

Quantized Histograms and the Statistical Detection of Hidden Distortions within Diffuse Spectra from Biological Systems

Raoul R. Nigmatullin^{1*}, Geoff Smith² and Paul C. Butler²

¹Theoretical Physics Department, Kazan State University, Kazan, 420008, Tatarstan, Russia.

²School of Pharmacy and Pharmaceutical Sciences, De Montfort University, The Gateway, Leicester, LE1 9BH UK.

* Correspondence and reprint requests to : Prof. R.R. Nigmatullin, Department of Theoretical Physics, Kazan State University, Kremlevskaya Str. 18, Kazan, 420008, Tatarstan, Russia. E-mail address: nigmat@knet.ru, fax +7.8432.764093

KEYWORDS: Statistical detection, fluctuation fork, hidden signals, dilute samples, near infrared spectroscopy, signal processing

ABSTRACT

Spectroscopy and hyperspectral imaging are ideally suited for astrobiological research because of their remote sensing capabilities. The detection of a small concentration of an unknown substance from spectroscopic data is an important problem in spectroscopy. Usually, this detection is based on the recognition of specific labels such as new resonance lines, or peaks, appearing in the spectrograms being analyzed. However, if the substance analyzed is present in very small concentration and visual labels of its presence are absent, then the concentration of substance becomes undetectable or biased to subjective interpretation. We suggest a new method of detection in spectroscopy based on the transformation of initial spectrograms into ordered quantized histograms. This transformation aids the statistical detection of otherwise undetectable signals within the spectrogram by comparing the so-called "fluctuation forks" (FF) of the unknown substance against that of a background measurement. The paper highlights the application of the new methodology to the detection of dilute concentrations of a type of white blood cell (mononuclear leukocytes) and a type of antibody (IgG) in solutions of phosphate buffered saline, using NIR spectroscopy. Detection limits of 500 leukocytes/mL and 0.01 g/mL IgG were achievable. Such detection limits for complex samples suggest that a sensor employing the method might be useful for astrobiological remote sensing. This technique may also have wide applicability to first pass analysis of highly diluted samples prior to more conventional analysis.

1. INTRODUCTION

The detection of small signals from random samplings is one of the main problems in modern science and technology. There are currently a number of signal-processing methods that are utilized to solve this problem, for example, Hurst analysis of fractal random samplings.¹ The powerful tool of wavelet-analysis, based on the generalization of the conventional Fourier transforms,²⁻⁵ has also been developed to extract weak signals from existing noise data. More recently, stochastic dynamics with discrete current time for correlation analysis of random samplings has been considered by Yulmetyev et al⁶. However, critical analysis of these current methods shows that we do not have a universal language for quantitative comparison of random samplings that possess different statistical characteristics. Moreover, many of the aforementioned techniques are relatively insensitive to the detection of weak signals with amplitude comparable with the level of noise. Researchers in microwave and optical SETI must regularly consider signal levels comparable to the noise. Spectroscopic remote sensing experiments conducted robotically on distant planets are likely to encounter high noise levels as well.

1.1 The basis of the new methodology

The deficiencies of current methods have been approached recently with new statistical methods of signal detection based on the recognition of corresponding histograms⁷ and signal-to-staircase transformations.⁸ These new approaches realize the possibility of a statistical method of detection for weak signals, which are hidden inside the random sampling analyzed but nevertheless distort the recognized histogram. For example, when the signal/noise ratio is close to unity, the signal-to-staircase method helps to find the location of the hidden signal and then to smooth it.

This earlier work has led to a new recognition method for statistical distributions of different characteristics, which transforms the random sampling considered into a quantum spectrum, or quantized histogram (QH), having analytical form. These QHs are obtained at certain ratios between the length, L , of the random sampling considered and the number of decompositions, K , of the original random samplings considered. QH analysis provides a universal quantitative language (adopted from quantum mechanics) for the comparison of random samplings of different nature. Moreover, because of the unique sensitivity of QHs it becomes possible to detect, statistically, any otherwise hidden distortions of the initial random sampling. QH analysis recently proved to be uniquely sensitive in the statistical detection of weak seismic tremors ($S/N < 0.1$) foregoing earthquakes.⁹ The basic idea of detection of statistical distortions is based on the calculation of the so-called fluctuation fork (FF), obtained from repeated measurement of QHs from a reference sample. If the statistical distortions in the corresponding noise spectra due to the presence of an additive are sufficient, then the FF of the original reference sample is distorted and the additive is easily detected. Statistical detection using quantized histograms can increase the sensitivity of existing equipment, especially in cases when visual evidence in the form of new resonance lines is absent in the corresponding spectrograms being analyzed. For example, this new method can be used for statistical detection of small concentrations of a new substance (e.g. particles or solutes) present in a solution. It might also find application in astrobiology for analysis of digital imaging and multi-dimensional spectral data, and for increasing sensitivity of existing radio and optical telescopes used for statistical detection of possible small signals received from the galaxy.

1.2 Application of the new methodology

The detection of small signals from spectroscopic (e.g. UV, IR, dielectric, NMR, ESR or fluorescence) data is an important challenge in the development of new sensors, imaging techniques and analytical methodologies. Usually, this detection is based on the recognition of specific labels such as new resonance lines, or peaks, appearing in the spectrograms being analyzed. However, if the substance analyzed is present in very small concentration and visual labels of its presence are absent, then the concentration of substance becomes undetectable or biased to subjective interpretation. For example, the detection of small concentrations of analytes or particles in large volumes of solvent is a problem in many fields such as environmental monitoring, healthcare, pharmaceutical processing, beverage production and many aspects of quality control in manufacturing industries. Current

procedures are time consuming and generally expensive due to sample concentration, derivatization or analysis by complex techniques that require considerable investment in equipment, time and skilled staff.

The new methodology can be applied to a wide class of spectroscopic data of different origin. The sensitivity of the statistical detection will be determined by the stability of the background noise (during long time of experimental observation) and the relative magnitude the background fluctuation fork. This paper explores the sensitivity of the new methodology in the detection of low concentrations of biological cells and an antibody in aqueous buffer solution.

Near-infrared (NIR) Fourier transform spectroscopy was chosen for this study for two reasons. The first reason is that there has been considerable interest in this the technique for in-situ and/or in-process measurements, in a broad range of applications, e.g. food¹⁰, polymers¹¹, non-invasive sensing of physiological metabolites¹². In many of these applications NIR spectra are obtained conveniently using transmission or diffuse reflectance probes (suitable for quantitative in-situ or in-process monitoring) and analyzed for components of interest by multivariate calibration techniques¹³, e.g. the Partial Least Squares method (PLS) and Principal Component Regression (PCR). In this statistical approach, the PLS algorithm is used for most applications. In order to build a reliable calibration model 30-50 samples are typically required. When analyzing complex systems, like blood cells in buffer solution, even more samples are required. Quantitative analysis is performed on absorbance spectra and typically up to 30 scans that are co-added to a single spectrum to reduce noise effects. In addition, several data pre-processing techniques may be used (for example, vector normalization, multiplicative scattering correction, or 1st and 2nd derivative) to further reduce disturbing effects.

The second reason for adopting NIR for this study is that the conventional methods for analyzing NIR spectra (i.e., PLS and PCR) are largely insensitive to concentrations of analyte much below 0.01 - 0.1%. Moreover, in the case of dilute aqueous systems, the absorbance of NIR light by the analyte is swamped by the absorbance from the high concentration of water in the surrounding media. For these reasons, biological cells and macromolecules are not measured directly by NIR spectroscopy for any routine biomedical application. The application of the new methodology to this form of non-optimal data therefore presents a considerable challenge. This paper demonstrates that the application of the new methodology to a minimal number of repeated samplings of simple optical spectra on highly dilute samples can give statistically meaningful results that cannot be obtained otherwise, without considerable cost and expertise.

2. MATERIALS AND METHODS

Unless otherwise stated, all reagents were purchased from Sigma Chemical Co. UK.

2.1 Preparation of Cell Suspensions

Peripheral blood mononuclear leukocytes (PBL) were prepared from normal human peripheral blood, obtained by venepuncture from a healthy volunteer, using a large gauge needle to prevent upregulation of the cells and anticoagulated using di-potassium EDTA. PBL were isolated by density gradient centrifugation over a layer of Histopaque 1077 (Ficoll/Hypaque, density=1.077) for 20 minutes, followed by removal of the interface band. This was washed (twice) in PBS before resuspension of the cells in PBS to a final concentration of 5×10^6 /mL. Viability, assessed by trypan blude exclusion, was $> 98\%$. The cells were held on ice and used within 6 h. PBS was used for the buffer, as it is a simple isoosmotic solution commonly used in life science applications to maintain short-term cell viability.

2.2 Preparation of Antibody

Murine IgG1 monoclonal antibody, clone LDS101 anti-cytochrome P4501B1 (CYP1B1), prepared in our laboratories was purified from spent hybridoma culture supernatants. The supernatants were centrifuged at 100,000g for 60 min, filtered through Whatman No.1 filter paper and adjusted to pH 8.0 before addition of 10% (w/v) NaCl and 10% (w/v) $\text{NH}_4(\text{SO}_4)_2$ as chaotropic salts to aid binding. This was then applied to a Sepharose 4B-Protein A (Pharmacia) affinity column at 1 ml/min, followed by washing extensively with phosphate buffered saline (PBS, pH 7.4) with 10% (w/v) NaCl, then PBS before elution with glycine pH 2.2 into 1M Tris.Cl pH 8.0. Absorbance of the eluate was monitored at 280nm. Antibody -containing fractions were pooled, then dialysed overnight against excess PBS (pH 7.4), concentrated by dry dialysis over polyethylene glycol (av. M.W. 10,000), redialysed against PBS and filtered through a 0.22 μ M filter. It was stored as small aliquots at 1mg/ml at -20°C . Purity was monitored by both reducing and non-reducing SDS-PAGE analysis, and was shown to be a single protein. Throughout the text the short abbreviation IgG for this antibody is used.

2.3 Experimental measurements

Serial dilutions in the range 5×10^5 to 5×10^2 leukocytes per mL and 100 to 0.1 μ g IgG1 per mL were prepared immediately before measurement by near-infrared spectroscopy. 700 μ L aliquots of each dilution were measured in a 2 mm path length quartz

cuvette, using a Bruker Vector 22/N-C FT-NIR spectrometer, with a thermoelectrically cooled indium arsenide detector. Twenty double-sided interferograms of each sample were recorded consecutively, and each scan saved as a separate file. High resolution (2 cm^{-1}), single-channel spectra (uncorrected for instrument response) were generated from each interferogram, and then analyzed with the help of the QHs technique.

3. DATA TREATMENT PROCEDURE

The statistical ensemble of the background noise comprises a number ($=M$) of single channel NIR spectra (where $j=1, 2, \dots, M$), measured at the same external experimental conditions (related presumably to local temperature and atmospheric pressure). In order to obtain the quantized histogram for the random sampling considered, it is first necessary to calculate the so-called *fluctuation noise*, in accordance with the following formula:

$$n(k, j) = S(k, j) - \langle S(k) \rangle, \quad (1)$$

where the average sampling $\langle S(k) \rangle$ is defined by the standard expression:

$$\langle S(k) \rangle = \frac{1}{M} \sum_{j=1}^M S(k, j) \quad (2)$$

$S(k, j)$ is a discrete data point in the recorded single channel spectrogram with co-ordinates defined by the wave number k (cm^{-1}), (on the abscissa) and an arbitrary measurement of intensity, S (on the ordinate). The index M defines the sampling volume (in our case $M = 20$, the number of spectra analyzed (where $j=1, 2, \dots, M$)). Following this transformation of the data set, one can then calculate a histogram for the fluctuation noise.

A conventional histogram defines the number of discrete amplitudes (defined in applied statistics as relative frequencies) ω_n located in the strip of the width $[n-1, n]$ ($n = 1, 2, \dots, K$) for the given level of column decomposition K and length L . In other words, the parameter K defines the total number of strips obtained for the given distribution of amplitudes located in vertical direction. In our case, it is convenient to measure L in terms of the number of registered data points, N , in each single channel spectrogram. The conventional histogram is *quantized* when the corresponding envelope forming the histogram is destroyed and transformed to a certain set, E_K , of discrete levels depending on the value of K . Each level keeps the same value N_K for a certain set of relative frequencies. The value of K for a normal histogram is chosen empirically from the interval ($5 < K < 20$).¹⁴

The quantized histogram is obtained by increasing the value of K until the situation is reached when the differences between the quantized levels are expressed by integer values. This requirement is not obligatory, one can choose another difference between the quantized levels formed but it is convenient if this difference is expressed in integer numbers. In our case, for the given length $L=8096$ (measured in number of registered data points per spectrum) the desired value of K is found empirically and equals 1010. ($K \geq 1010$ to be more precise). We want to stress here that a general formula for the finding of the optimal value of K cannot be found. Our model experiments, realized with different samplings, showed that besides the values of K and L , which play the decisive role in the forming of the quantized histogram, the level of quantization given by the analog-digital converter (ADC) is also important. In order to reduce this influence, it is necessary to choose an ADC that provides a number of significant digits ≥ 5 .

An *ordered quantified histogram* is defined as a function $E(n_K, N_K, \Omega_K)$, where n_K is the number of quantized levels, N_K is the number of relative frequencies on a given level, and Ω_K stands for a value of the K th level. The QHs can be read as quantitative characteristics of a quantum spectrum. These three above-mentioned statistical parameters can be calculated for the given value of K , which forms the desired QH, and can be applied in principle for transformation of any random sampling. This "universal" spectral language adopted from quantum mechanics can be used instead of various density distributions, which in most cases cannot be easily identified.

For the set of QH's (obtained for the total volume of sampling analyzed, i.e., M) it is then possible to calculate the so-called *fluctuation fork* (FF). This FF represents the calculated dispersion for the given set of QH's. It is calculated as

$$F_K = \pm \sqrt{\langle (\Delta E_K)^2 \rangle}, \quad (3)$$

where the dispersion of the quantized histogram is defined as:

$$\langle (\Delta E_K)^2 \rangle = M^{-1} \sum_{j=1}^M (\Delta E_K(j))^2. \quad (4)$$

Here $\Delta E_K(j) = \Omega_K(j) - \langle \Omega_K(j) \rangle$ defines the deviation of the j -th QH relative to its mean value. The value K , as before, determines the level of decomposition of the quantized histogram.

The fluctuation fork calculated with the use of expression (3) and (4) for buffer solution is depicted in Fig.3. Attentive analysis shows that each FF is characterized by at least by three parameters. These are

the maximal length of the fluctuation fork, D ($=K_{\max}-K_0$); the maximum width, W ; and the total area, A , occupied by of the fluctuation fork on the 2-dimensional plane.

The detection of statistical deviations is then based on the fact that (for relatively large sets of QHs, i.e., $M = 20$) all distortions associated with the background will be located inside the background FF. Other statistical deviations, caused by the presence of another component or predominant factor, will then distort the background FF and have a tendency to increase the values of D , W , and A . It is convenient to define the incremental increase in the parameters D , W , A relative to the corresponding parameter for the background noise, according to the following formula. For example, in the case of the area of the FF this incremental increase is defined by

$$\delta A = \left[\left(\frac{A_s}{A_b} \right) - 1 \right] \cdot 100\%. \quad (5)$$

Here A_b is an area formed by pure buffer solution, while A_s is a value of area occupied by FF caused by combined presence of the additive (i.e. leukocyte cells or IgG1) and the buffer solution.

4. RESULTS AND DISCUSSION

Single-channel spectra for PBS buffer solution and a suspension of leukocytes (500 cells/mL) display characteristic absorption bands of water, and little else from the sample itself (Fig.1).

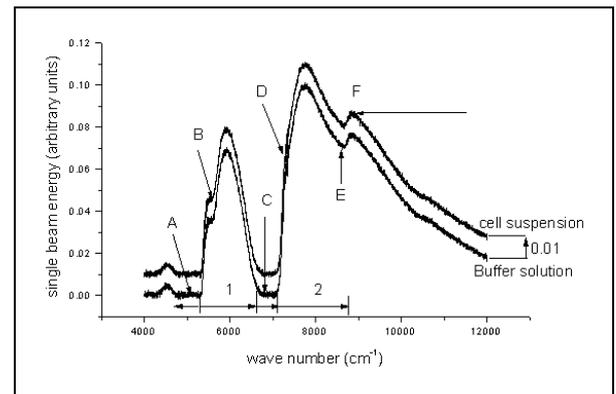


Figure 1. Two different single channel FT-NIR spectra (uncorrected for instrument response) are shown here. The superimposed spectra cannot be differentiated from each other visually, so it was necessary to shift them relative to each other in order to see possible distortions. The lower spectrum represents the FT-NIR spectrum of the *buffer* solution; the upper spectrum (shifted up by 0.01 units) corresponds the FT-NIR spectrum of a suspension of leukocyte cells (containing 500 cells/mL).

At 4000 cm^{-1} there is the high wavenumber end of the fundamental OH stretches. Centered on 5170 cm^{-1} (see region 'A' in Fig.1) is the combination band of OH stretch and OH deformation (region 'B'). At 'C' ($6800 - 7100\text{ cm}^{-1}$) there is the absorption due to the first overtone of the OH stretch. The sharp peaks at $\sim 7300\text{ cm}^{-1}$ (at 'D' on Fig.1) are due to atmospheric water. Above 9000 cm^{-1} (i.e., 'F' on Fig.1) there is very little spectral information from the sample itself. In this region, the higher harmonic overtones from various bond deformations of water are very weak, and what is measured is primarily due to the energy of the tungsten halogen lamp, modified by the efficiency curve of the beam splitter and the response curve of the detector and cell.

Overlaying spectra for the PBS solution and that from a leukocyte suspension containing 500 cells/mL shows that there are no apparent differences between the spectra (Fig.1). The deviations in each spectrum are randomly distributed and cannot be quantified easily. However, if each spectrum is first transformed into a QH (see Fig.2, where the corresponding QHs are depicted) then it is possible to detect some small deviations between any two spectra from more dilute samples.

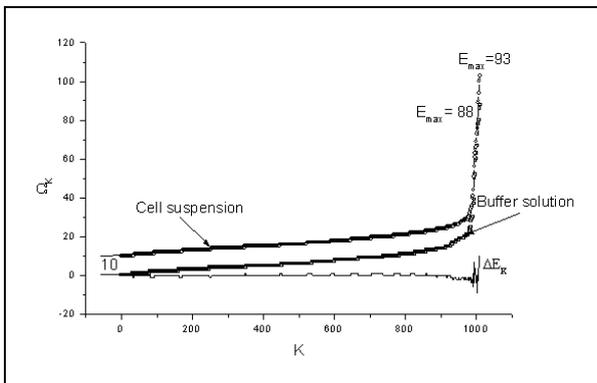


Figure 2. Ordered *quantized histogram* (QH), obtained for the two spectra in Fig.1 using the procedure described in the text, and based on a number of decompositions $K=1010$. For differentiation of these histograms, it was necessary to shift the QH belonging to the leukocyte solution (500 cells/mL) in the upward direction by 10 units. The number of quantized levels for buffer and cell suspension are equal to 88 and 93 respectively. The bottom curve shows the real difference between the two QHs.

In order to establish whether these deviations are statistically relevant, it is necessary to first calculate the FF of the background noise. This FF was obtained from an average of 20 QHs (according to expressions 3 and 4), with each QH being derived from an individual single channel NIR spectrum. Fig.3 shows the fluctuation fork for the background measurement on the buffer solution, and indicates how the parameters D , W , A are determined.

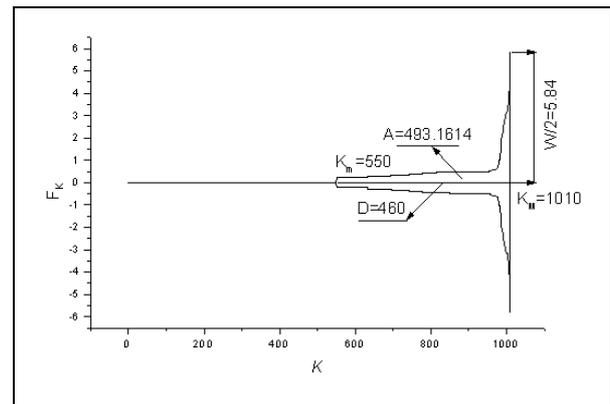


Figure 3. Fluctuation fork for the buffer solution, obtained from twenty averaged samplings. The basic parameters of this fork are the following: $D = K_M - K_m = 1010 - 550 = 460$; the width, $W/2 = 5.84038$ (defined by the maximum value of the FF defines); the total area of the FF, $A = 493.16144$.

All 'native' distortions should be localized inside the FF, whereas 'strange' distortions with other statistical characteristics should be located outside the FF, and therefore easily detected. The parameters δD , δW , δA were then calculated and plotted as a function of cell number, in order to establish the sensitivity of each parameter to the presence of the concentration of cells in the buffer solution. The area parameter (δA) was found to be the most sensitive and exhibits monotonic behavior (Fig.4).

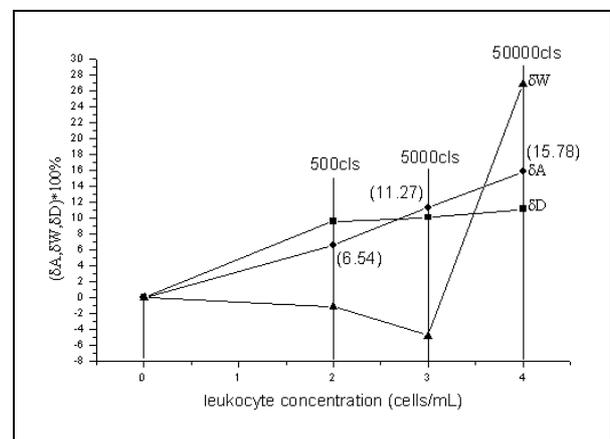


Figure 4. Corresponding dependencies of relative area ($\delta A = [A_{Cell}/A_{Bf} - 1] \cdot 100\%$), relative distance ($\delta D = [D_{Cell}/D_{Bf} - 1] \cdot 100\%$) and relative width ($\delta W = [W_{Cell}/W_{Bf} - 1] \cdot 100\%$) versus concentration of leukocytes. True values of the concentration are shown near the vertical lines and in the Tables. It follows from this figure that the most sensitive parameter exhibiting a monotonic behavior with concentration is the area (A). The corresponding values of the value δA are given in brackets. The absolute values of these parameters are given in the Table 1.

This figure demonstrates that the lowest concentration of cells detectable, against the buffer solution, is 500 cells per mL. In the same manner as for cell suspensions, we can also employ the procedure of

statistical detection for IgG solutions. The final result for the parameters δD , δW , δA as a function of concentration is shown in Fig.5. All calculated parameters for the two experimental situations analyzed are presented in Tables 1 and 2. Fig.5 demonstrates clearly that the lowest concentration of IgG detectable, against the given buffer solution, is 0.1 $\mu\text{g/mL}$.

Type of the files	Length of FF (D)	Width (W)	Total Area (A)
Buffer solution	460	11.68076	493.16144
0.1 $\mu\text{g/mL}$	490	13.16511	517.96335
1.0 $\mu\text{g/mL}$	502	12.62894	534.74623
10 $\mu\text{g/mL}$	509	16.67214	575.87653

Table 1. The absolute values of parameters of the FFs for leukocyte cells for different concentrations.

Type of the files	Length of FF (D)	Half of Width (W/2)	Total Area (A)
Buffer solution	460	5.84038	493.16144
$5 \cdot 10^2$ (cells/mL)	504	5.77256	525.40791
$5 \cdot 10^3$ (cells/mL)	506	5.55901	548.72829
$5 \cdot 10^4$ (cells/mL)	511	7.40338	570.96468

Table 2. The absolute values of parameters of the FFs for antibody IgG at different concentrations

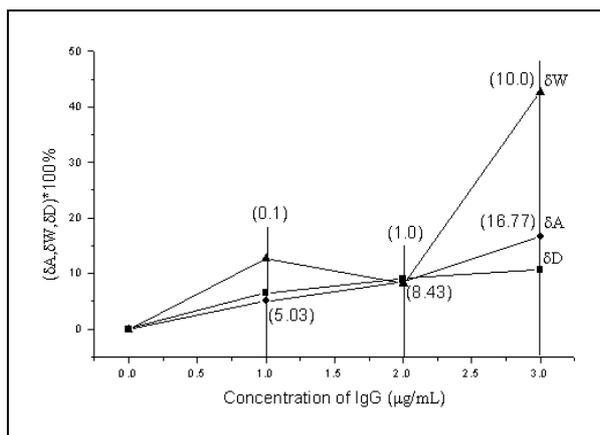


Figure 5. The dependencies of relative area δA , relative distance δD and relative width δW against relative concentration of antibody IgG. The true values of concentration (in $\mu\text{g/mL}$) are given near vertical lines. It follows from this figure the most sensitive parameters exhibiting a monotonic behavior with concentration are the area (A) and distance (D). The corresponding values of δA are given in brackets near vertical lines. The absolute values of these parameters are given in the Table2. The detected distortions exhibit a non-linear monotonic dependence of $\delta A(C)$ and $\delta D(C)$ with increasing of concentration of IgG.

In order to increase the sensitivity of the new methodology it is necessary to choose spectroscopic data that contains maximal information on possible statistical distortions caused by the presence of a new substance. For example, the new methodology would be most informative in detection of Brownian motion if the new substance comprises molecules that are geometrically differentiated from initial molecules forming the background measurement.

The results of these transformations show their possible utility in the analysis of previously undetectable compounds in highly dilute samples. The benefits of sample dilution (compared to sample extraction or derivatization) for valuable, concentrated materials (as are frequently encountered in pathological, prognostic and life science areas of research) indicate that this data handling technique may be a valuable screening tool prior to labor-intensive sample preparation for more expensive techniques. Moreover, the use of IR spectroscopy to analyze proteins is not one recommended by most biochemists when UV spectroscopy would be more appropriate. This further illustrates the utility of the data extraction technique in that it identifies analytes, *quantitatively*, even when the analysis method itself is non-optimal. It should be stressed that the application of new methodology does not require any specific understanding of the multicomponent spectrum (as shown in Fig.1). It is necessary to have relatively stable samplings of background noise for the application of the QH methodology. This sampling helps to differentiate the possible distortions, which are formed in the whole spectrogram, from the noise.

5. CONCLUSION

For the detection of small concentrations of a substance, when the visual labels are absent, one can apply a new methodology based on the quantification of statistical distortions in the form of a controllable fluctuation fork. This fluctuation fork can be calculated for background noise and should be stable to the influence of external factors. Moreover, the background FF should be of low area (A), width (W) and/or length (D), so that a sensitivity to low concentrations of additive may be realized. Of the parameters D, A, and W that characterize the properties of the fluctuation fork, it is the area, A, that provides greatest sensitivity to the presence of a low concentration of cells (i.e. 500 leukocytes/mL) and the presence of the antibody IgG1 in small concentrations (i.e. 0.1 $\mu\text{g/mL}$).

This new methodology will likely find applications in astronomy and in SETI analyzing spectra and in image processing. For the many-dimensional case, the

calculated FFs become two- and many-dimensional and possible distortions can occupy a certain region in a space with calculated volume V (instead of area A used) exceeding the initial volume V_0 occupied by 'blank' noise. Systematic observations over time help in recognition of possible dispersions of the initial forks. Furthermore, repeated observations help to increase the sensitivity to possible abrupt distortions that can appear in fluctuation spectral noise sampled systematically.

REFERENCES

1. Feder, E. Fractals; Plenum Press: New York, 1988.
2. Daubechies , I. Comm. Pure Appl. Math 1988, 41, 909-996.
3. Daubechies , I. IEEE Trans. Inform Theory 1990, 36, 961-1005.
4. Daubechies, I. CBMS Lecture Notes Series, Philadelphia, 1991.
5. Caufman, R. Wavelets and their Applications; John and Barlett Publishing: Boston, 1992.
6. Yulmetyev, R. M., Hanggi, P., Gafarov, F. Phys. Rev. E. 2000, 62, 6178-6194.
7. Nigmatullin, R.R. Physica A 2000, 285, 547-565.
8. Nigmatullin, R.R. Physica A 2001, 289, 18-36.
9. Nigmatullin, R.R. Physica A, in press.
10. Root, D. E., Hall, J. W. Meas. Control 1997, 181, 115-117.
11. Fischer, D., Eichhorn, K. J. Analisis 1998, 26, 58-61.
12. Heise, H.M., Bittner, A., Marbach, R. Clin. Chem. Lab. Med., 2000, 38, 137-145.
13. Sharaf, M. A., Illman, D. L., Kowalski, B. R., Chemometrics, John Wiley & Sons: NY, Chichester, Brisbane, Toronto, Singapore, 1986.
14. Johnson, N. L., Leone, F. C., Statistics and Experimental Design, vol.1, 2nd ed; John Wiley & Sons: New York, London, Sidney, Toronto, 1977.