

A statistical method (cross-validation) for bone loss region detection after spaceflight

Qian Zhao · Wenjun Li · Caixia Li ·
Philip W. Chu · John Kornak · Thomas F. Lang ·
Jiqian Fang · Ying Lu

Received: 27 February 2010 / Accepted: 6 June 2010 / Published online: 15 July 2010
© The Author(s) 2010. This article is published with open access at Springerlink.com

Abstract Astronauts experience bone loss after the long spaceflight missions. Identifying specific regions that undergo the greatest losses (e.g. the proximal femur) could reveal information about the processes of bone loss in disuse and disease. Methods for detecting such regions, however, remains an open problem. This paper focuses on statistical methods to detect such regions. We perform statistical parametric mapping to get t -maps of changes in images, and propose a new cross-validation method to select an optimum suprathreshold for forming clusters of pixels. Once these candidate clusters are formed, we use permutation testing of longitudinal labels to derive significant changes.

Keywords Suprathreshold · Cross-validation (CV) · Medical imaging · Permutation test

Introduction

Multiple images (e.g. X-rays) are often collected in epidemiological, medical, and other kinds of research in subjects over time or between conditions. Often the goal is to determine when and where statistically significant differences occur between conditions. Before evaluating changes between images, all images are co-registered both within and between subjects, e.g. to a single template [1–4]. Then, a statistical map [5–7] is created, consisting of t - (2 conditions) or F -statistics (>2 conditions) at each image unit, typically a pixel (2-dimensional) or voxel (3-dimensional). While direct comparison of the t -maps or F -maps, i.e. of changes of individual pixels, is possible using statistical parametric mapping (SPM) [8], such comparison may suffer from low statistical power after proper adjustment for multiple comparisons with family-wise error. More importantly, those pixels (or voxels) with significant changes can be distributed sparsely or be clinically or biologically irrelevant to a given application. Instead, a cluster of contiguous pixels or voxels is usually more informative and robust, in particular for the study of bone changes due to altered weight bearing conditions [9].

As an alternative, a suprathreshold cluster analysis (STCA) [10] determines the statistical significance of clusters with changes beyond a suprathreshold. STCA includes the following steps: First, it selects regions of interest (ROI), which are clusters of contiguous pixels with t - or F -values usually above the 95th percentile of the empirical distribution of the observed t - or F -statistics. Second, it uses permutation tests and selected cluster

Q. Zhao · C. Li · J. Fang (✉)
Department of Medical Statistics and Epidemiology, School
of Public Health, Sun Yat-Sen University,
Guangzhou 510080, China
e-mail: fangjq@mail.sysu.edu.cn

Q. Zhao · W. Li · C. Li · P. W. Chu · J. Kornak · T. F. Lang
Department of Radiology and Biomedical Imaging, University
of California, San Francisco, CA 94143-0946, USA

J. Kornak
Department of Epidemiology and Biostatistics, University
of California, San Francisco, CA, USA

T. F. Lang
UCSF/UC Berkeley Joint Graduate Group in Bioengineering,
University of California, San Francisco, CA, USA

Y. Lu
Cooperative Studies Program Coordinating Center, VA Palo
Alto Health Care System, Palo Alto, CA 94304-1290, USA

Y. Lu (✉)
Department of Health Research and Policy,
Stanford University, Stanford, CA 94305-5505, USA
e-mail: ylu1@stanford.edu

features (e.g. “size”) [5] to determine the family wise statistical significance of candidate clusters (ROIs). STCA has been successfully applied in neurological studies [11–13]. The main advantages of STCA over SPM are to avoid selecting isolated pixels due to extreme values and to reduce the number of comparisons from pixels to ROIs. By determining the ROIs, the chosen threshold heavily influences the overall conclusions [13, 14]. STCA still uses a fixed “primary” threshold to construct clusters and then determines their significance. The consistency of current method in selecting these thresholds, however, is not ideal [15]. Smith and Nichols [16] proposed an alternative to a fixed threshold approach using the average of p values from all possible thresholds according to their distributional weights as the summary p values [16]. However, their method is beyond scope of this paper, and debates about selecting thresholds and testing for differences continue [13, 14, 16].

The aims of our paper are to propose a new cross-validation (CV) method to select the optimal threshold for forming candidate clusters, and to assess the significance of those clusters via a permutation test. We demonstrate our new method with an application to a study of accelerated bone loss of astronauts during long-term spaceflight.

The paper is organized in the following way. First, we provide background to the astronaut pre- and post-spaceflight study that motivated this project (Sect. 2). Using the terminology set in Sect. 2, we present detailed descriptions of our theory and methods (Sect. 3), which we then apply to the astronaut data (Sect. 4). We present our conclusions and discussion in the final section.

A study of bone loss during long-duration spaceflight

A longitudinal study of bone loss

Our research was motivated by a study of accelerated bone loss of astronauts during spaceflight [17, 18]. Astronauts experience localized bone mineral loss during extended periods of weightlessness, for example in the proximal femur [18]. Methodologies for detecting regions that experience greatest bone loss due to spaceflight may inform the study of changes in bone density due to long-term physical inactivity, aging, disease, drug treatment, and other causes. We present this study first to provide details of STCA in order to understand our improvements in Sect. 3.

Quantitative computed tomography (QCT) scans of the hip were taken for 16 astronauts (44.6 ± 4.0 years old) prior to and after their 4–6 months spaceflight on the international space station (ISS). The study protocol was approved by the institutional review boards (IRBs) of the national aeronautics and space administration (NASA), Baylor College of Medicine, and the University of California, San Francisco (UCSF).

Pre-flight scans were performed 30–60 days prior to launch, and post-flight scans were performed within 7–10 days of landing. Helical CT images (GE Hispeed Advantage GE Medical System, Milwaukee, WI) were acquired at Methodist Hospital, Baylor College of Medicine, at a scan setting of 80 KVp, 280 mAs, 3-mm slice thickness, helical pitch of 1, and in-plane spatial resolution of 0.9375 mm. The pre- and post-flight scans of the 16 astronauts were co-registered (including rigid and non-rigid co-registration) to a common reference space so that the homologous tissue elements could be compared [19, 20]. After image co-registration, one middle coronal slice with $114 \times 151 = 17,214$ pixels in each scan was used for this study. Bone mineral density (BMD) was measured in Hounsfield Units (HU, a quantitative measure of radiodensity) under pre-flight (A) and post-flight (B) conditions, and then the matched pixel differences were compared between the two conditions. For this study, our analysis focused on the region of the proximal femur, which consisted of 3,948 pixels. Though we only have access to 2-D data, the methods described in this paper can be equally extended to 3-D voxel data.

Generation of space-flight statistical parametric maps

Let $A_i(k, l)$ and $B_i(k, l)$ be the pre- and post-flight BMD measured in HU, respectively, at pixel coordinate (k, l) for astronaut i ($i = 1, 2, \dots, I, I = 16$). In this study we assume that only bone loss (not gain) occurs during spaceflight and therefore only consider one-sided differences A–B. Hence, a difference image or difference map between pre- and post-flight scans is computed for individual astronauts at each pixel (k, l) as

$$\Delta_i(k, l) = A_i(k, l) - B_i(k, l) \quad (i = 1, \dots, 16). \quad (2.1)$$

General formulas for the mean and variance of difference images are the following:

$$\bar{\Delta}(k, l) = \frac{1}{I} \sum_{i=1}^I \Delta_i(k, l), \quad (2.2)$$

and

$$S^2(k, l) = \frac{1}{I-1} \sum_{i=1}^I (\Delta_i(k, l) - \bar{\Delta}(k, l))^2 \quad (2.3)$$

where I is the total number of astronauts. With the assumption that the true error variance is spatially smooth, we use a smoothed variance estimator, S^2 , based on the Gaussian kernel of full width at half-maximum (FWHM 1.5 pixel) to decrease the noise in variance estimation [21].

Our objective function here is the mean difference of two conditions with mean statistics or t -statistics. Using these smoothed variance estimates, we calculate pixel-level pseudo t -statistics as [10, 21]

$$T(k, l) = \bar{\Delta}(k, l) / \sqrt{S^2(k, l)}. \tag{2.4}$$

While the pseudo t -statistics used in this paper as an index of change are not independent between pixels, the conventional approach approximates them to t -statistics with degrees of freedom of $I - 1$. In the remainder of the paper we drop the “pseudo” qualifier and denote T as the range of t -statistics in T_0 .

Optimum suprathreshold selection

Cluster forming

A cluster is a set of spatially connected pixels sharing similar features and based on a t -map of I subjects. We consider a cluster as a set of connected pixels with $\{(k, l): T(k, l) \geq u\}$, where u is a certain threshold. The connected neighbour region of pixel (k, l) is $\{(k - 1, l - 1), (k, l - 1), (k + 1, l - 1), (k - 1, l), (k + 1, l), (k - 1, l + 1), (k, l + 1), \text{ and } (k + 1, l + 1)\}$.

Many discrete clusters can be formed within a t -map, even including a single isolated pixel. By altering the threshold u , we change the number and size distribution of clusters identified in a t -map. By determining which clusters become candidates for significance testing, threshold selection can have a strong influence on the results of any image analysis.

Cross-validation

Researchers face a constant challenge in trying to identify valid thresholds for constructing candidate clusters [10, 22]. Although a common approach arbitrarily uses the 95th percentile in t -statistics, there is no algorithm to provide an automatic threshold selection strategy to systematically identify candidate clusters. Clusters that experience true bone loss between conditions A and B should have higher values of $T(k, l)$. Thus, the optimum clusters derived from the current data also should have the largest mean difference value Δ for future astronauts in the same bone region.

We therefore propose the use of cross-validation methods to choose the optimum suprathreshold $u_c^* \in T$. The basic idea of cross-validation (CV) is to randomly split a data set D (of total size I) into K mutually exclusive subsets D_1, D_2, \dots, D_K of approximately equal size. The clusters based on a threshold u_c are then formed using $K - 1$ subsets by excluding D_i (denoted as $D \setminus D_i$). We can test the effect of these newly formed clusters on the excluded subset D_i that was not used to construct the clusters. Repeating the procedure K times, with each subset used exactly once for validation, constitutes a K -fold CV [23]. A

tenfold CV [23–25] is often considered sufficient. When K equals I , the number of observations in the original sample D , the procedure is known as leave-one-out cross-validation (LOOCV). Here, the full dataset D is the set of all images from all astronauts pre- and post-flight.

To expedite the search for the optimal suprathreshold, we first define a search range for u_c . Let u_L and u_H be the low and high bounds for u_c , respectively. We begin with an initial threshold of $u_1 = u_L$ and follow by iteration at $u_n = u_{n-1} + \Delta u$ until we reach u_H . Here Δu is an acceptable tolerance for error in the optimal u . In our example, we defined the 80th–99th percentiles from the original distribution of T as the search range and used a half percentile for Δu .

Proposed procedure for u_c^*

Our procedure for selecting the optimal suprathreshold u_c^* is as follows (Shown in Fig. 1):

- Step 0: Create the t -map T_0 for the full dataset D .
- Step 1: Partition the data D into K mutually exclusive subsets $D_i, i = 1, 2, \dots, K$.
- Step 2: Leave out subset D_i and use $D \setminus D_i$ to create the t -map T_{-i} . For the current u_n , define candidate clusters as all clusters C_j with t -statistics above u_n in T_{-i} . Calculate mean difference for all pairs of pixels in the clusters or ROIs(i, u_n) for subset D_i :

$$\text{ROIs}(i, u_n) = \{C_j : \forall (k, l) \in C_j, T_{-i}(k, l) \geq u_n\} \tag{3.1}$$

$$m_i(u_n) = \frac{\sum_{(k,l) \in \text{ROIs}(i,u_n)} \Delta_i(k, l)}{|\text{ROIs}(i, u_n)|} \tag{3.2}$$

where $\Delta_i(k, l)$ is defined in (2.1) and $|\text{ROIs}(i, u_n)|$ is the number of pixels in ROIs(i, u_n).

- Step 3: Repeat Steps 2 to 4 until the K th sample has been excluded.
- Step 4: Get the summary cross-validation (CV) statistics for the K models as objective function:

$$\text{CV}(u_n) = \frac{1}{K} \sum_{i=1}^K m_i(u_n) \tag{3.3}$$

Take u_0 to be the 50th percentile of T_0 as a baseline, normalize $\text{CV}(u_n)$ to accommodate for differences in scale between images using $\text{CV}(u_0)$ as:

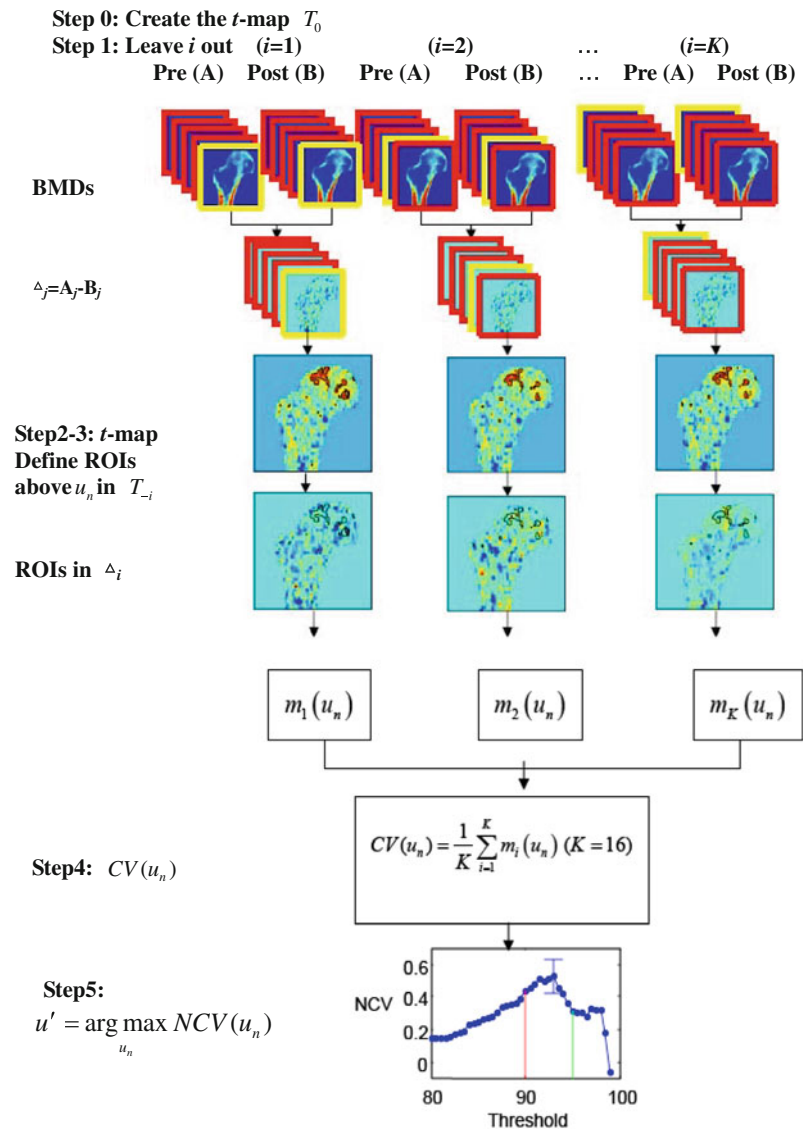
$$\text{NCV}(u_n) = \frac{\text{CV}(u_n) - \text{CV}(u_0)}{\text{CV}(u_0)} \tag{3.4}$$

- Step 5: Repeat Steps 2 to 7 over all candidate clusters (ROIs) and derive

$$u' = \arg \max_{u_n} \text{NCV}(u_n) \tag{3.5}$$

and finally choose

Fig. 1 Graphical illustration of the cross-validation procedure



$$u_c^* = \min \{u : NCV(u) \geq NCV(u') - SE(NCV(u'))\}$$

$$u_c^* = \min \{u : NCV(u) \geq NCV(u') - SE(NCV(u'))\} \tag{3.6}$$

where $SE(NCV(u'))$ is the standard error (SE) of $NCV(u')$ K -fold CV samples.

Here, the statistics $m_i(u_n)$ are used for clusters with single or multiple pixels. In this application we use a mean difference instead of a mean of the t -statistics because we will use LOOCV, and the pooled standard deviation of D_i is not available. With sufficient numbers of subjects in the CV subsets (>2), t -statistics for each pixel can be calculated, and the mean of t -statistics can be used to replace $\Delta_i(k, l)$ in (3.2). This percentage improvement measure in (3.6) is unitless and less dependent on different constructions of CV subsets. Equation (3.6) is the 1-SE rule originally recommended by

Brieman et al. for CVs [25] and adopted by many authors in evaluations of CV errors [26, 27], in particular in recursive partitioning analysis. It recognizes that candidate thresholds within 1-SE range from the optimal u' in (3.5) most likely will result in comparable $NCV(u)$'s to the optimal $NCV(u')$. By lowering the suprathreshold to 1-SE in (3.6), we will get slightly larger size clusters with more stable feature statistics, yet not sacrifice the efficiency to measure changes. Past experience and simulation studies suggested that the 1-SE rule can screen out noise in finite sample CVs [26, 28].

Once we identify the optimal suprathreshold, the remaining challenge is to determine which clusters represent significant change beyond chance. Traditional permutation tests [5, 6] could be used to derive the permutation distribution, thereby eliminating the need to assume a

specific distribution for the test statistics [10, 21]. Consider two conditions pre- and post-flight as A and B, and data from I subjects follows as ABABAB... Then rearrange the labels randomly within subjects to get another sequence maybe as BA AB BA ... And a new t -map T_r could be derived for each time ($r = 1, 2, \dots, R$).

Cluster size is the most sensitive to choice of thresholds, which is a simple statistics counting the number of connected pixels (in cluster) above the threshold u ,

$$Y_j^S(u) = \sum_{(k,l) \in C_j(u)} 1_{T(k,l) \geq u}, \tag{3.7}$$

where $1_{T(k,l) \geq u}$ is a binary indicator for pixels with t -statistics above u . For each permutation time r , find the cluster above u_c^* with maximum size Z_r in T_r . For each cluster $C_j(u_c^*)$ in T_0 , calculate how many Z_r is larger or equal to the size of cluster C_j and get the empirical p value as $p_j = \#(Z_r \geq Y_{0j})/R$. If $p_j \leq \alpha$ (e.g., $\alpha = 0.05$), we

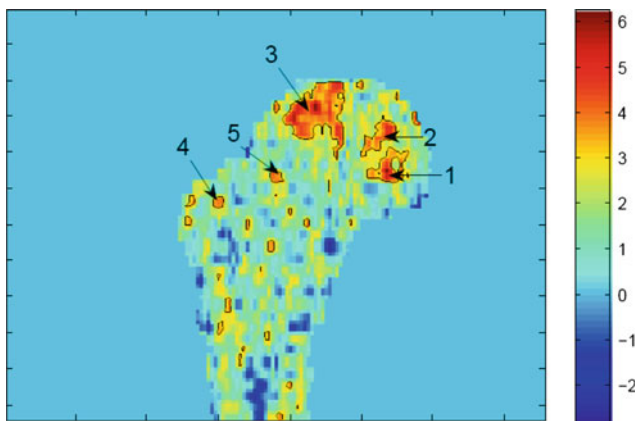
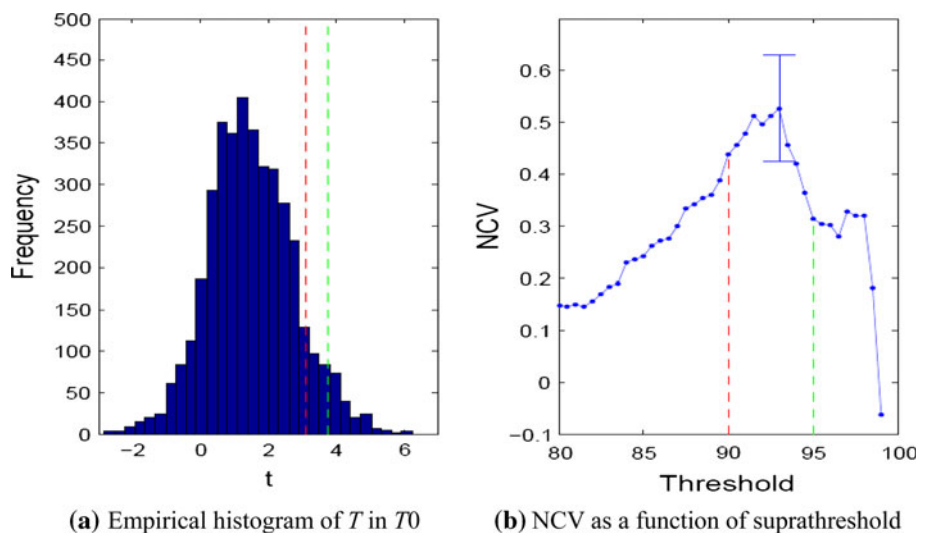


Fig. 2 Original T_0 and 28 candidate clusters by CV. Twenty-eight (28) candidate clusters were circled with red color for intensity in t statistics. The top five clusters are numbered in the figure

Fig. 3 Histogram of t -statistics T and NCV-threshold curve. Red and green dotted lines correspond to the 90th and 95th percentile t values, respectively, of T_0



conclude that cluster $C_j(u_c^*)$ displays statistically significant differences (in the feature statistics) between conditions A and B.

Application to the study of bone loss during long-duration spaceflight

As described in Sect. 2, scans of the hip were taken for 16 astronauts before and after their 4–6 months spaceflight. After co-registration of these images [19], paired t -statistics were calculated for each pixel to form an observed t -map T_0 in Fig. 2. This set of t -statistics had the empirical distribution T shown in Fig. 3a.

Because of the small sample size ($I = 16$ subjects), we used LOOCV (described in Sect. 3) to determine the optimal suprathreshold u_c^* , and a search range of 80th–99th percentiles and a tolerance level Δu of 0.5%. The maximum $NCV(u')$ was 0.52, achieved at the suprathreshold u' of 3.41, which corresponds to the 93rd percentile of the distribution of T . The standard error (SE) of $NCV(3.41)$ was 0.1 (Fig. 3b), which led to the optimum superathreshold u_c^* as 3.14, or the 90th percentile of T , according to Eq. (3.6).

With this optimal u_c^* , 28 clusters were formed within the original t -map T_0 , as represented by red regions in Fig. 2. Five major clusters ranged in size from 9 to 55 pixels. The remaining 23 smaller clusters had mean t -statistics 3.45 (SD 0.17) and mean size of 3.7 pixels. A permutation test based on 1,000 permutations was then used to yield p values for cluster size Y_j^S ($j = 1, \dots, 28$) that are shown in Table 1. Under control of type I error with $\alpha = 0.05$, we identified five clusters based on Y_j^S as sites with statistically significant bone loss. The remaining clusters were not statistically significant.

By comparison, using the conventional 95th percentile threshold for T produced 22 clusters. The p values for Y_j^S of

Table 1 Cluster size and p values for five major clusters and remaining smaller clusters

Method	Cluster #	Size	t values*		p value
			Mean	SD	
CV	1	55	3.84	0.67	0.002
	2	50	3.79	0.54	0.002
	3	185	4.04	0.62	0.001
	4	9	3.72	0.2	0.044
	5	11	3.69	0.29	0.021
Other $n = 23$	Mean	3.7	3.45	0.17	
	SD	1.89	0.23	0.14	
95%tile	1	20	4.51	0.66	0.002
	2	22	4.28	0.39	0.002
	3	113	4.4	0.52	0.001
	4	5	3.85	0.06	0.026
	5	5	3.94	0.18	0.026
Other $n = 17$	Mean	1.88	3.88	0.06	
	SD	0.86	0.11	0.07	

The cross-validation (CV) method used a suprathreshold 3.14, corresponding to a 90th percentile of the observed t -statistics

* t values were derived from the original t -map T_0 defined in Eq. (2.4)

these clusters are also presented in Table 1. Compared with the 95th percentile of T , the optimal suprathreshold u_c^* (90th percentile of T) produced the same significant results for cluster size for each cluster.

Discussion and conclusion

In this paper, we propose statistical improvements to the suprathreshold cluster analysis (STCA) framework for longitudinal image comparisons. While STCA has been used in neurological imaging research, particularly in functional brain imaging, its application to other imaging areas is less common. We hope our study of bone loss in astronauts during long-term spaceflight will support the general application of this statistical tool, and our extensions of it, to diverse biological systems.

As an alternative to STCA, Statistical Parametric Mapping (SPM) is a more commonly used method to compare longitudinal changes in images and identify clusters or ROIs. SPM assumes a Gaussian Random Field (GRF) and common variance structures across subjects, which are difficult to verify [21]. The main advantage of STCA is that it is a non-parametric method that does not require special assumptions about spatial or intra-subject longitudinal and biological correlation structures. Permutation tests have been widely used for high dimensional data, especially in genomics [29] and functional neurological image analysis [10, 30], and can be applied not only to longitudinal

changes in individuals, but also to (cross-sectional or longitudinal) comparisons of groups by permuting group assignments.

The main contribution of this paper was the use of cross-validation (CV) to select the optimum suprathreshold. CV methods are more independent of the type of input data and special image analysis than other methods that rely on untenable or unverifiable assumptions about statistical distributions. Comparing with traditional 95th percentile method, the clusters detected by CV tend to be with bigger size.

We performed a simulation study to demonstrate improved statistical power and efficiency of this method, which added artificial clusters with known intensity and size changes into an assigned region of the image. Compared with the conventional 95th percentile threshold, the clusters identified by CV tend to be larger in size. Especially for low intensity, the 95th percentile threshold sometimes divided one cluster into two or more sub-clusters, which decreased the homogeneity within clusters and resulted in insignificant changes. Results are not shown but are available from the authors.

In this application, we wanted to identify the specific location(s) of greatest bone loss within the hip during long-term spaceflight, and therefore used a one-sided change of bone loss as our CV metric. For more general longitudinal applications to a null hypothesis of no change and the alternative hypothesis of any change in either direction (e.g. gain or loss of bone), we can use the absolute t -map as the CV metric.

While the statistical methods of this paper can be used for either 2-D or 3-D images, our demonstration is confined to the 2-D (pixel-based) case. Extension to 3-D (voxel-based) images and other types of digital images, e.g. satellite remote sensing, photography, or astronomy, should be straightforward.

Astronauts incur bone loss during long-duration spaceflight, and it is reasonable to expect that the majority of bone loss occurs in areas that are subject to greatest mechanical stress under earth's gravity. Understanding the spatial heterogeneity in loss of proximal femoral bone tissue, in which the largest losses concentrate in the load-bearing subregions, is of interest to general mammalian biology as well as for the well-being of astronauts after their return to gravity. Some research has shown that bone adapts to earth's gravity by increasing the size of cortical bone but not necessarily trabecular bone [18]. Knowing the nature of most significant bone loss will help devise preventive measures during spaceflight as well as rehabilitation interventions post-spaceflight.

In summary, this paper proposed a cross-validation (CV) method to select the optimum suprathreshold and form candidate pixel clusters (or ROIs) of longitudinal changes

of images and provided one method solve the problem for a fixed “primary” threshold in STCA.

Acknowledgments The first author received a Chinese Scholarship for Joint Ph.D. Research, sponsored by China Scholarship Council to perform research at UCSF. The study is also supported by NASA grants NNJ04HC7SA and NNJ04HF78G.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Collins DL, Neelin P, Peters TM et al (1994) Automatic 3D intersubject registration of Mr. volumetric data in standardized talairach space. *J Comput Assist Tomogr* 18:192–205
- Collins DL, Holmes CJ, Peters RM et al (1995) Automatic 3-D model-based neuroanatomical segmentation. *Hum Brain Mapp* 3:190–208
- Collins DL, Evans AC (1997) Animal: validation and applications of nonlinear registration-based segmentation. *Int J Pattern Recogn Artif Intell* 11:1271–1294
- Studholme C, Hill DLG, Hawkes DJ (1999) An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recogn* 32:71–86
- Friston KJ, Holmes A, Poline JB et al (1996) Detecting activations in PET and fMRI: levels of inference and power. *NeuroImage* 4(3):223–235
- Friston KJ, Holmes AP, Poline JB et al (1995) Analysis of fMRI time-series revisited. *NeuroImage* 2(1):45–53
- Friston KJ, Holmes AP, Worsley KJ et al (1995) Statistical parametric maps in functional imaging: a general linear approach. *Hum Brain Mapp* 2:189–210
- Li W, Kornak J, Harris T et al (2009) Identify fracture-critical regions inside the proximal femur using statistical parametric mapping. *Bone* 44(4):596–602
- Li W, Kornak J, Harris TB et al (2009) Bone fracture risk estimation based on image similarity. *Bone* 45(3):560–567
- Nichols TE, Holmes AP (2002) Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp* 15(1):1–25
- Chung S, Pelletier D, Sdika M et al (2008) Whole brain voxel-wise analysis of single-subject serial DTI by permutation testing. *NeuroImage* 39(4):1693–1705
- Hayasaka S, Nichols TE (2003) Validating cluster size inference: random field and permutation methods. *NeuroImage* 20(4):2343–2356
- Heller R, Stanley D, Yekutieli D et al (2006) Cluster-based analysis of fMRI data. *NeuroImage* 33(2):599–608
- Genovese CR, Lazar NA, Nichols T (2002) Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage* 15(4):870–878
- Thirion B, Pinel P, Meriaux S et al (2007) Analysis of a large fMRI cohort: statistical and methodological issues for group analyses. *Neuroimage* 35(1):105–120
- Smith SM, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44(1):83–98
- Lang T, LeBlanc A, Evans H et al (2004) Cortical and trabecular bone mineral loss from the spine and hip in long-duration spaceflight. *J Bone Miner Res* 19(6):1006–1012
- Lang TF, Leblanc AD, Evans HJ et al (2006) Adaptation of the proximal femur to skeletal reloading after long-duration spaceflight. *J Bone Miner Res* 21(8):1224–1230
- Li W, Sode M, Saeed I et al (2006) Automated registration of hip and spine for longitudinal QCT studies: integration with 3D densitometric and structural analysis. *Bone* 38(2):273–279
- Li W, Kezele I, Collins DL et al (2007) Voxel-based modeling and quantification of the proximal femur using inter-subject registration of quantitative CT images. *Bone* 41(5):888–895
- Holmes AP, Blair RC, Watson G et al (1996) Nonparametric analysis of statistic images from functional mapping experiments. *J Cereb Blood Flow Metab* 16(1):7–22
- Suckling J, Bullmore E (2004) Permutation tests for factorially designed neuroimaging experiments. *Hum Brain Mapp* 22(3):193–205
- Hastie T, Tibshirani R, Friedman JH (2001) *The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations*. Springer-Verlag Inc, Berlin, New York
- Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. *International joint conference on artificial intelligence (IJCAI)*, pp 1137–1143
- Breiman L, Friedman JH, Olshen RA et al (1984) *Classification and regression trees*. Wadsworth International, Belmont, Ca
- Therneau T, Atkinson E (1997) An introduction to recursive partitioning using the RPART routines, in technical report #61. Department of Health Sciences Research, Section of Biostatistics, Mayo Clinic, Rochester: Rochester, MN
- Venables WN, Ripley BD (1999) *Modern applied statistics with S-PLUS*, 3rd edn. Springer, New York
- Atkinson EJ, Therneau TM (2000) An introduction to recursive partitioning using the RPART routines, in technical report, s.o. Biostatistics, editor. Mayo Clinic, Rochester, MN
- Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* 98(9):5116–5121
- Nichols T, Hayasaka S (2003) Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat Methods Med Res* 12(5):419–446