

O&GProfile : An automated method for attribution of satellite methane emissions detections to oil and gas sites and operators

Jade E Guisiano, Thomas Lauvaux, Claudio Cifarelli, Éric Moulines, Jérémie Sublime

▶ **To cite this version:**

Jade E Guisiano, Thomas Lauvaux, Claudio Cifarelli, Éric Moulines, Jérémie Sublime. O&GProfile : An automated method for attribution of satellite methane emissions detections to oil and gas sites and operators. International Conference on Machine Learning and Data Mining MLDM 2023, Jul 2023, New-York, United States. hal-04043407

HAL Id: hal-04043407
https://hal.science/hal-04043407v1

Submitted on 23 Mar 2023

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.



HAL Authorization

O&GProfile : An automated method for attribution of satellite methane emissions detections to oil and gas sites and operators

Jade E. Guisiano[1,2,3], Thomas Lauvaux[4], Claudio Cifarelli[3], Éric Moulines[2], and Jérémie Sublime[1]

[1] ISEP School of Engineering Paris, France
first.last@isep.fr
[2] École Polytechnique, Palaiseau, France
first.last@polytechnique.edu
[3] United Nations Environment Program, France
first.last@un.org
[4] Université de Reims, Reims, France
first.last@univ-reims.fr

Abstract. Methane emissions are the second leading cause of global warming and in order to be able to move towards adapted and personalized regulations, it is necessary to aim at new methods improving the current assessment of emission trends at the level of the whole oil and gas (O&G) supply chain. Emission profiles should be estimated by operator and by type of O&G site to design and to implement specific regulations. To establish the most complete emission profiles possible, a maximum of methane emissions measurements at various scales (satellites, UAVS, ground sensors) should be collected and associate to the different part of the O&G supply chain and concerned operators. We therefore propose in this paper "O&GProfile", the first automated method which permits to associate methane plumes detections to oil and gas sites type (Gathering & boosting, processing, production) and their respective operators. O&GProfile is based on the use of unsupervised machine learning methods for clustering purposes, and a semi-automated correction method. This is the first method which permits to associated automatically methane detection to the type of site and operator at the origin of the methane leak. Over the period 2020-2021 in the Delaware and Midland basins in the Permian, the O&GProfile was able to correctly associate point-source satellite detections to sites in 98.8% of cases, and the correction process permits to obtain 100% of correct associations. These associations allow to establish emissions profile by site and operator by combining point-source satellite and PermianMAP labeled detections.

Keywords: Spatial clustering · Satellite· methane· O&G

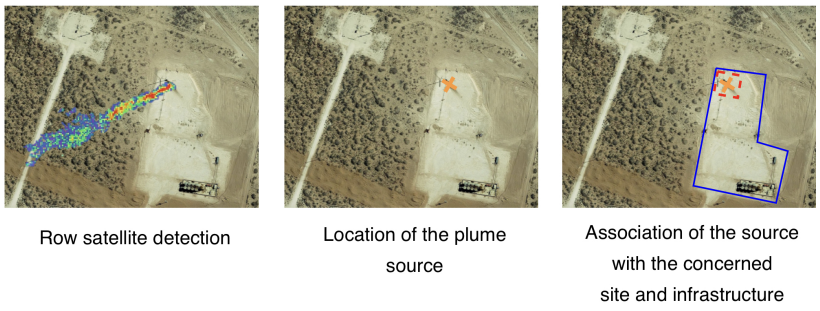


Fig. 1. Process from plume detection to its association to site and infrastructure concerned

1 Introduction

The oil and gas sector is the second largest anthropogenic methane emission source and emission reductions from this sector can be achieved using existing and proven technologies. Furthermore, in contrast to agriculture and wetlands the oil and gas industry has experience with reducing methane emissions due to safety concerns and has a certain financial benefit to reducing emissions as methane can be sold as natural gas. The International Energy Agency (IEA) estimates that the industry can reduce its global emissions by 75% and that up to 40% of these reductions can be achieved without financial compromise. However, the lack of reliable emissions data has made it hard for governments to carry out targeted action at the scale and speed needed to achieve the objectives of the Global Methane Pledge (GMP). In order to set up an efficient regulations for the reduction of methane emissions, it is necessary to first establish a new reliable and accurate inventory methodology that allows to establish methane emission profiles as close as possible to reality, and at several levels of granularity, from site to the national level.

There are currently about 10 satellites dedicated to methane emissions detection (from global-regional to point source view) which have various time interval between successive viewings of the same scene (between 1-14 days)[8]. Airbone campaigns for the detection of methane emissions are also conducted on a regular basis (over specific periods of time) in various targeted areas (region and basin scale). There are also other aerial means of emissions detection such as drones which are generally used for private studies on specific oil and gas sites. The collect of detection over a certain period from all these instruments in a precise zone (e.g O&G basin) can rapidly lead to big data configuration. For each of the methane leak detection represented by a plume, a geographical location is assigned to them[5] determining the potential location of the source of the leak as

[5] The determination of the origin of a methane plume is one of the results obtained by the inversion process (for global view- regional) and by various other methods for point-source view[8]

shown in the first two images of Figure 1. However, the geographic location of a detection in an O&G basin by itself does not directly indicate which type of site, infrastructure and operator are at the origin of the leak. In order to obtain a profile of emissions by type of site, infrastructure and operator it is first necessary to determine for each methane detection (from various instruments) what type of site and infrastructure is at the origin. This association of methane emission detections from various instruments to a multitude of sites/infrastructures (e.g. entire O&G basin scale) could rapidly become a huge manual task as there is currently no available automated method for this task. Indeed, this type of association of detections to the sites concerned and operator -for reasonable surveyed zone - is a long manual work carried out by experts.

For example, the PermianMAP[2] open-source project of Environment Defense Found (EDF) offers a database containing methane detections from the CarbonMapper[1] satellite and the Global Airborne Observatory (GAO) aerial campaign where each detection has been manually linked to a type of site, a concerned infrastructure and an operator in the Permian basin. In order to eliminate this time consuming manual task we propose O&GProfile, a first automated method based on the unsupervised machine learning algorithm "Density-based spatial clustering of applications with noise" (DBSCSAN)[6] able to associate satellite methane emissions detections to the O&G sector and link them to specific site type and operator. A study[10] proposes the use of the DBSCAN algorithm to gather methane plumes in areas of about 30 km, however the characteristics of the area concerned remain unknown and not directly qualifiable.

The PermianMAP database is integrated in the DBSCAN clustering in order to transfer site type and operator information to satellite point source detections. To do so, to inherits from PermianMap annotated detection, satellite point source detection need to be positioned on a site which also contain PermianMap detection. Therefore, clusters comprising PermianMap and satellite point source measurements will be targeted. The association of the detections to the infrastructures require high precision as the infrastructure could be very close to each others on a site. Then the obtaining of uncertainty related to the localization of the leak is required in order to associate a detection to these infrastructures. As we do not have this uncertainty parameter, the association at infrastructure level is voluntarily not treated. The O&GProfile also possess a correction step based on Voronoï[3]-Nearest Neighboor(NN) KDTree[5] algorithm which permit to identified and correct wrong associations from DBSCAN results based on PermianMap dataset. Site types be divided in 3 major categories of the oil and gas supply chain : production, gathering & boosting, processing. In addition to the interest of monitoring methane emissions by site and operator that our method makes possible, The O&GProfile also allows to labelled the raw satellite data by associating them to the O&G site type and operator where the leak comes from. This last aspect makes it possible to transform raw satellite detection to an informed detection database which can be used independently for any type of analysis.

The case study presented in this paper concerns the Permian area and more precisely the Delaware and Midland basins over the period 2021-2022. In the first part, we will present the characteristics of the CarbonMapper satellite data, the GAO aerial data - from the PermianMap- and the point-source satellite detections dataset. In a second part, the O&GProfile method, based on DBSCAN and Voronoï-NN KDTree algorithms will be detailed. And finally the results will be discussed in the last part and in the conclusion

2 Context and dataset description

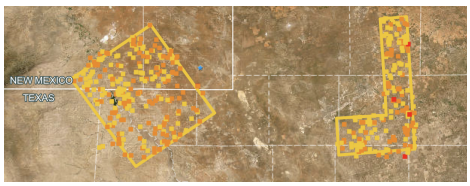


Fig. 2. Perimeter of the study zone of PermianMAP survey (left: Delaware basin, right: Midland basin)

2.1 Context

The Permian Basin is the largest O&G basin in the United States, it covers 86,000 square miles of land across West Texas and Southeast New Mexico and is home to more to tens of thousands of O&G sites. Many works, as for example [11][7][4] have been devoted to the study of methane emissions over different angle of view, which make the Permian one of the most informed basins in terms of data. The Environmental Defense Fund (EDF), through the PermianMap project, has launched two GAO air campaigns and the collection of data from the CarbonMapper satellite to survey its emissions in the Permian basin. All these detections, according to their location, have been assigned to operators, type of site and emitting infrastructure. These open-source data allow indirectly to have access to the geographical coordinates of the sites and to their information concerning their type and the concerned operator. This information is particularly sought after and useful in order to be able to inform detections from any other satellite measurements. In this study we will use the data from CarbonMapper and GAO to label raw satellite point source detections in order to obtain emissions profile.

2.2 Dataset description

Point-source satellites characteristics The point-source satellite data are acquired via 8 satellites: Landsat-8, WorldView-3, Sentinel-2, GHGSat, PRISMA,

EnMAP, EMIT and CarbonMapper. which are fine-pixel ($\leq 60m$) instruments designed to quantify individual point sources by imaging of the plume. Their detection thresholds in the 100–10 000 kg/h range that enable monitoring of small to large point sources.[8] The point source satellite used in this study will be name Satellite X for confidentially reasons.

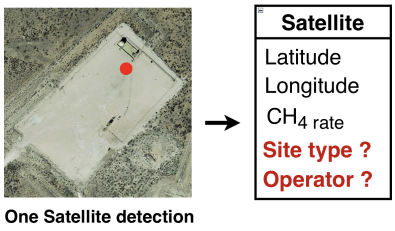


Fig. 3. Illustration of satellite X informations for one detection

Satellite X dataset description The satellite X data consists of a set of detections of methane emissions detected from various sources and around the world. To each detection is associated the estimated geographic coordinates of the detected plume (latitude and longitude)[6], the estimated amount of methane emitted and the uncertainty about this estimate (in %).This initial dataset covers the period 01/01/2021 to 30/09/2022.

PermianMAP project description Environmental Defense Fund launched the Permian Methane Analysis Project (PermianMAP), making near real-time methane monitoring data available to the public online enabling oil and gas companies, regulators and other stakeholders. The majority of measurements are taken in a 10,000 square-kilometer grid that spans the Delaware Basin and a portion of Midland Basin, a subsection of the broader Permian Region. Methane plume detections have been observed with the AVIRIS-NG instrument installed on the Global Atmospheric Observatory (GAO) aircraft in coordination with the satellite CarbonMapper. The measurements were made during two campaigns (summer and fall) in 2021 on predetermined regions in both Delaware (5000 km2) and Midland (2500 km2) sub-basins. Flights occurred on all days with acceptable atmospheric conditions with the intent to map all predetermined regions at least three times during the flight window. All detected emitting sources were then manually inspected using a combination of high-resolution imagery collected by the aircraft and Google Earth satellite imagery to determine the segment of O&G production responsible: Production, Gathering & Boosting and Processing. The observed plumes and emission sources were also attributed to the most likely responsible operators for each source by using a combination of open-source and

[6] It should be noted here that the uncertainty associated with the estimated location of each plume is not reported.

commercial databases. Since the Permian area is very dense and the sites are not widely spaced, the uncertainty as to the attrition of an operator to a site is present but not quantified.

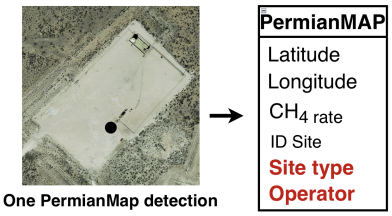


Fig. 4. Illustration of PermianMap informations for one detection

PermianMAP dataset description The selected data are distributed over two measurement periods, the first from 2021-07-26 to 2021-08-10 and the second from 2021-10-03 to 2021-10-17. These two campaigns have allowed the collection of 1696 detections in total. To each detection is associated the geographical location of the latter (latitude and longitude), the estimated quantity of methane, the uncertainty of the estimation (in %), the site ID, the type of site, the infrastructure and the operator at the origin of this emission. The number of sites involved in PermianMap detections is known for each period of the campaign and can be obtained by counting the IDs of these sites. However, the detections obtained during the second period can come from already visited sites during the first campaign. Site ID of both the campaign are sequence of letter and number starting by "S2" for the summer campaign and "F2" for the fall campaign. At this stage, the initial number of visited site is not possible to determine by considering campaign together. Since we have no uncertainty in the determination of the location of methane plumes detected by satellite, we will propose an automatic association of plumes only at the site level and not at the infrastructure level which, due to their fine spatial granularity, require more precision. It should also be noted that the absence of uncertainty measurements implies that a certain degree of unknown uncertainty is present in the association of our detections to sites.

2.3 Objective

The objective of the O&GProfile method is to provide an emission profile for each site but also to make it possible to associate satellite detections with the type of site and operator from which they come. Thus, all satellite detections can subsequently be studied independently of all other detections. In our case, the satellite X detections are initially not associated with the type of site or operator from which they come. PermianMAP detections are initially associated

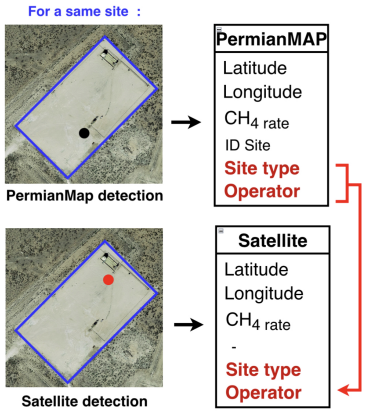


Fig. 5. Representation of the transfer of information from PermianMAP to satellite X for a same site

with the type of site and operators, so they will be used to label the satellite X data. However, in order for the PermianMAP information to be transferred to those of satellite X, it is necessary that they are on the same site in order to share the type of site and operator as illustrated in Figure 5. Once satellite X detections will be labelled, it will be possible to study the emission profile of a site from PermianMAP detections alone and/or from the combination of PermianMAP and satellite X detections. In the same way it will be possible to draw up an emission profile by type of site and operator.

3 Automated association of satellite X detection to O&G site type and operator

In order to obtain emission profiles by site and operator composed of satellite X and PermianMap quantified detection, we propose the O&GProfile which is mainly based on the use of the DBSCAN clustering algorithm[6]. The O&GProfile consists of 4 main steps:

1. Selection of satellite X data included in the PermianMAP data study perimeter selection of satellite X data included in the PermianMAP data study perimeter;
2. clustering of satellite X and PermianMap detections (DBSCAN);
3. Verification and correction of the obtained clusters/sites using Voronoï diagram and Nearest Neighboor KDtree;
4. Transfer of information from PermianMap detections to satellite X detections when PermianMap and satellite X data are located on the same site.

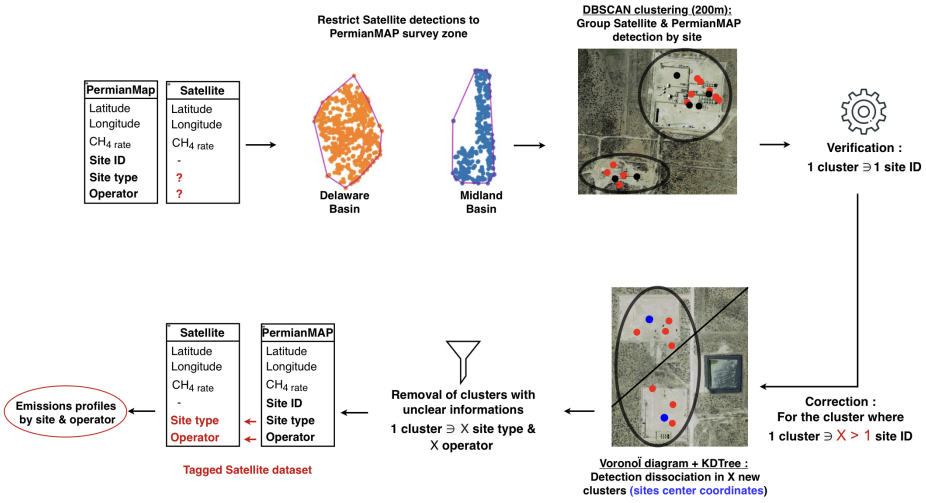


Fig. 6. O&GProfile method steps

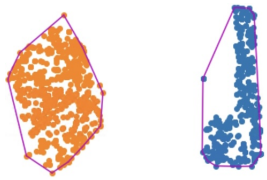


Fig. 7. Convex boundaries around 2 basins of PermianMAP study

3.1 Matching satellite X detection to PermianMAP survey zone

In order to benefit from PermianMap information, satellite X detections must be included in the PermianMap study area. The PermianMap study area is divided into two sub-basins, the Delaware basin and the Midland basin. In order to obtain the precise perimeter of each of them, the PermianMAP detections of the two basins were first separated via the use of the K-means[9] clustering algorithm.

The K-means clustering algorithm computes the centroïds and iterates until we it finds optimal centroïd. It is a parametric method which requires to know in advance the number of cluster k. In our case, the parameter k is fixed to k=2 in order to spatially dissociate 2 zones/basin (cluster 0 and 1). the detection points are then assigned to a cluster in such a manner that the sum of the squared distance between the data points and centroïd would be minimum. As output, each of the PermianMAP detection is now linked to one of the basin. For each of the two basins - which consist in two ensemble of detection - the ConvexHull function was applied to obtain convex polygon boundaries of each one. Finally,

the geographical coordinates of the satellite X detections were filtered according to their belonging or not to each of the two basins boundaries.

3.2 Automated grouping of PermianMAP and satellite X detections by site

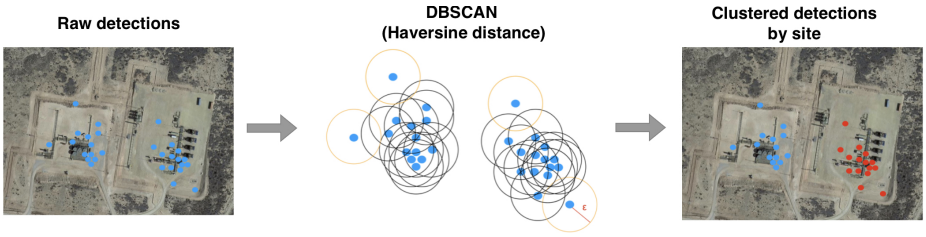


Fig. 8. O&GProfile : DBSCAN clustering with harvesine distance

Concerning satellite X detections the number of sites involved is completely unknown (as no sites are referenced in these data), so an initial number of sites from satellite X detection cannot be determined in advance. In order to obtain PermianMAP and satellite X detection grouped by site without a predetermined number of sites, spatial clustering method can be applied. The Density-based spatial clustering of applications with noise (DBSCAN) will be at the base of the O&GProfile method.

Contrary to other clustering methods (e.g. K-mean) DBSCAN is able to form clusters of arbitrary shape (not restricted to a spherical or convex shape), and can detect outliers, which makes it more suitable for our case where the data (detections) are spatially irregularly distributed over the sites. However, our study does not require the recognition and deletion of outliers, because each methane detection in our database represents important information about the quantity of methane emitted at a time t. DBSCAN works on the assumption that clusters are dense regions in space separated by regions of lower density. Indeed, by looking at the local density of the detections in large spatial dataset, DBSCAN groups densely grouped detections into a single cluster. The DBSCAN algorithm uses two main parameters:

- $minPts$: The minimal number of points to be present within the radius ε for the area to be considered "high density".
- Eps ε: A distance measure that will be used to locate the points in the neighborhood of any point.

The input data for the DBSCAN algorithm are the geographic coordinates of each detection (PermianMAP and satellite X): latitude and longitude. As used distance is geographic distance, the metric parameter used should be based on

the haversine formula which determines the great-circle distance between two points on a sphere given their longitudes and latitudes:

$$haversin\left(\frac{d}{r}\right) = haversin(\phi 2 - \phi 1) + cos(\phi 1)cos(\phi 2)haversin(\lambda 2 - \lambda 1)) \quad (1)$$

where r as radius of earth, d as the distance between two points, $\phi 1$, $\phi 2$ is latitude of two points and $\lambda 1$, $\lambda 2$ is longitude of two points respectively

With the aim of grouping these detections by site, the Eps parameter will therefore be set at 200m which is the general length and width of an O&G site. However, it will be necessary to fix the kilometers per radian fixed at $kms_per_rad = 6371.0088$ and to convert the 200m to radians: $0.2/kms_per_rad$.

The $MinPts$ is fixed at 1 a way to avoid that some detection can be considered by noise and then discard.

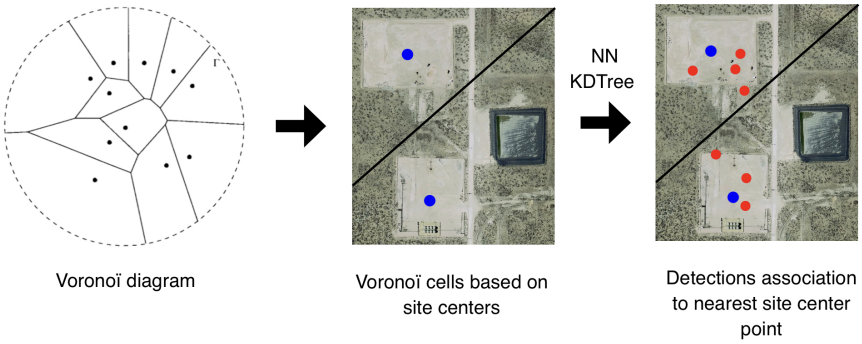


Fig. 9. Two examples of detections (red dots) forming a single cluster spread over two sites centers (blue dots)

Verification & correction : Voronoï diagram and KDtree algorithm Recall here that to transfer PermianMAP site type and operator information to satellite X detections it is required that they are positioned on the same site (cf Figure 10). Therefore, clusters comprising PermianMap and satellite X measurements will be targeted in this step, and it will be verified if each of these clusters corresponds to an O&G site.

The verification process consists in counting site IDs (from PermianMap detections) by cluster, 3 scenarios are possible here:

– If an ID is found in several clusters, it means that the Esp is too small and that a site is divided between several clusters as represented on the left image of the Figure10;



Fig. 10. Example of clustering errors: (left) *Esp* too small, site break in two clusters (right) *Esp* too big, 2 sites grouped in one cluster

- If a cluster contains several ID sites "S2" and/or "F2" it means that the *Esp* fixed at 200m is too large and that the cluster gathers several sites as illustrated on the right image of the Figure10;
- In the case where a cluster contains 1 unique ID then the cluster represents a site.

The last point represent the desired output, but in the case where the two others situations describe on the Figure 10 are identified then a correction process is require in order to correctly attribute detections to one site by cluster.

The correction procedure is based on the use of a Voronoï[3] diagram and the Nearest Neighboor (NN)[5] KDtree algorithm. The Voronoï diagram is a system created by scattering points at random on a Euclidean plane. The plane is then divided into tessellating polygons, known as cells (one around each point) consisting of the region of the plane closest to that point. The segments in a Voronoï Tessellation correspond to all points in the plane equidistant to the two nearest sites. The Voronoï diagram, taking as input the latitude and the longitude of the centers of sites established previously, permits to represent the possible division of the space according to the centers of site as represented on the Figure 9. The second step of the correction consists in linking the detections of each cluster to the centers of the two closest sites in order to constitute two new clusters for each site as illustrated on the last image of Figure 9. For this purpose, the nearest neighbor (NN) algorithm with the multi-dimensional search algorithm K-dimensional tree (KDTree) is used. The NN problem formulated here is to find the set of detections P_c in a data set P that are "closest" to a query point q (site center), as measured by a distance function $d(p, q)$. The KDTree is used partitioning spatial points using a tree-like data structure that is useful for searching for the nearest point or nearest neighbors to a given point.

4 Results

After the matching process the conserved satellite X detections belonging to the PermianMap study area are 103 of them spread over the two basins. These are added to the PermianMap detections and constitute a total dataset of 1799

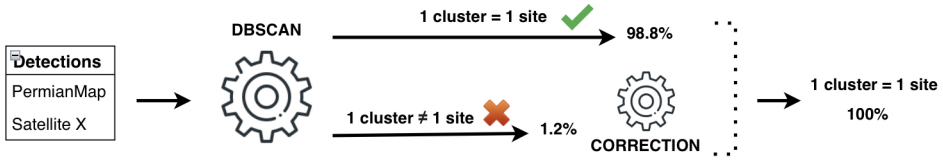


Fig. 11. O&GProfile results in % of correct association for each step

detections. The location (latitude and longitude) of each this detection database is then insert as input to the DBSCAN algorithm which permits to obtain 879 clusters. Each of these cluster is supposed to represent a single O&G site. Among these 879 clusters, there are :

- 793 cluster only containing PermianMAP detections;
- 32 cluster containing PermianMAP and satellite X detections;
- 54 clusters only containing satellite X detections.

Since satellite X detections must be co-located with PermianMAP data in order to be labeled and included in the study, the 54 clusters with only satellite X detections will be removed from the study. It will be considered therefore that 825 clusters (783 + 32).

After the verification procedure, the case where the fixed *Esp* does not seem to be too small (left picture of the Figure 10) was not found On the other hand, the second case where the *Esp* would be too large was detected on 21 clusters so 2,5% (21/825) of the initial clusters (right picture of Figure 10). In order to understand the origin of these errors, the 21 clusters were visually and manually inspected revealing that 11 of them (1,3% (11/825) of the initial cluster number) have anomalies related to the site IDs and 10 others (1,2%) directly related to the established *Eps*.

These 11 clusters with detected anomalies contains each several site IDs for a same and unique site. This error of annotation from the PermianMAP database implied that several IDs are attributed to each of these 11 sites. These 11 clusters will not be submitted to the correction process has the error not comes from the DBSCAN clustering. Concerning the 10 other clusters containing 2 sites each , a correction must be applied in order to dissociate the intra-cluster sites. The itinerary of this procedure starts with the extraction (manually) of the coordinates (latitude and longitude) of the centers of each site that the clusters contain. At the end of the correction process 10 anomalous clusters were replaced by 20 new clusters where each one represents a single site. The number of clusters(=sites) after this process is 835 (825-10+20).

Once the detections are correctly distributed by site, the membership of a single operator per site must also be checked. Note here that in the PermianMap study the manual association of sites to operators contains a certain amount of uncertainty that can lead to association errors. After verification, 12 clusters

(1,4% (12/835) of the initial clusters) contain multiple operators, only populated from the PermianMAP measure. These last one have been deleted not to introduce confusion by the presence of many operator in only one site which is not possible. The final number of obtain cluster is of 823 clusters composed of 32 sites containing PermianMap and satellite X detections and 791 sites only containing PermianMap detections.

As summarize in the Figure 11, the DBSCAN algorithm allowed the grouping of 835 clusters where 1.2% (10/835) of these clusters presented an error and the other 98.8% were correctly grouped. And in order to obtain a correct grouping on the totality of the cases the correction process to rectify the 1,2% of errors allowing to have a cluster associated with a single site in 100% of cases.

Table 1. PermianMap data : number of plume and average CH4 k/hr by site type

PermianMap

Site type	**Number of plume**	**Avg CH4 (kg/h)**
Production	446	372.04
Gathering & Boosting	494	423.06
Processing	74	514.09

Table 2. Satellite X data : number of plume and average CH4 k/hr by site type

Satellite X

Site type	**Number of plume**	**Avg CH4 (kg/h)**
Production	8	712.45
Gathering & Boosting	13	2061.98
Processing	12	1285.84

Thanks to the O&GProfile method, the raw satellite X detections can now be interpreted independently (Table 2 of the PermianMap data (Table 1 according to the type of site and the operator at their origin. For sites/operators where both satellite and PermianMap measurements are available, the detections data can be gathered to form a time series concerning the quantity of methane emissions in time for a precise site/operator (=site emission profile). It is also possible to directly compare the emissions quantities levels of PermianMap and satellite X instruments.

5 Conclusion

The O&GProfile, an automatic method based on clustering methods, is the first method which allow the grouping of detections by site for the study of emission profiles by type of site and operators. It also allows to automatically label raw detections in order to independently study and compare detections from different instruments. The O&GProfile has been tested in the Permian Basin (specifically the Midland and Delaware Basins) with labeled detections from the EDF PermianMAP project (CarbonMapper and GAO) and unlabeled satellite X detections. The association of these detections to sites via the use of the DBSCAN algorithm has a success rate of 98.8%. The 1.2% of errors could be corrected with the Vornoï-NN-KDTree algorithm, resulting in a success rate of 100%. Thanks to the O&GProfile, emission profiles for the period 2021-2022 can be obtained from 823 different O&G sites. Each site corresponds to a type of site (Production, Gathering & Boosting, Processing) it is also possible to establish emission profiles by site. This is also applicable for the operators of each site. The O&GProfile has also allowed the labeling of 103 satellite X detections divided into 33 clusters, which allows for independent analysis of satellite X data and/or comparison with PermianMAP data. The O&GP profile can be applied in study cases where the sources of unlabeled detections are unlimited, while the ground infrastructure data (type of site and operators) can also be easily replaced by others data (most complete and/or in other regions).The process of correction still contain a part of visual and manual inspection concerning for 1,2% of the case, where site have to be distinct by extracting their middle geographic position. This step could be replace in future studies by automated method able to dissociated site thank an image analysis and automatically extract for each of them their middle geographic position. Another possible and direct improvement of our method would be to provide an emission profile at the granularity of on-site infrastructure. In order to be able to link a detection to an on-site infrastructure in an accurate way it is necessary to have access not only to the geographical coordinates of the detection placement but also to the degree of uncertainty related to the determination of this location. Indeed, the space between two type of infrastructure could be very thin (e.g. less than 10m) and to associated a detection to one of these infrastructure, the uncertainty of the geographic location of the detection should be determined. Indeed the geographic position of a detection is sometimes not representative of the exact geographic position of the source of the detected plume. Point source satellite like satellite X permits to obtain a good precision concerning plume geographic localization (sufficient for site association), but as the uncertainty around the geographic positions of the detection is not know, there is a risk to try to associated them to a precise infrastructure without this information. A possible solution could be to determine a circle of uncertainty around each detection and determine which infrastructure has he highest probability to be at the origin of the detection.

References

1. https://carbonmapper.org
2. Permianmap, www.permianmap.org
3. Aurenhammer, F., Klein, R.: Chapter 5 - voronoi diagrams**partially supported by the deutsche forschungsgemeinschaft, grant k1 655 2-2. In: Sack, J.R., Urrutia, J. (eds.) Handbook of Computational Geometry, pp. 201–290. North-Holland, Amsterdam (2000). https://doi.org/https://doi.org/10.1016/B978-044482537-7/50006-1, https://www.sciencedirect.com/science/article/pii/B9780444825377500061
4. Cusworth, D.H., Duren, R.M., Thorpe, A.K., Olson-Duvall, W., Heckler, J., Chapman, J.W., Eastwood, M.L., Helmlinger, M.C., Green, R.O., Asner, G.P., Dennison, P.E., Miller, C.E.: Intermittency of large methane emitters in the permian basin. Environmental Science & Technology Letters **8**(7), 567–573 (2021). https://doi.org/10.1021/acs.estlett.1c00173, https://doi.org/10.1021/acs.estlett.1c00173
5. Dixon, P.M.: Nearest neighbor methods. Encyclopedia of environmetrics **3**, 1370–1383 (2002)
6. Ester, M., Kriegel, H.P., Sander, J., Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. p. 226–231. KDD'96, AAAI Press (1996)
7. Irakulis-Loitxate, I., Guanter, L., Liu, Y.N., Varon, D.J., Maasakkers, J.D., Zhang, Y., Chulakadabba, A., Wofsy, S.C., Thorpe, A.K., Duren, R.M., Frankenberg, C., Lyon, D.R., Hmiel, B., Cusworth, D.H., Zhang, Y., Segl, K., Gorroño, J., Sánchez-García, E., Sulprizio, M.P., Cao, K., Zhu, H., Liang, J., Li, X., Aben, I., Jacob, D.J.: Satellite-based survey of extreme methane emissions in the permian basin. Science Advances **7**(27), eabf4507 (2021). https://doi.org/10.1126/sciadv.abf4507, https://www.science.org/doi/abs/10.1126/sciadv.abf4507
8. Jacob, D.J., Varon, D.J., Cusworth, D.H., Dennison, P.E., Frankenberg, C., Gautam, R., Guanter, L., Kelley, J., McKeever, J., Ott, L.E., Poulter, B., Qu, Z., Thorpe, A.K., Worden, J.R., Duren, R.M.: Quantifying methane emissions from the global scale down to point sources using satellite observations of atmospheric methane. Atmospheric Chemistry and Physics Discussions **2022**, 1–44 (2022). https://doi.org/10.5194/acp-2022-246, https://acp.copernicus.org/preprints/acp-2022-246/
9. Jin, X., Han, J.: K-Means Clustering, pp. 563–564. Springer US, Boston, MA (2010). https://doi.org/10.1007/978-0-387-30164-8_425, https://doi.org/10.1007/978-0-387-30164-8_425
10. Schuit, B.J., Maasakkers, J.D., Bijl, P., Mahapatra, G., Van den Berg, A.W., Pandey, S., Lorente, A., Borsdorff, T., Houweling, S., Varon, D.J., McKeever, J., Jervis, D., Girard, M., Irakulis-Loitxate, I., Gorroño, J., Guanter, L., Cusworth, D.H., Aben, I.: Automated detection and monitoring of methane super-emitters using satellite data. Atmospheric Chemistry and Physics Discussions **2023**, 1–47 (2023). https://doi.org/10.5194/acp-2022-862, https://acp.copernicus.org/preprints/acp-2022-862/
11. Zhang, Y., Gautam, R., Pandey, S., Omara, M., Maasakkers, J.D., Sadavarte, P., Lyon, D., Nesser, H., Sulprizio, M.P., Varon, D.J., Zhang, R., Houweling, S., Zavala-Araiza, D., Alvarez, R.A., Lorente, A., Hamburg, S.P., Aben, I., Jacob, D.J.: Quantifying methane emissions from the largest oil-producing basin in the united states from space. Science Advances **6**(17), eaaz5120 (2020). https://doi.org/10.1126/sciadv.aaz5120, https://www.science.org/doi/abs/10.1126/sciadv.aaz5120