The critical regularization value: incorporating spatial smoothness to enhance signal detection in highly noisy fMRI data*

Xian Yang, Lei Nie, Paul M. Matthews, Valentina Tomassini, Zhiwei Xu and Yike Guo

Abstract— **Comparing serially acquired fMRI scans is a typical way to detect functional brain changes in different conditions. However, this approach introduces additional variation on physical and physiological conditions, which results in substantial noise. To improve sensitivity and accuracy of signal detection in such highly noisy fMRI data, potentially important information should be incorporated. Here we propose a new significance indicator, the critical regularization value (CR-value), which detects significantly changed voxels by taking both the magnitude of the voxel-wise signal variation and spatial smoothness into account. The CR-value allows voxels that survive in a stronger sparse constraint to be considered as more significant. We demonstrate our method using a simulation dataset and a real fMRI dataset collected from the previous study. The results show that CR-value more accurately detects the true activation than GLM P-value, Posterior Probability Maps (PPM) and the Threshold Free Cluster Enhancement (TFCE) in noisy datasets.**

I. INTRODUCTION

There is an increasing interest in using task-based fMRI to detect interactions of brain activation with differences in condition, often by contrasts of serial imaging sessions over time. Additional sources of noise are introduced into contrasts by factors such as day-to-day variation in the haemodynamic response [1-2] or subject movement [3]. Noise reduction of fMRI data either demands longer signal averaging times or averaging signal changes over larger numbers of subjects. For applications to clinical populations, e.g., in assessing learning [4] or drug effects [5-7], neither option may be practical.

To improve sensitivity and accuracy of signal detection from highly noisy fMRI data, potentially important information, such as spatial smoothness and connectivity constraints, therefore should be incorporated. Inspiring by the fact that active voxels are typically found adjacent to other active voxels in a cluster, several cluster-level inference methods have been proposed. These methods are commonly implemented in two steps: first, find groups of contiguous voxels (clusters) from the statistic image with a predefined threshold; second, calculate the significance level of each cluster [8]. A recent advance in this cluster-level inference is the Threshold Free Cluster Enhancement (TFCE) method that takes all threshold values and integrates over them to derive the cluster-level significance [9]. However, even with this framework, spatial information is used in the second processing step after single voxel inference.

In order to make direct use of spatial constraints, a regularized general linear model implemented in the Bayesian framework using spatial priors based on Sparse Spatial Basis Functions was proposed [10]. This approach used Posterior Probability Maps (PPM) to characterize regional specific effects [11]. As an exact Bayesian inference process is typically computationally intensive and frequently intractable, approximation approaches, such as variational Bayes method [12] and the type-II maximum likelihood method [13], were used to estimate the posterior distributions. Although empirical studies showed the mean of an approximate posterior distribution was relatively accurate [14-15], caution is needed in using the variance of an approximate posterior distribution as a significance indicator. This is because the approximate posterior distributions are usually more compact than real distributions [16].

In this paper, we propose an alternative index, the critical regularization value (CR-value), which detects significantly changed voxels by taking both the magnitude of the voxel-wise signal variation and spatial smoothness into account. A voxel that survives in a stronger sparse constraint has a greater CR-value indicating greater significance. To implement this idea, the general linear model (GLM) is solved using regularized optimization with various values of a penalty parameter λ. Fewer voxels survives with larger values of the parameter λ. If the voxel A is selected while the voxel B is not for a given value of the parameter λ, we can claim that the voxel A is relatively more significant than the voxel B. The CR-value of a voxel is defined as the largest value of the parameter λ to ensure this voxel being selected. Thus, a map of functional brain changes is constructed using the CR-values of all voxels. Every CR-value takes into account spatial relationships between voxels because the regularized optimization includes all voxels in the brain image. We also extend formulation of the CR-value for multi-subject analysis. We demonstrate the performance of our method using both the simulated data and a real fMRI dataset from a study of motor learning reported previously [17].

II. METHODS

In this section, fMRI data are modelled using the general linear model. Then, we define our significant indictor, critical

*Research is partially supported by the Strategic Priority Program of Chinese Academy of Sciences (No. XDA06010400), the Innovative R&D Team Support Program of Guangdong Province (No. 201001D0104726115) and the Imperial College Healthcare Trust Biomedical Research Centre.

X. Y. is with Imperial College London, London, SW7 2AZ, UK (e-mail: xian.yang08@imperial.ac.uk).

L. N. is with Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: nielei@ict.ac.cn).

P. M. is with Imperial College London, London, SW7 2AZ, UK (e-mail: p.matthews@imperial.ac.uk).

V. T. is with Cardiff University School of Medicine, Cardiff, CF14 4XN, UK (e-mail: TomassiniV@cardiff.ac.uk).

Z. X. is with Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100190, China (e-mail: zxu@ict.ac.cn).

Y. G. is with Imperial College London, London, SW7 2AZ, UK, and Shanghai University, Shanghai, 200444, China (corresponding author; phone: +44-20-75948182; e-mail: y.guo@imperial.ac.uk).

regularization value, for both single-subject and multi-subject analyses.

For each subject, we assume there are two fMRI images that are before and after the training process respectively. For each image s ∈ {1,2} , its corresponding general linear model is formulated as follows:

$$Y_s = DX_s + E_s. \quad (1)$$

The $T \times N$ matrix Y_s represents the fMRI data consisting of T time points at N voxels. These datasets are explained by effects included in the design matrix D. The design matrices for both images are identical, because subjects should be scanned under nearly the same conditions. The $K \times N$ matrix X_s is the regression coefficients. The element in the ith row and jth column of X_s indicates the association between the ith effect and the jth voxel. The matrix E_s represents the residuals. The objective is to find out the changes associated with these effects. We subtract the model of the first image from the model of the second one:

$$\Delta Y = D\Delta X + \Delta E \quad (2)$$

where $\Delta Y = Y_2 - Y_1$; $\Delta X = X_2 - X_1$; and $\Delta E = E_2 - E_2$. The matrix ΔY is the changes between the two fMRI images, while the matrix ΔX is the changes of the associations between effects and fMRI signals.

Equation 2 can be solved by the following optimization procedure:

$$min_{\Delta X} \|D\Delta X - \Delta Y\|_F^2 + \lambda \left(\frac{1}{\alpha} \|\Gamma \Delta X\|_F^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} |\Delta X_{ij}| \right) \quad (3)$$

where $\|\cdot\|_F$ represents the Frobenius norm of a matrix. The error term $\|D\Delta X - \Delta Y\|_F^2$ is to minimize the total square residuals, which is a standard objective in regression. The smooth term $\|\Gamma \Delta X\|_F^2$ is to make the regression coefficients spatially contiguous. The $H \times N$ matrix Γ represents spatial relations among voxels. H is the number of neighbourhood relationships in the image. Each row of Γ represents a neighbourhood relationship between two voxels. For example, suppose there are 4 voxels in an image. If the first and third voxels are neighbours, there is a row $(1 \ 0 \ -1 \ 0)$ in the matrix Γ to calculate the coefficient difference between these two voxels. There are several definitions of spatial relations. We use the typical definition that two voxels are related if they are neighbours sharing a common face. Therefore, there are 6 neighbours at most for each voxel. $\|\Gamma \Delta X\|_F^2$ counts the total square differences between neighbours. The smoothing parameter α controls the smoothness of the regression coefficients. Fixing the value of the smoothing parameter α, the coefficients ΔX is a function of the parameter λ, which is denoted as $\Delta X(\lambda)$.

The sparse term $\sum_{i=1}^{K} \sum_{j=1}^{N} |\Delta X_{ij}|$, similar to L1 regularization of a vector, aims at reducing the number of nonzero regression coefficients. The parameter λ both determines the level of sparseness and smoothness, but the ratio of sparsity to smoothness remains fixed to α. All the regression coefficients can be regulated to zero if the parameter λ is sufficiently large. With the decrease of λ, the coefficients ΔX will gradually become nonzero one by one. The larger the parameter λ is, the sparser the matrix ΔX will be. The regression coefficients are identical to the least square coefficients if the parameter λ is zero.

The critical regularization value (CR-value) of the ith effect and the jth voxel, which indicates the relative significance of association between the effect and the signal change of the voxel, is defined as the maximum of all possible values of the parameter λ that can regularize the coefficient ΔX_{ij} of this effect and voxel to be nonzero:

$$c_{ij} = max\{\lambda | \Delta X_{ij}(\lambda) \neq 0\}. \quad (4)$$

If $c_{ij} > c_{ik}$, then with λ being equal to c_{ik} the coefficient $\Delta X_{ik}(\lambda)$ of the ith effect and the kth voxel reaches zero while the coefficient $\Delta X_{ij}(\lambda)$ of the ith effect and the jth voxel is still maintained as non-zero. This indicates that the coefficient ΔX_{ij} is more likely to be non-zero than the coefficient ΔX_{ik}. In other words, for the ith effect, the jth voxel is more likely to be included in a model than the kth voxel. Thus, the jth voxel is more significant than the kth voxel for the ith effect. Fig. 1 provides a simple demonstration of how CR-values are calculated. After CR-values for all voxels have been calculated, they can be normalized to z-scores.

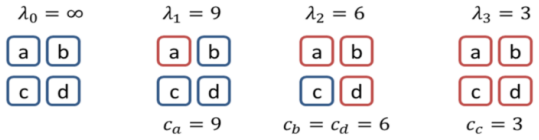

Figure 1. An illustration of CR-value calculation. There are 4 voxels that are denoted as a, b, c and d. If the coefficient of a voxel is zero, it is coloured in blue. Otherwise, it is coloured in red. All the coefficients are zero if the parameter λ is sufficiently large. When the parameter λ decreases to 9, the coefficient of voxel-a firstly becomes nonzero. Therefore, the CR-value of voxel-a is 9. As the parameter λ continually decreases to 6, the coefficients of voxel-b and voxel-d are nonzero for the first time. The CR-values of both voxel-b and voxel-d are 6. Similarly, the CR-values of voxel-c is 3.

It is worth noting that the CR-value is different from the linear regression with L1 regularization. For L1 linear regression, λ is a fixed parameter as a part of the input; the regression coefficients ΔX are the output. The values of λ are typically the same for all the regression coefficients. Differently, the CR-value aims at automatically finding the critical values of λ, which penalize regression coefficients into zero. The critical value of λ varies for different regression coefficients. These critical values of λ are the output of the CR-value procedure.

The CR-value can be extended to multi-subject analyses. Consider R subjects in a study. For each subject, there are two fMRI measurements that are taken before and after the training process respectively. According to Equation 2, we get the following equation for subject l:

$$\Delta Y^l = D\Delta X^l + \Delta E^l \quad (5)$$

where $1 \leq l \leq R$. Our objective is to find significant regression coefficients across all subjects. Utilizing multi-task learning [18], we get:

$$min_{\{\Delta X^l\}} \sum_{l=1}^{R} \|D\Delta X^l - \Delta Y^l\|_F^2 + \lambda \left(\frac{1}{\alpha} \sum_{l=1}^{R} \|\Gamma \Delta X^l\|_F^2 + \sum_{i=1}^{K} \sum_{j=1}^{N} \sqrt{\sum_{l=1}^{R} |\Delta X_{ij}^l|} \right). \quad (6)$$

Similar to Equation 3, the term $\sum_{l=1}^{R}\|D\Delta X^l - \Delta Y^l\|_F^2$ aims at minimizing the total square residuals. The second term $\lambda\left(\frac{1}{\alpha}\sum_{l=1}^{R}\|\Gamma\Delta X^l\|_F^2 + \sum_{i=1}^{K}\sum_{j=1}^{N}\sqrt{\sum_{l=1}^{R}|\Delta X_{ij}^l|}\right)$ is to make the regression coefficients spatially contiguous and simultaneously reduce the number of nonzero coefficients in a group level. Fixing the value of the smoothing parameter α, the coefficients ΔX^l for the lth subject is a function of the parameter λ, which is denoted as $\Delta X^l(\lambda)$. The grouped CR-value of the ith effect and the jth voxel is defined as follows:

$$c_{ij} = max\{\lambda | \exists l \in [1..R]\ \Delta X_{ij}^l(\lambda) \neq 0\}. \qquad (7)$$

Different from the single-subject analysis, the grouped CR-value is the maximum of all possible values of the parameter λ that can regularize at least one subject's coefficient is nonzero.

In the paper, we assume the smoothing parameter α is given and fixed. One typical method for justifying a parameter of a prediction algorithm is cross-validation. Cross-validation favours a smoothing parameter giving highest accuracy. However, a highly accurate model does not necessarily correspond to a "true model" [19]. For noisy fMRI data, for example, cross-validation favour large smoothing parameters, realizing little smoothness in the derived image. To overcome this problem, we designed a parameter tuning procedure to find a plausible value of the smoothing parameter α. The values of α in our experiments are chosen using this procedure. The details of this procedure will be presented and discussed in the full version of this paper.

III. Results

Here we demonstrated the performance of our method using both the simulated data and a real fMRI dataset. Due to space limitation, we only showed the results for single-subject analysis. More results and discussion will be presented in the full version of this paper.

A. Simulation study

In this sub-section, we firstly describe the design of the simulation study. The design matrix we used throughout our paper contained two regressors: a task related regressor and its corresponding derivative, which are shown in Fig. 2 (a). Both the task correlated regressor and its derivative were incorporated in the linear model to generate simulated data, but we only focused on investigating the performance of discovering regression coefficient for the task correlated regressor.

The spatially contiguous regression coefficients ΔX_1 for the task correlated regressor were generated as shown in Fig. 2 (b). ΔX_1 was represented by a 18×12 matrix with 25 non-zero pixels (in the black rectangular) grouped locally together. Each pixel had its colour ranging from blue to red and value ranging from 0 to 1 with a blue-red (low-high) colour scale. ΔX_2 in response to the derivative regressor was generated in a similar way (not shown). To mimic highly noisy fMRI data, our simulation experiment included randomly generated noise from a Gaussian distribution with mean 0 and standard deviation 1.5. For signal detection, locally grouped non-zero pixels should be found to be more significant than others.

CR-value was compared with GLM P-values, TFCE [9] and PPM [10]. For GLM P-value results, we did not apply the false discovery rate correction or Bonferroni correction methods. This is because these corrections only help to choose significant pixels using P-values for multiple comparisons. As the rank of pixels resulting from P-values is not influenced, these methods do not change the selection of top significant pixels.

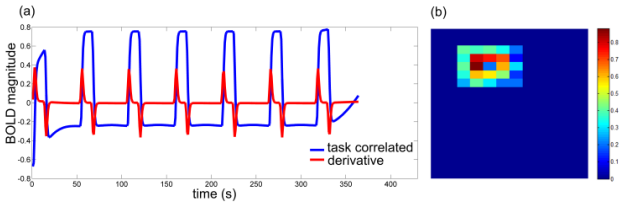

Figure 2. The design matrix and the regression coefficients for the simulation study.

Fig. 3 shows the single-subject analysis results. Subplots (a)-(c) show the results of selecting 25 pixels with the most significant GLM P-values, PPM-values and TFCE-values respectively; and subplots (d)-(i) show the results of selecting top 25 pixels using CR-values with the smoothing parameter α varying from 1 to 6. Intuitively, the pixels selected in subplots (a)-(c) are scattered; while some results of our method in subplots (d)-(i) with certain values of smoothing parameter α contain more locally grouped pixels that are from the black rectangular region in Fig. 2 (b). When the smoothing parameter α is equal to 1, the corresponding accuracy, specificity and sensitivity for CR-value were achieved as high as 0.98, 0.92 and 0.99 respectively. It is worth noting that the optimal value of smoothing parameter α can be selected by our parameter tuning procedure, which will presented in the full version of this paper.

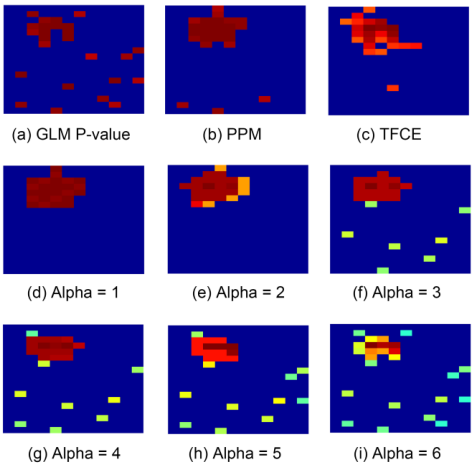

Figure 3. Single-subject analysis for simulation data. Subplots (a)-(i) are the top 25 pixels according to GLM P-values, PPM-values, TFCE-values and CR-values with the smoothing parameter α varying from 1 to 6. The ground truth is the black rectangular region in Fig. 2 (b).

B. Empirical study

The real fMRI dataset we used in this paper is taken from a recent study by [17]. 23 MS patients and 12 healthy volunteers

were studied with fMRI during a visuo-motor tracking task performed with the right hand. After 2 weeks of daily practise, the study was repeated. Here we aim to demonstrate our method rather than study the differences between healthy volunteers and patients, so we have analysed only the dataset for healthy controls. Please refer to V. Tomassini et al [17] for the details of the fMRI data acquisition and pre-processing protocols. As this real fMRI dataset was only used to demonstrate our method, we reduced its dimensionality to save computational time. The trilinear interpolation method was used here to down sample the dataset with the voxel size of 8mm in all three directions. Afterwards, we applied our method to find changes in functional brain regions by comparing the fMRI measurements between two different sessions. Our method was applied to the normalized fMRI measurements. The design matrix was same as the one used in the simulation study, reflecting the real dynamics of external motor practice. To save effort, in this study we showed the results when only the task correlated regressor was used.

Fig. 4 compares the results of applying GLM, TFCE and our method with a range of values of the smoothing parameter α. The regression coefficients indicated the strength of the relations between voxels and the design matrix rather than their significance. The GLM P-values are displayed in the second row (left) of Fig. 4 as a logarithmic transformation. The subplot next to the GLM P-value result in the second row is the scaled TFCE results (10 to the power of standardized TFCE-values). By comparing the subplots between the first two rows, we can see that voxels with large strength are not necessarily significant. Subplots from the third row to the last row in Fig. 4 shows the scaled CR-values (10 to the power of standardized CR-values) with the smoothing parameter varying from 2 to 16. As the smoothing parameter is reduced, CR-values tend to approximate results generated by GLM. When larger smoothness constraints (with consequent smaller smoothing parameter, α, values) were added, the significance values of neighbourhood voxels tended to be closer. As expected, some voxels were consistently more significant than others across different smoothness constraints.

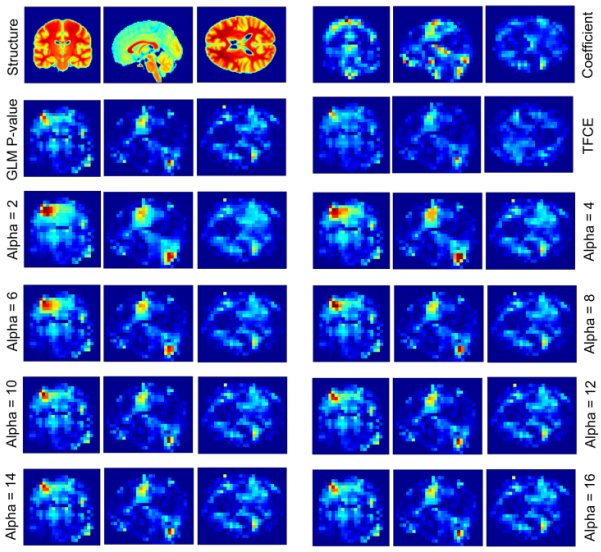

Figure 4. Single-subject analysis using coefficients, GLM P-values, TFCE-values and CR-values. The smoothing parameter α varies from 2 to 16. Alpha = X indicates the smoothing parameter α is equal to X.

IV. Conclusion

This paper introduced a method to find brain functional changes by comparing two task-related fMRI scans at different time points. The core of the method is a new significance indicator, which is known as CR-value. CR-value aims at improving sensitivity and accuracy of signal detection in the presence of substantial fMRI noise, which is introduced by variation on physical and physiological conditions at different time points. Both simulation study and empirical study demonstrated the improvement of signal detection compared with GLM P-value, PPM and TFCE.

References

[1] R. Bosnell, et al., Reproducibility of fMRI in the clinical setting: implications for trial designs. *Neuroimage*. vol. 42, pp. 603–610, 2008.

[2] R. A. Bosnell, et al., Motor practice promotes increased activity in brain regions structurally disconnected after subcortical stroke. *Neurorehabil. Neural Repair*. vol. 25, pp. 607–616, 2011.

[3] R. Liao, M. J. McKeown and J. L. Krolik, Isolation and minimization of head motion-induced signal variations in fMRI data using independent component analysis. *Magn. Reson. Med.* vol. 55, pp. 1396–1413, 2006.

[4] A. Floyer-Lea and P. M. Matthews, Distinguishable brain activation networks for short-and long-term motor skill learning. *J. Neurophysiol.* vol. 94, pp. 512–518, 2005.

[5] S. Cader, et al., Reduced brain functional reserve and altered functional connectivity in patients with multiple sclerosis. *Brain*. vol. 129, pp. 527–537, 2006.

[6] E. A. Rabiner, et al., Pharmacological differentiation of opioid receptor antagonists by molecular and functional imaging of target occupancy and food reward-related brain activation in humans. *Mol. Psychiatry*. vol. 16, pp. 826–835, 2011.

[7] B. A. Völlm, et al., Methamphetamine activates reward circuitry in drug naive human subjects. *Neuropsychopharmacology*. vol. 29, pp. 1715–1722, 2004.

[8] R. A. Poldrack, J. A. Mumford and T. E. Nichols, *Handbook of functional MRI data analysis*. Cambridge University Press, 2011.

[9] S. M. Smith and T. E. Nichols, Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*. vol. 44, pp. 83–98, 2009.

[10] W. D. Penny, N. J. Trujillo-Barreto and K. J. Friston, Bayesian fMRI time series analysis with spatial priors. *Neuroimage*. vol. 24, pp. 350–362, 2005.

[11] K. J. Friston, et al., Statistical Parametric Maps in Functional Imaging: A General Linear Approach. *Hum. Brain Mapp.* vol. 2, pp. 189–210, 1994.

[12] H. Attias, A variational Bayesian framework for graphical models. *Advances in Neural Information Processing Systems*. vol. 12, pp. 209–215, 2000.

[13] J. O. Berger, *Statistical Decision Theory and Bayesian Analysis*. Springer, 1985.

[14] S. Ji, Y. Xue and L. Carin, Bayesian compressive sensing. *IEEE Trans. Signal Process*.s vol. 56, pp. 2346–2356, 2008.

[15] M. E. Tipping, Sparse Bayesian Learning and the Relevance Vector Machine. *Journal Mach. Learn. Res*. vol. 1, pp. 211–244, 2001.

[16] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[17] V. Tomassini, et al., Relating brain damage to brain plasticity in patients with multiple sclerosis. *Neurorehabil. Neural Repair* vol. 26, pp. 581–93, 2012.

[18] A. Argyriou, T. Evgeniou and M. Pontil, Convex multi-task feature learning. *Mach. Learn.* vol. 73, pp. 243–272, 2008.

[19] K. Murphy, *Machine Learning: a Probabilistic Perspective*. MIT press, 2012.