Validation and deployment of greenhouse gas sensors (CH4) for the characterization of industrial and urban sources

Rodrigo Andres Rivera Martinez

université PARIS-SACLAY

Validation and deployment of greenhouse gas sensors (CH_4) for the characterization of industrial and urban sources

Validation et déploiement de capteurs de gaz à effet de serre (CH_4) pour la caractérisation de sources industrielles et urbaines

Thèse de doctorat de l'université Paris-Saclay

École doctorale n°129 Sciences de l'environnement d'Ile-de-France (SEIF)
Spécialité de doctorat: Sciences du climat, de l'atmosphère et des océans, terrestres et planétaire
Graduate School : Géosciences, climat, environnement et planètes
Référent : Université de Versailles-Saint-Quentin-en-Yvelines

Thèse préparée dans l'unité de recherche LSCE (Université Paris Saclay, CEA, CNRS, UVSQ), sous la direction de **Philippe CIAIS** (Directeur de recherche), le co-encadrement de **Olivier LAURENT** (Ingénieur de recherche), **Gregoire BROQUET** (Chargé de recherche), **Cécile MALLET** (Professeure des universités) et **Diego SANTAREN** (Chercheur).

Thèse présentée et soutenue à Gif-sur-Yvette, le 10 octobre 2022, par

Rodrigo Andres RIVERA MARTINEZ

Composition du jury

Membres du jury avec voix délibérative

Lilian JOLY Professeur des universités, Université de Reims Champagne-Ardenne	Président & Rapporteur
Jean SCIARE Professeur des universités, Cyprus Institute (CyI)	Rapporteur & Examinateur
Sabine CRUNAIRE Maîtresse de conférences, IMT Nord Europe	Examinatrice
Martina SCHMIDT Directrice de recherche, Universität Heidelberg	Examinatrice

THESE DE DOCTORAT

NNT : 2022UPASJ020

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences de l'Environnement d'Île-de-France (SEIF)

Titre: Validation et déploiement de capteurs de gaz à effet de serre (CH_4) pour la caractérisation de sources industrielles et urbaines

Mots clés: Méthane, Flux d'émission, Capteur «low cost», anthropique, Étalonnage, apprentissage automatique

Résumé: Cette thèse a pour objectif la reconstruction des concentrations de méthane (CH_4) atmosphérique à partir des signaux bruts (tension électrique) de capteurs d'oxyde de métal (MOS) à bas coût afin de surveiller les émissions de CH_4 sur des sites industriels. Notre stratégie se base essentiellement sur la colocalisation des mesures des capteurs à bas coût avec des instruments de référence de haute précision. Les données de concentration de CH_4 des instruments de haute précision sont utilisées comme variable cible pour être reconstruites à partir des mesures brutes de tension électrique issue des capteurs MOS et d'autres prédicteurs. La reconstruction est accomplie avec des modèles d'apprentissage automatique et paramétriques. Il est appliqué aux problèmes très complexes dont la reconstruction des : 1) faibles variations de CH_4 dans l'air ambiant du laboratoire, 2) variations rapides et de grandes amplitudes de CH_4 sous forme de pics générés artificiellement sur un bac d'essai en laboratoire reproduisant des signaux typiquement observés lors de fuites réelles sur le terrain, et 3) variations rapides et de grandes amplitudes de CH_4 mesurées sur le terrain durant une expérience de largage contrôlé de CH_4 dans l'atmosphère. La première expérience de laboratoire a révélé une forte influence de la vapeur d'eau sur les mesures des capteurs MOS, ainsi que les bonnes performances du modèle de perceptron multicouches (MLP) pour reconstruire le signal CH_4 à partir du signal brut du capteur (Chapitre 2). Les résultats de la deuxième expérience (Chapitre 3) ont montré l'importance du choix du type de capteur pour reconstruire le signal de concentration CH_4 avec une précision de 1 ppm (RMSE). Nous avons ainsi mis en avant que les capteurs TGS 2611-C00 ont permis une reconstruction plus précise des pics de CH_4 que les capteurs TGS 2611-E00. Une stratégie parcimonieuse d'apprentissage du modèle nous a permis de restreindre l'ensemble des données nécessaires à l'apprentissage de 70% à 25% des données totales sans dégradation des performances pour la reconstruction de pics de CH_4. Les performances sur les estimations des flux d'émissions de CH_4 lors des largages contrôlés dans l'atmosphère en utilisant les concentrations de CH_4 reconstruites à partir des mesures issus des capteurs MOS et une inversion avec un modèle gaussien (Chapitre 4) ont été similaires à celles obtenues à partir des données des instruments de référence à haute précision, avec une erreur moyenne d'estimation des flux d'émission de 25% sur 11 largages et une erreur moyenne de localisation des sources d'émission de 9.5 m. Les résultats de cette thèse constituent une base pour explorer les techniques avancées d'apprentissage automatique pour la reconstruction du signal de concentrations de CH_4 à partir du signal brut des capteurs MOS et étudier les informations requises par les modèles d'inversion afin d'offrir les meilleures estimations des flux d'émission et localisation des fuites de CH_4.

université PARIS-SACLAY
ÉCOLE DOCTORALE
Sciences de l'Environnement d'Île-de-France (SEIF)

Title: Validation and deployment of greenhouse gas sensors (CH_4) for the characterization of industrial and urban sources

Keywords: Methane, emission rate, Low cost sensors, anthropogenic, calibration, machine learning

Abstract: This PhD aims at reconstructing atmospheric CH_4 concentrations from voltage signals recorded by low-cost Metal Oxide Sensors (MOS) for the purpose of monitoring CH_4 leaks from an industrial facility. Our strategy is based on the collocation of measurements by low-cost sensors and expensive very high accuracy instruments. The CH_4 concentrations data from the high accuracy reference instrument are then used as a target variable to be reconstructed using MOS voltage and other predictors. The reconstruction is performed with both machine learning models and parametric models. It is applied to increasingly complex problems, including the reconstruction of: 1) smooth variations of ambient room air CH_4 in a laboratory, 2) fast and large variations of CH_4 during concentration spikes generated on a test bench in a laboratory with a typical signature similar to that observed for real world leaks, and 3) fast and large variations of CH_4 observed in the field during a dedicated controlled release experiment. The first laboratory experiments revealed a high influence of H2O in TGS sensors, and the reconstruction with a Multilayer perceptron model showed good agreement between the references and the reconstructed room air concentrations (Chapter 2). The results of the second experiment (Chapter 3) showed the importance of the selection of the sensor's type in order to produce reconstructions of CH_4 concentration with a 1 ppm accuracy (RMSE). We also discovered that the TGS 2611-C00 sensors provided a more accurate reconstruction of CH_4 spikes than the TGS 2611-E00 ones. A parsimonious model training strategy allowed us to restrict the training set from 70% to 25% of the data without a degradation in the reconstruction of the CH_4 spikes. Emission of controlled releases from reconstructed CH_4 concentrations measured by TGS sensors and the inversion of a Gaussian atmospheric model (Chapter 4) were similar to those inverted from high accuracy reference data, with an average emission rate estimation error of 25% over 11 controlled releases and a location error of 9.5 m. The results of this PhD are the basis to explore advanced techniques based on machine learning to produce accurate estimates of CH_4 concentrations, and study the information required by inversion modelling that produced best estimates of the emission rate and location of CH_4 leaks.

Résumé

Le Méthane (CH_4) est un gaz à effet de serre avec une capacité de réchauffement 28 fois plus puissant que le CO_2 dans une échelle du temps de 100 ans. La concentration de CH_4 dans l'atmosphère a incrémenté 2.6 fois depuis l'ère pré-industrielle (700 ppb) jusqu'à nos jours (1866 ppb). Le principal facteur qui contribue à l'incrément du CH_4 dans l'atmosphère est l'activité humaine (plus de 60% des émissions globales). Dans le secteur de la production et de l'utilisation des combustibles fossiles, 63% des émissions sont liées à la production de gaz et de pétrole. Des fuites fugitives de CH_4 se produisent dans toute la chaîne de production de gaz et de pétrole, depuis l'extraction, le transport, le stockage et la distribution. Pourtant, les fuites de méthane non détectées pourraient produire des émissions importantes aux taux élevés. Il existe aussi une grande variabilité, temporelle et spatiale, dans les estimations d'émissions dans les installations de gaz et de pétrole ou dans des sites de déchets. Il n'y a pas d'informations sur les émissions pour ces installations dans les inventaires et les facteurs d'émission sont difficiles d'appliquer. C'est la raison pour laquelle une étude lieu par lieu est nécessaire. Les estimations des émissions dans ces sites peuvent être améliorées à partir des mesures continues des concentrations de CH_4. Une alternative pour produire des mesures continues est le déploiement d'un réseau de stations à point fixe dans ces sites. Cependant, le déploiement d'un tel réseau avec des instruments d'haute précision, comme le CRDS (Cavity ring down spectrometers), est limitant à cause des prix élevés, les dimensions des instruments et les besoins au niveau d'énergie pour son correct fonctionnement. L'utilisation des capteurs à bas cout peut permettre de surmonter ces limitations. Ma thèse se focalise dans l'étude des capteurs d'oxide de métal (MOS) du fabricant Figaro® et je m'intéresse à déterminer une stratégie d'étalonnage pour trois versions de ces capteurs MOS et de fournir des estimations fiables des émissions de CH_4 lors d'un déploiement sur le terrain.

Trois expériences ont été conçus pour atteindre cet objectif. La première expérience a pour but de tester trois versions des capteurs MOS (TGS 2600, TGS 2611-C00 et TGS 2611-E00) dans des conditions contrôlées. Les capteurs ont été exposées à l'air d'une pièce de laboratoire pendant 47 jours où les paramètres environnementaux (température et humidité relative) ont été surveillés de façon continue. La première partie de cette étude consistait à étudier la sensitivité des capteurs à des variations de CH_4, H2O et CO. D'une part, on a pu observer l'influence importante du H2O dans les capteurs, ce qui limite la sensitivité au CH_4 ; et d'autre part, les relations non linéaires existantes avec le vapeur d'eau. La deuxième partie consistait à utiliser une approche d'apprentissage automatique (réseaux des neurones) pour reconstruire le signal de CH_4 à partir des variations de résistance mesurées avec les capteurs MOS et des informations des paramètres environnementaux comme la pression, la température, la fraction molaire d'eau et le CO. On a observé un bon accord entre le signal simulé par le modèle et la référence correspondent à des mesures prises par une capteur d'haute précision où on a pu atteindre notre objectif d'erreur (RMSE <= 0.2 ppm). Dans la deuxième expérience, on explore les capacités de deux versions de capteurs MOS (TGS 2611-C00 et TGS 2611-E00) pour détecter et quantifier des variations rapides et des grandes amplitudes de CH_4 sous forme des pics générés artificiellement reproduisant des signaux de fuites produits typiquement lors du terrain. Quatre chambres ont été développées et exposées pendant quatre mois aux pics de CH_4. Je me suis focalisé sur quatre questions dans cette expérience : i) la capacité de détection et de quantification des pics artificielles de CH_4 en comparant les deux types de capteurs sur cinq modèles différentes, paramétriques et non paramétriques (régression linéal et polynomial, random forest, hybrid random forest et multilayer perceptron) ; ii) la possibilité de réduire la taille de l'ensemble de l'apprentissage sans dégrader les performances dans la reconstruction du signal de CH_4 ; iii) l'évaluation de l'effet de l'âge dans les capteur après six mois des mesures continues d'air ambiante ; et iv) la capacité d'utiliser un modèle entrainé avec des données d'une chambre et un type de MOS sur une autre chambre avec le même type de capteur. Les résultats ont montré que les capteurs TGS 2611-C00 permettent d'avoir une meilleure reconstruction des pics de CH_4 que les capteurs TGS 2611-E00. Notre stratégie d'apprentissage parcimonieuse nous a permis de réduire l'ensemble des données

d'apprentissage de 70% à 25% sans dégradation des performances pour la reconstruction des pics de CH_4. La troisième expérience est consacrée à l'évaluation de la performance des capteurs MOS sur le terrain. Sept chambres, incluant deux versions des capteurs MOS (TGS 2611-C00 et TGS 2611-E00), ont été développées. Les chambres ont été exposées à une série des largages de CH_4 contrôlés pour simuler des fuites d'émission typiques dans des installations de pétrole et de gaz. La première partie de l'étude se focalise dans la reconstruction des plumes de CH_4 (concentrations) à partir des mesures de voltage des capteurs MOS dans un site où les paramètres environnementaux varient et où il existe aussi d'autres composantes qui peuvent affecter la capacité des capteurs à mesurer le CH_4. La deuxième partie se concentre à étudier l'estimation des flux d'émissions et la localisation de la source d'émission avec des modèles gaussiennes de dispersion de plume à partir des signaux de CH_4 reconstruites. Nos résultats ont montré une estimation de flux d'émission similaires à celles estimées en utilisant des données issues des capteurs à haute précision avec une erreur moyenne dans l'estimation de flux de 25% et une erreur moyenne dans la localisation de la source d'émission de 9.5 m sur 11 largages contrôlés. Le dernière chapitre de cette thèse se focalise dans la discussion des résultats de ces trois expériences. Je propose une discussion sur les développements futurs et les priorités de recherche des capteurs MOS pour étudier l'effet des sensitivités aux autres espèces et aux paramètres environnementaux, dans le développement des modèles d'étalonnage, lors du déploiement dans un réseaux de capteurs et pour améliorer l'estimation des flux et la localisation de sources d'émission.

To my dad Jesus, my mom Zaida,

my sister Cecilia,

and my love Blanca.

Acknowledgements

I would like to express my deepest gratitude to my PhD director and supervisors. Philippe, I cannot be more grateful for every lesson that you have taught me during these three years. You have shown me the way to be a good scientific researcher creating a positive impact on society. Olivier, thank you for your constant support through all these years. You have always had the time to answer my questions and find solutions to my (not always scientific) problems. Thank you, Cécile, for your endless support and advice, you introduced me to data science and I could not be more grateful as it has become my passion. Grégoire, thank you for all your comments, for your time reading and helping me improve all the papers. You have always brought interesting points to the discussion and helped me to become more critical of my work. Thank you, Diego, for your constant support, your passion for science and your joy which has always motivated me.

I would also like to thank Jean Sciare and Lilian Joly for agreeing to review my thesis. Many thanks to Sabine Crunaire and Martina Schmidth for being part of my Ph.D. jury. I wish to thank Laurent Barthes and Felix Voguel for the constructive and inspiring discussions during the committee meetings.

I wish to express my gratitude to the TRACE team, with a special mention to Pramod, Adil, Christopher and Ford. They have always been available to discuss my PhD among other topics. Thank you for your advice and constant support. Many thanks to the ICOS Team at LSCE, my time in the lab would not have been the same without the marvelous people working there who are always ready to help. Thank you for letting me be part of this team. Special thanks to Leo, Michel, Ivonne and Carole who were always available to answer my questions.

The PhD candidate office was always full of energy and enthusiasm thanks to Yunsong, Sara, Dro, Ivonne, Emeric and Yang. Thank you for making it an amazing place to work.

These lines are reserved for my two best friends, Jafet and Rene, thank you for always being there to cheer me on in every challenge I took.

A very warm thank you to my mother, Zaida, and to my father Jesus. Thank you for sending me on this adventure, where I found many friends and challenges, I learned a lot and I found love. Without your constant support, none of this would have happened, and please forgive me for this long absence. To my sister Cecilia, thank you for being an example to follow and showing me how to take challenges without fear. And thank you for always helping me to stand when I fall. To my love Blanca, thank you for supporting every crazy idea that crosses my mind and thank you for letting me support yours. Thank you for your patience and understanding, and sorry for the long nights working next to you.

This acknowledgment would not be complete if I do not mention my fluffy family. Thank you Khady, for taking care of mom and dad while I am abroad on this adventure. To Pandah, for the warm welcomes whenever I visit my sister. Part of this PhD thesis was written at night (during several nights) and I was always surrounded by the most incredible company. Even now that I am writing these last words, they are here resting next to me. Two of them have long ears and a fluffy little tail, Kuroka and Yuna, and the other is white as the snow and looks like a ferocious wolf but has one of the kindest hearts that I have ever known, Pancho. Thank you for bringing joy to my life.

I would like to reserve my very last words to the readers of this Thesis. Thank you for being interested in this topic. If you have learnt something from this work, if it helps you to have an idea or if it motivates you in some way, then the purpose of my work will have been fulfilled.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Energy balance, effective radiative forcing and global warming

The amount of energy received by the Earth from Sun, on a stable climate, is similar to the amount of energy emitted in the form of reflected sunlight and radiation (Forster et al., 2021). Nevertheless, anthropogenic processes affect the exchange of energy at the surface of Earth producing an imbalance in the global mean budget. Figure 1.1 presents a schematic of the best estimates of the incoming and outgoing energy at the surface of the Earth at beginning of 21st century. The estimated incoming shortwave radiation from the Sun is 340 W m^{-2}, and about 23% (80 W m^{-2}) is absorbed by the atmosphere, 29% (100 W m^{-2}) is reflected to the space by clouds, and the surface of Earth absorbs 47% (160 W m^{-2}) of the energy and it reflects $\sim$7% (25 W m^{-2}) to the space. Earth emits energy in the form of longwave radiation by evaporation (82 W m^{-2}) associated to latent heat and as sensible heat (21 W m^{-2}), which refers to exchange of heat between the surface of Earth and the atmosphere. Earth also emits infrared radiation (398 Wm^{-2}), from which 60% is emitted directly to the space and rest is absorbed by the greenhouse gases. Greenhouse gases, such as CO_2, CH_4 or H_2O, introduce feedbacks on the earth system emitting 342 Wm^{-2}, which creates a phenomenon called the "Greenhouse effect" and thus increasing the global mean temperature. On the global mean radiative budget it is observed that there is an estimated imbalance of 0.7 W m^{-2}, which leads to an increase of the mean temperature on Earth's surface.

There are different factors, called forcing agents, which allow the temperature of Earth to increase or decrease. The principal forcing agents are the solar radiation, tropospheric aerosols, greenhouse gases and alteration in land coverage (Hansen and Sato, 2001). Greenhouse gases are important for life development on Earth. Indeed, by trapping the longwave radiation emitted by Earth, it allows to raise the global mean temperature to 15°C, instead of -18°C without atmosphere (Mitchell, 1989).

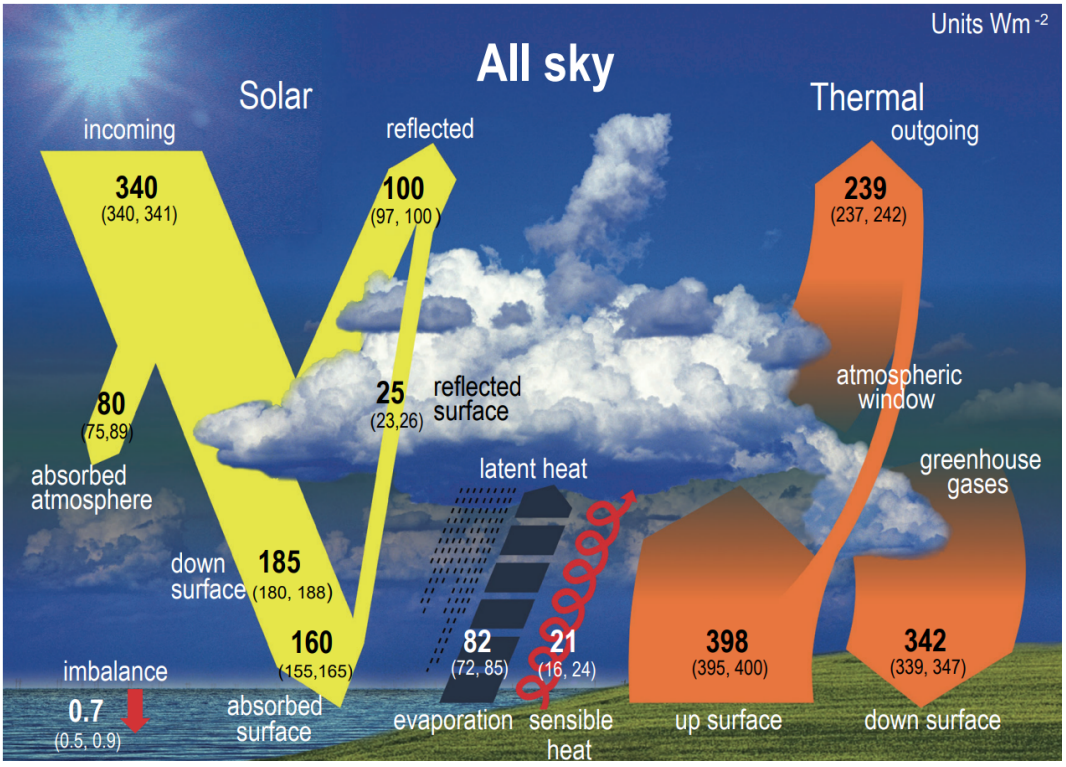


Figure 1.1: Global mean energy budget. Bold numbers are the mean magnitude estimate of the global energy balance in Wm^{-2}. In parenthesis is shown the uncertainty range (5 – 95%) (Source: Forster et al. (2021))

The Intergovernmental Panel on the Climate Change (IPCC) has assessed, since its creation in 1990, the impact of human activities on the Earth's climate and the outcomes on different case scenarios of increase in global mean surface temperature. Their latest report (6th Assessment Report) demonstrated a direct link between the anthropogenic activities and the increase in the average surface temperature on Earth (Eyring et al., 2021). Since its 5th assessment report, and following the current literature, IPCC highlights that radiative forcing represents a good predictor of the temperature change on the surface (Bellouin et al., 2020).

The radiative forcing is the change in the planetary radiation balance measured by the change in the net radiative flux, adjusted to equilibrium, in response to a perturbation. This perturbation can be due to a change in the solar radiation, the planetary surface or the composition of the atmosphere (Hansen et al., 1997). The adjustment time of the radiative flux in the troposphere after a perturbation is estimated between 2 to 3 months. The effective radiative forcing (ERF) was presented as a better predictor as it takes into account rapid adjustments. Furthermore, it considers the heating and cooling of the troposphere, and the heat change in absence of temperature variations in the ocean surface (Bellouin et al., 2020).

Figure 1.2a and 1.2b show the contribution of emitted components to the ERF and the change in the global surface temperature respectively. Amongst present greenhouse gases in the atmosphere, CO_2 and CH_4 contribute in larger proportions to increase the surface temperature. Since the pre-industrial era, CO_2 is responsible of the biggest contribution to the ERF reaching 2 W m^{-2}, which represents 1°C change in the global surface temperature. Whereas CH_4 contributes 1.2 Wm^{-2} to the ERF, which represents 0.6 °C change.

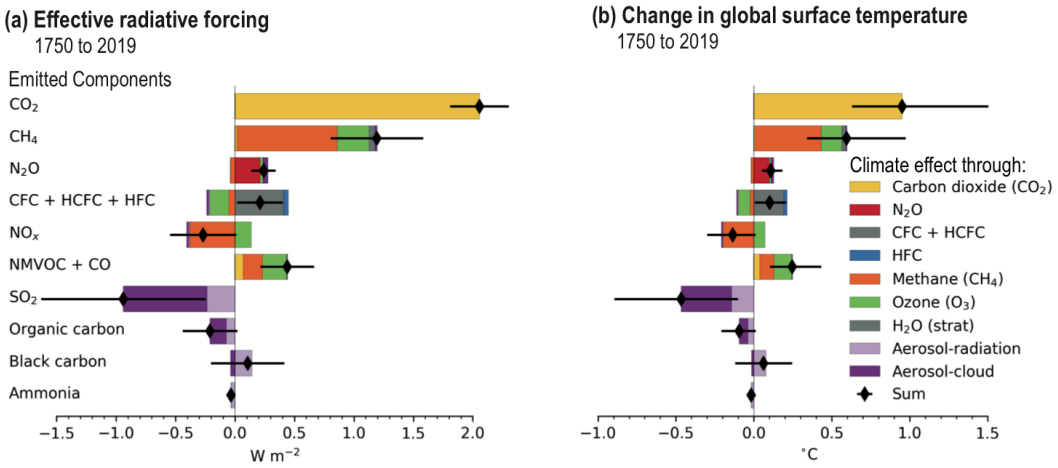


Figure 1.2: Contribution to (a) the effective radiative forcing and (b) the global surface temperature from component emissions for the period 1750 – 2019 (Source: (Arias et al., 2021))

Another metric used to evaluate the warming effects of greenhouse gases is the global warming potential (GWP). It includes information of the radiative forcing of a single pulse emission (usually 1 Kg) and its lifetime in the atmosphere, and it is compared to another forcing factor, usually CO_2 (Ramanswamy et al., 1991; Forster et al., 2021). It is estimated that CH_4 has a GWP 28

times larger than CO_2 on century time scale and a lifespan in the atmosphere of 9 years (Saunois et al., 2020). A reduction of the CH_4 anthropogenic emissions would lead to a fast stabilization of CH_4 in the atmosphere (Turner et al., 2019) and a reduction of its concentration in decades (Nisbet et al., 2020; Saunois et al., 2020).

In order to achieve adapted policies which promote the reduction of atmospheric CH_4 concentrations, it is important to understand and quantify the principal sources and sinks of CH_4 emissions at a local and global scale.

1.2 Global $\mathrm{CH_4}$ Budget

Since the pre-industrial era, CH_4 concentrations have increased 2.6 times, from 700 ppb to 1866 ppb in 2019 (Saunois et al., 2020). Ice core measurements showed that CH_4 concentrations in the atmosphere on the last millennium have been stable at around 700 ppb. Human activities have contributed to the increase in CH_4 concentrations especially from 1900 to 2000. This is due to intensive agriculture and a larger use of fossil fuels. Then, a stabilization period occurs during 7 years, with CH_4 concentrations around 1840 ppb. Finally, from 2007 to nowadays, CH_4 concentrations in the atmosphere have continuously increased, with an increase rate of 6 ppb/year at the end of 2007, and of 10 ppb/year in 2014. (Turner et al., 2019; Nisbet et al., 2020)

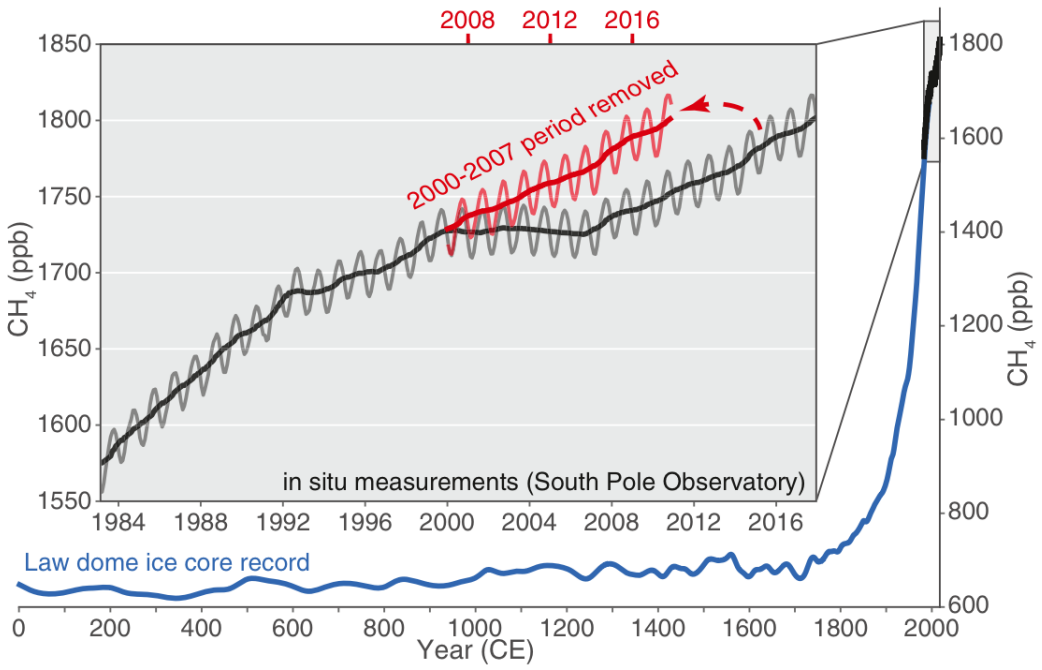


Figure 1.3: Atmospheric CH_4 concentrations observed for the past 2000 years. The inset plot corresponds to the period from 1984 to 2019 (black line). Removing the stabilization period of 7 years (red line of the inset plot), it is appreciated a continuous growth of CH_4 concentrations in the last decades. (Source: (Turner et al., 2019))

Figure 1.4 summarizes the global CH_4 budget showing the sources and sinks by category and the total estimates. The global CH_4 budget can give us an overview of the emissions estimate by sources and sinks. This budget can be estimated from Top-Down (TD) or Bottom-up (BU) approaches. Top-Down emission estimates are obtained by combining information of concentration measurements from a network of stations, and information of atmospheric transport of gases from each region. The Bottom-up emission estimates come from inventory datasets, although there are differences on detail level among datasets due to the information reported by each country. (Saunois et al., 2020; Turner et al., 2019)

1.2.1 Sources and sinks

Methane emissions can be from either natural or anthropogenic origin, and they are related to three main processes: i) biogenic: it is the result of decomposition of organic matter by methanogenic *Archaea*, which occurs in anaerobic environments such as wetlands, landfill sites, wastewater, ru-

minants, etc. ii) Thermogenic: it is originated on geological timescales by breakdown of buried organic matter under heat and pressure in Earth's crust. CH_4 and other gases reach the atmosphere through gas seeps in the sea or earth, or through human activities related to the exploitation and distribution of fossil fuels. iii) Pyrogenic: it is the product of incomplete combustion of biomass and organic matter. The largest sources of pyrogenic methane are wildfires, peat fires, biofuel and biomass burning of deforested/degrade areas. (Saunois et al., 2020)

In the last synthesis conducted by Saunois et al. (2020), it was reported that global methane emissions on the period between 2008 and 2017 account for 576 Tg CH_4 yr^{-1} (550 – 594 Tg CH_4 yr^{-1}, min - max respectively) from top-down estimates; and 737 Tg CH_4 yr^{-1} (594 – 881 Tg CH_4 yr^{-1}) from bottom-up estimates. Bottom-up and top-down estimates present a discrepancy of 30%. The overestimation of BU emissions are related to up-scaling local measurements and in some cases, double counting natural sources (Saunois et al., 2020). Figure 1.4 summarizes the global CH_4 budget showing the sources and sinks by category and the total estimates.

1.2.2 CH_4 sinks

The total mean loss of CH_4 in the atmosphere is estimated to 625 Tg CH_4 yr^{-1} [500 – 798, min – max in Tg CH_4 yr^{-1} respectively](BU) and 556 Tg CH_4 yr^{-1} [501 - 574](TD), and it can be regrouped in four categories. BU chemical sinks are estimated using the Chemistry Climate Model Initiative (CCMI) project (Morgenstern et al., 2017):

Tropospheric oxidation by hydroxyl radical (OH): it is the most important sink, representing 90% of the sink mechanism. The presence of OH in the atmosphere could give an indication of the CH_4 loss, although it is not directly linked (Saunois et al., 2020). Indeed, climate variability, biomass burning, and anthropogenic activities can alter the OH concentrations (Turner et al., 2018; Saunois et al., 2020). Saunois et al. (2020) reported an estimated loss of 553 Tg CH_4 yr^{-1} (BU) linked to OH oxidation.

Stratospheric loss: CH_4 is a reactive gas and its interactions with exited atoms of oxygen, atomic chlorine (Cl), atomic fluorine (F) and OH produce a loss of CH_4 in the stratosphere. An estimate of 31 Tg CH_4 yr^{-1} of chemical loss is reported for the period 2000 – 2010 (Saunois et al.,

2020).

Tropospheric reaction with atomic chlorine (Cl): it is a seasonally varying sink in the marine boundary layer which produces Cl after a reaction between NaCl and NO_2 and UV interactions. This reaction would be responsible of CH_4 loss. Despite the large amount of uncertainties in the quantification of Cl, Saunois et al. (2020) reported an estimated CH_4 loss of about 11 Tg CH_4 yr^{-1}.

Soil uptake: methanotrophic bacteria can consume methane as a source of energy representing a sink of methane in unsaturated oxic soils. The estimated CH_4 loss is of 30 Tg CH_4 yr^{-1} for periods 2000 – 2009 and 2008 – 2017 (Saunois et al., 2020).

1.2.3 CH_4 Sources

1.2.3.1 Natural sources

Organic matter decomposition in water-saturated or flooded ecosystems produces anaerobic conditions and methane. A combination of the following three processes can lead to CH_4 reaching the atmosphere: diffusive loss of dissolved CH_4 across the air-water boundary, ebullition flux from sediments, and plant transport (Saunois et al., 2020). Therefore, natural sources of methane include wetlands and water systems, land geological sources, wild animals and oceanic sources.

Wetlands are ecosystems with saturated water, or flooded soils or peats where anaerobic conditions can lead to methane production. Estimated emissions from Wetlands are of 149 Tg CH_4 yr^{-1} on the decade 2008 – 2017 (Saunois et al., 2020).

Inland water systems are referred to freshwater systems, like lakes, ponds or rivers on which the main CH_4 fluxes, between 50 to 90%, are attributed to ebullition and plant fluxes, specially to lakes and ponds. An estimation of 159 Tg CH_4 yr^{-1} is reported for freshwater systems(Saunois et al., 2020).

Geological sources refer to CH_4 originated from Earth's crust that migrate to the atmosphere by tectonic faults and fractured rocks. Global emissions are reported to be 45 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Wild animals refer to wild ruminants that produce CH_4 through microbial fermentation in

their rumen. Their contribution to the methane budget is estimated to be around 2 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Termites are also a natural source of methane, which is produced by anaerobic decomposition of biomass in their gut. The decadal value of CH_4 emissions is estimated to be 9 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Oceanic sources include coastal and open ocean methane releases. Some possible sources of methane production come from marine sediments, water column, leaks from geological marine seepage or emissions from destabilization of marine hydrates. The average emission estimates are 13 Tg CH_4 yr^{-1} for oceanic sources (Saunois et al., 2020).

Thawing of permafrost could generate direct and indirect emissions. Direct emissions occur when trapped CH_4 in the ice is released to the atmosphere accounting for less than 1 Tg CH_4 yr^{-1} in the global budget. Indirect emissions occur on three processes: i) when organic matter contained in the ice is released; ii) changes on the land surface hydrology; and iii) formation of thermokarst lakes. Indirect emissions are difficult to estimate and they overlap with wetland and freshwater emissions.

Vegetation could lead to three pathways of CH_4 production: i) plants produce CH_4 through an abiotic photochemical process induced by stress; ii) plants act as 'straws' releasing methane from soils; and iii) stems of living and dead wood are a suitable environment for microbial methanogenesis. These emissions are potentially larger but it remains uncertain or they are included in other flux categories.

1.2.3.2 Anthropogenic emissions

Agriculture and waste management

Agricultural and waste sectors accounts for 56% of the total anthropogenic emissions with an estimated emission of 206 Tg CH_4 yr^{-1} (191 – 223 Tg CH_4 yr^{-1}, min - max respectively). This category includes emissions linked to livestock production, rice cultivation, landfills and waste handling.

Like wild ruminants, domestic ruminants produce methane by microbial activity in their diges-

tive system. Livestock manure management is also related to the production of CH_4 in anaerobic conditions (liquid treated lagoons, ponds, tanks, etc.). Manure decomposition would lead to CH_4 production. However, if it is handled as a solid, in stacks or dry lots, there is little to no production of CH_4. Global emissions from livestock production are estimated to be 111 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Regarding rice cultivation, the most relevant factor which induces the production of CH_4 is the water management system, since rice cultivation is commonly done in flooded paddy fields where the environment is ideal for CH_4 production. The use of fertilizers such as urea, soil temperature and type, rice variety and cultivation practices are other factors which contribute to CH_4 emissions. The largest emissions correspond to Asia representing 30% to 50% of global emissions linked to rice production. The global emission estimate is 30 Tg CH_4 yr^{-1}, which represents 8% of total anthropogenic emissions (Saunois et al., 2020).

Waste management accounts for 12% of global anthropogenic emissions, and it includes emissions of managed and non-managed landfills and wastewater handling. The pH, moisture, and temperature of the materials are key factors to produce CH_4 from waste. For wastewater, the production of CH_4 is defined by the amount of degradable carried material. High organic material leads to anerobic processes generating more CH_4. Waste management emission is estimated to be 65 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Biomass and biofuel burning

Biomass and biofuel burning emits CH_4 by incomplete combustion conditions, like charcoal manufacturing. This category includes emissions from biomass burning in the forest, savannahs, grasslands, peats, agricultural residues and burning biofuels in residential areas.

Biomass burning accounts for 5% of total anthropogenic emissions, and arson fires represent the 90% of total fires in the Earth. Anthropogenic fires are originated for many reasons, such as, clearing lands for agricultural purposes or agricultural waste burning. The estimated emission on this sector is 17 Tg CH_4 yr^{-1} (Saunois et al., 2020).

Energy production from biomass burning, called biofuel burning, represents 3% of total global

anthropogenic CH_4 emissions. It comes mostly from domestic cooking or heating where materials like wood, charcoal, agricultural residues or animal dung are burned. This sector contributes with 12 Tg CH_4 yr^{-1} to the global CH_4 budget (Saunois et al., 2020).

Fossil fuels

This category includes emissions related to exploitation, transportation and usage of coal, oil and natural gas. The global mean emissions related to fossil fuel activities is 128 Tg CH_4 yr^{-1} for the decade 2008 – 2017.

In coal mining, CH_4 is produced mainly by ventilation ducts which introduce large volumes of air to maintain CH_4 mixing ratio below 0.5% and prevent accidental ignition. It is also emitted during post-mining handling, processing and transportation. Abandoned mines and coal piles also emit CH_4. The emissions estimate of coal mining represents 33% of fossil fuel related activities with 42 Tg CH_4 yr^{-1}.

Oil and natural gas systems represent 63% of total fossil fuel emissions, including conventional and shale oil and gas production. CH_4 is emitted in all the chain of gas production. Fugitive and planned emissions occur during drilling, extraction, transportation, storage, distribution, end use and incomplete combustion. Fugitive emissions can occur during the transportation stage, in oil tankers, fuel trucks or pipelines. Measurement campaigns showed that they can be originated in localized areas, storage facilities, pipeline pressurization/depressurization points, etc. However, they vary from one city or infrastructure to another. Undetected single point failures could lead to large emissions at high rate. Inventories, in general, rely on the same sources and magnitudes for activity data using emission factors, which are country specific or site specific. Emission factors produce large uncertainties usually underestimating emissions from oil and gas facilities. The estimated contribution to the global CH_4 budget is 80 Tg CH_4 yr^{-1}.

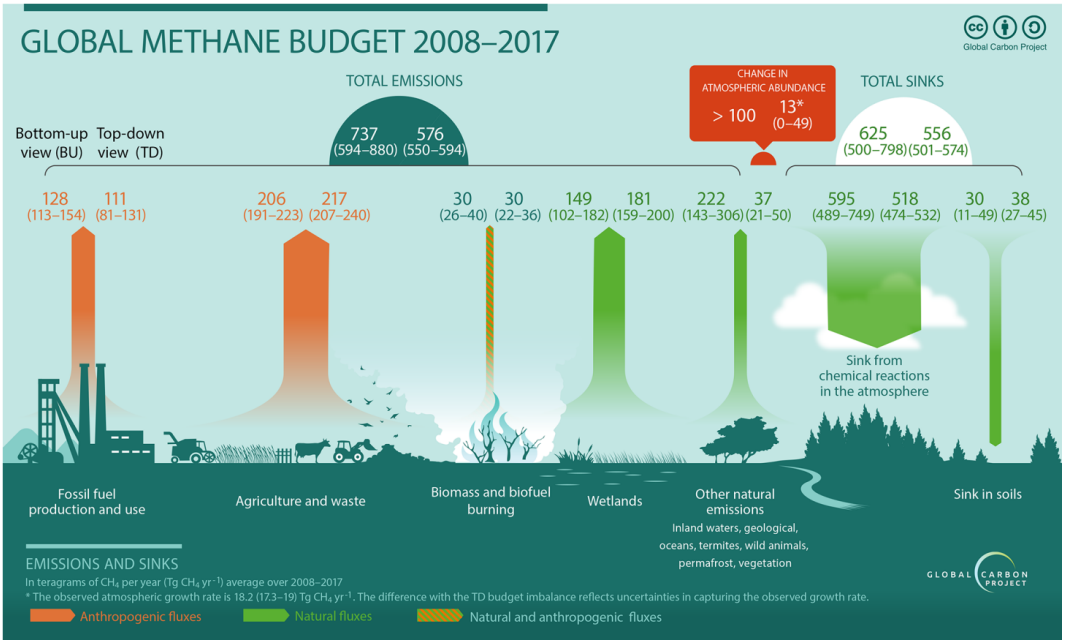


Figure 1.4: Global methane budget during 2008 – 2017 shows Bottom-up (left numbers) and Top-Down (right numbers) estimates and the min - max range (in parentheses) for each emission and sink categories and the total estimates. Units are in Tg CH_4 yr^{-1} (Source: (Saunois et al., 2020))

1.3 Improving emission estimates

From the BU perspective, inventory datasets from fossil fuel activities, livestock production, agriculture and waste management present large spatial and temporal variations (Saunois et al., 2020). They are also not available for many cities (Arzoumanian et al., 2019), and the uncertainty of these datasets increase due to omission or double counting in some sectors.

From the TD approach, a common practice is to measure the gradients of dry air greenhouse gases between the upwind and downwind vicinity and infer the emissions that are consistent with those gradients and uncertainties with atmospheric transport models (ATM). At a global or regional scale, it could provide accurate estimates, but at local scale, with few stations in the vicinity of an emitter zone, such as a city or an oil and gas facility, the uncertainties increase (Arzoumanian et al., 2019). TD emission estimates can be used as a verification tool for BU estimates on regions with large measurement networks like Europe (Bergamaschi et al., 2018).

Fugitive leaks of CH_4 also represent a safety issue that needs to be continuously monitored due to its episodic and spatial variability nature (Coburn et al., 2018). Leak detection and repair

(LDAR) surveys are conducted periodically to detect and quantify those leaks, specially at industrial facilities. It consists in detecting hotspots and quantify them with high precision analyzers. Even though they are conducted periodically, they present limitations such as, the portability of instruments or the accessibility of the site.

The idea of developing a high-density network of sensors to improve greenhouse gases fluxes monitoring in a region was studied by Wu et al. (2016); Turner et al. (2016) and Kumar et al. (2015). This study showed that uncertainties at local scale are largely reduced. In addition, such a network could allow to continuously monitor fugitive leaks at industrial facilities or landfills. Nevertheless, deploying research level instruments at this scale is cost prohibitive and it has minimum technical requirements to correctly operate, like a constant power supply, which is not always available for monitoring wastelands or for oil and gas facilities.

These limitations have motivated the research of instruments that are portable and cheap, can provide reliable measurements of greenhouse gases and can be deployed in dense networks.

In the following sections are explained the common CH_4 measurement techniques and the available technologies to quantify CH_4 concentrations.

1.4 Methane measurement techniques

There are different measurement techniques to quantify fugitive CH_4 emissions that can be conducted at the surface and spread out to several kilometers away from the point source. Their time scales also vary from minute resolutions to days or months, even sporadic measurements. We can distinguish two kind of techniques: qualitative and quantitative measurement. Figure 1.5 presents a diagram of the most common used qualitative and quantitative techniques with their approximate area coverage. On the following subsections we summarize the measurement techniques reported in the review paper of Mønster et al. (2019). Despite the techniques listed below are focused on ground-based sources, several techniques can also be applied to quantify fluxes on industrial facilities.

1.4.1 Qualitative measurement techniques

These techniques are used as a reconnaissance of a site and allow to detect hotspots.

1.4.1.1 Visual inspection

It is the first approach for routine inspection. It consists in conducting a visual and olfactory inspection on site. The detection of possible hotspots is linked to odors perception from additive odorous gases like THT (TetraHydroThiophene), and irregularities on the vegetation or material on site. Topography and wind conditions could alter the perception of the operator.

1.4.1.2 Portable CH_4 analyzer

Measurements are conducted with handheld analyzer that quantifies concentrations at surface level. The screening process is carried out by determining sections of the site and measuring the concentration by following grid points or by walking on accessible paths. The recorded concentrations at surface level are, theoretically, correlated with surface fluxes; although this relationship depends on different factors such as, the measurement distance from the source (e.g. ground), ground characteristics and weather conditions (atmospheric transport).

1.4.1.3 Field infrared survey

It uses thermal infrared devices, usually cameras, to detect changes in temperature, which are associated to gas emissions. It is easy to deploy and this can be done on a site, an unmanned aerial vehicle (UAV) or an aircraft. The main issues are: i) the origin of thermal anomalies, which can come from other processes than the gas emission; ii) the instrument will detect surface features such as materials with high heat absorption without leakage; and iii) the detection limit as it will need a minimum of CH_4 flux before the instrument can detect it.

1.4.1.4 Unmanned Aerial Vehicle (UAV)

UAV, such as drones, can be equipped with different type of CH_4 sensors. Their deployment is faster and simpler than traditional surveys. Screening could be controlled with more precision and cover larger areas. Limitations come from battery life (generally about 15 to 30 minutes), weather conditions and safety regulations on the site for UAVs. This technique was presented as qualitative by Mønster et al. (2019), although it can be categorized as a quantitative technique when using mass balance techniques or more complex inverse modeling (see next section).

1.4.2 Quantitative measurement techniques

The following techniques are arranged in ascending order from small-scale to large scale measurements.

1.4.2.1 Vertical soil gas concentration profiles

This method relies on the computation of the flux with equations of diffusional and pressure fluxes fed by single point measurements of CH_4 concentrations and pressure variations. It allows to understand the direction of the diffusional flux and the presence of subsurface CH_4 sources and sinks. Highly limited by the spatial variability, it also requires information of soil diffusivity. It is not recommended for upscaling in the quantification of emissions of an area or site.

1.4.2.2 Surface flux chambers

Mainly designed for ground sources, it consists in single point in-situ measurements using a chamber installed on site on the ground sources. It could be a 'closed' chamber if there is no exchange between the air in the chamber and the outside air; and an 'open' chamber when there is a continuous flow of air inside the chamber. Closed chambers can measure emission and uptake of CH_4, providing positive and negative fluxes. There are several chamber size configurations from small-size (< 0.1 m^2), mid-size (0.4 – 1 m^2) to over-size (> 15 m^2). Deployment is labor intensive and time consuming, and the detection of fluxes is linked to the chamber place-

ment on site. Upscaling the measured fluxes from such limited surface study depends on the spatial homogeneity of the sources on the considered site.

1.4.2.3 Eddy covariance

This method relies on the measurement of the turbulent transfer of gases between the surface and the atmosphere. It requires fast and high precision instruments to measure eddy fluxes. The footprint is linked to the height from which the measurements are taken, in the order of a radius equivalent to 100 times the height of the measurement point if sufficient information of wind is measured. It is sensitive to topography and wind conditions. The footprint study is a key parameter to interpret the estimated fluxes from such technique and it presents the same limitations of upscaling than other small area studies like surface flux chamber.

1.4.2.4 Stationary mass balance

This technique is based on measurements of gas concentrations at different heights and is combined with atmospheric vertical mixing (e.g. vertical diffusion coefficients) to determine the flux of CH_4. Advantages and limitations are in line with the eddy covariance method with the difference that it does not require fast instruments. Typically, measurements are taken at a maximum height of 10 m and cover a radius of 150 m (Scharff et al., 2003).

1.4.2.5 Radial plume mapping

It consist in using lasers and mirrors to measure downwind plume cross-sections, together with wind profile measurements and background (upwind) measurements. It provides estimations of emission on a localized area. This technique is limited by the line of sight between the lasers and the mirrors, the stable wind conditions and it requires two independent set of instruments. This makes its deployment costly and dependent on the site topography; although a deployment with a single instrument is possible depending on the disposition of the instrument and its scanning ability.

1.4.2.6 Mass balance using aerial measurements

This technique relies on high precision CH_4 sensor embedded on an aircraft or an UAV in order to measure the emission plume downwind, coupled with wind conditions to provide integrated emissions of the site. It is useful when access to the site is not available, but separate close sources is challenging.

1.4.2.7 Tracer gas dispersion

This technique consists of measuring CH_4 and a tracer gas simultaneously. The tracer gas is released at a fixed and known rate and at a known location (ideally collocated with the source to consider). The CH_4 emission rate is determined by the concentration ratio between both gases. It is called 'stationary' when measurements are on fixed point locations, and 'dynamic' when measurements are mobile. The latest one requires a vehicle to drive downwind the source in order to measure across the plume at a suitable distance. The advantage is that it does not rely on atmospheric conditions and provides integrated measurements of area fluxes. Common tracer gases employed are N_2O and C_2H_2, depending on the site co-emitted species and atmospheric background in the vicinity.

1.4.2.8 Differential absorption LiDAR (DIAL)

This method relies on emitting a pulsed laser radiation into the atmosphere at an appropriate absorption line of CH_4. Part of the emitted radiation is backscattered by the atmosphere and measured by a detector. The quantification is done by comparing it with a second detector that emits the radiation in a different wavelength that is not absorbed by CH_4 but is reflected in a similar way as the first laser. Measuring at several lines of sight gives information of the vertical distribution of CH_4 concentration. Combining the vertical CH_4 profile with wind measurements at different height provides the emission estimates. Data treatment is complex, and deployment is expensive.

1.4.2.9 Inverse modeling

It consists of providing measurements to an inverse transport model together with atmospheric parameters, such as wind speed and direction. It can be of different types depending on the nature of the measurements: 'stationary', if models are feed by measurements of fixed-point locations; 'dynamic', if they are captured downwind the plume; and 'aerial', if observations are obtained by aerial instruments like UAV or an aircraft. These approaches can give estimates of emissions for the entire site and for emissions from a specific area, depending on the good quality of the data and its density. These models requires a complex data treatment and they are sensitive to other interfering sources and to the topography of the site.

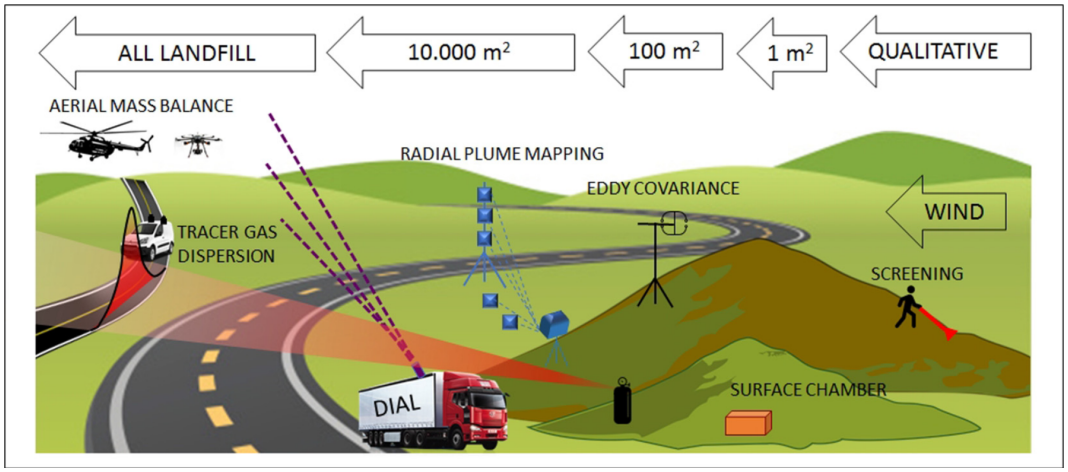


Figure 1.5: Common methods used to quantify CH_4 emissions (Source: (Mønster et al., 2019))

1.5 Technologies used for methane detection

There are six types of sensors that are typically used for methane detection and quantification: electroacoustic, calorimetric, pyroelectric, electrochemical, chemi-resistive, optical and electroacoustic sensors. On this thesis we will study Chemi-resistive sensors, specifically metal oxide semiconductors (MOS), and for the reference measurements we use optical sensors, cavity ring down spectrometers. Therefore, only a brief description will be made for the other types of CH_4 sensors.

1.5.1 Electroacoustic sensors

The quartz crystal microbalance (QCM) and the surface acoustic wave (SAW) sensors are the most commonly used to measure CH_4. QCM is composed of a resonating disk with metal electrodes. The principle of measurement is as follows: an oscillating signal is provided to the disk producing a resonation at a characteristic frequency. When CH_4 is in contact with the sensing material, it produces an increase in the mass of the device reducing the frequency of the resonance. The rate of change in frequency is inversely proportional to the mass adsorbed by the sensing material (Hong et al., 2020). SAW sensors are based on two principles: the first one is the piezoelectric effect that depends on the quality of some materials to produce electrical signal after a mechanical stress. The second one is the surface elastic waves, specifically Rayleigh waves (RW), which are characterized by propagating on the surface of a material and decaying into the depth of this material. The principle of sensing is to produce an excitation of the piezoelectric material with RW. The presence of the target gas will affect the modulation of the RW's characteristics and those changes in modulation indicate the concentration of the target gas (Devkota et al., 2017).

1.5.2 Calorimetric sensors

Calorimetric sensors that measure CH_4 are used in several applications, such as in coal mines, petroleum drilling and landfills (Aldhafeeri et al., 2020). Calorimetric sensors are composed of a temperature sensor, a catalytic combustor and a heater. The principle of sensing is based on a chemical reaction that absorbs or releases heat. The catalytic material when it is in contact with the target gas produces a chemical reaction, in the case of CH_4 it releases heat and changes the temperature on its surface. This is measured and converted to the gas concentration. Common catalysts are platinum, palladium and rhodium. Common problems of these sensors are cracking, catalyst poisoning, oversaturation from high concentration of gas, high power consumption, unsuitable for long term deployments and requirement of high temperatures to operate (Aldhafeeri et al., 2020).

1.5.3 Pyroelectric sensors

Pyroelectric sensors are another type of methane sensors, whose typical applications are as fire alarms, laser detectors, thermal or gas analyzers. They are composed of an electric or infrared laser heater that produces a thermal wave, which travels through a dielectric substrate and the gas, to finally arrives to a pyroelectric material. This produces an electrical signal, voltage, in proportion to the change on the heat measured that can be correlated to gas concentration. The presence of the gas will alter the heat measured by the pyroelectric material. This kind of sensors perform well at room temperature, since it does not alter the gas by any chemical reaction and the degradation of the material is low. The main drawbacks are the high cost of pyroelectric materials and high energy consumptions. (Aldhafeeri et al., 2020)

1.5.4 Electrochemical sensors

Electrochemical sensors are composed of three electrodes, called 'working', 'counter' and 'reference', and an electrolytic fluid that maintains charge neutrality in the system. The 'working' electrode will suffer a chemical reaction when it is in contact with the target gas. The 'counter' electrode balances the current on the 'working' electrode and the 'reference' electrode measures the electrode potential. The concentration of the target gas is proportional to the potential measured between the 'working' and the 'counter' electrode. The application of these sensors depend on the materials used for the electrodes and the electrolytic fluid (Aldhafeeri et al., 2020)

1.5.5 Chemi-resistive sensors

The principle of sensing is to measure the change of electrical properties when the sensing material is in contact with the target gas. The most common are the metal oxide semiconductors (MOS) and the conductive polymers/carbon material-based sensors.

MOS sensors are composed of a sensing material and a heater material. The heater increases the temperature of the sensing material between 200° to 500°C (Özgür Örnek and Karlik, 2012; Barsan et al., 2007; Hong et al., 2020) attracting oxygen molecules to the surface of the sensing

material. In presence of a deoxidizing gas, such as CH_4, a change in the conductivity is produced by removing some of the oxygen molecules in the surface of the sensing material. This change is measured by an external circuit (Özgür Örnek and Karlik, 2012; Hong et al., 2020), usually a voltage divider. Conductive polymer/carbon material sensors quantify CH_4 concentrations as changes in conductivity when there is direct interaction with the sensing material. (Hong et al., 2020).

The most common material used in MOS sensors is the tin oxide/stannic oxide (SnO_2) that presents high chemical and thermal stability in air. It has proved to distinguish different VOC gases such as carbon monoxide, hydrogen, nitrogen dioxide and methane (Hong et al., 2020). The effect of water vapor on CH_4 and CO was studied by Ionescu et al. (1999) and it was found that CH_4 and H_2O compete in reacting with oxygen at the surface affecting to the sensing performance in presence of humidity. A prolonged exposure to humid environments will degrade progressively the sensitivity of the sensor (Wang et al., 2010). In addition, the work of Tischner et al. (2008) has shown that after exposing the sensors to several humidity pulses, the resistance does not recover to its initial levels until it is heated up to temperatures up to 450°C for several minutes. Variations in temperature need to be taken into account when working with this kind of sensors, as the response of the sensors increases with temperature until it reaches a maximum at a temperature value, then it decreases rapidly on higher temperatures (Kolmakov et al., 2005). Another drawback of these sensors is the poor selectivity to various gases and several studies have tested ways of improving its selectivity and sensing response using catalyst, such as silver, iron, palladium, nickel, platinum or CuO. (Hong et al., 2020)

Chemi-resistive sensors are inexpensive and small which make them suitable for massive deployment in a network of sensors. Nevertheless, despite their wide range of detection they tend to be less precise than other technologies, such optical sensors. They also present cross-sensitivities to other electron donor gases and atmospheric parameters, such as temperature, humidity and pressure (Popoola et al., 2018). MOS sensors also present drift issues requiring develop continuous calibration models (Romain and Nicolas, 2010).

In this thesis we will focus on the study of MOS sensors from the manufacturer Figaro®, specif-

ically the Taguchi Gas Sensors (TGS) 2600, 2611-C00 and 2611-E00. All three sensors are sensitive to CH_4, but their design alter their sensitivities to CH_4. TGS 2600 was conceived as a smoke detector and it is sensitive to various air contaminants, such as, hydrogen, CO, iso-butane, ethanol and methane. TGS 2611-C00 and TGS2611-E00 are two versions with increased selectivity to gases presenting less response in presence of ethanol and iso-butane. The last one includes a carbon filter further increasing its selectivity.

Figure 1.6 and 1.7 illustrate typical values of the sensitivity response of the three types of TGS measured at air temperature at 20°C and 60% relative humidity (Figaro®, 2005, 2013).

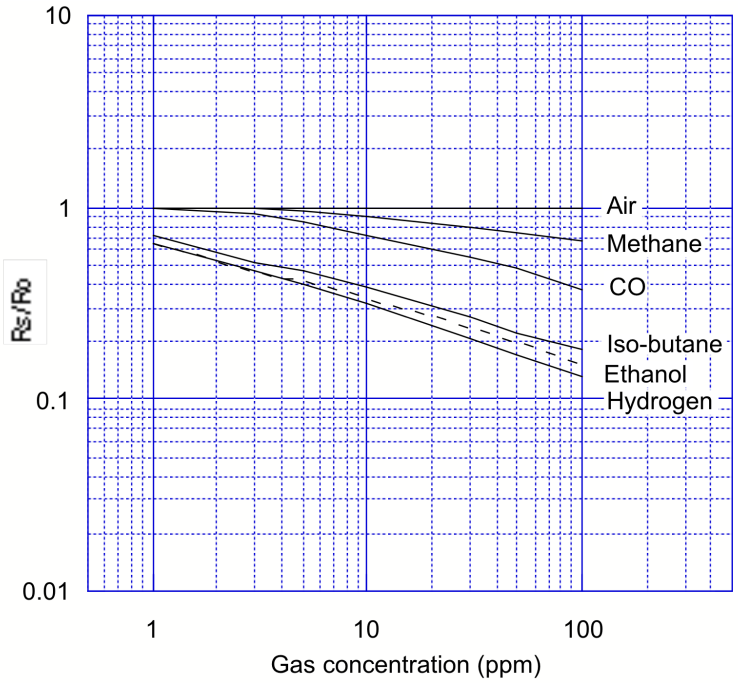


Figure 1.6: Sensitivity response of TGS 2600 to different gases on concentrations from 1 ppm to 100 ppm. (Source: (Figaro®, 2005))

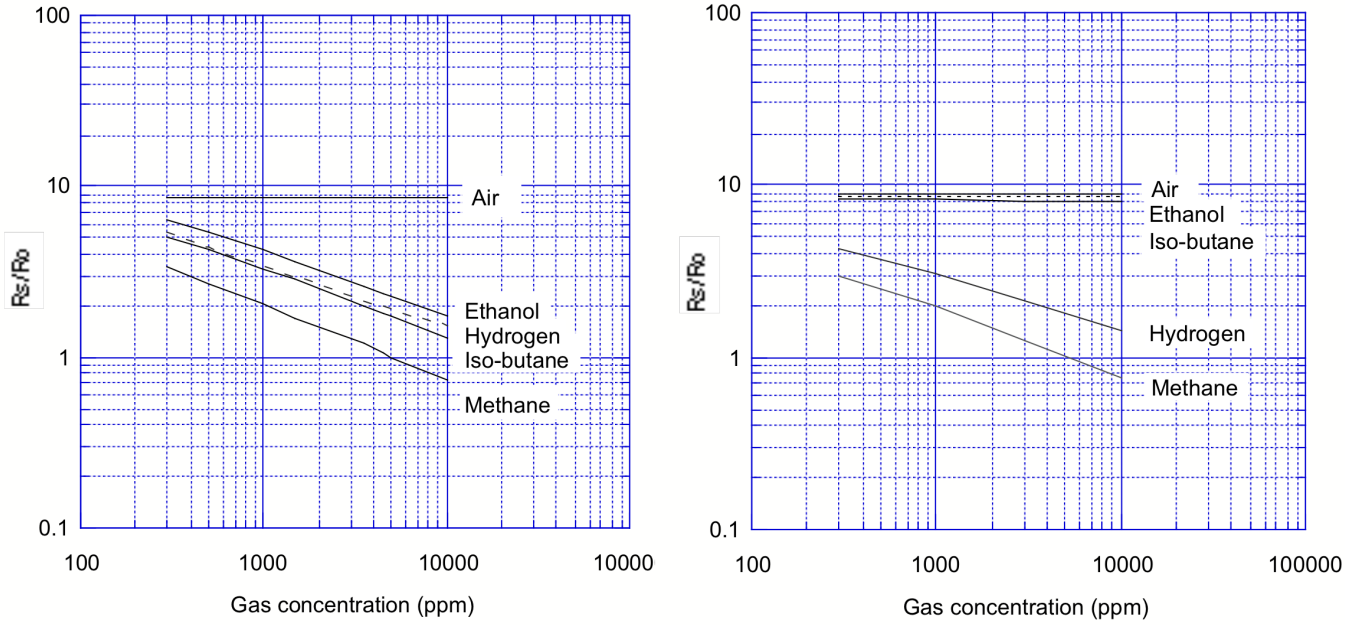


Figure 1.7: Sensitivity response of TGS 2611-C00 (left) and TGS2611-E00 (right) to different gases on concentrations from 200 ppm to 10000 ppm. (Source: (Figaro®, 2013))

1.5.6 Optical sensors

These sensors detect changes in visible light or electromagnetic waves resulting from an interaction of the target gas with the receptor. There are several optical sensing mechanisms, such as fluorescence, absorption, colorimetry, refractive index, surface plasmon resonance or luminescence. (jie Yin et al., 2018). Although, the most common technique is the infrared absorption spectroscopy to quantify CH_4 sensors. Depending on their specific molecular vibration (related to their chemical structure), each gas species interact with infrared radiation and in particular absorb infrared radiation at a specific wavelength. Methane presents two strong absorption bands around 1.3 and 3μm wavelengths. Cavity ring down spectroscopy (CRDS) is a technique based on infrared absorption with high sensitivity detection and accurate quantification of greenhouse gases able to meet WMO/GAW inter-laboratory compatibility goals (Rella et al., 2013).

The CRDS consists of a measurement cell bounded with high-reflectivity concave mirrors (usually > 99%), a laser and a photodetector. The working principle is as follows: the light from the laser is injected into the cavity through one of the mirrors, and then it is reflected in the whole cavity. Its light intensity increases over time and it is monitored by the detector placed outside

the cavity behind a mirror. An exponential decay in the intensity occurs when the laser is turned off with a known time constant τ. In the presence of an absorbing gas, the intensity decay is affected, shortening the decay time due to absorption and scattering of the gas. The mixing ratio of the target gas inside the mixture is computed comparing the decay time with and without the gas. The high sensitivity of CRDS instruments is due to the high reflectivity of the mirrors allowing the laser beam remains trapped inside the cavity for a long period of time, usually in the order of μs, representing an optical path length of few kilometers (Wheeler et al., 1998; Crosson, 2008) with precision reaching to parts per trillion depending on the analyte gas (Rella et al., 2013).

On this thesis we use high precision CRDS instruments as a reference for the CH_4 measurements to which we compare the MOS sensor measurements. We will briefly explain a few of the characteristics of the instruments used as reference in our studies. Picarro Inc. (Santa Clara, CA) is a manufacturer of research level CRDS instruments, equipped with telecom lasers, a setup of three high-reflective mirrors inside a cavity with a volume about 35 cm^3 and with an effective optical path length of 15 to 20 km (Rella et al., 2013). Their system allows to capture discrete spectrum on limited wavelength band on a millisecond timescale that convertes to gas concentrations with a precisions of 0.3 ppb for CH_4 at 5 seconds time resolution. Laboratory test (Yver-Kwok et al., 2015) pointed to high stability of the instruments with drifts over time in about 1 ppb CH_4 per month.

1.6 Current calibration methods used on TGS MOS sensors

In recent years, a focus on TGS MOS sensors have led to several studies of its capabilities to quantify not only CH_4 but other VOCs with promising results.

The use of empirical equations that link resistance measurements to CH_4 concentration have been used on several studies (Eugster et al., 2020; Eugster and Kling, 2012; Riddick et al., 2020) with good performances, although this approach is sensitive to the site and environmental conditions.

Linear and multilinear models have been also explored (Casey et al., 2019; Collier-Oxandale

et al., 2019, 2018), introducing interactions between environmental variables on the multilinear models have produced better results than simple linear models.

Artificial neural networks (ANN) have been also tested (Casey et al., 2019; Eugster et al., 2020) with interesting results: ANN outperformed linear models and provided relatively good estimates but failed at reproduce the range of the CH_4 signal. On warmer temperatures the ANN was no better than a linear model, but on colder temperatures its performance improved, and daily features were better captured by the ANN model than on linear models.

1.7 Objective and structure of this thesis

In recent years the focus on low-cost sensors, especially MOS sensors, has increased. Several studies partly assessed the capabilities of these sensors to quantify CH_4 concentrations and to be used in a network to detect and quantify fugitive emissions on industrial sites or landfill. This thesis aims to improve the knowledge of MOS sensors as a tool to detect and quantify fugitive emissions. The main objective of this PhD is to determine a calibration strategy for CH_4 MOS low-cost sensors from the manufacturer Figaro® and to provide reliable estimates of CH_4 emission computed from observations of these sensors when they are deployed in the field.

The approach to achieve this goal is to conduct three experiments on which the MOS sensors were tested. The manuscript is structured as follows:

In chapter 2, we present the first experiment, which consists in a 47 days deployment of a chamber with three types of MOS sensors (TGS 2600, TGS 2611-C00 and TGS 2611-E00) together with a high precision instrument (Picarro CRDS G2401) that will act as a reference, in a controlled space, room laboratory, where temperature and relative humidity are continuously monitored and presents low variations compared to a deployment in the field. Both chamber and reference instrument were exposed to room air. This first experiment in the study is important, since it explores the behavior of the MOS sensors under controlled conditions, allowing to understand the sensitivity to environmental parameters, notably H_2O, and it presents a first approach of data treatment for measurements of these sensors coupled with a machine learning model to

reconstruct the CH_4 signal.

Chapter 3 is dedicated to the second experiment on which we explore the capabilities of two versions of the MOS sensors (TGS 2611-C00 and TGS 2611-E00) to detect and quantify high frequency enhancements of CH_4 (spikes) under controlled conditions. Four chambers including both types of sensors were collocated with a reference instrument and exposed to artificial CH_4 spikes generated over variations of ambient air measurements (sampled from outside). On this experiment we have explored four questions related to the detection and quantification of CH_4 concentrations and the deployment of MOS sensors on an industrial site or landfill: i) the capability of detection and quantification of CH_4 concentrations of high frequency spikes comparing each type of sensor and employing five different models (parametric and non-parametric); ii) the possibility of reducing the training set size without losing performance on the reconstruction of the signal; iii) the effect of the ageing of the sensors in the reconstruction of the signal and; iv) the capability of sharing models between chambers. This study, as a second step, presents the basis of the calibration strategy defining the data preprocessing approach, the modeling selection and the discussion of several aspects of the sensors to consider in a field deployment.

On Chapter 4 we present the results of the third experiment, where we assess the performance of MOS sensors on a real case study in the field. Seven chambers containing two types of Figaro® TGS sensors (TGS 2611-C00 and TGS 2611-E-00) have been deployed on an experimental platform (Total Energies ® Anomaly Detection Initiatives, TADI) implemented with equipment simulating the typical environment of an industrial site. At this site, the chambers collocated each one with a high precision instruments have been exposed to a series of controlled releases of CH_4 with other compounds to simulate typical fugitive emissions at oil and gas facilities. In this experiment, we have treated two main questions: the first one is related to the capability of detection and quantification of CH_4 plumes (concentration) by the sensors on an environment where environmental factors vary and there is also the presence of other compounds that could interfere on the capability of the sensors to detect the interested species. The second question is related to the capability to quantify emissions from the reconstructed signal employing a Gaussian plume dispersion model and assessing the performance by comparing with the estimates

obtained with the reference instruments and the actual values for the same releases. We have employed the tools developed on the previous chapters to retrieve the information of the signal linked to the CH_4 enhancement. We have compared two models, statistical and machine learning, to reconstruct the CH_4 signal, and we have provided accurate estimates of the emissions for the releases exploiting the framework developed by Kumar et al. (2022).

Finally, on Chapter 5 we summarized the findings and provide an outlook of the possible research paths of these sensors.

Chapter 2

The Potential of Low-Cost Tin-Oxide Sensors Combined with Machine Learning for Estimating Atmospheric CH_4 Variations around Background Concentration

Rodrigo Rivera Martinez[1], Diego Santaren[1], Olivier Laurent[1], Ford Cropley[1], Cecile Mallet[2], Michel Ramonet[1], Christopher Caldow[1], Leonard Rivier[1], Gregoire Broquet[1], Caroline Bouchet[3], Catherine Juery[4] and Philippe Ciais[1]

[1] Laboratoire des Sciences du Climat et de l'Environnement (LSCE), LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
[2] Laboratoire Atmosphères Milieux, Observations Spatiales (LATMOS), UMR8190, CNRS/INSU, IPSL, Universite de Versailles Saint-Quentin-en-Yvelines (UVSQ), Quartier des Garennes, 11 Boulevard d'Alembert, 78280 Guyancourt, France
[3] SUEZ, Smart & Environmental Solutions, Tour CB21, 16 Place de l'Iris, 92040 La Defense, France
[4] Total Raffinage Chimie, Laboratoire Qualite de l'Air, 69360 Solaize, France

The full article was published in the journal Atmosphere 2021, 12, 107. https://doi.org/10.3390/atmos12010107.

Date of submission: 18 November 2020. Date of publication: 13 January 2021.

2.1 Summary of the publication

2.1.1 Introduction and motivation

Detection and quantification of CH_4 emissions from fossil fuel production or agricultural and waste facilities is challenging due to the nature of the emissions, temporally and spatially variable (Coburn et al., 2018). Emission estimates from inventories and atmospheric measurements have large inconsistencies (Duren et al., 2019). Some solutions to improve these estimates are proposed, like the development of a dense network of sensors that can continuously monitor at fixed locations on an area like a landfill or at oil and gas facilities (Kumar et al., 2015; Wu et al., 2016; Turner et al., 2016). Developing dense networks with research level instruments, such as Cavity Ring Down Spectrometers (CRDS), despite their high precision and low maintenance in terms of calibration, are a costly option, especially for a long-time deployment. The estimated prices are in the order of tens of thousands dollars per instrument. In recent years, the research of low-cost alternatives has been studied. Metal Oxide semiconductor sensors are a good alternative to implement in such dense networks, although as explained on Chapter 1 they are prone to drift over time and they are sensitive to environmental parameters such as H_2O and temperature and to other VOCs. In this chapter we present the results obtained in a first experiment involving three types of MOS sensors from the manufacturer Figaro®. We will focus specifically on the Taguchi Gas Sensors (TGS) 2600, 2611-C00 and 2611-E00 to measure CH_4 at background atmospheric levels. In order to assess the reconstruction accuracy, we defined an error target with a RMSE below 0.2 ppm of methane.

2.1.2 Design of the experiment

Two sensors of three different TGS types (2600, 2611-C00 and 2611-E00) were installed inside a stainless steel/glass chamber of 120 mL together with two sensors to measure environmental parameters (SHT75, that measures temperature and relative humidity; and BMP180 for temperature and pressure). The logging system was controlled by a Raspberry PI 3b+ and sampling was

made at 0.5 Hz. The chamber was collocated with a high precision instrument, Picarro G2401 CRDS, to have reference measurements for the CH_4 mole fractions.

We performed two sensitivity tests to measure the response of TGS voltage to variations of CH_4, H_2O and CO. The first test consisted in measuring the sensitivity of TGS to CO and H_2O. In this case, we exposed the sensors to air from a single high pressure air cylinder with a mixing of 1.5 ppm CO and 2 ppm CH_4 in natural air. The sampling line was splitted in two branches. In one of the branches, CO was removed (chemical scrubber) without affecting the H_2O mole fraction. Both branches included dedicated mass flow controllers and they were finally mixed at different ratios to generate eight CO levels (0, 0.07, 0.14, 0.3, 0.5, 0.8, 1.2 and 1.5 ppm). In addition, the H_2O mole fraction was controlled by a dew point generator producing three levels of H_2O (0, 1 and 2.3 %). In the second test we assessed the sensitivities of TGS resistance to CH_4 and H_2O, we exposed the MOS sensors to six different concentrations levels of CH_4, from 1.9 ppm up to 9 ppm, similar to concentration levels we can typically encounter in CH_4 enhancements around industrial sites. Each level was generated by the mixing of two high pressure dry air cylinders: one with High CH_4 mole fraction (9 ppm CH_4 and 0.08 ppm CO) and the other with Low CH_4 mole fraction (1.9 ppm CH_4 and 0.11 ppm CO). Then, it was humidified by a dew point generator to obtain four different H_2O levels (0.65, 1, 1.5 and 2.5 %). A stabilization time of the sensors of about 40 minutes was considered after each change of H_2O and CH_4 levels were increased at 20 minutes intervals.

The second part of the study was dedicated to the reconstruction of room air variations of CH_4, with magnitudes ranging from 1.9 to 2.4 ppm, from TGS resistance measurements using artificial neural networks (ANN), specifically the Multi-Layer Perceptron (MLP) model. Sensors were exposed to indoor air of an air-conditioned room for a period of 47 days. The reference measurements were measured by a Picarro G2401 CRDS instrument which took air from the same sampling line that was connected to the TGS chamber. We defined five predictors to train the reconstruction model: TGS resistance variations, CO, H_2O mole fraction, and air temperature and pressure. To train and assess the MLP model we have subset the dataset in a training set, corresponding to 70% of the observations, and a test set, for the remaining 30%. The influence of the choice of the training set in the model's performance was studied by creating 50 sliding

training and test sets that produced 50 different models, and four metrics were evaluated for each model: RMSE, bias, the ratio between the spread of the predicted outputs and the spread of the true values, and the correlation coefficient between the output of the models and the reference. We conducted a sensitivity test on the models with two approaches. The first was training the MLP models on two cases, i) by removing one predictor at a time and ii) testing the effect in the reconstruction error of the inclusion of different types of TGS or several versions of the same type of TGS in the models. The second approach was computing the partial dependence plots that shows the marginal effect of each predictor to the modelled signal.

2.1.3 Main results

2.1.3.1 Sensitivity to CO and $\mathbf{H_2O}$

The first test allowed us to determine the voltage contribution due to CO variations and was used to correct the measured voltage for CH_4 variations of the second test. We determined this voltage contribution caused by CO variations by fitting a multivariate quadratic model for each type of TGS on which we included the interaction between the CO and H_2O.

The corrected voltage, transformed to resistance, on the second experiment showed a non-linear relationship between the sensors' resistance variations and H_2O. This experiment, aligned with previous studies (Ionescu et al., 1999; Wang et al., 2010; Tischner et al., 2008), showed the high sensitivity to H_2O of MOS sensors, specifically in this study for the three types of TGS. In addition, it was observed that the TGS studied presents a small sensitivity to CH_4 with a slightly higher sensitivity for the TGS 2611-C00 sensor (slope: -1.85kW/ppm CH_4 at 1% H_2O).

2.1.3.2 Reconstruction of $\mathbf{CH_4}$ signal with Multi-Layer Perceptron (MLP)

The results observed from the reconstruction of CH_4 concentrations with the MLP model showed two critical aspects. The first one is related to the selection of the training period. It was observed that the performances of the models were higher for cases that had an overlapping in the distribution of the training and test set, where the range of variations were similar. The second one is

the difficulty in reproducing high frequency variations of CH_4 by the MLP, which showed a low-pass filter behavior benefiting a correction of the misfits for low frequencies when adjusting the weights during the training process.

The sensitivity tests led to interesting results regarding the importance of the predictors in the construction of the model. In the test of removing one predictor at a time, we observed that performances were significantly reduced when removing H_2O, and slightly increased when CO was not included as predictor. For the test with several versions and types of sensors, inconsistent information from several types of sensors affected the stability of the model and caused a larger spread of the test set error, especially when including the TGS 2611-E00 type as a predictor together with the other types. The use of two versions of the same type also led to a degraded performance of the MLP.

The results obtained from the partial dependence plots derived from the results of the MLP model showed a qualitative agreement between the sensitivity of reconstructed CH_4 to resistances and the sensitivities measured directly from the experimental tests. The positive sensitivities to H_2O resulting from the partial dependence plots, up to 1.6% of H_2O, were consistent with the experimental sensitivities. However, for larger H_2O values, the model-derived sensitivities presented a different behavior which could be explained by a poor training of the model on certain ranges of values.

2.2 Publication in the MDPI Atmosphere Journal

2.2.1 Introduction

Anthropogenic CH_4 emissions comprise 30% of the global source of this greenhouse gas, and are from various economic sectors (Saunois et al., 2016). Oil, gas and coal sector sources are localized in space, going from point scale (e.g., a single well, a compressor) to area sources (e.g., a refinery, a gas extraction field). The waste sector has area sources for waste-water treatment plants and landfills, although within a site there can be leaky equipment forming point sources.

Numerous campaigns estimated emissions from point and area sources using atmospheric measurements by deploying local dense networks of CH_4 instruments at fixed points and using mobile ground-based platforms and aircraft (Alvarez et al., 2018; Collier-Oxandale et al., 2018; Duren et al., 2019). The signal from a source in terms of CH_4 concentration at a nearby atmospheric measurement location depends on the magnitude of this source, on the wind speed, on the atmospheric turbulence and on the sampling distance. An excess of CH_4 mixing ratio going from a few tenth parts per billion (ppb) (Ars et al., 2017; Yver-Kwok et al., 2015) up to several parts per million (ppm) (Feitz et al., 2018) is typically recorded at a downwind distance from the source.

Research class CH_4 analyzers such as Cavity Ring Down Spectrometers (CRDS) used for background air monitoring have a precision higher than 1 ppb (Berhanu et al., 2016; Rella et al., 2013) but they are expensive. Such precision is needed to monitor the small CH_4 gradients between background stations, on the order of 10 to 50 ppb, that are used as an input of atmospheric inversion models to diagnose large-scale emissions (Pison et al., 2018). The deployment of multiple CRDS instruments in the vicinity of an industrial site for detecting and/or estimating its emissions is however a too costly option on a routine basis, especially when needing a very dense network to ensure precise location and quantification of a fugitive source. This has prompted research to develop low-cost sensors with a precision sufficient to characterize the signal of atmospheric plumes from industrial sites. From typical plume signals that are on the order of more than a ppm, a precision of 0.1 to 0.2 ppm on instantaneous measurements can be deemed to be suffi-

cient for a low cost CH_4 sensors. Low cost sensors are more likely to drift with time than CRDS analyzers, but the atmospheric signal used to quantify the emission of an industrial site is the near instantaneous difference between upwind and downwind concentrations (Feitz et al., 2018; Kumar et al., 2015). Therefore, constraining the drift of upwind and downwind sensors during a few hours to be less than 0.1 to 0.2 ppm would still be sufficient for monitoring CH_4 emissions from an industrial site.

Here, we formulate a target precision requirement of 0.1 to 0.2 ppm over a time scale of one hour for CH_4 low cost sensors to be deployed on dense networks around an emitting source. This requirement is suitable for detecting variations of CH_4 in background air, which are on the order of 0.1 ppm on an hourly time scale, and characterize CH_4 conditions upwind from an emitting source. We tested for this requirement solid-state tin-oxide (SnO_2) sensors models TGS 2600, TGS 2611-C00 and TGS 2611-E00 manufactured by Figaro®. We performed measurements of room air where CH_4 concentration varies from day to day by up to 0.5 ppm above a background value of 1950 ppb. The principle of those sensors is to measure changes in the tin-oxide resistance affected by electron donors in the air to which the tin-oxide is exposed. These sensors are cheap, with a unit cost of about 3 € to 25 € per sensor and they were shown to be sensitive to low CH_4 concentrations, thus being potentially suitable for emissions monitoring with a good characterization of background variations and plume amplitudes even for modest sources (Collier-Oxandale et al., 2018). On the other hand, low cost sensors are known to drift with time, to be sensitive to other reduced species than CH_4 and to factors such as water vapor, pressure and temperature (Collier-Oxandale et al., 2019). Therefore, cross sensitivities to these other species must be characterized in order to understand how they impact the retrieval of CH_4 from measured variations of resistance.

The first research question addressed in this study is to characterize the cross sensitivities of Figaro® resistances to CH_4 versus other factors known to influence the tin-oxide resistance: carbon monoxide (CO), water vapor, pressure and temperature (Chaiyboun et al., 2007). To address this question, we characterized in a laboratory facility under controlled conditions the resistances of six Figaro® sensors for a range of CH_4, CO, temperature and H_2O, and examined covariance

between their sensitivities, in a first step to diagnose cross-sensitivity effects from non-CH_4 related variables.

The second research question is to test whether measurements of low-cost sensors resistance combined with other cross-sensitivity variables allow for the reconstruction of CH_4 concentration and its variability to meet our precision requirement of 0.1 to 0.2 ppm, here in the case of CH_4 variations around background values of up to 0.5 ppm. To address this second question, we analyzed time series of Figaro® resistances continuously recorded by six Figaro® sensors and CH_4 measured with a high-precision CRDS analyzer for room air CH_4 variations, during a period of 47 days. This study is the first step to assess the potential of Figaro® sensors for measuring CH_4 concentrations close to current atmospheric levels; with small co-variations of water vapor and of a limited number of cross-sensitive species. On previous studies Collier-Oxandale et al. (2019) and on initial tests showing that there is a non-linear relation between CH_4, resistances and other variables affecting resistances such as temperature and H_2O mole fraction, we chose to construct and apply a machine learning model to reconstruct the true CH_4 concentration from the CRDS by using as predictors the resistances of the Figaro® sensors, as well as H_2O mixing ratio, carbon monoxide, temperature and pressure recorded by other sensors. The model is trained to optimally reconstruct the true CH_4 signal during a given period, and its results are evaluated against an independent subset of the data. The results are systematically evaluated varying the training and test periods, the number of ambient variables, and the addition of more than one type of Figaro® sensor resistance to reconstruct the true CH_4 time series.

2.2.2 Experimental Set-Up

2.2.2.1 Measurement of Low-Cost Sensors Sensitivities to CH_4, CO and H_2O

The cross sensitivities of the resistance of Figaro® sensors types TGS 2600, TGS 2611-C00, and TGS 2611-E00 were measured in the laboratory (measurements conducted at LSCE, Saclay, France). The sensors were incorporated into a low-cost sensor logger that featured a Raspberry Pi 3B+ single-board computer and Raspbian operating system, using bespoke software (coded in Python).

Figaro resistances for types TGS 2600, TGS 2611-C00 and TGS 2611-E00 were measured as voltages using a voltage divider Figaro® (2005, 2013) with precision resistor (5 kΩ, tolerance, temp coeff), and measured as single-ended voltages using an A/D board (ADCPiPlus, ABElectronics) with 17-bit resolution across the 5.06 V range. Air temperature and relative humidity were measured using a Sensirion SHT75 digital sensor which has an accuracy of $\pm$0.3 °C and $\pm$1.8%RH respectively and a repeatability of $\pm$0.1 °C and $\pm$0.1 %RH respectively. Air pressure was measured using a digital Bosch BMP180 pressure sensor (Adafruit, BMP180 breakout module), which has an accuracy of $\pm$0.12% across the range 950-1050 hPa. All sensors were mounted in a 120 mL stainless steel/glass sealed chamber (EIF 3S1NRGL), which provided a gas inlet and outlet and an air-tight port for the sensor cable (see Figure 2.1a). All measurements were made at 0.5 Hz and stored on the Raspberry Pi's SD card.

To assess the sensitivities of the sensor resistances to CH_4 and H_2O, we used air from two high pressure dry air cylinders, with a high CH_4 mole fraction of 8.999 ppm CH_4 and 0.08 ppm CO, and a low CH_4 mole fraction of 1.900 ppm CH_4 and 0.11 ppm CO, respectively. Air from the two cylinders was mixed using two mass flow controllers (see Figure 2.1b) to create six levels of CH_4 of 1.9, 2.985, 4.04, 6.17, 7.58 and 8.985 ppm in dry air. This range covers CH_4 mole fractions recorded in the atmosphere from background sites up to typical excess found in plumes from industrial sites (Xueref-Remy et al., 2019). The air with different CH_4 concentration was humidified by a dew-point generator (Licor, LI-610) in order to get four H_2O mixing ratios of 0.65, 1, 1.5 and 2.5% at stable atmospheric pressure and temperature. The experiment set up is illustrated in Figure 2.1b.

In the experiment, the dew point generator was set to one of the four H_2O mixing ratios. At each change of H_2O mixing ratio, the Figaro® reading was given 40 or more minutes to stabilize at the lowest CH_4 level, before CH_4 was increased in steps at 20 min intervals. Only the last 5 min' data of each step was used. Data from the sensors and the Picarro CRDS were merged and converted to one-minute medians.

The Figaro® sensors' sensitivity to CO and H_2O was measured in a similar manner, using a single high pressure dry air cylinder containing 1.5 ppm CO and 2 ppm CH_4. The sample line was

split into two branches, one equipped with Sofnocat 514, a hydrophobic CO oxidizing agent, to remove CO without changing the humidity. The air from the two lines was combined in different ratios thanks to dedicated mass flow controllers in order to produce CO mole fractions of 0, 0.07, 0.14, 0.29, 0.57, 0.87, 1.17 and 1.50 ppm, at H_2O mixing ratios of 0, 1.0 and 2.3% thanks to a dew point generator (Licor, LI-610) operated at constant temperature and pressure. The experimental configuration is shown in Figure 2.1c. The logging equipment and sampling procedure were the same as for the first experiment.

In this experiment, a Picarro G2401 CRDS was used as a reference high-precision instrument for CH_4, CO_2, CO and H_2O mole fraction. The CH_4 precision of a Picarro CRDS analyzer in dry air is below 1 ppb (Picarro Inc.: Santa Clara, CA, USA; Yver-Kwok et al., 2015) at instrument data acquisition rate (0.3 Hz) within the atmospheric range. CRDS calibration drift over time is usually better than 1 ppb CH_4 per month (Yver-Kwok et al., 2015).

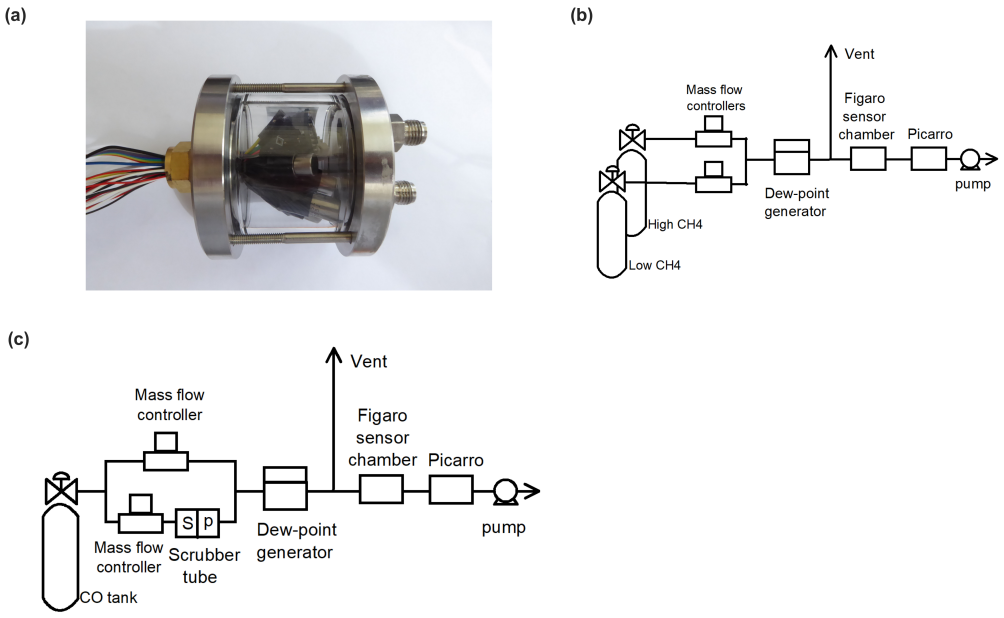


Figure 2.1: (**a**) Picture showing the sealed chamber with six Figaro® sensors of different types, and temperature and pressure sensors. (**b**) Scheme of the CH_4 and H_2O cross-sensitivity measurement set up. (**c**) Scheme of the CO and H_2O cross-sensitivity measurement set up.

2.2.2.2 Measurements of Room Air with Low Cost Sensors and CRDS

The resistances of six Figaro® sensors exposed to CH_4 variability in indoor air were monitored during 47 days (from 27 April to 12 June of 2018) in an air-conditioned room, with three versions of Figaro® sensors: TGS 2600, TGS 2611-C00 and TGS 2611-E00. Details of the data acquisition are as described in the previous section. Reference data was again provided by a Picarro G2401 gas analyzer. Air temperature and relative humidity was measured using a DHT22 digital sensor (Aosong Electronics) which has an accuracy of $\pm$0.5 °C and $\pm$0.5% RH respectively. The sensors were installed in a semi-open enclosure from which the Picarro CRDS took its intake, thus sampling the same air that the Figaro®s̃ensors.

2.2.3 Modeling CH_4 from Figaro Resistances and Other Predictors

Low-cost sensors, generally, present a non-linear dependency on environmental variables causing cross-sensitivities (Collier-Oxandale et al., 2018). There is no mathematical model of the relationship between the resistances and CH_4, given the dependency of resistances on other environmental variables (CO, H_2O, pressure and temperature). The analytical problem thus remains nonlinear and multi-dimensional. Therefore, an Artificial Neural Network model (ANN) was chosen to reconstruct CH_4 from observed time series of resistances, CO, H_2O, pressure and temperature. We chose a Multi Layer Perceptron (MLP) which is a classical supervised-based algorithm (Rumelhart et al., 1986). MLP models are generally considered to be the reference among machine learning methods because several theoretical results prove their ability as a universal approximator (Cybenko, 1989; Hornik et al., 1989), capable of learning from examples. For our problem, the advantages of a machine learning model such as MLP are the following: (i) it does not require any prior knowledge about I/O dependencies, (ii) it is able to construct arbitrary functions from noisy data (Bishop, 1995b), it makes no assumption on the distribution of data (Gardner and Dorling, 1998), and (iii) could produce reasonable outputs from entries that are not present in the learning set, i.e., generalization (Haykin, 1994). Over the past decade, deep networks such as MLP have demonstrated superior performance over a wide variety of tasks,

including function approximation. Recently, MLP have been proven to be more efficient than inverse linear methods in reconstructing the signals of trace gas species from low-cost sensors (Casey et al., 2019).

In a MLP model, unknown parameters (i.e., architecture and connection weights) are adjusted in order to obtain the best match between a dataset of model inputs (Figaro resistances, H_2O, CO, Temperature and Pressure) and corresponding outputs (Figaro CH_4). The connection weights are adjusted by using iterative learning processes such as the backpropagation (Rumelhart et al., 1986) or several algorithms that have been developed in order to achieve a good learning of the model (i.e., Stochastic gradient descent, Adam, etc., (Géron, 2019)). In our study, we chose to use a quasi-Newton method, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, which provide the optimal MLP weights in a limited number of iterations (300) due to its relatively fast convergence (Haykin, 1994). For our study, the architecture of the MLP producing the best results was found to be a four-layer network with 5 units in the input layer, 14 and 19 units with tanh activation function in the hidden layer and 1 unit with a linear activation function in the output layer. All models were constructed using the library scikit-learn (Varoquaux et al., 2015) on python 3.6.

The generalization error, also called test error, is the expected value of the error produced by new inputs (Goodfellow et al., 2016). This error is obtained from the performance of the MLP to match an independent test dataset. A central challenge of function approximation by MLPs is the risk of underfitting and overfitting (Goodfellow et al., 2016). Underfitting is referred to a high training error when, because of an inconvenient architecture or because of training inputs that are not explanatory enough for example, the MLP does not manage to efficiently fit the training data set. In our case, the risk of underfitting is mitigated by using a sufficiently complex model. Overfitting happens when the MLP learns features from the training data set, e.g., noise or biases, that are not relevant and do not generalize well to a different data set. To reduce the risk of overfitting we used the weight decay regularizer or L2 Norm that drives excess weights (weights of the network that does have little or no influence in the model) to values close to zero Haykin (1994). We use also an early stopping technique that constantly monitors the error produced

by the model with respect to an independent validation data set (validation error) during the learning process. When the validation error starts increasing, the training process is stopped in order to moderate the generalization error (Bishop, 1995b; Goodfellow et al., 2016). The best MLP model was selected as the one producing the lowest validation error from the results of many tests in which the number of neurons of the hidden layer were varied (see Section 2.2.4.4).

2.2.4 Results

2.2.4.1 Sensitivities of Low-Cost Sensors

To account for the systematic error of the Figaro® resistances sensitivities to CH_4 and H_2O, caused by the different CO levels in the two CH_4 target tanks, the sensors' sensitivities to CO and H_2O were separately measured and used to correct the data. In Figure 2.2, for each Figaro® type, the upper plot shows the measured voltage across the load resistor for changing CO mole fraction, at each of the three humidity levels 0, 1 and 2.3% mole fraction. The lower plot shows how much the Figaro® voltages increased above the baseline voltage, where the baseline is the voltage at zero CO and is a function of the humidity.

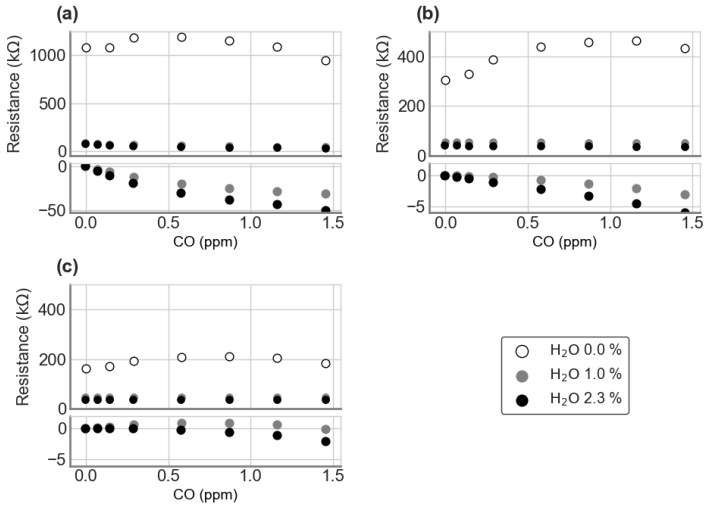


Figure 2.2: Measured sensitivity of Figaro® sensors (**a**) TGS 2600, (**b**) TGS 2611-C00 and (**c**) TGS 2611-E00 to CO, at different humidity levels. Upper plots show the measured resistance, while lower plots show the contribution to the resistance due to CO.

For each sensor type, these data were fitted with a multivariate quadratic model of the form:

$$f(x_1 = CO, x_2 = H_2O_{MoleFraction}) = ax_1 * x_2 + bx_1^2 + cx_2^2 \tag{2.1}$$

The voltage contribution due to CO was calculated as $f(x_1 = CO, x_2 = H_2O) - f(x_1 = CO = 0, x_2 = H_2O)$, where f is the fitted value. This model was used to correct the CH_4-H_2O voltage data to zero CO by subtracting the CO contribution from each point. These corrected values are show in Figure 2.3, converted to resistances.

The range of CH_4 mole fraction was 2 to 9 ppm, thus being larger than in the room air experiment where CH_4 varied only from 1.95 to 2.5 ppm. The range of H_2O mole fraction was from 0.5% to 2.5%, which is comparable to that of the room air experiment. In general, the sensors resistance presents a strong sensitivity to H_2O and a small sensitivity to CH_4. The TGS 2611-C00 is slightly the most sensitive sensor to CH_4, with a slope of -1.85 kΩ/ppm CH_4 at 1% H_2O.

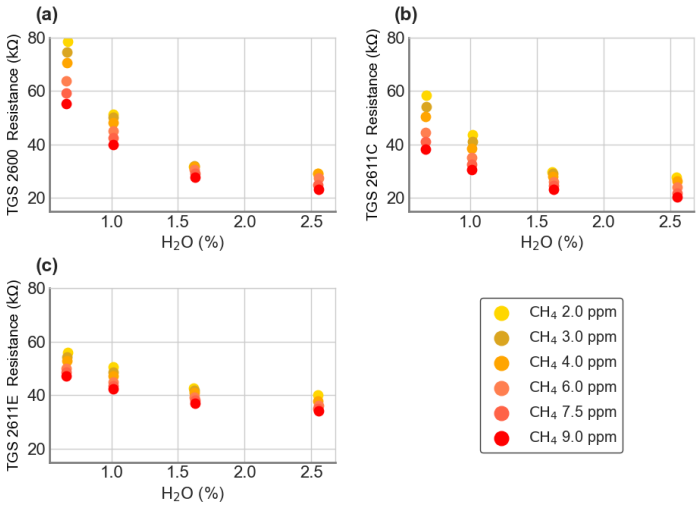


Figure 2.3: Resistances of Figaro® sensors (**a**) TGS 2600, (**b**) TGS 2611-C00 and (**c**) TGS 2611-E00 calculated from load voltages corrected for the cross-sensitivity to CO.

2.2.4.2 Data Pre-Processing for MLP Model

The data pre-processing scheme for the MLP model is summarized in Figure 2.4. We filtered the input and output data by removing NaN values and observations by an unknown source in the room that resulted in clear spikes of the TGS resistances. This resulted in 49,103 observations at

a resolution of 1-min or 34 days of measurements. For each of the independent input variables, a low pass Savitzky-Golay Filter has been applied to remove high frequencies corresponding to fluctuations of sensor measurements (1 min variations) and a median filter has been used to remove the effect produced by the low pass filter on gaps (Press and Teukolsky, 1990). These filters have been not applied to the output data because the CH_4 observations provided by the Picarro are not characterized by high frequency variations and because we wanted to keep the output data used for the training of the MLP as close as possible to the original data set. Figure A.1 shows a comparison between the raw signal and the filtered signal for one day of data. Because input variables have different units and scales that could affect the relative sensitivities of the MLP with respect to each variable, they were normalized with a robust scaler which considers the statistical dispersion of the observations by, removing the median and scaling the data according to a quantile range (Hagan et al., 2014). We used this scaler in order to prevent that outliers could affect the relative importance of each variable in the model, with that filtered dataset, we created two sub-sets of data to train and evaluate the MLP model. A training set always contained 70% of the entire dataset, the remaining 30% being the cross-validation dataset.

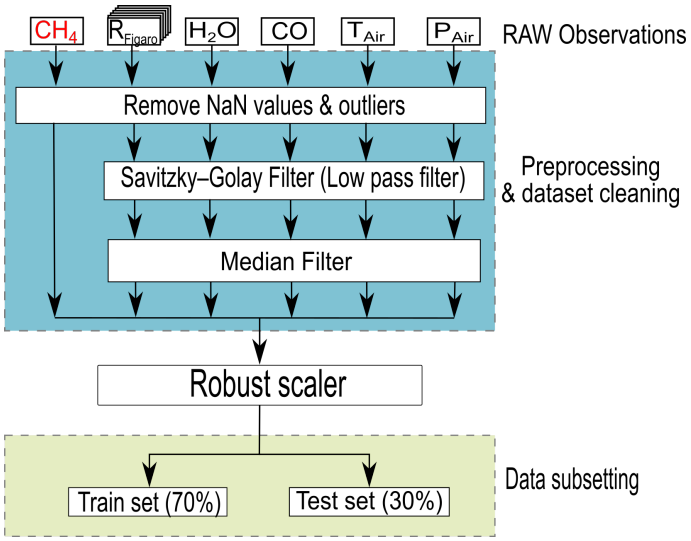


Figure 2.4: Data preprocessing and sub setting for the training and cross-validation of the Multi-Layer Perceptron model.

2.2.4.3 Room Air Measurements

In Figure 2.5 are shown the smoothed time series on a time step of 1 min after a low pass filtering (see Section 2.2.4.2) of room $\mathrm{CH_4}$ from the CRDS analyzer, resistances from each Figaro® CO from CRDS, room temperature of the DHT22 sensor, and pressure of the BMP180 sensor. $\mathrm{H_2O}$ mole fraction (in %) was computed from relative humidity and temperature of the DHT22 sensor and atmospheric pressure of the BMP180 sensor using Rankine's formula:

$$H_2O_{Mole\ Fraction} = 100 \times \left(\frac{\frac{RH}{100} \times e^{\frac{13.7-5120}{T+273.15}}}{\frac{P}{100000} - \frac{RH}{100} \times e^{\frac{13.7-5120}{T+273.15}}} \right) \tag{2.2}$$

where *RH* is the relative humidity in %, *P* the atmospheric pressure in Pa and *T* the temperature in °C.

The large spikes in the variations of temperature is due to the room air conditioning regulation system. Table 2.1 summarizes the principal statistics of the dataset.

Table 2.1: Summary of the statistics for each variable in the dataset.

	$\mathbf{CH_4}$ (ppm)	$\mathbf{TGS\ 2600_{01}}$ (Ω)	$\mathbf{TGS\ 2600_{02}}$ (Ω)	$\mathbf{TGS\ 2611-C_{01}}$ (Ω)	$\mathbf{TGS\ 2611-C_{02}}$ (Ω)
# of Obs.	49,103	49,103	49,103	49,103	49,103
mean	2.12	32,356.48	32,487.65	47,193.12	49,262.97
σ	0.11	5948.07	5969.96	4352.56	4891.01
min	1.94	18,446.51	18,871.92	37,504.39	37,768.43
max	2.45	47,262.67	47,418.24	57,590.59	60,616.80
25%	2.03	28,881.76	28,848.63	44,136.81	45,890.21
50%	2.10	31,584.97	31,633.92	46,706.42	48,884.94
75%	2.18	34,994.97	35,015.34	49,233.17	51,842.68
σ_{Rel}	5.35	18.38	18.38	9.22	9.93

	$\mathbf{TGS\ 2611-E_{01}}$ (Ω)	$\mathbf{TGS\ 2611-E_{02}}$ (Ω)	$\mathbf{H_2O_{Mole\ Fraction}}$ (%)	CO [ppm]	T (°C)	P (Pa)
# of Obs.	49,103	49,103	49,103	49,103	49,103	49,103
mean	60,425.14	63,378.21	1.58	0.11	25.53	99,709.67
σ	3010.45	6234.00	0.27	0.02	0.46	420.74
min	52,472.35	54,468.19	1.07	0.08	24.11	98,289.72
max	79,018.36	93,671.74	2.07	0.24	27.15	100,528.79
25%	58,255.57	59,549.05	1.38	0.10	25.29	99,406.22
50%	60,227.14	61,428.60	1.52	0.11	25.52	99,698.57
75%	61,792.62	64,557.91	1.87	0.12	25.74	100,004.34
σ_{Rel}	4.98	9.84	17.17	18.38	1.81	0.42

Before applying MLP models to reconstruct the (CRDS) reference $\mathrm{CH_4}$ time series, we analyzed the partial correlations between resistances and other predictors. The correlation matrix is show in Figure 2.6. This first analysis of linear correlations does not capture non-linear sensi-

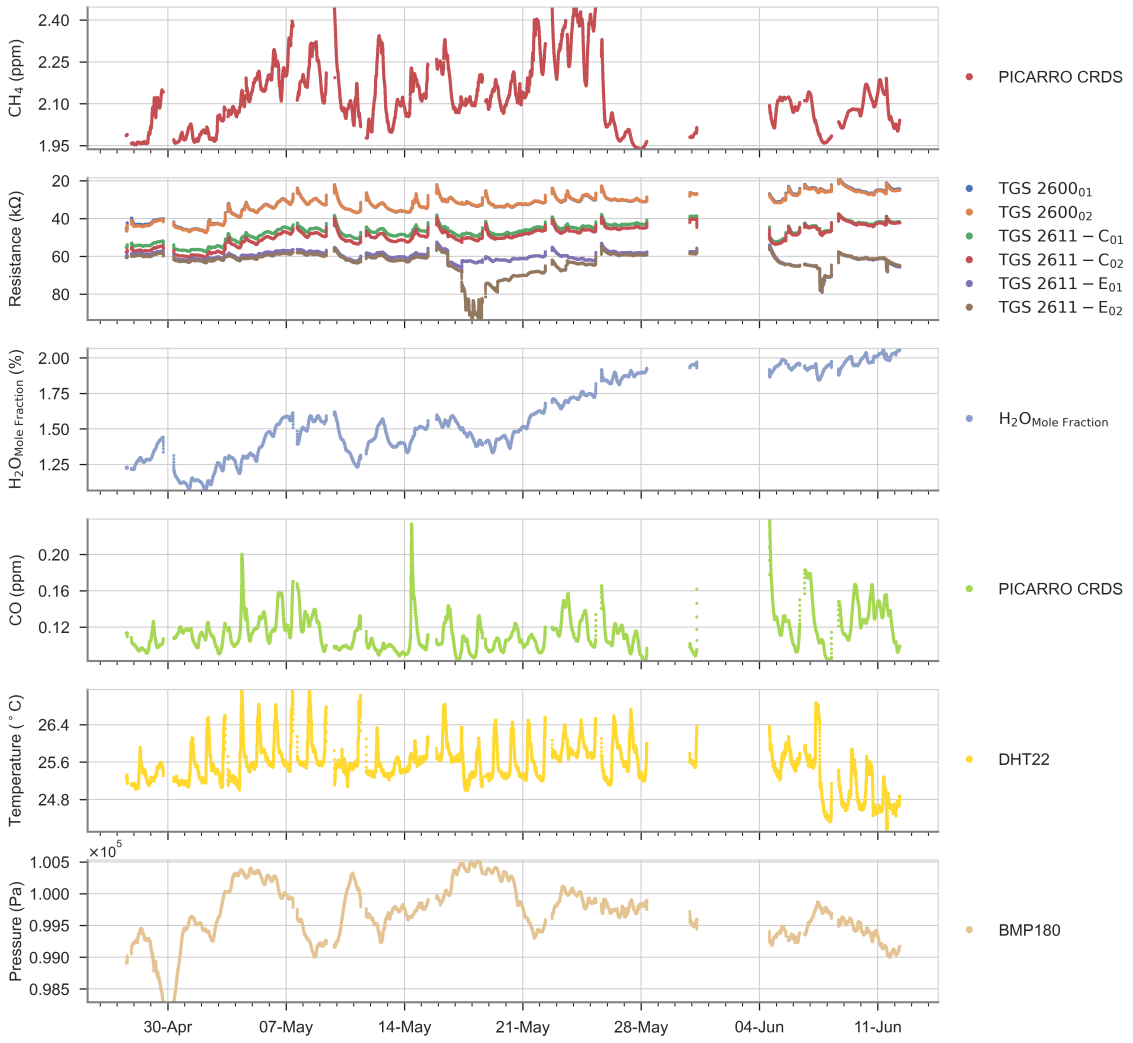


Figure 2.5: Time series of gas mole fractions, Figaro® sensors' resistances, temperature and pressure in room air during the room air experiment. The data was filtered as explained in Section 2.2.4.2.

tivities of Figaro® resistances to CH_4 and other predictors, but it is performed in order to identify the most influential predictors of CH_4 (i.e., those showing a higher positive or negative correlation with CH_4), the sign of the sensitivities (i.e., the sign of the correlation coefficient) and how stable is the influence of different predictors on CH_4 during the 47 days duration of the experiment (variability of correlations in time between 3-h intervals during the measurement period). The data in Figure 2.6 shows the upper triangular part the correlation matrix between variables, and in the lower triangular part the standard deviation of the correlation computed on bins of 3 days previously smoothed on 3-h intervals (the temporal resolution at which MLP is trained and applied). A stationary correlation would give a standard deviation close to zero. The Figaro®

resistances presented weak partial correlations with CH_4, the target variable, with r values of 0.25 and -0.27 for TGS 2600, 0.015 and -0.0098 for TGS 2611-C00, and -0.14 and 0.099 for TGS 2611-E00 types, respectively. Other variables also presented weak correlations with CH_4 of -0.17 (H_2O mole fraction), 0.16 (CO), 0.21 (T) and -0.18 (P), respectively. Correlations between of the resistances of two versions of sensors of the same type were strong for TGS 2600 (1) and TGS 2611-C00 (0.98) but not for TGS 2611-E00 (0.3). There were also appreciable strong correlations, positive and negative, between resistances of sensors from different types, in particular between TGS 2600 and TGS 2611-C00 types. The resistance of TGS 2611-E00 showed a weak correlation with TGS 2600 and a stronger one with TGS 2611-C00 (Figure 2.6). Again, as for the correlations with CH_4, we found a clear difference in the correlations of resistances from TGS 2600 and TGS 2611-C00 types on one hand, and from TGS 2611-E00 on the other. The resistances of all sensors were negatively correlated with H_2O on one version, with r values of. -0.31, -0.61 and -0.32 for version 2 of TGS 2600, TGS 2611-C00 and TGS 2611-E00 types, respectively. However, version 1 of the same sensors was instead positively correlated with H_2O (Figure 2.6). We also found weak correlations of the resistances with CO, but still larger in absolute values than the correlations with CH_4.

From the stability analysis (see lower triangular in Figure 2.6) we observe that most of the r (correlation) values are under 0.5 meaning that most of the data are consistent during a window of 3 days. We also noted relatively high values of r on the two versions of the TGS 2611-E00 with the TGS 2611-C00 (0.59), the H_2O with the two versions of TGS 2600 (0.64) and with the two versions of TGS 2611-C00 (0.63), high values are also present on the pressure with the CH_4 (0.6) and with the H_2O (0.59).

We conclude from this first analysis of correlations that, although there are correlations, between CH_4 and the resistances of TGS 2600 and TGS 2611 sensors, such correlations are small and vary with time, which will make it challenging to reconstruct CH_4 time series with linear models, and justifies a priori our choice of a MLP model. Same conclusions have been drawn from previous studies (Casey et al., 2019; Collier-Oxandale et al., 2018).

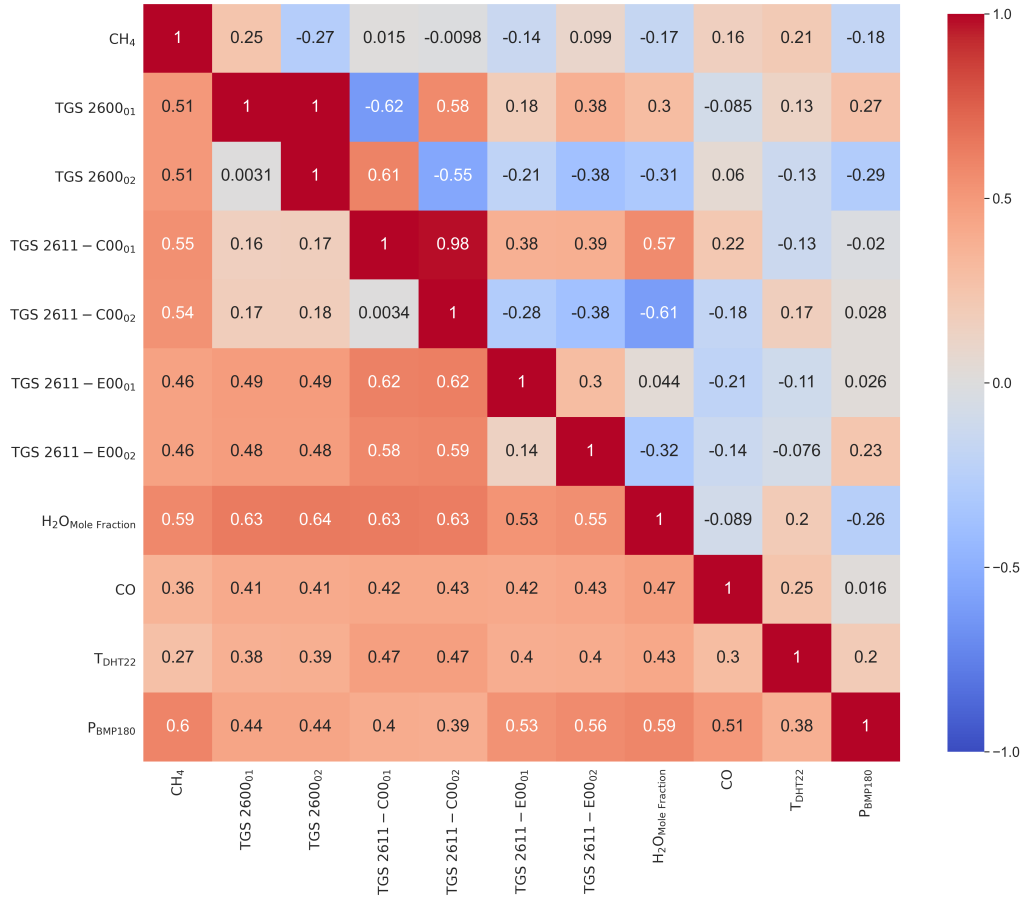


Figure 2.6: Partial correlation (r) matrix (upper triangular) and standard deviation of correlation for bins of 3 days previously smoothed at 20 min scale on 3 consecutive hours (lower triangular).

2.2.4.4 Evaluation of the MLP Model

To assess the influence of the choice of the training period in the performance of the MLP, we defined over the whole data set 50 sliding training periods which contain 70% of the observations. The corresponding test sets contain thus 30% of the observations and the results associated to the fits of the 50 MLPs are described in Figure 2.7.

The evaluation of the performance was based on 4 metrics: the RMSE on hourly data, the mean bias, the ratio between the spread of the predicted outputs from the model and the spread of the true values ($\sigma_{Model}/\sigma_{Data}$) and the correlation coefficient between the output of the model and the true values. On Figure 2.8 are presented two examples of the performance of the MLP with different test periods, selected to represents a bad (period 50) and a good test (period 7) performance.

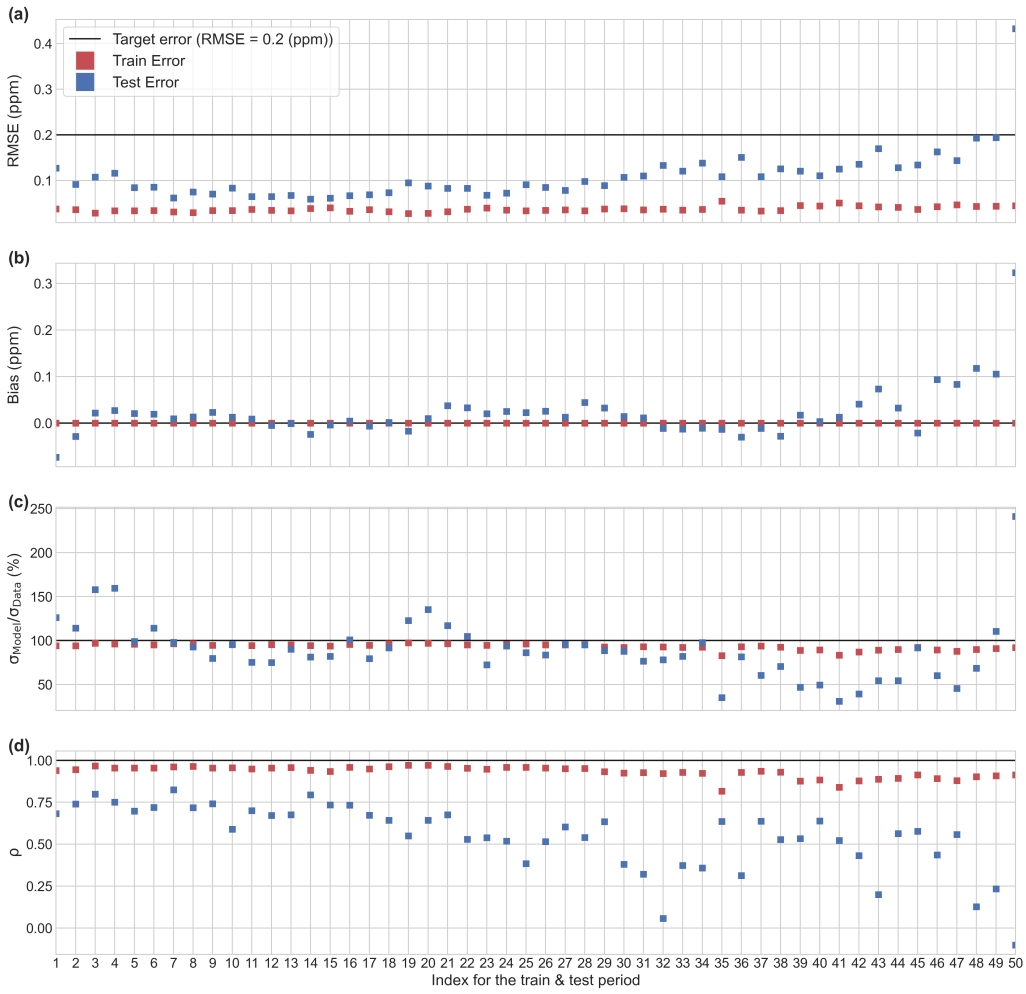


Figure 2.7: Performance of the MLP model for the 50 training and test periods. (**a**) RMSE on hourly data. (**b**) Mean bias. (**c**) Ratio between the spread of the predicted outputs (σ_{Model}) and the spread of the true values (σ_{Data}). (**d**) Correlation coefficient between the predicted outputs and the reference values.

In general, the RMSE, on the test set, was less than 0.2 ppm, except for one period in the end of the time series (50th). This value of RMSE meets our precision requirement target of 0.2 ppm posed in the introduction. The periods of lowest errors are periods 5 to 20 (from 30 April to 29 May), and we can observe that the worst case corresponds to a model learned only for low values of $\mathrm{H_2O}$ and tested on a test set which contains much higher values of this input variable. Likewise, many low temperature values are observed in the test set, while these values are missing in the learning set. In the best case, on the contrary, the ranges of variation of the input variables are narrower in the test set than in the set used for learning. (see Figures A.3 and A.4). A better performance in the model was observed when using the TGS 2611-C00 sensor. From the test

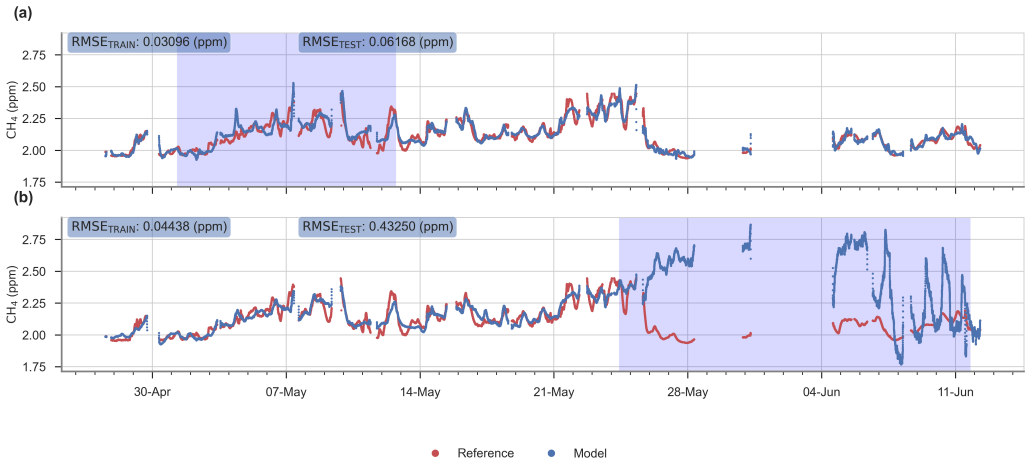


Figure 2.8: Time series showing a good (**a**) and a bad (**b**) performance of the MLP model for the test period 7 and 50 respectively. In red, time series of the reference instrument, and in blue the reconstructed signal given by the MLP model. White background: observations used for the training stage. Blue background: observations used for the test stage.

periods 31 to 50 we observed a larger increase of the test error than in the previous periods, the worst case being test period 50 (RMSE > 0.4 ppm).

Misfits are mostly due to a wrong simulation by the MLPs of the variability and/or the phasing of the data. For the periods 2, 3, 19 to 21 and from 35 to the end of the series, the difference between the standard deviations of the data and the MLP can reach 50% which points out a incorrect reconstruction of the amplitudes of the test data sets by the corresponding MLPs. For all the periods, the MLPs face difficulties to reproduce the phasing of the test data sets: the mean correlation coefficient is indeed of 0.54 and from the periods 30 to the end of the series, the phasing between the MLP and the observations is notably deteriorated ($r < 0.5$). At the contrary, the MLPs simulate correctly the average values of the data: the mean bias is under 0.1 ppm which represents less than 1% of the average values of the data. This is likely explained by a tendency of the MLP to act as a low-pass filter of the data: during the process of learning, the weights are indeed adjusted in order to minimize the misfits, over the entire training data set, between the MLP outputs and the Picarro $\mathrm{CH_4}$ data which thus favors a correction of the misfits for the low frequencies.

In summary, the results of the MLP model point out to several critical aspects. In the first place, the selection of the training period affects the performance of the model to reconstruct

CH_4, in which the covering of the same range of values in the training and test set is traduced by good performance. The choice of a 'good' training period results in higher cross validation scores. In particular, the period of CH_4 variation from 30 April to 29 May appears particularly critical and if it is not used for the training, the model cannot extrapolate the CH_4 data well (RMSE > 0.4 ppm). An overlapping of the data distribution of the training set and the test set, in the sense of similar variations observed in both sets, increases the performance of the model (see Figures A.3 and A.4). Secondly, we found that the model cannot reproduce well the magnitude of high frequency anomalies of CH_4, but tends to better reproduce the low frequency component of the signal which is consistent with low-pass filter behavior of the MLP described above (see Figure A.5).

2.2.4.5 Sensitivity of MLP Model to Input Variables

To understand the relative contribution of each input on the MLP model, the sensitivity to the number of inputs and to the number of TGS sensors were calculated and results are shown in Figure 2.9. For every case we compared the results with a reference model that has as inputs the Figaro TGS 2611-C00 resistance, H_2O mole fraction, Air temperature and pressure, and CO, corresponding to our best case from the previous section. For every configuration we trained 50 models in the same way as described in Section 2.2.4.4. For all the tests we kept the same MLP architecture of 4 layers with the same number of units for the hidden layers, and compared their performance for training datasets using the root mean square error (RMSE) of hourly data. For an identical architecture, the number of input variables has consequences on the complexity of the model (total number of parameters of the model) and therefore on the overfitting effects already mentioned.

We observe in Figure 2.9a that omitting the air pressure results in a better performance on the test set compared with the omission of other variables. On the other hand, omitting H_2O decreased the performance on the training periods RMSE increased from 0.036 to 0.046 ppm for the training set, and from 0.12 to 0.13 ppm for the test set. Omitting temperature gave a better performance on the test set, and omitting CO led to no appreciable differences. The Figure 2.9b shows the effect of increasing the number of Figaro® sensors in the MLP model. For this analysis,

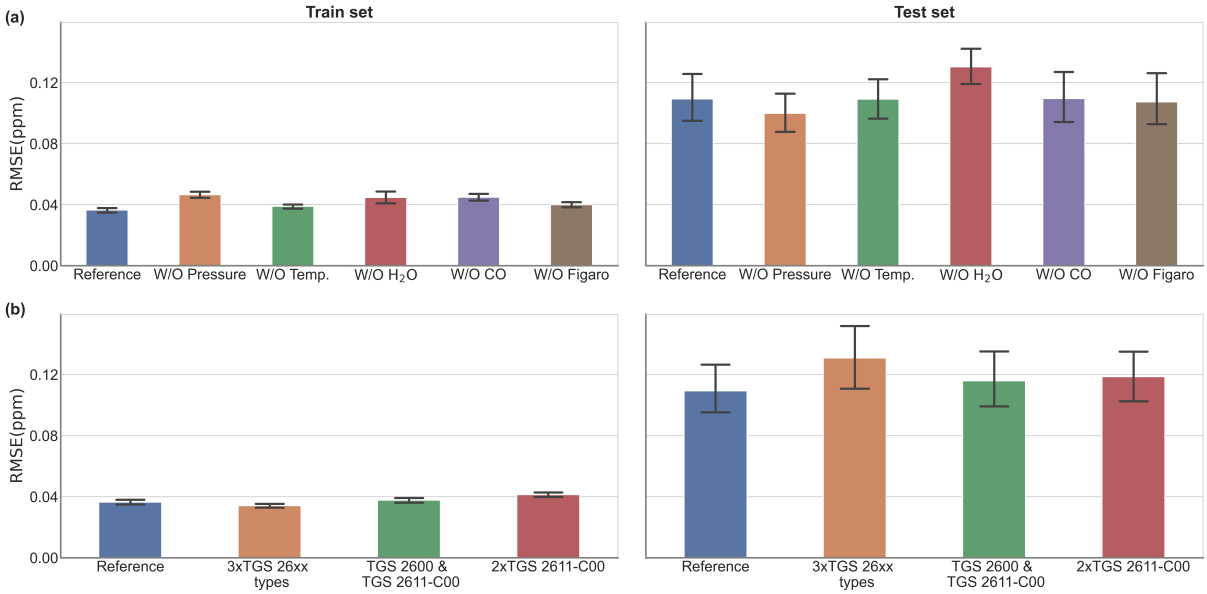


Figure 2.9: (**a**) Comparison of 5 models in which one input has been removed at each time (denoted 'W/O') with the reference model that has been built with the resistance data of the Figaro TGS 2611-C00 and the 4 other types of data (Reference). (**b**) Effect of increasing the number of Figaro® sensors in the model with no modification of the ambient variables in the input. The bar plot represents the mean error for every configuration and the error bar on top is the range of variation over the 50 validation periods.

the four environmental variables (Temperature, pressure, CO, and water vapor mole fraction) and four different combinations of resistances data from the three types of Figaro® sensors were used to train the MLP with one combination corresponding to the reference model (blue bar). A decrease in the performance (RMSE ∼0.15 ppm) was found in the test set when using resistances from three types of Figaro sensor: the TGS 2600, the TGS 2611-E00 and the TGS 2611-C00, brown bar. Using two versions of the same sensor decreased the MLP performance in the test set (RMSE ∼0.133 ppm for a combination of two TGS 2600 and TGS 2611-C00 and RMSE ∼0.13 ppm for two versions of TGS 2611-C00) in relation to the reference MLP model. For the training set, we found a decrease on the training error when using 3 different types of sensor (RMSE = 0.034 ppm). We found that all the models of this sensitivity test are matching our requirement as presented on Table A.1.

Several conclusions can be drawn from those tests using a reduced numbers of predictors. Firstly, the water vapor mole fraction affects significantly the predictive power of the model, and removing this variable produced a larger spread of the test error. This result is consistent with the

large sensitivity of resistance to H_2O shown in Figure 2.3 and indicates that this variable should be measured together with Figaro® resistances if using machine-learning models to reconstruct CH_4. An interesting result was that when CO is removed from the model predictors, the MLP performance even slightly improved. Using three types of sensor data in the training of the model increased the spread of the test error, that affected the stability of the model, because of the inconsistent information from different sensors, in particular the different behavior of type TGS 2611-E00 compared to the two others. Using data resistances from two versions of the same type of sensor degrades as well the model performance.

In addition to testing the sensitivity of the MLP results to the choice of predictors and sensors type, we analyzed the partial dependence of the target variable CH_4 to the marginal effect of each predictor in the reference model with 5 predictors included. The corresponding partial dependence plots were constructed using the scikit-learn package on python 3.6 (Varoquaux et al., 2015). The results for the type TGS 2611-C00 are show in Figure 2.10. In the case of the resistances of Figaro® sensors, we found a negative sensitivity of MLP-reconstructed CH_4 to resistances. This result is in qualitative agreement with the negative sensitivities measured experimentally in Figure 2.3. We could not compare however the values of experimental sensitivities with those inferred by the MLP because the range of CH_4 was much larger in the experiments (2 to 9 ppm) than in the room air dataset (2 to 2.5 ppm). Nevertheless, we noticed that the experimental sensitivity shown in Figure 2.3, ranging from -0.7 to -2.8 kΩ per ppm CH_4 over a 7 ppm CH_4 is much smaller (about twenty times less) than the sensitivity diagnosed from MLP partial dependence analysis in Figure 2.10. The reasons for this discrepancy may be due to sensors aging or to an over-estimation by the MLP model. The partial dependence of MLP-reconstructed CH_4 to H_2O showed a different behavior between 'good' and the 'bad' test datasets as seen in the data from Figure 2.10. For the good training period, the sensitivity to H_2O was rather constant and even slightly positive for H_2O values going from 1.2% to 1.6%, then a negative sensitivity was found until 1.9%. For the average of all training periods, the H_2O sensitivity peaks and declines with a humped shape curve reaching a maximum at 1.6%. The positive sensitivity below 1.6% is consistent with experimental sensitivities of Figure 2.3, in which the resistance decreases for an

increase of H_2O and decreases when CH_4 increases. Since some values are not or hardly represented in the training set for the worst case (see the comment in Section 2.2.4.4 for the variable H_2O) the humped-shaped dependence of CH_4 to H_2O in Figure 2.10 may be linked to poor MLP learning in certain ranges of values. We found from the partial dependence analysis a positive sensitivity of reconstructed CH_4 to CO. For temperature and the pressure, we found a quadratic shaped sensitivity for the worst case, with negative curvature for temperature and positive for pressure.

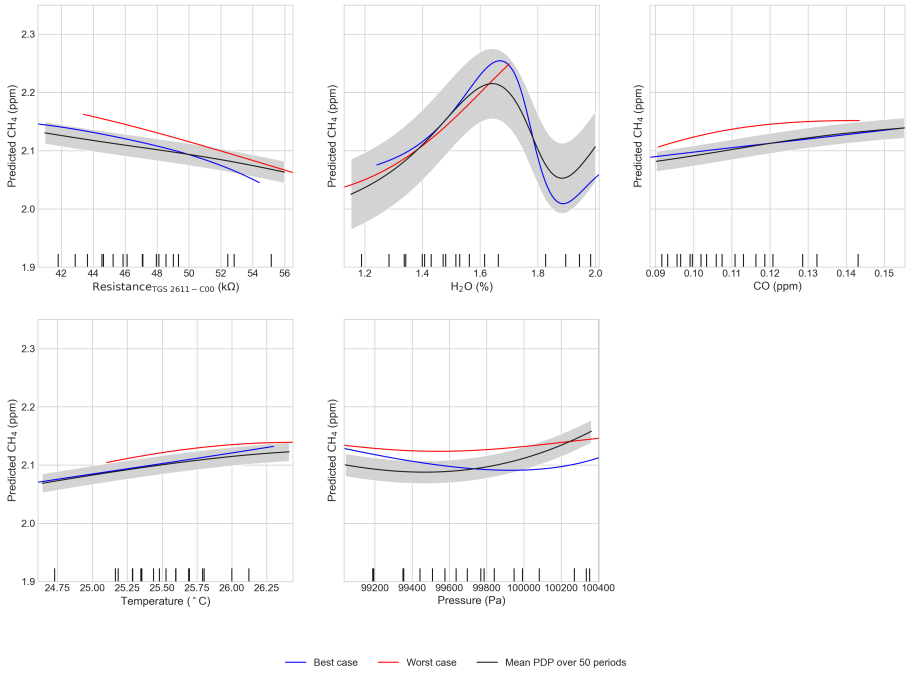


Figure 2.10: Partial dependence plot for the best (blue) and worst (red) case and mean Partial dependence plot computed over the 50 periods (black), the shaded gray area is the uncertainty (1σ) for the 50 periods. The inputs of the model were the Figaro 2611-C00 resistance, water vapor mole fraction, CO, air temperature and pressure. Ticks on the x axes of the figures are the deciles of the input variables.

The bivariate partial dependence plots in Figure 2.11a show the dependence of the MLP-reconstructed CH_4 on the joint values of resistance and the other variables for the best test case (see Figure 2.11b for the worst case). On this best case, we observe that there is a high dependence to resistance of the MLP-reconstructed CH_4 for values of H_2O under 1.6%, whereas for values between 1.6% and 1.8% the dependence flattens off. Considering CO and temperature, we found that the MLP model is highly dependent of resistances for values under 0.15 ppm of

CO and temperature under 26.5 °C, for the best and the worst cases for CO. Finally, the model seems to be sensitive to pressure when the resistance varies over 48 kΩ and under 44 kΩ.

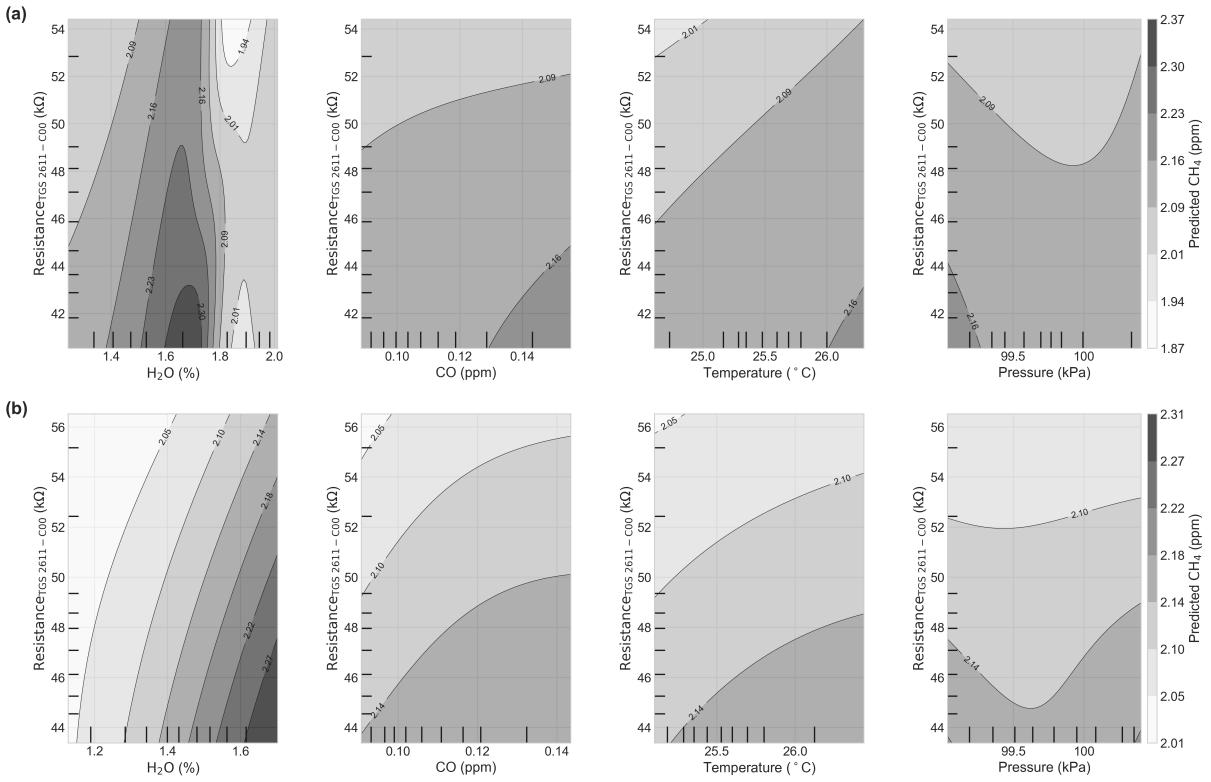


Figure 2.11: Bivariate partial dependence plot for the TGS 2611-C00 sensor versus H_2O mole fraction, CO, air temperature and pressure. (**a**) Partial Dependence Plot (PDP) for the model trained in the best case scenario and (**b**) Partial Dependence Plot (PDP) for the worst case scenario. Ticks on the x and y axes of the figures are the deciles of the input variables.

2.2.5 Discussion

Few studies have tried to use machine-learning models to reconstruct the variability and the concentration of greenhouse gases from low cost sensors (Shahid et al., 2018). In the work of Spinelle et al. (2015,?, 2017)) several field calibration methods for low cost sensors were explored: Linear and multilinear regression and Artificial Neural Network (ANN), for five trace gases (O_3, NO_2, NO, CO and CO_2) measured by metal oxide, electrochemical and miniaturized infrared sensors over five months. They concluded that the best calibration method was ANN and that the use of different types of sensors could help the ANN to solve the cross sensitivities. Here, we found that only using Figaro® TGS sensors, even from different types, it was not possible to make a good

reconstruction of the signal without concurrent measurements of other environmental variables, because the sensors had a high cross sensitivity to water vapor, aggravated by differences on the distribution of the H_2O density for the training and test set.

The study of Esposito et al. (2016) compared the performance of feed forward neural networks (FFNN) with dynamical neural network (DNN) in the calibration of three trace gases (NOx, NO_2 and O_3) measured with low cost sensors over five weeks. They found that DNN was significantly more accurate than FFNN in the reconstruction of high variations of concentrations. As explained on Section 2.2.3, the high capacity problem in which the more complex models tend to overfit needs to be treated carefully, thus we tested a series of combinations of number of units and number of layers obtaining an architecture of 2 hidden layers the more adapted to this problem, our limited dataset also restraints a more complex model.

Cordero et al. (2018) worked on a two-step calibration process of NO_2, NO and O_3 from low cost sensors. They applied a first multilinear regression considering all the predictors, then the error of the multilinear regression was introduced as an input, in addition to the others predictors, to a supervised machine learning algorithm (Support Vector Machine—SVM, random forest or ANN) to reconstruct the concentration of trace gases. They concluded that globally SVM and ANN performed well in the reconstruction of the concentrations in all the cases over a threshold (40 $\mu g/m^3$). For data below that threshold, the random forest was the best model to reconstruct the signal. As a universal approximator, we decided to use MLP in our study for the reconstruction of small variations of CH_4 measurements at levels around ambient air values, and we did not test the limits of this type of model in presence of high variations of our signal, such as CH_4 spikes of several ppm encountered when measuring air at a point nearby an industrial site. This question remains open for a future study with a specific dataset of CH_4 that contains spikes.

Casey et al. (2019) compared the performance of direct linear models, inverse linear models and ANN models over three months of data of ambient air in a region influenced by oil and gas production. Their main results pointed that the ANN model, when applied to CH_4 and CO observations, gave better performance (RMSE = 0.13 ppm over a month) than the direct and inverse linear models, due to the smaller dynamic range from their observations. For our study,

a linear model could not be applied due to nonlinear relationships between predictors with the target CH_4 signal. With a careful selection of the MLP model, our results indicate that the MLP model provided performances that meet our target requirement of an error of 0.1 to 0.2 ppm for hourly average CH_4, except during periods when the distribution of training data was too different from the one of the test data (80% on the last test period of the cross validation). This illustrates the critical aspect for MLP and other machine learning models to use large datasets, with all the space of predictors being covered by training datasets, to reach good cross validation performance.

Eugster et al. (2020) conducted a long term evaluation of the Figaro® TGS 2600 over seven years at Toolik Lake in Alaska; they proposed a multilinear model to calibrate the voltage signal from the sensor including other environmental variables such as air temperature and absolute humidity. The calibration methods were assessed under summer and winter conditions and compared their proposed model with an ANN. Eugster et al. (2020) found satisfying agreement on 30 min average observations for the multilinear model ($R^2 = 0.424$). They reported a more balanced performance of the ANN on cold conditions (winter), but they not find a substantial difference between their proposed model and the ANN. They concluded that ANN would outperform linear models if other driving variables were included to the model.

Riddick et al. (2020) conducted an experiment to investigate the potential of Figaro® TGS 2600 to measure CH_4 mixing ratios in ranges between 2 and 10 ppm, assess the long term measurements over 3 months and estimate the emissions from a natural gas point source. Calibration of the sensor was derived from a non linear relationship giving the best agreement with the reference measurements when computing the time averaged concentration with a uncertainty of $\pm$0.01 ppm. The authors observed that reliably measurements of CH_4 was in the range of 1.8 to 6 ppm and suggest that calibrations need to be derived for each individual sensor.

From the results of the sensitivity tests to removing predictors one at a time, and the partial dependence plots providing the sensitivities of the MLP modeled CH_4 to individual predictors we could observe the importance of the water vapor as a critical input for the models. This is mainly due to the high sensitivity for the TGS sensors to H_2O confirmed by our experimental data.

Variations of H_2O in the field are typically larger than the ones covered by our experiment and they have an important impact on the model's performance. Refining models to further separate the H_2O and CH_4 signal will be needed to meet the target error when increasing the range of H_2O and CH_4 variations in future experiments. For the temperature, pressure and CO we found that those predictors have a lower influence on CH_4 in our room air dataset, and for similar type of data, they could be ignored as concurrent measurements. The influence of CO on the model should be studied in depth as well as that of other cross-influencing compounds being electron donors such as ethane, hydrogen or H_2S, whose concentrations in industrial environments are likely larger than the ones during our idealized experiments. This is the second critical topic that we should address in our following assessments of low-cost tin-oxide sensors.

2.2.6 Conclusions

The theoretical contribution of this study is to demonstrate the potential of Artificial Neural Networks models for the reconstruction of atmospheric CH_4 variations based on tin-oxide sensors resistances, within a small CH_4 variation range around mean levels similar to current atmospheric concentrations, achieving a target RMSE $\leq$ 0.2 ppm. The selection of the training and test periods was shown to be a critical factor to obtain good performance, because our dataset was relatively short and some training periods included predictor distributions that strongly differ from that of the test periods. The practical contribution of this study is a detailed characterization of CO and H_2O cross influences on tin-oxide sensors resistances, from laboratory tests. We also found that adding different combinations of Figaro tin-oxide sensors versions did not produce better results. Using only the TGS 2611-C00 sensor version led to better results in regard to the others types.

Chapter 3

Reconstruction of high-frequency atmospheric methane concentration peaks using metal oxide low-cost sensors

Rodrigo Rivera Martinez[1], Diego Santaren[1], Olivier Laurent[1], Gregoire Broquet[1], Ford Cropley[1], Cécile Mallet[2], Michel Ramonet[1], Adil Shah[1], Leonard Rivier[1], Caroline Bouchet[3], Catherine Juery[4], Olivier Duclaux[4] and Philippe Ciais[1].

[1] Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
[2] Université de Versailles Saint-Quentin, UMR8190 – CNRS/INSU, LATMOS-IPSL, Laboratoire Atmosphères Milieux, Observations Spatiales, Quartier des Garennes, 11 Boulevard d'Alembert, 78280 Guyancourt, France
[3] SUEZ - Smart & Environmental Solutions; Tour CB21, 16 place de l'Iris, 92040 La Défense France
[4] TOTAL Energies - Raffinage chimie, Laboratoire Qualité de l'Air – 69360 SOLAIZE FRANCE

The manuscript is currently in open discussion in the journal of Atmospheric Measurement Techniques, Atmos. Meas. Tech. Discuss. [preprint], https://doi.org/10.5194/amt-2022-200, in review, 2022.

Date of submission: 29 June 2022. Discussion started on: 08 July 2022.

3.1 Summary of the publication

3.1.1 Introduction and motivation

CH_4 emissions from oil and gas production accounts for 63% of the total fossil fuel production (Saunois et al., 2020). Fugitive emissions can occur in all the chain of oil and gas productions, from extraction, transport and stockage. Periodical campaigns of LDAR (Leak Detection And Repair) have the intention of detecting and quantifying the sources of leaks. Nevertheless, they are not sufficient to detect all the fugitive emissions due to the episodic and spatially variable nature of leaks. Portability of instruments and accessibility of sites are some of the main problems that occurs in this kind of campaigns.

As mentioned on chapter 1, the deployment of a dense network of measurement sensors can increase the coverage of an emitting area, and thus allowing to better detect and quantify leaks. Current instruments used in surveys to detect leaks have high precision, but are too expensive to deploy in a network of measurement nodes. Low cost sensors provide an alternative to deploy such networks, but they have known issues, like cross-sensitivity to other VOC and environmental parameters.

In this chapter we performed a second experiment that assessed the performance of two types of MOS sensors from Figaro® manufacturer to reconstruct CH_4 enhancements (spikes) over background conditions. The sensors are the TGS 2611-C00 and the TGS 2611-E00, which were previously studied in chapter II for the reconstruction of room air CH_4 variations. In this study we addressed four questions: i) the performance of parametric and non-parametric calibration models to reconstruct CH_4 from TGS sensors and other variables, ii) the possibility of reducing the training length of the models, iii) the reduction in performance linked to aging in the sensors and iv) the possibility of use one calibration model across several sensors of the same type.

3.1.2 Design of the experiment

The experiment consisted of exposing four chambers, each containing three types of TGS to high concentration CH_4 enhancements generated over background conditions. This study was limited to reconstruct CH_4 enhancements from two out of the three TGS types. Indeed, TGS 2600 showed small signal enhancements to the spike CH_4 signal respect to the baseline, and thus it was excluded from this study. Each chamber included one version of TGS 2600, 2611-C00 and a 2611-E00. In addition, two sensors to monitor environmental variables (DHT22, SHT75 or SHT85 for relative humidity and temperature and BMP280 or BMP180 for temperature and pressure) were also included. Sensors were placed inside an acrylic/glass chamber of 100 ml on a circuit board that minimize heating influence from TGS on temperature sensors. The logging system was controlled by a Raspberry PI 3b+ and the sampling frequency was of 0.5 Hz (2s). A high precision instrument, Picarro G2401 CRDS, was used as a reference for the measurements.

3.1.2.1 Artificial spike experiment

CH_4 enhancements (spikes) were generated during 130 days, from 28 October 2019 to 5 March 2020. The automatic system in charge of the artificial spikes generation consisted in adding periodically small amounts of CH_4, from a cylinder containing 5% CH_4 in argon, on the ambient air flushed from the roof of the building. Spikes were produced automatically at least 3 times a day. The magnitudes and duration of these CH_4 spikes were predefined and controlled with two mass flow controllers to be ranging from 3 to 24 ppm, with durations varying between 1 to 7 minutes. Two types of spikes were generated during the experiment: the first type, from 28 October 2019 to 9 December 2019, has amplitudes comprised between 20 to 24 ppm; and the second type, from 9 December 2019 to 5 March 2020, has amplitudes comprised between 5 to 10 ppm and a higher number of peaks during a given period. A buffer effect was observed on the measurements from the chamber despite setting a high flow rate through the chambers (2.5 L/min). In consequence, an exponential weighted moving average (EWMA) has been applied on the Picarro G2401 CRDS data set in order to simulate the buffering effect from the chamber with

a time constant of 10s.

3.1.2.2 Separating CH_4 spikes from background variations in ambient air

Being able to separate spike conditions from background variations is a pre-requisite for CH_4 detection and reconstruction from TGS. Here we summarize the three-step approach we followed to achieve this goal. The first step consisted in removing the contribution of H_2O variations to the sensor voltage signals. We determined a linear relationship between H_2O variation and TGS voltage from a 32 day period with no spikes in the dataset to derive H_2O sensitivities of the voltage of each TGS in mV/ppm H_2O. Then, the linear model was applied to TGS voltage time series during spike periods. The second step consisted in separating background and spike conditions from voltage variations. We tested two methods. The first one consisted in applying a peak detection algorithm that iteratively compared neighboring observations to assign voltage measurements to spikes. The background signal is determined by a linear interpolation between non-spikes values at the start and at the end of each detected spike. The second method consisted in applying the Robust Extraction of Baseline Signal (REBS) algorithm, in which the spike conditions are defined as observations that are far away from a modelled background based on a threshold. This background is determined by computing local regressions over small moving time windows. The last step was to remove background observations and only keep data from spikes.

In this study we addressed four questions related to the reconstruction of CH_4 concentrations: TGS measurements, training of the models, aging effects and transfer of models between chambers.

In first place, we assessed five models, which can be parametric and non-parametric (linear, polynomial, random forest, hybrid random forest and multi-layer perceptron (MLP)), to reconstruct the CH_4 spikes. They were trained in two configurations of input variables (only TGS and TGS + environmental variables) and for two training lengths (70% and 50% of the total observations).

Secondly, we wanted to reduce the length of the training set while keeping good performances in the reconstruction of the spikes. This is done by a two-step approach: first we stratify the

data in different types of spikes with an unsupervised hierarchical clustering algorithm, then we construct different training sets by randomly selecting spikes from each cluster.

Thirdly, we analyzed aging effect of the sensors by applying the two best models on a second dataset generated six months after the first experiment under the same conditions and checking if these models showed degraded performances after six months.

Finally, we looked at the possibility of applying a trained model for one TGS type from one chamber to reconstruct CH_4 for the same type of sensor in another chamber.

To assess the goodness of reconstruction we consider that an error under 10% of the amplitude of the spikes we are interested in reconstructing as an acceptable target for TGS sensors. This corresponds to a RMSE $\leq$ 2 ppm for spikes with amplitudes of 24 ppm.

3.1.3 Main results

3.1.3.1 Reconstruction of CH_4 spikes

We compared the performances of five models using the mean square deviation (MSD) error decomposition. The MSD is decomposed in three main components: errors linked to incorrect reconstruction of magnitudes (SDSD), wrong reconstruction of phase or shape (LCS) and the bias (SB). Our results showed that the selection of the sensors' type was more critical than the selection of the model, and that the use of the TGS 2611-C00 produced the best reconstruction of the spikes. Regardless the training set size, the main source of errors corresponds to incorrect phase reconstruction (LCS) of the spikes, which increased when using TGS 2611-E00 sensors. Inclusion of environmental parameters produced little effects in the performance of the models. The magnitude reconstruction, that is, the SDSD component of the MSD, was affected when reducing the size of the training set, increasing the bias (SB) also for the non-parametric models, and it was partially mitigated by the inclusion of the environmental variables. Regarding the reconstruction models, better performances were observed when using the polynomial and MLP models (RMSE = 0.88 ppm and 0.85 ppm respectively).

3.1.3.2 Parsimonious training

We obtained nine clusters after applying a hierarchical clustering based on similarity between the spikes. The results of the training strategy using different sets of the data, on which we compared the effect of favorizing one set of clusters with the others, provided interesting insights in the performance of the training set. In first place, we observed that the smallest error obtained from models did not correspond to the training configuration with the largest training set configuration (accounting for 70% of the observations in the dataset). In addition, the most parsimonious training set configuration (with only 10% of the observations in the dataset) did not produce the best performances neither, although, it provided better performances than most of the cases. This points to a redundancy of information introduced by some of the clusters. Training set configurations composed with spikes characterized by having complex shapes produced the best performances allowing us to reduce the length of the training set size to 25% of the total observations with an error of RMSE = 0.89 ppm.

3.1.3.3 Effect of ageing on the reconstruction of CH_4 spikes

When reconstructing spikes on a second dataset generated six months after the first experiment, and based on the models established in the first experiment, we observed an increase in the reconstruction error from 0.57 to 0.85 ppm. This reflects that a possible source of error in the model after a long-time deployment of the sensors is caused by the drift on the sensors. An interesting result was observed on TGS 2611-E00, on which its degradation over time is not as large as on the TGS 2611-C00 type. In all configurations, the increase of the RMSE error was less than 1 ppm after six months showing that even with aging, the proposed models meet our requirement target.

3.1.3.4 Generalized models

We assessed the performance of two models, polynomial and MLP, for each type of TGS across different chambers. For each chamber, the reconstruction error on the test data set from the same chamber meet our requirement target. When reconstructing spikes from other chambers

we found that chambers that shared the same distribution of clusters, obtained good performances meeting the requirement target error as well. However, the models trained on data from one chamber that had less spike clusters than in the other chambers (due to a reduced number of observations) did not perform well to reconstruct CH_4 in other chambers. Regarding the influence of the type of sensor in the transferability of the models, TGS 2611-C00 produced better performances when reconstructing spikes from other chambers.

3.2 Publication in the journal of Atmospheric Measurement Techniques (AMT)

3.2.1 Introduction

Methane (CH_4) is a greenhouse gas 28 times more potent than carbon dioxide considering its warming potential over 100 years (Travis et al., 2020). Anthropogenic CH_4 emissions account for 60% of global emissions (Saunois et al., 2016). Fugitive leaks of natural gas at industrial facilities also present a safety hazard. Emissions from such facilities need to be continuously monitored, due to the episodic and spatially variable nature of leaks (Coburn et al., 2018). Leaks can be detected and quantified by LDAR surveys (Leak Detection And Repair) to detect high concentrations caused by a leak. Those surveys are periodical and have limitations related to the portability of instruments or accessibility of sites.

A possible solution to overcome these limitations is to deploy a network of sensors that continuously measure methane concentrations around an emitting area (Kumar et al., 2015). Deploying such a network with highly precise instruments, using techniques such as cavity ring down spectrometry (CRDS) is, however, cost prohibitive. Low cost sensors such as low power metal oxide semiconductor (MOS) sensors for methane are an alternative. Recent studies (Riddick et al., 2020; Casey et al., 2019; Collier-Oxandale et al., 2018; Jørgensen et al., 2020; Rivera Martinez et al., 2021; Eugster et al., 2020) tested the ability of MOS sensors to monitor methane concentrations in natural and controlled conditions and showed a fair agreement between the concentrations derived from the sensors and those from high precision reference instruments. MOS sensors are composed of a semiconducting metal oxide sensing element heated at a temperature between 200° to 400 °C (Özgür Örnek and Karlik, 2012; Barsan et al., 2007). When the semiconducting material is in contact with an electron donor gas like CH_4, a change in the conductivity occurs, measured by an external electrical circuit (Özgür Örnek and Karlik, 2012). MOS sensors are known to be less precise than CRDS to CH_4 variations, although they can detect small variations in concentrations. Most MOS sensors have cross-sensitivities to other electron donors and to environmental vari-

ables such as absolute humidity, pressure and temperature (Popoola et al., 2018) with non-linear interactions (Rivera Martinez et al., 2021).

Biases affect CH_4 measurements derived from low-cost sensors because of cross sensitivities to other gases, dependence to environmental factors and internal drifts e.g. due to aging. Figaro Taguchi Gas Sensors (TGS) are a particular series of MOS capable of measuring CH_4. In order to limit biases of these sensors, several studies proposed a calibration model against a high precision reference instrument. Casey et al. (2019) compared different calibration approaches with inverse and direct linear models and artificial neural networks to quantify O_3 from an SGX Corporation MiCS-2611 sensor, CO from a Baseline Mocon PID sensor, CO_2 from an ELT S-100 non-dispersive infrared (NDIR) sensor and CH_4 from observations of a Figaro TGS 2600 sensors. Collier-Oxandale et al. (2018, 2019) applied multilinear models, including interactions from environmental variables, to predict CH_4 concentrations and to detect and quantify VOCs from Figaro TGS 2600 and TGS 2602 MOS sensors at two sites with active oil and gas operations. Eugster et al. (2020) used empirical functions and Artificial Neural Networks (ANN) to derive CH_4 concentrations from 6 years of data collected with Figaro TGS 2600 sensors at a field site in the Arctic. Riddick et al. (2020) derived nonlinear empirical relationships for Figaro TGS 2600 sensors from three experiments with durations varying from one day to one month. Rivera Martinez et al. (2021) reconstructed CH_4 concentrations variations in room air from Figaro TGS 2611-C00 sensors using ANN models and co-variations of temperature, water mole fraction and pressure. Nevertheless, those comparisons were limited by the choice of a specific reconstruction model and restricted to only one type of sensor.

There is a need for a more thorough comparison of different calibration approaches for Figaro MOS sensors applied to measure CH_4. In addition, there is a need to assess the performances of MOS sensors to detect and quantify CH_4 spikes typical of industrial emission. This study aims to compare several parametric (linear and polynomial) and non-parametric models (random forest, hybrid random forest and ANN) applied to different combinations of Figaro TGS sensors to reconstruct the CH_4 signals of repeated atmospheric spikes, based on the observed voltage of each sensor and other variables. The CH_4 signal we aim to reconstruct is representative of variations

observed in the atmosphere from leaks that occur within or close to an emitting industrial facility, i.e. short duration CH_4 enhancements (spikes) lasting between 1 to 7 minutes and ranging between 1 to several tenth of ppm above an atmospheric background concentration of around 2 ppm (Kumar et al., 2021). In this study, we performed a laboratory experiment where a CRDS instrument and many TGS sensors of different types were exposed to a controlled air flow with artificially created CH_4 concentration spikes (section 2). The spikes were composed of pure CH_4 and did not contain any VOCs, although those species could be present in natural gas leaks from oil and gas facilities. The experiment lasted four months and provided 838 spikes, which give us a dense and complex dataset to train and test different models for reconstructing CH_4 variations.

For low-cost sensors, a collocation is often required with a highly precise reference instrument to train an empirical calibration model. This training phase should be as effective (parsimonious) as possible. The strategy is to reduce the time and maintenance costs of having a reference instrument on site if the purpose is to bring it in the field for future studies where low-cost sensors would have to be calibrated. We investigate the problem of 'parsimonious training' by testing different configurations (model and inputs) to establish the minimum amount of reference data needed to obtain good performances with the low-cost sensors (section 3.2.3.2 and 3.2.3.3). Secondly, since the performance of low-cost sensors may change with time, it is important to understand if their measurements could be affected by a drift of their sensitivity over time. We address this problem of 'non-stationary training' by comparing different calibration models for a second spike experiment conducted six months after the first one (section 3.2.3.4). Thirdly, sensitivities may vary from one sensor to another and may require a sensor-specific calibration model, which becomes a problem when a large number of sensors are deployed. Finding a robust calibration model that could be trained using data from one or several sensors and applied to others remains an open question. We bring some insight to this problem of 'generalized calibration' by training models to reconstruct the CH_4 signal from a group of sensors located in the same chamber and applying them to other groups of sensors in a different chamber (section 3.2.3.5). To assess the performance of the calibration models and particularly their capability to reconstruct spikes of several ppm occurring upon a background CH_4 level, we define here as an acceptable

performance to be an error less than the 10% of the maximum amplitude of the peaks we aim to reconstruct. In our case, this requirement is an RMSE of 2 ppm between the reconstructed CH_4 data from low cost sensors and the true data from a reference instrument at a time resolution of 5 seconds.

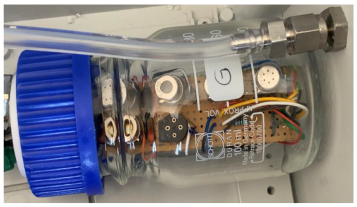
3.2.2 Methods

3.2.2.1 Experimental set-up

Low-cost CH_4 sensors

For the experiment, four independent sampling chambers were assembled. Each chamber contained a Figaro TGS 2600, TGS 2611-C00 and TGS 2611-E00 sensor, alongside a relative humidity and temperature sensor (DHT22 or Sensirion SHT75), and a temperature and pressure sensor (Bosch BMP280, see Table 3.1 for details). Issues with the logger system produced gaps in environmental variables data, thus observations information from an external chamber (E, see figure 3.1b and Table 3.1 for details) was used in the correction of the sensitivity across all chambers. The sensors were disposed on a circuit board to minimise the direct heating influence of the TGS sensors on temperature measurements. The sampling chamber was made of acrylic/glass with a gas inlet and outlet and a port for the electrical cables (Figure 3.1a). Each sensor was connected in series with a high-precision load resistor which controlled sensitivity (Figaro®, 2005, 2013). The voltage across each load resistor was recorded by an AB Electronics PiPlus ADC board, mounted on a Raspberry Pi 3b+ logging computer, sampling at a frequency of 0.5 Hz (2s). This voltage measurement was used in our characterisation algorithms, referred to hereafter as the sensor voltage. We focus on the reconstruction of CH_4 using only the TGS 2611-C00 and TGS 2611-E00 data.

(a)



(b)

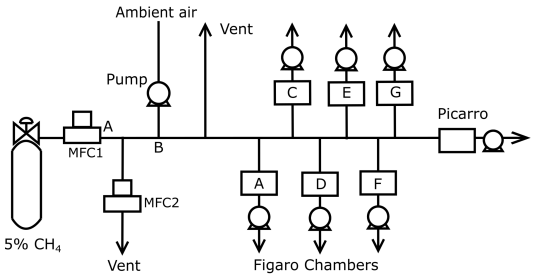


Figure 3.1: (a) Example of a chamber with three sensors inside, (b) Scheme of the spike creation experiment.

Table 3.1: Summary of the sensors included on each logger box.

Chamber	Figaro TGS	Temperature & Relative Humidity sensor	Temperature & Pressure sensor	Load Resistor
A	TGS 2600 TGS 2611-C00 TGS 2611-E00	DHT22	BMP280	50 KΩ
C	TGS 2600 TGS 2611-C00 TGS 2611-E00	SHT75	BMP280	50 KΩ
F	TGS 2600 TGS 2611-C00 TGS 2611-E00	SHT75	BMP280	50 KΩ
G	TGS 2600 TGS 2611-C00 TGS 2611-E00	SHT85	BMP280	50 KΩ
E	TGS 2600* TGS 2611-C00* TGS 2611-E00*	SHT75 & DHT22	BMP180	5 KΩ

* Two sensors of this type

Generation of methane spikes on top of ambient air

The experiment lasted 130 days from 28 October 2019 to 5 March 2020. During this period, the four chambers containing TGS sensors sampled ambient air pumped from the roof of the laboratory. Relative humidity, air pressure and temperature were measured in the ambient air flux, as well as CH_4 , using a Picarro CRDS G2401 reference instrument. No calibration was considered on the CRDS instrument during the experimental period due to its high precision and low drift over time (less than 1 ppb per month; (Yver-Kwok et al., 2015)).

To expose the TGS sensor chambers to CH_4 enhancements (spikes) of different durations and amplitude comparable with typical enhancements observed around industrial sites (Kumar et al., 2021), we designed an automatic system to add small amounts of CH_4 on top of the ambient air acquired from our roof. The system presented on figure 3.1b consists of an ambient air flow to which was periodically added a small amount of a gas from a cylinder containing 5% of CH_4 (in argon), controlled by two mass flow controllers denoted MFC1 and MFC2 in Figure 3.1b.

The occurrences of the spikes were programmed to be automatically generated, with at least three spikes each day. The duration and magnitude of the spikes were predefined and controlled by varying the flows of MFC1 and MFC2, the two mass flow controllers being programmed to add an amount of CH_4 to produce spikes of an expected amplitude ranging between 3 and 24 ppm. Two different types of spikes were generated. The first type, with large amplitudes between 20 and 24 ppm, was generated from 28 October 2019 to 9 December 2019. The second type, with smaller amplitudes ranging between 5 ppm and 10 ppm but with a higher number of spikes during a given period of time, was generated from 9 December 2019 to 5 March 2020. The typical duration of the spikes of both types varied between 1 and 7 minutes, which is longer than the known response time of the TGS sensors. Gas from the 5% CH_4 cylinder that persisted on the air flow after a spike in segment A-B (Fig. 1b) was expulsed though MFC2, preventing very high CH_4 concentrations to remain in the air flow following a spike. We verified that the amount of gas with 5% of CH_4 added to the air flow measured by the TGS sensors did not affect the air pressure, temperature and relative humidity in the chambers.

The volume of each chamber is 100 ml and the flow rate through the chambers was fixed to 2.5 L per minute. We did not test the effect of increasing the flow rate on the TGS measurements. Instead, we decided to choose a high enough flow rate to reduce the buffering effect of the chamber volume that would systematically smooth the CH_4 spikes. Despite this set-up, a buffering effect was still present in the chamber, evidenced by the fact that after stopping the injection of air with 5% CH_4 , the CH_4 draw down in the chamber was observed to be smooth and lagged the drawdown of the CRDS instrument by a time constant of 10s, consistent with previous measurements on buffer volumes acting as a low-pass filter (Cescatti et al., 2016).

To determine the time constant (τ) of the buffer effect of the chambers, we applied an exponential weighted moving average to the CRDS data with different values of τ and compared them with the shape of the response of the TGS sensor (see Fig. B.1). A similar approach was employed by (Jørgensen et al., 2020) to compensate for effects of micro turbulent mixing of subglacial air with atmospheric observations. Before applying this temporal smoothing on the CRDS data, we resampled both signals, the reference CRDS and the TGS, from their original time resolutions (1s and 2s, respectively) to a common time resolution of 5s.

3.2.2.2 Separating CH_4 spikes from background variations in ambient air

Different algorithms have been proposed to identify short term variations of atmospheric signals (Ruckstuhl et al., 2012) from slower variations of background variations in atmospheric composition. These approaches were applied to low-cost sensors for the detection of local events (Heimann et al., 2015), and for the removal of diurnal periodical signals to identify peaks of air pollutants (Collier-Oxandale et al., 2020). In this study, we want to separate the background of slowly varying CH_4 in the outside air pumped from the laboratory roof from the signal of the CH_4 spikes by using an algorithm.

We followed a three-step approach. The first step was to remove the impact of H_2O variations on the sensor voltage signals, given that H_2O changes in the background air. Previous studies (Eugster and Kling, 2012; Rivera Martinez et al., 2021) demonstrated a direct dependence between the voltage/resistance of metal oxide sensors and H_2O concentration. In order to determine this relationship, we used the background H_2O mole fraction and TGS voltage measurements in ambient air during a period of 32 days with no CH_4 spikes, and regressed both variables to derive the H_2O-sensitivity of the voltage of each TGS sensor in mV per ppm H_2O. This linear model of the Voltage - H_2O sensitivity was applied to voltage time series of the TGS sensors during the spike measurement period.

The second step was to separate background and spike conditions from voltage variations in the time series. We tested two approaches. The first approach applied the peak detection algorithm of (Coombes et al., 2003) to detect the voltage associated to spikes and separate the

background signal by a linear interpolation between non-spikes values at the start and the end of each spike. The second approach applied the Robust Extraction of Baseline Signal (REBS) algorithm from Ruckstuhl et al. (2012) to separate voltage observations associated to background from those during the spikes. The principle of REBS is to compute local regressions over the time series on small moving time windows (60 minutes) and to iteratively identify outliers that are far from the modelled background, based on a threshold. Here, the detected outliers are considered to belong to a spike. The threshold or scale parameter, β, defines a range in number of standard deviations around the modelled baseline. A value of β = 3.5 ppm was used. The third step was to remove observations corresponding to baseline and keep only the data classified as spikes, which form the signal of interest in this study.

3.2.2.3 Modelling CH_4 spikes from TGS sensor voltages and environmental variables

The impact of different magnitudes of the variables used as predctors are prone to affect the parameters of the models in the training stage. Thus, to reduce this impact we standardize the inputs before training the models. We chose a robust scaler unaffected by outliers by removing the median and scaling the data to a quantile range (Demuth et al., 2014). To reconstruct CH_4 spikes from TGS sensor voltages, we applied linear and polynomial regressions, ANN and Random Forest models, all trained using the CRDS measurements. We assessed the performance of the different models using a k-fold cross validation, here with k = 20. A fraction of the data was used for the training of each model and the rest for evaluation. We repeated this training and evaluation process with a moving window to make a robust assessment of each model performance considering all data available. We specified two cases for the relative sizes of the training and evaluation (test) sets. The first case used training and test set fractions of 70% and 30% of the observations, respectively, and the second one used 50% and 50%. We focus in the example below on the spike data from one chamber (chamber A) using different models as test inputs: 1) voltages of TGS C or E sensors separately, 2) voltages of a single sensor type and measurements of H_2O, temperature and pressure, 3) combined TGS C and E voltages, and 4) combined TGS C and E voltages, plus H_2O, temperature and pressure.

Linear and multilinear regression models

Linear regressions between dry air CH_4 concentrations from the CRDS and TGS sensor voltages are the simplest models, used in studies with similar low-cost sensors by others (Collier-Oxandale et al., 2018; Casey et al., 2019; Cordero et al., 2018; Spinelle et al., 2015, 2017; Malings et al., 2019). We derived linear regressions between the reference CH_4 from the CRDS instrument and the sensor voltage, as well as a multi-linear regression including voltage, H_2O, air pressure and temperature, as given by:

$$\hat{y}_{CH_4}(x_1 = V_{TGS}) = \beta_0 x_1 + \beta_1 \tag{3.1}$$

and

$$\hat{y}_{CH_4}(x_1 = V_{TGS}, x_2 = H_2O, x_3 = P_{Air}, x_4 = T_{Air}) = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4 + \alpha_5 \tag{3.2}$$

Where $\hat{y}_{CH_4}$ is the predicted methane concentration in ppm, V_{TGS} the observed sensor voltage in V, H_2O is the water vapor mole fraction in %, P_{Air} the air pressure in kPa and T_{Air} is the air temperature in °C.

Polynomial regression models

The second type of models are second degree polynomials, for which we considered as predictors either TGS sensor voltage alone, or TGS voltage plus environmental variables, as given by:

$$\hat{y}_{CH_4}(x_1 = V_{TGS}) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \tag{3.3}$$

and

$$\hat{y}_{CH_4}(x_1 = V_{TGS}, x_2 = H_2O, x_3 = P_{Air}, x_4 = T_{Air}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1^2 + \beta_6 x_1 x_2 + \beta_7 x_1 x$$
$$\beta_8 x_1 x_4 + \beta_9 x_2^2 + \beta_{10} x_2 x_3 + \beta_{11} x_2 x_4 + \beta_1 2 x_3^2 + \beta_{13} x_3 x_4 + \beta_1$$
(3.4)

Random forest and hybrid random forest models

Random forest regressors (Breiman, 2001) are an ensemble learning method consisting of creating several decision trees to fit complex data. Each tree is composed of leaves defined hierarchically based on thresholds that group values of input variables, constructed from a subset of predictors randomly chosen, a process known as 'feature bagging'. The prediction is made by averaging the outputs of all the trees. As a non-parametric method, the generalization of Random Forests is limited by the range of values present in the training set. A methodology proposed by Malings et al. (2019) to boost the generalization of random forest models is to 'hybridize' them with a parametric model to be able to predict values that are out of those present in the training set. The principle of hybridization consists in training a random forest model with about 80 to 90% of the observations, and reserving 20 to 10% of the higher observations to train a linear or polynomial model. This approach allows us to benefit, on one hand, from the capability to derive nonlinear relationships from the inputs, while on the other hand boosting the prediction outside of the range present in the training set, here with a linear or polynomial model. Here we used both traditional and hybrid random forest models. For the hybrid models, we reserved, for each cross-validation fold, the higher 10% concentrations to train a polynomial fit, the remaining observations being fitted by the random forest. The same four cases of input combinations explained in section 3.2.2.3 were used for the training of traditional and hybrid random forest models.

Artificial neural networks (ANN)

In recent studies with low-cost sensors (Rivera Martinez et al., 2021; Casey et al., 2019) ANN models have proven to be powerful models to derive CH_4 concentrations from sensor signals. We chose here a Multilayer Perceptron (MLP) model due to its ability to provide a universal approximator (Hornik et al., 1989) and generalization capabilities (Haykin, 1998). No prior knowledge of relationships between variables is required to produce model outputs. Our MLP is composed of a series of units (neurons) arranged in fully connected layers, each unit being a weighted sum of its inputs to which an activation function (tanh, ReLU) is applied. The last layer of the network when used as a regressor usually has one unit and a linear activation function. As a supervised learning algorithm, MLP requires examples (the training set) and an iterative learning algorithm to adjust the weights of its connections. The main challenges for training an MLP are: 1) underfitting, when the model is not able to fit the training set and 2) overfitting, when the model is not capable of generalizing new examples. Underfitting can be mitigated by increasing the complexity of the MLP, and overfitting can be partly mitigated by weight decay regularization or early stopping (Bishop, 1995a; Goodfellow et al., 2016).

We built different MLP models using the BFGS algorithm (Bishop, 1995a). The optimal number of layers and units was determined using a grid search technique (Géron, 2019), resulting in 1) 4 hidden layers when using the voltage from a single TGS sensor as input, with 2, 3, 5 and 2 units per layer, 2) when using TGS sensor voltages and other variables, 4, 2, 5 and 5 units per layer, 3) 5 layers when combining both TGS sensor voltages together (5, 3, 5, 5 and 4 units), and 4) 4 layers when using both TGS sensor voltage types and other variables (5, 3, 5 and 5 units). The ReLU activation function was used on units of the hidden layer and early stopping was used to prevent overfitting.

3.2.2.4 Finding a parsimonious model training strategy

To determine the minimum number of training observations to obtain a model with satisfactory performances, given our 2 ppm RMSE requirement posed in the introduction, we followed a two-step approach. First, we stratified the data into different types of spikes using an unsuper-

vised hierarchical clustering algorithm (Johnson, 1967). Secondly, we constructed training sets by randomly selecting spikes inside each cluster, in different proportions given in Table 3.2, and evaluated our models against the remaining spikes used as a test set. This evaluation strategy helped us to understand the clusters that have the most influential impact on to increase the models performance. This allowed us to reduce the length of the training set by sampling the training data preferentially in the most influential clusters.

The clusters of spikes were defined using the ward distance to determine a matrix measuring the degree of similarity between spikes using Dynamic Time Warping (DTW) (Sakoe and Chiba, 1978), and to construct a dendrogram. A threshold on the dendrogram allowed us to determine 9 different clusters from our dataset. For the second step, we defined 11 cases to construct training sets. Cases 1 and 11 correspond to sampling 70% and 10% of the data for training, respectively, equally distributed across the clusters. Cases 2 to 10 correspond to preferential sampling one cluster over the others for training, by selecting 70% of the spikes in this cluster and 10% in all others. The purpose of this stratified data selection is to determine the type of spikes that best allows for reconstruction of the variations of CH_4 when training a model. At this stage we are not interested in the temporal dependency between observations since we train models with instant values. On a practical application side, a parsimonious model training strategy will require users to expose their sensors to specific type of 'highly influential' spikes on a shorter period from, e.g., a laboratory experiment like the one described above, then train the models upon those spikes and apply them to data collected in the field.

3.2.2.5 Assessing ageing effects of the sensors

To assess the effect of ageing sensors on the reconstruction of CH_4, we conducted a 33-day experiment from 11 August to 12 September 2020, six months after the first experiment described in section 3.2.2.1. The spike generation system was the same. Between the two experiments, the chambers containing TGS sensors had been measuring ambient air pumped from our laboratory roof. To assess the ageing effect on the TGS sensors during the six-month interval, we selected the two models that gave the highest performance for the first experiment and applied them to

Table 3.2: Percentage of spikes in each cluster (C1 to C9) considered for training different models.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	# Spikes in total	% of data in the training
Case 1	70%	70%	70%	70%	70%	70%	70%	70%	70%	587	70.0%
Case 2	70%	10%	10%	10%	10%	10%	10%	10%	10%	122	12.5%
Case 3	10%	70%	10%	10%	10%	10%	10%	10%	10%	150	13.7%
Case 4	10%	10%	70%	10%	10%	10%	10%	10%	10%	147	14.5%
Case 5	10%	10%	10%	70%	10%	10%	10%	10%	10%	198	19.3%
Case 6	10%	10%	10%	10%	70%	10%	10%	10%	10%	105	17.8%
Case 7	10%	10%	10%	10%	10%	70%	10%	10%	10%	166	18.5%
Case 8	10%	10%	10%	10%	10%	10%	70%	10%	10%	150	16.7%
Case 9	10%	10%	10%	10%	10%	10%	10%	70%	10%	94	11.9%
Case 10	10%	10%	10%	10%	10%	10%	10%	10%	70%	124	24.5%
Case 11	10%	10%	10%	10%	10%	10%	10%	10%	10%	84	10.0%

simulate the spikes generated during the second experiment.

3.2.2.6 Finding generalized models that can be used for other sensors of the same type

We were interested in understanding to what extent a model trained with the outputs of a given TGS sensor type in a given chamber could be applied to other sensors of the same type in other chambers. The experiment consisted of training a model per sensor and chamber with the best configuration subset based on the cluster classification outlined in section 3.2.2.4. The trained model is then used to reconstruct the CH_4 spikes using data from the TGS in other chambers and compare their performances. For this, we used data from chambers A, C, F and G to train chamber-specific models, and used each chamber-specific model to reconstruct CH_4 spikes using data from other chambers, as shown in Table 1. The four chambers have a load resistor of 50 kΩ and contain three TGS sensors each. We did not use data from chamber D and E because it has a load resistor of 5 kΩ and on the chamber E contains two of each TGS sensor.

3.2.2.7 Metrics for performance evaluation

The performance of the models to reconstruct the dry CH_4 concentrations observed by the CRDS instrument using TGS sensors was assessed using a decomposition of the mean squared deviation (MSD) of the misfits between reconstructed and true CH_4 (Kobayashi and Salam, 2000), to

separate the main source of errors when comparing different models. MSD was decomposed into the sum of the Square Bias (SB), the difference in the magnitude fluctuation (SDSD) and the lack of positive correlation weighted by the standard deviation (LCS). A large SDSD indicates an incorrect reconstruction of CH_4 spike magnitudes. A large LCS indicates an incorrect reconstruction of spike phase or shape. The equations for each error term according to Kobayashi and Salam (2000) are given by:

$$\mathrm{SB} = (\overline{\hat{y}_{\mathrm{CH_4}}} - \overline{y_{\mathrm{CH_4}}})^2 \tag{3.5}$$

$$\mathrm{SDSD} = (\sigma_{\mathrm{Model}} - \sigma_{\mathrm{Ref}})^2 \tag{3.6}$$

$$\mathrm{LCS} = 2\sigma_{\mathrm{Model}}\sigma_{\mathrm{Ref}}(1 - \rho) \tag{3.7}$$

$$\mathrm{MSD} = \mathrm{SB} + \mathrm{SDSD} + \mathrm{LCS} \tag{3.8}$$

With $\overline{\hat{y}_{\mathrm{CH_4}}}$ the mean of the prediction, $\overline{y_{\mathrm{CH_4}}}$ the mean of the reference observations, σ_{Model} the standard deviation of the modelled CH_4 time series, σ_{Ref} the standard deviation of the reference one and ρ their correlation coefficient. All results presented below are using metrics computed for the test set only.

3.2.3 Results

3.2.3.1 Data pre-processing and baseline correction

Figure 3.2 shows the pre-processing steps of the dataset, with the identification and removal of the background signal from the spikes in the time series. We removed outliers and the first 30 minutes of observations in case of a reboot of the data loggers of each chamber, i.e. during stabilization of the sensors. The original observations on a time step of 2 s were resampled to means of 5 s. Time shift due to incorrect clock synchronization between the reference CRDS instrument and the loggers of the TGS sensors were partly corrected with a search of the maximum correlation on non-overlapping windows of 6 hours and a manual inspection of the agreement between

TGS voltages and CH_4 observations of the reference CRDS instrument.

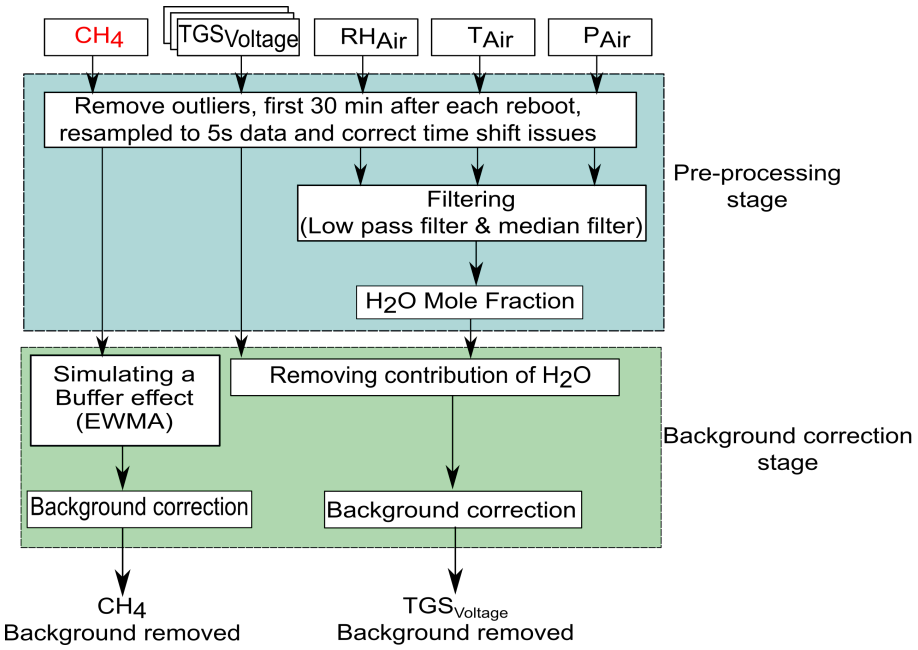


Figure 3.2: Data pre-processing diagram, correction of H_2O effects, and separation of the spikes from background data in the time series.

Environmental variables (H_2O, temperature, pressure) were filtered using a low pass filter (Press and Teukolsky, 1990) to remove high frequency noise from the sensors and circuit connections. The water vapour mole fraction was calculated with Rankine's formula (Eq. 9) from relative humidity (RH) in % and temperature (T) in °C from the DHT22 sensors and pressure (P) in Pa from the BMP180 sensors in each chamber, according to:

$$H_2O_{Mole\ Fraction} = 100 \times \left(\frac{\frac{RH}{100} \times e^{\frac{13.7-5120}{T+273.15}}}{\frac{P}{100000} - \frac{RH}{100} \times e^{\frac{13.7-5120}{T+273.15}}} \right) \tag{3.9}$$

An example of several spikes obtained after the pre-processing and background signal removal is shown in figure 3.3 (see also figure B.3). The entire spike dataset contains 838 spikes, representing 1.6% (35536 5s observations) of the full dataset.

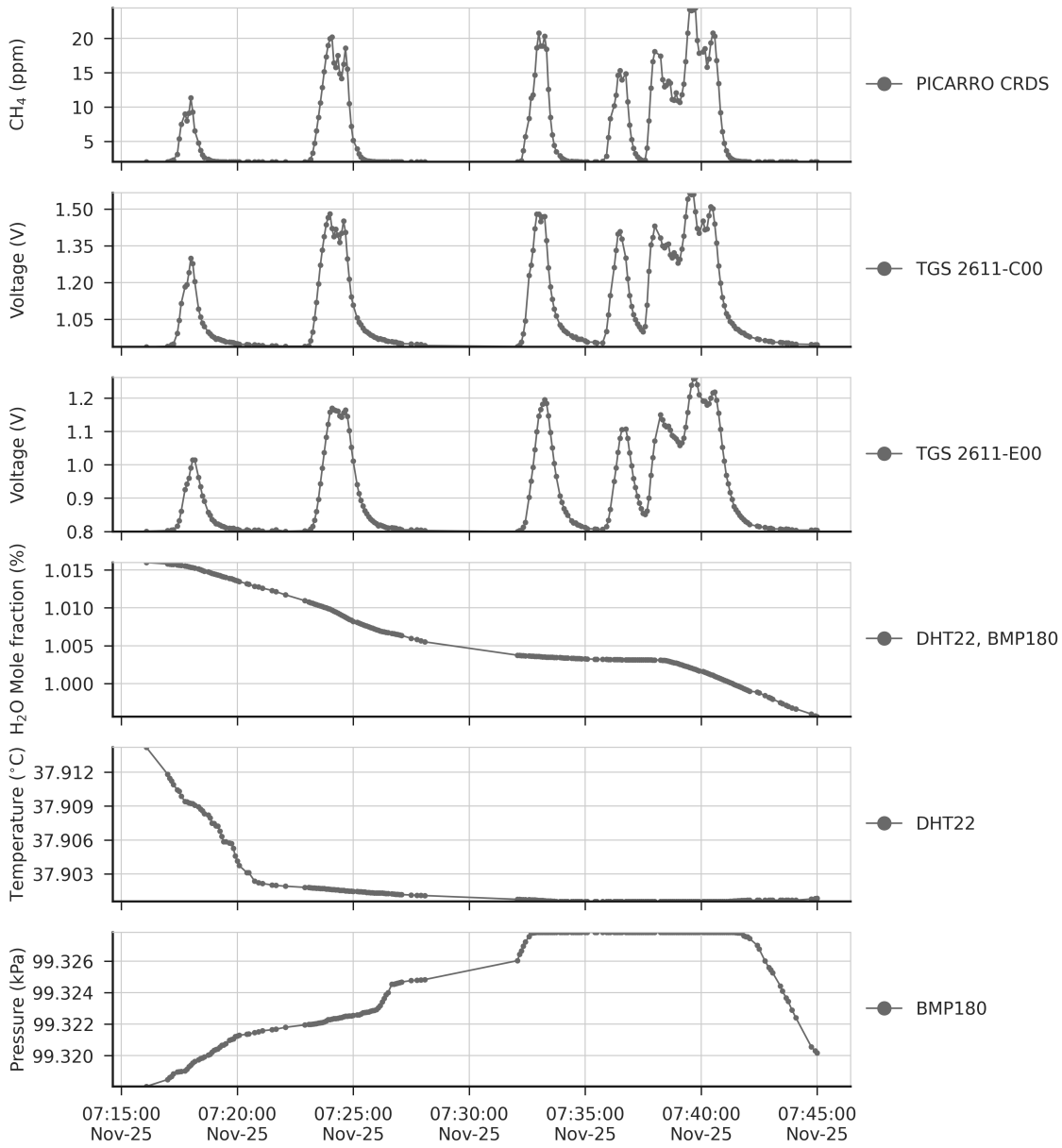


Figure 3.3: (top to bottom) Time series of reference CH_4 signal from CRDS, voltage from TGS sensors, H_2O, temperature and pressure during a period of 30 minutes, after removing from the time series the variations of background signals, and applying the H_2O correction to the voltage signal of TGS sensors. Dots on panels represent actual observations and lines between dots are drawn to show the shape of the signals.

3.2.3.2 Reconstruction of CH_4 spikes

Figure 3.4 shows the reconstruction of several spikes by the linear, polynomial, random forest (RF), random forest hybrid (RFH) and MLP models using data from the type C TGS sensor in chamber A. Figure 5 shows the reconstruction results using data from the type E TGS sensor in chamber

A. In both figures, the model training set contains 70% of the total observations available. The spikes reconstructed by the different models show good agreement with the reference CH_4 signal for the type C sensors, but not for type E ones which are associated with phase errors and greater noise in the reconstructed CH_4. The linear model, RF and the RFH models broadly capture the mean amplitude of spikes, but they are less capable of reconstructing small CH_4 variations on the top of the spikes. The RF and RFH models (the latter with a polynomial model) provided very similar outputs, with a small enhancement of the amplitude for RFH during some spikes and noise, especially with type E sensors (Fig. 3.5). The MLP model showed a constant underestimation of the spike magnitudes and produced smoother spike shapes, presenting a low pass filter behaviour. The polynomial fit models appeared to perform better. Despite the phase misfit of models with type E sensors, for all the models, both type C and type E meets our requirement target of an RMSE $\leq$ 2 ppm (MSD $\leq$ 4 ppm^2). With a stricter requirement of an error less than the 5% of the maximum amplitude of the peaks (RMSE $\leq$ 1 ppm) only Type C is adequate.

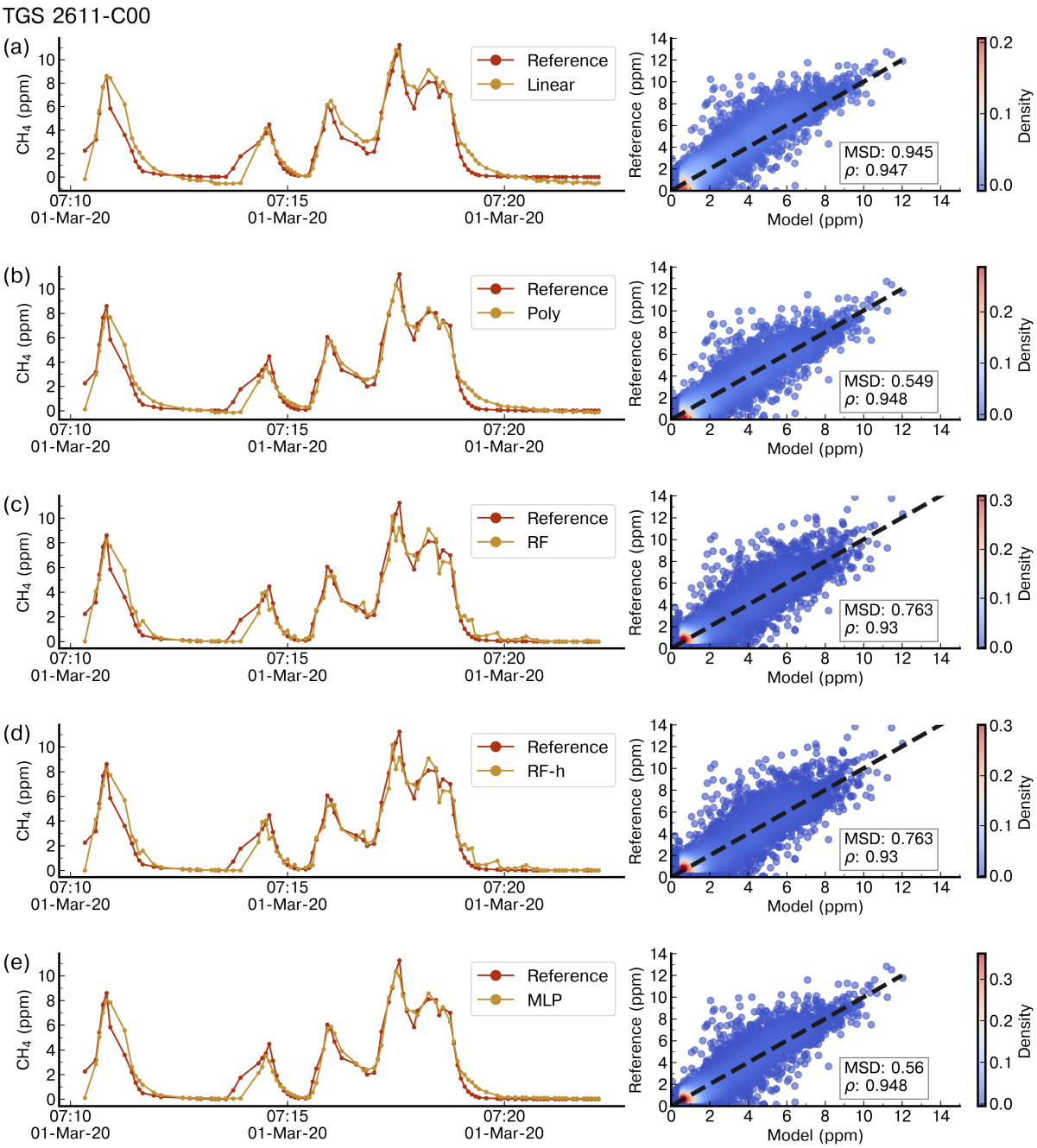


Figure 3.4: Example of reconstruction of the CRDS reference CH_4 signal on a time step of 5 s for a few spikes in the test set by (a) a Linear model, (b) a polynomial model, (c) a Random Forest model, (d) a Random Forest Hybrid model and (e) a Multi-Layer Perceptron model trained with 70% of data and using as input data from the TGS 2611-C00 sensors only. The right panels show scatter plots between the reference CH_4 signal and the modelled outputs. The colour code is the density of observations.

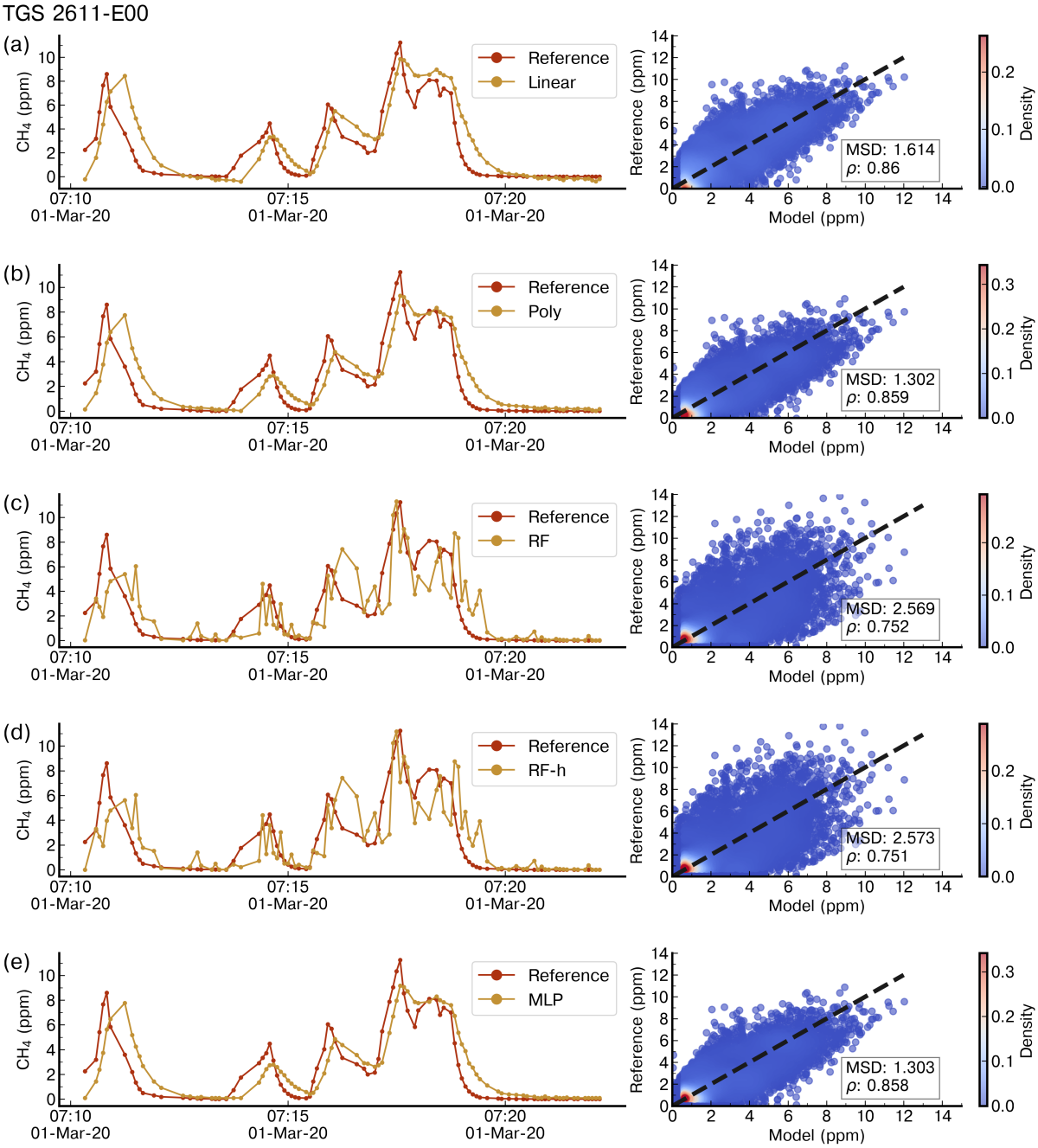


Figure 3.5: Same as Figure 3.4 but for data from the TGS 2611-E00 sensors only.

Figure 3.6 shows the distributions of the correlations (ρ) between modelled and observed CH_4 spikes for the 20-fold validations periods (test sets) for different models. We distinguished two groups of models, based on median values of ρ. The first group corresponds to models trained with type E sensors data only, characterized by $\rho_{\text{Median}} \leq 0.93$. The second group corresponds to models trained with type C sensor data only, or with data from both types of sensors, char-

acterized by a higher $\rho_{\mathrm{Median}} \geq 0.96$. Among the models in the first group, the Polynomial Model gave the largest correlations (ρ_{Median} = 0.92, interquartile range (IQ) = 0.001). Among the models in the second group, the Polynomial Model also showed the largest correlation, especially with both types of sensors, and a training set of 50% of the observations (ρ_{Median} = 0.98, IQ = 0.004), closely followed by the MLP model with the same inputs and the same training set size (ρ_{Median} = 0.98, IQ range=0.006). The Random Forest, Random Forest Hybrid and MLP models also showed high correlations when input data are from the type C sensors and the training uses 70% of the observations (RF ρ_{Median} = 0.982, RFH ρ_{Median} = 0.983 and MLP ρ_{Median} = 0.982). These three models however had lower correlations when input data are from type E sensors (RF ρ_{Median} = 0.893, RFH ρ_{Median} = 0.894 and MLP ρ_{Median} = 0.92). Phase errors were reduced when training the models with either type C sensor data or data from both sensors. The length of the training set had an important impact on the spread of the correlations across the 20-fold periods. With 70% of observations in the training set, the IQ of the correlations increased, whereas for a smaller training set, the IQ was smaller but the distribution of the correlations showed more outliers. The inclusion of environmental variables (Fig. 3.6b) as input to models, in addition to voltages from TGS sensors, reduced significantly the phase error in the Random Forest models but produced little improvements in the results from other models.

Figure 3.7 shows the MSD error decomposition for the different models and for the two training set sizes of 70% and 50%, respectively. We observed that the LCS component of the MSD (related to a phase misfit of the modelled series) is the principal source of error across the different models, regardless of the input used or the size of the training set, meaning that models have more difficulties to reproduce the phase of the spikes than their amplitude. A systematically higher LCS error was obtained when data from type E sensors are used as input, and there is also a larger SDSD error with this type of sensor. For example, the largest LCS error was found with a training set of 70% for the Random Forest models ($\mathrm{LCS_{RF}}$ = 4.67 ppm^2, $\mathrm{LCS_{RF}}$=1.24 ppm^2, $\mathrm{LCS_{RF}}$=0.79 ppm^2 with type E sensor data, type C sensor data and both types respectively) as well as for the RFH models, when compared with other models. Additionally, the inclusion of environmental variables had little effect on the model performance. This was clearly shown for the LCS

error of the polynomial model, for a training set of 70% of the data, which was identical with and without environmental variables as input ($\mathrm{LCS_{poly}}$=3.0 ppm^2 for the type E sensor, $\mathrm{LCS_{poly}}$=0.84 ppm^2 for the Type C sensor and $\mathrm{LCS_{poly}}$=0.7 ppm^2 for both types). Reducing the size of the training set affected mostly the SDSD component, by slightly lowering the capability of models to reconstruct the amplitude of the CH_4 spikes. For the non-parametric models, reducing the size of the training set also increased the bias error (SB), an effect that was partially mitigated with the inclusion of environmental variables. Amongst the non-parametric models, the MLP obtained similar performance than the parametric polynomial model ($\mathrm{MSD_{MLP}}$= 3.2 ppm^2, $\mathrm{MSD_{MLP}}$=0.85 ppm^2 and $\mathrm{MSD_{MLP}}$=0.7 ppm^2 for type E sensors, type C sensor and both types together, respectively). To summarize, the choice of the sensor type used to train the models affected more the reconstruction error than the selection of the model. The type C sensor data produced the lowest error compared to type E, irrespective of the model used. Overall, the polynomial model gave better performance than the non-parametric models. More detailed statistics are summarized in tables B.1 and B.2.

3.2.3.3 Results of parsimonious training tests

Figure 3.8 shows the result of the spike clustering. Based on spike similarity, we found 9 clusters. The peaks with short durations (under 50s) and containing only one spike were grouped into cluster C1 (signal amplitude (sa) $\leq$ 6 ppm) and cluster C3 (6 ppm $\leq$ sa $\leq$ 12 ppm). Peaks with longer duration (over 50 s) were grouped in clusters C2 (sa $\leq$ 4 ppm) and C4 (4 ppm $\leq$ sa $\leq$ 8ppm). Peaks with very long duration (between 50s to 1.5 min) were grouped in cluster C5. Peaks with a small concentration at the beginning (around 6 ppm) followed by a larger peak (up to 12 ppm) were grouped in cluster C6. Peaks with larger concentrations ($\geq$ 12 ppm) and complex in shape were grouped in clusters C7, C8 and C9, respectively. The cluster regrouping the largest number of spikes (191) was C4, and the one with the smallest number of spikes (17) is C8.

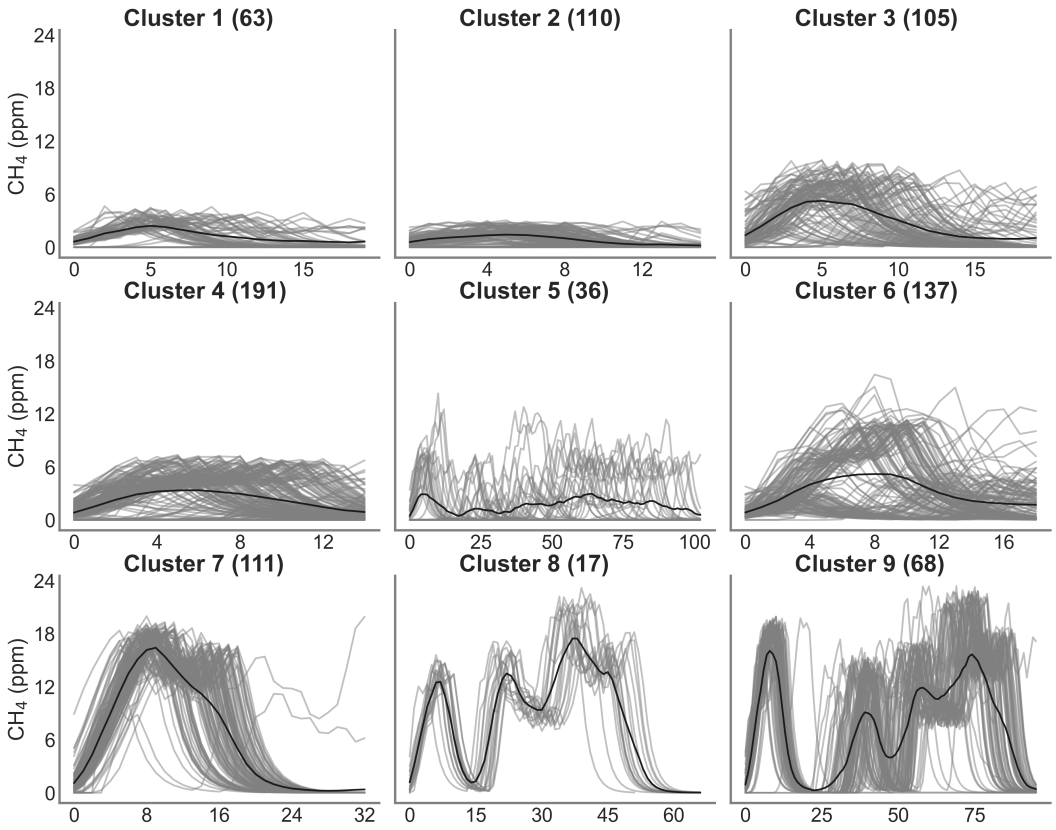


Figure 3.8: Clustering of peaks using DTW on the reference instrument. On the title of each plot the number inside the parentheses corresponds to the number of spikes attributed to each cluster. Thin grey lines represent all the peaks inside each cluster and the black line is the mean of all the peaks corresponding to each class.

Figure 3.9 shows the error of the models against the test set, for each of the training cases listed in Table 3.2, based on spikes chosen from different clusters for doing the training (see Section 3.2.2.4). The results are summarized in Tables B.3 and B.4. First, the polynomial and MLP models performed consistently better than the other models, the MLP being slightly better for most of the cases. In contrast, the linear, random forest and random forest hybrid models had the highest error, regardless of the sensor type or the addition of environmental variables. To compare the performances of the models trained by spikes from different clusters (Table 3.2), we ranked them by their error. The MLP model with type C sensor data as input, and training with spikes from Case 10 (124 spikes) produced the smallest error (MSD = 0.79 ppm^2), followed by the same model for Case 8 (MSD = 0.85 ppm^2, 150 spikes), Case 9 (MSD = 0.86 ppm^2, 94 spikes) and Case 11 (MSD = 0.87 ppm^2, 84 spikes). For the MLP model, Case 4 (147 spikes), Case 1 (587

spikes) and Case 7 (166 spikes) performed slightly worse, with a MSD = 0.89 ppm^2. Finally, Case 2 (MSD = 0.9 ppm^2, 122 spikes), Case 3 (MSD = 0.91 ppm^2, 150 spikes), Case 6 (MSD = 0.93 ppm^2, 105 spikes) and Case 5 (MSD = 0.95 ppm^2, 198 spikes) showed lower performances. From the model ranking, we derived the following conclusions. Firstly, the smallest error did not correspond to the most parsimonious training set (Case 11) but to a larger training set (Case 1, 70% of the data). Nevertheless, we found that Case 11, which was constructed with an even selection of spikes from all the clusters, each in a modest proportion (10% from each cluster) provided better performance than most of the other training cases. This result shows that some clusters introduce less information or have redundancy. Overall, the best performances corresponded to Cases 10, 8 and 9, which all included spikes with complex shapes from clusters C7, C8 and C9. Training models with a sample of those spikes thus ensured better model performances.

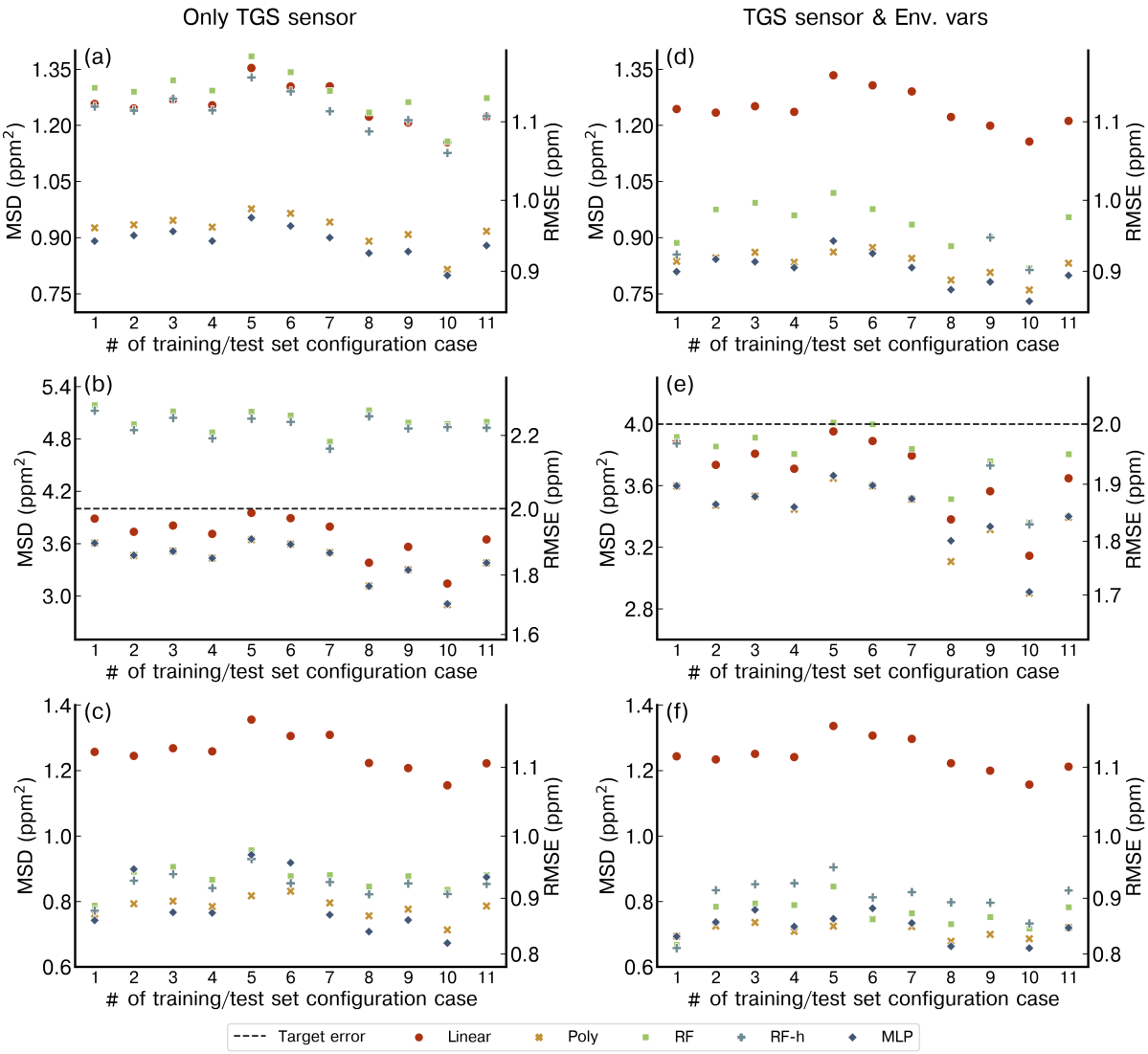


Figure 3.9: Performance of each model for the different configurations of training and test set (1 to 11 in the x-axis) considering the identified clusters. (a) Only Figaro TGS 2611-C00 data as input. (b) Only TGS 2611-E00 data as input. (c) Both Figaro sensors data as input. (d) TGS 2611-C00 data and environmental variables. (e) TGS 2611-E00 and environmental variables. (f) Both TGS sensors and environmental variables. Note the different y-axis for the figure (b) and (e).

3.2.3.4 Results for possible ageing effect on model performance

To test for a possible ageing effect of the sensors, we selected the two best models (polynomial regression and MLP) found in the previous section, and trained them following the best training configuration (Case 10). After being trained using data from the first experiment, these two models were applied to reconstruct the spikes of the second experiment, six months later. A summary of the results is presented on Table 3.3. We observed that after six months, the RMSE

error produced by the models increased from 0.57 to 0.85 ppm. The models trained with type E sensor data showed a smaller degradation (higher RMSE) after six months compared to those trained with the type C sensor data. Considering the amplitude of the peaks that we aim to reconstruct ($\sim$ 24 ppm), a possible drift caused by ageing effects on the sensors appeared to be a small source of error in the reconstruction of CH_4 spikes during the second experiment. Assuming that the error of the sensors increased linearly with time, we determined an error 'drift rate' by computing the ratio of the difference in the error from both experiments divided by the time between them. We observed that for all the cases, the difference in the error is less than 1 ppm after six months and the mean RMSE on the second experiment is less than 2 ppm in all cases, except for the models trained with only the type E sensor. Thus, even with aging the type E sensors would still meet our requirement of a RMSE smaller than 2 ppm. This shows the capability of our models to reconstruct spikes despite possible ageing effects of the sensors.

Table 3.3: Comparison of error for reconstructing spikes in experiment 2, using the two best models (polynomial and MLP) trained with the best training set configuration during experiment 1.

	Mean RMSE$_1$* (ppm)	**Mean RMSE$_2$** (ppm)**	**Difference (ppm)**	**Monthly RMSE increase (ppm month^{-1})**
Poly (C)	0.96	1.82	0.85	0.14
MLP (C)	0.95	1.75	0.80	0.13
Poly (E)	1.84	2.53	0.69	0.11
MLP (E)	1.84	2.41	0.57	0.09
Poly (C+E)	0.89	1.58	0.69	0.11
MLP (C+E)	0.86	1.51	0.64	0.10

* For spikes reconstructed during experiment 1.
** For spikes of experiment 2 reconstructed with models trained on experiment 1.

3.2.3.5 Generalized models

In this section, we address the comparison of model performances when we train a model on a subset of the sensor data from one chamber and reconstruct the spikes of the other chambers. Table 3.4 presents a summary of the number of spikes, observations and clusters analysed for each chamber. The number of clusters, as well the number of spikes, were not equally captured by all the chambers. Only three chambers, A, F and G, shared the same number of clusters.

Chambers C had a more limited number of peaks, due to a reduced sampling period.

Table 3.4: Summary of spikes, observations and clusters detected following the procedure explained on Section 3.2.2.4 for chambers A, C, F, and G.

Chamber	Number of Observations	Number of Spikes	Number of Clusters
A	35536	836	9
C	35499	902	7
F	50089	861	9
G	50569	612	9

To illustrate the performance of models for their ability to be generalized from one chamber to another, we selected the polynomial model with input data from the type C sensor (Fig. 3.10) and from the type E sensor (Fig. 3.11). The same results with the MLP model are shown in figures B.7 and B.8, respectively. The data in Figure 3.10 indicate that the error was lower for the test set of the chamber on which the model was trained than for the test sets of other chambers, as expected. In Figure 3.10a, c and d, we observed that the models trained with the data from chambers A, F or G produced good performances for reconstructing the spikes of another chamber, and met the requirement target of an RMSE $\leq$ 2 ppm. The models trained with the data from chamber C (fig 3.10b) however, performed poorly in reconstructing the spikes from the other chambers and met the target requirement only when trained using data from the same chamber. The performances of the MLP model were similar to those of the polynomial model in terms of generalization from one chamber to another. When trained by data from the type E sensor, our models were found to be less transferable from one chamber to another, meaning they had a larger error for the test sets of another chamber than for the one used for training (Figure 3.11). We inferred that the reconstruction of spikes from models of other chambers needs to be coherent with the number of clusters of the chamber used for training in order to ensure transferability of the models. This is the case for chambers A, F and G for which nine clusters were detected and the distribution of peaks within the clusters was similar (Figure 3.8, 3.9 and B.10). On the other hand, if the clusters are not similar between chambers, the transferability of models is lower.

3.2.4 Discussion

Our results show that a pre-processing of the data to remove H_2O effects and separate spikes from ambient air CH_4 variations, followed by a careful definition of the training set provides capabilities for different models to reconstruct the CH_4 spikes on a 5 s time step, across a large range of concentration variations and spike durations, meeting our requirement o a target error of RMSE $\leq$ 2 ppm. The TGS 2611-E00 (Type E) was the sensor with the poorest performance, regardless of the model employed, or of the subset of data used to train models, as shown by our tests with 5 chambers, each containing 5 different sensors. The model performances for TGS 2611-E00 were thus always poorer than for TGS 2611-C00 (Type C), with a degradation in the reconstruction coming from the larger misfit of the phase of the spikes signal than with the TGS 2611-E00 sensors. This probably is related to the carbon filter that is integrated within this type of sensor, to improve the selectivity. An additional step of the pre-processing algorithm could help to correct problems due to the carbon filter. The MSD error decomposition showed that the sources of error in the reconstruction were mainly from an inaccurate reconstruction of the phase, followed by a misfit of the magnitude of the spikes. The inclusion of environmental variables reduced the LCS component of the MSD, especially for non-parametric models. Nevertheless, for the Type E sensor, adding environmental variables increased the error in the reconstruction of the magnitude. Finally, we found that the error always increased with the reduction of the length of the training set, as previously shown by Rivera Martinez et al. (2021). This sensitivity to the training set mainly affected the non-parametric models due to their limited capability of extrapolation and their requirement of large datasets to keep good performances.

How do our approach and results compare with previous studies? Malings et al. (2019) performed a comparison of different calibration approaches including linear, quadratic, Gaussian models, clustering models, ANN and hybrid random forest models across low-cost sensors measuring different species (CO_2, CO, NO_2, SO_2 and NO) with the aim to calibrate Real-time Affordable Multi-Pollutant monitors (RAMP) to assess the air quality within a city, using a network of sensors. Their set of sensors included an NDIR CO_2 sensor, an Alphasense photoionization

detector and an Alphasense electrochemical unit. They found that a quadratic regression and a hybrid RF model produced the best performance across different pollutants for training sets with durations between 21 and 28 days, and observations with a resolution of 15 minutes. Our results showed that the hybrid random forest model did not perform as well as the polynomial model or the MLP for the reconstruction of CH_4 spikes using data from TGS sensors, and that these models were sensitive to the length of the training set for the k-fold cross validation. An improvement of our models' performances could be achieved with a selection of the proportion of observations used for the parametric model. Nevertheless, the polynomial model gave consistently better results regardless of the inclusion of environmental variables.

Casey et al. (2019); Rivera Martinez et al. (2021) and Eugster et al. (2020) used ANN models to derive CH_4 concentration from observations of TGS sensors and obtained good performances. Casey et al. (2019) suggested that the inclusion of correlated species (e.g. CO_2) rather than the type of sensor led to better performance for their MLP model to reconstruct CH_4. The performance of their ANN model to reconstruct CH_4 variations provided an RMSE of 0.13 ppm for a range of variation between 1.5 and 4.5 ppm. Eugster et al. (2020) also found that the inclusion of other driving variables could increase the performance of ANN models. Their overall model performances for seven years of continuous CH_4 monitoring on ambient air in northern Alaska (range of variation between 1.7 and 2.1 ppm) with a Figaro TGS2600 gave an RMSE of the residuals of 0.043 μmol mol^{-1} (0.69 ppm). Our results showed that different types of TGS sensors used with the same model gave complementary information by reducing the error of the reconstruction and should be used, especially with non-parametric models. The performance of our best model for CH_4 spikes with concentrations much larger than those measured by Eugster et al. (2020), produced under controlled laboratory conditions, provides a mean RMSE of 0.9 ppm for a range of CH_4 variation between 3 and 24 ppm, thus rather comparable results. Regarding the calibration strategy, the clustering approach allowed us to determine nine clusters of spikes in our dataset, with three of them regrouping the largest peaks with complex shapes. This classification allowed us to understand the impact of each cluster in the training. Cluster 9, composed with peaks of complex shape and a range of variation between 3 and 24 ppm was the one that

provided the best information for training the models, due to the fact that spikes present on this cluster include information of larger and shorter peaks, medium peaks and larger peaks with patterns on top of the peaks. With the parsimonious training using Case 10, corresponding to a high proportion of peaks from cluster 9, we were able to reduce the length of the training dataset from 70% to 25% while maintaining similar performance. This approach has a strong potential to reduce the length of the training set by selecting only observations from specific clusters defined from the data, and which represent the entire dataset.

Concerning the ageing effect of the sensors, after six months, we observed only small increases in the RMSE of our models, between 0.6 to 0.8 ppm corresponding to an error increase rate of 0.1 ppm per month. Our results also showed the capability to transfer the models from one chamber to another, provided that the chamber used for testing contains data with the same range of CH_4 variations as the chamber used for training, which can be assessed by our clustering analysis of the data.

3.2.5 Conclusions

We performed a systematic comparison of different parametric and non-parametric models to reconstruct atmospheric CH_4 spikes under laboratory conditions, based on the voltages recorded by low cost metal oxide sensors. Other environmental variables such as temperature, pressure and water vapor were used. The true CH_4 time series comes from a high precision instrument run alongside the low-cost sensors. The best models were a 2nd-degree polynomial function and a multi-layer-perceptron model. These two models both meet our requirements of a RMSE smaller than 2 ppm. We found that the main limitation was the large fraction of data (70%) needed to train the model. This would limit the use of low-cost sensors in the field, as they would need to be frequently trained with an expensive instrument at the same location. This limitation was partly overcome by adopting a stratified training strategy, namely to perform the training on fewer but more influential spikes selected into orthogonal clusters applied to the whole dataset. This parsimonious training allows to use only 25% of the data to keep a model performance compliant with our 2 ppm RMSE threshold. We also showed that sensors' ageing effects after six months

did not degrade too much the performances of our models. Finally TGS 2611C-00 was superior to TGS 2611-E00 model. For this experiment we generated about 800 peaks with some predefined shapes, future implementations should consider increase the diversity of shapes and durations of the generated peaks. Regarding the models employed, we assessed the performances of models that considers no time dependency in the signal, more complex models that allows to include the time dependence such as Recursive neural networks (RNN) should be tested.

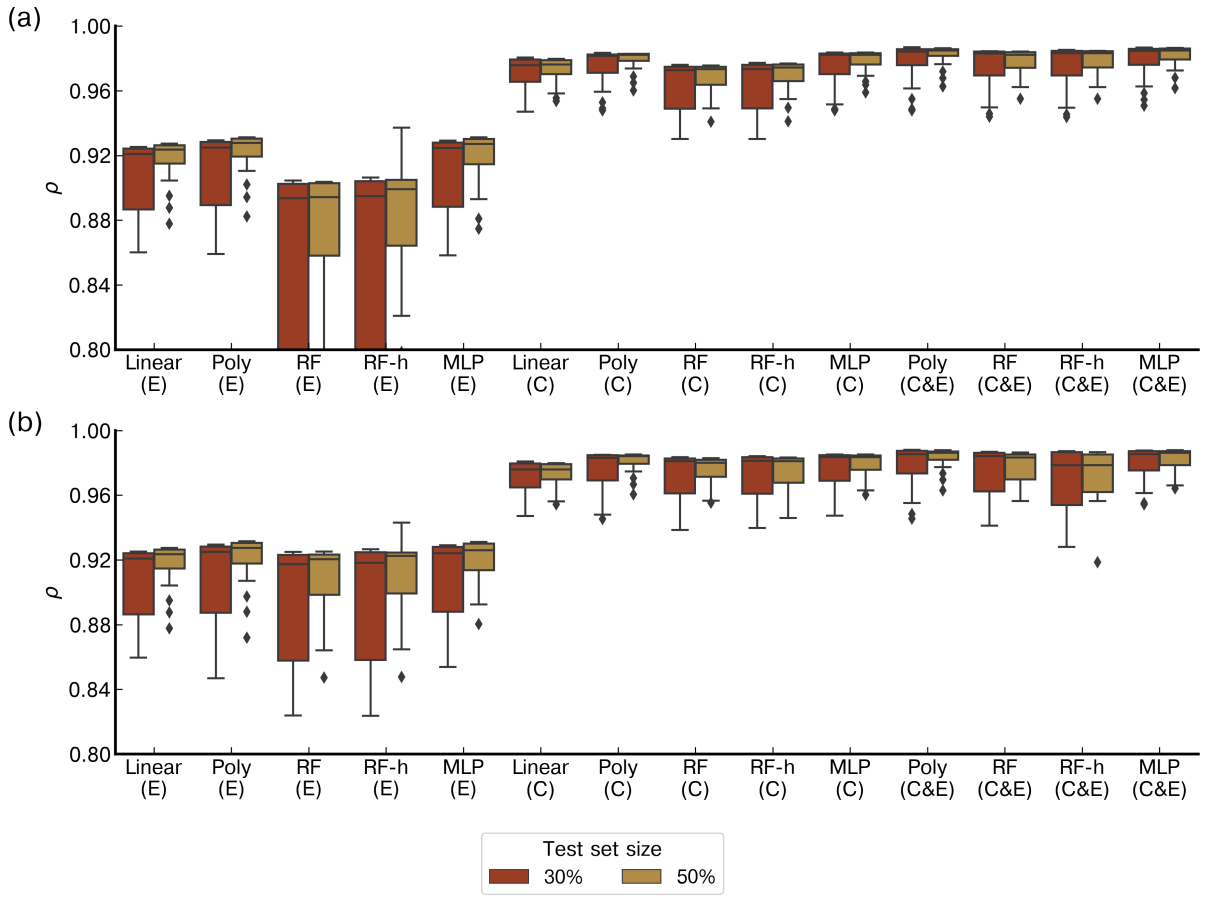


Figure 3.6: Comparison of the Pearson correlation coefficient (ρ) distributions between models on the test set for a 20-fold cross validation. The boxes are the inter-quartile of the distribution of ρ, the whiskers are the 5^{th} and 95^{th} percentiles, and the black line is the median (a) Models in which the inputs are only voltage from the Figaro TGS sensors, (b) Models in which the inputs include voltage from low-cost sensors and environmental variables (H_2O, Temperature and Pressure). 'Linear' represents the linear or multilinear model, 'Poly' the polynomial model, 'RF' the random forest model, 'RF-h' the random forest hybridized with a polynomial regression, and 'MLP' the multilayer perceptron. Under each model, labels denote which TGS sensor was used; 'C' is the TGS 2611-C00, 'E' the TGS 2611-E00 and 'C & E' both sensors at the same time. The red boxplots represent the results of models with a test set size of 30% of the total observations and the yellow ones a test set size of 50%. Note that the y-axis was limited in a range to distinguish the different models.

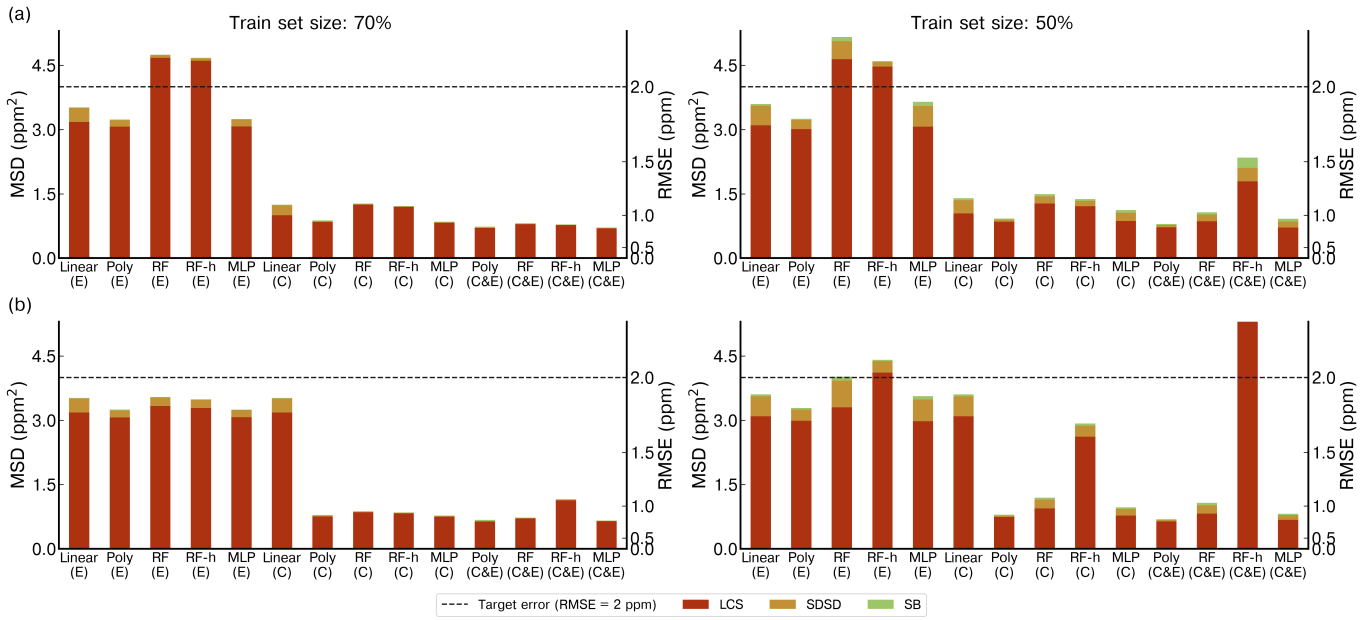


Figure 3.7: Comparison of the Mean standard deviation (MSD) across the different models on the test set for a 20-fold cross validation. (a) Models with only voltage of TGS sensors as input. (b) Models including environmental variables and voltage of TGS sensors in the input. Left panels show the performances on a train set size of 70% and right panels a train set size of 50% of the total observations. The stacked bars show the contribution of each component of the MSD to the total error, the Lack of Positive Correlation weighted by σ (LCS) in red, the difference in the magnitude fluctuation (SDSD) in orange and the simulation bias (SB) in green. Notation for the models is the same as for the figure 3.6.

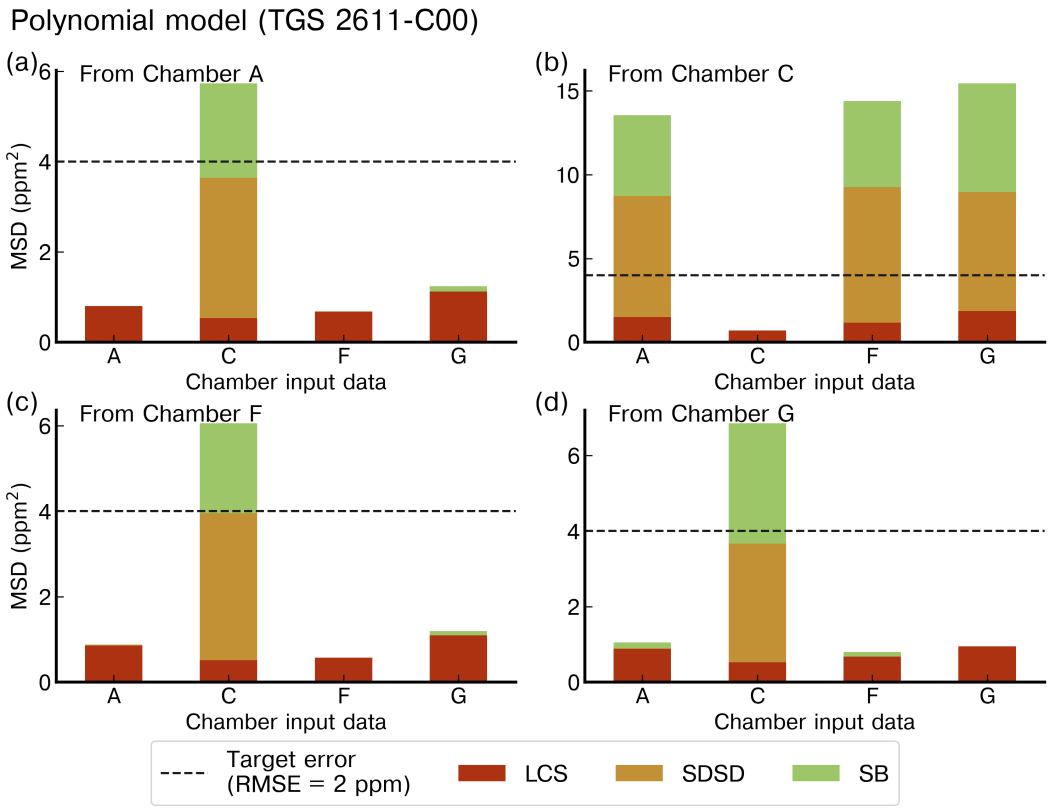


Figure 3.10: Reconstruction error of the peaks for the polynomial model with TGS 2611-C00 as input using the best stratified training case from (a) Chamber A, (b) Chamber C, (c) Chamber F and (d) Chamber G to reconstruct the peaks from the other chambers (listed on the x-axis) with data from the same type of sensor. Note the different ranges of the y-axis for the panels (b) and (c).

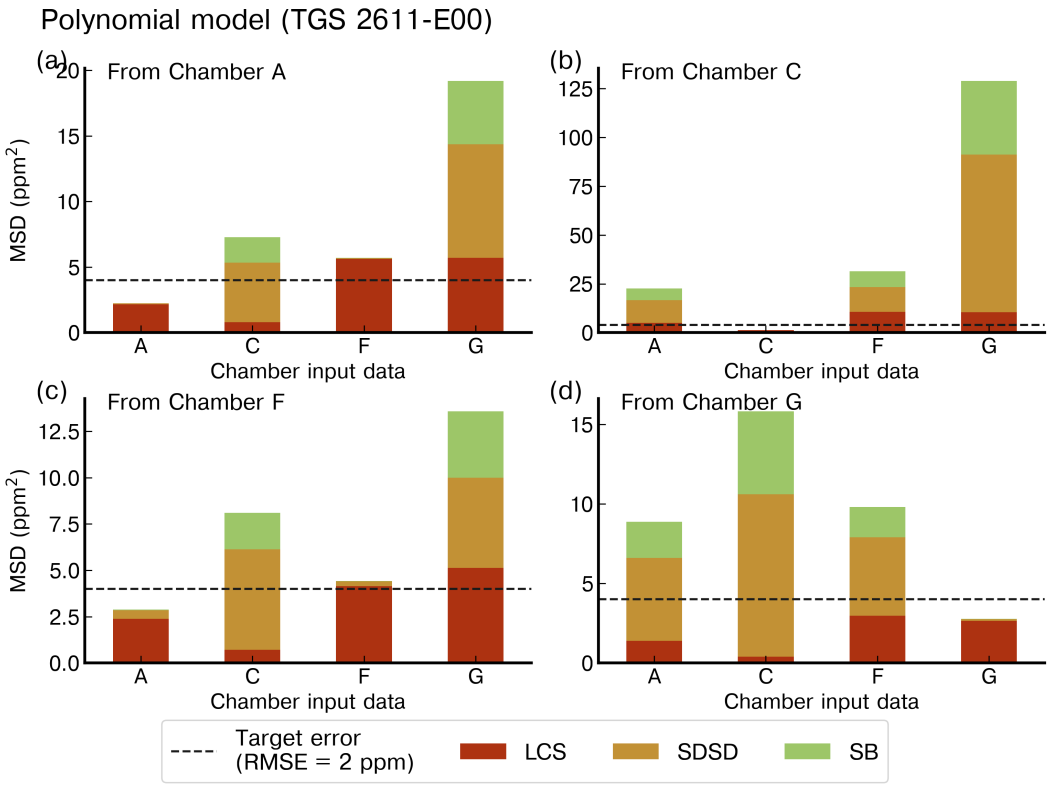


Figure 3.11: Reconstruction error of the peaks for the polynomial model with TGS 2611-E00 as input using the best stratified training case on (a) Chamber A, (b) Chamber C, (c) Chamber F and (d) Chamber G to reconstruct the peaks from the other chambers with data from the same type of sensor. Note the different ranges of the y-axis.

Chapter 4

Using metal oxide gas sensors for the estimate of methane controlled releases: reconstruction of the methane mole fraction time-series and quantification of the release rates and locations

Rodrigo Rivera Martinez[1], Pramod Kumar[1], Olivier Laurent[1], Gregoire Broquet[1], Christopher Caldow[3], Ford Cropley[1], Diego Santaren[1], Adil Shah[1], Cécile Mallet[2], Michel Ramonet[1], Leonard Rivier[1], Catherine Juery[4], Olivier Duclaux[4], Caroline Bouchet[5], Elisa Allegrini[5], Hervé Utard[5], and Philippe Ciais[1]

[1]Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France
[2]Université de Versailles Saint-Quentin, UMR8190 – CNRS/INSU, LATMOS-IPSL, Laboratoire Atmosphères Milieux, Observations Spatiales, Quartier des Garennes, 11 Boulevard d'Alembert, 78280 Guyancourt, France
[3]Clean Energy Regulator, Discovery House, 47 Bowes St, Phillip ACT 2606, Australia
[4]TotalEnergies - OneTech, Laboratoire Qualité de l'Air – 69360 SOLAIZE FRANCE
[5]SUEZ - Smart & Environmental Solutions; Tour CB21, 16 place de l'Iris, 92040 La Défense France.

Awaiting for submission. Expected submission: December 2022

4.1 Summary of the publication

4.1.1 Introduction and motivation

CH_4 emissions can occur in all the chain of oil and gas production, from the extraction, transportation, storage, distribution and use (Saunois et al., 2020). Atmospheric measurements are usually used with local dispersion models to infer the source location and the emission rates of CH_4leaks at industrial facilities. Measurements are usually obtained from surveys as close as possible to the source location. Although these surveys are labor intensive and costly, they do not provide a continuous monitoring of the sources (Travis et al., 2020). Deploying a permanent network of sensors to continuously monitor CH_4 leaks at industrial facilities would overcome current limitations of surveys and provide more accurate estimates thanks to a continuous sampling of the plumes.

Current surveys use research level instruments, such as CRDS. They are highly precise and require low maintenance but they are costly, limiting the deployment of such networks. Metal oxide semiconductors (MOS) are one of the most common technologies of low-cost sensors studied for the detection and quantification of CH_4. They present some advantages, such as being compact and suitable for long time deployment, making them very appropriate for deployments of dense networks. However, they present low accuracy, low selectivity and they drift with time requiring periodic calibrations.

In this chapter we report the results from the third experiment where we test the ability of MOS sensors, specifically two versions of Figaro® TGS, to reconstruct an actual signal from a source in open air corresponding to a series of controlled releases. We also assess the accuracy of the emission estimates from the reconstructed CH_4 concentrations of these sensors.

4.1.2 Design of the experiment

We participated in the TOTAL Energies® Anomaly Detection Initiative (TADI) campaign 2019 from October 2nd to October 10th. It consisted of 41 controlled releases of CH_4 and CO_2 covering

emissions between 0.15 to 150 g CH_4 s^{-1} and durations ranging between 25 to 75 minutes. The site was equipped with common elements that can be found in oil and gas facilities, such as, pipes and valves, to simulate typical leaks in a real case scenario. Releases were emitted at different heights and locations inside an ATEX zone of 40 m x 50 m.

We installed 16 sampling lines circling the ATEX zone. Only seven sampling lines were active to measure each release and they were activated depending on wind conditions. This strategy provides us with several measurement points within and outside the plume despite weather conditions. Each active sampling line was connected to a high precision instrument, a CRDS or a LGR MGGA, and to a chamber containing at least three types of TGS sensors (TGS 2600, 2611-C00 and 2611-E00) and two sensors that measured environmental parameters, such as, relative humidity, temperature and pressure.

From the 41 releases, eight releases are excluded from this study and correspond to one of three cases: i) they were not measured by the TGS sensors, ii) the enhancements were too small to be detected by TGS sensors or iii) the amplitude of peaks measured by the TGS sensors does not allow to distinguish large peaks from medium and small ones.

Each release is characterized by a series of CH_4 spikes at high frequency over the background produced by the plume reaching and leaving the sampling inlet of the sensors due to the turbulence and high frequency variations of the wind. Therefore, to separate the TGS voltage enhancements linked to the variations of the CH_4 spikes from the reference signal, we used an iterative algorithm. It considers neighboring observations to determine baseline conditions from the spike ones.

The definition of the training and test sets used in the reconstruction models was constrained by the number of releases available per chamber and the number of releases measured simultaneously by several chambers. These considerations led to a ratio of releases per chamber: 60% in the training set, and 40% in the test set. Reconstructed concentrations of the test set were used to feed the inverse modelling in order to estimate the rate and the emissions location.

For the reconstruction of CH_4 concentrations, we assessed the performances of a 2nd degree polynomial and a Multilayer perceptron (MLP) under three input configurations: i) using only TGS

2611-C00, ii) using only TGS 2611-E-00, and iii) using both sensors at the same time. In order to evaluate the release signal reconstruction performance, we considered a notional target error of an adjusted RMSE $\leq$ 0.1 ppm, computed by the RMSE and weighted by the inverse of the maximum peak present in the release.

The inverse modeling framework used in this study relies on the framework developed by Kumar et al. (2022). It uses a gaussian model and its adjoint to simulate the CH_4 mole fraction enhancement above the background at fixed measurement stations corresponding to a release rate and location. The optimal horizontal and vertical location and rate are derived in terms of misfits between the averages of the observed and simulated CH_4 enhancements over the background. The optimal release location and rate is determined simultaneously by looping on a large ensemble of potential locations and by identifying the optimal rate and location by the smallest misfits. The ATEX zone is then discretized in a 3D grid (1 m x 1 m x 0.5 m) to define the potential locations. The bins of the measurements and of the simulated mole fractions correspond to sectors of wind direction of equal ranges during the release. The computation of the inverse modelling on the reference data was restrained to the station and time corresponding to the available data from TGS sensors.

4.1.3 Main results

This study allowed to prove the capability of measurements from MOS sensors to produce estimates of the emission rate and location from controlled CH_4 releases, similar to typical leaks at industrial facilities.

We were capable to derive a baseline signal, used to extract the voltage enhancements of TGS linked to spike conditions, by using an iterative process of successive comparisons between neighboring observations without information of environmental parameters. However, this approach is only useful for signals composed of spikes at high frequency produced by the plume reaching and leaving the inlet tube of the sensors due to turbulence and wind conditions.

The two models assessed, a 2nd degree polynomial and MLP, to reconstruct CH_4 concentrations from voltage measurements of TGS sensors, show similar performances. However, some

differences are observed between them. In first place, the polynomial model produced a slightly better reconstruction with a small training set corresponding to observations on Chamber D. Secondly, the present noise in voltage observations on some releases was reduced using the MLP models, but only when the input was the Type C sensor. Regarding the configuration of inputs on the models, Type C sensor produced better reconstruction of the concentrations. The combination of both sensors produced a similar reconstruction to the case using only type C sensor. Only chamber H met our target error (0.03 ppm $\leq$ adjusted RMSE < 0.1 ppm).

Concerning the inversion of emission rates and locations, in general, we obtained good estimates from inversion of reconstructed CH_4 spikes from TGS data. The simulated gradients were close to the observed and simulated gradients from the reference data, where observations were restrained to the data availability from the TGS sensors. In most of the cases it produced an emission rate error of 25% compared to 30% for the estimations using the reference data, and a location error of 9.5 m for TGS data compared to 7.8 m using the reference.

4.2 Publication

4.2.1 Introduction

Fossil fuel anthropogenic methane (CH_4) emissions related to the production, exploitation and transport of coal, oil and natural gas, account for 35% of global anthropogenic emissions (Saunois et al., 2020). Emissions from natural gas production occur along the chain from extraction, transportation, storage, distribution and use. Emissions estimates reported by inventories rely on information from activity data and emission factors. Emission factors are different between sites, technologies, operating modes and are not stationary, which makes the upscaling of fugitive CH_4 emissions highly uncertain (Alvarez et al., 2018). For instance, emissions from the oil and gas supply chain in the US constrained from ground based and aircraft measurements were found to be 60% higher than the EPA inventory (Alvarez et al., 2018). More generally, the characterization of CH_4 emissions from complex processes based on static emission factors can be challenged when the best practices are not followed by operators (Riddick et al., 2020).

Atmospheric measurements are increasingly used to detect and quantify CH_4 leaks from industrial facilities. The measurements are often interpreted with local-scale dispersion models using atmospheric inversion methods to infer the CH_4 source location and emission rates, see e.g. (Kumar et al., 2022). Current approaches generally consist in conducting atmospheric surveys of the enriched concentration plume created by the emitting source. Difficulties are the accessibility to sample the plumes from emitting locations, labor and instrument costs given that surveys currently employ expensive high precision research-level CH_4 instruments, such as Cavity ring-down spectrometers (CRDS). Further, surveys do not provide continuous monitoring of the sources (Travis et al., 2020). The deployment and functioning of mini-networks of continuous monitoring sensors for CH_4 mole fractions is an alternative to surveys, but the costs of each instrument remain a limitation. Advances in the development of low-cost sensors facilitates the deployment of dense sensors' networks to increase the coverage of a site (Kumar et al., 2015; Mead et al., 2013). Permanent deployment of a network of sensors can overcome limitations in

the quantification of leaks and help to better characterize the plumes by limiting the impact of atmospheric transport modelling uncertainties. In addition, the theoretical study of Chamberland and Veeravalli (2006) proved that performance is improved in differentiation of known signals from noise by increasing the sensor density in an area.

In later years, an increase in the interest in low cost and low power sensors to be used in dense networks led to the study of different kinds of sensors to measure pollutants and trace gases like CO_2 or CH_4. One of the most common low-cost sensors technologies for the detection and quantification of CH_4 emissions is metal oxide semiconductors (MOS). MOS sensors are composed of a metal oxide sensing material and a heater ensuring that the sensing material reaches temperatures between 300 to 500 °C. A chemical reaction affects the electrical conductivity of the sensing material in the presence of an electron donor gas such as CH_4 (Özgür Örnek and Karlik, 2012). The advantages of MOS sensors are that they are compact and very well suited to long time deployment due to their resilience to extreme weather conditions. However, their sensitivity is affected by environmental parameters (temperature and relative humidity) (Popoola et al., 2018) and VOCs; they also present low accuracy and drift with time (in the form of a decrease in the conductance of the sensing material), requiring periodic re-calibrations, and the need of constant power supply due to the heater material.

The Taguchi Gas Sensors (TGS) commercial MOS from the Figaro® manufacturer, were widely tested in different environments under controlled conditions and field deployment due to their sensitivity to CH_4 (Eugster et al., 2020; Eugster and Kling, 2012; Riddick et al., 2020; Collier-Oxandale et al., 2018; Bastviken et al., 2020; van den Bossche et al., 2017). The standard technique to derive a calibration methodology is to collocate these MOS sensors with a high precision instrument used as a reference, then apply empirical equations or data-driven approaches (Eugster et al., 2020; Eugster and Kling, 2012; Casey et al., 2019; Bastviken et al., 2020; Collier-Oxandale et al., 2018, 2019). In a previous work (Rivera Martinez et al., 2021) we have studied the possibility of using Artificial neural Networks (ANN) to reconstruct the variations of CH_4 mole fractions in room air under controlled conditions from three types of Figaro sensors (TGS 2600, TGS 2611-C00 and TGS 2611-E00). A following study (Rivera Martinez et al., 2022) analyzed the potential to recon-

struct spikes of CH_4 generated on top of ambient air observations, that corresponded to typical signals from leaks at industrial sites, employing two types of Figaro sensors (TGS 2611-C00 and TGS 2611-E00). That study made a thorough comparison of the performance of five models for the reconstruction of CH_4.

The next logical step is to test the performances of the same sensors to reconstruct CH_4 from real leaks, and to use the reconstructed mole fractions to quantify emission rates. To our knowledge, only one study attempted to do so. Riddick et al. (2020) quantified emissions of a gas terminal using a Figaro TGS 2600 included in a logger system controlled by an Arduino Uno. The logging system was located 1.5 m from a point source. To reconstruct CH_4 mole fractions from voltage observations, Riddick et al. (2020) developed an empirical equation considering the measured voltage, temperature and the humidity. Then, a Gaussian plume model was used to quantify the emission rate using information from the reconstructed CH_4 mole fractions and wind information from a nearby meteorological station. Their estimates of the emissions rates had an average value of 9.6 g CH_4 s^{-1} and reached a maximum of 238 g CH_4 s^{-1}, given corresponding to enhancements of the CH_4 mole fractions between 2 ppm to 5.4 ppm within the plume. Their estimates based on a Figaro sensor were not confronted with high precision instruments nor with an independent knowledge of emission rate.

In this study, we test the ability of a network of several Figaro sensors to reconstruct the CH_4 atmospheric enhancements from a series of controlled releases of known magnitudes and duration in open air at a facility called TADI (see Methods), and to infer the emission rate of each release by an inverse modeling approach. The accuracy of the CH_4 reconstruction is evaluated against collocated accurate CH_4 measurements from high precision CRDS instruments. The accuracy of the inverted emission locations and rates is evaluated against the known (controlled) location and magnitude using the inversion model of Kumar et al. (2022)

For 33 controlled releases at the TADI facility, we utilized fixed-point measurements from CRDS instruments and low-cost TGS. A large fraction of the TGS measurements were used for training models to reconstruct CH_4 mixing ratios from measured TGS resistance and other variables. We pose a target minimum accuracy for CH_4 reconstruction models equal to 15% of the

amplitude of the largest observed excess within a release. This corresponds to accuracies going from 0.3 ppm for a release causing a maximum excess of 2.4 ppm, to 18 ppm for a maximum excess of 120 ppm. This accuracy is consistent with the accuracy requirement set in our previous study where we used TGS sensors to reconstruct spikes of CH_4 created in a laboratory experiment (Rivera Martinez et al., 2022).

The plan of the study is as follows. Section 4.2.2 presents the TADI 2019 controlled releases campaign, the logger systems, the models employed to reconstruct CH_4 from TGS data, and the atmospheric inversion approach. The data treatment, comparison of the models for the reconstruction of CH_4 and the inversion results for rates and locations of different releases are analyzed in section 4.2.3. Results are discussed in Section 4.2.4, and conclusion are given in section 4.2.5.

4.2.2 Methods

4.2.2.1 Sampling strategy at the TADI-2019 campaign

In October 2019, TotalEnergies® conducted an experiment with multiple controlled releases at the TotalEnergies Anomaly Detection Initiative (TADI), to investigate the capability of detection and quantification of different technologies for local emissions produced on industrial facilities. The TADI test site is designed and operated by TotalEnergies® to test different technologies and methodologies of detection and quantification of gas leaks in an industrial environment, such as oil and gas production facilities. The platform is located northwest of Pau, France, with an approximate area of 200 m × 200 m. The site is equipped with a series of pipes, valves, tanks, , and other equipment commonly found on oil and gas facilities to simulate 'realistic' leaks. The terrain is flat but includes different obstacles that can affect the dispersion of the gases released to the atmosphere. This experiment consisted of 41 controlled releases of CH_4 and CO_2 covering a wide range of emissions between 0.15 and 150 g CH_4 s^{-1} and durations ranging between 25 to 75 minutes. We participated to this experiment to develop and test inverse modelling frameworks within the TRAcking Carbon Emissions (TRACE, https://trace.lsce.ipsl.fr/) program for the estimation of emission location and rates based on CH_4 mole fractions from high preci-

sion instruments (Kumar et al., 2022). We presented the inversion results for 26 releases from single point sources based on two inversion approaches, one relying on fixed-point measurements, and the other one on mobile near-surface measurements (the latter had already been documented in (Kumar et al., 2021)). In both cases, the estimates of the emissions relied on CH_4 mole fractions from high precision instruments, and on a Gaussian plume model to simulate the local atmospheric dispersion of CH_4. The results from Kumar et al. (2022) proved to be relatively good, with an error in the release rate estimates from fixed-point measurements between $\sim$23 to $\sim$30 % and an error in the location of the point sources (within a 40 m × 50 m area) of between 8 and 10 m.

The controlled releases were emitted at different heights up to 6 m above the ground, and inside the 40 m × 50 m ATEX zone of the TADI facility (see Fig. 4.1). More information on the site infrastructure and on these experiments in October 2019 are presented respectively in Kumar et al. (2021) and Kumar et al. (2022).

The multiple controlled releases experiment was conducted from October 2, 2019, to October 10, 2019. Our atmospheric sampling configuration for measuring CH_4 is shown in figure 2. It consisted of placing 16 sampling lines on the ground connected on one end to air intakes in tripods at heights between 2.75 to 3.50 m around the ATEX zone and on the other end to a pump flushing at 6 LPM (KNF N811 with PTFE diaphragm). The lengths of the sampling lines varied from 10 m to 100 m connecting each tripod air intake to CH_4 measurement instruments located inside a tent. The pump was connected upstream from the high precision instruments (Picarro CRDS or LGR), a chamber containing a series of TGS CH_4 sensors, and other sensors measuring environmental parameters such as relative humidity, pressure and temperature. To maintain the inline pressure at atmospheric pressure, a vent was also connected to each sampling line (Fig. 4.2).

Table 4.1 summarizes the species measured, and the identifiers of the reference high precision instruments. All reference instruments measured H_2O to provide dry mole fractions of the species. The analyzers' sampling frequency ranges between 0.3 to 1 Hz. In a previous study by Yver-Kwok et al. (2015), it was proven that those analyzers ensure high precision measurements

and a low drift over time, less than one ppb per month. Yet, two calibrations were conducted before and after the campaign. On average 6-7 sampling lines were active for each release, each active line being connected to a high precision instrument and a TGS chamber. The lines were activated depending on wind direction. The strategy behind the distribution of the tripods around the emitting area and for the inversion was to acquire continuously several measurements points within the plume generated by each release, in addition to one or few measurements points outside the plume (to characterize the background level upon which plumes enhancements can be assessed) for each release regardless of the wind conditions (Kumar et al., 2022).

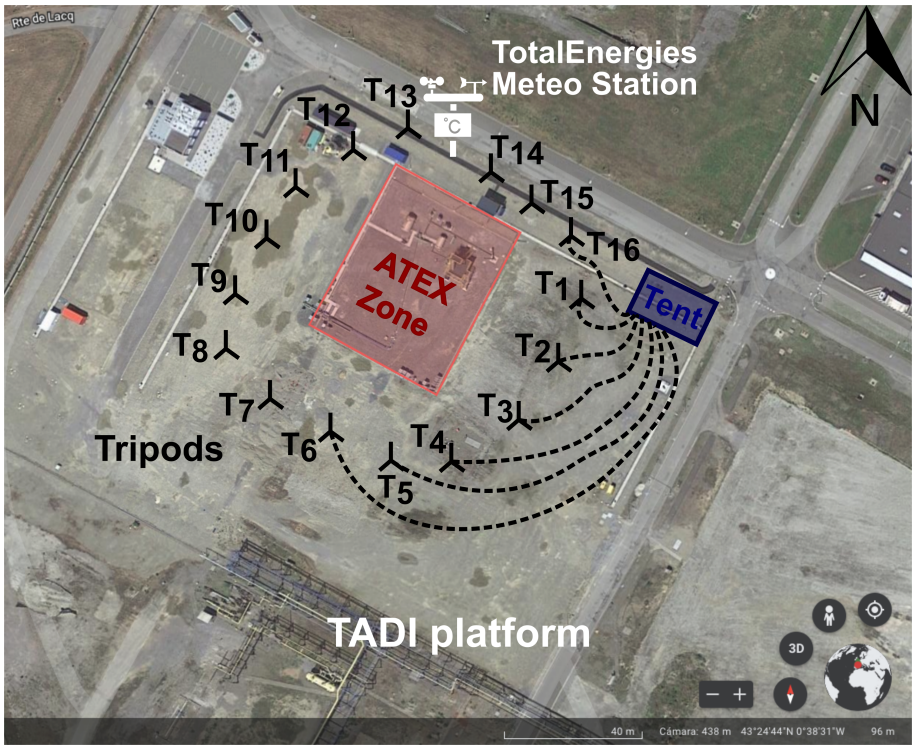


Figure 4.1: Diagram of the experimental setup on top of a satellite image of the TADI platform (source: Google Earth©). The locations of the releases are inside the red rectangle (ATEX zone). The locations of the 16 tripods are presented as black symbols and denoted with a Tx where x is the index of the tripod from 1 to 16. The blue rectangle indicates the tent location. Examples of the sampling lines connecting the tripods to the tent are shown as dashed lines, only showing 7 of 16 in total. The white symbol shows the location of the Meteorological station installed by TotalEnergies® .

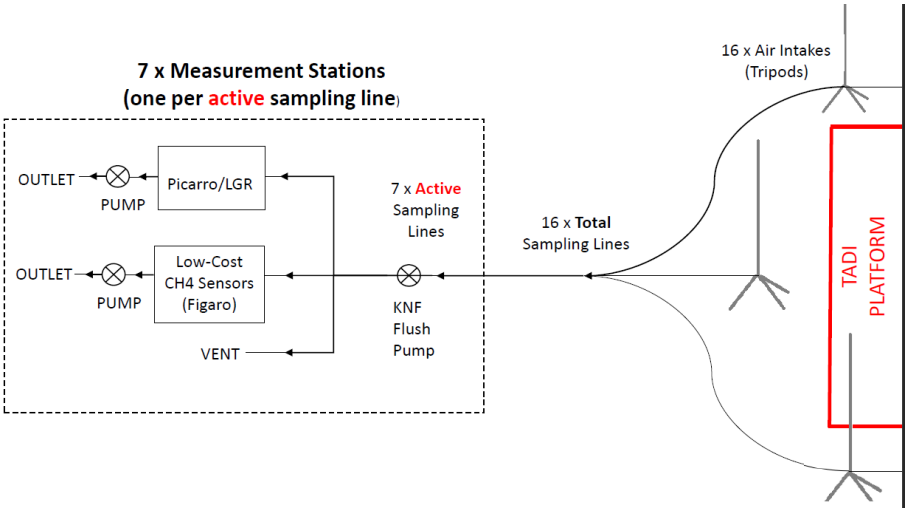


Figure 4.2: Diagram of the measurement stations and their connection to the sampling lines.

Table 4.1: Summary of the species measured by each reference instrument.

Serial number / Code	Identifier	Species measured
CFKADS2286 / Picarro 1	Picarro CRDS G2401	CH_4, CO_2, CO
CFKADS2301 / Picarro 2	Picarro CRDS G2401	CH_4, CO_2, CO
CFKADS2194 / Picarro 3	Picarro CRDS G2401	CH_4, CO_2, CO
CFKADS2131 / Picarro 4	Picarro CRDS G2401	CH_4, CO_2, CO
CFIDS2067 / Picarro 5	Picarro CRDS G2201-i Isotopic	$^{13}CH_4$, $^{12}CH_4$, $^{13}CO_2$, ^{12}CO
CFIDS2072 / Picarro 6	Picarro CRDS G2201-i Isotopic	$^{13}CH_4$, $^{12}CH_4$, $^{13}CO_2$, ^{12}CO
LGR MGGA	Los Gatos Micro-portable Greenhouse gas analyzer	CH_4, CO_2

Low-cost CH_4 sensors logger system

Seven chambers were assembled for the campaign. Table 4.2 shows the sensors included in each chamber, the environmental variable sensors included, and the type of chamber. Each chamber contained at least three TGS with voltage measurements sensitive to CH_4 and two sensors other sensors measuring relative humidity/temperature and pressure/temperature. All these sensors were inserted inside an acrylic/glass or steel/glass chamber with volumes of 100 ml and 120 ml, respectively. The sensitivity of TGS was controlled by a load resistor connected in series to the sensor (Figaro®, 2013, 2005), two values of resistor were used, 5KΩ and 50KΩ (see table 4.2 for details). And AB Electronics PiPlus ADC board mounted on a Raspberry Pi 3B+ recorded the voltage on the load resistor providing observations every 2s. This voltage is used for the characterization and reconstruction of the CH_4 signal. Consistency was observed between the

two TGS 2611-E00 sensors installed on chamber E, and only one sensor of this type is used in this study.

Measurements of environmental parameters from the other chambers, besides chamber E, had data gaps for extended periods or bad recordings occurring at releases and were not included. This study focuses on reconstructing CH_4 using data from TGS 2611-C00 and TGS 2611-E00 from chambers A, C, D, E, F, and H. Data from TGS 2600 were discarded since this sensor did not respond to most of the CH_4 peaks during the releases (see Figure C.1).

Table 4.2: Summary of the specifications of the chambers, the tripods to which each chamber was connected, the captured releases and the reference instrument collocated with each chamber.

Chamber	Figaro TGS sensors	Load resistor (Ω)	Other sensors	Chamber type	Tripod	# of measured releases	Reference instrument
A	2600 2611-C00 2611-E00	50K	DHT22 BMP280	Acrylic/glass	1, 4, 6 8, 9, 10 11, 14, 15	28	Picarro CFKADS2286
C	2600 2611-C00 2611-E00	50K	SHT75 BMP280	Acrylic/glass	2, 7, 9 14, 15, 16	24	Picarro CFIDS2072
D	2600 2611-C00 2611-E00	5K	SHT75 BMP280	Steel/glass	2, 3, 9 10, 11, 12 13, 16	14	Picarro CFKADS2301
E	2600* 2611-C00* 2611-E00*	5K	DHT22 SHT75 BMP180	Steel/glass	1, 3, 4 5, 10, 11 12, 13, 16	24	Picarro CFKADS2131
F	2600 2611-C00 2611-E00	50K	SHT75 BMP280	Acrylic/glass	2, 3, 4 10, 11, 12 13, 14, 15	25	Picarro CFKADS2194
H	2600 2611-C00 2611-E00	50K	SHT75 BMP180	Acrylic/glass	4, 5, 6 7, 12, 13 14, 15	22	LGR MGGA

* Two versions of each type.

TADI controlled releases and meteorological data

A total of 41 controlled releases were conducted during the seven days of experiment, with release durations varying between 25 minutes to 75 minutes. Because low wind conditions (Ur < 0.6 m s^{-1}) are not suitable for the atmospheric inverse modelling, six releases corresponding to such low wind conditions have been excluded for the inversion modelling here such as in Kumar et al. (2022). However, they are used in the training of the CH_4 reconstruction models. There was no TGS measurements during the five releases corresponding to the last day of the campaign. Two largest releases produced high CH_4 mole fraction plumes that affected the amplitude mea-

sured by the TGS sensors on which it was not possible to distinguish large spikes from medium and small ones on the measured voltage (see fig C.3) and they are removed from the study. One release was aborted due to technical problems at the site and is as well removed from this work. This study is thus focused on 33 out of the 41 controlled releases. A summary of these releases is presented in table C.1, where is marked with an 'x' where data measured by the chambers are invalid due to some releases yielded small peaks over the background signal (with enhancements of less than 4 ppm) and TGS sensors were not able to detect such small variations.

The protocol followed in the selection of the releases used in the training and test set for the reconstruction models is explained in section 2.2.2. A meteorological station was installed on the TADI platform by TotalEnergies® with a sonic 3D anemometer at 5 m height above the ground surface (see Fig. 4.1), providing 1-minute averages of wind speed (Ur), wind direction (θ) and of the standard deviation of wind speed on the three axes (σ_u, σ_v and σ_w) amongst other parameters. The data of turbulence and meteorological conditions are used in the dispersion model. Table 4.3 gather general information for each of the 33 controlled release during which we have valid TGS measurements: the duration of the release, the actual release rate, the average wind speed over the duration of the release and an indication showing if it was selected for the inverse modelling.

4.2.2.2 Reconstruction of spikes in CH_4 mixing ratios caused by the releases

The chambers full of TGS sensors captured different portions of the plume with variations at high frequencies due to the distribution of the tripods with regards to the variable wind direction and due to the turbulence. The typical signal measured by the chambers is a series of spikes, ranging between 1 and 15 minutes, corresponding to the plume lying over a slowly varying background signal associated to remote emissions. The targeted signal is that of the difference between the spikes and the background (Kumar et al., 2022). As an example, Figure 4.3 shows 1-minute averages CH_4 mole fractions measured by the reference instruments and the voltage from the TGS 2611-C00 at six tripods during release 25 (Q_s = 5 g s^{-1}). We can observe that CH_4 of the reference instrument and TGS voltage show good consistency at this temporal resolution. Chamber A, C and D were in the trajectory of the plume or very close to it measuring peaks up to 30 ppm,

Table 4.3: Summary of the information for the controlled releases with single CH_4 point sources during the TADI 2019 campaign. Rows in gray shows the releases with low wind speed conditions.

Release number	Duration (hh:mm)	Emission rate (Q_s ($g\ s^{-1}$))	Average wind speed (U_r ($m\ s^{-1}$))	Used in the atmospheric inverse modelling
1	00:58	CH_4: 10	2.76	No
2	00:32	CH_4: 1	3.31	Yes
3	00:33	CH_4: 0.5	3.56	No
4	00:33	CH_4: 5	3.91	No
5	00:35	CH_4: 3, CO_2: 85	0.65	Yes
6	00:39	CH_4: 0.5	0.45	No
7	00:46	CH_4: 5.0	0.80	No
8	00:50	CH_4: 0.5 & 0.75 & 0.5 *	1.41	No
9	00:38	CH_4: 1, C_2H_6: 0.5	1.46	Yes
10	00:38	CH_4: 0.5	2.17	Yes
11	00:30	CH_4: 0.16	2.39	No
12	00:46	CH_4: 1	0.93	Yes
13	00:44	CH_4: 0.2	0.26	No
14	00:55	CH_4: 0.5 & 1.0 *	0.07	No
15	01:01	CH_4: 2	3.50	No
16	00:44	CH_4: 2	1.83	No
17	00:50	CH_4: 4	1.45	No
18	00:48	CH_4: 0.3	0.13	No
19	00:40	CH_4: 2.0	0.41	No
20	00:58	CH_4: 2 & 4 *	0.47	No
21	00:44	CH_4: 1	1.31	Yes
22	00:33	CH_4: 1, C_2H_6: 0.2	1.11	No
23	00:50	CH_4: 2	1.84	No
24	00:43	CH_4: 150	2.63	No
25	00:35	CH_4: 5	3.12	Yes
26	00:48	CH_4: 0.4	2.73	Yes
27	00:37	CH_4: 0.5	3.12	No
28	00:45	CH_4: 0.5 & 0.5 *	1.04	No
29	00:44	CH_4: 0.6	1.07	Yes
30	00:44	CH_4: 1	1.51	No
31	00:24	CH_4: 2	1.70	No
32	00:34	CH_4: 4	3.58	Yes
33	00:45	CH_4: 2	2.49	Yes

* Multiple source releases.

chambers E and F only captured one peak of ∼10 ppm and chamber H one large peak of 30 ppm. The mean wind speed during this release was of 3.12 m s^{-1} and the wind direction had little variations, ranging between 270° to 272°.

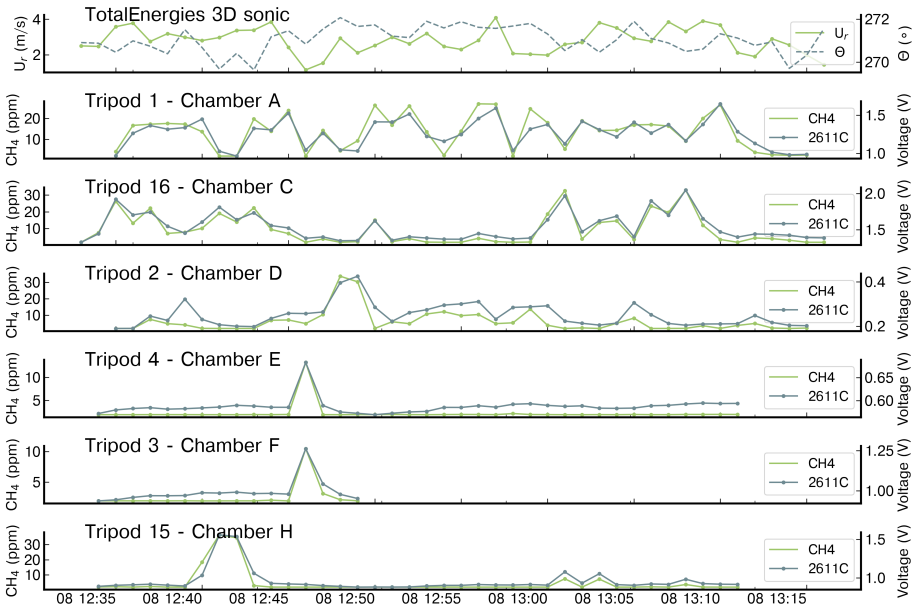


Figure 4.3: An example of 1-minute averaged CH_4 mole fraction (ppm) and voltage (V) measurements respectively measured by six high precision instruments and one type of TGS sensor (TGS 2611-C00) for release 25 (Q_s = 5 g s^{-1}). CH_4 measurements from the high precision instruments are denoted as 'CH4' and the voltage measurements from TGS sensor are denoted as '2611C'. The top panel shows the 1-minute averaged wind speed (*Ur*) and wind direction (θ) measured by the 3D sonic anemometer.

TGS sensors are known to be sensitive to variations of H_2O and T, affecting mainly the reconstruction of CH_4 baseline, and thus the characterization of peaks above this baseline (Rivera Martinez et al., 2021, 2022). Two approaches can be used to correct the effect of variable H_2O and T on the TGS signals baseline and separate the spikes from the baseline data in the time series. The first one is the use of information from H_2O and T to correct the TGS baseline signals correspond to these drivers. The second approach is to detect the voltage peaks associated to CH_4 spikes and derive a baseline with a linear interpolation on non-peak voltages. For some chambers due to logging system faults, we lost H_2O and T data and the corresponding gaps in the H_2O and T time series prevent us from defining a correction model. Therefore, in this study we have employed the second approach. To justify our choice, we have trained a multilinear regression model to determine a baseline signal on TGS 2611-C00 from Chamber E corresponding to

H_2O and T. The regression model was trained on using observations from midnight to 6:00 in the morning on the first day and we attempted to reconstruct baseline variations of TGS voltage from observations comprised between 18:00 to midnight on the same day. The results of the multilinear model are presented on Figure C.2 (in addition to the derived baseline when using the second approach). The second approach which produce a better detection of the baseline signal is also shown (see Fig C.2) where we do not need a training set or environmental variables because it consists in the detection of peaks based on an iterative process on fixed rolling windows and a comparison with neighboring observations.

To reconstruct CH_4 mole fraction, we calibrated empirical models that derive relationships between TGS voltage and other input variables and true CH_4 observed by the high-precision instruments. The models are calibrated (training) and evaluated (testing) using two independent subsets of the data. Following the widespread practice in the training of data-driven models to standardize the input variables to prevent difference in the range of magnitudes from conditioning the determination of model parameters, we applied a robust transformation consisting in removing the median and dividing the observations by their 1-99th quantile range. We selected the two reconstruction models that gave the best performances in our previous study (Rivera Martinez et al., 2022), namely a polynomial regression and a Multilayer Perceptron (MLP) model, described below.

Second-degree polynomials have proven to be robust to derive relationships between the TGS voltage signal related to spikes and the corresponding CH_4 concentration (Rivera Martinez et al., 2022). Its formulation is of the form:

$$\hat{y}_{CH_4}(x_1) = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 \tag{4.1}$$

Where $\hat{y}_{CH_4}$ is the predicted CH_4 concentration, x_1 is the Corrected voltage of the TGS after removing the effects of the baseline.

Artificial neural networks have been widely used to derive non-linear relationships between predictors and independent variables in many applications, as a universal approximator method (Hornik et al., 1989) and for their generalization capabilities (Haykin, 1998). In previous studies

(Casey et al., 2019; Rivera Martinez et al., 2021, 2022) ANN was employed to derive CH_4 concentrations from TGS observations on different sampling configurations (field and laboratory conditions) with good agreement between the reference observations and the outputs produced from the models.

The simplest architecture of an ANN is the multi-layer perceptron (MLP), conformed of a series of units (neurons) in fully connected layers. The inputs of any unit will be the weighted sum of the outputs of the previous layer, to which an activation function (ReLU, tanh, etc.) is applied. As a machine learning approach, it requires a training basis to learn the relationships, adjusting the weights of its connections, between the inputs and outputs using an iterative process known as optimization. Problems of MLP models are either underfit of data, producing a high error on the train set which can be mitigated with a sufficiently large network, or overfitting, producing a high test error when they cannot generalize to new examples. Regularizations terms and early stopping techniques are helpful to prevent overfitting (Bishop, 1995a; Goodfellow et al., 2016).

Here, we have trained the MLP model using the Adam optimizer (Kingma and Ba, 2014; Géron, 2019) resulting in 50, 10 and 5 units per layer with ReLU as the activation function for the hidden units. A regularization factor of α=0.05 and early stopping was used to prevent overfitting.Three configurations of the input variables were tested: i) only with the TGS 2611-C00, ii) only with the TGS 2611-E00, and iii) with both TGS sensors at the same time. The results are shown in section 4.2.3.2.

Metrics for evaluation of the reconstruction

To assess the performance of the models to provide dry CH_4 concentration enhancements (above the background) from voltage observations of the low-cost sensors we use a normalized root mean square error (NRMSE) per release, including information from the spikes and the background occurring in the duration of the release, defined in equation 4.2, the RMSE being weighted by the inverse of the maximum peak present in the release:

$$NRMSE = \frac{\sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}}}{h_{max}} \tag{4.2}$$

where y_i is the actual concentrations (provided by the high precision instrument),$\hat{y}_i$ the predicted concentration, n the number of observations present in the release and h_{max} the amplitude of the maximum peak present in the release after removing the background. The normalization allows to compare the performances across the different releases.

As mentioned earlier in section 4.2.2.2, the target signal on this study is of CH_4 enhancements above the atmospheric background. We obtain this signal by subtracting the raw signal of the release from an inferred baseline computed using the pic detection algorithm and a linear interpolation. We consider as an acceptable notional target error for the reconstruction models to be under the 15% of the amplitude of the maximum peak inside the release, this error corresponds to a NRMSE $\leq$ 0.15 ppm.

Selection of the training and test subsets for the reconstruction of CH_4 mole fractions as input of the atmospheric inversion of emissions

Defining the appropriate training set is important to allow reconstruction models to derive sufficient information to generalize and obtain good performances in the test set. As well, the test set should be chosen to allow evaluating the performances of the models under a wide variety of conditions. Regarding the inverse modelling, in order to provide a meaningful assessment of the estimation of emission rates and locations, inversion should be conducted using reconstructed CH_4 mole fractions that are not from the training data set to avoid introducing bias in the evaluations of errors. Furthermore, depending on magnitude of release rates, the atmospheric turbulence, and the locations/distances of the downwind active tripods from the emission sources, the six chambers did not measure CH_4 mole fractions in all the releases, therefore a separate training and test set needs to be defined for each chamber.

The previous considerations constrain the selection of the training and test sets from the data of each chamber. The test set of the releases for inversions was defined based on two criteria:

1) the releases which have the reconstructed CH_4 mole fractions by at least three chambers simultaneously, and 2) the releases corresponding to the more favorable wind speed conditions ($Ur \geq 1.4$ m s^{-1}) for inversions. We determined seven releases that meet these considerations: release #2, #9, #10, #25, #26, #32 and #33. Because this test set was not sufficiently large for all the chambers, we decided to increase it by data from four more releases with low wind speed conditions ($0.65 \leq Ur \leq 1.31$) (release #5, #12, #21 and #29). This selection led to a test set of 40% of the releases. All remaining data were used as a training set (Table 4.4). The reconstruction models are trained and tested only once per chamber following the distribution of the releases from table 4.4.

Table 4.4: Summary of the releases included in the training set and test set of the CH_4 reconstruction models. The mixing ratios modelled for the test set are also used as input of the inversion model to infer the emission rate of CH_4 and their location.

Chamber	Releases in the training set	Releases in the test set	Number of releases in the training set	Number of releases in the test set	Percentage of releses in the training/test set
A	6, 7, 8, 11, 14, 15, 16, 17, 18, 19, 20, 24, 27, 28, 30	2, 5, 9, 10, 21, 25, 26, 29, 32	15	9	62.5 % / 37.5 %
C	14, 15, 17, 18, 19, 20, 22, 24, 27, 28, 30, 31	9, 10, 21, 25, 26, 29, 32, 33	12	8	60 % / 40 %
D	6, 7, 8, 13, 14	5, 9, 12, 25	5	4	55.5 % / 44.5 %
E	3, 4, 6, 7, 8, 13, 14, 19, 20, 22, 23	2, 5, 9, 12, 21, 25, 26, 32, 33	11	9	55.5 % / 44.5 %
F	3, 4, 6, 7, 8, 13, 14, 15, 18, 19, 20, 22, 24	2, 5, 9, 10, 12, 21, 25, 29	13	8	62 % / 38 %
H	1, 3, 4, 13, 14, 18, 19, 20, 23, 24, 28, 30	2, 21, 25, 26, 29, 32, 33	12	7	63 % / 37 %

4.2.2.3 Atmospheric inversion of the release locations and emission rates

Our derivation of the release location and rates relies on the inversion framework developed and tested by Kumar et al. (2022) on the measurements of the high precision instruments. This framework uses adjoint of a gaussian plume model to simulate the sensitivity of the CH_4 mole fraction enhancements above the background at a measurement location to the emissions at all potential source locations. For each release, the optimal horizontal and vertical location and rate are derived based on the minimization of the root sum square (RSS) misfits between averages of the observed and simulated CH_4 mole fraction enhancements above the background. The bins

of the measurements and of the simulated mole fractions for the averages correspond to sectors of wind directions of equal ranges during the release. The optimal release location and rates are searched simultaneously, looping on a finite but large ensemble of potential locations, using an analytical formulation of the problem to derive the optimal rate and corresponding RSS misfits for each potential location and then identifying the optimal location and rate providing the smallest RSS misfits. The 40 m × 50 m (horizontally) × 8 m (vertically) volume above the ATEX zone is discretized with a high resolution (1 m × 1 m horizontally and 0.5 m vertically) 3D grid to define the finite ensemble of potential locations. The inversion exploits the change of wind direction during a release and the corresponding variations and spatial gradients in average mole fractions respectively at and between the different measurement locations crossed by the plumes to triangulate the release location. The amplitude of the enhancements directly constrains the release rate estimate.

The Gaussian model and its adjoint are driven by averaged wind directions and averaged turbulence parameters derived from 3D sonic measurements, using the same bins for these averages as for the mole fractions. Those bins are defined during each release based on 1-min averaged wind directions. These bins partition the lower and upper range of potential wind directions, and they have equal width in terms of range of wind directions. The total number of bins during this initial partition is defined as the rounding integer of the division of the release duration (in min) by approximately 7 min. However, only bins gathering at least four 1-min averages are retained. The aim is that the mole fraction and meteorological averages are representative of a timescale that is long enough for use in or comparison to the Gaussian model. Depending on the releases, the number of bins ranges between 2 and 7.

Here, we slightly revise the reference computations of release location and rate estimates based on the high precision instruments from Kumar et al. (2022). Indeed, in order to compare the release location and rate estimates from such a reference and the one derived here based on the TGS sensors, we restrain the set of high precision observations that are used in the reference computation to the station and time corresponding to the data availability from the TGS sensors.

4.2.3 Results

4.2.3.1 Pre-processing of the data from the low-cost CH_4 sensors

Original observations with a time step of 2 s were resampled to 5 s. We corrected the time offset corresponding to delays of the air travel through the air intake from the tripods to the instruments, time delay from synchronization between analyzers and chambers. Also, we removed invalid data produced by the logging system on each chamber. The baseline correction was then applied for each sensor chamber considering the entire campaign. As an illustration of the impact of the baseline correction Figure 4.4 shows the signal corresponding to one release for the chamber A after these pre-processing steps. The corrected signal in the TGS voltage measurements showed better agreement with the reference between the occurrences of spikes and phases.

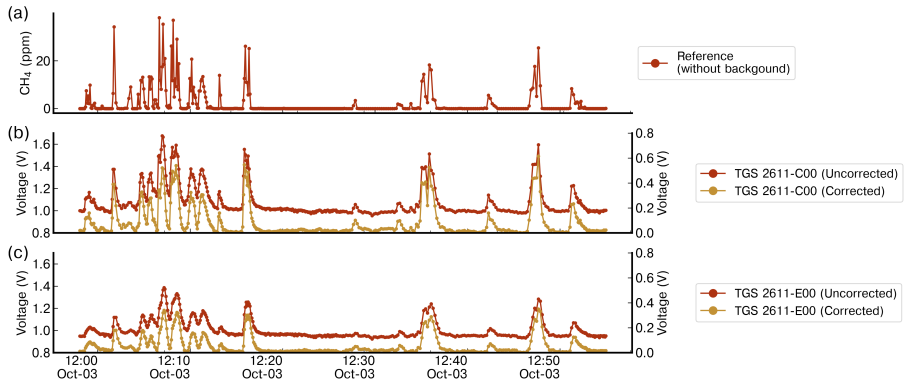


Figure 4.4: Comparison of the voltage signal for one release (#8) from Chamber A before (Uncorrected) and after (Corrected) the baseline correction on (b) TGS 2611-C00 and (c) TGS 2611-E00, on which it is appreciated the correction of the offset preserving the amplitude enhancements linked to CH_4 variations. (a) Reference CH_4 mole fractions, also corrected using the spike correction algorithm.

4.2.3.2 Reconstruction of CH_4 mole fractions

Due to the diversity of the releases, environmental conditions, distribution of the tripods and selection of the training and test sets for each chamber, there is no single release that can be viewed as representative for the test set across the chambers. Yet, we chose release #25 as an example of the signal measured across the chambers and the reconstructed signal for each chamber using the MLP model (Figure 4.5) and the 2nd order polynomial model (Figure 4.6), for

each chamber we shown the reconstructed CH_4 mole fractions estimated using only the type C sensors (red), the type E sensor (yellow) and both sensors used as inputs for the models at the same time (green).

We found that the MLP and 2nd degree polynomials gave similar performances across the releases regardless of the chamber used for the CH_4 reconstruction. For two releases on chamber A (release #10 and #26, see Fig C.6 and C.9 for MLP model and C.11 and C.12 for the polynomial model respectively) where amplitudes are below 10 ppm, the polynomial model provides a noisy signal as output regardless the configuration of the inputs used. There were however some cases on which the polynomial model produced better outputs than the MLP, for example the four releases on chamber D where MLP model produced a systematic underestimation of the reconstructed CH_4 on the three configurations of inputs.

Regarding the TGS types, the type C sensor gave better reconstructions than the type E or both types used as the same time as inputs for the model. The reconstruction of CH_4 with the type E sensor shows phasing errors in the form of a slow decay after large spikes. In addition, there are some cases where type E sensors showed a response whereas no spikes were measured by the reference instrument. For example, release #9 (Figure C.5 and C.10, for the MLP and the polynomial model respectively) of chamber D shows few spikes between 10 to 30 ppm reconstructed from data of the type E sensor with the polynomial model which are not present on the reconstructed data from the type C sensor. Using Type C and E sensors at the same time as training data for models produced outputs closer to models trained only with type C sensor. Some cases of reconstruction with MLP model produced a saturation of the outputs (release #9, #12 and #25 for chamber D (Figure C.5, C.7 and 4.5), release #21 for chamber H (Figure C.8)) or a systematic bias (releases #2, #10 and #26, see Figure C.4, C.6 and C.9). For releases with peaks' amplitudes above 40 ppm a systematic underestimation is observed regardless the model or the sensor's type used as input.

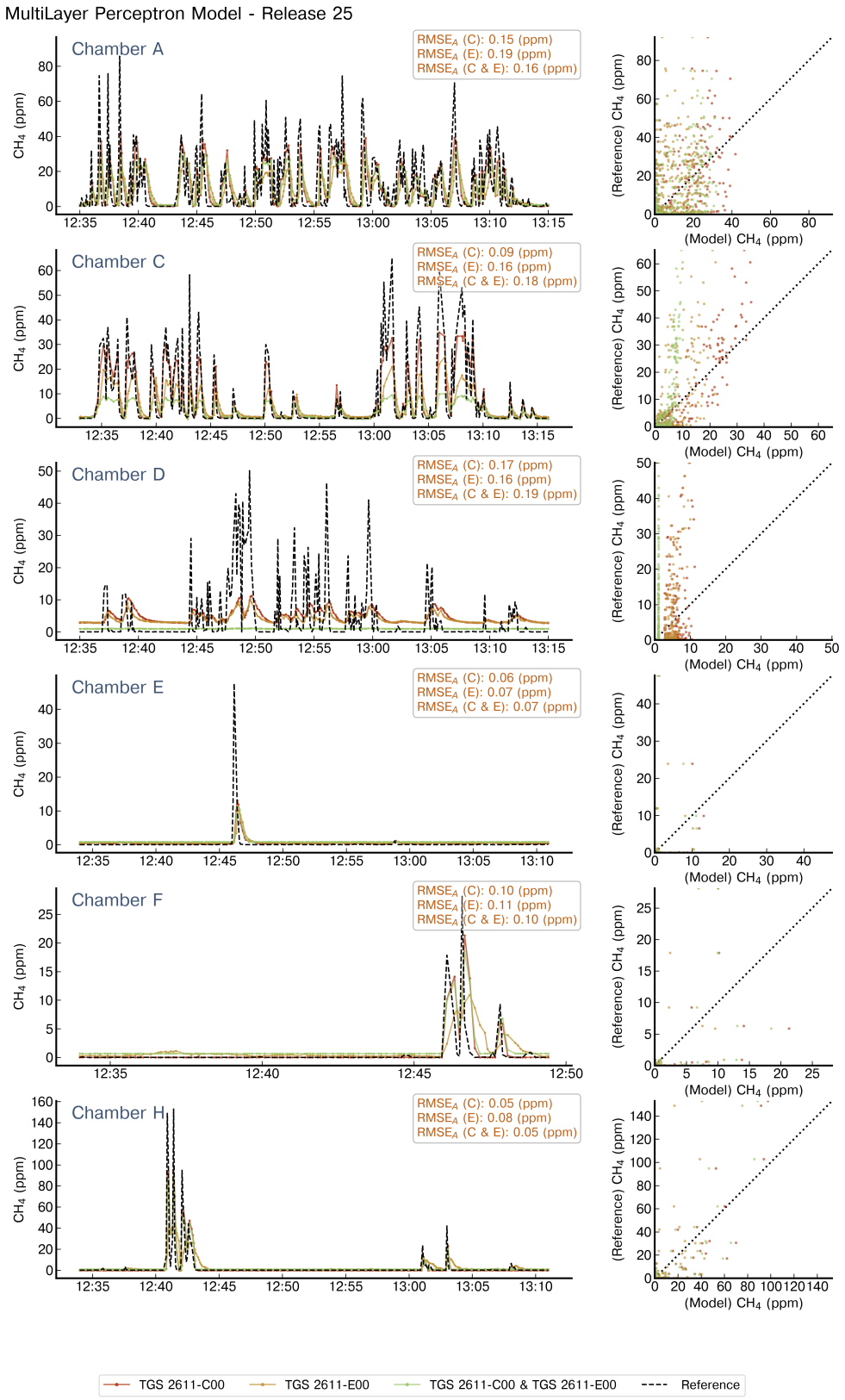


Figure 4.5: Example of reconstruction of release #25 using an MLP model. On left panels are shown the reconstructed CH_4 mole fractions for each chamber that captured the release, we present the reference signal (black dotted line), the reconstructed CH_4 mole fractions when the model has as input the TGS 2611-C00 sensor (red), the TGS 2611-E00 (yellow) or both types at the same time (green). The right panels show the 1:1 plot of the reference against the output of the model for the three configurations of inputs. Note the difference in the x-axis for Chamber F.

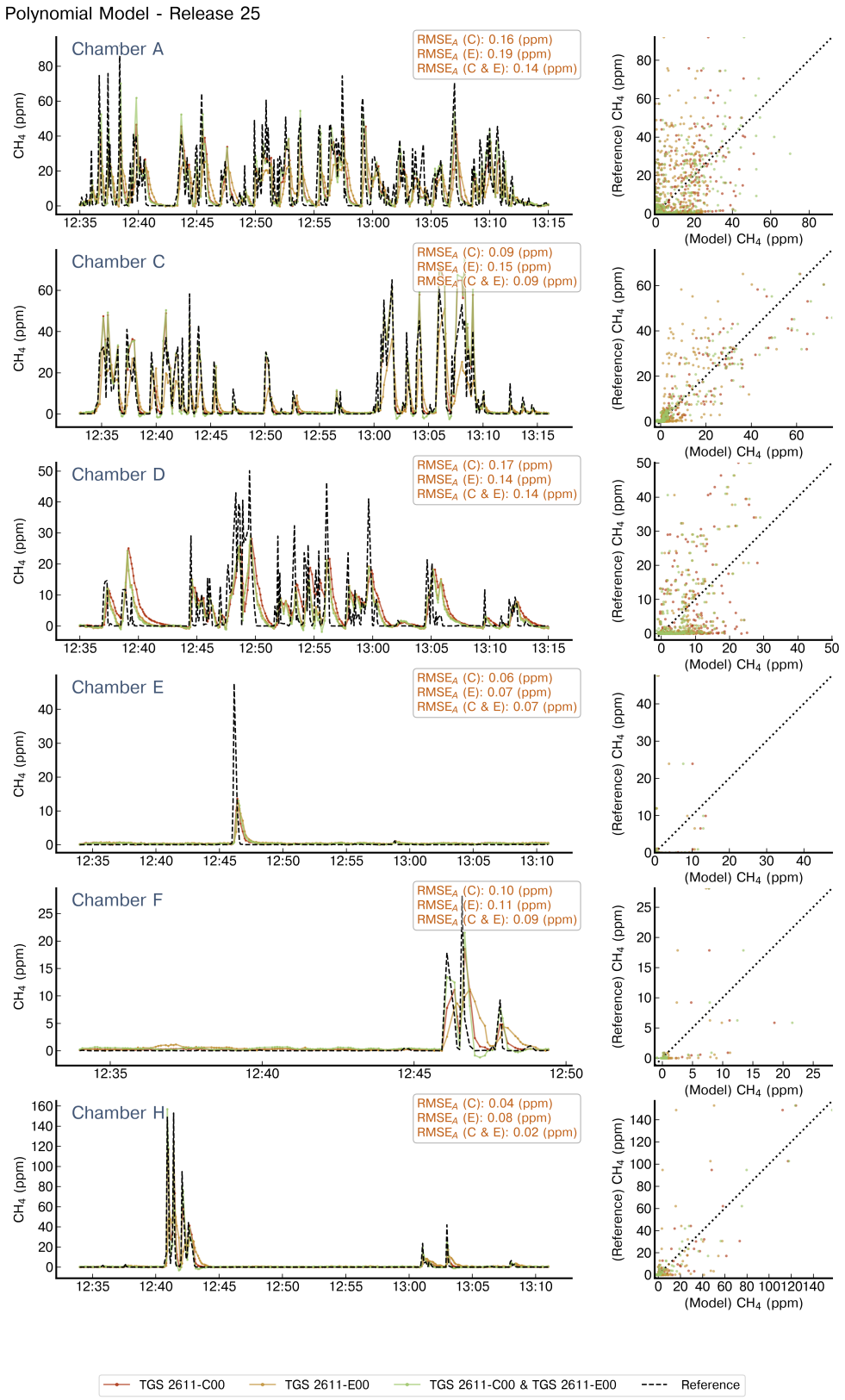


Figure 4.6: Example of reconstruction of release #25 using a Polynomial model. Notations are the same as in Figure 4.5.

On figure 4.7, we present a summary of the performance of the reconstruction of the signal on the test set, given the NRMSE error defined in eq. (3). All chambers have reached our target error of NRMSE $\leq$ 0.15 ppm, except for Chamber A with the polynomial model using as input the type E sensor and the MLP model for chambers A and C as well for the type E sensor. With a stricter target requirement of NRMSE $\leq$ 0.1 ppm, only Chamber H met the target error regardless of the model or sensor used. Performances are similar when using the type C sensor as input regardless the model across all the chambers. When used both types at the same time as input, the 2nd degree polynomial provide better reconstruction than the MLP specially on chambers C, D and H (NRMSE = 0.09, 0.09 and 0.04 ppm for the polynomial model and 0.11, 0.13 and 0.07 ppm for the MLP). Chamber D, where there is little training data available, produced a systematic lower error with the polynomial model than with the MLP regardless the input variable used.

In summary, the model used in the reconstruction is important only for the cases where little information is available for the training. This was the case for chamber D where the polynomial model provides better performances than the MLP model. We also found that Type C sensors produced a better reconstruction of CH_4 spikes than Type E ones, and a combination of data from both types of sensors did not improve the reconstruction producing similar outputs than the other types.

4.2.3.3 Release rate and location estimates based on the observations from the TGS sensors

Averages of mole fractions enhancements above the background and their spatial gradients are displayed for release #25 in figure 4.8. The figure compares the values of reconstructions from the low-cost sensors (with the MLP model; see figure **??** for the values corresponding to the polynomial model), with the high precision measurements, and of the simulations resulting from the inversions assimilating either the reference high precision data or the reconstructions from the low-cost sensors. Since the best reconstruction performances were obtained when using the type C sensor, the inversion results presented here are based on the reconstructions from those sensors only. For the release #25, used as an example here, the procedure to define average

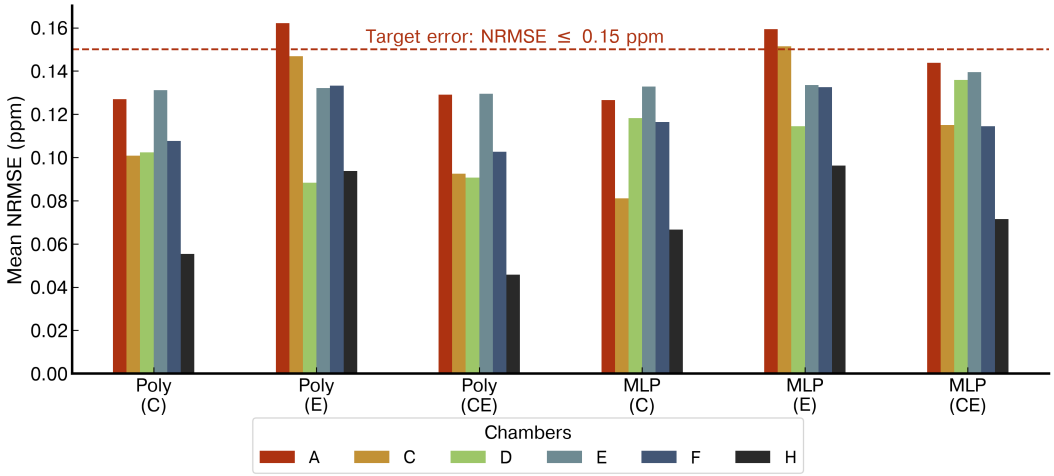


Figure 4.7: Comparison of the mean NRMSE of the two types of models trained with the three configurations of the inputs. The 2nd degree polynomials are denoted as 'Poly' and the multilayer perceptron as 'MLP'. The three input configurations are denoted inside parentheses, 'C' when the model's input was only the TGS 2611-C00, 'E' for the TGS 2611-E00 and 'CE' when both sensors were used as inputs at the same time. The color code of the bars corresponds to the chambers.

values per wind sectors has resulted in four bins of wind sectors with an approximate size of 10°. Average mole fractions are derived from the six chambers. To simplify the numbering when mentioning the reference instrument or the TGS, we refer to the chamber identifier X (REF-X and TGS-X respectively, with X the name of the chamber).

In general, the observed spatial CH_4 gradients between the different stations are similar when considering the reference measurements and the estimates of the TGS, except for few cases where the reference is more consistent to the expected signal. For example for release 25 (see Fig. 4.8) observed gradient from TGS-D data underestimate the actual gradients given by REF-D for θ = 308.3° and overestimate them for θ = 279.2°, where θ is the average direction of the wind sector.

The modelled average mole fractions enhancements and thus the modelled gradients assimilating reference data are very close, in general, to the ones from these reference data, although some discrepancies can occur, e.g., for release #25, for REF-H with θ = 279.2°, REF-C with θ =301.4° and θ =289.1° and REF-A with θ =301.4° and 308.3°. For most of the cases, the modelled gradients assimilating the TGS data are closer to the modelled gradients assimilating the reference data than to the observed TGS ones. In addition, the observed TGS data, for some cases, is closer to

the observed reference one than to the modelled gradients assimilating either reference or TGS data, highlighting the higher impact of the model error on the inversion than the reconstruction error of CH_4 mole fractions.

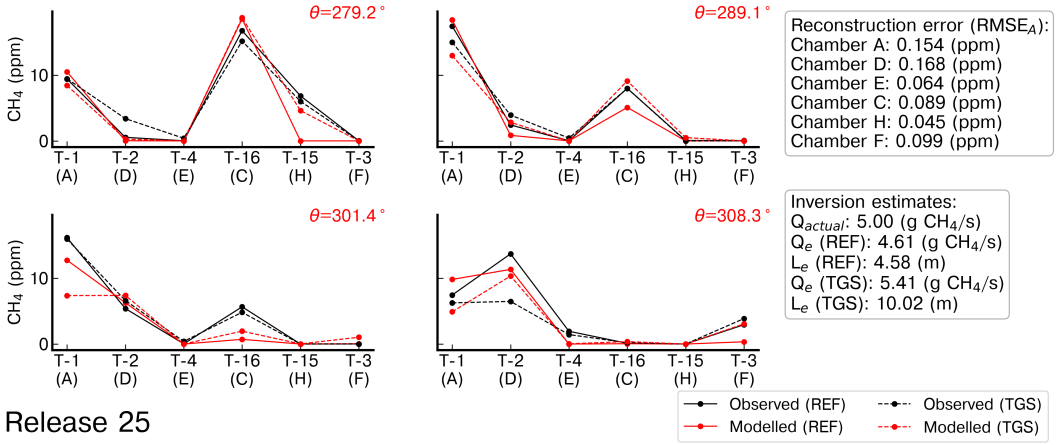


Figure 4.8: Observed and modelled average CH_4 mole fractions from the reference, denoted 'REF', and low-cost sensor, denoted TGS, corresponding to the release #25. The reconstructed CH_4 was computed using the MLP model. The index of the tripods is denoted as T-x and the average wind direction (θ) for the binning of wind sectors is shown on the top right of each panel in red.

Figure 4.9 shows the comparison of the emission rate estimates with corresponding errors, and of the location errors for the different inversions across the eleven releases. In this figure, estimates assimilating CH_4 mole fractions from the TGS using the reconstruction with the MLP models (see Figure C.14 for the results when assimilating the reconstruction based on the 2nd degree polynomial model).

Regarding the release rate estimates, those from inversions assimilating the reference mole fractions bear an average error of 30% and those from the inversion assimilating data from the TGS sensors bear an average error of 25%.

In the case of the estimation of the release location, the assimilation of the reference data produces a slightly smaller average error location of 7.86 m (σ = 5.47 m) compared to 9.49 m (σ = 4.58 m) from the assimilation of TGS data. For five releases (#2, #10, #12, #25 and #26) the assimilation of reference data yields a better estimate of the location and for one release (#21) both inversions yield similar location errors.

In general, estimates of the emission rate (see fig 4.9a) from reference data and TGS data are

similar. For three releases (#12, #25 and #32), we observe large errors in the estimate of the release rate. Inversion assimilating TGS data or reference data highly underestimate the rate for release #5 (1.41 and 1.34 g CH_4 s^{-1} respectively, with an actual emission rate of 3.0 g CH_4 s^{-1}) and strongly overestimate the rate for release #32 (5.14 and 6.55 g CH_4 s^{-1} respectively, with an actual emission rate of 4.0 g CH_4 s^{-1}). Reference data provide a slightly better estimation of the location of releases than the TGS. Only for releases #29 and #33, the inversion assimilating TGS observations provide a slightly better location of the source. Conversely, for releases #2, #12, #25 and #26, the location error from the inversion assimilating TGS observations is almost double than the one of the reference. The errors on the emission rate estimate from both inversions was smaller than 30% for most of the releases, except on four cases, where errors reached 80% for the inversion assimilating TGS data and 65% for the inversion assimilating reference data, respectively. There were two cases, the release #26 and #33, when the inversion assimilating TGS observations produced a much lower error (2.5% and 5.3% respectively) in the quantification of the emission rate than the inversion assimilating reference observations (20.9% and 22.7% respectively). The fact that the assimilation of the TGS reconstructed CH_4 data can yield better results than when using accurate CH_4 mole fractions measured by the reference instrument highlights the impact of the transport model error (associated to the simulation of the average mole fractions with the Gaussian model) in the inversion process. These errors dominate the resulting errors in the estimates of the release rate and location when assimilating the reference data Kumar et al. (2022). They appear to have a weight larger than that of the errors in the reconstructed mole fraction from TGS data when assimilating these data.

4.2.4 Discussion

Our study showed the capability of the signal from metal oxide sensors to produce estimates of the emission rate and location from controlled CH_4 releases typical of those expected from leaks in industrial facilities. The used baseline correction algorithm allows to extract the variations of voltages from the TGS signal related to the high frequency variation of the plume across the different sensors' inlets. We compared the performances of two models, 2nd degree polynomials and

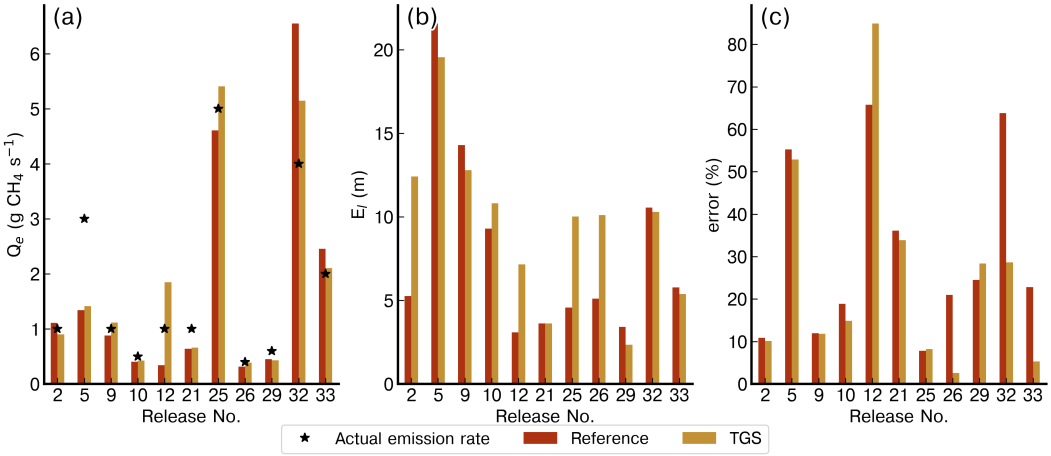


Figure 4.9: Comparison of the emission rate estimate (*Qe*) (a), of the location error (*El*) (b) and of the relative error in the emission rate estimate (c) from the inversions assimilating the Reference data (in red) and the reconstruction of the CH_4 mole fraction from the TGS sensors (in orange). The reconstructed CH_4 mole fractions used in these inversions are computed with the MLP model

MLP, to reconstruct CH_4 mole fractions during the controlled releases for three configurations of inputs. The reconstructed CH_4 mole fractions were used as input to an inversion modelling framework to estimate the emission rate and location for each release. Results of inversions assimilating TGS data were compared with those assimilating reference (CRDS) data.

The correction of baseline in TGS sensors assumes that the targeted signal measured by the sensors corresponds to a series of spikes at high frequency produced by the plume reaching and leaving the inlet tube of the sensors, due to the atmospheric turbulence and high frequency variations of the wind. Our approach of deriving a baseline signal from observations surrounding the spikes in an iterative process, offers a suitable alternative to correct the TGS observations when little or insufficient information is available to derive a baseline correction model (e.g. from observations of H_2O and temperature). This approach is interesting for conditions when the environmental parameters are highly variable or models does not dispose of sufficient observations to derive robust relationships to correct the effects of environmental variables on the sensors' baseline signal. This corresponds well with the measurements presented in this study. However, in some cases, the plume can touch the inlet tube of the sensors during a prolonged period producing a signal not only having high frequency spikes but also continuous varying enhancements above the background. For those cases, this method would not be able to distinguish the

Table 4.5: Comparison of the emission rate estimates (*Qe*), location error (*El*) and relative error on the rate estimates for the inversions assimilating the reference data and the reconstruction of the CH_4 from the TGS low-cost sensor based on the MLP model

Release	**Actual emission**	**Reference**			**TGS**		
N°	$(\mathrm{g\ CH_4\ s^{-1}})$	$\mathrm{Q_e\ (g\ CH_4\ s^{-1})}$	$\mathrm{E_l\ (m)}$	**error (%)**	$\mathrm{Q_e\ (g\ CH_4\ s^{-1})}$	$\mathrm{E_l\ (m)}$	**error (%)**
2	1.0	1.10	5.26	10.8	0.89	12.40	10.1
5	3.0	1.34	21.57	55.2	1.41	19.55	52.8
9	1.0	0.88	14.29	11.9	1.11	12.78	11.8
10	0.5	0.40	9.29	18.9	0.42	10.80	14.8
12	1.0	0.34	3.08	65.7	1.84	7.15	84.9
21	1.0	0.63	3.61	36.1	0.66	3.61	33.8
25	5.0	4.61	4.57	7.8	5.41	10.02	8.2
26	0.4	0.31	5.10	20.9	0.39	10.10	2.5
29	0.6	0.45	3.40	24.5	0.43	2.34	28.3
32	4.0	6.55	10.55	63.8	5.14	10.28	28.6
33	2.0	2.45	5.77	22.7	2.10	5.37	5.3
Average error			7.86	30.7		9.49	25.5
σ_{error}			5.47	20.3		4.58	23.6

enhancements on sensors' voltage corresponding to the CH_4 plume from the background and then we need to reconsider the derivation of a baseline based on environmental parameters (H_2O and T). Regarding the type of sensor used in the reconstruction of CH_4 mole fractions, we obtain best performances using only with Type C sensor as input for the models. The fast decay observed on the reconstructed CH_4 after the spikes was attributed to the response time of the TGS sensor. The slow decay observed on Type E sensors was probably due to a combination of the response time and the carbon filter added on top of the sensitive material to improve the selectivity of gases. Concerning the reconstruction models, the polynomial and the MLP models in general produced equivalent results with few differences. It confirms our previous study (Rivera Martinez et al., 2022) in which we observed that the performances of models to reconstruct CH_4 mole fractions were mainly driven by the type of used sensor, rather than from the model chosen for the reconstruction. With a low content of information (only few spikes, limited range and variability of the spike magnitude, frequency and duration) in the training set (e.g. when reconstructing the CH_4 mole fractions of chamber D), the 2nd degree polynomial provides more accurate estimates than the MLP. This is probably due to the distribution of the data in the training set that MLP used to compute its parameters, which does not represent the same range of variations than the one in the test set. For spikes with enhancements under 5 ppm, the MLP

model with the Type C sensor signal as input, produced a more accurate reconstruction than the Type E or both sensor's types when used as inputs at the same time. The noise present in the voltage signal on some releases, for example release #26 on chamber A, were not correctly removed in the reconstruction with the Polynomial model. However for the type C sensor, the MLP model reduces the noise on the signal producing a more accurate reconstruction.

Regarding the inversion of emission rates and locations using the gaussian plume model framework developed by Kumar et al. (2022), we obtained good estimates and performances with the reconstructed timeseries of CH_4 spikes from voltage measurements of TGS sensors and the results are comparable to those obtained when assimilating the reference data. We observed that the simulated gradients of the gaussian model assimilating observation from the TGS chambers were close to the simulated gradients of the reference inversions (assimilating high precision measurements), even if the observed gradients were sometimes in a different direction. In most cases errors from both inversions ranges between 2.5% and 55% except for release #12 and #32 where error reached 65% and 63% for simulated gradients assimilating the reference data respectively and release #12 with an error of 85% for simulated gradients assimilating the TGS data. The overall inversion performance assimilating TGS data and reference data are good and consistent. The slightly better average performance in the release rate estimates using TGS data (25% error) than the estimates using reference data (30% error) is not significant in regards of the variability of the results. It highlights the weight of the model errors associated to the simulation of the average mole fraction with a Gaussian model. The results demonstrate that the errors in the release rate and location estimations from inversions using both reference and TGS data are dominated by these model errors. The errors in the reconstruction of the CH_4 spikes from the TGS data are thus sufficiently low for use in the inverse modelling problem analyzed here. One should note that, as mentioned on section 4.2.2.3, in this study, the reference inversions rely on a restrained subset of the reference data that match the available data from TGS sensors. Results from Kumar et al. (2022) considering the full dataset available on the reference instruments obtained highly better results.

4.2.5 Conclusions

This study presents different techniques to reconstruct CH_4 mole fractions from the voltage signal measured by metal oxide low cost TGS sensors deployed downwind an area of points of controlled releases during a campaign at TADI in 2019. The data from this reconstruction are assimilated in an inverse modelling framework to quantify the rate (ranging from 0.4 to 5 g CH_4 s^{-1}) and location of these controlled release. The approach employed to extract the baseline signal on TGS voltage measurements based on surrounding observations allowed us to derive and successfully correct the baseline signal on TGS sensors without the need of using other environmental parameters. The reconstruction of CH_4 mole fraction from voltage observations measured during controlled releases showed good agreement with observed CH_4 mole fractions from the reference instruments. The reconstruction was consistently better with TGS 2611-C00 sensor regardless the reconstruction model used. Both models had met our requirement target of NRMSE of reconstructed CH_4 lower than 0.15 ppm across all chambers when trained with the TGS 2611-C00 sensor. Emission rate and source location estimates using an inversion based on a gaussian plume model produced similar results using reconstructed CH_4 mole fraction from TGS sensors data to those obtained with high precision instruments, with an average estimate rate error of 25.5% and a mean source location error of 9.5 m from TGS data. In this study, the reconstruction of the CH_4 mole fractions was conducted independently from the inversion modelling. The estimation error could probably be reduced with a better understanding of inverse modelling sensitivity to the misfits from the reconstruction models. In consequence, a sensitivity study is encouraged to determine the best approach for the reconstruction of the observations from TGS sensors.

Chapter 5

Conclusions and perspectives

5.1 Conclusions

This PhD thesis aimed to determine a calibration strategy for CH_4 using MOS low-cost sensors in the laboratory, specifically TGS from the manufacturer Figaro®, as well as, to provide estimates of source rate and location from reconstructed CH_4 variations based on measurements of these sensors in the field. Controlled laboratory experiments presented in Chapters 2 and 3 allowed to test the capability of TGS sensors to measure near background variations and CH_4 enhancements of several ppm over background. The controlled release experiment of Lacq allowed to evaluate the source rate and location estimates from reconstructed signal measured by TGS and to compare with estimates from high precision instruments on similar conditions.

5.1.1 Reconstruction of room air variations of CH_4 and sensitivities to CO and H_2O

The first study of this thesis was dedicated to develop and apply a machine learning model to reconstruct CH_4 variations in room air under controlled conditions and to determine the sensitivities of three types of TGS to CO and H_2O. Our assessment of sensitivities to CO and H_2O showed a nonlinear relationship of Figaro® TGS. A multilinear quadratic function explained the contribution of CO variations to voltage measured by the sensors. The H_2O sensitivity test revealed the high influence of H_2O in the sensors in presence of high H_2O mole fractions reducing

its sensitivity to CH_4. These results are in line with previous works (Ionescu et al., 1999; Wang et al., 2010; Tischner et al., 2008) and they explained the competition of CH_4 and H_2O molecules in reacting with oxygen, which affects the sensitivity of the sensor. TGS 2611-E00 sensor, which included a carbon filter to improve selectivity, provided a slight nonlinear response to H_2O variations, although sensitivity to CH_4 was equally reduced across the different H_2O levels. On the other hand, TGS 2600 and 2611-C00 types presented a pronounced nonlinear relationship where CH_4 sensitivity was higher for lower H_2O levels.

The sensitivity test derived from the modelled CH_4 and the training removing one predictor at a time, confirmed the importance of H_2O as a predictor. In the case of CO, our models found a lower influence in the reconstruction of CH_4 levels. This can be explained by the controlled conditions of the environment where concentrations are lower that on industrial environments, and thus it does not reflect a real case scenario. The inclusions of different types of sensors as predictors to the models reduced the MLP performances, contrary to what was expected. This behavior can be explained by the introduction of inconsistent information from other sensor types due to different sensitivities and response time, especially for TGS 2611-E00 type.

Regarding the reconstruction of CH_4 concentrations from TGS resistance variations, the MLP model provided good agreement between the simulated signal and the reference, meeting our target requirement of RMSE $\leq$ 0.2 ppm for nearly all the cross-validation periods. The length and selection of the training period was a key element to ensure good performances from the MLP model, and only similar distribution of the training and test periods led to better reconstruction of CH_4 concentrations. A low-pass filter behavior of the MLP was also observed through our test where the approximation of low frequencies is favorized in the learning process of the model. Clearly, a very long training period and the need to choose representative variations to ensure a good accuracy would be a strong limit for a user who would like to “install and forget” sensors in the field and let them record CH_4 variations.

5.1.2 Baseline correction and reconstruction of CH_4 enhancements (spikes) over background variations

In the second study we evaluated the performance of two TGS types to reconstruct CH_4 spikes generated over background variations on laboratory conditions, to mimic typical variations of CH_4 in the vicinity of industrial leaks. An automatic artificial spike generation system was developed in order to produce such 'realistic' CH_4 enhancements over background. We have developed a two-step baseline correction algorithm that corrected the cross-sensitivities of H_2O to environmental parameters on the TGS voltage measurements.

The first test consisted in a systematic comparison of parametric and non-parametric models to reconstruct the CH_4 spikes generated in laboratory. We obtained the best performances from the 2nd degree polynomial regressor and the MLP model, meeting our error requirement of an RMSE $\leq$ 2 ppm. These performances were obtained with a large training set consisting of 70% of the observations in the dataset, which is not very appealing for a user who aims to minimize the training length.

In the second test, we evaluated a stratified training strategy aiming to keep the performance of the models with a reduced training set. We define a new training set composed of a selection of the most influential spikes previously regrouped into clusters based on its similarity. The proposed parsimonious training allowed to reduce the length of the training set size from 70% to 25% of the data keeping our target error requirement. This is clearly more encouraging for any future 'operational' applications of the TGS sensors.

Ageing effects of TGS were studied in the third test. We observed a slight decrease in the performances after six months of continuous functioning of the sensors, although the degraded performances when we applied a model that was trained six months before still met our target error requirement.

Finally, the capability of transfering a trained model per sensor and per chamber into other sensors of the same type from other chambers was assessed. We observed that the use of generalized models is possible when the target signal, CH_4 spikes in our case, varies in the same

range that the one used during the training of the model. In the same line of the previous assessment of TGS in chapter 2, the TGS 2611-C00 type showed better performance in detecting and quantifying CH_4 variations than the TGS 2611-E00 one.

5.1.3 Reconstruction of CH_4 spikes during controlled releases and source rate and location estimation

The third study of this thesis consisted in the reconstruction of CH_4 concentrations and the estimation of the release rate and location from voltage measurements of TGS sensors during the TADI 2019 campaign. This campaign consisted of a series of 41 controlled gas releases, from which we focused on 33. We deployed 16 sampling lines in the field, from which only 7 were active for each release, around a 40x50 m ATEX zone where releases were emitted. Each active sampling line was connected to a high precision instrument and a chamber containing three types of TGS. This study focused on two types of TGS (TGS 2611-C00 and TGS 2611-E00) due to low sensitivity of CH_4 enhancements on TGS 2600.

We assessed the performance to reconstruct CH_4 concentrations of a parametric and a non-parametric model on 11 controlled releases measured by six chambers including two types of TGS. Those releases were used later to estimate the source emission rate and location, and for this we used an atmospheric inversion model. Then, they were compared to estimations from high precision instruments and the actual rates and location. Although I did not develop myself the inversion model, I adapted it, tuned it and ran it to invert emissions rates and their locations.

The first part of the study was dedicated to the preprocessing of the signal and the reconstruction of CH_4 concentrations. We employed the automatic spike detection algorithm of Chapter 3 to derive and correct the baseline signal on voltage measurements from TGS sensors, that are highly influenced by environmental parameters, such as, H_2O and T. Concerning the reconstruction models, similar performances are seen in general for the polynomial and MLP model, although in configurations with a reduced training set size, the 2nd degree polynomials produced slightly better performances. The configuration of the inputs of the reconstruction models showed to be of much greater importance than the model itself. TGS 2611-C00 produced the lowest errors

in general across the chambers where the training set was sufficiently large. The combination of both types as inputs increased the error for most of the chambers and on few cases it produced an output similar to TGS 2611-C00. Only one of the assembled chambers (H) meet our requirement error target. Nevertheless the distribution of peaks inside the releases measured by this chamber could have diluted the summary statistics.

The results obtained from the inverse modelling showed good agreement between TGS and the inversion using the same dispersion model with CRDS data and obtaining an average emission rate error of 25% and 30% respectively. Regarding the estimation of the source location, the TGS performed worse than the reference inversion with CRDS data, with an average location error of 9.5 m compared to 7.8 m obtained from the reference inversion. Simulated CH_4 gradients from the TGS data were closer to the observed and simulated gradients of the reference even if observed gradients from TGS data were on some cases in opposite direction. The way errors from the observed/reconstructed TGS data are considered from the inverse modelling is still unknown and requires further analysis. Overall, this result led us to be optimistic about the use of TGS sensors for monitoring leaks in the field, but CRDS data will be still needed for training (at least at the beginning) and for ensuring robust inversion results.

5.2 Perspectives

This PhD thesis aimed to improve the knowledge of MOS low-cost sensors in the detection and quantification of CH_4 emissions. It remains an open research area to explore the capabilities and limitations of these sensors. Here we present some perspectives for future research involving MOS sensors.

5.2.1 Cross sensitivity to other species and environmental parameters

When we conducted the sensitivity analysis under laboratory conditions in the first experiment, we found nonlinear relationships between CO, H_2O and the resistance measurements from TGS sensors. Previous studies (Tischner et al., 2008; Kolmakov et al., 2005) showed that relative hu-

midity and temperature affects directly the sensitivity of MOS sensors. It is necessary to conduct a more thorough assessment of the influence of environmental parameters on the sensitivity of TGS, especially for temperature and H_2O.

The sensitivity analysis on the MLP model trained with room air variations showed that the contribution of CO was not of importance to derive a calibration model on TGS. This is counterintuitive since CO is also an electron donor gas and should compete with CH_4 affecting its sensitivity. More tests should be conducted under real conditions to determine the real contribution of CO to the calibration model.

The idea of adding metallic particles as a membrane, known as catalyst, over the sensing material of MOS sensors acting as a filter to other compounds is an alternative to increase the selectivity of the sensors. For example, the work of Zhao et al. (2021) developed a Pt-Al_2O_3 catalyst film that was printed on top of Pd-In_2O_3 sensors acting as a filter reducing cross-sensitivities to VOCs without affecting the sensing mechanism. The use of catalyst is recommended when working with MOS sensors, due to their high cross sensitivities to other VOC. Our past experiments compared mainly two types of TGS, one of them including a carbon filter, on which a degradation of the sensitivity is observed as a tradeoff for an increase of the selectivity to CH_4. Therefore, two main considerations need to be studied. In the first place, the selection of the adapted materials for the catalyst considering the final application of the sensors, since harsh environments like oil and gas facilities could produce a fast degradation in the material of the catalyst. Secondly, the effect on the sensitivity and the response time of the catalyst on the sensors.

5.2.2 Calibration models

All the calibration models developed for the three experiments considered instantaneous measurements from the sensors to estimate CH_4 concentrations. The inclusion of temporal relationships into the reconstruction models would provide insights on the evolution of the signal, hence this will result in robust models. Recursive neural networks (RNN) are a type of neural networks that proved to be useful in derive relationships on sequential data. For MOS sensors we recommend to study one special type of RNN called Long-short term memory (LSTM) that allows to

provide outputs considering values of inputs of several steps past in time.

Autoencoders are another special type of artificial neural network which are trained to copy the inputs onto its outputs. They are composed by two main structures the encoder and the decoder. Autoencoders are designed to produce an approximative representation of the inputs forcing them to prioritize the learning of useful properties of the data (Goodfellow et al., 2016). Several tests can be conducted using autoencoders, such as: i) exploiting their feature to encode the data to a latent space, and then applying some transformations, like filtering variables to remove the noise component of the TGS signal, and decoding the transformed data into a representation in the original domain; ii) correcting the cross-sensitivities or drifting directly in the latent space (Yan and Zhang, 2016); iii) defining relationships between the predictors and the target variables in this restricted representation of the data and retrieving the reconstructed signal applying the decoder.

LSTM networks and autoencoders are powerful tools that can help to improve the quality of the reconstructed CH_4 concentrations from TGS. However, they need large amounts of data to train the models to prevent overfitting and to determine the relationships between the variables. Therefore, before implementing such models, it is important to provide large and good quality datasets to train the models.

5.2.3 Considerations in the deployment of network of sensors

Low cost metal oxide semiconductors are an alternative to high precision instruments to be deployed in dense network of sensors. The work of Chamberland and Veeravalli (2006) tackled the problematic of designing a network of sensors for the detection, and they specifically studied the effect on the performance of increasing the density of sensors in a finite space. Nevertheless, their assessment was based on theoretical data. Therefore, an evaluation for a real deployment should be studied. This specially brings questions about the maintenance of the network, the optimal distribution and the number of low-cost nodes that are required to provide good detection and reliable information of the fugitive emissions in a site.

Those networks should be designed considering the optimal calibration approaches for the

network, since recalibration of each node together with a high precision instrument become more inefficient as the network increases. Two calibration techniques are of particular interest for dense networks. The first was approached by the work of Malings et al. (2019) for electrochemical sensors. It consisted in determining the calibration model considering a subset of measurements of a cluster of nodes deployed in the same region. Information from different sensors were treated as a 'virtual instrument' and a model was trained. The study reported a bigger decrease in the performances of the calibration than using individualized models by nodes. This tradeoff in performance could be acceptable for larger clusters of nodes. The second calibration technique tested on MOS sensors for O_3 by Okorn and Hannigan (2021) is called 'multi-hop calibration'. Developed for mobile sensors, it consisted in having only one low-cost sensor collocated with a high precision reference instrument on which it will be defined a calibration model. This first step is called 'one-hop'. The calibrated sensor will be the reference for a second sensor and a new calibration model will be defined. This process will be repeated chaining all the sensors in the network. This methodology has its own challenges, being the most important one the propagation of errors.

5.2.4 Source rate and location estimation

Regarding the inverse modelling, our reconstruction and inverse models were computed independently in different stages. The next step would be the definition of calibration models that optimize the reconstruction of the signal for this inversion modelling. Therefore, it is important to understand the way misfits from the reconstructed signal affect the computation of gradients in the inverse model.

The work of Travis et al. (2020) showed the possibility to provide estimations of the release rate and location from measurements from the sensors. They used MIR sensors with artificial neural networks. The experiment provided information of CH_4 enhancements from several sensors together with environmental parameters to a multilayer perceptron model. This model had two units as output which provided the estimation of the release rate and location respectively.

The data used in the training on the MLP model comes from simulated data from a LES model. This approach is interesting because it opens the possibility to generate a large set of examples to train the models under different configurations of receptors and environmental conditions, increasing the generalization of the models. Models trained with simulated data and fine-tuned with data from controlled releases or measurement campaigns could improve their accuracy. It is important to consider that complex topography needs to be introduced on LES simulations to provide useful data for the ANN models.

Appendix A

Appendix Chapter 2

Table A.1: Mean MSD and RMSE for the 50 training and test periods of the sensitivity test.

		Mean MSD (ppm^2)	**Mean RMSE (ppm)**
Training set	Reference	0.001352331	0.036774055
	W/O Pressure	0.002216097	0.047075444
	W/O Temperature	0.001535907	0.039190651
	$\mathrm{W/O\ H_2O_{Mole\ Fraction}}$	0.002176811	0.046656307
	W/O CO	0.002071878	0.045517882
	W/O Figaro	0.001626768	0.040333216
	3xTGS 26xx types	0.001183233	0.034398159
	TGS 2600 & TGS 2611-C00	0.001441292	0.037964357
	2xTGS 2611-C00	0.001723121	0.04151049
Test set	Reference	0.014911814	0.122113937
	W/O Pressure	0.012041034	0.109731645
	W/O Temperature	0.014275558	0.119480365
	$\mathrm{W/O\ H_2O_{Mole\ Fraction}}$	0.018681443	0.136680075
	W/O CO	0.015550217	0.124700508
	W/O Figaro	0.015273629	0.123586523
	3xTGS 26xx types	0.0224715	0.14990497
	TGS 2600 & TGS 2611-C00	0.0178823	0.133724719
	2xTGS 2611-C00	0.01768717	0.132993119

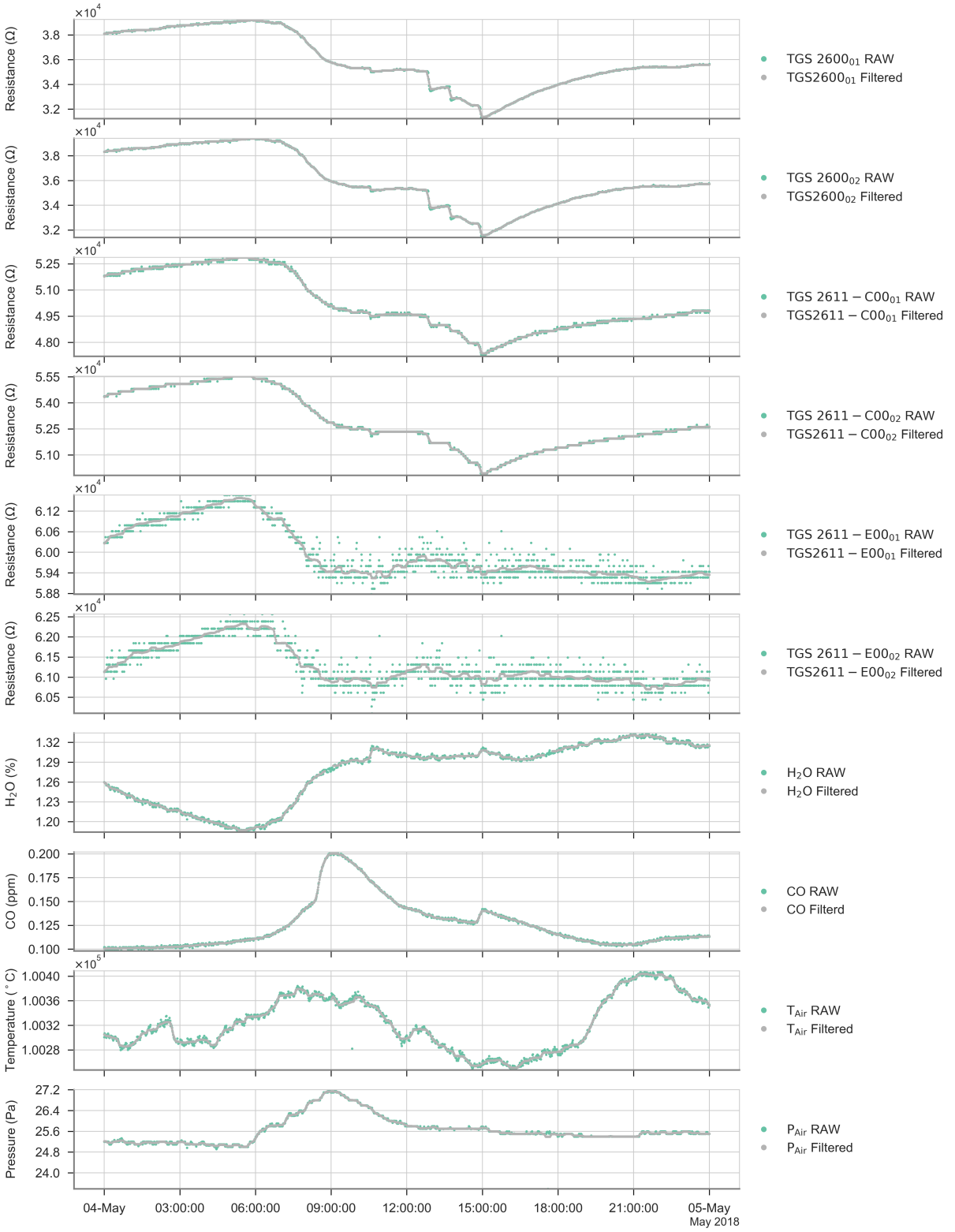


Figure A.1: Comparison between the raw (green) and the filtered signal (gray) over one day.

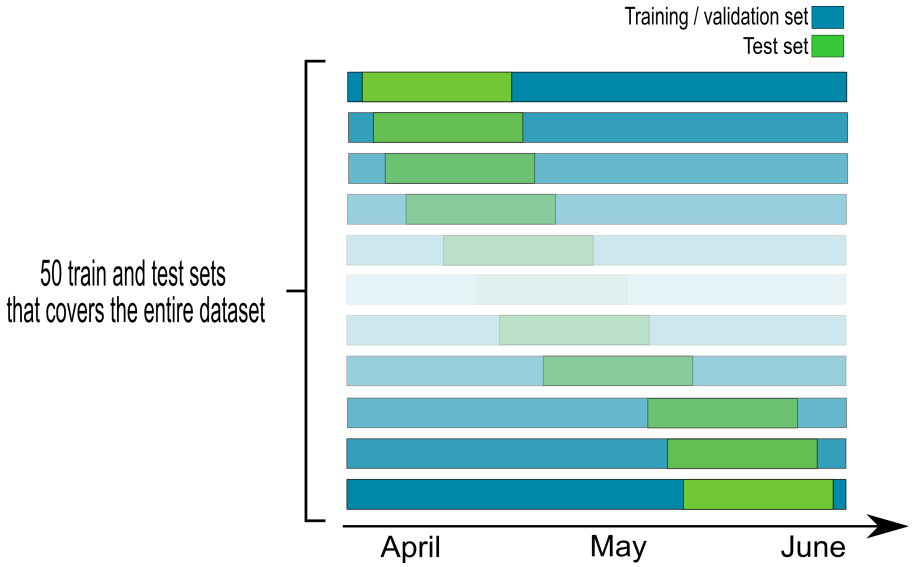


Figure A.2: Diagram showing the process of training and evaluation of the model with 50 training and test sets covering the entire time series.

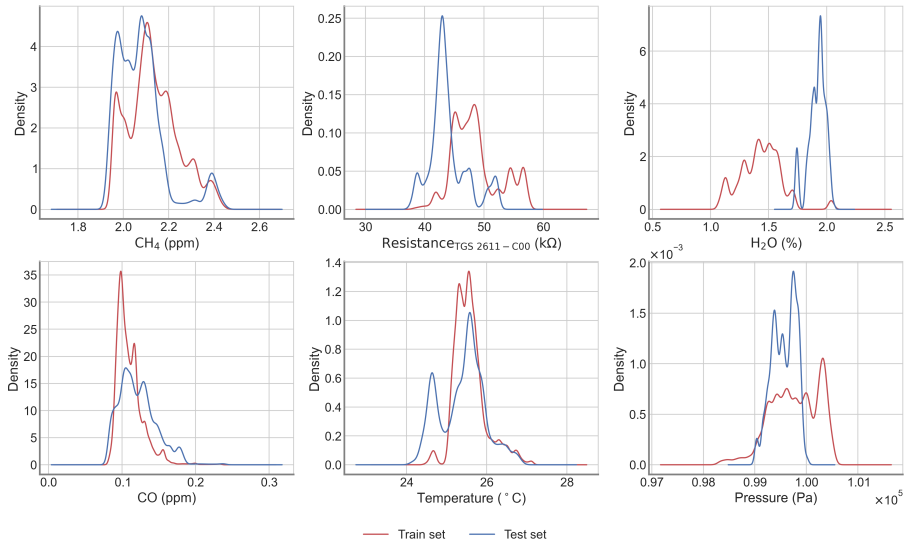


Figure A.3: Density distribution of the training (red) and test (blue) set for the worst case (50).

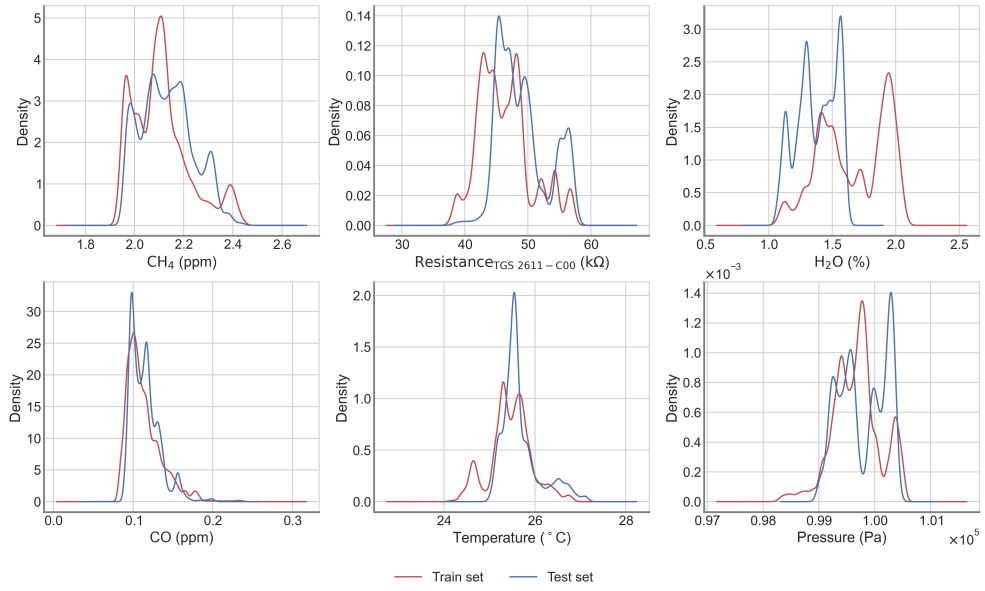


Figure A.4: Density distribution of the training (red) and test (blue) set for the best case (7).

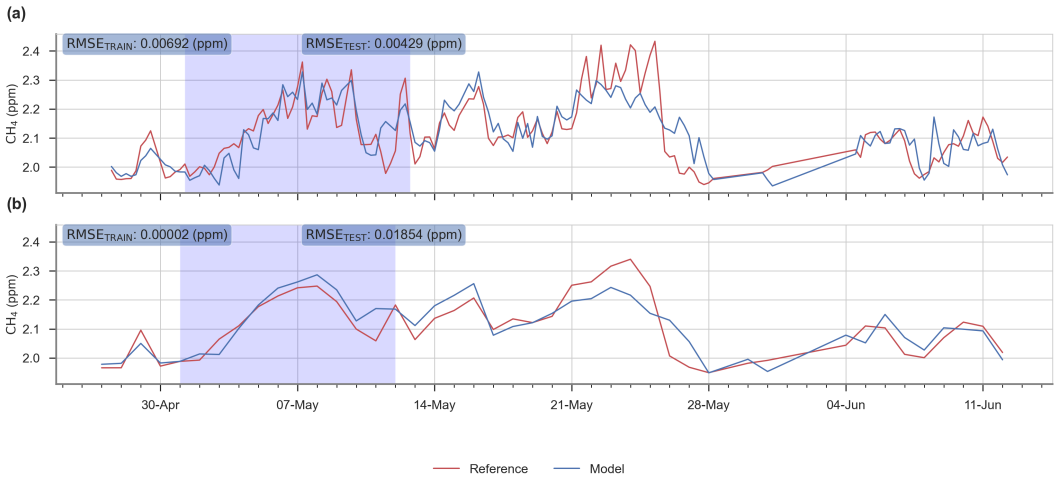


Figure A.5: Output of the model for a smoothed signal of 12 h (**a**) and 24 h (**b**).

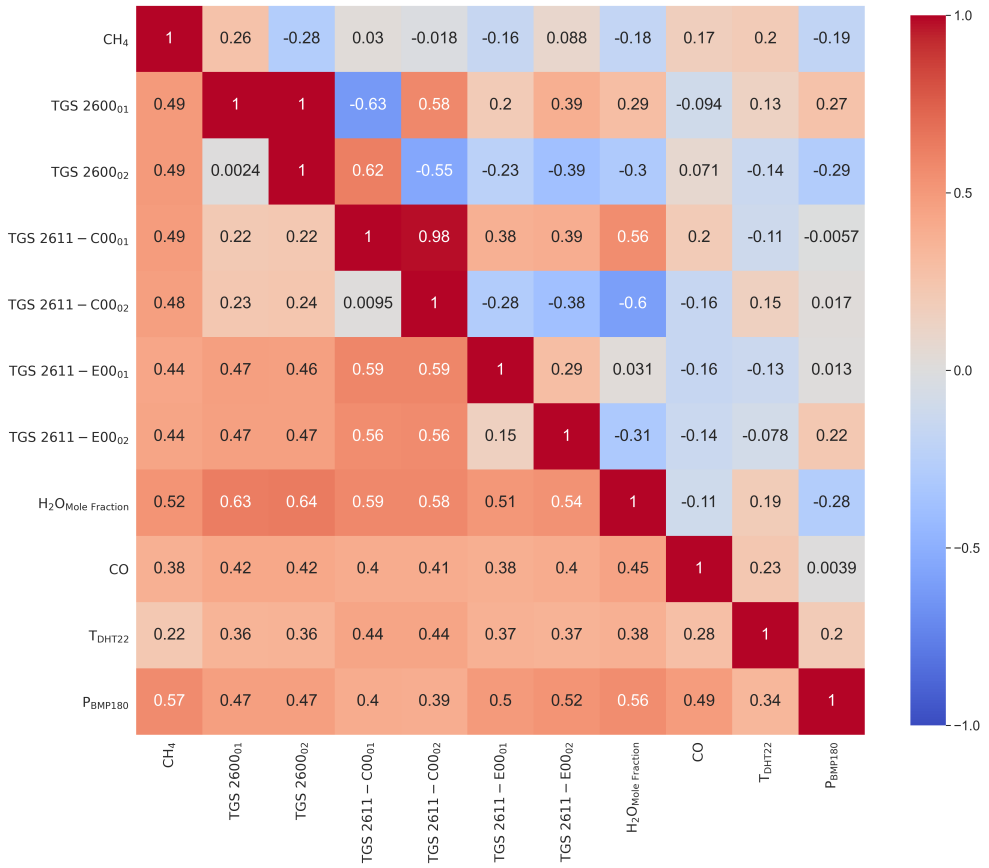


Figure A.6: Partial correlation (r) matrix (upper triangular) and standard deviation of correlation for bins of 3 days previously smoothed at an hourly scale (lower triangular).

Appendix B

Appendix Chapter 3

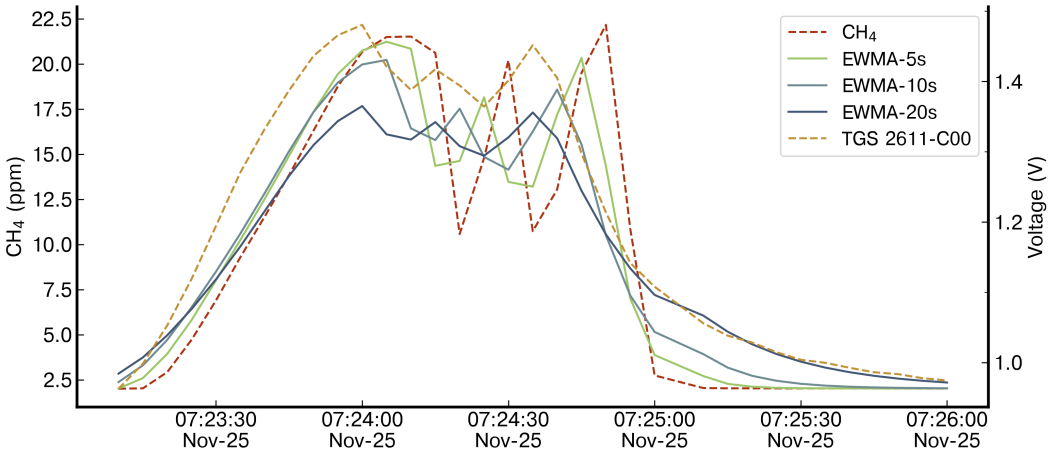


Figure B.1: Different time constant values of the exponential weighted moving average (EWMA) applied to the reference instrument. The reference instrument in red dotted lines, the applied smoothing for three values of time constant (5s, 10s, and 20s) denoted 'EWMA' for one peak and the TGS 2611-C00 voltage from logger A to compare the smoothing effect in yellow dotted.

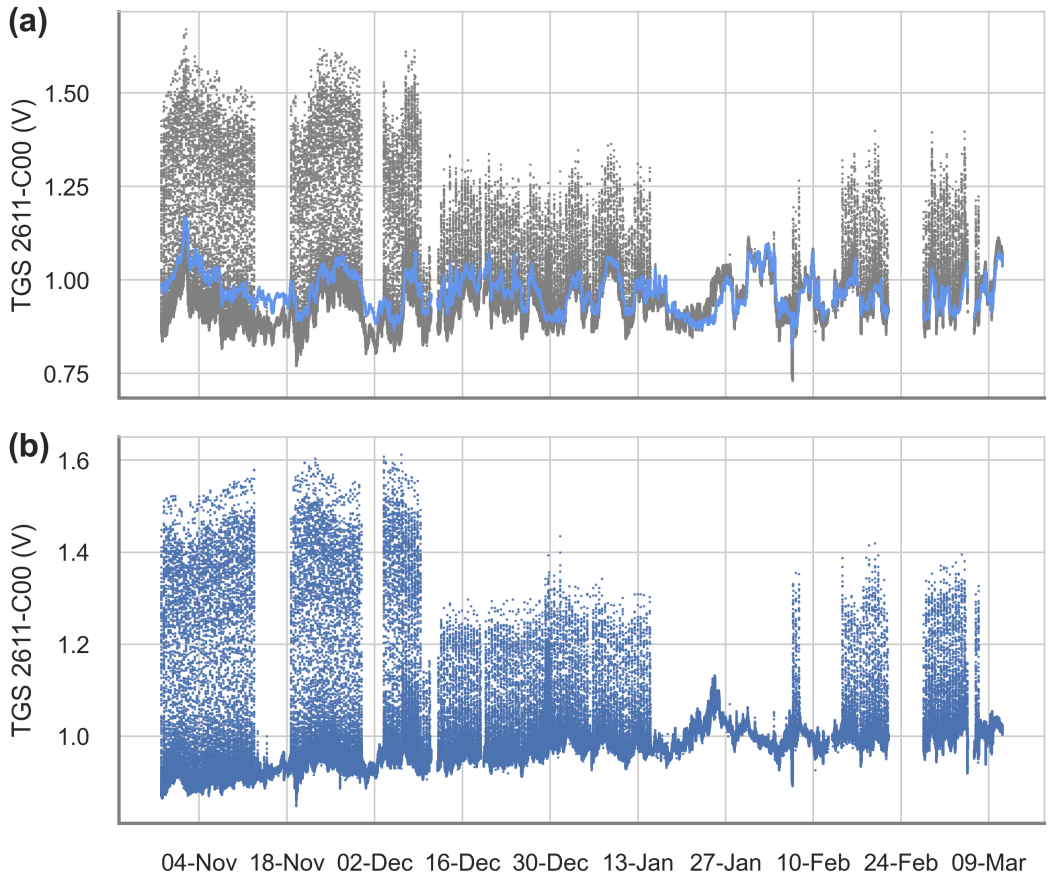


Figure B.2: Derived contribution and correction of water vapor for the Figaro TGS 2611-C00. (a) The raw voltage signal (gray) and the derived cross-sensitivities to H_2O (blue). (b) The cross-sensitivity corrected signal.

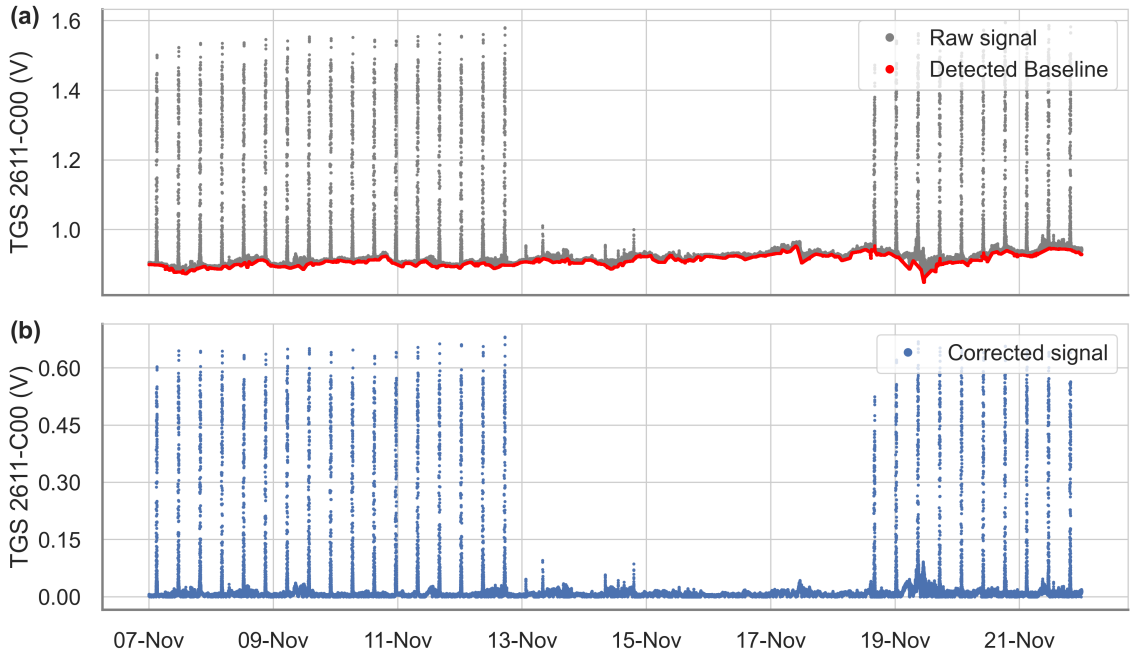


Figure B.3: Example of the baseline extraction and correction for the Figaro TGS 2611-C00 over 15 days. (a) Raw signal (gray) and detected baseline with the spike detection algorithm (red). (b) Voltage signal with the corrected baseline.

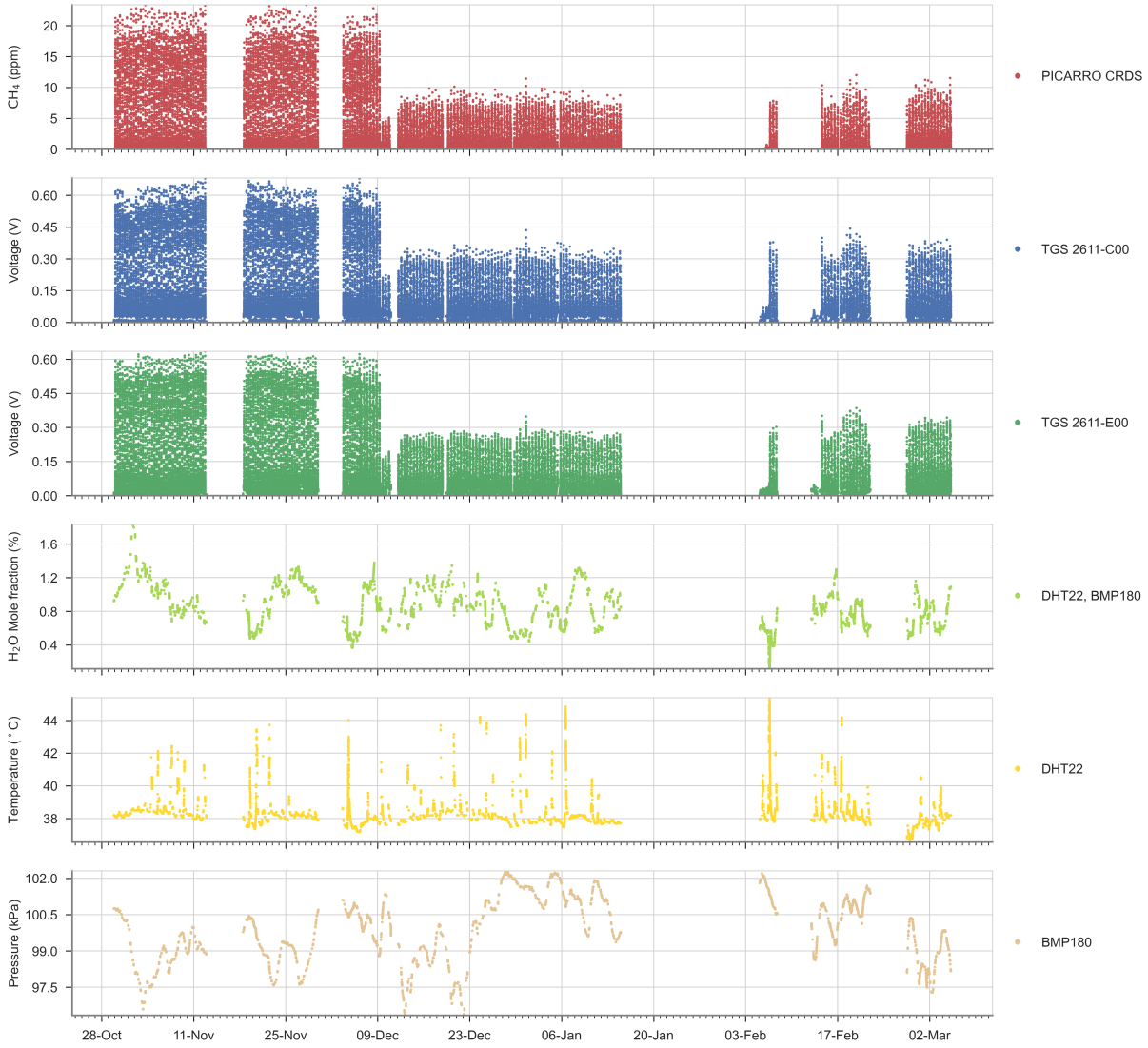


Figure B.4: Time series of the reference CH_4 signal, Figaro® TGS sensor, and environmental variables for the entire experiment.

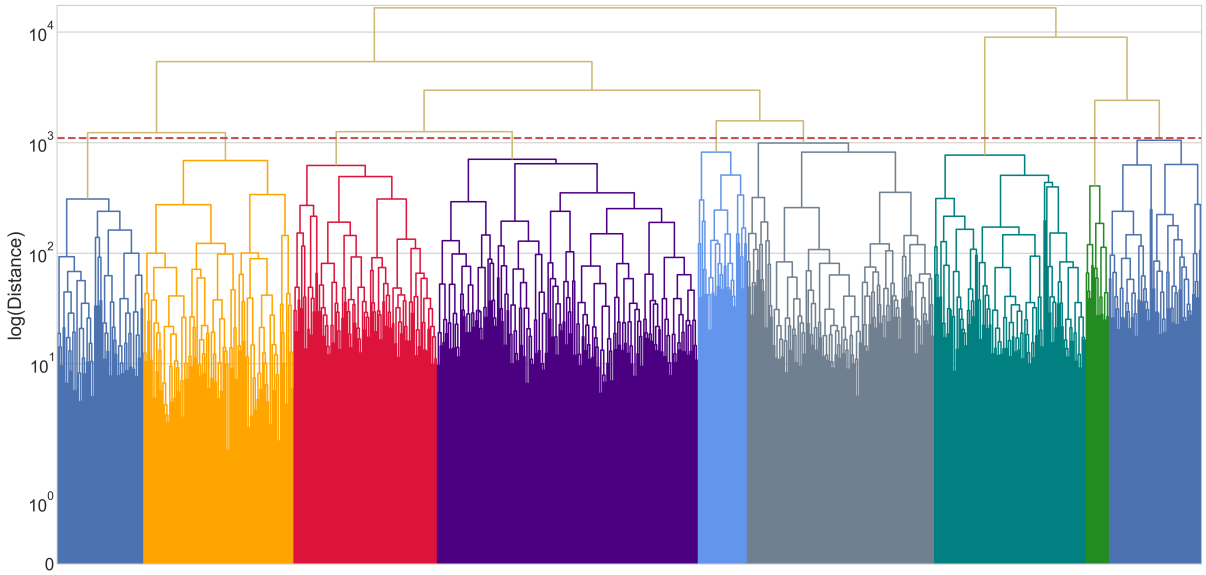


Figure B.5: Dendrogram constructed from the distance matrix computed using the DWT metric. Red dotted line represents the threshold used to determine the clusters. Each color under the threshold line represents one cluster of peaks. Note that y-axis was rescaled to the logarithm of the 'ward' distance to appreciate better the threshold and the clusters.

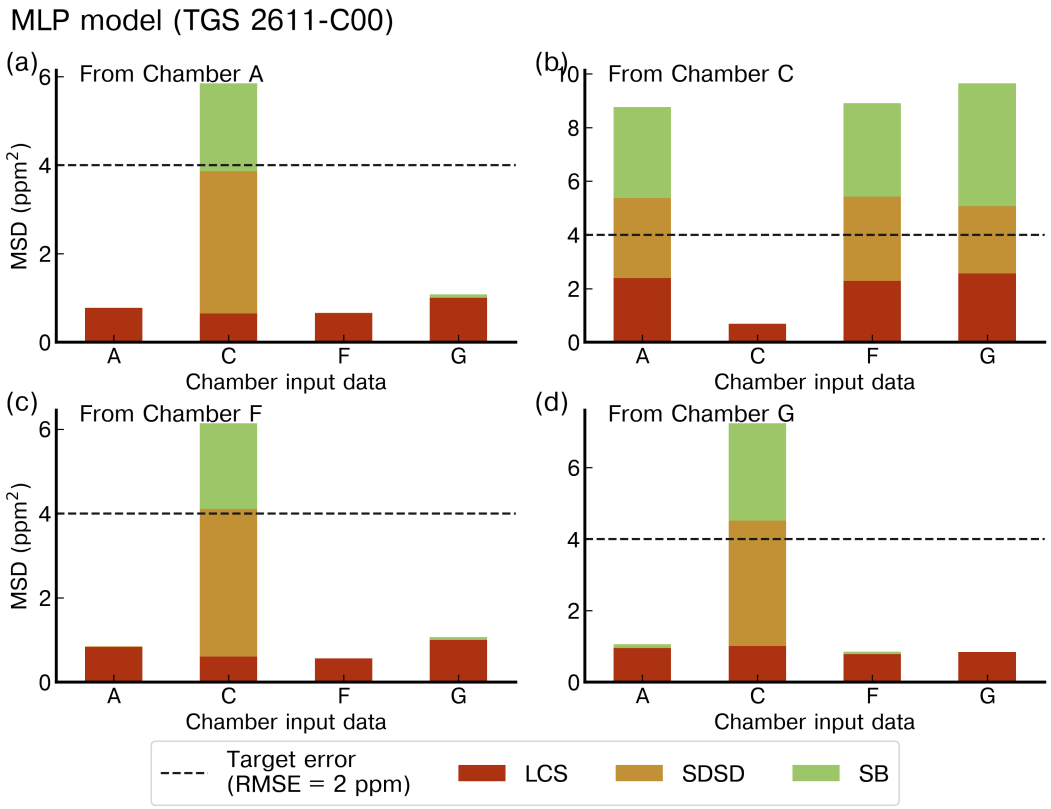


Figure B.6: Reconstruction error of the peaks for the MLP model with the TGS 2611-C00 as input and using the best stratified training on the (a) Chamber A, (b) Chamber C, (c) Chamber F and (d) Chamber G. The first column on each panel is the reconstruction error on the test set of the chamber on which the training was made, the other columns are the reconstruction on the whole dataset for that chamber on the same sensor. Note the different ranges of the y-axis.

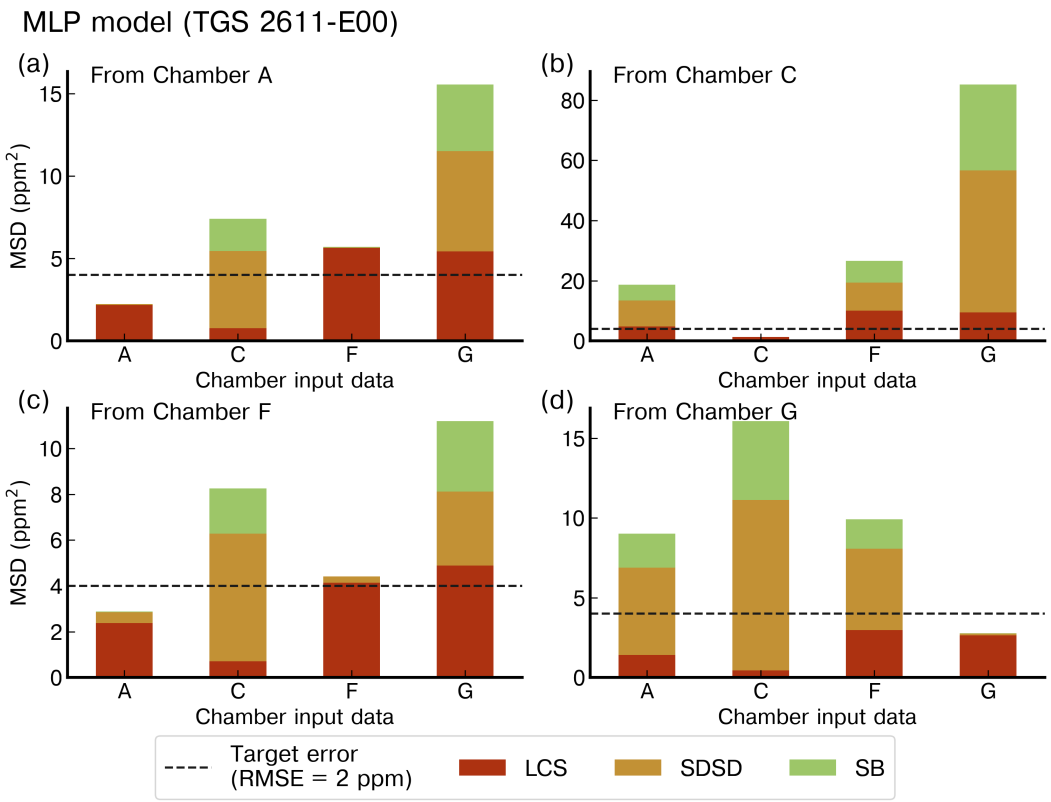


Figure B.7: Reconstruction error of the peaks for the MLP model with the TGS 2611-E00 as input and using the best stratified training on the (a) Chamber A, (b) Chamber C, (c) Chamber F and (d) Chamber G. The first column on each panel is the reconstruction error on the test set of the chamber on which the training was made, the other columns are the reconstruction on the whole dataset for that chamber on the same sensor. Note the different ranges of the y-axis.

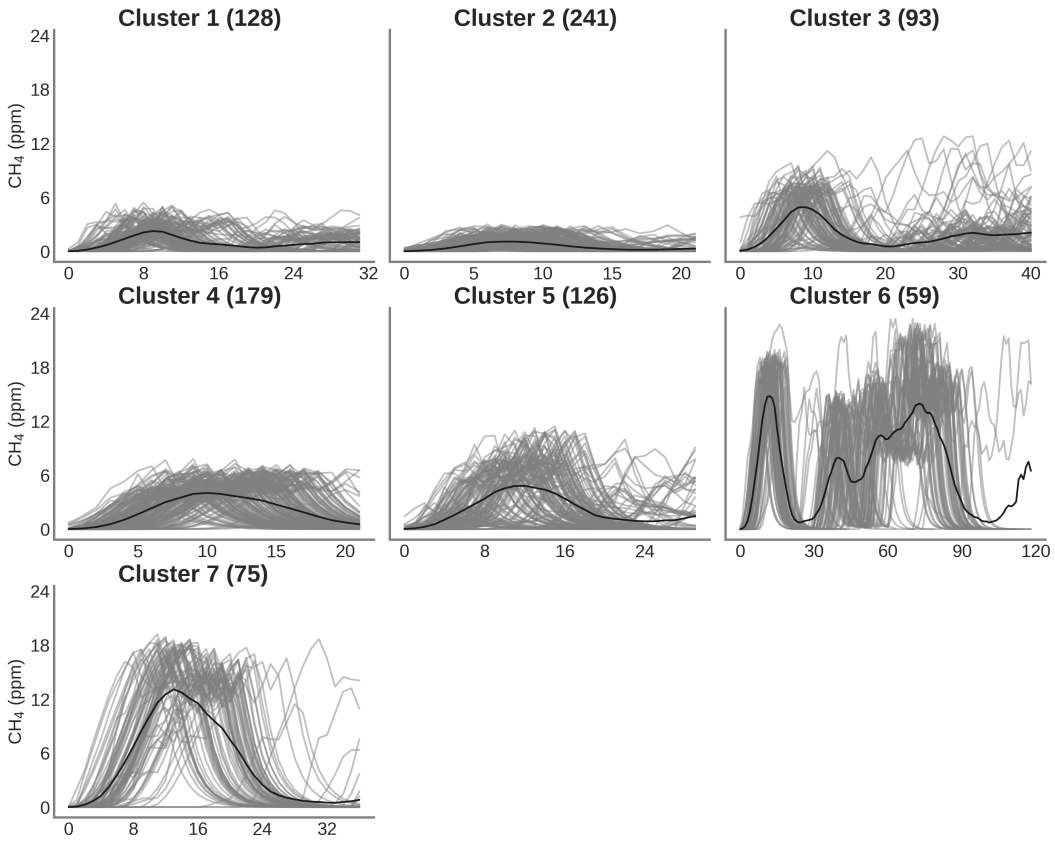


Figure B.8: Clustering of peaks using DTW on the reference instrument for the same spikes detected by sensors on Chamber C. On the title of each plot the number inside the parentheses correspond to the number of spikes attributed to each cluster. Thin gray lines represent all the peaks inside each cluster and the black line is the mean of all the peaks corresponding to each class.

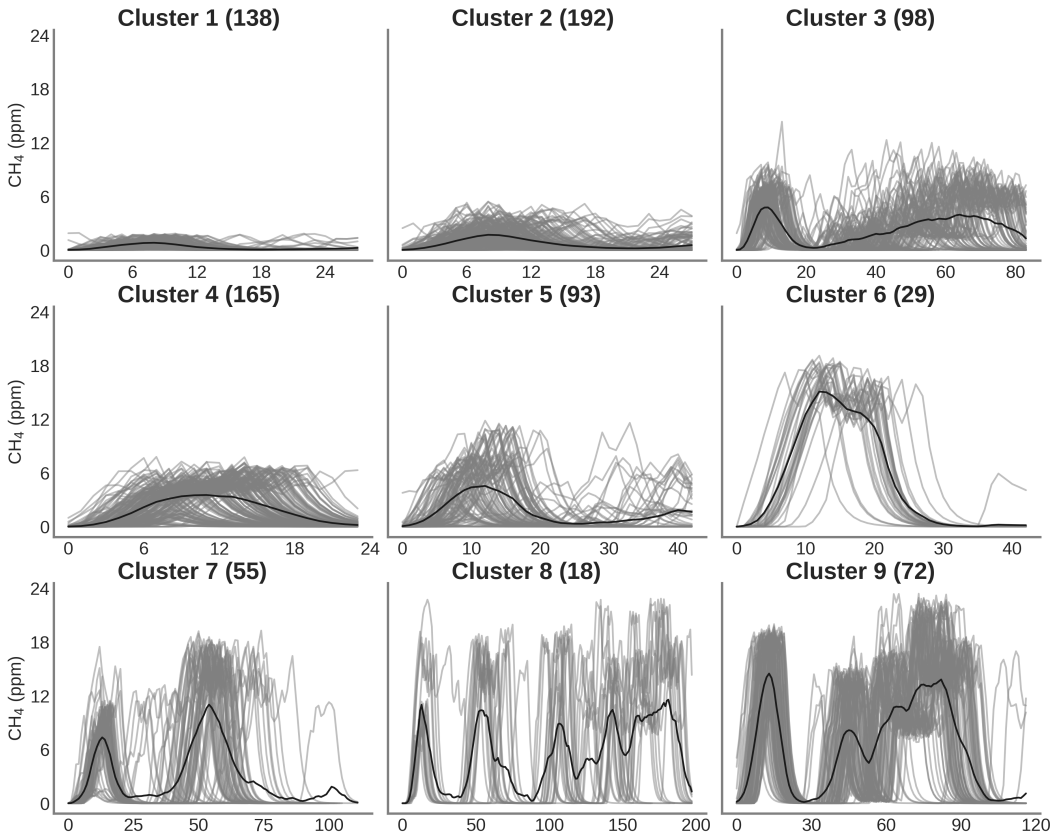


Figure B.9: Clustering of peaks using DTW on the reference instrument for the same spikes detected by sensors on Chamber F. On the title of each plot the number inside the parentheses correspond to the number of spikes attributed to each cluster. Thin gray lines represent all the peaks inside each cluster and the black line is the mean of all the peaks corresponding to each class.

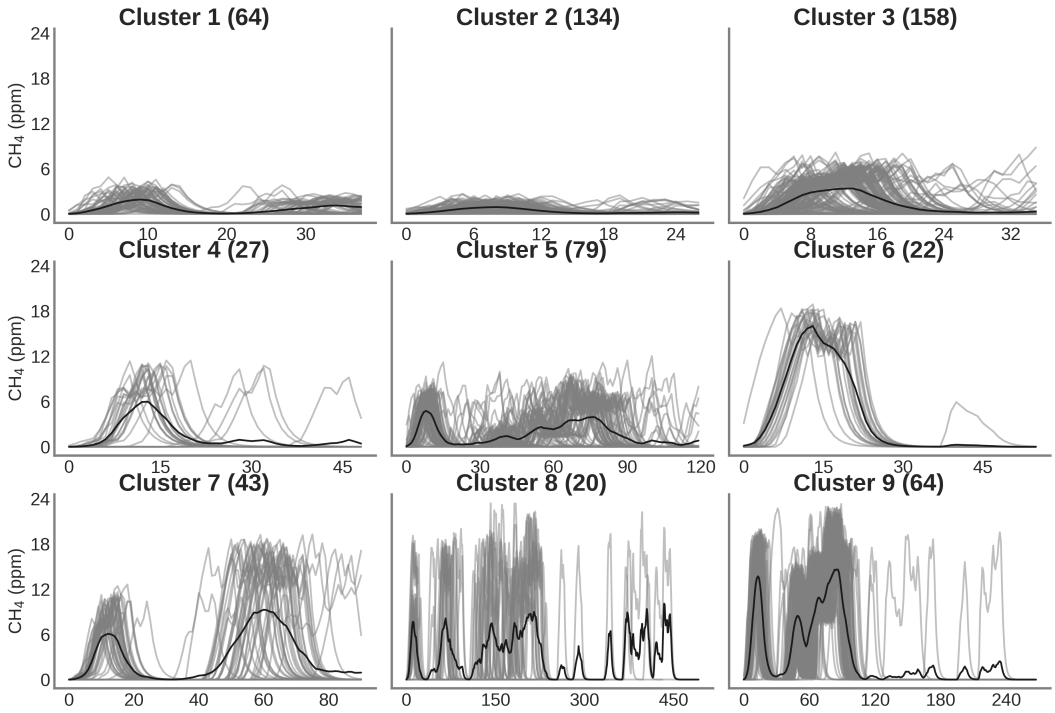


Figure B.10: Clustering of peaks using DTW on the reference instrument for the same spikes detected by sensors on Chamber G. On the title of each plot the number inside the parentheses correspond to the number of spikes attributed to each cluster. Thin gray lines represent all the peaks inside each cluster and the black line is the mean of all the peaks corresponding to each class.

Table B.1: MSD decomposition for the different configurations and both test set sizes considering only Figaro® TGS sensors as input. Letters inside parentheses indicates the sensor used: TGS 2611-E00 denoted 'E', TGS2611-C00 denoted 'C' and both sensors as input denotes 'C&E'

Test set size	Model	MSD (ppm^2)	LCS (ppm^2)	SDSD (ppm^2)	SB (ppm^2)
30%	Linear (E)	3.51	3.17	0.326	0.013
	Poly (E)	3.23	3.06	0.155	0.014
	RF (E)	4.74	4.67	0.061	0.011
	RF-h (E)	4.67	4.60	0.057	0.010
	MLP (E)	3.24	3.07	0.163	0.011
	Linear (C)	1.24	1.00	0.229	0.015
	Poly (C)	0.88	0.84	0.016	0.021
	RF (C)	1.27	1.24	0.014	0.012
	RF-h (C)	1.21	1.19	0.012	0.011
	MLP (C)	0.85	0.82	0.014	0.014
	Poly (C&E)	0.73	0.70	0.010	0.023
	RF (C&E)	0.81	0.79	0.009	0.009
	RF-h (C&E)	0.78	0.76	0.008	0.009
	MLP (C&E)	0.71	0.69	0.009	0.011
50%	Linear (E)	3.59	3.09	0.457	0.042
	Poly (E)	3.24	3.00	0.218	0.022
	RF (E)	5.16	4.64	0.415	0.103
	RF-h (E)	4.59	4.46	0.115	0.014
	MLP (E)	3.64	3.06	0.481	0.097
	Linear (C)	1.39	1.04	0.304	0.050
	Poly (C)	0.92	0.84	0.047	0.025
	RF (C)	1.49	1.27	0.164	0.054
	RF-h (C)	1.37	1.20	0.127	0.040
	MLP (C)	1.11	0.86	0.192	0.063
	Poly (C&E)	0.79	0.71	0.051	0.030
	RF (C&E)	1.07	0.85	0.158	0.054
	RF-h (C&E)	2.34	1.78	0.314	0.243
	MLP (C&E)	0.91	0.71	0.144	0.059

Table B.2: MSD decomposition for the different configurations and both test set sizes considering Figaro® TGS sensors and environmental variables as input. Notation is the same as on Table B.1

Test set size	Model	MSD (ppm^2)	LCS (ppm^2)	SDSD (ppm^2)	SB (ppm^2)
30%	Linear (E)	3.59	3.09	0.45	0.042
	Poly (E)	3.24	3.00	0.21	0.022
	RF (E)	5.16	4.64	0.41	0.103
	RF-h (E)	4.59	4.46	0.11	0.014
	MLP (E)	3.64	3.07	0.48	0.097
	Linear (C)	1.39	1.04	0.30	0.050
	Poly (C)	0.92	0.84	0.04	0.025
	RF (C)	1.49	1.27	0.16	0.054
	RF-h (C)	1.37	1.20	0.12	0.040
	MLP (C)	1.11	0.86	0.19	0.063
	Poly (C&E)	0.79	0.71	0.05	0.030
	RF (C&E)	1.07	0.85	0.15	0.054
	RF-h (C&E)	2.34	1.78	0.31	0.243
	MLP (C&E)	0.91	0.71	0.14	0.059
50%	Linear (E)	3.60	3.09	0.46	0.045
	Poly (E)	3.28	2.98	0.26	0.034
	RF (E)	4.02	3.30	0.62	0.094
	RF-h (E)	4.41	4.11	0.26	0.032
	MLP (E)	3.56	2.97	0.50	0.077
	Linear (C)	3.60	3.09	0.46	0.045
	Poly (C)	0.79	0.74	0.03	0.020
	RF (C)	1.19	0.94	0.20	0.051
	RF-h (C)	2.92	2.61	0.25	0.049
	MLP (C)	0.96	0.77	0.15	0.035
	Poly (C&E)	0.69	0.64	0.02	0.023
	RF (C&E)	1.07	0.82	0.19	0.056
	RF-h (C&E)	33933.88	566.99	33319.58	47.296
	MLP (C&E)	0.81	0.67	0.11	0.031

Table B.3: RMSE in ppm for the different configurations of subsetting based on the selected clusters of peaks. Only Figaro® sensors were used to compute this errors. Each configuration is denoted 'CX' with X the number of the configuration. On each row the models are denoted with a letter inside parentheses to indicate the sensors used. 'C' for the TGS 2611-C00, 'E' for the TGS 2611-E00 and 'C&E' for both sensors.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**
Linear (C)	1.121	1.115	1.126	1.119	1.163	1.141	1.142	1.105	1.098	1.074	1.105
Poly (C)	0.962	0.966	0.972	0.963	0.988	0.982	0.970	0.943	0.952	0.902	0.957
RF (C)	1.140	1.135	1.149	1.137	1.176	1.158	1.136	1.111	1.123	1.075	1.128
RF-h (C)	1.118	1.113	1.127	1.113	1.152	1.136	1.112	1.088	1.101	1.061	1.106
MLP (C)	0.943	0.951	0.957	0.943	0.976	0.964	0.948	0.926	0.928	0.893	0.937
Linear (E)	1.971	1.932	1.951	1.926	1.988	1.972	1.947	1.838	1.887	1.772	1.909
Poly (E)	1.899	1.861	1.874	1.852	1.909	1.896	1.870	1.764	1.817	1.703	1.838
RF (E)	2.277	2.229	2.261	2.207	2.261	2.251	2.184	2.264	2.233	2.230	2.235
RF-h (E)	2.263	2.214	2.245	2.192	2.243	2.235	2.165	2.249	2.218	2.221	2.219
MLP (E)	1.898	1.861	1.874	1.853	1.910	1.895	1.869	1.764	1.816	1.705	1.838
Linear (C&E)	1.121	1.115	1.126	1.121	1.164	1.142	1.144	1.105	1.098	1.074	1.105
Poly (C&E)	0.872	0.890	0.895	0.885	0.904	0.911	0.891	0.869	0.881	0.844	0.886
RF (C&E)	0.887	0.943	0.951	0.931	0.978	0.936	0.938	0.919	0.936	0.914	0.938
RF-h (C&E)	0.878	0.929	0.940	0.917	0.964	0.925	0.927	0.906	0.924	0.907	0.924
MLP (C&E)	0.861	0.948	0.875	0.874	0.970	0.958	0.871	0.841	0.862	0.820	0.935

Table B.4: RMSE in ppm for the different configurations of subsetting based on the selected clusters of peaks. Figaro® sensors and environmental variables were used to compute this errors. Notation is the same as on Table B.3.

	C1	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**	**C8**	**C9**	**C10**	**C11**
Linear (C)	1.115	1.111	1.118	1.112	1.155	1.143	1.136	1.106	1.095	1.075	1.101
Poly (C)	0.915	0.920	0.928	0.913	0.928	0.935	0.919	0.887	0.898	0.872	0.912
RF (C)	0.941	0.988	0.997	0.980	1.010	0.988	0.967	0.937	0.950	0.905	0.977
RF-h (C)	0.925	2721.300	2738.400	2752.600	2842.700	2800.200	2827.700	2788.600	0.949	0.902	2681.000
MLP (C)	0.899	0.918	0.914	0.905	0.944	0.926	0.905	0.872	0.884	0.854	0.894
Linear (E)	1.971	1.932	1.951	1.926	1.988	1.972	1.948	1.839	1.888	1.773	1.910
Poly (E)	1.897	1.863	1.879	1.856	1.910	1.897	1.874	1.763	1.820	1.704	1.842
RF (E)	1.979	1.963	1.978	1.951	2.003	2.000	1.959	1.874	1.939	1.834	1.951
RF-h (E)	1.969	1370.000	13792.900	13864.300	14317.900	14103.900	14242.700	14045.700	1.932	1.830	13506.000
MLP (E)	1.897	1.865	1.879	1.860	1.914	1.898	1.875	1.801	1.826	1.706	1.844
Linear (C&E)	1.115	1.111	1.118	1.114	1.156	1.143	1.139	1.106	1.095	1.076	1.101
Poly (C&E)	0.833	0.851	0.858	0.842	0.851	0.864	0.850	0.824	0.836	0.828	0.849
RF (C&E)	0.817	0.886	0.891	0.888	0.920	0.864	0.874	0.855	0.867	0.847	0.885
RF-h (C&E)	0.811	0.914	0.924	0.925	0.951	0.902	0.911	0.893	0.893	0.856	0.913
MLP (C&E)	0.833	0.858	0.880	0.850	0.864	0.883	0.857	0.814	1.524	0.811	0.848

Appendix C

Appendix Chapter 4

Table C.1: Distribution of the releases by chamber. For each chamber is denoted with an 'o' the releases for which the TGS sensors produced valid measurements and with an 'x' the invalid ones.

Release	Chamber						Release	Chamber					
number	A	C	D	E	F	H	number	A	C	D	E	F	H
1	-	-	-	x	x	o	19	o	o	-	o	o	o
2	o	-	-	o	o	o	20	o	o	-	o	o	o
3	-	-	-	o	o	o	21	o	o	-	o	o	o
4	-	-	-	o	o	o	22	-	o	-	o	o	-
5	o	-	o	o	o	-	23	-	-	-	o	-	o
6	o	-	o	o	o	-	24	o	o	x	-	o	o
7	o	-	o	o	o	-	25	o	o	o	o	o	o
8	o	-	o	o	o	-	26	o	o	x	o	-	o
9	o	o	o	o	o	-	27	o	o	-	-	-	x
10	o	o	x	x	o	-	28	o	o	-	-	-	o
11	o	x	-	-	-	-	29	o	o	-	-	o	o
12	x	x	o	o	o	-	30	o	o	-	-	-	o
13	-	x	o	o	o	o	31	-	o	-	-	-	-
14	o	o	o	o	o	o	32	o	o	-	o	-	o
15	o	o	-	-	o	-	33	x	o	-	o	-	o
16	o	-	-	x	-	-							
17	o	o	-	-	-	x							
18	o	o	-	-	o	o							

Table C.2: Comparison of the emission rate estimates ($\mathrm{Q_s}$), location error ($\mathrm{L_e}$) and percentage of error of the rate estimates ($\mathrm{Q_e}$) for the reference instrument and the TGS low-cost sensor from reconstructed CH_4 of the 2nd degree polynomial model.

Release	**Actual emission**	**Reference**			**TGS**		
N°	$(\mathrm{g\ CH_4\ s^{-1}})$	$\mathrm{Q_e\ (g\ CH_4\ s^{-1})}$	$\mathrm{E_l\ (m)}$	**error (%)**	$\mathrm{Q_e\ (g\ CH_4\ s^{-1})}$	$\mathrm{E_l\ (m)}$	**error (%)**
2	1.0	1.10	5.26	10.8	1.19	12,40	19,1
5	3.0	1.34	21.57	55.2	1.53	19,55	48,8
9	1.0	0.88	14.29	11.9	1.03	13,60	2,9
10	0.5	0.40	9.29	18.9	0.34	7,74	30,7
12	1.0	0.34	3.08	65.7	2.99	9,55	199,2
21	1.0	0.63	3.61	36.1	0.64	3,61	35,7
25	5.0	4.61	4.57	7.8	6.50	10,02	30,0
26	0.4	0.31	5.10	20.9	0.43	10,04	9,1
29	0.6	0.45	3.40	24.5	0.41	2,34	30,8
32	4.0	6.55	10.55	63.8	5.42	10,28	35,6
33	2.0	2.45	5.77	22.7	2.11	5,37	5,5
Average error			7.86	30.7		9.5	40.6
σ_{error}			5.46	20.3		4.6	51.9

Table C.3: Summary of the tripods that were connected to each chamber.

Chamber	**Tripod N°**
A	1, 4, 6, 8, 9, 10, 11, 14, 15
C	2, 7, 9, 14, 15, 16
D	2, 3, 9, 10, 11, 12, 13, 16
E	1, 3, 4, 5, 10, 11, 12, 13, 16
F	2, 3, 4, 10, 11, 12, 13, 14, 15
H	4, 5, 6, 7, 12, 13, 14, 15

Table C.4: Comparison between TGS sensors included on the low-cost logging systems during the TADI 2019 campaign.

Type	**Target gas**	**Approximate price**	**Comments**
2600	C_2H_5OH, C_4H_{10}, CO, H_2, CH_4	15 $us	Designed as a smoke detector.
2611-C00	CH_4, C_2H_5OH, C_4H_{10}, CO, H_2	20 $us	Designed for CH_4 detection. Fast response.
2611-E00	CH_4, H_2	20 $us	Designed for CH_4 detection. Increased selectivity due to a carbon filter installed on top of the sensing material.
2602	C_7H_8, H_2S, C_2H_5OH, NH_3, H_2	17 $us	High sensitive to VOC and odor gases.

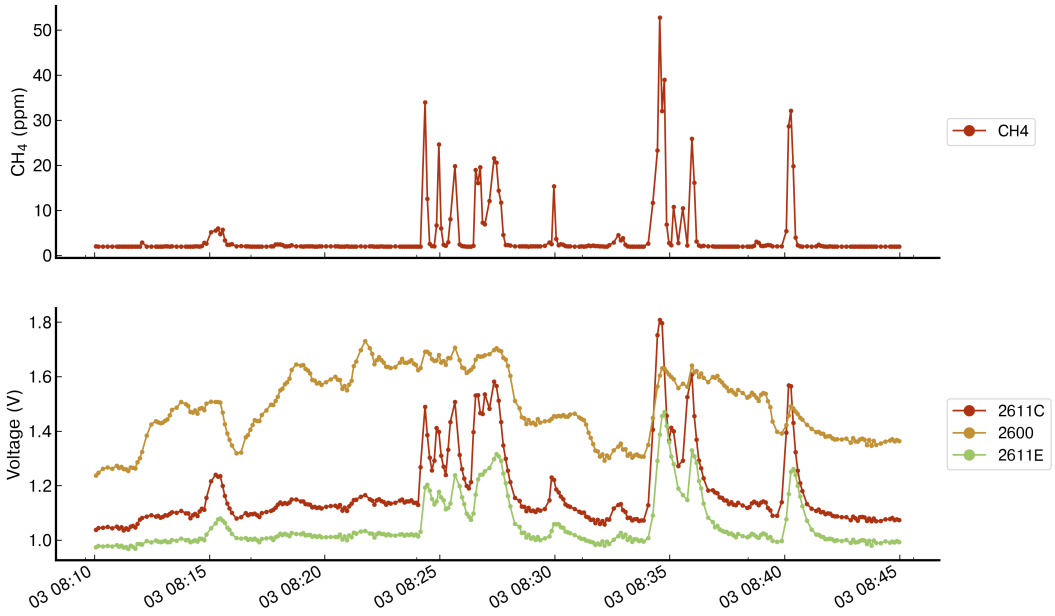


Figure C.1: Comparison of the voltage measurements from three types of TGS included on chamber A. Upper plot shows the reference CH_4 observations measured from the reference instrument. Lower plot shows the voltage observations from TGS 2611-C00, 2600 and 2611-E00.

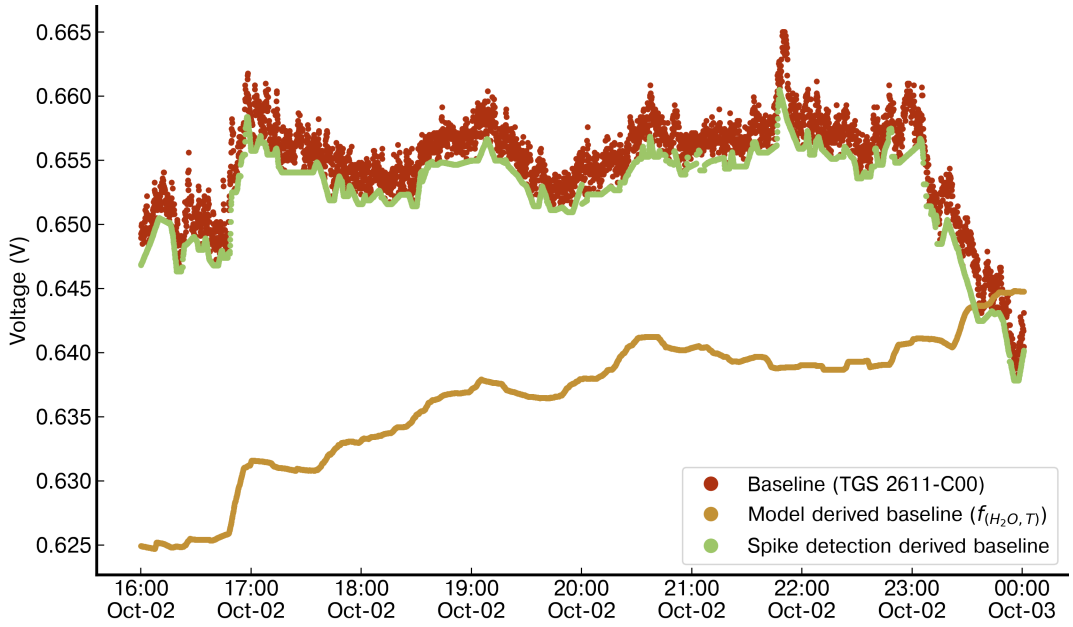


Figure C.2: Comparison of the performance in deriving a baseline signal for the TGS 2611-C00 (red) of Chamber E between a function of H_2O and Temperature (yellow) and a spike detection algorithm (green). The multilinear model derived baseline was trained on six hours of non-release periods at the start of the first day of the campaign and evaluated on the last eight hours of the same day (shown in the figure). The Spike detection algorithm, an iterative function, does not need any prior training and detects the baseline based on neighboring observations and fixed parameters.

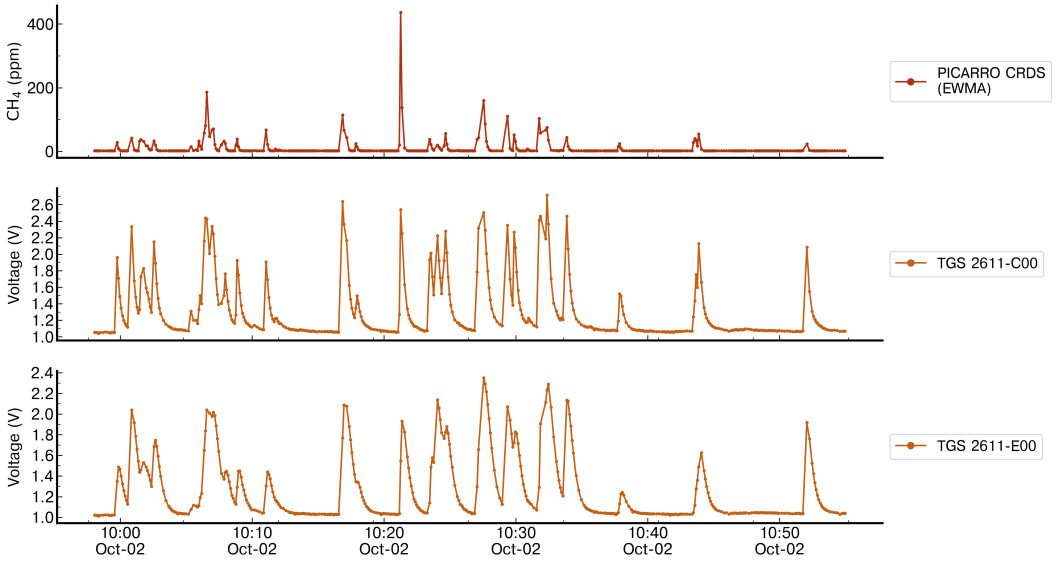


Figure C.3: Comparison of the response of the TGS 2611-C00 and TGS 2611-E00 sensors with CH_4 measurements from the reference instrument for the release #2 which contains spikes with high concentration. The spikes observed on the TGS sensors corresponding from amplitudes between 100 ppm to more than 200 ppm are not distinguishable from spikes with amplitudes lower than 50 ppm.

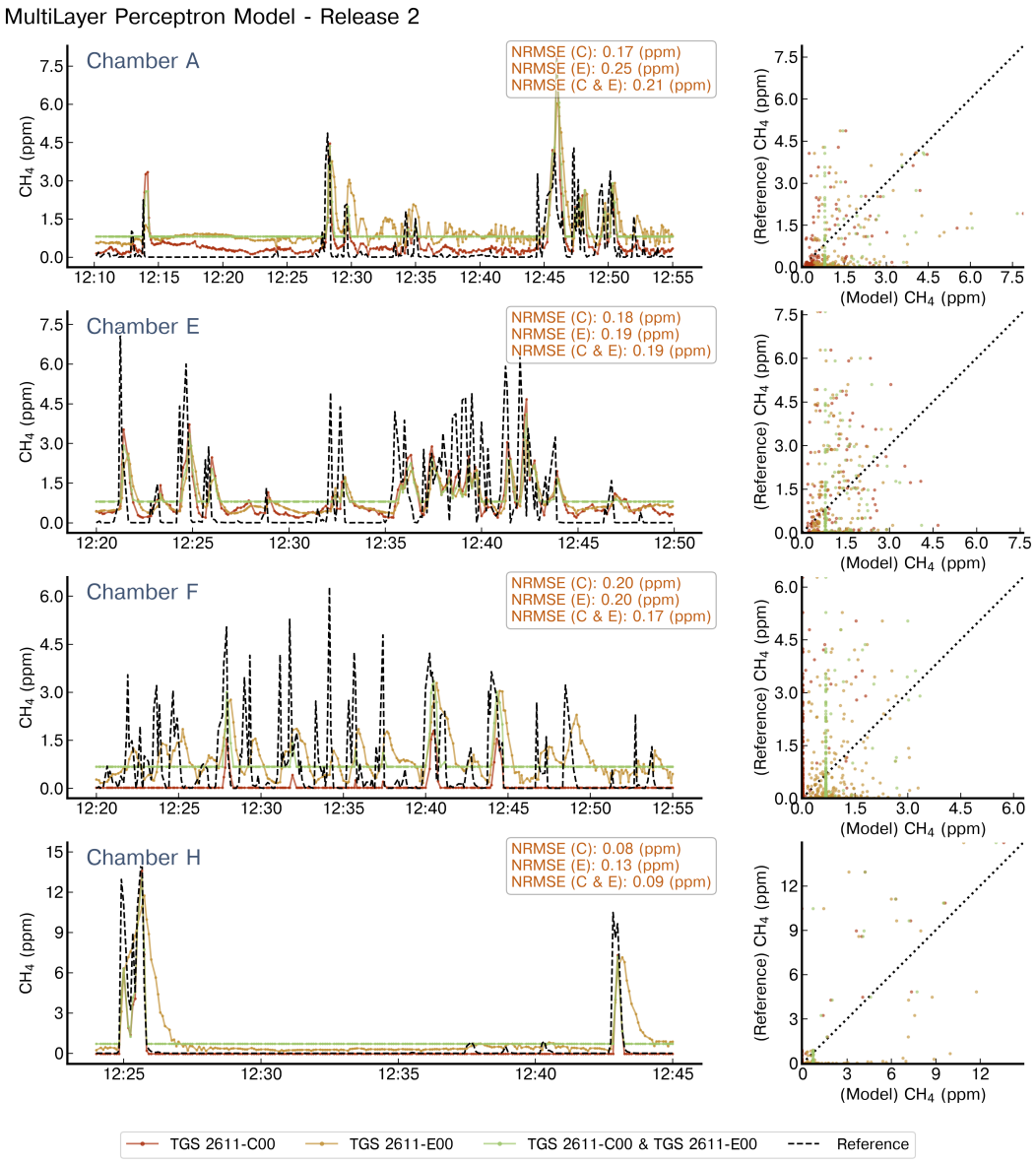


Figure C.4: Reconstruction of release #2 using a MLP model. On left panels are shown the reconstructed CH4 mole fractions for each chamber that captured the release, we present the reference signal (black dotted line), the reconstructed CH_4 mole fractions when the model has as input the TGS 2611-C00 sensor (red), the TGS 2611-E00 (yellow) or both types at the same time (green). The right panels show the 1:1 plot of the reference against the output of the model for the three configurations of inputs. Note the difference in the x-axis for the chambers.

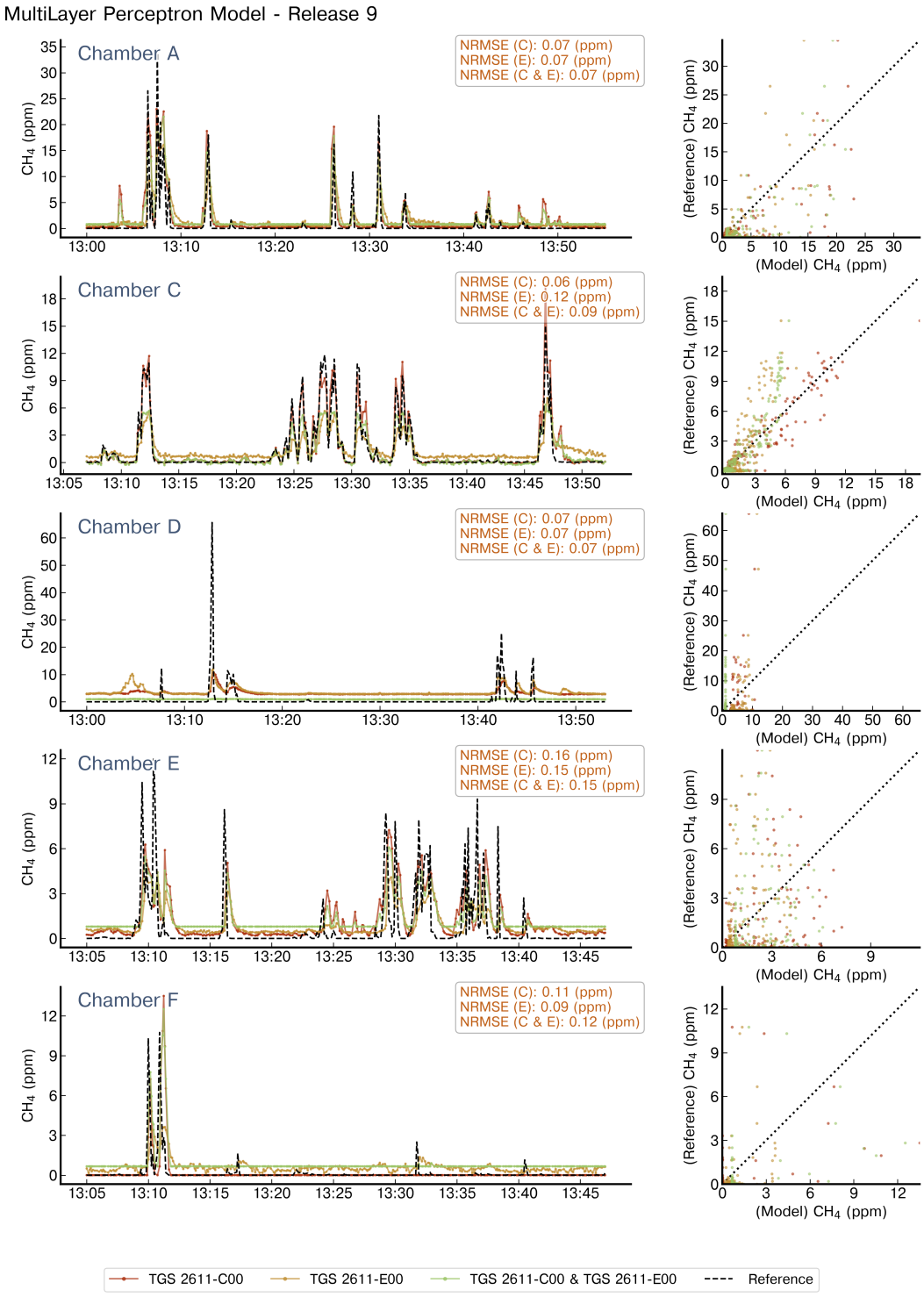


Figure C.5: Reconstruction of release #9 using a MLP model. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

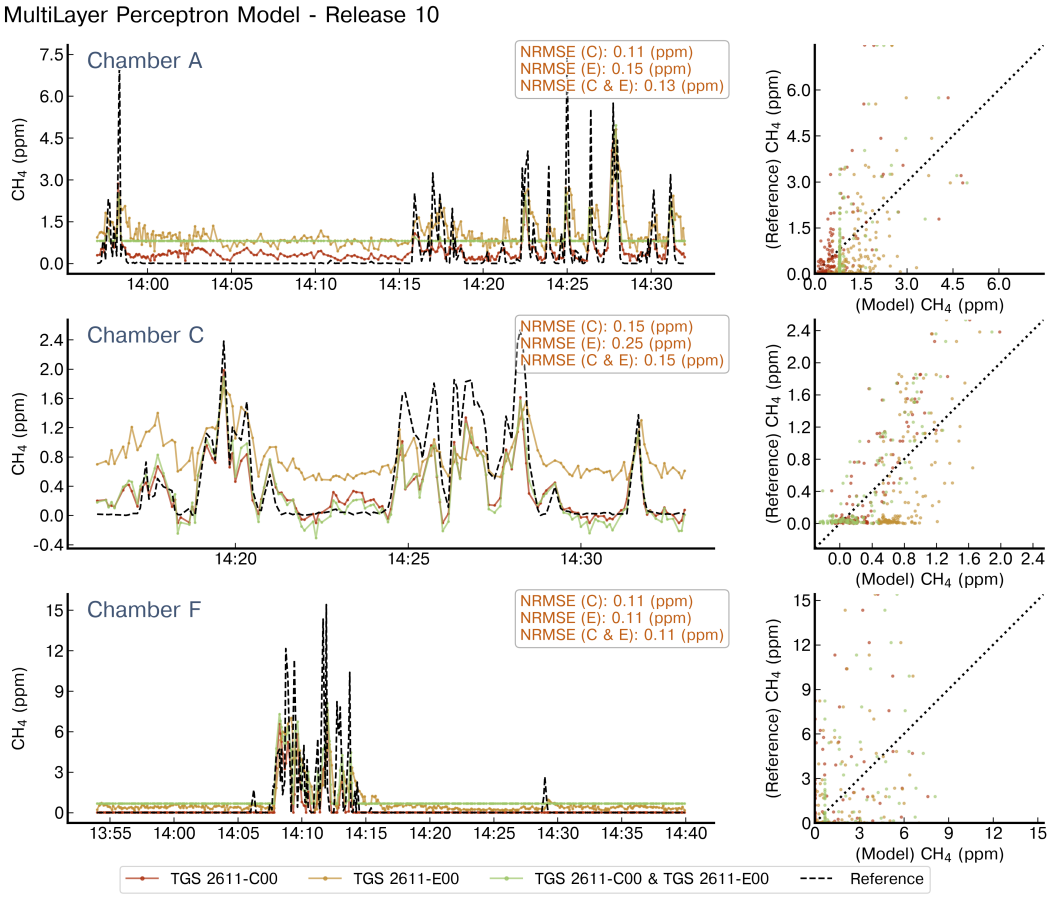


Figure C.6: Reconstruction of release #10 using a MLP model. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

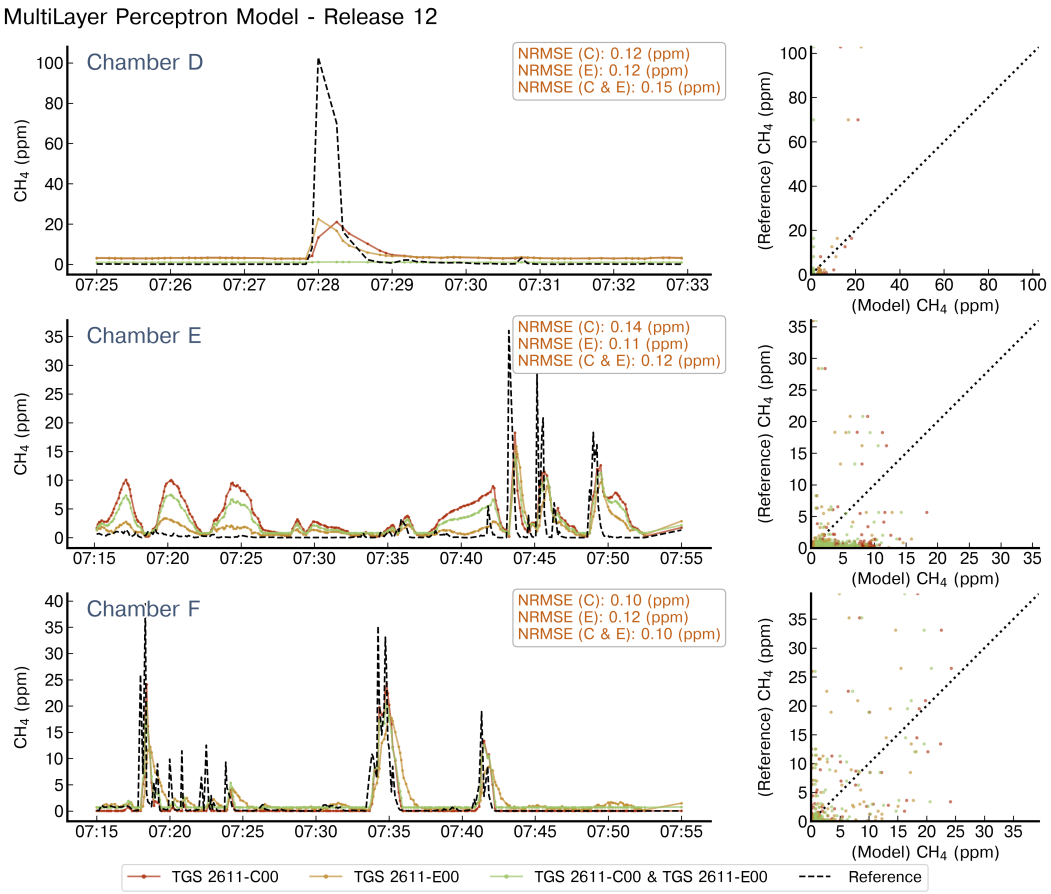


Figure C.7: Reconstruction of release #12 using a MLP model. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

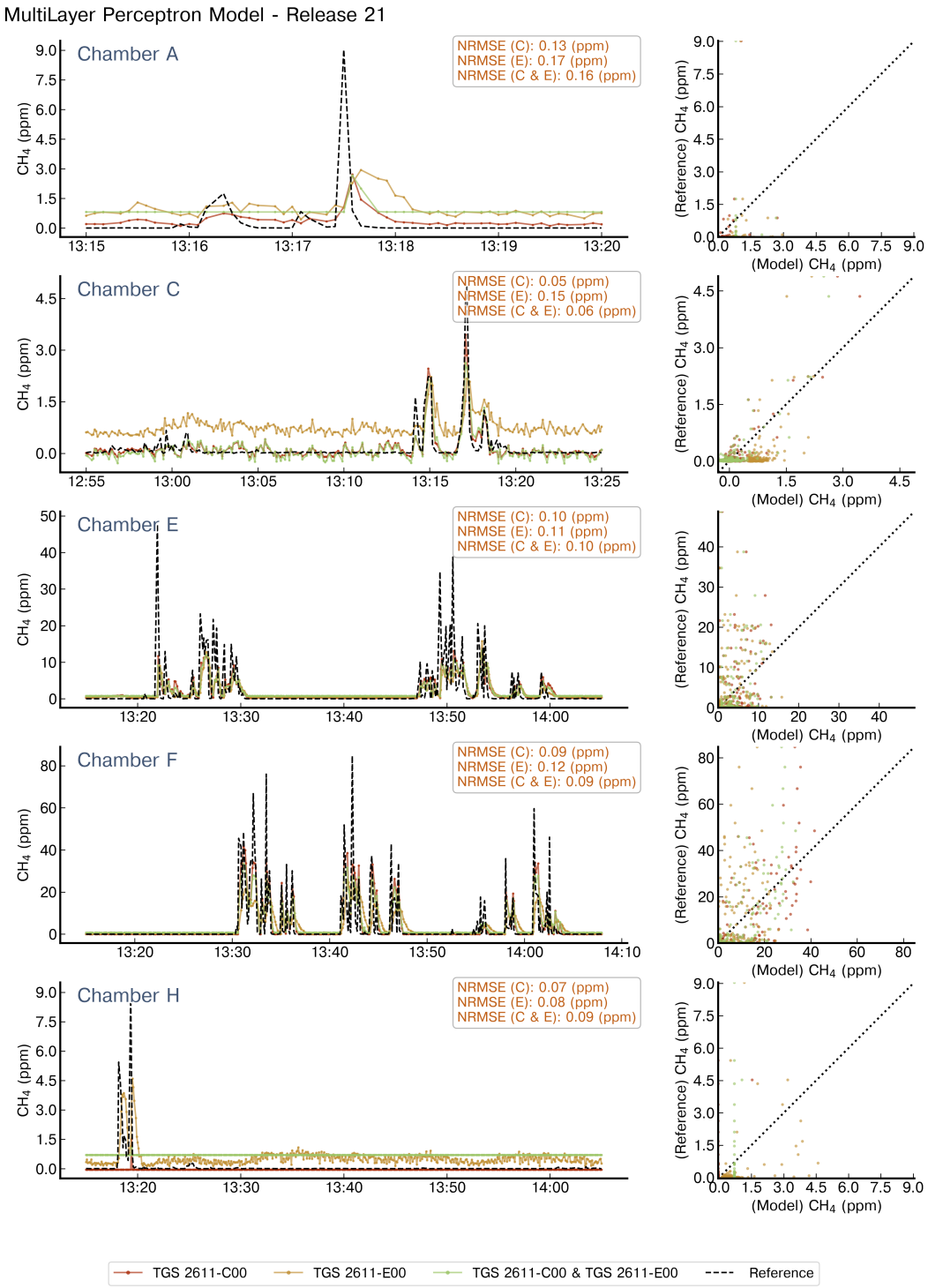


Figure C.8: Reconstruction of release #21 using a MLP model. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

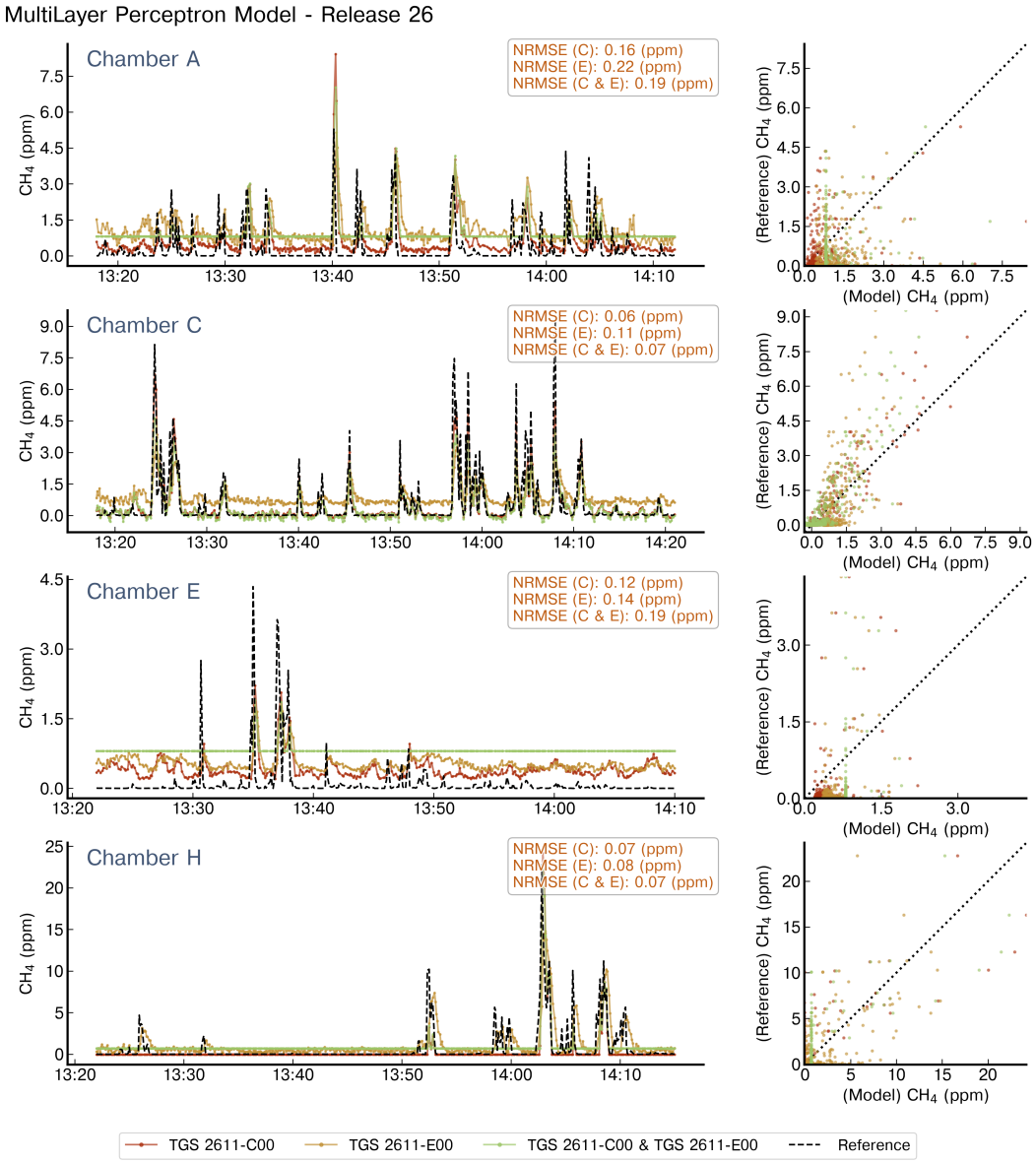


Figure C.9: Reconstruction of release #26 using a MLP model. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

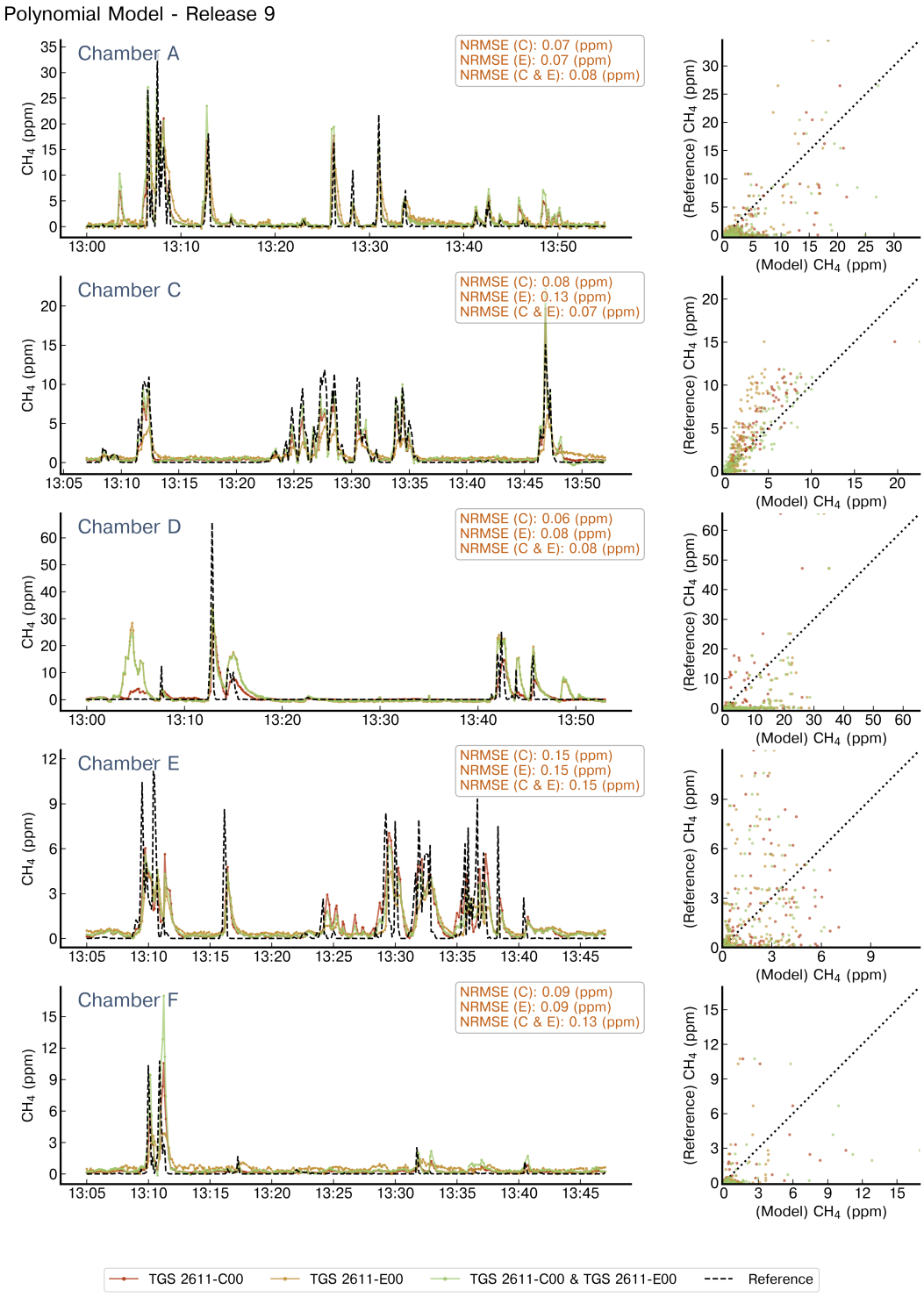


Figure C.10: Reconstruction of release #9 using 2nd degree polynomials. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

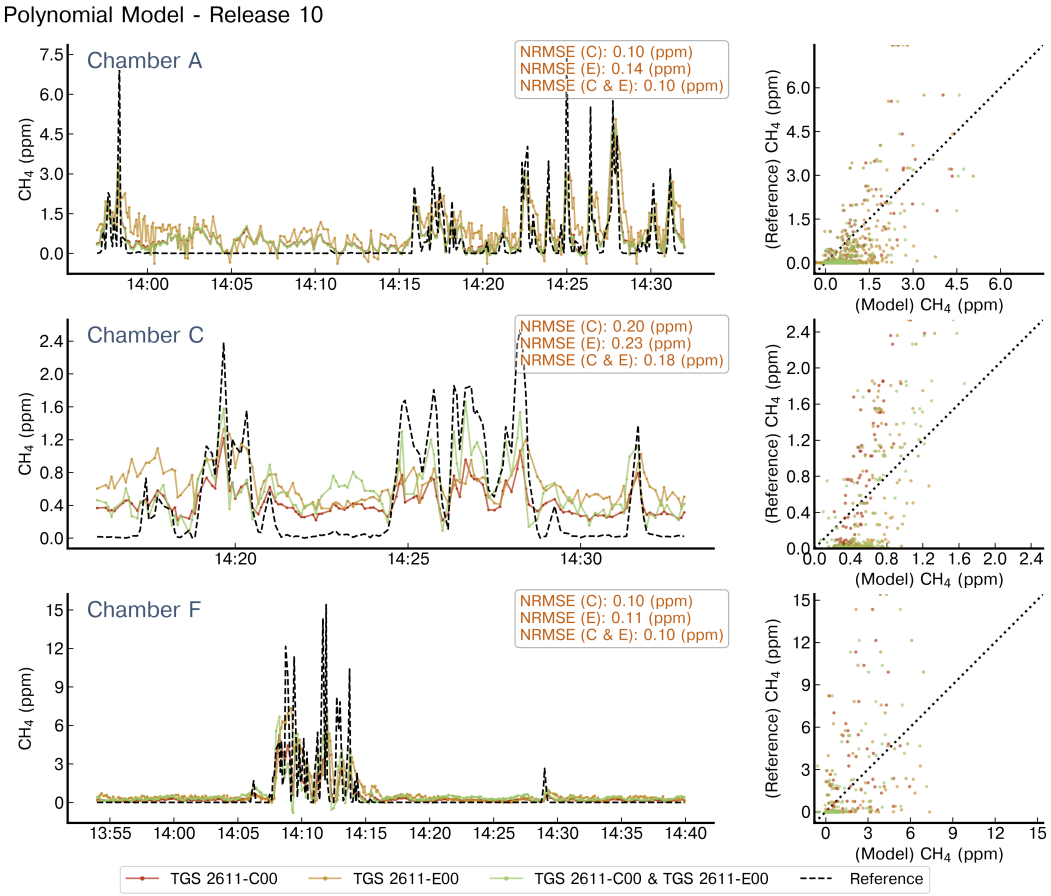


Figure C.11: Reconstruction of release #10 using 2nd degree polynomials. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

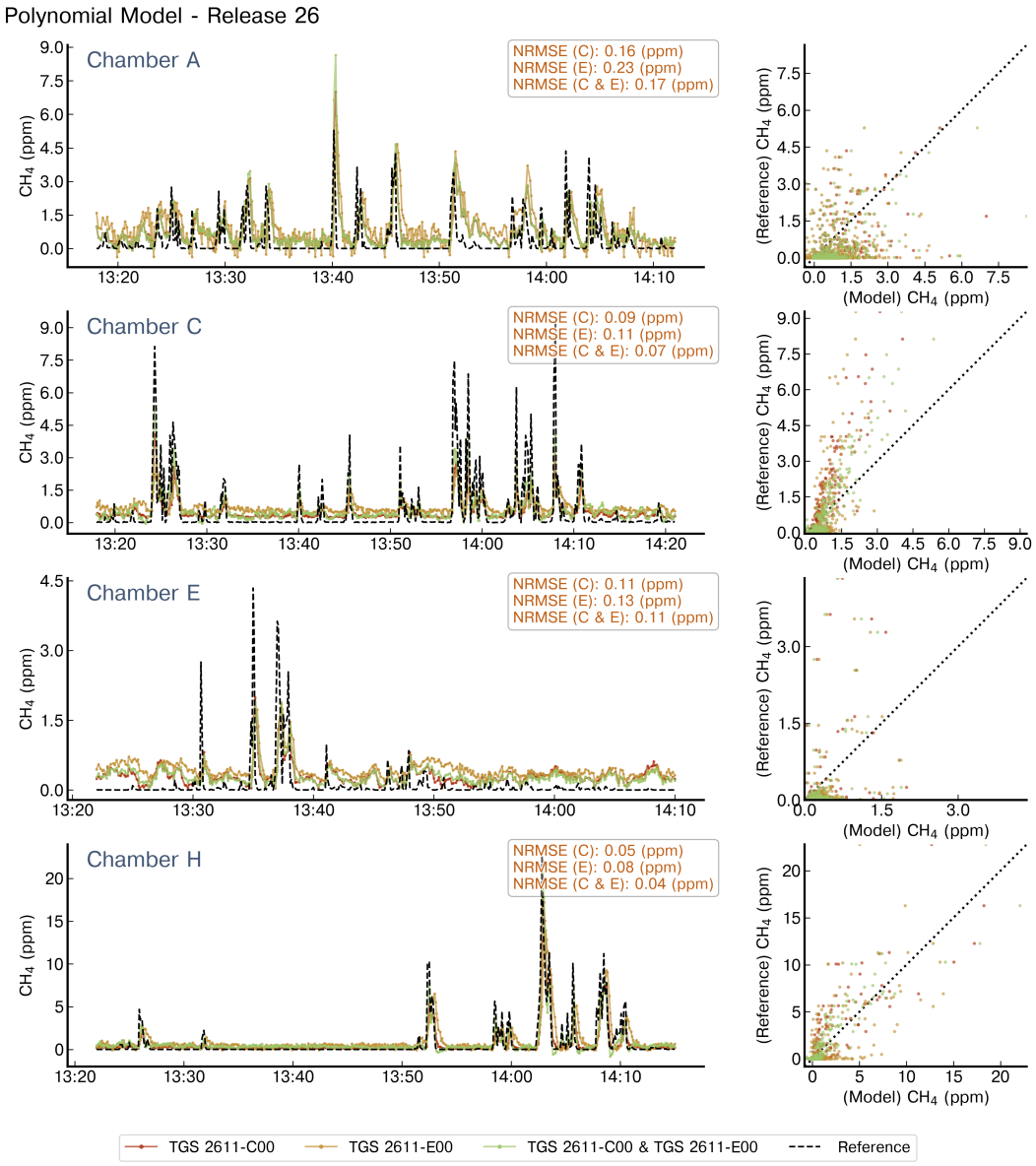


Figure C.12: Reconstruction of release #26 using 2nd degree polynomials. Notations are the same as in Figure C.4. Note the difference in the x-axis for the chambers.

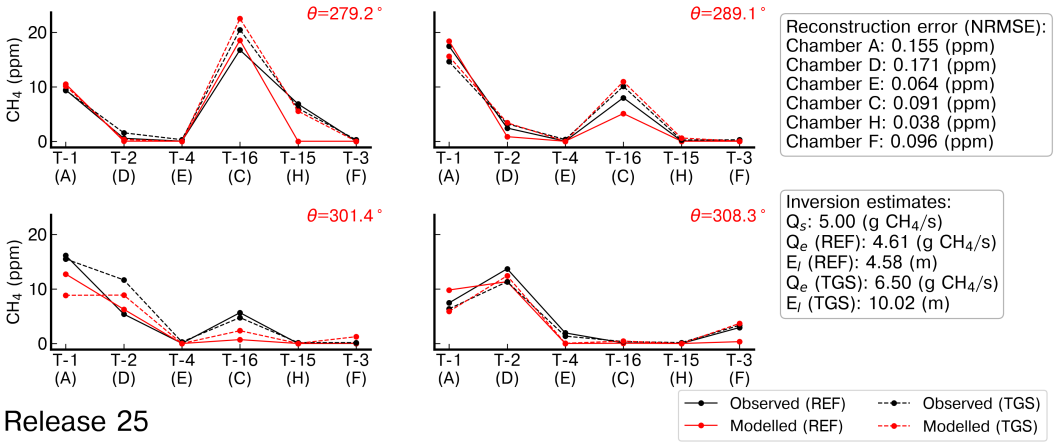


Figure C.13: Observed and modelled average CH_4 mole fractions from the reference, denoted 'REF', and low-cost sensor, denoted TGS, corresponding to the release #25. The reconstructed CH_4 was computed using the polynomial model of 2nd degree. The index of the tripods is denoted as T-x and the average wind direction (θ) for the binning of wind sectors is shown on the top right of each panel in red.

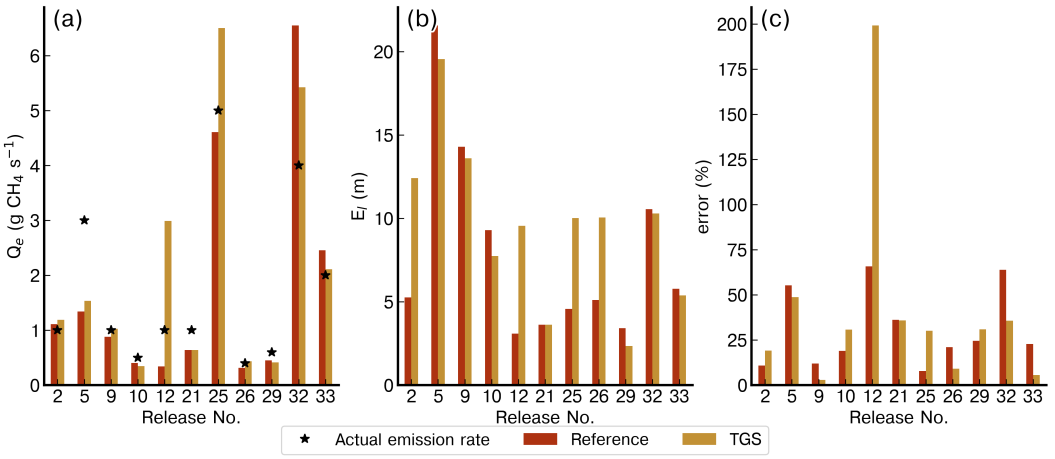


Figure C.14: Comparison of the emission rate estimates (*Qe*), location error (*El*) and relative error on the rate estimates for the inversions assimilating the reference data and the reconstruction of the CH_4 from the TGS low-cost sensor based on the Polynomial model of 2nd degree.

Bibliography

T. Aldhafeeri, M.-K. Tran, R. Vrolyk, M. Pope, and M. Fowler. A review of methane gas detection sensors: Recent developments and future perspectives. *Inventions*, 5:28, 7 2020. ISSN 2411-5134. doi: 10.3390/inventions5030028.

R. A. Alvarez, D. Zavala-Araiza, D. R. Lyon, D. T. Allen, Z. R. Barkley, A. R. Brandt, K. J. Davis, S. C. Herndon, D. J. Jacob, A. Karion, E. A. Kort, B. K. Lamb, T. Lauvaux, J. D. Maasakkers, A. J. Marchese, M. Omara, S. W. Pacala, J. Peischl, A. L. Robinson, P. B. Shepson, C. Sweeney, A. Townsend-Small, S. C. Wofsy, and S. P. Hamburg. Assessment of methane emissions from the u.s. oil and gas supply chain. *Science*, page eaar7204, 6 2018. ISSN 0036-8075. doi: 10.1126/science.aar7204.

P. A. Arias, N. Bellouin, E. Coppola, R. G. Jones, G. Krinner, J. Marotzke, V. Naik, M. D. Palmer, G.-K. Plattner, J. Rogelj, M. Rojas, J. Sillmann, T. Storelvmo, P. W. Thorne, B. Trewin, K. A. Rao, B. Adhikary, R. P. Allan, K. Armour, G. Bala, R. Barimalala, S. Berger, J. G. Canadell, C. Cassou, A. Cherchi, W. Collins, W. D. Collins, S. L. Connors, S. Corti, F. Cruz, F. J. Dentener, C. Dereczynski, A. D. Luca, A. D. Niang, F. J. Doblas-Reyes, A. Dosio, H. Douville, F. Engelbrecht, V. Eyring, E. Fischer, P. Forster, B. Fox-Kemper, J. S. Fuglestvedt, J. C. Fyfe, N. P. Gillett, L. Goldfarb, I. Gorodetskaya, J. M. Gutierrez, R. Hamdi, E. Hawkins, H. T. Hewitt, P. Hope, A. S. Islam, C. Jones, D. S. Kaufman, R. E. Kopp, Y. Kosaka, J. Kossin, S. Krakovska, J.-Y. Lee, J. Li, T. Mauritsen, T. K. Maycock, M. Meinshausen, S.-K. Min, P. M. S. Monteiro, T. Ngo-Duc, F. Otto, I. Pinto, A. Pirani, K. Raghavan, R. Ranasinghe, A. C. Ruane, L. Ruiz, J.-B. Sallée, B. H. Samset, S. Sathyendranath, S. I. Seneviratne, A. A. Sörensson, S. Szopa, I. Takayabu, A.-M. Tréguier, B. van den Hurk, R. Vautard, K. von Schuckmann, S. Zaehle, X. Zhang, and K. Zickfeld. *Technical Summary*. Cambridge University Press, 2021. doi: 10.1017/9781009157896.002.

S. Ars, G. Broquet, C. Yver Kwok, Y. Roustan, L. Wu, E. Arzoumanian, and P. Bousquet. Statistical atmospheric inversion of local gas emissions by coupling the tracer release technique and local-scale transport modelling: a test case with controlled methane emissions. *Atmospheric Measurement Techniques*, 10(12):5017–5037, 2017. doi: 10.5194/amt-10-5017-2017. URL https://amt.copernicus.org/articles/10/5017/2017/.

E. Arzoumanian, F. R. Vogel, A. Bastos, B. Gaynullin, O. Laurent, M. Ramonet, and P. Ciais. Characterization of a commercial lower-cost medium-precision non-dispersive infrared sensor for atmospheric co2 monitoring in urban areas. *Atmospheric Measurement Techniques*, 12:2665–2677, 5 2019. ISSN 1867-8548. doi: 10.5194/amt-12-2665-2019.

N. Barsan, D. Koziej, and U. Weimar. Metal oxide-based gas sensor research: How to? *Sensors and Actuators B: Chemical*, 121:18–35, 1 2007. ISSN 09254005. doi: 10.1016/j.snb.2006.09.047. URL https://linkinghub.elsevier.com/retrieve/pii/S0925400506006204.

D. Bastviken, J. Nygren, J. Schenk, R. Parellada Massana, and N. T. Duc. Technical note: Facilitating the use of low-cost methane (ch_4) sensors in flux chambers – calibration, data processing, and an open-source make-it-yourself logger. *Biogeosciences*, 17(13):3659–3667, 2020. doi: 10.5194/bg-17-3659-2020. URL https://bg.copernicus.org/articles/17/3659/2020/.

N. Bellouin, W. Davies, K. P. Shine, J. Quaas, J. Mülmenstädt, P. M. Forster, C. Smith, L. Lee, L. Regayre, G. Brasseur, N. Sudarchikova, I. Bouarar, O. Boucher, and G. Myhre. Radiative forcing of climate change from the copernicus reanalysis of atmospheric composition. *Earth System Science Data*, 12:1649–1677, 7 2020. ISSN 1866-3516. doi: 10.5194/essd-12-1649-2020. URL https://essd.copernicus.org/articles/12/1649/2020/.

P. Bergamaschi, U. Karstens, A. J. Manning, M. Saunois, A. Tsuruta, A. Berchet, A. T. Vermeulen, T. Arnold, G. Janssens-Maenhout, S. Hammer, I. Levin, M. Schmidt, M. Ramonet, M. Lopez, J. Lavric, T. Aalto, H. Chen, D. G. Feist, C. Gerbig, L. Haszpra, O. Hermansen, G. Manca, J. Moncrieff, F. Meinhardt, J. Necki, M. Galkowski, S. O'Doherty, N. Paramonova, H. A. Scheeren, M. Steinbacher, and E. Dlugokencky. Inverse modelling of european ch4

emissions during 2006–2012 using different inverse models and reassessed atmospheric observations. *Atmospheric Chemistry and Physics*, 18:901–920, 1 2018. ISSN 1680-7324. doi: 10.5194/acp-18-901-2018.

T. A. Berhanu, E. Satar, R. Schanda, P. Nyfeler, H. Moret, D. Brunner, B. Oney, and M. Leuenberger. Measurements of greenhouse gases at beromünster tall-tower station in switzerland. *Atmospheric Measurement Techniques*, 9(6):2603–2614, 2016. doi: 10.5194/amt-9-2603-2016. URL https://amt.copernicus.org/articles/9/2603/2016/.

C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, Inc., 1995a. ISBN 0198538642.

C. M. Bishop. Neural Networks for Pattern Recognition CLARENDON PRESS • OXFORD 1995. 1995b. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.679.1104&rep=rep1&type=pdf.

L. Breiman. Random forests. *Machine Learning*, 45, 2001. ISSN 08856125. doi: 10.1023/A:1010933404324.

J. G. Casey, A. Collier-Oxandale, and M. Hannigan. Performance of artificial neural networks and linear models to quantify 4 trace gas species in an oil and gas production region with low-cost sensors. *Sensors and Actuators B: Chemical*, 283, 3 2019. ISSN 09254005. doi: 10.1016/j.snb.2018.12.049.

A. Cescatti, B. Marcolla, I. Goded, and C. Gruening. Optimal use of buffer volumes for the measurement of atmospheric gas concentration in multi-point systems. *Atmospheric Measurement Techniques*, 9:4665–4672, 9 2016. doi: 10.5194/amt-9-4665-2016.

A. Chaiyboun, R. Traute, T. Haas, O. Kiesewetter, and T. Doll. A logarithmic multi-parameter model using gas sensor main and cross sensitivities to estimate gas concentrations in a gas mixture for sno2 gas sensors. *Sensors and Actuators B: Chemical*, 123(2):1064–1070, 2007. ISSN 0925-4005. doi: https://doi.org/10.1016/j.snb.2006.11.012. URL https://www.sciencedirect.com/science/article/pii/S0925400506007623.

J.-F. Chamberland and V. Veeravalli. How dense should a sensor network be for detection with correlated observations? *IEEE Transactions on Information Theory*, 52:5099–5106, 11 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.883551.

S. Coburn, C. B. Alden, R. Wright, K. Cossel, E. Baumann, G.-W. Truong, F. Giorgetta, C. Sweeney, N. R. Newbury, K. Prasad, I. Coddington, and G. B. Rieker. Regional trace-gas source attribution using a field-deployed dual frequency comb spectrometer. *Optica*, 5, 4 2018. ISSN 2334-2536. doi: 10.1364/OPTICA.5.000320.

A. Collier-Oxandale, J. G. Casey, R. Piedrahita, J. Ortega, H. Halliday, J. Johnston, and M. P. Hannigan. Assessing a low-cost methane sensor quantification system for use in complex rural and urban environments. *Atmospheric Measurement Techniques*, 11, 6 2018. ISSN 1867-8548. doi: 10.5194/amt-11-3569-2018.

A. Collier-Oxandale, N. Wong, S. Navarro, J. Johnston, and M. Hannigan. Using gas-phase air quality sensors to disentangle potential sources in a los angeles neighborhood. *Atmospheric Environment*, 233, 7 2020. ISSN 13522310. doi: 10.1016/j.atmosenv.2020.117519.

A. M. Collier-Oxandale, J. Thorson, H. Halliday, J. Milford, and M. Hannigan. Understanding the ability of low-cost mox sensors to quantify ambient vocs. *Atmospheric Measurement Techniques*, 12, 3 2019. ISSN 1867-8548. doi: 10.5194/amt-12-1441-2019.

K. R. Coombes, H. A. J. Fritsche, C. Clarke, J.-N. Chen, K. A. Baggerly, J. S. Morris, L.-C. Xiao, M.-C. Hung, and H. M. Kuerer. Quality control and peak finding for proteomics data collected from nipple aspirate fluid by surface-enhanced laser desorption and ionization. *Clinical chemistry*, 49:1615–1623, 10 2003. ISSN 0009-9147 (Print). doi: 10.1373/49.10.1615.

J. M. Cordero, R. Borge, and A. Narros. Using statistical methods to carry out in field calibrations of low cost air quality sensors. *Sensors and Actuators B: Chemical*, 267, 8 2018. ISSN 09254005. doi: 10.1016/j.snb.2018.04.021.

E. Crosson. A cavity ring-down analyzer for measuring atmospheric levels of methane, carbon

dioxide, and water vapor. *Applied Physics B*, 92:403–408, 9 2008. ISSN 0946-2171. doi: 10.1007/s00340-008-3135-y.

G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals, Syst.*, 2(4):303–314, dec 1989. ISSN 0932-4194. doi: 10.1007/BF02551274. URL http://link.springer.com/10.1007/BF02551274.

H. B. Demuth, M. H. Beale, O. D. Jess, and M. T. Hagan. *Neural Network Design*. Martin Hagan, 2nd edition, 2014. ISBN 0971732116.

J. Devkota, P. Ohodnicki, and D. Greve. Saw sensors for chemical vapors and gases. *Sensors*, 17: 801, 4 2017. ISSN 1424-8220. doi: 10.3390/s17040801.

R. M. Duren, A. K. Thorpe, K. T. Foster, T. Rafiq, F. M. Hopkins, V. Yadav, B. D. Bue, D. R. Thompson, S. Conley, N. K. Colombi, C. Frankenberg, I. B. McCubbin, M. L. Eastwood, M. Falk, J. D. Herner, B. E. Croes, R. O. Green, and C. E. Miller. California's methane super-emitters. *Nature*, 575: 180–184, 11 2019. ISSN 0028-0836. doi: 10.1038/s41586-019-1720-3.

E. Esposito, S. De Vito, M. Salvato, V. Bright, R. L. Jones, and O. Popoola. Dynamic neural network architectures for on field stochastic calibration of indicative low cost air quality sensing systems. *Sensors Actuators, B Chem.*, 231:701–713, 2016. ISSN 09254005. doi: 10.1016/j.snb.2016.03.038. URL http://dx.doi.org/10.1016/j.snb.2016.03.038.

W. Eugster and G. W. Kling. Performance of a low-cost methane sensor for ambient concentration measurements in preliminary studies. *Atmospheric Measurement Techniques*, 5, 8 2012. ISSN 1867-8548. doi: 10.5194/amt-5-1925-2012.

W. Eugster, J. Laundre, J. Eugster, and G. W. Kling. Long-term reliability of the figaro tgs 2600 solid-state methane sensor under low-arctic conditions at toolik lake, alaska. *Atmospheric Measurement Techniques*, 13, 5 2020. ISSN 1867-8548. doi: 10.5194/amt-13-2681-2020.

V. Eyring, N. P. Gillett, K. M. A. Rao, R. Barimalala, M. B. Parrillo, N. Bellouin, C. Cassou, P. J. Durack,

Y. Kosaka, S. McGregor, S. Min, O. Morgenstern, and Y. Sun. *Human Influence on the Climate System*. Cambridge University Press, 2021. doi: 10.1017/9781009157896.005.

A. Feitz, I. Schroder, F. Phillips, T. Coates, K. Negandhi, S. Day, A. Luhar, S. Bhatia, G. Edwards, S. Hrabar, E. Hernandez, B. Wood, T. Naylor, M. Kennedy, M. Hamilton, M. Hatch, J. Malos, M. Kochanek, P. Reid, J. Wilson, N. Deutscher, S. Zegelin, R. Vincent, S. White, C. Ong, S. George, P. Maas, S. Towner, N. Wokker, and D. Griffith. The ginninderra ch4 and co2 release experiment: An evaluation of gas detection and quantification techniques. *International Journal of Greenhouse Gas Control*, 70:202–224, 2018. ISSN 1750-5836. doi: https://doi.org/10.1016/j.ijggc.2017.11.018. URL https://www.sciencedirect.com/science/article/pii/S1750583617306862.

Figaro®. Figaro (2005), 2005. URL https://www.figaro.co.jp/en/product/entry/tgs2600.html. accessed on 10 February 2020 accessed on 10 February 2020.

Figaro®. Figaro (2013), 2013. URL https://www.figaro.co.jp/en/product/entry/tgs2611-c00.html. accessed on 10 February 2020.

P. Forster, T. Storelvmo, K. Armour, W. Collins, J.-L. Dufresne, D. Frame, D. J. Lunt, T. Mauritsen, M. D. Palmer, M. Watanabe, M. Wild, and H. Zhang. *The Earth's Energy Budget, Climate Feedbacks, and Climate Sensitivity*. Cambridge University Press, 2021. doi: 10.1017/9781009157896.009.

M. W. Gardner and S. R. Dorling. Artificial neural networks (the multilayer perceptron) - a review of applications in the atmospheric sciences. *Atmos. Environ.*, 32(14-15):2627–2636, 1998. ISSN 13522310. doi: 10.1016/S1352-2310(97)00447-0.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

A. Géron. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019.

M. T. Hagan, H. B. Demuth, M. H. Beale, and O. De Jesús. *Neural Network Design*. 2014. ISBN 9780971732117.

J. Hansen, M. Sato, and R. Ruedy. Radiative forcing and climate response. *Journal of Geophysical Research: Atmospheres*, 102:6831–6864, 3 1997. ISSN 01480227. doi: 10.1029/96JD03436. URL http://doi.wiley.com/10.1029/96JD03436.

J. E. Hansen and M. Sato. Trends of measured climate forcing agents. *Proceedings of the National Academy of Sciences*, 98:14778–14783, 12 2001. ISSN 0027-8424. doi: 10.1073/pnas.261553698.

S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.

S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, 2nd edition, 1998. ISBN 0132733501.

I. Heimann, V. Bright, M. McLeod, M. Mead, O. Popoola, G. Stewart, and R. Jones. Source attribution of air pollution by spatial scale separation using high spatial density networks of low cost air quality sensors. *Atmospheric Environment*, 113, 7 2015. ISSN 13522310. doi: 10.1016/j.atmosenv.2015.04.057.

T. Hong, J. T. Culp, K.-J. Kim, J. Devkota, C. Sun, and P. R. Ohodnicki. State-of-the-art of methane sensing materials: A review and perspectives. *TrAC Trends in Analytical Chemistry*, 125:115820, 4 2020. ISSN 01659936. doi: 10.1016/j.trac.2020.115820.

K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2, 1 1989. ISSN 08936080. doi: 10.1016/0893-6080(89)90020-8.

R. Ionescu, A. Vancu, C. Moise, and A. Tomescu. Role of water vapour in the interaction of sno2 gas sensors with co and ch4. *Sensors and Actuators B: Chemical*, 61:39–42, 12 1999. ISSN 09254005. doi: 10.1016/S0925-4005(99)00277-4.

M. jie Yin, B. Gu, Q.-F. An, C. Yang, Y. L. Guan, and K.-T. Yong. Recent development of fiber-optic chemical sensors and biosensors: Mechanisms, materials, micro/nano-fabrications and applications. *Coordination Chemistry Reviews*, 376:348–392, 12 2018. ISSN 00108545. doi: 10.1016/j.ccr.2018.08.001.

C. J. Jørgensen, J. Mønster, K. Fuglsang, and J. R. Christiansen. Continuous methane concentration measurements at the greenland ice sheet–atmosphere interface using a low-cost, low-power metal oxide sensor system. *Atmospheric Measurement Techniques*, 13, 6 2020. ISSN 1867-8548. doi: 10.5194/amt-13-3319-2020.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

K. Kobayashi and M. U. Salam. Comparing simulated and measured values using mean squared deviation and its components. *Agronomy Journal*, 92, 3 2000. ISSN 0002-1962. doi: 10.2134/agronj2000.922345x.

A. Kolmakov, D. O. Klenov, Y. Lilach, S. Stemmer, and M. Moskovits. Enhanced gas sensing by individual sno ₂ nanowires and nanobelts functionalized with pd catalyst particles. *Nano Letters*, 5:667–673, 4 2005. ISSN 1530-6984. doi: 10.1021/nl050082v.

P. Kumar, L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. D. Sabatino, M. Bell, L. Norford, and R. Britter. The rise of low-cost sensing for managing air pollution in cities. *Environment International*, 75, 2 2015. ISSN 01604120. doi: 10.1016/j.envint.2014.11.019.

P. Kumar, G. Broquet, C. Yver-Kwok, O. Laurent, S. Gichuki, C. Caldow, F. Cropley, T. Lauvaux, M. Ramonet, G. Berthe, F. Martin, O. Duclaux, C. Juery, C. Bouchet, and P. Ciais. Mobile atmospheric measurements and local-scale inverse estimation of the location and rates of brief ch$_{4}$ and co$_{2}$ releases from point sources. *Atmospheric Measurement Techniques*, 14:5987–6003, 2021. doi: 10.5194/amt-14-5987-2021. URL https://amt.copernicus.org/articles/14/5987/2021/.

P. Kumar, G. Broquet, C. Caldow, O. Laurent, S. Gichuki, F. Cropley, C. Yver-Kwok, B. Fontanier, T. Lauvaux, M. Ramonet, A. Shah, G. Berthe, F. Martin, O. Duclaux, C. Juery, C. Bouchet, J. Pitt, and P. Ciais. Near-field atmospheric inversions for the localization and quantification of controlled methane releases using stationary and mobile measurements. *Quarterly Jour-*

nal of the Royal Meteorological Society, 4 2022. ISSN 0035-9009. doi: 10.1002/qj.4283. URL https://onlinelibrary.wiley.com/doi/10.1002/qj.4283.

C. Malings, R. Tanzer, A. Hauryliuk, S. P. N. Kumar, N. Zimmerman, L. B. Kara, and A. A. Presto. Development of a general calibration model and long-term performance evaluation of low-cost sensors for air pollutant gas monitoring. *Atmospheric Measurement Techniques*, 12:903–920, 2 2019. doi: 10.5194/amt-12-903-2019.

M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayes, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. Saffell, and R. Jones. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmospheric Environment*, 70:186–203, 5 2013. ISSN 13522310. doi: 10.1016/j.atmosenv.2012.11.060.

J. F. B. Mitchell. The "greenhouse" effect and climate change. *Reviews of Geophysics*, 27:115, 1989. ISSN 8755-1209. doi: 10.1029/RG027i001p00115. URL http://doi.wiley.com/10.1029/RG027i001p00115.

O. Morgenstern, M. I. Hegglin, E. Rozanov, F. M. O'Connor, N. L. Abraham, H. Akiyoshi, A. T. Archibald, S. Bekki, N. Butchart, M. P. Chipperfield, M. Deushi, S. S. Dhomse, R. R. Garcia, S. C. Hardiman, L. W. Horowitz, P. Jöckel, B. Josse, D. Kinnison, M. Lin, E. Mancini, M. E. Manyin, M. Marchand, V. Marécal, M. Michou, L. D. Oman, G. Pitari, D. A. Plummer, L. E. Revell, D. Saint-Martin, R. Schofield, A. Stenke, K. Stone, K. Sudo, T. Y. Tanaka, S. Tilmes, Y. Yamashita, K. Yoshida, and G. Zeng. Review of the global models used within phase 1 of the chemistry–climate model initiative (ccmi). *Geoscientific Model Development*, 10:639–671, 2 2017. ISSN 1991-9603. doi: 10.5194/gmd-10-639-2017.

J. Mønster, P. Kjeldsen, and C. Scheutz. Methodologies for measuring fugitive methane emissions from landfills – a review. *Waste Management*, 87:835–859, 3 2019. ISSN 0956053X. doi: 10.1016/j.wasman.2018.12.047.

E. G. Nisbet, R. E. Fisher, D. Lowry, J. L. France, G. Allen, S. Bakkaloglu, T. J. Broderick, M. Cain, M. Coleman, J. Fernandez, G. Forster, P. T. Griffiths, C. P. Iverach, B. F. J. Kelly, M. R. Man-

ning, P. B. R. Nisbet-Jones, J. A. Pyle, A. Townsend-Small, A. Al-Shalaan, N. Warwick, and G. Zazzeri. Methane mitigation: Methods to reduce emissions, on the path to the paris agreement. *Reviews of Geophysics*, 58, 3 2020. ISSN 8755-1209. doi: 10.1029/2019RG000675. URL https://onlinelibrary.wiley.com/doi/abs/10.1029/2019RG000675.

K. Okorn and M. Hannigan. Improving air pollutant metal oxide sensor quantification practices through: An exploration of sensor signal normalization, multi-sensor and universal calibration model generation, and physical factors such as co-location duration and sensor age. *Atmosphere*, 12:645, 5 2021. ISSN 2073-4433. doi: 10.3390/atmos12050645.

Picarro Inc.: Santa Clara, CA, USA.

I. Pison, A. Berchet, M. Saunois, P. Bousquet, G. Broquet, S. Conil, M. Delmotte, A. Ganesan, O. Laurent, D. Martin, S. O'Doherty, M. Ramonet, T. G. Spain, A. Vermeulen, and C. Y. Kwok. How a European network may help with estimating methane emissions on the French national scale. *Atmos. Chem. Phys.*, 18(5):3779–3798, 2018. ISSN 16807324. doi: 10.5194/acp-18-3779-2018.

O. A. Popoola, D. Carruthers, C. Lad, V. B. Bright, M. I. Mead, M. E. Stettler, J. R. Saffell, and R. L. Jones. Use of networks of low cost air quality sensors to quantify air quality in urban settings. *Atmospheric Environment*, 194, 12 2018. ISSN 13522310. doi: 10.1016/j.atmosenv.2018.09.030.

W. H. Press and S. A. Teukolsky. Savitzky-golay smoothing filters. *Computers in Physics*, 4:669–672, 1990.

V. Ramanswamy, K. Shine, C. Leovy, W.-C. Wang, H. Rodhe, D. J. Wuebbles, M. Ding, J. Lelieveld, J. A. Edmonds, and M. P. Mccormick. Radiative forcing of climate. *NASA, Washington, Scientific Assessment of Ozone Depletion: 1991*, 1991.

C. W. Rella, H. Chen, A. E. Andrews, A. Filges, C. Gerbig, J. Hatakka, A. Karion, N. L. Miles, S. J. Richardson, M. Steinbacher, C. Sweeney, B. Wastine, and C. Zellweger. High accuracy measurements of dry mole fractions of carbon dioxide and methane in humid air. *Atmospheric Measurement Techniques*, 6:837–860, 3 2013. ISSN 1867-8548. doi: 10.5194/amt-6-837-2013.

S. N. Riddick, D. L. Mauzerall, M. Celia, G. Allen, J. Pitt, M. Kang, and J. C. Riddick. The calibration and deployment of a low-cost methane sensor. *Atmospheric Environment*, 230:117440, 6 2020. ISSN 13522310. doi: 10.1016/j.atmosenv.2020.117440.

R. Rivera Martinez, D. Santaren, O. Laurent, F. Cropley, C. Mallet, M. Ramonet, C. Caldow, L. Rivier, G. Broquet, C. Bouchet, C. Juery, and P. Ciais. The potential of low-cost tin-oxide sensors combined with machine learning for estimating atmospheric ch4 variations around background concentration. *Atmosphere*, 12, 1 2021. ISSN 2073-4433. doi: 10.3390/atmos12010107.

R. A. Rivera Martinez, D. Santaren, O. Laurent, G. Broquet, F. Cropley, C. Mallet, M. Ramonet, A. Shah, L. Rivier, C. Bouchet, C. Juery, O. Duclaux, and P. Ciais. Reconstruction of high-frequency methane atmospheric concentration peaks from measurements using metal oxide low-cost sensors. *Atmospheric Measurement Techniques Discussions*, 2022:1–45, 2022. doi: 10.5194/amt-2022-200. URL https://amt.copernicus.org/preprints/amt-2022-200/.

A. Romain and J. Nicolas. Long term stability of metal oxide-based gas sensors for e-nose environmental applications: An overview. *Sensors and Actuators B: Chemical*, 146:502–506, 4 2010. ISSN 09254005. doi: 10.1016/j.snb.2009.12.027.

A. F. Ruckstuhl, S. Henne, S. Reimann, M. Steinbacher, M. K. Vollmer, S. O'Doherty, B. Buchmann, and C. Hueglin. Robust extraction of baseline signal of atmospheric trace species using local regression. *Atmospheric Measurement Techniques*, 5, 11 2012. ISSN 1867-8548. doi: 10.5194/amt-5-2613-2012.

D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323, 10 1986. ISSN 0028-0836. doi: 10.1038/323533a0.

H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE transactions on acoustics, speech, and signal processing*, 26:43–49, 1978.

M. Saunois, P. Bousquet, B. Poulter, A. Peregon, P. Ciais, J. G. Canadell, E. J. Dlugokencky, G. Etiope, D. Bastviken, S. Houweling, G. Janssens-Maenhout, F. N. Tubiello, S. Castaldi, R. B. Jackson,

M. Alexe, V. K. Arora, D. J. Beerling, P. Bergamaschi, D. R. Blake, G. Brailsford, V. Brovkin, L. Bruhwiler, C. Crevoisier, P. Crill, K. Covey, C. Curry, C. Frankenberg, N. Gedney, L. Höglund-Isaksson, M. Ishizawa, A. Ito, F. Joos, H.-S. Kim, T. Kleinen, P. Krummel, J.-F. Lamarque, R. Langenfelds, R. Locatelli, T. Machida, S. Maksyutov, K. C. McDonald, J. Marshall, J. R. Melton, I. Morino, V. Naik, S. O'Doherty, F.-J. W. Parmentier, P. K. Patra, C. Peng, S. Peng, G. P. Peters, I. Pison, C. Prigent, R. Prinn, M. Ramonet, W. J. Riley, M. Saito, M. Santini, R. Schroeder, I. J. Simpson, R. Spahni, P. Steele, A. Takizawa, B. F. Thornton, H. Tian, Y. Tohjima, N. Viovy, A. Voulgarakis, M. van Weele, G. R. van der Werf, R. Weiss, C. Wiedinmyer, D. J. Wilton, A. Wiltshire, D. Worthy, D. Wunch, X. Xu, Y. Yoshida, B. Zhang, Z. Zhang, and Q. Zhu. The global methane budget 2000–2012. *Earth System Science Data*, 8, 12 2016. ISSN 1866-3516. doi: 10.5194/essd-8-697-2016.

M. Saunois, A. R. Stavert, B. Poulter, P. Bousquet, J. G. Canadell, R. B. Jackson, P. A. Raymond, E. J. Dlugokencky, S. Houweling, P. K. Patra, P. Ciais, V. K. Arora, D. Bastviken, P. Bergamaschi, D. R. Blake, G. Brailsford, L. Bruhwiler, K. M. Carlson, M. Carrol, S. Castaldi, N. Chandra, C. Crevoisier, P. M. Crill, K. Covey, C. L. Curry, G. Etiope, C. Frankenberg, N. Gedney, M. I. Hegglin, L. Höglund-Isaksson, G. Hugelius, M. Ishizawa, A. Ito, G. Janssens-Maenhout, K. M. Jensen, F. Joos, T. Kleinen, P. B. Krummel, R. L. Langenfelds, G. G. Laruelle, L. Liu, T. Machida, S. Maksyutov, K. C. McDonald, J. McNorton, P. A. Miller, J. R. Melton, I. Morino, J. Müller, F. Murguia-Flores, V. Naik, Y. Niwa, S. Noce, S. O'Doherty, R. J. Parker, C. Peng, S. Peng, G. P. Peters, C. Prigent, R. Prinn, M. Ramonet, P. Regnier, W. J. Riley, J. A. Rosentreter, A. Segers, I. J. Simpson, H. Shi, S. J. Smith, L. P. Steele, B. F. Thornton, H. Tian, Y. Tohjima, F. N. Tubiello, A. Tsuruta, N. Viovy, A. Voulgarakis, T. S. Weber, M. van Weele, G. R. van der Werf, R. F. Weiss, D. Worthy, D. Wunch, Y. Yin, Y. Yoshida, W. Zhang, Z. Zhang, Y. Zhao, B. Zheng, Q. Zhu, Q. Zhu, and Q. Zhuang. The global methane budget 2000–2017. *Earth System Science Data*, 12:1561–1623, 7 2020. ISSN 1866-3516. doi: 10.5194/essd-12-1561-2020.

H. Scharff, A. Martha, D. M. M. V. Rijn, A. Hensen, C. Flechard, H. Oonk, R. Vroon, A. de Visscher, and P. Boeckx. A comparison of measurement methods to determine landfill methane emissions. *Afvalzorg Deponie BV, Dutch ministry of Housing, spatial Planning and Environment*, 2003.

A. Shahid, J. H. Choi, A. U. H. S. Rana, and H. S. Kim. Least squares neural network-based wireless E-nose system using an SnO2 sensor array. *Sensors (Switzerland)*, 18(5), 2018. ISSN 14248220. doi: 10.3390/s18051446.

L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola. Field calibration of a cluster of low-cost available sensors for air quality monitoring. part a: Ozone and nitrogen dioxide. *Sensors and Actuators B: Chemical*, 215, 8 2015. ISSN 09254005. doi: 10.1016/j.snb.2015.03.031.

L. Spinelle, M. Gerboles, M. G. Villani, M. Aleixandre, and F. Bonavitacola. Field calibration of a cluster of low-cost commercially available sensors for air quality monitoring. part b: No, co and co2. *Sensors and Actuators B: Chemical*, 238, 1 2017. ISSN 09254005. doi: 10.1016/j.snb.2016.07.036.

A. Tischner, T. Maier, C. Stepper, and A. Köck. Ultrathin sno2 gas sensors fabricated by spray pyrolysis for the detection of humidity and carbon monoxide. *Sensors and Actuators B: Chemical*, 134:796–802, 9 2008. ISSN 09254005. doi: 10.1016/j.snb.2008.06.032.

B. Travis, M. Dubey, and J. Sauer. Neural networks to locate and quantify fugitive natural gas leaks for a mir detection system. *Atmospheric Environment: X*, 8:100092, 12 2020. ISSN 25901621. doi: 10.1016/j.aeaoa.2020.100092.

A. J. Turner, A. A. Shusterman, B. C. McDonald, V. Teige, R. A. Harley, and R. C. Cohen. Network design for quantifying urban co2; emissions: assessing trade-offs between precision and network density. *Atmospheric Chemistry and Physics*, 16:13465–13475, 11 2016. ISSN 1680-7324. doi: 10.5194/acp-16-13465-2016.

A. J. Turner, I. Fung, V. Naik, L. W. Horowitz, and R. C. Cohen. Modulation of hydroxyl variability by enso in the absence of external forcing. *Proceedings of the National Academy of Sciences*, 115: 8931–8936, 9 2018. ISSN 0027-8424. doi: 10.1073/pnas.1807532115.

A. J. Turner, C. Frankenberg, and E. A. Kort. Interpreting contemporary trends in atmospheric

methane. *Proceedings of the National Academy of Sciences*, 116:2805–2813, 2 2019. ISSN 0027-8424. doi: 10.1073/pnas.1814297116.

M. van den Bossche, N. T. Rose, and S. F. J. De Wekker. Potential of a low-cost gas sensor for atmospheric methane monitoring. *Sensors and Actuators B: Chemical*, 238:501–509, 2017. ISSN 0925-4005. doi: https://doi.org/10.1016/j.snb.2016.07.092. URL https://www.sciencedirect.com/science/article/pii/S0925400516311273.

G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller. Scikit-learn. *GetMobile Mob. Comput. Commun.*, 19(1):29–33, 2015. ISSN 23750529. doi: 10.1145/2786984.2786995.

C. Wang, L. Yin, L. Zhang, D. Xiang, and R. Gao. Metal oxide gas sensors: Sensitivity and influencing factors. *Sensors*, 10:2088–2106, 3 2010. ISSN 1424-8220. doi: 10.3390/s100302088.

M. D. Wheeler, S. M. Newman, A. J. Orr-Ewing, and M. N. R. Ashfold. Cavity ring-down spectroscopy. *Journal of the Chemical Society, Faraday Transactions*, 94:337–351, 1998. ISSN 09565000. doi: 10.1039/a707686j.

L. Wu, G. Broquet, P. Ciais, V. Bellassen, F. Vogel, F. Chevallier, I. Xueref-Remy, and Y. Wang. What would dense atmospheric observation networks bring to the quantification of city co2 emissions? *Atmospheric Chemistry and Physics*, 16:7743–7771, 6 2016. ISSN 1680-7324. doi: 10.5194/acp-16-7743-2016.

I. Xueref-Remy, G. Zazzeri, F. Bréon, F. Vogel, P. Ciais, D. Lowry, and E. Nisbet. Anthropogenic methane plume detection from point sources in the Paris megacity area and characterization of their δ13C signature. *Atmos. Environ.*, page 117055, 2019. ISSN 1352-2310. doi: 10.1016/j.atmosenv.2019.117055. URL https://doi.org/10.1016/j.atmosenv.2019.117055.

K. Yan and D. Zhang. Correcting instrumental variation and time-varying drift: A transfer learning approach with autoencoders. *IEEE Transactions on Instrumentation and Measurement*, 65(9): 2012–2022, 2016. doi: 10.1109/TIM.2016.2573078.

C. Yver-Kwok, O. Laurent, A. Guemri, C. Philippon, B. Wastine, C. W. Rella, C. Vuillemin, F. Truong, M. Delmotte, V. Kazan, M. Darding, B. Lebègue, C. Kaiser, I. Xueref-Rémy, and M. Ramonet. Comprehensive laboratory and field testing of cavity ring-down spectroscopy analyzers measuring h_2o, co_2, ch_4 and co. *Atmospheric Measurement Techniques*, 8:3867–3892, 9 2015. doi: 10.5194/amt-8-3867-2015.

Y. Zhao, S. Wang, W. Yuan, S. Fan, Z. Hua, Y. Wu, and X. Tian. Selective detection of methane by pd-in2o3 sensors with a catalyst filter film. *Sensors and Actuators B: Chemical*, 328:129030, 2021. ISSN 0925-4005. doi: https://doi.org/10.1016/j.snb.2020.129030. URL https://www.sciencedirect.com/science/article/pii/S0925400520313770.

Özgür Örnek and B. Karlik. An overview of metal oxide semiconducting sensors in electronic nose applications. volume 2, pages 506–515, 2012.