Model averaging to determine reaction process endpoint

using infrared spectroscopy as process analytical technology

Tatsiana Khamiakova | Manufacturing and Applied Statistics
Adriaan Blommaert | Open Analytics | 4/10/2018

Rhonda Fenwick, *Time is Now I*
Through her art, Rhonda has explored psoriasis, a chronic skin disorder she has lived with since the age of six.

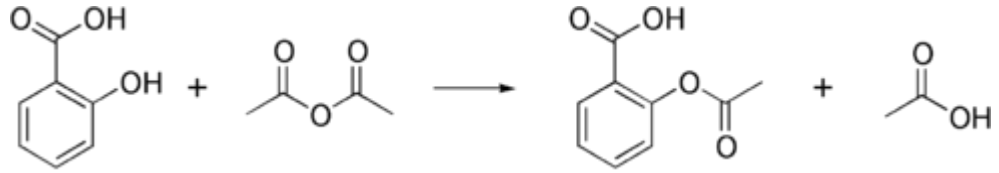
janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson

Outline

- Introduction to Process analytical technology (PAT)
- IR spectroscopy: endpoint determination
- Model averaging approach to reaction trends
- Discussion

API synthesis: reaction monitoring

- Reaction for the synthesis of aspirin*:



- Goal of the process chemist:

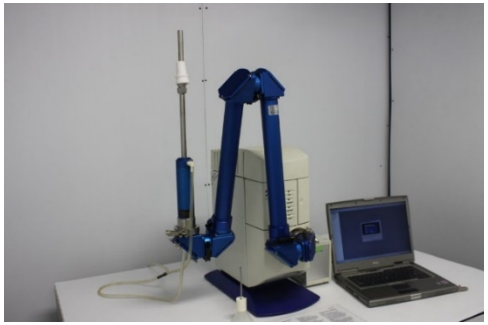
✓ Optimize reaction conditions
✓ Maximize yield
✓ Minimize side products
✓ Maintain robustness of reaction for long period of time (up to 24 h)

* https://en.wikipedia.org/wiki/Aspirin

PAT for reaction monitoring

PAT is a system to monitor and control the process, e.g. synthesis of API in the lab

'Offline' way: take a sample and measure in a separate lab



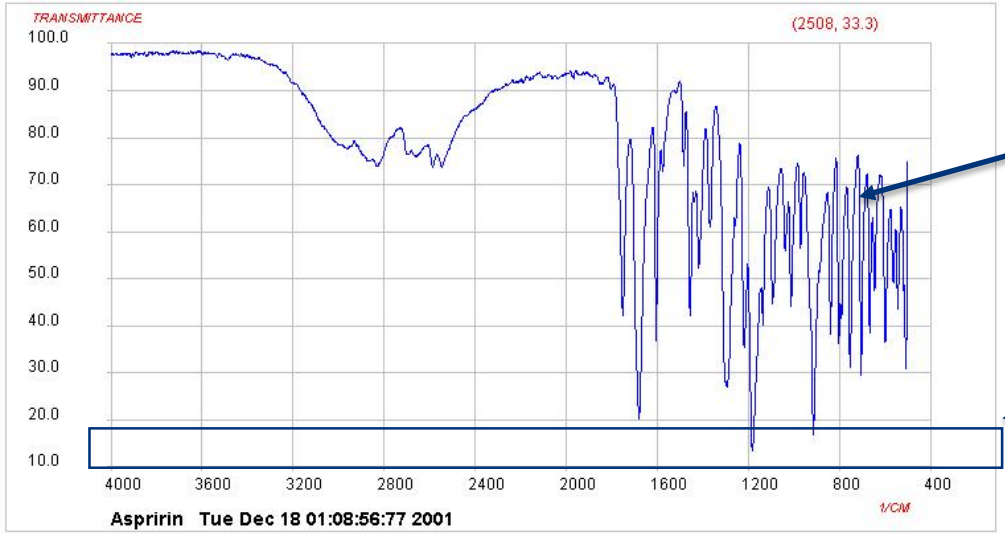
'Online' way: measure during reaction

Advantages of PAT:

✓Fast
✓Not laborious
✓Continuous monitoring

PAT: (mid)infrared (m-IR) spectroscopy

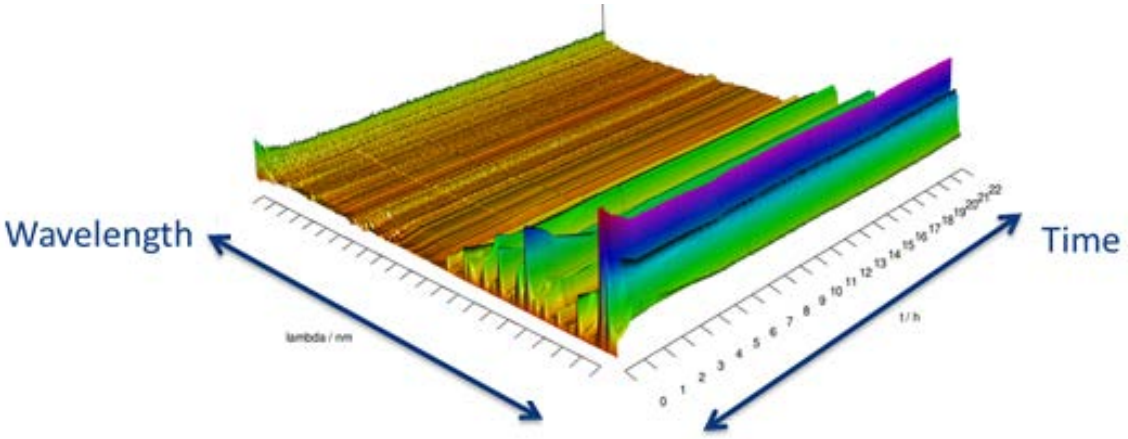
- m-IR continuously measures relative concentrations of reaction components
- For each measurement a spectrum of values is provided:



Peaks (height of the peak proportional to concentration)

Wave numbers

PAT data: a 3D plot



(3D image is created by `spectralAnalysis` R package).

Data: motivating example

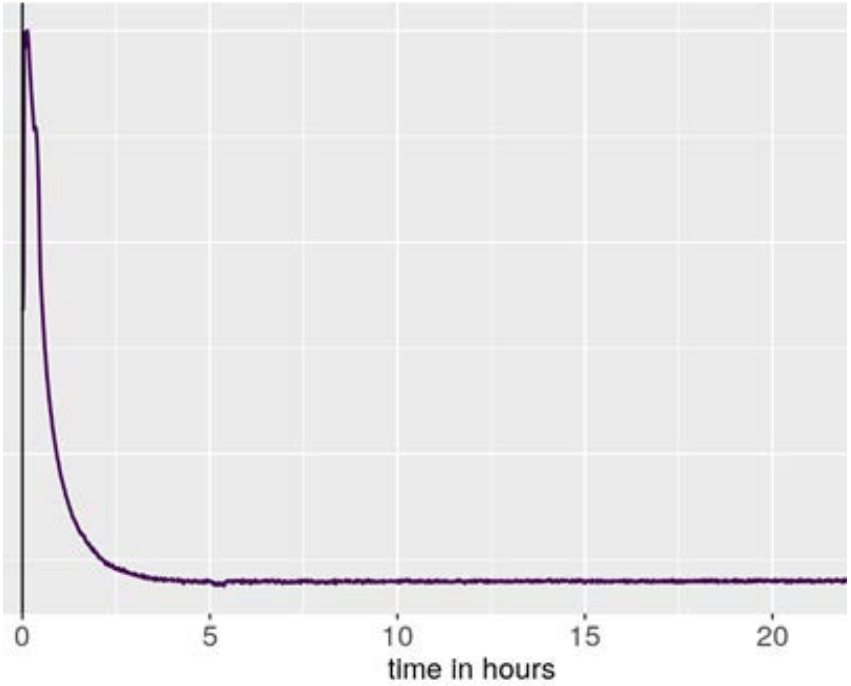
- IR spectra of ~24 hours reaction (in total >1000 timepoints).
- Preprocessing:

✓ baseline corrected

✓ normalized

✓ reaction start time determination

✓ Extracted known peak of starting material (univariate):



Motivating research question: Endpoint

- **Definition of endpoint**

time when all starting material (SM) has been consumed and no more final product can be formed.

- **Data-driven definition of endpoint**

Criterium	Data Type	Challenge
SM < 1%	Offline	Online processing require extra calibration
No change in the trend of SM	Online	Threshold dependent
SM consumed >99%	Online	Rescaling of the trend

Methodology

- An experiment-wise method:

1. Fit parametric non-linear model.
2. Construct (a) confidence interval on asymptote (end value) and (b) prediction interval on trend
3. The uncertainty in model shape/choice of the model is accounted by model averaging.

Non-linear modelling

- Models for non-linear degradation kinetics:

1. $Y_t = A + be^{-rt}$: exponential decay of the 1st order.
2. $Y_t = A + be^{-rt^n}$: exponential decay of the n^{th} order.
3. $Y_t = A + b_1 e^{-rt} + b_2 e^{-(r+s)t}$: biexponential decay (reparametrized to avoid label switching).

- Other functions can be added depending on the knowledge of kinetics.

Model averaging and interval estimation

The method of Burnham and Anderson (2003) based on AIC is applied:

1. Model averaged estimator of a model parameter

$$\hat{\theta} = \sum w_i \, \theta_i \ ,$$

where w_i are weights calculated over the model.

2. If AIC is used for weights (as in our case)

$$w_i = \mathrm{e}^{-0.5\Delta_i} / \sum \mathrm{e}^{-0.5\Delta_i},$$

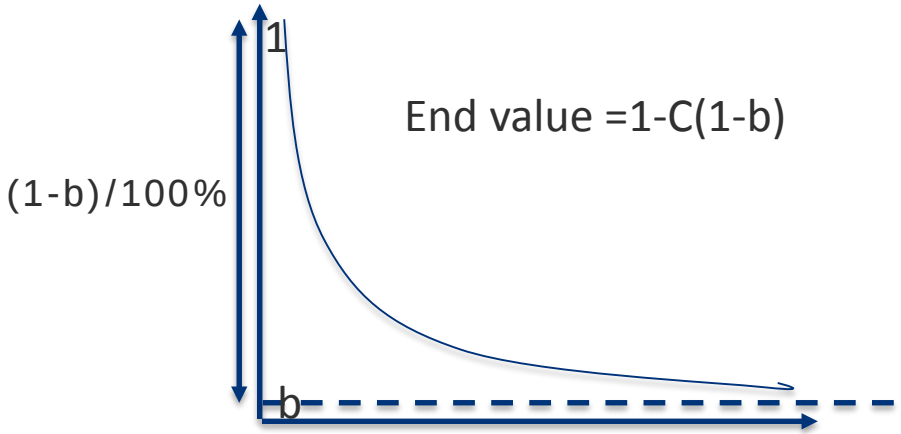
where Δ_i is a difference in weight with best model.

3. The variance of the model-averaged estimator is then

$$\widehat{Var}(\hat{\theta}) = \left[\sum w_i \sqrt{var(\hat{\theta}_i | g_i) + (\theta_i - \bar{\theta})^2)} \right]^2 .$$

4. Based on the estimate and its variance the 95% confidence interval will be constructed using Z-method.

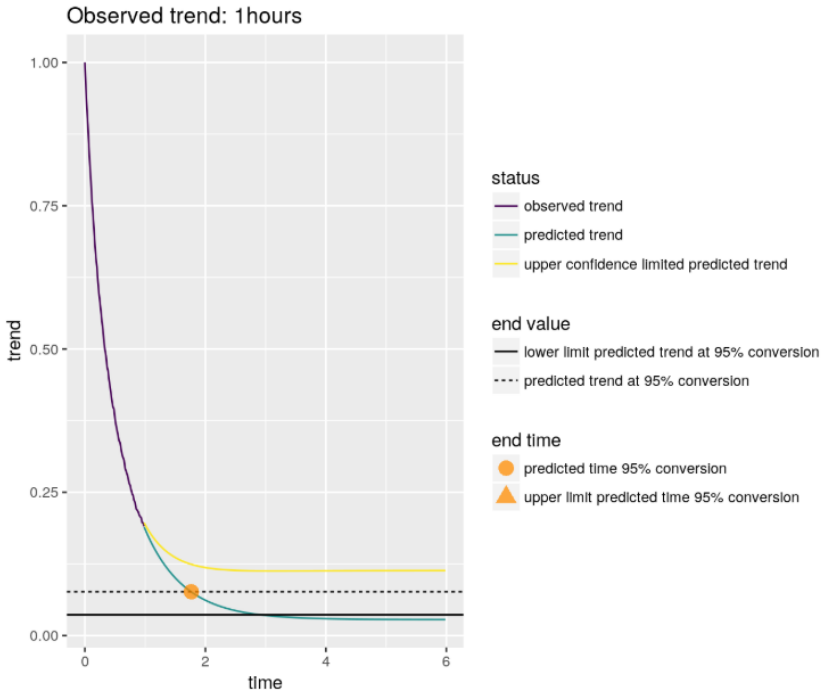
Endpoint calculation based on the interval estimates



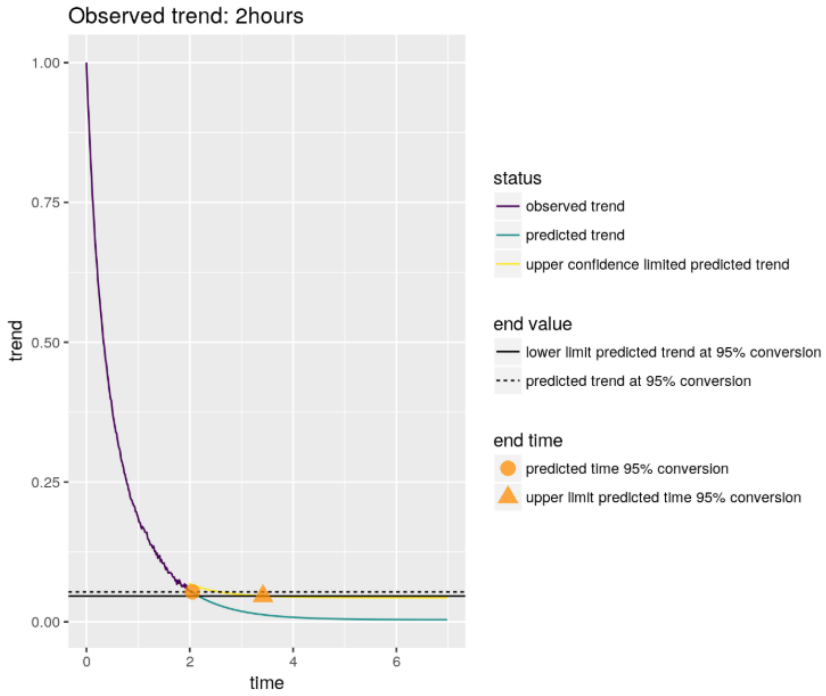
1. Obtain 95% lower confidence limit for the asymptote of the process (estimate of the end value).
2. Obtain 95% upper prediction limit for the trend using delta-method.
3. Obtain the estimate of the conversion time at desired level (e.g. C=95%, 97%, 99%...) using direct calculation.
4. Get the predicted time and the upper 95% confidence limit of the time as follows:

- endpoint time = time when predicted trend has reached 1% (in case observed did not reach this limit);
- upper 95% limit of the endpoint time = the first time when the lower 95% LCL of the asymptote is reached by the 95% UPL of the trend.

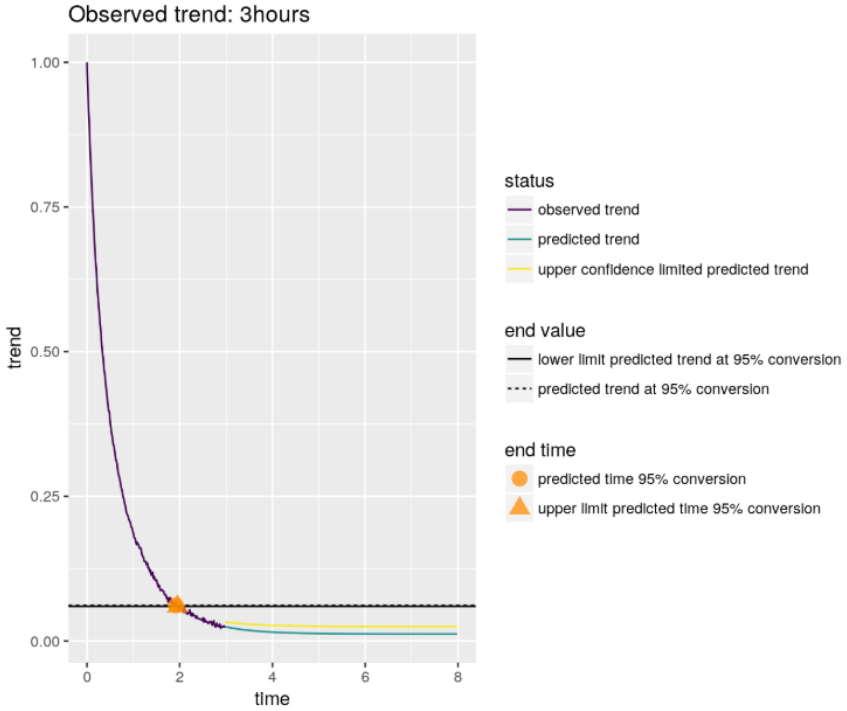
Endpoint detection: illustration



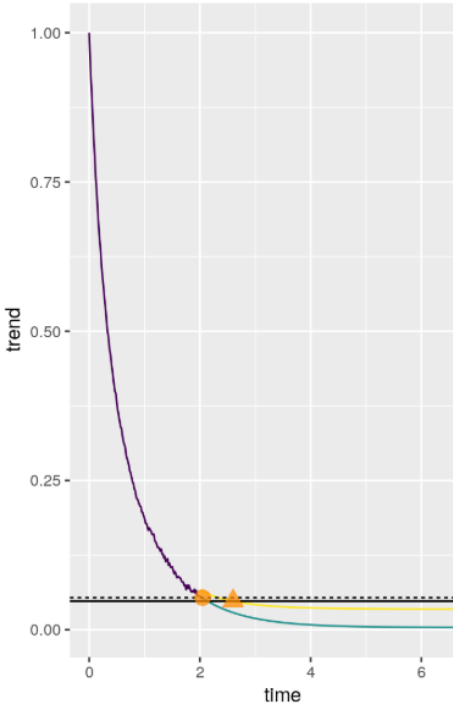
Endpoint detection: illustration



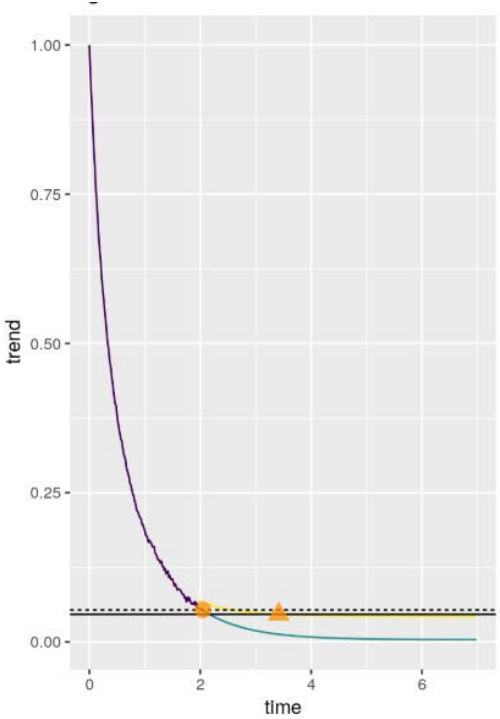
Endpoint detection: illustration



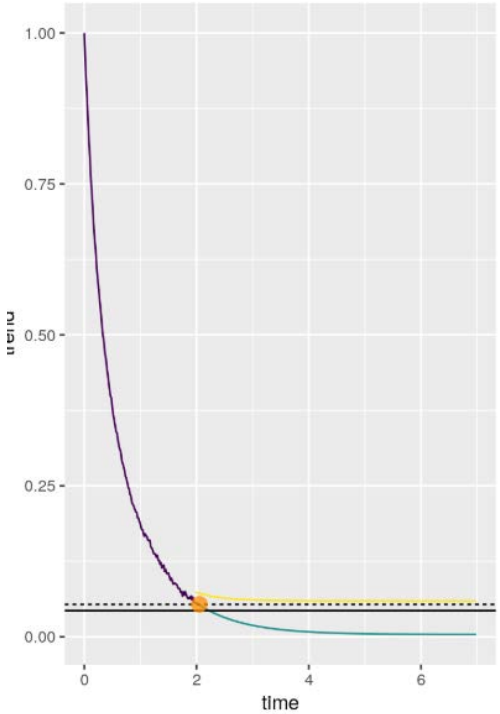
Endpoint detection wrt Confidence Level



90% confidence level

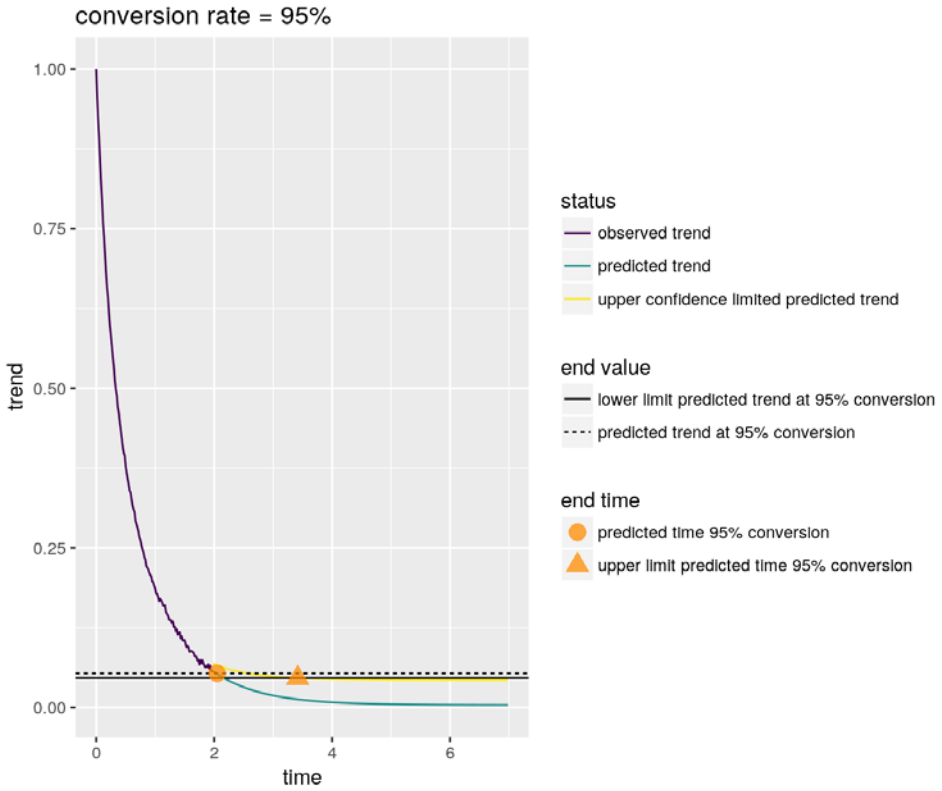


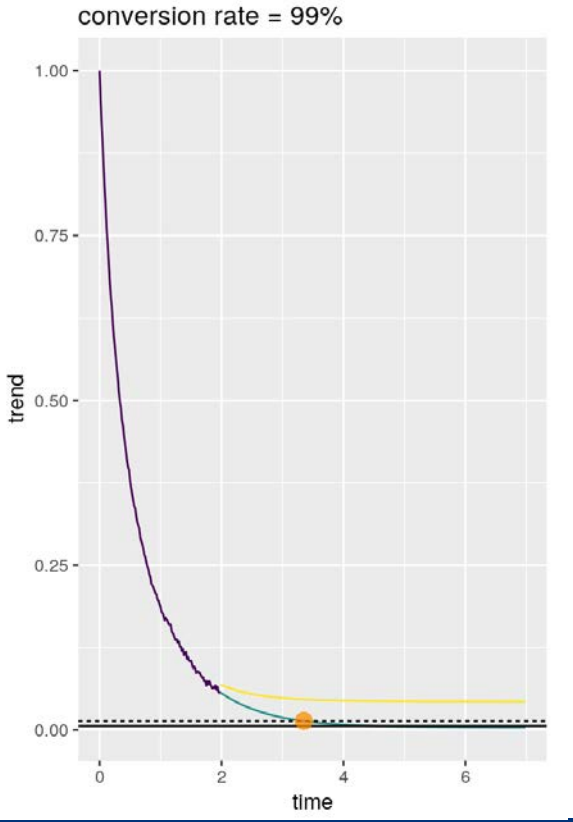
95% confidence level



99% confidence level

Endpoint detection wrt endpoint definition





Discussion: Advantages

- Real-time data processing by non-linear modelling.
- No external calibration needed.
- No extra experimental runs required.
- Flexible set of kinetic models.

Discussion: Limitations

- Reaction start point uncertainty.
- Sensitivity to choice of conversion percentage.
- Noise in online PAT measurements.

Assumptions and further research

- Conversion is reached at asymptote.
- No drift in measurements (cfr. heat-generating process).
- Normal distribution of the residuals.
- Separate construction of the intervals.
- Bayesian paradigm (simultaneous estimation) may be unfeasible for online application.
- Parametric bootstrap.

Conclusions

- IR spectroscopy allows for online monitoring of chemical processes.
- The generated IR data can be used for real time endpoint detection in development phase.
- Non-linear modelling together with model averaging are tools for real-time prediction which are experiment-specific.
- Some further research and extensions are needed to make this methodology applicable for a variety of processes.

Acknowledgments

Open Analytics:
Adriaan Blommaert
Nicolas Sauwen

Janssen, Manufacturing and Applied statistics:
Michel Thiel
Hans Coppenolle

Janssen, DPDS SMPD API:
Tor Maes
Jan Dijkmans

References

(1) Assessment of Recent Process Analytical Technology (PAT) Trends: A Multiauthor Review. (2015) Org. Process Res. Dev. 19, 3−62. dx.doi.org/10.1021/op500261y.

(2) Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance (2004). FDA guidance: Pharmaceutical CGMPs.

(3) Burnham, Kenneth P, and David R Anderson. 2003. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer Science & Business Media.

(4) R package spectralAnalysis: https://cran.r-project.org/web/packages/spectralAnalysis/index.html

Thank you!

tkhamiak@its.jnj.com

04/10/2018

Rhonda Fenwick, *Time is Now I*
Through her art, Rhonda has explored psoriasis, a chronic skin disorder she has lived with since the age of six.

janssen | PHARMACEUTICAL COMPANIES OF Johnson & Johnson


janssen
PHARMACEUTICAL COMPANIES OF
Johnson & Johnson