

README

Dataset: Gas Permeability and Polymer Properties for Pure Polymer Membranes

Associated publication: From FTIR Spectra to Gas Permeability: Machine Learning for Polymer Membrane Screening

Authors: Mehryar Jaafri, Khashayar Salchi, Bernardo Castro-Dominguez

Institution: University of Bath

Licence: Creative Commons Attribution 4.0 International (CC BY 4.0)

Overview

This dataset supports a CatBoost gradient-boosted machine learning model trained to predict gas permeability of dense polymer membranes from ATR-FTIR spectroscopy combined with polymer-family metadata. It contains gas permeability values (Barrer) for six gases compiled from the peer-reviewed literature for 204 pure polymers, alongside polymer names, SMILES strings, thermal properties, morphological category, and chemical family classification. ATR-FTIR spectra are not included in this deposit due to source licence restrictions. Spectra can be accessed directly from SpectraBase (spectrabase.com). A spectrum-to-polymer mapping file (`graphs.csv`) is provided to link SpectraBase spectrum IDs to polymer entries in `Dataset.csv`.

Files

1. Dataset.csv

204 rows (one per polymer), 15 columns.

Item -- Unique integer polymer ID. Used as the join key with `graphs.csv`.

Polymer -- Polymer name.

SMILES -- SMILES string representation of the repeat unit. One entry is missing.

Atoms -- Elemental composition of the repeat unit (e.g. C, H, O).

Category -- Morphological classification: Amorphous or Semi-crystalline.

Tm (K) -- Melting temperature in Kelvin. Missing for amorphous polymers (153 entries).

Tg (K) -- Glass transition temperature in Kelvin.

Chemical Family -- Polymer chemical family classification. 11 classes: Acrylic, Fluoropolymer, Other, Phenolic, Polyacetal, Polyamide, Polyester, Polyolefin, Polystyrene, Polysulfone, Polyvinyl.

He (Barrer) -- Helium permeability (Barrer).

H2 -- Hydrogen permeability (Barrer).

CO2 -- Carbon dioxide permeability (Barrer).

N2 -- Nitrogen permeability (Barrer).

O2 -- Oxygen permeability (Barrer).

CH4 -- Methane permeability (Barrer). One entry is missing.

Polyinfo name -- Polymer name as listed in the PolyInfo database, where applicable. 200 entries are blank.

2. graphs.csv

This file allows users who obtain spectra from SpectraBase to map each spectrum to its corresponding polymer entry in `Dataset.csv`. **Column 1 (new_key)** corresponds to the spectrum identifier as used in SpectraBase. **Column 2 (Item)** matches the Item column in `Dataset.csv`."

How to use this dataset

To reproduce the machine learning pipeline described in the associated publication:

1. Load `Dataset.csv` to obtain permeability targets, thermal properties, and polymer-family metadata keyed by Item.

2. To use FTIR spectra, obtain them from SpectraBase using the spectrum IDs in graphs.csv (Column 1) and match them to polymers via the Item column. 3. Load individual FTIR CSV files from FTIR_Spectra/ using the new_key values.

Notes

All permeability values are sourced from peer-reviewed literature and reported in Barrer ($1 \text{ Barrer} = 10^{-10} \text{ cm}^3(\text{STP}) \text{ cm} / (\text{cm}^2 \text{ s cmHg})$).

FTIR spectra are raw ATR-FTIR absorbance measurements. In the published pipeline, spectra were interpolated to a common wavenumber grid and compressed using PCA before use as model inputs.

Polymers are pure homopolymers only; blends and composites are not included.

Not all polymers have measurements for all six gases. Missing values appear as blank cells.

Citation

If you use this dataset, please cite the associated publication and this data deposit. Full citation details will be added upon acceptance.

Contact

University of Bath Research Data Archive. For enquiries contact the corresponding author via the associated publication.