Electronic Supplementary Information

Predictive Analysis of Organic Semiconductor Photodegradation by FTIR Spectroscopy with Multivariate Analysis

Sarah M. Tyler,[a] Thomas J. Blackburn,[a] Kellen T. Schneider,[a]

and Jeanne E. Pemberton[a]*

[a]Department of Chemistry and Biochemistry
University of Arizona
1306 East University Boulevard
Tucson, AZ 85721
United States

*Correspondence: pembertn@arizona.edu

Table of Contents

I. PTB7 ATR-FTIR Data – 1000-3400 cm^{-1}, Averaged……S3

II. PTB7 Vibrational Assignments……S3

III. A Brief Primer on 2D-COS……S5

IV. Statistical Analysis of Chemometric Results……S11

V. Individual Test Dataset Accuracy Results……S13

VI. References……S14

I. PTB7 ATR-FTIR Data – 1000-3400 cm^{-1}, Averaged

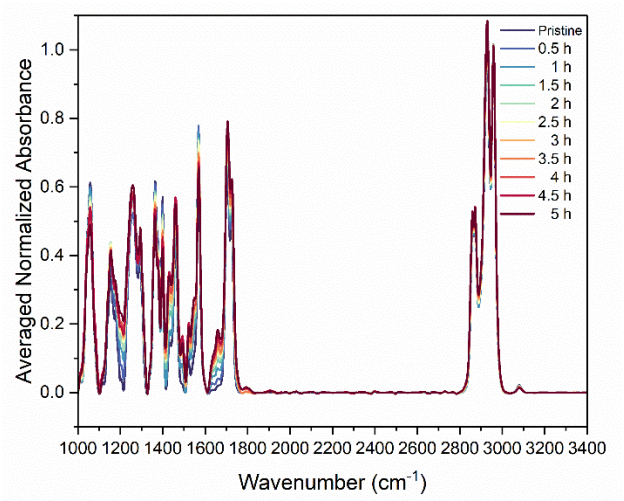

Figure S1. Averaged and normalized ATR-FTIR spectra of PTB7 spin cast on a ZnSe IRE (512 scans, 4 cm^{-1} resolution, sample size n = 4 collected in triplicate) under 30 min interval exposures to photo-illumination from a solar simulator over the spectral range of 1000-3400 cm^{-1}.

II. PTB7 Vibrational Assignments

Parametric Model #7 Hamiltonian (PM7) available in the MOPAC interface in Chem3D was chosen as the semi-empirical quantum method for prediction of vibrational assignments based on its history of comparable predictions to density functional theory (DFT) using B3LYP methods for a wide range of organic species.[1] Separately, each subunit (BDT or FTT) was minimized either through PM7 or RM1 minimization until imaginary frequencies were no longer predicted. The predicted frequencies were then corrected with appropriate scaling factors.[1]

Table S1: Characteristic Frequencies and Assignments for Pristine PTB7.[a]

Predicted Frequency (cm^{-1})	Observed Frequency (cm^{-1})	Assignment[b]	Subunit
1008 1067	1057	ν_s(C-O-C)$_{ether}$ + δ(C-C-C)$_{Th}$	FTT
1144 1166	1153 1176	ν(C-C)$_{alkyl}$ + ν_s(C-O-C)$_{ester}$	BDT + FTT
1220 1279	1263	ν_{as}(C-O-C)$_{ether}$ + ν_{as}(C-O-C)$_{ester}$ + ν(C-C-C)$_{TP}$	FTT + BDT
1303 1318	1294	ν_{as}(C-C-O)$_{ester}$ + δ(CH$_2$)$_{alkyl}$	BDT + FTT
1348 1357	1365	ν(C=C)$_{TP}$	BDT
1380 1386	1400 1429	ν(C=C)$_{\phi\ ring}$ + ν(C=C)$_{TP}$	BDT
1452	1461	δ(CH$_2$)$_{alkyl}$	FTT + BDT
1505 1589	1569	ν_s(C-C-F)$_T$ + ν(C=C)$_{\phi\ ring}$ + ν(C=C)$_{TP}$ + ν(C=C)$_T$	BDT + FTT
1630 1708 1734	1705	ν(C=O)$_{ester}$	FTT
1778	1728	ν(C=O)$_{ester}$ + ν_{as}(C-C-F)$_T$	FTT
1806			
2856	2860	ν(CH$_2$)$_{alkyl}$	BDT + FTT
2876	2871	ν(CH$_3$)$_{alkyl}$	BDT + FTT
2892	2929	ν(CH$_2$)$_{alkyl}$	BDT + FTT
2960	2960	ν(CH$_3$)$_{alkyl}$	BDT + FTT

[a]Vibrational assignments from references [2], [3], [4], [5], [6], and [7] and supplemented by PM7 semi-empirical quantum mechanical predictions performed separately on the BDT and FTT subunits in the MOPAC interface in Chem3D.

[b]Assignment notations: ν = stretch, δ = bend; ether = BDT alkyl ether side chains, ester = FTT alkyl ester side chains, alkyl = ether or ester-linked alkyl groups, Th = thieno ring of FTT, T = thiophene ring of FTT, TP = thiophene ring of BDT.

III. A Brief Primer on 2D-COS

The flexibility and generality of correlation analysis allow the application of 2D-COS to any set of spectral data for which the system has been exposed to an external perturbation, such as radiant or thermal exposure, wherein the perturbation causes the magnitude (e.g. intensity, frequency) of spectral variables to change. Consider a photodegradation experiment of a polymer system wherein a series of FTIR spectra of the polymer is acquired at specific time intervals of radiant exposure, as shown in Figure S2. In this thought experiment, let's limit the region of interest to two spectral bands, one at 1050 cm^{-1} and the other at 1150 cm^{-1}. Qualitative observation of the spectra reveals that the band at 1050 cm^{-1} begins to decrease in intensity after the first interval of light exposure while the band at 1150 cm^{-1} increases in intensity starting after the second light exposure interval.

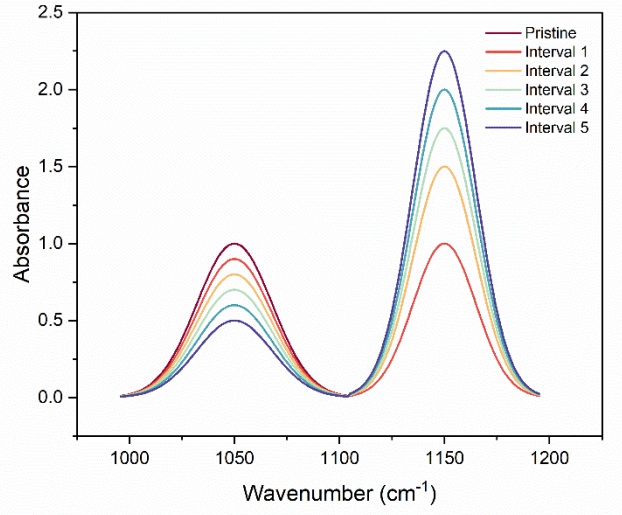

Figure S2. Artificially generated FTIR spectral data of a hypothetical polymer system exposed to cumulative intervals of light. The band centered at 1050 cm starts to decrease in intensity after the first interval of light exposure while the band centered at 1150 cm increases in intensity starting after the second light exposure. For this second band, the maroon trace representing the pristine spectrum is masked by the dark orange trace of representing the first light interval as no spectral change has occurred yet.

A dynamic spectrum ($\tilde{y}(\nu, t)$) of this system can be formally defined as

$$\tilde{y}(\nu, t) = \begin{cases} y(\nu, t) - \bar{y}(\nu), & for\ T_{min} \leq t \leq T_{max} \\ 0, & otherwise \end{cases} \quad \text{(S.1)}$$

where ν is the wavenumber, t is time, $\bar{y}(\nu)$ is the reference spectrum, T_{min} is the beginning of spectral measurement, and T_{max} is the end of spectral measurement. Although the choice of the reference spectrum is not strictly defined and historically has been treated somewhat arbitrarily, selecting a reference spectrum too dissimilar from the spectra in the dataset complicates 2D-COS interpretation and can even lead to erroneous conclusions.[8] The average spectrum of the dataset often

serves as a good estimate of the "static" or constant component of the system with respect to the external perturbation; thus, choosing the average spectrum as the reference allows subtraction of the static component from the data, thereby facilitating 2D-COS interpretation.[8, 9]

Once the dynamic spectrum has been calculated, the cross-correlation analysis can proceed. The generic definition of the 2D correlation intensity ($\mathrm{X}(\nu_1, \nu_2)$) is

$$\mathrm{X}(\nu_1, \nu_2) = \langle \tilde{y}(\nu_1, t) \cdot \tilde{y}(\nu_2, t') \rangle \qquad \text{(S.2)}$$

where the brackets denote the cross-correlation function. $\mathrm{X}(\nu_1, \nu_2)$ is treated as a complex number function with two orthogonal components and is expressed as

$$\mathrm{X}(\nu_1, \nu_2) = \Phi(\nu_1, \nu_2) + i\psi(\nu_1, \nu_2) \qquad \text{(S.3)}$$

where the real ($\Phi(\nu_1, \nu_2)$), or synchronous, component represents the coincidental intensity trends at any two given spectral variables, while the imaginary ($i\psi(\nu_1, \nu_2)$), or asynchronous, component represents the out-of-phase intensity variations at any two given spectral variables.

To generate the synchronous and asynchronous spectra via generalized 2D correlation analysis, the synchronous and asynchronous intensities are formally defined as

$$\Phi(\nu_1, \nu_2) + i\Psi(\nu_1, \nu_2) = \frac{1}{\pi(T_{max} - T_{min})} \int_0^\infty \tilde{Y}_1(\omega) \cdot \tilde{Y}_2^*(\omega)\, d\omega \qquad \text{(S.4)}$$

Simply put, this is a Fourier transformation of the dynamic spectrum carried out with respect to the external variable, e.g. time. For a given pair of spectral variables (ν_1, ν_2), $\tilde{Y}_1(\omega)$ represents the forward Fourier transform of spectral intensity changes occurring at ν_1, while $\tilde{Y}_2^*(\omega)$ is the complex conjugate of the spectral intensity changes occurring at ν_2. These Fourier transformations are given by

$$\tilde{Y}_1(\omega) = \int_{-\infty}^{\infty} \tilde{y}(\nu_1, t) e^{-i\omega t}\, dt = \tilde{Y}_1^{Re}(\omega) + i\tilde{Y}_1^{Im}(\omega) \qquad \text{(S.5)}$$

and

$$\tilde{Y}_2^*(\omega) = \int_{-\infty}^{\infty} \tilde{y}(\nu_2, t) e^{i\omega t}\, dt = \tilde{Y}_2^{Re}(\omega) - i\tilde{Y}_2^{Im}(\omega) \qquad \text{(S.6)}$$

where $\tilde{Y}_1^{Re}(\omega)$ and $\tilde{Y}_2^{Re}(\omega)$ are the real components of each Fourier transform and $\tilde{Y}_1^{Im}(\omega)$ and $\tilde{Y}_2^{Im}(\omega)$ are the imaginary components of each Fourier transform.

Given the large frequency or energy range of typical spectral data and the need to calculate a correlation at every point of a 2D spectral plane, even fast Fourier transforms can take up large amounts of computational resources. To save computational time, different "shortcuts" can be taken to satisfactorily estimate the true 2D correlation intensity. For the example described here for which n spectral traces are collected at evenly spaced time intervals, the synchronous intensity is calculated through a cross-product calculation between the dynamic intensity changes at any two given wavenumbers ν_1 and ν_2.

$$\Phi(\nu_1, \nu_2) = \frac{1}{n-1} \sum_{j=1}^{n} \tilde{y}_j(\nu_1) \cdot \tilde{y}_j(\nu_2) \qquad \text{(S.7)}$$

Calculating the intensity of the asynchronous spectrum is a little more complex, but the Hilbert transformation of the dynamic spectrum can be utilized in lieu of a Fourier transform:

$$\Psi(\nu_1, \nu_2) = \frac{1}{n-1} \sum_{j=1}^{n} \tilde{y}_j(\nu_1) \cdot \sum_{k=1}^{n} N_{jk} \cdot \tilde{y}_k(\nu_2) \qquad \text{(S.8)}$$

where N_{jk} is the element in the discrete Hilbert-Noda transformation matrix corresponding to the jth row and the kth column. This matrix is given by

$$N_{jk} = \begin{Bmatrix} 0, & if\ j = k \\ \frac{1}{\pi(k-i)}, & otherwise \end{Bmatrix} \qquad \text{(S.9)}$$

These calculations facilitate the creation of 2D synchronous and asynchronous plots. Although both plots provide insight into the relationship between coordinate frequency pairs (ν_1, ν_2) upon exposure to the external perturbation, the information gained from each plot is complementary.

The synchronous plot provides information about correlated intensity changes due to the external perturbation at specific coordinate frequency pairs. These plots are symmetric across the diagonal line ($\nu_1 = \nu_2$) and exhibit two types of contour peaks: autopeaks located on the diagonal and cross peaks located in frequency regions off the diagonal. Autopeaks are always positive in sign, and those with high intensity values represent frequencies where significant spectral changes occur in the form of peaks increasing or decreasing in intensity. The cross peaks located off the diagonal represent coordinate pairs of different frequencies where spectral intensity changes occur simultaneously or coincidentally. These peaks are positive when the spectral intensity changes at both frequencies in the same direction (both increasing or decreasing) and are negative when the spectral intensity changes occur in the opposite direction.

The synchronous plot generated from the FTIR spectral data of the hypothetical polymer system of this example is shown in Figure S3. For clarity, the spectral traces are also shown along both X and Y axes. In this example, coordinate pairs with positively correlated intensity changes are red, pairs with negatively correlated intensity changes are blue, and pairs with no correlated intensity changes are yellow. Two autopeaks are present in this plot at (1050, 1050 cm^{-1}) and (1150, 1150 cm^{-1}), which correspond to the two peaks of interest in this hypothetical spectrum. The autopeak at (1150, 1150 cm^{-1}) is larger in magnitude, which implies that upon radiant exposure, the degree of spectral intensity change at 1150 cm^{-1} is greater than the

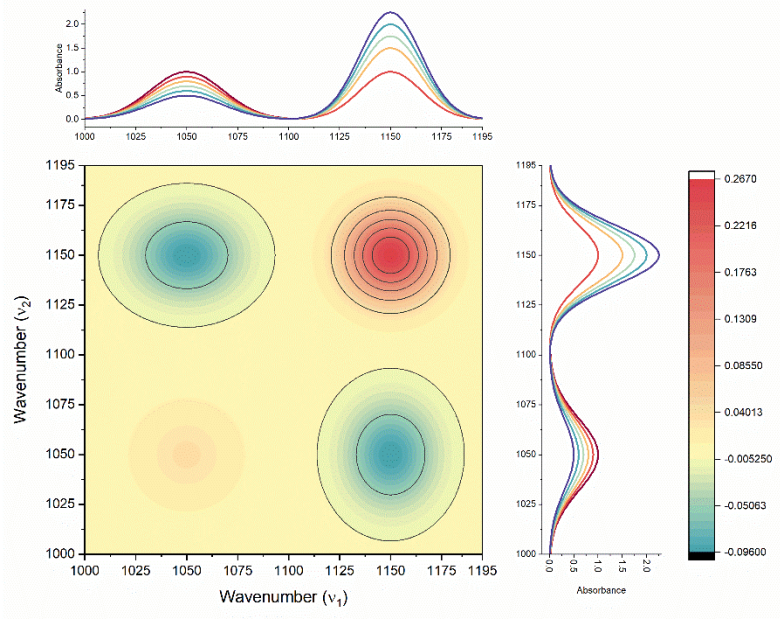

Figure S3. Synchronous plot generated from the simulated FTIR spectra.

spectral intensity change at 1050 cm^{-1}. For this simplified example, the 1150 cm^{-1} band does indeed undergo more dramatic spectral intensity change upon irradiation as shown in Figure S2. As expected, the cross peaks in this system are mirror images of each other at (1150, 1050 cm^{-1}) and (1050, 1150 cm^{-1}). These peaks are negative because the two peaks change intensity in opposite directions. For cumulative radiant exposure intervals, the band at 1050 cm^{-1} decreases in intensity while the band at 1150 cm^{-1} increases. This simple example illustrates how 2D-COS can be utilized for systems with several peaks changing upon exposure to a perturbation and facilitate easier analysis of spectral changes.

The asynchronous plot of generalized 2D-COS provides information about the time-dependent order of spectral changes. These plots only contain off-diagonal cross peak pairs that are mirror images of opposite sign across the diagonal. Asynchronous cross peaks result only from out-of-phase changes between two spectral variables, i.e. changes at a frequency pair that are delayed or accelerated with respect to one another. According to Noda's rule,[10] for an off-diagonal coordinate pair in the synchronous plot, a positive cross peak in the asynchronous plot indicates an intensity change at ν_1 that occurs largely before an intensity change at ν_2. Conversely, a negative cross peak in the asynchronous plot is indicative of a change occurring predominantly at ν_2 before ν_1. These rules are reversed if the corresponding off-diagonal coordinate pair in the synchronous plot has a negative value. While the asynchronous plots provide sensitive analysis for the differentiation of overlapped bands, they are more sensitive to noise and fluctuations in spectral baseline than their synchronous counterparts, which can make interpretation of complex systems difficult.[11]

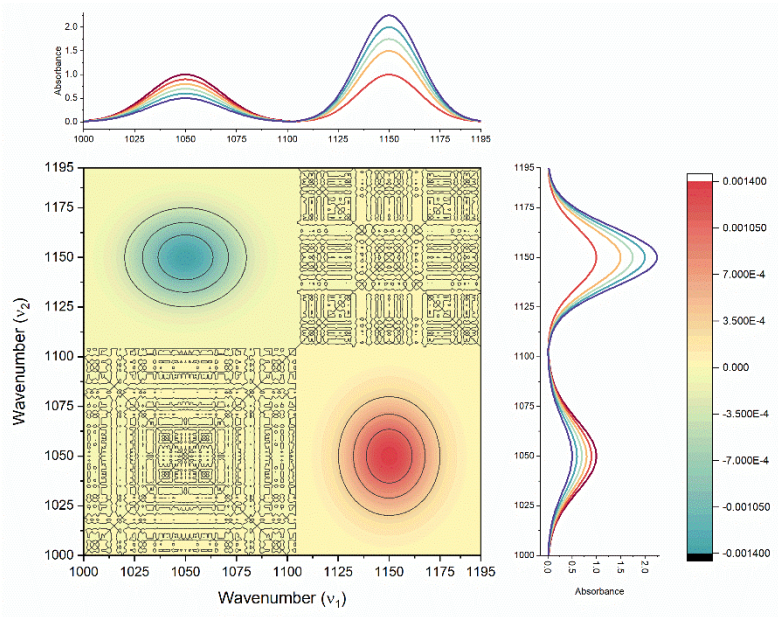

Figure S4. Asynchronous plot generated from the simulated FTIR data.

For this example, the asynchronous plot is shown below in Figure S4. The same color-coding system is utilized as above, with red representing positive values, blue representing negative values, and yellow representing no out-of-phase intensity changes. Predictably, there are no autopeaks in this plot since intensity changes occurring at a given frequency cannot change out of phase with itself. Unlike the synchronous plot, noisy contour lines are observed which demonstrate the increased sensitivity of the asynchronous calculations to random baseline fluctuations. The cross peaks present occur at (1150, 1050 cm^{-1}) and (1050, 1150 cm^{-1}), although they are opposite in sign. Analyzing the positive peak below the diagonal at (1150, 1050 cm^{-1}) and following Noda's rule, the positive peak should indicate spectral intensity change occurring first at 1150 cm^{-1} before 1050 cm^{-1}. However, since in the synchronous plot, the coordinate pair (1150, 1050 cm^{-1}) is negative in sign, the rule reverses, and instead, the correct conclusion is that intensity changes in the 1050 cm^{-1} band occur before the intensity changes in the 1150 cm^{-1} band. Indeed, upon close examination of the simulated dataset shown in Figure S2, the band at 1050 cm^{-1} starts decreasing in intensity after the first radiant exposure interval while the band at 1150 cm^{-1} doesn't start to increase in intensity until after the second exposure interval. This subtle difference would be difficult to parse out in a system where multiple bands change in intensity; this example again demonstrates the utility of 2D-COS.

IV. Statistical Analysis of Chemometric Results

Using the training dataset, 40 principle components (PCs) were used to calculated two linear discriminants (LDs) which were then used to create a model to predict the degradation extent of PTB7 in unlabeled test datasets. Two spectral frequencies were utilized – 1300-1900 cm^{-1} and 1000-3400 cm^{-1}.

1300-1900 cm^{-1}

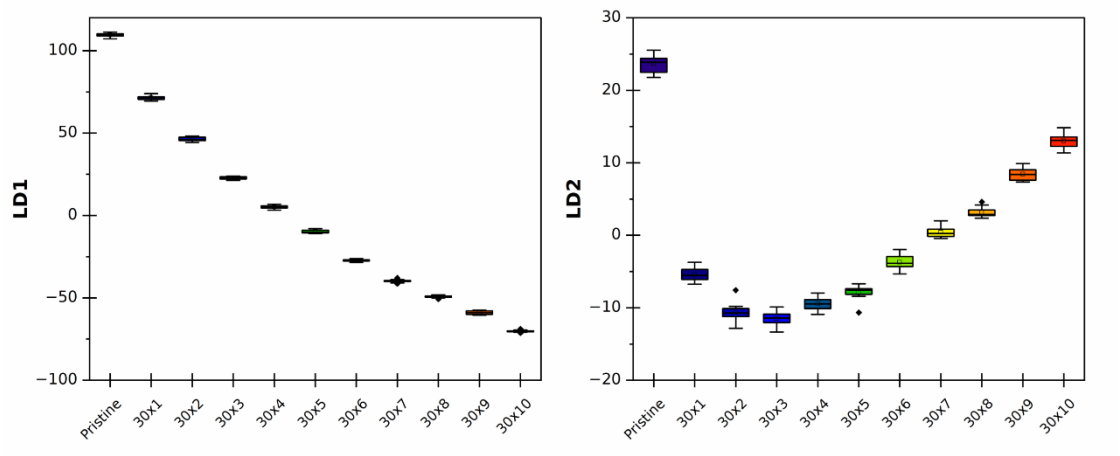

Figure S5. Box plot analysis on LD1 and LD2 scores after PCA treatment from ATR-FTIR spectral data between 1300-1900 cm^{-1} for PTB7 films under sequential 30 min radiant exposures.

Table S2: ANOVA of LD1 and LD2 scores calculated after PCA treatment from the FTIR spectral data between 1300-1900 cm^{-1} for PTB7 films under sequential 30 min radiant exposures.

	LD1					LD2				
	DF	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**
Model	10	396867.05357	39686.70536	39686.70536	<0.0001	10	14995.4941	1499.54941	1499.54941	<0.0001
Error	121	121	1			121	121	1		
Total	131	396988.05357				131	15116.4941			

Null Hypothesis: The means of all levels are equal.
Alternative Hypothesis: The means of one or more levels are different.
At the 0.05 level, the population means are significantly different.

Table S3: Post hoc Tukey's test results for mean LD1 and LD2 values after PCA treatment between 1300-1900 cm^{-1} for ATR-FTIR spectral data between 1300-1900 cm^{-1} of PTB7 films under sequential 30 min radiant exposures.

LD1													LD2										
Sample	**Mean**	**Groups**											**Sample**	**Mean**	**Groups**									
Pristine	109.6167	A											**Pristine**	23.66228	A									
0.5 h	71.3999		B										**5 h**	12.98877		B								
1 h	46.31365			C									**4.5 h**	8.40632			C							
1.5 h	22.89186				D								**4 h**	3.15531				D						
2 h	5.13377					E							**3.5 h**	0.39786					E					
2.5 h	-9.79012						F						**3 h**	-3.72026						F				
3 h	-27.30272							G					**0.5 h**	-5.42689							G			
3.5 h	-39.73141								H				**2.5 h**	-7.83553								H		
4 h	-49.35226									I			**2 h**	-9.46868									I	
4.5 h	-58.93424										J		**1 h**	-10.64665									I	J
5 h	-70.24514											K	**1.5 h**	-11.51253										J

Means that do not share a letter are significantly different.

1000-3400 cm^{-1}

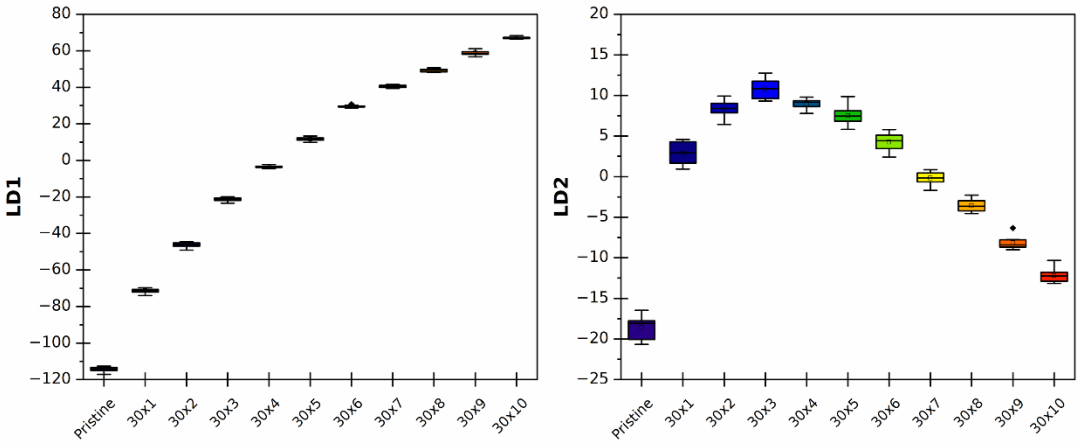

Figure S6. Box plot analysis on LD1 and LD2 scores after PCA treatment from ATR-FTIR spectral data between 1000-3400 cm^{-1} for PTB7 films under sequential 30 min radiant exposures.

Table S4: ANOVA of LD1 and LD2 scores calculated after PCA treatment from the FTIR spectral data between 1000-3400 cm^{-1} for PTB7 films under sequential 30 min radiant exposures.

	LD1					LD2				
	DF	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**
Model	10	405474.23547	40547.42355	40547.42355	<0.0001	10	11083.37454	1108.33745	1108.33745	<0.0001
Error	121	121	1			121	121	1		
Total	131	405595.23547				131	11204.37454			

Null Hypothesis: The means of all levels are equal.
Alternative Hypothesis: The means of one or more levels are different.
At the 0.05 level, the population means are significantly different.

Table S5: Post hoc Tukey's test results for mean LD1 and LD2 valuesafter PCA treatment between 1000-3400 cm^{-1} for ATR-FTIR spectral data between 1300-1900 cm^{-1} of PTB7 films under sequential 30 min radiant exposures.

LD1												LD2										
Sample	**Mean**	**Groups**										**Sample**	**Mean**	**Groups**								
5 h	67.11138	A										**1.5**	10.79532	A								
4.5 h	58.7491		B									**2 h**	8.9696		B							
4 h	49.3025			C								**1 h**	8.39373		B	C						
3.5 h	40.49622				D							**2.5 h**	7.49526			C						
3 h	29.45885					E						**3 h**	4.25203				D					
2.5 h	11.58379						F					**0.5 h**	2.83081					E				
2 h	-3.46536							G				**3.5 h**	-0.18296						F			
1.5 h	-21.29135								H			**4 h**	-3.60119							G		
1 h	-46.28844									I		**4.5 h**	-8.1692								H	
0.5 h	-71.3891										J	**5 h**	-12.202									I
Pristine	-114.26759										K	**Pristine**	-18.5814										J

Means that do not share a letter are significantly different.

V. Individual Test Dataset Accuracy Results

Table S6: Accuracy of PCA-LDA predictions grouping by total radiant exposure time for TD1 collected and processed by Researcher 1.

1300-1900 cm^{-1}				1000-3400 cm^{-1}			
±0 min	**±30 min**	**±60 min**	**±90 min**	**±0 min**	**±30 min**	**±60 min**	**±90 min**
51.5%	86.4%	98.5%	100.0%	68.2%	90.9%	100.0%	100.0%

Table S7: Accuracy of PCA-LDA predictions grouping by total radiant exposure time for TD2 processed by Researcher 2.

1300-1900 cm^{-1}				1000-3400 cm^{-1}			
±0 min	**±30 min**	**±60 min**	**±90 min**	**±0 min**	**±30 min**	**±60 min**	**±90 min**
83.3%	100.0%	100.0%	100.0%	56.1%	78.8%	83.3%	90.9%

Table S8: Accuracy of TD1 data for binary distinction between pristine and degraded PTB7 fims for 1300-1900 cm^{-1} and 1000-3400 cm^{-1}.

1300-1900 cm^{-1}	**1000-3400 cm^{-1}**
100%	100%

Table S9: Accuracy of TD2 data for binary distinction between pristine and degraded PTB7 films for 1300-1900 cm^{-1} and 1000-3400 cm^{-1}.

1300-1900 cm^{-1}	**1000-3400 cm^{-1}**
100%	78.8%

VI. References

1. X. Rozanska, J. J. P. Stewart, P. Ungerer, B. Leblanc, C. Freeman, P. Saxe and E. Wimmer, *J. Chem. Eng. Data*, 2014, **59**, 3136−3143.
2. J. Razzell-Hollis, J. Wade, W. C. Tsoi, Y. Soon, J. Durrant and J.-S. Kim, *J. Mater. Chem. A*, 2014, **2**, 20189-20195.
3. S. Shah, R. Biswas, T. Koschny and V. Dalal, *Nanoscale*, 2017, **9**, 8665-8673.
4. S. Kim, M. A. M. Rashid, T. Ko, K. Ahn, Y. Shin, S. Nah, M. H. Kim, B. Kim, K. Kwak and M. Cho, *J. Phys. Chem. C*, 2020, **124**, 2762-2770.
5. D. M. Schwaiger, W. Lohstroh, M. Wolf, C. J. Garvey and P. Müller-Buschbaum, *J. Polym. Sci.*, 2023, **61**, 1660–1674.
6. S. Holliday and C. K. Luscombe, *Adv. Electron. Mater.*, 2018, **4**, 1700416.
7. B. J. Tremolet de Villers, K. A. O'Hara, D. P. Ostrowski, P. H. Biddle, S. E. Shaheen, M. L. Chabinyc, D. C. Olson and N. Kopidakis, *Chem. Mater.*, 2016, **28**, 876-884.
8. M. A. Czarnecki, *Appl. Spectrosc.* , 2003, **57**, 991-995.
9. P. D. B. Harrington, A. Urbas and P. J. Tandler, *Chemom. Intell. Lab. Syst.*, 2000, **50**, 149-174.
10. I. Noda, *Appl. Spectrosc.*, 1993, **47**, 1329-1336.
11. M. A. Czarnecki, *Appl. Spectrosc.*, 1998, **52**, 1583-1590.