Master Thesis

Evaluation of intensity normalization methods for MR images

Author
M.J.A. Jansen

Examination Committee
Prof. dr. ir. C.H. Slump
Dr. H.J. Kuijf
Dr. ir. F. van der Heijden
Ir. E.E.G.Hekman

July 2015

UNIVERSITY OF TWENTE.

Contents

1 List of Abbreviations

CSF	Cerebrospinal fluid
DSC	Dice score
GM	Grey matter
HD	Hausdorff distance
KL	Kullback-Leibler
NAWM	Normal appearing white matter
V	Volume
WM	White matter
WML	White matter lesion

2 Abstract

MRI intensities do not have a tissue specific value, meaning that the same tissue has a wide range of intensities, even within the same protocol, subject, and MR scanner. This intensity variation might affect image processing, since many segmentation methods and other tasks are based on intensity. Several methods to normalize MRI intensities have been proposed over the years. In this study, six intensity normalization methods for MRI are evaluated. The evaluation is based on the ability to create similar tissue intensities within a group of subjects, between repeated scans of the same subjects, and between scanners with different magnetic field strengths. Also the effect on a simple k-Nearest Neighbour (kNN) segmentation is tested quantitatively. The evaluations are performed on various data sets of T1 weighted images.

3 Samenvatting

De grijswaarden van een MRI scan hebben geen weefselspecifieke waarde. Dit betekent dat de grijswaarden, ook wel intensiteiten, van eenzelfde weefsel verschillen binnen hetzelfde scanprotocol, hetzelfde subject, en dezelfde MRI scanner. Deze intensiteitverschillen hebben invloed op de beeldverwerking, omdat veel segmentatie methoden en andere beeldverwerkingsmethoden zijn gebaseerd op intensiteit. De afgelopen jaren zijn er verschillende methoden ontwikkeld om MRI intensiteiten te normaliseren. In dit onderzoek worden zes verschillende intensiteitsnormalisatiemethoden voor de MRI geëvalueerd. De evaluatie is gebaseerd op het vermogen van de methoden om vergelijkbare weefselintensiteiten te creëren binnen een group subjecten, tussen herhaalde scans van dezelfde subjecten, en tussen scanners met een verschillende magnetische veldsterkte. Ook wordt de invloed van de normalisatie op een simpele k-Nearest Neighbour (kNN) segmentatie getest. De evaluaties zijn uitgevoerd op verschillende datasets van T1 gewogen beelden.

4 Introduction

Magnetic Resonance Imaging (MRI) is an advanced, non-invasive imaging technique that gives an excellent contrast between soft tissues [1]. However, one of the major disadvantages of MRI is that the tissues do not have a specific intensity, such as in computer tomography. Similar protocols show different intensities for the same tissue type, even within the same subject. See Figure 1 These variations are machine-dependent and cannot be corrected with bias field correction [2]. These intensity variations make segmentation and image analysis difficult[3]. Therefore, intensity normalization is an important pre-processing step for MR image analysis. Segmentation methods can benefit from intensity normalization and produce accurate and consistent results with less errors [4, 5, 6].

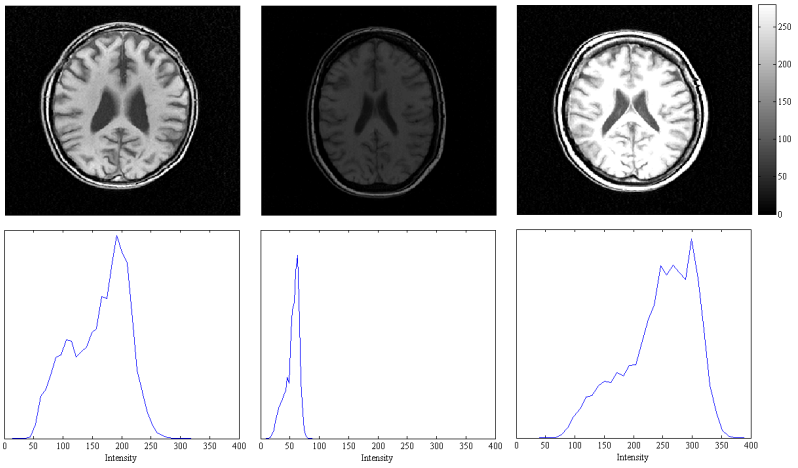

Figure 1: **Three images of three different subjects, scanned with the same protocol and displayed with the same window settings, and their intensity histograms.**

Several intensity normalization methods have been developed to reduce the intensity variety with various degrees of success. Most methods are based on histogram matching, with different approaches for landmark selection, e.g. mean intensity [7], even-order derivatives of white matter peak [8], or the median intensity in the foreground [4]. Some methods match the (mean) intensities of one or more tissues to a reference model [2, 3, 5]. Most of these proposed methods have been evaluated on healthy tissue, whereas these methods are mostly applied in the presence of pathological tissue. Only a small number of studies have investigated the effect of presence of e.g. white matter lesions (WMLs) on the normalization methods. The aim of this study is to evaluate several intensity normalization methods for T1 weighted brain MR images with the presence of WMLs due to cerebrovascular disease. A second aim is to provide recommendations about the usage of intensity normalization in image processing. The methods are selected based on reviews, references by articles, and originality, to cover the whole field of intensity normalization methods.

The normalization methods will be assessed on four aspects. The first aspect is the ability to create similar tissue intensities within a group of subjects with different amounts of WMLs. In cross-sectional studies the scans of different subjects are compared and a consistent intensity distribution might be beneficial for accurate and consistent image processing tasks. The second aspect is the ability to create similar tissue intensities within repeated scans of the same subject, for subjects with and without cerebrovascular disease. For longitudinal studies, a similar intensity distribution for different scans of the same subject is beneficial when comparing the scans. The effect of the presence of pathologies on the methods will be evaluated as well. The third as-

pect is the ability to create similar tissue intensities in repeated scans of the same subject made in MRI scanners with different magnetic field strengths. The tissue intensities differ between MRI scanners of different magnetic field strength or from different vendors [9]. For longitudinal studies a similar intensity distribution for different scans of the same subject is beneficial when comparing the scans, especially when the intensities are so different. The last aspect is the effect on a simple k-Nearest Neighbour (kNN) segmentation. This evaluation reflects on the benefits an intensity normalization method can have on intensity based image processing tasks. Another study already proofed the benefit of intensity normalization for a more advanced, atlas based segmentation method [10] when scans from different MR scanner platforms are used [6]. Intensity normalization is therefore beneficial in image processing tasks for multicentre studies.

5 Materials and Methods

Various intensity normalization methods have been proposed. Six methods were selected for evaluation. The selection was based on reviews, references of other articles, and the originality of the method. The methods vary in complexity, computational time, the usage of a reference model, and dependence on other pre-processing steps. For each method the advantages and limitations are described below and in subsection 8.3. The computational time was measured on an Intel Core2 Quad CPU processor @ 2.83 GHz for a scan with the matrix size of 38×256×256.

5.1 Intensity normalization methods

5.1.1 Gaussian

This method, referred to as Gaussian, is the most simple method. It rescales the intensities by $I_{new} = I/SD$. Where I is the intensity and SD is the standard deviation of the whole scan. The method assumes that each scan has the same intensity distribution .The rescaling is done based on this assumption [7]. The computational time is approximately 2 seconds.

5.1.2 Z-score

This method, referred to as Z-score, is also known as zero mean unit variance. It rescales and shifts the intensities by $I_{new} = (I - \mu)/SD$. Where I is the intensity, μ is the mean intensity and SD is the standard deviation of the whole scan. The method assumes that each scan has the same intensity distribution. The rescaling and shifting is done based on this assumption [7]. All means are rescaled to zero, but the means do not automatically correspond to the same tissue type. The computational time is approximately 2 seconds.

5.1.3 Histogram matching on median (HM_median)

This method, referred to as HM_median, was first described by Nyúl and Udupa (2000) [4]. This two-stage method consists of a training step and a transformation step. The principle is to find the white matter (WM) peak by taking the mean intensity of all the slices of the scan. This mean is used as threshold between background and foreground. The mode of the foreground is taken as the median landmark. Two other landmarks are the 0^{th} and 99.8^{th} percentile of the entire intensity histogram. During the transformation step these three landmarks are derived from the intensity histogram and piece-wise linear mapped to the standard scale landmarks.

The standard scale landmarks are determined during the training step with the use of multiple scans. Standard scale landmarks s_1 and s_2 are chosen in such a way that the mapping is one-to-one and that the histogram is not compressed. These landmarks correspond to the 0^{th} and 99.8^{th} percentile of the entire intensity histogram, respectively. The mode of the foreground is mapped on this new standard scale and the average remapped mode of the training scans is

taken as the standard scale median landmark. The computational time for the transformation step is approximately 6 seconds.

5.1.4 Histogram matching on generalized ball scale (HM_ballscale)

This method was first described by Madabhushi and Udupa (2006) [3] and is a variation of the HM_median method. The difference is the way to retrieve the WM peak. With the use of a manual selection of the WM and the generalized ball scale, described by Madabhushi, Udupa and Souza (2006) [11], the largest connected homogeneous area is determined. See Appendix 8.3. This area is considered to be the normal appearing white matter (NAWM) area. The median intensity of the intensity histogram of this area is selected as landmark. The other two landmarks are again the 0^{th} percentile and 99.8^{th} percentile of the entire intensity histogram. The intensities are piece-wise linear mapped to the standard scale landmarks. The standard scale landmarks are obtained during the training step. This training step is the same as for the HM_median method.

In the method described the whole 3D scan is included for determination of the median WM landmark, but due to the long computational time, only the middle slice of the scan is used in this study for determination of the landmark. The computational time is approximately 80 seconds, including the manual selection of NAWM.

5.1.5 Standardization of Intensities (STI)

This method is called Standardization of Intensities (STI), and was proposed by Robitaille et al. (2011) [2]. The subject scan is nonlinear registered to a reference scan. The intensities of the reference scan are rescaled to intensities between 0 and 100. The masks of white matter (WM), grey matter (GM), and cerebrospinal fluid (CSF) of the reference scan are applied to both the reference and the subject scan. A joint histogram is made of the subject WM area and the reference WM area. The mode of the joint histogram is selected as landmark for WM. This is repeated for GM and CSF. The minimum intensity and the maximum intensity of the subject scan are set to 0 and 100 respectively and are also used as landmarks. The transformation is a linear piece-wise mapping between the five landmarks.

A proper reference scan has a similar intensity distribution as the subject scan. The computational time is approximately 80 seconds.

5.1.6 MIMECS

This method was developed by Roy, Carass and Prince (2013) and the software is called MIMECS [5, 12]. The intensities of the subject scan are normalized to the intensities of a reference scan with the use of $3 \times 3 \times 3$ voxel patches. Both scans are normalized by dividing by the WM peak intensity, in order to have an almost similar intensity range. This WM peak intensity is predetermined by the user. Then for each voxel a surrounding patch is defined. The best matching reference patch for each subject patch is defined by the maximum likelihood and found using the expectation maximization algorithm [5]. The central voxel intensity of the subject is replaced with the matching central voxel intensity of the reference. The reference scan should contain all the pathologies that are expected in the subject scan for a good intensity normalization. The computational time is approximately 4 hours.

5.2 Data

The data of three different MRI scanners and sequences are used to evaluate the six normalization methods. All the scans are bias field corrected and segmented by SPM12[13]. Each data set has

Table 1: **Training set and reference scans for the normalization methods.**

Method	Data set A	Data set B	Data set C
HM_median	Training set of 8 randomly chosen scans of first subject.	Training set of 4 random scans from both SMART-MR and SMS study.	Training set of 10 randomly selected scans.
HM_ballscale	Training set 8 randomly chosen scans of first subject.	Training set of 4 random scans from both SMART-MR and SMS study.	Training set of 10 randomly selected scans.
STI	First scan of first subject; reference masks originate from MNI-152 atlas.	Scan with no brain pathologies, of average age, and average brain volume; reference masks originate from MNI-152 atlas.	Scan with least amount of pathology; reference masks originate from SPM12 segmentation.
MIMECS	First scan of first subject.	Scan with average amount of WMLs present.	Scan with medium load of WMLs present.

its own training and reference scans used for normalization.

Data set A In data set A, three healthy subjects are scanned twice a day for 20 days within a 30-day period on a GE MR750 3 T scanner (software version DV22.0_V02_1122.a, XRMB gradient set) and with the ADNI-recommended T1 weighted imaging protocol for this system (accelerated sagittal 3D IR-SPGR, TR: 7.3 ms, TE: 3 ms, TI: 400 ms, flip angle: 11 deg., FOV: 270 mm × 270 mm, matrix size: 256×256×196, voxel size: 1.055 mm × 1.055 mm × 1.2 mm, standard 8-channel phased array head coil, acquisition time: 5 min 37 s) [14].

Data set B The scans are from the SMART-MR study and were made on a 1.5 T whole-body system (Gyroscan ACS-NT, Philips Medical Systems, Best, the Netherlands) with a transversal T1 weigthed gradient-echo sequence (TR: 235 ms, TE: 2 ms, flip angle: 8 deg., FOV: 230 mm × 230 mm, matrix size: 256×256×38, voxel size: 0.898 mm × 0.898 mm × 4.0 mm) [15]. Within the SMS follow-up study the subjects were scanned for the second time with the same protocol, approximately five years after the baseline. A total of 1309 subjects were scanned within the SMART-MR study.

Data set C In data set C, the subjects were scanned on a 3 T whole-body system (Philips Medical Systems, Best, the Netherlands) at the UMC Utrecht with a T1 weighted sequence (TR: 7.9 ms, TE: 4.5 ms, FOV: 230 mm × 230 mm, matrix size: 240×240×48, voxel size: 0.958 mm × 0.958 mm × 3.0 mm) [16]. A total of 207 subjects were scanned with this protocol.

All the scans were normalized according to the normalization method described above. In Table 1 the training sets and reference scans properties are given for the methods using a training or reference set. The first scan of subject one from Data set A is used as reference for the STI and MIMECS method and is the only scan used as reference and as subject in the study.

5.3 Evaluation methods

The six normalization methods are evaluated on four aspects. These aspects are;

Cross-sectional evaluation: The ability to create similar tissue intensities within a group of subjects with different white matter lesion loads.

Longitudinal evaluation: The ability to create similar tissue intensities within repeated scans of the same subject, for subjects with and without cerebrovascular disease.

Interscanner reproducibility: The ability to create similar tissue intensities in repeated scans of the same subject made in MRI scanners with different magnetic field strengths.

Segmentation: The effect on a simple kNN segmentation.

5.3.1 Cross-sectional

Data selection For the cross-sectional evaluation, 30 scans were selected from data set C. Selection was based on the white matter lesion load and the absence of large (cortical) infarcts and artefacts. The lesion load is defined as the volume of WMLs. Three groups of 10 subject scans were made, in order to compare the effect of WMLs on the intensity normalization method. The lesion load of the scans determined the groups. The lesion load of the low lesion load group is 0.6 ± 0.3 ml. The lesion load of the medium lesion load group is 4.6 ± 0.4 ml. The lesion load of the high lesion load group is 30.7 ± 11.6 ml.

Evaluation At first a visual examination of the intensity distributions of WM, GM, CSF, and WML, before and after normalization is given. The distributions of the intensities of the 30 scans were plotted in one image. A good normalization method provides similar intensity distributions across the subjects.

For each scan, the mean intensity of each tissue was determined. Using the 30 mean intensities a mean intensity and coefficient of variance (CV) was calculated. The CV was also calculated for each lesion load group. A smaller CV means that the mean intensities are more similar. A visual representation of some slices with a fixed colour window is given as well. The window is set on the first scan and applied to the other scans.

The Kullback-Leibler (KL) divergence between a Gaussian and the intensity distribution for WM and GM was calculated as well. The means and the standard deviations of the intensities of WM and GM of the first scan of the medium lesion load group were used to compute the reference Gaussian intensity distribution for each tissue type. The KL divergence between this Gaussian and the normalized intensity distribution is a measure for the similarity of the intensity distribution and therefore for the performance of the normalization method [17]. A Gaussian distribution was used as reference since the WM and GM intensity distributions should be represented by Gaussians [18]. This calculation was also performed on each lesion load group separately. The KL divergence was also calculated between the intensity distributions of different tissues. In Figure 2A, a schematic overview of the evaluation method is given.

Analysis The nonparametric Wilcoxon signed rank test was applied to test for a significant difference between the KL divergence outcomes of the original scans and the normalized scans. This test was applied because this data was not normally distributed. A Bonferroni corrected significance level of $p < 0.05$ was set for the level of significance. The mean intensities are normally distributed.

5.3.2 Longitudinal

Data selection For the longitudinal evaluation, scans of data set A and B were used. Data set A consists of healthy subjects and data set B consists of subjects with WMLs and infarcts. Ten combinations of a baseline and a follow-up scan were selected from data set A.

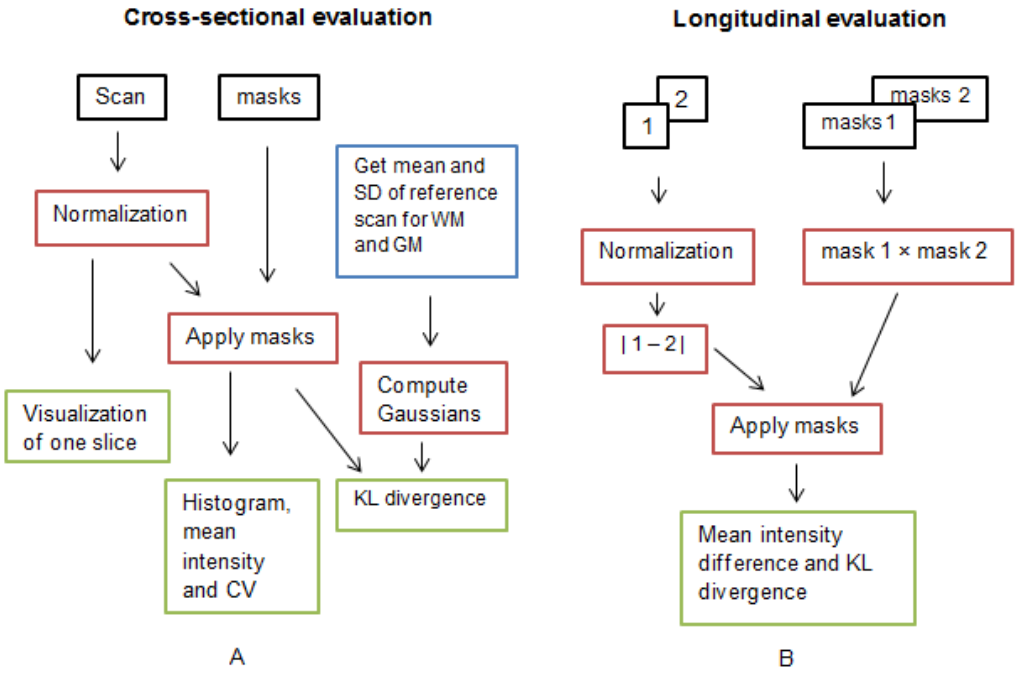

Figure 2: **A) Schematic overview of the cross-sectional evaluation. The reference scan is the first scan of the medium lesion load group. B) Schematic overview of the longitudinal evaluation. Numbers 1 and 2 represent the baseline and follow-up scan, respectively. This overview also represents the interscanner reproducibility evaluation.**

Twenty-five combinations of a baseline and a follow-up scan were randomly selected from data set B, where subjects with WMLs and infarcts were included. The mean volume of the WMLs in the baseline scan is 6.76 ± 8.55 ml and in the follow up scans 6.94 ± 7.85 ml. The mean volume of the infarcts in the first scan is 9.93 ± 17.60 ml and in the follow up scans 18.71 ± 29.03 ml.

All the scans for the longitudinal evaluation were linear registered to the MNI-152 atlas using elastix [19]. Therefore the repeated scans of the same subject and can be compared. White matter lesion segmentations were provided by the SMART-MR study [15].

Evaluation In order to assess the intensity distribution similarity, the absolute intensity difference between baseline and follow-up scan was calculated after normalization. For WM, GM, and CSF the mean absolute intensity difference was calculated, to assess the normalization performance for each tissue type. The resulting intensity range differs between the normalization methods, so in order to compare the results, the values were normalized. This was done by dividing the outcome by the intensity range of interest (IOI). The IOI is defined as the 99.8^{th} percentile intensity minus the minimum intensity.

The KL divergence between the intensity probability distributions of the brain volume (WM and GM) of both scans was also calculated, using 121 bins. A smaller KL divergence corresponds to more similar intensity distributions [17]. In Figure 2B, a schematic overview of the evaluation is shown.

The subjects with cerebrovascular disease were scanned with an interval of five years and the same evaluation was applied as described above. Due to changes in the brain during these five years, there could be a large difference in intensities after normalization when subtracting

the two brain volumes. Therefore only the voxels with the same tissue types were taken into account. Also the mean WML intensity was calculated for each scan and then the absolute difference between the lesion intensities is calculated. This is done, because the WMLs were not necessarily located on the same spots in both scans. The KL divergence between the intensity histograms of the lesions, using 121 bins, was also obtained.

Analysis The nonparametric Wilcoxon signed rank test was applied to test for a significant difference between the evaluation outcomes of the original scans and the normalized scans. This test was applied because the data was not normally distributed. A Bonferroni corrected significance level of $p < 0.05$ was set for the level of significance.

5.3.3 Interscanner reproducibility

Data selection Ten subject scans were acquired with the MR protocol of data set B, on a 1.5 T scanner, and data set C, on a 3 T scanner, on the same day. The scans were linear registered to the MNI-152 atlas using elastix [19]. In order for the two corresponding scans to have a similar intensity distribution, all the scans were normalized using the training and reference scans of data set C.

Evaluation The evaluation was the same as for section 5.3.2 Longitudinal evaluation. WMLs might be present in the subjects brain, but no information and segmentation about these lesions is known. The possible lesions were therefore treated as WM tissue.

Analysis The nonparametric Wilcoxon signed rank test was applied to test for a significant difference between the evaluation outcomes of the original scans and the normalized scans. This test was applied because not all the data was normally distributed. A Bonferroni corrected significance level of $p < 0.05$ was set for the level of significance.

5.3.4 Segmentation

Data selection Twenty subject scans of data set C were manually segmented [20]. These scans have an average white matter lesion load of 7.7 ± 10.0 ml. Five scans are used to train a kNN-classifier and fifteen scans are used as test scans. A simple kNN classifier, with k =23 is trained with 40,000 randomly selected voxels for each tissue (GM, WM, and CSF). Segmentation and evaluation were performed according to the guidelines of the MRBrainS Challenge [20].

Evaluation The outcome of the segmentation by the kNN-classifier was evaluated within the MRBrainS platform [20]. The outcome was compared with the manual segmentations using the Dice score, the 95^{th} percentile of the Hausdorff distance, and the absolute volume difference.

The segmentation by the kNN-classifier is visualized for one slice for each normalization method.

Analysis Paired t-tests between original and the normalized results were done. The outcomes were also ranked; a score of 7 was assigned to the method with the worst result and a score of 1 was assigned to the method with the best result. This was done for all the outcomes.

6 Results

6.1 Cross-sectional

In Figure 3 the intensity distribution of WM for the 30 subjects for each normalized method is visualized. The lines represent the low (orange), medium (purple), and high (green) lesion load groups. The intensity distributions for GM, CSF, and WML are shown in Appendix 8.1. In Figure 4 the intensity distribution of the intracranial structures (WM, GM, and CSF) is visualized. It clearly shows that the HM_ballscale, STI and MIMECS methods provide a more similar intensity distribution than the other methods.

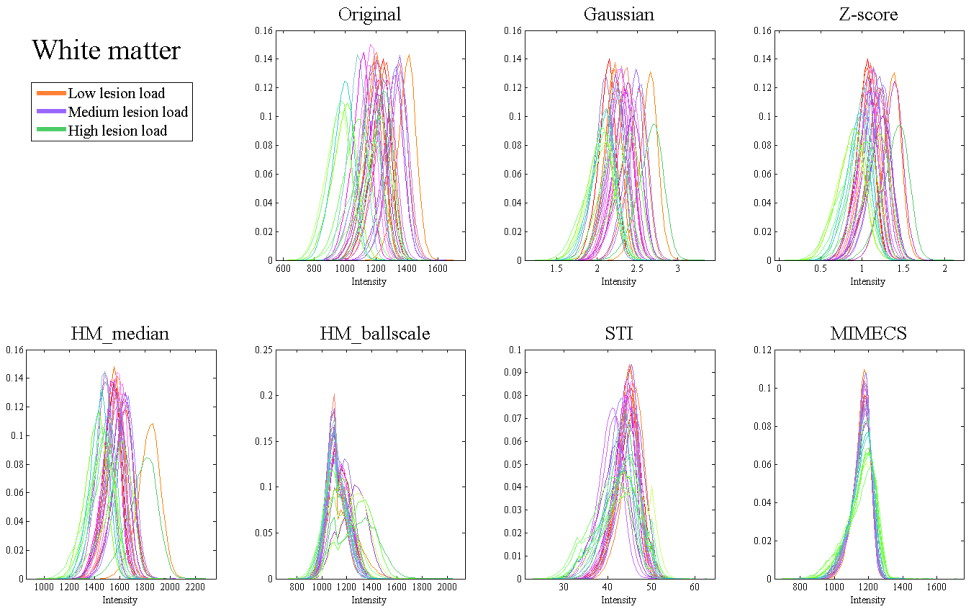

Figure 3: **Intensity distributions of the white matter area of 30 subjects. The three lesion load groups can be recognized by the colour of the histogram.**

In Table 2 the mean intensity and the coefficient of variation (CV) of the different tissue types are given. In Table 3 the CV is given for each tissue and each lesion load group. In general the performance of the normalization methods gets worse when the lesion load is high. The Z-score method has the worst performance when it comes to the similarity of the means of intensities. The MIMECS method performs best together with the STI, HM_median, and the HM_ballscale method. The latter two methods are negatively affected by a high lesion load.

In Figure 5 a slice of six different scans are shown with a fixed window. For each lesion load group two scans are shown. The window was determined on the first scan, row 1 in Figure 5, for each normalization method separately. The colours represent intensity values. The effect of the lesion load is clearly visible. A proper normalization method yields the same colour for the same tissue for all the scans. This is seen in MIMECS, STI and HM_ballscale. The other methods improve the consistency of the colours in comparison with the original data, but variation in the colours is visible.

In Figure 6 the KL divergence outcomes between a Gaussian and the real intensity distribution are shown for WM and GM. A lower KL divergence indicates a smaller difference between the Gaussian and the intensity distribution of the given tissue type.

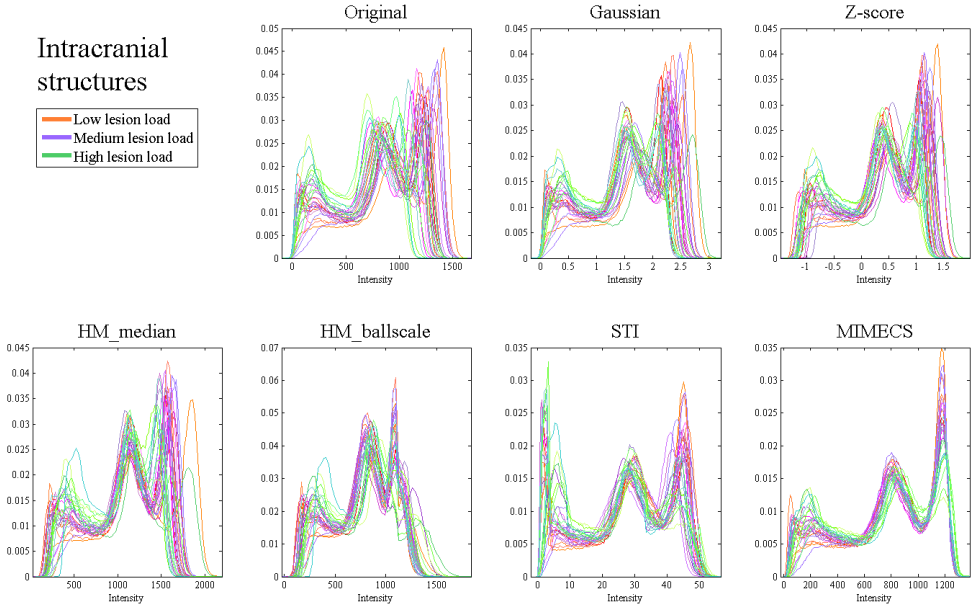

Figure 4: **Intensity distributions of the intracranial structures of 30 subjects. The three lesion load groups can be recognized by the colour of the histogram.**

Table 2: **Mean intensity of the different tissues types for each normalization method and the coefficient of variation (CV). A lower CV represents a more similar intensity distribution.**

	WM intensity		GM intensity		CSF intensity		WML intensity	
	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)	Mean	CV (%)
Original	1180	9.8	845	8.9	328	14.3	950	13.9
Gaussian	2.26	7.6	1.62	6.9	0.63	14.1	1.82	11.2
Z-score	1.09	12.9	0.45	20.5	-0.54	15.9	0.65	27.1
HM_median	1549	6.1	1155	3.3	563	10.4	1278	8.0
HM_ballscale	1146	4.4	840	4.6	411	12.7	926	4.9
STI	43.4	3.4	29.2	4.8	9.68	18.9	33.5	9.3
MIMECS	1157	0.3	830	2.1	329	13.0	941	6.3

Table 3: **Coefficient of variance (CV) of the mean intensity for each tissue, and each lesion load group. A lower CV represents a more similar intensity distribution.**

	WM CV (%)			GM CV (%)			CSF CV (%)			WML CV (%)		
Lesion load	**Low**	**Med.**	**High**	**Low**	**Med.**	**High**	**Low**	**Med.**	**High**	**Low**	**Med.**	**High**
Original	6.3	6.4	9.2	5.9	7.2	8.5	13.3	16.4	11.4	10.0	12.4	10.5
Gaussian	6.7	5.6	9.1	6.4	5.6	8.6	13.2	14.0	14.2	10.4	11.2	9.3
Z-score	10.3	6.3	16.7	17.0	17.2	25.1	13.2	19.3	16.3	18.1	26.0	25.7
HM_median	6.0	3.2	7.3	3.2	2.4	4.1	10.3	7.6	11.1	7.5	8.0	5.3
HM_ballscale	1.6	3.7	6.4	2.6	3.3	5.2	11.2	7.7	12.6	4.1	6.6	3.3
STI	2.7	3.8	2.6	4.1	5.3	4.5	12.6	18.9	20.5	7.3	10.7	3.7
MIMECS	0.1	0.2	0.4	1.5	1.3	1.9	9.0	10.7	15.3	5.4	7.8	3.4

The KL divergences were also calculated between the intensity distributions of different tis-

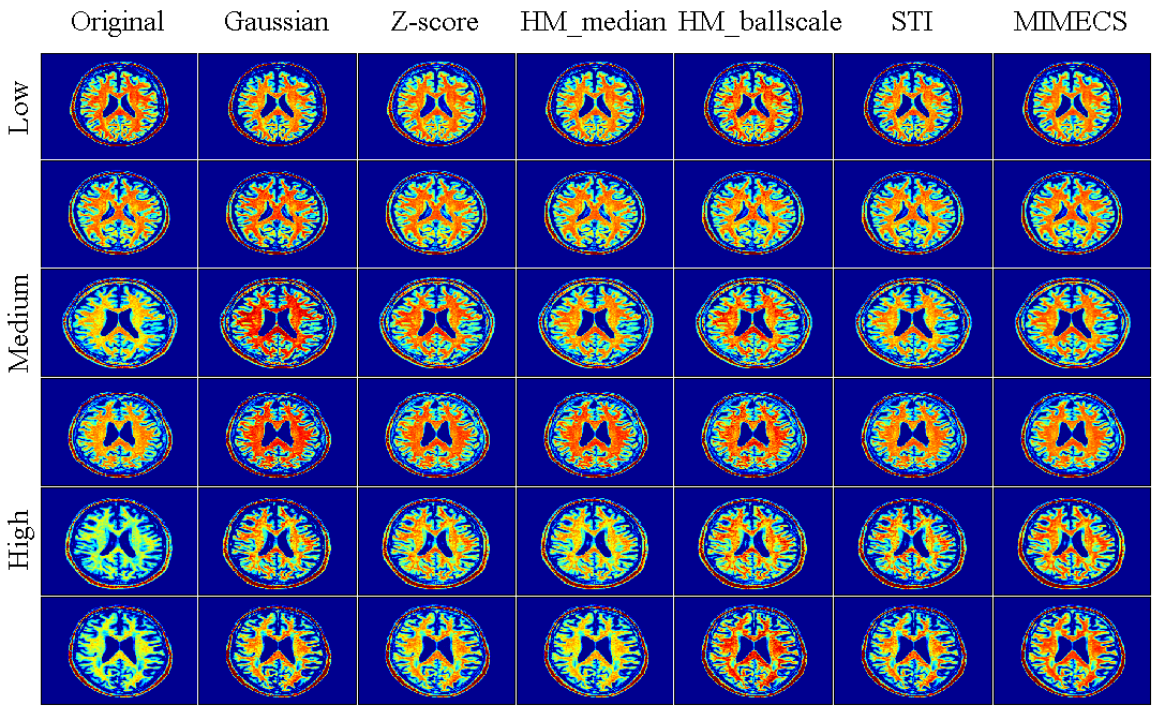

Figure 5: **Slices displayed at fixed window from the low, medium and high lesion load group. Each normalization method has its own fixed window. A consistent colour distribution suggests a better intensity normalization.**

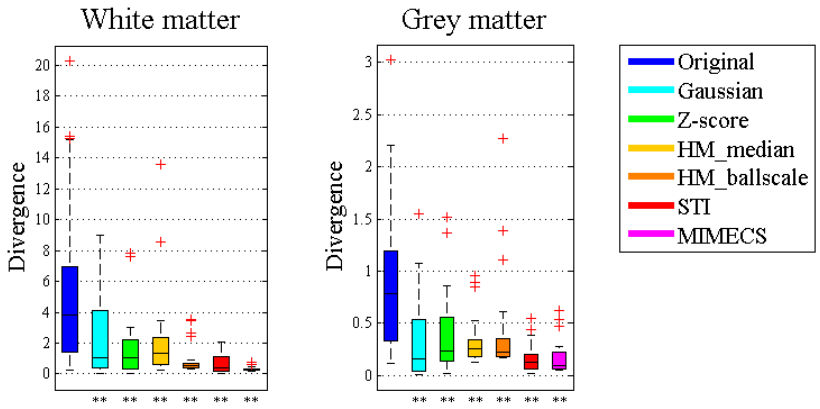

Figure 6: **KL divergence between a normal Gaussian and the intensity distribution of scan for WM and GM. Lower KL divergence suggests a better intensity normalization. Double asterisks indicate a significant difference between the original and normalized scan (Bonferroni corrected p-value<0.05).**

sue types. The difference in the outcomes is too small to be relevant. The results are given in Appendix 8.2, Table 5.

The KL divergence was also calculated for the three different groups separately, see Figure 7. The effect of the amount of lesions on the normalization method is made clear. The differences between the original intensity distribution and the normalized intensity distribution become

smaller when more WMLs are present. There are less significant differences in the high lesion load group than in the low lesion load group.

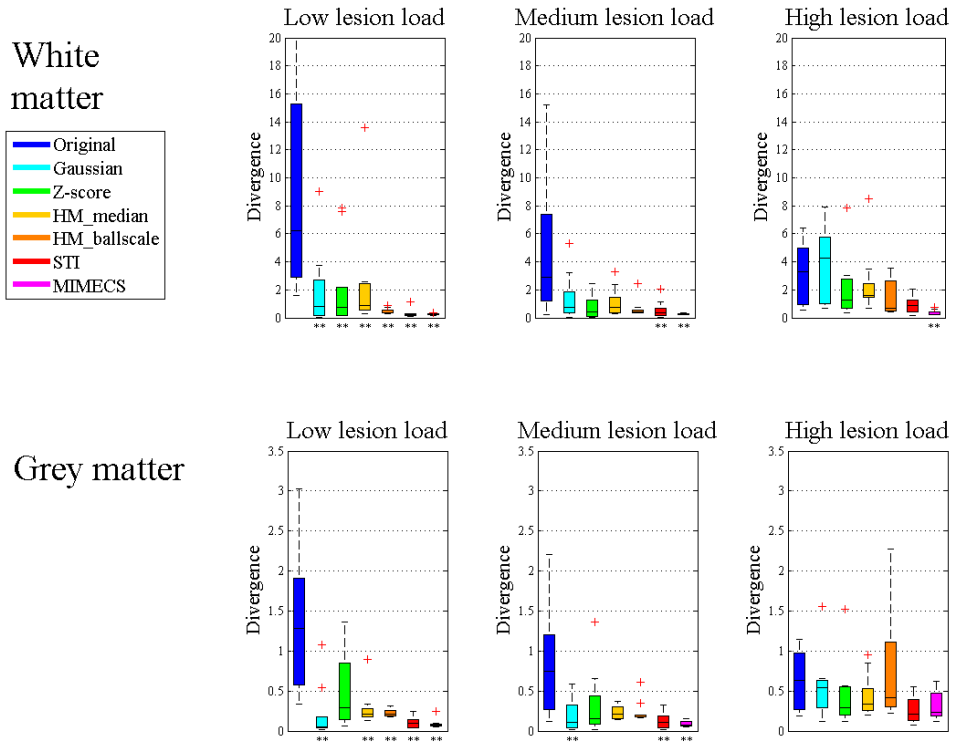

Figure 7: **KL divergence between a normal Gaussian and the intensity distribution of WM and GM for the different groups. Lower KL divergence suggests a better intensity normalization. Double asterisks indicate a significant difference between the original and the normalized scan (Bonferroni corrected p-value**< 0.05**).**

6.2 Longitudinal

The results of the longitudinal evaluation for the healthy subjects are shown in Figure 8. Wilcoxon signed rank tests were applied between the original outcomes of the evaluation and the normalized outcomes of the evaluation. Double asterisks below the bar indicate a significant difference between the original outcome and the normalized outcome. The effect of the intensity normalization is clearly visible. Especially for WM and GM the difference in intensities is smaller for all normalization methods.

In Figure 9 the results of the longitudinal evaluation of the pathologic subjects are shown. The effect of the intensity normalization is clearly visible for WM, GM and WML. Especially the more advanced methods (HM_ballscale, STI, and MIMECS) provide a smaller difference in intensities. The performance of the Gaussian and Z-score normalization methods are affected by the WMLs.

6.3 Interscanner reproducibility

The results of the evaluation of the normalization of 1.5 T and 3 T scanner scans are given in Figure 10. Wilcoxon signed rank tests were applied between the original outcomes of the evaluation and the normalized outcomes of the evaluation. The effect of the intensity normalization

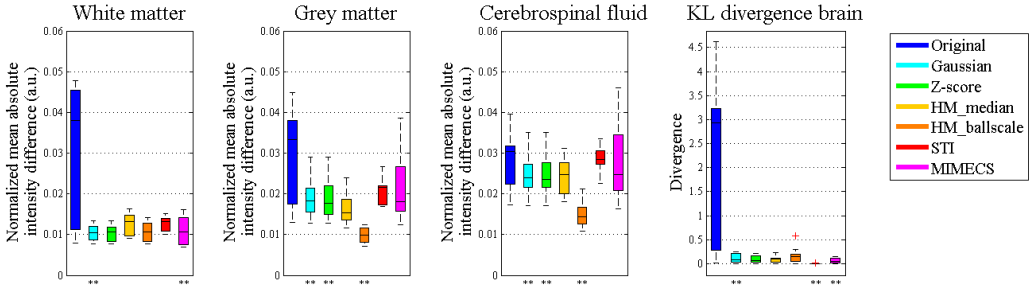

Figure 8: **Results longitudinal evaluation for healthy subjects. Lower outcomes suggest a better normalization, for both the normalized mean absolute intensity difference and the KL divergence. The double asterisks below a bar indicate a significant difference between the original scan and the normalized scan (Bonferroni corrected p-value<0.05).**

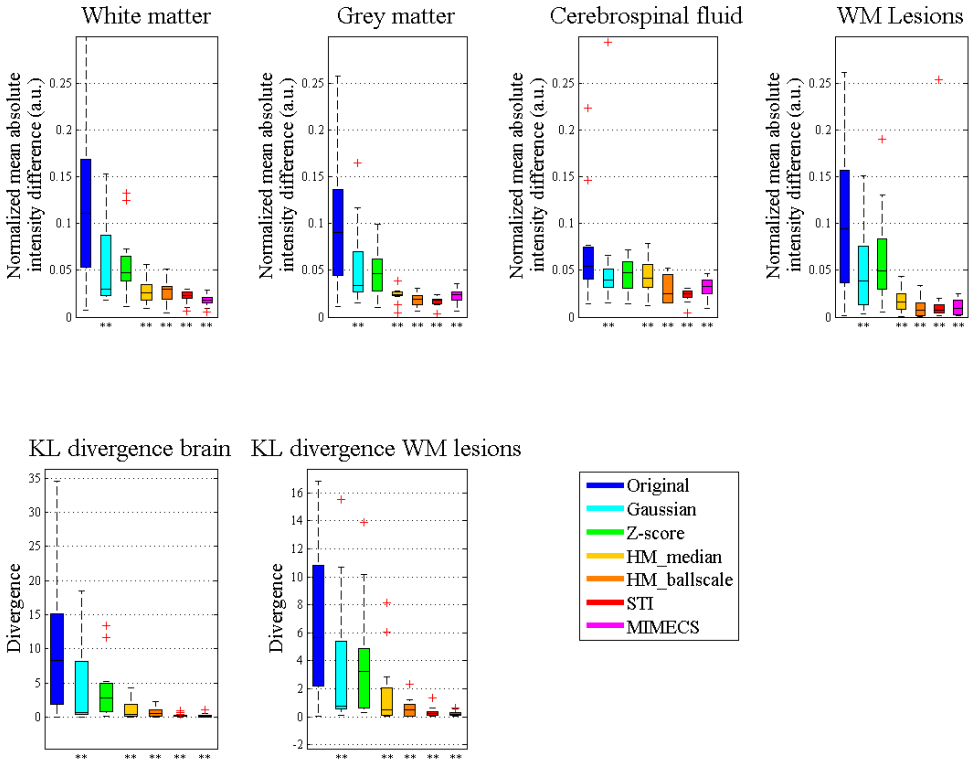

Figure 9: **Results longitudinal evaluation for pathological subjects. Lower outcomes suggest a better normalization, for both the normalized mean absolute intensity difference and the KL divergence. Double asterisks below the bar indicate a significant difference between the original scan and the normalized scan (Bonferroni corrected p-value<0.05).**

is clearly visible. All methods improve the similarity of the intensity distribution. HM_ballscale performs the best, especially in the CSF.

6.4 Segmentation

The results of the Dice score (DSC), Hausdorff distance (HD) and volume difference (V) between the manual segmentation and the kNN segmentation for each tissue type and each normalization method are shown in Figure 11.

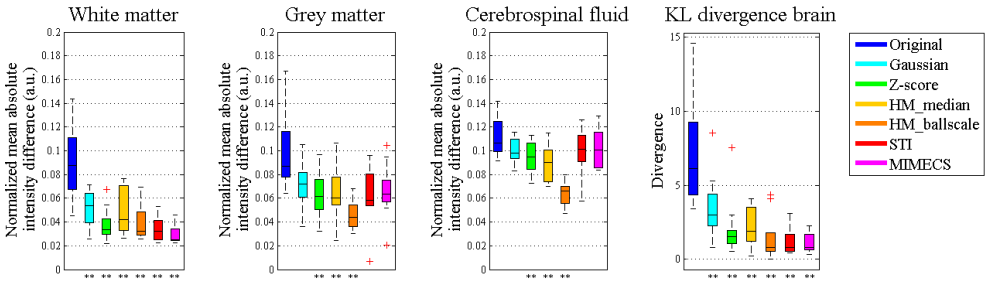

Figure 10: **Results evaluation interscanner reproducibility between 1.5 T and 3 T scanner. Lower outcomes suggest a better normalization, for both the normalized mean absolute intensity difference and the KL divergence. Double asterisks indicate a significant difference between the original and the normalized scan (Bonferroni corrected p-value**< 0.05**).**

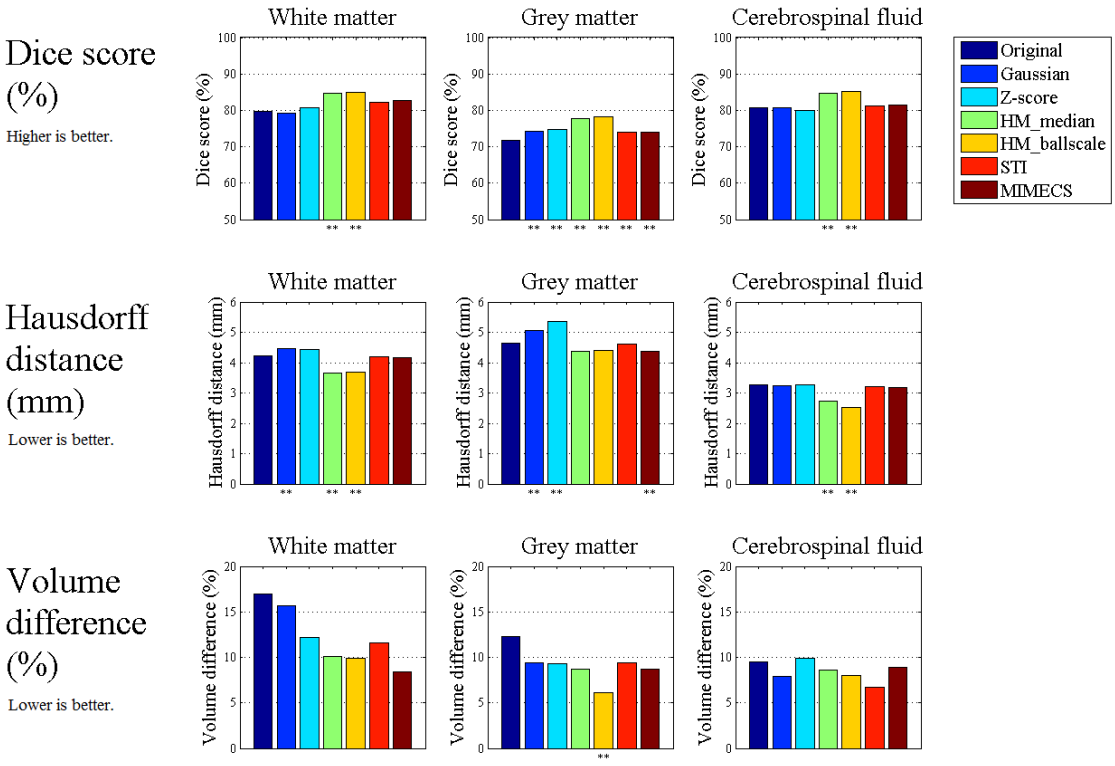

Figure 11: **Results segmentation evaluation; Dice score, Hausdorff distance and volume difference between the manual segmentation and the kNN segmentation for different tissues and normalization methods. Double asterisks indicate a significant difference between the original and the normalized scan according to a paired t-test (Bonferroni corrected p-value**< 0.05**).**

The results of Figure 11 were ranked and this can be found in Table 4. The HM_ballscale method is overall the best method to use for intensity normalization when it comes to kNN segmentation.

In Figure 12, the result of the segmentation by the kNN classifier is given for slice 25 of test-subject 1 for each normalization method.

Table 4: **Ranking performance of the normalization methods. DSC: Dice coefficient score (%), HD: Hausdorff distance (mm), V: absolute volume difference (%), Tot: Total of Dice, HD and V for that specific tissue.**

Rank	White matter				Grey matter				Cerebrospinal fluid				Total
	DSC	HD	V	Tot	DSC	HD	V	Tot	DSC	HD	V	Tot	
Original	7	5	7	**19**	6	5	7	**18**	5	7	6	**18**	**55**
Gaussian	4	6	5	**15**	7	7	6	**20**	6	5	2	**13**	**48**
Z-score	3	7	4	**14**	5	6	5	**16**	7	6	7	**20**	**50**
HM_median	2	1	3	**6**	2	1	3	**6**	2	2	4	**8**	**20**
HM_ballscale	1	3	1	**5**	1	2	2	**5**	1	1	3	**5**	**15**
STI	6	4	6	**16**	4	4	4	**12**	4	4	1	**9**	**37**
MIMECS	5	2	2	**9**	3	3	1	**7**	3	3	5	**11**	**27**

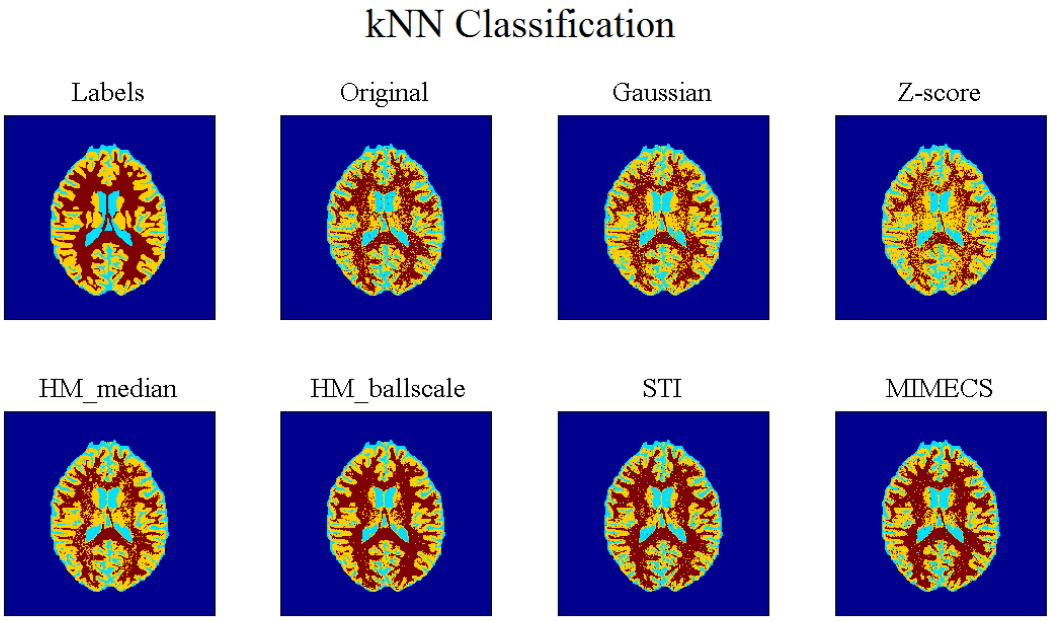

Figure 12: **kNN Classifier segmentation for the different normalization methods, k=23. The ground truth is the manual segmentation, denoted here as "Labels".**

7 Discussion

In this study, six intensity normalization methods were evaluated on several aspects. The results demonstrate the potential capabilities of these methods on creating a more consistent intensity distribution, within a group of subjects, in repeated scans of the same subject, and in repeated scans of the same subject on MR scanners with different magnetic field strengths. The effect of the amount of WMLs on these methods is also demonstrated.

Although the methods were only evaluated on T1 weighted images, the normalization methods can also be applied on other MR sequences. For each MR sequence the methods need to have a reference scan or training set, which are obtained with that specific sequence.

Cross-sectional This part of the evaluation clearly showed the effect of the amount of WMLs on the normalization methods. All methods are affected by the presence of large WML loads,

but STI and MIMECS are affected least. Gaussian and Z-score are the most affected by the high lesion load, because it changes the intensity distribution. Furthermore these methods assume that the intensity distribution is similar. The HM_ballscale method has difficulties finding the NAWM in high lesion load scans, because this area is not always the largest connected area. In three cases with high lesion load, the area detected was not the NAWM area and therefore the KL divergence had a large variety, see Figure 7. The method performs well when the lesion load is of low or medium volume.

Longitudinal In general, all normalization methods provide a more similar intensity distribution within the same subject; especially for WM and GM the normalized mean intensity difference is lower than for the original scans. For CSF, it is harder to provide a more similar intensity distribution, mainly because this is already similar in the original images. A larger time span between the two scans can give different intensity distributions, due to real anatomical changes, e.g. atrophy or WMLs. Methods primarily based on intensity distribution, i.e. Gaussian and Z-score, are affected by this. These methods perform less when there is a large anatomical difference between the baseline and the follow-up scan. Besides this, the Z-score method is having difficulties normalizing in the presence of WMLs, due to the change in the intensity distribution. The mean intensity is shifted owing to the presence of these lesions, which results in a shifted normalization and a larger KL divergence, as can be seen in Figure 9.

Interscanner reproducibility The difference in intensities between scanners with a different magnetic field strength is made smaller by all the normalization methods. The results are similar to the longitudinal evaluation, but the normalized mean intensity differences and the KL divergences are larger than for the longitudinal evaluation. This is probably due to the fact that the 1.5 T and the 3 T MRI scanners produce different intensity values. So the initial intensity difference was already larger for the interscanner reproducibility evaluation than for the longitudinal evaluation.

Segmentation A quantitative comparison of the effect of intensity normalization on kNN segmentation was performed. Although MIMECS provided the most similar scans in the cross-sectional evaluation, the segmentation was not superior. The reason is probably that voxels of different tissue types can have the same intensity. This results in an overlap of the different tissue histograms. This overlap remained after normalization with MIMECS. The methods HM_median and HM_ballscale decreased the overlap in intensity histograms by enlarging the difference between the intensities of the tissue types and provide therefore a better outcome for segmentation (cf Table 5 in Appendix 8.2). The STI method decreased the difference between the intensities of the tissue types. The significantly larger Hausdorff distance of the Gaussian and Z-score data, is caused by incorrect and noisy segmentations.

7.1 Limitations of the study

Cross-sectional The visualization of the intensity distribution by presenting the histograms, the CV, and the slices depicted on a fixed window, provided a clear presentation of the effect of the normalization methods on the intensity distribution and the effect of the amount of WMLs on the normalization methods, see Figure 4 and Figure 5.

The KL divergence between a Gaussian and the WM/GM intensity distribution in the cross-sectional evaluation is not the optimal measure to assess the similarity of the intensity distributions. The WM/GM distributions are not normally distributed and so an increase in the KL divergence can originate from this or from a misalignment in the intensities by the method. A distinction between the two causes is not possible. This can be improved by calculating the KL

divergence between the average tissue intensity probability distributions of all the scans and the tissue intensity probability distributions.

Longitudinal and Interscanner reproducibility In the longitudinal and interscanner reproducibility evaluation, the scans were all registered to the MNI-152 atlas, because the probability masks of the MNI-152 atlas were already available for the STI method. The result was a transformation and resampling from 38 slices to 193 slices. This resulted in an interpolation of the intensities and blurring of the scan, which deteriorated the quality of the scan. However the comparison is still valid, since all the scans were affected equally by this. The registration to the MNI-152 atlas is probably the best solution to make a fair comparison between the scans and to have good probability masks available.

The tissue probability masks obtained by SPM (WM, GM, and CSF)and WML masks were not optimal. A high threshold was set to the probability masks to produce the binary masks with a high specificity. Still, voxels from a different tissue type might be included in a mask. Especially WMLs were hard to distinguish from other tissue types for automatic software.

Segmentation This study suggests that some methods (HM_median and HM_ballscale) improve the outcome of a simple, intensity based kNN segmentation. This does not automatically mean that these methods also improve the outcomes of other image processing tasks. For example, a more consistent intensity distribution provided by STI or MIMECS might provide better outcomes for other, more advanced, segmentation methods (i.e. SPM or FreeSurfer [21]), than HM_median and HM_ballscale. Further research can give more conclusive answers on this topic.

Also the WMLs were classified as WM tissue and therefore the ability to make a distinction between these two tissues on intensity basis was not tested.

7.2 Intensity normalization methods

Several intensity normalization methods were evaluated during this study. The simplistic methods, as **Gaussian** and **Z-score**, are fast but perform less than other methods and are not robust to the presence of pathologies, such as WML.

HM_median is a fast method and increases the relative distance between intensity peaks of different tissue types, but it gives mediocre results when it comes to similarity of the intensity distributions. The ability to improve the intensity consistency and the segmentation was already verified, even when pathologies are present in the brain [4, 9, 21].

The **STI** method provides a more consistent intensity distribution than the original data and the HM_median method, which is concordant to Robitaille et al. [2]. The limitations of this method are the reliance on registration, and the need of a representative reference scan.

HM_ballscale gives good results and is semi-robust to the presence of pathologies. It was already shown that this method provides consistent intensities better than the original data and the HM_median method [3]. It also increases the relative difference between the intensity peaks of different tissues. The limitations are that the determination of the landmark in the subject scan takes some computational time and is not fully automatic. The implementation of this method was based on the method described by Madabhushi and Udupa [3], with a major difference in the selection of the landmark. In our study only one slice was used to generate the landmark and in the original method the whole 3D scan was used. The use of the whole scan might improve the stability of the landmark selection. Another deviation from the described method is the training part. The median landmark was defined by taking the average of the

median of the determined NAWM area of the original intensity scale, instead of the median on the standard scale.

The **MIMECS** method computes the same intensity distribution as the reference scan and creates therefore very similar intensity distributions. The pathologies present in the subject scan need to be present in the reference scan. A careful choice of a reference scan is recommended for the method to assign the right intensity to the right pathology. Other limitations are the large computational time (approximately 4 hours) and the need to provide the WM peak when the scan is not skull stripped.

7.3 Conclusion

Intensity normalization is a difficult task, however the methods HM_ballscale, STI, and MIMECS provided the best results for our study. MIMECS creates the most similar intensity distributions; for longitudinal and cross-sectional purposes. This method is recommended when the pathologies present are known and computational time is not an issue. HM_ballscale is the best semi-automatic method, had the best effect on the kNN segmentation, and yields the most similar intensity distributions in the interscanner reproducibility. It is recommended when image processing tasks are based on the intensity distributions and a large volume of NAWM is available in the scan. STI is recommended in other cases.

It is still unclear whether other (non-intensity based) image processing tasks benefit from the intensity normalization. Concerning this part, further research is recommended.

References

[1] WM Wells, WEL Grimson, R Kikinis, and FA Jolesz. Adaptive Segmentation of MRI Data. *IEEE transactions on medical imaging*, 15(4):429–442, 1996.

[2] Nicolas Robitaille, Abderazzak Mouiha, Burt Crépeault, Fernando Valdivia, Simon Duchesne, and The Alzheimer's Disease Neuroimaging Initiative. Tissue-based MRI intensity standardization: application to multicentric datasets. *International journal of biomedical imaging*, January 2012.

[3] Anant Madabhushi and Jayaram K. Udupa. New methods of MR image intensity standardization via generalized scale. *Medical Physics*, 33(9):3426, 2006.

[4] LG Nyúl, Jayaram K Udupa, and Xuan Zhang. New variants of a method of MRI scale standardization. *IEEE transactions on medical imaging*, 19(2):143–150, 2000.

[5] Snehashis Roy, Aaron Carass, and Jerry Prince. Patch based Intensity Normalization of brain MR images. *IEEE international symposium on biomedical imaging*, 2013.

[6] Xiao Han and Bruce Fischl. Atlas renormalization for improved brain MR image segmentation across scanner platforms. *IEEE transactions on medical imaging*, 26(4):479–86, April 2007.

[7] Benjamin M Ellingson, Taryar Zaw, Timothy F Cloughesy, Kourosh M Naeini, Shadi Lalezari, Sandy Mong, Albert Lai, Phioanh L Nghiemphu, and Whitney B Pope. Comparison between intensity normalization techniques for dynamic susceptibility contrast (DSC)-MRI estimates of cerebral blood volume (CBV) in human gliomas. *Journal of magnetic resonance imaging : JMRI*, 35(6):1472–7, June 2012.

[8] James D. Christensen. Normalization of brain magnetic resonance images using histogram even-order derivative analysis. *Magnetic Resonance Imaging*, 21(7):817–820, September 2003.

[9] Mohak Shah, Yiming Xiao, Nagesh Subbanna, Simon Francis, Douglas L Arnold, D Louis Collins, and Tal Arbel. Evaluating intensity normalization on MRIs of human brain with multiple sclerosis. *Medical image analysis*, 15(2):267–82, April 2011.

[10] Bruce Fischl, David H. Salat, Evelina Busa, Marilyn Albert, Megan Dieterich, Christian Haselgrove, Andre Van Der Kouwe, Ron Killiany, David Kennedy, Shuna Klaveness, Albert Montillo, Nikos Makris, Bruce Rosen, and Anders M. Dale. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355, 2002.

[11] Anant Madabhushi, Jayaram K. Udupa, and Andre Souza. Generalized scale: Theory, algorithms, and application to image inhomogeneity correction. *Computer Vision and Image Understanding*, 101(2):100–121, February 2006.

[12] Snehashis Roy, Aaron Carass, and Jerry Prince. A Compressed sensing approach for MR Tissue Contrast Synthesis. *Inf Process Med Imaging*, (22):371–383, 2011.

[13] J. Ashburner. Spm12, 2011. http://www.fil.ion.ucl.ac.uk/spm/software/spm12/.

[14] Julian Maclaren, Zhaoying Han, Sjoerd B Vos, Nancy Fischbein, and Roland Bammer. Reliability of brain volume measurements : A test-retest dataset. pages 1–9, 2014.

[15] Mirjam I. Geerlings, Auke P.a. Appelman, Koen L. Vincken, Ale Algra, Theo D. Witkamp, Willem P.T.M. Mali, and Yolanda van der Graaf. Brain volumes and cerebrovascular lesions on MRI in patients with atherosclerotic disease. The SMART-MR study. *Atherosclerosis*, 210(1):130–136, 2010.

[16] Hugo J. Kuijf, Manon Brundel, Jeroen de Bresser, Susanne J. van Veluw, Sophie M. Heringa, Max a. Viergever, Geert Jan Biessels, and Koen L. Vincken. Semi-Automated Detection of Cerebral Microbleeds on 3.0 T MR Images. *PLoS ONE*, 8(6), 2013.

[17] S Kullback and R A Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[18] K Van Leemput, F Maes, D Vandermeulen, a Colchester, and P Suetens. Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE transactions on medical imaging*, 20(8):677–688, 2001.

[19] S. Klein, M. Staring, K. Murphy, M.a. Viergever, and J. Pluim. elastix: A Toolbox for Intensity-Based Medical Image Registration. *IEEE Transactions on Medical Imaging*, 29(1):196–205, 2010.

[20] Image Science Institute. Mrbrains challenge, 2013. http://mrbrains13.isi.uu.nl/, Accessed on 20-Apr-2015.

[21] Bruce Fischl. FreeSurfer. *NeuroImage*, 62:774–781, 2012.

[22] L G Nyúl and Jayaram K Udupa. On Standardizing the MR Image Intensity Scale. *Magnetic Resonance in Medicine*, 42:1072–1081, 1999.

8 Appendices

8.1 Intensity distributions for GM, CSF, and WML

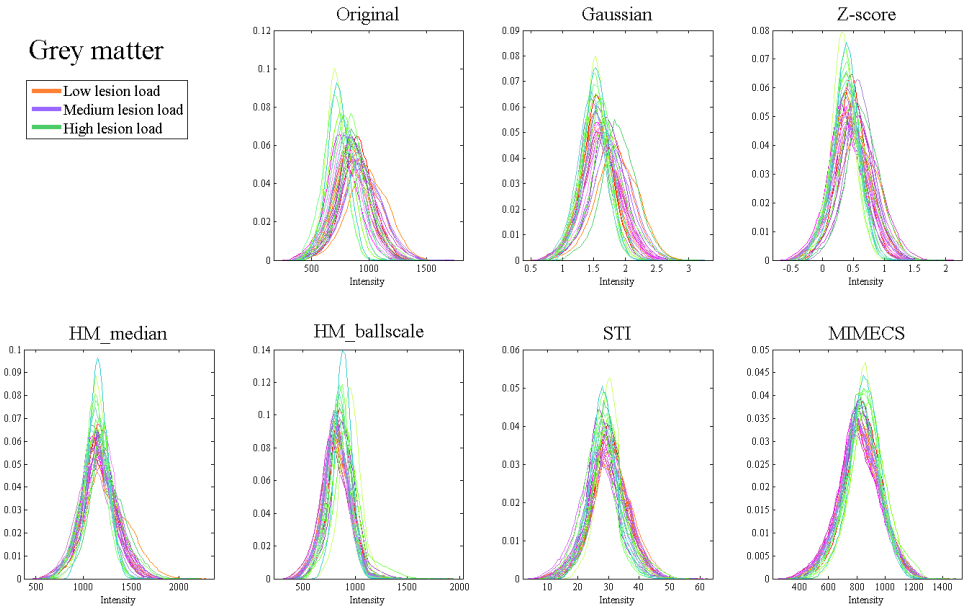

Figure 13: **Intensity distributions of the GM of 30 subjects. The three lesion load groups can be recognized by the colour of the histogram.**

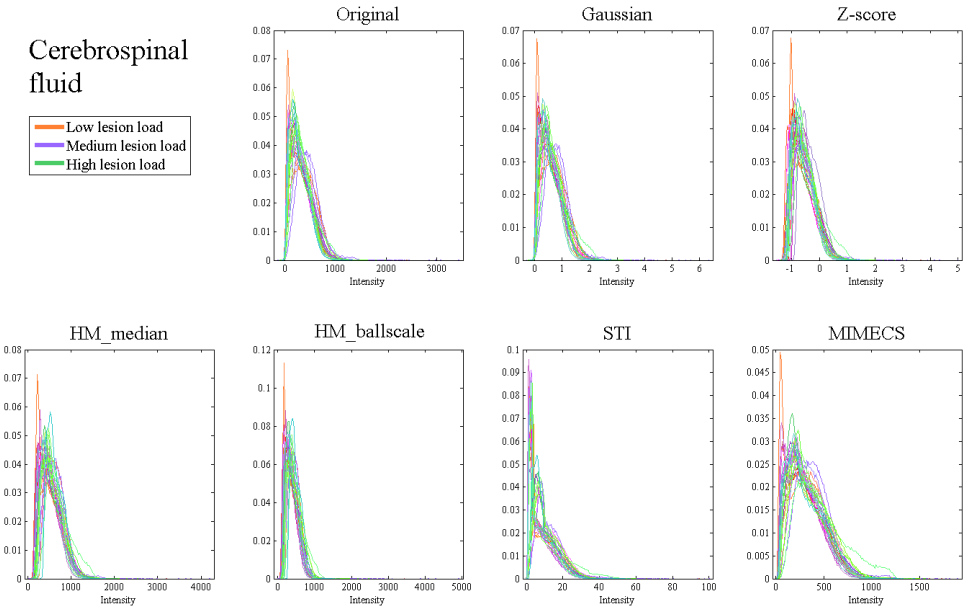

Figure 14: **Intensity distributions of the CSF of 30 subjects. The three lesion load groups can be recognized by the colour of the histogram.**

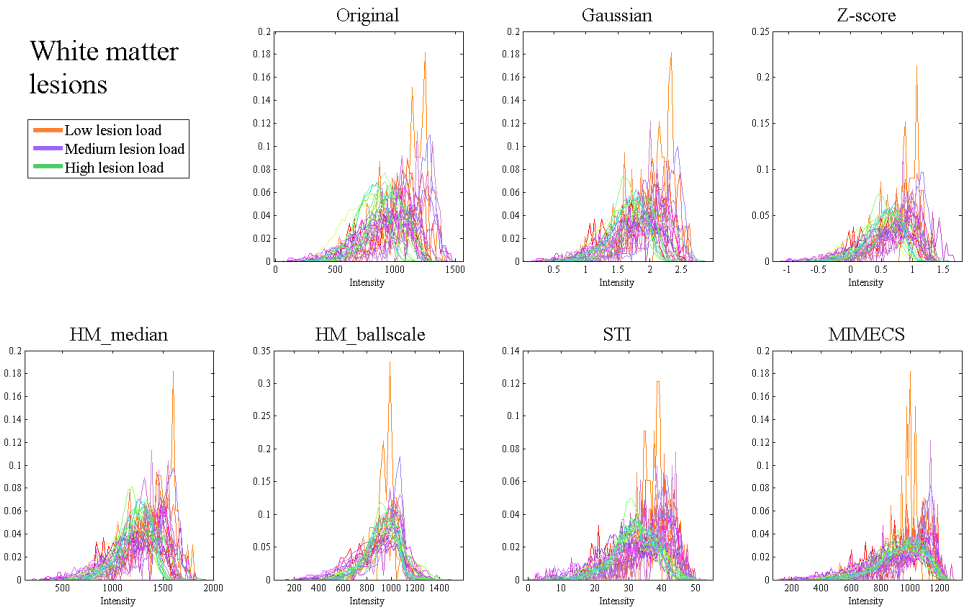

Figure 15: **Intensity distributions of the WML of 30 subjects. The three lesion load groups can be recognized by the colour of the histogram.**

8.2 KL divergences between different tissue intensity distributions.

Table 5: **Average Kullback-Leibler divergences between tissue types. A higher value suggests a larger difference between the tissue intensity distributions and is therefore better. A bold printed value indicates a difference from the original.**

KL divergence	WM - GM	WM - WML	GM - CSF	GM - WML
Original	13.5	7.9	13.1	2.6
Gaussian	13.5	7.9	13.1	2.6
Z-score	13.5	7.9	**13.2**	2.6
HM_median	13.5	7.9	13.1	2.6
HM_ballscale	**13.6**	7.9	**13.3**	**2.7**
STI	**13.1**	**7.7**	13.1	**2.4**
MIMECS	13.5	7.9	13.1	2.6

8.3 Normalization methods

A more extended explanation of the six intensity normalization methods applied in this study is given in this part. An overview of the properties of the methods is also given at the end.

Gaussian [7] This method, referred to as Gaussian, is the simplest and the fastest method. It rescales the intensities by a global linear scaling using $I_{new} = I/SD$. Where I is the intensity and SD is the standard deviation of the whole scan. The intensities are scaled based on the assumption that each scan has the same intensity distribution [7]. The presence of pathologies can alter the intensity distribution and the consequence is a shift in the normalization. The computational time is approximately 2 seconds.

Z-score or zero mean, unit variance [7] This method, referred to as Z-score, is also known as zero mean unit variance. It rescales the intensities by a global linear shift and scaling, using $I_{new} = (I - \mu)/SD$. Where I is the intensity, μ is the mean intensity and SD is the standard deviation of the whole scan. The intensities are scaled and shifted based on the assumption that each scan has the same intensity distribution. All means are shifted to zero, but the means do not automatically correspond to the same tissue type in different scans, especially when a large amount of WML is present. The computational time is approximately 2 seconds.

Histogram matching on median [4, 22] This method, referred to as HM_median, was first described by Nyúl and Udupa (2000) [4]. This two-stage method consists of a training step and a transformation step.

Transformation
Take the overall mean intensity of the scan and threshold the scan by this value to get the foreground. Take the median of the foreground as landmark μ_{50}. Get the 0^{th} and 99.8^{th} percentile intensity of the scan (p_1 and p_2 respectively). Map the intensity values piece wise linearly with the obtained landmarks (i.e.p_1, p_2 and μ_{50}) to the standard landmarks (s_1, s_2, and s_{50}, respectively). See Figure 16. The standard landmarks are obtained during the training step.

Training
In this step the histogram is rescaled to the pre-set landmarks s_1 and s_2. These values are chosen by the user in such a way that the histogram is not compressed after the normalization, so the s_1 and s_2 should be larger than the expected minimum and maximum intensity of the subject histogram, respectively. These values of s_1 and s_2 depend on the MRI sequence and brand of the MRI scanner, since the intensity ranges can differ between them. The intensity values of p_1 and p_2 are determined and the rescaling is applied by mapping [p_1, p_2] to [s_1, s_2] linearly. Next, the mean of the intensities is set as threshold and the median of the foreground is taken as landmark. This procedure is repeated for several times and the rounded mean of the determined median landmarks is taken as the standard median landmark s_{50}.

This method is semi-robust to WML. The computational time for the transformation step is approximately 6 seconds.

Generalized ball scale histogram matching [3, 11] This method was first described by Madabhushi and Udupa (2006) [3] and is a variation of the HM_median method. The difference is the way to retrieve the WM peak/landmark. For each voxel a radius of a ball is determined. This radius is growing, as long as 83 percent of the outer voxels satisfy a predefined homogeneity criterion. The homogeneity criterion is determined by the manually selected voxels (the value is the mean difference between adjacent voxels in the selected region plus three times the standard

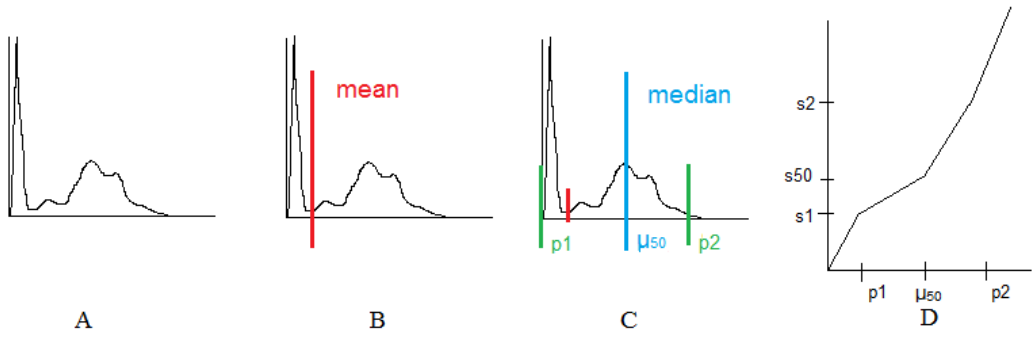

Figure 16: **Transformation step of HM_median. A) Compute intensity histogram of scan. B) Set the mean intensity as threshold for the foreground. C) Get the median intensity μ_{50}, p_1, and p_2 as landmarks. D) Linear piece-wise mapping to standard landmarks.**

deviation of this difference). The 83 percent is the adviced threshold.

The result is an image in which the voxel values represent the largest radius for that voxel, see Figure 17. This image is thresholded by 2 voxels in this study. The largest connected volume of this thresholded image is considered NAWM in brain scans. The median intensity of this volume is taken as a landmark. The other two landmarks are again the 0^th percentile and the 99.8^th percentile of the entire intensity histogram. The intensities are piece-wise linear mapped to the standard scale landmarks. The standard landmarks are obtained during the training step with the use of multiple scans. This training part is the same as for the HM_median method, described above.

In the method described the whole scan is included, but due to an otherwise long computational time, only the middle slice is used in this study. When large areas of WMLs are present in the brain, the algorithm might not detect the NAWM, because this is no longer the largest connected area. The computational time is approximately 80 seconds, including the manual selection of NAWM.

STI [2] This method is called Standardization of Intensities (STI), and was proposed by Robitaille et al. (2011)[2]. The subject scan is nonlinear registered to a reference scan. The intensities of the reference scan are rescaled to intensities between 0 and 100. The masks of WM, GM and CSF of the reference scan are applied to both the reference and the subject scan. A joint histogram is created for each tissue separately and this is filtered with a Gaussian filter. The mode of the joint histogram is selected as landmark. The minimum intensity and the maximum intensity of the subject scan are set to 0 and 100 respectively and are also used as landmarks. The transformation is a linear piece-wise mapping between the five landmarks. See Figure 18.

A proper reference scan has a similar intensity distribution as the subject scan. This reference scan needs to be carefully selected. STI is a WML robust method and can handle atrophy. Unfortunately it yields loss of intensity information. The computational time is approximately 80 seconds.

MIMECS [5] This method was developed by Roy, Carass, and Prince (2013) and the software is called MIMECS [5, 12]. The intensities of the subject scan are normalized to the intensities

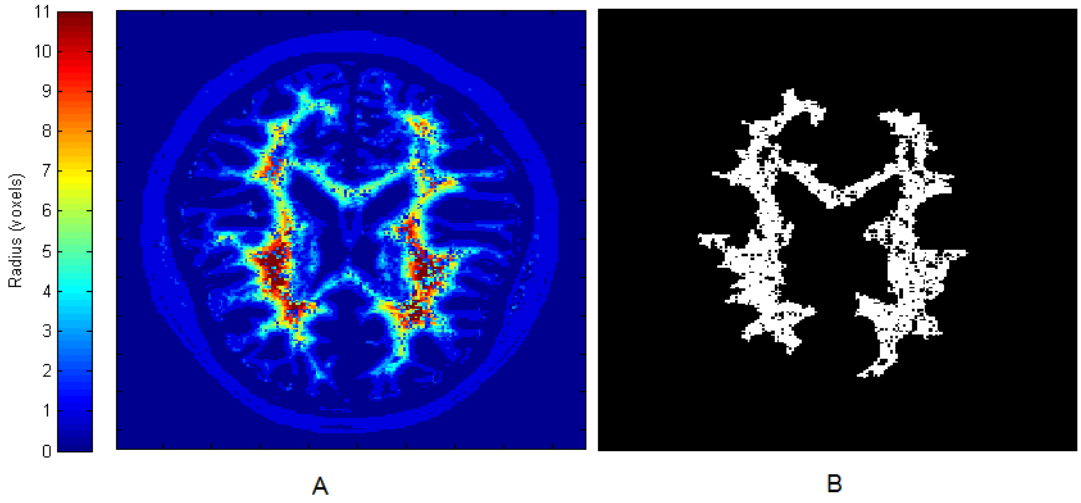

Figure 17: **Steps in HM_ballscale method. A) For all voxels the largest ball radii for which the homogeneity criterion still holds is given. B) Largest connected volume after thresholding at a radius of 2 voxels.**

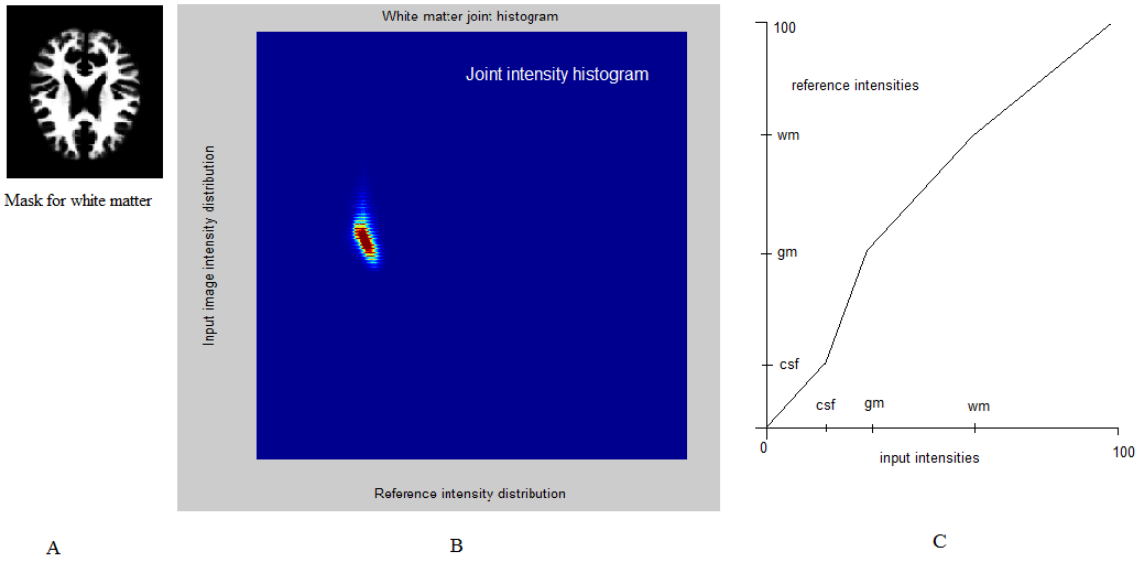

Figure 18: **Steps for STI. A) WM masks from reference to apply on the subject and reference scan. B) Gaussian filtered joint histogram for WM. C) Linear piece-wise mapping between the five landmarks.**

of a reference scan with the use of $3 \times 3 \times 3$ voxel patches. At first a similar intensity range is needed for a better matching of voxels, so both scans are normalized by the WM peak. This WM peak is predetermined by the user. This can be done fast, accurate and automatically, when tissue masks of the reference and of the subject scan are available. The masks are applied to the scans. Then all the voxels of WM can be obtained and the median intensity of these voxels is the WM peak. When the scan is skull stripped, this WM peak does not have to be given to the method, since it determines the WM peak automatically.

Next for each voxel a surrounding patch is defined. The best matching reference patch for each subject patch is defined by the maximum likelihood and found using the expectation max-

imization algorithm [5]. The central voxel intensity of the subject is replaced with the matching central voxel intensity of the reference.

The reference scan should contain all the pathologies that are expected in the subject scan for a good intensity normalization. The reference scan should therefore be carefully chosen. The computational time is approximately 4 hours, which makes it a very time consuming method.

Table 6: **Overview of properties of intensity normalization methods described in this study.**

	Robust to WML	**Reference needed**	**Computational time**	**Fully automatic**	**Pre-processing steps needed**	**Assumption of similar intensity distribution**
Gaussian	No	No	2 seconds	Yes	No	Yes
Z-score	No	No	2 seconds	Yes	No	Yes
HM_median	Partly	Training set needed	6 seconds	Yes, during the transformation step	No	Yes
HM_ballscale	Partly	Training set needed	80 seconds	No, manual selection of NAWM	No	No
STI	Yes	Yes, with a comparable intensity distribution and no pathologies	80 seconds	Yes	Yes, registration	Yes
MIMECS	Yes, when pathologies are also present in reference.	Yes, the expected pathologies in the subject, need to be present in the reference scan.	4 hours	Yes, when the WM peak selection is automatic.	No	No