ELSEVIER

Catalysis Today 81 (2003) 359–367

CATALYSIS TODAY

www.elsevier.com/locate/cattod

Optimal on-line sampling of parallel reactions: general concept and a specific spectroscopic example

Gadi Rothenberg [a,*], Hans F.M. Boelens [a], David Iron [b], Johan A. Westerhuis [a]

[a] *Chemical Engineering Department, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*
[b] *KdV Institute, Universiteit van Amsterdam, Nieuwe Achtergracht 166, 1018 WV Amsterdam, The Netherlands*

Received 16 June 2002; received in revised form 18 September 2002; accepted 26 September 2002

Abstract

The 'gap' between the discovery and the optimisation phases in high-throughput systems is described and the problems that stem from the centralised nature of parallel reactor set-ups are outlined. It is proposed that on-line analysis methods (e.g. fast spectroscopic analysis) can be adapted to perform kinetic studies in high-throughput environments. Strategies for performing time-resolved analysis are examined using a (pseudo)first-order reaction that is monitored by on-line FT-NIR as an example. Stochastic modelling is used to compare the efficiency of the various strategies in the determination of accurate first-order kinetic constants. The effects of the sampling time, the 'sample window' size, and the number of samples on the error in k_{obs} are studied. A new sampling strategy is introduced that is based on equidistant sampling along the concentration axis. This strategy is adapted to realistic conditions using a simple iterative algorithm. The application of this iterative approach to automatic intelligent on-line analysis of chemical reaction kinetics is discussed.

Keywords: High-throughput experimentation; Sampling strategies; Split/pool methods; Combinatorial catalysis; Iterative algorithms

1. Introduction

Combinatorial chemistry and high-throughput experimentation (HTE), the powerful tools of the pharmaceutical industry, have quickly spread in the last decade also to the areas of catalysis (combicat) and materials research. Several excellent reviews cover the application of HTE in these fields.[1] Following the example set by the pharmaceutical screening approach, many ingenious and elegant methods for the synthesis and testing of libraries of new catalysts and materials have been reported [2]. Common to all these new systems is the basic two-step approach comprised of primary screening of a large number of candidates (discovery stage) followed by optimisation of a small number of leads.

One problem is that this two-stage approach dictates a low information content per reaction (often only one binary parameter is scanned in the discovery stage). Thus, for example, good catalysts may be overlooked if they do not exhibit, for any reason, high activity in the initial tests. A possible solution to this problem would be to adapt time-resolved analysis to high-throughput environments, establishing a bridge between the discovery and optimisation stages.

* Corresponding author.
E-mail address: gadi@science.uva.nl (G. Rothenberg).

[1] For reviews on the application of combinatorial techniques to heterogeneous catalysis see [1a,1b]. For applications to materials science see [1c,1d,1e,1f]. For applications to biocatalysis see [1g].

0920-5861/03/$ – see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0920-5861(03)00134-2

Unfortunately, the centralised structuring of high-throughput reaction set-ups has one critical drawback as far as quantitative time-resolved analysis is concerned: most robotic systems consist of many (small and inexpensive) reactors and one (expensive) analyser. Thus the time of the analyser is of essence not only in the economical respect, but also simply because time does not wait for parallel experimentation.

It is unlikely that the fundamental structure of such laboratory systems is going to change (somehow robots consisting of 300 reactors and 300 GC instruments do not seem like an attractive solution). Coupling HTE and on-line analysis seems more practical, and one attractive analysis option is using on-line spectroscopy. State-of-the-art Fourier-transform spectrophotometers can, in theory at least, be used to track a large number of kinetic profiles simultaneously. However, the development of novel parallel experimental set-ups must go hand-in-hand with the research of new concepts in analysis [4].[2] In this paper, we propose a novel sampling strategy and compare it to existing laboratory practice. We use stochastic modelling to evaluate various sampling strategies and demonstrate that the estimation of reaction rate constants in high-throughput set-ups can be based on a small number of measurements and still give accurate results. Although the model parameters are based on spectroscopy the results are general and require only that the analysis method be fast, quantitative, and performed on-line.

2. Results and discussion

2.1. Concept and reaction parameters

Measuring the extent of a chemical reaction over time yields always a rate constant k_{obs} that is only an *estimate* of the true rate constant of the process k_{true}, and has a certain error value Δk that is comprised of systematic (bias) and random errors [5]. If spectroscopic methods are used, this Δk depends on the experimental set-up and on the way that the spectra are processed. In other words, Δk is influenced by both the measurement itself[3] and by the calibration model that is applied.[4]

Let us consider as a model reaction the (pseudo)first-order reaction $\mathbf{A} \rightarrow \mathbf{B}$, where the initial concentration of **A** is $[\mathbf{A}]_0$, the initial concentration of **B** is 0, the reaction rate constant is k_{true} and $[\mathbf{A}]_t$ is measured from time t_0 to time t_∞ using vibrational spectroscopy (e.g. FT-NIR or Raman).[5] The measurements are sample windows of certain width N_{scans} (i.e. each sample is an average over N_{scans} scans at a given wavenumber resolution). Each of these windows yields an average time value and an average concentration value (see Fig. 1a).

To assess the effectiveness of different sampling strategies, a large number of stochastic white noise vectors is generated to simulate the noise that is observed in the determination of the concentration. The model then examines the accuracy for different numbers of samples that are distributed according to a given strategy.[6] The window size n is also varied for each given number of samples, resulting in a three-dimensional matrix that details the *relative* accuracy of k as a function of the number of samples and the window size. The simulations are described in detail in the methods section.

2.2. Traditional kinetic studies versus high-throughput kinetics

The typical recipe followed when studying chemical reactions is to sample the reaction at regular intervals

[2] For a discussion on the experimental design of HTE see [3a]. For an application of ATR spectroscopy to high-throughput studies of polymers see [3b].

[3] Spectra can be measured either with dispersive or with Fourier-transform spectrometers. In both cases the spectrum is based on a window comprised of a number of scans averaged over time. This time averaging increases the signal-to-noise (S/N) ratio of the spectrum. However, the disadvantage of increasing the number of scans per window is that the total time needed to collect one spectrum also increases and meanwhile the reaction continues and concentrations change.

[4] A calibration model must be applied to the measured spectrum to extract concentration estimates. The error of this estimates depends on the model because any calibration model is based on spectroscopic and reference measurements pertaining to a finite number of calibration samples. For a discussion see [6].

[5] Here t_8 is defined as the time when conversion >99.5%.

[6] Accuracy is taken as the degree of closeness between the test unit and an accepted reference value (see [5], p. 41). There are various ways to define this closeness, and here we use |bias| + standard deviation as the measure of accuracy, where the bias is defined as the difference between the estimated average concentration and the true concentration. Similar results are obtained when the root mean square accuracy is used.

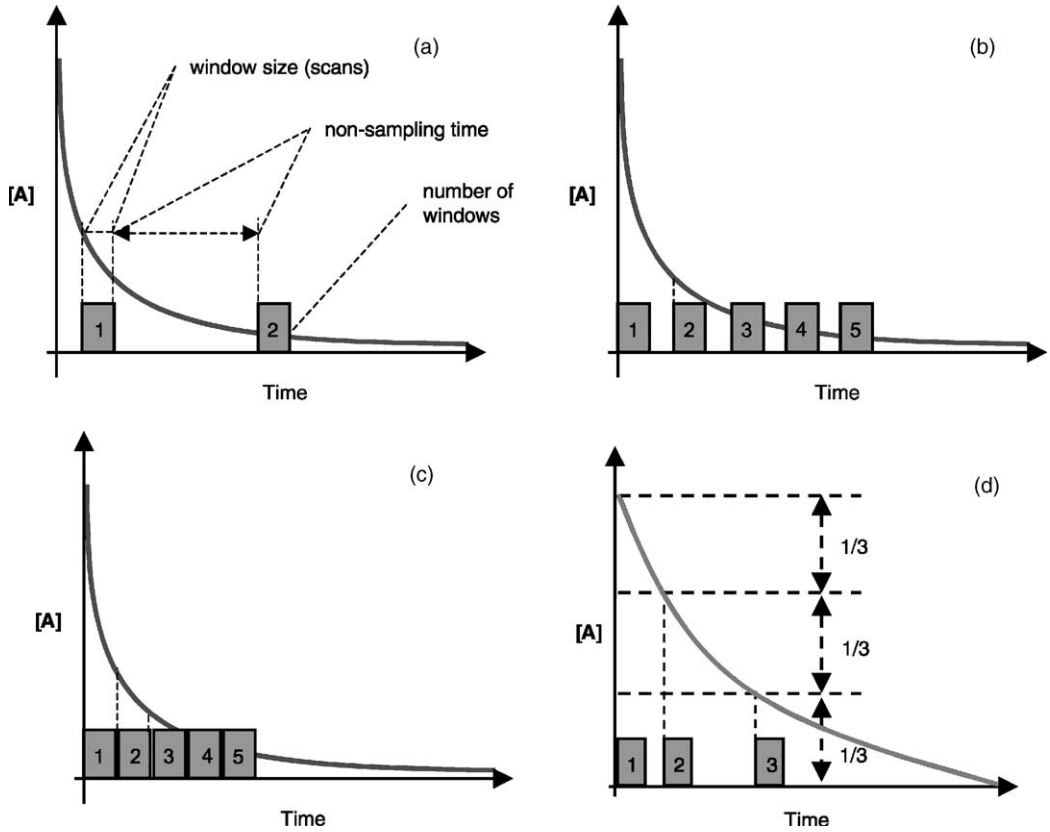

Fig. 1. System parameters and sampling strategies for following a first-order reaction. (a) System parameters. (b) Sampling strategy **S1**—*even distribution of samples along the time axis*. (c) Sampling strategy **S2**—*pack the samples at the start*. (d) New sampling strategy **S3**—*even distribution of samples along the concentration axis* (example shown with three equidistant samples, the green concentration curve has been shifted for clarity).

(e.g. to take 10 samples over 1 h you sample every 6 min). We will refer to this method as strategy **S1**, the *even distribution of samples along the time axis* (see Fig. 1b). Since in many investigations the comparison of initial reaction rates is of interest, a second common strategy is to sample frequently at the beginning of the reaction, i.e. to *pack the samples at the start* (strategy **S2**, shown in Fig. 1c). In both cases taking more samples should decrease the magnitude of Δk (in practice, 5–20 samples over the reaction period are considered adequate to obtain 'accurate' k_{obs} values).[7]

In sharp contrast, kinetic experiments in high-throughput set-ups have to be based on as few samples as possible in order to get the best utilisation of the measuring instrument. With conventional sampling this would lead to inaccurate results but the combination of "intelligent" sampling and on-line analysis can solve this problem.

Figs. 2 and 3 show the simulation results for sampling strategies **S1** and **S2** (note the difference in scale along the *z*-axis). The results affirm the rule of thumb for taking more samples (in both cases the error decreases as larger and/or more sample windows are taken). Note too that in both cases increasing the window size and taking more samples has a similar effect on Δk.

[7] Regrettably in numerous cases Δk is not reported. Based on our own experience the magnitude of Δk in organic chemistry/catalysis studies is ca. ±0.5–2.0%. See, for example [7].

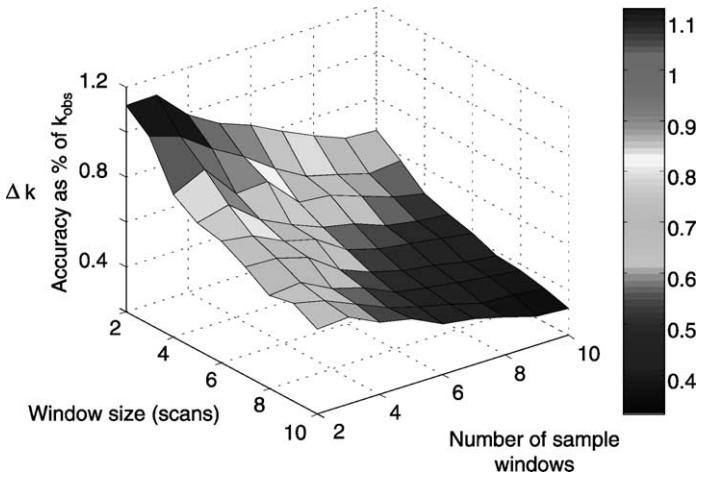

Fig. 2. Relative accuracy (% of k_{obs}) of the reaction rate constant using sampling strategy **S1**. Conditions: $[\mathbf{A}]_0 = 1\,\mathrm{M}$, $k_{true} = 1/60\,\mathrm{min}^{-1}$, $t_\infty = 120\,\mathrm{min}$, 1000 simulations.

Strategy **S1** relies too heavily on information at low concentrations, where measurement errors can lead to large errors in k. On the other hand, **S2** yields measurements with high concentration values, but since they are *all* placed close together in time it is difficult to derive an accurate kinetic profile (it can be seen by comparing Figs. 2 and 3 that using method **S2** more measurements are required to reach similar error margins). Thus, a sampling strategy combining the advantages of **S1** (good spread over the time axis) and **S2** (measurements with high signal values) could be beneficial.

We propose here a new sampling strategy, **S3**, which is a compromise between **S1** and **S2**. This strategy spreads the samples in time so that the change in concentration, $\Delta[\mathbf{A}]$, is equal (Fig. 1d, *equidistant distribution of the samples along the concentration axis*). The results, shown in Fig. 4, are striking—the highest

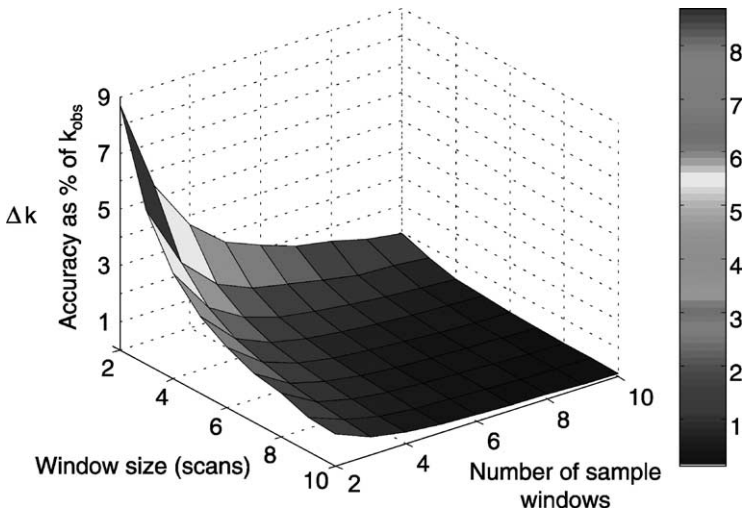

Fig. 3. Relative accuracy (% of k_{obs}) of the reaction rate constant using sampling strategy **S2**. Conditions: $[\mathbf{A}]_0 = 1\,\mathrm{M}$, $k_{true} = 1/60\,\mathrm{min}^{-1}$, $t_\infty = 120\,\mathrm{min}$, 1000 simulations.

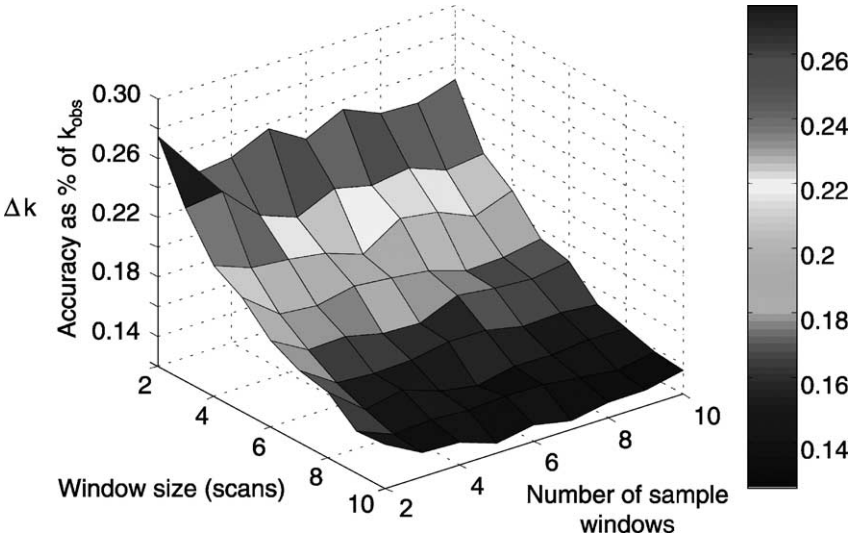

Fig. 4. Relative accuracy (% of k_{obs}) of the reaction rate constant using sampling strategy **S3** under ideal conditions, i.e. assuming that the value of k_{true} is known. Conditions: $[\mathbf{A}]_0 = 1\,M$, $k_{true} = 1/60\,min^{-1}$, $t_\infty = 120\,min$, 1000 simulations.

value obtained for Δk is only ±0.27% (cf. again the different scales in Figs. 2–4). Furthermore, increasing the number of sampling windows in this case has very little effect on the relative error in k.[8] Increasing the sample window size, on the other hand, improves the accuracy considerably.

2.3. An iterative approach to time-resolved sampling

The problem in the above implementation of **S3** is that you have to know exactly when to sample your reaction. This means that you have to know k_{true} in advance, and this is ultimately the value that you want to obtain from the kinetic studies! A much more realistic set-up would be to use a model that does not "know" k_{true} but is capable of "self-adaptation". The simple iterative algorithm shown in Fig. 5 fulfils these requirements. This algorithm begins with an initial guess for the rate constant (k_{init}) that is much higher than k_{true}.[9] It then "measures" a given number of samples, and

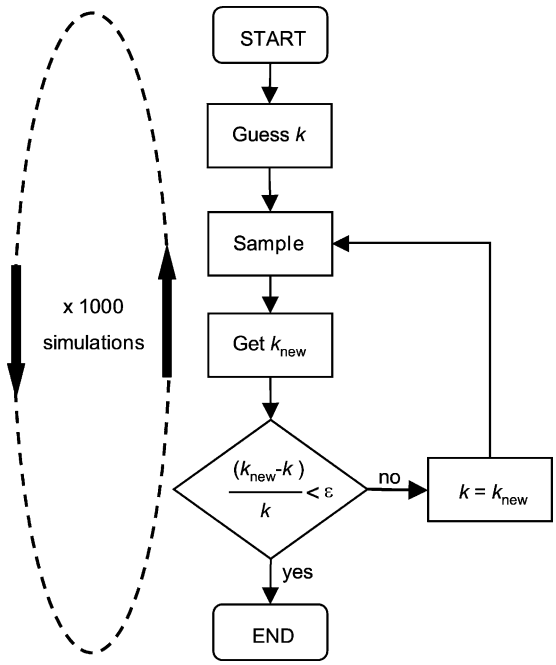

Fig. 5. Iterative sampling algorithm that guesses a large k_{init} value and converges on the real rate constant k_{true}.

[8] The small imperfections observed along the number of samples axis are a result of the discrete structure of the time vector that does not have to be an integer multiple of the number of windows.

[9] In the example shown in Fig. 7 the initial guess for k is $1\,min^{-1}$ and the true k value is $1/60\,min^{-1}$.

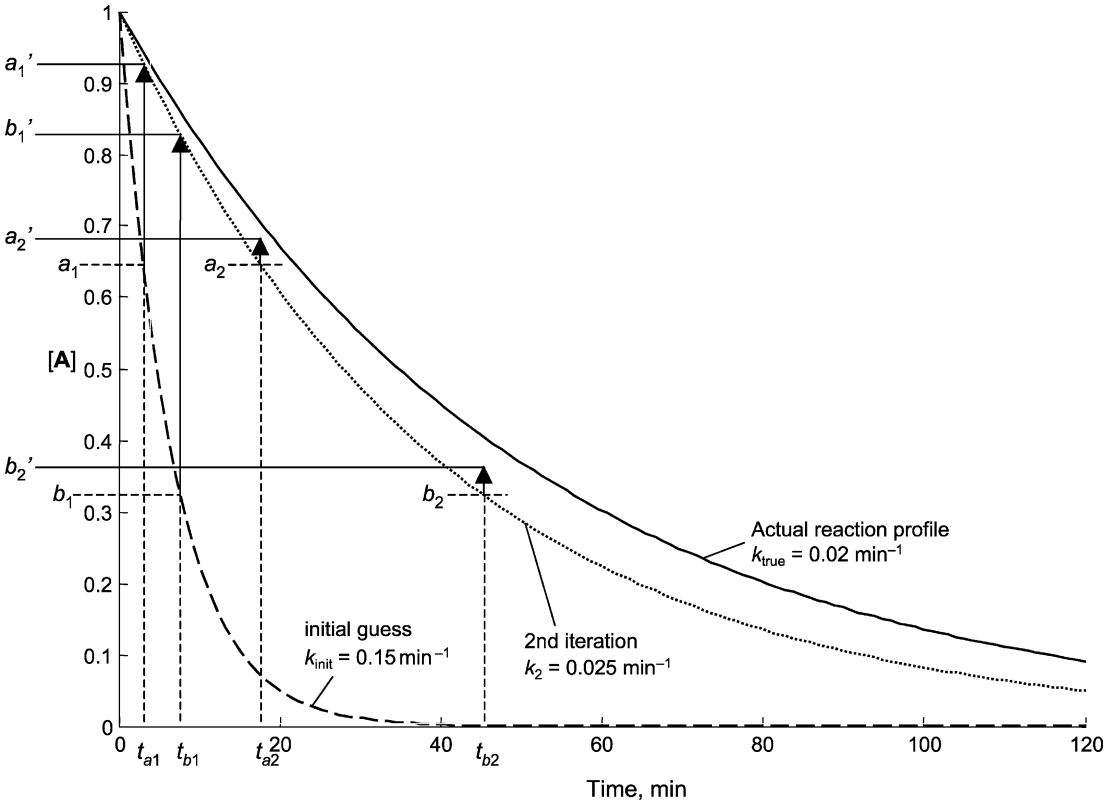

Fig. 6. Example detailing the progress of the iterative sampling algorithm shown in Fig. 5, using two samples and with $k_{init} = 0.15\,min^{-1}$ (dashed line). In this example, sampling at times t_a and t_b is expected to yield the concentrations a_1 and b_1. Instead, the concentrations a_1' and b_1' are obtained, that are within the instrumental measurement error. A second iteration (dotted line) is performed and two more samples are taken, this time at times t_{a_2} and t_{b_2}, and so forth until the results converge on the actual reaction profile (continuous line). The above arbitrary k values have been chosen to provide a clear pictorial representation (in the simulation convergence is faster, i.e. the difference between k_2 and k_{true} is smaller).

based only on the sampled information generates a new k value and compares it to the previous guess. The reaction is then started again[10] and the process is repeated until the k value has converged (typically 3–4 cycles). This value is stored and a new simulation is started. After 1000 simulations have been performed, the average k is compared with k_{true} to obtain the bias and standard deviation values.

Note that from the practical point of view it is beneficial to choose a large k_{init} value if no prior knowledge is available. This is because a high k_{init} value would result in sampling at earlier points on the time axis, and even if the initial guess is wrong a rough approximation for k_{true} could be obtained after a short time span. Fig. 6 illustrates this by way of example, showing how the first and second iterations progress ($k_{init} = 0.15\,min^{-1}$, $k_{true} = 0.02\,min^{-1}$, two samples). In the first iteration in this example the "optimal" sampling times t_{a_1} and t_{b_1} are calculated using k_{init} as a guess (dashed line), with respective expected concentrations a_1 and b_1. The measurements yield the true concentrations a_1' and b_1', that are within the measurement error margin, and a new k_2 value (in this example $k_2 = 0.025\,min^{-1}$) is generated using these values. In the second iteration (dotted line) k_2 is considered to be the reaction constant, and two more samples are taken, this time at times t_{a_2} and t_{b_2}. The expected

[10] It is assumed that in the HTE set-up it is possible to reproduce the reaction conditions accurately.

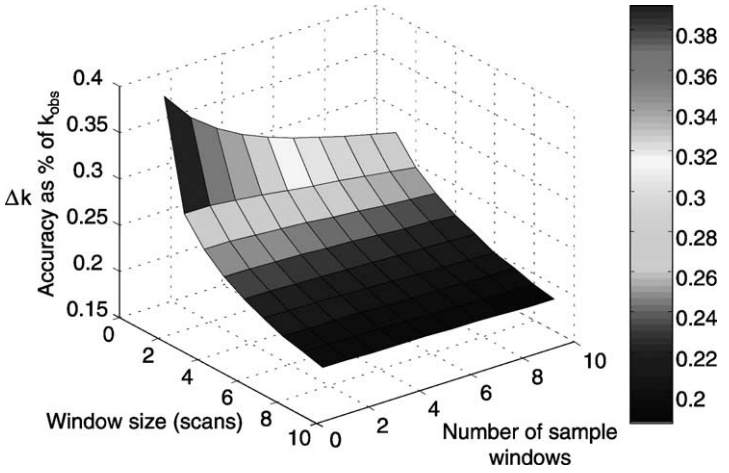

Fig. 7. Relative error (% of k_{obs}) of the reaction rate constant using the iterative algorithm shown in Fig. 5 to run strategy **S3**, where the value of k_{true} is unknown and the only data available to the model is what has been sampled. Conditions: $[\mathbf{A}]_0 = 1\,\mathrm{M}$, $k_{true} = 1/60\,\mathrm{min}^{-1}$, initial guess for $k_{init} = 1\,\mathrm{min}^{-1}$, $t_\infty = 120\,\mathrm{min}$, 1000 simulations, with each simulation limited to three iterations to enable quantitative comparison.

concentrations are a_2 and b_2, and the real concentrations measured are a'_2 and b'_2. All of the four samples a'_1, b'_1, a'_2, and b'_2 are then used in the estimation of the new constant, k_3, for the third iteration, etc. The iterations stop when the new value of k_n is within a small ε of the current value k_{n-1}.

The results for this iterative approach, shown in Fig. 7, are remarkable. Even with three iterations using two samples of two scans each the error is already on the same scale as taking 10 samples of 10 scans each using the 'conventional' strategies **S1** or **S2** (cf. Fig. 7 with Figs. 2 and 3).[11]

The important difference between the non-iterative models and the iterative one is that the latter does not have any access to "true" data unless it has sampled it. This is similar to the real situation in the laboratory, and means that on-line spectroscopic analysis combined with iterative sampling can be a practical solution for high-throughput kinetics. For a given library of parallel reactions, fast analysis of three to five samples from every reaction should suffice to obtain reaction profiles that would be as accurate as those obtained with 10–20 samples using conventional strategies. What is more, after sampling the first 2–3 points from a given vessel, the control program of the spectrophotometer could "decide" on the optimal time to take the following samples from it. In this way, the allocation of the instrument time could be easily optimised.

Another interesting point is that this iterative sampling approach allows the user to specify the end accuracy of the kinetic constant. This is a key step towards bridging the discovery and the optimisation phases, as one could shift between taking "quick and dirty" reaction profiles (discovery) and more accurate profiles for chosen candidates (optimisation).

3. Computational methods

All simulations were performed in MATLAB version 6.1.[12] Resolution along the time axis and standard deviation magnitude values were taken as those of a Perkin-Elmer FT-NIR Spectrum GXIII instrument used in our laboratory to monitor concentration changes in (pseudo)first-order reactions. In all cases, the total number of data points was at least 10 times as large than the actual number of points used for

[11] In the specific case of the simulation shown in Fig. 7, the number of iterations has been fixed at three to allow a quantitative comparison between this simulation and the results shown in Figs. 2–4.

[12] MATLAB is commercially available from MathWorks Inc., Natick, MA, USA.

Table 1
Overview of simulation parameters

Symbol	Meaning	Value	Unit
$[\mathbf{A}]_0$	Initial reactant concentration	1	M
k_{true}	Kinetic constant	1/60	min^{-1}
Δt	Time increment	10[a]	s
t_∞	End time reaction	120	min
N_{sim}	Number of simulations	1000	–
σ_c	Standard deviation of concentration error	0.002[b]	M
$N_{windows}$	Maximum number of samples	10	–
N_{scans}	Maximum number of scans	10	–
k_{init}	Initial guess for kinetic constant[c]	1	min^{-1}
ε	Relative change in k (kinetic constant)[c]	1	Percent

[a] Value based on one scan, absorption spectra (resolution $2\,cm^{-1}$) collected on a Perkin-Elmer FT-NIR Spectrum GXIII spectrometer.
[b] Value derived from an initial NIR calibration model for a Heck reaction on the same spectrometer (transmission spectroscopy, light path length 2 mm, cuvette).
[c] Iterative sampling only.

sampling. That is, the total time segment sampled was never more than 10% of t_∞. It is assumed that the errors introduced by the calibration model and by the estimation procedure are fixed and independent of the sampling strategy. The settings for the simulation parameters are listed in Table 1.[13]

3.1. Generation of the true concentration profile

The true concentration profile $[\mathbf{A}]_t$ was generated using the kinetic constant k_{true}. The concentration profile was calculated from $t = 0$ to t_∞ using a time increment of Δt. The total number of time points generated, N_{time}, was equal to $(t_\infty/\Delta t) + 1$ and the initial concentration $[\mathbf{A}]_0$ was assumed to be known.

3.2. Generation of simulated concentration profiles

N_{sim} simulated concentration profiles were generated. The concentration measurement errors were modelled as independent, Gaussian distributed, $N(0, \sigma_c{}^2)$ random numbers. A sequence of such random numbers is a white noise sequence. Each simulated concentration profile is the sum of the true concentration profile and a new white noise sequence.

3.3. Examination of the various sampling strategies

Each simulated concentration profile can be considered as a 'real' concentration profile originating from exactly the same reaction. A number of 'spectroscopic measurements' is performed during this reaction. One full spectroscopic measurement is a window (see Fig. 1a above) that consists of a given number of sub-measurements, (scans). The number of scans is the window size. In the simulations the number of spectroscopic measurements (windows) is varied between 2 and $N_{windows}$ and the window size between 1 and N_{scans}.

Note that the number of time points (N_{time}) is usually not an integer multiple of the number of windows. In such cases the algorithm moves the start of the sample window to the nearest time point. As both in these simulations and in real life N_{time} is much larger than $N_{windows}$ the resulting rounding error is negligible.

3.4. Implementation of the iterative form of strategy *S3*

The iterative algorithm requires two additional parameters. The initial value for the kinetic constant (k_{init}), and the tolerance for the relative change in k that halts the iterative procedure (ε). k_{init} should always be taken larger than k_{true}. In the simulation $\varepsilon =$ 1%, meaning that if the relative change in k is lower than 1% of the previous k value the iterations stop.

4. Conclusions

We have demonstrated here that the combination of high-throughput experimental set-ups and fast on-line analysis can in principle be applied to studying chemical reaction kinetics in parallel. To attain this goal, new approaches to chemical kinetics are required, and just like some HTE techniques (e.g. split/pool methods), these can seem "contradictory to common sense" at first glance. As shown here, iterative sampling algorithms could play a key role in the development of intelligent laboratory systems capable

[13] A detailed description of the simulations and the original MATLAB source code files are available on the Internet at http://www.science.uva.nl/~gadi.

of high-throughput on-line analysis. The application of these methods to real experimental libraries is currently under study in our laboratory.

References

[1] (a) S. Senkan, Angew. Chem. Int. Ed. 40 (2001) 312–329;
(b) A. Holzwarth, P. Denton, H. Zanthoff, C. Mirodatos, Catal. Today 67 (2001) 309–318;
(c) B. Jandeleit, H.W. Turner, T. Uno, J.A.M. Van Beek, W.H. Weinberg, Cattech 2 (1998) 101–123;
(d) T. Bein, Angew. Chem. Int. Ed. 38 (1999) 323–326;
(e) F. Schüth, C. Hoffmann, A. Wolf, S. Schunk, W. Stichert, A. Brenner, Combinatorial Chem. (1999) 463–477;
(f) H.B.J. Kagan, Organomet. Chem. 567 (1998) 3–6;
(g) M.T. Reetz, Angew. Chem. Int. Ed. 40 (2001) 284–310.

[2] (a) P. Claus, D. Honicke, T. Zech, Catal. Today 67 (2001) 319–339;
(b) C. de Bellefon, N. Tanchoux, S. Caravieilhes, P. Grenouillet, V. Hessel, Angew. Chem. Int. Ed. 39 (2000) 3442–3445;
(c) C. de Bellefon, N. Tanchoux, S.J. Caravieilhes, J. Organomet. Chem. 567 (1998) 143–150;
(d) M.T. Reetz, K.M. Kuhling, S. Wilensek, H. Husmann, U.W. Hausig, M. Hermes, Catal. Today 67 (2001) 389–396.

[3] (a) M.W. Lutz, J.A. Menius, T.D. Choi, R.G. Laskody, P.L. Domanico, A.S. Goetz, D.L. Saussy, Drug Discovery Today 1 (1996) 277–286;
(b) A. Tuchbreiter, J. Marquardt, J. Zimmermann, P. Walter, R. Muelhaupt, B. Kappler, D. Faller, T. Roths, J. Honerkamp, J. Comb. Chem. 3 (2001) 598–603;
(c) C.M. Snively, G. Oskarsdottir, J.A. Lauterbach, Catal. Today 67 (2001) 357–368.

[4] (a) F.W.J. van den Berg, H.F.M. Boelens, H.C.J. Hoefsloot, A.K. Smilde, AIChE J. 47 (2001) 2503–2514;
(b) A.K. Smilde, H.C.J. Hoefsloot, F.W.J. van den Berg, Anal. Chem. 74 (2002) 3105–3111.

[5] D.L. Massart, B.G.M. Vandeginste, L.M.C. Buydens, S. de Jong, J. Smeyers-Verbeke, Handbook of Chemometrics and Qualimetrics: Part A, Elsevier, Amsterdam, 1997, pp. 393–396.

[6] H. Martens, T. Naes, Multivariate Calibration, Wiley, New York, 1989.

[7] (a) G. Rothenberg, A.P. Downie, C.L. Raston, J.L. Scott, J. Am. Chem. Soc. 123 (2001) 8701–8708;
(b) G. Rothenberg, L. Feldberg, H. Wiener, Y. Sasson, J. Chem. Soc., Perkin Trans. 2 (1998) 2429–2434;
(c) S. Mukhopadhyay, G. Rothenberg, Y. Sasson, Adv. Synth. Catal. 343 (2001) 274–278.