Published in final edited form as:

Neuroimage. 2011 July 15; 57(2): 423–430. doi:10.1016/j.neuroimage.2011.04.046.

Biological Parametric Mapping WITH Robust AND Non-Parametric Statistics

Xue Yang^a, Lori Beason-Held^b, Susan M. Resnick^b, and Bennett A. Landman^{a,c,d,*}

^a Electrical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^b National Institute on Aging, National Institutes of Health, Baltimore, MD, 21224

^c Biomedical Engineering, Vanderbilt University, Nashville, TN, USA 37235

^d Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA 21218

Abstract

Mapping the quantitative relationship between structure and function in the human brain is an important and challenging problem. Numerous volumetric, surface, regions of interest and voxelwise image processing techniques have been developed to statistically assess potential correlations between imaging and non-imaging metrics. Recently, biological parametric mapping has extended the widely popular statistical parametric mapping approach to enable application of the general linear model to multiple image modalities (both for regressors and regressands) along with scalar valued observations. This approach offers great promise for direct, voxelwise assessment of structural and functional relationships with multiple imaging modalities. However, as presented, the biological parametric mapping approach is not robust to outliers and may lead to invalid inferences (e.g., artifactual low p-values) due to slight mis-registration or variation in anatomy between subjects. To enable widespread application of this approach, we introduce robust regression and non-parametric regression in the neuroimaging context of application of the general linear model. Through simulation and empirical studies, we demonstrate that our robust approach reduces sensitivity to outliers without substantial degradation in power. The robust approach and associated software package provide a reliable way to quantitatively assess voxelwise correlations between structural and functional neuroimaging modalities.

Keywords

Structure-Function Relationships; Statistical Parametric Mapping; Biological Parametric Mapping; Robust Regression; Non-Parametric Regression

INTRODUCTION

The assessment of structure-function relationships plays an important role in developing understanding of the human brain. These relationships appear to transcend scales, with connectivity dynamics appearing at both a micro- (intra-voxel) and macro- (the whole brain)

© 2011 Elsevier Inc. All rights reserved.

Corresponding author: Bennett A. Landman, PhD, Vanderbilt University EECS, 2301 Vanderbilt Pl., PO Box 351679 Station B, Nashville, TN 37235-1679, Work: 410-917-6166, bennett.landman@vanderbilt.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

scales (Honey, 2010). While there is firm evidence that structure shapes neural dynamics and function, the quantitative relationships and how such relationships might be altered in aging and disease remain an area of active research (Rubinov, 2009). Statistical parametric mapping and its implementation in SPM software has emerged as a powerful tool to simultaneously study changes within single voxels and spatial patterns of activity or tissue volume changes (Penny et al., 2006; Worsley et al., 2004). Recently, biological parametric mapping (BPM) (Casanova R., 2007) and incorporation of voxel-wise covariates (Oakes et al., 2007) has enabled multi-modal images to be analyzed in a manner consistent with traditional voxel-based morphometry. BPM enables regression of multi-modal images in the same manner that SPM enables regression of images on scalars. In BPM, both the outcome variables (regressands, “y”) and the explanatory variables (e.g., regressors, “x”) may be images, so the design matrices vary voxel-by-voxel. With the BPM approach, one can directly assess the relationships between images containing functional and structural information. Alternatively, one could assess the explanatory power of one set of imaging modalities for another.

Violations of the distributional assumptions (such as the presence of outliers) can influence the results of a statistical analysis considerably, especially when the sample size is small (Penny et al., 2006). In neuroimaging, outliers can appear because of imaging artifacts, data processing anomalies, and non-modeled anatomical differences between brains. In SPM, smoothing is used to compensate for “uninteresting” residual anatomical variability after registration. However, these anatomical differences are especially problematic in BPM because they introduce systematic and highly correlated variation in both regressors and regressands. Herein, we implement and evaluate robust and non-parametric statistical approaches for neuroimaging inference in the presence of outliers and distributional violations within the BPM regression model.

First, we consider the use of robust statistics for inference. These methods have been well-established in the statistical literature (Holland and Welsch, 1977; Huber, 1981) and are commonly available in statistical analysis packages (Chen, 2002). The general principle of robust statistics is to use an estimator that is consistent with a traditional, least-squares estimator, but is less sensitive to distributional assumptions – for example the median is a robust version of the mean. Herein, we extend the BPM approach to use robust inference with M-estimators (Huber, 1964) in place of ordinary least squares regression. The resulting statistical parametric maps are compatible with the SPM software and are amenable to standard multiple comparison corrections approaches including random field theory family-wise error and false discovery rate approaches (Penny et al., 2006).

Second, we investigate non-parametric regression in the context of BPM. Non-parametric mapping requires minimal assumptions for validity and has been particularly advocated for small-sample size analysis in the statistical non-parametric mapping (SnPM) implementation (Nichols and Holmes, 2002). We apply the SnPM non-parametric regression method within the BPM regression model to develop biological non-parametric mapping (BnPM). As with robust BPM, the uncorrected statistical maps are compatible with SPM software and are amenable to standard multiple comparison corrections approaches if appropriate field theory assumptions are valid. Additionally, the non-parametric approach offers the ability to perform non-parametric multiple comparison correction based on the maximal statistic.

This manuscript is organized as follows. First, we introduce the robust and non-parametric inference theory. Then we describe how these approaches can be used within a voxel-wise framework to develop robust BPM and BnPM. Simulation examples compare the results from original, robust BPM and BnPM. We close with a discussion of the advantages and limitations of each method and the potential opportunities for continued innovation.

Methods

BPM enables the use of image on image regression by solving the general linear model (GLM) with different design matrices for each voxel. This contrasts with conventional SPM (statistical parametric mapping) which can only use non-image regressors as it uses the same design matrix for a whole image. By choosing “regression” in the BPM interface, we can evaluate one set of images relative to another set of images. BPM has been shown to be a promising technique (Casanova R., 2007) for multi-parametric analysis.

To illustrate the potential short coming, consider two groups of data from one set of subjects where one subject is mis-registered relative to the remainder of the subjects; Figure 1 shows one set of positron emissions tomography (PET) images and one set of smoothed gray matter (GM) images. If we examine one voxel within the misregistration region, simultaneous BPM analysis of both datasets suffers from greatly increased sensitivity to the outlier data as compared to separate SPM of the individual datasets. In BPM, both regressors and regressands are images which are not fixed, while in SPM, one regresses scalar values on images, which ensures the regressors are fixed. Both robust regression and nonparametric regression are promising methods for addressing this outlier problem. To reduce the effects of outliers, we proposed replacing the ordinary least squares regression method by robust regression, and using the robust variance to perform robust T-tests and F-tests.

Theory: Robust Regression

M-estimators represent a broad class of techniques which can be considered as the generalizations of the maximum likelihood framework (Fox, 2002). Under the Gaussian distribution and independence assumptions, ordinary least squares regression is the maximum likelihood estimator. When there are outliers, the measurement errors are not normally distributed. Since parameters are fitted to minimize the squared error, a single outlier observation can have very large impact. In the design of an M-estimator, one assumes a form for the conditional probability of the data that allows for large deviations (Press W. H., 2007):

$$P(\text{data}|\text{model}) = \prod_{i=0}^{N-1} \{ \exp \left[-\rho \left(\frac{y_i - y(x_i|\beta)}{\sigma_i} \right) \right] \Delta y \} \quad (1)$$

where y_i is the measurement, $y(x_i|\beta)$ is the predicted value based on the model, σ_i is the weighting factor for observation, i , ρ is the influence function, and Δy is a small fixed factor for each data point.

The ρ function must be chosen appropriately to achieve the desired effect. If $\rho(r) = \frac{1}{2}r^2$, the M-estimator is equal to the least squares estimator. In this case, when an outlier occurs, $y(x_i|$

$\beta)$ is pulled towards the outlier to minimize $\sum_{i=0}^{N-1} \rho \left(\frac{y_i - y(x_i|\beta)}{\sigma_i} \right)$, which makes the least squares estimate method sensitive to outliers. In robust regression, $\rho(r)$ does not increase or increases slowly when r is greater than a predefined constant, thus assigning a small weight to outliers which avoids pulling $y(x_i|\beta)$ towards the outlier. Hence, one seeks a ρ that has lower values in the extreme than a squared term. There are many weight functions that have been provided and evaluated. The Bisquare weight function and Huber weight function are both widely used and continuous weight functions. The Huber weight function is monotone, but assigns a bigger weight to possible outliers. Based on preliminary evaluations of sensitivity to outliers (not shown), we choose the Bisquare weight function (Beaton, 1974):

$$\rho(r) = \begin{cases} \left(\frac{C^2}{2}\right) \left[1 - \left[1 - \left(\frac{r}{C}\right)^2\right]^3\right] & |r| \leq C \\ \frac{C^2}{2} & |r| > C \end{cases} \quad (2)$$

where C is the tuning constant, usually the value is chosen for 95% asymptotic efficiency when the distribution of observation error is Gaussian distribution.

Given the robust likelihood model in Eqs. (1) and (2), we can use the maximum likelihood algorithm to estimate β . To maximize $P(\text{data}|\text{model})$, we can equivalently minimize $-\log P(\text{data}|\text{model})$:

$$\sum_{i=0}^{N-1} \rho\left(\frac{y_i - y(x_i|\beta)}{\sigma_i}\right) \quad (3)$$

Using maximum likelihood method, β should satisfy

$$\sum_{i=0}^{N-1} \rho'\left(\frac{y_i - y(x_i|\beta)}{\sigma_i}\right) \left(-\frac{1}{\sigma_i}\right) \left(\frac{dy(x_i|\beta)}{d\beta}\right) = 0 \quad (4)$$

To solve this equation, we can use the iteratively reweighted least squares techniques (Holland and Welsch, 1977), which can be applied to the general linear model (GLM). A specific algorithm to compute the estimated solution of β is provided in (Street, 1988).

The least-squared formulas for statistical inference and confidence intervals are not meaningful when the parameters are computed using robust regression. Thus, a robust estimation of the covariance structure is required (Huber, 1981):

$$\text{cov}(\hat{\beta}) = K^2 \frac{\left[\frac{1}{n-m}\right] \sum \psi(r_i)^2}{\left[\left(\frac{1}{n}\right) \sum \psi'(r_i)\right]^2} (X^T X)^{-1} \quad K = 1 + \frac{m}{n} \frac{\text{var}(\psi')}{(E\psi')^2} \quad \psi(r) \stackrel{\text{def}}{=} \frac{d\rho(r)}{dr} \quad (5)$$

and a robust projection matrix can be used in place of $X^T X$:

$$\frac{\sum w_i x_{ij} x_{ik}}{\left(\frac{1}{n}\right) \sum w_i} \quad w_i \stackrel{\text{def}}{=} \frac{\psi(r_i)}{r_i} \quad (6)$$

where n is the number of observation y , m is the number of independent β parameters, w_i is the weight for data i calculated from the ρ function. The robust covariance formula is based on the assumption that $1 \ll m \ll n$. Details are discussed in (Huber, 1973). In most cases for normal errors, the criteria of $\frac{m}{n} < 1/4$ can ensure agreement of Monte Carlo experiments with theoretical models. With limited degrees of freedom, a more conservative approach which restricts the robust variance estimation to be at worst similar to the variance of the least squares estimation may be appropriate (Dumouchel, 1992). However, this can result in greatly reduced power in the presence of large outliers, and the assumption $m \ll n$ should be satisfied in the practical environment. The robust covariance structure and robust parameter estimates can be combined to perform robust T-tests and F-tests.

If there are no outliers, all distances are in the range of predefined threshold, the results from robust regression are the same as or close to least squares regression. When there are

outliers, robust regression reduces the effect of outliers so that these methods suffer from reduced power (e.g., ability to detect true extreme values) compared with least squares regression which considers all data points equally. The assumptions used in SPM to estimate the spatial smoothing kernel are not as problematic with the outliers as information is pooled across the whole image. However, exploration of robust smoothness estimation would be a fantastic area of future research.

Theory: Non-parametric Regression

The least-squares approach is based upon the parameterization of the error distribution as independent and Gaussian. These assumptions are violated by outliers. In contrast, non-parametric methods make minimal assumptions about the error probability distributions and may be applied in situations where such assumptions would be inappropriate. Permutation tests have become a widely popular nonparametric test in neuroimaging through the SnPM software (Nichols and Holmes, 2002). Permutation testing is relevant when data are associated with labels. The hypothesis is that if there is no effect in the data attributable to the labels, then the probability of the actual label order is equal to that of any other label permutation. Hence, the statistics associated with each of the possible labelings are also equally likely. The probability of an outcome as extreme or more than the one observed, the p-value, is the proportion of statistic values in the permutation distribution greater than or equal to that observed. The actual labeling used in the experiment is one of the possible labelings, so if the observed statistic is the largest of the permutation distribution, the p-value is $1/N$, where N is the number of possible labelings of the initial randomization scheme (including the observed one). Herein, we use the same observed statistic that has been implemented in SnPM5 (Wellcome Department of Cognitive Neurology).

The large number of possible label permutations can be problematic, due to the computations involved. In situations where it is not feasible to compute the statistic images for all the labelings, a random subsample of labelings can be used. The set of N possible labelings is reduced to a more manageable N' consisting of the true labeling and $N' - 1$ randomly chosen from the set of $N-1$ possible relabeling. Such a test is sometimes known as an approximate permutation test, since the permutation distribution is approximated by a subsample, leading to approximate p-values and critical thresholds. (Holmes, 1995)

Random field theory family wise error (FWE) (Worsley, 1996) and false discovery rate (FDR) (Benjamini Y., 1995) corrections for multiple comparisons can be inappropriate for non-parametric methods since there are limited distributional assumptions. Holmes et al. (Holmes, 1996) provided a single threshold test for permutation multiple comparisons.

Variance estimation for t-statistic can be a noisy process, especially with limited sample sizes. Since one would expect the noise field to be relatively smooth, one can use a locally pooled variance to improve accuracy. For parametric methods, use of smoothing would be problematic because it would require knowledge of the variance-covariance of the voxel level variance structure. Permutation testing does not suffer from this limitation.

Implementation

The BPM software package was extended to include support for robust regression with Bisquare M-estimators and non-parametric techniques. In this manner, we extend BPM to form robust BPM (rBPM) and Biological non-Parametric Mapping (BnPM). BnPM uses non-parametric regression for each voxel, based on the permutation tests theory. Non-parametric regression calculates p-value based on randomized permutations (considering the proportion of the numbers with extreme value in permutations) but not based parametric assumptions. Thus, the direct result of non-parametric regression is a p-value rather than a T

or F statistic. To report BnPM significance results within the SPM software, we invert the p-value to calculate a T map which can then be mapped back to a p-value in the visualization tools. The p-value is computed according to the sign of T test (i.e., proportion of the larger numbers for positive t-test and proportion of the smaller numbers for negative t-test), so the signs of T scores are preserved.

It is important to realize that the computational burden of non-parametric methods is very high. For a typical image size (e.g., $85 \times 128 \times 64$), a modern, single-threaded CPU may require more than thirty days to compute 10000 permutations. To address this situation, we integrated BnPM with a cluster processing environment. With BnPM, a regression task (e.g., “job”) is divided into many smaller sub-jobs, each with a specific set of permutation orders for the whole job. Every sub job is considered a new job submitted to the cluster. The number of sub jobs and the execution wall time (i.e., maximum time allowed for each sub-job completion) is defined by the user and depends on the image size and the performance of the cluster. The description of each job is written in a format that provides compatibility to diverse cluster environments and scheduling priorities.

Experiments

First, to demonstrate BPM, rBPM and BnPM in a framework containing outliers, we created 20 simulated images as imaging regressors and another 20 images as the main modality (regressands). The imaging regressor images were created from one smoothed gray matter image by adding different noise fields which are generated from Gaussian distribution with zero mean. We define SNR as the mean signal divided by the standard deviation of noise. We define CNR as the magnitude of the range of voxel intensities divided by the standard deviation of noise. Figure 2 illustrates representative simulated datasets. The main modality images are created from the regressor images. Inside the head of the caudate region, the simulated true-intensity of the main modality images is “ $1.5 \times \text{regressor images} - \text{constant}$ ”. Inside the putamen region, the simulated true-intensity of main modality images is “ $(-1) \times \text{regressor images} + \text{constant}$.” Everywhere else the simulated true-intensity equals a constant. All constants are chosen to make $\text{SNR} = 10$ and $\text{CNR} = 4$. To simulate an artifact by mis-registration, one pair of images (e.g., representing one simulated subject) are rotated around the x axis (i.e., right-left) by 15 degrees. Then, a linear model ($y = \beta x + \mu$) was fit to evaluate correlations between the pairs of simulated images, and the null hypothesis of $\beta = 0$ was tested.

Second, to evaluate the relative performance of each method on empirical data, we tested the non-robust BPM, robust BPM and BnPM (with 10000 permutations) using a sub-set of the normal aging study of the BLSA neuroimaging project (Resnick et al., 2000). Herein, we used a sub-cohort of 23 healthy participants (12 M/8 F, 60–85 y/o at baseline). Each subject was studied annually for eight years with T1-weighted MRI sequence (1.5T, GE Medical Systems, SPGR, $0.9375 \times 0.9375 \times 1.5$ mm, $256 \times 256 \times 124$ mm field of view) and PET data (^{15}O , GE 4096+ scanner, 15 slices of 6.5mm thickness, 60s). The baseline scan was denoted as year 1 and the last scan was denoted as year 9. The data were preprocessed with SPM 5 (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5>) following steps described in (Beason-Held et al., 2008). The structural scans were normalized to MNI-space, segmented and smoothed (12 mm Gaussian kernel) to obtain smooth gray matter (GM) density images. PET images were normalized to SPM’s spatial normalization, smoothed (12 mm Gaussian kernel) to MNI-space, and calibrated for global blood flow measurements to form quantitative cerebral blood flow measurements. The smoothed gray matter structure images were regressed on the PET data using non-robust BPM, robust BPM and BnPM to explore the potential for gray matter changes to explain changes in PET activation levels. Differences between estimated model fits and significance levels of the covariate were plotted.

Results

As Figure 3 illustrates, rBPM and BnPM were incorporated into the BPM software. With the new user interface, the user has a choice about robust or non-robust regression. If the user chooses robust regression, we use the iteratively reweighted least squares method to perform M-estimation and compute robust variances for the GLM model. Otherwise the BPM analysis method remains the same as the original BPM package. If the user selects BnPM analysis, sub-jobs are created using non-parametric regression method and submitted to the supercomputing cluster. After all sub-jobs are finished, users can call BnPM results to calculate either uncorrected p-value or corrected p-value.

Table 1 summarizes the regression methods studied herein and presents a relative comparison of analysis time for the empirical example. All toolboxes that have been implemented are listed in the table with running time compared to the SPM running time. Note that non-parametric methods were set to use 10000 Monte Carlo permutations rather than exhaustive search. For multi-modality regression, BPM is the shortest while the BnPM running is the longest. Computation times are listed in terms of a single threaded CPU. For example, with a grid of 33 computers, the running time for BnPM would be equivalent to that of rBPM. The robust BPM running time is about 20 minutes for 20 subjects with 79*95*69 size images. The RobustWLS toolbox for SPM performs robust estimation to compensate artifacts in fMRI time series data using weighted least squares estimates; this method was not evaluated as this manuscript focuses on data with a single time point (Diedrichsen and Shadmehr, 2005).

Table 2 shows the sensitivity and specificity of BPM, rBPM and BnPM on simulated dataset based on uncorrected p-value ($p < 0.001$). The type I error is under control when there are no outliers (specificity $\approx 99.9\%$). When there are outliers, the specificity of BPM is excessively low (lower than 70%) and the sensitivity is also lower. The specificity of rBPM is slightly lower than the theoretical value. Note that the robust statistics are asymptotically valid, so ability to precisely control type I error improves with sample size. If rBPM is performed over larger sample size or higher tolerated type I error, the specificity can be better controlled. BnPM also results slightly decreased specificity, but has lower sensitivity than rBPM.

After correcting for multiple comparisons, each of the three regression approaches would lead to a different conclusion based on inter-modality statistical parametric mapping as illustrated in Figure 4. With traditional BPM, one would see trends towards significant differences in both the region with true difference (i.e., caudate and putamen) and in the region where outliers were present after multiple comparisons. The rBPM approach would observe only the true difference. Finally, the BnPM found only a small significant region inside the true difference region after correction for multiple comparisons with the maximum statistic.

Both rBPM and BnPM control type I error in the presence of artifacts as seen in Figure 4; however, BnPM loses sensitivity and requires enormous computation relative to rBPM. For data without outliers, BnPM is more powerful (more significant voxels seen). In analysis with small sample sizes, the asymptotic Gaussian distribution assumptions may be inappropriate and BnPM should be applied (as opposed to rBPM or BPM) (Arndt et al., 1996). Yet, rBPM provides an efficient means to deal with outliers when the sample size is sufficient, and as such BnPM and rBPM can be understood as complementary approaches.

To a large extent, the results found by BPM, rBPM, and BnPM were qualitatively similar using the empirical dataset as seen in Figure 5. However, there were several substantial (and

significant) clusters where the methods reported dissimilar results, two of which are detailed in Figure 5.

Conclusions

The presence of outliers and distributional violations in neuroimaging is not a new phenomenon. Assessing the validity of statistical parametric models in a tractable manner has been paramount in the design of analysis software (Luo and Nichols, 2003). Concerns over the validity of parametric models have motivated the development of alternative non-parametric approaches (Nichols and Holmes, 2002). These approaches have become widely used for analysis of experiments with small sample sizes. While manual data inspection and outlier exclusion are common place, systematic incorporation of these criteria in the estimation process (e.g., through M-estimators) has not been as thoroughly pursued in neuroimaging.

In this study, we incorporated robust regression (M-estimation) into BPM to replace ordinary least-squares regression and implemented BnPM by applying non-parametric permutation tests and using cluster computation. BPM, robust BPM and BnPM are all integrated in one SPM toolbox so that users can choose an appropriate regression method for the task at hand. Both simulation and empirical data suggest increased resilience to outliers in accordance to the theoretical motivation for these methods.

In the simulation and empirical demonstrations, we demonstrate that these analysis approaches may lead to differing results. rBPM explicitly tolerates outliers by using robust regression, which makes the tool reliable against large outliers. However, rBPM still relies on Gaussian assumptions once outliers are excluded, so the results may not be appropriate for simulations with tenuous distributional assumptions. Meanwhile BnPM is intrinsically robust to distributional assumptions, but does not explicitly handle outliers, so it may suffer for lost power in these cases. Note that it would be straightforward to combine non-parametric and robust regression, i.e., perform permutation testing based on robust statistics, but this would dramatically increase the computation time.

In practical use, the assessment of structure-function relationships will likely involve the presence of outliers. As significant findings or surprising lack of findings are encountered, it is important to probe these datasets with robust tools to reveal the potential influence of outliers or distributional violations. For example, one would want to know if the outliers seen on Figure 5 (near 0.8, 20 at left and 0.2, 38 on right) are noise, systematic mis-registration, or seemingly valid data points. If these appear valid, is an important factor missing from the model design? As John Tukey, one of the founders of modern statistics commented, “that just which robust/resistant methods you use is not important-what is important is that you use some. It is perfectly proper to use both classical and robust/resistant methods routinely, and only worry when they differ enough to matter. But when they differ, you should think hard” (Huber, 2002). Here we have provided two very different robust approaches for use in the neuroimaging community. The described software is available in open source under the Lesser GNU Public License via the NITRC project “rbpm” (URL: <http://www.nitrc.org/projects/rbpm>).

Acknowledgments

This project was supported in part by grants NIH/NINDS 1R01NS056307 and NIH/NINDS 1R21NS064534, the Intramural Research Program of the NIH, National Institute on Aging and by Research and Development Contract N01-AG-3-2124.

References

- Arndt S, Cizadlo T, Andreasen NC, Heckel D, Gold S, O'Leary DS. Tests for comparing images based on randomization and permutation methods. *Journal of Cerebral Blood Flow & Metabolism*. 1996; 16:1271–1279. [PubMed: 8898701]
- Beason-Held L, Kraut M, Resnick S. II. Temporal patterns of longitudinal change in aging brain function. *Neurobiology of aging*. 2008; 29:497–513. [PubMed: 17178430]
- Beaton AE, Tukey JW. The fitting of power series, meaning polynomials, illustrated on band-spectroscopics data. *Technometrics*. 1974; 16:147–185.
- Benjamini YHY. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc Series B, Methodological*. 1995; 57:289–300.
- Casanova RSR, Baer A, Laurienti PJ, Burdette JH, Hayasaka S, Flowers L, Wood F, Maldjian JA. Biological parametric mapping: a statistical toolbox for multimodality brain image analysis. *NeuroImage*. 2007; 34:137–143. [PubMed: 17070709]
- Chen, C. Robust Regression and Outlier Detection with the ROBUSTREG Procedure. the Twenty-seventh Annual SAS Users Group International Conference; Cary, NC: SAS Institute Inc; 2002.
- Diedrichsen J, Shadmehr R. Detecting and adjusting for artifacts in fMRI time series data. *NeuroImage*. 2005; 27:624–634. [PubMed: 15975828]
- Dumouchel, W.; O'Brien, F. Computing and graphics in statistics. Springer-Verlag New York, Inc; New York, NY, USA: 1992. Integrating a robust option into a multiple regression computing environment; p. 41-48.
- Fox, J. An R and S-PLUS Companion to Applied Regression. 2002. Robust Regression.
- Holland PW, Welsch RE. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*. 1977; A6:813–827.
- Holmes, AP. Statistical Issues in functional Brain Mapping. 1995.
- Holmes AP, Blair RC, Watson JDG, Ford I. Nonparametric Analysis of Statistic Images from Functional Mapping Experiments. *Journal of Cerebral Blood Flow and Metabolism*. 1996; 16:7–22. [PubMed: 8530558]
- Honey CJ, et al. Can structure predict function in the human brain? *NeuroImage*. 2010; 52:766–776.
- Huber PJ. Robust Estimation of a Location Parameter. *Ann Math Statist*. 1964; 35:73–101.
- Huber PJ. Robust Regression: Asymptotics, Conjectures and Monte Carlo. *The Annals of Statistics*. 1973; 1:799–821.
- Huber, PJ. Robust Statistics. Hoboken, NJ: John Wiley & Sons, Inc; 1981.
- Huber PJ. John W. Tukey's Contributions to Robust Statistics. *The Annals of Statistics*. 2002; 30:1640–1648.
- Luo WL, Nichols TE. Diagnosis and exploration of massively univariate neuroimaging models. *Neuroimage*. 2003; 19:1014–1032. [PubMed: 12880829]
- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum Brain Mapp*. 2002; 15:1–25. [PubMed: 11747097]
- Oakes TR, Fox AS, Johnstone T, Chung MK, Kalin N, Davidson RJ. Integrating VBM into the general linear model with voxelwise anatomical covariates. *NeuroImage*. 2007; 34:500–508. [PubMed: 17113790]
- Penny, WD.; Friston, KJ.; Ashburner, JT.; Kiebel, SJ.; Nichols, TE., editors. Statistical Parametric Mapping: The Analysis of Functional Brain Images. Academic Press; New York, NY: 2006.
- Press, WHTSA.; Vetterling, WT.; Flannery, BP. Numerical recipes: the art of scientific computing. 3. 2007.
- Resnick SM, Goldszal AF, Davatzikos C, Golski S, Kraut MA, Metter EJ, Bryan RN, Zonderman AB. One-year age changes in MRI brain volumes in older adults. *Cereb Cortex*. 2000; 10:464–472. [PubMed: 10847596]
- Rubinov M, Sporns O, van Leeuwen C, Breakspear M. Symbiotic relationship between brain structure and dynamics. *BMC Neuroscience*. 2009; 10
- Street JO, Carroll RJ, Ruppert D. A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares. *The American Statistician*. 1988; 42:152–154.

- Worsley KJ, Marrett S, Neelin P, et al. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*. 1996; 4:58–73. [PubMed: 20408186]
- Worsley KJ, Taylor JE, Tomaiuolo F, Lerch J. Unified univariate and multivariate random field theory. *Neuroimage*. 2004; 23(Suppl 1):S189–195. [PubMed: 15501088]

Figure 1.

Increased sensitivity to outliers with BPM. Considered a simple one regressor situation, model $y = \beta x + \mu$ is used to fit data points in voxel wise analysis. In SPM, the imaging data (left column) lies along the y axis. (i.e., imaging data is treated as y in the fitted model.) The left column shows two sets of images data, PET blood flow maps (top row) and gray matter (GM) tissue density maps (lower row). The middle column shows the design matrix for each regression. Considered one voxel within mis-registration region, in separated analyses of these datasets with SPM (top and lower rows), a mis-registration induced outlier impacts only on data point in the y-direction. In a joint analysis with BPM, the outlier impacts both the x-axis and y-axis, and can result in dramatically different model fits. In the simulation the true β is a small negative value, but the outliers increase the estimated value of β to make it approximately zero or even a large positive number.

Figure 2. Simulation design. The normal image (A) shows the regressor images that are created from one image. The normal image (B) is one of the regressand images. The outlier images (C,D) are rotated around x (left-right axis) by 15 degrees.

Figure 3.

Software graphical user interface (GUI) and flowchart. With the new software, the user has a choice between robust regression, non-parametric regression, and traditional least squares regression (shown at left). Both options make slight modifications to the analysis workflow (right) with the end results of all methods prepared for visualization within the SPM software package.

Figure 4.

Simulation results. Results are displayed in SPM in glass brains. The statistical significance is calculated based on a selected p-value and then projected to three dimensions. Here we display one dimension of glass brains. The results are shown according to corrected p-value based on random field theory with $p < 0.05$ and 5 voxels extent threshold to exclude noise. The red region is where there is a positive relationship and the blue region is where there is a negative relationship. Some points seem to overlap because the glass brain map collapses the out-of-plane dimension.

Figure 5.

Empirical Results. Gray matter structural data of 23 participants from year 1 and year 9 (46 images) were regressed on the PET data of the same subjects. The statistical significance maps are under uncorrected p-value ($p < 0.001$) and 5 voxels extent threshold to exclude noise. The two plots are the data for one voxel. The green arrow shows one voxel in the region where non-robust regression indicates non-significance while robust regression indicates significance. The blue arrow shows one voxel in the region where non-robust regression indicates significance while robust regression indicates non-significance. The x axis shows the value of GM structure data and the y axis shows the value of PET data. The red line is the fit of the non robust regression while the green line is the fit of the robust regression. The maps in the lower right are the significance maps with family wise error correction of $p < 0.05$ and 5 voxels cluster threshold.

Table 1
Comparison of statistical regression methods and available software for neuroimaging

	<u>Regression Technique</u>		
	Ordinary Least-Squares	Non-parametric: Permutation Test	Robust: M-estimator
Single Modality Software Package	SPM (Penny et al., 2006)	SnPM (Nichols and Holmes, 2002)	rSPM (Diedrichsen and Shadmehr, 2005)
<i>Assumptions</i>	Modality specific analysis, data are normally distributed.	Modality specific analysis, labels have exchangeability	fMRI time series data analysis
<i>Relative computation time</i>	Time 1.0 [*]	Time 110	-
Inter- Modality Software Package	BPM (Casanova R., 2007)	BnPM [‡]	rBPM [‡]
<i>Assumptions</i>	Multi-modality analysis, data are normally distributed		
<i>Relative computation time</i>	Time 10		

^{*} The total computational time for performing each regression analysis with data corresponding to the experiment shown in Figure 4 (20 subjects, whole brain, 1 mm isotropic resolution, etc.) is reported relative to the computation time for the traditional SPM approach (approximately 15 seconds on our hardware).

[‡] The highlighted methods represent contributions of this work.

[‡] BnPM runs in approximately five hours “real time” on our cluster.

Table 2

Sensitivity and Specificity on simulated dataset

	Caudate		Putamen	
	Sensitivity	Specificity	Sensitivity	Specificity
BPM on data without outliers	99.95% $\pm$ 0.067%	99.89% $\pm$ 0.11%	97.21% $\pm$ 0.23%	99.87% $\pm$ 0.14%
BPM on data with outliers	66.38% $\pm$ 0.43%	67.41% $\pm$ 0.84%	81.53% $\pm$ 0.34%	58.03% $\pm$ 1.04%
rBPM on data with outliers	99.88% $\pm$ 0.12%	99.79% $\pm$ 0.12%	96.63% $\pm$ 0.41%	99.51% $\pm$ 0.20%
BnPM on data with outliers	72.08% $\pm$ 0.45%	98.77% $\pm$ 0.30%	89.50% $\pm$ 0.36%	98.54% $\pm$ 0.34%

The sensitivity is calculated in the voxels in the true mask and the specificity is calculated in the voxels outside the true mask but inside the rotated ("outlier") mask. They are computed based on uncorrected p-value ($p < 0.001$). Here we show the mean value and standard deviation of sensitivity and specificity from 10 simulations (with different Gaussian noise) for each method.