

# Investigation of structure of noise component in molecular spectra

M E Sibgatullin<sup>1</sup>, D Z Galimullin<sup>1</sup>, D I Kamalova<sup>1</sup> and M Kh Salakhov<sup>1,2</sup>

<sup>1</sup> Kazan Federal University, Kazan 420008, Russia

<sup>2</sup> Tatarstan Academy of Sciences, 420111, Bauman St, 20, Kazan, Russia

E-mail: sibmans@mail.ru

**Abstract.** In the work the noise component of molecular spectra is investigated by use the Hurst index which is the quantitative parameter of memory effects. It was shown that the low-frequency noise component predominates. The Hurst index is from 0.6 to 0.7 for the studied FTIR spectra registered at different temperatures.

## 1. Introduction

In applied spectroscopy the presence of an experimental noise influences on the mathematical processing of weak spectral absorption bands. An existence of the noise caused by both the nature of the studied object and the recorder device, in the case of the small signal/noise ratio does not allow to restore the spectral contour and, therefore, to estimate the important physical parameters: an intensity, a width, a maximum, a number of components of complex spectral contours. The investigation of the noise structure is an important step of the spectroscopic experiment and it is directed at a reduction of errors to obtain reliable data about a studied object.

For successful and effective use of different mathematical methods of processing of molecular spectra the approaches which allow to take account of a priori information on the nature of useful signal and noise, are of important significant. We would remind you that a noise is usually defined by the recorder noise which does not depend on the signal value in infrared spectral range. The recorders for visible range are much better than ones for infrared range: separate photons can be registered even. For interferometers which are a basis of the Fourier-spectrometers the noise is proportional to a square root of source intensity [1].

In this work, to investigate the noise component structure we used the Hurst index ( $H$ ) which is the quantitative characteristic of memory effects. It characterizes the noise from the point of view of a prevalence of a low-frequency or high-frequency component in this experimental noise. We studied the noise which exists at the IR spectra of 1,2-dichloroethane introduced in branched copolymer of methylmethacrylate as a conformational-inhomogeneous probe to study of local mobility of macromolecules [2]. IR spectra were registered by a FTIR spectrometer Spectrum 100 Perkin Elmer in the range of 400 - 4000  $\text{cm}^{-1}$  at the various temperatures from 100 to 300 K. The mathematical treatment of 44 spectra was carried out.

## 2. Separation and analysis of the noise component from molecular spectra

For separation of the noise component of spectra it is necessary to carry out the denoising procedure of the low-frequency component from a signal. Considering that fact that molecular spectra are smooth, it



is possible to carry out smoothing procedure with the subsequent removal of the low-frequency component. In this work the addition of the smallest distortions of spectral contours which are characterized by the high signal/noise ratio was used as a criterion of smoothing.

One of the simplest and widespread methods of mathematical processing of the signals distorted by a random noise is the Savitzky-Golay algorithm. The computing algorithm or the filter for smoothing represents an operation of linear weighing of values of entrance data. Weight multipliers are simply filter coefficients. The general procedure of smoothing can be presented by the following form [3, 4]:

$$\varphi_n = \sum_{k=-N_B}^{N_B} a_k f_{n-k}, \quad (1)$$

where  $a_k$  are the weight coefficients of the filter,  $2N_B$  is a filter width,  $\varphi_n$  represents the convolution of an experimental signal  $f_n$  with sequence of coefficients  $\{a_k\}$ . In the Savitzky-Golay algorithm the coefficients of the filter are defined from the condition:

$$\min \|f - Va\|_{L_2}, \quad (2)$$

where  $f$  is the vector with length of  $n$ ,  $a$  is the vector with length of  $m+1$ ,  $V$  is the Vandermonde matrix  $(n \times (m+1))$ , determined by the equation:  $V_{ik} = i^{k-1}$ .

In figure 1a one of the experimental spectra is given and the fragment of the spectrum in the range of  $800-525 \text{ cm}^{-1}$  is shown by the increased scale. As seen from figure 1b, for the absorption band at  $750 \text{ cm}^{-1}$  the signal/noise ratio is big, and the distortions are shown generally on its contour wings while for the absorption bands at  $650$  and  $710 \text{ cm}^{-1}$  the signal/noise ratio is small, and there is a distortion of the spectral contour as a whole. The example of the noise component at the interval of  $2075-1775 \text{ cm}^{-1}$ , for which the Hurst index was calculated, is given in figure 1b.

One of the methods of the optimum estimation of  $H$  is the statistical method of rated scope or  $R/S$  - analysis which was offered by Hurst [3]. Let  $\{x(t)\}$  is a random process. For the analyzed selection  $\{x_i = x(t_i)\} (i = 0, \dots, T)$  a sum of random values is calculated as

$$X_\tau(t) = \sum_{i=1}^t (x_i - \langle x \rangle_\tau), \quad 1 \leq t \leq \tau \leq T, \quad (3)$$

where  $\langle x \rangle_\tau$  means an average for the selection at the interval  $[0, \tau]$ . Then, the scope function  $R(\tau)$  as the difference between the maximum and the minimum of  $X_\tau(t)$  is calculated by

$$R(\tau) = \max_{t \in [1, \tau]} X_\tau(t) - \min_{t \in [1, \tau]} X_\tau(t). \quad (4)$$

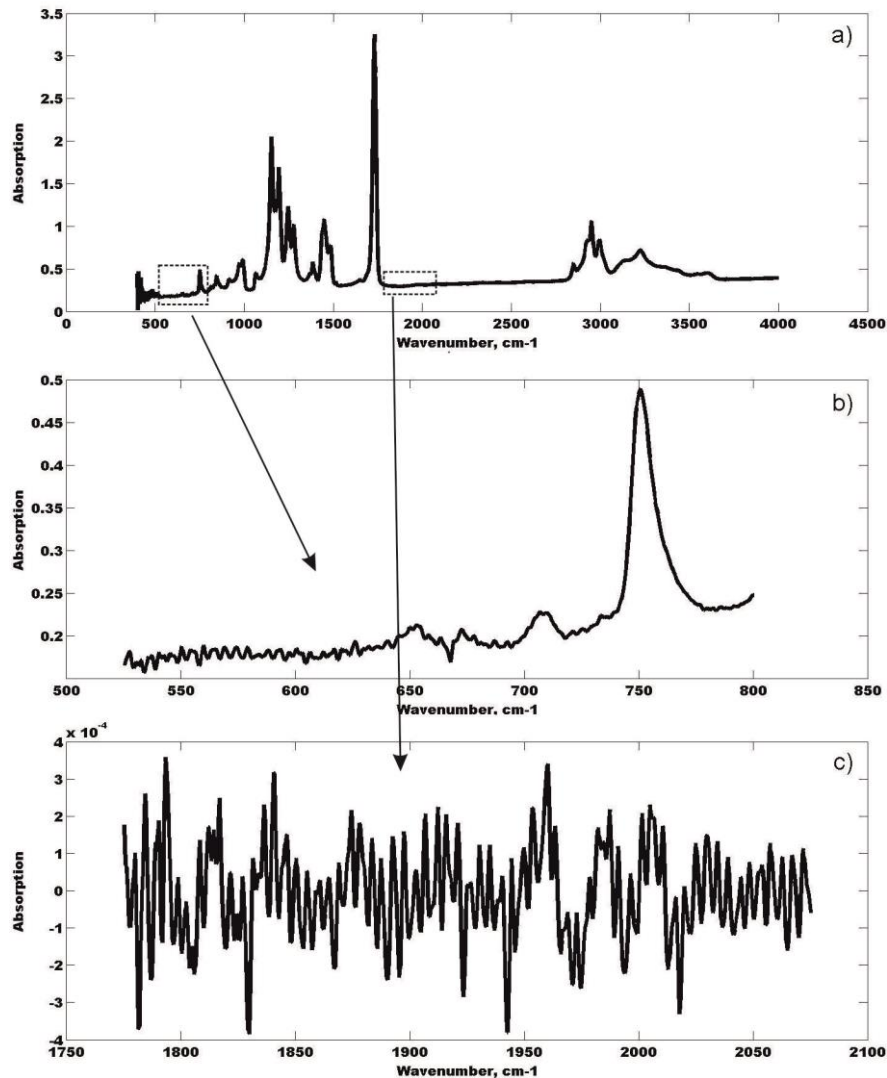
Hurst showed empirically that  $R/S$  ratio is written by the power law:

$$R/S = (a\tau)^H, \quad (5)$$

where  $a$  is any parameter,  $H$  is the Hurst index,  $S$  is the mean-square distance which is determined by

$$S(\tau) = \left[ \frac{1}{\tau} \sum_{i=1}^{\tau} (x_i - \langle x \rangle_\tau)^2 \right]^{1/2}. \quad (6)$$

Thus, it is possible to estimate the Hurst index when the experimental data are approximated by equation (5).



**Figure 1.** Experimental spectrum (a), distortion of spectral contour by noise (b), noise component (c).

To provide a statistical homogeneity of the results of  $R/S$  measurements we used the procedure given in [4].  $N$  random sequences (length  $\tau$ ) are selected from the analyzed selection with equal probabilities. For each sequence the relation  $R/S$  is calculated. From initial time series at the fixed  $\tau$  we obtain  $N$  independent  $R/S$  values. The results corresponding to the same value of  $\tau$ , are averaged, and the values  $R/S$  are postponed in double logarithmic scale as a function of  $\tau$ . Further, the obtained values  $R/S$  are approximated according to the law (5) by the least-squares method, following which the estimation of the parameter  $H$  is carried out. Since the length of the analyzed noise is rather small (600 points), it is necessary to estimate an accuracy of calculations for the determination of a reliability of the obtained results. The calculation error was defined as follows. With the Student distribution the confidential interval for parameter  $H$  was defined

$$\left( \bar{H} - t_{\gamma} \frac{s}{\sqrt{n}}, \bar{H} + t_{\gamma} \frac{s}{\sqrt{n}} \right), \quad (7)$$

where  $\bar{H}$  is an average value of  $H$ ,  $t_\gamma$  is the Student coefficient,  $n$  is an amount of sampling,

$s = \sqrt{\frac{\sum_{i=1}^k n_i (H_i - \bar{H})^2}{n-1}}$  is a dispersion. The confidence interval covers the  $H$  with the reliability  $\gamma$ .

We used the following values of the parameters:  $n=10$ ,  $\gamma = 99$ . The maximum error of the calculations was not in excess of 5%.

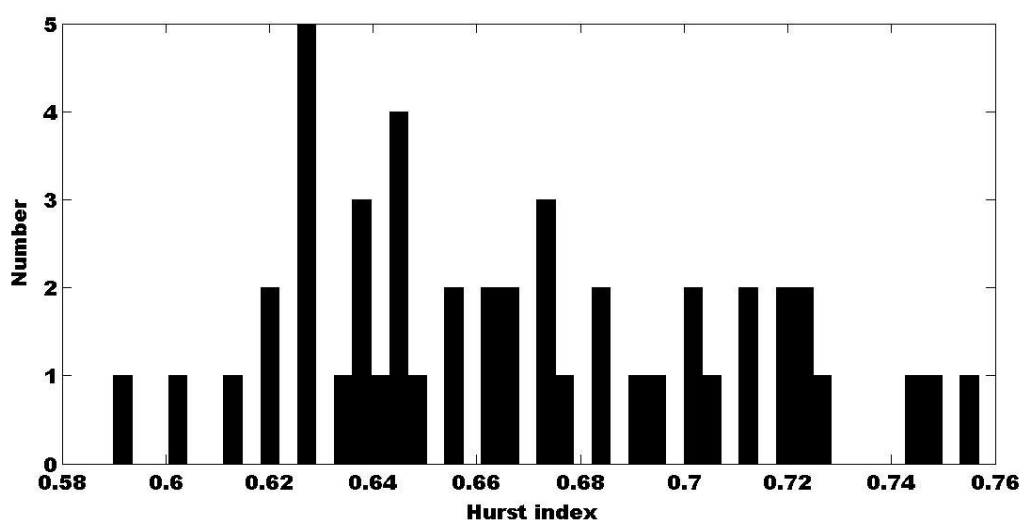


Figure 2. Histogram of the Hurst index.

The histogram of the Hurst index calculated for the noise component of the spectra for different temperatures is given in figure 2. We would remind you that noises can be classified formally as follows: the high-frequency noise ( $0 < H < 0.5$ ), uncorrelated (white) noise ( $H = 0.5$ ) and low-frequency noise ( $0.5 < H < 1$ ). The low-frequency noise component becomes predominate when  $H > 0.5$ . Our investigations showed the presence of the low-frequency noise structure at FTIR spectra of 1,2-dichloroethane introduced in branched copolymer of methylmethacrylate. The Hurst index is in the range from 0.6 to 0.7 generally for all spectra registered at different temperatures. Therefore, when mathematical processing of experimental data it is necessary to take account of the low-frequency noise component, for example, to execute the procedure of the preliminary removal of a noise using the wavelet-analysis with the Hurst index obtained from the analysis of experimental data.

## References

- [1] Bell R J 1972 *Introductory Fourier Transform Spectroscopy* (New York and London: Academic Press)
- [2] Kamalova D I et al 2008 *Konformatsionnye zondy v izuchenii lokalnoi podvizhnosti polimerov* (Moscow: Fizmatkniga)
- [3] Savitsky A 1964 *Anal. Chem.* **36** 1627
- [4] Salakhov M Kh and Kharintsev S S 2001 *Matematicheskaya obrabotka i interpretatsiya spektroskopicheskogo eksperimenta* (Kazan, KSU)