

Simultaneous PET Image Reconstruction and Feature Extraction Method using Non-negative, Smooth, and Sparse Matrix Factorization

Kazuya Kawai*, Hidekata Hontani*, Tatsuya Yokota*, Muneyuki Sakata† and Yuichi Kimura‡

* Nagoya Institute of Technology, Aichi, Japan,

E-mail: kawai@iu.nitech.ac.jp, {hontani,t.yokota}@nitech.ac.jp

† Tokyo Metropolitan Institute of Gerontology, Tokyo, Japan

E-mail: sakata@pet.tmig.or.jp

‡ Kindai University, Wakayama, Japan

E-mail: ukimura@ieee.org

Abstract—Positron emission tomography (PET) is an important imaging technique to visualize a number of functions in the brain or human body. For reconstructing PET images from the sinogram data, an inverse problem has to be solved using numerical optimizations such as expectation-maximization (EM)-based methods. However, the standard EM method suffers from measurement noise added in the sinogram data. In this paper, we propose a new simultaneous PET image reconstruction and parts extraction method using constrained non-negative matrix factorization. In contrast that the many existing methods reconstruct a single PET image independently, we reconstruct the time-series of PET images simultaneously from the time-series of sinograms using non-negative matrix factorization. Furthermore, we impose the smoothness constraint for the temporal feature, and the exclusive LASSO-based sparseness constraint for the spatial feature for robust image reconstruction and physically meaningful feature extraction.

I. INTRODUCTION

Positron emission tomography (PET) is a powerful tool in the biomedical analysis of brain functions, body functions, and drug effects [14], [15]. Some typical applications of PET imaging are the diagnosis of Alzheimer’s disease by visualizing the distributions of amyloid beta in the brain [5], and the diagnosis of cancers by visualizing glucose metabolism in the body [3].

PET images are obtained by the following procedures: First, radioactive ligand is injected to a human subject. The ligand is transmitted and distributed in into tissues, and retained by the receptors. Thus, the concentration of ligand is affected by the distributions of the receptors in the human body/brain. The PET scanner detects the gamma rays from the ligand and records them as sinograms. It is noteworthy that the sinograms do not directly show the distributions of the radioactive ligand in the body/brain, and that it can be shown only after the computational imaging. Finally, PET images are obtained from the sinograms using computational image reconstruction methods such as filtered back-projection [1], [13], [16], and the expectation-maximization (EM) algorithm [9], [17].

As sinogram data suffer from Poisson noise with a low signal-to-noise ratio (SNR), it directly affects the reconstructed

PET images. Thus, a robust reconstruction algorithm is important. The aim of this study is to propose a new PET image reconstruction algorithm that is tough on the noise in PET data. A typical approach to preventing noise is to introduce prior knowledge for observations or images.

Maximum likelihood estimation based on Poisson noise is the EM algorithm [9], [17] as mentioned above. Herein, we consider some additional priors/assumptions based on the compartment model of pharmacokinetic analysis [7], [8]. In other words, unlike conventional methods, we reconstruct the time-series of all PET images at once from the time-series of sinograms by modeling the smooth time curves of the ligand concentration. Furthermore, we introduce the spatial sparseness priors for the parts of PET images by assuming that the brain image can be separated into several anatomical parts. For introducing the parts priors, we consider the constrained non-negative matrix factorization (NMF) model [10], [18] to represent the PET images. Hence, the matrix of the PET images is modeled by the multiplication of two matrices: a non-negative and sparse spatial pattern matrix, and a non-negative and smooth temporal pattern matrix. We employ the exclusive LASSO [6] penalty for the spatial sparse pattern matrix, and the quadratic variation penalty for the temporal smooth pattern matrix. The optimization problem can be solved by a multiplicative update algorithm similar to the EM algorithm. The experimental results show the advantages of the proposed method for both objectives: noise-robust reconstruction and parts extraction.

II. THEORY AND METHOD

When a pharmaceutical is injected and atomic collapse occurs, radioactivity is emitted in two directions angled at 180° . This radiation is observed as a pair of events by two detectors of the PET scanner. The observed data accumulating the event pairs is called a sinogram.

Let V and B be numbers of voxels in the PET image and sinogram, respectively. We denote a PET image at a time frame f by $d_f \in \mathbb{R}^{V \times 1}$, and the sinogram measured by the

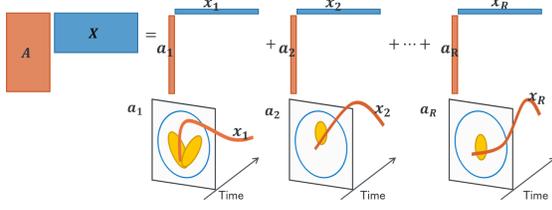


Fig. 1. Concept of NMF of PET image [4]

PET scanner by $\mathbf{y}_f \in \mathbb{R}^{B \times 1}$. From the physical model of the PET scanner, a sinogram is obtained by $\mathbf{y}_f = \mathbf{P}\mathbf{d}_f$, where $\mathbf{P} \in \mathbb{R}^{B \times V}$ is a projection matrix. Considering all time frames $f \in \{1, 2, \dots, F\}$, we have $\mathbf{Y} = \mathbf{P}\mathbf{D}$, where $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_F] \in \mathbb{R}^{B \times F}$ and $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_F] \in \mathbb{R}^{V \times F}$ are a sinogram matrix and a PET matrix, respectively. In contrast to classical existing methods that reconstruct \mathbf{d}_f from \mathbf{y}_f for each f , our approach reconstructs \mathbf{D} from \mathbf{Y} simultaneously.

A. Reconstruction model

The proposed PET image reconstruction model is given by

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{X}} D_{KL}(\mathbf{Y}, \mathbf{P}\mathbf{D}) + \alpha J_1(\mathbf{A}) + \beta J_2(\mathbf{X}) \quad (1) \\ \text{s.t. } \mathbf{D} = \mathbf{A}\mathbf{X}, \mathbf{A} \geq 0, \mathbf{X} \geq 0. \end{aligned}$$

where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_R] \in \mathbb{R}^{V \times R}$ is a basis matrix for constructing spatial patterns, and $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_R]^T \in \mathbb{R}^{R \times F}$ is a basis matrix for constructing time activity curves. From the compartment model of the dynamics of radioactive ligand [7], [8], the matrix of PET image \mathbf{D} can be assumed as low rank. Therefore, we consider the number of patterns R is smaller than $\min(V, F)$. Fig. 1 shows the concept of our matrix factorization in PET images. Each column of matrix \mathbf{A} represents a spatial pattern that correspond to anatomical regions exhibiting the same temporal pattern. Each row of matrix \mathbf{X} represents a temporal basis function. The matrix of PET image \mathbf{D} can be considered linear sums and decomposed into the R image bases.

In Problem (1), $D_{KL}(\cdot, \cdot)$ is the Kullback–Leibler divergence, as shown in the following formula:

$$\begin{aligned} D_{KL}(\mathbf{Y}, \mathbf{P}\mathbf{A}\mathbf{X}) \\ = \sum_{i,j} \left(Y_{ij} \log \frac{Y_{ij}}{(\mathbf{P}\mathbf{A}\mathbf{X})_{ij}} - Y_{ij} + (\mathbf{P}\mathbf{A}\mathbf{X})_{ij} \right). \quad (2) \end{aligned}$$

The functions $J_1(\mathbf{A})$ and $J_2(\mathbf{X})$ are the penalty terms used to enforce the constraints on the solution of (1), and α and β are their corresponding regularization parameters.

From the compartment model of time activity curves, we assume that the brain region can be separated into several compartments based on the clusters of the time activity curves. Further, we consider that each column vector of spatial matrix

\mathbf{A} can be separated into meaningful image patterns. Therefore, we introduce the following regularizer:

$$J_1(\mathbf{A}) = \sum_{v=1}^V \left(\sum_{r=1}^R |a_{vr}| \right)^2 \Big|_{\mathbf{A} \geq 0} = \|\mathbf{A}\mathbf{1}\|_2^2, \quad (3)$$

where a_{ij} is the i -th and j -th elements of matrix \mathbf{A} , and $\mathbf{1} \in \{1\}^R$ is an R -dimensional vector of ones. We introduce l_1 -norm to impose row of matrix \mathbf{A} on sparseness and l_2 -norm to the combined elements of the same brain regions. We refer to the regularizer above as *exclusive lasso* [6]. It is noteworthy that the exclusive LASSO penalty can be rewritten as the right hand side of Eq. (3) because of the non-negativity constraint of \mathbf{A} .

To enforce a smooth temporal pattern matrix \mathbf{X} , we consider the following penalty:

$$J_2(\mathbf{X}) = \sum_{r=1}^R \left(\sum_{f=1}^F (x_{rf} - x_{r(f+1)})^2 \right) = \|\mathbf{L}\mathbf{X}^T\|_F^2, \quad (4)$$

where x_{ij} is the i -th and j -th elements of matrix \mathbf{X} , and matrix \mathbf{L} is the matrix takes the difference of adjacent elements along the row direction of matrix \mathbf{X} . We refer to the regularizer above as the *quadratic variation*.

Using the exclusive LASSO and quadratic variation as regularizers, can write the overall optimization problem:

$$\begin{aligned} \min_{\mathbf{A}, \mathbf{X}} D_{KL}(\mathbf{Y}, \mathbf{P}\mathbf{A}\mathbf{X}) + \frac{\alpha}{2} \|\mathbf{A}\mathbf{1}\|_2^2 + \frac{\beta}{2} \|\mathbf{L}\mathbf{X}^T\|_F^2 \quad (5) \\ \text{s.t. } \mathbf{A} \geq 0, \mathbf{X} \geq 0. \end{aligned}$$

B. Optimization method

Problem (5) can be solved using multiplicative update rules for NMF [11], which is based on the gradient descent with the auxiliary function approach [2]. First, we explicitly define the cost function as follows:

$$\begin{aligned} L(\mathbf{A}, \mathbf{X}) \\ = D_{KL}(\mathbf{Y}, \mathbf{P}\mathbf{A}\mathbf{X}) + \frac{\alpha}{2} \text{tr}(\mathbf{A}^T \mathbf{A} \mathbf{1} \mathbf{1}^T) + \frac{\beta}{2} \text{tr}(\mathbf{X}^T \mathbf{X} \mathbf{L}^T \mathbf{L}), \quad (6) \end{aligned}$$

where $\text{tr}(\mathbf{Z})$ is the trace of \mathbf{Z} . When we consider the standard gradient decent algorithm such as $\mathbf{A} \leftarrow \mathbf{A} - \epsilon \frac{\partial L(\mathbf{A}, \mathbf{X})}{\partial \mathbf{A}}$, it could violate the non-negativity constraint $\mathbf{A} \geq 0$ with some large step-size ϵ . By contrast, multiplicative update rules such as $a_{vr} \leftarrow a_{vr} w_{vr}$, do not violate the non-negativity constraint as both the optimization parameter and multiplicative weight are non-negative, $\mathbf{A} \geq 0$, $\mathbf{W} \geq 0$, respectively. The multiplicative weight should be decided based on the gradient such that $w_{vr} \leq 1$ for $\left[\frac{\partial L(\mathbf{A}, \mathbf{X})}{\partial \mathbf{A}} \right]_{vr} \geq 0$, and $w_{vr} \geq 1$ for $\left[\frac{\partial L(\mathbf{A}, \mathbf{X})}{\partial \mathbf{A}} \right]_{vr} \leq 0$. The following update rules can be used for minimizing the (6) with each non-negative variable, \mathbf{A} or \mathbf{X} .

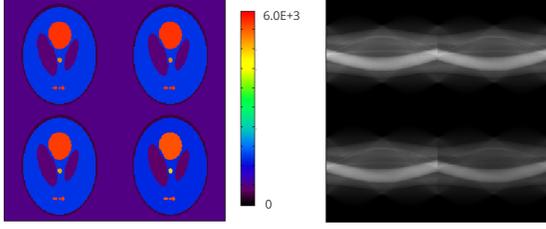


Fig. 2. Simulation image (left) and its sinogram (right)

1) *Update rule for A*: The differential of (6) with respect to \mathbf{A} can be obtained as follow:

$$\frac{\partial L(\mathbf{A}, \mathbf{X})}{\partial \mathbf{A}} = \mathbf{P}^T \mathbf{1}_{B \times F} \mathbf{X}^T - \mathbf{P}^T (\mathbf{Y} \oslash \mathbf{P} \mathbf{A} \mathbf{X}) \mathbf{X}^T + \alpha \mathbf{A} \mathbf{1} \mathbf{1}^T, \quad (7)$$

where \oslash means element-wise division; $\mathbf{1}_{B \times F}$ is a $(B \times F)$ -matrix of ones.

Therefore, we can obtain the following update rule:

$$\mathbf{A} \leftarrow \mathbf{A} \circ \left\{ \mathbf{P}^T (\mathbf{Y} \oslash \mathbf{P} \mathbf{A} \mathbf{X}) \mathbf{X}^T \oslash \left[\mathbf{P}^T \mathbf{1}_{B \times F} \mathbf{X}^T + \alpha \mathbf{A} \mathbf{1} \mathbf{1}^T \right] \right\}, \quad (8)$$

where \circ represents the Hadamard (element-wise) product.

2) *Update rule for X*: The differential of (6) with respect to \mathbf{X} can be obtained as follows:

$$\begin{aligned} \frac{\partial L(\mathbf{A}, \mathbf{X})}{\partial \mathbf{X}} &= \mathbf{A}^T \mathbf{P}^T \mathbf{1}_{B \times F} - \mathbf{A}^T \mathbf{P}^T (\mathbf{Y} \oslash \mathbf{P} \mathbf{A} \mathbf{X}) + \beta \mathbf{X} \mathbf{L}^T \mathbf{L} \\ &= \mathbf{A}^T \mathbf{P}^T \mathbf{1}_{B \times F} - \mathbf{A}^T \mathbf{P}^T (\mathbf{Y} \oslash \mathbf{P} \mathbf{A} \mathbf{X}) \\ &\quad + [\beta \mathbf{X} \mathbf{L}^T \mathbf{L}]_+ - [\beta \mathbf{X} \mathbf{L}^T \mathbf{L}]_-, \end{aligned} \quad (9)$$

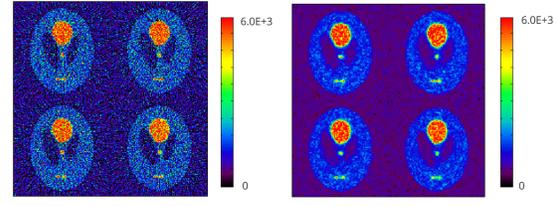
where $[\cdot]_+ := \max(0, \cdot)$ and $[\cdot]_- := |\min(0, \cdot)|$. Let \mathbf{X} be the update value when (9) becomes zero. Therefore, we can obtain the following update rule:

$$\mathbf{X} \leftarrow \mathbf{X} \circ \left\{ \mathbf{A}^T \mathbf{P}^T (\mathbf{Y} \oslash \mathbf{P} \mathbf{A} \mathbf{X}) + [\beta \mathbf{X} \mathbf{L}^T \mathbf{L}]_- \oslash \left\{ \mathbf{A}^T \mathbf{P}^T \mathbf{1}_{B \times F} + [\beta \mathbf{X} \mathbf{L}^T \mathbf{L}]_+ \right\} \right\}. \quad (10)$$

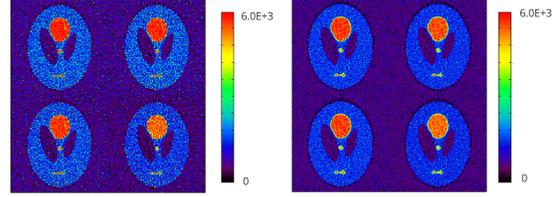
The PET image is reconstructed by updating the variables repeatedly using update rules above.

III. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed PET image reconstruction using the simulation data because true PET images are necessary. Fig. 2 shows the simulated parts of the PET image and sinogram data that we used in the experiments. The artificial sinogram data were created by simulations based on the compartment model [7], [8] with blood sampling data clinically measured from a real patient. Initially, the spatio-temporal distribution of the radio-pharmaceutical ligand in a virtual brain was simulated using the compartment model; subsequently, the time-series of the sinogram data were artificially created from the simulated ligand distribution by artificially measuring the gamma rays emitted from the ligand followed



(a) FBP (left) and EM (right)



(b) EM (all frame)(left) and Proposed (right)

Fig. 3. Examples of reconstruction images applied to each method

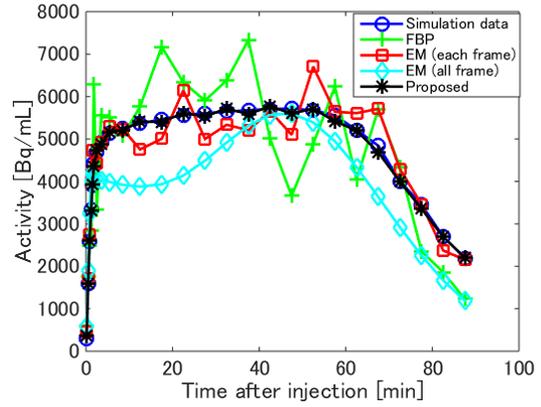


Fig. 4. Graph of time variation at a typical voxel in simulation image

with the Poisson distribution. The SNR of the sinogram data were controlled by varying the artificial measuring process. We applied three conventional methods and the proposed method to the temporal series of the sinogram data of various SNRs, and evaluated the SNR of each reconstructed PET image. One conventional method is the filtered back-projection [1], [13], [16]. Another conventional method is using EM algorithm at each frame. Further, another conventional method is the EM-based method [12] using basis functions that represent the temporal change of the time activity curves.

Fig. 3 shows results of the reconstructed images applied for each method when the SNR of the series of the sinogram data is 18.43 dB. As shown in Fig. 3, the results of the simulation images applied with the proposed method has less noise than the results of the simulation image applied with conventional methods.

Fig. 5 shows the graph of time variation at a typical voxel in the simulation image when the SNR of the sinogram series

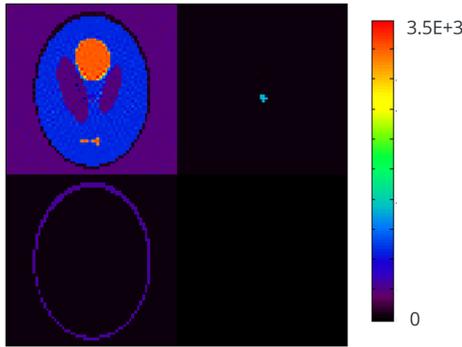


Fig. 5. Examples of results of spatial pattern

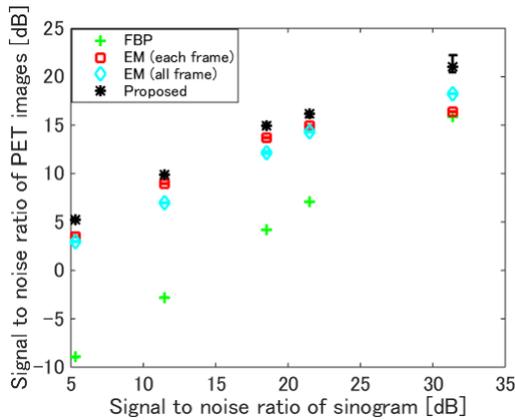


Fig. 6. Signal-to-noise ratio between constructed image and correct image

is 18.43 dB. As shown in Fig. 4, the time variations of the FBP and EM algorithm, which are conventional methods, include some noise and are different from their simulation data. However, the time variation of the proposed method is smooth and consistent with its simulation data.

Fig. 5 shows examples of the spatial pattern results. As shown in Fig. 5, each pattern is meaningful, and represents the outline of the brain and a partial area.

Fig. 6 shows the SNR graph between a reconstructed PET image and a true PET image using each method. The horizontal axis shows the SNR of the sinogram and the vertical axis shows the SNR of the reconstructed simulation images. As shown in the graph, the proposed method reconstructed the simulation images with the highest SNR among the applied conventional methods.

IV. CONCLUSIONS

In this paper, we proposed a new method that reconstructs the time-series of all PET images simultaneously from the time-series of sinograms. Physically meaningful features were extracted using constrained NMF. The smoothness constraint for the temporal feature was imposed, as well as the *exclusive*

LASSO-based constraint for the spatial feature. We experimentally demonstrated that our method could reconstruct PET images with a higher SNR than the conventional methods. Further, our method could separate several anatomical parts with different time variations. Future works entail performing experiments and evaluations with clinical data.

ACKNOWLEDGMENT

This work was supported in part by JSPS Grant-in-Aid for Scientific Research on Innovative Areas (Multidisciplinary Computational Anatomy) JSPS KAKENHI Grant Number 26108003, and THE HORI SCIENCES AND ARTS FOUNDATIONS.

REFERENCES

- [1] H. H. Barrett and W. Swindell. *Radiological Imaging: the Theory of Image Formation, Detection, and Processing*, volume 2. Academic Press, 1996.
- [2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977.
- [3] S. S. Gambhir. Molecular imaging of cancer with positron emission tomography. *Nature Reviews Cancer*, 2(9):683, 2002.
- [4] K. Kawai, J. Yamada, H. Hontani, T. Yokota, M. Sakata, and Y. Kimura. A robust PET image reconstruction using constrained non-negative matrix factorization. In *Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), 2017*, pages 1815–1818. IEEE, 2017.
- [5] W. E. Klunk, H. Engler, A. Nordberg, Y. Wang, G. Blomqvist, D. P. Holt, M. Bergström, I. Savitcheva, G.-F. Huang, S. Estrada, et al. Imaging brain amyloid in Alzheimer’s disease with pittsburgh compound-b. *Annals of Neurology*, 55(3):306–319, 2004.
- [6] D. Kong, R. Fujimaki, J. Liu, F. Nie, and C. Ding. Exclusive feature learning on arbitrary structures via $l_{1,2}$ -norm. In *Advances in Neural Information Processing Systems*, pages 1655–1663, 2014.
- [7] A. Lammertsma, C. Bench, S. Hume, S. Osman, K. Gunn, D. Brooks, and R. Frackowiak. Comparison of methods for analysis of clinical [^{11}C] raclopride studies. *Journal of Cerebral Blood Flow & Metabolism*, 16(1):42–52, 1996.
- [8] A. A. Lammertsma and S. P. Hume. Simplified reference tissue model for PET receptor studies. *Neuroimage*, 4(3):153–158, 1996.
- [9] K. Lange, R. Carson, et al. EM reconstruction algorithms for emission and transmission tomography. *Journal of Computer Assisted Tomography*, 8(2):306–16, 1984.
- [10] D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.
- [11] D. D. Lee and H. S. Seung. Algorithms for non-negative matrix factorization. In *Advances in Neural Information Processing Systems*, pages 556–562, 2001.
- [12] J. Matthews, D. Bailey, P. Price, and V. Cunningham. The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction. *Physics in Medicine and Biology*, 42(6):1155, 1997.
- [13] F. Natterer. *The Mathematics of Computerized Tomography*, volume 32. Siam, 1986.
- [14] M. E. Raichle. Behind the scenes of functional brain imaging: a historical and physiological perspective. *Proceedings of the National Academy of Sciences*, 95(3):765–772, 1998.
- [15] M. E. Raichle. A brief history of human functional brain mapping. In *Brain Mapping: The Systems*, pages 33–75. Elsevier, 2000.
- [16] A. G. Ramm and A. I. Katsevich. *The Radon Transform and Local Tomography*. CRC press, 1996.
- [17] L. A. Shepp and Y. Vardi. Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.
- [18] T. Yokota, R. Zdunek, A. Cichocki, and Y. Yamashita. Smooth nonnegative matrix and tensor factorizations for robust multi-way data analysis. *Signal Processing*, 113:234–249, 2015.