



Détection et classification des signaux sismo-volcaniques de la Soufrière de Guadeloupe par apprentissage supervisé

Alexis Falcin

► To cite this version:

Alexis Falcin. Détection et classification des signaux sismo-volcaniques de la Soufrière de Guadeloupe par apprentissage supervisé. Volcanologie. Université Paris Cité, 2022. Français. NNT : 2022UNIP7318 . tel-04520418

HAL Id: tel-04520418

<https://theses.hal.science/tel-04520418>

Submitted on 25 Mar 2024

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

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

Université Paris Cité

École doctorale STEP'UP - ED n°560

Institut de Physique du Globe de Paris - Équipe de sismologie

Détection et classification des signaux sismo-volcaniques de la Soufrière de Guadeloupe par apprentissage supervisé

Alexis Falcin

Thèse de doctorat de Sciences de la Terre et de l'Environnement

Dirigée par Jean-Philippe Métaxian
Et par Éléonore Stutzmann

Présentée et soutenue publiquement le 14 Avril 2022

Devant un jury composé de :

- **David Marsan**, *Professeur des Universités, ISTerre, Rapporteur*
- **Alessia Maggi**, *Professeur des Universités, UNISTRA, Rapporteur*
- **Jean-Paul Montagner**, *Professeur des Universités, IPGP, Examineur*
- **Marielle Malfante**, *PhD, CEA, Examinatrice*
- **Jérôme Mars**, *Professeur des Universités, GIPSA-LAB, Membre invité*
- **François Beauducel**, *Phycisien, IPGP, Membre invité*
- **Jean-Philippe Métaxian**, *HDR, IPGP, Directeur de thèse*
- **Éléonore Stutzmann**, *Phycisienne, IPGP, Co-directrice de thèse*

Remerciements

Je tiens à exprimer ma reconnaissance envers mes directeurs de thèse, Jean-Philippe Métaxian, Éléonore Stutzmann et Jérôme Mars. Je les remercie pour leur bienveillance, pour leur patience, pour l'aide qu'ils ont pu m'apporter au cours de ces trois années.

Je remercie toutes les personnes de l'équipe de sismologie et de volcanologie de l'IPGP ou d'ailleurs qui ont, à un moment ou à un autre, participé à ma réflexion. Notamment Jean-Christophe Komorowski, François Beauducel et Glenn Thompson qui ont montré un véritable enthousiasme quant à mon travail.

Je veux remercier tous ceux qui ont pu évaluer mon travail et donner des retours constructifs lors de cette thèse, les membres du comité de thèse, Claudio Satriano, Mauro Dalla-Mura, Clément Hibert et Jean-Christophe Komorowski ainsi que les membres du jury, Alessia Maggi, David Marsan, Marielle Malfante, Jean-Paul Montagner et François Beauducel.

Merci à mes collègues doctorants pour l'ambiance de travail très agréable, particulièrement Pablo Poulain qui aura été mon voisin de bureau du premier au dernier jour.

Merci également à Sylvie Contamina et Geneviève Molguinny pour leur aide précieuse. Merci à l'IPGP FC, pour ces matchs épiques.

Enfin je veux remercier ma famille et mes amis pour le soutien qu'ils ont pu m'apporter au cours de ces trois années.

Bonne continuation à tous.

Abstract

Seismic activity at La Soufrière volcano of Guadeloupe is composed of various transient signals, which are classified manually by the Observatoire Volcanologique et Sismologique de Guadeloupe (OVSG-IPGP) considering waveforms recorded at several stations. Three main classes readily distinguishable on seismic traces during the daily analytical protocol have been catalogued : Volcano-Tectonic events, Long-Period events and Nested events, each related to a distinct physical process. Automatic detection and classification of volcano-seismic signals of La Soufrière was performed by using an architecture based on supervised learning, available at github.com/malfante/AAA. Seismic waveforms are transformed into a large set of features (34 features for each representation domain) computed from three representation domains of the signal (time, frequency, quefrequency). The resulting vectors of features are then used for modeling. We are using the Random Forest Classifier algorithm from the scikit-learn library.

At first, we trained the model with the dataset given by the OVSG consisting of 845 available labeled events (542 VT, 217 nested and 86 LP) recorded in the period 2013-2018. We obtained an average classification rate of 72%. We determined that the VT class includes a variety of signals covering the LP, Nested and VT classes. After reviewing in detail waveforms and spectral characteristics of the signals belonging to the 3 classes we decided to introduce 2 new classes, Hybrid events and also a monochromatic class (so-called Tornillo) of LP signals, thus matching the full description of signals provided in Moretti et al. (2020).

Then, using the new information, a new model was trained with 5 new classes. We obtained a much better classification average rate of 84%. The classification is excellent for Nested events (93% of recall and precision) and Tornillo events (93% of recall and precision). The classification of VT events (90% recall, 89% precision) and LP events (86% recall, 82% precision) were also very good. The most difficult class to recognize is the Hybrid class (64% recall, 69% precision). Hybrid events are often mixed with VT and LP events. This may be explained by the nature of this class and the physical process that includes both a fracturing and a resonating component with different modal frequencies.

Moreover, by using a supervised machine learning model to distinguish the events from the background noise, we were able to detect three times more events as the observatory with an STA / LTA method.

Machine learning is a powerful tool to handle large datasets. We were able to improve the classification, correct some misclassifications and detect more events.

Keywords : Seismology, Volcano, La Soufrière, Classification, Detection, Machine Learning

Résumé

L'activité sismique du volcan de la Soufrière de Guadeloupe se compose de différents signaux transitoires qui sont classés manuellement par l'Observatoire Volcanologique et Sismologique de Guadeloupe (OVSG-IPGP) en visualisant les formes d'ondes enregistrées sur plusieurs stations. Trois principales classes d'événements facilement identifiables sur les traces sont considérés durant le protocole journalier de classification : les événements Volcano-Tectonique, les événements Longue-Période et les événements Emboîtés, chacune reliées à des processus physiques distincts.

La détection et la classification automatique des signaux sismo-volcaniques de la Soufrière ont été réalisées en utilisant une méthode basée sur l'apprentissage supervisé, disponible sur github.com/-malfante/AAA. Les formes d'ondes sismiques sont transformées en des jeux de descripteurs (34 pour chaque domaine de représentation) calculés à partir de trois domaines de représentation des signaux (temps, fréquence, quéfrencence). Le vecteur descripteur obtenu est alors utilisé pour construire le modèle. Nous utilisons le Random Forest Classifier disponible dans la bibliothèque scikit-learn.

Dans un premier temps, nous avons entraîné le modèle avec un jeu de données construit par l'OVSG se composant de 845 événements labélisés disponibles (542 VT, 217 Emboîtés et 86 LP) enregistrés sur la période 2013-2018. Nous avons obtenu un taux de classification moyen de 72%. Nous avons déterminé que la classe VT comprenait une variété de signaux recouvrant les classes VT, LP et Emboîtés. Après avoir manuellement revu en détails les formes d'ondes et les caractéristiques spectrales de ces signaux, nous avons introduit les classes d'événements Hybride et Longue-Période Monochromatique (également appelée Tornillo) correspondant à la description des signaux présentée par Moretti et al. (2020).

Ensuite, en utilisant ces nouvelles informations, un nouveau modèle a été construit avec 5 classes et testé. Nous avons obtenu un bien meilleur taux de classification de 84%. La classification est excellente pour les événements Emboîtés (93% de recall et precision) et Tornillos (93% de recall et precision). La classification des VT (90% recall, 89% precision) et des LP (86% recall, 82% precision) sont également très bonnes. La classe la plus difficile à reconnaître est la classe Hybride (64% recall, 69% precision). Les séismes Hybrides sont souvent confondu avec les signaux LP et VT. Cela peut être expliqué par la nature mme de cette classe et le processus physique qui inclut à la fois de la fracturation mais également une composante de résonance avec différents modes de fréquence.

De plus, en utilisant un modèle d'apprentissage supervisé pour distinguer les événements du bruit de fond, nous avons été capable de détecter trois fois plus d'événements que l'observatoire avec une méthode de STA/LTA.

Le machine learning est un outil très puissant pour traiter de grands jeux de données. Nous avons été capable d'améliorer la catalogue, corriger quelques erreurs de cassification et détecter plus d'événements.

Mots clefs : Sismologie, Volcan, La Soufrière, Classification, Détection, Machine Learning

Table des matières

Remerciements	I
Abstract	II
Résumé	III
Table des matières	VI
Table des figures	VII
Liste des tableaux	IX
1 Introduction	1
1.1 La surveillance volcanique	1
1.2 Le machine learning pour la surveillance volcanique	2
1.3 La Soufrière de Guadeloupe	3
1.3.1 Contexte	3
1.3.2 La détection et la classification de l'observatoire volcanique	4
1.3.3 Catalogue	6
2 Méthode d'apprentissage supervisé	8
2.1 Introduction	8
2.1.1 L'apprentissage automatique	8
2.1.2 La classification supervisée	8
2.1.3 La classification supervisée appliquée aux signaux sismiques d'origine volcanique	9
2.2 État de l'art du machine learning en sismologie	10
2.2.1 État de l'art de la classification d'événements sismique	10
2.2.2 État de l'art de la détection d'événements sismique	12
2.3 Prétraitements	13
2.3.1 Filtrage	13
2.3.2 Contrôle qualité	14
2.3.3 Normalisation	15
2.4 Extraction des descripteurs	15
2.4.1 Transformation des signaux en descripteurs	15
2.4.2 Descripteurs utilisés	16
2.5 Random Forest	17

2.6	Validation du modèle	20
2.6.1	Validation croisée	20
2.6.2	Métriques et Matrice de confusion	21
2.7	Résumé	23
3	Classification automatique des signaux sismo-volcaniques de la Soufrière de Guadeloupe	24
3.1	Sélection d'un modèle	24
3.2	Performances détaillées de la classification [2013-2018]	25
3.3	Performances de la classification affinée [2013-2018]	26
3.3.1	Nouvelle classification	26
3.3.2	Performances détaillées de la nouvelle classification	28
3.4	Analyse des descripteurs [2013-2018]	30
3.4.1	Performance des descripteurs	30
3.4.2	Réduction des descripteurs pour le modèle à 5 classes	32
3.5	Application du modèle à une nouvelle période [2018-2020]	35
3.5.1	Nouveau jeu de données	35
3.5.2	Premier résultat après application du modèle sur la période 2018-2020 et identification d'une erreur	37
3.5.3	Correction du modèle	37
3.5.4	Application du modèle corrigé sur 2018-2020	38
3.5.5	Recherche d'un changement de VT	40
3.6	Résumé	42
4	Détection et classification d'événements sismo-volcaniques	43
4.1	Introduction	43
4.2	Application du modèle sur le signal continu	43
4.2.1	Méthode	43
4.2.2	Résultats	44
4.2.3	Discussion	49
4.3	Introduction d'une classe bruit	51
4.3.1	Ajout de la classe bruit dans la phase d'apprentissage	51
4.3.2	Classe Signal contre classe Bruit appliqué sur deux semaines	58
4.3.3	Classification appliquée aux événements détectés	61
4.3.4	Discussion	62
5	Autres Résultats	66
5.1	Classification multistation	66

5.1.1	Modèles globaux	67
5.1.2	Modèles par vote	68
5.1.3	Modèles encapsulés	69
5.1.4	Résultats et discussion	70
5.2	Application au volcan Merapi	71
6	Conclusion et Perspectives	74
	Bibliographie	76
A	Annexes	83
B	Publication	85

Table des figures

1.1	<i>Carte de la Guadeloupe</i>	3
1.2	<i>VT, LP et Emboité</i>	5
1.3	<i>Distribution événements 2013-2018</i>	6
2.1	<i>Illustration du concept de classification</i>	9
2.2	<i>Filtrage</i>	14
2.3	<i>Contrôle qualité</i>	15
2.4	<i>Données séparées en 2 classes selon : a) Un seul descripteur. b) Deux descripteurs. c) Une combinaison de deux descripteurs</i>	16
2.5	<i>Construction arbre de décision</i>	19
2.6	<i>Random Forest</i>	21
2.7	<i>Validation croisée</i>	22
3.1	<i>Exemple d'événement Tornillo et Hybride</i>	28
3.2	<i>Performance des features</i>	31
3.3	<i>Représentation de la valeurs des meilleurs des descripteurs</i>	33
3.4	<i>Performances N meilleurs descripteurs</i>	34
3.5	<i>Distribution 2013-2020</i>	35
3.6	<i>Seuil Signal sur Bruit</i>	36
3.7	<i>Mauvais LP</i>	38
3.8	<i>Distribution des familles de VT</i>	40
3.9	<i>Probabilité des événement VT étant VT</i>	41
4.1	<i>Durée des événements</i>	44
4.2	<i>Détection de VT avec modèle à 5 classes</i>	45
4.3	<i>Détection de LP avec modèle à 5 classes</i>	46
4.4	<i>Détection de LP avec modèle à 5 classes</i>	47
4.5	<i>Détection de Hybride avec modèle à 5 classes</i>	48
4.6	<i>Détection d'Emboité avec modèle à 5 classes</i>	49
4.7	<i>Détection de Tornillo avec modèle à 5 classes</i>	50
4.8	<i>Représentation spatiale des meilleurs descripteurs de la classe bruit</i>	52
4.9	<i>Modèle à 6 classes appliqué à la détection de VT</i>	53
4.10	<i>Modèle à 6 classes appliqué à la détection de LP</i>	54

4.11	<i>Modèle à 6 classes appliqué à la détection d'Hybride</i>	56
4.12	<i>Modèle à 6 classes appliqué à la détection d'Emboîté</i>	57
4.13	<i>Modèle à 6 classes appliqué à la détection de Tornillo</i>	59
4.14	<i>Détection de Hybride avec modèle à Signal contre Bruit</i>	60
4.15	<i>Durée des détection</i>	61
4.16	<i>Ecart entre temps des détections OVSG et modèle</i>	65
5.1	<i>Modèle Globale Simple</i>	67
5.2	<i>Modèle Globale Combine</i>	68
5.3	<i>Modèle Vote 3 sous-modèle</i>	69
5.4	<i>Modèle Vote 1 sous-modèle</i>	70
5.5	<i>Modèles encapsulés</i>	73

Liste des tableaux

2.1	Liste des descripteurs utilisés pour représenter les signaux	18
2.2	Exemple d'une matrice de confusion.	23
3.1	Score par algorithme	24
3.2	Matrice de confusion de la classification à trois classes	25
3.3	Nouvelles classes	28
3.4	Matrice de confusion de la classification à cinq classes	29
3.5	Répartition classes selon SNr	36
3.6	Première application du modèle sur 2018-2020	37
3.7	Correction classe LP	38
3.8	Première application du modèle sur 2018-2020	39
3.9	Comparaison Emboîté OVSG et modèle	40
4.1	Matrice de confusion de la classification à cinq classes + Bruit	51
5.1	Accuracy multistation par station	66
5.2	Accuracy multistation par stratégie	71
5.3	Matrice de confusion de la classification des signaux du Merapi	72
A.1	Matrice de confusion de la classification à cinq classes à la station AMC pour le jeu de données restreint au multistation	83
A.2	Matrice de confusion de la classification à cinq classes à la station AMC pour le jeu de données restreint au multistation	83
A.3	Matrice de confusion de la classification à cinq classes à la station AMC pour le jeu de données restreint au multistation	84
A.4	Matrice de confusion de la classification multistation globale simple	84

1 Introduction

1.1 La surveillance volcanique

Les volcans sont des systèmes non-linéaires très complexes dont le comportement reflète une diversité de processus internes cachés pouvant entraîner une activité éruptive. La surveillance des volcans repose sur le paradigme selon lequel divers paramètres observables ont un lien de causalité plus ou moins direct avec ces processus complexes et variables internes du système ainsi qu’avec une série de processus de forçage. Parfois, la simple absence ou la présence de certaines observables peut fournir des informations très précieuses sur la dynamique du système. La surveillance des volcans est cruciale pour évaluer l’état d’activité d’un volcan actif et pour évaluer la probabilité d’une éruption future et la fenêtre temporelle probable d’occurrence. Elle constitue donc l’un des piliers des méthodes scientifiques qui contribuent à atténuer les risques sur les sociétés humaines. Parmi les nombreuses disciplines utilisées dans la surveillance des volcans, dont la géodésie, la géochimie, la gravimétrie, l’imagerie géophysique, la surveillance thermique, l’acoustique, la sismologie reste la discipline la plus largement utilisée. Elle fournit souvent une richesse de données haute résolution en continu et en temps réel à partir de réseaux de capteurs déployés sur les volcans, sur terre et sur les fonds marins.

Une surveillance sismique volcanique optimale requiert la détection, la localisation, la quantification de la magnitude et la classification d’un grand nombre de signaux sismo-volcaniques. Celle-ci doit être réalisée de façon journalière dans les observatoires volcanologiques. Avec les progrès en sismo-volcanologie des dernières décennies et une surveillance de haute résolution sur de nombreux volcans très actifs, il existe désormais un corpus de connaissances qui permet d’associer, avec moins d’incertitude, les différents types de signaux sismiques enregistrés sur les volcans avec différents processus sources (McNutt et al., 2005 ; Chouet et Matoza, 2013 ; McNutt and Roman, 2015). Ainsi, avec d’autres données de surveillance, la sismicité volcanique constitue un outil fondamental pour suivre, dans le temps et dans l’espace, l’évolution de la dynamique de l’activité volcanique afin de prévoir en temps opportun la probabilité de scénarios futurs, dans un contexte d’incertitude, intrinsèque au comportement des systèmes non-linéaires complexes. L’occurrence de certains événements peut être précurseuse d’éruptions volcaniques et peut être utilisée dans les systèmes d’alerte précoce. Alors que, de nos jours, le travail de détection est principalement automatique, la tâche de classification est souvent effectuée manuellement. La tâche de classification manuelle est limitée et plusieurs facteurs peuvent diminuer la robustesse de cette labellisation. La classification est basée sur le jugement subjectif de l’opérateur, donc si la tâche est effectuée par plusieurs personnes, les critères de classification peuvent parfois varier d’une personne à l’autre. Ce travail doit être effectué en temps quasi-réel, avec une énorme quantité de données à traiter, en particulier pendant les périodes de crise volcanique où l’interprétation scientifique doit être rapide et des conseils opportuns doivent être fournis par les scientifiques aux autorités pour les aider dans leur processus décisionnel pour atténuer les risques sur la population et les infrastructures. En raison du volume de données en constante augmentation et des limites de la classification manuelle, le développement d’outils automatiques utilisant une

approche d'apprentissage automatique pour classer rapidement et aussi précisément que possible ces signaux sismiques volcaniques est essentiel à l'interprétation spatio-temporelle des données. Cette méthodologie a donc le potentiel d'améliorer considérablement le suivi et l'analyse en temps réel des phases d'instabilité et la prévision de leur évolution possible.

À la Soufrière de Guadeloupe, la tâche de classification des signaux sismiques est actuellement réalisée de manière manuelle par différents experts à l'Observatoire Volcanologique et Sismologique de Guadeloupe (OVSG-IPGP) en regardant les formes d'ondes enregistrées sur un sous-jeu de stations du réseau. Les séismes d'origine volcanique sont de bons indicateurs de l'état du système hydrothermal superficiel et du système magmatique plus profond. Et, même si cinq types de signaux sont reconnus dans l'analyse des données par l'observatoire (Moretti et al., 2020), seulement trois classes principales sont facilement distinguables sur les traces sismiques au cours du protocole analytique quotidien et donc cataloguées : les événements volcano-tectoniques (VT), les événements longue-période (LP) et les événements emboîtés (les VT et les LP peuvent former des événements individuels distincts, mais ils peuvent arriver tellement rapprochés dans le temps qu'ils peuvent souvent être emboîtés ensemble temporellement). Donc, pour la première fois au volcan de la Soufrière, nous allons tenter de classer automatiquement les événements détectés. Et également, nous allons tenter de réaliser les étapes de détection et de classification en simultanée.

1.2 Le machine learning pour la surveillance volcanique

L'apprentissage supervisé consiste à entraîner un algorithme à réaliser une tâche, ici assigner une classe sismique à un événement d'origine volcanique. On parle de supervisé, car nous possédons déjà une base de données composée de couples événements-classes à partir de laquelle l'algorithme va apprendre à faire le lien entre ces deux éléments. Si nous n'avions pas la classe associée à chaque événement, mais que l'on voudrait tout de même entraîner un algorithme à séparer les événements en différentes classes alors on parlerait de non-supervisé. Mais plutôt que d'analyser la forme d'onde du signal, nous transformons ce signal en un ensemble de descripteurs qui porteront ses caractéristiques. Ainsi, le rôle du machine learning est de construire des frontières dans l'espace des descripteurs correspondantes spécifiquement à chacune des classes considérées. Pour différencier au mieux chacune des classes, les descripteurs doivent être choisis avec attention et doivent représenter la variabilité des signaux. Ici, nous choisissons d'utiliser un jeu de descripteur proposé par Malfante et al. (2018), qui donne une description générale et précise des signaux transitoires et qui s'est révélée efficace. Ceci nous permettra également de vérifier si les descripteurs sélectionnés sur un volcan peuvent être utilisés pour un autre volcan.

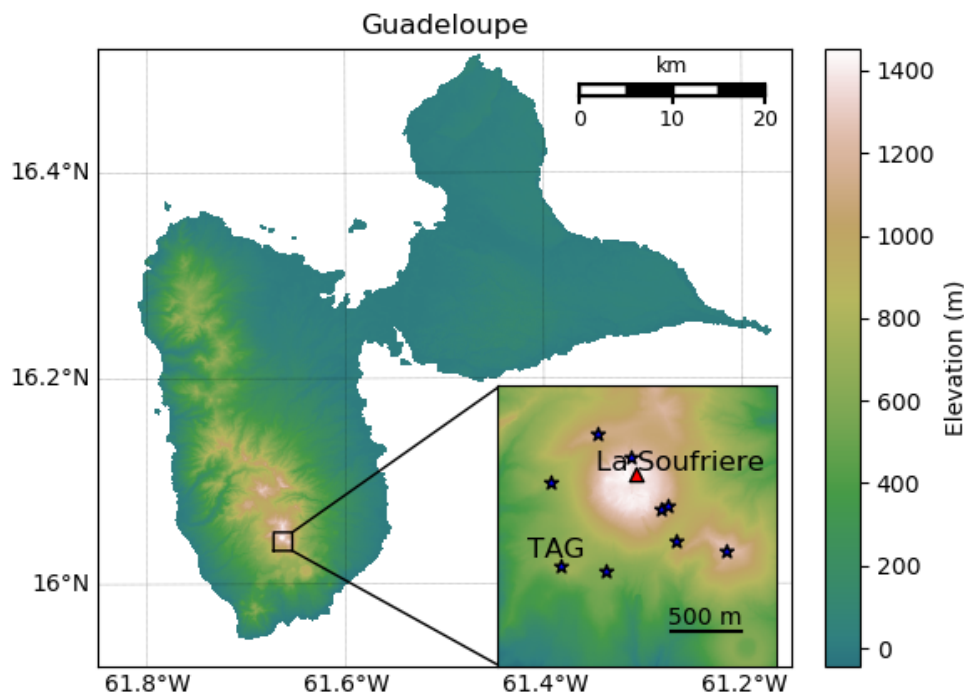


FIGURE 1.1 – Carte de la Guadeloupe et position des stations sismologiques

1.3 La Soufrière de Guadeloupe

1.3.1 Contexte

La Soufrière de Guadeloupe est un volcan actif situé sur l'île de Basse-Terre en Guadeloupe dans les Antilles (Figure 1). Elle est actuellement sous la surveillance de l'Observatoire Volcanologique et Sismologique de Guadeloupe" (OVSG-IPGP). De nombreuses éruptions magmatiques et non-magmatiques ont eu lieu dans son histoire. L'actuel dôme andésitique a été formé par le dernier événement magmatique majeur en 1530 (Boudon, 1988; Boudon, 2008). Le volcan a aussi subi quelques éruptions phréatiques en 1690, 1797-98, 1809-12, 1836-37, 1956, 1976-77 (Komorowski et al., 2005). La dernière en date, a été particulièrement violente et a demandé l'évacuation de plus de 70 000 personnes pendant 6 mois. Cet événement a été associé à une importante controverse dans la communauté scientifique (Komorowski et al., 2005; Hincks et al., 2014; Komorowski et al., 2018). Après cette crise, l'activité a lentement diminué, atteignant en 1990 son niveau d'activité le plus bas depuis le début de la surveillance en 1950 (Komorowski et al., 2005; Villemant et al., 2014; Jessop et al., 2019; Moretti et al., 2020). Depuis 1992, les activités sismique, thermique et fumerollienne ont augmenté à la Soufrière (Komorowski et al., 2005; Villemant et al., 2014) pour atteindre le plus haut niveau d'énergie sismique le 27 Avril 2018 avec le séisme volcano-tectonique le plus fort ressenti depuis la crise de 1976-77 (M4.1). Même si l'énergie sismique relâchée a baissé depuis Janvier-Avril 2018, période de grosse activité, la sismicité de la Soufrière de Guadeloupe a été caractérisée en Août 2018 par l'occurrence d'essaim sismique de basse énergie avec de nombreux séismes VTs et Emboités et quelques rares LPs. Entre Août 2018 et la fin Avril 2020, 37 de ces essaims ont été enregistrés par l'OVSG pour un total de 7 114 signaux identifiés avec une moyenne de 192 événements par essaim (min : 13, max : 1014; Moretti et al., 2021). Cette sismicité, croissante bien que peu énergétique, présente

des défis très importants pour l'identification des événements, leur classification et leur interprétation par l'équipe de l'observatoire pour suivre de manière efficace l'activité du volcan. Mais voyons tout d'abord comment sont réalisées les étapes de détection et de classification à l'OVSG.

1.3.2 La détection et la classification de l'observatoire volcanique

Actuellement à l'OVSG, le travail de détection est effectué principalement de manière automatique en utilisant un algorithme de type STA/LTA. Les opérateurs sur site analysent les détections en considérant plusieurs stations se situant à différentes positions autour du sommet de l'édifice volcanique (Figure 1.1) et déterminent la magnitude et la localisation de l'événement. En observant les sismogrammes à différentes stations, ils assignent une classe à l'événement détecté selon la nomenclature utilisée par l'observatoire. La grande majorité des signaux possède une très faible magnitude (souvent négative), une durée très courte et sont souvent distinguables sur deux stations seulement.

Même si cinq principaux types de signaux sont reconnus dans l'analyse des données par l'observatoire (Moretti et al., 2020), seulement trois classes principales facilement distinguables sur les traces sismiques continues (SEFRAN) au cours du protocole analytique quotidien sont labellisées : les événements Volcano-Tectoniques (VT), les événements Longues-Périodes (LP) et les événements Emboîtés. Les deux autres classes, les événements Hybrides et les événements Longues-Périodes Monochromatiques (ou Tornillos) sont classés respectivement avec les événements VT et LP. La Figure 1.2 montre les formes d'ondes, les spectrogrammes et les spectres typiques de ces événements. On les définit de la manière suivante :

Séismes Volcano-Tectoniques : Ces événements ont un contenu haute fréquence similaire aux séismes tectoniques. Chouet et Matoza (2013) définissent leur gamme de fréquence entre 5 et 15 Hz. Dans le cas de la Soufrière, un pic caractéristique est souvent observé entre 10 et 15 Hz. La forme d'onde présente une arrivée d'onde P très impulsive. La source de ces événements est une rupture en profondeur de la roche associée aux changements de contrainte dus aux mouvements du magma (Chouet et Matoza, 2013).

Séismes Longue-Période ou Basse Fréquence : Ces signaux présentent une gamme de fréquence généralement entre 0.5 et 5 Hz (Chouet et Matoza, 2013). À la Soufrière, ces événements basse fréquence sont caractérisés par un contenu fréquentiel plutôt étroit autour de 4 Hz. Leurs formes d'ondes sont caractérisées par une arrivée d'onde P émergente alors que la phase S n'est pas identifiable. Il est communément accepté que le processus qui génère les événements LP est lié à la résonance de fractures, dykes, conduits ou de cavités durant la propagation de fluide magmatique ou hydrothermal (Chouet et al., 1988 ; Ucciani, 2015).

Séismes Emboîtés : Ces événements apparaissent comme de petits paquets de plusieurs signaux sismiques dans lesquels des événements successifs apparaissent dans la coda des autres. Ils ne sont pas concomitants ou précurseurs d'un phénomène particulier (Moretti et al., 2020).

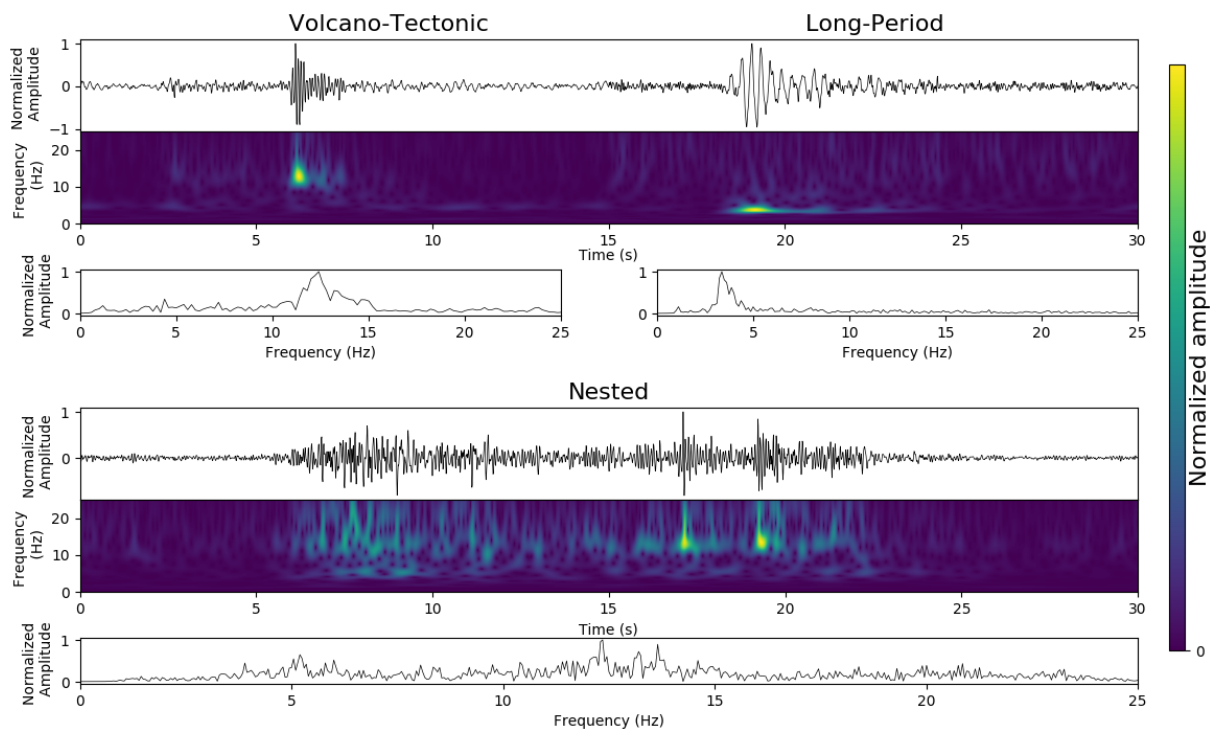


FIGURE 1.2 – Exemple de forme d'onde filtrée entre 0.8-25 Hz, Spectrogramme et Spectre de Fourier des événements VT (en haut à gauche), LP (en haut à droite) et Emboîtés (en bas) enregistrés à la station TAG à la Soufrière de Guadeloupe sur la composante verticale

Les événements Emboîtés consistent en une séquence de plusieurs événements sismiques avec des inter-temps très courts, avec très souvent plus de 6 événements sismiques dans une courte séquence (10 s) (Ucciani, 2015 ; Moretti, 2020). Ces signaux emboîtés sont souvent des VT, mais pas toujours, il peut également y avoir des LP Emboîtés. Les différents signaux dans un événement Emboité ne sont pas toujours similaires en terme d'amplitude ou de forme d'onde. Le processus source relié à cette classe n'est pas encore très bien compris et ces événements se retrouvent à la Soufrière de Guadeloupe et à la Montagne Pelée en Martinique. Bell et al. (2017) ont souligné la présence de séismes LP "drumbeat" sur le volcan Tungurahua en Équateur, ces événements sont caractérisés par des gammes de temps inter-événements et des amplitudes très restreintes comparées aux formes d'onde et à l'activité plus classique du volcan. Ces propriétés nécessitent un emplacement de source fixe, un mécanisme non-destructif ou à renouvellement rapide et un système physique oscillant autour d'un état proche de l'équilibre. Ces événements sont similaires aux événements Emboîtés de la Soufrière de Guadeloupe, mais alors que les événements Emboîtés sont majoritairement des VT et peuvent varier en terme d'amplitude et de forme d'onde, les "Drumbeat" sont des LP similaires en amplitude et forme d'onde. De plus, les "Drumbeat" sont des séismes associés à la croissance d'un dôme, or il n'y a pas de dôme actif à la Soufrière, donc même si en apparence ces deux types d'événements sont similaires, ils restent assez éloignés en termes de processus génératif.

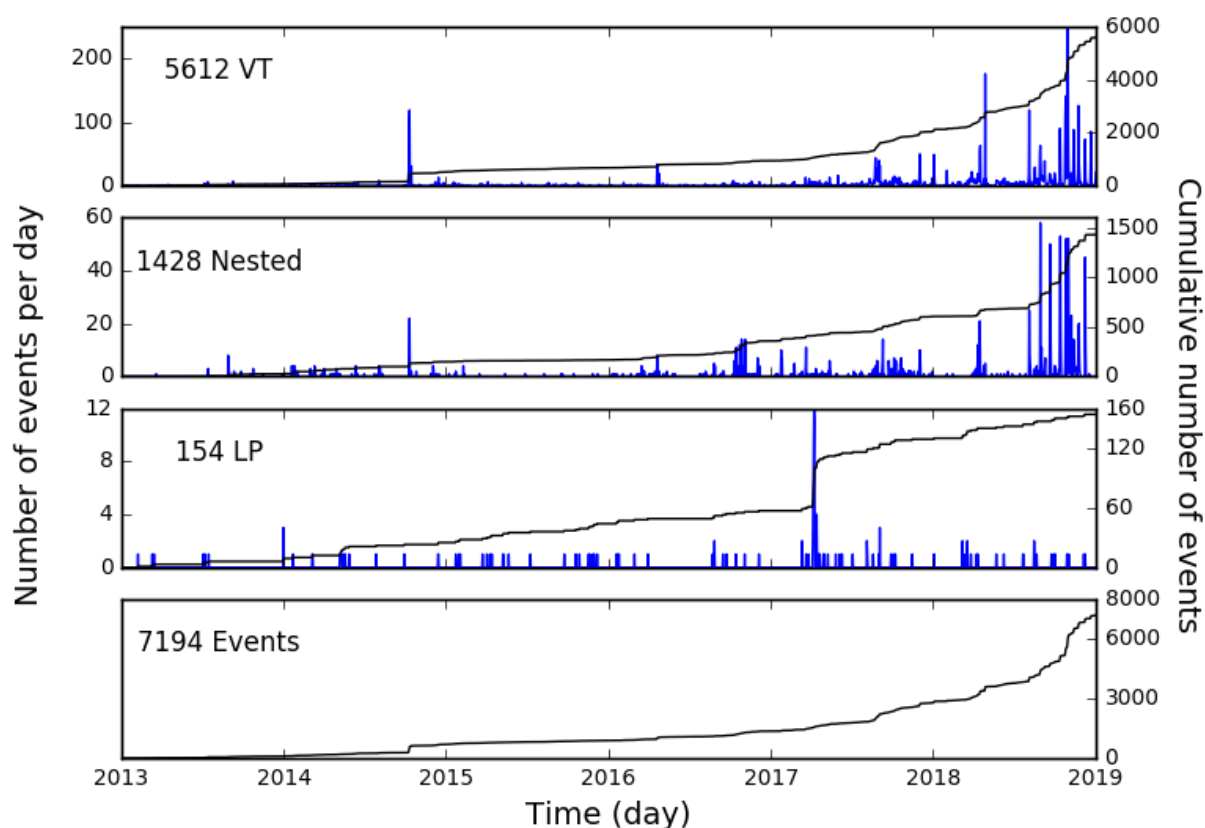


FIGURE 1.3 – Distribution (Bleu) et distribution cumulée (Noir) temporelle du nombre d'événements journaliers entre le 1er Janvier 2013 et le 31 Décembre 2018, pour (de haut en bas) les événements VT, Emboités et LP avec la distribution cumulée de toutes les classes (données OVSG-IPGP)

1.3.3 Catalogue

Une fois les étapes de détection et de classification réalisées, nous avons accès au catalogue. La base de données d'apprentissage utilisée pour l'analyse vient du catalogue construit par l'OVSG et disponible sur WebObs. WebObs est un outil web qui effectue une surveillance centralisée et automatisée des volcans en temps réel. Ce système propose une base de données pour la gestion des équipements du réseau, une douzaine de tâches périodiques dédiées à chaque technique de surveillance comme la sismologie, la géodésie et la géochimie avec des exécutions automatisées des tâches et des interfaces web pour la saisie manuelle des données et l'exportation (Beauducel, 2020). Notre période d'étude correspond à la période 2013-2018. Les informations disponibles sur WebObs indiquent l'heure et la date de première arrivée, la durée de l'événement et la classe de l'événement proposée par les analystes. Dans cette thèse, on se concentre majoritairement sur les signaux enregistrés sur une seule station. Nous avons sélectionné la station à trois composantes Piton Tarade (TAG, Nanometric Taurus Trillum compact 120S at 100 Hz), une station historique du réseau de surveillance de la Soufrière située à moins d'un km du sommet. Cette station, protégée des perturbations extérieures, n'est pas sujet aux vents forts et nous avons des données pour 84% des jours de la période 2013-2018. Le catalogue d'événements sur la période 2013-2018 contient 7 149 événements, 78% sont des événements VT, 20% sont des Emboités et les 2% restants sont des événements LP. La figure 1.3 montre la distribution de ces événements au

cours du temps. Les distributions des VT et des Emboités sont plutôt corrélées, mais avec un nombre bien plus important de VT que d'Emboités. Les événements LP sont plus uniformément répartis dans le temps excepté un essaim d'une dizaine d'événements qui se sont produits en Avril 2017. La distribution des événements VT et Emboités n'est pas constante au cours du temps, on observe une grosse augmentation de leurs occurrences à partir de la fin 2018.

Dans cette thèse, nous présenterons tout d'abord la méthode d'apprentissage supervisé, le workflow et les algorithmes. Ensuite, dans le chapitre 3, nous verrons comment à partir du catalogue d'événement de l'OVSG on a pu construire un modèle de classification automatique fiable et quelles sont les paramètres les plus utiles. Dans le chapitre 4, nous nous intéresserons à la détection et la classification des événements à partir du signal continu. Et enfin, dans le chapitre 5, nous verrons différents résultats obtenus au cours de la thèse sur une approche multistation de la classification et l'application de la méthode présentée au chapitre 2 sur le volcan Merapi en Indonésie.

2 Méthode d'apprentissage supervisé

2.1 Introduction

2.1.1 L'apprentissage automatique

L'apprentissage automatique, ou *machine learning*, est un champ de recherche à la croisée de l'intelligence artificielle, de l'informatique et des statistiques. Le but est de donner à un algorithme la capacité d'apprendre à partir de données. L'algorithme doit apprendre à réaliser une tâche sans avoir été explicitement programmée pour. Les algorithmes de machine learning s'appuient, pour construire les modèles, sur les patterns issus des données et ajustent leurs paramètres pour maximiser leurs performances, c'est-à-dire, ici le taux de reconnaissance le taux de reconnaissance des différentes classes d'événements. Les applications du machine learning sont multiples, on peut citer la reconnaissance automatique d'image (Simonyan et al., 2015), la description automatique d'image (Karpathy et al., 2015) ou encore la colorisation automatique d'image en niveau de gris (Sousa et al., 2013). Parmi les applications du machine learning les plus connues, on peut évoquer le logiciel de commande vocal d'Apple, Siri (Capes et al., 2017) ou encore, le logiciel de reconnaissance musicale de chansons Shazam (Wang et al., 2003). Les applications peuvent être également un peu plus extravagante, comme la transformation d'une photo pour la faire correspondre au style d'un grand peintre (Gatys et al., 2015). Dans ce chapitre, nous expliciterons les différentes étapes de la construction d'un modèle ainsi que les difficultés pouvant être rencontrées. Mais tout d'abord, commençons par présenter le concept de classification supervisée qui est au centre de cette thèse.

2.1.2 La classification supervisée

Dans le domaine du machine learning, plusieurs types de problèmes peuvent être posés, des problèmes de régression où la donnée que l'on cherche à prédire est quantitative, c'est-à-dire des valeurs numériques ou bien un problème de classification où l'on cherche à prédire une classe. Dans le cas de cette thèse, il s'agit d'un problème de classification, car on veut prédire le type d'événement sismique à partir de sa forme d'onde. Imaginons un problème consistant à associer les éléments d'un espace X aux éléments d'un espace Y . On a un couple de données $(x, y) \in X \times Y$. La tâche de l'algorithme consiste à trouver la fonction f tel que $f(x) = y$. Dans un problème de classification, y est une donnée catégorique (s'exprimera mathématiquement sous la forme d'un entier $y \in \{0, \dots, 9\}$). Dans notre couple (x, y) , x , l'entrée sera appelée descripteur ou *feature*, y sera appelé classe ou *label* et enfin f est la fonction de décision.

La classification est dite supervisée, car nous connaissons déjà des couples (x, y) . Nous construisons un modèle de classification automatique supervisée à partir de ces couples (x, y) qui constituent notre jeu d'apprentissage. Notre objectif est donc de faire des prédictions y à partir de nouvelles données x qui n'ont pas été rencontrées. Les problèmes d'apprentissage supervisé nécessitent souvent à la base un travail humain pour construire le jeu d'apprentissage, mais ensuite,

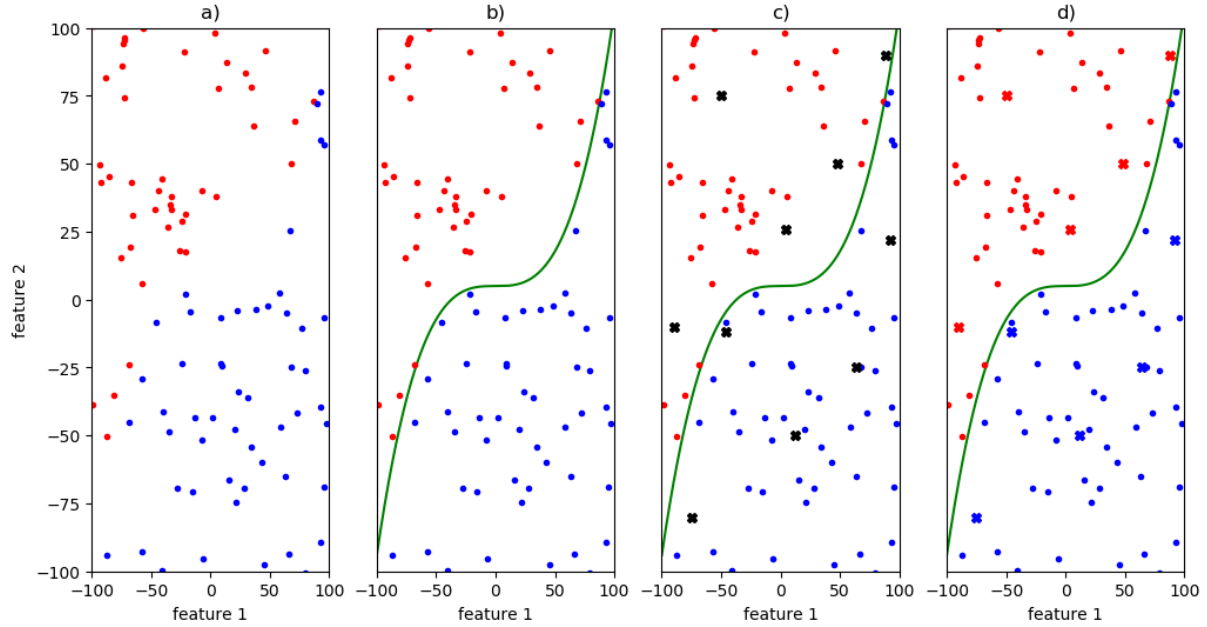


FIGURE 2.1 – Illustration du concept de classification. a) Un jeu de données comportant 2 classes et représenté par deux features est considéré, b) l'apprentissage permet de séparer ces deux classes dans l'espace des features (b). c) De nouvelles données (croix noires) peuvent être automatiquement classées par le modèle. d) Les nouvelles données sont classées.

automatisent et accélèrent une tâche qui pourrait être fastidieuse, chronophage, voire impossible.

La figure 2.1 illustre de façon générale le workflow suivie par les méthodes de machine learning. La figure 2.1(a) représente un jeu de données composé de deux classes (la première représentée par des points rouges et la seconde représentée par des points noirs). À chacune des données sont associées deux descripteurs (ou feature). Comme le montre la figure 2.1(b), construire un modèle consiste à construire la ou les frontières qui séparent au mieux les données dans l'espace des features. Cette étape est réalisée par l'algorithme d'apprentissage. Ici, la frontière séparant le mieux les données est représentée par la courbe verte. Enfin, une fois que ces frontières sont construites, que le modèle est entraîné, on peut l'appliquer à de nouvelles données entrantes (figure 2.1(c)) pour les classer dans une des classes considérées par le modèle, figure 2.1(d).

2.1.3 La classification supervisée appliquée aux signaux sismiques d'origine volcanique

Dans cette thèse, les données considérées sont des signaux sismiques d'origine volcanique, ces signaux sont de durées parfois très variables entre quelques secondes jusqu'à plusieurs dizaines de secondes. Les classes de ces signaux sont associées à la physique génératrice de ces événements sismiques, donc différents processus physiques généreront différentes formes d'ondes. On doit alors trouver des descripteurs x capables de décrire ces formes d'ondes et de les discriminer en fonction de la classe y de l'événement.

2.2 État de l'art du machine learning en sismologie

Les applications du machine learning en sismologie sont nombreuses depuis quelques années, la détection d'événements (Tiira et al., 1999), le pointé de phase (Dai et al., 1995), la classification d'événements sismiques (Dysart et al., 1990), la mise en place de système d'early warning (Kong et al., 2016), la prediction de mouvements du sol (Alavi et al., 2011), la tomographie (Moya et Irikura, 2010) ou encore en géodésie (DeVries et al., 2017). Un inventaire des différentes application du machine learning en sismologie a été réalisé par Kong et al. (2019). Nous nous intéresserons ensuite plus en détail sur les études qui ont pu être réalisé sur les aspect de détection et de classification d'événements sismiques en contexte volcanique notamment.

2.2.1 État de l'art de la classification d'événements sismique

Beaucoup d'études récentes ont eu pour sujet la classification automatique d'événements sismo-volcaniques en utilisant différentes techniques.

Benitez et al. (2006) a utilisé les chaînes de Markov cachées (HMM) pour la classification en continu de quatre classes à l'île de la Déception et a obtenu un taux de reconnaissance de 90%. Les chaînes de Markov caché ont également été utilisées par Alasonati et al. (2006) au Merapi, par Gutiérrez et al. (2009) au San Cristóbal, par Ibañez et al. (2009) au Stromboli et à l'Etna ou encore Cortès et al. (2009) au Colima et au Popocatepetl et ils ont tous obtenu des taux de reconnaissance autour de 80% ou supérieur. Ohrnberger (2001) a également appliqué un HMM au Merapi obtenant un taux de reconnaissance moyen de 67% mais ayant des résultats très différents selon les classes. Bayreuther et al. (2012) ont également appliqué un HMM au Merapi et ont obtenu des résultats encourageant notamment pour les événements Rockfall.

Hibert et al. (2014) ont utilisé de la logique floue et une méthode basée sur des règles de décision pour distinguer les éboulements des événement volcano-tectoniques au Piton de la Fournaise et ont obtenu un taux de reconnaissance de 92%. Mais cette méthode a besoin de l'intervention de l'opérateur pour choisir les fonctions et les seuils utilisés.

Langet (2014) a utilisé la régression logistique et une machine à vecteur de support (SVM) et obtenu 90% et 92% de bonne reconnaissance respectivement sur le même jeu de données que Hibert et al. (2014). Ces méthodes ont pour avantage d'utiliser très peu de descripteurs, mais les jeux de données utilisés dans les deux études présentaient un très bon rapport signal sur bruit et les deux classes d'événements reconnues dans l'analyse, VT et éboulement, sont facilement identifiables. Le SVM a également été testé au Mont Etna par Langer et al. (2009) pour un taux de reconnaissance de 94.8% et par Masotti et al. (2006) pour l'identification de tremor avec un taux de reconnaissance supérieur à 94%. Enfin, Apolloni et al. (2009) a appliqué un SVM et un perceptron multicouche aux événements du Stromboli obtenant respectivement une overall accuracy de 99.54% et de 98.39%.

Titos et al. (2018) a utilisé des réseaux de neurones profonds pour la classification de sept classes d'événements sismiques au "Volcàn de Fuego" et a obtenu 94% de bonne classification. Les réseaux de neurones ont également été testés par Langer et al. (2006) à Soufrière Hills, Ibs-Von Seht (2008) au Krakatoa, Curilem et al. (2009) au Villarica, Falsaperla et al. (1996) et Giacco

et al. (2009) au Stromboli, Scarpetta et al. (2005) au Vésuve ou encore Langer et al. (2009) au Vésuve et ont donné des résultats assez variés entre 70% et 99% de taux de bonne reconnaissance. Mais cette méthode demande beaucoup de tests pour savoir quelles couches de neurones utiliser et le nombre optimal.

Moussavi et al. (2016) ont comparé un réseau de neurones et une régression logistique pour déterminer la profondeur d'un séisme (profond ou superficiel) obtenant respectivement des taux de reconnaissance de 90.7% et 88%.

Curilem et al. (2018) a utilisé des cross-corrélations de spectrogrammes en accord avec un algorithme des k plus proches voisins pour la classification de cinq classes au volcan LLaima au Chili et a obtenu un score de 95% pour trois des cinq classes considérées. Mais les données utilisées avaient un haut rapport signal sur bruit, facilitant la comparaison entre spectrogrammes.

Bueno et al. (2019) ont utilisé un réseau de neurones bayésien pour la classification de cinq classes sur deux volcans, avec lequel ils ont obtenu une excellente performance de 92.1% quand les deux jeux de données sont réunis, ils ont aussi montré que les incertitudes étaient liées à l'activité des volcans.

Des études ont également utilisé des algorithmes de classification non supervisés tels que du partitionnement de données (Langer et al., 2009), des cartes auto-adaptatives (Esposito et al., 2008 ; Langer et al., 2009 ; Köhler et al., 2010 ; Unglert et al., 2016) et des analyses en composantes principales (Unglert et al., 2016) avec de bons résultats.

Au volcan Ubinas (Pérou), Malfante et al. (2018) ont utilisé un Random Forest (RF) ou un SVM et ont obtenu 92.5% et 92.1% de bonne reconnaissance. Cela montre que les différences de performances selon les types d'algorithmes ne sont pas extrêmement importantes et l'outil le plus important pour pouvoir faire de la reconnaissance automatique est le jeu de descripteurs utilisé pour représenter les signaux.

L'algorithme utilisé dans cette thèse est un Random Forest (RF), car il permet de traiter avec un grand nombre de descripteurs sans sur-apprentissage et peut également donner des informations sur l'importance des descripteurs dans le modèle final. Ce genre d'algorithme a déjà été testé et s'est montré puissant dans un contexte volcanique. Provost et al. (2016) ont proposé une classification automatique se basant sur le calcul de 71 descripteurs et ont utilisé un algorithme RF. Ils se sont concentrés sur la classification d'événements enregistrés près du glissement de terrain de Super-Sauze (Sud-Est de la France). Ils ont obtenu une performance de 93% de bonne classification. En utilisant plusieurs stations Maggi et al. (2017) a proposé un classifieur opérationnel se basant sur le RF pour la surveillance du volcan du Piton de la Fournaise. Ils ont défini huit classes, ont utilisé une approche multistation et ont atteint une performance de 92% de bonne classification en combinant les informations de trois stations. Pour combiner au mieux les informations des différentes stations, ils définissent pour chaque station les meilleurs descripteurs et utilisent les meilleurs descripteurs de plusieurs stations en entrée de la nouvelle analyse multistation. Hibert et al. (2017) ont aussi proposé un classifieur RF pour la classification d'éboulements et de séismes volcano-tectoniques au Piton de la Fournaise et ont obtenu des performances entre 90% et 99% selon la taille du jeu de données d'apprentissage. Malfante

et al. (2018) ont proposé un large set de descripteurs pour la représentation des signaux sismo-volcaniques, en effet 34 descripteurs calculés depuis trois domaines de représentation ont été utilisés. En utilisant un algorithme RF, ils ont pu réussir à bien classer 92.5% des événements du volcan Ubinas au Pérou. Malfante et al. (2018) ont également montré qu'avec seulement les trois meilleurs descripteurs, ils pouvaient atteindre une performance de 84% et avec les 13 meilleurs descripteurs, ils pouvaient atteindre un taux de bonne classification de 90%.

Des études ont réussi à améliorer leur classification initiale. Langer et al. (2006) ont classé cinq classes d'événements volcaniques à la Soufrière de Montserrat. Ils ont dans un premier temps obtenu un taux de bonne classification de 70%. Mais après analyse des événements mal classés par le modèle, ils ont trouvé que, pour la plupart d'entre eux, la classe initialement affectée aux signaux était incorrecte. Ils ont manuellement reclassé ces événements et ont entraîné un nouveau modèle de classification et obtenu un nouveau score de 80% de bonnes classifications. Hammer et al. (2013) ont proposé un modèle de Markov caché pour la détection et la classification d'événement dans un contexte alpin. Ils ont appliqué leur modèle sur le signal continu et ont pu détecter 97% des événements et en classer correctement 87%. Au début, ils distinguaient deux classes mais ont pu mettre en évidence la présence d'une troisième classe lorsque leur critère seuil de ressemblance avec les deux classes existantes n'était pas atteint. Langet (2014) a montré que la classification utilisée au volcan Kawah Ijen avec huit classes n'était pas adaptée. En utilisant un algorithme de classification non-supervisé, ils ont montré qu'avec les données et descripteurs disponibles, seulement cinq classes étaient réellement distinguables.

2.2.2 État de l'art de la détection d'événements sismique

La plupart du temps la tâche de détection se fait automatiquement (Alasonati et al., 2006; Köler et al., 2010; Scarpetta et al., 2005). Certaines techniques de détection automatique se sont révélées efficaces et sont fréquemment utilisées de façon opérationnelle. L'une d'entre elles est le STA/LTA (Allen, 1978; 1982), le rapport d'énergie d'une courte fenêtre sur une plus longue fenêtre du signal temporel. D'autres systèmes de détection basés sur le kurtosis du signal (Baillard et al., 2014) ou sur du template matching (Gibbons et al., 2006) ont été proposés. Bell et al. (2021) utilisent une combinaison de STA/LTA et de template matching pour détecter les événements liés à l'activité du volcan.

Cependant depuis quelques années des méthodes de détection novatrices utilisant du machine learning ont été mis en place.

Wang & Tien (1995) ont entraîné deux réseaux de neurones, le premier utilisant en entrée le STA/LTA du signal continu, le second utilise les spectrogrammes d'une fenêtre glissante sur le signal continu. Ils ont montré que ces détecteurs basés sur des réseaux de neurones sont plus efficace qu'un STA/LTA notamment pour les signaux avec un faible rapport signal sur bruit. En utilisant un jeu de données pour lequel un détecteur basé sur le STA/LTA détecte 50% des événements, Madureira & Ruano (2009) ont pu obtenir des valeurs de recall et de precision supérieures à 95% en utilisant un système d'apprentissage supervisé basé sur un classifieur Perceptron

Multicouche. Puis, toujours avec le même jeu de données, ils ont obtenu un score quasi parfait en utilisant un classifieur SVM, cependant, le modèle ainsi construit présente une forte complexité. Wu et al. (2018) ont développé un détecteur DeepDetect basé sur les réseaux de neurones et ont montré que leur méthode est plus efficace qu'une méthode plus classique par template matching notamment lorsque la taille des événements varie beaucoup. D'autres outils basés sur des réseaux de neurones ont été développés, par exemple ConvNetQuake présenté par Perol et al. (2018), a permis de détecter 17 fois plus d'événements que dans le catalogue initial. Ou encore CRED, présenté par Mousavi et al. (2019) qui après comparaison avec du STA/LTA, du template matching ou un algorithme FAST (Fingerprint And Similarity Thresholding) montrent que cette méthode peut être efficace pour baisser les seuils de détection tout en minimisant le taux de fausses détections. On peut encore citer Wiszniowski et al. (2014), Ross et al. (2018) ou encore Seydoux et al. (2020) qui ont utilisé des réseaux de neurones pour la détection automatique d'événements sismiques.

En matière de détection automatique d'événements sismique en contexte volcanique, Lapins et al. (2021) ont appliqué un modèle utilisant l'apprentissage par transfert sur les données du volcan Nabro et ont montré que leur méthode permettait d'obtenir plus de détections, moins d'erreur sur les détections et cela pour des coûts de calculs moindres. Maeda et al. (2021) ont proposé un modèle basé sur l'amplitude des signaux sur plusieurs stations, du deep learning et un SVM pour la détection d'événement au Mont Ontake au Japon et ont réussi à détecter 97% des événements.

La plupart des études se concentrant sur la détection automatique d'événements sismiques se font en contexte tectonique et en utilisant les réseaux de neurones profonds.

Maintenant que l'on a vu ce qu'il est possible de faire en matière de classification et de détection automatique dans le domaine de la sismologie et notamment de la sismo-volcanologie, entrons plus en détail dans la méthode et commençons par les prétraitements que l'on applique aux signaux pour faciliter le travail de classification.

2.3 Prétraitements

Les différentes étapes de prétraitements vont servir à harmoniser les signaux, faire en sorte que les événements d'une même classe soient le plus proche possible et éliminer ceux qui pourraient parasiter la reconnaissance automatique par la suite. Pour cela, on va donc faire du filtrage, un contrôle de la qualité des données et normaliser les signaux.

2.3.1 Filtrage

Le premier prétraitement que l'on applique aux signaux est un filtrage. On veut éliminer des signaux parasites qui sont liés à du bruit haute-fréquence ou basse-fréquence et qui ne correspondent donc pas à la signature de l'événement que l'on cherche à classer. Cela retire également l'offset (figure 2.2) en cas de dérive instrumentale. Dans le cas de la Soufrière de Guadeloupe, les fréquences des signaux caractéristiques d'une classe ou d'une autre sont entre 3 Hz et 15 Hz. On a donc choisi d'appliquer un filtre passe-bande butterworth entre 0.8 Hz et 25 Hz, plus large que les fréquences d'intérêts pour éviter les effets de bord dus au filtre.

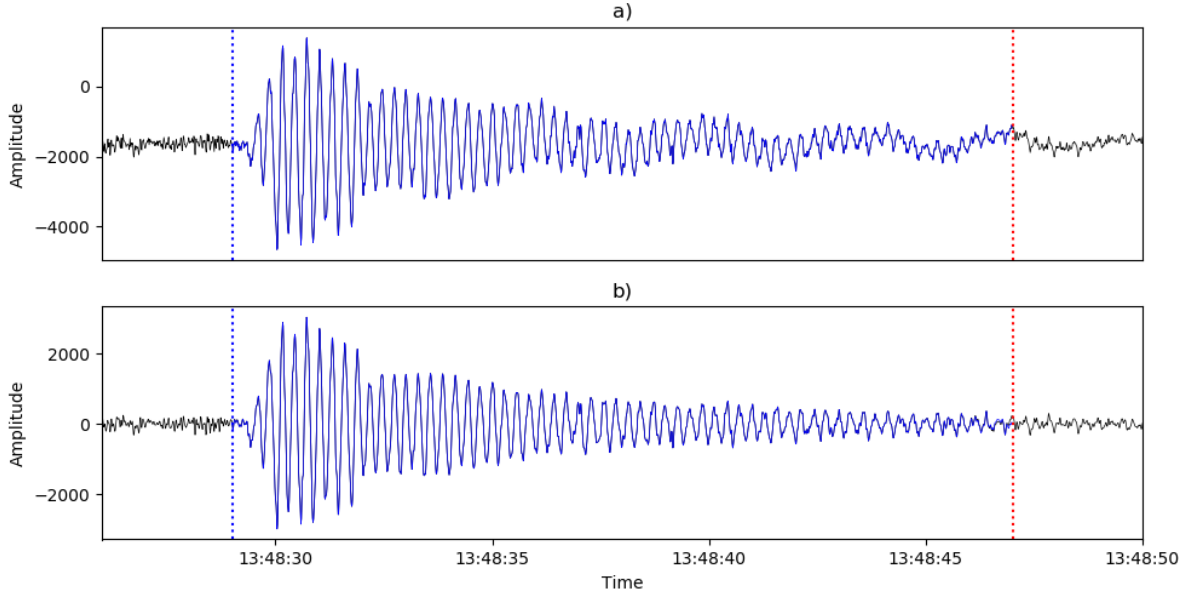


FIGURE 2.2 – a) Forme d’onde non filtrée. b) Forme d’onde filtrée entre 0.8 Hz et 25 HZ

2.3.2 Contrôle qualité

Dans le domaine du machine learning, les données sont primordiales. Pour entraîner le modèle, on a besoin de données de bonne qualité et dont les classes sont bien différenciées les unes des autres. Ici on ne travaille qu’à partir de données enregistrées à une seule station (TAG), or, on sait que tous les événements ne sont pas forcément bien observés à cette seule station. On aura besoin d’observer manuellement les signaux pour vérifier leur classe, cette tâche est plus facile si les signaux ont un rapport signal sur bruit élevé. Pour cela, dans un premier temps, on élimine les signaux de mauvaise qualité. On veut donc appliquer un critère qualité aux signaux pour ne conserver que ceux ayant un rapport signal/bruit prédéfini. Le rapport signal sur bruit est calculé en faisant le rapport entre la moyenne quadratique (RMS) de l’événement sur la moyenne quadratique de 20s de signal avant le temps d’arrivée déterminé par l’observatoire (figure 2.3, équation 2.1).

On a choisi 20 secondes pour calculer la valeur de la RMS du bruit pour deux raisons. La première, le signal bruit doit être suffisamment long pour que sa RMS ne soit pas contrôlée par l’éventuelle présence d’un bout d’événement (donc potentiellement plus fort en amplitude) dans le bruit dû à une erreur sur le temps d’arrivée déterminé par l’observatoire. La seconde raison est d’avoir une durée d’un ordre de grandeur comparable à la durée des événements de la base de données. Après visualisation des signaux et des valeurs de SNR, on établit empiriquement le signal de bonne qualité si le rapport dépasse 1.5, en dessous les signaux sont considérés comme trop bruités pour déterminer avec fiabilité la classe de l’événement et ce même de façon visuelle.

$$SNR = \frac{RMS_{Event}}{RMS_{Noise}} \quad (1)$$

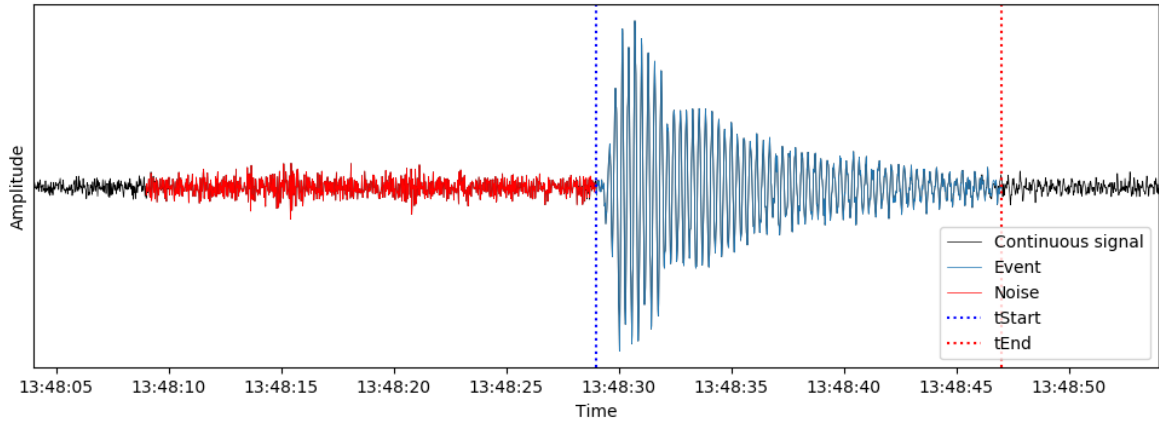


FIGURE 2.3 – Rapport signal sur bruit calculé à partir du bruit 20 secondes (en rouge) avant l'événement (en bleu). Le temps d'arrivée de l'événement (pointillés bleus) et le temps de la fin de l'événement (pointillés rouges) donnés par l'observatoire volcanologique.

2.3.3 Normalisation

Enfin, un dernier prétraitement que l'on pourra choisir ou non d'appliquer est la normalisation des signaux (équation 2.2). On peut choisir de l'appliquer pour, par la suite être capable d'appliquer le modèle à de nouvelles données sans prendre en compte le niveau d'amplitude et donc que le modèle puisse fonctionner quelle que soit l'intensité des événements détectés. On peut également ne pas l'appliquer pour se servir du niveau d'amplitude du signal comme d'un descripteur du signal.

$$S_{norm}[i] = \frac{S[i]}{\sqrt{\sum_{i=1}^n S[i]^2}} \quad (2)$$

2.4 Extraction des descripteurs

Une fois les prétraitements effectués, nous allons présenter les descripteurs qui sont des éléments essentiels pour la classification d'événements sismique.

2.4.1 Transformation des signaux en descripteurs

Les algorithmes de machine learning et notamment de classification peuvent être appliqués à une grande variété de types de données. On peut classer des images, des vidéos, des signaux sonores, etc. Néanmoins, la plupart du temps, on n'utilisera pas la donnée brute pour la classification, soit parce que son format n'est pas forcément adapté aux algorithmes, soit parce que l'information marqueur de la classe n'est pas directement accessible à l'algorithme. Donc, une étape préliminaire avant tout apprentissage est de transformer les signaux bruts dans une nouvelle représentation pouvant être lue par l'algorithme de classification. On appellera ce nouvel espace de représentation, l'espace des features et cette étape de transformation des données, l'extraction des features. Notre base de données n'est dès lors plus un catalogue d'événements sismiques, mais une collection de vecteurs descripteurs de taille D (D étant le nombre de descripteurs utilisés

pour représenter les signaux).

Ainsi, l'algorithme d'apprentissage n'a accès aux données qu'à travers ces vecteurs descripteurs. Les descripteurs doivent donc être soigneusement choisis pour représenter au mieux la variabilité des signaux et doivent être capables de discriminer les classes entre elles dans l'espace des features. Cette tâche d'extraction des features est hardue car elle suppose que l'utilisateur sait sur quels critères les données doivent être séparées.

Pour illustrer l'importance du choix des descripteurs, prenons l'exemple de trois jeux de données simples représentés par deux descripteurs. Sur la figure 2.4a, les données peuvent être séparées grâce au feature 2 mais ne sont pas séparables selon le feature 1 donc le feature 2 est un bon choix de descripteurs mais le feature 1 non. Sur la figure 2.4b les features 1 et 2 seules ne suffisent pas pour séparer en deux classes les données mais ensemble les données deviennent séparables dans l'espace des features. Enfin sur la figure 2.4c, on voit que les deux features ensemble permettent de séparer en deux groupes les données mais la frontière entre les deux classes dans l'espace des features peut être plus complexe qu'une simple droite.

L'enjeu est donc de sélectionner un ensemble de descripteurs qui vont séparer de manière efficace les données selon leurs classes.

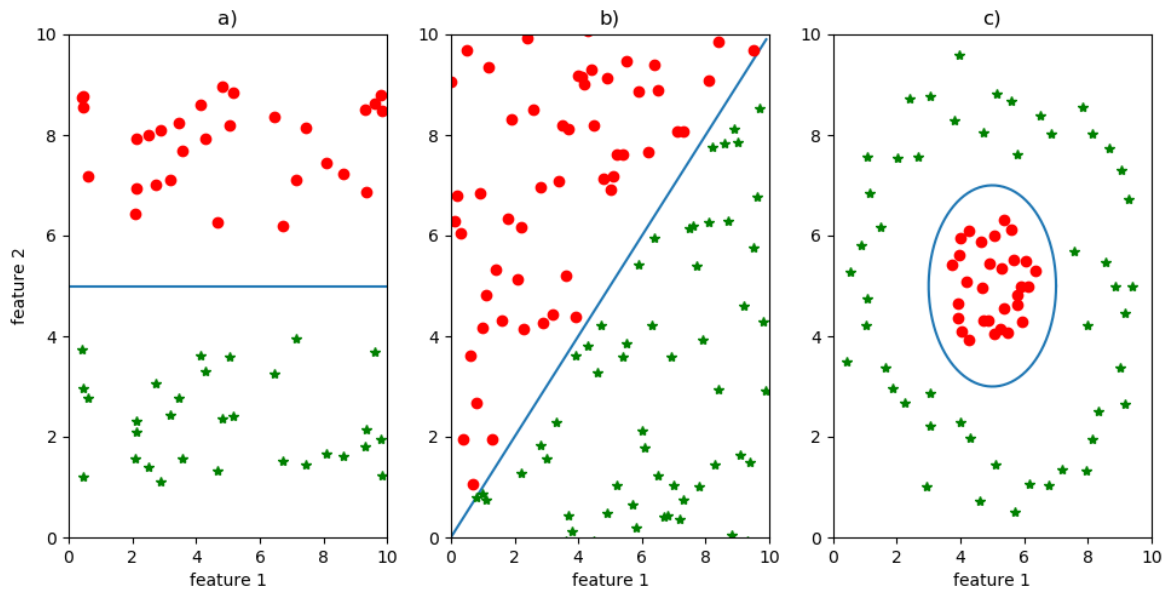


FIGURE 2.4 – Données séparées en 2 classes selon : a) Un seul descripteur. b) Deux descripteurs. c) Une combinaison de deux descripteurs

2.4.2 Descripteurs utilisés

Comme dit plus haut, les données utilisées pour l'apprentissage ne sont pas les formes d'ondes des événements de la base de données mais des descripteurs calculés à partir de ces mêmes formes d'ondes. Ces descripteurs permettent une description physique des données et portent donc des informations liées à la physique des phénomènes donc à la classe des événements. Ici, nous utilisons le jeu de 34 descripteurs proposé par Malfante et al. (2018). La description des descripteurs ainsi que leurs expressions mathématiques sont disponibles dans la Table 2.1. On peut regrouper

ces descripteurs en trois catégories, 9 descripteurs statistiques comme la moyenne, l'écart-type ou encore le kurtosis, 9 descripteurs basés sur l'entropie du signal tel que l'entropie de Shannon ou de Rényi et 16 descripteurs de formes comme le taux d'attaque ou bien des rapports d'amplitude. Ces 34 descripteurs sont calculés dans trois domaines de représentation du signal, le domaine temporel, le domaine fréquentiel et le domaine quéfrentiel. Le domaine quéfrentiel décrit les propriétés périodiques du signal, et est couramment utilisé dans le domaine du traitement de la parole et est obtenu en calculant la transformée de Fourier du logarithme du spectre du signal.

Utiliser un grand nombre de descripteurs permet au signal d'être représenté d'un grand nombre de façon différente, tout en gardant autant d'informations sur le signal d'origine que possible, que ce soit la longueur des signaux, leur contenu fréquentiel ou bien leurs énergies, mais surtout des informations marqueurs des classes que l'on cherche à distinguer. L'autre intérêt majeur de l'extraction des features réside dans la réduction de la dimension des données tout en conservant les informations marqueuses de la classe. Ici le vecteur descripteurs a une dimension de $34 \times 3 = 102$, alors que pour un signal équivalent de 10 secondes échantillonné à 100 Hz, le signal temporel serait de taille $10 \times 100 = 1000$ échantillons. Néanmoins, nous ne nous attendons pas à ce que tous les descripteurs utilisés ici soient utiles. Certains peuvent donner des informations redondantes ou inutiles, par exemple la longueur du signal qui sera la même quel que soit le domaine de représentation du signal.

Une fois que l'on sait quels descripteurs utiliser, on peut passer à l'étape d'apprentissage. Pour cela, on utilise un algorithme d'apprentissage, ici le *Random Forest* qui est particulièrement efficace pour des classifications multiclasses et dans des espaces avec beaucoup de dimensions.

2.5 Random Forest

L'étape d'apprentissage consiste à définir les frontières entre les différentes classes dans l'espace des features. Pour cela, l'algorithme d'apprentissage sélectionne les descripteurs les plus discriminants parmi les 102 descripteurs (34×3). L'algorithme utilisé ici est une forêt aléatoire d'arbres de décisions, ou *Random Forest*, cet algorithme est disponible dans la bibliothèque python scikit-learn (`sklearn.ensemble.RandomForestClassifier`). En section 3.1, la Table 3.1 montre les résultats de différents algorithmes classifieur. Les résultats étant équivalents, nous n'explicitons que le RF car c'est celui majoritairement utilisé au cours de cette thèse pour des raisons expliquées ci-dessous.

L'algorithme Random Forest (RF) se base sur des arbres de décision binaire présentés par Quinlan (1986) qui a proposé l'algorithme CART (*Classification And Regression Trees*) pour séparer l'espace des features en de simples régions. Chaque région étant associée à la classe la plus représentée à l'intérieur. Des tests itératifs sont appliqués aux descripteurs représentant les données à classer, et ainsi, l'algorithme détermine à quelle région de l'espace des features correspond cette donnée et donc sa classe. La figure 2.5 illustre le fonctionnement d'un arbre de décision.

Le concept derrière l'algorithme est de minimiser l'impureté de chaque région, c'est à dire de

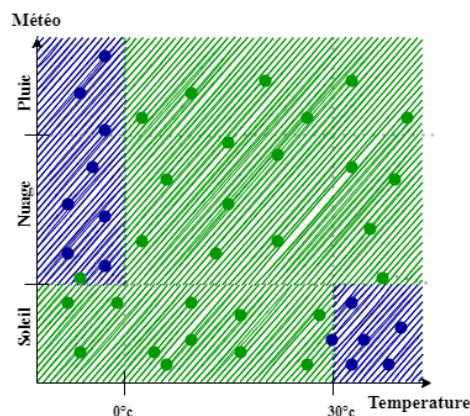
TABLE 2.1 – Descripteurs utilisés pour représenter les signaux, set de descripteurs provenant de Malfante et al. (2018).

Features	Definition	Index
Statistic features		
Length	$n = \text{length}(s)$	1
Mean	$\mu_s = \frac{1}{n} \sum_i s[i]$	2
Standard deviation	$\sigma_s = \sqrt{\frac{1}{n-1} \sum_i (s[i] - \mu_s)^2}$	3
Skewness	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^3$	4
Kurtosis	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^4$	5
i of central energy	$\bar{i} = \frac{1}{E} \cdot \sum_i E_i \cdot i$	6
RMS bandwidth	$B_i = \sqrt{\frac{1}{E} \sum_i i^2 \cdot E_i - \bar{i}^2}$	7
Mean skewness	$\sqrt{\frac{\sum_i (i - \bar{i})^3 E_i}{E \cdot B_i^3}}$	8
Mean kurtosis	$\sqrt{\frac{\sum_i (i - \bar{i})^4 E_i}{E \cdot B_i^4}}$	9
Entropy features (with $p(s_j)$ the probability of amplitude level s_j)		
Shannon entropy ^a	$-\sum_j p(s_j) \log_2(p(s_j))$	10 to 12
Rényi entropy ^b	$\frac{1}{1-\alpha} \log_2(\sum_j p(s_j)^\alpha)$	13 to 18
Shape descriptors features		
Rate of attack	$\max_i \left(\frac{s[i] - s[i-1]}{n} \right)$	19
Rate of decay	$\max_i \left(\frac{s[i] - s[i+1]}{n} \right)$	20
Ratios	min/mean and max/mean	21 to 22
Energy descriptors	Signal Energy, maximum, average, standard deviation, skewness and kurtosis	23 to 28
Specific values	min, max, i of min, i of max, threshold crossing rate and silence ratio	29 to 34

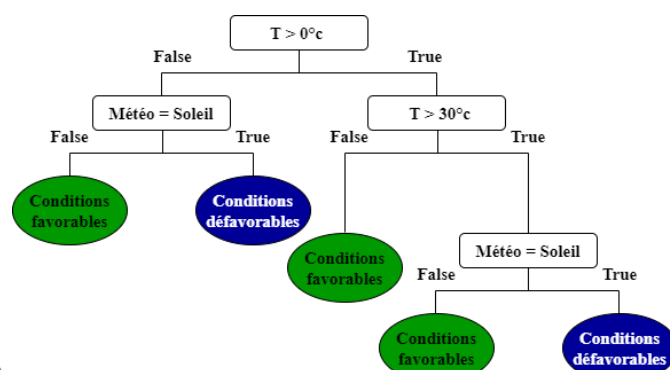
Note. Features computed for a signal $s[i]_{i=1}^n$ (in which i might correspond to a temporal, frequency or cepstral sample). $E = \sum_{i=1}^n s[i]^2$ and $E_i = s[i]^2$ describe the signal energy and the energy at sample i , respectively. Some features have a dimension greater than others; *e.g.*, entropy measurements are made on three different estimations of the amplitude probability (*i.e.*, different histogram bin numbers).

^aBin numbers for probability estimation : 5, 30 and 500.

^bBin numbers for probability estimation : 5, 30, 500, $\alpha = 2$, inf.



(a) Séparation au sein de l'espace des descripteurs



(b) Arbre de décision

FIGURE 2.5 – Illustration d'un arbre de décision prédisant si les conditions atmosphériques sont bonnes ou mauvaises pour faire un match de football à partir de deux descripteurs, la météo et la température. Les données correspondantes à de bonnes conditions sont représentées en Vert et en Rouge lorsque les conditions sont mauvaises. La séparation au sein de l'espace des features en a) et l'arbre de décision associé en b).

prioriser des régions dans lesquelles les données sont de la même classe aux régions dans lesquelles il y a des données de différentes classes. La fonction de coût associée à cet algorithme est donc une mesure de l'impureté des différentes régions, on pourra utiliser l'*impureté de Gini*.

Le processus de construction d'un arbre est itératif et est décrit dans l'algorithme 1. Tant que le critère d'arrêt n'est pas atteint une nouvelle séparation est créée dans l'arbre. Pour trouver la meilleure séparation, on utilise tous les descripteurs. Pour chaque descripteur, tous les points de séparation sont considérés. Le descripteur sélectionné est celui pour lequel la nouvelle région a le score d'impureté le plus bas. Le critère d'arrêt peut être la profondeur de l'arbre, le taux d'erreur dû à cette séparation sous un certain seuil ou encore la taille d'une région atteignant un certain seuil.

Les arbres de décision sont populaires car simple d'interprétation. Là où certains algorithmes d'apprentissage sont vus comme des boîtes noires, les arbres de décision peuvent faciliter l'interprétation des données ou du modèle. Les temps de calcul des arbres de décision sont assez courts. L'algorithme est déterministe, donc les mêmes jeux de données donneront les mêmes modèles. Si l'arbre est assez profond, le modèle pourra être sans biais, il ne fera pas d'erreur systématique. Mais la variance peut être importante, deux jeux de données d'apprentissage légèrement différents peuvent donner deux modèles très différents. Ce dernier point est cependant une limitation majeure de l'algorithme, un modèle se doit d'être stable malgré de légères modifications du jeu d'apprentissage.

C'est pour répondre à ce dernier point que le Random Forest a été développé (Breimann, 2001). L'idée principale consiste à créer un grand nombre d'arbres de décision et la prédiction finale sera donnée par un vote des arbres de la forêt. On pourra utiliser des dizaines, voire des centaines d'arbres de décision. Augmenter le nombre d'arbres permet la généralisation du modèle jusqu'à

Algorithm 1 Random Forest algorithm

while *Stopping Criteria* = *False* **do** **for** each feature x^j **do** **for** each split point s **do** Define a pair of half plane (two regions R_1 and R_2) in the features space :

$$R_1(j, s) = \{\mathbf{x}_i | x_i^j \leq s\}$$

$$R_2(j, s) = \{\mathbf{x}_i | x_i^j > s\}$$

 Define the associate class proportions $pR_{1,c}(j, s)$ and $pR_{2,c}(j, s)$:

$$\forall c \in [1, C], pR_{1,c}(j, s) = \frac{1}{N_{R1}} \cdot \sum_{\mathbf{x}_i \in R_1} (y_i = c)$$

$$\forall c \in [2, C], pR_{2,c}(j, s) = \frac{1}{N_{R2}} \cdot \sum_{\mathbf{x}_i \in R_2} (y_i = c)$$

end for

Find the best split point :

$$s = \operatorname{argmax}_s (\max_c pR_{1,c}(j, s) + \max_c pR_{2,c}(j, s))$$

end for Find the best feature x^j :

$$j = \operatorname{argmax}_j (\max_{c1} pR_{1,c}(j, s) + \max_{c2} pR_{2,c}(j, s))$$

end while

un certain point, mais augmente également les coûts de calcul. Chaque arbre est construit en utilisant un sous-jeu du jeu d'apprentissage original, donc chaque arbre est différent.

Un autre avantage du RF est de pouvoir obtenir des informations sur l'importance des descripteurs dans la classification. Cette information est obtenue en calculant le *mean decrease in impurity*, on s'intéresse à la baisse d'impureté moyenne due à chaque descripteur. Cela permet de réduire le nombre de descripteurs en sélectionnant seulement les plus importants pour l'apprentissage mais aussi de faire le lien entre les descripteurs et la physique génératrice de ces événements.

Enfin, pour avoir le meilleur modèle possible, on fait une recherche de grille sur les hyper-paramètres du random forest et on sélectionne le set d'hyper-paramètres qui donne des résultats bons et stables, c'est-à-dire que les résultats sont peu influencés par le jeu d'apprentissage et les hyper-paramètres peuvent être ajustés sans modifications excessives des résultats.

Maintenant que le modèle est construit il faut le tester et voir s'il est efficace.

2.6 Validation du modèle

2.6.1 Validation croisée

Pour tester les performances du modèle, on peut effectuer une validation croisée. On partitionne aléatoirement les données entre un jeu d'apprentissage et un jeu de test selon différents rapports

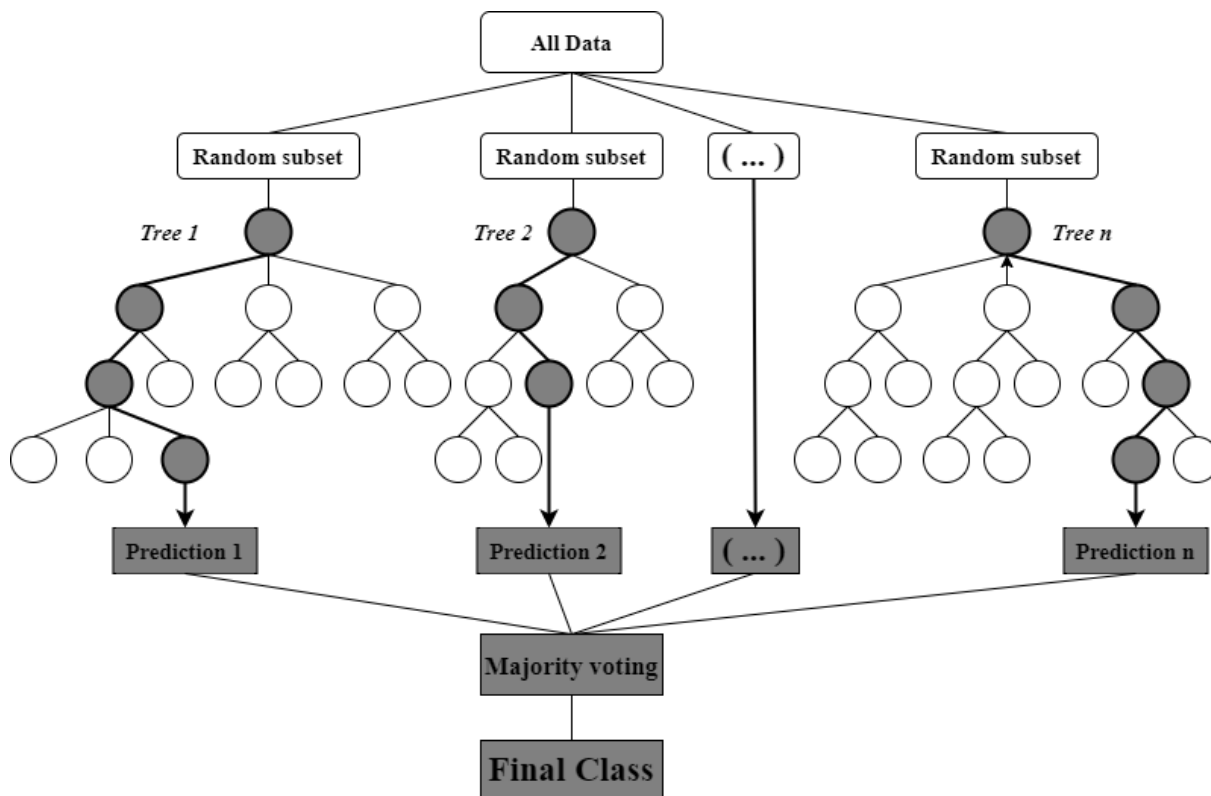


FIGURE 2.6 – Random Forest : Un sous jeu de données pour chaque arbre. La prédiction finale est obtenue en moyennant les résultats de chaque arbre.

entre ces deux jeux, 20/80 jusqu'à 80/20 pour voir les effets de la taille du jeu d'apprentissage sur les performances. Après apprentissage d'un modèle, on fait des prédictions de classe pour le jeu de test, et on compare ces prédictions avec la vraie classe des événements labellisés du jeu de test. On répète ces étapes de partitionnement des données et de prédictions plusieurs fois, en ayant à chaque fois des jeux d'apprentissage et de test différents (Figure 2.6) On fera attention lorsqu'une classe est sous représentée dans le jeu de données initial que le nombre d'événements de cette classe soit équilibré dans chaque partition. Ensuite, on peut analyser les résultats de cette validation croisée à l'aide de différentes métriques. Cette étape de validation de croisée nous permettra également de vérifier si notre modèle a des problèmes de sous-apprentissage ou de sur-apprentissage. Le sous-apprentissage correspondant à un modèle incapable de fournir des prédictions précises. Et le sur-apprentissage apparaît quand un modèle est trop associé à certains jeux de données et qu'il ne peut pas prédire de manière fiable de nouvelles observations.

2.6.2 Métriques et Matrice de confusion

Après l'étape de validation croisée, on obtient une matrice de confusion moyenne qui indique l'organisation des classes prédites par rapport à la vraie classe des données de test.

Avec cette matrice de confusion (Tableau 2.2), on peut étudier en détail les résultats classe par classe, mais on peut également étudier quelques tendances générales dans les résultats grâce à différentes métriques, accuracy, recall, precision :

- Accuracy : Mesure le nombre de prédictions correctes par rapport au nombre de prédic-

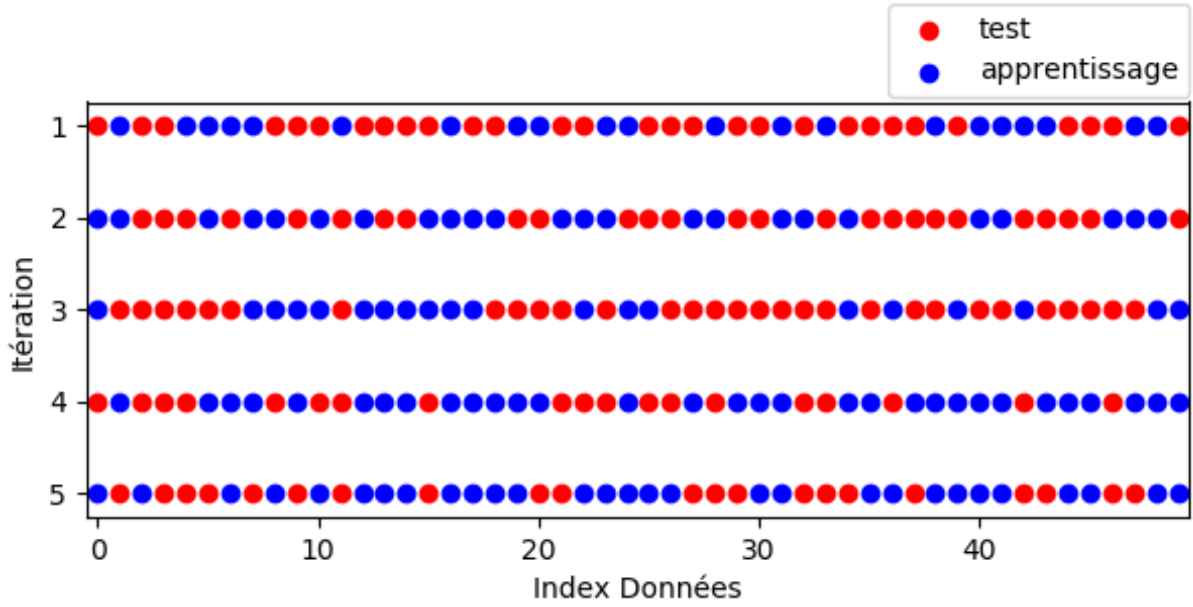


FIGURE 2.7 – Validation croisée : Une partie des données labellisées est aléatoirement sélectionnée puis utilisée pour l'apprentissage et l'autre partie pour tester les résultats. On répète cette étape plusieurs fois pour avoir des résultats statistiquement consistant, ici 50% des données sont utilisées pour l'apprentissage et 50% pour le test.

tions total, donne des informations sur les performances générales du modèle. Comme on a préalablement fait une validation croisée, on obtient cette métrique pour chaque réalisation et donc un écart-type sur ces différentes réalisations. Si cet écart-type est élevé alors il y a une disparité entre les différentes réalisations et il peut exister des données abaissant les performances du modèle.

$$Accuracy = \frac{\#GoodPrediction}{\#TotalEventTestDataset} \quad (3)$$

- Recall : Mesure au sein d'une classe α la proportion d'événement bien classée. Si cette métrique est basse alors les événements d'une classe ont été mal classés et il y a trop de faux négatifs parmi les résultats.

$$Recall = \frac{\#GoodPredictionClass_i}{\#TotalTrueClass_i} \quad (4)$$

- Precision : C'est une métrique donnant parmi les classes prédites, la proportion de prédictions correctes. Si cette métrique est basse alors il y a trop peu de bonnes prédictions par rapport aux mauvaises, il y a trop de faux positifs parmi les résultats.

$$Precision = \frac{\#GoodPredictionClass_i}{\#TotalPredictedClass_i} \quad (5)$$

TABLE 2.2 – *Exemple d'une matrice de confusion.*

True class	Predicted Class			Recall (%)
	Class 1	Class 2	Classe n	
Class 1	Good	Bad	Bad	
Class 2	Bad	Good	Bad	
Class n	Bad	Bad	Good	
Precision (%)				
Accuracy (%)				±

2.7 Résumé

Dans ce chapitre, on a vu les grandes étapes de la construction d'un modèle de classification supervisée :

- Le jeu de données : Tout d'abord, on a besoin d'un jeu de données labellisées de bonne qualité représentant la variabilité des classes que l'on peut rencontrer.
- Prétraitements : On effectue quelques étapes de prétraitements (filtrage, fenêtrage, contrôle qualité et normalisation) pour mettre en avant dans les signaux les informations liées à la classe de l'événement.
- Extraction des features : L'algorithme d'apprentissage ne peut pas apprendre directement à partir des formes d'ondes, il faut donc trouver une représentation des données qui peut être lue par l'algorithme et conservant les informations liées à la classe des événements, ce nouvel espace de représentation est l'espace des descripteurs. Ici, on utilise un set de 102 descripteurs calculés à partir de la forme d'onde dans trois domaines de représentation (domaine temporel, fréquentiel et quéfrentiel).
- Apprentissage : Un algorithme de machine learning (ici le Random Forest) va lire nos données d'entraînement et va construire des frontières dans l'espace des descripteurs pour discriminer chacune des classes.
- Validation : On répète l'étape d'apprentissage en utilisant à chaque réalisation des données différentes et on compare les prédictions faites par le modèle avec la vraie classe des événements. Et à l'aide de différentes métriques et de la matrice de confusion, on peut connaître les performances générales du modèle ainsi que classe par classe.
- Application du modèle : Une fois que l'on a construit un modèle fiable, on peut l'appliquer à de nouvelles données non labellisées et déterminer leur classe.

Maintenant que la méthode est claire, on peut l'appliquer sur le volcan de La Soufrière de Guadeloupe et analyser la classification actuelle et potentiellement l'améliorer.

3 Classification automatique des signaux sismo-volcaniques de la Soufrière de Guadeloupe

Dans ce chapitre, nous construirons un modèle en entraînant un RF à partir du catalogue établi par l'OVSG, puis nous ajusterons ce catalogue pour maximiser le taux de reconnaissance du modèle. Ensuite, nous verrons les descripteurs les plus utiles pour représenter les signaux de la Soufrière. Enfin, nous appliquerons le modèle construit sur la période 2013-2018, sur la période 2018-2020.

Une partie des résultats présentés dans ce chapitre a fait l'objet d'une publication dans le Journal of Volcanology and Geothermal Research, celle-ci est disponible en annexe.

3.1 Sélection d'un modèle

Dans cette section, nous testons différents algorithmes d'apprentissage pour déterminer lequel sera le plus pertinent par la suite. Le Tableau 3.1 recense les différents types d'algorithme utilisés, le score associé ainsi que les meilleurs hyper-paramètres pour la classification. On a choisi d'utiliser 845 événements ayant passé le contrôle qualité décrit en section 2.3.2, les événements sont répartis en trois classes différentes, telles que décrite en Partie 1. On choisit d'entraîner les modèles avec 50% des données et de tester les modèles avec les 50% restants. On a effectué une recherche de grille pour trouver les meilleurs hyperparamètres de chaque type d'algorithme.

On voit que dans notre cas les résultats dépendent peu du type d'algorithme de classification choisi, en effet, le Support Vector Machine nous donne un score de 72% tout comme le Random Forest. Le K-Nearest Neighbors atteint un score de 71% comme le Quadratic Discriminant Analysis. Et enfin, le meilleur score est de 74% pour le Linear Discriminant Analysis. Tous ces algorithmes sont disponibles dans la bibliothèque scikit-learn de python.

Étant donné le peu d'écart entre les différents résultats, nous choisirons d'utiliser un Random Forest pour les raisons expliquées dans la partie 2.5. Maintenant que nous avons sélectionné un type d'algorithme, nous allons nous intéresser aux résultats de la classification plus en détail.

TABLE 3.1 – Score obtenu pour différent algorithmes de classification, SVM : Support vector machine, KNN : K nearest neighbors, QDA : Quadratic discriminant analysis, LDA : Linear discriminant analysis, RF : Random Forest

Algorithme d'apprentissage	Score	Meilleurs hyperparamètres
SVM	72%	'C' : 1.0, 'gamma' : 0.0625, 'kernel' : 'rbf'
KNN	71%	'n_neighbors' : 13
QDA	71%	'reg_param' : 0.0034482758620689659
LDA	74%	'shrinkage' : 0.17241379310344829
RF	72%	'n_estimators' : 140

3.2 Performances détaillées de la classification [2013-2018]

Ici, nous présentons la performance de la classification automatique de trois classes d'événements sismo-volcaniques LP, VT et Emboîtés en utilisant un algorithme de classification de type Random Forest. A la station TAG (Figure 1.1), beaucoup d'événements détectés ne respectent pas le critère qualité de l'étape de prétraitement expliqué en section 2.3.2. L'OVSG a détecté 7 149 événements sur la période 2013-2018 mais seulement 845 ont passé cette phase de sélection. Cette phase de sélection nous permet par la suite de pouvoir facilement vérifier visuellement si les événements que nous analysons sont correctement labellisés. Si on laissait tous les événements même ceux très bruités qui sont difficiles à classer à la station TAG, on ne devrait se fier qu'à la classification construite par l'OVSG or cela nous intéresse de repérer si des événements ont pu être mal classés par l'observatoire. Par ailleurs, pour des raisons de coût de calcul, il sera plus facile de traiter un peu moins de données. L'analyse se fait avec 542 événements VT, 217 emboîtés et 86 LP. Le Tableau 3.2, montre la matrice de confusion moyenne obtenue après dix réalisations de l'analyse avec des rapports de 50-50% entre les jeux d'entraînement et les jeux de test. On rappelle que pour chaque réalisation, on sélectionne de façon aléatoire les données utilisées pour l'entraînement et on teste le modèle avec les données restantes.

TABLE 3.2 – Matrice de confusion obtenue avec un rapport de 50 :50 entre les jeux d'entraînement et de test pour le catalogue de l'OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest (n_estimators=100, criterion='entropy', bootstrap=True, class_weight=None), Score moyen obtenu après 10 réalisations

True class	Predicted Class			Recall (%)
2-4	VT	Nested	LP	
VT	251	16	4	92.6
Nested	71	33	4	30.6
LP	16	5	2	8.70
Precision (%)	74.3	61.1	20.0	
Accuracy (%)				72.5 ±1.0

Nous obtenons une accuracy de 72.5 +/- 1%, ce qui est plutôt bas comparé aux scores obtenus dans la littérature qui peut atteindre 80% ou 90%. Différents rapports entre le jeu d'entraînement et le jeu de test ont également été testés. Entre des rapports de 80-20% et de 30-70%, les résultats étaient quasiment les mêmes, autour de 72.5 +/- 1%. Au-delà de 80% des données utilisées pour l'entraînement et en dessous de 30% l'accuracy diminue et des problèmes de sur-apprentissage et de sous-apprentissage commencent à apparaître. Quand on analyse les résultats plus en détails, on peut détecter certaines tendances.

L'accuracy de 73% est largement dû à la taille de la classe VT comparé au reste du jeu de données. Cela masque un peu les mauvais résultats des autres classes. Le recall pour la classe VT est de 93% donc la plupart des vrais VT sont bien classés, mais la précision pour cette classe est de seulement 74%, cela veut dire que le modèle prédit comme VT des événements qui sont considérés comme Emboîtés ou LP par l'observatoire. Si on regarde un peu plus en détail, 71 des 108 Emboîtés utilisés pour le test sont classés comme VT par le modèle, de la même façon 15 des 23 LP sont classés comme des événements VT. La mauvaise précision de la classe VT peut

s'expliquer de deux manières. La classe VT est trop large et comprend une variété de signaux qui couvre les caractéristiques des classes Emboîtés et LP ou au contraire, il y a des signaux dans les classes LP et Emboîtés qui présentent des caractéristiques de la classe VT.

La précision pour la classification de la classe Emboîtés est de 61%, donc les Emboîtés prédit par le modèle ne sont pas très fiables. Le recall, lui est de 31%, donc la plupart des événements considérés comme Emboîtés par l'OVSF ne sont pas bien reconnus par le modèle. En effet, 71 des 108 Emboîtés sont classés comme des événements VT par le modèle et 4 des 108 Emboîtés sont considérés comme LP.

La précision de la classification pour la classe LP est de 20%, cela signifie que les prédictions LP du modèle sont peu fiables. De plus, le recall de 9% pour cette classe montre que la plupart des événements considérés comme LP par l'OVSF ne sont pas reconnus comme tel par le modèle. La plupart des mauvaises classifications (21 sur 23) de la classe LP sont classées comme VT par le modèle (16 sur 21).

Ces résultats montrent que les événements VT peuvent être confondus avec les événements LP et Emboîtés, alors que les événements LP et Emboîtés sont rarement confondus entre eux. La confusion dans la classification des VT et des Emboîtés est partiellement due au fait que les événements Emboîtés se composent de plusieurs événements VT individuels qui sont séparés par de très courts délais inter-événements (tel que la coda du premier signal est mélangée avec l'arrivée du signal suivant). Dans d'autres cas, beaucoup de VT présentent ce qui semble être une phase P émergente sur la trace sismique de basse résolution. Mais, quand on zoome sur la forme d'onde dans l'environnement Seiscomp3 pendant la phase de pointé, il est clair que cette phase P apparente est en fait un premier signal VT de basse amplitude rapidement suivi par un autre événement VT d'amplitude plus forte avec des phases P et S. Ainsi si les événements Emboîtés sont des groupes de signaux VT, il est possible que la discrimination entre ces deux classes puisse être difficile, car la description donnée par les descripteurs des Emboîtes correspond à celle des VT.

A l'OVSF, l'outil de visualisation utilisé pour classer manuellement les événements rend ce travail compliqué car les spectrogrammes ne sont pas immédiatement accessibles aux opérateurs, de plus, les plots des traces sont basse résolution et il est compliqué de zoomer sur les formes d'onde car on ne fait que zoomer sur des pixels. Donc certains événements pourraient avoir été mal classés. Cela expliquerait la large variété de signaux au sein de la classe VT. C'est pourquoi nous avons décidé d'analyser plus en détail la labellisation des données à travers une nouvelle visualisation des signaux en s'aidant du spectrogramme et une nouvelle phase de labélisation basée sur les domaines fréquentiels des VT (5-20 Hz) et des LP (0.1-5 Hz).

3.3 Performances de la classification affinée [2013-2018]

3.3.1 Nouvelle classification

Dans le but d'améliorer les performances de notre modèle, nous avons visuellement revu tous les 845 événements pour vérifier si certains événements étaient mal classés, si nous observions

des signaux montrant des caractéristiques différentes des classes d'événement déjà considérées et si de nouvelles classes devaient être introduites dans notre analyse. Après la revue manuelle de tous les signaux, nous avons décidé d'introduire deux nouvelles classes, la classe Hybride et la classe LP monochromatique également appelée Tornillo (Figure 3.1). Ces types d'événements sont considérés par l'OVSG sous la forme d'un commentaire mais ne sont pas pris en compte durant le travail journalier de classification. Ici, nous allons rajouter spécifiquement ces deux classes dans notre étape d'apprentissage.

Séismes hybrides : De nombreuses analyses détaillées de séismes volcaniques durant de récentes éruptions (Redoubt, Lahr et al. 1994 ; Soufriere Hills, Miller et al. 1998 ; Redoubt, White 1998) ont montré qu'une catégorie distincte de séismes possède des caractéristiques appartenant à la fois aux classes VT et LP. Ces événements sont appelés Hybrides (Chouet et Matoza, 2013). Typiquement, ils présentent une arrivée impulsive haute-fréquence entre 10 Hz et 20 Hz qui est caractéristique des séismes VT, alors que la coda est dominée par des ondes de plus basses fréquences comme celles des LP (1.2-2.5 Hz, au Redoubt, Lahr et al. 1994, à Soufriere Hills, Montserrat, White et al. 1998). Cette composante LP est observée depuis le début jusqu'à la fin de ces événements. A la Soufrière, les séismes Hybrides (Figure 3.1) ont une coda basse fréquence autour de 5-6 Hz et une première arrivée plus haute fréquence aux alentours de 12-14 Hz. Différentes études ont cherché à comprendre ces événements Hybrides. Lahr et al. (1994) ont proposé un modèle à la croisée de celui des VT et de celui des LP avec en premier lieu un processus de fracturation fragile produisant des hautes fréquences, suivi de la propagation de fluide responsable d'un phénomène de résonance produisant de basses fréquences. Neuberg et al. (2000) ont montré un clair continuum entre les événements LP et Hybrides. Harrington et al. (2007) ont proposé un modèle dans lequel un simple processus de fracturation avec une vitesse de rupture très lente suffisait pour expliquer ce type d'événement.

Séismes LP monochromatiques (ou Tornillos) : Ces Tornillos sont une sous-catégorie des LP, mais leur forme d'onde particulière (qui ressemble à une vis, tornillo en espagnol) les rendent facilement distinguables. Ils ont été décrits pour la première fois sur le volcan Galeras par Gomez et al. (1997), et ont par la suite été identifiés lors de nombreuses éruptions. Ces événements présentent une arrivée d'onde émergente sur une durée allant jusqu'à plusieurs dizaines de secondes, un signal quasiment sinusoïdal et une coda dont l'amplitude décroît très lentement et de façon presque linéaire. Leur spectre de Fourier présente un pic caractéristique aux alentours de 4 Hz pour la Soufrière. Sur le spectrogramme (Figure 3.1), on voit que la résonance dure plus longtemps que celle des signaux LP. Un modèle d'oscillation de fluide remplissant une cavité a été proposé comme processus physique pouvant générer ces Tornillo (Gomez et al., 1997 ; Konstantinou et al. 2015). Dans une étude expérimentale récente, Fazio et al. (2019) ont suggéré que le gradient de pression du gaz déclenche des événements qui régulent la faible décroissance de la coda, alors que la résonance de fluide dans de petites structures contrôle le contenu fréquentiel du signal. Ce type d'événement a été introduit en 2001 à l'OVSG, mais a par la suite été retiré de la routine de classification.

Le Tableau 3.3 montre les classes qui ont été nouvellement assignées dans notre catalogue. Premièrement, 89 événements Emboîtés et 15 événements LP ont été reclassés comme des VT. Ensuite, 39 VT et 22 Emboîtés ont été reclassés comme des événements LP. Seulement 9 VT et

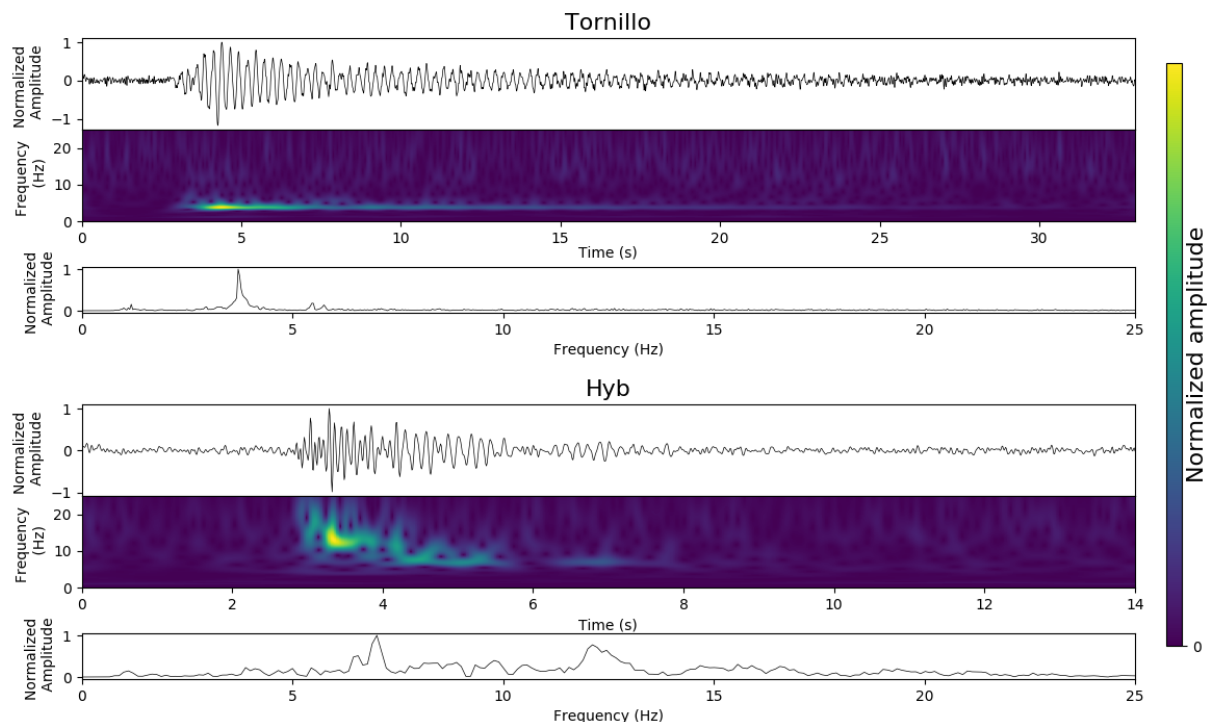


FIGURE 3.1 – Exemple de forme d’onde filtrée entre 0.8-25 Hz, Spectrogramme et Spectre de Fourier des événements Tornillo (en haut) et Hybride (en bas) enregistrés à la station TAG à la Soufrière de Guadeloupe sur la composante verticale

1 LP ont été reclassés comme des événements Emboîtés. Également, 2 Emboîtés et 26 LP ont été reclassés comme des événements Tornillo. Et enfin, 151 VT, 45 Emboîtés et 8 LP ont été reclassés comme événements Hybrides.

TABLE 3.3 – Nouvelles classes d’événements ayant été manuellement réassignés après visualisation basée sur la forme d’onde, le spectrogramme et le spectre de Fourier.

Ancienne classe	Nouvelle classe					Total
2-6	VT	Emboîté	LP	Hyb	Tor	
VT	343	9	39	151	0	542
Emboîté	89	59	22	45	2	217
LP	15	1	36	8	26	86
Total	447	69	97	204	28	845

3.3.2 Performances détaillés de la nouvelle classification

On peut noter que très peu des événements Emboîtés sont restés dans cette classe (59 sur 217), donc beaucoup ont été reclassés comme des événements VT (89 sur 217). Les événements Emboîtés sont une classe de séisme volcanique plutôt unique, car ils n’ont été observés qu’à la Soufrière et à la Montagne Pelée en Martinique. Il n’y a pas d’événement similaire aux Emboîtés dans la littérature, donc, pour maintenir une certaine robustesse dans notre nouvelle labellisation et pour ne pas confondre ces événements avec des VT, nous avons choisi des critères explicites pour décider à quelle classe ces événements appartenaient. Nous considérons les événements comme Emboîtés quand ils présentent clairement plusieurs arrivées dans la forme d’onde. Comme la revue des événements s’est faite sur une seule station, il est possible qu’un événement puisse

présenter plusieurs arrivées sur une station, mais pas sur une autre, car ces signaux Emboîtés peuvent être différents en terme d’amplitude et de forme d’onde selon le site d’acquisition.

Lorsque nous avons créé un modèle avec ces nouvelles classes et le même rapport entre jeu d’entraînement et jeu de test que précédemment, c’est à dire 50 :50, notre accuracy a augmenté de 73% à 82% $\pm$ 2%. Cela montre que cette nouvelle classification est meilleure pour le traitement automatique. De plus, en regardant les résultats plus en détail (Tableau 3.4), on peut voir que la classification pour certaines classes a été nettement améliorée. Nous avons testé différents rapports entre jeu d’entraînement et jeu de test tel que 90 :10, 70 :30, 30 :70, et 10 :90, les scores obtenus sont respectivement : 83% $\pm$ 4%, 83% $\pm$ 2%, 82% $\pm$ 1%, and 77% $\pm$ 2%; *i.e.* cela reste similaire.

TABLE 3.4 – Matrice de confusion obtenue avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue de l’OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest (n_estimators=100, criterion=’entropy’, bootstrap=True, class_weight=None), Score moyen obtenu après 10 réalisations

True class	Predicted Class					Recall (%)
	VT	Nested	LP	Hyb	Tor	
VT	203	2	1	18	0	90.6
Nested	5	23	1	5	0	67.7
LP	1	1	41	6	0	83.7
Hyb	25	2	7	69	0	67.0
Tor	0	1	1	0	13	86.7
Precision (%)	86.8	79.3	80.4	70.4	100	
Accuracy (%)						82.1 $\pm$ 1.6

Ensuite, nous avons regardé les résultats classe par classe pour analyser les erreurs. Elles étaient nettement plus faibles pour la classe Emboîté, avec 68% et 79% de recall et de précision au lieu de 31% et 61% pour la classification précédente. Un seul vrai événement LP a été classé comme un événement Emboîté, et *vice versa*. Cinq Emboîtés ont été classés comme des événements VT et cinq comme des événements Hybride. Ces confusions ont été observées lorsqu’il y avait peu d’événements emboîtés dans le signal (*i.e.* 2 ou 3) et quand un de ces signaux était plus énergétique que les autres et donc que le modèle classe ce signal énergétique. Les six événements qui ont été mal classés comme des événements Emboîtés étaient des événements particulièrement bruités.

Les événements Tornillos sont très bien classés par le modèle, avec 100% et 87% de recall et de précision. Il y a eu un vrai Tornillo classé comme LP. Une fois encore, ce genre d’erreur se comprend facilement, car la classe Tornillo est une sous-classe de la classe LP. La plus grosse différence entre ces deux classes est la longue coda monochromatique résonante dans les événements Tornillos. Selon la forme de la coda, on peut choisir de classer un événement comme LP ou Tornillo et il arrive que cette distinction soit dure à faire. L’autre Tornillo mal classé a été labellisé par Emboîté par le modèle, c’était le Tornillo le plus bruité du catalogue. Enfin, les différents tests effectués avec différents rapports entre jeu d’entraînement et jeu de test ont montré qu’on n’avait pas besoin de beaucoup d’événements dans la base d’entraînement pour reconnaître de manière fiable les événements Tornillo.

Le recall des prédictions des événements LP a augmenté de 51% à 84%. Un vrai VT, un vrai Emboîté, un vrai Tornillo et sept vrais Hybrides ont été classés comme des événements LP par le modèle. La précision des LP a augmenté de 73% à 80%. Un seul vrai LP a été classé comme VT, un comme Emboîté et six comme Hybride. Les événements LP peuvent être confondus avec des VT, des Hybrides et des Tornillos. La confusion entre LP et VT a énormément baissé par rapport à l'ancienne classification avec 16 événements LP sur 43 classés comme VT auparavant et seulement 1 sur 49 avec cette nouvelle classification.

Avec la nouvelle classification, la classe VT présente un recall de 91% et une précision de 87%. Si le recall a légèrement baissé de 6%, la précision a augmenté de 74% à 87%. Ces résultats montrent que la variété de signaux au sein de la précédente classe VT a diminué.

La classe la plus compliquée à classer est la classe Hybride. Le recall est de 67% et la précision est de 70%. Les événements Hybrides sont seulement confondus avec les VT et les LP, avec 25 vrais Hybrides classés comme VT et 7 comme LP. Ici, 25 Hybrides prédits sont des VT et 6 Hybrides prédits sont des LP. Les événements Hybrides sont plus souvent confondus avec les VT qu'avec les LP. Cela peut être expliqué par le caractère impulsif et hautes fréquences des premières arrivées des événements Hybrides, ce qui résulte dans des confusions plus fréquentes entre Hybride et VT qu'entre Hybride et LP. Ces résultats renforcent l'idée que les événements Hybrides sont un continuum entre les événements LP et VT. Même si les scores de la classe Hybride sont plus faibles que le score des autres classes, cela reste encourageant car ces événements sont les plus difficiles à reconnaître pour un opérateur humain.

Cette nouvelle classification semble être très efficace et permet à notre algorithme automatique de distinguer les différentes classes d'événements plus facilement que la précédente classification. De plus, deux nouvelles classes d'événement, les Hybrides et les Tornillo ont été introduites pour mieux caractériser la sismicité. On peut noter cependant un certain déséquilibre dans le nombre d'événements de chacune des classes utilisées pour construire le modèle, avec notamment 447 VT contre seulement 28 Tornillos. Des méthodes existent pour compenser ces problèmes de classes déséquilibrées. Rahmani et al. (2017) ont montré qu'utiliser une méthode de clustering pour séparer la classe la plus nombreuse en de plus petites classes pouvait être efficace. Cieslak Chawla (2017) ont montré que d'autres critères plus appropriés pouvaient être utilisés pour la construction des arbres de décision. Cependant, on jugera les performances du modèle satisfaisantes. Maintenant que l'on a un modèle avec des performances satisfaisantes, on veut faire le lien entre la physique de la source et les descripteurs utilisés.

3.4 Analyse des descripteurs [2013-2018]

3.4.1 Performance des descripteurs

La sélection des descripteurs est une étape décisive pour obtenir de bons résultats de la classification. Ici, nous nous sommes intéressés à la possibilité de réduire le nombre de descripteurs, tout en gardant les plus représentatifs, et aussi de déterminer quels descripteurs sont importants et pour quelle classe. Il est également possible que certains descripteurs ne soient pas optimaux pour la classification, car fortement corrélés par exemple. Pour conserver les propriétés physiques

intrinsèques des descripteurs, nous n'avons pas transformé ou projeté de descripteurs dans un nouvel espace (*i.e.*, avec une analyse en composantes principales). Le RF permet de déterminer le poids de chaque descripteur dans une classification, en utilisant le score d'impureté. Le score d'impureté est lié à la perte de précision lorsqu'un descripteur est supprimé. Plus un descripteur est important dans la classification, plus la perte sera importante dans la performance de la classification lors de sa suppression. Cependant, en suivant cette méthode, un descripteur qui a une faible importance n'est pas automatiquement dénué de sens, car deux descripteurs peuvent être corrélés, et un seul aura un score élevé car l'information est redondante. Plus d'informations sur l'importance des descripteurs peuvent être trouvées dans Menze et al. (2009).

Dans la Figure 3.2, nous montrons le poids des 102 descripteurs (Table 2.1) pour la classification binaire (*i.e.*, chacune des classes contre toutes les autres, ligne 2 à 5) ainsi que pour la classification à 5 classes (ligne 1). Les meilleurs descripteurs pour reconnaître toutes les classes (ligne All Class de la Figure 3.2) sont le minimum du signal (29T), l'index d'énergie centrale (6T), l'énergie maximum (24T), le taux d'attaque (19T), le maximum du signal (30T) dans le domaine temporel, le kurtosis moyen (9F), la RMS (7F), l'écart-type (26F) dans le domaine fréquentiel et le taux de décroissance (20C) dans le domaine quéfrentiel. Ces neuf descripteurs ont des poids supérieurs à 2% dans la classification et proviennent de trois domaines de représentations différents (*i.e.*, domaines temporel, fréquentiel et quéfrentiel), cela donc souligne la pertinence de l'utilisation de ces trois domaines de représentation.

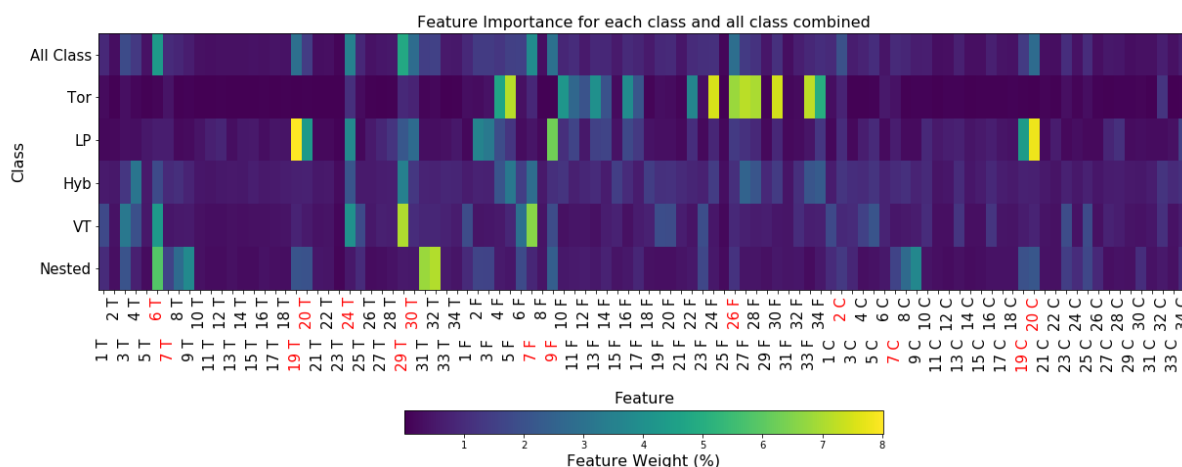


FIGURE 3.2 – Poids en pourcentage de tout les descripteurs, pour chaque classe et pour toutes les classes combinées pour la sismicité volcanique détectée à la Soufrière de Guadeloupe (1er Janvier 2013 jusqu'au 31 Décembre 2018; OVSG-IPGP) déterminé avec la fonction "feature_importance" dans la bibliothèque scikit-learn. Bleu, peu de poids dans la classification; Jaune : Poids important dans la classification; En rouge : Les 14 descripteurs les plus utiles avec lesquels le score atteint est similaire à celui avec les 102 descripteurs. En abscisse : L'index du descripteur disponible au Tableau 2.1 et son domaine de représentation (T : temporel, F : fréquentiel, C : quéfrentiel)

Parmi les 10 meilleurs descripteurs des classes les moins nombreuses, tels que les Emboîtés et les Tornillos, il y a seulement respectivement 3 et 1 de ces descripteurs que l'on retrouve parmi les 10 meilleurs descripteurs de la classification à 5 classes. Cela signifie que les descripteurs les plus discriminants d'une classe n'ont pas nécessairement un poids très important dans la classification générale et cela dépend du nombre d'événements des classes. Inversement, 7 des

10 meilleurs descripteurs de la classe VT, 6 des 10 meilleurs descripteurs de la classe LP et 5 des 10 meilleurs descripteurs de la classe Hybride sont parmi les 10 meilleurs descripteurs de la classification générale à 5 classes. En particulier, le maximum d'énergie (24T, Figure 3.2) dans le domaine temporel fait partie des 10 meilleurs descripteurs des classes VT, LP, Hybride et de la classification à 5 classes. Cela montre que ce descripteur est utile pour distinguer ces classes entre elles.

On voit que pour la classe LP, le rate of attack dans le domaine temporel (19T) a un poids très important pour distinguer cette classe des autres types d'événements. Cela concorde avec le fait qu'on ait une arrivée d'onde émergente pour cette classe. On voit également que les meilleurs descripteurs de la classe tornillo sont liés au niveau d'énergie dans le domaine fréquentiel donc souligne le phénomène de résonance produisant des Tornillos. On peut voir sur la Figure 3.3a une représentation des valeurs des meilleurs descripteurs de ces deux classes (19T, 24F). On observe que les LP, les Tornillo et le reste des signaux sont bien séparés dans l'espace des descripteurs.

La Figure 3.3b représente les meilleurs descripteurs des classes VT et Emboîtés (29T, 6T), on voit ici que les données ne sont pas beaucoup séparables dans cet espace, on a donc besoin de plus d'informations pour discriminer au mieux ces événements. Les figure 3.3c et 3.3d représentent les 4 meilleurs descripteurs pour la classification à 5 classes (19T, 24T, 9F, 20C). On peut voir que certains des meilleurs descripteurs peuvent être très corrélés entre eux (19T, 24T). On voit également sur la figure 3.3d, que ces descripteurs isole plutôt bien dans l'espace les classes LP et Tornillo des trois autres.

Pour la classe Hybride, les meilleurs descripteurs ont un poids moins important comparé aux meilleurs descripteurs des autres classes, avec un poids de 4.9% contre 8.2%, 8.7%, 8.2%, et 7.5% pour respectivement les classes VT, Emboîtés, LP et Tornillo. Aussi, le meilleur descripteur de la classe hybride, le minimum du signal (29T), est moins discriminant que les meilleurs descripteurs des autres classes dans la classification à 5 classes. C'est en accord avec le fait que la classe Hybride est la plus difficile à classer correctement pour le modèle.

3.4.2 Réduction des descripteurs pour le modèle à 5 classes

Sur la Figure 3.4, nous montrons les accuracy moyennes obtenues après 10 réalisations de classification utilisant seulement le meilleur descripteur, puis les deux meilleurs descripteurs et ainsi de suite jusqu'à utiliser chacun des 102 descripteurs. Avec les 9 meilleurs descripteurs ayant chacun un poids supérieur à 2% dans la classification, le score moyen converge rapidement au-dessus de 80%. Pour atteindre le score moyen obtenu en utilisant le jeu complet de 102 descripteurs, 82.5%, on a besoin des 14 meilleurs descripteurs. Ajouter de nouveaux descripteurs ne modifie pas sensiblement les performances et porte peu d'information additionnelle utile.

Ce jeu de descripteurs a déjà été utilisé par Malfante et al. (2018) au volcan Ubinas au Pérou. Malfante et al. (2018) ont obtenu une accuracy de 84.4% avec leurs trois meilleurs descripteurs et ont eu besoin des 13 meilleurs descripteurs pour atteindre un score (90.3%) qui est proche de celui atteint avec tous les descripteurs (92.5%). Ici, nous comparons les 13 meilleurs descripteurs obtenus par Malfante et al. (2018) avec les 14 meilleurs descripteurs que nous avons déterminé avec notre analyse. Il est intéressant de noter que parmi les 13 meilleurs descripteurs du volcan

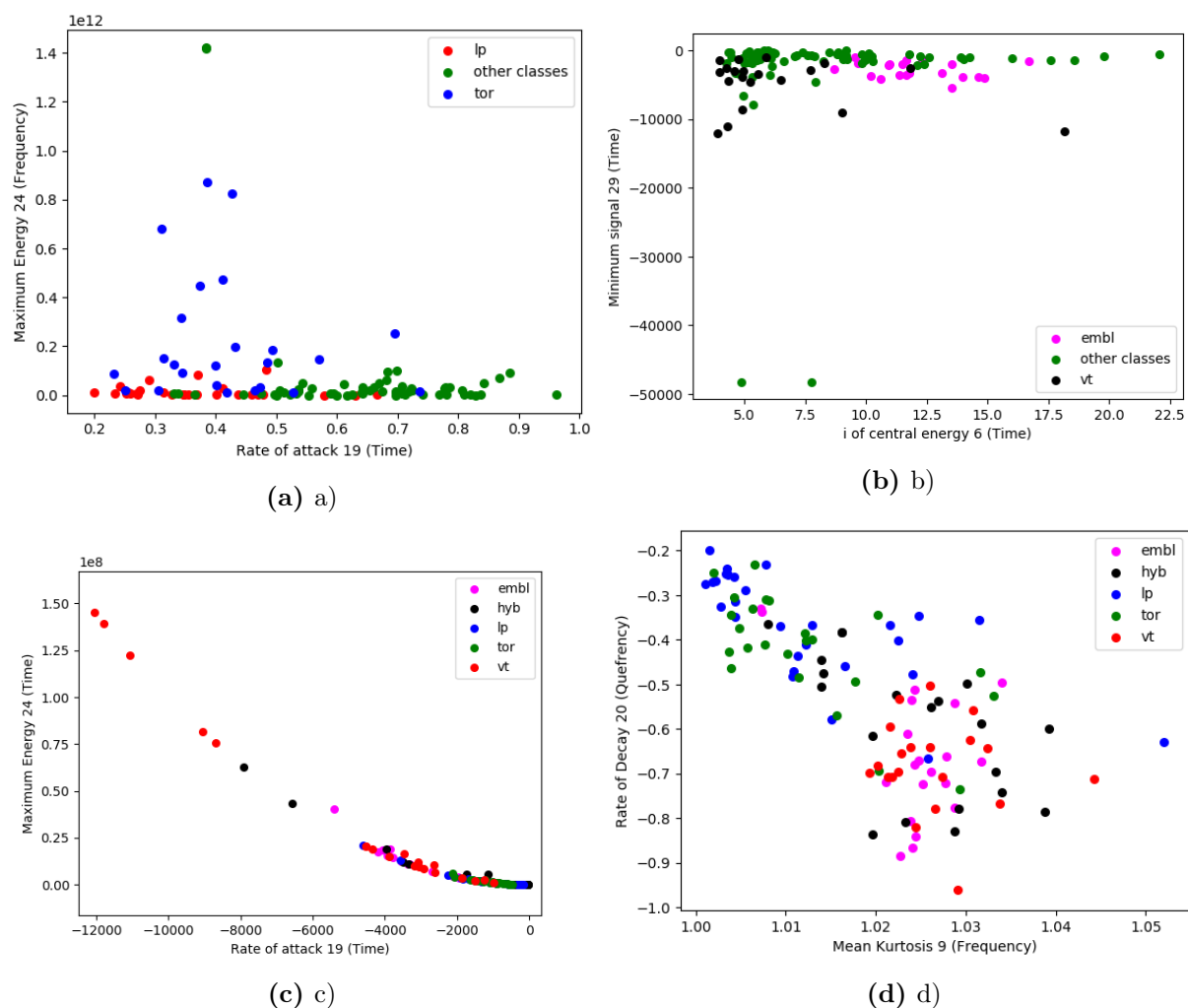


FIGURE 3.3 – a) Valeurs des meilleurs descripteurs des classes LP et Tornillos ; b) Valeurs des meilleurs descripteurs des classes Emboîtés et VT ; c) Valeurs des 2 des 4 meilleurs descripteurs de la classification à cinq classes (19T, 24T) ; d) Valeurs des 2 des 4 meilleurs descripteurs de la classification à cinq classes (9F, 20C)

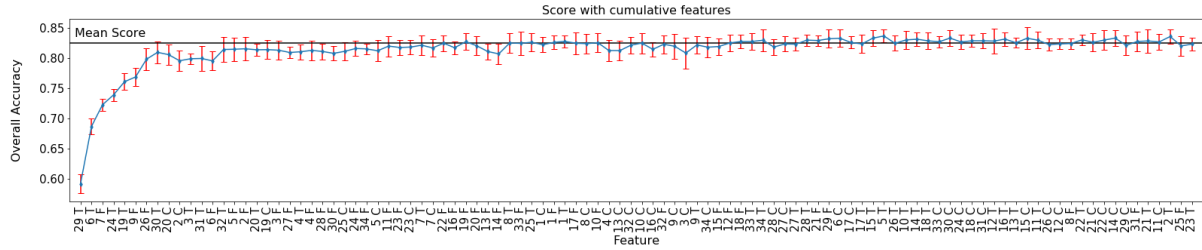


FIGURE 3.4 – Performances moyennes obtenues avec chaque n^{me} meilleurs descripteurs successifs; rapport entraînement :test = 50 :50; Modèle entraîné avec un Random Forest Classifieur (n_estimators=100, criterion='entropy', bootstrap=False, class_weight=None), performance moyenne après 10 réalisations. Ligne noire : Score moyen obtenu avec les 102 descripteurs après 10 réalisations, pour la sismicité de la Soufrière de Guadeloupe du 1er Janvier 2013 au 31 Décembre 2018 (OVSG-IPGP)

Ubinas, seulement un seul est dans nos 14 meilleurs descripteurs. C'est la moyenne du signal dans le domaine qu'éfrentiel. Cette grande différence peut être expliquée par le type de sismicité de l'Ubinas qui est différente de celle de la Soufrière de Guadeloupe. La sismicité de la Soufrière est dominée par les événements VT et Hybrides, alors que la sismicité de l'Ubinas est dominée par des séismes LP et des Trémor. Une autre différence notable est que l'Ubinas était en éruption magmatique alors que la Soufrière est un système hydrothermal avec des signaux qui ne sont pas d'origine magmatique. Cela suggère que le modèle qui nous avons construit est spécifique à la Soufrière ou au moins au type de sismicité que présente la Soufrière. Il serait intéressant de tester les performances des descripteurs sur un volcan présentant le même type de sismicité que le volcan de la Soufrière, qui est caractérisé par un système hydrothermal très actif et extensif. Une autre explication possible de la différence entre les meilleurs descripteurs pourrait être le fait que nos 14 meilleurs descripteurs obtenus à la Soufrière sont plus spécifiques à la position de la station qu'au volcan lui-même. Sur la base des structures internes complexes des volcans, la station TAG peut montrer des effets de site forts et spécifiques qui déforment les signaux (*e.g.*, forte atténuation, forte anisotropie, effets de guide d'ondes, contrastes de densité). Le très petit nombre d'événements ayant passé la phase de sélection (*i.e.*, seuls 845 des 7 194 détectés par l'OVSG répondaient au critère de qualité à la station TAG) est probablement une illustration de ce phénomène.

Pour surmonter ce problème, différentes approches peuvent être utilisées. La première, bien que peu efficace et chronophage, serait d'appliquer la même approche à station unique à d'autres stations du volcan la Soufrière et de voir si les meilleurs descripteurs sont identiques pour toutes les stations individuelles. Une deuxième approche serait d'utiliser une approche multistation, en utilisant les descripteurs calculés à partir de différentes stations pour prendre en compte la variabilité du même signal sur tout le volcan, comme dans Maggi et al. (2017). Ces auteurs ont atteint leur meilleure performance en utilisant une combinaison de descripteurs de trois stations. Une troisième approche consisterait à utiliser une approche multistation de niveau supérieur, avec des descripteurs multistation. Ces descripteurs seraient dérivés de signaux obtenus à partir d'analyses basées sur le réseau qui impliquent des données provenant de plusieurs capteurs, telles que des corrélations croisées entre des signaux de différentes stations. Nous pourrions calculer les descripteurs basés sur ces corrélations croisées dans les trois domaines de représentation (*i.e.*, domaines temporel, spectral, cepstraux).

On a donc construit un modèle de classification plutôt efficace avec une accuracy de $82.1\% \pm 1.6$. On peut faire en sorte de réduire le nombre de paramètres du modèle pour résoudre des problèmes de coût de calcul ou de sur-apprentissage. Mais maintenant on veut savoir si notre modèle reste pertinent lorsqu'il est appliqué à des signaux d'une nouvelle période temporelle.

3.5 Application du modèle à une nouvelle période [2018-2020]

Nous avons construit un modèle à partir d'événements de la période allant du 1er Janvier 2013 au 31 Décembre 2018. Pour construire le modèle, nous avons utilisé des événements respectant un certain niveau de rapport signal sur bruit. Nous allons appliquer ce modèle sur une nouvelle période pour déterminer s'il est toujours fiable ou non et nous allons tenter de détecter des changements dans le comportement, la physique du volcan.

3.5.1 Nouveau jeu de données

Sur la Figure 3.5, nous voyons la distribution des événements sismo-volcanique sur la période du 1er Janvier 2013 au 31 Décembre 2020 (en bleu), et également la distribution temporelle des événements utilisés pour l'entraînement de notre modèle (en orange). On observe ici une augmentation de la sismicité à partir de fin 2017, début 2018, il y a même eu au cours du mois d'avril 2018 le séisme le plus fort depuis la crise de 1976. Nous voulons appliquer notre modèle sur la période du 1er Janvier 2018 au 31 Décembre 2020 car il y a une augmentation du nombre de signaux par rapport à la période précédente qui pourrait être causé par une migration de fluide. Nous conservons un recouvrement entre la période d'entraînement et la période d'application pour vérifier s'il y a une modification des résultats lorsqu'on l'applique à de nouvelles données. Il y a eu 17 929 événements sur la période 2018-2020 dont 13 193 sont disponibles à la station TAG. Parmi ceux-ci, il y a selon l'OVSG : 8 295 VT, 46 LP et 4 852 Emboîtés.

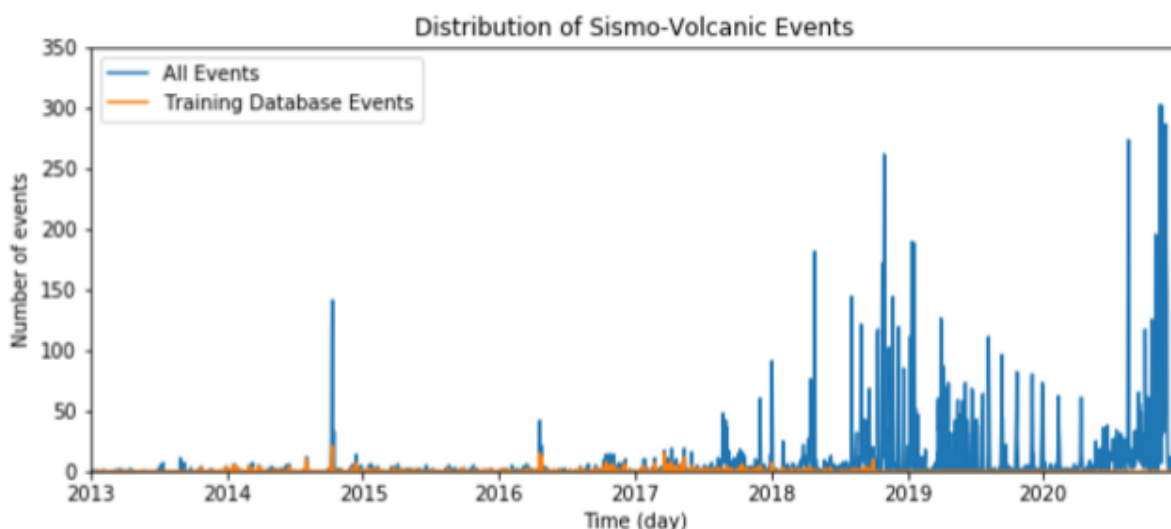


FIGURE 3.5 – Distribution des événements sismo-volcaniques sur la période 2013-2020 ; En Bleu : Tout les événements détectés par l'OVSG-IPGP ; En Orange : Les événements utilisés pour l'entraînement du modèle

Sur la figure 3.6, nous voyons les rapports signal sur bruit des événements selon les périodes. Pour notre jeu d'entraînement, nous avons sélectionné des événements ayant un rapport signal sur bruit supérieur au seuil de 1.5. Nous appliquerons notre modèle sur les événements de la période 2018-2020 respectant également ce critère. On cherche à appliquer notre modèle sur les événements respectant ce critère de qualité pour faciliter les interprétations et éviter les événements bruités pour lesquels la prédiction faite par le modèle serait moins fiable.

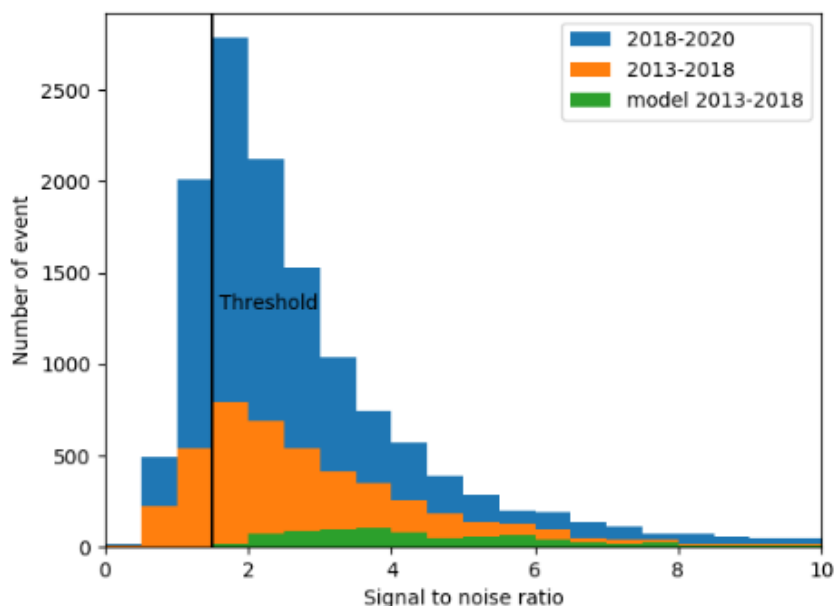


FIGURE 3.6 – Distribution des rapports signal sur bruits. Bleu : Événements de la période 2018-2020 ; Orange : Événements de la période 2013-2018 ; Vert : Événements de la base d'entraînement ; Ligne noire : Seuil du critère qualité, au-delà les événements sont utilisés dans l'analyse

Sur le Tableau 3.5, nous voyons la répartition des événements dans les trois classes selon le rapport signal sur bruit. Nous avons 13 193 événements disponibles à la station TAG, parmi ceux-là, 10 680 passent le critère qualité. En terme de proportion de chacune des classes, la classe VT représente un peu plus de 60% des événements quel que soit l'échantillon auquel on s'intéresse. De la même façon, les séismes Emboîtés représentent toujours un peu moins de 40% des données et enfin les LP représentent environ 0.4% des événements. Les proportions entre les différentes classes sont stables quels que soient les rapports signal sur bruit. On suppose donc que les interprétations faites avec les événements respectant le critère de qualité peuvent être étendues à l'ensemble des données.

TABLE 3.5 – Nombre d'événements dans les classes selon les rapports signal sur bruits atteints

	Tous les événements		Événements respectant le critère qualité	
	Nombre d'événement	%	Nombre d'événement	%
2-5				
VT	8295	63.0	6604	61.8
LP	46	0.3	42	0.4
Emboîté	4852	36.7	4034	37.8
Total	13193		10880	

3.5.2 Premier résultat après application du modèle sur la période 2018-2020 et identification d'une erreur

On applique donc le modèle à 5 cinq classes entraînées avec des événements de la période 2013-2018 sur les événements de la période 2018-2020 ayant passé le contrôle qualité. Le premier résultat qui saute aux yeux sur le Tableau 3.6 est le nombre d'événements LP prédits par le modèle. 851 LP ont été prédit par le modèle sur la période alors que selon l'OVSG il n'y a seulement que 42 événements LP dans ce jeu de données. On pourrait s'attendre à trouver plus de LP que l'OVSG mais 20 fois plus semble étonnant comme résultat. On a donc deux hypothèses, la première étant que beaucoup d'événements LP ont été mal classés par l'OVSG, la seconde est qu'il y a des erreurs dans notre jeu d'entraînement notamment dans nos événements LP. On va donc vérifier les LP utilisés dans notre base d'apprentissage et voir si des erreurs de labellisation ont été commises.

TABLE 3.6 – Matrice comparant les classes données par l'OVSG et celles prédites par le modèle construit sur 2013-2018 et appliqué sur la période 2018-2020. On s'intéresse également aux résultats par physique similaire, les VT sont groupés avec les Hybrides et les emboîtés, les LP sont groupés avec les Hybrides et les Tornillos, les Hybrides sont groupés avec les LP, les VT et les Emboîtés

2-6 Classe OVSG	Classe prédite					Total OVSG	$\frac{\%Physiquesimilaire}{TotalClasseOVSG}$
	VT	Emboîté	LP	Hyb	Tor		
VT	4809	408	520	865	2	6604	92.10
Emboîté	2981	276	325	450	2	4034	91.89
LP	29	2	6	3	2	42	26.19
Total Prédite	7819	686	851	1318	6	10680	
$\frac{\%Physiquesimilaire}{TotalClassePrdite}$	99.63	99.71	0.71	100	33.33		

3.5.3 Correction du modèle

La première phase de visualisation correspondant à la section 3.3 du manuscrit a été réalisée lors de la première année de thèse. Cette nouvelle visualisation a été réalisée durant la troisième année de la thèse. Il est probable que l'oeil de l'opérateur (c'est-à-dire mon oeil) se soit aguerri et que l'on puisse corriger des erreurs qui pourraient s'être glissées dans la première phase de visualisation.

Après visualisation des 97 événements LP de la base d'entraînement (donc de la période 2013-2018), on a repéré des événements n'étant pas des LP. La Figure 3.7 représente un exemple d'événement classé comme LP dans un premier temps, mais qui après réflexion correspond à une autre classe. Ici, l'opérateur s'est trop fié au spectrogramme et au spectre de Fourier et avait conclu à un événement LP à cause du contenu fréquentiel plutôt basse fréquence. Mais la forme d'onde plutôt impulsive et haute fréquence semble suggérer un événement hybride. De plus la partie basse fréquence de ce signal se situe à environ 6 Hz alors que la fréquence caractéristique

des LP de la Soufrière est aux alentours de 4 Hz.

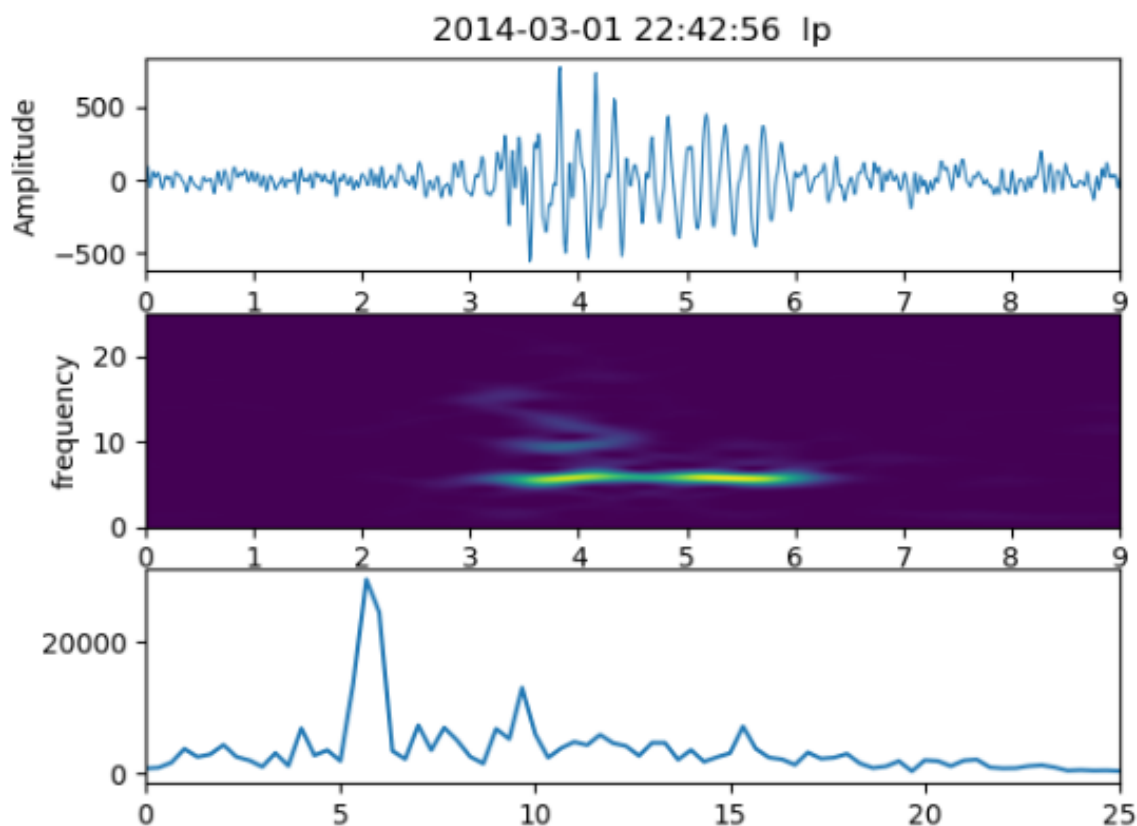


FIGURE 3.7 – Forme d’onde, spectrogramme et spectre de Fourier d’un événement classé faussement comme LP

Le Tableau 3.7, montre les corrections que l’on a apportées à notre base d’entraînement. Sur les 97 LP de départ, 34 sont des VT, 7 des Emboîtés, 22 des Hybrides et il reste donc 34 LP. Le modèle est donc actualisé.

TABLE 3.7 – Assignation de nouvelles classes après visualisation des événements VT de la base d’entraînement

Ancienne classe	Nouvelle classe			
	VT	Emboîté	Hybride	LP
97 LP	34	7	22	34

3.5.4 Application du modèle corrigé sur 2018-2020

Le Tableau 3.8 présente comment les événements de la période 2018-2020 sont classés par le modèle tout en comparant avec les classes assignées par l’OVSG.

Dans un premier temps, on voit que le modèle trouve 9 événements Tornillo. Ce chiffre nous permet de pouvoir directement vérifier à l’œil si les prédictions du modèle sont bonnes. Et, en effet, ces 9 événements sont bien des Tornillo. 5 d’entre eux étaient classés comme LP par l’OVSG ce qui est logique car l’OVSG considère les Tornillo comme LP mais 2 étaient classés comme VT

et 2 comme Emboîtés. Il y a donc eu des erreurs et le modèle a été capable de les retrouver. Tous les résultats jusqu'ici semblent montrer que les Tornillo prédits par le modèle sont très fiables.

TABLE 3.8 – Matrice comparant les classes données par l'OVSG et celle prédite par le modèle corrigé sur la période 2018-2020. On s'intéresse également aux résultats par physique similaire, les VT sont groupés avec les Hybrides et les emboîtés, les LP sont groupés avec les Hybrides et les Tornillos, les Hybrides sont groupés avec les LP et les Emboîtés

2-6 Classe OVSG	Classe prédite					Total OVSG	$\frac{\%Physiquesimilaire}{TotalClasseOVSG}$
	VT	Emboîté	LP	Hyb	Tor		
VT	5053	460	66	1023	2	6604	98.97
Emboîté	3159	307	50	516	2	4034	98.71
LP	10	2	23	2	5	42	71.43
Total Prédite	8222	769	139	1541	9	10680	
$\frac{\%Physiquesimilaire}{TotalClassePrdite}$	99.88	99.74	16.55	100	55.56		

Ensuite, notre modèle prédit 139 événements LP contre 42 pour l'observatoire. Encore une fois, ce chiffre nous permet de vérifier à l'oeil si les 139 LP prédits sont corrects, ce qui est le cas. Sur ces 139 LP, seulement 23 étaient considérés comme LP par l'OVSG alors que 66 étaient classés comme VT et 50 comme emboîtés. Sur les 42 LP considérés par l'OVSG seulement 71% ont été classés comme LP ou Tornillo, les 29% restant ont été classés comme VT, Emboîté et Hybride et à raison par le modèle, il y a également des erreurs dans les LP assignées par l'OVSG. L'application du modèle sur la période nous a permis de détecter des LP qui ont été mal classés et de tripler le nombre d'événements LP de la période, ce qui n'est pas négligeable étant donné que ce type d'événement est assez rare.

4 034 Emboîtés ont été détectés par l'OVSG contre 769 par notre modèle. La différence entre les deux résultats semble grande, mais on rappelle que lors de l'étape de visualisation décrite en partie 3.3, nous avons redéfini la classe Emboîtés et que le nombre d'événements de cette classe a été réduit. On peut comparer comment les événements considérés Emboîté par l'OVSG ont été relabellisés par notre phase de visualisation en 3.3 et par le modèle sur la période 2018-2020.

Le Tableau 3.9 montre comment lors de la première phase de visualisation (sur la période 2013-2018), les événements classés comme Emboîtés par l'OVSG ont été réorganisés dans d'autres classes et la même chose avec les événements classés comme Emboîté par l'observatoire durant la période 2018-2020. On voit que les événements classés comme Emboîté par l'OVSG sont reclassés de la même façon par le modèle et notre visualisation avec plus de la moitié qui sont reclassés comme VT, entre 10% et 20% sont reclassés comme Hybride. Environ 10% sont conservés comme emboîtés. Les différences entre les proportions peuvent être dues à la différence de taille des deux échantillons, mais en terme d'ordre de grandeur les deux réorganisations sont équivalentes. Une visualisation de certains des événements prédits comme Emboîté par le modèle sont confirmés comme Emboîté après visualisation, même ceux venant de la classe VT ou les deux événements qui ont été auparavant classés comme LP.

Enfin, pour la classe VT, 6 604 VT étaient détectés par l'observatoire contre 8 222 par le modèle

dont 5 053 événements classés de la même façon. 99% des VT classés par l'Observatoire ont été reclassés dans des classes de physique similaire à savoir VT, Emboîté ou Hybride.

En conclusion, l'application du modèle à une nouvelle période nous a permis dans un premier temps et à cause d'un résultat aberrant de corriger des erreurs dans notre base d'entraînement du modèle. Puis, on a pu repérer des événements LP (et Tornillo) qui n'ont pas été correctement classés lors du processus de classification de l'observatoire. Enfin, les classes VT et Emboîté sont réorganisés en VT, Emboîtés ou Hybride donc des classes de physiques similaires. Donc le modèle de classification d'événements sismo-volcaniques construit sur la période 2013-2018 reste pertinent sur la période 2018-2020. Malgré cela, maintenant, on veut voir si on est capable de détecter des changements aux sein des signaux d'une même classe.

TABLE 3.9 – Répartition des Emboîté détecté par l'OVSG dans d'autre classe par le modèle

	Classe modèle					
	2-6	VT	LP	Emboîté	Tor	Hyb
OVSG Emboîté		125	22	28	2	40
%		57.6	10.1	12.9	0.9	18.4
OVSG Emboîté 2018-2020		3159	50	307	2	516
%		78.3	1.2	7.6	0.0	12.8

3.5.5 Recherche d'un changement de VT

Arnaud Burtin (communication personnelle) a montré qu'il y avait plusieurs types de VT à la Soufrière de Guadeloupe. Grâce à une méthode de template matching il a pu détecter 4 familles de VT avec un taux de détection nettement supérieur à celui utilisé de manière journalière à l'OVSG.

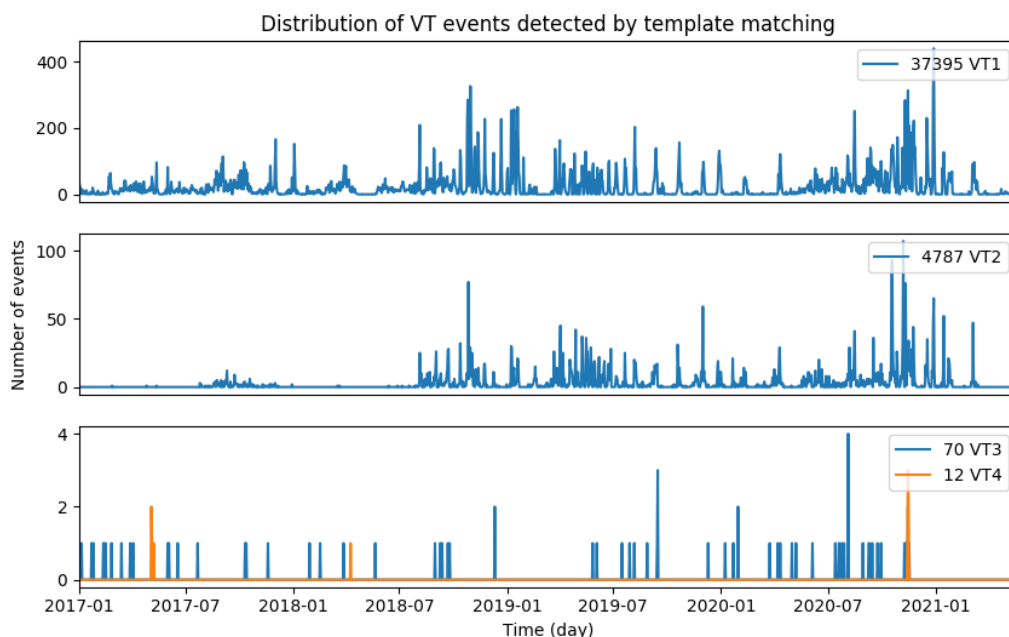


FIGURE 3.8 – Distribution temporelle journalière de quatre familles de VT détectées par template matching (A. Burtin, communication personnelle)

La Figure 3.8 montre la distribution temporelle journalière de chacune des familles de VT. On voit que sur la période de Janvier 2017 à Mai 2021, on peut compter 37 395 VT1, 4 787 VT2, 70 VT3 et 12 VT4. Ici, on s'intéressera aux deux premières familles. On voit que les VT1 sont les plus nombreux et sont présent tout le temps. La famille de VT2 présente quelques événements lors de la deuxième moitié de 2017 mais connaît une forte augmentation à la fin de l'année 2018. On cherche à voir si avec notre modèle construit sur 2013-2018 donc potentiellement avec une forte proportion de VT1 s'il est possible de détecter un changement dans la physique de VT correspondant à cette augmentation de VT2.

Malfante et al. (2018) ont pu détecter un changement dans la physique des LP associé à un reflux de la colonne de magma dans le conduit de l'Ubinas grâce à une baisse du taux de reconnaissance des LP sur la période. En suivant la même approche, on cherche à voir si ce changement dans la classe VT à la Soufrière peut se traduire par une modification des performances de la reconnaissance des VT entre 2018 et 2019.

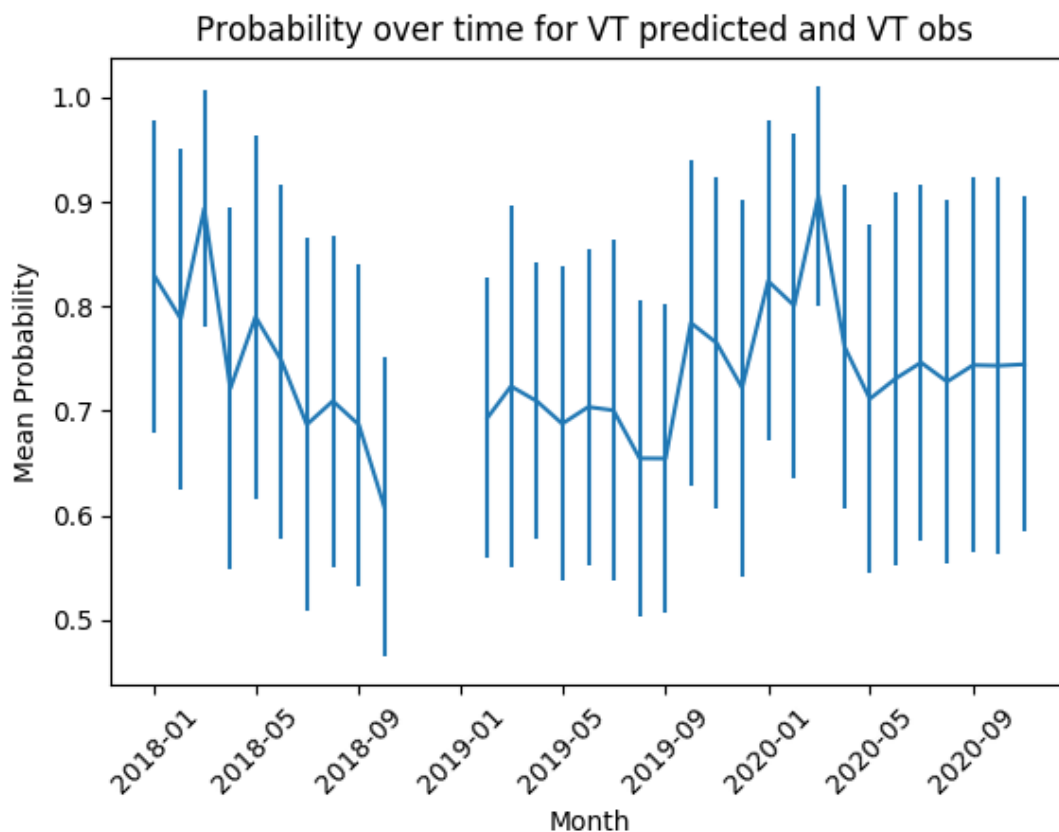


FIGURE 3.9 – Evolution mensuelle de la probabilité moyenne que les VT considérés par l'OVSG et le modèle soient des VT et écart-type; Le trou dans la figure correspond à une absence de données pour la station TAG durant cette période

La Figure 3.9 représente les événements considérés comme VT par l'OVSG et par le modèle, c'est à dire les 5 053 du Tableau 3.8. On utilise cette fois le modèle, non pas pour avoir une prédiction, mais pour déterminer la probabilité c'est à dire le niveau de confiance sur le fait que chacun de ces 5 053 VT soient bel et bien des VT. Comme on a une accélération du nombre de VT2 à partir de Juillet 2018 et que notre modèle est entraîné majoritairement avec des VT1, on peut imaginer

que ce changement de signaux puisse se traduire par une baisse de la probabilité d'appartenance à la classe VT. Pour cela, on regarde de façon mensuelle cette probabilité et l'écart-type de ces probabilités sur le mois. On voit que les variations de cette probabilité sont faibles comparées aux écart-types et qu'il n'y a pas de corrélation entre une potentielle baisse de cette probabilité et l'augmentation de VT2 à partir de fin 2018.

En analysant les évolutions de performances du modèle au cours du temps, nous n'avons pas pu détecter de changements dans la physique du volcan. Faire une classification VT1 contre VT2 pourrait nous permettre de voir quels sont les descripteurs les plus discriminants et donc portant des informations sur la physique de la source. Mais ce n'était pas l'objectif ici, on voulait voir si à partir d'un changement dans les performances, on pouvait détecter un changement dans la physique et pas l'inverse.

3.6 Résumé

- A partir d'un catalogue fourni par l'OVSG, nous avons construit un modèle de classification automatique d'événements sismo-volcaniques.
- Dans le but d'améliorer les performances du modèle, nous avons défini de nouvelles classes et nous avons relabellisé le jeu de données. Nous avons ainsi affiné le catalogue fourni par l'OVSG
- Nous avons construit un modèle de classification automatique fiable à 84%. Le modèle est capable de reconnaître avec une fiabilité d'environ 80% ou plus les VT, LP, Tornillo et Emboîté, mais est moins fiable pour la reconnaissance des événements hybrides (précision de 70%) car souvent confondus avec les LP ou les VT.
- Une analyse des descripteurs nous a permis de voir qu'un subset de ceux-ci était suffisant pour obtenir de bons résultats mais ils sont spécifiques à notre jeu de données.
- Le modèle construit a été appliqué à une autre période de temps, le nombre trop grand de LP reconnus par le modèle sur cette nouvelle période nous a incité à relabelliser les LP de notre base d'entraînement pour corriger le modèle. Après cela, nous avons vu que le modèle restait fiable quelle que soit la période d'application et permettait de corriger des erreurs de classification.
- Nous n'avons pas pu mettre en évidence de changements au sein de la physique du volcan et notamment repérer les différentes familles de VT.

Nous avons montré que l'apprentissage supervisé peut se révéler puissant pour la classification automatique d'événement sismo-volcaniques. On a pu corriger des erreurs commises dans le processus de classification. Cette analyse nécessite d'avoir des événements déjà détectés pour pouvoir fonctionner et donc leur assigner une classe. Dans la partie suivante, nous allons nous intéresser à l'apprentissage supervisé pour traiter le signal continu et faire le travail de détection et de classification sans intervention manuelle.

4 Détection et classification d'événements sismo-volcaniques

4.1 Introduction

Actuellement, le travail de détection à l'OVSIG est fait automatiquement grâce à un algorithme de type STA/LTA (Allen, 1978). C'est un type d'algorithme de détection très communément utilisé en sismologie. Il consiste à calculer de manière continue la moyenne de la valeur absolue d'un signal sismique dans deux fenêtres temporelles simultanées, la petite fenêtre temporelle (short time window, STA) correspondant à une portion de signal à la fin de la longue fenêtre temporelle (long time window, LTA). La fenêtre STA est sensible aux événements sismiques pendant que la fenêtre LTA donne des informations sur le niveau de bruit enregistré sur le site. Quand le ratio des deux moyennes dépasse un seuil déterminé par l'opérateur, un événement est détecté.

Nous disposons d'un modèle de classification automatique capable de reconnaître 5 classes d'événements sismo-volcanique, on veut savoir si, en appliquant le modèle sur un signal continu, il est capable de détecter et classer des événements.

4.2 Application du modèle sur le signal continu

4.2.1 Méthode

Notre première approche est donc d'appliquer notre modèle de façon brute sur le signal continu. Pour cela, on applique une fenêtre glissante sur le signal continu et, pour chaque fenêtre, nous calculons les descripteurs correspondant au signal, ainsi le modèle peut donner une prédiction. Mais plutôt que de demander au modèle de prédire une classe, nous cherchons la probabilité d'appartenance à chacune des classes. Comme le RF est un algorithme se basant sur une collection d'arbres de décision, la prédiction finale se base sur un système de vote de chacun des arbres et donc la classe prédite par la majorité des arbres. Ici, nous nous intéressons à la prédiction faite par chacun des arbres et nous occultons donc cette phase de vote. Nous attribuerons une classe à une fenêtre glissante si la probabilité d'appartenance à la classe en question dépasse un certain seuil, autrement dit s'il y a suffisamment d'arbres de décision qui donnent la même prédiction.

Maintenant, il nous faut définir la taille de notre fenêtre glissante. La Figure 4.1 montre la distribution de la longueur des événements sur la période 2013-2018. On voit qu'une majorité des événements ont une durée très courte, la moyenne de la durée des événements est de 4.49 secondes tandis que la médiane est de 3 secondes. 80% des événements ont une durée inférieure à 6 secondes et 95% des événements ont une durée inférieure à 11 secondes. La sismicité est majoritairement composée d'événements VT ou Hybride, c'est-à-dire d'événements de courte durée, généralement inférieure à 5 secondes. Les événements qui ont une durée plus longue sont des séismes Emboîtés ou bien des Tornillos, deux classes que notre modèle reconnaît assez facilement. Pour nous, ce qui est intéressant, c'est d'être capable de détecter ces événements courts donc pour cela, on choisira une fenêtre glissante de 5 secondes. On décalera notre fenêtre de 1 seconde à chaque pas

de glissement. Pour commencer nous nous intéresserons à des cas plutôt simples avec des enregistrements assez courts sur lesquels l'OVSG a détecté un ou plusieurs événements appartenants à chacune des classes de notre catalogue.

Il ne faut pas oublier que notre modèle de classification automatique est capable de distinguer des événements de 5 classes différentes, or, sur le signal continu il y a au moins une autre classe qui est le bruit sismique. C'est pour cela que nous ne voulons pas que notre modèle nous donne des prédictions de classe, mais la probabilité d'appartenance à chacune des classes. Ainsi, on peut fixer un seuil de probabilité au-delà duquel le signal contenu dans la fenêtre correspond à une de nos 5 classes et en dessous duquel le signal correspond à du bruit sismique. Pour le moment, nous ne fixerons pas de seuil, mais nous nous intéresserons aux comportements de ces probabilités d'appartenance.

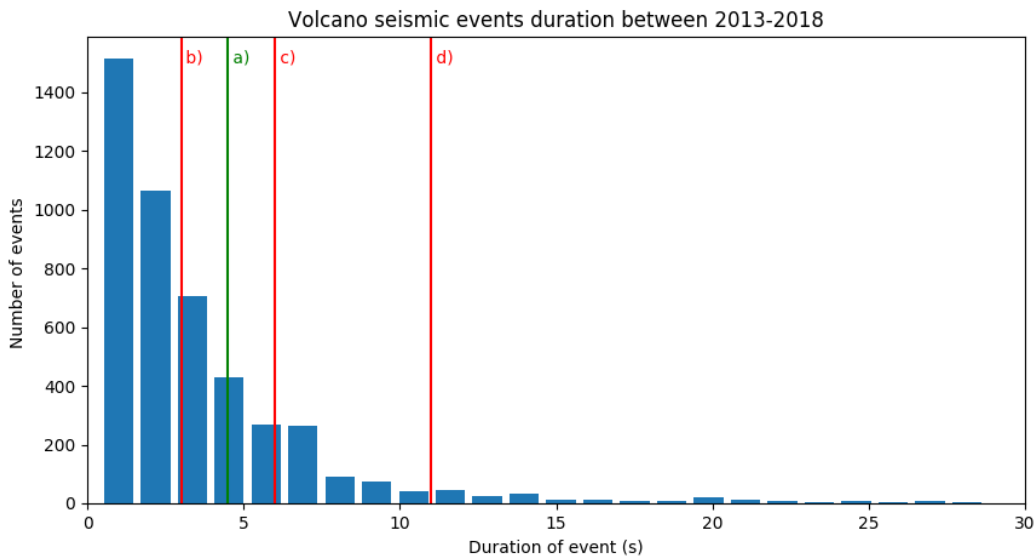


FIGURE 4.1 – Histogramme représentant la durée des événements sismo-volcaniques détectés sur la période 2013-2018 ; a : Moyenne = 4.49s ; b : médiane = 3.0s ; c : 80ème percentile = 6.0s ; d : 95ème percentile = 11.0s

4.2.2 Résultats

VT

Sur la Figure 4.2, nous observons 2 minutes de signal sur lesquelles nous avons appliqué notre modèle de classification à 5 classes. Entre 14h03 et 14h05, le 9 septembre 2018 deux séismes de type VT ont été détectés par l'OVSG, tous deux ayant une durée d'environ 3 secondes. On peut voir ici plusieurs choses, la première est qu'on a réussi à détecter les deux séismes VT. Sur la figure, on voit qu'au moment de chaque événement il y a plusieurs fenêtres pour lesquelles la probabilité d'avoir un VT est très forte, supérieure à 0.9. En-dehors de ces deux événements, pendant que le bruit sismique est dominant, les probabilités de classes semblent se partager entre VT, LP et Hybride. La classe LP étant quasiment toujours la plus probable. Alors que les probabilités pour qu'on ait un Tornillo ou un Emboité est toujours zéro ou proche de zéro. Dans

le cas de ces deux minutes, on pourrait fixer un seuil de détection à 0.9 pour obtenir un modèle de détection et de classification efficace, mais regardons d'autres signaux pour voir ce qui peut se passer.

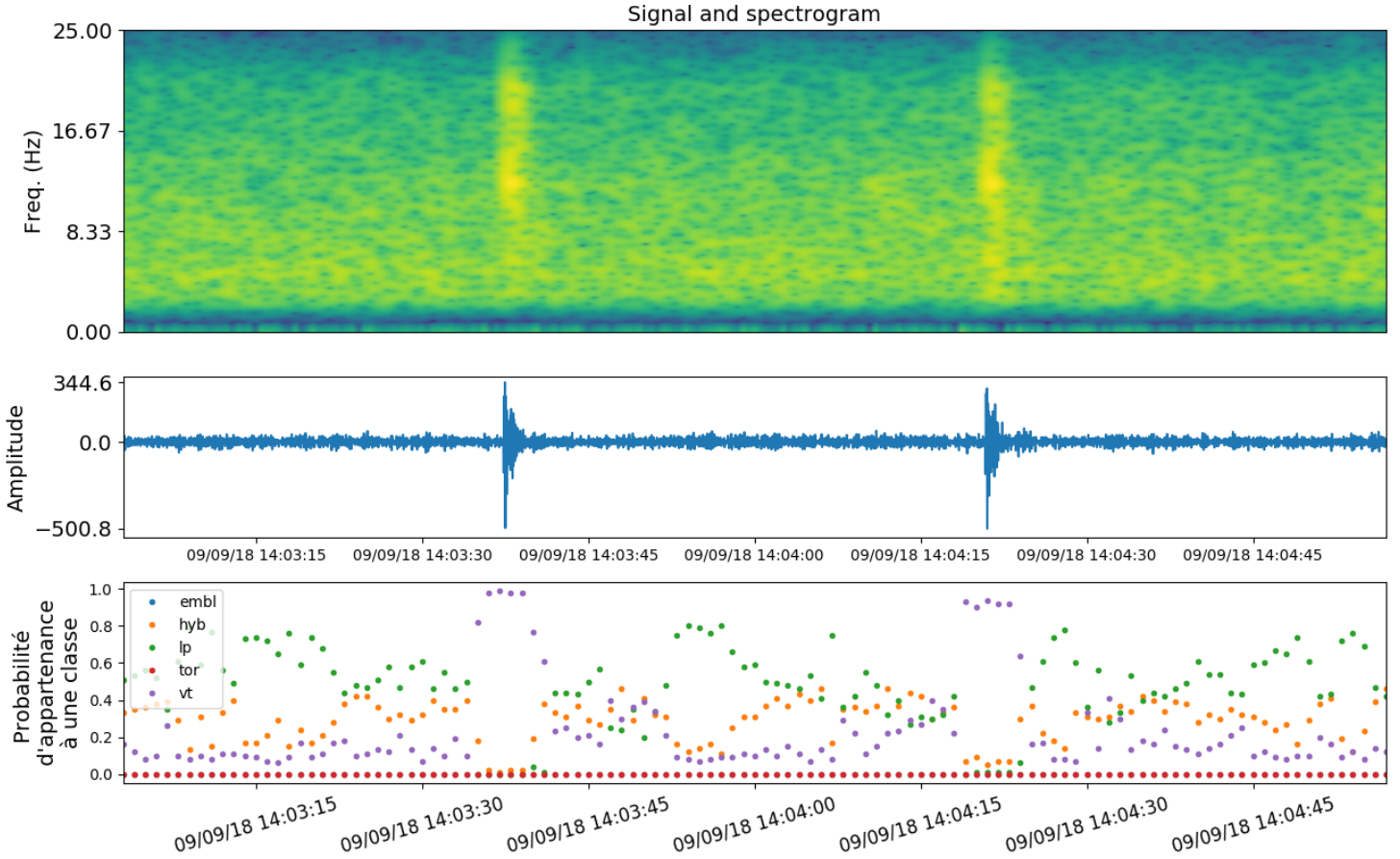


FIGURE 4.2 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 2 VT ont été détectés par l'OVSG ; Emboité : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

LP

Sur la Figure 4.3, nous observons 2 minutes de signal sur lesquelles nous avons appliqué notre modèle de classification à 5 classes. Durant cette période, un séisme LP a été observé. Sur cette figure, comme sur la Figure 4.2, lorsque le signal est composé de bruit sismique, la classe avec la probabilité la plus forte est la classe LP. Mais ici, la probabilité oscille autour de 0.8 quand elle était entre 0.4 et 0.8 sur la figure précédente. Il est important de noter qu'on observe sur le spectrogramme de l'énergie quasiment constante autour de 4Hz, ce bruit constant pourrait être produit par le système hydrothermal de la Soufrière. On peut aussi voir sur la figure qu'avant l'événement la probabilité d'un LP oscille autour de 0.8 puis décroît jusqu'à 0.6 environ pendant 2 fenêtres consécutives, puis remonte à 0.8 pendant 1 fenêtre, décroît de nouveau aux alentours de 0.6 pendant 3 fenêtres et enfin remonte autour du niveau de bruit à 0.8. Si l'événement LP n'est pas centré dans la fenêtre sur laquelle on calcule les descripteurs alors le modèle reconnaît moins le LP. On pourrait expliquer ça par le fait que le taux d'attaque est le meilleur descripteur

pour reconnaître des LP et que selon la position de l'événement dans la fenêtre glissante alors le taux d'attaque est modifié.

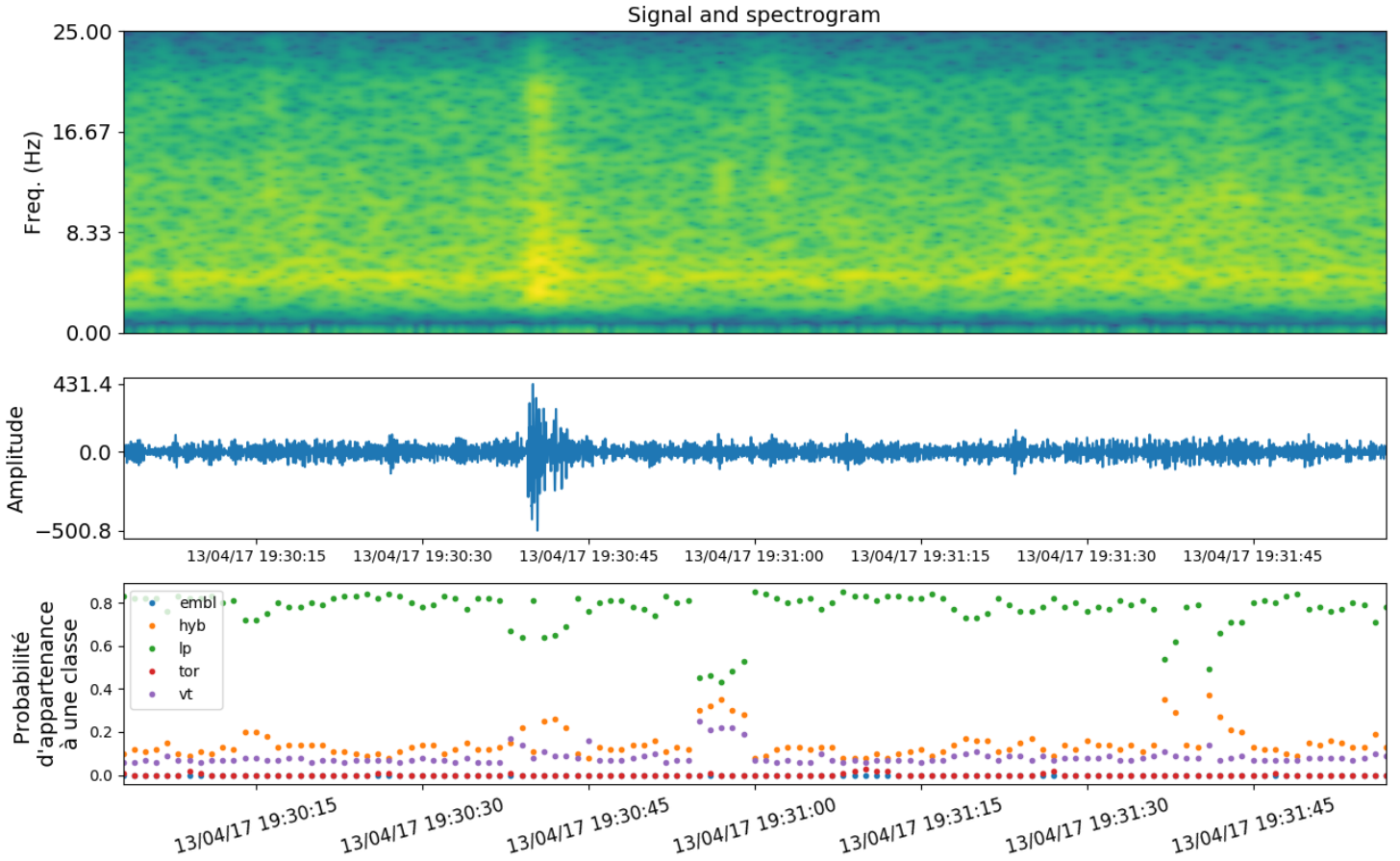


FIGURE 4.3 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 1 LP a été détecté par l'OVSF ; Emboité : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

Sur la Figure 4.4, nous avons une nouvelle fois 2 minutes de signal pendant lesquelles un événement LP s'est produit. Comme sur la figure 4.2, durant le bruit sismique les probabilités sont partagées entre VT, LP et Hybride. Mais dans ce cas-ci, la probabilité dominante n'est pas celle associée aux LP, mais celle des VT oscillant aux alentours de 0.6, cela peut s'expliquer par le fait que le bruit ambiant est moins énergétique dans les basses fréquences (aux alentours de 4Hz) que dans les deux cas précédents. On peut noter qu'au moment du LP, la probabilité associée à cette classe ne dépasse pas beaucoup le niveau de bruit de la probabilité VT avec deux fenêtres ayant une probabilité entre 0.7 et 0.8 et 6 fenêtres aux alentours de 0.6. Néanmoins la probabilité qu'on ait un LP a beaucoup augmenté quand on la compare à sa valeur quand il n'y a pas d'événement (entre 0.1 et 0.4).

Hybride

Sur la Figure 4.5, nous observons 2 minutes de signal, dans lequel se trouvent 2 événements. Le premier événement à 15 :43 :50 a une faible amplitude. Il n'a pas été détecté par l'observatoire

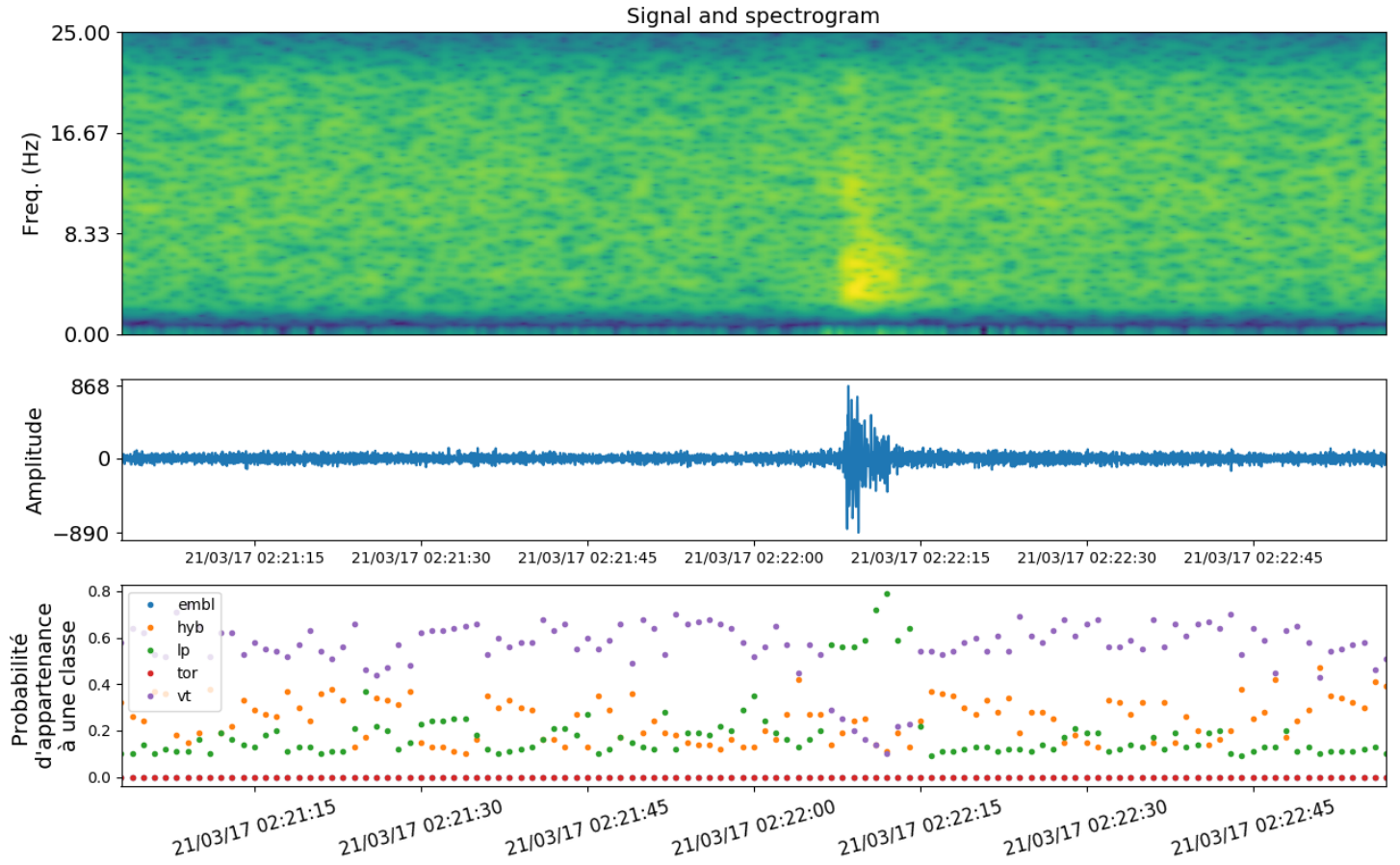


FIGURE 4.4 – Spectrogramme, forme d’onde et probabilités d’appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 1 LP a été détecté par l’OVSG ; Emboîté : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

et semble être un VT le second à 15 :44 :22 a été détecté par l’OVSG et est considéré comme Hybride par notre modèle. Comme sur les Figures 4.2 et 4.3 le bruit sismique est classé comme LP. On considère l’événement de 15 :43 :50 comme détecté par le modèle car la probabilité d’être un VT monte au-dessus de 0.6 pendant plusieurs fenêtres consécutives alors que pendant le bruit sismique, la probabilité VT oscille entre 0.1 et 0.4. Pour le second événement, on voit que l’on déclenche d’abord un VT avec une probabilité très forte (>0.9) puis sur les fenêtres suivantes, la probabilité maximale est celle d’être un Hybride (>0.5). On voit également que sur la fin du signal la probabilité d’être un LP augmente d’environ 0 à 0.2. Cette évolution dans la valeur des probabilités et dans l’évolution des classes les plus probables pourrait s’expliquer par la combinaison de l’Hybride avec la fenêtre glissante, car l’Hybride possède des caractéristiques des LP et des VT. En effet, les premières fenêtres dans lesquelles le signal Hybride apparaît possèdent les caractéristiques d’un VT avec une arrivée impulsive à haute fréquence. Ensuite, quand une plus grande partie du signal est située dans la fenêtre glissante, la composante basse fréquence du signal, caractéristique des Hybrides, commence à apparaître ce qui est cohérent avec une augmentation de la probabilité d’être un Hybride. Enfin quand la fenêtre glisse sur la coda de l’événement où la composante basse fréquence du signal est plus marquée, la probabilité

d'être un LP augmente. On voit donc que selon la position de la fenêtre glissante sur le signal la classe prédite par le modèle change. De plus, le fait que le modèle hésite entre plusieurs classes entraîne une baisse de la valeur des probabilités et ne facilite pas le réglage d'un seuil à partir duquel on peut considérer qu'il y ait ou pas un événement.

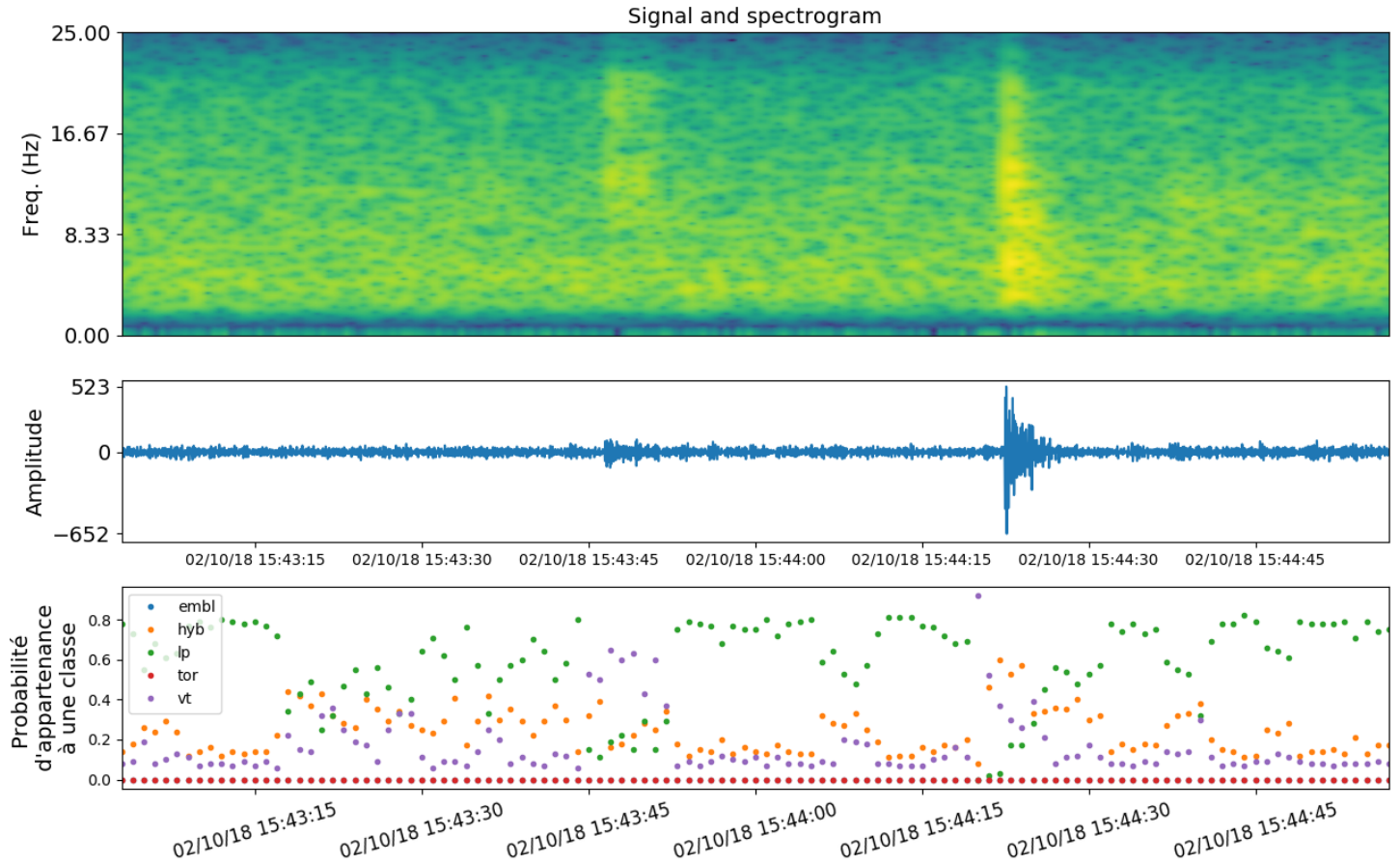


FIGURE 4.5 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Hybride a été détecté par l'OVSG ; Emboîté : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

Emboîté

La Figure 4.6 représente 2 minutes de signal pendant lesquelles un séisme Emboîté a été détecté. Encore une fois, la probabilité maximale caractérisant le bruit correspond à la classe LP. Mais intéressons nous aux probabilités associées à l'événement. Le modèle donne pour type d'événement le plus probable la classe VT. Comme le modèle a été entraîné avec des signaux de taille variable et que les Emboîtés sont majoritairement des événements plus longs que les VT (jusqu'à plusieurs dizaines de secondes), alors l'utilisation d'une fenêtre courte par rapport à la durée des événements Emboîtés pour la détection peut entraîner une perte de précision dans la reconnaissance. De plus, on considère les Emboîtés comme des enchevêtrements de VT. Par conséquent, si on partitionne un Emboîté en plusieurs signaux de petite taille alors il peut être logique que le modèle le classe comme VT.

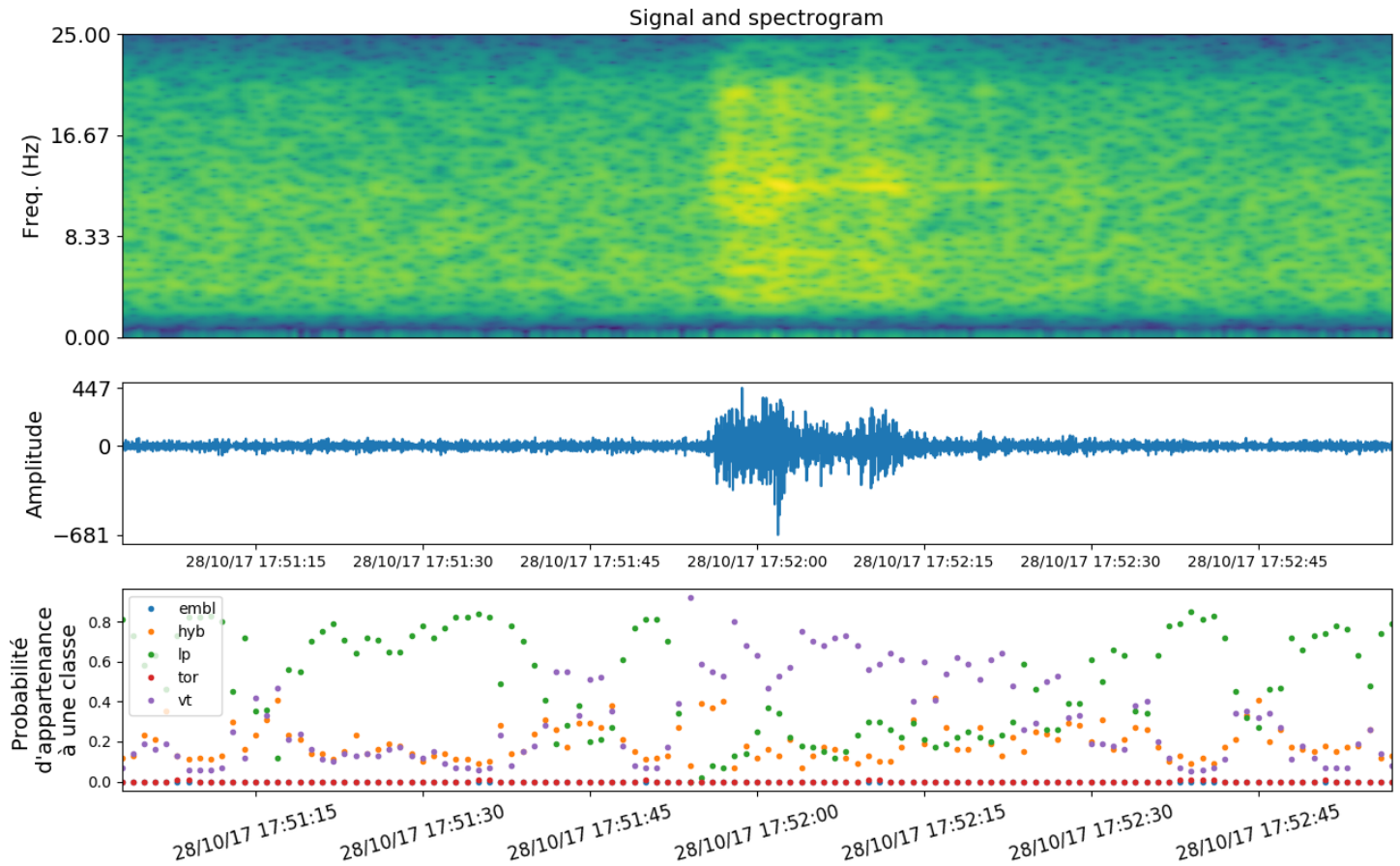


FIGURE 4.6 – Spectrogramme, forme d’onde et probabilités d’appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Emboité a été détecté par l’OVSG ; Emboité : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

Tornillo

Sur la Figure 4.7, nous voyons une dernière fois un enregistrement de 2 min incluant un événement de type Tornillo. Cet événement a été détecté par l’OVSG mais classé comme LP. Notre modèle, ici, a tendance à considérer le bruit sismique comme des VT avec des probabilités qui oscillent autour de 0.6. Pendant l’événement Tornillo, la probabilité maximale est celle d’être un LP (autour de 0.8). Mais alors que pour les exemples précédents, la probabilité d’être un Tornillo n’est jamais allée au-delà de 0, ici elle monte jusqu’à 0.2. Dans ce cas, le modèle échoue à classer correctement l’événement.

4.2.3 Discussion

Appliquer le modèle construit lors du chapitre 3 directement sur le signal continu nous permet de détecter les événements que l’OVSG a lui-même détectés et peut même en détecter d’autres. Néanmoins, le travail de détection fait ici conserve un caractère manuel. Mettre en place un système de seuil au-delà duquel on déclenche automatiquement une détection est encore périlleux.

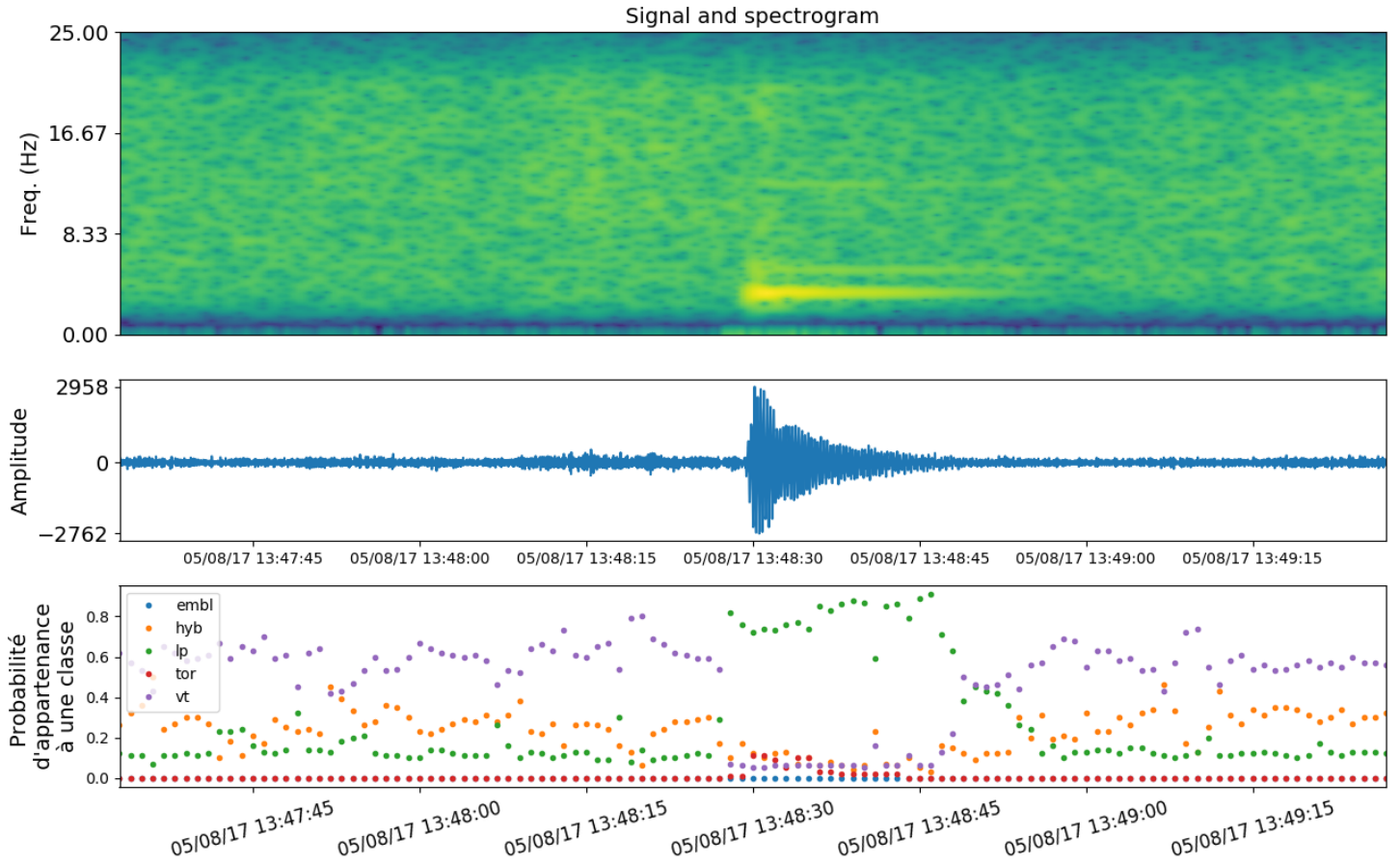


FIGURE 4.7 – Spectrogramme, forme d’onde et probabilités d’appartenance à chacune des 5 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Tornillo a été détecté par l’OVSG ; Emboîté : Bleu, Hybride : Orange, LP : Vert, Tornillo : Rouge, VT : Violet

En effet, on a vu sur la Figure 4.3 que le modèle peut considérer que certaines fenêtres sont des LP alors que c’est du bruit sismique et parfois le modèle classe le bruit ambiant comme VT. Mettre un système de seuil pour le déclenchement automatique dans ce genre de cas ne serait pas efficace et provoquerait beaucoup de trop fausses détections. Ensuite, on a vu qu’avoir plusieurs classes dans notre classification peut faire hésiter notre système, comme sur la Figure 4.5 où le modèle hésite entre Hybride et VT. Encore une fois, si on mettait en place un système de seuil ici, bien que le modèle comprenne qu’on est plus dans du bruit (à travers la baisse de la probabilité LP), il ne saurait pas quelle classe attribuer. Les classes VT et Hybride s’annulent entre elles et il n’y aurait pas de détections. Et enfin, le dernier problème rencontré est lié à la taille des événements Tornillos et Emboîtés. Pendant la phase d’apprentissage, les descripteurs ont été calculés à partir des durées renseignées par l’OVSG donc des fenêtres de durée variable alors que pendant la phase de détection, la fenêtre glissante sur laquelle on effectue notre prédiction est d’une taille fixe de 5 secondes. Et comme le montre la Figure 4.1, 80% des événements ont des durées inférieures à 6 secondes, mais il y a néanmoins des événements de durée de plusieurs dizaines de secondes comme les Tornillos ou les Emboîtés. Et donc à travers ces fenêtres de 5 secondes, le modèle reconnaît les Tornillo comme LP et les Emboîtés comme VT. Il faut voir si en augmentant la taille de la fenêtre

glissante, on arrive à mieux reconnaître certaines classes. À partir de maintenant, on comparera les résultats obtenus avec une fenêtre de 5 secondes avec ceux obtenus pour une fenêtre de 15 secondes.

4.3 Introduction d'une classe bruit

4.3.1 Ajout de la classe bruit dans la phase d'apprentissage

Pour remédier au problème du bruit de fond qui est reconnu comme LP ou VT par le modèle, nous introduisons une classe bruit dans notre phase d'apprentissage. De cette façon, si l'apprentissage est bien fait, c'est à dire que l'on a dans notre base d'apprentissage une variété de Bruits correspondant à la variété que l'on peut retrouver tout au long de notre période d'étude, alors le bruit sismique sera reconnu comme tel par le modèle. Pour sélectionner, les "événements" que nous considérerons comme bruit, nous recensons tous les événements sismo-volcaniques, tectoniques ou même anthropiques détectés par l'OVSG et estimons que tout ce qui n'est pas un de ces événements est du bruit de fond sismique.

TABLE 4.1 – Matrice de confusion obtenue avec un rapport de 50 :50 entre les jeux d'entraînement et de test pour le catalogue de l'OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest (n_estimators=100, criterion='entropy', bootstrap=True, class_weight=None), Score moyen obtenu après 10 réalisations

True class	Predicted Class						Recall (%)
	VT	Nested	LP	Hyb	Tor	Noise	
VT	203	2	1	18	0	0	90.6
Nested	5	23	1	5	0	0	67.7
LP	1	1	41	6	0	0	83.7
Hyb	25	2	7	69	0	0	67.0
Tor	0	1	1	0	13	0	86.7
Noise	0	0	0	0	0	61	100
Precision (%)	86.8	79.3	80.4	70.4	100	100	
Accuracy (%)							84.4 ±1.0

On voit sur la table 4.1 que l'introduction de la classe bruit dans notre modèle ne baisse pas les performances du modèle avec une accuracy de 84.4%. Le modèle parvient à classer parfaitement les signaux de classe bruit avec 100% de recall et de precision. On peut voir sur la figure 4.8 une représentation de la valeurs des meilleurs descripteurs de la classe bruit, à savoir les entropies de Shannon et de Rényi. Ces deux descripteurs permettent de bien séparer la classe bruit de l'ensemble des autres classes.

De plus, nous avons choisi de créer un modèle à partir de signaux de la même taille que notre fenêtre glissante. Donc si notre fenêtre est de 15 secondes les signaux utilisés pour l'entraînement auront une durée de 15 secondes. Si la durée réelle de l'événement est plus courte alors il y aura un petit peu de bruit de fond dans la fenêtre utilisée pour l'entraînement. En revanche, si la durée réelle de l'événement est plus longue que les 15 secondes considérées alors ne sera pris en compte que le début de l'événement. De cette façon, on imagine être plus précis dans nos détections.

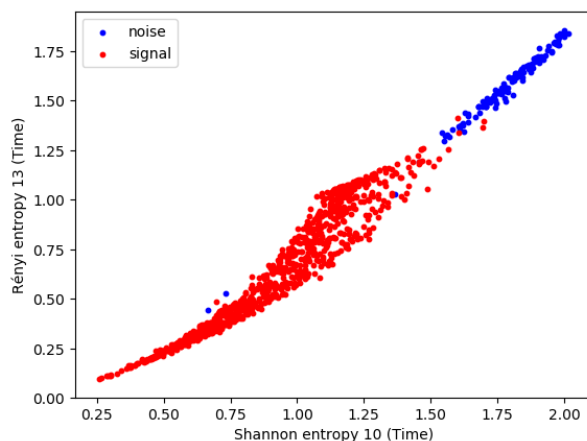


FIGURE 4.8 – Représentation spatiale des meilleurs descripteurs de la classe bruit. Abscisse : Entropie de Shannon ; Ordonnée : Entropie de Rényi ; Rouge : Signal ; Bleu : Bruit

La seule perte que cette méthode pourrait engendrer serait que les fins des événements les plus longs soient considérées comme du bruit par le modèle. De plus, en utilisant des fenêtres de même tailles pour tous les événements de la base d'apprentissage ainsi que pour la fenêtre glissante sur le signal continu alors le descripteur correspondant à la taille des signaux sera inutile, car toujours égal à 1500 échantillons dans cet exemple (15 s x 100 Hz).

VT

La figure 4.9 représente l'application du modèle à 6 classes sur le signal continu avec des fenêtres de 5 secondes sur la figure 4.9.a et des fenêtres de 15 secondes sur la figure 4.9.b. Sur la première, on voit que les deux événements ont été détectés et bien classés avec une probabilité forte (>0.8) ce qui était déjà le cas sur la figure 4.2 mais ici, le reste du temps la probabilité la plus forte est celle d'avoir du bruit de fond.

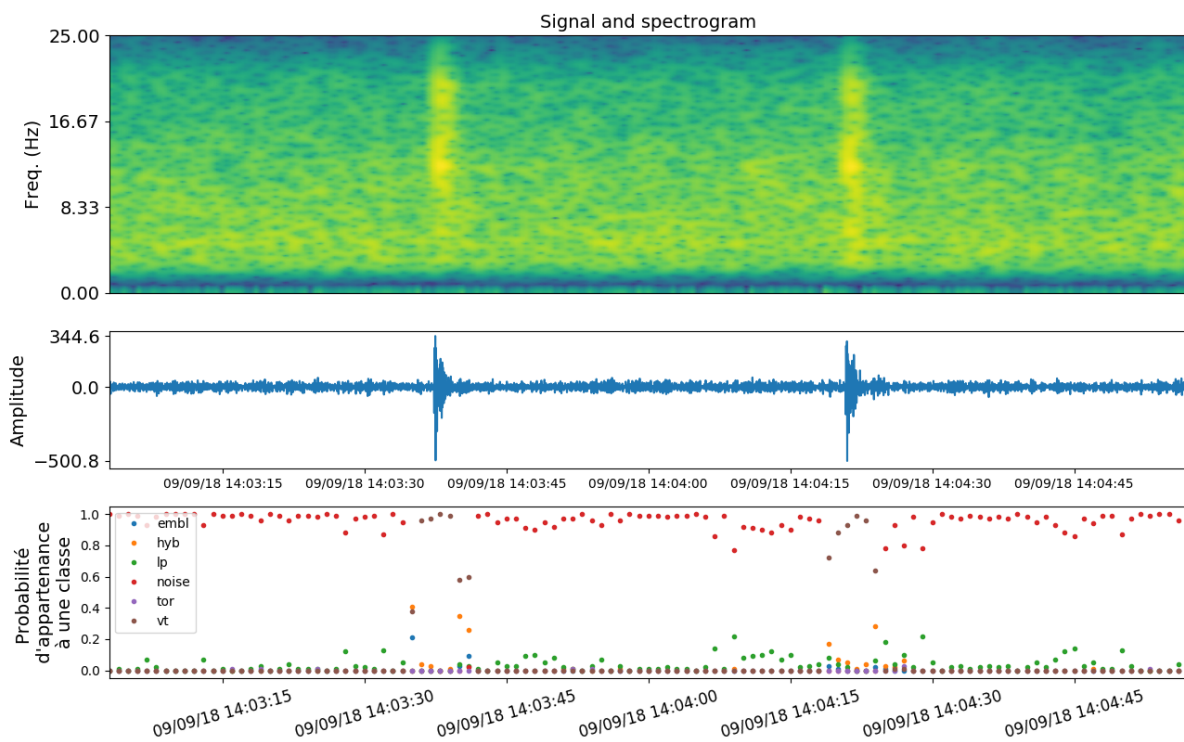
Sur la figure 4.9.b, on observe à peu près la même chose à savoir, lors des événements une probabilité forte (>0.8) d'avoir un VT et en dehors des événements une probabilité forte d'avoir du bruit. La différence entre les deux méthodes est que lorsqu'on utilise des fenêtres de 15 secondes alors il y a plus de fenêtre pour lesquelles on détecte un événement (si on considère un seuil de 0.8 pour déclencher la détection). Ce qui est logique, plus la fenêtre est longue, plus il y aura de fenêtres consécutives dans lesquelles le signal sera présent.

LP

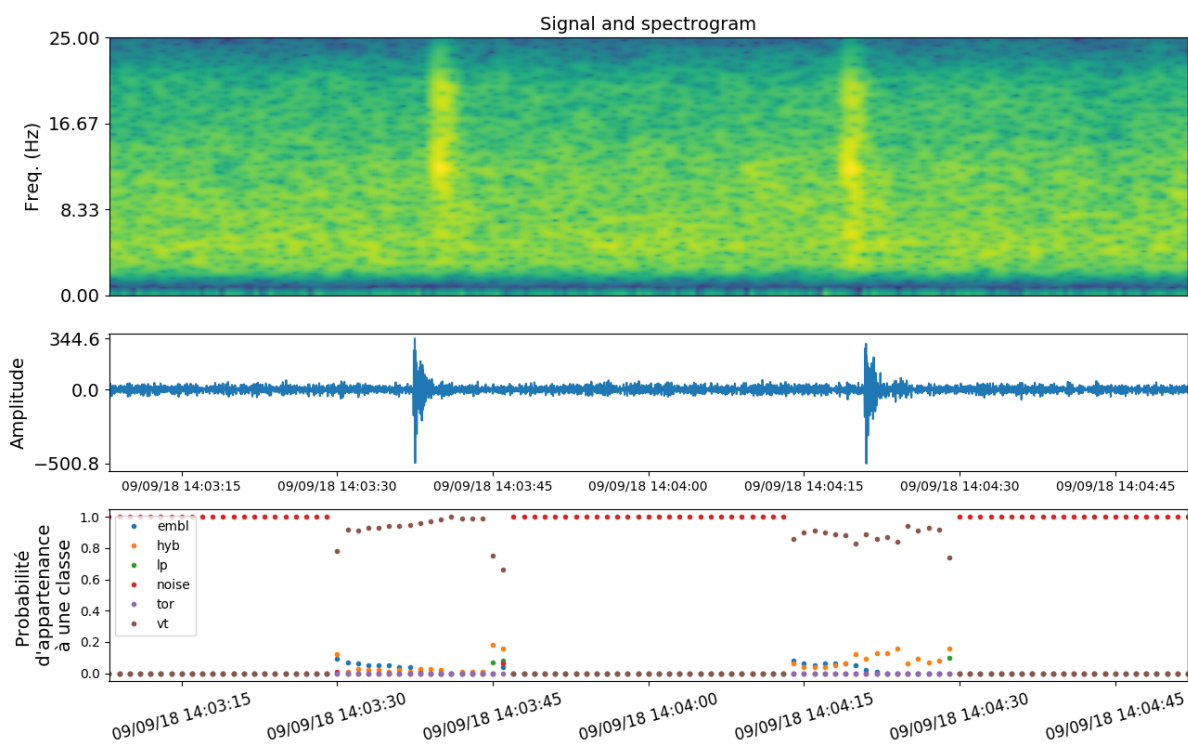
La figure 4.10 représente l'application du modèle à 6 (5+bruit) classes sur le même signal que la figure 4.3, avec en a) des fenêtres de 5 secondes et en b) des fenêtres de 15 secondes. Dans les deux cas, le modèle a été capable de différencier l'événement LP du bruit de fond ce que le modèle à 5 classes n'a pas pu faire. De plus on voit qu'en appliquant une fenêtre glissante de 15 secondes la probabilité d'avoir un événement LP est plus forte, plus longtemps avec 13 fenêtres consécutives ayant une probabilité d'être un LP supérieure à 0.8 contre 2 fenêtres consécutives avec la fenêtre de 5 secondes.

Hybride

La figure 4.11 représente l'évolution de la probabilité d'appartenance aux 6 (5+bruit) classes

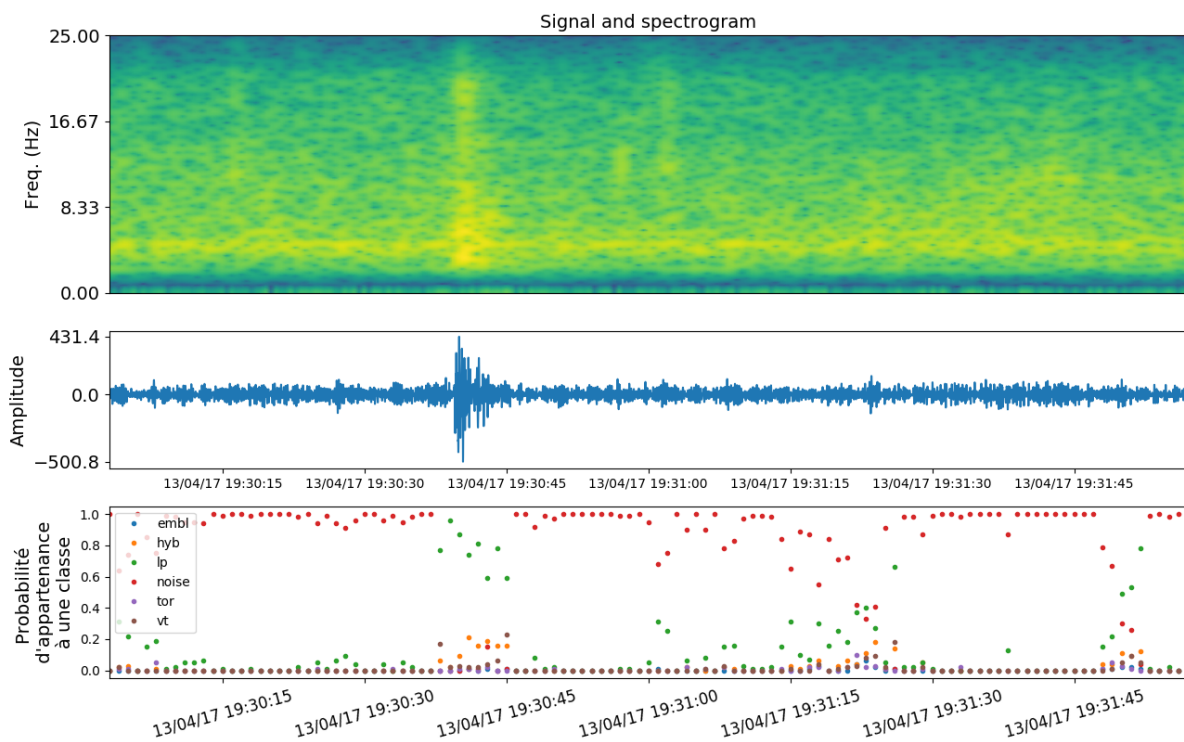


(a) Fenêtre de 5 secondes

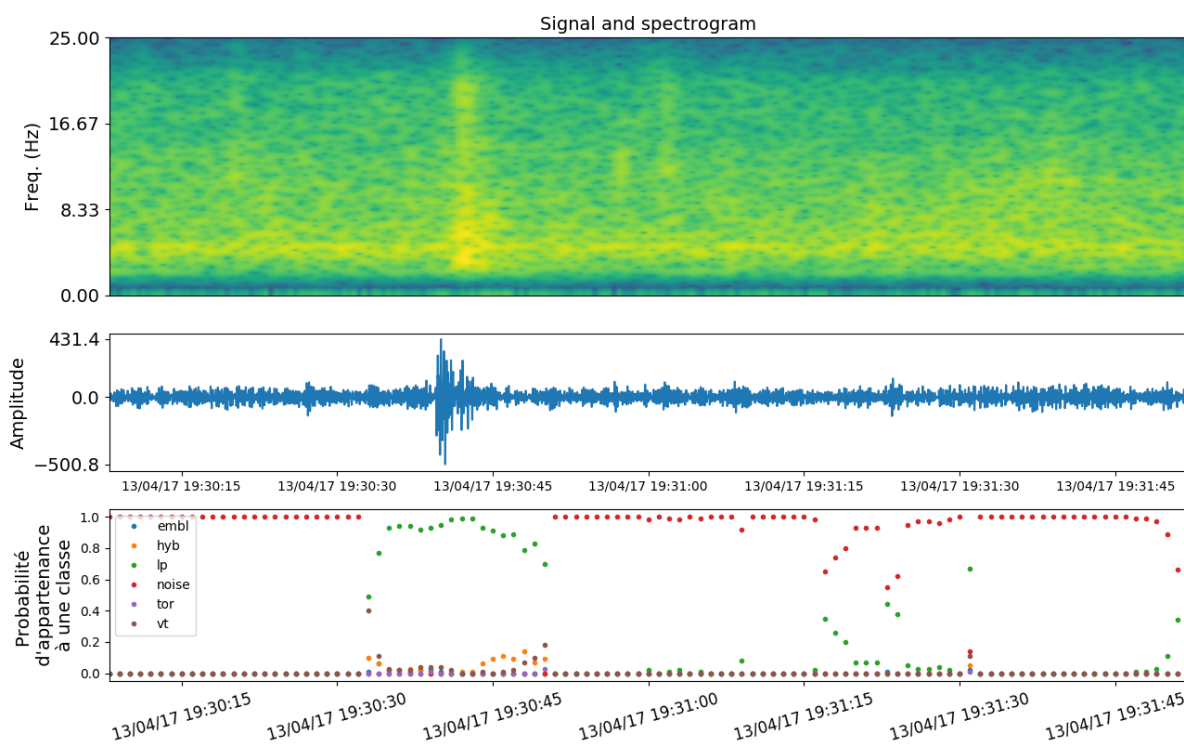


(b) Fenêtre de 15 secondes

FIGURE 4.9 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 6 classes du modèle sur 2 minutes de signal pendant lesquelles 2 VT ont été détectés par l'OVSG ; Emboîté : Bleu, Hybride : Orange, LP : Vert, Bruit : Rouge, Tornillo : Violet, VT : Brun



(a) Fenêtre de 5 secondes



(b) Fenêtre de 15 secondes

FIGURE 4.10 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 6 classes du modèle sur 2 minutes de signal pendant lesquelles 1 LP a été détecté par l'OVSG ; Emboité : Bleu, Hybride : Orange, LP : Vert, Bruit : Rouge, Tornillo : Violet, VT : Brun

dans 2 minutes de signal continu contenant un événement Hybride. Ici peu de différence entre les deux sous-figures. Sur les premières fenêtres, quand l'événement commence à apparaître dans celle-ci et que le modèle est capable de dire qu'on n'est plus dans du bruit de fond alors les probabilités d'appartenance sont partagées entre VT et Hybride, ensuite la probabilité qu'on ait un Hybride augmente jusqu'à atteindre 0.8 dans le cas à 5 secondes et 0.9 dans le cas à 15 secondes. Ensuite quand l'événement sort de la fenêtre glissante la probabilité d'être un Hybride diminue et celle d'être un LP augmente. Encore une fois, cela est cohérent avec ce que l'on sait des signaux hybrides avec une arrivée d'onde qui tient plus du VT et une coda qui ressemble plus à celle d'un LP. Le problème dans ces cas de probabilités partagées entre plusieurs classes est qu'il est difficile de déterminer automatiquement le début et la fin de l'événement.

Une chose intéressante à noter sur ces deux minutes de signal est qu'on a à environ 15 :43 :45 un autre événement qui n'est pas recensé par l'observatoire. Lors de cet événement, les deux cas, pour une fenêtre de 5 secondes et de 15 secondes, présentent une baisse de la probabilité d'être du bruit et une hausse de la probabilité d'être un LP. La probabilité d'être un LP ne monte pas très haut dans les deux cas (environ 0.4) donc un système de seuil n'aurait peut-être pas détecté cet événement, mais le modèle voit tout de même qu'il se passe quelque chose.

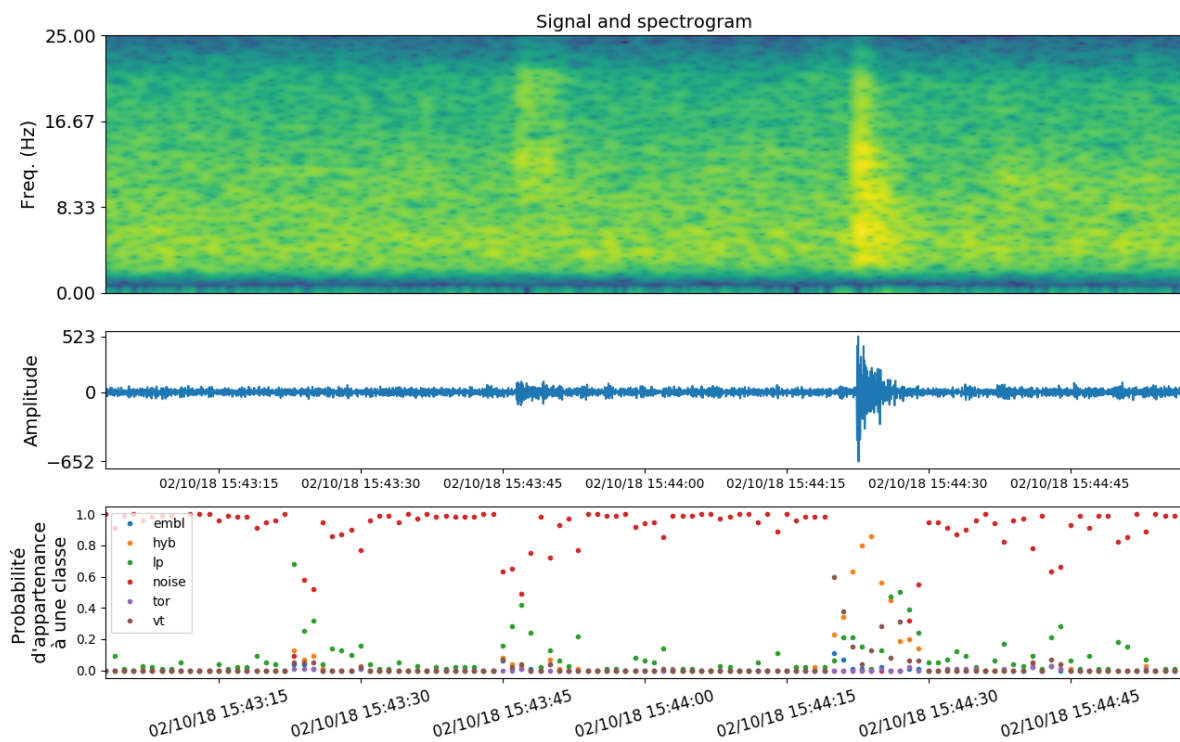
Emboîté

La figure 4.12 représente l'évolution de la probabilité d'appartenance aux 6 classes dans 2 minutes de signal continu contenant un événement Emboîté. Ici, on note des différences entre les deux sous-figures dans le comportement des probabilités d'appartenance au moment de l'événement emboîté. En effet sur la figure 4.12.a, lorsqu'on utilise des fenêtres de 5 secondes les probabilités sont partagées entre toutes les classes. Mais il y a une nette amélioration par rapport à la figure 4.6. En effet, ici, la probabilité d'appartenance à la classe Emboîté est montée jusqu'à 0.6 alors qu'elle était restée à 0 dans le cas précédent. Mais au cours de l'événement, la probabilité de bruit augmente également au-dessus de 0.9. Il est donc difficile de déduire une prédiction avec ces résultats. Au contraire, les fenêtres de 15 secondes donnent de meilleurs résultats avec 13 fenêtres consécutives dans lesquels la probabilité maximale est celle d'avoir un Emboîté. Il n'y a également qu'une seule fenêtre pour laquelle la probabilité d'avoir du bruit est maximale. Néanmoins, sur la fin de l'événement, le modèle prédit un LP puis un VT. Donc, pour ce genre d'événements longs, le cas avec des fenêtres de 15 secondes est plus efficace car il a moins de risque de confondre l'événement avec du bruit, il est capable de bien reconnaître le début de l'événement, cependant les prédictions sont fausses sur la fin de l'événement.

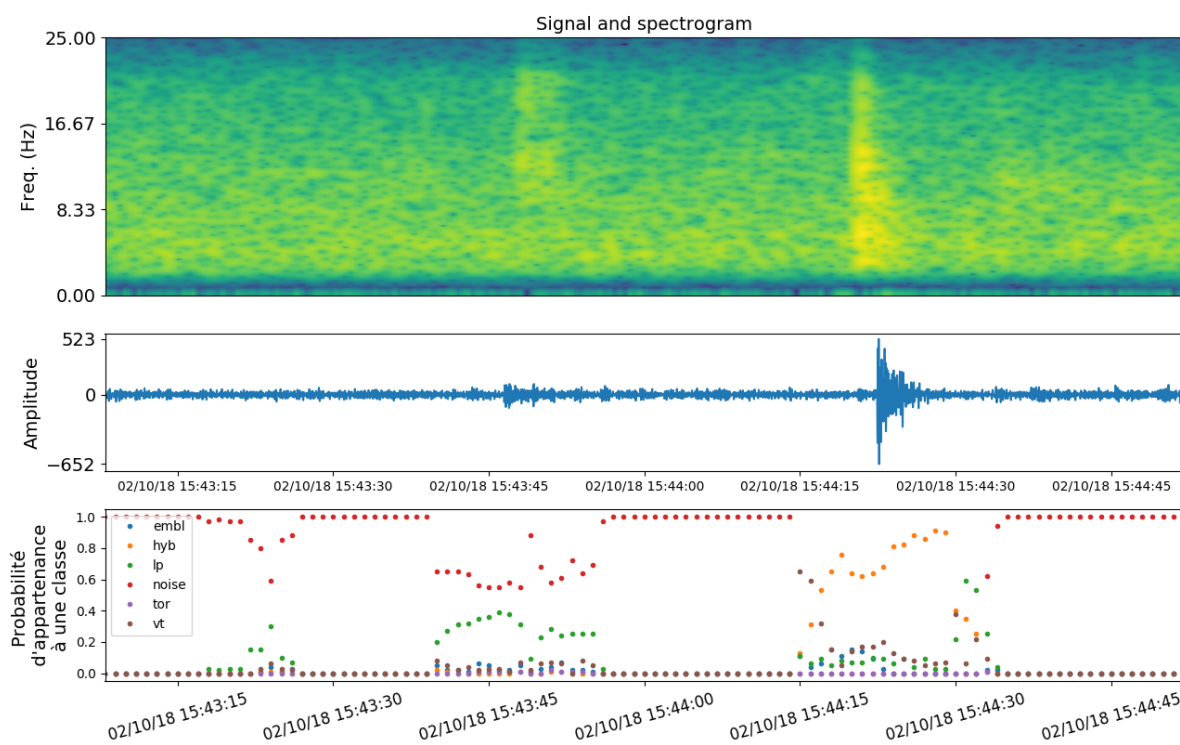
Tornillo

La figure 4.13 représente l'évolution de la probabilité d'appartenance aux 6 classes dans 2 minutes de signal continu contenant un événement Tornillo. Encore une fois, pour un signal long la détection avec des fenêtres de 5 secondes (4.13.a) se révèle moins efficace. Tout d'abord, le modèle voit un événement LP avec une forte probabilité (environ 0.8) puis une fenêtre pour laquelle la probabilité d'avoir un Tornillo est très forte (> 0.8) puis les probabilités se partagent entre Tornillo, LP et Bruit (les trois aux alentours de 0.3) jusqu'à la fin de l'événement où la probabilité de bruit remonte jusqu'à 1.

Pour le cas avec des fenêtres de 15 secondes (4.13.b), au début de l'événement, il y a également des fenêtres pour lesquelles la probabilité maximale est celle des LP (> 0.8) puis pendant 23

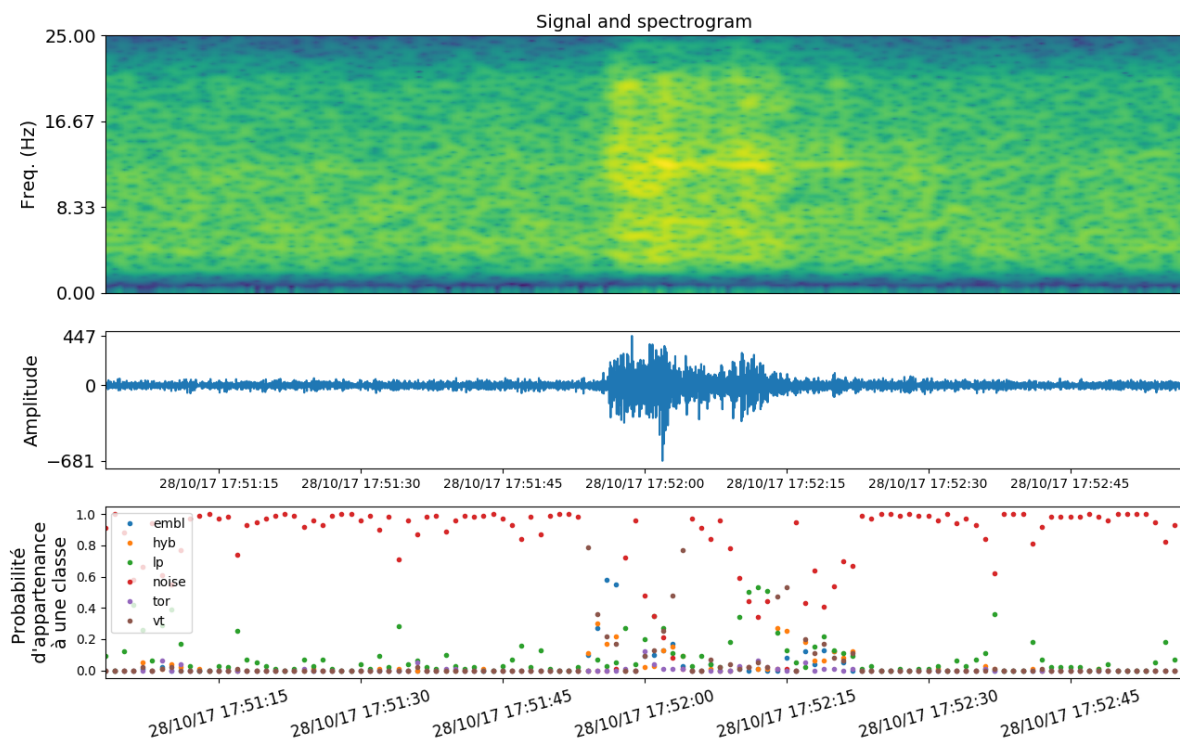


(a) Fenêtre de 5 secondes

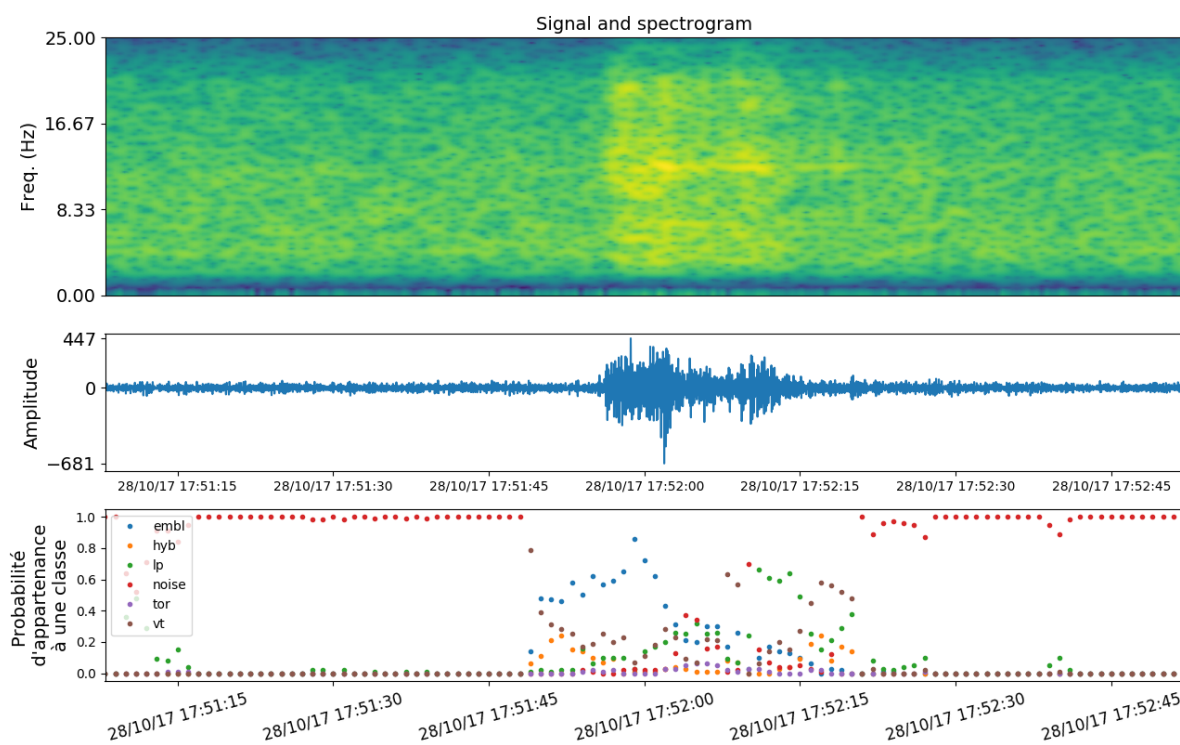


(b) Fenêtre de 15 secondes

FIGURE 4.11 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 6 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Hybride a été détecté par l'OVSG ; Emboité : Bleu, Hybride : Orange, LP : Vert, Bruit : Rouge, Tornillo : Violet, VT : Brun



(a) Fenêtre de 5 secondes



(b) Fenêtre de 15 secondes

FIGURE 4.12 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 6 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Emboîté a été détecté par l'OVSG ; Emboîté : Bleu, Hybride : Orange, LP : Vert, Bruit : Rouge, Tornillo : Violet, VT : Brun

fenêtres consécutives, c'est la probabilité d'avoir un Tornillo qui est maximale, atteignant parfois 1. À la fin de l'événement, la probabilité de bruit remonte progressivement.

Discussion

L'ajout de la classe Bruit dans notre modèle a permis de distinguer plus facilement le bruit de fond des événements, notamment dans le cas de la reconnaissance de LP, car là où la distinction LP/Bruit de fond se faisait difficilement par le modèle à 5 classes, elle est ici évidente. De plus, l'harmonisation de la taille des fenêtres lors de l'apprentissage avec la taille de la fenêtre glissante a permis d'avoir une meilleure reconnaissance des événements de longue durée. Cette reconnaissance a même été accrue lorsqu'on augmente la taille de ces fenêtres à 15 secondes, en effet de cette façon le modèle apprend plus de détails sur ces événements longs, là où les fenêtres courtes pouvaient laisser des confusions notamment entre LP et Tornillos ou entre VT et Emboîtés. On va donc pour la suite garder la taille de 15 secondes pour les fenêtres glissantes. Enfin, la dernière amélioration à faire est d'automatiser le système de détection. On voudrait mettre en place un système de seuil sur les probabilités au-delà duquel on considère qu'il y a un événement. Mais plusieurs problèmes se posent à nous, premièrement le fait d'avoir plusieurs classes Non-Bruit fait que les probabilités associées à ces classes peuvent s'annuler entre elles comme les cas de la figure 4.12.b où au cours de l'événement plusieurs classes sont prédites. Ou le cas de la figure 4.11.b où le fait que la probabilité d'avoir un VT et celle d'avoir un LP ne sont pas minimales baissent la probabilité d'avoir un événement Hybride.

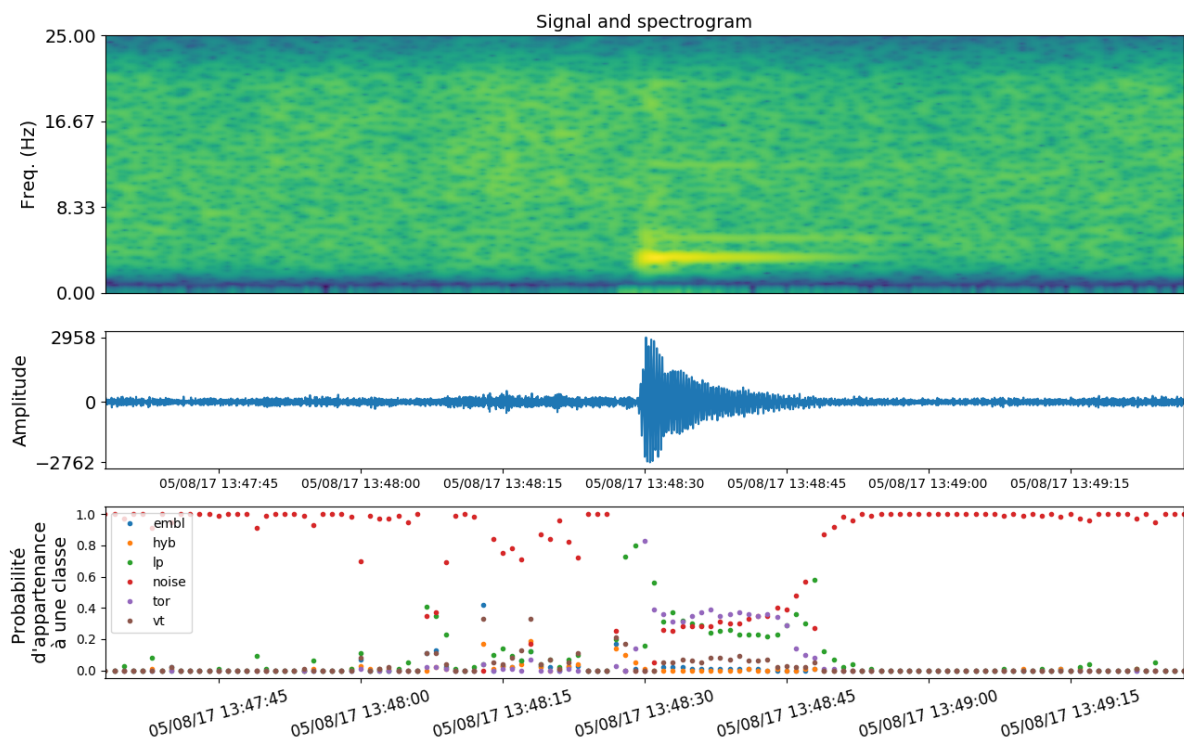
4.3.2 Classe Signal contre classe Bruit appliqué sur deux semaines

Pour résoudre la difficulté d'avoir plusieurs classes possibles faisant baisser la probabilité de la meilleure classe et de ce fait rendant difficile l'utilisation d'un seuil pour le déclenchement des détections. Nous choisissons de construire un modèle Signal contre Bruit pour déterminer les moments où le modèle considère qu'il y a un événement. Pour construire ce modèle nous utilisons la même base de données qu'avant, les mêmes événements, à la seule différence que les classes VT, LP, Hybrides, Emboîtés et Tornillo sont regroupés dans la même classe "Signal".

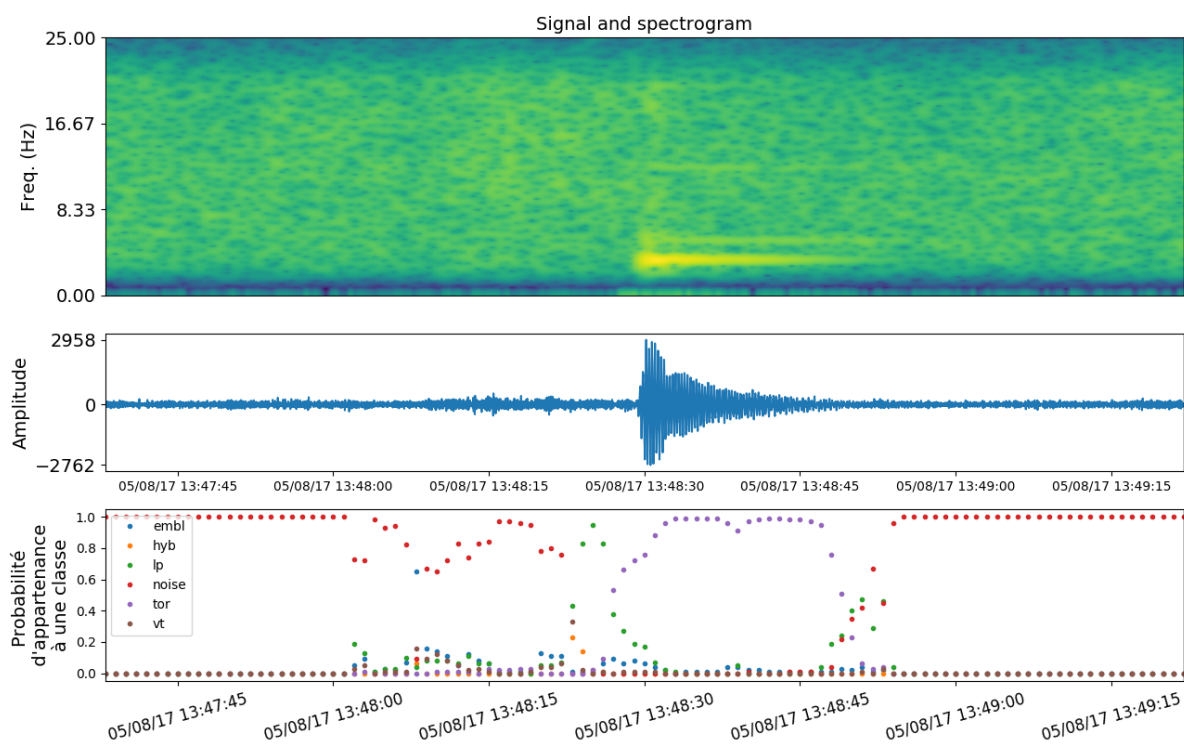
La figure 4.14 présente les mêmes deux minutes de signal que les figures 4.5 et 4.11 et les mêmes tailles de fenêtres que la figure 4.11.b. On voit qu'au moment de l'événement la probabilité d'avoir du signal est entre 0.9 et 1 alors que sur la figure 4.11 la probabilité maximale pour que l'événement détecté soit un Hybride oscille entre 0.6 et 0.9. Pour l'événement faible aux alentours de 15h43, la probabilité que ce soit du bruit donnée par le modèle baisse jusqu'à moins de 0.5 alors que sur la figure 4.11.b, cette probabilité restait supérieure à 0.6. On a légèrement amélioré les résultats mais pour un événement de faible amplitude le modèle reste confus entre les classes bruit et signal.

Nous appliquons notre modèle Bruit contre Signal à deux semaines de données puis nous comparons les détections obtenues avec celles de l'OVSG. Du 14 Septembre 2018 au 30 Septembre 2018, l'OVSG a détecté 194 événements (103 VT, 89 Emboîté et 2 LP). On détermine un seuil tel que si la probabilité qu'une fenêtre soit du signal est supérieure à 0.8 alors on déclenche une détection.

Sur la période, 20144 fenêtres ont dépassé le seuil de déclenchement. Dans une première étape,



(a) Fenêtre de 5 secondes



(b) Fenêtre de 15 secondes

FIGURE 4.13 – Spectrogramme, forme d'onde et probabilités d'appartenance à chacune des 6 classes du modèle sur 2 minutes de signal pendant lesquelles 1 Tornillo a été détecté par l'OVSG ; Emboité : Bleu, Hybride : Orange, LP : Vert, Bruit : Rouge, Tornillo : Violet, VT : Brun

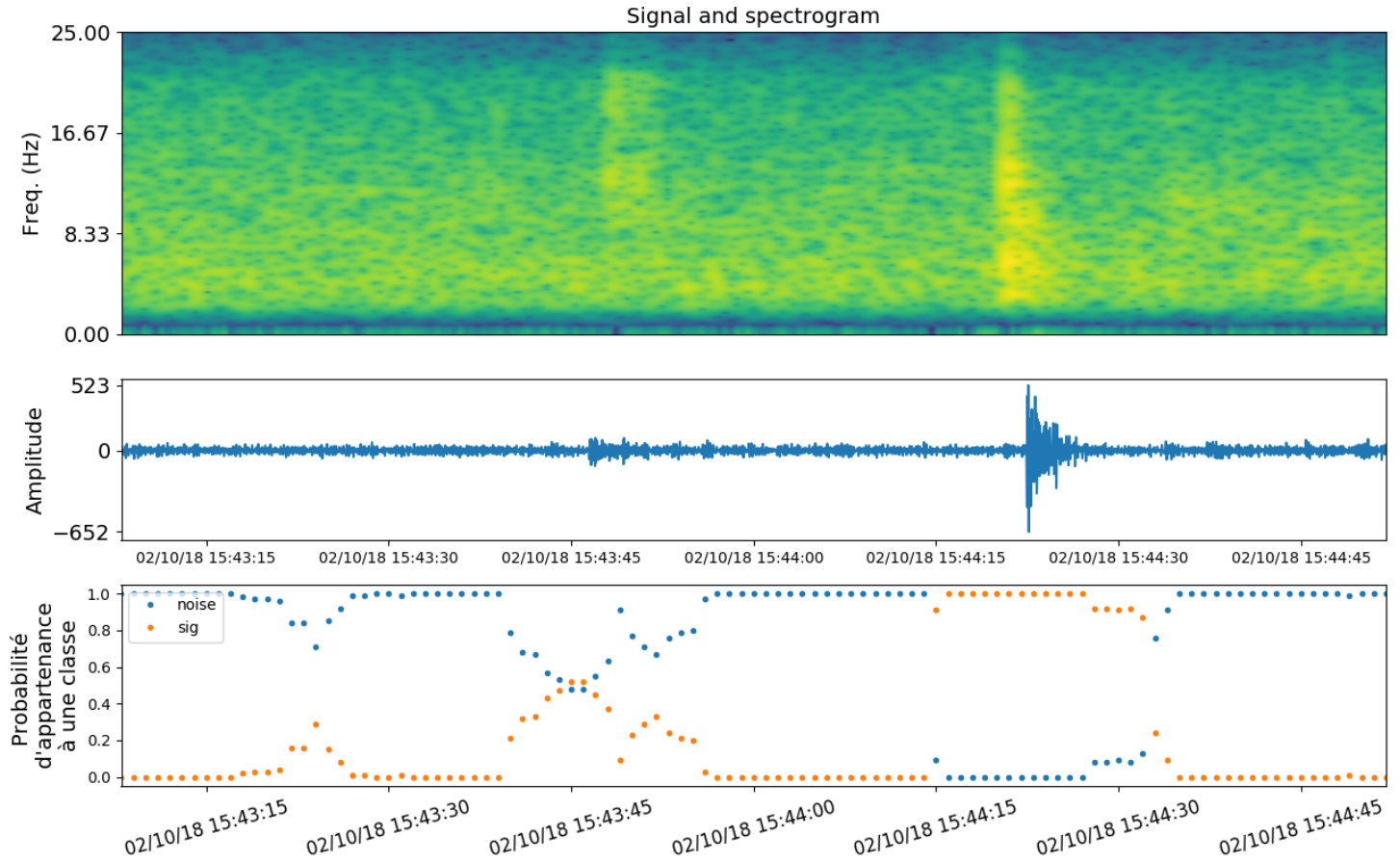


FIGURE 4.14 – Spectrogramme, forme d'onde et probabilités d'appartenance Signal contre Bruit sur 2 minutes de signal pendant lesquelles 1 Hybride a été détectés par l'OVSF ; Bruit : Bleu, Événement : Orange

nous regroupons les fenêtres consécutives qui ont déclenché sur le même événement. Après cette étape, on a 5667 événements selon notre détecteur. Seuls 16 événements détectés par l'observatoire n'ont pas été détectés par notre système donc 92% des événements détectés par l'observatoire l'ont aussi été par notre système. On a vu que les événements non détectés s'étaient produits à des moments où le bruit de fond sismique était très important à la station TAG.

La seconde étape consiste à enlever les fausses détections de nos résultats pour appliquer notre modèle de classification aux événements restants, on a vu sur les figure 4.2 à 4.14 que lorsqu'il y a un événement dans le signal continu alors plusieurs fenêtres consécutives sont déclenchées. La figure 4.15 représente la durée des déclenchements de la détection de notre système, autrement dit, la distribution du nombre de fenêtres consécutives ayant déclenché. On voit qu'il y a beaucoup de déclenchements de 15 secondes, c'est à dire qu'une seule fenêtre est au-dessus du seuil, celle d'avant et la suivante sont en dessous de 0.8. Ensuite, plus la durée de déclenchement est longue moins il y a de déclenchements, jusqu'à 28 secondes, c'est à dire 13 fenêtres consécutives ayant déclenché. Au-delà de 28 secondes de déclenchements successifs, le nombre d'événements détectés augmente, 30 événements de 28 secondes contre 158 événements de 29 secondes.

Parmi les 178 événements détectés à la fois par l'observatoire et par notre système, 40 correspondent à des durées de détections inférieures à 28 secondes et 138 correspondent à des durées de détection supérieures à 28 secondes, soit 71% des événements détectés par l'OVSG. On classera les événements au-dessus de cette limite de 29 secondes. Les événements en dessous seront considérés comme des fausses détections car ils n'auront pas déclenché suffisamment de fenêtres consécutives, et ce, malgré la présence d'événements également détectés par l'observatoire. On considère que le seuil utilisé pour la détection est imparfait, car si au cours de l'événement une fenêtre a une probabilité d'être un événement inférieur à 0.8 même de très peu alors la détection est coupée.

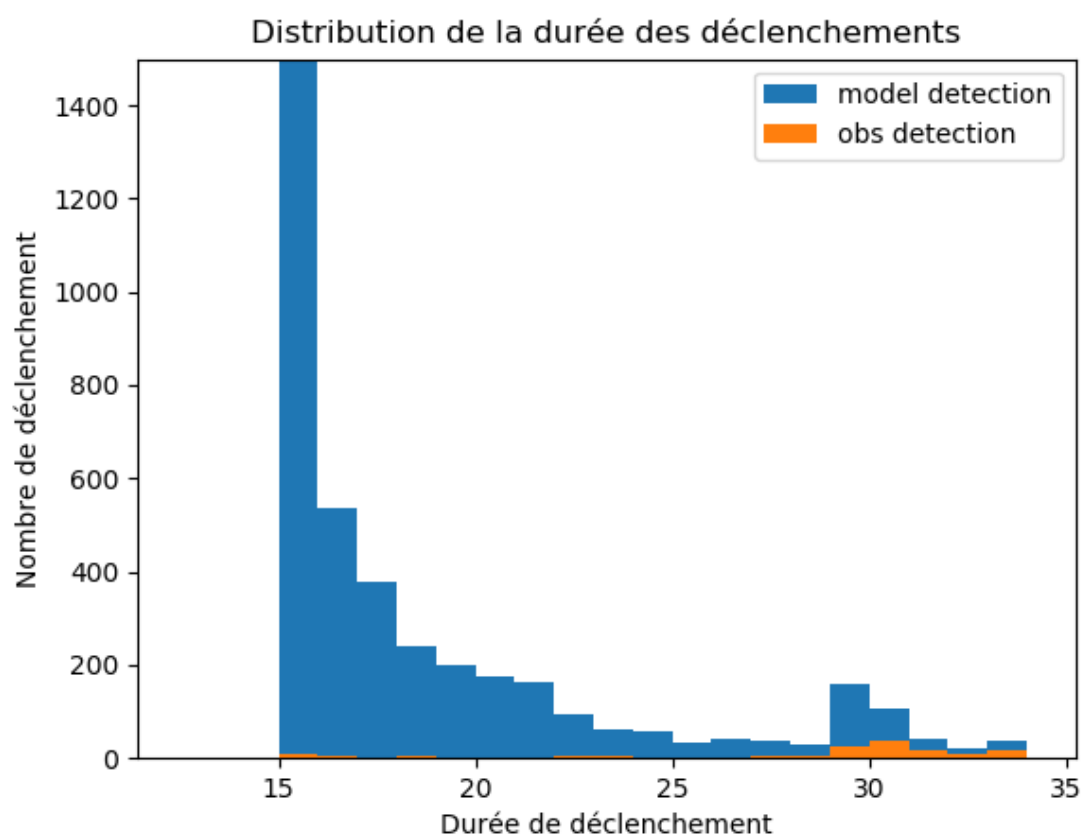


FIGURE 4.15 – Distribution de la durée des détections, les détections en orange correspondent aux détections de l'OVSG

4.3.3 Classification appliquée aux événements détectés

On va donc s'intéresser aux événements au-dessus des 29 secondes de déclenchement, car 71% des événements détectés par l'OVSG sur la période s'y trouve. Il y a 403 détections avec une durée supérieure à 29 secondes dont 138 ont été également détectés par l'OVSG. On classera ces événements, mais on veut réduire la taille des fenêtres pour avoir des signaux de tailles correspondants à ce qu'on peut trouver à la Soufrière, en effet 71% des événements détectés par l'observatoire sur la période n'ont pas une durée supérieure à 28 secondes.

La figure 4.16 représente l'écart entre les temps d'arrivée déterminés par l'OVSG et les temps de déclenchement de la première fenêtre pour les 138 événements détectés par l'OVSG et notre modèle avec une durée de déclenchement supérieure à 28 secondes. Pour plus de la moitié des événements, l'écart entre la fenêtre de déclenchement et le temps d'arrivée est de 6 secondes. Et une majorité des événements ont un écart entre les temps de fin déterminés par l'OVSG et les temps de fin de dernière fenêtre autour de 20 secondes.

Pour raccourcir la durée des événements détectés par le modèle et avoir des tailles de signaux correspondant à ceux avec lesquels nous avons entraîné le modèle, nous enlevons 3 secondes de signal au début de la première fenêtre pour correspondre aux 6 secondes d'écart déterminé plus haut, car le modèle de classification est également construit en rajoutant 3 secondes de signal avant le temps d'arrivée déterminé par l'observatoire. Nous enlèverons 20 secondes de signal à la fin de nos détections pour raccourcir les événements et ainsi avoir une fin de signal correspondant à la fin de l'événement. Maintenant, nous passons à l'étape de classification, nous allons appliquer notre modèle aux événements ayant une durée de déclenchement supérieur à 29 secondes. Nous comparerons nos résultats avec la classification de l'OVSG et nous regarderons les signaux n'ayant pas été détectés par l'observatoire.

Sur les 138 événements détectés par l'OVSG, 68 sont classés comme VT, 69 comme Emboîté et 1 comme LP. Lorsqu'on applique notre modèle de classification à ces événements, la classe prédite est la classe VT pour les 138 événements. Et lorsqu'on applique notre modèle à l'ensemble des 403 signaux détectés ayant une longueur supérieure à 29 secondes, la classe prédite est également toujours VT. En regardant les signaux en détail, on voit que notre modèle se trompe et que ces signaux ne sont pas exclusivement des VT. Le problème vient de la taille de la fenêtre des signaux que l'on cherche à classer, notre étape précédente voulant raccourcir les signaux pour ne conserver que ce qui correspond à un événement s'est révélé inefficace. Il y a encore beaucoup trop de bruit sismique dans nos signaux comparé à la longueur des événements et notre modèle les considère comme VT.

On a vu que sur les 403 détections faites par le modèle, 138 ont également été détectées par l'observatoire, il reste donc 265 événements détectés par notre modèle, mais pas par l'observatoire. On a donc manuellement visualisé ces 265 événements et déterminer que 247 sont effectivement des événements soit un taux de bonnes détection de 93%.

4.3.4 Discussion

La détection d'événement par apprentissage supervisé est plutôt encourageante. En effet, l'application directe d'un modèle de classification à 5 classes d'événements sismo-volcaniques en continu permet de repérer des changements dans la nature du signal continu. Il subsiste le problème des prédictions faites par le modèle pendant les phases de bruit de fond. Si le bruit de fond ressemble à des LP selon le modèle alors il aura plus de mal à repérer les LP, si le bruit de fond ressemble à des VT alors le modèle reconnaîtra difficilement les VT. De plus, selon la taille choisie pour

la fenêtre glissante sur le signal continu pour la phase de calcul des descripteurs, certains types de signaux seront plus difficilement reconnus par le modèle. C'est le cas pour les Tornillos et les Emboîtés qui sont difficilement reconnaissables par le modèle lorsqu'on utilise une fenêtre glissante de 5 secondes alors que c'est plus simple avec une fenêtre de 15 secondes.

Pour bien caractériser le bruit de fond, on a voulu introduire la classe bruit dans notre modèle pour bien différencier le bruit de fond des événements LP ou VT. Cela se révèle efficace et la confusion entre VT, LP et bruit de fond ne se fait plus. Le problème de la reconnaissance des classes de signaux longs comme Tornillos ou Emboîtés peut se résoudre en utilisant une fenêtre glissante plus longue notamment 15 secondes. Mais un problème apparaît lorsqu'on a un événement ambigu pour lequel le modèle hésite ou alors si la classe prédite évolue au cours du signal évolue. Les probabilités s'annulent entre elles et avec un système de seuil de détection, même si le modèle repère un événement qui se manifeste par une augmentation de probabilité le seuil n'est pas forcément atteint.

Pour s'affranchir de ce problème où le fait qu'avoir plusieurs classes possibles parasite la détection, on a choisi de traiter le cas simple avec un classifieur Événement contre Bruit pour ensuite extraire ce qui est considéré comme un événement par ce modèle puis appliquer dans un deuxième temps un second modèle de classification, à 5 classes (celui présenté au chapitre 3), sur le signal extrait. Mais la nouvelle limitation de la méthode ici est qu'à cause de la longueur de notre fenêtre glissante les signaux considérés comme des événements sont très long comparés aux signaux utilisés pour l'apprentissage et de ce fait les prédictions faites par le modèle sont faussées par la quantité de bruit dans ces signaux à classer et ne prédit que des VT. Cela semble incohérent avec ce que montre les Figures 4.2, 4.3, 4.5 ou 4.6 à savoir que lorsque le signal est dominé par du bruit le modèle à cinq classes a tendance à classer les signaux comme LP.

Au final, on voit que la détection par apprentissage supervisé est plutôt puissante et est capable de détecter des signaux difficilement détectables par d'autres méthodes, le STA/LTA utilisé par l'OVSG par exemple. Mais les limitations de la méthode viennent plus du système de seuil utilisé pour définir une détection que de l'apprentissage supervisé lui-même. Utiliser un système de déclenchement plus complexe qu'une valeur de probabilité au-delà de laquelle il y a une détection et en dessous de laquelle il n'y en a pas pourrait être efficace. Ainsi, le modèle à 6 classes pourrait être viable pour détecter et classer les événements en continu.

Le deuxième problème lui est lié aux fenêtres et est multiple. Premièrement, selon la taille de la fenêtre glissante utilisée, les prédictions sont différentes. Deuxièmement, notre système est incapable de déterminer avec précision l'arrivée des événements et la fin des événements, à cause de cela, la classification des événements extraits dans le cas Événement contre Bruit est fausse.

Actuellement avec les résultats de la détection et de la classification, on est dans un cas semi-automatique qui aurait besoin d'un opérateur pour observer l'évolution des probabilités au cours du temps dans le cas à 6 (5+bruit) classes. Ou alors on devrait, après que le modèle ait fait une détection avec le modèle Bruit contre Événement, définir le temps d'arrivée et de fin du signal pour appliquer le modèle à 5 classes. Dans tous les cas un système de déclenchement plus complexe qu'un simple seuil doit être utilisé.

Pour conclure, la détection par apprentissage supervisé est très puissante et peut détecter plus d'événements que la méthode actuellement utilisée en observatoire, mais est encore imparfaite, un meilleur système de déclenchement doit être trouvé. Et cette méthode est encore incapable de trouver seule les temps de début et de fin des événements utiles pour classer efficacement ces événements, mais aussi pour la localisation des sources.

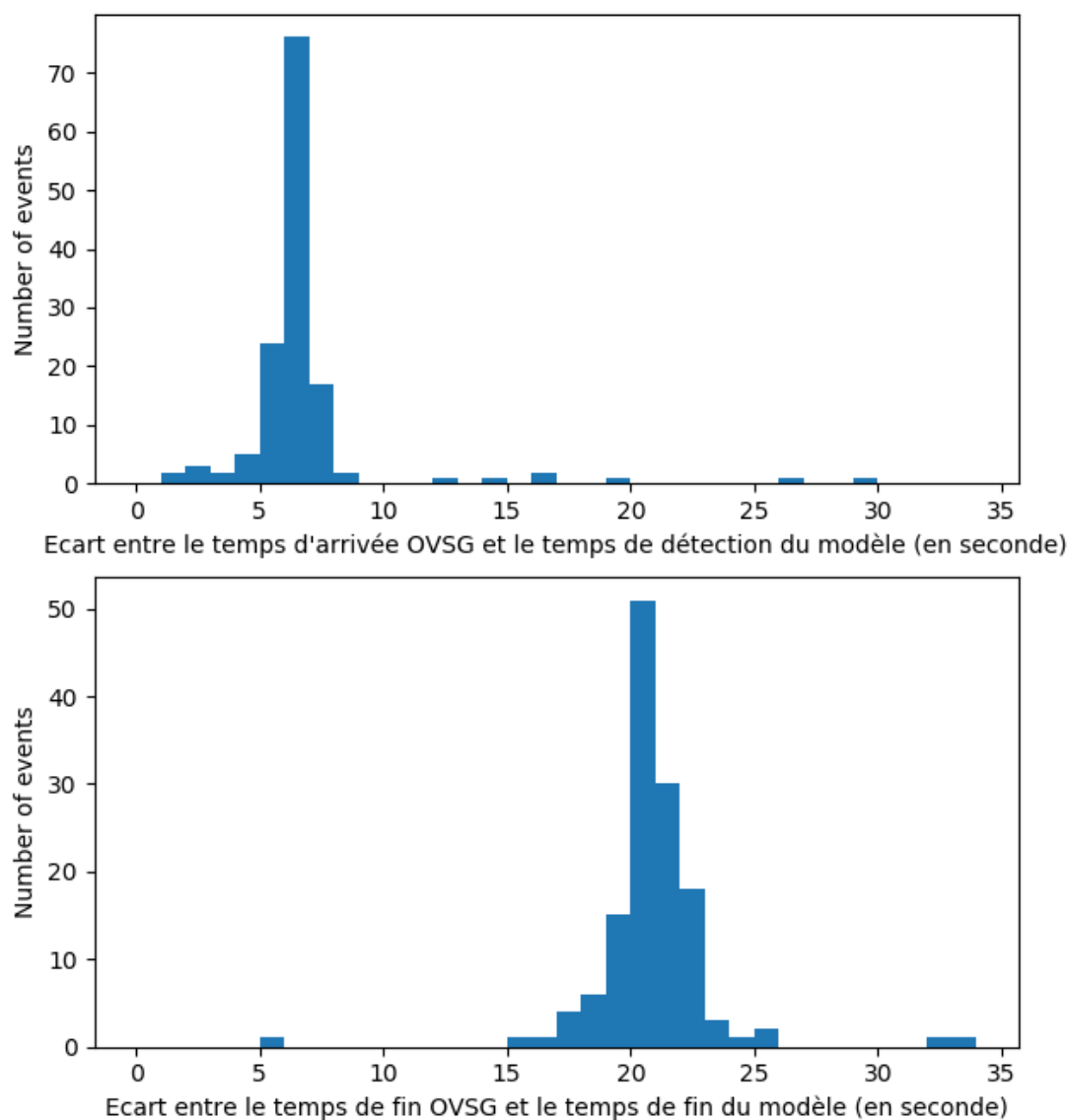


FIGURE 4.16 – En haut : Ecart temporel entre temps d'arrivée déterminé par l'OVSG et le temps de la première fenêtre déclenchée; En bas : Ecart temporel entre le temps de fin de l'événement déterminé par l'OVSG et le temps de fin de la dernière fenêtre déclenchée

5 Autres Résultats

Au cours de ces trois années de thèses, plusieurs pistes de recherche ont été suivies, toutes n'ont pas été scientifiquement fructueuses, mais cela montre néanmoins qu'il reste de nombreux sujets à explorer dans le domaine de l'intelligence artificielle appliquée à la surveillance sismo-volcanique. Dans ce chapitre, nous présenterons quelques travaux qui n'étaient pas assez étoffés pour représenter un chapitre à eux seuls, mais qui peuvent servir de piste de réflexion pour de futurs projets de recherche. Nous verrons quelques stratégies mises en place pour réaliser une classification automatique multistation puis nous présenterons quelques résultats en lien avec la classification automatique des signaux sismo-volcaniques du volcan Merapi en Indonésie et enfin, on s'intéressera à des perspectives un peu plus générales.

5.1 Classification multistation

Le réseau de surveillance sismique du volcan de la Soufrière comporte 24 stations, or dans cette thèse jusqu'ici, nous n'avons utilisé qu'une seule station pour la classification automatique des événements sismo-volcaniques. L'utilisation d'une stratégie multistation pourrait s'avérer très utile pour classer les signaux. En effet, la prédiction finale devrait être plus fiable, car elle aurait plus d'informations pour prendre une meilleure décision. De plus, en contexte volcanique, il peut arriver que des stations soient endommagées et donc que des données soient indisponibles, il est donc crucial de pouvoir continuer le processus de classification dans ces conditions.

Ici nous avons voulu étendre notre classification monostation et utiliser deux stations supplémentaires dans l'analyse. On a donc choisi d'utiliser les stations Morne Amic (AMC) et Mamelle (MML) - en plus de la station TAG que l'on utilisait précédemment - deux stations large-bande à l'abri des perturbations extérieures, se situant à moins de deux kilomètres du sommet et présentant souvent des signaux de bonnes qualités avec un bon rapport signal sur bruit.

Le jeu de données utilisé pour construire les modèles multistations est le même que précédemment, mais nous ôtons de la base de données les signaux qui ne sont pas disponibles en même temps sur les trois stations. Nous avons donc après cette phase 362 VT, 131 Hybrides, 24 Emboîtés, 14 LP et 12 Tornillos. Les performances obtenues avec chacune de ces trois stations individuellement et un algorithme RF sont disponibles Tableau 5.1.

TABLE 5.1 – Accuracy obtenue avec le jeu de données multistation pour chaque station, Score moyen obtenu après 10 réalisations

Station	Accuracy (%)
AMC	77.46 +/- 1.87
MML	75.22 +/- 1.55
TAG	79.23 +/- 1.41

Maintenant nous allons présenter les différentes stratégies mises en place pour construire les

modèles et les performances associées.

5.1.1 Modèles globaux

Modèle global simple

Ce premier modèle multistation est assez simple (Figure 5.1), il considère un même événement enregistré sur deux stations différentes comme deux événements distincts. Cette construction de modèles permet de prendre en compte la variabilité des événements sur le réseau, mais ne prend pas en compte la variabilité d'un seul et même événement qui pourrait être dû à un effet de site, un effet de propagation ou bien même une station soumise aux conditions météorologiques. Un autre avantage de ce modèle est qu'il permet de s'affranchir d'un problème de disponibilité de données sur une station.

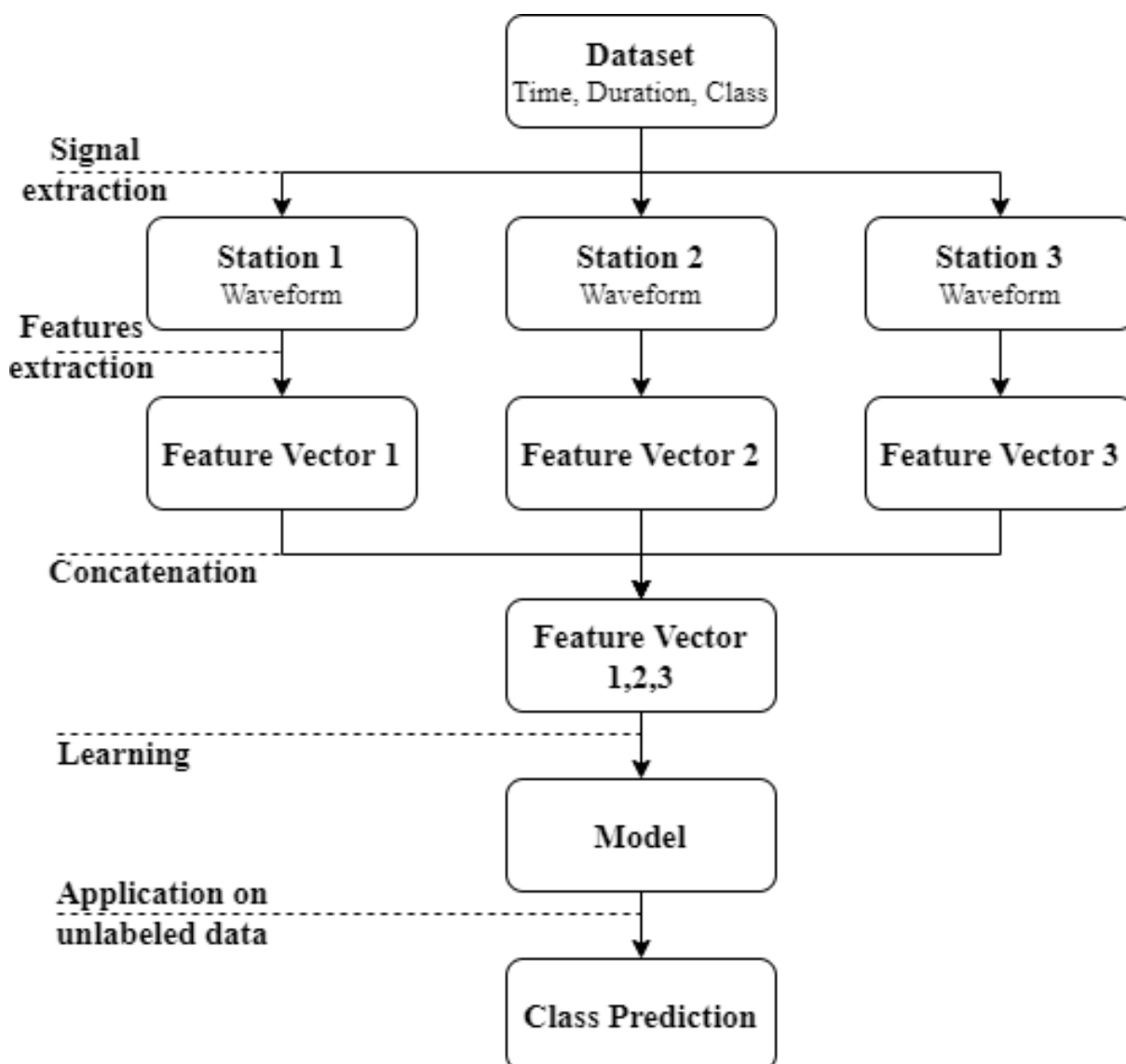


FIGURE 5.1 – Schéma représentant le modèle multistation simple où un même événement enregistré sur plusieurs stations peut être utilisé plusieurs fois dans l'apprentissage

Modèle global combiné

Ce deuxième modèle est, lui aussi, plutôt simple (Figure 5.2), il consiste à utiliser les signaux de toutes les stations pour l'apprentissage. On calcule les descripteurs des signaux pour chaque station et puis on les associe pour former un vecteur descripteur autant de fois plus grand qu'il y a de stations. Cela ressemble beaucoup à ce qu'a proposé Maggi et al. (2017) à la différence que dans leur étude, un modèle était préalablement construit pour chaque station afin de déterminer les descripteurs les plus discriminants, pour ensuite ne garder que les 10 meilleurs descripteurs de chaque station pour l'apprentissage. Notre méthode par rapport à celle de Maggi et al. (2017), utilise beaucoup de descripteurs et prend en compte la variabilité d'un même événement selon la station. Néanmoins, cette méthode demande d'avoir tout le temps les mêmes stations opérationnelles, car si on manque de données pour une seule station alors le vecteur descripteur ne peut plus être construit.

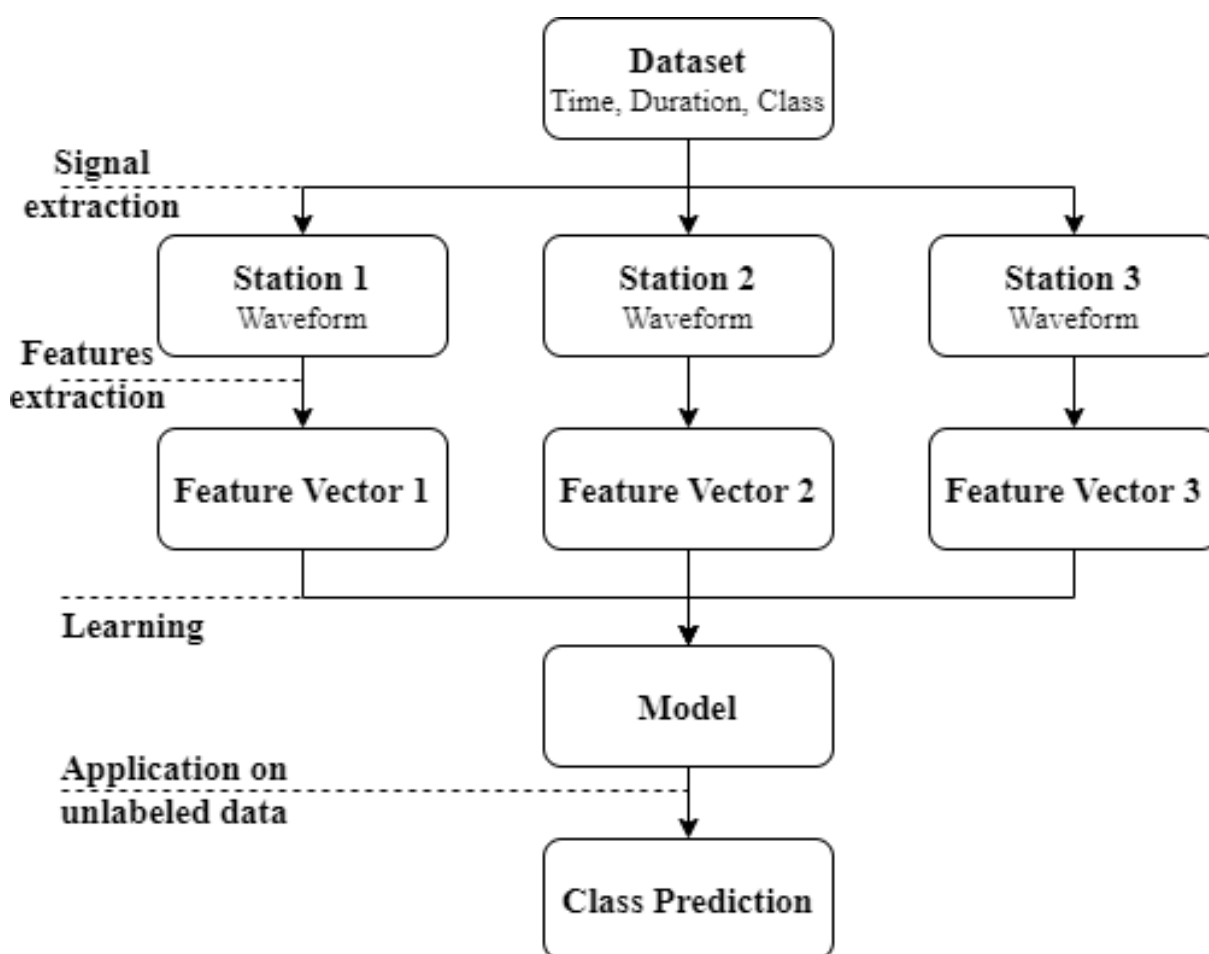


FIGURE 5.2 – Schéma représentant le modèle multistation combiné où les descripteurs associés aux événements enregistrés sur plusieurs stations sont combinés pour former un grand vecteur descripteur

5.1.2 Modèles par vote

Modèle par vote à 3 sous-modèles

La construction de ce modèle (Figure 5.3) demande de créer un sous-modèle pour chaque station et à chaque nouvelle donnée, on veut que l'algorithme nous donne la probabilité d'appartenance à chacune des classes. On moyenne ces probabilités sur toutes les stations et la probabilité

maximale sera notre prédiction. Ce modèle permet de s'affranchir du problème de disponibilité des stations et permet même d'utiliser autant de stations que l'on veut. Un inconvénient de cette construction peut survenir lorsque l'événement n'est visible que sur une seule station, les résultats des mauvaises stations pourraient fausser les résultats généraux.

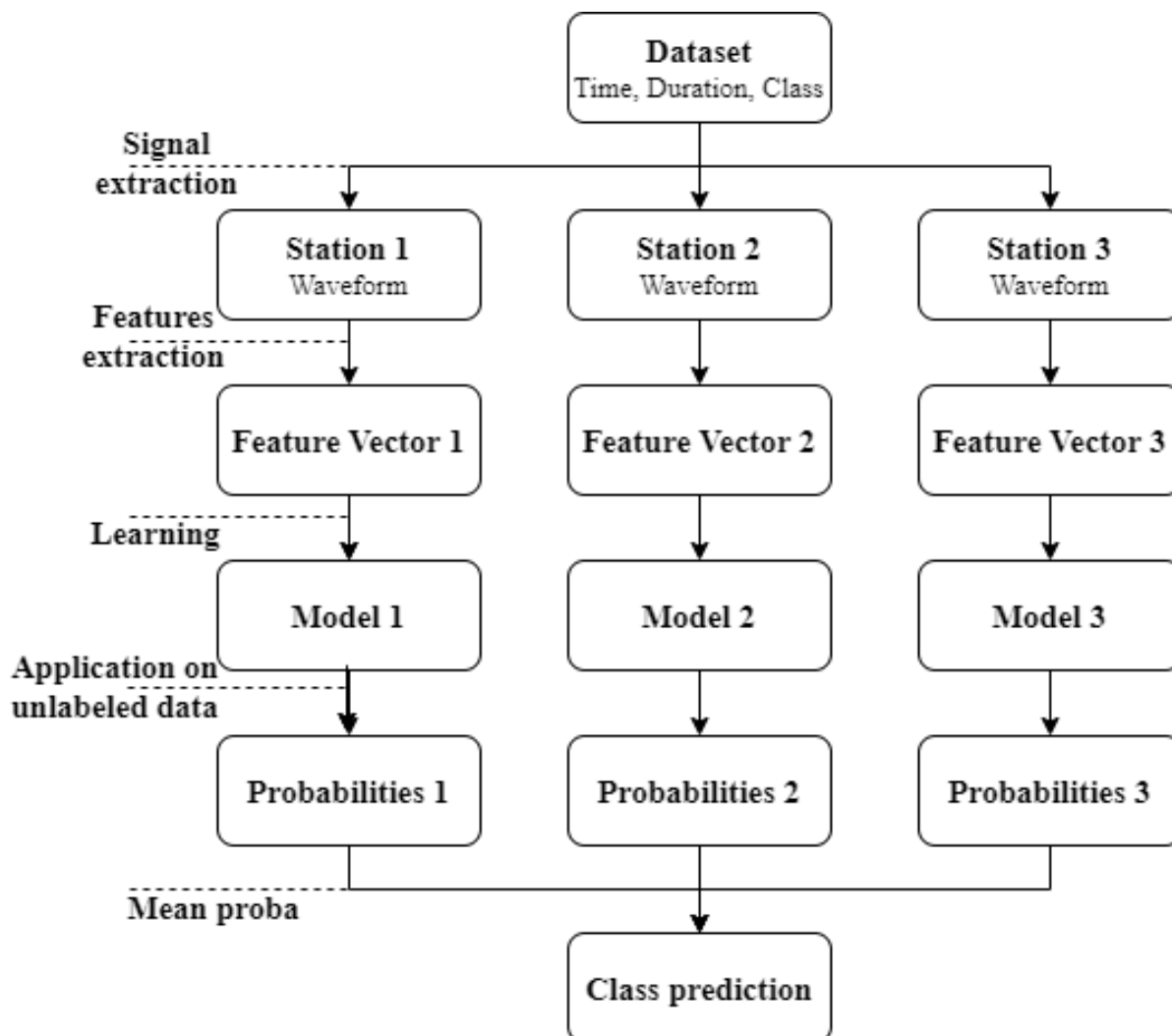


FIGURE 5.3 – Schéma représentant le modèle multistation par vote avec 3 sous-modèles, un par station

Modèle par vote et 1 sous-modèle

Cette construction (Figure 5.4) mixe le modèle par vote à 3 sous-modèles et le modèle global combiné. Au lieu de créer un modèle par station, on crée un seul modèle en combinant les descripteurs de chaque station. A contrario de la construction avec 3 sous-modèles, celui-ci a besoin que toutes les stations soient disponibles en même temps, mais a l'avantage de prendre en compte la variabilité d'un même événement.

5.1.3 Modèles encapsulés

Le dernier modèle présenté (Figure 5.5) est un peu plus complexe, il consiste à utiliser les probabilités d'appartenance à chacune des classes, données par un modèle correspondant à une seule station, comme descripteur pour le second modèle. Cette méthode a pour avantage de prendre

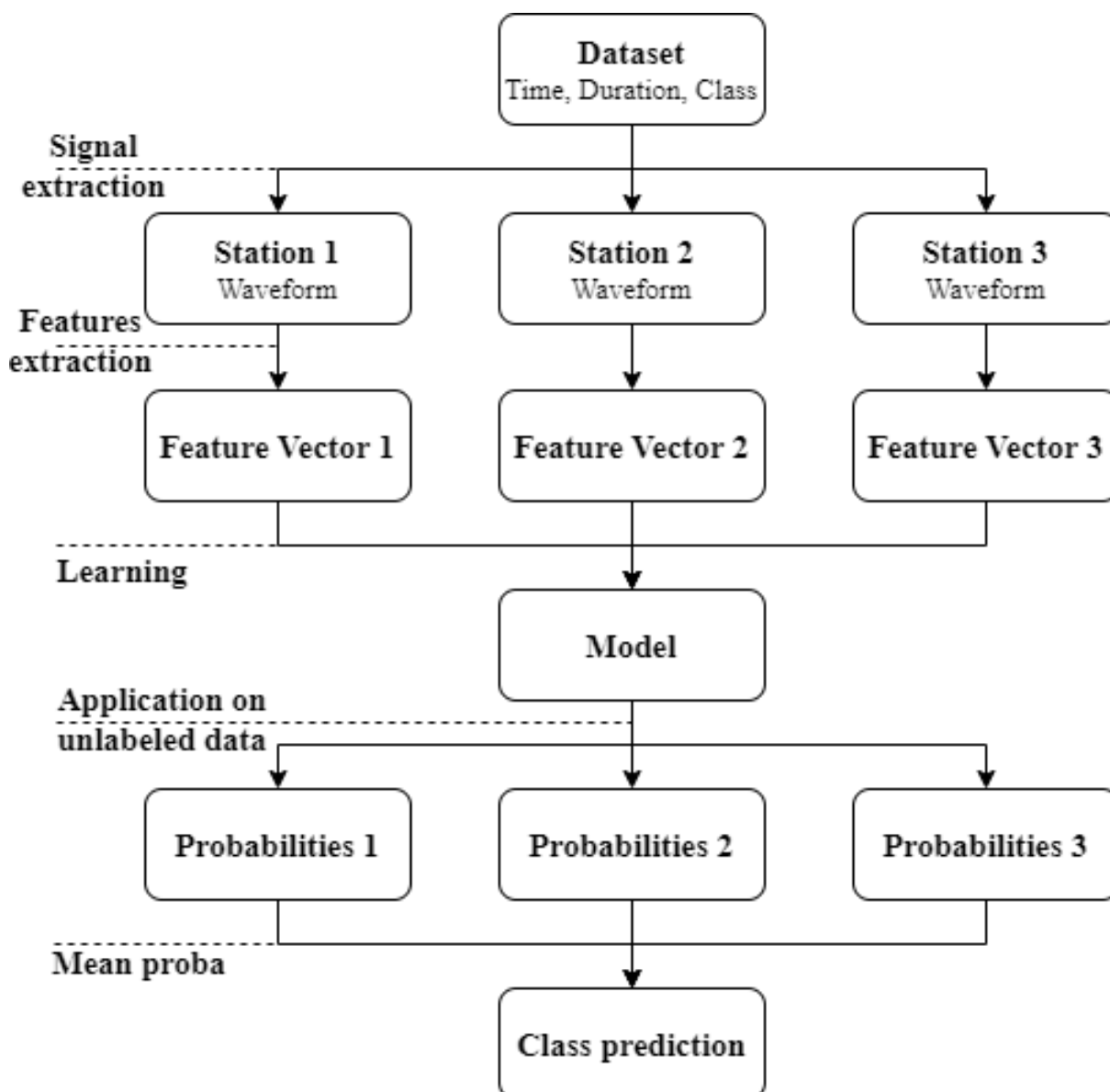


FIGURE 5.4 – Schéma représentant le modèle multistation par vote avec 1 sous-modèle

en compte la variabilité d'un même événement selon la station, mais est sensible à un problème de disponibilité de station.

5.1.4 Résultats et discussion

Les résultats des performances des différentes stratégies sont disponibles Tableau 5.2 et les matrices de confusion associées à chacun des modèles sont disponibles en annexe. En comparant les classes prédites avec la vraie des événements, on voit que l'accuracy obtenue avec les stratégies multistations (79% au mieux) n'est pas meilleure que celle obtenue avec la station TAG seule (79% également). La principale cause de ces résultats décevants est la qualité des signaux enregistrés à AMC et MML sur la période. S'ils sont très distincts sur TAG, ils sont moins forts en amplitude sur les deux autres stations. Le fait que la taille du jeu de données se soit réduit à cause du problème de disponibilité n'aide pas à conclure sur l'efficacité ou l'inefficacité de ces méthodes multistations. Il faudrait étendre le nombre d'événements dans notre base de données

TABLE 5.2 – Accuracy obtenue avec le jeu de données multistation pour chaque stratégie, Score moyen obtenu après 10 réalisations

Stratégie	Accuracy (%)
Modèle global simple	79.5 +/- 0.9
Modèle global combiné	79.3 +/- 1.8
Modèle par vote à 3 sous-modèles	69.7 +/- 2.9
Modèle par vote à 1 sous-modèle	76.4 +/- 1.9
Modèle encapsulé	50.2 +/- 2.9

et avoir des classes moins déséquilibrées en termes de nombre pour bien analyser les résultats. De plus, la Soufrière est un volcan difficile pour tester ces méthodes, car le rapport signal sur bruit est généralement faible sur ce volcan. Il faudrait tester ces méthodes sur un volcan plus actif tel le Merapi avant de revenir sur la Soufrière appliquer ces méthodes.

5.2 Application au volcan Merapi

Le volcan Merapi est un volcan situé en Indonésie sur l'île de Java et l'un des plus actifs du pays. Il est observé en continu par l'observatoire local, le Balai Penyelidikan dan Pengembangan Teknologi Kebencanaan Geologi (BPPTKG). Ce volcan entre en éruption assez souvent (tous les 5 ans environ) et peut entraîner des croissances et des explosions de dômes causant des coulées pyroclastiques. La région autour du volcan est très peuplée avec notamment la ville de Yogyakarta et ses 400 000 habitants à 25 km de l'édifice. La plus forte éruption récente s'est produite en 2010 et a causé la mort de 353 personnes. Il est donc essentiel de surveiller son activité pour atténuer les risques sur les populations locales.

Le volcan est surveillé depuis 1924 par le BPPTKG et son activité est plutôt bien connue des experts. Les différentes classes de séismes liés à l'activité volcaniques sont bien définies (Budi-Santoso et al., 2013) :

- Séismes volcano-tectoniques : Associés à une fracture en profondeur, ils sont ici distingués en 2 types, profonds (VTA) et superficiels (VTB). Ils sont très énergétiques, très fréquents et durent entre 5 et 20 secondes. 326 VA et 415 VTB entre 2018 et 2020. Ils sont souvent identifiés comme précurseurs d'une éruption.
- Séismes longues périodes (ou basse-fréquence) : Ils sont associés à des mouvements de fluide, sont marqueurs de l'état du volcan et sont de faible amplitude. 1439 événements entre 2018 et 2020.
- Séismes Multi-Phase (ou Hybride) : Événements causés par la friction du magma sur les parois du conduit. Ils sont composés de plusieurs phases difficilement identifiables et durent environ 5 secondes. Les VT de faible amplitude peuvent être considérés Multi-Phase. 2478 événements entre 2018 et 2020.
- Eboulement (ou Guguran) : Ce sont des signaux émergents et énergétiques durant de quelques secondes à plusieurs minutes selon le parcours suivi par les roches éboulées. 12 995 événements entre 2018 et 2020.
- Nuées ardentes (ou Awanpanas) : D'une durée de plusieurs minutes, avec des formes d'ondes

TABLE 5.3 – Matrice de confusion obtenue avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue du BPPTKG entre le 1er Janvier 2018 et le 31 Décembre 2019, avec un modèle entraîné en utilisant un classifieur de type Random Forest (`n_estimators=100`, `criterion=’entropy’`, `bootstrap=True`, `class_weight=None`), Score moyen obtenu après 10 réalisations

True class	Predicted Class						Recall
2-7	AwanPanas	LP	MP	RF	VTA	VTB	(%)
AwanPanas	27	0	0	5	0	0	84.3
LP	0	92	8	10	9	1	76.7
MP	0	5	90	3	12	10	75.0
RF	4	6	4	102	3	1	85.0
VTA	0	6	15	4	59	3	67.8
VTB	0	1	16	0	1	76	80.9
Precision (%)	87.1	83.6	67.7	82.2	70.2	83.5	
Accuracy (%)							78.1 ±0.8

différentes selon la cause. 105 événements entre 2018 et 2020.

- Tremors : Liés à des flux de gaz ou de magma constant sortant du dôme. 15 événements entre 2018 et 2020.

Lors d’une collaboration avec François Beauducel débuté par Marielle Malfante (Malfante, 2018), il avait été émis l’idée d’implémenter un système de classification automatique des signaux du Merapi dans le WebObs local. Le travail de classification étant fait manuellement jusqu’alors.

On a voulu tester la robustesse du catalogue et de la classification actuelle. Pour cela on a sélectionné aléatoirement 400 événements (on utilise tous les événements d’une classe si moins de 400 sont disponibles) de chaque classe entre 2018 et 2020, les résultats sont disponibles Tableau 5.3.

On obtient une bonne accuracy de 78% avec notamment les nuées ardentes, les éboulements et les VTB qui ont une accuracy et une precision supérieures à 80%. Les LP prédits sont plutôt fiables (precision de 83.6%) mais on en rate quelques-uns (76.7% de recall). Les deux classes avec les moins bonnes accuracy ou precision sont VTA et MP. Or les deux sont des événements qui se produisent dans la partie superficielle de la croute. De plus, on sait que les VT de faible amplitude peuvent être considérés comme MP.

Le catalogue fourni par le BPPTKG est donc plutôt fiable, les potentielles erreurs au sein du catalogue sont très difficiles à corriger car ambiguës et sujettes à l’interprétation de l’opérateur. Finalement, un modèle de classification automatique a été implémenté dans le WebObs. Celui-ci permet lorsqu’un événement est détecté de pouvoir avoir une indication sur la classe grâce à la probabilité d’appartenance aux classes utilisées lors de la construction du modèle. Une architecture similaire pourrait être implémentée dans le WebObs de l’OVSG et pourrait faciliter le travail des opérateurs sur place.

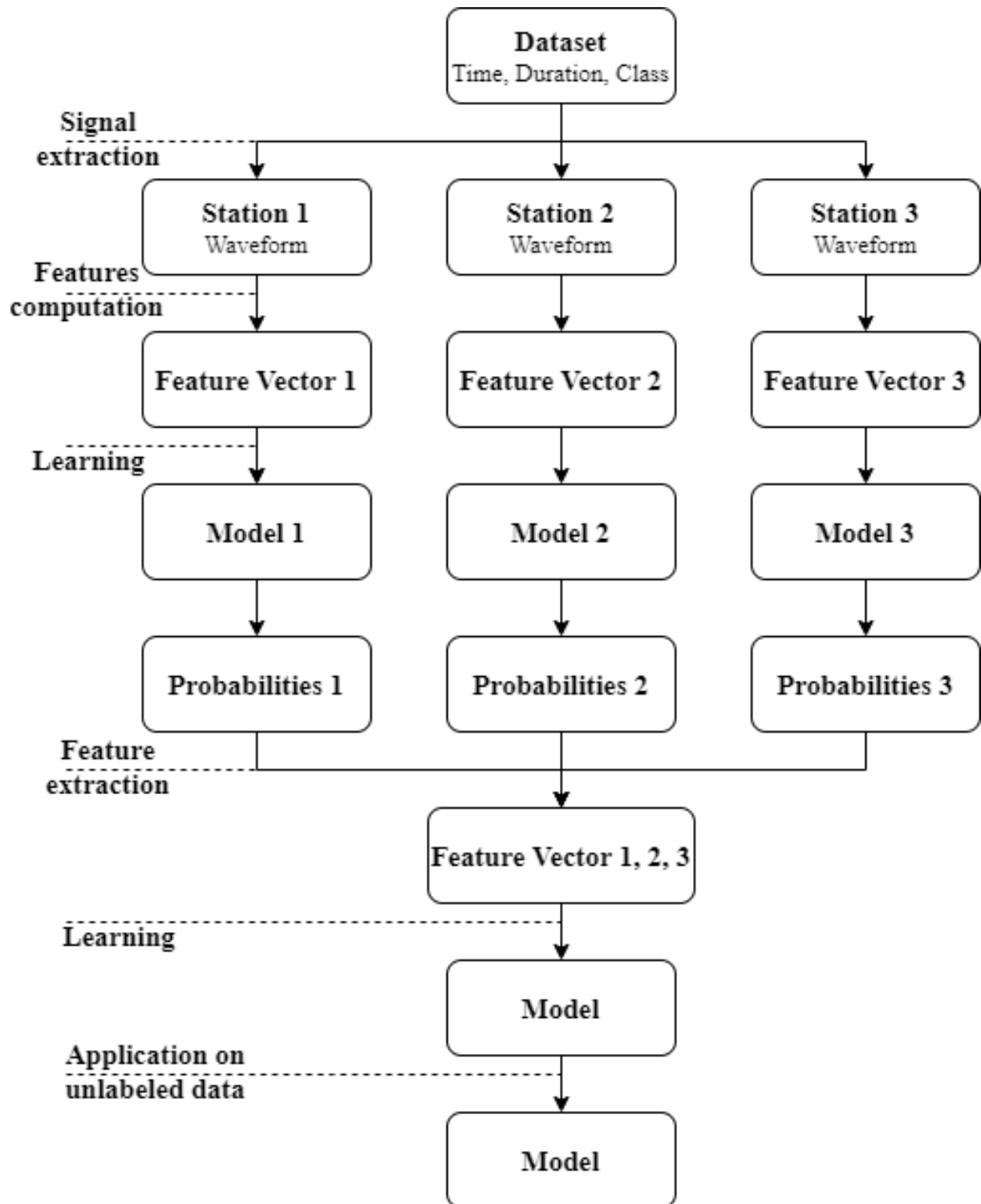


FIGURE 5.5 – Schéma représentant les modèles encapsulés

6 Conclusion et Perspectives

Cette thèse a pour objet l'application de l'apprentissage automatique pour la classification et la détection de signaux sismiques liés à l'activité volcanique, notamment, de la Soufrière de Guadeloupe.

Dans le chapitre 2, nous avons explicité la méthode basée sur l'apprentissage supervisé. Les formes d'ondes sismiques sont transformées en un large jeu de descripteurs (34 descripteurs pour chaque domaine de représentation) calculés à partir de trois domaines de représentation du signal (domaines temporel, fréquentiel et quéfrentiel). Les vecteurs qui résultent de cette transformation représentent au mieux les signaux et la variabilité des signaux et sont utilisés pour la modélisation. Pour l'apprentissage, nous avons choisi d'utiliser le Random Forest Classifier de la bibliothèque scikit-learn.

On a commencé par entraîner un modèle à partir d'un catalogue construit par l'OVSG comprenant 845 événements réparti en trois classes, VT, LP et Emboîtés et enregistrés entre 2013 et 2018. On a obtenu un taux de bonne classification de 72%. On en a conclu que la classe VT contenait une variété de signaux couvrant les LP et les Emboîtés. Après avoir revu et reclassé manuellement tous les signaux en observant leurs formes d'onde, spectrogrammes et spectres de Fourier, on a introduit deux nouvelles classes, Hybride et Tornillo. En utilisant ces nouvelles informations, un nouveau modèle a été entraîné et on a atteint un bien meilleur score de 84%. La plupart des erreurs étant dues aux événements Hybride qui sont confondus avec les VT et les LP. Mais cela peut être expliqué par la nature même de cette classe qui est un continuum entre LP et VT avec des processus physiques qui comportent de la fracturation et de la résonnance. Ensuite, on a montré quels étaient les meilleurs descripteurs pour chacune des classes. On a vu que parmi les meilleurs descripteurs, il y en avait dans chaque domaine de représentation. Les 14 meilleurs descripteurs sont suffisants pour obtenir des performances sensiblement identiques à celle avec tous les descripteurs. Par la suite, on a vu que notre modèle était applicable à une autre période, car le type de sismicité de la Soufrière n'avait pas beaucoup changé.

Dans le chapitre 4, nous nous sommes intéressés à la détection et la classification d'événements sur le signal continu. On y a vu qu'en utilisant un modèle signal contre bruit la détection pouvait être particulièrement puissante, mais que classer les événements ainsi détectés est très difficile, car on n'arrive pas à borner le début et la fin des événements. On a essayé de classer le signal continu, mais les résultats étaient trop dépendant de la taille de la fenêtre glissante utilisée, plus la fenêtre est petite moins on reconnaît les événements de grande durée et vice-versa.

On a donc vu à travers l'aspect multistation qu'il est possible d'aller plus loin dans la classification automatique des signaux. On a également vu que rendre opérationnel en observatoire ce type d'algorithme est tout à fait possible. Mais beaucoup d'autres aspects peuvent encore être explorés.

On a vu dans le chapitre sur la détection qu'il y avait encore des améliorations possibles. On pourrait mettre en place des systèmes de seuil plus complexes pour détecter le début et la fin des événements. Une autre limitation pour pouvoir détecter les débuts et fins de signaux est

la taille fixe de la fenêtre glissante. En effet, nos signaux sont de tailles variables (de quelques secondes à plusieurs dizaines de secondes) et notre fenêtre glissante est de taille fixe donc on aura des difficultés à détecter des événements avec des longueurs très différentes de notre fenêtre. Les chaînes de Markov pourraient s'avérer utiles pour ce genre de problème, car elles permettent de définir l'état du signal en tout point.

On a vu qu'on était capable d'améliorer une classification existante, mais cette amélioration s'est faite en grande partie à la main en visualisant les formes d'ondes et les spectrogrammes. Cette visualisation manuelle n'est pas fiable à 100% et peut également changer selon l'opérateur. Il serait intéressant de voir comment un algorithme non supervisé classerait les signaux, est-ce qu'il respecterait nos classes ou alors en construirait-il d'autres ? On pourrait également utiliser des algorithmes semi-supervisés dans le cas où une classe se modifierait avec le temps.

Il y a également des choses à faire avec les descripteurs utilisés, certains sont redondants comme la longueur dans les trois domaines de représentations. Certains descripteurs sont importés du traitement de la parole et sont peut-être moins pertinents en sismologie. L'utilisation de descripteurs liés aux spectrogrammes (Hibert et al., 2017) pourrait s'avérer utile.

La comparaison entre les meilleurs descripteurs sur deux volcans différents, mais de sismicité similaire pourrait renforcer le lien entre la physique du volcan et les descripteurs utilisés pour représenter les signaux. Ou alors quels sont les meilleurs descripteurs associés à différents types de volcanisme, par exemple, est-ce que les meilleurs descripteurs du Merapi sont identiques à ceux de l'Ubinas.

Les travaux présentés dans cette thèse ont pu mettre en évidence que l'apprentissage supervisé peut-être utile dans le cadre d'une surveillance volcanique opérationnelle et plus précisément pour l'étude de la sismicité. Cet outil permet d'automatiser le travail de classification et de détection des signaux sismiques liés à l'activité volcanique. Sur l'aspect classification, l'apprentissage supervisé permet d'automatiser un travail fastidieux notamment en période de crise durant laquelle le nombre d'événements à classer peut être très grand, jusqu'à plusieurs centaines d'événements par jour à La Soufrière par exemple, et requerrait plusieurs opérateurs pour être réalisé. De plus, comme l'apprentissage supervisé se base sur des critères objectifs pour classer les événements, la classification est plus fiable que lorsqu'elle est faite par plusieurs opérateurs différents pour lesquels les critères peuvent légèrement varier d'un œil à l'autre. Cependant, pour que cette fiabilité puisse être établie, elle requiert des tests du modèle ainsi que la construction d'une base de données d'apprentissage correcte dans laquelle les classes sont bien séparées et à partir de laquelle un algorithme classifieur va pouvoir apprendre à classer les nouveaux événements inconnus. Cette étape peut demander de faire des aller-retour entre des tests du modèle et la construction de la base d'apprentissage dans les cas où des erreurs peuvent s'être glissées dans le catalogue ou lorsque deux classes sont très proches l'une de l'autre selon le modèle. Pour l'aspect détection, l'apprentissage supervisé permet de bien différencier le bruit de fond d'un signal d'intérêt et détecte plus d'événements qu'un STA/LTA classique, mais il manque encore de la précision sur le temps d'arrivée de l'événement pour pouvoir ensuite directement faire un travail de localisation à partir des temps d'arrivée déterminés automatiquement.

Références

- ALASONATI, P., WASSERMANN, J. & OHRNBERGER, M. (2006). Signal classification by wavelet-based hidden Markov models : application to seismic signals of volcanic origin. *Statistics in Volcanology*. Geological Society of London. <https://doi.org/10.1144/IAVCEI001.13>
- ALAVI, A. H. & GANDOMI, A. H. (2011). Prediction of principal ground-motion parameters using a hybrid method coupling artificial neural networks and simulated annealing. *Computers & Structures*, 89(23-24), 2176-2194.
- ALLEN, R. (1982). Automatic phase pickers : Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B), S225-S242.
- ALLEN, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the seismological society of America*, 68(5), 1521-1532.
- APOLLONI, B. et al. (2009). Support vector machines and MLP for automatic classification of seismic signals at Stromboli volcano. *Proc. 19th Italian Workshop Neural Nets*, 204, 116.
- BAILLARD, C., CRAWFORD, W. C., BALLU, V., HIBERT, C. & MANGENEY, A. (2014). An automatic kurtosis-based P-and S-phase picker designed for local seismic networks. *Bulletin of the Seismological Society of America*, 104(1), 394-409.
- BEAN, C., LOKMER, I. & O'BRIEN, G. (2008). Influence of near-surface volcanic structure on long-period seismic signals and on moment tensor inversions : Simulated examples from Mount Etna. *Journal of Geophysical Research : Solid Earth*, 113(B8).
- BEAN, C. J., DE BARROS, L., LOKMER, I., MÉTAXIAN, J.-P., O'BRIEN, G. & MURPHY, S. (2014). Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes. *Nature geoscience*, 7(1), 71-75.
- BEAUDUCEL, F., LAFON, D., BÉGUIN, X., SAUREL, J.-M., BOSSON, A., MALLARINO, D., BOISSIER, P., BRUNET, C., LEMARCHAND, A., ANTÉNOR-HABAZAC, C. et al. (2020). WebObs : The volcano observatories missing link between research and real-time monitoring. *Frontiers in Earth Science*, 8, 48.
- BELL, A. F., HERNANDEZ, S., MCCLOSKEY, J., RUIZ, M., LAFEMINA, P. C., BEAN, C. J. & MÖLLHOFF, M. (2021). Dynamic earthquake triggering response tracks evolving unrest at Sierra Negra volcano, Galápagos Islands. *Science advances*, 7(39), eabh0894.
- BENITEZ, M. C., RAMIREZ, J., SEGURA, J. C., IBANEZ, J. M., ALMENDROS, J., GARCIA-YEGUAS, A. & CORTES, G. (2006). Continuous HMM-based seismic-event classification at Deception Island, Antarctica. *IEEE Transactions on Geoscience and remote sensing*, 45(1), 138-146.
- BEYREUTHER, M., HAMMER, C., WASSERMANN, J., OHRNBERGER, M. & MEGIES, T. (2012). Constructing a Hidden Markov Model based earthquake detector : application to induced seismicity. *Geophysical Journal International*, 189(1), 602-610.
- BOUDON, G., DAGAIN, J., SEMET, M. & WESTERCAMP, D. (1987). Carte géologique 1 : 20 000 du massif volcanique de la Soufrière (Département de la Guadeloupe, Petites Antilles). *BRGM, Paris*.
- BOUDON, G., KOMOROWSKI, J.-C., VILLEMANT, B. & SEMET, M. P. (2008). A new scenario for the last magmatic eruption of La Soufrière of Guadeloupe (Lesser Antilles) in 1530

- AD Evidence from stratigraphy radiocarbon dating and magmatic evolution of erupted products. *Journal of Volcanology and Geothermal Research*, 178(3), 474-490.
- BREIMAN, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.
- BUDI-SANTOSO, A., LESAGE, P., DWIYONO, S., SUMARTI, S., JOUSSET, P., METAXIAN, J.-P. et al. (2013). Analysis of the seismic activity associated with the 2010 eruption of Merapi Volcano, Java. *Journal of Volcanology and Geothermal Research*, 261, 153-170.
- BUENO, A., BENITEZ, C., DE ANGELIS, S., MORENO, A. D. & IBÁÑEZ, J. M. (2019). Volcano-seismic transfer learning and uncertainty quantification with Bayesian neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 58(2), 892-902.
- CAPE, T., COLES, P., CONKIE, A., GOLIPOUR, L., HADJITARKHANI, A., HU, Q., HUDDLESTON, N., HUNT, M., LI, J., NEERACHER, M. et al. (2017). Siri on-device deep learning-guided unit selection text-to-speech system. *Interspeech*, 4011-4015.
- CHOUET, B. (1988). Resonance of a fluid-driven crack : Radiation properties and implications for the source of long-period events and harmonic tremor. *Journal of Geophysical Research : Solid Earth*, 93(B5), 4375-4400.
- CHOUET, B. A. & MATOZA, R. S. (2013). A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *Journal of Volcanology and Geothermal Research*, 252, 108-175.
- CIESLAK, D. A. & CHAWLA, N. V. (2008). Learning decision trees for unbalanced data. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 241-256.
- CORTÉS, G., ARÁMBULA, R., GUTIÉRREZ, L., BENÍTEZ, C., IBÁÑEZ, J., LESAGE, P., ALVAREZ, I. & GARCIA, L. (2009). Evaluating robustness of a HMM-based classification system of volcano-seismic events at colima and popocatepetl volcanoes. *2009 IEEE International Geoscience and Remote Sensing Symposium*, 2, II-1012-II-1015. <https://doi.org/10.1109/IGARSS.2009.5418275>
- CURILEM, G., VERGARA, J., FUENTEALBA, G., ACUÑA, G. & CHACÓN, M. (2009). Classification of seismic signals at Villarrica volcano (Chile) using neural networks and genetic algorithms. *Journal of Volcanology and Geothermal Research*, 180(1), 1-8. <https://doi.org/https://doi.org/10.1016/j.jvolgeores.2008.12.002>
- CURILEM, M., de MELLO, R. F., HUENUPAN, F., SAN MARTIN, C., FRANCO, L., HERNÁNDEZ, E. & RIOS, R. A. (2018). Discriminating seismic events of the Llaima volcano (Chile) based on spectrogram cross-correlations. *Journal of Volcanology and Geothermal Research*, 367, 63-78.
- DAI, H. & MACBETH, C. (1995). Automatic picking of seismic arrivals in local earthquake data using an artificial neural network. *Geophysical journal international*, 120(3), 758-774.
- DEVRIES, P. M., THOMPSON, T. B. & MEADE, B. J. (2017). Enabling large-scale viscoelastic calculations via neural network acceleration. *Geophysical Research Letters*, 44(6), 2662-2669.
- DYSART, P. S. & PULLI, J. J. (1990). Regional seismic event classification at the NORESS array : seismological measurements and the use of trained neural networks. *Bulletin of the Seismological Society of America*, 80(6B), 1910-1933.

- ESPOSITO, A. M., GIUDICEPIETRO, F., D'AURIA, L., SCARPETTA, S., MARTINI, M. G., COLTELLI, M. & MARINARO, M. (2008). Unsupervised neural analysis of very-long-period events at Stromboli volcano using the self-organizing maps. *Bulletin of the Seismological Society of America*, 98(5), 2449-2459.
- FALSAPERLA, S., GRAZIANI, S., NUNNARI, G. & SPAMPINATO, S. (1996). Automatic classification of volcanic earthquakes by using Multi-Layered neural networks. *Natural Hazards*, 13, 205-228. <https://doi.org/10.1007/BF00215816>
- FAZIO, M., ALPARONE, S., BENSON, P. M., CANNATA, A. & VINCIGUERRA, S. (2019). Genesis and mechanisms controlling tornillo seismo-volcanic events in volcanic areas. *Scientific reports*, 9(1), 1-11.
- GATYS, L. A., ECKER, A. S. & BETHGE, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv :1508.06576*.
- GIBBONS, S. J. & RINGDAL, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation. *Geophysical Journal International*, 165(1), 149-166.
- GÓMEZ M, D. M. & TORRES C, R. A. (1997). Unusual low-frequency volcanic seismic events with slowly decaying coda waves observed at Galeras and other volcanoes. *Journal of volcanology and geothermal research*, 77(1-4), 173-193.
- GUTIÉRREZ, L., IBAÑEZ, J., CORTÉS, G., RAMÍREZ, J., BENÍTEZ, C., TENORIO, V. & ISAAC, Á. (2009). Volcano-seismic signal detection and classification processing using hidden Markov models. Application to San Cristóbal volcano, Nicaragua. *2009 IEEE International Geoscience and Remote Sensing Symposium*, 4, IV-522-IV-525. <https://doi.org/10.1109/IGARSS.2009.5417428>
- HAMMER, C., OHRNBERGER, M. & FÄH, D. (2013). Classifying seismic waveforms from scratch : a case study in the alpine environment. *Geophysical Journal International*, 192(1), 425-439.
- HARRINGTON, R. M. & BRODSKY, E. E. (2007). Volcanic hybrid earthquakes that are brittle-failure events. *Geophysical Research Letters*, 34(6).
- HIBERT, C., MANGENEY, A., GRANDJEAN, G., BAILLARD, C., RIVET, D., SHAPIRO, N. M., SATRIANO, C., MAGGI, A., BOISSIER, P., FERRAZZINI, V. et al. (2014). Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. *Journal of Geophysical Research : Earth Surface*, 119(5), 1082-1105.
- HIBERT, C., PROVOST, F., MALET, J.-P., MAGGI, A., STUMPF, A. & FERRAZZINI, V. (2017). Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm. *Journal of Volcanology and Geothermal Research*, 340, 130-142.
- IBAÑEZ, J. M., BENITEZ, C., GUTIÉRREZ, L. A., CORTÉS, G., GARCIA-YEGUAS, A. & ALGUACIL, G. (2009). The classification of seismo-volcanic signals using Hidden Markov Models as applied to the Stromboli and Etna volcanoes. *Journal of Volcanology and Geothermal Research*, 187(3-4), 218-226.
- IBS-VON SEHT, M. (2008). Detection and identification of seismic signals recorded at Krakatau volcano (Indonesia) using artificial neural networks. *Journal of Volcanology and Geothermal Research*, 176(4), 448-456. <https://doi.org/https://doi.org/10.1016/j.jvolgeores.2008.04.015>

- JESSOP, D., MORETTI, R., MOUNE, S., BONIFACIE, M., BURTIN, A., DE CHABALIER, J.-B., DEROUSSE, S., KOMOROWSKI, J.-C., ROSAS-CARBAJAL, M., BOSSON, A. et al. (2019). Les risques volcaniques au XXI^e siècle : surveiller et prévoir ; l'exemple de l'OVSG. *Géologues*.
- KARPATHY, A. & FEI-FEI, L. (2015). Deep visual-semantic alignments for generating image descriptions. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3128-3137.
- KÖHLER, A., OHRNBERGER, M. & SCHERBAUM, F. (2010). Unsupervised pattern recognition in continuous seismic wavefield records using self-organizing maps. *Geophysical Journal International*, 182(3), 1619-1630.
- KOMOROWSKI, J., BOUDON, G., SEMET, M., BEAUDUCEL, F., ANTÉNOR-HABAZAC, C., BAZIN, S., HAMMOUYA, G., LINDSAY, J., ROBERTSON, R., SHEPHERD, J. et al. (2005). Volcanic hazard atlas of the Lesser Antilles : Guadeloupe. *Volcanic Hazard Atlas of the Lesser Antilles*, 65-102.
- KOMOROWSKI, J.-C., MORIN, J., JENKINS, S. & KELMAN, I. (2016). Challenges of volcanic crises on small islands states. *Observing the volcano world* (p. 353-371). Springer.
- KONG, Q., ALLEN, R. M. & SCHREIER, L. (2016). MyShake : Initial observations from a global smartphone seismic network. *Geophysical Research Letters*, 43(18), 9588-9594.
- KONG, Q., TRUGMAN, D. T., ROSS, Z. E., BIANCO, M. J., MEADE, B. J. & GERSTOFT, P. (2019). Machine learning in seismology : Turning data into insights. *Seismological Research Letters*, 90(1), 3-14.
- KONSTANTINO, K. I. (2015). Tornillos modeled as self-oscillations of fluid filling a cavity : Application to the 1992–1993 activity at Galeras volcano, Colombia. *Physics of the Earth and Planetary Interiors*, 238, 23-33.
- LAHR, J. C., CHOUET, B. A., STEPHENS, C. D., POWER, J. A. & PAGE, R. A. (1994). Earthquake classification, location, and error analysis in a volcanic environment : Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, 62(1-4), 137-151.
- LANGER, H., FALSAPERLA, S., MASOTTI, M., CAMPANINI, R., SPAMPINATO, S. & MESSINA, A. (2009). Synopsis of supervised and unsupervised pattern classification techniques applied to volcanic tremor data at Mt Etna, Italy. *Geophysical Journal International*, 178(2), 1132-1144. <https://doi.org/10.1111/j.1365-246X.2009.04179.x>
- LANGER, H., FALSAPERLA, S., POWELL, T. & THOMPSON, G. (2006). Automatic classification and a-posteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat. *Journal of volcanology and geothermal research*, 153(1-2), 1-10.
- LANGET, N., MAGGI, A., MICHELINI, A. & BRENGUIER, F. (2014). Continuous kurtosis-based migration for seismic event detection and location, with application to Piton de la Fournaise Volcano, La Reunion. *Bulletin of the Seismological Society of America*, 104(1), 229-246.
- LAPINS, S., GOITOM, B., KENDALL, J.-M., WERNER, M. J., CASHMAN, K. V. & HAMMOND, J. O. (2021). A little data goes a long way : Automating seismic phase arrival picking at Nabro volcano with transfer learning. *Journal of Geophysical Research : Solid Earth*, 126(7), e2021JB021910.

- MADUREIRA, G. & RUANO, A. E. (2009). A neural network seismic detector. *IFAC Proceedings Volumes*, 42(19), 304-309.
- MAEDA, Y., YAMANAKA, Y., ITO, T. & HORIKAWA, S. (2021). Machine-learning-based detection of volcano seismicity using the spatial pattern of amplitudes. *Geophysical Journal International*, 225(1), 416-444.
- MAGGI, A., FERRAZZINI, V., HIBERT, C., BEAUDUCEL, F., BOISSIER, P. & AMEMOUTOU, A. (2017). Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano. *Seismological Research Letters*, 88(3), 878-891.
- MALFANTE, M., DALLA MURA, M., MARS, J. I., MÉTAXIAN, J.-P., MACEDO, O. & INZA, A. (2018). Automatic classification of volcano seismic signatures. *Journal of Geophysical Research : Solid Earth*, 123(12), 10-645.
- MALFANTE, M., DALLA MURA, M., MÉTAXIAN, J.-P., MARS, J. I., MACEDO, O. & INZA, A. (2018). Machine learning for volcano-seismic signals : Challenges and perspectives. *IEEE Signal Processing Magazine*, 35(2), 20-30.
- MASOTTI, M., FALSAPERLA, S., LANGER, H., SPAMPINATO, S. & CAMPANINI, R. (2006). Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy. *Geophysical research letters*, 33(20).
- MCNUTT, S. R. (2005). Volcanic seismology. *Annu. Rev. Earth Planet. Sci.*, 32, 461-491.
- MCNUTT, S. R. & ROMAN, D. C. (2015). Volcanic seismicity. *The encyclopedia of volcanoes* (p. 1011-1034). Elsevier.
- MENZE, B. H., KELM, B. M., MASUCH, R., HIMMELREICH, U., BACHERT, P., PETRICH, W. & HAMPRECHT, F. A. (2009). A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC bioinformatics*, 10(1), 1-16.
- MILLER, A., STEWART, R., WHITE, R., LUCKETT, R., BAPTIE, B., ASPINALL, W., LATCHMAN, J., LYNCH, L. & VOIGHT, B. (1998). Seismicity associated with dome growth and collapse at the Soufriere Hills Volcano, Montserrat. *Geophysical Research Letters*, 25(18), 3401-3404.
- MORETTI, R., KOMOROWSKI, J.-C., UCCIANI, G., MOUNE, S., JESSOP, D., de CHABALIER, J.-B., BEAUDUCEL, F., BONIFACIE, M., BURTIN, A., VALLEE, M. et al. (2020). The 2018 unrest phase at La Soufrière of Guadeloupe (French West Indies) andesitic volcano : Scrutiny of a failed but prodromal phreatic eruption. *Journal of Volcanology and Geothermal Research*, 393, 106769.
- MOUSAVI, S. M., HORTON, S. P., LANGSTON, C. A. & SAMEI, B. (2016). Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression. *Geophysical Journal International*, 207(1), 29-46.
- MOUSAVI, S. M., ZHU, W., SHENG, Y. & BEROZA, G. C. (2019). CRED : A deep residual network of convolutional and recurrent units for earthquake signal detection. *Scientific reports*, 9(1), 1-14.
- MOYA, A. & IRIKURA, K. (2010). Inversion of a velocity model using artificial neural networks. *Computers & geosciences*, 36(12), 1474-1483.

- NEUBERG, J., LUCKETT, R., BAPTIE, B. & OLSEN, K. (2000). Models of tremor and low-frequency earthquake swarms on Montserrat. *Journal of Volcanology and Geothermal Research*, 101(1-2), 83-104.
- OHRNBERGER, M. (2001). *Continuous automatic classification of seismic signals of volcanic origin at Mt. Merapi, Java, Indonesia* (thèse de doct.). Potsdam, Univ., Diss., 2001.
- PEROL, T., GHARBI, M. & DENOLLE, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, 4(2), e1700578.
- PROVOST, F., HIBERT, C., MALET, J.-P., STUMPF, A. & DOUBRE, C. (2016). Automatic classification of endogenous seismic sources within a landslide body using random forest algorithm. *EGU General Assembly Conference Abstracts*, EPSC2016-15705.
- QUINLAN, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- RAHMANI, M. E., AMINE, A. & HAMOU, R. M. (2017). Clustering Based Sampling for Learning from Unbalanced Seismic Data Set. *International Journal of Geotechnical Earthquake Engineering (IJGEE)*, 8(2), 1-22.
- ROSS, Z. E., MEIER, M.-A., HAUSSON, E. & HEATON, T. H. (2018). Generalized Seismic Phase Detection with Deep Learning Short Note. *Bulletin of the Seismological Society of America*, 108(5A), 2894-2901.
- SCARPETTA, S., GIUDICEPIETRO, F., EZIN, E. C., PETROSINO, S., DEL PEZZO, E., MARTINI, M. & MARINARO, M. (2005). Automatic Classification of Seismic Signals at Mt. Vesuvius Volcano, Italy, Using Neural Networks. *Bulletin of the Seismological Society of America*, 95(1), 185-196. <https://doi.org/10.1785/0120030075>
- SEYDOUX, L., BALESTRIERO, R., POLI, P., HOOP, M. d., CAMPILLO, M. & BARANIUK, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. *Nature communications*, 11(1), 1-12.
- SIMONYAN, K. & ZISSERMAN, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*.
- SOSA, U., KABIRZADEH, R. & BLAES, P. (2013). Automatic colorization of grayscale images. *Department of Electrical Engineering, Stanford University*.
- TIIRA, T. (1999). Detecting teleseismic events using artificial neural networks. *Computers & Geosciences*, 25(8), 929-938.
- TITOS, M., BUENO, A., GARCÍA, L. & BENÍTEZ, C. (2018). A Deep Neural Networks Approach to Automatic Recognition Systems for Volcano-Seismic Events. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 11(5), 1533-1544. <https://doi.org/10.1109/JSTARS.2018.2803198>
- UCCIANI, G. (2015). *Caractérisation spatiale et temporelle de la sismicité volcanique de la Soufrière de Guadeloupe : relation avec le système hydrothermal* (thèse de doct.) [2015USPCC198]. <http://www.theses.fr/2015USPCC198>
- UNGLERT, K., RADIĆ, V. & JELLINEK, A. M. (2016). Principal component analysis vs. self-organizing maps combined with hierarchical clustering for pattern recognition in volcano seismic spectra. *Journal of Volcanology and Geothermal Research*, 320, 58-74.
- VILLEMANT, B., KOMOROWSKI, J., DESSERT, C., MICHEL, A., CRISPI, O., HAMMOUYA, G., BEAUDUCÉL, F. & DE CHABALIER, J. (2014). Evidence for a new shallow magma intrusion at La Soufrière of Guadeloupe (Lesser Antilles) : Insights from long-term geochemical

- monitoring of halogen-rich hydrothermal fluids. *Journal of Volcanology and Geothermal Research*, 285, 247-277. <https://doi.org/https://doi.org/10.1016/j.jvolgeores.2014.08.002>
- WAITE, G. P., CHOUET, B. A. & DAWSON, P. B. (2008). Eruption dynamics at Mount St. Helens imaged from broadband seismic waveforms : Interaction of the shallow magmatic and hydrothermal systems. *Journal of Geophysical Research : Solid Earth*, 113(B2). <https://doi.org/https://doi.org/10.1029/2007JB005259>
- WANG, A. et al. (2003). An industrial strength audio search algorithm. *Ismir*, 2003, 7-13.
- WANG, J. & TENG, T.-L. (1995). Artificial neural network-based seismic detector. *Bulletin of the Seismological Society of America*, 85(1), 308-319.
- WHITE, R. A., MILLER, A. D., LYNCH, L. & POWER, J. (1998). Observations of hybrid seismic events at Soufriere Hills Volcano, Montserrat : July 1995 to September 1996. *Geophysical Research Letters*, 25(19), 3657-3660. <https://doi.org/https://doi.org/10.1029/98GL02427>
- WISZNIEWSKI, J., PLESIEWICZ, B. M. & TROJANOWSKI, J. (2014). Application of real time recurrent neural network for detection of small natural earthquakes in Poland. *Acta Geophysica*, 62(3), 469-485.
- WU, Y., LIN, Y., ZHOU, Z., BOLTON, D. C., LIU, J. & JOHNSON, P. (2018). DeepDetect : A cascaded region-based densely connected network for seismic event detection. *IEEE Transactions on Geoscience and Remote Sensing*, 57(1), 62-75.

A Annexes

TABLE A.1 – Matrice de confusion obtenue pour la station AMC avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue de l’OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest (n_estimators=100, criterion=’entropy’, bootstrap=True, class_weight=None), Score moyen obtenu après 10 réalisations

True class	Predicted Class					Recall
2-6	VT	Nested	LP	Hyb	Tor	(%)
VT	164	1	1	15	0	90.6
Nested	1	10	0	1	0	83.3
LP	3	0	1	2	1	14.3
Hyb	33	0	0	32	0	49.2
Tor	0	1	1	0	4	66.7
Precision (%)	81.6	83.3	33.3	64	80	
Accuracy (%)						77.5 ±1.9

TABLE A.2 – Matrice de confusion obtenue pour la station MML avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue de l’OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest (n_estimators=100, criterion=’entropy’, bootstrap=True, class_weight=None), Score moyen obtenu après 10 réalisations

True class	Predicted Class					Recall
2-6	VT	Nested	LP	Hyb	Tor	(%)
VT	164	0	1	15	0	91.1
Nested	1	10	0	1	0	83.3
LP	2	0	2	2	1	28.6
Hyb	40	1	0	40	0	37.9
Tor	0	1	1	0	4	66.7
Precision (%)	79.2	83.3	50	58.1	80	
Accuracy (%)						75.2 ±1.6

TABLE A.3 – Matrice de confusion obtenue pour la station TAG avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue de l’OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle entraîné en utilisant un classifieur de type Random Forest ($n_estimators=100$, $criterion='entropy'$, $bootstrap=True$, $class_weight=None$), Score moyen obtenu après 10 réalisations

True class 2-6	Predicted Class					Recall (%)
	VT	Nested	LP	Hyb	Tor	
VT	162	1	1	17	0	89.5
Nested	1	10	0	1	0	83.3
LP	2	1	3	1	0	14.3
Hyb	28	0	2	36	0	54.6
Tor	0	1	0	0	5	83.3
Precision (%)	83.9	76.9	50	65.5	100	
Accuracy (%)						79.2 $\pm$1.4

TABLE A.4 – Matrice de confusion obtenue avec un rapport de 50 :50 entre les jeux d’entraînement et de test pour le catalogue de l’OVSG entre le 1er Janvier 2013 et le 31 Décembre 2018, avec un modèle multistation global simple entraîné en utilisant un classifieur de type Random Forest ($n_estimators=100$, $criterion='entropy'$, $bootstrap=True$, $class_weight=None$), Score moyen obtenu après 10 réalisations

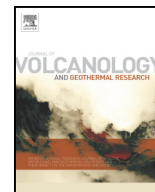
True class 2-6	Predicted Class					Recall (%)
	VT	Nested	LP	Hyb	Tor	
VT	499	1	1	42	0	91.9
Nested	3	30	0	2	1	83.3
LP	6	1	8	5	2	36.4
Hyb	99	0	1	96	0	49.0
Tor	0	3	1	0	14	77.8
Precision (%)	82.2	85.7	72.7	66.2	82.4	
Accuracy (%)						79.5 $\pm$0.9

B Publication



Contents lists available at ScienceDirect

Journal of Volcanology and Geothermal Research

journal homepage: www.elsevier.com/locate/jvolgeores

A machine-learning approach for automatic classification of volcanic seismicity at La Soufrière Volcano, Guadeloupe

Alexis Falcin^{a,*}, Jean-Philippe Métaxian^{a,b}, Jérôme Mars^c, Éléonore Stutzmann^a, Jean-Christophe Komorowski^a, Roberto Moretti^{a,d}, Marielle Malfante^e, François Beauducel^a, Jean-Marie Saurel^a, Céline Dessert^a, Arnaud Burtin^a, Guillaume Ucciani^{a,d}, Jean-Bernard de Chaballier^a, Arnaud Lemarchand^a

^a Université de Paris, Institut de physique du globe de Paris, CNRS, F-75005 Paris, France^b Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, IFSTTAR, ISTERRE, 38000 Grenoble, France^c Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France^d Observatoire volcanologique et sismologique de la Guadeloupe, Institut de physique du globe de Paris, F-97113 Gourbeyre, Guadeloupe^e Univ. Grenoble Alpes, CEA, LIST, F-38000 Grenoble, France

ARTICLE INFO

Article history:

Received 7 July 2020

Received in revised form 7 December 2020

Accepted 10 December 2020

Available online 14 December 2020

Keywords:

Seismology

Volcano

La Soufrière

Classification

Machine Learning

ABSTRACT

The classification of seismo-volcanic signals is performed manually at La Soufrière Volcano, which is time consuming and can be biased by subjectivity of the operator. We propose here a machine-learning-based model for classification of these signals, to handle large datasets and provide objective and reproducible results. To describe the properties of the signals, we used 104 statistical, entropy, and shape descriptor features computed from the time waveform, the spectrum, and the cepstrum. First, we trained a random forest classifier with a dataset provided by the *Observatoire Volcanologique et Sismologique de Guadeloupe* that consisted of 845 labeled events that were recorded from 2013 to 2018: 542 volcano-tectonic (VT); 217 Nested; and 86 long period (LP). We obtained an overall accuracy of 72%. We determined that the VT class includes a variety of signals that cover the VT, Nested and LP classes. After visual inspection of the waveforms and spectral characteristics of the data set, we introduced two new classes: Hybrid and Tornillo. A new random forest classifier was trained with this new information, and we obtained a much better overall accuracy of 82%. The model is very good for recognition of all event classes, except Hybrid events (67% accuracy, 70% precision). Hybrid events are often considered to be a mix of VT and LP events. This can be explained by the nature of this class and the physical processes that include both fracturing and resonating components with different modal frequencies. By analyzing the feature weights and by training a model with the most important features, we show that a subset of the 14 best features is sufficient to obtain a performance that is close to that of the model with the whole feature set. However, these best features are different from the 13 best features obtained for another volcano in Peru, with only one feature common to both sets of best features. Therefore, the model is not universal and it must be trained for each volcano, or it is too specific to the one station used here.

© 2020 Elsevier B.V. All rights reserved.

1. Introduction

Volcanoes are highly complex nonlinear systems where their behavior reflects a diversity of internal hidden processes that can lead to their unrest and eruptive activity. Volcano monitoring rests on the paradigm that various observable parameters have a more or less direct causal link to these complex processes and to the internal variables of the system, as well as to a series of forcing and modulating processes. Although unrest is often linked to rate-dependent processes, this is not always the

case, as sometimes just the absence or presence of some observables can provide very valuable insight into the dynamics of these systems. Volcano monitoring is crucial to evaluate the state of activity of a restless volcano, and to determine the likelihood of a future eruption, along with the probable temporal window of such an occurrence. Hence, volcano monitoring constitutes one of the pillars of the scientific methods that contribute to mitigate the risks to human societies. Among the many disciplines used in volcano monitoring, which include geodesy, geochemistry, gravimetry, geophysical imaging, thermal monitoring, acoustics, and physical volcanology, seismology remains one of the most widely developed. Seismology can provide an often overwhelming wealth of high-resolution continuous real-time data from networks of sensors deployed on volcanoes, both on land and on the sea floor.

* Corresponding author.

E-mail address: falcin@ipggp.fr (A. Falcin).

Optimal seismic monitoring requires the detection, location, magnitude quantification, and classification of large numbers of volcano-seismic signals, and it must be performed daily at volcano observatories. Given the vast progress in volcano seismology in recent decades and the high-resolution monitoring of numerous very active volcanoes, there is now a corpus of knowledge that allows the association of the different types of seismic signals recorded at volcanoes to different source processes with greater certainty (e.g., McNutt (2005); Chouet and Matoza (2013); McNutt and Roman (2015)). Thus, together with other monitoring data, volcanic seismicity constitutes a fundamental tool to track the evolution of the dynamics of volcanic activity in time and space, to provide timely forecasts of the likelihood of future scenarios in a context of uncertainty, which is intrinsic to the behavior of such complex non-linear systems. The occurrence of some events might be precursors of volcanic eruptions, and as such, can be used in early warning systems.

However, while the detection work is now mostly automatic, the classification task is often performed manually. The manual classification tasks are limited, and there are several factors that can decrease the robustness of the labeling. The classification is based on the subjective judgement of the operator, so if the task is performed by several people, the resulting classification criteria can vary from one person to another. This work also has to be carried out in near real-time, while there can be huge amounts of data to deal with, in particular during periods of volcanic crises, when the scientific interpretation needs to be rapid and timely, so that the required advice can be provided by scientists to the authorities to assist in their decision-making processes that are designed to mitigate the risks to the population and the critical infrastructures. With the constantly increasing volumes of data and the limitations of the manual classification, the development of automatic tools using a machine-learning approach to rapidly classify these volcanic seismic signals as accurately as possible is essential for spatio-temporal interpretation and process modeling of the data, as well as for pattern recognition. This methodology thus has the potential to significantly improve the timely tracking and analysis of volcanic unrest and the forecasting of its possible evolution.

The machine learning methodology consists of training an algorithm to do a task, here assigning a seismic class to a volcanic event. However, rather than analyzing the waveform of the signal, we transform this signal into a set of features that describes its characteristics. Then, the role of the machine-learning algorithm is to define boundaries in the feature space that specifically correspond to each class considered. To best differentiate each of the classes, the features must be carefully chosen and must represent the variability of the signals. Here, we chose to use the feature set proposed by Malfante et al. (2018a, b), which provides a general and precise description of transient signals. These have been shown to be efficient, and they also allow us to investigate whether features selected at one volcano can be used for another volcano.

One of the first attempts to automatically recognize earthquakes was made by Allen (1978), who built a decision tree, through which 70% of the events could be recognized. Many studies in recent years have considered such automatic classification of volcano seismic events, and many techniques have been used. Benítez et al. (2006) used hidden Markov modeling for continuous classification of four classes, and they reached an overall accuracy of 90%, where the overall accuracy is the percentage of correct predictions out of all of the predictions made. Hibert et al. (2014) used fuzzy logic and a method based on decision rules to distinguish rockfalls from volcano-tectonic events, and they obtained a success rate of 92%. Langet et al. (2014) used logistic regression and a support vector machine, through which they obtained 90% and 92% good results, respectively. At Ubinas Volcano, Malfante et al. (2018a, b) tried a random forest (RF) and support vector machine classifiers, and here they obtained 92.5% and 92.1% overall accuracy, respectively. Titos et al. (2018) used a deep neural network approach with seven classes of volcanic events, and they reached 94% overall accuracy. At the same time, Curilem et al. (2018) used spectrogram

cross-correlations in conjunction with the K-nearest neighbors algorithm with five classes of volcanic events at Llama Volcano (Chile), and they reached an overall accuracy of 95%. Finally here, Bueno et al. (2019) used a Bayesian neural network for classification of five classes on two volcanoes, through which they achieved an excellent performance of 92.1% when the two datasets were merged, and they also showed that their uncertainties were linked to the state of unrest of the volcanoes.

The algorithm used in the present study is a RF classifier, because this allows a wide range of features to be handled without overfitting and producing a model that is too specific to the data used for the training, and it also gives information about the feature importance. This kind of algorithm has already been tested and has been shown to be powerful not only in a volcanic context. Provost et al. (2016) proposed an automatic classification method based on the computation of 71 features, and they used a RF classifier. They focused on classification of events recorded near the Super-Sauze landslide (southeastern France). They obtained 93% sensitivity for this classification, where the sensitivity defines the proportion of the true events that are correctly classified. Using several stations, Maggi et al. (2017) presented an operational automatic classifier based on a RF algorithm for monitoring at the Piton de la Fournaise Volcano. They defined eight classes of seismic signals used a multi-station approach, and they reached peak performance when running on a three-station combination (92% good classification). Hibert et al. (2017) also proposed a RF classifier for binary classification of seismic signals (rockfall, volcano-tectonic) that were recorded at the Piton de la Fournaise Volcano. They obtained results from 90% to 99%, which depended on the size of their training dataset. Malfante et al. (2018a, b) proposed a large set of features for representation of the seismic signals, as 34 features in three representation domains: temporal, spectral, and cepstral domains (where cepstral describes the periodic properties of a signal, as commonly used in speech processing, and as obtained by computing the Fourier transform of the logarithm of the signal spectrum). Hence they defined 102 features, which they used to train a RF classifier with signals from Ubinas Volcano in Peru, and with which they obtained 92.5% overall accuracy. Malfante et al. (2018a, b) also showed that with only the three best features included, they obtained 84% overall accuracy, while with the 13 best features they obtained 90% overall accuracy with their RF classifier.

Some studies have refined their initial classification into a second step of automatic classification. Langer et al. (2006) classified five volcanic classes at Soufrière Hills Volcano, in Montserrat. They obtained a fair performance, where 70% of the automatic signal classification was consistent with the manual classification of the signals. From an analysis of the misclassified events, however, they found that for most of them, the original *a-priori* classification was incorrect. They carried out a re-analysis by hand of the seismic traces recorded at different seismic stations. Then they trained the classifier again using this new information, and here obtained a success rate of 80% good classification. Hammer et al. (2013) proposed a hidden Markov model for detection and classification of events in an alpine context. They applied their model to the continuous signal and they managed to detect 97% of the events and correctly classified 87% of them. At first they distinguished two classes, as earthquake and quarry-blast, although they were also able to detect rockfalls using a threshold criterion to flag a poor match between the incoming signal and all of the defined classes. Langet et al. (2014) showed that the current classification for Kawah Ijen Volcano with eight classes was not adapted. By using an unsupervised classification, they showed that with the data and features available, only five classes were clearly distinguishable.

In this paper, we have applied a method following that developed for the first time by Malfante et al. (2018a, b), where we have automatically classified volcano-seismic events recorded from 2013 to 2018 for La Soufrière Volcano. In Section 2, we present La Soufrière Volcano, in terms of its historical eruptive activity, then we detail how the detection and classification work was performed at the observatory, and finally

we discuss the dataset used in our analysis. Then, in Section 3, we present the method and discuss the large set of statistical features that were calculated in the three domains to represent the seismic signals (*i.e.*, temporal, spectral, cepstral domains). In Section 4, we test the robustness of the actual classification, and then we refine the classification by adding new classes in the analysis. We show which features are the most useful to differentiate classes of volcanic earthquakes at La Soufrière, and finally, we compare these important features with the most valuable features defined by Malfante et al. (2018a, b) for Ubinas Volcano.

2. Data

2.1. La Soufrière de Guadeloupe

La Soufrière de Guadeloupe is an active and hazardous volcano located on the island of Basse-Terre in Guadeloupe in the Lesser Antilles (Fig. 1). It is currently monitored by the *Observatoire Volcanologique et Sismologique de Guadeloupe - Institut de Physique du Globe de Paris* (OVSG-IPGP). Numerous magmatic and nonmagmatic eruptions have occurred in its history. The current andesitic dome was formed in the last major magmatic eruption in 1530 CE (Boudon et al., 1988, 2008). The volcano also had several historical nonmagmatic phreatic or hydrothermal explosive eruptions in 1690, 1797–98, 1809–12, 1836–37, 1956, 1976–77 CE (Komorowski et al., 2005). The last one of these was particularly violent, and required the evacuation of more than 70,000 people for 6 months; it was also associated with a major controversy in the scientific community (Komorowski et al., 2005; Hincks et al., 2014; Komorowski et al., 2016). After this crisis, the volcanic activity gradually declined, and in 1990 it reached its lowest level since the start of seismic monitoring in 1950 (Komorowski et al., 2005; Villemant et al., 2014; Jessop et al., 2019; Moretti et al., 2020). Since 1992, its seismic fumarolic and thermal unrests has increased gradually (Komorowski et al., 2005; Villemant et al., 2014) (OVSG-IPGP, 1999–2020), to reach its highest level of seismic energy on April 27, 2018. This resulted in the strongest volcano-tectonic earthquake recorded (M4.1) since the phreatic eruption of 1976–77, with an associated episode of deep-sourced magmatic degassing (Moretti et al., 2020). Although the release of seismic energy has decreased markedly since the January–April 2018 period of elevated unrest, in August 2018, the seismicity of La Soufrière de Guadeloupe was characterized by the periodic occurrence of earthquake swarms of low energy, with numerous volcano-tectonic (VT) and Nested signals, and only rare long-period (LP) signals (OVSG, 2018–2020). Between

August 2018 and the end of April 2020, 37 of these swarms were recorded by the OVSG (based on the swarm quantitative criteria defined by the OVSG), for a total of 7114 identified signals with a mean of 192 ± 187 events per swarm (minimum, 13; maximum, 1014). Although of low energy, this continuous heightened seismicity represents a very significant challenge for event identification, classification, and interpretation by the observatory staff, to provide tracking in an efficient and timely manner during this unrest.

2.2. The OVSG classification

Currently at OVSG, the detection work is carried out mostly automatically using a short-time average over long-time average algorithm. The operators on site review the detected signals daily, considering the several stations located at different distances from the volcano summit, and they determine the magnitude and location of the signals. By observing the seismograms at different stations, they can assign a class to detected events according to the nomenclature used by the OVSG. The vast majority of volcanic signals have very low magnitudes (*i.e.*, mostly negative) and short durations, and they are very often discernible by only two stations, unlike tectonic earthquakes, which are generally recorded by more stations and by stations from larger networks.

Although five main types of signals are recognized in the data analysis by the observatory (Moretti et al., 2020), only three main classes are readily distinguishable on the continuous seismic traces during the daily analytical protocol, which are defined as: VT events, Nested events, and LP events. The two other classes, as Hybrid events and LP monochromatic events (or Tornillos) are included as VT and LP events, respectively. Fig. 2 shows the typical waveforms, spectrograms, and Fourier spectra of these events.

2.2.1. Volcano-tectonic, or high frequency, earthquakes

These events have a high frequency content similar to tectonic earthquakes. Chouet and Matoza (2013) defined their frequency range as 5 Hz to 15 Hz. For La Soufrière, a characteristic peak is often observed between 10 Hz and 15 Hz. The waveform has a very impulsive P-wave arrival. These are brittle failure events that are associated with stress changes due to magma movement (Chouet and Matoza, 2013).

2.2.2. Nested earthquakes

These events appear as small packets of several seismic signals in which successive events appear within the coda of each other. They are not concomitant or precursors to any particular phenomenon (Moretti et al., 2020). Nested events consist of a sequence of several volcano-seismic events with very short inter-times, very often as more than 6 seismic events in a short sequence (10 s) (Ucciani, 2015; Moretti et al., 2020). These Nested events are mainly composed of VT high-frequency events, but not always, as they can also sometimes be Nested LPs. The different signals in a Nested event are not always similar, in terms of amplitude or waveform. The source process related to this class is not well understood, and indeed, these events are specific to La Soufrière Volcano. Waite et al. (2008) indicated the presence of ‘drumbeat’ LP earthquakes at Mount St. Helens Volcano, which are characterized by a restricted range of inter-event times and amplitudes compared to the more typical activity and highly similar waveforms. These properties require a persistent source location, a nondestructible or rapidly renewing mechanism, and a physical system that involves small oscillatory deviations from near equilibrium conditions. These events of Mount St. Helens Volcano are similar to the Nested events of La Soufrière de Guadeloupe; although the Nested events are mostly VT and the drumbeats are LP, the different signals are similar in terms of amplitude and waveform.

2.2.3. Long-period or low-frequency earthquakes

The frequency range for these events is generally taken as 0.5 Hz to 5 Hz (Chouet and Matoza, 2013). At La Soufrière, these LP events are

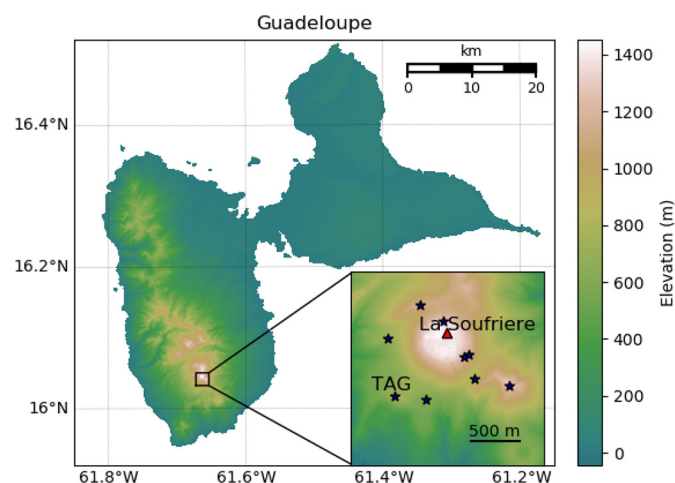


Fig. 1. Map of Guadeloupe. Red triangle, La Soufrière summit; blue stars, seismic stations. TAG station was used in this study.

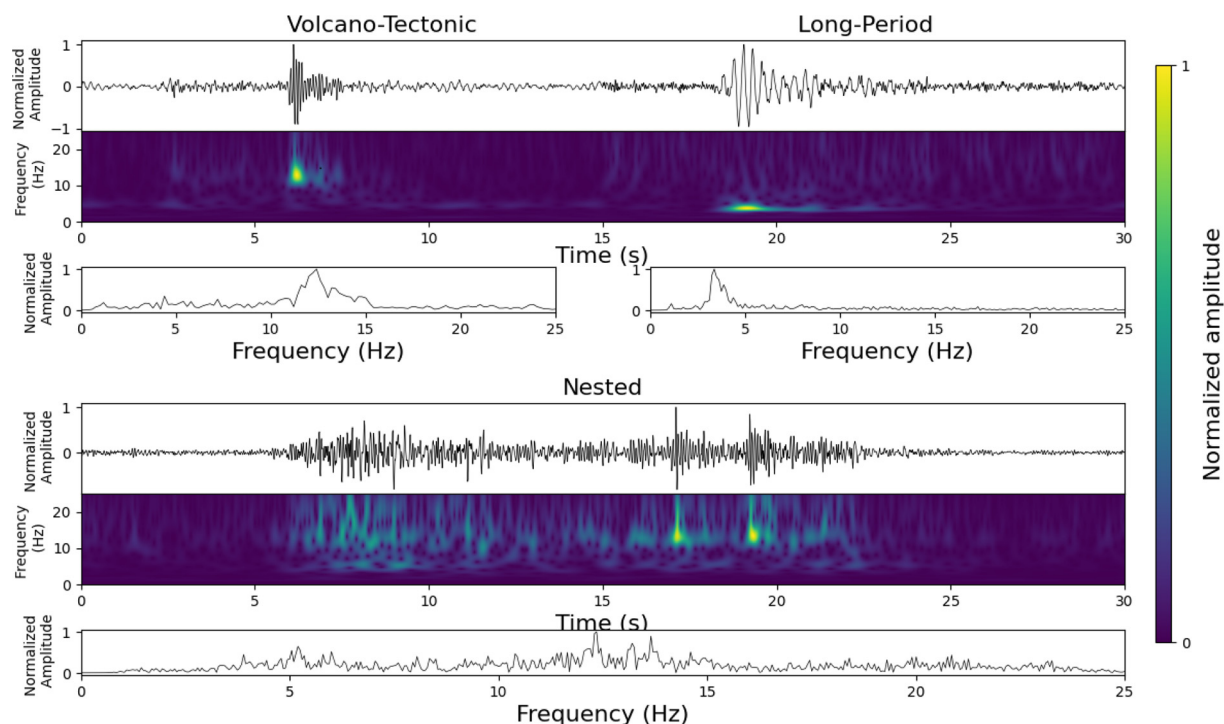


Fig. 2. Examples of waveforms filtered between 0.8 Hz and 25 Hz, for spectrogram and Fourier spectra of volcano-tectonic events (top left), long-period events (top right), and Nested events (bottom) recorded at station TAG, for the vertical component, from January 1, 2013, to December 31, 2018.

characterized by a relatively narrow spectral content, at around 4 Hz. Their waveforms are characterized by the emerging arrival of P-waves, with the S-phases not identifiable. LP events can be generated by the resonances of fractures, dykes, conduits, or cavities during the propagation of magmatic or hydrothermal fluids (Chouet, 1988). Some other models without fluids or resonance have been proposed, such as trapped waves in loosely consolidated and shallow layers of the crust (Bean et al., 2008). Another model proposed by Bean et al. (2014) indicated that LPs might be generated by slow rupture of the nonconsolidated volcanic material itself.

2.3. Catalog

The learning base used for this analysis comes from the catalog built by the OVSG, which is available on WebObs (Beauducel et al., 2020). WebObs is a web-based tool that performs integrated, centralized, and automated real-time volcano monitoring. This system mainly offers a modular database for equipment network management, with a dozen evolving dedicated periodic tasks for each monitoring technique, such as for seismology, geodesy, and geochemistry, with automated execution of periodic tasks and web-form interfaces for manual data input/editing and export (Beauducel et al., 2020). Our study period extended from 2013 to 2018. The information available on WebObs indicates the first arrival times, the event durations, and the classes.

In this paper, we focus on the data recorded by one component (i.e., the vertical component) of one station. We selected the three-component Piton Tarade station (TAG; nanometric Taurus Trillum compact 120S, at 100 Hz), a historic station of the La Soufriere monitoring network that is located closer than 1 km from the summit. This station is protected from external disturbance and is not subject to strong winds, and we have the data for 84% of the days for the period from 2013 to 2018. The catalog contains 7149 events, 78% of which are VT events, 20% are Nested events, and 2% are LP events, and hence these three classes are highly unbalanced. Fig. 3 shows the distribution of these events over time. The distributions of the VT and Nested events

are correlated with a much greater number of VT events than Nested events. The LP events are evenly distributed over time, except for the swarm of a few dozen events that occurred in April 2017. Conversely, the distributions of the VT and Nested events are not constant over time, and we observed a strong increase in their occurrence at the end of 2018.

3. Processing methodology

In this section, we detail our workflow (Fig. 4) to automatically classify the three classes of events; namely, VT, Nested, and LP events. There are several main steps involved here.

3.1. Data

Here our data consist of labeled windowed seismic waveforms. The waveforms are extracted from the continuous signals, with the detection time and duration provided by the OVSG. These waveforms are associated with a label linked to the physical phenomenon that triggers the event.

3.2. Pre-processing

First, we need to eliminate the poor quality signals from our dataset. Not all of the events recorded by the network are observed well at station TAG. We need to visually check if the events are correctly labeled, and this is easier if the signals are strong compared to the ambient noise. A criterion was therefore applied to keep the signals with a significant signal-to-noise ratio at TAG. The criterion we applied was to compare the mean squared amplitude of 20 s of noise recorded before the event with the mean squared amplitude of the event considered. We kept those signals with a signal-to-noise ratio greater than 1.5. Then, we applied a bandpass filter between 0.8 Hz and 25 Hz to keep only the signals related to the event. Finally, signals were normalized by

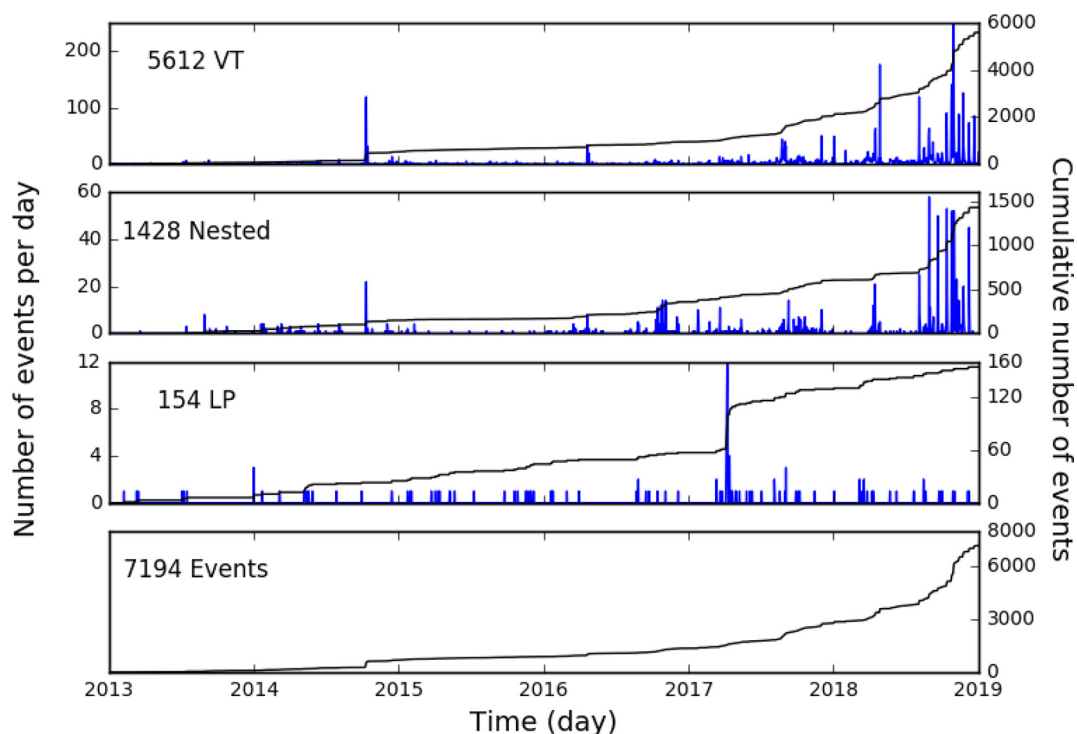


Fig. 3. Distributions (blue) and cumulative distributions (black) of the numbers of events between January 1, 2013, and December 31, 2018, for (top to bottom) volcano-tectonic (VT), Nested, and long period (LP) events, along with the cumulative distribution of all of the classes (data from OVSG-IPGP).

the maximum of 1.0, so that the model can be applied to all observations regardless of their amplitudes.

3.3. Feature calculation

The data used for learning are not the windowed waveforms, but a set of descriptors that are extracted from the windowed waveforms. The features are important as they carry the information related to the classes. Here, we used the features proposed by Malfante et al. (2018a, b). The description of the features is given in Table 1. We grouped these into three categories, as statistical features ($n = 9$), entropy-based features ($n = 9$), and shape descriptors ($n = 16$). These 34 features were computed in three representation domains of the signal: the temporal domain, the spectral domain, and the cepstral domain. The cepstral domain describes the periodic properties of the signal, as commonly used in speech processing, and it is obtained by computing the Fourier transform of the logarithm of the signal spectrum.

Using a large number of features allows the signals to be represented in many ways, while keeping as much of the information about the signal as possible. Another advantage of using features is to reduce the dimension of the data, while keeping the information contained in the signals. Here, the input vector has a dimension of $34 \times 3 = 102$, whereas dealing with the corresponding 10-s-long waveforms sampled at 100 Hz would imply a vector size of $10 \times 100 = 1000$. We do not expect to need all of the features, as some provide redundant or useless information; e.g., the length of the signal is the same whatever the representation domain.

3.4. Training

The training step consists of defining the boundaries between the different classes in the feature space. For this, the learning algorithm automatically selects the most discriminative features among the 102

features. We chose to train the model from a given percentage of the labeled data, and we used the rest of the data to test the reliability of the model constructed. The algorithm used here is the RF classifier, which is available in the python scikit-learn library (`sklearn.ensemble.RandomForestClassifier`).

This algorithm involved decision trees that split the parameter space into simple regions. A decision tree (Quinlan, 1986) is an algorithm with a tree-like structure where each internal node corresponds to a test on a subset of the features, each branch represents the outcome of the test, and each leaf node is a class label. The random forest algorithm (Breiman, 2001) consists of a collection of decision trees. This works as follows: first subsets are randomly selected from the learning dataset, then a decision tree is constructed for each subset, and a prediction result is obtained for each tree. Then, the most frequent prediction is selected as the final prediction. The biggest advantage of the RF over a single tree is that the overfitting problem is attenuated by averaging out the predictions from all of the trees. With this algorithm, we also gain information about the feature importance. This information is obtained by computing the relevance score of each feature in the training phase. This allowed us to reduce the number of features by selecting only the most important. We performed a grid search in the hyperparameter space to determine which features were the best, and we selected the hyperparameters around which the classification score was stable, which are given in the captions to Tables 2 and 4.

3.5. Test

To test the model, we performed cross-validation, by randomly partitioning the labeled dataset between a training dataset and a testing dataset. The random selection of the training dataset and the testing dataset was performed several times, to have statistically valid results. Due to the low number of LPs compared to the other classes, we made sure that the number of LPs was balanced in all partitions. We can

modify the partition coefficient, and thus see the effects of the size of the training set on the performance of the model. To analyze the results, we computed the mean confusion matrix obtained after several trials (*i.e.*, at least 10) by comparing the predicted classes with the real classes of the test dataset. We also considered three metrics: the accuracy, the precision, and the overall accuracy:

$$\text{Accuracy} = \frac{\# \text{GoodPredictionClass}_i}{\# \text{TotalTrueClass}_i} \quad (1)$$

$$\text{Precision} = \frac{\# \text{GoodPredictionClass}_i}{\# \text{TotalPredictedClass}_i} \quad (2)$$

$$\text{OverallAccuracy} = \frac{\# \text{GoodPrediction}}{\# \text{TotalEventTestDataset}} \quad (3)$$

These metrics provide information on the specificity (*i.e.*, false-positives) and sensitivity (*i.e.*, false-negatives) of the model. If the precision is too low, it means that there are too many bad predictions compared to good ones, and thus it means there are too many false-positive outcomes. If the accuracy is low, it means that most of the data of a class have been badly predicted, and thus it means there are too many false-negative outcomes. The overall accuracy gives information on the general performance of the model.

3.6. Classification

Once the model is trained and reliable, we can apply it on a new dataset, as either labeled or not labeled. This is the operational part of the process, which we do not include as part of this study.

4. Results and analysis

4.1. Performance of initial classification

Here we present the performance of the automatic classification using the three classes of events: VT, Nested, and LP events. At the TAG station, many of the events detected did not meet the quality criterion of the pre-processing step of the analysis described earlier. The OVSG detected 7149 events over the period from 2013 to 2018, although only 845 passed our selection phase. The selection phase allows visual checking of whether the events we analyze were correctly labeled, and for computational reasons, it is easier to process less data. This analysis was thus performed with 542 VT, 217 Nested, and 86 LP events. Table 2 gives the average confusion matrix of the classification that was obtained after 10 trials, with a ratio between the training dataset and the testing dataset of 50:50. For each trial, we randomly selected the data used for the training, and tested the model with the remaining data.

We obtained an overall accuracy of $73\% \pm 1\%$, which is relatively low compared to the scores obtained in the literature, some of which reached up to 80% or 90%. Different ratios between the training and the test dataset were also tested. Between the 80:20 and 30:70 ratios, the data were almost the same, at around $73\% \pm 1\%$. With more than 80% of the data for the training and less than 30% for the testing, the overall accuracies decreased, and overfitting and underfitting problems started to appear. We detected several trends after analysis of these data.

The score of 73% was largely due to the size of the VT class compared to the rest of the dataset, which masks the poor classification. The accuracy of the VT class was 93%, so most of the true VT events are well classified. However, the precision of the VT class was only 68%, which means that some events of the VT class are also assigned to the true Nested and LP events. In more detail, 71 of the 108 Nested events used for the test were classified as VT events, and 16 of the 43 LP events were classified as VT events. The bad precision of the VT class can be

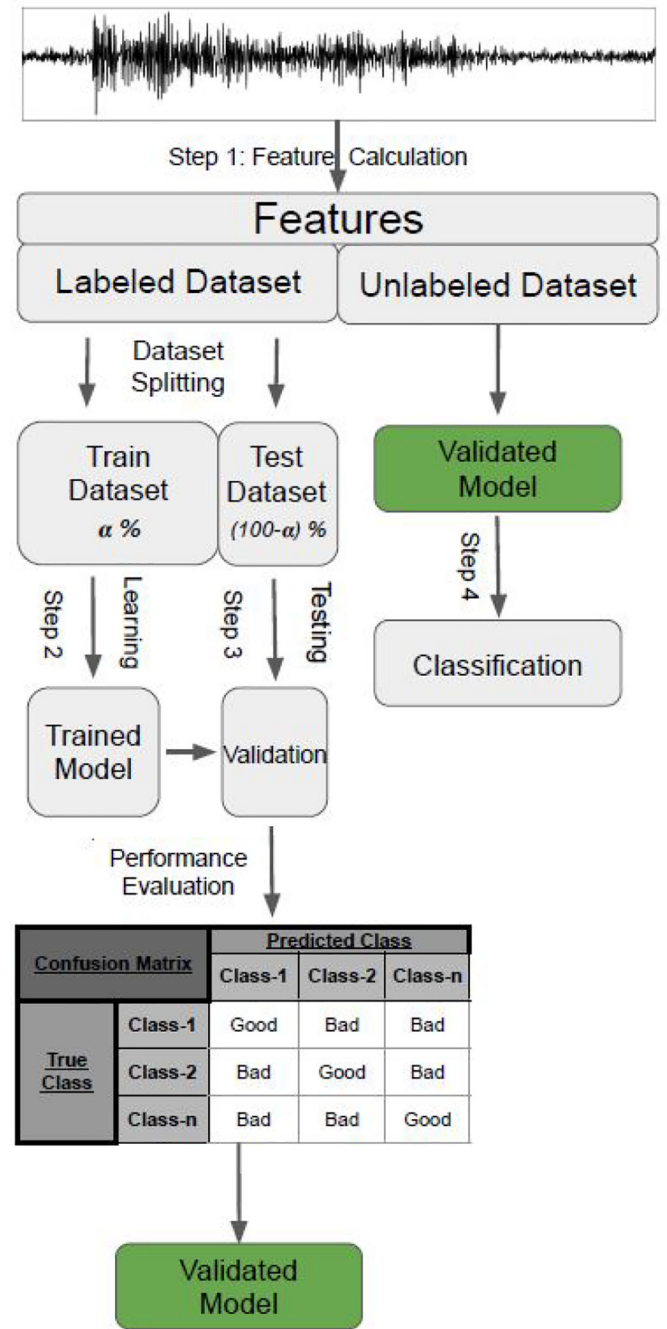


Fig. 4. Workflow of the machine learning process. Step 1: Feature calculation step to represent the signals. Step 2: Learning step on a given proportion of the dataset. Step 3: Testing step to evaluate the performance of the model. Step 4: Classification step on new, unlabeled data with a validated model, as the operational part of the process.

explained in two ways. The VT class is too broad, and contains a variety of signals that cover the characteristics of the Nested and LP events, or on the contrary, there are signals in the LP and Nested classes that have characteristics of the VT class.

The precision of the Nested classification was 61%, so the predicted nested events are not reliable. The accuracy was 31%, so most of the true Nested events are not well recognized. Indeed, 71 of the 108 true Nested events were classified as VT events, and only 4 of the 108 were classified as LP events.

The precision of the LP classification was 73%, which means that the LP prediction was quite reliable, although the accuracy of 51% shows

Table 1

Features used to represent each transient signal, as the feature set from Malfante et al. (2018).

Features	Definition	Ref.
Statistic features		
Length	$n = \text{length}(s)$	1
Mean	$\mu_s = \frac{1}{n} \sum_i s[i]$	2
Standard deviation	$\sigma_s = \sqrt{\frac{1}{n-1} \sum_i (s[i] - \mu_s)^2}$	3
Skewness	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^3$	4
Kurtosis	$\frac{1}{n} \sum_i \left(\frac{s[i] - \mu_s}{\sigma_s} \right)^4$	5
i of central energy	$\bar{i} = \frac{1}{E} \cdot \sum_i E_i \cdot i$	6
RMS bandwidth	$B_i = \sqrt{\frac{1}{E} \sum_i i^2 \cdot E_i - \bar{i}^2}$	7
Mean skewness	$\sqrt{\frac{\sum_i (i - \bar{i})^3 E_i}{E B_i^4}}$	8
Mean kurtosis	$\sqrt{\frac{\sum_i (i - \bar{i})^4 E_i}{E B_i^4}}$	9
Entropy features		
Shannon entropy ^a	$-\sum_j p(s_j) \log_2(p(s_j))$	10 to 12
Rényi entropy ^b	$\frac{1}{1-\alpha} \cdot \log_2 \left(\sum_j p(s_j)^\alpha \right)$	13 to 18
Shape descriptors features		
Rate of attack	$\max_i \left(\frac{s[i] - s[i-1]}{n} \right)$	19
Rate of decay	$\max_i \left(\frac{s[i] - s[i+1]}{n} \right)$	20
Ratios	min/mean and max/mean	21 to 22
Energy descriptors	Signal Energy, maximum, average, standard deviation, skewness and kurtosis	23 to 28
Specific values	min, max, i of min, i of max, threshold crossing rate and silence ratio	29 to 34

Note. Features computed for a signal $s[i]_{i=1}^n$ (in which i might correspond to a temporal, frequency or cepstral sample). $E = \sum_{i=1}^n s[i]^2$ and $E_i = s[i]^2$ describe the signal energy and the energy at sample i , respectively. Some features have a dimension greater than others; e.g., entropy measurements are made on three different estimations of the amplitude probability (i.e., different histogram bin numbers).

^a Bin numbers for probability estimation: 5, 30 and 500.

^b Bin numbers for probability estimation: 5, 30, 500, $\alpha = 2$, inf.

Table 2

Confusion matrix obtained with a ratio of 50:50 between the training and testing datasets for the catalog from the OVSG for January 1, 2013, to December 31, 2018, with the model trained using a random forest classifier (n_estimators = 100, criterion = 'entropy', bootstrap = True, class_weight = None), mean score after 10 trials.

True class	Predicted Class			Accuracy (%)
	VT	Nested	LP	
VT	251	16	4	92.6
Nested	71	33	4	30.6
LP	16	5	2	51.2
Precision (%)	74.3	61.1	73.3	
Overall accuracy (%)				72.5 ± 1.0

that half of the true LP events are not well recognized by the model. Almost a third of these misclassifications of LP (16 of 43) were classified as VT (48 of 54).

These results show that VT events can be confused with LP events and Nested events, while Nested and LP events are rarely confused with each other. The confusion in the classification of the VT and Nested events was partly because the Nested events actually consisted of several individual VT events that were separated by very short inter-event times, such that the coda of the first signal was mixed with the onset of the next signal. In the other cases, many VTs appeared to

show what looks like an emergent P-phase on the low resolution seismic trace. However, upon zooming in on the waveforms in the Seiscomp3 environment when picking phases, it was clear that this apparent P-wave was a low amplitude initial VT signal that was quickly followed by another larger amplitude VT signal with a P-phase and S-phase. Thus, if Nested events are groups of VT signals, it is highly possible that the discrimination between these is difficult because many of the description features of VT events are also in the Nested events.

At OVSG, the visualization tool used to manually classify events makes this work very complicated, because the spectrograms are not readily available to all of the operators, so some events might have been misclassified. This would explain the broad variety of signals in the VT class. New visualization of the signals with a spectrogram and new labeling based on the definition of the VT (5–20 Hz) and LP (0.1–5 Hz) frequency domains given here appears to be necessary to improve the performances and the reliability of the models.

4.2. Performance of the refined classification

We visually reviewed all of the 845 events, to check whether some were misclassified and whether we observed signals that showed characteristics different from the classes of events already considered, and if new classes should be introduced into our analysis.

After reviewing all of the signals by hand, two new classes were introduced, the Hybrid class and the LP monochromatic, or Tornillos, class (Fig. 5). These types of events were considered by the observatory in the form of a comment, but were not taken into account during the routine daily classification work. Here we now add those two classes into our learning step.

4.2.1. Hybrid earthquakes

Numerous detailed analyses of volcanic earthquakes during recent eruptions (Redoubt, Lahr et al. (1994); Soufriere Hills, Miller et al. (1998); White et al. (1998)) have shown that a distinct category of earthquakes has spectral characteristics of both VT (or high frequency) and LP earthquakes. These have been called Hybrid earthquakes (e.g., Chouet and Matoza, 2013). Typically, they have a high frequency impulsive arrival between 10 Hz and 20 Hz, as typical of VT earthquakes, while the coda is dominated by lower frequency waves in the upper range of the LP frequency spectrum (1.2–2.5 Hz, as at Redoubt, Lahr et al. (1994) and Soufriere Hills, Montserrat, White et al. (1998)). Also, the LP component is observed from the beginning to the end of these events. This indicates a fluid component in the mechanism of rupture. At La Soufrière, the Hybrid earthquakes (Fig. 5) had a lower frequency coda, at around 5 Hz to 6 Hz, and a higher frequency onset, at around 12 Hz to 14 Hz. Various studies have sought to understand Hybrid events. Lahr et al. (1994) proposed a model halfway between VT and LP, with fragile fracturing processes producing high frequencies, and then the propagation of fluid responsible for the resonance phenomena that produces low frequencies. Neuberg et al. (2000) showed a clear continuum between LP and Hybrid events. Harrington and Brodsky (2007) proposed a model where a simple fracturing process with a very slow rupture velocity was enough to explain these events.

4.2.2. LP monochromatic earthquakes

These Tornillos are a subcategory of LPs; however, their particular waveform (which looks like a screw) makes them easily distinguishable. They were first described as such for Galeras Volcano (e.g., Gomez et Torres, 1997), and have subsequently being identified for many eruptions. We therefore have an emerging wave arrival, with a duration of a few tens of seconds, an almost sinusoidal signal, and a coda that decreases very slowly, and almost linearly. Their Fourier spectrum has a characteristic peak, which was at around 4 Hz for La Soufrière. On the spectrogram we saw that the resonance lasted longer than for the LPs. A model of self-oscillations of fluid filling a cavity was proposed as the physical process that generates Tornillos seismic signals

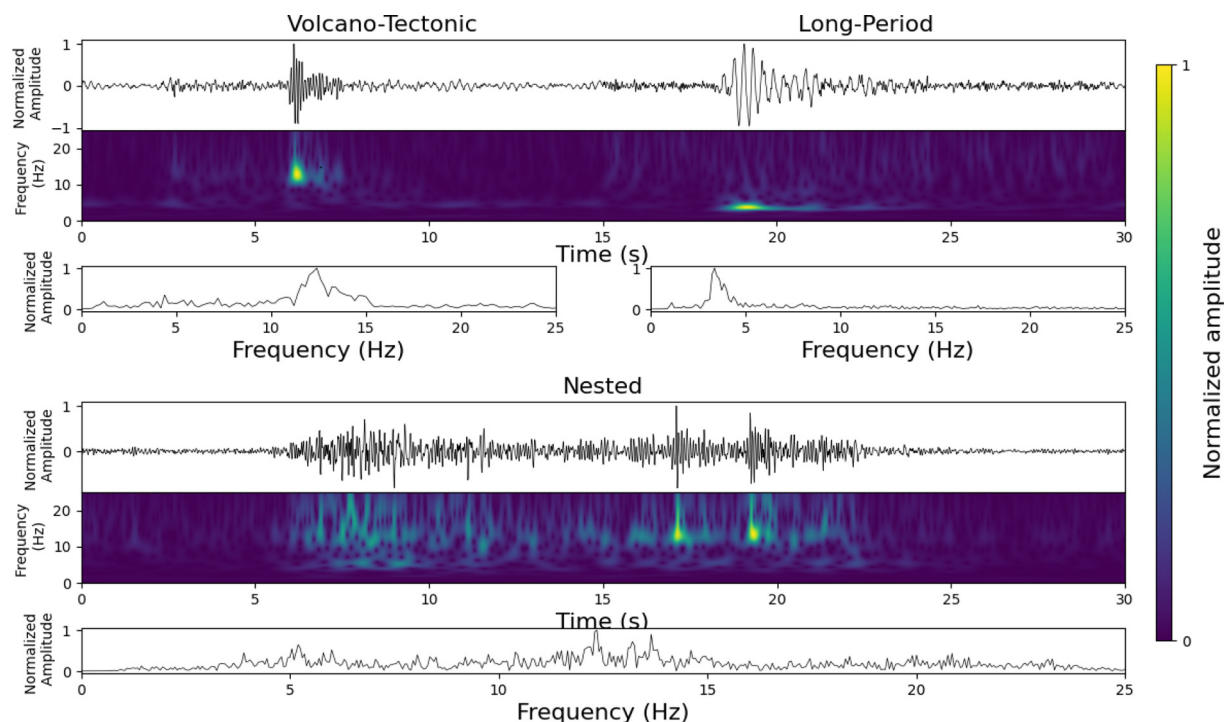


Fig. 5. Example of a waveform filtered between 0.8 Hz and 25 Hz, with the spectrogram and Fourier spectrum for Tornillo events (top) and Hybrid events (bottom) recorded at station TAG at La Soufrière de Guadeloupe, for the vertical component, for January 1, 2013, to December 31, 2018.

(Gómez, 1997; Konstantinou, 2015). In a recent experimental study, Fazio et al. (2019) suggested that gas pressure gradients trigger the event that regulates the slow decay of the coda, whereas fluid resonance in small structures controls the frequency content of the signal. This type of event was introduced into the analysis for the observatory in 2001, but was removed from the classification routine later.

Table 3 shows the classes that were newly assigned to our catalog. First, 89 Nested events and 15 LP events were reclassified as VT events. Then 39 VT events and 22 Nested events were reclassified as LP events. Only 9 VT events and 1 LP event were reclassified as Nested events. Also, 2 Nested events and 26 LP events were reclassified as Tornillos events. Finally, 151 VT events, 45 Nested events, and 8 LP events were reclassified as Hybrid events.

We note that very few of the Nested events remained in this class (59 of 217), with a lot reclassified as VT events (89 of 217). The Nested events are a relatively unique type of seismo-volcanic event, because these are specific to the La Soufrière seismicity. Nothing exactly like this type of event has been reported in the literature, so to maintain a certain robustness in the new labeling, and not to confuse this type of event with the VT events, we chose an explicit criterion to decide whether these belonged to one or the other class. We considered the events as Nested when they clearly presented several signals in the waveform. As the reviewing of the events was carried out using only one station, it is possible that an event clearly presents several signals

at one station but not at another, because these Nested signals can be different in terms of their amplitudes and waveforms.

When we created a model with these new classes and the same ratio between the training dataset and the test dataset of 50:50, as previously here, the overall accuracy increased from 73% to $82\% \pm 2\%$. This shows that this new classification was much better for the automatic process. Furthermore, by looking at the results in detail (Table 4), we see that the classification of certain classes was significantly improved. With different ratios between the training dataset and the test dataset, as 90:10, 70:30, 30:70, and 10:90, the accuracies obtained were (respectively): $83\% \pm 4\%$, $83\% \pm 2\%$, $82\% \pm 1\%$, and $77\% \pm 2\%$; i.e. they remained similar.

Then we looked at the results class by class, to analyze the errors. These were significantly lower for the Nested class, with 68% and 79% for the accuracy and precision, instead of the 31% and 61%, respectively, for the previous classification. One true LP event was classified as a Nested event, and *vice versa*. Five Nested events were classified as VT and five as Hybrid. This confusion was seen when there were few Nested signals in the event (i.e. 2 or 3) and when one of these signals was much more energetic than the others, whereby the model

Table 3

New event classes that were manually assigned after visual review based on the waveform, the spectrogram, and the Fourier spectrum.

Old class	New class					Total
	VT	Nested	LP	Hyb	Tor	
VT	343	9	39	151	0	542
Nested	89	59	22	45	2	217
LP	15	1	36	8	26	86
Total	447	69	97	204	28	845

Table 4

Confusion matrix obtained with a ratio of 50:50 between the training and testing datasets for the refined catalog for January 1, 2013, to December 31, 2018, with the model trained using a random forest classifier ($n_{\text{estimators}} = 100$, $\text{criterion} = \text{'entropy'}$, $\text{bootstrap} = \text{True}$, $\text{class_weight} = \text{None}$), mean score after 10 trials.

True class	Predicted Class					Accuracy (%)
	VT	Nested	LP	Hyb	Tor	
VT	203	2	1	18	0	90.6
Nested	5	23	1	5	0	67.7
LP	1	1	41	6	0	83.7
Hyb	25	2	7	69	0	67.0
Tor	0	1	1	0	13	86.7
Precision (%)	86.8	79.3	80.4	70.4	100	
Overall accuracy (%)						82.1 ± 1.6

recognized the energetic signal. The six events that were incorrectly classified as Nested events were also particularly noisy events.

The Tornillos events were well classified by the model, with 100% and 87% accuracy and precision. There was 1 true Tornillo event classified as an LP event. Once again, this kind of error in the model is easily understandable, because the Tornillo class is a subclass of the LP class. The biggest difference between these two classes is the long monochromatic resonating coda in the Tornillos events. Depending on the shape of the coda, we can choose to classify an event as LP or Tornillo, and sometimes this distinction was hard to make. The other badly classified Tornillo was labeled as Nested by the model; it was the most noisy Tornillo event of the catalog. In addition, the various tests carried out with different sizes of training sets showed that there was no need to have a large training database to effectively recognize the Tornillos events.

The accuracy of the predicted LP events was increased from 51% to 84%. One true VT event, one true Nested event, one true Tornillo event, and seven true Hybrid events were classified as LP events. The precision of the LP events was increased from 73% to 80%. One true LP event was classified as a VT event, and one as a Nested event, with six classified as Hybrid events. LP events can be confused with VT, Hybrid, and Tornillo events. The confusion between the LP and VT events decreased a lot compared to the previous classification, with 16 LP events of 43 classified as VT events before, and only 1 of 49 with this new classification.

With the new classification, the VT class showed an accuracy of 91% and a precision of 87%. While the accuracy was slightly decreased from 93%, the precision was increased from 74% to 87%. These results show that the variety of the signals in the previous VT class was reduced.

The hardest events to classify were the Hybrid events. The accuracy was 67% and the precision was 70%. Hybrid events were only confused with VT and LP events, with 25 true Hybrid events classified as VT events, and 67 as LP events. Here, 25 predicted Hybrid events were true VT events, and six predicted Hybrid events were true LP events. Hybrid events were more often confused with VT events than LP events. This might be explained by the impulsive and high frequency onset of the Hybrid events, which resulted in confusion between Hybrid and VT events more frequently than for LP events. We saw that the Hybrid events with a low frequency component that was much more energetic than the high frequency component tended to be classified as LP events. These results reinforce the idea that Hybrid events are a continuum between VT and LP events. Although the scores of the Hybrid classification were weaker than the scores of the other classes, this is still encouraging because these events are the most difficult for a human operator to recognize.

This new classification appears to be very effective, and allowed the automatic algorithm to distinguish the different earthquake classes more easily than the previous classification. Moreover, two new earthquake classes, as Hybrid earthquakes and LP monochromatic earthquakes (Tornillos), were introduced to better characterize the seismicity. There is a wealth of literature data and well-documented active volcanic sequences that have culminated in eruptive activity that underscore the fundamental roles that these Hybrid earthquakes and Tornillos earthquakes have to understand and track the complex nature and dynamics of seismogenic processes for an active volcano.

4.3. Feature performance

Selection of the features is a decisive step to obtain good classification results. Here, we were interested in the possibility to reduce the number of features, while keeping the most representative features, and also to determine which features were important for which class. It is also possible that some features were not optimal for the classification, if they were highly correlated. To keep the intrinsic physical properties of the features, we did not transform or project features in a new space (e.g., with principal component analysis). RF allows determination of the weight of each feature in a classification, using the impurity

score. The impurity score is related to the loss of accuracy when a feature is removed. The greater the importance of a feature in the classification, the higher the loss will be in the performance of the classification upon its removal. However, following this method, a feature that has a low importance is not automatically meaningless, as two features can be correlated, and only one will have a high score. More information about the feature importance can be found in Menze et al. (2009).

In Fig. 6, we show the weights of the 102 features for the binary classifications (i.e., each class against all of the other classes) and for the classification with all of the classes. We can see that the most important features changed depending on the class. The best features for recognizing all of the classes are the minimum of the signal (29), the index of central energy (6), the maximum energy (24), the rate of attack (19), the maximum signal (30) in the time domain, the mean kurtosis (9), the root mean square bandwidth (7), the standard deviation in the frequency domain, and the rate of decay (20) in the cepstral domain. These nine features had weights greater than 2% in the classification, and came from the three representation domains (i.e., temporal, spectral, cepstral domains), so they underscore the relevance of using these representation domains.

For the Hybrid class, the best features had lower weight compared to those of the best features obtained for the other classes, at 4.9% importance compared to 8.2%, 8.7%, 8.2%, and 7.5% importance for the VT class, the Nested class, the LP class, and the Tornillo class. Also, the best feature of the Hybrid class was less discriminant than the best feature of the other classes. This is consistent with this class being the most difficult to classify correctly.

Among the 10 best features of the least numerous classes, as the Nested class and the Tornillo class, there were only 3 and 1 of these features, respectively, in common with the 10 best features of the classification with all of the classes. This means that the most discriminant features of a class do not necessarily have great importance in the general classification, and this depends on the number of events in these classes. Conversely, 7 of the 10 best features of the VT class, 6 of the 10 best features of the LP class, and 5 of the 10 best features of the Hybrid class were among the 10 best features of the general classification. In particular, the maximum of energy (24) in the temporal domain was in the 10 best features of the VT, the LP, and the Hybrid classes, and also for the classification with all of the classes. This means that this feature is useful to distinguish these classes from the rest of the dataset, and also to discriminate between these three classes.

In Fig. 7, we show the mean overall accuracies obtained after 10 trials of the classification using only the best feature, and then adding the second best feature, and so on until every features has been added in the classification analysis. With the nine best features having an individual weight greater than 2%, the mean accuracy rapidly converged above 80% after the use of these features. To reach the mean accuracy score of 82.1% of the full set of 102 features, the 14 best features need to be used for the learning: after 14, adding new features did not change the score significantly.

This set of features has already been used by Malfante et al. (2018a, b) for Ubinas Volcano in Peru. Malfante et al. (2018a, b) obtained 84.4% accuracy with only the three best features, and needed the 13 best features to reach a score (90.3%) that was close to the accuracy obtained with all of the features (92.5%). Here, we compared the 13 best features obtained by Malfante et al. (2018a, b) with the 14 best features that we determined in our analysis. It is interesting to note that among the 13 best features of the Ubinas Volcano analysis by Malfante et al. (2018a, b), only one of their features was in our set of 14 best features. This was the mean of the signal (2) in the cepstral domain. This difference can be explained by the different type of seismic activity recorded at Ubinas Volcano compared to the seismic activity of La Soufrière de Guadeloupe. The seismicity of La Soufrière Volcano was dominated by VT and Hybrid events, whereas Ubinas Volcano was dominated by LP and Tremor events. It also suggests that the model we built is specific

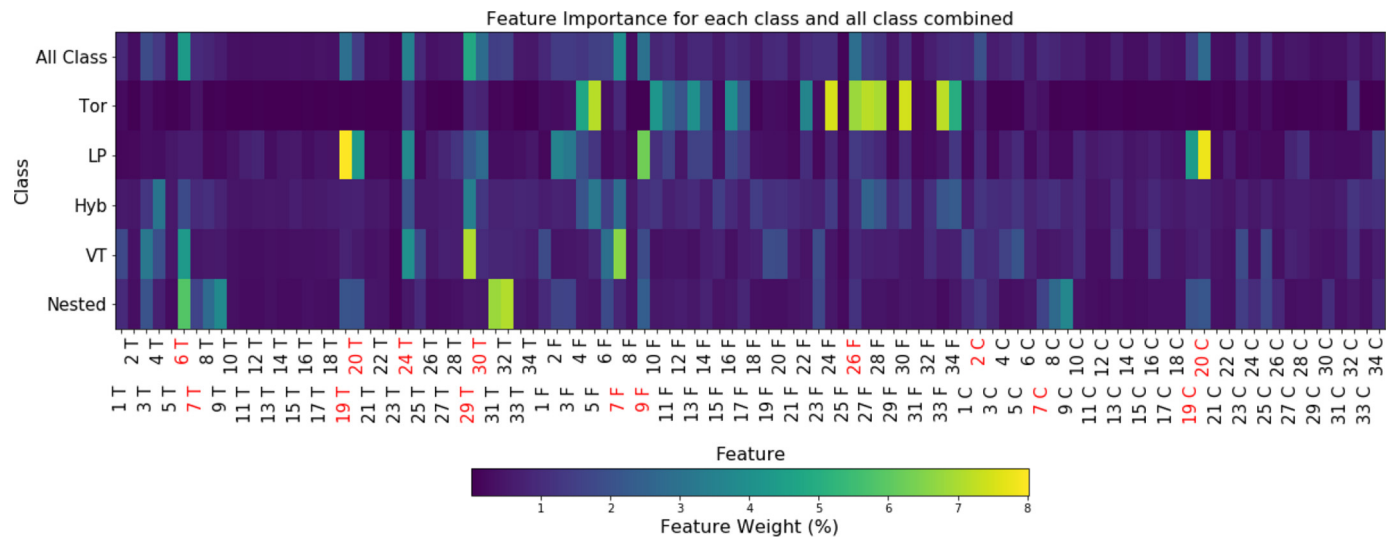


Fig. 6. Weights as percentages for all of the features, for each class, and for all classes combined for the volcanic seismicity detected at La Soufrière de Guadeloupe (January 1, 2013, to December 31, 2018; OVSG-IPGP) determined with the feature importance function in the scikit-learn library. Blue, poor importance in the classification; yellow, important for the classification; red, the 14 most useful features with which the classification is similar to the mean score obtained for the whole feature set.

to La Soufrière Volcano, or at least the kind of seismicity shown by La Soufrière Volcano. It would be interesting to test the performances of the features on a volcano with the same kind of seismicity as La Soufrière de Guadeloupe, that is characterized by a very active and extensive hydrothermal system. Another possible explanation for the difference in the best features might be that the 14 best features obtained at the TAG station for La Soufrière Volcano are more specific to the station position than to the volcano itself. On the basis of the complex internal structures of volcanoes, the TAG station might show strong and specific site effects that distort the signals (e.g., high attenuation, strong anisotropy, waveguide effects, high density contrasts). The very small number of events that passed the selection phase (i.e., only 845 of the 7194 detected by the OVSG met the quality criterion at station TAG) is probably an illustration of this phenomenon.

To overcome this issue, different approaches can be used. The first one, although not very efficient and time consuming, would be to apply the same single-station approach to other stations at La Soufrière Volcano and see whether the best features are identical for all of the individual stations. A second approach would be to use a multi-station approach, using the features computed from different stations to take into account the variability of the same signal all over the volcano, as in Maggi et al. (2017). These authors reached their best performance by using a combination of features from three stations. A third approach would be to use a multi-station approach of a higher level, with multi-station specific features. These features would be derived from signals

obtained from network-based analyses that involve data from multiple sensors, such as cross-correlations or higher-order cross-correlations. We could compute the features based on these cross-correlation in the three representation domains (i.e., temporal, spectral, cepstral domains).

5. Conclusions and prospects

We have applied an automatic classification method based on machine learning, and more specifically on a RF classifier, for the analysis of the volcanic seismicity of La Soufrière Volcano of Guadeloupe. The principle is based on the calculation of a large set of descriptors from the waveforms of the volcanic seismic events that satisfy the quality criteria. This set of descriptors, or features, is then used to train a RF classifier using all of the seismic events for their classification into the main classes of seismic signals that have causal links to specific dynamic volcanic processes in the system. After training of the model on a subset of the seismic signals from the catalog of volcanic seismic events built by the OVSG, we obtained a good classification rate of 73% on the subset of seismic signals that remained to be classified. However, after a review of the results class by class, we showed that the variability in the VT class was too broad, and that it caused distortion for the recognition of the two other classes. We therefore manually reviewed and relabeled all of the events, which resulted in the addition of two new classes that had been recognized by the observatory, but were not taken into

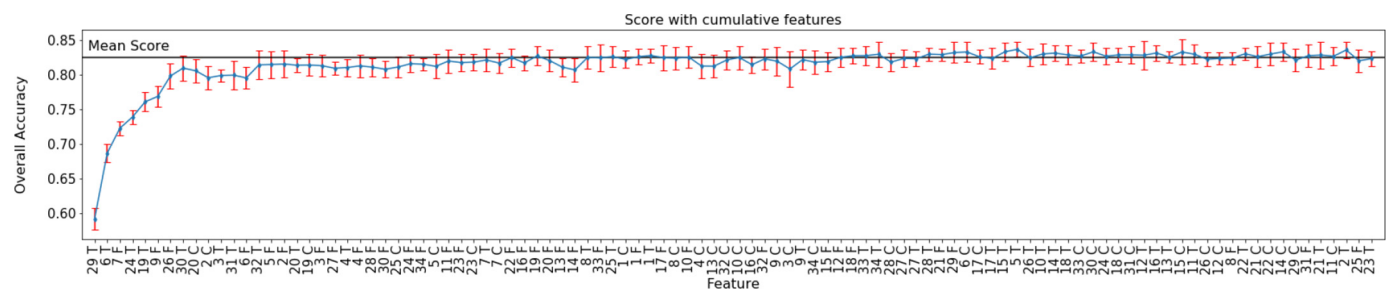


Fig. 7. Mean accuracies for each successive n^{th} best feature; learning rate = 50%; model trained using a random forest classifier ($n_{\text{estimators}} = 100$, $\text{criterion} = \text{'entropy'}$, $\text{bootstrap} = \text{False}$, $\text{class_weight} = \text{None}$), as mean score after 10 trials. Black line, mean score obtained with all 102 features after 10 trials, for the La Soufrière de Guadeloupe volcanic seismicity from January 1, 2013, to December 31, 2018 (OVSG-IPGP).

account during the daily classification protocol. After the introduction of these new Hybrid and monochromatic LP (Tornillo) classes, the performance of the automatic classification increased to 82.1%. Most of the errors made by the model were due to the confusion between Hybrid events and VTs and LPs. This result reinforces the idea that Hybrid events are a continuum between VTs and LPs. The relatively poor scores obtained with the initial automatic classifier highlighted the presence of inaccuracies in the classification. The much improved accuracy that we achieved after refining the classification shows that the machine learning helps to build a robust catalog of volcanic earthquake signals that are classified into process-related classes. We explored the importance of the features in our model to determine which features of the full set of 102 performed best for the recognition of these classes of volcanic earthquakes at La Soufrière de Guadeloupe that characterized the period analyzed (*i.e.*, January 1, 2013, to December 31, 2018). We showed which features are the most useful depending on the class we want to recognize for La Soufrière Volcano. We also showed that with only a subset of the features (*i.e.*, 14 of 102), we can obtain a substantially similar score. Finally, we compared our best features for La Soufrière Volcano with the best features presented by Malfante et al. (2018a, b) for Ubinas Volcano, and showed that the most important features are different (except one) between these two volcanoes. Therefore, it is important to maintain this whole feature set to test the method on another volcano.

In the future, we will apply this methodology to data obtained from the complete volcanic seismic network for La Soufrière, which is composed of about 20 sensors, to improve the accuracy. Furthermore, it will be interesting to repeat the exploration of these features on other stations or for other volcanoes that have activities that are similar to that of La Soufrière de Guadeloupe, to see if the same features are among the most important, or at least if the same features are useful for the same classes of volcanic seismic events. More importantly, by training a classification model on a specific time period for La Soufrière Volcano or any volcano and applying this to a subsequent time period, it will be possible to detect any changes in the nature, dynamics, and/or patterns of seismicity by quantifying and following with time the progressive or rapid loss in accuracy of the classification algorithm. This approach was reported by Hibert et al. (2017) at Piton de la Fournaise Volcano, and by Malfante et al. (2018a, b) at Ubinas Volcano. We will test this at La Soufrière Volcano to monitor for any changes in the pattern of seismicity that preceded the major unrest phase of January to April 2018 (Moretti et al., 2020) with the seismicity that has been recorded from after April 2018 that showed characteristic swarm activity of numerous low-energy events (OVSG-IPGP, 1999–2020). It will be possible to test whether all swarms show similar patterns, or whether, as qualitatively observed, some swarms are dominated by certain classes of events, and also whether this pattern changes within the swarm.

In this paper, we have proposed a new machine learning classification scheme that is based on direct human examination and analysis of each signal. It is fundamental to test the performance of a supervised classification using the machine learning algorithms to discriminate the different signals. Our final goal is to implement this model in the OVSG in Guadeloupe, and also to test it for other active volcanoes that are monitored by IPGP, such as Montagne Pelée in Martinique (OVSM-IPGP, 2020), and even in the context of the major submarine eruption offshore of Mayotte that has been ongoing since 2018 (REVOSIMA, 2019–2020).

Our results show that these machine learning tools can represent a major component of any volcanic monitoring system for an active volcano. By providing rapidly significant real-time insight into the nature, style, and patterns of seismicity at active volcanoes in a state of unrest while optimizing human resources, these tools will allow scientists to gain access to new knowledge to understand and track the complex nature and dynamics of the nonlinear processes of active volcanoes. Hence, this approach contributes to improved forecasts of the likelihood of future eruptive activity, and to improve the societal responses to

crises and risk mitigation in a context of increasing exposed assets and vulnerability.

Declaration of Competing Interest

None.

Acknowledgements

The code used in this work is available at <https://github.com/malfante/AAA>.

We would like to acknowledge the numerous staff members that have identified and localized seismicity at La Soufrière Volcano over the last three decades: S. Acounis, D. Amorèse, C. Antenor-Habazac, S. Bazin, D. Bertil, M.-P. Bouin, J. Delacaze, S. Deroussi, T. Didier, A. Domiquin, M. Feuillard, B. Figaro, G. Hammouya, D. Jessop, T. G. Kitou, C. Lambert, F. Massin, S. Moune, J.-Y. Nadeau, A. Nercessian, and J. Pierre, V. Robert. We also thank X. Béguin, A. Bosson, D. Lafon, Didier Mallarino, P. Boissier, C. Brunet, and A. Ali Fahmi for computer engineering, code development, and data management for the WebObs platform.

The Soufrière de Guadeloupe full waveform seismicity data is available from the IPGP Data Centre (<https://www.ipgp.fr/fr/centredonnees/centre-de-donnees-ipgp>) and <http://volobsis.ipgp.fr/> as well as from RESIF (<https://www.resif.fr/>). Seismic catalogs are available upon reasonable request at: infos@ovsg.univ-ag.fr

We thank the IPGP for general funding to *Observatoires Volcanologiques et Sismologiques* (OVS), INSU-CNRS for the funding provided by *Service National d'Observation en Volcanologie* (SNOV), and the *Ministère pour la Transition Ecologique et Solidaire* (MTES) for financial support, and the FEDER INTERREG Caraïbes program for funding the West Indies regional seismic and geodetic monitoring network with the VSAT satellite communications. Part of G. Ucciani's postdoc was funded by the project *Vers la Plateforme Régionale de Surveillance Tellurique du futur* (PREST), which was co-funded by INTERREG Caraïbes for the European Regional Development Fund. This study was also supported by the European Union Horizon 2020 Research and Innovation Programme, under grant agreement No 731070 (EUROVOLC project). This study contributes to IdEx *Université de Paris* ANR-18-IDEX-0001.

References

- Allen, R.V., 1978. Automatic earthquake recognition and timing from single traces. *Bull. Seismol. Soc. Am.* 68 (5), 1521–1532.
- Bean, C., Lokmer, I., O'Brien, G., 2008. Influence of near-surface volcanic structure on long-period seismic signals and on moment tensor inversions: simulated examples from Mount Etna. *J. Geophys. Res. Solid Earth* 113 (B8).
- Bean, C.J., De Barros, L., Lokmer, I., Métaxian, J.P., O'Brien, G., Murphy, S., 2014. Long-period seismicity in the shallow volcanic edifice formed from slow-rupture earthquakes. *Nat. Geosci.* 7 (1), 71–75.
- Beauducel, F., Lafon, D., Béguin, X., Saurel, J.M., Bosson, A., Mallarino, D., Nercessian, A., 2020. WebObs: the volcano observatories missing link between research and real-time monitoring. *FrEaS* 8, 48.
- Benítez, M.C., Ramírez, J., Segura, J.C., Ibanez, J.M., Almendros, J., García-Yeguas, A., Cortes, G., 2006. Continuous HMM-based seismic-event classification at Deception Island, Antarctica. *IEEE Trans. Geosci. Remote Sens.* 45 (1), 138–146.
- Boudon, G., Dagain, J., Semet, M., Westercamp, D., 1988. Carte géologique au 1/20000e du Massif volcanique de la Soufriere, 1 sheet Notice explicative de la carte géologique au 1/20000e du Massif volcanique de la Soufriere-Carte Géologique. BRGM-CNRS-DRM-IPGP, BRGM, Orléans, pp. 1–43.
- Boudon, G., Komorowski, J.C., Villemant, B., Semet, M.P., 2008. A new scenario for the last magmatic eruption of La Soufrière of Guadeloupe (Lesser Antilles) in 1530 AD evidence from stratigraphy radiocarbon dating and magmatic evolution of erupted products. *J. Volcanol. Geotherm. Res.* 178 (3), 474–490.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32.
- Bueno, A., Benítez, C., De Angelis, S., Moreno, A.D., Ibáñez, J.M., 2019. Volcano-Seismic transfer Learning and uncertainty Quantification with Bayesian Neural Networks. *IEEE Trans. Geosci. Remote Sens.* 58 (2), 892–902.
- Chouet, B., 1988. Resonance of a fluid-driven crack: Radiation properties and implications for the source of long-period events and harmonic tremor. *J. Geophys. Res. Solid Earth* 93 (B5), 4375–4400.
- Chouet, B.A., Matoza, R.S., 2013. A multi-decadal view of seismic methods for detecting precursors of magma movement and eruption. *J. Volcanol. Geotherm. Res.* 252, 108–175.

- Curilem, M., de Mello, R.F., Huenupan, F., San Martin, C., Franco, L., Hernández, E., Rios, R.A., 2018. Discriminating seismic events of the Llaima volcano (Chile) based on spectrogram cross-correlations. *J. Volcanol. Geotherm. Res.* 367, 63–78.
- Fazio, M., Alparone, S., Benson, P.M., Cannata, A., Vinciguerra, S., 2019. Genesis and mechanisms controlling tornillo seismo-volcanic events in volcanic areas. *Sci. Rep.* 9 (1), 1–11.
- Gómez M, Torres C, R. A. (1997). Unusual low-frequency volcanic seismic events with slowly decaying coda waves observed at Galeras and other volcanoes. *J. Volcanol. Geotherm. Res.*, 77(1–4), 173–193.
- Hammer, C., Ohrnberger, M., Faeh, D., 2013. Classifying seismic waveforms from scratch: a case study in the alpine environment. *Geophys. J. Int.* 192 (1), 425–439.
- Harrington, R.M., Brodsky, E.E., 2007. Volcanic hybrid earthquakes that are brittle-failure events. *Geophys. Res. Lett.* 34 (6).
- Hibert, C., Mangeney, A., Grandjean, G., Baillard, C., Rivet, D., Shapero, N.M., Crawford, W., 2014. Automated identification, location, and volume estimation of rockfalls at Piton de la Fournaise volcano. *J. Geophys. Res. Earth Surf.* 119 (5), 1082–1105.
- Hibert, C., Provost, F., Malet, J.P., Maggi, A., Stumpf, A., Ferrazzini, V., 2017. Automatic identification of rockfalls and volcano-tectonic earthquakes at the Piton de la Fournaise volcano using a Random Forest algorithm. *J. Volcanol. Geotherm. Res.* 340, 130–142.
- Hincks, T.K., Komorowski, J.C., Sparks, S.R., Aspinall, W.P., 2014. Retrospective analysis of uncertain eruption precursors at La Soufrière volcano, Guadeloupe, 1975–77: volcanic hazard assessment using a Bayesian Belief Network approach. *J. Appl. Volcanol.* 3 (1), 3.
- Jessop, D., Moretti, R., Moune, S., Bonifacie, M., Burtin, A., De Chabalier, J.B., Didier, T., 2019. Les risques volcaniques au XXI^e siècle: surveiller et prévoir; l'exemple de l'OVSG. *Géologues: revue de l'Union française des géologues*.
- Komorowski, J.C., Boudon, G., Semet, M.P., Beauducel, F., Anténor-Habazac, V., Bazin, S., Cheminée, J.L., 2005. Volcanic Atlas of the Lesser Antilles. University of the West Indies, pp. 63–100.
- Komorowski, J.C., Morin, J., Jenkins, S., Kelman, I., 2016. Challenges of Volcanic Crises on Small Islands States. In *Observing the Volcano World* (Pp. 353–371) Springer, Cham.
- Konstantinou, K.I., 2015. Tornillos modeled as self-oscillations of fluid filling a cavity: Application to the 1992–1993 activity at Galeras volcano, Colombia. *Phys. Earth Planet. Inter.* 238, 23–33.
- Lahr, J.C., Chouet, B.A., Stephens, C.D., Power, J.A., Page, R.A., 1994. Earthquake classification, location, and error analysis in a volcanic environment: Implications for the magmatic system of the 1989–1990 eruptions at Redoubt Volcano, Alaska. *J. Volcanol. Geotherm. Res.* 62 (1–4), 137–151.
- Langer, H., Falsaperla, S., Powell, T., Thompson, G., 2006. Automatic classification and a posteriori analysis of seismic event identification at Soufriere Hills volcano, Montserrat. *J. Volcanol. Geotherm. Res.* 153 (1–2), 1–10.
- Langet, N., Maggi, A., Michelini, A., Brenguier, F., 2014. Continuous Kurtosis-based Migration for Seismic Event Detection and Location, with Application to Piton de la Fournaise Volcano, La Réunion. *Bull. Seismol. Soc. Am.* 104 (1), 229–246.
- Maggi, A., Ferrazzini, V., Hibert, C., Beauducel, F., Boissier, P., Amemoutou, A., 2017. Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano. *Seismol. Res. Lett.* 88 (3), 878–891.
- Malfante, M., Dalla Mura, M., Mars, J.I., Métaxian, J.P., Macedo, O., Inza, A., 2018a. Automatic classification of volcano seismic signatures. *J. Geophys. Res. Solid Earth* 123 (12), 10–645.
- Malfante, M., Dalla Mura, M., Métaxian, J.P., Mars, J.I., Macedo, O., Inza, A., 2018b. Machine learning for volcano-seismic signals: challenges and perspectives. *IEEE Signal Process. Mag.* 35 (2), 20–30.
- McNutt, S.R., 2005. Volcanic seismology. *Annu. Rev. Earth Planet. Sci.* 32, 461–491.
- McNutt, S.R., Roman, D.C., 2015. Volcanic seismicity. *The Encyclopedia of Volcanoes*. Academic Press, pp. 1011–1034.
- Menze, B.H., Kelm, B.M., Masuch, R., Himmelreich, U., Bachert, P., Petrich, W., Hamprecht, F.A., 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinforma.* 10 (1), 213.
- Miller, A.D., Stewart, R.C., White, R.A., Luckett, R., Baptie, B.J., Aspinall, W.P., Voight, B., 1998. Seismicity associated with dome growth and collapse at the Soufriere Hills Volcano, Montserrat. *Geophys. Res. Lett.* 25 (18), 3401–3404.
- Moretti, R., Komorowski, J.C., Ucciani, G., Moune, S., Jessop, D., de Chabalier, J.B., Deroussi, S., 2020. The 2018 unrest phase at La Soufrière of Guadeloupe (French West Indies) andesitic volcano: Scrutiny of a failed but prodromal phreatic eruption. *J. Volcanol. Geotherm. Res.* 106769.
- Neuberg, J., Luckett, R., Baptie, B., Olsen, K., 2000. Models of tremor and low-frequency earthquake swarms on Montserrat. *J. Volcanol. Geotherm. Res.* 101 (1–2), 83–104.
- Provost, F., Hibert, C., Malet, J.P., Stumpf, A., Doubre, C., 2016, April. Automatic classification of endogenous seismic sources within a landslide body using random forest algorithm. *EGU Gen. Assembly Conf. Abstracts* 18.
- Quinlan, J.R., 1986. Induction of decision trees. *Mach. Learn.* 1 (1), 81–106.
- Titos, M., Bueno, A., Garcia, L., Benitez, C., 2018. A deep neural networks approach to automatic recognition systems for volcano-seismic events. *IEEE J. Selected Topics Appl. Earth Observ. Remote Sens.* 11 (5), 1533–1544.
- Ucciani, G., 2015. Caractérisation spatiale et temporelle de la sismicité volcanique de la Soufrière de Guadeloupe: relation avec le système hydrothermal. Doctoral dissertation. Sorbonne Paris Cité.
- Villemant, B., Komorowski, J.C., Dessert, C., Michel, A., Crispi, O., Hammouya, G., De Chabalier, J.B., 2014. Evidence for a new shallow magma intrusion at La Soufrière of Guadeloupe (Lesser Antilles): insights from long-term geochemical monitoring of halogen-rich hydrothermal fluids. *J. Volcanol. Geotherm. Res.* 285, 247–277.
- Waite, G.P., Chouet, B.A., Dawson, P.B., 2008. Eruption dynamics at Mount St. Helens imaged from broadband seismic waveforms: Interaction of the shallow magmatic and hydrothermal systems. *J. Geophys. Res. Solid Earth* 113 (B2).
- White, R.A., Miller, A.D., Lynch, L., Power, J., 1998. Observations of hybrid seismic events at Soufriere Hills volcano, Montserrat: July 1995 to September 1996. *Geophys. Res. Lett.* 25 (19), 3657–3660.