

Characterization of microplastics and natural gas by infrared spectrometry and multivariate modelling

Author: Borja de la Fuente Ferreiro

Doctoral thesis by compendium of publications UDC
2023

Director: José Manuel Andrade Garda

Programa de doctorado en Ciencia y Tecnología Ambiental



UNIVERSIDADE DA CORUÑA



This work is licensed under a [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License](https://creativecommons.org/licenses/by-nc-nd/4.0/). This license does not apply to the aspects covered by the copyright transfers made to MDPI (article: “New Ways for the Advanced Quality Control of Liquefied Natural Gas”), ACS Publications (article: “Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas”) and Elsevier (articles: “Fast quality control of natural gas for commercial supply and transport utilities”, “Monitorization of polyamide microplastics weathering using attenuated total reflectance and microreflectance infrared spectrometry”, “Standardization of the minimum information for publication of infrared-related data when microplastics are characterized” and “Weathering-independent differentiation of microplastic polymers by reflectance IR spectrometry and pattern recognition”).

A mis compañeros, familia
y todos los que han hecho posible esta tesis.

Gracias.



Química Analítica
Departamento de Química
UNIVERSIDADE DA CORUÑA

D. JOSÉ MANUEL ANDRADE GARDA, Catedrático de Universidad de Química Analítica de la Universidad de A Coruña, Departamento de Química

CERTIFICA

Que la presente memoria, presentada como compendio de publicaciones, titulada “*Characterization of microplastics and natural gas by infrared spectrometry and multivariate modelling*” ha sido realizada por D. BORJA DE LA FUENTE FERREIRO bajo su dirección en el Departamento de Química (grupo de Química Analítica Aplicada, QANAP) de la Universidad de A Coruña, constituyendo su Tesis Doctoral para optar al grado de Doctor en Ciencia y Tecnología Ambiental por la Universidad de A Coruña con Mención Internacional.

Y para que así conste a los efectos oportunos, firmo la presente en A Coruña, a 15 de junio de 2023

Abstract

The main objective of this work is to explore the use of infrared spectrometry combined with the application of multivariate chemometric models to the quantification of the major components of natural gas and to the identification of plastic samples, both artificially aged and collected from coastal ecosystems. The articles presented here deal with different themes related with these topics. In particular: the application of IR-inert gases to improve the spectra of other gases of interest; a report evidencing a general absence of important specifications of the instrumental setup of environmental studies on microplastics, suggesting a minimum of information to be offered; and a study of the effects of aging in both the surface morphology and the spectral characteristics of polyamide 6.6. Finally, chemometric models have been developed to identify the main constituent polymers of microplastics and to quantify the major components of natural gas samples, as well as their Wobbe index.

Resumen

El principal objetivo de este trabajo es explorar el uso de la espectrometría infrarroja combinada con la aplicación de modelos quimiométricos multivariados para cuantificar de la composición mayoritaria de muestras de gas natural y para identificar muestras de plásticos, tanto envejecidos de forma artificial como recolectados de ecosistemas costeros. Los artículos presentados tratan sobre diversos aspectos relacionados con estos temas. En concreto: el uso de gases inertes a la radiación infrarroja para mejorar los espectros de otros gases de interés; un informe que evidenciando una importante ausencia de especificaciones instrumentales básicas en los estudios medioambientales sobre microplásticos, donde se sugiere un mínimo de información a aportar; y un estudio de los efectos del envejecimiento en la morfología y características espectrales de la poliamida 6.6. Finalmente, se han desarrollado modelos quimiométricos capaces de identificar los principales polímeros constituyentes de microplásticos y de cuantificar los componentes mayoritarios de muestras de gas natural, así como su índice Wobbe.

Resumo

O obxectivo principal deste traballo é explorar o uso da espectrometría infravermella combinada coa aplicación de modelos quimiométricos multivariabeis para a cuantificación da composición maioritaria de mostras de gas natural e a identificación de plástico, tanto envellecidos artificialmente como recollido de ecosistemas costeiros. Os diferentes artigos presentados tratan diversos aspectos relacionados con estes temas. Nomeadamente: a aplicación de gases inertes á radiación infravermella para mellorar os espectros doutros gases de interese; unha recompilación onde se constata a ausencia xeral especificacións instrumentais fundamentais nos estudos medioambientais de microplásticos, suxerindo un mínimo de información a aportar; e un estudo dos efectos do envellecemento na morfoloxía e características espectrais na poliamida 6.6. Finalmente, desenvolvéronse modelos quimiométricos capaces de identificar os polímeros constituíntes de microplásticos e de cuantificar os principais compoñentes de mostras de gas natural, ademais do seu índice de Wobbe.

Contents

List of abbreviations	1
Foreword	3
Objectives	5
Publications	7
Chapter 1: General Introduction.	11
1. Overview	13
2. Natural gas	15
2.1. Introduction	15
2.2. Global market	16
2.3. Sources and composition	16
2.4. Processing	18
2.5. Spectrometric measurements	20
3. Microplastics	21
3.1. Introduction	21
3.2. Environmental effects	24
3.3. Spectrometric measurements	27
4. IR Spectrometry	28
4.1. History	28
4.2. Fundamentals	30
4.3. State of the art	36
5. Chemometrics	38
5.1. Fundamentals	39
5.2. Variable selection methods	44

5.3.	Applications in natural gas analysis	48
5.4.	Applications in microplastic analysis	49
6.	Overview of the NG-related publications	52
7.	Overview of the microplastic-related publications	53
8.	General conclusion	54
9.	References	55
Chapter 2: Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas.		75
Chapter 3: New Ways for the Advanced Quality Control of Liquefied Natural Gas.		83
Chapter 4: Fast Quality Control of Natural Gas for Commercial Supply and Transport Utilities.		103
Chapter 5: Standardization of the Minimum Information for Publication of Infrared-related Data when Microplastics are Characterized.		131
Chapter 6: Monitorization of Polyamide Microplastics Weathering Using Attenuated Total Reflectance and Microreflectance Infrared Spectrometry.		149
Chapter 7: Weathering-independent Differentiation of Microplastic Polymers by Reflectance IR Spectrometry and Pattern Recognition.		165
Appendix: Summary in Spanish		197

List of abbreviations

1D-CNN: one-dimensional neural network.

ATR: attenuated total reflectance.

BDT: Boosted Decision Tree.

CART: Classification and Regression Trees.

CNN convolutional neural network.

DBSCAN: Density-based Spatial Clustering of Applications with Noise.

DEHP: di-(2-ethylhexyl)phthalate.

DT: Decision tree.

FTIR: Fourier-transform IR.

GC: gas chromatography.

GC-MS: gas chromatography-mass spectrometry.

HDPE: High-density polyethylene.

HIS: hyperspectral imaging.

HPLC: high-performance liquid chromatography.

iPLS: interval partial least squares.

IR: infrared.

LDA: Linear discriminant analysis.

LDIR: Laser Direct IR.

LDPE: Low-density polyethylene.

LIBS: Laser-induced breakdown spectroscopy.

MALDI-MS: Matrix-assisted laser desorption/ionisation mass spectroscopy.

MCR-ALS: Multivariate curve resolution-alternating least squares.

MCT: Mercury Cadmium Telluride.

MCUVE: Monte-Carlo uninformative variable elimination.

MD: Mahalanobis distance.

ML: maximum likelihood.

NG: natural gas.

NGL: natural gas liquids.

PCA: Principal component analysis.

PE: Polyethylene.

PET: Polyethylene terephthalate.

PLS-DA: Partial least squares discriminant analysis.

PLS-PM: Partial least squares path modelling.

PLS-R: Partial Least Squares regression.

PP: Polypropylene.

PUR: Polyurethane.

PUR: Polyurethane.

PVC: Polyvinylchloride.

QANAP: Grupo de Química Analítica Aplicada.

QCL: Quantum cascade lasers.

RBF: radial basis function.

RMSEC: Root-mean-square of calibration.

RMSECV: root-mean-square of cross-validation.

RMSEP: Root-mean-square of prediction.

S/N: Sygnal-Noise ratio.

SIMCA: soft independent modelling of class analogy.

SRI: Selectivity ratio index.

SVM: Support Vector Machines.

THAID: Theta Automatic Interaction Detection.

TRWP: Tyre and road wear particles.

UMAP: uniform manifold approximation and projection.

VIP: Variable Importance in Projection.

Foreword

Our Society faces analytical chemists with very diverse challenges. Some could be considered as more methodological, like the search for new sensors or analytical procedures, while others have a more pragmatic scope, like the quality control of industrial products.

In this Doctoral Thesis both aspects are considered, and the suggested solutions combine vibrational spectrometry and chemometric multivariate methods to process raw data into useful information, so that adequate decisions can be proposed. The two studied fields are very different, as a consequence of the various interests and projects in which the *Grupo de Química Analítica Aplicada* (QANAP) of the University of A Coruña is, or has been, involved with. Both are somehow related to the petrochemical industry and are environmentally relevant for different reasons:

- 1) The efficient monitoring of the quality and composition of natural gas for its use as an energy source. This is especially important in Galicia, having one of the most important regasification plants in Spain: Reganosa, located in Mugardos (Ferrol). The QANAP group has collaborated closely with them including a participation in the *LNG 3* (16ENG09, “*Metrological support for LNG and LBG as transport fuel*”) European project, within the EMPIR III program.
- 2) The study of the occurrence of microscopic polymeric fragments (microplastics) in environmental matrices is nowadays a relevant topic. However, their characterization in such complex matrices can be cumbersome, in particular because of the difficulties that their weathering poses in their identification. QANAP has been involved in several national and international projects that studied how to address this issue (*BASEMAN* (JPI Oceans PCIN2015-170C0202), *MICROPLASTIX* (PCI2020-112145), *LABPLAS* (H2020,101003954)).

These are the two analytical fields of study that shaped the main objectives of this Doctoral Thesis.

Objectives

The major objective of this Doctoral Thesis can be summarized as the implementation of reliable and validated analytical methods capable of offering useful information by combining mid-infrared spectrometry and several chemometric tools. The two fields of study and their particular working objectives are:

- a) Natural gas
 - To improve the intensity and sensitivity of the infrared spectra of natural gas samples by adding inert gases.
 - To develop chemometric regression models able to predict the chemical composition and the Wobbe index of different blends of commercial natural gas samples from their spectra.
- b) Microplastics:
 - To identify polymers in various weathering stages using chemometric classification models developed from their reflectance and ATR infrared spectra.
 - To propose a way to standardize published information regarding the mid-infrared instrumental setup employed for microplastic studies.
 - To study the effects of weathering on the spectra of polyamide 6.6 (or nylon 6.6).

Publications

This Doctoral Thesis presents the following articles (constituting Chapters 2 to 7), as they contain the most relevant results associated with the stated objectives:

Article 1: Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas.

Authors: Borja Ferreiro¹, José Manuel Andrade¹, Carlota Paz-Quintáns¹, Purificación López-Mahía¹, Soledad Muniategui-Lorenzo¹, María Rey-Garrote², Cristina Vázquez-Padín², Carlos Vales².

Affiliation:

¹ Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

² REGANOSA (Regasificadora del Noroeste, S.A.), Punta Promontoiro s/n, Mugardos, 15071 A Coruña, España.

Journal: Energy & Fuels

Impact factor JCR: 3.421 (2019) (Q2)

Reference (APA):

Ferreiro, B., Andrade, J. M., Paz-Quintáns, C., López-Mahía, P., Muniategui-Lorenzo, S., Rey-Garrote, M., Vázquez-Padín, C., Vales, C. (2019). Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas. *Energy & Fuels*, 33(8), 6929–6933. <https://doi.org/10.1021/acs.energyfuels.9b00549>

Article 2: New Ways for the Advanced Quality Control of Liquefied Natural Gas.

Authors: Borja Ferreiro, José Andrade, Carlota Paz-Quintáns, Purificación López-Mahía, Soledad Muniategui-Lorenzo.

Affiliation: Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

Journal: Energies

Impact factor JCR: 3.252 (2021) (Q3)

Reference (APA):

Ferreiro, B., Andrade, J., Paz-Quintáns, C., López-Mahía, P., Muniategui-Lorenzo, S. (2022). New Ways for the Advanced Quality Control of Liquefied Natural Gas. *Energies*, 15(1), 359. <https://doi.org/10.3390/en15010359>

Article 3: Fast Quality Control of Natural Gas for Commercial Supply and Transport Utilities.

Authors: Borja Ferreiro¹, José Manuel Andrade¹, Purificación López-Mahía¹, Soledad Muniategui-Lorenzo¹, Cristina Vázquez-Padín², Andrés Pérez, María Rey-Garrote², Carlos Vales².

Affiliation:

¹ Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

² REGANOSA (Regasificadora del Noroeste, S.A.), Punta Promontoiro s/n, Mugarbos, 15071 A Coruña, España.

Journal: Fuel

Impact factor JCR: 8.035 (2022) (Q1)

Reference (APA):

Ferreiro, B., Andrade, J., López-Mahía, P., Muniategui, S., Vázquez, C., Pérez, A., Rey, M., Vales, C. (2021). Fast quality control of natural gas for commercial supply and transport utilities. *Fuel*, 305. <https://doi.org/10.1016/j.fuel.2021.121500>

Article 4: Standardization of the Minimum Information for Publication of Infrared-Related Data when Microplastics are Characterized.

Authors: José Andrade, Borja Ferreiro, Purificación López-Mahía, Soledad Muniategui-Lorenzo

Affiliation: Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

Journal: Marine pollution bulletin

Impact factor JCR: 5.553 (2020) (Q1)

Reference (APA):

Andrade, J. M., Ferreiro, B., López-Mahía, P., Muniategui-Lorenzo, S. (2020). Standardization of the minimum information for publication of infrared-related data when microplastics are characterized. *Marine Pollution Bulletin*, 154 (February). <https://doi.org/10.1016/j.marpolbul.2020.111035>

Article 5: Monitorization of Polyamide Microplastics Weathering Using Attenuated Total Reflectance and Microreflectance Infrared Spectrometry.

Authors: Verónica Fernández-González, José Andrade, Borja Ferreiro, Purificación López-Mahía, Soledad Muniategui-Lorenzo

Affiliation: Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

Journal: Spectrochimica Acta Part A: Molecular and Biomolecular Spectrometry

Impact factor JCR: 4.831 (2021) (Q1)

Reference (APA):

Fernández-González, V., Andrade, J. M., Ferreiro, B., López-Mahía, P., Muniategui-Lorenzo, S. (2021). Monitorization of polyamide microplastics weathering using attenuated total reflectance and microreflectance infrared spectrometry. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 263, 120162. <https://doi.org/10.1016/j.saa.2021.120162>

Article 6: Weathering-Independent Differentiation of Microplastic Polymers by Reflectance IR Spectrometry and Pattern Recognition.

Authors: Borja Ferreiro, José M. Andrade, Carlota Paz-Quintáns, Verónica Fernández-González, Purificación López-Mahía, Soledad Muniategui.

Affiliation: Grupo de Química Analítica Aplicada (QANAP), Facultade de Ciencias, Universidade de A Coruña, Campus da Zapateira s/n, 15071 A Coruña, España.

Journal: Marine Pollution Bulletin

Impact factor JCR: 7.001 (2021) (Q1)

Reference (APA):

Ferreiro, B., Andrade, J. M., Paz-Quintáns, C., Fernández-González, V., López-Mahía, P., & Muniategui, S. (2022). Weathering-independent differentiation of microplastic polymers by reflectance IR spectrometry and pattern recognition. *Marine Pollution Bulletin*, 181(July), 113897. <https://doi.org/10.1016/j.marpolbul.2022.113897>

Chapter 1: General Introduction.



1. Overview

The common thread of the publications in this Doctoral Thesis is the use of infrared (IR) spectrometry in the medium region ($4000\text{-}400\text{ cm}^{-1}$) combined with multivariate regression methods for the analysis of complex samples. Some of them are related to an industrial setting, where this methodology is uncommon, while the others correspond to a relatively recent topic in environmental studies. Various IR techniques and instruments were used: “traditional” transmission Fourier-transform IR (FTIR) spectrometry and FTIR microscopy, in reflectance and attenuated total reflectance (ATR) modes.

The explored applications are:

- The determination of the main chemical components and the Wobbe index in different blends of natural gas (NG), as an alternative to gas chromatography (GC); which is a slower and more expensive method (Chapters 2, 3 and 4).
- The analysis of micro- and macroplastics, both in field samples and samples weathered in laboratory controlled conditions, simulating seawater and sunlight exposition (Chapters 5, 6 and 7).

Regarding NG, its analysis and monitoring is usually performed by GC. This is a well-established technique, with an ISO standard¹, showing good accuracy, and low limits of detection and quantification. However, it has a few drawbacks, like its high analysis time, the required staff dedication, the consumption of reagents, and its associated economic cost. IR spectrometry is, however, an interesting alternative, cheaper and faster. Its main disadvantage, however, is the high complexity of the resulting spectra, generally yielding higher errors than GC in the quantifications and the impossibility to predict minor components (pentanes and hexanes), because of their low concentrations in the samples. To mitigate this problem, the “gas broadening effect” was exploited, which establishes that after mixing a gaseous sample with a gas inert to IR radiation (like nitrogen or argon), the recorded spectra are more defined and more intense².

Chemometric models were developed to accurately predict the composition of the NG, as well as the Wobbe index, an important parameter that relates to the miscibility of blends of NG with different origins and/or compositions. Variable reduction methods were applied, greatly reducing the wavenumbers required to develop successful models.

In microplastic analysis, a collection of nine of the most common plastics (both in production and applications) were studied. They were specifically fabricated for the BASEMAN project³. The objectives are to develop chemical identification methods for microplastics using mid-IR spectrometry, especially for those encountered in marine environments, as well as to standardize the information required for an adequate publication of the experimental setup in this field. In addition, models were developed to differentiate the nine types of polymers using their IR spectra and several multivariate chemometric methods.

Reviewing the literature referring to microplastic analysis (up to 2020) a remarkable lack of homogeneity was observed in the description of the experimental sections. In some cases, important parameters were omitted, like the instrument or data processing used. This is probably due to the lack of standardized methods or well defined guidelines, for being a field of relatively recent creation and the arrival of many researchers with from different backgrounds with no experience in chemistry or spectrometry.

Finally, chemometric variable reduction methods were applied with the intention of reducing the necessary wavenumber range and suppressing redundant variables that can hinder the correct identification or quantification of the polymers.

2. Natural gas

2.1. Introduction

Natural gas (NG) is a colourless, odourless fossil gas used mainly as fuel for domestic and industrial heating, and as a reagent in the manufacturing of products like ethylene, ammonia (mainly for fertilizers), sulphur or black carbon. It is formed by the decomposition of biologic matter under extreme conditions of temperature and pressure, under the Earth's cortex, similarly to petroleum or coal. In fact, it is common for them to appear in the same or adjacent deposits. It can also be generated from volcanic activity.

It was first discovered in Sichuan, China, around 500 BC, but the first exploitations were performed by William Hart in Fredonia, USA in 1821, where it was used mainly as a light source⁴. Nowadays it is a viable and environmentally interesting alternative to other fossil fuels in many applications, as it burns with low emissions of CO₂, NO_x and SO_x. In **Figure 1** a comparison between the most common emissions from the combustion of fossil fuels per kiloJoule is shown⁵.

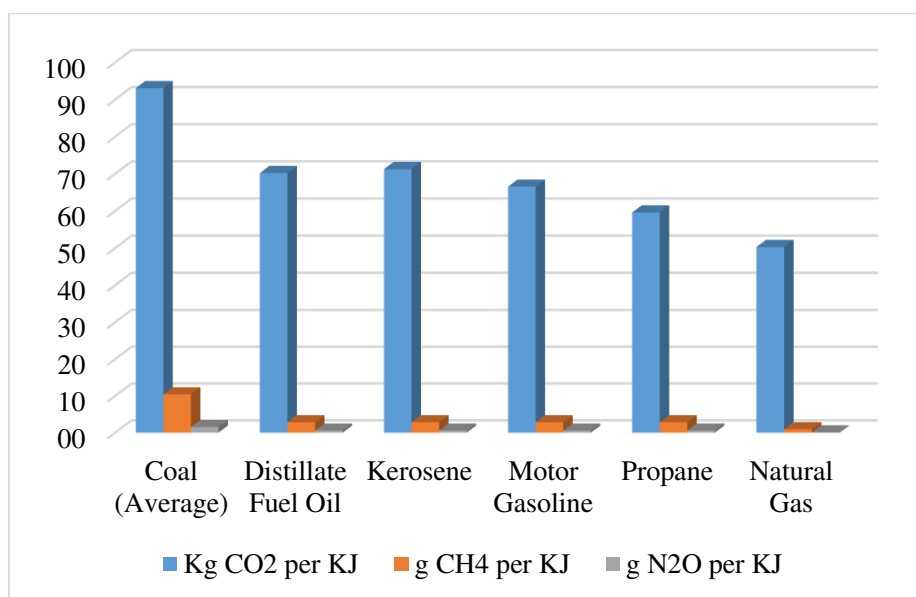


Figure 1: Common fossil fuel emissions per unit of energy (kiloJoule) generated in combustion⁵.

2.2. Global market

Global NG production has been steadily increasing for at least 70 years (with the exceptions of the 2008 economic crisis and the 2020 COVID pandemic). The largest estimated reserves of NG are currently located mainly in Asia and North America. Currently, the United States is the largest producer and consumer of NG with 1006 billion cubic meters produced in 2022, consuming almost 90% of it (ca. 887 billion cubic meters). Its low pollution impact and general abundance made it a suitable option for the EU decarbonisation plans⁶. However, the main exporter, mainly to Europe and China, is Russia. Since the onset of their war and the EU support to Ukraine, gas prices have increased significantly, as it has been used as a strategic resource, and the markets have become unstable. In response, the EU is steadily reducing its imports, with the intention of fully stopping them by 2027⁷.

2.3. Sources and composition

NG is found mainly in subterranean deposits, confined by impermeable stone, generally in a dome-shaped structure above the gas, which blocks its escape into the atmosphere. In these deposits NG is generated following one or several of three processes: thermogenesis, biogenesis and abiogenesis⁴.

NG produced by **thermogenic processes** (also denominated thermogenic methane) is generated by the high pressures and temperatures present in the Earth's cortex. Generally, this kind of deposits occur due to the transformation of biomass buried by sedimentation processes. Under these conditions, and in the absence of oxygen, organic matter is slowly decomposed, generating a kerogene residue that is accumulated in the sediments. This material can be degraded following several processes, depending on temperature, generating carbohydrates, as well as CO₂, H₂S, H₂ and N₂ and pyrobitumen (a mineraloid composed mainly of carbon).

Biogenic processes involve methanogenic bacteria in the degradation of organic matter. These bacteria are usually found in anoxic sediments

(anaerobic digesters, anoxic sediments or flooded soils) or in the gastrointestinal tract of the fauna. This process has different phases, which involve different kinds of microorganisms. During the *acidogenesis* stage, complex molecules are degraded into simpler products, like sugars, fatty acids or alcohols. In *acetogenesis*, those products are fermented, forming acetates, H₂ and CO₂. During *methanogenesis*, these products react to produce mainly methane⁴.

Unlike the previous cases, NG produced by **abiogenic processes** doesn't have organic matter as a precursor, but volcanic gases. These gases, rich in CO₂, are chemically reduced when the magma is cooled, notably when it interacts with seawater, generating NG.

Those formation processes are relevant, as they have an impact on the composition of the generated gas. Another critical factor is the type of deposit where NG is found: as an independent reservoir, associated with coal or petroleum (either dissolved or as two separate phases), enmeshed in low-permeability minerals, as a hydrate, etc.

NG is mainly composed of a mixture of organic gases in different concentrations: methane (75-99 %), ethane (0-20 %), propane, butanes, pentanes and hexanes (<10 %). It can also contain other eluent gases (CO₂, N₂, H₂, O₂, H₂O or noble gases), contaminants (Hg, As, or sulphur compounds) and solids in suspension.

During the NG industrial processing it is desirable to remove all these additional compounds, leaving only the lighter hydrocarbons, as they are the main source of energy for its combustion. The concentrations of the different components are important, as they affect the combustion properties. This is especially relevant in motors, as they are optimized for a specific mixture of gases. An inadequate composition can provoke *knocking* effects^{8,9}, i.e. an incomplete combustion of the fuel in the cylinders of the engine, generating high-pressure oscillations¹⁰. This does not only affect its power, but also generates vibrations, noise and can damage the engine, reducing its durability. This effect is measured by the Methane number, an analogous parameter to the octane number in gasoline or the cetane number in diesel, and it is an important characteristic to monitor, especially in vehicular applications¹¹.

Another important property of natural gas is the Wobbe index, a parameter that serves as a criterion for the interchangeability between combustion gases, especially relevant when blending natural gases with different origins and compositions. It can be interpreted as the heating value of the fuel that would flow through an orifice of a given size (as in a burner, for example) in a given amount of time¹². It varies with the composition of the blend and the presence of eluent gases like CO₂ and N₂. Typical values in commercial liquefied natural gas range between 13.1-15.2 kWh/Nm³.

2.4. Processing

NG cannot be used directly from the gas reservoir. It usually carries high amounts of contaminants or other products, generally associated with other types of fossil fuels, which need to be separated. The most common contaminants are⁴:

- Gaseous water, helium and nitrogen, which affect the calorific output of NG combustion, decreasing its efficiency as a fuel.
- In associated reservoirs it needs to be separated from petroleum and other fuels.
- Hydrogen sulphide and other sulphur-related compounds (well-known atmospheric pollutants).
- Mercury, a very toxic substance, which can also deteriorate some of the equipment of industrial plants.
- Higher hydrocarbons or natural gas liquids (NGL).
- Naturally occurring radioactive material, like radium-226 and -228, lead-210 or radon can accumulate in the storage tanks and processing units of industrial facilities, endangering exposed employees.

During processing, most contaminants are eliminated or reduced to an acceptable concentration and several products and by-products, like sulphur, mercury or different hydrocarbons can be obtained. **Figure 2**

offers a general idea of a NG processing workflow, with examples of some of the treatments that can be employed^{4,13-15}.

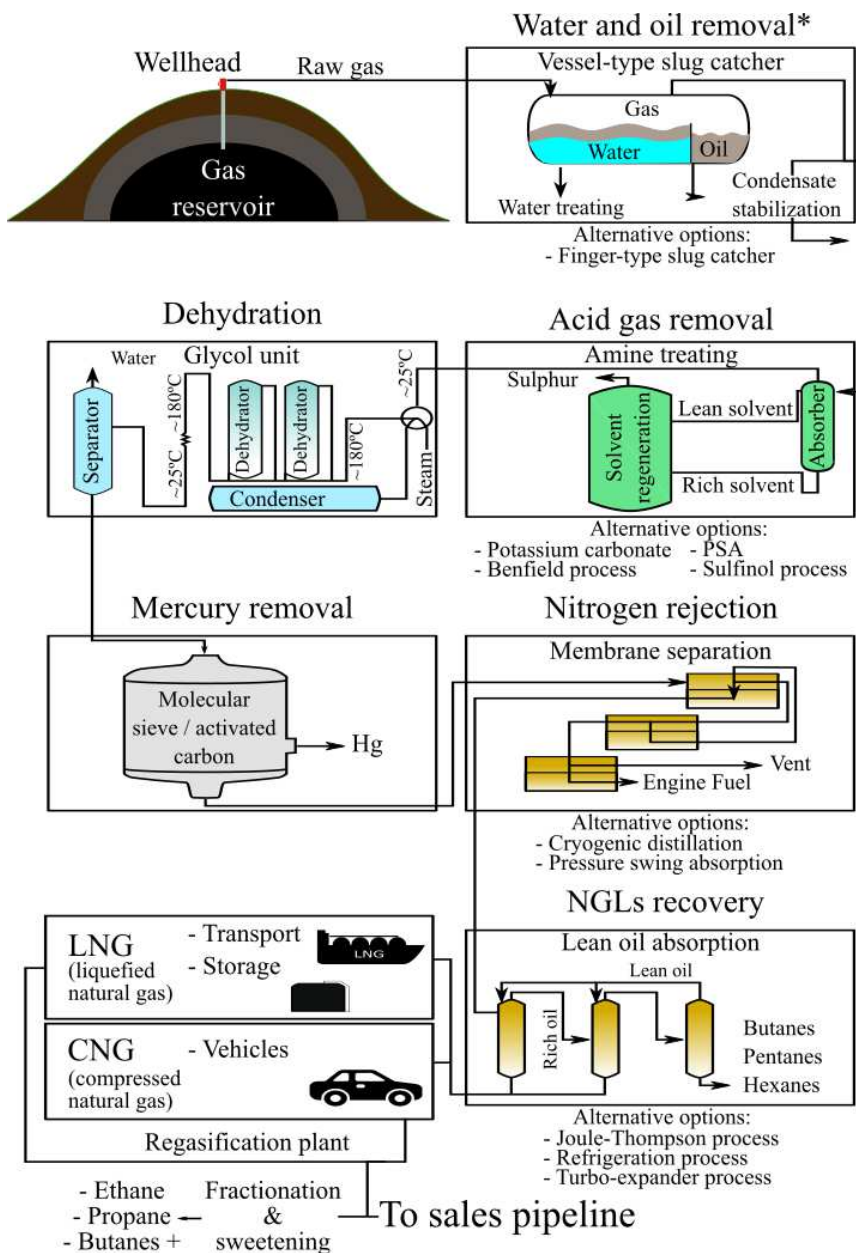


Figure 2: General workflow of NG industrial processing^{4,13-15}. *Optional, mainly used for associated reservoirs. PSA: Pressure swing absorption; NGL: Natural gas liquids

2.5. Spectrometric measurements

Traditionally, the most common technique for NG chemical characterization and monitoring is GC¹. Its high sensitivity allows for more thorough quantifications than spectrometry in its current state. However, GC instruments are usually voluminous, with high running costs (as they require more reagents, staff dedication, frequent calibration) and maintenance and are not very suitable for *on-line* systems (although dedicated micro-GC systems exist, they still need a rack of carrier gases close to the measuring site). For these reasons, IR spectrometry has become an interesting alternative, being able to reduce the analysis times from ~150 min to less than 30 min, with a negligible consumption of reagents, if any.

To improve the sensitivity of IR spectrometry when measuring gases, the “gas broadening” effect was exploited, a physical phenomenon discovered in the 1950s. When a sample gas is “dissolved” in another gas, invisible to IR radiation, its spectral features are magnified, yielding more intense spectra, not altering the results otherwise^{16,17}.

3. Microplastics

3.1. Introduction

“Plastic” is a general term that encompasses an extensive collection of materials used widely in modern Society. Their high stability, versatility, malleability, and low cost of production makes them very attractive for a large variety of applications^{18–20}. Plastics (polymers in general) are composed of long chains of hydrocarbons, in which every link is a “monomer”. They can be as simple as a repetition of an ethane molecule in polyethylene (PE), or have complex structures with heteroatoms and aromatic groups as in polyurethane (PUR). These monomers determine most of their physical properties, like physical and thermal resistance, density or flexibility.

Around 391 million tonnes of plastics were produced worldwide in 2021²¹, of which, around 44 % were destined to packaging, mostly single-use. In terms of production, the most prevalent polymers are PP (19.3 %) and PE (LDPE: 14.4 %; HDPE: 12.5 %). In **Figure 3** the most produced polymers worldwide are shown.

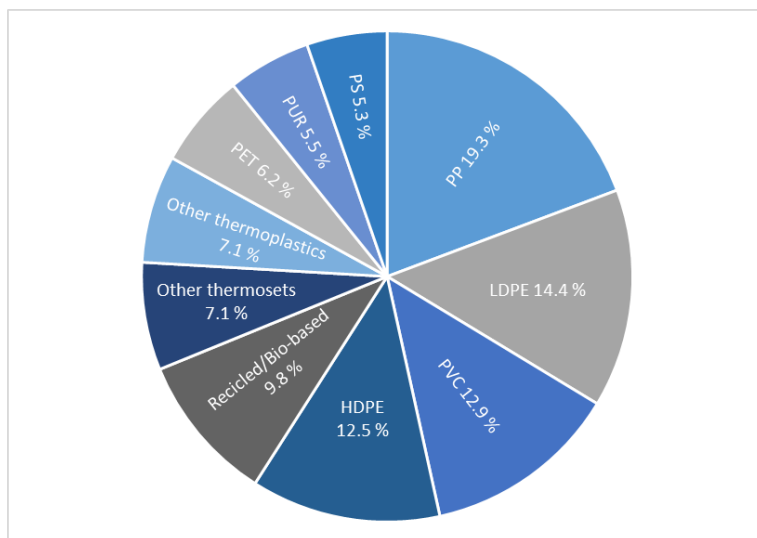

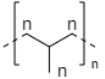


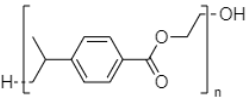
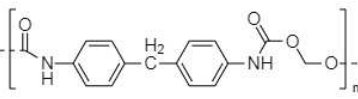
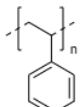
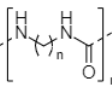

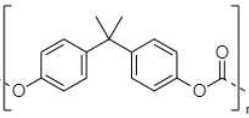
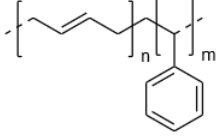
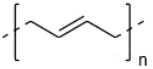


Figure 3: Global plastic production by polymer²¹.

Some of the most common plastics, their monomers and some examples of their commercial applications are summarized in **Table 1**.

Table 1: Most produced polymers, with their chemical structure and some examples of their applications.

Polymer	Structure	Applications
Polypropylene (PP)		Rope, nets, stationery supplies
Low-density polyethylene (LDPE)		Single use bags, electric cable insulation, straws, fishing nets
Polyvinyl chloride (PVC)		Food, household cleaning, medicine, hygiene products packaging
High-density polyethylene (HDPE)		Household cleaning and food packaging, bottles
Polyethylene terephthalate (PET)		Bottles and food packaging
Polyurethane (PUR)		Construction (sealants, insulators, firestoppers), vehicle parts, sponges, adhesives
Polystyrene (PS)		Disposable cutlery, food and medicine packaging, insulation
Polyamide (PA)		Textile, mechanical engineering supplies, medical applications
Polymethyl-methacrylate (PMMA)		Windows, bulkheads, screens
Polycarbonate (PC)		Electronic and construction materials, 3D printing
Styrene-butadiene rubber (SBR)		Tyres
Butadiene rubber (BR)		Tyres

Initially, in the 1930's and 40's, most plastics were developed and produced for military applications during WWII, ranging from parachutes and ropes to airplane windows. Having proved its applicability, their mass production began in the decade of the 1950s for commercial purposes, substituting other materials, like wood, steel, horn or glass for a large number of products²², and their demand has increased exponentially since then.

This huge plastic production, along with the fact that an important number of applications are single-use, generates a high quantity of residues. Many of them are not collected or recycled properly, ending up in domestic garbage, landfills or, simply, spilled into the environment. As a pollutant, plastics have some concerning properties: they degrade very slowly, often lasting for hundreds or thousands of years, which derives in a high prevalence in the environment, and can break down into "microplastics", which generate a new set of associated problems.

Specifically, microplastics (MPs) are defined by ISO as polymeric fragments or fibres with sizes between 5 mm and 1 μm ²³ and they can appear as primary or secondary microplastics. Primary microplastics are produced in certain industries, such as cosmetics or pharmaceutical (personal care and cosmetic products, like sunscreen, toothpaste, eyeshadow, etc.), and also for industrial abrasion/polishing. However, relevant measures are being taken to reduce them in personal care products in some countries, like UK, USA, Thailand or the EU²⁴. Also, it is common to find small pre-production pellets or "nurdles" in the environment, of around 3-5 mm diameter. Those are the fundamental constituents for most industrial applications where thermoplastics are used, by melting and casting them into different appliances. Accidents or mismanagement in storage or during transport²⁵⁻²⁷ can spill large quantities of these particles, often ending up in sediments and beaches.

Secondary microplastics are produced by the fragmentation of bigger plastic items (bags, fishing equipment, packaging, bottles, etc.), mostly after photodegradation²⁸ (reactions produced by the long exposure to UV radiation), but also biological action, erosion and other factors. Landfills, construction sites, littering, mismanaged waste and equipment from fishing and transport vessels as well as polymeric fibres from clothing

and TRWP are some of the major sources of secondary microplastics in the environment²⁹.

Weathering occurs mostly on the surface and affects every polymer differently, but some common mechanisms are involved³⁰. Thermal or photochemical reactions can break chemical bonds and release atoms or functional groups, depolymerize into its monomers or suffer scissions at random points in its chain. They can also oxidize or hydrolyse in contact with water (especially polymers with esters, amides or acetyl groups in their structures). All these reactions modify the superficial structure of the plastic item and weaken its integrity, favouring the occurrence of microplastics.

Further degradation generates nanoplastics (particles of less than 1 μm). Their smaller size allows them to penetrate cellular membranes, potentially affecting their functions, and can transport more adsorbed contaminants by mass than their larger counterparts, due to their higher surface/volume ratio. Even though there is not extensive literature about them due to the lack of adequate methods of analysis and detection, there are already some articles documenting their effects on various organisms, like zebrafish³¹, sea urchins³², zooplankton (*Daphia magna*³³ and *Daphia pulex*³⁴) and algae (*Chlorella*, *scenedesmus*^{35,36}).

3.2. Environmental effects

Around 77 % of the EU plastic waste is recycled or used for energy recovery²¹, while the remaining 23 % end up in landfills. If poorly managed, these plastics can escape via wastewater, wind and torrential transport. It is estimated that ca. 10 % of all produced plastics ultimately reach the oceans³⁷. Tyre and road wear particles (TRWP) are an important contributor to polymer pollution, mostly as suspended particulate matter. Flakes of old paint, fishing equipment fragments or textile fibres are also potential polymeric residues that can easily enter different ecosystems.

As plastic residues tend to wash out into seas and oceans, a great deal of research has been undergone to study them in this specific

ecosystem^{19,38-42}. Other studies have focused on the effects of plastics and microplastics in other media, like agricultural soil⁴³⁻⁴⁵ or the atmosphere^{46,47}, both indoor and outdoor.

MPs are present in most environmental compartments worldwide. It has been proven that they have detrimental effects in soil and fauna^{48,49}, and can easily enter and spread through the marine trophic chain and even reach humans. Several studies argue that the ingestion of microplastics has a detrimental effect on the nervous system. However, the most harmful properties are not attributed to the polymers themselves⁵⁰⁻⁵², at least in current concentrations in the environment. The main concern are the multiple chemicals that are used during their production, like additives, flame retardants or pigments, as well as the contaminants and biota that get adsorbed at their surface while in the environment. In certain environments, it is common for plastics to loose additives like plasticizers or hardness and resistance enhancers, further accelerating their fragmentation and increasing their pollution potential. For example, for PC production and other resins it is common to use Bisphenol A as a co-monomer, which is a potential endocrine disruptor and carcinogen⁵³. Some phthalates (esters of phthalic acid), like di-(2-ethylhexyl)phthalate (DEHP), mostly used in PVC production, are widely used as plasticizers and additives, used to modify different physical properties in plastic, like durability or flexibility, and have been proved to have detrimental effects in the endocrine and reproductive systems⁵⁴.

Most research is centred on the marine environment, as most residues tend to end up there via rivers, floods, rain and atmospheric transport. Even, due to their small size, they can pass through most wastewater treatment facilities, so it is virtually impossible to avoid them nowadays. There are plastics in all oceans and seas, and in most lakes and rivers, even in remote locations like lake Hobsgol, in Mongolia⁵⁵, lake Dimon, in the Carnic Alps (Italy)⁵⁶, sparsely populated shorelines⁵⁷, the Arctic and Antarctic coasts⁵⁸⁻⁶¹ or deep sea sediments⁶². One of the most striking examples of the magnitude of plastic pollution is the great oceanic garbage patches found in the Pacific, Atlantic and Indian oceans⁶³, as floating solid matter tends to accumulate in “gyres” (circular oceanic currents). Specifically in the North Pacific Gyre, the largest of these patches, an estimated 79000 tonnes of plastic are floating at the

surface of the ocean, of which around 8 % in mass are thought to be microplastics⁶⁴.

Depending on their density, plastics and microplastics can float at the water's surface or sink and embed themselves in the sediments. Thus they become easily accessible for the local fauna and it is common for small fragments to be ingested by plankton and fish, entering the trophic chain. The amount of literature addressing the occurrence of microplastics in marine flora and fauna is very extensive: algae, fish, marine mammals and birds, bivalves and other macro- and microbiota have been studied in over 1500 publications dealing with microplastic occurrence in marine organisms since 2010⁶⁵.

Microplastics can affect organisms that regulate organic matter decomposition in soils and sediments. This poses an ecological risk, as decomposers recycle nutrients necessary for plant growth and development, potentially disrupting an important part of the ecosystem^{44,66}. Also, as in the previous scenario, they can be easily introduced into the trophic chain through different animals. Some authors studied the effects and occurrence of microplastics in fauna, like earthworms⁶⁷, springtails⁶⁸, snails⁶⁹, nematodes⁷⁰, isopods⁷¹, enchytraeid worms⁷² and oribatid mites⁷³. Some of these publications show some negative effects, like histopathological damage, oxidative stress, genetic and reproductive damage, neurotoxicity and other metabolic and digestive disorders⁴⁹. The occurrence of microplastics in soil is mostly derived from agricultural activity and waste treatment. In agricultural soil the most common sources of microplastic pollution are film mulching (a practice consisting of covering the crops with a plastic film, often PE or PP, to better regulate temperature and humidity and avoid weed proliferation), the use of sewage sludge as compost, atmospheric deposition, runoff, littering, etc^{43,49,74}. In urban and peri-urban soils, TRWP derived from vehicular traffic has a larger impact.

Atmospheric microplastics are held in suspension in the air, being transported by the wind. In this medium, like oceanic microplastics, they can travel long distances, as proved by the presence of microplastic particles in remote places with no nearby population nuclei, like mountains, glaciers or polar regions^{59,75,76}. There are numerous

anthropogenic sources, the most important ones being textile fibres from clothing, both indoor and outdoor, releasing mostly PS, PA and acrylic. Also, TRWP and other urban and industrial activities can add to this type of pollution. Eventually, the largest and heaviest airborne particles deposit, substantially adding to the previously discussed soil and marine pollution. Their permanence in the atmosphere depends also on several external factors, like wind, rain, the presence of physical obstacles (like buildings), temperature, etc^{46,77}, being more prevalent in urban areas.

A source of concern are particles smaller than $<20\ \mu\text{m}$, which can go through cell membranes, potentially disrupting its functions, but then again, more information is needed to assess the nature and extent of these effects.

3.3. Spectrometric measurements

Nowadays, vibrational spectrometry is one of the most common techniques to identify and monitor environmental (micro)plastic residues²⁰. As for most carbon based-compounds, each polymer has a unique infrared spectrum. Weathering has also an impact on the shape of the spectra as it modifies the polymeric chains in the surface, although it is hard to identify the specific processes taking place. This is due to the different environmental conditions they can be subject to and the chemical reactions they have undergone, not only on their polymeric chain, but also on the pigments, additives or plasticizers present on their surface.

A problem that arose while studying the existing literature in the microplastics field was that, being a relatively young field of study, the information reported about the infrared instrumental setup in publications was often unclear, incomplete or insufficient, as they often obviate important information, like the measuring technique, instrumental characteristics, conditions or data treatment used. These shortcomings showed the need to establish some sort of standardization in the information provided in the publications, so the studies can be reproducible and comparable.

4. Infrared spectrometry

4.1. History

Infrared (IR) spectrometry has been developed throughout several centuries^{2,78-83}, after William Herschel discovered IR radiation in 1800. Using a prism to divide the different wavelengths of sunlight and with highly sensible thermometers, he observed that the temperature outside the visible part of the spectrum increased. He named this effect “calorific rays”. In **Figure 4**, a diagram of this experiment is shown, where it can be seen that the temperature of the thermometer augments as the wavelength decreases (to the left of the picture).

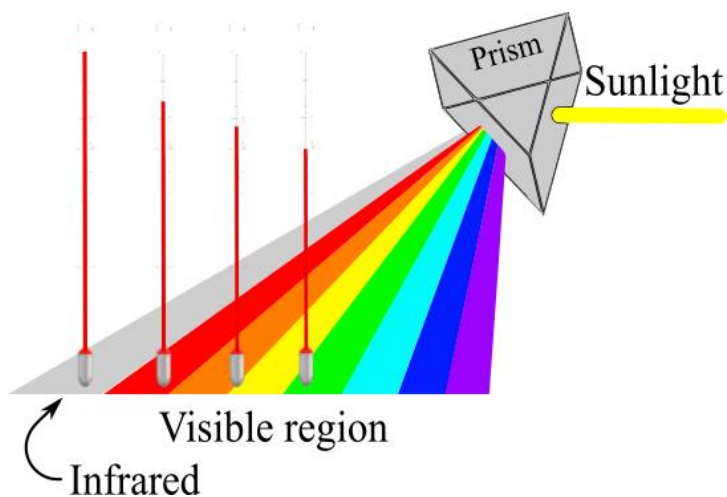


Figure 4: Schematic of Herschel’s experiment.

However, it was almost a century later that this effect began to be studied more thoroughly. Several scientists made some advancements in the technique, especially after the decade of the 1880s, summarized in **Figure 5**^{81,84,85}. It wasn't until the mid-20th century, bolstered by WW-II, that the technology was developed for more commercial applications, like the industrial analysis of endorphins and synthetic rubber. Since then, applications and publications employing IR spectrometry have increased exponentially. Because of that, new methods and techniques were developed, allowing IR spectrometry to be applied to an immense quantity of fields like medicine, pharmacy, environmental studies or different industrial applications.

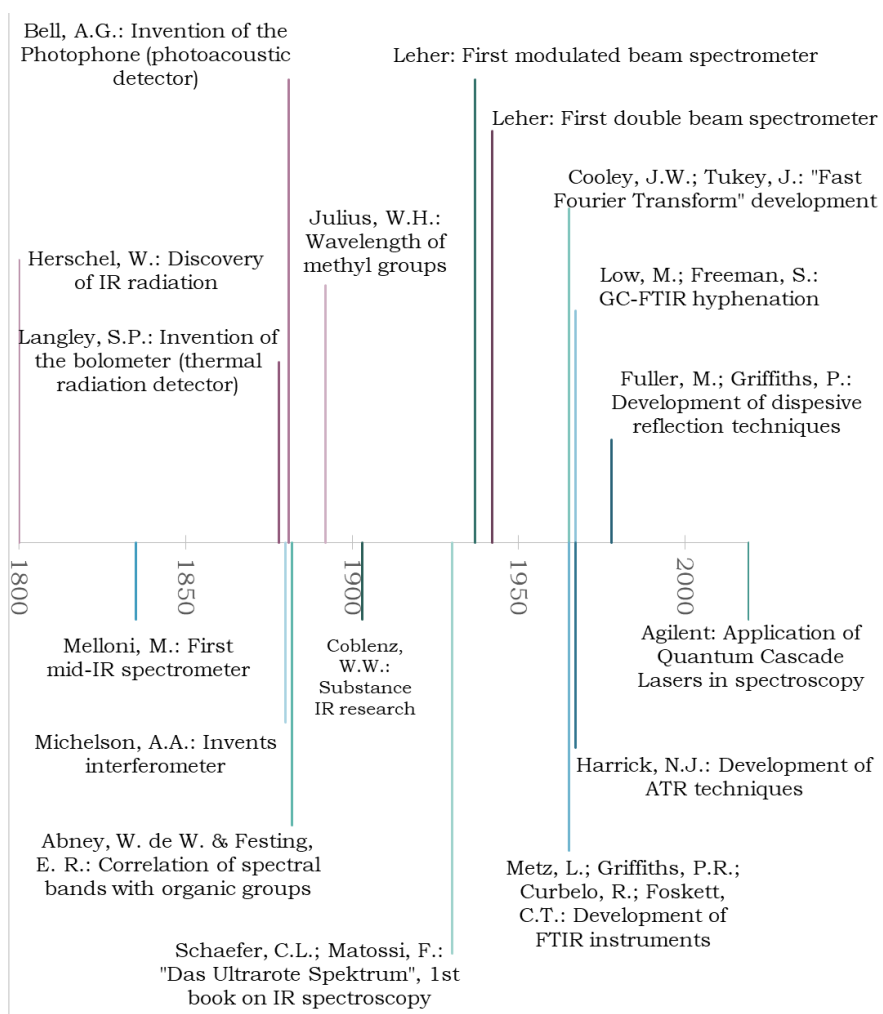


Figure 5. Some major milestones in vibrational spectrometry.

4.2. Fundamentals

Spectrometry is based on the measurement of the interaction of electromagnetic radiation with matter. However, different kinds of radiation have different interactions with it, with several wavelength (λ) intervals being especially interesting from an analytical point of view. Thus, different types of spectrometry have been developed: IR, UV-Vis, X-ray, etc., each of them with different applications, advantages and limitations. The IR region is usually measured in wavenumbers ($1/\lambda$) and is situated between 14000 and 4 cm^{-1} , subdivided into Far-IR (400-4 cm^{-1}), Mid-IR (4000-400 cm^{-1}) and Near-IR (14000-4000 cm^{-1}).

IR spectrometry is one of the most used analytical techniques worldwide, as it can be applied to almost any kind of sample: solid, liquid or gaseous, both organic and inorganic (though, for the latter, their spectra can be difficult to interpret, being X-Ray spectrometry more convenient). Absorbed IR radiation excites the vibrational and rotational levels of the bonds in the molecules of the sample. The only molecules invisible in the IR region are homopolar diatomic gases (O_2 , H_2 , N_2) and noble gases, and it can't differentiate optical isomers. Besides those exceptions, an IR spectrum is unique for each substance, which makes IR spectrometry a very useful tool.

Molecules are not static structures, they are in continuous movement, both the constituent atoms and the length and orientation of its chemical bonds. There are different kinds of movements in these bonds: stretching, bending, rocking and other deformations. Those are called "vibrations", and, when the molecule is struck with IR radiation, their energy states change. A molecule with different kinds of bonds and a high number or diversity of atoms tends to exhibit complex vibrational and rotational transitions. This usually translates into a higher number of peaks or bands in the spectra. The so-called normal modes of vibration are $3n-6$ ($3n-5$ in linear molecules), being n the number of atoms. In addition, there are combination bands (when the same wavenumber excites two different kinds of vibrations), overtones (multiples of a previous active wavelength) or "hot" bands (a transition between two excited states)⁸².

When radiation interacts with a sample, the different wavelengths are either reflected or absorbed to a certain degree. By measuring their

relative variations and plotting them against some frequency scale spectra can be obtained.

In IR spectrometry, transmitted or reflected radiation is directed towards a detector, which measures its energy after interacting with the sample. This is generally expressed in % of transmittance, reflectance, or, most commonly, absorbance (that part of the radiation retained by the sample to change its energy state). Depending on how the radiation interacts with the sample several kinds of IR techniques exist. Some of the most usual ones are shown in **Figure 6**.

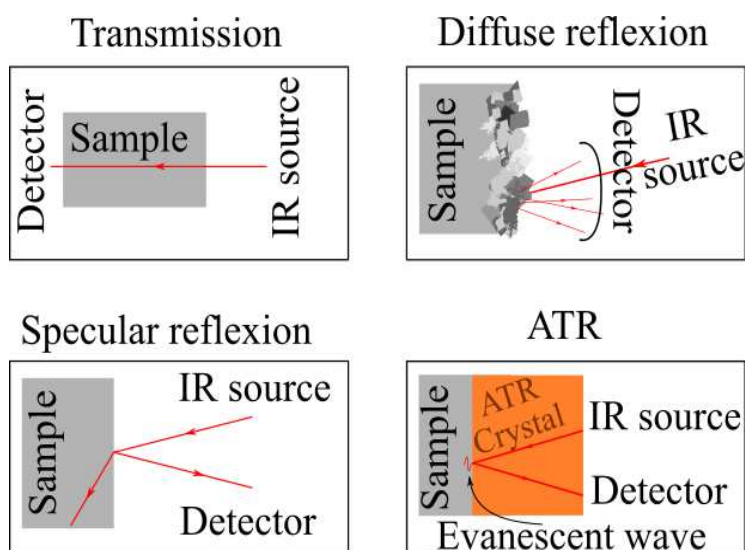


Figure 6: Typical IR working modes.

They are:

Transmission: The most traditional measuring mode. The radiation beam goes through the sample to the detector. In this mode, the most difficult part is often to prepare the sample. It can be in several different configurations, like in solution, forming a thin film, embedded in KBr pellets or as mulls. In the case of gases, a cell with IR transparent windows is often used.

In any case, the objective is to prepare a sample with a given thickness or concentration that allows the passage of the radiation through it, and

allows to establish a relationship between the measured spectrum and the analyte(s) to be identified or its concentration. This technique has been used to analyse microplastics, although its main problem is that the light paths of the particles are different and not necessarily homogeneous through them.

Reflectance: It is a very useful mode to characterize opaque or thick solid samples. Instead of the transmitted light, it is the reflected radiation what is measured. Depending on the surface morphology of the sample, the radiation can behave in two different ways: samples with an irregular surface or in powder form lead to diffuse reflectance, while samples with a bright, flat and homogeneous surface produce specular reflectance.

In diffuse reflectance the IR beams are scattered by the sample's surface in different directions. This hinders the ability to establish a correlation between the intensity of the peaks and the actual quantity of reflected radiation, often yielding almost unidentifiable spectra. To mitigate this effect, a Kubelka–Munk mathematical correction can be applied. In essence, it modifies the raw spectrum taking into account the absorption coefficients and the scattering factor of the sample. This transforms the reflectance spectrum to a transmission-like spectrum, in which a direct relation between the spectrum and the analyte (or its concentration) can be established (following Bouguer-Lambert-Beer's law). Diffuse radiation usually requires the use of an adequate accessory to collect all radiation.

In the case of specular reflectance, the radiation reflects in a singular direction, as it would in a mirror. In this situation the angle of reflectance is identical to the angle of incidence. Spectra obtained this way have the appearance of a 1st derivative, with a positive maximum and a negative minimum for each peak, known as “*restrahlen*”. Very briefly, this is due to the partially “imaginary” nature of the equation of the refractive index: $N = n + ik$, where N is the complex refractive index, n the real refractive index, k the absorption coefficient and $i = \sqrt{-1}$. A spectrum is the representation of N with the wavenumber. However, n varies drastically at wavenumbers where the sample absorbs radiation, generating this derivative shape, while a k vs wavenumber plot is identical in shape to the absorbance spectrum. To correct this effect a

Kramers-Kronig transformation is used, so the n and k influence on the spectrum can be separated⁸¹.

Unfortunately, real samples don't have an exclusively specular or diffuse reflectance, because their surface will neither be perfectly flat nor will disperse the radiation homogeneously in every direction, so it is the user's responsibility to decide which processing will be more adequate in each situation.

Attenuated total reflectance (ATR): This technique is somewhat more complex than the previous ones. It is based on the behaviour of the radiation when it hits in an interphase between two materials with different refractive indexes⁸⁵. When a beam traverses through a material h and impacts in another material j part of the radiation is reflected back and part is transmitted to j , following the equation:

$$R = \frac{(n_j - n_h)^2}{(n_h + n_j)^2}$$

Where R (reflectivity) is the portion of the radiation that is reflected back to material h and n_h and n_j are the refraction indexes of both materials. Also, the angle of the transmitted beam can be calculated using Snell's law:

$$n_h * \text{sen } \theta = n_j * \text{sen } \phi$$

Being θ and ϕ the angles in which the radiation is reflected to the material h and transmitted to the material j , respectively, with a plane perpendicular to the interphase between them. When $\phi = 90^\circ$ most of the radiation is not transmitted to j , as it would follow the plane of the interphase, and the corresponding θ angle is denominated critical angle θ_c . At values of $\theta \geq \theta_c$ only internal reflection occurs, and the radiation is not transmitted to the material j . This preserves virtually all the radiation in each reflection within h , allowing a high quantity of reflections, as the only energy losses would be due to the absorption of the material h . However, even though the beam is reflected at the surface of the interphase, part of the electrical field of the radiation penetrates slightly in the material j , about 0.1 to 5 μm , before being reflected. This phenomenon is called *evanescent wave*, and is where the sample

(material j in this case) interacts with the radiation. In ATR spectrometry material j is the sample, and material h a medium with a high refraction index (diamond, Germanium, ZnSe or ZnS). As the evanescent wave occurs at the surface of the sample, part of the radiation is absorbed, thus enabling the acquisition of its spectrum.

As the only energy absorbed is part of this *evanescent wave*, ATR spectra usually have less intensity than the transmittance or reflectance spectra. However, as the sample is pressed against the crystal, no specular or diffuse effects appear and, unlike transmittance, opaque or thick samples pose no measuring problem.

Besides these different modes of operation, it is also common to hyphenate FTIR spectrometry with other techniques to gain more information, obtain better results or measure different kinds of samples. The most common combinations are with chromatographic and thermogravimetric techniques. Here, the spectra of the different molecules, already separated by the chromatograph or thermal events are registered, offering additional insight into the sample composition. These combinations, however, will not be discussed in this Doctoral Thesis.

Another hyphenated technique is **IR microscopy**⁸¹. The combination of a microscope with a FTIR spectrometer is especially useful for the visualization and analysis of small particles (down to $10\ \mu\text{m}^2$ in modern instruments). IR microscopy, or micro-IR, can work in either transmittance, reflectance and/or ATR modes. This enables the measurement of not only small items but very heterogeneous ones (e.g. composite materials), greatly increasing the applicability of the technique. Some examples are biological and forensic applications, quality control, or, as in this Doctoral Thesis, microplastic analysis.

In **Figure 7** the modes of operation of an IR microscope are depicted. The light is reflected from the source by a toroidal mirror to another mirror that aligns it vertically. For transmittance mode, it is reflected upwards, while in reflectance and ATR, downwards. In any case, the beam goes through a Cassegrain, a pair of concave and convex mirrors that focuses the beam on the surface of the sample or the ATR crystal. Once transmitted or reflected after interacting with the sample, the beam

is redirected via mirrors and Cassegrains to the detector, usually a Mercury Cadmium Telluride (MCT), encased in a deposit with liquid nitrogen.

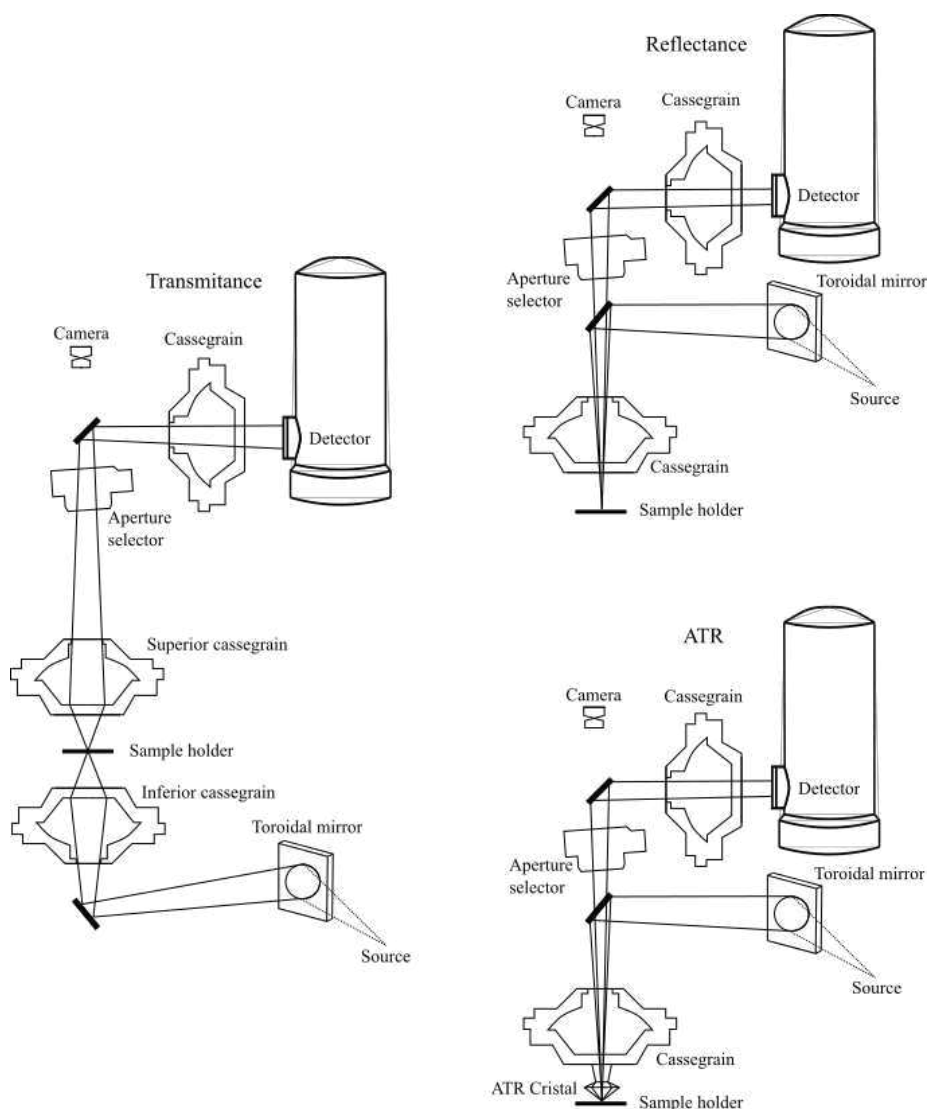


Figure 7. Schematics of the different modes of operation of an IR microscope. Adapted from Spotlight 200 User's Guide⁸⁶

A practical limitation of most IR microscopes nowadays is related to the focusing capabilities of the IR beams, as they can be focused only up to ca. 5-10 μm (their wavelength size). Besides, the edges of the mechanical

mask (apertures) that have to be adapted to the dimensions of the sample before measurement, which can create polychromatic radiation, known as “slit effect” in typical monochromators. In most instruments, this implies that particles lower than 20 μm cannot be measured accurately, as the S/N ratio decreases considerably. The number of scans can be increased to reduce the noise, but the signal is usually still of low quality, and the measurement times can increase considerably, making it impractical when the number of particles to be measured is large. As a result, particles lower than 30 μm are not easy to measure by classical micro-IR spectroscopy.

4.3. State of the art

Vibrational spectrometry is a highly innovative field. From quantum laser radiation sources to detectors able to register a spectrum for every pixel in an image, IR spectrometry has come a long way since its inception⁸⁷.

Initially developed for satellites, hyperspectral imaging (HSI)⁸⁸ cameras (detectors) are designed to register a full spectrum of each pixel from an image of a sample. The result is a tridimensional data space, a hypercube, with a spatial x - y image and a z dimension as the spectra of a specific wavenumber range (which can be in the visible or IR regions). This is useful not only to determine the composition of an analyte, but to ascertain its spatial distribution in the case of heterogeneous samples. Its main drawback at present is its resolution, limited to 300-100 μm for laboratory applications. Nevertheless, they have been successfully applied in fields like biomedicine^{89,90}, product quality and contaminant monitoring^{91,92} or military surveillance^{93,94}.

Quantum cascade lasers (QCL) are a novel kind of semiconductor laser emitting diodes in the mid- to far- IR radiation range. The laser is composed of a series of adjacent thin films of different semiconductor materials. This creates a succession of quantum wells (a potential well with discreet energy values) where an electron can go from a high energy state to a fundamental state in a quantum well, then tunnel to a contiguous quantum well, where it would be in a high energy state,

transition to a fundamental state, and so on, emitting a photon in each of these “jumps”⁹⁵, thus generating a beam of directed, collimated, high intensity light. By selecting the width and material of the different films it is possible to establish the spectral range of the laser, which is much easier than in traditional lasers. It is also smaller, yields more intense radiation than traditional sources and can be used at room temperature in combination with a Peltier cooling system (so no liquid nitrogen is necessary). Its use as a radiation source for vibrational spectrometry for commercially available instruments has increased, e.g. Agilent’s Laser Direct IR chemical imaging system (LDIR)⁹⁶, Bruker’s Hyperion II microscope⁹⁷ or Daylights’ Solutions ChemDetect Analyzer⁹⁸ and Spero Chemical Imaging Microscope⁹⁹. A basic diagram of a LDIR instrument is shown in **Figure 8**.

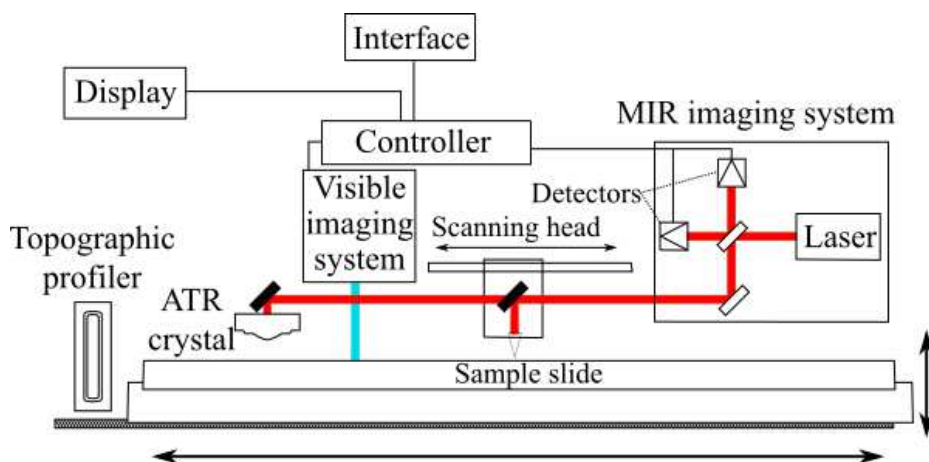


Figure 8. Conceptual diagram of an LDIR imaging system.

5. Chemometrics

Even though IR spectrometry is a very versatile and useful technique, sometimes it is insufficient by itself to solve a problem. For example, in the analysis of NG, the spectra are too complex to be interpreted without adequate tools to generate information. NG is composed of a mixture of different gases, very similar structurally and spectrally, making the resulting spectra hard to interpret, with a multitude of overlapping peaks belonging to different gas components, some of them with concentrations lower than 0.1 %. Chemometrics can help mitigate this problem.

Chemometrics is defined as the application of statistical, mathematical and formal logic techniques to both experimental design and interpretation of chemical data¹⁰⁰. As for IUPAC, “Chemometrics is the application of statistics to the analysis of chemical data (from organic, analytical or medicinal chemistry) and design of chemical experiments and simulations”¹⁰¹. In this work it is used in combination with IR spectrometry, but it can be used in other fields: mass spectrometry, atomic spectrometry, high-performance liquid chromatography (HPLC), gas chromatography-mass spectrometry (GC-MS), etc.

Its development started in the 1970s, with pioneers like Herman and Svante Wold, Bruce Kowalski, Paul Geladi or Michele Forina. In the beginning it was a purely theoretical discipline, and was intimately tied to computational science and the development of analytical instruments. In the 80s the first publications dedicated to applied chemometrics appeared, solving simple problems, mainly in IR spectrometry and chromatographic applications.

In subsequent years its relevance grew, both in applications and the complexity of the problems being solved, attracting funding from pharmaceutical or alimentary industries. Nowadays its applicability also encompasses petrochemistry, metabolomics, genetics, forensics, biology, medicine, environmental data, etc.

There are two main branches of chemometrics depending on their area of application: experimental design and multivariate analysis. The first

focuses mainly on the optimization of parameters in an experimental process (a synthesis, for example). Its main objective is to obtain the most information with the minimum number of experiments possible. This not only allows for a reduction in time and costs of the process but offers insight into the synergies between the different parameters.

Multivariate analysis, on the other hand, is used to extract as much information as possible from datasets. Once the experimental part of a study is performed, a high quantity of information might be obtained. For instance, in spectrometry or chromatography it is not uncommon to have datasets containing several thousand variables and hundreds of samples. This immense quantity of data might be difficult to interpret, and relevant details might be overlooked. Whether the objective of the study is to identify a compound, ascertain the concentration of a certain reagent or detect anomalous samples, the problems associated with the gathering of information from the raw data can be mitigated by using chemometric techniques which help extract relevant information and interpret it.

5.1. Fundamentals

In recent decades, a large number of chemometric techniques have been developed for classification, quantification and pattern recognition. In the publications presented in this Doctoral Thesis (Chapters 2 to 7) four methods were used: principal component analysis, partial least squares regression, classification and regression trees and support vector machines¹⁰²⁻¹¹¹. They are to be introduced briefly in the next paragraphs:

Principal component analysis (PCA) is one of the most used techniques in chemometrics¹⁰⁷⁻¹⁰⁹. It is an unsupervised exploratory data analysis method developed by Pearson in 1901¹¹², although the fundamentals had already been developed in the fields of mathematics and physics decades before. However, it was barely used in chemometric applications until the decade of the 1970s, thanks to the increase in the use of computers and the subsequent improvements in computation power. It is applied in unsupervised multivariate analysis to a matrix composed of several independent variables (p) (or X-block) measured on a collection of specimens of interest (n). The conceptual basis of the technique is that it is possible to concentrate the information of the X-

block in a reduced set of new mathematical variables. These are denominated Principal Components (PCs) which, in essence, are linear combinations of the measured variables. Each PC explains part of the information, or variance, of the original dataset.

Initially, mean centring of the data is usually performed ($\mathbf{X}_{centered} = \mathbf{X} - \bar{\mathbf{X}}$). Then the variance is calculated for each of the variables ($var(X) = \frac{1}{n-1} \sum_i^n (X_i - \bar{X}_i)^2$) and the covariance between them ($cov(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_i)(X_j - \bar{X}_j)$), obtaining a square matrix (of size p,p):

$$\begin{pmatrix} varX_1 & covX_1X_2 & \cdots & covX_1X_p \\ covX_2X_1 & varX_2 & \cdots & covX_2X_p \\ \vdots & \vdots & \ddots & \vdots \\ covX_pX_1 & covX_pX_2 & \cdots & varX_p \end{pmatrix}$$

Its z ($z=\min(n,p)$) *eigenvalues* (ev) and associated *eigenvectors* ($\vec{e}\vec{v}$) (with size (1,p)) are calculated (for more technical details some references can be consulted¹¹³). Here, each element of the *eigenvectors* is the *weight* (w) of the corresponding p variable in that specific PC, and it represents (in absolute value) the importance of each of the variables to define that specific PC. Being a symmetrical matrix, all $\vec{e}\vec{v}$ are orthogonal to each other. Then, the $\vec{e}\vec{v}$ are ordered from highest to lowest based on their ev and are normalized to 1. Being extracted from the covariance matrix, it is implied that the higher this value is, the higher the variance it explains, thus offering more information. These $\vec{e}\vec{v}$ are compiled into a matrix and they can be multiplied times the original (centered) dataset. The columns of the resulting matrix correspond to the so-called principal components, being each of the values the “score” (s) for a given sample on a specific PC:

$$\begin{pmatrix} X_{1,1} & \cdots & X_{1,p} \\ \vdots & \ddots & \vdots \\ X_{n,1} & \cdots & X_{n,p} \end{pmatrix} \begin{pmatrix} \vec{e}\vec{v}_{1,1} & \cdots & \vec{e}\vec{v}_{z,1} \\ \vdots & \ddots & \vdots \\ \vec{e}\vec{v}_{1,p} & \cdots & \vec{e}\vec{v}_{z,p} \end{pmatrix} = \begin{pmatrix} S_{1,PC1} & \cdots & S_{1,PCz} \\ \vdots & \ddots & \vdots \\ S_{n,PC1} & \cdots & S_{n,PCz} \end{pmatrix}$$

If the PC_i weights are multiplied by the square root of their corresponding *eigenvalue*, a new “loadings” vector can be obtained ($\vec{Loadings} = \vec{e}\vec{v} * \sqrt{ev}$), and its elements correspond to the Pearson’s

correlation coefficient between each variable and PC_z . Thus, a PC is composed of an eigenvalue, a vector of loadings and a vector of scores.

The user has to select the number of PCs to be included in the model by considering, among other factors, the variance explained by each PC. A graphical representation of the scores of two or three principal components (usually those with the most explained variance) would reveal clusters of data points corresponding to the samples, if there are real differences between them and the model is adequate. Samples with similar characteristics will be close to each other, thus allowing their grouping or ascertaining of patterns in the dataset.

Partial Least Squares regression (PLS-R) was developed initially by Herman Wold as an empirical algorithm in 1975¹¹⁰. It is a supervised multivariate analysis technique, useful for applications where it is required to correlate two sets of data corresponding to the same specimens of study. PLS-R is especially suitable to those cases where the number of measured X variables (predictors) is significantly larger than the number of predictands (i.e. the parameters to be predicted), like in spectroscopy or chromatography. The versatility of PLS-R has granted it myriads of applications, not only in chemometry, but bioinformatics and machine learning¹¹⁴.

Some variability exists in the PLS-R algorithms currently in use, but the most common one, and the basis for many others, is the *orthogonal score PLS-R*¹¹⁴. There is also a difference in the algorithm if there are multiple predictands (PLS2) or just one (PLS1, or just PLS). In the publications presented in this thesis only PLS1 will be considered. The process is somehow similar to a PCA decomposition, in the sense that it reduces the number of experimental variables to a few uncorrelated combinations, in this case called “latent variables” (LV) or “PLS components”. However, to define this LVs, PLS takes into account the Y value instead of the variance of the predictors when calculating the *weights*. This is done by the algorithm in such a way that the LVs establish the maximum possible covariance between the X-block and the Y-block. Once the weights are calculated, the LVs scores are calculated analogously to the PC scores in PCA. This results in a series of LVs, ordered from highest to lowest covariance with the Y values to be

predicted. The final number of LVs has to be selected by the users. Then a regression equation is created:

$$Y = \beta_0 + \beta_1 LV_1 + \cdots \beta_z LV_z$$

There, β_0 would be the y-intercept ($\beta_0 = 0$ in mean centred data) and β_z the parameters defining the multivariate slopes. Once the model is developed it is important to reassess the number of selected LVs. If there are enough samples, it is recommended to prepare a validation subset to check the errors in the predictions. In most programs, a cross-validation protocol is applied automatically. It develops models with the calibration samples, but separating one or several samples, which will then be used to evaluate the model. Then, those samples are reincorporated in the overall dataset and another subset is extracted. A model is constructed again and the error in the predictions is evaluated again. This procedure continues until all samples have been used for calibration and validation. In general, the adequacy of a model is assessed by looking at three average error parameters: the root-mean-square of prediction (RMSEP), when using a validation subset, the root-mean-square of cross-validation (RMSECV), when cross-validation is applied, and the root-mean-square of calibration (RMSEC), which is the error in the predictions of the samples used to calibrate the models. For a model to be acceptable, all those error figures should be low and similar to each other. Optimally, the number of selected LVs would also be low as to avoid “overfitting” the model to the calibration samples.

PLS-R has become a *de-facto* standard for multivariate regression, very useful for predicting properties when they are more difficult to measure directly (or if the technique necessary is costlier, time-consuming, prone to errors, etc.), using other experimental data like chromatograms or spectra of samples used for environmental studies or quality control.

Classification and Regression Trees (CART) can be used as a simple method of classification, based on the Theta Automatic Interaction Detection (THAID) algorithm, the first tree-based classification algorithm¹¹⁵. CART is a supervised technique based on the creation of binary partitions between the specimens of interest in a tree-like structure, in which, each partition, or “node” is divided in two different branches. In each of these nodes a condition is described¹¹⁶, e.g. a

wavenumber and a specific intensity. The samples are separated between those which surpass that intensity value and those which do not, assigning each of them to one or several categories. The process is repeated until, ideally, all categories are separated or other user-established conditions are met. The conditions of each node are defined by an impurity function I , being the optimal condition the one that minimizes I in the sub-nodes.

For regression, the process is similar, but considering continuous Y values instead of discrete classes. For regression purposes, the node impurity function I is calculated as the sum of squared deviations of the means of the node prediction and the real value of Y ¹¹⁶.

Support Vector Machines (SVM) is a supervised classification algorithm (though it can also be used for regression and outlier detection)^{111,117–119} developed by Boser, Guyon and Vapnik¹²⁰ in 1992. Originally, the algorithm develops just a binary separation (with categories coded as +1 and -1). Hence, in cases with multiple classes, a model would be developed for each of them in a one-vs-all approach. There are SVM algorithms capable of separating multiple classes at the same time (present in programs like Multid's Genex¹²¹ or PLS Toolbox¹²²), however, in most cases, they tend to yield slightly worse results than the classic, binary approach. SVM works by developing a separating boundary between the different categories (a line, a plane or a hyperplane when higher dimensions are considered) in a space defined by the original p variables (thus, with p dimensions). The boundary is placed so that its distance (margin) from the nearest data points (of each class) is maximized. A peculiarity of this algorithm is that only the datapoints closest to the boundary (the "support vectors" that inspire the name of the technique) are taken into account, being the ones that define both the boundary and the margins. The optimal hyperplane is obtained by calculating the *weights* of the support vectors and creating linear combinations: $wX_i + b = y$, where b represents the bias of the model^{117,118}. This equation, however, has to maximize the margins ($margin = \frac{2}{\|w\|}$) while respecting the constraints:

- $f(X) = w^T X_i + b \geq 1$ For samples with class +1

- $f(X) = w^T X_i + b \leq 1$ For samples with class -1

This defines what is called a “hard margin SVM”, however, this is not very adequate for datasets with outliers or overlapping, as it does not allow for errors in the models, which is not feasible in most real problems. To mitigate this, new “slack variables” (a sort of error variables, ξ_i) and a cost parameter (C) are included in the algorithm. The insertion of ξ_i allows some flexibility in the constraints of the margins, while C is a user-selected value that, in essence, takes into account the number of errors in classification and penalizes them. In this new “soft margin SVM” model the margins have to be maximized under the constraint that the empirical classification error must be minimized^{117,118}:

$$margin = \frac{2}{\|w\|} + C \sum_{i=1}^n \xi_i$$

Despite working well with linearly separable datasets, however its full potential is found when considering non-linearly separable groups. For this, the original dimensionality (i.e. the number of original variables) is augmented mathematically by using the so-called “kernel trick”. This is based on the calculation of new mathematical variables using a function of the original variables. There are several options, with linear, polynomial or radial basis (RBF) functions being the most common. Hence, conceptually, SVM is a mathematical method developed to add more dimensions to the original space by combining the original variables. This is done in hopes that a higher dimensional space would allow for a separation among the groups, even when that may not have been possible using only the original dimensions.

5.2. Variable selection methods

Another relevant application of chemometrics is to determine the most important variables of a dataset for a given purpose. Many techniques can be used for this. The ones used in the studies presented in this Doctoral Thesis (Chapters 3, 4 and 7) are:

In **interval partial least squares (iPLS)**¹²³ the original variables are divided into intervals of a width selected by the user, and a PLS model is performed with each one. The interval with the best prediction capabilities is selected and kept through. Then, a new interval is added using the same methodology, but using a combination of the first selection and each of the other intervals to develop the PLS models. The process is repeated recursively and the number of retained intervals increase as far as the result improve or the maximum number of intervals is reached. The selection can be performed forward or backwards, if the intervals are added successively (and so, the number of variables increases after each step) or reduced continuously (i.e. intervals of variables are discarded after each iteration), respectively. Each strategy may yield different selections. The final set of variables should reflect the most relevant information to address the problem or the interferences conveyed by each interval¹²³. The user can set the total number of intervals and the interval width, thus allowing the reduction of a large dataset to a controlled number of variables. The major drawback of this approach is that it may need high processing capabilities and time. This is especially acute when selecting small intervals of variables, as it forces the computer to perform a higher number of PLS models. iPLS can be a very useful and powerful technique with an adequate selection of parameters.

The Selectivity ratio index (SRI) measures the ratio between the explained and the residual variance for each variable^{103,124}. SRI can be considered as a measure of the capability of a variable to predict a specific property under study in a PLS model¹²⁵. A *scores* (t_{TP}) and a *loadings* (p_{TP}) vectors are calculated as: $t_{TP} = X * b / \|b\|$ and $p_{TP} = X^T * t_{TP} / (t_{TP}^T * t_{TP})$ where b is the regression coefficient vector (calculated as $b = (X^T * X)^{-1} * X$), the operator $\| \|$ represents the module or magnitude and the ^T superscript denotes a transposed matrix or vector. The SRI is then calculated as the ratio of variance that t_{TP} can explain (v_{expl}) versus the variance that cannot (v_{res}) for every variable j :

$$SRI = v_{expl,j} / v_{res,j} = \|t_{TP} * p_{TP,j}^T\|^2 / \|E_{TP,j}\|^2$$

Where $E_{TP,j}$ is the residuals vector. The SRI is thus proportional to the importance of a specific variable in a PLS prediction. This is not a variable selection tool itself, but a parameter that evaluates the relevance of the different variables for a model, which then can be sorted and selected manually. The variables with a SRI over 1 have a positive ratio of explained variance, thus making them eligible to be selected. The biggest downside of this method is, precisely, that all variables are evaluated individually, ignoring the possible interactions that can occur between two or more variables.

A **Variable Importance in Projection (VIP)**¹²⁶ score is calculated for each variable j as the sum of its PLS weights values weighted by the percentage of explained Y variance of each specific latent variable^{127,128}. It was first published as a variable selection method by Svante Wold et al.¹²⁹ It is calculated following the following equation:

$$VIP_j = \sqrt{\frac{\sum_{f=1}^F w_{jf}^2 * SSY_f * p}{SSY_{total} * F}}$$

Where F is the total number of latent variables in the model, w is the weight value of variable j for latent variable f , SSY_f is the sum of squares of explained variance for the latent variable f and p is the total number of variables.

This score is used to rank the variables according to their importance in predicting a property in PLS: values greater than 1 highlight very informative variables while values close to 0 indicate irrelevant ones. As for SRI, the variables with the highest VIPs are selected *ad-hoc*.

Monte-Carlo uninformative variable elimination (MCUVE)^{103,130} is a modified version of the UVE method¹³⁰, using a Monte Carlo approach to randomly create a large number of subsets of samples from the calibration dataset. PLS models are developed for each of the subsets. The regression coefficients associated with each variable in each model are collected into a vector c , and, then a reliability index is calculated for each of the variables: $RI = mean(c)/SD(c)$. The ones with the highest RI values are then selected as more informative. This approach is useful in reducing the possible overfitting of the model when the whole

calibration dataset is used, as it develops different models with different subsets of samples.

Random Frog is based on the comparison of several PLS models developed with randomly selected variables¹⁰³. An initial model is developed using w variables, and its cross-validation error is calculated. Then, a new model with a subset of those variables is developed. If the cross-validation error is superior to the initial, the new model is discarded; if not, it replaces the previous. This process is repeated a fixed number of iterations, selected by the user (preferably as high as possible, usually no less than 2000, though this increases the processing time) and a selection probability index is calculated for each variable depending on the number of times it was included in an accepted model.

iPLS, SRI and VIP are included in PLS_Toolbox MATLAB package, while MCUVE and random frog are from the libPLS package¹⁰³.

While a **Manual selection of the highest loadings** is not a variable reduction method in itself, it is possible to use the PCA loadings matrix to ascertain the most important variables in a dataset. If a PCA is performed with the full dataset, the scores plots allow the user to see which PCs separate better each category. Then, it is possible to identify the highest loadings of these PCs and retain the corresponding variables. This hopefully yields a dataset with only the variables that best differentiate the different categories. The main drawback of this approach is the inherent subjectivity of the selection and the fact that the loadings are obtained considering all variables.

Dynamic PCA is based on performing a Student's t -test for each variable. The test evaluates the significance (p -value) of each variable when a group of interest is compared against all other samples (one-vs-all approach). Then, all statistically significant variables are used to perform a new PCA.

Traditional scores plots will show immediately after the selection the separation between the categories, allowing the scientist to evaluate the goodness of the selection and, so, select the minimum number of variables that yield a satisfactory separation between both classes.

This algorithm is available in the commercial software GenEx (MultiD, Göteborg, Sweden).

5.3. Applications in natural gas analysis

In NG analysis, the primary objective is to determine the concentration of the different components of the gas mixture, as well as the Wobbe index. The chemometric method applied in this study was PLS regression on gas phase mid-IR spectra, combined with several variable reduction strategies: MCUVE, iPLS, random frog and VIP.

However, in IR spectrometry it is impossible to determine the N₂ concentration in the samples, which is inversely proportional to the Wobbe index, hindering its prediction. NG blends with less nitrogen show higher Wobbe index values so acceptable predictions were possible, but at higher N₂ concentrations the errors in the predictions increased significantly. Therefore, it was only possible to predict the Wobbe index in a reduced experimental range (14.8-15.3 kWh/Nm³). Besides, given the lower sensitivity of IR spectrometry when compared to GC (the most used analytical technique in NG quality control), it was not possible to develop methods to predict the concentrations of the components with the lowest concentrations (pentanes and hexanes, <1 %). Nevertheless, it was possible to adequately ascertain the concentrations of the other NG components. As IR spectrometry is not a common method of analysis for natural gas, related literature is scarce. The few publications regarding chemometric methods combined with vibrational spectroscopy in NG analysis are summarized in **Table 2**.

Table 2: Publications combining spectrometric measurements and chemometric models for the analysis of natural gas

Algorithm	Technique	Sample (analyte)	Ref.
PLS	FT-NIR	Natural gas	131
PLS	Handheld NIR	Natural gas, biogas	132
PLS	iHWG*- μ NIR	Synthetic mix. (CH ₄ -C ₄ H ₁₀)	133
PLS	FT-NIR	Natural gas (CO ₂)	134
PLS, PCR	FT-MIR and FT-NIR	Synthetic mix. (CH ₄ , C ₂ H ₆ , C ₃ H ₈)	135

*Substrate-integrated hollow waveguide

5.4. Applications in microplastic analysis

For plastic and microplastic analysis, several multivariate pattern recognition and classification techniques were tried (PCA, SVM and CART) with the purpose of creating models able to identify the different plastics independently to their aging, both in seawater and dry (shoreline) conditions.

In recent years, there has been a significant increase in the number of publications regarding the evaluation of microplastic determination in combination with the aid of different chemometric techniques. Their scope however encompasses a variety of environmental matrices and/or artificial scenarios as well as theoretical studies. In **Table 3** some relevant publications regarding this trend are summarized.

Table 3: Examples of publications combining spectrometric measurements and chemometric models for the identification or characterisation of microplastics

Algorithm	Technique	Sample	Ref.
PCA	ATR-FTIR and Raman micro-spectrometry	Fish stomach content	136
PCA	FTIR, Scanning electron microscopy/energy dispersive spectrometry	Seawater	137
PCA	ATR-FTIR	Chicken meat	138
PCA	Raman imaging	Printed microplastics	139
PCA and UMAP*	Hyperspectral imaging	Soil and wastewater	140
SVM, kNN, LDA and combinations with PCA	ATR-FTIR, NIR spectrometry, LIBS*, and X-Ray fluorescence	Consumer plastics, marine plastic debris	141
PCA, SVM and LDA	Raman spectrometry	Consumer products breakdown microplastics	142
PCA	ATR-FTIR	Artificially aged plastics	143
PCA	Raman imaging	Pristine microplastic and garden trimmer lines	144
PCA	Raman imaging and MALDI-MS*	Ink	145
PCA	Micro-FTIR	Microplastic biofilms	146
PCA, PLS	Micro-NIR, Raman spectrometry	Consumer plastics	147
LDA, collinear analysis	Metabolomics	Plastic mulch residues	148
1D-CNN*	Raman spectrometry	Pristine microplastics	149
PLS-DA, SVM	NIR spectrometry	Soil	150
PLS-DA, SIMCA	μ -FTIR	Sediment	151
SVM, NN, LS-SVM, PLS-DA	Vis. Hyperspectral imaging	In-situ underwater microplastics	152

PLS-DA	Hyperspectral imaging	Marine microplastics	153
PLS-DA	ATR-FTIR	Seafood	154
PLS-PM*	UV-Vis Raman and X-ray photoelectron spectrometry	Mulch film microplastics	155
ANN, kNN, LDA, PCA, PLS-DA, SIMCA, SVM	ATR-FTIR	Pristine microplastics	156
PCA, PLS-R	ATR-FTIR	Pristine and recycled microplastics	157
Two-dimensional correlation analysis	Fluorescence spectrometry	Dissolved organic matter	158
SVM, MD*, ML*	Hyperspectral imaging	Soil	159
Hierarchical cluster analysis	ATR-FTIR	Cellulose fibres	160
SVM	Hyperspectral imaging	Artificially weathered microplastics	161
DT*, SVM, CNN*	Hyperspectral imaging	Farmland soil	162
SIMCA*	Hyperspectral imaging	Environmental samples	163
MCR-ALS*	Raman hyperspectral imaging	Sand and fish faeces	164
Sub-kNN, BDT*, DBSCAN*	LDIR	Pristine and weathered microplastics	165

* UMAP: uniform manifold approximation and projection. LIBS: Laser-induced breakdown spectroscopy. MALDI-MS: Matrix-assisted laser desorption/ionisation mass spectrometry. 1D-CNN: one-dimensional neural network. LDA: Linear discriminant analysis. PLS-DA: Partial least squares discriminant analysis. PLS-PM: Partial least squares path modelling. DT: Decision tree. CNN: convolutional neural network. MD: Mahalanobis distance. ML: maximum likelihood. SIMCA: soft independent modelling of class analogy. MCR-ALS: Multivariate curve resolution-alternating least squares. BDT: Boosted Decision Tree. DBSCAN: Density-based Spatial Clustering of Applications with Noise.

6. Overview of the NG-related publications in this Doctoral Thesis

In Chapter 2, the use of several inert gases was studied in order to take advantage of the *gas broadening effect*. More defined and intense spectra were obtained, where smaller peaks can be visualized, increasing the information that can be surmised. The working hypothesis was that this would allow the development of more efficient multivariate chemometric regression models. That was precisely the objective of Chapters 3 and 4, where PLS regression was used to create models in order to predict the composition of a series of real, commercial NG samples. There, the use of variable selection methods (MCUVE, iPLS, SR and *random frog*) was explored in order to get rid of variables with no useful information or that hindered the models.

7. Overview of the microplastic-related publications in this Doctoral Thesis

The general aim of the papers presented here, besides their specific objectives, is to maximize the usefulness of the spectrometric information used to identify microplastics gathered from environmental samples. Because of this, it is of paramount importance to offer reliable methods of analysis to characterize, identify and quantify them. Specifically, the premise of Chapter 5 is to address the aforementioned lack of instrumental information presented in most microplastic-related publications, as well as to propose a unified way to present it. Chapter 6 studies the effects of weathering in the mid-IR spectra of PA. Artificially weathered pellets and powder were analysed by ATR, micro-reflectance and Scanning Electron Microscopy (SEM). There, the effects of photodegradation are studied considering both the evolution of the spectral profile of IR spectra and the morphologic characteristics seen in SEM images. In Chapter 7, the spectra of 9 of the most commonly-produced polymers were obtained in various instrumental configurations and weathering stages. Then, variable reduction methods and multivariate chemometric modelling tools were applied to predict the major polymeric constituent of several fragments, in both artificially-aged and field samples.

8. General conclusion

As a general conclusion, the publications that constitute this Doctoral Thesis focus on applying different FTIR spectrometry and chemometric techniques to address relevant questions in two environmentally relevant fields of study, with the intention of improving already existing measuring processes, or suggesting new ones. In general, novel analytical procedures were developed, with varying degrees of success. In the case of NG, the minor components could not be quantified. The Wobbe index couldn't be determined in the full range of the commercial products due to its nitrogen dependency. The concentrations of the main components of commercial NG have been determined with acceptable errors, with a methodology that offers faster results and less cost of reagents and workload than the chromatographic standards.

For plastic and microplastic characterisation, several models were developed, able to predict the polymeric composition of artificially aged plastics. Even though the results for environmental samples were not satisfactory using reflectance techniques, those for ATR show promising performance parameters. The weathering behaviour of polyamide was also studied and compared with previous partial studies found in literature. Finally, an outline for the standardization of the information reported in publications dealing with microplastic analysis was proposed in order to address a serious lack of information observed in many articles.

In both fields more work can be done, but this Doctoral Thesis demonstrates that the combination of IR spectrometry and chemometrics is a powerful tool to open new working methods with relatively moderate costs, short turnaround times and very low consumption of reagents.

9. References

- (1) ISO 6975:1997 Natural Gas — Extended Analysis — Gas-Chromatographic Method. International Organization for Standardization: Geneva 1997.
- (2) Kolthoff, I. M.; Elving, P. J. Infrared Spectroscopy. In *Treatise on Analytical Chemistry, vol. 7*; Interscience: Geneva, 1965; pp 3535–3743.
- (3) JPI Oceans. BASEMAN <http://www.jpi-oceans.eu/baseman/main-page> (accessed Jan 17, 2022).
- (4) Faramawy, S.; Zaki, T.; Sakr, A. A.-E. Natural Gas Origin, Composition, and Processing: A Review. *J. Nat. Gas Sci. Eng.* **2016**, *34*, 34–54. <https://doi.org/10.1016/J.JNGSE.2016.06.030>.
- (5) United States Environmental Protection Agency (EPA). Emission Factors for Greenhouse Gas Inventories [https://nepis.epa.gov/Exe/ZyNET.exe/P1013XRF.txt?ZyActionD=ZyDocument&Client=EPA&Index=2016 Thru 2020&Docs=&Query=&Time=&EndTime=&SearchMethod=1 &TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=](https://nepis.epa.gov/Exe/ZyNET.exe/P1013XRF.txt?ZyActionD=ZyDocument&Client=EPA&Index=2016%20Thru%202020&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&UseQField=&IntQFieldOp=0&ExtQFieldOp=) (accessed Dec 21, 2022).
- (6) European Research Institute for Gas and Energy Innovation (ERIG). Theses for the european energy future <https://erig.eu/wp-content/uploads/2021/05/ERIG-Theses-EU-energy-future.pdf> (accessed May 6, 2023).
- (7) International Energy Agency (IEA). Gas Market Report, Q1-2022 <https://iea.blob.core.windows.net/assets/4298ac47-e19d-4ab0-a8b6-d8652446ddd9/GasMarketReport-Q12022.pdf> (accessed Apr 5, 2023).
- (8) Leiker, M.; Cartellieri, W.; Christoph, K.; Pfeifer, U. Evaluation of Antiknocking Property of Gaseous Fuels by Means of Methane Number and Its Practical Application to Gas Engines. *ASME-paper 72-DGP-4* **1972**.
- (9) Eilts, P.; Klare, L. Investigations on the Determination of the Service Methane Number of LNG. In *SAE Technical Papers*; 2018; Vol. April, pp 1–11. <https://doi.org/10.4271/2018-01-1143>.

- (10) Zhang, Q.; Yang, Y.; Jia, D.; Li, M. Knocking Characteristics of a High Pressure Direct Injection Natural Gas Engine Operating in Strati Fi Ed Combustion Mode. *Open Phys.* **2021**, No. 19, 534–538. <https://doi.org/10.1515/phys-2021-0064>.
- (11) Ponte, S.; Andrade, J. M.; Vázquez, C.; Ferreiro, B.; Cobas, C.; Pérez, A.; Rey, M.; Vales, C.; Pellitero, J.; Santacruz, B.; Muniategui, S.; López-Mahía, P.; Shu, B.; Bettin, H.; Klaus, D.; Anders, B.; Betz, M.; Kühne, U.; Meier, C.; Eilts, P. Modeling the Natural Gas Knocking Behaviour Using Gas-Phase Infrared Spectra and Multivariate Calibration. *J. Nat. Gas Sci. Eng.* **2021**, *90*, 103944. <https://doi.org/10.1016/j.jngse.2021.103944>.
- (12) Emerson Process Management. The Wobbe Index and Natural Gas Interchangeability <https://www.emerson.com/documents/automation/white-paper-wobbe-index-natural-gas-interchangeability-ras-en-133638.pdf> (accessed Dec 15, 2022).
- (13) Speight, J. G. *Natural Gas A Basic Handbook*, 2nd ed.; Gulf Professional Publishing: Wyoming, 2018.
- (14) Lokhandwala, K. A.; Pinnau, I.; He, Z.; Amo, K. D.; DaCosta, A. R.; Wijmans, J. G.; Baker, R. W. Membrane Separation of Nitrogen from Natural Gas: A Case Study from Membrane Synthesis to Commercial Deployment. *J. Memb. Sci.* **2010**, *346* (2), 270–279. <https://doi.org/10.1016/j.memsci.2009.09.046>.
- (15) Mokhatab, S.; Mak, J. Y.; Valappil, J. V.; Wood, D. A. *Handbook of Liquefied Natural Gas*; Elsevier, 2014. <https://doi.org/10.1016/C2011-0-07476-8>.
- (16) Anderson, P. W. Pressure Broadening in the Microwave and Infra-Red Regions. *Phys. Rev.* **1949**, *76* (5), 647–661. <https://doi.org/10.1103/PhysRev.76.647>.
- (17) Jabłoński, A. General Theory of Pressure Broadening of Spectral Lines. *Phys. Rev.* **1945**, *68* (3–4), 78–93. <https://doi.org/10.1103/PhysRev.68.78>.
- (18) Silva, A. B.; Bastos, A. S.; Justino, C. I. L.; da Costa, J. P.; Duarte, A. C.; Rocha-Santos, T. A. P. Microplastics in the Environment: Challenges in Analytical Chemistry-A Review. *Anal. Chim. Acta* **2018**, *1017*, 1–19. <https://doi.org/10.1016/j.aca.2018.02.043>.

-
- (19) Bayo, J.; Rojo, D.; Olmos, S. Abundance, Morphology and Chemical Composition of Microplastics in Sand and Sediments from a Protected Coastal Area: The Mar Menor Lagoon (SE Spain). *Environ. Pollut.* **2019**, *252*, 1357–1366. <https://doi.org/10.1016/J.ENVPOL.2019.06.024>.
- (20) Shim, W. J.; Hong, S. H.; Eo, S. E. Identification Methods in Microplastic Analysis: A Review. *Anal. methods* **2017**, *9* (9), 1384–1391. <https://doi.org/10.1039/C6AY02558G>.
- (21) Plastics Europe. Plastics – the Facts 2022 <https://plasticseurope.org/knowledge-hub/plastics-the-facts-2022/> (accessed Dec 20, 2022).
- (22) Amato, J. A. Plastic: A Toxic Love Story. By Susan Freinkel (Boston: Houghton Mifflin Harcourt, 2011. 324 Pp.). *J. Soc. Hist.* **2013**, *46* (3), 811–814. <https://doi.org/10.1093/jsh/shs055>.
- (23) ISO. ISO/TR 21960:2020 Plastics — Environmental Aspects — State of Knowledge and Methodologies. International Organization for Standardization: Geneva 2020.
- (24) Li, Y. Legislation and Policy on Pollution Prevention and the Control of Marine Microplastics. *Water (Switzerland)* **2022**, *14* (18). <https://doi.org/10.3390/w14182790>.
- (25) de Vos, A.; Aluwihare, L.; Youngs, S.; DiBenedetto, M. H.; Ward, C. P.; Michel, A. P. M.; Colson, B. C.; Mazzotta, M. G.; Walsh, A. N.; Nelson, R. K. The M/V X-Press Pearl Nurdle Spill: Contamination of Burnt Plastic and Unburnt Nurdles along Sri Lanka’s Beaches. *ACS Environ. Au* **2021**, *2* (2), 128–135. <https://doi.org/10.1021/acsenvironau.1c00031>.
- (26) National Transportation Safety Board. Severe Winds Cause Ship Breakaway, Damage at Port in New Orleans <https://www.nts.gov/news/press-releases/Pages/NR20210916.aspx> (accessed Dec 20, 2022).
- (27) South Africa’s ecological “nightmare” after plastic pellets spill <https://news.sky.com/story/south-africas-ecological-nightmare-after-plastic-pellets-spill-11264554> (accessed Dec 20, 2022).
- (28) Chamas, A.; Moon, H.; Zheng, J.; Qiu, Y.; Tabassum, T.; Jang, J. H.; Abu-omar, M.; Scott, S. L.; Suh, S. Degradation Rates of Plastics in the Environment. **2020**. <https://doi.org/10.1021/acssuschemeng.9b06635>.

- (29) Sundt, P.; Schulze, P.-E.; Syversen, F. Sources of microplastic-pollution to the marine environment Project <https://www.miljodirektoratet.no/globalassets/publikasjoner/M321/M321.pdf> (accessed Dec 20, 2022).
- (30) Allara, D. L. Aging of Polymers. *Environ. Health Perspect.* **1975**, *11* (June), 29–33. <https://doi.org/doi.org/10.1289/ehp.751129>.
- (31) Lu, Y.; Zhang, Y.; Deng, Y.; Jiang, W.; Zhao, Y.; Geng, J.; Ding, L.; Ren, H. Uptake and Accumulation of Polystyrene Microplastics in Zebrafish (*Danio Rerio*) and Toxic Effects in Liver. *Environ. Sci. Technol.* **2016**, *50* (7), 4054–4060. <https://doi.org/10.1021/acs.est.6b00183>.
- (32) Della Torre, C.; Bergami, E.; Salvati, A.; Faleri, C.; Cirino, P.; Dawson, K. A.; Corsi, I. Accumulation and Embryotoxicity of Polystyrene Nanoparticles at Early Stage of Development of Sea Urchin Embryos *Paracentrotus Lividus*. *Environ. Sci. Technol.* **2014**, *48* (20), 12302–12311. <https://doi.org/10.1021/es502569w>.
- (33) Mattsson, K.; Johnson, E. V; Malmendal, A.; Linse, S.; Hansson, L.-A.; Cedervall, T. Brain Damage and Behavioural Disorders in Fish Induced by Plastic Nanoparticles Delivered through the Food Chain. *Sci. Rep.* **2017**, *7* (1), 11452. <https://doi.org/10.1038/s41598-017-10813-0>.
- (34) Liu, Z.; Cai, M.; Wu, D.; Yu, P.; Jiao, Y.; Jiang, Q.; Zhao, Y. Effects of Nanoplastics at Predicted Environmental Concentration on *Daphnia Pulex* after Exposure through Multiple Generations. *Environ. Pollut.* **2020**, *256*, 113506. <https://doi.org/10.1016/j.envpol.2019.113506>.
- (35) Bhattacharya, P.; Lin, S.; Turner, J. P.; Ke, P. C. Physical Adsorption of Charged Plastic Nanoparticles Affects Algal Photosynthesis. *J. Phys. Chem. C* **2010**, *114* (39), 16556–16561. <https://doi.org/10.1021/jp1054759>.
- (36) Besseling, E.; Wang, B.; Lürling, M.; Koelmans, A. A. Nanoplastic Affects Growth of *S. Obliquus* and Reproduction of *D. Magna*. *Environ. Sci. Technol.* **2014**, *48* (20), 12336–12343. <https://doi.org/10.1021/es503001d>.
- (37) Coyle, R.; Hardiman, G.; Driscoll, K. O. Microplastics in the

- Marine Environment: A Review of Their Sources, Distribution Processes, Uptake and Exchange in Ecosystems. *Case Stud. Chem. Environ. Eng.* **2020**, *2*, 100010. <https://doi.org/10.1016/j.cscee.2020.100010>.
- (38) Wagner, S.; Hüffer, T.; Klöckner, P.; Wehrhahn, M.; Hofmann, T.; Reemtsma, T. Tire Wear Particles in the Aquatic Environment - A Review on Generation, Analysis, Occurrence, Fate and Effects. *Water Res.* **2018**, *139*, 83–100. <https://doi.org/10.1016/j.watres.2018.03.051>.
- (39) Denuncio, P.; Bastida, R.; Dassis, M.; Giardino, G.; Gerpe, M.; Rodríguez, D. Plastic Ingestion in Franciscana Dolphins, *Pontoporia Blainvillei* (Gervais and d'Orbigny, 1844), from Argentina. *Mar. Pollut. Bull.* **2011**, *62* (8), 1836–1841. <https://doi.org/10.1016/j.marpolbul.2011.05.003>.
- (40) Gewert, B.; Plassmann, M. M.; Macleod, M. Pathways for Degradation of Plastic Polymers Floating in the Marine Environment. *Environ. Sci. Process. Impacts* **2015**, *17* (9), 1513–1521. <https://doi.org/10.1039/c5em00207a>.
- (41) Wayman, C.; Niemann, H. The Fate of Plastic in the Ocean Environment-a Minireview. *Environ. Sci. Process. Impacts* **2021**, *23* (2), 198–212. <https://doi.org/10.1039/d0em00446d>.
- (42) Lamb, J. B.; Willis, B. L.; Fiorenza, E. A.; Couch, C. S.; Howard, R.; Rader, D. N.; True, J. D.; Kelly, L. A.; Ahmad, A.; Jompa, J.; Harvell, C. D. Plastic Waste Associated with Disease on Coral Reefs. *Science (80-.)*. **2018**, *359* (6374), 460 LP – 462. <https://doi.org/10.1126/science.aar3320>.
- (43) Bläsing, M.; Amelung, W. Plastics in Soil: Analytical Methods and Possible Sources. *Sci. Total Environ.* **2018**, *612*, 422–435. <https://doi.org/10.1016/J.SCITOTENV.2017.08.086>.
- (44) Tian, L.; Jinjin, C.; Ji, R.; Ma, Y.; Yu, X. Microplastics in Agricultural Soils: Sources, Effects, and Their Fate. *Curr. Opin. Environ. Sci. Heal.* **2022**, *25*, 100311. <https://doi.org/10.1016/j.coesh.2021.100311>.
- (45) Kumar, M.; Xiong, X.; He, M.; Tsang, D. C. W.; Gupta, J.; Khan, E.; Harrad, S.; Hou, D.; Ok, Y. S.; Bolan, N. S. Microplastics as Pollutants in Agricultural Soils. *Environ. Pollut.* **2020**, *265*, 114980.

- <https://doi.org/10.1016/j.envpol.2020.114980>.
- (46) Prata, J. C. Airborne Microplastics: Consequences to Human Health? *Environ. Pollut.* **2018**, *234*, 115–126.
<https://doi.org/10.1016/J.ENVPOL.2017.11.043>.
- (47) Levermore, J. M.; Smith, T. E. L.; Kelly, F. J.; Wright, S. L. Detection of Microplastics in Ambient Particulate Matter Using Raman Spectral Imaging and Chemometric Analysis. *Anal. Chem.* **2020**, *92* (13), 8732–8740.
<https://doi.org/10.1021/acs.analchem.9b05445>.
- (48) Egbeocha, C. O.; Malek, S.; Emenike, C. U.; Milow, P. Feasting on Microplastics : Ingestion by and Effects on Marine Organisms. **2018**, *27*, 93–106. <https://doi.org/10.3354/ab00701>.
- (49) Wang, Q.; Adams, C. A.; Wang, F.; Sun, Y. Interactions between Microplastics and Soil Fauna : A Critical Review. *Crit. Rev. Environ. Sci. Technol.* **2022**, *52* (18), 3211–3243.
<https://doi.org/10.1080/10643389.2021.1915035>.
- (50) Barua, R.; Banerjee, D.; Bhowmik, S. Study of the Potential Impact of Microplastics and Additives on Human Health. In *Assessing the Effects of Emerging Plastics on the Environment and Public Health*; IGI Global, 2022; pp 128–147.
<https://doi.org/10.4018/978-1-7998-9723-1.ch007>.
- (51) Diamanti-Kandarakis, E.; Bourguignon, J.-P.; Giudice, L. C.; Hauser, R.; Prins, G. S.; Soto, A. M.; Zoeller, R. T.; Gore, A. C. Endocrine-Disrupting Chemicals: An Endocrine Society Scientific Statement. *Endocr. Rev.* **2009**, *30* (4), 293–342.
<https://doi.org/10.1210/er.2009-0002>.
- (52) North, E. J.; Halden, R. U. Plastics and Environmental Health: The Road Ahead. *Rev. Environ. Health* **2013**, *28* (1), 1–8.
<https://doi.org/10.1515/reveh-2012-0030>.
- (53) Hafezi, S. A.; Abdel-Rahman, W. M. The Endocrine Disruptor Bisphenol A (BPA) Exerts a Wide Range of Effects in Carcinogenesis and Response to Therapy. *Curr. Mol. Pharmacol.* **2019**, *12* (3), 230–238.
<https://doi.org/10.2174/1874467212666190306164507>.
- (54) Rowdhwal, S. S. S.; Chen, J. Toxic Effects of Di-2-Ethylhexyl Phthalate: An Overview. *Biomed Res. Int.* **2018**, *2018*, 1750368.
<https://doi.org/10.1155/2018/1750368>.

- (55) Free, C. M.; Jensen, O. P.; Mason, S. A.; Eriksen, M.; Williamson, N. J.; Boldgiv, B. High-Levels of Microplastic Pollution in a Large, Remote, Mountain Lake. *Mar. Pollut. Bull.* **2014**, *85* (1), 156–163. <https://doi.org/https://doi.org/10.1016/j.marpolbul.2014.06.001>.
- (56) Pastorino, P.; Pizzul, E.; Bertoli, M.; Anselmi, S.; Kušće, M.; Menconi, V.; Prearo, M.; Renzi, M. First Insights into Plastic and Microplastic Occurrence in Biotic and Abiotic Compartments, and Snow from a High-Mountain Lake (Carnic Alps). *Chemosphere* **2021**, *265*. <https://doi.org/10.1016/j.chemosphere.2020.129121>.
- (57) Hirai, H.; Takada, H.; Ogata, Y.; Yamashita, R.; Mizukawa, K.; Saha, M.; Kwan, C.; Moore, C.; Gray, H.; Laursen, D.; Zettler, E. R.; Farrington, J. W.; Reddy, C. M.; Peacock, E. E.; Ward, M. W. Organic Micropollutants in Marine Plastics Debris from the Open Ocean and Remote and Urban Beaches. *Mar. Pollut. Bull.* **2011**, *62* (8), 1683–1692. <https://doi.org/10.1016/j.marpolbul.2011.06.004>.
- (58) Bergmann, M.; Klages, M. Increase of Litter at the Arctic Deep-Sea Observatory HAUSGARTEN. *Mar. Pollut. Bull.* **2012**, *64* (12), 2734–2741. <https://doi.org/https://doi.org/10.1016/j.marpolbul.2012.09.018>.
- (59) Bergmann, M.; Mützel, S.; Primpke, S.; Tekman, M. B.; Trachsel, J.; Gerdts, G. White and Wonderful? Microplastics Prevail in Snow from the Alps to the Arctic. *Sci. Adv.* **2019**, *5* (8), eaax1157. <https://doi.org/10.1126/sciadv.aax1157>.
- (60) Barnes, D. K. A.; Walters, A.; Gonçalves, L. Macroplastics at Sea around Antarctica. *Mar. Environ. Res.* **2010**, *70* (2), 250–252. <https://doi.org/10.1016/j.marenvres.2010.05.006>.
- (61) Alurralde, G.; Isla, E.; Fuentes, V.; Olariaga, A.; Maggioni, T.; Rimondino, G.; Tatián, M. Anthropogenic Microfibres Flux in an Antarctic Coastal Ecosystem: The Tip of an Iceberg? *Mar. Pollut. Bull.* **2022**, *175*, 113388. <https://doi.org/10.1016/j.marpolbul.2022.113388>.
- (62) Van Cauwenberghe, L.; Vanreusel, A.; Mees, J.; Janssen, C. R. Microplastic Pollution in Deep-Sea Sediments. *Environ. Pollut.* **2013**, *182*, 495–499. <https://doi.org/10.1016/j.envpol.2013.08.013>.

- (63) Marine Debris Program Office of Response and Restoration. Garbage Patches <https://marinedebris.noaa.gov/info/patch.html> (accessed Dec 14, 2022).
- (64) Lebreton, L.; Slat, B.; Ferrari, F.; Sainte-Rose, B.; Aitken, J.; Marthouse, R.; Hajbane, S.; Cunsolo, S.; Schwarz, A.; Levivier, A. Evidence That the Great Pacific Garbage Patch Is Rapidly Accumulating Plastic. *Sci. Rep.* **2018**, *8* (1), 1–15. <https://doi.org/10.1038/s41598-018-22939-w>.
- (65) Elsevier B.V. Scopus <https://www.scopus.com/search/form.uri?display=basic#basic> (accessed Jan 11, 2023).
- (66) Ng, E. L.; Lin, S. Y.; Dungan, A. M.; Colwell, J. M.; Ede, S.; Huerta Lwanga, E.; Meng, K.; Geissen, V.; Blackall, L. L.; Chen, D. Microplastic Pollution Alters Forest Soil Microbiome. *J. Hazard. Mater.* **2021**, *409* (August 2020), 124606. <https://doi.org/10.1016/j.jhazmat.2020.124606>.
- (67) Cao, D.; Wang, X.; Luo, X.; Liu, G.; Zheng, H. Effects of Polystyrene Microplastics on the Fitness of Earthworms in an Agricultural Soil. In *IOP conference series: earth and environmental science*; IOP Publishing, 2017; Vol. 61, p 12148. <https://doi.org/10.1088/1755-1315/61/1/012148>.
- (68) Ju, H.; Zhu, D.; Qiao, M. Effects of Polyethylene Microplastics on the Gut Microbial Community, Reproduction and Avoidance Behaviors of the Soil Springtail, *Folsomia Candida*. *Environ. Pollut.* **2019**, *247*, 890–897. <https://doi.org/10.1016/j.envpol.2019.01.097>.
- (69) Song, Y.; Cao, C.; Qiu, R.; Hu, J.; Liu, M.; Lu, S.; Shi, H.; Raley-Susman, K. M.; He, D. Uptake and Adverse Effects of Polyethylene Terephthalate Microplastics Fibers on Terrestrial Snails (*Achatina Fulica*) after Soil Exposure. *Environ. Pollut.* **2019**, *250*, 447–455. <https://doi.org/10.1016/j.envpol.2019.04.066>.
- (70) Lei, L.; Wu, S.; Lu, S.; Liu, M.; Song, Y.; Fu, Z.; Shi, H.; Raley-Susman, K. M.; He, D. Microplastic Particles Cause Intestinal Damage and Other Adverse Effects in Zebrafish *Danio Rerio* and Nematode *Caenorhabditis Elegans*. *Sci. Total Environ.* **2018**, *619*, 1–8. <https://doi.org/10.1016/j.scitotenv.2017.11.103>.

- (71) Kokalj, A. J.; Horvat, P.; Skalar, T.; Kržan, A. Plastic Bag and Facial Cleanser Derived Microplastic Do Not Affect Feeding Behaviour and Energy Reserves of Terrestrial Isopods. *Sci. Total Environ.* **2018**, *615*, 761–766.
<https://doi.org/10.1016/j.scitotenv.2017.10.020>.
- (72) Lahive, E.; Walton, A.; Horton, A. A.; Spurgeon, D. J.; Svendsen, C. Microplastic Particles Reduce Reproduction in the Terrestrial Worm *Enchytraeus Crypticus* in a Soil Exposure. *Environ. Pollut.* **2019**, *255*, 113174.
<https://doi.org/10.1016/j.envpol.2019.113174>.
- (73) Selonen, S.; Dolar, A.; Kokalj, A. J.; Skalar, T.; Dolcet, L. P.; Hurley, R.; van Gestel, C. A. M. Exploring the Impacts of Plastics in Soil—the Effects of Polyester Textile Fibers on Soil Invertebrates. *Sci. Total Environ.* **2020**, *700*, 134451.
<https://doi.org/10.1016/j.scitotenv.2019.134451>.
- (74) Corradini, F.; Meza, P.; Eguiluz, R.; Casado, F.; Huerta-Lwanga, E.; Geissen, V. Evidence of Microplastic Accumulation in Agricultural Soils from Sewage Sludge Disposal. *Sci. Total Environ.* **2019**, *671*, 411–420.
<https://doi.org/10.1016/j.scitotenv.2019.03.368>.
- (75) Ambrosini, R.; Azzoni, R. S.; Pittino, F.; Diolaiuti, G.; Franzetti, A.; Parolini, M. First Evidence of Microplastic Contamination in the Supraglacial Debris of an Alpine Glacier. *Environ. Pollut.* **2019**, *253*, 297–301.
<https://doi.org/10.1016/j.envpol.2019.07.005>.
- (76) Zhang, Y.; Gao, T.; Kang, S.; Allen, S.; Luo, X.; Allen, D. Microplastics in Glaciers of the Tibetan Plateau: Evidence for the Long-Range Transport of Microplastics. *Sci. Total Environ.* **2021**, *758*, 143634.
<https://doi.org/10.1016/j.scitotenv.2020.143634>.
- (77) Kaur, S.; Nieuwenhuijsen, M. J.; Colvile, R. N. Fine Particulate Matter and Carbon Monoxide Exposure Concentrations in Urban Street Transport Microenvironments. *Atmos. Environ.* **2007**, *41* (23), 4781–4810.
<https://doi.org/10.1016/j.atmosenv.2007.02.002>.
- (78) Svehla, G. *Comprehensive Analytical Chemistry; Volume 6: Analytical Infrared Spectroscopy*, 1st ed.; Elsevier Publishing Company: Amsterdam, Holland, 1976.

- (79) Cantellano, M. A. G.; Zetina, L. M. M. La Espectroscopia y Su Tecnología: Un Repaso Histórico y Su Importancia Para El Siglo XXI. *Lat. Am. J. Phys. Educ.* **2015**, 9 (4), 1–14.
<https://doi.org/10.1119/1.2895684>.
- (80) Gómez, R.; Murillo, R. Espectroscopía Infrarroja
<https://sistemas.fcencias.unam.mx/~fam/Infrarroja.pdf> (accessed Nov 14, 2022).
- (81) Smith, B. C. *Fundamentals of Fourier Transform Infrared Spectroscopy*, 1st ed.; CRC press: Boca Raton, 2011.
- (82) *Near-Infrared Spectroscopy: Principles, Instruments, Applications*, 1st ed.; Siesler, H. W., Ozaki, Y., Kawata, S., Heise, H. M., Eds.; John Wiley & Sons: Weinheim, 2008.
- (83) *Introduction to Experimental Infrared Spectroscopy Fundamentals and Practical Methods*; Tasumi, M., Sakamoto, A., Eds.; Wiley: Chichester, UK.
- (84) Bougeard, D.; Buback, M.; Cao, A.; Gerwert, K.; Heise, H. M.; Hoffmann, G. G.; Jordanov, B.; Kiefer, W.; Korte, E.-H.; Kuzmany, H.; Leipertz, A.; Lentz, E.; Liquier, J.; Roseler, A.; Schnockel, H.; Schrader, B.; Schrotter, H. W.; Spiekermann, M.; Taillandier, E.; Willner, H. *Infrared and Raman Spectroscopy Methods and Applications*; Schrader, B., Ed.; CRC Press: Weinheim, 1996.
- (85) Smith, A. L. *Applied Infrared Spectroscopy. Fundamentals, Techniques and Analytical Problem Solving.*, 1st ed.; John Wiley & Sons: Chichester, UK, 1979.
- (86) Perkin Elmer. Perkin Elmer Spotlight 200 User ' s Guide
<https://www.s4science.at/wordpress/wp-content/uploads/2019/04/L1050036-Spotlight-200-Users-Guide.pdf> (accessed Apr 1, 2023).
- (87) Xu, J.; Thomas, K. V; Luo, Z.; Gowen, A. A. Trends in Analytical Chemistry FTIR and Raman Imaging for Microplastics Analysis : State of the Art , Challenges and Prospects. *Trends Anal. Chem.* **2019**, 119, 115629.
<https://doi.org/10.1016/j.trac.2019.115629>.
- (88) Cramer, J.; Vogt, F.; Booksh, K. S. Smart Sensors. In *Comprehensive Chemometrics*; Elsevier, 2009; pp 357–376.
<https://doi.org/10.1016/B978-044452701-1.00015-6>.

-
- (89) Gutiérrez-Gutiérrez, J. A.; Pardo, A.; Real, E.; López-Higuera, J. M.; Conde, O. M. Custom Scanning Hyperspectral Imaging System for Biomedical Applications: Modeling, Benchmarking, and Specifications. *Sensors* **2019**, *19* (7), 1692. <https://doi.org/10.3390/s19071692>.
- (90) Stergar, J.; Hren, R.; Milanič, M. Design and Validation of a Custom-Made Hyperspectral Microscope Imaging System for Biomedical Applications. *Sensors* **2023**, *23* (5), 2374. <https://doi.org/10.3390/s23052374>.
- (91) Bonifazi, G.; Palmieri, R.; Serranti, S. Evaluation of Attached Mortar on Recycled Concrete Aggregates by Hyperspectral Imaging. *Constr. Build. Mater.* **2018**, *169*, 835–842. <https://doi.org/10.1016/j.conbuildmat.2018.03.048>.
- (92) Palmieri, R.; Bonifazi, G.; Serranti, S. Recycling-Oriented Characterization of Plastic Frames and Printed Circuit Boards from Mobile Phones by Electronic and Chemical Imaging. *Waste Manag.* **2014**, *34* (11), 2120–2130. <https://doi.org/10.1016/j.wasman.2014.06.003>.
- (93) Ke, C. Military Object Detection Using Multiple Information Extracted from Hyperspectral Imagery. In *2017 International Conference on Progress in Informatics and Computing (PIC)*; IEEE, 2017; pp 124–128. <https://doi.org/10.1109/PIC.2017.8359527>.
- (94) Ardouin, J.-P.; Levesque, J.; Rea, T. A. A Demonstration of Hyperspectral Image Exploitation for Military Applications. In *2007 10th International Conference on Information Fusion*; IEEE, 2007; pp 1–8. <https://doi.org/10.1109/ICIF.2007.4408184>.
- (95) Skoog, D. A.; Holler, F. J.; Crouch, S. R. *Principles of Instrumental Analysis*, 7th ed.; Cengage learning: Boston, Massachusetts, 2017.
- (96) Agilent Technologies Inc. Agilent 8700 LDIR Chemical Imaging System Brochure https://www.agilent.com/cs/library/brochures/5994-0275EN_8700_LDIR_Chemical_Imaging_System_brochure.pdf (accessed Feb 8, 2023).
- (97) Bruker Optics. Hyperion II <https://www.bruker.com/es/products-and-solutions/infrared-and-raman/ft-ir-microscopes/hyperion-ii->

- ft-ir-and-qcl-microscope.html (accessed Dec 2, 2022).
- (98) DRS Daylight Solutions. ChemDetect Analyzer <https://daylightsolutions.com/product/chemdetect/> (accessed Dec 2, 2022).
- (99) DRS Daylight Solutions. Spero Chemical Imaging Microscope <https://daylightsolutions.com/product/spero/> (accessed Dec 2, 2022).
- (100) Héberger, K. Chemoinformatics—Multivariate Mathematical—Statistical Methods for Data Evaluation. In *Medical Applications of Mass Spectrometry*; Vékey, K., Telekes, A., Vertes, A. B. T.-M. A. of M. S., Eds.; Elsevier: Amsterdam, 2008; pp 141–169. <https://doi.org/10.1016/B978-044451980-1.50009-4>.
- (101) IUPAC. *The IUPAC Compendium of Chemical Terminology*, 2nd ed.; Gold, V., Ed.; International Union of Pure and Applied Chemistry (IUPAC): Research Triangle Park, NC, 2019. <https://doi.org/10.1351/goldbook>.
- (102) Febrero-Bande, M.; Galeano, P.; González-Manteiga, W. Functional Principal Component Regression and Functional Partial Least-Squares Regression: An Overview and a Comparative Study. *Int. Stat. Rev.* **2017**, *85* (1), 61–83. <https://doi.org/10.1111/insr.12116>.
- (103) Li, H.-D.; Xu, Q.-S.; Liang, Y.-Z. LibPLS: An Integrated Library for Partial Least Squares Regression and Linear Discriminant Analysis. *Chemom. Intell. Lab. Syst.* **2018**, *176* (November 2017), 34–43. <https://doi.org/10.1016/j.chemolab.2018.03.003>.
- (104) Westerhuis, J. A.; Kourti, T.; MacGregor, J. F. Analysis of Multiblock and Hierarchical PCA and PLS Models. *J. Chemom.* **1998**, *12* (5), 301–321. [https://doi.org/10.1002/\(SICI\)1099-128X\(199809/10\)12:5<301::AID-CEM515>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1099-128X(199809/10)12:5<301::AID-CEM515>3.0.CO;2-S).
- (105) Ballabio, D.; Consonni, V. Classification Tools in Chemistry. Part 1: Linear Models. PLS-DA. *Anal. Methods* **2013**, *5* (16), 3790. <https://doi.org/10.1039/c3ay40582f>.
- (106) Gómez-Carracedo, M. P.; Andrade, J. M.; Rutledge, D. N.; Faber, N. M. Selecting the Optimum Number of Partial Least Squares Components for the Calibration of Attenuated Total Reflectance-Mid-Infrared Spectra of Undesigned Kerosene

- Samples. *Anal. Chim. Acta* **2007**, 585 (2), 253–265.
<https://doi.org/10.1016/j.aca.2006.12.036>.
- (107) Marini, F. *Chemometrics in Food Chemistry*, 1st ed.; Elsevier: Oxford, UK, 2013.
- (108) Joliffe, I.; Morgan, B. Statistical Methods in Medical Research. *Stat Methods Med Res* **1992**, 1, 69–95.
<https://doi.org/10.1177/096228029200100105>.
- (109) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: Chichester, UK, 2003.
- (110) Mevik, B.-H.; Wehrens, R. The PLS Package: Principal Component and Partial Least Squares Regression in R. *J. Stat. Softw.* **2007**, 18 (2), 1–23.
- (111) Brereton, R. G. *Applied Chemometrics for Scientists*; John Wiley & Sons, 2007.
- (112) Pearson, K. Principal Components Analysis. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1901**, 6 (2), 559.
- (113) Otto, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*; John Wiley & Sons: Chichester, UK, 2016.
- (114) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. A Review of Variable Selection Methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, 118, 62–69.
<https://doi.org/10.1016/j.chemolab.2012.07.010>.
- (115) Loh, W. Classification and Regression Trees. *WIREs Data Min. Knowl. Discov.* **2011**, 1 (1), 14–23.
<https://doi.org/10.1002/widm.8>.
- (116) Lewis, R. J.; Ph, D.; Street, W. C. An Introduction to Classification and Regression Tree (CART) Analysis. *2000 Annu. Meet. Soc. Acad. Emerg. Med.* **2000**, No. 310, 14p.
<https://doi.org/10.1.1.95.4103>.
- (117) Noble, W. S. What Is a Support Vector Machine? *Nat. Biotechnol.* **2006**, 24 (12), 1565–1567.
<https://doi.org/10.1038/nbt1206-1565>.
- (118) Christopher J.C. Burges. A Tutorial on Support Vector Machines for Pattern Recognition. *Data Min. Knowl. Discov.* **1998**, 2, 121–

167. <https://doi.org/10.1023/A:1009715923555>.
- (119) Andrade-Garda, J. *Basic Chemometric Techniques in Atomic Spectroscopy*, 1st ed.; Royal Society of Chemistry: Cambridge, 2013.
- (120) Boser, B. E.; Guyon, I. M.; Vapnik, V. N. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the Fifth Annual Workshop on Computational Learning Theory; COLT '92*; Association for Computing Machinery: New York, NY, USA, 1992; pp 144–152.
<https://doi.org/10.1145/130385.130401>.
- (121) MultiD Analyses BD. Genex <https://multid.se/genex/> (accessed Nov 24, 2021).
- (122) Eigenvector Research, I. PLS Toolbox. Manson, WA USA 2017.
- (123) Nørgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. Interval Partial Least-Squares Regression (IPLS): A Comparative Chemometric Study with an Example from near-Infrared Spectroscopy. *Appl. Spectrosc.* **2000**, *54* (3), 413–419. <https://doi.org/10.1366/0003702001949500>.
- (124) Kvalheim, O. M. Variable Importance: Comparison of Selectivity Ratio and Significance Multivariate Correlation for Interpretation of Latent-Variable Regression Models. *J. Chemom.* **2020**, *34* (4), 1–10. <https://doi.org/10.1002/cem.3211>.
- (125) Rajalahti, T.; Arneberg, R.; Berven, F. S.; Myhr, K. M.; Ulvik, R. J.; Kvalheim, O. M. Biomarker Discovery in Mass Spectral Profiles by Means of Selectivity Ratio Plot. *Chemom. Intell. Lab. Syst.* **2009**, *95* (1), 35–48.
<https://doi.org/10.1016/j.chemolab.2008.08.004>.
- (126) Favilla, S.; Durante, C.; Vigni, M. L.; Cocchi, M. Assessing Feature Relevance in NPLS Models by VIP. *Chemom. Intell. Lab. Syst.* **2013**, *129*, 76–86.
<https://doi.org/10.1016/j.chemolab.2013.05.013>.
- (127) Eriksson, L.; Byrne, T.; Johansson, E.; Trygg, J.; Vikström, C. *Multi- and Megavariate Data Analysis Basic Principles and Applications*, 3rd ed.; Umetrics Academy, 2013; Vol. 1.
- (128) Cocchi, M.; Biancolillo, A.; Marini, F. Chemometric Methods for Classification and Feature Selection. In *Comprehensive*

Analytical Chemistry; Elsevier, 2018; Vol. 82, pp 265–299.

- (129) Wold, S.; Johansson, E.; Cocchi, M. PLS: Partial Least Squares Projections to Latent Structures. In *3D-QSAR in drug design Theory, methods and applications.*; H, K., Ed.; Kluwer ESCOM Science Publisher, 1993; pp 523–550.
- (130) Cai, W.; Li, Y.; Shao, X. A Variable Selection Method Based on Uninformative Variable Elimination for Multivariate Calibration of Near-Infrared Spectra. *Chemom. Intell. Lab. Syst.* **2008**, *90* (2), 188–194. <https://doi.org/10.1016/j.chemolab.2007.10.001>.
- (131) Haghi, R. K.; Yang, J.; Tohidi, B. Fourier Transform Near-Infrared (FTNIR) Spectroscopy and Partial Least-Squares (PLS) Algorithm for Monitoring Compositional Changes in Hydrocarbon Gases under In Situ Pressure. *Energy & Fuels* **2017**, *31* (9), 10245–10259. <https://doi.org/10.1021/acs.energyfuels.7b01677>.
- (132) Barbosa, M. F.; Santos, J. R. B.; Silva, A. N.; Soares, S. F. C.; Araujo, M. C. U. A Cheap Handheld NIR Spectrometric System for Automatic Determination of Methane, Ethane, and Propane in Natural Gas and Biogas. *Microchem. J.* **2021**, *170* (106752). <https://doi.org/10.1016/j.microc.2021.106752>.
- (133) Rohwedder, J. J. R.; Pasquini, C.; Fortes, P. R.; Raimundo, I. M.; Wilk, A.; Mizaikoff, B. IHWG-MNIR: A Miniaturised near-Infrared Gas Sensor Based on Substrate-Integrated Hollow Waveguides Coupled to a Micro-NIR-Spectrophotometer. *Analyst* **2014**, *139* (14), 3572. <https://doi.org/10.1039/c4an00556b>.
- (134) Yusop Nurida, M.; Norfadilah, D.; Siti Aishah, M. R.; Zhe Phak, C.; Saleh, S. M. Monitoring of CO₂ Absorption Solvent in Natural Gas Process Using Fourier Transform Near-Infrared Spectrometry. *Int. J. Anal. Chem.* **2020**, *2020*. <https://doi.org/10.1155/2020/9830685>.
- (135) Makhoukhi, N. Determination of the Composition and the Gross Heating Value of a Mixture of Gases by Infrared Spectroscopy and Chemometric Methods. In *International Gas Union 23rd World Gas Conference 2006*; 2006; Vol. 5, pp 2731–2740.
- (136) Halstead, J. E.; Smith, J. A.; Carter, E. A.; Lay, P. A.; Johnston, E. L. Assessment Tools for Microplastics and Natural Fibres

- Ingested by Fish in an Urbanised Estuary. *Environ. Pollut.* **2018**, *234*, 552–561. <https://doi.org/10.1016/j.envpol.2017.11.085>.
- (137) Jiang, Y.; Yang, F.; Zhao, Y.; Wang, J. Greenland Sea Gyre Increases Microplastic Pollution in the Surface Waters of the Nordic Seas. *Sci. Total Environ.* **2020**, *712*. <https://doi.org/10.1016/j.scitotenv.2019.136484>.
- (138) Huang, Y.; Chapman, J.; Deng, Y.; Cozzolino, D. Rapid Measurement of Microplastic Contamination in Chicken Meat by Mid Infrared Spectroscopy and Chemometrics: A Feasibility Study. *Food Control* **2020**, *113*. <https://doi.org/10.1016/j.foodcont.2020.107187>.
- (139) Fang, C.; Luo, Y.; Zhang, X.; Zhang, H.; Nolan, A.; Naidu, R. Identification and Visualisation of Microplastics via PCA to Decode Raman Spectrum Matrix towards Imaging. *Chemosphere* **2022**, *286*. <https://doi.org/10.1016/j.chemosphere.2021.131736>.
- (140) Wander, L.; Vianello, A.; Vollertsen, J.; Westad, F.; Braun, U.; Paul, A. Exploratory Analysis of Hyperspectral FTIR Data Obtained from Environmental Microplastics Samples. *Anal. Methods* **2020**, *12* (6), 781–791. <https://doi.org/10.1039/c9ay02483b>.
- (141) Michel, A. P. M.; Morrison, A. E.; Preston, V. L.; Marx, C. T.; Colson, B. C.; White, H. K. Rapid Identification of Marine Plastic Debris via Spectroscopic Techniques and Machine Learning Classifiers. *Environ. Sci. Technol.* **2020**, *54* (17), 10630–10637. <https://doi.org/10.1021/acs.est.0c02099>.
- (142) Jin, N.; Song, Y.; Ma, R.; Li, J.; Li, G.; Zhang, D. Characterization and Identification of Microplastics Using Raman Spectroscopy Coupled with Multivariate Analysis. *Anal. Chim. Acta* **2022**, *1197*. <https://doi.org/10.1016/j.aca.2022.339519>.
- (143) Zvekcic, M.; Richards, L. C.; Tong, C. C.; Krogh, E. T. Characterizing Photochemical Ageing Processes of Microplastic Materials Using Multivariate Analysis of Infrared Spectra. *Environ. Sci. Process. Impacts* **2022**, *24* (1), 52–61. <https://doi.org/10.1039/d1em00392e>.
- (144) Luo, Y.; Zhang, X.; Zhang, Z.; Naidu, R.; Fang, C. Dual-

- Principal Component Analysis of the Raman Spectrum Matrix to Automatically Identify and Visualize Microplastics and Nanoplastics. *Anal. Chem.* **2022**, *94* (7), 3150–3157. <https://doi.org/10.1021/acs.analchem.1c04498>.
- (145) Luo, Y.; Sobhani, Z.; Zhang, Z.; Zhang, X.; Gibson, C. T.; Naidu, R.; Fang, C. Raman Imaging and MALDI-MS towards Identification of Microplastics Generated When Using Stationery Markers. *J. Hazard. Mater.* **2022**, *424*. <https://doi.org/10.1016/j.jhazmat.2021.127478>.
- (146) Battulga, B.; Kawahigashi, M.; Oyuntsetseg, B. Characterization of Biofilms Formed on Polystyrene Microplastics (PS-MPs) on the Shore of the Tuul River, Mongolia. *Environ. Res.* **2022**, *212*. <https://doi.org/10.1016/j.envres.2022.113329>.
- (147) Marchesi, C.; Rani, M.; Federici, S.; Alessandri, I.; Vassalini, I.; Ducoli, S.; Borgese, L.; Zacco, A.; Núñez-Delgado, A.; Bontempi, E.; Depero, L. E. Quantification of Ternary Microplastic Mixtures through an Ultra-Compact near-Infrared Spectrometer Coupled with Chemometric Tools. *Environ. Res.* **2023**, *216*. <https://doi.org/10.1016/j.envres.2022.114632>.
- (148) Wu, C.; Ma, Y.; Wang, D.; Shan, Y.; Song, X.; Hu, H.; Ren, X.; Ma, X.; Cui, J.; Ma, Y. Integrated Microbiology and Metabolomics Analysis Reveal Plastic Mulch Film Residue Affects Soil Microorganisms and Their Metabolic Functions. *J. Hazard. Mater.* **2022**, *423*. <https://doi.org/10.1016/j.jhazmat.2021.127258>.
- (149) Zhang, W.; Feng, W.; Cai, Z.; Wang, H.; Yan, Q.; Wang, Q. A Deep One-Dimensional Convolutional Neural Network for Microplastics Classification Using Raman Spectroscopy. *Vib. Spectrosc.* **2023**, *124*. <https://doi.org/10.1016/j.vibspec.2022.103487>.
- (150) Paul, A.; Wander, L.; Becker, R.; Goedecke, C.; Braun, U. High-Throughput NIR Spectroscopic (NIRS) Detection of Microplastics in Soil. *Environ. Sci. Pollut. Res.* **2019**, *26* (8), 7364–7374. <https://doi.org/10.1007/s11356-018-2180-2>.
- (151) da Silva, V. H.; Murphy, F.; Amigo, J. M.; Stedmon, C.; Strand, J. Classification and Quantification of Microplastics (<100 Mm) Using a Focal Plane Array–Fourier Transform Infrared Imaging System and Machine Learning. *Anal. Chem.* **2020**, *92* (20),

- 13724–13733. <https://doi.org/10.1021/acs.analchem.0c01324>.
- (152) Huang, H.; Sun, Z.; Liu, S.; Di, Y.; Xu, J.; Liu, C.; Xu, R.; Song, H.; Zhan, S.; Wu, J. Underwater Hyperspectral Imaging for in Situ Underwater Microplastic Detection. *Sci. Total Environ.* **2021**, 776. <https://doi.org/10.1016/j.scitotenv.2021.145960>.
- (153) Serranti, S.; Fiore, L.; Bonifazi, G.; Takeshima, A.; Takeuchi, H.; Kashiwada, S. Microplastics Characterization by Hyperspectral Imaging in the SWIR Range. In *Proceedings of SPIE - The International Society for Optical Engineering; Department of Chemical Engineering, Materials and Environment (DICMA), Sapienza - University of Rome, Via Eudossiana 18, Rome, 00184, Italy, 2019; Vol. 11197.* <https://doi.org/10.1117/12.2542793>.
- (154) Owen, S.; Cureton, S.; Szuhan, M.; McCarten, J.; Arvanitis, P.; Ascione, M.; Truong, V. K.; Chapman, J.; Cozzolino, D. Microplastic Adulteration in Homogenized Fish and Seafood - a Mid-Infrared and Machine Learning Proof of Concept. *Spectrochim. Acta - Part A Mol. Biomol. Spectrosc.* **2021**, 260. <https://doi.org/10.1016/j.saa.2021.119985>.
- (155) Jiao, M.; Wang, Y.; Li, T.; Li, R.; Liu, B. Riverine Microplastics Derived from Mulch Film in Hainan Island: Occurrence, Source and Fate. *Environ. Pollut.* **2022**, 312. <https://doi.org/10.1016/j.envpol.2022.120093>.
- (156) Yan, X.; Cao, Z.; Murphy, A.; Qiao, Y. An Ensemble Machine Learning Method for Microplastics Identification with FTIR Spectrum. *J. Environ. Chem. Eng.* **2022**, 10 (4). <https://doi.org/10.1016/j.jece.2022.108130>.
- (157) da Silva, D. J.; Wiebeck, H. ATR-FTIR Spectroscopy Combined with Chemometric Methods for the Classification of Polyethylene Residues Containing Different Contaminants. *J. Polym. Environ.* **2022**, 30 (7), 3031–3044. <https://doi.org/10.1007/s10924-022-02396-3>.
- (158) Chen, W.; Ouyang, Z.-Y.; Qian, C.; Yu, H.-Q. Induced Structural Changes of Humic Acid by Exposure of Polystyrene Microplastics: A Spectroscopic Insight. *Environ. Pollut.* **2018**, 233, 1–7. <https://doi.org/10.1016/j.envpol.2017.10.027>.
- (159) shan, J.; Zhao, J.; Liu, L.; Zhang, Y.; Wang, X.; Wu, F. A Novel

- Way to Rapidly Monitor Microplastics in Soil by Hyperspectral Imaging Technology and Chemometrics. *Environ. Pollut.* **2018**, 238, 121–129. <https://doi.org/10.1016/j.envpol.2018.03.026>.
- (160) Comnea-Stancu, I. R.; Wieland, K.; Ramer, G.; Schwaighofer, A.; Lendl, B. On the Identification of Rayon/Viscose as a Major Fraction of Microplastics in the Marine Environment: Discrimination between Natural and Manmade Cellulosic Fibers Using Fourier Transform Infrared Spectroscopy. *Appl. Spectrosc.* **2017**, 71 (5), 939–950. <https://doi.org/10.1177/0003702816660725>.
- (161) Bonifazi, G.; Fiore, L.; Pelosi, C.; Serranti, S. Evaluation of Plastic Packaging Waste Degradation in Seawater and Simulated Solar Radiation by Spectroscopic Techniques. *Polym. Degrad. Stab.* **2023**, 207. <https://doi.org/10.1016/j.polymdegradstab.2022.110215>.
- (162) Ai, W.; Liu, S.; Liao, H.; Du, J.; Cai, Y.; Liao, C.; Shi, H.; Lin, Y.; Junaid, M.; Yue, X.; Wang, J. Application of Hyperspectral Imaging Technology in the Rapid Identification of Microplastics in Farmland Soil. *Sci. Total Environ.* **2022**, 807. <https://doi.org/10.1016/j.scitotenv.2021.151030>.
- (163) Vidal, C.; Pasquini, C. A Comprehensive and Fast Microplastics Identification Based on Near-Infrared Hyperspectral Imaging (HSI-NIR) and Chemometrics. *Environ. Pollut.* **2021**, 285. <https://doi.org/10.1016/j.envpol.2021.117251>.
- (164) Tian, M.; Morais, C. L. M.; Shen, H.; Pang, W.; Xu, L.; Huang, Q.; Martin, F. L. Direct Identification and Visualisation of Real-World Contaminating Microplastics Using Raman Spectral Mapping with Multivariate Curve Resolution-Alternating Least Squares. *J. Hazard. Mater.* **2022**, 422. <https://doi.org/10.1016/j.jhazmat.2021.126892>.
- (165) Tian, X.; Beén, F.; Bäuerlein, P. S. Quantum Cascade Laser Imaging (LDIR) and Machine Learning for the Identification of Environmentally Exposed Microplastics and Polymers. *Environ. Res.* **2022**, 212. <https://doi.org/10.1016/j.envres.2022.113569>.

**Chapter 2: Improved
Sensitivity of Natural Gas
Infrared Measurements Using a
Filling Gas.**



Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas

Borja Ferreiro,[†] Jose M. Andrade,^{*,†} Carlota Paz-Quintáns,[†] Purificación López-Mahía,[†] Soledad Muniategui-Lorenzo,[†] María Rey-Garrote,[‡] Cristina Vázquez-Padín,[‡] and Carlos Vales[‡]

[†]QANAP Research Group, Analytical Chemistry and Facultade de Ciencias, University of A Coruña, Campus da Zapateira s/n, 15071 A Coruña, Spain

[‡]REGANOSA (Regasificadora del Noroeste, S.A.), Punta Promontoiro s/n, Mugardos, 15071 A Coruña, Spain

ABSTRACT: European trends to get greener cities and protect the environment imply substituting traditional diesel/gasoline engines for gas (gas-hybrid) powered engines. To accomplish this, straightforward quality control of liquefied and/or compressed natural gas is needed. This communication shows that the broadening effect of an auxiliary inert gas (Ar) enhances their infrared (IR) gaseous spectra and improves usual analytical performance parameters by 50%, which paves the way to use IR routinely to assess the composition of natural gas samples.

1. INTRODUCTION

The direct measurement of gaseous samples is becoming increasingly important due to society's needs for monitoring atmospheric pollutants,¹ industrial applications,² and the steady commercial use of energetic gaseous resources (biogas, natural gas, hydrogen, etc.).

Focusing on the latter issue, a typical field where accurate and fast measuring methods are needed is the natural gas industry. The European Commission seeks to replace current oil-derived fuels for alternative fuels, such as compressed natural gas, electricity, hydrogen, etc. In particular, when heavy-duty vehicles and maritime transport are considered, the EU objective is to change traditional diesel engines for liquefied natural gas (LNG) propellers, and it was planned to deploy the corresponding facilities on maritime ports and several points on land along the Trans-European Transport Networks by the end of year 2025.³ Also relevant, the trend for changing traditional personal vehicles to alternative fuel vehicles (considering both compressed natural gas and gas-hybrid engines) is gaining momentum rapidly as most European capitals strongly limit the circulation of traditional vehicles through their city centers.

The key parameters of natural gas are important to know for transportation and efficient consumption.⁴ Therefore, the properties of natural gas must be determined in order for large-scale shipments to be used at energetic facilities, particularly for LNG energy custody transfer purposes. To achieve this goal, it is critical to accurately measure the composition of the gas so its properties can be assessed. A paradigmatic example is the determination of the “methane number” (MN), whose concept is analogue to the octane number in gasolines or the cetane number in diesel oils.⁵ It is used to determine the “knock” rating of the natural gas when used in a motor (“engine knocking” occurs when small pockets of air and fuel ignite irregularly out of the optimum moment in the cylinder, potentially damaging engines⁶). As an alternative to the use of engines, this property is estimated using the chemical

composition of the natural gas, although the equations themselves are a matter of debate.⁶

The most used technique for gas quality assessment is gas chromatography (GC), combined with detection by either mass spectrometry (MS)⁷ or, more usually, thermal conductivity (ISO 6974-5:2014⁸). GC allows for highly accurate quantitative measurements, although it has several drawbacks when used offline (the size of the instruments, although modern in-line CG devices became very small, the need for dedicated installations, supply of auxiliary gases, the need for a sampling stage, etc.). Several alternatives were proposed instead, such as pellistors, semiconductor gas sensors, or electrochemical devices; however, they still have several problems, such as cross-responses, sensitivity to humidity, and drifts.⁹

In contrast, optical measuring systems, specifically Fourier transform infrared (FTIR) spectrometry, do not require special setups and additional reagents, are rather cheap, fast, and very stable, and yield high turnaround times (some practical details are given in the last section of this paper). Hence, they seem an appealing alternative for measuring gases both for industrial and/or environmental applications.

The main disadvantage of FTIR when analyzing natural gas samples is the complexity of the resulting spectra and so their detailed chemical interpretation. This is because the spectral bands of the components (from methane to hexane plus some isomers) overlap hugely because, after all, the unique functional groups are those related to C–H moieties. In such cases, quantitative analysis of the samples requires multivariate chemometric algorithms.^{10–12} However, for these methodologies to be really useful, we need to first measure the best possible IR spectra.

A way to get this is to enhance the weak IR spectral bands exploiting the “pressure broadening effect”. Reported as early as in 1965,¹³ this effect arises as a consequence of the collision

Received: February 22, 2019

Revised: July 5, 2019

Published: July 5, 2019

between molecules, which degenerates the energetic rotational and vibrational levels of the molecules, and it also reduces the lifetime of the excited states of the molecules.¹⁴ When the instrument operates at resolutions wider than the theoretical individual absorption line widths, as it is common for most commercial spectrophotometers, the polychromatic limitation of the Beer's law makes the absorption band to be weak. However, intermolecular collisions degenerate the transitions (absorption widths become broader with increasing pressure), enhancing the signal. The increase depends on the total pressure within the cell, the wavenumber, and, to some extent, the accompanying inert gas.¹³

Some authors applied this effect successfully using nitrogen as the filling gas, although without detailed explanations in the final reports.¹⁰ Varanasi et al.¹⁵ studied the broadening effect when analyzing the ν_3 -fundamental band (asymmetric stretching, ca. 3019 cm^{-1}) in methane standards using H_2 and He. The broadening of the P-branch of the same band due to N_2 , O_2 , H_2 , He, Ar, and CO_2 was also studied by Es-sebbar and Farooq.¹⁶ McDermitt et al.¹⁷ considered N_2 , O_2 , H_2 , and other filling gases when measuring atmospheric CO_2 . More recently, Frankenberg et al. studied the broadening of the NIR bands of methane using N_2 as the filling gas.¹⁸ However, these applications do not deal with routine natural gas measurements at an industrial scale.

The objective of this paper is to develop an FTIR-based analytical procedure that maximizes the signals recorded for LNG samples in industrial laboratories. This is the first step in an ongoing effort to set mid-IR spectrometry as a simple-to-use and fast framework to predict the chemical composition of liquefied natural gases. To set the analytical procedure, the influence of three common and readily available gases (nitrogen, helium, and argon) as pressurizing media in FTIR measurements is studied.

2. MATERIALS AND METHODS

2.1. Equipment. All measurements were made with an 8400S Shimadzu FTIR spectrometer and a 10 cm fixed-path gas cell. The cell setup required customizing a temperature-controlled, 10 cm path, stainless steel Harrick gas cell (Harrick Scientific, USA) with 2 mm-thick, 25 mm-diameter ZnSe windows. Figure 1 shows a conceptual

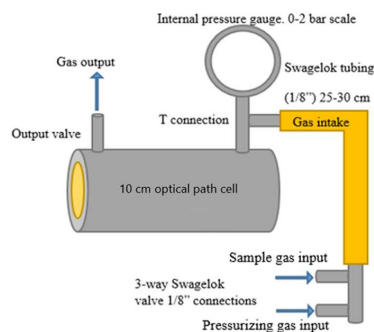


Figure 1. Scheme of the sample cell and its setup.

scheme of the cell setup designed in our laboratory ready for use (we added two external Viton O-rings to assure tightness without compromising the optical path). The exhaust tube of the gas cell was connected to a laboratory fume cupboard, which helps evacuating the cell. The software used throughout was the Shimadzu IR Solutions 1.30. The operating conditions of the spectrometer are given in Table 1.

2.2. Reagents and Samples. All pressurizing gases (Ar, He, and N_2) were obtained from Carburros Metálicos (Barcelona, Spain) with

Table 1. Spectrometer Operational Settings

mode	absorbance
apodization	Happ–Genzel
scans/spectrum	200
nominal resolution	8 cm^{-1}
range	5500–480 cm^{-1}
background ^a	Ar/He/ N_2 0.5 bar

^aThe corresponding filling gas was used to record the background.

99.9992% purity (the terms “filling”, “pressurizing”, or “broadening” are used here as synonymous).

The sample used to perform the studies was a standard developed to mimic a natural gas mixture (ready for its final use) provided by Reganosa (Regasificadora del Noroeste, Mugardos, Spain). Its composition is shown in Table 2.

Table 2. Composition (as %) of the Natural Gas Standard Used To Check the Pressure Effect

component	concentration (%)
N_2	1.002
CO_2	0.500
methane	92.000
ethane	4.997
propane	1.000
<i>i</i> -butane	0.150
<i>n</i> -butane	0.200
neopentane	0.000
<i>i</i> -pentane	0.049
<i>n</i> -pentane	0.049
<i>n</i> -hexane	0.049

3. RESULTS AND DISCUSSION

The influence of each broadening gas was studied by comparing a series of spectra obtained at different total pressures (between 0.2 and 1.8 bar; see Table 3). All experiments were performed at

Table 3. Gas Mixtures Used in the Experiments (Given as the Partial Pressure of each Gas)^a

sample (bar)	added Ar, He, or N_2 (bar)
0.2	0.0
	0.3
	0.8
	1.3
	1.6

^aThe total final pressure was the sum of both pressures.

20 ± 1 °C in an air-conditioned room (humidity: 35 ± 2 %). The protocol used to obtain the spectra had five steps (in all the experiments, the broadening gas and the samples were already thermostated at room temperature, and no further stabilization was required):

1. Purge the system with the broadening gas for 1 min (using a 0.5 bar pressure, the background did not change after this time).
2. Record the background with the pressurizing gas at 0.5 bar.
3. Purge with the natural gas sample for 1 min.
4. Measure the sample at 0.2 bar.

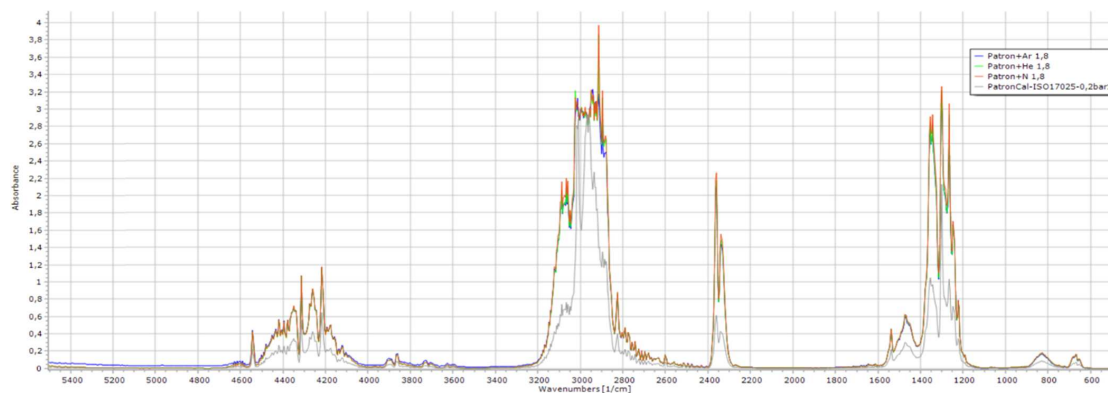


Figure 2. Superposition of the spectra of a 0.2 bar standard (gray), without the filling gas, and the same standard with three pressurizing gases (total pressure: 1.8 bar): N₂ (red), He (green), and Ar (blue).

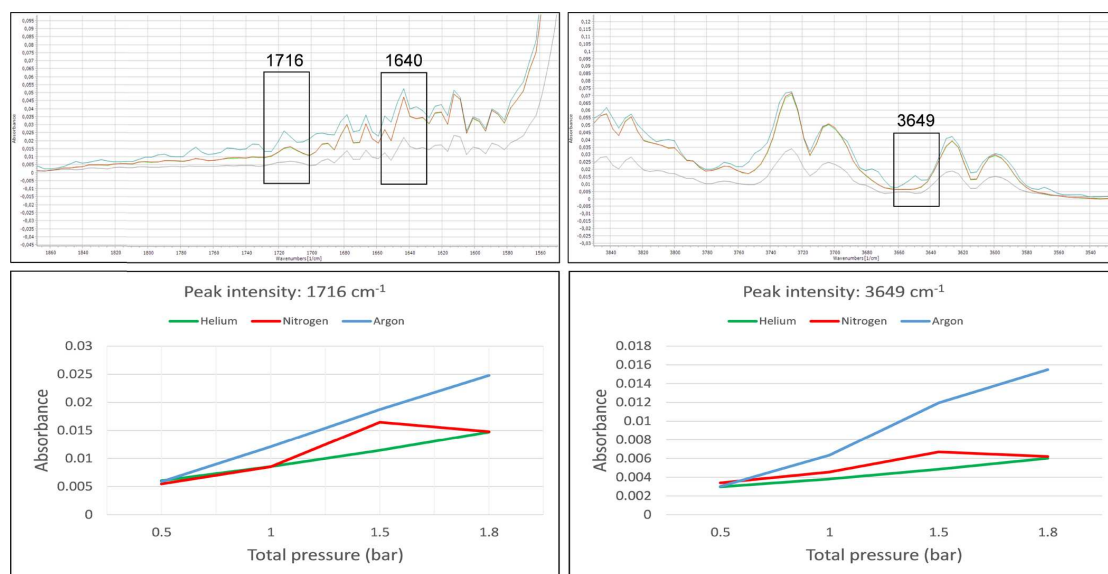


Figure 3. Examples of IR bands whose intensity increased most after adding the broadening gases—He (green), N₂ (red), and Ar (blue)—to the natural gas sample (top subplots) and graphical resume of the absorbance recorded as a function of the total pressure (lower subplots).

5. Add the broadening gas to the sample up to the target pressure (Table 3) and record the spectrum of the mixture.

The spectra thus obtained showed a clear general increment on their peak intensities, larger as the pressure increased. In Figure 2, the spectrum of the original natural gas sample is compared to the spectra obtained after adding the broadening gases (1.8 bar total pressure). The broadening effects are different for each band and also vary from the center of the band to its sides.¹⁹ They were more dramatic at the C–H bending region (ca. 1600–1200 cm⁻¹), where the bands at 1250 and 1350 cm⁻¹ doubled their absorbance. However, despite that the three filling gases yielded essentially the same increments in the most intense bands, there were differences at some weak bands (not visible in the general overview of Figure 2). The relevant

issue is that some of those bands may relate to specific vibrations of particular components of the sample, as a matter of example:

1. The band from 1820 to 1750 cm⁻¹ corresponds to the characteristic signal of methane (ν_2 , C–H asymmetric bending) although interfered by the C–H bending of other gas components (all typical ones but ethane).
2. Some peaks at the 3900 to 3500 cm⁻¹ range relate to methane ($3\nu_4$, third overtone of the C–H symmetric bending).²⁰
3. The peaks at the 1200 to 1120 cm⁻¹ range can be attributed to a suite of components: isopentane, isobutene, butane, and propane (as for comparison against NIST public databases²¹).

To assess the effects of the different broadening gases, some of the weakest recorded spectral peaks were chosen. These are

Table 4. Repeatability, Standard Deviation (within Brackets), and Relative Standard Deviations (%RSD, in Parentheses) at Different Peak Positions^a

band (cm ⁻¹)	sample gas at 0.2 bar, <i>n</i> = 7	sample + pressurizing gas at 1.8 bar, <i>n</i> = 7
4216.5	6.2 × 10 ⁻² [2.2 × 10 ⁻²] (3.1%)	5.3 × 10 ⁻² [1.9 × 10 ⁻²] (2.4%)
2824	5.3 × 10 ⁻² [1.9 × 10 ⁻²] (4.3%)	4.5 × 10 ⁻² [1.6 × 10 ⁻²] (3.1%)
1470	6.8 × 10 ⁻² [2.4 × 10 ⁻²] (7.8%)	3.1 × 10 ⁻² [1.1 × 10 ⁻²] (3.3%)

^a*r* and SD are in absorbance units (a.u.).

mostly secondary bands (P or R bands), with less intensity than the Q bands.¹⁶ As they originally have very small absorbance, it is relevant to improve the sensitivity of the measurements here by means of a broadening gas before proceeding with further quantitation studies. It was observed (Figure 3) that Ar yielded higher increments than the other two filling gases whenever very weak peaks (<0.05 absorbance units (a.u.) in the original natural gas sample) were considered.

For example, a ca. 3-fold improvement was obtained at 1640 cm⁻¹ (Figure 3, top left panel), a band that can be attributed to the C–H bending of methane.

Since this region may be influenced by atmospheric humidity, we made all efforts to avoid such a negative participation. First, samples do not contain water. They proceed from expanded liquefied natural gas (under controlled pressure and absence of air using conditions that avoid condensation). On the other hand, even if the broadening gases contain trace humidity, the signals are recorded in the background and subtracted automatically from the samples. Atmospheric air is also considered in the background because the extremes of the gas cell do not adjust tightly to the sample compartment walls. The air conditioning system of the room controls the humidity, so we were confident that no relevant changes occurred there.

Furthermore, some minute peaks, which do not appear while using other broadening gases (in the original sample, they would hardly be considered as “peaks”), became quite clear when Ar was selected instead (see augmented views at the top of Figure 3). This was attributed to the higher atomic mass of Ar,¹³ which makes the collisions between the molecules of the natural gas constituents and the broadening gas more efficient, and so, the degeneration of the energetic rotational and vibrational levels of the former.

To illustrate how these signal enhancements can improve the analytical performance of future measurements, precision (as repeatability) and the classical limits of detection were evaluated before and after using the broadening gas; see Table 4 (at the selected 1.8 bar total pressure).

To calculate the repeatability ($r = 2 \cdot SD \cdot \sqrt{2}$), seven repeats were measured on the same working session on the original natural gas sample and on the sample plus the filling gas mixture, and the SD was calculated at different points. The most intense signal was at 4216.5 cm⁻¹ (ca. 0.7 a.u.); that at 2824 cm⁻¹ was next (ca. 0.48 a.u.), and the weakest one was at 1470 cm⁻¹ (ca. 0.33 a.u.), all of them before using the broadening gas. It can be seen (Table 4) that the repeatability reduced when the filling gas was used up to a remarkable 50% when the weak signal at 1470 cm⁻¹ was considered. The RSDs (relative standard deviations, %) at those wavenumbers are not too different among them, although it is seen that the more intense the signal is, the better the relative precision becomes, as expected.

Table 5 resumes the classical limits of detection (LOD) of some natural gas components (high, medium, and low concentration levels). They were calculated considering the background noise as closely as possible to their peaks and the

Table 5. Classical Limits of Detection (as %) for Some Natural Gas Components^a

component (cm ⁻¹)	LOD ^b	LOD ^c
methane (2959 cm ⁻¹)	1.8 × 10 ⁻³ (7.9 × 10 ⁻³)	1.6 × 10 ⁻³ (3.1 × 10 ⁻²)
ethane (2365 cm ⁻¹)	2.2 × 10 ⁻⁵ (1.1 × 10 ⁻²)	1.0 × 10 ⁻⁵ (4.0 × 10 ⁻¹)
<i>n</i> -pentane (1470 cm ⁻¹)	4.2 × 10 ⁻⁶ (5.9)	2.1 × 10 ⁻⁶ (11.7)

^aThe slopes of the one-point calibrations are given in parentheses (a.u./% of component). ^bConsidering the natural gas sample alone. ^cConsidering the natural gas sample plus the filling gas.

slope of a one-point calibration line (the certified standard shown in Table 2).

Despite that this approach yields narrow calibration ranges, it is acceptable (indeed it can be better than other approaches, see ref²² and others cited therein) as long as the one-point calibration is used to quantify very close signals. Only those compounds for which a peak can be assigned unambiguously (by comparison with NIST spectra²¹) were considered. The calculations shown here correspond to the 1.8 bar total pressure situation (Table 2) as this was the setup selected finally.

The results show that, as expected, the LOD improved in all cases. Not too significantly for methane, because the enlargement of its intense associated band was small, but by a remarkable 50% for ethane and *n*-pentane (owing to the increase in the sensitivity on their weak bands, and so on the slope).

4. CONCLUSIONS

The use of a broadening, inert gas presents relevant advantages for the FTIR measurement of natural gas samples since the absorbance of otherwise weak bands (e.g., <0.05 a.u.) may double. This, in turn, increases the sensitivity and reduces both the repeatability and the limits of detection even by ca. 50%. Those improvements were more remarkable when the heaviest filling gas (argon) was employed as it even allowed the visualization of tiny (in the original sample) spectral characteristics. In addition, the presence of Ar in natural gas samples is negligible (contrary to N₂, which might amount up to 5.5%), and it is cheaper than He. Therefore, Ar seems to be the best choice to enhance the FTIR measurement of natural gas samples. All these characteristics suggest that the use of a filling (or broadening) gas to measure commercial natural gas samples opens ways for more sensitive determinations of their composition and, eventually, related properties. A positive point is that inert gases were employed satisfactorily. This contrasts with other proposals where H₂ and O₂ were suggested,¹⁶ which might present hazards concerns.

Finally, some comments on the practical and financial aspects of the chromatographic (GC) and infrared (FTIR) systems can be presented. The laboratory space occupied by the two systems is similar. The initial purchase costs are higher for the standardized GC system, and it needs more gas installments (ca. three times the cost of the FTIR device used here). The

turnaround time (offline analyses) is ca. 150 and 10 min for GC and FTIR, respectively (only instrumental analysis and basic processing; GC includes a blank, quality controls, and initial and final instrument verification). Routine GC measurements need a permanent carrier and detector gas (He), whereas FTIR only needs a small amount of Ar per sample (a bottle lasted for >100 samples). A major difference between GC and FTIR is due to staff dedication to calibrate and sample signal treatment (more exhaustive for GC). Roughly speaking, our estimations showed that a chromatographic quantification of hydrocarbons costs €600/sample, whereas less than €200/sample can be established for FTIR (equipment and amortization excluded; in FTIR, the cost of developing multivariate chemometric models was considered).

AUTHOR INFORMATION

Corresponding Author

*E-mail: andrade@udc.es. Fax: 34 981 167065.

ORCID

Jose M. Andrade: [0000-0003-1020-5213](https://orcid.org/0000-0003-1020-5213)

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The EU EMPIR (European Metrology Programme for Innovation and Research) Program in collaboration with EURAMET (European Association of National Metrology Institutes) is acknowledged for its financial support (grant 16ENG09). The Galician Government ("Xunta de Galicia") is acknowledged for its support to the QANAP group (Programa de Consolidación y Estructuración de Unidades de Investigación Competitiva. Ref. ED431C 2017/28) partially financed by FEDER (RDEF funds). Naturgy (Gas Natural Fenosa), Reganosa, and Mestrelab are acknowledged for their financial support to the UDC-QANAP research group. Three anonymous referees are acknowledged for their useful suggestions.

REFERENCES

- (1) Laj, P.; Klausen, J.; Bilde, M.; Plaß-Duelmer, C.; Pappalardo, G.; Clerbaux, C.; Baltensperger, U.; Hjorth, J.; Simpson, D.; Reimann, S.; Coheur, P.-F.; Richter, A.; De Mazières, M.; Rudich, Y.; McFiggans, G.; Torseth, K.; Wiedensohler, A.; Morin, S.; Schulz, M.; Allan, J. D.; Attié, J.-L.; Barnes, I.; Birmili, W.; Cammas, J. P.; Dommen, J.; Dorn, H.-P.; Fowler, D.; Fuzzi, S.; Glasius, M.; Granier, C.; Hermann, M.; Isaksen, I. S. A.; Kinne, S.; Koren, I.; Madonna, F.; Maione, M.; Massling, A.; Moehler, O.; Mona, L.; Monks, P. S.; Müller, D.; Müller, T.; Orphal, J.; Peuch, V.-H.; Stratmann, F.; Tanré, D.; Tyndall, G.; Abo Riziq, A.; Van Roozendaal, M.; Villani, P.; Wehner, B.; Wex, H.; Zardini, A. A. Measuring atmospheric composition change. *Atmos. Environ.* **2009**, *43*, 5351–5414.
- (2) Willer, U.; Saraji, M.; Khorsandi, A.; Geiser, P.; Schade, W. Near- and mid-infrared laser monitoring of industrial processes, environment and security applications. *Opt. Lasers Eng.* **2006**, *44*, 699–710.
- (3) European Commission. COM/2013/017, 24.1.2013 *Final Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions. Clean Power for Transport: A European alternative fuels strategy.* 2013.
- (4) Oliveira, E. C.; Aguiar, P. F. Data Reconciliation in the Natural Gas Industry: Analytical Applications. *Energy Fuels* **2009**, *23*, 3658–3664.
- (5) Malenshek, M.; Olsen, D. B. Methane number testing of alternative gaseous fuels. *Fuel* **2009**, *88*, 650–656.
- (6) Callahan, T. J.; King, S. R. Engine Knock Rating of Natural Gases—Methane Number. *J Eng Gas Turbines Power.* **1993**, *115*, 769–776.
- (7) Tang, X.; Li, Y.; Zhu, L.; Zhao, A.; Liu, J. On-line multi-component alkane mixture quantitative analysis using Fourier transform infrared spectrometer. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 371–377.
- (8) Technical Committee ISO/TC 193/SC 1. ISO 6974-5:2014. *Natural gas: Determination of composition and associated uncertainty by gas chromatography – Part 5: Isothermal method for nitrogen, carbon dioxide, C1 to C5 hydrocarbons and C6+ hydrocarbons;* 2014.
- (9) Hodgkinson, J.; Tatam, R. P. Optical gas sensing: a review. *Meas. Sci. Technol.* **2013**, *24*, 012004.
- (10) Fodor, G. E. Analysis of natural Gas by FT-IR; Calibrations & Validation. Interim Report to the US Army TARDEC Mobility Technology Center (TFLRF No. 324). 1997.
- (11) García-Mencia, M.; Andrade, J.; López-Mahía, P.; Prada, D. An empirical approach to update multivariate regression models intended for routine industrial use. *Fuel* **2000**, *79*, 1823–1832.
- (12) Makhoukhi, N.; Péré, E.; Creff, R.; Pouchan, C. Determination of the composition of a mixture of gases by infrared analysis and chemometric methods. *J. Mol. Struct.* **2005**, *744*, 855–859.
- (13) Smith, A. L. Infrared spectroscopy, in Kolthoff, I. M., Elving, P. J., Eds.; *Treatise on Analytical Chemistry (Part I, 2nd Edition, Vol 7, Section H, Chapter 5)*, Interscience (John Wiley and Sons, New York), 1981.
- (14) Nyquist, R. A. *Interpreting Infrared, Raman and Nuclear Magnetic Resonance Spectra.* Volume 1. Academic Press (San Diego). 2001.
- (15) Varanasi, P.; Sarangi, S.; Pugh, L. Measurements on the Infrared Lines of Planetary Gases at Low Temperatures. *Astrophys J.* **1973**, *197*, 977–982.
- (16) Es-sebbar, E.; Farooq, A. Intensities, broadening and narrowing parameters in the ν_3 band of methane. *J Quant Spectrosc Radiat Transf.* **2014**, *149*, 241–252.
- (17) McDermitt, D. K.; Welles, J. M.; Eckles, R. D. Effects of temperature, pressure and water vapor on gas phase infrared absorption by CO₂. *LI-COR Tech Publ.* **1993**, *116*, 1–5.
- (18) Frankenberg, C.; Warneke, T.; Butz, A.; Aben, I.; Hase, F.; Spietz, P.; Brown, L. R. Methane spectroscopy in the near infrared and its implication on atmospheric retrievals. *Atmos Chem Phys Discuss.* **2008**, *8*, 10021–10055 hal-00304207 (Open Source).
- (19) Aousgi, F.; Haddad, S.; Aroui, H. On the Pressure and Temperature Dependence of the Absorption Coefficient of NH₃. *Int J Spectrosc.* **2011**, 816548.
- (20) Doyennette, L.; Menard-Bourcin, F.; Menard, J.; Boursier, C.; Camy-Peyret, C. Vibrational energy transfer in methane excited to 2v₃ in CH₄-N₂/O₂ Mixtures from Laser-Induced Fluorescence Measurements. *J Phys Chem A.* **1998**, *102*, 3849–3855.
- (21) NIST Public IR Spectra Database. webbook.nist.gov/cgi/cbook.cgi. Consulted November 2018.
- (22) Andrade, J. M.; Terán-Baamonde, J.; Soto-Ferreiro, R. M.; Carlosena, A. Interpolation in the standard additions method. *Anal. Chim. Acta* **2013**, *780*, 13–19.

Chapter 3: New Ways for the Advanced Quality Control of Liquefied Natural Gas.



Article

New Ways for the Advanced Quality Control of Liquefied Natural Gas

Borja Ferreiro, Jose Andrade ^{*}, Carlota Paz-Quintáns, Purificación López-Mahía and Soledad Muniategui-Lorenzo

Group of Applied Analytical Chemistry, Campus da Zapateira, University of A Coruña, 15071 A Coruna, Spain; borja.ferreiro@udc.es (B.F.); carlota.paz@udc.es (C.P.-Q.); purmahia@udc.es (P.L.-M.); smuniat@udc.es (S.M.-L.)

* Correspondence: andrade@udc.es; Fax: +34-981167065

Abstract: Currently, gas chromatography is the most common analytical technique for natural gas (NG) analysis as it offers very precise results, with very low limits of detection and quantification. However, it has several drawbacks, such as low turnaround times and high cost per analysis, as well as difficulties for on-line implementation. With NG applications rising, mostly thanks to its reduced gaseous emissions in comparison with other fossil fuels, the necessity for more versatile, fast, and economic analytical methods has augmented. This work summarizes the latest advances to determine the composition and physico-chemical properties of regasified liquid natural gas, focusing on infrared spectroscopy-based techniques, as well as on data processing (chemometric techniques), necessary to obtain adequate predictions of NG properties.

Keywords: natural gas composition; natural gas analysis; quality control; FTIR; multivariate regression; PLS



Citation: Ferreiro, B.; Andrade, J.; Paz-Quintáns, C.; López-Mahía, P.; Muniategui-Lorenzo, S. New Ways for the Advanced Quality Control of Liquefied Natural Gas. *Energies* **2022**, *15*, 359. <https://doi.org/10.3390/en15010359>

Academic Editors: Lorenzo Ferrari and Andrea Baccioli

Received: 26 November 2021

Accepted: 29 December 2021

Published: 4 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The increasing number of new natural gas (NG) applications, such as in maritime transport and large trucks, demand faster and more versatile ways to determine its quality. Today, the standardized analytical characterization of the NG composition requires gas chromatography (GC). This is a well-established analytical methodology, capable of measuring accurately and reliably the NG components at very low concentrations (e.g., ca. 0.01% (v/v) [1]). On-line gas chromatographs are currently being implemented for routine measurements and quality control. They are fast (ca. 5 min/sample) but still rely on benchtop instruments for their validation (and frequent revalidation). That is time-consuming and not straightforward. In addition, contractual agreements usually focus on standardized benchtop chromatographs, whose overall running cost per sample is high [2–5].

In light of the rise of natural gas consumption in the last few decades (though mitigated recently by the COVID-19 pandemic) [6] and the new applications being proposed [7], it is expected that NG production will keep increasing for, at least, the next two decades [8], even considering the most optimistic predictions for the use of renewable energies.

This favors the development of new, faster, and versatile quality control methods. Some of them are based on physical measurements, such as combustion energy, speed of sound, or density, and they have been applied to determine the heating value [9], the lower heating value plus the Wobbe index (WI), and the stoichiometric air–fuel ratio [10], as well as the Methane number, WI, and CH₄, C₂H₆, and CO₂ concentrations [5,9–11]. Noteworthy, these methods require complex instrumentation.

Small electrochemical sensors offer a very appealing and convenient solution (likely, portable in a near future) for routine measurements, although they are still under development and have not been commercialized. Some of them are based on capacitive transducers whose signals depend on the presence (and quantity) of specific NG compositional gases [12–14].

In addition, different kinds of convenient spectroscopic measurements (although not fully portable, still) were proposed for NG quality control, mostly based on the use of infrared radiation [3]. They ranged from tunable-filter near infrared radiation [12] to common Fourier transform IR spectrometry (FTIR), both in the near region (NIR, 12500–4000 cm^{-1} , [15,16] because it contains the 1st and 2nd overtone regions –5700 and 8300 cm^{-1} , respectively, plus the 1st overtone of the combination mode of hydrocarbons, 7150 cm^{-1}), and in the medium region (MIR, 4000–450 cm^{-1}). Although NIR was initially suggested as an optimal spectral region, this has been shown not to be the case always and MIR provided very advantageous, as well [2].

Raman spectrometry has also been suggested to measure the chemical composition of gases [4,17,18] and was applied to evaluate the composition of NG and biogas [19] and fuel gas [20] (from CH_4 to C_4H_{10}).

It is worth noting that several difficulties arise when using IR and Raman spectrometry. In effect, as NG is composed of a variety of molecules, the resulting spectra are a complex combination of overlapping peaks and/or broad bands corresponding, more often than not, to several compounds (for example, the peaks around 2950 cm^{-1} correspond to the C-H stretching vibrational mode, which belongs to every aliphatic molecule in the gas). The use of chemometric methods has been demonstrated to be useful to circumvent this problem [2].

The main objective of this work is to offer tutorial guidelines on how to develop an advanced MIR-based quality control methodology to estimate NG quality control parameters. The approach detailed here is based on measuring gaseous infrared spectra of regasified liquefied natural gas (LNG) and coupling them with multivariate statistical modeling. Details are given on the experimental setup used to gather reliable results and the workflow to get a sound prediction model.

2. Materials and Methods

2.1. Instrumentation

A relevant issue when analyzing NG is the gas cell. The most important factors defining it are the pressure it can withstand, the optical length, and the material of the windows. The first may affect the seals and the cell windows since an excessive pressure may cause leaks and damage the usually expensive and fragile cell crystals (windows), also creating security concerns, as NG is flammable. In our experience, a sample pressure around 0.2 bar is enough to obtain well-defined spectra, although the signals are improved using additional an inert gas, as explained in Section 2.2.

The optical path impacts directly on the intensity of the spectral peaks (so, on the Lambert-Beer-Bouguer's law). Single-pass, multireflection, and hollow-core fiber cells can be used to study NG. The first type is very straightforward, as it only consists of a capsule (the simplest one made of glass) with IR-transparent windows, which the IR beam traverses linearly. Such a cell tends to be not very long, often not exceeding 10 cm, thus yielding a relatively short optical pathlength. More robust 10-cm single-pass cells are of stainless steel and might be thermostated, as well [21–23].

Multireflection cells contain a series of internal mirrors that reflect the IR incident beam, augmenting the pathlength. The limiting factor in these cells is the reflection coefficient of the mirrors, which determines the maximum number of reflections. Typical pathlengths can range between 1 and 15 m [21–24].

A hollow-core fiber cell consists of a hollow internal-reflection fiber filled with the gas (Figure 1). A laser beam (usually from a quantum cascade laser [25]) is focused on the hollow core space and travels inside the coil to the detector, with a pathlength equal to the longitude of the fiber. The main limitation is the loss of energy in the beam as it travels through, which is affected by both the length and the curvature of the fiber [24]. Typical pathlengths are between 0.1 and 5 m. Even though there have been some studies with methane samples [26], hollow-core fibers are not still of common usage; so, we will focus on traditional cells.

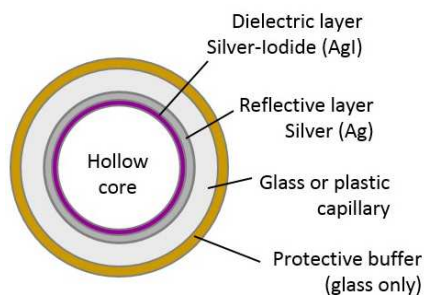


Figure 1. Basic diagram of a hollow-core fiber cell internal structure. Source: [guidingphotonics.com](https://www.guidingphotonics.com) (accessed on 20 November 2021) [24].

At first, multireflection cells may appear as a good choice because of the many reflections they have, which increase the S/N (signal to noise) ratio. However, they might not be the best option, as they are very expensive and need large sample quantities, not only for the measurement itself but for the previous purge required to avoid cross-contamination. In addition, major sample components might be an issue because their signals can saturate the spectrum. In addition, those cells are quite fragile because of their glass body, and the mirrors need regular realignment (not difficult, but time-consuming).

Figure 2 compares two spectra of a regasified LNG sample measured with a 10 cm pathlength single-pass cell and with a 1 m pathlength multireflection cell. It can be seen that even though the intensity is considerably higher for the latter (indeed, too high, with absorbances up to 2.5), the general spectral profiles are identical, and no peaks were lost when the short cell was used.

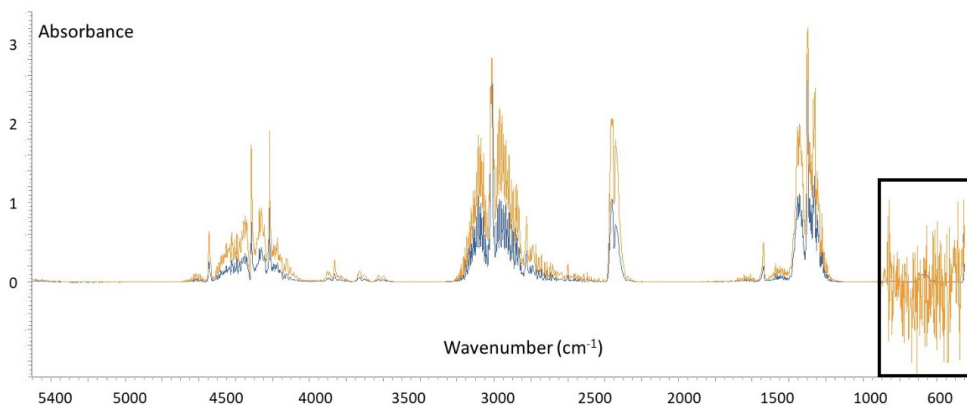


Figure 2. Comparison of spectra obtained using a multireflection gas cell (1 m pathlength) with CaF_2 windows (yellow) and a single-pass gas cell (10 cm) with ZnSe windows. The region in the box shows the influence of the windows (see text for details).

Nevertheless, for a simple and straightforward working protocol, the single-pass cell can (and should) be customized so that the filling and emptying of the cell become simple, and the backgrounds are recorded easily. Figure 3 shows an example that we have used quite successfully. It consists of a 10 cm path, stainless steel Harrick gas cell (Harrick Scientific, Pleasantville, NY, USA) with 2 mm-thick, 25 mm-diameter ZnSe windows, an input tube with a Swagelock 3-way valve, an exhaust tube, and an internal pressure gauge.

As for any traditional IR application, the material of the windows is a very important parameter, as it can reduce the range of wavenumbers in which the analysis can be performed. This is also exemplified in Figure 2. The spectra obtained for the short cell with ZnSe windows is well defined in the 500–950 cm^{-1} region, which is totally obscure

when CaF_2 windows were used in the multireflection cell. This region is relevant when studying hydrocarbons. For the sake of information, the most common window materials are summarized in Table 1, along with their effective wavenumbers working ranges [27–29].

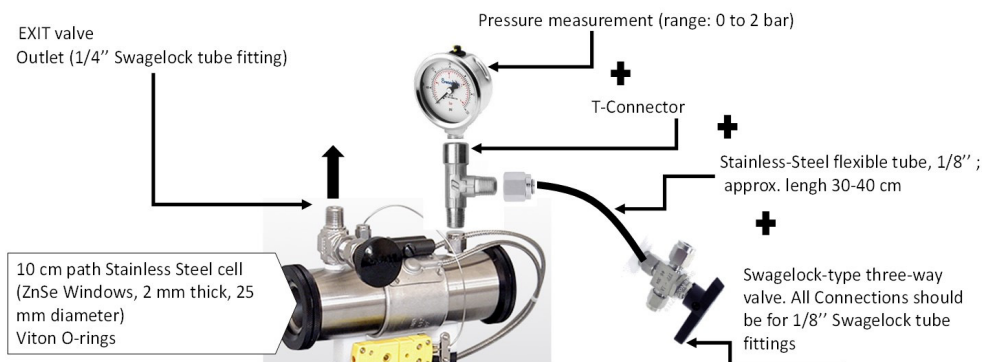


Figure 3. Conceptual scheme of the 10 cm-path gas cell tailored for measuring regasified LNG samples.

Table 1. Some common gas cell window materials and their transmission working range [27–29] (there may be slight discrepancies in the wavenumbers among the references).

Window Material	Effective Range (cm^{-1})	Window Material	Effective Range (cm^{-1})
AgCl	25,000–360	KBr	40,000–340
AMTIR (GeAsSe)	11,000–625	KRS-5 (TlBr + TlI)	16,600–250
BaF ₂	50,000–740	NaCl	40,000–625
CaF ₂	50,000–1025	Sapphire (Al ₂ O ₃)	50,000–1525
Chalcogenide (AsSeTe)	4000–900	Si	8000–660
CsI	33,000–200	SiO ₂	50,000–2500
Diamond	40,000–12.5	ZnS	17,000–690
Ge	5500–475	ZnSe	10,000–550

When selecting an instrument to perform IR measurements several important parameters need to be addressed. A first factor to take into consideration is the wavenumbers range, ideally from 5000 to 450 cm^{-1} so it includes the 3rd methane overtone around 4400 cm^{-1} and the aliphatic C-C skeletal vibrations around 600 cm^{-1} . A basic setup might be the one depicted in Figure 4, with a Shimadzu 8400S Spectrometer (Shimadzu, Kyoto, Japan) and the gas cell depicted in Figure 3.

A second issue is the nominal spectral resolution. In gas analysis, this is a particularly relevant parameter as gas spectra contain very narrow rotovibrational peaks very close to each other. Resolution can be defined as the wavenumber interval at which two consecutive datapoints are recorded. A high resolution means small intervals (closely adjacent intensity signals), more data, well defined spectra, and less overlap between peaks. However, it also decreases the speed of the analysis and reduces the S/N ratio, which implies that a higher number of scans is required to get a satisfactory spectrum [30]. A minimum 1 cm^{-1} resolution is required to resolve the signals; otherwise, the spectral bands will overlap in excess (see Figure 5). Such a resolution is easily attainable by most modern FTIR instruments, and they even offer much better ones (current benchtop instruments offer up to 0.25 cm^{-1} resolution). However, one must be sure that the wealth of data per spectrum can be retrieved and saved fast (more than 15,000 digitized values per spectrum for highest resolutions) and handled by statistical software, which is not always the case. So, depending on the available instruments and computers, a compromise might be needed.

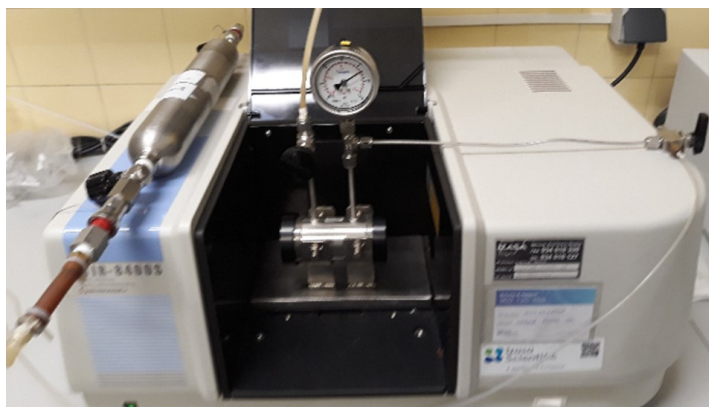


Figure 4. Instrumental setup with a gas cylinder connected to the 3-way valve so that the sample and the broadening gas can be introduced into the cell.

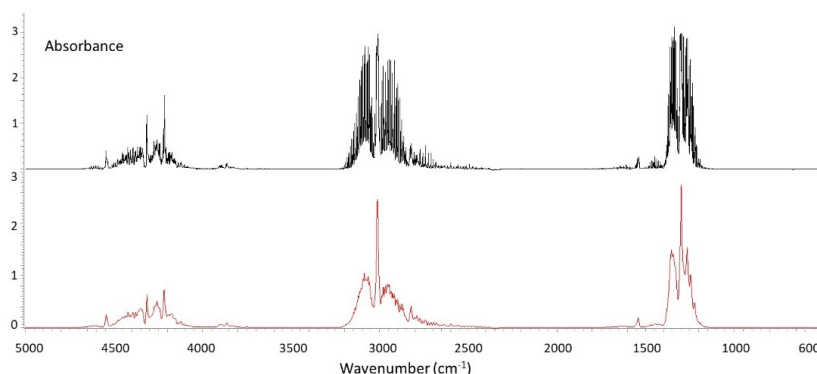


Figure 5. Effect of the nominal resolution on the IR natural gas spectra. Black: 1 cm^{-1} ; Red: 8 cm^{-1} resolution.

There are more thorough definitions of resolution, based on the Rayleigh or the full width at half-height (FWHH) criteria. The Rayleigh criterion states that two adjacent spectral lines (with a $\text{sinc}^2(x)$ form and equal intensity) “are considered to be just resolved when the center of one line is at the same frequency as the first zero value of the other” [31]. Visually, a slight dip (approximately 20%) between the two maxima can be seen. On the other hand, the FWHH criterion states that two triangular shaped lines are resolved when the spacing between the lines is greater than the FWHH of either line [31]. This latter definition, however, is not very appropriate for FTIR devices [31]. The nominal resolution is going to affect the required time of analysis, so a trade-off between time and quality of the spectra must be obtained.

A third issue is the number of scans recorded by the instrument to get the final spectrum. It influences the time required to retrieve the final spectrum, as well as its quality. As for any instrumental device, there is always a certain degree of variability in the signal due, in part, to the motion of the gaseous molecules in the sample, which causes many IR bands/peaks to be of low intensity [30], but also due to instrumental noise [32,33]. Too few scans will yield a spectrum with a poor S/N ratio, whereas a high amount of scans helps in removing this variability by averaging out random noisy signals, thus offering a more precise result. However, recall that noise gets only reduced by a factor of $1/\sqrt{n}$ (n = number of scans) [32,33]. The optimal number of scans depends on several factors, such as the physical state of the sample (gases need more scans than solids), temperature control of the room, externally caused bench vibrations, real-time processing capabilities

of the microprocessors, etc. In our experience (room temperature controlled to ± 1 °C), around 200 scans per spectrum offer an optimal result, though 100 scans yielded suitable final spectra many times.

A well-known point for all IR practitioners is the need to background-correct the spectra. However, the stressing environment of industrial laboratories, with tight delivering times, much workload, and adjusted personnel, may tempt analysts to measure gas samples in a ‘continuous mode’ without recording new backgrounds. Now, taking into account that 200 scans (1 sample) need ca. 15 min to be recorded (this obviously depends on the instrument being used and the number of scans and resolution), it is clear that the atmosphere of the laboratory is prone to fluctuate; thus, the background after 15, 30, . . . min will definitely not be constant. Therefore, it is critical to perform a background before every sample is measured, especially whenever extensive recording times are needed. Failure in doing so will lead to ‘tilted’ spectra and more intense peaks associated to environmental water, CO, and CO₂ (at least, but any organic volatile solvent from the laboratory can also be detected). This will be more dramatic as time goes by (Figure 6) and becomes very important as soon as after 60 min. A way to avoid this is to tightly close the space where the gas cell was installed and continuously purge with N₂, which is not always feasible as it happens in Figure 4.

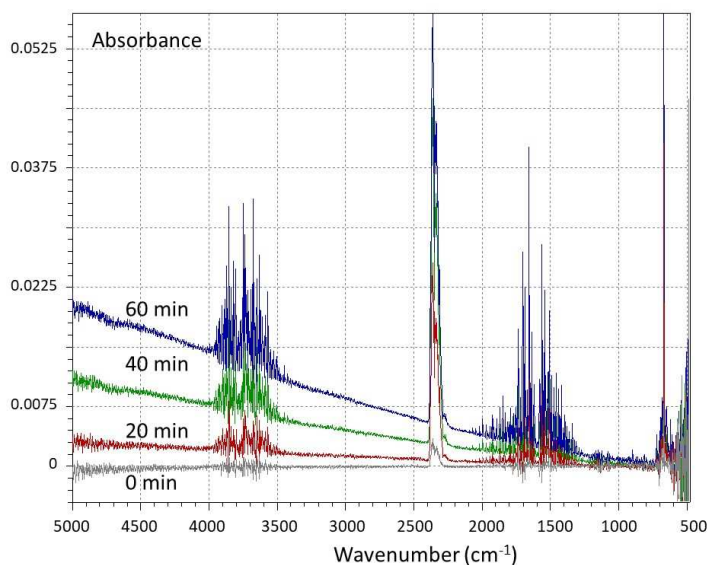


Figure 6. Baseline evolution of the spectrum of a sample during 1 h without background correction.

Finally, apodization is a mathematical operation applied after the Fourier transform that minimizes the appearance of secondary peaks at both sides of the actual peak [34]. For natural gas analysis, medium-strong functions, such as Happ-Genzel or Beer-Norton, are used commonly. The strongest the function is, the lower the secondary, undesired, peaks are, but, also, the required signal of interest decreases, as well. To select the best level of apodization preliminary trials have to be performed considering the various possibilities of the instrument-controlling software. Different vendors offer different algorithms, although the ‘strong’ functions are good many times. It is recommended to report the apodization function applied in your particular study and maintain the selection throughout all the measurements.

2.2. Gas Broadening

A typical problem when measuring LNG is that the product contains major constituents (>75%, *v/v* concentrations, e.g., methane) and minor components (<1%, *v/v* con-

centrations, e.g., butanes and pentanes), all of them being of importance for quality control purposes. Hence, there is some difficulty in obtaining representative signals for each compound, despite the overall spectrum has visually nice characteristics (low noise, well-shaped bands, etc.). A way to address this problem consists of boosting the less intense bands using a so-called broadening gas (or filling gas). That consists of an inert gas that, when added to the gas sample, enhances the signal of weakest peaks. The effect has been known since the first half of the 20th century [35–37] but not too applied in routine quality control measurements. Briefly, the molecules of the inert gas collide with those of the sample compounds, momentarily blocking the rotation of the atoms in the moment of the collision. This enhances the analytical signal of the rotational modes, increasing the intensity of the spectral peaks.

The operational procedure when dealing with a broadening gas is not difficult and we found nice results considering the next steps [38]: (i) purge the cell with the filling gas (0.5 bar, 1 min); (ii) record background; (iii) purge with a small flow of the sample gas; (iv) fill the cell with the sample (0.2 bar), and, then, add the broadening gas until a total pressure of 1.5 bar; and (v) record spectra.

In previous studies [38] a suite of inert gases (N_2 , He, and Ar) was evaluated to determine which improves most the sensitivity of selected weak bands when measuring commercial LNG samples. Out of them, Ar arose as the most adequate one to enhance the low intensity peaks (as shown in Figure 7), while N_2 and He just increased the signal of the overall spectra equally.

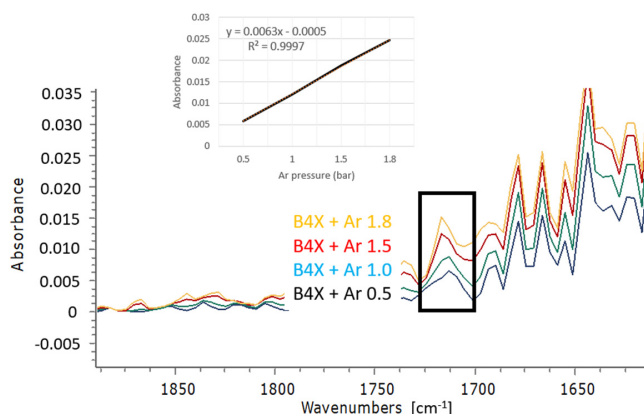


Figure 7. Exemplification of the effect of a broadening gas (Ar) on the intensity of a weak peak of an LNG sample (named B4X) at 1716 cm^{-1} (attributed to CH bending). The inset shows the ratio of variation of the signal as a function of the pressure of Ar. The pressure of the sample was always 0.2 bar, whereas the total pressure is indicated in the labels.

2.3. Peak Identification

Determining the components of LNG is a major issue for industries, customers, and interested parties. They are relevant to evaluate the combustion properties for its use in engines (measured experimentally by the motor methane number); for trade and energy custody activities and billing; and to evaluate the interchangeability of different gas mixtures (measured experimentally by the Wobbe index). IR spectra contain much information on the nature of the chemical constituents but, unfortunately, their spectral peaks overlap strongly because all of them have just unsubstituted hydrocarbon chains (methane, ethane, propane, butanes (i- and n-), pentanes, and, scarcely, hexane). Therefore, it is not easy to differentiate each constituent or characterize each and every compound by a unique distinctive wavenumber. Some chemical interpretation of the spectra can be obtained using reference databases, as that from NIST [39]. For example, the band at $2800\text{--}3200\text{ cm}^{-1}$ (Figure 8, blue rectangle) has contributions from all the aliphatic molecules, as it corresponds to the

C-H bond stretching, which is present in all components. The band at 1200–1380 cm^{-1} (Figure 8, green rectangle) corresponds only to methane, while the band at 1380–1550 cm^{-1} (Figure 8, red rectangle) has, again, contributions from all the other aliphatic compounds as it corresponds to the CH bending.

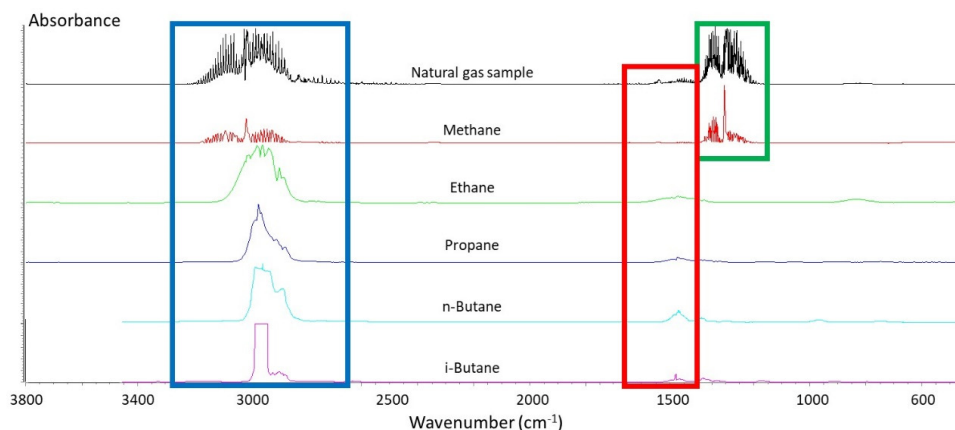


Figure 8. Comparison between the spectra of regasified LNG and its main aliphatic components (as stated in Reference [39]).

In the close-up image of the 1300–500 cm^{-1} region (Figure 9), the peaks corresponding to ethane and propane can be seen. The band between 800–750 cm^{-1} is mainly due to ethane, but for a small peak at 795 cm^{-1} corresponding to isobutane, which is not possible to appreciate visually in the LNG spectrum. For propane, the small peaks around 750 cm^{-1} (CH_2 and CH_3 , twisting, rocking), 950–900 cm^{-1} (CH_3 , CCH deformation), and 1075–1025 cm^{-1} (C-C, asymmetric stretching) can be seen. It is very difficult to visualize the butane peaks in the NG spectra due to its very low concentrations although, nevertheless, a small band can be ascertained at 1000–950 cm^{-1} (corresponding to n-butane), but it could only be observed in those samples with the highest concentrations of butane (usually not exceeding 0.4%). Table 2 compiles the usual positions (wavenumbers) of some relevant vibrational peaks of the constituents of a LNG sample.

2.4. Chemometric Predictive Models

The IR spectra of complex gas mixtures are too complicated to interpret by usual means. The large number of peaks and their interactions makes it almost impossible to apply simple calibration techniques and the typical Lambert-Beer-Bouguer's direct correlation is not feasible. Some of the minor components cannot be quantified directly using a specific peak, and even those with highest concentrations have a considerable number of overlapping spectral features, as seen at Figures 8 and 9. Thus, considering both the full spectrum and multivariate calibration to gather relevant information is a must.

Here, the statistical and data treatment methods included in the 'chemometrics' field take the lead. Unfortunately, introducing the various multivariate regression methods (even only the most frequent ones) would be too lengthy, and it is out of the scope of this manuscript. Interested readers are kindly directed to some introductory texts [45–48]. It is worth citing a few important methods, namely Principal Components regression (PCR), Locally Weighted Regression (LWR), Support Vector Machines Regression (SVM(R)), or Artificial Neural Networks (ANN). Noticeably, ANNs [49] and SVM [50] were applied for quality control of natural gas, though none was applied to determine its composition. Our focus here will concentrate on the commonest, widely applied, one: PLS (partial least squares) regression, which has become a standard [2,15,51,52]. As a matter of example, PLS regression allowed the prediction of a mechanical/engine property, the methane number [53] with average prediction errors (or RMSEP) < 0.2 [16], the concentrations of

ethane (error < 0.19 [16]), propane (error < 0.03 [2]), and iso and n-butane (error < 0.012 [16]), as well as the Wobbe index, a combustion property [2]. Multivariate regression models can be developed using different commercial software, such as, for instance, MATLAB's PLS_Toolbox [54], Grams [55], Unscrambler [56], GenEx [57], SPSS [58], or Statgraphics [59], to cite some common ones. Other packages are available as freeOpen Source, e.g., CAT [60].

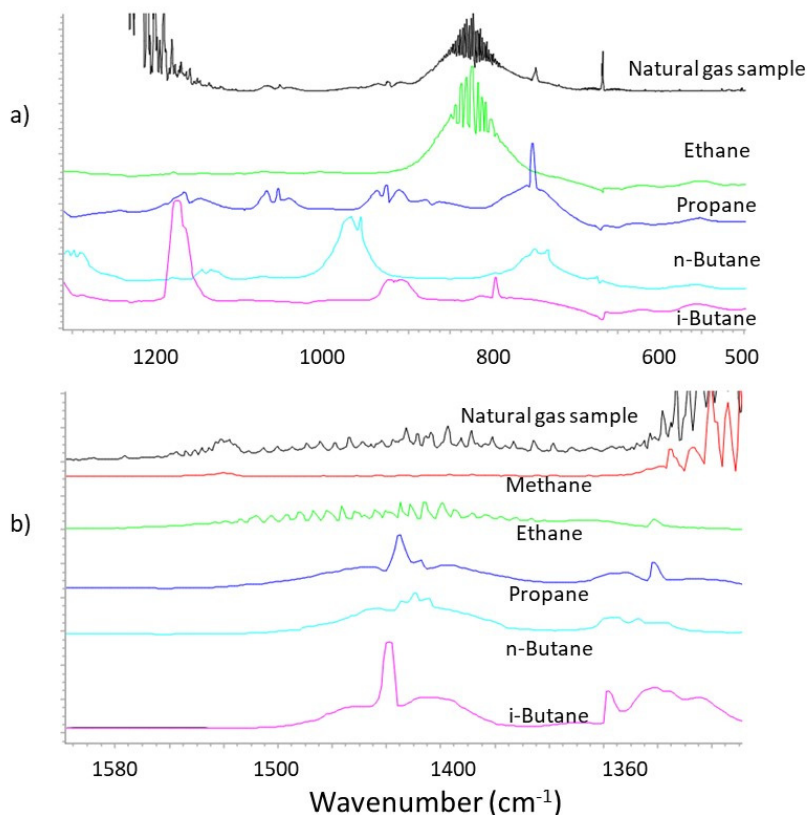


Figure 9. Close-up view of the 1300–500 cm⁻¹ (a) and 1600–1340 cm⁻¹ (b) regions of a LNG spectrum (depicted in Figure 8).

2.4.1. Samples

Every multivariate regression method requires a collection of samples to develop the model. In industrial environments, this can be achieved by collecting samples during a period broad enough to get a sound representation of the typical samples received/processed by the units. Different raw materials and finalized products are also welcome. This collection is then subdivided in, at least, two datasets: one for modeling and one for validating the proposed model (it is crucial to remember that the latter should always be used independently of the former). If enough samples are available, a third independent test set is advised, in order to determine the optimal number of factors. These samples have to be characterized using both the infrared procedure and the official analytical methods associated to the properties we want to predict. The absorbance of the different IR wavenumbers will act as the predictors within the models, whereas the results of the official methods will be the predictands (i.e., the values of the property to be predicted). As a rule of thumb, the proportion of samples should be around 70% for calibration, 20% for validation, and 10% for the test dataset. No doubt, it is strongly advised that all datasets represent the concentration working ranges (values of the property of interest).

Table 2. Collection of the most relevant MIR spectral peaks, and their chemical assignments, for typical hydrocarbon constituents of LNG.

Methane [40]		Ethane [41]		Propane [42,43]			
Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)
CH asymmetric stretch	3019	CH ₃ asymmetric stretch	2985	CH stretch	2957	CH ₃ & CH ₂ rocking	1186
CH symmetric stretch	2917	CH ₃ symmetric stretch	2895	CH stretch	2870	CH ₃ wagging, deformation	1155
CH symmetric bend	1543	CH ₃ Asymmetric deformation	1469	CH ₃ & CH ₂ scissoring	1466	C-C asymmetric stretch	1051
CH asymmetric bend	1311	CH Symmetric deformation	1379	CH ₃ & CH wagging (in phase)	1384	CH ₃ & CCH deformation	919
		CH ₃ rocking	821	CH ₃ & CH wagging (out of phase)	1368	C-C symmetric stretching	869
				CH ₂ & CH wagging	1331	CH ₂ & CH ₃ twisting & rocking	746
n-Butane [41]				i-Butane [44]			
Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)	Description	$\bar{\nu}$ (cm ⁻¹)
CH ₃ asymmetric stretch	2968	CH ₃ asymmetric stretch	2968	CH ₃ asymmetric stretch	2968		
CH ₃ asymmetric stretch	2965	CH ₃ symmetric stretch	2965	CH ₃ symmetric stretch	2956		
CH ₃ asymmetric stretch	2912	CH ₃ symmetric stretch	2912	CH ₃ symmetric stretch	2894		
CH ₃ symmetric stretch	2872	CH asymmetric stretch	2872	CH asymmetric stretch	2872		
CH ₂ symmetric stretch	2853	CH ₃ asymmetric stretch	2853	CH ₃ asymmetric stretch	2748		
CH ₃ asymmetric deformation	1460	CH ₃ symmetric stretch	1460	CH ₃ symmetric stretch	2629		
CH ₃ scissoring	1442	CH ₃ asymmetric stretch	1442	CH ₃ asymmetric stretch	1477		
CH ₃ twisting	1300	CH ₃ asymmetric bend	1300	CH ₃ asymmetric bend	1379		
CH ₃ rocking	1151	CH asymmetric bend	1151	CH asymmetric bend	1334		
CC stretching	1059	CCH ₃ bend	1059	CCH ₃ bend	1177		
CC stretching	837	CC stretch	837	CC stretch	925		
		CC bend		CC bend	797		

2.4.2. Simplified Workflow for Model Development

In order to depict how to get a reasonable PLS model, it must be stressed from the very beginning that, although we tried to separate the workflow in individual steps for the sake of simplicity, developing a predictive regression model is an iterative and ongoing process where decisions are taken and models reformulated until a satisfactory one is obtained. This is nicely explained in public documents related to the pharmaceutical arena, where these methods are quite common today. Some of them can be of interest as introductory texts [61–63]. The basic steps when developing a predictive PLS model are schematized below:

- Stage 1: Preliminary assays: it is always important to visualize the spectra in order to evaluate gross differences among them, unusual signals, potential range of variables to be considered, presence of outlying samples, etc. A preliminary model can also be done to feel what its results look like and whether the samples spread through the working range of interest.
- Stage 2: spectra usually need what is called data pre-processing. This step attempts to get rid of useless information or undesired characteristics that may be detrimental for the predictions. For instance, baseline correction or noise filtering are typical steps before developing models. Very common pre-processings are the first derivative [16,64] (sometimes the 2nd derivative is also used) and mean centering [2]. Combinations of pre-processings are also common. Several options used for LNG modeling are shown in Table 3. This stage of model development is in general recommended, although some applications argued that no pre-processing was needed [4]. There is not a definitive answer to this issue, and the unique ‘true’ advice is to try different pre-processings and see which one improves the models.

- Stage 3: A critical point to develop a satisfactory PLS model (also critical for many other techniques that use abstract factors) is to ascertain the number of factors that must be included in the model (this is called model dimensionality). Although including many factors to take account of the majority of the variance (consider variance and information as synonymous for this purpose) contained into the spectra might seem to be a good idea, it is worth considering that those factors might be unrelated to the property of interest. They might be related to baseline effects or some spurious behavior of a peak, contributing to bad predictions. Of course, we also assume that too few factors will not be sufficient to extract enough relevant information to get a sound model. So, we need to equilibrate overfitting (too many factors) with underfitting (too few factors). One of the best options for this, though not perfect, consists of performing cross-validation. This empirically evaluates a cost function so that a minimum in the error is searched for [65]. Cross-validation is iterative and sometimes takes some computer time. Its conceptual idea can be resumed in a pseudo-code as follows:
 1. Fix a number of factors, let us say 1.
 2. Extract, momentarily, a reduced set of samples (spectra) from the data set assigned to ‘calibration’. Several approaches can be used for this purpose, such as ‘leave-one-out’, ‘venetian blinds’, and ‘random selection of the subsets’ [66,67].
 3. Develop a model, and test how well it predicts those samples left out of it. The error can be stored in the computer memory.
 4. Reintegrate those spectra to the calibration set and extract another small subset of samples.
 5. Develop a model, test it with the second set of left-out spectra, and sum the error to the previous one.
 6. Continue the process until all samples (or possible subsets of samples) are excluded from the calibration stage and predicted afterwards. The summed errors yield the overall prediction error (=RMSECV).
 7. Return to 1 and increase the number of factors by one, and repeat the process again.
 8. At the end of a number of factors (for example, 20), a plot of the prediction errors will show which number of factors leads to the lowest overall prediction error (Figure 10). More details can be consulted elsewhere [65].

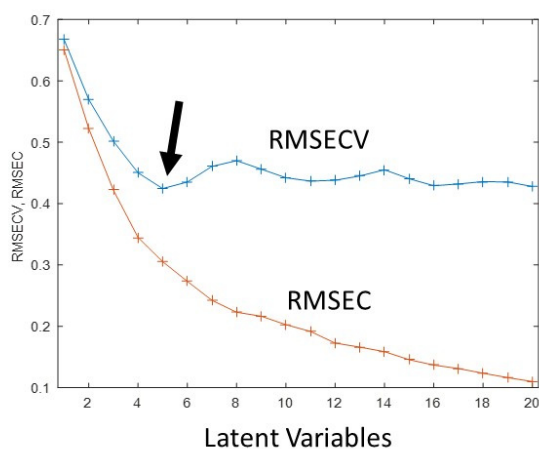


Figure 10. Examples of RMSEC-RMSECV plots for decision-making. The best number of factors (Latent Variables) is indicated with an arrow. Note that the RMSEC (root mean square error in calibration) diminishes monotonically; so it is useless to select the dimensionality of the models. RMSECV stands for root mean square error in cross-validation.

Table 3. Pre-processing options used by several authors before LNG multivariate modeling.

Author	Pre-Processing
Haghi et al. [16]	1st derivative (Savitzky–Golay algorithm), smoothing over five points plus orthogonal signal correction (OSC).
Ferreiro et al. [2]	Iterative baseline correction (automatic weighted least squares, using a polynomial of order 2), spectral normalization (total area = 1), and mean centring.
Barbosa et al. [4]	2nd derivative with Savitzky-Golay smoothing using 21-point windows and a 2nd order polynomial.
Rohwedder et al. [64]	1st derivative (Savitzky–Golay algorithm) employing a 7-points window, 2nd order polynomial for baseline correction and smoothing.
Nurida et al. [52]	Baseline correction and mean center (for CO ₂ absorption).

The procedure thus schematized is also termed ‘internal validation’, but it can still yield models that overfit the calibration set. The best way to avoid this problem is to use an external test set, if it is possible, or an external validation set (mentioned at the very beginning of this section). If the errors obtained for the internal validation (RMSECV) and the external validation (RMSEP) are similar, we can assume that the model does not overfit. That would be our best number of factors for the model (or latent variables, as it is a more correct denomination for PLS). Unfortunately, in many occasions, a clear minimum is not seen, and an inflection point where from the error do not decrease significantly has to be selected.

A nice and conceptually simple alternative to avoid overfitting without resorting to cross-validation or excluding samples consists of performing randomization. In essence, it breaks the ‘natural’ relation between spectra and their corresponding reference values (obtained after analyzing the samples with the reference methodologies), so that they are now related randomly. Then, PLS regression models that should reflect the absence of a real association between the X- and Y-variables are calculated. The randomization step is repeated a large number of times to get a sound significance, and a test statistic is calculated from all of them. Then, a critical value is derived from the distribution of those statistics. Finally, the statistic obtained for the model with the (not randomized) original data for a given number of factors is calculated. Both values are compared. A significant factor should lead to a statistic greater than the critical one. This is done successively for each factor being introduced into the model so that it is possible to decide when new factors are useless [68,69].

- Stage 4: A reason why developing a model is an iterative process is because we have to check for the existence of outlying samples. If they are present, all the previous stage is biased, and the model is not reliable. The problem is how to detect them. Likely, in the same way as you detect wrong points in a traditional calibration plot: by calculating statistics to evaluate the behavior of the samples into the model. Two of the most important and useful ones are the ‘Q residuals’ and the ‘Hotelling’s T²’ statistics [70] (another usual statistic is the leverage, although it is closely related to the T² one). The former detects whether a spectrum has some new or different spectral characteristic(s) that could not be modeled with the present model (note: ‘different’ refers to its comparison with the residuals of the other calibration spectra in the model), whereas the second evaluates how close the spectrum is to the average spectrum of the calibration set. Clearly, we would like samples with spectra close to the average and without new spectral characteristics.

From a practical viewpoint inspect graphical outputs such as, typically, the ‘Q residuals versus Hotelling’s T² plot’ and the ‘Y_{residuals} (predicted minus actual values for the property) versus leverage plot’. Briefly, an outlier at the former plot (a point further away from the statistical limits), such as that shown in Figure 11a (red point), will have either

spectral characteristics not considered at the calibration set or much higher/lower amounts of one/several components. An outlier on the latter plot (Figure 11b) will have a high residual value, either positive or negative, which denotes that it contains more unmodeled spectral characteristics than the calibration samples do. Finally, samples deemed to be outliers should be deleted and the model reformulated (and validated) again.

- Stage 5: The external validation set of samples is of use not only to assess that the model yields no overfitting but to evaluate how good the predictions of new samples are. A typical plot, such as that in Figure 12, is of most information as it yields insight on the closeness of the individual predictions to the true values, and samples predicted badly (possible outliers?), in addition to the overall average error.

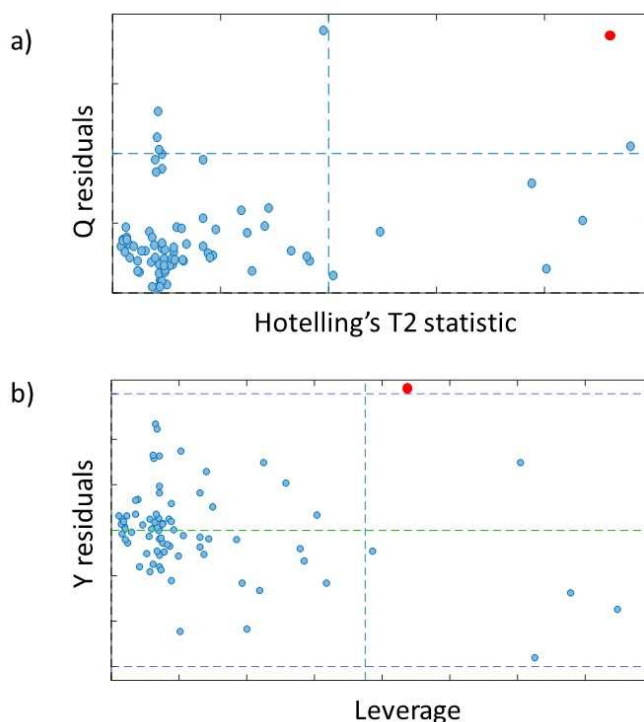


Figure 11. (a) Hotelling's T^2 versus Q residuals plot and (b) $Y_{\text{residuals}}$ versus leverage plot. One of the outlying samples is shown in red.

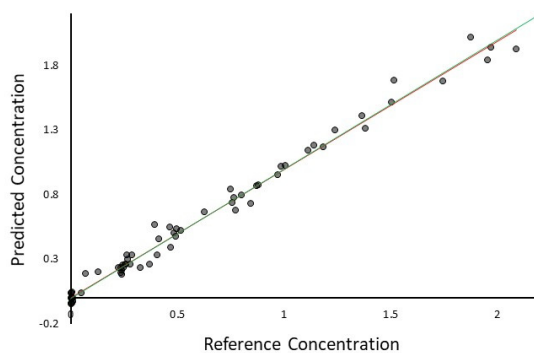


Figure 12. Typical comparison between the PLS-predicted values and the reference or 'true' ones. See text for the different studies that can be derived from it.

If the slope of the regression line calculated using the pairs of data in Figure 12 is statistically equal to one, both methods lead to statistically comparable results, but, for a possible background, which is given by the intercept (although it may not always hold a true significance, for instance, when large extrapolations to the zero are required). If the intercept is statistically different from zero, it indicates that there is a general offset in the predictions towards higher or lower values (when compared to the reference ones).

Furthermore, the ‘actual versus predicted’ regression line (Figure 12) is of use to calculate the procedural bias and the average precision of the multivariate model. Bias can be, simply, $\frac{\sum(x_{\text{pred}} - x_{\text{true}})}{m}$, (m = number of samples). Repeatability precision can be evaluated quite easily as the standard error of the regression line ($s_{y/x}$), just assuming that the error in the reference values is negligible in comparison to the PLS predictions [65,71]. That is true when using standardized chromatographic methodologies but not so evident when complex methods, such as the motor methane number, are used. Reproducibility precision can be roughly the average prediction error, calculated as $\text{RMSEP} = \left(\sqrt{\frac{\sum(x_{\text{pred}} - x_{\text{true}})^2}{m}} \right)$, m = number of validation samples, and ‘pred’ stands for predicted. This term can be bias-corrected considering $\text{RMSEP}_{\text{corrected}} = \text{RMSEP} - \text{bias}$ (if bias is zero, the usual statement that RMSEP is the reproducibility precision holds fine).

It is also worth noting that Figure 12 can be used to calculate figures or merit related to the modern limits of detection (where the risks of false positives and false negatives are controlled [72–76]) and to derive a sample-specific confidence interval associated to each prediction. Calculations for the former topic are not trivial, and readers are kindly forwarded to the references given above. As a snapshot, the classical, old (1970s decade) limit of detection is today termed critical limit, and it only takes account for the risk of false positives. It can be estimated as $x_c = (t_{(\alpha, N-2)} \cdot w_0 \cdot s_{y/x}) / b$, with t = one-tail Student’s t statistic, and $w_0 \cdot s_{y/x}$ = standard error at zero concentration derived from the ‘predicted versus actual’ regression; b is the slope [72,74]. The ‘limit of detection’ (indeed, this term has been superseded, and it should be now called capability of detection or minimum detectable net concentration) has a very similar equation, although using a non-central t distribution instead of the current Student’s one in order to consider the risks of false positives and false negatives simultaneously (both at 95% confidence level).

With regard to the second topic, in simple terms, the sample-specific confidence interval can be formulated as [predicted value $\pm t \cdot S_{x_0}$], where t is the Student’s t value for ν degrees of freedom at a 95% confidence level. S_{x_0} is the specific standard error of the prediction [77,78], which can be estimated as $\left[(1 + h_i) \cdot \text{RMSEC}^2 - S_{\text{ref}}^2 \cdot \left(\frac{\nu}{X^2} \right) \right]^{1/2}$, where h_i is the leverage for sample i , RMSEC is the standard error of the calibration, S_{ref} is the measurement error in the reference value, and X^2 is the X^2 statistic (ν degrees of freedom and 95% confidence level) used to avoid overcorrection of the bias term [79]. The fact that the equation includes the leverage of the sample (which depends on its position in the factorial space of the model) particularizes the equation for each sample. Hence, the denomination sample-specific as it calculates a different confidence interval for each sample. Note also that it takes account of the error in the reference values.

3. Conclusions

The new analytical methodology proposed in this work for routine LNG quality control is composed of two major steps: infrared spectral measurement of a gaseous sample and chemometric model development. It was reviewed that the most important parameters that affect the measurements are the instrument (spectral range, operational characteristics, such as resolution, number of scans, apodization, and background collection) and the gas cell (pathlength, internal pressure, broadening gas, windows). The tiered development of a chemometric model was reviewed, and it was found to be composed of several intertwined steps: pre-processing, outlier detection, regression method, cross-validation, and validation.

The combination of the spectrometric and chemometric methods has some interesting advantages for routine quality control: high speed, reduced costs, reduced use of reagents, and high flexibility (as the same spectrum can be used to predict several LNG properties). Hence, the hybridization of FTIR and multivariate regression (in particular, PLS) yields a reliable, powerful alternative to the traditional, standardized Gas Chromatography approach for routine measurements.

Author Contributions: Conceptualization, B.F. and J.A.; Data curation, P.L.-M. and S.M.-L.; Formal analysis, B.F., J.A. and C.P.-Q.; Investigation, B.F., J.A. and C.P.-Q.; Methodology, B.F., J.A., P.L.-M. and S.M.-L.; Resources, P.L.-M. and S.M.-L.; Validation, J.A. and C.P.-Q.; Writing—original draft, B.F.; Writing—review & editing, B.F., J.A., P.L.-M. and S.M.-L. All authors have read and agreed to the published version of the manuscript.

Funding: Part of this work was performed under the EMPIR 16ENG09 project ‘Metrological support for LNG and LBG as transport fuel (LNG III)’. This project has received funding from the EMPIR programme co-financed by the Participant States and from the European Union’s Horizon 2020 Research and Innovation programme. Mestrelab, Reganosa and Naturgy are acknowledged for hiring the services of the Group of Applied Analytical Chemistry for FTIR method development. The Program “Consolidación e Estructuración de Unidades de Investigación Competitivas” of the Galician Government (Xunta de Galicia) is also acknowledged (Grant ED431C 2021/56).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: NIST infrared spectra can be obtained from <https://webbook.nist.gov/chemistry/name-ser/> (accessed on 25 November 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. ISO 6974-1: 2013 Natural Gas—Determination of Composition and Associated Uncertainty by Gas Chromatography—Part 1: General Guidelines and Calculation of Composition; ISO: Geneva, Switzerland, 2013.
2. Ferreiro, B.; Andrade, J.; López-Mahía, P.; Muniategui, S.; Vázquez, C.; Pérez, A.; Rey, M.; Vales, C. Fast quality control of natural gas for commercial supply and transport utilities. *Fuel* **2021**, *305*, 121500. [CrossRef]
3. Kiefer, J. Recent advances in the characterization of gaseous and liquid fuels by vibrational spectroscopy. *Energies* **2015**, *8*, 3165–3197. [CrossRef]
4. Barbosa, M.F.; Santos, J.R.B.; Silva, A.N.; Soares, S.F.C.; Araujo, M.C.U. A cheap handheld NIR spectrometric system for automatic determination of methane, ethane, and propane in natural gas and biogas. *Microchem. J.* **2021**, *170*, 106752. [CrossRef]
5. Koturbash, T.; Karpash, M.; Darvai, I.; Rybitskyi, I.; Kutcherov, V. Development of new instant technology of natural gas quality determination. In Proceedings of the ASME 2013 Power Conference, Boston, MA, USA, 29 July–1 August 2013; Volume 1. [CrossRef]
6. Looney, B. Statistical Review of World Energy 2021. *Rev. World Energy Data* **2021**, *70*, 8–20.
7. Hosseini, M.; Dincer, I.; Ozbilen, A. *Expert Opinions on Natural Gas Vehicles Research Needs for Energy Policy Development*; Elsevier: Amsterdam, The Netherlands, 2018; ISBN 9780128137352.
8. BP, P.L.C. Natural Gas Demand. Available online: <https://www.bp.com/en/global/corporate/energy-economics/energy-outlook/demand-by-fuel/natural-gas.html> (accessed on 20 October 2021).
9. Rahmouni, C.; Tazerout, M.; Le-Corre, O. Determination of the combustion properties of natural gases by pseudo-constituents. *Fuel* **2003**, *82*, 1399–1409. [CrossRef]
10. Roy, P.S.; Ryu, C.; Park, C.S. Predicting Wobbe Index and methane number of a renewable natural gas by the measurement of simple physical properties. *Fuel* **2018**, *224*, 121–127. [CrossRef]
11. Karpash, O.; Darvai, I.; Karpash, M. New approach to natural gas quality determination. *J. Pet. Sci. Eng.* **2010**, *71*, 133–137. [CrossRef]
12. Sweelssen, J.; Blokland, H.; Rajamäki, T.; Boersma, A. Capacitive and Infrared Gas Sensors for the Assessment of the Methane Number of LNG Fuels. *Sensors* **2020**, *20*, 3345. [CrossRef]
13. Sweelssen, J.; Blokland, H.; Rajamäki, T.; Sarjonen, R.; Boersma, A. A versatile capacitive sensing platform for the assessment of the composition in gas mixtures. *Micromachines* **2020**, *11*, 116. [CrossRef]
14. Boersma, A.; Sweelssen, J.; Blokland, H. Gas Composition Sensor for Natural Gas and Biogas. *Procedia Eng.* **2016**, *168*, 197–200. [CrossRef]
15. Makhoukhi, N.; Péré, E.; Creff, R.; Pouchan, C. Determination of the composition of a mixture of gases by infrared analysis and chemometric methods. *J. Mol. Struct.* **2005**, *744–747*, 855–859. [CrossRef]

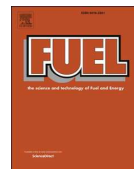
16. Haghi, R.K.; Yang, J.; Tohidi, B. Fourier Transform Near-Infrared (FTNIR) Spectroscopy and Partial Least-Squares (PLS) Algorithm for Monitoring Compositional Changes in Hydrocarbon Gases under In Situ Pressure. *Energy Fuels* **2017**, *31*, 10245–10259. [CrossRef]
17. Ribessi, R.L.; Neves, T.D.A.; Rohwedder, J.J.R.; Pasquini, C.; Raimundo, I.M.; Wilk, A.; Kokoric, V.; Mizaikoff, B. IHEART: A miniaturized near-infrared in-line gas sensor using heart-shaped substrate-integrated hollow waveguides. *Analyst* **2016**, *141*, 5298–5303. [CrossRef]
18. Dąbrowski, K.M.; Kuczyński, S.; Barbacki, J.; Włodek, T.; Smulski, R.; Nagy, S. Downhole measurements and determination of natural gas composition using Raman spectroscopy. *J. Nat. Gas Sci. Eng.* **2019**, *65*, 25–31. [CrossRef]
19. Eichmann, S.C.; Kiefer, J.; Benz, J.; Kempf, T.; Leipertz, A.; Seeger, T. Determination of gas composition in a biogas plant using a Raman-based sensorsystem. *Meas. Sci. Technol.* **2014**, *25*, 075503. [CrossRef]
20. Sieburg, A.; Knebl, A.; Jacob, J.M.; Frosch, T. Characterization of fuel gases with fiber-enhanced Raman spectroscopy. *Anal. Bioanal. Chem.* **2019**, *411*, 7399–7408. [CrossRef]
21. Specac Product Catalogue. Available online: <https://www.specac.com/en/documents/catalogues> (accessed on 25 November 2021).
22. Microptik Gas Sampling. Available online: <https://www.microptik.eu/product/gas-sampling> (accessed on 25 November 2021).
23. Jasco Gas Cells. Available online: <https://jascoinc.com/products/spectroscopy/ftir-spectrometers/ftir-accessories/gas-cells/> (accessed on 25 November 2021).
24. Guiding Photonics Gas Cells. Available online: <https://guidingphotonics.com/gas-cells/> (accessed on 21 September 2021).
25. Kriesel, J.M.; Gat, N.; Bernacki, B.E.; Erikson, R.L.; Cannon, B.D.; Myers, T.L.; Bledt, C.M.; Harrington, J.A. Hollow core fiber optics for mid-wave and long-wave infrared spectroscopy. In Proceedings of the International Society for Optics and Photonics, Orlando, FL, USA, 26–28 April 2011; Volume 8018, p. 80180V.
26. Li, N.; Tao, L.; Yi, H.; Kim, C.S.; Kim, M.; Canedy, C.L.; Merritt, C.D.; Bewley, W.W.; Vurgafman, I.; Meyer, J.R. Methane detection using an interband-cascade LED coupled to a hollow-core fiber. *Opt. Express* **2021**, *29*, 7221–7231. [CrossRef]
27. Pike Technologies. Choice of Window Materials for Transmission Sampling of Liquids in the Mid-IR Spectral Region. Available online: https://www.piketech.com/skin/fashion_mosaic_blue/application-pdfs/CrystalChoiceForTransmission.pdf (accessed on 25 November 2021).
28. Shimadzu Corporation Safety of Windows and Prisms Used in FTIR. Available online: <https://www.shimadzu.com/an/service-support/technical-support/analysis-basics/tips-ftir/safety.html> (accessed on 20 September 2021).
29. Spectra-Tech. How to Select an Infrared Transmission Window 2012. Available online: https://kinecat.pl/wp-content/uploads/2012/11/crystal_ref.pdf (accessed on 25 November 2021).
30. Smith, B.C. *Fundamentals of Fourier Transform Infrared Spectroscopy*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2011; ISBN 1420069306.
31. Griffiths, P.R.; De Haseth, J.A. *Fourier Transform Infrared Spectrometry*; John Wiley & Sons: Chichester, UK, 2007; Volume 171, ISBN 0470106298.
32. Harris, D.C. *Análisis Químico Cuantitativo*; Reverté: Barcelona, Spain, 2007; ISBN 8429172246.
33. Skoog, D.A.; Holler, F.J.; Nieman, T.A. *Principios de Análisis Instrumental*, 7th ed.; Cengage Learning: Boston, MA, USA, 2018; ISBN 978-607-481-390-6.
34. Stuart, B. Infrared Spectroscopy. In *Kirk-Othmer Encyclopedia of Chemical Technology*, 1st ed.; John Wiley & Sons: Chichester, UK, 2015; pp. 73–101. ISBN 0471238961.
35. Anderson, P.W. Pressure Broadening in the Microwave and Infra-Red Regions. *Phys. Rev.* **1949**, *76*, 647–661. [CrossRef]
36. Jabłoński, A. General Theory of Pressure Broadening of Spectral Lines. *Phys. Rev.* **1945**, *68*, 78–93. [CrossRef]
37. Margenau, H.; Watson, W.W. Pressure effects of foreign gases on the sodium D-lines. *Phys. Rev.* **1933**, *44*, 92–98. [CrossRef]
38. Ferreiro, B.; Andrade, J.M.; Paz-Quintáns, C.; López-Mahía, P.; Muniategui-Lorenzo, S.; Rey-Garrote, M.; Vázquez-Padín, C.; Vales, C. Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas. *Energy Fuels* **2019**, *33*, 6929–6933. [CrossRef]
39. NIST Public. IR Spectra Database. Available online: <https://webbook.nist.gov/chemistry/name-ser/> (accessed on 25 November 2021).
40. Es-Sebar, E.; Farooq, A. Intensities, broadening and narrowing parameters in the ν_3 band of methane. *J. Quant. Spectrosc. Radiat. Transf.* **2014**, *149*, 241–252. [CrossRef]
41. Shimanouchi, T. Tables of Molecular Vibrational Frequencies: Part 6. *J. Phys. Chem. Ref. Data* **1973**, *2*, 121–162. [CrossRef]
42. Gough, K.M.; Murphy, W.F.; Raghavachari, K. The harmonic force field of propane. *J. Chem. Phys.* **1987**, *87*, 3332–3340. [CrossRef]
43. Hudson, R.L.; Gerakines, P.A.; Yarnall, Y.Y.; Coones, R.T. Infrared spectra and optical constants of astronomical ices: III. Propane, propylene, and propyne. *Icarus* **2021**, *354*, 114033. [CrossRef]
44. Evans, J.C.; Bernstein, H.J. The vibrational spectra of isobutane and isobutane-d1. *Can. J. Chem.* **1956**, *34*, 1037–1045. [CrossRef]
45. Brereton, R.G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley & Sons: Chichester, UK, 2003; ISBN 0470845740.
46. Otto, M. *Chemometrics: Statistics and Computer Application in Analytical Chemistry*; John Wiley & Sons: Chichester, UK, 2016; ISBN 3527340971.
47. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
48. Martens, H. Reliable and relevant modelling of real world data: A personal account of the development of PLS Regression. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 85–95. [CrossRef]

49. Gupta, S.K.; Mittal, M. Predicting the methane number of gaseous fuels using an artificial neural network. *Biofuels* **2019**, *12*, 1191–1198. [CrossRef]
50. Bai, P.; Duan, X.; He, C.; Li, Y. Natural gas infrared spectrum analysis based on multi-level and SVM-subset. In Proceedings of the 2009 IEEE International Conference on Virtual Environments, Human-Computer Interfaces and Measurements Systems, Hong Kong, 11–13 May 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 336–339.
51. Udina, S.; Carmona, M.; Pardo, A.; Calaza, C.; Santander, J.; Fonseca, L.; Marco, S. A micromachined thermoelectric sensor for natural gas analysis: Multivariate calibration results. *Sens. Actuators B Chem.* **2012**, *166–167*, 338–348. [CrossRef]
52. Nurida, M.Y.; Norfadilah, D.; Aishah, M.R.S.; Phak, C.Z.; Saleh, S.M. Monitoring of CO₂ Absorption Solvent in Natural Gas Process Using Fourier Transform Near-Infrared Spectrometry. *Int. J. Anal. Chem.* **2020**, *2020*, 1–9. [CrossRef]
53. Ponte, S.; Andrade, J.M.; Vázquez, C.; Ferreira, B.; Cobas, C.; Pérez, A.; Rey, M.; Vales, C.; Pellitero, J.; Santacruz, B.; et al. Prediction of the methane number of commercial liquefied natural gas samples using mid-IR gas spectrometry and PLS regression. *J. Nat. Gas Sci. Eng.* **2021**, *90*, 103944. [CrossRef]
54. Wise, B.M.; Gallagher, N.B.; Bro, R.; Shaver, J.; Windig, W.; Koch, J. *PLS_Toolbox*; Eigenvector Research Inc.: Manson, WA, USA, 2006.
55. ThermoFisher Scientific. Available online: <https://www.thermofisher.com/order/catalog/product/INF-15004> (accessed on 25 November 2021).
56. Aspen Technology Inc. Unscrambler. Available online: <https://www.aspentech.com/en/products/msc/aspens-unscrambler> (accessed on 24 November 2021).
57. MultiD Analyses, B.D. Genex. Available online: <https://multid.se/genex/> (accessed on 24 November 2021).
58. IBM Co. IBM SPSS Software. Available online: <https://www.ibm.com/es-es/analytics/spss-statistics-software> (accessed on 24 November 2021).
59. Statgraphics Technologies, Inc. Available online: <https://www.statgraphics.com/> (accessed on 24 November 2021).
60. Leardi, R.; Melzi, C.; Polotti, G. CAT (Chemometric Agile Tool), Freely. Available online: <http://gruppochemiometria.it/index.php/software> (accessed on 4 November 2021).
61. Broad, N.; Graham, P.; Hailey, P.; Hardy, A.; Holland, S.; Hughes, S.; Lee, D.; Prebble, K.; Salton, N.; Warren, P. Guidelines for the Development and Validation of Near-Infrared Spectroscopic Methods in the Pharmaceutical Industry. In *Handbook of Vibrational Spectroscopy*; John Wiley & Sons: Chichester, UK, 2002; Volume 5, pp. 3590–3610.
62. Validation of Analytical Procedures: Text and Methodology. In *ICH Harmonised Tripartite Guideline*; Somatek Inc.: San Diego, CA, USA, 2014.
63. European Medicines Agency Guideline on the Use of Near Infrared Spectroscopy (NIRS) by the Pharmaceutical Industry and the Data Requirements for New Submissions and Variations. 2012. Available online: <https://www.ema.europa.eu/en/use-near-infrared-spectroscopy-nirs-pharmaceutical-industry-data-requirements-new-submissions> (accessed on 25 November 2021).
64. Rohwedder, J.J.R.; Pasquini, C.; Fortes, P.R.; Raimundo, I.M.; Wilk, A.; Mizaikoff, B. iHWG- μ NIR: A miniaturised near-infrared gas sensor based on substrate-integrated hollow waveguides coupled to a micro-NIR-spectrophotometer. *Analyst* **2014**, *139*, 3572. [CrossRef]
65. Andrade-Garda, J.M.; Carlosena-Zubieta, A.; Boqué-Martí, R.; Ferré-Baldrich, J. Partial Least Squares Regression. In *Basic Chemometric Techniques in Atomic Spectroscopy*; Andrade-Garda, J.M., Ed.; Royal Society of Chemistry: Cambridge, UK, 2013; pp. 280–347. ISBN 1849737967.
66. Malinowski, E.R.; Howery, D.G. *Factor Analysis in Chemistry*; Wiley: Hoboken, NJ, USA, 1980; ISBN 0471058815.
67. Eigenvector Research Eigenvector Wiki. Available online: <https://www.wiki.eigenvector.com/index.php?title=Confusionmatrix> (accessed on 14 May 2021).
68. Faber, N.M.; Rajkó, R. How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Anal. Chim. Acta* **2007**, *595*, 98–106. [CrossRef]
69. Wiklund, S.; Nilsson, D.; Eriksson, L.; Sjöström, M.; Wold, S.; Faber, K. A randomization test for PLS component selection. *J. Chemom.* **2007**, *21*, 427–439. [CrossRef]
70. Vinzi, V.E.; Chin, W.W.; Henseler, J.; Wang, H. *Handbook of Partial Least Squares*; Springer: Berlin/Heidelberg, Germany, 2010; ISBN 978-3-540-32825-4.
71. Sanz, M.B.; Sarabia, L.A.; Herrero, A.; Ortiz, M.C. Multivariate analytical sensitivity in the determination of selenium, copper, lead and cadmium by stripping voltammetry when using soft calibration. *Anal. Chim. Acta* **2003**, *489*, 85–94. [CrossRef]
72. Ortiz, M.C.; Sarabia, L.A.; Sánchez, M.S. Tutorial on evaluation of type I and type II errors in chemical analyses: From the analytical detection to authentication of products and process control. *Anal. Chim. Acta* **2010**, *674*, 123–142. [CrossRef]
73. Ortiz, M.C.; Sarabia, L.A.; Herrero, A.; Sánchez, M.S.; Sanz, M.B.; Rueda, M.E.; Giménez, D.; Meléndez, M.E. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemom. Intell. Lab. Syst.* **2003**, *69*, 21–33. [CrossRef]
74. ISO 11843. *Capability of Detection—Part 2: Methodology in the Linear Calibration Case*; ISO: Geneva, Switzerland, 2008.
75. SANCO/2004/2726 Guidelines for the Implementation of Decision 2002/657/EC. Rev 4-December 2008. Available online: https://ec.europa.eu/food/system/files/2016-10/cs_vet-med-residues_cons_2004-2726rev4_en.pdf (accessed on 25 November 2021).
76. European Commission. Commission Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Off. J. Eur. Communities* **2002**, *50*, 8–36.

77. Faber, K.; Kowalski, B.R. Prediction error in least squares regression: Further critique on the deviation used in The Unscrambler. *Chemom. Intell. Lab. Syst.* **1996**, *34*, 283–292. [[CrossRef](#)]
78. Faber, N.M.; Schreutelkamp, F.H.; Vedder, H.W. Estimation of prediction uncertainty for a multivariate calibration model. *Spectrosc. Eur.* **2004**, *16*, 17–21.
79. Faber, K.; Kowalski, B.R. Improved prediction error estimates for multivariate calibration by correcting for the measurement error in the reference values. *Appl. Spectrosc.* **1997**, *51*, 660–665. [[CrossRef](#)]

Chapter 4: Fast Quality Control of Natural Gas for Commercial Supply and Transport Utilities.





Full Length Article

Fast quality control of natural gas for commercial supply and transport utilities

Borja Ferreiro^a, José Andrade^{a,*}, Purificación López-Mahía^a, Soledad Muniategui^a,
Cristina Vázquez^b, Andrés Pérez^b, María Rey^b, Carlos Vales^b

^a Group of Applied Analytical Chemistry. University of A Coruña. Campus da Zapateira, s/n, E-15071, A Coruña, Spain

^b Regasificadora del Noroeste, S.A. Punta Promontorio, Mugarbos (Ferro), Spain



ARTICLE INFO

Keywords:

Natural gas composition
Wobbe index
Infrared gas measurement
PLS prediction
Spectral variable Selection

ABSTRACT

Quality control of natural gas frequently relies on off-line slow standardized chromatographic techniques. Previous implementations of new measurement approaches focused of synthetic mixtures without extensive industrial validation. Here, a fast alternative based on infrared spectra is presented to predict the gas constituents and a physical parameter, the Wobbe index. Commercial samples instead of synthetic mixtures were used to develop predictive models. Method performance parameters were calculated and ca. 100 % of the sample-specific confidence intervals for the predictions overlapped with those of the reference values and the approach was unbiased and precise. The limits of detection and quantification (classical and considering errors of type I and II) outperformed other approaches. Validation included commercial samples and primary mixtures. Furthermore, prediction models considering reduced sets of variables were sought for using Markov-chain Monte Carlo guided searches (uninformative variable elimination and random frog) and common (iPLS, UVE and SR) approaches. The prediction errors and limits of detection of these 'reduced' models outperformed those from other approaches. The methodology takes only minutes to analyse a sample, requires few sample and no reagents (only some argon), making this approach cost-effective and environmentally-friendly.

1. Introduction

Despite petroleum being a non-renewable resource its distillates are still critical as energetic and raw material sources. Unfortunately, their combustion lead to atmospheric pollution [1], which constitutes one of the strongest causes of global climate change. Many countries adopted policies to reduce the carbon footprint and pollution by empowering the use of greener energies. Obvious renewable energy sources (wind, sun, etc.) should be complemented with natural gas (NG) to change smoothly some energetic paradigms as its combustion leads to reduced emissions of CO₂, NO_x and SO_x [2]. Besides, massive deposits have been discovered, many of which are still under-exploited. Nowadays NG is employed mainly to generate electric power and to industrial and domestic applications. Its use in transportation is not as prevalent, mainly due to the low energy density (for liquefied NG, ca. half that of gasoline) which would require bigger fuel tanks, so that hybrid and electric engines are preferred in most cases [3]. However, the EU promoted it as a suitable fuel for heavy-duty vehicles and maritime transport [4]. The EU

objective is to change traditional diesel engines for liquefied natural gas (LNG) propellers, and it was planned to deploy the corresponding facilities on maritime ports and on land along the Trans-European Transport Networks by the end of 2025. The trend of changing traditional personal vehicles to alternative fuel vehicles (considering compressed natural gas and gas-hybrid engines) is also rising. The EU directive [5] is gaining momentum rapidly as most European capitals strongly limit the circulation of traditional vehicles through their city centers.

The growing use of NG requires reliable quality control methodologies to determine its composition and energetic properties. They depend on the geological deposit where the NG comes from. NG is composed mainly of methane (75–99 %), mixed with other light hydrocarbons, like ethane (0–20 %), propane, butanes and pentanes, plus nitrogen and/or carbon dioxide (all of them < 10 %). Sometimes, hexane isomers could be present at trace levels. Hence, the composition of the NG must be known to evaluate its physical properties and combustion power. In particular:

* Corresponding author.

E-mail address: andrade@udc.es (J. Andrade).

<https://doi.org/10.1016/j.fuel.2021.121500>

Received 15 April 2021; Received in revised form 30 June 2021; Accepted 17 July 2021

Available online 28 July 2021

0016-2361/© 2021 The Author(s).

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

- i) The methane number, MN, is an analogue to the octane number in gasoline or the cetane number in diesel. Its determination is controversial [6–8] and there exist different algorithms to estimate its value once the NG composition is known. Recently an IR-based alternative without that need has been proposed [9].
- ii) The Wobbe index, WI, evaluates the combustion energy of the NG using standardized formulae and its composition and represents the volume-basis calorific value, at specified reference conditions, divided by the square root of the relative density at the same specified metering reference conditions. The WI is said to be gross or net (sometimes, lower or higher) according to whether the calorific value used is the gross or net calorific value (in common usage, the WI as employed here refers to the gross index) [10,11]. The WI is the primary gas interchangeability criterion for residential and commercial appliances as well as for some large-scale combustion equipment in industry and power generation.

Currently, the NG composition is measured by standardized gas chromatography (GC) procedures [12]. However, GC is difficult to adapt to online devices and small facilities [13], and is costly (overall, ca. 600 € per sample) [14]. Modern on-line devices are relatively rapid (ca. 4 min/sample) but they still need frequent time-consuming calibrations and verifications. GC off-line systems need ca. 150 min to get a full, validated result [14], which may delay pipeline distribution and energy custody transfer activities (shipments, storage, etc.). Thus, faster and cheaper methods are required to simplify and accelerate NG routine quality control.

Published alternatives rely mostly on IR (infrared) measurements (although molecular fluorescence was also proposed [15]), with notable reductions in costs (e.g., less than 200 € per sample, in total [14]). Note that although the IR region can reflect accurately the NG composition [16,17] gas spectra are pretty hard to interpret due to the strong overlap of the peaks and so multivariate chemometric treatments are required to get predictions, typically by partial least squares regression (PLS) [13,18].

Very scarce papers applied multivariate regression to deal with the prediction of NG properties. To the best of the authors' knowledge, the most relevant examples are reviewed herein. The two typical IR regions, near and medium (NIR and MIR) were used, although the former was more common.

Likely, one of the very first studies considered methane-ethane-propane ternary mixtures and NIR and MIR measurements combined with PLS and PCR (principal components regression) [18]. Only two pipeline NG samples were considered to test the models, with good agreements for methane and ethane and worst results for propane. The MIR models doubled the errors of the NIR ones and the latter were selected. On the contrary, another report proposed MIR to predict the concentrations of 12 light hydrocarbons (C₁-C₄) when studying the catalytic degradation of butane [19], with most errors around ± 10 % (relative standard deviation). Noteworthy, ethane and propane could not be determined individually but their sum was predicted accurately.

A fast screening NIR-based method for methane in NG was implemented considering classification by SIMCA (soft independent modeling of class analogy) instead of a regression because the interest was on ascertaining whether the samples had a minimum methane content [16]. Models were developed using synthetic gas mixtures and validated with a collection of 55 commercial NG samples. Similar to this approach, a very fast microNIR system using hollow waveguides was proposed to determine methane, ethane, propane and total butane in synthetic mixtures using PLS regression [20]. That study was amplified next considering a microNIR advanced system and an acoustic-optical tunable filter device [21]; although only synthetic mixtures were considered, not truly NG samples. In another report, the NIR spectra of 31 synthetic mixtures were used to develop PLS models that were validated using synthetic mixtures and one additional certified gas mixture [13]. The authors studied the impact of temperature and pressure

fluctuations in the models and derived some performance parameters (figures-of-merit). In these studies no commercial NG samples were analyzed.

The flue gas of a NG-fired generator was studied to determine methane, CO and CO₂ using NIR spectra, non-linear PLS regression and real samples of flue gas (reference values determined by GC) [22]. Finally, a US patent [23] was issued for a NIR-based system (coupled to a PCA-PLS computing module) to be deployed in gas fields and/or transmission infrastructures to monitor the gas composition and WI of several wells.

Very recently two different sensors comprising either six electrochemical detectors and a tunable mid-IR photometer were proposed to determine the composition of NG-like standard mixtures [7], although they are not still on the market.

With regards to the WI, to the best of our knowledge, only a handful of publications evaluated it without resorting to the composition of the gas [24] or other physical parameters [25–27]. A relevant, seminal approach predicted the energy content of NG by combining NIR and PLS [28]. There, synthetic mixtures of the NG components (including N₂ and CO₂) were used to get a model that was validated using a certified standard gas mixture. In another work, a micromachined thermoelectric sensor was proposed to measure several NG properties [29], including the WI, methane and ethane using PLS. Synthetic mixtures composed of methane, ethane, CO₂ and N₂ were used for calibration, and this involved a degradation on the performance of the Wobbe models when typical levels of propane were considered. Also, fiber-enhanced Raman spectroscopy [30] was employed to determine a partial composition of the sample (C₁-n-C₄, N₂ and CO₂) used subsequently to calculate the WI in real time.

From this review it is concluded that true commercial and/or industrial NG samples have scarcely been considered into the studies. From the pragmatic viewpoint of industrial and quality control laboratories, this is a relevant drawback that should be addressed. In addition, almost no synthetic samples contained N₂ nor heavier compounds than butane (although they are indeed present in NG) and most reports did not differentiate between i- and n-butane.

The major aim of this paper is to develop and validate a methodology to determine the composition and the Wobbe index of NG samples by hybridizing their gas-phase MIR spectra with multivariate regression. Our working hypothesis was that having enough industrial samples to build a sound model, the composition of unknown samples can be predicted with remarkable accuracy and, even, complex physical parameters (the Wobbe index) can be addressed. Further, should the approach be satisfactory enough, industrial laboratories can avoid the preparation (purchase) of a huge number of synthetic gas mixtures to develop calibration models. A second objective is to reduce the time required for the analyses as much as possible and make them robust to uninformative variables. For this, a suite of spectral variable reduction methods have been applied to select the most relevant wavenumbers, although without compromising the predictions.

2. Experimental

2.1. Samples

The routine-operation NG samples used throughout correspond to a one-year-collection of ca. 120 samples in the Reganosa gasification plant (Mugardos, A Coruña, Spain) in 500 cm³ stainless steel cylinders. They were NG, vaporized liquefied NG (LNG), and boil-off-gas (BOG) samples. LNG samples from tanks at the terminal harbour were also taken. In the following only the term NG will be used for all samples for the sake of simplicity. Most samples had been employed in the European EMPIR LNGIII project [9] although it did not include the objectives addressed here.

As an additional validation set, 27 mixtures specially developed by Nippon Gases (formerly Praxair) and Linde Gas Benelux B.V. during the

European EURAMET-EMRP LNG II and EURAMET-EMPIR LNG III research projects were employed. Their compositions were determined by standardized methods [9].

2.2. Apparatus

Gas-phase IR measurements were made with an 8400S Shimadzu FTIR spectrometer. The cell setup consisted of a 10 cm path, stainless steel Harrick gas cell (Harrick Scientific, USA) with 2 mm- thick, 47 mm-diameter ZnSe windows, an input tube with a Swagelock 3 way valve, an exhaust tube and an internal pressure gauge (see Fig. 1). The resolution used was 1 cm^{-1} , with a spectral range from 5500 to 480 cm^{-1} and Happ-Genzel apodization. A background was made before each sample, using 0.5 bar of the broadening gas (Argon, Carburios Metálicos (Barcelona, Spain), 99.9992 % purity). All measurements were done at $25 \pm 1\text{ }^\circ\text{C}$, mixing 0.2 bar of the selected sample (ca. 500 mL were enough to perform the studies) and 1.3 bar of argon, employed as broadening gas to enhance the intensity of the spectral features, more comprehensive details and explanations on this issue and the use of broadening gases were detailed previously [14]. The chemical composition of the samples was determined by the ISO-17025-accredited Reganosa laboratory using a protocol based on ISO 6974-4 [12] (more details can be found elsewhere [14]).

2.3. Software and chemometrics

The spectrometer was controlled by the Shimadzu IR Solutions software, v.1.30. The raw spectra were digitized to 9375 data points per spectrum (from 5000 to 480 cm^{-1} , 1 datum per ca. 0.5 cm^{-1}). The spectral treatments and developments of multivariate models used the PLS Toolbox (Eigenvector Co, WA, USA). Different preprocessings were studied and the selected one consisted of an iterative baseline correction (automatic weighted least squares, using a polynomial of order 2), followed by spectral normalization (total area = 1) and mean centring. Variable selection was made using the PLS_Toolbox and a collection of routines for Monte Carlo natural computation presented recently [31], the latter complemented with in-house Matlab routines.

For the purposes of this manuscript, PLS can be described conceptually as a powerful regression method where both the spectral information (X-space) and each of the properties to be predicted (Y-space) are maximally related by means of a set of abstract factors (termed latent variables or, just, factors). PLS has become a de-facto standard and, hence, it was considered here; technical details can be found elsewhere [32,33]. Roughly, PLS modelling consists of a calibration (training) and a validation stage. The latter verifies that neither underfitting (i.e., a

model with a lack of predictive capability due to a lack of information) nor overfitting (i.e., a model with a lack of predictive capability due to memorization of too particular information associated to the training samples, e.g., some minor spectral characteristics) occur. Noteworthy, the latter problem is pretty much frequent than the former [34]. See [Supplementary Material](#) for practical details. The usual practice by which a minimum in the internal validation procedure ([Supplementary Material](#)) is used to fix the number of latent variables can lead to overfitting [34] and, so, it is highly advisable to compare the average error in cross-validation (RMSECV, root mean square error of cross-validation when calibrating) and the average error in a 'true' external prediction set of new samples $-\text{RMSEP}$ - (root mean square error of prediction). Sometimes, this set of samples is called the fine-tuning set, as it is used to refine the selection of the number of LV. Ideally, RMSECV and RMSEP should be of the same order, which results from a trade-off among fitting and prediction [35]. A final external validation set of samples is needed to accurately check how the model behaves (sometimes it is called testing set). Hence, two 'validation' steps were undergone here, both with external new samples: a first one to fine-tune the model after a number of factors was suggested by internal cross validation, and a final one to validate the model.

2.3.1. Variable reduction methods and performance parameters

A model can be refined by avoiding those spectral variables that do not contribute to the predictions. Variable selection methods, thus, play an important role in potentially improving model robustness and/or allowing for the development of dedicated instruments. In this paper we selected five strategies. The first three are applied frequently and are broadly available; the other two are based on exhaustive 'natural computation algorithms' (population-driven or Bayesian approaches) and constitute quite new developments [31]; however, they were not applied in quality control. A conceptual overview of each of them is presented in the [Supplementary Material](#).

After developing the models several statistical performance parameters [34] must be calculated to assess their adequacy (details are given in the [Supplementary Material](#)): the coefficient of determination (R^2), bias, the standard error of performance (SEP) (if bias is statistically negligible, the value of SEP equals the RMSEP and, so the latter is interpreted as a standard deviation), and the ratio of prediction to deviation (RPD) [34]. In addition, the modern IUPAC, EU and ISO limits of detection and quantification (i.e., including both the risks of type I –false positives- and type II -number of false negatives) [36–39,40,41] were employed.

Here, both the classical and the modern limits of detection were calculated for ethane, propane, n-butane and i-butane because they have

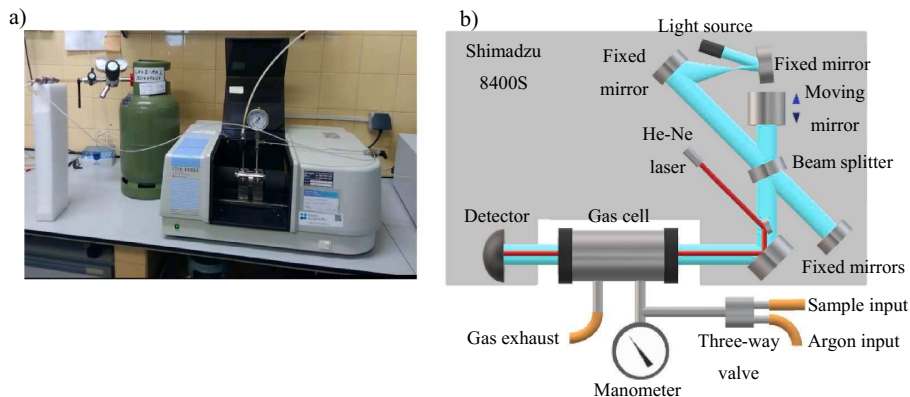


Fig. 1. a) Measurement setup, 8400S Shimadzu FTIR spectrometer and connections to a synthetic NG-like mixture container. b) Experimental setup diagram (the grey background indicates the parts of the Shimadzu 8400S spectrometer).

concentrations close to zero, whereas methane and the WI values remain far apart from the origin and, so, they would require large extrapolations which would render unreasonable figures. Noteworthy, the modern limits have not still been reported for usual NG analysis.

3. Results and discussion

3.1. Predictive model for % of methane

The model to predict the percentage molar volume concentration, vol%, of Methane was developed from a calibration set of 71 samples whose concentrations spread evenly between 89 % and 99 %. During our

sampling period (ca. one year) only 4 samples with percentages between 68 % and 85 % methane were collected and, thus, the models could not predict them reliably and they were discarded. The model selected finally considered 5 latent variables (LV, see Fig. 2, and was quite satisfactory. Note that this methodology can definitely be adequate to other ranges of values (typically, around 75 % which are common in many deposits worldwide) as long as enough samples are available to get a calibration, which was not the case here.

The external validation set to fine-tune the model consisted of 13 external samples, not included in the model at all, which confirmed its good predictive properties. It takes account of a large amount of information in the spectra and in the parameter of interest (methane), and

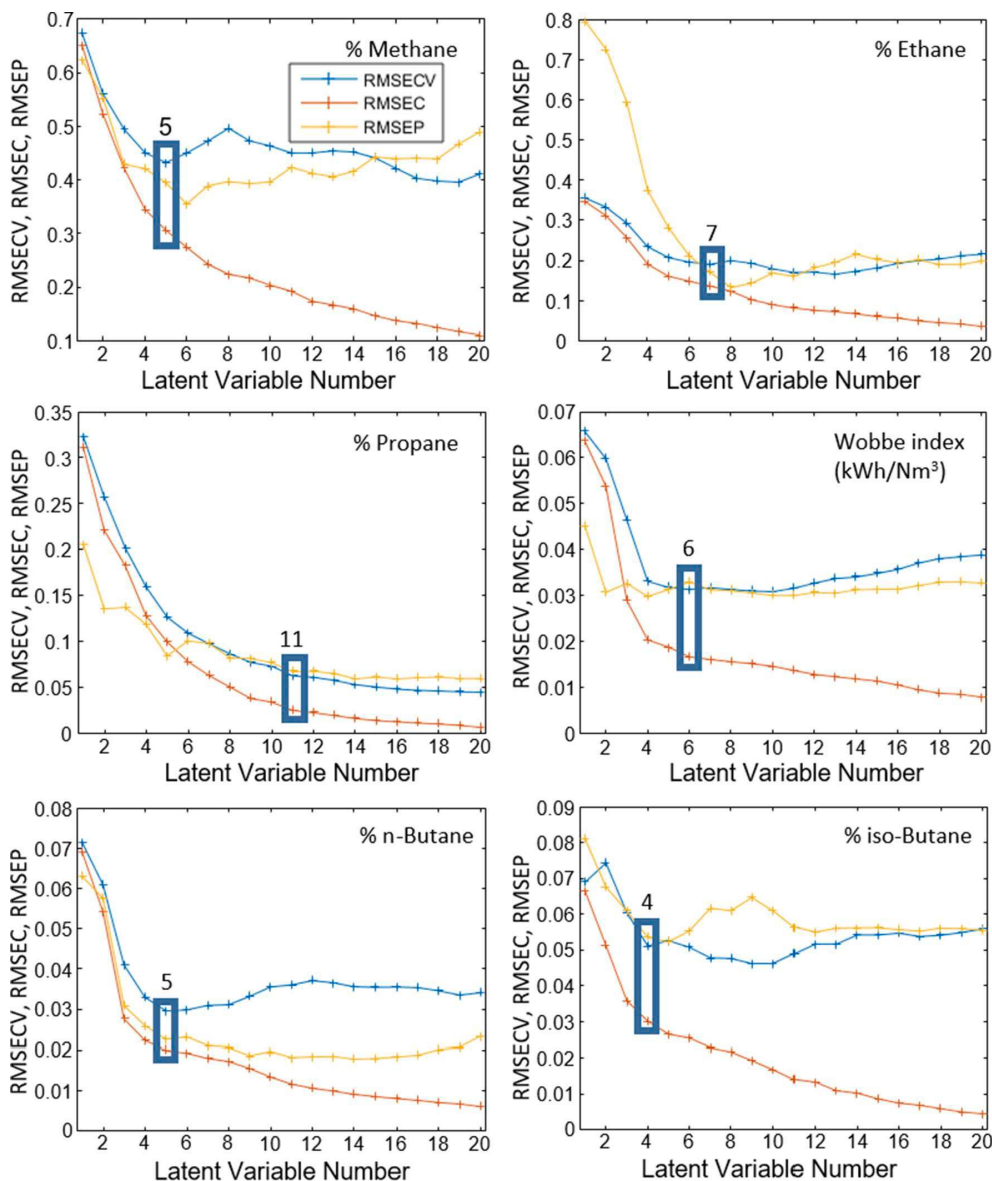


Fig. 2. Global average errors for different PLS models on calibration and validation for each studied parameter. The boxes indicate the number of latent variables selected for each model. RMSEC = root mean square error of calibration, RMSECV = root mean square error of cross-validation, RMSEP = root mean square error of prediction (fine-tuning external set), all them in vol%.

yields a high coefficient of determination, 98.5 % (Table 1), so that a huge amount of information related to methane is explained.

Fig. 3 presents the typical accuracy line ('actual vs predicted' line) for calibration, which summarizes the predictive performance of the model. The average overall error for calibration, measured as RMSEC, was ca. 0.3 vol% methane. The model has not a statistical bias (Table 1 and Fig. 3) nor relevant outliers, as seen at Figure SM1 (supplementary material), which shows two common diagnostic statistics, the Q residuals and the Hotelling T² statistics (they two yield the so-called, applicability domain) [13].

Further validation with an external test set comprising 20 NG samples distributed throughout the working range was also satisfactory. Fig. 4 presents its predictions and Table 1 resumes some relevant associated statistics. The RPD index (>3) shows that accurate predictions can be expected throughout the working range, without bias and a predictive average standard error ca. 0.4 vol% of methane. The RMSEC and RMSEP values can be compared directly with the precision of the GC method (method-reproducibility) thanks to the absence of bias. They are of the same order as the 0.37–0.39 vol% reproducibility values obtained for methane using the chromatographic method in this range (Table 2).

Indeed, 95 % of the external validation samples yield prediction differences (=predicted – reference) smaller than the maximum GC reproducibility allowed by the ISO 6975:1997 [42]. Fig. 5 depicts the reference compositional values derived from the GC measurement, along with their 95 % confidence intervals, superimposed to the MIR + PLS predicted values, also with their sample-specific 95 % PLS confidence intervals (calculated as formulated elsewhere [43]). The key idea is that to get statistically unbiased predictions both confidence intervals must overlap and this occurs for 100 % of the validation samples. A note about the interpretation of Table 2 and Fig. 5 is in order: note that the reproducibility values given in table 2 act as confidence intervals of the GC values alone. When the number of predictions that are excluded from this range are counted, a 'worst scenario' situation is considered. In effect, as the PLS predictions themselves have associated confidence intervals it may happen that even when a PLS prediction is out of the GC reproducibility range the PLS-predicted and GC-values do agree (statistically), when both confidence intervals overlap. Hence, the use of sample-specific intervals for the predictions are of most importance

Table 1

Performance parameters associated to each selected model for calibration and validation; N = number of samples used in the model after removing outliers. X represents the spectra and Y represents the parameter of interest, see text for more details. The parameters specified for the selected variable reduction method are shown between brackets.

		C1	C2	C3	n-C4	i-C4	Wobbe Index
Calibration	N	71	73	65	60	50	65
	LVs (PLS)	5	7	11	5	4	9
	RMSEC (vol%)	0.3	0.1	0.03	0.02	0.03	0.02
		[SR: 0.4]	[SR: 0.2]	[SR: 0.04]	[iPLS: 0.007]	[iPLS: 0.004]	[SR: 0.05]
	RMSECV (vol%)	0.4	0.2	0.06	0.03	0.05	0.03
		[SR: 0.4]	[SR: 0.4]	[SR: 0.05]	[iPLS: 0.009]	[iPLS: 0.009]	[SR: 0.07]
	Total % info explained in X	93.40	98.86	99.44	81.58	78.83	93.42
	Total % info explained in Y	98.49	99.46	99.80	95.63	92.48	99.16
	R ²	0.985	0.994	0.998	0.956	0.894	0.992
	BIAS	0	0	0	0	0	0
	LOD (vol%)	–	0.264	0.052	0.020	0.024	–
			[SR: 0.161]	[SR: 0.033]	[iPLS: 0.0072]	[iPLS: 0.0064]	
	LOQ (vol%)	–	0.880	0.174	0.067	0.082	–
		[SR: 0.458]	[SR: 0.097]	[iPLS: 0.019]	[iPLS: 0.018]		
x _d (vol%)	–	1.11	0.27	0.107	0.108	–	
x _q (vol%)	–	3.24	0.80	0.331	0.336	–	
1st Validation (fine-tuning)	N	13	15	12	11	16	14
	RMSEP (vol%)	0.2	0.2	0.1	0.02	0.04	0.05
	RPD	2.84	11.01	3.40	4.80	2.79	3.38
	R ²	0.993	0.993	0.932	0.962	0.873	0.913
External Validation	N	20	24	17	18	27	20
	RMSEP (vol%)	0.4	0.2	0.03	0.03	0.05	0.03
		[SR: 0.4]	[SR: 0.3]	[SR: 0.03]	[iPLS: 0.008]	[iPLS: 0.007]	[SR: 0.03]
	RPD	5.83	11.8	15.7	3.38	2.38	4.85
	R ²	0.971	0.991	0.990	0.919	0.847	0.959

(Fig. 5).

The average predictive error (RMSEP) obtained in this model compares nicely to literature, see Table 3, mostly considering that we used true NG samples whereas the other approaches considered mostly synthetic mixtures.

3.2. Predictive model for % of ethane

The model established to predict the vol% of ethane considered a calibration set with 73 industrial LNG samples in the 0 – 6 % volume molar range concentration and 7 LV (see Fig. 2). It yielded a good calibration (Fig. 3) with reasonably good statistics (Table 1). Although a slight improvement could be seen for the fine-tuning test set when 8 LV were considered (Fig. 2) it was only marginal and was not observed for the external validation set. The model explained ca. 99 % of the information in the spectral and concentration domains, and the coefficient of determination was very high (99.4 %), as well as the RPD index (>11).

No obvious suspicious samples were seen on the applicability domain plot (Figure SM1, supplementary material). No bias was observed neither for calibration (Table 1 and Fig. 3), nor for validation (Table 1 and Fig. 4) and, so, it is worth noting that the average prediction errors for the samples of the calibration, fine-tuning and external validation sets (i.e., 0.2 vol% ethane, Table 1) were comparable to the GC method-reproducibility range (Table 2). When the external validation set of samples was considered, only 37.5 % of them yielded predictions within the ISO 6975:1997 maximum GC reproducibility values. However, Fig. 5 indicates that 96 % of the 95 % confidence intervals associated to the reference and predicted values overlap. Two samples gave bad predictions although their behaviour was very good for the other parameters so we could not find a reason for that point. In addition, the average error (RMSEP) obtained when validating this model is of the same order as the best ones reported using NIR (Table 3).

Both the modern and classical limits of detection and quantification (x_d and x_q, and LOD and LOQ, respectively; see Supplementary Material) were calculated (see Table 1). The classical limits were calculated following the Eurachem Guide [44] which recommends using the 10 samples with the lowest concentration values, along with blanks or samples without the analyte. As there were no industrial samples with

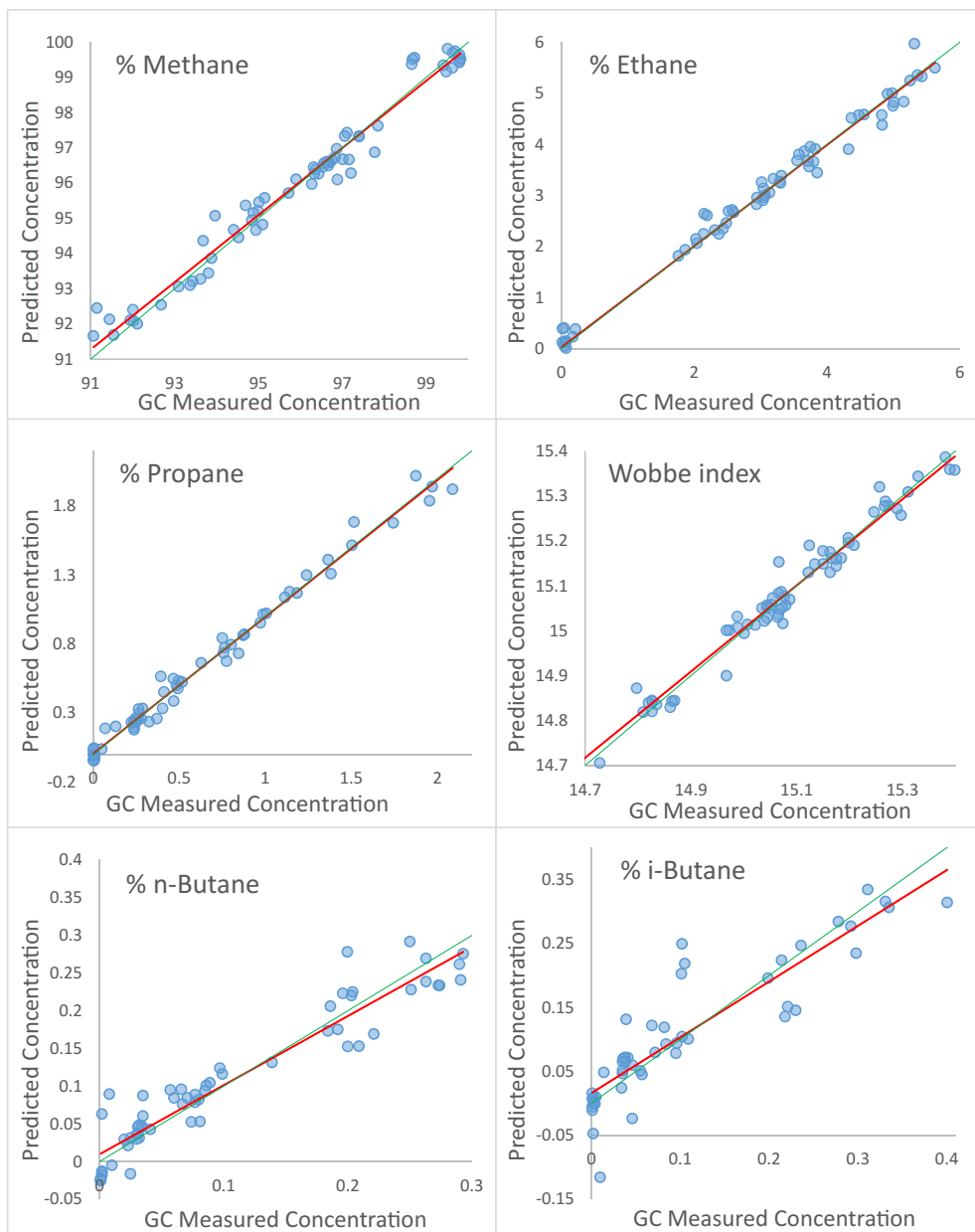


Fig. 3. Calibration accuracy lines for the PLS models selected for each predicted parameter. The green line depicts the theoretical perfect prediction (45° line) whereas the red one shows the actual predictive behaviour of the model. The units for the x and y axes are % vol molar concentration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0 % ethane concentration the samples used as 'zero' had concentrations ranging from 0.01 to 0.027 vol% ethane. With respect to the classical limit of detection (LOD) very few comparisons could be established with literature because of the scarcity of reported values, and none of them reported the modern definitions (x_d , x_q). All LODs are clearly higher than the chromatographic ones (Table 3), which is the reference methodology. However, positively enough, the mid-IR approach leads to lower LODs than the NIR ones.

3.3. Predictive model for % of propane

The model to predict the concentration of propane was developed using a calibration set consisting of 78 LNG samples in a range between 0 and 2.1 % volume molar concentration of propane. A model with 11 LVs (see Fig. 2) yielded a good calibration (Fig. 3 and Table 1). The need for this rather high number of LVs was attributed to the requirement for considering minor spectral signals which can correlate positively with propane (likely because its spectral bands overlap strongly with those from other NG components). In fact, a bit more than 99 % of the

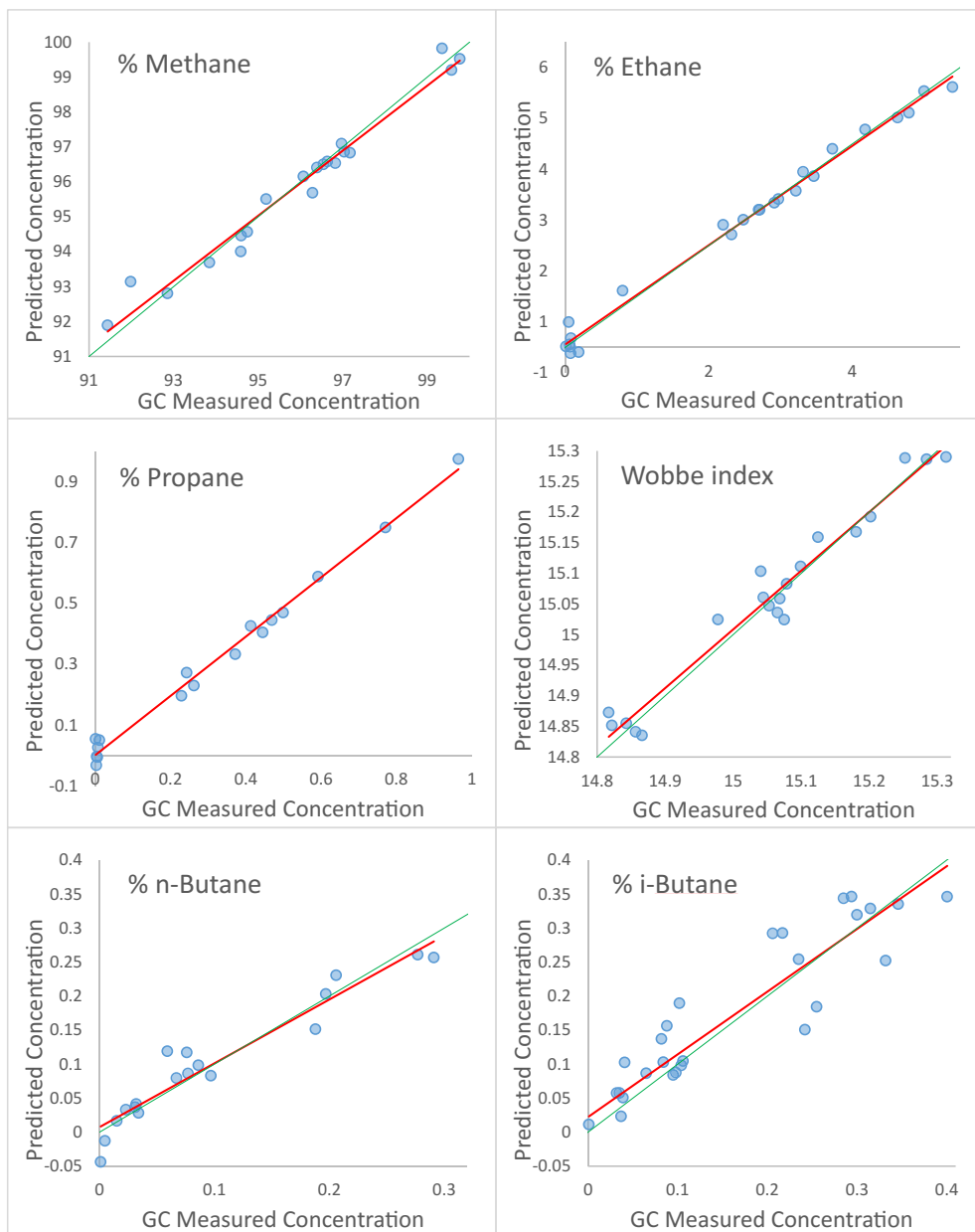


Fig. 4. Predictions for the external test set using the PLS selected model for each parameter. The green line depicts the theoretical perfect prediction (45° line) whereas the red one shows the actual predictive behaviour of the model. The units for the x and y axes are % vol molar concentration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

information in the spectral domain was required to explain the concentration of propane ($R^2 = 0.998$, Table 1). No obvious suspicious samples were seen on the applicability domain plot (Figure SM1, supplementary material) and no relevant bias was observed for calibration or validation (Figs. 3 and 4, and Table 1).

The model was validated using external NG samples. The RMSEP values were similar to the RMSECV one (0.03 vol% – 0.1 vol%) so no overfitting occurred. The model was not biased (Table 1) and its precision was comparable to the GC reproducibility (Table 2). In fact, 73.3 % of the predictions of the validation samples were within the ISO

6975:1997 maximum GC reproducibility and, also, 100 % of them overlapped their confidence sample-specific intervals with the GC ones (Fig. 5).

The limits of detection and quantification were calculated (Table 1). As for ethane, there were not enough industrial samples with zero propane concentration. Thus, the LOD and LOQ were calculated considering samples with concentrations ranging from 0 to 0.004 vol%. The LOD was higher than the GC ones although quite similar to that reported by Haghi et al [13] and much better than that of Rivessi et al [21] (Table 3). The average error (RMSEP) obtained when validating this model is clearly

Table 2

Range-dependent overall-method reproducibility values calculated for the gas chromatography reference method (indicated as the reproducibility for the lowest and highest concentrations of each component). The percentages between parentheses indicate the number of external samples whose predictions using gas-phase FTIR + PLS are within these reproducibility values, for the full-spectrum and reduced models.

	GC calibration range(Vol.% or kWh/Nm ³)	GC method ISO 6975:1997 reproducibility values(% of predictions within the reproducibility ranges)
Methane	99.850–93.220	0.37–0.39 (PLS: 95; SR: 100)
Ethane	3.002–0.010	0.0092–0.11 (PLS: 37.5; SR: 42)
Propane	3.115–0.003	0.0032–0.56 (PLS: 73.3; SR: 53)
n-Butane	0.717–0.002	0.0032–0.034 (PLS: 77.8; iPLS: 83)
i-Butane	0.722–0.002	0.0032–0.038 (PLS: 63.0; iPLS: 96)
Wobbe index	14.866–14.552	0.04 (PLS: 80; SR: 85)

better than those reported for NIR (Table 3).

3.4. Predictive model for % of Butane.

The models to predict the concentrations of n- and i-butane were developed using calibration sets including 60 and 50 industrial LNG samples, respectively, whose concentration ranges were between 0 % and 0.3 vol%, and between 0 and 0.4 vol%, each. The optimized models considered 5 LVs for n-butane, and 4 LVs for i-butane (see Fig. 2). They yielded good calibrations, without bias and quite good performance parameters (Fig. 3 and Table 1). No obvious outliers were seen on the models (Figure SM1, supplementary material). The models were validated using external fine-tuning and test sets. The RMSEP values were similar to the RMSECV ones (0.02 vol%–0.03 vol%, n-butane, and 0.05 vol%–0.06 vol%, i-butane) so no overfitting occurred. In addition, those values were similar to the ISO 6975:1997 maximum GC reproducibility precision, being the number of predictions within those figures 77.8 % of the external validation samples for n-butane, and 63.0 % for i-butane (Table 2). Fig. 5 shows that 100 % of the confidence intervals overlap for the validation samples.

Despite quite good models were obtained, they have low RPDs (ca. 3, Table 1), which seems to be caused mainly by the reduced working ranges. The average errors (RMSEP) obtained when validating this model with true NG samples are slightly higher (although of the same order) than those reported using NIR using synthetic mixtures [13] although much better than those reported for the sum of i + n butanes [20,21] (Table 3). The limits of detection and quantification were also calculated (see Table 1). In this case there were 15 production samples without butane, so that they constituted a true zero. The LODs were better than those reported for NIR [13,21]; all IR-based approaches had bigger LODs than the GC ones (Table 3).

3.5. Predictive model for the Wobbe index

The model developed for the Wobbe index considered 65 NG samples ranging from 14.5 to 15.5 kWh/Nm³. This interval does not include the low values associated to some samples, that can be as low as 12.6 kWh/Nm³ in some countries [45]. Despite some very few samples had WIs between 13.0 and 14.5 kWh/Nm³ they disrupted the models. This was attributed to their contents on N₂ as it decreases dramatically the WI. The problem here is that N₂ is transparent to the mid-IR radiation, thus making it hard for the models to take it into account. This yielded a broad dispersion of the predictions at the lowest values of the calibration. Therefore, those particular samples were not considered in the models.

After several preliminary studies (Fig. 2 was not conclusive by itself) with the fine-tuning dataset 6 LV were fixed, which yielded a good calibration (Fig. 3 and Table 1), without bias and no outliers (Figure SM1, supplementary material). The RMSECV and RMSEP average errors for the calibration and the fine-tuning and external validation sets (Fig. 4) were similar (ca. 0.03–0.05 kWh/Nm³), suggesting that overfitting did not occur. These errors (0.28 % as relative

error, for the experimental range of values) compare very well to other publications reporting relative errors between 1 % and 0.03 % [25]. Other authors reported even higher relative errors, like 1.5 % [29] or 14 % [27]. Although Brown et al [28] reported average errors around 0.5 %, those corresponded to energy predictions, using BTU units.

As ISO 6975:1997 does not contemplate the WI, the precision figures calculated for this parameter using the MIR+PLS approach were compared to the experimental reproducibility (strictly, intermediate precision) obtained by Reganosa using GC (ca. 0.04 kWh/Nm³). They are similar (Table 2), and 80 % of the predictions of the validation samples became within the reproducibility range. Notwithstanding, 100 % of the sample-specific confidence intervals of the validation samples overlapped with the reference ones (Fig. 5).

3.6. Validation with synthetic gas mixtures

As mentioned in the experimental part, 27 gas mixtures from two European projects were considered. Some of them were used to check two new sensors, one based on electrochemical membranes and a TFIR system [7]. Despite some of these mixtures were only binary or ternary ones (and, so, quite different from the industrial NG samples employed in the models above), it was considered interesting to predict them as a benchmark activity to see whether the MIR + PLS approach could predict them. Table 3 summarizes the average prediction errors (as RMSEP) and despite they are of somewhat lower quality than those from the real NG samples (as expected because of the spectral differences), they are very encouraging. The RMSEPs (as they had not bias, they can be immediately compared with standard deviations) for methane and ethane are midway between those of the other two methods [7] while for propane, the MIR-PLS approach yields slightly better results. The predictions for the two isomers of butane yielded only semiquantitative results because many synthetic mixtures were out of the calibration range of the models and they contained much more butanes (0–2 vol% range) than our usual NG samples (0–0.45 vol% range).

3.7. Variable reduction methods

As mentioned in the experimental section, the spectra measured in this work contain as many as 9375 variables/spectrum because of the need to register the sharp IR spectral peaks of the gaseous components at high resolution. This can be a problem when time is an issue (here, at least 30 min/spectrum were needed to get a high signal/noise ratio, including the background), when implementing probes or portable equipment, or when some of those wavenumbers offer no relevant information and degrade the predictions. Hence, spectral variable reduction appears as a nice option. Disappointingly, as mentioned in the Supplementary Material, many variable selection strategies do not set absolute thresholds for their statistics [46] and these have to be established ad-hoc. Here, we took advantage of the many trials the algorithms performed to set a strategy rooted on classical quality control. As for traditional quality control charts, average values of the statistics associated to all the variables (e.g., the reliability (of the MCUVE approach)

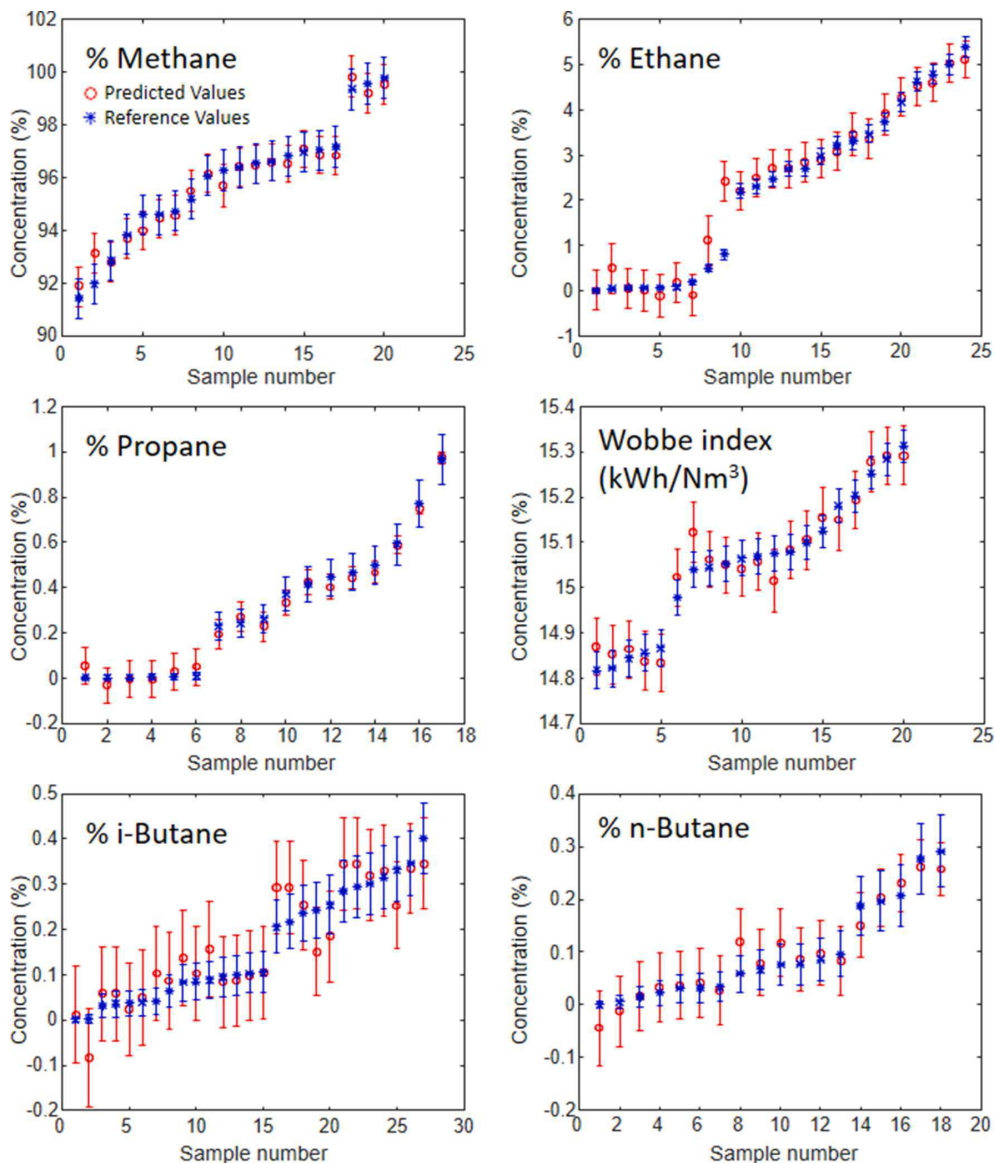


Fig. 5. Reference (blue asterisk) with the ISO 6975:1997 max. reproducibility (Reganosás in the case of the Wobbe index) and predicted (red circle) values with their respective 95 % confidence intervals for the external calibration samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

and the selection probability (for random frog) can be calculated, along with their 2 and 3 standard deviation upper limits. Hence, any variable outside the upper control limit is deemed to be selected; see next paragraphs for an example.

The number of selected spectral variables was varied from 10 to 400 for all five approaches and all analytical parameters (composition plus Wobbe index). Further, the number of latent variables was also optimized for each trial. As this generates a very large number of models, to choose a final one we proceeded as follows (for each analytical parameter): first, for each selection strategy the number of spectral variables leading to best predictions was chosen among the different trials (the number of latent variables in each PLS models were optimized as well); second, the best candidates of the five variable reduction methods were

compared; third, whenever several 'best' candidate models performed approximately the same, that with less spectral variables and/or with the spectral variables more concentrated in particular spectral regions would be preferred. This criterion was applied at the final stage of the selection because one of our objectives was to simplify the measuring stage in the industrial laboratory, which is trivial whenever the selected variables become close to each other.

As a general result, it was found that many models considering reduced suites of variables outperformed those considering full-spectra. However, the 'best' models disagreed on the number of wavenumbers they considered, as expected because of the different criteria involved in the approaches (see [Supplementary Material](#)), and they varied also with the property under consideration (as it is logical because relevant

Table 3

Average errors obtained in this work (RMSEP) compared to those reported in literature using NIR and other methods. The LODs are shown between parentheses (when available) and the calibration ranges, as percentage, between brackets. The chromatographic LODs are shown as a reference. All values represent vol% molar concentration of the constituent. Note that RMSEPs can be related to SD because the models in this work are not biased (see text for details). SD means standard deviation.

	Range		C1	C2	C3	i-C4	n-C4
LNG samples in this work	MIR:(PLS)	RMSEP	0.4	0.2	0.03	0.03	0.05
		(LOD)	– [99–89]	(0.26)[6–0]	(0.05)[2–0]	(0.02)[0.4–0]	(0.02)[0.4–0]
	MIR (reduced)	RMSEP	SR: 0.4[99–89]	SR: 0.3[6–0]	SR: 0.04[2–0]	iPLS: 0.008[0.4–0]	iPLS: 0.007[0.4–0]
		(LOD)	– [99–89]	SR: 0.161[6–0]	SR: 0.033[2–0]	iPLS: 0.0072[0.4–0]	iPLS: 0.0064[0.4–0]
Synthetic Mixtures in this work	MIR:	RMSEP	1.04[95–89]	0.49[14–0]	0.28[5–0]	–	–
		(LOD)	– [99–89]	–	–	–	–
Electrochemical sensor (SD) [7]	MIR:	SD	0.58[95–89]	0.62[10–0]	0.34[5–0]	0.15[2–0]	0.22[2–0]
TF-IR sensor (SD) [7]	MIR	SD	1.23[95–89]	0.13[10–0]	0.44[5–0]	0.34[2–0]	0.07[2–0]
Haghi et al [13]	NIR	RMSEP	0.20	0.19	0.16	0.012	0.012
		(LOD)	(0.90)[100–80]	(0.39)[12–0]	(0.33)[8–0]	(0.13)[2–0]	(0.15)[2–0]
Rohwedder et al[20]	NIR	RMSEP	0.37[82–69]	0.36[17–6]	0.67[14–5]	0.37[7–2]	–
Makhoukhi et al[18]	MIR/NIR	RMSEP	0.2[96–85]	0.2[10–2]	0.1[4–0]	–	–
Ribessi et al [21]	NIR	RMSEP	0.77	0.75	1.63	0.54	–
		(LOD)	(0.42)[90–30]	(0.59)[50–10]	(1.42)[30–3]	(1.67)[20–1]	–
Cao et al [22]	NIR	RMSEP	0.95[0.46–0 ppm]	–	–	–	–
Udina et al [29]	Thermic sensor	RMSEP	0.6[100–72]	1.0[20–0]	–	–	–
GC method	–	LOD	– [99–68]	0.01[12–0]	0.003[3–0]	0.002[0.7–0]	0.002[0.7–0]

variables for methane may not be so for –e.g. i-butane). This complicates any comparison and makes it difficult to decide which approach is best in an absolute sense. As a consequence, a trade-off was made for each analytical parameter.

Table 1 depicts the variable reduction method of choice for each property, their RMSEC, RMSECV, RMSEP, and limits of detection and quantification. First of all, note the improvements achieved when the variable subsets are used as those performance parameters became half or a third those of the full-spectrum ones. Of most importance is the very large enhancement associated to i- and n-butane, which are particularly difficult to address because their bands overlap not only among them but with methane and ethane.

The feature reduction method selected for methane, ethane, propane and WI was SR (the selectivity ratio index, see [Supplementary Material](#)), with 3 LV (50 wavenumbers), 6 LV (100 wavenumbers), 6 LV (50 wavenumbers) and 6 LV (250 wavenumbers), respectively (see [Fig. 6](#) and [Figure SM2, supplementary material](#)). Despite the wavenumbers selected for methane became not too grouped ([Fig. 7](#)), they indeed formed nice, definite regions for ethane, propane and WI ([Figure SM3, supplementary material](#)).

A good alternative for the former three properties may be VIP (variable importance in projection, [Supplementary Material](#)) as it chose those limited spectral ranges ([Fig. 7](#) and [Figure SM3, supplementary material](#)) although with slightly higher prediction errors. For the WI good

alternatives might be random frog or common PLS (although all the spectrum needs to be measured). iPLS performs quite well although at the expense of selecting very sparse sets of variables (so all the full spectrum would still be required). Random frog yielded also sparse selections while MCVUE (Monte Carlo uninformative variable elimination, [Supplementary Material](#)) and VIP lead to two very well defined groups around variables 5000–5500 and 1500–1800 (i.e., 2650–2900 and 4430–4580 cm^{-1} , respectively). See [Supplementary Material](#) for a general chemical interpretation of the most relevant regions. [Fig. 8](#) exemplifies how thresholds for MCVUE and random frog were set following the ‘quality control chart’ criterion depicted above.

Note the very stable predictions SR yields for ethane and Wobbe index, where the dimensionality is not a critical factor ([Figure SM2, Supplementary Material](#), in the figure it is called SRI to stress its indexing nature). This is a very positive result for quality control purposes as parsimony can be applied, however in this paper we only focused on absolute minima to simplify the comparisons and discussions.

Analogous situations were found for the other analytical parameters ([Supplementary Material, Figures SM2 and SM3](#)), i- and n-butane are predicted nicely by iPLS (200 wavenumbers and 6 LV in both cases; 4500–2450 cm^{-1} for i-butane and 4100–950 cm^{-1} for n-butane). The alternative for i-butane (although with higher RMSEP) was random frog (10 wavenumbers), with a very stable behaviour when dimensionality is

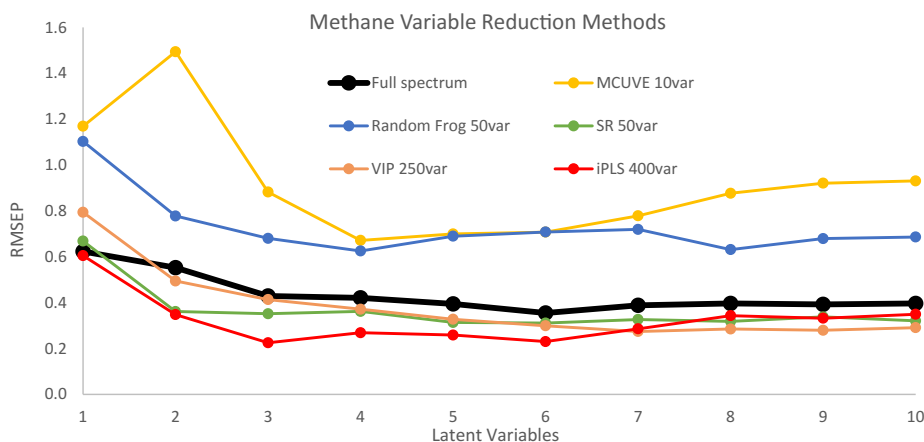


Fig. 6. Errors of prediction (RMSEP, vol%) for each latent variable and variable reduction method for the methane determination.

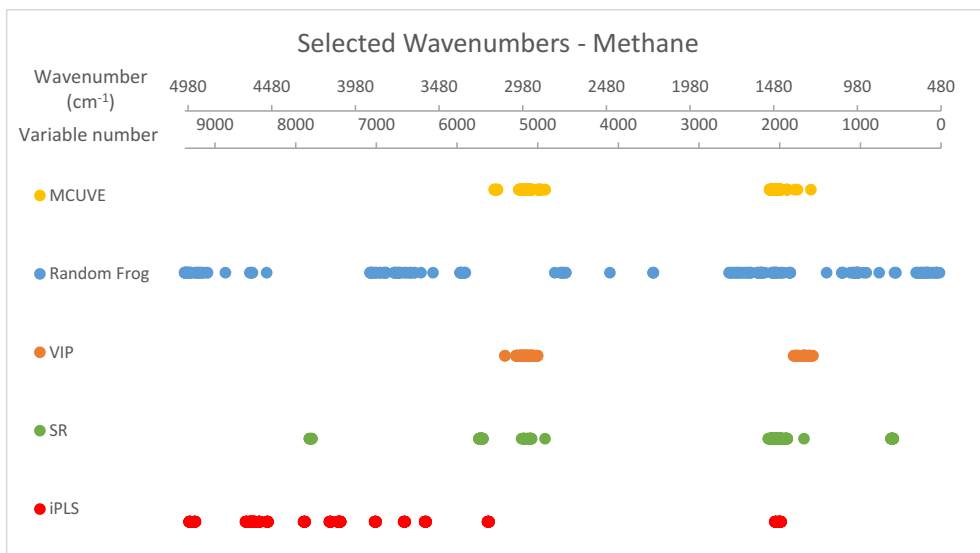


Fig. 7. Selected variables for each of the tested methods for the methane determination.

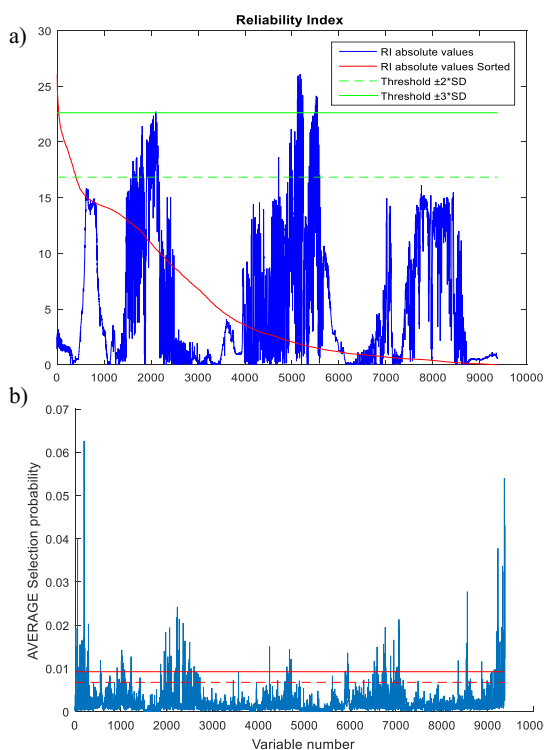


Fig. 8. Exemplification of how a selection threshold can be set considering the average and classical upper quality control limits ($\pm 3SD$) for the MCUVE (a) and random frog (b) indices calculated for methane (SD stands for standard deviation). See text for details.

considered and an average error almost as good. Nevertheless, those 10 variables spanned throughout the overall original spectrum, whereas for iPLS the intervals were collateral. For n-butane, SR (50 wavenumbers)

may be a reasonable alternative, although with ca. 50 % more error.

As a final remark, it is worth noting the very 'homogeneous' overall performance of iPLS throughout the various analytical parameters as it had a quite good behaviour almost always, but for the Wobbe index. However, it required high processing times ($>8h$ for a 9375 to 400 variables reduction) and it can only select intervals of variables. SR and VIP performed rather similarly to iPLS although much faster and, indeed, SR was the method of choice for methane, ethane and propane, mostly because it 'clustered' the selected wavenumbers (see, e.g. Fig. 7) and it appeared also as a good alternative for n-butane.

MCUVE and Random frog demonstrated an intermediate behaviour, often similar to iPLS, with the advantages that they required much reduced computing times and can select discreet variables. The former tends to group them slightly more than iPLS, while random frog tended to select much more dispersed wavenumbers.

4. Conclusions

This work demonstrates that the combination of mid-IR spectrometry and PLS regression yields regression models that predict the primary composition of natural gas reliably. The approach is cost effective because it requires standard laboratory instrumentation, industrial samples (not standard mixtures) and requires ca. 45 min to get the main composition of natural gas, along with an important physicochemical property, the Wobbe index (this included recording the sample spectrum, ca. 35 min, plus the software application to run the models). Further, it can be considered a green method because it does not demand chemicals (but some argon), uses very little sample (ca. 500 mL) and does not generate residues nor cleaning steps.

The models were unbiased in the typical industrial working ranges, with precisions comparable to the chromatographic reference method. The mid-IR-PLS approach compares advantageously to other NIR-based methods, with better average errors and better limits of detection (or, in the worst case, similar to the best NIR ones).

The models developed with reduced sets of variables clearly decreased the average predictions errors and limits of detection and quantification and, so, outperformed previous reports using other spectral regions. As a trade-off solution, the selectivity ratio index, SR, seemed the most satisfactory alternative because it was selected for four parameters and could be a good alternative for another one. iPLS

behaved also very well, achieving the lowest errors to determine the butanes (although it selects variables across all the spectrum). The prediction capabilities of the advanced random frog method were good but it selected very sparse variables.

Declaration of Competing Interest

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

Acknowledgement

The Program 'Consolidación e Estructuración de Unidades de Investigación Competitiva' of the Galician Government (Xunta de Galicia) is acknowledged (Grant: ED431C-2017/28). Funding for Open Access charges was granted by Universidade da Coruña/CISUG.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.fuel.2021.121500>.

References

- Hausfather Z. Analysis: Fossil-fuel emissions in 2018 increasing at fastest rate for seven years 2018. <https://www.carbonbrief.org/analysis-fossil-fuel-emissions-in-2018-increasing-at-fastest-rate-for-seven-years> (accessed March 10, 2021).
- Banawan AA, El Gohary MM, Sadek IS. Environmental and economical benefits of changing from marine diesel oil to natural-gas fuel for short-voyage high-power passenger ships. *Proc Inst Mech Eng Part M J Eng Marit Environ* 2010;224:103–13. <https://doi.org/10.1243/14750902JEME181>.
- Hosseini M, Dincer I, Ozbilin A. Expert Opinions on Natural Gas Vehicles Research Needs for Energy Policy Development. Elsevier 2018. <https://doi.org/10.1016/B978-0-12-813734-5.00041-X>.
- European Commission. COM/2013/017 Communication from the Commission to the European Parliament, the Council, the Economic and Social Committee and the Committee of the Regions Clean Power for Transport: A European alternative fuels strategy; 2013.
- European Commission. Directive 2014/94/EU - Deployment of alternative fuels infrastructure. *Off J Eur Union* 2014;L307:20.
- Eilts P, Klare L. Investigations on the Determination of the Service Methane Number of LNG. SAE Tech. Pap., vol. April, 2018, p. 1–11. <https://doi.org/10.4271/2018-01-1143>.
- Sweelssen J, Blokland H, Rajamäki T, Boersma A. Capacitive and Infrared Gas Sensors for the Assessment of the Methane Number of LNG Fuels. *Sensors* 2020;20:3345. <https://doi.org/10.3390/s20123345>.
- Malenshek M, Olsen DB. Methane number testing of alternative gaseous fuels. *Fuel* 2009;88:650–6. <https://doi.org/10.1016/j.fuel.2008.08.020>.
- Ponte S, Andrade JM, Vázquez C, Ferreira B, Cobas C, Pérez A, et al. Prediction of the methane number of commercial liquefied natural gas samples using mid-IR gas spectrometry and PLS regression. *J Nat Gas Sci Eng* 2021.
- Emerson Process Management. The Wobbe Index and Natural Gas Interchangeability; 2007.
- ISO 6976:2016 Natural gas — Calculation of calorific values, density, relative density and Wobbe indices from composition; 2016.
- ISO 6974-4: Natural gas—Determination of composition with defined uncertainty by gas chromatography—Part 4: Determination of nitrogen, carbon dioxide and C1 to C5 and C6+ hydrocarbons for a laboratory and on-line measuring system; 2000.
- Haghi RK, Yang J, Tohid B, Fourier Transform Near-Infrared (FTNIR) Spectroscopy and Partial Least-Squares (PLS) Algorithm for Monitoring Compositional Changes in Hydrocarbon Gases under In Situ Pressure. *Energy Fuels* 2017;31:10245–59. <https://doi.org/10.1021/acs.energyfuels.7b01677>.
- Ferreiro B, Andrade JM, Paz-Quintás C, López-Mahía P, Muniategui-Lorenzo S, Rey-Garrote M, et al. Improved Sensitivity of Natural Gas Infrared Measurements Using a Filling Gas. *Energy Fuels* 2019;33:6929–33. <https://doi.org/10.1021/acs.energyfuels.9b00549>.
- Dong C, O'Keefe M, Elshahawi H, Hashem M, Williams S, Stensland D, et al. New downhole-fluid-analysis tool for improved reservoir characterization. *SPE Reserv Eval Eng* 2008;11:1107–16. <https://doi.org/10.2118/108566-pa>.
- Dantas HV, Barbosa MF, Nascimento ECL, Moreira PNT, Galvão RKH, Araújo MCU. Screening analysis of natural gas with respect to methane content by near-infrared spectrometry. *Microchem J* 2014;114:210–5.
- Van Aghoven MA, Fujisawa G, Rabbito P, Mullins OC. Near-infrared spectral analysis of gas mixtures. *Appl Spectrosc* 2002;56:593–8.
- Makhoukhi N, Péré E, Creff R, Pouchan C. Determination of the composition of a mixture of gases by infrared analysis and chemometric methods. *J Mol Struct* 2005; 744–747:855–9. <https://doi.org/10.1016/J.MOLSTRUC.2005.01.021>.
- Hakuli A, Kyttkivi A, Lakomaa E, Krause O. FT-IR in the Quantitative Analysis of Gaseous Hydrocarbon Mixtures. *Microchem Ltd* 1995;67:1881–6. <https://doi.org/10.1021/ac00107a019>.
- Rohwedder JJR, Pasquini C, Fortes PR, Raimundo IM, Wilk A, Mizaikoff B. iHWG- μ NIR: a miniaturised near-infrared gas sensor based on substrate-integrated hollow waveguides coupled to a micro-NIR-spectrophotometer. *Analyst* 2014;139:3572. <https://doi.org/10.1039/c4an00556b>.
- Ribessi RL, Neves TDA, Rohwedder JJR, Pasquini C, Raimundo IM, Wilk A, et al. IHEART: A miniaturized near-infrared in-line gas sensor using heart-shaped substrate-integrated hollow waveguides. *Analyst* 2016;141:5298–303. <https://doi.org/10.1039/c6an01027j>.
- Cao H, Yan X, Li Y, Wang Y, Zhou Y, Yang S. A component prediction method for flue gas of natural gas combustion based on nonlinear partial least squares method. *Sci World J* 2014;2014. <https://doi.org/10.1155/2014/418674>.
- Patent Application Publication (10) Pub. No.: US 2010 / 0035098 A1 Patent Application Publication 2010;1:1–5.
- Loubar K, Rahmouni C, Le Corre O, Tazerout M. A combustionless determination method for combustion properties of natural gases. *Fuel* 2007;86:2535–44. Doi: 10.1016/j.fuel.2007.02.024.
- Roy PS, Ryu C, Park CS. Predicting Wobbe Index and methane number of a renewable natural gas by the measurement of simple physical properties. *Fuel* 2018;224:121–7. Doi: 10.1016/j.fuel.2018.03.074.
- Rahmouni C, Tazerout M, Corre O Le. Determination of the combustion properties of natural gases by pseudo-constituents. *Fuel* 2003;82:1399–409. Doi: 10.1016/S0016-2361(03)00029-2.
- Venturim RS, Ferreira LM, Pereira FG, Resende CZ. Desenvolvimento de um sensor virtual para o monitoramento de índice Wobbe na planta do LTQ da AcerloMittal Tubarão. *ABM Proc., São Paulo: Editora Blucher*; 2018, p. 149–57. Doi: 10.5151/2237-0234-31585.
- Brown CW, Lo SC. Feasibility of on-line monitoring of the BTU content of natural gas with a near-infrared fiber optic system. *Appl Spectrosc* 1993;47:812–5. <https://doi.org/10.1366/0003702934066974>.
- Udina S, Carmona M, Pardo A, Calaza C, Santander J, Fonseca L, et al. A micromachined thermoelectric sensor for natural gas analysis: Multivariate calibration results. *Sensors Actuators, B Chem* 2012;166–167:338–48. <https://doi.org/10.1016/j.snb.2011.11.086>.
- Sandfort V, Trabold B, Abdolvand A, Bolwien C, Russell P, Wollenstein J, et al. Monitoring the Wobbe Index of Natural Gas Using Fiber-Enhanced Raman Spectroscopy. *Sensors* 2017;17:2714. <https://doi.org/10.3390/s17122714>.
- Li HD, Xu QS, Liang YZ. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemom Intell Lab Syst* 2018;176: 34–43. <https://doi.org/10.1016/j.chemolab.2018.03.003>.
- Otto M. Chemometrics: statistics and computer application in analytical chemistry. John Wiley & Sons; 2016.
- Breton RG. Chemometrics: data analysis for the laboratory and chemical plant. John Wiley & Sons; 2003.
- Andrade-Garda JM, Carloseña-Zubieta A, Boqué-Martí R, Ferré-Baldrich J. Partial Least Squares Regression. In: Andrade-Garda JM, editor. *Basic Chemom. Tech. At. Spectrosc.* 2nd ed., Cambridge: Royal Society of Chemistry; 2013, p. 280–347.
- Gómez-Carracedo MP, Andrade JM, Rutledge DN, Faber NM. Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undesigned kerosene samples. *Anal Chim Acta* 2007;585:253–65. Doi: 10.1016/j.aca.2006.12.036.
- European Commission. COMMISSION DECISION of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results; 2002.
- Currie LA. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure Appl Chem* 1995;67:1699–723.
- Courtene-Jones W, Quinn B, Ewins C, Gary SF, Narayanaswamy BE. Consistent microplastic ingestion by deep-sea invertebrates over the last four decades (1976–2015), a study from the North East Atlantic. *Environ Pollut* 2019;244: 503–12. <https://doi.org/10.1016/j.envpol.2018.10.090>.
- ISO 11843-2:2000 Capability of detection — Part 2: Methodology in the linear calibration case; 2000.
- Ortiz MC, Sarabia LA, Herrero A, Sánchez MS, Sanz MB, Rueda ME, et al. Capability of detection of an analytical method evaluating false positive and false negative (ISO 11843) with partial least squares. *Chemom Intell Lab Syst* 2003;69: 21–33.
- Ortiz MC, Sarabia LA, Sánchez MS. Tutorial on evaluation of type I and type II errors in chemical analysis: From the analytical detection to authentication of products and process control. *Anal Chim Acta* 2010;674:123–42.
- ISO 6975:1997 Natural gas — Extended analysis — Gas-chromatographic method; 1997.
- Faber K, Kowalski BR. Improved prediction error estimates for multivariate calibration by correcting for the measurement error in the reference values. *Appl Spectrosc* 1997;51:660–5.

- [44] Magnusson B, Örnemark U. Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics. *Eur J Pharmacol* 2014;183–9.
- [45] European Network of Transmission System Operators for Gas. Wobbe Index and Gross Calorific Value in European networks - Analysis of ranges and variability; 2017:21.
- [46] Nørgaard L, Saudland A, Wagner J, Nielsen JP, Munck L, Engelsen SB. Interval partial least-squares regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy. *Appl Spectrosc* 2000;54:413–9. <https://doi.org/10.1366/0003702001949500>.

FAST QUALITY CONTROL OF NATURAL GAS FOR COMMERCIAL SUPPLY AND TRANSPORT UTILITIES

Borja Ferreiro¹, José Andrade^{1*}, Purificación López-Mahía¹, Soledad Muniategui¹, Cristina Vázquez², Andrés Pérez², María Rey², Carlos Vales²

¹ *Group of Applied Analytical Chemistry. University of A Coruña. Campus da Zapateira, s/n, E-15071, A Coruña, Spain.*

² *Regasificadora del Noroeste, S.A. Punta Promontorio, Mugar dos (Ferrol), Spain.*

(*) Corresponding author: andrade@udc.es; fax: +34981167065

SUPPLEMENTARY MATERIAL

SOFTWARE AND PARTICULAR CHEMOMETRIC DETAILS

From an operative viewpoint, **validation** can be internal (i.e., using a part of the samples employed during model development; typically this involves some form of *cross-validation*) or external (i.e., using new samples whose properties are known in advance). Both approaches were employed throughout this paper.

Undoubtedly, the critical step in PLS modelling is to set the adequate number of latent variables or factors (LV) to model the system. An efficient approach is to check for a minimum in the internal cross-validation procedure (where the number of LV is systematically varied). That was made by splitting randomly the training dataset in 10 subsets so that each one is predicted after all other nine

subsets were used to construct a PLS model. Dissapointingly this still may lead to overfitting models and truly new samples are required for a correct validation.

When it comes about **spectral variable selection**, five methods were studied here:

- Interval partial least squares (iPLS) works by dividing the data in spectral intervals (of a given width, that needs to be optimized or selected in advance) and performs a PLS model on each one. The interval that shows the best prediction capabilities is selected and ‘retained’. Then, other intervals can be added to the former and new models are studied to see whether the synergies between two intervals improve the model [46]. The process can be continued in the same way with more intervals. The final selection reflects the amount of information provided or the interferences conveyed by each interval (spectral region). The user can select a specific number of intervals and the interval width, thus allowing the reduction of a full spectrum to a controlled number of variables.
- Selectivity ratio index (SR or SRI) is the ratio between the explained and the residual variance for each variable. It shows the capability of a variable to discriminate specific properties of the analyte under study [SM1]. The larger the SR parameter is, the more important the variable becomes for the prediction. The variables with the highest SRs are selected *ad-hoc*.
- Variable Importance in Projection (VIP) is used to rank the variables according to their importance to predict a property; values greater than 1 highlight important variables while values close to 0 indicate not important ones. A VIP score is calculated for each variable as the sum, over the latent variables considered into the model, of its PLS-weight value weighted by the percentage of explained Y variance by each specific LV [SM2,SM3]. As for SR, the variables with the highest VIPs are selected *ad-hoc*.

- Monte-Carlo uninformative variable elimination (MCUVE) builds a very large number of models by randomly selecting the calibration samples (following a Monte Carlo sampling approach). Then, the stability of the correlation coefficient of each variable is studied by means of a reliability index, which evaluates the amount of systematic information that each variable offers. Uninformative variables are those with the more unstable coefficients [31,SM4].
- Random Frog is based on a special form of Monte Carlo-based search called ‘reversible jump Markov Chain Monte Carlo’ that allows this approach to search in the model space through moves between different models (dimensionality, etc.). The algorithm can be visualized as composed of three major stages [31]. First, a variable subset $[VS]_0$ containing w wavenumbers from the overall spectrum is initiated randomly ($1 < w < p$, p = total number total of wavenumbers). Second, propose another candidate variable subset $[VS]^*$ with w' wavenumbers (w' is chosen randomly from a normal distribution centred on w) and evaluate whether $[VS]^*$ can substitute $[VS]_0$. For this, compute the cross-validation prediction error using both subsets and if $\text{error}([VS]^*) < \text{error}([VS]_0)$ set $[VS]^*$ as $[VS]_1$ and replace $[VS]_0$. Even if $\text{error}([VS]^*) > \text{error}([VS]_0)$ by a maximum percentage $[VS]^*$ can still be accepted as $[VS]_1$. The process is repeated many times (e.g. 10000). Finally, calculate a selection probability for each variable (e.g. number of models that include such a variable divided by the total number of iterations).

iPLS, SR and VIP were from the PLS_Toolbox, while MCUVE and random frog were from the libPLS package [31].

A relevant problem when reducing the number of variables is to set a sort of threshold to consider only the most important ones. None of the five selection strategies employed here establish a final number of variables *per se* (although VIP can be the exception if all variables with $VIP > 1$ are considered important) and, so, different trials varying the metaparameters controlling the algorithms

and the number of selected variables must be done and the prediction capabilities of the different models tested and compared. This is especially cumbersome for the Monte Carlo-based approaches. The original developments did not define selection thresholds (see [46] and references cited therein) and, so, a pragmatic one was established here taking advantage of the many trials the algorithms perform. We used the concept of control charts, as detailed in section 3.7 of the main text.

The **performance of the prediction models** can be evaluated using the following statistics:

- The coefficient of determination (R^2) measures the proportion of variance of the dependent variable (here, the vol% of a compound in the mixture or the Wobbe index) that can be predicted from the predictor (or independent) variables (here, the spectra).
- Bias is the average difference between the predicted and reference 'y' values of the property under investigation. It measures the tendency of the model to predict the average component concentration over or under the average measured concentration.
- The standard error of performance (SEP) evaluates the overall precision of the predictions. It is a bias-corrected error of prediction.
- The ratio of prediction to deviation (RPD) is the quotient between the *standard deviation* of the predicted values and the *standard error of performance* ($RPD = \frac{SD}{SEP}$ or $\frac{SD}{RMSEP}$). It gives insight on the capability of the model to give an accurate prediction in the working range, being a value <3 an indicative of a weak model [34].
- The so-called limits of detection and quantification are two traditional parameters linked to calibration tasks. However, there was controversy on how to calculate them. Put simply, the - formerly called- limit of detection (LOD) and limit of quantification (LOQ) represent '...the lowest concentration of the analyte that can be detected by a method at a specified level of confidence' and '...the lowest level at which the performance of the prediction is acceptable

for a typical application', respectively [44]. Although still quite common nowadays, those old IUPAC definitions (that can be formulated as $LOD=3S_0$ and $LOQ=10S_0$, being S_0 the standard deviation of a blank) were superseded on 2002 (harmonized by IUPAC, ISO and the EU) and nowadays the risk type II -number of false negatives- must be taken into consideration [36–39]. The new limits (x_d and x_q , limits of decision and quantification –respectively-, considering the risk of false positives and false negatives simultaneously) are more realistic, comprehensive and secure for most applications, although they yield higher numerical values than the simpler, classical equations where only the risk for false positives is taken into account. Of most relevance here is that those calculations can be applied straightforwardly to multivariate regressions models using the simple accuracy line (the traditional 'real vs predicted' plot), as demonstrated elsewhere [40,41]. The only requirement is to have a collection of samples for which the reference values of the property under investigation are regressed against the model-predicted ones.

With respect to the **chemical interpretation of the most relevant spectral regions** for the reduced models, some general comments can be given although it is very difficult to interpret every selected wavenumber (as explained in the main text), not only because all compounds contain only C and H but because the IR band tables do not detail the positions of the bands for gas-phase compounds.

- i- For the Wobbe Index, only two small spectral regions were needed; essentially related to the classical CH regions: $1399-1524\text{ cm}^{-1}$, linked to CH_3 asymmetric bending overlapped with the CH_2 bending scissoring; $2877-2997\text{ cm}^{-1}$, the symmetric and asymmetric CH stretching peaks for CH_2 and CH_3 moieties.
- ii- For methane, the spectral areas selected for the model are: ca. 750 cm^{-1} , corresponding to the CH rocking; $1303-1306\text{ cm}^{-1}$, associated to the CH_2 bending (wagging); $1340-1380\text{ cm}^{-1}$

¹, related to the CH₃ symmetric bending; 1440-1480 cm⁻¹, corresponding to the CH₃ symmetric bending overlapped with the CH₂ bending scissoring; 2925-2985 cm⁻¹, linked to the CH₂ and CH₃ symmetric and asymmetric stretching, whose bands overlap (ca. 2950 cm⁻¹ asymmetric stretching and ca. 2860 cm⁻¹ symmetric stretching). In addition, the model uses some wavenumbers from a broad band centred at ca. 4300 cm⁻¹ which seems a 1st overtone of the CH₂ asymmetric stretching plus CH₂ bending [SM5,SM6].

- iii- For ethane, the model selected the following bands: ca. 750-780 cm⁻¹, corresponding to the CH rocking; 1430-1522 cm⁻¹, associated to the CH₃ symmetric bending overlapped with the CH₂ bending scissoring; 2996-3034 cm⁻¹, CH₃ asymmetric stretching
- iv- For propane, four bands were used: ca. 745-760 cm⁻¹, corresponding to the CH rocking; a wide region ca. 1389-1501 cm⁻¹, related to the CH₃ asymmetric bending overlapped with the CH₂ bending scissoring, and the symmetric bending of CH₃ (ca. 1380 cm⁻¹); 2837-2992, CH₃ and CH₂ symmetric and asymmetric stretching. The model considered also the 3185-3188 cm⁻¹ region, without a clear chemical assignment, although it might be a 2nd overtone of the 1132-1141 cm⁻¹ CH₃ bending (rocking).
- v- For i-butane, the relevant bands were: 2428-2683 cm⁻¹, associated to the 1st overtone of the CH₃ symmetric bending (ca. 1306 cm⁻¹); 2828-2833 cm⁻¹, related to the CH stretching of the C(CH₃)₂ moiety [SM7]; 3730-3797 and 3860-3864 cm⁻¹, which corresponds to the 3rd overtone of the CH symmetric bending [SM8]; 4313-4317 cm⁻¹, which corresponds to the broad band centred at ca. 4300 cm⁻¹, that can be interpreted as the 1st overtone of the CH₂ asymmetric stretching plus CH₂ bending [SM5,SM6].
- vi- For n-butane, the model selected a collection of small ranges in many regions: 1083-1203 cm⁻¹, which corresponds to the CH₃ bending (rocking) overlapped with the C-C stretching [SM7]; 1377-1381 cm⁻¹, associated to the CH₃ symmetric bending; 1473-1478 cm⁻¹, related to CH₃ asymmetric bending overlapped with the CH₂ bending scissoring; 3585-3864 cm⁻¹,

mainly related to the 3rd overtone of the CH symmetric bending [SM8]; 4313-4318 cm⁻¹, which corresponds to the broad band centred at ca. 4300 cm⁻¹, that can be interpreted as the 1st overtone of the CH₂ asymmetric stretching plus CH₂ bending [SM5,SM6].

REFERENCES

- [SM1] Rajalahti T, Arneberg R, Berven FS, Myhr KM, Ulvik RJ, Kvalheim OM. Biomarker discovery in mass spectral profiles by means of selectivity ratio plot. *Chemom Intell Lab Syst* 2009;95:35–48. <https://doi.org/10.1016/j.chemolab.2008.08.004>.
- [SM2] Eriksson L, Byrne T, Johansson E, Trygg J, Vikström C. Multi-and megavariate data analysis basic principles and applications. vol. 1. 3rd ed. Umetrics Academy; 2013.
- [SM3] Cocchi M, Biancolillo A, Marini F. Chemometric methods for classification and feature selection. *Compr. Anal. Chem.*, vol. 82, Elsevier; 2018, p. 265–99.
- [SM4] Cai W, Li Y, Shao X. A variable selection method based on uninformative variable elimination for multivariate calibration of near-infrared spectra. *Chemom Intell Lab Syst* 2008;90:188–94.
- [SM5] Workman, Jr. J, Weyer L. *Practical Guide to Interpretive Near-Infrared Spectroscopy*. 1st ed. CRC Press; 2007.
- [SM6] Socrates G. *Infrared and Raman characteristic group frequencies: tables and charts*. 3rd ed. Chichester, UK: John Wiley & Sons; 2004.
- [SM7] Conley RT. *Infrared spectroscopy*. 2nd ed. Boston, Massachusetts: Allyn and Bacon Inc.; 1972.
- [SM8] Doyennette L, Menard-Bourcin F, Menard J, Boursier C, Camy-Peyret C. Vibrational energy transfer in methane excited to 2v₃ in CH₄-N₂/O₂ Mixtures from Laser-Induced Fluorescence Measurements. *J Phys Chem A* 1998;102.

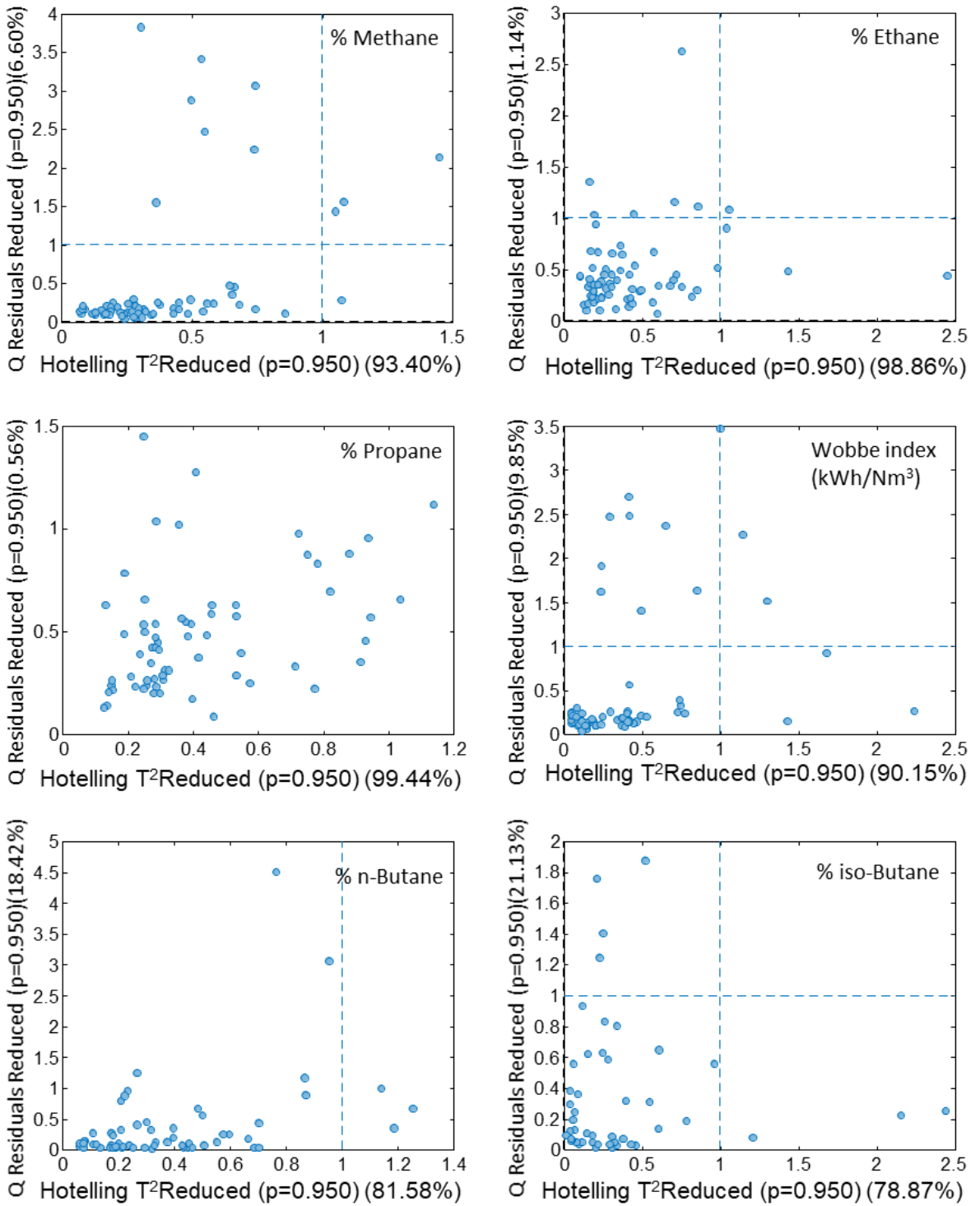
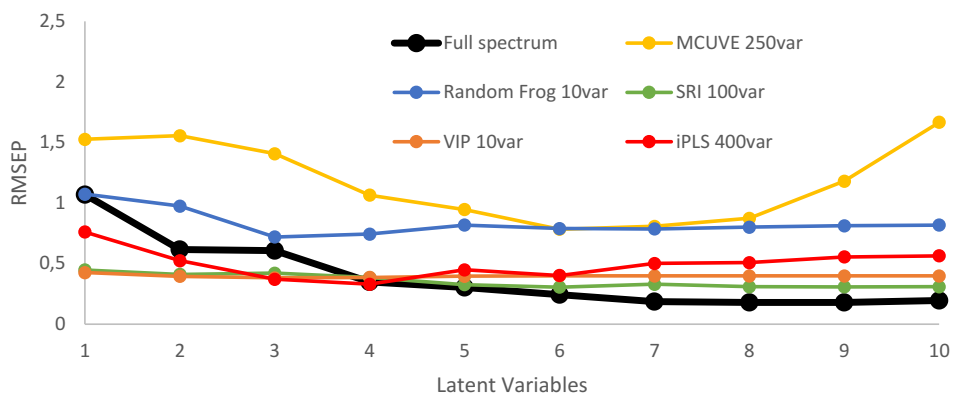
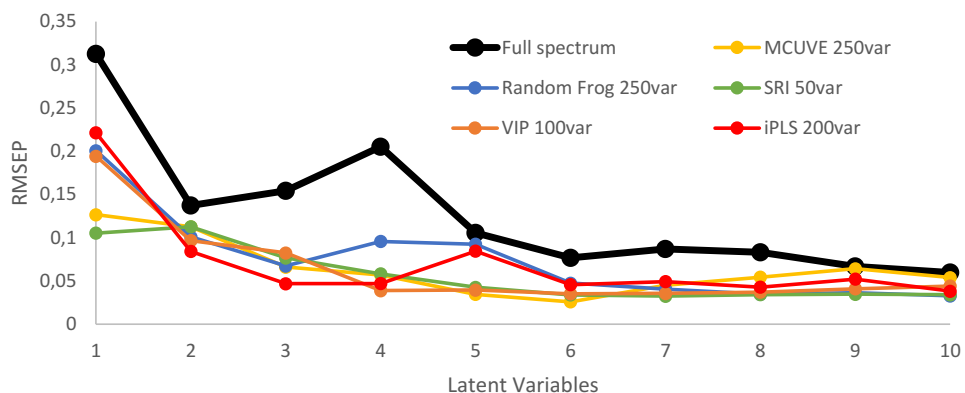


Figure SM 1: Applicability domain for each of the studied parameters (Q vs. Hotelling- T^2 statistics, at 95% confidence level) for the studied parameters.

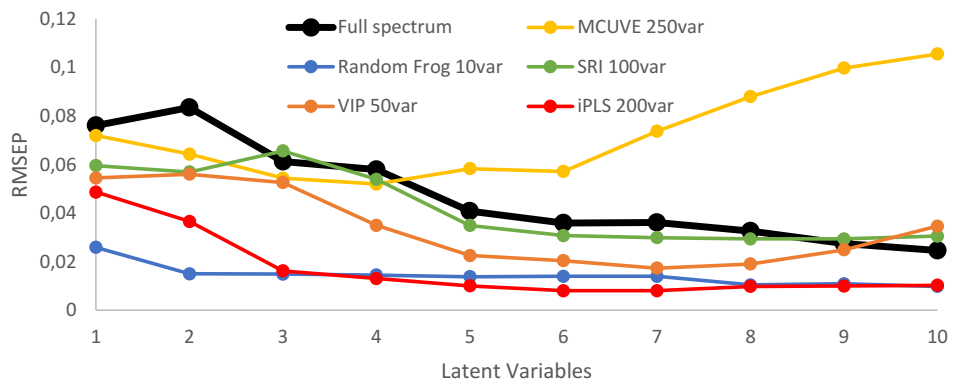
Ethane Variable Reduction Methods



Propane Variable Reduction Methods



iButane Variable Reduction Methods



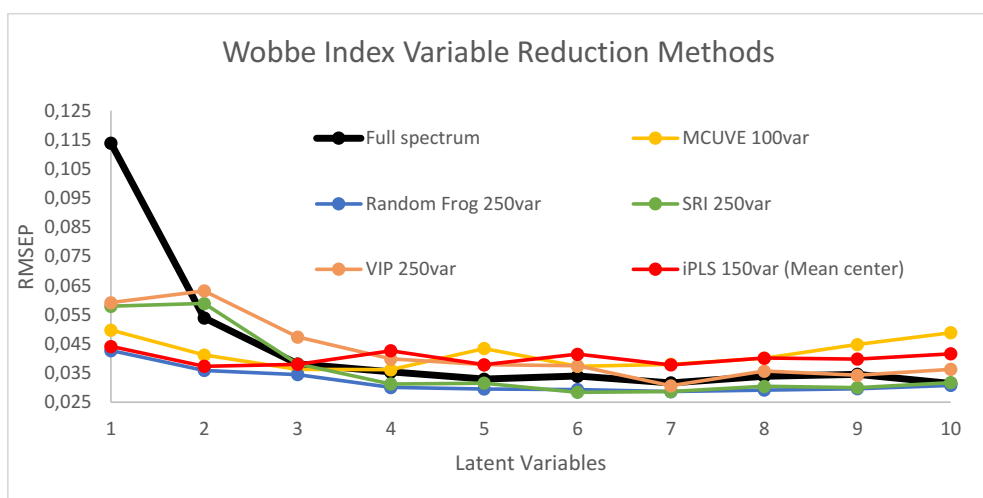
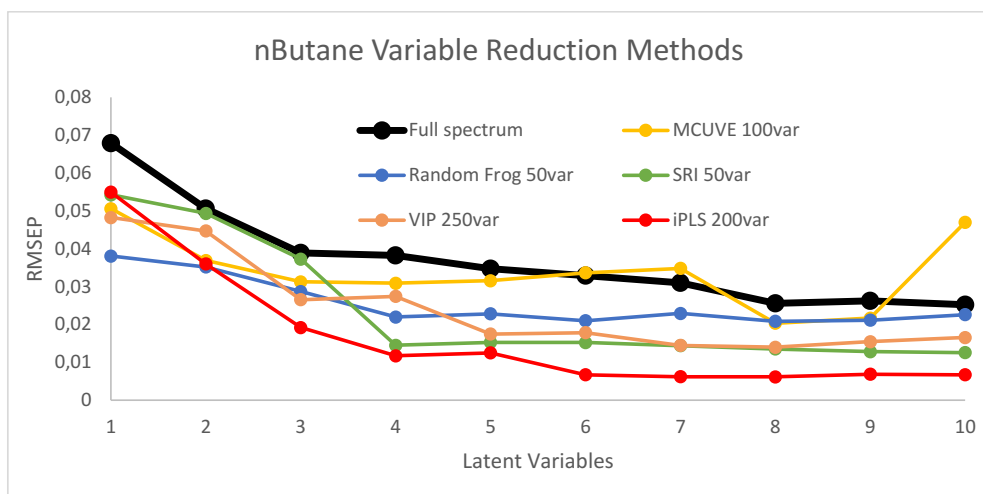
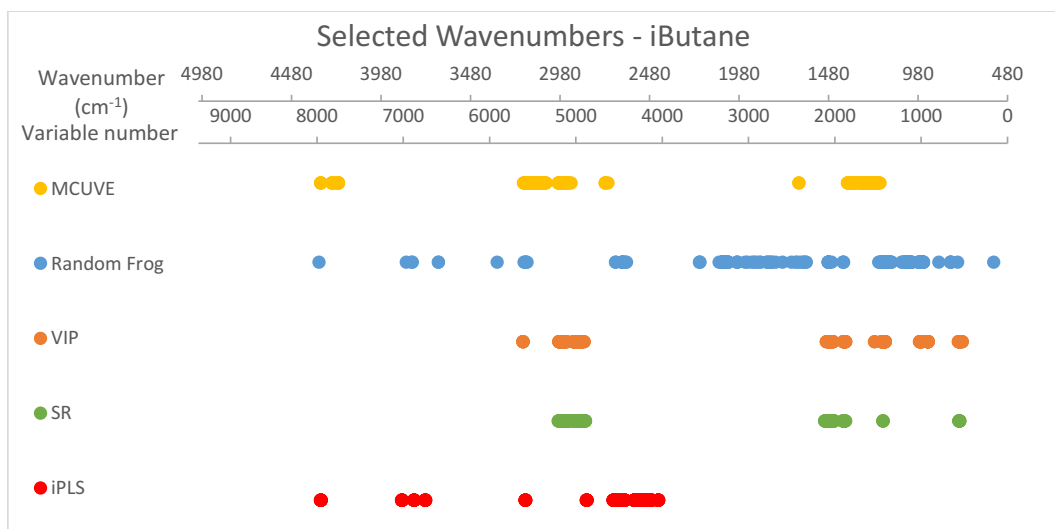
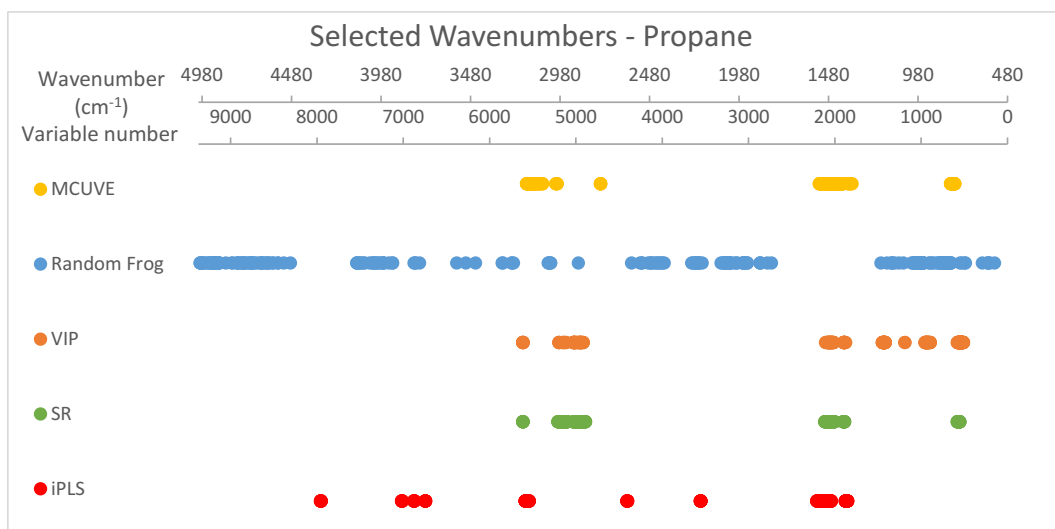
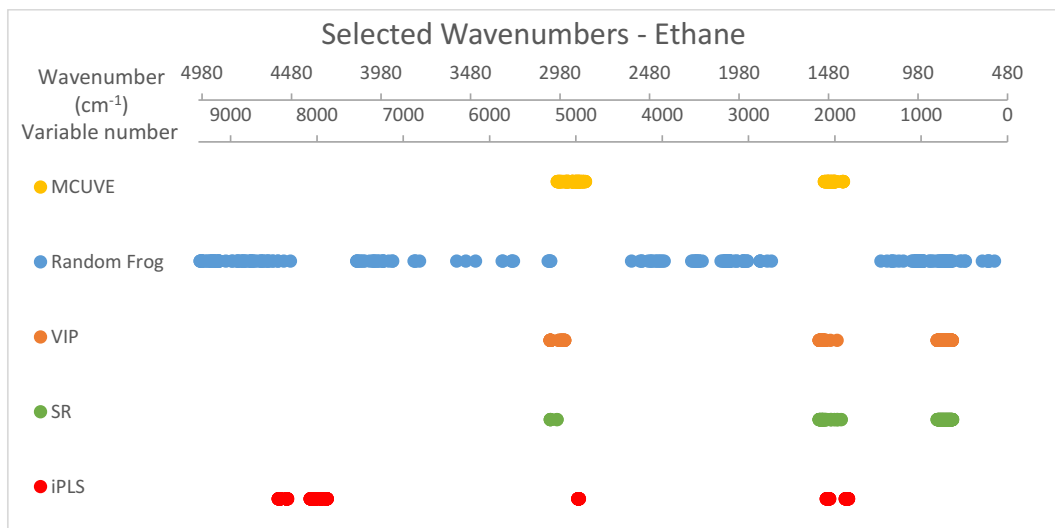


Figure SM2: Errors of prediction (RMSEP, as vol%) as a function of the model selected from each variable reduction strategy and the dimensionality.



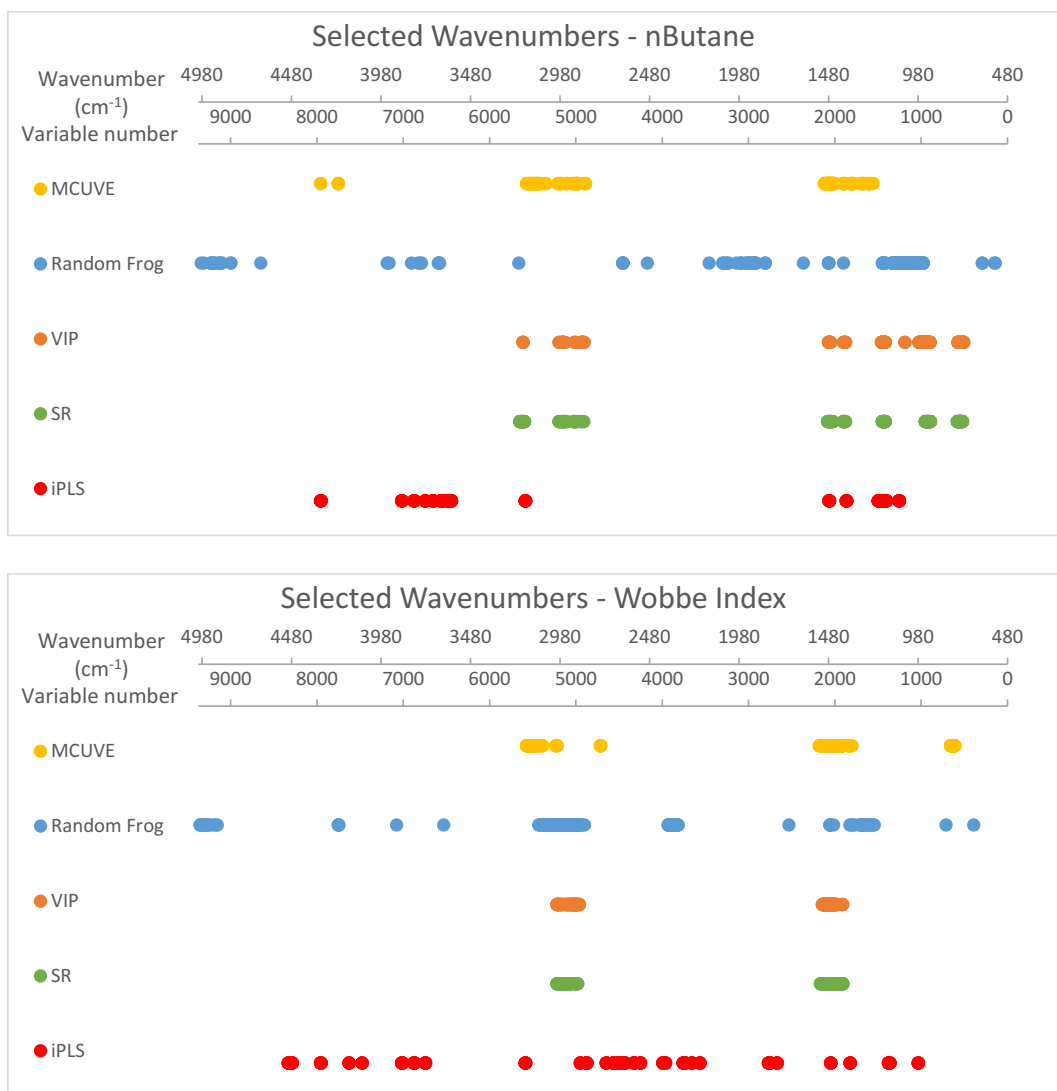


Figure SM3: Distribution of the best 250 variables chosen by each feature selection method.

**Chapter 5: Standardization of
the Minimum Information for
Publication of Infrared-related
Data when Microplastics are
Characterized.**





Review

Standardization of the minimum information for publication of infrared-related data when microplastics are characterized



Jose M. Andrade*, Borja Ferreiro, Purificación López-Mahía, Soledad Muniategui-Lorenzo

Group of Applied Analytical Chemistry, University of A Coruña, Campus da Zapateira s/n, 15071 A Coruña, Spain

ARTICLE INFO

Keywords:

Microplastics
Infrared spectrometry
Standardization
Information
Reproducibility crisis
Quality control

ABSTRACT

Infrared spectrometry (IR) became a workhorse to characterize microplastics (MPs) worldwide. However, reports on the experimental conditions to measure them decreased alarmingly. As complete, relevant information on the instrumental setup determining IR spectra is crucial for scientific reproducibility, ca. 50% of the papers that reported FTIR to measure MPs were evaluated and it was found that most studies cannot be replicated due to missing experimental details. To ameliorate this, the most critical parameters influencing IR spectra are depicted, their impact when matching a spectrum against databases exemplified, and, following efforts from other scientific fields, a minimum information for publication of IR-related data on MPs characterization (MIPIR-MP) is proposed, along with a brief, simple paragraph to resume the most critical information to be reported. This can be used to improve the worrying figures that point out to a reproducibility crisis in the field, as disclosed by the survey.

1. Introduction

The ever increasing use of plastics in most commodities and the elevated time required for them to (bio)degrade (hundreds of years depending of the item) makes them one of the most common residues of mankind nowadays. Between 4.8 and 12.7 million metric tons of plastic enter the marine environment annually (Wang et al., 2016), which is one of the most affected ecosystems. This will continue to happen until drastic legislation is established (like the ban adopted by different countries on one-use plastics (Nielsen et al., 2019)). However even then oceans will remain affected for years because of the plastics slow degradation there (Andrady, 2003). This long residence time of plastics in the marine environment favors their fragmentation, generating microplastics (< 5 mm particles) (Löder and Gerdtz, 2015) which amount to the already spilled primary microplastics (Huppertsberg and Knepper, 2018), like those from certain toothpastes, deodorants, lotions, etc.

Therefore, plastic and microplastic (MP) monitoring needs to be addressed in routine environmental assessments. Regardless of how this is done, validated analytical methodologies can never be given up. However, unfortunately, validation is a difficult issue because at present none of the usual metrological ways to assess trueness (Harris, 2003) is available: certified reference materials (CRMs) have not been fabricated, standardized analytical protocols have not been agreed upon and analytical recoveries or studies on the biota rely on nice

“synthetic” particles (mostly spherical, white/transparent, made of homogeneous and of only some specific plastic types, etc.) (Kühn et al., 2018).

No doubt, the scientific community working at the so mass-media-resonant “microplastics field” is aware of how anxious Society is about reliable and useful information. Scientists cannot deny our responsibility on the final decisions taken by politicians with the data we generate and, unfortunately, the real fact is that data on MPs is not too adequate for decision-making (SAPEA, 2018). Letting aside that undoubtedly more studies need to be done to understand many aspects of this relatively recent problem, there is a growing concern among politicians and scientists on the lack of reliable and comparable data about the amount, types, mass, etc. of MPs in the environment (SAM, 2019; SAPEA, 2018). Any rough estimation in a sample (or set of samples) may lead to conclusions that may differ in tens of millions of tons. The plethora of analytical approaches to measure MPs from visual counting to cutting-edge instrumental systems and the huge variety of sample treatments need strict quality control, validation and standardization. This was at the root of the first JPI-Oceans European microplastics research program, and other ongoing standardization and research efforts in the European Union. If you ever tried unsuccessfully to replicate a published methodology and you thought it was not 100% your fault ... then you experienced a lack of validation problem. Now, think about decision-makers and how big their problems might be (SAPEA, 2018).

* Corresponding author.

E-mail address: andrade@udc.es (J.M. Andrade).

At present, the most widely accepted instrumental analytical technique to characterize plastics and MPs is infrared spectrometry (IR). Many papers support their results on some form of IR measurement and, so, it is imperative to reflect on the minimum information for publication that should be given when reporting scientific data on microplastics based on IR measurements (MIPIR-MP). Hence, the objective of this paper is to review the main instrumental parameters affecting common IR measurements, show how they can affect the characterization of MPs and draw attention on the need for reporting them. The paper is structured in 5 parts. Section 2 reflects on the ultimate reason for gathering scientific data and how important correct reporting is for scientific comparisons and adequate decision-making. Current publishing practices got controversial in many scientific fields, which degraded the quality of the reported information, and some efforts are trying to ameliorate this. In Section 3 a survey on the reported instrumental IR parameters when MPs were analyzed is done and conclusions are gathered from it. As a natural follow-up, Sections 4 and 5 review conceptually the relevance of some instrumental parameters on IR measurements, with some examples. Finally, a paragraph that complies with MIPIR-MP is suggested.

2. Reported scientific data can (should?) be questioned

The term “reproducibility crisis” or “replication crisis” has been coined to summarize a complex phenomenon by which some top journals withdrew a number of papers whose results were impossible to reproduce. Some were discovered to be partially invented and others did not yield conclusive results when repeated by the authors (Berg, 2018). The term(s) appeared by 2012 (Fidler and Wilcox, 2018) and it expanded since then.

Indeed, a seminal survey from Baker –ca. 1500 respondents– was a big “burst” to the scientific community (as a whole) that highlighted the problem of not being able to replicate published studies (Baker, 2016). Recently, in October 2018, Nature collected a series of reflections, papers and discussions in an on-line special issue titled “Challenges in irreproducible research” that comprises discussions from the high-energy physics community to biology (Nature, 2018).

In the analytical field the issue is not unknown unfortunately and some discussions have already been presented. In 2017, Krull focused on the lack of method validation (and lack of statistical data) in analytical chemistry papers –where this problem should not exist! (Krull, 2017). Not long ago, in a nice paper discussing preanalytical errors in the clinic laboratories (where an analytical error can be dramatic), Compton (Compton, 2018) recalled that up to 50% published biomedical data cannot be reproduced due to either a lack of reproducibility or bad (wrong) data. It is clear that non reported experimental data can indeed be a reason for not repeating the findings. An argument developed also by Stark (Stark, 2018) who stated an interesting idea: “science should be show me not trust me”. He also suggested the use of a new word “preproducibility” to denote scientific papers that do describe the analysis properly with sufficient detail, because this is a prerequisite for reproducibility.

“Justifications” to this problem range from fraud to inconclusive results which are interpreted poorly (Kissinger, 2019). But there are several issues which relate to the instrumental measurement side of the problem: instruments that are trusted but not calibrated, the use of inadequately validated methods, little training, important variables which are not accounted for because researchers are unaware of them, lack of care regarding the measurement experimental setup or the blind adoption of black-box methods (Bean, 2019). As many readers may have noticed, the methods sections (including instrumental details) have been shortened in many journals (and their typographical size diminished so that it is even hard to read them!) and this means that “it is often impossible to reproduce a published protocol without some degree of guesswork” (Kissinger, 2019).

The problem is complex and, hence, solutions are not trivial. A way

to reduce it consists of reinforcing the need for an epistemological and ethical metrological approach in every science. Publishers should be more specific in their standards for methods sections and reviewers must critically read those sections and ask for missing information (Bean, 2019). Despite in Analytical Chemistry uncertainty calculations are gaining momentum, this is not so in other fields. Thus, an introductory metrological perspective was presented to biological-related fields (Plant et al., 2018) where the uncertainty terms are not simple to apply following the GUM guide (ISO, 2008).

An exemplifying case study is worth noting: the ever-growing biochemical and genomics field, which received multi-million investments, faced the challenge of reaching a consensus on how to perform and interpret the experiments undergone with a relevant workhorse (q-PCR, quantitative real-time polymerase chain reaction). Then, the idea of requesting a collection of data for minimum information for publication arose (Bustin et al., 2009). Quoting one of their reflections about lack of reliable data “[...] is exacerbated by the lack of information that characterizes most reports of studies that have used this technology, with many publications not providing sufficient experimental detail to permit the reader to critically evaluate the quality of the results presented or to repeat the experiments”. Nowadays, the “Minimum Information about a Biomedical or Biological Investigation” project (Taylor et al., 2008), links ca. 40 “minimum information standards”; specific details can be searched by free at (Sansone et al., 2009).

The prestigious journals Science and Nature adopted new rules. The former set out an editorial principle by which “all data and materials necessary to understand, assess, and extend the conclusions of the manuscript must be available” (Berg, 2018). This is included in a set of “transparency and openness promotion” guidelines (Nosek et al., 2015) and, although they were developed primarily by social scientists, their essentials can be applicable to many other fields. Nature has also set some guidelines for the life sciences, behavioural & social sciences and ecology, evolution & environmental sciences fields (Nature, 2020). In relation to the environmental field, Nature stresses the need to report carefully on experimental and analytical design. The guidelines again point out the need for available and clear supporting information to editors and readers. The main chapters of the reporting summary are sampling strategy (including samples, field collection and transport), data collection, data exclusions, reproducibility studies, statistics, software and code and availability of data. Noteworthy, preprocessing, statistical modelling and inference, and multivariate analyses are also considered.

Despite relevant journals are reacting to reduce the reproducibility crisis, others merely look for profit in a difficult situation. The so-called *predatory journals* flourished after the open-source wave but they do not perform serious peer-reviewing, the quality of the reported information is poor and, even, prejudice the scientific level of developing countries (AlAhmad et al., 2019).

Likely, in our view, the main conclusion gathered from these (and other) studies is that enough data should be presented so that the reader can visualize what it was really done and decide (not figure out or guess) whether a/some relevant variable/s was/were not properly addressed.

The situation in the “MPs field” is not much better, as it will be shown in Section 3 below by means of an empirical survey carried out on 100 published papers that reported infrared spectrometry as their main analytical technique. The results found here are compared to other studies discussing on the quality of MPs results.

3. A survey on the reported IR parameters

3.1. Working methodology

In order to evaluate whether the MPs research field suffers from the reproducibility (or replication) crisis, an extensive search comprising 100 papers was done. To select them, the Web of Science search

algorithms were used (years 2006–2020). First, the key word microplastics was searched for into the title, obtaining 1247 papers (the majority in the last 4 years); they were refined to contain also FTIR (and FT-IR), with only 112 entries. To widen the search both key words were searched for in the theme field, which yielded 241 entries. Most of them (206) were classified in the environmental sciences and ecology field, as expected. The journals showing most publications were: *Mar. Poll. Bull.* (57), *Environ. Poll.* (42), *Sci. Tot. Environ.* (36), *Environ. Sci. Poll. Res.* (28), *Environ. Sci. Technol.* (14), *Chemosphere* (11), *Anal. Methods* (9) and up to 44 papers were published in other journals.

The number of publications including the two key words showed an exponential raising trend. From only one or two publications per year (2006 to 2013), to 36 in 2017, 68 in 2018 and 109 during 2019 (this year has not still been closed). After the overall search, we randomly selected 10 publications of each of the most relevant journals cited above, plus a number of papers from other sources so as to cover ca. 50% of all papers included in the Web of Science results. The overall search with the associated raw data is presented in the Supplementary material.

The 100 papers considered in this study fit the only condition that FTIR had been used to derive results and conclusions. This was assessed by reviewing each paper (sometimes the supplementary materials) and whenever a paper did not fulfil this condition it was not considered (in most occasions that corresponded to comments or quotes from literature).

Further, all the posters of an international conference devoted only to studying MPs were visited to assess whether relevant FTIR information was offered to conferees (Micro2018, 19–23 November 2018, Lanzarote (Canary Islands), Spain).

We definitely agree with Hermesen et al. (Hermesen et al., 2018) in that this kind of studies do not judge the value of the reports. They only reflect the compliance of the papers/presentations with reliability criteria as perceived by the authors. Misjudgements and/or misinterpretations, as well as mistakes during data collection were avoided as far as possible, but they cannot be totally excluded; should that be the case, our most sincere and deepest apologies.

3.2. Results and discussion

Table 1 resumes the overall results of the survey where the basic FTIR instrumental parameters were searched for. A brief comment about FTIR microspectrometry: despite a specific entry is not given into the table, it was seen that almost 100% of the reports mentioned the numerical aperture employed to register the spectra; however, on the

Table 1

General statistics derived from the 100 reviewed papers and 16 poster communications presented at two international conferences to assess the lack of information on the experimental setup used when FTIR spectrometry is used to characterize microplastics.

Parameter reported?	Papers		Conferences	
	Yes	No	Yes	No
Equipment (brand, model, etc.)	97	3	3	6
Working technique (transmission, ATR, reflectance, anvil cell, etc.)	86	14	6	3
Nominal spectral resolution	59	41	2	7
Number of scans per spectrum	66	34	3	6
Background	17	83	2	7
Background per sample (background periodicity)	8	92	1	8
Apodization mode	6	94	1	8
Spectral treatment: ATR correction/Kubelka-Munk/Kramers-Kronig, other	31 ^a	69	1	8
No information at all	2 ^b		7	

^a Including 17 papers reporting transmittance measurements where no dedicated transformations are required.

^b 10 cases reported only brand and model.

contrary, it was rarely specified whether it changed throughout the different particles (and this is a relevant issue), and it was not always clear the measurement mode (transmittance or reflectance), this latter issue was considered under the heading “working technique”.

Positive facts were that 97% of the reports specified the brand and model of the equipment and that 86% of the reports specified the measurement technique employed throughout. The latter figure should nevertheless be increased to avoid the readers to figure out experimental conditions or to spend time looking for what was done (sometimes this is absolutely impossible from the only available data).

Not so nice percentages of good practices were found for reports on the number of scans per spectrum (66%) and nominal resolution (59%). Although these two figures are not totally good, it is strange that these parameters are given in so “many” occasions while others (at least as important as them) are not even reported. This suggests that they were given just by routine and, so, that there might be a lack of understanding of the real importance of the other variables in the final results. Thus, only an incredibly low 17% of the reports specified how the background was made (even, whether it was made at all) and, worst, only 8% of the papers indicated whether the background was repeated periodically and/or per sample. Given the sensitivity of FTIR to the ambient conditions avoiding information on the backgrounds is a relevant bad practice to be avoided (see next section for an example). Despite the apodization is, likely, a variable that exerts less influence on the measurement of solids than on gases or liquids, it has been totally neglected by the authors as only 2% reported on its use and/or the function employed to perform it (in another 4 times this function is not required).

After the IR spectra are obtained, many times some kind of spectral treatment is needed, as detailed in the next sections. Again, we have very bad figures here as only 31% of the papers delivered information; with scarcely 14% ones reporting clearly the spectral treatments (another 4% avoided the problem by using ATR databases). We can sum up here 17 cases where transmittance measurements were explicitly done (N/A in the column headed as ‘transformation’ in Table S1, Supplementary Information) which do not require dedicated transformations, although even in this case baseline treatments might be applied, which in general is not indicated. Absolutely worrying is the fact that out of the 59 papers that used ATR only 7 declared to correct the spectra for radiation penetration (this information is not explicit in Table S1, Supplementary Information). This is of huge concern because many authors claimed that the ATR spectra were compared directly with databases, whose usual spectra are in transmittance (absorbance) units, and they based the assignments on the correlation coefficient. Further, in one case transmittance is said to be used throughout the paper instead of the correct term reflectance, which is a serious mistake and that, unfortunately, was overlooked throughout the publication process. Therefore, serious concerns appear on the correctness of the MPs identifications and the corresponding mass- or counting-balances; and scientists have to reflect seriously on how useful the publications can be for good science and decision-making.

Finally, despite only 2% of the papers have no information on the instrumental setup at all (this is a good figure, per se) the real percentage is much worse as 10% of the reports supplied only the brand and model of the equipment (no experimental details), and another 13% of reports include only the measurement technique. So in total, up to 25% of the papers have no useful information at all about the instrumental setup. This high value (also very cumbersome for the conference, as discussed next) does not seem an oversight, but a bad practice that should be avoided at all cost. To state that FTIR measurements were done with an equipment is not enough, even though spectra are presented.

With regards to the poster communications in a conference devoted to microplastics studies, despite we acknowledge that the overall quantity of FTIR-based posters was very limited (16 in total), the figures can be useful to—at least—envision the situation. Table 1 indicates that

the circumstances are not much better than for the papers and, in some issues, much worse. There was a huge lack of information on basic instrumental parameters so that scientific reproducibility is totally compromised. In total, a shocking 78% of the communications reported no measurement information at all. The less affected parameter was the measurement technique with “only” 19% of the total posters lacking it (or one third of the posters which contain other IR information). The omission of other parameters amounts from two thirds of the posters (e.g., not even the instrument brand and model was mentioned) up to 89% (e.g., background periodicity or ATR apodization are almost not even mentioned). The basic reason for this problem is not space as most of the information can be covered with only two or three sentences, as it will be exemplified in the final section.

The worrying situation depicted by the results above agrees quite well with a handful of very recent surveys of published papers reporting on the quality of MPs studies. In one of them the reliability of the data about the ingestion of microplastics by biota was investigated (Hermesen et al., 2018) and a relevant conclusion was the lack of essential information that made some studies irreproducible. The authors pointed out the need for a standardized way to report information. In other study the authors (Koelmans et al., 2019) reviewed 50 papers reporting plastics on different matrices and proposed some working practices to sample, extract and detect microplastics, and provide some quality assessment. Only 4 out of 50 studies (8%) received positive scores for all proposed quality criteria and, so, they concluded that more high quality data is needed on the occurrence of microplastics in water. Another bibliographical study considered four major stages when quantifying MPs in sediments; namely, sampling, extraction, quantitation and quality control (Hanvey et al., 2017), and found out that only 6 papers out of 43 (ca. 14%) reported laboratory control samples or validation trials, with variable recoveries (higher for larger plastics); and only 3 papers (ca. 7%) reported laboratory blanks.

Some other papers dealt with lack of information on quality control/quality assurance issues. A study provided guidance to improve the reliability and relevance of MPs ecotoxicological studies for regulatory and broader environmental assessments (Connors et al., 2017). They stated that from an objective viewpoint “*much of the existing hazard literature is unusable in a risk assessment framework [...] because of [...] limited methodological details*”. Some reports were closely related to the analytical side of the work, as the analytical characterization of the test particles and analytical verification of test concentrations, for which details of the instrumental setup are indeed a must. Wesch et al. (Wesch et al., 2017) focused on how to avoid aerial microfibre contamination. Finally, Silva et al. (Silva et al., 2018) reviewed different steps of current methodologies to characterize MPs and reflected on major difficulties that nowadays have to be confronted to improve analytical performance when characterizing them.

Lack of reported information on method validation is an issue in its own, as reported by Hanvey et al. (Hanvey et al., 2017), who pointed out the need for both extensive validation trials and laboratory control samples. In particular, they stressed that, as a general rule, inter-laboratory proficiency testing should be undergone. We fully agree with this idea, as it is common practice in Analytical Chemistry. This approach is not so expensive and, on the contrary, is highly rewarding in terms of reported information and knowledge management. Inter-laboratory exercises are envisaged as a must, especially as far as we still do not have certified reference materials nor standardized analytical methods for MPs.

Those issues support the opinion of several experts that analytical approaches applied to measuring plastic debris in environmental samples are underdeveloped (GESAMP, 2015). Somehow this could be due to the huge drawback of the (very) heterogeneous distribution of MPs in the samples and, of course, in natural systems (GESAMP, 2015). However, all the reports cited above concluded that analytical protocols used commonly to characterize MPs are far from being able to deal with the EU statements (Silva et al., 2018) to set the performance

characteristics (European Commission, 2002) of many analytical methods.

4. IR spectrometry in a nutshell

The results above make it necessary a brief, conceptual resume on the basics behind the FTIR technique (specific details can be found in any classical introductory textbook; e.g., (Smith, 2011)) as they are crucial to understand why some parameters have to be optimized/reported. The following section reviews typical parameters that affect the quality of IR spectra and shows that when overlooked, they can prejudice the visualization of weak spectral bands/peaks and affect the interpretation of a spectrum and/or its match with spectral databases. Therefore, it should be mandatory to report them.

IR spectrometry (denoted as such because nowadays benchtop dispersive instruments are not used so no confusion is possible) has become an established and powerful analytical workhorse in many qualitative and quantitative applications, both at industry and research (Doyle, 1992). It is a ubiquitous tool for the study and quality control of industrial materials (ingredients, primary and secondary products, residues, etc.); food, forensics, environmental and pharmaceutical analyses; organic molecules (Chalmers and Griffiths, 2002; Lin-Vien et al., 1991); and thermal imaging (Vollmer and Möllmann, 2017). It has also been used to develop sensing techniques in atmospheric composition analysis (Laj et al., 2009); surface analysis (Belfer et al., 2000); biological and medical applications (Krafft and Sergo, 2006).

Indeed, IR spectrometry is one of the best options to characterize the structure and degradation of polymers and MPs. At present IR and Raman spectrometry are the most used techniques to identify MPs (Mukherjee and Gowen, 2015). Some of its advantageous characteristics are its high turnaround time, selectivity and its relatively straightforward combination with microscopes (Renner et al., 2017).

Unfortunately, when a scientific field develops as rapidly as MPs analysis did, and an analytical technique becomes a workhorse to get relevant information there is a risk that the basic and/or fundamental parameters become overlooked or partially neglected. This may have detrimental consequences not only for the experimental work at hand, but for others willing to carry out a similar study.

IR spectrometry is based on the interaction of an infrared light beam with the sample molecules by modifying their rotational and vibrational energy levels when the Planck's equation is satisfied. When multifrequency radiation is used, each bond of each molecule in the sample can attenuate the beam (as far as its dipole changes when the rotational or vibrational mode changes, making nonpolar molecules transparent to IR radiation). As a consequence, when the recorded beam is processed it contains information about each and every molecule in the sample. Accordingly, be cautious when interpreting the spectrum, unless it is known that only one type of molecules is present in the sample.

The core of any IR spectrometer is the interferometer, where a continuously moving mirror is the key. Its purpose is to create wave interferences, most of them destructive, so that the polychromatic radiation gets modulated and the total power reaching the detector can be registered as a function of the displacement (distance or time) of the moving mirror. Each time this mirror moves back and forth the instrument registers a full interferogram, or *scan*, which is a very complex power fringe, not interpretable visually (hence the need for the computerized Fourier transform, FT). The interferograms are averaged to increase the signal-to-noise ratio (S/N). The difference between the interferograms obtained before and after the sample is placed at the IR radiation path yields the net interferogram. Next, the FT algorithm leads to the net spectrum of the sample, expressed in classical units. A misunderstanding appearing in several reports has to be clarified: classical units to set the “x-axis” of the spectra are cm^{-1} and this corresponds to wavenumbers, $\tilde{\nu}$, not wavelengths, λ .

Different IR sources yield different light intensities and different detectors have different efficiencies, so it is worth specifying them

when reporting results (although most times this is implicit when specifying the brand and model).

5. Common basic parameters for IR measurements

To organize this section as simply as possible, the parameters required to inform on the IR experimental set up employed in a MPs study can be classified in two types: instrumental- and software-related (spectral treatments).

5.1. Instrumental

5.1.1. Wavenumber range

The frequency range or, more commonly, the wavenumber range defines the minimum and maximum frequency points used in the analysis. Its selection is of utmost importance for the measurements as each analyte has a range in which the characteristic peaks caused by its bonds and functional groups appear. The medium range, Mid-IR (or simply, IR), 4000–400 cm^{-1} is most commonly considered because the near region (NIR, 4000–14,000 cm^{-1}) yields very broad, overlapping bands not easy to interpret and differentiate. The far or terahertz IR region (FAR, 400–10 cm^{-1}) has not still been applied to determine MPs as instruments have been developed recently. Declaring the working range in the reports is necessary and almost always done.

5.1.2. Number of scans

The number of scans (NoS) is not considered in many works, as it should be. It is the number of times the interferogram (scan) is repeated to get a spectrum. Currently the software averages all the interferograms and then applies FFT (Fast Fourier Transform) to yield the spectrum. As noise is random, increasing the NoS per spectrum improves the S/N ratio (by a factor of $\sqrt{\text{NoS}}$ (Smith, 2011)), although at the expense of correspondingly increasing the time needed for the analysis. From a practical viewpoint, the analyst should decide the NoS

after a preliminary study where the S/N ratio is plotted against the NoS for a typical sample and setup configuration. Fig. 1 shows an example where a polystyrene pellet was measured by ATR. Of course, the more scans we register the better the signal will be, so a trade-off between S/N ratio and time has to be sought for. In the figure, a clear stabilization of the S/N ratio is seen after at least 30 scans, further improvements are not highly relevant and they consume time (e.g., 200 scans in this particular configuration took around 20 min per sample, while 30 scans required only 5 min). Note also that the S/N ratio changes for different absorption peaks so it is worth testing several, or the one of interest.

5.1.3. Background

Every spectroscopic measurement is affected by many contributions besides that from the sample. Ambient conditions like temperature, humidity, CO_2 and other vapors present in the laboratory can largely affect the IR spectrum. To correct them it is necessary to record them (as a background) and subtract them from the gross spectrum of the sample (this final step is done automatically by the controlling software). In general, when the measurements take time (e.g., due to a high NoS) or whenever a series of samples have to be measured it is not a good practice to record only one background for all the samples (for it only represents the initial conditions in the laboratory) but, instead, it is recommended to make several backgrounds through the working session; ideally, a background per sample. When the analyst has to stay near the instrument during the measurements, as is often the case, it is mandatory to make a background before each sample to ensure accurate results.

Fig. 2a shows the spectrum of air recorded over a reduced time frame (< 1 h) in a small room without the presence of working staff. The baseline and the peaks associated to H_2O and CO_2 (vapors) change substantially, as it is usual. Unless this is corrected for, the baselines of the spectra registered will show a sloppy trend with time that will difficult the identification of the item under study with the spectra of the databases. Recall again that the reflectance measurements may

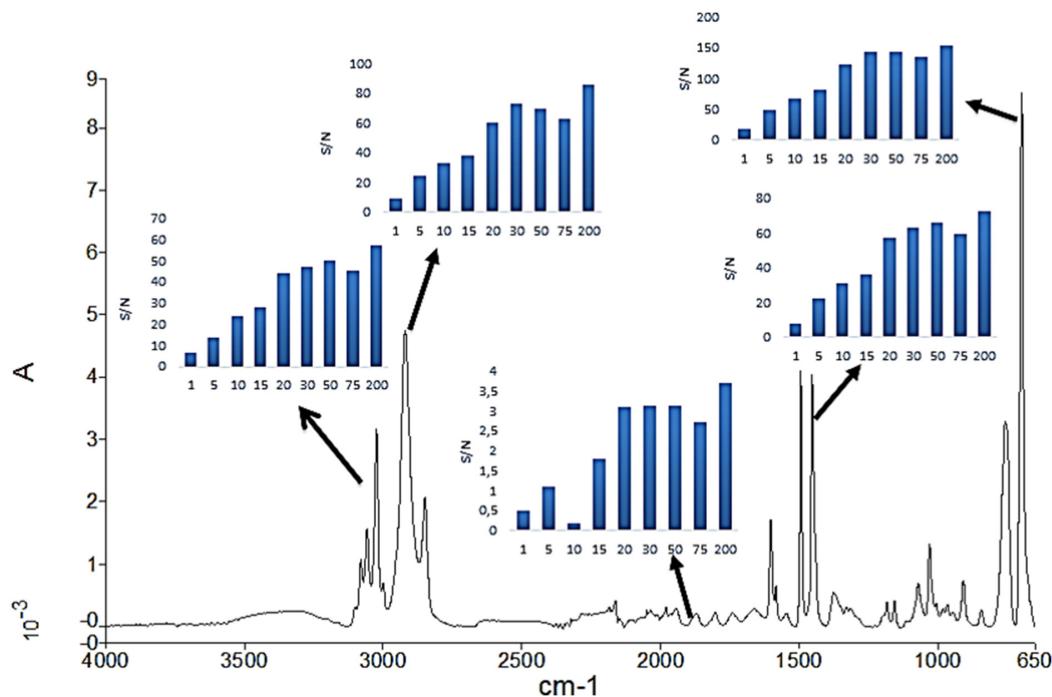


Fig. 1. Measurement of a polystyrene pellet by ATR and evaluation of the number of scans to register a spectrum (see text for details).

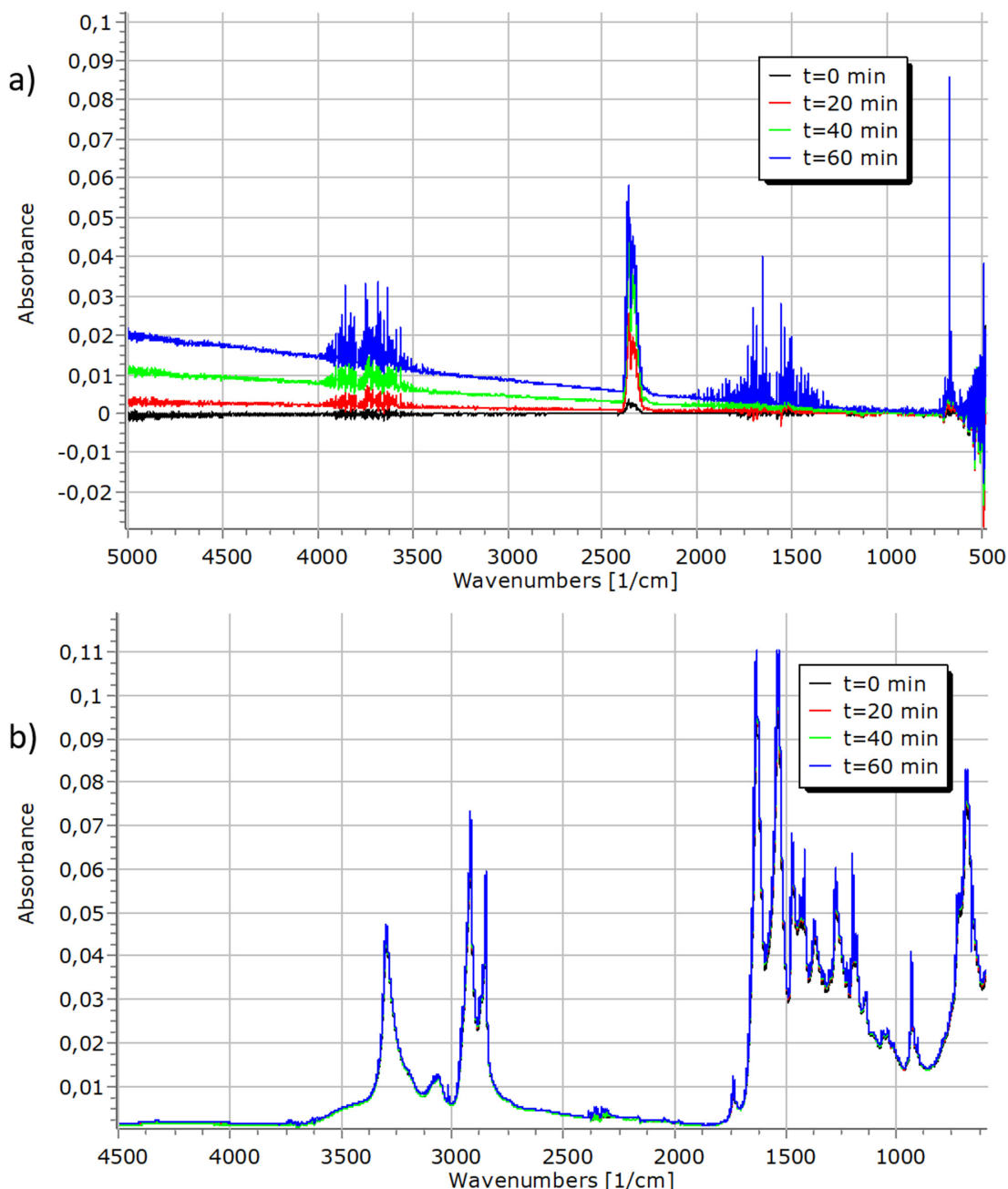


Fig. 2. (a) Typical sloppy trend of the IR baseline when the background is not corrected for during a working period of 1 h; (b) Spectra of a PA6.6 pellet registered each 20 min, background uncorrected, whose match against the database degraded from 0.9109 ($t = 0$), to 0.9084 ($t = 20$) and 0.9007 ($t = 60$), see text for details.

require different backgrounds (see next sections).

Fig. 2b shows the spectra of a PA6.6 pellet measured throughout 1 h (at 20 min intervals) and whose comparison against a commercial database degraded with time (from 0.9109, original measurement, to 0.9007 at 60 min) because of a lack of background correction (recall that no analyst was breathing in the room, so this is an optimistic result).

5.1.4. Resolution

Resolution defines the frequency with which the spectrometer takes an experimental readout. The higher the resolution is, the lower the interval between successive experimentally-registered data points becomes (Smith, 2011). This generates spectra with more information, but it also increases spectral complexity, decreases the speed of the analysis and reduces the S/N ratio. It is worth noting that “high resolution” means setting lower figures. For instance, 1 cm^{-1} denotes an 8-fold

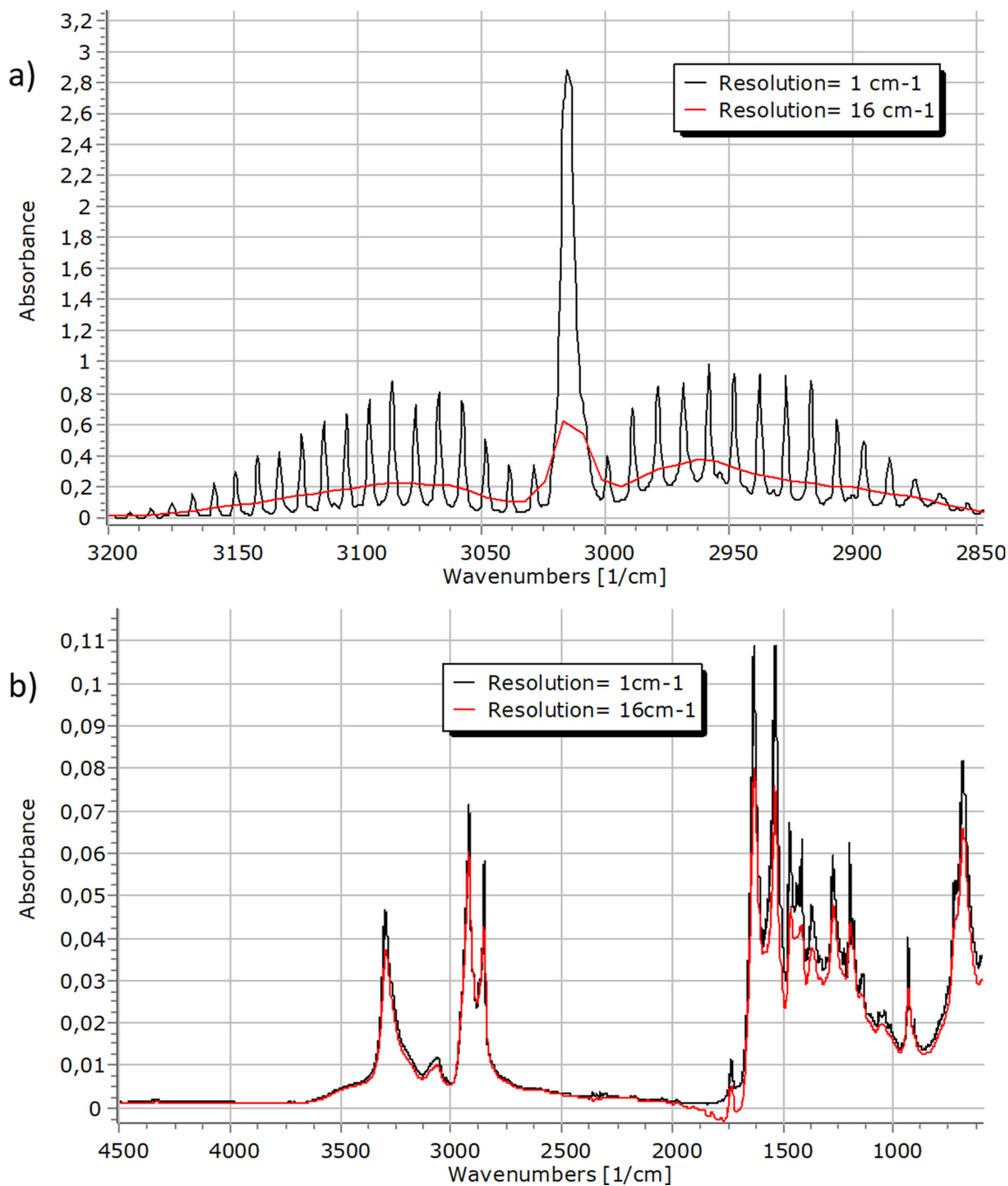


Fig. 3. (a) Spectra of a gaseous sample recorded at nominal resolutions of 1 cm⁻¹ (black) and 16 cm⁻¹ (red) denoting clear differences in the less intense bands; (b) resolution affected the identification of a PA6.6 pellet whose spectrum was registered at 2 (r = 0.9190) and 16 cm⁻¹ (r = 0.8707). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

better resolution than 8 cm⁻¹. As a consequence, the selection of the resolution will determine the shape of a spectrum. From a pragmatic viewpoint, resolution can be defined as the ability of an instrument to differentiate two different spectral features that are close together (Harris, 2003). This is an especially important parameter when spectral peaks of interest are close and there is a need to differentiate them. No doubt, its relevance on the study of gaseous samples is paramount, as their spectrum is usually composed of a multitude of very narrow peaks.

As an example Fig. 3a shows the difference between the spectra of the same gaseous sample using two different resolutions. A preliminary evaluation is required to select the best resolution for a particular study. Common nominal resolutions for liquid and solid samples are 4 and 8 cm⁻¹ in order not to miss the sharpest spectral peaks. Note that the lower the resolution is (e.g., 32 cm⁻¹), the higher the risk of missing sharp peaks becomes. Fig. 3b shows the spectra of a PA6.6 pellet measured at two nominal resolutions and whose identification against

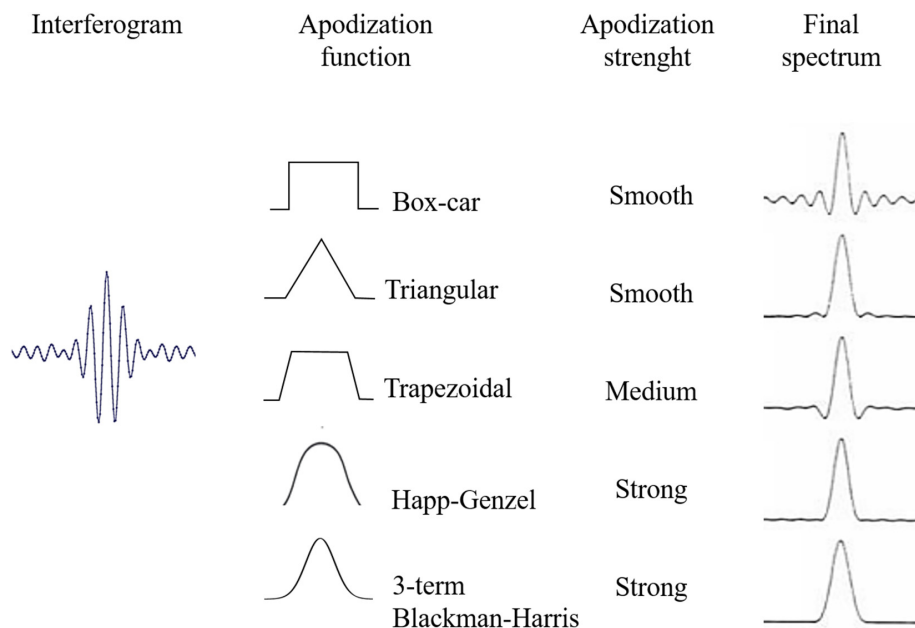


Fig. 4. Graphical concept of different apodization functions and their effects on an IR peak obtained after FT transformation of the interferogram. (Adapted from (Bruker Optics, 2020).

commercial database became degraded when the resolution increased too much. Unfortunately, the resolutions employed in the databases are unknown so we have to consider that they used the most common ones (typically, from 2 to 8 cm^{-1}), but this remains a bit obscure. Hence, it is relevant to indicate the experimental nominal working resolution.

5.1.5. Apodization

Strictly speaking, the computation of the FFT involves a complex integral whose limits are $\pm \infty$. However the mobile mirror can move only a finite distance ($\pm d$), thus the limits of the integral are restricted to $\pm d$. This prejudices the computations and alters the resulting spectrum (even when modern algorithms to compute the Fourier transform are used) by generating secondary peaks, or lobes, adjacent to the main spectral peak. Those side peaks difficult the interpretation of the spectrum, as they can overlap with the true signals, broadening them or modifying their shape. To correct this, a so-called “apodization” function is used. It consists of a mathematical function applied after the FFT one to remove the sides of the peaks in the spectrum. There are numerous apodization functions, some of them proprietary of the instrumental companies. In general, the most aggressive functions are used for gas analysis. Fig. 4 presents some typical apodization functions and their effect on a spectral peak. The user can change the apodization function in the controlling software and, so, it is important to report the function used in a study and its related parameters, when they can be modified.

5.1.6. Measuring technique

The enormous versatility of infrared spectrometry led to a wealth of different IR-based working techniques, with different experimental setups and parameters to fix. As a consequence, it is of paramount importance to report all the variables involved in the measurement at hand. For the sake of simplicity, the major techniques are:

5.1.6.1. Transmission. No doubt this is the most common approach to measure gases, pure liquids, solutions and flat, thin, transparent solids. It corresponds to the most basic configuration of any IR spectrometer

and it records the total energy exiting the sample layer. To make quantitative measurements the pathlength should be fixed although normalization by the thickness of the absorbing medium can be done (if this is unknown, it can be calculated easily from the interference fringes of the spectra (Smith, 2011)).

5.1.6.2. Reflectance. In order for too thick, too absorbing or opaque samples be studied surface reflectance methods need to be used, such as Attenuated Total Reflectance (ATR), Diffuse Reflectance Infrared Transform Spectrometry (DRIFTS) or photoacoustic spectrometry (PAS). All three require additional accessories, increasing the cost of the instrument.

ATR is the general choice when measuring MPs and, following, we will consider here some more details than for the other options. The sample is put in close contact with an internal reflection element (IRE), usually a crystal of a material with a high refractive index (like silicon, germanium, zinc selenide or diamond). Depending on the experimental setup the IR beam can be reflected back multiple times (commonly, 1, 5 or 10 reflections) at the crystal-sample interface. Each time it impacts on the sample, part of the electromagnetic energy of the beam (the so-called *evanescent wave*) penetrates into it. The interaction between the wave and the sample allows us to obtain a quite good IR spectrum in otherwise unpractical conditions. The depth of penetration (d_p) of the wave depends of the specific sample and the wavelength under consideration, and it can be calculated easily, Eq. (1) (Harris, 2003). The pathlength is the number of reflections times d_p .

$$d_p = \frac{\lambda}{2\pi n_c \sqrt{\sin^2 \theta - (n_s/n_c)^2}} \quad (1)$$

There, λ is the wavelength, in mm; n_c is the refractive index of the crystal (for diamonds is ca. 2.4 at 1000 cm^{-1}); n_s is the refractive index of the sample (for common transparent plastics –PC, PS, PMMA, etc.– it is ca. 1.5) and θ is the incidence angle of the radiation on the crystal. An excellent compilation of refractive indexes of polymers can be found at (Polyanskiy, 2020).

The final spectrum depends on the experimental setup and,

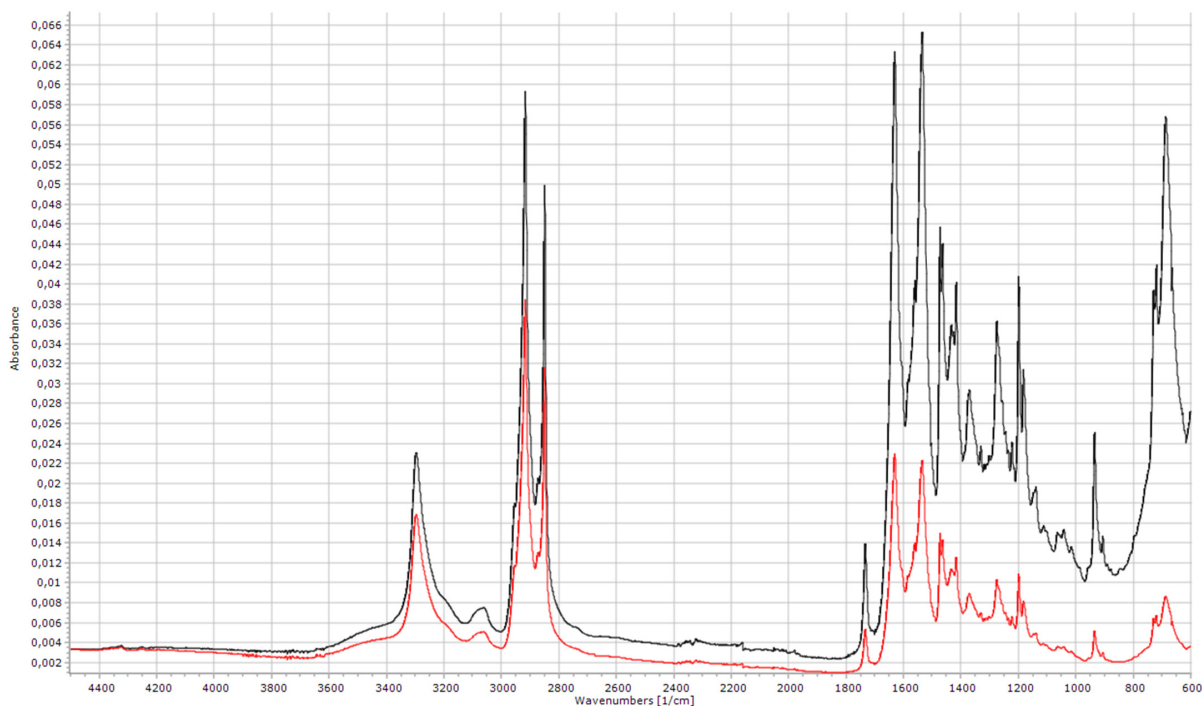


Fig. 5. Example of the influence of the ATR correction on the appearance of a spectrum from a PA6.6 pellet. Correction by depth penetration yielded a 0.9900 match with an in-house database whereas no correction reduced the match to 0.9421.

therefore it is important to report the nature of the crystal and the number of reflections of the beam in the sample, along with some value for the penetration depth in our sample. The reflectance spectrum thus obtained suffers the problem that the intensity of the recorded signal varies with the frequency. Longer wavelengths (lower wavenumbers) penetrate more than shorter wavelengths (higher wavenumbers), see Fig. 5. Further, as the refractive index of a sample changes sharply close to the absorption peaks, the peaks can be shifted (Smith, 2011) regarding their positions in absorbance spectra. As the reflectance spectrum is not directly comparable to a typical absorbance one, it is common practice to modify the ATR spectrum so that it looks like an absorbance spectrum and it can be compared to typical databases. Although different commercial brands have slightly different proprietary algorithms, in general the correction modifies each original ATR reflectance by a factor inversely proportional to the wavelength and, eventually, a penalty parameter. Such a penalty stems from the fact that for this technique to work effectively the crystal and the sample must be perfectly in contact with each other so a piston is usually needed. For liquid samples the penalty parameter is set to zero (perfect contact between sample and crystal), but if improper contact is suspected (like in the case of plastic pellets or fragments), it might be adequate to study different penalty values (and report the selection).

Therefore, report always whether the ATR spectrum was corrected and the penalty parameter (even if it is zero). Note that the identification of a polymer against a database can be affected seriously if the ATR correction is not considered (see Fig. 5 for an example).

For solids with a rough surface, fibers or powders it is more appropriate to use DRIFTS. Here, the light that falls upon the sample is reflected back in multiple directions, which is called diffuse reflectance. The DRIFTS accessory is designed to redirect all the light reflected from the sample to the detector. The diffuse reflectance spectra may potentiate certain peaks in comparison with a usual absorption/transmission spectra. To correct for this effect the Kubelka-Munk equation

can be used (Stuart, 2015). The background here can be made either with an aluminum-coated silicon abrasive pad or with KBr or KCl powder.

If the surface of the sample is flat and smooth, the reflectance of light is more like than that of a polished surface and it is said to be “specular” reflectance. If the thickness of the sample allows the IR light to traverse all the sample layer and reflect back on the solid holder and exit the sample, transmittance (when the films are rather thick) or reflection-absorption (when the films are thin) are more suited terms. This can happen when flat films of polymers are deposited over reflective (e.g., gold) surfaces. Band intensity ratios and band positions can be severely affected depending on the thickness of the sample, although these effects are generally overlooked (see (Mayerhöfer et al., 2018) for a more profound and technical discussion). Note that reflectance measurements need a proper background, which is generally made using a reflective, flat, golden polished surface.

5.1.6.3. Microspectrometry. This technique is applied whenever microscopic samples are to be analyzed. It consists of a FTIR device whose modulated beam is diverted to an adjacent microscope setup, adapted for the IR light. Micrometers-sized samples can be analyzed and chemically mapped. The microscope is situated over the sampling plate, on a mobile stage. IR radiation is focused onto the sample using a combination of convex and concave mirrors (Cassegrains) that redirect the image to the barrel of the microscope (Larkin, 2017; Perkin Elmer, 2015). These instruments generally allow for transmission, reflectance and ATR measurements so clarifying which mode was used to perform the MPs characterization is a must in any report. If ATR is used in a microscope (also termed micro-ATR), report how, and how often, cleaning of the ATR tip is done.

The quality of the spectra and the spectral resolution attainable with an IR microscope relies strongly on the amount of energy received by the sample, which in turns depends on the size of the area of the sample

illuminated by the IR beam. This is controlled by a mechanical system that has to be modified by the analyst: the aperture. Using a traditional MCT detector, the spatial resolution is related to the infrared beam aperture dimension which cannot be lower than the theoretical diffraction limit of about 10 μm , although for some systems this could be up to 20 μm (Kanno, 2008). However, the S/N ratio at this limit will be quite poor and so at least a 20 \times 20 μm aperture is needed to obtain enough energy (Bhargava and Levin, 2007; Prati et al., 2010). The optimal aperture depends also on the detector type under use; between 10 and 100 μm for MCT detectors, between 10 and 200 μm for InGaAs detectors and between 50 and 200 μm for DTGS detectors (Perkin Elmer, 2015). Accordingly, reporting on the aperture (or apertures) used to study the particle(s) and the detector employed therein is needed.

5.1.7. IR hyphenation

A relevant, general advantage of FTIR spectrometers is their relatively simple combination (hyphenation) with dedicated accessories or other analytical techniques, such as chromatography or thermal analysis. Both GC-FTIR and TGA-FTIR require the spectrometer to continually scan the gas cell and to combine/integrate the band(s) during some time. For quantitative purposes the Gram-Schmidt orthogonalization algorithms are really needed (Namba, 1990; Robertson and Lawson-wood, 2017; Sparks et al., 1982).

The GC-IR hyphenation leads to evolving gas analysis as the gases eluting from the chromatographic column are introduced in a gas flow cell of the IR spectrometer to identify them, yielding information on the components that exit the column (Villberg et al., 1997) or volatiles entrained in polymers (Schmidt et al., 1988).

The main purpose of the hyphenation between IR spectrometry and thermal analysis is to obtain information on the chemical processes that take place on a sample when subdued to a temperature variation. The TGA instrument records the loss of mass as a function of time or temperature while IR spectrometry generates information about the evolved gases of the combustion/pyrolysis. Physical properties, chemical changes and degradation pathways of polymers and copolymers can be studied (Lin, 2016; Loría-Bastarrachea et al., 2011); for instance, to characterize PVC and PS microplastics (Yu et al., 2019).

5.2. Spectral treatment parameters

In addition to the parameters above, spectra are usually treated to improve their quality and/or avoid undesired artifacts. Some typical treatments are resumed below (note that they are not the unique possibilities, although they are probably the most frequent ones):

5.2.1. Baseline correction

Currently, the general shape of the IR baseline (which ideally should be flat) can be modified by the dispersion of the IR beam, changes in the atmosphere of the laboratory, temperature variations, instrumental drifts, the sample itself (scattering powder, inhomogeneous composition, etc.) or other unwanted effects. Hence, it is necessary to suppress unwanted slopes and curvatures, while maintaining the band intensities. The process of baseline correction varies greatly between different software and in some cases it can be corrected manually, selecting the points of the spectrum that will be bound to zero (be careful and set always the same baseline correction points). Both the correction points (number and position) and the correction function (linear, polynomial, etc.) are relevant information that should be reported. Whenever the baseline is not corrected for its effect on the identification of eventual polymers using databases will be similar to that of not correcting for the background.

5.2.2. Smoothing

This correction is applied to spectra with low S/N ratios so that noise is reduced, making it easier to discern the different peaks while

creating more visually appealing plots. There are different smoothing functions, such as Box-Car, triangle, Savitsky-Golay (this is the most common option), etc. Be aware that while smoothing can help differentiating a real signal from noise, the intensities of relevant spectral features for your problem may be reduced (smoothed) as well because smoothing decreases the intensity of the overall spectrum and, in extreme cases, the signal of interest can be completely eliminated. Also, neighbor weak peaks or bands can merge and, so, the resolution of the spectrum reduces. Preliminary trials are needed to fix the smoothing parameters, where the key point is not to affect the intensity and shape of the peaks of interest. Changing the parameters involved in the smoothing function can largely affect the spectrum and so, they must be reported (e.g. the selected function and the window width).

5.2.3. Kramers-Kronig

When using a specular reflectance accessory, spectra containing regions with derivative shapes (called *reststrahlen*) (Smith, 2011) are obtained frequently. This is so because the refractive index of the sample depends of the wavelength. To correct for this the so-called Kramers-Kronig equations are applied to transform a specular reflectance spectrum in an absorbance-like spectrum. This is useful because most IR databases contain transmission/absorbance spectra. This transformation is only advised on samples with a smooth and shiny surface. The samples with rough surface can have a reflectance that is partly specular and part diffuse, which can distort the resulting spectra (termed *k*). The set of equations are rather complex and exceeds the scope of this tutorial; more interested readers can consult (Warwick and Eda, 2014) and the basic ideas depicted nicely at (Shimadzu, 2008), which are exemplified with two spectra of black rubber and a plastic slab.

5.2.4. Kubelka Munk

This treatment is applied when diffuse reflectance occurs primarily and/or when using a DRIFTS accessory, generally on spectra with low absorption, so that the transformed spectrum obeys the Bouguer-Lambert-Beer's law. Spectra that undergo this transformation (see Eq. (2)) are expressed in Kubelka-Munk (K-M) units, where *KM* is the spectrum in K-M units and *R* is the absolute reflectance of the sample (Smith, 2011).

$$KM = \frac{(1 - R)^2}{2R} \quad (2)$$

This equation can also be expressed as

$$KM = \frac{2.303 \cdot a \cdot c}{s} \quad (3)$$

where *s* is the scattering factor, *a* is the absorptivity and *c* the concentration of the sample. In this equation the only unknown variable is the scattering factor, which depends on the distribution of the sample (size, shape, packing, etc.) (Smith, 2011). As the IR beam penetrates between 1 and 10 μm in the sample, this equation assumes the thickness of the sample to be infinite.

5.2.5. Deconvolution

It is a function used to enhance the resolution of the spectra by allowing the differentiation of overlapping peaks mathematically when it is not possible on the original spectra. It uses the Fourier transform to diminish the broadening of the band. While it can enhance the resolution of a band or peak, it can never be better than the spectral resolution used to obtain the data. Deconvolution comprises 3 steps (Smith, 2011): The first consists of selecting the band to be modified. Then, a Fourier transform is applied, generating a partial interferogram called *cepstrum*. This interferogram is altered to increase the intensity of its "wings". This way, when the Fourier transform is applied again the peaks constituting the overlapped band are narrowed and, hopefully, differentiated. The main problems this treatment can produce are due

to overdeconvolution, by increasing the resolution so much that it can generate secondary lateral peaks as well as increment noise, baseline fluctuations and creation of other non-existing spectral features. Deconvolution is a relevant issue when defining functional group indexes unambiguously; e.g., to study PA6.6 degradation (Vasanthan and Salem, 2000), interactions between components in polyester/poly-styrene blends (Gordon et al., 2015), to quantify the crystalline phase in poly(vinyl)alcohol (Tretinnikov and Zagorskaya, 2012), study cellulose crystallinity after enzymatic digestion (Kruer-Zerhusen et al., 2018) or evaluate mixtures of plastics (Fuller and Gautam, 2016; Miller et al., 2017).

6. Conclusions

This paper demonstrated the huge importance of adequately detailing the minimum experimental conditions required to replicate the IR studies undergone to seek for the occurrence of microplastics in environmental matrices. A survey was made covering approximately 50% of the papers where microplastics were assessed by FTIR spectroscopy and a very important lack of information regarding the experimental setup was found. In addition, the results gathered from the survey suggest that some microplastics identifications might be compromised, thus affecting mass- or counting-balances. For instance, up to 25% of the published papers do not reported relevant instrumental operational details, and only 12% of the papers that used ATR reported that the difference on the depth of penetration of the IR radiation was corrected for, which is a serious conceptual pitfall. Similarly, the typical IR background drift has been largely overlooked as scarcely 17% of the papers reported on it and, worse, only 8% of the papers indicated whether the background was periodic (as it should be).

As a consequence, a brief tutorial on the most important fundamentals of FTIR and the influence of the most common parameters on the spectra was presented, where from it was concluded that some kind of standardization of the Minimum Information for Publication of IR-related data when microplastics (MP) are characterized (MIPIR-MP) is required urgently, as in other scientific fields. This would allow interested parties to gain confidence in the reported data and decision-making can be done according to trustworthy scientific data. The situation depicted in this study agreed quite well with other previous reports dealing with MPs analyses where a poor 8% of studies received positive scores for quality criteria and only 14% reported control samples and validation trials.

We encourage authors, reviewers and editors to demand MIPIR-IR data in every study of microplastics as this will translate in social advancements and avoiding discredit of the scientists involved in the study of this major pollution issue. In our opinion, a viable simple paragraph that covers most requisites could be something like (in each case, tailor it):

Measurements were made using a FTIR spectrometer (Model XXX, Commercial Brand), equipped with a [...] (Model and Brand –if needed- of the device used to measure, including detector and aperture when microspectrometry is used), operating in the AAA-BBB cm^{-1} mid-IR (or NIR) region, YY scans/sample, apodization ABC (type, check software), ZZ cm^{-1} nominal resolution. All spectra were corrected for light-reflectance penetration and baseline displacement. Background (empty and clean system) was done before [...] (e.g., measuring each sample). When needed, mention whether Kubelka-Munk or Kramers-Kronig corrections were applied; also specify spectra normalization, spectral treatment and other details that assure reproducibility of the measurements.

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.marpolbul.2020.111035>.

Declaration of competing interest

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

Acknowledgements

This work was partially supported by the Spanish Ministry of Economy and Competitiveness, JPI-Oceans BASEMAN project -partially financed by the European Regional Development Fund program- (Grants: PCIN-2015-170-C02-01 and CTM2016-77945-C3-3-R, ARPA-ACUA). The Program 'Consolidación e Estructuración de Unidades de Investigación Competitiva' of the Galician Government (Xunta de Galicia) is also acknowledged (Grant: ED431C 2017/28).

References

- AlAhmad, Y., Abdelhafez, I., Skenderi, F., Vranić, S., 2019. When progress becomes predation. *The Analytical Scientist* 79, 42–46.
- Andrady, A.L., 2003. *Plastics and the Environment*. John Wiley & Sons, New York, USA.
- Baker, M., 2016. Is there a reproducibility crisis. *Nature* 533, 452–455.
- Bean, H., 2019. How committed are we to reproducibility in science? *The Analytical Scientist* 82, 14–15.
- Belfer, S., Fainchtein, R., Purinson, Y., Kedem, O., 2000. Surface characterization by FTIR-ATR spectroscopy of polyethersulfone membranes-unmodified, modified and protein fouled. *J. Memb. Sci.* 172, 113–124.
- Berg, J., 2018. Progress on reproducibility. *Science* 359, 9–10.
- Bhargava, R., Levin, I.W., 2007. Fourier transform mid-infrared spectroscopic imaging. In: Bhargava, R., Levin, I. (Eds.), *Spectrochemical Analysis Using Infrared Multichannel Detectors*. Blackwell Publishing Ltd, Oxford, UK, pp. 1–24.
- Bruker Optics, 2019. IR tutorial: introduction to FTIR spectroscopy. <https://www.bruker.com/es/products/infrared-near-infrared-and-raman-spectroscopy/landing-pages/general/ir-tutorial.html>.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., Vandesompele, J., Wittwer, C.T., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin. Chem.* 55, 611–622.
- Chalmers, J., Griffiths, P. (Eds.), 2002. *Handbook of Vibrational Spectroscopy*. John Wiley and Sons, Chichester, UK.
- Compton, C., 2018. Garbage in, garbage out. *The Analytical Scientist* 64, 33–39.
- Connors, K.A., Dyer, S.D., Belanger, S.E., 2017. Advancing the quality of environmental microplastic research. *Environ. Toxicol. Chem.* 36, 1697–1703.
- Doyle, W.M., 1992. Principles and applications of Fourier transform infrared (FTIR) process analysis. *Process. Control. Qual.* 2, 11–41.
- European Commission, 2002. COMMISSION DECISION of 12 August 2002 Implementing Council Directive 96/23/EC Concerning the Performance of Analytical Methods and the Interpretation of Results.
- Fidler, F., Wilcox, J., 2018. Reproducibility of scientific results. In: Zalta, E.N. (Ed.), *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University.
- Fuller, S., Gautam, A., 2016. A procedure for measuring microplastics using pressurized fluid extraction. *Environ. Sci. Technol.* 50, 5774–5780.
- GESAMP, 2015. Sources, fate and effects of microplastics in the marine environment: a global assessment. In: Kershaw, P.J. (Ed.), (IMO/FAO/UNESCO-IOC/UNIDO/WMO/IAEA/UN/UNEP/UNDP Joint Group of Experts on the Scientific Aspects of Marine Environmental Protection). Rep. Stud. GESAMP 90.
- Gordon, S.H., Mohamed, A.A., Harry-Okuru, R.E., Biresaw, G., 2015. Identification and measurement of intermolecular interaction in polyester/polystyrene blends by FTIR-photoacoustic spectrometry. *J. Polym. Environ.* 23, 459–469.
- Hanvey, J.S., Lewis, P.J., Lavers, J.L., Crosbie, N.D., Pozo, K., Clarke, B.O., 2017. A review of analytical techniques for quantifying microplastics in sediments. *Anal. Methods* 9, 1369–1383.
- Harris, D.C., 2003. *Quantitative Chemical Analysis*, 6th Edition. W. H. Freeman, New York, USA.
- Hermesen, E., Mintenig, S.M., Besseling, E., Koelmans, A.A., 2018. Quality criteria for the analysis of microplastic in biota samples: a critical review. *Environ. Sci. Technol.* 52, 10230–10240.
- Huppertsberg, S., Knepper, T.P., 2018. Instrumental analysis of microplastics—benefits and challenges. *Anal. Bioanal. Chem.* 410, 6343–6352.
- ISO, 2008. *Evaluation of Measurement Data — Guide to the Expression of Uncertainty in Measurement*. (Geneva (Switzerland)).
- Kanno, M., 2008. Application Note: Development and Applications of a Sample Compartment FTIR Microscope. <https://jascoinc.com/wp-content/uploads/2014/12/Development-and-Applications-of-a-Sample-Compartment-FTIR-Microscope.pdf> (accessed 24.01.2020).
- Kissinger, P., 2019. Replicating failure. *The Analytical Scientist* 77, 37–39.
- Koelmans, A.A., Mohamed Nor, N.H., Hermesen, E., Kooi, M., Mintenig, S.M., De France, J., 2019. Microplastics in freshwaters and drinking water: critical review and assessment of data quality. *Water Res.* 155, 410–422.
- Krafft, C., Sergio, V., 2006. Biomedical applications of Raman and infrared spectroscopy to

- diagnose tissues. *J. Spectrosc.* 20, 195–218.
- Kruer-Zerhusen, N., Cantero-Tubilla, B., Wilson, D.B., 2018. Characterization of cellulose crystallinity after enzymatic treatment using Fourier transform infrared spectroscopy (FTIR). *Cellulose* 25, 37–48.
- Krull, I., 2017. Who's to blame for the reproducibility crisis? *The Analytical Scientist* 49, 33–35.
- Kühn, S., van Oyen, A., Booth, A.M., Meijboom, A., van Franeker, J.A., 2018. Marine microplastic: preparation of relevant test materials for laboratory assessment of ecosystem impacts. *Chemosphere* 213, 103–113.
- Laj, P., Klausen, J., Bilde, M., Plaß-Duelmer, C., Pappalardo, G., Clerbaux, C., Baltensperger, U., Hjorth, J., Simpson, D., Reimann, S., Coheur, P.-F., Richter, A., De Mazière, M., Rudich, Y., McFiggans, G., Torseth, K., Wiedensohler, A., Morin, S., Schulz, M., Allan, J.D., Attié, J.-L., Barnes, I., Birmili, W., Cammas, J.P., Dommen, J., Dorn, H.-P., Fowler, D., Fuzzi, S., Glasius, M., Granier, C., Hermann, M., Isaksen, I.S.A., Kinne, S., Koren, I., Madonna, F., Maione, M., Massling, A., Moehler, O., Mona, L., Monks, P.S., Müller, D., Müller, T., Orphal, J., Peuch, V.-H., Stratmann, F., Tanré, D., Tyndall, G., Abo Riziq, A., Van Roozendael, M., Villani, P., Wehner, B., Wex, H., Zardini, A.A., 2009. Measuring atmospheric composition change. *Atmos. Environ.* 43, 5351–5414.
- Larkin, P., 2017. *Infrared and Raman Spectroscopy: Principles and Spectral Interpretation*. Elsevier, Amsterdam (Netherlands).
- Lin, S.-Y., 2016. An overview of advanced hyphenated techniques for simultaneous analysis and characterization of polymeric materials. *Crit. Rev. Solid State Mater. Sci.* 41, 482–530.
- Lin-Vien, D., Colthup, N.B., Fateley, W.G., Grasselli, J.G., 1991. *The Handbook of Infrared and Raman Characteristic Frequencies of Organic Molecules*. Elsevier, Amsterdam (Netherlands).
- Löder, M.G.J., Gerds, G., 2015. Methodology used for the detection and identification of microplastics—a critical appraisal. In: Bergmann, M., Gutow, L., Klages, M. (Eds.), *Marine Anthropogenic Litter*. Springer.
- Loria-Bastarrachea, M.I., Herrera-Kao, W., Cauich-Rodríguez, J.V., Cervantes-Uc, J.M., Vázquez-Torres, H., Ávila-Ortega, A., 2011. A TG/FTIR study on the thermal degradation of poly (vinyl pyrrolidone). *J. Therm. Anal. Calorim.* 104, 737–742.
- Mayerhöfer, T.G., Pahlow, S., Hübner, U., Popp, J., 2018. Removing interference-based effects from the infrared transmittance spectra of thin films on metallic substrates: a fast and wave optics conform solution. *Analyst* 143, 3164–3175.
- Miller, M.E., Kroon, F.J., Mott, C.A., 2017. Recovering microplastics from marine samples: a review of current practices. *Mar. Pollut. Bull.* 123, 6–18.
- Mukherjee, S., Gowen, A., 2015. A review of recent trends in polymer characterization using non-destructive vibrational spectroscopic modalities and chemical imaging. *Anal. Chim. Acta* 895, 12–34.
- Namba, R., 1990. Industrial applications of GC/FT-IR. In: Ferraro, J.R., Krishnan, K. (Eds.), *Practical Fourier Transform Infrared Spectroscopy*. Elsevier, Amsterdam, Netherlands.
- Nature, 2018. *Challenges in irreproducible research*. <https://www.nature.com/collections/prbfkwmvzv>, Accessed date: 11 December 2019.
- Nature, 2020. *Reporting standards and availability of data, materials, code and protocols*. <https://www.nature.com/nature-research/editorial-policies/reporting-standards> (accessed 24.01.2020).
- Nielsen, T.D., Holmberg, K., Stripplé, J., 2019. Need a bag? A review of public policies on plastic carrier bags – where, how and to what effect? *Waste Manag.* 87, 428–440.
- Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D.P., Hesse, B., Humphreys, M., Ishiyama, J., Karlan, D., Kraut, A., Lupia, A., Mabry, P., Madon, T., Malhotra, N., Mayo-Wilson, E., McNutt, M., Miguel, E., Paluck, E.L., Simonsohn, U., Soderberg, C., Spellman, B.A., Turitto, J., VandenBos, G., Vazire, S., Wagenmakers, E.J., Wilson, R., Yarkoni, T., 2015. Promoting an open research culture. *Science* 348, 1422–1425.
- Perkin Elmer, 2015. *Perkin Elmer Spotlight 200 User's Guide*.
- Plant, A.L., Becker, C.A., Hanisch, R.J., Boisvert, R.F., Possolo, A.M., Elliott, J.T., 2018. How measurement science can improve confidence in research results. *PLoS Biol.* 16, e2004299.
- Polyanskiy, M.N., 2020. *Refractive Index Database*. <https://refractiveindex.info/> (accessed 21.01.2020).
- Prati, S., Joseph, E., Scutti, G., Mazzeo, R., 2010. New advances in the application of FTIR microscopy and spectroscopy for the characterization of artistic materials. *Acc. Chem. Res.* 43, 792–801.
- Renner, G., Schmidt, T.C., Schram, J., 2017. A new chemometric approach for automatic identification of microplastics from environmental compartments based on FT-IR spectroscopy. *Anal. Chem.* 89, 12045–12053.
- Robertson, I., Lawson-wood, K., 2017. Application Note, Perkin Elmer. *Hyphenated Technology – Advanced Data Analysis of Evolved Gases from TG-IR Hyphenation Studies of Polymers*.
- SAM, 2019. *Environmental and Health Impacts of Microplastic Pollution - From Scientific Evidence to Policy Advice*. An Expert Workshop of the Group of Chief Scientific Advisors of the European Commission's Scientific Advice Mechanism.
- Sansone, S.-A., Roca-Serra, P., McQuilton, P., Thurston, M., Lister, A., Izzo, M., Granel, R., Batista, D., Dauga, D., 2009. <https://fairsharing.org/standards> (accessed 21.01.20).
- SAPEA, 2018. *A Scientific Perspective on Microplastics in Nature and Society*. (Berlin, Germany).
- Schmidt, S., Blomberg, L., Wännman, T., 1988. Dynamic headspace enrichment/re-injection for open tubular GC/FTIR analysis of volatiles in polymers. *J. High Resolut. Chromatogr.* 11, 242–247.
- Shmadzu, 2008. *Application Note: Kramers-Kronig Transform and Applications*.
- Silva, A.B., Bastos, A.S., Justino, C.I.L., da Costa, J.P., Duarte, A.C., Rocha-Santos, T.A.P., 2018. Microplastics in the environment: challenges in analytical chemistry-a review. *Anal. Chim. Acta* 1017, 1–19.
- Smith, B.C., 2011. *Fundamentals of Fourier Transform Infrared Spectroscopy*. CRC press, Boca Raton, (USA).
- Sparks, D.T., Lam, R.B., Isenhour, T.L., 1982. Quantitative gas chromatography/Fourier transform infrared spectrometry with integrated Gram-Schmidt reconstruction intensities. *Anal. Chem.* 54, 1922–1926.
- Stark, P.B., 2018. No reproducibility without preproducibility. *Nature* 557, 613.
- Stuart, B., 2015. *Infrared spectroscopy*. In: Kirk-Othmer Encyclopedia of Chemical Technology. John Wiley and Sons, Chichester, UK.
- Taylor, C.F., Field, D., Sansone, S.-A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C.A., Binz, P.-A., Bogue, M., Booth, T., 2008. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.* 26, 889.
- Tretinnikov, O.N., Zagorskaya, S.A., 2012. Determination of the degree of crystallinity of poly (vinyl alcohol) by FTIR spectroscopy. *J. Appl. Spectrosc.* 79, 521–526.
- Vasanthan, N., Salem, D.R., 2000. Infrared spectroscopic characterization of oriented polyamide 66: band assignment and crystallinity measurement. *J. Polym. Sci. Part B Polym. Phys.* 38, 516–524.
- Villberg, K., Vejjanen, A., Gustafsson, I., Wickström, K., 1997. Analysis of odour and taste problems in high-density polyethylene. *J. Chromatogr. A* 791, 213–219.
- Vollmer, M., Möllmann, K.-P., 2017. *Infrared Thermal Imaging: Fundamentals, Research and Applications*. John Wiley & Sons, New York (USA).
- Wang, J., Tan, Z., Peng, J., Qiu, Q., Li, M., 2016. The behaviors of microplastics in the marine environment. *Mar. Environ. Res.* 113, 7–17.
- Warwick, B.C., Eda, K.E., 2014. *Keysight Technologies Understanding the Kramers-Kronig Relation Using a Pictorial Proof*, pp. 1–10. <https://www.keysight.com/us/en/assets/7018-02435/white-papers/5990-5266.pdf>.
- Wesch, C., Elert, A.M., Wörner, M., Braun, U., Klein, R., Paulus, M., 2017. Assuring quality in microplastic monitoring: about the value of clean-air devices as essentials for verified data. *Sci. Rep.* 7, 5424.
- Yu, J., Wang, P., Ni, F., Cizdziel, J., Wu, D., Zhao, Q., Zhou, Y., 2019. Characterization of microplastics in environment by thermal gravimetric analysis coupled with Fourier transform infrared spectroscopy. *Mar. Pollut. Bull.* 145, 153–160.

<i>Environmental Pollution</i> , 2014, 195, 163-166		X	X	X	X				
<i>Environmental Pollution</i> , 2015, 207, 190-195	X	X	X	X	X				
<i>Environmental Pollution</i> , 2017, 227, 167-174	X	X	X	X				X	
<i>Environmental Pollution</i> , 2019, 113440	X								
<i>Environmental Pollution</i> , 2019, 113450	X	X	X	X					
<i>Environmental Pollution</i> , 2019, 113658	X	X		X					N/A
<i>Environmental Pollution</i> , 2019, 13391	X	X	X	X	X	X			
<i>Environmental Pollution</i> , 2019, 231-2, 1552-1559	X								
<i>Environmental Pollution</i> , 2019, 244, 503-512	X	X	X	X	X	X			N/A
<i>Environmental Pollution</i> , 2019, 244, 675-684	X	X			X		N/A		N/A
<i>Environmental Pollution</i> , 2019, 244, 675-684	X	X							N/A
<i>Environmental Pollution</i> , 2019, 244, 827-833	X	X	X	X					
<i>Environmental Pollution</i> , 2019, 245, 1107-1116	X	X	X	X					
<i>Environmental Pollution</i> , 2019, 246, 174-182	X	X							
<i>Environmental Pollution</i> , 2019, 246, 26-33	X								
<i>Environmental Pollution</i> , 2019, 250, 292-299	X	X							N/A
<i>Environmental Pollution</i> , 2019, 252, 1357-1366	X	X		X			N/A		N/A
<i>Environmental Pollution</i> , 2019, 252, 1357-1366	X	X	X	X					X
<i>Environmental Pollution</i> , 2019, 252B, 1719-1729	X	X							
<i>Environmental Pollution</i> , 2019, 253, 297-301	X	X		X					
<i>Environmental Pollution</i> , 2019, 253, 297-301	X	X		X					
<i>Environmental Pollution</i> , 2019, 254B, 113104	X								
<i>Environmental Pollution</i> , 2019, doi j.envpol.2019.113450	X	X	X	X					
<i>Environmental Pollution</i> , 2019, doi j.envpol.2019.113658	X	X		X			N/A		N/A
<i>Environmental Pollution</i> , 2020, doi j.envpol.2020.113948	X	X		X	X	X			X
<i>Environmental Science and Technology</i> , 2016, 50, 5668-5675	X	X	X	X					X
<i>Environmental Science and Technology</i> , 2017, 51, 4368-4376	X	X		X					
<i>Environmental Science and Technology</i> , 2017, 51, 810-818	X	X	X	X					
<i>Environmental Science and Technology</i> , 2017,51, 1100-11010	X	X	X	X					
<i>Environmental Science and Technology</i> , 2018, 52, 12188-12197	X								
<i>Environmental Science and Technology</i> , 2018, 52, 13279-13288	X	X	X	X					X
<i>Environmental Science and Technology</i> , 2018, 52, 14480-14486	X								
<i>Environmental Science and Technology</i> , 2019, 53, 12218-12226	X	X	X	X					
<i>Environmental Science and Technology</i> , 2019, 53, 12300-12310	X	X							
<i>Environmental Science and Technology</i> , 2019, 53, 14204-14211	X	X	X	X					
<i>Estuarine, Coastal and Shelf Science</i> , 2013, 130, 54-61	X	X	X	X					
<i>Fibers and Polymers</i> , 2010, 11(5), 710-718	X	X	X	X					N/A
<i>Journal of Polymer Research</i> , 2017, 24:133	X								
<i>Journal of Polymer Research</i> , 2019, 26, 242	X	X	X	X					
<i>Journal of Spectroscopy</i> , 2015:586514	X	X	X						
<i>Marine Environmental Research</i> , 2012, 81, 70-77	X	X							N/A
<i>Marine Environmental Research</i> , 2014, 95, 89-95	X	X	X	X					X
<i>Marine Environmental Research</i> , 2015, 111, 18-26	X	X	X	X	X	X			
<i>Marine Environmental Research</i> , 2016, 114, 24-30	X	X	X	X					N/A
<i>Marine Environmental Research</i> , 2016, 120, 1-8	X	X	X	X					
<i>Marine Environmental Research</i> , 2019, 151, 104768	X	X							
<i>Marine Pollution Bulletin</i> , 2018, 133, 553-561	X	X	X		X	X			
<i>Marine Pollution Bulletin</i> , 2011, 62, 2199-2204	X	X	X	X					
<i>Marine Pollution Bulletin</i> , 2014, 79, 278-283	X	X	X	X					
<i>Marine Pollution Bulletin</i> , 2016, 110, 299-308	X	X	X						
<i>Marine Pollution Bulletin</i> , 2018, 133, 553-561	X	X	X	X	X				
<i>Marine Pollution Bulletin</i> , 2018, 135, 187-194	X	X	X	X	X	X			X
<i>Marine Pollution Bulletin</i> , 2019, 140, 262-273	X	X	X	X					
<i>Marine Pollution Bulletin</i> , 2019, 142, 112-118	X	X							
<i>Marine Pollution Bulletin</i> , 2019, 142, 263-273	X								

<i>Marine Pollution Bulletin, 2019, 142, 510-519</i>	X	X						N/A	
<i>Marine Pollution Bulletin, 2019, 145, 334-342</i>	X	X	X	X					
<i>Polymer Degradation and Stability, 2013, 98, 22-36</i>	X	X	X					N/A	
<i>Polymer Degradation and Stability, 2015, 112, 145-459</i>	X	X	X	X					
<i>Science of the Total Environment, 2012, 416, 455-463</i>	X	X	X	X	X			X	
<i>Science of the Total Environment, 2017, 586, 10-15</i>	X	X	X		X				
<i>Science of the Total Environment, 2018, 628-629, 740-747</i>	X	X		X					
<i>Science of the Total Environment, 2018, 645, 1388-1399</i>	X	X	X	X	X	X		X	
<i>Science of the Total Environment, 2019, 687, 1186-1196</i>	X	X	X	X					
<i>Science of the Total Environment, 2019, 689, 341-346</i>	X	X	X	X				N/A	
<i>Science of the Total Environment, 2019, 690, 821-830</i>	X	X	X	X					
<i>Science of the Total Environment, 2019, 694, 133712</i>	X	X	X					N/A	
<i>Science of the Total Environment, 2020, 703, 134768</i>	X		X	X					
<i>Science of the Total Environment, 2020, 703, 134947</i>	X	X	X	X	X	X		N/A	
<i>Water Research, 2016, 91, 174-182</i>	X	X							
<i>Water Research, 2017, 112, 93-99</i>	X	X	X	X					
<i>Water Research, 2018, 133, 236-246</i>	X	X	X	X					
<i>Water Research, 2018, 147, 382-392</i>	X	X	X	X					
<i>Water Research, 2019, 157, 228-237</i>	X	X							
<i>Water Research, 2019, 158, 392-400</i>	X	X							
<i>Water Research, 2019, 161, 560-569</i>	X	X	X	X					
<i>Water Research, 2019, 163, 114909</i>	X	X		X					
<i>Water Research, 2019, 164, 114968</i>	X	X	X	X					
<i>Water, Air, & Soil Pollution, 2017, 228:470</i>									X
<i>Water, Air, & Soil Pollution, 2019, 230:257</i>	X	X							
<i>Water, Air, & Soil Pollution, 2014, 225:1853</i>	X	X	X	X					

**Chapter 6: Monitorization of
Polyamide Microplastics
Weathering Using Attenuated
Total Reflectance and
Microreflectance Infrared
Spectrometry.**





Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Monitorization of polyamide microplastics weathering using attenuated total reflectance and microreflectance infrared spectrometry



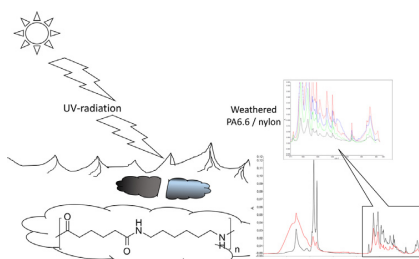
Verónica Fernández-González, Jose Manuel Andrade*, Borja Ferreiro, Purificación López-Mahía, Soledad Muniategui-Lorenzo

Group of Applied Analytical Chemistry (QANAP) and Instituto Universitario de Medio Ambiente (IUMA), University of A Coruña, E-15071 A Coruña, Spain

HIGHLIGHTS

- Nylon microplastics identification may be defective if weathering is neglected.
- Usual databases do not contain spectra of environmentally relevant polymers.
- Polyamide weathering was monitored and interpreted chemically.
- A collection of IR indexes reflect overall weathering routes.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 15 January 2021
 Received in revised form 2 July 2021
 Accepted 4 July 2021
 Available online 10 July 2021

Keywords:

Polyamide 6.6
 Nylon
 Photooxidation
 Weathering
 Microplastic
 Marine environment

ABSTRACT

The EU goal to reduce marine plastic litter by ca. 30% by 2020 stressed the need to deploy analytical methods to ascertain the polymeric nature of a residue. Furthermore, as plastics age under natural conditions and usual databases do not include their weathered spectra, (micro)plastics in environmental samples may be unidentified. In this paper, polyamide (nylon) microplastics weathering was monitored because of its ubiquity in household commodities, clothes, fishery items and industry, whose residues end up frequently in the environment. Infrared spectra (ATR and microreflectance) and Scanning Electron Microscopy (SEM) images were collected periodically while exposing nylon to controlled weathering. It was seen that ATR was more sensitive than microreflectance to monitor the structural evolution of polyamide and that the spectra and the surface of weathered microplastics showed remarkable differences with the pristine material, which stresses the need for considering its evolution when identifying microplastics in environmental studies. The evolution of six band ratios related to the chemical evolution of this polymer are presented. SEM images revealed the formation of secondary microplastics at the most advanced weathering stages of polyamide.

© 2021 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The so many advantages that plastics brought to human technology and wellbeing cannot be denied. Unfortunately, their success might also bring about a (so far hidden) planetary boundary

threat if we consider the Persson's *et al* studies [1,2]. Their presence in the environment is worldwide recognized, even in very large quantities (think about the oceanic gyres where thousands of plastic items and microplastics accumulate), although it is not still clear the effects they may cause in the different compartments of Nature.

The European Global Strategy on Circular Economy [3,4] targeted actions to reduce plastic consumption, increase its

* Corresponding author.

E-mail address: andrade@udc.es (J.M. Andrade).

recyclability and reduce its disposal, and it also proposed the development of innovative plastics (e.g., biodegradable). At present, around 50% of plastic residues go to landfill (although this figure varies throughout Europe) and only 25% of the collected plastic is recycled. A major objective of the EU is to reduce plastic marine litter, about 30% by 2020, and, so, the Green paper on plastic waste was deployed [5].

A good example of a plastic commodity is polyamide –PA- (best known commercially as nylon). This is the general term for a broad range of chemical formulations, out of which PA 6 and PA 6.6 outstand for textile and plastic industries. The former gives rise to a commercial product known as Perlon [6] or, more common, Nylon 6 (CAS number 25038–54–4, polycaprolactam) highly similar to PA 6.6 although it absorbs more humidity, and has lower strength and stiffness than that. The latter, PA 6.6 or Nylon 6.6 (CAS number 32131–17–2, poly(hexamethylene adipamide)), is obtained through a condensation polymerization of two monomers each containing 6 carbon atoms –hexamethylenediamide and adipic acid- (from which the name Nylon 6.6 arose). Crystallization is due to strong hydrogen bonding between the chains [7]. The uses and properties of both polymers are almost the same (many times they are mentioned indistinctly) and the choice between the two polymers is often made on non-technical grounds, e.g. local availability, price or familiarity [8].

The worldwide production of PA 6 and PA 6.6 (in the following, PA) raised to 7.8 million tons in 2016 and found many industrial applications [9]. It was reported that 5.7 million tons of polyamide fibers were produced worldwide in 2017, which is about 8% of the worldwide production of synthetic fibers [10]. The most important sales were related to carpet and staple fibers, textiles and industrial filaments (including fishing nets, lines, etc.). Carpets and rugs account for 16% of the PA consumed and are expected to grow at 0.7% per year. North America and Western and Central Europe are the major producers of carpets and rugs, accounting for about 80% of the nylon fiber consumed [11].

The overall consumption of nylon fiber is expected to grow at about 3.5% per year for the next five years, with Northeast Asia accounting for nearly two-thirds of the world's nylon fiber output. PA is also a very relevant engineering plastic for automotive industries (ca. 38% of the PA production), electrical and electronics (21%) [12].

Contrary to other plastics, like PET (whose ubiquitous worldwide presence is obvious due to its use to get disposable containers, like bottles), the PA environmental impact might be not so evident as it tends to sink in seawater and is mostly related to fishing, household and industrial activities, including dry-cleaning, which –most times- end up in wastewater treatment plants [13,14]. As a token of the relevance of PA in the marine environment, some papers can be cited. Thus, the majority of microplastics (MPs) found in a natural protected area at Southern Spain corresponded to PA, likely from a nearby harbour and fishing activities [15]. A river basin polluted by anthropogenic discharges showed PA as the most frequent polymer [16] and it was the second most frequent polymer in fibers collected from fish samples off North-east Greenland [17] (considering all MPs identified by infrared spectrometry, it was the third most abundant polymer there). PA constituted the second most common type of MPs found in the water column of the Gulf of Lions [18] and two papers discovered also big quantities of PA microparticles in shrimps [19] and in reefs at the South China Sea (second polymer after polypropylene) [20]. Finally, PA was third among the selected MPs that underwent chemical characterization in samples from the Mediterranean area [21], the Baltic sea [22] and the coastal area of Tamil Nadu (India) [23], in the latter case due to recreational and fishing activities.

Despite the number of papers related to the environmental presence of PA in the environment is rising, the evolution of PA

microplastics has not been considering in depth and, so, this paper attempts to shed some light on this issue. Not in vain, weathering might be a key factor to study the ad/absorption of pollutants on plastics and their potential impacts in the biota [2].

Infrared spectrometry (IR) has become a *de facto* standard to rigorously identify polymer fragments in the environment. However, as plastics degrade under solar and marine conditions (see, e.g. [2,24,25]). Their IR spectra have to be compared to a collection of known degraded polymer spectra to find out the best match. Unfortunately, to best of the authors' knowledge, almost no commercial database contains series of degraded polymers and, so, many plastic fragments may be misidentified in environmental samples. This topic was discussed in recent papers [25,26,27,28]. Notwithstanding, it is critical to keep in mind that this problem is insidious because even if the evolution of a plastic is known the variety of commercial brands (including different compositions and additives) and the complex natural conditions might modify the weathering routes/products for that plastic (see, e.g. [28] for a good discussion on this topic).

The major objectives of this paper are twofold: first, to use ATR –attenuated total reflection- and reflectance IR microspectrometry to monitor PA 6.6 weathering (the latter being applied first time for this purpose). Second, to relate the changes observed into the spectra to polymeric structural modifications. As this polymer is not studied frequently in the microplastics literature it is expected that this work will help both scientists focused on studying the evolution of MP in the sea (for whom some band ratios are proposed) and those performing routine measurements aimed at identifying and quantifying microplastics in environmental samples (for whom the evolution of the spectral characteristics are vital).

2. Materials and methods

2.1. Samples

The PA 6.6 polymer used in this study corresponds to the commercial name 'Ultramid', from BASF, whose density was 1.13 g. cm⁻³; melting temperature 260 °C. Two presentations were available: pellets, ca. 4 mm diameter, and powder, with a size Gaussian distribution centred around 100 µm (standard deviation ca. 80 µm). The polyamide was fabricated so as to contain as few additives as possible, and it was included in a collection of microplastics developed for the JPI-Oceans-funded project BASEMAN.

A device for the standardized simulated weathering of powdered and pelletized PA microplastics was used. In brief, it resembles the natural seawater conditions by setting continuous agitation, sand erosion, sunlight irradiation and oxidative conditions. It uses two metal halide lamps whose emission spectra matches that of the solar radiation at the Earth surface (medium latitude). The overall UV/VIS illuminance was 12,200 lx in a continuous mode (i.e., no day–light cycles). The total irradiation time was eleven weeks (ca. 1850 irradiation hours). More details can be found elsewhere [29].

For the seawater weathering conditions 10 g of PA powder and 20 g of PA pellets were placed in 1000 mL Pyrex cylinders, along with siliceous sea-sand and natural, filtered (10 µm) seawater. For the dry conditions (emulating solar irradiation at the shoreline) the same quantities were placed in 12 cm diameter Petri dishes. Control samples were also considered (pellets and powder submerged in seawater at dark). The contents of the dishes were stirred manually each two or three days. Seawater, dry and control aliquots were taken each fortnight (15, 30, 45, 60 and 75 days; corresponding to 360, 720, 1080, 1440 and 1800 irradiation hours, respectively).

2.2. Apparatus and equipments

A Spectrum 400 FTIR spectrometer, Perkin Elmer, equipped with a horizontal one-bounce single-reflection ATR (attenuated total reflection) diamond (MIRacle, Pike), operating in the 4000–600 cm^{-1} mid-IR region, 30 scans/sample, apodization Beer-Norton strong, 4 cm^{-1} nominal resolution, was employed throughout. All ATR spectra were corrected for light-reflectance penetration and baseline displacement.

A Perkin Elmer Spotlight 200i IR microscope was used to perform the reflectance measurements. Each item (pellet or granule) was measured twice (changing its position) and the resulting spectra were averaged. The experimental setup was: nominal resolution: 4 cm^{-1} ; number of scans: 200; spectral range: 3500–600 cm^{-1} ; nominal aperture: 100 μm (adjusted whenever the scanned part of the grain required it); apodization: Beer-Norton, strong; spectral processing: Kubelka-Munk transformation (for pellets) and normalization 10% plus Kramers-Kronig (for powders). All spectra were baseline corrected using multipoint baseline correction. Note that the surface characteristics of the item under measurement affect data processing due to their possible main interaction with the IR beam: pellets allow for measuring almost flat surfaces, making the reflection almost specular, whereas for powders the irregular shape of the grain's surface makes reflection more diffuse.

Reflectance measurements were selected instead of transmittance ones because the former are the most common ones in literature. Also, reflectance spectra are independent of the thickness of the particles (transmittance can be affected by the effective pathlength) and can be applied straightforwardly. Further, as reflectance is a surface-characterization technique it can detect polymer ageing quite accurately (a superficial phenomenon). Note also that the size of the particles handled in this work ($>70 \mu\text{m}$) is far from the physical limitations that can occur when very small ($<20 \mu\text{m}$) particles are measured. This allows reflectance to be applied safely to most common environmental monitoring studies (for instance, when Neustonic or manta-trawl nets are used their pore sizes are usually around 200–330 μm).

A JEOL JSM-6400 Scanning Electron Microscope (SEM), coupled to an Energy-dispersive X-ray spectroscopic microanalysis device (EDXA, Oxford INCA Energy 200) was employed. All samples were covered with a gold film using a cathodic spraying system (BAL-TEC SCD 004) prior to the SEM measurements.

3. Results and discussion

3.1. Chemical interpretation of the IR spectra

In this section, PA spectra are interpreted in detail and related to the chemical processes PA undergoes during weathering. This is in order here as most previous reports focused only on partial aspects of the spectrum and a full interpretation was not found. Thus, a brief review is presented. With respect to PA degradation, although the cleavage of N-C bonds at short radiations (254 nm) to yield amines and aldehydes was clearly studied, it is not evident what happens within the solar spectral region (300 nm and higher) [30]. Four principal weathering routes for PA are described, of which the first two (see next paragraphs) are the major ones, however notice that all them may occur almost conjunctly:

- (i) Photooxidation can be initiated by chromophoric impurities from fabrication (unbounded or partially bounded functional groups). They lead to an initiation step that can

involve either the photolysis of hydroperoxides or ketones formed during fabrication (which is a common mechanism for most polyolefins –Norrish I and II mechanisms–) [31].

- (ii) Photooxidation can also be due to oxygen-induced reactions because of charge transfer complexes [30,32]. This seems particularly feasible for PA6.6 and involves direct photolysis of the amide bond [32].
- (iii) The hydrolytic degradation route is attributed to the especial susceptibility of the amide bond to acid- and basic-catalysed hydrolysis [30]. The catalysing metals might be those present naturally in seawater and those used at the production catalysts or embedded salts [33].
- (iv) Finally, thermal oxidation –a non photooxidative process– is also possible by means of the abstraction of hydrogen atoms on the methylene groups close to the nitrogen atom of the amide group [34].

Roughly, photodegradation would lead to aldehydic and acidic groups, whereas hydrolysis would lead to carboxylic acid and amine groups [30]. In essence, the mechanisms of thermal and photochemical oxidation are identical, but for their onset [31]. Although thermal oxidation should not be a major weathering route in this study, because neither the water nor the air temperatures surpassed 30 °C, it cannot be disregarded totally (as it will be shown later on).

The ATR spectra of pristine (*as received*) pellets and powder and last aliquots of weathered PA under the three weathering setups (seawater plus irradiation; dry plus irradiation and control-dark) are shown in Fig. 1. First, it is observed that no new distinctive spectral bands appear throughout weathering, as reported for outdoor weathering of ropes [30]. However, this might be a consequence of the broad spectral bands and their mutual overlaps.

All PA spectra show the typical CH-related spectral bands (stretching, at 2863 and 2934 cm^{-1} ; and bending, 1450 and 1375 cm^{-1}). The amide group yields three major sets of characteristic bands [7,35,36,37]:

1. The NH stretch-free motion at ca. 3400 cm^{-1} ; the NH stretch H-bonded, around 3300 cm^{-1} ; and a NH bending overtone around 3070 cm^{-1} (which might coincide with a weak NH stretch H-bonded band [35,36] and C-N stretching [38]). Here, the most distinguishable band is the peak at 3300 cm^{-1} , which increases roughly with ageing, likely because of two main processes: i) the photolysis of the amide bond to yield amines [33,34], and ii) the introduction/appearance of hydroxyl groups in the polymer. This band is much broader for seawater-submerged microplastics than for dry weathering, where a relatively sharp peak (associated to free N-H stretching, ca. 3400 cm^{-1}) is more alike that reported by Ksouri *et al.* [35]. The hydroxyl groups (broad band between 3000–3500 cm^{-1}) would correspond, mostly, to water absorbed in the plastic (despite secondary or tertiary alcohols or even hydrogen bonds between carboxylic groups cannot be disregarded). In effect, it was reported that PA 6.6 absorbs more than 8% of its weight in 100% relative humidity at room temperature [35]. The hygroscopic introduction of the OH groups in the structure is due to the presence of the polar functional groups and it occurs, mainly, into the amorphous regions of the polymer [35]. Note that in our case the two experimental setups that involved immersion in seawater yielded spectral profiles different from the dry weathering setup (see Fig. 1).
2. A strong sharp band around 1633 cm^{-1} corresponding to the carbonyl (C = O) stretching (it is also called Amide I band). Its overall increase (see Fig. 1) characterizes the final products of both the photodegradation and hydrolysis reactions because

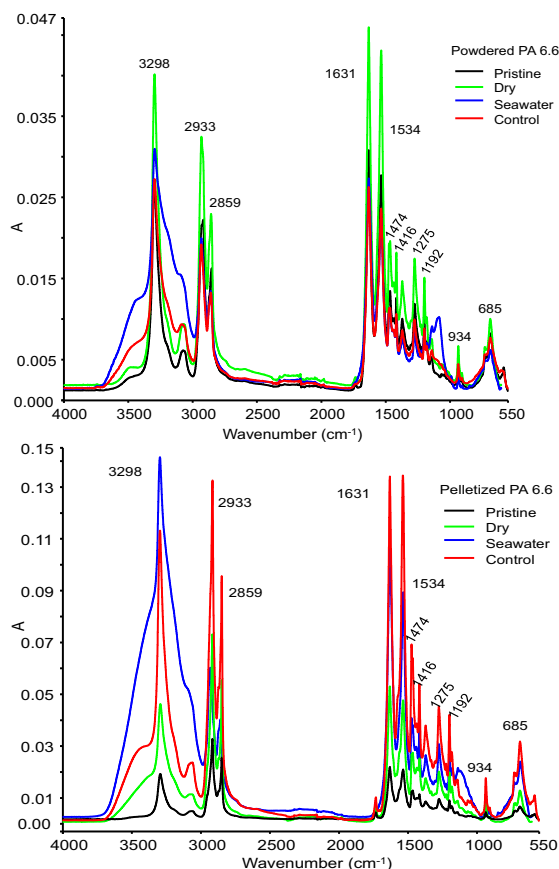


Fig. 1. Comparison of the PA spectra for powder and pellets obtained by ATR: pristine (as received), and last aliquots of controls, dry weathering and seawater weathering.

aldehydes, ketones and acid groups display highly close strong absorption bands in this region. Therefore, any assignment to a particular functional group is troublesome.

3. The band at 1537 cm^{-1} (Amide II band) is associated to the NH mono-substituted amide bending plus the C-N stretching. Another band could appear around 1650 cm^{-1} , although overlapped with the Amide I band [35,36]. The Amide II band rises clearly with ageing (as well as the 696 cm^{-1} one, related to end NH_2 groups [39]) and, hence, indicates the polymer chain scission. Observe (Fig. 1) that the bands are more intense for the PA samples weathered under dry conditions, likely suggesting a participation of thermal degradation processes (contrary to the PA seawater-immersed ones).

Some other interesting bands can be considered:

- i. The unique band in the spectrum that decreased clearly with ageing was at 1735 cm^{-1} , with two interpretations:
- a. The presence of cyclopentanone derivatives originated at fabrication process, whose content decreases during thermo-oxidation processes [40]. Noteworthy, this peak is only seen for pellets weathered under dry conditions, which might sup-

port the reported steady disappearance of the residual cyclopentanone. That peak is neither present in powdered samples nor in seawater submerged ones and, so, we hypothesized that cyclopentanone disappeared there much faster thanks to the much smaller size of the solid grains (in seawater it would leach to the water).

- b. Another possibility is to relate this band to a vibration form of hydrogen-bonded carbonyl groups. PA has charged regions in the polymer chain becoming attracted to each other, which causes the individual polymer chains to fold back over on itself and for other polymer chains to similarly be attracted to each other [41]. As this occurs mostly in the amorphous regions, the increase on crystallinity (to be discussed later) would lead to the disappearance of this band.
 - ii. The band centred at 1150 cm^{-1} (not visible in the reflectance spectra) is very prominent for the seawater conditions and it could suggest tertiary or secondary alcohols, ketones (asymmetric C-C stretching, with central C containing the O), ethers (symmetric and asymmetric stretching of C-O-C), esters (C-O-R bending) and/or the typical combination band of the carboxylic acids (C-O stretching plus OH bending). All these structures are compatible with reported end products for photooxidation and/or thermo-oxidation of PA [31,32,34,38].
 - iii. Two weak bands appear around 930 and 920 cm^{-1} , related to the CO-NH in-plane vibration [42], and another one at 696 cm^{-1} , related to end NH_2 groups [39].
 - iv. A relevant property of any polymer is crystallinity, which was studied extensively by Li and Goddard [43] and Vasanthan and Salem [42]. They established that out of four possible forms only two are stable; α and γ [43]. The γ form (amorphous) tends to convert into the former (crystalline), even when only mechanical stress is applied at room temperature. In the α form the plane of the amide group and that of the $(\text{CH}_2)_5$ group are parallel, while in the γ form they are approximately perpendicular. H-bonds are formed between the two forms. Vasanthan and Salem established [7,42] that the bands at 936 and 1200 cm^{-1} are due to the crystalline conformation. They correspond to the CO - NH in-plane vibration [7] (alternately, to the C-CO amide axial stretching [44]) and the symmetrical CH_2 twist-wag angular deformation out of plane [7], respectively). A shoulder at 924 cm^{-1} (CO - NH in-plane stretching) and 1136 cm^{-1} (C - C stretching) correspond to the amorphous structure [7]. The bands at 936 and 1200 cm^{-1} are usually sharper than the other two. The crystalline fraction can be calculated from the IR spectrum as $8.8 \cdot (A_{936}/A_{1630})$ because the C = O band at 1630 cm^{-1} does not depend on crystallinity [7]. Other possible bands to characterize crystallinity are those at 976 , 1030 , and 1074 cm^{-1} [39] but they are used less. Unfortunately, these bands are very weak in the reflectance spectra so these calculations become compromised.

With regards to the reflectance spectra of pellets although those characteristics hold on (Figure 2), some interesting differences can be underlined:

- i. Broad bands due to the -OH groups are not seen, which simplifies the identification of several peaks (mostly, the NH stretch-free motion at ca. 3400 cm^{-1} ; the NH stretch H-bonded, around 3300 cm^{-1} ; and a NH bending overtone

around 3070 cm^{-1}). However, it is worth noting that the band ca. 3400 cm^{-1} was not clearly visible until the most weathered stages of the polymer.

- ii. Spectra are dominated by strong and quite narrow peaks which, in essence, correspond to the amide functional groups. On the contrary, other bands related to -e.g.- the CH and OH moieties of the polymeric skeleton yield weak bands. Thus, despite the reflectance spectra of the pellets appear less noisy than their ATR counterparts (Fig. 2), they are less sensitive and, unfortunately, this will have consequences in some band ratios, as it will be shown in next sections.

The microreflectance spectra of the small granules ($70\text{--}300\text{ }\mu\text{m}$) were much less intense than the ATR ones and also than those for pellets measured by microreflectance (Fig. 2). This is attributed to the diffuse scattering and radiation interactions caused by the rough and irregular surface of the granules. This adds to the intrinsic reduced intensity of the microreflectance technique. The most intense and clear bands are associated to the typical amide group, whereas the other bands are less clear. This hampers both the chemical interpretation and the calculations related to the less intense bands. Therefore, they will not be considered in more detail in next sections. Nevertheless, it is derived immediately from these facts that the use of adequate (in this case, home-made) databases is of absolute, paramount importance to identify polymers from field samples. In effect, current spectral databases will not be of

use for microreflectance spectra as they are based on transmittance or absorbance units.

3.2. Monitoring PA6.6 evolution using spectral indexes

It is common practice in weathering studies to monitor the evolution of the polymer using band ratios. They normalize bands associated to functional groups of interest against a reference peak that ideally is unaffected by the evolution of that functional group and remains constant with time. Here, the C-H bending peak located around 2915 cm^{-1} was selected. To the best of our knowledge weathering ratios have not still been established for PA 6.6 and, hence, several options are evaluated here, including some employed to monitor other polymers [45,46]. Another approach to monitor weathering may be to consider spectral differences between the spectra of the aged aliquots and that of the pristine polymer, as a referee suggested. This option was not explored in the present manuscript but we acknowledge its feasibility. To simplify readability, hereinafter comments and studies will be given for the ATR spectra. Particular details will be added for the reflectance spectra whenever appropriate (in general, as a comparison with the ATR results).

3.2.1. The carbonyl index

Defined as the A_{1640}/A_{2914} integrated areas ratio, the carbonyl index considers the C=O stretching band and has been of widespread use when studying different polymers [28,45,46]. The only similar application for PA is a work from Gijssman *et al.* [32], although they used a subtraction, not really a ratio.

In Fig. 3a the evolution of the index is shown for both pellets and powder. It raises with time for all weathering setups, being higher for the simulated marine conditions (either because of the kinetics being faster into seawater or because of the displacement of the chemical equilibria due to the release of the products to the aqueous phase). A steady, smooth increment is seen even for the control samples (pellets or powder kept at dark, immersed in seawater), which confirms that the mere presence of water interpenetrates the bonds between the -NH and C=O groups and degrades the polymer [35,41]. No doubt, this process is much slower when light is absent, but it occurs. Indeed, note that the samples weathered without water led to a lower carbonyl formation than the controls.

Although the general rising trend is clear, three stages can be seen in the plots, and they agree with previous studies [39,40]. A first, fast increase indicates the oxidative degradation (mainly in the amorphous phase, where oxygen can diffuse fast). Then, a plateau is observed probably because the oxidative process slows down due to a reduction in the amount of amorphous phase; finally, a continuous increase because of the degradation of the crystalline phase [40]. For powdered microplastics, the dry and control conditions lead to almost coincident situations and despite an initial weathering the trends are very smooth. This agrees with Goodridge *et al.* [41] and they attributed the evolution to a time-dependent change of the amorphous phase and reported that this is a general phenomenon for all polymeric materials.

Finally, a comparison of the spectra in Fig. 3a reveals that weathering seems more pronounced for the pellets, likely because the plastic powder might have undergone already some previous degradation and, so, their initial states are not exactly the same. This seems reinforced by the carbonyl ratio for the 'pristine' (as received) samples, 1.08 for powder and 0.52 for pellets. Recall that powder fabrication implies a huge stress on the polymer and this has been shown to increase the physical and chemical degradation [34]. The final sudden increment for the dry powder after the eighth week can only be explained by a 'collapse' (i.e., strong degradation) of the polymer. This hypothesis was confirmed by

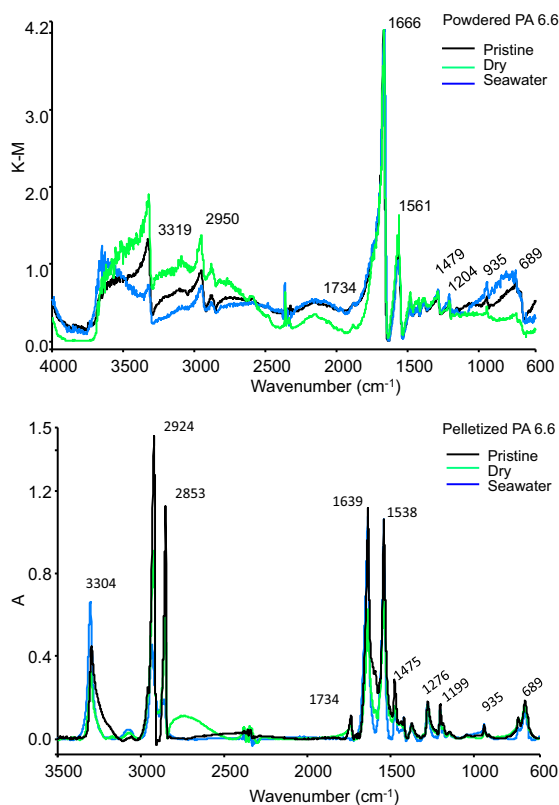


Fig. 2. Comparison of the PA spectra for powder (Kubelka-Munk ordinate units) and pellets obtained by reflectance microspectrometry: pristine (as received), and last aliquots of controls, dry weathering and seawater weathering.

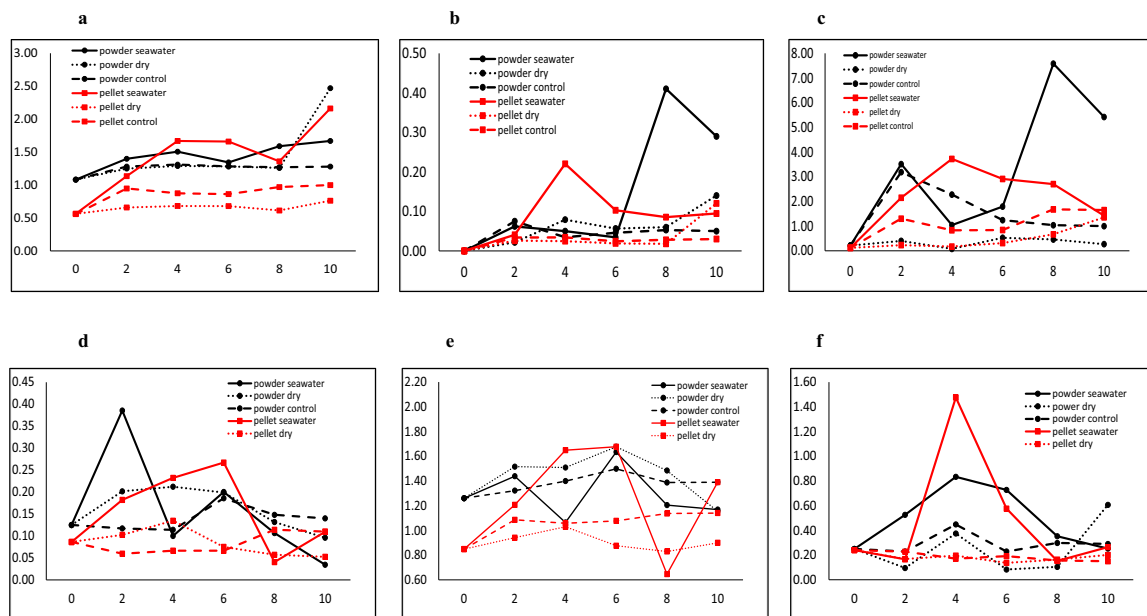


Fig. 3. Evolution of several functional indexes to monitor PA weathering using ATR: a) C=O; b) C-O; c) O-H; d) crystallinity; e) amide II; and f) long chains. The x-axis represents the number of weathering weeks (0 = means 'pristine' -as received- polymer).

scanning electron microscopy (see **Figure SM1f** in the [supplementary material](#) and explanations at the corresponding section below).

Calculations of this index for the reflectance spectra of pellets allowed to derive analogous conclusions (**Fig. 4a**).

3.2.2. The C-O index

This index, defined as the ratio of the $A_{1000-1200}/A_{2914}$ integrated bands, is intended to characterize carboxylic acids and/or esters. Unfortunately, for PA 6.6 the region between 1000 and 1200 cm^{-1} includes several peaks from different functional groups (see discussions above), which makes it unspecific for this bond. Note that for the reflectance spectra this is not the case but the peaks seen there became weak or very weak. The overall behaviour of the index parallels the C=O one (**Fig. 3b**). Powdered PA in seawater showed a remarkable increment after the sixth weathering week (which is not seen so dramatically for the C=O group). The reflectance-derived index for pellets showed the same general behaviour (**Fig. 4b**), although without the remarkable increase after the 6th week that was seen for the ATR spectra. The overall evolution profile for the C-O index is also very similar to the OH index, discussed below.

3.2.3. The O-H index

The evolution of the O-H index (A_{3300}/A_{2914} integrated bands ratio) represents changes in the intermolecular interactions among carboxylic groups and the introduction of OH groups in the polymer. However, the ATR broad band at ca. 3300 cm^{-1} is not specific for the O-H group as the sharp band in its centre corresponds to the NH vibration and H-bonds between the C=O and NH groups of different chains when water is absent [44]. In our case, we do have aqueous/humid conditions and, so, from a practical point of view for current environmental studies, we decided to consider the overall area as a 'proxy' for the overall OH groups.

As for the C=O index, three major stages can be seen for the OH evolution during PA weathering (**Fig. 3c**). Despite the reflectance

bands ca. 3300 cm^{-1} were narrower than the ATR ones, the index had essentially the same behaviour (but for a slight increase in the reflectance index at the 8th week for pellets in seawater, which is not seen for ATR data), **Fig. 4c**. A first rise attributed to the initial degradation of the polymer, followed by a plateau and a slight final increment. This latter stage is not seen for the pellets submerged into seawater using ATR although it is clear in the reflectance spectra (**Fig. 4c**). After the sixth week, the ATR indicates that the degradation of the powdered PA in seawater is much higher than for any other assay. This may be explained because the powder had large floating times in the water surface despite the probes were agitated continuously by an air stream. Every day the powder in the surface was sent back manually to the water. The pellets, on the contrary, were at the surface marginally. Submerged pellets and powder became more degraded than their dry counterparts.

3.2.4. The crystallinity index

The index was defined as the A_{936}/A_{1640} ratio [42], although in order to use a unique reference peak, the alternative A_{936}/A_{2914} ratio is proposed in this work (it was verified that both calculations yielded the same graphical pattern). **Fig. 3d** shows the evolution of this index during weathering. The overall profile agrees very well with previous discussions for the other indexes. Crystallinity has not a uniform evolution throughout weathering. Some authors reported on its diminution with weathering, likely due to a severe chain scission after prolonged exposure of PA [31,39,40]. On the contrary, Arrieta et al. [30] informed about crystallinity increments. This discrepancy may be explained by a first increase on crystallinity due to rearrangements of the molecular fragments caused by photooxidation in the amorphous region; then a stabilization and finally a decrease, because even the crystalline phase is degraded and cleavages chains from the crystal surface [31,39,40]. This effect would be more relevant with prolonged ageing as the entanglement network reduces and only Van der Waals interactions maintain the polymer in the final stages [35]. This

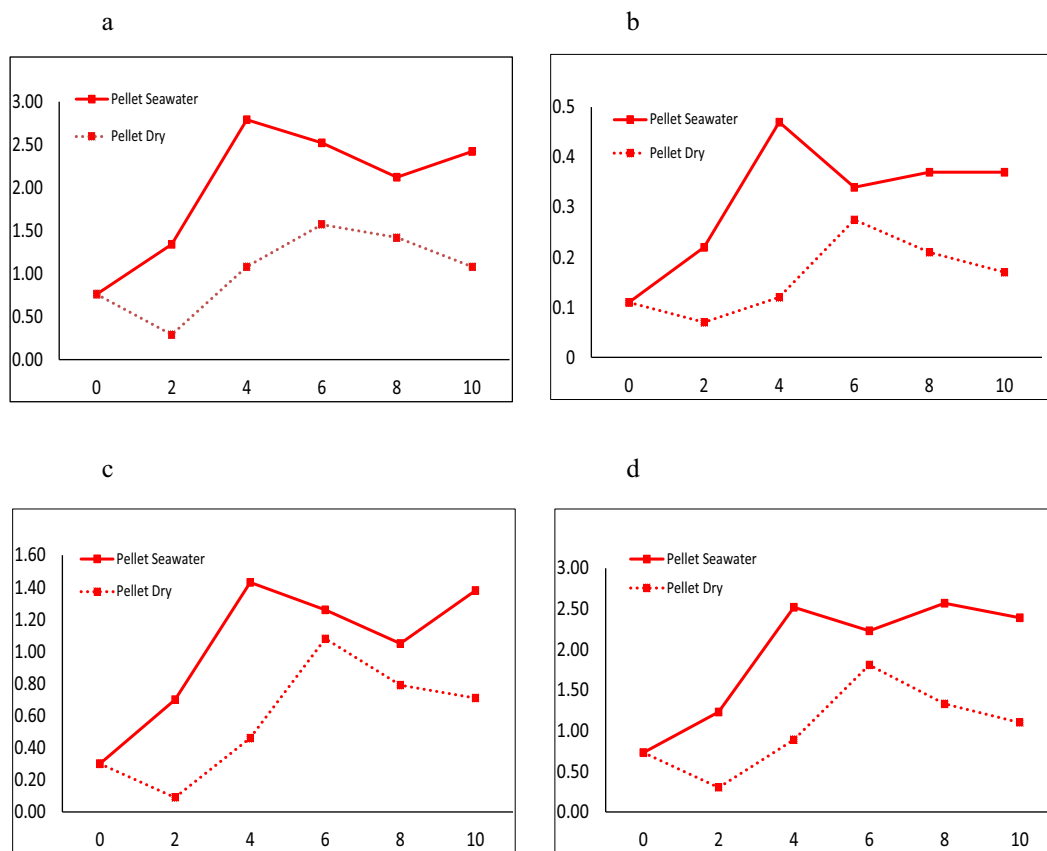


Fig. 4. Evolution of several functional indexes to monitor PA weathering using reflectance spectrometry: a) C=O; b) C-O; c) O-H; d) amide II. The x-axis represents the number of weathering weeks (0 = means 'pristine' -as received- polymer).

complex sequence in crystallinity is seen also in our experimental results:

It increases quite sharply during the first fortnight (which corresponds to a rearrangement of the scissored chains/fragments in the amorphous regions), then it reaches a plateau (but for the powder in seawater) and, finally, decreases clearly after the sixth week (ca. 1000 h of irradiation). The relevant drop in PA crystallinity for powder in seawater around four weeks can be attributed (as for the OH index) to its flotation (which means it received more irradiation than submerged pellets), combined with a release of molecular fragments from the degraded polymer to the aqueous phase. Then, a second rearrangement of the amorphous phase seems to occur (an increase on the crystallinity index) which finally gets degraded as well. Note that although both control samples remain quite stable their crystallinity increased slightly at the end, likely due to a reorganization of the amorphous phase after some hydrolysis/thermal processes.

When the reflectance spectra for pellets were considered, the index showed only a smooth, steady increasing pattern, likely because the bands involved in this index were weak and, so, scarcely sensitive to the minor variations. In particular, the drop after the 6th week was not seen, only a plateau was observed instead (figure not shown here).

3.2.5. The amide II index

This index had not been defined previously in the literature, and is proposed as the A_{1537}/A_{2914} ratio. Despite the initial ratios for

pellets and powder are different (Fig. 3e) their behaviours are similar and resemble the crystallinity index. The powdered samples show the three major steps explained for the indexes above (ATR spectra). The pellets submerged in seawater present two clear stages, considering both ATR and microreflectance (a steady increase until the sixth week, after which a degradation occurs, leading to a new reorganization after the eighth week, Fig. 4d). The pellets under dry irradiation had the same pattern as the submerged ones although less intense.

3.2.6. The Branched- and long-chains indexes

Finally, two indexes are of potential interest here. They evaluate how the amount of branched and long hydrocarbon chains evolve [47,48,49,50]. The **branched-chain index** is calculated as the $A_{1376}/A_{1450+1376}$ ratio whereas the long-chains index is the $A_{724}/A_{1450+1376}$ ratio. The former has not been too informative in this study because it showed a relative random pattern, maybe because the band at 1376 cm^{-1} is not only associated to the symmetric bending of the CH bond (in CH_3 groups) and the asymmetric bending of CH (in CH_2 units), but to other groups like the CN stretching plus the in-plane NH deformation [44].

On the contrary, the **long chain index** (which, in essence, measures the number of linearly bonded CH_2 groups) was more informative. All assays show (Fig. 3f) a decrease on the length of the CH_2 chain during the first two weeks (but for the powdered PA submerged into the seawater, which shows a steady increase). This opposes to the increment in the C=O, C-O and O-H indexes (initial

oxidative degradation in the amorphous phase, [39,40]) and it confirms that the total amount of longer fragments of the polymer ($-(CH_2)_n-$) decreases (and, so, that oxidation occurs throughout chain scission, as Arrieta *et al* [30] suggested). At the fourth week a sudden and sharp increase of the long chain index points towards the recrystallization process (which coincides with the decrease on the C=O, C-O and O-H indexes). Finally, after the eighth week, the long chain index stabilizes and it does not recover the previous values, probably because the oxidation proceeds faster. The powder weathered under dry conditions have a slightly different behaviour and after the eighth week it increases again, coinciding with the slight decrease in the C-O and O-H indexes, pointing towards a second recrystallization. These two indexes were not very informative when reflectance spectra was considered, likely because the bands involved are very weak.

3.3. Scanning electron microscopy and energy-dispersive X-ray analyses

SEM (scanning electron microscopy) studies were also done in order to characterize further the weathering processes. First, the weathering for pellets will be discussed.

Original pellets *as received* showed the typical fibrils associated to the ductile fracture occurring at the cutting stage of the polymer [31]. Some exfoliation linked to the fracture planes and a relevant void (supposedly a gas bubble) were observed (**Figure SM1, a and b, supplementary material**).

After 10 weeks of simulated sea conditions, the surface revealed clearly eroded, with a porous appearance in some parts of the pellet, cracks, scratches, wells and, of course, salt deposits (**Figure SM1, c and d**). The latter are of importance here because we visualized many sites where the surface was clearly eroded and highly affected by the presence of salt crystals. The Energy-dispersive X-ray analyses, EDXA, showed that they were mostly NaCl or CaCl₂. Indeed the surface was clearly carved and the crystals grown at those voids (see zoomed view in **Figure SM1c**).

Scratches and cracks can be explained by thermal stress and mechanical impacts but also by the internal stress caused by the water molecules entering the structure and causing hydrolysis and swelling [35]. Carving could be explained partly by the fact that some metals were reported to accelerate PA degradation (e.g., Co and Ni) [33,40]. Although those studies were done with metals included in the polymer formulations, they are indeed present in seawater and their pro-oxidation effect should not be disregarded [31]; more specific studies would be needed to clarify this.

With respect to the dry weathering setup, PA pellets showed also clear grooves, flakes (see zoomed view in **Figure SM1e**) and small scratches. However, the most dramatic visualization was a region where the polymer was totally fragmented following parallel planes (**Figure SM1f**), similar to those seen in **Figure SM1e**. Recall that PA is mostly constituted by a lamellar structure [43]. Despite it might appear a collection of salt deposits, the EXDA elemental analysis indicated that their elemental composition was dominated by C; in addition, their shape do not match the cubic appearance of salt crystals. Hence, it was concluded that it was degraded PA.

When it comes to the *as received* powdered PA the typical fibrils and fracture planes can be seen (**Figure SM2a and SM2b, supplementary material**); as well as sharp angles and well defined borders. After the 10-weeks seawater weathering the borders appeared fragmented, rounded and without the original fibrils, denoting clear erosion patterns. The fragmentation planes were blurred by erosion and many grooves. In some pictures (not shown here) small cracks and wells appeared. Noteworthy, those voids were quite often associated to included salt crystals. Their elemen-

tal composition revealed Cl, Mg, K, Ca and O; whose most probable matches are salts of the corresponding chlorides and MgO.

A number of small fragments was also seen in most degraded granules, some of them still unreleased from their main body (**Figure SM2c**). This would evidence the well-known reported formation of secondary microplastics from larger pieces. [2]. In addition, clear crazing and ploughing were seen in some granules (**Figure SM2d**).

With regard to the dry conditions, the mechanical erosion and degradation of the borders of the granules were much less relevant (as expected). However, more flakes could be observed in turn, along with crazing and voids (**Figure SM2e**) even in a stratified way (**Figure SM2f**) where some very small particles can be observed, likely due to superficial mechanical and thermal stress. The humid conditions of the room due to water evaporation from the adjacent probes containing seawater might facilitate water absorption which stresses the structure of the polymer [35].

4. Conclusions

The ageing of polyamide 6.6 was monitored under simulated natural weathering conditions. With regards to the qualitative identification of particles in environmental samples, it was found that the evolution of PA does not depend on its form (pellets or powder) nor on the experimental setup (seawater-submerged or dry conditions), but for some more pronounced evolution when the dried setup was considered, mostly for powder. A remarkable finding was that the evolution of the spectral bands do not follow a steady pattern. Instead, the spectral indexes denote cyclic evolutions, as explained in literature. The IR spectral profiles of the original and weathered PA microplastics were different and this stresses the importance of considering this evolution whenever microplastics from the environment are characterized. The simplest option is to include weathered spectra of PA in the databases driving the searches. This is of most importance whenever reflectance spectra are measured.

About the evolution of the polymer, ATR and microreflectance spectra put forward the two major photooxidative weathering pathways for PA. They show in general three stages. First, a fast increase indicates the oxidative degradation (mainly in the amorphous phase, where oxygen can diffuse fast); then a plateau probably due to a slow down on the oxidative process in the amorphous phase; finally, a continuous increase because even the crystalline phase gets degraded. It was seen that reflectance microspectrometry had less complex spectra although they were less sensitive than ATR, and this caused that only the monitoring indexes calculated from the most intense bands had the same general behaviour.

The overall large increase of the so-called Amide I band characterizes the final products of the photodegradation and hydrolysis reactions. The Amide II band rises clearly with ageing (as well as the 696 cm⁻¹ one, related to terminal NH₂ groups) and, hence, characterizes the polymer chain scission. The outstanding general increase of the broad band centred ca. 3300 cm⁻¹ points towards the introduction/appearance of hydroxyl groups in the polymer. They would correspond mostly to water absorbed in the plastic, mainly into the amorphous regions of the polymer.

The carbonyl index raises for all weathering setups with time, being higher for the simulated marine conditions. The same happens for the C-O index and, less obvious, the O-H one. The crystallinity index reinforces those conclusions as it increases quite sharply during the first fortnight (rearrangement of the scissored chains/fragments in the amorphous regions), then it reaches a plateau (but for the powder in seawater) and, finally, decreases clearly after the sixth week.

Regarding the SEM measurements, the surface of the most weathered pellets and powder grains revealed clearly eroded, with a porous appearance in some parts, cracks, scratches, wells and salt deposits. The surface was clearly carved and many crystals were seen at those voids, which might suggest that superficial salts may potentiate mechanical particle degradation. The most eroded grains revealed the formation of secondary microplastics.

CRedit authorship contribution statement

Verónica Fernández-González: Conceptualization, Methodology, Formal analysis. **Jose Manuel Andrade:** Conceptualization, Methodology, Software, Validation, Formal analysis. **Borja Ferreira:** Methodology, Software, Validation, Formal analysis. **Purificación López-Mahía:** Conceptualization, Validation. **Soledad Muniategui-Lorenzo:** Conceptualization, Validation.

Declaration of Competing Interest

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

Acknowledgements

This work was supported through the JPI-Oceans BASEMAN and MicroplastiX projects, sponsored by the Spanish Ministry of Science and Innovation (Agencia Estatal de Investigación) - partially financed by the European Regional Development Fund program- (Grants: PCIN-2015-170-C02-01; PCIN- PCI2020-112145 and CTM2016-77945-C3-3-R, ARPA-ACUA). The Program 'Consolidación e Estructuración de Unidades de Investigación Competitiva' of the Galician Government (Xunta de Galicia) is also acknowledged (Grant: ED431C-2017/28)

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.saa.2021.120162>.

References

- [1] L.M. Persson, M. Breitholtz, I.T. Cousins, C.A. de Wit, M. MacLeod, M.S. McLachlan, Confronting unknown planetary boundary threats from chemical pollution, *Environ. Sci. Technol.* 47 (22) (2013) 12619–12622, <https://doi.org/10.1021/es402501c>.
- [2] A. Jahnke, H.P.H. Arp, B.J. Escher, B. Gewert, E. Gorokhova, D. Kühnel, M. Ogonowski, A. Potthoff, C. Rummel, M. Schmitt-Jansen, E. Toorman, M. MacLeod, Reducing Uncertainty and Confronting Ignorance about the Possible Impacts of Weathering Plastic in the Marine Environment, *Environ. Sci. Tech. Lett.* 4 (2017) 85–90, <https://doi.org/10.1021/acs.estlett.7b00008>.
- [3] Communication from the commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Closing the loop - an EU action plan for the circular economy, COM (2015) 614
- [4] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European Strategy for Plastics in a Circular Economy, COM (2018) 28
- [5] Green Paper on a European Strategy on Plastic Waste in the Environment, COM (2013) 123.
- [6] www.textileschool.com/467/polyamide-fibres-manmade-artificial-fibres/; last access August 1, 2018.
- [7] N. Vasanthan, Crystallinity determination of Nylon 66 by density measurement and Fourier Transform infrared (FTIR) spectroscopy, *J. Chem. Educ.* 89 (2012) 387–390, <https://doi.org/10.1021/ed200398m>.
- [8] <https://www.azom.com/article.aspx?ArticleID=442>, last access November 15, 2020.
- [9] <https://www.plasticsinsight.com/resin-intelligence/resin-prices/polyamide/>, last access November 15, 2020.
- [10] <https://www.statista.com/statistics/649908/polyamide-fiber-production-worldwide/>, last access November 15, 2020.
- [11] <https://ihsmarkit.com/products/nylon-fibers-chemical-economics-handbook.html>, last access November 15, 2020.
- [12] M. Scheibitz, R. Kaneko, P. Spies, Polyamide 6 and 66 (PA6 and PA66), *Kunststoffe International* 10 (2016) 40–44.
- [13] E.A. Gies, J.L. LeNoble, M. Noël, A. Etamadifar, F. Bishay, E.R. Hall, P.S. Ross, Retention of microplastics in a major secondary wastewater treatment plant in Vancouver, Canada, *Mar. Poll. Bull.* 133 (2018) 553–561, <https://doi.org/10.1016/j.marpolbul.2018.06.006>.
- [14] I.E. Napper, R.C. Thompson, Release of synthetic microplastic plastic fibres from domestic washing machines: effects of fabric type and washing conditions, *Mar. Pollut. Bull.* 112 (2016) 39–45, <https://doi.org/10.1016/j.marpolbul.2016.09.025>.
- [15] V.M. León, I. García-Aguiera, V. Moltó, V. Fernández-González, L. Llorca-Pérez, José M. Andrade, S. Muniategui-Lorenzo, J.A. Campillo, PAHs, pesticides, personal care products and plastic additives in plastic debris from Spanish Mediterranean beaches. *Science of the Total Environment* 670 (2019), pp 672–684. Doi: 10.1016/j.scitotenv.2019.03.216
- [16] M. Yan, H. Nie, K. Xu, Y. He, Y. Hu, Y. Huang, J. Wang, Microplastic abundance, distribution and composition in the Pearl River along Guangzhou city and Pearl River estuary, China, *Chemosphere* 217 (2019) 879–886, <https://doi.org/10.1016/j.chemosphere.2018.11.093>.
- [17] S. Morgana, L. Ghigliotti, N. Estévez-Calvar, R. Stifanese, A. Wieczorek, T. Doyle, J.S. Christiansen, M. Faimali, F. Garaventa, Microplastics in the Arctic: A case study with sub-surface water and fish samples off Northeast Greenland, *Environmental Pollution* 242 (2018) 1078–1086, <https://doi.org/10.1016/j.envpol.2018.08.001>.
- [18] C. Lefebvre, C. Saraux, O. Heitz, A. Nowaczyk, D. Bonnet, Microplastics FTIR characterisation and distribution in the water column and digestive tracts of small pelagic fish in the Gulf of Lions, *Mar. Poll. Bull.* 142 (2019) 510–519, <https://doi.org/10.1016/j.marpolbul.2019.03.025>.
- [19] M.S. Hossain, M.S. Rahman, M.N. Uddin, S.M. Sharifuzzaman, S.R. Chowdhury, S. Sarker, M.S.N. Chowdhury, Microplastic contamination in Penaeid shrimp from the Northern Bay of Bengal, *Chemosphere* 238 (2020), <https://doi.org/10.1016/j.chemosphere.2019.124688>.
- [20] Y. Huang, M. Yan, K. Xu, H. Nie, H. Gong, J. Wang, Distribution characteristics of microplastics in Zhubi Reef from South China Sea, *Environ. Poll.* 255 (2019), <https://doi.org/10.1016/j.envpol.2019.113133>.
- [21] A.V. Filgueiras, J. Gago, J.A. Campillo, V.M. León, Microplastic distribution in surface sediments along the Spanish Mediterranean continental shelf, *Environ. Sci. Poll. Res.* 26 (21) (2019) 21264–21273, <https://doi.org/10.1007/s11356-019-05341-5>.
- [22] M.B. Zobkov, E.E. Esiukova, A.Y. Zyubin, I.G. Samusev, Microplastic content variation in water column: The observations employing a novel sampling tool in stratified Baltic Sea, *Mar. Poll. Bull.* 138 (2019) 193–205, <https://doi.org/10.1016/j.marpolbul.2018.11.047>.
- [23] N. Sathish, K.I. Jeyasanta, J. Patterson, Abundance, characteristics and surface degradation features of microplastics in beach sediments of five coastal areas in Tamil Nadu, India, *Mar. Poll. Bull.* 142 (2019) 112–118, <https://doi.org/10.1016/j.marpolbul.2019.03.037>.
- [24] A.L. Andrade, Microplastics in the marine environment, *Mar. Poll. Bull.* 62 (2011) 1596–1603, <https://doi.org/10.1016/j.marpolbul.2011.05.030>.
- [25] A.B. Silva, A.S. Bastos, C.I.L. Justino, J.P. da Costa, A.C. Duarte, T.A.P. Rocha-Santos, Microplastics in the environment: challenges in analytical chemistry, a review, *Anal. Chim. Acta.* 1017 (2018) 1–19, <https://doi.org/10.1016/j.aca.2018.02.043>.
- [26] A. Karami, A. Golieskardi, Y.B. Ho, V. Larat, B. Salamatinia, Microplastics in eviscerated flesh and excised organs of dried fish, *Sci. Rep.* 7 (2017) 5473, <https://doi.org/10.1038/s41598-017-05828-6> (9 pages).
- [27] P. Ribeiro-Claro, M.M. Nolasco, C. Araújo, Chapter 5-characterization of microplastics by Raman spectroscopy, in *Comprehensive Analytical Chemistry Handbook*, Vol 75 (Characterization and analysis of microplastics). Elsevier, 2017
- [28] A. ter-Halle, L. Ladirat, M. Martignac, A.F. Mingotaud, O. Boyron, E. Perez, To what extent are microplastics from the open ocean weathered?, *Environ. Poll.* 227 (2017) 167–174, <https://doi.org/10.1016/j.envpol.2017.04.051>.
- [29] J. Andrade, V. Fernández-González, P. López-Mahía, S. Muniategui, A low-cost system to simulate environmental microplastic weathering, *Mar. Poll. Bull.* 149 (2019), <https://doi.org/10.1016/j.marpolbul.2019.110663>.
- [30] C. Arrieta, Y. Dong, A. Lan, T. Vu-Khanh, Outdoor weathering of polyamide and polyester ropes used in fall arrest equipment, *J. Appl. Polym. Sci.* 130 (5) (2013) 3058–3064, <https://doi.org/10.1002/app.39524>.
- [31] P. Cerruti, M. Lavorgna, C. Carfagna, L. Nicolais, Comparison of photo-oxidative degradation of polyamide 6,6 films stabilized with HALS and CuCl₂+KI mixtures, *Polym.* 46 (2005) 4571–4583, <https://doi.org/10.1016/j.polymer.2005.03.065>.
- [32] P. Gijssman, G. Meijers, G. Vitarelli, Comparison of the UV-degradation chemistry of polypropylene, polyethylene, polyamide 6 and polybutylene terephthalate, *Polym. Degrad. Stab.* 65 (1999) 433–441, [https://doi.org/10.1016/S0141-3910\(99\)00033-6](https://doi.org/10.1016/S0141-3910(99)00033-6).
- [33] E. Dümichen, U. Braun, R. Kraemer, P. Degelmann, R. Senz, Thermal extraction combined with thermal desorption: a powerful tool to investigate the thermo-oxidative degradation of polyamide 66 materials, *J. Anal. Appl. Pyrol.* 115 (2015) 288–298, <https://doi.org/10.1016/j.jaap.2015.08.006>.
- [34] X. Zhao, X. Li, L. Ye, G. Li, Stress-accelerated photothermal oxidative aging behaviour of polyamide 6,6, *J. Thermoplast. Compos. Mater.* 27 (11) (2014) 1573–1586, <https://doi.org/10.1177/0892705713495437>.

- [35] I. Ksouri, O. De Almeida, N. Haddar, Long term ageing of polyamide 6 and polyamide 6 reinforced with 30 % of glass fibers: physicochemical, mechanical and morphological characterization, *J. Polym. Res.* 24 (8) (2017) 133–145, <https://doi.org/10.1007/s10965-017-1292-6>.
- [36] J. Morcillo, R. Madroño. Aplicaciones prácticas de la espectroscopia infrarroja. Edited by Universidad de Madrid, Madrid, 1962.
- [37] R.T. Conley, *Espectroscopia infrarroja*, Edit. Alhambra, Madrid, 1979.
- [38] C. El-Mazry, M.B. Hassine, O. Correc, X. Colin, Thermal oxidation kinetics of additive free polyamide 6–6, *Polym. Degrad. Stabil.* 98 (2013) 22–36, <https://doi.org/10.1016/j.polydegradstab.2012.11.002>.
- [39] A. Bienik, B. Lipp-Symonowicz, S. Sztajnowski, Influence of the structure of polyamide 6 fibers on their ageing under intensive insolation conditions, *Polimery* 54 (11–12) (2009) 840–844, <https://doi.org/10.14314/polimery.2009.840>.
- [40] P. Cerruti, C. Carfagna, Thermal-oxidative degradation of polyamide 6.6 containing metal salts, *Polym. Degrad. Stabil.* 95 (2010) 2405–2412, <https://doi.org/10.1016/j.polydegradstab.2010.08.014>.
- [41] R.D. Goodridge, R.J.M. Hague, C.J. Tuck, Effect of long-term ageing on the tensile properties of a polyamide 12 laser sintering material, *Polym. Test.* 29 (2010) 483–493, <https://doi.org/10.1016/j.polymertesting.2010.02.009>.
- [42] N. Vasanthan, D.R. Salem, Infrared spectroscopic characterization of oriented polyamide 66: band assignment and crystallinity measurement, *J. Polym. Sci. (Part B)* 38 (2000) 516–524, [https://doi.org/10.1002/\(SICI\)1099-0488\(20000215\)38:4<516::AID-POLB3>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-0488(20000215)38:4<516::AID-POLB3>3.0.CO;2-Y).
- [43] Y. Li, W.A. Goddard III, Nylon 6 Crystal Structures, Folds, and Lamellae from Theory, *Macromolecules* 35 (2002) 8440–8455, <https://doi.org/10.1021/ma020815n>.
- [44] N.K. Pramanik, M.S. Alam, R.K. Khandal, Electron beam irradiation of Nylon 66: characterization by IR spectroscopy and viscosity studies, *Int. J. Innovat. Res. Sci. Engineer. Tech.* 4 (1) (2015) 18547–18555, <https://doi.org/10.15680/IJRSET.2015.0401019>.
- [45] K. Rajakumar, V. Sarasvathy, A. Thamarai Chelvan, R. Chitra, C.T. Vijayakumar, Natural Weathering Studies of Polypropylene, *J. Polym. Environ.* 17 (2009) 191–202, <https://doi.org/10.1007/s10924-009-0138-7>.
- [46] J. Brandon, M. Goldstein, M.D. Ohman, Long-term aging and degradation of microplastic particles: Comparing in situ oceanic and experimental weathering patterns, *Mar. Poll. Bull.* 110 (2016) 299–308, <https://doi.org/10.1016/j.marpolbul.2016.06.048>.
- [47] O. Abbas, N. Dupuy, C. Rebufa, L. Vrielynck, J. Kister, A. Permanyer, Prediction of source rock origin by chemometric analysis of Fourier Transform infrared-attenuated total reflectance spectra of oil petroleum: evaluation of aliphatic and aromatic fractions by self-modeling mixture analysis, *Appl. Spectrosc.* 60 (3) (2006) 304–314, <https://doi.org/10.1366/000370206776342508>.
- [48] A. Permanyer, L. Douifi, A. Lahcini, J. Lamontagne, J. Kister, FTIR and SUVF spectroscopy applied to reservoir compartmentalization: a comparative study with gas chromatography fingerprints results, *Fuel* 81 (2002) 861–866, [https://doi.org/10.1016/S0016-2361\(01\)00211-3](https://doi.org/10.1016/S0016-2361(01)00211-3).
- [49] P. Fresco Rivera, R. Fernández Varela, M.P. Gómez Carracedo, F. Ramírez Villalobos, D. Prada, S. Muniategui, J.M. Andrade, Development of a fast analytical tool to identify oil spillages employing infrared spectral indexes and pattern recognition techniques, *Talanta* 74 (2007) 163–175, <https://doi.org/10.1016/j.talanta.2007.05.047>.
- [50] J.M. Andrade, P. Fresco, S. Muniategui, D. Prada, Comparison of oil spillages using mid-IR indexes and 3-way Procrustes rotation, matrix-augmented principal components analysis and parallel factor analysis, *Talanta* 77 (2008) 863–869, <https://doi.org/10.1016/j.talanta.2008.07.044>.

1 **MONITORIZATION OF POLYAMIDE MICROPLASTICS**
2 **WEATHERING USING ATTENUATED TOTAL REFLECTANCE**
3 **AND MICROREFLECTANCE INFRARED SPECTROMETRY**

4

5 Verónica Fernández-González; Jose Manuel Andrade (*); Borja Ferreiro; Purificación
6 López-Mahía; Soledad Muniategui-Lorenzo

7 Group of Applied Analytical Chemistry (QANAP) and Instituto Universitario de Medio
8 Ambiente (IUMA), University of A Coruña, E-15071, A Coruña, Spain.

9 Corresponding author: andrade@udc.es; fax: +34-981167065

10

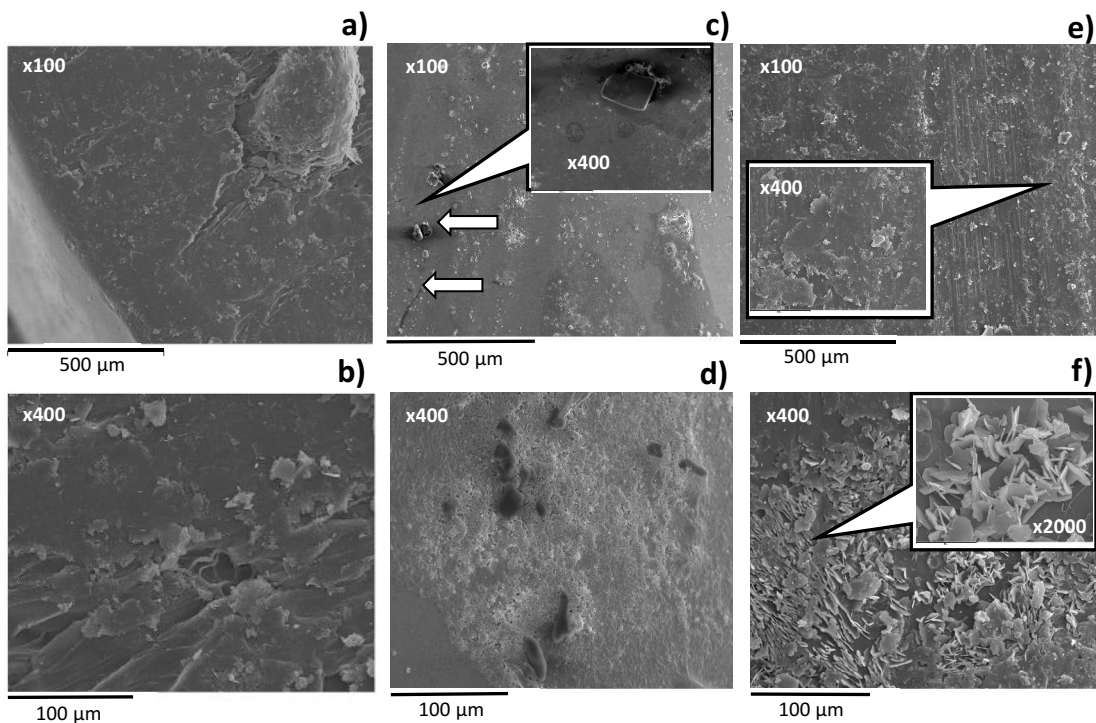
11 **SUPPLEMENTARY MATERIAL**

12

13 This Supplementary Material is composed of two figures:

14

15



16

17 **Figure SM1:** SEM microphotographs showing the evolution of weathered pellets.

18 Original pellets (a and b), 10-weeks seawater weathering (c and d) and pellets weathered

19 under 10-weeks dry conditions (e and f). See text for explanations.

20

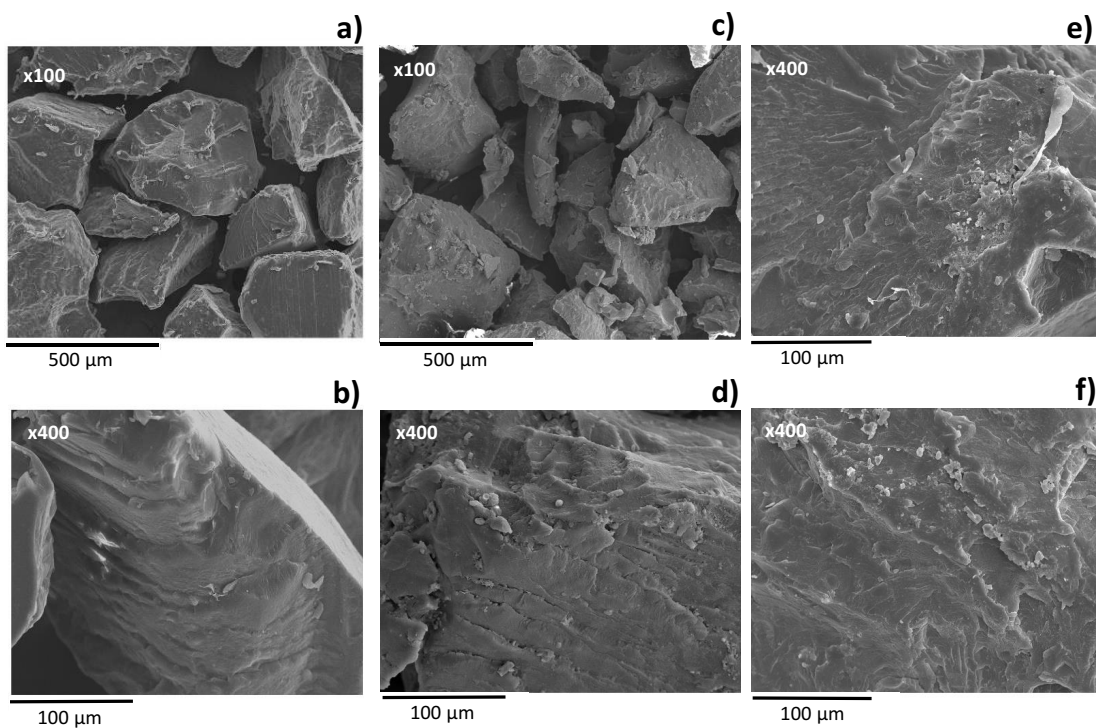
21

22

23

24

25



26

27

28 **Figure SM2:** SEM microphotographs showing the evolution of weathered PA powder.
29 Original grains (a and b), 10-weeks seawater weathering (c and d) and grains weathered
30 under 10-weeks dry conditions (e and f). See text for explanations.

**Chapter 7: Weathering-
independent Differentiation of
Microplastic Polymers by
Reflectance IR Spectrometry
and Pattern Recognition.**





Weathering-independent differentiation of microplastic polymers by reflectance IR spectrometry and pattern recognition

Borja Ferreiro, Jose M. Andrade^{*}, Carlota Paz-Quintáns, Verónica Fernández-González, Purificación López-Mahía, Soledad Muniategui

Grupo Química Analítica Aplicada (QANAP), Instituto Universitario de Medio Ambiente (IUMA), Universidade da Coruña, 15071 A Coruña, Spain

ARTICLE INFO

Keywords:

Microplastics
Infrared spectrometry
Reflectance
Weathering
Pattern recognition
Variable selection

ABSTRACT

The presence and effects of microplastics in the environment is being continuously studied, so the need for a reliable approach to ascertain the polymer/s constituting them has increased. To characterize them, infrared (IR) spectrometry is commonly applied, either reflectance or attenuated total reflectance (ATR). A common problem when considering field samples is their weathering and biofouling, which modify their spectra. Hence, relying on spectral matching between the unknown spectrum and spectral databases is largely defective. In this paper, the use of IR spectra combined with pattern recognition techniques (principal components analysis, classification and regression trees and support vector classification) is explored first time to identify a collection of typical polymers regardless of their ageing. Results show that it is possible to identify them using a reduced suite of spectral wavenumbers with coherent chemical meaning. The models were validated using two datasets containing artificially weathered polymers and field samples.

1. Introduction

Plastics constitute a durable, lightweight, and versatile family of materials from which an immense variety of products are created. Their applications range from food packaging to sports, electronics, construction and transport. Currently, plastics are required in so many industrial fields that they became an indispensable material for many commodities and complex products (cars, planes, etc.), with their global production reaching 367 million tonnes in 2020, of which Europe accounted for 55 million tonnes (Mt) (PlasticEurope, 2021). Polypropylene and polyethylene are the most demanded polymers (19.7 % PP, 17.4 % LDPE, 12.9 % HDPE), followed by PVC (9.6 %) and PET (8.4 %), mostly for packaging. China accounts for the highest plastic consumption and production, with up to 32 % of the World's plastic production in 2020 (PlasticEurope, 2021).

A consequence of the massive use and inadequate recovery and recycling of plastics is that they have become one of the most ubiquitous anthropogenic contaminants in the World's environments. They have been found in soil (Bläsing and Amelung, 2018), airborne particles (Prata, 2018), water (Koelmans et al., 2019) and food (Van Cauwenberghe and Janssen, 2014). However, the problem is specially serious on aquatic environments (Jiang, 2018). It was estimated that in

2018 alone 5.5 to 14.5 Mt entered the oceans (Wayman and Niemann, 2021) and, as a consequence, seas and oceans have plastic debris both at their surface and seabed worldwide. There are many regions affected by this kind of pollution, even in deep sea trenches (Chiba et al., 2018). The most affected zone is the North Pacific area (Howell et al., 2012) as oceanic currents drag the plastic debris to this zone, generating plastic 'fields' swirling in the oceanic surface. This has been demonstrated to have a detrimental effect on different species of marine flora and fauna (Coffin et al., 2019; Lamb et al., 2018; Markic et al., 2020; Wang et al., 2016; Wilcox et al., 2018; Young et al., 2009).

Microplastics (MPs) can be defined as 'any solid plastic particle insoluble in water with any dimension between 1 μ m and 1 mm', being larger particles –between 1 mm and 5 mm- 'large microplastics' (ISO, 2020). The smaller sizes in particular, along with nanoplastics, can enter the trophic chain through plankton (Botterell et al., 2019), fish (Bellas et al., 2016) and other marine species by ingestion (Jiang, 2018), and it is common to find microplastics in the digestive system of different aquatic species (Bellas et al., 2016; Compa et al., 2018). This fact, along with an improved capacity to adsorb other contaminants in comparison with meso- and macroplastics, is known to have detrimental effects (Gewert et al., 2015), which might affect human health through ingestion of –as a matter of example- shellfish (Van Cauwenberghe and Janssen, 2014),

^{*} Corresponding author.

E-mail address: andrade@udc.es (J.M. Andrade).

<https://doi.org/10.1016/j.marpolbul.2022.113897>

Received 12 April 2022; Received in revised form 23 June 2022; Accepted 25 June 2022

Available online 7 July 2022

0025-326X/© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

fish (Battaglia et al., 2016; Bellas et al., 2016) and table salt (Iniguez et al., 2017). In fact, in recent studies, plastics have been found in human blood and lung tissue (Jenner et al., 2022; Leslie et al., 2022).

Therefore, a successful identification of the polymers found in the environment is of utmost importance to evaluate their distribution, origin, or subsequent behaviour there. This can be done using different analytical techniques, a common one being vibrational spectrometry. Its speed, high selectivity, and low demand of sample quantity make it a suitable technique for MPs analysis. Besides, it is easily combined with microscopic techniques and attenuated total reflectance (ATR) accessories. Although ATR spectrometry is well suited for measuring big and medium particles (let's say $>500 \mu\text{m}$) it is not suited for smaller ones. Here it is where reflectance gains momentum and overcome the ATR limitations as it does not need direct contact between the sample and the IR focusing device. Two nice reports on the use and limitations of various types of IR spectrometry are those from Veerasingam et al. (2021) –that includes some interpretation of spectral bands– and Primpke et al. (2020) –that included a comparison with other analytical methodologies–.

Respectively of the analytical measuring technique, there are many physical, biological and chemical processes that plastics undergo while in the environment (e.g., mechanical erosion, photodegradation due to the UV light and biological colonization). They can modify the original polymers by breaking their molecules into smaller ones, oxidizing their chains (usually due to chain attack by UV-originated radicals), giving rise to/modifying oxygenated functional groups (carboxylic acids, ketones, peroxides, etc.), crosslinking alterations, changing crystallinity, etc. Many times this leads to smaller plastic fragments. All these effects are known collectively as ‘weathering’ and more details can be found elsewhere (Göpferich, 1996; Gewert et al., 2015; Raddadi and Fava, 2019; Chamas et al., 2020; Ali et al., 2021; Zhang et al., 2021). The final consequence is that weathering alters the surface of the (micro)plastics and, so, their spectra. Many IR spectral bands evolve with weathering, some others appear, others broaden and/or overlap with neighbouring signals, etc. All this hinders the interpretation of the spectra. Many practitioners rely on spectral matching between the unknown spectrum and spectral databases to account for polymer identification. However, often the latter cannot properly match the former due to the aforementioned problems, as reported frequently; e.g., Fernández-González et al. (2021a) and Mecozzi et al. (2016).

A complete relation of the evolution of each of the polymers used in this work is out of the scope of this paper and only some general details will be given. Interested readers are kindly forwarded to the references cited next, and those referred to therein. The most relevant changes observed in the spectra were compatible with damages caused by the UV radiation that gave rise to chain attacks by radicals (Norris-type reactions). A general introduction to those modifications was presented by Gewert et al. (2015). Specific changes on the spectra and surface of the five most common packaging polymers –LDPE, HDPE, PS, PP and PET– and PA6.6 were studied previously (Fernández-González et al., 2021a, 2021b). In the former paper, a table resumed which bands increased or decreased with time, per plastic. Also, dramatic changes on the polymeric structure (leakage of Cl atoms and appearance of C=C bonds) were reported for weathered PVC particles (Fernández-González et al., 2022), which might explain the usually low reports of PVC microplastics in environmental samples. Particular details can be found elsewhere for PET (Gok et al., 2019; Oreski and Wallner, 2005; Renner et al., 2017; Venkatachalam et al., 2012), LDPE (Brandon et al., 2016; Hirsch et al., 2017; Luo et al., 2020), PC (Shi et al., 2021), PP (Brandon et al., 2016; ter Halle et al., 2017), and PS (Yousif and Haddad, 2013).

A natural way to reduce this problem is to include spectra of weathered polymers into the databases (Fernández-González et al., 2021a). In this sense, a low-cost weathering system was proposed recently to resemble natural conditions (Andrade et al., 2019) and standardize this task. The need for complete databases and the risk of relying on current correlation coefficients to identify unknown particles

were presented by Mecozzi et al. (2016) and they proposed a suite of three similarity indexes and the use of independent components analysis to search the database.

Another novel and totally different route to mitigate this problem is explored in this work using unsupervised chemometric pattern recognition. The main objective of this paper is to explore a new way to get rid of the spectral information related to the weathering processes of the materials constituting the MPs. If so, the remaining spectral characteristics, which in essence will not be affected by weathering, would simplify the identification of the polymers and, therefore, open up new possibilities for MPs studies and environmental monitoring. It is worth noting that this working hypothesis does not correspond to typical pattern recognition studies where the most important patterns (linked to the first statistical factors that explain most of the variance into the spectral data) are used to, precisely, visualize the sets of samples. Here we look in the other way around; i.e., how the influence of weathering on the spectra – which constitute their principal source of variation – can be avoided and, so, weathered and unweathered specimens of a polymer appear close together after a statistical study (e.g. in a given plot of the samples). As a referee pointed out, a similar though conceptually different approach would be to look for parameters not evolving with time, but such an approach was not considered in this work.

2. Experimental

2.1. Samples

The polymers used throughout this study were provided by the Universität of Bayreuth (Germany), within the framework of the JPI-Oceans-funded BASEMAN project. They were fabricated with the lowest possible amount of additives. Two sample forms were studied: powder (average size ca. $300 \mu\text{m}$) and pellets (average size ca. 3mm), more details can be found in the Supplementary material.

Small quantities (10–20 g) of all polymers were aged for 10 weeks in a dedicated ad-hoc weathering system designed for standardizing the weathering of MPs at geographical medium latitude (Andrade et al., 2019). Aliquots were either submerged in seawater to simulate weathering in the superficial oceanic layer or kept dry to simulate weathering at the shoreline (e.g., upper part of beaches and dunes). An aliquot of each polymer was withdrawn weekly from each weathering container and measured by ATR and micro reflectance infrared spectrometry. The experimental weathering conditions and related details can be found elsewhere (Andrade et al., 2019). Fig. SM1 (Supplementary material) exemplifies the reflectance spectra obtained for the pelletized and powdered samples.

In the present work, the pristine (*as received*) and ten aliquots withdrawn from the weathering system weekly were considered. Half the samples, the odd ones, were employed to calibrate the models whereas the even were used for validation. In addition a second field dataset composed of 67 field plastic fragments collected from three Mediterranean beaches studied previously (León et al., 2019) was used to test the models under field monitoring conditions.

2.2. Equipment

A Spectrum 400 FT-IR Perkin-Elmer spectrometer coupled with a Perkin Elmer IR Spotlight 200i microscope and a horizontal one-bounce diamond ATR (Miracle, Pike, USA) were employed. Each particle was measured twice (changing the position of the pellet) and the resulting spectra were averaged. The MIPIR (minimum information for publication of IR-related data on MPs characterization (Andrade et al., 2020)) experimental setup was: resolution: 4cm^{-1} ; number of scans: 200; spectral range: $3500\text{--}600 \text{cm}^{-1}$; background recording before measuring each particle; aperture: $100 \mu\text{m}$ (adjusted to smaller apertures whenever the granule was smaller); apodization: Beer-Norton, strong; spectral processing: multipoint baseline correction plus

normalization 10 % plus Kubelka-Munk (for pellets), and multipoint baseline correction plus Kramers-Kronig (for powders). ATR spectra were corrected for depth penetration. The sample form affects the required spectral processing due to the different interaction the surface of the granule has with the IR beam. For pellets the surface selected for measurements was mostly flat, making the reflection mostly specular. For powders, however, the reflection is mostly diffuse, due to the irregularities of the powder grains.

Reflectance instead of transmittance was considered because the former is very common to deal with MPs characterization; mainly, attenuated total reflectance, ATR, but also specular and diffuse reflectance (Fernández-González et al., 2021b). Reflectance spectra do not depend in essence on particle thickness and can be applied straightforwardly. On the other hand, as reflectance characterizes the surface of the particles it is affected by polymer ageing and, so, the studies presented in this paper are worth of interest for routine use.

2.3. Software

The multivariate statistical software employed throughout was GenEx7 (MultiD Analysis AB, Göteborg, Sweden), Matlab's PLS Toolbox (Eigenvector Co, USA) and Orange (University of Ljubljana, Slovenia).

2.4. Chemometric tools

The evolution of IR spectra during weathering makes this source of variance (statistical information) dominate the distribution of the samples in multivariate analyses. This difficult or impedes the correct assignation of a material to a type of polymer. Such an information is what it is expected to avoid here.

The three chemometric techniques applied in this paper are depicted briefly in the Supplementary material. They are principal components analysis (PCA), classification and regression trees (CART) and support vectors machine (SVM). They correspond to standard algorithms available in any chemometric statistical package, no implementation changes have been done. Therefore, the explanations refer to the conceptual bases of the techniques, without going into mathematical details (which are available in the references given therein). Non skilled readers are kindly encouraged to review the Supplementary material for the basic meaning of the terms scores and loadings as they are used throughout the next sections. They also constitute the basis of two variable selection approaches.

The major output of a PCA study is a set of combinations of variables, each of which is a principal component, PC (see Supplementary material for a simplified explanation on how they are calculated and interpreted from a chemical viewpoint). To avoid confusion with polycarbonate (whose acronym is also PC) references to the principal components will be always associated to an ordinal (i.e., the number of the principal component under discussion; e.g. PC4), or to the plural form, PCs.

SVM was performed using a reduced set of variables derived from the most relevant PCA loadings. In particular, those from the PCs that differentiated the most among the polymers. The data pretreatment for the SVM models that used micro reflectance spectra was the same as for the PCA model from which the variables were selected (automatic baseline correction plus first derivative). The data pretreatment for SVM models developed with ATR spectra required baseline correction (automatic weighted least squares, 2nd order), normalization (area = 1), and standard normal variate (SNV).

For the CART models no pretreatment was applied, but the spectra were reduced to the 1800–600 cm^{-1} range, in order to avoid variables without real information (atmospheric CO_2 peaks, baseline and noise) and to reduce the computational burden.

To evaluate the performance of the models a series of straightforward statistics can be calculated: the *ratio of false positives* (calculated as (false positives) / (true negatives + false positives)); and the *ratio of false negatives*, calculated as (false negatives) / (true positives + false

negatives). Also, the Mathews' Correlation coefficient (MCC) was calculated to accurately summarize the behaviour of the models (Cuadros-Rodríguez et al., 2016), following the equation below:

$$MCC = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)}}$$

where TP is the number of true positives in the sample, TN the number of true negatives, FP the number of false positives and FN the number of false negatives. A perfect model would yield a MCC value of 1. In all models, the criterion used to assign a sample to a group was the 'most probable' (i.e., the sample is included in the group to which it shows the highest probability).

3. Results and discussions

To simplify the discussions and comparisons among the studies, the results for the pelletized form of the polymers are presented first for seawater and second for dry conditions, for both the micro reflectance and ATR measurements, each. Then, those for the powder form will be given in the same order. The results are presented in the following order: 1st, the PCA study (scores and loadings), with the identification and interpretation of the most relevant loadings; 2nd, dynamic PCA; 3rd, PCA using only the variables associated to the most relevant loadings; 4th, SVM model; and 5th, CART model.

The chemical interpretation of the loadings is presented graphically for the sake of simplicity and to avoid repetitive discussions. Specific details on the interpretation of the IR spectra for the different polymers and their functional groups can be found in some previous works; for example, those from Arrieta et al. (2013), Brandon et al. (2016), Mecozzi et al. (2016) -although they focused on the most relevant bands to identify the polymers-, Tiwari et al. (2019), Vasanthan (2012), Veerasingham et al. (2021) and the exhaustive recopilation of Jung et al. (2018).

In addition to the spectral processing detailed in Section 2.2, the data pretreatment for the PCA was: automatic baseline correction (using a second order polynomial function) plus normalization (total area = 1) plus mean centring. The pretreatment for powders measured by micro reflectance was: automatic baseline correction plus first derivative. These treatments were selected after several preliminary trials as they yielded the best groups of samples.

3.1. Pellets weathered in seawater

The PCA carried out on the micro reflectance spectra of the pellets weathered in seawater conditions revealed that they could be differentiated nicely using the PC1-PC4-PC6 subspace (56.2 % explained variance), see Fig. 1a. PC2, PC3 and PC5 did not differentiate them because even though they could group some polymers the others led to widely dispersed scores (likely, due to weathering), yielding overlaps between the groups and impeding their separation.

PC1 (39.5 % explained information) in essence opposes PC, PET and PMMA (negative scores) to the other polymers (with positive or close-to-zero scores), the former being the polymers with the most complex chemical structures. Fig. 2 shows that the variables contributing more to this factor are clearly associated to HDPE and LDPE, with positive loadings, and also related with highest positive scores in PC1. The most important negative loadings characterize PC, PET and PMMA.

PC4 (11.2 % explained variance) separates PA (negative scores) and PS (positive scores, Fig. 1b). The loadings are mostly related to PA (Fig. 2, negative loadings) but for a band at 698 cm^{-1} , which can be attributed mainly to PS due to its aromatic nature and positive loading.

PC6 (5.5 % explained information) separates basically PET (characterized by the most negative scores, Fig. 1b; and negative loadings, Fig. 2) from PP plus PMMA (with positive scores, Fig. 1b, and positive loadings associated to the paraffinic characteristics of PP and PMMA,

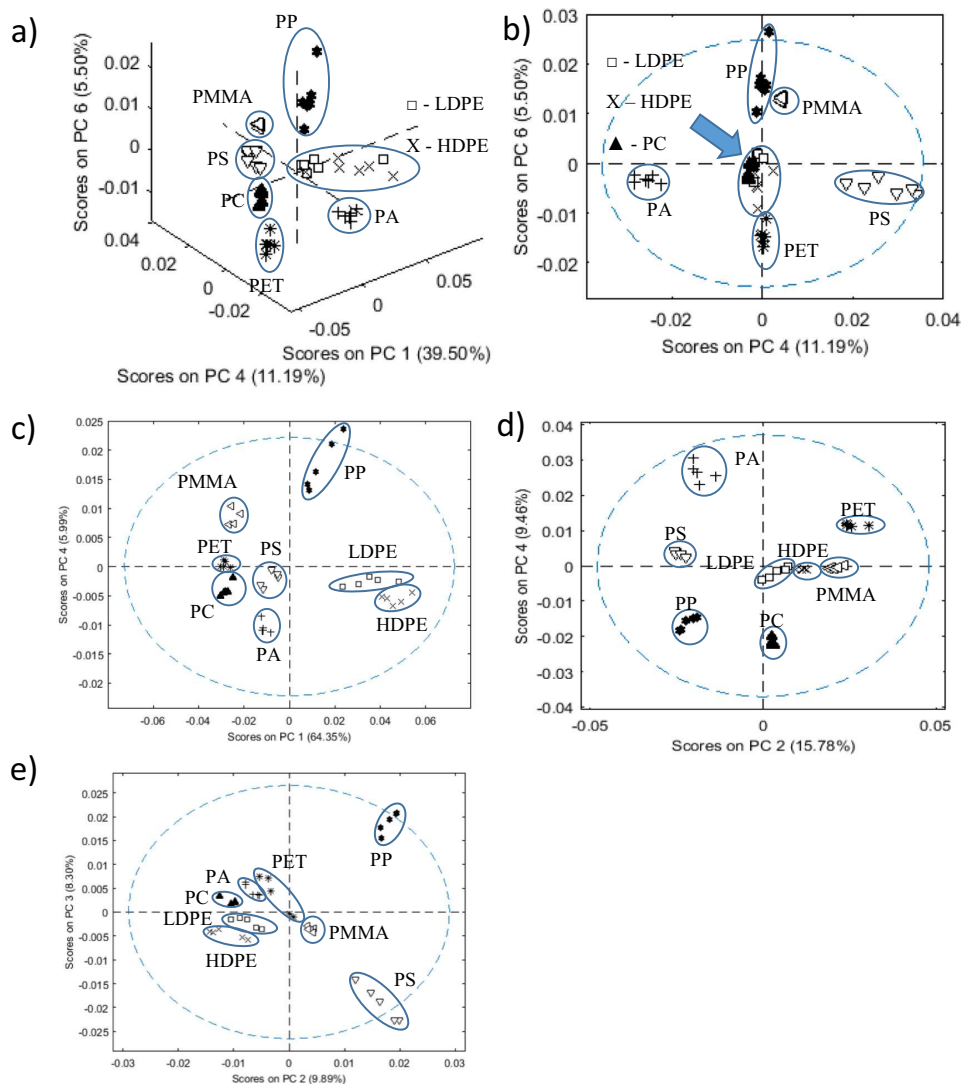


Fig. 1. Representation of the most relevant principal component subspaces to identify the groups of polymers. Seawater-aged pellets measured by micro reflectance (a) and (b), and by ATR (c). Dry-aged pellets measured by micro reflectance (d) and by ATR (e).

Fig. 2).

Using dynamic PCA it was possible to reduce the overall number of wavenumbers from 3400 to 10 and they still allowed visualizing the groups of polymers: 3455, 3024, 2918, 1775, 1631, 1268, 1262, 1230, 1156, and 706 cm^{-1} . All the variables are coherent with those illustrated in the loadings figures (Fig. 2, and Supplementary material), with the exception of 3455 and 3024 cm^{-1} , that are associated with the humidity of the samples (Brandon et al., 2016).

When the variables associated with the maxima of the loadings were used (only 11 wavenumbers corresponding to the peaks labelled in Fig. 2) to develop a dedicated PCA the groups of polymers could also be differentiated. SVM models considering these wavenumbers yielded satisfactory classifications, with only partially erroneous models when considering HDPE and LDPE. Fig. 3a exemplifies the separation of HDPE in a model, where the similarity between HDPE and LDPE is clear (indeed, they have minor spectral differences). Fig. 3b exemplifies the optimization of the cost and gamma parameters required to get the SVM

model.

Finally, CART studies required only 7 divisions (14 branches) to successfully separate the polymers (Fig. 4a). This implied that only 7 variables were required (i.e., 1797, 1749, 1663, 1602, 1512, 1475 and 1268 cm^{-1}), whose interpretation agrees quite well with the previous relevant vector loadings (the C—O stretching, N—H bending plus C—N stretching, and the C=O typical functional groups can be seen rather clearly here).

When the ATR spectra were considered, the polymers could be differentiated using just two PCs: PC1 and PC4 (70.34 % explained variance, Fig. 1c). PC2 and PC3 could not separate HDPE and LDPE. In this study the 4000 to 3000 cm^{-1} region (mostly attributed to water) was excluded, as well as a PMMA sample, due to its anomalous behaviour. It is worth noting that PC1 and PC4 were also required when using the micro reflectance dataset above.

PC1 (64.35 % explained information) in essence opposes the most 'simple' polymers HDPE, LDPE and PP (positive scores, Fig. 1c,

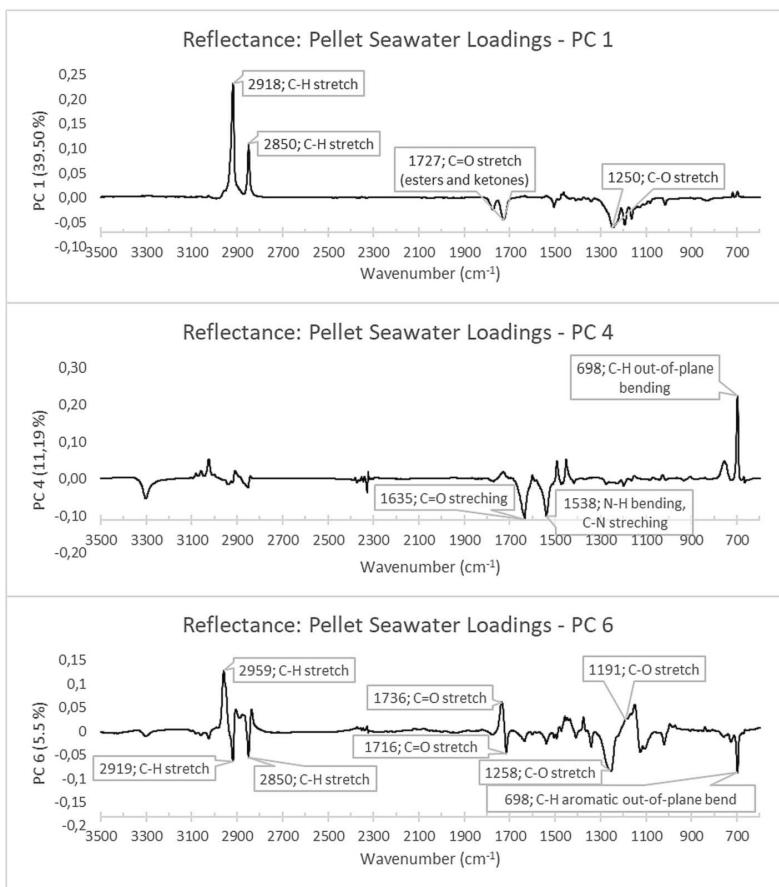


Fig. 2. Seawater-aged pellets, micro reflectance loadings.

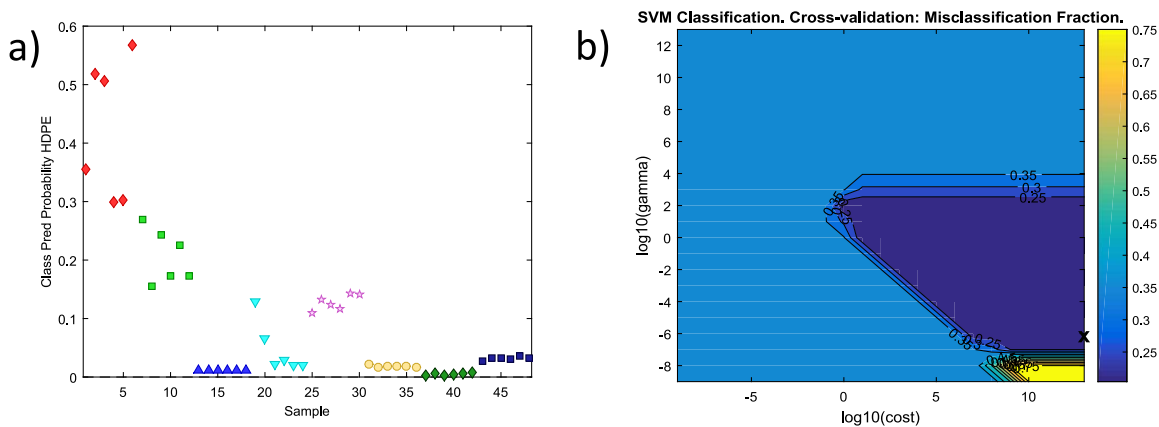


Fig. 3. a) Example of a SVM classification model, corresponding to pelletized HDPE aged in seawater and measured by micro reflectance IR spectrometry. b) Optimization of the cost and gamma parameters (the optimal value is marked with an X at the low, right corner).

characterized by the most relevant positive loadings in PC1, Fig. 5) to the other polymers, mainly PC, PET and PMMA, described by negative loadings (Fig. 5).

PC4 (6 % explained variance) separates mostly PA6.6, with negative

scores and negative loadings, from PP which has positive scores and positive loadings (Figs. 1c and 5, respectively). The maximum loadings agree with characteristic spectral peaks for these polymers.

The use of dynamic PCA reduced the variables needed to keep the

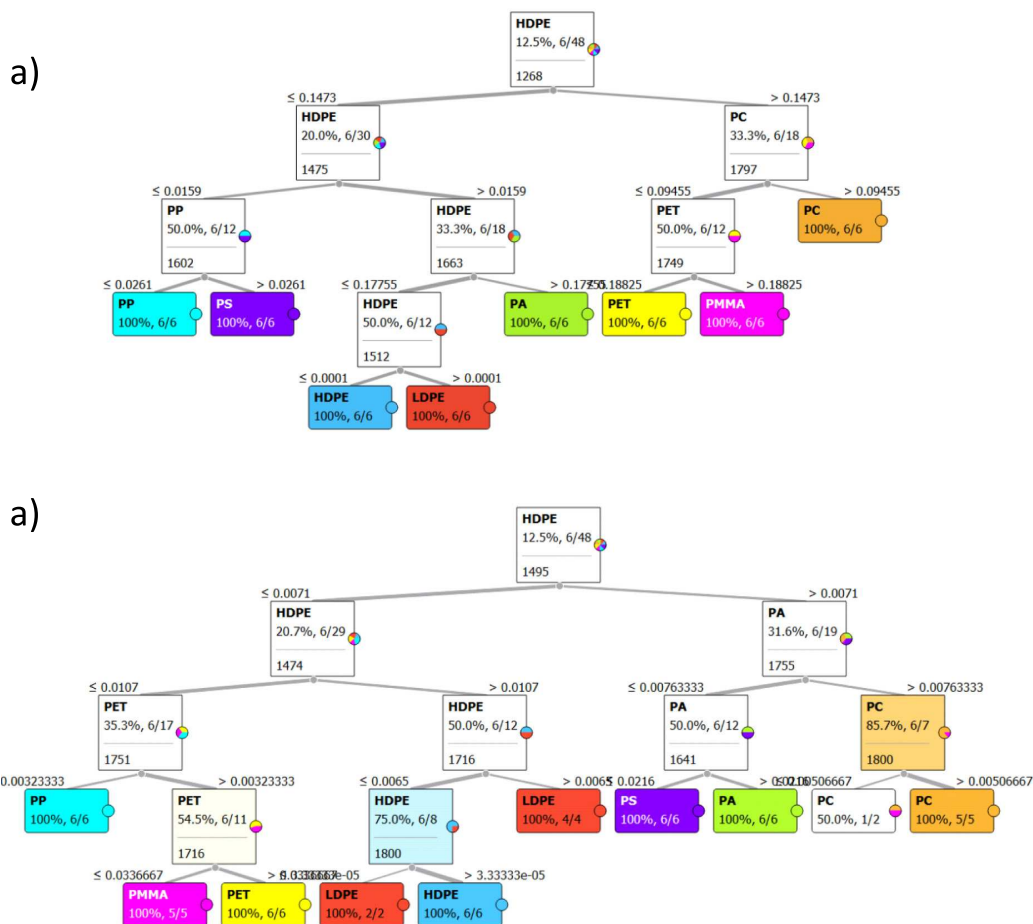


Fig. 4. Pellets aged in seawater, schematic representation of the CART decision tree obtained from the micro reflectance (a) and ATR (b) measurements. The final groups (each in a different colour) are 100 % homogeneous. The values at the bottom of the decision boxes correspond to the discriminating wavenumber whereas the values over the boxes are the absorbance of the wavenumber that separates the branches (e.g., in the first node, if the absorbance at 1268 cm^{-1} is lower than or equal to 0.1473 a.u. the polymer may be HDPE, LDPE, PA, PP or PS).

polymer groups to 10 wavenumbers, at: 2959, 2847, 1736, 1544, 1420, 827, 872, 702, 699 and 688 cm^{-1} .

When the variables associated to the maximum loadings of both factors were used (6 in total), the groups of polymers could also be differentiated, with the exception of a PMMA outlying sample (as noted above). Application of SVM considering only those wavenumbers yielded good separations, with the exception of some HDPE and PS samples.

Finally, CART required only 9 divisions (18 branches) to get a good separation of the polymers (Fig. 4b). The 7 variables required for the sequential decisions were 1800 (twice), 1755, 1751, 1716 (twice), 1495, 1474 and 1641 cm^{-1} , whose general interpretation (CH_2 bending and $\text{C}=\text{O}$ stretching) mostly agrees with the regions selected by dynamic-PCA (but for the $3000\text{--}2800\text{ cm}^{-1}$ spectral region, not included in CART, as explained above).

However, the classification is not as nice as that derived from micro reflectance measurements. The final groups for PP, PET, HDPE, PS and PA are 100 % homogeneous and form differentiated groups. However, LDPE needed two criteria (at 1716 and 1800 cm^{-1}) to be differentiated from HDPE and a PC sample mixes with another PMMA one, although the other pellets of these polymers became well separated.

3.2. Pellets weathered under dry conditions

When micro reflectance measurements were made, the factors that lead to the best differentiation among polymers were PC2 and PC4 (25.2 % explained information, Fig. 1d).

PC2 (15.8 % explained variance) differentiates PMMA and PET (positive scores) from PP, PS and PA (negative scores). Fig. SM2 (Supplementary material) shows the most relevant loadings defining this factor. PMMA and PET became characterized by their ester characteristics (the most positive loadings) while PA, PS and PP were characterized by negative loadings. No distinctive loadings can be attributed to PP, but for the typical CH stretching at ca. 2850 and 2950 cm^{-1} .

PC4 (9.5 % explained variance) is dominated by the Amide II band (NH monosubstituted amide bending plus the CN stretching) and differentiates basically PA (positive scores and positive loadings) from PC and PP (negative scores and their associated loadings, Fig. SM2). The negative loadings seem mostly linked to PC as the PP characteristics (stretching and bending of the CH bonds) cannot be seen, probably, masked by the CH bonds of PA6.6.

Using dynamic PCA the overall number of wavenumbers needed to visualize the groups of polymers got reduced from 2900 to 15 variables.

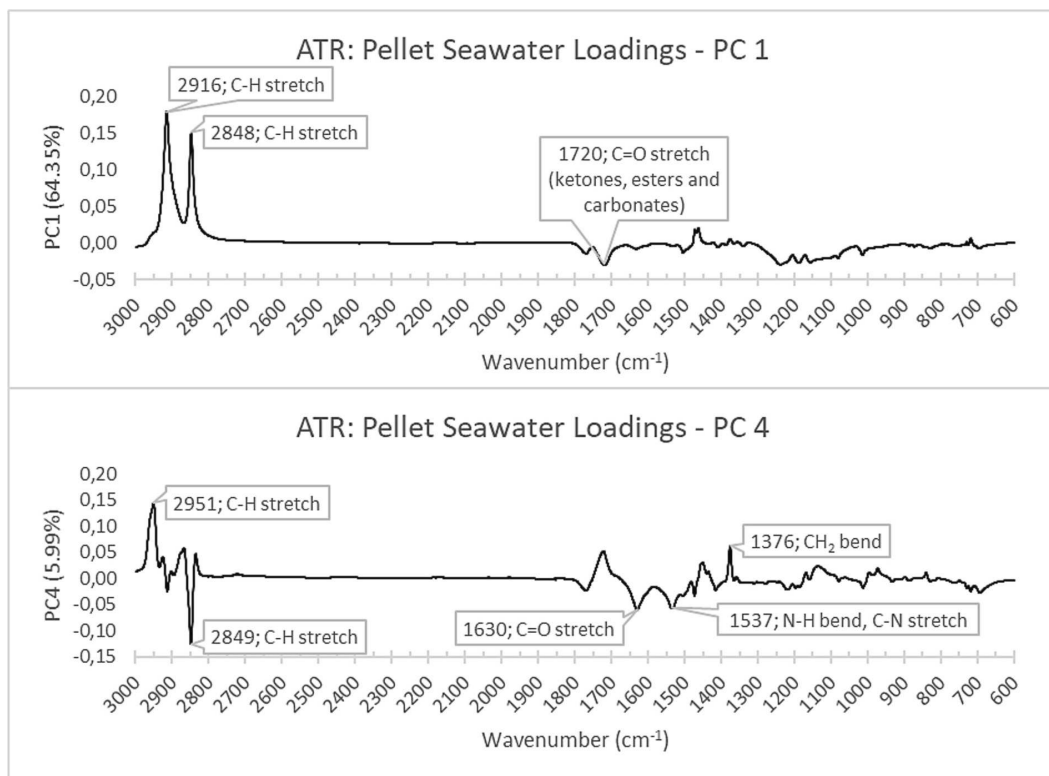


Fig. 5. Seawater-aged pellets, ATR loadings.

These were: 3457–3455, 2918, 1917, 1775, 1732, 1731, 1636–1634, 1269, 1268, 698, and 697 cm^{-1} wavenumbers.

When the 12 variables associated to the most relevant loadings of the principal components were used to perform a dedicated PCA the groups of polymers could still be differentiated. Using SVM, however, it was not possible to categorize the HDPE and LDPE samples. Finally, CART required 10 divisions (20 branches) to successfully separate the polymers (Fig. SM3), using only 9 variables (i.e., 1800 (twice), 1795, 1740, 1677, 1651, 1474, 1468, 1465 and 600 cm^{-1}).

For ATR spectra, the factors that differentiated the polymers best were PC2 and PC3 (18.2 % explained information, Fig. 1e). PC1 could not separate several polymers despite it takes account of more information (likely related to ageing). This result is similar to that with reflectance data above. However, the interpretation of the loadings is a bit more complicated here.

Positive scores and loadings in PC2 (9.89 % explained variance) differentiated clearly PP and PS from the other polymers. The most important positive loadings were linked to PS and PP (the peaks at 2950 and 1452 cm^{-1}) (Fig. SM4). The peaks at 2950 and 1452 related to PP, whereas the others are characteristic of PS. The most relevant negative loadings relate to the C–H stretching of HDPE and LDPE, whose samples outstand in the negative scores. The loading at 1720 cm^{-1} points towards the C=O groups of PC, PA and PET.

PC3 (8.30 % explained variance) separates PP from PS (Fig. 1e), whose scores in PC2 are almost the same. It also helps differentiating LDPE, HDPE and PMMA from PC, PA and PET. Here, it is difficult to assign loadings to specific monomers. The most important loadings relate to C–H stretching, CH₂ bending, C–H bending and aromatic out-of-plane bending (see Fig. SM4).

With the use of dynamic PCA the number of variables can be reduced

to only 4, still with a nice polymer differentiation: 2925 and 2924 cm^{-1} (C–H stretch) and 1714 and 1713 cm^{-1} (mostly C=O bands).

When the variables associated only to the maximum loadings (positive and negative) were used (7 in total) to carry out a PCA, the groups of polymers were still differentiated. When SVM was applied to those variables it was possible to visualize several groups of polymers, but LDPE and PS offered unsatisfactory results.

Finally, CART required only 8 divisions (16 branches) to successfully separate the polymers (Fig. SM5) employing 8 variables (i.e., 1800, 1752, 1733, 1507, 1496, 1474, 1456 and 1378 cm^{-1}).

3.3. Powder aged in seawater

The typical PC1-PC2-PC3 scores subspace (51.9 % explained variance) is adequate to separate the polymers when their powders are weathered in seawater and measured by micro reflectance spectrometry (Fig. 6a and b).

PC1 (21.4 % variance) is devoted to differentiate the PS samples (the most negative scores, Fig. 6a and b). It is not surprising thus that the most remarkable loadings in this factor are associated to PS (Fig. SM6); even the typical four peaks of similar intensity that characterize the typical monosubstituted aromatic pattern between 2000 and 1700 cm^{-1} can be seen.

PC2 opposes the 'simplest' hydrocarbon structures of PP, HDPE and LDPE (negative scores) to the other polymers (which contain heteroatoms and aromatic and more complex monomeric structures), in particular to PS (highest positive scores and whose bands can be linked to the largest positive loadings, Fig. SM6). The most relevant positive loadings are associated to –likely– the more linear structures: CH₂ bending and rocking.

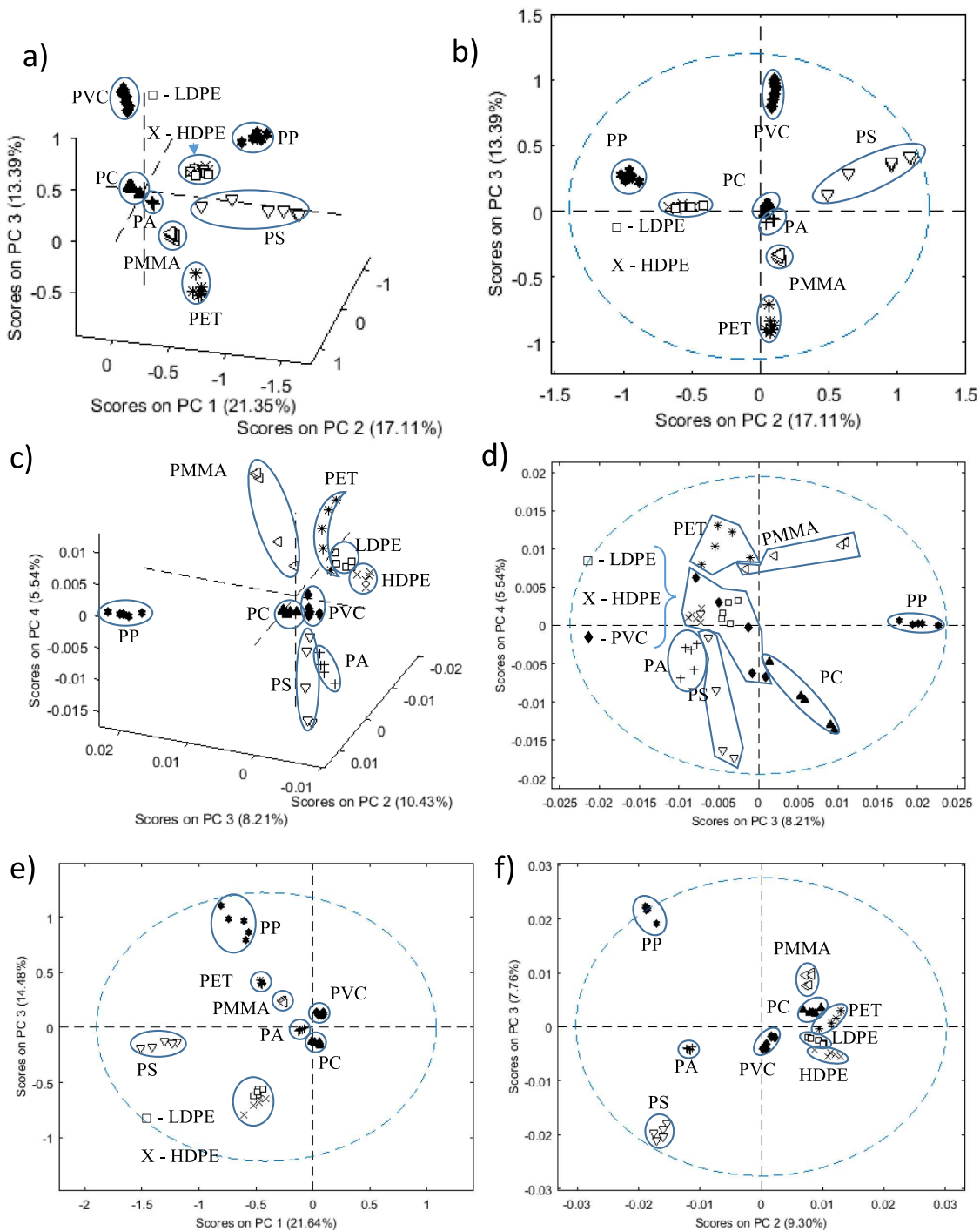


Fig. 6. Representation of the most relevant principal component subspaces to identify the groups of polymers. Seawater-aged powder measured by microreflectance (a) and (b); and by ATR (c) and (d). Dry-aged powder measured by microreflectance (e) and ATR (f).

PC3 separates mainly PVC (highest positive scores) from PET (highest negative scores), Fig. 6a and b. The loadings reflect also this point (Fig. SM6) as PET is characterized by positive loadings, mainly C=O stretch. On the contrary, PVC –an essentially linear structure– is reflected in the most relevant negative loadings, including the typical C–Cl bending.

The variables associated to those loadings were not enough to differentiate the groups of samples when a dedicated PCA was done, likely because of the complexity of the loadings (Fig. SM6) which make it hard any reasonable manual selection. However, after using SVM with these wavenumbers an acceptable separation of the polymers was found.

Using the dynamic PCA algorithm it was possible to reduce the amount of wavenumbers to 11: 2849, 2292, 1745, 1740, 1660, 1443, 1287, 1257, 905, 737 and 702 cm^{-1} . Finally, CART required 8 divisions (16 branches) to get a good separation of the polymers (Fig. SM7) using only 7 variables (i.e., 1800 (twice), 1787, 1776, 1743, 1507, 1392, and 1106 cm^{-1}).

The results obtained when ATR spectra were used are depicted in Fig. 6c and d. The PC2-PC3-PC4 scores subspace (24.1 % explained information) was chosen. Even though the groups appear quite separated from each other the results are not as good as for the study above using micro reflectance because the groups presented quite large dispersions (in fact, PC1 was not selected because the samples there become even more disperse, making it impossible to discern the groups). It also happens that the interpretation of the loadings is not straightforward. In PC2 (10.43 % explained variance) the only polymers that are sharply separated are PA and PP and HDPE and LDPE (figure not shown). The most important positive loadings for this factor (Fig. SM8) characterize HDPE and LDPE, whereas the negative loadings correspond to typical PA spectral bands.

PC3 (8.2 % explained variance) allows a slight separation of HDPE and LDPE (although they overlap with other polymers, Fig. 6c and d) and it clearly differentiates PP. The most important loadings that differentiate it from the other polymers (Fig. SM8) are linked to the C–H structures (positive loadings). The negative loadings point towards the other more complex structures with C=O groups.

PC4 (5.5 % explained variance) separates PET and PMMA (positive scores) from PC and PS (negative scores), Fig. 6d, though the groups are highly dispersed. The latter polymer becomes differentiated by the C=O stretching and the CH out-of-plane aromatic ring stretching (positive loadings, see Fig. SM8).

As for the micro reflectance spectra, the use of the variables associated to the most relevant loadings alone yielded bad results when a PCA was made and that was attributed to the difficulty in selecting the most relevant wavenumbers. However, applying SVM to these variables all the samples became well classified except for some LDPE ones.

On the other hand, using dynamic PCA it was possible to reduce the total amount of wavenumbers required to visualize the groups to 36, in the 2950–2840, 1780–1738 cm^{-1} ranges and 1535, 1427–1424 955–953, 829, 828, 725–723 cm^{-1} wavenumbers.

CART required only 10 divisions (20 branches) to get a good separation of the polymers (Fig. SM9) with only 9 variables (i.e., 1800 (twice), 1795, 1740, 1677, 1651, 1474, 1468, 1465 and 600 cm^{-1}), all of them in the surroundings of the typical C=O carbonyl region –as it happened with dynamic PCA, but for the 600 cm^{-1} wavenumbers.

3.4. Powder weathered under dry conditions

In essence, powdered samples weathered in dry conditions and measured by micro reflectance spectrometry behaved as their seawater-aged counterparts. In this case the separation between the polymers can be achieved with just two principal components (Fig. 6e); i.e. the PC1-PC3 subspace (ca. 36.5 % explained variance). PC2 was not selected because some PP samples got intertwined with those from HDPE and LDPE, yielding worst results. It was found that a PMMA sample behaved anomalously, so it was excluded from the final studies. There, PC1

(21.64 % variance) differentiates PS (extreme negative scores, Fig. 6e). The negative loadings influencing this factor the most were related to PS (Fig. SM10). The peak at 2924 cm^{-1} has no clear assignment and it may be a mixture of the C–H aromatic stretching from PS and the aliphatic CH stretching of HDPE and LDPE.

PC3 (ca. 14.5 % variance) in essence opposes PP (positive scores) to HDPE and LDPE (negative scores), with all other polymers showing close-to-zero scores, although each polymer forms very tight clusters, regardless of the degree of weathering during the study (Fig. 6e). The most important loadings are positive (Fig. SM10), but they cannot be associated to specific structures.

Dynamic PCA reduced the number of variables needed to visualize the groups of polymers from 3400 to 15 ones: 3457–3455, 2918, 1917, 1775, 1732, 1731, 1636–1634, 1269, 1268, 698 and 697 cm^{-1} . Using SVM with the wavenumbers associated with the highest loadings an acceptable separation among the polymers was obtained. Finally, CART employed only 9 divisions (18 branches) to get a good separation of the polymers (Fig. SM11) with only 8 variables (i.e., 1800 (twice), 1713, 1507, 1414, 1276, 1245, 1155 and 717 cm^{-1}).

Finally, two components separate the polymers when ATR spectra are considered (Fig. 6f): PC2 and PC3, being PC1 useless for this purpose. PC2 (9.3 % explained variance) opposes PA, PP and PS (negative scores) to the other polymers, most notably HDPE, LDPE, and PET (Fig. 6f), as it happened previously with the micro reflectance measurements. Relevant positive loadings for this factor coincide with typical HDPE and LDPE bands (Fig. SM12) while highest negative loadings correspond mainly to PP.

PC3 (7.8 % explained variance) separates clearly PS and PP, Fig. 6f. The PS-related negative loading at 695 cm^{-1} (aromatic CH out-of-plane bending) opposes to the most relevant positive loadings (typical C–H stretching, likely from PP, ca. 2950 cm^{-1} , Fig. SM12).

Dynamic PCA required only 3 variables to visualize the groups, corresponding to the C–H stretching region (i.e., 2924, 2923 and 2922 cm^{-1}).

Combining SVM with the wavenumbers associated with the highest loadings a perfect differentiation between the polymers were obtained. Finally, CART made 9 divisions (18 branches) to get a good separation of the polymers (Fig. SM13) with 8 variables (i.e., 1800, 1659, 1605, 1468, 1467, 1136, 722 and 600 (twice) cm^{-1}).

3.5. General comparison of the selected wavenumbers

The wavenumbers selected by the different approaches were highly consistent among them, of course with the logical small differences due to their different fundamentals, and they allowed differentiating among the polymers at the model development (or training) stage. Fig. SM14 depicts graphically all the variables selected from the different models studied above (recall that SVM does not select variables by itself, but we included here a previous step that chooses those linked to the most relevant loadings). The figure shows clearly that the most relevant wavenumbers accumulate almost always in the fingerprint and the C–H stretching (~2800–3000 cm^{-1}) regions. Note that even despite CART considered only the fingerprint region, the variables selected for the decision nodes agree very well with the variables selected (in that region) using the other techniques.

The main question now is to decide which option is best. This involves a twofold choice: it has to be decided among ATR and reflectance measurements, and among the variable selection strategies. Unfortunately, this is a complex decision because although a variable might be useful to separate some polymers using (e.g.) ATR it might be detrimental when reflectance is considered. So, it is not only about ATR or reflectance, it is also about which particular combination of wavenumbers is considered. Many models discussed in the previous section had no false positives nor false negatives (detailed statistics are not shown here) but one has to be aware that this may be an “artefact” caused by overfitting the models. Therefore, it is critical to test each and every

model with samples not used to develop the models at all. This is done in the next section.

3.6. Model validation

To validate the models, two set of samples were used, as detailed in Section 2.1; one considers particles extracted from the weathering setup during the period under study and the other considers field plastics collected at beaches. The latter are much more complex because of their unknown weathering extent, physical and chemical degradation, amount of additives, morphology, etc. Because of that, and after preliminary assays, the “HDPE” and “LDPE” categories will be combined into one, with the general denomination “PE”.

PCA projection of the field samples into the scores subspace didn't yield satisfactory identifications, so it will not be considered further. This is explained because PCA and dynamic-PCA are not true classification methods as the final assignments are made subjectively, as per visual observation of the location of the samples. Hence, they are fast, although rough approaches to ascertain a polymer type. However, using the loadings from the PCA models it was possible to develop useful SVM models with a reduced number of variables Tables 1 and 2.

Table 1 presents the validation statistics for the SVM models. For the samples weathered artificially the results are encouraging for both spectrometric techniques. Micro reflectance models are, in general, slightly better than their ATR-based counterparts since 29 models (out of 30) had MCCs ≥ 0.8 (this happened for 26 ATR ones), with almost no false positives and with only one model showing poor classification

capabilities. The worst models were found for PMMA measured by ATR, although those developed using powders became slightly better than those for pellets. In general, the models differentiate the polymers really well.

When field samples are considered (Table 1, bottom part) the final assignments are not so satisfactory because the spectra became affected more by the different weathering processes they might had undergone: biofouling, physical erosion, chemical degradation, etc. Their spectra are noisier, with broader and less defined bands. Most field samples (>90 %) were thin sheet fragments, from food and one-use wrappings. This produces detrimental effects in reflectance measurements, as it often provokes an etalon-like effect: sinusoidal waves along the spectrum produced by the light passing through several parallel surfaces, which complicates the characterizations. This is reflected in the poor polymer identifications in the micro reflectance models, none of them surpassing a 0.5 MCC value. However, the ATR models showed more acceptable identifications, most of them having MCCs ≥ 0.8 (7 models out of 10), and low ratios of false positives and negatives. This seems very good news for garbage and plastic litter monitoring because they usually concentrate on relatively 'big' fragments, mostly when following OSPAR recommendations.

CART yielded less satisfactory models than SVM (Table 2), as evidenced by slightly worse statistics. For artificially weathered samples, reflectance-based models are preferred to ATR-based ones because up to 18 models lead to satisfactory ≥ 0.8 MCC values (vs. only 12 satisfactory ones using ATR). However, figures for field samples are not adequate. In our view, this might be because of the simple classification strategy of

Table 1
Validation statistics obtained for SVM models. The first value is the MCC statistic; those within the parenthesis represent the ratios of false positives and negatives (respectively). The colors in the boxes indicate the success of the model, from green (adequate model) to orange (unsuccessful model). (PVC was not available in pellets).

Polymer	Weathered samples							
	ATR				Reflectance			
	Pellets		Powder		Pellets		Powder	
Sea	Dry	Sea	Dry	Sea	Dry	Sea	Dry	
PE	1 (0; 0)	1 (0; 0)	1 (0; 0)	1 (0; 0)	1 (0; 0)	0.8 (0.1; 0)	1 (0; 0)	0.9 (0.03; 0)
PA	1 (0; 0)	1 (0; 0)	0.9 (0; 0.3)	0.7 (0; 0.5)	1 (0; 0)	0.8 (0; 0.4)	1 (0; 0)	0.9 (0.03; 0)
PC	1 (0; 0)	0.9 (0; 0.3)	1 (0; 0)	0.9 (0; 0.3)	1 (0; 0)	1 (0; 0)	1 (0; 0)	1 (0; 0)
PET	1 (0; 0)	1 (0; 0)	0.9 (0; 0.3)	0.9 (0.03; 0)	1 (0; 0)	1 (0; 0)	0.9 (0; 0.2)	1 (0; 0)
PMMA	0.5 (0; 0.8)	0.7 (0; 0.5)	1 (0; 0)	0.7 (0; 0.5)	1 (0; 0)	1 (0; 0)	1 (0; 0)	0.8 (0; 0.4)
PP	0.9 (0; 0.3)	1 (0; 0)	1 (0; 0)	1 (0; 0)	0.9 (0; 0.2)	0 (0; 1)	1 (0; 0)	1 (0; 0)
PS	1 (0; 0)	1 (0; 0)	1 (0; 0)	0.9 (0; 0.3)	1 (0; 0)	1 (0; 0)	1 (0; 0)	0.9 (0; 0.2)
PVC	--	--	1 (0; 0)	0.9 (0; 0.3)	--	--	1 (0; 0)	0.9 (0; 0.2)
Field samples								
Polymer	ATR		Reflectance					
	Sea	Dry	Sea	Dry				
PE	0.95 (0; 0.07)	0.95 (0; 0.07)	0.4 (0; 0.7)	0.3 (0.5; 0.1)				
PET	0.8 (0.03; 0.2)	0.9 (0.02; 0.1)	0.3 (0; 0.9)	0.3 (0; 0.9)				
PP	0.8 (0; 0.3)	0.8 (0; 0.3)	0.5 (0; 0.7)	0 (0; 1)				
PS	0.8 (0; 0.5)	0.5 (0.07; 0.25)	0 (0; 1)	-0.04 (0.06; 1)				
PVC	0.4 (0.02; 0.7)	0.6 (0; 0.6)	0.5 (0.02; 0.7)	0.5 (0.02; 0.7)				

Table 2

Validation statistics obtained for CART models. The first value corresponds to the MCC statistic; those within the parenthesis represent the ratios of false positives and negatives (respectively). The colors in the boxes indicate the success of the model, from green (adequate model) to orange (unsuccessful model). (PVC was not available in pellets).

Weathered samples								
Polymer	ATR				Reflectance			
	Pellets		Powder		Pellets		Powder	
	Sea	Dry	Sea	Dry	Sea	Dry	Sea	Dry
PE	1 (0; 0)	0.9 (0.2; 0)	0.9 (0; 0.1)	0.9 (0; 0.1)	0.3 (0.6; 0.2)	0.6 (0.5; 0.1)	0.9 (0; 0.1)	0.9 (0.1; 0)
PA	0.85 (0; 0.3)	0.7 (0; 0.5)	0.2 (0.7; 0.8)	0.3 (0.7; 0.5)	0.6 (0; 0.6)	0 (0; 1)	0.9 (0.2; 0)	1 (0; 0)
PC	1 (0; 0)	0.9 (0; 0.3)	0.5 (0; 0.8)	0.7 (0.4; 0)	0.8 (0; 0.4)	0.4 (0; 0.8)	0 (0; 1)	0.9 (0.2; 0)
PET	0 (0; 1)	-0.07 (1; 1)	0 (0; 1)	0.7 (0.3; 0.3)	1 (0; 0)	0.8 (0.4; 0)	0.8 (0.4; 0)	0 (0; 1)
PMMA	-0.06 (1; 1)	0.5 (0.6; 0.3)	1 (0; 0)	0.5 (0; 0.75)	0.4 (0; 0.8)	0.8 (0; 0.4)	0.2 (0.7; 0.6)	0.6 (0.2; 0.5)
PP	0.5 (0.7; 0)	0.8 (0.3; 0)	0.9 (0.3; 0)	-0.059 (0; 0)	0.1 (0.8; 0.8)	0.5 (0.5; 0.4)	0.9 (0.2; 0)	0.9 (0; 0.2)
PS	1 (0; 0)	0.9 (0; 0.3)	0 (0; 1)	0.5 (0.3; 0.5)	0.9 (0; 0.2)	0.8 (0.2; 0.2)	1 (0; 0)	0.9 (0; 0.2)
PVC	--	--	0.4 (0.8; 0)	0.7 (0; 4)	--	--	0.9 (0; 0.2)	0.9 (0.2; 0)
Field samples								
Polymer	ATR		Reflectance					
	Powder		Powder					
	Sea	Dry	Sea	Dry				
PE	0.4 (0.5; 0.2)	0.4 (0.5; 0.2)	0.3 (0.5; 0.2)	0.03 (0.6; 0.7)				
PET	0.4 (0; 0.8)	0.6 (0.4; 0.4)	0.2 (0.5; 0.9)	0.1 (0.8; 0.8)				
PP	0 (0; 1)	0.5 (0.2; 0.6)	-0.2 (1; 1)	-0.15 (0.9; 0.9)				
PS	0 (0; 1)	-0.05 (1; 1)	0 (0; 1)	0.7 (0; 0.5)				
PVC	0.4 (0.6; 0.3)	0.3 (0.6; 0.7)	0.4 (0.6; 0.3)	0.1 (0.8; 0.6)				

the CART method. Dichotomic decision rules work well as long as similar spectra are considered for both calibration and validation. However, field samples were exposed to many weathering phenomena that, finally, distort the original spectrum so that the dichotomic absorbance decisions at the different nodes can be misleading.

Therefore, in the authors' opinion the final tradeoff analysis points out that ATR analysis (of both pellets and powders) combined with the chemometric support vector machine tool lead to the most satisfactory differentiation among the nine polymers considered in this work, along with satisfactory classification ratios of new samples. Although particular details were given in the previous sections, Table SM1 resumes the variables employed for each model and it can be seen that the three most relevant structural moieties involved in the differentiation and classification of the polymers were the C—H stretching, C—H bending, C=O stretching and the C—H out-of-plane bending (likely of aromatic rings); see Table SM2 (compiled from Brandon et al. (2016), Tiwari et al. (2019), Vasanthan (2012), Veerasingam et al. (2021), Jung et al. (2018)). Somehow this suggests that the models try to find out the signals of the main structural chains of the polymers (note that—for instance—the broad O—H stretching band typical of polymer weathering in the 3000–3200 cm⁻¹ region does not appear at all within any variable selection).

4. Conclusions

The results obtained in this study indicate that polymer weathering complicates the formation of clear groups of samples in multivariate studies. In many cases, high-order principal components were needed to visualize more or less definite groups. This means that the most relevant spectral variance is not related to the different polymers but to other

factors, mostly weathering. In general, the micro reflectance spectra of pellets were more interpretable and lead to better groups than those of powders, likely because reflected radiation is mostly specular, which yields simpler spectra than the diffuse radiation from the powders.

The studies presented here suggest that pattern recognition models developed with a reduced suite of selected variables may be a good way to address this problem. All the variable selection strategies allowed for significant reductions in the number of wavenumbers required to visualize the groups of polymers; being the most extreme situation that for dynamic PCA in the dry weathering of powdered polymers by ATR which required only three wavenumbers to differentiate the groups.

The simplest feature selection implies selecting wavenumbers associated to the maximum loadings of the principal components. This yields satisfactory differentiations when the spectra are well defined, mostly with pellet configurations. When this option is combined with SVM it is possible to differentiate all polymers with only some few marginal errors in some models (mostly for HDPE and LDPE, due to their spectral similarity).

Dynamic PCA is also a very simple methodology and it leads to very satisfactory results although the number of wavenumbers required to visualize the groups of polymers vary among the assays. In most cases, only 10 wavenumbers (out of the 3400 initial ones) were sufficient. Unfortunately, the projections of the field samples led to poor assignments because they have to be done visually.

CART behaved very consistently and required around 8 wavenumbers to separate the polymer groups. However, they were not always 100 % homogeneous and results were not too satisfactory for field samples. This was attributed to the spectral differences between the field and artificially-weathered plastics (which in turn yielded good validations), which make the dichotomic CART decisions suboptimal for field

samples. The classification models for SVM are more satisfactory when ATR spectra are considered. In particular, SVM performed clearly better than CART for field samples.

In our view, this work opens up the possibility of identifying polymers in spite of their weathering level by considering a reduced number of IR wavenumbers, although more studies are needed to improve the modelization of tiny particles and to ascertain how to incorporate this approach into the databases.

A final reflection on the limitations and future work to refine this approach is in order. The artificial weathering process employed here yields physical and chemical degradation but biofouling was not considered. The SVM models worked well for field samples without obvious biofouling but it remains to be assessed whether they also work in other circumstances where more bio-materials are present at the surface of the plastics. Such a study is being done in our laboratory as a part of a more comprehensive project (the JPI-Oceans EU-funded MicroplasticX). As a referee pointed out, it also has to be studied how the models behave when big amounts of additives are added to the core polymer so that they affect its spectrum.

It is worth noting that this approach can be used for on-site field MPs identification. Right now there are commercial portable (car battery-powered) ATR-FTIR systems, even from major brands, and since they are controlled by laptops it is indeed possible to include the chemometric models there. However, micro reflectance instrumental systems are not still portable. So, currently the portability option depends on the size of the microplastics under scrutiny (likely, 500 µm would be the very minimum acceptable size for current ATR devices).

CRedit authorship contribution statement

Conceptualization: B.F.; J.M.A.; S.M.: Methodology, Software and Formal analysis: B.F.; J.M.A.; C. P-Q.; V.F-G.; Validation: B.F.; J.M.A.; V. F-G.; P. L-M.; S.M.; Resources and Funding acquisition: S.M.; P. L-M; Writing: B.F.; J.M.A.; C. P-Q.; V.F-G.; P.L-M.; S.M.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Prof. Dr. Soledad Muniategui reports financial support was provided by EU Framework Programme for Research and Innovation Societal Challenges. Soledad Muniategui reports financial support was provided by Spain Ministry of Science and Innovation. Soledad Muniategui reports financial support was provided by Government of Galicia Department of Culture Education and Universities.

Data availability

The authors do not have permission to share data.

Acknowledgments

This work was supported by the EU Horizon2020-JPI Oceans Project “LAnd-Based Solutions for PLAstics in the Sea” (Grant No. 101003954, LABPLAS), and the “Integrated approach on the fate of MicroPlastics towards healthy marine ecosystems” (MicroplastiX, Grant PCI2020-112145, supported by the JPI Oceans Program and by Spanish Government, MCIN/AEI/10.13039/501100011033 and the European Union “Next Generation EU/PRTR program”). The Galician Government (“Xunta de Galicia”) is acknowledged for its support to the QANAP group (Programa de Consolidación y Estructuración de Unidades de Investigación Competitiva. Ref. ED431C 2021/56). Funding for open access charge: Universidade da Coruña/CISUG.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.marpolbul.2022.113897>.

References

- Ali, S.S., Elsamahy, T., Koutra, E., Kornaros, M., El-Sheekh, M., Abdelkarim, E.A., Zhu, D., Sun, J., 2021. Degradation of conventional plastic wastes in the environment: a review on current status of knowledge and future perspectives of disposal. *Sci. Total Environ.* 771, 144719 <https://doi.org/10.1016/j.scitotenv.2020.144719>.
- Andrade, J., Fernández-González, V., López-Mahía, P., Muniategui, S., 2019. A low-cost system to simulate environmental microplastic weathering. *Mar. Pollut. Bull.* 149, 110663 <https://doi.org/10.1016/j.marpolbul.2019.110663>.
- Andrade, J.M., Ferreira, B., López-Mahía, P., Muniategui-Lorenzo, S., 2020. Standardization of the minimum information for publication of infrared-related data when microplastics are characterized. *Mar. Pollut. Bull.* 154 <https://doi.org/10.1016/j.marpolbul.2020.111035>.
- Arrieta, C., Dong, Y., Lan, A., Vu-Khanh, T., 2013. Outdoor weathering of polyamide and polyester ropes used in fall arrest equipment. *J. Appl. Polym. Sci.* 130, 3058–3065. <https://doi.org/10.1002/app.39524>.
- Battaglia, P., Pedà, C., Musolino, S., Esposito, V., Andaloro, F., Romeo, T., 2016. Diet and first documented data on plastic ingestion of *Trachinotus ovatus* L. 1758 (Pisces: Carangidae) from the strait of Messina (central Mediterranean Sea). *Ital. J. Zool.* 83, 121–129. <https://doi.org/10.1080/11250003.2015.1114157>.
- Bellas, J., Martínez-Armental, J., Martínez-Cámara, A., Besada, V., Martínez-Gómez, C., 2016. Ingestion of microplastics by demersal fish from the Spanish Atlantic and Mediterranean coasts. *Mar. Pollut. Bull.* 109, 55–60. <https://doi.org/10.1016/j.marpolbul.2016.06.026>.
- Bläsing, M., Amelung, W., 2018. Plastics in soil: analytical methods and possible sources. *Sci. Total Environ.* 612, 422–435. <https://doi.org/10.1016/j.scitotenv.2017.08.086>.
- Botterell, Z.L.R., Beaumont, N., Dorrington, T., Steinke, M., Thompson, R.C., Lindeque, P. K., 2019. Bioavailability and effects of microplastics on marine zooplankton: a review. *Environ. Pollut.* 245, 98–110. <https://doi.org/10.1016/j.envpol.2018.10.065>.
- Brandon, J., Goldstein, M., Ohman, M.D., 2016. Long-term aging and degradation of microplastic particles: comparing in situ oceanic and experimental weathering patterns. *Mar. Pollut. Bull.* 110, 299–308. <https://doi.org/10.1016/j.marpolbul.2016.06.048>.
- Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J.H., Abu-omar, M., Scott, S.L., Suh, S., 2020. Degradation rates of plastics in the environment. *Sustain. Chem. Eng.* 8 (9), 3494–3511. <https://doi.org/10.1021/acsuschemeng.9b06635>.
- Chiba, S., Saito, H., Fletcher, R., Yogi, T., Kayo, M., Miyagi, S., Ogido, M., Fujikura, K., 2018. Human footprint in the abyss: 30 year records of deep-sea plastic debris. *Mar. Policy* 96, 204–212. <https://doi.org/10.1016/j.marpol.2018.03.022>.
- Coffin, S., Huang, G.-Y., Lee, I., Schlenk, D., 2019. Fish and seabird gut conditions enhance desorption of estrogenic chemicals from commonly-ingested plastic items. *Environ. Sci. Technol.* 53, 4588–4599.
- Compa, M., Ventero, A., Iglesias, M., Deudero, S., 2018. Ingestion of microplastics and natural fibres in *Sardina pilchardus* (Walbaum, 1792) and *Engraulis encrasicolus* (Linnaeus, 1758) along the Spanish Mediterranean coast. *Mar. Pollut. Bull.* 128, 89–96. <https://doi.org/10.1016/j.marpolbul.2018.01.009>.
- Cuadros-Rodríguez, L., Pérez-Castaño, E., Ruiz-Samblás, C., 2016. Quality performance metrics in multivariate classification methods for qualitative analysis. *TrAC - Trends Anal. Chem.* 80, 612–624. <https://doi.org/10.1016/j.trac.2016.04.021>.
- Fernández-González, V., Andrade-Garda, J.M., López-Mahía, P., Muniategui-Lorenzo, S., 2021a. Impact of weathering on the chemical identification of microplastics from usual packaging polymers in the marine environment. *Anal. Chim. Acta* 1142, 179–188. <https://doi.org/10.1016/j.aca.2020.11.002>.
- Fernández-González, V., Andrade, J.M., Ferreira, B., López-Mahía, P., Muniategui-Lorenzo, S., 2021b. Monitorization of polyamide microplastic weathering using attenuated total reflectance and microreflectance infrared spectrometry. *Spectrochim. Acta, Part A* 263, 120162. <https://doi.org/10.1016/j.saa.2021.120162>.
- Fernández-González, V., Andrade-Garda, J.M., López-Mahía, P., Muniategui-Lorenzo, S., 2022. Misidentification of PVC microplastics in marine environmental samples. *Trends Anal. Chem.* 153, 116649 <https://doi.org/10.1016/j.trac.2022.116649>.
- Gewert, B., Plassmann, M.M., Macleod, M., 2015. Pathways for degradation of plastic polymers floating in the marine environment. *Environ. Sci. Process. Impacts* 17, 1513–1521. <https://doi.org/10.1039/c5em00207a>.
- Gok, A., Fagerholm, C.L., French, R.H., Bruckman, L.S., 2019. Temporal evolution and pathway models of poly(ethylene-terephthalate) degradation under multi-factor accelerated weathering exposures. *PLoS One* 14, 1–23. <https://doi.org/10.1371/journal.pone.0212258>.
- Göpferich, A., 1996. In: Williams, D.F.B.T.-T.B.S.J.C. (Ed.), *Mechanisms of Polymer Degradation and Erosion*. Elsevier Science, Oxford, pp. 117–128. <https://doi.org/10.1016/B978-008045154-1.50016-2>.
- ter Halle, A., Ladirat, L., Martignac, M., Mingotaud, A.F., Boyron, O., Perez, E., 2017. To what extent are microplastics from the open ocean weathered? *Environ. Pollut.* 227, 167–174. <https://doi.org/10.1016/j.envpol.2017.04.051>.
- Hirsch, S.G., Barel, B., Shpapper, D., Segal, E., Gazit, O.M., 2017. Correlating chemical and physical changes of photo-oxidized low-density polyethylene to the activation

- energy of water release. *Polym. Test.* 64, 194–199. <https://doi.org/10.1016/j.polymertesting.2017.10.005>.
- Howell, E.A., Bograd, S.J., Seki, M.P., Polovina, J.J., 2012. On North Pacific circulation and associated marine debris concentration. *Mar. Pollut. Bull.* 65, 16–22. <https://doi.org/10.1016/j.marpolbul.2011.04.034>.
- Íñiguez, M.E., Conesa, J.A., Fullana, A., 2017. Microplastics in spanish table salt. *Sci. Rep.* 7, 8620. <https://doi.org/10.1038/s41598-017-09128-x>.
- ISO, 2020. ISO/TR 21960:2020 Plastics — Environmental Aspects — State of Knowledge and Methodologies. ISO, Geneva (Switzerland).
- Jenner, L.C., Rotchell, J.M., Bennett, R.T., Cowen, M., Tentzeris, V., Sadofsky, L.R., 2022. Detection of microplastics in human lung tissue using μ FTIR spectroscopy. *Sci. Total Environ.* 831, 154907. <https://doi.org/10.1016/j.scitotenv.2022.154907>.
- Jiang, J.-Q., 2018. Occurrence of microplastics and its pollution in the environment: a review. *Sustain. Prod. Consum.* 13, 16–23. <https://doi.org/10.1016/j.spc.2017.11.003>.
- Jung, M.R., Horgen, F.D., Orski, S.V., Rodríguez, C.V., Beers, K.L., Balazs, G.H., Jones, T. T., Work, T.M., Brignac, K.C., Royer, S.-J., Hyrenbach, K.D., Jensen, B.A., Lynch, J. M., 2018. Validation of ATR FT-IR to identify polymers of plastic marine debris, including those ingested by marine organisms. *Mar. Pollut. Bull.* 127, 704–716. <https://doi.org/10.1016/j.marpolbul.2017.12.061>.
- Koelmans, A.A., Mohamed Nor, N.H., Hermesen, E., Kooi, M., Mintenig, S.M., De France, J., 2019. Microplastics in freshwaters and drinking water: critical review and assessment of data quality. *Water Res.* 155, 410–422. <https://doi.org/10.1016/j.watres.2019.02.054>.
- Lamb, J.B., Willis, B.L., Fiorenza, E.A., Couch, C.S., Howard, R., Rader, D.N., True, J.D., Kelly, L.A., Ahmad, A., Jompa, J., Harvell, C.D., 2018. Plastic waste associated with disease on coral reefs. *Science* (80-) 359. <https://doi.org/10.1126/science.aar3320>, 460 LP – 462.
- León, V.M., García-Agüera, I., Moltó, V., Fernández-González, V., Llorca-Pérez, L., Andrade, J.M., Muniategui-Lorenzo, S., Campillo, J.A., 2019. PAHs, pesticides, personal care products and plastic additives in plastic debris from spanish Mediterranean beaches. *Sci. Total Environ.* 670, 672–684. <https://doi.org/10.1016/j.scitotenv.2019.03.216>.
- Leslie, H.A., Van Velzen, M.J., Brandsma, S.H., Vethaak, A.D., Garcia-Vallejo, J.J., Lamoree, M.H., 2022. Discovery and quantification of plastic particle pollution in human blood. *Environ. Int.* 163, 107199. <https://doi.org/10.1016/j.envint.2022.107199>.
- Luo, H., Xiang, Y., Zhao, Y., Li, Y., Pan, X., 2020. Nanoscale infrared, thermal and mechanical properties of aged microplastics revealed by an atomic force microscopy coupled with infrared spectroscopy (AFM-IR) technique. *Sci. Total Environ.* 744. <https://doi.org/10.1016/j.scitotenv.2020.140944>.
- Markic, A., Gaertner, J.-C., Gaertner-Mazouni, N., Koelmans, A.A., 2020. Plastic ingestion by marine fish in the wild. *Crit. Rev. Environ. Sci. Technol.* 50, 657–697. <https://doi.org/10.1080/10643389.2019.1631990>.
- Mecozzi, M., Pietroletti, M., Monakhova, Y.B., 2016. FTIR spectroscopy supported by statistical techniques for the structural characterization of plastic debris in the marine environment: application to monitoring studies. *Mar. Pollut. Bull.* 106, 155–161. <https://doi.org/10.1016/j.marpolbul.2016.03.012>.
- Oreski, G., Wallner, G.M., 2005. Aging mechanisms of polymeric films for PV encapsulation. *Sol. Energy* 79, 612–617. <https://doi.org/10.1016/j.solener.2005.02.008>.
- PlasticEurope, 2021. Plastics - The Facts 2021: An Analysis of European Plastics Production, Demand and Waste Data.
- Prata, J.C., 2018. Airborne microplastics: consequences to human health? *Environ. Pollut.* 234, 115–126. <https://doi.org/10.1016/j.envpol.2017.11.043>.
- Primpke, S., Christiansen, S.H., Cowger, W., et al., 2020. Critical assessment of analytical methods for the harmonized and cost-efficient analysis of microplastics. *Appl. Spectrosc.* 74 (9), 1012–1047. <https://doi.org/10.1177/0003702820921465>.
- Raddadi, N., Fava, F., 2019. Biodegradation of oil-based plastics in the environment: existing knowledge and needs of research and innovation. *Sci. Total Environ.* 679, 148–158. <https://doi.org/10.1016/j.scitotenv.2019.04.419>.
- Renner, G., Schmidt, T.C., Schram, J., 2017. A new chemometric approach for automatic identification of microplastics from environmental compartments based on FT-IR spectroscopy. *Anal. Chem.* 89, 12045–12053. <https://doi.org/10.1021/acs.analchem.7b02472>.
- Shi, Y., Liu, P., Wu, X., Shi, H., Huang, H., Wang, H., Gao, S., 2021. Insight into chain scission and release profiles from photodegradation of polycarbonate microplastics. *Water Res.* 195, 116980. <https://doi.org/10.1016/j.watres.2021.116980>.
- Tiwari, M., Rathod, T.D., Ajmal, P.Y., Bhangare, R.C., Sahu, S.K., 2019. Distribution and characterization of microplastics in beach sand from three different indian coastal environments. *Mar. Pollut. Bull.* 140, 262–273. <https://doi.org/10.1016/j.marpolbul.2019.01.055>.
- Van Cauwenbergh, L., Janssen, C.R., 2014. Microplastics in bivalves cultured for human consumption. *Environ. Pollut.* 193, 65–70. <https://doi.org/10.1016/j.envpol.2014.06.010>.
- Vasanthan, N., 2012. Crystallinity determination of nylon 66 by density measurement and fourier transform infrared (FTIR) spectroscopy. *J. Chem. Educ.* 89, 387–390. <https://doi.org/10.1021/ed200398m>.
- Veerasingam, S., Ranjani, M., Venkatachalapathy, R., Bagaev, A., Mukhanov, V., Litvinyuk, D., Mugilarasan, M., Gurumoorthy, K., Guganathan, L., Aboobacker, V.M., Vethamony, P., 2021. Contributions of fourier transform infrared spectroscopy in microplastic pollution research: a review. *Crit. Rev. Environ. Sci. Technol.* 51, 2681–2743. <https://doi.org/10.1080/10643389.2020.1807450>.
- Venkatachalam, S., Nayak, S.G., Labde, J.V., Gharal, P.R., Rao, K., Kelkar, A.K., 2012. Degradation and recyclability of poly (ethylene terephthalate). In: *Polyester*. InTech Rijeka, Croatia, pp. 75–98. <https://doi.org/10.5772/48612>.
- Wang, J., Tan, Z., Peng, J., Qiu, Q., Li, M., 2016. The behaviors of microplastics in the marine environment. *Mar. Environ. Res.* 113, 7–17. <https://doi.org/10.1016/j.marenvres.2015.10.014>.
- Wayman, C., Niemann, H., 2021. The fate of plastic in the ocean environment—a minireview. *Environ Sci Process Impacts* 23, 198–212. <https://doi.org/10.1039/d0em00446d>.
- Wilcox, C., Puckridge, M., Schuyler, Q.A., Townsend, K., Hardesty, B.D., 2018. A quantitative analysis linking sea turtle mortality and plastic debris ingestion. *Sci. Rep.* 8, 1–11. <https://doi.org/10.1038/s41598-018-30038-z>.
- Young, L.C., Vanderlip, C., Duffy, D.C., Afanasyev, V., Shaffer, S.A., 2009. Bringing home the trash: do Colony-based differences in foraging distribution Lead to increased plastic ingestion in laysan Albatrosses? *PLoS One* 4, e7623. <https://doi.org/10.1371/journal.pone.0007623>.
- Yousif, E., Haddad, R., 2013. Photodegradation and photostabilization of polymers, especially polystyrene: review. *Springerplus* 2, 1–32. <https://doi.org/10.1186/2193-1801-2-398>.
- Zhang, K., Hamidian, A.H., Tubić, A., Zhang, Y., Fang, J.K.H., Wu, Ch., Lam, P.K.S., 2021. Understanding plastic degradation and microplastic formation in the environment: a review. *Environ. Pollut.* 274, 116554. <https://doi.org/10.1016/j.envpol.2021.116554>.

WEATHERING-INDEPENDENT DIFFERENTIATION OF MICROPLASTIC POLYMERS BY REFLECTANCE FT-MID IR SPECTROMETRY AND PRINCIPAL COMPONENTS ANALYSIS

Borja Ferreiro, Jose M. Andrade*, Carlota Paz-Quintáns,
Verónica Fernández-González, Purificación López-Mahía, Soledad Muniategui

Group of Applied Analytical Chemistry, Department of Chemistry, University of A Coruña,
Campus da Zapateira, 15071, A Coruña, España

*Corresponding author: fax: +34-981-167065; *andrade@udc.es*

SUPPLEMENTARY MATERIAL

INDEX:

Supplementary Text	pp 2
Supplementary Tables	pp 5
Supplementary Figures	pp 7

MATERIAL AND METHODS

SAMPLES

For this work we considered HDPE (high density polyethylene; from LyondellBasell, commercial name Lupolen 4261 AG UV, density and melting temperature were $0.945 \text{ g}\cdot\text{cm}^{-3}$ and $130 \text{ }^\circ\text{C}$); LDPE (low density polyethylene; from LyondellBasell, commercial name Lupolen 1800P, density and melting temperature were $0.918 \text{ g}\cdot\text{cm}^{-3}$ and $105 \text{ }^\circ\text{C}$); PA 6.6 (Nylon, polyamide, or poly(hexamethylene adipamide; from BASF; commercial name 'Ultramid', density and melting temperature were $1.13 \text{ g}\cdot\text{cm}^{-3}$ and $260 \text{ }^\circ\text{C}$); PC (polycarbonate; from Bayer, commercial name Makrolon 2558, density was $1.2 \text{ g}\cdot\text{cm}^{-3}$); PET (Polyethylene Terephthalate; from Neogroup, commercial name Neopet 80, melting temperature was $248 \text{ }^\circ\text{C}$); PMMA Poly(methyl methacrylate; from Plexiglas, commercial name Plexiglas 7N, density was $1.19 \text{ g}\cdot\text{cm}^{-3}$); PP (polypropylene; from Borealis, commercial name HL508FB, melting temperature $158 \text{ }^\circ\text{C}$); PS (polystyrene; from INEOS Styrolution, commercial name, Styrolution PS 158N/L; density $1.04 \text{ g}\cdot\text{cm}^{-3}$); and PVC (Polyvinyl chloride; from Vinnolit, commercial name Vinnolit S3268).

RESULTS AND DISCUSSION

INTRODUCTION TO THE CHEMOMETRIC TECHNIQUES

Principal Components Analysis (PCA, also known as factor analysis despite mathematically they are not strictly the same) aims to define directions of maximum variance (principal components or factors) so that they explain most of the information in the wavenumbers defining the collection of spectra. The principal components (PCs) can be visualized as combinations of variables whose directions are those on which the samples spread the most.

A given collection of n spectra, where w wavenumbers have been measured on each spectrum, can be arranged in a raw data matrix $\mathbf{X}_{(n \times w)}$. The basic equation for PCA states that a raw data matrix can be decomposed in two simpler matrices. One contains information on the samples (\mathbf{T} , the scores matrix) while the other contains information about the variables (\mathbf{P} , the loadings matrix); i.e., $\mathbf{X}_{(n,w)} = \mathbf{T}_{(n,k)} \cdot \mathbf{P}_{(k,w)}$. k represents the number of principal components relevant for the problem at hand (which has to be decided by the scientist). The first column of \mathbf{T} and the first row of \mathbf{P} constitute the first principal component; and so forth for the second, the third, etc. By definition, the first PC (termed PC1) explains the largest amount of variance (information), the second (PC2) explains less (but more than the third, PC3), and so forth.

For the purpose of the simple understanding of the results of this paper, any principal component can be represented as $PC_k = p_{(k,1)} \cdot \mathbf{x}_1 + p_{(k,2)} \cdot \mathbf{x}_2 + \dots + p_{(k,w)} \cdot \mathbf{x}_w$; where k denotes the principal component under consideration, p is the variable-specific coefficient or loading (from matrix $\mathbf{P}_{(k,w)}$ above), and \mathbf{x} denotes each of the $(1, \dots, w)$ experimentally measured spectral variables (i.e., the intensities organized as column vectors). Hence, for a given principal component, high loadings denote important variables whereas loadings close to zero denote not relevant ones. More details can be found at Brereton (2006) and Otto (2017).

Hence, each PC can be interpreted chemically by studying the more relevant coefficients defining them (loadings) and recalling that variables whose coefficients have the same sign vary in the same way. A note is in order here: since the loadings derive from the overall spectra of all samples, the chemical assignments may sometimes be not straightforward because some loadings may become slightly displaced from typical well-known spectral peaks. This may be accentuated in wider bands and very close neighbouring peaks.

To reduce the spectral range required to identify the polymers after PCA two approaches were applied:

- i) Dynamic PCA removes variables iteratively while assessing the effect of their removal on a PCs scores plot subspace. In essence, it evaluates whether a variable can differentiate between a group of samples and all other samples ('one vs all approach'), according to a Student's t -test. Those variables not contributing statistically to the differentiation will be avoided. The procedure is repeated for each group of samples and, then, all relevant variables are considered altogether. This was done using GenEx's proprietary algorithm (MultiD Analysis AB, Göteborg, Sweden). The initial spectral range was $3000 - 600 \text{ cm}^{-1}$.
- ii) Selection of the wavenumbers associated to the most relevant loadings (either positive or negative). This approach was applied to both the pelletized and powdered samples.

CART (classification and regression trees) looks for individual wavenumbers that allow for decision rules that when applied to the collection of spectral data characterise a group of samples against the other groups. CART works by creating binary recursive partitions or 'branches' based on a tree-like structure where each branch divides the dataset in two different groups. For every branch there is a 'node' that sets a condition for the spectra (Lewis (2000); Otto (2017)). For example, 'If wavenumber at $2900 \text{ cm}^{-1} > 0.5$ u.a., the spectrum corresponds to HDPE, otherwise it belongs to the others group'. Then, a new condition is developed for a next branch until –hopefully– every polymer is identified. In order to reduce the computing intensity only the fingerprint region ($1800 - 600 \text{ cm}^{-1}$) will be included into the calculations associated to this technique.

Support Vector Machine Classification (SVM) was also considered. The difference between this technique and PCA and CART is that it only separates a group of samples from all the other samples (the ‘others’ group), making it necessary to develop a model per polymer. Despite this being the usual working approach, there are studies that consider simultaneously various SVM models for different classes, although they require specific algorithms (Belousov et al., 2002) and will not be considered here.

SVM was used to enhance the classification capabilities of the wavenumbers associated with the more intense loadings selected in PCA, option ii) above. The conceptual basis of the SVM method is to separate the datapoints (spectra) of a polymer from the other polymers by means of a hyperplane. This is a border or frontier that is defined mathematically in a dimensionality higher than that where the samples were measured (in this case the wavenumbers of the IR spectra). The rationale behind SVM is that if the original variables are not good enough to differentiate classes, they can be operated (by means of the so-called kernels) so that they yield additional mathematical dimensions that, potentially, can differentiate the groups best. Two major parameters have to be optimized for each model: the cost (or penalty) and gamma constants. Interested readers are encouraged to consult additional literature for more technical details; e.g., Cortes and Vapnik (1995), and Brereton and Lloyd (2010).

REFERENCES

- Belousov, A.I.; Verzakov, S.A.; Von Frese, J., 2002. Applicational aspects of support vector machines. *J. Chemom.* 16, 482–489, doi:10.1002/cem.744.
- Brereton, R.G. *Chemometrics: data analysis for the laboratory and chemical plant*. Wiley, 2006.
- Brereton, R.G., Lloyd, G.R., 2010. Support Vector Machines for classification and regression, *Analyst* 135, 230-267.
- Cortes, C.; Vapnik, V., 1995. Support-Vector Networks. *Mach. Learn.*, 20, 273–297.
- Lewis, R.J.; Ph, D.; Street, W.C., 2000. An Introduction to Classification and Regression Tree (CART) Analysis. *Annu. Meet. Soc. Acad. Emerg. Med.* 2000, 14p, doi:10.1.1.95.4103.
- Otto, M. *Chemometrics: statistics and computer application in Analytical Chemistry*. Wiley-VCH, 2017.

SUPPLEMENTARY TABLES

Table SM1: Comparison of wavenumbers used to develop SVM models using ATR spectra.

ATR wavenumbers (cm ⁻¹)			
Pellets		Powder	
Sea	Dry	Sea	Dry
	3027		3024
2951	2951	2950	2950
	2950	2949	
2916			2913
			2867
2849	2849	2848	2848
2848			
		1769	
1720		1718	1718
1630		1630	
1537			
	1492	1492	1492
	1461		
1447	1452		
1376	1376	1375	1375
		1230	
		1187	
		1158	
	695	694	694

Table SM2: Resume of the most relevant wavenumbers selected to characterize each polymer. Data extracted from Brandon et al. (2016), Tiwari et al. (2019), Vasanthan (2012), Veerasingam et al. (2021), Jung et al. (2018).

Polymer	Wavenumber (cm ⁻¹)	Moiety	Polymer	Wavenumber (cm ⁻¹)	Moiety
HDPE	2915	C-H stretch	PVC	1427	CH ₂ bend
	2845	C-H stretch		1331	CH bend
	1472	CH ₂ bend		1255	CH bend
	1462	CH ₂ bend		1099	C-C stretch
	730	CH ₂ rock		966	CH ₂ rock
717	CH ₂ rock	616		C-C stretch	
LDPE	2915	C-H stretch	PS	3024	Aromatic C-H stretch
	2845	C-H stretch		2847	C-H stretch
	1467	CH ₂ bend		1601	Aromatic ring stretch
	1462	CH ₂ bend		1492	Aromatic ring stretch
	1377	CH ₂ bend		1451	CH ₂ bend
	730	CH ₂ rock		1027	Aromatic CH bend
717	CH ₂ rock	694		Aromatic CH out-of-plane bend	
PA	3298	N-H stretch	537	Aromatic ring out-of-plane bend	
	2932	C-H stretch	PP	2950	C-H stretch
	2858	C-H stretch		2915	C-H stretch
	1634	C-O stretch		2838	C-H stretch
	1538	N-H bend, C-N stretch		1455	CH ₂ bend
	1464	CH ₂ bend		1377	CH ₃ bend
	1372	CH ₂ bend		1166	CH bend, CH ₃ rock, C-C stretch
	1274	N-H bend, C-N stretch		997	CH ₂ rock, CH ₃ bend, CH bend
1199	CH ₂ bend	972		CH ₃ rock, C-C stretch	
687	N-H bend, C=O stretch	840		CH ₂ rock, C-CH ₃ stretch	
PC	2966	C-H stretch		808	CH ₂ rock, C-C stretch, C-CH stretch
	1768	C=O stretch	PET	1713	C=O stretch
	1503	Aromatic ring stretch		1241	C-O stretch
	1409	Aromatic ring stretch		1094	C-O stretch
	1364	CH ₃ bend		720	Aromatic CH out-of-plane bend
	1186	C-O stretch	PMMA	2992	C-H stretch
	1158	C-O stretch		2949	C-H stretch
1013	Aromatic CH out-of-plane bend	1721		C=O stretch	
828	Aromatic CH out-of-plane bend	1433		CH ₂ bend	
		1386		CH ₃ bend	
		1238		C-O stretch	
		1189		CH ₃ rock	
		1141		C-O stretch	
		985		CH ₃ rock	
		964		CH ₂ rock, C-H bend	
		750	C=O bend		

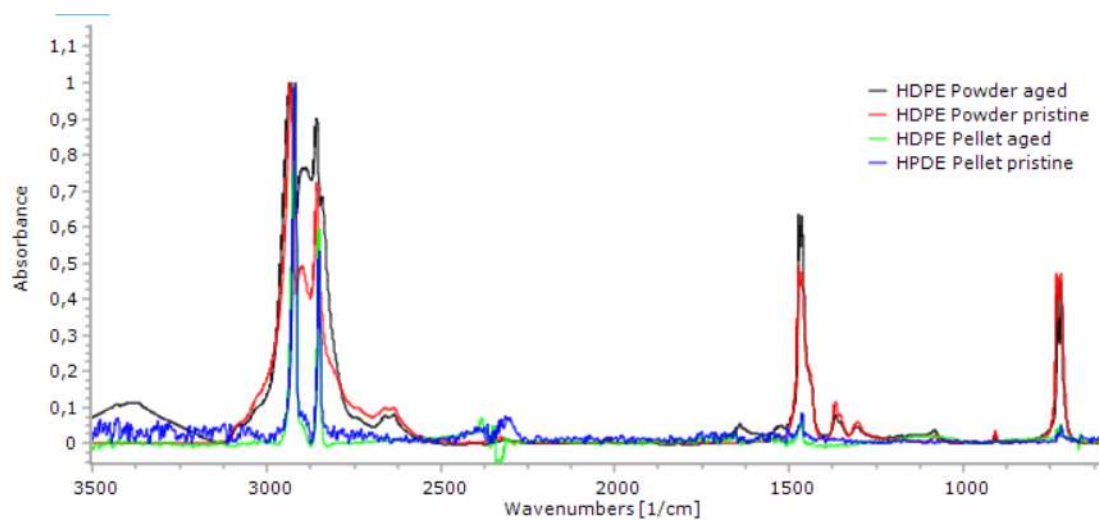
SUPPLEMENTARY FIGURES

Figure SM 1. Comparison between normalized (maximum absorbance=1) micro reflectance spectra of HDPE disposed in pellets and powder configurations.

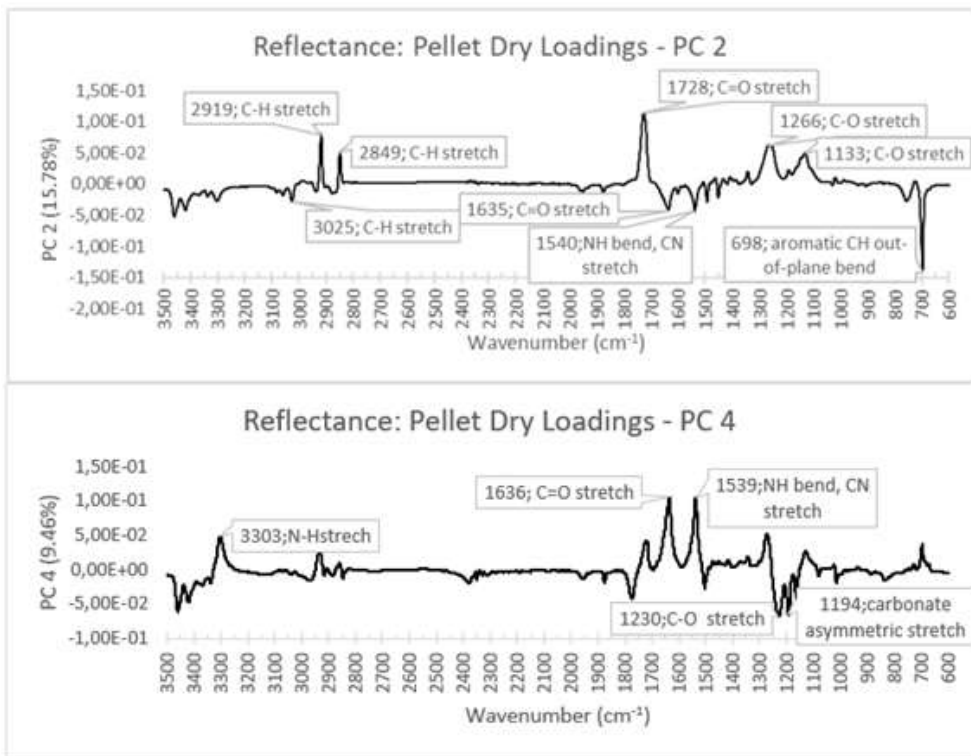


Figure SM 2. Loadings of the PCA performed for pellets weathered in dry conditions, micro reflectance spectrometry.

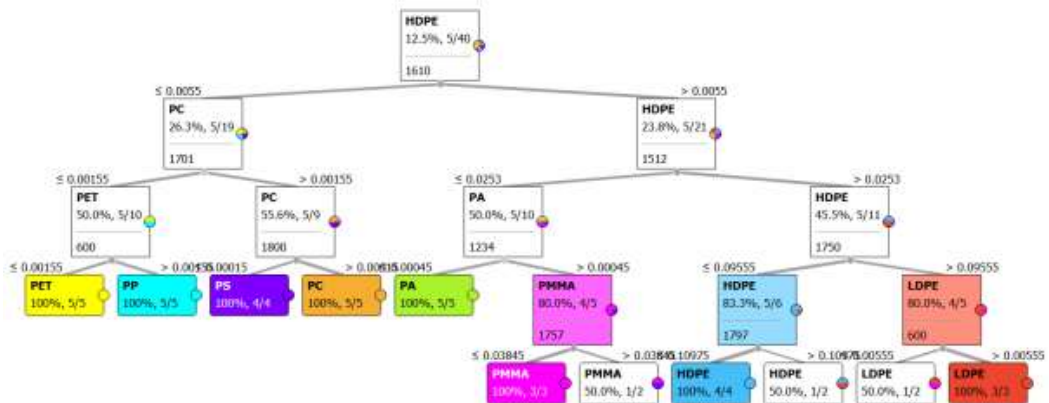


Figure SM 3. Pellets aged in dry conditions, schematic representation of the CART decision tree obtained from the micro reflectance measurements

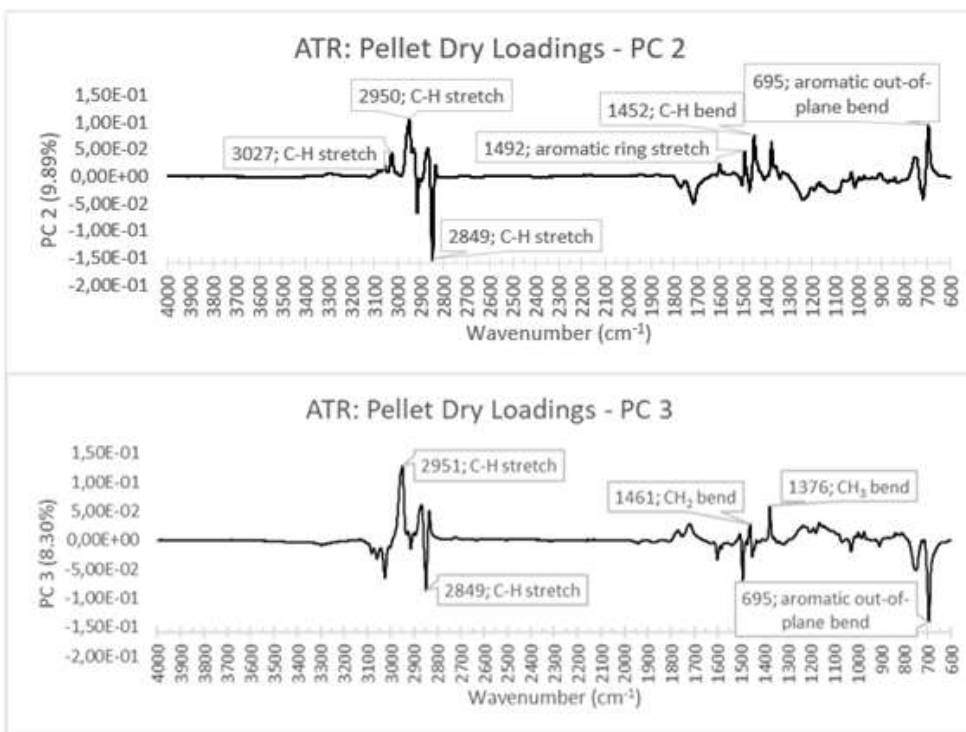


Figure SM 4. Loadings of the PCA performed for pellets weathered in dry conditions, ATR spectrometry

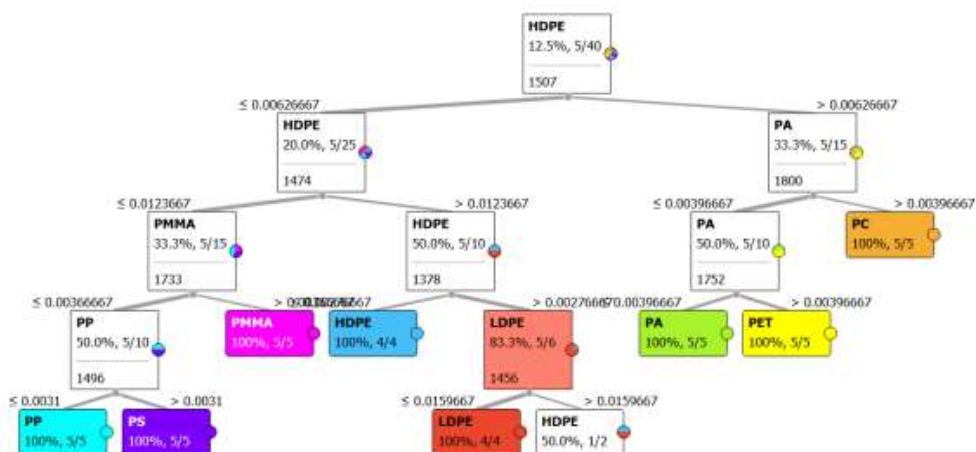


Figure SM 5. Pellets aged in dry conditions, schematic representation of the CART decision tree obtained from the ATR measurements.

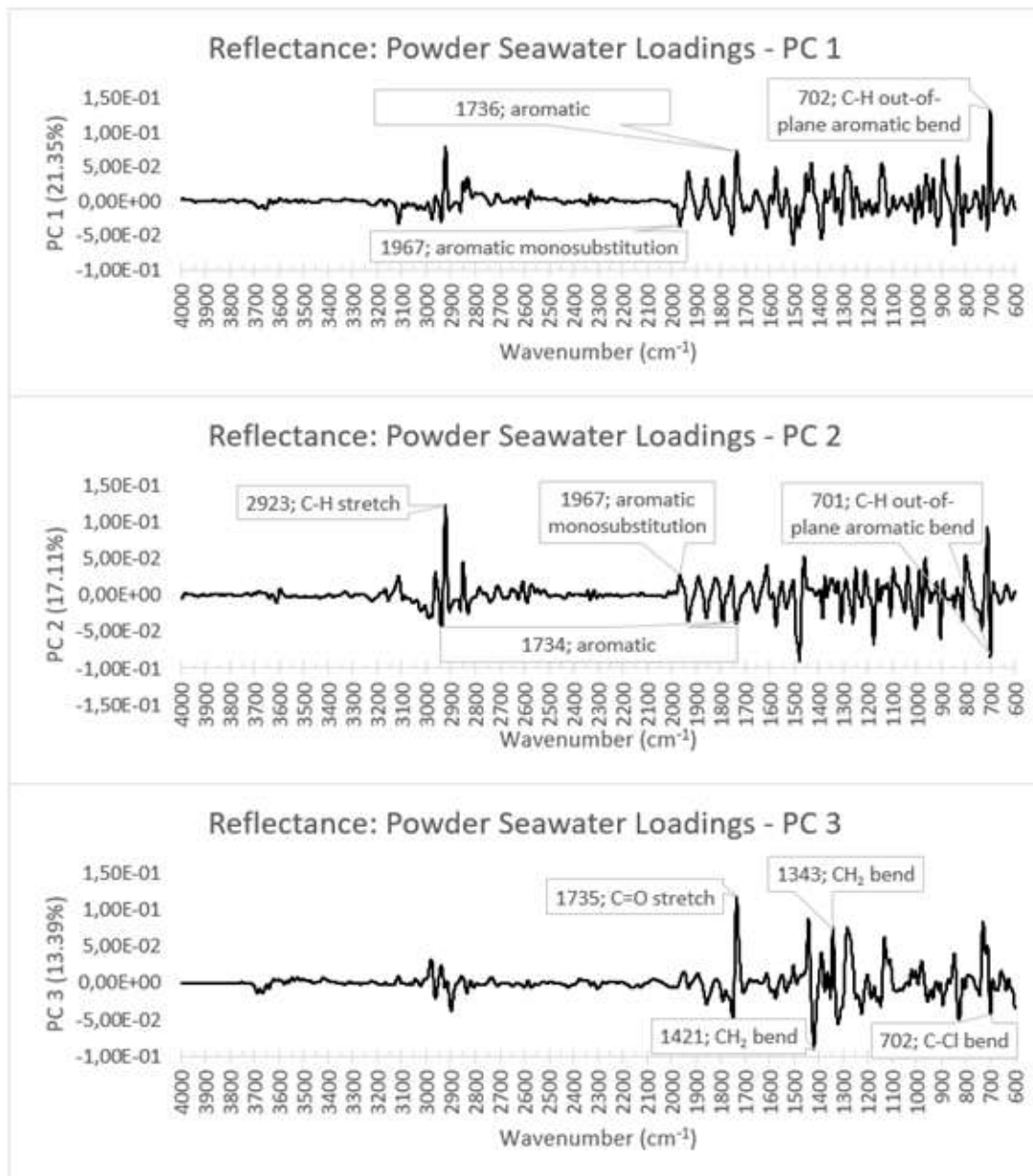


Figure SM 6. Loadings of the PCA performed for powder weathered in seawater, micro reflectance spectrometry.

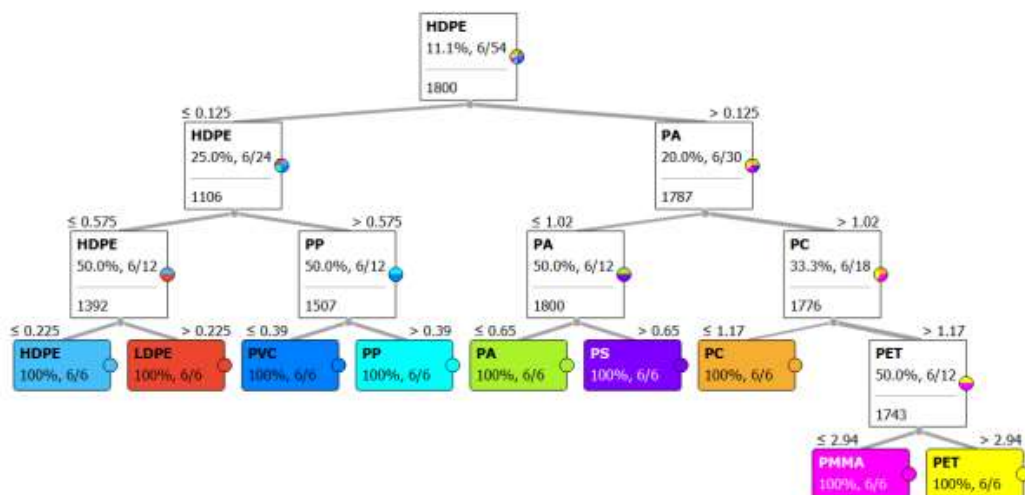


Figure SM 7. Powder aged in seawater, schematic representation of the CART decision tree obtained from the micro reflectance measurements.

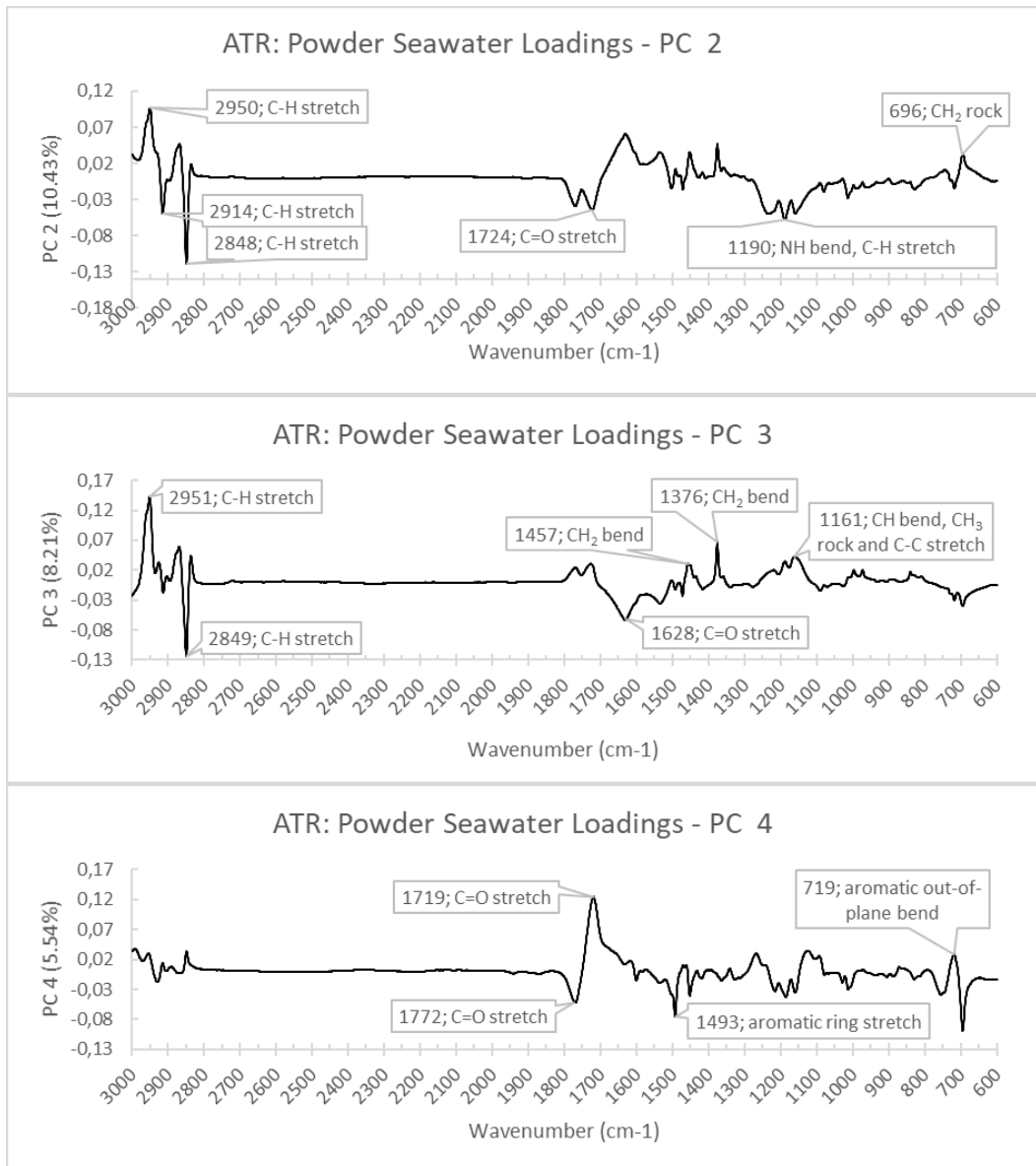


Figure SM 8. Loadings of the PCA performed for powder weathered in seawater, ATR spectrometry.

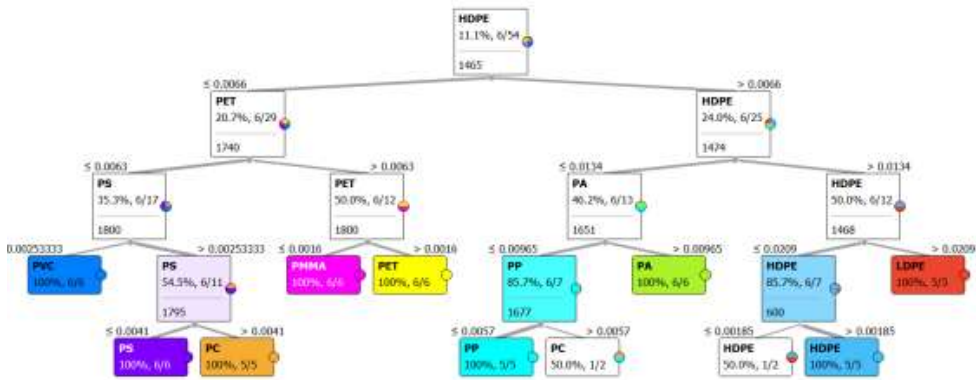


Figure SM 9. Powder aged in seawater, schematic representation of the CART decision tree obtained from the ATR measurements.

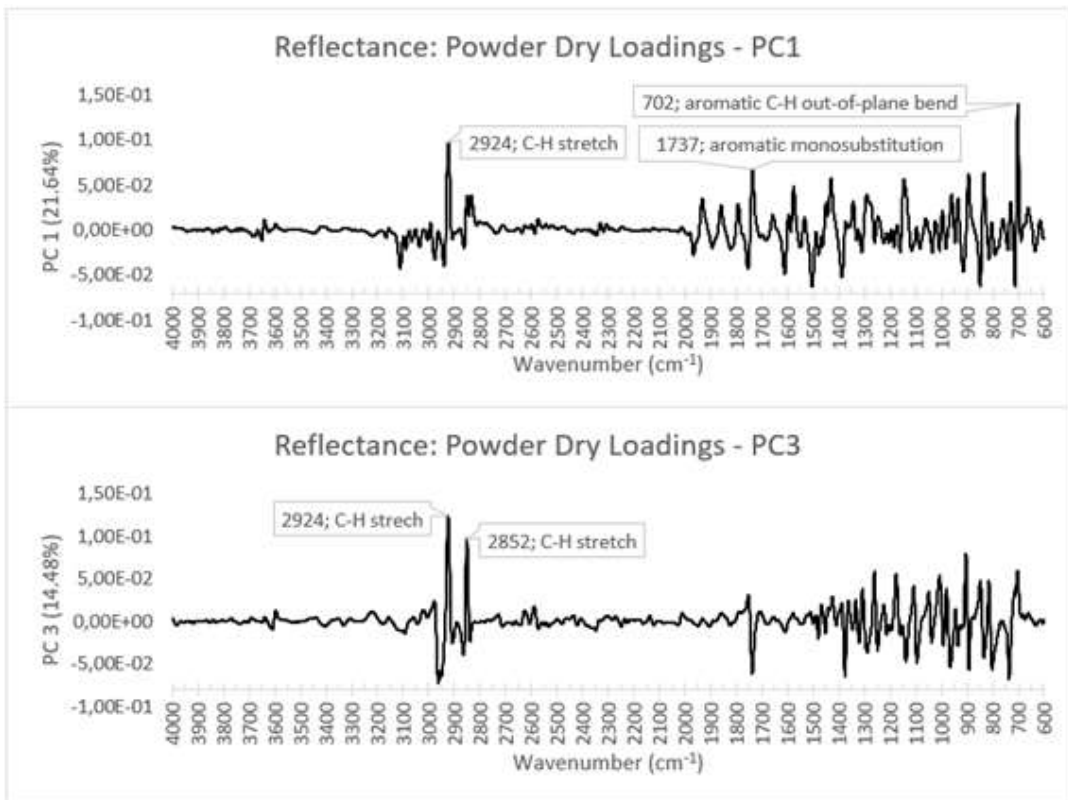


Figure SM 20. Loadings of the PCA performed for powder weathered in dry conditions, micro reflectance spectrometry.

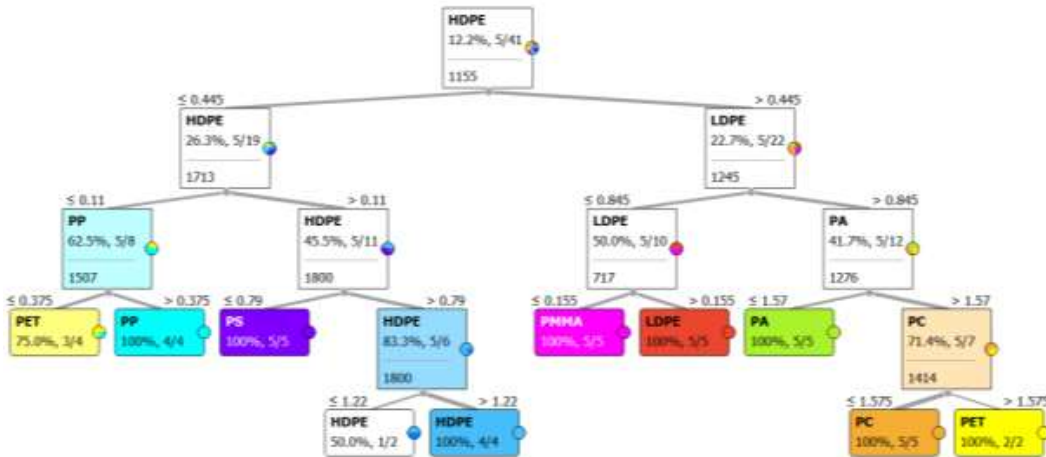


Figure SM 11. Powder aged in dry conditions, schematic representation of the CART decision tree obtained from the micro reflectance measurements.

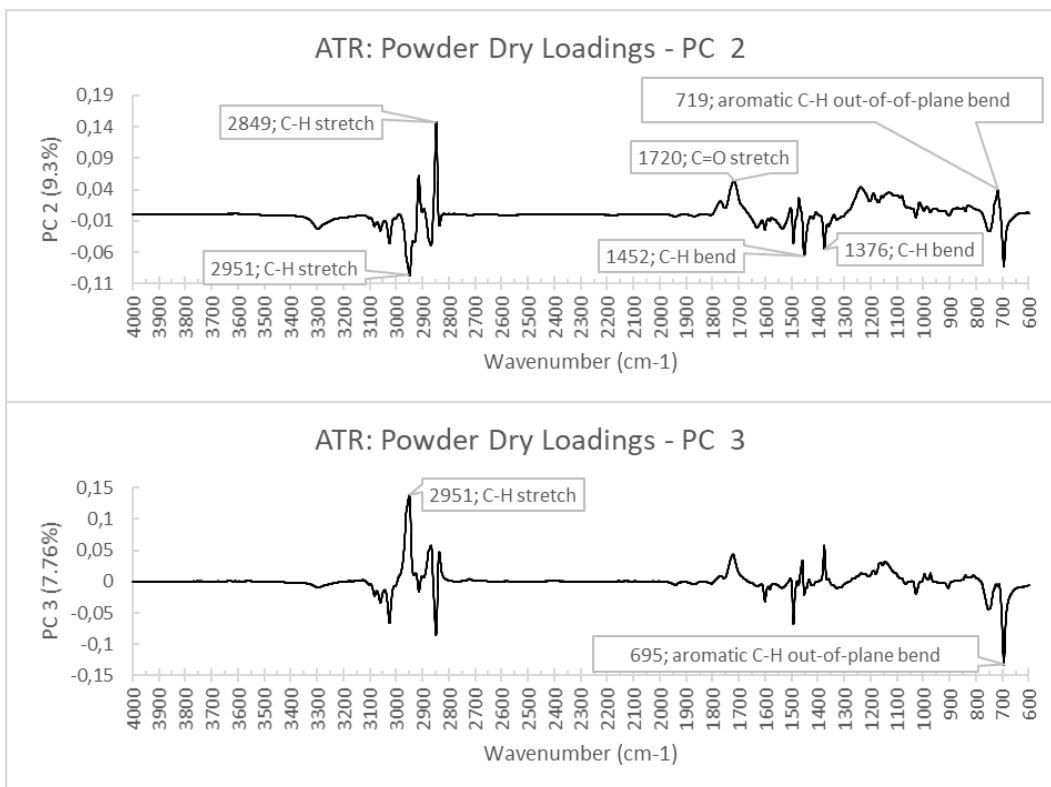


Figure SM 12. Loadings of the PCA performed for powder weathered in dry conditions, ATR spectrometry.

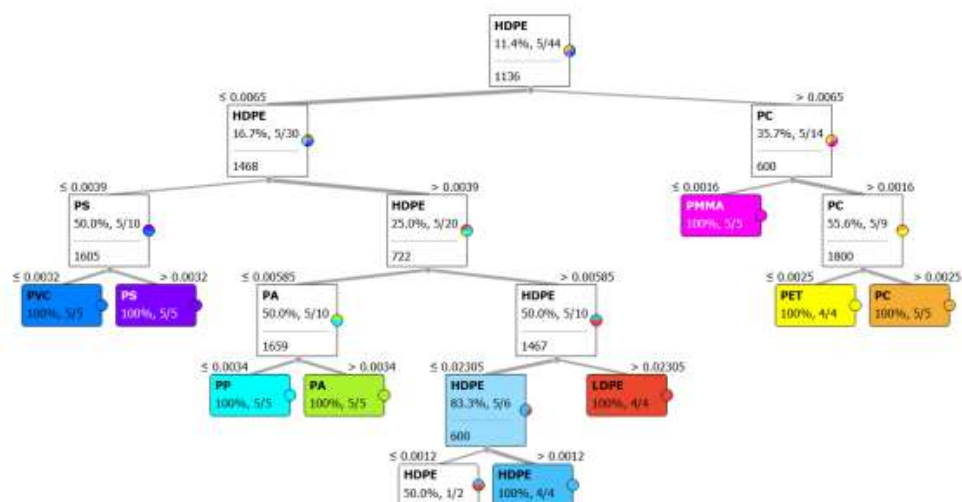


Figure SM 3. Powder aged in dry conditions, schematic representation of the CART decision tree obtained from the ATR measurements.

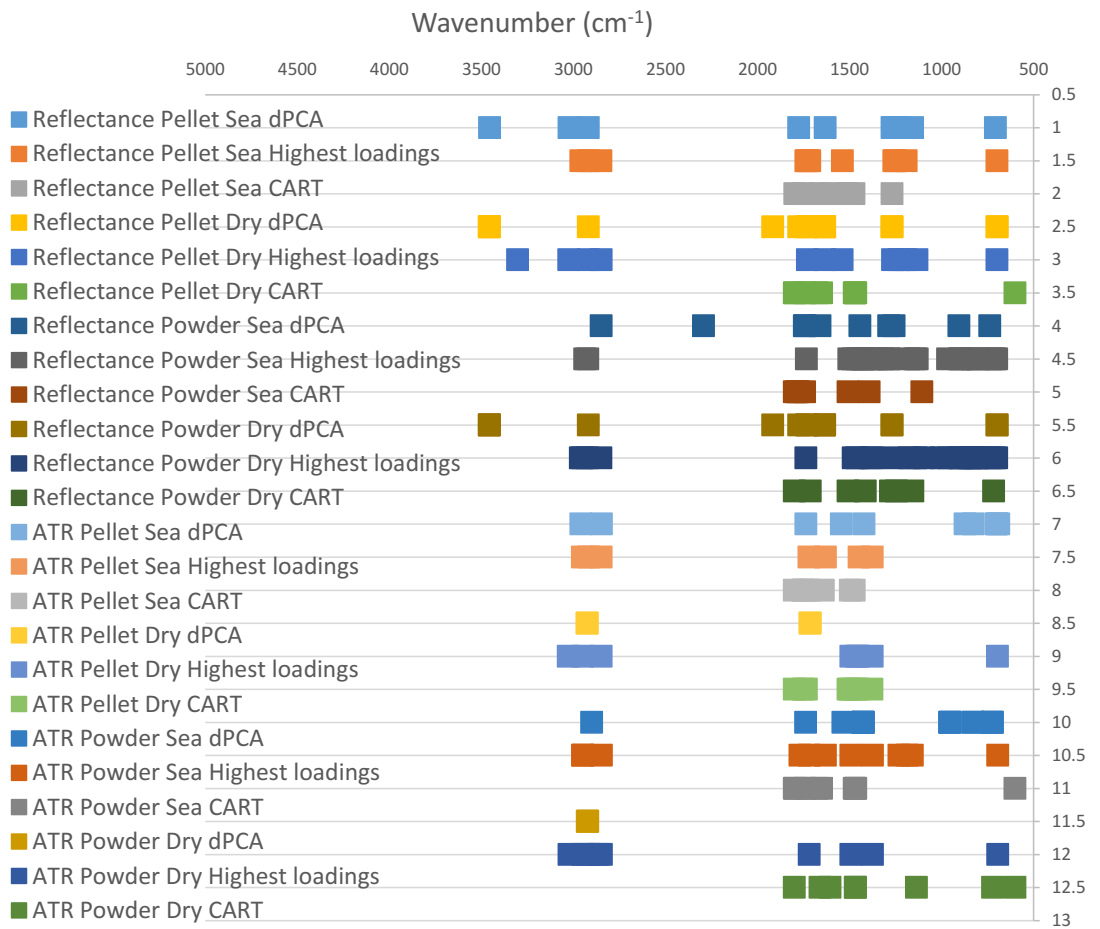


Figure SM14: Comparison of the wavenumbers selected as a function of the different weathering setups and variable selection methods (CART considered wavenumbers only from 1800 to 600 cm⁻¹).

Appendix: Summary in Spanish.



Appendix: Summary in Spanish/Resumen en castellano

En los artículos seleccionados para conformar esta Tesis Doctoral por compendio de publicaciones se busca desarrollar nuevos métodos analíticos para muestras procedentes de dos campos de aplicación diferentes, pero de gran repercusión social. El hilo conductor de todos ellos es la combinación de distintas modalidades de espectrometría infrarroja (IR) en la zona media con técnicas quimiométricas. En primer lugar, se desarrollaron modelos para determinar la composición de muestras comerciales de gas natural (GN), así como su Índice Wobbe (un parámetro relacionado con la miscibilidad de gases con distinta composición relativa), cuyo análisis y monitorización se lleva a cabo, tradicionalmente, mediante cromatografía de gases.

El segundo campo abordado es el de los microplásticos, fragmentos poliméricos de pequeño tamaño (≤ 5 mm), que se han convertido en uno de los tipos de contaminación más estudiados de los últimos años, especialmente en el medio marino. En este caso el objetivo es desarrollar métodos que permitan la identificación de los principales polímeros constituyentes de los microplásticos en muestras reales, con independencia de su nivel de envejecimiento en el medio o los distintos aditivos que puedan tener.

1. Gas Natural

El GN es un combustible fósil muy utilizado para calefacción tanto doméstica como industrial, para la producción de etileno, amoníaco (generalmente para fertilizantes), azufre o carbono negro y, en menor medida aplicaciones para el transporte. Sus procesos de formación son similares a los de otros combustibles fósiles, derivados de la descomposición de materia orgánica enterrada en la corteza terrestre y sometida a condiciones elevadas de presión y temperatura (procesos termogénicos) o procesos biológicos (procesos biogénicos), aunque también pueden derivarse de actividad volcánica (procesos abiogénicos). Generalmente se encuentra en yacimientos delimitados por roca

impermeable, ya sea de forma independiente, en yacimientos asociados de petróleo o carbón, filtrado en minerales de baja permeabilidad o en forma de hidratos.

El GN está compuesto principalmente por metano (75-99 %), etano (0-20 %), propanos, butanos, pentanos y hexanos (<10 %). También puede contener otros componentes indeseados, como gases eluyentes (CO₂, N₂, H₂, O₂, H₂O o gases nobles), contaminantes (Hg, As, o compuestos de azufre) y sólidos en suspensión, que serán eliminados o reducidos durante el procesado industrial del gas. También puede contener trazas de sustancias radiactivas. Esta variabilidad en la composición relativa depende principalmente del yacimiento de origen y los procesos industriales que se le apliquen.

Para su uso, sobre todo en aplicaciones de automoción, es importante conocer en detalle su composición, ya que afecta a su correcta combustión. Una composición inapropiada puede provocar problemas como el picado de bielas o autoencendido, una combustión incompleta en el motor, generando ruido, vibraciones, daños en el motor y pérdida de potencia. Para evitar este efecto es importante que el GN tenga un número de metano apropiado, un parámetro físico-químico análogo al octanaje en gasolina o el índice de cetano en diésel. Otro parámetro importante del GN es el índice Wobbe, que se relaciona con la miscibilidad de mezclas de GN con diferentes composiciones relativas.

Su principal atractivo como fuente de energía es su relativa abundancia y la baja cantidad de gases contaminantes (CO₂, NO_x y SO_x) que emite en comparación con el carbón o los derivados de petróleo. Por ello, se ha considerado una opción viable para los objetivos de decarbonización europeos. Sin embargo, con la invasión de Rusia a Ucrania las importaciones de GN a Europa se han reducido significativamente, ya que Rusia es el principal exportador, siendo uno de los países con más producción y reservas confirmadas. El uso del GN como recurso estratégico y la reticencia de la Unión Europea a financiar la maquinaria bélica rusa ha llevado a que se aspire a reducir el uso del GN como fuente de energía, con la intención de reemplazarlo completamente antes de 2027.

Normalmente los análisis de GN se llevan a cabo mediante cromatografía de gases, ya que es una técnica bien establecida con normas definidas a nivel internacional. De hecho, permite realizar análisis más precisos y sensibles, con límites de detección y cuantificación más bajos que las técnicas espectrométricas en su estado actual. Sin embargo, la cromatografía de gases tiene una serie de desventajas, como sus altos costes de reactivos, calibración, mantenimiento, elevado tiempo de dedicación de los operarios y su poca adaptabilidad para sistemas *on-line*. Si bien, a nivel comercial, se han desarrollado sistemas portátiles para medidas *in-situ*, se necesita disponer los gases portadores cerca de la localización donde se realice la medida. Por ello, la espectrometría IR podría considerarse una alternativa atractiva: el análisis de una muestra de gas puede realizarse en ~30 minutos, es fácilmente adaptable a sistemas *on-line*, y no requiere reactivos (aunque se pueden añadir gases no activos en la zona IR para mejorar la calidad de los espectros).

2. Microplásticos

El segundo campo de trabajo abordado en la Tesis está constituido por los microplásticos, fragmentos poliméricos de entre 5 mm y 1 μm , que han llamado la atención tanto de la comunidad científica como de la población general debido a la ingente cantidad que se ha generado en las últimas décadas.

Los plásticos en sus inicios se desarrollaron para aplicaciones militares durante la Segunda Guerra Mundial, en paracaídas, cuerdas o ventanas de aviones. En la década de los 50 se comenzó su producción en masa para aplicaciones comerciales, sustituyendo en gran medida a otros materiales como madera, el cuerno, el acero o el vidrio. En décadas posteriores su uso se fue aumentando exponencialmente, llegando a producirse 391 millones de toneladas en 2021.

Los polímeros están formados por cadenas regulares de moléculas orgánicas, llamadas monómeros, que definen, en gran medida, sus propiedades físicas y químicas. Pueden ser tan simples como cadenas lineales de etilenos en el polietileno (PE), o contener moléculas más

complejas con heteroátomos y grupos aromáticos como en el poliuretano (PUR).

Como material, el plástico resulta tremendamente atractivo para la sociedad: su gran maleabilidad, versatilidad y estabilidad, así como sus grandes posibilidades de funcionalización *ad-hoc* o bajos costes de producción hacen que tenga un amplio abanico de aplicaciones. Por ejemplo: en envasado, embotellado, textiles, construcción, industria automovilística, electrónica, adhesivos, etc. Esto provoca que la producción de plásticos se haya disparado en los últimos 70 años, con la consiguiente aparición de grandes cantidades de residuos. De hecho, la contaminación por desechos de plástico es especialmente preocupante: el hecho de que muchas de sus aplicaciones sean de un solo uso conduce a que se generen grandes cantidades de residuos y su estabilidad provoca que permanezcan en el medio natural durante decenas o incluso cientos de años. Además, son susceptibles de fragmentarse en piezas cada vez más pequeñas generando micro y nanoplásticos, que pueden acarrear otros problemas añadidos en relación a su toxicidad para con los seres vivos.

Existen dos tipos de microplásticos, dependiendo de su origen. Los microplásticos primarios son plásticos que se fabrican específicamente en tamaño reducido. Generalmente se usan en cosméticos, productos de higiene corporal, o abrasivos industriales. En vista de la gran cantidad de residuos que generan, varios países han regulado su producción (especialmente en la industria cosmética), como Reino Unido, Estados Unidos, la Unión Europea, o Tailandia. También se consideran microplásticos primarios los pellets de preproducción, que se usan como material base en la producción de diversos objetos, y que pueden verterse al medio en grandes cantidades por accidente en su transporte o almacenaje.

Los microplásticos secundarios se derivan de la fragmentación de macropelásticos en el medio. Generalmente se producen por fotodegradación (reacciones químicas generadas por exposiciones prolongadas a la luz ultravioleta), aunque también por erosión o acción biológica. Algunas de las fuentes de microplásticos secundarios se dan

en vertederos, zonas de construcción o derivadas de la fragmentación de redes y aparejos de pesca, textiles o desgaste de neumáticos.

La contaminación por microplásticos está presente en la mayoría de compartimentos medioambientales a nivel mundial. Además, se ha probado que tienen potenciales efectos nocivos en la fauna y flora y que pueden introducirse muy fácilmente en la cadena trófica, pudiendo alcanzar a la alimentación humana. Sin embargo, actualmente las propiedades más nocivas no se atribuyen al polímero en sí, sino a los distintos tipos de aditivos usados en su producción o los contaminantes adsorbidos en el medio en el que se encuentren. Por ejemplo, el Bisfenol A o el di-(2-etilhexil)ftalato (DEHP), aditivos comunes en la producción de diversos plásticos, tienen efectos adversos en el sistema endocrino y la salud reproductiva, así como efectos carcinogénicos.

La mayoría de las publicaciones que tratan el tema de la contaminación por microplásticos se centran en sus efectos en el medio marino. Esto se debe a que la mayoría de los residuos acaban en mares y océanos, ya sea mediante el transporte por ríos, arrastrados por la lluvia o por transporte atmosférico, siendo capaces de atravesar las plantas de tratamiento de aguas. Una vez en el medio, dependiendo de su densidad pueden flotar cerca de la superficie o hundirse, mezclándose con el sedimento, pasando en ambos casos a ser fácilmente accesibles para la fauna local. De hecho, desde 2010 ya existen más de 1500 estudios que han encontrado partículas de microplásticos en organismos marinos.

También se han evidenciado sus efectos en suelos, donde afectan a los organismos que regulan la descomposición de la materia orgánica, necesarios para el reciclaje de nutrientes, imprescindibles para el desarrollo de la flora. Las fuentes de contaminación más comunes de microplásticos en suelos derivan de la agricultura y la gestión de residuos. En suelos destinados a la agricultura las fuentes más comunes son el uso de mantillos de plástico, partes de invernaderos o el uso de lodos de depuradora como fertilizante. En áreas urbanas muchos microplásticos se generan por el desgaste de neumáticos.

Los microplásticos también influyen en la contaminación atmosférica, ya que pueden ser transportados fácilmente por el aire, pudiendo permanecer en suspensión grandes distancias. Las partículas más

pequeñas son susceptibles de ser inhaladas. Sus fuentes principales son partículas textiles, desgaste de neumáticos y actividades industriales. Su permanencia en la atmósfera depende de varios factores, como su tamaño, su densidad, su morfología, la velocidad del viento, precipitaciones u obstáculos físicos. En general son más prevalentes en zonas urbanas que en rurales.

En la actualidad, la técnica más utilizada para el análisis y monitorización de microplásticos en el medio ambiente es la espectrometría vibracional. Tiene, como ventaja, que cada polímero presenta un espectro específico, por lo que, a priori, es fácil identificarlos si no han sufrido modificaciones muy significativas. Sin embargo, es difícil identificar los procesos específicos de degradación o envejecimiento que tienen lugar en los polímeros debido a las distintas condiciones medioambientales. Además, esas alteraciones pueden tener lugar no solo en la cadena polimérica sino también en sus aditivos. Todo ello complica la identificación de los fragmentos sospechosos de ser microplásticos.

3. Espectrometría

Como ya se ha comentado, la técnica a utilizar en esta Tesis Doctoral para el análisis de ambos tipos muestras es la espectrometría IR. Esta es una técnica que se ha ido desarrollando durante varios siglos, desde que el astrónomo William Herschel descubrió la radiación IR de la luz solar en 1800. Sin embargo, no fue hasta mediados del siglo XX que la espectrometría IR se empezó a utilizar, primero en aplicaciones militares y, posteriormente, de forma comercial, principalmente para el análisis de endorfinas y caucho sintético. Desde entonces sus aplicaciones han aumentado significativamente, empleándose de manera sistemática en campos tan diversos como medicina, farmacia, estudios medioambientales y múltiples aplicaciones industriales.

En la actualidad, la espectrometría IR es una de las técnicas analíticas más utilizadas. Puede ser usada en casi todo tipo de muestras, con la excepción de gases nobles, O₂, H₂ y N₂, que son invisibles a la radiación IR. A su éxito ha contribuido mucho el hecho de que cada sustancia

tiene un espectro diferente (excepto isómeros ópticos), convirtiendo la espectrometría IR en una herramienta muy versátil en cuanto a sus aplicaciones.

Conceptualmente, la espectrometría se basa en medir la interacción de la radiación con la materia. En el caso de la espectrometría vibracional, se usan números de onda de 14000 a 4 cm^{-1} (400-4 cm^{-1} : infrarrojo lejano, 4000-400 cm^{-1} : infrarrojo medio, 14000-4000 cm^{-1} : infrarrojo cercano).

Cuando la radiación impacta en una muestra, las diferentes longitudes de onda son reflejadas o absorbidas. Es midiendo estas interacciones como se obtienen los espectros. En el caso de la radiación IR absorbida, esa energía se emplea para excitar los niveles vibracionales y rotacionales de los enlaces en las moléculas. Estas no son estructuras estáticas, si no que están en continuo movimiento, tanto sus átomos como la longitud y orientación de los enlaces químicos. Cuando un haz de radiación con la longitud de onda adecuada impacta con la molécula sus niveles de energía cambian, modificando las tensiones y flexiones de los enlaces de un estado fundamental a uno excitado (más energético). Las moléculas con un mayor número o diversidad de átomos tienden a exhibir transiciones rotacionales y vibracionales más diversas, lo que se traduce en espectros más complejos con más picos o bandas.

Existen distintos tipos de modalidades de medida, dependiendo de cómo se determine la interacción entre la muestra y la radiación.

El modo de transmisión es el más tradicional. En este caso el objetivo es que el haz de radiación IR pueda atravesar la muestra e impactar en el detector. En este modo de medida la tarea más complicada es, generalmente, la preparación de la muestra, la cual puede adoptar distintas configuraciones, ya sea como disolución, película fina, integrada en pastillas de KBr o en una suspensión de Nujol. En el caso de los gases, se usa una celda con ventanas transparentes al IR.

La reflectancia es un método más apropiado para la medida de muestras opacas o sólidos gruesos, ya que se mide la radiación reflejada en su superficie. Por ello, los espectros obtenidos son muy susceptibles a la morfología de la muestra, pudiendo darse una reflexión difusa (en caso de polvo o superficies irregulares) o una reflexión especular (en

superficies lisas y homogéneas). Cada uno de estas situaciones requiere un modo distinto de tratamiento de los datos para obtener un espectro interpretable. Precisamente, uno de los retos para el analista es decidir cuál es el más apropiado, ya que no suele haber muestras con reflexión puramente especular o difusa.

Los fundamentos de la técnica de reflectancia total atenuada (ATR) son técnicamente más complejos, ya que se basa en la interacción del campo electromagnético evanescente del haz de radiación que se genera en la interfase entre dos materiales (siendo uno de ellos la muestra). Cuando un haz de radiación atraviesa un medio e incide en otro, parte de la radiación se refleja y parte se transmite en un determinado ángulo. Si el haz incidente supera un ángulo crítico toda la radiación es reflejada internamente, por lo que no existe pérdida de energía por transmisión, pero aún hay una pequeña pérdida de energía en forma de “onda evanescente” que se disipa en el segundo medio, en nuestro caso, la muestra. Para conseguir este efecto, la muestra se coloca sobre un material con alto índice de refracción (ZnSe, germanio, diamante, etc.) que es atravesado por la radiación, siendo la interacción de la muestra con la onda evanescente la que permita la adquisición del espectro.

Además de estos modos de operación, se suele hibridar la espectrometría IR con otras técnicas analíticas para obtener información adicional. Una de ellas es la microscopía IR (micro-IR), que combina la espectrometría IR con la microscopía. Es especialmente útil para la visualización y análisis de partículas de hasta $10\ \mu\text{m}^2$ (aunque a partir de $30\ \mu\text{m}^2$ hay gran cantidad de ruido instrumental). La mayoría de los equipos de micro-IR pueden trabajar en modo transmitancia, reflectancia o ATR, redireccionando el haz de radiación mediante una serie de espejos móviles. Esta técnica permite el análisis no solo de partículas muy pequeñas, sino de muestras muy heterogéneas o hechas de materiales compuestos, como en aplicaciones biológicas, forenses, de control de calidad o en microplásticos.

A pesar de la gran utilidad de la información que aportan las distintas técnicas de espectrometría IR, en muchos casos la información que brindan es incompleta o es difícil de interpretar. Por ejemplo, el GN, al ser una mezcla de varios gases relativamente similares tanto en

composición como en características espectrales, genera espectros muy complejos, con una gran cantidad de picos superpuestos correspondientes a distintos analitos. En el caso de los microplásticos, el envejecimiento y los materiales adsorbidos o las distintas reacciones de degradación que el polímero pueda haber sufrido en el medio puede dificultar su identificación. Para mitigar estos problemas en esta Tesis Doctoral se propone el uso de distintos modelos quimiométricos desarrollados sobre los datos espectrales.

4. Quimiometría

La quimiometría se define como la aplicación de técnicas matemáticas, estadísticas y de lógica formal al diseño de experimentos e interpretación de datos químicos. La IUPAC la define como la aplicación de la estadística al análisis de datos químicos (de química orgánica, analítica o medicinal) y al diseño de experimentos y simulaciones químicas.

Su desarrollo comenzó en la década de los 70 con pioneros como Herman y Svante Wold, Bruce Kowalski, Paul Geladi o Michele Forina. En sus inicios era una disciplina puramente teórica, íntimamente ligada con la computación y el desarrollo de instrumentos analíticos. A partir de la década de los 80 empezaron a aparecer las primeras publicaciones que empleaban la quimiometría para resolver problemas simples, principalmente de espectrometría y cromatografía.

En los años siguientes su relevancia creció, tanto en aplicaciones como en complejidad de los problemas que solucionaba. Esto atrajo industrias, como la farmacéutica o alimentaria, que propiciaron muchos avances y nuevas estrategias de trabajo. Actualmente también se aplica en petroquímica, metabolómica, genética, biología, medicina, datos medioambientales, ciencia forense, etc.

Existen dos ramas principales de la quimiometría, dependiendo de su área de aplicación: el diseño de experimentos y el análisis multivariante de datos. El primero se centra en la optimización de los parámetros de un proceso experimental (una síntesis, por ejemplo). Su objetivo principal es la obtención de la mayor cantidad de información posible

realizando el menor número de experimentos. Con ello no solo se reducen el tiempo y costes de los procesos, si no que se genera información acerca de las sinergias entre los distintos parámetros experimentales.

El análisis multivariante de datos, por otra parte, se usa para extraer la mayor cantidad posible de información de tablas de resultados. Por ejemplo, en espectrometría o cromatografía pueden obtenerse tablas de datos de miles de variables para cientos de muestras. Esta ingente cantidad de datos pueden ser difíciles de interpretar.

En los últimos años se han desarrollado un gran número de técnicas para clasificación, cuantificación y reconocimiento de patrones. En esta Tesis se usaron cuatro metodologías: el análisis de componentes principales, regresión parcial por mínimos cuadrados, arboles de regresión y clasificación y máquinas de vectores de soporte.

El análisis de componentes principales (PCA) es una de las técnicas más usadas en quimiometría. Se usa para realizar análisis no supervisado (solo se tienen en cuenta las variables experimentales). Esta técnica se basa en concentrar la información inherente a los datos en un nuevo conjunto reducido de variables (denominadas componentes principales) calculadas mediante combinaciones lineales de las variables originales. Cada componente principal explica parte de la información o varianza de los datos. El analista tiene que seleccionar un número adecuado de componentes teniendo en cuenta, entre otras cosas, su varianza explicada. La representación gráfica de los scores (parámetros calculados para cada una de las muestras) de dos o tres componentes principales permitirá evaluar la existencia de grupos de muestras o pautas

La regresión parcial por mínimos cuadrados (PLS-R), es un método de regresión multivariante supervisado, especialmente útil para aplicaciones donde hay un gran número de variables descriptoras X (predictores) que se relacionarán con otro/s parámetro/s Y (predictandos). Su versatilidad ha hecho que se aplique no solo en quimiometría, también en bioinformática o aprendizaje automatizado (machine learning).

Existen distintos algoritmos para calcular modelos PLS-R pero el más común, y la base de muchas otras opciones es el *orthogonal score PLS-R*. El proceso es similar al de PCA, en el sentido de que se proyectan las variables hacia un número menor de factores, en este caso denominados variables latentes (LVs) o componentes de PLS que, a diferencia de PCA, tienen en cuenta los valores de los predictandos. PLS-R se ha convertido en una técnica estándar *de-facto*, muy útil para predecir propiedades difíciles de medir instrumentalmente de otra forma.

Los árboles de regresión y clasificación son técnicas supervisadas basadas en la creación de particiones binarias en estructuras tipo árbol, en la que cada partición, o *nodo* se divide en dos subramas. En cada uno de los nodos se describe una condición, por ejemplo, un número de onda y una intensidad específica, de modo que las muestras se dividen entre las que superan esa intensidad y las que no, asignando cada una de las muestras a una o varias categorías. El proceso se repite hasta que, idealmente, todas las muestras son separadas en sus correspondientes categorías o se cumplen otras condiciones marcadas por el usuario. Existen variantes destinadas a realizar modelos predictivos (de regresión).

Las máquinas de vectores de soporte (SVM) son un algoritmo de clasificación supervisado (aunque también se pueden usar para regresión y detección de anómalos). Originalmente fueron desarrollados para separaciones de dos categorías, por lo que en escenarios con múltiples clases se establece un modelo para cada una de ellas, separándola del resto (aproximación 1-vs-todos). Posteriormente se generan algoritmos capaces de separar múltiples clases simultáneamente, pero suelen dar resultados ligeramente peores. Fundamentalmente, SVM genera una división mediante el cálculo de una “frontera” (una línea, plano o hiperplano en dimensiones más altas). Esta frontera se coloca de forma que la distancia (márgenes) con los puntos de datos más cercanos se maximice. Una peculiaridad de este algoritmo es que solo tiene en cuenta las muestras más cercanas a la separación.

En problemas reales lo más común es que no se pueda establecer una división lineal perfecta entre las clases. En casos donde los grupos no son separables linealmente (justamente aquellos para los que se indica

esta técnica), se aumenta la dimensionalidad matemática, Para ello se añaden nuevas variables calculadas aplicando una función “kernel” (o función de núcleo) a las originales, siendo los kernel más comunes funciones lineales, polinomiales o radiales (gaussianas). La idea conceptual es que, al trabajar en una dimensionalidad mayor, se pueda crear una separación entre grupos que no habría sido posible en la dimensionalidad original.

Otra aplicación relevante de la quimiometría es determinar cuáles son las variables más importantes de un conjunto de datos para un propósito específico. Aunque existe una gran cantidad de técnicas disponibles para esta finalidad, las que se han aplicado en esta Tesis Doctoral son:

Regresión parcial por mínimos cuadrados en intervalos (i-PLS): Consiste en dividir el conjunto total de variables originales en intervalos de una determinada amplitud (seleccionada por el analista). Se realiza una regresión PLS en cada uno de ellos. El intervalo que genere un mejor modelo es seleccionado. El proceso se repite, combinando el intervalo previamente seleccionado con cada uno de los demás y, si el modelo mejora para alguna de esas combinaciones, se añade un intervalo adicional a la selección. Siguiendo esta fórmula se pueden seleccionar más intervalos hasta que el modelo deje de mejorar o se llegue a un límite seleccionado por el analista. La selección final reflejaría las variables que ofrecen una mayor cantidad de información. La desventaja más acusada de este método reside en la capacidad de computación necesaria y los elevados tiempos necesarios para realizar el estudio. Esto es especialmente acusado si se seleccionan intervalos pequeños, ya que fuerza al ordenador a hacer más modelos PLS. A pesar de ello, con una sección de parámetros adecuada, iPLS puede ser una técnica útil para realizar una selección de variables satisfactoria.

El *Selectivity Ratio Index*, traducido como “relación de selectividad”, (SRI): Es un parámetro que mide la relación entre la varianza explicada y la residual por cada variable en un modelo de PLS. Aunque no es un método de selección en sí mismo, es un parámetro numérico que permite establecer un ranking de las variables, lo que permite establecer una selección manual. Valores de SRI superiores a 1 indican una relación positiva de varianza explicada e importancia en ese modelo. La mayor

desventaja de este método es que cada variable se evalúa de forma individual, ignorando las posibles sinergias entre ellas.

Importancia de la variable en la proyección (VIP): Es un parámetro calculado para cada variable en un modelo PLS. Se define como la suma de los “pesos” de las variables en cada factor PLS, ponderados con el porcentaje de varianza explicada en Y en cada variable latente (factor). De esta forma las variables originales se jerarquizan de acuerdo con su capacidad para predecir mediante el modelo PLS. Del mismo modo que para SRI, los VIPs más altos se seleccionan *ad-hoc*.

Eliminación de variables no informativas usando muestreo reiterado mediante Monte-Carlo: Se generan de forma aleatoria una gran cantidad de subgrupos de muestras, desarrollándose modelos PLS para cada uno de ellos. Se calcula un índice de fiabilidad para cada variable a partir del número de veces que presenta un valor positivo elevado en los coeficientes de regresión de los modelos, de forma que los índices más altos corresponden a las variables que contienen más información útil para la regresión de los modelos. Al realizar distintos modelos con muestras aleatoriamente seleccionadas se disminuye el riesgo de sobreajuste al conjunto de datos original.

Random frog: Se basa en comparar los modelos PLS obtenidos empleando subconjuntos de variables seleccionadas aleatoriamente. Se desarrolla un primer modelo con w variables y se calculan sus errores de validación cruzada. Se elige aleatoriamente un conjunto de esas variables y también se calcula un modelo PLS y su error de validación. Si es mayor se descarta ese subconjunto; sino, se acepta y sustituye a la selección anterior. El proceso se repite un número seleccionado de iteraciones (cuantas más, mejor) y se calcula una probabilidad de selección para cada variable basada en el número de veces que aparece en un modelo aceptado.

iPLS, SRI y VIP se incluyen en el paquete PLS-Toolbox de MATLAB, mientras que MCUVE y *Random frog* se incluyen en el paquete de libPLS.

Selección manual: Al realizar un PCA, es posible utilizar los loadings para determinar las variables más importantes de cada conjunto de datos.

El proceso se establece de la siguiente forma: al desarrollar un PCA con todas las variables experimentales la representación de los scores de los componentes principales permite ver qué componente/s principal/es separa/n mejor cada categoría. Los “loadings” más altos de esos componentes principales corresponderían a las variables que separan mejor cada categoría. La mayor desventaja de este método es la inherente subjetividad de la selección y la falta de un criterio numérico relevante.

PCA dinámico: Está basado en realizar un test *t* de Student para cada variable. El test evalúa la importancia estadística de cada variable para diferenciar un grupo de muestras del resto. La idea es que solo las variables estadísticamente significativas se usen para desarrollar un nuevo modelo.

La representación gráfica de los “scores” del modelo PCA realizado con las variables seleccionadas (las estadísticamente significativas) permite al analista evaluar la separación de las muestras inmediatamente después de la selección, pudiendo incluso elegir un número menor de variables que permitan una diferenciación aceptable entre los grupos.

Este algoritmo forma parte del programa Genex de Multid.

No existe un método universal de reducción de variables, ya que los resultados obtenidos dependen mucho de los problemas a considerar y la naturaleza y las características de los datos. Por ello es importante seleccionar un método adecuado para cada situación específica.

Para el análisis de gas natural se utilizaron modelos de PLS combinados con técnicas de selección de variables: MCUVE, iPLS, Random frog, SRI y VIP.

En el caso de los microplásticos, se probaron varias técnicas con el objetivo de identificar distintos polímeros, con independencia de su envejecimiento, tanto en agua de mar como en seco. Finalmente, los resultados publicados son se obtuvieron mediante PCA, SVM y CART, con selección de variables por PCA dinámico y selección manual.

5. Conclusión

Como conclusión general, en esta Tesis Doctoral se han aplicado diferentes técnicas espectrométricas en la zona infrarroja media y quimiométricas en dos campos socialmente relevantes. El objetivo fundamental es mejorar procedimientos analíticos existentes o sugerir otros nuevos. El Capítulo 2 estudia el efecto en los espectros de muestras de GN al mezclarlas con varios gases inertes, obteniendo picos espectrales más intensos y definidos. En los Capítulos 3 y 4 se crearon modelos para predecir la composición química e índice Wobbe de diversas muestras comerciales de GN.

La premisa del Capítulo 5 es señalar la falta de información instrumental encontrada en muchas publicaciones relacionadas con el análisis de microplásticos, así como proponer una forma unificada de transmitirla. El Capítulo 6 estudia el efecto del envejecimiento ambiental sobre la poliamida 6.6. Para ello se analiza la evolución de su espectro IR y la morfología de las partículas, usando un microscopio de barrido electrónico (SEM). En el Capítulo 7 se obtuvieron los espectros de los 9 polímeros más producidos a nivel global en varias formas (gránulos y polvo) y grados de envejecimiento. Usando varias técnicas quimiométricas, los espectros permitieron desarrollar modelos que predijeron el polímero que actuaba como componente principal de fragmentos obtenidos en muestras reales.

Aunque en los dos campos considerados se pueden realizar nuevas aplicaciones o buscar nuevos objetivos, los trabajos presentados en esta Tesis Doctoral demuestran que la combinación de la espectrometría IR en la zona media, empleando diversas modalidades de medida, y distintas técnicas quimiométricas puede generar nuevos métodos de análisis fiables para el control de calidad con costes relativamente moderados, bajo consumo de reactivos y reducción en carga de trabajo y tiempos de análisis.