

# 3D MRI Denoising Using Rough Set Theory and Kernel Embedding Method

Ashish Phophalia<sup>(✉)</sup> and Suman K. Mitra

Dhirubhai Ambani Institute of Information and Communication Technology,  
Gandhinagar 382007, Gujarat, India  
{ashish\_phophalia, suman\_mitra}@daiict.ac.in

**Abstract.** In this paper, we have presented a manifold embedding based method for denoising volumetric MRI data. The proposed method via kernel mapping tries to find linearity among data in the projection/feature space. Prior to kernel mapping, a Rough Set Theory (RST) based clustering technique has been used with extension to volumetric data. RST clustering method groups similar voxels (3D cubes) using class and edge information. The basis vector representation of each cluster is then explored in the Kernel space via Principal Component Analysis (known as KPCA). The work has been compared with state-of-the-art methods under various measures for synthetic and real databases.

**Keywords:** Image denoising · Magnetic resonance imaging · Rough set theory

## 1 Introduction

Being a non-invasive technique, Magnetic Resonance Imaging (MRI) is widely used modality (along with X-Ray, CT, etc.) for clinical diagnosis. The acquisition process of medical images is highly sensitive to get accumulated with noise or undesired signals. It has been shown that the noise in Magnetic Resonance (MR) Image is Rician in nature [7]. It has been shown that the intensities of MR images are magnitude of underlying complex data following Rice distribution [6, 7]. The real and imaginary parts are modeled as being independently distributed Gaussian with means  $a_r$  and  $a_i$  respectively, with same variance  $\sigma^2$ . The rician random variable  $y$  with PDF can be defined as follows:

$f_Y(y|a, \sigma) = \frac{y}{\sigma^2} e^{\left(-\frac{y^2+a^2}{2\sigma^2}\right)} I_0\left(\frac{ya}{\sigma^2}\right), y > 0$ , where  $a = \sqrt{a_r^2 + a_i^2}$  is underlying noise free signal amplitude and  $I_n(z)$  is  $n^{\text{th}}$  order modified bessel function of first kind. Let  $SNR$  be the signal to noise ratio (here, it is  $a/\sigma$ ). When  $SNR$  is high, the Rician distribution approaches to Gaussian; when  $SNR$  approaches to zero (that is only noise is present,  $a \rightarrow 0$ ) the Rician distribution becomes Rayleigh distribution and the PDF becomes  $f_Y(y|a \rightarrow 0, \sigma) = \frac{y}{\sigma^2} e^{\left(-\frac{y^2}{2\sigma^2}\right)}$ .

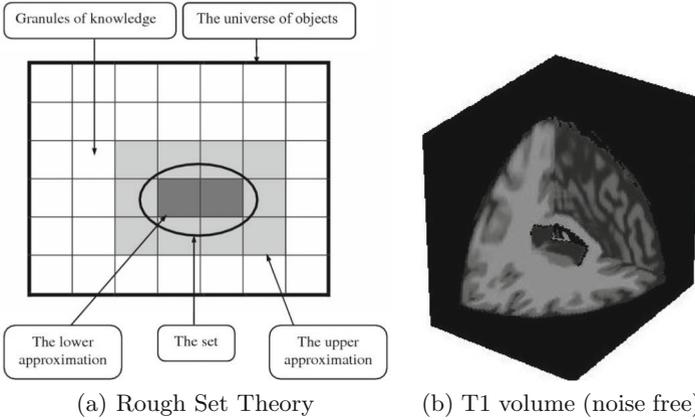
Many state-of-the-art methods for 3D image denoising are extension of their 2D counterpart version. However, computational complexity becomes a crucial

factor during the extension. The 3D MR image denoising was introduced in modern literature in [5] and then followed by [4, 8–10, 14, 16] etc. Instead of playing with patches in 2D images, here a voxel is defined as 3D cube centered at location  $(i, j, k)$  in  $\mathcal{R}^3$ . Hence, a voxel is simply counterpart of a patch with size  $w \times w \times w$ . Consequently, exploring relationship among intensity values in  $\mathcal{R}^3$  is highly sought and thus leads to computational intensive process. There have been many denoising/restoration methods proposed for Medical Images ranging from diffusion filters to dictionary and clustering based filters.

The Non Local Means (NLM) method [1] has been extended to Optimized Blockwise NLM (OBNLM, [5]) for volumetric data. It tweaks the computation of similarity weight between voxels and constrained with predefined criteria for mean and variance of both the voxels. At the same time, it adopted the block-wise strategy to drop restoration of adjacent  $n$  voxels which effectively reduce the computation load by  $n^3$  times instead of processing each voxel in the image space. The same has been adopted in [4, 8]. In case of ABONLM method [4], denoising is performed under Wavelet framework using adaptive soft wavelet coefficient mixing (ASCM) approach. A non-parametric kernel regression framework has been adopted for 3D MR image denoising in Unbiased Kernel Regression (UKR) method [14]. UKR is rooted on a zeroth order 3D kernel regression and similarity weight between voxels is derived on small sized feature vectors based image intensity and gradient information. The sparseness and self-similarity has been unified in PRI-NLM method [10]. It incorporates rotational invariant NLM and discrete cosine transform hard thresholding for sparsity.

The well-known BM3D method has been extended to 3D MR image denoising as BM4D method [8]. Similar to BM3D, BM4D is also equipped with collaborative filtering notion where similar voxels are arranged in fourth dimension. It is also a two stage method where the output of first stage guides the second stage and uses hard thresholding in first pass and wiener filtering in second pass. Ideally, BM4D is designed for Gaussian noise whereas Variance Stabilization Technique (VST) has been adopted to deal with Rician noise in 3D MRI data. The VST based scheme is also adopted in two phased HOSVD-R method [16]. Recently, another two stage method PRI-NL-PCA [9] was proposed based on sparsity and self-similarity of voxels. It is encompassed with PCA thresholding strategy in first stage where rotational invariant NLM method is deployed in second stage. In this paper, we present manifold embedding based method for MR image denoising. A nonparametric variant of PCA, known as Kernel Principal Component Analysis (KPCA) has been explored for Rician noise removal. The KPCA tries to explore the structure in the data in Feature space instead of data space itself and tries to capture higher-order dependencies in the data.

This paper is organized as follows: Sect. 2 discusses RST based clustering approach followed by overview of Kernel Principal Component Analysis. The proposed method in presented last in Sect. 2. Section 3 presents simulation results of proposed method along with state-of-the-art-method on phantom and real MRI data. The manuscript is concluded in Sect. 4.



**Fig. 1.** (a) Image granules with upper and lower approximation of an object as conceptualized in Rough Set Theory and (b) Noise free T1 volume data from BrianWeb Database [3].

## 2 Material and Methods

### 2.1 Rough Set Based Cluster Formation

Rough Set Theory (RST) can be utilized to explore structural similarity between pixel or set of pixels even in the presence of noise. This classification is defined on predefined attribute(s),  $\Theta$ , by forming granules within the image space,  $\Omega$ . This indeed establishes an equivalence classes among the data based on attributes. Rough sets define a class by approximating two sets, namely lower approximation and upper approximation sets of a class with respect to an attribute(s). The lower approximation set of class w.r.t. attribute(s),  $\underline{C}_\Theta$ , consists of certainly classified granules of  $\Omega$  whereas the upper approximation set of class with respect to attribute(s),  $\overline{C}_\Theta$ , constructed by possible granules of that class defined by attribute(s),  $\Theta$  [12]. The figurative description is given in Fig. 1.

Rough set based derivation of class label (RCL) information and edge details (REM) have been derived in [13]. For given attribute(s), granules can be classified in either lower or upper approximation of an object. The attribute considered is the intensity values at each location in the image space. The objects present in the image are categorized in intensity range by optimizing image histogram. The Rough set based entropy criteria [12] was used in optimizing intensity thresholds. The class label can be assigned by comparing intensity value at each location against the intensity ranges of all the objects. A granule (set of adjacent pixels) is assigned to an object's lower if all the pixels fall in its intensity range. Otherwise it will assign in object's upper approximation only if any pixel in that granule belongs to intensity range. The difference of both the approximations of any objects will generate pixels which are possible edges of the object in the image. Thus, union of all such edges will generate edge map of the image.

We have extended method proposed in [13] to 3D imaging. A voxel is converted to vector representation. We use the same approach for deriving lower and upper approximation of all the objects. For  $K$  number of objects present in the image, the number of pool constructed would be  $\sum_{i=1}^K \binom{K}{i}$ . There will be  $K$  clusters corresponding to  $K$  lower approximations of each object and rest are union approximations of combinations of the objects.

## 2.2 Kernel Principal Component Analysis

In KPCA, the non-linearity is introduced by first mapping the data into another space  $F$  using a nonlinear map  $\Phi : R^N \rightarrow F$ , before a standard linear PCA is carried out in  $F$  using the mapped samples  $\phi(x_k)$ . The map  $\Phi$  and the space  $F$  are determined implicitly by the choice of a kernel function  $k$ , which acts as a similarity measure. This mapping computes the dot product between two input samples  $x$  and  $y$  mapped into  $F$  via

$$k(x; y) = \Phi(x) \cdot \Phi(y) \quad (1)$$

One can show that if  $k$  is a positive definite kernel, then there exists a map  $\Phi$  into a dot product space  $F$  such that Eq. 1 holds. The space  $F$  then has the structure of a Reproducing Kernel Hilbert Space (RKHS) [2]. Equation 1 is important for KPCA since PCA in  $F$  can be formulated entirely in terms of inner products of the mapped samples. This has two important consequences: first, inner products in  $F$  can be evaluated without computing  $\Phi(x)$  explicitly. This allows to work with a very high-dimensional, possibly infinite-dimensional RKHS  $F$ . Second, if a positive definite kernel function is specified, we need to know neither  $\Phi$  nor  $F$  explicitly to perform KPCA since only inner products are used in the computations.

In PCA, the covariance matrix is defined as  $C = \frac{1}{N-1} X^t X$  where is  $X$  is called data matrix containing samples in columns. The covariance matrix in case of KPCA of size  $M \times M$ , calculated by  $C_F = \frac{1}{N} \sum_{i=1}^N \phi(\mathbf{x}_i) \phi(\mathbf{x}_i)^T$ . Its eigenvalues and eigenvectors are given by  $C_F \mathbf{v}_k = \lambda_k \mathbf{v}_k$ , where  $k = 1, 2, \dots, M$ . Mathematical simplification leads to  $\mathbf{v}_k = \sum_{i=1}^N a_{ki} \phi(x_i)$  and hence  $\mathbf{a}_k$  ( $N$ -dimensional column vector of  $a_{ki}$ ) can be solved by  $C_F \mathbf{a}_k = \lambda_k N \mathbf{a}_k$ . If projected dataset  $\phi(\mathbf{x}_i)$  does not have zero mean, one can use Gram matrix  $\tilde{C}_F$  to substitute the kernel matrix  $C_F$  which is given by  $\tilde{C}_F = C_F - \mathbf{1}_N C_F - C_F \mathbf{1}_N + \mathbf{1}_N C_F \mathbf{1}_N$ , where  $\mathbf{1}_N$  is the  $N \times N$  matrix with all elements equal to  $1/N$ .

## 2.3 Proposed Method

MRI data is corrupted with rician noise which is not additive in nature, hence, it is expected in this work that transformation of data into high dimensional space may rise to linearity of data. This work cascades clustering method and manifold embedding method. We derive clusters of similar voxels based on classes present in those. Then KPCA finds the linearity in the higher dimensional space. To denoising voxels, each voxel is projected on the corresponding class basis vectors

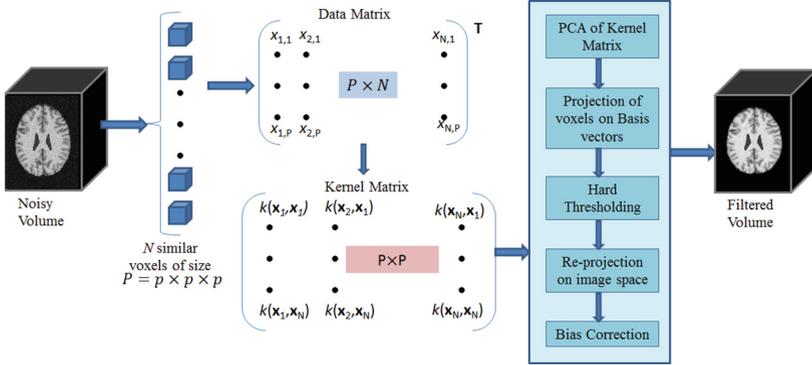


Fig. 2. Flowchart of proposed method

and coefficients below threshold are truncated assumed as noise part in the data. The voxels are reprojected in the image space via invertible basis vectors and bias correction is performed. Figure 2 shows flow of proposed method. The outline of present work can be described as follows:

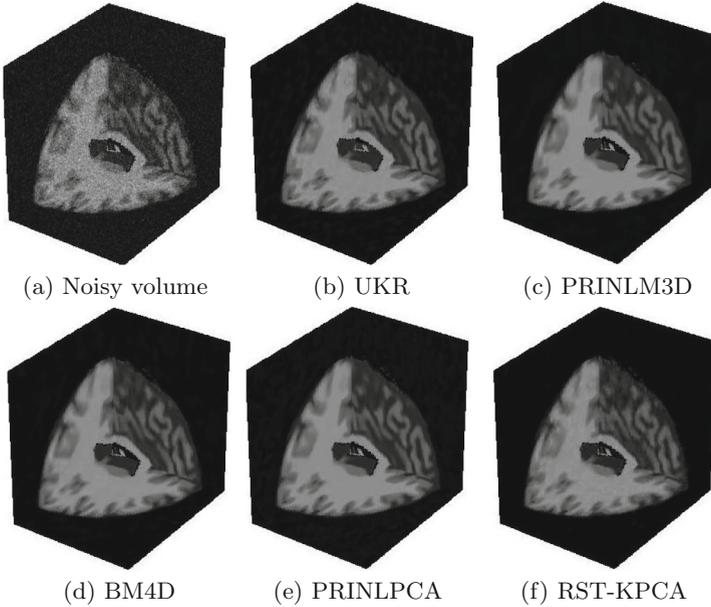
1. Get the clusters of voxels ( $p \times p \times p$ ) from the given noisy image using Rough set based method (as described in [13]).
2. For each cluster, get the basis vectors using KPCA method along pixel positions. For cluster matrix of size  $p^3 \times N$ , kernel matrix would be of size  $p^3 \times p^3$ , where  $N$  is number of voxels in the cluster.
3. Project the noisy image voxels on the obtained basis vectors in the KPCA domain. Apply coefficient shrinkage method on these projected voxels to get the denoised voxels. Transform them back to image space.
4. Remove the bias term from each pixel of the denoised image i.e.  $\hat{I}_{unbiased}(i, j, k) = \sqrt{\max(\hat{I}(i, j, k)^2 - 2h^2, 0)}$ , where  $h$  is the standard deviation of noise and  $\hat{I}$  is the image obtained by step (4).

### 3 Experimental Section

This Section encompasses the qualitative and quantitative evaluation of the proposed method along with some of the state-of-the-art methods. The experiments have been carried out on 3D monochrome phantom human brain MRI images obtained from Brain Web Database [3]. The parameters are as follows: Modality = T1, RF = 0, protocol = ICBM, slice thickness = 1 mm, volume size =  $181 \times 217 \times 181$  (shown in Fig. 1b). The skull portion have been from the volume and considered four classes: (a) White Matter, (b) Gray matter, (c) Cerebrospinal Fluid and (d) Background. The evaluation measures used are Peak-Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), Structural

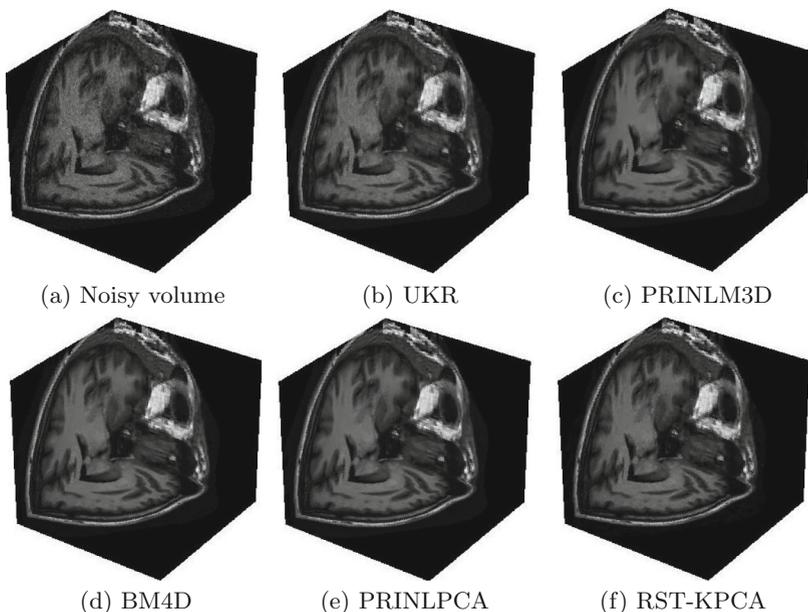
**Table 1.** Results of state-of-the-art Methods for T1 modality (represented row-wise against each method). In each  $2 \times 2$  block, top-left figure is PSNR, top-right is RMSE, bottom-left is SSIM and bottom-right is BC measure. The best figure against each noise level is represented in Bold face.

$h \rightarrow$	5		10		15		20		25	
Noise	31.78	06.57	25.76	13.13	22.25	19.69	19.75	26.24	17.82	32.77
	0.3473	0.3497	0.2170	0.2910	0.1583	0.2859	0.1209	0.3033	0.0949	0.3207
UKR	36.04	04.02	30.59	07.53	33.72	05.25	32.11	06.32	32.85	05.81
	0.5084	0.5373	0.3359	0.2760	0.6507	0.7809	0.6364	0.8210	0.7762	0.8902
PRI-NLM	<b>45.16</b>	<b>01.41</b>	<b>39.95</b>	<b>02.56</b>	<b>36.91</b>	<b>03.64</b>	<b>34.84</b>	<b>04.62</b>	<b>33.27</b>	<b>05.53</b>
	0.9527	0.9009	0.8769	0.8291	0.8129	0.8173	0.7734	0.8276	0.7297	0.8225
BM4D	43.81	06.57	39.17	02.81	36.50	03.81	34.61	04.74	33.14	05.62
	0.9452	0.9008	0.8729	0.8690	0.8167	0.8597	0.7735	0.8549	0.7390	0.8518
PRI-NL-PCA	40.36	02.45	35.01	04.53	32.0	06.41	29.90	08.16	28.32	09.78
	0.6640	0.4699	0.4918	0.5332	0.4243	0.5705	0.3864	0.5930	0.3662	0.6173
RST-KPCA	44.09	01.59	39.45	02.72	36.83	03.67	32.93	05.76	32.11	06.33
	<b>0.9904</b>	<b>0.9818</b>	<b>0.9770</b>	<b>0.9760</b>	<b>0.9535</b>	<b>0.9641</b>	<b>0.9446</b>	<b>0.9505</b>	<b>0.9391</b>	<b>0.9446</b>



**Fig. 3.** Comparison of various methods on T1 from BrainWeb Database

Similarity Index (SSIM) [15] and Bhattacharya Coefficient (BC). The methods used for comparisons are: (a) UKR [14], (b) PRI-NLM [10], (c) BM4D [8], (d) PRI-NL-PCA [9]. In all the experiment, voxels of size  $p = 3$  are considered. We have used *simple/linear kernel* only i.e.  $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^T \cdot \mathbf{y}$ . Table 1 shows quantitative results over varying level of Rician noise. It can be observed that our methods outperforms state-of-the-art methods in terms of SSIM and BC



**Fig. 4.** Comparison of various methods on subject 018 from OASIS Database

measures. However, our method is behind PRI-NLM method and better than BM4D (at lower noise levels) in terms of PSNR and RMSE measures. Figure 3 shows cross sectional views of denoised volumes from all methods. The proposed method with single threaded MATLAB implementation takes around 45 min on core i7 processor, 2.10 GHz and 8 GB RAM machine. The UKR is observed to have same computational time whereas BM4D with MATLAB/C implementation takes 11 min and others run in less than five minutes.

The subject details from OASIS dataset [11] are as follows: Subject ID: 018, Age: 39 (male) respectively, scan number: mpr-1, type: MPRAGE, voxel resolution:  $1.0\text{ mm} \times 1.0\text{ mm} \times 1.25\text{ mm}$ , Orientation: Sagittal, TR (ms) = 9.7, TE (ms) = 4.0, TI (ms) = 20.0, Flip angle = 10. The results are shown in Fig. 4 for subject 018 as cross sectional view.

## 4 Conclusion

Kernel method is explored in this work to deal with Rician noise present in the MRI data. Being signal dependent noise, applicability of linear denoising operation such as PCA is not advisable. It is expected in the present work that kernel method may project the nonlinear data in the linear feature space. We have an extended Rough Set based clustering method to collect similar voxels conditioned on class and edge information. These similar voxels are then used to define kernel matrix via kernel function. However, it can be exercised with

other known kernels with suitable parameter estimation method or data adaptive kernel for rician model can be thought of.

The proposed method is non-iterative and single stage method in comparison to some of predecessor methods like BM4D, PRI-NL-PCA etc. In this work, intensity values are used as feature in clustering step and in kernel space however more features can be considered like gradient information as in UKR. The predecessor methods restrict the search space for searching similar voxels. However, current method exploits the whole volume and thereby form clusters of similar voxels.

## References

1. Buades, A., Coll, B., Morel, J.M.: A non local algorithm for image denoising. In: IEEE Computer Vision and Pattern Recognition, pp. 60–65 (2005)
2. Charpiat, G., Hofmann, M., Schölkopf, B., et al.: Kernel methods in medical imaging. In: Handbook of Biomedical Imaging (2010)
3. Collins, D.L., Zijdenbos, A., Kollokian, V., Sled, J., Kabani, N., Holmes, C., Evans, A.: Design and construction of a realistic digital brain phantom. *IEEE Trans. Med. Imaging* **17**(3), 463–468 (1998)
4. Coupé, P., Manjón, J.V., Robles, M., Collins, L.D., et al.: Adaptive multiresolution non-local means filter for 3d mr image denoising. *IET Image Process.* **6**, 558–568 (2011)
5. Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C.: An optimized blockwise non local means denoising filter for 3d magnetic resonance images. *IEEE Trans. Med. Imaging* **27**(4), 425–441 (2008)
6. Foi, A.: Noise estimation, removal in mr imaging: the variance-stabilization approach. In: ISBI, pp. 1809–1814 (2011)
7. Gudbjartsson, H., Patz, S.: The rician distribution of noisy mri data. *Magn. Reson. Med.* **34**(6), 910–914 (1995)
8. Maggioni, M., Katkovnik, V., Egiazarian, K., Foi, A.: Nonlocal transform domain filter for volumetric data denoising and reconstruction. *IEEE Trans. Image Process.* **22**(1), 119–133 (2013)
9. Manjón, J.V., Coupé, P., Buades, A.: Mri noise estimation and denoising using non-local pca. *Med. Image Anal.* **22**(1), 35–47 (2015)
10. Manjón, J.V., Coupé, P., Buades, A., Louis Collins, D., Robles, M.: New methods for mri denoising based on sparseness and self-similarity. *Med. Image Anal.* **16**(1), 18–27 (2012)
11. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* **19**(9), 1498–1507 (2007)
12. Pal, S.K., Shankar, B.U., Mitra, P.: Granular computing, rough entropy and object recognition. *Pattern Recogn. Lett.* **26**, 2509–2517 (2005)
13. Phophalia, A., Rajwade, A., Mitra, S.K.: Rough set based image denoising for brain mr images. *Sig. Process.* **103**, 24–35 (2014)
14. Rubio, E.L., Nunez, M.N.F.: Kernel regression based feature extraction for 3d mr image denoising. *Med. Image Anal.* **15**, 498–513 (2011)

15. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
16. Zhang, X., Xu, Z., Jia, N., Yang, W., Feng, Q., Chen, W., Feng, Y.: Denoising of 3d magnetic resonance images by using higher-order singular value decomposition. *Med. Image Anal.* **19**(1), 75–86 (2015)