

# Multivariate Statistics for Detection of MS Activity in Serial Multimodal MR Images

Sylvain Prima, Douglas L. Arnold, and D. Louis Collins

McConnell Brain Imaging Centre, Montreal Neurological Institute  
3801 University Street, Montreal, Quebec, Canada  
{prima,doug,louis}@bic.mni.mcgill.ca

**Abstract.** We present multivariate statistics to detect intensity changes in longitudinal, multimodal, three-dimensional MRI data from patients with multiple sclerosis (MS). Working on a voxel-by-voxel basis, and considering that there is at most one such *change-point* in the time series of MR images, two complementary statistics are given, which aim at detecting disease activity. We show how to derive these statistics in a Neyman-Pearson framework, by computing ratios of data likelihood under null and alternative hypotheses. Preliminary results show that it is possible to detect both lesion activity and brain atrophy in this framework.

## 1 Introduction

In a previous work [1], we proposed to consider the detection of disease activity in MRI as a *change-point problem*. After spatial and intensity normalization of MR data acquired on a given patient over time, we proposed to apply a one-sided [2] and a two-sided [3] univariate statistical test for the detection of at most one change-point in the intensity profile of each image voxel over time. Such a change-point was hypothesized to convey an actual biological change, eventually related to MS activity. The first test aimed at detecting directional changes (increase or decrease of the intensity), whereas the second one aimed at detecting changes in either of the two directions. The limitation of these tests is twofold. First, they assume that the variance of the observations is known. Second, their application is limited to monomodal images, whereas complementary information about the pathology is often available in multimodal data (classically,  $T_1$ -,  $T_2$ -weighted and  $PD$  MR images). In this paper, we show how to derive two complementary multivariate statistics which do not assume that the variance matrix is known. These statistics are given in Section 2.2, after the description of the preprocessing pipeline in Section 2.1. Preliminary results are presented in Section 3.

## 2 Methods

### 2.1 Preprocessing Steps

Before voxel-by-voxel statistical analysis of a time series of MR images, spatial and intensity normalization must be performed, to reduce intensity variations

due to imaging artefacts and to ensure spatial correspondence of the voxels under study. These preprocessing tasks have been extensively described in our previous work [1]. The whole MRI analysis pipeline is summarized in Figure 1. Briefly, it consists in:

- Intensity non-uniformity correction [4].
- Intensity normalization [1].
- Affine registration in the stereotaxic space [5].

## 2.2 Statistical Analysis

**Problem Formulation.** In this paper, we extend our previous work [1] by proposing two multivariate tests for the detection of a unique change-point. After the MR data have been spatially- and intensity-normalized, we perform a statistical analysis on a voxel-by-voxel basis (see Figure 2). Let  $x_i$  be the  $p$ -dimensional vector gathering the information available at time  $i$  for a given voxel (for example,  $p = 3$  if we have 3 modalities). In a probabilistic framework, the vector  $x_i$  can be seen as the realization of a random variable. For the sake of simplicity, this random variable and its realization will be named the same way ( $x_i$ ) in the following.

It is common to assume that  $x_i$  has a normal distribution with mean  $\mu_i$  (depending on the brain structure the voxel belongs to) and covariance matrix  $\Sigma$ , considered as unknown but common to all the  $x_i$ 's (this matrix mostly conveying the image acquisition noise). This can be summarized as:  $x_i \sim N(\mu_i, \Sigma)$ . Given these hypotheses, an active pathological process occurring at this voxel is likely to translate into a change in the mean  $\mu_i$  (for example, a white matter area

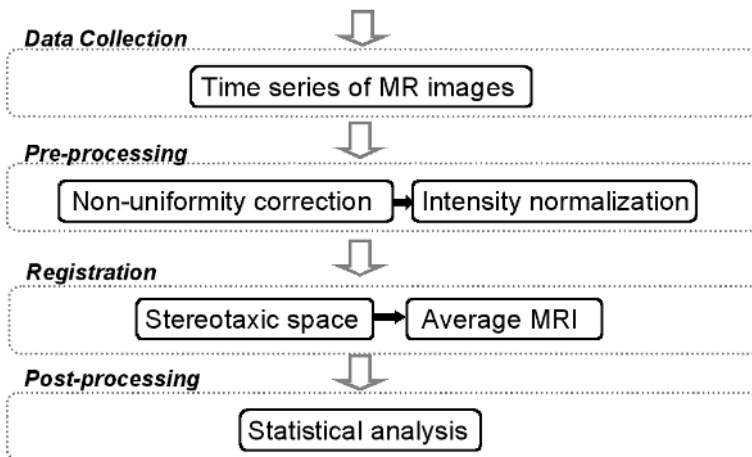


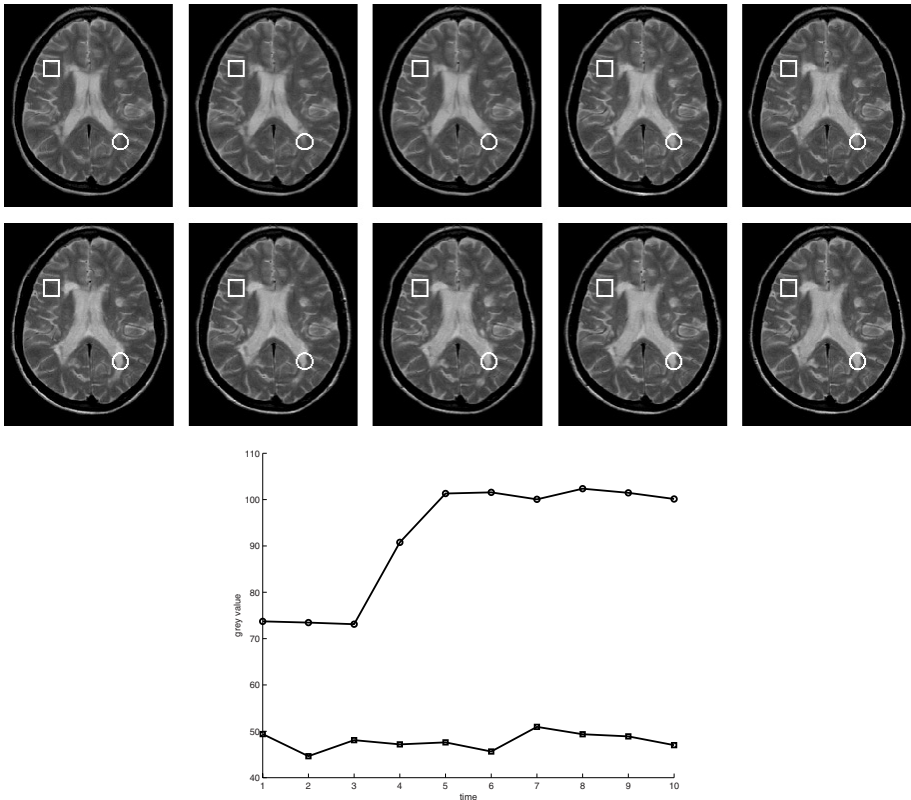
Fig. 1. MRI analysis pipeline.

becomes lesional after time  $m$ ). In the Neyman-Pearson framework, reasonable null ( $H_0$ ) and alternative ( $H_1$ ) can then be simply stated as follows:

$$\begin{aligned}
 H_0: & \quad \mu_i = \mu, & i = 1, \dots, n \\
 H_1: & \quad \mu_i = \mu, & i = 1, \dots, m \\
 & \quad \mu_i = \mu^* \neq \mu, & i = m + 1, \dots, n
 \end{aligned}$$

$\mu$  and  $\mu^*$  are unknown mean vectors before and after the unknown change-point  $m$ . In the following, we show how to derive statistics to test  $H_0$  against  $H_1$ . Two approaches are used: the likelihood ratio (LR) statistic and a Bayesian statistic are given in the next two sections.

**Likelihood Ratio Statistic.** Under  $H_0$ ,  $\mu$  and  $\Sigma$  are unknown. Under  $H_1$ ,  $\mu^*$  and  $m$  are additional unknown nuisance parameters. Let  $L_0(x)$  (resp.  $L_1(x)$ ) be



**Fig. 2.** A time series of registered  $T_2$ -weighted MR images. Top: the white square and white circle indicate, respectively, an area in the white matter that visually seems unaffected, and an area close to the lateral ventricles where a lesion appears over time. Left: the intensity profiles of the centers of these two areas indicate, respectively, a flat profile conveying image noise, and an intensity jump after time-point 3.

the likelihood of the data  $x = (x_1, \dots, x_n)$  under  $H_0$  (resp.  $H_1$ ). Both  $L_0(x)$  and  $L_1(x)$  depend on the nuisance parameters. As  $H_0$  is a nested hypothesis within  $H_1$ , computing the (sometimes referred to as *generalized*) likelihood ratio is a natural way to derive a statistic to test  $H_0$  against  $H_1$ . The LR is classically defined as the ratio between the *profile likelihoods*  $L_h^p(x)$  of the data [6] under both hypotheses, which can be written as:

$$LR = \frac{L_0^p(x)}{L_1^p(x)} = \frac{\sup_{\mu, \Sigma} L_0(x; \mu, \Sigma)}{\sup_{\mu, \mu^*, \Sigma, m} L_1(x; \mu, \mu^*, \Sigma, m)} \tag{1}$$

Computing the LR consists in replacing the unknown parameters by their maximum likelihood estimates under the null and the alternative hypotheses in the numerator and the denominator, respectively. Intuitively, a LR close to 1 (resp. 0) implies that  $H_0$  is fully consistent (resp. inconsistent) with the sample information. A test based on the LR test is intuitively sensible, and even optimal for simple hypotheses. Most standard tests are LR tests (*t*-, *F*-tests, *etc.*). In our case, the two hypotheses are composite, and the LR test is not necessarily optimal. If the vectors  $x_1, \dots, x_n$  are independent random variables (which is an hypothesis we make in this paper), it can be shown that the LR leads to a test using the following statistic [7]:

$$T^2 = \max_{m=1, \dots, n-1} T_m^2, \text{ with } T_m^2 = y'_m W_m^{-1} y_m \tag{2}$$

where:

$$y_m = \frac{m(n-m)}{n} (\bar{x}_m - \bar{x}_m^*), \text{ with } \bar{x}_m = \frac{1}{m} \sum_{i=1}^m x_i \text{ and } \bar{x}_m^* = \frac{1}{n-m} \sum_{i=m+1}^n x_i$$

$$W_r = \frac{1}{n-2} \left( \sum_{i=1}^m (x_i - \bar{x}_m)(x_i - \bar{x}_m)' + \sum_{i=m+1}^n (x_i - \bar{x}_m^*)(x_i - \bar{x}_m^*)' \right)$$

The exact distribution of  $T^2$  under  $H_0$  is known in the univariate case ( $p = 1$ ) [8, 9]. In the multivariate case, the distribution appears to be intractable. The classical asymptotic theory would imply a  $\chi^2$  limiting distribution for  $-2 \log LR$ , as  $n \rightarrow \infty$ . This theory is not applicable here, as the likelihood function (under  $H_1$ ) is not continuously differentiable in  $m$ . Instead, the simple Bonferroni inequality can be applied, as the distribution of each  $T_r^2$  can be computed [7].

**Bayesian Statistic.** An alternative, Bayesian approach to the LR has been proposed to test  $H_0$  against  $H_1$ . It consists in selecting *a priori* distributions  $p(\cdot)$  for the nuisance parameters. Then, the *integrated likelihoods*  $L_h^i(x)$  of the data [6] under both hypotheses are computed, by integrating  $L_0(x)$  and  $L_1(x)$  over these nuisance parameters. Finally, the ratio of the integrated likelihoods, called Bayesian ratio (BR), is computed; its interpretation is close to that of the

LR given in the previous section. The BR can be written:

$$BR = \frac{L_0^i(x)}{L_1^i(x)} = \frac{\int_{\mu, \Sigma} L_0(x|\mu, \Sigma)p(\mu)p(\Sigma)d\mu d\Sigma}{\sum_m \int_{\mu, \mu^*, \Sigma} L_1(x|\mu, \mu^*, \Sigma, m)p(\mu)p(\mu^*)p(\Sigma)p(m)d\mu d\mu^* d\Sigma} \quad (3)$$

Two univariate approximate tests proposed in our previous work [1] were derived based on the BR, with  $\Sigma$  being considered fixed. Following the same procedure, considering the covariance matrix to be known (assumed to be the identity matrix, without loss of generality) and provided the  $x_1, \dots, x_n$  are independent, an analogous approximate two-sided multivariate test can be derived [10], whose statistic can be written as:

$$U = \frac{1}{n^2} \sum_{i=1}^{n-1} \left( \sum_{j=i}^{n-1} (x_{j+1} - \bar{x}) \right)' \left( \sum_{j=i}^{n-1} (x_{j+1} - \bar{x}) \right) \quad (4)$$

The exact distribution of  $U$  can be computed [10]. However, the covariance matrix is generally not known. To overcome this difficulty in the univariate case, Sen and Srivastava [11,12] have proposed to divide the Gardner's and the Chernoff and Zack's statistics [2,3] by an unbiased estimate of the variance. The sample variance could be used, but in case of a change, its value will be larger than the true searched value. Instead, the mean square successive difference, originally proposed by von Neumann [13], is much less sensitive to such intensity variations. Its expression is:

$$\delta^2 = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2$$

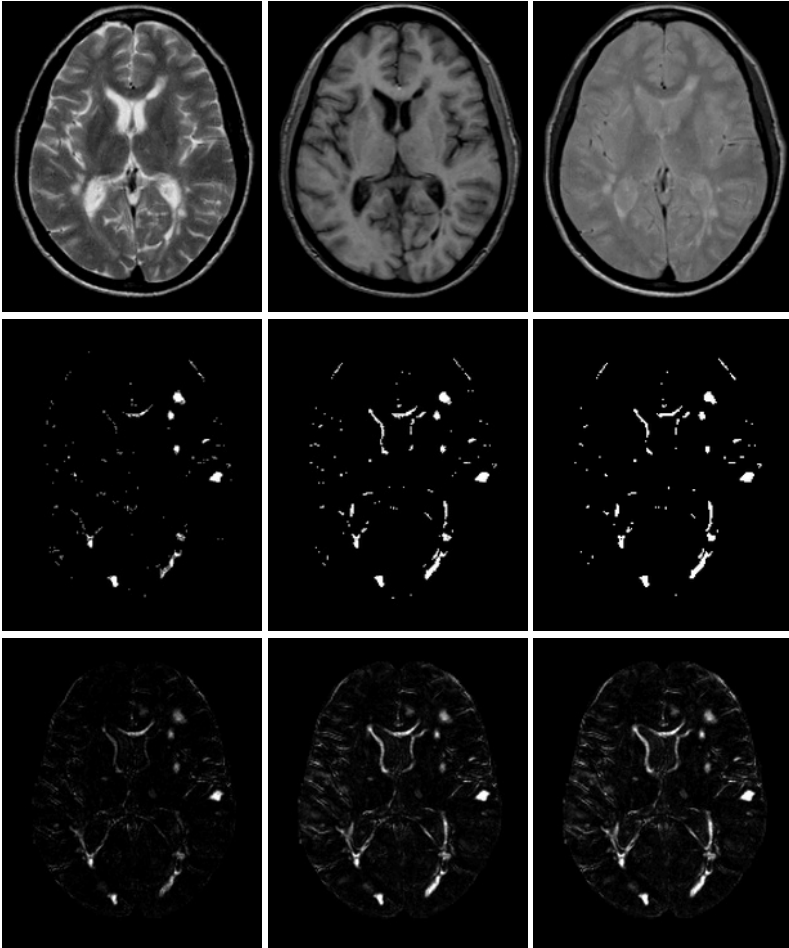
The normalized Gardner's and the Chernoff and Zack's statistics have known distributions [11,12]. In the multivariate case, an analogous mean square successive difference  $\Delta$  can be computed.  $\Delta$  is an unbiased estimate of  $\Sigma$  that reads:

$$\Delta = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (x_{i+1} - x_i)(x_{i+1} - x_i)'$$

By analogy with the univariate case, we propose that  $U$  can be normalized by using  $\Delta$ , which yields the following statistics:

$$V = \frac{1}{n^2} \sum_{i=1}^{n-1} \left( \sum_{j=i}^{n-1} (x_{j+1} - \bar{x}) \right)' \Delta^{-1} \left( \sum_{j=i}^{n-1} (x_{j+1} - \bar{x}) \right)$$

Unfortunately, to our knowledge, no exact, approximate or asymptotic formula is known for the distribution of  $V$ . Thus, in Section 3, we use the statistics  $T^2$  and  $U$ , respectively defined in Equations 2 and 4, and whose distributions can be at least approximated, which allows to compute significance levels.



**Fig. 3.**  $T^2$  and  $U$  statistics applied on a time series. First row: transverse MR images of a patient with MS at the level of the lateral ventricles; the first time point of the series for each of the 3 modalities is displayed (from left to right,  $T_2$ ,  $T_1$ ,  $PD$ ). Second row:  $T^2$  statistic applied on the series, using only  $T_2$  data (left),  $T_2$  plus  $T_1$  (middle), and the 3 modalities (right). Third row: same display as the second row for the statistic  $U$ . Saturated white voxels are significant at the 0.1 level.

### 3 Results

In this section, we give preliminary results on a time series of multimodal images. Ten image volumes over a four year period were acquired on a patient with very active disease, at the MS clinic of the Montreal Neurological Institute. We applied the two statistics  $T^2$  and  $U$  on this data, and the results are displayed in Figure 3. Saturated white voxels are significant at the 0.1 level. For both statistics, using more modalities seems to improve the detection of disease activity. Qualitatively,

the detected voxels are mainly located around the ventricles, where most of the lesions are. These significant voxels also convey the brain atrophy that occurs during the MS course, which translates into a decrease of ventricular size.

In the univariate case, it has been experimentally demonstrated that both statistics perform best (*i.e.*, have their best power) when  $m$  is close to  $n/2$  [14]. At this value,  $U$  is superior to  $T^2$ . Otherwise, most of the time,  $T^2$  performs better than  $U$ . To our knowledge, no experiment has been led to compare the power of these statistics in the multivariate case. However, a qualitative interpretation of Figure 3 suggests that this conclusion may be also valid in the multimodal case. For example, a active lesion in the right frontal lobe is much well detected with  $T^2$  than with  $U$ . It turns out that this lesion is only present at the first time point of the series: the change-point occurs very early, which favors  $T^2$ . Further experiments will be necessary to determine which statistic, on average, is the best.

## 4 Conclusion and Future Work

In this paper, we have presented two complementary multivariate statistics to detect intensity changes in longitudinal, multimodal, three-dimensional MRI data. A preliminary result has been presented, which suggests that is possible to detect both lesion activity and brain atrophy in this framework. Qualitative comparison of the two statistics has been given, but further experiments will be necessary to determine which statistic, on average, performs best. The significance levels of the statistical maps we obtained should also be corrected for multiple comparisons across all voxels of the volume. Bonferroni correction, or less conservative approximations [15] could be used. These statistics can be viewed as activity indices whose use, together with global or local atrophy metrics, may result in a better surrogate of disease activity, with potential applications for better diagnosis, prognosis and treatment of the disease.

## References

1. S. Prima, N. Ayache, A. Janke, S.J. Francis, D.L. Arnold, and D.L. Collins. Statistical Analysis of Longitudinal MRI Data: Applications for Detection of Disease Activity in MS. In T. Dohi and R. Kikinis, editors, *Fifth International Conference on Medical Image Computing and Computer-Assisted Intervention, MICCAI'2002*, volume 2488-9 of *Lecture Notes in Computer Science*, pages 363–371, Tokyo, Japan, September 2002. Springer.
2. H. Chernoff and S. Zacks. Estimating the Current Mean of a Normal Distribution which is Subjected to Changes in Time. *Annals of Mathematical Statistics*, 35(3): 999–1018, September 1964.
3. L.A. Gardner. On Detecting Changes in the Mean of Normal Variates. *Annals of Mathematical Statistics*, 40(1): 116–126, February 1969.
4. J.G. Sled and A.P. Zijdenbos. A Nonparametric Method for Automatic Correction of Intensity Nonuniformity in MRI Data. *IEEE Transactions on Medical Imaging*, 17(1): 87–97, February 1998.

5. D.L. Collins, P. Neelin, T.M. Peters, and A.C. Evans. Automatic 3D Intersubject Registration of MR Volumetric Data in Standardized Talairach Space. *J. of Computer Assisted Tomography*, 18(2): 192–205, March 1994.
6. J.O. Berger, B. Liseo, and R.L. Wolpert. Integrated likelihood methods for eliminating nuisance parameters. *Statistical Science*, 14(1): 1–28, 1999.
7. M.S. Srivastava and K.J. Worsley. Likelihood Ratio Tests for a Change in the Multivariate Normal Mean. *Journal of The American Statistical Association*, 81(393): 199–204, March 1979.
8. D.M. Hawkins. Testing a Sequence of Observations for a Shift in Location. *Journal of The American Statistical Association*, 72(357): 180–186, March 1977.
9. K.J. Worsley. On the Likelihood Ratio Test for a Shift in Location of Normal Populations. *Journal of The American Statistical Association*, 74(366): 365–7, June 1979.
10. A.K. Sen and M.S. Srivastava. On Multivariate Tests for Detecting Change in Mean. *Sankhya: The Indian Journal of Statistics*, 35(2): 173–186, 1973.
11. A.K. Sen and M.S. Srivastava. On Tests for Detecting Change in Mean When Variance Is Unknown. *Annals of the Institute of Statistical Mathematics*, 27: 479–486, 1975.
12. A.K. Sen and M.S. Srivastava. Some One-Sided Tests for Change in Level. *Technometrics*, 17: 61–64, 1975.
13. J. von Neumann, R.H. Kent, H.R. Bellinson, and B.I. Hart. The mean square successive difference. *Annals of Mathematical Statistics*, 12: 153–62, 1941.
14. A.K. Sen and M.S. Srivastava. On Tests for Detecting Change in Mean. *Annals of Statistics*, 3(1): 98–108, January 1975.
15. K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach or determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4: 58–73, 1996.