doi:10.1017/S1431927617001180

Microsc. Microanal. 23 (Suppl 1), 2017

An Information Theoretic Approach for Creating 3D Spatial Images from 4D Time Series Data

Willy Wriggers[1], Julio Kovacs[1], Federica Castellani[2], P. Thomas Vernier[2], and Dean J. Krusienski[3]

[1] Dept. of Mechanical & Aerospace Engineering and Institute of Biomedical Engineering, Old Dominion University, Norfolk, VA 23452, USA

[2] Frank Reidy Center for Bioelectrics and Institute of Biomedical Engineering, Old Dominion University, Norfolk, VA 23452, USA

[3] Dept. of Electrical & Computer Engineering and Institute of Biomedical Engineering, Old Dominion University, Norfolk, VA 23452, USA

We are developing statistical analytics for generating spatial heat maps from the information matching of fast and slow degrees of freedom. The resulting multi-timescale heat maps can be used to transform highly resolved time domain data into spatial features. For example, in imaging applications, the transformation can extract 3D heat map images from 4D time series. We are currently applying this technology to the detection of hinges and contact breaking hot spots in the 3D structure of unfolding proteins [1], to the detection of membrane-weakening lipids in electroporation simulations (Figure 1), and to the mapping of spoken language production and perception patterns in the human brain (Figure 2). Since the method is of general applicability, we expect it to be useful in multi-modal and big data microscopy applications where time-dependent signals should be transformed to 2D and 3D images.

The general goal of our analytics is to discover a statistical dependence between fast/local and slow/global processes in the multichannel time series. The essential, slow processes of each channel i are characterized by a user-defined activity function a(t) that measures the global rate of change in the system (Figure 1), or by an external stimulus of functional importance such as an acoustic speech signal (Figure 2). In the example in Figure 1, the fast/local time series in each channel X_i(t) is ranked by a coefficient $R_{X,a}(i) = I\left(\left|\dot{X}_i\right|, a\right)$, where I is a statistical measure of dependence, such as the Pearson correlation coefficient. This ranking provides a measure of the relevance or contribution of each of the local processes X_i(t) to the global process a(t) by means of the statistical dependence of the fast and slow rates of change. The Pearson cross-correlation (Figure 1) detects a signed linear statistical dependence and is efficient to compute, but it does not differentiate slopes or nonlinear structure in the joint distribution $p(\left|\dot{X}_i\right|, a)$. Mutual information, on the other hand, measures how similar the joint distribution is to the products of factored marginal distribution. It is a positive measure in bit units and captures nonlinear dependence, which in our tests has provided significant performance benefits over Pearson cross-correlation [1]. Mutual information (Figure 2) is usually expensive to compute, but we recently developed a fast and accurate Balanced Adaptive Density Estimation (BADE) that will be distributed with our *TimeScapes* analytics package [2].

Our BADE approach solves the general problem of how to efficiently estimate the probability density function of very unevenly sampled random variables arising in our mutual information calculations. BADE effectively optimizes the amount of smoothing at each point by means of an efficient nearest-neighbor search that results in good scaling for large data sizes. Our tests on simulated data show that BADE exhibits equal or better accuracy than existing methods that involve kernel functions, and the results are also aesthetically pleasing. In our software, BADE will replace the older fixed-bandwidth solver originally proposed in [1].

Applications of BADE and the heat map analysis at our institution are focused on molecular dynamics analysis and brain mapping (Figures 1 and 2). We will also discuss possible future applications of interest in the video processing of microscopy data [3].

References:

[1] J Kovacs and W Wriggers, J. Phys. Chem. B **120** (2016), p. 8473
[2] http://timescapes.biomachina.org
[3] The authors acknowledge funding from the Batten Endowment at ODU and NIH R01GM62968 (to W.W.), Air Force Office 311 of Scientific Research (FA9550-15-1-0517 and FA9550-14-1-0123 to P.T.V.), National Science Foundation 1451028 and 1608140 (to D.J.K.)

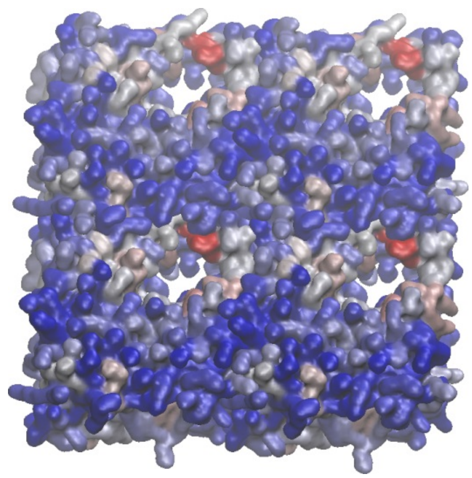

Figure 1. The heat map analysis of lipid interaction networks in molecular dynamics simulations of electroporation (under periodic boundary conditions) shows that the disruption of the bilayer, as measured by the breaking activity of lipid contacts, is associated with the externally imposed pore formation. Shown is the Pearson cross-correlation of the contact distance geometry with the contact breaking activity (see [1]).

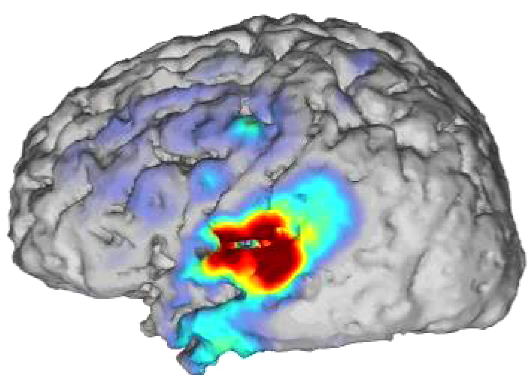

Figure 2. Brain map of spoken language processes. The heat map locates areas of speech production and perception on the brain of an epilepsy patient who had an intracranial EEG electrode array implanted on the cortex. The intracranial EEG recording was compared to the external acoustic speech signal by mutual information analysis with BADE (see text).

https://doi.org/10.1017/S1431927617001180