Studying Embedded Human EEG Dynamics Using
Generative Topographic Mapping
El-Deredy, W. and Lisboa, P.J.G. and Vellido, A.
Research Report LSI-04-8-R

Departament de Llenguatges i Sistemes Informàtics

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Studying Embedded Human EEG Dynamics Using Generative Topographic Mapping

Vellido, A.[1*], El-Deredy, W.[2], and Lisboa, P.J.G.[2]

[1] *Departamento de Lenguajes y Sistemas Informáticos. Universidad Politécnica de Cataluña.C. Jordi Girona, 1-3. 08034, Barcelona, Spain*

[2] *School of Computing and Mathematical Sciences, Liverpool John Moores University. Byrom St. L3 3AF, Liverpool, UK*

Abstract

A method has recently been proposed [1] to extract multiple signal source information from single-channel electroencephalogram (EEG) recordings. A dynamical systems approach is used to analyze the resulting EEG time series, and its dynamics are captured by the transformation of the original data into an *embedding matrix* residing in a Euclidean *embedding space*. Measurements in [1] are taken to be of ongoing unbounded EEG recordings. Many experiments concerning the study of cognitive tasks, though, are developed in a multi-subject repetitive setting where time-boundaries are defined in relation to the onset time of certain stimuli. Each repetition of an experiment is known as a trial and, although the experimental setting might induce to expect little variability amongst responses, the reality usually yields high inter-trial and inter-subject variability. Pooling all responses may mislead their interpretation. In this paper we resort to the Generative Topographic Mapping (GTM, [2]), a neural-network inspired but statistically principled unsupervised model, to achieve the following goals: First, the definition of groups of trials with intra-group similarities and inter-group differences in order to improve the interpretability of the results in the aforementioned experimental settings; second, the visualization of embedded EEG dynamics in a 2-dimensional latent space; finally, the study of the trajectories of these EEG dynamics over the GTM latent space representation, showing that transitions and stationary states in these trajectories correspond to special features in the time-power and time-frequency representations of the EEG data.

Keywords: Electroencephalogram (EEG); Generative Topographic Mapping (GTM); Dynamical Embedding (DE); Dynamical trajectories; Time-frequency representation of EEG signals.

---

[*] *Corresponding autor.* avellido@lsi.upc.es ; *URL:* www.lsi.upc.es/~avellido

1. Introduction

The Self-Organizing Map (SOM, [3,4]) is the most widely successful and frequently used neural network unsupervised model, inspiring thousands of studies over the last fifteen-odd years[1]. It was, by conception, an attempt to model certain brain processing features ([3], [5]). Over the years, it has veered, as a model, towards more conventional problems such as clustering and visualization. Only in the last few years it has been used to peer back from these knowledge niches to cognitive problems, as in reinforcement learning ([6], [7]), language brain processing [8], visual cortex activity ([9],[10]), spiking neurons analysis ([11],[12]), or temporal cortex categorization [13]. Despite these examples, neural network, or neural network-inspired unsupervised models have not regularly been applied to problems in the area of neurosciences.

To the best of our knowledge, there have been to date no applications of the GTM in the field of neurosciences. This is a model somehow inspired by the SOM and, while the data visualization possibilities of both models are akin, it has some advantageous features, all of them stemming from its probabilistic formulation: when probability theory lays at the foundation of a learning algorithm, the risk that the reasoning performed in it be inconsistent in some cases is lessened ([14],[15]). The main advantage is that the GTM generates a density distribution in data space in such a way that the model can be described and developed within a principled probabilistic framework in which all the modelling assumptions are made explicit. The GTM also provides a well-defined objective function (something that the SOM lacks) and its optimisation, using either non-linear standard techniques or the EM-algorithm, has been proved to converge. As part of this process, the calculation of the GTM learning parameters is grounded in a sound theoretical basis. Bayesian theory can be used in the GTM to calculate a posterior probability of each point in latent space being responsible for each point in data space, instead of the SOM sharp map unit membership attribution for each data point.

There are several non-invasive methods for the analysis of brain signals. Most research on this area has resorted to functional Magnetic Resonance Imaging (fMRI) and EEG. The former is most suited to achieve spatial resolution and signal source localization, but is very limited when it comes to explore the signal time-course. On the contrary, EEG recordings have a fine time resolution, but the problem of signal source localization is a hard and still open one. A novel method has recently been proposed [1] to extract multiple signal source information from single-channel EEG recordings. This is quite a departure from existing methods that make use of rather complex experimental arrangements of over a hundred

[1] See http://www.cis.hut.fi/research/refs/ for a very complete index of references up to 2000.

recording electrodes (channels). It resorts to a dynamical systems approach to the analysis of sampled time-series, which states that, "given a time-series in a n-dimensional dynamical system, delay-coordinate embedding lets one reconstruct a useful version of the internal dynamics of that system" [16]: the dynamics of the EEG are shown to be captured by means of a transformation of the original EEG data into an *embedding matrix* made up of overlapping successive intervals of the data (*delay vectors*) residing in a Euclidean *embedding space*. The illustration of this method in [1] uses ongoing EEG recordings, but many experiments concerning the study of cognitive tasks are designed to encompass multiple trials and involve multiple subjects. A variation on the *dynamical embedding* matrix is used in the present study to analyse the dynamics of this type of experimental settings, using the GTM.

The second section of this paper provides a brief summary of the dynamic embedding procedure. This is followed by a proposal to extend this to multi-trial experimental settings, suggesting the definition of homogeneous sub-groups of trials to improve the interpretability of the results in the aforementioned settings. Section 4 attempts to succinctly describe the GTM and its recent developments, and its application to the clustering and visualization of EEG dynamics is proposed in section 5. Section 6 compiles the results of applying these theories in a specific human cognitive problem and a discussion of these results follows.

2. Embedded dynamics of ongoing EEG

Different techniques are used to gauge brain activity. These can be roughly classified into two groups: invasive (intracraneal recordings, mostly used in non-human brains) and non-invasive techniques. Amongst the latter fMRI and Positron Emission Tomography (PET), and also EEG which resorts to measurements of electric potentials in the scalp with a setting of electrodes. EEG provides the means to explore brain dynamics through its fine time-resolution capability.

Unfortunately, time-resolution is not accompanied by resolution in space: signal source localization is a classic and still open problem in the EEG literature. Recently, James and Lowe [1] provided a method for source localization with the rather unique feature that it resorts to single-electrode (channel) recordings. Based on a dynamical systems approach to the analysis of the underlying signal generators, it makes use of the technique of dynamical embedding (DE) and assumes that the signal is due to the non-linear interaction of a few degrees of freedom with added noise. A DE matrix, built by concatenation of successive, overlapping fragments (or delay vectors) of the original EEG signal over time, and residing in a Euclidean embedding space, is shown to "reconstruct" the unknown dynamical signal generator. Further details of the DE method can be found in ([1],[17]).

In the examples of the application of DE methods for source localization of EEG signals in [1], unbounded ongoing EEG recordings are used. Although useful in many practical instances, it leaves aside many experimental settings characterised by a design in which several individuals are invited to perform repeatedly certain precisely defined tasks over a number of trials. Bounded intervals of EEG signal are then isolated for analysis, usually encompassing pre-stimulus and post-stimulus periods.

The present study aims to modify this DE approach to be used in such type of experimental settings. Consequently, an alternative way to design the DE matrix is required; it will be described in section 6.1. As mentioned above, there are also different goals for its analysis in this study, which are summarised next.

3. From ongoing EEG to multiple-trial, multiple-subject dynamics analysis

Even in very narrowly defined and scrupulously planned EEG measurement experiments involving human high-level brain processing tasks, the presence of variability in the results between subjects and within-subject different trials seems unavoidable ([18],[19],[20]). Some standard techniques, such as Event Related Potentials (ERPs) resort to straight averaging. This has the undesired effect of cancelling out any observations that are not stimuli-locked, including most induced signals that are the result of top-down high-level brain processing. A number of techniques that attempt to circumvent this shortcoming have been developed, for instance those based on simultaneous visualization of multiple trials such as the *ERP Image* [16], or those based on wavelet transformations of the original EEG signal resulting on time-frequency plots ([21], [22]).

These methods still involve the use of all trials on an equal basis. Similar brain processes might be expected to entail similar measured brain activities. Nevertheless, trial variability might be the result of actually different ongoing processes happening during the execution of the experimental tasks. If that is the case we might expect to observe different *subgroups of dynamics*, similar within subgroup but dissimilar amongst subgroups. The final analysis of the responses to these experimental tasks might benefit from the isolation of trials belonging to certain subgroups.

The use of DE techniques, with the modifications in the construction of the DE matrix described in section 6, will help in this work to study several issues: The visualization of the trajectories of those dynamics represented by the delay vectors over a latent, 2-dimensional visualization space; as a result, different types of dynamics, such as transitions and stationary states, might be isolated and associated to magnitude and frequency patterns in the original signal. Equally, experimental trials might be grouped according to their particular dynamic

pattern. Also, fragments of each trial or even complete trails might be singled out as outliers according to such dynamic patterns.

4. The GTM as a tool for the visualization and analysis of EEG dynamics

According to the dynamical systems view put forward in [1], brain inner sources of the signal, captured in the scalp through EEG recordings, generate data that lie in an unobservable manifold that can be represented by the DE matrix and its corresponding Euclidean embedding. We attempt to visualize this unobservable manifold and the trajectories described by its delay vectors by using the latent two-dimensional space of the GTM that *generates* it.

4.1. Introduction to the GTM

The GTM is a latent variable model introduced by Svensen [2] as probabilistic formulation of Kohonen's SOM ([3],[4]). GTM goes beyond projection (from high-dimensional to low-dimensional spaces) methods as it generates a non-linear and topographically preserving mapping from a low-dimensional visualization latent space onto the high-dimensional space in which the data reside. The mapping is carried through by an intermediate set of basis functions generating a mixture density distribution in the data space. An objective function can be defined for the GTM, whose optimization using either standard non-linear techniques or the EM-algorithm, has been shown to converge.

In short, GTM describes a non-linear low-dimensionality latent variable model that generates a probability density in the multi-dimensional data space through a linear mixture of basis functions defined as:

$$\mathbf{y} = \mathbf{W}\phi(\mathbf{u}) \,, \tag{1}$$

where $\mathbf{y}$ is a point in data space, $\mathbf{u}$ is an L-dimensional point in latent space, $\mathbf{W}$ is the matrix that generates the explicit mapping from latent space to an L-dimensional manifold embedded in data space (in the Euclidean embedding space defined by the DE matrix, in this case), and ϕ is a set of R basis functions. The prior distribution of $\mathbf{u}$ in latent space is usually constrained to form a uniform discrete grid of points, units or nodes, analogous to the layout of the SOM, in the form:

$$p(\mathbf{u}) = \frac{1}{M}\sum_{i=1}^{M}\delta(\mathbf{u} - \mathbf{u}_i), \tag{2}$$

where M is the number of nodes.

The log-likelihood of fitting the data from the latent variables is given by

$$LL(\mathbf{W},\beta) = \sum_{n=1}^{N} \ln p\left(\mathbf{x}^n \mid \mathbf{W},\beta\right) \tag{3}$$

which is empirically estimated by summing over the input data set $\{\mathbf{x}^n\}$. The optimisation of the objective function entails finding optimal values for $\mathbf{W}$ and β, which can be accomplished using the Expectation-Maximization (EM) algorithm [23]. Details of this calculation can be found in [2].

4.2. Automatically regularized GTM

An advantage of the probabilistic setting of the GTM is the possibility of introducing regularization in the mapping from latent to data space. This procedure automatically regulates the level of map smoothing necessary to avoid overfitting the data, which are assumed to be the product of underlying generators and likely to contain noise.

The use of a single regularization term entails a homogeneous smoothing of the mapping and can be formulated through the modification of the objective function (3) in the form of a *penalized* log-likelihood:

$$LL_{PEN}(\mathbf{W},\beta) = \sum_{n=1}^{N} \ln p\left(\mathbf{x}^n \mid \mathbf{W},\beta\right) - \frac{1}{2}\alpha\|\mathbf{w}\|^2 \tag{4}$$

where the regularization coefficient α is now an adaptive parameter that can be estimated using the *evidence approximation* [24]. For details on this procedure, see [25].

4.2.1. Selective Mapping Smoothing

Alternatively, the smoothing of the mapping can be adapted locally. The complexity of the mapping is mostly controlled by the number and form of the intermediate basis functions. One regularization term can this time be associated to each of them, so that eq.(4) becomes:

$$LL_{PEN}(\mathbf{W},\beta) = \sum_{n=1}^{N} \ln p\left(\mathbf{x}^n \mid \mathbf{W},\beta\right) - \sum_{r=1}^{R} \frac{1}{2}\alpha_r \|\mathbf{w}_r\|^2 \tag{5}$$

where α_r is the regularization coefficient associated to basis function r, and $\mathbf{w}_r$ is the vector of weights from $\mathbf{W}$ corresponding to α_r. The multiple adaptive regularization parameters $\{\alpha_r\}$ can be estimated using a variation of the aforementioned *evidence approximation*. This procedure, named *Selective Mapping Smoothing* (SMS) and inspired in the technique of *Automatic Relevance Determination* (ARD, [26]) is described in more detail in [27].

The selective smoothing accomplished by SMS should provide mappings of optimum complexity for a given common width of the intermediate basis functions. Zones of data space requiring strong smoothing will have the activity of their corresponding basis functions diminished through the reduction of the values of their associated weights caused by high regularization coefficients' values. On the contrary, zones that do not require smoothing will keep their corresponding basis functions active.

5. Mapping the EEG dynamics into two dimensions

This section develops several new ideas concerning the representation of the DE matrix, and therefore the representation of the EEG dynamics, on the latent space for visualization of the GTM. It includes hypothesis on the clustering of embedded dynamics, trajectories over the latent space, and the relation of these features with time-power and time-frequency-power representations of the EEG signal.

5.1. Clustering of embedded dynamics

Each trial in an experiment corresponds to one performance of a scheduled task. As previously pointed out, even for narrowly defined tasks the results stubbornly show high within-subject inter-trial variability and, indeed, high variability between subjects. Despite this overall variability, we should also expect to find regularities between trials. In other words, we might expect subgroups of trials to exist, showing a degree of homogeneity that also discriminated them from other subgroups. This is especially important for tasks that are likely to increase the variability phenomenon by their own nature: figure 1 illustrates one of these cases, where variability is enhanced by the unexpectedness of the stimuli involved.

Assuming that the hypothesis that distinct areas of the GTM two-dimensional latent space correspond to different types of dynamics of the EEG holds true, we would expect that the projections of the delay vectors corresponding to the same intervals across trials fell into a reduced number of map units that resided in a reasonably limited neighbourhood. But this can only expected to be true provided there is homogeneity across the set of trials of the experiment. Therefore, should the results show that different sub-groups of trials exist, the aforementioned clustering of equivalent delay vectors across trials should be much more compact for each sub-group than for the whole set of trials.

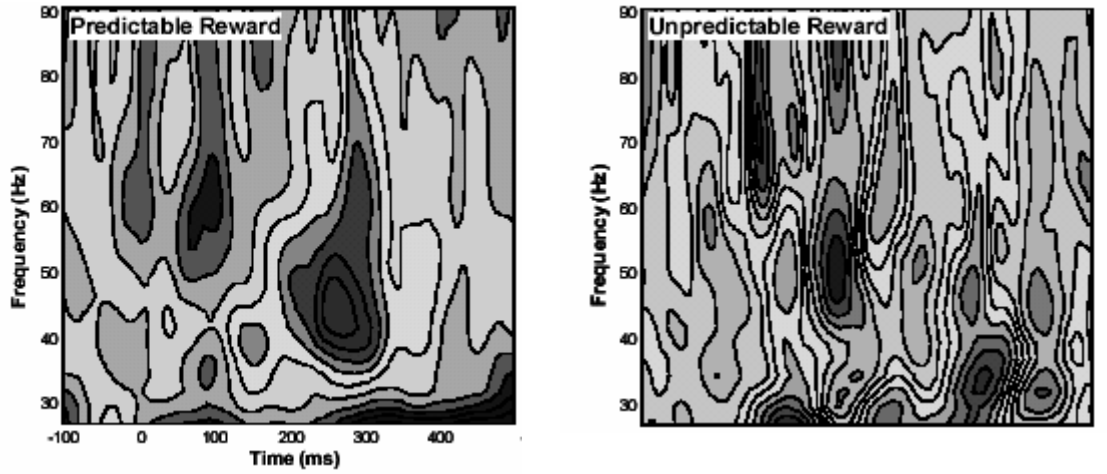

Figure 1: Time-frequency plots of signal power, obtained using the procedure described in section 6.1, from an experimental task analysed in [28], showing power EEG measurements before and after a rewarding stimulus. In the left-hand plot, the reward was delivered in a predictable way, whereas in the right-hand plot the reward was delivered in an unexpected fashion. Although bursts of activity across the frequency range are visible in both cases, they are much more regular in the predictable experiment. Although not necessarily the case, this is likely to be the result of a much higher variability over trials in the unpredicted experiment due to the own nature of the experiment. Power values are colour-coded from high (positive or negative, dark) to low (near 0, light); the scale has been omitted.

Even if these "clusters of trials" exist, it would be interesting to find out whether some delay vectors show more homogeneity than others. This is akin to asking if homogeneity over trials is higher at certain times in the experiment. Should they exist: do they correspond to delay vectors of special interest for the experiment at hand?

5.2 Intra-trial dynamic trajectories

Following a dynamical systems view, it is suggested in [1] that the DE matrix represents the dynamics of the EEG signal in a Euclidean embedding space of a dimension corresponding to the length of the delay vectors. These delay vectors would trace a *trajectory* on the manifold generated by the Euclidean embedding. The dimensionality of this space, though, renders the visualization of these trajectories impossible.

The GTM provides a way around this problem through the visualization of the DE matrix on its latent space. The trajectory traced by the delay vectors in the discrete two-dimensional latent space of the GTM would be expected to show certain features: First, certain level of orientation manifested in the prevalence of latent space contiguity for successive delay vectors; second, the existence of stationary states in the form of successive delay vectors

being mapped in the same latent space units; third, the occurrence of neat transitions between areas of the latent space, corresponding to transitions in the dynamics of the EEG recordings.

Assuming that different trials might show different characteristic trajectories through the latent space map and that these might be organized into different subgroups, we now outline a procedure to group GTM nodes in order to compare the resulting clusters (or macro-clusters, if we accept that each GTM node can be considered a cluster by itself) with the trajectories traced by the trials. It is based in a *contiguity-constrained agglomerative* algorithm proposed in [29] for the SOM and its pseudo-code is shown in figure 2. In short, this procedure assumes that delay vectors that are close in the DE space are also close in their latent space representation (contiguity condition). The distances in the algorithm are taken to be Euclidean and there is no *a priori* selection of *cluster seeds*. The contiguity or neighbouring condition consists of only considering, as candidates to be merged, those (macro)clusters that contain neighbour nodes in the GTM latent space. For nodes neither in the edges, nor in the corners of the latent space, the eight surrounding nodes are considered neighbours.

There are of course diverse alternatives to this procedure. One of those might, for instance, use *cumulative responsibility* peaks [30] or maxima of mapping distortion (*magnification factor*, [31]) as (macro)cluster initial *seeds*. Comparing clustering algorithms is beyond the scope of this study, though. The number of final (macro)clusters defining the GTM latent space partition is also an open theoretical problem, but a subjective and parsimonious enough solution suffices here, given that we only aim to characterize trial trajectories over the map in an interpretable way.

```
Initialise each GTM node as a macrocluster seed
Intialize these macroclusters' centroids with their reference vectors.
Repeat until a given termination condition is met
       Merge the two closest (Euclidean distance) macroclusters
       Replace their centroids by their mean vector
End
```

figure 2: Pseudo-code of the contiguity-constrained agglomerative algorithm described in the text, adapted from [29].

5.3. Correspondence between cluster dynamics and EEG signal representations

The previous ideas are nothing but different aspects of a main core hypothesis, stating that different dynamics in the EEG signal, through the design of a DE matrix, will be mapped into distinct areas of the GTM latent space. Given that each of the delay vectors that shape the DE matrix is defined as a time-window in the signal, we can check the correspondences between the delay vector representations on the GTM discrete latent space and any EEG signal representation over time. Therefore, correspondences with the raw EEG signal can be explored, as well as correspondences with time-frequency displays of the signal.

6. Experiments

The ideas put forward in the previous section are now set to test using data from a specific experimental task, which is described first. Details of the GTM settings are also provided and several results are displayed and analysed.

6.1. Experimental task, EEG data and its processing

In order to study the ideas advanced in the previous sections, we will resort to data from a specific experimental task. The schema of this experimental paradigm is shown and described in figure 3. Twelve healthy right-handed subjects, ages 23 – 38 (mean age = 27), participated in the study. EEG was recorded continuously with an EGI (Electrical Geodesics Inc.) 129-electrode array. The vertex (position Cz) was used as reference. Sampling rate was 500 Hz, and all channels were processed on-line by means of a 0.1 to 200 Hz band-pass filter. Eye movements were monitored with a subset of the 128 electrodes. EEG was segmented into epochs containing data from 500 ms prior to stimulus onset to 1500 ms following it.

In this study, data from only one of the experimental conditions described in figure 3 are selected: that of *Unexpected Reward* (UR), given that its unpredictability condition should highlight inter-trial variability. Data from a single channel (Pz), corresponding to the parietal area are selected. This area and, particularly, this channel are known to be relevant for the experimental task at hand. There are 30 trials of data, each containing 1000 values / time points. All trials are concatenated to create the DE matrix. Two parameters are relevant to its construction: the lag (τ), controlling the amount of overlapping between consecutive delay vectors, and the number of lags (m) or *embedding dimension*, which controls the length of the delay vectors. In this study, we select their values to be $\tau = 5$ (10ms) and $m = 60$ (equivalent to intervals of 120ms), given that any relevant changes in the experiment at hand are likely to happen at such time-scale.

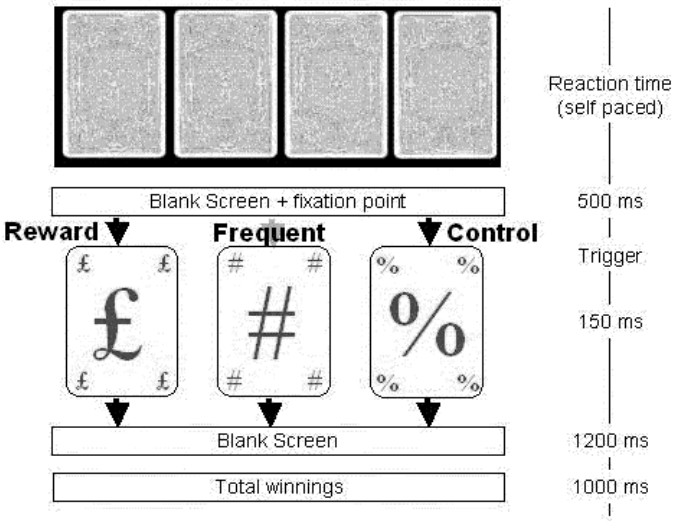

Figure 3: Schematic representation of the experimental task. On each trial, subjects selected one of four cards displayed on a computer screen positioned 1 meter in front of them. Subjects responded pressing one of 4 keys marked on a computer keyboard. On rewarded trials, a card with a **£** symbol appeared for 150 ms. Each **£** sign was worth £0.25, and subjects saw a running total of their earnings after each trial. In the Unexpected Reward condition (UR), subjects were unaware of when they would receive a reward or which card was rewarding. In the Expected Reward condition (XR), subjects were informed that every fourth card choice would result in a **£** sign regardless of which card they picked. Two control conditions were used: Subjects selected the cards in the same way as above. In the Unexpected Control (UC), a card with **%** sign appeared with probability 0.25 while in the Expected Control (XC), every fourth card selection resulted in a **%** sign regardless of the card is selected. Subjects were informed that no reward was associated with the control conditions.

Time-frequency power plots have previously been mentioned, and its calculation will now be briefly described (for more details see, for instance, [22]). Using a wavelet analysis of the epochs, spectral changes in frequency bands activity can be analysed. Complex Morlet wavelets are used in this study for frequency analysis in order to overcome limitations associated with the constant Fast Fourier Transform window length and, thus, achieve a trade-off between time and frequency resolutions. The method provides a magnitude of the signal for each frequency, over a time range and, therefore, a time-frequency representation of the signal is generated. In this study, the frequency range analysed spans from 4.84 to 97.66 Hz in 0.49 Hz steps. An epoch from 400 ms to 100 ms prior to stimulus onset was used as a baseline, and its mean was subtracted from each time-frequency point.

6.2. GTM initialisation

The GTM-SMS model [27] was designed as follows: the latent space arrangement is a square lattice of 100 nodes, the width of the basis functions was set to a fixed value of 1, and the

initial values of the $\{\alpha_r\}$ hyperparameters, considering uninformative priors, were set in all cases to be 0.5. In order to achieve consistent results over experiments, the weight matrix $\boldsymbol{W}$ of the GTM was initialised so that the model starts from an approximation to Principal Components Analysis (PCA), a method that is described in [32].

6.3. Results

The results shown in this subsection focus on intra-trial dynamic trajectories, across-trials clustering of delay vectors, and the correspondences between these and features of different time-related representations of the original EEG data.

6.3.1 Intra-trial dynamic trajectories and its correspondence with EEG raw signal and time-frequency representations

The EEG recordings that are the origin of the analysed DE matrix contain data of 30 different trials of the same experimental task. Only a limited number of representative trajectories are plotted here in figure 4: this might suffice, as several subgroups of trajectories are neatly differentiated by the GTM latent space areas they evolve through.

These areas roughly coincide with the (macro)cluster partition of the GTM latent space obtained with the contiguity-constrained agglomerative procedure outlined in section 5.3. The latter can also be seen in figure 4. Each of these representative trajectories is neatly distinct in terms of the areas of the latent space and (macro)clusters of the map nodes involved.

It should be possible to link these results with alternative representations of the EEG data. More specifically, we would like to ascertain: first, if transition areas in the dynamic trajectories correspond to transition areas in the signal and, equally, if stationary stages of these trajectories correspond to uneventful periods both in the raw EEG and on its time-frequency representation.

First, the resulting trajectories (some of which are shown in the next few figures) manifest a clear level of "orientation", understood as the predominance of *paths* in relatively close neighbourhoods for successive delay vectors (therefore with a predominance of gradual changes), with little presence of criss-crossing. Second, stationary states are quite commonly found. Let us isolate some of these trajectories: take for instance *trials 11* and *20*, seen in figure 5. In this figure, the highlighted nodes correspond to rather long time intervals formed by several consecutive delay vectors of the DE matrix. These *stationary nodes* seem to correspond to relatively uneventful periods in the EEG signal where signal amplitude does not sharply vary.

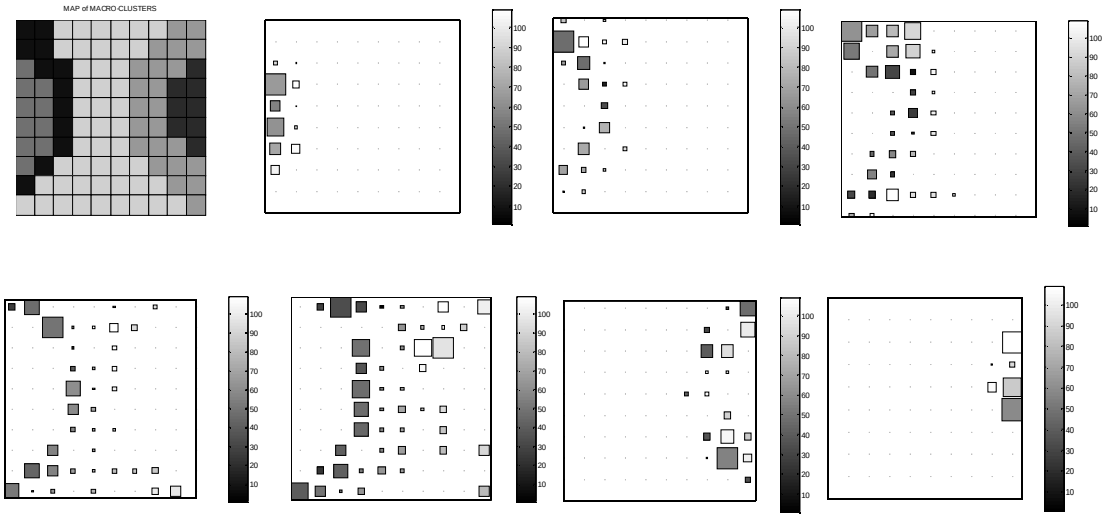

Figure 4: On the left-hand side of the top-row: a cluster partition of the GTM map obtained by application of the contiguity-constrained algorithm described in section 5.2. The rest of the figures are representative trajectories of delay vectors over complete trials. The colour-coding hints the relative time within a trial at which a delay vector has last been mapped onto it: from white, corresponding to the end of the time-span of the trial, to black, corresponding to its beginning, through all grades of grey. It must be noted that nodes responsible for several delay vectors are colour coded only according to the delay vector corresponding to the later time-interval. Notice that these trajectories (as well as the ones not depicted here) follow quite closely the limits of the obtained clusters. The relative size of the GTM nodes, represented by squares, corresponds to the number of delay vectors within a trial that are mapped onto each of them.

Third, neat transitions between areas of the latent space can be found: For the sake of brevity, consider just *trial 1*, shown in figure 6. There is a neat point of transition, between the fourth and fifth columns of the map, where the delay vectors leave the left-hand side of the map not to return, crossing towards the centre of the map. The accompanying raw EEG recording interval shows that such transition in the GTM map corresponds to a sudden shift towards high positive activity in the vicinity of 200ms after stimulus, which might be the result of a first wave of induced brain activity. Most interestingly, the later "climb" through the rightmost column of nodes towards yet a different area at the centre-top of the map corresponds to a second sudden rise of electric activity. The accompanying time-frequency plot reveals that these two transition periods immediately precede changes in the frequency pattern of activity, especially in the over-20Hz Gamma frequency band, which is commonly associated with induced brain activity.

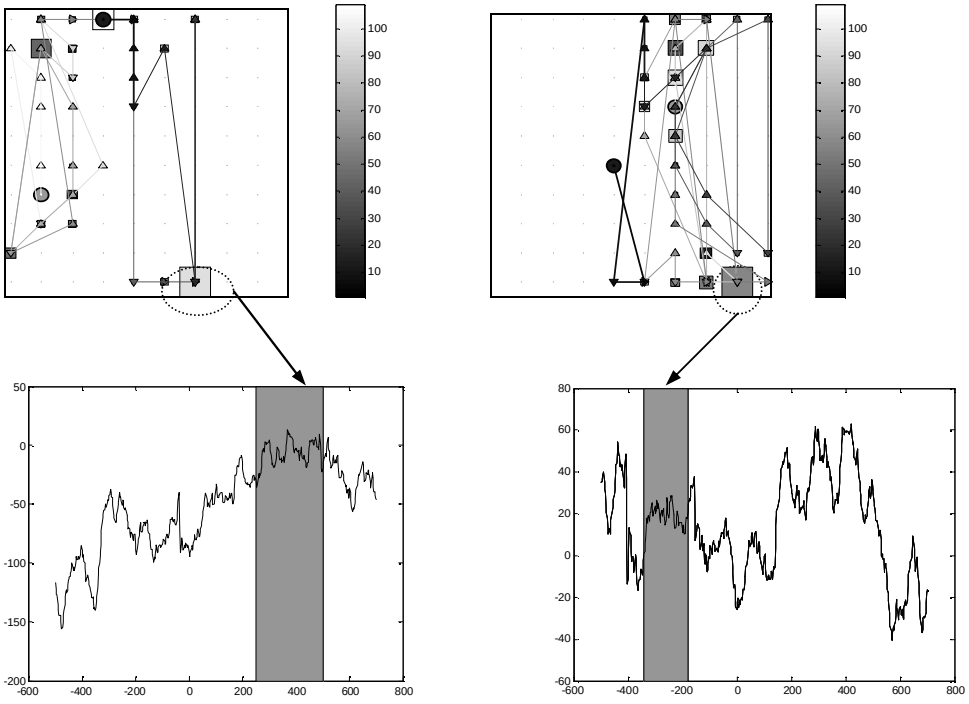

Figure 5: GTM representation of trials 11 (right) and 20 (left) and the trajectories traced by the delay vectors. Relative square (node) size and colour coding as in figure 4, but this time trajectory lines have been drawn between consecutive nodes following a similar colour pattern, with lighter shades of grey for nodes corresponding to earlier delay vectors, and darker shades corresponding to later ones. A grey circle identifies the starting node, whereas a black circle identifies the finishing one. Stationary states, characterised by several consecutive delay vectors being mapped onto the same GTM node, are highlighted. The raw EEG signal corresponding to these trials is also shown, and the time intervals corresponding to the "stationary nodes" mentioned above are also highlighted. Notice that these stationary states seem to correspond to rather uneventful periods of the signal in terms of its power amplitude.

6.3.2 Across-trials location of delay vectors and the advantages of trial subgroup isolation

As outlined in section 5.1, should the different subgroups of trajectories previously found have an influence on the distribution of the delay vectors across trials, the homogeneity of the delay vectors representation over trials would increase by isolating those delay vectors corresponding to specific trial subgroups. It would be expected that the projections of the delay vectors that correspond to the same time-windows across trials fell into a reduced number of map units that resided in a reasonably limited neighbourhood. The results generally confirm this hypothesis, as illustrated in figure 7. Each of the maps in this figure, corresponding to delay vector type 9 (from –420ms to –300ms), highlights delay vector

instances belonging to a different subgroup of trials (out of the 30 existing ones for each delay vector type) and occupying close neighbourhoods.

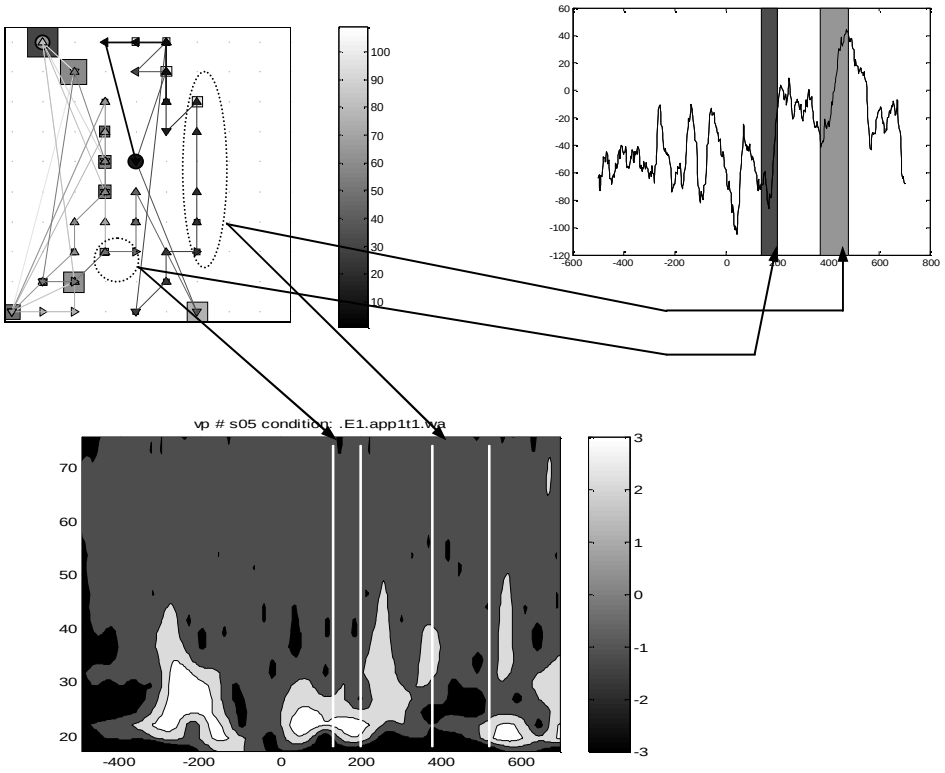

Figure 6: Similar arrangement to that in figure 5, but highlighting transition periods in the trajectory of delay vectors corresponding to trial 1 across the GTM latent space. These transition periods are shown to correspond to sharp magnitude variations in the EEG time-power signal, as well as to periods immediately preceding neat changes in the time-frequency-power patterns.

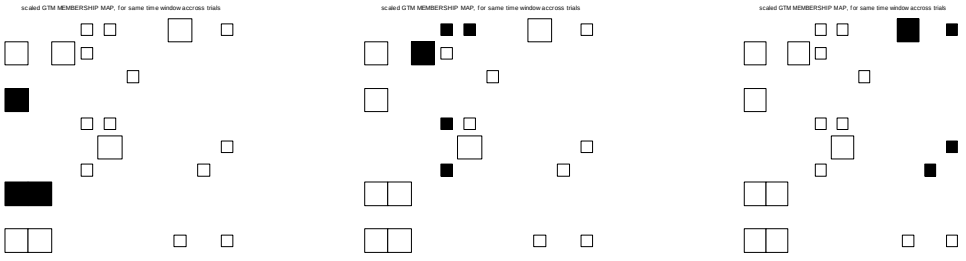

Figure 7: These GTM latent space plot represent all delay vector instances (30) for delay vector type 9 (with size-coding as in previous figures). On each of the plots, subgroups of delay vectors have been highlighted in black. On the left-hand side plot, delay vectors from two subgroups of trials (represented by the second and third plots from the left in the top row of figure 4) are highlighted. On the centre plot, delay vectors from the subgroup of trials represented by the first plot from the left in the bottom row of figure 4 are highlighted. In the right-hand side plot, delay vectors from two subgroups of trials (represented by the first and second plots from the right in the bottom row of figure 4) are highlighted.

The level of homogeneity is quantified in a simple way by calculating (and plotting in figure 8), for every delay vector, the mean distance between its instances over trials for every subgroup of trials as well as for the whole set of trials:

$$D_{sg} = {}^{1}\!\Big/\!_{2(N_{sg}-1)^2} \sum_{j=1}^{N_{sg}-1} \sum_{i=2,i>j}^{N_{sg}} \left\| x_i - x_j \right\|^2 \qquad (6)$$

for $\{x\}_{sg}$, the N_{sg} data points (delay vectors) whose representation in the GTM map belongs to each of the defined subgroups or the whole set of trials.

As expected, the mean distances (6) are much lower for any of the subgroups of trials, for any representative delay vector than for the whole set of 30 trials. There is a further interesting feature in the shape of the mean distances for the whole set of trials: it evolves quite smoothly, suggesting a gradual transition over time for the variability over trials. In fact trial homogeneity seems at its maximum at pre-stimulus periods (minimum mean distance, roughly between –260ms and stimulus onset) and at its minimum at post-stimulus times (0ms to 480ms) with the mean distance peaking at intervals usually related to top-down induced brain activity ([22],[33]) (maximum mean distance, roughly between 200ms and 320ms), suggesting that high-level cognitive processes are accompanied of higher variability over trials.

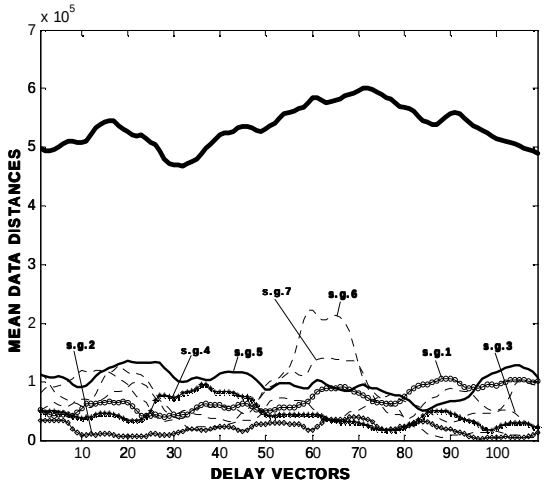

Figure 8: Plots of the mean distance (6) over the delay vector types. Values are plotted for all trials as well as for the seven subgroups of trials represented by the plots in figure 4. The much higher homogeneity of the subgroups is clearly visualised. Also the smooth and progressively changing shape of the values for all trials is indicative of time intervals for which trial variability is higher or lower. A functional interpretation is suggested in the text.

Even more interesting is the fact that the subgroups of trials correspond to gradually changing but distinctly different time-frequency representations. This can be appreciated in figure 9,

where time-frequency representations of the *Unexpected Reward* (UR) condition, following the procedure outlined in section 6.1, have been obtained for the 7 subgroups of trials represented in figure 4, as well as for the whole set of 30 trials. The latter would be the type of representation found in a typical study with multi-trial, multi-subject set-up and it turns out to be rather uninformative, with only very vague bursts of activity in the area around 20Hz, at –200ms and before 200ms. The reason behind this is evidenced by the trial subgroup representations: all of them are much more informative, with neater bursts of activity, but also very different amongst them, so that the representation for the whole set of 30 trials has most of the relevant activity eclipsed, providing a misleading clue to real activity.

Also, a clear pattern of dynamic evolution can be observed if we follow this trial subgroup representations "from left to right" in the GTM latent space (from left to right and from the top to the bottom in figure 4): The first subgroup shows narrow bursts of high-frequency gamma band activity between –400ms and –300ms, together with vague low-gamma (~20Hz) activity at around –200ms and 600ms. In the second subgroup, the high-frequency activity has decreased whereas the low-gamma activity of the previous subgroup intensifies, suggesting foci below the gamma band and also at around 100ms. All this low-frequency activity disappears in the third subgroup, with higher-frequency activity at around –200ms. This is much more intense in the fourth sub-group and seems centred around 40Hz, also active at around 100ms. The active frequencies get lowered again towards 20Hz, but more or less at the same times, although a pattern of sustained activity is clear after stimulus onset. This sustained activity is more focused in the sixth sub-group, but also an intense pre-stimulus activity appears at –400ms with the activity at –200ms extending to high-gamma frequencies. Finally, activity in the seventh subgroup increases at times over 600ms whereas almost disappears at pre-stimulus times shifting again to lower frequencies.

All these results support the hypothesis that different signal dynamics are represented by different areas of the GTM, and that these dynamics gradually evolve throughout the topographic mapping. They also come to show the usefulness of defining subgroups of trials, as this procedure unmasks different dynamics and limits the undesired effects of trial variability in the experiments.

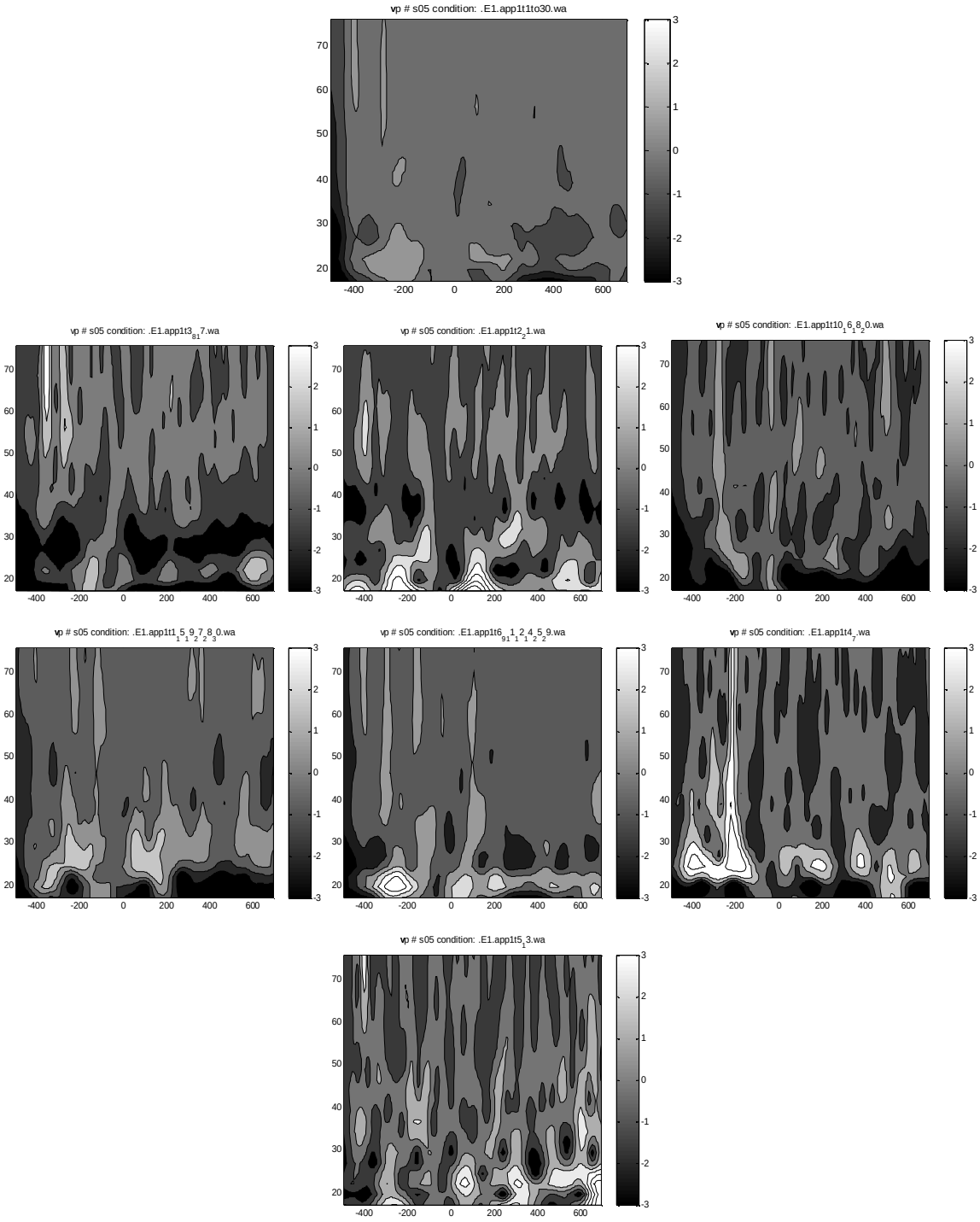

Figure 9: Time-frequency-power plots, obtained following the procedure outlined in section 6.1. The first one corresponds to the whole set of trials, whereas the following seven correspond to the subgroups of trials represented by the plots in figure 4. Much neater activations (high power values colour coded as clear shades) are seen for the subgroups of trials, as well as gradually changing patterns of activation following an imaginary path from left to right in the GTM latent space (for which functional interpretation is given in the text).

7. Discussion

The main contributions of this study can be summarised as follows: it has been shown that, through a variation on the EEG signal dynamical embedding techniques proposed in [1], the main dynamics of bounded time-interval EEG signals can be represented, to intuitive advantage, in the two-dimensional latent space of the GTM model. Signal transitions and stationary states can be neatly visualized in this representation. Many experiments involving the study of brain responses to defined tasks resort to their repetition in what are known as trials, according to the expectation that brain responses will be replicated. Experience indicates that trial variability is a common phenomenon. The GTM has been shown to discriminate between different subgroups of the multiple trials in which the experimental tasks are organized. Isolation of these subgroups has been shown to benefit the coherence of their analysis.

The GTM is a very useful tool for exploratory data mining. The present study is mainly exploratory in nature, and it is acknowledged that further research is due before the tenets in it suggested can be fully validated. Many interesting issues can be targeted by following this line of research, including the use of a similar approach to address multi-subject and multi-channel analyses. Given that the replication of a task usually entails within-subject variability, even greater inter-subject variability may be expected. It would therefore be beneficial to explore the possible existence of subgroups of individuals showing similar responses. The extension to the analysis of multi-channel recordings is of a different nature: the goal would be the discovery of subgroups of channels with similar underlying signal dynamics, and it would complement existing techniques of channel activity representation over the scalp topology. It would be of utmost interest qualifying whether the independent component analysis (ICA) of the DE matrix (following the procedure in [1]), associated to different subgroups of trials, indicates the existence of different subgroups of brain signal source combinations.

References

[1] James, C.J., and Lowe, D. (2003). Extracting multisource brain activity from a single electromagnetic channel. *Artificial Intelligence in Medicine*, 28, 89-104.

[2] Svensén, M. (1998) GTM: The Generative Topographic Mapping. PhD thesis. Aston University, UK.

[3] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics,* 43(1), 59-69.

[4] Kohonen, T. (2000). *Self-organizing Maps.* (3rd edition) Berlin: Springer-Verlag.

[5] Kohonen, T., and Hari, R. (1999). Where the abstract feature maps of the brain might come from. *Trends in Neurosciences*, 22, 135-139.

[6] Touzet, C. (1997). Neural reinforcement learning for behaviour synthesis. *Robotics and Autonomous Systems. Special Issues on Learning Robots: The New Wave*, 22, 251-281.

[7] Smith, A.J. (2002). Applications of the self-organising map to reinforcement learning. *Neural Networks*, 15, 1107-1124.

[8] Assadollahi, R., and Pulvermuller, F. (2001). Neural network classification of word evoked neuromagnetic brain activity. In *Emergent neural computational architectures based onneuroscience. Towards neuroscience-inspired computing.* Springer-Verlag, Berlin, Germany, 311-319.

[9] Kalarickal, G.J., and Marshall, J.A. (1999) Models of receptive-field dynamics in visual cortex. *Visual Neuroscience*, 16(6), 1055-1081.

[10] Miikkulainen, R., Bednar, J. A., Choe, Y., and Sirosh, J. (1999). Modeling self-organization in the visual cortex. In *Kohonen Maps* (Oja, E., and Kaski, S., eds.), Amsterdam, Elsevier.

[11] Ruf, B., and Schmitt, M. (1998). Self-organizing maps of spiking neurons using temporal coding. In *Computational Neuroscience: Trends in Research* (Bower, J.M., ed.) Plenum Press, New York, 509-514.

[12] Xu, Z.M., Ivanusic, J.J., Bourke, D.W., Butler, E.G., and Horne, M.K. (1999). Automatic detection of bursts in spike trains recorded from the thalamus of a monkey performing wrist movements. *Journal of Neuroscience Methods*, 91, 123-33.

[13] Thomas, E., Van Hulle, M.M., and Vogels, R. (2001). Encoding of categories by noncategory-specific neurons in the inferior temporal cortex. *Journal Of Cognitive Neuroscience,* 13(2), 190-200.

[14] Jaynes, E. (2003). Probability Theory: The Logic of Science. Cambridge University Press.

[15] Cerquides, J. (2003). Improving Bayesian Network Classifiers. PhD Thesis. U.P.C. Barcelona. Spain.

[16] Bradley, E. (1999). Analysis of time series. In *Intelligent Data Analysis* (Berthold, M. & Hand, D., eds.) Springer-Verlag, Berlin.

[17] James, C.J., and Lowe, D. (2000). Using dynamical embedding to isolate seizure components in the ictal EEG. *IEE Proceedings –Science, Measurement and Technologies*- 147(6), 1350-2344.

[18] Jung,T.-P., Makeig, S., Westerfield, M., Townsend, J., Courchesne, E., and Sejnowski, T.J. (2001). Analysis and visualization of single-trial event-related potentials. *Human Brain Mapping,* 14,166–185.

[19] Schlögl, A., Slater, M., and Pfurtscheller, G. (2002). Presence research and EEG. In *Presence 2002: 5th Annual Workshop on Presence*, Porto, Portugal.

[20] Duann, J.-R., Jung, T.-P., Kuo, W.-J., Yeh, T.-C., Makeig, S., Hsieh, J.-C., and Sejnowski, T.J. (2002). Single-trial variability in event-related BOLD signals. *Neuroimage,* 15, 823-835.

[21] Sinkkonen, J., Tiitinen, H., and Naatanen, R. (1995). Gabor filters: an informative way for analysing event-related brain activity. *Journal of Neuroscience Methods*, 56, 99-104.

[22] Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3, 151-162.

[23] Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, B*, 39, 1-38.

[24] Mackay, D.J.C. (1991). Bayesian Methods for Adaptive Models, PhD thesis, California Institute of Technology.

[25] Bishop, C.M., Svensén, M., and Williams, C.K.I. (1998) Developments of the Generative Topographic Mapping. *Neurocomputing,* 21(1-3), 203-224.

[26] Mackay, D.J.C. (1995). Probable networks and plausible predictions: a review of practical Bayesian methods for supervised neural networks. *Network: Computation in Neural Systems,* 6, 469-505.

[27] Vellido, A., El-Deredy, W., and Lisboa, P.J.G. (2003). Selective smoothing of the Generative Topographic Mapping. *IEEE Transactions on Neural Networks*, 14(4), 847-852.

[28] Vellido, A., El-Deredy, W., Gruber, T., Zald, D.H., McGlone, F.P., and Müller, M.M. (2003). When the penny drops: Reward predictability and oscillatory brain processes. *Computational Neuroscience Meeting, CNS 2003,* Alicante, Spain.

[29] Murtagh, F. (1995). Interpreting the Kohonen self-organizing feature map using contiguity-constrained clustering. *Pattern Recognition Letters,* 16(4), 399-408.

[30] Vellido, A., Lisboa, P.J.G., and Meehan, K. (2000). The Generative Topographic Mapping as a principled model for data visualization and market segmentation: an electronic commerce case study. *International Journal of Computers, Systems and Signals,* 1(2), 119-138.

[31] Bishop, C.M., Svensén, M., and Williams, C.K.I. (1997). Magnification factors for the GTM algorithm. In *Proceedings IEE Fifth International Conference on Artificial Neural Networks*, Cambridge, U.K., 64-69.

[32] Bishop, C.M., Svensén, M., and Williams, C.K.I. (1998). GTM: the Generative Topographic Mapping. *Neural Computation,* 10(1), 215-234.

[33] Pulvermuller, F., Birbaumer, N., Lutzenberger, W., and Mohr, B. (1997). High-frequency brain activity: its possible role in attention, perception and language processing. *Progress in Neurobiology,* 52(5), 427-445.