Electronic Supplementary Material (ESI) for Reaction Chemistry & Engineering.

Supporting information for:

A method for real time detection of reaction endpoints using a moving window t-test of *in situ* time course data

Elyse N. T. DiMaso, Robert W. Bondi, John Guo, Alexander G. O'Brien

Table of Contents

S1. Experimental Details

Starting materials **A** and **C** were added to a reactor with reagent **D**. The temperature was then increased to the set temperature and FTIR data collection was started.

FTIR spectra were acquired using either an ic10 or 45P ReactIR instrument. FTIR spectra were acquired using either a ReactIR ic10 or 45P Fourier Transform infrared (FTIR) spectrometer equipped with a 9.5mm diameter DiComp (Diamond) attenuated total reflectance (ATR) probe connected to either a 2m or 4m fiber optic conduit (all from Mettler Toledo AutoChem, USA). The ReactIR ic10 spectrometer had a liquid nitrogen cooled MCT detector and was allowed to stabilize for at least 1 hour before use. The ReactIR 45P spectrometer had a Stirling Engine cooled MCT detector. Both instruments were purged with nitrogen gas or dry compressed air for at least 24 hours prior to use to minimize contributions from water vapor in the spectra. Prior to starting the experiments performance tests were completed to ensure the instrument was working as expected, including a signal to noise test, acetone standard, and comparison of a new background to last clean background. A clean background was acquired before each reaction and the probe was inserted directly into the reaction vessel. A spectrum was acquired every 2 minutes with 256 scans averaged and a resolution of 8 cm-1. Spectra were exported in real time from the icIR software (version 4.3, Mettler Toledo Autochem, USA) in the form of .csv files. At a specified rate, the files were read by MATLAB (version R2018B, Mathworks, Natick, USA) where the analysis was completed according to equations 1-4 in the main text by a bespoke algorithm. When the confidence interval was less than zero the algorithm signaled the reaction was within the specification window, the reaction was cooled, and a sample was taken for HPLC analysis.

S2. IR spectra

To determine if IR could be used to monitor the condensation reaction, IR spectra were collected of the pure reaction components in the laboratory with a ReactIR ic10. The normalized raw spectra of the reaction components are shown in Figure S1. Due to the short depth of penetration of ATR-FTIR, typically only species dissolved in solution can be monitored. Due to the limited solubility of starting material **A** and the product **F** in the reaction mixture, the reaction was a slurry. At lower concentration the starting material **A** dissolved in the reaction mixture and could be monitored through the IR peak between 1560 and 1583 cm^{-1}. Though the product could not be tracked with IR, reaction progress could be monitored through the soluble intermediate **E** at 1455 cm^{-1} because the reaction was complete when all the intermediate was consumed. To best resolve the signal at 1455 cm^{-1}, the reaction spectra underwent Savitzky-Golay second derivative preprocessing. The second derivative spectra were inverted and the IR trend used in the t-test algorithm was the peak area of the signal at 1455 cm^{-1}.

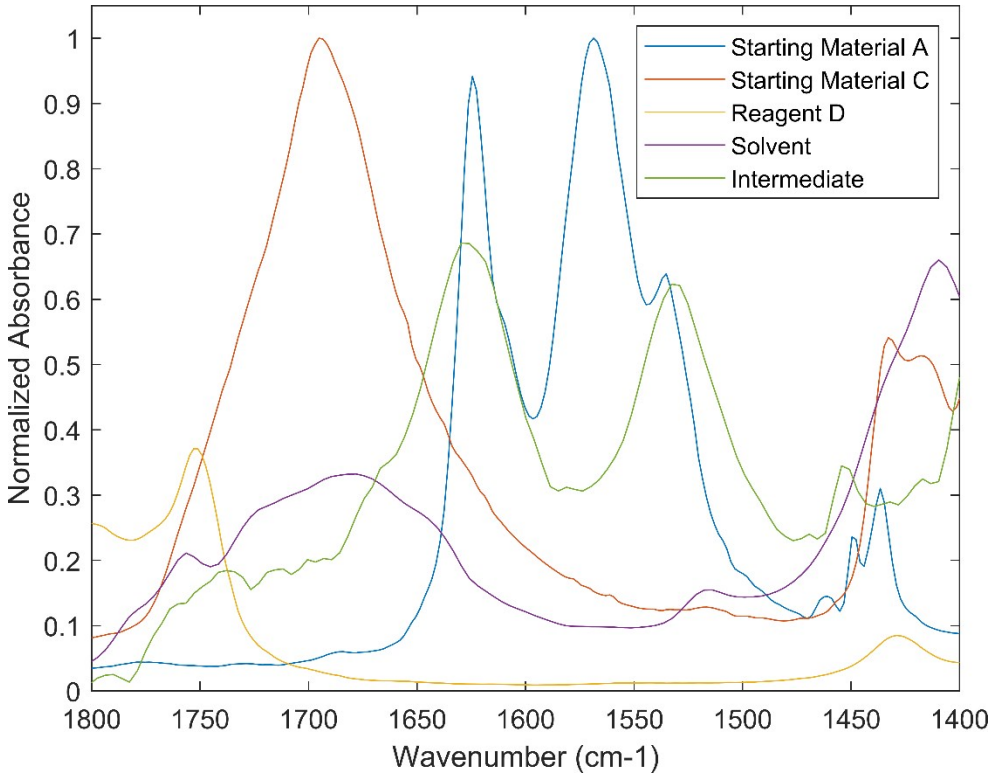
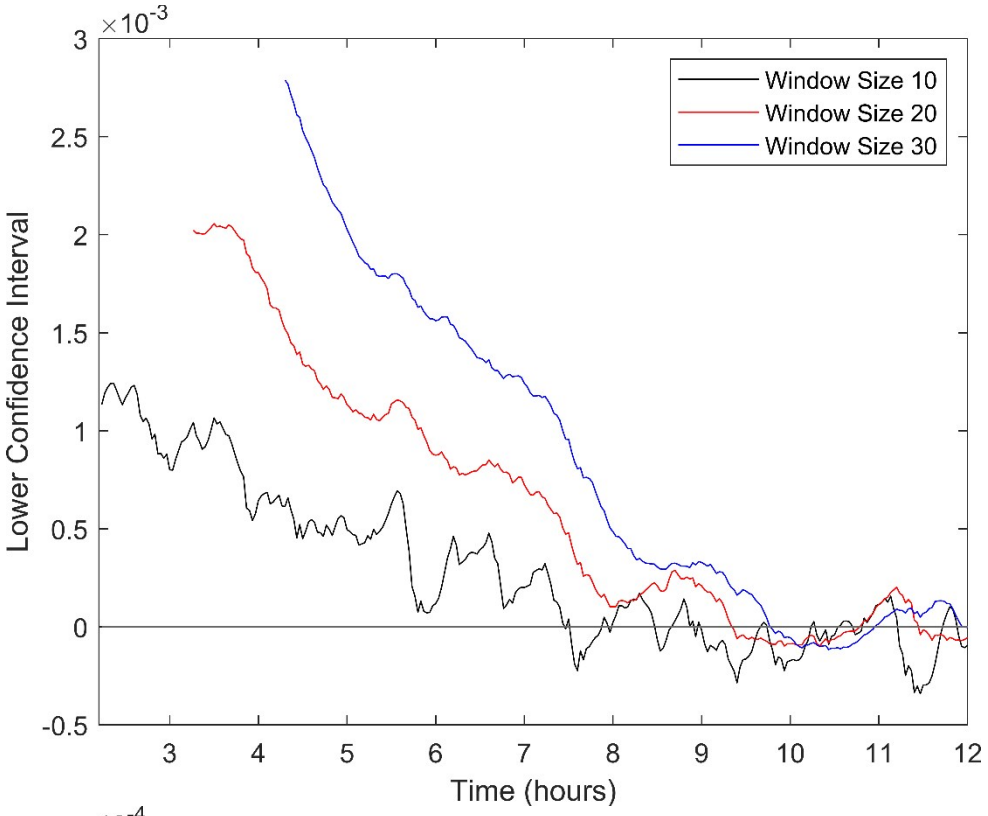


Figure S1. Normalized IR spectra of the different reaction components.

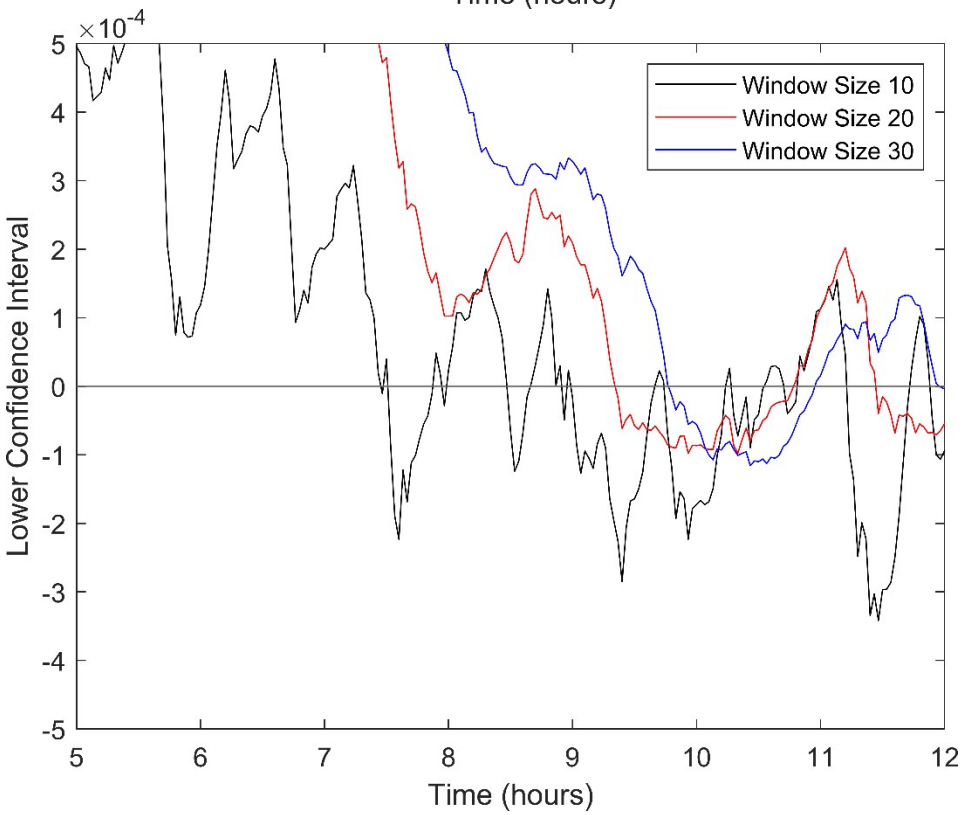
S3. Impact of t-test window size and IR trend smoothing on endpoint

The two parameters that were found to have the largest impact on when the t-test algorithm calculated the endpoint were the t-test window size and the degree of smoothing of the IR trend. The larger the t-test windows size, the more data that must be collected before a result is calculated by the algorithm. After double the t-test window size data are collected, the first calculation can be done. If the IR spectra were collected every 2 minutes and the t-test window size was 20, then the chemist or operator would have to wait 80 minutes before receiving the first result of the t-test calculation. It is ideal to have a calculation as quickly as possible to decrease time the chemist or operator are waiting for a result. Therefore, the smallest t-test window size that estimates the reaction endpoint in the correct range should be selected.

Figure S2a shows the calculated lower confidence interval for different t-test window sizes. When the lower confidence interval is less than zero, the reaction is deemed to be complete with the t-test algorithm. The same data are plotted in Figure S2b to better see where the confidence intervals intersect with zero. As the t-test window size increases, so does the estimated endpoint by the intersection of the confidence interval with zero. The difference between the estimated endpoint for a window size of 10 and 30 was ~2.3 hours.


×10^-3
3
2.5
2
1.5
1
0.5
0
-0.5
Lower Confidence Interval
3
4
5
6
7
8
9
10
11
12
Time (hours)
Window Size 10
Window Size 20
Window Size 30

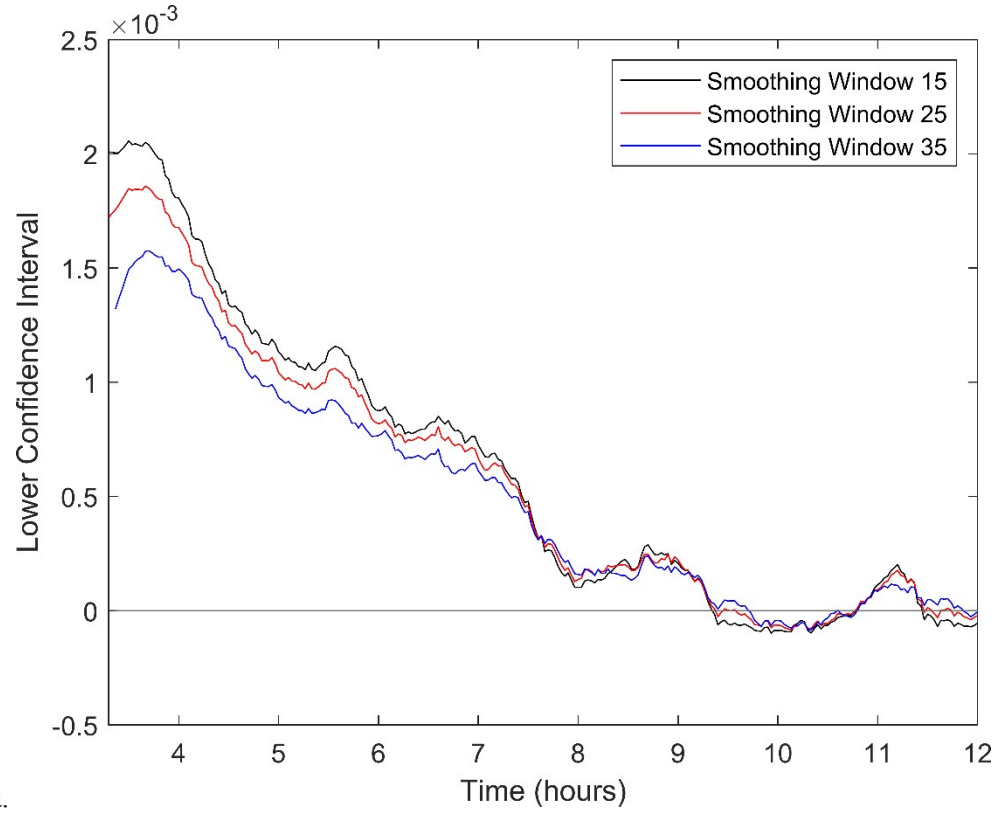
a.


×10^-4
5
4
3
2
1
0
-1
-2
-3
-4
-5
Lower Confidence Interval
5
6
7
8
9
10
11
12
Time (hours)
Window Size 10
Window Size 20
Window Size 30

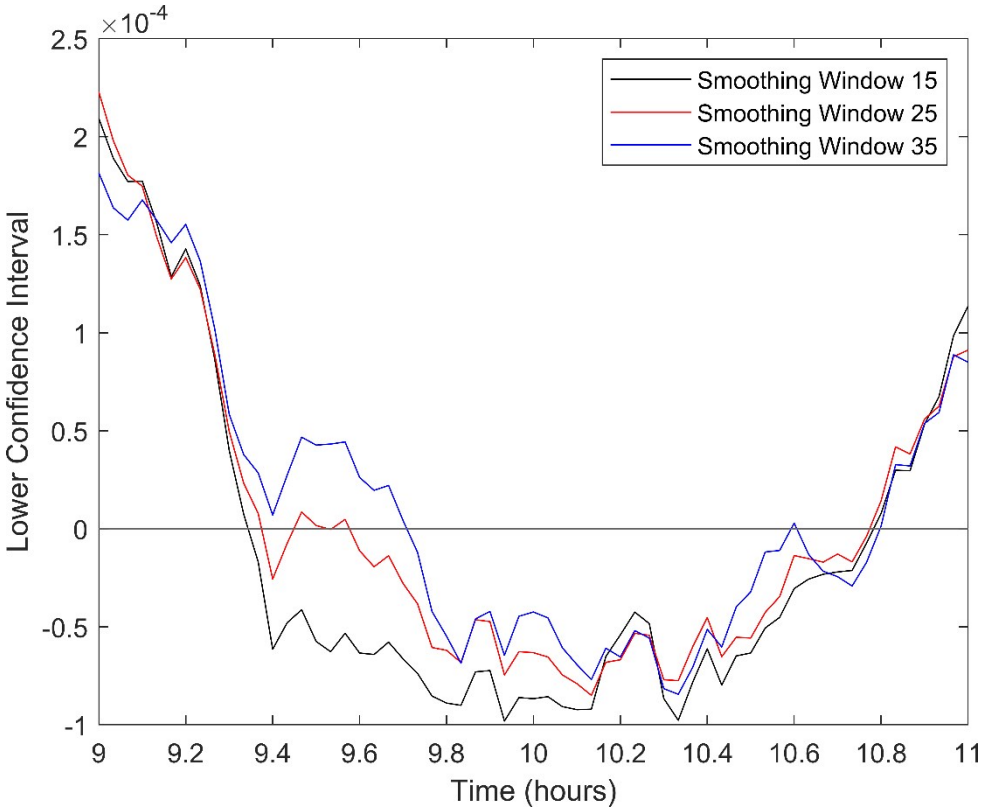
b.

Figure S2. (a) Lower confidence interval calculated from the t-test with different window sizes and (b) zoomed in on intercept with zero.

The smoothing of the IR intermediate **E** trend also impacted when the reaction endpoint was estimated. A moving average was used to smooth the raw IR trend. Figure S3a shows the lower confidence interval for three different moving average window sizes and Figure S3b is zoomed in on where the lower confidence interval crosses zero. Compared to the t-test window, the moving average window has a lower impact on the estimated endpoint. The endpoint was estimated to be around the same place for moving average windows of 15 and 25. However, due to the fluctuations in the difference in means caused by the noise of the IR data, the estimated endpoint with the most smoothing shown, a moving average of 35, was ~0.3-0.4 minutes later.



a.



b.

Figure S3. (a.) Lower confidence interval calculated from the t-test with different window sizes and (b.) zoomed in on intersection with zero.

The impact of the t-test window size and moving average window size of the IR trend shows that these parameters should be carefully chosen before using the t-test algorithm. Prior to using the algorithm to inform HPLC sampling, the t-test window size and smoothing window size were optimized by varying the windows sizes and determining the estimated endpoint on historical data collected with different IR models and under different process conditions. This allowed us to select the appropriate window sizes that would determine the endpoint in the correct range for batches under different process conditions and with data collected with different instruments.