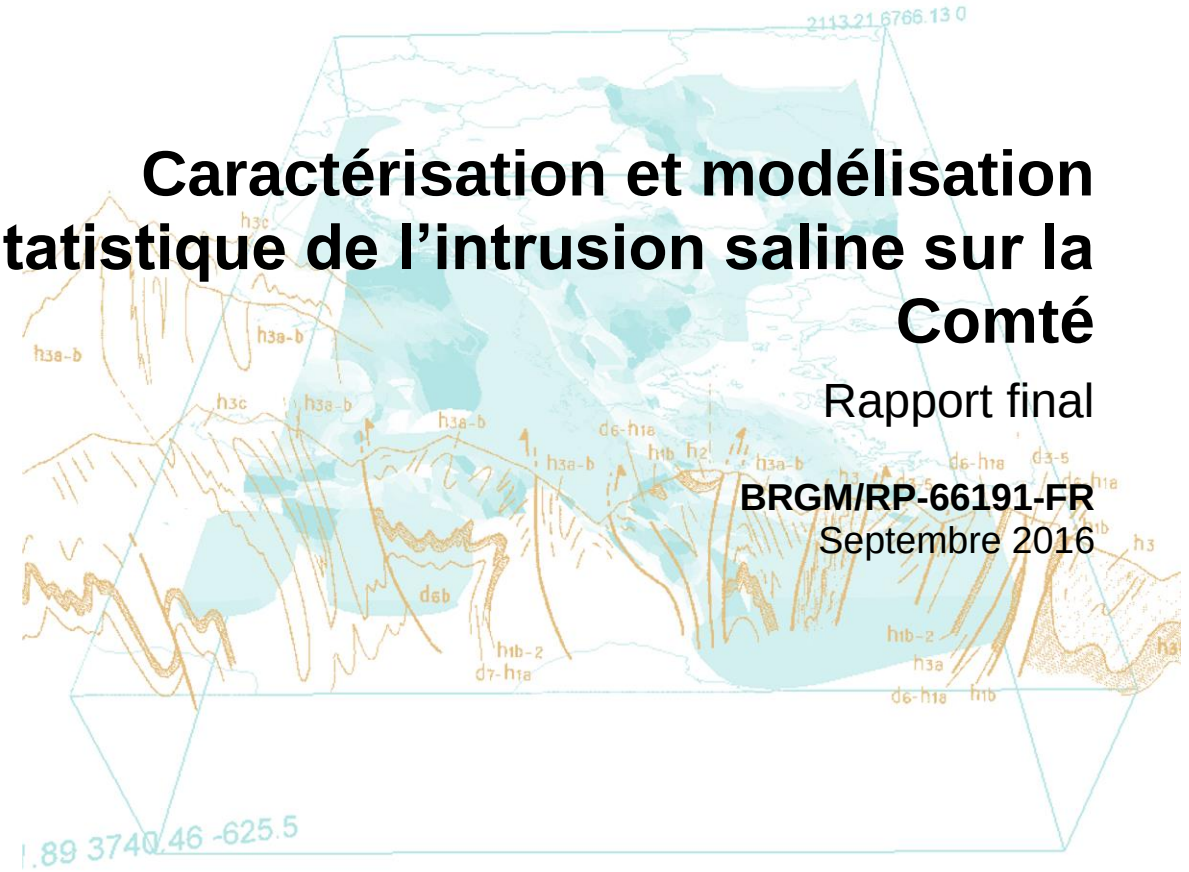
Document public



Caractérisation et modélisation statistique de l'intrusion saline sur la Comté

Rapport final

BRGM/RP-66191-FR
Septembre 2016

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Direction de l'Environnement, de l'Aménagement et du Logement
GUYANE

Géosciences pour une Terre durable
brgm

Document public

Caractérisation de l'intrusion saline sur la Comté et analyse statistique de ses conditions

Rapport final

BRGM/RP-66191-FR
Septembre 2016

Étude réalisée dans le cadre des projets
de Service public du BRGM AP15GUY024

J. Rohmer et N. Brisset
Avec la collaboration de
B. Joseph, L. Bechelen et S. Drouhin

Vérificateur :	**Approbateur :**
Nom : Olivier Douez	Nom : Verneyre Laure
Fonction : Responsable Scientifique	Fonction: Directrice
Date : 23/09/2016	Date : 06/10/2016
Signature :	

Le système de management de la qualité et de l'environnement est certifié par AFNOR selon les normes ISO 9001 et ISO 14001.

Liberté • Égalité • Fraternité
RÉPUBLIQUE FRANÇAISE
Direction de l'Environnement, de l'Aménagement et du Logement
GUYANE

Géosciences pour une Terre durable
brgm

Mots-clés : Intrusion saline, conductivité, corrélation, SVM, Roura, La Comté

En bibliographie, ce rapport sera cité de la façon suivante :

Rohmer J., Brisset N. (2016) – Caractérisation de l'intrusion saline sur la Comté et analyse statistique de ses conditions. Rapport final. BRGM/RP-66191-FR, 41 p., 26 ill.

Synthèse

Le captage de la Comté permet d'assurer deux tiers de l'alimentation en eau potable de la Communauté d'Agglomération du Centre Littoral (CACL) avec des débits de pompage de l'ordre de 1200 m^3/h. Bien que situé à 40 km de l'estuaire du fleuve, il est parfois sujet, durant les saisons sèches marquées, à des intrusions d'eau dont la salinité dépasse la limite de qualité des eaux brutes pour la production d'eau destinée à la consommation humaine de 1000 µS/cm[1].

Le suivi de la conductivité en place sur la Comté depuis 2007 n'est actuellement pas conçu pour assurer une prévision sur le moyen et le court terme. Un modèle de vigilance permettant d'appréhender le phénomène à échéances multiples reste à définir.

La connaissance de la forme du front salin étant déterminante pour établir des modèles conceptuels ou hydrodynamiques, et donc pour améliorer la gestion de l'unité de potabilisation et du captage associé, des campagnes de mesures de la conductivité de l'eau étaient nécessaires. Des sections transversales composées de plusieurs verticales de mesures ont ainsi été réalisées pendant la saison sèche 2015 sur cinq missions. Les résultats obtenus ont permis de mettre en évidence les points suivants :

- Lorsque le débit de la rivière est très faible (observé durant l'étude en octobre, novembre, janvier) et que la hauteur d'eau à la pleine mer est élevée (octobre et janvier), on observe un front salin relativement homogène sur les sections de mesures, autant sur la hauteur que d'une rive à l'autre. Il n'y a ainsi pas d'allure biseautée observée dans ce contexte, qui est celui le plus problématique car favorisant la plus forte remonté du front.

- Lorsque le cours d'eau a un débit relativement plus élevé (observé en septembre et décembre), une stratification de la conductivité se dessine sur les verticales de mesures. Ce phénomène est encore plus prononcé lors d'un faible coefficient de marée comme en septembre, où la salinité peut tripler entre la surface et le fond de la rivière

Ces analyses montrent que dans le contexte critique pour la gestion du captage (débit faible et coefficient de marée élevé), le front a une allure pseudo-verticale. Ces observations permettent d'orienter le choix futur d'un modèle numérique décrivant la distance du front salin au captage en fonction de l'évolution du débit du cours d'eau et des niveaux marins.

Afin de mieux comprendre le phénomène et les facteurs caractéristiques, une étude statistique des principales variables mises en jeu a été entreprise. Ainsi, une étude de la corrélation temporelle entre la conductivité électrique σ et deux facteurs explicatifs, à savoir le niveau d'eau donné par le marégraphe de l'île Royale et le débit de la Comté Q mesuré à Saut Bief a été réalisée. Puis, un modèle statistique de type SVM a été construit afin de prédire l'occurrence d'une intrusion saline mise en évidence par un pic de conductivité électrique.

La capacité de ce modèle prédictif a été étudiée via un exercice de validation basée sur la sélection aléatoire des données d'apprentissage du modèle et de validation : cela a confirmé la bonne performance du modèle SVM. Le même exercice a été effectué pour prédire si la durée du pic dépasse 2 heures et s'est révélé également concluant. Cet aspect de durée de l'évènement est d'une grande importance pour le gestionnaire.

[1] Arrêté du 11 janvier 2007 relatif aux limites et références de qualité des eaux brutes et des eaux destinées à la consommation humaine

Sommaire

Liste des illustrations

1. Introduction

1.1. CONTEXTE

Depuis 2007, le suivi de la remontée du front salin réalisé par le BRGM a pour objectif d'anticiper une éventuelle non-conformité de la qualité de l'eau prélevée sur la Comté pour la production d'eau potable et de fournir, par l'acquisition de données régulières, des tendances d'évolution du front salé, notamment au cours de la saison sèche. Si ce suivi permet d'anticiper d'éventuelle dégradation de l'eau prélevée, il n'est pas aujourd'hui conçu pour assurer une prévision sur le moyen et le court terme. La surveillance porte par ailleurs sur le respect de la norme de 1 000 µS/cm fixée pour le prélèvement d'eaux brutes destinées à la consommation humaine. Des seuils d'alertes intermédiaires et un modèle de vigilance restent à définir pour anticiper suffisamment la mise en place d'actions de prévention.

Le captage de la Comté permet d'assurer les deux tiers de l'alimentation en eau potable de la Communauté d'Agglomération Centre Littoral avec des débits de pompage de l'ordre de 1200 m^3/h. L'installation se situe à environ 40 km (Illustration 1), en longueur de cours d'eau linéaire, de l'estuaire du Mahury, fleuve prolongeant la rivière.

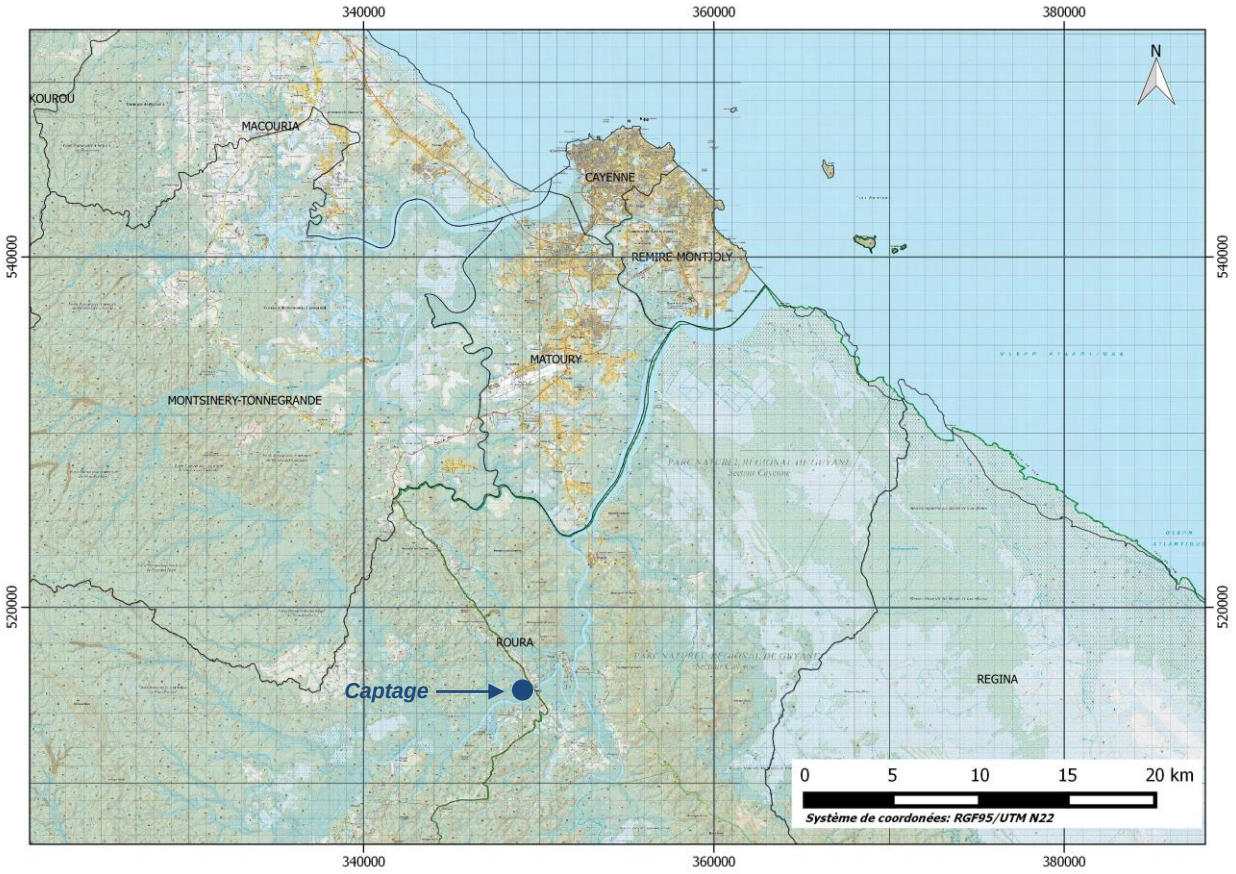


Illustration 1 : Localisation du captage d'eau potable sur la Comté

1.2. OBJECTIFS

Afin de mieux anticiper d'éventuelles contaminations du captage de la Comté par des eaux saumâtres, le BRGM propose de caractériser et modéliser l'intrusion saline dans la Comté.

Les objectifs sont, d'une part d'établir une échelle de risque, en croisant les enjeux et l'aléa liés au phénomène (ce dernier sera décrit et caractérisé par les variables qui le représentent le plus et d'autre part de définir un modèle de vigilance permettant d'appréhender le phénomène à échéances multiples. Un premier modèle probabiliste, basé sur les données de débit du cours d'eau et de marée, est ainsi développé pour débuter ce travail de vigilance.

A partir de la caractérisation du front salin, un modèle numérique sera développé par la suite. Ce modèle permettra à terme de fournir la distance à l'estuaire (xl) du front salé en fonction : du débit de la Comté, du coefficient de la marée et de la forme de l'estuaire. Ces relations mathématiques seront ensuite testées sur les données acquises depuis 2007 pour validation. Couplée à l'étude statistique, des indices de vigilance pourront être définis afin de gérer les situations de crise en saison sèche marquée.

A la suite de ces travaux, en collaboration avec la cellule de veille hydrologique, un schéma de gestion sera proposé et mis à disposition de l'exploitant. Ce schéma lui permettra d'anticiper les problèmes d'intrusion d'eau salée et de gérer de façon optimale l'alimentation en eau du territoire de la CACL durant la saison sèche.

2. Caractérisation du front salin

2.1. METHODOLOGIE

Un des buts premiers de cette étude fut de connaître la forme que pouvait avoir le front salin de 1000 µS/cm. Afin de gérer correctement les infrastructures impactées par ce phénomène, il est en effet nécessaire de savoir si la conductivité mesurée en sub-surface, ce qui se fait de façon usuelle, correspond à celle au cœur et au fond du fleuve.

Au sein d'un volume d'eau, si l'on mélange sans turbulences des eaux douces et des eaux salées, naturellement plus denses, il se créé une stratification naturelle de la salinité, avec de l'eau plus conductrice vers le fond du volume d'eau. Cela se traduit, lors d'une remonté d'eau marine dans un cours d'eau, par une forme biseauté de l'intrusion d'eau saline. Le débit du cours d'eau ainsi que le coefficient de marée vont contribuer à accentuer ou limiter ce phénomène comme on peut le voir dans l'Illustration 2 ci-dessous (*Savenije 2012*).

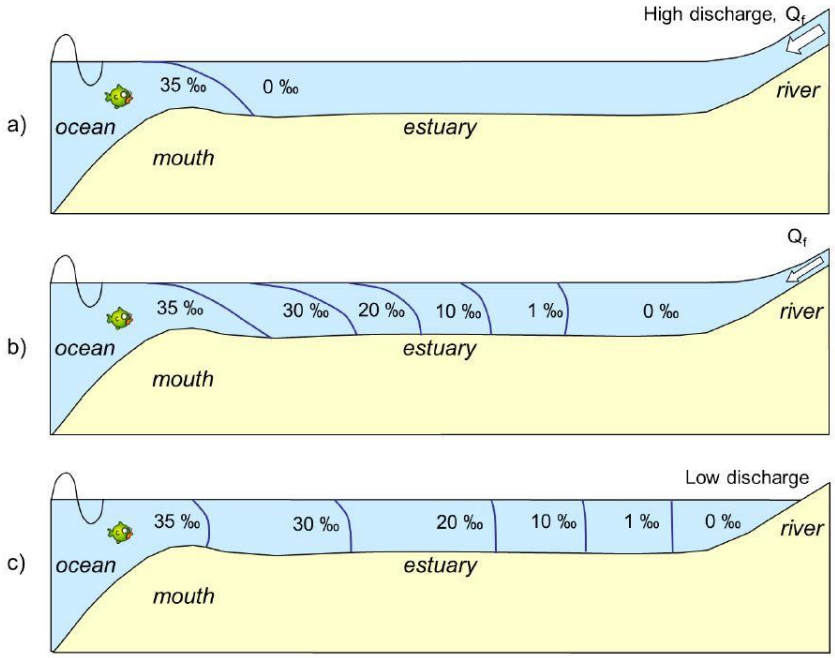


Illustration 2 : Distribution longitudinale de la salinité pour un estuaire avec des eaux stratifiées (a), partiellement mélangées (b) et bien mélangées (c),. (Savenije 2012)

Ce schéma montre que d'une façon générale, en plus d'un décalage du front vers l'aval, plus l'intrusion marine est faible face au volume provenant du cours d'eau, plus le phénomène de stratification, et donc de front en forme de biseau, sera accentué. Le littoral guyanais étant sujet à une forte amplitude de la marée au cours de l'année (hauteur d'eau de 0,7 à 3,7 m sur l'île Royale), l'étude s'est étalée sur plusieurs journées de terrain afin d'observer le phénomène à divers coefficients de marée durant la saison sèche.

Afin de caractériser la forme du front, des verticales de mesures de conductivité ont été réalisées suivant des sections allant d'une rive à l'autre. Pour chacune des missions, des mesures de surface ont d'abord servis à déterminer approximativement la position du front salin de 1000 µS/cm, puis des verticales de mesures espacées tous les 15 à 20 mètres avec des mesures à chaque mètre en profondeur ont été effectuées (Illustration 3 : Principe des

verticales de mesures). Une telle répartition des points de mesure permet non seulement d'avoir une idée de la spatialisation de la salinité de la surface vers le fond sur l'ensemble d'une section du cours d'eau mais également de mettre en lumière les hétérogénéités d'une rive à l'autre.

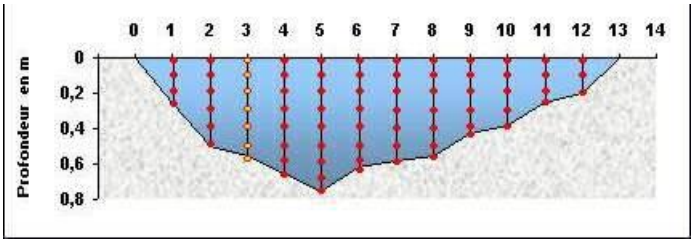


Illustration 3 : Principe des verticales de mesures

La réalisation de sections de mesure d'une rive à l'autre nécessite de stabiliser l'embarcation au niveau de chaque nouvelle verticale de mesure. Une ancre, suffisamment lourde pour empêcher le déplacement du bateau et assez légère pour être remontée le plus rapidement possible, était donc indispensable pour arriver à cet objectif.

Le matériel utilisé pour les mesures de conductivité est une sonde *WTW 1970i* composée d'un câble de 20 m gradué chaque mètre. Afin d'assurer des mesures le long de lignes les plus verticales possible, un poids de 600 g a permis de lester la sonde (Illustration 4). Pour chaque verticale, la première mesure démarre à un mètre sous la surface de l'eau et la dernière mesure a été réalisée au niveau du fond, permettant par la même occasion d'avoir une idée du profil transversale du fond du cours d'eau.

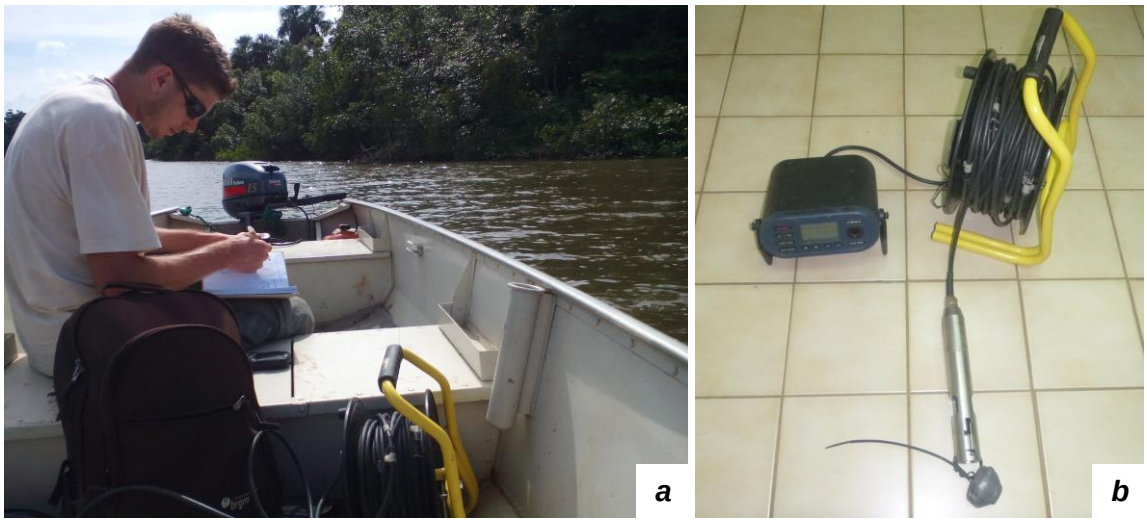


Illustration 4 : Opération de mesure de la conductivité (a) à l'aide d'une sonde lestée (b)

Les phénomènes d'intrusion saline dans le fleuve Comté survenant durant les périodes de hautes eaux, les missions se sont déroulées chaque mois de la saison sèche juste avant, et pendant, l'heure de hauteur d'eau maximale. L'Illustration 5 montre la position des différentes sections de mesures juste avant et au moment de l'étale le jour de la mission. La situation la plus critique des cinq missions a été observée le 28 octobre 2016, où le maximum de la remonté du front salin (1000 µs/cm) est arrivé à moins de 500 m du captage.

Au-delà des variations de conductivité au sein d'une section il a été constaté que l'étale, le moment où le niveau est au plus haut et le courant est nulle, est beaucoup plus courte lors des

remontées moins importantes du front. De fait, les missions de septembre et novembre 2015 ne se composent que d'une seule section, le front étant plus loin à aller chercher et l'eau du fleuve reprenant son sens d'écoulement normale quelques instants après le moment de plus haute eaux, ne permettant pas une section de mesures à ce moment particulier. L'Illustration 6 expose la hauteur d'eau, donnée par le marégraphe de l'île royale, au moment de la pleine mer de chacune des missions ainsi que le débit journalier de la rivière, mesuré à Saut Bief, en amont de la zone influencée par les marées.

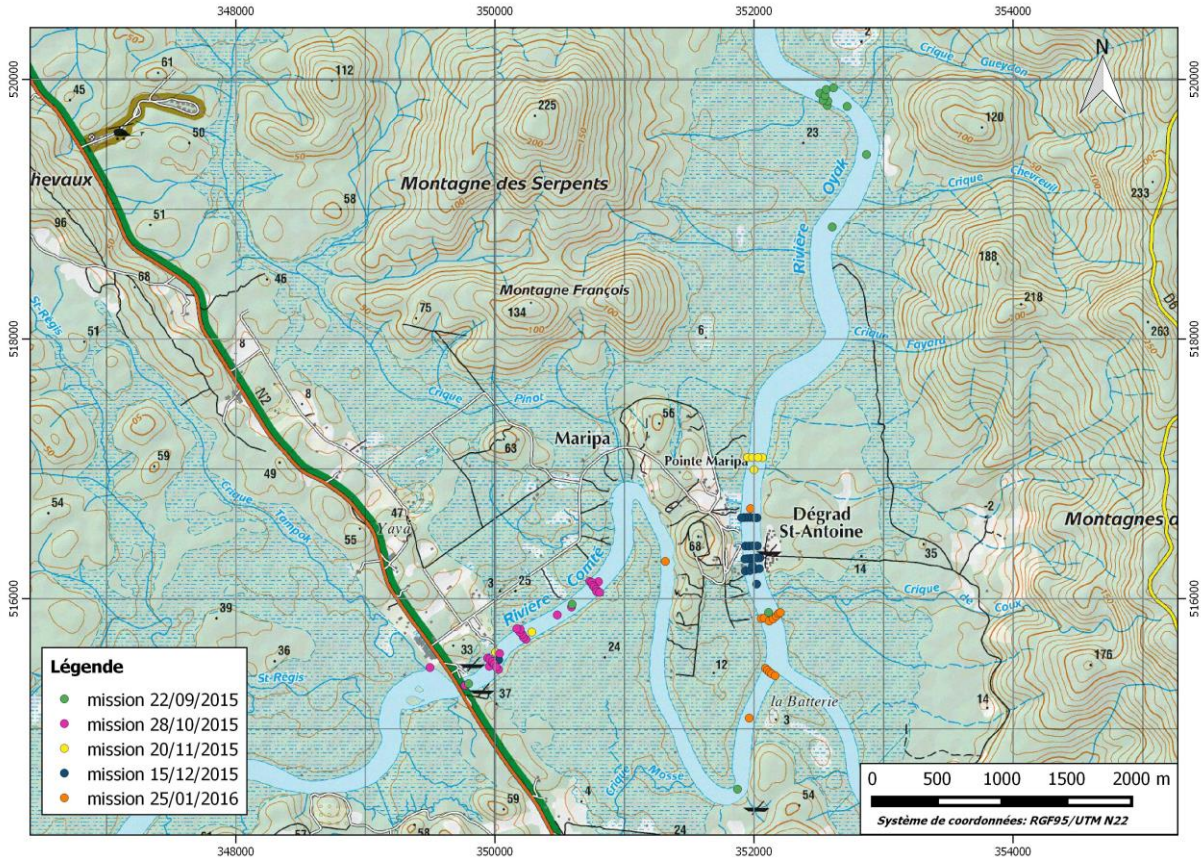


Illustration 5 : Position des verticales de mesure sur l'ensemble des cinq missions

Date	Hauteur de la pleine mer aux îles du Salut (en m)	Débit à Saut Bief (en m3/j)
22/09/2015	2,57	32 800
28/10/2015	3,64	17 600
20/11/2015	2,88	16 000
15/12/2015	2,99	24 400
25/01/2016	3,27	14 700

Illustration 6 : Etat de la marée et du débit du fleuve le jour des missions

2.2. RESULTATS

Une représentation en deux dimensions, avec un gradient de couleur, de l'ensemble des points de mesure de chaque section permet d'apprécier la variation de la conductivité de la lame d'eau, à la fois en profondeur mais aussi sur la largeur du cours d'eau (Illustration 7). Comme les sections doivent être réalisées rapidement, afin qu'elles soient les plus représentatives possibles de la lame d'eau, le nombre de verticales a été adapté en fonction du contexte géomorphologique de la rivière et de la dynamique de la remonté du front salin.

Tous les profils présentés ci-dessous sont repositionnés de façon à ce que la rive gauche soit du côté gauche de l'illustration et la rive droite du côté droit. Lorsque plusieurs profils de mesures ont été réalisés lors d'une même mission, les premiers sont ceux les plus en aval.

Mission du 22 septembre 2015

Lors de la première mission de caractérisation du front salin, la saison sèche démarrait et la hauteur d'eau était de 2,57 m aux îles du Salut, ce qui est une valeur faible pour la pleine mer en cet endroit. Il fut nécessaire de remonter la Comté presque jusqu'au bourg de Roura pour arriver au niveau du front salin des 1000 µS/cm, avec des mesures régulières en surface au milieu de la rivière. La section réalisée à ce niveau a mis en lumière les phénomènes suivants : une forte variation de conductivité, de 2000 à 600 µS/cm existe de la rive gauche à la rive droite ainsi qu'une forte stratification de la salinité, avec des valeurs mesurées allant de 860 à 1960 µS/cm entre -1 m et -13 m sur la seconde verticale de la section.

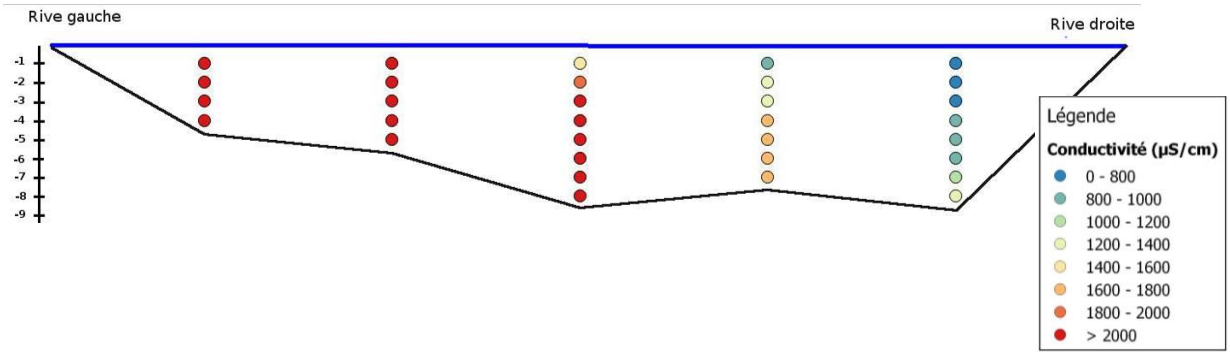


Illustration 7 : Répartition de la conductivité sur la section S1 réalisée le 22 septembre 2015

Ces mesures traduisent non seulement une forme biseautée de l'intrusion saline mais également un front non homogène d'une rive à l'autre. Ce constat nous a montré l'intérêt de réaliser des mesures à différentes profondeurs ainsi que sur toute la largeur du cours d'eau. Les mesures de conductivité habituellement réalisées en sub-surface sur la Comté, ne représentent ainsi pas forcément la conductivité globale de la rivière.

Mission du 28 octobre 2015

La mission du mois d'octobre 2015 s'est déroulée durant un très fort coefficient de marée, avec une hauteur de 3,64 m durant la pleine mer. Parallèlement, le débit de la rivière Comté a nettement diminué à cette période avec un débit presque deux fois inférieur à celui de la précédente mission. Ce jour-là le front salin a pu remonter à moins de 500 m du captage (Illustration 5). Les sections de S2 à S4 ont été réalisées de l'aval vers l'amont (Illustration 8 a, b et c) avant l'étale.

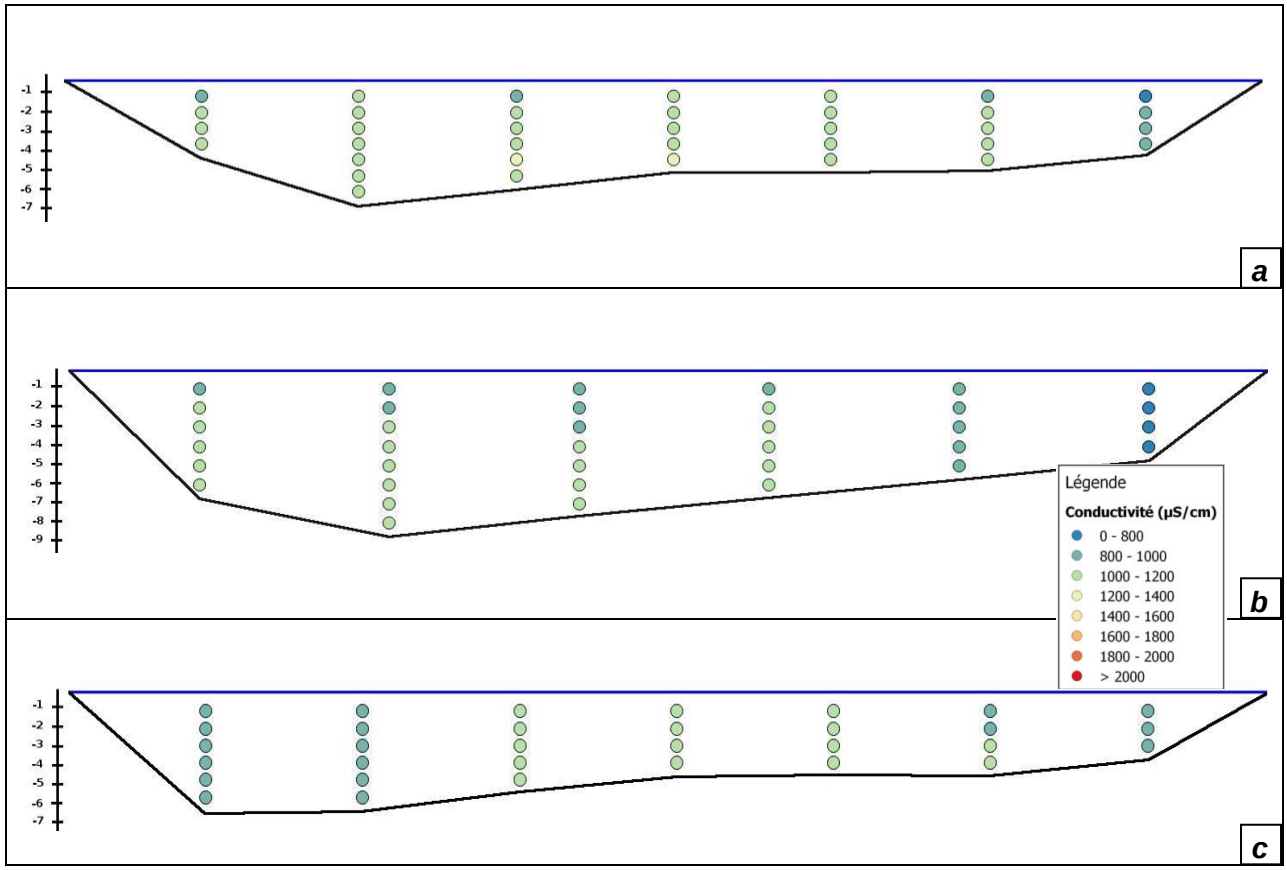


Illustration 8 : Répartition de la conductivité sur les section S2 (a) S3 (b) et S4 (c) réalisées le 28 octobre 2015

On remarque cette fois-ci, à l'inverse de la mission du mois de septembre, que les mesures de conductivité sont plus homogènes sur la lame d'eau, à la fois sur la hauteur que sur la largeur. De la surface vers le fond, bien que l'on observe une légère augmentation, les valeurs restent entre 900 et 1100 µS/cm. On remarque également sur S2 et S3 des valeurs légèrement plus faibles proche de la rive droite. Sur S4, il apparait une conductivité un peu plus élevée au centre du cours d'eau que proche des berges. Dans cette situation de fort coefficient de marée et faible débit du fleuve, l'intrusion saline ne montre ainsi pas ou peu de forme biseautée.

Mission du 20 novembre 2015

Bien que le débit de la rivière soit relativement faible lors de la mission du mois de novembre, la hauteur d'eau à la pleine mer de seulement 2,88 m crée un contexte où le front salin remonte au maximum en aval du dégrad St Antoine (village Favard) comme on peut le voir sur l'Illustration 5. Une seule section a pu être réalisée ce jour due à la rapidité du retour au sens d'écoulement normal du cours d'eau.

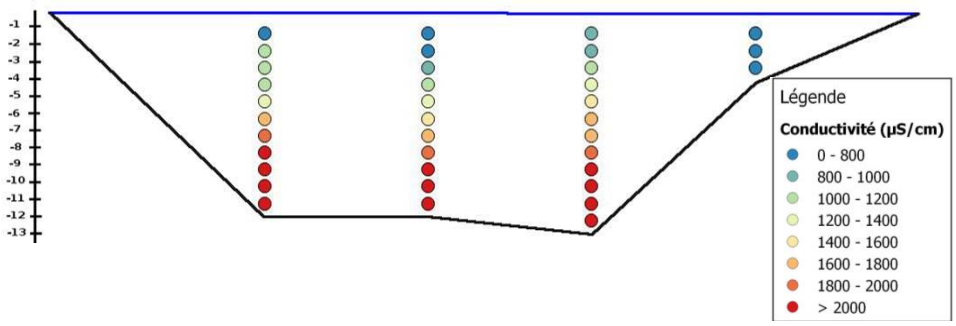


Illustration 9 : Répartition de la conductivité sur la section S5 réalisée le 20 novembre 2015

La section de mesure réalisée (Illustration 9) met en évidence une très forte stratification de la conductivité, avec des valeurs multipliées par quatre entre la surface et le fond du cours d'eau (de 600 à 2800 µS/cm). Ces observations traduisent, comme pour la mission de septembre, une allure très biseautée du front, cette fois-ci dans un contexte de faible débit avec un faible coefficient de marée.

Mission du 15 décembre 2015

Durant le mois de décembre 2015, des épisodes pluvieux ont marqués la saison sèche, provoquant une hausse du débit de la rivière (de 16 000 m^3/j le 20 novembre à 24 400 m^3/j). Bien que la hauteur d'eau à la pleine mer soit de 2,99 m lors de la mission, qui est une valeur moyenne, la position du front salin lors du maximum de sa remonté était au niveau du Dégrad St-Antoine (Illustration 5).

Quatre sections ont pu être réalisées ce jour (Illustration 10). En revanche la première (S6), qui est la plus en aval, a été obtenue sur un temps trop long (>25 min) ce qui ne permet pas d'avoir une image représentative de la répartition de la conductivité. En effet les fortes valeurs de conductivité obtenues à partir du milieu de ce profil proviennent très probablement de mesures réalisées après le passage du front salin des 1000 µS/cm. Néanmoins cette section permet de voir la forte stratification de la conductivité au moment de l'arrivée du front. Les profils S7 à S9 confirment ce constat avec des valeurs passant de 900 à 1300 µS/cm de la surface vers le fond du fleuve. Cette stratification, et donc l'allure biseautée du front, est cependant bien moins marquée que celles observées durant les missions de septembre et novembre 2015.

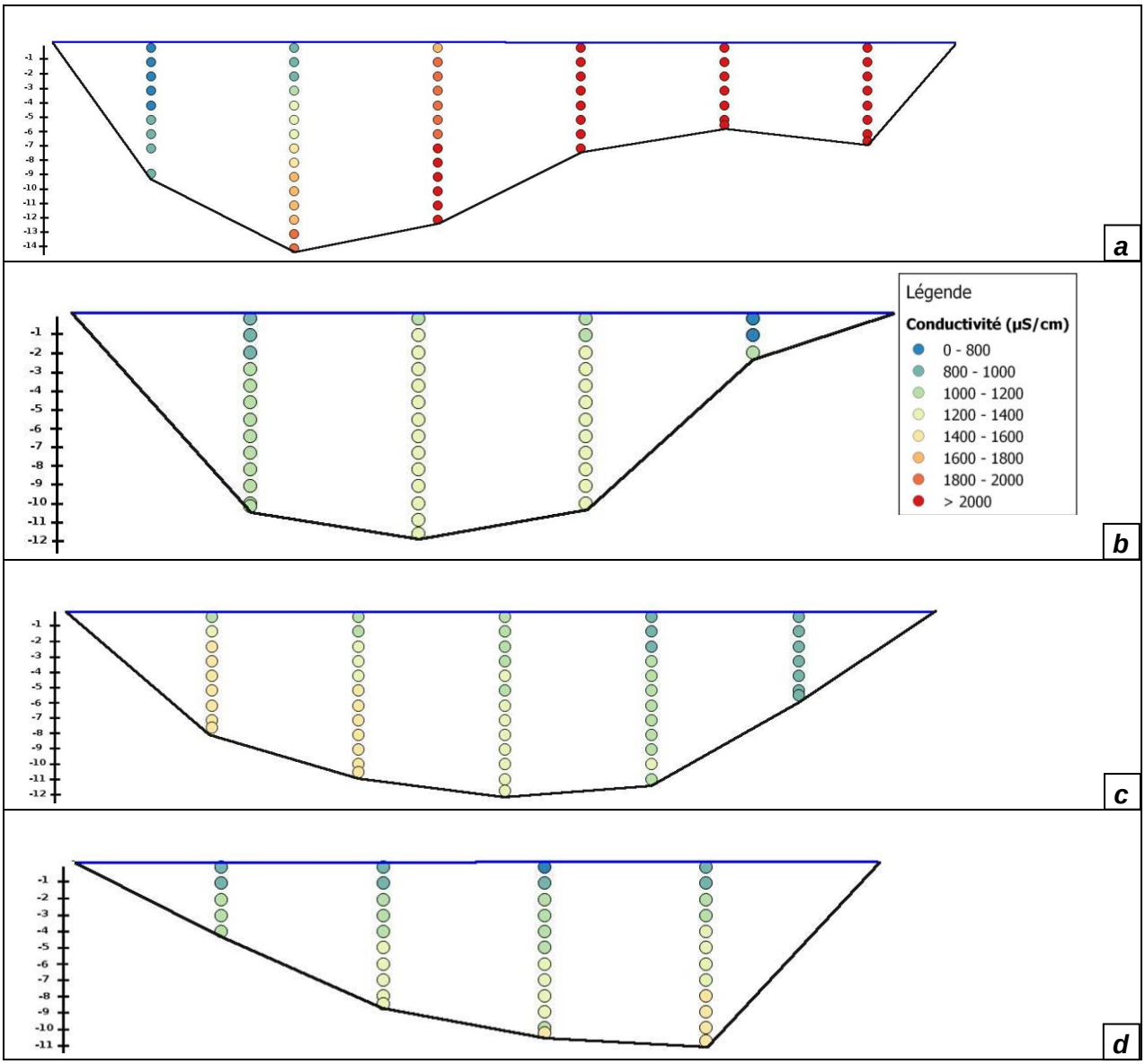


Illustration 10 : Répartition de la conductivité sur les section S6 (a) S7 (b), S8 (c) et S9 (d) réalisées le 15 décembre 2015

Mission du 25 janvier 2016

La dernière mission s'est déroulée fin janvier 2016, période à laquelle débute normalement la saison des pluies. Mais la pluviométrie étant restée très faible début 2016, il en est de même pour le débit journalier mesuré à Saut Bief ce jour (14 700 m^3/j). Le coefficient de marée est relativement élevé avec une hauteur d'eau de 3,27 m à la pleine mer. L'intrusion saline est remontée cette fois-ci juste en amont de la confluence entre la rivière Comté et l'Orapu (Illustration 5).

Comme lors de la journée du 28 octobre 2015 où le coefficient de marée était élevé, on observe des profils de conductivité assez homogène (Illustration 11), avec une légère stratification (globalement de 900 µS/cm à 1100 µS/cm) de la surface vers le fond.

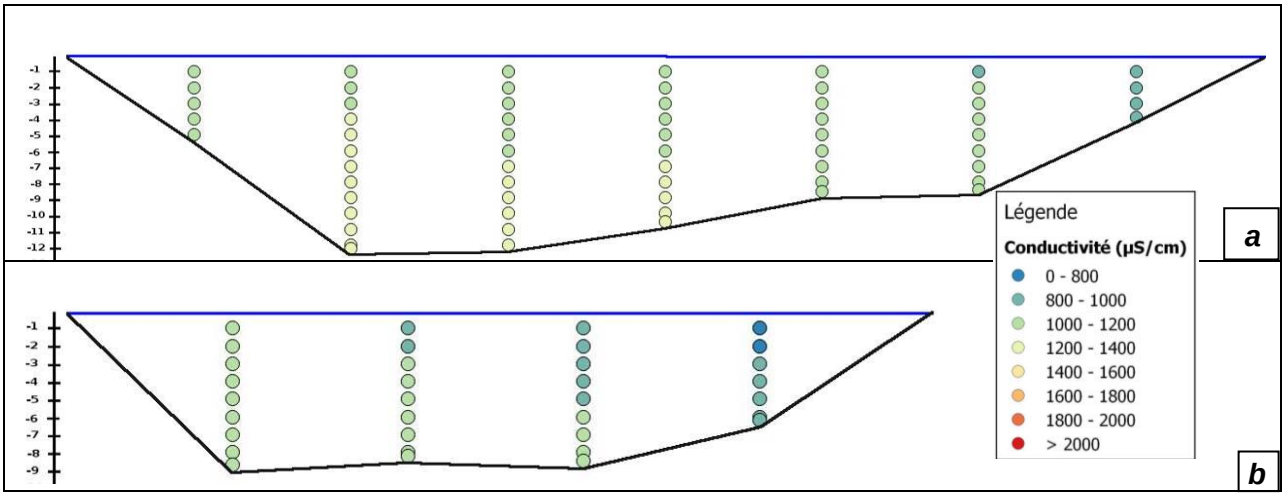


Illustration 11 : Répartition de la conductivité sur les section S10 (a) et S11 (b) réalisées le 25 janvier 2016

2.3. CONCLUSION SUR LA CARACTERISATION DU FRONT SALIN

La connaissance de la forme du front salin étant déterminante pour établir des modèles conceptuels ou hydrodynamiques, et donc pour améliorer la gestion de l'unité de potabilisation et du captage associé, des campagnes de mesures de la conductivité de l'eau étaient nécessaires.

Afin d'apprécier l'influence de la marée sur la position et l'allure du front, cinq missions ont été réalisées de septembre 2015 à janvier 2016, à des coefficients de marée faibles, moyens et forts, durant une saison sèche qui fut relativement longue et marquée.

Il en ressort les éléments suivants :

- Lorsque le débit de la rivière est très faible (octobre, novembre, janvier) et que la hauteur d'eau à la pleine mer est élevée (octobre et janvier), on observe un front salin relativement homogène sur les sections de mesures, autant sur la hauteur que d'une rive à l'autre. Il n'y a ainsi pas d'allure biseautée observée dans ce contexte, qui est celui le plus problématique car favorisant la plus forte remonté du front.

- Lorsque le cours d'eau a un débit relativement plus élevé (septembre et décembre), une stratification de la conductivité se dessine sur les verticales de mesures. Ce phénomène est encore plus prononcé lors d'un faible coefficient de marée comme en septembre, où la salinité peut tripler entre la surface et le fond de la rivière.

Ces constats sont en accord avec les travaux de Savenije (2012) qui définissent la distribution de la salinité dans les estuaires en fonction du débit du cours d'eau. Les campagnes de mesures réalisées sur la Comté mettent également en lumière l'importance de l'amplitude de la marée dans cette distribution de la conductivité. En revanche ces analyses montrent que dans le contexte critique pour la gestion du captage, où le débit est faible et le coefficient de marée élevé, le front a une allure pseudo-verticale. Ces observations vont permettre d'orienter le choix du futur modèle numérique qui décrira la distance du front salin au captage en fonction du débit et de la hauteur d'eau à la pleine mer.

3. Analyses statistiques et modèle probabiliste

Ca chapitre décrit les méthodes statistiques mises en œuvre pour : 1. Etudier la corrélation temporelle entre la conductivité électrique σ et deux facteurs explicatifs, à savoir le niveau d'eau à la côte *SWL* (Sea Water Level) et le débit de la Comté *Q* (Section 2.1); 2. Construire un modèle statistique afin de prédire l'occurrence d'une intrusion saline mise en évidence par un pic de conductivité électrique (Sect. 2.2). Un focus est fait sur la prédiction de deux évènements, à savoir : A. la future valeur maximale du pic dépasse la valeur critique de 900 µS/cm ; B. la durée pendant laquelle le pic dépasse la valeur critique est supérieure à 2 heures.

3.1. ANALYSE DES SERIES TEMPORELLES

3.1.1. Description des données

L'Illustration 12 donne un aperçu des séries temporelles de conductivité σ (logarithme base 10) mesurées pendant les périodes sèches de septembre à décembre de chaque année (de 2009 à 2015) sur la Comté, à 1,4 km en aval du captage. A titre indicatif, le seuil à 500 µS/cm est également donné. Ces séries temporelles (ainsi que celles des niveaux d'eau et de débit) sont toutes exprimées en UT-3 (temps local).

Dans la suite de l'étude seules les périodes de 2009-2012 ont été reprises, car : i. en 2013, une erreur de capteur semble être présente ; ii. en 2014, il n'y a aucun évènement avec σ dépassant le seuil et iii. en 2015, aucune donnée sur le débit de la rivière n'était disponible au moment de l'analyse.

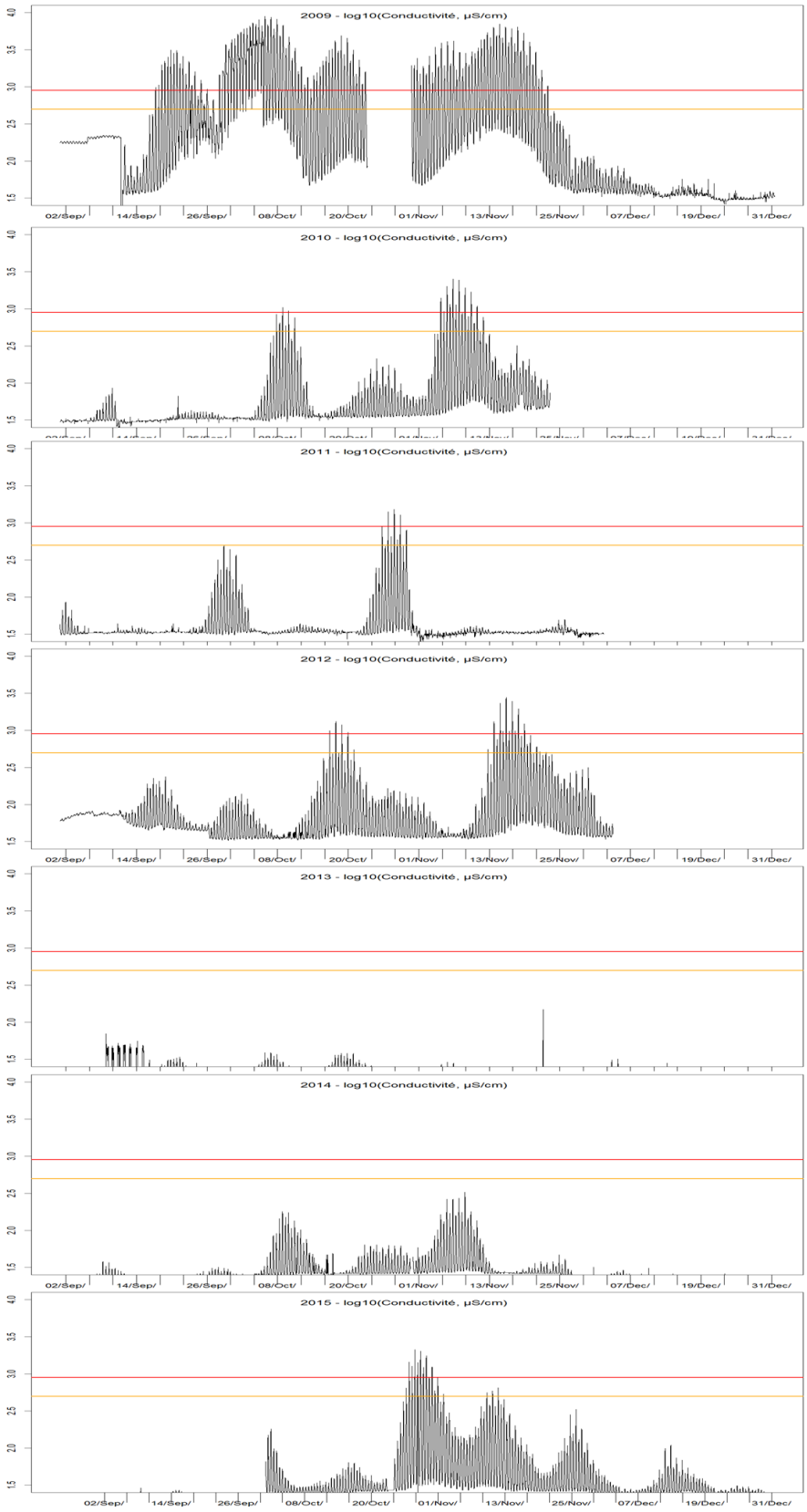


Illustration 12 - Séries temporelles des conductivités électriques (logarithme base 10, µS/cm) mesurées sur La Comté de 2009 à 2015. Les seuils à 900 et 500 µS/cm sont respectivement indiqués par un trait horizontal rouge et orange.

Les séries temporelles sont caractérisées par des « clusters » (groupes) de pics plus ou moins espacées de façon uniforme (Illustration 13). Dans la suite, un pic est assimilé à l'occurrence d'une intrusion saline plus ou moins conséquente ; l'intensité de l'évènement étant mesurée par l'amplitude du pic. L'Illustration 13A donne le détail pour la série temporelle complète de 2010 ainsi que le détail de deux évènements, i.e. deux pics (Illustration 13B et C qui sont respectivement indiquée sur l'Illustration 13A).

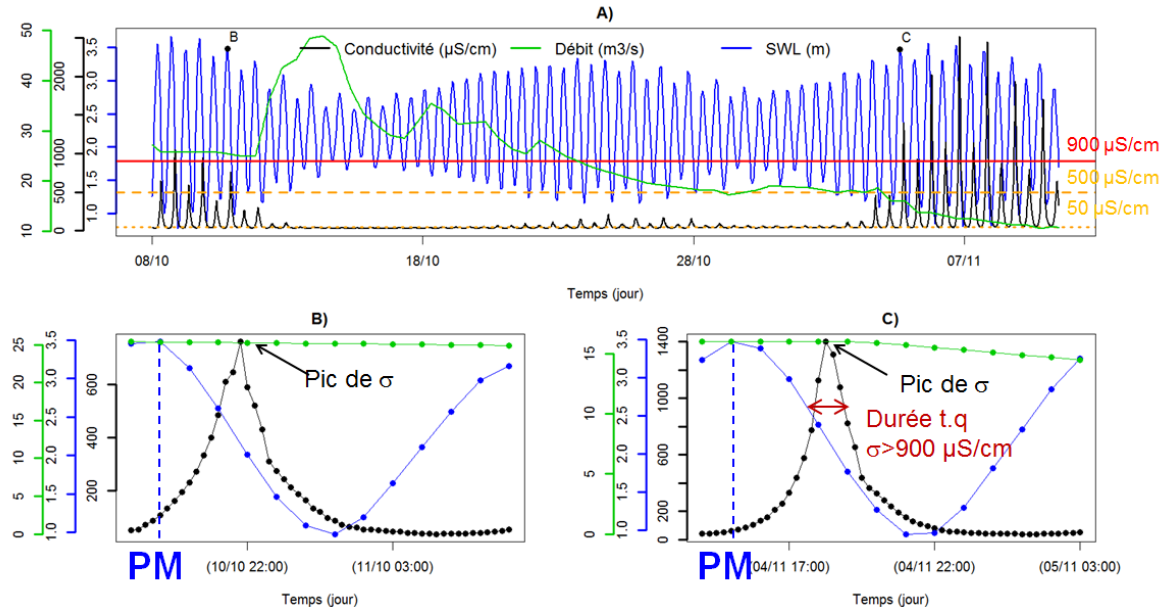


Illustration 13 – Série temporelle de conductivité électrique (noir), débit (vert) et de niveau d'eau à la côte SWL (bleu) pour l'année 2010; B) and C) deux exemples de pic de conductivité. Les conditions à pleine mer (PM) sont indiquées par un trait pointillé bleu. La durée critique t.q. σ > 900 µS/cm est indiquée en rouge.

- Le

montre que :

- le nombre de pics> 50 µS/cm (correspondant à la conductivité électrique moyenne de la pluie en milieu tropical) varie entre quelques dizaines à une centaine selon les années Le nombre total est de 313 ;

- le nombre de pics> 900 µS/cm est de l'ordre de 10-20 à l'exception de l'année 2009 où ce nombre atteint 89 avec des amplitudes max jusqu'à ~9,000 µS/cm ;

- le nombre de pics qui dépassent 900 µS/cm pendant plus de 2 heures restent très modéré (de l'ordre de 0-5 pour 2010- 2012) à l'exception de 2009 où ce nombre atteint 84.

Année	Période**	Nombre de pics > 900 µS/cm	Nombre de pics > 50 µS/cm	1st Quartile – Médiane - Moyenne- 3rd quartile*	Nombre de pics qui dépassent 900 µS/cm pendant plus de 2 heures
2009	17/09 - 22/11	88	128	112.5 – 1690.0 2101.0 – 3749.0	84
2010	08/10 - 10/11	10	53	74.25 - 141.0 - 484.9 - 663.5	5
2011	26/10 - 28/10	3	36	89.0 – 219.0 – 349.5 – 489.0	0
2012	16/10 - 19/11	13	96	75.0 - 119.0 349.6 – 393.0	4

**calculé pour l'ensemble des pics > 50 µS/cm*
*** entre le 1er pic > 900 µS/cm et le dernier*

Tableau 1. *Description de la séquence des pics de conductivité.*

La possibilité de l'occurrence d'un pic est étudiée selon deux facteurs explicatifs principaux :

- Le niveau d'eau à la côte *SWL* mesuré par le marégraphe de l'Ile Royale (données du SHOM accessible sous refmar.shom.fr) ; fréquence d'acquisition : toutes les heures (pour la période 2009-2012) ;

- Le débit de la rivière *Q* mesuré à la station "Saut Bief" (données de la CVH) à 38 kilomètres en amont de la station de pompage ; fréquence d'acquisition : toutes les heures.

3.1.2. Analyse des corrélations

En premier lieu la périodicité de la série temporelle de pics de conductivité est analysée: l'analyse des périodogrammes (construit à partir d'une analyse Fast Fourier Transform FFT) montre que la séquence des pics est caractérisée par une période maximale à 12.3 – 12.5 heures (voir Illustration 13 pour l'année 2010) qui correspond à la période des niveaux d'eau à la côte *SWL* comme le montre l'Illustration 14A pour l'année 2010. Ce constat est également confirmé pour les autres années.

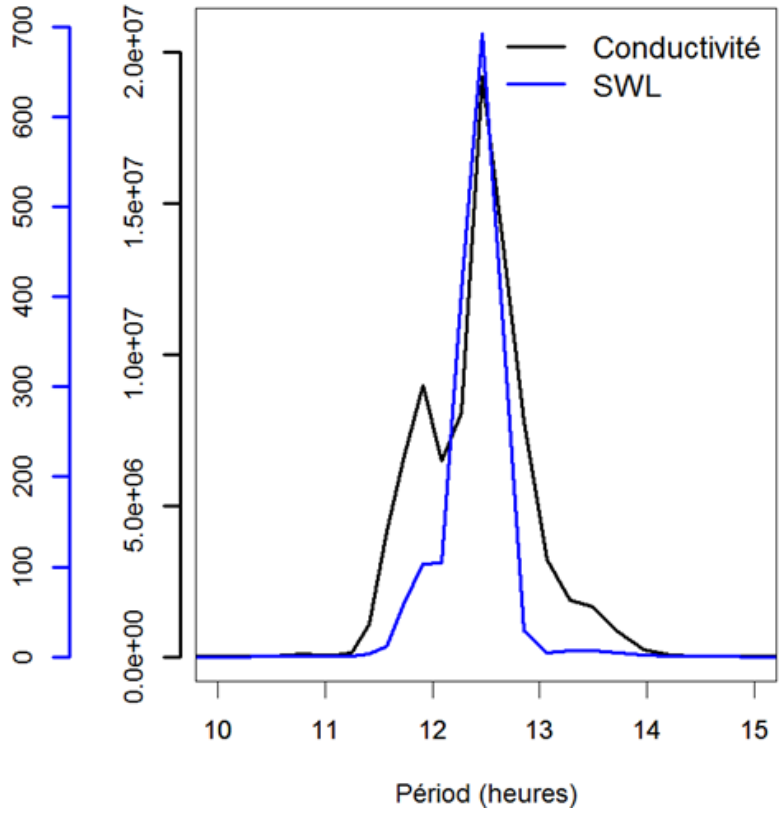


Illustration 14 –Périodogramme pour la série temporelle de conductivité ainsi que celle des niveaux d'eau pour l'année 2010;

Une possible corrélation temporelle entre σ et *SWL* est analysée en calculant la fonction de corrélation croisée $r_{\sigma\text{-}Xe}$ comme suit :

$$c_{\sigma-SWL} = \frac{1}{N}\sum_{t=1}^{N}(\sigma(t)-\overline{\sigma})(SWL(t+k)-\overline{SWL}) \qquad (1)$$

$$r_{\sigma-SWL} = \frac{c_{\sigma-SWL}}{\sqrt{c_{\sigma-\sigma}c_{SWL-SWL}}}$$

Où les termes $\overline{\sigma}, \overline{SWL}$ sont les valeurs moyennes et *N* est la longueur des séries temporelles.

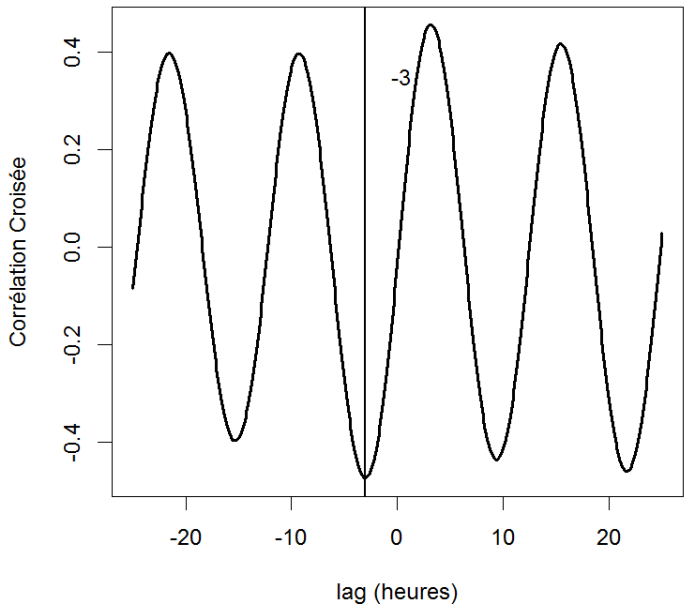


Illustration 15 –Analyse par corrélation croisée pour les deux séries temporelles σ et SWL en 2010: une anticorrlétation est mise en évidence pour un déclage de -3 heures.

L'Illustration 15 montre une anti-corrélation maximale (~coefficient de -0.47) pour un décalage de -3 heures pour 2010 : cela signifie que le pic de *SWL* (i.e. condition de pleine mer PM) apparait lorsque σ est minimale. Cette analyse est conduite pour toutes les années et confirme l'anti-corrélation pour un décalage temporel variant entre 2.9 – 3.1 heures. Il est intéressant de noter que cette dépendance temporelle est également confirmée par les observations sur le terrain.

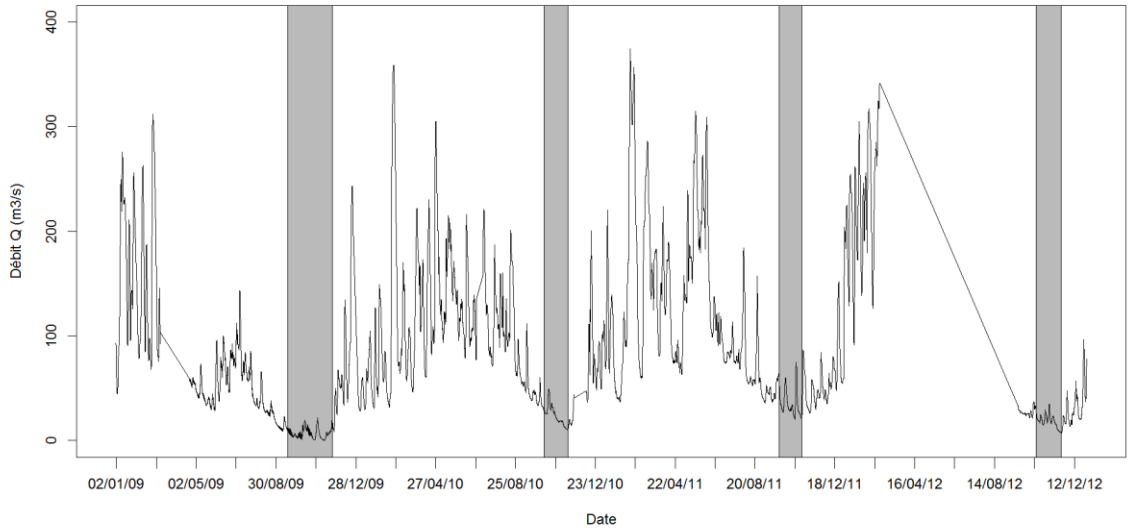


Illustration 16 – Série temporelle de débit de 2009 à 2012. Les périodes où les pics de conductivité sont > 900 µS/cm sont indiquées par une enveloppe grise.

L'analyse de l'influence de *Q* est plus complexe et réalisée de manière qualitative. L'analyse à grande échelle temporelle (à l'échelle pluri annuelle) (Illustration 16) montre que les « périodes de crise », i.e. d'occurrence des pics σ coïncident avec les périodes où *Q* est faible : cela confirme les études faites dans cette région dans le passé (e.g., Lambs et al., 2007).

Une analyse à résolution temporelle plus fine est faite en inspectant les séries temporelles pour chaque année. L'analyse visuelle pour l'année 2010 (Illustration 13A), semble suggérer que la variation du débit se fait à l'échelle temporelle de plusieurs évènements (i.e. plusieurs pics) : la fin du « cluster » de pics à mi-octobre (partie la plus à gauche de l'illustration) semble coïncider avec l'accroissement abrupte de *Q* (ligne en vert). Le début du 2ème cluster de pics au début de novembre (partie la plus à droite de l'illustration) semble coïncider avec la forte décroissance de *Q*. A l'échelle temporelle plus fine d'un évènement, nous noterons que les variations de *Q* restent faibles. Le contrôle du débit est étudié plus en détails dans la suite.

3.2. MODELE STATISTIQUE DE PREDICTION

Dans cette section, les méthodes statistiques mises en œuvre afin de prédire les deux évènements sont décrites: A. la future valeur maximale du pic dépasse la valeur critique de 900 µS/cm ; B. la durée pendant laquelle le pic dépasse la valeur critique est supérieure à 2 heures.

L'analyse précédente montre une dépendance temporelle avec un décalage de 3 heures entre le maximum de *SWL* et σ. Il est proposé de développer un modèle pour prédire si dans un horizon temporel de 3 heures, si l'un des deux évènements A ou B survient. Les valeurs des 2 facteurs explicatifs *SWL* et *Q* dont les valeurs sont prises aux conditions de pleine mer PM sont retenues. Le problème est abordé sous l'angle de la classification i.e. en prédisant si le prochain pic σ appartient à la classe +1 (évènement A (ou B) survient) ou -1 (évènement A (ou B) n'apparait) à partir de la connaissance de la combinaison $\boldsymbol{x}=(SWL_{PM} ; Q_{PM})$.

Dans la littérature, un nombre très important de méthodes existent afin de traiter ce problème (voir par exemple une revue récente par Kuhn et Johnson (2013)). Ici, un focus est fait sur la technique de machine à vecteurs de support (Support Vector Machine SVM, Vapnik, 1998) : cette technique a été utilisée pour plusieurs problèmes en hydrologie (e.g., Yu et al., 2006; Wu et al., 2008; Wang et al., 2009).

3.2.1. Principes de SVM

L'idée de base est de trouver la meilleure frontière (dénommée frontière de décision *f*) qui sépare de manière optimale les deux classes {-1 ; 1}. Afin d'illustrer les principes, un cas fictif 2d (Illustration 17A) est retenu pour la suite : l'objectif est donc de séparer au mieux les deux ensembles de points, chaque ensemble étant associé à une classe différente {-1 ; 1}.

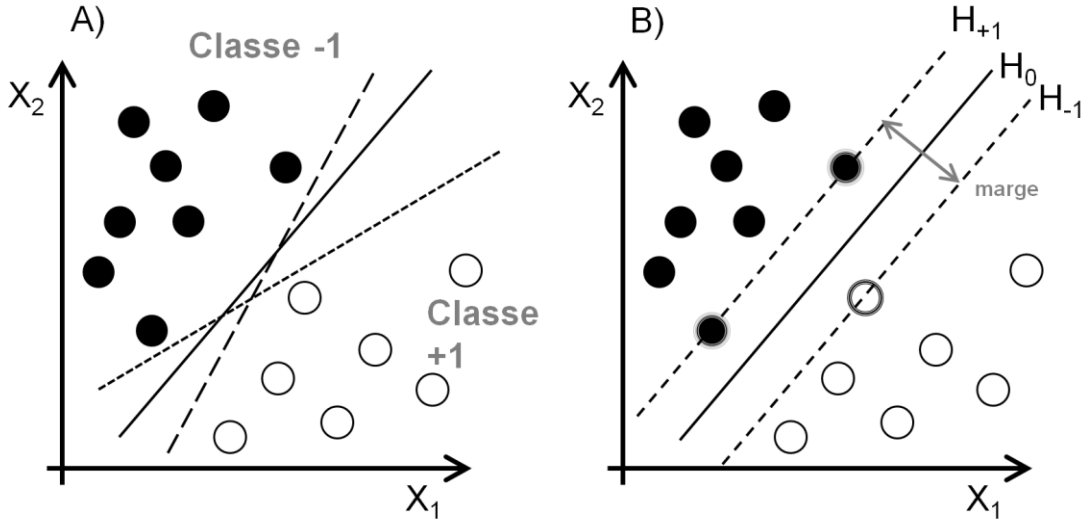


Illustration 17 – A) Identification des différentes frontières linéaires (hyperplans) qui séparent les deux ensembles de points; B) Résultat du modèle SVM : hyperplan de décision H_0 et hyperplans associés aux marges, H_{-1} and H_{+1}.

Dans le cas de frontières linéaires, le problème de classification peut être résolu en approximant *f par* un hyperplan H_0 auquel deux hyperplans parallèles peuvent être associés et définissant la « marge » (H_{-1} et H_1) qui séparent les deux ensembles de points :

$$\begin{aligned} H_0 &: \boldsymbol{w}\cdot\boldsymbol{x}+b=0 \\ H_{-1} &: \boldsymbol{w}\cdot\boldsymbol{x}+b=-1 \\ H_{+1} &: \boldsymbol{w}\cdot\boldsymbol{x}+b=+1 \end{aligned} \qquad (2)$$

Le vecteur $\boldsymbol{w}$ et le scalaire b (biais) sont déterminés via une procédure d'apprentissage supervisé à partir des n observations (données d'apprentissage) : ($\boldsymbol{x}_1$,y_1), ($\boldsymbol{x}_2$,y_2), …, ($\boldsymbol{x}_n$,y_n): cela consiste en la maximisation de la marge sous la contrainte que cet espace ne contienne aucune observation. En pratique, lorsque ce problème est difficile à résoudre, l'hyperplan H_0 peut être trouvé en « permettant » qu'un faible nombre de points soient mal classés : cela passe par l'introduction d'un facteur dit de « régularisation » C, qui permet de régler l'équilibre entre le taux d'erreur (nombre de points mal classés) et distance entre les deux hyperplans H_{-1} et H_1.

Dans le cas non linéaire, une fonction « noyau » K peut être utilisée au préalable pour projeter (transformer) les données sur un nouvel domaine mathématique où la linéarité est valide. Plus de détails peuvent être trouvés dans (Schölkopf et Smola 2002). L'équation de la frontière devient :

$$\Phi(\boldsymbol{x}) = b + \sum_{i=1}^{n} \lambda_i y_i K(\boldsymbol{x}_i, \boldsymbol{x}) \qquad (3)$$

où b est un terme constant (biais), λ_i sont les multiplieurs Lagrangien obtenus en résolvant le problème d'optimisation décrit ci-avant. La fonction K peut prendre différentes formes : polynomiale, sigmoïde, fonction de lissage, etc. Dans la suite, nous nous focalisons sur le dernier modèle dit « gaussien » :

$$K(\boldsymbol{u},\boldsymbol{v}) = \exp(-\gamma\|\boldsymbol{u}-\boldsymbol{v}\|^2) \quad (4)$$

où le paramètre γ est la longueur de corrélation. En practice, γ (ainsi que le terme de régularisation *C*) sont estimés de telle manière à minimiser le d'erreur obtenu par approche bootstrap (e.g., Fan et al., 2005).

3.2.2. Validation

Une difficulté majeure lorsque le SVM est utilisé dans une démarche prédictive est de soigneusement vérifier sa capacité à prévoir la classe d'un futur évènement pour une configuration ***x*** « jamais vue » i.e. qui n'a pas été utilisée pour l'apprentissage (construction) du SVM. Afin de valider cette capacité prédictive, un jeu de données (configurations ***x*** ; classes du pic) indépendant de celui utilisé pour la construction du SVM doit être défini : cela peut se faire en retirant 50% des observations du jeu de départ et de les utiliser comme ensemble de validation. En comparant prédictions et « vraies » classes, différents indicateurs de validation (e.g. Powers, 2007) peuvent être estimés, à savoir :

1) la précision (ratio du nombre d'évènements (pics) correctement classés (classe +1 ou -1) sur le nombre total de pics) ;

2) le taux de « vrais positifs » *tpr* (ratio du nombre d'évènements prédits par le modèle comme appartenant à la classe « +1 » et appartenant réellement à cette classe sur le nombre total de pics classés en « +1 ») ;

3) le taux de « faux positifs » *fpr* (ratio du nombre d'évènements prédits par le modèle comme appartenant à la classe « +1 » mais appartenant réellement à la classe « -1 » sur le nombre total de pics classés en « -1 » ;

4) un dernier indicateur peut être défini à partir de la « courbe d'efficacité du récepteur » (Receiver Operating Characteristic ROC, Metz 1978), qui relie *tpr* et *fpr*: Un exemple est donné dans la section suivante. L'aire sous cette courbe *auc* mesure à quel point le modèle peut distinguer les deux classes. Plus cette aire est proche de 1.0, meilleure est cette capacité à distinguer les deux classes.

Enfin, il faut souligner que le modèle SVM est construit avec un nombre limité de points, la classification peut donc être associée à une erreur. Afin d'apprécier cette incertitude classification, une probabilité de classification peut être estimée. Cela peut être basée sur l'approche par fonction sigmoïde (Vapnik 1998; Platt 1999). Pour le point ***x***, the probabilité d'appartenir à la classe +1 est :

$$P(y=+1|\boldsymbol{x}) = \frac{1}{1+\exp(A\cdot\Phi(\boldsymbol{x})+B)} \quad (5)$$

où les constantes *A* et *B* sont déterminées par maximisation de la vraisemblance (Platt 1999).

En pratique, si $P(y=+1|\boldsymbol{x})$ >50%, cela signifie que le point ***x*** devrait appartenir à la « classe +1 » pour ce niveau de confiance (la probabilité d'appartenir à la « classe -1 » étant dans ce cas $P(y=-1|\boldsymbol{x})=1-P(y=+1|\boldsymbol{x})$ < 50%). Si $P(y=+1|\boldsymbol{x})$ est proche 100%, on peut considérer que la classification est associée à un fort degré de confiance. Par contre, si $P(y=+1|\boldsymbol{x})$ est proche de 50% (par exemple < 75%), le résultat de la classification devrait être pris avec précaution. La

suite du rapport montre comment cette information sur la probabilité de classification peut être utilisée pour aider à la décision dans une démarche prédictive.

3.3. APPLICATION

3.3.1. Evènement A : « pic > 900 µS/cm »

Tous les pics de conductivités électriques dépassant 50 µS/cm ont été extraits des séries temporelles 2009-2012, ainsi que les conditions à pleine mer du niveau d'eau et du débit (i.e. ~3 heures avant le pic): SWL_{PM} and Q_{PM}. Au total, 313 des combinaisons (SWL_{PM} ; Q_{PM} ; σ) ont été extraites.

Dans un premier temps, la relation entre l'occurrence de l'évènement A et les facteurs SWL_{PM} et Q_{PM} est analysée: l'Illustration 18 donne l'évolution de la probabilité empirique (fréquence) de l'évènement A sur la plage de variation respective de SWL_{PM} et Q_{PM}. Plus SWL_{PM} est grand, plus cette probabilité est grande, i.e. plus la chance d'avoir un pic dépassant le seuil est grande, alors que plus Q_{PM} est grand, plus cette probabilité est petite. Notons également que pour SWL_{PM} < ~3m ou Q_{PM}> ~30 m^3/s, cette probabilité est nulle.

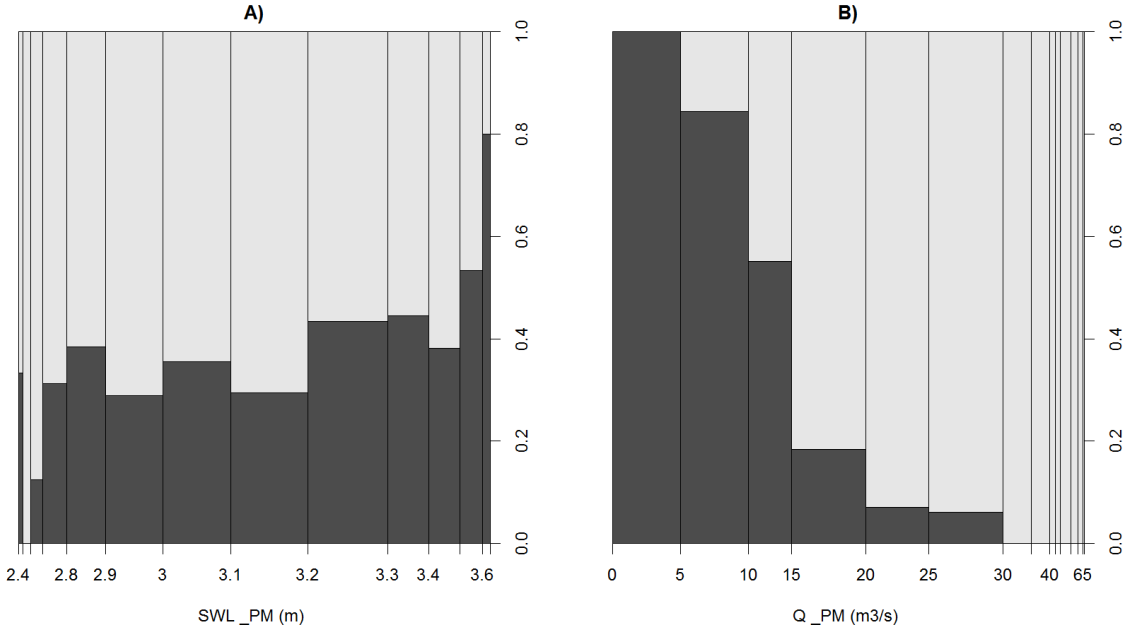


Illustration 18 – A) Fréquence des pics > 900 µS/cm (en noir) versus le niveau d'eau à conditions de pleine mer (SWL_PM); B) Fréquence des pics > 900 µS/cm (en noir) versus le débit de la rivière à conditions de pleine mer (Q_PM).

Dans un premier temps, un modèle SVM est construit en n'utilisant que 50 % des combinaisons (156) : cela comprend les évènements de 2009 (128) et ~50% de ceux pour 2010 (28). Pour ce jeu de données d'apprentissage, le nombre de pics dépassant 900 µS/cm est de 17 (~11% du nombre total des évènements). Le reste (157 évènements) a été utilisé (2010-2012) comme ensemble de validation en suivant la procédure décrite dans la section précédente.

La forme du graphique de l'Illustration 18 suggère une relation non linéaire entre (SWL_{PM} ; Q_{PM}) et la probabilité d'être dans la classe « +1 »: une forme non linéaire de SVM est alors choisie en utilisant une fonction noyau K de type gaussien. Les paramètres sont estimés avec une méthode par bootstrap (250 échantillons sont générés aléatoirement), ce qui donne γ=0.01 et

C=500. En pratique, le package "caret" développé par Kuhn (2008) dans le logiciel R (R Development Core Team, 2015) a été utilisé.

L'Illustration 19A donne l'évolution de la probabilité d'appartenir à la classe « +1 » estimée avec le modèle SVM sur tout le domaine de (SWL_{PM}; Q_{PM}). La frontière de décision *f* est ici donnée par l'iso-contour à niveau de probabilité de 50% (ligne verte). Notons que deux points sont mal classés : cela est lié à l'utilisation d'un terme de régularisation C facilitant la construction du modèle.

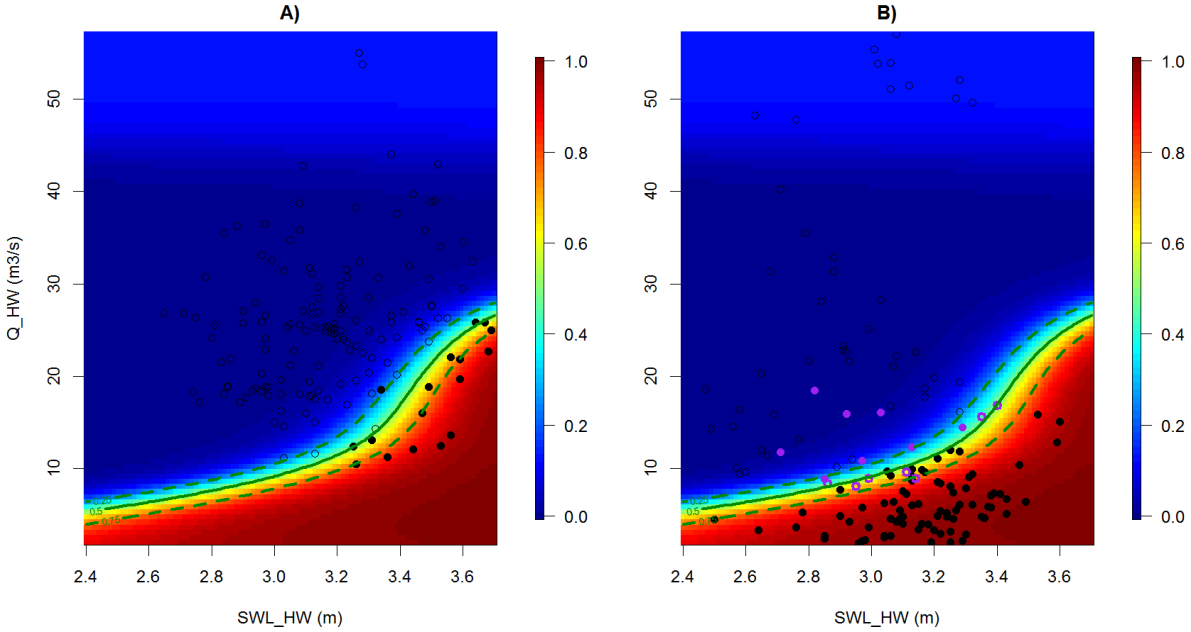


Illustration 19 – A) Probabilité d'appartenir à la classe «+1 : pic > 900 µS/cm » estimée par le modèle SVM construit à partir de 50% des données. Les ronds noirs sont ceux en classe « +1 » et les cercles sont ceux en classe « -1 »; B) Localisation des points de validation; les points mal classés sont indiqués en violet.

L'Illustration 19B donne la localisation des points de validation par rapport à la frontière de décision estimée par le SVM. Dans l'ensemble, la classification est très satisfaisante à l'exception de 15 points. En utilisant ce jeu de validation, les différents indicateurs peuvent être évalués : *précision*=90.4% (15 points mal classés sur 157); *tpr*=91.75% (8 pics dépassant réellement le seuil sont mal classés); *tnr*=88.33% (7 pics réellement en dessous du seuil sont mal classés); *auc*=90.1%. L'Illustration 20 donne la courbe ROC, qui confirme la bonne capacité prédictive du modèle SVM : la courbe est proche du coin en haut à gauche.

L'analyse des probabilités de classification révèlent que ~50% des 15 évènements mal classés (violet sur l'Illustration 19B) ont une probabilité proche de 50% (entre 50 et 75%) : cela indique que le modèle SVM n'est pas « sûre » de la classe de ces évènements. Pour ces évènements, un message d'avertissement pourrait être émis afin de pointer cette incertitude. En d'autres termes, si cette information sur les probabilités était intégrée, le taux d'erreur pourrait être réduit.

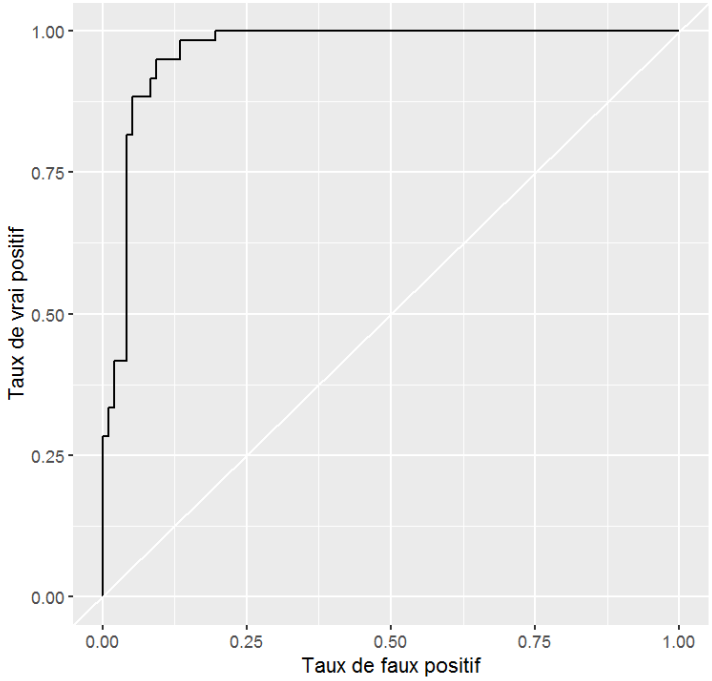


Illustration 20 – courbe ROC du modèle SVM pour l'évènement B construit avec 50 % des données. Plus la courbe est proche du coin en haut à gauche, meilleure est la classification.

3.3.2. Evènement B : « durée du pic >900 µS/cm est supérieure à 2 heures »

Dans un premier temps, la relation entre l'occurrence de l'évènement B et les facteurs SWL_{PM} et Q_{PM} est analysée: l'Illustration 21 donne l'évolution de la probabilité empirique (fréquence) de l'évènement B sur la plage de variation respective de SWL_{PM} et Q_{PM}. Il est plus difficile de détecter une relation claire avec SWL_{PM}. D'un autre côté, plus Q_{PM} est grand, plus cette probabilité est petite. Notons également que pour Q_{PM}> ~20 m^3/s, cette probabilité est nulle.

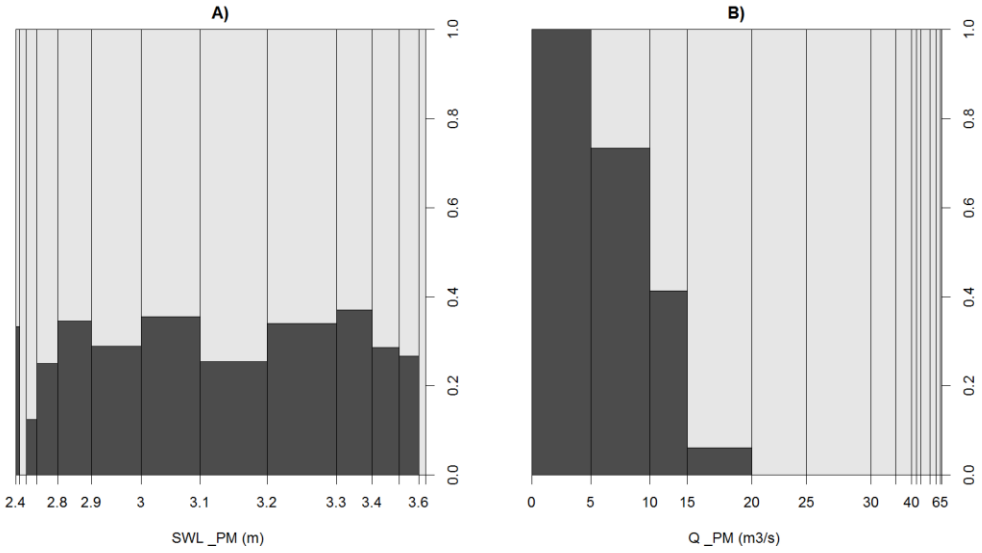


Illustration 21 – A) Fréquence de l'évènement B (en noir) versus le niveau d'eau à conditions de pleine mer SWL_PM; B) Fréquence de l'évènement B (en noir) versus le débit de la rivière à conditions de pleine mer Q_PM.

A l'instar de l'évènement A, un modèle SVM en n'utilisant que 50 % des combinaisons (156) est construit. Pour ce jeu de données d'apprentissage, le nombre de pics dépassant 900 µS/cm pendant plus de 2 heures est de 5 (~3% du nombre total des évènements). Notons ici le faible nombre d'évènements « positifs » utilisés pour réaliser la cosntruciton du SVM. Le reste (157 évènements) a été utilisé (2010-2012) comme ensemble de validation.

Sur ce jeu de données, les paramètres du modèle SVM (avec noyau gaussien) sont estimés à γ=0.02 et C=500. L'Illustration 22 donne le modèle SVM pour l'apprentissage et la validation. Le modèle est plus linéaire que celui pour l'amplitude des pics : la frontière de décision (en vert sur l'Illustration 22) est beaucoup moins convexe et se rapproche pratiquement d'une droite. Notons aussi sur l'Illustration 22A que l'apprentissage se fait sur peu d'échantillon « positif » (seulement 4 points noirs foncés)

Dans l'ensemble, la classification reste satisfaisante bien que moins performante que pour le modèle SVM sur l'amplitude. En utilisant le jeu de validation, les différents indicateurs peuvent être évalués: *précision*=89.2% (17 points mal classés sur 157); *tpr*=~92% (7 pics mal classés); *tnr*=~85.5% (10 pics mal classés); *auc*=88.8% (courbe ROC sur l'Illustration 23).

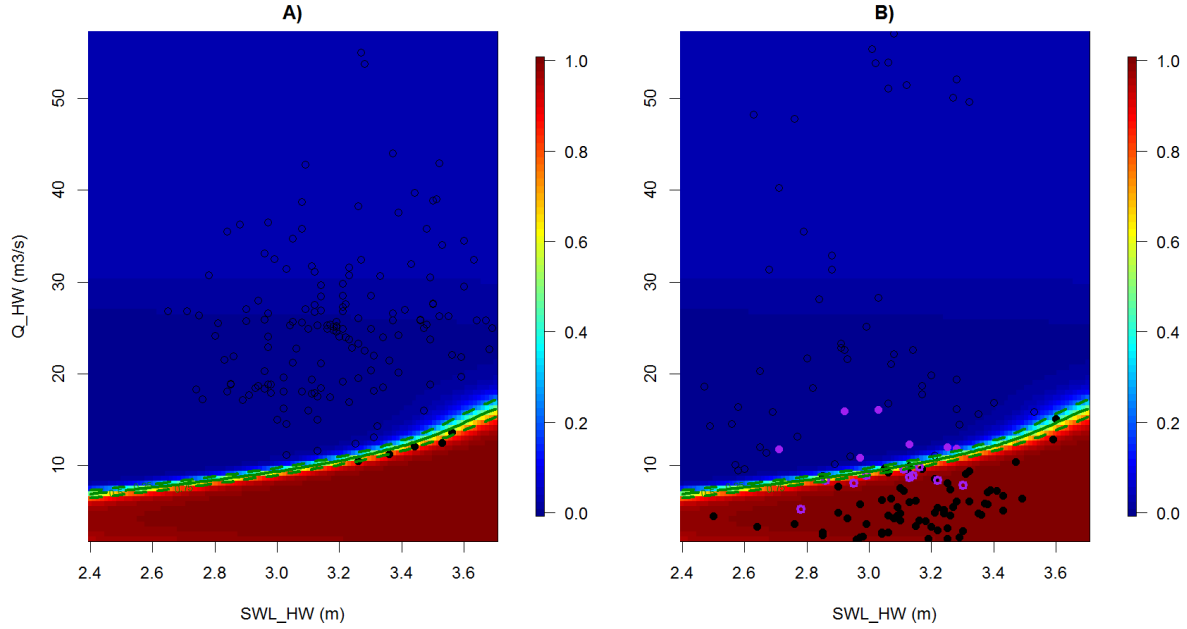


Illustration 22 – A) Probabilité d'appartenir à la classe «+1 : durée des pics dont la conductivité > 900 µS/cm supérieure à 2 heures» estimée par le modèle SVM construit à partir de 50% des données. Les ronds noirs sont ceux en classe « +1 » et les cercles sont ceux en classe « -1 »; B) Localisation des points de validation; les points mal classés sont indiqués en magenta.

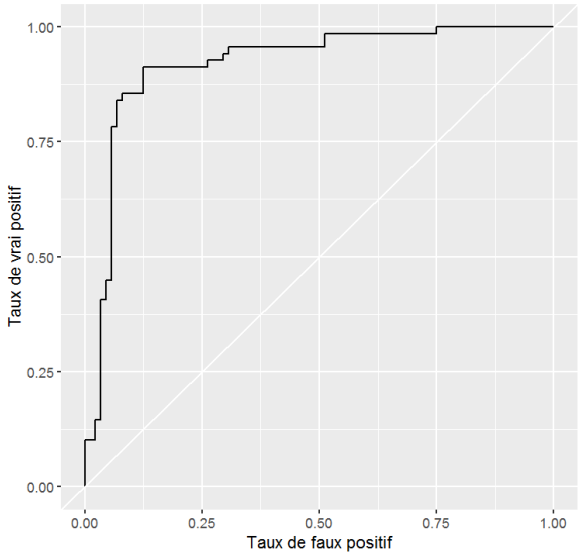


Illustration 23 – courbe ROC du modèle SVM pour l'évènement B construit avec 50 % des données. Plus la courbe est proche du coin en haut à gauche, meilleure est la classification.

3.3.3. Tests de performance

La performance du modèle SVM décrite ci avant ne suffit pas à valider complètement la capacité prédictive du modèle. Une exploration plus poussée de la dépendance de la performance au jeu de données d'apprentissage doit être faite. Dans ce but, le processus d'apprentissage en tirant aléatoirement 250 fois, soit un jeu de données des données de 2009 à 2012, est réitéré. En d'autres termes, à chaque itération, 156 données sont tirées aléatoirement (échantillonnage aléatoire avec remplacement) et les autres données restantes sont utilisées comme données de validation.

En considérant l'évènement A, l'Illustration 24 donne l'histogramme pour les 250 jeux d'indicateurs de validation. Cela confirme la très bonne capacité du modèle SVM à prédire pour des configurations (SWL_{PM} ; Q_{HW}) non encore « vues » avec *auc* moyen de ~93%; *précision* moyenne *de* 92.5%; un nombre moyen de pics mal classés : ~10; avec des taux moyens de classification: ~90% (vrais positifs) and ~95% (vrais négatifs). Par ailleurs, l'analyse des probabilités associées aux pics mal classés révèle qu'en moyenne, ~6 évènements ont des probabilités entre 50 et 75%. Pour ces évènements un avertissement pourrait être émis pour souligner un degré de confiance moindre sur la prédiction (classification).

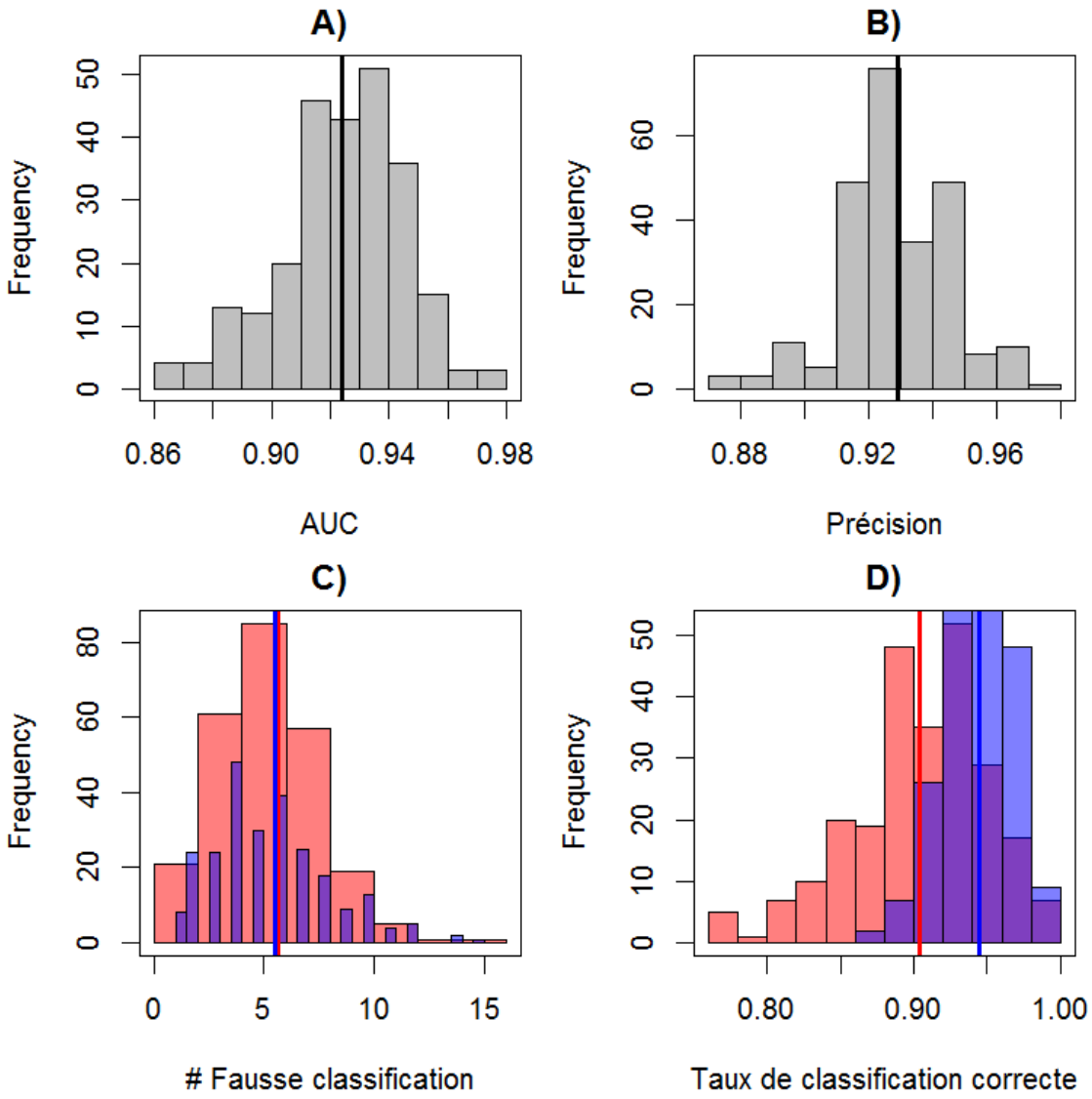


Illustration 24 – Histogrammes des indicateurs de validation pour les 250 tests de performance pour le classement de l'amplitude pics (évènement A). A) aire sous la courbe ROC auc*; B) précision; C) nombre de pics mal classés (rouge: faux positifs; bleu: faux négatifs); D) taux de classification pour les deux classes (rouge: positif; bleu: négatif). Les valeurs moyennes sont indiquées par une barre verticale.*

En considérant l'évènement B, l'Illustration 25 donne les résultats des indicateurs de validation pour 250 tests aléatoires. Ici, la performance du modèle SVM est satisfaisante, mais moins bonne que celle pour l'amplitude des pics avec *auc* moyen de <92%; *précision* moyenne *de* ~92%; un nombre moyen de pics mal classés : ~12; avec des taux moyens de classification: ~89% (vrais positifs) and ~94% (vrais négatifs). Notons aussi que l'étalement du nombre de pics mal classés est plus important que pour le modèle SVM pour l'amplitude : le nombre de pics positifs mal classés est plus important ici, ce qui est en lien avec le faible nombre d'observations de ce type (i.e. ce qui rend difficile la construction du SVM).

A l'instar de l'évènement A, une amélioration peut tirer parti de l'analyse des probabilités associées aux pics mal classés : cela révèle qu'en moyenne, ~7 évènements ont des

probabilités entre 50 et 75%. Pour ces évènements un avertissement pourrait être émis pour souligner un degré de confiance moindre sur la prédiction (classification).

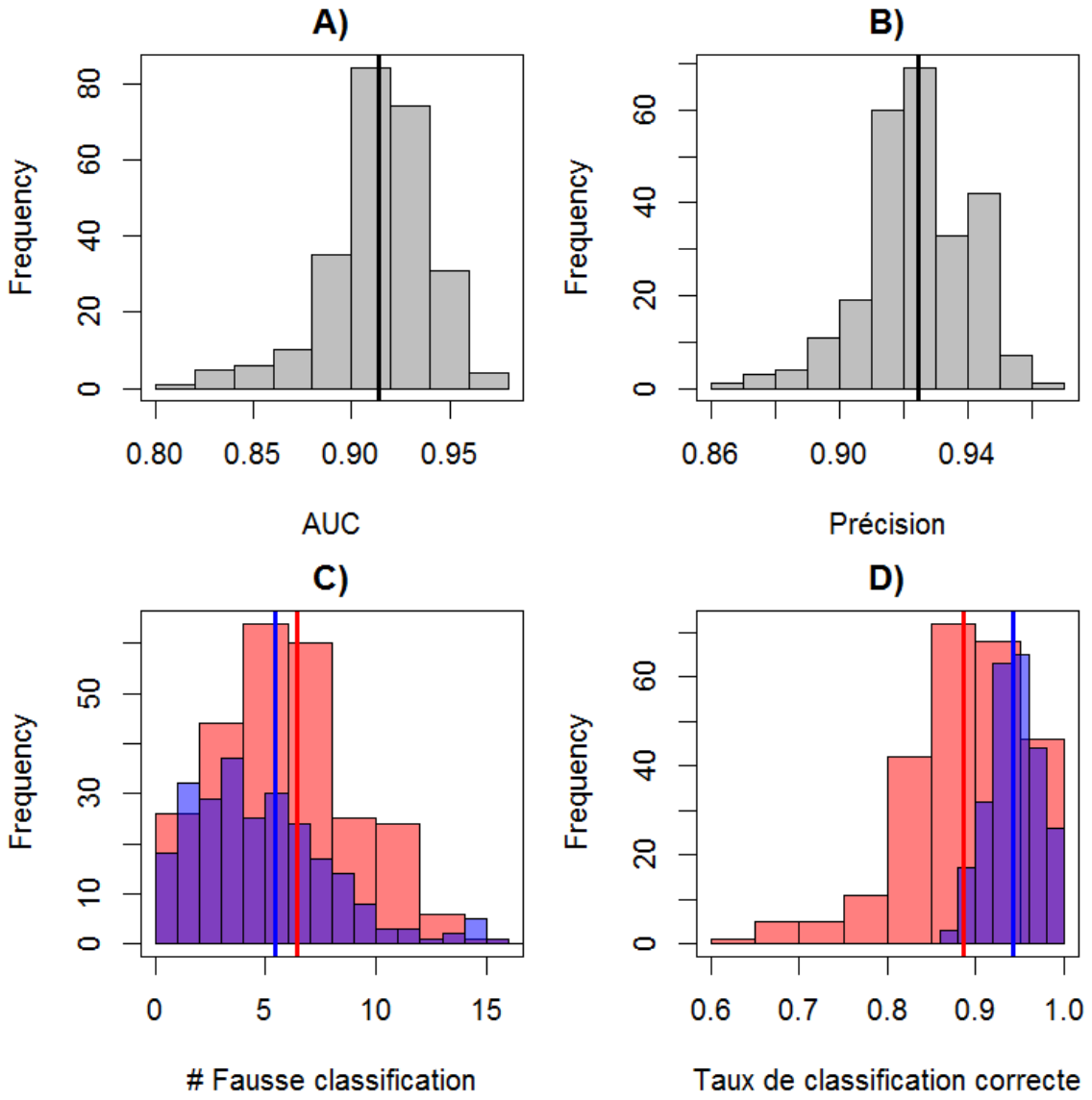


Illustration 25 – Histogrammes des indicateurs de validation pour les 250 tests de performance pour le classement des durées de pics (évènement B). A) aire sous la courbe ROC auc*; B) précision; C) nombre de pics mal classés (rouge: pics dépassant le seuil mal classés; bleu: pics en dessous du seuil mal classés); D) taux de classification pour les deux classes (rouge: positif; bleu: négatif). Les valeurs moyennes sont indiquées par une barre verticale.*

3.4. CONCLUSION SUR L'ANALYSE STATISITIQUE ET LE MODELE PROBABILISTE

Dans cette étude, les séries temporelles de conductivité électrique (entre 2009 et 2012 pour les périodes sèches de septembre à décembre) mesurées à La Comté ont été utilisées pour étudier leur dépendance temporelle avec le niveau d'eau à la côte (mesuré au marégraphe de l'Ile Royale), ainsi que le débit de la rivière. Un décalage de 3 heures entre le maximum de niveau d'eau (pleine mer) et celui de conductivité électrique par corrélation croisée a été mis en évidence. La co-occurrence des périodes d'intrusion saline avec celles où les débits de rivière sont faibles a aussi été révélée.

Sur cette base, un modèle statistique pour prédire à court terme (3 heures en avance) l'amplitude du pic de conductivité en fonction des valeurs de niveaux d'eau et de débit pris à la date de pleine mer a été proposé. Ce problème a été abordé sous l'angle de la classification à l'aide de la technique de machine à support vecteur, i.e. en cherchant à prédire si l'amplitude du pic suivant la pleine mer dépasse 900 µS/cm (classe +1) ou non (classe -1). La capacité de ce modèle à prédire a été étudiée via un exercice de validation basée sur la sélection aléatoire des données d'apprentissage du modèle et de validation : cela a confirmé la bonne performance du modèle SVM.

Le même exercice a été effectué pour prédire si la durée pendant laquelle le pic>900 µS/cm dépasse 2 heures et s'est révélé également concluant, mais la performance du modèle reste moins bonne que celle du modèle sur l'amplitude des pics à cause du faible nombre d'observations.

Ces résultats peuvent ainsi contribuer à l'établissement de seuils de vigilance pour le suivi de ce phénomène ainsi qu'à une meilleure gestion du captage.

4. Bibliographie

Fan, R.E., Chen, P. H., Lin, C.J., 2005. Working set selection using second order information for training support vector machines. The Journal of Machine Learning Research 6, 1889-1918.

Kuhn, M., & Johnson, K. (2013). Applied predictive modeling (pp. 389-400). New York: Springer.

Metz, C.E., 1978. Basic principles of ROC analysis. In: Seminars in nuclear medicine. WB Saunders, pp. 283-298.

Platt, J.C., 1999 Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods, in: Smola, A.J., Bartlett, P., Schölkopf, B., Schuurmans, D. (Eds.), Advances in large margin classifiers. MIT Press, Cambridge, pp 61–74.

Powers, D.M., 2007. Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation, School of Informatics and Engineering, Flinders University of South Australia Adelaide, technical report SIE-07-001.

Savenije H.G., 2012. Salinity and tides in alluvial estuaries. Second edition.

Schölkopf, B., Smola, A.J., 2002. Learning with kernels: support vector machines, regularization, optimization, and beyond. MIT Press, Cambridge.

R Development Core Team, 2015. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, http://www.R-project.org/.

Vapnik V., 1998. Statistical learning theory. Wiley, New York.

Wang, W.C., Chau, K.W., Cheng, C.T., Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of hydrology 374(3), 294-306.

Wu, C.L., Chau, K.W., Li, Y.S., 2008. River stage prediction based on a distributed support vector regression. Journal of Hydrology 358(1), 96-111.

Yu, P.S., Chen, S.T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. Journal of Hydrology 328(3), 704-716.

Géosciences pour une Terre durable
brgm

Centre scientifique et technique
3, avenue Claude-Guillemin
BP 36009
45060 – Orléans Cedex 2 – France
Tél. : 02 38 64 34 34 - **www.brgm.fr**

Direction régionale Guyane
Domaine de Suzini – route de Montabo
BP 10552
97333 – CAYENNE – France
Tél. : 05 94 30 06 24