

# Age-Related Changes in Probability Density Function of Pairwise Euclidean Distances between Multichannel Human EEG Signals

Mikhail Trifonov\*, Vladimir Rozhkov

Sechenov Institute of Evolutionary Physiology and Biochemistry of the Russian Academy of Sciences,  
Saint-Petersburg, Russia  
Email: \*[mtrifonov@mail.ru](mailto:mtrifonov@mail.ru), [vlrozhkov@mail.ru](mailto:vlrozhkov@mail.ru)

Received April 2014

---

## Abstract

The probability density functions (pdf's) and the first order structure functions (SF's) of the pairwise Euclidean distances between scaled multichannel human EEG signals at different time lags under hypoxia and in resting state at different ages are estimated. It is found that the hyper gamma distribution is a good fit for the empirically derived pdf in all cases. It means that only two parameters (sample mean of EEG Euclidean distances at a given time lag and relevant coefficient of variation) may be used in the approximate classification of empirical pdf's. Both these parameters tend to increase in the first twenty years of life and tend to decrease as healthy adults getting older. Our findings indicate that such age-related dependence of these parameters looks like as age-related dependence of the total brain white matter volume. It is shown that 15 min hypoxia (8% oxygen in nitrogen) causes a significant (about 50%) decrease of the mean relative displacement EEG value that is typical for the rest state. In some sense the impact of the oxygen deficit looks like the subject getting older during short-term period.

## Keywords

EEG Development, Hypoxia, Probability Density Functions, Hyper Gamma Distribution

---

## 1. Introduction

The progress in information theory, nonlinear dynamics, deterministic chaos theory, and random fractal theory caused a wave of researches where the analysis of complexity EEG signals is done on the base on the using of various complexity measures derived from them. Since the foundations of these theories are fundamentally different one can get a variety of complexity measures concerning the same EEG process. Detailed examination of a number of such measures given in [1] shows that their variations with time are either similar or reciprocal, but behaviors some of them are counter-intuitive and puzzling. The attempt to understand such behaviors is done in [1] through a new multiscale complexity measure of EEG. Despite of very promising findings in these studies

---

\*Corresponding author.

some of complexity measures adopted from nonlinear dynamics and chaos theory domain should be used with caution for EEG classification, since the brain isn't completely deterministic and the stochasticity may influence its function in some cases [2]. This means that it is reasonable just now to avoid any speculations about what types of deterministic and/or stochastic processes govern the EEG signals and to use some characteristics of these signals that are insensitive to any process underlying them. The one-dimensional probability density functions (pdf's) of the human EEG relative displacements may be used as one of such characteristics. It is not sufficient to infer the EEG dynamics but it is enough to capture some of its features.

## 2. Methods

### 2.1. EEG Data Collection

The eye closed resting state EEG data were recorded from 46 healthy subjects (12 adults and 34 children of school-age and preschool-age) with 16 Ag/AgCl scalp electrodes placed according to the international 10-20 system over both hemispheres at a sampling frequencies  $f_s$  of 185 Hz and 250 Hz. Three adults were also engaged in night time EEG recording and one was involved in 15 minutes EEG recording under hypoxia (8% oxygen in nitrogen). All school-age children took part in a longitudinal study of the EEG activity started at age 8.8 - 11.5 in 2005 and ended at age of 16.3 - 17.4 in 2011. The artifact free epochs selected for our analysis vary in length from a several seconds to one minute.

### 2.2. EEG Data Analysis

Let all EEG data samples be represented as  $m$ -dimensional vectors ( $m$ -dimensional time series)

$$X(t) = \{X_1(t), X_2(t), \dots, X_m(t)\}^T$$

where  $m = 16$  is a number of channels (electrodes),  $X_j(t)$  is the signal amplitude on the channel  $j$  at the discrete integer valued time moment  $t = 1, 2, \dots, N$ ,  $N$  is the number of samples received, and the superscript  $T$  denotes the matrix transpose operation. Since signals reveal significant spread of amplitude values from subject to subject and for different sleep stages within a single sleep recording the original EEG data was centered by subtracting their mean in every channel first and then scaled by the  $[\det(\mathbf{R})]^{(1/2m)}$ , where  $\det$  denotes determinant,  $\mathbf{R} = E[\delta\mathbf{X}\delta\mathbf{X}^T]$  is the sampling covariance matrix,  $\delta\mathbf{X} = \mathbf{X} - E[\mathbf{X}]$ , and  $E$  denotes statistical expectation. As the result any new vector  $\mathbf{Y}(t) = \delta\mathbf{X}(t)/[\det(\mathbf{R})]^{(1/2m)}$  has the same generalized variance independently on the subject since the determinant of the covariance matrix  $\mathbf{\Sigma} = E[\delta\mathbf{Y}\delta\mathbf{Y}^T]$  is equal here to 1. Geometrically the quantity  $[\det(\mathbf{\Sigma})]^{1/2}$  determines the volume of the confidence ellipsoid for any particular confidence level and the scaling proposed here makes the distributions of any vectors  $\mathbf{Y}$  to be equivalent in the sense that they occupy the same hypervolume in the  $m$ -space. It means that the ellipsoids with different orientations and different semi-axes but having the same generalized variance will be considered here as equivalent.

The vector sequence  $(\mathbf{Y}(1), \mathbf{Y}(2), \dots, \mathbf{Y}(N))$  specifies a personal EEG trajectory in initial  $m$ -dimensional space. Our aim is to estimate the probability density functions (pdf's) of the EEG relative displacements  $\Delta Y_\tau$  (the Euclidean distances between scaled multichannel human EEG signals pairs at given time lag  $\tau$ ) along this trajectory, defined by

$$\Delta Y_\tau = |\mathbf{Y}(t+\tau) - \mathbf{Y}(t)| = \left[ \sum_{j=1}^m (Y_j(t+\tau) - Y_j(t))^2 \right]^{1/2} \quad (1)$$

The idea to use such pdf's for describing and discriminating EEG patterns was inspired by the approach [3] proposing to reduce the shape matching problem to the comparison of probability distributions of the distance between two random points on a surface provide a robust method for discriminating between some classes of 3D objects. Since the relative displacement  $\Delta Y_\tau$  value, being  $m$ -dimensional Euclidean distance, doesn't depend on how the axes of the initial space are chosen and how many of principal components exhaust a given amount of the total variance there is no need to care about the proper state space reconstruction for pdf's deriving. But unlike artificial objects having rather stable shapes, the EEG trajectory forms a "living" shape that may evolve in time, *i.e.* may change its geometrical characteristics in dependence on epoch selection during a given EEG recording. Our aim is not only to analyze the empirically derived pdf's but also to find some types of theoretical distributions that can fit them. Here we will follow the general assumption borrowed from [4] "that if the me-

chanism (experiment) to generate the samples is the same, then the distribution type that describes the datasets will also be the same". In our case it means that we need to identify a single type of theoretical distribution that can fit the different datasets of  $\{\Delta Y_\tau\}$  by altering its parameters. Most likely the best candidate for such distribution may be a hyper gamma distribution proposed in [5] that is written here as

$$f(\Delta Y_\tau) = \gamma \cdot \Delta Y_\tau^{\nu-1} \exp(-\beta \Delta Y_\tau^\alpha) \quad (2)$$

where  $\gamma = \alpha \beta^{\nu/\alpha} \Gamma(\nu/\alpha)$ ,  $\Gamma(\cdot)$  is the gamma function,  $\alpha$  and  $\nu$  are shape parameters, and  $\beta$  is the scale parameter. It follows from the fact that the "random EEG" relative displacement  $\Delta Y_\tau$  (obtained from  $m$ -dimensional vector  $\mathbf{X}(t)$  whose components are independent normal random variables having zero mean and variance  $\sigma^2$ ) is distributed according to the chi distribution that is a special case of (2) when  $\nu = m$ ,  $\alpha = 2$ , and  $\beta = 0.25/\sigma^2$ .

The pdf's  $f(\Delta Y_\tau)$  are obtained in the standard way by fitting an empirical probability functions  $f_0(\Delta Y_\tau)$  represented by normalized histograms. The histogram bin size  $W_\tau$  is selected here as the half-sum of the  $W_{1\tau} = 2(IQR_\tau)(N - \tau)^{-1/3}$  and  $W_{2\tau} = 3.73\sigma_\tau(N - \tau)^{-1/3}$  calculated according to [6] and [7] suggestions respectively, where  $IQR_\tau$  is the interquartile range defined as the 75th percentile minus the 25th percentile, and  $\sigma_\tau$  is the standard deviation of the distribution. The number of bins  $h_\tau$  is defined as  $\text{entire}(\max(\Delta Y_\tau)/W_\tau) + 1$ .

The fitting procedure is based here on the estimation of parameters  $\nu$ ,  $\alpha$ , and  $\beta$  by the moment method. Since any covariance matrix  $\Sigma$  for our EEG data under analysis has full column rank, we set  $\nu = m$ . Then the estimates of  $\alpha$  is defined by the squared coefficient of variation ( $CV_\tau$ ) according to the following equation

$$(\sigma_\tau/E[\Delta Y_\tau])^2 = \Gamma((m+3)/\alpha)\Gamma((m+1)/\alpha)/[\Gamma((m+2)/\alpha)]^2 - 1 \quad (3)$$

Getting the estimate of  $\alpha$  using the Equation (3) one can obtain the estimate of  $\beta$  according to the following relation

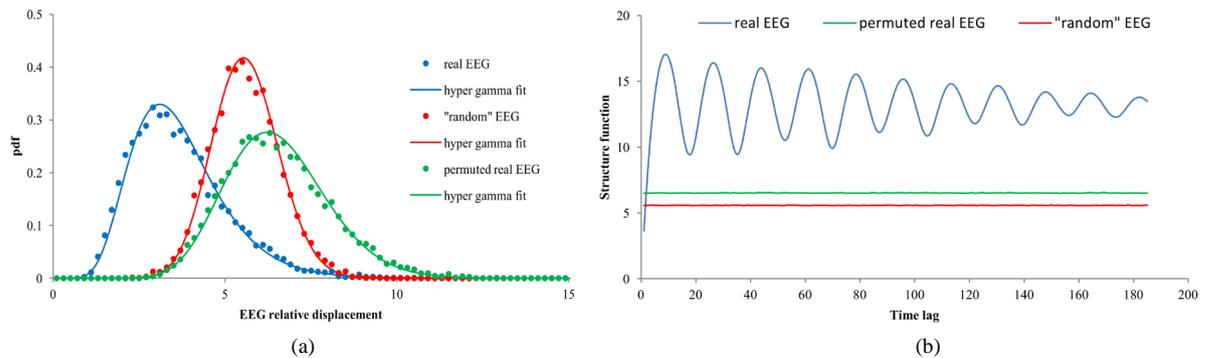
$$\beta = [\Gamma((m+2)/\alpha)]^\alpha / [E[\Delta Y_\tau] \cdot \Gamma((m+1)/\alpha)]^\alpha \quad (4)$$

The fitting procedure is limited here by the fitting range ( $FR_\tau$ ) endpoints that are defined as 1st percentile and the 99th percentile respectively. The criterion of fitting quality evaluation is based on the relative error of the first three moments and the percentage root mean square difference ( $PRD$ ) suggested in [8]

$$PRD_\tau = 100\% \cdot \sqrt{\sum_{k=1}^{h_\tau} (f_0(\Delta Y_{\tau,k}) - f(\Delta Y_{\tau,k}))^2} / \sum_{k=1}^{h_\tau} (f_0(\Delta Y_{\tau,k}))^2$$

### 3. Results

Examples of the empirical pdf  $f_0(\Delta Y_\tau)$  and its hyper gamma fit  $f(\Delta Y_\tau)$  for one of the subjects are shown in **Figure 1(a)**. In this case the value of  $N$  is equal to 8060 that corresponds to about 43.6 second EEG recording with the sampling frequency  $f_s = 185$  Hz. Time lag  $\tau = 1$  corresponds to time interval of 0.0054 sec and  $W_1 = 0.2$ . For comparison the pdf's for the "random" EEG with  $\sigma_n = 1$  and for the original EEG with randomly permuted data samples within each channel are shown on the same **Figure 1(a)**. Visually the single type of theoretical  $f(\Delta Y_\tau)$



**Figure 1.** (a) The examples of the empirical pdf's  $f_0(\Delta Y_\tau)$  and hyper gamma fits  $f(\Delta Y_\tau)$  for the real (blue), real, but randomly permuted (green), and "random EEG" data (red) and (b) relevant structure functions  $SF_\tau$ .

provides a reasonably good fit for all three empirical  $f_0(\Delta Y_1)$ . The maximum value of  $PRD_1$  here refers to the case of the original EEG and is less than 8.8% while the maximum relative error in the first three moments does not exceed 0.5%.

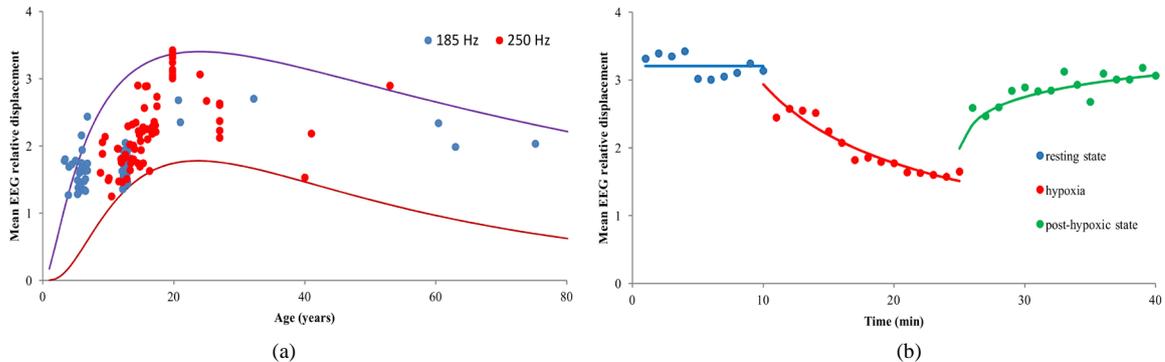
In general there is a noticeable inter-individual variation in  $PRD_\tau$  value even for the same sample size  $N$  but in average  $PRD_\tau$  decreases proportional to  $N^{-0.17}$  as  $N$  increases. As usual the relatively high  $PRD_\tau$  values occur when the empirical pdf's  $f_0(\Delta Y_\tau)$  have a rather long tails. It might mean that they can represent a mixture of hyper gamma distributions with different time scales. Since the mean  $PRD_\tau$  value here is about 9.1% and visually any empirical pdf is fitted well enough with the single hyper gamma pdf, we restrict our attention to this case. According to Equations (3) and (4) it is enough to analyze here only two parameters- $E[\Delta Y_\tau]$  and  $\sigma_\tau/E[\Delta Y_\tau]$ .

The first one defines the individual first order structure function ( $SF_\tau$ ) that rapid increases for small  $\tau$  as  $\tau^\delta$  ( $\delta < 1$ ) leading to shifting the  $f_0(\Delta Y_\tau)$  from  $f_0(\Delta Y_1)$  to the right. The  $CV_\tau$  also reveals the individual power law dependence on  $\tau$  for small  $\tau$ . But starting with some  $\tau^*$  value the  $SF_\tau$  begins to oscillate around its sill level defined by the variance of stochastic component of  $\Delta Y_\tau$ . It causes the oscillatory behavior of  $f_0(\Delta Y_\tau)$  that depends on the subject's individuality and is most conspicuous in the resting state. It was found that for all our subjects in this state the oscillation frequency lies in the alpha range (see **Figure 1(b)**). For comparison, both  $SF_\tau$  and  $CV_\tau$  corresponding to the "random" EEG do not depend on  $\tau$  and for  $\sigma_n = 1$  are equal to 5.57 and 0.99 respectively. The  $SF_\tau$  and  $CV_\tau$  for the original EEG with randomly permuted data samples within each channel do not depend on  $\tau$  as well.

The age dependence of  $E[\Delta Y_1]$  in the eye-closed resting state is shown in **Figure 2(a)**. The original  $E[\Delta Y_1]$  values getting at  $f_s = 185$  Hz were multiplied by 0.74 to approximately adopt them to  $f_s = 250$  Hz. The blue dots show such adopted data representing in fact the upper bounds for  $E[\Delta Y_1]$  that actually could be in this case. The presented data reflect high variability of  $E[\Delta Y_1]$  both within some subjects and between different subjects. Nevertheless there is some tendency of  $E[\Delta Y_1]$  increasing in the first twenty years of life and of  $E[\Delta Y_1]$  decreasing as adults getting older. This tendency can be approximately described by log-normal curves in **Figure 2(a)** depicted by colored lines. We are going to define more exactly the type of curve fitting at a later time.

Having conducted multiple longitudinal studies, we find that the  $CV_1$  reveals noticeable inter-individual differences being rather stable within an individual subject over long period of time. The analysis of 10 min resting state, following 15 min hypoxia and 15 min after hypoxia study did not show significant variation in  $CV_1$  as well. The relative standard deviation of  $CV_1$  for this case is about 4.4%.

The time-specific dependence of  $E[\Delta Y_1]$  in 15 min hypoxia (red dots) was followed by 10 min resting state (blue dots) is shown in **Figure 2(b)**. The hypoxia causes a dramatic decreasing of  $E[\Delta Y_1]$  relative to the resting state. During the last minute of the hypoxia the initial  $E[\Delta Y_1]$  value is reduced by 50%. In some sense the impact of the oxygen deficit looks like as the subject get older during short-term period (see age-dependence of  $E[\Delta Y_1]$  on **Figure 2(a)**). During the 15 min period of time after the hypoxia  $E[\Delta Y_1]$  tends to reach the value inherent in the resting state. It was found that changes of  $E[\Delta Y_1]$  have a strong positive correlation ( $r = 0.89$ ) with changes in oxygen saturation.



**Figure 2.** (a) Age dependence of  $E[\Delta Y_1]$  in the eye-closed resting state and (b) time-specific dependence of  $E[\Delta Y_1]$  in 15 min hypoxia (red dots) was followed by 10 min resting state (blue dots) and 15 min post-hypoxic recovery (green dots). The different color lines on the panel (a) corresponding to log-normal curves and the green line and curve on the panel (b) are used for a fitting of the empirical data.

## 4. Conclusions and Future Work

The results of this study allow us to make a preliminary conclusion that one-dimensional pdf's of EEG relative displacements can be used for understanding of the real EEG dynamics in various functional states and different subjects. To a first approximation, in each case the empirically derived pdf are fitted quite well by the single hyper gamma distribution. It means that only two parameters (sample mean of EEG relative displacements and coefficient of variation) may be taken into account. Both these parameters exhibit subject's individuality. The first one reveals age and state dependence while the second one stays rather stable for a given subject over long period of time except sleep stages. It is interesting to note that age-related dependence of  $E[\Delta Y_j]$  looks like as age-related dependence of the total brain white matter volume given in [9]. In addition the non-linear age effect on  $E[\Delta Y_j]$  is consistent with the suggestion that during late childhood period there is a shift of topological organization of brain white matter toward a more randomized configuration [10].

In our future research we are going to analyze much more EEG records to investigate in details the age and time-specific dependence of the parameters mentioned above. The next interesting aspect of our future work is the improvement of the fitting quality of the empirically derived pdf of EEG relative displacements by using distribution that consists of a mixture of hyper gamma distribution.

## Acknowledgements

The authors would like to express their sincere thanks to Prof. M. N. Tsitseroshin, Dr. E. A. Panasevich, and Dr. S. S. Bekshaev for providing EEG data.

## References

- [1] Gao, J., Hu, J. and Tung, W. (2011) Complexity Measures of Brain Wave Dynamics. *Cognitive Neurodynamics*, **5**, 171-182. <http://dx.doi.org/10.1007/s11571-011-9151-3>
- [2] Masquelier, T. (2013) Neural Variability, or Lack There of. *Frontiers in Computational Neuroscience*, **7**, 1-7. <http://dx.doi.org/10.3389/fncom.2013.00007>
- [3] Osada, R., Funkhouser, T., Chazelle, B. and Dobkin, D. (2002) Shape Distributions. *ACM Transactions on Graphics*, **21**, 807-832. <http://dx.doi.org/10.1145/571647.571648>
- [4] Nikolova, N.D., Toneva-Zheynova, D., Kolev, K. and Tenekedjiev, K. (2013) Monte Carlo Statistical Tests for Identity of Theoretical and Empirical Distributions of Experimental Data. In: Chan, V., Ed., *Theory and Applications of Monte Carlo Simulations*, InTech, 1-26. <http://dx.doi.org/10.5772/53049>
- [5] Suzuki, E. (1964) Hyper Gamma Distribution and Its Fitting to Rainfall Data. *Papers in Meteorology and Geophysics*, **15**, 31-51. [http://www.mri-jma.go.jp/Publish/Papers/DATA/VOL\\_15/15\\_031.pdf](http://www.mri-jma.go.jp/Publish/Papers/DATA/VOL_15/15_031.pdf)
- [6] Izenman, A.J. (1991) Recent Developments in Nonparametric Density Estimation. *Journal of the American Statistical Association*, **86**, 205-224. <http://www.jstor.org/stable/2289732>
- [7] Scott, D.W. (2004) Multivariate Density Estimation and Visualization. Papers/Humboldt-Universität Berlin, Center for Applied Statistics and Economics (CASE), No. 2004, 16. <http://hdl.handle.net/10419/22190>
- [8] Alfaouri, M., Daqrouq, K., Abu-Isbeih, I.N., Khalaf, E.F., Al-Qawasmi, A-R. and Al-Sawalmeh, W. (2009) Quality Evaluation of Reconstructed Biological Signals. *American Journal of Applied Science*, **5**, 187-193. <http://dx.doi.org/10.3844/ajas.2009.187.193>
- [9] Sowell, E.R., Peterson, B.S., Thompson, P.M., Welcome, S.E., Henkenius, A.L. and Toga, A.W. (2003) Mapping Cortical Change across the Human Life Span. *Nature Neuroscience*, **6**, 309-315. <http://dx.doi.org/10.1038/nn1008>
- [10] Chen, Z., Liu, M., Gross, D.W. and Beaulieu, C. (2013) Graph Theoretical Analysis of Developmental Patterns of the White Matter Network. *Frontiers in Human Neuroscience*, **7**, 1-13. <http://dx.doi.org/10.3389/fnhum.2013.00716>