



HAL
open science

Discriminating industrial and smallholder oil palm plantations in Indonesia using Sentinel-1 textural features

Carl Bethuel, Julien Pellen, Damien Arvor, Samuel Corgne, Elise Colin, Jérémie Gignoux

► **To cite this version:**

Carl Bethuel, Julien Pellen, Damien Arvor, Samuel Corgne, Elise Colin, et al.. Discriminating industrial and smallholder oil palm plantations in Indonesia using Sentinel-1 textural features. IEEE IGARSS 2024, Jul 2024, Athènes, Greece. hal-04718575

HAL Id: hal-04718575

<https://hal.science/hal-04718575v1>

Submitted on 2 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

DISCRIMINATING INDUSTRIAL AND SMALLHOLDER OIL PALM PLANTATIONS IN INDONESIA USING SENTINEL-1 TEXTURAL FEATURES

*Carl Bethuel**, *Julien Pellen*,
Damien Arvor, *Samuel Corgne*

CNRS, Univ Rennes 2
LETG-UMR 6554
F-35000 Rennes
France

Elise Colin

DTIS, ONERA &
Université Paris Saclay
Palaiseau, France

Jérémy Gignoux

Paris School
of Economics
Paris, France

ABSTRACT

In recent decades, oil palm plantations have expanded significantly in Indonesia. Palm oil production currently relies on two plantation types with different economic, social and environmental impacts: (1) industrial plantations and (2) smallholder plantations. While efforts have been made to characterize these plantation types, this objective remains challenging for the remote sensing community. Consequently, this study assesses the potential of Sentinel-1 textural metrics in distinguishing between industrial and smallholder plantations, offering potential solutions to this persistent challenge. We used machine learning algorithms (Random Forest vs. eXtreme Gradient Boosting Tree) with textural features calculated by the Grey Level Co-occurrence Matrix. The results confirmed the potential of Sentinel-1 textural metrics to discriminate OP plantation types. The XGBTree model achieved a higher Kappa (0.71) than the Random Forest model (0.63). Moreover, the contrast, dissimilarity, GLCM-Mean, and GLCMVariance metrics were the most explanatory for discriminating industrial and smallholder plantations. This study possesses limitations, especially concerning its applicability on a broader scale. However, leveraging cloud computing tools like Google Earth Engine could aid in scaling up the methodology.

Index Terms— Oil Palm plantation, Sentinel-1, Time-series, Textural parameters, Classification.

1. INTRODUCTION

In recent decades, oil palm (OP) plantations have expanded significantly in Southeast Asia, especially in Indonesia, the world's leading producer. Due to high yields (three to four times higher than other oil crops such as olives, sunflowers,

or rapeseed), OP is the world's leading oil crop, involving significant economic, social, and environmental consequences. Indeed, OP plantations are often installed in forest or peatland areas, threatening numerous species in a region identified as a biodiversity hot-spot [1]. Moreover, the land rights of native peoples are regularly endangered by OP plantations, causing social and cultural challenges [2]. However, expanding oil palm plantations has also contributed to develop Indonesian agriculture and economy [3]. Today, palm oil production relies on several plantation types, which we have grouped into two classes for this work:

- highly capitalized industrial plantations corresponding to nucleus estates plantations and plasma smallholders [3],
- smallholder plantations representing independent smallholders [3].

Accurately characterizing these plantations is essential but is a challenging topic for the remote sensing community due to the high cloud cover in Indonesia and the similarities between the plantation types [4]. Despite the high cloud cover, optical data were widely used to discriminate OP plantations [4, 5, 6]. Yet, other studies used radar images [7, 8, 9] and highlighted their potential for characterizing OP plantation types. Textural metrics derived from radar annual composites have effectively detected OP plantations [10]. In the continuity of these findings, the objective of this work is to assess the potential of Sentinel-1 textural metrics for discriminating oil palm plantation types, i.e. industrial and smallholder plantations.

2. DATA

2.1. Study area

The study was conducted on an area of 170 km² (Figure.1) located in the province of North Sumatra. This province is located on Sumatra island (area 473,481 km²), which is the

*Thanks to the Région Bretagne and the French National Research Agency (ANR) through the Palmexpand project (grant agreement: ANR-20-CE03-0004) for funding.

Contact : carl.bethuel@univ-rennes2.fr

island most affected by OP expansion (9 million ha in Sumatra in 2019 vs 6 million ha in Kalimantan [11]). The study area was chosen to represent both industrial and smallholder plantations.

2.2. Sentinel-1 data

This work relied on a five-year radar time series from the Sentinel-1A satellite (February 25, 2017 to December 31, 2022). This time series of 176 dates (i.e., about 30 images per year) was automatically downloaded from the European Space Agency hub in Ground Range Detected (GRD) format. The images were acquired at C-band every 12 days in the Interferometric Wide mode on ascending orbit n°143 with dual VV/VH polarization at 10-meter spatial resolution.

2.3. Reference data

The reference data is a map of industrial and smallholder plantations provided by a two-step approach [11]. First, the mapping of industrial plantations has been performed through visual interpretation of Landsat annual median composites between 2000 and 2019. Second, smallholder plantations were mapped by the AURIGA non-governmental organization based on the visual interpretation of SPOT-6 images at a 1.5m spatial resolution for 2016. This 2016 map, approved by the Indonesian government, was then updated until 2019 through change detection techniques using Sentinel-1 composites, whereas the Tree Loss [12] product was used to estimate the planting years up to 2000. The maps of industrial plantations and smallholders were then merged and corrected where conflicts arose.

Based on this reference map, we located all pixels of the Sentinel-1 image corresponding to oil palm areas and then selected 10% of these pixels (i.e., 68,929 pixels for industrial plantations and 31,934 pixels for smallholder plantations) to create a reference dataset, further divided into a training set (70,607 pixels) and a validation set (30,256 pixels).

3. METHOD

The method is divided in two stages: (1) calculation of textural metrics and (2) classification models.

3.1. Calculation of textural metrics

From the GRD images with linear Sigma0 values, we computed ten textural metrics (i.e., Contrast, Dissimilarity, Homogeneity, Angular Second Moment (ASM), Energy, Maximum Probability, Entropy, GLCM Mean, GLCM Variance, and GLCM Correlation) using the Grey Level Co-occurrence Matrix (GLCM) for each polarisation (i.e., VV and VH). These metrics were calculated with a 5 x 5 window to derive 2060 textural features considering :

- each of the 176 dates, i.e. 1760 bands corresponding to 176 VH images * 10 textural metrics,
- 10 mean seasonal composites (i.e., dry season from May to October vs. wet season from November to April), i.e. 200 images corresponding to (10 VH composites + 10 VV composites) * 10 textural metrics,
- 5 mean annual composites, i.e. 100 images corresponding to (5 VH composites + 5 VV composites) * 10 textural metrics.

For the 176 dates, we focused only on the VH polarization as they generally provide significant information for vegetation characterization [7]. For the mean composite images (seasonal and annual), VH and VV are used as they provide despeckle information useful for textural parameters extraction.

3.2. Classification models

Once the textural features were calculated, we used supervised classification models to discriminate the OP plantation types. First, we selected the most explanatory textural features by using the Gini index (in this work, Gini index > 23) of Random Forest (RF). Thus, 195 features were identified and further used to compare two classification models : a Random Forest [13] and an eXtreme Gradient Boosting Tree (XGBTree) [14, 15]. On the one hand, Random Forest, regularly used with remote sensing data, is recognized for its accuracy. Its operation, based on independent regression trees (bagging model), is robust to the noise present in the data. On the other hand, XGBTree is a recent model iterating on mutually dependent decision trees (boosting model). XGBTree has several advantages, such as low data-processing capacity, significantly increased algorithm speed by reducing computational memory in large-scale data training, and high-performance results, which have recently been very attractive. [16, 17]. The models were computed with the same reference training set (i.e., 70,607 pixels) and validation set (i.e., 30,256 pixels) divided according to the two classes (i.e., smallholder vs. industrial). Finally, several tests were calculated using the grid-search cross-validation approach to identify the best hyper-parameters and optimize the results.

4. RESULTS

First, the Table.1 presents the most explanatory textural features occurrences (i.e., 195 features). We note that GLCM Variance, GLCM Mean, Contrast and Dissimilarity metrics appear respectively 40, 39, 32 and 31 times (i.e., 20.5%, 20%, 16.4% and 15.9% of the 195 textural features). According to Haralick et al., [18], the GLCM Variance and GLCM Mean metrics correspond to statistical parameters providing denoised information on pixel neighbourhood (in this

case, mean or variance of neighbouring pixels). For contrast and dissimilarity, Haralick et al. identify them as contrast variables that provide information (local or global) on the variation of intensity in pixels. Differences in organization between industrial (homogeneous organization) and smallholder (heterogeneous organization) plantations significantly impact the contrast features, with marked intensity variations discriminating between OP plantation types.

Metrics	Number of occurrences & (Percentage)
Contrast	32 (i.e., 16.4%)
Dissimilarity	31 (i.e., 15.9%)
Homogeneity	15 (i.e., 7.7%)
ASM	9 (i.e., 4.6%)
Energy	8 (i.e., 4.1%)
Maximum probability	1 (i.e., 0.5%)
Entropy	10 (i.e., 5.1%)
GLCMMean	39 (i.e., 20%)
GLCMVariance	40 (i.e., 20.5%)
GLCMCorrelation	10 (i.e., 5.1%)

Table 1. Textural features occurrences from the Random Forest (i.e., 195 features)

Then, the Table.2 present the most explanatory temporal features occurrences. It is worth mentioning that the temporal metrics emphasize annual and seasonal composite images (83 and 92 times respectively, i.e., 42.6% and 47.2% of the 195 features). We also note that the importance of annual and seasonal composites is similar. On the contrary, the single-date images appear only 20 times (10.3%), testifying that high temporal resolution (i.e., every 12 days for Sentinel-1 data) is not necessary to discriminate industrial and smallholder plantations since oil palms have a long phenological cycle (around 25 years). Moreover, temporal composites despeckle the radar data, providing finer textural information.

Metrics	Number of occurrences & (Percentage)
Annual composite images	83 (i.e., 42.6%)
Seasonal composite images	92 (i.e., 47.2%)
Single-date images	20 (i.e., 10.3%)

Table 2. Temporal features occurrences from the Random Forest (i.e., 195 features)

The classification (Figure.1) shows a salt-and-pepper effect (e.g., in the large area of smallholder plantations to the East of our study area). Errors are also frequently observed on the edges of plantations, showing that it is still challenging to discriminate plantation types or between plantations and other land uses (e.g. forests, roads and rivers). Statistical validation of the classifications presented in Table 3 highlights that the highest Kappa is achieved by the boosting model (XGBTree

Kappa =0.71), testifying the under-performance of the bagging model (RF Kappa = 0.63) to discriminate industrial and smallholder plantations. These results can be explained as follows:

- Discrimination of OP plantation types is a difficult task that can be better achieved with sequentially trained trees with adjustments made from the error of the previous tree,
- Random forest models are prone to overfitting, whereas XGBTree models can counter this by creating shallower trees,
- XGBTree works well on a wide variety of large datasets.

Furthermore, the User and Producer Accuracy show that smallholder plantations are poorer detected than industrial, with more errors of commission than errors of omission, i.e., Producer Accuracy (0.75) < User Accuracy (0.85) (Table.3). This finding is explained by the heterogeneous organization of smallholders planted in small areas, directly impacting image texture with high contrast and dissimilarity values. Finally, we note that the RF model has more errors of omission than the XGBTree, particularly for industrial plantations (Table.3). In other words, the Random Forest classifies more industrial plantations as smallholders than the XGBTree.

Classification models	Kappa	User Accuracy		Producer Accuracy	
		Ind	SH	Ind	SH
Random Forest (bagging)	0.63	0.85	0.84	0.94	0.65
xgbTree (boosting)	0.71	0.89	0.85	0.94	0.75

Table 3. Statistical validation of classifications

5. DISCUSSION/CONCLUSION

The aim of this article, to distinguish smallholders from industrial plantations, remains a challenge for the remote sensing community. Indeed, the similarities between plantation types make them challenging to distinguish, and grouping them into just two classes (i.e. smallholders vs. industrial plantations) is questionable as it accentuates similarities. However, our reference data and various studies [19, 20, 11] are based on these two types of plantation. Hence, to identify OP plantation types, we computed machine learning algorithms (RF vs. XGBTree) with textural features. Four textural metrics (Contrast, dissimilarity, GLCM mean, and GLCM variance) of composite images are the most explanatory in discriminating smallholders from industrial plantations. Even if the best Kappa (0.71) was achieved by the XGBTree model, the Producer and User Accuracies show

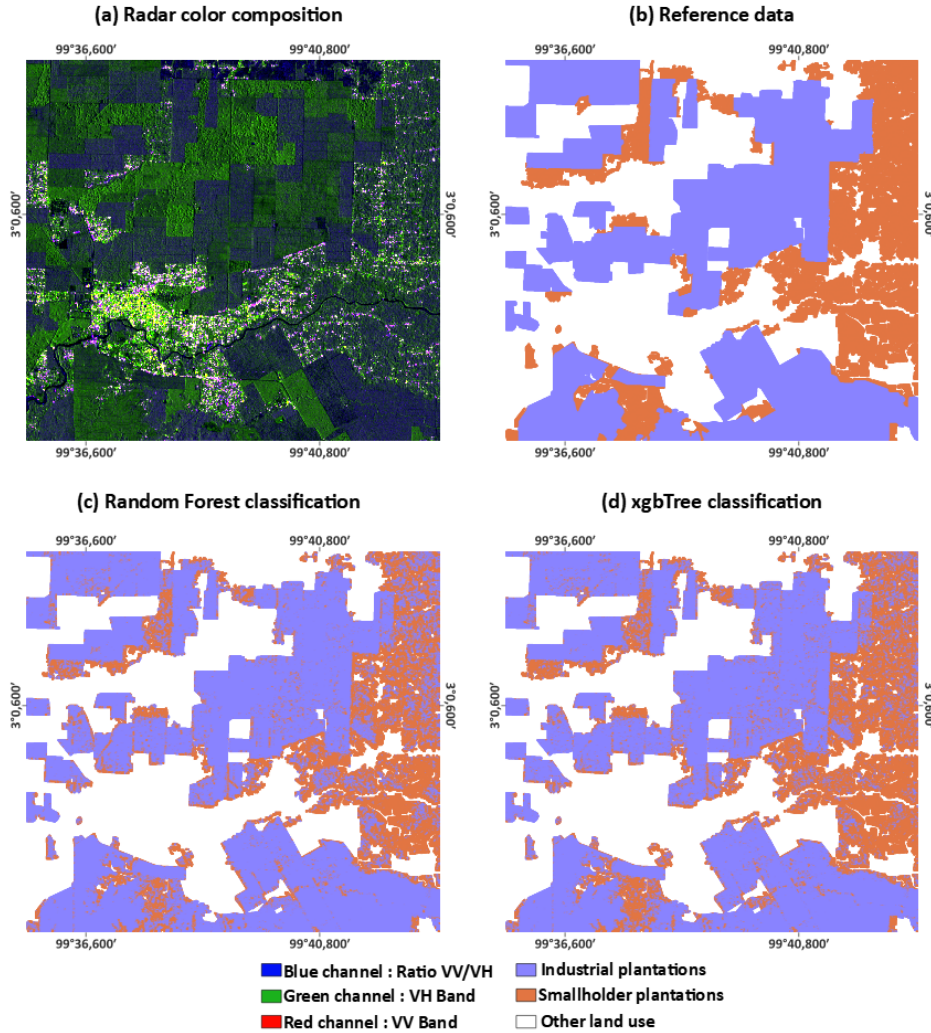


Fig. 1. Characterization of Oil palm plantations by (a) Radar color composite of 2022 mean composite, (b) Reference data, (c) Random Forest classification, and (d) XGBTree classification

that smallholders have more commission errors than omission errors due to their heterogeneous organization impacting the textural information.

These limits could be reduced by using additional optical data. Even if such data are limited in Indonesia due to very high cloud cover rates, they offer vegetation information, especially in the Red and Near-Infrared bands [20], and the synergy between optical and radar data has already been recognized to identify the types of oil palm plantations [21]. Concerning the radar, we could investigate more complex radar features such as polarimetric parameters (e.g., degree of polarization [22]). Taking the temporal dimension into account could also improve results or at least demonstrate the relevance of using a specific type of temporal data (e.g. composite images).

In parallel with these developments on the data, we could also use different classification models, in particular, deep

learning models according to two approaches: a spatial approach [19] and a temporal approach [23]. In addition, post-processing steps such as a filtering operation (e.g. a median filter) or ancillary decision criteria such as roads, oil mills or hydrological networks could reduce "salt and pepper" and plantation edges effects (figure 1). From an application standpoint, the results remain consistent within our 170 km² study area. However, scaling up to a larger extent could provide spatial information to address environmental, social, and economic questions. To achieve this, using cloud computing platforms such as Google Earth Engine [24] would be essential to circumvent the time-consuming steps of data downloading and processing.

6. REFERENCES

- [1] David S. Wilcove and Lian Pin Koh, "Addressing the threats to biodiversity from oil-palm agriculture," *Biodivers. and Conserv.*, vol. 19, no. 4, pp. 999–1007, Apr. 2010.
- [2] Marcus Colchester, "Palm oil and indigenous peoples in South East Asia Land acquisition, human rights violations and indigenous peoples on the palm oil frontier," Tech. Rep., IUCN, Washington D.C, Sept. 2010.
- [3] Rob Cramb and John F. McCarthy, *The Oil Palm Complex: Smallholders, Agribusiness and the State in Indonesia and Malaysia*, NUS Press, 2016.
- [4] Khai Loong Chong, Kasturi Devi Kanniah, Christine Pohl, and Kian Pang Tan, "A review of remote sensing applications for oil palm studies," *Geo-spatial Inf. Sci.*, vol. 20, no. 2, pp. 184–200, Apr. 2017.
- [5] Samuele De Petris, Piero Boccardo, and Enrico Borgogno-Mondino, "Detection and characterization of oil palm plantations through MODIS EVI time series," *Int. J. of Remote Sens.*, vol. 40, no. 19, Oct. 2019.
- [6] Aslinda Oon, Helmi Zulhaidi Mohd Shafri, Alex Mark Lechner, and Badrul Azhar, "Discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands using vegetation indices and supervised classification of LANDSAT-8," *Int. J. of Remote Sens.*, vol. 40, no. 19, pp. 7312–7328, Oct. 2019.
- [7] Aslinda Oon, Khanh Duc Ngo, Rozilah Azhar, Adham Ashton-Butt, Alex Mark Lechner, and Badrul Azhar, "Assessment of ALOS-2 PALSAR-2L-band and Sentinel-1 C-band SAR backscatter for discriminating between large-scale oil palm plantations and smallholdings on tropical peatlands," *Remote Sens. Applications: Soc. and Environ.*, vol. 13, pp. 183–190, Jan. 2019.
- [8] Fatwa Ramdani, "Recent expansion of oil palm plantation in the most eastern part of Indonesia: feature extraction with polarimetric SAR," *Int. J. of Remote Sens.*, vol. 40, no. 19, pp. 7371–7388, Oct. 2019.
- [9] Lissa Fajri Yayusman and Ryota Nagasawa, "ALOS-Sensor data integration for the detection of smallholder's oil palm plantation in Southern Sumatra, Indonesia," *J. of the Jpn. Agric. Syst. Soc.*, vol. 31, no. 2, pp. 27–40, 2015.
- [10] Olga Danylo, Johannes Pirker, Guido Lemoine, Guido Ceccherini, Linda See, Ian McCallum, Hadi, Florian Kraxner, Frédéric Achard, and Steffen Fritz, "A map of the extent and year of detection of oil palm plantations in Indonesia, Malaysia and Thailand," *Scientific Data*, vol. 8, no. 1, pp. 96, Mar. 2021.
- [11] David L. A. Gaveau, Bruno Locatelli, Mohammad A. Salim, Husnayaen, Timer Manurung, Adria Descals, Arild Angelsen, Erik Meijaard, and Douglas Sheil, "Slowing deforestation in Indonesia follows declining oil palm expansion and lower oil prices," *PLOS ONE*, vol. 17, no. 3, Mar. 2022.
- [12] M. C. Hansen, P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend, "High-Resolution Global Maps of 21st-Century Forest Cover Change," *Sci.*, vol. 342, no. 6160, pp. 850–853, Nov. 2013.
- [13] Leo Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [14] Tianqi Chen and Carlos Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, Aug. 2016, KDD '16, pp. 785–794, Association for Computing Machinery.
- [15] Jerome H. Friedman, "Greedy function approximation: A gradient boosting machine.," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, Publisher: Institute of Mathematical Statistics.
- [16] Nan Zhang, Mingjie Chen, Fan Yang, Cancan Yang, Penghui Yang, Yushan Gao, Yue Shang, and Daoli Peng, "Forest Height Mapping Using Feature Selection and Machine Learning by Integrating Multi-Source Satellite Data in Baoding City, North China," *Remote Sensing*, vol. 14, no. 18, pp. 4434, Jan. 2022, Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.
- [17] Yuzhen Zhang, Jingjing Liu, Wenhao Li, and Shunlin Liang, "A Proposed Ensemble Feature Selection Method for Estimating Forest Aboveground Biomass from Multiple Satellite Data," *Remote Sensing*, vol. 15, no. 4, pp. 1096, Jan. 2023, Number: 4 Publisher: Multidisciplinary Digital Publishing Institute.
- [18] Robert M. Haralick, K. Shanmugam, and Its'hak Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973, Conference Name: IEEE Transactions on Systems, Man, and Cybernetics.
- [19] Adria Descals, Serge Wich, Erik Meijaard, David L. A. Gaveau, Stephen Peedell, and Zoltan Szantoi, "High-resolution global map of smallholder and industrial closed-canopy oil palm plantations," *Earth Syst. Sci. Data*, vol. 13, no. 3, pp. 1211–1231, Mar. 2021.

- [20] Adrià Descals, Zoltan Szantoi, Erik Meijaard, Harsono Sutikno, Guruh Rindanata, and Serge Wich, "Oil Palm (*Elaeis guineensis*) Mapping with Details: Smallholder versus Industrial Plantations and their Extent in Riau, Sumatra," *Remote Sens.*, vol. 11, no. 21, pp. 2590, Jan. 2019.
- [21] Thuan Sarzynski, Xingli Giam, Luis Carrasco, and Janice Ser Huay Lee, "Combining Radar and Optical Imagery to Map Oil Palm Plantations in Sumatra, Indonesia, Using the Google Earth Engine," *Remote Sens.*, vol. 12, no. 7, pp. 1220, Jan. 2020.
- [22] Elise Colin, "PolTimeSAR: Link between the measurement of a time series of N Jones vectors, and the measurement of a single Stokes vector integrating all these measurements. Application to (HH,HV) or (VH,VV) SAR time series," Mar. 2023.
- [23] Charlotte Pelletier, Geoffrey I. Webb, and François Petitjean, "Temporal Convolutional Neural Network for the Classification of Satellite Image Time Series," *Remote Sensing*, vol. 11, no. 5, pp. 523, Jan. 2019, Number: 5 Publisher: Multidisciplinary Digital Publishing Institute.
- [24] Noel Gorelick, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. of Env.*, vol. 202, pp. 18–27, Dec. 2017.