International Journal of Bioelectromagnetism
Vol. 9, No. 4, pp. 223 - 232, 2007

www.ijbem.org

Geodesic and functional K-means algorithms. A bounding and multistart procedure to recover functional networks from MEG/EEG recordings.

Anael DOSSEVI[ab], Line GARNERO[a], Habib AMMARI[b]

[a]*Cognitive Neuroscience & Brain Imaging Lab CNRS UPR 640., Paris, Franc*

[b]*Center of Applied Mathematics, Ecole Polytechnique/CNRS UMR 7641, Paris, France*

Anael DOSSEVI. E-mail: anael.dossevi@normalesup.org

Abstract We propose a method based on the K-means algorithm to recover correlated areas from MEG/EEG reconstructions at the source level. We use a mixt distance based on both anatomical and functional information in order to parcellise the cortex in several areas. We restrict our distance so that parcels are always of small extent and low dimensionality, and so that local correlations can be computed more precisely. We finally select sources of interest through a multistart procedure and find the relevant networks. We applied this method on both simulated data and real recordings from a visual task.

Keywords: Functional networks, K-means, geodesic distance, MEEG, correlation, coherence, phase synchrony

1. Introduction

Recent methods have proposed to estimate the coupling between distant areas of the brain using the coherence, temporal correlation or synchrony analysis combined to source reconstruction from MEG and EEG signals ([1], [2], [3]). The major problem with finding these networks is the high number of sources to test. It is computationally impossible to test all pairs of sources (a cortical distribution of N = 10000 sources would give 50 millions of correlation coefficients to compute and sort). In the present work, we have explored the idea of gathering sources in correlated and local patches and then compute the correlation or synchronies between these patches. The principal limitation with using classification methods to parcellise big data sets is the *curse of dimensionality* ([5]) which explains that classifying data sets of big dimension is a very unstable procedure that can lead to very different classifications when data slightly change. In this context, using K-means or K-nearest neighbors algorithms can fail to reveal significant classes of functionally close sources.

We decided to use the well-known K-means algorithm ([4]) based on both geometrical and functional information to subdivide the cortex in correlated areas. However, in order to cope with the problems caused by the high dimension of the data, we decided to associate two additional constraints to the classical K-means algorithm. First, we chose to classify the activity of the cortex *locally*: we added a *bounding* constraint for classification so that too spatially distant sources cannot be associated. Indeed, sources spatially close are supposed to have close activities ([1], [6]) thus the dimensionality in areas of small extent is generally smaller than for the entire data set. Secondly, given a data set, we used a *multistart* procedure: we computed series of K-means parcellisations with different initial parameters and selected sources belonging recurrently to correlated parcels.

Our method is divided in three steps. First, we perform a multistart of K-means parcellisations series (changing the number and location of the original seeds) based on a mixt anatomo-functional distance. Then, given a correlation criterion, we select the recurrently correlated sources. Finally, we find the networks within the selected sources, based on a purely functional K-means on the selected sources. We applied this method to simulated as well as real data in order to find functional networks of correlated sources.

2. Material and Methods

2.1. K-means algorithm.

The K-means is a particular case of EM algorithms (Expectation Maximization). Given $\mathbf{M}=\{x_1,\ldots, x_N\}$ a set of N data equipped with a distance *d*, a number of classes K and $\{m_1,\ldots, m_K\}$ K elements of **M**, called the *seeds*, the K-means algorithm subdivides the data in K classes minimizing the total intra-class variance.

Given a subdivision of the data set **M** in K classes, denoted $\{C_1,\ldots, C_K\}$, and their K seeds $\{m_1,\ldots, m_K\}$ the total intra-class variance is given by $\mathbf{V}=\sum_{k=1}^{K}\sum_{x\in C_k} d(x,m_k)^2$. The process to minimize recursively the total intra-class variance is divided in two steps: the first step consists in, given the seeds, associating each element of the data set to its closest seed; the second step is, for each class C_k, to define the new seed as the data element $\mathbf{x}$ of C_k minimizing the intra-class variance (given by $\sum_{x\in C_k} d(x,\mathbf{x})^2$). This procedure is a simple gradient-descent of **V** with regard to the parameters *seeds* and *classes*. Consequently, the algorithm converges to a local minimum of **V**.

2.2. Distances.

In order to parcellise the cortex in correlated areas, we based the K-means algorithm on a distance using the geometrical as well as the functional information of the sources. The distance was a weighting between a distance on the activities of the sources, such that two correlated sources have a zero distance, and a geometrical distance adapted to the cortex convoluted surface.

Functional distance.

The criterion we chose to estimate the coupling between two sources, with activities j_1 and j_2, is the temporal correlation given by:

$$\mathbf{cor}=\left|\frac{\int j_1(t)j_2(t)dt}{\|j_1\|_2\|j_2\|_2}\right|=\left|\left\langle \frac{j_1}{\|j_1\|_2},\frac{j_2}{\|j_2\|_2}\right\rangle\right| \tag{1}$$

where $\|j\|_2^2=\int j(t)^2dt$ is the classical L_2-norm, and $\langle .,.\rangle$ is the associated scalar product. From now, we will always consider that the activities of all the sources have been normalized so that $\|j_s\|_2=1$ for all source s, and consequently that the temporal correlation between two sources is given simply by $|\langle j_1, j_2\rangle|$. The correlation is a scalar between 0 and 1, such that the closest to 1 is the correlation the more similar are the signals. Consequently, the functional distance we decided to choose to quantify the temporal correlation between 2 activities j_1 and j_2 was given by

$$d_f(j_1,j_2)=2-2|\langle j_1,j_2\rangle|=\min(\|j_1-j_2\|_2,\|j_1+j_2\|_2)\,. \tag{2}$$

This distance permits to consider that 2 correlated activities but with opposite signs have a zero distance.

Spatial distance.

As the cortex is a very complex and highly convoluted surface, the Euclidean distance is not a relevant tool to fit its geometry. In order to be able to describe more precisely the distance between sources on the cortical surface, one has to compute the geodesic distance. This measure was the distance we decided to use.

The computation of the geodesic distance at the cortex level between a source s_1 and a source s_2 has to be done step by step by computing the Euclidean distance between s_1 and its neighbors, and then gradually to its second neighbors, third neighbors, and so on till reaching the source s_2. The method used to compute this distance is based on the hierarchical queue (HQ) process exposed precisely in [7].

Bounding procedure.

We explained in the introduction that the problem with classifying large data sets is that when the dimensionality of the data is too high, classification algorithms like K-means or K-NN generally fail to reveal significant classes. This limitation is known as the *curse of dimensionality* and some solutions have been proposed in the past decades in order to solve it ([6], [8], [9]). A first idea to circumvent this problem is to apply K-means on data subsets supposed to have a low dimensionality. It is commonly admitted ([1], [6]) that the cortical activity is not purely punctual, and that, if a source is activated in a region, contiguous sources belonging to the same functional area will be active, and with similar time courses. Consequently, we can consider that on areas of small spatial extent the dimension of the activities will be significantly lower.

That is why we decided that a source could be associated only to one of its *p* spatially closest seeds. This constraint prevents a source from being associated with a remote seed and leads to applying the K-means algorithm on bounded areas, where the bounding limit is adapted to each source. This is equivalent to saying that the geodesic distance between a source and the p+1th to Kth seeds (sorted in the geodesic distance order), given by d_g (s, $\mathbf{x}_k$), is set to infinity.

Choice of the mixt distance

Finally, given the functional and the geodesic distance between two sources, the mixt distance was given by

$$d_\alpha(s_1, s_2) = d_g(s_1, s_2) + \alpha d_f(j_1, j_2) \tag{3}$$

The parameter α is not intrinsic and depends on the weight one wants to give to the functional distance. We will see that we tested different weights in the treatment of simulated as well as real data to test the efficiency of each type of weighting.

2.3. Multistart procedure and sources selection.

In order to overcome the dimensionality of the data, and in particular the dependency of the K-means on the choice of the original seeds, we decided to compute a series of K-means where we changed the number as well as the location of the original seeds. These K-means lead to very different parcellisations of the cortex activity and we need to find the recurrent patterns of these parcellisations. The main pattern of interest we want to focus on is the correlation within the parcels.

Consequently, for each K-means, we chose to compute in each parcel a correlation criterion and then average it across the results of the series of K-means. Given a fixed parcel, the correlation criterion, assigned to each source of the parcel, is defined by the following formula:

$$\mathbf{cc} = 1 - \frac{s_2}{s_1} \tag{4}$$

where s_1 and s_2 are the first two singular values given by the singular value decomposition (SVD) of the signals of the sources in the parcel. When all the sources have similar activities s_2 equals 0 and **cc** equals 1, while on the contrary sources have very distinct activities s_2 is close to s_1 so **cc** is close to 0. This criterion permits to know if the sources within a parcel are highly correlated or not, and we can show that it is a better discriminator than the average correlation within the parcel. We finally decided to average it across the series of parcellisations to have a quantification of the correlation of each source inside its parcels. From the value of this average correlation criterion, we can select the sources of interest, being the most correlated.

2.4. From sources to networks.

Once we have selected the sources of interest, we still need to find if these sources belong to a functional network and determine these networks. The method we have adopted at this step is to apply again a K-means algorithm, yet this time based only on a functional distance. Once we have selected a small ratio of the most correlated sources, we can admit that the dimensionality of the selected data set is low, and consequently that a usual K-means is stable enough to classify precisely the groups of correlated sources, with no regard to the location of the sources.

As the number of functional networks to be detected can be considered as small (we hope no more than a few networks to be active in a given task on a fixed time window), we initialized the K-means with a small number of seeds, in general lower than 10. The SVD of the selected sources permits to know the dimension of interest. We showed in a previous article ([3]) that when different networks are implied in a task, if they are not too correlated, the number of networks equals the number of singular values significantly bigger than the following ones. Thus, displaying the first singular values can give an idea of the number of seeds to be taken. The optimal number should not be precisely equal to the number of significant singular values, so that we can classify more accurately the sources belonging to not functionally orthogonal networks, as well as the sources selected via our previous process but not belonging to any network (this means to create a parcel of outliers).

Let denote p the number of seeds necessary, we decided to define the first p-1 seeds as those most correlated to the first p-1 singular vectors respectively, and on the contrary, the last seed as the source whose projection in the subspace spanned by the p-1 first singular vectors was the smallest.

Once the parcels have been computed, we compute again the SVD of each parcel. We first remove the parcels whose correlation criterion is too small; typically, we remove the parcel for which it is lower than 0.5 (which means that $s_1 < 2s_2$). Then, we merge the parcels whose first singular vectors are highly correlated (over 0.9). Finally, we eliminate the parcels containing too few sources, given the fact that a relevant functional network should not contain a restricted number of sources and should extend on several cm².

2.5. Algorithm steps

In order to clarify the algorithm process, we recall and show a few more details about its successive steps. The letters correspond to those displayed in figure 1.

First step: parcellisation and correlation criterion (A-B)

We perform a series of N parcellisations of the same data set, varying each time the number and location of the initial sources. For each parcellisation and each parcel, we compute the first 2 singular values s_1 and s_2 of the sources signals and give to each source of the parcel the criterion value $1-(s_2/s_1)$.

Second step: selection of the sources of interest (C-D)

We compute for each source its mean correlation criterion value. From the statistical distribution of these mean values or given a ratio of sources to be selected, we select the sources most correlated. In the simulations and real data processing, we chose to take ten percent of the most correlated sources.

Third step: determination of the functional networks (E-F)

We compute a K-means based only on the functional distance in order to classify the selected data in a small number of classes. We compute the correlation criterion of each class and select the most correlated classes. Then we compute the characteristic signal of each correlated parcel, given by the first singular vector of its signals, and compute the correlation between the characteristic signals of the parcels to merge the highly correlated ones. We finally get rid of the too small parcels.

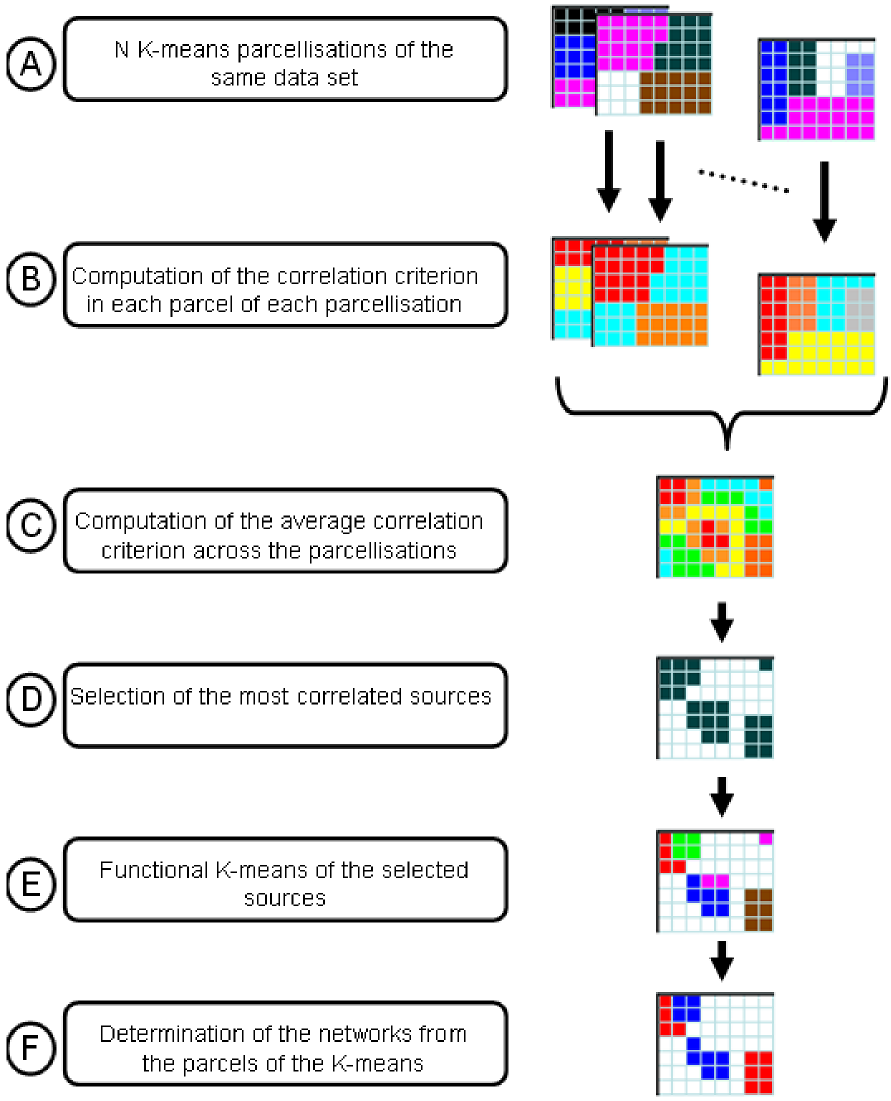

Figure 1. Detailed outline of the different steps of our functional networks classification algorithm.

3. Results

We applied this method on simulated data as well as on real recordings in a 151 channels MEG of a visual task (see [1]). In each case, we performed 3 series of 100 K-means parcellisations, applying the algorithm with the number of neighbor seeds being p=4. Each series corresponded to a different weighting between geodesic and functional distance: the first giving more importance to the geodesic distance, the second giving same weight to both and third giving more emphasis on the functional distance. We also performed a series of the classical K-means at whole cortex level, based on the purely functional distance.

3.1. Results on simulated data

Simulation settings

We used a distributed dipoles model in order to simulate MEG data. A segmented cortex obtained from an MRI volume was used. The boundary between white and grey matter was extracted and approximated by a triangle-based tessellation. The corresponding vertices provided approximately 11200 dipole positions covering the entire cerebral cortex.

We located two *target* networks of perfectly correlated sources. The sources of each target network were distributed between 2 or 3 patches (see fig. 2(a)). The rest of the cortex was subdivided into 500 patches such that the sources not included in the target networks were uncorrelated to the sources of the networks, although locally partially correlated within each of these 500 patches. In these series of simulations, the power of the activity of the sources in the network was equal to 1, while the sources in the other 500 patches had random power between 0 and 1, but constant in each patch.

Knowing the location of the vertices of the cortex and of the virtual MEG sensors, we computed the gain matrix using the Brainstorm software[1]. We computed the recordings obtained on the sensors and then reconstructed the cortex activity using the Minimum Norm operator ([10]). Although this inverse operator is not the most accurate and despite the controversial aspect of choosing the solution of minimal norm, it gives a good estimation of the major centers of activity on the cortex and, furthermore, is linear and consequently computationally very quick to give the inverse estimation.

After reconstructing the activity, we normalized the activities of each sources (so that the L_2-norm is 1), and applied our functional networks classification algorithm.

Choice of the distances

In order to estimate the influence of the weighting parameter in the mixt distance (3), we performed a series of 100 hundreds K-means for several values of α. We chose the values 1, 10 and 100 corresponding respectively to a distance almost equivalent to the geodesic distance, a distance giving the same weight to both geometric and functional parts, and a distance putting the most emphasis on the functional information. Furthermore, in order to test the influence of the bounding procedure we also computed a purely functional K-means at the whole cortex level. In this last K-means, the distance is almost equivalent to the third mixt distance but the difference lies in the cancellation of the bounding procedure whose purpose is to give a more accurate estimation of local correlations as well as to try to overcome the problem of the high dimensionality of the data.

In order to be able to compare properly the results of each type of distance, we used the same sets of initial seeds for all the distances in each series.

Results

The results of the simulations are displayed in figure 2. The results given by the three mixt distances (a-b-c) seem to be globally similar and the different patches of both networks are detected. However, the results given by the unbounded K-means based on a purely functional distance (d) gives poor results: the extent of the red network is far too important and the blue one is hardly detected on the right hemisphere.

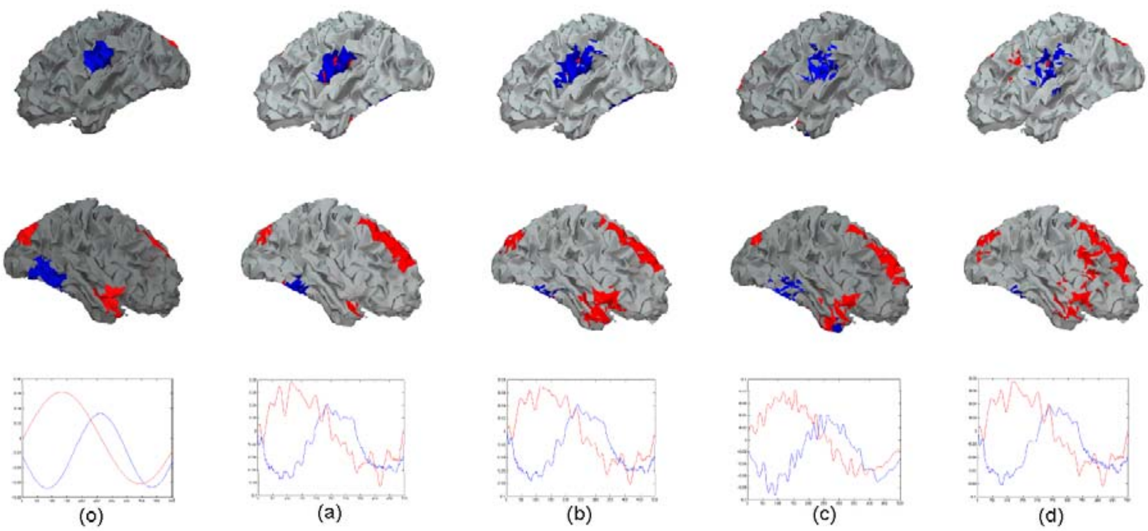

Figure 2. (o) Original target networks and their associated signals. (a-b-c) For each weighting in the mixt distance, networks reconstructed after application of the algorithm, after inverse problem. (d) Networks reconstructed via the purely functional K-means. Bottom: the characteristic signals of the reconsctructed networks.

[1] Freely available at http://neuroimage.usc.edu/brainstorm

3.1. Results on real data

Experimental protocol

In order to test the efficiency of our approach under actual experimental conditions, we used data from a simple visual stimulation in one test subject. Neuromagnetic recordings were acquired on 151-SQUID sensor VSM MEG (see [1]) at the MEG-center of the Hôpital de la Salpétrière, Paris. The subject was instructed to fixate on the center of an expanding checkerboard ring (see fig. 3). The data we treated were the average signal from 33 recordings of 1.7sec long: the first 500ms corresponded to a rest period and the following 1.2sec corresponded to a period of expanding rings. The rings spanned at 5Hz, and we consequently filtered our data between 3 and 7Hz in order to isolate the networks driven by the visual stimulation.

We focused and applied our method on a particular window of interest. This window begins 800ms after the apparition of the rings and corresponds to a steady-state period for the response of the subject. We applied as well our method on the transitory response (corresponding to the N1/P1 waveforms), but we do not display the results here.

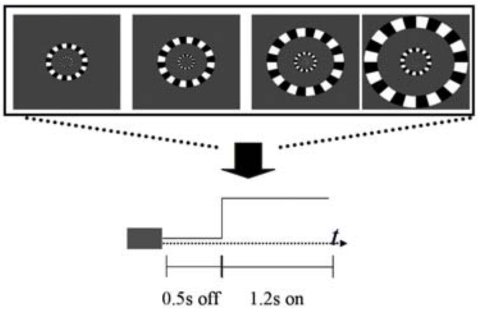

Figure 3. Expanding checkerboard used in the real experiment.

Results on the real data.

The figure 4 represents the networks detected for each kind of distance on the 4 external and mesial faces of the cortex. Each color represents a different network. The graphs at the bottom of figure 3 represent the time courses of each network, and in the case where several patches have been merged into one single network, the time course of each patch is presented.

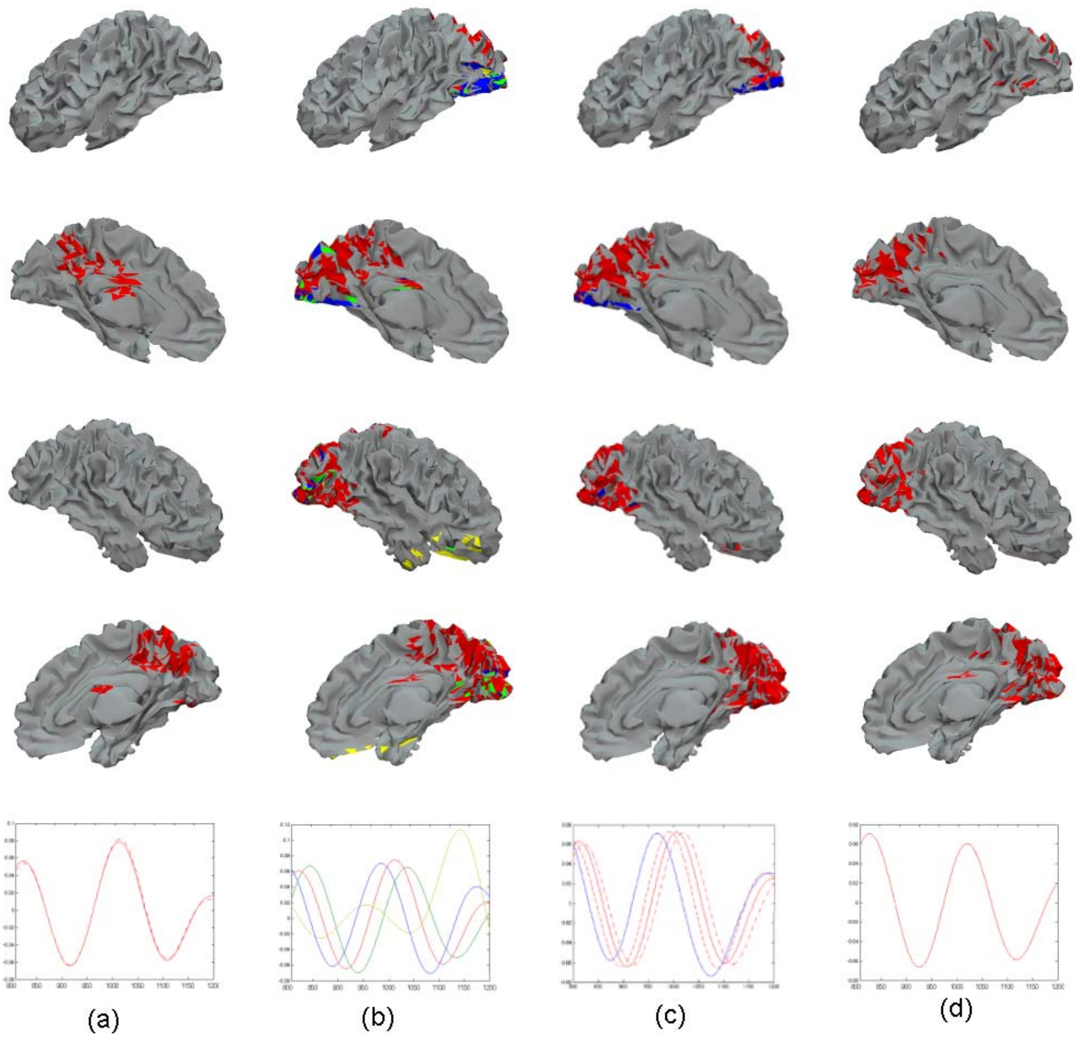

Figure 4. (a-b-c) Networks given by the algorithm for each mixt distance (respectively with weights 1, 10 and 100). (d) Network given by the unbounded K-means based on the purely functional distance. Bottom: the characteristic signal of each network. When several patches have been merged into one single network, the signals af all the patches are displayed in the same color (eventually in doted lines).

In the case of the real data study, the networks given by each mixt distance are quite different. In particular, the mainly geodesic distance does not detect sources in the occipital visual cortex which is quite unexpected for a visual stimulation task. A few sources are detected in the occipital lobe but principally around the parieto-occipital junction, far from the primary areas of the visual cortex. On the contrary, the other K-means manage to detect networks in the visual cortex. A big network (in red) is detected by both the second and third mixt distances bounded K-means as well as by the purely functional one. This major network is equivalent for these three K-means, and similar to the network found in a previous article (see [3]). The medial occipital regions are bilaterally involved, as is the dorsal extension of the occipito-parietal areas. A right hemisphere ventral (inferior lateral occipital and posterior infero-temporal regions) stream is clearly appreciated. As expected, this main network is strongly localized to the primary and higher level visual cortices. For the second and third mixt distances, other networks are also detected. A network (in blue) is detected in the higher areas of the visual cortex, especially in the V4 and V5 areas of the left hemisphere. The second mixt distance reveals also two other networks, one (in green) in the occipital lobe and another (in yellow) in the inferior part of the frontal lobe and in the most frontal part of the temporal region in the right hemisphere, that could correspond to the infero-temporal visual areas. This *yellow* network is also present in the case of the third distance but, as the number of sources was too small in this case, we did not select it.

Furthermore, the time courses of the networks, especially for the second and third mixt distances (b-c), give interesting information on the networks. Figure 4(b) (bottom) shows that the first three networks (in red, blue and green) are not correlated over 0.9 but present a good phase locking. All of them show a steady state response at the 5Hz frequency (the frequency of the checkerboard) but the blue network is in advance on the red one while the green is late. For the third distance, the time courses of figure 4(c) also bring the conclusions that the different networks are not correlated but have good phase synchrony. The 3 patches merged in one network (in red) show quite similar activities to those of the blue and green one of figure 4(b). The phase difference is lower but one is also in advance and one is late. Although some patches are grouped under one single network, the different parcels given by the functional K-means applied after selecting the sources are quite similar for the two last mixt distances.

4. Discussion on the results

Whereas results on simulated data seemed to tell that results where mostly equivalent (even if not exactly identical) without regard to the chosen distance, the application of our method and the analysis of the results on a real data set leads to more constructive conclusions.

It appears that the bounded K-means based on the quasi-geodesic distance (see fig. 3(a)) fails to detect sources in the occipital lobe, which is preposterous as we are studying a purely visual task response. This kind of distance seems to be irrelevant in the case of the real data. On the contrary, the two other mixt distances and the functional K-means show much more coherent results.

The difference between the results on the simulated data and on the real experiment is the following. The first distance is almost equal to the geodesic distance, thus, the K-means based on this distance gives *connected* parcels. In the case of our simulations, we set several *connected* patches, and consequently, this distance is efficient to detect these big *connected* correlated areas, while the other distances select some more correlated sources even if they are not physically connected to the others. On the contrary, in the case of real data, the functional networks are not necessarily composed by big connected regions. The figure 4(b) shows that the networks are tangled in the occipital cortex, and that the blue, green and yellow networks are not constituted with big connected areas. As the first distance is not able to separate in different patches close sources even if they are functionally distant, sources from different networks are grouped together. Then, the step of selection of sources fails to detect these sources because the value of the average correlation criterion in their parcels is too low. As the second and third mixt distances, putting more emphasis on functional information, do not necessary lead to connected parcels, they are able to separate these sources, then give good correlation criterions in these same areas, and finally bring results that are more relevant.

The purely functional and not bounded K-means is also able to detect the big networks but fails to detect the networks constituted of a small number of sources. This shows the interest of the bounding procedure that allows detecting local high correlations. In the case of a functional K-means at the whole cortex level, correlation of sources in small areas are covered by higher long distance correlations and, consequently, only the sources belonging to the biggest networks can be selected.

Although it seems to be a big difference (owing to the colors of the networks) between the results of the second and third mixt distance, the precise analysis of the time courses of each sub-patch of each network shows that the functional characteristics of each patch are quite similar. However, at this point of the study, the mixt distance giving equivalent weights to both geometrical and functional parameters seems to give slightly better results.

5. Conclusions

We developed a new approach to classify sources activity at the cortex level, by applying a K-means based on both geometrical and functional distance. The first major contribution in this work is the use of a bounding procedure that permits to compute more precisely the local correlations than a classical K-means at the whole cortex level. This allows detecting small networks even when bigger and highly correlated networks should hide the correlation of small extent networks. The second interest in this method is the multistart procedure that allows finding the sources belonging to recurrently correlated parcels over trials. From the results obtained on the real data, it appears that the simultaneous use of these two properties is efficient to overcome the *curse of dimensionality* which usually prevents any simple K-means from giving relevant classification of the data.

Although the multistart procedure is computationally long, a good choice of the weighting between the distances allows this method to reconstruct relevant networks. The results obtained on simulated data where quite good for all the distances we used, but the analysis of real MEG recordings of a visual stimulation proved that the results were much better when using distances incorporating an important part of functional information. We do not give a definitive conclusion on which precise weighting should be used, but it appears that functional information should be at least as important as the geometrical part.

We applied our method using the temporal correlation as a coupling criterion, but other perspectives are possible: changing the functional distance the appropriate way can allow one to use other coupling measures such as coherence and even phase synchrony.

Acknowledgements
AD thanks Diego Cosmelli for letting us use his recordings of the visual stimulation task, and Michel Besserve for his advice and references on the *curse of dimensionality*.

References

[1] David O & al., IEEE Trans Biomed Eng. 2002,49(9) :975-87.
[2] Gross J & al., Proc Natl Acad Sci USA. 2001 ;98(2) :694-9.
[3] Dossevi A & al., IEEE Trans Biomed Eng. 2007, to appear
[4] McQueen J, Proc 5th Berk Symp on Math Stat & Prob, 1967, 1: 281-297
[5] Donohoo DL, Lecture Note Math Challenges of the 21st century, 2000
[6] Varela F & al, Neuroscience 2001, 22 : 145-173
[7] Cuisenaire O, Proc 9th Int Conf Image Analysis and Processing, 1997, I: 263-270
[8] Weber R & al, Proc 24rd Int Conf Very Large Data Bases, 1998: 194-205
[9] Korn F & al, IEEE Trans Knowledge and Data Eng, 2001, 13: 96-111
[10] Baillet S. & al, IEEE Signal Processing Magazine, 2001, 18(6) : 14-30