Supplementary information.

Data pre-processing steps

Data pre-processing is vital in the generation of infrared images. The steps used in this paper were (a) 1st order baseline corrected and (b) vector normalised. The effect of these processes on both the spectra and the resultant peak height images is outlined in Figure S1.

Figure S1(a) shows typical raw spectra taken from the interface highlighted by the grey arrow in Figure 3(a). The images below this line of spectra have been obtained by plotting the peak height of the 1745 cm^{-1} PLGA carbonyl peak (upper image) and the 1635 cm^{-1} peak (water bending mode) for the lower image. The contrast between the polymer particle and the water is quite poor in the water image, due to in part to baseline effects. Applying a 1st order baseline correction to the data (Figure S1(b) and images beneath) improve the contrast significantly especially noticeable in the water distribution image, it is most likely here that the lack of contrast is the result of a variation in the response of each pixel within the image. Vector normalisation (shown in Figure S1(c)) reduced the effect of a variation in the pixel response, resulting in images with a good contrast for both the PLGA and water distribution.

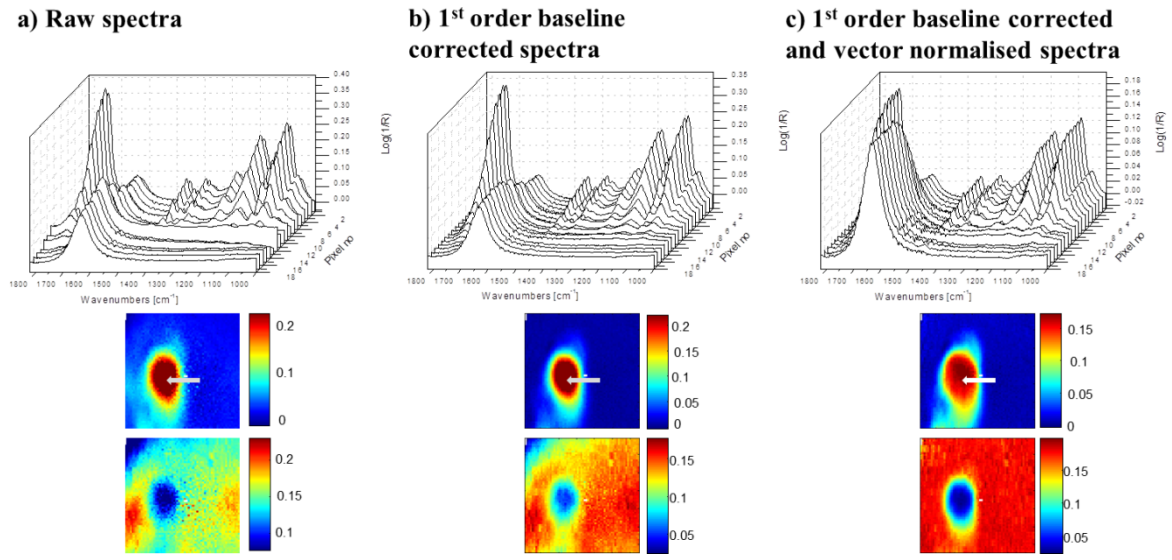

Figure S1: (a) typical raw spectra and the resulting images, (b) the effect of 1st order baseline correction on the same data and (c) the effect of vector normalisation on the 1st order baseline corrected data.

Selection of number of MCR-ALS factors

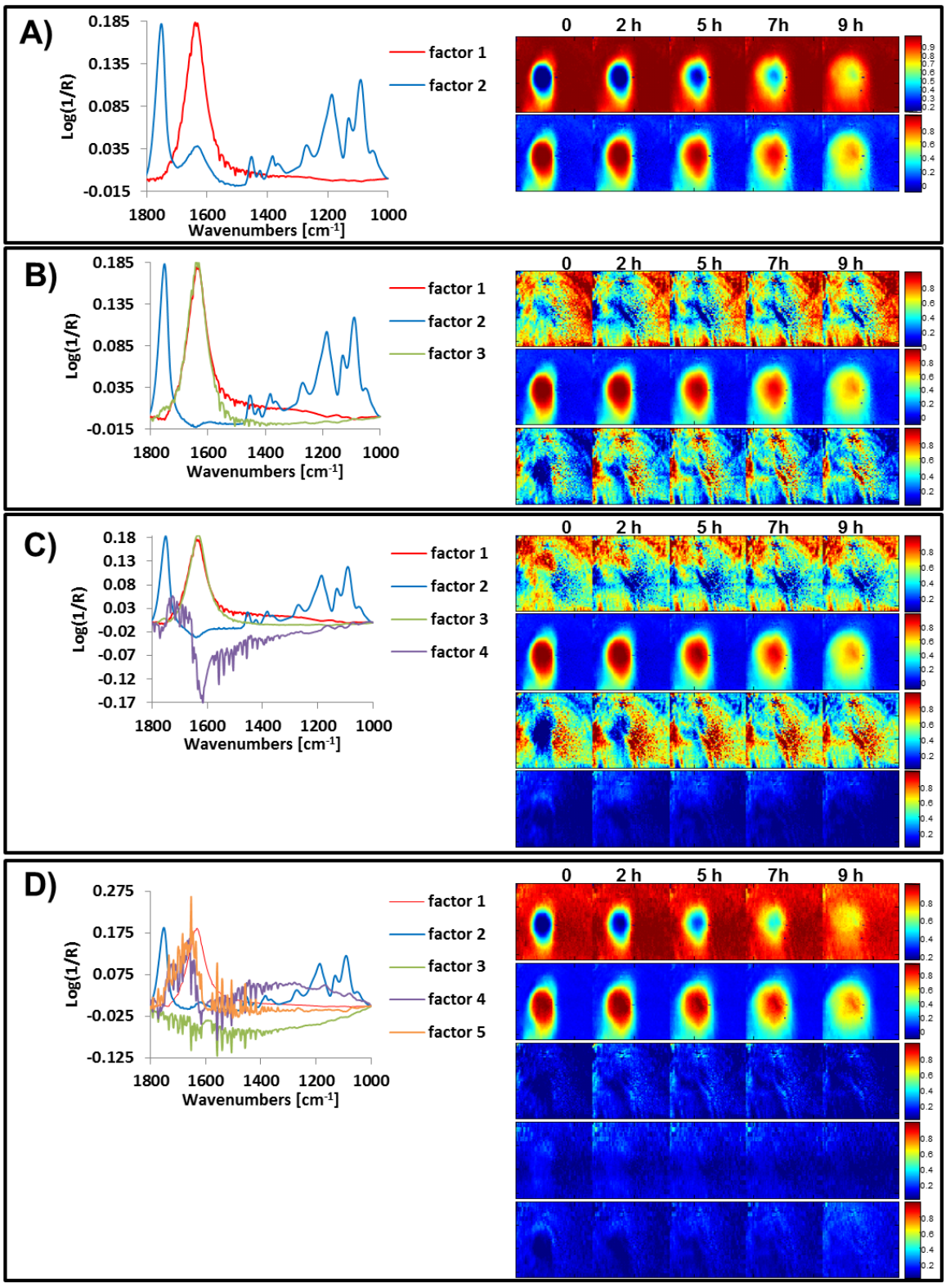

Figure S2: MCR-ALS output analogous to Figure 4 using MCRv3.19 software developed by Andrew and Hancewicz [32] utilising (a) 2 factors, (b) 3 factors, (c) 4 factors and (d) 5 factors.

Figure S2 shows the comparison of a set of images that have been processed using MCR-ALS and, keeping all other parameters the same, changing the number of factors used in the alternating least squares fitting process. Figure S2(a) shows the output where two factors have been selected. Here the factors can clearly be ascribed to belonging to water (factor 1) and PLGA (factor 2), although there is still evidence of the water δ(OH) mode in the factor 2. Increasing the number of factors to 3 (Figure S2(b)) has a significant impact on the quality of the images. Factors 1 and 3 can be ascribed to water and factor 2 to PLGA; interestingly factor 2 no longer has a contribution from water within it. Plotting the loading intensity of factor 2 shows an image very similar to that of factor 2 in Figure S2(a). However the images derived from factors 1 and 3 show no clear anti-correlation with the image derived from factor 2 and have clearly not improved the image resolution over. Increasing the number of factors used to 4 also does not seem to improve the quality of the images produced. Once more the factors can be related to water (factors 1 and 3) and PLGA (factor 3), whilst factor 4 looks to have a component from atmospheric water vapour and a baseline factor. Increasing the number of factors used to 5, leads to the generation of a factor associated with water (factor 1) and PLGA (factor 2) and other factors associated with water vapour and baseline effects. There is no discernible improvement in the S/N of the images produced using 5 factors when compared to those generated using 2 factors and it was still not possible to distinguish between LA and GA regions. Based on these findings we used 2 factors for comparison with the other data analysis approaches.

Outline of the Non-linear Curve Fitting protocol

To obtain a set of non-linear curve fitted peaks for each image the following procedure was performed in MATLAB:

1. For each pixel (i.e. each individual spectrum), the first task was to detect the peaks and for each detected peak store the peak centre ($k0$), maximum intensity $I(k0)$ and calculate and store the full width values at half maximum intensity ($I(k0)/2$). This was achieved by firstly smoothing the pre-processed data by employing a Savitzky-Golay filter [s1] with a polynomial order of 3 and filter length of 13, in order to improve SNR and therefore avoid detecting noise as a peak. The imaging pixel data was sifted and the pixel with the highest intensity between 1820 and 1700 cm^{-1} was identified

and the peak detection algorithm was performed. The output was visually assessed by checking the peak detection output (examples shown in figure S3), and validated to ensure that each peak had been detected. We expected to detect 11 peaks for spectra from wet samples and 10 peaks for spectra from dry samples. If the algorithm failed to detect all the peaks then the parameter 'Delta' (or peak threshold and related to height difference between peak and its surrounding) were adjusted to help the code identify peak maxima and minima. The calculated variables were stored and used later as starting the point when running the gradient search algorithm to fit peaks to the spectrum obtained from each pixel (#5).

2. Peak detection for all pixels was performed by visually assessing whether the expected PLGA peaks ai peak were detected in smoothed spectra of pixels from the core of the partıcıe, tne water rich region and interface regions using same detection criteria. Peak detection was optimised such that all expected peaks in these 3 types of spectra could be detected using the same detection criteria. When all peaks from each of the 18 selected pixels were correctly identified (once more this was checked visually, see figure S3), 'Delta' was deemed to be optimised and fixed. This parameter was then applied to all 4096 pixels to detect all peaks in the image.

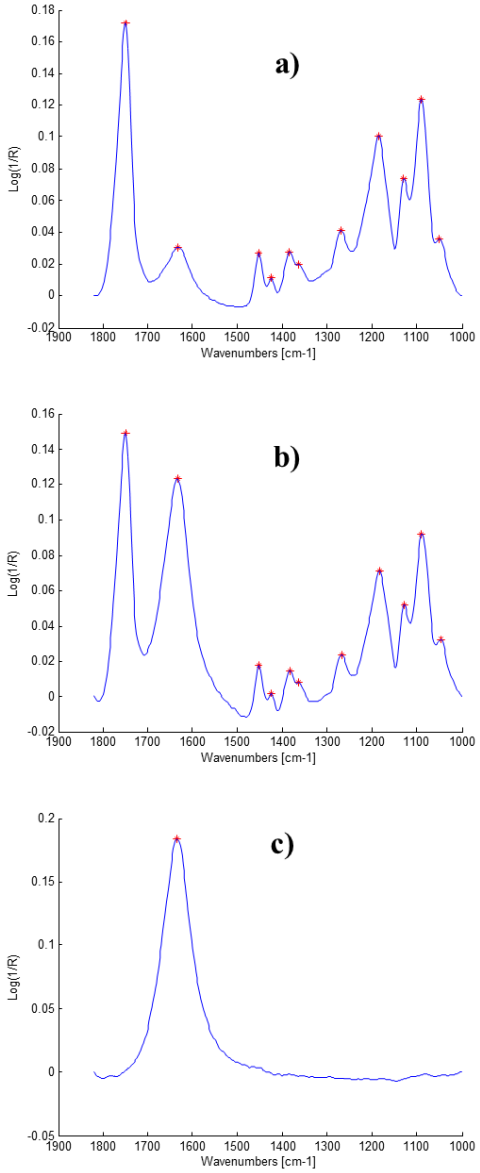

Figure S3: Spectra extracted from the grey line shown on the '0m wet' image in Figure 3(a). Red stars indicate detected peaks using the same X and 'delta' values, (a) typical polymer rich region, (b) typical interface spectrum and (c) typical water only spectrum.

3. Once the number of peaks at each pixel have been detected and stored, then upper and lower limits were defined for the rest of the parameters required for the fitting process namely *M* (the Pearson parameter) which defines the proportion of Gaussian, Lorentzian character for each peak and υ, which is the asymmetry parameter. The mean value of the upper and lower limits of these parameters were used as starting points during the optimisation process.

4. The peak fits for the spectra in each pixel were optimised by using a gradient search algorithm ("trust-region", an improved version of Levenberg-Marquardt algorithm which provides a numerical solution to the problem of minimizing a nonlinear function, over a space of parameters of the function) defined in MATLAB library to perform a non-linear least squares fit to the measured data with $k0$, $I(k0)$, Δ, M and υ for each peak as free parameters to be varied until the sum of squares due to error is reached to its minimum or 500 iterations is reached.
5. The fitting was started in the top right hand corner pixel (pixel #64) and the optimised values for the matching detected peaks outlined in #3 were used as the starting point for the fit (described in #4). The second pixel fitted was pixel #1 i.e. the top left hand corner. If the number of peaks detected by the peak detection algorithm for the current pixel (pixel #1), matched the number of peaks used in the preceding fit, then the optimised preceding fit was used as the starting point for the fitting process. If there is a discrepancy in the number of peaks, then the optimised values for the matching detected peaks outlined in #3 was used as the starting point for the fit. This process was continued with the fitting order being pixel #64, pixel #1, pixel #63, pixel #2 etc. Once a row was completed then the pixels in the row below were fitted following the same protocol; working from the edge of the image to the centre, using the preceding image fit as the reference point.
6. Once the fitting for all 4096 pixels was complete, the peak centres, peak intensities and other parameters for each pixel were collated and the relevant parameters for water (peak centre close to ~1635 cm^{-1}) was removed from each pixel. The output for each pixel was therefore an equation containing up to 10 Pearson IV components (equation 8), the sum of which represents the pure spectrum of PLGA at each pixel.
7. This process (#6) was replicated for the water peak (~1635 cm^{-1}) using calculated parameters in Equation (8) to represent the pure water spectrum.
8. False colour images for the PLGA and water components were generated by plotting the total intensity for each function over 1820-1000 cm^{-1} range at each pixel.

Spectral maps across an interface

Figure S4 shows waterfall plots of the spectra extracted from the area marked with a grey arrow in Figure 3 for (a) a dry PLGA microparticle (image labelled 'dry') and (b) a PLGA

microparticle surrounded by water (image labelled 'wet'). From Figure S4 it is possible to show clear differences in the step-edge measurement between the wet and dry samples, with the gradient of the 'dry' sample being somewhat shallower than that associated with the data from the 'wet' sample.

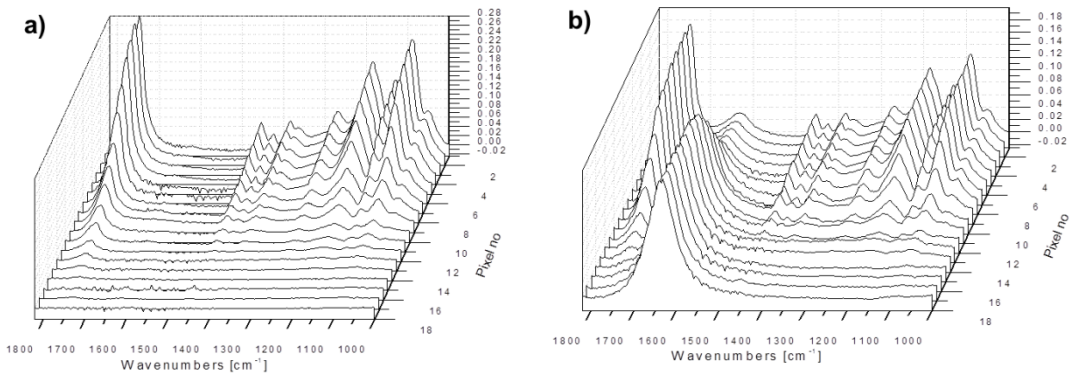

Figure S4. (a). ATR-FTIR spectra along the grey line towards the particle shown in left handside 'dry' image in Figure 3(a), with decreasing pixel number respectively and (b).Pre-processed ATR-FTIR spectra along the gray line shown in Figure 3(a), from the particle towards water rich area with increasing pixel number respectively for the 'wet' image

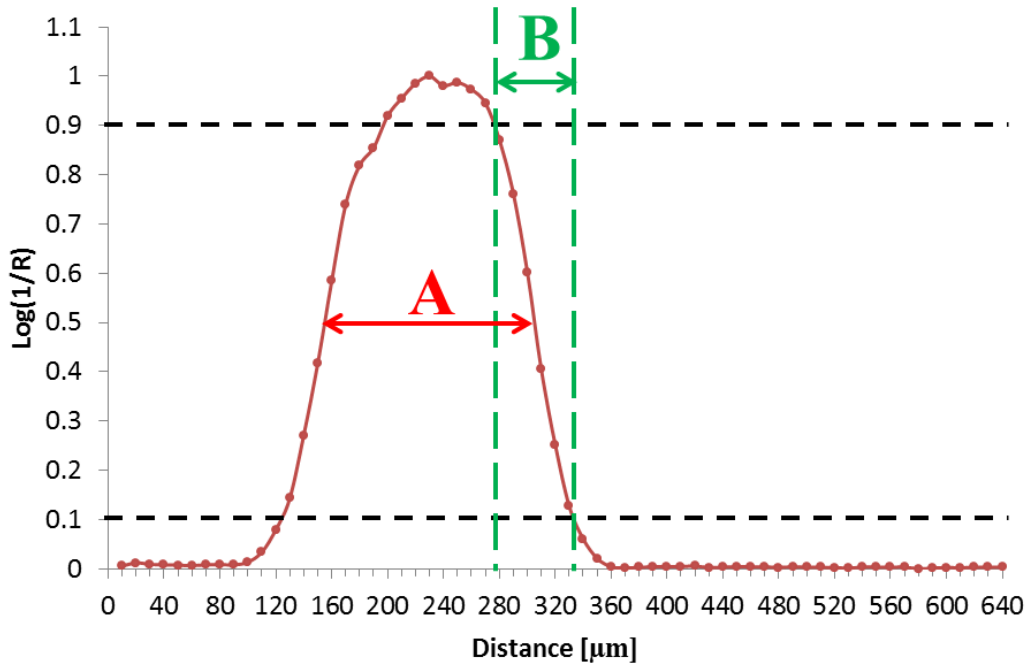

Figure S5. A PLGA line profile for NLCF data showing the derivation of the parameters (A) full width at half maximum height of the normalised PLGA signal and (B) the dimension of the right hand side interface.

To facilitate comparison between data analysis approaches we have defined two parameters that can readily be extracted from each polymer response profile; namely the full width at half maximum height (A) and the region of the right hand side of the profile where the polymer intensity is <90% of its maximum value and >10% of its minimum value (B). Figure S5 shows a typical polymer response profile and the parameters A and B derived from that profile.

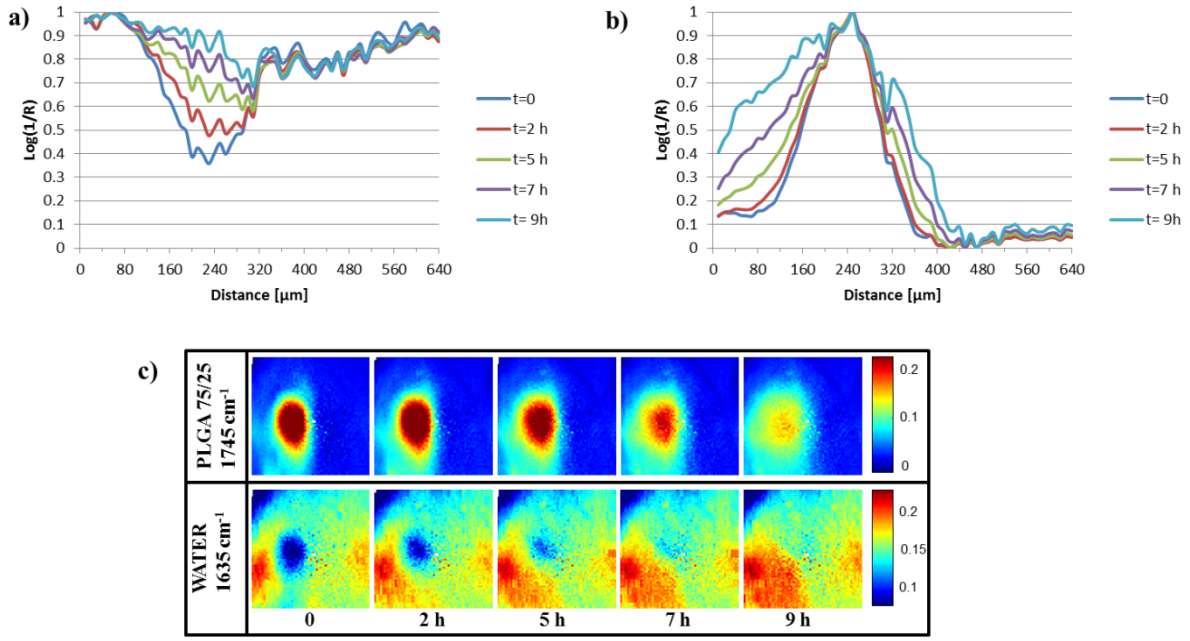

Figure S6. Profiles analogous to figures 5(a) and 5(b) generated for the peak height processed data without vector normalisation and baseline correction. (c) The corresponding peak height images.

Vector normalisation and baseline correction have been shown to be essential in obtaining useful image data using peak heights as the data analysis tool. Figure S6 shows the extracted profiles analogous to figures 5, 6 and 7 (c) and (d) extracted from peak height images S6(c) generated without any data pre-processing. The extracted line profile associated with the water peak is heavily influenced by the fluctuating water vapour during the course of the experiment. The instrument is purged using dried air, but even so we have experienced difficulties in fluctuating water vapour levels over the course of an experiment lasting tens of hours. Data pre-processing significantly improves the quality of the generated images; compare data figure S6(c) with the data shown in figure 4(a).

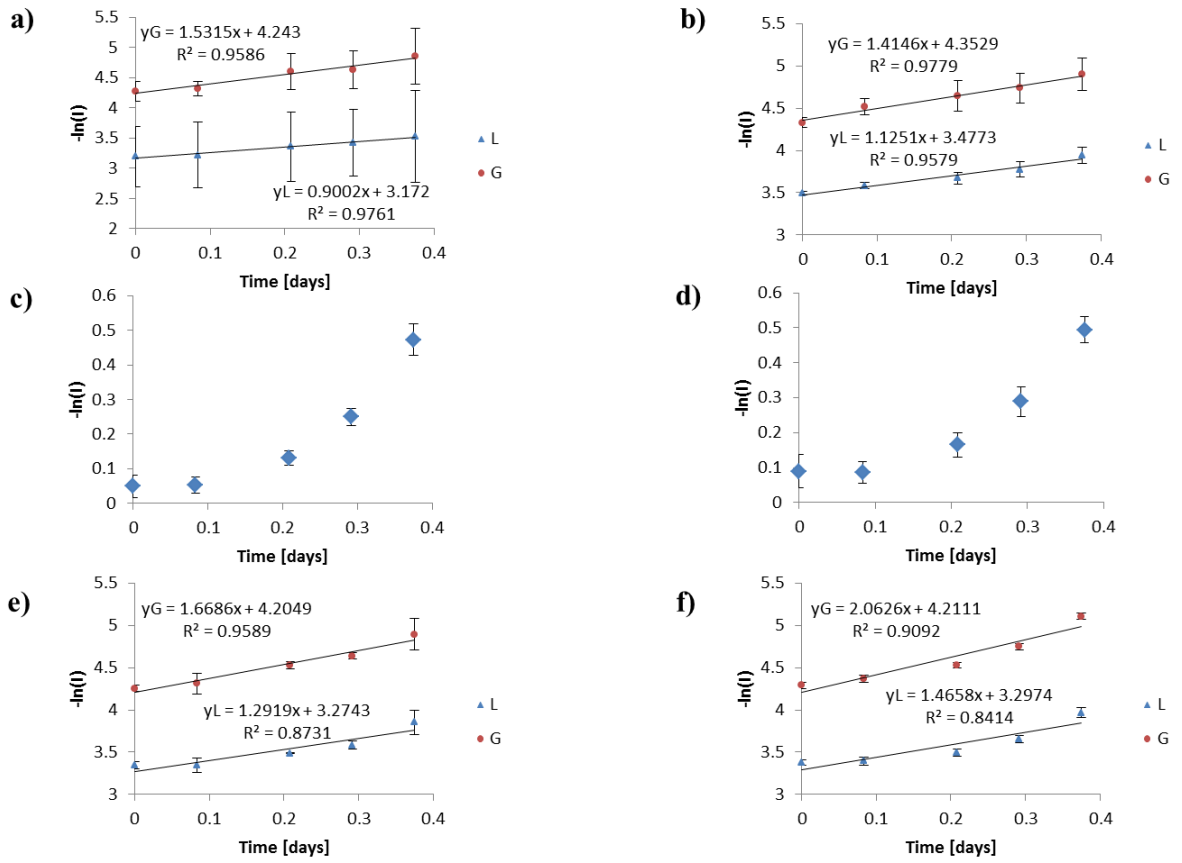

Figure S7. First order kinetic plots used to determine the hydrolysis rates for (a) peak height data using single pixels, (b) peak height data using binned 5x5 areas, (c) MCR-ALS data from single pixels, (d) MCR-ALS data using binned 5x5 areas, (e) NLCF data using single pixels and (f) NLCF data using binned 5x5 areas.

Assuming 1st order kinetics we have plotted the log of the integrated intensity of the band in question (~1452 cm^{-1} for the lactic units and ~1424 cm^{-1} for the glycolic units for peak heights and NLCF) and the entire PLGA factor for the MCR-ALS versus time in days. The slope of the line of best fit using the standard linear regression algorithm in Microsoft Excel 2010, was used to calculate the hydrolysis rate constant. The data shown are the mean and standard deviation of 3 separate measurements from the same 3 random regions within the evolving images. There appears to be a deviation from the expected linear relationship in the plots used to generate k values from the MCR data. This could be due to the inability of the technique to distinguish between the lactic acid and glycolic acid segments necessitating in the entire spectral range being used to determine the k values. As is shown in the peak height and NLCF plots, and the degradation rate studies conducted by Vey *et al.* [47] on mm sized

PLGA discs using FTIR and Raman spectroscopic data fitted to Voigt line shape functions and the study by Tracy *et al.* [48] who calculated degradation rate constants probing the changes in M_w as a function of time using GPC, we are confident that the PLGA hydrolytic degradation is best described by first order kinetics. Here, the MCR factor also contains a water component which is not increasing linearly during the experiment, this issue could be magnified more obviously in the MCR derived data at the t=0 time point, leading to a response which is higher than expected. Consequently we have not attempted to generate a 1^{st} order rate constant for this data. The MCR-ALS analysis could be enhanced by implementing hard constraints [s2] such as pure components spectra in to the algorithm or cropping down to smaller spectral regions, which are both not considered in this study as the aim of employing MCR-ALS was to demonstrate the use of an automated soft modelling tool without hard constraints and using the same spectral range considered in rest PH and NLCF analyses.

[s1] A. Savitzky and M. J. E. Golay, *Anal. Chem*., 1964, **36**, 1627–1639.

[s2] L. Blanchet, C. Ruckebusch, J.P. Huvenne and A. de Juan, *J. Chemometrics,* 2008, **22**, 666–673.