SUPERVISED PLANETARY UNMIXING WITH OPTIMAL TRANSPORT

Sina Nakhostin[(a,b)], Nicolas Courty[(a,b)], Rémi Flamary[(d)], Thomas Corpetti[(a,c)]

[(a)] : IRISA, UMR 6074, Vannes, France
[(b)] : Université de Bretagne-Sud
[(c)] : CNRS, UMR 6554 LETG - Rennes COSTEL
[(d)] : Laboratoire Lagrange, Université de Nice Sophia-Antipolis

ABSTRACT

This paper is focused on spectral unmixing and present an original technique based on Optimal Transport. Optimal Transport consists in estimating a plan that transports a spectrum onto another with minimal cost, enabling to compute an associated distance (Wasserstein distance) that can be used as an alternative metric to compare hyperspectral data. This is exploited for spectral unmixing where abundances in each pixel are estimated on the basis of their projections in a Wasserstein sense (Bregman projections) onto known endmembers. In this work an over-complete dictionary is used to deal with internal variability between endmembers, while a regularization term, also based on Wasserstein distance, is used to promote prior proportion knowledge in the endmember groups. Experiments are performed on real hyperspectral data of asteroid 4-Vesta.

Index Terms— Optimal Transport, Spectral Unmixing, Endmember Variability, Wassertein Distance, Bregman Projection

1. INTRODUCTION

Unmixing is an active field in hyperspectral data analysis and consists in estimating the abundance of pure materials (named endmembers) that are part of a pixel spectrum. These pure elements are either estimated from the image (unsupervised methods) or directly given in a prior knowledge (supervised approach).

Unmixing spectral information. A large number of algorithms have been proposed to deal with hyperspectral unmixing (HSU) (see [1, 2] for a comprehensive survey) and existing methods can roughly be distinguished by their mixing process assumption (linear or not). Under linear assumption, associated techniques have good relative balance between computational complexity and acceptable accuracy. However in practice such linear assumption is violated in many cases [1] for example because of *multiple scattering* phenomenon [3] that occurs due to the interaction of light with several materials (in macroscopic scale). It can also be due to *intimate mixing* of materials which refers to the simultaneous interaction of light with all the endmembers (in microscopic scale) [4]. These phenomena have raised a number of different nonlinear mixing models such as *intimate mixtures*, *bilinear models* and other physics-based nonlinear mixing models in order to give an approximate formulation of the real mixing process. Intimate mixture models rely in general in the formulation and inversion of the *radiative transfer* [5]. However all physical quantities involved, often depend on unknown parameters in the observed scene, yielding a difficult estimation process. This is the reason why (non-linear) analytical models such as *bilinear models* have been applied with success to represent the multiple scattering phenomenon. One can for example mention the non-linear model proposed by Nascimento [6], the general bilinear model [7], linear-quadratic mixing. [8]. Also a comparison of these models in a forestry context can be seen in [9]. To deal with nonlinearity, kernel-based algorithms which project the input data in a feature space of (possibly very) high dimension where linearity is improved also offer interesting alternatives [3, 10].

Endmember variability. In addition to abundance estimation, the problem of endmember quality has been the focus of several recent studies. As a matter of fact in many cases endmembers are either estimated up to some uncertainties or characterized not only by a single spectrum but rather by a set of different spectra. Using only one spectrum to decompose the information is then hazardous and likely to decrease the efficiency of the unmixing process. In order to cope with these issues, over complete dictionaries of hyperspectral signatures have reveal their efficiency [11, 12]. The general idea is to construct a set/dictionary of endmembers larger than the effective number of pure materials to deal with the internal heterogeneity inside a particular material (which should correspond to only one endmember in classical approaches). This principle will be explored in this paper.

Spectra comparison. & optimal transport In all techniques mentioned above, a metric between spectra is required. Several criteria have been used in a hyperspectral context (spectral angle mapper and Euclidean minimum distance being the two most popular ones), however, few approaches consider spectra as probability distributions in the spectral domain. We consider in this work an original metric, designed between probability density functions to compare spectra. This metric, based on optimal transport (OT) is called Wasserstein distance, and can be designed to be mostly sensitive to shifts in the frequency domains. To exploit OT, each spectrum should be normalized (the integration along spectral values should be 1) in order to be viewed as a probability distribution. By doing this normalization we make the analysis insensitive to quantities of materials in each pixel and we rather focus on proportions. This enables for example to improve robustness against shadows or other large radiance changes and as discussed in [13], can prevent degenerate solutions.

Outline of the paper. In the next section, we explain the basics of OT together with the entropy-based regularization term which is used for calculation of abundance coefficients. Experiments on the challenging 4-Vesta asteroid dataset are then shown and discussed in section 3.

2. UNMIXING USING PRIORS AND ITERATIVE BREGMAN PROJECTIONS

We first introduce optimal transport, then we present in section 2.2 a regularized version more efficient from a computational point of view and section 2.3 proposes an unmixing model based on optimal transport.

2.1. Optimal transport of spectra

Optimal Transport (OT) can be seen as the search for a plan that moves (transports) a spectrum μ_1 onto another spectrum μ_2 with a minimum cost. This cost is usually related to a metric of the embedding space. In our case, we will generally use a squared Euclidean distance L_2^2 as this cost, but it should be mentioned that other cost functions, specific to a given problem, might be used. In the relaxed formulation of Kantorovitch [14], OT seeks for an optimal coupling that can be seen as a joint probability distribution between μ_1 and μ_2. In other words, if we define $\Pi(\mu_1, \mu_2)$ the space of probability distributions over $\mathbb{R}^2$ with prescribed marginals μ_1 and μ_2, the optimal transport is the coupling $\gamma \in \Pi(\mu_1, \mu_2)$ which minimizes the following quantity:

$$W_c(\mu_1, \mu_2) = \inf_{\gamma \in \Pi(\mu_1, \mu_2)} \int_{\mathcal{R}^2} c(\mathbf{x}_1, \mathbf{x}_2) d\gamma(\mathbf{x}_1, \mathbf{x}_2) \quad (1)$$

where $c(\mathbf{x}_1, \mathbf{x}_2)$ is the cost between $\mathbf{x}_1$ and $\mathbf{x}_2$ (issued from distributions μ_1 and μ_2 respectively). In the discrete versions of the problem, *i.e.* when μ_1 and μ_2 are sampled by a sensor and expressed as vectors of $\mathbb{R}^d$, the previous equation reads:

$$W_{\mathbf{C}}(\mu_1, \mu_2) = \min_{\gamma \in \Pi(\mu_1, \mu_2)} < \gamma, \mathbf{C} >_F, \quad (2)$$

where $< ., . >_F$ is the Frobenius dot product and $\mathbf{C} \geq 0$ is a cost matrix (of size $d \times d$), which gathers all the costs for moving from bin i to bin j, and γ is now a matrix of size $d \times d$, with marginals defined as μ_1 and μ_2. We note that this distance is also known as the Earth Moving Distance in the computer vision community [15]. Solving for equation (2) is a simple linear programming problem with equality constraints, but its dimensions scales quadratically with the number of samples in the spectra. This might generate computational problems when the dimension of distributions is high, as in hyperspectral data. For this reason, we now consider a regularized version of the problem which has the extra benefit of being faster to solve.

2.2. Entropy regularized Optimal Transport and Bregman projections

The idea is here to put a regularization term over γ that controls the smoothness of the coupling through its entropy. The entropy regularized version of the discrete optimal transport problem reads [16]:

$$W_{\mathbf{C},\epsilon}(\mu_1, \mu_2) = \min_{\gamma \in \Pi(\mu_1, \mu_2)} < \gamma, \mathbf{C} >_F - \epsilon h(\gamma), \quad (3)$$

where $h(\gamma)$ is the entropy of γ, reading

$$h(\gamma) = -\sum_{ij} \gamma_{ij} \log \gamma_{ij} = - < \gamma, \log \gamma >_F,$$

if $\log$ is applied component-wise to γ. Denoting the Kullback Leibler divergence (KL), *i.e.* $\mathrm{KL}(\gamma, \rho) = \sum_{ij} \gamma_{ij} \log \frac{\gamma_{ij}}{\rho_{ij}} =< \gamma, \log \frac{\gamma}{\rho} >_F$ where again the $\log$ and the division are taken component wise, one can establish the following link between OT and Bregman projections:

Proposition 2.1 [OT as a Bregman projection [17, Eq. (6)]] *The minimizer $\gamma^\star$ of* (3) *is the solution of the following Bregman projection*

$$\gamma^\star = \operatorname*{arg\,min}_{\gamma \in \Pi(\mu_1, \mu_2)} \mathrm{KL}(\gamma, \zeta), \quad (4)$$

where $\zeta = \exp(-\frac{C}{\epsilon})$.

Interestingly enough, the entropy regularized version of OT admits a simple algorithm for the resolution of this problem, based on the successive projections over the two marginal constraints. More specifically, solving (4) when one of the two marginals is unknown admits a closed form solution.

Proposition 2.2 [Closed form solution of OT with one unknown marginal [17, Eq. (7)]] *For an undefined μ_2, $\gamma^\star$ is solely constrained by one marginal μ_1 and is the solution of the following closed-form projection:*

$$\gamma^\star = \text{diag}\left(\frac{\mu_1}{\zeta \mathbb{1}}\right)\zeta \tag{5}$$

where again the division has to be understood componentwise. Consequently, by iterating over successive Bregman projections on the set of the two marginal constraints, one can derive an efficient algorithm to solve for the OT problem. We refer the reader to [17] for a more complete treatment on this subject. Finally, when several OT distances are implied in the same optimization problem, this iterative projection strategy is still applicable. We use this property in our unmixing model, described in the following.

2.3. Unmixing model

We start here by recalling our initial assumptions. We assume a linear mixture for each spectrum $\mu : \mu = \mathbf{E}\alpha$ where matrix $\mathbf{E}$ of size $d \times q$ contains q (known) endmembers of dimension d and vector $\alpha > 0$ of size $q \times 1$ has to be estimated under the sum-to-one constraint, *i.e.* $\alpha^T \mathbb{1} = 1$. As such, it corresponds to discrete probability distribution function or histogram. From the image analysis, we seek to extract p abundances per pixels of distinct materials of interest. In practice the number of endmembers q exceeds the number of proportions p we aim at estimating ($q \geq p$). This allows to consider several distinct versions of spectra for the same material of interest, thus accounting for endmembers variability. In addition we assume that we have a prior knowledge $\alpha_\mathbf{0}$ (p-dimensional) over the abundances.

The unmixing of a spectrum μ is the solution of the following optimization:

$$\alpha = \arg\min_{\alpha} \quad \underbrace{W_{\mathbf{C}_0,\epsilon_0}(\mu, \mathbf{E}\alpha)}_{\text{data fitting}} + \tau \underbrace{W_{\mathbf{C}_1,\epsilon_1}(\alpha, \alpha_\mathbf{0})}_{\text{prior}}. \tag{6}$$

As it can be seen, the problem is subdivided in a data fitting term (which tries to find the best decomposition from observations) and a regularization term (which enforces the compliance of the solution with the prior) balanced by parameter $\tau \in \mathbb{R}^+$. From these two terms, regarding equation (3), two cost functions $\mathbf{C}_0$ and $\mathbf{C}_1$ are involved. Matrix $\mathbf{C}_0$ is linked with a metric in the spectral dimensions of the data (and is usually related to the L_2-norm along frequencies) while $\mathbf{C}_1$ is a cost matrix of size $q \times p$ with a specific structure which contains information about the classes of endmembers. Two endmembers related to the same element will share a very low cost with the corresponding material in $\alpha_\mathbf{0}$ but a higher cost with other materials. An easy way to handcraft $\mathbf{C}_1$ is to set $\mathbf{C}_1(i,j) = 0$ if endmember i is related to material j, and 1 elsewhere. Finally, two regularization parameters

Algorithm 1 Supervised unmixing procedure with OT

Input: μ
Parameters: $\alpha_\mathbf{0}$,$\mathbf{E}$,τ,$\mathbf{C}_0$,$\mathbf{C}_1$,ϵ_0,ϵ_1,maxIter
$cpt \leftarrow 0$, err $\leftarrow \infty$
$\alpha \leftarrow \mathbb{1}/q$
$\zeta_0 \leftarrow \exp(-\frac{C_0}{\epsilon_0})$, $\zeta_1 \leftarrow \exp(-\frac{C_1}{\epsilon_1})$
while cpt <maxIter **or** err>threshold **do**
  $\zeta_0 \leftarrow \text{diag}(\frac{\mu}{\zeta_0 \mathbb{1}})\zeta_0$
  $\zeta_1 \leftarrow \text{diag}(\frac{\alpha_\mathbf{0}}{\zeta_1 \mathbb{1}})\zeta_1$
  $\mu_{tmp} \leftarrow \exp(\log(\zeta_0 \mathbb{1}) + \tau \log(\mathbf{E}\zeta_1 \mathbb{1}))$
  $\zeta_0 \leftarrow \zeta_0 \,\text{diag}(\frac{\mu_{tmp}}{\zeta_0^T \mathbb{1}})$
  $\zeta_1 \leftarrow \zeta_1 \,\text{diag}(\frac{\mathbf{E}^T \mu_{tmp}}{\zeta_1^T \mathbb{1}})$
  err$\leftarrow ||\alpha^{(cpt)} - \alpha^{(cpt-1)}||_2$
  cpt$\leftarrow$cpt+1
end while
return α

ϵ_0 and ϵ_1 related to the regularized version of the transport have to be set. The resolution of the optimization problem in (6), inspired from the weighted Wasserstein barycenter problem [17], is done using iterative Bregman projections. Algorithm 1 presents the corresponding procedure in pseudo-code.

3. EXPERIMENTS AND ANALYSIS

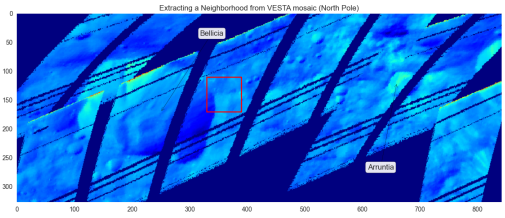


Fig. 1. Mosaic of northern hemisphere of Vesta-4 asteroid. The window in red (61×61) is used as the test bed.

3.1. 4-Vesta properties and dataset

A mosaic of Vesta-4 asteroid is visible in Figure 1. The image is taken by *visible and infrared mapping spectrometer* (VIR) which comes with two arrays of detectors: One array covering the visible range (between $.25\mu m$ and $1.05\mu m$) with spectral sampling of $1.8nm$ and the other covering the infrared range (between $1.0\mu m$ and $5.0\mu m$) with spectral sampling of $9.8nm$. The total number of considered bands after removing the noisy ones are $d = 383$ covering the range of $.55\mu m$ to $2.4\mu m$. Although there is no exact information about the components of the *crust* and *mantel* of this protoplanet asteroid, some expert knowledge is available (see [18] and [19]). From this we know that the major surface is dominated by three rocks: *Eucrite*, *Howardite* and *Diogenite*. It is believed

that roughly the southern hemisphere of 4-Vesta is Deogenite-rich while the northern hemisphere crust is mainly composed of *Eucrite* and different types of *pyroxene* with recently detected clusters of *Olivine*.

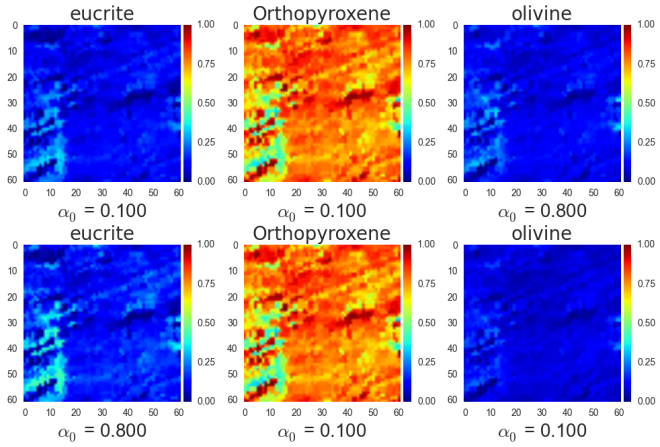
3.2. Experimental setup

We used a total number of $q = 20$ spectral signatures of different lithologies, found ubiquitously in the composition of the meteorites, collected on Earth. we could categorize these signatures in 3 groups: *Olivine*, *Orthopyroxene*, *Eucrit*. Consequently we set $p = 3$ in order to identify "maximally pure" endmembers which are believed to exist within the scene. As the consequence, matrix $\mathbf{C}_1$ involved in equation (6) will be of dimension 20×3. As mentioned in previous section, this cost matrix contains information about the classes of endmembers (two endmembers related to the same material will share a very low cost with the corresponding material in $\alpha_\mathbf{0}$ but a higher cost with other materials). In particular we set $\mathbf{C}_1(i, j) = 0$ if endmember i is related to material group j, and 1 elsewhere.

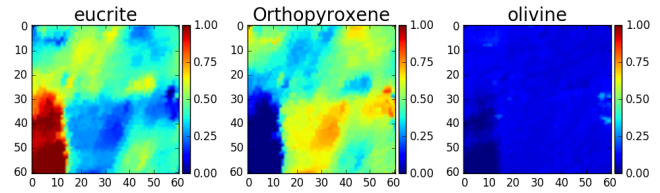
As for the cost matrix $\mathbf{C}_0$ in relation (6), we simply use the L_2^2 norm between each spectral value. The regularization parameters ϵ_0 and ϵ_1 have been set to 1 and 10^3 and the balance parameter is $\tau = 0.9$. This is set manually from data but more advanced approaches with cross-correlation could be used. For setting prior knowledge of the expected endmembers different combinations of non-negative 3-values for $\alpha_\mathbf{0}$ with non-negativity and additvity constraints were tested. By $\alpha_\mathbf{0}$ we reflect our prior expectation of the proportions of the 3 expected groups of lithologies.

3.3. Unmixing results and discussion

We made some experiments with setting different combinations of priors for each group of materials. By this we observe abundance maps with some nuances. These are represented in Fig. 2 a). It is interesting to observe that in all cases, spatially consistent areas extracted and correspond to existing spatial patterns of rocks. This is in contrast to the unmixing results we achieved by using constrained least squares (CLS) method which is illustrated in 2 b). As for the quantities of pure elements, our results are consistent with very recent studies [19] on this part of 4-Vesta. In particular the fact that different compositions of *Orthopyroxene* are dominant while *Olivine* patches are spatially much more isolated and sparse goes well with the latest observations on the composition of the materials of this area. As the role of priors one can see the change in susceptibility/sensitiveness towards a particular material when changing the prior value over that material. This in particular is visible w.r.t the Olivine maps.



a) Estimated Abundance Maps with different priors $\alpha_\mathbf{0}$.



b) Estimated Abundance Maps with Constrained Least Squares.

Fig. 2. Estimated Abundance Maps using OT (top) and CLS (bottom)

4. DISCUSSION/CONCLUSION

In this work we have proposed a hyperspectral unmixing approach based an optimal transport. This metric, devoted to distributions, is mostly senstive to shifts in the frequency domains and therefore differs from most usual criteria as spectral angle mapper or Euclidean distance. The problem of spectral variability and the confusion in selecting the best endmembers among the existing dictionary is addressed through an overcomplete dictionary. This results in a specific cost-function to optimize where both observation and regularization terms are based on OT. Our primarily experiments on the real 4-Vesta dataset show a good accordance of our results with the unmixing issued from the latest geological findings about the composition of lithologies on this asteroid.

5. ACKNOWLEDGMENTS

The Authors would like to thank *Conseil Régional de Bretagne* and *Université de Bretagne sud* for their support of the project *ARED -RB OPTIMAP*. They would like also to thank *Harold Clenet* for proving the VESTA-4 dataset and spectral measurements of the meteorites used in this work.

6. REFERENCES

[1] J.M. Bioucas-Dias and A. Plaza, "An overview on hyperspectral unmixing: Geometrical, statistical, and sparse regression based approaches," in *Geoscience and Remote Sensing Symposium (IGARSS), 2011 IEEE International*, July 2011, pp. 1135–1138.

[2] J.M. Bioucas-Dias, A. Plaza, N. Dobigeon, M. Parente, Q. Du, P. Gader, and J. Chanussot, "Hyperspectral unmixing overview: Geometrical, statistical, and sparse regression-based approaches," *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, vol. 5, no. 2, pp. 354–379, 2012.

[3] Jie Chen, C. Richard, and P. Honeine, "Nonlinear unmixing of hyperspectral data based on a linear-mixture/nonlinear-fluctuation model," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 480–492, Jan 2013.

[4] Nicolas Dobigeon, Jean-Yves Tourneret, Cédric Richard, Jose Bermudez, Steve Mclaughlin, and Alfred O Hero, "Nonlinear unmixing of hyperspectral images: Models and algorithms," *Signal Processing Magazine, IEEE*, vol. 31, no. 1, pp. 82–94, 2014.

[5] Bruce Hapke, *Theory of reflectance and emittance spectroscopy*, Cambridge University Press, 2012.

[6] José M. P. Nascimento and José M. Bioucas-Dias, "Nonlinear mixture model for hyperspectral unmixing," 2009, vol. 7477, pp. 74770I–74770I–8.

[7] Abderrahim Halimi, Yoann Altmann, Nicolas Dobigeon, and Jean-Yves Tourneret, "Nonlinear unmixing of hyperspectral images using a generalized bilinear model," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 49, no. 11, pp. 4153–4162, 2011.

[8] Ines Meganem, Ph Deliot, Xavier Briottet, Yannick Deville, and Sepehr Hosseini, "Linear–quadratic mixing model for reflectances in urban environments," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 52, no. 1, pp. 544–558, 2014.

[9] Wenyi Fan, Baoxin Hu, John Miller, and Mingze Li, "Comparative study between a new nonlinear model and common linear model for analysing laboratory simulated-forest hyperspectral data," *International Journal of Remote Sensing*, vol. 30, no. 11, pp. 2951–2962, 2009.

[10] Gustavo Camps-Valls and Lorenzo Bruzzone, "Kernel-based methods for hyperspectral image classification," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 43, no. 6, pp. 1351–1362, 2005.

[11] J. Bieniarz, E. Aguilera, X.X. Zhu, R. Muller, and P. Reinartz, "Joint sparsity model for multilook hyperspectral image unmixing," *Geoscience and Remote Sensing Letters, IEEE*, vol. 12, no. 4, pp. 696–700, April 2015.

[12] J. Bieniarz, R. Muller, Xiaoxiang Zhu, and P. Reinartz, "On the use of overcomplete dictionaries for spectral unmixing," in *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2012 4th Workshop on*, June 2012, pp. 1–4.

[13] Jakob Sigurdsson, Magnus O Ulfarsson, and Johannes R Sveinsson, "Endmember constrained semi-supervised hyperspectral unmixing," in *Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, 2014, p. 11.

[14] L. Kantorovich, "On the translocation of masses," *C.R. (Doklady) Acad. Sci. URSS (N.S.)*, vol. 37, pp. 199–201, 1942.

[15] Y. Rubner, C. Tomasi, and L.J. Guibas, "A metric for distributions with applications to image databases," in *Proc. ICCV*, Jan 1998, pp. 59–66.

[16] Marco Cuturi, "Sinkhorn distances: Lightspeed computation of optimal transport," in *Advances in Neural Information Processing Systems*, 2013, pp. 2292–2300.

[17] Jean-David Benamou, Guillaume Carlier, Marco Cuturi, Luca Nenna, and Gabriel Peyré, "Iterative bregman projections for regularized transportation problems," *SIAM Journal on Scientific Computing*, vol. 37, no. 2, pp. A1111–A1138, 2015.

[18] E Ammannito, MC De Sanctis, E Palomba, A Longobardo, DW Mittlefehldt, HY McSween, S Marchi, MT Capria, F Capaccioni, A Frigeri, et al., "Olivine in an unexpected location on vesta/'s surface," *Nature*, vol. 504, no. 7478, pp. 122–125, 2013.

[19] Jean-Philippe Combe, Thomas B McCord, Lucy A McFadden, Simone Ieva, Federico Tosi, Andrea Longobardo, Alessandro Frigeri, Maria Cristina De Sanctis, Eleonora Ammannito, Ottaviano Ruesch, et al., "Composition of the northern regions of vesta analyzed by the dawn mission," *Icarus*, vol. 259, pp. 53–71, 2015.