

# Region-Based Data-Driven Intensity Normalization for Group Comparison of Functional Brain Images

Zhiyong Xie, Aijun Zhu, Laigao Chen and Timothy McCarthy

**Abstract** Intensity normalization is widely used to remove the confounding effect of global change exhibited in PET or SPECT brain images such that the local activity can be detected. Improper estimate of global change may induce a biased normalization. To separate the global change from local measurements, an iterative method is proposed to identify reference regions that are not affected by the local activity. From more than one hundred predefined anatomical regions, the reference regions are selected based on their intensity similarity between two groups. Weighted least squares regression is used to compute linear intensity transformations to align intensities of corresponding reference regions across all subjects. Studies with simulated data demonstrated that the proposed method performed better in recovering real intensity change comparing with global mean normalization and with Andersson's data-driven method.

**Keywords** Brain imaging · Glucose metabolism · Intensity normalization · PET · SPECT

## 1 Introduction

Positron emission tomography (PET) and single photon emission computed tomography (SPECT) has been used to assess cerebral blood flow or metabolic activity at voxel level. With images acquired at different experimental conditions, statistical comparison can be performed between groups to discover their functional difference voxel-wisely or over a region of interest [1, 5, 9]. In a PET or SPECT image, the functional activity measured at each voxel is the confounding effect of a local activity and a global change. The global change is regional independent and always exhibits

---

Z. Xie (✉) · L. Chen · T. McCarthy  
Pfizer Inc., Cambridge, MA, USA  
e-mail: Zhiyong.Xie@pfizer.com

A. Zhu  
Massachusetts General Hospital, Cambridge, MA, USA

© Springer International Publishing Switzerland 2015  
F. Gao et al. (eds.), *Computational Methods for Molecular Imaging*,  
Lecture Notes in Computational Vision and Biomechanics 22,  
DOI 10.1007/978-3-319-18431-9\_12

large inter- and intra-subject variation. To localize and quantify the regional activity, the global change needs to be removed to increase the statistical power of the group comparison. The global change can be estimated based on a predefined reference region which is assumed to have no local activity [4, 13]. However, selecting an appropriate reference region is very challenging in some studies. Using different brain regions may lead to different or even conflicting results [10, 12].

Besides the reference-based estimation, the average of all intracerebral voxels may be used as a measurement of the global change. In the widely used global mean normalization (GMN), the global effect is eliminated by dividing the local measurements by the whole-brain average [8]. This method works well in studies where a few small regions are activated by a task or cognitive challenge. In these kinds of studies, the magnitude of the activation is moderate and the whole-brain average is not affected by the local change. In studies with a pharmacological challenge however, a relatively large region may be activated and the local change may alter the whole-brain average. In this case, the GMN proportionally reduces the magnitude of the true activation and induces a bias in the opposite direction, thus decreasing the sensitivity and specificity in detecting the real functional change.

To address this issue, Andersson proposed an iterative method to identify voxels that are not affected by local activity and normalize images with the average intensity of identified voxels [2]. The method uses the GMN as an initial step. Then a voxel-wise comparison is performed between groups and all voxels with  $p > 0.05$  are used to normalize the image. This procedure is repeated until there is no further change in the global estimate. The Andersson method reduced the biased global estimate comparing to the GMN. However, in a study with a large inter-subject variation, voxels with moderate activation may be included in the global estimate due to the initial step of GMN and the insufficient criterion of  $p > 0.05$ . If the region with moderate activity is large enough, it will change the global estimate thus leads to a biased normalization in the similar way as the GMN. Such issue had been observed in a comparative study performed by Borghammer et al. [3]. Yakushev et al. proposed a non-iterative two-step method where the global change was estimated by only including “hypermetabolic” voxels [3]. This method only works for studies where the intensities of all activated regions change in a same pre-known direction. Moreover, the result of this method is very sensitive to the selection of the threshold of “hypermetabolic” voxels. Global change can be easily over or under estimated if the threshold is inappropriate [3].

In this paper, an automated region-based method is proposed by improving upon several drawbacks of the Andersson method: first, the proposed method removes the initial step of the GMN. Second, anatomical regions instead of individual voxels are identified for intensity normalization. Third, both p-value and percentage difference are used in identifying the reference regions. Fourth, the proposed method employs linear model to align the intensities of corresponding regions rather than using ratio only. Finally, linear transformations are computed using weighted least squares regression where the contribution of each identified region is determined based on its size and the intensity similarity between groups. To compare the proposed method with the GMN and the Andersson method, FDG-PET images from

normal rats were divided into two groups. Artificial intensity change was added in one group of images while the other group was used as the control. All images were normalized using three methods to test their capabilities of recovering the real intensity change in statistical comparisons.

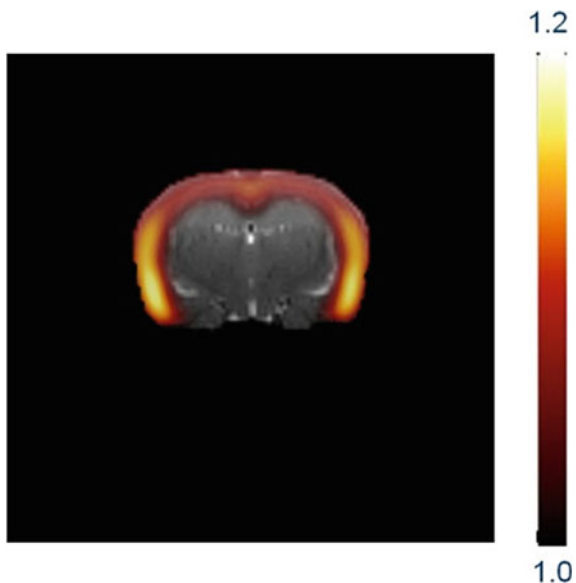
## 2 Materials and Methods

### 2.1 Image Acquisition and Data Simulation

Twenty two male Sprague-Dawley rats were scanned using a Focus-220 PET scanner (Siemens Medical Solutions). Images were reconstructed into a  $128 \times 128 \times 95$  volume with in-plane resolution of 0.6 mm and slice thickness of 0.8 mm with corrections for detector normalization, decay, attenuation, and scatter. The SUV (Standard Uptake Value) of FDG uptake was calculated and used for group comparison. All animal usage and experimental procedures were reviewed and approved by local Animal Usage Committee (IACUC).

All SUV images were evenly assigned into two groups so that the difference of their whole-brain average between groups was minimized (the difference was less than 1 % with p-value of 0.89 in a two tailed t-test). One group was used as control while images of another group were multiplied with a predefined scale image. In the first simulation, the SUV was increased by up to twenty percent in cortical regions. Figure 1 shows a representative slice of the scale image overlaid on a MRI template.

**Fig. 1** A representative slice of the predefined scale image. The *color bar* represents the scale of intensity increase. The scale image is served as the ground truth in the evaluation of different normalization methods



After the data manipulation, the whole-brain average of the manipulated group was 3.93% higher than the control group ( $p = 0.27$  in a two tailed t-test).

In the second simulation study, SUV in cerebellum was increased 7% in the manipulated group besides the SUV increase in cortical regions. This simulation was designed to assess the performance of different intensity normalization methods when there was a moderate signal change in a relatively large region. With the data manipulation, the difference of the whole-brain average between the two groups was 4.59% with p-value of 0.21 in a two tailed t-test.

## 2.2 Normalization Method

All images were spatially aligned to a template on which more than one hundred anatomical regions had been delineated. The mean image  $G$  was computed by averaging all aligned images. Denotes the mean intensity of  $i$ th region on the  $G$  as  $G_i$ . The proposed algorithm is summarized as follows:

1. At every voxel of the brain, compute the absolute percentage difference between the two groups. The median percentage difference (denotes as  $\Delta$ ) is used as the threshold of percentage difference in step 3b.
2. For each anatomical region  $i = 1:m$ ,
  - a. Compute the mean intensity  $F_{ij}$  for each image  $j = 1:n$ ,
  - b. Compare the intensity difference between the two groups.
  - c. Get p-value and percentage difference of intensity between groups.
3. The unaffected regions (denotes as  $\Omega$ ) are selected based on following criteria,
  - a. The p-value of the region is greater than 0.1.
  - b. The percentage difference of the regional intensity is less than the  $\Delta$  defined in step 1.
4. For  $j$ th image  $F_j$ ,
  - a. Compute  $\alpha_j$  and  $\beta_j$  by minimizing

$$\sum_{i \in \Omega} w_i (\alpha_j + \beta_j F_{ij} - G_i)^2$$

Where the  $w_i = s_i * p_i$  are weights defined by the size ( $s_i$ ) and p-value ( $p_i$ ) of the  $i$ th region.

- b. Update the intensity of image  $F_j$  by  $F_j = \alpha_j + \beta_j F_j$ .
5. Repeat step 1 to step 4 until  $|\beta_j - 1| < 0.01$  and  $|\alpha_j| < \theta_j$  for all images.  $\theta_j$  is defined as one percent of the whole-brain average. For simplicity, this method is referred as **ReDIN (Region-based Data-driven Intensity Normalization)** in the rest of the paper.

### 2.3 Image Analysis

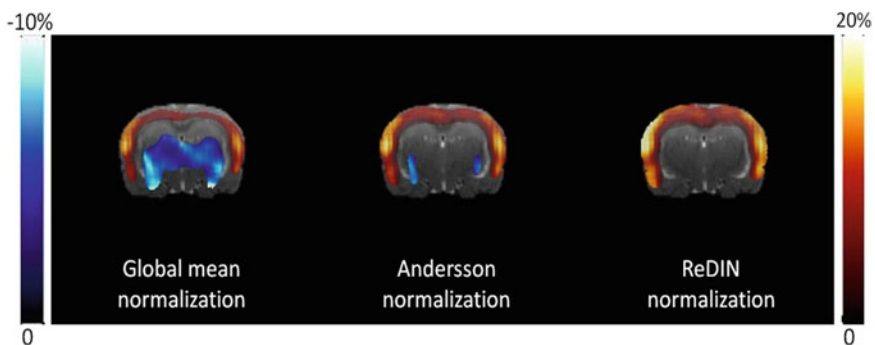
After being aligned to the template, the intensities of all SUV images were normalized using the GMN, the Andersson, and the ReDIN separately. Statistical comparisons were performed on normalized images to detect the difference between the two groups at voxel level or over the anatomical region of interest (ROI). The Voxel-wise comparison was carried out using the software package AFNI (Analysis of Functional NeuroImages, <http://afni.nimh.nih.gov>) where two-tailed t-test was performed after all images were smoothed with a Gaussian filter (FWHM = 1.5 mm). The family-wise error rate was controlled to 0.05 when detecting clusters with significant metabolic differences.

In ROI analysis, the mean SUV of each ROI was computed on normalized images and compared between groups. The percentage differences and p-value of corresponding ROI were reported. The percentage differences recovered from different normalization methods were compared with the real change of the manipulation.

## 3 Results

### 3.1 Simulation I Study

The results of voxel-wise analysis is displayed in Fig. 2 where the detected clusters are highlighted by color. The red color represents the SUV increase of the manipulated group compared to the control group and the blue color shows the SUV decrease. Color bars indicate the percentage difference. After the GMN, only a part of SUV increase can be detected and an artificial SUV decrease was found in a large region



**Fig. 2** The influence of different normalization methods on voxel-wise comparison in simulation I. The identified clusters were highlighted by colors that represented the detected percentage change of the manipulated group over the control group. *Red color* shows the regions with SUV increase and *blue color* indicates the SUV decrease

(left panel). The Andersson method improved the GMN method by recovering more regions with real SUV increase and inducing smaller region with the artificial SUV decrease (middle panel). ReDIN recovered most regions with SUV increase without inducing SUV decrease (right panel).

With ROI analysis, the influence of different normalization methods can be compared quantitatively. Table 1 shows the percentage difference and p-value between the two groups after normalizing data with different methods. The true change is the real percentage change applied to images of the manipulated group. With the GMN, the magnitude of SUV increase was reduced and around 5% artificial signal decrease was induced in hippocampus and thalamus. With Andersson and ReDIN methods, the results were close to the real change.

### 3.2 Simulation II Study

In the second simulation study, a moderate SUV increase was introduced in cerebellum. Because the whole-brain average of the manipulated group is 4.59% higher than the control group, the GMN reduced the intensity increase of the manipulated group and made the intensity change in cerebellum undetectable. In the Andersson normalization, cerebellum may be included in the global estimate after the initial step of the GMN. Considering the size of cerebellum, the global change was overestimated, therefore pushing the intensity of all voxels to an opposite direction of the real SUV change. Such influence is demonstrated by Fig. 3 and Table 2. Both GMN and Andersson method recovered only a part of real SUV increase and induced a large artificial SUV decrease. ReDIN did not suffer such issue.

## 4 Discussion

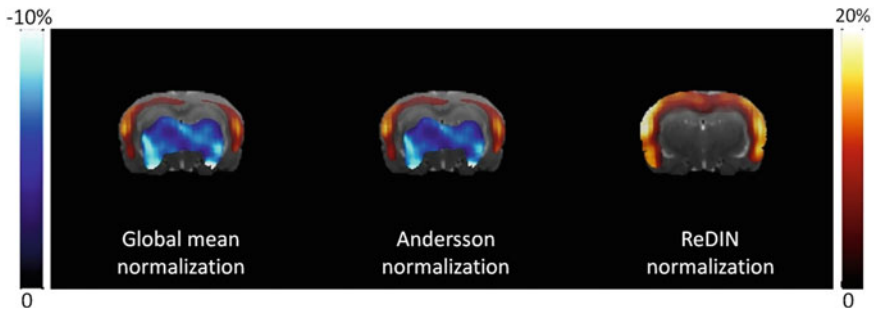
We presented an image normalization method and compared the proposed method with the GMN and the Andersson method. Simulation studies demonstrated that the proposed method yielded the best result in recovering the real metabolic change. The GMN suffered problems when the whole-brain average was affected by the local activities. The performance of the Andersson method is somewhere in between the GMN and the proposed method.

The propose method identifies anatomical regions instead of individual voxels for intensity normalization. The region-based method has two benefits: it reduced the computational cost by reducing the least squares regression from more than ten thousand voxels to less than a hundred of regions. Secondly, we can report the name of anatomical regions instead of using pictures to illustrate the region used for normalization. When preferred, some regions can be excluded based on prior knowledge.

**Table 1** Quantitative comparison of different normalization methods in simulation study I

ROI	True change (%)	Global mean		Andersson		ReDIN	
		Percentage	P-value	Percentage	P-value	Percentage	P-value
CUL	1.05	-4.03	0.0508	-1.56	0.4352	0.95	0.5879
AINS	15.44	13.55	0.0007	16.58	0.0002	17.24	0.0002
AUD	12.13	10.79	<0.0001	13.67	<0.0001	13.58	<0.0001
MOP	9.77	6.43	0.0028	9.20	0.0002	9.34	<0.0001
MOS	8.37	4.15	0.0884	6.86	0.0082	7.51	0.0035
PIR	12.61	7.75	0.0127	10.60	0.0020	12.13	0.0011
RSP	7.40	4.19	<0.0001	6.90	<0.0001	7.34	<0.0001
SSp	9.88	7.15	<0.0001	9.93	<0.0001	10.00	<0.0001
SSs	12.83	8.95	0.0006	11.80	0.0001	11.79	0.0001
TE	12.30	6.41	0.0006	9.18	0.0003	9.83	0.0002
VIS	8.23	6.18	0.0678	8.89	0.013	9.44	0.0069
DG	1.06	-4.45	0.0011	-1.96	0.118	-0.36	0.7589
TH	0.13	-4.45	0.0008	-1.97	0.1109	-1.54	0.0946

CENT-central lobule cerebellum, SIM-simple lobule cerebellum, CUL-culmen cerebellum, AINS-agranular insular ctx, AUD-auditory ctx, MOP-motor ctx primary, MOS-motor secondary ctx, PIR-piriform ctx, RSP-retrosplenial ctx, SSp-somatotemporal ctx primary, SSs-somatotemporal ctx secondary, TE-temporal ctx, VIS-visual ctx, DG-dentate gyrus hippocampus, TH-thalamus



**Fig. 3** In simulation II, moderate signal increase was added to some regions of cerebellum. Both the Andersson's method and the GMN created large areas of false SUV decrease (showed in *blue color*) where ReDIN method was not affected

Besides the applications in functional image analysis, the proposed method can be used in anatomical MRI analysis as well. For example, in tensor-based morphometry, Jacobian determinant is used to characterize the structural difference between groups at voxel level. To reduce the inter-subject variation, some studies calculated the Jacobian determinant only based on the deformation field and ignored the structural difference captured in affine registration. In Alzheimer's disease (AD) research, Affine transformation needs to be included in Jacobian determinant because there may be a whole brain atrophy in AD patients. In this case, the proposed method can be used to reduce the inter-subject variation in group comparison.

Although ReDIN normalization is developed based on preclinical data, it is straightforward to extend this method to clinical applications. In presented studies, a rigid body registration was used for spatial normalization due to the low resolution of PET image relative to the brain size of the rat. In clinical studies, more accurate deformable registration may be used if anatomical MRI of same subject is available. Software packages have been developed by different groups and were evaluated by Klein et al. [11].



**Table 2** Quantitative comparison of different normalization methods in simulation study II

ROI	True change (%)	Global mean		Andersson		ReDIN	
		Percentage	P-value	Percentage	P-value	Percentage	P-value
CENT	7.17	3.04	0.0756	3.50	0.04170	7.25	0.0001
CUL	7.58	1.92	0.3548	2.36	0.2505	7.81	0.0081
SIM	8.56	1.32	0.6346	1.75	0.5241	7.96	0.0372
AINS	15.44	12.75	0.0012	13.28	0.0010	18.07	<0.0001
AUD	12.13	10.01	<0.0001	10.50	<0.0001	14.08	<0.0001
MOP	9.77	5.67	0.0069	6.14	0.0046	9.80	<0.0001
MOs	8.37	3.41	0.1556	3.87	0.1061	8.02	0.0025
PIR	12.61	6.99	0.0225	7.48	0.0170	13.06	0.0002
RSP	7.40	3.45	<0.0001	3.91	<0.0001	7.76	<0.0001
SSp	9.88	6.39	0.0003	6.85	0.0003	10.42	<0.0001
SSs	12.83	8.18	0.0013	8.69	0.0012	12.29	<0.0001
TE	12.30	5.66	0.0133	6.12	0.0073	10.37	0.0002
VIS	8.23	5.43	0.1038	5.86	0.0785	9.85	0.0075
DG	1.06	-5.12	0.0003	-4.70	0.0011	0.14	0.8973
TH	0.13	-5.12	0.0002	-4.70	0.0009	-1.25	0.1597

## References

1. Aanerud, J., Borghammer, P., et al.: Brain energy metabolism and blood flow differences in healthy aging. *J. Cereb. Blood Flow Metab.* **32**(7), 87–1177 (2012)
2. Andersson, J.L.: How to estimate global activity independent of changes in local activity. *Neuroimage* **6**(4), 237–44 (1997)
3. Borghammer, P., Aanerud, J., et al.: Data-driven intensity normalization of PET group comparison studies is superior to global mean normalization. *Neuroimage* **46**(4), 981–8 (2009)
4. Boussion, N., Ryvlin, P., et al.: Towards an optimal reference region in single-photon emission tomography difference images in epilepsy. *Eur. J. Nucl. Med.* **27**(2), 60–155 (2000)
5. Camargo, E.E.: Brain SPECT in neurology and psychiatry. *J. Nucl. Med.* **42**(4), 23–611 (2001)
6. Cohen, B.D., Rosenbaum, G., et al.: Comparison of phencyclidine hydrochloride (Sernyl) with other drugs. Simulation of schizophrenic performance with phencyclidine hydrochloride (Sernyl), lysergic acid diethylamide (LSD-25), and amobarbital (Amytal) sodium; II. Symbolic and sequential thinking. *Arch. Gen. Psychiatry* **6**, 395–401 (1962)
7. Duncan, G.E., Miyamoto, S., et al.: Comparison of brain metabolic activity patterns induced by ketamine, MK-801 and amphetamine in rats: support for NMDA receptor involvement in responses to subanesthetic dose of ketamine. *Brain Res.* **843**(1–2), 83–171 (1999)
8. Fox, P.T., Mintun, M.A., et al.: Enhanced detection of focal brain responses using intersubject averaging and change-distribution analysis of subtracted PET images. *J. Cereb. Blood Flow Metab.* **8**(5), 53–642 (1988)
9. Friston, K.J., Frith, C.D., et al.: Comparing functional (PET) images: the assessment of significant change. *J. Cereb. Blood Flow Metab.* **11**(4), 9–690 (1991)
10. Gracia Marco, R., Aguilar Garcia-Iturrospe, E.J., et al.: Hypofrontality in schizophrenia: influence of normalization methods. *Prog. Neuropsychopharmacol. Biol. Psychiatry* **21**(8), 56–1239 (1997)
11. Klein, A., Andersson, J., et al.: Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *Neuroimage* **46**(3), 786–802 (2009)
12. Lee, S.J., Lee, W.Y., et al.: Apparent relative hypermetabolism of selective brain areas in Huntington disease and importance of reference region for analysis. *Clin. Nucl. Med.* **37**(7), 8–663 (2012)
13. Talbot, P.R., Lloyd, J.J., et al.: Choice of reference region in the quantification of single-photon emission tomography in primary degenerative dementia. *Eur. J. Nucl. Med.* **21**(6), 8–503 (1994)
14. Yakushev, I., Hammers, A., et al.: SPM-based count normalization provides excellent discrimination of mild Alzheimer's disease and amnesic mild cognitive impairment from healthy aging. *Neuroimage* **44**(1), 43–50 (2009)