Available online at www.sciencedirect.com
ScienceDirect
Catalysis Today 121 (2007) 269–281
CATALYSIS TODAY
www.elsevier.com/locate/cattod

Noise in temporal analysis of products (TAP) pulse responses

Raf Roelant [a], Denis Constales [b], Gregory S. Yablonsky [c], Roger Van Keer [b], Michael A. Rude [c,1], Guy B. Marin [a,*]

[a] *Laboratorium voor Petrochemische Techniek, Department of Chemical Engineering, Ghent University, Krijgslaan 281 (S5), B-9000 Gent, Belgium*
[b] *Department of Mathematical Analysis, Ghent University, Galglaan 2, B-9000 Gent, Belgium*
[c] *Department of Chemical Engineering, Washington University, Campus Box 1198, One Brookings Drive, St.-Louis, MO 63130-4899, USA*

Available online 25 July 2006

Abstract

Pulse responses recorded by the mass spectrometer of a TAP apparatus are subjected to noise. A general theoretical framework for noise studies and classification has been developed and applied to experimental data obtained at Ghent University (Belgium) and Washington University in St. Louis (USA). Two types of noise have been distinguished: Gaussian noise and spectrally localized noise. While the first one is believed to be essential, the second one is due to mains frequency interference. These types of noise are clearly distinguished from uncertainty caused by a variable pulse intensity. Analytical expressions are presented, and a regression is found to enable the experimentalist to analyze the noise quantitatively. This can be used as troubleshooting tool for the functioning of the TAP-setup. Increasing the sampling frequency can improve the precision by which physico-chemical parameters are estimated by regression of the pulse responses. When Gaussian noise is dominant, its correlation time determines up to which frequency this improvement is possible.

Keywords: Temporal analysis of products; Noise; Stochastic processes; Errors; Statistics

1. Introduction

Kinetic dependences obtained in non-steady-state chemical experiments are liable to noises. TAP studies [1,2] provide an opportunity to perform a thorough stochastic and statistical analysis, particularly a noise analysis based on large amounts of data that can be collected in a short amount of time. In fact, such an opportunity is unique in chemical kinetics. Multi-pulse experiments are performed using anywhere from 1 to 30,000 pulses, each pulse being separated by 1–30 s. The signal recorded by the mass-spectrometer may be looked upon as a superposition of the 'pure' pulse response related to the studied processes and noise. The noise may be a superposition of multiple noise types.

Mostly, the noise constrains the experimentalist to average multiple pulse responses originating from subsequently performed experiments. Typically, 20–50 pulse responses are averaged to a single pulse response with an acceptable (optimal) *signal to noise ratio*, SNR.

The goal of this paper is to present the theoretical framework for primary noise analysis of TAP-data and illustrate it by some useful examples. The noise analysis will significantly improve quantitative information on the precision of diffusion/kinetic parameters obtained in TAP-experiments.

Noise actually limits the experimental conditions at which meaningful conclusions can be drawn. Indeed, whenever these conditions cause noise to become dominant, any conclusions would be illegitimate. This paper should present a first step in the discovery of the experimental latitude. Especially, this experimental latitude imposes a window of measurable rate coefficients. Depending on the experimental setup used, this may further limit the window previously derived from purely physico-chemical considerations [3].

In the future, the theoretical framework presented in this paper could be a part of the systematic analysis for distinguishing

* Corresponding author.
E-mail address: Guy.Marin@UGent.be (G.B. Marin).
[1] Present address: ExxonMobil Research and Engineering, 3225 Gallows Road, 7A0405, Fairfax, VA 22037, USA.

0920-5861/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.cattod.2006.05.087

Nomenclature

Roman symbols

$C_Y(p_j, p_k)$	autocovariance: covariance of stochastic variables $Y(p_j)$ and $Y(p_k)$ of a random sequence: Covar $[Y(p_j), Y(p_k)]$
Covar $[A, B]$	covariance of stochastic variables A and B: Covar $[A, B] = E[(A - E[A])\cdot\overline{(B - E[B])}]$
$E[A]$	expected value of stochastic variable A
$(\mathcal{F}Y(\omega_j))$	discrete Fourier transform of a random sequence $(Y(t_j))$
i	imaginary unit: $i^2 = -1$
M	number of pulse responses in a set
N	number of samples per pulse response
t_j	time between the pulse and the measurement of sample j (s)
T	total collection time of a pulse response: $T = N\,\Delta t$ (s)
$x_k(t_j)$	sample j of the kth pulse response of a set (V)
$(X(t_j))$	stochastic process, model for a sampled pulse response (V)

Greek symbols

Δt	sampling interval (sampling period): time between two consecutive samplings by the mass spectrometer signal (s)
Δz	reactor bed length (m)
δ_{xy}	Kronecker delta: $\delta_{xy} \equiv \begin{cases} 1 & \text{when } x = y \\ 0 & \text{when } x \neq y \end{cases}$
ε_b	reactor bed porosity
$\mu_Y(p_j)$	mean (expected value) of stochastic variable $Y(p_j)$ of a random sequence: $\mu_Y(p_j) = E[Y(p_j)]$
ν_j	frequency: $\nu_j = j/N\,\Delta t$ (Hz)
θ	correlation time (s)
$\rho_Y(p_j, p_k)$	coefficient of correlation of stochastic variables $Y(p_j)$ and $Y(p_k)$ of a random sequence: $\rho_Y(p_j, p_k) = C_Y(p_j, p_k)/(\sigma_Y(p_j)\sigma_Y(p_k))$
$(\sigma^2)_M$	time average variance of $X(t_j)$ (V^2)
$\sigma_Y^2(p_j)$	variance of stochastic variable $Y(p_j)$ of a random sequence: $\sigma_Y^2(p_j) = C_Y(p_j, p_j)$
ω_j	pulsation (angular frequency): $\omega_j = 2\pi\nu j$ (Hz)

Mathematical symbols

$a \in A$	a is an element of set A
$\{a, b, \ldots\}$	set containing the elements a, b, ...
$]a, b[$	set of real numbers x, with $a < x < b$
$\overline{a + ib}$	complex conjugate of $a + ib$: $a - ib$ (a and b real)
$\lvert a + ib\rvert$	modulus of the complex number $a + ib$ (a and b real): $\sqrt{a^2 + b^2}$ if $b = 0$, $\lvert a\rvert$ a is the absolute value of the real variable a

noises caused by different physical and physico-chemical reasons, particularly for distinguishing process noises and noises of the measuring device. Also based on the results of this noise analysis, the concrete noises related to different subsystems of the device (valve, mass-spectrometer, etc.) have to be identified. The noise analysis will play a special role in performing the so-called Y-procedure proposed in [4] for extracting model-free reaction rate data from the results of TAP-experiments. The Y-procedure applied to data obtained on a thin-zone TAP-reactor (TZTR) [5] involves Fourier analysis. This necessitates a transfer from the time to the frequency domain by Fourier Transform. The noise, and its manifestation in the frequency domain particularly, has to be taken into account while performing the Y-procedure and interpreting its results.

Experimental data obtained at the TAP-1 setup at Ghent University and the TAP-2 setup at Washington University are the object of the present study.

2. Stochastic processes

The description of noise is best done in terms of the theory of stochastic processes. Consider a sequence $(X(p_0), X(p_1), \ldots, X(p_{P-1}))$ of random variables, where $p_0 < p_1 < \ldots < p_{P-1}$ are values adopted by a certain parameter p. Shortly, the sequence can be written as $(X(p_j))$. Such a sequence is called a *random sequence* [6,7] or a *discrete-parameter stochastic process* [8,9]. Consider the most general case where the random variables are complex.

Each *mean* $\mu_X(p_j)$ is the expected value of the sample at $p = p_j$:

$$\mu_X(p_j) = E[X(p_j)]. \tag{1}$$

$(\mu_X(p_j))$ may thus be called the *mean sequence*.

The *autocovariance* $C_X(p_j, p_k)$ is the covariance of the corresponding samples $X(p_j)$ and $X(p_k)$:

$$C_X(p_j, p_k) = \mathrm{Covar}[X(p_j), X(p_k)] \tag{2}$$

$$C_X(p_j, p_k) = E[\overline{(X(p_j) - \mu_X(p_j))} \cdot (X(p_k) - \mu_X(p_k))] \tag{3}$$

$$C_X(p_j, p_k) = E[\overline{X(p_j)} \cdot X(p_k)] - \overline{\mu_X(p_j)} \cdot \mu_X(p_k). \tag{4}$$

Note that $C_X(p_j, p_k)$ and $C_X(p_k, p_j)$ form a pair of complex conjugates.

A random sequence is called *white noise* if different random variables in the sequence are uncorrelated:

$$j \neq k \Rightarrow C_X(p_j, p_k) = 0. \tag{5}$$

Random sequences that are not white noise are called *colored noise*.

The *standard deviation* $\sigma_X(p_j)$ of sample j is given by

$$\sigma_X(p_j) = \sqrt{C_X(p_j, p_j)}. \tag{6}$$

If the standard deviation of all its elements is the same, the random sequence is called *homoskedastic*. If not, it is called *heteroskedastic*.

Finally, the dimensionless *correlation coefficient* $\rho_X(p_j, p_k)$ of the jth and pth sample, is obtained as

$$\rho_X(p_j, p_k) = \frac{C_X(p_j, p_k)}{\sigma_X(p_j)\sigma_X(p_k)}. \tag{7}$$

3. Procedure

3.1. Estimation of stochastic properties

Consider a recorded pulse response subjected merely to baseline correction. Taken from an electric signal produced by a mass spectrometer, they therefore have the dimension of an electric charge. As it is liable to stochastic noise, a pulse response may be modeled as a random sequence $(X(t_0), X(t_1), \ldots, X(t_{N-1}))$, where the parameter t is the time progressed since the feed gas was pulsed. Traditionally, the samples are taken at equidistant times:

$$t_j = j\,\Delta t, \tag{8}$$

where the time Δt between two consecutive samplings is called the *sampling interval*.

The mean or expected pulse response is obtained by estimating

$$\mu_X(t_j) = \frac{1}{M}\sum_{k=1}^{M} x_k(t_j), \tag{9}$$

where $x_k(t_j)$ is the sample taken at time t_j in the kth experiment (pulse response). The autocovariance is estimated by

$$C_X(t_l, t_m) = \frac{1}{M-1}\sum_{k=1}^{M}\Big(x_k(t_l) - \mu_X(t_l)\Big)\cdot\Big(x_k(t_m) - \mu_X(t_m)\Big) \tag{10}$$

This estimation is also used for the standard deviation profile ($\sigma_X(t_j)$), using Eq. (6).

Each pulse response is subjected to the discrete Fourier transform. The discrete Fourier transform of the random sequence ($X(t_j)$) is a random sequence itself:

$$(\mathcal{F}X(\omega_0), \mathcal{F}X(\omega_1), \ldots, \mathcal{F}X(\omega_{N-1})),$$

where by definition,[1]

$$\mathcal{F}X(\omega_j) = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1} X(t_k)\,\mathrm{e}^{-i\omega_j t_k} \tag{11}$$

in which

$$\omega_j = \frac{2\pi j}{N\,\Delta t}. \tag{12}$$

is a pulsation, corresponding to a frequency

$$\nu_j = \frac{\omega_j}{2\pi} = \frac{j}{N\,\Delta t}. \tag{13}$$

The discrete Fourier transform of a sequence may also simply be called its *spectrum*.

As all $X(t_j)$ are real stochastic variables, definition (11) implies that

$$\mathcal{F}X(\omega_{N-j}) = \overline{\mathcal{F}X(\omega_j)}. \tag{14}$$

For this reason, beyond the so-called *Nyquist pulsation* $\omega_{N/2} = \pi/\Delta t$, no extra information can be extracted from the discrete Fourier transform.[2] The Nyquist pulsation is the highest pulsation at which the discrete Fourier transform has physical meaning.

Estimations similar to (9) and (10) are performed for $\mu_{\mathcal{F}X}(\omega_j)$, $C_{\mathcal{F}X}(\omega_j, \omega_k)$ and $\sigma_{\mathcal{F}X}(\omega_j)$ in the frequency domain.

The typical statistical point of view is that the higher the number M of repeat experiments, the better all estimations become. However, in reality this concept has to be tested experimentally because of presence of additional factors/processes. During the adsorption/reaction process, which occurs in multi-pulse experiments, the surface coverage and, therefore, catalyst state change gradually. Such changes impose an upper limit to the number M of experiments needed for characterization of the chosen catalyst state (state-defining characterization). Even data of simple diffusion processes with no interaction between the gas substance and material surface may be shadowed by trends of some experimental characteristics (pulse intensity, temperature, etc.).

3.2. Experimental data

As a basis of the current statistical study, a proper dataset should result from a lot of experiments where the catalyst is left net unaffected by each pulse to avoid the pulse responses to change shape. This is the case if the feed gas is inert, if the catalyst is in fact an inert material or if in another way the net adsorption of the feed gas on the catalyst is zero during each experiment.

The data set from Ghent University contains 600 pulse responses from a TAP-1 setup. An estimated average number of 1.2×10^{15} oxygen molecules were pulsed over a three-zone TAP reactor at 500 °C. In central position was a bed of Ti_HV5_H catalyst [10], vanadia based and obtained by DC magnetron sputtering on a ZrO_2/SiO_2 support. This zone was 21.78 mm long and sandwished between two beds of quartz beads, the first one of which 1.34 mm long and the second one 5.3 mm long. Both the quartz beads and the catalyst support particles had diameters of 250–425 μm. The catalyst was not further oxidized by the oxygen pulses. Therefore, the oxygen responses are uniform and fit for the statistical analysis. Each pulse response is stored as a sequence of 1000 samples taken every 0.4 ms. The collection time per pulse thus adds up to 0.4 s.

[1] Alternatively, the discrete Fourier transform is often defined as $\mathcal{F}X(\omega_j) = \sum_{k=0}^{N-1} X(t_k)\,\mathrm{e}^{-i\omega_j t_k}$. However, definition (11) provides the maximum degree of symmetry between the time and the frequency domain.

[2] Corresponding Nyquist frequency: $\nu_{N/2} = \frac{1}{2\Delta t}$.

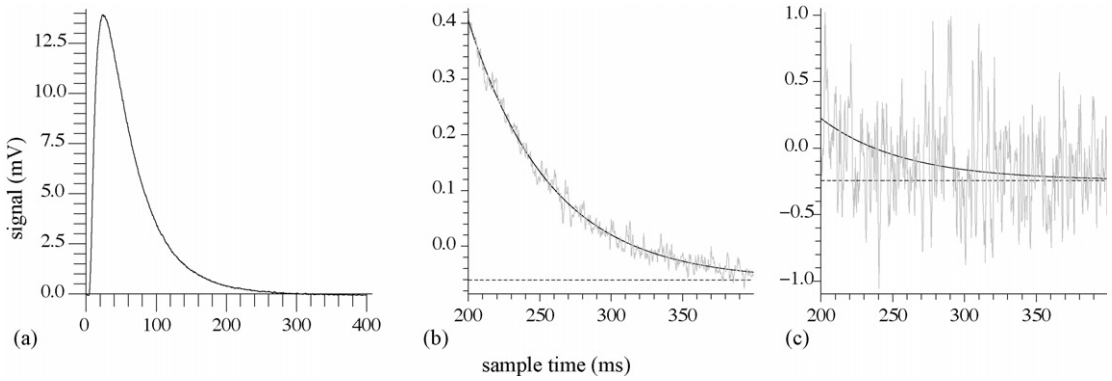

Fig. 1. (a) Average of non-baseline-corrected oxygen pulse responses obtained at Ghent University. These were obtained on a TAP-1 system loaded with a vanadia-based catalyst [10] sandwiched between two zones packed with quartz beads of the same diameter as the catalyst particles. The bed was at 500 °C. The catalyst is not further oxidized by the pulses. (b) The grey curve represents the second half of (a). Clearly, the signal is not yet completely extinct at the end of the pulse recording. For this reason, the location of the baseline cannot, as is customary, be estimated as the average of the last couple of hundreds of samples. Rather, the grey curve is non-linearly regressed with an exponential profile $a\,e^{-bt} + c$. The result is shown as the black full line. The black dashed line represents its horizontal asymptote. The location of the baseline of all separate pulse responses is now estimated as the average of the last 500 samples (from 200 to 400 ms) less $a\,e^{-bt}$. (c) shows the last 500 samples of an example pulse response in grey. The black dashed line shows the estimated location of the baseline for this particular pulse response. The full black line represents $a\,e^{-bt}$ superposed on this baseline.

An extra time gap of 0.5 s was left between the end of the collection of each pulse response and the next pulse.

Fig. 1(a) shows the average non-baseline-corrected oxygen pulse response. Fig. 1(b) is a magnification of the second half of this response, close to extinction. It is apparent that the signal is not yet completely extinct by the time the response is stopped being recorded. Therefore, the location of the baseline cannot, as is customary, be estimated as the average of the last couple of hundreds of samples. A more refined approach consists of regressing the second half of the average pulse response with an expression $a\,e^{-bt} + c$. Subsequently, amounts $a\,e^{-bt}$ are subtracted from the last 500 samples of all pulse responses prior to their being averaged to yield the estimated baseline position. Thus, all pulse responses are separately baseline corrected. As an extra time gap of 0.5 s was left between any two subsequent pulses, it can be assumed that the former response is extinct by the time the recording of the latter is started.

The data set from Washington University contains 1000 pulse responses from a TAP-2 setup. Here, an estimated number of 2.8×10^{14} krypton atoms were pulsed over a 31.8 mm long bed of quartz particles with a diameter of 210–250 μm. The temperature of the bed was 50 °C. The collection time was 4 s and 1000 samples are taken per pulse response, making the sampling interval 4 ms. An extra time gap of about 0.1 s was applied between the collection of subsequent pulses.

To perform the baseline correction of all pulse responses, the method used for the data from Ghent University was also applied here. As was the case there, the last 500 samples out of 1000 were used for the estimation of the baseline location. This time only an extra 0.1 s is left between the end of a pulse recording and the next pulse. Therefore, some krypton atoms of the former pulse can be expected still to be present in the bed when the latter is given. However, their number can be assumed to have a negligible effect.

From this point on, the data sets from Ghent and Washington University will be referred to as the UGent set and the WUStL set, respectively. They contain the baseline-corrected pulse responses.

Fig. 2 depicts semilogarithmic graphs of the average baseline-corrected pulse responses of the UGent set and the

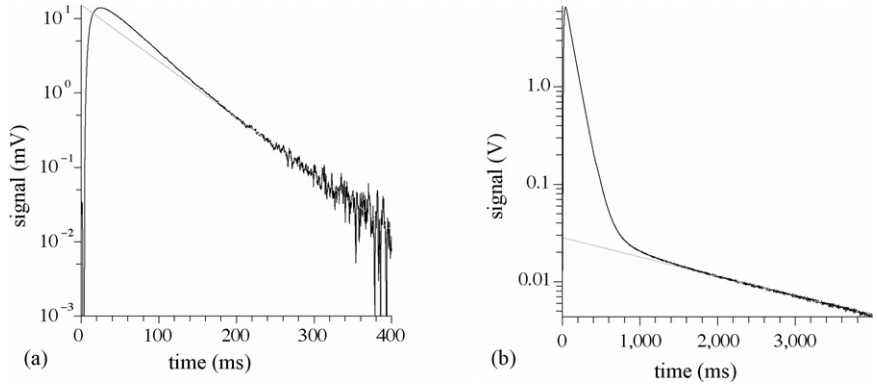

Fig. 2. (a) Semilogarithmic graph (in black) of the average of 600 baseline-corrected oxygen pulse responses obtained at Ghent University, over a three-zone TAP-reactor with central vanadia-based catalytic zone. (b) Semilogarithmic graph (in black) of the average of 1000 baseline-corrected krypton pulse responses obtained at Washington University, over a single quartz zone. The grey lines in (a) and (b) represent the final exponential decay. Both pulse responses show a convex tail, which indicates that some molecules have interacted reversibly with the TAP reactor bed.

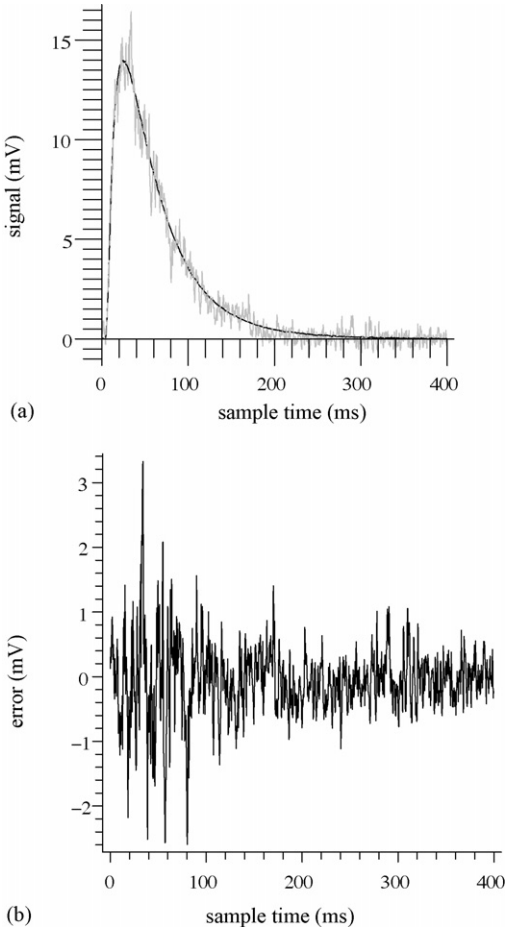

Fig. 3. (a) An oxygen pulse response (in grey) and the average of 600 such pulses from the UGent data set (in black). (b) The example response's deviation from average.

WUStL set. Fig. 2(b) for the WUStL set clearly shows how the tail of the pulse response is the superposition of two logarithmic decays. The first is due to the simple diffusive release of krypton atoms that have not interacted with the surface of the quartz bed. The second is due to the slow release of krypton atoms by the surface. Apparently, some krypton interacts reversibly with the quartz. The corresponding exponential

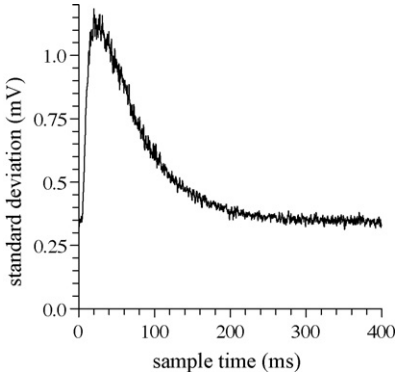

Fig. 5. Standard deviation of all samples in a pulse response from the UGent set vs. their sample time.

decay $a\,e^{-bt}$ is the one found in preparation for the baseline correction. It is represented by the grey line. Albeit less clearly, Fig. 2(a) shows similar behaviour of the oxygen molecules constituting the pulse responses from the UGent set. This is less surprising as oxygen can indeed be expected to undergo reversible adsorption on the vanadia-based catalyst.

4. Spectrum of TAP-noise

4.1. Gaussian noise

4.1.1. Normal distribution and standard deviation

Fig. 3 depicts the mean pulse response ($\mu_X(t_j)$), estimated from all pulse responses of the UGent set as in Eq. (9). It also shows an example pulse response and its deviation from the mean.

Traditionally, the deviations from the mean would be assumed normally distributed around zero. For the UGent data set, this assumption is verified in Fig. 4. It shows histograms of the deviations from the average of samples taken before, near and after the peak of the pulse response, believed to be representative for the whole pulse response. These are compared to normal (Gaussian) probability density curves with mean zero and standard deviation estimated from the

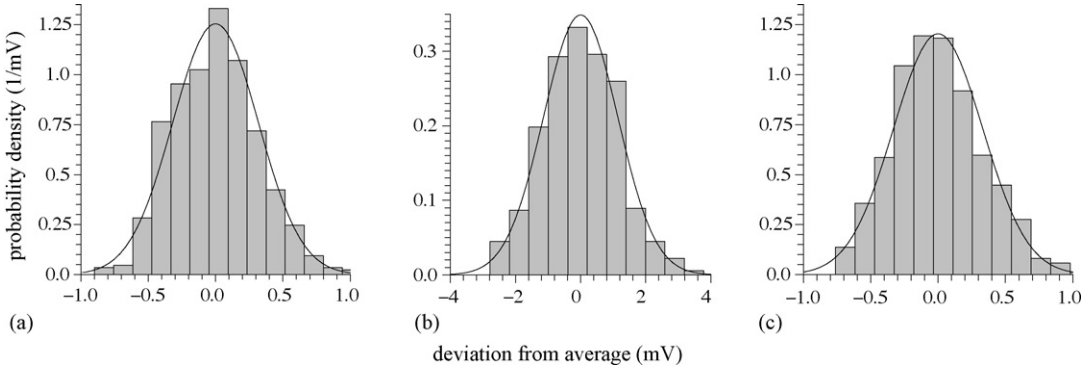

Fig. 4. Histograms of the deviation from average of samples in the pulse responses from the UGent data set. (a) An early sample taken at 0.4 ms after the pulse. (b) A sample taken at 26 ms, about the peak of the pulse response. (c) A late sample taken at 384 ms. Comparison to normal probability density functions (black lines) around 0 mV and with standard deviations estimated from the data shows that all deviations from average can be considered normally distributed in good approximation.

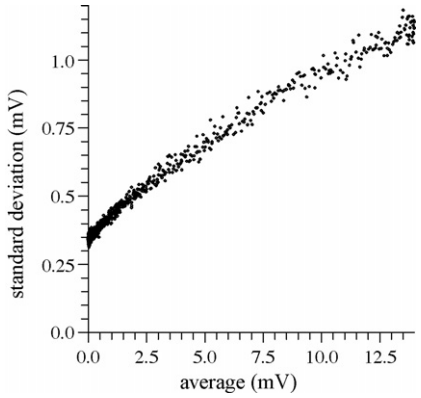

Fig. 6. Standard deviation of the samples as a function of their mean for the UGent set, resulting from elimination of the sample times. The graph shows that the standard deviation is an increasing function of the signal strength. An analytic form of this function could be derived from appropriate interpolation.

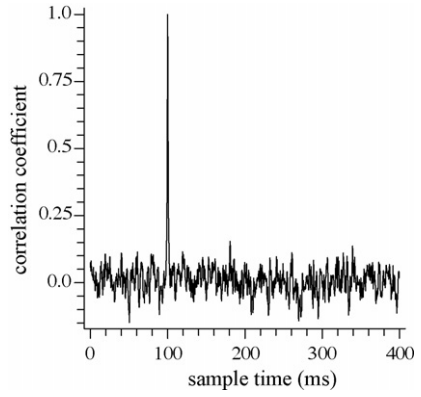

Fig. 7. Correlation coefficient between the sample taken at 100 ms after the pulse and all samples from the UGent set. The graph shows a steep peak with height 1 around 100 ms. The rest of the profile is noisy with average 0.

experimental data. Because of the apparent resemblance of the histograms and the gaussian curves, all samples $X(t_j)$ along the pulse response are considered normally distributed around the mean $\mu_X(t_j)$. The Gaussian deviation from this mean is symbolized as $\Phi(t_j)$.

The estimated standard deviation profile ($\sigma_X(t_j)$) is shown in Fig. 5. The pulse response is heteroskedastic: a part of the standard deviation is proportional to the strength of the signal, while another part is constant. This becomes clearer if the standard deviation σ_X is plotted versus the mean μ_X, eliminating time. The result is shown in Fig. 6. Clearly, an appropriate interpolation allows to predict the standard deviation from the strength of the signal.

4.1.2. Colored noise: Ornstein–Uhlenbeck model

Fig. 7 shows the profile of the correlation coefficient between a base sample taken at 100 ms, in the tail of the curve, and all samples. This figure shows a single steep peak at 100 ms amid noise about the zero of correlation. It is trivial that the correlation coefficient between the sample taken at 100 ms and itself is 1. However, zooming in on the peak would yield a correlation peak such as these shown in Fig. 8, which reveal that the noise is colored, i.e. the correlation coefficient does not drop to zero immediately around the base sample. Specifically, these graphs, showing correlation profiles with base samples chosen at the beginning, the peak and the end of the peak response, suggest an exponential correlation relaxation towards past and future. This, as well as the normality, is a typical feature of an *Ornstein–Uhlenbeck stochastic process* (also called a *Gauss–Markov stochastic process*):

$$\rho_\Phi(t_j, t_k) = \mathrm{e}^{-|t_k - t_j|/\theta}, \tag{15}$$

where θ is the correlation time (see Doob's theorem [11]).

The Ornstein–Uhlenbeck process was originally presented as a model for the velocity of a particle in Brownian motion [12,13]. Here, it is a suitable model for the noise divided by its standard deviation: $(\Phi(t_j)/\sigma_\Phi(t_j))$.

Even though the TAP reactor operates under high vacuum, enough molecules are collected by the mass spectrometer to consider the non-sampled electronic signal a continuous one. Hence, to ensure this continuity, correlation between very rapidly succeeding samples must approach 1 as the sampling

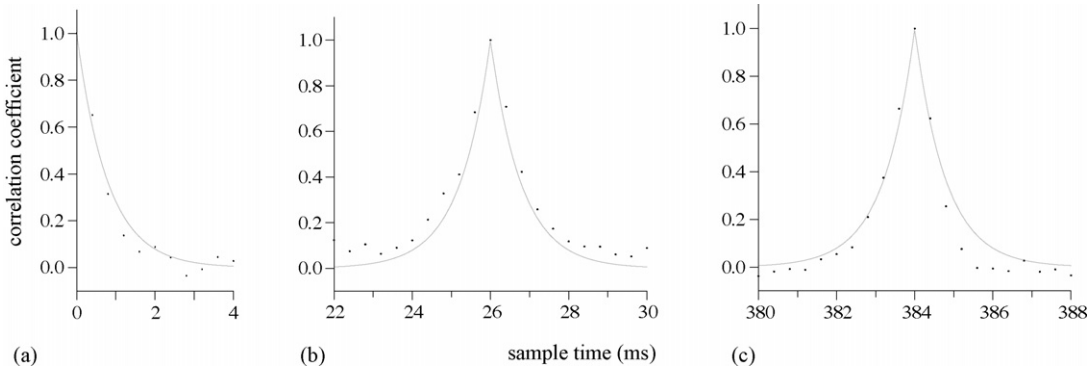

Fig. 8. Fragments of correlation profiles derived for the UGent data set. (a) Coefficient of correlation between the first sample (taken as the pulse was given) and neighbouring samples (black dots). (b) Correlation coefficient between the sample taken at 26 ms after the pulse and neighbouring samples. (black dots). (c) Correlation coefficient between the sample taken at 384 ms after the pulse and neighbouring samples (black dots). These graphs suggest an exponential decrease of the correlation coefficient towards past and future. The grey lines in (a–c) represent such an exponential relation with correlation time θ estimated from regression of the standard deviation profile in the frequency domain (see Fig. 10) with Eq. (20). The correspondence is satisfactory.

interval goes to zero. Therefore, there must always exist a certain correlation time θ, that can only be neglected if Δt has been chosen well over θ, in which case a white noise Gaussian model is satisfactory. In the other case however, as will be clear from the following, a colored noise model, and more precisely an Ornstein–Uhlenbeck model is in order.

4.1.3. The time autocorrelation effect

Call $(\sigma^2)_M$ the time average variance of the pulse response. Then

$$N \cdot (\sigma^2)_M = \sum_{j=0}^{N-1} \sigma_X^2(t_j) = \sum_{k=0}^{N-1} \sigma_{\mathcal{F}X}^2(\omega_k), \tag{16}$$

where the second equality is a consequence of *Parseval's theorem.* In words, the variance in the time domain is *redistributed* in the frequency domain. As a consequence $(\sigma^2)_M$ is the frequency average as well as the time average variance.

It can be proved that the autocovariance of the discrete Fourier transform of the Gaussian noise is approximated very well by

$$C_{\mathcal{F}\Phi}(\omega_p, \omega_q) = \frac{1}{\sqrt{N}} \frac{(1-\rho^2)\mathcal{F}\sigma^2(\omega_p - \omega_q)}{1 - 2\rho\cos(((\omega_p + \omega_q)/2)\Delta t) + \rho^2} \tag{17}$$

where it is assumed that $\omega_p \geq \omega_q$.[3] ρ is the coefficient of correlation of two consecutive samples. Δt is the sampling interval. The expression involves the discrete Fourier transform $(\mathcal{F}\sigma^2(\omega_j))$ of the variance profile in the time domain. Respecting (15), ρ is related to θ as

$$\rho = \mathrm{e}^{-\Delta t/\theta}. \tag{18}$$

A special case of (17) is given by

$$\sigma_{\mathcal{F}\Phi}^2(\omega_p) = \frac{(1-\rho^2)(\sigma^2)_M}{1 - 2\rho\cos(\omega_p \Delta t) + \rho^2}. \tag{19}$$

It can be proved that expression (19) obeys Eq. (16) following from Parseval's theorem. Following directly from (19),

$$\sigma_{\mathcal{F}\Phi}(\omega_p) = \sqrt{\frac{1-\rho^2}{1 - 2\rho\cos(\omega_p \Delta t) + \rho^2}} (\sigma^2)_M^{1/2}. \tag{20}$$

Observe that (20) implies that the heteroskedasticity of the pulse response (as apparent in Fig. 5) is not reflected in the standard deviation pattern in the frequency domain, but only in the covariance pattern via $\mathcal{F}\sigma^2$ in (17).

Limit values of (19) are given by

$$\sigma_{\mathcal{F}\Phi}^2(0\,\mathrm{Hz}) = \frac{1+\rho}{1-\rho}(\sigma^2)_M = \coth\left(\frac{\Delta t}{2\theta}\right)(\sigma^2)_M \tag{21}$$

$$\sigma_{\mathcal{F}\Phi}^2\left(\frac{\pi}{\Delta t}\right) = \frac{1-\rho}{1+\rho}(\sigma^2)_M = \tanh\left(\frac{\Delta t}{2\theta}\right)(\sigma^2)_M, \tag{22}$$

[3] In the opposite case $\omega_p < \omega_q$, $C_{\mathcal{F}\Phi}(\omega_p, \omega_q)$ is found with $\mathcal{F}\sigma^2(\omega_p - \omega_q)$ replaced by $\overline{\mathcal{F}\sigma^2(\omega_q - \omega_p)}$ in (17).

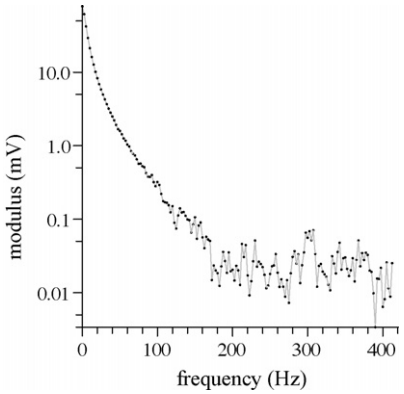

Fig. 9. Semilogarithmic graph of the modulus of the average pulse response's discrete Fourier transform, for the UGent set. This graph shows that no meaningful information can be extracted from the spectrum beyond 200 Hz, where noise becomes dominant.

where $\pi/\Delta t$ is the Nyquist pulsation. The second equalities in (21) and (22) are achieved by substituting (18).

Correlation ρ between subsequent samples taken by the mass spectrometer causes an increase of the variance of the discrete Fourier transform at low frequencies and a decrease at high frequencies. This contrast becomes more pronounced if ρ increases. For example, as it may be derived from (21), in case the sampling interval Δt is half the correlation time θ, the standard deviation at 0 Hz is doubled by the correlation effect. This effect can become very important when sampling at high frequencies is required by e.g. fast diffusion of the pulse through the reactor bed.

Consider the profile of the modulus of the average signal in the frequency domain, as depicted in Fig. 9. Clearly, the relevant information contained in the discrete Fourier transform is concentrated in the frequency domain from 0 to 200 Hz, while the Nyquist frequency is 1250 Hz. Generally spoken, at reasonably low sampling intervals Δt, the information contained in the spectrum of a recorded TAP-signal is restricted to the low frequency region. This is precisely the region where autocorrelation causes an increase of the standard deviation. It is concluded that this autocorrelation causes an unfavourable increase of the level of uncertainty. As was already stated, this effect becomes important when sampling is carried out at high frequencies.

It can be derived from Eq. (21) that if the sampling interval is higher than 2.5 times the correlation time θ, the low frequency standard deviation stays within 10% of $(\sigma^2)_M^{1/2}$. In this case, the unfavourable time autocorrelation effect can be neglected. Increasing the number of samples N by increasing the sampling frequency $1/\Delta t$ while keeping the total collection time $T = N\,\Delta t$ constant essentially does not change the mean variance $(\sigma^2)_M$ in the time domain. As a consequence of Parseval's theorem (see Eq. (16)) $(\sigma^2)_M$ is also equal to the mean variance in the frequency domain and also, as there is assumed not to be a time autocorrelation effect, the mean variance at 0 Hz:

$$\sigma_{\mathcal{F}\Phi}^2(0\,\mathrm{Hz}) \approx (\sigma^2)_M. \tag{23}$$

On the other hand, consider the limit case in which the sampling interval is much smaller than the correlation time. In that case, Eq. (21) can be approximated by

$$\sigma^2_{\mathcal{F}\Phi}(0\,\mathrm{Hz}) \approx \frac{2\theta}{\Delta t}(\sigma^2)_M = N \cdot \frac{2\theta}{T}(\sigma^2)_M. \tag{24}$$

As the interesting frequency region is around 0 Hz, the relative error of $\mathcal{F}X(0\,\mathrm{Hz})$ gives a good idea of the quality of the estimation of any physico-chemical parameters carried out by regression of the mean pulse response with suitable expressions. Firstly, it follows from definition (11) that

$$\mu_{\mathcal{F}X}(0\,\mathrm{Hz}) = \frac{1}{\sqrt{N}}\sum_{k=0}^{N-1}\mu_X(t_k) \tag{25}$$

$$\mu_{\mathcal{F}X}(0\,\mathrm{Hz}) \approx \frac{A_0}{\sqrt{N}\,\Delta t} = \sqrt{N}\cdot\frac{A_0}{T}, \tag{26}$$

where A_0 is the area under the ideal continuous pulse response. In the low sampling frequency limit, the relative error of $\mathcal{F}X$ (0 Hz) is found from Eqs. (23) and (26):

$$\frac{\sigma_{\mathcal{F}\Phi}(0\ \mathrm{Hz})}{\mu_{\mathcal{F}X}(0\ \mathrm{Hz})} \approx \frac{1}{\sqrt{N}}\cdot\frac{T}{A_0}\sqrt{(\sigma^2)_M}. \tag{27}$$

The relative error is inversely proportional to the square root of the number of samples N. This shows that as long as the sampling interval Δt is kept well above the correlation time θ, increasing the sampling frequency while keeping the total collection time T constant will improve the precision of parameter estimations by regression. In the high sampling frequency limit however, combination of (24) and (26) yields

$$\frac{\sigma_{\mathcal{F}\Phi}(0\ \mathrm{Hz})}{\mu_{\mathcal{F}X}(0\ \mathrm{Hz})} \approx \frac{\sqrt{2T\theta}}{A_0}\sqrt{(\sigma^2)_M}. \tag{28}$$

The relative error is no longer dependent on the number of samples N. This means that the precision of parameter estimations by regression will hardly improve by increasing the sampling frequency even further.

In summary, increasing the sampling frequency will improve the precision of parameter estimations by regression. However, once the sampling interval Δt has been made smaller than the correlation time θ, this quality will approach its upper limit. The memory required to store the pulse responses is proportional to the sampling frequency. Assume that the sampling frequency is already high enough for the spectrum to cover even the fastest of the observable physico-chemical phenomena. Then increasing the number of recorded pulse responses rather than the number of samples per pulse (keeping T constant) is a better strategy to improve parameter estimations. Indeed, this improvement will be at least as high as if the extra memory needed to do so was used to increase the sampling frequency. Of course, the number of recorded pulse responses is still limited to the point where the change of the catalyst state can no longer be considered insignificant (see Section 3.1).

Suppose an experimentalist plans to record a series of TAP-pulse responses and attaches great importance to the precision of the resulting average pulse response. In this case, he should collect a considerable number of responses taking care, however, to avoid a significant change of the catalyst state. Secondly, he should establish a sampling interval Δt about the correlation time θ or less. There is little point to applying even higher sampling frequencies.

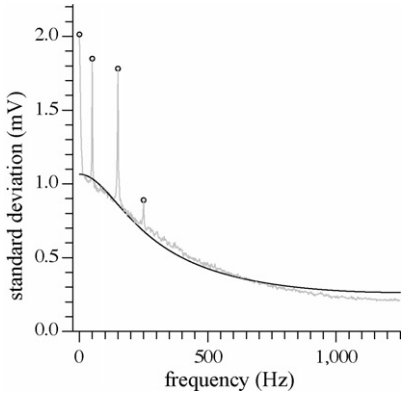

Fig. 10. Estimated standard deviation vs. frequency of the discrete spectrum of the pulse responses from the UGent set (grey curve). Steep peaks about 0, 50, 150 and 250 Hz are marked by a small black circle. Abstracting these peaks, regression with Eq. (20) gives a reasonably good correspondence to the experimental profile. The curve resulting from the regression is shown in black.

4.1.4. Noise characterization

The parameters $(\sigma^2)_M^{1/2}$ and ρ can be estimated from regressing the observed standard deviation pattern in the frequency domain, $(\sigma_{\mathcal{F}X}(\omega_j))$, with the right hand side of (20). Fig. 10 shows this profile. Steep peaks about 0, 50, 150 and 250 Hz attract attention. These are due to other random phenomena which will be discussed in subsequent sections. For now, it suffices to note that these effects can easily be abstracted by deleting the narrow peaks from the profile. Such an abstraction is impossible in the time domain, which is why it would be imprecise to estimate $(\sigma^2)_M$ directly as the average of $C_X(t_j, t_j)$ using (10) for all j from 0 to $N-1$. For the same reason, it would not be advisable to estimate ρ directly from correlation graphs such as these in Fig. 8.

Non-linear regression of the experimental sequence $(\sigma_{\mathcal{F}X}(\omega_j))$ with Eq. (20) is carried out. The resulting curve is shown in Fig. 10 and the corresponding parameter estimations are $\rho = 0.603 \pm 0.006$ and $(\sigma^2)_M^{1/2} = 0.531 \pm 0.003$ mV (95% confidence intervals).

The correlation time θ can be derived from ρ using Eq. (18). This yields $\theta = 1.97\cdot\Delta t\,(= 0.789$ ms). This value can be verified by plotting the theoretical prediction (15) of the correlation profile $(\rho_X(t_j))$ together with its experimental estimation, which has been done in Fig. 8. As is clear from the foregoing, the deviations are largely due to the presence of the non-Gaussian random processes that will be discussed in the subsequent sections.

4.2. Spectrally localized noise: mains frequency interference

Consider the estimated standard deviation profile in the frequency domain of the WUStL set, shown in Fig. 11. A large

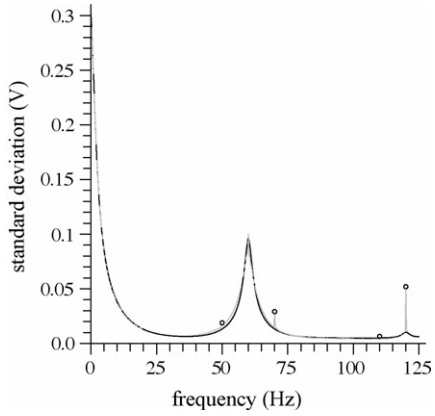

Fig. 11. Estimated standard deviation vs. frequency of the discrete spectrum of the pulse responses from the WUStL set (grey curve). Small black circles mark steep peaks at 50, 70, 110 and 120 Hz. The black curve shows the result of regression with Eq. (47), after abstraction of the marked peaks. The result is satisfactory, although the regression of the peak about 60 Hz is not perfect. This suggests that a simple proportional relation is not the perfect model for the interdependence of the amplitude of the 60 Hz noise and the signal strength.

peak at 60 Hz and a smaller one at 120 Hz attract the attention. Similar peaks, at 50, 150 and 250 Hz, were present in the standard deviation profile of the UGent set (see Fig. 10). The mains frequency in America, where the WUStL data set was collected, is 60 Hz. The mains frequency in Europe, where the UGent data set was collected, is 50 Hz. Apparently, some noise present is spectrally localized about the harmonics of the frequency of the electric energy transmission.[4] This must be due to an interfering effect exerted by electric devices near the mass spectrometer, or an undesired direct influence of the mass spectrometer's ac power supply. Small peaks about 50, 70 and 110 Hz are also present in the standard deviation profile of the WUStL set in the frequency domain. This shows that noise can also be spectrally localized at frequencies that are no harmonics of the mains frequency. However, interfering effects exerted by electric devices present near the mass spectrometer can still be believed to underly this noise. These devices may be part of the TAP apparatus itself.

As a model for the transient process, consider that a pulse response ($X(t_j)$) is the superposition of the ideal pulse response ($\mu_X(t_j)$) and a stochastic process ($\Gamma(t_j)$) respresenting an oscillating noise with fixed pulsation $\omega_p = 2\pi p/N\,\Delta t \in]0, \pi/\Delta t$ [but amplitude ($A(t_j)$) variable along the pulse response, as ($\sigma_\Phi(t_j)$) was variable in the Gaussian noise.

$$\Gamma(t_j) = A(t_j)\cos(\omega_p t_j + \Theta_m), \tag{29}$$

where Θ_m is a random variable representing the random initial phase of the oscillation. Autocorrelation in the time domain is readily found to be

$$C_\Gamma(t_j, t_k) = \frac{A(t_j)A(t_k)}{2}\cos(\omega_p(t_j - t_k)). \tag{30}$$

Clearly, the TAP-noise is colored.

The variance profiles in the time and frequency domain are given by

$$\sigma^2_\Gamma(t_j) = \frac{A(t_j)^2}{2} \tag{31}$$

$$\sigma^2_{\mathcal{F}\Gamma}(\omega_j) = \frac{1}{4}(|\mathcal{F}A(\omega_{|p-j|})|^2 + |\mathcal{F}A(\omega_{f(p+j)})|^2), \tag{32}$$

respectively, where

$$f(s) = \begin{cases} s & \text{when } s < N \\ s - N & \text{when } s \geq N \end{cases}. \tag{33}$$

In a first approximation, it is useful to assume that the amplitude is linearly related to the signal strength:

$$A(t_j) = \alpha\mu_X(t_j) + \beta \tag{34}$$

so that

$$\mathcal{F}A(\omega_j) = \alpha\mathcal{F}\mu_X(\omega_j) + \beta\sqrt{N}\delta_{j0}, \tag{35}$$

where α and β are positive. The Kronecker delta symbol δ_{ab} is 1 if a is equal to b but 0 else. Substitution of (35) into (32) yields after some manipulations

$$\sigma^2_{\mathcal{F}\Gamma}(\omega_j) = \frac{1}{4}[\alpha^2(\underbrace{|\mathcal{F}\mu_X(\omega_{|p-j|})|^2}_{a} + \underbrace{|\mathcal{F}\mu_X(\omega_{f(p+j)})|^2}_{b}) + N\delta_{pj}(\beta^2 + 2\alpha\beta\mu_{X,M})], \tag{36}$$

where $\mu_{X,M}$ is the time average mean signal strength. If the amplitude $A(t_j)$ is constant, $\alpha = 0$, there is a discrete standard deviation peak at the pulsation ω_p of the oscillation. If $A(t_j)$ is purely proportional to the signal strength, $\beta = 0$, there is a continuous peak about ω_p. In other words, variability of the amplitude causes broadening of the standard deviation peak in the frequency domain. Mostly, term b in (36) can be neglected with respect to term a. Therefore, the flanks of the peak have the shape of the modulus profile in the frequency domain. If both α and β in (34) differ from zero, the standard deviation peak consists of a discrete peak at ω_p superposed on a continuous peak around ω_p. The model presented will be validated in Section 5.

If possible, spectrally localized noise should be reduced by preventing the electrical interference at its root. However, its complete elimination might prove unattainable. The presence of spectrally localized noise reduces the quality of parameter estimations carried out by regression if it overlaps spectrally with the ideal signal. Specifically, noises localized around low frequencies are most disadvantageous.

[4] As far as the data from the UGent set are concerned, it is notable that only the odd harmonics are present. The cause of the absence of the even multiples is unknown. However, even multiples were observed beside the odd multiples in other pulse response sets collected at Ghent University.

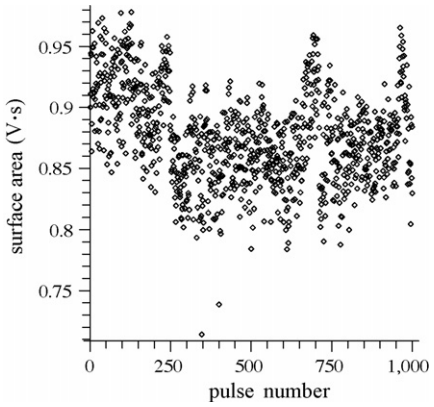

Fig. 12. Surface area under the pulse response curve vs. the number of the pulse for the WUStL data set. As the area is proportional to the intensity of the pulse, this gives an impression of the variability of this intensity. The graph shows short-term random behaviour and slow trends.

4.3. Pulse intensity and mass spectrometer sensitivity

In many cases, the intensity of the TAP reactant pulses is poorly reproducible. In reality, different trends may be observed and overlapped, e.g.

A. The drop in pulse intensity over a long pulse series could be due to a decrease in pressure in the blend tank with each pulse, though the number of molecules per pulse is small (10^{13}–10^{15} molecules).
B. The temperature of the pulse valve tends to increase with use so the resistance in the wire coils increases, causing a slight decrease in the pulse intensity.
C. The pulse intensity drifts if the manifold temperature changes significantly [1].

Fig. 12 gives an impression of the variability of the pulse intensity for the WUStL data set. Each point corresponds to the surface area under the curve of a pulse response. The graph shows short term random behaviour superposed on slower fluctuations. No periodic phenomena have been revealed by taking the discrete Fourier transform.

In the case where no reactions or only first order or quasi-linear reactions take place in the reactor, the shape of the pulse response is not affected by the intensity of the pulse. Therefore, the influence of the variability of the pulse intensity can be modelled by thinking the mean pulse response ($\mu_X(t_j)$) is multiplied by a random factor $1 + \zeta\Lambda$, so that the pulse response ($X(t_j)$) becomes the superposition of ($\mu_X(t_j)$) and the stochastic process ($\Psi(t_j)$):

$$\Psi(t_j) = \zeta\Lambda\mu_X(t_j), \tag{37}$$

ζ is a dimensionless constant much smaller than 1 and Λ is a stochastic variable with mean 0 and standard deviation 1 but unknown distribution. Obviously, ζ must be expected different for each pulse valve. In fact, as the pulse intensity is amenable to trends, ζ must be expected different for each experimental series.

Although the uncertain contribution ($\Psi(t_j)$) mostly has to do with an imperfect pulse gas delivery, as reflected in the enumeration A–C above, it can also be rooted in slow fluctuations of the mass spectrometer's sensitivity. In either case it would not be appropriate to denote ($\Psi(t_j)$) as noise as it is no time-varying stochastic phenomenon but a phenomenon varying from pulse to pulse.

It is easily verified that the autocovariances in the time and the frequency domain are given by $C_\Psi(t_j, t_k) = \zeta^2\mu_X(t_j)\cdot\mu_X(t_k)$ and $C_{\mathcal{F}\Psi}(\omega_p, \omega_q) = \zeta^2\mathcal{F}\mu_X(\omega_p)\cdot\overline{\mathcal{F}\mu_X(\omega_q)}$, respectively. Especially, the variance profiles ($\sigma^2_\Psi(t_j)$) and ($\sigma^2_{\mathcal{F}\Psi}(\omega_j)$) are defined by

$$\sigma^2_\Psi(t_j) = \zeta^2\mu_X(t_j)^2 \tag{38}$$

and

$$\sigma^2_{\mathcal{F}\Psi}(\omega_j) = \zeta^2|\mathcal{F}\mu_X(\omega_j)|^2. \tag{39}$$

Variability of the pulse intensity is reflected in a standard deviation peak about 0 Hz. Fig. 11 indeed shows such a peak. As is clear from (39), its flank has the shape of the modulus profile in the frequency domain.

In practice, area normalization after baseline correction rules out the uncertainty caused by the variability of the pulse intensity or mass spectrometer sensitivity. However, as to some extent it would deform the noise present, area normalization was not applied to the data sets presently used with the purpose to characterize the noise.

4.4. Absence of cross-effects

Consider the natural case where multiple noise types are present in the pulse responses. In the most general case, a pulse response must be considered a superposition of the ideal pulse response and stochastic processes due to the uncertainty concerning the intensity of the pulses (type Ψ, Section 4.3), an Ornstein–Uhlenbeck noise (type Φ, Section 4.1) and multiple spectrally localized noises (type Γ, Section 4.2) with different pulsations ω_p.

Stochastic processes of type Ψ, Φ and Γ have different origins. Therefore, they can be considered stochastically independent. As a result, they are uncorrelated so that their autocovariances can simply be added up. In the time domain:

$$C_X(t_j, t_k) = C_\Psi(t_j, t_k) + C_\Phi(t_j, t_k) + C_\Gamma(t_j, t_k) \tag{40}$$

$$\sigma^2_X(t_j) = \sigma^2_\Psi(t_j) + \sigma^2_\Phi(t_j) + \sigma^2_\Gamma(t_j), \tag{41}$$

and in the frequency domain:

$$C_{\mathcal{F}X}(\omega_j, \omega_k) = C_{\mathcal{F}\Psi}(\omega_j, \omega_k) + C_{\mathcal{F}\Phi}(\omega_j, \omega_k) + C_{\mathcal{F}\Gamma}(\omega_j, \omega_k) \tag{42}$$

$$\sigma^2_{\mathcal{F}X}(\omega_j) = \sigma^2_{\mathcal{F}\Psi}(\omega_j) + \sigma^2_{\mathcal{F}\Phi}(\omega_j) + \sigma^2_{\mathcal{F}\Gamma}(\omega_j). \tag{43}$$

Multiple spectrally localized noises (type Γ) with different pulsations ω_p have a common origin: the electric energy transmission. Therefore, it is not a priori ensured that they are stochastically independent. However, if their mutual phase differences at the start of each new pulse response recording are

completely random, i.e. have no expected value, they are uncorrelated. This is a fair assumption, even if there is a constant phase difference between the oscillations in the global time frame.[5] The mutual phase differences between spectrally localized noises with equal pulsation can be expected to be constant. Therefore, cross-effects cannot be excluded in this case.

It can be concluded that the autocovariances of all stochastic processes can simply be added to give the autocovariance of the pulse response, both in the time and the frequency domain, as long as there are no multiple spectrally localized noises with equal pulsation.

4.5. Deviation from normality

Of the three forms of uncertainty discussed, only the first is Gaussian. The distribution of the samples of a spectrally localized noise ($\Gamma(t_j)$) defined by (29) is known, but non-Gaussian, U-shaped:

$$P[x < \Gamma(t_j) < x + \mathrm{d}x] = \frac{\mathrm{d}x}{\pi\sqrt{A(t_j)^2 - x^2}}, \quad \text{for all } -A(t_j) < x < A(t_j). \tag{44}$$

The distribution of the samples of a stochastic process ($\Psi(t_j)$) defined by (37), due to uncertain pulse intensity or mass spectrometer sensitivity, is unknown. Only the standard deviation profile ($\sigma^2_\Psi(t_j)$) can be estimated and the mean is known to be zero everywhere, by definition. As this form of uncertainty is liable to drifts, the stochastic process ($\Psi(t_j)$)

- can be expected to differ from experimental series to experimental series,
- must be expected to be non-Gaussian.

As a consequence, the distribution of the deviation from the mean

$$X(t_j) - \mu_X(t_j) = \Phi(t_j) + \sum_{k=1}^{S} \Gamma_k(t_j) + \Psi(t_j) \tag{45}$$

is generally not normal. As known, the traditional mean-square regression method for determining model parameters is based on the assumption of normal distribution [14]. Therefore, its direct application to TAP-data analysis is generally not valid. Thus, in principle, the deviations from average have to be transformed to become normally distributed prior to the regression. During the regression, the heteroskedasticity of the pulse response must be taken into account.

[5] As an example, consider as noise an oscillation with frequency 60 Hz, the American mains frequency. Say a pulse is given every 5 s. This is exactly 300 times the period of the oscillation. Therefore, the phase of the oscillation is expected to be the same at the moment each pulse recording is started. However, the mains frequency is not 100% constant and 100% precision is neither attained by the pulse timer. For these reasons, the initial phase of the oscillation can be considered random. To gain absolute certainty about this, a small random time interval could be left between the end of each pulse recording and the next one.

5. Full noise analysis

It was justified in Section 4.4 that the variance profiles due to the different uncertainty effects could simply be added to obtain the total variance profile. In the most general case where there is Gaussian noise, an uncertain pulse intensity and S spectrally localized noises, an expression derived from Eqs. (19), (39) and (36) applies:

$$\sigma^2_{\mathcal{F}X}(\omega_j) = \frac{(1-\rho^2)(\sigma^2)_M}{1 - 2\rho\cos(\omega_p \Delta t) + \rho^2} + \zeta^2 |\mathcal{F}\mu_X(\omega_j)|^2 + \frac{1}{4}\sum_{k=1}^{S}\Big[\alpha_k^2(|\mathcal{F}\mu_X(\omega_{|p_k - j|})|^2 + |\mathcal{F}\mu_X(\omega_{f(p_k+j)})|^2) + N\delta_{p_k j}(\beta_k^2 + 2\alpha_k\beta_k\mu_{X,M})\Big]. \tag{46}$$

As an example, the noise present in the WUStL data set has been analyzed. Here, the sampling interval is 4 ms. If it is assumed that the correlation time θ is about the same as the one found for the UGent set: $\theta \approx 0.789$ ms, the ratio $\Delta t/\theta$ is about 5. As this value is higher than 2.5, the unfavourable autocorrelation effect on the variance profile in the frequency domain can be neglected (see Section 4.1). This amounts to putting $\rho = 0$ in (46).

Fig. 11 shows five spectrally localized noises: at 50, 60, 70, 110 and at 120 Hz.[6] The noises with frequencies of 50 and 110 Hz are neglected. Therefore, in (46), $S = 3$ and, from (13), $p_1 = N\,\Delta t{\cdot}60$ Hz, $p_2 = N\,\Delta t{\cdot}70$ Hz and $p_3 = N\,\Delta t{\cdot}120$ Hz.

Eight parameters remain unknown: $(\sigma^2)_M$, ζ, α_1, β_1, α_2, β_2, α_3 and β_3. Multiple regression attempts have shown that the most meaningful correspondence between the observed and theoretical standard deviation profile is obtained assuming that the amplitude of the oscillation of 60 Hz is perfectly proportional to the signal strength, $\beta_1 = 0$, and that the amplitude of the noise about 70 Hz is constant, $\alpha_2 = 0$.

If the values $j \in \{p_2, p_3\}$ are excluded from the domain of interest, it is derived from (46) that

$$\sigma_{\mathcal{F}X}(\omega_j) = \sqrt{(\sigma^2)_M + \zeta^2|\mathcal{F}\mu_X(\omega_j)|^2 + \tfrac{1}{4}\textstyle\sum_{k\in\{1,3\}} \alpha_k^2\left(|\mathcal{F}\mu_X(\omega_{|p_k-j|})|^2 + |\mathcal{F}\mu_X(\omega_{f(p_k+j)})|^2\right)}. \tag{47}$$

Eq. (14) stated that the elements of the discrete Fourier transform beyond the Nyquist pulsation can be derived from those before, corresponding to the fact that they do not add physical meaning to it. For that matter

$$\sigma^2_{\mathcal{F}\mu_X}(\omega_j) = \sigma^2_{\mathcal{F}\mu_X}(\omega_{N-j}). \tag{48}$$

Therefore, the domain of interest is further reduced to the first half of the discrete Fourier transform.

Eq. (47) has four unknown parameters: $(\sigma^2)_M^{1/2}$, ζ, α_1 and α_3. These are estimated by non-linear regression of the observed

[6] The peak about 0 Hz is due to a variable pulse intensity.

Table 1
Parameter estimates with their 95% confidence intervals obtained by non-linear regression of the observed standard deviation profile in the frequency domain with Eq. (47)

Parameter	Value	Unit
$(\sigma^2)_M^{1/2}$	0.00497 ± 0.00029	V
ζ	0.0434 ± 0.0002	–
α_1	0.0281 ± 0.0003	–
α_3	0.00266 ± 0.00039	–

The regression is shown in Fig. 11.

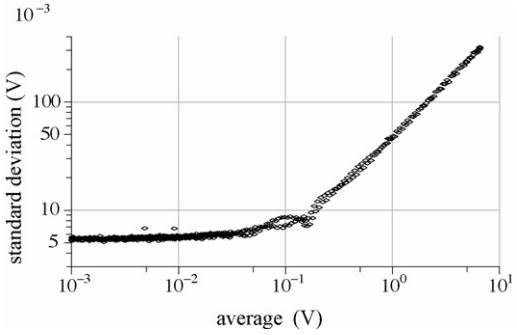

Fig. 13. Logarithmic graph of the standard deviation of the samples as a function of their mean for the WUStL set, resulting from elimination of the sample times. The graph shows that the standard deviation is more or less proportional to the signal strength, although the standard deviation axis intercept differs from zero. This means that some noise remains present once the pulse response is near extinction.

standard deviation profile with (47), abstracting the singular points at 50, 70, 110 and at 120 Hz to improve the regression. The result is shown in Fig. 11 and yields the results listed in Table 1.

The standard deviation profile resulting from this regression may be symbolized as $\sigma_{\mathcal{F}X,\mathrm{fit}}(\omega_j)$. From comparision of (46) and (47), it is clear that there is still a deficit at the frequencies 70 and 120 Hz:

$$\sigma_{\mathcal{F}X}(\omega_{p_2})^2 - \sigma_{\mathcal{F}X,\mathrm{fit}}(\omega_{p_2})^2 = \frac{N}{4}\beta_2^2 \tag{49}$$

$$\sigma_{\mathcal{F}X}(\omega_{p_3})^2 - \sigma_{\mathcal{F}X,\mathrm{fit}}(\omega_{p_3})^2 = \frac{N}{4}(\beta_3^2 + 2\alpha_3\beta_3\mu_{X,M}). \tag{50}$$

Estimations of β_2 and β_3 can be obtained simply by solving the Eqs. (49) and (50). This yields estimates

$$\beta_2 = 0.00166 \ \mathrm{V} \tag{51}$$

and

$$\beta_3 = 0.00270 \ \mathrm{V}. \tag{52}$$

In conclusion, the uncertainty the pulse responses from the WUStL set are subjected to, can be modelled as the stochastic process $(X(t_j) - \mu_X(t_j))$:

$$\begin{aligned} X(t_j) - \mu_X(t_j) &= (\sigma^2)_M^{1/2} G(t_j) + \zeta \Lambda \mu_X(t_j) + \alpha_1 \mu_X(t_j)\cos(\omega_{p_1} t_j + \Theta_1) \\ &\quad + \beta_2 \cos(\omega_{p_2} t_j + \Theta_2) + (\alpha_3 \mu_X(t_j) + \beta_3)\cos(\omega_{p_3} t_j + \Theta_3) \end{aligned} \tag{53}$$

where

- $(G(t_j))$ is white Gaussian noise with mean 0 and time average standard deviation 1;
- Λ is a stochastic variable with mean 0 and standard deviation 1;
- $\omega_{p_1} = 2\pi \cdot 60$ Hz, $\omega_{p_2} = 2\pi \cdot 70$ Hz and $\omega_{p_3} = 2\pi \cdot 120$ Hz;
- Θ_1, Θ_2 and Θ_3 are stochastic variables with uniform distribution over $[0, 2\pi]$;
- the parameters $(\sigma^2)_M^{1/2}$, ζ, α_1, and α_3 are listed in Table 1;
- the parameters β_2 and β_3 are given by (51) and (52), respectively.

Consider the experimental interdependence of the standard deviation and the mean, shown in Fig. 13. In good approximation, the standard deviation is proportional to the signal strength, although it differs slightly from zero at zero signal strength. This corresponds to the result of the noise analysis. Indeed, the most important uncertainty effects, about 0 and 60 Hz, are found to be more or less proportional to the signal strength, while less important ones depend less on this signal strength.

The Gaussian noise is not believed to be due to avoidable interference. Therefore, it may be called the "essential" noise. This analysis shows the standard deviation of the essential noise in the WUStL case to amount to about 0.1% of the peak response. In the case of the UGent set, this was about 4% (see Section 4.1.4). This difference is not surprising, as the UGent data set was collected on a TAP-1 setup, while the WUStL set was the result of experiments performed on a TAP-2 setup. Compared to the TAP-1 system, the TAP-2 has the advantage that the mass spectrometer collects a higher fraction of the molecules leaving the reactor. Therefore, the signal is stronger and less sensitive to essential noise. In the case of the presently analyzed noise (WUStL data set), the Gaussian noise is clearly negligible with respect to the spectrally localized noise and the uncertainty caused by a variable pulse intensity. From the parameter estimates listed in Table 1, or even the mere observation of Fig. 11, it is clear that the spectrally localized noise is rather important. For example, the amplitude of the 60 Hz oscillation is about 3% of the signal strength. Clearly it would be interesting to examine if the interference causing the spectrally localized noise could be reduced in a physical manner. The present noise analysis, applied to data collected at Washington University, revealed a pulse intensity variability of about 4%. This is a typical number [2] and similar variabilities have been derived from data collected at Ghent University.

6. Conclusions

Noise limits the precision by which physico-chemical parameters can be estimated by regression of pulse responses obtained on TAP-setups. These pulse responses are assumed to be liable to two types of noise. The first is Gaussian with standard deviation increasing as a function of the signal

strength. When this type of noise is dominant, which is likely in pulse responses recorded on TAP-1 setups, its colored character puts a limit to the improvement of the precision of the parameter estimations achievable by increasing the sampling frequency of the signal. The second type of noise is spectrally localized, mostly around harmonics of the mains frequency. This noise is caused by interference by electric devices near the mass spectrometer, possibly part of the TAP apparatus itself, or directly by the mass spectrometer's ac power supply.

Analytical expressions have been presented for the autocovariance of the pulse response, both in the time and the frequency domain. Specifically, regression of the observed standard deviation profile in the frequency domain with the analytical expression, allowed to derive quantitative information about the fine structure of the noise. The presently developed error analysis can be applied to single-pulse TAP-data resulting from diffusion/reaction phenomena.

Acknowledgments

The authors would like to express their gratitude to Veerle Balcaen for placing the Ghent University experimental data at their disposal. They also thank Maria Olea for fruitful discussions.

R. Roelant is grateful to the Fund for Scientific Research-Flanders (F.W.O.-Vlaanderen) for a Research Assistantship.

D. Constales was supported by project BOF/GOA 01GA0405 of Ghent University.

References

[1] J.T. Gleaves, J.R. Ebner, T.C. Kuechler, Catal. Rev. Sci. Eng. 30 (1988) 49.
[2] J.T. Gleaves, G.S. Yablonskii, P. Phanawadee, Y. Schuurman, Appl. Catal. A: Gen. 160 (1997) 55.
[3] J.P. Huinink, J.H.B.J. Hoebink, G.B. Marin, Can. J. Chem. Eng. 74 (1996) 580.
[4] G.S. Yablonsky, D. Constales, J.T. Gleaves, SAMS 42 (2002) 1143.
[5] S.O. Shekhtman, G.S. Yablonsky, S. Chen, J.T. Gleaves, Chem. Eng. Sci. 54 (1999) 4371.
[6] P.Z. Peebles, Probability, random variables, and random signal principles, in: S.W. Director (Ed.), McGraw-Hill Series in Electrical Engineering, second ed., McGraw-Hill, Singapore, 1987, p. 142.
[7] C.W. Therrien, Discrete random signals and statistical signal processing, in: A.V. Oppenheim (Ed.), Prentice Hall Signal Processing Series, Prentice-Hall, Englewood Cliffs, 1992, p. 86.
[8] J.H. Seinfeld, L. Lapidus, Process modeling, estimation, and identification, in: N.R. Amundsen (Ed.), Mathematical Methods in Chemical Engineering, vol. 3, Prentice-Hall, Englewood Cliffs, 1974, p. 197.
[9] A. Papoulis, Probability, random variables, and stochastic processes, in: S.W. Director (Ed.), McGraw-Hill Series in Electrical Engineering, third ed., McGraw-Hill, Singapore, 1991, p. 285.
[10] I. Sack, V. Balcaen, M. Olea, H. Poelman, G.B. Marin, Catal. Today 112 (2006) 68.
[11] J.L. Doob, Ann. Math. 43 (1942) 351.
[12] G.E. Uhlenbeck, L.S. Ornstein, Phys. Rev. 36 (1930) 823.
[13] M.C. Wang, G.E. Uhlenbeck, Rev. Mod. Phys. 17 (1945) 323.
[14] J. Mateu, Environ. Manage. 21 (1997) 767.