Data in Brief 55 (2024) 110612

Contents lists available at ScienceDirect

Data in Brief

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

Data Article

FTIR-Plastics: A Fourier Transform Infrared Spectroscopy dataset for the six most prevalent industrial plastic polymers

Octavio Villegas-Camacho[a], Roberto Alejo-Eleuterio[a,*], Iván Francisco-Valencia[a], Everardo Granda-Gutiérrez[b], Sonia Martínez-Gallegos[a], Javier Illescas[a]

[a] Tecnológico Nacional de México/Instituto Tecnológico de Toluca, Av. Tecnológico S/N. Colonia Agrícola Bellavista C.P. 52149, Metepec, Estado de México, México
[b] Centro Universitario UAEM Atlacomulco, Universidad Autónoma del Estado de México, Autopista Toluca-Atlacomulco km 60, C.P. 50400, Atlacomulco, Estado de México, México

A R T I C L E I N F O

Article history:
Received 5 March 2024
Revised 22 May 2024
Accepted 4 June 2024
Available online 9 June 2024

Dataset link: FTIR-Plastics: a Fourier Transform Infrared Spectroscopy dataset for the six most prevalent industrial plastic polymers (Original data)

Keywords:
FTIR spectrophotometer
Synthetic polymers
PET, HDPE, PVC, LDPE, PP, PS

A B S T R A C T

This work introduces two datasets: FTIR-Plastics-C4 (Fourier Transform Infrared Spectroscopy, in plastics, at a wavenumber spectral resolution of 4 cm^{-1}) and FTIR-Plastics-C8 (Fourier Transform Infrared Spectroscopy, in plastics, at a wavenumber spectral resolution of 8 cm^{-1}), each comprising 3,000 spectra corresponding to the most used synthetic polymers worldwide. The main contribution of this work lies in the selection and FTIR characterization of the six polymers commonly used in everyday life and industry, namely Polyethylene Terephthalate (PET), High-Density Polyethylene (HDPE), Polyvinyl Chloride (PVC), Low-Density Polyethylene (LDPE), Polypropylene (PP), and Polystyrene (PS). FTIR-Plastics-C4 consists of 3,000 spectra obtained with a configuration of 32 scans and a resolution of 4 cm^{-1}, covering a range from 4000 to 400 cm^{-1}. The FTIR-Plastics-C8 dataset also contains 3,000 spectra obtained with 32 scans and a resolution of 8 cm^{-1} within the same range. A cleaning stage was applied to the FTIR-Plastics datasets, removing the header containing 19 lines and a footer with 34 lines from the original file. Additionally, a standardization process as-

* Corresponding author.
E-mail address: ralejoe@toluca.tecnm.mx (R. Alejo-Eleuterio).

https://doi.org/10.1016/j.dib.2024.110612

signs 15 lines in the files to highlight information regarding the equipment used (based on the information provided by a Jasco spectrophotometer, model FT/IR-6700 PRO 4x, used for polymer characterization). The final dataset is in tabular .csv file format. The dataset is available on an open repository, and its application was designed to identify microplastics extracted from the environment and enable comparisons between commercial polymers.

Specifications Table

Subject	Material Science, Chemistry
Specific subject area	Polymers and Plastics, Material Characterization, Analytical Chemistry: Spectroscopy
Type of data	Raw, Filtered
Data collection	The dataset comprises the six most common synthetic polymers: PET, HDPE, PVC, LDPE, PP, and PS. Samples from these materials were prepared for analysis using Jasco FTIR-6700 PRO 4x spectrophotometer, with parameters adjusted for optimal performance. Spectra captured information on the wave number and percentage of transmittance, and they were organized into a comprehensive database for the selected polymers, with resulting data stored in .csv files comprising header, value, and footer sections. Data mining techniques were applied to clean the files, removing unnecessary sections, and ensuring uniformity. Finally, the database was structured into FTIR-Plastics-C4 and FTIR-Plastics-C8 datasets. These configurations represented variations in parameters such as scan number and resolution. Each configuration contained 3,000 spectra, totaling 6,000 spectra across both configurations. The database columns were organized to include identifiers for each polymer, details of the analysis technique, sample numbers, scan numbers, resolutions, and spectral FTIR data points. The datasets were saved in .csv format, ensuring standardization and easy access.
Data source location	The authors obtained the dataset from a specifically designed methodology carried out in the Advanced Instrumentation Lab, Environmental Engineering Research Lab, and Artificial Intelligence and Open-source Software Lab, belonging to The National Technological Institute of Mexico (Tecnológico Nacional de México), campus Toluca, located on Av. Tecnológico S/N, Colonia Agrícola Bellavista, C.P. 52149, Metepec, Estado de México, México.
Data accessibility	Repository name: FTIR-Plastics-C4, FTIR-Plastics-C8 Data identification number: https://zenodo.org/doi/10.5281/zenodo.10736649 Direct URL to data: https://zenodo.org/records/11176740

1. Value of the Data

- The primary objective of analyzing the most used polymers is to distinguish their characteristic bands in the infrared spectra to construct a specific, valuable, and understandable database for the chemical identification of plastic materials. This approach directly benefits the scientific community involved in using and managing materials made from synthetic polymers, specifically plastics: PET, HDPE, PVC, LDPE, PP, and PS.
- Through FTIR (Fourier Transform Infrared Spectroscopy) analysis, detailed information about the spectra corresponding to each plastic is sought, with their characteristic bands contributing to identify these polymers in samples containing mixtures of them, such as microplastics. This significantly enhances knowledge, understanding, and classification of these materials in several scientific and technological fields.

- The most widely used commercial plastics, namely PET, HDPE, PVC, LDPE, PP, and PS, often contain the reference polymer predominantly, with some mixtures of other polymers. A database created with FTIR, a common technique for polymer identification due to its simplicity and effectiveness, can help identify one or several polymers within their mixtures.
- FTIR spectra typically exhibit vibrational peaks with low intensity. However, comparing these small peaks with an extensive database can facilitate their identification and subsequent classification.
- The creation of this database offers significant advantages to other researchers and professionals in the field. The compiled and structured information can be leveraged by the scientific community for a variety of purposes, from validating and comparing results to driving new research and discoveries in polymer characterization.
- The dataset structure, as *.csv formatted files, facilitates access to plastic-related information, saving time and resources for those involved in similar projects. Moreover, it is helpful for the scientific community at large, including other studies and disciplines related to Big Data, Data Mining, and Artificial Intelligence. The data are helpful for experimentation and testing of many classification and grouping algorithms, such as K-means, DBSCAN, or even other architectures based on neural networks.
- The information in the datasets can be used in the analysis, identification, classification, and/or characterization of microplastics which can be found in bodies of water, in crop fields, in sediments, or in the food industry, among other areas where microplastics are defined as contaminants or unwanted substances.
- The datasets presented here use the transmittance percentage as the "Y" variable. This is significant because transmittance directly measures how much light passes through the sample, which can offer more intuitive information about the vibrational characteristics of the molecular bonds in polymers than absorbance. Absorbance data often requires additional interpretation to understand the intensity of molecular vibrations.
- In the "X" axis, our datasets utilize a spectral range of 4000 cm^{-1} to 400 cm^{-1}. This wider wavenumber range allows for a more comprehensive analysis of the polymer structures by capturing a broader spectrum of vibrational frequencies. This is advantageous over other datasets that typically use a narrower range of 3600 cm^{-1} to 1250 cm^{-1} [1] or 1000 nm to 1700 nm [2], as it enables the identification of a more comprehensive array of molecular features and potential contaminants.

2. Background

This dataset arises from the need for a comprehensive collection of infrared spectra for chemical identification of common synthetic polymers used mainly for packaging in many industries due to their flexible and rigid properties [3,4]. The polymers selected for database creation were chosen based on their high frequency of use, attributed to their excellent mechanical properties. They have unique characteristics that make them relevant in various applications; therefore, they simultaneously reveal a particular challenge in identifying and classifying polymers because they are often mixed with other materials or contain more than one variety of polymers.

Below is a list of the six polymers selected for the database development; the choice is based on their wide usage due to mechanical properties directly related to the adhesion created within polymers [5].

- Polyethylene Terephthalate (PET).
- High-Density Polyethylene (HDPE).
- Polyvinyl Chloride (PVC).
- Low-Density Polyethylene (LDPE).
- Polypropylene (PP).
- Polystyrene (PS).

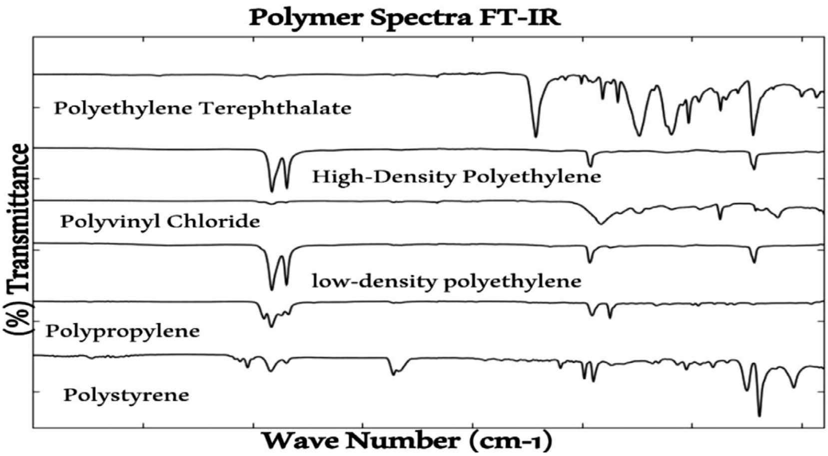


Fig. 1. Example of the FTIR spectra of the six most used industrial plastics: PET, HDPE, PVC, LDPE, PP, and PS/figure of its own creation.

Infrared spectroscopy is most employed in the field of chemistry for its stability in material analysis, as the excitation of vibrational energy provides information about molecular structures, interactions, and connections (in terms of individual functional groups), properties recognized in the characteristic bands represented in FTIR spectroscopy analysis [6,7].

3. Data Description

Fig. 1 presents an example of FTIR spectra of the six plastics chosen to create the database. As observed, each spectrum shows different vibrational peaks, allowing for the distinction and identification of each one.

The variables manipulated during the acquisition of FTIR spectra are the equipment resolution, the number of scans, and the reading range for the wavenumber. These variables must be tested iteratively until the combination shows the most defined peaks, proceeding with additional sample readings and analysis. In this work, special effort has been made to carefully select parameters to obtain representative and repetitive FTIR spectra to perform the chemical characterization of the most used synthetic polymers in the industry. This careful selection not only expands the applicability of the database in practical contexts but also provides valuable information about the specific properties of each polymer, thus contributing to a deeper understanding of these materials in the scientific and technological fields. The data in this article serves as a crucial reference for those exploring spectral characteristic analysis in synthetic polymers with wavenumber and transmittance derived from FTIR spectroscopy.

The FTIR analysis was conducted in five sections of each polymer, randomly chosen in the polymeric piece. These sections were labeled as: section 1, section 2, section 3, section 4, and section 5. The details of the sections used for the analysis of the six polymers are presented in Table 1, consisting of 8 columns described as follows:

- Number assigned by the Society of the Plastics Industry to each type of plastic.
- Acronyms of the synthetic polymer.
- From the third to the seventh column, the considered areas for the analysis are section 1, section 2, section 3, section 4, and section 5.
- Total number of files obtained.

As an essential remark, we used data mining and augmentation to clean and correct the original data acquired with the FITR spectrophotometer. These are widely adopted technologies

Table 1
Number of analyzed FTIR data from each sample section to generate the total number of files per dataset.

SPI No.	Plastic	Section1	Section2	Section3	Section4	Section5	No. files
1	PET	100	100	100	100	100	500
2	PEHD	100	100	100	100	100	500
3	PVC	100	100	100	100	100	500
4	PELD	100	100	100	100	100	500
5	PP	100	100	100	100	100	500
6	PS	100	100	100	100	100	500
				Total of files per dataset			3,000

Table 2
Name and description of the metadata corresponding to the original files.

Metadata	Description
TITLE SAMPLE NAME:	The name of the analyzed polymer.
DATA TYPE:	Specifying the characterization technique.
MEASUREMENT INFORMATION:	Equipment used for data collection.
MODEL NAME:	Name of the used equipment.
SERIAL No:	Serial number assigned to the equipment.
ACCESSORY:	Complementary device integrated into the equipment.
LIGHT SOURCE:	Standardized light source related to the DLATGS detector.
RESOLUTION:	Parameter used to distinguish the wavenumber in the analyzed materials.
XUNIT/HORIZONTAL AXIS:	Referring to the unit's title assigned on the x-axis.
YUNITS/VERTICAL AXIS:	Referring to the unit's title designated on the y-axis.
FIRSTX:	Initial value set for the x-axis.
FIRSTY:	Initial value set for the y-axis.
LASTX:	Final value set for the x-axis.
LASTY:	Final value set for the y-axis.
NPOINTS:	Total data points in the file.

to enhance models used in machine learning, increasing randomness and data sample size for the initial development of the dataset [8,9].

The FTIR-Plastics-C4 and FTIR-Plastics-C8 datasets result from adjusted configurations in the parameters of the spectrophotometer used for spectral characterization, such as the number of scans and resolution. Each dataset has specific values, as the assigned resolution determines the interaction of the infrared light beam with the material. FTIR-Plastics-C4 consists of 3,751 values, while FTIR-Plastics-C8 has 1,884 values. Both configurations include a standardized header.

Files generated by the Jasco FTIR PRO 4x spectrophotometer are retrieved in a .csv (comma-separated values) format. The files are structured into two columns and three distinct parts: header, values, and footer, described as follows:

- Header: Comprises 18 rows and two columns. It provides primary information such as the title of the analyzed sample, peak maximum, peak minimum, date, and time, among others.
- Values: Represent the collected data with two columns: column one corresponds to the "x" axis (wavenumber), and column two corresponds to the "y" axis (transmittance percentage or intensity). These data, originally in decimal format with four decimal places, start from 399.9162 cm^{-1} to 4000.1206 cm^{-1} for the "x" axis, and the intensity on the "y" are in a range from 0 to 120.
- Footer: Comprising 27 rows containing specific information about the used equipment, analysis location, date, time, resolution, scan speed, and other details already included in the header at the beginning of the file.
- Also, the files contain metadata (see Table 2) with relevant information about the equipment and testing conditions.

Table 3
The header section of the standardized file in the FTIR-Plastics-C4 and FTIR-Plastics-C8 datasets.

Metadata	Description
TITLE SAMPLE NAME	Polyethylene High Density (commercial packaging)
DATA TYPE	Infrared Spectrum
MEASUREMENT INFORMATION	Spectrophotometer
MODEL NAME	FT/IR ATR-4x Jasco Type A
SERIAL No.	A009662111
ACCESORY	ATR PRO4x
LIGHT SOURCE	DLATGS Standard
RESOLUTION	4 cm^{-1} (or 8cm^{-1} depending on selected configuration)
XUNITS/HORIZONTAL AXIS	Wavenumber (cm^{-1})
YUNITS/VERTICAL AXIS	Transmittance (%)
FIRSTX	400
FIRSTY	0
LASTX	4000
LASTY	120

Table 4
The numerical value section in the FTIR-Plastics-C4 and FTIR-Plastics-C8 datasets' standardized file represents the FITR spectrum.

Metadata	Description
NPOINTS	1884 -3751 (depending on selected configuration)
399.1927	95.4227
400.1569	95.632
401.1211	95.7494
402.0854	95.8077
403.0496	95.8137
……..	……..
The rest of the numerical data till the end of file	The rest of the numerical data till the end of file

A data mining technique is applied for cleaning each of the original files retrieved from the spectrophotometer, aiming to standardize the header of each file, adding a total of 15 rows. In the second part of the file, only numerical values are included (corresponding to the wavenumber and transmittance of the FTIR signal).

By integrating 500 standardized files of each sample, the FTIR-Plastics-C4 and FTIR-Plastics-C8 databases are generated, each composed of 3,000 samples for each configuration. Each database collection is exclusively carried out using the values from each file. Subsequently, a transpose operation is applied to both columns to combine the data into a single row. This row is read in pairs, where the first data corresponds to the "x" axis (wavenumber), while the second data represents the infrared signal transmittance on the "y" axis (expressed as a percentage).

The 6,000 files comprising the databases are structured into two collections named FTIR-Plastics-C4 and FTIR-Plastics-C8; these have a standardized header for each of the polymers. The labels of the file's header section are detailed in Table 3, while the numerical values section structure is found in Table 4.

The FTIR-plastic-c4 and FTIR-plastic-c8 datasets are in the Zenodo repository [10], both in a Zip format, containing the following structure:

- FTIR-PLASTIC-c4:
  - HDPE_c4.zip: Contains the 500 files corresponding to HDPE
  - LDPE_c4.zip: Contains the 500 files corresponding to LDPE
  - PET_c4.zip: Contains the 500 files corresponding to the PET
  - PP_c4.zip: Contains the 500 files corresponding to the PP
  - PS_c4.zip: Contains the 500 files corresponding to the PS
  - PVC_c4.zip: Contains the 500 files corresponding to the PVC
  - FTIR_PLASTIC_c4.csv: The file corresponds to the general base

Table 5
Final integration of the FTIR-Plastics-C4 dataset.

IDE	Polymer	Technique	Sample	BR	RST	Data (x)	Data(y)	Data (x)	Data(y)
PETFTIR001324	PET	FTIR	001	32	**4**	399.1927	87.9136	400.1569	88.0659
HDPEFTIR001324	HDPE	FTIR	001	32	**4**	399.1927	94.7263	401.1211	94.7092
PVCFTIR001324	PVC	FTIR	001	32	**4**	399.1927	92.4165	402.0854	94.7665
LDPEFTIR001324	LDPE	FTIR	001	32	**4**	399.1927	94.7937	403.0496	98.1111
PPFTIR001324	PP	FTIR	001	32	**4**	399.1927	102.691	404.0138	103
PSFTIR001324	PS	FTIR	001	32	**4**	399.1927	96.8123	404.0496	97.6158

Table 6
Final integration of the FTIR-Plastics-C8 dataset.

IDE	Polymer	Technique	Sample	BR	RST	Data (x)	Data(y)	Data (x)	Data (y)
PETFTIR500328	PET	FTIR	001	32	**8**	399.1927	71.8845	412.6919	73.6194
HDPEFTIR500328	HDPE	FTIR	001	32	**8**	399.1927	92.633	410.7635	94.725
PVCFTIR500328	PVC	FTIR	001	32	**8**	399.1927	99.014	408.8350	96.2656
LDPEFTIR500328	LDPE	FTIR	001	32	**8**	399.1927	99.3487	403.0496	100.564
PPFTIR500328	PP	FTIR	001	32	**8**	399.1927	91.0639	408.8350	94.6687
PSFTIR500328	PS	FTIR	001	32	**8**	399.1927	87.5917	406.9065	88.789

- FTIR-PLASTIC-c8:
  - HDPE_c8.zip: Contains the 500 files corresponding to HDPE
  - LDPE_c8.zip: Contains the 500 files corresponding to LDPE
  - PET_c8.zip: Contains the 500 files corresponding to the PET
  - PP_c8.zip: Contains the 500 files corresponding to the PP
  - PS_c8.zip: Contains the 500 files corresponding to the PS
  - PVC_c8.zip: Contains the 500 files corresponding to the PVC
  - FTIR_PLASTIC_c8.csv: The file corresponds to the general base

The FTIR-Plastics database consists of two configurations: FTIR-Plastics-C4 (Table 5) and FTIR-Plastics-C8 (Table 6), each with a total of 3,000 FTIR spectra of the six polymer materials analyzed (500 for each), thus obtaining a total of 6,000 FTIR spectra by integrating the two configurations into a single file.

4. Experimental Design, Materials and Methods

This section addresses the experimental approach planning and is systematically divided into three distinct phases. These phases provide an organizational structure to the research and allow a deeper and more comprehensive exploration of the applied methods. As each stage is broken down, meticulous planning and execution supporting the experimental design are presented, highlighting scientific rigor and coherence throughout the utilized process steps applied (see Fig. 2).

4.1. Phase 1: Data Acquisition

Collecting samples of PET, HDPE, PVC, LDPE, PP, and PS plastics was performed considering the recycling standards established by the Society of the Plastics Industry (SPI). According to SPI, each type of plastic is identified by an assigned number, showing a clear connection between the material and its classification within the recycling system (see Fig. 3).

The followed steps for sample preparation, according to SPI recommendations, and subsequent data acquisition are described below:

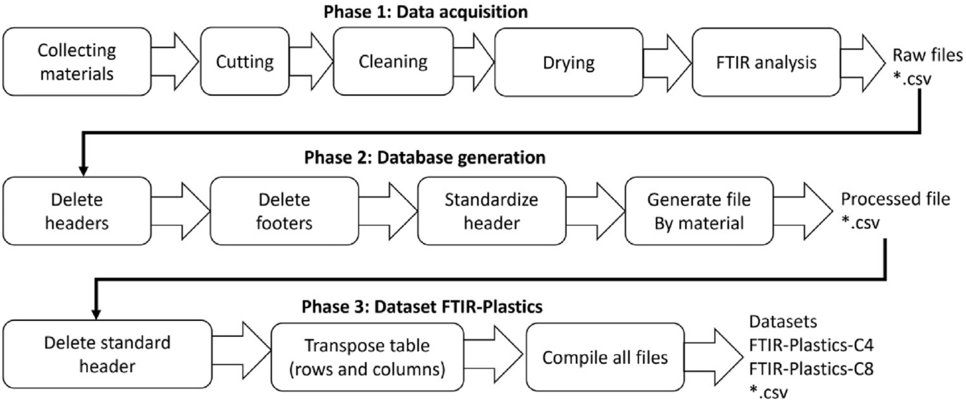


Fig. 2. Summary of the proposed experimental design to perform the polymer characterization, the cleaning and pre-processing of the database, and the final dataset generation.

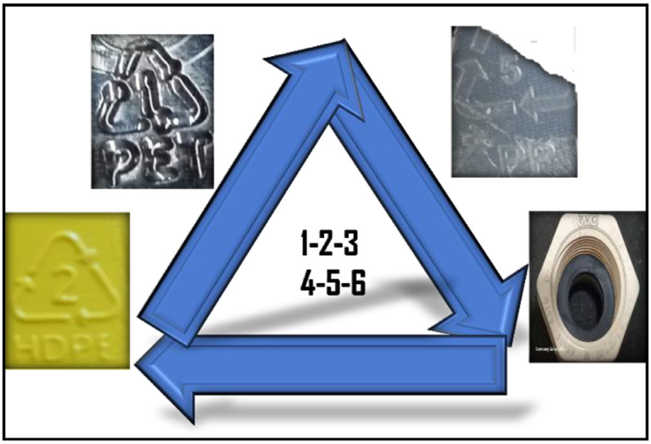


Fig. 3. Triangle of recycling symbol used for plastics: 1 for PET, 2 for HDPE, 3 for PVC or vinyl, 4 for LDPE, 5 for PP, 6 for styrene or PS, and 7 for other plastics not specified/ figure of its own creation.

Step 1. Cutting: The material was carefully cut into pieces no larger than 5 mm to emulate the distinctive characteristics of microplastics. Measurement was carried out using a Vernier caliper to maintain the same dimensional precision in each cut sample. The objective was to obtain semi-square or rectangular pieces as long as the material allowed (see Fig. 4).

Step 2. Cleaning: The pieces obtained in the previous step underwent a washing process to eliminate surface contaminants that could lead to erroneous measurements. A neutral chemical detergent, Hyclin Plus with pH = 7 (Hycel brand), was used, followed by an initial rinse with tap water and then Milli-Q grade distilled water. It's worth noting that Milli-Q grade refers to a type of ultra-pure water produced by a Milli-Q water purification system commonly used in laboratories. This water is highly purified, free from impurities, and meets the standards required for sensitive analytical techniques such as FTIR. Milli-Q grade water ensures no contaminants interfere with the sample analysis, providing more accurate and reliable results. Additionally, It is important to note that the samples were handled with gloves and other plastic materials were avoided during manipulation to prevent cross-contamination.

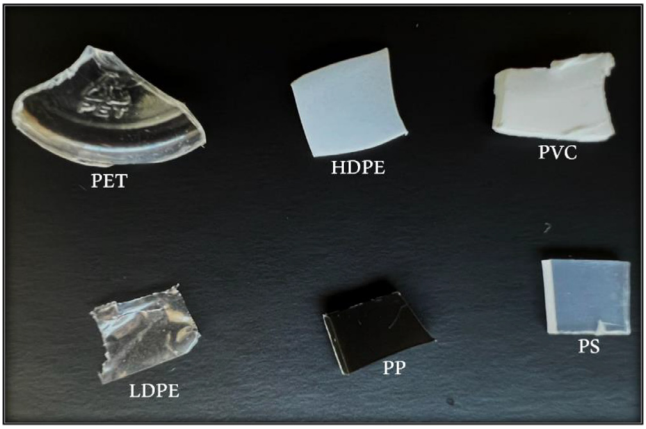


Fig. 4. Samples of each plastic material to be analyzed by FTIR were cut into small pieces.

Step 3. Drying: The pieces obtained from the six plastics were subsequently subjected to a drying process. The samples were kept for 4 hours at room temperature without air currents to avoid contamination between plastics and to reduce moisture. This also minimized exposure to environmental factors such as light, carbon dioxide, humidity, and temperature, which could affect the FTIR analysis.

Step 4. Analysis: The final step involves analysis using the Jasco FTIR PRO 4x spectrophotometer. Prior to this, the user must ensure the proper energy conditions using an uninterruptible regulated power supply. Additionally, environmental conditions must be maintained at a maximum of 5% relative humidity and standard room temperature, not exceeding 38°C.

To ensure measurement accuracy and minimize potential errors in polymer sampling, we utilized the auto-review module of the Spectra Manager Ver. 2.5 software provided with the spectrophotometer at the start of each sampling session. The results of this process are illustrated in Fig. 5. The upper graph displays the transmittance percentage as a function of the wavenumber, forming an "m"-shaped curve, indicating that the equipment is correctly calibrated. Similarly, the lower graph shows a near-straight line without significant variations, confirming the stability of the equipment's energy signal and ensuring precise measurements. Additionally, we have included the calibration certificate provided by the equipment maker (calibration_certificate.pdf) in the Zenodo repository, certifying that the spectrophotometer is calibrated according to ISO 9001 quality standards. It is important to note that, as per standard practice, the equipment is under periodic calibration to maintain accuracy and reliability.

Before each batch analysis, after 3 or 4 h of the equipment use, or when modifying some of the reading conditions, a background revision was performed to register and subtract environmental gases from each spectrum in the laboratory (background is the reference signal of the instrument). The parameters adjusted in the background and the subsequent sample reading were the number of scans, resolution, and reading range for the analysis (see Table 3, where details about the FTIR spectrophotometer configuration are included). The spectra of the samples were obtained in transmittance (%) as a function of the wavelength of the infrared light (expressed as wavenumber in cm^{-1}). Once these parameters were established, the dry sample could be placed on the spectrophotometer's interaction or reading window.

The analysis of the obtained results from the plastic sample are automatically generated as files in *.jws (JSON web signature) format. However, the equipment software allows selecting from several format extensions to store a copy of the file, such as *.txt (plain text files), *.csv (comma-separated values), and *.json (javascript object notation). In this case, the *.csv format was selected to continue with data processing.

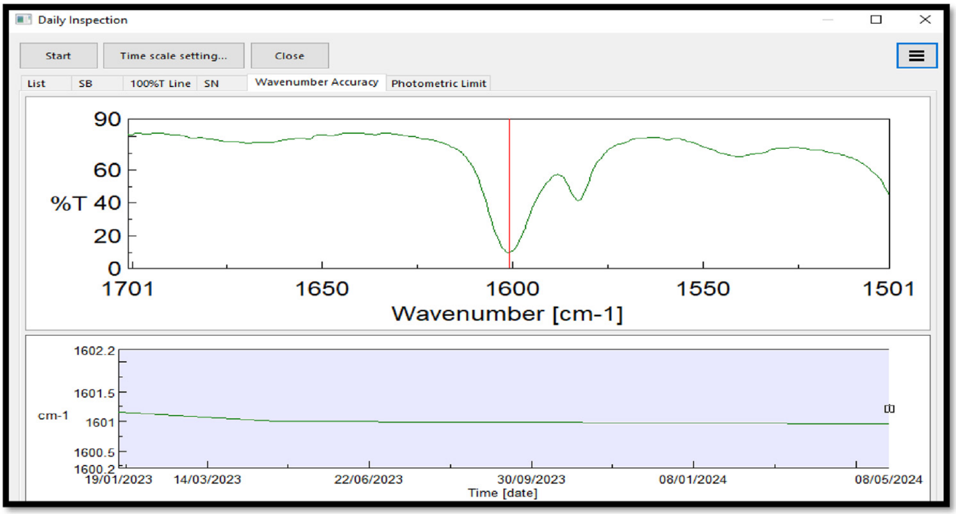
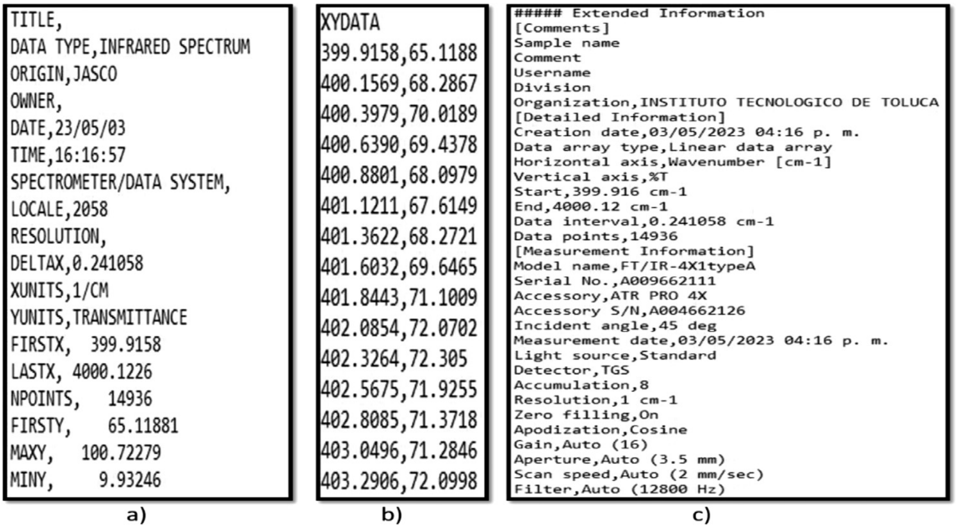


Fig. 5. Example screen of the auto-verification routine to ensure proper operation of the spectrophotometer.



TITLE,
DATA TYPE,INFRARED SPECTRUM
ORIGIN,JASCO
OWNER,
DATE,23/05/03
TIME,16:16:57
SPECTROMETER/DATA SYSTEM,
LOCALE,2058
RESOLUTION,
DELTAX,0.241058
XUNITS,1/CM
YUNITS,TRANSMITTANCE
FIRSTX, 399.9158
LASTX, 4000.1226
NPOINTS, 14936
FIRSTY, 65.11881
MAXY, 100.72279
MINY, 9.93246

a)

XYDATA
399.9158,65.1188
400.1569,68.2867
400.3979,70.0189
400.6390,69.4378
400.8801,68.0979
401.1211,67.6149
401.3622,68.2721
401.6032,69.6465
401.8443,71.1009
402.0854,72.0702
402.3264,72.305
402.5675,71.9255
402.8085,71.3718
403.0496,71.2846
403.2906,72.0998

b)

Extended Information
[Comments]
Sample name
Comment
Username
Division
Organization,INSTITUTO TECNOLOGICO DE TOLUCA
[Detailed Information]
Creation date,03/05/2023 04:16 p. m.
Data array type,Linear data array
Horizontal axis,Wavenumber [cm-1]
Vertical axis,%T
Start,399.916 cm-1
End,4000.12 cm-1
Data interval,0.241058 cm-1
Data points,14936
[Measurement Information]
Model name,FT/IR-4X1typeA
Serial No.,A009662111
Accessory,ATR PRO 4X
Accessory S/N,A004662126
Incident angle,45 deg
Measurement date,03/05/2023 04:16 p. m.
Light source,Standard
Detector,TGS
Accumulation,8
Resolution,1 cm-1
Zero filling,On
Apodization,Cosine
Gain,Auto (16)
Aperture,Auto (3.5 mm)
Scan speed,Auto (2 mm/sec)
Filter,Auto (12800 Hz)

c)

Fig. 6. File extracted from the Spectra Analysis Software, as *.csv file: a) original header, b) FTIR spectrum numerical values, and c) footer.

4.2. Phase 2. Database Generation

The next step involves creating the database of the six selected synthetic polymers. From the characterization carried out in the previous phase, 500 files were obtained for each material.

The original files are extracted in *.csv format and integrated into three sections: header, value content, and footer, generating a total of 3795 rows and two columns (see Fig. 6):

- Header: contains specific data of the employed equipment, the date and time of creation, and the maximum and minimum values of the variables in the “x” axis (wavenumber) and the “y” axis (transmittance percentage).

- Values: corresponds to the obtained when applying the FTIR analysis in a range of 4000 to 400 cm^{-1}.
- Footer: contains similar data to the file header and the resolution and number of scans used in the analysis.

In this phase, a cleaning process was applied to each file to remove categorical data such as the analysis date, sample name, data type, spectrophotometer model, serial number, and accessory name. This was done to focus solely on the relevant spectral data. No normalization techniques were applied, meaning the data are presented precisely as initially obtained from the equipment for each synthetic polymer. The standardization process refers solely to the homogenization of the structure and information contained in the header and footer of each file in the dataset.

4.3. Phase 3. FTIR-Plastics Database

This phase is crucial for generating the structure that makes up the database, composed of two columns: the first represents the values of the "x" axis (wavenumber), while the second contains the values of the "y" axis associated with the intensities of the molecular vibrations, indicating the percentage of transmittance, and differentiating each material.

A transposition is carried out to form the FTIR-Plastics-C4 and FTIR-Plastics-C8 databases, inverting the columns by rows representing each file. At this point, experimentation is fundamental to adjust parameters such as the number of scans, resolution, and range, discarding those that show too much electronic noise. The transmittance percentage is established in a range from 0 to 120%. The decision not to modify this range is because the analysis is qualitative, meaning that vibration peaks correspond to molecular vibrations produced by the polymers and do not quantify the amount of that molecule in each polymer.

The FTIR-Plastics databases are composed of individual files in *.csv format, corresponding to the six materials analyzed and standardized during the cleaning process to ensure coherence and easy identification of variables associated with each type of polymer. This organization allows an efficient data manipulation to facilitate accessibility.

The creation of the database comprising 6,000 FTIR spectra from commonly used synthetic polymers presents a valuable resource for various scientific and technological applications. This extensive dataset facilitates the identification and characterization of polymers, enabling more profound insights into their properties and behaviors. Using advanced analytical techniques and meticulous data collection processes, this database offers researchers and practitioners a comprehensive tool for material characterization, quality control, and product development across diverse industries. Moreover, the standardized format and accessibility of the dataset enhance its usability and effectiveness in advancing research endeavors and addressing real-world challenges related to polymer science and engineering. As an indispensable asset in materials science, this database serves as a cornerstone for innovation, discovery, and progress in synthetic polymers. Finally, FTIR-Plastics databases represent a valuable tool for data science applied in material characterization, as this methodology can be applied to other techniques used for the same purpose.

Limitations

The authors and the technical staff have tried their best to characterize the polymers using Fourier Transform Infrared Spectroscopy (FTIR). However, it is important to acknowledge certain limitations and potential biases in this dataset:

- Calibration of Measurement Equipment: The accuracy of the data within the dataset is highly dependent on the calibration of the measurement equipment. Any deviations in the calibration process can influence the precision of the results. Although we have taken measures to ensure proper calibration, slight inaccuracies may still affect the data.

- Spectrophotometer Parameters: The precision of the obtained data is directly linked to the specific configurations employed in the spectrophotometer's parameters. Any variations in these settings can lead to significant discrepancies in the data. Consequently, the resulting spectrum might not accurately represent the analyzed polymer or could exhibit notable differences if a reader attempts to replicate the obtained spectrum.
- Sample Size Limitation: The FTIR-Plastics-C4 and FTIR-Plastics-C8 datasets were explicitly designed with a focus on microplastics, meaning that all analyzed samples have dimensions no larger than 5 mm. While this size restriction is crucial to ensure that the results align with the characteristic behavior of microplastics, it also limits the applicability of the data to larger-sized samples.

Ethics Statement

The authors have read and followed the ethical requirements for publication in Data in Brief and confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

FTIR-Plastics: a Fourier Transform Infrared Spectroscopy dataset for the six most prevalent industrial plastic polymers (Original data) (Zenodo).

CRediT Author Statement

Octavio Villegas-Camacho: Conceptualization, Investigation, Data curation, Writing – original draft; **Roberto Alejo-Eleuterio:** Conceptualization, Methodology, Writing – review & editing; **Iván Francisco-Valencia:** Methodology, Writing – review & editing; **Everardo Granda-Gutiérrez:** Supervision, Resources, Writing – review & editing; **Sonia Martínez-Gallegos:** Supervision, Resources, Writing – review & editing; **Javier Illescas:** Writing – review & editing.

Acknowledgements

Octavio Villegas Camacho thanks CONAHCYT for his scholarship number 934178 and to the Instituto Tecnológico de Toluca for proving the facilities to carry out this project.

To the technical and support personnel from Advanced Instrumentation Lab, Environmental Engineering Research Lab, and Artificial Intelligence and Open-source Software Lab at The National Technological Institute of Mexico (Tecnológico Nacional de México), campus Toluca.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Competing Interests

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

References

[1] S. Primpke, M. Wirth, C. Lorenz, G. Gerdts, Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared (FTIR) spectroscopy, Anal. Bioanal. Chem. 410 (2018) 5131–5141.
[2] G. Bonifazi, F. Di Maio, F. Potenza, S. Serranti, FT-IR spectroscopy and hyperspectral imaging applied to post-consumer plastic packaging characterization and sorting, in: SENSORS, 2014 IEEE, IEEE, 2014, pp. 633–636.
[3] Y-J. Weng, Forming of Dynamic microstructure of flexible polymer, Materials 12 (20) (2019) 3332, doi:10.3390/ma12203332.
[4] M. Dendisová, A. Jeništová, A. Parchaňská-Kokaislová, P. Matějka, V. Prokopec, M. y Švecová, The use of infrared spectroscopic techniques to characterize nanomaterials and nanostructures: A review, Anal. Chimica Acta 1031 (2018) 1–14, doi:10.1016/j.aca.2018.05.046.
[5] C. Fan, Y.-Z. Huang, J.-N. Lin, J. y Li, Microplastic constituent identification from admixtures by Fourier-transform infrared (FTIR) spectroscopy: The use of polyethylene terephthalate (PET), polyethylene (PE), polypropylene (PP), polyvinyl chloride (PVC) and nylon (NY) as the model constituents, Environ. Technol. Innov. 23 (2021) 101798, doi:10.1016/j.eti.2021.101798.
[6] X. Yan, Z. Cao, A. Murphy, Y. Ye, X. Wang, Y. y Qiao, FRDA: Fingerprint Region based data augmentation using explainable ai for FTIR-based microplastics classification, Sci. Total Environ. 896 (2023) 165340, doi:10.1016/j.scitotenv.2023.165340.
[7] U. Riaz, S.M. y Ashraf, Characterization of polymer blends with FTIR spectroscopy, John Wiley & Sons, Ltd, 20 (2014) 625–678, doi:10.1002/9783527645602.ch20.
[8] Smirnova Margarita, Achim Kohler, & Shapaval Volha. (2020). FTIR Dataset. doi:10.5281/zenodo.4297950.
[9] S. Primpke, M. Wirth, C. Lorenz, et al., Reference database design for the automated analysis of microplastic samples based on Fourier transform infrared (FTIR) spectroscopy, Anal. Bioanal. Chem. 410 (2018) 5131–5141, doi:10.1007/s00216-018-1156-x.
[10] O. Villegas Camacho, R. Alejo Eleuterio, I. Francisco Valencia, E. Granda Gutiérrez, S. Mártinez Gallegosy, J. Illescas, FTIR-Plastics: a Fourier Transform Infrared Spectroscopy dataset for the six most prevalent industrial plastic polymers, Zenodo (2024), doi:10.5281/zenodo.10736650.