Voxel-Based Morphometry

F Kurth and E Luders, UCLA School of Medicine, Los Angeles, CA, USA
C Gaser, Jena University Hospital, Jena, Germany

Introduction

The human brain is in a state of constant change and adaptation. This may be driven either by normal developmental or aging processes or by the effects of learning, training, and new occurrences in daily life. In addition to these aforementioned changes, more systematic influences such as gender, disease, and genes affect the brain's structure. Using magnetic resonance imaging, brain changes and differences can be measured noninvasively and *in vivo*, making them particularly interesting for both basic research and clinical research. The easiest way to assess brain changes (or group differences) is to measure whole-brain volume. However, assessing the volume of the entire brain is rather unspecific. So-called region-of-interest (ROI) analyses are more sensitive to local changes than are whole-brain assessments but are also subject to several limitations. For example, if one specific region is measured, other brain structures are ignored, and possible effects remain undetected elsewhere in the brain. Moreover, ROIs are usually created based on individual protocols and depend on rater-specific 'judgment calls,' thus requiring a clearly definable and unambiguous structure. For large parts of the brain, however, it may be difficult to precisely define (or identify) unambiguous boundaries. Finally, if an ROI is only partially different, this will lower the sensitivity to detect any effects in this region. This is where voxel-based morphometry (VBM) comes into play, as VBM allows for the examination of brain changes and/or group differences across the entire brain with a high regional specificity (i.e., voxel by voxel), without requiring the a priori definition of particular ROIs (Ashburner & Friston, 2000, 2001, 2007).

VBM: An Overview

VBM is an objective approach that enables a voxel-wise estimation of the local amount of a specific tissue. Most commonly, VBM is directed at examining gray matter but it can also be used to examine white matter. In the latter case, however, the sensitivity is limited, for white matter areas are characterized by large homogeneous regions with only subtle changes in intensity. The concept of VBM comprises three basic preprocessing steps: (1) tissue classification, (2) spatial normalization, and (3) spatial smoothing, which are followed by the actual statistical analysis. That is, if we know exactly what tissue can be found at a specific voxel, we can quantify and analyze it. This can be achieved by tissue classification. Furthermore, if we know that a specific voxel is at exactly the same anatomical location across all subjects (e.g., at the tip of the Sylvian fissure), we can compare voxel values across subjects. This is achieved by spatial normalization. Each brain, however, is unique; sulcal or gyral patterns, for example, vary greatly across subjects (some sulci are even missing in some brains). Thus, the success of spatial normalization is limited and depends on the accuracy of the applied registration method. In addition, parametric tests assume a Gaussian distribution of the residuals, which is not necessarily true for normalized tissue segments. Fortunately, these limitations can be addressed by applying a Gaussian blurring to the normalized tissue segment. This is achieved by convolving with a Gaussian function, which is commonly referred to as spatial smoothing. The smoothed normalized tissue segments are then entered into a statistical model to map changes within brains over time and/or differences between brains. The subsequent sections will further discuss these steps in detail; an overview of the basic workflow is illustrated in Figure 1.

Tissue Classification

Tissue classification is based on intensity values and basically serves to segment the brain into gray matter, white matter, and cerebrospinal fluid after removing any nonbrain parts (Ashburner & Friston, 1997, 2005; Rajapakse, Giedd, & Rapoport, 1997). However, intensity values in structural brain scans are not exclusively attributable to different tissue types, as an intensity-based tissue classification would assume. Rather, inhomogeneities of the magnetic field will lead to inhomogeneities in image intensity as well. This effect is even more pronounced with high-field scanners, since it is more difficult to keep the magnetic field homogeneous for higher field strengths. As shown in Figure 1 (T1-weighted image), the intensity inhomogeneity looks like a field of smoothly varying brightness, which results in different intensities for the same tissue at different locations. Thus, image intensity inhomogeneities need to be corrected before applying the actual tissue classification. This correction process is usually referred to as bias correction. The bias-corrected T1-weighted image can then be classified into any set of tissue types (usually three different tissue types for the brain plus one or more background types).

As shown in Figure 2 (left panel), the distributions of intensities for each tissue class overlap, even after a bias correction is applied. One reason for this overlap is that at a common voxel size of $1 \times 1 \times 1$ mm^3, any given voxel can contain more than one tissue. This is generally the case at the border between the brain parenchyma and cerebrospinal fluid, at boundaries between gray matter and white matter, and in structures where white matter fibers cross the gray matter. Thus, even in a bias field-corrected image, signal intensities for different tissues will vary and result in a considerable overlap and so-called partial volumes. Partial volumes can be modeled explicitly in order to more accurately classify the tissues and calculate local volumes (Tohka, Zijdenbos, &

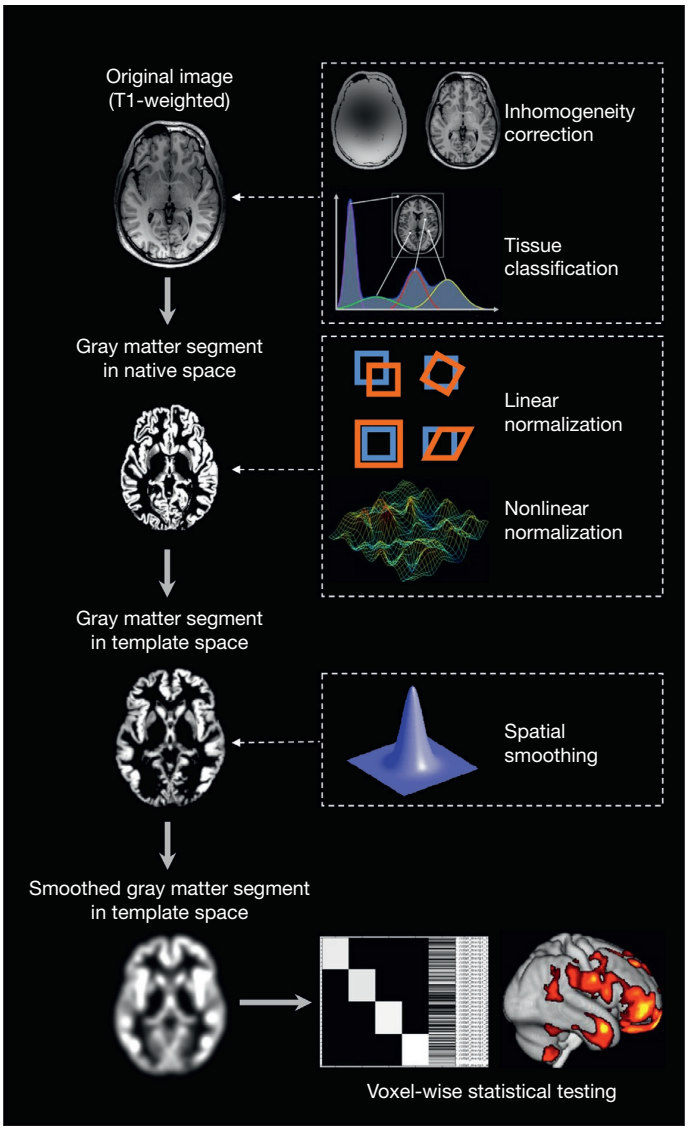

Figure 1 Workflow of a voxel-based morphometry (VBM) analysis. The analysis is based on high-resolution structural brain images. First, the T1-weighted images are corrected for inhomogeneities and classified into different tissue types, such as gray matter, white matter, and cerebrospinal fluid. The gray matter segment (i.e., the tissue of interest) is then spatially normalized to match a common template. Subsequently, the normalized gray matter segment is smoothed with an isotropic Gaussian kernel. Finally, the smoothed normalized gray matter segments are entered into a statistical model to conduct voxel-wise statistical tests and map significant effects.

Evans, 2004). To guide tissue classification, additional tissue probability maps can be used to apply prior knowledge of where in the brain different tissues can be expected (Ashburner & Friston, 2005). This means that for each tissue, a map of how probable it is to be represented by a certain voxel in the image is used to drive and restrict the tissue classification algorithm. While this may be valuable as long as the tissue probability maps match the subject's tissue distribution, it can lead to misclassifications in all populations that deviate from these maps (e.g., child data) (Wilke, Holland, Altaye, & Gaser, 2008). **Figure 2** (right panel) depicts the results of tissue classification. Since an algorithm that accounts for partial volumes was used, the given segments encode a local volume estimate of tissue content for every voxel.

Spatial Normalization

In addition to tissue classification, the individual brains – or the native gray matter segments (**Figure 3(a)**) – must be spatially normalized in order to ensure a voxel-wise comparability. Spatial normalization can be divided in linear and

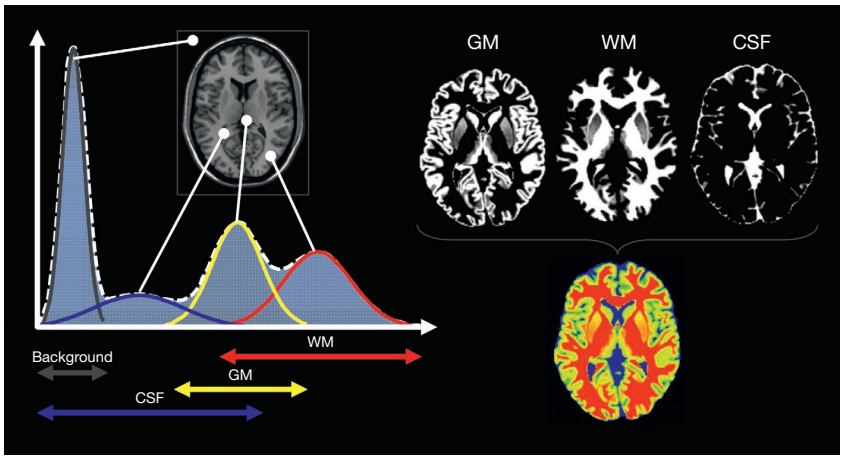

Figure 2 Tissue classification. Left panel: Whole-brain images can be segmented into background and different tissue classes, such as gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) based on their intensity. Note that these tissue-specific intensity distributions overlap, which can be due to partial volume effects. Right panel: The GM, WM, and CSF segments (top) were obtained using a partial volume estimation, which allows for more than one tissue per voxel. The partial volume estimation label (bottom) depicts the voxel values as transitions between tissue contents. GM is shown in yellow, WM in red, and CSF in blue. Voxels containing both GM and WM are shown in varying shades of orange, depending on the mixture of both tissues at this location. Voxels containing both GM and CSF are shown in varying shades of green, depending on the mixture of both tissues at this location. As both voxel size and tissue content per voxel are known, proper estimations of local tissue volumes can be made.

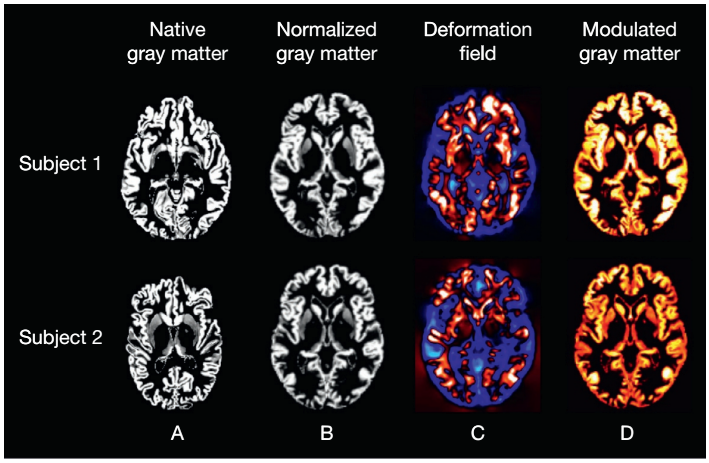

Figure 3 Spatial normalization. For visualization, two very different examples are depicted. Subject 1 is a 23-year-old male, while subject 2 is a 64-year-old female. (a) While local gray matter volumes can be measured in native space in both brains, a voxel-wise comparison is not easily possible. (b) After spatial normalization, both brains have the same size, shape, and overall pattern of major sulci and gyri. The local amount of gray matter can be directly compared in voxel-wise statistical tests. (c) The Jacobian determinants derived from the deformation fields that were applied for spatial normalization indicate different patterns of volume change for both subjects. The deformation forces needed to transform each subject's brain image to the template and highlight regions that were expanded (blue/cyan) or compressed (red/yellow) to match the respective areas in the template. Analyzing these deformation fields or the Jacobian determinants constitutes what is known as tensor-based or deformation-based morphometry. (d) Multiplying these deformation fields (or more precisely, the Jacobian determinants) with the original normalized gray matter segments corrects for the volume changes that occurred during the spatial normalization and is known as modulation. Voxel-wise statistical testing applied to these segments will analyze the local gray matter volume as estimated in native space. Note that although both brains are very similar, the second subject's smaller and probably slightly atrophic brain shows less local volume (evident as darker shades of orange).

nonlinear components. Linear normalization alters every part of the image in exactly the same way and includes translation, rotation, scaling, and shearing for each dimension (Ashburner & Friston, 1997). Translation and rotation (each in the *x*-axis, *y*-axis, and *z*-axis, yielding a total of six parameters) change the position in space but do not alter shape or size of the brain. This six-parameter transformation (also known as rigid body transformation) is frequently used to realign images of the same brain to each other and can be used, for example, to detect changes over time in the same subject. The addition of

scaling and shearing (each again in the *x*-axis, *y*-axis, and *z*-axis, yielding a total of 12 parameters) will alter the size and global shape of the brain. A 12-parameter transformation (also known as affine transformation) is frequently used to register brains to a template space.

While linear transformations can correct for interindividual differences in brain size, they cannot model local differences in size and shape as the same transformation is applied to every voxel. In contrast, nonlinear transformations allow the application of different changes in position, size, and shape locally and thus correct for interindividual differences on a local scale (Ashburner, 2007; Ashburner & Friston, 1999, 2005). Still, a perfect match between any two brains is very unlikely because brains are highly individual in their local anatomy (e.g., some sulci and gyri cannot be found in all brains). Nevertheless, in spite of minor remaining interindividual differences within the normalized gray matter segments (**Figure 3(b)**), modern normalization techniques result in brains with a reasonable local comparability (Ashburner, 2007).

All spatial transformations result in a deformation field (**Figure 3(c)**) that describes how local structures were adjusted to match two brains to each other (i.e., indicating if a part of the brain had to be enlarged or compressed). The exact voxel-wise volume changes can be easily derived from these deformation fields as Jacobian determinants. Analyzing these Jacobian determinants or the deformation fields themselves constitutes what is known as tensor-based morphometry or deformation-based morphometry. The Jacobian determinant may also be used to correct resulting gray matter segments for volume changes that occurred due to the spatial normalization. More specifically, suppose a structure of the brain with a certain amount of gray matter becomes bigger during normalization. Consequently, this structure will seem to have larger local gray matter values than are truly present. If the difference between true gray matter and apparent gray matter can be quantified – which is exactly what the Jacobian determinants do – the measured gray matter can simply be corrected (i.e., basically 'undoing' the unwanted effects of the normalization). This way, the amount of original gray matter is preserved in the new space and reflected as so-called modulated gray matter (**Figure 3(d)**).

Spatial Smoothing

The reason to smooth the images before statistical analysis is threefold: First of all, parametric tests assume that the residuals follow a Gaussian distribution. Simple smoothing of the images satisfies this assumption by the central limit theorem (after smoothing, the data are more normally distributed) and thus makes a parametric test a valid choice (Ashburner & Friston, 2000; Nichols & Hayasaka, 2003). Second, as outlined earlier, the spatial normalization is not perfect and small interindividual differences remain. Smoothing accounts for these residual small interindividual differences in local anatomy (Ashburner & Friston, 2000). Finally, according to the matched filter theorem, smoothing renders the analysis sensitive to effects that approximately match the size of the smoothing kernel (Ashburner & Friston, 2000). As smoothing kernels usually have a full width at half maximum of 4–16 mm, this means that very small differences, which are possibly due to noise, are not picked up by the analysis. Consequently, after smoothing, each voxel represents a sphere similar to the smoothing kernel or, in other words, a weighted mean of its own and its neighbors' values.

Statistical Analysis

The smoothed normalized tissue segments can be analyzed in statistical models using parametric tests, although non-parametric tests are also common. Usually, these tests will be applied in a mass-univariate approach, which means that the same test is applied for each voxel simultaneously. As in most other neuroimaging analyses, this entails a severe multiple comparison problem and an appropriate correction has to be applied. In neuroimaging, two major levels of correction are frequently used that are both based on Gaussian random field theory (Worsley et al., 1996): a correction on a voxel level and a correction on a cluster level (though a set-level correction is also possible) (Friston, Holmes, Poline, Price, & Frith, 1996). Assume the results are to be corrected controlling the family-wise error (FWE) at $p \leq 0.05$. At the voxel level, an FWE correction will assure that only in 1 out of 20 images a finding will have reached significance by chance. This is a perfectly legitimate way of correcting the results. To apply an FWE correction at cluster level, an arbitrary cluster-forming threshold must be applied, say at $p \leq 0.001$ uncorrected (Friston et al., 1996). Given the smoothness of the data, smaller clusters are likely to occur by chance thus constituting false positives. Larger clusters, however, are less likely to occur and cluster-forming thresholds will produce clusters that constitute real effects. Controlling the FWE at the cluster level therefore means that only in 1 out of 20 images a cluster of this extent will occur by chance. This correction will consequently result in a spatial extent threshold expressed as the minimum number of voxels comprising the significance cluster.

Unfortunately, statistical parametric maps from structural analyses vary considerably in local smoothness, meaning that the appropriate extent threshold varies locally as well. In other words, within the same image, there might be very smooth regions where large clusters may occur by chance and relatively rough regions where true effects may manifest as very small clusters. Applying one single extent threshold for the whole image is therefore inappropriate (Ashburner & Friston, 2000; Hayasaka, Phan, Liberzon, Worsley, & Nichols, 2004). A possible solution is to correct each voxel individually based on the local smoothness by rendering smoothness isotropic, which results in locally varying extent thresholds. Another possibility is to use a correction based on threshold-free cluster enhancement (TFCE) (Smith & Nichols, 2009). This method estimates a voxel value that represents the accumulative cluster-like local spatial support at a range of cluster-forming thresholds. TFCE has a variety of advantages that make it an elegant solution to correct for multiple comparisons in structural analyses. First of all, it does not need an arbitrary cluster-forming threshold, making it more objective. Second, it combines statistics based on the local significance as well as the spatial extent of this effect. However, because the distribution of the TFCE values is not known, permutation tests must be used to assess thresholds.

See also: **INTRODUCTION TO METHODS AND MODELING**: Diffeomorphic Image Registration; Nonlinear Registration Via Displacement Fields; Rigid-Body Registration; Tensor-Based Morphometry; Tissue Classification.

References

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *NeuroImage*, *38*, 95–113.

Ashburner, J., & Friston, K. (1997). Multimodal image coregistration and partitioning—A unified framework. *NeuroImage*, *6*, 209–217.

Ashburner, J., & Friston, K. J. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, *7*, 254–266.

Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry—The methods. *NeuroImage*, *11*, 805–821.

Ashburner, J., & Friston, K. J. (2001). Why voxel-based morphometry should be used. *NeuroImage*, *14*, 1238–1243.

Ashburner, J., & Friston, K. J. (2005). Unified segmentation. *NeuroImage*, *26*, 839–851.

Ashburner, J., & Friston, K. (2007). Voxel-based morphometry. In K. Friston, J. Ashburner, S. Kiebel, T. E. Nichols & W. D. Penny (Eds.), *Statistical parametric mapping: The analysis of functional brain images*. London: Elsevier.

Friston, K. J., Holmes, A., Poline, J. B., Price, C. J., & Frith, C. D. (1996). Detecting activations in PET and fMRI: Levels of inference and power. *NeuroImage*, *4*, 223–235.

Hayasaka, S., Phan, K. L., Liberzon, I., Worsley, K. J., & Nichols, T. E. (2004). Nonstationary cluster-size inference with random field and permutation methods. *NeuroImage*, *22*, 676–687.

Nichols, T., & Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, *12*, 419–446.

Rajapakse, J. C., Giedd, J. N., & Rapoport, J. L. (1997). Statistical approach to segmentation of single-channel cerebral MR images. *IEEE Transactions on Medical Imaging*, *16*, 176–186.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*, 83–98.

Tohka, J., Zijdenbos, A., & Evans, A. (2004). Fast and robust parameter estimation for statistical partial volume models in brain MRI. *NeuroImage*, *23*, 84–97.

Wilke, M., Holland, S. K., Altaye, M., & Gaser, C. (2008). Template-O-Matic: A toolbox for creating customized pediatric templates. *NeuroImage*, *41*, 903–913.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., & Evans, A. C. (1996). A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, *4*, 58–73.