{MS_{\text{REG}}}{MS_{\text{RES}}}$$

Equation 2.46

- **Validity:** also known as goodness-of-fit [15], the validity of the regression model provides information about the error deriving from approximation. In this case, the Fisher's F statistic involves the so-called variance of *lack-of-fit* (Equation 2.47) that can be derived from the sum of squares of the residuals and from the sum of squares of the pure experimental error.

$$F = \frac{MS_{\text{LOF}}}{MS_{\text{PE}}}$$

Equation 2.47

If the model is not affected by a significant lack-of-fit, it means that the error deriving from the model approximation can be explained in terms of pure experimental variability.

2.2.3. Precision

In a validation study, precision should be evaluated by performing replicated experiments to provide an estimate of the typical variability that could be expected within (and between) the results that are produced by a method. According to the ISO/IEC Guide 99:2007^{xxv} and the ISO 5725-5:1998^{xxvi}, precision refers to an estimate on the closeness of the results produced by a measurement under specific conditions. Standard deviation, relative standard deviation and interquartile range are typical measures of spread utilized in univariate statistics.

Depending on the measurement conditions, different levels of precision can be estimated. The measurement conditions must be specified when reporting the results of a precision study.

- **Repeatability:** s_r , it represents the typical variation that can be expected in the results when the measurement is performed in a short period of time, typically by the same operator in the same laboratory.

^{xxv} <https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:99:ed-1:v2:en> (accessed 21/10/2022)

^{xxvi} <https://www.iso.org/obp/ui/#iso:std:iso:5725:-5:ed-1:v1:en> (accessed 21/10/2022)

- **Reproducibility:** s_R , it represents the variation that can be expected in the results when the measurement is performed over a long period of time by different laboratories that use the same method, or even different ones.

Repeatability and reproducibility represent the two extremes of precision conditions. Between the two there is the so-called *intermediate precision* s_I that represents the variability that can be expected over a longer period within the same laboratory under more variable conditions than repeatability. Note that repeatability and intermediate precision can be estimated with the so-called *one laboratory validation*, whereas reproducibility requires *collaborative interlaboratory validation*, i.e., different laboratories working together in order to validate the same method.

In evaluating precision, the replicates should be independent and numerous enough to provide a reliable estimate of the standard deviation. Typically, 6–15 replicates are recommended [40]. In addition, it should be considered that for quantitative methods precision should be evaluated on different concentration levels over the working and/or linear range as precision may be dependent on concentration. For the same reason, generally the results of a precision study are expressed as relative standard deviation ($RSD\%$). Repeatability and intermediate precision can be studied separately or in a single study, which is convenient as it allows to reduce the number of experiments. This second strategy, described extensively in the ISO 5725-3^{xxvii}, involves performing small groups of independent measurements so that each group is measured under repeatability conditions and different groups are measured under intermediate precision conditions (e.g., on different days, with a different stock solution, by a different operator, etc.). The standard deviations for repeatability s_r and intermediate precision s_I can be derived by carrying out one-way Analysis of Variance (ANOVA):

- Repeatability can be derived directly from the variance within groups MS_W (Equation 2.48):

$$s_r = \sqrt{MS_W}$$

Equation 2.48

- Intermediate precision accounts both from the variability that can be expected between the means of the groups of measurements (MS_B) and within the same group (Equation 2.49):

^{xxvii} <https://www.iso.org/obp/ui/#iso:std:iso:5725:-3:ed-1:v1:en> (accessed 21/10/2022)

$$s_I = \sqrt{MS_W + \frac{MS_B - MS_W}{n}}$$

Equation 2.49

Where n is the number of measurements *per* group performed under repeatability conditions.

ANOVA also suggests whether the means of the groups are significantly different from each other. In the case there are significant differences between the groups, this experimental design does not allow to understand which factor is responsible from the variation that was observed under intermediate precision conditions.

2.2.4. Trueness

The *trueness* of a measurement reflects the closeness between the mean of an infinite number of results to the true population mean (also known as reference value or true value) [40]. The term *accuracy* is often improperly used as a synonym of trueness. The difference between the two terms lays in the definition of accuracy: according to the ISO/IEC Guide 99:2007^{xxviii}, accuracy represents how close the result of a single measurement is to the true population mean. In other words, accuracy accounts both from systematic deviations and random variation of a result:

- The entity of the random variations is accounted by estimating the precision of the method, as discussed earlier.
- The systematic deviation from the true population mean is accounted by trueness.

It must be noted that trueness cannot be estimated, as it is not possible to perform an infinite number of measurements. *Bias* is considered a suitable measurable estimate of trueness. Bias represents the deviation of the sample mean \bar{x} of the results from a suitable reference value x_{ref} .

Bias b can be expressed in absolute terms, as a percentage value with respect to x_{ref} , or as recovery rate $RR\%$ (Equation 2.50):

$$b = \bar{x} - x_{\text{ref}}$$

Equation 2.50a

^{xxviii} <https://www.iso.org/obp/ui/#iso:std:iso-iec:guide:99:ed-1:v2:en> (accessed 22/10/2022)

$$b_{\%} = \frac{\bar{x} - x_{\text{ref}}}{x_{\text{ref}}} \times 100$$

Equation 2.50b

$$RR_{\%} = \frac{\bar{x}}{x_{\text{ref}}} \times 100$$

Equation 2.50c

In the case the method is able to estimate exactly x_{ref} (meaning that $\bar{x} = x_{\text{ref}}$), the expected bias $b_{\%}$ will be 0% and the recovery rate $RR_{\%}$ will be 100%. Therefore, note that $b_{\%} < 0\%$ and $RR_{\%} > 100\%$ are allowed when the method underestimates and overestimates, respectively, the concentration of the reference x_{ref} . To estimate bias in this way the x_{ref} value must be known. This is the case, for instance, of certified reference materials. In this case, those are measured in independent replicates and the mean of the results is utilized to assess the bias. Nevertheless, such materials can be expensive and not always available for the intended scope of the method. Another option is that x_{ref} has been estimated with a different method in the same laboratory or by an independent laboratory.

Otherwise, spiking experiments can be carried out. In spiking, a blank matrix or a sample with known concentration of the analyte is fortified with a known concentration of analyte, denoted as x_{spike} . In this instance, bias is expressed as relative spike recovery rate $RR'_{\%}$ (Equation 2.51):

$$RR'_{\%} = \frac{\bar{x}' - \bar{x}}{x_{\text{spike}}} \times 100$$

Equation 2.51

Where \bar{x}' is the estimated concentration of the spiked sample with the method under validation and \bar{x} is the estimated concentration in the sample that necessarily has been determined with another method in the same laboratory or by an independent laboratory. Note that in the case of a spiked blank matrix $\bar{x} = 0$, therefore Equation 2.51 simplifies in Equation 2.52:

$$RR'_{\%} = \frac{\bar{x}'}{x_{\text{spike}}} \times 100$$

Equation 2.52

In either case, enough replicated measurements must be performed and $n = 10$ is the suggested number by the EURACHEM guidelines [40]. As for precision, in quantitative analysis bias should be estimated on multiple levels to check whether

the method is able to provide reliable results at different concentration levels. The results for bias can be reported as is, or by also reporting the confidence interval.

2.2.5. Enrichment capability

This quality parameter is typically not reported in guidelines for the validation of analytical methods, but its estimation can be important in the case the proposed method involves non-standard devices or techniques. The enrichment capability is a quality parameter that expresses the magnitude of the concentration factor of an analytical method, including any sample preparation step that is involved.

Enrichment Factors (EFs) provide an estimate of the enrichment capability the devised method has on the investigated analytes. EFs can be calculated as the ratio between the signal of the analyte obtained by submitting a sample (typically a spiked blank matrix at a known concentration level) to the whole analytical method ($S_{\text{processed}}$) and the signal of the analyte obtained by liquid injection of a solution at the theoretical concentration that would be reached if the method is unbiased (S_{standard}) [44,45].

EFs can be calculated according to Equation 2.53:

$$EF = \frac{S_{\text{processed}}}{S_{\text{standard}}}$$

Equation 2.53

Note of the author

This **Chapter** was meant to introduce the basic methodologies that were utilized for the experimental planning and analysis of the results in the following **Chapters**. In addition to that, it was intended to underline the importance of Chemometrics in the field of Material Science, allowing for the rational design of novel materials through Design of Experiments and Pattern Recognition, as well as optimizing procedures and processes in the context of sample preparation based both on standard and new materials.

References

- [1] H. Martens, Quantitative Big Data: where chemometrics can contribute, J Chemom. 29 (2015) 563–581. <https://doi.org/10.1002/CEM.2740>.
- [2] R. Kitchin, Big Data, new epistemologies and paradigm shifts, [Http://Dx.Doi.Org/10.1177/2053951714528481](http://Dx.Doi.Org/10.1177/2053951714528481). <https://doi.org/10.1177/2053951714528481>. 1 (2014).

- [3] B.R. Kowalski, Chemometrics: Views and Propositions, *J Chem Inf Comput Sci.* 15 (1975) 201–203. https://doi.org/10.1021/C160004A002/ASSET/C160004A002.FP.PNG_V03.
- [4] S. Wold, Chemometrics; what do we mean with it, and what do we want from it?, *Chemometrics and Intelligent Laboratory Systems.* 30 (1995) 109–115. [https://doi.org/10.1016/0169-7439\(95\)00042-9](https://doi.org/10.1016/0169-7439(95)00042-9).
- [5] K. Varmuza, P. Filzmoser, *Introduction to Multivariate Statistical Analysis in Chemometrics*, 1st Edition, CRC Press, Boca Raton, Florida, 2009. <https://doi.org/10.1201/9781420059496>.
- [6] J.W. Tukey, *Exploratory data analysis*, 1st Edition, Addison-Wesley Publishing Company, Reading, Massachusetts, 1977. <https://zbmath.org/0409.62003> (accessed September 19, 2022).
- [7] M. Li Vigni, C. Durante, M. Cocchi, *Exploratory Data Analysis*, in: *Data Handling in Science and Technology*, Elsevier Ltd, 2013: pp. 55–126. <https://doi.org/10.1016/B978-0-444-59528-7.00003-X>.
- [8] R. Bro, A.K. Smilde, Principal component analysis, *Analytical Methods.* 6 (2014) 2812–2831. <https://doi.org/10.1039/c3ay41907j>.
- [9] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems.* 2 (1987) 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- [10] S. Wold, A. Ruhe, H. Wold, I. W. J. Dunn, The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses, *SIAM Journal on Scientific and Statistical Computing.* 5 (1984) 735–743. <https://doi.org/10.1137/0905052>.
- [11] L.A. Berrueta, R.M. Alonso-Salces, K. Héberger, Supervised pattern recognition in food analysis, *J Chromatogr A.* 1158 (2007) 196–214. <https://doi.org/10.1016/J.CHROMA.2007.05.024>.
- [12] M. Cocchi, A. Biancolillo, F. Marini, *Chemometric Methods for Classification and Feature Selection*, in: *Comprehensive Analytical Chemistry*, Elsevier B.V., 2018: pp. 265–299. <https://doi.org/10.1016/bs.coac.2018.08.006>.
- [13] P. Oliveri, C. Malegori, E. Mustorgi, M. Casale, Application of Chemometrics in the Food Sciences, *Comprehensive Chemometrics.* (2020) 99–111. <https://doi.org/10.1016/B978-0-12-409547-2.14748-1>.
- [14] P. Oliveri, C. Malegori, E. Mustorgi, M. Casale, Qualitative pattern recognition in chemistry: Theoretical background and practical guidelines, *Microchemical Journal.* 162 (2021) 105725. <https://doi.org/10.1016/J.MICROC.2020.105725>.
- [15] A.M. Committee, Uses (proper and improper) of correlation coefficients, *Analyst.* 113 (1988) 1469–1471. <https://doi.org/10.1039/AN9881301469>.
- [16] F. Westad, F. Marini, Validation of chemometric models - A tutorial, *Anal Chim Acta.* 893 (2015) 14–24. <https://doi.org/10.1016/j.aca.2015.06.056>.
- [17] R.W. Kennard, L.A. Stone, Computer Aided Design of Experiments, *Technometrics.* 11 (1969) 137–148. <https://doi.org/10.1080/00401706.1969.10490666>.
- [18] R.D. Snee, Validation of Regression Models: Methods and Examples, *Technometrics.* 19 (1977) 415–428. <https://doi.org/10.1080/00401706.1977.10489581>.
- [19] J. Djuris, S. Ibric, Z. Djuric, Chemometric methods application in pharmaceutical products and processes analysis and control, *Computer-Aided Applications in Pharmaceutical Technology.* (2013) 57–90. <https://doi.org/10.1533/9781908818324.57>.
- [20] G. Hanrahan, F. Udeh, D.G. Patil, CHEMOMETRICS AND STATISTICS | Multivariate Calibration Techniques, *Encyclopedia of Analytical Science: Second Edition.* (2005) 27–32. <https://doi.org/10.1016/B0-12-369397-7/00077-7>.
- [21] A.J. Burnham, J.F. MacGregor, R. Viveros, Latent variable multivariate regression modeling, *Chemometrics and Intelligent Laboratory Systems.* 48 (1999) 167–180. [https://doi.org/10.1016/S0169-7439\(99\)00018-0](https://doi.org/10.1016/S0169-7439(99)00018-0).

- [22] R.X. Liu, J. Kuang, Q. Gong, X.L. Hou, Principal component regression analysis with spss, *Comput Methods Programs Biomed.* 71 (2003) 141–147. [https://doi.org/10.1016/S0169-2607\(02\)00058-5](https://doi.org/10.1016/S0169-2607(02)00058-5).
- [23] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: A basic tool of chemometrics, in: *Chemometrics and Intelligent Laboratory Systems*, Elsevier, 2001: pp. 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [24] U.G. Indahl, The geometry of PLS1 explained properly: 10 key notes on mathematical properties of and some alternative algorithmic approaches to PLS1 modelling, *J Chemom.* 28 (2014) 168–180. <https://doi.org/10.1002/cem.2589>.
- [25] S. Wold, H. Martens, H. Wold, The multivariate calibration problem in chemistry solved by the PLS method, in: Springer, Berlin, Heidelberg, 1983: pp. 286–293. <https://doi.org/10.1007/bfb0062108>.
- [26] S. de Jong, SIMPLS: An alternative approach to partial least squares regression, *Chemometrics and Intelligent Laboratory Systems.* 18 (1993) 251–263. [https://doi.org/10.1016/0169-7439\(93\)85002-X](https://doi.org/10.1016/0169-7439(93)85002-X).
- [27] S. Wold, E. Johansson, M. Cocchi, PLS: partial least squares projections to latent structures, in: H. Kubinyi (Ed.), *3D QSAR in Drug Design*, ESCOM Science Publishers, Leiden, Netherlands, 1993: pp. 523–550. <https://iris.unimore.it/handle/11380/607273> (accessed October 3, 2022).
- [28] S. Favilla, C. Durante, M.L. Vigni, M. Cocchi, Assessing feature relevance in NPLS models by VIP, *Chemometrics and Intelligent Laboratory Systems.* 129 (2013) 76–86. <https://doi.org/10.1016/J.CHEMOLAB.2013.05.013>.
- [29] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA?, *BMC Bioinformatics.* 21 (2020) 2. <https://doi.org/10.1186/s12859-019-3310-7>.
- [30] R.A. Fisher, *The Design of Experiments*, 9th ed., Macmillan, London, UK, 1971.
- [31] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, A. Nystrom, J. Pettersen, R. Bergman, Experimental Design and Optimization, *Chemometrics and Intelligent Laboratory Systems.* 42 (1998) 3–40. <https://doi.org/10.1007/978-3-540-49148-4-3>.
- [32] B. Benedetti, V. Caponigro, F. Ardini, *Experimental Design Step by Step: A Practical Guide for Beginners*, <https://doi.org/10.1080/10408347.2020.1848517>. 52 (2020) 1015–1028. <https://doi.org/10.1080/10408347.2020.1848517>.
- [33] R. Leardi, Experimental design in chemistry: A tutorial, *Anal Chim Acta.* 652 (2009) 161–172. <https://doi.org/10.1016/J.ACA.2009.06.015>.
- [34] L. Vera Candioti, M.M. de Zan, M.S. Cámara, H.C. Goicoechea, Experimental design and multiple response optimization. Using the desirability function in analytical methods development, *Talanta.* 124 (2014) 123–138. <https://doi.org/10.1016/j.talanta.2014.01.034>.
- [35] F. Dong, On the identification of active constraints in unreplicated fractional factorials, *Statistica Sinica.* 3 (1993) 209–217.
- [36] R. Leardi, D-Optimal Designs, *Encyclopedia of Analytical Chemistry.* (2018) 1–11. <https://doi.org/10.1002/9780470027318.A9646>.
- [37] J.J. Borkowski, Center Points, *Encyclopedia of Statistics in Quality and Reliability.* (2008). <https://doi.org/10.1002/9780470061572.EQR030>.
- [38] M.A. Bezerra, V.A. Lemos, C.G. Novaes, R.M. de Jesus, H.R.S. Filho, S.A. Araújo, J.P.S. Alves, Application of mixture design in analytical chemistry, *Microchemical Journal.* 152 (2020) 104336. <https://doi.org/10.1016/J.MICROC.2019.104336>.
- [39] G. Derringer, R. Suich, Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology.* 12 (1980) 214–219. <https://doi.org/10.1080/00224065.1980.11980968>.

- [40] B. Magnusson, U. Örnemark, eds., Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topic, 2nd ed., 2014. [https://doi.org/10.1016/S0014-2999\(99\)00500-2](https://doi.org/10.1016/S0014-2999(99)00500-2).
- [41] M. Valcárcel, G.D. Christian, R. Lucena, Teaching social responsibility in analytical chemistry, *Anal Chem.* 85 (2013) 6152–6161. https://doi.org/10.1021/AC400323M/ASSET/IMAGES/LARGE/AC-2013-00323M_0007.JPEG.
- [42] P. Konieczka, Validation and Regulatory Issues for Sample Preparation, *Comprehensive Sampling and Sample Preparation: Analytical Techniques for Scientists.* (2012) 699–711. <https://doi.org/10.1016/B978-0-12-381373-2.00064-8>.
- [43] J.M. Andrade, M.P. Gómez-Carracedo, Notes on the use of Mandel’s test to check for nonlinearity in laboratory calibrations, *Analytical Methods.* 5 (2013) 1145–1149. <https://doi.org/10.1039/C2AY26400E>.
- [44] A. Arcoleo, F. Bianchi, M. Careri, A sensitive microextraction by packed sorbent-gas chromatography-mass spectrometry method for the assessment of polycyclic aromatic hydrocarbons contamination in Antarctic surface snow, *Chemosphere.* 282 (2021) 131082. <https://doi.org/10.1016/J.CHEMOSPHERE.2021.131082>.
- [45] N. Riboni, J.W. Trzcinski, F. Bianchi, C. Massera, R. Pinalli, L. Sidisky, E. Dalcanale, M. Careri, Conformationally blocked quinoxaline cavitand as solid-phase microextraction coating for the selective detection of BTEX in air, *Anal Chim Acta.* 905 (2016) 79–84. <https://doi.org/10.1016/j.aca.2015.12.005>.

Chapter 3 | Evaluating the potential of carbon nanotubes for the solid-phase microextraction of personal care products from water samples

3.1. Introduction

3.1.1. Emerging contaminants

Emerging contaminants (ECs) comprise a pool of chemicals (either synthetic or natural-identical, including their metabolites) and biological entities that could damage the environment and ecosystems, but whose presence in environmental matrices is usually not monitored [1,2]. It has been suggested that municipal wastewater can be considered as the primary source of ECs; other sources include grey- and blackwaters from households and industries, and effluents of sewage treatment facilities. All these sources are known also as point-pollution sources [3]. Other diffuse-pollution sources include atmospheric deposition and wastes from farms and livestock [4].

Currently, an international monitoring program on ECs has not been proposed yet for many reasons [2,3]:

- Their impact on the ecosystems and toxicological behavior are still a matter of debate and, often, not well understood.
- There are limited data about their fate once they enter the environment.
- The number of chemicals/biological entities that can be considered as ECs is constantly increasing. This could happen, e.g., with the incorporation of a new chemical in a commercial formulation, or with the diffusion of a new infectious disease.
- The discovery of these substances was possible only thanks to the advancements in analytical sciences. In fact, the concentration of ECs in environmental matrices ranges from $\mu\text{g/L}$ to ng/L .

Given the abovementioned considerations, it can be stated that the fraction of regulated pollutants that are conventionally monitored in water bodies and other environmental matrices represents only a minor fraction of the total number of potential pollutants [5]. In addition, the lack of knowledge in regard of the

ecotoxicological effects and potential adverse consequences on the human health makes ECs a seemingly unsurmountable issue.

In 2005, a project chartered by the European Commission known as the NORMAN project^{xxix} was established with the aim of generating knowledge regarding ECs, as well as promoting the harmonization and validation of analytical methodologies for the identification of new ECs and their quantitation in environmental matrices [2,6]. Thus, regulating agencies could promptly manage and communicate risks related to such substances. The contributions to the NORMAN project are made possible thanks to an interlaboratory network comprising universities, industries, regulating agents and research centers across 20 countries [6]. The information that has been gathered in the framework of the NORMAN project has been made available in the NORMAN SusDat^{xxx}, one of the largest databases about ECs.

Up to now, the NORMAN SusDat accounts of more than one hundred thousand entries, including ECs and their metabolites, classified in more than twenty categories depending on their origin. Figure 3.1 reports few examples of such classes.

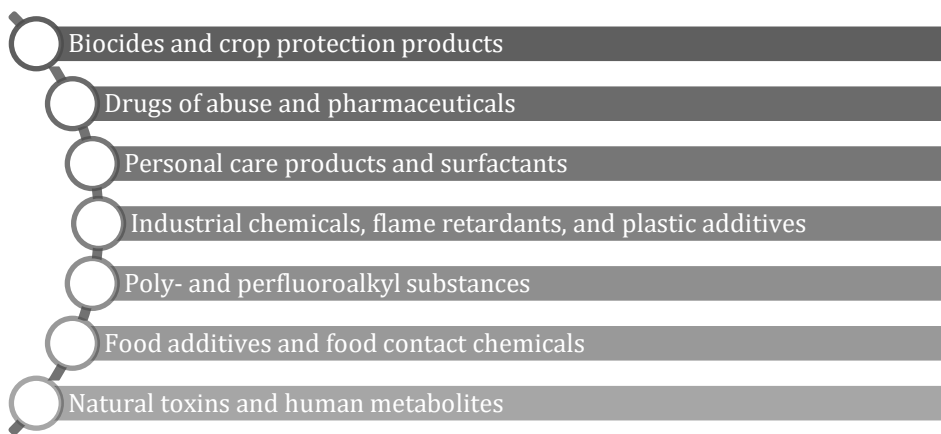


Figure 3.1. Classes of ECs according to the NORMAN SusDat.

Other examples worth to mention include disinfection by-products, siloxanes, pesticides, and nanomaterials [7–9].

3.1.2. Personal care products

Personal Care Products (PCPs) is an umbrella term enclosing active and functional ingredients that can be found in a variety of commodities and in every household

^{xxix} <https://www.norman-network.net/> [accessed 19/08/2022]

^{xxx} <https://www.norman-network.com/nds/susdat/> [accessed 19/08/2022]

including cleaning products, soaps, cosmetics, detergents, disinfectants, as well as veterinary products and goods used for pest control [2,10]. Such formulations are utilized and disposed daily down the drains of every household and workplace.

PCPs represent the vast majority of the ECs present in the environment. Due to their complex molecular structure, most PCPs resist to most of the unitary operations in sewage and wastewater treatment facilities and can access the environment intact or only partially degraded. In fact, detectable amounts of ECs have been found in effluents of water treatment facilities, as well as in groundwater and surface water [11,12], in soil and sediments [13], and also in drinking water [14,15].

Given the abovementioned considerations, the presence of PCPs and their metabolites in water poses a risk to human health and environmental safety as they can go back into the water cycle thus persisting into the environment. In addition, their diverse chemical structure makes them able to access different biochemical pathways, and, up till now, there is limited information available regarding their combined toxic effects [2]. Moreover, few PCPs are suspected to cause sensitization, allergies, and to disrupt the endocrine system [16].

The main sub-classes of PCPs are reported below:

- **Surfactants:** surfactants represent the major ingredient in cleaning products, including detergents, soaps, and body and face cleansers. Surfactants are amphiphilic molecules that lower the surface tension of water and increases the wettability of surfaces so that fat-based substances can more easily detach [17]. The environmental risks connected to surfactants are mostly under control since most surfactants are biodegradable and/or can be mineralized by the unitary operations in water treatment facilities [18].
- **UV filters:** these substances are mainly found in sun protection products and in skincare formulations, and their function is to reduce the amount of UV radiation that reaches the skin. UV filters can be grouped in two main classes, i.e., physical and chemical filters [19]. Physical filters are inorganic compounds, like titanium dioxide, able to backscatter the UV radiation. Chemical filters are organic compounds with high molar attenuation coefficients in the UV region of the electromagnetic spectrum. Most of chemical filters resist biodegradation and their ecotoxicological profile suggests that acute effects on the environment can be observed at concentration around 1 mg/L [18]. Besides the effects on the environment,

awareness towards this class of compounds has increased as evidence suggested that certain chemical UV filters can act as endocrine disruptors [18,20].

- **Antimicrobial agents and preservatives:** this class of ingredients comprises substances with different effects depending on their concentration. For products specifically intended for disinfection, antimicrobial agents can represent the major fraction of the formulation, whereas if they have a functional role in a formula, they represent a minor fraction. The functional role of preservatives and antimicrobial agents is to extend the shelf-life of the formulation, preventing microbial growth and oxidative degradation processes [10]. The most known preservatives and antimicrobial agents that can be found in formulations are triclosan, parabens and butylated hydroxytoluene. Most of these substances are readily biodegradable, so they do not constitute a hazard for the environment [18]. Triclosan resists biodegradation and is currently under evaluation as it is suspected as a bioaccumulating substance that can have adverse effects on the endocrine system^{xxxi}.
- **Pest repellents:** this class comprises substances involved in regulating the growth and/or eliminating unwanted organisms [21,22]. Pest repellents are biocides, so they are policed by the Regulation EU 528/2012^{xxxii}. The major fraction of pesticides found into the environment derives from agriculture and aquaculture, but a minor fraction could derive also from domestic activities, like gardening or from the use of insect-repellent lotions. The environmental concerns related to these substances are related to their persistence and capability to bioaccumulate along the food chain [23].
- **Fragrances:** fragrances are volatile organic compounds widely used in the manufacturing of cleansers, house-cleaning products, cosmetics, and perfumes [24,25]. Over 3000 synthetic and natural-identical substances are used for this purpose. For what concerns cosmetics, the use of fragrances and other ingredients must follow the Regulation EU 1223/2009^{xxxiii}. Many substances used as fragrance agents have been reported in having adverse effects on the human health, including sensitization, allergic reaction, as well as contact dermatitis and disrupting

^{xxxi} <https://echa.europa.eu/it/brief-profile/-/briefprofile/100.020.167> [accessed 24/08/2022]

^{xxxii} <http://data.europa.eu/eli/reg/2012/528/oj> [accessed 24/08/2022]

^{xxxiii} <http://data.europa.eu/eli/reg/2009/1223/2022-07-31> [accessed 24/08/2022]

effects towards the endocrine system [16,25]. Since such compounds are widely used in everyday commodities in 2012 the Scientific Committee on Consumer Safety declared 82 fragrances as contact allergens for the safeguard of consumers [25]. Many fragrances are persistent into the environment and resist both chemical and biological processes in water treatment facilities and, as bioaccumulation phenomena concerning such substances have been reported [16,24,25], they constitute a threat for human health and the safety of the environment.

3.1.3. Solid-phase microextraction

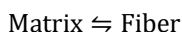
As mentioned earlier, since ECs and, therefore, PCPs can be present in the environment at trace levels [2,3], proper sample preparation techniques are required for an effective clean-up and preconcentration of these analytes [26].

No official analytical methodologies are prescribed for the identification and quantitation of ECs in environmental matrices. To be mindful about the environmental impact of analytical chemistry, the development and validation of methods of analysis should be carried out in accordance with the principles of Green Analytical Chemistry [27]. The combination of multiple steps of the analytical process into a single one, the reduction of sample size and the limited use of solvents are key concept to be considered in the development of analytical methods with a reduced impact on the environment. Miniaturization has become arguably the response to fulfill the requirements of Green Analytical Chemistry [28].

Solid-phase microextraction (SPME) was formerly introduced by Janusz Pawliszyn and coworkers in 1989 [29,30]. SPME is a solventless sample preparation technique, that combines extraction, clean-up, and preconcentration in a single step. In SPME (Figure 3.2) a narrow silica or metal needle is covered with an appropriate coating material that actively participates to the extraction process. This abovementioned device is called “fiber”.

SPME can be operated in:

- **Direct immersion mode (DI):** the fiber is submerged into the liquid sample. The equilibrium involved in the extraction process is:



Equation 3.1

The DI mode is mandatory with nonvolatile analytes.

- **Headspace mode (HS):** the fiber is exposed to the vapors generated by a solid or a liquid sample, without a direct contact with it. The equilibrium involved in the extraction process is:



Equation 3.2

HS mode is particularly suitable for the extraction of volatile analytes. Since no contact with the matrix is required, HS sampling also allows the clean-up as nonvolatile interfering compounds are not collected during the extraction process.

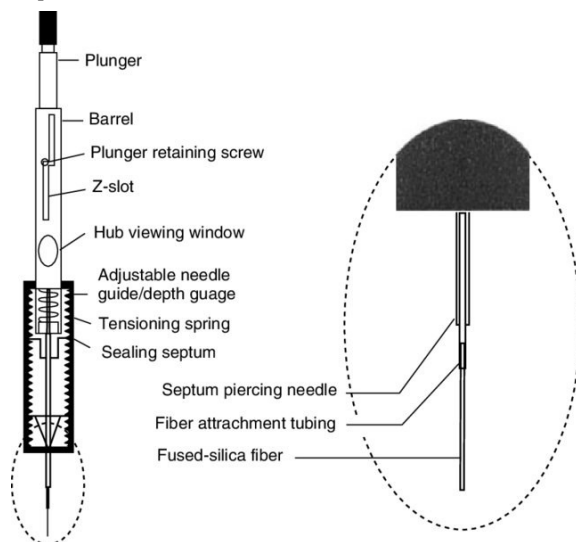


Figure 3.2. Schematic representation of a SPME manual sampling device. Reprinted with permission from [31].

During the extraction, analytes interact with the coating material and concentrate on it. After extraction, the analytes are detached from the coating by placing the fiber into the injection port of a gas chromatograph for thermal desorption or into a proper interface of a liquid chromatograph for solvent desorption.

The extraction mechanism is different depending on the nature of the coating. Two main mechanisms can be identified:

- **Partitioning:** in partitioning, the coating behaves like a bonded liquid phase, solubilizing the analytes. The loading capacity of such fibers depends only on the thickness of the coating. Polydimethylsiloxane (PDMS) is an example of coating material whose extraction mechanism is partitioning.

- **Adsorption:** the coating is constituted by a high-porous solid that non-covalently interacts with the analytes. Adsorptive coatings are typically characterized by lower loading capacities than coatings whose extraction mechanism is partitioning. In fact, once the active sites have been saturated, no more analytes can be adsorbed. Carboxen™ is a carbon-based coating material whose extraction mechanism is adsorption.

There are different fibers available on the market, with different coatings, thickness, and length, suitable for the vast majority of routine applications. Since SPME have become popular, important innovations have been made thanks to the advancements in the field of Materials Science. In fact, SPME devices with a plethora of new coatings have been proposed to face more difficult analytical challenges [32,33], like the detection of analytes at trace and ultra-trace levels in complex matrices. Novel materials that have been proposed as SPME coating include ionic liquids [34], Metal-Organic Frameworks [35], carbon nanotubes [36], and molecularly imprinted polymers [37].

3.1.4. Carbon nanotubes

Carbon nanotubes (CNTs) are an allotropic state of carbon, serendipitously discovered by Sumio Iijima in 1991 [38]. CNTs are constituted by a single or multiple layers of graphene rolled up into a tubular shape whose length is in the order of tens of micrometers. Depending on how many concentric graphene tubes are in the structure [39], two main categories of CNTs can be identified (Figure 3.3):

- **Single-walled carbon nanotubes (SWCNTs):** they are constituted by a single sheet of graphene rolled on itself. Depending on the direction of rolling, different configurations can be produced. SWCNTs have a diameter around 1 nm.
- **Multi-walled carbon nanotubes (MWCNTs):** they are constituted by multiple layers of graphene rolled up together. MWCNTs can reach diameters in the order of few hundreds of nanometers and the graphene layers are coaxial, typically 0.3–0.4 nm apart from each other.

CNTs have a large surface area and therefore they have potential as sorbent materials. CNTs have been proposed as sorbents both in sample preparation [39,40] and in the context of sewage treatment, to scavenge both organic and inorganic contaminants [41,42].

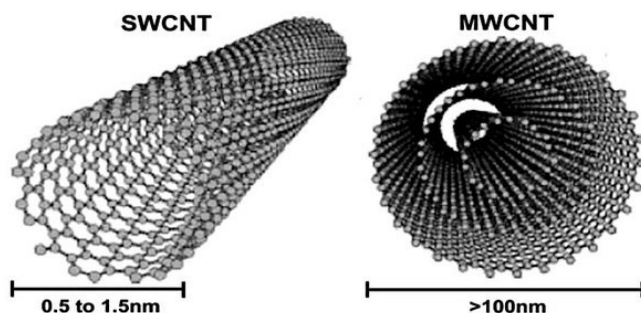



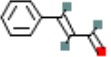
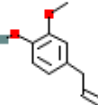

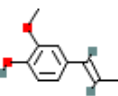
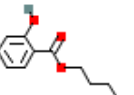
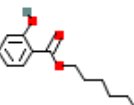
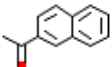
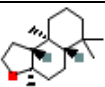
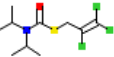
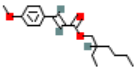
Figure 3.3. Schematic representation of a SWCNT (left) and a MWCNT (right). Reprinted with permission from [43].

CNTs can establish interaction with other molecules mainly thanks to the delocalized electrons on their entire surface. Such interactions are π - π stacking, CH- π and π -cation interactions, and London dispersion forces. In addition, dipole-dipole interactions and hydrogen bonding can be exploited after proper modification of CNTs with the introduction of polar functional groups [39,40].

In the framework of a research activity devoted to the development of miniaturized extraction techniques, this study aimed at evaluating the extraction capability of various CNTs as SPME coating for the determination of PCPs in water samples. A pool of 12 PCPs were selected as model compounds (Table 3.1^{xxxiv}), including the substances commonly found in the environment [12,44–46]. The SPME procedure was optimized by evaluating the effect of four different types of CNTs in a first step, and the effect of extraction time and temperature thereafter. The validated method was successfully utilized for the determination of PCPs in surface water samples, collected in various water bodies in Northern Italy.

^{xxxiv} Molecular structures were retrieved from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>, accessed 29/08/2022) and the properties of concern from the European Chemical Agency (<https://echa.europa.eu/it/home>, accessed 29/08/2022)

Table 3.1. Overview of the analytes included in this study.

Analyte	Molecular structure ^a	PCP category	Properties of concern in EU ^a
(+)-limonene		Fragrance	Ss PBT
Cinnamaldehyde		Fragrance	Ss
Eugenol		Fragrance	Ss
Butylated hydroxytoluene		Antioxidant	<u>ED</u>
<i>i</i> -eugenol		Fragrance	Ss
Amyl salicylate		Fragrance UV filter	n.n.
Hexyl salicylate		Fragrance UV filter	Ss
2'-acetonaphthone		Fragrance	n.n.
(-)-ambroxide		Fragrance	n.n.
Triallate		Plant protection product	Ss
2-ethylhexyl-4-methoxycinnamate		UV filter	n.n.

^a ■ hydrogen; ■ oxygen; ■ nitrogen; ■ sulfur; ■ chlorine. ^b **bold**: recognized property; plain: broad agreement about the property; underlined: property under evaluation; grey: property reported in a minority of notifications; Ss: skin sensitizing; PBT: persistent, bioaccumulative, and toxic; ED: endocrine disruptor; n.n.: not notified.

3.2. Materials and methods

3.2.1. Chemicals and materials

Methyl alcohol (> 99%) and HF (48% in H₂O) were obtained from Merck (Milan, Italy). NaCl (> 99.99%) was obtained from VWR Chemicals (Milan, Italy).

Amyl salicylate (AmS), 2-ethylhexyl-4-methoxycinnamate (EHMC), triallate (Tri, all analytical grade), butylated hydroxytoluene (BHT), cinnamaldehyde (Cin), eugenol (Eug), hexyl salicylate (HexS, all > 99%), isoeugenol (iEug, mixture of *cis-/trans*-isomers, > 98%), and perdeuterated naphthalene (\geq 98%; deuteration grade > 99%) were obtained from Merck. 2'-acetonaphthone (2AcN, > 99%) was obtained from Acros Organics (Milan, Italy). Benzyl salicylate (BzS, > 99%) was obtained from Alfa Esar (Kandel, Germany). (-)-ambroxide (Amb, > 98%) was obtained from TCI Chemicals (Zwijndrecht, Netherlands). (+)-limonene (Lim, > 98%) was obtained from Fluka (Seelze, Germany).

Helical multi-walled carbon nanotubes (HMWCNTs; average diameter: 100–200 nm; length: 10–30 μ m), multi-walled carbon nanotubes with an average diameter of 50 nm (MWCNTs50; length: 10–20 μ m), multi-walled carbon nanotubes with an average diameter < 8 nm (MWCNTs8; length: 10–30 μ m), and carboxylated graphitized multi-walled carbon nanotubes (MWCNTs–COOH; average diameter: 50 nm; length: 10–20 μ m) were obtained from Cheap Tubes Inc. (Massachusetts, USA). Bare fused silica SPME fibers (length: 1 cm) and 50/30 μ m DVB/Carboxen™-PDMS fibers (length: 1 cm) were obtained from Supelco (Bellefonte, USA). Duralco 4460 epoxy glue was obtained from Cotronics Corp. (Brooklyn, USA).

Milli-Q water was produced by means of a Millipore Milli-Q Element A10 water purification system (Merck-Millipore, Milan, Italy).

Working solutions were prepared by proper dilution of the analytes and N-d₈ in methanol. Diluted solutions were kept at 4 °C in the dark until use.

3.2.2. Preparation and characterization of the fibers

3.2.2.1. Fiber fabrication

The CNTs-coated fibers were fabricated by following a dipping procedure [40]. Bare fused silica fibers were submerged into an aqueous HF solution (40% v/v) for 10 s and then washed several times in Milli-Q water. The etched fibers were let dry into a desiccator overnight.

Once dried, the fibers were vertically dipped into the Duralco 4460 epoxy glue and, after 2 min into the CNT powder for three times. The coatings were let harden overnight. For each coating (HMWCNTs, MWCNTs50, MWCNTs8, and MWCNTs-COOH) three fibers were fabricated.

Each fiber was conditioned in the GC injection port at 320 °C for 1 h under a constant flow of helium.

3.2.2.2. Fiber characterization

Morphological investigation was carried out by using a Quanta FEG 250 (Thermo Fisher Scientific, Waltham, USA) Environmental Scanning Electron Microscope (ESEM) equipped with a XFlash 6 | 30 (Bruker, Billerica, USA) detector.

Thermal stability was assessed *via* Thermo-Gravimetric Analysis (TGA) by heating at 10 °C/min under nitrogen atmosphere in the 40–400 °C range. The investigation was carried out on a TGA 7 instrument by PerkinElmer (Waltham, USA). Fiber bleeding was investigated by placing the fibers in the GC injection port at 320 °C for 2 min.

3.2.3. Optimization of the extraction procedure

Optimization of the SPME procedure was carried out by performing extraction experiments on 19.5 mL of aqueous solutions containing the analytes. SPME was operated in Direct Immersion mode (DI). The response variable was the peak area of each analyte.

3.2.3.1. Preliminary evaluation of the coating performance

The performance of each CNTs-based coating was tested by carrying out extraction experiments on aqueous samples at the concentration of 2 µg/L of each analyte. Each sample was equilibrated for 10 min at 45 °C. Thereafter, the fiber was immersed into the sample solution for 30 min at 45 °C under constant agitation (250 rpm). Then, the fiber was placed in the GC injection port at 320 °C for 2 min. For each fiber, three independent replicated extractions were performed ($N = 3$ fibers *per* coating, $n = 3$ extractions *per* fiber).

Differences in the average responses for each coating were highlighted with one-way ANOVA. Homoscedasticity and normality were tested beforehand with Bartlett's test and Shapiro-Wilk test, respectively. For statistically significant results, the effect size was estimated in terms of Cohen's η^2 [47] and, *post-hoc* multiple pairwise comparisons were carried out with Student's *t*-tests with a Bonferroni correction. The confidence level was 95%.

3.2.3.2. Optimization

Two main factors ($k = 2$) were investigated at three levels within their respective experimental domain according to a Box-Wilson Central Composite Design (CCD): extraction temperature (X_1 , 30–60 °C) and extraction time (X_2 , 15–45 min). In total, $N = 12$ experiments were carried out, including $n_0 = 4$ experiments in the center of the experimental domain to estimate the pure experimental variance.

A full second-order model was postulated. The significance of the regression coefficients was evaluated with a backward-stepwise variable selection algorithm ($\alpha_{\text{to remove}} = 0.05$). Each model was evaluated in terms of fraction of explained variance (R^2) and validity was evaluated by carrying out a lack-of-fit F -test ($\alpha = 0.05$).

The global optimal conditions were identified according to Derringer's method [48]. Single desirability functions d_i were defined for each analyte according to Equation 3.1:

$$d_i = \begin{cases} 0, & y_i < 0 \\ \frac{y_i}{1.2 \times y_i^{\text{MAX}}}, & 0 \leq y_i \leq 1.2 \times y_i^{\text{MAX}} \\ 1, & y_i > 1.2 \times y_i^{\text{MAX}} \end{cases}$$

Equation 3.1

y_i^{MAX} is the maximum predicted response within the experimental domain. The optimal conditions were in correspondence of the maximum global desirability (Equation 2.42), computed as unweighted geometric mean of the single desirability functions. The maximum was found with a derivative-free search algorithm.

3.2.4. Operating procedure and instrumental conditions

3.2.4.1. Sample preparation

A volume of 19.5 mL of sample was introduced into 20 mL amber glass vials. The aliquot was spiked with the internal standard (IS) solution so that its final concentration was 250 ng/L. The IS was N-d₈.

Each sample was equilibrated for 10 min at 60 °C. Thereafter, the HMWCNTs-based fiber was immersed into the sample solution for 45 min at 60 °C under constant agitation (250 rpm). Then, the fiber was placed in the GC injection port at 320 °C for 2 min. These operations were carried out with the aid of a PAL COMBI-xt autosampler (CTC Analytics AG, Zwingen, Switzerland).

3.2.4.2. GC-MS

Instrumental analyses were carried out with a HP 6890 Series Plus gas chromatograph hyphenated with an MSD 5973 mass spectrometer (both by Agilent Technologies, Milan, Italy).

The split/splitless injector temperature was 320 °C. The injection was executed in splitless mode into a 5190-4056 ultra-inert liner (Agilent Technologies). An Rxi-17Sil MS capillary column (30 m length × 0.25 mm i.d., 0.25 μm film thickness; Restek, Bellefonte, USA) was used for the chromatographic separation with the following temperature program: 60 °C, 15 °C/min to 120 °C, 3 °C/min to 180 °C, 20 °C/min to 280 °C, 5 °C/min to 290 °C, held for 1.00 min (runtime: 32.00 min). The carrier gas was helium (constant flow rate: 1.2 mL/min). The transfer line was maintained at 280 °C.

The single quadrupole mass spectrometer was operated in electron ionization mode (EI; 70 eV). The ion source and the quadrupole were held at 230 and 150 °C, respectively. Mass spectra were registered in selected ion monitoring (SIM) mode (dwell time: 30 ms; electron multiplier voltage: 2.470 kV). The acquisition started 2.00 min after the injection. One microliter of a solution containing both the analytes and the IS (concentration: 1 mg/L) in methyl alcohol was injected and the mass range 50–350 *m/z* was recorded to properly set the time scheduled SIM program. The retention times and the monitored *m/z* ratios for each compound are reported in Table 3.2.

Table 3.2. Retention times and monitored ions of the compounds involved in the present study. Qualifier ions are reported in plain text, quantifier ions are reported in bold text.

Compound	Retention time (min)	Monitored <i>m/z</i> ratios
Lim	3.57	68 , 93
N-d ₈	7.05	108, 136
Cin	9.71	103, 131
Eug	10.56	149, 164
BHT	12.88	205 , 220
iEug	13.63	131, 164
AmS	15.59	120 , 208
HexS	18.61	120 , 222
2AcN	20.97	127, 155
Amb	21.35	137, 221
Tri	24.77	86 , 268
BzS	26.51	91 , 228
EHMC	29.29	161, 178

3.2.5. Method validation

The SPME–GC–MS method was validated by working under the optimized conditions according to the EURACHEM guidelines [49]. The validation study was carried out with matrix-matched standards that were subjected to the whole analytical process. Tap water was used as blank matrix.

Limits of detection and limits of quantitation (LODs and LOQs, respectively) were evaluated by estimating the standard deviation s_0 of $m = 10$ independent replicated measurements of blank matrix fortified with a detectable amount of analytes. The average signal of the blank y_0 was estimated by submitting to analysis $n = 5$ procedural blanks. Detection and quantitation limits, expressed in the signal domain, were calculated according to Equation 2.43 and 2.44, respectively, considering that $n = 3$ replicated measurements will be averaged when reporting the results. The corresponding concentrations were calculated by their projection on the x axis. Calibration functions were evaluated on $k = 6$ levels exploring two orders of magnitude from the LOQ ($n = 3$ independent replicated measurements *per* level) for all the analytes. The intercept was tested for significance with the Student's t -test ($\alpha = 0.05$). Linearity and validity were evaluated, by applying Mandel's test and lack-of-fit test ($\alpha = 0.01$ and $\alpha = 0.05$, respectively).

Repeatability and intermediate precision were evaluated in a single study by carrying out $n = 6$ independent measurements *per* day for three days by using a different fiber each day. Repeatability and intermediate precision standard deviations were estimated according to Equation 2.48 and 2.49 and expressed as relative standard deviations ($RSD\%$). ANOVA was applied to assess whether the results obtained under intermediate precision conditions were significantly different among each other ($\alpha = 0.05$). Trueness was expressed as spike recovery rate ($RR'\%$, Equation 2.52) by submitting $n = 10$ independent replicated measurements to the whole analytical process. Precision and bias were evaluated on three levels for each analyte (Table 3.3).

Enrichment capabilities of the HMWCNTs-based coating were evaluated in terms of enrichment factors (EFs). EFs were calculated according to Equation 2.53 by submitting to SPME aqueous solutions spiked with each analyte at 200 ng/L and by injecting 1 μ L of a solution containing all the analytes in methyl alcohol at 50 μ g/L. The EFs of the proposed coating were compared with the ones obtained by using the DVB/Carboxen™-PDMS fiber ($n = 3$ independent replicated measurements). For the latter, a desorption temperature of 250 °C was used. The responses were

adjusted by the corresponding response factor as they were carried out at different concentration levels.

Table 3.3. Concentration levels for the precision and bias studies. All the values are reported in ng/L.

Analyte	Precision			Trueness		
	L1^a	L2	L3	L1^a	L2	L3
Lim	3.2	250	1800	3.2	500	1500
Cin	42.3	500	5000	42.3	1000	3000
Eug	20.6	500	5000	20.6	1000	3000
BHT	2.9	125	900	2.9	250	750
iEug	20.1	500	5000	20.1	1000	3000
AmS	4.1	125	900	4.1	250	750
HexS	4.4	50	450	4.4	100	350
2AcN	4.6	500	1800	4.6	750	1500
Amb	4.1	250	1800	4.1	500	1500
Tri	7.8	250	1800	7.8	500	1500
BzS	0.7	50	450	0.7	100	350
EHMC	3.0	50	450	3.0	100	350

^a evaluation carried out at the LOQ.

3.2.6. Analysis of real samples

Eleven samples were collected during a monitoring campaign (February 2020) of different water bodies in the Northern Italy (Table 3.4). Samples were collected keeping a proper distance from the bank. Samples were stored in 100 mL decontaminated clear glass bottles filled to the top. Samples were preserved at 4 °C and analyzed within 15 days after the sampling. Samples were submitted to analysis with the proposed analytical method after proper dilution in triplicate.

3.2.7. Software

GC-MS data were handled with the HP ChemStation (Agilent Technologies) software. The significance of the coefficients was evaluated by means of SPSS Statistics v.23.0 (IBM, Milan, Italy) software. Multicriteria optimization was carried out by running a custom macro in Excel (Microsoft, Washington, USA).

Table 3.4. Locations of the samples that were collected and analyzed in the framework of the present study.

Sample	Area	Province	Water body	Name
A	Sirmione	Brescia	Lake	Garda
B	Rivoltella del Garda	Brescia	Lake	Garda
C	Montichiari	Brescia	River	Chiese
D	Sassuolo	Modena	River	Secchia
E	S. Vittoria di Gualtieri	Reggio Emilia	River	Tassone
F	Traversetolo	Parma	River	Termina
G	Fiorenzuola d'Arda	Piacenza	River	Arda
H	Finale Emilia	Modena	Outlet of a wastewater treatment facility	-
I	S. Quirico	Parma	River	Taro
J	Colorno	Parma	River	Po
K	Colorno	Parma	River	Naviglio

3.3. Results and discussion

3.3.1. Characterization of the CNTs-based coatings

As mentioned earlier, CNTs have interesting topological, geometric, and physicochemical properties that make them attractive as SPME coating. In the framework of this study, four different types of CNTs were taken into consideration for the fabrication of SPME devices.

HMWCNTs, MWCNTs50, and MWCNTs8 have different diameters and spatial configurations and differ in surface area. These three types of CNTs can interact with the analytes *via* π - π stacking, CH- π interactions, and London dispersion forces. MWCNTs-COOH had the same size of MWCNTs50, but their surface was both graphitized and functionalized with -COOH groups. Graphitization exposes graphene leaves on the surface of the CNTs [50], thus increasing both the surface area and, the possibility of interacting *via* π - π stacking, CH- π interactions, and London dispersion forces. The functionalization with -COOH groups allows hydrogen bonding and dipole-dipole interactions with polar functional groups present in the molecular structure of the analytes.

All the investigated CNTs proved being thermally stable up to 400 °C with a mass-loss below 3%. For temperatures higher than 160 °C, MWCNTs-COOH showed a

steeper mass loss than the other CNTs. This was most likely ascribable to the loss of -COOH functional groups through decarboxylation [51]. Thermal stability was further investigated by placing the fibers into the GC injector for 2 min at 320 °C after proper conditioning. TGA curves and gas chromatograms (SIM-MS mode, see Table 3.2) are reported in Figure 3.4. The morphology of the fibers was investigated by ESEM. All the fabricated fibers showed a homogeneous coating. The thicknesses of the coatings were: 41.1 (\pm 3.7) μm for the HMWCNTs-based fiber, 40.7 (\pm 3.5) μm for the MWCNTs50-based fiber, 40.6 (\pm 4.4) μm for the MWCNTs8-based fiber, and 40.5 (\pm 4.1) μm for the MWCNTs-COOH-based fiber (results are reported as mean \pm one standard deviation; $N = 3$ fibers *per* coating, $n = 3$ replicated measurements *per* fiber). As an example, the ESEM micrograph of the HMWCNTs-based fiber is reported in Figure 3.5.

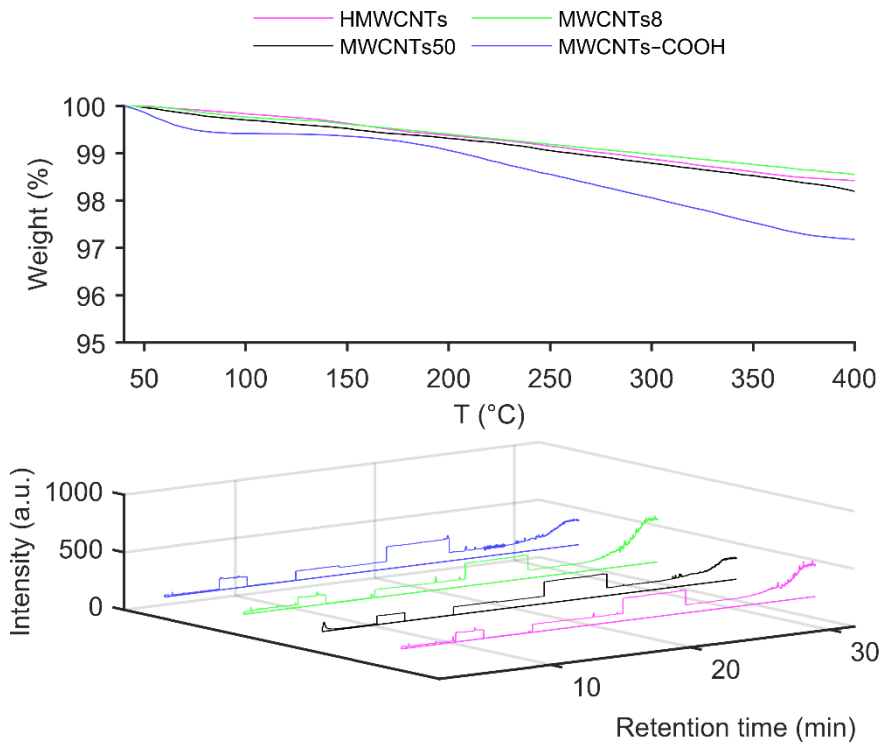


Figure 3.4. Results of the TGA (left) and GC-SIM-MS chromatograms obtained *via* thermal desorption (right).

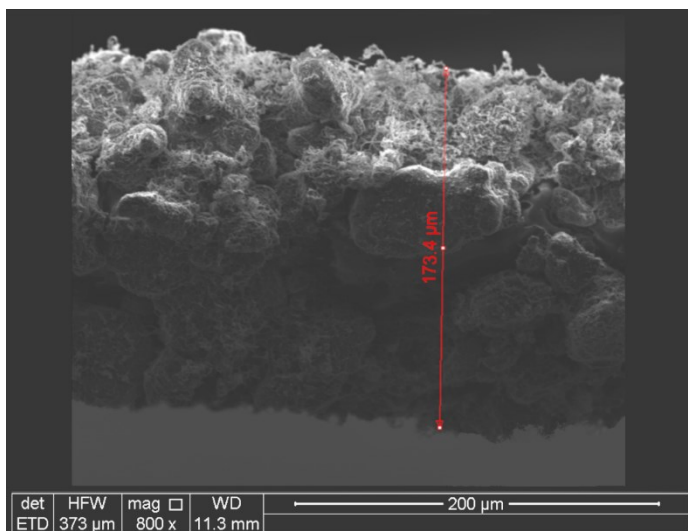


Figure 3.5. ESEM micrograph of the HMWCNTs-based coating. Reprinted from [52].

3.3.2. Optimization of the SPME procedure

3.3.2.1. Preliminary evaluation of the coating performance

The analytical performance of the investigated CNTs-based coatings was evaluated before the proper optimization stage. The desorption temperature and the desorption time were set at 320 °C and 2 min, respectively, according to what was observed from the TGA measurements and the evaluation of bleeding. The responses were transformed by taking the square root to fulfill the homoscedasticity assumption. ANOVA was carried out on the transformed responses: statistically significant differences were found in the average responses for all the analytes ($p < 0.05$; Cohen's $\eta^2 \geq 0.88$). The results are depicted in Figure 3.6. HMWCNTs and MWCNTs8 performed the best in terms of repeatability, with $RSD\% < 20\%$, whereas for the MWCNTs50 and MWCNTs-COOH $RSD\%$ within 24% and 23%, respectively, were observed.

As a general comment, HMWCNTs showed better or not significantly different performance than the other investigated coatings ($p > 0.05$, Bonferroni adjusted). It has been reported in the literature that the shape and topological features of the analytes are the primary factor that influences their ability to interact with MWCNTs [53], but the reasons behind the superior sorptive capabilities of HMWCNTs are still a matter of debate.

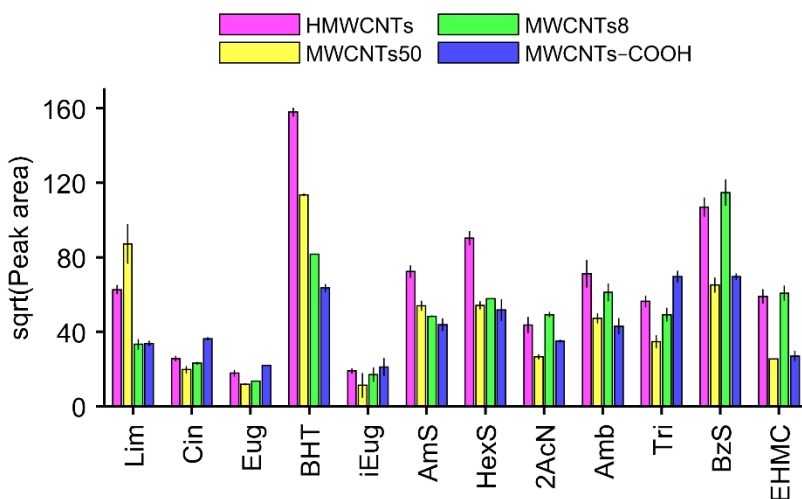


Figure 3.6. Coating effect on the extraction of the analytes. Results are plotted as mean \pm one standard deviation ($N = 3$ fibers *per* coating, $n = 3$ independent replicated measurements *per* fiber). Reprinted from [52].

The results observed in the present study could find an explanation in the lack of functional groups on the surfaces on HMWCNTs as well as in their helicoidal shape, as these two features yields a high surface area. Other studies have found similar results in evaluating the performance of different CNTs on the extraction of pharmaceuticals and polycyclic aromatic hydrocarbons [54–56].

In only two instances MWCNTs–COOH showed the best performance ($p < 0.05$, Bonferroni adjusted), particularly towards Cin and Tri. This could be ascribed to the presence of strongly polar groups in the structure of the abovementioned analytes. MWCNTs–COOH did not appear suitable for the extraction of the other analytes, as the carboxylic groups can interact *via* hydrogen bonding with water, hindering the interaction with the analytes. In addition, hydrophobic and π – π interactions are weakened by the presence of carboxylic groups [54]. Furthermore, lower desorption temperatures had to be used to ensure adequate reusability for the MWCNTs–COOH-based fiber avoiding decarboxylation [51]. In this case, the main drawback resulted in the incomplete desorption of the analytes.

Based on these findings, HMWCNTs were selected as the coating material for further method development.

3.3.2.2. Optimization

Among the several parameters that could potentially affect the efficacy of the SPME procedure, extraction temperature and extraction time are among the most critical

ones [29,30]. Extraction temperature and extraction time were, therefore, studied as main factors. The experimental domains were set by taking into consideration operative limitations.

- **Extraction temperature:** temperatures lower than 30 °C were difficult to stabilize, whereas temperatures higher than 60 °C were not considered to avoid the desorption of the analytes from the fiber.
- **Extraction time:** extraction times higher than 15 min were considered to ensure a good analytical response from all the analytes, whereas extraction times higher than 45 min were not taken into consideration to guarantee a good balance between the time required for the extraction and the analysis time.

Ionic strength was not investigated as preliminary experiments (data not shown) revealed that no significant improvement ($p > 0.05$) was produced when NaCl was added in a concentration range of 0–10% w/w. Moreover, when the concentration was 20% w/w the responses decreased. This is most likely ascribable to the increased viscosity of sample solutions, making more difficult for the analytes to diffuse to the coating interface. Similar results have been previously reported [57,58].

The responses were left untransformed, and the models were calculated. ANOVA showed that all the models were valid ($p > 0.05$), meaning that the error deriving from approximation is not significantly greater than the variance that it could be expected experimentally. The models showed R^2 in the 0.65–0.91 range. For the sake of brevity, Lim, HexS, and EHMC surface plots were shown to exemplify the typical shapes of response surfaces (Figure 3.7), whereas the regression models are reported in Table 3.5.

All the terms in the regression models had a positive coefficient. As far as the main factors are taken into consideration, it can be stated that temperature had a positive influence on the response, as it increases the diffusion coefficients of the analytes towards the fiber interface [29,30]. As for extraction time, salicylate esters and EHMC responses resulted in being positively influenced by a protracted exposure of the fiber into the sample solution. Salicylate esters and EHMC models also included the interaction term, and EHMC model also retained the quadratic term for extraction temperature.

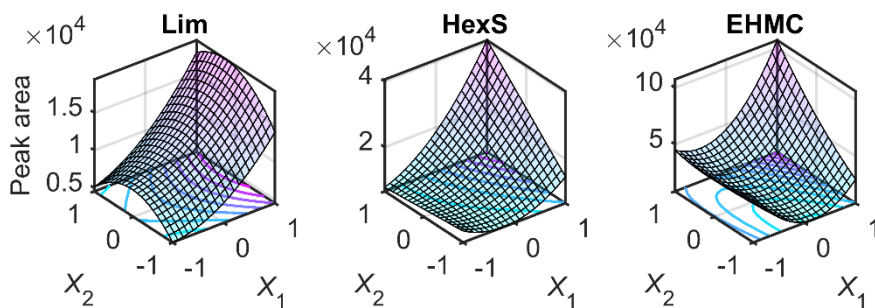


Figure 3.7. Response surfaces of Lim, HexS, and EHMC. The predicted response for each analyte is plotted as a function of extraction temperature and extraction time (X_1 and X_2 , respectively).

Table 3.5. Regression models calculated for each analyte.

Analyte	Equation ^a
Lim	$y = 10960 (\pm 880) + 5500 (\pm 1200)X_1$
Cin	$y = 3100 (\pm 370) + 2280 (\pm 350)X_1$
Eug	$y = 540 (\pm 60) + 420 (\pm 80)X_1$
BHT	$y = 12700 (\pm 1200) + 7900 (\pm 1700)X_1$
iEug	$y = 730 (\pm 60) + 720 (\pm 80)X_1$
AmS	$y = 14800 (\pm 1600) + 9900 (\pm 2300)X_1 + 5900 (\pm 2300)X_2 + 6900 (\pm 2900)X_1X_2$
HexS	$y = 15500 (\pm 1600) + 9400 (\pm 2300)X_1 + 5900 (\pm 2300)X_2 + 7100 (\pm 2800)X_1X_2$
2AcN	$y = 4610 (\pm 490) + 2900 (\pm 700)X_1$
Amb	$y = 15700 (\pm 1400) + 8100 (\pm 1900)X_1$
Tri	$y = 6020 (\pm 570) + 3380 (\pm 810)X_1$
BzS	$y = 23000 (\pm 2500) + 14100 (\pm 3600)X_1 + 8500 (\pm 3600)X_2 + 10800 (\pm 4400)X_1X_2$
EHMC	$y = 23400 (\pm 4200) + 10600 (\pm 4200)X_1 + 18200 (\pm 4200)X_2 + 20600 (\pm 5200)X_1X_2 + 32400 (\pm 6000)X_1^2$

^a only the terms that were retained by the variable selection process are reported ($\alpha_{to_remove} = 0.05$), the coefficients (X_1 : extraction temperature; X_2 : extraction time) are reported as coefficient (\pm standard error) rounded at two significant digit.

The optimal conditions required an extraction temperature of 60 °C and a extraction time of 45 min. The global desirability corresponding to such experimental conditions was $D = 0.91$, with single desirability values $d_i \geq 0.90$. A non-zero global desirability value means that a set of conditions able to fulfill the optimization criteria for all the responses could be found, whereas its high value denotes a high degree of accordance between the single desirability values. The

identified experimental conditions were valid, as the lack-of-fit was not statistically significant ($p > 0.05$).

3.3.3. Method validation

The SPME–GC–MS method was validated operating under optimized conditions. The LODs, LOQs, calibration range, and equation of the calibration functions are reported in Table 3.6.

The method exhibited LODs in the 0.2–12.7 ng/L range, suitable for the detection of the investigated analytes at trace and ultra-trace levels, whereas LOQs were in the 0.7–42.3 ng/L range. All the calibration functions proved to be linear ($p > 0.01$) at least within two orders of magnitude starting from the LOQ and statistically significant ($p < 0.05$). In addition, the approximation error was not significantly greater than the expected pure experimental variation ($p > 0.05$).

Satisfactory results were obtained from the precision and trueness studies, with the results summarized in Table 3.7 and in Figure 3.8, respectively. There were not statistically significant differences in the results obtained under intermediate precision conditions.

Table 3.6. LODs, LOQs, calibration range, and calibration curve equations of the SPME–GC–MS analytical method. LODs, LOQs, and calibration range values are reported in ng/L.

Analyte	LOD	LOQ	Range	Calibration curve equation ^a	
				$b_0 \pm SE$ ^b	$b_1 \pm SE$ ^b
Lim	1.0	3.2	LOQ–2000	$7 (\pm 2) \times 10^{-2}$	$1.26 (\pm 0.02) \times 10^{-3}$
Cin	12.7	42.3	LOQ–7500	– ^c	$3.24 (\pm 0.03) \times 10^{-4}$
Eug	6.2	20.6	LOQ–7500	–	$4.13 (\pm 0.03) \times 10^{-4}$
BHT	0.9	2.9	LOQ–1000	–	$2.81 (\pm 0.03) \times 10^{-3}$
iEug	6.0	20.1	LOQ–7500	–	$7.45 (\pm 0.09) \times 10^{-4}$
AmS	1.2	4.1	LOQ–1000	–	$6.13 (\pm 0.04) \times 10^{-3}$
HexS	1.3	4.4	LOQ–500	–	$1.38 (\pm 0.01) \times 10^{-2}$
2AcN	1.4	4.6	LOQ–2000	–	$2.17 (\pm 0.04) \times 10^{-4}$
Amb	1.2	4.1	LOQ–2000	$0.21 (\pm 0.05)$	$1.32 (\pm 0.04) \times 10^{-3}$
Tri	2.3	7.8	LOQ–2000	–	$5.48 (\pm 0.08) \times 10^{-4}$
BzS	0.2	0.7	LOQ–500	–	$1.55 (\pm 0.01) \times 10^{-2}$
EHMC	0.9	3.0	LOQ–500	–	$8.04 (\pm 0.11) \times 10^{-3}$

^a the equation of the calibration curve is $y = b_0 + b_1x$. ^b the intercept and slope are reported as coefficient \pm standard error rounded at one significant digit. ^c – not significant ($p > 0.05$).

Table 3.7. Results of the precision study. $RSD\%$ referring to repeatability is marked with an r in apex and the one referring to intermediate precision is marked with an I in apex. Results are rounded at the nearest integer.

Analyte	L1 ^a		L2 ^a		L3 ^a	
	$RSD^r\%$	$RSD^I\%$	$RSD^r\%$	$RSD^I\%$	$RSD^r\%$	$RSD^I\%$
Lim	6	19	7	9	3	4
Cin	6	18	10	12	1	3
Eug	9	17	2	4	1	2
BHT	2	19	4	5	3	4
iEug	7	18	2	13	3	5
AmS	12	19	9	11	1	2
HexS	12	19	11	12	2	3
2AcN	12	13	12	17	4	5
Amb	2	19	5	6	6	8
Tri	6	20	9	11	3	5
BzS	12	19	5	14	1	2
EHMC	1	20	12	15	4	5

^a the concentration levels are reported in Table 3.3.

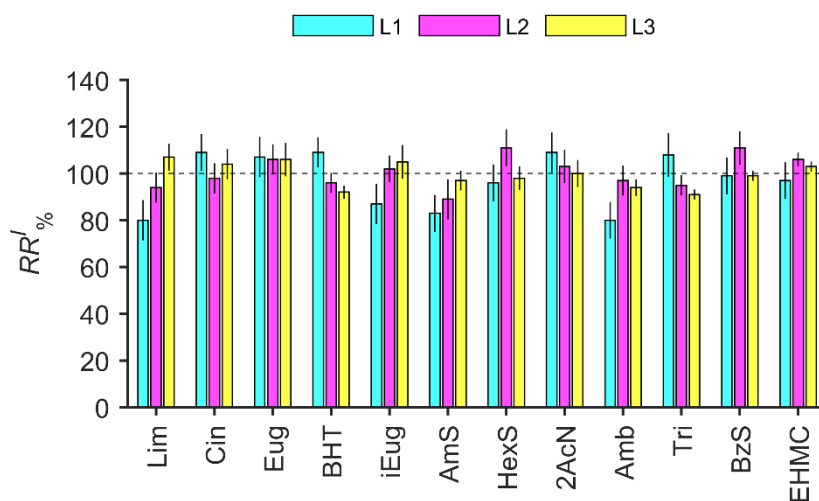


Figure 3.8. Results of the trueness study. Results are reported as mean \pm semi-amplitude of the 95% confidence interval ($n = 10$). The horizontal dashed line corresponds to a $RR\% = 100\%$. The concentration levels L1, L2, and L3 are reported in Table 3.3.

The most hydrophobic compounds and 2AcN showed the highest EFs (Figure 3.9). In the case of 2AcN, this behavior could find an explanation in the structure of the analyte: in fact, 2AcN is characterized by an electron-poor naphthalene scaffold that can interact with electron-rich HMWCNTs. The performance of the proposed

HMWCNTs-based coating was compared with that of a commercially available SPME fiber, i.e., DVB/Carboxen™-PDMS fiber. The latter is an adsorptive coating whose characteristics make it suitable for the vast majority of routine applications. The HMWCNTs-based coating provided better or not significantly different EFs than the ones achieved by the DVB/Carboxen™-PDMS fiber, with an average improvement of 6 standard deviations [59].

The proposed analytical methodology provided LODs substantially lower than the ones reported in other studies [60,61] dealing with the determination of this class of analytes also by using more selective and sensitive techniques, such as MS/MS [12,44–46]. It is worth mentioning that the volume of the aliquots requested by the proposed SPME–GC–MS method were considerably lower than the ones requested by methodologies reported in other studies dealing with the determination of such analytes *via* solid phase extraction followed by GC–MS/MS [12,44–46]. In addition, the devised method is solventless and requires minimum sample handling, as SPME can be completely automated.

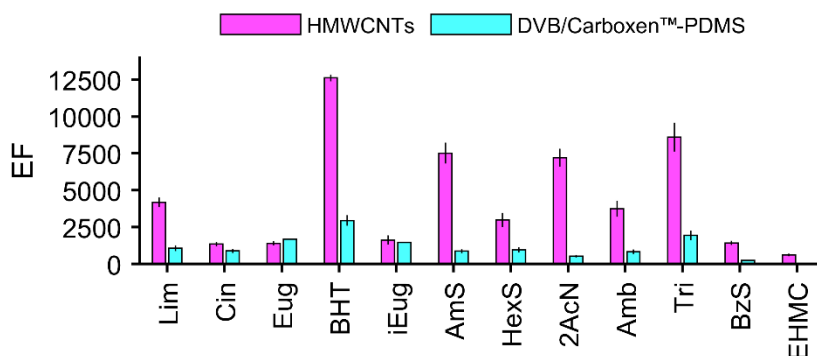


Figure 3.9. Enrichment factors for the HMWCNTs-based fiber and the DVB/Carboxen™-PDMS fiber. Results are reported as mean \pm one standard deviation ($n = 3$). Reprinted from [52].

3.3.4. Analysis of real samples

The proposed analytical method was used for the analysis of 11 surface water samples collected during a sampling campaign, in February 2020. The samples were collected from rivers and lakes belonging to the drainage basin of the Po River. Po is 652 km long and its drainage basin extends for about 71000 km^{2xxxv}. These characteristics make the Po River the longest river in Italy and the one with the largest drainage basin. The Po River flows, for the vast majority of its length, across the Padan Plain, a highly populated area, but the distribution of the population is not uniform. Within this frame of reference, the monitoring of PCPs in

xxxv <https://www.adbpo.it> (accessed 03/09/2022)

the drainage basin of the Po River could provide useful information related to the anthropic impact of the population on the water quality.

The results obtained in this study are summarized in Table 3.8. With the exception of sample B, in which none of the investigated analytes was detected, the other ones showed various concentration levels of the analytes, ranging from few ng/L to few hundreds ng/L. BHT was quantified in 9 out of 11 samples: the widespread presence of this compound could find an explanation in the fact is widely used as an antioxidant in the food and pharmaceutical industries as well as in the manufacturing of plastic materials [62]. Lim and Amb were the second most widely detected analytes, being present in 4 samples out of 11. AmS, HexS, 2AcN, and Tri were not detected in any of the analyzed samples.

Sample H was characterized by the highest overall concentration of PCPs, i.e., 420 (± 10) ng/L, with BHT being the most abundant analyte. The high concentration of analytes detected in this samples agrees with what has already been reported in the literature: high concentration levels of PCPs can be found in sample collected in the neighboring of wastewater treatment facilities [1–3,15]. This could be an indication about the persistence in the environment of these compounds and their non-exhaustive removal by the unitary operations involved in wastewater treatment processes.

3.4. Conclusions

This study was devoted to the evaluation of the potential of four different CNTs for the fabrication of SPME coatings. The fabricated fibers were used for the SPME-GC-MS analysis of a pool of 12 PCPs in surface water samples. HMWCNTs proved to be the most promising material for the adsorption of the investigated compounds and, therefore, MNWCNTs-based fibers were utilized for further method development. The method was validated according to the EURACHEM guidelines, providing LODs and LOQs suitable for the analysis of the investigated PCPs at trace and ultra-trace levels. The enrichment factors were noticeable, and they were compared with the ones obtained with a commercially available SPME fiber, namely the DVB/Carboxen™-PDMS. The HMWCNTs-based coating showed enrichment capabilities, on average, 6 standard deviations higher than the ones obtained for the commercial fiber, with the only exception of 2 analytes. On a closing note, the proposed methodology can be considered suitable for the determination of PCPs at trace and ultra-trace levels in full respect of the principles of Green Analytical Chemistry, as no extraction solvents were involved in the operating procedures.

Table 3.8. Results of the analyses on the collected samples. All the results are in ng/L expressed as mean \pm one standard deviation (n = 3). Results are rounded at one significant digit.

Analyte	Sample										
	A	B	C	D	E	F	G	H	I	J	K
Lim	39 \pm 5	n.d.	n.d.	18 \pm 2	97 \pm 5	n.d.	42 \pm 4	n.d.	n.d.	n.d.	n.d.
Cin	n.d.	n.d.	n.d.	126 \pm 4	22 \pm 2	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Eug	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	89 \pm 6	n.d.	n.d.	58 \pm 4
BHT	4.0 \pm 0.3	n.d.	n.d.	16 \pm 1	4.9 \pm 0.2	6.3 \pm 0.4	30 \pm 3	310 \pm 10	5.9 \pm 0.7	40 \pm 3	8.4 \pm 0.9
iFug	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	15 \pm 3	n.d.	n.d.	n.d.
Ams	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
HexS	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
ZAcN	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Amb	n.d.	n.d.	n.d.	11 \pm 1	5.6 \pm 0.7	n.d.	8.9 \pm 0.7	n.d.	n.d.	n.d.	4.4 \pm 0.2
Tri	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
BzS	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	7.1 \pm 0.8	n.d.	n.d.	n.d.
EHMC	n.d.	n.d.	13.0 \pm 0.4	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Σ^{12} PCP ^b	43 \pm 5	n.d.	13.0 \pm 0.4	172 \pm 5	129 \pm 5	6.3 \pm 0.4	81 \pm 5	420 \pm 10	5.9 \pm 0.7	40 \pm 3	71 \pm 4

a not detected, b the sum accounts only for the quantitated analytes.

Note of the author

The results presented in this **Chapter** are published on *Talanta* 224 (2021) 121891 [52].

The author would like to express gratitude to Cesare D'Angelo and Enrico Marraffa for participating to this study as a part of the internship required to obtain the Master of Science and Bachelor of Science degrees, respectively.

References

- [1] A. Grobelak, A. Kowalska, Emerging environmental contaminants—current status, challenges, and technological solutions, *Emerging Contaminants in the Environment: Challenges and Sustainable Practices*. (2022) 39–53. <https://doi.org/10.1016/B978-0-323-85160-2.00010-X>.
- [2] S. Dey, F. Bano, A. Malik, Pharmaceuticals and personal care product (PPCP) contamination—a global discharge inventory, *Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology Emerging Contaminants and Micro Pollutants*. (2019) 1–26. <https://doi.org/10.1016/B978-0-12-816189-0.00001-9>.
- [3] A. Gogoi, P. Mazumder, V.K. Tyagi, G.G. Tushara Chaminda, A.K. An, M. Kumar, Occurrence and fate of emerging contaminants in water environment: A review, *Groundw Sustain Dev*. 6 (2018) 169–180. <https://doi.org/10.1016/J.GSD.2017.12.009>.
- [4] V. Geissen, H. Mol, E. Klumpp, G. Umlauf, M. Nadal, M. van der Ploeg, S.E.A.T.M. van de Zee, C.J. Ritsema, Emerging pollutants in the environment: A challenge for water resource management, *International Soil and Water Conservation Research*. 3 (2015) 57–65. <https://doi.org/10.1016/J.ISWCR.2015.03.002>.
- [5] G. Hanke, V. Dulio, T. Ternes, F.P.J. Lamé, D. Kotzias, V.P. De, E. Brorström-Lundén, W. Giger, R. Owen, S. Eisenreich, N. Cartwright, N. Theobald, J. Volz, C. Carlon, D. Briggs, H. Lokke, M. Holt, A.M. Fouillac, J. Slobodnik, D. Schwesig, D., H. Leslie, M. Coquery, R. Loos, P. Lepom, W. Pejinenburg, M. McLachlan, Report of the NORMAN Workshop on Emerging Environmental Pollutants: Key Issues and Challenges, (2006). <https://publications.jrc.ec.europa.eu/repository/handle/JRC37428> (accessed August 22, 2022).
- [6] V. Dulio, B. van Bavel, E. Brorström-Lundén, J. Harmsen, J. Hollender, M. Schlabach, J. Slobodnik, K. Thomas, J. Koschorreck, Emerging pollutants in the EU: 10 years of NORMAN in support of environmental policies and regulations, *Environ Sci Eur*. 30 (2018) 1–13. <https://doi.org/10.1186/S12302-018-0135-3/FIGURES/2>.
- [7] O.M. Rodriguez-Narvaez, J.M. Peralta-Hernandez, A. Goonetilleke, E.R. Bandala, Treatment technologies for emerging contaminants in water: A review, *Chemical Engineering Journal*. 323 (2017) 361–380. <https://doi.org/10.1016/J.CEJ.2017.04.106>.
- [8] M.L. Richardson, J.M. Bowron, The fate of pharmaceutical chemicals in the aquatic environment, *Journal of Pharmacy and Pharmacology*. 37 (2011) 1–12. <https://doi.org/10.1111/J.2042-7158.1985.TB04922.X>.
- [9] S. Yan, S.B. Subramanian, R.D. Tyagi, R.Y. Surampalli, T.C. Zhang, Emerging Contaminants of Environmental Concern: Source, Transport, Fate, and Treatment, *Practice Periodical of Hazardous, Toxic, and Radioactive Waste Management*. 14 (2009) 2–20. [https://doi.org/10.1061/\(ASCE\)HZ.1944-8376.0000015](https://doi.org/10.1061/(ASCE)HZ.1944-8376.0000015).
- [10] Ş. Sungur, Pharmaceutical and personal care products in the environment: occurrence and impact on the functioning of the ecosystem, *Emerging Contaminants in the Environment: Challenges and Sustainable Practices*. (2022) 137–157. <https://doi.org/10.1016/B978-0-323-85160-2.00009-3>.

- [11] D. Relić, A. Popović, D. Đorđević, J. Čáslavský, Occurrence of synthetic musk compounds in surface, underground, waste and processed water samples in Belgrade, Serbia, *Environmental Earth Sciences* 2017 76:3. 76 (2017) 1–10. <https://doi.org/10.1007/S12665-017-6441-Z>.
- [12] M. Vecchiato, S. Cremonese, E. Gregoris, E. Barbaro, A. Gambaro, C. Barbante, Fragrances as new contaminants in the Venice lagoon, *Science of The Total Environment*. 566–567 (2016) 1362–1367. <https://doi.org/10.1016/J.SCITOTENV.2016.05.198>.
- [13] B.F. da Silva, A. Jelic, R. López-Serna, A.A. Mozeto, M. Petrovic, D. Barceló, Occurrence and distribution of pharmaceuticals in surface water, suspended solids and sediments of the Ebro river basin, Spain, *Chemosphere*. 85 (2011) 1331–1339. <https://doi.org/10.1016/J.CHEMOSPHERE.2011.07.051>.
- [14] P.E. Stackelberg, E.T. Furlong, M.T. Meyer, S.D. Zaugg, A.K. Henderson, D.B. Reissman, Persistence of pharmaceutical compounds and other organic wastewater contaminants in a conventional drinking-water-treatment plant, *Science of The Total Environment*. 329 (2004) 99–113. <https://doi.org/10.1016/J.SCITOTENV.2004.03.015>.
- [15] Y. Yang, Y.S. Ok, K.H. Kim, E.E. Kwon, Y.F. Tsang, Occurrences and removal of pharmaceuticals and personal care products (PPCPs) in drinking water and water/sewage treatment plants: A review, *Science of The Total Environment*. 596–597 (2017) 303–320. <https://doi.org/10.1016/J.SCITOTENV.2017.04.102>.
- [16] K. Jamil, Health effects of pharmaceuticals and personal care products, *Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology Emerging Contaminants and Micro Pollutants*. (2019) 115–128. <https://doi.org/10.1016/B978-0-12-816189-0.00005-6>.
- [17] Y. Nakama, Surfactants, *Cosmetic Science and Technology: Theoretical Principles and Applications*. (2017) 231–244. <https://doi.org/10.1016/B978-0-12-802005-0.00015-X>.
- [18] J. Tolls, H. Berger, A. Klenk, M. Meyberg, R. Müller, K. Rettinger, J. Steber, Environmental safety aspects of personal care products—A European perspective, *Environ Toxicol Chem*. 28 (2009) 2485–2489. <https://doi.org/10.1897/09-104.1>.
- [19] A. Chisvert, A. Salvador, UV Filters in Sunscreens and other Cosmetics. *Regulatory Aspects and Analytical Methods, Analysis of Cosmetic Products*. (2007) 83–120. <https://doi.org/10.1016/B978-044452260-3/50028-0>.
- [20] J. Wang, L. Pan, S. Wu, L. Lu, Y. Xu, Y. Zhu, M. Guo, S. Zhuang, Recent Advances on Endocrine Disrupting Effects of UV Filters, *International Journal of Environmental Research and Public Health* 2016, Vol. 13, Page 782. 13 (2016) 782. <https://doi.org/10.3390/IJERPH13080782>.
- [21] R.D. Horsak, P.B. Bedient, M.C. Hamilton, F. ben Thomas, *Pesticides, Environmental Forensics: Contaminant Specific Guide*. (1964) 143–165. <https://doi.org/10.1016/B978-012507751-4/50030-6>.
- [22] B. Freedman, *PESTICIDES, Environmental Ecology*. (1995) 213–277. <https://doi.org/10.1016/B978-0-08-050577-0.50013-3>.
- [23] S. Panthi, A.R. Sapkota, G. Raspanti, S.M. Allard, A. Bui, H.A. Craddock, R. Murray, L. Zhu, C. East, E. Handy, M.T. Callahan, J. Haymaker, P. Kulkarni, B. Anderson, S. Craighead, S. Gartley, A. Vanore, W.Q. Betancourt, R. Duncan, D. Foust, M. Sharma, S.A. Micallef, C. Gerba, S. Parveen, F. Hashem, E. May, K. Kniel, M. Pop, S. Ravishankar, A. Sapkota, *Pharmaceuticals, herbicides, and disinfectants in agricultural water sources, Environ Res*. 174 (2019) 1–8. <https://doi.org/10.1016/J.ENVRES.2019.04.011>.
- [24] D.T. Salvito, M.G.H. Vey, R.J. Senna, *Fragrance materials and their environmental impact, Flavour Fragr J*. 19 (2004) 105–108. <https://doi.org/10.1002/FFJ.1398>.
- [25] European Commission, Directorate-General for Health and Consumers, *Opinion on fragrance allergens in cosmetic products*, (2013). <https://doi.org/doi/10.2772/77628>.
- [26] R.M. Smith, Before the injection - Modern methods of sample preparation for separation techniques, *J Chromatogr A*. 1000 (2003) 3–27. [https://doi.org/10.1016/S0021-9673\(03\)00511-9](https://doi.org/10.1016/S0021-9673(03)00511-9).

- [27] M. de la Guardia, S. Garrigues, Past, Present and Future of Green Analytical Chemistry, in: M. de la Guardia, S. Garrigues (Eds.), *Challenges in Green Analytical Chemistry*, 2nd ed., The Royal Society of Chemistry, 2020: pp. 1–18. <https://doi.org/10.1039/9781788016148-00001>.
- [28] J. Płotka-Wasyłka, N. Szczepańska, M. de la Guardia, J. Namieśnik, Miniaturized solid-phase extraction techniques, *TrAC - Trends in Analytical Chemistry*. 73 (2015) 19–38. <https://doi.org/10.1016/j.trac.2015.04.026>.
- [29] J. Pawliszyn, *Solid Phase Microextraction: Theory and Practice*, Wiley-VCH, New York, 1997. <https://www.wiley.com/en-us/Solid+Phase+Microextraction%3A+Theory+and+Practice-p-9780471190349> (accessed August 25, 2022).
- [30] J. Pawliszyn, *Handbook of Solid Phase Microextraction*, 2012. <https://doi.org/10.1016/B978-0-12-416017-0.00011-5>.
- [31] Z. Zhang, M.J. Yang, J. Pawliszyn, Solid-Phase Microextraction, *Anal Chem*. 66 (2012) 844A–853A. <https://doi.org/10.1021/AC00089A716>.
- [32] E. Gionfriddo, E. Boyaci, J. Pawliszyn, New Generation of Solid-Phase Microextraction Coatings for Complementary Separation Approaches: A Step toward Comprehensive Metabolomics and Multiresidue Analyses in Complex Matrices, *Anal Chem*. 89 (2017) 4046–4054. <https://doi.org/10.1021/acs.analchem.6b04690>.
- [33] J. Zheng, J. Huang, Q. Yang, C. Ni, X. Xie, Y. Shi, J. Sun, F. Zhu, G. Ouyang, Fabrications of novel solid phase microextraction fiber coatings based on new materials for high enrichment capability, *TrAC - Trends in Analytical Chemistry*. 108 (2018) 135–153. <https://doi.org/10.1016/j.trac.2018.08.021>.
- [34] Y. Ai, M. Wu, L. Li, F. Zhao, B. Zeng, Highly selective and effective solid phase microextraction of benzoic acid esters using ionic liquid functionalized multiwalled carbon nanotubes-doped polyaniline coating, *J Chromatogr A*. 1437 (2016) 1–7. <https://doi.org/10.1016/j.chroma.2016.01.072>.
- [35] H.B. Shang, C.X. Yang, X.P. Yan, Metal–organic framework UiO-66 coated stainless steel fiber for solid-phase microextraction of phenols in water samples, *J Chromatogr A*. 1357 (2014) 165–171. <https://doi.org/10.1016/j.chroma.2014.05.027>.
- [36] J. Feng, M. Sun, Y. Bu, C. Luo, Facile modification of multi-walled carbon nanotubes–polymeric ionic liquids-coated solid-phase microextraction fibers by on-fiber anion exchange, *J Chromatogr A*. 1393 (2015) 8–17. <https://doi.org/10.1016/j.chroma.2015.03.022>.
- [37] D. Djozan, T. Baheri, Preparation and evaluation of solid-phase microextraction fibers based on monolithic molecularly imprinted polymers for selective extraction of diacetylmorphine and analogous compounds, *J Chromatogr A*. 1166 (2007) 16–23. <https://doi.org/10.1016/j.chroma.2007.08.003>.
- [38] S. Iijima, Helical microtubules of graphitic carbon, *Nature*. 354 (1991) 56–58. <https://doi.org/10.1038/354056a0>.
- [39] A. Merkoçi, Carbon Nanotubes in Analytical Sciences, *Microchimica Acta*. 152 (2005) 157–174. <https://doi.org/10.1007/S00604-005-0439-Z>.
- [40] F. Ghaemi, A. Amiri, R. Yunus, Methods for coating solid-phase microextraction fibers with carbon nanotubes, *TrAC Trends in Analytical Chemistry*. 59 (2014) 133–143. <https://doi.org/10.1016/j.trac.2014.04.011>.
- [41] Y. Wang, J. Ma, J. Zhu, N. Ye, X. Zhang, H. Huang, Multi-walled carbon nanotubes with selected properties for dynamic filtration of pharmaceuticals and personal care products, *Water Res*. 92 (2016) 104–112. <https://doi.org/10.1016/j.watres.2016.01.038>.
- [42] Ihsanullah, A. Abbas, A.M. Al-Amer, T. Laoui, M.J. Al-Marri, M.S. Nasser, M. Khraisheh, M.A. Atieh, Heavy metal removal from aqueous solution by advanced carbon nanotubes: Critical review of adsorption applications, *Sep Purif Technol*. 157 (2016) 141–161. <https://doi.org/10.1016/j.seppur.2015.11.039>.

- [43] P.A. Martins-Júnior, C.E. Alcântara, R.R. Resende, A.J. Ferreira, Carbon nanotubes: Directions and perspectives in oral regenerative medicine, *J Dent Res.* 92 (2013) 575–583. <https://doi.org/10.1177/0022034513490957>.
- [44] M. Vecchiato, E. Gregoris, E. Barbaro, C. Barbante, R. Piazza, A. Gambaro, Fragrances in the seawater of Terra Nova Bay, Antarctica, *Science of The Total Environment.* 593–594 (2017) 375–379. <https://doi.org/10.1016/j.scitotenv.2017.03.197>.
- [45] M. Vecchiato, E. Barbaro, A. Spolaor, F. Burgay, C. Barbante, R. Piazza, A. Gambaro, Fragrances and PAHs in snow and seawater of Ny-Ålesund (Svalbard): Local and long-range contamination, *Environmental Pollution.* 242 (2018) 1740–1747. <https://doi.org/10.1016/j.envpol.2018.07.095>.
- [46] M. Vecchiato, C. Turetta, B. Patti, C. Barbante, R. Piazza, T. Bonato, A. Gambaro, Distribution of fragrances and PAHs in the surface seawater of the Sicily Channel, Central Mediterranean, *Science of The Total Environment.* 634 (2018) 983–989. <https://doi.org/10.1016/j.scitotenv.2018.04.080>.
- [47] J. Cohen, Eta-squared and partial eta-squared in fixed factor ANOVA designs, *Educ Psychol Meas.* 33 (1973) 107–112. <https://doi.org/10.1177/001316447303300111>.
- [48] G. Derringer, R. Suich, Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology.* 12 (1980) 214–219. <https://doi.org/10.1080/00224065.1980.11980968>.
- [49] B. Magnusson, U. Örnemark, eds., *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topic*, 2nd ed., 2014. [https://doi.org/10.1016/S0014-2999\(99\)00500-2](https://doi.org/10.1016/S0014-2999(99)00500-2).
- [50] K. Yu, G. Lu, Z. Bo, S. Mao, J. Chen, Carbon nanotube with chemically bonded graphene leaves for electronic and optoelectronic applications, *Journal of Physical Chemistry Letters.* 2 (2011) 1556–1562. https://doi.org/10.1021/JZ200641C/SUPPL_FILE/JZ200641C_SI_001.PDF.
- [51] A.N. Laguta, N.O. McHedlov-Petrosyan, S.M. Kovalenko, T.O. Voloshina, V.I. Haidar, D.Y. Filatov, P. v. Trostyanko, V.L. Karbivski, S.I. Bogatyrenko, L. Xu, O. v. Prezhdo, Stability of Aqueous Suspensions of COOH-Decorated Carbon Nanotubes to Organic Solvents, Esterification, and Decarboxylation, *Journal of Physical Chemistry Letters.* (2022) 10126–10131. https://doi.org/10.1021/ACS.JPCLETT.2C02902/ASSET/IMAGES/LARGE/JZ2C02902_0003.JPEG.
- [52] N. Riboni, F. Fornari, F. Bianchi, M. Careri, A simple and efficient Solid-Phase Microextraction – Gas Chromatography – Mass Spectrometry method for the determination of fragrance materials at ultra-trace levels in water samples using multi-walled carbon nanotubes as innovative coating, *Talanta.* 224 (2021) 121891. <https://doi.org/10.1016/j.talanta.2020.121891>.
- [53] Chayawan, Vikas, Quantum-mechanical parameters for the risk assessment of multi-walled carbon-nanotubes: A study using adsorption of probe compounds and its application to biomolecules, *Environmental Pollution.* 218 (2016) 615–624. <https://doi.org/10.1016/j.envpol.2016.07.045>.
- [54] A. Jakubus, K. Godlewska, M. Gromelski, K. Jagiello, T. Puzyn, P. Stepnowski, M. Paszkiewicz, The possibility to use multi-walled carbon nanotubes as a sorbent for dispersive solid phase extraction of selected pharmaceuticals and their metabolites: Effect of extraction condition, *Microchemical Journal.* 146 (2019) 1113–1125. <https://doi.org/10.1016/j.microc.2019.02.051>.
- [55] M. Paszkiewicz, M. Caban, A. Bielicka-Giełdoń, P. Stepnowski, Optimization of a procedure for the simultaneous extraction of polycyclic aromatic hydrocarbons and metal ions by functionalized and non-functionalized carbon nanotubes as effective sorbents, *Talanta.* 165 (2017) 405–411. <https://doi.org/10.1016/j.talanta.2016.10.049>.
- [56] A. Arcoleo, F. Bianchi, M. Careri, Helical multi-walled carbon nanotube-coated fibers for solid-phase microextraction determination of polycyclic aromatic hydrocarbons at ultra-trace levels in ice and snow samples, *J Chromatogr A.* 1631 (2020) 461589. <https://doi.org/10.1016/j.chroma.2020.461589>.

- [57] T.E. Wang, M. Guo, W.L. Song, Y. da Zhang, X.Z. Du, A new nitrogen-containing carbon nanoparticle coated stainless steel fiber for selective solid-phase microextraction of ultraviolet filters, *Analytical Methods*. 7 (2015) 3385–3394. <https://doi.org/10.1039/C5AY00009B>.
- [58] W.H. Chung, S.H. Tzing, W.H. Ding, Dispersive micro solid-phase extraction for the rapid analysis of synthetic polycyclic musks using thermal desorption gas chromatography–mass spectrometry, *J Chromatogr A*. 1307 (2013) 34–40. <https://doi.org/10.1016/J.CHROMA.2013.07.074>.
- [59] S. Vandekar, R. Tao, J. Blume, A Robust Effect Size Index, *Psychometrika*. 85 (2020) 232–246. <https://doi.org/10.1007/S11336-020-09698-2/FIGURES/4>.
- [60] E. Becerril, J.P. Lamas, L. Sanchez-Prado, M. Llompert, M. Lores, C. Garcia-Jares, Analysis of regulated suspected allergens in waters, *Talanta*. 83 (2010) 464–474. <https://doi.org/10.1016/J.TALANTA.2010.09.038>.
- [61] A. Godayol, E. Besalú, E. Anticó, J.M. Sanchez, Monitoring of sixteen fragrance allergens and two polycyclic musks in wastewater treatment plants by solid phase microextraction coupled to gas chromatography, *Chemosphere*. 119 (2015) 363–370. <https://doi.org/10.1016/J.CHEMOSPHERE.2014.06.072>.
- [62] E. Fries, W. Püttmann, Analysis of the antioxidant butylated hydroxytoluene (BHT) in water by means of solid phase extraction combined with GC/MS, *Water Res*. 36 (2002) 2319–2327. [https://doi.org/10.1016/S0043-1354\(01\)00453-5](https://doi.org/10.1016/S0043-1354(01)00453-5).

Chapter 4 | Development of a magnetic Metal-Organic Framework composite for the dispersive micro solid-phase extraction of polycyclic aromatic hydrocarbons from water samples

4.1. Introduction

4.1.1. Polycyclic aromatic hydrocarbons

Polycyclic aromatic hydrocarbons (PAHs) are organic compounds constituted of a minimum of two aromatic rings fused together. Depending on the production mechanism and on how they enter the environment, PAHs can be classified in three main categories [1–3]:

- **Pyrogenic:** most PAHs present in the environment are pyrogenic. They are produced by the incomplete combustion of organic matter at temperatures typically higher than 350 °C.
- **Petrogenic:** this class of PAHs have formed through over million years through the geochemical conversion of organic matter at temperatures of 100–150 °C and high pressures. The result of this conversion is what today it is called fossil fuels, i.e., petroleum and coal. Petrogenic PAHs have a relative high fraction of alkylated compounds in respect to pyrogenic and biogenic ones.
- **Biogenic:** biogenic PAHs represent the minoritarian fraction. They are produced by the biological conversion of substrates of various nature. Perylene and retene are examples of biogenic PAHs [4].

Besides the production mechanisms, both natural processes and anthropic activities can introduce these compounds into the environment [1]. Pyrogenic PAHs can enter the ecosystems, e.g., *via* volcanic eruptions or wildfires, as well as domestic and industrial activities that involve combustion. Erosion in proximity of coal and petroleum reservoirs or accidental leakages represents examples on how petrogenic PAHs can access the environment.

PAHs are ubiquitous, as that they can be found in all environmental matrices, and persist into the environment, meaning that they resist to degradation processes

and are subjected to bioaccumulation phenomena. With the Regulation EU 2019/1021^{xxxvi}, the European Union included PAHs in the persistent organic pollutants (POPs) list. The Regulation aims at protecting human health and the environment by banishing or restricting the use and the emission in the environment of such substances. The need for regulation and monitoring in environmental matrices is due to their well-known adverse effects on the health of humans and other organisms. In fact, many PAHs are suspected or certainly carcinogenic, mutagenic, or toxic for the reproductive system^{xxxvii}. Given their persistence in the environment and the adverse effects they have on health, the United States Environmental Protection Agency (US-EPA) included 16 PAHs in the Priority Pollutants List (Table 4.1), published at 40 CFR Part 423^{xxxviii}.

Table 4.1. PAHs included in the US-EPA Priority Pollutants List along with their classification as carcinogenic, mutagenic or reprotoxic according to the Regulation (EC) 1272/2008^{xxxix}.

PAH	Carcinogenic	Mutagenic	Reprotoxic
naphthalene	<u>2</u> ^a	n.n. ^b	n.n.
acenaphthylene	n.n.	n.n.	n.n.
acenaphthene	n.n.	n.n.	n.n.
fluorene	n.n.	n.n.	n.n.
phenanthrene	2	n.n.	n.n.
anthracene	1B/2	n.n.	n.n.
fluoranthene	n.n.	n.n.	n.n.
pyrene	n.n.	n.n.	n.n.
benz[<i>a</i>]anthracene	<u>1B</u>	n.n.	n.n.
chrysene	<u>1B</u>	<u>2</u>	n.n.
benzo[<i>b</i>]fluoranthene	<u>1B</u>	n.n.	n.n.
benzo[<i>k</i>]fluoranthene	<u>1B</u>	n.n.	n.n.
benzo[<i>a</i>]pyrene	<u>1B</u>	<u>1B</u>	<u>1B</u>
indeno[1,2,3- <i>c,d</i>]pyrene	2	n.n.	n.n.
dibenz[<i>a,h</i>]anthracene	<u>1B</u>	n.n.	n.n.
benzo[<i>g,h,i</i>]perylene	n.n.	n.n.	n.n.

^a underlined: the classification has been harmonized; plain: the classification has only been notified. ^b not notified.

PAHs included in the Priority Pollutants Lists are intended to be monitored specifically in water bodies, as evidence suggests that there are several ways PAHs

^{xxxvi} <http://data.europa.eu/eli/reg/2019/1021/oj> (accessed 05/08/2022).

^{xxxvii} <https://echa.europa.eu/it/home> (accessed 05/08/2022).

^{xxxviii} <https://www.ecfr.gov/current/title-40/chapter-I/subchapter-N/part-423> (accessed 05/08/2022).

^{xxxix} <http://data.europa.eu/eli/reg/2008/1272/oj> (accessed 05/08/2022).

can reach water bodies, including run-offs, deposition from the atmosphere, and direct discharge [5,6]. In addition, since water is one of the main supplies for the sustenance of agriculture and husbandry, the use of contaminated water poses a risk not only towards the exposure of water itself, but also towards food deriving from those activities [6].

4.1.2. Magnetic dispersive micro solid-phase extraction

The detection and quantitation of PAHs in environmental matrices is challenging due to their low concentration levels. To reach the required sensitivity for the quantitation of such analytes, official analytical methodologies, such as the EPA Method 610^{xl}, typically involves the treatment of large volumes of sample with chlorinated solvents.

With the advent of Green Analytical Chemistry [7], new challenges have been posed in the development of analytical methodologies paying attention to their impact on the environment. In this context, reduction of sample size, utilization of reduced amounts of solvents, and integration of multiple steps of the analytical process (e.g., preconcentration and clean-up), represent few valuable strategies to meet the requirements of Green Analytical Chemistry. The abovementioned approaches are the core of miniaturized sample preparation techniques.

Dispersive micro solid-phase extraction (D- μ SPE) [8] can be considered the miniaturized version of the sample preparation technique formerly introduced by Michelangelo Anastassiades and coworkers in 2003 [9]. In D- μ SPE (Figure 4.1):

1. the sorbent is dispersed into the liquid sample and, after an adequate amount of time has elapsed,
2. it is recovered (e.g., by centrifugation) and the supernatant is discarded.
3. An appropriate amount of organic solvent is added to elute the analytes and, after the sorbent has been removed,
4. the extract is subjected to analysis.

^{xl} <https://www.epa.gov/sites/files/2015-10/documents/method610.pdf> (accessed 06/08/2022)

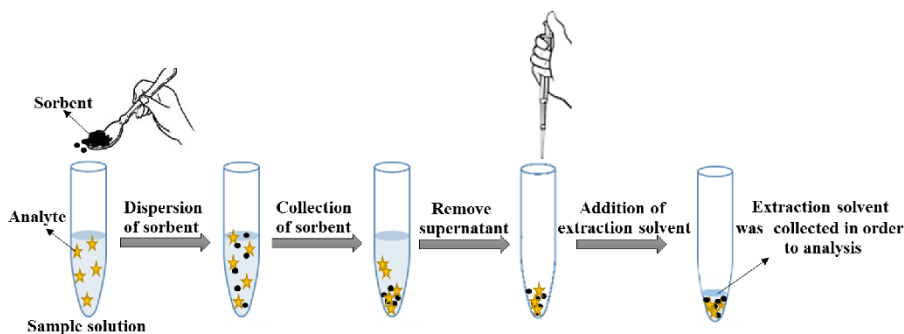


Figure 4.1. Graphical representation of the main steps of the D- μ SPE procedure. Reprinted from [10].

Compared to other sample preparation techniques, such as liquid-liquid extraction and solid phase extraction, D- μ SPE offers the advantage of smaller sample size and solvent consumption, as well as a reduced extraction time [8]. The latter is because the dispersion of the sorbent increases the surface contact between the sorbent itself and the sample solution and the eluent. The most critical step in D- μ SPE is sorbent recovery [8]. A poor sorbent recovery during the extraction phase results in analyte loss, whereas during the elution step a poor sorbent separation results in a turbid extract which is problematic for further instrumental analysis. Within this frame of reference, in 1999, Mirka Šafaříková and Ivo Šafařík brought substantial improvements [11]. In fact, their contribution is the first reporting the use of a magnetic sorbent in sample preparation in the field of analytical chemistry: this means that the sorbent can be easily recovered just by the application of an external magnetic field.

The conjugation between the magnetic properties of magnetic nanoparticles (NPs) and the structural characteristics of nanomaterials (e.g., high surface area, porosity, and tunable affinity) lead to a generation of magnetic sorbents with high affinity toward target analytes. Few examples include carbon nanotubes, silica, graphene oxide, molecularly imprinted polymers, and ionic liquids [12,13]. Within this frame of reference, such magnetic sorbents allowed the development of analytical methodologies based on the magnetic dispersive micro solid-phase extraction (MD- μ SPE).

4.1.3. Metal-Organic Frameworks

Metal-Organic Frameworks (MOFs) are porous crystals originating by the self-assembly of polytopic organic ligands and metallic nodes [14,15]. The inorganic part of MOFs can be either ions or polymetallic clusters, and it is the so-called secondary building unit (SBU) [16]. The proper selection of the SBU and the organic

ligands allows for the design of MOFs with unique characteristics in terms of porosity of the crystals, and geometry and polarity of the cavities [17]. As easily tunable, MOFs can be specifically designed to best fit their field of application, ranging from catalysis to gas storage and fractionation [18].

4.1.3.1. MOFs as sorbents for sample preparation

Given the abovementioned characteristics, is it not surprising that MOFs have found application also in the framework of analytical chemistry. Examples of applications in this field include sorbents for sample preparation, stationary phases for gas chromatography, and recognition elements in sensing [19–21].

In the context of sample preparation, the interactions between the sorbent and the analyte should be maximized. This can be achieved, for instance, by introducing functional groups on the ligand that can interact with the analytes of interest. This strategy was followed by Iván Taima-Mancera and coworkers [22], who used variously substituted terephthalic acids for the synthesis of UiO66 for the D- μ SPE of a pool of contaminants from water samples. Another valuable strategy is pillaring [23,24]; in pillared MOFs, parallel bidimensional lattices are held together by bridging ligands, oriented perpendicularly to the lattices: in other words, the overall structure resembles the floors of a skyscraper. Pillared MOFs are heteroleptic, as they involve two different ligands in the overall structure. The presence of multiple ligands offers few possibilities in the frame of designing MOFs as suitable sorbent materials capable of targeting analytes with different polarity and size, because:

- Ligands with different functional groups can be selected.
- The careful choice of the two different ligands allows the modulation of the framework flexibility.

Providencia González-Hernández and coworkers [25] took advantage of the pillaring strategy, using the MOF CIM-81 as a sorbent for the D- μ SPE of different personal care products from wastewater. Another application of pillared MOFs in sample preparation is the one reported by Federica Bianchi and coworkers [26], who used the mixed-ligand MOF PUM210 for the solid-phase microextraction of PAHs from environmental water samples.

4.1.3.2. Fabrication of magnetic MOF composites for MD- μ SPE

The vast majority of known MOFs are not magnetic themselves. Therefore, most of MOFs are not directly applicable in MD- μ SPE. In order to make them magnetic, few strategies have been proposed. All of them have in common the incorporation or

the decoration of MOF crystals with magnetic NPs: the resulting material is, therefore, a composite.

A summary of the available approaches for the synthesis of magnetic MOF composites is given below [19,27]:

- **Direct mixing:** this is by far the most straightforward and simple to execute among the fabrication strategies. MOF and NPs are synthesized separately, and the magnetic composite is obtained by mixing the two components in solution. In the composite, NPs and MOF interact with each other most likely via electrostatic forces.
- ***In situ* magnetic nanoparticles growth:** MOF is synthesized separately and, then, introduced in the reaction environment for the preparation of magnetic NPs. The magnetic NPs will form on the surface of the MOF crystals and/or inside the cavities, depending both on the size of the cavities and the size of NPs. In order to actuate this synthetic protocol, MOF must be stable under the reaction conditions for the synthesis of NPs.
- ***In situ* MOF growth:** this synthetic way is the opposite of the one discussed previously. In this case, NPs are synthesized separately and, then, introduced in the reaction environment containing the salt and the ligand for the synthesis of the MOF. NPs can be functionalized with a suitable spacer able to coordinate the SBU: if it is the case, the fabrication strategy can be called encapsulation. Two subgroups of this strategy can be identified:
 - **Single step:** after NPs are added to the reaction mixture for MOF synthesis, the components are left reacting until the endpoint. The spatial distribution of magnetic NPs inside the MOF crystals can be modulated by adding them after a proper amount of time has elapsed, or by making multiple additions.
 - **Layer-by-layer:** this strategy involves sequential self-assembly cycles. NPs are firstly introduced in an environment containing only the inorganic part of the MOF, so that the SBU can be coordinated. Thereafter, the NPs are put in an environment containing the organic ligand. The abovementioned procedure, describes a growth cycle. The composite can be subjected to several growth cycles, until the desired thickness of the crystalline coating around the NPs is reached.

- **MOF carbonization:** this synthetic path is suitable only for MOFs whose SBU converts into zerovalent metallic NPs that are magnetic. This is the case, for instance, of the ZIF-67, whose carbonization produces cobalt nanoparticles trapped in a porous carbonaceous residue [28]. Carbonization takes place under inert atmosphere and at high temperature, so combustion do not occur.

4.1.3.3. PUM198

PUM198 is a zinc-based MOF, firstly synthesized at the University of Parma by Davide Balestri and coworkers [29]. PUM198 is the result of the combination of a bispyridine-bisamide ligand (L1, [30]), terephthalic acid, and $\text{Zn}(\text{NO}_3)_2$ in N,N' -dimethylformamide (DMF).

PUM198 is a pillared MOF, in which 2D planes, containing Zn^{2+} ions and terephthalate ions, are connected by the bridging L1 ligand. The SBU has formula $[\text{Zn}_2(\kappa^1\text{-COO})_2(\mu\text{-COO})_2(\text{L1})_4]$ and the four carboxylate anions behave differently:

- two of them act as bridging ligands between the Zn^{2+} ions in the SBU,
- the other two are monodentate and interact *via* hydrogen bonding with the amide groups of the L1 ligand.

The framework is doubly interpenetrated and it is microporous, with a pore size of $5.6 \text{ \AA} \times 17.4 \text{ \AA}$. After the synthetic procedure, PUM198 channels are filled with the reaction solvent, i.e., DMF. The removal of the electronic density of DMF results in a 28% void volume *per* cell unit (corresponding to 2125 \AA^3 , Figure 4.2).

The flexibility of the framework, as well as the presence of aromatic rings both on terephthalate ions and L1 ligands makes PUM198 a potential sorbent material for aromatic analytes, such as PAHs.

As in the case of carbon nanotubes presented in **Chapter 3**, this research activity is in the framework of the development of miniaturized extraction techniques. The aim is the fabrication of a magnetic composite based on PUM198 for the MD- μ SPE of the 16 US-EPA PAHs from environmental water samples. The most critical factors affecting the extractive performance were optimized with a focus on reducing the sample size and the solvent needed for elution. The validated method was successfully utilized for the determination of PAHs in underground water samples, collected in a contaminated site in the province of Parma.

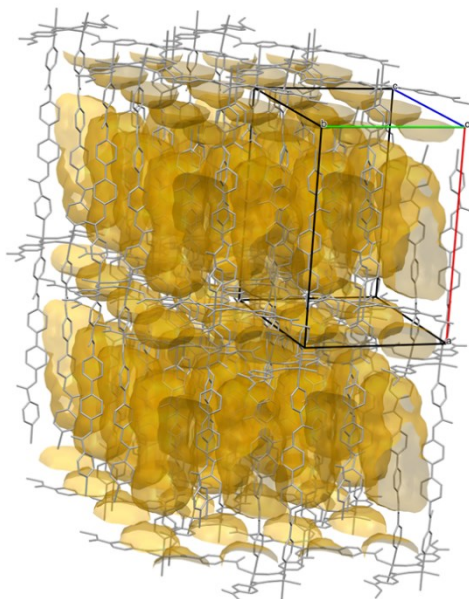


Figure 4.2. Representation of the voids (depicted in yellow) and the pores of PUM198. The channels of PUM198 runs along the crystallographic *a* axis (depicted in red). Reprinted from [31].

4.2. Materials and methods

4.2.1. Chemicals

Acetonitrile (> 99.9%), acetone (> 99.5%), $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$, *n*-hexane, toluene, $\text{Zn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ (all > 99%), (3-aminopropyl)triethoxysilane (APTES), terephthalic acid (both > 98%), FeCl_3 (> 97%), HCl (> 37% in H_2O), NH_4OH (33% in H_2O) were obtained from Sigma-Aldrich (Milan, Italy). DMF (99.8% purity) was from VWR International (Milan, Italy). Methylene chloride (> 99.9%) and methyl alcohol (> 99.8%) were obtained from Honeywell International Inc. (Morristown, USA). Cyclohexane (> 99.8%) was obtained from Carlo Erba Reagents (Milan, Italy), whereas *i*-propyl alcohol (> 99.98%) was obtained from Fisher Chemicals (Milan, Italy).

PAH-Mix 16 (EPA 550, EPA 610) containing naphthalene (N), acenaphthylene (Acy), acenaphthene (Ace), fluorene (F), phenanthrene (P), anthracene (Ant), fluoranthene (Fl), pyrene (Pyr), benz[*a*]anthracene (BaAnt), chrysene (Chr), benzo[*b*]fluoranthene (BbFl), benzo[*k*]fluoranthene (BkFl), benz[*a*]pyrene (BaPyr), indeno[1,2,3-*c,d*]pyrene (IPyr), dibenz[*a,h*]anthracene (DBahAnt), benzo[*g,h,i*]perylene (BghiPer) (nominal concentration: 100 mg/L in acetonitrile), and PAH-Dx-Mix 16 containing their perdeuterated surrogates (nominal

concentration: 10 mg/L in acetonitrile) were obtained from NEOCHEMA GmbH (Bodenheim, Germany).

Working solutions were prepared by proper dilution of PAH-Mix 16 and PAH-Dx-Mix 16 in acetonitrile. Diluted solutions were kept at 4 °C in the dark until use.

4.2.2. Preparation and characterization of the sorbent

4.2.2.1. Synthetic procedures

4.2.2.1.1. Iron oxide nanoparticles. Iron oxide nanoparticles (IONPs) were prepared by coprecipitation [32]. Under nitrogen atmosphere, 1.00 g of $\text{FeCl}_2 \cdot 4\text{H}_2\text{O}$ and 1.56 g of FeCl_3 were dissolved in 12 mL of an aqueous solution of HCl 2 M. The reaction mixture was brought to 85 °C under constant and fierce stirring and 10 mL of aqueous NH_4OH (33%) were added dropwise. The mixture was let react for 1 h and let cool at room temperature. The black precipitate was collected by magnetic decantation, and it was washed i) with distilled water until neutrality and ii) two times with 25 mL of methanol. The product was stored in methanol under nitrogen atmosphere until use.

4.2.2.1.2. Silanization. IONPs were encapsulated with APTES [33]. IONPs were dried under a nitrogen flow overnight. One hundred milligrams of dried IONPs were dispersed in 15 mL of anhydrous toluene by vigorous stirring at room temperature under nitrogen atmosphere. After 45 minutes had elapsed, 150 μL of APTES were introduced and the mixture was brought at 120 °C and let react for 18 h. After cooling at room temperature, the product was collected by magnetic decantation and washed i) two times with 25 mL of *n*-hexane, ii) two times with 25 mL of acetone, and iii) two times with 25 mL of methanol. The product was stored in methanol until use.

4.2.2.1.3. PUM198. The synthesis followed the procedure reported on the original publication [29]. In a 70 mL glass vessel were introduced 41 mL of DMF, 120 g of $\text{Zn}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$, 66 mg of terephthalic acid, and 80 mg of the L1 ligand [30]. The vessel was closed with a screw cap and maintained still for 4 days at 80 °C. After slow cooling at room temperature, yellow crystals were collected by Büchner filtration and washed with DMF. The crystals were stored under DMF until use.

4.2.2.1.4. Magnetic composite. The grafting was carried out by following the procedure reported by Yuling Hu and coworkers [34], with some modifications. PUM198 crystals were dried on a vacuum filter and encapsulated IONPs were dried under a nitrogen flow prior the use. Twenty milligrams of PUM198 crystals and 6.5 mg of encapsulated IONPs were dispersed in 3.5 mL of DMF. The mixture was

stirred fiercely at 120 °C for 12 h. After cooling at room temperature, the product was collected with a magnet and washed two times with 5 mL of DMF. The composite sorbent was activated by following a solvent exchange procedure. The supernatant DMF was removed, and the composite was soaked in acetone for 24 h. Thereafter, acetone was replaced with methylene chloride. After additional 24 h had elapsed, the supernatant was discarded, and the sorbent was kept at 50 °C and under vacuum overnight to remove the residual methylene chloride.

4.2.2.2. Characterization

Vibrational spectra were recorded in the 400–4000 cm⁻¹ region using a Spectrum Two FT-IR Spectrometer (Perkin Elmer, Milan, Italy) by operating in Attenuated Total Reflectance (ATR) mode. The collection of the Powder X-Ray Diffraction (PXRD) traces was carried out in Bragg-Brentano geometry using a Cu-K α radiation on a Rigaku Smartlab XE diffractometer (Tokyo, Japan) equipped with a 2D Hypix3000 solid-state detector.

Morphological investigation and Energy Dispersive Spectroscopy (EDX) were carried out on the magnetic composite with a Quanta FEG 250 (Thermo Fisher Scientific, Waltham, USA) Environmental Scanning Electron Microscope (ESEM) equipped with a XFlash 6 | 30 (Bruker, Billerica, USA) detector.

The magnetic properties of the IONPs and the composite were investigated by Vibrating Sample Magnetometry (VSM) using the VSM 7400 instrument by Lake Shore (Westerville, USA). The maximum magnetic field that was applied was 1.8 T.

4.2.3. Optimization of the extraction procedure

Optimization of the MD- μ SPE procedure was carried out by performing extraction experiments on aqueous solutions at the concentration of 500 ng/L of each analyte. After extraction, 50 μ L of extract were recovered. The extract was spiked with the internal standard (IS) solution containing perdeuterated PAHs so that the final concentration was 10 μ g/L. This last step was for compensating instrumental fluctuations. One microliter of spiked extract was submitted to analysis. The response variable was defined as the ratio between the area of each analyte and the area of the respective perdeuterated surrogate (Equation 4.1):

$$\text{Area ratio} = \frac{A_{\text{analyte}}}{A_{\text{IS}}}$$

Equation 4.1

Extraction experiments were performed by means of an ultrasonicated bath DU-32 (Argolab, Modena, Italy). The ultrasound frequency was 40 kHz.

4.2.3.1. Preliminary evaluation of the elution solvent

The effect of acetonitrile, *i*-propyl alcohol, cyclohexane, and a mixture 10% v/v of *i*-propyl alcohol in cyclohexane was evaluated in terms of signal intensity, number of extracted analytes and repeatability. Three milliliters of sample were processed with 10 mg of sorbent and 130 μ L of solvent. Extraction and elution time were both set at 10 minutes. The temperature was left at ambient conditions. Three independent extractions ($n = 3$) were carried out for each solvent.

Differences in the average responses for each solvent were highlighted with one-way ANOVA. Homoscedasticity and normality were tested beforehand with Bartlett's test and Shapiro-Wilk test, respectively. For statistically significant results, the effect size was estimated in terms of Cohen's η^2 [35] and, *post-hoc* multiple pairwise comparisons were carried out with Student's *t*-tests with a Bonferroni correction. The confidence level was 95%.

4.2.3.2. Screening for the important factors

Seven factors ($k = 7$) were investigated at two levels within their respective experimental domain according to a Plackett-Burman Design (PBD) with $N = 12$ runs. The following factors were studied: sorbent amount (*A*, 3–12 mg), sample volume (*B*, 2–5 mL), solvent volume (*C*, 70–150 μ L), extraction time (*D*, 5–15 min), elution time (*E*, 5–15 min), ultrasonication power (*F*, 48–98 W), and temperature (*G*, ice bath–ambient condition). The design was saturated with four dummy variables.

A first-order model was postulated. A factor was considered critical if the absolute value of its regression coefficient was greater than the semi-amplitude of its 95% confidence interval. The pure experimental variance was estimated from the regression coefficients of the dummy variables.

4.2.3.3. Optimization of the critical factors

Three main factors ($k = 3$) were investigated at three levels within their respective experimental domain according to a Box-Wilson Central Composite Design (CCD): sample volume (X_1 , 1–5 mL), solvent volume (X_2 , 50–150 μ L), and sorbent amount (X_3 , 2–14 mg). In total, $N = 26$ experiments were carried out, including $n_0 = 6$ experiments in the center of the experimental domain to estimate the pure experimental variance.

A full second-order model was postulated. A coefficient was considered significant if the absolute value of its regression coefficient was greater than the semiamplitude of its 95% confidence interval. Each model was evaluated in terms of fraction of explained variance (R^2) and predictive capability in leave-one-out cross-validation (Q^2). The validity was evaluated by carrying out a lack-of-fit F -test ($\alpha = 0.01$).

The global optimal conditions were identified according to Derringer's method [36]. Single desirability functions d_i were defined for each analyte according to Equation 4.2:

$$d_i = \begin{cases} 0, & y_i < L_i \\ \frac{y_i - L_i}{U_i - L_i}, & \text{otherwise} \end{cases}$$

Equation 4.2

L_i is the upper limit of the 95% confidence interval of the minimum predicted response and U_i is the maximum predicted response within the experimental domain. The optimal conditions were in correspondence of the maximum global desirability (Equation 2.42), computed as unweighted geometric mean of the single desirability functions. The maximum was found with a derivative-free search algorithm.

4.2.4. Operating procedure and instrumental conditions

4.2.4.1. Sample preparation

Five milliliters of sample were introduced into a centrifuge glass test tube and spiked with 10 μL of IS solution (final concentration: 200 ng/L). Thereafter, 5.4 (± 0.1) mg of sorbent were added, and the sample was ultrasonicated for 5 min. The sorbent was collected by magnetic decantation and the supernatant was discarded. The elution step was carried out by adding 50 μL of acetonitrile to the sorbent and the suspension was ultrasonicated for 15 min. Before each ultrasonication step, the test tube was vortexed for 10 s. The ultrasonication power was set at 48 W.

After magnetic decantation, the extract was transferred into 2 mL clear glass vials equipped with a conical high-recovery glass insert. One microliter of extract was addressed to the GC injector with the aid of a PAL COMBI-xt autosampler (CTC Analytics AG, Zwingen, Switzerland) equipped with a 10 μL syringe.

4.2.4.2. GC-MS

Instrumental analyses were carried out with a HP 6890 Series Plus gas chromatograph hyphenated with an MSD 5973 mass spectrometer (both by Agilent Technologies, Milan, Italy).

The split/splitless injector temperature was 300 °C. The injection was executed in pulsed splitless mode (pressure pulse: 20.0 psi; pulse duration: 0.3 min; purge flow: 1.2 mL/min @ 1.20 min) into a 5190-2293 ultra-inert liner (Agilent Technologies). An Rxi-5Sil MS capillary column (30 m length × 0.25 mm i.d., 0.25 μm film thickness; Restek, Bellefonte, USA) was used for the chromatographic separation with the following temperature program: 70 °C, held for 0.50 min, 15 °C/min to 290 °C, held for 4.00 min (runtime: 19.17 min). The carrier gas was helium (constant flow rate: 1.0 mL/min). The transfer line was maintained at 280 °C.

The single quadrupole mass spectrometer was operated in electron ionization mode (EI; 70 eV). The ion source and the quadrupole were held at 230 and 150 °C, respectively. Mass spectra were registered in selected ion monitoring (SIM) mode (dwell time: 30 ms; electron multiplier voltage: 1.153 kV). The acquisition started 3.80 min after the injection. One microliter of a solution containing both PAHs and perdeuterated PAHs (concentration: 1 mg/L) in acetonitrile was injected and the mass range 45-350 m/z was recorded to properly set the time scheduled SIM program (Table 4.4).

Table 4.4. Time scheduled SIM program.

Time (min)	Compound	Monitored m/z ratios
3.80–4.90	N	128
	N-d ₈ ^a	136
4.90–7.60	Acy	152
	Ace	153
	Acy-d ₈	160
	Ace-d ₁₀	162
7.60–8.50	F	166
	F-d ₁₀	176
8.50–10.50	P, Ant	176
	P-d ₁₀ , Ant-d ₁₀	188
10.50–12.50	Fl, Pyr	202
	Fl-d ₁₂ , Pyr-d ₁₂	212
12.50–14.50	BaAnt, Chr	228
	BaAnt-d ₁₂ , Chr-d ₁₂	240
14.50–16.50	BbFl, BkFl, BaPyr	252
	BbFl-d ₁₂ , BkFl-d ₁₂ , BaPyr-d ₁₂	264
16.50–19.17	IPyr, BghiPer	276
	DBahAnt	278
	IPyr-d ₁₂ , BghiPer-d ₁₂	288
	DBahAnt-d ₁₄	292

^a perdeuterated surrogates are marked as such.

4.2.5. Method validation

The MD- μ SPE-GC-MS method was validated by working under the optimized conditions according to the EURACHEM guidelines [37]. The validation study was carried out with matrix-matched standards that were subjected to the whole analytical process. Drinking water was used as blank matrix.

Limits of detection and limits of quantitation (LODs and LOQs, respectively) were evaluated by estimating the standard deviation s_0 of $m = 10$ independent replicated measurements of blank matrix fortified with 30 ng/L of each analyte. The average signal of the blank y_0 was estimated by submitting to analysis $n = 4$ procedural blanks. Detection and quantitation limits expressed in the signal domain, were calculated according to Equation 2.43 and 2.44, respectively, considering that $n = 2$ replicated measurements will be averaged when reporting the results. The corresponding concentrations were calculated by their projection on the x axis. Calibration functions were evaluated on $k = 9$ levels in the LOQ–600 ng/L range (n

= 2 independent replicated measurements *per* level) for all the analytes. The intercept was tested for significance with the Student's *t*-test ($\alpha = 0.05$). Linearity and validity were evaluated, respectively, with Mandel's test and lack-of-fit test ($\alpha = 0.01$).

Repeatability and intermediate precision were evaluated in a single study by carrying out independent duplicated measurements *per* day for five days by five different operators. Repeatability and intermediate precision standard deviations were estimated according to Equation 2.48 and 2.49 and expressed as relative standard deviations ($RSD\%$). ANOVA was applied also to assess whether the results obtained under intermediate precision conditions were significantly different among each other ($\alpha = 0.05$). Trueness was expressed as spike recovery rate ($RR'\%$, Equation 2.52) by submitting $n = 10$ independent replicated measurements to the whole analytical process. Precision and trueness were evaluated at 60, 250, and 550 ng/L.

Enrichment capabilities of the composite sorbent and bare IONPs were evaluated in terms of enrichment factors (EFs). EFs were calculated according to Equation 2.53 by submitting to MD- μ SPE aqueous solutions containing the analytes at the concentration of 500 ng/L and by injecting 1 μ L of a solution of the analytes in acetonitrile at 50 μ g/L. The EFs of the sorbent composite and bare IONPs were compared ($n = 3$ independent replicated measurements were performed). The responses were adjusted by the corresponding response factor as they were carried out at different concentration levels.

4.2.6. Analysis of real samples

Nineteen samples were collected near a company devoted to the production of tar and other chemicals in the province of Parma (Emilia-Romagna, Italy). Samples were stored in 20 mL amber glass vials sealed with polytetrafluoroethylene septa and kept in the dark at 4 °C until analysis. Samples were submitted to analysis with the proposed analytical method after proper dilution in duplicate.

4.2.7. Software

GC-MS data were handled with the HP ChemStation (Agilent Technologies) software. Multilinear regression and multicriteria optimization were carried out by using custom scripts in MATLAB environment (v. R2019a, Mathworks, Massachusetts, USA).

4.3. Results and discussion

4.3.1. Characterization of the magnetic composite

As reported earlier, PUM198 can be considered a good candidate for adsorbing PAHs for sample preparation purposes. Even if the framework can establish nonspecific interactions with potential guests, the presence of aromatic rings in the ligands can promote π - π interactions and London dispersion forces with PAHs.

In order to fabricate a magnetic composite based on PUM198, a post-synthetic protocol for its modification was followed:

1. PUM198 and IONPs were synthesized separately.
2. IONPs were modified with APTES to yield encapsulated IONPs with NH_2 groups exposed on their surface (IONPs@ NH_2).
3. IONPs@ NH_2 were grafted onto PUM198 crystals.

Vibrational spectra and diffraction traces of all the different intermediates are depicted in Figure 4.3. The vibrational spectrum of IONPs showed the distinctive bands of Fe_3O_4 systems, i.e., a broad weak band around 3400 cm^{-1} and a stronger one around 550 cm^{-1} ascribable to O-H and Fe-O stretching vibrations [33]. The presence of absorption bands around 1220 cm^{-1} and 990 cm^{-1} , ascribable to the Si-O-Si and Si-O-Fe vibrations [38], respectively, in the IONPs@ NH_2 spectrum suggested that the encapsulation of IONPs with APTES was successful. The bands in the spectrum of PUM198 were at 1660 cm^{-1} and at 1670 cm^{-1} , related to κ^1 -COO- groups in the SBU and the amidic C=O of L1 [30], respectively. The spectrum of the magnetic composite comprised all the bands of PUM198 and IONPs. The band at 1660 cm^{-1} resulted slightly attenuated: this finding could be ascribed to the participation of the monodentate κ^1 -COO- groups at the frontier PUM198 crystals in grafting IONPs@ NH_2 .

This hypothesis was supported by the ESEM micrograph (Figure 4.4), showing IONPs@ NH_2 grafted on the coarse surface of PUM198 crystals. Yuling Hu and coworkers [34] found a similar result by grafting APTES-encapsulated magnetite on the surface of MOF-5 crystals by exploiting the free carboxylate groups of the MOF.

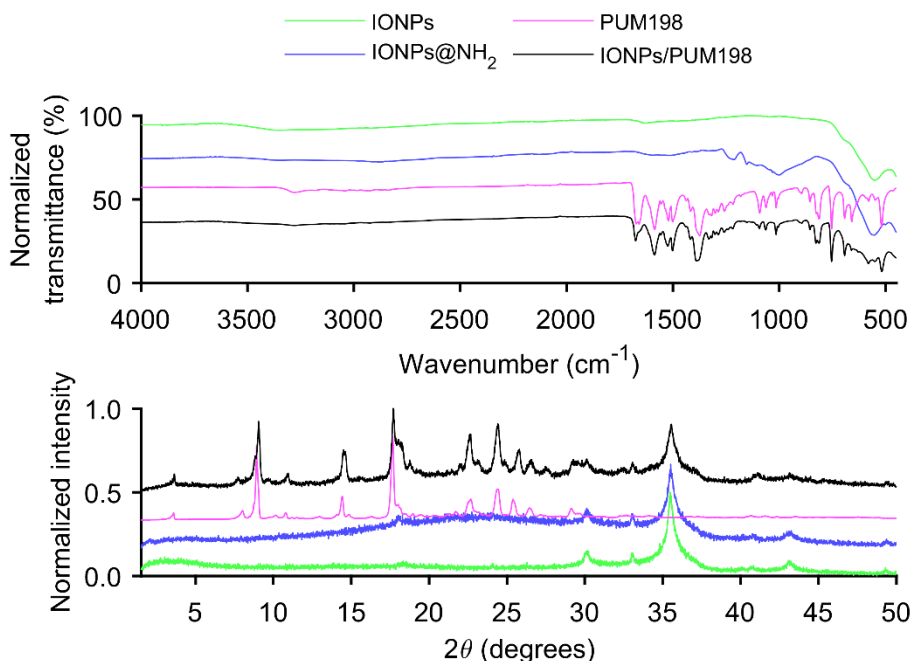


Figure 4.3. Vibrational spectra (top) and PXRD patterns (bottom) of all the intermediates (IONPs: green; IONPs@NH₂: blue; PUM198: magenta; magnetic composite: black). Each spectrum was shifted upwards or downwards for comparison purposes. Reprinted from [31].

The Si–O–Fe vibration band was not visible in the spectrum of the magnetic composite, but the presence of Si and other elements emerged by recording the EDX spectrum (Figure 4.4). In addition, the broad band around $2\theta \approx 23^\circ$ in the diffractograms of the magnetic composite and IONPs@NH₂ imputable to the amorphous APTES coating suggested that the encapsulation and the fabrication of the magnetic composite were successful.

All the diffraction traces of the intermediates containing IONPs showed an intense reflection at $2\theta \approx 36^\circ$, suggesting the successful grafting of the encapsulated IONPs on MOF crystals. In addition, the diffraction bands of PUM196 were also found in the magnetic composited diffractogram, suggesting that the crystal structure was retained through the synthetic procedure.

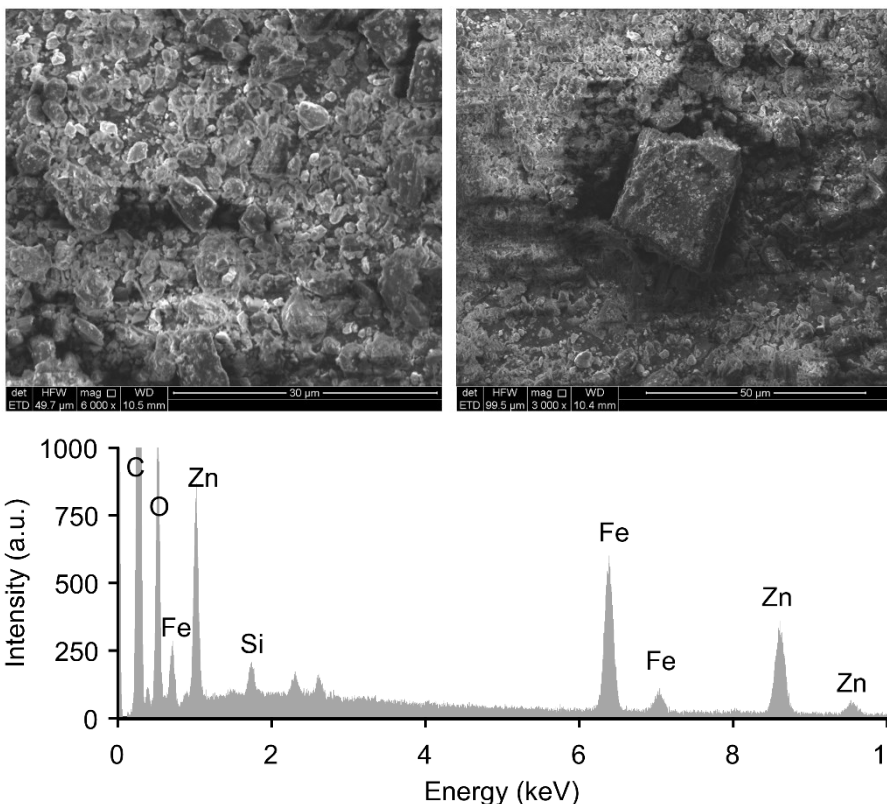


Figure 4.4. ESEM micrographs (top) and EDX spectrum (bottom). Reprinted from [31].

Results deriving from magnetometry revealed, as per the other characterization techniques, the successful fabrication of the magnetic composites. Magnetograms (Figure 4.5) suggested that the magnetic behaviors of both bare IONPs and the magnetic composite were predominantly superparamagnetic. The presence of bigger aggregates could be hypothesized due to the presence of a small hysteresis when low fields were applied. IONPs were characterized by a specific magnetization value of $39 \text{ A} \times \text{m}^2/\text{kg}$, whereas the specific magnetization value magnetic composite was $12 \text{ A} \times \text{m}^2/\text{kg}$ (considering the total mass of the sample), both at 1.8 T and at room temperature. The normalization of the specific magnetization value of the magnetic composite by the weight fraction of bare IONPs yielded to a value of $40 \text{ A} \times \text{m}^2/\text{kg}$, in accordance with the IONPs one. The decrease of the specific magnetization value evidenced the successful grafting of IONPs@NH₂ on PUM198 crystals.

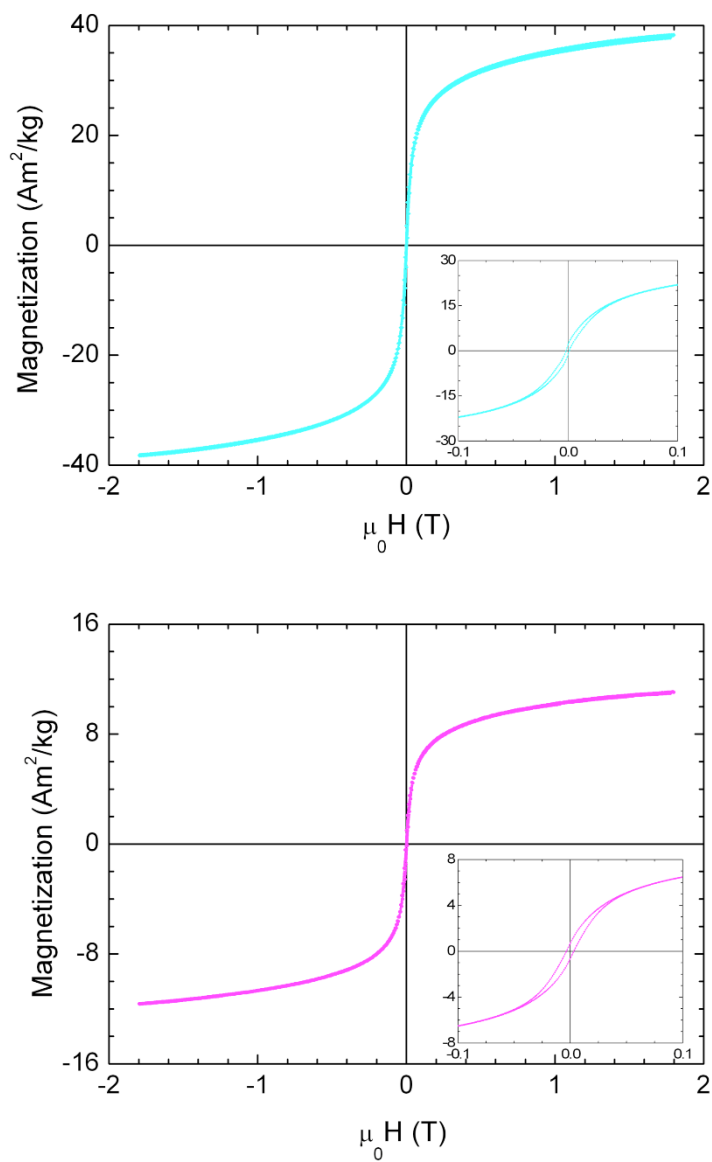


Figure 4.5. Magnetization loops measured at room temperature on bare IONPs (top) and the magnetic composite (bottom). The insets highlight the slight hysteresis when low fields are applied. Reprinted from [31].

4.3.2. Optimization of the MD- μ SPE procedure

4.3.2.1. Preliminary evaluation of the elution solvent

Solvent plays a key role in all the sample preparation techniques that involves a partitioning and/or an elution step. Aspects like polarity, volatility, and viscosity are crucial to provide a reliable chromatographic behavior and good extraction efficiency. Acetonitrile, isopropanol, cyclohexane, and a mixture 10% v/v of isopropanol in cyclohexane were tested in this study. The results are depicted in Figure 4.6. Statistically significant differences were found in the average responses for all the analytes ($p < 0.05$; Cohen's $\eta^2 \geq 0.77$).

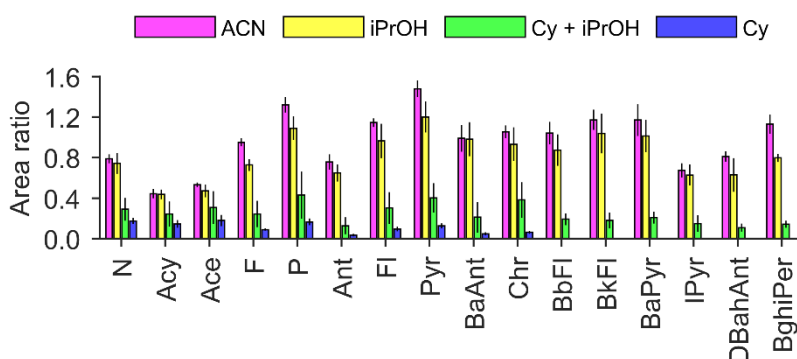


Figure 4.6. Solvent effect on the extraction of the analytes (ACN: acetonitrile; iPrOH: isopropanol; Cy: cyclohexane). Results are plotted as mean \pm one standard deviation ($n = 3$). Reprinted from [31].

Acetonitrile and isopropanol provided the best performance, with acetonitrile showing better or not significantly different extractive capability than isopropanol ($p > 0.05$, Bonferroni adjusted) and good repeatability ($RSD\% < 13\%$). Isopropanol, on the contrary, provided poor repeatability, with $RSD\% < 26\%$, most likely due to its viscosity obstructing the diffusion of the analytes.

The extracts in cyclohexane provided weak GC-MS responses and high-molecular weight PAHs were not recovered. This finding was unexpected, as cyclohexane is non-polar as the analytes of interest. This behavior could be explained by considering that cyclohexane is able to interact with the framework of PUM198 only *via* London dispersion forces, whereas isopropanol and acetonitrile can also interact *via* hydrogen-bonding. Another explanation could be ascribed to residual water coating the sorbent after the extraction step and in the higher lipophilicity of high-molecular weight PAHs [39]. These compounds could have remained preferably adsorbed rather than partitioning in water to reach cyclohexane. A similar behavior was observed by Angela Arcoletto and coworkers [5], who

experienced difficulties in the elution of PAHs with methylene chloride in the development of a microextraction by packed sorbent procedure. In that case, the presence of a small amount of isopropanol allowed the extraction of the high-molecular weight PAHs, somehow supporting the hypotheses stated earlier.

However, in this case, the addition of 10% v/v of isopropanol, did not significantly improve the recovery of the analytes ($p > 0.05$, Bonferroni adjusted), so acetonitrile was then selected for further method development.

4.2.3.2. Screening for the important factors

Factors to be investigated and their experimental domain were selected according to previous studies concerning the development of D- μ SPE-based analytical methods [39–41], preliminary experiments (data not shown), and the following considerations:

- Since a fine control of temperature was not possible, temperature was studied as a dichotomous variable.
- Extraction and elution times greater than 15 min were not considered to maintain a good balance between the analysis time and the time needed to process the samples.
- Salt addition was not investigated, as it has been reported that high concentration of Na⁺ could hinder the interactions between the analytes and the framework, worsening its performance [26,42,43].

As a multitude of factors was involved, the optimization was preceded by a screening step so that only the critical factors can be more finely studied during the optimization stage, and the experimental domain of the critical factors can be adjusted consequently, whereas unimportant ones can be fixed at a reasonable level.

For the sake of brevity, Ace, Pyr, and BaPyr coefficient plots were shown to exemplify the behavior of low-, mid- and high-molecular weight PAHs, respectively (Figure 4.7). The inspection of the coefficient plots revealed that three general behaviors could be outlined:

- The amount of sorbent had a positive effect on the extraction of low-molecular weight PAHs.
- The volume of treated sample and the eluent volume had a positive and negative effect, respectively, on the responses of both mid-molecular

weight analytes. The same behavior was also observed for high-molecular weight analytes.

- The amount of sorbent had a negative effect on the responses of high-molecular weight PAHs.

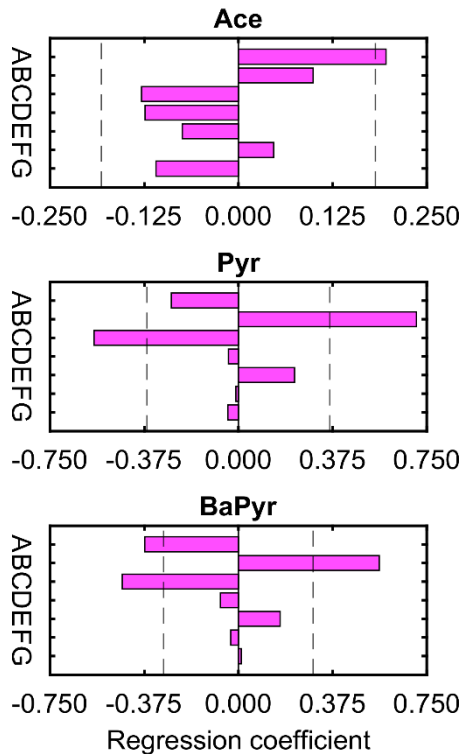


Figure 4.7. Coefficient plot of Ace, Pyr, and BaPyr. The sign of each regression coefficient (A: sorbent amount; B: sample volume; C: solvent volume; D: extraction time; E: elution time; F: ultrasonication power; G: temperature) defines how the response changes when the setting value of a given factor is switched from the lower to the upper limit of its experimental domain. The dashed vertical lines show the limits of the 95% confidence interval. Reprinted from [31].

For what concerned the remaining factors, they were not critical. Their value was set according to the sign of their coefficients. Extraction and elution times were fixed at 5 and 15 minutes, respectively, as they had a negative and a positive coefficient, respectively, for most of the analytes. Ultrasonication power was fixed at 48 W as its impact on the responses was very little, and ice was added to the ultrasonication bath to reduce the evaporation rate of the solvent and to prevent the loss of low-molecular weight PAHs.

4.2.3.3. Optimization of the critical factors

Prior the execution of the experiments for the estimation of the response surfaces, the experimental domains of the critical factors, i.e., sample volume, solvent volume, and sorbent amount, were adjusted according to the results that emerged from the screening stage:

- **Sample volume:** its effect was explored in the 1–5 mL range, to orientate method development towards smaller volumes.
- **Solvent volume:** its effect was explored in the 50–150 μL range, as all the analytes presented a negative coefficient in the screening stage. Volumes smaller than 50 μL were not considered as they were difficult to recover.
- **Sorbent amount:** its effect was explored in the 2–14 mg range, as low- and high-molecular weight PAHs showed an opposite behavior. Amounts outside this range were not considered both to limit sorbent consumption and to avoid a large error in the mass measurement.

The responses were logarithmically transformed [44], and the models were calculated. ANOVA showed that all the models were valid ($p > 0.01$), meaning that the error deriving from approximation is not significantly greater than the variance that it could be expected experimentally. The models were good in terms of explained variance and predictive capability in cross-validation, with $R^2 \geq 0.95$ and $Q^2 \geq 0.75$. For the sake of brevity, Ace, Pyr, and BaPyr surface plots were shown to exemplify the behavior of low-, mid- and high-molecular weight PAHs, respectively (Figure 4.8), whereas the regression models are reported in Table 4.5.

The results agreed with the main idea behind preconcentration: the greater the treated sample and the smaller the solvent volume, the highest the greater the preconcentration factor. In fact, as far as the main effect of sample volume and solvent volume are taken into consideration, they were present in most of the regression models with a positive and negative coefficient, respectively. The main effect of sorbent mass had a negative and a positive influence on high- and low-molecular weight PAHs, respectively. These findings agreed with those obtained earlier in the screening stage. Quadratic effects related to the amount of treated sample and sorbent used for the extraction were present with a negative coefficient. This could be ascribed to saturation phenomena involving the adsorption sites.

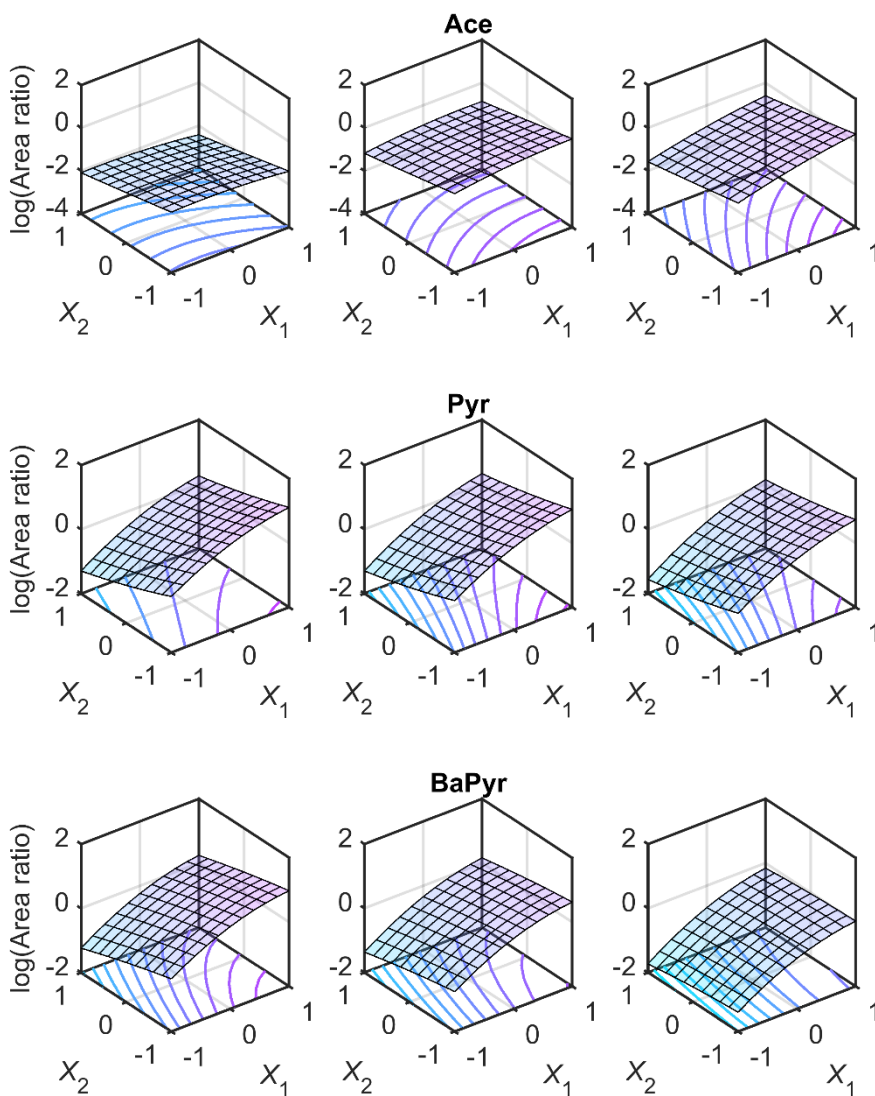


Figure 4.8. Response surfaces of Ace, Pyr, and BaPyr. The predicted response for each analyte is plotted as a function of sample volume and solvent volume (X_1 and X_2 , respectively) at three different levels of the amount of sorbent used for the extraction (X_3): -1 (left column), 0 (middle column), and +1 (right column).

The optimal conditions required to treat an aliquot of 5 mL of sample with 5.4 mg of sorbent and 50 μL . The global desirability corresponding to such experimental conditions was $D = 0.90$, with single desirability values $d_i \geq 0.71$. A non-zero global desirability value means that a set of conditions able to fulfill the optimization criteria for all the responses could be found, whereas its high value denotes a high degree of accordance between the single desirability values.

Table 4.5. Regression models calculated for each PAH.

Analyte	Equation ^a
N	$y = -0.89 (\pm 0.03) + 0.32 (\pm 0.02)X_1 - 0.35 (\pm 0.02)X_2 + 0.23 (\pm 0.02)X_3 - 0.34 (\pm 0.05)X_3^2$
Acy	$y = -0.63 (\pm 0.03) + 0.16 (\pm 0.02)X_1 - 0.52 (\pm 0.02)X_2 + 0.61 (\pm 0.02)X_3 + 0.30 (\pm 0.03)X_1X_3 + 0.07 (\pm 0.03)X_2X_3 - 0.34 (\pm 0.05)X_3^2$
Ace	$y = -0.46 (\pm 0.03) + 0.19 (\pm 0.03)X_1 - 0.47 (\pm 0.03)X_2 + 0.56 (\pm 0.03)X_3 + 0.33 (\pm 0.03)X_1X_3 - 0.65 (\pm 0.05)X_3^2$
F	$y = -0.41 (\pm 0.05) + 0.57 (\pm 0.05)X_1 - 0.35 (\pm 0.05)X_2$
P	$y = -0.15 (\pm 0.03) + 0.54 (\pm 0.03)X_1 - 0.35 (\pm 0.03)X_2 - 0.14 (\pm 0.05)X_3^2$
Ant	$y = -0.36 (\pm 0.02) + 0.62 (\pm 0.02)X_1 - 0.32 (\pm 0.02)X_2 - 0.23 (\pm 0.02)X_3 + 0.06 (\pm 0.02)X_1X_2 + 0.12 (\pm 0.02)X_2X_3 - 0.19 (\pm 0.03)X_1^2 - 0.16 (\pm 0.03)X_3^2$
Fl	$y = -0.10 (\pm 0.04) + 0.74 (\pm 0.03)X_1 - 0.39 (\pm 0.03)X_2 - 0.17 (\pm 0.02)X_3 - 0.22 (\pm 0.06)X_1^2$
Pyr	$y = -0.10 (\pm 0.03) + 0.76 (\pm 0.02)X_1 - 0.41 (\pm 0.02)X_2 - 0.16 (\pm 0.02)X_3 - 0.23 (\pm 0.05)X_1^2 - 0.13 (\pm 0.05)X_3^2$
BaAnt	$y = 0.72 (\pm 0.03)X_1 - 0.34 (\pm 0.03)X_2 - 0.22 (\pm 0.02)X_3 + 0.12 (\pm 0.04)X_2X_3 - 0.25 (\pm 0.06)X_1^2$
Chr	$y = 0.74 (\pm 0.04)X_1 - 0.34 (\pm 0.04)X_2 - 0.21 (\pm 0.04)X_3 + 0.12 (\pm 0.04)X_2X_3 - 0.32 (\pm 0.07)X_1^2$
BbFl	$y = -0.08 (\pm 0.02) + 0.69 (\pm 0.02)X_1 - 0.20 (\pm 0.02)X_2 - 0.41 (\pm 0.02)X_3 + 0.20 (\pm 0.02)X_2X_3 - 0.28 (\pm 0.03)X_1^2 - 0.14 (\pm 0.03)X_3^2$
BkFl	$y = 0.76 (\pm 0.02)X_1 - 0.27 (\pm 0.02)X_2 - 0.37 (\pm 0.02)X_3 + 0.16 (\pm 0.03)X_2X_3 - 0.30 (\pm 0.04)X_1^2 - 0.15 (\pm 0.04)X_3^2$
BaPyr	$y = 0.73 (\pm 0.03)X_1 - 0.27 (\pm 0.03)X_2 - 0.36 (\pm 0.03)X_3 + 0.14 (\pm 0.03)X_2X_3 - 0.26 (\pm 0.06)X_1^2$
IPyr	$y = -0.42 (\pm 0.03) + 0.76 (\pm 0.03)X_1 - 0.14 (\pm 0.03)X_2 - 0.64 (\pm 0.03)X_3 + 0.27 (\pm 0.03)X_2X_3 - 0.27 (\pm 0.05)X_1^2$
DBahAnt	$y = -0.78 (\pm 0.04) + 0.75 (\pm 0.04)X_1 - 0.74 (\pm 0.03)X_3 + 0.31 (\pm 0.04)X_2X_3 - 0.28 (\pm 0.07)X_1^2$
BghiPer	$y = -0.41 (\pm 0.03) + 0.66 (\pm 0.02)X_1 - 0.65 (\pm 0.02)X_3 + 0.35 (\pm 0.03)X_2X_3 - 0.20 (\pm 0.05)X_1^2$

^a only the significant terms at the 95% confidence level are reported, the coefficients (X_1 : sample volume; X_2 : solvent volume; X_3 : sorbent amount) are reported as coefficient (\pm standard error) rounded at one significant digit.

The identified experimental conditions were valid, as the lack-of-fit was not significant, and practicable, allowing for the extraction of all the analytes by limiting the amount of sample to be treated as well as solvent and sorbent to be consumed.

4.3.3. Method validation

The MD- μ SPE-GC-MS method was validated operating under optimized conditions. The equation of the calibration functions, LODs and LOQs are reported in Table 4.6.

Table 4.6. LODs, LOQs and calibration curve equations of the MD- μ SPE-GC-MS analytical method. LODs and LOQs values are reported in ng/L.

Analyte	LOD	LOQ	Calibration curve equation ^a	
			$b_0 \pm SE$ ^b	$b_1 \pm SE$ ^b
N	27	73	2.1 ± 0.2	2.2 ± 0.1
Acy	15	41	- ^c	1.19 ± 0.02
Ace	17	46	-	1.29 ± 0.03
F	14	38	0.16 ± 0.05	1.05 ± 0.03
P	9.8	27	-	1.47 ± 0.02
Ant	12	32	-	1.41 ± 0.01
Fl	8.5	23	-	1.212 ± 0.008
Pyr	6.9	19	-	1.34 ± 0.01
BaAnt	6.7	18	-	1.34 ± 0.01
Chr	8.3	23	-	1.25 ± 0.01
BbFl	21	57	-	1.25 ± 0.02
BkFl	13	36	-	1.43 ± 0.01
BaPyr	15	40	-	1.52 ± 0.01
IPyr	14	36	-	1.28 ± 0.02
DBahAnt	20	54	0.16 ± 0.04	1.29 ± 0.02
BghiPer	20	56	0.15 ± 0.05	1.59 ± 0.02

^a the equation of the calibration curve is $y = b_0 + b_1x$. ^b the intercept and slope are reported as coefficient \pm standard error rounded at one significant digit. ^c - not significant ($p > 0.05$).

The method exhibited LODs in the low ng/L range, suitable for the detection of PAHs at trace levels, whereas LOQs were in the 18–73 ng/L range. All the calibration functions proved to be linear ($p > 0.01$) in the LOQ–600 ng/L range and statistically significant ($p < 0.05$). In addition, the approximation error was not significantly greater than the expected pure experimental variation ($p > 0.01$).

Satisfactory results were obtained from precision and trueness studies, with the results summarized in Table 4.7 and in Figure 4.9, respectively. There were not statistically significant differences in the results obtained under intermediate precision conditions.

Table 4.7. Results of the precision study. $RSD\%$ referring to repeatability is marked with an r in apex and the one referring to intermediate precision is marked with an l in apex. Results are rounded at the nearest integer.

Analyte	L1 ^a		L2 ^b		L3 ^c	
	$RSD^r\%$	$RSD^l\%$	$RSD^r\%$	$RSD^l\%$	$RSD^r\%$	$RSD^l\%$
N	n.d. ^d	n.d.	18	26	17	24
Acy	10	14	7	9	6	7
Ace	11	15	8	16	11	11
F	11	12	7	10	13	23
P	9	9	4	6	11	11
Ant	7	10	8	9	8	8
Fl	6	7	5	5	3	5
Pyr	5	6	4	5	4	6
BaAnt	9	14	11	12	7	7
Chr	4	5	5	5	3	3
BbFl	10	18	15	16	12	14
BkFl	4	4	6	6	2	3
BaPyr	9	9	7	11	10	13
IPyr	6	6	6	7	5	8
DBahAnt	16	19	9	9	9	11
BghiPer	13	15	10	11	5	8

^a 60 ng/L. ^b 250 ng/L. ^c 550 ng/L. ^d not determined.

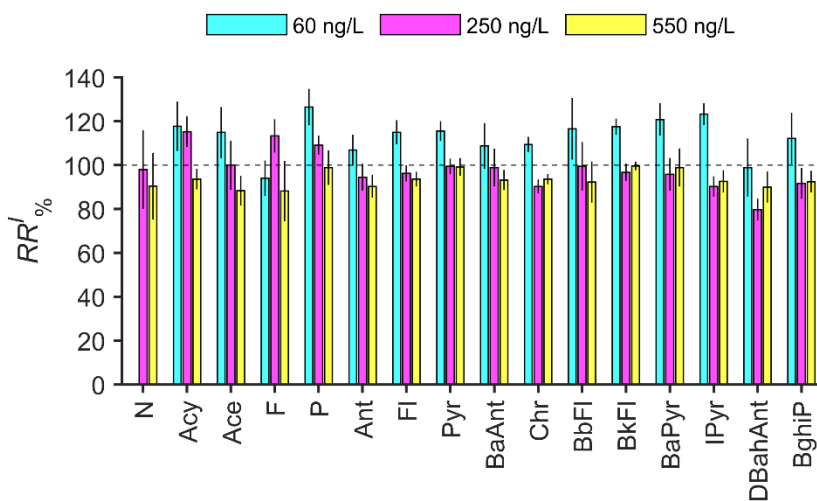


Figure 4.9. Results of the trueness study. Results are reported as mean \pm semi-amplitude of the 95% confidence interval ($n = 10$). The horizontal dashed line corresponds to a $RR\% = 100\%$.

For what concerns EFs (Figure 4.10), the lowest values were reached for low-molecular weight PAHs, mid-range values for high-molecular weight PAHs, and the highest values were obtained for the mid-molecular weight PAHs.

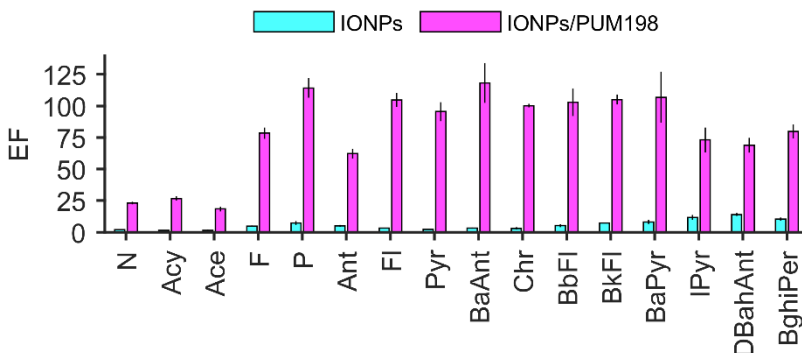


Figure 4.10. Enrichment factors for the bare IONPs and the magnetic composite. Results are reported as mean \pm one standard deviation ($n = 3$).

These results could be related to the different steric hindrance of the analytes:

- Smaller PAHs could have been less retained, being able to diffuse inside and outside the framework more easily.
- High-molecular weight PAHs could have interacted preferably with the surface of the framework rather than diffusing inside the pores. In fact, on average, the observed EFs for the high-molecular weight PAHs were 20–30 % lower than the ones of mid-molecular weight PAHs.
- The highest EFs observed for mid-molecular weight PAHs suggested that they could benefit both from the adsorption on the surface of the framework and from the diffusion inside the pores.

The trend observed with EFs was consistent with the one of LODs and LOQs. The comparison between the EFs obtained for bare IONPs and the magnetic composite furtherly confirmed that the fabrication of the sorbent was successful. In fact, the modification of bare IONPs improved their EFs of, on average, 11 standard deviations [45].

The proposed analytical methodology provided LODs comparable or lower than the ones reported in previous studies [38–40,46–52] dealing with the development of sorbents for MD- μ SPE. In addition, the preconcentration achieved with the proposed method was noteworthy, as only 5 mL of sample are required: this amount is roughly 3–20 times lower than the one requested by other methods.

4.3.4. Analysis of real samples

The proposed analytical method was used for the qualitative-quantitative determination of the 16 US-EPA PAHs in underground water samples, collected in a contaminated site in the province of Parma. Out of 19 samples, only six were characterized by the presence of PAHs at various concentration levels. The results are summarized in Table 4.8. High-molecular weight PAHs were the less abundant compounds, with DBahAnt and BghiPer being undetected in all the processed samples. Low-molecular weight analytes were the most abundant, with sample A and sample F being the only exception. Sample B and E were massively contaminated, showing the highest Σ_{16} PAHs concentration, i.e., $830 (\pm 30)$ and $197 (\pm 5) \mu\text{g/L}$, respectively, with Ace representing the major fraction of the quantitated analytes.

4.4. Conclusions

This study was devoted to the development of a magnetic composite for the MD- μSPE of the 16 PAHs included in the US-EPA list from environmental water samples. The synthetic procedure was based on the post-synthetic modification of PUM198, by decorating it with APTES-encapsulated IONPs. Thanks to the presence of aromatic ligands and the high porosity of PUM198, the magnetic composite proved to be suitable for the enrichment of PAHs. The method was validated according to the EURACHEM guidelines, providing LODs and LOQs suitable for the analysis of PAHs at trace levels. Other figures of merit include good precision and bias also near the LOQs. Noticeable enrichment capabilities were also reached, especially considering that the method requires the treatment of only 5 mL and the consumption of limited amounts of organic solvents, i.e., 50 μL . This well counterbalances the resources spent in the synthesis of the material, making a step forward towards the requirements of Green Analytical Chemistry. On a closing note, the potential of the proposed method could be exploited for the treatment of samples whose availability is limited, and a good sample throughput can be forecasted, being the time *per* sample preparation being close to the time *per* analysis.

Table 4.8. Results of the quali- quantitative analysis. All the results are in $\mu\text{g/L}$ expressed as mean \pm one standard deviation ($n = 2$). Results are rounded at one significant digit.

Analyte	Sample A	Sample B	Sample C	Sample D	Sample E	Sample F
N	n.q. ^a	4.5 \pm 0.2	0.45 \pm 0.02	0.50 \pm 0.02	9.6 \pm 0.4	n.d. ^b
Acy	0.09 \pm 0.02	60 \pm 2	0.23 \pm 0.01	0.15 \pm 0.02	5.5 \pm 0.1	n.d.
Ace	0.08 \pm 0.02	670 \pm 30	0.60 \pm 0.02	n.d.	126 \pm 5	n.d.
F	n.q.	93 \pm 4	0.14 \pm 0.02	0.06 \pm 0.02	11.5 \pm 0.5	0.50 \pm 0.04
P	n.q.	n.d.	0.19 \pm 0.01	0.16 \pm 0.01	13.6 \pm 0.3	0.060 \pm 0.001
Ant	n.q.	n.d.	0.05 \pm 0.01	0.04 \pm 0.01	3.37 \pm 0.06	n.d.
Fl	n.q.	0.343 \pm 0.007	0.024 \pm 0.008	0.508 \pm 0.008	15.5 \pm 0.2	0.50 \pm 0.04
Pyr	n.q.	0.76 \pm 0.01	0.029 \pm 0.009	0.224 \pm 0.008	11.2 \pm 0.1	n.d.
BaAnt	0.14 \pm 0.01	0.42 \pm 0.01	0.36 \pm 0.01	0.18 \pm 0.01	0.19 \pm 0.01	0.20 \pm 0.02
Chr	n.q.	0.16 \pm 0.01	0.03 \pm 0.01	n.q.	0.27 \pm 0.01	0.070 \pm 0.008
BbFl	0.13 \pm 0.02	n.d.	0.06 \pm 0.02	0.33 \pm 0.02	0.08 \pm 0.02	n.q.
BkFl	n.d.	n.d.	0.04 \pm 0.01	n.q.	n.d.	n.q.
BaPyr	0.05 \pm 0.01	0.29 \pm 0.01	0.09 \pm 0.01	0.12 \pm 0.01	n.q.	n.q.
lPyr	n.d.	n.d.	n.d.	n.d.	n.d.	0.50 \pm 0.03
DBahAnt	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
BghiPer	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
$\Sigma_{16}\text{PAH}^c$	0.48 \pm 0.04	830 \pm 30	2.30 \pm 0.05	2.28 \pm 0.04	197 \pm 5	1.83 \pm 0.07

^a not quantitated. ^b not detected. ^c the sum accounts only for the quantitated analytes.

Note of the author

The results presented in this **Chapter** are published on *Journal of Chromatography A* 1671 (2022) 463010 [31].

The author would like to express gratitude to Massimo Savazzi and Davide Agostini for participating to this study as a part of the internship required to obtain the Master of Science degree.

References

- [1] O.O. Alegbeleye, B.O. Opeolu, V.A. Jackson, Polycyclic Aromatic Hydrocarbons: A Critical Review of Environmental Occurrence and Bioremediation, *Environ Manage.* 60 (2017) 758–783. <https://doi.org/10.1007/s00267-017-0896-2>.
- [2] H.I. Abdel-Shafy, M.S.M. Mansour, A review on polycyclic aromatic hydrocarbons: Source, environmental impact, effect on human health and remediation, *Egyptian Journal of Petroleum.* 25 (2016) 107–123. <https://doi.org/10.1016/j.ejpe.2015.03.011>.
- [3] C. Achten, J.T. Andersson, Overview of Polycyclic Aromatic Compounds (PAC), *Polycycl Aromat Compd.* 35 (2015) 177–186. <https://doi.org/10.1080/10406638.2014.994071>.
- [4] Y.L. Tan, A. Kong, M.A. Monetti, Biogenic Polycyclic Aromatic Hydrocarbons in an Alaskan Arctic Lake Sediment, *Polycycl Aromat Compd.* 9 (2006) 185–192. <https://doi.org/10.1080/10406639608031217>.
- [5] A. Arcoleo, F. Bianchi, M. Careri, A sensitive microextraction by packed sorbent-gas chromatography-mass spectrometry method for the assessment of polycyclic aromatic hydrocarbons contamination in Antarctic surface snow, *Chemosphere.* 282 (2021) 131082. <https://doi.org/10.1016/j.chemosphere.2021.131082>.
- [6] C.A. Menzie, B.B. Potocki, S. Joseph, Exposure to Carcinogenic PAHs in The Environment, *Environ Sci Technol.* 26 (1992) 1278–1284. <https://doi.org/10.1021/ES00031A002>.
- [7] M. de la Guardia, S. Garrigues, Past, Present and Future of Green Analytical Chemistry, in: M. de la Guardia, S. Garrigues (Eds.), *Challenges in Green Analytical Chemistry*, 2nd ed., The Royal Society of Chemistry, 2020: pp. 1–18. <https://doi.org/10.1039/9781788016148-00001>.
- [8] M. Ghorbani, M. Aghamohammadhassan, M. Chamsaz, H. Akhlaghi, T. Pedramrad, Dispersive solid phase microextraction, *TrAC - Trends in Analytical Chemistry.* 118 (2019) 793–809. <https://doi.org/10.1016/j.trac.2019.07.012>.
- [9] M. Anastassiades, S.J. Lehotay, D. Štajnbaher, F.J. Schenck, Fast and easy multiresidue method employing acetonitrile extraction/partitioning and “dispersive solid-phase extraction” for the determination of pesticide residues in produce, *J AOAC Int.* 86 (2003) 412–431. <https://doi.org/10.1093/jaoac/86.2.412>.
- [10] K. Moyakao, Y. Santaladchaiyakit, S. Srijaranai, J. Vichapong, Preconcentration of Trace Neonicotinoid Insecticide Residues Using Vortex-Assisted Dispersive Micro Solid-Phase Extraction with Montmorillonite as an Efficient Sorbent, *Molecules.* 23 (2018) 883. <https://doi.org/10.3390/MOLECULES23040883>.
- [11] M. Šafaříková, I. Šafařík, Magnetic solid-phase extraction, *J Magn Magn Mater.* 194 (1999) 108–112. [https://doi.org/10.1016/S0304-8853\(98\)00566-6](https://doi.org/10.1016/S0304-8853(98)00566-6).
- [12] M.L. Castillo-García, M.P. Aguilar-Caballeros, A. Gómez-Hens, Nanomaterials as tools in chromatographic methods, *TrAC - Trends in Analytical Chemistry.* 82 (2016) 385–393. <https://doi.org/10.1016/j.trac.2016.06.019>.

- [13] S. Büyüktiryaki, R. Keçili, C.M. Hussain, Functionalized nanomaterials in dispersive solid phase extraction: Advances & prospects, *TrAC - Trends in Analytical Chemistry*. 127 (2020) 115893. <https://doi.org/10.1016/j.trac.2020.115893>.
- [14] S.L. James, Metal-organic frameworks, *Chem Soc Rev*. 32 (2003) 276–288. <https://doi.org/10.1039/b200393g>.
- [15] H.- Cai Joe Zhou, S. Kitagawa, Metal-Organic Frameworks (MOFs), *Chem. Soc. Rev*. 43 (2014) 5415. <https://doi.org/10.1039/c4cs90059f>.
- [16] M. Eddaoudi, D.B. Moler, H. Li, B. Chen, T.M. Reineke, M. O’Keeffe, O.M. Yaghi, Modular Chemistry: Secondary Building Units as a Basis for the Design of Highly Porous and Robust Metal–Organic Carboxylate Frameworks, *Acc Chem Res*. 34 (2001) 319–330. <https://doi.org/10.1021/AR000034B>.
- [17] O.M. Yaghi, M. O’Keeffe, N.W. Ockwig, H.K. Chae, M. Eddaoudi, J. Kim, Reticular synthesis and the design of new materials, *Nature* 2003 423:6941. 423 (2003) 705–714. <https://doi.org/10.1038/nature01650>.
- [18] Inamuddin, R. Boddula, M.I. Ahamed, A.M. Asiri, Applications of Metal–Organic Frameworks and Their Derived Materials, Wiley, 2020. <https://doi.org/10.1002/9781119651079>.
- [19] F. Maya, C. Palomino Cabello, R.M. Frizzarin, J.M. Estela, G. Turnes Palomino, V. Cerdà, Magnetic solid-phase extraction using metal-organic frameworks (MOFs) and their derived carbons, *TrAC - Trends in Analytical Chemistry*. 90 (2017) 142–152. <https://doi.org/10.1016/j.trac.2017.03.004>.
- [20] Z.Y. Gu, J.Q. Jiang, X.P. Yan, Fabrication of isoreticular metal-organic framework coated capillary columns for high-resolution gas chromatographic separation of persistent organic pollutants, *Anal Chem*. 83 (2011) 5093–5100. https://doi.org/10.1021/AC200646W/SUPPL_FILE/AC200646W_SI_001.PDF.
- [21] P. Kumar, A. Deep, K.H. Kim, Metal organic frameworks for sensing applications, *TrAC - Trends in Analytical Chemistry*. 73 (2015) 39–53. <https://doi.org/10.1016/j.TRAC.2015.04.009>.
- [22] I. Taima-Mancera, P. Rocío-Bautista, J. Pasán, J.H. Ayala, C. Ruiz-Pérez, A.M. Afonso, A.B. Lago, V. Pino, Influence of Ligand Functionalization of UiO-66-Based Metal-Organic Frameworks When Used as Sorbents in Dispersive Solid-Phase Analytical Microextraction for Different Aqueous Organic Pollutants, *Molecules*. 23 (2018) 2869. <https://doi.org/10.3390/MOLECULES23112869>.
- [23] D. Balestri, P.P. Mazzeo, R. Perrone, F. Fornari, F. Bianchi, M. Careri, A. Bacchi, P. Pelagatti, Deciphering the Supramolecular Organization of Multiple Guests Inside a Microporous MOF to Understand their Release Profile, *Angewandte Chemie International Edition*. 60 (2021) 10194–10202. <https://doi.org/10.1002/ANIE.202017105>.
- [24] S. Pullen, G.H. Clever, Mixed-Ligand Metal–Organic Frameworks and Heteroleptic Coordination Cages as Multifunctional Scaffolds—A Comparison, *Acc Chem Res*. 51 (2018) 3052–3064. <https://doi.org/10.1021/ACS.ACCOUNTS.8B00415>.
- [25] P. González-Hernández, A.B. Lago, J. Pasán, C. Ruiz-Pérez, J.H. Ayala, A.M. Afonso, V. Pino, Application of a Pillared-Layer Zn-Triazololate Metal-Organic Framework in the Dispersive Miniaturized Solid-Phase Extraction of Personal Care Products from Wastewater Samples, 24 (2019) 690. <https://doi.org/10.3390/MOLECULES24040690>.
- [26] F. Bianchi, A. Pankajakshan, F. Fornari, S. Mandal, P. Pelagatti, A. Bacchi, P.P. Mazzeo, M. Careri, A zinc mixed-ligand microporous metal-organic framework as solid-phase microextraction coating for priority polycyclic aromatic hydrocarbons from water samples, *Microchemical Journal*. 154 (2020) 104646. <https://doi.org/10.1016/j.microc.2020.104646>.
- [27] S. Yadav, R. Dixit, S. Sharma, S. Dutta, K. Solanki, R.K. Sharma, Magnetic metal–organic framework composites: structurally advanced catalytic materials for organic transformations, *Mater Adv*. 2 (2021) 2153–2187. <https://doi.org/10.1039/D0MA00982B>.
- [28] N.L. Torad, M. Hu, S. Ishihara, H. Sukegawa, A.A. Belik, M. Imura, K. Ariga, Y. Sakka, Y. Yamauchi, Direct Synthesis of MOF-Derived Nanoporous Carbon with Magnetic Co Nanoparticles toward

- Efficient Water Treatment, *Small*. 10 (2014) 2096–2107. <https://doi.org/10.1002/SMLL.201302910>.
- [29] D. Balestri, D. Costa, A. Bacchi, L. Carlucci, P. Pelagatti, Linker dependent dimensionality in Zn(II)-coordination polymers containing a flexible bis-pyridyl-bis-amide ligand, *Polyhedron*. 153 (2018) 278–285. <https://doi.org/10.1016/j.poly.2018.07.025>.
- [30] D. Balestri, A. Bacchi, P. Scilabra, P. Pelagatti, Extension of the Pd-catalyzed C–N bond forming reaction to the synthesis of large polydentate ligands containing N–H functions, *Inorganica Chim Acta*. 470 (2018) 416–422. <https://doi.org/10.1016/j.ica.2017.06.029>.
- [31] F. Fornari, F. Bianchi, N. Riboni, F. Casoli, A. Bacchi, P.P. Mazzeo, P. Pelagatti, M. Careri, Metal-organic framework-based magnetic dispersive micro-solid-phase extraction for the gas chromatography–mass spectrometry determination of polycyclic aromatic compounds in water samples, *J Chromatogr A*. 1671 (2022) 463010. <https://doi.org/10.1016/j.CHROMA.2022.463010>.
- [32] P. Berger, N.B. Adelman, K.J. Beckman, D.J. Campbell, A.B. Ellis, G.C. Lisensky, Preparation and Properties of an Aqueous Ferrofluid, *J Chem Educ*. 76 (1999) 943–948. <https://doi.org/10.1021/ed076p943>.
- [33] M. Esmaeilzadeh, A composite prepared from a metal-organic framework of type MIL-101(Fe) and morin-modified magnetite nanoparticles for extraction and speciation of vanadium(IV) and vanadium(V), *Microchimica Acta*. 186 (2019) 1–7. <https://doi.org/10.1007/s00604-018-3093-y>.
- [34] Y. Hu, Z. Huang, J. Liao, G. Li, Chemical bonding approach for fabrication of hybrid magnetic metal-organic framework-5: High efficient adsorbents for magnetic enrichment of trace analytes, *Anal Chem*. 85 (2013) 6885–6893. <https://doi.org/10.1021/ac4011364>.
- [35] J. Cohen, Eta-squared and partial eta-squared in fixed factor ANOVA designs, *Educ Psychol Meas*. 33 (1973) 107–112. <https://doi.org/10.1177/001316447303300111>.
- [36] G. Derringer, R. Suich, Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology*. 12 (1980) 214–219. <https://doi.org/10.1080/00224065.1980.11980968>.
- [37] B. Magnusson, U. Örnemark, eds., *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topic*, 2nd ed., 2014. [https://doi.org/10.1016/S0014-2999\(99\)00500-2](https://doi.org/10.1016/S0014-2999(99)00500-2).
- [38] Y. Tong, Q. Zhou, Y. Sun, X. Sheng, B. Zhou, J. Zhao, J. Guo, Magnetic polyamidoamine dendrimer grafted with 4-mercaptobenzoic acid as an adsorbent for preconcentration and sensitive determination of polycyclic aromatic hydrocarbons from environmental water samples, *Talanta*. 224 (2021) 121884. <https://doi.org/10.1016/j.TALANTA.2020.121884>.
- [39] A. Azizi, F. Shahhoseini, C.S. Bottaro, Magnetic molecularly imprinted polymers prepared by reversible addition fragmentation chain transfer polymerization for dispersive solid phase extraction of polycyclic aromatic hydrocarbons in water, *J Chromatogr A*. 1610 (2020) 460534. <https://doi.org/10.1016/j.chroma.2019.460534>.
- [40] A. Asfaram, E.A. Dil, P. Arabkhani, F. Sadeghfard, M. Ghaedi, Magnetic Cu: CuO-GO nanocomposite for efficient dispersive micro-solid phase extraction of polycyclic aromatic hydrocarbons from vegetable, fruit, and environmental water samples by liquid chromatographic determination, *Talanta*. 218 (2020) 121131. <https://doi.org/10.1016/j.talanta.2020.121131>.
- [41] A. Gutiérrez-Serpa, A.I. Jiménez-Abizanda, F. Jiménez-Moreno, J. Pasán, V. Pino, Core-shell microparticles formed by the metal-organic framework CIM-80(Al) (Silica@CIM-80(Al)) as sorbent material in miniaturized dispersive solid-phase extraction, *Talanta*. 211 (2020) 120723. <https://doi.org/10.1016/j.TALANTA.2020.120723>.
- [42] J. Gao, C. Huang, Y. Lin, P. Tong, L. Zhang, In situ solvothermal synthesis of metal-organic framework coated fiber for highly sensitive solid-phase microextraction of polycyclic aromatic hydrocarbons, *J Chromatogr A*. 1436 (2016) 1–8. <https://doi.org/10.1016/j.chroma.2016.01.051>.
- [43] X.F. Chen, H. Zang, X. Wang, J.G. Cheng, R.S. Zhao, C.G. Cheng, X.Q. Lu, Metal-organic framework MIL-53(Al) as a solid-phase microextraction adsorbent for the determination of 16 polycyclic

- aromatic hydrocarbons in water samples by gas chromatography-tandem mass spectrometry, *Analyst*. 137 (2012) 5411–5419. <https://doi.org/10.1039/c2an35806a>.
- [44] G.E.P. Box, D.R. Cox, *An Analysis of Transformations*, *Journal of the Royal Statistical Society. Series B*. 26 (1964) 211–252. <https://www.jstor.org/stable/2984418> (accessed September 3, 2021).
- [45] S. Vandekar, R. Tao, J. Blume, A Robust Effect Size Index, *Psychometrika*. 85 (2020) 232–246. <https://doi.org/10.1007/S11336-020-09698-2/FIGURES/4>.
- [46] S.H. Huo, X.-P.P. Yan, Facile magnetization of metal-organic framework MIL-101 for magnetic solid-phase extraction of polycyclic aromatic hydrocarbons in environmental water samples, *Analyst*. 137 (2012) 3445–3451. <https://doi.org/10.1039/c2an35429b>.
- [47] Q. Zhou, M. Lei, Y. Wu, Y. Yuan, Magnetic solid phase extraction of typical polycyclic aromatic hydrocarbons from environmental water samples with metal organic framework MIL-101 (Cr) modified zero valent iron nano-particles, *J Chromatogr A*. 1487 (2017) 22–29. <https://doi.org/10.1016/j.chroma.2017.01.046>.
- [48] P. Rocío-Bautista, V. Pino, J.H. Ayala, J. Pasán, C. Ruiz-Pérez, A.M. Afonso, A magnetic-based dispersive micro-solid-phase extraction method using the metal-organic framework HKUST- and ultra-high-performance liquid chromatography with fluorescence detection for determining polycyclic aromatic hydrocarbons in waters and fruit tea i, *J Chromatogr A*. 1436 (2016) 42–50. <https://doi.org/10.1016/j.chroma.2016.01.067>.
- [49] F. Du, Q. Qin, J. Deng, G. Ruan, X. Yang, L. Li, J. Li, Magnetic metal-organic framework MIL-100(Fe) microspheres for the magnetic solid-phase extraction of trace polycyclic aromatic hydrocarbons from water samples, *J Sep Sci*. 39 (2016) 2356–2364. <https://doi.org/10.1002/JSSC.201600100>.
- [50] A. Mehdinia, N. Khodaei, A. Jabbari, Fabrication of graphene/Fe₃O₄@polythiophene nanocomposite and its application in the magnetic solid-phase extraction of polycyclic aromatic hydrocarbons from environmental water samples, *Anal Chim Acta*. 868 (2015) 1–9. <https://doi.org/10.1016/j.aca.2014.12.022>.
- [51] M. Majd, S. Nojavan, Determination of polycyclic aromatic hydrocarbons in soil, tree leaves, and water samples by magnetic dispersive solid-phase extraction based on β -cyclodextrin functionalized graphene oxide followed by GC-FID, *Microchemical Journal*. 171 (2021) 106852. <https://doi.org/10.1016/j.MICROC.2021.106852>.
- [52] N. Manousi, E.A. Deliyanni, E. Rosenberg, G.A. Zachariadis, Ultrasound-assisted magnetic solid-phase extraction of polycyclic aromatic hydrocarbons and nitrated polycyclic aromatic hydrocarbons from water samples with a magnetic polyaniline modified graphene oxide nanocomposite, *J Chromatogr A*. 1645 (2021) 462104. <https://doi.org/10.1016/j.CHROMA.2021.462104>.

Chapter 5 | Exploring the potential of aquatic plants and carbon-based materials in detoxifying polluted sediments

5.1. Introduction

5.1.1. Eutrophication of water bodies and bioremediation

The recent increase in the worldwide population^{xli} produced an acceleration of economic development, intensifying manufacturing activities to satisfy the worldwide demand^{xlii}. In this context, the food system had serious difficulties in keeping pace: in fact, although more sustainable approaches to agriculture and livestock have been proposed [1,2], the effects of the inconsiderate use of mineral fertilizers, pesticides and intensive farming are still a matter to deal with [3–5]. As a matter of fact, such practices produced adverse effects, depleting precious resources like soil, reducing biodiversity, and producing effects on water ecosystems [3–9].

Eutrophication is a phenomenon that occurs in water ecosystems in response to an increased level of nutrients in water bodies deriving from anthropic activities [6–9]. Nutrients are substances with fertilizing effects and include essential components for plant metabolism among which P, N, and K are the most important ones [10]. The higher availability of such substances, deriving, for instance, from agricultural activities, husbandry, and the improper disposal of grey- and blackwaters from households and industries, produces a rapid and uncontrollable algal proliferation, producing undesirable consequences on the aquatic ecosystem sustenance and on the quality of the water itself [7,8,11,12] (Figure 5.1).

The increased availability of nutrients promotes the rapid blooming of aquatic plants, including both microscopic and macroscopic algae and epiphyte plants. This represents only the initial response of the ecosystem to the perturbation [6–9]. The large algal biomass causes the loss of transparency of the water body subjected to eutrophication, hindering the photosynthesis to the aquatic plants that are part of the benthos. This depletes the dissolved oxygen extending the anoxic zone, causing the loss of submerged aquatic vegetation and fish kill with an overall reduction in terms of biodiversity [7–9,11,12].

^{xli} <https://unric.org/en/8-billion-people-10-facts-on-the-worlds-population/> (accessed 07/01/2023)

^{xlii} <https://www.un.org/en/dayof8billion> (accessed 07/01/2023)

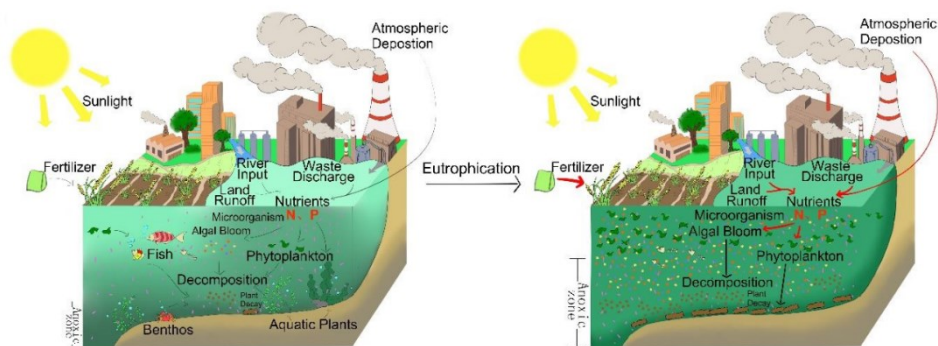


Figure 5.1. Schematic representation of the eutrophication process. Adapted after Yan Zhang and coworkers [7].

5.1.1.1. Bioremediation

In December 2019, the European Commission announced a series of policies that aim at making Europe the first climate-neutral continent by 2050, reducing by 55% the emission of greenhouse gasses in the atmosphere by 2030 (with respect to 1990 levels), to create a toxic-free environment. These are just a few of the objectives that are reported in the so-called European Green Deal^{xliii}. In order to concretize the climatic objectives, it is fundamental to re-design the economic model from a linear to a more circular one^{xliiv}. Additionally, along with the introduction of technologies with a reduced impact on the climate and the environment, it is fundamental to implement ones able to remediate polluted ecosystems to effectively reach a depolluted environment.

Within this frame of reference, bioremediation provides a sustainable and cost-effective approach for the removal of pollutants from contaminated ecosystems [13–18]. The term bioremediation comprises a series of biotechnological processes that involves organisms from different taxonomic Kingdoms, including *Bacteria* [19], *Plantae* [14,19], *Fungi* [20] (see **Chapter 7**) and even *Animalia* [21], for the detoxification of contaminated ecosystems. The detoxification process is due to the metabolic activity of the involved organisms, that promotes the degradation, bioadsorption, and immobilization of contaminants [13].

There is evidence reported in the literature about the efficacy of bioremediation processes applied to different environmental matrices towards different class of pollutants, including metals [19], pesticides [17], aliphatic and aromatic

^{xliii} <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1576150542719&uri=COM%3A2019%3A640%3AFIN> (accessed 31/12/2022)

^{xliiv} <https://www.europarl.europa.eu/news/en/headlines/economy/20151201STO05603/circular-economy-definition-importance-and-benefits> (accessed 30/12/2022)

hydrocarbons [22], and also emerging contaminants (see **Chapter 3**) like organic dyes [20] and antibiotics [23].

5.1.1.2. Carbon-based materials in remediation processes

It is known that the remediation process can be improved using active materials such as biochar and hydrochar^{xlv} [18,24].

Hydrochar is a material similar to lignite. It is poor in ashes and rich in carbon [25–27]. The composition and its surface chemistry can change depending on the starting material. Hydrochar is a notoriously poorly porous material [28], with surface areas ranging from few tens to few hundreds m²/g depending on the starting material [29] and the post-production thermal treatments [25]. Nevertheless, its surface chemistry still makes it appealing as a sorbent material. Indeed, the surface of hydrochar is rich in oxygenated functional groups [27], including –OH (both hydroxylic and phenolic), –COOH, and C=O, making it suitable in cation exchange processes for removing metals from liquid matrices [24,25], and also promoting the immobilization of microbial communities and biofilm formation [18].

Hydrochar is produced by the so-called hydrothermal carbonization (HTC). HTC is a thermochemical process that utilizes biomass as feedstock and converts it into different fractions in the presence of water. The work conditions are milder with respect to the ones of other well-known technologies, such as pyrolysis and gasification [27,30]. In addition, the HTC process takes place in presence of water, thus not making necessary biomass drying, saving a large amount of energy as the vaporization enthalpy of water is fairly high [25–27,30]. HTC is usually carried out by operating into a 180–280 °C temperature range, under autogenous pressure and non-oxidant atmosphere. The carbonization process lasts from few minutes to several hours and produces a solid fraction (i.e., the hydrochar), a liquid fraction, and a smaller proportion of gaseous fraction [25].

The typical feedstock is biomass that can derive from agriculture [26,27], but it has been reported that sludges produced by sewage treatment facilities can constitute an appealing feedstock to produce hydrochar [31] in the context of circular economy^{xlvi, xlvii}.

^{xlv} For the sake of clarity, different technologies are available to produce biochar, like pyrolysis. The term hydrochar comprises the biochars that are produced *via* hydrothermal carbonization.

^{xlvi} <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1576150542719&uri=COM%3A2019%3A640%3AFIN> (accessed 02/01/2023)

^{xlvii} <https://www.europarl.europa.eu/news/en/headlines/economy/20151201STO05603/circular-economy-definition-importance-and-benefits> (accessed 02/01/2023)

5.1.1.3. *Ecophysiology of submerged macrophytes: an outline*

The term macrophyte refers to macroscopic plants that evolved and adapted to live in aquatic environments. Among them, rooted aquatic macrophytes are of particular interest for their unique characteristic of transporting variable amounts of oxygen through the roots [32,33]. This process is possible thanks to vascular structure called aerenchyma that connects the leaves to the roots through the stem. This intercellular structure allows the diffusive transport of oxygen to the roots, permitting respiration also in strictly anoxic substrates, such as sediments.

The most interesting aspect of this peculiar characteristic is that part of the oxygen that is transported to the roots can be transferred in the rhizosphere (i.e., the volume of soil in close proximity to the roots), causing the oxygenation of the vegetated substrate [32,33]. This oxygen transfer goes under the name of radial oxygen loss (ROL) and represents the fraction of oxygen that is transferred to the rhizosphere with respect to the oxygen produced during the photosynthesis [34]. The ability of submerged macrophytes of oxygenating the sediments has important ecological implications in shallow water ecosystems, including the sustenance of aerobic microorganisms surrounding the rhizosphere [35], and the promotion of the nitrification process, i.e., the oxidation of NH_4^+ to NO_2^- and NO_3^- [36,37].

Vallisneria spiralis (Figure 5.2) is a macrophyte with ribbon-like leaves that colonizes the vast majority of shallow water bodies in the North of Italy, including rivers, irrigation canals, and Alpine lakes [37]. This plant has been extensively studied by Prof. Marco Bartoli and coworkers (University of Parma), who investigated the dynamics of oxygen in the rhizosphere and the consequent effects on nutrients, as well as the chemical modifications the plants can exert on the pore water [33,36–38].



Figure 5.2. Specimens of *V. spiralis*. Photograph taken by Arianna Soci.

The capability of *V. spiralis* in diminishing the concentration of nutrients, such as PO_4^{3-} and NH_4^+ , from the rhizosphere [33,36–38], as well as immobilizing and/or depleting metal ions [33,38], i.e., Fe^{2+} and Mn^{2+} , along with the exceptional tolerance towards a wide range of environmental conditions [38], make this organism a good candidate for being included in bioremediation processes.

5.1.2. Mantua: the city shaped by water

Mantua (45° 9' 23" N, 10° 47' 28" E; Mantua, Italy) is a city located in Northern Italy, more precisely in the South-East area of Lombardy. The complex hydrographic network of Mantua has origins in the late centuries of the Middle Ages, and it transformed into the city that is known today during the Renaissance [39].

In the late XII century, the Bergamasque engineer Alberto Pitentino directed a series of hydraulics projects that aimed at protecting the city by surrounding it with water. This has been accomplished by channeling the water of the Mincio River into four artificial lakes, i.e., Lower Lake, Middle Lake, Upper Lake, and Paiolo Lake, which surrounded the city completely [39,40]. Although the defense system designed by Pitentino required few adjustments over the years, it remained unchanged across many centuries (Figure 5.3).

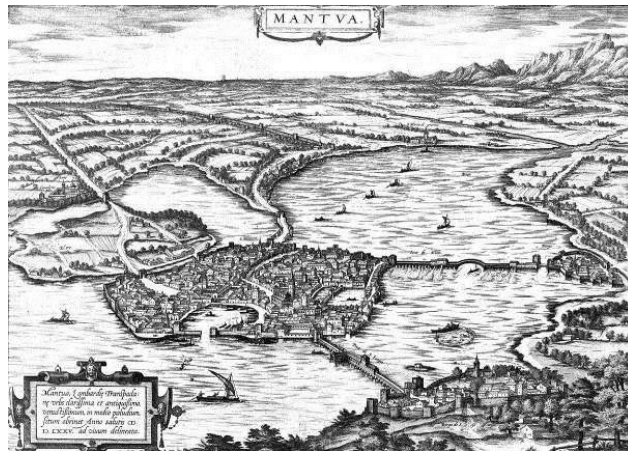


Figure 5.3. Map of the Duchy of Mantua in the XVI century. Reprinted from Damien Claeys and Claude Lambert [41] (originally by Franz Hogenberg and edited by Georg Braun; from *Civitates Orbis Terrarum*, vol. II, 50, 1575).

It was not until the early XVIII century, during the last years of dominance of the House of Gonzaga [42,43], that the Paiolo Lake was drained transforming Mantua into a peninsula, to promote further urban development [39,40].

What remains today are the three artificial lakes, i.e., Lower Lake, Middle Lake, and Upper Lake, which with the regional law LR 47/1984^{xlviii} were recognized as a natural conservation area. With the regional law DGR 7/193/2000^{xlix} the area was finally institutionalized as the Mincio Regional Park.

5.1.2.1. Rise of the industries¹

With the city staying on one side of the lakes, the opposite one became a site of intense industrial activity starting from mid '40s. The rise of the Italian Chemical Industry for Petroleum (ICIP) refinery in 1946 and the petrochemical facility under the governance of Edison S.p.A. in 1956 signed the beginning of the industrialization process of Mantuan surroundings.

During the '60s, the ICIP refinery encountered a rapid expansion of the plant, allowing to produce 2.6 million tons/year of fuels. After having been sold to the Italian Energy and Services S.p.A., the company was acquired in 2007 by the Hungarian company MOL Plc. Under the governance of the latter, in 2013 the facility stopped the production, and it has been converted into a logistic center, that started its activity in 2015.

In the meantime, the petrochemical plant after years of conversions and joint ventures, ended under the governance of different companies. The activities devoted to the production of energy were absorbed by Enipower Mantua in 2000. The petrochemical facility was acquired two years later by Europe Polymers, today known as Versalis S.p.A. The areas of the industrialized site not devoted to productive activities and landfills remained under the governance of Enichem since 1991, now under the name of Eni Rewind S.p.A. The plants are today involved in the transformation of feedstock like benzene, ethylbenzene, ethylene, cumene, acrylonitrile, and pentane into basic bulk chemicals and polymers, including phenol and styrene and polystyrene, respectively. Originally, the facility plants were dedicated to petroleum cracking, to the production of maleic anhydride and to the integrated production of NaOH and Cl₂ by electrolyzing molten NaCl. Such plants were closed during the years for ecological reasons: the cracking plant was closed in 1978, whereas all the other ones in 1991.

During the years of industrial production, the soil beneath the ex-refinery as well as the groundwater suffered of a massive contamination of hydrocarbons due to pipeline leakages. Additionally, before a sewage treatment plant was properly

^{xlviii} <https://normelombardia.consiglio.regione.lombardia.it/normelombardia/Accessibile/main.aspx?view=showdoc&elnode=0&idoc=lr001984090800047> (accessed 27/12/2022)

^{xlix} <http://www.parcodelmincio.it/parco/documents/norme.pdf> (accessed 27/12/2022)

¹ <https://www.arpalombardia.it/Pages/Bonifica/Mantova/Storia.aspx?firstlevel=Mantova> (accessed 28/12/2022)

installed, the effluents deriving from the petrochemical plant caused the contamination of the sediments of the Mincio River. Moreover, despite the chloroalkali electrochemical plant has been dismissed and the Hg-based electrolytic cells removed, during the years mercury percolated first in the soil, and then into the aquifer.

Along with the on-site industrial activities, starting from the '90s, agricultural activities and husbandry in upstream areas of the Mincio River and the discharge of partially treated wastewaters contributed to the eutrophication of Mantuan water bodies^{li}.

5.1.2.2. *The Mantuan contaminated site*

The environmental concerns deriving from the contamination of Mantuan water bodies brought to the emanation of the law L 179/2002^{lii}. With this action, the “Lakes of Mantua and Chemical Hub” have been included in the National Remediation Program.

The contaminated site was institutionalized as a Site of National Interest (SNI) in relation to its characteristics, the quantity and danger of pollutants, the impact on the surrounding environment in health and ecological terms as well as damage to cultural and environmental heritage, in compliance to Ronchi’s legislation (DLGS 192/1997^{liii}), ministerial decree DM 471/1999^{liv}, and the Single Text on the Environment (DLGS 152/2006^{lv}). With the ministerial decree released on February 3rd 2003^{lvi}, the area was circumscribed for a total surface of nearly 10.3 km², including both public and private districts. Ten years later, as a further measure, it has been forbidden to farm in the areas included in the SNI^{lvii}.

The area was segmented in three main compartments^{lviii} (Figure 5.4), i.e., Middle and Lower Lakes (~ 2.5 km²), Chemical Hub (~ 3.5 km²), and the natural conservation area “La Vallazza” along with a tract of the Mincio River and the wetlands (~ 4.5 km²). The last compartment is the environmental target of the contamination deriving from the Chemical Hub.

^{li} https://www.arpalombardia.it/sites/DocumentCenter/Documents/Stato%20delle%20acque%20superficiali%20-%202014-2019/Rapporto_2014_2019_MANTOVA_def.pdf (accessed 30/12/2022)

^{lii} <https://www.gazzettaufficiale.it/eli/id/2002/08/13/002G0218/sg> (accessed 28/12/2022)

^{liii} <https://www.gazzettaufficiale.it/eli/id/1997/02/15/097G0043/sg> (accessed 28/12/2022)

^{liv} <https://www.gazzettaufficiale.it/eli/id/1999/12/15/099G0540/sg> (accessed 28/12/2022)

^{lv} <https://www.gazzettaufficiale.it/dettaglio/codici/materiaAmbientale> (accessed 28/12/2022)

^{lvi} https://www.gazzettaufficiale.it/atto/serie_generale/caricaDettaglioAtto/originario?atto.dataPubblicazioneGazzetta=2003-04-12&atto.codiceRedazionale=03A04689&elenco30giorni=false (accessed 28/12/2022)

^{lvii} <https://www.comune.mantova.it/index.php/area-documentale/file/4593-ordinanza-divieto-di-coltivazione> (accessed 28/12/2022)

^{lviii} <https://www.arpalombardia.it/Pages/Bonifica/Mantova.aspx> (accessed 28/12/2022)

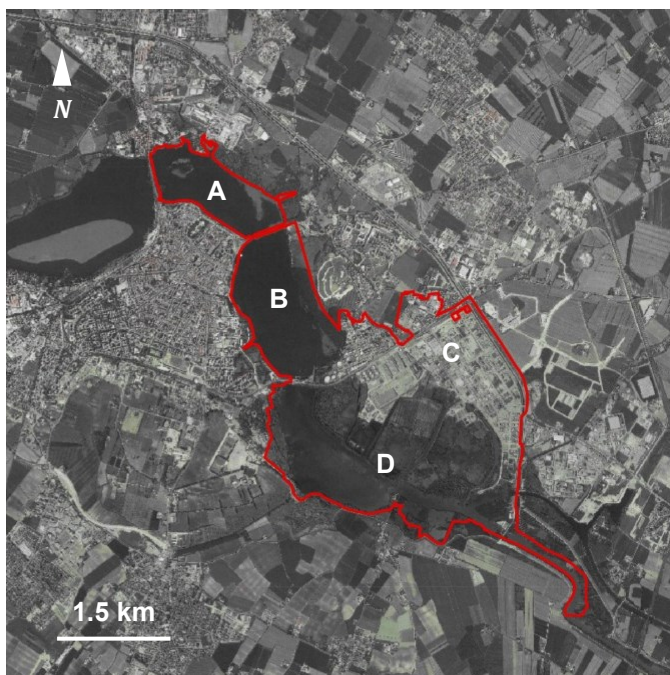


Figure 5.4. Delimitation of the Mantuan SNI. The red line denotes the delimitation, and the different letters denotes the different environmental compartments: Middle and Lower Lakes (A and B, respectively), Chemical Hub (C), and natural conservation area "La Vallazza" along with a tract of the Mincio River and the wetlands (D). Adapted after BLOM C.G.R. S.p.A., flight IT2000 in 2007^{lix}.

Starting from the '70s, the areas within the SNI have been characterized by the local companies and the Regional Environmental Protection Agency of Lombardy (ARPA Lombardia)^{lx, lxi}.

Since 2013, ARPA Lombardia carries out sampling campaigns across the site every other year, to evaluate the evolution over times of contamination within the SNI areas^{lxii}. The polluted environmental matrices include:

- **Soil and subsoil:** heavy contamination of hydrocarbons, aromatic organic compounds (benzene, toluene, ethylbenzene, xylenes, and styrene), metals (especially Hg), polychlorobiphenyls, dioxins and furans.
- **Groundwater:** heavy contamination of hydrocarbons, aromatic organic compounds (benzene, toluene, ethylbenzene, and xylenes), styrene and cumene (in the neighboring of the petrochemical facility), methyl-*t*-butyl ether and ethyl-*t*-butyl ether (in the neighboring of the ex-refinery), and

^{lix} <https://www.comune.mantova.it/index.php/area-documentale/file/818-perimetrazione-sin> (accessed 29/12/2022)

^{lx} <https://www.comune.mantova.it/index.php/area-documentale/file/805-scheda-sin> (accessed 28/12/2022)

^{lxi} <https://www.arpalombardia.it/Pages/Bonifica/Mantova/Bonifica.aspx?firstlevel=Mantova> (accessed 28/12/2022)

^{lxii} <https://www.arpalombardia.it/Pages/Bonifica/Mantova/Dati-Ambientali.aspx?firstlevel=Mantova> (accessed 29/12/2022)

metals (especially Hg). In addition, the groundwater drawn from piezometers around the ex-refinery and the petrochemical plant has shown a layer of organic phase floating on top of the water (*supernatant*). The latter has been identified as the primary source of contamination.

- **Sediments:** the sediments on the bottom of the canals, the Mincio River and the wetlands are contaminated with heavy hydrocarbons (C > 12), dioxins and metals (especially Hg).

Hydraulic barriers have been installed around the companies within the Chemical Hub to prevent the leakage of contaminated groundwater.

In the last technical board in 2020^{lxiii}, ARPA Lombardia, the ex-Italian Ministry of the Environment and the Protection of the Territory and the Sea, the Regional Council of Lombardy, and Edison S.p.A. have decided to extend the investigation area to thoroughly characterize both the contaminated site and areas right outside the SNI to better plan an intervention program.

5.1.3. Aromatic organic compounds in the Mantuan SNI

As already stated, in the context of the Mantuan SNI^{lxiv}, high concentration levels of aromatic organic compounds have been detected.

Such compounds comprise the low-molecular weight derivatives of benzene. This class of volatile organic compounds, usually known with the acronym BTEX (benzene, toluene, ethylbenzene, and xylenes), poses serious risks on human health, as few of them are certainly recognized as mutagenic and/or carcinogenic [44,45]. BTEX have been found with various concentration levels and in different locations in the groundwater drawn from the piezometers within the Chemical Hub.

High concentrations of styrene and cumene have been found in the neighboring of the petrochemical facility, whereas the concentration of benzene ranged from under control to concentration more than 1000 times higher with respect to the contamination threshold fixed by the DLGS 152/2006^{lxv}. The data collected during the monitoring campaigns across the years suggest that the distribution of the abovementioned compounds reflects the production activities of the petrochemical facility (see Section 5.1.1.1.).

This study is part of the MACHY (*MACrophyte-HYdrochar bioremediation technology applied to contaminated sediments*) project, a multidisciplinary project carried out

^{lxiii} <https://www.mite.gov.it/bonifiche/conferenze-dei-serviz-contenuti/11845> (accessed 29/12/2022)

^{lxiv} <https://www.arpalombardia.it/sites/DocumentCenter/Documents/SIN%20Mantova/Relazione%20campagna%20a%20cque%202019.pdf> (accessed 29/12/2022)

^{lxv} <https://www.gazzettaufficiale.it/dettaglio/codici/materiaAmbientale> (accessed 28/12/2022)

at the University of Parma in collaboration with the University of Trento. The aim is the development of a clean and cost-effective bioremediation technology based on the combination of macrophytes (*V. spiralis*, see Section 5.1.2.) and hydrochar to remove nutrients and micropollutants (both organic and inorganic) from contaminated water and sediments of the Mantuan SNI. The approach aims at being circular, producing a precious resource, i.e., hydrochar, starting from the biomass deriving from the eutrophicated site of the Mantuan SNI. The produced hydrochar will be utilized in the bioremediation process itself for the adsorption of contaminants, but it can exploit its potential as a fuel and in many applications that requires a sorbent material, ranging from sewage treatment to sample preparation in analytical chemistry.

What is presented in this **Chapter** are the optimization and validation of a method of analysis based on solid-phase microextraction (SPME, see **Chapter 3**) followed by GC–MS for the quantitation of a pool of BTEX that have been found in past analytical records produced by ARPA Lombardia along with others detected during the preliminary qualitative characterization (Table 5.1). The SPME–GC–MS method will be applied for the evaluation of the remediation capabilities of the proposed technology by analyzing the pore water of microcosms set up in the laboratory.

Table 5.1. Overview of the analytes included in this study^{lxvi}.

Analyte	Properties of concern in EU ^a	Contamination threshold ($\mu\text{g/L}$)	
		Surface water ^b	Groundwater
Benzene	C, M	10	1
Toluene	C, M, <u>R</u>	5	15
Ethylbenzene	C, M	– ^c	50
<i>o</i> -Xylene	n.n.	5	–
<i>m</i> -Xylene	n.n.	5	10 ^d
Styrene	<u>R</u>	–	25
Cumene	n.n.	–	–
Durene	n.n.	–	–

^a **bold**: recognized property; plain: broad agreement about the property; underlined: property under evaluation; **grey**: property reported in a minority of notifications; C: carcinogenic; M: mutagenic; R: toxic to reproduction; n.n.: not notified. ^b annual average value. ^c unregulated. ^d referred to *p*-xylene.

^{lxvi} The properties of concern were retrieved from the European Chemical Agency (<https://echa.europa.eu/it/home>, accessed 30/12/2022), whereas the contamination threshold values were retrieved from the DLGS152/2006 (<https://www.gazzettaufficiale.it/dettaglio/codici/materiaAmbientale>, accessed 30/12/2022)

5.2. Materials and methods

5.2.1. Chemicals and materials

Methyl alcohol (> 99.8%) was obtained from Honeywell (Charlotte, USA) and NaCl (> 99.99%) was obtained from VWR Chemicals (Milan, Italy).

Benzene (B, > 99.8%) and ethylbenzene (E, > 99.5%) were obtained from Carlo Erba (Milan, Italy). *m*-xylene (*m*X, > 99%), *o*-xylene (*o*X, > 98%), styrene (S, > 99%), and perdeuterated *m*-xylene (*m*X-d₁₀, ≥ 99%; deuteration grade > 98%) were obtained from Sigma-Aldrich (Milan, Italy). Cumene (C, > 99%) and durene (D, 98%) were obtained from Merck (Darmstadt, Germany). Toluene (T, > 99.7%) was obtained from Honeywell.

Seventy-five micrometers Carboxen™-PDMS SPME fibers (length: 1 cm) were obtained from Supelco (Bellefonte, USA). Before use, the fiber was conditioned by placing it in the GC injection port under a constant flow of helium according to the manufacturer instructions.

Working solutions were prepared by proper dilution of the analytes and *m*X-d₁₀ in methanol. Diluted solutions were kept at -20 °C in the dark until use.

5.2.2. Optimization of the extraction procedure

Optimization of the SPME procedure was carried out by performing extraction experiments on 4 mL of aqueous solutions containing the analytes at a concentration of 500 ng/L. SPME was operated in headspace mode (HS). The response variable was the peak area of each analyte.

Three main factors ($k = 3$) were investigated at three levels within their respective experimental domain according to a Box-Behnken Design (BBD): extraction time (X_1 , 5–15 min), extraction temperature (X_2 , 40–60 °C), and NaCl addition (X_3 , 0–1 g). In total, $N = 17$ experiments were carried out, including $n_0 = 5$ experiments in the center of the experimental domain to estimate the pure experimental variance.

A full second-order model was postulated. A coefficient was considered significant if the absolute value of its regression coefficient was greater than the semiamplitude of its 95% confidence interval. Each model was evaluated in terms of fraction of explained variance (R^2) and validity expressed in terms of lack-of-fit ($\alpha = 0.01$).

The global optimal conditions were identified according to Derringer's method [46]. Single desirability functions d_i were defined for each analyte according to Equation 5.1:

$$d_i = \begin{cases} 0, & y_i < L_i \\ \frac{y_i - L_i}{U_i - L_i}, & \text{otherwise} \end{cases}$$

Equation 5.1

L_i is the upper limit of the 95% confidence interval of the minimum predicted response and U_i is the maximum predicted response within the experimental domain. The optimal conditions were in correspondence of the maximum global desirability (Equation 2.42), computed as unweighted geometric mean of the single desirability functions. The maximum was found with a derivative-free search algorithm.

5.2.3. Operating procedure and instrumental conditions

5.2.3.1. Sample preparation

A volume of 4 mL of sample was introduced into 10 mL clear glass vials. One gram of NaCl was introduced into the vial. The aliquot was spiked with the internal standard (IS) solution so that its final concentration was 25 ng/L. The IS was $mX-d_{10}$.

Each sample was equilibrated for 5 min at 57 °C. Thereafter, the fiber was exposed to the headspace above the sample for 11 min at 57 °C under constant agitation (250 rpm). Then, the fiber was placed in the GC injection port at 300 °C for 2 min. These operations were carried out with the aid of a PAL COMBI-xt autosampler (CTC Analytics AG, Zwingen, Switzerland).

5.2.3.2. GC-MS

Instrumental analyses were carried out with a HP 6890 Series Plus gas chromatograph hyphenated with an MSD 5973 mass spectrometer (both by Agilent Technologies, Milan, Italy).

The split/splitless injector temperature was 300 °C. The injection was executed in split mode (split ratio: 1:1) into a 5190-4056 ultra-inert liner (Agilent Technologies). An Rxi-5Sil MS capillary column (30 m length × 0.25 mm i.d., 0.25 μm film thickness; Restek, Bellefonte, USA) was used for the chromatographic separation with the following temperature program: 40 °C held for 2 min, 4 °C/min to 50 °C, 15 °C/min to 200 °C, held for 1.50 min (runtime: 16.00 min). The carrier

gas was helium (constant flow rate: 1.4 mL/min). The transfer line was maintained at 280 °C. The single quadrupole mass spectrometer was operated in electron ionization mode (EI; 70 eV). The ion source and the quadrupole were held at 230 and 150 °C, respectively. Mass spectra were registered in selected ion monitoring (SIM) mode (dwell time: 30 ms; electron multiplier voltage: 1.047 kV). The acquisition started 0.50 min after the injection. One microliter of a solution containing both the analytes and the IS (concentration: 1 mg/L) in methyl alcohol was injected and the mass range 50–350 m/z was recorded to properly set the time scheduled SIM program. The retention times and the m/z ratios that were monitored for each compound are reported in Table 5.2.

Table 5.2. Time scheduled SIM program. Qualifier ions are reported in plain text, quantifier ions are reported in bold text.

Time (min)	Compound	Monitored m/z ratios
0.50–1.80	B	78
1.80–3.00	T	91, 92
3.00–4.10	E, mX $mX-d_{10}$	91, 106 98
4.10–4.80	oX S	91, 106 103, 104
4.80–6.00	C	105, 120
6.00–16.00	D	119, 134

5.2.4. Method validation

The SPME–GC–MS method was validated by working under the optimized conditions according to the EURACHEM guidelines [47]. The validation study was carried out with matrix-matched standards that were subjected to the whole analytical process. Drinking water was the blank matrix.

Limits of detection and limits of quantitation (LODs and LOQs, respectively) were evaluated by estimating the standard deviation s_0 of $m = 10$ independent replicated measurements of blank matrix fortified with a detectable amount of analytes. The average signal of the blank y_0 was estimated by submitting to analysis $n = 4$ procedural blanks. Detection and quantitation limits, expressed in the signal domain, were calculated according to Equation 2.43 and 2.44, respectively, considering that $n = 2$ replicated measurements will be averaged when reporting the results. The corresponding concentrations were calculated by their projection on the x axis. Calibration functions were evaluated on $k = 9$ levels in the LOQ–1000 ng/L range ($n = 2$ independent replicated measurements *per* level) for all the

analytes. The intercept was tested for significance with the Student's *t*-test ($\alpha = 0.05$). Linearity and validity were evaluated, respectively, with Mandel's test and lack-of-fit test ($\alpha = 0.01$).

Repeatability and intermediate precision were evaluated in a single study by carrying out independent triplicated measurements *per* day for three days by three different operators. Repeatability and intermediate precision standard deviations were estimated according to Equation 2.48 and 2.49 and expressed as relative standard deviations (*RSD*_%). ANOVA was applied also to assess whether the results obtained under intermediate precision conditions were significantly different among each other ($\alpha = 0.05$). Trueness was expressed as spike recovery rate (*RR*'_%, Equation 2.52) by submitting $n = 9$ independent replicated measurements to the whole analytical process. Precision and trueness were evaluated at 40, 350, and 800 ng/L.

5.2.5. Evaluation of bioremediation capability

5.2.5.1. Plants collection and hydrochar production

V. spiralis plants were harvested from an eutrophicated site of the Mincio river in the neighboring of Massimbona (45° 16' 42.6" N, 10° 43' 4.8" E; Mantua, Italy). A sample of 100 individuals was randomly collected, taking care in avoiding any damage to the roots. The plants were thoroughly rinsed with river water to remove residuals of sediments and epiphyte plants.

Hydrochar was produced with the pilot plant located at the University of Trento (Trento, Italy) by utilizing various vegetable scraps (*Phragmites australis*, *Nelumbo nucifera*, and *Vallisneria spiralis*) in variable proportions as starting material.

5.2.5.2. Sediment collection and processing

Sediments (1 m depth) and the water going along with them were collected from multiple locations from the Lower Lake of the Mantuan SNI.

Few hours after sampling, the subsamples were reconstituted for a total of 20 L of sample. The reconstituted sample was passed through a screening panel (mesh: 1 mm) and collected in a bucket. Thereafter, the sample was artificially contaminated with a mixture of the investigated analytes so that their final concentration in the sediments was approximately 5 mg/L.

5.2.5.3. Experimental setup

Microcosms were setup into cylindrical PVC tubes (10 cm height × 8 cm i.d.) filled for the 90% of their volume with contaminated sediment. Microcosms were transferred into an aquarium tank filled with tap water maintained at 28 °C. The tank was exposed continuously to a halogen lamp (2000 W) producing an illuminance of $400 \mu\text{mol} \times \text{s}^{-1} \times \text{m}^{-2}$ to mimic the average photosynthetically active radiation at the sampling site. Aquarium pumps were utilized to guarantee water recirculation. The whole system was maintained under a hood.

Four treatments were set up, for a total of $N = 48$ microcosms:

- **Control (C):** the control experiment was represented by the contaminated sediment itself.
- **Hydrochar (H):** 10 g of hydrochar were added to the contaminated sediment.
- **Plant (P):** two plants of *V. spiralis* were planted into the contaminated sediment.
- **Hydrochar + Plant (HP):** 10 g of hydrochar were added to the contaminated sediment and two plants of *V. spiralis* were planted.

The microcosms were incubated for 1, 15, and 30 days. In total, $n = 4$ replicates were set up *per* treatment and *per* incubation time. After the incubation period had elapsed, $n = 4$ microcosms *per* treatment were randomly removed from the tank and 10 mL of pore water were drawn with a plastic syringe equipped with a 2.5 mm Rhizon with an average pore size of $0.15 \mu\text{m}$ (Rhizosphere Research Products B.V., Wageningen, Netherlands). The filtered samples were transferred into 10 mL clear glass vials so that they were filled to their maximum capacity and sealed with poly(tetrafluoroethylene) (PTFE)/silicone septa. The samples were kept in the dark at 4 °C and were analyzed within 24 hours from sampling. Samples were treated according to what is reported in Section 5.2.3 after proper dilution with drinking water.

5.2.5.4. Data analysis

The results deriving from the experimentation were organized into a data matrix **X** (48×8): each row representing a microcosm and each column representing the concentration of each analyte (expressed as ng/L) deriving from the SPME–GC–MS determination. Concentration values below the LOQ were removed and treated as

missing data. Each row was enriched with additional information, i.e., total concentration, incubation time, and treatment.

Exploratory Data Analysis (EDA) was carried out. Principal Component Analysis (PCA) [48–50] was the chosen technique for this purpose (see **Chapter 2**, Section 2.1.1.1). Before model computation, data were transformed by taking their natural logarithm and were mean centered.

5.2.6. Software

GC–MS data were handled with the HP ChemStation (Agilent Technologies) software. Multilinear regression and multicriteria optimization were carried out by using custom scripts in MATLAB environment (v. R2022a, Mathworks, Massachusetts, USA).

PCA was carried out by utilizing the built-in function in MATLAB environment. Among the available algorithms, Alternating Least Squares (ALS) was chosen as it tolerates missing values [51].

5.3. Results and discussion

5.3.1. Optimization of the SPME procedure

As this part of the project focused on the determination of VOCs, the SPME fiber had to be chosen accordingly [52,53]. Based on the ISO 17943:2016^{lxvii} for the quantitation of VOCs in water, either the Carboxen™-PDMS or DVB/Carboxen™-PDMS fibers can be utilized. In a preliminary evaluation both coatings were tested, but difficulties were encountered when utilizing the DVB/Carboxen™-PDMS fiber as highly persistent carryover effect was observed (data not shown). Accordingly, the Carboxen™-PDMS fiber was chosen for further method development.

As it was discussed earlier in **Chapter 3**, different parameters can affect the efficiency of the SPME procedure. Besides extraction time and temperature, ionic strength can also play a role, especially when operating in HS-mode [52,53]. The abovementioned factors were selected to be investigated in optimizing the SPME procedure. The experimental domains were set by taking into consideration operative limitations:

- **Extraction time:** extraction times shorter than 5 min were not considered as poor GC–MS signals were produced, whereas a maximum extraction

^{lxvii} <https://www.iso.org/standard/61076.html> (accessed 26/12/2022)

time of 15 min was selected to guarantee a good balance between the time required for the extraction and the analysis time to maximize the sample throughput.

- **Extraction temperature:** poor responses were observed for D below 40 °C, therefore it was selected as the lowest extreme of the experimental domain for this factor. On the other hand, temperatures higher than 60 °C were not considered as high of temperatures promote the desorption of the analytes from the fiber.
- **Ionic strength:** in HS-mode, the addition of NaCl to the sample solution can promote the volatilization of nonpolar analytes in the headspace due to the augmented ionic strength of the medium. On the contrary, too high of concentration of NaCl increases the viscosity of the sample solution, hindering the diffusion of the analytes towards the headspace. The lower extreme of the experimental domain corresponded to no addition of NaCl, whereas the upper extreme was set at 1 g.

The responses were left untransformed, and the models were calculated. ANOVA showed that all the models were valid ($p > 0.01$), meaning that the error deriving from approximation was not significantly greater than the variance that it could be expected experimentally. The models showed R^2 in the 0.71–0.87 range. Cross-validation was not carried out as when the center point was left out the matrix determinant was zero and regression coefficient could not be calculated. For the sake of brevity, B, ρX , and D surface plots were shown to exemplify the typical shapes of response surfaces (Figure 5.5), whereas the regression models are reported in Table 5.3.

As far as only the main factors were concerned, they had always positive coefficients, with the only exception of X_2 on B, suggesting that working at too high of a temperature could have promoted the desorption of the analyte from the fiber before the injection. As for X_1 and X_3 , when significant they had always a positive coefficient, meaning that the analytes benefitted from longer extraction times and the addition of NaCl produced an effective salting-out [52,53].

Table 5.3. Regression models calculated for each analyte.

Analyte	Equation ^a
B	$y = 1350000 (\pm 21000) - 106000 (\pm 15000)X_2 + 196000 (\pm 15000)X_3 + 196000 (\pm 21000)X_2X_3 - 202000 (\pm 21000)X_1^2 - 205000 (\pm 21000)X_2^2$
T	$y = 1453000 (\pm 24000) + 59000 (\pm 19000)X_1 + 57000 (\pm 19000)X_2 + 53000 (\pm 19000)X_3 - 79000 (\pm 27000)X_1X_3 - 101000 (\pm 27000)X_1^2 - 114000 (\pm 27000)X_2^2$
E	$y = 439700 (\pm 7600) + 20100 (\pm 6000)X_2 + 30900 (\pm 8500)X_1X_2 - 36000 (\pm 8300)X_1^2 - 37400 (\pm 8300)X_2^2$
mX	$y = 1218000 (\pm 20000) + 45000 (\pm 16000)X_1 + 58000 (\pm 16000)X_2 + 46000 (\pm 16000)X_3 + 74000 (\pm 22000)X_1X_2 - 89000 (\pm 21000)X_1^2 - 102000 (\pm 21000)X_2^2$
oX	$y = 504700 (\pm 7300) + 23700 (\pm 5200)X_1 + 42500 (\pm 5200)X_2 + 26800 (\pm 7300)X_1X_2 - 28400 (\pm 7300)X_1X_3 + 33100 (\pm 7300)X_2X_3 - 35800 (\pm 7300)X_1^2 - 35200 (\pm 7300)X_2^2 - 26700 (\pm 7300)X_3^2$
S	$y = 143600 (\pm 6300) + 16500 (\pm 4900)X_2 + 17600 (\pm 4900)X_3$
C	$y = 468600 (\pm 5300) + 20000 (\pm 3700)X_1 + 21800 (\pm 3700)X_2 + 20700 (\pm 3700)X_3 + 30200 (\pm 5300)X_1X_2 - 45500 (\pm 5300)X_1^2 - 30800 (\pm 5300)X_2^2$
D	$y = 89400 (\pm 2100) + 6200 (\pm 1700)X_1 + 9000 (\pm 1700)X_2 - 9500 (\pm 2300)X_1^2$

^a only the significant terms at the 95% confidence level are reported, the coefficients (X_1 : extraction time; X_2 : extraction temperature; X_3 : salt addition) are reported as coefficient (\pm standard error) rounded at two significant digit.

Various first-order interactions were present, involving different main factors and with different signs depending on the analyte. The interaction X_1X_2 was significant for E, mX, oX, and C, always with a positive coefficient, and such analytes benefitted of long extraction times and high extraction temperatures. Similarly, the interaction X_2X_3 was significant with a positive coefficient for B and oX, meaning that better extraction was achieved at high temperatures and at high NaCl concentrations. The interaction X_1X_3 was significant with a negative sign for T and oX, meaning that longer extraction times were the most ideal when lower amounts of NaCl were added and *vice versa*. In addition, X_1 and X_2 were present in most of the models as squared terms with a negative sign, meaning that the responses increased until a certain point and then reached a plateau or decreased.

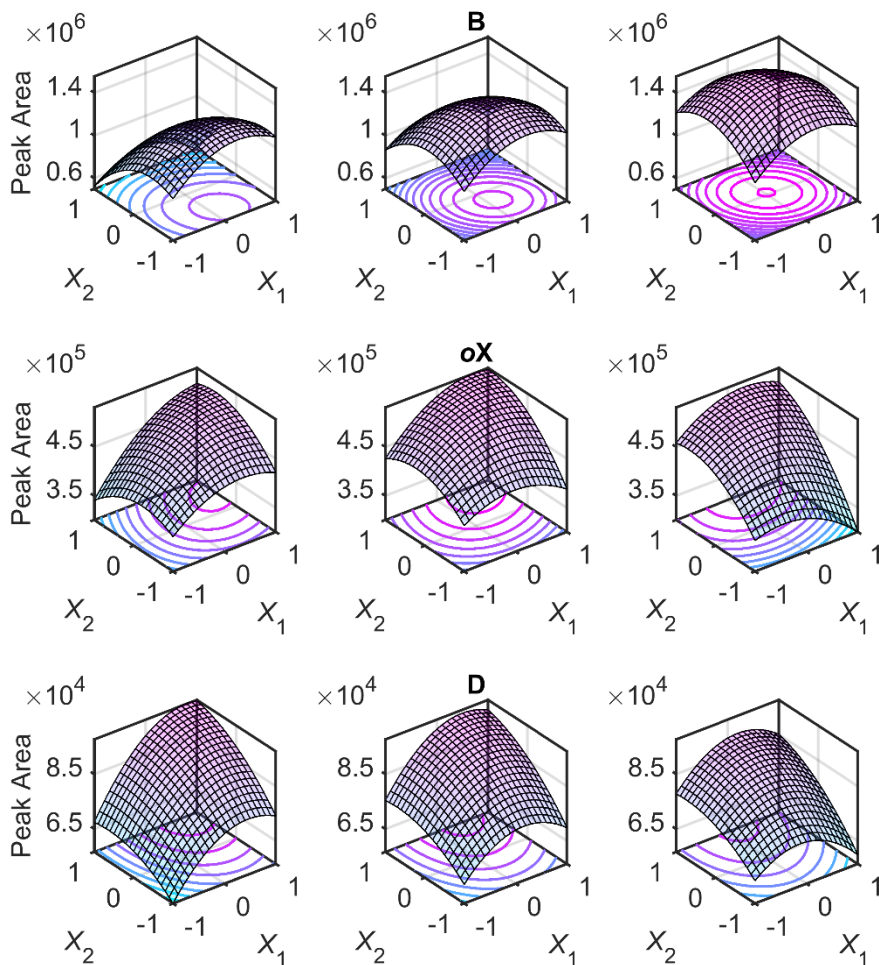


Figure 5.5. Response surfaces of B, oX, and D. The predicted response for each analyte is plotted as a function of extraction time and extraction temperature (X_1 and X_2 , respectively) at three different levels of the amount of the amount of NaCl added to the sample solution (X_3): -1 (left column), 0 (middle column), and +1 (right column).

The optimal conditions required an extraction time of 11 min and an extraction temperature of 57 °C, with the addition of 1 g of NaCl to the sample solution. The global desirability corresponding to such experimental conditions was $D = 0.92$, with single desirability values $d_i \geq 0.71$. A non-zero global desirability value means that a set of conditions able to fulfill the optimization criteria for all the responses could be found, whereas its high value denotes a high degree of accordance between the single desirability values. The identified experimental conditions were valid, as the lack-of-fit was not statistically significant ($p > 0.01$).

5.3.2. Method validation

The SPME–GC–MS method was validated operating under optimized conditions. The LODs, LOQs, and equation of the calibration functions are reported in Table 5.4. The method exhibited LODs in the 15.1–49.8 ng/L range, suitable for the detection of the investigated analytes at trace and ultra-trace levels, whereas LOQs were in the 20.3–76.3 ng/L range. In terms of LODs and LOQs, the results were extremely satisfactory, allowing the determination of the regulated analytes far below the contamination threshold established by the DLGS 152/2006 (see Table 5.1), both for surface water and groundwater.

Table 5.4. LODs, LOQs, and calibration curve equations of the SPME–GC–MS analytical method. LODs and LOQs are reported in ng/L.

Analyte	LOD	LOQ	Calibration curve equation ^a	
			$b_0 \pm SE^b$	$b_1 \pm SE^b$
B	28.1	55.2	– ^c	4.30(±0.04)
T	42.6	62.9	–	5.06 (±0.05)
E	15.1	22.1	–	2.68 (±0.02)
<i>mX</i>	49.4	76.3	–	4.32 (±0.05)
<i>oX</i>	36.6	51.5	–	2.42 (±0.02)
S	49.8	52.8	-2.1 (±0.5)	1.52 (±0.03)
C	22.2	24.0	-4 (±1)	5.60 (±0.07)
D	19.5	20.3	-1.7 (±0.4)	2.41 (±0.02)

^a the equation of the calibration curve is $y = b_0 + b_1x$. ^b the intercept and slope are reported as coefficient \pm standard error rounded at one significant digit. ^c – not significant ($p > 0.05$).

All the calibration functions proved being linear ($p > 0.01$) in the LOQ–1000 ng/L range and statistically significant ($p < 0.05$), with linearity holding nearly 2.5 orders of magnitude. In addition, the approximation error was not significantly greater than the expected pure experimental variation ($p > 0.01$). Satisfactory results were obtained from the precision and trueness studies, with the results summarized in Table 5.5 and in Figure 5.6, respectively. There were not statistically significant differences in the results obtained under intermediate precision conditions.

Another figure of merit of the developed analytical method was the good balance between the time required for extraction (5 min equilibration + 11 min extraction) and chromatographic separation (16 min). This results in a good overlap between chromatographic runs guaranteeing a fast sample throughput, desirable for routine application and for many samples to be processed. In addition to this, the method did not require the utilization of any organic solvent for extraction, as all the application with SPME hyphenated to GC [52,53].

Table 5.5. Results of the precision study. $RSD\%$ referring to repeatability is marked with an r in apex and the one referring to intermediate precision is marked with an I in apex. Results are rounded at the nearest integer.

Analyte	L1 ^a		L2 ^b		L3 ^c	
	$RSD^r\%$	$RSD^I\%$	$RSD^r\%$	$RSD^I\%$	$RSD^r\%$	$RSD^I\%$
B	n.d. ^d	n.d.	11	13	9	13
T	n.d.	n.d.	18	20	13	19
E	11	12	11	12	7	8
mX	n.d.	n.d.	13	13	14	20
oX	n.d.	n.d.	14	15	14	18
S	n.d.	n.d.	10	14	7	10
C	11	13	11	13	11	14
D	17	20	15	16	15	16

^a 40 ng/L. ^b 350 ng/L. ^c 800 ng/L. ^d not determined.

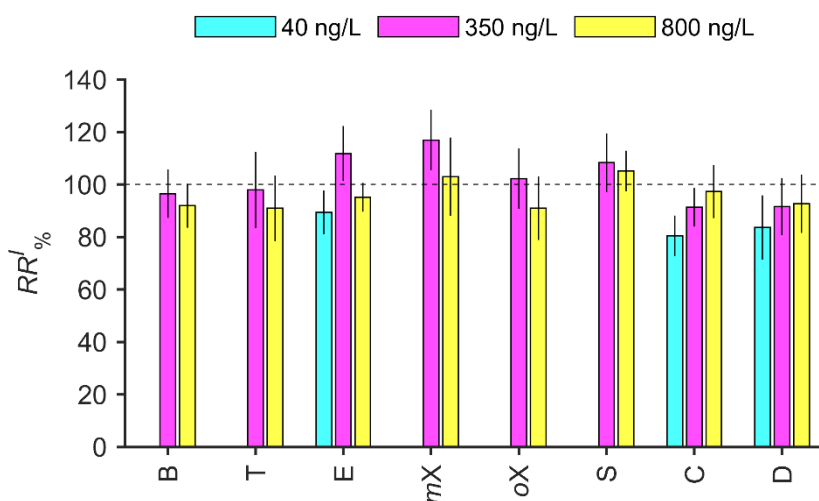


Figure 5.6. Results of the trueness study. Results are reported as mean \pm semiamplitude of the 95% confidence interval ($n = 9$). The horizontal dashed line corresponds to a $RR' \%$ = 100%.

5.3.3. Evaluation of bioremediation capability

The optimized and validated SPME-GC-MS method was utilized in evaluating the bioremediation capabilities of *V. spiralis* and the inclusion of hydrochar in the sediment. After pretreating the sediment and artificially contaminating it, it was subjected to four treatments, including the control. The control experiment was represented by the contaminated sediment itself subjected to the same experimental conditions of the remaining treatments to assess the variation in concentration of the pollutants over time.

This aspect is fundamental, since a loss in concentration could be expected due to:

- Dilution, through the diffusion of the analytes from the sediment towards the water contained in the aquarium tank.
- Volatilization, through the diffusion of the analytes from the water to the atmosphere, as the system was open to guarantee gas exchange.

Since the sampling distribution of concentrations for environmental data is notoriously log-normal, the results were logarithmically transformed (base: e) to make them more normal [54]. Figure 5.7 shows a summary of the transformed results for each analyte.

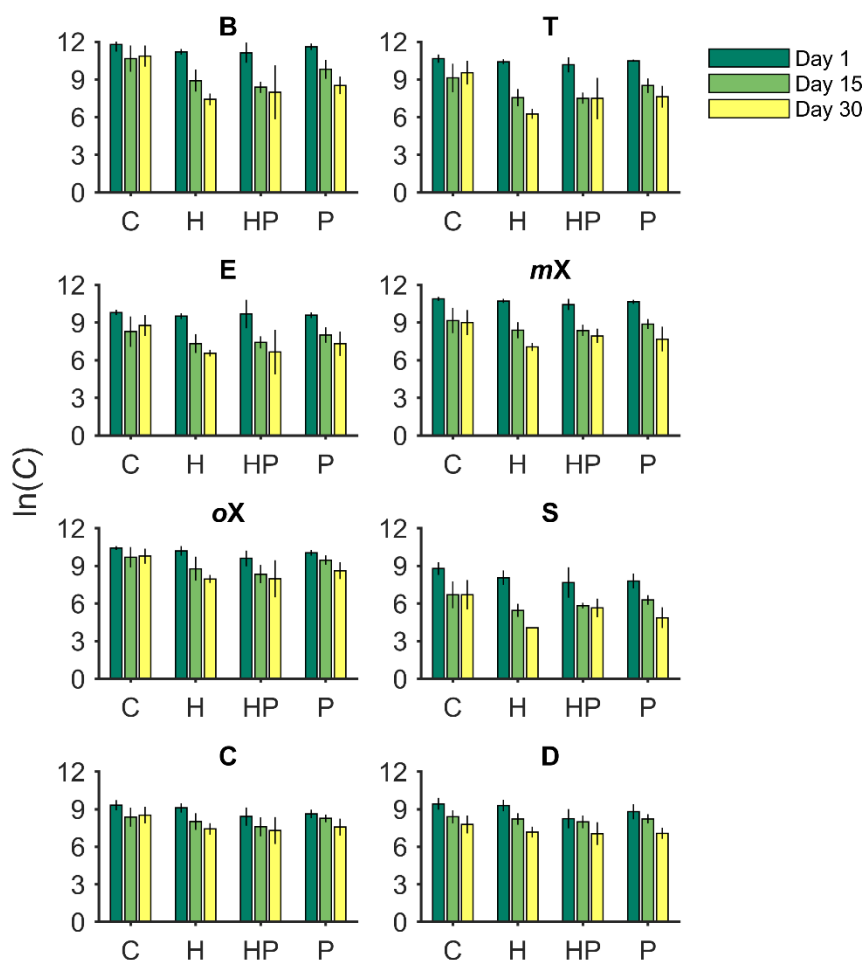


Figure 5.7. Summary of the results of the bioremediation study. Results were logarithmically transformed (original measurement unit: ng/L). Results are plotted as mean \pm one standard deviation ($n = 4$). Each plot is referred to a different analyte, reported in bold in the title. The x axis reports the treatments (C: control; H: hydrochar-enriched sediment; HP: vegetated hydrochar-enriched sediment; P: vegetated sediment). Each bar represents a different sampling time as shown in the legend.

5.3.3.1. Exploratory data analysis

The harvested data were rationalized by carrying out EDA. This was accomplished by subjecting the dataset to PCA in order to highlight the presence of potential data structures such as trends and clusters.

Figure 5.8 shows the decomposition on the two low-order Principal Components (PCs).

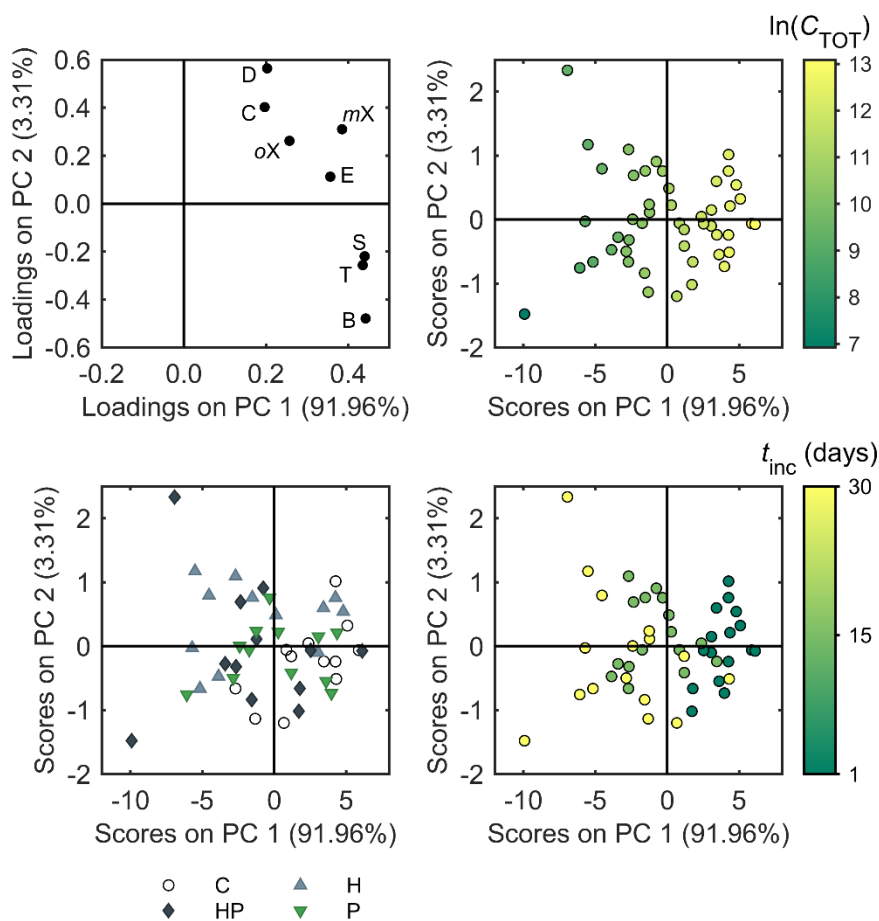


Figure 5.8. Results of the PCA decomposition for the two low-order PCs (the fraction of total variance that a given PC explains is reported on the corresponding axis). The top-left plot reports the distribution of the variables in the loadings space, whereas all the others report the distribution of objects in the score space, marked according to different additional information. Top-right: objects are marked according to the logarithm of total concentration (original measurement unit: ng/L) of pollutants quantified in the pore water. Bottom-left: objects are marked according to the treatment (empty circles (C): control; grey triangles (H): hydrochar-enriched sediment; charcoal diamonds (HP): vegetated hydrochar-enriched sediment; inverted green triangles (P): vegetated sediment). Bottom-right: objects are marked according to the incubation time (days).

The first two low-order PCs accounted for a little more than 95% of the total variance, with PC 1 capturing nearly 92% of it. The loadings related to all the analytes were positive with respect to PC 1. This provided an indication that the effect with the largest magnitude, i.e., the one captured mostly by PC 1, was the variation in concentration of pollutants in the pore water. This information can be retrieved by inspecting the score plot, where the objects were colored according to the overall concentration, shading from yellow (high concentration) to dark green (low concentration) moving from positive to negative scores on PC 1.

A similar tendency was observed for the incubation time: microcosms incubated for 1 day (dark green) were on positive scores for PC 1, whereas microcosms incubated for 30 days (yellow) were on negative scores for PC 1, and microcosms incubated for 15 days (light green) were around the origin of the axes. This trend was in contrast with the one observed for concentration and, according to the joint interpretation of score and loading plots, it suggests that a reduction in concentration could be expected as more and more times had elapsed. This fact could be related both to the volatilization of the analytes over time since, as it was stated earlier, the system was open and also to an actual reduction thanks to the treatment with *V. spiralis* and hydrochar.

As for the treatment, despite no clear clusters being visible, it is worth noting that most of microcosms that received no treatment, i.e., control, were located on positive scores on PC 1, and all the remaining treatments were, on average, at negative scores on the same PC. According to the loading plot, this means that the control experiments had, on average, a higher concentration than the microcosms subjected to any of the treatments, independently of the incubation time. The only exception was represented by the microcosms that were incubated only for 1 day, which were all similar in terms of concentration regardless the treatment. Therefore, the treatments had an effect in reducing the concentration of pollutants in the pore water. In terms of which pollutants were removed, by the inspection of the loading plot it was observed that the loadings related to C and D were lower than the others on PC 1, suggesting that they were more resistant to the treatments in diminishing their concentration.

The effect of the treatment with respect to the control experiment appeared even more evidently when the offset with respect to the incubation time was removed. This was accomplished by mean centering the groups of data with respect to the mean of each incubation time.

Figure 5.9 shows the decomposition of the block-centered data on the two low-order PCs, together capturing a little more than 95% of the total variance.

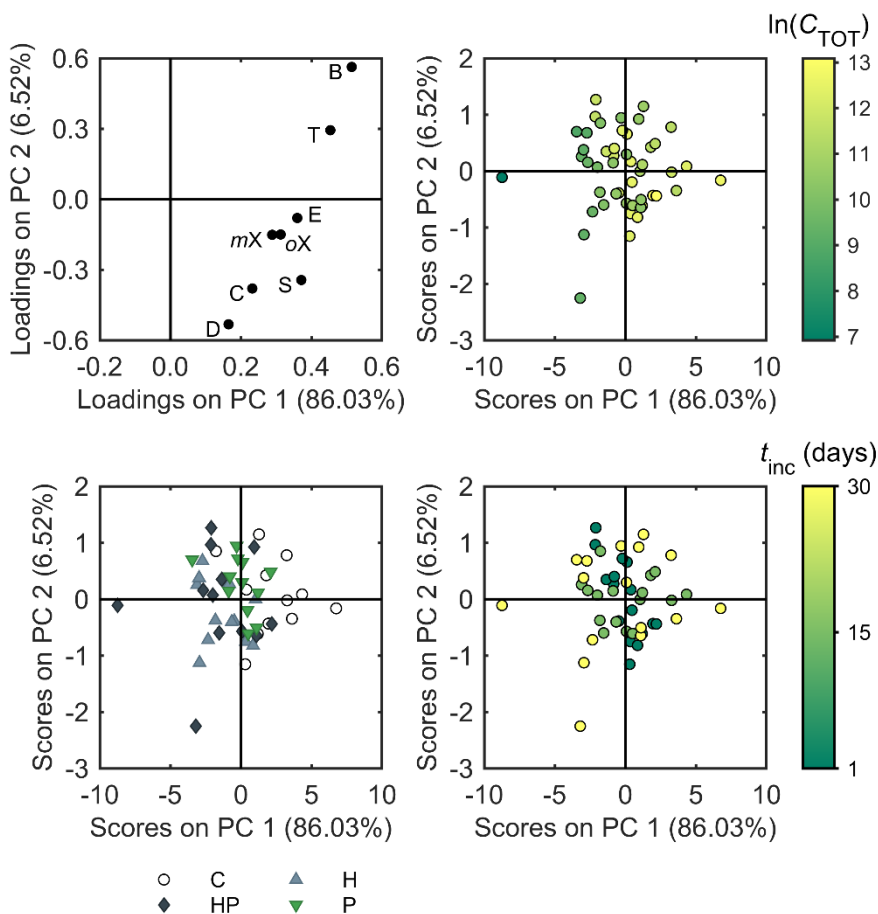


Figure 5.9. Results of the PCA decomposition of the block-centered data for the two low-order PCs (the fraction of total variance that a given PC explains is reported on the corresponding axis). The top-left plot reports the distribution of the variables in the loadings space, whereas all the others report the distribution of objects in the score space, marked according to different additional information. Top-right: objects are marked according to the logarithm of total concentration (original measurement unit: ng/L) of pollutants quantified in the pore water. Bottom-left: objects are marked according to the treatment (empty circles (C): control; grey triangles (H): hydrochar-enriched sediment; charcoal diamonds (HP): vegetated hydrochar-enriched sediment; inverted green triangles (P): vegetated sediment). Bottom-right: objects are marked according to the incubation time (days).

As it was for mean centered data (with respect to the global mean), all the variables had positive loadings on PC 1, meaning that the microcosms whose scores were negative on PC 1 (capturing the 86% of the total variance) were characterized by an overall lower concentration of pollutants. With the samples centered around the origin of the axes with respect to the incubation time, still remained evident how

the control experiments were all located on positive scores of PC 1 but one and the remaining treatments mostly on the negative side of the same PC. Within this frame of reference, the hydrochar-enriched sediments and the vegetated hydrochar-enriched sediments were more on the negative side of PC 1, whereas the vegetated sediments were closer to the origin. As it was observed for globally mean centered data, C and D were the analytes that resisted the most to the treatments, as they were closer to the origin with respect to the other loadings on PC 1.

These findings suggested that the remediation process benefitted from the inclusion of hydrochar into the sediments, that aids the removal of the investigated pollutants by adsorbing them: in fact, thanks to its surface chemistry [24,25,27], hydrochar could interact with aromatic analytes *via* π - π interaction, London dispersion forces, and temporary dipole interactions [55].

Given the fact that *V. spiralis* is capable, thanks to the ROL, in immobilizing certain metal ions, such as Fe^{2+} and Mn^{2+} [33,36–38], the reduced removal capabilities observed towards the investigated compounds were unexpected. The finding might still be relevant, as there is evidence in scientific literature of certain members of the *Bacteria* Kingdom being able to degrade BTEX both under aerobic and anaerobic conditions [56–58]: such bacteria could, therefore, benefit from an oxygen-enriched environment thanks to the eco-physiological characteristics of *V. spiralis* and other macrophytes. Nevertheless, the algal biomass could still be recovered and utilized for the production of hydrochar, that demonstrated its effectiveness in reducing the concentration of the investigated pollutants.

The interaction of different bacterial species can be intricate [58], especially in a complex matrix like sediments in the presence of environmental contaminants. This matter will be further investigated by the units of Analytical Chemistry and Environmental Microbiology of the University of Parma.

5.4. Conclusions

The results presented in this **Chapter** are the very first findings of the research activities involved in the MACHY project (*MACrophyte-HYdrochar bioremediation technology applied to contaminated sediments*). A method based on SPME-GC-MS was devised for the quantitation of BTEX, styrene, cumene, and durene micropollutants found in the frame of the Mantuan SNI. After optimization through the experimental design methodology, the method was validated. The quality parameters were suitable for the quantitation of the investigated analytes at trace and ultra-trace levels way below the contamination thresholds established by the

DLGS 152/2006. The developed analytical method was utilized to evaluate the capabilities of *V. spiralis*, hydrochar, and a combination of them in reducing the concentration of the investigated pollutants in sediments over a period of one month. The analytical results were rationalized by subjecting them to PCA, that highlighted how the incorporation of hydrochar, produced starting from vegetation infesting the eutrophicated Mantuan SNI, into the sediments produced a reduction in the content of the examined micropollutants. Also, the cultivation of *V. spiralis* had a depolluting effect, but it was smaller in magnitude. Further investigation will be carried out to clarify the effect of *V. spiralis* on the decontamination of sediments from these analytes, taking into consideration also the microbial population inhabiting the sediments. The approach outlined in this **Chapter** will be followed by the evaluation of the bioremediation capabilities towards other classes of micropollutants, including PAHs (see **Chapter 4**), linear hydrocarbons, and metals, starting with the development of methods based on GC-MS and ICP-OES, respectively. By far the results suggested that the bioremediation strategy proposed in this study has a potential depolluting effect. Hopefully, the MACHY project will provide a green, circular, and cost-effective approach to restore the ecological condition of the intricate hydrogeological network of Mantua.

Note of the author

The results presented in this **Chapter** are matter of ongoing research.

The author would like to express gratitude to Arianna Soci and Margherita Curti for participating to this study as a part of the internship required to obtain the Master of Science and Bachelor of Science degrees, respectively.

References

- [1] H. Sandhu, Bottom-Up Transformation of Agriculture and Food Systems, Sustainability. 13 (2021) 2171. <https://doi.org/10.3390/SU13042171>.
- [2] M. Tampaki, G. Koutouzidou, A. Ragkos, K. Melfou, I.A. Giantsis, Eco-Value and Public Perceptions for Indigenous Farm Animal Breeds and Local Plant Varieties, Focusing on Greece, Sustainability. 14 (2022) 11211. <https://doi.org/10.3390/SU141811211>.
- [3] P. Krasilnikov, M.A. Taboada, Amanullah, Fertilizer Use, Soil Health and Agricultural Sustainability, Agriculture. 12 (2022) 462. <https://doi.org/10.3390/AGRICULTURE12040462>.
- [4] P. Sharma, A. Gedanken, O. Sarti, F. el Mansouri, E. Otal, J. Morillo, A. Ouassini, J. Brigui, M. Saidi, Assessing the Effect of Intensive Agriculture and Sandy Soil Properties on Groundwater Contamination by Nitrate and Potential Improvement Using Olive Pomace Biomass Slag (OPBS), C (Basel). 9 (2022) 1. <https://doi.org/10.3390/C9010001>.
- [5] H.J. Beckie, K.C. Flower, M.B. Ashworth, Farming without Glyphosate?, Plants. 9 (2020) 96. <https://doi.org/10.3390/PLANTS9010096>.

- [6] D.A. Lemley, J.B. Adams, Eutrophication, in: B. Fath (Ed.), *Encyclopedia of Ecology*, 2nd ed., Elsevier, Amsterdam, Netherlands, 2019: pp. 86–90. <https://doi.org/10.1016/B978-0-12-409548-9.10957-1>.
- [7] M. Li, J. Dong, Y. Zhang, H. Yang, L. van Zwieten, H. Lu, A. Alshameri, Z. Zhan, X. Chen, X. Jiang, W. Xu, Y. Bao, H. Wang, A Critical Review of Methods for Analyzing Freshwater Eutrophication, *Water (Basel)*. 13 (2021) 225. <https://doi.org/10.3390/W13020225>.
- [8] M. le Moal, C. Gascuel-Oudou, A. Ménesguen, Y. Souchon, C. Étrillard, A. Levain, F. Moatar, A. Pannard, P. Souchu, A. Lefebvre, G. Pinay, Eutrophication: A new wine in an old bottle?, *Science of the Total Environment*. 651 (2019) 1–11. <https://doi.org/10.1016/J.SCITOTENV.2018.09.139>.
- [9] M.T. O'Hare, A. Baattrup-Pedersen, I. Baumgarte, A. Freeman, I.D.M. Gunn, A.N. Lázár, R. Sinclair, A.J. Wade, M.J. Bowes, Responses of aquatic plants to eutrophication in rivers: A revised conceptual model, *Front Plant Sci*. 9 (2018) 451. <https://doi.org/10.3389/FPLS.2018.00451/BIBTEX>.
- [10] Y. Wang, Y.F. Chen, W.H. Wu, Potassium and phosphorus transport and signaling in plants, *J Integr Plant Biol*. 63 (2021) 34–52. <https://doi.org/10.1111/JIPB.13053>.
- [11] W.K. Dodds, W.W. Bouska, J.L. Eitzmann, T.J. Pilger, K.L. Pitts, A.J. Riley, J.T. Schloesser, D.J. Thornbrugh, Eutrophication of U. S. freshwaters: Analysis of potential economic damages, *Environ Sci Technol*. 43 (2009) 12–19. <https://doi.org/10.1021/ES801217Q>.
- [12] N.J. Anderson, H. Bennion, A.F. Lotter, Lake eutrophication and its implications for organic carbon sequestration in Europe, *Glob Chang Biol*. 20 (2014) 2741–2751. <https://doi.org/10.1111/GCB.12584>.
- [13] P. Singh, V.K. Singh, R. Singh, A. Borthakur, S. Madhav, A. Ahamad, A. Kumar, D.B. Pal, D. Tiwary, P.K. Mishra, Bioremediation: a sustainable approach for management of environmental contaminants, in: P. Singh, A. Kumar, A. Borthakur (Eds.), *Abatement of Environmental Pollutants: Trends and Strategies*, Elsevier, Amsterdam, Netherlands, 2020: pp. 1–23. <https://doi.org/10.1016/B978-0-12-818095-2.00001-1>.
- [14] V. Ancona, I. Rascio, G. Aimola, A.B. Caracciolo, P. Grenni, V.F. Uricchio, D. Borello, Plant-assisted bioremediation: Soil recovery and energy from biomass, *Assisted Phytoremediation*. (2022) 25–48. <https://doi.org/10.1016/B978-0-12-822893-7.00012-4>.
- [15] M. Salari, V. Rahmanian, S.A. Hashemi, W.H. Chiang, C.W. Lai, S.M. Mousavi, A. Gholami, Bioremediation Treatment of Polyaromatic Hydrocarbons for Environmental Sustainability, *Water (Basel)*. 14 (2022) 3980. <https://doi.org/10.3390/W14233980>.
- [16] Q. Xu, B. Wu, X. Chai, In Situ Remediation Technology for Heavy Metal Contaminated Sediment: A Review, *International Journal of Environmental Research and Public Health* 2022, Vol. 19, Page 16767. 19 (2022) 16767. <https://doi.org/10.3390/IJERPH192416767>.
- [17] A. Gonzalez-Martinez, P.S. Goh, W.J. Lau, A. Fauzi Ismail, Z. Samawati, Y.Y. Liang, D. Kanakaraju, Microalgae-Enabled Wastewater Treatment: A Sustainable Strategy for Bioremediation of Pesticides, *Water (Basel)*. 15 (2022) 70. <https://doi.org/10.3390/W15010070>.
- [18] A. Sinharoy, B. Jyoti Deka, L.E. Amabilis-Sosa, E.I. Valenzuela, J.A. Quezada-Renteria, A.M. Pat-Espadas, Biochar-Assisted Bioengineered Strategies for Metal Removal: Mechanisms, Key Considerations, and Perspectives for the Treatment of Solid and Liquid Matrixes, *Sustainability*. 14 (2022) 17049. <https://doi.org/10.3390/SU142417049>.
- [19] M. Rajkumar, S. Sandhya, M.N.V. Prasad, H. Freitas, Perspectives of plant-associated microbes in heavy metal phytoremediation, *Biotechnol Adv*. 30 (2012) 1562–1574. <https://doi.org/10.1016/J.BIOTECHADV.2012.04.011>.
- [20] M. Tekere, A.Y. Mswaka, R. Zvauya, J.S. Read, Growth, dye degradation and ligninolytic activity studies on Zimbabwean white rot fungi, *Enzyme Microb Technol*. 28 (2001) 420–426. [https://doi.org/10.1016/S0141-0229\(00\)00343-4](https://doi.org/10.1016/S0141-0229(00)00343-4).

- [21] T. Lleó, E. Albacete, R. Barrena, X. Font, A. Artola, A. Sánchez, Home and vermicomposting as sustainable options for biowaste management, *J Clean Prod.* 47 (2013) 70–76. <https://doi.org/10.1016/j.jclepro.2012.08.011>.
- [22] W. Xia, Z. Du, Q. Cui, H. Dong, F. Wang, P. He, Y.C. Tang, Biosurfactant produced by novel *Pseudomonas* sp. WJ6 with biodegradation of n-alkanes and polycyclic aromatic hydrocarbons, *J Hazard Mater.* 276 (2014) 489–498. <https://doi.org/10.1016/j.jhazmat.2014.05.062>.
- [23] A.C. Reis, M. Čvančarová, Y. Liu, M. Lenz, T. Hettich, B.A. Kolvenbach, P.F.X. Corvini, O.C. Nunes, Biodegradation of sulfamethoxazole by a bacterial consortium of *Achromobacter denitrificans* PR1 and *Leucobacter* sp. GP, *Appl Microbiol Biotechnol.* 102 (2018) 10299–10314. <https://doi.org/10.1007/S00253-018-9411-9/FIGURES/6>.
- [24] I. Budiman, D. Hermawan, F. Febrianto, G. Pari, Subyakto, Char properties and pollutant adsorption capability of oil palm shell using hydrothermal process, *Biomass Convers Biorefin.* 9 (2019) 681–688. <https://doi.org/10.1007/S13399-019-00394-5/TABLES/5>.
- [25] C.W. Purnomo, D. Castello, L. Fiori, Granular Activated Carbon from Grape Seeds Hydrothermal Char, *Applied Sciences.* 8 (2018) 331. <https://doi.org/10.3390/APP8030331>.
- [26] M. Volpe, L. Fiori, From olive waste to solid biofuel through hydrothermal carbonisation: The role of temperature and solid load on secondary char formation and hydrochar energy properties, *J Anal Appl Pyrolysis.* 124 (2017) 63–72. <https://doi.org/10.1016/j.jaap.2017.02.022>.
- [27] S. Kang, X. Li, J. Fan, J. Chang, Characterization of Hydrochars Produced by Hydrothermal Carbonization of Lignin, Cellulose, d-Xylose, and Wood Meal, *Ind Eng Chem Res.* 51 (2012) 9023–9031. <https://doi.org/10.1021/IE300565D>.
- [28] S. Masoumi, V.B. Borugadda, S. Nanda, A.K. Dalai, Hydrochar: A Review on Its Production Technologies and Applications, *Catalysts.* 11 (2021) 939. <https://doi.org/10.3390/CATAL11080939>.
- [29] A. Jain, R. Balasubramanian, M.P. Srinivasan, Hydrothermal conversion of biomass waste to activated carbon with high porosity: A review, *Chemical Engineering Journal.* 283 (2016) 789–805. <https://doi.org/10.1016/j.cej.2015.08.014>.
- [30] Y. Wang, L. Qiu, M. Zhu, G. Sun, T. Zhang, K. Kang, Comparative Evaluation of Hydrothermal Carbonization and Low Temperature Pyrolysis of *Eucommia ulmoides* Oliver for the Production of Solid Biofuel, *Sci Rep.* 9 (2019) 1–11. <https://doi.org/10.1038/s41598-019-38849-4>.
- [31] R. Ferrentino, F. Merzari, L. Fiori, G. Andreottola, Coupling Hydrothermal Carbonization with Anaerobic Digestion for Sewage Sludge Treatment: Influence of HTC Liquor and Hydrochar on Biomethane Production, *Energies (Basel).* 13 (2020) 6262. <https://doi.org/10.3390/EN13236262>.
- [32] J.G. Ehrenfeld, Plant–Soil Interactions, in: S.A. Levin (Ed.), *Encyclopedia of Biodiversity* (Second Edition), Second Edition, Academic Press, Waltham, 2013: pp. 109–128. <https://doi.org/https://doi.org/10.1016/B978-0-12-384719-5.00179-9>.
- [33] U. Marzocchi, S. Benelli, M. Larsen, M. Bartoli, R.N. Glud, Spatial heterogeneity and short-term oxygen dynamics in the rhizosphere of *Vallisneria spiralis*: Implications for nutrient cycling, *Freshw Biol.* 64 (2019) 532–543. <https://doi.org/10.1111/FWB.13240>.
- [34] W. Armstrong, E.J. Wright, Radial Oxygen Loss from Roots: The Theoretical Basis for the Manipulation of Flux Data Obtained by the Cylindrical Platinum Electrode Technique, *Physiol Plant.* 35 (1975) 21–26. <https://doi.org/10.1111/j.1399-3054.1975.tb03861.x>.
- [35] S. Bai, J. Chen, M. Guo, N. Ren, X. Zhao, Vertical-scale spatial influence of radial oxygen loss on rhizosphere microbial community in constructed wetland, *Environ Int.* 171 (2023) 107690. <https://doi.org/10.1016/j.envint.2022.107690>.
- [36] E. Soana, M. Bartoli, Seasonal regulation of nitrification in a rooted macrophyte (*Vallisneria spiralis* L.) meadow under eutrophic conditions, *Aquat Ecol.* 48 (2014) 11–21. <https://doi.org/10.1007/S10452-013-9462-Z/TABLES/3>.

- [37] E. Soana, M. Naldi, S. Bonaglia, E. Racchetti, G. Castaldelli, V. Brüchert, P. Viaroli, M. Bartoli, Benthic nitrogen metabolism in a macrophyte meadow (*Vallisneria spiralis* L.) under increasing sedimentary organic matter loads, *Biogeochemistry*. 124 (2015) 387–404. <https://doi.org/10.1007/S10533-015-0104-5/TABLES/4>.
- [38] E. Soana, M. Naldi, M. Bartoli, Effects of increasing organic matter loads on pore water features of vegetated (*Vallisneria spiralis* L.) and plant-free sediments, *Ecol Eng*. 47 (2012) 141–145. <https://doi.org/10.1016/J.ECOLENG.2012.06.016>.
- [39] B. Furlotti, G. Rebecchini, *The art of Mantua: power and patronage in the Renaissance*, 1st ed., Getty Publications, Los Angeles, USA, 2008.
- [40] T. Bonomi, D. Sartirana, L. Toscani, G.A. Stefania, C. Zanotti, M. Rotiroti, A. Redaelli, L. Fumagalli, Modeling groundwater/surface-water interactions and their effects on hydraulic barriers, the case of the industrial area of Mantua (Italy), *Acque Sotterranee - Italian Journal of Groundwater*. 11 (2022) 43–55. <https://doi.org/10.7343/AS-2022-569>.
- [41] D. Claeys, C. Lambert, La chute des géants: De la crise à l'émergence, *Acta Europaea Systemica*. 8 (2020) 395–412. <https://doi.org/10.14428/aes.v8i1.56403>.
- [42] T.F. Arnold, Gonzaga Fortifications and the Mantuan Succession Crisis of 1613-1631, *Mediterr Stud (Kirksv Mo)*. 4 (1994) 113–130. <http://www.jstor.org/stable/41166883>.
- [43] T.F. Arnold, Fortifications and the Military Revolution: The Gonzaga Experience, 1530-1630, in: C.J. Rogers (Ed.), *The Military Revolution Debate*, 1st ed., Routledge, New York, USA, 2018: pp. 201–226. <https://doi.org/10.4324/9780429496264-9>.
- [44] A. Garg, N.C. Gupta, A comprehensive study on spatio-temporal distribution, health risk assessment and ozone formation potential of BTEX emissions in ambient air of Delhi, India, *Science of The Total Environment*. 659 (2019) 1090–1099. <https://doi.org/10.1016/J.SCITOTENV.2018.12.426>.
- [45] N. Qin, Y. Zhu, Y. Zhong, J. Tian, J. Li, L. Chen, R. Fan, F. Wei, External Exposure to BTEX, Internal Biomarker Response, and Health Risk Assessment of Nonoccupational Populations near a Coking Plant in Southwest China, *Int J Environ Res Public Health*. 19 (2022) 847. <https://doi.org/10.3390/IJERPH19020847/S1>.
- [46] G. Derringer, R. Suich, Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology*. 12 (1980) 214–219. <https://doi.org/10.1080/00224065.1980.11980968>.
- [47] B. Magnusson, U. Örnemark, eds., *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topic*, 2nd ed., 2014. [https://doi.org/10.1016/S0014-2999\(99\)00500-2](https://doi.org/10.1016/S0014-2999(99)00500-2).
- [48] R. Bro, A.K. Smilde, Principal component analysis, *Analytical Methods*. 6 (2014) 2812–2831. <https://doi.org/10.1039/c3ay41907j>.
- [49] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems*. 2 (1987) 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- [50] M. Li Vigni, C. Durante, M. Cocchi, *Exploratory Data Analysis*, in: *Data Handling in Science and Technology*, Elsevier Ltd, 2013: pp. 55–126. <https://doi.org/10.1016/B978-0-444-59528-7.00003-X>.
- [51] F.W. Young, Y. Takane, J. de Leeuw, The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling features, *Psychometrika*. 43 (1978) 279–281. <https://doi.org/10.1007/BF02293871/METRICAL>.
- [52] J. Pawliszyn, *Solid Phase Microextraction: Theory and Practice*, Wiley-VCH, New York, 1997. <https://www.wiley.com/en-us/Solid+Phase+Microextraction%3A+Theory+and+Practice-p-9780471190349> (accessed August 25, 2022).
- [53] J. Pawliszyn, *Handbook of Solid Phase Microextraction*, 2012. <https://doi.org/10.1016/B978-0-12-416017-0.00011-5>.

- [54] A. Andersson, Mechanisms for log normal concentration distributions in the environment, *Sci Rep.* 11 (2021) 1–7. <https://doi.org/10.1038/s41598-021-96010-6>.
- [55] H. Rajabi, M. Hadi Mosleh, T. Prakoso, N. Ghaemi, P. Mandal, A. Lea-Langton, M. Sedighi, Competitive adsorption of multicomponent volatile organic compounds on biochar, *Chemosphere.* 283 (2021) 131288. <https://doi.org/10.1016/j.chemosphere.2021.131288>.
- [56] D.T. Gibson, J.R. Koch, R.E. Kallio, Oxidative Degradation of Aromatic Hydrocarbons by Microorganisms. I. Enzymatic Formation of Catechol from Benzene, *Biochemistry.* 7 (1968) 2653–2662. https://doi.org/10.1021/BI00847A031/ASSET/BI00847A031.FP.PNG_V03.
- [57] S.A.B. Weelink, M.H.A. van Eekert, A.J.M. Stams, Degradation of BTEX by anaerobic bacteria: physiology and application, *Reviews in Environmental Science and Bio/Technology* 2010 9:4. 9 (2010) 359–385. <https://doi.org/10.1007/S11157-010-9219-2>.
- [58] M. Yoshikawa, M. Zhang, K. Toyota, Biodegradation of Volatile Organic Compounds and Their Effects on Biodegradability under Co-Existing Conditions, *Microbes Environ.* 32 (2017) 188–200. <https://doi.org/10.1264/j SME2.ME16188>.

Chapter 6 | Cocystals for agriculture and food packaging applications: a chemometric approach for their discovery

6.1. Introduction

6.1.1. Cocrystallization as a crystal engineering strategy

The formulation of drugs and pharmaceuticals is a process that requires a deep evaluation of the characteristics of the active pharmaceutical ingredient (API). Solubility, dissolution rate, absorption rate, and bioavailability are just a few of the matters of interest in the formulation process, especially in the case of poorly water-soluble APIs [1–3]. In this context, the main drawback is related to the high dose of API required to achieve the desired therapeutic effect [2].

The aforementioned properties can be adjusted with proper formulative practices including salification, formation of complexes, and encapsulation. Obviously, all these approaches might impact the overall stability of the formulation [4].

Another solution is represented by the manipulation of the crystal structure of the API, so that all the properties that depend on reticular energy can be modified accordingly. This is the core idea behind a branch of solid-state chemistry called crystal engineering, that can be defined as the rational design of molecular solids [1,5,6]. One of the most relevant examples in which the formulation of a drug was aided by crystal engineering is represented by itraconazole (Figure 6.1), an antimycotic. Being this API practically insoluble in water in its crystalline form, to increase its bioavailability it is marketed in its amorphous form [2,4].

This represents one of the very few examples of API commercialized in amorphous form, as the high content of internal energy causes the conversion of the amorphous form into the crystalline one [2,7]. Another strategy in crystal engineering includes the conversion of a crystalline form of an API into another one, i.e., a polymorph characterized by an enhanced solubility. This strategy can be very inconvenient as the process that leads to a polymorph instead of another requires fine tuning and it is not always predictable [2,8].

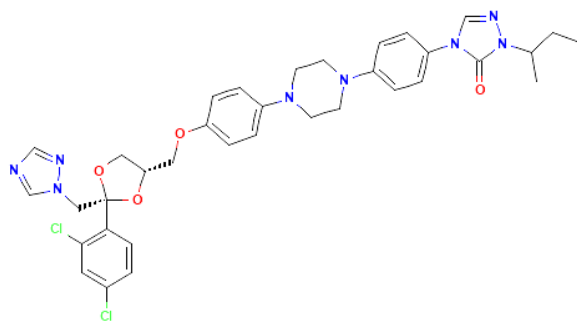


Figure 6.1. Molecular structure of itraconazole^{lxviii}.

In the last decades, the cocrystallization strategy has become more and more popular in the field of pharmaceuticals to enhance the solubility of APIs [2,4]. A cocrystal is a crystalline solid in which the API interacts non-covalently with one or more coformers (i.e., chemical entities such as ions or molecules) in a well-defined stoichiometric ratio [9,10]. The interactions involved in cocrystals are non-covalent, and include electrostatic interactions, hydrogen-bonding and halogen-bonding, as well as interactions involving π electrons, such as π - π stacking [9,10]. Within the frame of reference of itraconazole, it has been reported that its cocrystals with dicarboxylic acids, i.e., L-malic acid and L-tartaric acid, showed a dissolution profile similar to the one of the amorphous API, reaching concentration up to 20-folds higher than the crystalline form of itraconazole [4].

The cocrystallization strategy opens the possibility of synthesizing a variety of crystalline materials of the same API by properly selecting a suitable coformer (in the case of binary cocrystals), characterized by unique properties depending on the interaction energy between the molecular partners [11–13]. Therefore, *ad hoc* cocrystals can be produced depending on the intended application. In addition, the synthesis of such crystalline solids does not involve any cleavage or formation of covalent bonds and, in the case of mechanochemical synthesis, can be solventless [14]. A noteworthy implication of cocrystallization as a crystal engineering strategy is that the outcome of the process is a solid, allowing for the conversion of liquid APIs into crystalline solids. This aspect might seem negligible, but it has strong technological implication, as solids are easier to transport and store, and are safer to handle than liquids.

Ever since cocrystals gained popularity in pharmaceuticals, their use has been exploited also in agrochemistry, nutraceuticals, and cosmetics [1,11,15–17].

^{lxviii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/55283> (accessed 02/11/2022)

6.1.2. Essential oils and Generally Recognized As Safe substances

According to ISO 9235:2021^{lxix}, essential oils (EOs) are “*products obtained from a natural raw material of plant origin, by steam distillation, by mechanical processes from the epicarp of citrus fruits or by dry distillation, after separation of the aqueous phase -if any- by physical processes*”. EOs are, therefore, mixtures of compounds that are produced by the plants themselves as secondary metabolites to address biological functions such as reproduction and defense [18–21].

Chemically, EOs are complex mixtures whose 20–70% is typically represented by up to three main compounds responsible for the biological activity [20,22]. The remaining portion contains less abundant components. It has to be stated that the chemical composition of a given EO is not constant, but it is subjected to variation depending on the maturation stage, stress conditions, and the extraction method that it is utilized to recover the EO itself [22]. Their composition can be investigated with appropriate analytical techniques, often hyphenated, such as GC–MS and LC–MS.

The constituents of EOs can be classified depending on their chemical nature. The main classes are [18,22]:

- **Terpenes and their derivatives:** with more than 80000 known compounds, terpenes and their derivatives represent the most numerous, chemically, and functionally diverse class of compounds that are part of the metabolome of living organisms [21,23,24]. Chemically, terpenes are hydrocarbons constituted by the repetition of a core unit called isoprene (Figure 6.2), a diene with 5 carbon atoms.

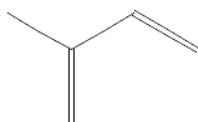


Figure 6.2. Molecular structure of isoprene^{lxx}.

Terpenes found in EOs are normally characterized by a low molecular weight and are constituted by two or three isoprene units bonded together to yield monoterpenes and sesquiterpenes, respectively [22]. Such substances can undergo to biochemical reactions so that the product is a

^{lxix} <https://www.iso.org/obp/ui/#iso:std:iso:9235:ed-3:v1:en> (accessed 22/11/2022)

^{lxx} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/6557> (accessed 22/11/2022)

positional isomer with respect to methyl groups or an oxygenated derivative. These compounds are called terpenoids and are almost always found in EOs as well. Figure 6.3 shows the molecular structure of examples of terpenes and terpenoids.

- **Phenylpropenes:** these compounds are produced in the first step of the phenylpropenes biosynthesis [22,25]. Compounds belonging to this class have in common a phenolic ring and the vast majority of them have the aromatic ring substituted with a chain with 3 carbon atoms. Figure 6.3 shows the molecular structure of examples of phenylpropenes.
- **Flavonoids:** flavonoids belong to the polyphenol chemical class [26]. The core unit of flavonoids is constituted by: i) a benzene ring that shares 2 contiguous carbon atoms with an oxygenated 6-membered heterocyclic ring and ii) a benzene ring σ -bonded with the heterocycle. Figure 6.3 shows the molecular structure of examples of flavonoids.

EOs and their pure components are highly valuable chemicals for the industry. As it was discussed earlier in **Chapter 3**, terpenes and terpenoids are widely used in the manufacturing of perfumes, fragrances, and personal care products [18,27,28]. In addition, few phenylpropenes are important building blocks for the pharmaceutical industry [29,30], whereas flavonoids, having displayed many beneficial effects on the human health, found important application in the field of nutraceuticals [26]. Besides their application in the manufacturing industry, EOs and their active components are of biological and technological relevance. In fact, thanks to antimicrobial and antioxidant activity [18,25,31] they can find application in the manufacturing of active food packaging, as well as in agriculture for controlling pests [11,19,32]. The possibility of applying such substances also in the food industry and agriculture relies on the fact that EOs and many of their active components are included in the Generally Recognized As Safe (GRAS) list under the regulation of the Food and Drug Administration^{lxxi} (FDA). In Europe, these chemicals are also regulated according to the Regulation EC 1334/2008^{lxxii}. As already stated in **Chapter 3**, these substances can have adverse effects on human health and on the environment [27,28]. Therefore, their dosage must attain the limits imposed by the regulating authorities and a good balance between beneficial effects and toxicity has to be found [22]. In addition, the products formulated with

^{lxxi} <https://www.fda.gov/food/food-ingredients-packaging/generally-recognized-safe-gras> (accessed 23/11/2022)

^{lxxii} <https://eur-lex.europa.eu/eli/reg/2008/1334/oj> (accessed 23/11/2022)

EOs and their active components need to be handled with care, to avoid the dispersion of such substances in the environment.

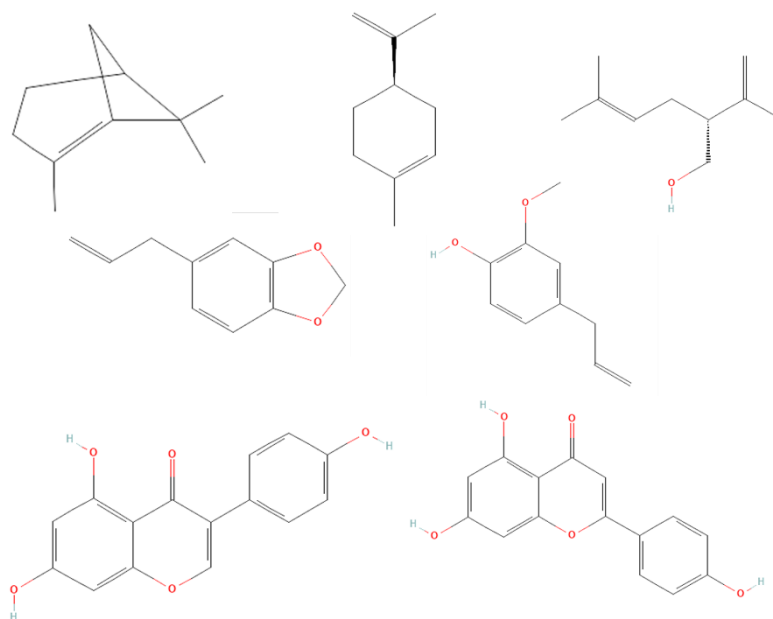


Figure 6.3. Examples of terpenes and terpenoids (top row, from left to right: (-)-pinene^{lxxiii}, (+)-limonene^{lxxiv}, and (-)-lavandulol^{lxxv}), phenylpropenes (middle row, from left to right: safrole^{lxxvi} and eugenol^{lxxvii}), and flavonoids (bottom row, from left to right: genistein^{lxxviii} and apigenin^{lxxix}).

Despite EOs and their active components have attractive properties, their applicability for the manufacturing of active packaging or their use in agrochemistry suffers from technological limitations, mainly related to their physicochemical properties [11,19,32–34]. In fact, most of these compounds are volatile liquids or low-melting solids at ambient conditions and lacks thermal and light stability. Other than making their transportation and storage difficult, volatility can represent a health risk for operators involved in their handling. Within this frame of reference, cocrystallization represents a valuable strategy to turn these chemicals from volatile liquids to more stable solids, making them more appealing for technological applications. Cocrystals of terpenoids have been proposed as ingredients for active packaging materials able to extend the shelf-life

^{lxxiii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/15837102> (accessed 23/11/2022)

^{lxxiv} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/440917> (accessed 23/11/2022)

^{lxxv} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/5280961> (accessed 23/11/2022)

^{lxxvi} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/5144> (accessed 23/11/2022)

^{lxxvii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/3314> (accessed 23/11/2022)

^{lxxviii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/5464156> (accessed 23/11/2022)

^{lxxix} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/5280443> (accessed 23/11/2022)

of foodstuff, as well as greener alternatives to agrochemicals in the control of pests [11,32,35].

6.1.3. Computational approaches to cocrystal discovery

As it was mentioned earlier, cocrystallization offers great advantages allowing to design a crystalline material in which the properties of the API have been improved by properly cocrystallizing it with a suitable partner molecule^{lxxx}. Obviously, in order to achieve cocrystallization, a certain degree of compatibility and complementarity between the synthons involved in the synthesis of a cocrystal has to be present [3,36–38]. Therefore, both the conformer and the API need to be carefully selected.

This fact has strong implications in the framework of cocrystal design and cocrystal discovery for two main reasons:

- It is not guaranteed that the selected partner molecules will yield a cocrystal as a product.
- It is not guaranteed that the product of the supramolecular synthesis will have the forecasted properties.

In this context, it is easy to understand that the selection of suitable partner molecules, the synthesis attempt and the characterization of the products require a great effort, representing an out-and-out bottleneck, that can grow overwhelmingly, especially if many combinations between partner molecules needs to be tested is large.

Computational methods relying on the calculation of useful parameters to estimate cocrystallization propensity have been proposed. Some examples include the estimation of solubility [39], lattice and interaction energies [37,38,40], as well as the degree of complementarity between the synthons [41]. Despite the effort that has been put in the development of such methodologies the lack of reliable and/or easy to apply methodologies is still present.

Very recently attention has been paid to the discovery of cocrystals with approaches based on Machine Learning using different supervised approaches to predict the formation of new cocrystals. Jan-Joris Devogelaer and coworkers developed a model based on Artificial Neural Networks [42]: the model was calibrated by taking the input data directly from the Cambridge Structural Database (CSD) [43]. Aikaterini Vriza and coworkers evaluated the capability of different

^{lxxx} From now on, with the term cocrystal the author is going to refer to binary cocrystals, i.e., resulting from the combination of two partner molecules, as they are the matter of this study.

supervised methods such as Support Vector Machine and k -Nearest Neighbors with a class-modelling approach (see **Chapter 2**) [44]. Both these approaches utilized for the calibration only cocrystals known in the literature, whereas attempts of failed cocrystallization experiments (i.e., the two synthons yielded a eutectic or a binary mixture) were overlooked.

The approach proposed by Dingyan Wang and coworkers involved also failed cocrystallization data [45]. In the study, a Random Forest classifier based on consensus was trained by including in the calibration set random combinations of synthons as failed cocrystallization data. The underlying assumption was that randomly selected partner molecules were unlikely to yield a cocrystal as the product of their combination. Other research groups retrieved both successful and failed cocrystallization attempts from the literature. For instance, Maciej Przybyłek and coworkers developed supervised models based on regression to predict the formation of cocrystals based on phenolic and dicarboxylic acids: in that study both successful and unsuccessful cocrystallization data were retrieved from the literature [46,47]. The approach proposed by Jerome G. P. Wicker and coworkers involved a Support Vector Machine model calibrated with successful and failed cocrystallization experiments of variously substituted benzoic acids and benzamides [12]. One of the last contributions in the field of Machine Learning-assisted cocrystallization was the one by Medard Edmund Mswahili and coworkers, who calibrated an Artificial Neural Networks classifier on successful and failed cocrystallization attempts retrieved from the literature [48].

6.1.3.1. Quantitative Structure-Property Relationship

Besides the source of cocrystallization data source and the inclusion or not of failed cocrystallization attempts, all the aforementioned studies utilize a pool of molecular and topological features as the variables to describe the molecules involved in the study. Therefore, the aim is to find a relationship that links the property of interest (cocrystallization propensity in this case) with the molecular features of the involved synthons.

The last statement represents the core idea behind the so-called Quantitative Structure-Property Relationship (QSPR) [49,50]. In a figurative way, QSPR is the bridge between the molecular features and a property that can be measured or observed. This approach is widely utilized in drug development, as it allows to design a molecule with the desired activity once sufficient data about other molecules have been collected. In the specific case in which the desired property is

the activity (e.g., IC₅₀, EC₅₀, MIC, etc.), the approach is abbreviated as QSAR, which stands for Quantitative Structure-Activity Relationship [51–53].

6.1.4. Cocrystal-based active food packaging

As previously stated, cocrystals can play a key role also in field of active packaging. The PAC (*Packaging Attivo Cristallino*) project, funded in 2018 by the ex-Italian Ministry of Agricultural, Food, Forestry Policies and Tourism (Principal Investigator: Prof. Alessia Bacchi, University of Parma, with the cooperation of the units of Crystal Engineering and Analytical Chemistry of the same University), aimed at developing cocrystals containing active components of EOs (terpenoids) to be incorporated in active packaging materials for the prevention of food waste.

In the framework of the PAC project, a research study carried out in cooperation with the section of Infectious Diseases of Animals (University of Parma) was focused on the investigation of the antimicrobial activity of active components of EOs on four bacterial strains commonly involved in food deterioration, namely *Escherichia coli*, *Salmonella Typhimurium*, *Staphylococcus aureus* and a methicillin resistant *S. aureus*. Among the tested compounds, thymol, carvacrol, and cinnamaldehyde showed the most promising antimicrobial activity. The next step involved the evaluation of the potential synergistic effect of these compounds in the simultaneous inhibition of the investigated organisms. A Quality by Design approach was followed, allowing the systematic evaluation of the effects of involved factors (i.e., the concentrations of the active compounds) towards the inhibition of the investigated strains, and the identification of experimental conditions allowing robust results, with a success rate of 95%.

Thereafter, three cocrystals were synthesized by utilizing GRAS molecules as cofomers. The products, i.e., isonicotinamide/carvacrol, hexamethylenetetramine/thymol, and, 4-hydroxybenzoic acid/cinnamaldehyde, were thoroughly characterized in terms of crystallographic properties, thermal behavior, and controlled release of the active compound. Then, the developed cocrystals were utilized to produce a multilayer material: the support material was polyethylene, and it was coated with a 7.02 (± 1.00) μm ($n = 5$) layer of chitosan with the proper amount of cocrystal anchored on it. The developed material showed a promising antimicrobial activity *in vitro* towards the investigated organisms, producing an antimicrobial effect both in direct and indirect contact with the microbial culture. Additionally, the proposed cocrystal based-active food packaging material was able to exert control on the total mesophilic charge of grapes samples, with an inhibition of 86% after one week of incubation with respect to the control.

The study presented in this **Chapter** emerged as a collateral consequence of other activities involved in the cocrystallization of active ingredients contained in EOs, such as the one involved in the framework of the PAC project [32], along with others involved in the development of more sustainable agrochemicals [11,35]. In this context, additional studies were carried out on a pool of both successful and failed cocrystallization data collected both experimentally and from the literature. The novelty of this research activity relies on the fact that the use of chemometrics to aid the discovery of cocrystals had not been explored just yet, therefore the aim of the study was to train a discriminant model based on Partial Least Squares to distinguish binary cocrystals (CC) from binary mixtures (BM) based on the molecular features of the two synthons. The workflow followed the philosophy of QSPR, with the difference that the modelled property is not quantitative, but categorial instead. To ensure the reliability of the model, it was subjected to multiple validation steps, including an external test set.

6.2. Materials and methods

6.2.1. Data collection

6.2.1.1. Class assignation

The molecules involved in this study are all included in the GRAS list approved by the FDA^{lxxxi}. Binary combinations of such molecules were selected from the list and either assigned to the CC or BM class, i.e., cocrystal or binary mixtures, respectively. Cocrystals already known in the scientific literature were retrieved from the CSD [43] and undoubtedly assigned to the CC class. For all the remaining binary combinations, crystallization experiments were attempted, and the products were carefully characterized.

All the syntheses were performed mechanochemically by mixing equimolar amounts of each supramolecular synthon and by grinding them manually into a mortar for 10–15 min. No solvents were added. The products were collected and kept in closed vials.

Powder X-Ray Diffraction (PXRD) was utilized for the characterization. The samples were placed on a glass sample holder and exposed to the Cu K α radiation in θ - θ Bragg-Brentano configuration. The vast majority of samples were analyzed on a Rigaku Smartlab XE instrument (Tokyo, Japan) equipped with a 2D Hypix 3000

^{lxxxi} <https://www.fda.gov/food/food-ingredients-packaging/generally-recognized-safe-gras> (accessed 23/11/2022)

solid state detector. A 5° soller slit, beam stopper and an anti-scatter air component were utilized to acquire better signals and low angles. The scan program was $1.5^\circ \leq 2\theta \leq 50^\circ$ at 10°/min scan rate. Occasionally, samples were characterized with an ARL X'TRA diffractometer (Thermo Fisher Scientific, Massachusetts, USA) with the following programs: $3^\circ \leq 2\theta \leq 30^\circ$ at 5°/min scan rate or $3^\circ \leq 2\theta \leq 40^\circ$ at 0.3°/min scan rate.

Both the products and the supramolecular synthons were submitted to PXRD analysis. Each product was classified either in the CC class or the BM class according to its PXRD trace:

- The presence of new diffraction peaks not ascribable to the presence of unreacted synthons was considered as evidence that a cocrystal was obtained and the product was classified as a CC.
- Otherwise, the product was classified as a BM.

Polymorphic transitions were excluded by comparing the PXRD trace of the synthons with the calculated traces of their known crystalline forms.

Extensive details about the characterization of the cocrystals and binary mixtures involved in this study, as well as the references of the cocrystals retrieved from the literature are reported in the original publication and in its Supporting Material [54].

6.2.1.2. *Molecular modelling*

A set of 31 molecular descriptors (Table 6.1) was calculated for each molecule present in the products.

The 3D structures of the molecules were retrieved from the PubChem database and were utilized in SDF format to calculate molecular weight, number of atoms, number of bonds, and the number of hydrogen bond donor and acceptor sites at pH = 7. Such structures were also utilized to calculate number of rotatable bonds, number of rings, number of hydrophobic centers, logP, molecular volume, and the total molecular dipole moment (calculated by using the principal axes of the molecular graph). The protonation state of the molecule was considered in the calculation of these descriptors.

The molecular volume was defined as the volume enclosed in a water-accessible surface at the frontier of which a water probe experiences a repulsive interaction energy of 0.20 kcal/mol. The strain energy of the molecule was calculated by considering the atomic charges without carrying out any geometry optimization of the structure. SASA was calculated automatically with a dot density of 4.

An *in vacuo* Density Functional Theory optimization was carried out on all the molecules by using the hybrid functional B3LYP and the People double-z basis set 6-31+g(d,p). After optimization, the average isotropic polarizability, the polarizability anisotropy, and the Molecular Electrostatic Potential (MEP) were evaluated. Post-processing on the MEP data was carried out in order to extract the minimum and the maximum. The map cube was sampled along the three directions with a sampling density of 6 points/Bohr and only the MEP values corresponding to regions of electron density 0.002 ± 0.001 a.u. were taken into consideration. As the procedure produced a large number of candidate points, a three-step filtering was applied sequentially:

1. Step 1: a candidate point was considered a local minimum (maximum) if at least 4 out of 6 nearest neighbors had a MEP value greater (lower) or equal than the one of the candidate point.
2. Step 2: the ratio between the MEP value of the local minima (maxima) and the global minimum (maximum) MEP value were calculated. The points for which such ratio was below a threshold of 0.1 were discarded. This allowed for the identification of the portion of the molecules involved in non-covalent interactions in the formation of a potential cocrystal. This means that in those regions, a multitude of candidate points could be very close to each other.
3. Step 3: the last step is dedicated to univocally identify the minimum (maximum) MEP value. The candidate points that were closer than 1.0 Bohr to each other were iteratively merged remaining candidate points were merged with an iterative procedure and only the point corresponding to the lower (higher) MEP value were finally retained.

Extensive details about the calculation of the molecular descriptors and about the theoretical background of most of them can be found in the original publication and in its Supporting Material [54].

The molecular volume was defined as the volume enclosed in a water-accessible surface at the frontier of which a water probe experiences a repulsive interaction energy of 0.20 kcal/mol. The strain energy of the molecule was calculated by considering the atomic charges without carrying out any geometry optimization of the structure. SASA was calculated automatically with a dot density of 4.

17	Isotropic polarizability (α_{iso})	Physical property	3D	a.u. ^d
18	Anisotropic polarizability ($\Delta\alpha$)	Physical property	3D	a.u.
19	Maximum of the molecular electrostatic potential surface	Physical property	3D	kJ/mol
20	Minimum of the molecular electrostatic potential surface	Physical property	3D	kJ/mol
21	Hall & Kier alpha (α_{HK})	Topological	2D	d.q.
22	Topological polar surface area (TPSA)	Topological	2D	Å ²
23	Hall & Kier alpha-corrected first-order shape index ($^1\kappa_d$)	Topological	2D	d.q.
24	Hall & Kier zeroth-order connectivity index ($^0\chi$)	Topological	2D	d.q.
25	Hall & Kier RDKit alternative valence connectivity index ($^0\chi^m$)	Topological	2D	d.q.
26	Hall & Kier zeroth-order valence connectivity index ($^0\chi^v$)	Topological	2D	d.q.
27	Number of heteroatoms	Constitutional	1D	d.q.
28	Number of valence electrons	Constitutional	1D	d.q.
29	Labute' s approximate surface area (LabuteASA)	Surface area-based	2D	Å ²
30	PEOE_VSA	Surface area-based	2D	Å ²
31	SMR_VSA	Surface area-based	2D	Å ²

^a identification numbers. ^b dimensionless quantity. ^c debye. ^d atomic units.

Table 6.1 Molecular descriptors involved in this study.

Id. ^a	Molecular descriptor	Typology	Dimensionality	Unit
1	Molecular weight	Constitutional	1D	g/mol
2	Number of atoms	Constitutional	1D	d.q. ^b
3	Number of bonds	Constitutional	1D	d.q.
4	Number of rotatable bonds	Constitutional	1D	d.q.
5	Number of rings	Constitutional	1D	d.q.
6	Number of hydrogen bond donor sites	Constitutional	1D	d.q.
7	Number of hydrogen bond acceptor sites	Constitutional	1D	d.q.
8	Hydrophobicity	Physical property	1D	d.q.
9	Molecular volume	Geometric	3D	Å ³
10	Logarithm of octanol/water partition coefficient (logP)	Physical property	2D	d.q.
11	Component of the dipole along the X-axis	Physical property	3D	D ^c
12	Component of the dipole along the Y-axis	Physical property	3D	D
13	Component of the dipole along the Z-axis	Physical property	3D	D
14	Total dipole	Physical property	3D	D
15	Energy	Physical property	3D	kcal/mol
16	Solvent accessible surface area (SASA)	Geometric	3D	Å ²

An *in vacuo* Density Functional Theory optimization was carried out on all the molecules by using the hybrid functional B3LYP and the People double-z basis set 6-31+g(d,p). After optimization, the average isotropic polarizability, the polarizability anisotropy, and the Molecular Electrostatic Potential (MEP) were evaluated. Post-processing on the MEP data was carried out in order to extract the minimum and the maximum. The map cube was sampled along the three directions with a sampling density of 6 points/Bohr and only the MEP values corresponding to regions of electron density 0.002 ± 0.001 a.u. were taken into consideration. As the procedure produced a large number of candidate points, a three-step filtering was applied sequentially:

4. Step 1: a candidate point was considered a local minimum (maximum) if at least 4 out of 6 nearest neighbors had a MEP value greater (lower) or equal than the one of the candidate point.
5. Step 2: the ratio between the MEP value of the local minima (maxima) and the global minimum (maximum) MEP value were calculated. The points for which such ratio was below a threshold of 0.1 were discarded. This allowed for the identification of the portion of the molecules involved in non-covalent interactions in the formation of a potential cocrystal. This means that in those regions, a multitude of candidate points could be very close to each other.
6. Step 3: the last step is dedicated to univocally identify the minimum (maximum) MEP value. The candidate points that were closer than 1.0 Bohr to each other were iteratively merged remaining candidate points were merged with an iterative procedure and only the point corresponding to the lower (higher) MEP value were finally retained.

Extensive details about the calculation of the molecular descriptors and about the theoretical background of most of them can be found in the original publication and in its Supporting Material [54].

6.2.2. Data cleaning and validation strategy

In the matrices **X** and **Y**, each object is represented by a couple of partner molecules, i.e., A and B. As for the **X** matrix, the elements of the Manhattan distance [55] between A and B were calculated for each object and utilized as variables, i.e., the absolute value of the difference between a given molecular descriptor D_j of the two partner molecules (i.e., A and B) was calculated (Equation 6.1) and utilized as variable.

$$\Delta D_j = |D_j^A - D_j^B|$$

Equation 6.1

Therefore, each of the $I = 181$ object of the data matrix \mathbf{X} is defined by $J = 31$ variables and the dimensions of \mathbf{X} are 181×31 . As for the \mathbf{Y} matrix, the classes are codified as dummy variables (0: not belonging to the class; 1: belonging to the class) in the response matrix \mathbf{Y} (181×2).

Three validation strategies were adopted for the supervised modelling:

- **Cross-validation:** a leave-more-out cross validation approach was utilized in the model optimization stage. A Venetian blinds cancellation scheme with $C = 10$ cancellation groups and a blind thickness of 1 was adopted. Nearly 10% of the data was left out at each iteration.
- **Single-evaluation test set:** a subset of \mathbf{X} , and the corresponding responses in \mathbf{Y} , was extracted. This subset contained 30% of the original data and it was utilized to validate the model at the very end of the model development stage. Therefore, the remaining 70% of data were utilized as the calibration set to build the model. The selection of the sample to be sent to the single-evaluation test set was carried out with the Duplex algorithm [56] by balancing the classes: this means that the class distribution in the single-evaluation test set is the same as in the calibration set.
- **External test set:** a further validation step was carried out by involving an external test set. This dataset accounted for 58 objects, corresponding both to new cocrystallization trials attempted with a mechanochemical approach and to cocrystals retrieved from the CSD. To capture extra variability, in this dataset were also included objects that do not involve at least one of the two partner molecules already present in the calibration set.

The entries *per set* are reported in Table 6.2. The full list of the objects included in each set is available in Table 6A.1–6A.3.

Table 6.2. Number of objects for each class for the calibration set, the single-evaluation test set and external test set. The last row contains the total number of objects for each set.

	Calibration set	Single-evaluation test set	External test set
CC	71	30	31
BM	56	24	27
<i>I</i>	127	54	58

6.2.3. Unsupervised modelling

Exploratory Data Analysis (EDA) was carried out on the **X** block of the calibration set to evaluate the underlying potential data structures present in the data. Principal Component Analysis (PCA) [57–59] was the chosen technique for this purpose (see **Chapter 2**, Section 2.1.1.1). Before the model computation, the data were preprocessed by mean centering the columns and by scaling to unit variance.

6.2.4. Supervised modelling

Supervised data analysis was carried out in the context of discriminant modelling. For this purpose, the technique of choice was Partial Least Squares–Discriminant Analysis (PLS–DA; see **Chapter 2**, Section 2.1.1.2) [60–65]. Before the model computation, the **X** matrix was pretreated by carrying out mean centering on the columns and by scaling to unit variance. The **Y** matrix was mean centered on the columns.

Model optimization was carried out by evaluating the results deriving from cross-validation. The number of Latent Variables (LVs) to be retained was the one able to provide the maximum Non-Error Rate (*NER*, Equation 2.8). Model diagnostics was carried out by inspecting the residuals and the leverage of the samples. The functional relationship between the variables in **X** and the responses in **Y** was assessed by evaluating the PLS coefficients as well as the Variable Importance on Projection (VIP score, Equation 2.24) [64,66,67]. A variable was considered important if its VIP score was greater than 1.

The capability of the model in making prediction was expressed in terms of *NER* and Sensitivity (*S_n*, Equation 2.7a) with respect of both classes on the fitted data, on the single-evaluation test set, and on the external test set.

6.2.5. Software

The Cambridge Crystallographic Data Center (CCDC) and ConQuest were utilized to search for the crystal structures of the known cocrystals. Such structures were visualized with Mercury (CCDC, Cambridge, UK).

FLAP [68] was utilized for the calculation of molecular weight, number of atoms, number of bonds, and the number of hydrogen bond donor and acceptor sites. Sybyl (v. 8.1, Tripos International, Missouri, USA) was utilized for the calculation of number of rotatable bonds, number of rings, number of hydrophobic centers, logarithm of octanol/water partition coefficient, molecular volume, total molecular dipole moment, and strain energy. For calculating the latter, the Tripos force field

[69] was utilized, whereas a specific SPL script was used for calculating the molecular volume and dipole moment. PyMol (v. 2.0) [70] was utilized to calculate the Solvent Accessible Surface Area. A custom code under Python v. 3.7 environment and the chemoinformatics toolkit RDKit Q4 2013 [71] were utilized for the computation of the number of heteroatoms and of valence electrons, the connectivity indexes ${}^0\chi$, ${}^0\chi^n$, and ${}^0\chi^v$, the Topological Polar Surface Area as well as α_{HK} and ${}^1\kappa_\alpha$, the average isotropic polarizability, the polarizability anisotropy and the MEP were calculated with Gaussian (v. 16, Pennsylvania, USA). MEP postprocessing was performed with a custom script in Python v. 3.6.1 environment. Data analysis was carried out with PLS_Toolbox (v. 8.7.1, Eigenvector Research Inc., Washington, USA) and custom routines by working under MATLAB environment (v. R2019a, Mathworks, Massachusetts, USA).

6.3. Results and discussion

6.3.1. Data cleaning

The very first step after data were collected was to find a way to effectively organize the data to be analyzed in the subsequent steps. This is undoubtedly a key point, as any manipulation and preprocessing before analysis can severely affect the degree of interpretability of the model.

The objects are the results of combination of a pair of molecules (i.e., A and B), each one described by its own set of descriptors. The first strategy that was attempted was concatenation. This consisted in the combination of the two descriptor matrices (31 columns each) to yield a single matrix with 62 columns: the first 31 columns were the molecular descriptors of A and the remaining 31 were the molecular descriptors of B. This strategy appeared not to be optimal because:

- The order in which the variables were listed is arbitrary. In fact, the attribution of which of the two molecules should act as the API and which one should act just as a cofomer represents a forced *a priori* decision.
- From preliminary evaluations, carried out by inspecting the score space in PCA (not shown for the sake of brevity), it emerged that different order in the concatenation yielded to different distribution of the scores, making not only the results unreliable, but also difficult to interpret.

The data were, therefore, aggregated by following the strategy proposed by Maciej Przybyłek and coworkers [46,47]. The absolute value of the difference between the

molecular descriptors of A and B was calculated (Equation 6.1). The operation produces a new matrix (with 31 columns) whose elements are the terms of the row-wise Manhattan distance [55] between the partner molecules involved in the formation of a cocrystal or a binary mixture. By aggregating the data in this way:

- The result is independent on the order as the operation that was followed is commutative.
- The result is still rich of information to be extracted and it is easily accessible to interpretation, as the variables are related to the degree of dissimilarity between the two partner molecules.

Before proceeding with the unsupervised modelling, a single data subset was extracted to be utilized as a single-evaluation test set after supervised modelling. This operation was carried out by maintaining the same class distribution of the whole dataset so that the bias towards one of the two classes is minimized, as well as overfitting.

In order to effectively interpret and discuss the results both the objects and the variables will be defined as their identification number: variables are reported in underlined text (Table 6.1^{lxxxii}), whereas objects are reported in plain text (Table 6A.1–6A.3).

6.3.2. Unsupervised modelling

EDA was carried out on the calibration set by subjecting it to PCA. This operation was carried out before the supervised modelling to check for potential data structures, as well as correlation between the variables.

Figure 6.4 shows the results of the PCA decomposition of the first four low-order Principal Components (PCs) that account approximately for 67% of the total variance. Most of the CCs are characterized by negative scores on PC 2 and PC 3 and are generally less dispersed than BMs. The loadings are depicted in Figure 6.5.

^{lxxxii} Note that Table 6.1 reports the identification numbers of the molecular descriptors that has been calculated. In the discussion the same numbers were utilized for the variables already subjected to the aggregation (e.g., 1 refers to the dissimilarity in the molecular weight between the two partner molecules and not to the molecular weight itself).

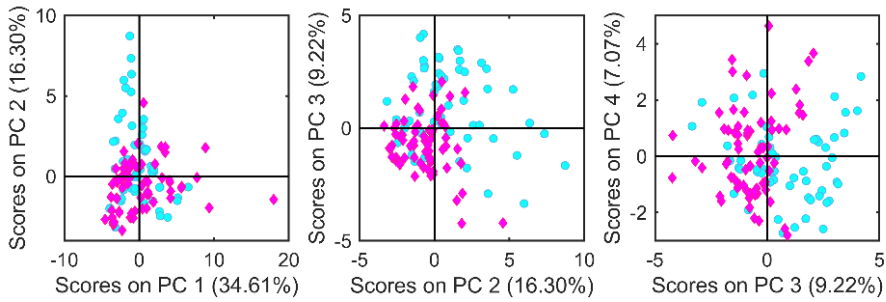


Figure 6.4. Distribution of the objects in the score space of the first four PCs (the fraction of total variance that a given PC explains is reported on the corresponding axis). Left: PC 1 vs. PC 2; center: PC 2 vs. PC 3; right: PC 3 vs. PC 4. Objects are marked according to their class (magenta diamonds: CC; cyan

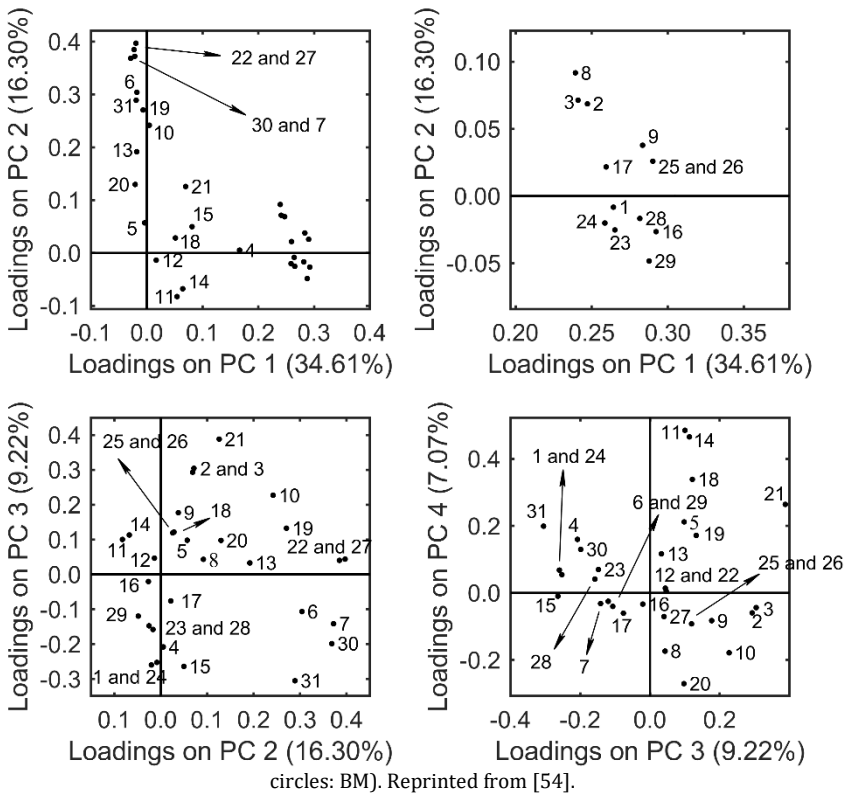


Figure 6.5. Distribution of the variables in the loading space of the first four PCs (the fraction of total variance that a given PC explains is reported on the corresponding axis). Top-left: PC 1 vs. PC 2; top-right: magnification of the PC 1 vs. PC 2 loading plot; bottom-left: PC 2 vs. PC 3; bottom-right: PC 3 vs. PC 4. Reprinted from [54].

Features related to the dissimilarity in molecular dimensions (e.g., [1](#), [2](#), [3](#), [9](#), [16](#), [29](#)) and connectivity ([23](#), [24](#), [25](#), [26](#)) are the major sources of variation within the data, as they were captured by PC 1 (explaining nearly 35% of the total variance). The second most important source of variation (PC 2, capturing nearly 16% of the total variance) was represented by the diversity in electronic properties (e.g., [6](#), [7](#), [19](#), [22](#), [30](#)) as well as in the number of heteroatoms that are present in the molecule ([27](#)). PC 3 and PC 4 captured nearly 9% and 7% of the total variance and accounted for more specific features. Differences in molecular complexity ([21](#)), molecular refractivity ([31](#)) and energy ([15](#)) were the most contributing ones to PC 3, whereas dissimilarity in the component of the dipole along the X-axis ([11](#)), total dipole ([14](#)) and the minimum of the MEP surface ([20](#)) were the most contributing ones on PC 4. By the joint interpretation of the score and the loading plots it can be stated that partner molecules forming CCs are similar to each other in terms of molecular complexity ([21](#)) and logP ([10](#)).

6.3.3. Supervised modelling

Among the different supervised approaches that can be utilized for classification, in the present study the focus was on basic linear modelling instead of nonlinear methodologies such as Artificial Neural Network [72,73]. This choice was operated primarily to keep the computation of the model as simple as possible and to ensure a certain level of interpretability of the outcomes. In addition, the number of entries in the calibration set is insufficient to properly set the hyperparameters of Artificial Neural Network.

PLS-DA was then utilized to seek the correlation between the dissimilarity between partner molecules, i.e., the matrix **X**, and the class membership encoded in the matrix **Y**. The maximum *NER* in cross-validation was achieved with a model with 6 LVs. The model calculated with 6 LVs captured nearly the 72% of the total variance of **X** and the 69% of the total variance of **Y**.

A summary of the performance is reported in Table 6.3 and in Figure 6.6.

The results were promising both in fitting and in cross-validation, with *NER* values similar to each other. In addition, *Sn* values were also satisfactory and were very close to each other, meaning that the model can correctly identify a given object regardless of its class, at least on the fitted data and on the predicted ones in cross-validation.

Table 6.3. Confusion matrix for the fitted data and for the data predicted in cross-validation. The last two lines show the sensitivity and the non-error rate expressed as percentages.

True class	Predicted class			
	Fitting		Cross-validation	
	CC	BM	CC	BM
CC	66	5	61	10
BM	5	51	7	49
<i>Sn</i>	93%	91%	86%	88%
<i>NER</i>	92%		87%	

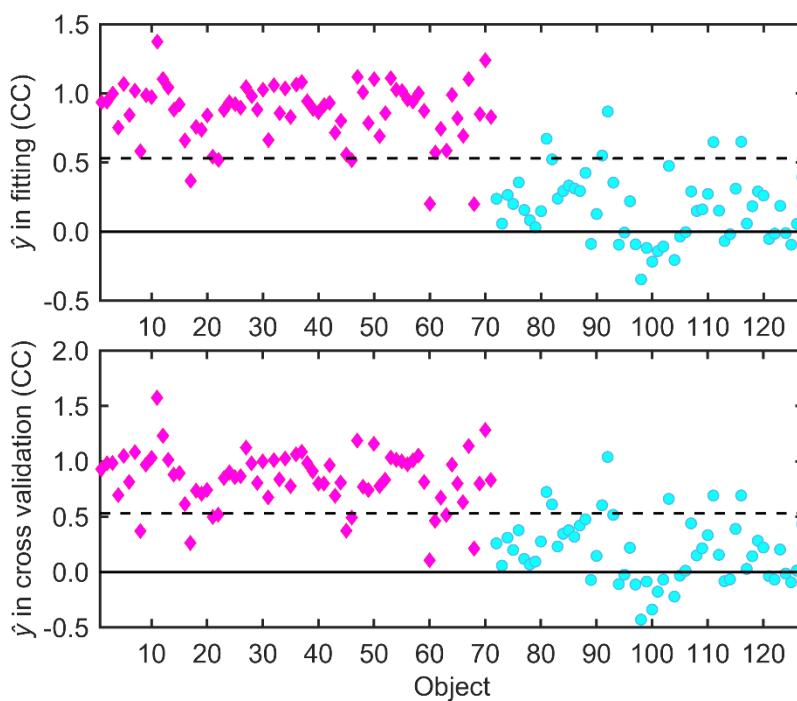


Figure 6.6. Predicted response \hat{y} (with respect to the CC class) for the training set on the fitted data (top) and in cross-validation (bottom). The dashed horizontal line shows the hard classification threshold of 0.53. Samples are marked according to their class (magenta diamonds: cocrystals; cyan circles: binary mixtures). Reprinted from [54].

6.3.3.1. Model diagnostics

The two low-order LVs provided the maximum separation between the modelled classes and explained nearly 30% of the total variance of the \mathbf{X} matrix. The score plot of the two low-order LVs is reported as an example in Figure 6.7. BMs were characterized, on average, by positive scores on LV 1 and negative scores on LV 2, whereas CCs were characterized mainly by negative scores on LV 1.

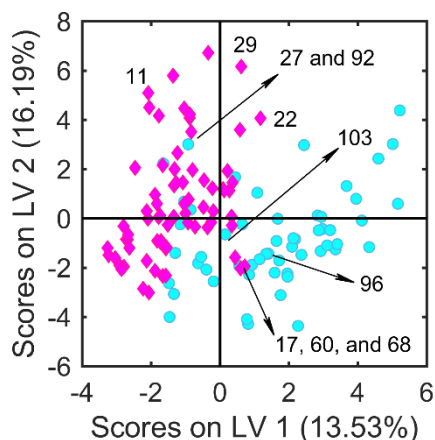


Figure 6.7. Distribution of the objects in the score space of the first two LVs (the fraction of total variance that a given LV explains is reported on the corresponding axis). Objects are marked according to their class (magenta diamonds: CC; cyan circles: BM). Reprinted from [54].

Information regarding the presence of outliers and/or objects with a high influence on the model can be deduced by inspecting few statistical parameters:

- **Hotelling's T^2 vs. squared Q residuals:** these two statistical parameters denote if an object conforms to the variability that has been captured by the model. An object that holds a high Hotelling's T^2 has a variance-covariance structure similar to the other objects, but its features are extreme. An object that holds a high squared Q residuals has a variance-covariance structure different from the other objects and, therefore, it is not well described by the model.
- **Leverage vs. y residuals:** as discussed earlier in **Chapter 2**, the leverage accounts for the influence that an object has on the model, whereas the residuals (i.e., the term \mathbf{F}' in Equation 2.23) account for the deviations from the model. In general, if an object is characterized by high residuals (in absolute value) and its leverage is high, it would be worth considering its exclusion from the dataset as its presence could rotate of the LV space.

Such parameters, useful for the model diagnostics, are plotted in Figure 6.8. The most relevant objects are marked with their identification number and are discussed below^{lxxxiii}.

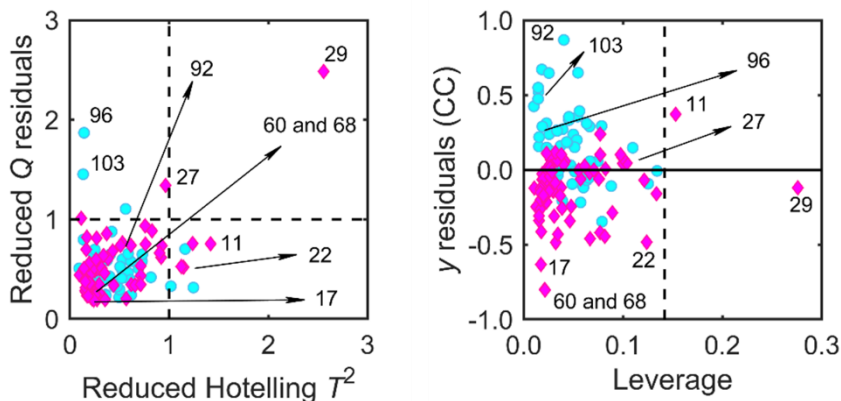


Figure 6.8. Left: reduced Hotelling's T^2 vs. squared Q residuals, the dashed lines show the amplitude of the 95% confidence interval for both statistical parameters. Right: leverage vs. y residuals (related to the CC class), the dashed vertical line shows the leverage threshold above which a sample can be considered influential. Objects are marked according to their class (magenta diamonds: CC; cyan circles: BM). Reprinted from [54].

CCs 27 and 29 have been prepared by combining fatty acids, i.e., lauric acid (27) and palmitic acid (29), with nicotinamide. Object 27 showed high squared Q residuals, whereas object 29 was influential and it was characterized by Hotelling's T^2 and squared Q residuals both outside their 95% confidence interval. The features responsible for the peculiar behavior of these CCs were those related to the difference in molecular dimensions between the partner molecules.

Three objects that involved hexamethylenetetramine had a peculiar behavior. The object with adipic acid (11), belonging to the CC class, held high Hotelling's T^2 and leverage, and it was, therefore, influential. The high Hotelling's T^2 denoted that the variance-covariance structure of this object conformed with the one of the other modelled objects, but certain features, i.e., the difference of number of rings and number of routable bonds, were more extreme with respect to the average. The BMs with limonene and menthone (96 and 103, respectively) held features that were not captured by the model and showed, in fact, high squared Q residuals.

Differences in electronic features, connectivity, molecular weight, and surface area were the ones responsible for the high Hotelling's T^2 of object 22, i.e., ferulic acid pyrazine.

^{lxxxiii} Information regarding which variables are responsible for the peculiar behavior for the discussed objects can be retrieved by inspecting the contribution plots, that are not reported for the sake of brevity.

In addition to what was stated earlier for objects 22 and 103, they held high y residuals (with respect to CC class) and were among the wrong predicted objects in cross validation. In addition to them, cinnamaldehyde/4-hydroxybenzoic acid (17), carvacrol/nicotinamide (60), thymol/tetramethylpyrazine (68), and eugenol/pyrazine (92) held high y residuals (with respect to CC class) and were wrongly assigned to their respective class in cross-validation despite they were in the unsuspecting sample zone (i.e., both low Hotelling's T^2 and squared Q residuals). The misclassification was due to the features of the abovementioned objects, that were related contrarywise with the ones of their respective class.

6.3.3.2. Analysis of the pseudo-regression coefficients

The evaluation of the PLS weights can provide information regarding the correlation between the class membership and the modelled features. As an example, the PLS weights on the two low-order LVs are reported in Figure 6.9.

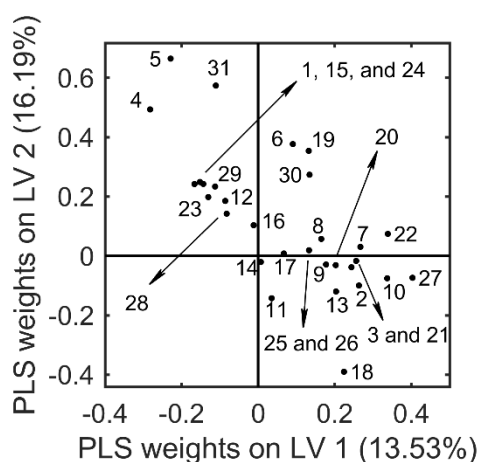


Figure 6.9. PLS weights of the variables for LV 1 vs. LV 2 (the fraction of total variance that a given LV explains is reported on the corresponding axis). Reprinted from [54].

The contribution that each feature has on the discrimination can be retrieved from the evaluation of the pseudo-regression coefficients, whereas the importance of each variable can be evaluated by checking its VIP score. As mentioned earlier in **Chapter 2**, the VIP score denotes the importance a given modelled variable in \mathbf{X} has in explaining the response in \mathbf{Y} . This parameter can guide variable selection and the proposed threshold above which a variable can be considered important is 1.

The pseudo-regression coefficients and the VIP scores (both with respect to the CC class) are plotted in Figure 6.10. The signs of the pseudo-regression coefficients suggested that relevant dissimilarities in the polarizability and exposed surface between the two molecular partners are likely to prevent cocrystallization.

Particularly, this can be inferred considering the negative sign of the coefficients related to differences in atom and bond count (2 and 3, respectively), logP (10), topological polar surface area (22) and number of heteroatoms (27). These findings match with the general rules of thumb of crystal engineering [11,17,35,74] and emerged partially also from EDA.

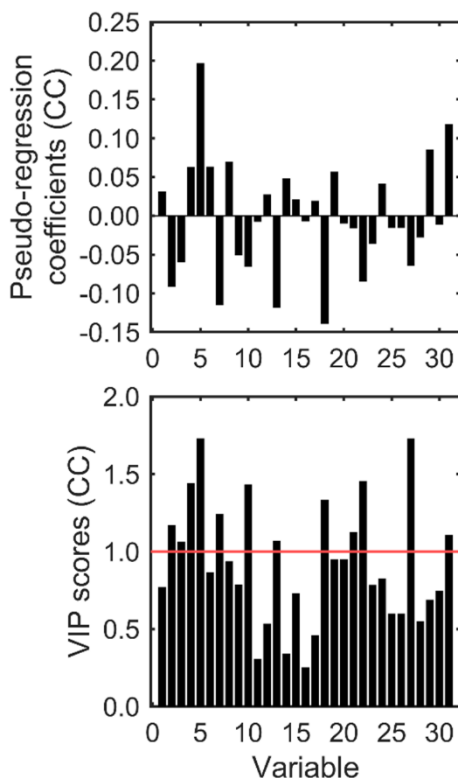


Figure 6.10. Top: pseudo-regression coefficients (with respect to the CC class). Bottom: VIP scores (with respect to the CC class) for the modelled variables, the red horizontal line denotes the significance threshold set to 1. Reprinted from [54].

6.3.4. Validation

Validation was carried out by involving two datasets. The first one was a single subset of data properly extracted from the whole dataset at the very beginning of data analysis, i.e., the single-evaluation test set. An external test set was also involved, to provide a better estimate of the predictive capability of the model when extra variability is included. In this case, the external validation set contained also objects in which at least one of the two partner molecules was not present in the calibration set. The results are summarized in the confusion matrix, reported in Table 6.4.

As far as the single-evaluation test set is concerned, the performance of the model in terms of *NER* was aligned with respect to the one observed in fitting and in cross-validation. This means that in general the model was not affected by a considerable overfitting, as the performance did not change substantially as more and more variability was introduced. In addition to this, a similar performance in terms of *Sn* was observed with respect to the CC class. On the contrary, on the BM class the model appeared to lose in *Sn* on the single-evaluation test set.

Table 6.4. Confusion matrix for the predicted data on the single-evaluation test set and on the external test set. The last two lines show the sensitivity and the non-error rate expressed as percentages.

	Predicted class			
	Single-evaluation test set		External test set	
True class	CC	BM	CC	BM
CC	27	3	29	2
BM	5	19	13	14
<i>Sn</i>	90%	79%	94%	52%
<i>NER</i>	85%		74%	

This phenomenon was much more evident on the external test set, for which the *Sn* on BMs dropped at 52% with 13 out of 27 objects wrongly identified as CCs. This fact is not necessarily negative: in fact, the model appeared fairly conservative in rejecting the possibility of cocrystallization, meaning that fewer potential cocrystals could be missed. On the other hand, the *Sn* for the CC class was totally similar to the one observed in all the other instances, with 29 out of 31 objects correctly identified as CCs. Globally, the *NER* was 74%, which can be considered satisfactory to aid the selection of the cofomer on a computational basis.

Since a consistent portion of objects belonging to the BM class were misclassified in the external test set, the presence of trends and systematic behaviors was assessed. As it is shown in Figure 6.11, not all the samples that do not conform to the model were misclassified, meaning that the behavior is not systematic. The finding could find an explanation in the preparation method that was utilized: in fact, all the BMs were prepared *via* mechanochemical grinding, whereas different methods have been utilized for the preparation of CCs, as many of them were retrieved from the CSD. Therefore, those 13 misclassified BM that were experimentally prepared might yield a CC if synthesized with different methodologies.

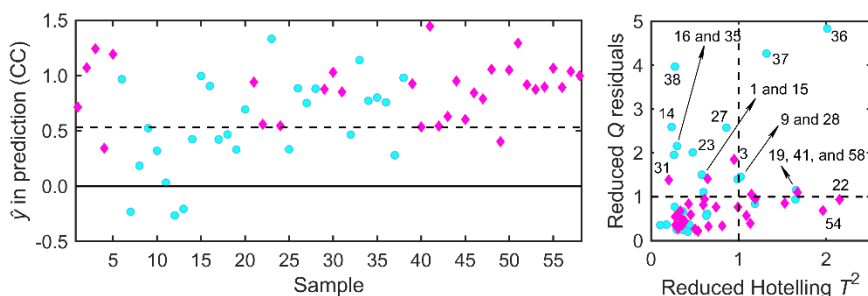


Figure 6.11. Left: predicted response \hat{y} (with respect to the CC class) for the external validation set, the dashed horizontal line shows the hard classification threshold of 0.53. Right: reduced Hotelling's T^2 vs. squared Q residuals, the dashed lines show the amplitude of the 95% confidence interval for both statistical parameters. Samples are marked according to their class (magenta diamonds: cocrytals; cyan circles: binary mixtures). Reprinted from [54].

6.4. Conclusions

This study explored the potential of one of the most known classification techniques, i.e., PLS-DA, in predicting the formation of cocrytals with a *NER* of 74% on external data. The model was trained on both successful and unsuccessful cocrytallization data based on a chemically diverse class of compounds, i.e., active components of EOs and other GRAS molecules. Therefore, it can be stated that the model is suitable for a quite general applicability. The workflow followed the philosophy of QSPR, rationalizing how similar or dissimilar the molecular features of the two synthons should be for cocrytall production. By following a similar approach, other features can be modelled, to aid the synthesis of the desired cocrytals. The model suffered of a bias due to the preparation method that was utilized. Such bias could be reduced if more data regarding failed cocrytallization experiments, with a variety of methods, would be present in the scientific literature.

Note of the author

The results presented in this **Chapter** are published on *Chemometrics and Intelligent Laboratory Systems* 226 (2022) 104580 [54] and on *Food Chemistry* 347 (2021) 129051 [32].

References

- [1] G.R. Desiraju, G.R. Desiraju, *Crystal Engineering: A Holistic View*, Angewandte Chemie International Edition. 46 (2007) 8342–8356. <https://doi.org/10.1002/ANIE.200700534>.

- [2] R. Thakuria, A. Delori, W. Jones, M.P. Lipert, L. Roy, N. Rodríguez-Hornedo, Pharmaceutical cocrystals and poorly soluble drugs, *Int J Pharm.* 453 (2013) 101–125. <https://doi.org/10.1016/J.IJPHARM.2012.10.043>.
- [3] M. Karimi-Jafari, L. Padrela, G.M. Walker, D.M. Croker, Creating cocrystals: A review of pharmaceutical cocrystal preparation routes and applications, *Cryst Growth Des.* 18 (2018) 6370–6387. <https://doi.org/10.1021/acs.cgd.8b00933>.
- [4] J.F. Remenar, S.L. Morissette, M.L. Peterson, B. Moulton, J.M. MacPhee, H.R. Guzmán, Ö. Almarsson, Crystal engineering of novel cocrystals of a triazole drug with 1,4-dicarboxylic acids, *J Am Chem Soc.* 125 (2003) 8456–8457. https://doi.org/10.1021/JA035776P/SUPPL_FILE/JA035776PSI20030611_082112.PDF.
- [5] G.R. Desiraju, Crystal engineering: From molecule to crystal, *J Am Chem Soc.* 135 (2013) 9952–9967. <https://doi.org/10.1021/ja403264c>.
- [6] G.R. Desiraju, Crystal engineering: A brief overview, *Journal of Chemical Sciences.* 122 (2010) 667–675. <https://doi.org/10.1007/s12039-010-0055-2>.
- [7] B.C. Hancock, G. Zografi, Characteristics and significance of the amorphous state in pharmaceutical systems, *J Pharm Sci.* 86 (1997) 1–12. <https://doi.org/10.1021/JS9601896>.
- [8] D.A. Bardwell, C.S. Adjiman, Y.A. Arnautova, E. Bartashevich, S.X.M. Boerrigter, D.E. Braun, A.J. Cruz-Cabeza, G.M. Day, R.G. della Valle, G.R. Desiraju, B.P. van Eijck, J.C. Facelli, M.B. Ferraro, D. Grillo, M. Habgood, D.W.M. Hofmann, F. Hofmann, K.V.J. Jose, P.G. Karamertzanis, A. v. Kazantsev, J. Kendrick, L.N. Kuleshova, F.J.J. Leusen, A. v. Maleev, A.J. Misquitta, S. Mohamed, R.J. Needs, M.A. Neumann, D. Nikylov, A.M. Orendt, R. Pal, C.C. Pantelides, C.J. Pickard, L.S. Price, S.L. Price, H.A. Scheraga, J. van de Streek, T.S. Thakur, S. Tiwari, E. Venuti, I.K. Zhitkov, Towards crystal structure prediction of complex organic compounds - A report on the fifth blind test, *Acta Crystallogr B.* 67 (2011) 535–551. <https://doi.org/10.1107/S0108768111042868/BK5106SUP4.PDF>.
- [9] J.W. Steed, The role of co-crystals in pharmaceutical design., *Trends Pharmacol Sci.* 34 (2013) 185–93. <https://doi.org/10.1016/j.tips.2012.12.003>.
- [10] N.K. Duggirala, M.L. Perry, Ö. Almarsson, M.J. Zaworotko, Pharmaceutical cocrystals: along the path to improved medicines, *Chemical Communications.* 52 (2015) 640–655. <https://doi.org/10.1039/C5CC08216A>.
- [11] P.P. Mazzeo, C. Carraro, A. Monica, D. Capucci, P. Pelagatti, F. Bianchi, S. Agazzi, M. Careri, A. Raio, M. Carta, F. Menicucci, M. Belli, M. Michelozzi, A. Bacchi, Designing a Palette of Cocrystals Based on Essential Oil Constituents for Agricultural Applications, *ACS Sustain Chem Eng.* 7 (2019) 17929–17940. <https://doi.org/10.1021/acssuschemeng.9b04576>.
- [12] J.G.P. Wicker, L.M. Crowley, O. Robshaw, E.J. Little, S.P. Stokes, R.I. Cooper, S.E. Lawrence, Will they co-crystallize?, *CrystEngComm.* 19 (2017) 5336–5340. <https://doi.org/10.1039/C7CE00587C>.
- [13] A. Bacchi, D. Capucci, M. Giannetto, M. Mattarozzi, P. Pelagatti, N. Rodriguez-Hornedo, K. Rubini, A. Sala, Turning Liquid Propofol into Solid (without Freezing It): Thermodynamic Characterization of Pharmaceutical Cocrystals Built with a Liquid Drug, *Cryst Growth Des.* 16 (2016) 6547–6555. <https://doi.org/10.1021/acs.cgd.6b01241>.
- [14] T. Carstens, D.A. Haynes, V.J. Smith, Cocrystals: Solution, Mechanochemistry, and Sublimation, *Cryst Growth Des.* 20 (2020) 1139–1149. https://doi.org/10.1021/ACS.CGD.9B01450/ASSET/IMAGES/LARGE/CG9B01450_0011.JPEG.
- [15] A. Bacchi, P.P. Mazzeo, Cocrystallization as a tool to stabilize liquid active ingredients, *Crystallogr Rev.* (2021) 1–22. <https://doi.org/10.1080/0889311X.2021.1978079>.
- [16] D. Braga, F. Grepioni, L. Maini, P.P. Mazzeo, K. Rubini, Solvent-free preparation of co-crystals of phenazine and acridine with vanillin, *Thermochim Acta.* 507–508 (2010) 1–8. <https://doi.org/10.1016/j.tca.2010.04.021>.
- [17] E. Stoler, J.C. Warner, Non-Covalent Derivatives: Cocrystals and Eutectics, *Molecules.* 20 (2015) 14833–14848. <https://doi.org/10.3390/MOLECULES200814833>.

- [18] F. Bakkali, S. Averbeck, D. Averbeck, M. Idaomar, Biological effects of essential oils - A review, *Food and Chemical Toxicology*. 46 (2008) 446–475. <https://doi.org/10.1016/j.fct.2007.09.106>.
- [19] S. Sharma, S. Barkauskaite, A.K. Jaiswal, S. Jaiswal, Essential oils as additives in active food packaging, *Food Chem.* (2020) 128403. <https://doi.org/10.1016/j.foodchem.2020.128403>.
- [20] R. Pavela, Essential oils for the development of eco-friendly mosquito larvicides: A review, *Ind Crops Prod.* 76 (2015) 174–187. <https://doi.org/10.1016/j.indcrop.2015.06.050>.
- [21] E. Pichersky, R.A. Raguso, Why do plants produce so many terpenoid compounds?, *New Phytologist*. 220 (2018) 692–702. <https://doi.org/10.1111/NPH.14178>.
- [22] Y. Kourkoutas, M. Angane, S. Swift, K. Huang, C.A. Butts, S.Y. Quek, Essential Oils and Their Major Components: An Updated Review on Antimicrobial Activities, Mechanism of Action and Their Potential Application in the Food Industry, *Foods* 2022, Vol. 11, Page 464. 11 (2022) 464. <https://doi.org/10.3390/FOODS11030464>.
- [23] D.W. Christianson, Structural and Chemical Biology of Terpenoid Cyclases, *Chem Rev.* 117 (2017) 11570–11648. https://doi.org/10.1021/ACS.CHEMREV.7B00287/ASSET/IMAGES/LARGE/CR-2017-00287U_0045.JPEG.
- [24] F. Zhou, E. Pichersky, More is better: the diversity of terpene metabolism in plants, *Curr Opin Plant Biol.* 55 (2020) 1–10. <https://doi.org/10.1016/j.pbi.2020.01.005>.
- [25] F. Nazzaro, F. Fratianni, L. de Martino, R. Coppola, V. de Feo, Effect of Essential Oils on Pathogenic Bacteria, *Pharmaceuticals* 2013, Vol. 6, Pages 1451–1474. 6 (2013) 1451–1474. <https://doi.org/10.3390/PH6121451>.
- [26] A.N. Panche, A.D. Diwan, S.R. Chandra, Flavonoids: an overview, *J Nutr Sci.* 5 (2016) e47. <https://doi.org/10.1017/JNS.2016.41>.
- [27] J. Tolls, H. Berger, A. Klenk, M. Meyberg, R. Müller, K. Rettinger, J. Steber, Environmental safety aspects of personal care products—A European perspective, *Environ Toxicol Chem.* 28 (2009) 2485–2489. <https://doi.org/10.1897/09-104.1>.
- [28] S. Dey, F. Bano, A. Malik, Pharmaceuticals and personal care product (PPCP) contamination—a global discharge inventory, *Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology Emerging Contaminants and Micro Pollutants.* (2019) 1–26. <https://doi.org/10.1016/B978-0-12-816189-0.00001-9>.
- [29] R.C. Maia, L.L. Silva, E.F. Mazzeu, M.M. Fumian, C.M. de Rezende, A.C. Doriguetto, R.S. Corrêa, A.L.P. Miranda, E.J. Barreiro, C.A.M. Fraga, Synthesis and analgesic profile of conformationally constrained N-acylhydrazone analogues: Discovery of novel N-arylideneamino quinazolin-4(3H)-one compounds derived from natural safrole, *Bioorg Med Chem.* 17 (2009) 6517–6525. <https://doi.org/10.1016/j.bmc.2009.08.009>.
- [30] E.J. Barreiro, M.E.F. Lima, The Synthesis and Anti-inflammatory Properties of a New Sulindac Analogue Synthesized from Natural Safrole, *J Pharm Sci.* 81 (1992) 1219–1222. <https://doi.org/10.1002/JPS.2600811219>.
- [31] A.C. Guimarães, L.M. Meireles, M.F. Lemos, M.C.C. Guimarães, D.C. Endringer, M. Fronza, R. Scherer, Antibacterial Activity of Terpenes and Terpenoids Present in Essential Oils, *Molecules.* 24 (2019) 2471. <https://doi.org/10.3390/MOLECULES24132471>.
- [32] F. Bianchi, F. Fornari, N. Riboni, C. Spadini, C.S. Cabassi, M. Iannarelli, C. Carraro, P.P. Mazzeo, A. Bacchi, S. Orlandini, S. Furlanetto, M. Careri, Development of novel cocrystal-based active food packaging by a Quality by Design approach, *Food Chem.* 347 (2021) 129051. <https://doi.org/10.1016/j.foodchem.2021.129051>.
- [33] J. Wiczyńska, I. Cavoski, Antimicrobial, antioxidant and sensory features of eugenol, carvacrol and trans-anethole in active packaging for organic ready-to-eat iceberg lettuce, *Food Chem.* 259 (2018) 251–260. <https://doi.org/10.1016/j.foodchem.2018.03.137>.
- [34] L. Pavoni, D.R. Perinelli, G. Bonacucina, M. Cespi, G.F. Palmieri, An overview of micro- and nanoemulsions as vehicles for essential oils: Formulation, preparation and stability, *Nanomaterials.* 10 (2020) 135. <https://doi.org/10.3390/nano10010135>.

- [35] F. Montisci, P.P. Mazzeo, C. Carraro, M. Prencipe, P. Pelagatti, F. Fornari, F. Bianchi, M. Careri, A. Bacchi, Dispensing Essential Oil Components through Cocrystallization: Sustainable and Smart Materials for Food Preservation and Agricultural Applications, *ACS Sustain Chem Eng.* 10 (2022) 8388–8399. https://doi.org/10.1021/ACSSUSCHEMENG.2C01257/SUPPL_FILE/SC2C01257_SI_001.CIF.
- [36] M.D. Perera, J. Desper, A.S. Sinha, C.B. Aakerøy, Impact and importance of electrostatic potential calculations for predicting structural patterns of hydrogen and halogen bonding, *CrystEngComm.* 18 (2016) 8631–8636. <https://doi.org/10.1039/c6ce02089e>.
- [37] M.C. Etter, Encoding and Decoding Hydrogen-Bond Patterns of Organic Compounds, *Acc Chem Res.* 23 (1990) 120–126. <https://doi.org/10.1021/ar00172a005>.
- [38] C.A. Hunter, Quantifying intermolecular interactions: Guidelines for the molecular recognition toolbox, *Angewandte Chemie - International Edition.* 43 (2004) 5310–5324. <https://doi.org/10.1002/anie.200301739>.
- [39] M.A. Mohammad, A. Alhalaweh, S.P. Velaga, Hansen solubility parameter as a tool to predict cocrystal formation, *Int J Pharm.* 407 (2011) 63–71. <https://doi.org/10.1016/j.ijpharm.2011.01.030>.
- [40] N. Issa, P.G. Karamertzanis, G.W.A. Welch, S.L. Price, Can the formation of pharmaceutical cocrystals be computationally predicted? I. Comparison of lattice energies, *Cryst Growth Des.* 9 (2009) 442–453. <https://doi.org/10.1021/cg800685z>.
- [41] L. Fábián, Cambridge Structural Database Analysis of Molecular Complementarity in Cocrystals, *Cryst Growth Des.* 9 (2009) 1436–1443. <https://doi.org/10.1021/cg800861m>.
- [42] J. Devogelaer, H. Meeke, P. Tinnemans, E. Vlieg, R. Gelder, Co-crystal Prediction by Artificial Neural Networks, *Angewandte Chemie International Edition.* 59 (2020) 21711–21718. <https://doi.org/10.1002/anie.202009467>.
- [43] C.R. Groom, I.J. Bruno, M.P. Lightfoot, S.C. Ward, The Cambridge structural database, *Acta Crystallogr B Struct Sci Cryst Eng Mater.* 72 (2016) 171–179. <https://doi.org/10.1107/S2052520616003954>.
- [44] A. Vriza, A.B. Canaj, R. Vismara, L.J. Kershaw Cook, T.D. Manning, M.W. Gaultois, P.A. Wood, V. Kurlin, N. Berry, M.S. Dyer, M.J. Rosseinsky, One class classification as a practical approach for accelerating π - π co-crystal discovery, *Chem Sci.* 12 (2021) 1702–1719. <https://doi.org/10.1039/d0sc04263c>.
- [45] D. Wang, Z. Yang, B. Zhu, X. Mei, X. Luo, Machine-Learning-Guided Cocrystal Prediction Based on Large Data Base, *Cryst Growth Des.* 20 (2020) 6610–6621. <https://doi.org/10.1021/acs.cgd.0c00767>.
- [46] M. Przybyłek, P. Cysewski, Distinguishing Cocrystals from Simple Eutectic Mixtures: Phenolic Acids as Potential Pharmaceutical Cofomers, *Cryst Growth Des.* 18 (2018) 3524–3534. <https://doi.org/10.1021/acs.cgd.8b00335>.
- [47] M. Przybyłek, T. Jeliński, J. Słabuzewska, D. Ziółkowska, K. Mroczńska, P. Cysewski, Application of Multivariate Adaptive Regression Splines (MARSplines) Methodology for Screening of Dicarboxylic Acid Cocrystal Using 1D and 2D Molecular Descriptors, *Cryst Growth Des.* 19 (2019) 3876–3887. <https://doi.org/10.1021/acs.cgd.9b00318>.
- [48] M.E. Mswahili, M.J. Lee, G.L. Martin, J. Kim, P. Kim, G.J. Choi, Y.S. Jeong, Cocrystal prediction using machine learning models and descriptors, *Applied Sciences.* 11 (2021) 1–12. <https://doi.org/10.3390/app11031323>.
- [49] R. Todeschini, V. Consonni, P. Gramatica, Chemometrics in QSAR, *Comprehensive Chemometrics.* 4 (2009) 129–172. <https://doi.org/10.1016/B978-044452701-1.00007-7>.
- [50] S. Yousefinejad, B. Hemmateenejad, Chemometrics tools in QSAR/QSPR studies: A historical perspective, *Chemometrics and Intelligent Laboratory Systems.* 149 (2015) 177–204. <https://doi.org/10.1016/J.CHEMOLAB.2015.06.016>.

- [51] K.E. Hevener, D.M. Ball, J.K. Buolamwini, R.E. Lee, Quantitative structure–activity relationship studies on nitrofuranyl anti-tubercular agents, *Bioorg Med Chem.* 16 (2008) 8042–8053. <https://doi.org/10.1016/j.BMC.2008.07.070>.
- [52] N. Edraki, B. Hemmateenejad, R. Miri, M. Khoshneviszade, QSAR study of phenoxy pyrimidine derivatives as potent inhibitors of p38 kinase using different chemometric tools, *Chem Biol Drug Des.* 70 (2007) 530–539. <https://doi.org/10.1111/J.1747-0285.2007.00597.X>.
- [53] N. Basant, C. Durante, M. Cocchi, M.C. Menziani, Modeling the Binding Affinity of p38 α MAP Kinase Inhibitors by Partial Least Squares Regression, *Chem Biol Drug Des.* 80 (2012) 455–470. <https://doi.org/10.1111/J.1747-0285.2012.01419.X>.
- [54] F. Fornari, F. Montisci, F. Bianchi, M. Cocchi, C. Carraro, F. Cavaliere, P. Cozzini, F. Peccati, P.P. Mazzeo, N. Riboni, M. Careri, A. Bacchi, Chemometric-assisted cocrystallization: supervised pattern recognition for predicting the formation of new functional cocrystals, *Chemometrics and Intelligent Laboratory Systems.* 226 (2022) 104580. <https://doi.org/10.1016/J.CHEMOLAB.2022.104580>.
- [55] J. F \ddot{u} rnkranz, P.K. Chan, S. Craw, C. Sammut, W. Uther, A. Ratnaparkhi, X. Jin, J. Han, Y. Yang, K. Morik, M. Dorigo, M. Birattari, T. St \ddot{u} tzle, P. Brazdil, R. Vilalta, C. Giraud-Carrier, C. Soares, J. Rissanen, R.A. Baxter, I. Bruha, R.A. Baxter, G.I. Webb, L. Torgo, A. Banerjee, H. Shan, S. Ray, P. Tadepalli, Y. Shoham, R. Powers, Y. Shoham, R. Powers, G.I. Webb, S. Ray, S. Scott, H. Blockeel, L. De Raedt, Manhattan Distance, *Encyclopedia of Machine Learning.* (2011) 639–639. https://doi.org/10.1007/978-0-387-30164-8_506.
- [56] R.D. Snee, Validation of Regression Models: Methods and Examples, *Technometrics.* 19 (1977) 415–428. <https://doi.org/10.1080/00401706.1977.10489581>.
- [57] R. Bro, A.K. Smilde, Principal component analysis, *Analytical Methods.* 6 (2014) 2812–2831. <https://doi.org/10.1039/c3ay41907j>.
- [58] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemometrics and Intelligent Laboratory Systems.* 2 (1987) 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9).
- [59] M. Li Vigni, C. Durante, M. Cocchi, *Exploratory Data Analysis*, in: *Data Handling in Science and Technology*, Elsevier Ltd, 2013: pp. 55–126. <https://doi.org/10.1016/B978-0-444-59528-7.00003-X>.
- [60] S. Wold, M. Sj \ddot{o} str \ddot{o} m, L. Eriksson, PLS-regression: A basic tool of chemometrics, in: *Chemometrics and Intelligent Laboratory Systems*, Elsevier, 2001: pp. 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1).
- [61] U.G. Indahl, The geometry of PLS1 explained properly: 10 key notes on mathematical properties of and some alternative algorithmic approaches to PLS1 modelling, *J Chemom.* 28 (2014) 168–180. <https://doi.org/10.1002/cem.2589>.
- [62] S. Wold, H. Martens, H. Wold, The multivariate calibration problem in chemistry solved by the PLS method, in: Springer, Berlin, Heidelberg, 1983: pp. 286–293. <https://doi.org/10.1007/bfb0062108>.
- [63] S. Wold, A. Ruhe, H. Wold, I. W. J. Dunn, The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses, *SIAM Journal on Scientific and Statistical Computing.* 5 (1984) 735–743. <https://doi.org/10.1137/0905052>.
- [64] M. Cocchi, A. Biancolillo, F. Marini, Chemometric Methods for Classification and Feature Selection, in: *Comprehensive Analytical Chemistry*, Elsevier B.V., 2018: pp. 265–299. <https://doi.org/10.1016/bs.coac.2018.08.006>.
- [65] D. Ruiz-Perez, H. Guan, P. Madhivanan, K. Mathee, G. Narasimhan, So you think you can PLS-DA?, *BMC Bioinformatics.* 21 (2020) 2. <https://doi.org/10.1186/s12859-019-3310-7>.
- [66] S. Wold, E. Johansson, M. Cocchi, PLS: partial least squares projections to latent structures, in: H. Kubinyi (Ed.), *3D QSAR in Drug Design*, ESCOM Science Publishers, Leiden, Netherlands, 1993: pp. 523–550. <https://iris.unimore.it/handle/11380/607273> (accessed October 3, 2022).

- [67] S. Favilla, C. Durante, M.L. Vigni, M. Cocchi, Assessing feature relevance in NPLS models by VIP, *Chemometrics and Intelligent Laboratory Systems*. 129 (2013) 76–86. <https://doi.org/10.1016/J.CHEMOLAB.2013.05.013>.
- [68] M. Baroni, G. Cruciani, S. Sciabola, F. Perruccio, J.S. Mason, A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and application, *J Chem Inf Model*. 47 (2007) 279–294. https://doi.org/10.1021/CI600253E/SUPPL_FILE/CI600253E-FILE002.PDF.
- [69] M. Clark, R.D. Cramer, N. van Opdenbosch, Validation of the general purpose tripos 5.2 force field, *J Comput Chem*. 10 (1989) 982–1012. <https://doi.org/10.1002/JCC.540100804>.
- [70] L.L.C. Schrödinger, The PyMOL molecular graphics system, (2010). https://scholar.google.com/scholar_lookup?title=The%20PyMOL%20molecular%20graphics%20system&author=L.L.C.%20Schr%C3%B6dinger&publication_year=2010 (accessed November 21, 2022).
- [71] G. Landrum, Rdkit: Open-source cheminformatics software, (2006). https://scholar.google.com/scholar_lookup?title=RdKit%3A%20Open-Source%20Cheminformatics%20Software&author=G.%20Landrum&publication_year=2021#d=g_s_cit&t=1669048018046&u=%2Fscholar%3Fq%3Dinfo%3Af7fGBvDBPugl%3Ascholar.google.com%2F%26output%3Dcite%26scirp%3D1%26hl%3Dit (accessed November 21, 2022).
- [72] L.E. Juarez-Orozco, O. Martinez-Manzanera, S. v. Nesterov, S. Kajander, J. Knuuti, The machine learning horizon in cardiac hybrid imaging, *Eur J Hybrid Imaging*. 2 (2018) 1–15. <https://doi.org/10.1186/S41824-018-0033-3/FIGURES/5>.
- [73] A. Singh, N. Thakur, A. Sharma, A review of supervised machine learning algorithms, in: 3rd International Conference on Computing for Sustainable Global Development (INDIACom), Institute of Electrical and Electronics Engineers, New Delhi, India, 2016: pp. 1310–1315. <https://ieeexplore.ieee.org/abstract/document/7724478>.
- [74] G. Bolla, B. Sarma, A.K. Nangia, Crystal Engineering of Pharmaceutical Cocrystals in the Discovery and Development of Improved Drugs, *Chem Rev*. 122 (2022) 11514–11603. https://doi.org/10.1021/ACS.CHEMREV.1C00987/ASSET/IMAGES/MEDIUM/CR1C00987_0117.GIF.

Appendix

Table 6A.1. Objects included in the calibration set, each identified by its own identification number and class membership.

Id. ^a	Class ^b	Molecule A	Molecule B
1	CC	2,4-dihydroxybenzoic acid	nicotinamide
2	CC	2,6-dimethylpyrazine	urea
3	CC	2-methylpyrazine	urea
4	CC	2-phenylpropanoic acid	isonicotinamide
5	CC	3-hydroxybenzoic acid	isonicotinamide
6	CC	3-hydroxybenzoic acid	tetramethylpyrazine
7	CC	3-methoxybenzoic acid	isonicotinamide
8	CC	4-hydroxybenzoic acid	hexamethylenetetramine
9	CC	4-hydroxybenzoic acid	nicotinamide
10	CC	4-methoxybenzoic acid	isonicotinamide
11	CC	adipic acid	hexamethylenetetramine
12	CC	adipic acid	pyrazine
13	CC	azelaic acid	isonicotinamide
14	CC	benzoic acid	nicotinamide
15	CC	carvacrol	tetramethylpyrazine
16	CC	catechol	hexamethylenetetramine
17	CC	cinnamaldehyde	4-hydroxybenzoic acid
18	CC	ethyl gallate	isonicotinamide
19	CC	ethyl paraben	nicotinamide
20	CC	eugenol	hexamethylenetetramine
21	CC	ferulic acid	nicotinamide
22	CC	ferulic acid	pyrazine
23	CC	fumaric acid	isonicotinamide
24	CC	fumaric acid	nicotinamide
25	CC	glutaric acid	isonicotinamide
26	CC	glutaric acid	urea
27	CC	lauric acid	nicotinamide
28	CC	2-methylpyridine	urea
29	CC	palmitic acid	nicotinamide
30	CC	phenylacetic acid	hexamethylenetetramine
31	CC	phenylacetic acid	nicotinamide
32	CC	propanoic acid	isonicotinamide
33	CC	resorcinol	isonicotinamide

34	CC	resorcinol	tetramethylpyrazine
35	CC	resorcinol	urea
36	CC	salicylic acid	nicotinamide
37	CC	succinic acid	hexamethylenetetramine
38	CC	succinic acid	isonicotinamide
39	CC	succinic acid	nicotinamide
40	CC	succinic acid	pyrazine
41	CC	succinic acid	urea
42	CC	tartaric acid	pyrazine
43	CC	tartaric acid	urea
44	CC	thymol	2,3-dimethylquinoxaline
45	CC	thymol	hexamethylenetetramine
46	CC	thymol	pyrazine
47	CC	<i>E</i> -2-hexenoic acid	isonicotinamide
48	CC	vanillic acid	hexamethylenetetramine
49	CC	vanillic acid	isonicotinamide
50	CC	vanillic acid	urea
51	CC	vanillin	4-hydroxybenzoic acid
52	CC	vanillin	isonicotinamide
53	CC	3-hydroxybenzoic acid	2,3-dimethylpyrazine
54	CC	3-hydroxybenzoic acid	nicotinamide
55	CC	3-hydroxybenzoic acid	pyrazine
56	CC	4-hydroxybenzoic acid	isonicotinamide
57	CC	adipic acid	nicotinamide
58	CC	azelaic acid	nicotinamide
59	CC	carvacrol	2,3-dimethylquinoxaline
60	CC	carvacrol	nicotinamide
61	CC	carvacrol	pyrazine
62	CC	cinnamic acid	nicotinamide
63	CC	eugenol	isonicotinamide
64	CC	glutaric acid	nicotinamide
65	CC	malonic acid	isonicotinamide
66	CC	phenylacetic acid	isonicotinamide
67	CC	salicylic acid	isonicotinamide
68	CC	thymol	isonicotinamide
69	CC	thymol	tetramethylpyrazine
70	CC	urea	salicylic acid
71	CC	vanillin	nicotinamide

72	BM	1,8-cineol	4-hydroxybenzoic acid
73	BM	1,8-cineol	cinnamic acid
74	BM	1,8-cineol	isonicotinamide
75	BM	1,8-cineol	kojic acid
76	BM	1,8-cineol	salicylic acid
77	BM	carvacrol	4-hydroxybenzoic acid
78	BM	carvacrol	cinnamic acid
79	BM	carvacrol	kojic acid
80	BM	cinnamaldehyde	ascorbic acid
81	BM	cinnamaldehyde	cinnamic acid
82	BM	cinnamaldehyde	ferulic acid
83	BM	cinnamaldehyde	kojic acid
84	BM	cinnamaldehyde	nicotinamide
85	BM	cinnamaldehyde	pyrazine
86	BM	cinnamaldehyde	salicylic acid
87	BM	cinnamic acid	ascorbic acid
88	BM	eugenol	4-hydroxybenzoic acid
89	BM	eugenol	ascorbic acid
90	BM	eugenol	kojic acid
91	BM	eugenol	nicotinamide
92	BM	eugenol	pyrazine
93	BM	ferulic acid	4-hydroxybenzoic acid
94	BM	limonene	4-hydroxybenzoic acid
95	BM	limonene	ascorbic acid
96	BM	limonene	hexamethylenetetramine
97	BM	limonene	kojic acid
98	BM	menthol	cinnamic acid
99	BM	menthol	isonicotinamide
100	BM	menthol	nicotinamide
101	BM	menthone	4-hydroxybenzoic acid
102	BM	menthone	ascorbic acid
103	BM	menthone	hexamethylenetetramine
104	BM	menthone	kojic acid
105	BM	menthone	nicotinamide
106	BM	menthone	salicylic acid
107	BM	thymol	menthol
108	BM	thymol	nicotinamide
109	BM	thymol	salicylic acid

110	BM	vanillin	ascorbic acid
111	BM	vanillin	cinnamic acid
112	BM	dehydrozingerone	4-hydroxybenzoic acid
113	BM	dehydrozingerone	ascorbic acid
114	BM	zingerone	ascorbic acid
115	BM	dehydrozingerone	isonicotinamide
116	BM	zingerone	isonicotinamide
117	BM	Z-nerol	ascorbic acid
118	BM	Z-nerol	cinnamic acid
119	BM	Z-nerol	salicylic acid
120	BM	1,8-cineol	nicotinamide
121	BM	carvacrol	ascorbic acid
122	BM	carvacrol	ferulic acid
123	BM	carvacrol	salicylic acid
124	BM	limonene	isonicotinamide
125	BM	menthone	isonicotinamide
126	BM	thymol	cinnamic acid
127	BM	dehydrozingerone	nicotinamide

^a identification numbers. ^b CC: cocrystals; BM: binary mixtures.

Table 6A.2. Objects included in the single-evaluation test set, each identified by its own identification number and class membership.

Id. ^a	Class ^b	Molecule A	Molecule B
1	CC	4-hydroxybenzoic acid	pyrazine
2	CC	adipic acid	isonicotinamide
3	CC	adipic acid	urea
4	CC	ascorbic acid	nicotinamide
5	CC	carveol	isonicotinamide
6	CC	catechol	tetramethylpyrazine
7	CC	cinnamic acid	hexamethylenetetramine
8	CC	cinnamic acid	isonicotinamide
9	CC	ferulic acid	isonicotinamide
10	CC	ferulic acid	tetramethylpyrazine
11	CC	fumaric acid	urea
12	CC	glutaric acid	pyrazine
13	CC	lauric acid	pyrazine
14	CC	malonic acid	nicotinamide
15	CC	malonic acid	pyrazine
16	CC	<i>o</i> -cresol	hexamethylenetetramine
17	CC	<i>o</i> -cresol	urea
18	CC	octanoic acid	pyrazine
19	CC	propanoic acid	pyrazine
20	CC	salicylic acid	2,3-dimethylpyrazine
21	CC	syringic acid	nicotinamide
22	CC	vanillic acid	nicotinamide
23	CC	vanillin	hexamethylenetetramine
24	CC	vanillin	urea
25	CC	2,5-dimethylpyrazine	urea
26	CC	4-hydroxybenzoic acid	tetramethylpyrazine
27	CC	benzoic acid	isonicotinamide
28	CC	carvacrol	hexamethylenetetramine
29	CC	carvacrol	isonicotinamide
30	CC	lauric acid	isonicotinamide
31	BM	1,8-cineol	ascorbic acid
32	BM	4-hydroxybenzoic acid	ascorbic acid
33	BM	4-hydroxybenzoic acid	cinnamic acid
34	BM	4-hydroxybenzoic acid	kojic acid
35	BM	cinnamaldehyde	isonicotinamide

36	BM	cinnamaldehyde	menthol
37	BM	eugenol	salicylic acid
38	BM	geraniol	menthol
39	BM	limonene	cinnamic acid
40	BM	limonene	nicotinamide
41	BM	limonene	salicylic acid
42	BM	linalool	4-hydroxybenzoic acid
43	BM	linalool	ascorbic acid
44	BM	menthol	hexamethylenetetramine
45	BM	menthone	cinnamic acid
46	BM	thymol	ascorbic acid
47	BM	thymol	ferulic acid
48	BM	urea	hexamethylenetetramine
49	BM	zingerone	4-hydroxybenzoic acid
50	BM	zingerone	nicotinamide
51	BM	dehydrozingerone	salicylic acid
52	BM	Z-nerol	kojic acid
53	BM	thymol	4-hydroxybenzoic acid
54	BM	thymol	kojic acid

^a identification numbers. ^b CC: cocrystals; BM: binary mixtures.

Table 6A.3. Objects included in the single-evaluation test set, each identified by its own identification number and class membership.

Id. ^a	Class ^b	Molecule A	Molecule B
1	CC	ethyl gallate	1,8-cineol
2	CC	ethyl gallate	hexamethylenetetramine
3	CC	lauric acid	hexamethylenetetramine
4	CC	tetramethylpyrazine	cinnamaldehyde
5	CC	urea	4-hydroxybenzoic acid
6	BM	2,3-dimethylquinoxaline	cinnamaldehyde
7	BM	ascorbic acid	menthol
8	BM	ethyl gallate	ascorbic acid
9	BM	ethyl gallate	cinnamaldehyde
10	BM	ethyl gallate	menthone
11	BM	kojic acid	carvacrol
12	BM	kojic acid	menthol
13	BM	kojic acid	menthone
14	BM	lauric acid	dehydrozingerone
15	BM	lauric acid	vanillin
16	BM	lauric acid	zingerone
17	BM	salicylic acid	zingerone
18	BM	urea	dehydrozingerone
19	BM	urea	limonene
20	BM	urea	zingerone
21	CC	urea	isonicotinamide
22	CC	urea	lauric acid
23	BM	lauric acid	1,8-cineol
24	CC	3-phenylpropanal	tetramethylpyrazine
25	BM	3-phenylpropanal	4-hydroxybenzoic acid
26	BM	isonicotinamide	maltol
27	BM	leucine	4-hydroxybenzoic acid
28	BM	leucine	isonicotinamide
29	CC	4-hydroxybenzoic acid	maltol
30	CC	isonicotinamide	sorbic acid
31	CC	sorbic acid	tyramine
32	BM	3-phenylpropanal	maltol
33	BM	sorbic acid	2-hydroxyacetophenone
34	BM	sorbic acid	maltol
35	BM	leucine	2-hydroxyacetophenone

36	BM	leucine	maltol
37	BM	leucine	sorbic acid
38	BM	leucine	tyramine
39	CC	glutaric acid	4-hydroxybenzoic acid
40	CC	hexamethylenetetramine	3-hydroxybenzoic acid
41	CC	hexamethylenetetramine	azelaic acid
42	CC	hexamethylenetetramine	benzoic acid
43	CC	hexamethylenetetramine	resorcinol
44	CC	hexamethylenetetramine	syringic acid
45	CC	isonicotinamide	ascorbic acid
46	CC	isonicotinamide	nicotinamide
47	CC	malonic acid	4-hydroxybenzoic acid
48	CC	nicotinamide	ethylene glycol
49	CC	nicotinamide	sinapic acid
50	CC	pyrazine	fumaric acid
51	CC	pyrazine	nonanoic acid
52	CC	urea	2,3-dimethylpyrazine
53	CC	urea	catechol
54	CC	urea	ferulic acid
55	CC	urea	malonic acid
56	CC	urea	nicotinamide
57	CC	urea	pyrazine
58	CC	urea	syringic acid

^a identification numbers. ^b CC: cocrystals; BM: binary mixtures.

Chapter 7 | Novel strategies to produce secondary packaging material from cosmetic waste

DISCLAIMER: IN WRITING THIS CHAPTER MANY PIECES OF INFORMATION HAD TO BE OMITTED AS THEY ARE PROTECTED BY A NON-DISCLOSURE AGREEMENT SIGNED WITH THE PARTNER OF THIS RESEARCH PROJECT, I.E., DAVINES S.P.A.

7.1. Introduction

7.1.1. Chemistry of hair dyes

Hair dyeing is a practice known to humankind at least for 4000 years [1,2]. There is evidence that hair dyeing was a practice known since Ancient Egypt, as mummies were found with hair dyed with henna [3]. In addition, lead combs dipped in acidic solutions, such as vinegar, were utilized to darken hair during the Roman Empire [3]. Nowadays, millions of people dye their hair at home or in hair salons for aesthetics, self-awareness, and self-affirmation [1,2,4].

According to the Regulation EC 1223/2009^{lxxxiv}, hair dyes can be classified in three major categories, depending on how long the color lasts on the hair.

Temporary dyes remain adsorbed on hair fibers after application and have a low capability of penetrating the cortex due to their high molecular weight [2,5]. Temporary dyes are highly water-soluble compounds, often sodium salts of sulfonated molecules [2,3,6]. Commercial formulations such as dyeing shampoos, gels, and lotions, are high in concentration of dyes (0.1–2.0% w/w) since they are formulated to give the desired coloration in one application [2]. Due to their high molecular weight and high solubility in water, temporary hair dyes typically last one wash, but they can last up to three to six washes on bleached hair [2,7]. Acid Yellow 1 [2] represent an example of compound utilized in the formulation of temporary hair dyes (Figure 7.1).

^{lxxxiv} <http://data.europa.eu/eli/reg/2009/1223/2019-08-13> (accessed 09/12/2022)

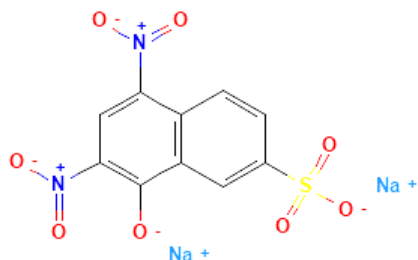


Figure 7.1. Molecular structure of Acid Yellow 1^{lxxxxv}.

Unlike temporary dyes, semi-permanent dyes are on the low-molecular weight spectrum and have a higher capability of penetration towards the cortex [2,5,8]. In terms of formulation, many commercial products contain ammonia or ethanolamine and have a pH value around 9: this ensures the cuticles to open for a better penetration of the product [2]. On average, the coloration is retained for three to six washes. Two subclasses can be identified, depending on the chemical nature of the dye and, consequently, on the nature of the formulation [2,3]:

- **Nitroanilines:** these dyes comprise nitrated derivatives of variously substituted aromatic amines or anthraquinones. Once the product has been applied on the scalp, nitroanilines interacts with the cuticles and cortex through non-covalent forces [2,5]. Nitroanilines are less soluble in water with respect to the salts utilized in temporary dyes, and they require the addition of solvents such as glycerol and benzyl alcohol in the formula to be effective in dyeing [2]. Their poor solubility in water is what makes such dyes more resistant to washing out than temporary ones.
- **Cationic dyes:** these dyes are quaternary ammonium salts. The quaternary ammine is normally located on an aromatic ring, bonded to another aromatic scaffold *via* a variety of strategies, often an azo bond [2,3]. Despite being highly water soluble such as temporary dyes, cationic dyes allow for a better color resistance. In fact, as most of the interfaces, the surface of the hair is negatively charged, allowing for these dyes to electrostatically interact with it [2,5]. Within this frame of reference, the best efficacy of this dyes occurs on damaged hair, as their surface area is higher.

^{lxxxxv} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/2724063> (accessed 10/12/2022)

Washing-resistance can be improved by introducing in the formula also an oxidative dye precursor, and the product requires activation with hydrogen peroxide. These products are called demi-permanent hair dyes and can last up to 20 washes [2]. HC Blue n. 2 and Basic Yellow 57 [2] are examples of nitroaniline and cationic dye, respectively (Figure 7.2).

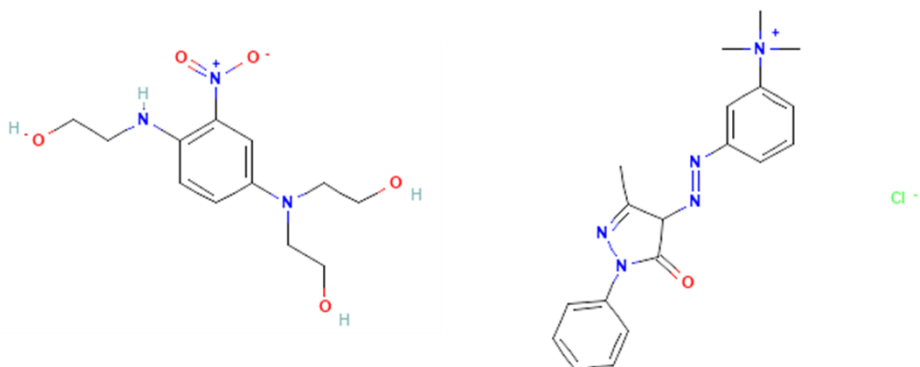


Figure 7.2. Molecular structures of HC Blue n.2^{lxxxvi} (left) and Basic Yellow 57^{lxxxvii} (right).

Permanent dyes include compounds also known as oxidative dyes. Oxidative dyes represent the majority of the hair dye products sold on the market and provide the best performance in terms of washing-resistance [2]. Unlike temporary and semi-permanent dyes, oxidative dyes are not colored by themselves, but require a chemical transformation in order to obtain adducts with absorption bands in the visible region of the electromagnetic spectrum [2,3]. In fact, permanent dyes are often sold as kits of two separate components, i.e., a product containing the dye precursors and a color developer [2,9]. The key ingredients of a formulation as such are:

- **Precursors:** as it was stated above, the dyes in a permanent hair dye are not colored by themselves, but need an oxidant to react. The precursors are of two different kinds and have different functions [2,3,9]:
 - Primary dye: often an aniline substituted with $-NH_2$ or $-OH$ in *p*- or *o*- position, it represents the component of the formulation susceptible to oxidation. After the addition of the oxidant, the primary dye converts into a quinone-like intermediate that is the reactive species. Compounds such as *p*-phenylenediamine and *p*-toluenediamine (Figure 7.3) are among the most frequently utilized primary dyes [3].

^{lxxxvi} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/HC-Blue-no.-2> (accessed 10/12/2022)

^{lxxxvii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/Basic-yellow-57> (accessed 10/12/2022)

- **Coupling dye:** after the reactive intermediate was formed, it can react with other molecules present in the formulation called coupling dyes (or couplers). The reaction yields high-molecular weight products responsible for the coloration. Note that the same primary dye can combine with different couplers to yield different shades. Resorcinol (Figure 7.3) is an example of coupler [3].
- **Retardants:** retardants act as antioxidants. They are added to guarantee the shelf-life of the formulation but play a key role in color development. The aims of retardants are: i) avoiding the conversion of the primary dye into the reactive species due to the oxygen present in the air and ii) retarding the reaction between the primary dye and the developer, so that the precursors have a chance to penetrate the cuticles. The most frequently used retardants include sodium metabisulfite, erythorbic acid, and *t*-butylquinone [2].
- **Oxidants:** oxidants are contained only in the developer product and typically consists of a solution of stabilized hydrogen peroxide, or other compound able to release reactive oxygen species, such as sodium perborate or persulfate [2,3]. The aim of the oxidant is to initiate the conversion of the primary dye into its reactive form.

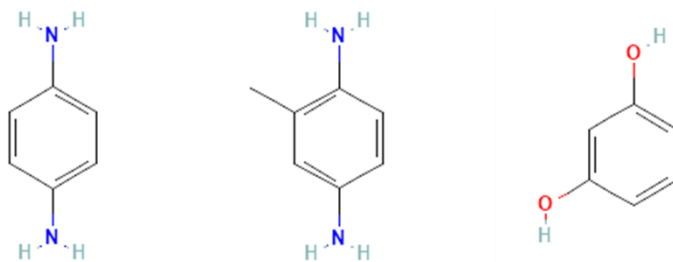


Figure 7.3. Molecular structures of *p*-phenylenediamine^{lxxxviii} (left), *p*-toluenediamine^{lxxxix} (middle) and resorcinol^{xc} (right).

Before the application, the dye with precursors and the developer are mixed according to the manufacturer's direction. The precursors are typically highly soluble and have a low molecular weight, so that, in synergy with the alkaline environment, they can penetrate through the cuticle to the cortex. The coupling reaction takes place after the product has penetrated thanks to the presence of the retardant. The products are higher in molecular weight and remain trapped inside

^{lxxxviii} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/7814> (accessed 11/12/2022)

^{lxxxix} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/7252> (accessed 11/12/2022)

^{xc} Retrieved from <https://pubchem.ncbi.nlm.nih.gov/compound/5054> (accessed 11/12/2022)

the hair fibers. This strategy of color development is what makes oxidative dyes highly resistant to washing [2,3,8,9].

7.1.1.1. Environmental and health concerns

As it was discussed thoroughly in **Chapter 3**, ingredients of cosmetic formulations, including hair dyes, can be identified as Personal Care Products [10,11] and, therefore, as emerging contaminants. As such, their monitoring in the environment is rarely under regulation. The major highways that such substances follow to enter the environment are the drains of households and hair salons [12]. As it holds true for other emerging contaminants, the effect of dyes on the ecosystems and their toxicological profiles are complex and often still matter of debate [10,13,14]. One natural consequence on aquatic environments to think about is related to the chemical composition of dyes commonly used in hair product. As it was discussed earlier, many compounds are rich in nitrogen, and their occurrence in water bodies can lead to eutrophication (see **Chapter 5**) [14–17].

In Europe the utilization of dyes in hair products is under regulation to guarantee the safety of the consumers. Currently, more than two hundred compounds are banned^{xcvi},^{xcvii} and more than one hundred compounds are allowed for restricted use^{xcviii}. In addition, many substances are under surveillance for their potential effects on human health. For instance, *p*-phenylenediamine and *p*-toluenediamine are classified as skin sensitizing substances and there is a major agreement about their toxicity^{xcix},^{xcv}.

Advancements in sewage treatment technologies have been made to effectively remove dyes from wastewater. The proposed strategies include physical (e.g., adsorption), chemical (e.g., photocatalytic degradation and anodic oxidation), and biological approaches [14]. Although such strategies might result effective in the framework of a sewage treatment facility, difficulties may be encountered when dealing with manufacturing waste of hair dye companies. This is mainly because the concentration of dyes is fairly higher than the one found in wastewater.

^{xcvi} <https://ec.europa.eu/docsroom/documents/13210/attachments/1/translations> (accessed 11/12/2022)

^{xcvii} <https://ec.europa.eu/docsroom/documents/13209/attachments/1/translations> (accessed 11/12/2022)

^{xcviii} <https://ec.europa.eu/docsroom/documents/22242> (accessed 11/12/2022)

^{xcix} <https://echa.europa.eu/it/brief-profile/-/briefprofile/100.003.096> (accessed 11/12/2022)

^{xcv} <https://echa.europa.eu/it/brief-profile/-/briefprofile/100.002.221> (accessed 11/12/2022)

7.1.2. Biology of *Fungi*

With an estimated number of 1.5 million species, *Fungi* represent one of the seven taxonomical Kingdoms [18–20]. Constituting a Kingdom on their own, they have a set of unique characteristics that differentiate them from other living organisms [18,19].

Fungi, like organisms belonging to *Animalia* and *Plantae* Kingdoms, are eukaryotic. This means that the genetic material, i.e., the DNA, is confined inside the nucleus of the cell [18,19]. The unique characteristic that distinguishes *Fungi* from all the other eukaryotic organisms is ploidy: in fact, their somatic cells are normally haploid. Like *Plantae*, the cells in *Fungi* are characterized by the presence of a cell wall, but with a substantially different chemical composition. The major components of the fungal cell walls are chitin, a linear polymer of β -(1,4)-N-acetylglucosamine units, and glucans, β -(1,3) and β -(1,6) polymers of glucose, while cellulose represents a minor component [21].

Fungi can exist as unicellular and pluricellular organisms [18,19]. Pluricellular *Fungi* grows in filamentous structures, called hyphae. A hypha can be constituted by one single cell or multiple cells; in both cases, cells can present multiple haploid nuclei instead of only one. What differentiates pluricellular *Fungi* with respect to other organisms that produce filamentous structures, is that hyphae can augment their length only through apical growth. As the cells divide, a hypha can grow linearly or create branches. The filamentous structure constituted by the hyphae is also known as mycelium, that can extend from few μm^2 to even km^2 [22], playing an important role in forests and woods, allowing for the transfer of nutrients between plants and trees through the soil [23,24]. On the contrary, unicellular *Fungi*, also known as yeasts, live as single celled organisms, usually with a diploid set of chromosomes. Certain yeasts, such as *Candida albicans*, present a dimorphic behavior, meaning that they can live both as single cells, but can also produce hyphae under specific environmental conditions.

In terms of reproduction, *Fungi* can reproduce asexually or sexually [18,19,25]:

- **Asexual reproduction:** asexual reproduction is typical both of unicellular and pluricellular *Fungi*. In pluricellular *Fungi*, asexual reproduction can occur through the liberation of spores and mycelial fragmentation, whereas for unicellular ones it happens by budding. In either case, the process is mitotic, meaning that the new organisms have the same DNA of the parent. Figure 7.4 reports an example of asexual life cycle.

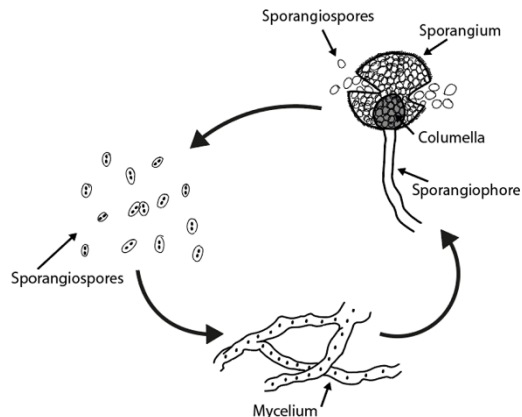


Figure 7.4. Asexual reproductive cycle of Mucoromycetes. The haploid sporangiospores germinate to form hyphae that then develop into a mycelium. In the mature mycelium, hyphae can grow into sporangiophores whose apical extremities develop reproductive structures called sporangia. The sporangium is a structure designated for the maturation of new sporangiospores, that, once mature, are dispersed and the cycle starts over. Adapted after Claire Mary Lee and coworkers [25] (who in turn adapted after Ann Bell [26]).

- Sexual reproduction:** in sexual reproduction, two haploid cells with opposite sexual polarity, i.e., the gametes, combine into a single cell (plasmogamy) and then the two nuclei fuse together (karyogamy) to generate a new diploid organism, genetically different from the parents. Thereafter, meiosis can take place to generate new haploid cells for reproduction. These steps always happen in sequence in all the organisms able to sexually reproduce, but in *Fungi* they can happen at different stages of the life cycle depending on the Phylum. Another important aspect is that sexual spores can be released as a form of safeguard towards unfavorable environmental conditions. The spores remain dormant until favorable circumstances for growth are met, and then germination occurs. Figure 7.5 reports an example of sexual life cycle.

Like *Animalia*, *Fungi* requires outer nutrients as a source of building blocks and energy to survive. They are, in fact, chemoheterotrophic organisms and rely on enzymes for the digestion of macromolecules, whereas soluble nutrients can be directly absorbed by the cells [18,19]. Another aspect *Fungi* have in common with animals is found in the compounds utilized by such organisms to store energy. These compounds are mainly soluble carbohydrates [21], including glycogen, trehalose and few alditols.

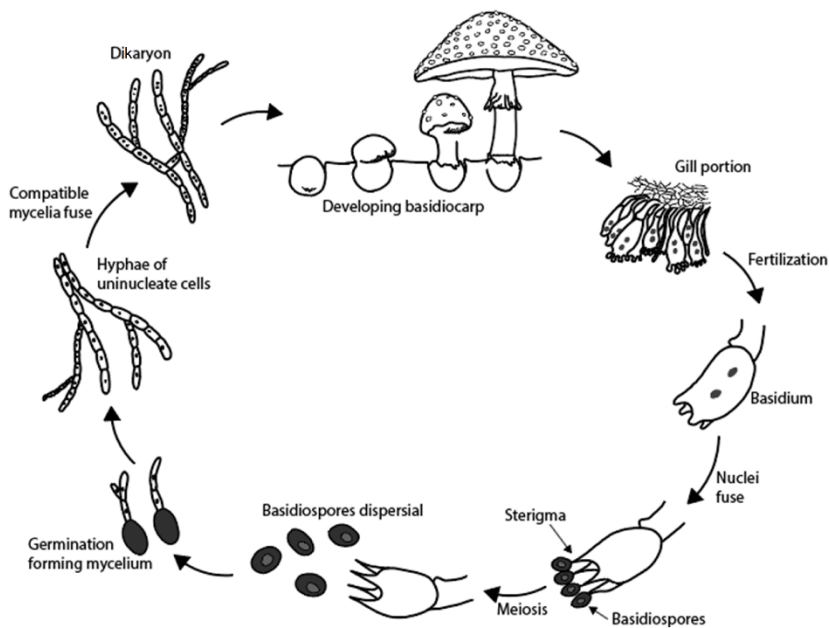


Figure 7.5. Sexual reproductive cycle of Basidiomycetes. The haploid basidiospores germinate to form hyphae that then develop into a mycelium. Mycelia with opposite sexual polarity can fuse together to form the dikaryon, i.e., a new mycelium whose cells contain two haploid nuclei. After the development of the fruiting body, the two haploid nuclei combine into a single diploid one into the reproductive structures called basidia. Immediately after recombination, meiosis takes place to yield four haploid basidiospores (i.e., the gametes). Once mature, basidiospores are released and the cycle starts over. Adapted after Claire Mary Lee and coworkers [25] (who in turn adapted after Thomas N. Taylor and coworkers [27]).

7.1.2.1. Growth factors

The environmental conditions can affect growth, development, and reproduction of *Fungi*. These aspects are of particular importance when *in vitro* growth is attempted [18,19].

In general, *Fungi* tolerate a wide range of temperatures, from few degrees below 0 °C to few degrees above 60 °C depending on the species [18,19]. In addition, different species have different optimal temperatures and different ranges of temperature tolerance. For instance, mesophile *Fungi* are commonly grown in the 10–40 °C range, with the optimum near the room temperature (22–25 °C), whereas human pathogens have an optimal growth temperature around the physiological temperature of the human body.

Most *Fungi* are strict aerobe organisms, meaning that they require oxygen as the terminal electron acceptor in respiration [18,19]. This means that air circulation is necessary when attempting to grow such organisms *in vitro*. There are, however, pluricellular *Fungi* able to utilize different substrates as terminal electron acceptors

(anaerobic respiration). Other species, mainly yeasts, are facultative, meaning that they can grow either in presence or absence of oxygen: in presence of oxygen the predominant metabolic pathway is respiration, in absence of oxygen they can grow by fermenting the substrate.

In terms of pH, different species have their own tolerance and optimal intervals. In general, it has been reported that the optimal pH interval suitable for growth is fairly large, about 5–7 [18,19]. Certain species, mainly yeasts, are acid-tolerant and acidophilic, meaning that can survive and grow, respectively, at pH values of about 2. Strong alkaline environments are tolerated only by very few specialized species.

As it was stated earlier, *Fungi* are chemoheterotrophic organisms, meaning that they require an outer source of nutrients to survive and grow. The main nutrient sources are discussed below:

- **Carbon:** *Fungi* require organic carbon as a source of energy. The spectrum of compounds that can be utilized is very wide, ranging from methane to longer-chain hydrocarbons, or even more complex polymers [18,19]. However, monosaccharides represent the preferred source of organic carbon. The utilization of complex polymers, such as starches and cellulose, as a source of carbon requires the production of exoenzymes. These enzymatic complexes operate outside the cell so that high-molecular weight substrates can be degraded into low-molecular weight molecules that can be more easily transported inside the cell. Certain Basidiomycetes, also known as white-rot mycetes, developed enzymatic complexes able to degrade highly cross-linked polymers like lignin [28].
- **Nitrogen:** nitrogen is among the mineral nutrients (e.g., nitrogen, phosphorus, iron) that are required for fungal growth. Among mineral nutrients, nitrogen is the one required in the highest proportions. As non-nitrogen-fixing organisms, *Fungi* require an outer source of nitrogen [18,19]. Amino acids are the favorite source of nitrogen, but most *Fungi* can directly utilize ammonia as a nitrogen source. Few species can utilize nitrates, by converting them first into nitrites, and then into ammonia. It has been reported that certain species are able to utilize more complex molecules as a source of nitrogen, such as organic dyes [28–32].

7.1.2.2. *The potential of laccases in bioremediation*

White-rot *Fungi* comprise several species involved in the degradation of wood. In particular, white-rot mycelia are the only ones capable of fully degrade lignin [19]. This peculiarity is due to the secretion of a pool of highly specialized exoenzymes.

Among the enzymatic complexes secreted by such fungal species, laccases received an increasing attention in the last twenty years [29,31–35]. The interest towards these enzymes is due to their potential applicability in biotechnological processes. Evidence suggests the suitability of fungal and bacterial laccases as highly selective catalysts for the synthesis of heterocyclic molecules [33,34], but also their effectiveness in degrading pollutants like organic dyes [28–32], paving the way for the utilization of laccase-producing organisms for bioremediation or in sewage treatment facilities.

Laccases are metalloproteins belonging to the family of oxidases [36]. Laccases involve copper ions as cofactors [33]. Figure 7.6 shows the structure of laccase TaLcc1 [37] produced by the Ascomycete *Thielavia arenaria*.

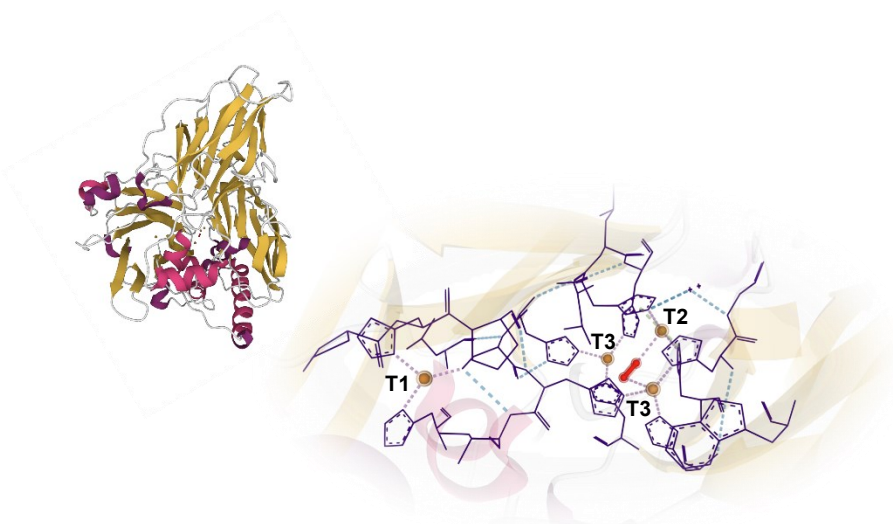


Figure 7.6. Top-left: 3D representation of laccase TaLcc1 with secondary structures highlighted in different colors (α -helices are depicted in strong pink; β -sheets are depicted in tangerine yellow). Bottom-right: focus on the active site with aminoacidic residues shown in skeletal representation, copper ions (copper spheres), and the oxygen molecule (red ball-stick model) in the TnC. Pictures were retrieved from the Protein Data Bank (entry 3PPS^{xcvi}).

Two distinct centers comprising a total of four copper ions are implicated in the active site: the mononuclear center (MnC) constituted by one copper ion Cu T1, and a trinuclear center (TnC) constituted by one copper ion Cu T2 and two copper ions Cu T3 [33]. The distance between the two centers is around 13 Å.

In a laccase-mediated oxidation [33,38]:

^{xcvi} <http://dx.doi.org/10.2210/pdb3pps/pdb> (accessed 15/12/2022)

1. The substrate is oxidized in the MnC and the Cu T1 is reduced from +2 to +1. The cysteine residue binding the Cu T1 re-oxidizes the copper from +1 to +2.
2. The reduced cysteine re-oxidizes itself by transferring one electron to histidine residues binding the Cu T3 causing the reduction of one copper ion in the TnC from +2 to +1. The steps 1 and 2 are repeated until all the four copper ions are in the +1 oxidation state.
3. The oxygen in the TnC is reduced into two water molecules, re-oxidizing the copper ions from +1 to +2, and the cycle starts over.

7.1.3. Mycelium-based materials

In the last few years, the European Parliament focused the attention on changing perspective of the productive cycles, from the traditional linear approach, to a more circular one^{xcvii}. On average, 2.5 billion of tons of waste are produced in Europe every year^{xcviii}. In this context, effective strategies such as more sustainable approaches to waste management and low-carbon footprint processes are demanded to reduce the overall waste production.

The utilization of biocomposite and bio-based materials could play a crucial role in the context of circular economy, as they can be produced starting from biological wastes [39–43]. Biocomposites are composite materials where a biopolymeric or a bio-derived polymeric matrix is structurally reinforced with natural fibers, whereas bio-based materials are materials in which at least one of their constituents is biologically produced and fully biodegradable [39].

In the early '90s, a Japanese scientist named Shigeru Yamanaka moved the first steps in exploring the potential of pluricellular *Fungi* as a resource for the manufacturing of bio-composites [39,44]. With properties similar to expanded polystyrene and polyurethane [39–42], further research investigated the development of mycelium-based composites, with applications ranging from materials for packaging, thermal and acoustic insulation, and building [39,43].

Pluricellular *Fungi* are so attractive for the manufacturing of biocomposite materials for two main reasons:

- Their filamentous structure is ideal as, during growth, the hyphae grow around the substrate leading to the cohesion of loose material [18,19,39].

^{xcvii} <https://www.europarl.europa.eu/news/en/headlines/economy/20151201ST005603/circular-economy-definition-importance-and-benefits> (accessed 15/12/2022)

^{xcviii} <https://www.europarl.europa.eu/news/en/headlines/society/20180328ST000751/eu-waste-management-infographic-with-facts-and-figures> (accessed 15/12/2022)

- The fact that many *Fungi* produces exoenzymes specialized in the digestion of certain substances, waste materials deriving from the wood and paper industries can be utilized as a substrate [18,27–32,39].

The latter could exploit its potential in integrated processes whose feedstock is constituted by toxic waste [28–32] so that bioremediation and the production of biocomposite materials can be carried out in a single step.

In this context, the aim of this study is to develop a mycelium-based material by utilizing cosmetic production wastes, i.e., permanent hair dyes, and cardboard scraps as starting material. In the context of circular economy, this strategy could provide a more sustainable solution to waste management. The results presented in this **Chapter** are related to the very first stages of this research, carried out in close collaboration with Davines S.p.A. (Parma, Italy). The composition of the culture medium for the *in vitro* growth of three fungal species was optimized *via* Mixture Design. In addition, a preliminary evaluation of the degradative capability towards primary dyes and the respective couplers was carried out.

7.2. Materials and methods

7.2.1. Chemicals and materials

Toluene-2,5-diamine sulfate (PTD), 4-amino-*m*-cresol (4AMC), N,N'-bis(2-hydroxyethyl)-*p*-phenylenediamine sulfate (NNB), resorcinol (RES, all 100%), *m*-aminophenol (MAF, 99.8%), and 2-methylresorcinol (2MR, 99.5%) were all kindly provided by Davines S.p.A. (Parma, Italy). Lactic acid ($\geq 88\%$) and *i*-propyl alcohol ($\geq 99.9\%$) were obtained from Carlo Erba Reagents (Milan, Italy). MgSO₄ (98.4%), NaCl (99.9%), and sodium citrate ($\geq 99.0\%$) were obtained from VWR International (Milan, Italy). Glacial acetic acid (99–100%), sodium acetate ($\geq 99.0\%$), and acetonitrile ($> 99.9\%$) were obtained from Sigma-Aldrich (Milan, Italy). Methyl alcohol ($> 99.8\%$) and citric acid ($\geq 99.5\%$) were obtained from Honeywell International Inc. (Morristown, USA).

Milli-Q water was produced by means of a Millipore Milli-Q Element A10 water purification system (Merk-Millipore, Milan, Italy).

Mycelium of species A, B, and C were acquired from Mycelia (Deinze, Belgium). Cardboard and the ingredients for the formulation of hair dyes were all kindly provided by Davines S.p.A.

QuEChERS PSA/C18 + MgSO₄ and PSA/C18/GCB + MgSO₄ dSPE devices were obtained from Restek (Bellefonte, USA), whereas the Supel™ QuE Z-Sep/C18 dSPE devices were obtained from Supelco (Bellefonte, USA).

7.2.2. Culture preparation

Proper amounts of finely divided cardboard and dye were weighted into a porcelain capsule and homogenized with a mortar. Thereafter, 10 mL of an aqueous solution of lactic acid were added to neutralize the alkaline components present in the hair dye. After neutralization, the mycelium was added, and the culture was homogenized. The capsule was covered with perforated Parafilm® (Bemis Company Inc., Wisconsin, USA). Before the incubation, the culture was pressed with a smaller capsule to make it assume a coupe-like shape.

The cultures were kept in aerobic atmosphere at 28 °C, at 85% relative humidity, and in absence of light into a PID system incubator (Instruments s.r.l., Milan, Italy) for 10 days. After incubation, the cultures were thermically treated at 120 °C for 2 h to inhibit further growth.

Before the preparation of the cultures, all the instrumentation was sterilized in a POLEAX EC 10 L autoclave (Quirumed, Valencia, Spain) at 121 °C for 18 min. Between the preparation of one culture and the other, the instrumentation was disinfected with an aqueous solution of 70% v/v *i*-propyl alcohol.

7.2.3. Optimization of the growth factors

The effect of the composition of the culture medium on growth was investigated. For this purpose, blank dye (i.e., dye without the addition of pigments) was utilized. Three main factors ($k = 3$) were investigated, i.e., cardboard, dye, and mycelium. For each culture, always 10 mL of an aqueous solution of lactic acid were added, but the concentration varied proportionally to the amount of dye present in the culture itself.

Two response variables were defined as follows:

- **Germination time** (t_G): amount of time (expressed as h) that had elapsed until visible growth was present. This response was evaluated separately for each species.
- **Shape factor** (S_j): qualitative evaluation that accounts for the ease of mixing and the capability of the culture after thermal treatment in retaining its shape as it was taken out of the capsules. The evaluation scale

goes from a minimum of 1 point to a maximum of 6 points. This response was evaluated separately for each species.

In addition to them, the mixture factor related to the amount in terms of mass (expressed as g) of dye (m_d) in the formulation of the growth medium was evaluated. Therefore, $R = 7$ responses were considered in total.

The experiments were planned in pseudo-component domain (X_1 , X_2 , and X_3), whose composition is reported in Table 7.1, whereas the experimental matrix is reported in Table 7.2. For each trial, a total of 9.50 g of culture were prepared, not including the 10 mL of an aqueous solution of lactic acid.

Table 7.1. Composition of the pseudo-components expressed as weight fraction.

Pseudo-component	Cardboard	Dye	Mycelium
X_1	0.10	0.10	0.80
X_2	0.10	0.50	0.40
X_3	0.50	0.10	0.40

For each species, a total of $N = 15$ experiments were carried out in random order, including $n_0 = 3$ experiments in the center of the experimental domain to estimate the pure experimental variance.

A first-order model (Equation 2.40) was postulated for t_G and S_f . A coefficient was considered significant if the absolute value of its regression coefficient was greater than the semiamplitude of its 95% confidence interval. Each model was evaluated in terms of fraction of explained variance (R^2) and predictive capability in leave-one-out cross-validation (Q^2). The validity was evaluated by carrying out a lack-of-fit F -test ($\alpha = 0.01$).

The global optimal conditions were identified according to Derringer's method [45]. Single desirability functions d_i were defined for t_G as in Equation 7.1a, whereas for S_f and m_d as in Equation 7.1b:

$$d_i = \begin{cases} 0, & y_i > U_i \\ \frac{U_i - y_i}{U_i - L_i}, & \text{otherwise} \end{cases}$$

Equation 7.1a

$$d_i = \begin{cases} 0, & y_i < L_i \\ \frac{y_i - L_i}{U_i - L_i}, & \text{otherwise} \end{cases}$$

Equation 7.1b

Table 7.2. Experimental matrix. The composition of each experiment is reported in as weight fraction in the pseudo-component domain.

Standard order	X_1	X_2	X_3
1	1.00	0.00	0.00
2	0.00	1.00	0.00
3	0.00	0.00	1.00
4	0.67	0.33	0.00
5	0.67	0.00	0.33
6	0.33	0.67	0.00
7	0.00	0.67	0.33
8	0.33	0.00	0.67
9	0.00	0.33	0.67
10	0.67	0.17	0.17
11	0.17	0.67	0.17
12	0.17	0.17	0.67
13	0.33	0.33	0.33
14	0.33	0.33	0.33
15	0.33	0.33	0.33

The optimization boundaries are reported below:

- t_G : L_i was identified as the upper limit of the 95% confidence interval of the minimum predicted response within the experimental domain and of the and U_i was set at 240 h.
- S_f : L_i was set at 2, whereas U_i was set at 6.
- m_d : L_i corresponded to the minimum amount of dye utilized in the formulation of the culture (i.e., 0.95 g), whereas U_i corresponded to the maximum amount of dye utilized in the formulation of the culture (i.e., 4.75 g).

The optimal conditions were in correspondence of the maximum global desirability (Equation 2.42), computed as weighted geometric mean of the single desirability functions. The weights r_i for t_G , S_f and m_d were 3, 1, and 2, respectively. The maximum was found with a derivative-free search algorithm.

7.2.4. Optimization of the QuEChERS procedure

QuEChERS extraction [46] was optimized in terms of extraction salts for salting-out partitioning and dSPE sorbent for clean-up to recover the organic dyes from the culture medium.

Cultures were prepared by using 2.96 g of cardboard, 2.73 g of dye and 10 mL of an aqueous solution of 2.05% v/v lactic acid. No mycelium was added for this evaluation. Three combinations of pigments were tested as model compounds, i.e., PTD + 2MR, 4AMC + RES, NNB + MAF.

The experiments were planned according to a unreplicated two-ways ANOVA experimental design.

- **Factor A – Extraction salts:** three different extraction salts mixtures were tested, according to the three validated variants of the QuEChERS method.
 - **Unbuffered (UB):** 4 g of MgSO₄ and 1 g of NaCl [47].
 - **AOAC:** 6 g of MgSO₄ and 1.5 g of sodium acetate^{xcix}.
 - **EN:** 4 g of MgSO₄, 1 g of NaCl, 1 g of sodium citrate and 0.32 g of citric acid^c.
- **Factor B – dSPE sorbent:** three different clean-up sorbents were tested, i.e., PSA/C18 + MgSO₄, PSA/C18/GCB + MgSO₄, and Z-Sep/C18.

All the other factors were kept constant^{ci}. Therefore, a total of $N = 9$ combinations were tested.

The extracts were diluted 1:20 with Milli-Q water and then submitted to UV-vis analysis. For each dye, the absorbance at the following wavelengths was utilized as the response variable (Table 7.3).

For each treatment, a culture with blank dye also submitted to extraction so that the spectrum of the blank matrix could be utilized for background correction.

^{xcix} http://www.aocofficialmethod.org/index.php?main_page=product_info&products_id=2155 (accessed 09/12/2022)

^c <https://www.en-standard.eu/csn-en-15662-foods-of-plant-origin-multimethod-for-the-determination-of-pesticide-residues-using-gc-and-lc-based-analysis-following-acetonitrile-extraction-partitioning-and-clean-up-by-dispersive-spe-modular-quechers-method/> (accessed 09/12/2022)

^{ci} The extractions that were carried out with the AOAC method utilized a 1% v/v acetic acid in acetonitrile solution for the extraction. The agitation time during the dSPE clean-up varied according to manufacturer directions as follows: PSA/C18 + MgSO₄ required 0.5 min, PSA/C18/GCB + MgSO₄ required 2 min, Z-Sep/C18 required 1 min.

Table 7.3. Utilized wavelength for each compound.

Dye	Compound	Wavelength (nm)
Primary dyes	PTD	256
	4AMC	236
	NNB	250
Couplers	2MR	279
	RES	275
	MAF	288

Two-ways ANOVA was carried out to highlight differences in the average responses for each studied factor. Homoscedasticity and normality were tested beforehand with Hartley's test and Shapiro-Wilk test, respectively. For statistically significant results, the effect size was estimated in terms of Cohen's η^2 [48] and, *post-hoc* multiple pairwise comparisons were carried out with Student's *t*-tests with a Bonferroni correction. The confidence level was 95%.

7.2.5. Operating procedure and instrumental conditions

7.2.5.1. Sample preparation

Ten grams of sample (8.05 g if mycelium was absent) were weighted into a 50 mL polypropylene centrifuge tube and were integrated with 2.500 mL (4.625 mL if mycelium was absent) of cold Milli-Q water so that approximately 10 g of water were present in the sample.

Ten milliliters of cold acetonitrile were introduced, and the sample was vigorously shaken for 1 min. A pre-weighted amount of extraction salts, i.e., 4 g of MgSO_4 and 1 g of NaCl, were added and the tube was vigorously shaken again for 1 min. The tubes were kept at $-20\text{ }^\circ\text{C}$ for 1 h to aid the separation of fats and proteins. Thereafter, the tubes were centrifuged at 5000 rpm for 5 min at $4\text{ }^\circ\text{C}$.

One milliliter of supernatant was introduced in the PSA/C18/GCB + MgSO_4 dSPE clean-up tube and it was agitated for 2 min and then centrifuged at 5000 rpm for 5 min at $4\text{ }^\circ\text{C}$. The extract was, finally, diluted 1:20 with Milli-Q water.

7.2.5.2. UV-vis spectroscopy

The purified and diluted extracts were analyzed in the 210–800 nm range with the aid of an Evolution 260 Bio spectrophotometer (Thermo Fisher Scientific, Massachusetts, USA). The sample was exposed to the radiation by operating in transmittance mode. The sample was placed into a quartz semi-micro cuvette

(Hellma Analytics, Milan, Italy) with a light path of 10 mm. The instrument was operated with a band width and a resolution both set at 1 nm. The integration time was 250 ms.

Before acquiring the spectra, the background signal was registered against a reagent blank, i.e., a 5% v/v solution of acetonitrile in Milli-Q water.

7.2.6. Preliminary evaluation of the degradative capability

A preliminary degradation experiment was set only with species A and with the dye containing PTD + 2MR as dyes. Cultures were prepared by using 2.96 g of cardboard, 2.73 g of dye, 3.81 g of mycelium, and 10 mL of an aqueous solution of 2.05% v/v lactic acid.

Two sets of experiments were prepared in independent duplicates ($n = 2$):

- **Control:** no mycelium was added in the preparation of the culture.
- **Treatment:** mycelium was added in the preparation of the culture.

The cultures were then incubated for 1 h and for 7 days and they were treated according to what is reported in Section 7.2.5. The absorbance at the respective wavelengths was utilized as a response variable to evaluate the degradative capabilities.

An additional set of cultures was prepared in the same way by using blank dye so that the spectrum of the blank matrix could be utilized for background correction.

7.2.7. Software

UV-vis data were acquired with the Thermo Insight (Thermo Fisher Scientific) software.

Multilinear regression and multicriteria optimization were carried out by using custom scripts in MATLAB environment (v. R2022a, Mathworks, Massachusetts, USA). Ternary contour plots were plotted with a custom function in MATLAB environment developed by Joel Lynch^{cii} after adaptation of Ulrich Theune's one^{ciii}.

^{cii} Joel Lynch (2022). ternary_plots (https://github.com/lynch4815/ternary_plots/releases/tag/v1.3.3), GitHub. Retrieved December 8, 2022.

^{ciii} Ulrich Theune (2021). Ternary Plots (<https://www.mathworks.com/matlabcentral/fileexchange/7210-ternary-plots>), MATLAB Central File Exchange. Retrieved May 31, 2021.

7.3. Results and discussion

7.3.1. Preliminary experiments

As discussed earlier in Section 7.1.2.1., environmental conditions are key for the *in vitro* cultivation of living organisms. In the case of this study, the environmental conditions had to be ideal to promote fungal growth and laccase production [18,19,32]. Preliminary experiments were, therefore, carried out to primarily assess environmental conditions and nutritional factors suitable for fungal growth, according with what has been reported in the literature.

These preliminary evaluations were carried out by preparing duplicate cultures for each species by utilizing 12.00 g of mycelium as is, with the only addition of a wheat flour-water slurry as the solely source of nutrients.

It has been reported that temperatures in the 25–30 °C range are generally suitable laccase production depending on the light conditions [18,19,32,36]. Particularly, the optimal temperature for laccase production in the presence of light is reported to be around 25 °C, whereas for cultures grown in the dark the optimal temperature is around 30 °C. In any case, it has been reported that above 30 °C laccases show a reduction in their activity. Based on these considerations, a temperature of 28 °C was chosen and cultures were incubated in the dark. Most of *Fungi* are aerobic organisms, meaning that they require oxygen as the final acceptor of electrons in the electron transport chain. In addition, as the activity of laccases involves molecular oxygen, all the cultures were incubated in aerobic conditions. The capsules were covered with perforated Parafilm® as a compromise between avoiding contamination and guaranteeing air circulation. The relative humidity was around 50%. After 7 days of incubations, barely noticeable growth was observed for species A and B, whereas no growth was observed for species C.

A new set of experiments was carried out by introducing a container filled with distilled water into the incubator. To further increase relative humidity, the internal walls of the incubators were further insulated with a perforated polyethylene plastic bag to guarantee air circulation. Under this experimental conditions, values of relative humidity around 85% were achieved and further growth could be observed.

After temperature and relative humidity were addressed, a new set of preliminary experiments was carried out by utilizing pulverized carbon and blank dye as the sources of carbon and nitrogen. The cultures were prepared by utilizing 6.00 g of mycelium, 1.50 g of pulverized cardboard and 2.00 g of blank hair dye. Under these

conditions, none of the species were able to grow. This was probably due to the alkaline components contained in the hair dye. To confirm this hypothesis, a new set of experiments was carried out by adding to the cultures 10 mL of an aqueous solution containing 1.50% v/v lactic acid to aid neutralization. In these conditions, growth was observed for all the species. Neutralization was of paramount importance, as evidence suggests that strongly alkaline environments not only hinder fungal growth, but also laccase production and activity [32,36,49].

These experimental conditions were adopted for further investigation.

7.3.2. Optimization of the growth factors

The composition of the culture was further investigated. Both nutrients and inoculum size were taken into consideration. The source of carbon is represented by both the cardboard and hair dye, with the latter also constituting the nitrogen source. The amount of cardboard and dye, as well as the inoculum size (i.e., the amount of mycelium to be introduced into the cultures) could also have an impact on the texture of the outcoming material.

The initial composition was 16% of cardboard, 21% of hair dye, and 63% of mycelium (all as % w/w). The experimental domain to be explored for each factor was selected according to preliminary experiments and few constraints:

- For sake of ease in experimental planning, the trials were planned in the pseudo-component domain.
- For all the experimental trials, all the components had to be present, as the goal is to utilize both hair dye and cardboard wastes. In addition, the absence of hair dye in the culture would have led to a nitrogen deficit, being, therefore, not ideal for fungal growth. For these reasons, only combinations in which at least 10% of hair dye and 10% of cardboard were considered.
- As a consequence of the previous constraint, only trials with 80% or less mycelium were studied. Additionally, the minimum amount of inoculum was set at 40% and, accordingly, the maximum amounts of hair dye and cardboard were set at 50%

A total of $N = 15$ (including $n_0 = 3$ replicates at the center of the experimental domain) were carried out with respect to the bare minimum $N = 10$ (replicates included). The chosen arrangement in the pseudo-component experimental domain is summarized in Table 7.2 and shown in Figure 7.7.

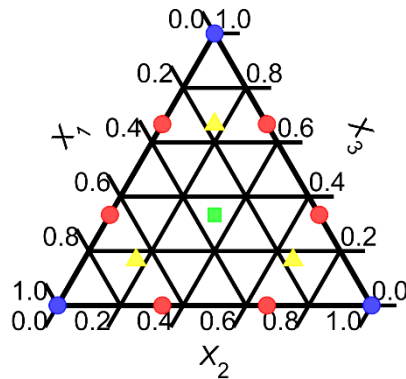


Figure 7.7. Arrangement of the experimental runs in the pseudo-component experimental domain.

The choice of the experimental plan was made to effectively counteract potential experimental failures with the possibility of replanning the experiments in an easy way.

The responses were left untransformed, and the models were calculated. ANOVA showed that all the models were valid ($p > 0.01$), meaning that the error deriving from approximation is not significantly greater than the variance that it could be expected experimentally.

The groups of responses are discussed one at a time.

7.3.2.1. Germination time

The models were good in terms of explained variance and predictive capability in cross-validation, with $R^2 \geq 0.98$ and $Q^2 \geq 0.86$, with the latter being exceptionally good considering that a biological process is taken into consideration [50]. Figure 7.8 depicts the ternary contour plots, whereas the regression models are reported in Table 7.4.

Table 7.4. Regression models calculated for each species with respect to the germination time.

Species	Equation ^a
A	$y = 19 (\pm 3)X_1 + 68 (\pm 3)X_2 + 20 (\pm 3)X_3 - 70 (\pm 20) X_2X_3$
B	$y = 40 (\pm 10)X_1 + 160 (\pm 10)X_2 - 190 (\pm 60) X_2X_3$
C	$y = 50 (\pm 20)X_1 + 180 (\pm 20)X_2 - 200 (\pm 80) X_2X_3$

^a only the significant terms at the 95% confidence level are reported, the coefficients are reported as coefficient (\pm standard error) rounded at one significant digit.

From the contour plots and the regression models it can be noticed that all the species behaved similarly to each other.

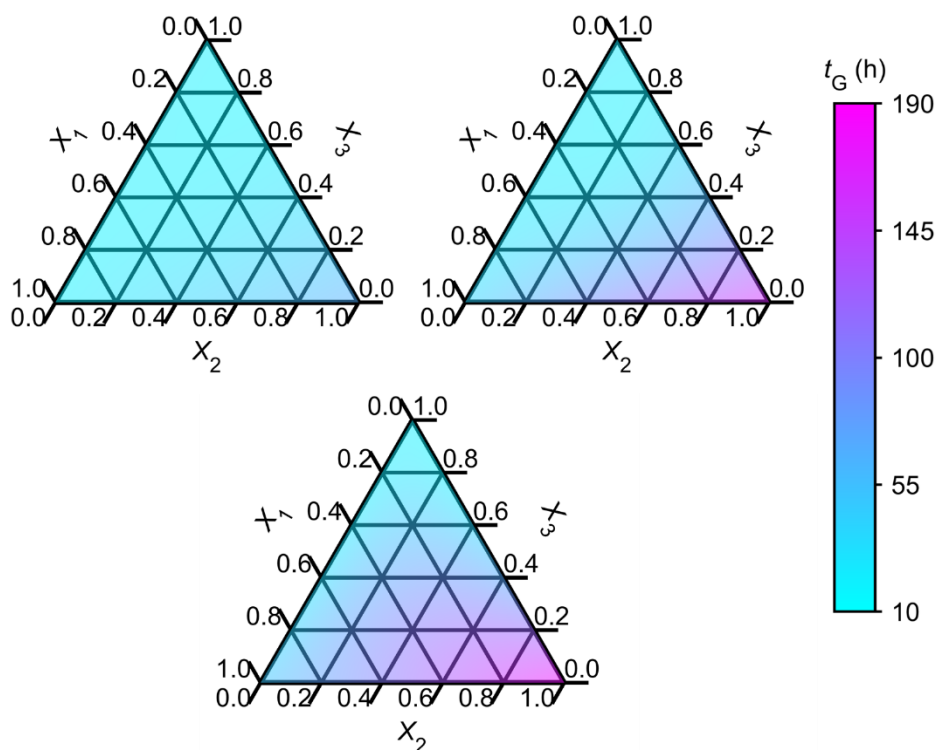


Figure 7.8. Ternary contour plots for germination time for species A (top-left), B (top-right), and C (bottom). All the plots were rescaled to the same colorbar for comparison purposes.

The terms related to pseudo-component X_1 , pseudo-component X_2 and the interaction term between pseudo-components X_2 and X_3 were present in all the models, whereas the term related to pseudo-component X_3 was present only for species A. The effect related to pseudo-component X_2 had the highest magnitude with respect to the others, as the largest variation in response was observed when considering the t_G at a 100% of X_2 with respect to the one at 50% of X_1 and 50% of X_3 . Despite being always statistically significant, the effect related to pseudo-component X_1 was the weakest.

For all the species, visible growth was observed after many hours of incubation when higher proportions of pseudo-component X_2 (rich in hair dye) were involved in the culture. On the contrary, faster growth was observed at high contents of pseudo-component X_3 (rich in cardboard). This was most likely unrelated to the ratio of nutrients, as evidence suggests that high levels of nitrogen are positively correlated with biomass production [51].

This result could find an explanation in the texture of the culture: higher proportions of pseudo-component X_2 resulted in a more liquid and compact consistency, whereas high contents of pseudo-component X_3 yielded a dryer and more porous appearance. The latter allows for a better air circulation with respect to a liquid consistency that could produce an anoxic environment, not compatible with fungal proliferation.

As a general remark, species A resulted more tolerant than species B and C towards the composition of the culture medium, as it showed visible growth in shorter times. On the contrary, longest germination times were observed for species C for a broader zone of the pseudo-component experimental domain. Species B had an intermediate behavior.

7.3.2.2. Shape factor

The models were good in terms of explained variance, whereas the predictive capability in cross-validation was acceptable [50], with $R^2 \geq 0.93$ and $Q^2 \geq 0.51$. Figure 7.9 depicts the ternary contour plots, whereas the regression models are reported in Table 7.5.

As it is shown in the contour plots, the three species behaved differently in terms of ease of culture preparation and the texture of resulting material looked after thermal treatment. This was most likely ascribable to the different texture of mycelia themselves, both during inoculation and after growth.

Table 7.5. Regression models calculated for each species with respect to the shape factor.

Species	Equation ^a
A	$y = 5 (\pm 1)X_1$
B	$y = 6 (\pm 1)X_1 + 17 (\pm 7) X_2X_3$
C	$y = 4 (\pm 1)X_1$

^a only the significant terms at the 95% confidence level are reported, the coefficients are reported as coefficient (\pm standard error) rounded at one significant digit.

From the contour plot it can be observed that the zones corresponding to the highest predicted values for S_f were found at low contents of pseudo-components X_2 and X_3 . Despite being easier to homogenize, cultures at higher proportion of pseudo-component X_2 (rich in hair dye) were lacking cohesion after the thermal treatment. In fact, most of them broke during the extraction from the capsule. The lack of cohesion, again, was ascribed to the poor fungal growth observed in such conditions. In the meanwhile, cultures prepared at a higher proportion of pseudo-component X_3 (rich in cardboard) had a drier appearance and were difficult to

homogenize. This could lead to weak spots in the final material with a consequent deterioration of the mechanical properties.

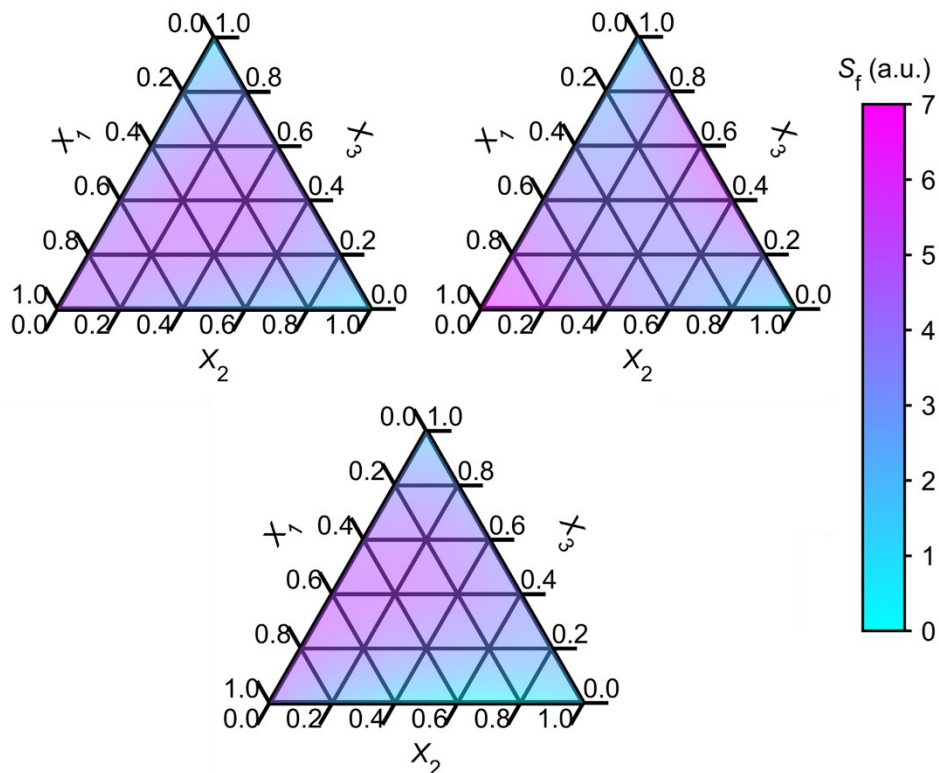


Figure 7.9. Ternary contour plots for shape factor for species A (top-left), B (top-right), and C (bottom). All the plots were rescaled to the same colorbar for comparison purposes.

Generally, the best performances were achieved at higher proportions of pseudo-component X_1 (rich in mycelium) -always statistically significant-, especially at combinations that allowed a fast and a thorough fungal growth. In fact, such trials were easily released from the capsule after the thermal treatment and did not crumble. This fact could be related to the binding effect of the mycelium that was mentioned earlier [18,19,39].

As it was stated earlier, different species produced materials with different texture and appearance:

- The mycelium of species A was easier to incorporate while inoculating the culture and, in the end, easier to release from the capsules as species A showed poor growth only at higher proportions of pseudo-component X_2 .

- Difficulties were encountered in terms of homogenization of species B during inoculum preparation. In addition, good performance in terms of S_f was achieved both near 0% and 100% ratios of pseudo-component X_1 . At high proportions of pseudo-components X_2 poor growth was observed.
- Species C incorporated well during inoculum preparation. In terms of performance, the best S_f were achieved at higher proportions of pseudo-components X_1 and X_3 . For these combinations only, the material was easily removed from the capsule without breaking.

7.3.2.3. Amount of hair dye

The amount of dye utilized for culture preparation was taken into consideration as additional response. Since this response corresponded to one of the mixture factors, the equation of the model was known, and it did not have to be postulated.

The regression model is reported in Equation 7.2.

$$y = 0.95X_1 + 4.75X_2 + 0.95X_3$$

Equation 7.2

The contour plot, reported in Figure 7.10 shows how the amount of hair dye increases linearly by moving towards higher proportions of pseudo-component X_2 (i.e., the richest in hair dye) as it should.

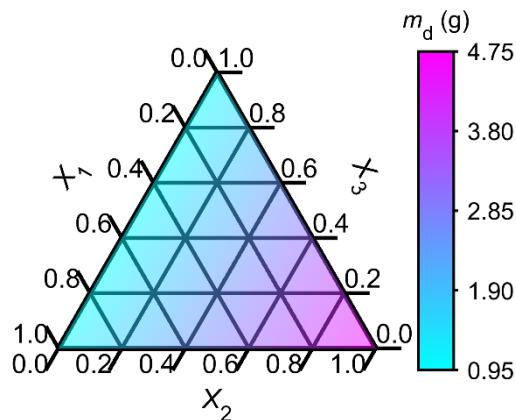


Figure 7.10. Ternary contour plots for the amount of dye involved in the preparation of different cultures.

7.3.2.4. Multicriteria optimization

The Derringer's method [45] was utilized to identify the growth conditions able to guarantee the best compromise between the investigated responses. The optimization boundaries and criteria were discussed with Davines S.p.A.:

- t_G : these responses were to be minimized in order to achieve a fast growth. The upper boundary was set at 240 h, i.e., the typical amount of time required to grow one of the investigated species^{civ}. As for the lower boundary, the upper limit of the 95% confidence interval of the predicted response was chosen. Between the boundaries, the desirability function was varied linearly. Since growth was a fundamental requirement in order to both obtain a suitable material and to potentially produce noticeable degradation of the hair dyes, a weight of $r_i = 3$ was applied.
- S_f : these responses have to be maximized. The upper boundary was set at 6, i.e., the maximum score that was attributed. As for the lower boundary, it was set at 2, representing a condition in which the most trials could be extracted from the capsule without risk of breaking. Between the boundaries, the desirability function was varied linearly. A weight of $r_i = 1$ was applied as this aspect was considered not to be crucial at such an early stage of the research.
- m_d : the amount of hair dye utilized in the cultures was introduced as a mixture factor. Since the manufacturing methodology is batch-based, the amount of dye *per* batch had to be maximized for a more efficient waste management. This reasoning involved only the hair dye and not the cardboard, as the first one is more problematic under the waste management perspective. The upper and lower boundaries were set at 4.75 g and 0.95 g, respectively, i.e., the maximum and the minimum mass of hair dye utilized in the culture media. Between the boundaries, the desirability function was varied linearly. Since this aspect was fairly important, a weight of $r_i = 2$ was applied.

By inspecting the contour plots (Figures 7.8–7.10) it has to be noticed that the optimal conditions have to represent a compromise. In fact, the optimal setting for m_d were not optimal at all for S_f and t_G . The results of the optimization are summarized in Tables 7.6 and 7.7.

A global desirability of $D = 0.79$ was obtained, which was satisfactory considering that, as it was mentioned earlier, few of the optimization criteria had opposite directions. Single desirability values close to 1 were obtained for the t_G of all the investigated species, whereas $d_i \geq 0.42$ and a $d_i = 0.47$ were obtained for the S_f

^{civ} This evidence was found in the literature, but the reference had to be omitted in compliance to the non-disclosure agreement.

values and for m_d , respectively. The identified optimal conditions were valid, since lack-of-fit was not statistically significant.

Table 7.6. Optimization results. The last column reports the single desirability values d_i , whereas the last row reports the global desirability value D .

Response	Weight	Species	d_i
t_G	3	A	1.00
		B	0.96
		C	1.00
S_f	1	A	0.55
		B	0.87
		C	0.42
m_d	2		0.47
D			0.79

Table 7.7. Composition related to the maximum global desirability D , expressed in the pseudo-component domain. Additional columns and the last row show the back-conversion into the explicit mixture factors. The compositions are reported as weight fractions.

Pseudo-component	Optimum	Cardboard	Dye	Mycelium
X_1	0.00	0.00	0.00	0.00
X_2	0.47	0.05	0.23	0.19
X_3	0.53	0.26	0.05	0.21
Σ	1.00	0.31	0.29	0.40

The initial composition accounted for the 37% of incoming materials deriving from waste, whereas the newly identified composition accounted for the 60% of incoming materials deriving from waste, still providing a suitable substrate for fungal growth. This is noteworthy in the context of circular economy.

7.3.3. Optimization of the QuEChERS procedure

In order to evaluate the degradative capability of the fungal species towards the investigated dyes, a sample preparation step had to be involved due to the high complexity of the matrix. QuEChERS extraction [46] was selected among possible other sample preparation techniques as the typical sample amount to be processed is comparable to the sample availability and it allows for extraction and clean-up in a relatively short time.

In this study, the performance of QuEChERS extraction was evaluated by varying the extraction salts mixtures and the sorbent for dSPE. Different extraction salts

have an influence on the ionic strength and on the final pH of the extract, therefore attention has to be paid when analytes are pH sensitive.

The three different validated variants of the QuEChERS method involve an unbuffered (UB) and two buffered methods, one with acetic buffer (AOAC method) and one with citrate buffer (EN method). In the meanwhile, different dSPE sorbents aid the removal of different interfering compounds:

- Primary-secondary amine (PSA) aids the removal of acidic interferences.
- Octadecylsilane (C18) aids the removal of fats and nonpolar interferences.
- Graphitized Carbon Black (GCB) aids the removal of planar pigments, such as chlorophyll.
- Z-Sep is a proprietary phase by Supelco based on zirconium oxide that aids the removal of polar interferences.

In this study, three combinations of sorbents were evaluated, i.e., PSA/C18 + MgSO₄, Z-Sep/C18, and PSA/C18/GCB + MgSO₄.

In real permanent hair dye formulations, multiple primary dyes may be present to achieve the desired shade of color. In the context of this study, and in agreement with Davines S.p.A., three different combinations of primary dye and couplers were evaluated as model compounds.

It must be stated that separative techniques such as HPLC–DAD and LC–MS would be more appropriate for this kind of analytical problem. In this stage of the research UV-vis spectroscopy was utilized to evaluate the effectiveness of the QuEChERS extraction. This was carried out as extracts deriving from salts/sorbent combinations producing potentially a poor-clean up would cause high backpressure or clogging.

The preliminary evaluations were carried out *via* UV-vis spectroscopy was utilized Figure 7.11 reports the UV-vis spectra of the selected components.

The experiments were planned according to an unreplicated two-ways ANOVA, meaning that all the possible combinations salts/sorbents were explored. The purified blanks showed a red coloration, more or less intense depending on the buffer and the dSPE sorbent utilized. In addition, particles in suspension were noticed, depending on the dSPE sorbent utilized.

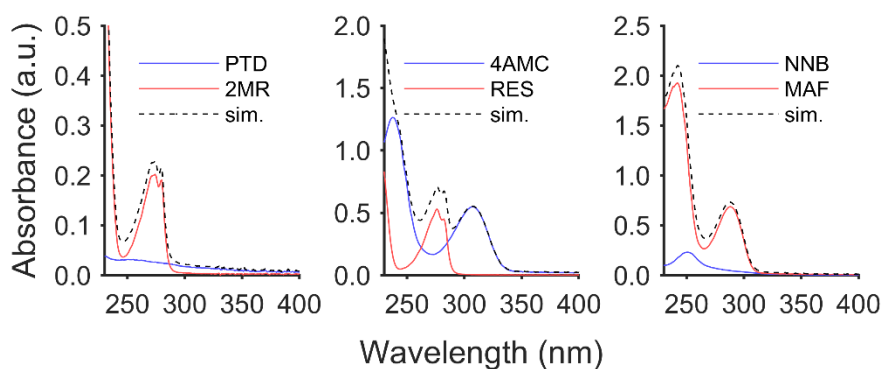


Figure 7.11. UV-vis spectra of PTD + 2MR (left), 4AMC + RES (middle), NNB + MAF (right) in the 230–400 nm region (concentration: 270 μ M; solvent: acetonitrile). Primary dyes are reported as blue lines, and couplers are reported as red lines, whereas simulated spectra of the respective mixtures are reported as black dashed lines.

As shown in Figure 7.12, both factors had a negligible effect on the extraction of the investigated dyes ($p > 0.05$). The only exception was represented by the dyes 4AMC + RES, for whom the buffer had a significant effect ($p < 0.05$; Cohen's $\eta^2 \approx 0.67$). Pairwise comparisons highlighted that for both compounds, the UB and EN methods provided a statistically significant difference ($p < 0.05$, Bonferroni adjusted), with the UB method providing the best performance for both.

The reasons behind this behavior could find an explanation in the salification of the amino group in 4AMC making it much more soluble in water with respect to acetonitrile, and, in general, in the higher background spectra that were registered when the extraction was carried out by applying the EN method. In fact, the blanks processed with the EN method were the ones showing the more intense red coloration. Therefore, the UB method was selected as the most promising one.

Despite no significance differences were detected in terms of dSPE sorbent, it was noticed that the blanks processed with PSA/C18/GCB + MgSO_4 showed the lightest red pigmentations and were less opalescent than the others. Therefore, the PSA/C18/GCB + MgSO_4 was selected as the most promising sorbent for dSPE clean-up.

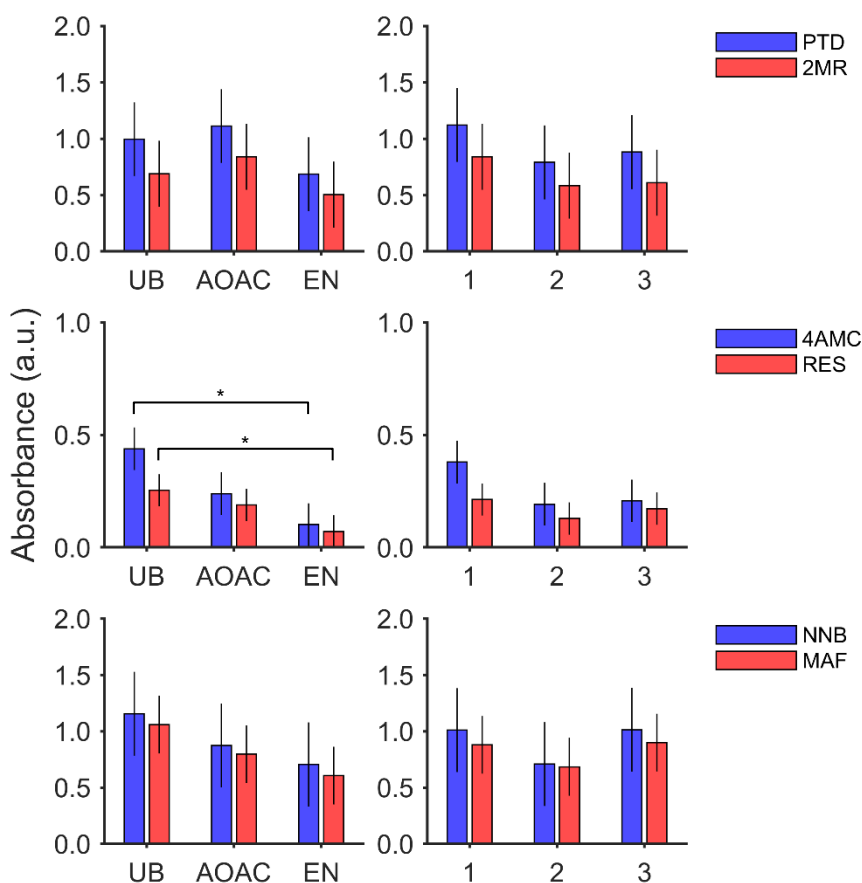


Figure 7.12. Summary of two-way ANOVA after background correction. Left: effect of the buffer (UB: unbuffered; AOAC: acetic buffer; EN: citrate buffer). Right: effect of the dSPE clean-up (1: PSA/C18 + MgSO₄; 2: Z-Sep/C18; 3: PSA/C18/GCB + MgSO₄). The results are reported as means with respect to each treatment for the three investigated combinations of dyes: PTD + 2MR (top), 4AMC + RES (center), NNB + MAF (bottom). Absorbance values were recorded at the wavelengths reported in Table 7.3. Primary dyes are reported in blue, and couplers are reported in red. The error bars show the amplitude of the 95% confidence interval estimated with respect to the residual variance.

7.3.4. Evaluation of the degradative capability

As for the optimization of QuEChERS extraction, also this step was carried out preliminarily by analyzing the extracts with UV-vis spectroscopy. Two experimental conditions were tested:

- **Control:** no mycelium was added in the preparation of the culture. This experimental condition had the purpose of providing information regarding possible changes with respect to oxidative processes with consequent formation of the pigmented complex due to the oxygen present

in air. A set of experiments with blank dye were prepared for background correction.

- **Treatment:** mycelium was added in the preparation of the culture. This experimental condition had the purpose of providing information regarding the degradative capabilities of *Fungi* towards the dyes. A set of experiments with blank dye were prepared for background correction.

The cultures were incubated for 1 h and for 7 days. For each experimental conditions, independent duplicated experiments were carried out, blanks included. These initial evaluations were carried out only with species A and by using PTD + 2MR as model compounds. The preliminary results are summarized in Figure 7.13.

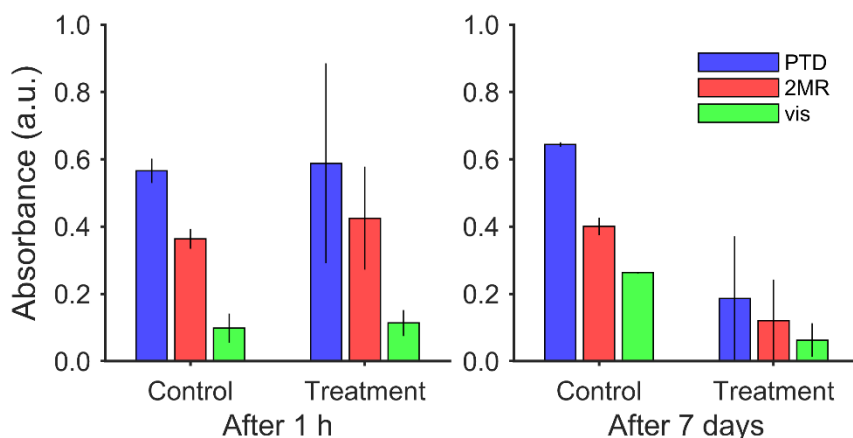


Figure 7.13. Summary of the preliminary evaluation of the degradative capabilities of species A. Left: absorbance values after 1 h of incubation. Right: absorbance values after 7 days of incubation. The results are reported as mean \pm standard deviation ($n = 2$). Absorbance values were recorded at the wavelengths reported in Table 7.3, whereas the one in the visible region was recorded at 414 nm. Primary dyes are reported in blue, couplers are reported in red, and the band observed in the visible region is depicted in green.

High variability was observed for the sample subjected to the fungal treatment. By speaking in terms of average absorbance only, it was noticed that for PTD and 2MR similar results were obtained for both the control and the treatment after 1 h of incubation and for the control after 7 days of incubation. After 7 days of incubation, samples subjected to the fungal treatment experienced an attenuation of the signals with respect to the ones subjected to the fungal treatment after 1 h of incubation. This attenuation was about 1.6 and 2.0 standard deviations [52] for PTD and 2MR, respectively.

A band around 414 nm was noticed, probably ascribable to the reaction abduct between PTD and 2MR. The intensity of this band increased for the control during

the 7 days of incubation, probably related to the initiation of the reaction due to the oxygen present in the air. As for the samples subjected to the fungal treatment, a decrease of 1.1 standard deviations [52] was observed. This result could be an indication of the capability of the fungal species in degrading also the reaction adducts given by the precursors commonly utilized in permanent hair dyes.

These experiments will be repeated with more replicates in the near future to confirm or disprove what these initial observations. Further investigation will be carried out in order to identify the compounds responsible for the absorption band observed in the visible region.

7.4. Conclusions

This study, carried out in collaboration with Davines S.p.A., aims at the production of secondary packaging materials based on mycelium and to waste materials, such as cardboard and hair dyes, by following the principles of circular economy. So far, the composition of the growth medium was optimized through the experimental design methodology. The conditions that were identified allowed the preparation of cultures that involved a total of 60% w/w of waste materials, and the composition was still suitable for fungal growth. QuEChERS extraction was tuned in terms of buffer salts and dSPE sorbent allowing for the recovery of six dyes selected as model compounds for the future evaluation of the degradative capabilities of *Fungi* towards hair dye formulations. A preliminary evaluation of the degradative capabilities of species A towards PTD and 2MR was carried out, obtaining promising results. In this context, more experiments will be carried out to confirm or avert the preliminary results. In addition, mixture design will be applied to evaluate the potential synergistic effects of the three species in degrading dyes contained in hair products, as well as the mechanical properties of the materials that will be produced after thermal treatment. Further method development will be carried out by involving liquid chromatography, potentially hyphenated to high-resolution mass spectrometry, for a more thorough evaluation of the degradative capabilities and for the investigation of the degradation products.

Note of the author

The results presented in this **Chapter** are matter of ongoing research.

The author would like to express gratitude to Erika Ribezzi for participating to this study as a part of the internship required to obtain the Master of Science degree.

This research project is financially supported by Davines S.p.A.

References

- [1] G.T. Corrêa, J.C. de Souza, J.P. Silva, M.I. Pividori, M.V.B. Zanoni, Determination of temporary dye Basic Red 51 in commercial hair dye, river water and wastewater from hairdressing salon using graphite-epoxy composite electrode modified with magnetic nanoparticles, *Microchemical Journal*. 159 (2020) 105485. <https://doi.org/10.1016/J.MICROC.2020.105485>.
- [2] S.A. da França, M.F. Dario, V.B. Esteves, A.R. Baby, M.V.R. Velasco, Types of Hair Dye and Their Mechanisms of Action, *Cosmetics*. 2 (2015) 110–126. <https://doi.org/10.3390/COSMETICS2020110>.
- [3] P. Ghosh, A.K. Sinha, Hair Colors: Classification, Chemistry and a Review of Chromatographic and Electrophoretic Methods for Analysis, *Anal Lett*. 41 (2008) 2291–2321. <https://doi.org/10.1080/00032710802352605>.
- [4] J. Manning, The Sociology of Hair: Hair Symbolism Among College Students, *Social Science Journal*. (2010) 35–48. <https://repository.wcsu.edu/cgi/viewcontent.cgi?article=1072&context=ssj> (accessed December 9, 2022).
- [5] C.R. Robbins, *Chemical and Physical Behavior of Human Hair*, 2nd ed., Springer New York, New York, USA, 1988. <https://doi.org/10.1007/978-1-4757-2009-9>.
- [6] J.A. Wenninger, Gerald N. McEwen, *International cosmetic ingredient dictionary*, Cosmetic, Toiletry and Fragrance Association, Washington, USA, 1993.
- [7] M.Y.M. Wong, The Kinetics of Dye Rinse, *J Soc Cosmet Chem*. 25 (1972) 165–170.
- [8] C. Bouillon, J. Wilkinson, eds., *The Science of Hair Care*, 1st ed., CRC Press, Boca Raton, USA, 2005. <https://doi.org/10.1201/b14191>.
- [9] A.O. Barel, H.I. Maibach, eds., *Handbook of Cosmetic Science and Technology*, 1st ed., CRC Press, Boca Raton, USA, 2001. <https://doi.org/10.1201/9780824741396>.
- [10] S. Dey, F. Bano, A. Malik, Pharmaceuticals and personal care product (PPCP) contamination—a global discharge inventory, *Pharmaceuticals and Personal Care Products: Waste Management and Treatment Technology Emerging Contaminants and Micro Pollutants*. (2019) 1–26. <https://doi.org/10.1016/B978-0-12-816189-0.00001-9>.
- [11] Ş. Sungur, Pharmaceutical and personal care products in the environment: occurrence and impact on the functioning of the ecosystem, *Emerging Contaminants in the Environment: Challenges and Sustainable Practices*. (2022) 137–157. <https://doi.org/10.1016/B978-0-323-85160-2.00009-3>.
- [12] J.C. de Souza, B.F. da Silva, D.A. Morales, G. de A. Umbuzeiro, M.V.B. Zanoni, Assessment of p-aminophenol oxidation by simulating the process of hair dyeing and occurrence in hair salon wastewater and drinking water from treatment plant, *J Hazard Mater*. 387 (2020) 122000. <https://doi.org/10.1016/J.JHAZMAT.2019.122000>.
- [13] A. Gogoi, P. Mazumder, V.K. Tyagi, G.G. Tushara Chaminda, A.K. An, M. Kumar, Occurrence and fate of emerging contaminants in water environment: A review, *Groundw Sustain Dev*. 6 (2018) 169–180. <https://doi.org/10.1016/J.GSD.2017.12.009>.
- [14] R. Al-Tohamy, S.S. Ali, F. Li, K.M. Okasha, Y.A.G. Mahmoud, T. Elsamahy, H. Jiao, Y. Fu, J. Sun, A critical review on the treatment of dye-containing wastewater: Ecotoxicological and health concerns of textile dyes and possible remediation approaches for environmental safety, *Ecotoxicol Environ Saf*. 231 (2022) 113160. <https://doi.org/10.1016/J.ECOENV.2021.113160>.
- [15] M. Berradi, R. Hsissou, M. Khudhair, M. Assouag, O. Cherkaoui, A. el Bachiri, A. el Harfi, Textile finishing dyes and their impact on aquatic environs, *Heliyon*. 5 (2019) e02711. <https://doi.org/10.1016/J.HELIYON.2019.E02711>.

- [16] E. de Raús Maúre, G. Terauchi, J. Ishizaka, N. Clinton, M. DeWitt, Globally consistent assessment of coastal eutrophication, *Nature Communications* 2021 12:1. 12 (2021) 1–9. <https://doi.org/10.1038/s41467-021-26391-9>.
- [17] M. Ajaz, S. Shakeel, A. Rehman, Microbial use for azo dye degradation—a strategy for dye bioremediation, *International Microbiology*. 23 (2020) 149–159. <https://doi.org/10.1007/S10123-019-00103-2>.
- [18] J.B. Morton, Fungi, in: T.J. Gentry, J.J. Fuhrmann, D.A. Zuberer (Eds.), *Principles and Applications of Soil Microbiology*, 3rd ed., Elsevier, Amsterdam, Netherlands, 2021: pp. 149–170. <https://doi.org/10.1016/B978-0-12-820202-9.00006-X>.
- [19] J.W. Deacon, *Fungal biology*, 4th ed., Blackwell Pub, Malden, USA, 2005.
- [20] M.A. Ruggiero, D.P. Gordon, T.M. Orrell, N. Bailly, T. Bourgoin, R.C. Brusca, T. Cavalier-Smith, M.D. Guiry, P.M. Kirk, A Higher Level Classification of All Living Organisms, *PLoS One*. 10 (2015) e0119248. <https://doi.org/10.1371/JOURNAL.PONE.0119248>.
- [21] E.H. Song, J. Shang, D.M. Ratner, Polysaccharides, in: K. Matyjaszewski, M. Möller (Eds.), *Polymer Science: A Comprehensive Reference*, Elsevier, Amsterdam, Netherlands, 2012: pp. 137–155. <https://doi.org/10.1016/B978-0-444-53349-4.00246-6>.
- [22] J. Dikec, A. Olivier, C. Bobée, Y. D’Angelo, R. Catellier, P. David, F. Filaine, S. Herbert, C. Lalanne, H. Lalucque, L. Monasse, M. Rieu, G. Ruprich-Robert, A. Véber, F. Chapeland-Leclerc, E. Herbert, Hyphal network whole field imaging allows for accurate estimation of anastomosis rates and branching dynamics of the filamentous fungus *Podospora anserina*, *Sci Rep*. 10 (2020) 1–16. <https://doi.org/10.1038/s41598-020-57808-y>.
- [23] T. Klein, R.T.W. Siegwolf, C. Körner, Belowground carbon trade among tall trees in a temperate forest, *Science* (1979). 352 (2016) 342–344. https://doi.org/10.1126/SCIENCE.AAD6188/SUPPL_FILE/KLEIN-SM.PDF.
- [24] M.G.A. van der Heijden, Underground networking, *Science* (1979). 352 (2016) 290–291. https://doi.org/10.1126/SCIENCE.AAF4694/ASSET/D5F16707-5D46-47AE-A022-942F094F907A/ASSETS/GRAPHIC/352_290_F2.JPEG.
- [25] C.M. Lee, B. van Geel, W.D. Gosling, On the Use of Spores of Coprophilous Fungi Preserved in Sediments to Indicate Past Herbivore Presence, *Quaternary*. 5 (2022) 30. <https://doi.org/10.3390/QUAT5030030>.
- [26] A. Bell, *Dung fungi: an illustrated guide to coprophilous fungi in New Zealand*, Victoria University Press, Wellington, New Zealand, 1983. https://books.google.com/books?hl=it&lr=&id=7XJQxILUvm4C&oi=fnd&pg=PA8&ots=wRZICND7If&sig=0_DQ9zUIN8WnEsmFD9KADSLY-Wg (accessed December 13, 2022).
- [27] T.N. Taylor, E.L. Taylor, M. Kings, Fungi, Bacteria, and Lichens, in: E.L. Taylor, T.N. Taylor, M. Kings (Eds.), *Paleobotany: The Biology and Evolution of Fossil Plants*, 2nd ed., Academic Press, Cambridge, USA, 2009: pp. 71–119.
- [28] M. Tekere, A.Y. Mswaka, R. Zvauya, J.S. Read, Growth, dye degradation and ligninolytic activity studies on Zimbabwean white rot fungi, *Enzyme Microb Technol*. 28 (2001) 420–426. [https://doi.org/10.1016/S0141-0229\(00\)00343-4](https://doi.org/10.1016/S0141-0229(00)00343-4).
- [29] S.K. Sen, S. Raut, P. Bandyopadhyay, S. Raut, Fungal decolouration and degradation of azo dyes: A review, *Fungal Biol Rev*. 30 (2016) 112–133. <https://doi.org/10.1016/J.FBR.2016.06.003>.
- [30] A. Pandi, G. Marichetti Kuppaswami, K. Numbi Ramudu, S. Palanivel, A sustainable approach for degradation of leather dyes by a new fungal laccase, *J Clean Prod*. 211 (2019) 590–597. <https://doi.org/10.1016/J.JCLEPRO.2018.11.048>.
- [31] P. Kaushik, A. Malik, Fungal dye decolourization: Recent advances and future potential, *Environ Int*. 35 (2009) 127–141. <https://doi.org/10.1016/J.ENVINT.2008.05.010>.
- [32] B. Viswanath, B. Rajesh, A. Janardhan, A.P. Kumar, G. Narasimha, Fungal laccases and their applications in bioremediation, *Enzyme Res*. 2014 (2014). <https://doi.org/10.1155/2014/163242>.

- [33] A.C. Sousa, L.O. Martins, M.P. Robalo, Laccases: Versatile Biocatalysts for the Synthesis of Heterocyclic Cores, *Molecules*. 26 (2021) 3719. <https://doi.org/10.3390/MOLECULES26123719>.
- [34] I. Bassanini, E.E. Ferrandi, S. Riva, D. Monti, Biocatalysis with Laccases: An Updated Overview, *Catalysts*. 11 (2020) 26. <https://doi.org/10.3390/CATAL11010026>.
- [35] M. Bilal, T. Rasheed, F. Nabeel, H.M.N. Iqbal, Y. Zhao, Hazardous contaminants in the environment and their laccase-assisted degradation – A review, *J Environ Manage*. 234 (2019) 253–264. <https://doi.org/10.1016/j.jenvman.2019.01.001>.
- [36] C.F. Thurston, The structure and function of fungal laccases, *Microbiology (N Y)*. 140 (1994) 19–26. <https://doi.org/10.1099/13500872-140-1-19/CITE/REFWORKS>.
- [37] J.P. Kallio, C. Gasparetti, M. Andberg, H. Boer, A. Koivula, K. Kruus, J. Rouvinen, N. Hakulinen, Crystal structure of an ascomycete fungal laccase from *Thielavia arenaria* – common structural features of asco-laccases, *FEBS J*. 278 (2011) 2283–2295. <https://doi.org/10.1111/j.1742-4658.2011.08146.x>.
- [38] S.M. Jones, E.I. Solomon, Electron transfer and reaction mechanism of laccases, *Cellular and Molecular Life Sciences*. 72 (2015) 869–883. <https://doi.org/10.1007/S00018-014-1826-6/FIGURES/12>.
- [39] C. Girometta, A.M. Picco, R.M. Baiguera, D. Dondi, S. Babbini, M. Cartabia, M. Pellegrini, E. Savino, Physico-Mechanical and Thermodynamic Properties of Mycelium-Based Biocomposites: A Review, *Sustainability*. 11 (2019) 281. <https://doi.org/10.3390/SU11010281>.
- [40] R. Abhijith, A. Ashok, C.R. Rejeesh, Sustainable packaging applications from mycelium to substitute polystyrene: a review, *Mater Today Proc*. 5 (2018) 2139–2145. <https://doi.org/10.1016/j.matpr.2017.09.211>.
- [41] J. Jose, K.N. Uvais, T.S. Sreenadh, A. v. Deepak, C.R. Rejeesh, Investigations into the Development of a Mycelium Biocomposite to Substitute Polystyrene in Packaging Applications, *Arab J Sci Eng*. 46 (2021) 2975–2984. <https://doi.org/10.1007/S13369-020-05247-2/TABLES/1>.
- [42] S. Sivaprasad, S.K. Byju, C. Prajith, J. Shaju, C.R. Rejeesh, Development of a novel mycelium biocomposite material to substitute for polystyrene in packaging applications, *Mater Today Proc*. 47 (2021) 5038–5044. <https://doi.org/10.1016/j.matpr.2021.04.622>.
- [43] K. Joshi, M.K. Meher, K.M. Poluri, Fabrication and Characterization of Bioblocks from Agricultural Waste Using Fungal Mycelium for Renewable and Sustainable Applications, *ACS Appl Bio Mater*. 3 (2020) 1884–1892. <https://doi.org/10.1021/ACSABM.9B01047>.
- [44] S. Yamanaka, R. Kikuchi, Complex of fibers and fungi and a process for preparation thereof, 5,074,959, 1991.
- [45] G. Derringer, R. Suich, Simultaneous Optimization of Several Response Variables, *Journal of Quality Technology*. 12 (1980) 214–219. <https://doi.org/10.1080/00224065.1980.11980968>.
- [46] M. González-Curbelo, B. Socas-Rodríguez, A. v. Herrera-Herrera, J. González-Sálamo, J. Hernández-Borges, M. Rodríguez-Delgado, Evolution and applications of the QuEChERS method, *TrAC Trends in Analytical Chemistry*. 71 (2015) 169–185. <https://doi.org/10.1016/j.trac.2015.04.012>.
- [47] M. Anastassiades, S.J. Lehotay, D. Štajnbaher, F.J. Schenck, Fast and easy multiresidue method employing acetonitrile extraction/partitioning and “dispersive solid-phase extraction” for the determination of pesticide residues in produce, *J AOAC Int*. 86 (2003) 412–431. <https://doi.org/10.1093/jaoac/86.2.412>.
- [48] J. Cohen, Eta-squared and partial eta-squared in fixed factor ANOVA designs, *Educ Psychol Meas*. 33 (1973) 107–112. <https://doi.org/10.1177/001316447303300111>.
- [49] C. Novoa, G. v. Dhoke, D.M. Mate, R. Martínez, T. Haarmann, M. Schreiter, J. Eidner, R. Schwerdtfeger, P. Lorenz, M.D. Davari, F. Jakob, U. Schwaneberg, KnowVolution of a Fungal Laccase toward Alkaline pH, *ChemBioChem*. 20 (2019) 1458–1466. <https://doi.org/10.1002/CBIC.201800807>.

- [50] T. Lundstedt, E. Seifert, L. Abramo, B. Thelin, A. Nystrom, J. Pettersen, R. Bergman, Experimental Design and Optimization, *Chemometrics and Intelligent Laboratory Systems*. 42 (1998) 3–40. <https://doi.org/10.1007/978-3-540-49148-4-3>.
- [51] D.P. di Lonardo, A. van der Wal, P. Harkes, W. de Boer, Effect of nitrogen on fungal growth efficiency, *Plant Biosyst.* 154 (2020) 433–437. https://doi.org/10.1080/11263504.2020.1779849/SUPPL_FILE/TPLB_A_1779849_SM3872.DOCX.
- [52] S. Vandekar, R. Tao, J. Blume, A Robust Effect Size Index, *Psychometrika*. 85 (2020) 232–246. <https://doi.org/10.1007/S11336-020-09698-2/FIGURES/4>.

Chapter 8 | Remarks and perspectives

This Thesis, in the framework of the PhD Program in Materials Science and Technology, aimed at exploring the potential of novel materials for applications in the field of Analytical Chemistry, with particular attention towards packaging, and bioremediation technologies.

In **Chapter 3**, the capabilities of four typologies of carbon nanotubes (CNTs) as coating materials for solid-phase microextraction (SPME) were evaluated, with the aim of quantifying a pool of 12 personal care products, in environmental water samples. A preliminary evaluation showed that the performance of the selected material, i.e., helical multi-walled carbon nanotubes was superior to the other investigated CNTs. This allowed the development and validation of a SPME–GC–MS suitable for the detection of the investigated analytes at trace and ultra-trace levels. The developed SPME fiber showed enrichment capabilities, on average, 6 standard deviations higher than those achieved with commercially available coatings, thus proving its reliability for the extraction of environmental pollutants at trace levels from water samples. Given the low detection and quantitation limits achieved in the study, the devised method will be applied for monitoring the investigated analytes in samples taken from remote regions of the planet, such as polar regions, to provide chemical records about the anthropic impact on the environment. Future studies will investigate the adsorption capabilities of carbon nanotubes towards other classes of analytes, potentially with applications for different sample preparation techniques.

Chapter 4 was devoted to the rational design of a composite magnetic sorbent based on the Metal-Organic Framework PUM198 to be applied as novel material for miniaturized sample treatment techniques with regards towards dispersive micro solid-phase extraction (MD- μ SPE). In this case, the determination of polycyclic aromatic hydrocarbons in water samples was the main goal of the research activity. The extraction procedure was thoroughly studied by applying an experimental design to study the effect of the elution solvent, the effect of a multitude of factors on the recovery of the analytes, and the effect of the most critical ones. The MD- μ SPE–GC–MS method was validated, achieving detection limits in the low ng/L range. The noteworthy achievement was the drastic reduction in sample size and the low solvent consumption with respect to conventional sample treatment techniques such as solid-phase extraction: the proposed methodology allowed the treatment of aliquots of 5 mL of sample consuming only 50 μ L of extraction solvent. Taking advantage of the low sample consumption, further development will be

focused in evaluating the reliability of the proposed method for the analysis of clinical samples, like oral fluid, plasma, and urine.

Regarding the bioremediation technology proposed in **Chapter 5**, the results achieved in this study showed that the incorporation of hydrochar into contaminated sediments produced a reduction in the content of the examined micropollutants, i.e., BTEX and styrene. Cumene and durene, were, apparently, more resistant to the treatment. Additionally, the cultivation of *V. spiralis* had a depolluting effect, although smaller in magnitude. Further investigation will be carried out to clarify the interaction between *V. spiralis* and the microbial communities in the removal of micropollutants from contaminated sediments. In the near future, the study is going to involve the validation of methodologies based on GC-MS and ICP-OES to evaluate the efficacy of the proposed technology in removing other classes of pollutants usually found in the Mantuan Site of National Interest, including polycyclic aromatic hydrocarbons, linear hydrocarbons, and metals.

Owing to the pivotal importance of a rational design of novel materials, in **Chapter 6** a PLS-DA model was successfully trained to predict the formation of binary cocrystals of active components commonly found in essential oils and other Generally Recognized As Safe molecules based on their molecular descriptors. Overall, the trained model showed a Non-Error Rate of 74% on an external test set. The result was more than satisfactory, especially taking into consideration the chemical diversity of the compounds included in the study. Hopefully, the results found in this research activity will provide guidance to cocrystal engineers in selecting suitable partner molecules for cocrystallization, reducing the efforts required for experimentation and paving the way towards a plethora of interesting applications in the field of Materials Science. Future studies will take advantage of the developed model to design cocrystals with potential applications in agrochemistry and in the development of innovative food packaging materials. In addition, by following the reported methodology, further studies could develop PLS-R models to predict other features of interest of such functional cocrystals, such as the melting point.

Finally, **Chapter 7** summarized the preliminary results of a study carried out in collaboration with Davines S.p.A., aimed at the fabrication of a biocomposite materials starting from waste deriving from the cosmetic industry. A Mixture Design was applied to optimize the growth of 3 species of *Fungi*, allowing the identification of experimental conditions able to reuse a total of 60% *w/w* of waste materials. A preliminary evaluation of the degradative capabilities of one of the

species towards two model compounds was carried out with UV-vis spectroscopy, obtaining promising results. Future studies will be carried out to investigate the potential synergistic effect of the three different species in degrading organic dyes, as well as to evaluate the mechanic properties of the biocomposite material resulting from fungal growth. Additionally, the development of methods of analysis based on HPLC-DAD and LC-HRMS will be object of future advancement, for a more thorough understanding of the degradation capabilities and for the characterization of the degradation products.

This page was intentionally left blank.

