Modélisation de scénarios futurs de température de l'eau en milieu côtier et implications sur les infections potentielles par *Vibrio Parahaemolyticus et Vibrio Vulnificus* : application aux bancs coquilliers de l'estuaire et du Golfe du St-Laurent

> André St-Hilaire (INRS-ETE) Juin 2019





- Institut national de santé publique
- Ministère de la Santé et des Services sociaux

Les résultats et opinions présentés dans cette publication sont entièrement la responsabilité des auteurs et n'engagent pas Ouranos ni ses membres.

## **Table des matières**

Table des matièresI
Table des FiguresIII
Table des tableauxIV
Liste des abréviations et acronymesV
1.0. Introduction1
2.0. Mise en contexte2
3.0. Problématique et objectifs3
4.0. Revue de littérature4
4.1. Les facteurs influençant la prolifération des souches pathogènes des <i>V.</i>
Parahaemolyticus et V. Vulnificus4
4.1.1. Effet de la température de l'eau et la salinité de l'eau
<b>4.1.2.</b> Effet du changement climatique sur le taux d'infection par les <i>Vibrios</i> 9
<b>4.1.3.</b> Effet des téléconnexions10
5.0. Méthodologie et base de données11
5.1. Sélection des variables explicatives de la variation de la température de l'eau 11
5.2. Zone d'étude et base de données14
5.3. La modélisation de la température de l'eau par l'intermédiaire des modèles
statistiques17
5.3.1. Régression linéaire multiple17

	5.3.3.	Les Forêts Aléatoires	22
	5.3.4.	Facteur d'inflation de la variance	25
	5.3.5.	Sélection de variables (attributs)	25
	5.3.6.	Comparaison des modèles statistiques	27
5.	4. Acq	uisition des scénarios climatiques et modélisation des températures de l'e	au
fu	tures		29
5.	5. Car	tographie des zones à risque d'infection	31
6.0.	Résul	tats	.33
6.	1. Séle	ection des prédicteurs de la variation de la température de l'eau	33
6.	2. Per	formance des modèles testés	35
6.	3. Moo	lélisation des futures températures de l'eau	38
6.	4. Car	tographie des futures zones à risque d'infections	39
7.0.	Discu	ssion	.44
8.0.	Concl	usions et recommandations	.47
9.0.	Référ	ences	.49
Anr	iexe A		.54
Anr	iexe B		62

# **Table des Figures**

Figure 5-1 : Courbe de marée au niveau du marégraphe Charletown (IPE)15
Figure 5-2: Lissage de courbe de marée au niveau du marégraphe Cap aux meules (Iles de la Madeleine) 15
Figure 5-3: Localisation des stations de mesure de l'estuaire maritime et Golfe du Saint-Laurent
Figure 5-4: Illustration d'architecture de ANN de type MLP20
Figure 5-5 : Principe de base de k-fold-cross validation
Figure 5-6: Principe de Bagging
Figure 5-7: Structure des forêts aléatoires de régression <sup>60</sup>
Figure 6-1: Matrice de corrélations au niveau de la station Grande –Rivière
Figure 6-2 : Comparaison de performance des modèles des Réseaux des Neurones Artificiels (ANN) et des
Forêts Aléatoires (RF) au niveau de la station Grande Rivière
Figure 6-3: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois d'Août
Figure 6-4: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois d'Août
Figure 6-5: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et
Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois
d'Août
Figure B-1 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois du Juille
Figure B-2 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois du Juillet
Figure B-3: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et
Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois du
Juille
Figure B-4: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois de Septembre

Figure B-5 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et
Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois de Septembr
Figure B-6: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et
Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois de
Septembre

## **Table des tableaux**

Tableau 4-1: Tableau récapitulatif des seuils thermiques et de salinité en fonction de la prolifération des
vibrios pathogènes7
Tableau 5-1: Caractéristiques des scénarios RCP 66
Tableau 6-1: Résultats des critères de performance des modèles sur les données du test
Tableau 6-2 : Le nombre de jours au-dessus du seuil 15°C pour le scénario RCP8.5 des mois de juin et octobre
durant la période (2080-2100)
Tableau A-1: Liste des stations dans le Golfe et estuaire du Saint Laurent
Tableau A-2: Tableau récapitulatif des corrélations entre la température de l'eau et les prédicteurs
sélectionnées pour l'ensemble des stations étudiées56
Tableau A-3: Résultats des critères des performances des modèles testés suite à la sélection des meilleurs
attributs parmi tous prédicteurs sélectionnées57
Tableau A-4: Résultats des critères des performances des modèles testés en utilisant suite à la sélection des
meilleurs attributs parmi la Tair et la vitesse du vent58
Tableau A-5: Résultats de test de la tendance sur les futures températures de l'eau en utilisant le test du
Mann Kendall Modifié

## Liste des abréviations et acronymes

ANN	Artificial Neural Network						
Bagging	Bootstrap Aggregating						
CART	Classification And Regression Tree						
CDC	Centers for Disease Control and Prevention						
C.S.A.S	Canadian Science Advisory Secretariat						
ENSO	EL Nino-Southern Oscillation						
FDA	Food and Drug Administration						
Flow	La MM-120 jours du débit moyen journalier du fleuve						
	Saint-Laurent						
IDW	Inverse Distance Weighting						
INSPQ	Institut National de la Santé Publique du Québec						
IPE	Île-du-Prince-Édouard						
MEI	Multivariate ENSO Index						
MLP	Multi-Layer-Percepton						
MLR	Multiple Linear Regression						
MM-X jours	Moyenne Mobile (glissante) sur X jours						
NAO	North Atlantic Oscilation						
Nash	Le critère de Nash-Sutcliff						
ppt	parts per thousand, ou parties par millier						
R <sup>2</sup>	Le coefficient de détermination						
rBiais	Le biais moyen relatif						
RCP	Representative Concentration Pathways						
RF	Random Forest						
RFE	Recursive Feature Elimination						
RMSE	Root Mean Square Error (l'erreur quadratique moyenne)						
rRMSE	relative Root Mean Square Error (l'erreur relative						
	quadratique moyenne)						
SRES	Special Report on Emissions Scenarios						
SSTP	La température moyenne journalière de la surface d'eau						
Tair	La MM-3jours de la température moyenne journalière de						
	l'air						
Tidal_range	La MM-30 jours du marnage maximal journalier						
V. Parahaemolyticus	Vibrio Parahaemolyticus						
V. Vulnificus	Vibrio Vulnificus						
WSPD	La MM-30 jour de la vitesse moyenne journalière du vent						

## **1.0.Introduction**

Les changements climatiques risquent de menacer la santé humaine puisqu'ils vont possiblement influencer l'incidence et la prévalence de certaines maladies. Par exemple, le risque d'infection humaine par certaines bactéries marines nommées «*vibrios*», suite à la consommation des mollusques crus, pourrait s'accroitre durant le 21<sup>e</sup> siècle<sup>1-6</sup>. En effet, l'infection par ces pathogènes opportunistes provoque des gastroentérites dues aux *Vibrio Parahaemolyticus* et des infections sévères, notamment des septicémies, dues aux *Vibrio Vulnificus*. Dans ce dernier cas, le taux de fatalité peut être important<sup>7-9</sup>. Ces bactéries sont présentes dans les eaux chaudes et moins salées des estuaires et les zones côtières. En fait, ces *vibrios* sont très sensibles à la variation de la température de l'eau (voir section 4). C'est pourquoi les changements anticipés du climat et donc l'augmentation de la température pourrait avoir un impact positif sur leur prolifération.

Aux États-Unis, plus de 95 % de décès, liés à la consommation des fruits de mer, sont causés par le *V.Vulnificus*<sup>10</sup>. Le CDC (Centers for Disease Control and Prevention) estime une moyenne annuelle d'environ 100 cas d'infections par *V. Vulnificus*, entrainant 50 décès par an.

Au Canada, une étude de surveillance sur la diversité et la dynamique de la communauté des *vibrios* a été effectuée récemment au niveau des côtes canadiennes (Colombie britannique (côte pacifique), Nouvelle Écosse (côte atlantique) et la Gaspésie). Cette étude met en évidence d'une part l'émergence de *Vibrio.Cholerae* dans les estuaires tempérés du Canada et l'augmentation de la détection des souches pathogènes des *V.Parahaemolyticus* au niveau des mollusques bivalves récoltées au Canada (palourdes, moules et huitres) durant les mois les plus chauds de 2006-2016<sup>11</sup>. Cette étude a révélé la présence de deux tendances ascendantes du risque des *V.Cholerae et V.Parahaemolyticus* dans les eaux côtières canadiennes<sup>11</sup>.

Étant donné que la récolte des mollusques (pétoncles, moules, huitres) est en pleine expansion en Québec et à l'Ile-Prince-Édouard, plus précisément dans le Golfe du Saint Laurent, l'analyse du risque d'infection est primordiale pour protéger la santé humaine et la qualité des produits maritimes.

Le présent projet de recherche a pour but de modéliser les futures températures de l'eau au niveau des bancs coquillers du Golfe et de l'estuaire du Saint Laurent pour qu'on puisse cartographier les futures zones à risque d'infection. En premier lieu, on a essayé de comprendre la relation entre la température de l'eau et la croissance des souches pathogènes des *vibrios*, puis on a identifié les autres facteurs influençant leurs proliférations. En deuxième lieu, on a cherché les variables explicatives de la variation de la variable d'intérêt (température de l'eau). En troisième lieu, on a procédé à la modélisation de la température de l'eau par des modèles statistiques. Finalement, on a précisé les métriques thermiques qui vont être utilisées pour l'identification des zones à risque d'infection.

Sous la direction d'André St-Hilaire, Taha Ouarda et Benoit Levesque, ce projet a été mené dans le cadre de la gestion du risque de la variation de la température de l'eau dans un contexte de changements climatiques et ses implications potentielles sur la prolifération des *vibrios* pathogènes.

### 2.0.Mise en contexte

Les bactéries de type *Vibrio* sont susceptibles de causer des infections zoonotiques. Elles sont présentes naturellement dans le milieu maritime et plus particulièrement dans les eaux estuariennes et côtières. On peut distinguer deux espèces de *Vibrio*: Les espèces cholériques causant le choléra et dont l'agent pathogène sont les souches des *Vibrio*. *Cholerae*, et les espèces non cholériques notamment *les Vibrio Parahaemolyticus* et *Vibrio Vulnificus*.

Ces *vibrios* non cholériques sont des pathogènes associés à la consommation des mollusques crus sujets à être contaminés (les huîtres, moules, myes, pétoncles etc....).

Les signes cliniques de l'infection par *Vibrio* se manifestent généralement 12 heures après l'ingestion de l'aliment contaminé. Pour *V.Vulnificus*, le taux de fatalité peut être important <sup>12</sup>. Les *V.Parahaemolyticus* sont à l'origine de gastroentérites alors que les *V.Vulnificus* provoquent des infections graves (septicémies, infection hépatique...) qui peuvent être mortelles.

L'abondance de ces bactéries marines est fortement corrélée avec la température de l'eau, avec des concentrations croissantes et proportionnelles au réchauffement saisonnier des eaux <sup>13</sup>. Étant donné que la prolifération de ces *vibrios* dans leur environnement naturel dépend de la température, ce groupe de pathogènes représente un baromètre important et tangible du changement climatique au niveau des systèmes maritimes <sup>14</sup>. De plus, plusieurs recherches prédisent que le changement climatique va augmenter le nombre d'infections par ce type de *vibrios* <sup>15,16</sup>.

## **3.0.Problématique et objectifs**

La récolte des mollusques (pétoncles, moules, huitres..) est une activité commerciale importante au Québec et à l'Île-du-Prince-Édouard. Les bancs coquillers sont localisés sur les zones côtières de l'estuaire et de Golfe du Saint Laurent notamment Rimouski, la Baie des Chaleurs, la Côte Nord, Iles de la Madeleine et à l'Île-du-Prince-Édouard. Étant donné que les *vibrios* pathogènes menace la santé humaine d'une part et la qualité des industries de la conchyliculture à long terme, la modélisation de la température de l'eau est primordiale pour les protéger.

Pour atteindre cet objectif, on a utilisé des modèles statistiques paramétriques (régression linéaire multiple) et des modèles d'apprentissage automatique (Forêts Aléatoires et Réseaux des Neurones Artificiels) pour modéliser la température de l'eau en fonction des variables météorologiques (température d'air, vitesse du vent), océanographiques (niveau des marées), atmo-océanographiques (téléconnexions) et hydrologiques (débit). Une fois que ces modèles ont été calés, des scénarios climatologiques anticipés ont été utilisés pour simuler les futures températures de l'eau afin de pouvoir cartographier les futures zones à risque d'infection. Les objectifs principaux du présent projet étaient :

- L'évaluation et la compréhension de la relation entre la prolifération des *vibrios* pathogènes et les seuils thermiques.
- La modélisation de la température de l'eau aux sites de suivi de la température à l'aide de thermographes en fonction des variables explicatives (océanographiques, météorologiques, indices climatiques).
- La comparaison de modèles statistiques et la sélection du modèle le plus performant.
- La cartographie des futures zones à risque en fonction des seuils thermiques, et possiblement d'autres variables comme la salinité.

## 4.0. Revue de littérature

#### 4.1. Les facteurs influençant la prolifération des souches pathogènes des

#### V. Parahaemolyticus et V. Vulnificus

#### 4.1.1. Effet de la température de l'eau et la salinité de l'eau

La recension bibliographique de la relation entre la prolifération des *vibrios* pathogènes, et les seuils de température de l'eau est une étape cruciale pour éventuellement cartographier les zones à risque actuelles et futures.

Plusieurs études ont affirmé que les *vibrios* préfèrent les eaux chaudes et moins salées <sup>1,17-20</sup>. En effet, Motes, et al. ont trouvé que la densité des *V.Vulnificus* dans les huitres, récoltés à partir des sites de la côte du Golfe et la côte atlantique des États-Unis, est fortement corrélée avec la température, qui explique la majorité de la variation de sa densité <sup>19</sup>. En fait, ils ont construit un modèle de régression linéaire pour chercher la liaison de variation de la densité des *V.Vulnificus* des *vibrios* (organisme/g) dans les huitres récoltées en fonction de la température et la salinité. Suite à ce modèle, ils ont trouvé qu'à peu près 60 % de la variation de la densité des *vibrios* est expliquée par la variation de la température et 10 % de la variation des *vibrios* est expliquée par la variation de la salinité.

À partir d'un certain seuil thermique variant entre 13 et 26°C, les souches des *vibrios* croissent rapidement <sup>17</sup>. En plus de la température, le pouvoir pathogène de ces bactéries n'est déclenché qu'en salinité relativement faible inférieure à 25 ppt <sup>18,19</sup>. Notons que pour une température optimale de 37°C, correspondant à une croissance optimale des *vibrios*, et une salinité relativement élevée (supérieure à 30 ppt) on constate une décroissance de la concentration des *vibrios* et un arrêt de leur prolifération pour une salinité très élevée (supérieure à 80 ppt *in vitro*) <sup>19,21,22</sup>

En effet, Motes, et al. <sup>19</sup> n'ont pas détecté des cas d'infection des huîtres par les *V.Vulnificus* pour un climat froid (T<15°C) au niveau des sites de culture aux États-Unis, localisés sur la côte Nord du Golfe du Mexique et la côte Atlantique. En plus de la température, ces bactéries sont plus abondantes dans les sites les moins salés (<25ppt) <sup>19</sup>. En outre, plus de 95% des décès chez l'humain, suite à la consommation d'huîtres crues aux États-Unis, se produisent lorsque la température de l'eau de culture des coquillages dépasse 20°C <sup>23,24</sup>. Ces bactéries ne peuvent pas supporter des salinités élevées, elles sont donc concentrées aux eaux saumâtres <sup>23,24</sup>. Ainsi, on trouve les *V. Vulnificus* principalement dans les zones côtières et les estuaires.

Il en va de même pour les *V.Parahaemolyticus*. Ils sont concentrés dans les estuaires et les eaux côtières (Liu, 2016). La température minimale d'eau de mer qui déclenche l'infection des huîtres par ce type de *vibrio* est 15°C, malgré que cette espèce peut croître dans une T<10°C <sup>25</sup>.

Le tableau ci-dessous collige les résultats, confirmés par des recherches antérieures, qui précisent des métriques thermiques et de salinité en fonction des cas d'études de la prolifération des *vibrios* pathogènes soit *in situ* ou *in vitro*.

Type de Vibrios	Type de							
	culture	Source	Température de l'eau (°C)			Salinité ( <i>ppt</i> )		
			Min	Max	Optimale	Min	Max	Optimale
	In vitro	Kelly <sup>26</sup>	13 : absence des <i>vibrios</i> pathogènes	42 : arrêt de la croissance des <i>vibrios</i>	37		85 arrêt de la croissance des <i>vibrios</i>	≤20
		Leonard <sup>27</sup>	8	43			50	
		Kaspar and Tamplin <sup>18</sup>	0 à 4	≥ 30	Entre 13 et 22	5ppt	>25 réduction de la concentration des <i>vibrios</i>	Entre 5 et 25
V.Vulnificus	In situ	Kelly <sup>26</sup>	12 : absence des <i>vibrios</i> pathogènes		31 (> 25)		>16 réduction de la concentration des <i>vibrios</i>	<16
		Motes and DePaola <sup>28</sup>	patriogenes		$\geq 20$			<u></u> 10 ≤30
		Motes, et al. <sup>19</sup>	<15 : absence des <i>vibrios</i> -		26 (entre 15 et 26)		>32 réduction de la concentration des <i>vibrios</i>	<25nnt
			pathogenes		0(20)		>30	PPt
		WHO <sup>23</sup>			≥20 Présence		réduction de la	<30
		Nishibuchi and DePaola <sup>24</sup>			des <i>vibrios</i> pathogènes		concentration des vibrios	

Tableau 4-1: Tableau récapitulatif des seuils thermiques et de salinité en fonction de la prolifération des vibrios pathogènes

	In vitro	Leonard <sup>27</sup>	5	45,3			100 : arrêt de la croissance des <i>vibrios</i>	
							90 : arrêt de la	
		Liu, et al. <sup>22</sup>			37	5	vibrios	30
V.Parahaemolyticus		McLaughlin <sup>3</sup>			≥15 Apparition des <i>vibrios</i> pathogènes et infections			
	In situ	Gilliss, et al. <sup>15</sup>	10		>10	10	>34 : réduction de la concentration des <i>vibrios</i>	23
		Baker-Austin, et al. <sup>17</sup>			≥15			≤25
		Motes, et al. <sup>19</sup>	<15 : absence des <i>vibrios</i>		Entre 15 et			
			pathogènes		26			≤25

Dans le cadre de cette étude, 15°C sera considéré comme une limite inférieure d'alerte d'un risque d'infection. Quant à la température optimale pour un maximum de croissance de ces *vibrios*, la majorité des études et des analyses *in vi*tro affirment que ce seuil est atteint à une T=37°C, température qui n'a jamais été observée dans le Golfe du Saint-Laurent. Une température de 26 °C *in si*tu a été observée lors d'étude de l'abondance des *vibrios* dans les huitres cultivés aux États-Unis sur la côte Nord du Golfe de Mexique et la côte atlantique <sup>19</sup>.

#### 4.1.2. Effet du changement climatique sur le taux d'infection par les Vibrios

Il y a plusieurs indications que les changements climatiques ont engendré des évènements climatiques extrêmes incluant les canicules et une hausse de l'intensité des précipitations, en particulier durant le dernier demi-siècle. Ces événements et deviennent de plus en plus fréquents<sup>29</sup>.

Plusieurs études récentes ont mis en évidence le rôle de ces extrêmes climatiques dans l'augmentation du risque d'infection par les *vibrios* <sup>5</sup>. Les auteurs ont mentionné que les canicules dans le nord européen sur les trois décennies passées ont amené des dépassements du seuil de 18 °C qui correspondent à une augmentation significative des infections des plaies causées par les *V. Vulnificus*.

Prenant le cas des canicules, les plus intenses et anormales jamais enregistrées en Scandinavie se sont produites en 2014 et correspondent aux nombres d'infections humaines par les *vibrios* les plus élevés en Finlande et Suède <sup>30</sup>. De même, pour la Nouvelle-Calédonie, des précipitations intenses ont engendré une diminution de la salinité donc des conditions favorables pour le développement des bactéries, ce qui a mené à un premier cas d'infection humaine par les *V*. *Vulnificus* en 2008 <sup>31</sup>.

Même un petit réchauffement peut affecter le taux de croissance des *V. Parahaemolyticus* et réduira la restriction d'impact associée à la période hivernale sur leur cycle de croissance. Une augmentation moyenne de 1,5 °C de la température de l'eau va étendre la saison et la zone géographique propices à des proliférations avec une abondance plus répandue et un risque plus élevé <sup>32</sup>.

#### 4.1.3. Effet des téléconnexions

Les téléconnexions représentent des anomalies qui influencent la variabilité de la circulation atmosphérique et océanographique à grande échelle spatiale et temporelle. Ces oscillations climatiques engendrent des anomalies de la température de l'air, de la température de surface de la mer ainsi que des précipitations. On peut les décrire au moyen de certains indices climatiques. Plusieurs oscillations climatiques ont été responsables de l'apparition de cas d'infections par ces *vibrios*, en particulier certaines phases de l'ENSO (EL Nino-Southern Oscillation), provoquant des anomalies climatiques dans l'océan pacifique.

Pour la première fois, une épidémie associée à *V. Parahaemolyticus* au Pérou en 1997 a provoqué des infections humaines inattendues dues à la formation des colonies pandémiques qui ont occupé toute la zone côtière du pays<sup>33</sup>.

Les investigations subséquentes ont montré que les conditions océanographiques pendant l'apparition des épidémies péruviennes associées à *V. Choléra* et à *V. Parahaemolyticus* ont coïncidé avec les épisodes d'El Nino<sup>4,33</sup>. Ce phénomène était caractérisé par le déplacement des eaux chaudes et moins salées du Pacifique Ouest vers les zones côtières d'Amérique du Sud. Similairement, en Alaska en 2004, plusieurs gastroentérites ont été signalées pour la première fois pendant le mois de juillet suite à la consommation d'huîtres crues<sup>25</sup>. Ces cas d'infections ont été déclenchés suite à l'arrivée extraordinaire des eaux chaudes (>15°C) de la côte du Pacifique vers le pôle en passant par les côtes d'Alaska<sup>25</sup>. Aucune infection n'a été détectée depuis 2005 suite au retour des étés typiques et frais d'Alaska<sup>4</sup>. De plus, il est important de mentionner que l'absence d'infection peut être due au changement de la culture des huîtres sous des températures inférieures à 10°C <sup>4</sup>.

Pour les États-Unis, le nombre d'infections humaines par les *vibrios* augmentent depuis 2000 pour plusieurs raisons. D'abord, la fréquence et la sévérité des extrêmes des températures et les précipitations s'élèvent de plus en plus surtout dans le sud-ouest à cause des anomalies climatiques et de l'élévation de températures dans l'ouest du Pacifique <sup>34</sup>. Ces anomalies climatiques coïncident aussi avec les variations de deux phases d'ENSO : El Nino et la Nina <sup>35</sup>. Suite à La Nina en 1998, 11 cas d'infection par les *V. Vulnificus* ont été détectés durant le mois de novembre dans le Golfe de Mexique <sup>23,36</sup>.

### 5.0.Méthodologie et base de données

#### 5.1. Sélection des variables explicatives de la variation de la température de

#### l'eau

L'océanographie physique et le climat dans l'estuaire et le Golfe du Saint-Laurent relie la température de la surface d'eau (SSTP), la température d'air et la vitesse du vent suivant un cycle saisonnier. En effet, en été et printemps, le rayonnement solaire est une source de chaleur pour la surface des eaux, en plus des eaux douces arrivant de l'estuaire du Saint-Laurent (le débit d'eau douce du fleuve Saint Laurent et la fonte de neige). Par la suite, la température de la surface d'eau atteint son maximum au mois d'août ou juillet. Par contre, en automne et en hiver, la faible température de l'air et la vitesse du vent élevée dissipent la chaleur de la couche superficielle.

Le fleuve du Saint Laurent est le deuxième plus grand fleuve du l'Amérique du Nord<sup>37</sup>, son débit moyen est d'environ 12100 m<sup>3</sup>/s<sup>38</sup>. IL contribue avec plus de 80% des eaux douces entrant dans le Golfe à travers l'estuaire. L'écoulement de l'eau douce diminue la salinité de l'estuaire et augmente la température de la surface de l'eau au printemps et au début de l'été.

Concernant les marées, le type de marées dans l'estuaire et le Golfe du Saint Laurent est semi diurne à inégalité diurne. Il y a donc deux pleines mers et deux basses mers avec des hauteurs différentes par jour. En basse mer, le volume d'eau est moindre, ce qui peut conduire à une température de surface plus élevée que celle en pleine mer au niveau des zones côtières. Pour prendre cette relation en considération, on a essayé d'introduire le facteur du marnage qui représente la différence entre une pleine mer et une basse mer successives pour détecter la variation de la température de la surface en fonction du niveau d'eau. Puisqu'il y a deux marnages par jour, on a décidé d'extraire deux variables qui peuvent influencer la température de l'eau : le marnage moyen journalier et le marnage maximal journalier qui peuvent influencer la variation de la température de l'eau journalière.

Au sujet des indices climatiques, une recherche bibliographique a été complétée pour sélectionner les indices qui peuvent influencer le climat au Canada et plus précisément la température de l'eau.

La variabilité et les tendances du climat canadien sont influencées par les oscillations océaniques et atmosphériques à grande échelle spatiale et temporelle. En effet, les principales oscillations qui ont probablement un effet sur le climat canadien sont EL-Nino oscillation australe (ENSO –EL Nino-Southern Oscillation), l'Oscillation Décennale du Pacifique (PDO), la téléconnexion Pacifique-Amérique du Nord (PNA), l'oscillation de l'Atlantique du Nord (NAO), l'oscillation de l'Arctique (AO), l'oscillation multi décennale de l'Atlantique (AMO), l'oscillation d'Est Pacifique / Nord Pacifique (EP/NP), l'oscillation d'Ouest Pacifique (WP) et l'oscillation d'Est Pacifique (EP) <sup>39</sup>. Plusieurs autres recherches <sup>39-41</sup> ont confirmé l'influence d'un ou de ces indices mentionnés sur le climat canadien (précipitations extrêmes, variations de débits, anomalies des températures d'air).

Dans son rapport sur les conditions océanographiques dans la zone atlantique, le Secrétariat Canadien de Consultation Scientifique a mentionné que le Golfe du Saint-Laurent est affecté par la variation de l'oscillation du Nord Atlantique (NAO)<sup>42</sup>. En effet, les hivers chauds couplés avec les vents sud-ouest les plus forts, coïncident avec la phase négative de l'indice NAO, ce qui engendre la réduction de la couverture de glace au sud des Iles-de-la-Madeleine.

Certains indices sont calculés au pas du temps journalier et d'autres au pas de temps mensuel. On a sélectionné seulement les indices journaliers dans le but de tester la corrélation entre la température de l'eau journalière et les moyennes mobiles de ces indices.

Afin de bien identifier les meilleurs prédicteurs, on a testé les corrélations croisées linéaires (Pearson) et non linéaires (Spearman), en prenant en considération le décalage du temps (lags), entre les Moyennes Mobiles (MM) sur différentes durées (3,7, 15, 30, 60, 90, 120 jours) des variables explicatives (Température de l'air, la vitesse du vent, le marnage maximal journalier, le marnage moyen journalier et les indices climatiques) et notre variable d'intérêt (la température moyenne journalière de la surface d'eau). Les prédicteurs sélectionnés étaient significativement et plus fortement corrélés avec la température de l'eau.

La prolifération des *Vibrios* pathogènes se fait pendant la période estivale (pour une température supérieure à 15°C), puis ils entrent dans une phase de VNC (Viable Non Cultivable), c'est-àdire les cellules ne se multiplient pas mais restent vivantes pendant la période froide (T<5°C). Donc dans cette étude, on va se concentrer seulement sur les températures estivales, i.e. entre les mois de juin et octobre.

#### 5.2. Zone d'étude et base de données

Notre zone d'étude comprend les côtes de l'estuaire maritime (Rimouski) et le Golfe du Saint-Laurent (Côte Nord, Gaspésie, la Baie des Chaleurs, Île-du-Prince-Édouard, Iles-de-la-Madeleine).

Les données existantes des thermographes côtiers dans le Golfe et l'estuaire du Saint-Laurent ainsi que celles des bouées sont fournies par l'Institut de Maurice-Lamontagne et Pêches et Océans Canada. La profondeur moyenne de ces thermographes est d'environ 1.5m.

Grâce à une collaboration avec Merinov, centre d'innovation de l'aquaculture et des pêches de Québec, nous avons accès à une base de données des températures de l'eau des thermographes situés au niveau des sites de bancs coquilliers. La liste des thermographes et des bouées est présentée dans le tableau A-1 de l'annexe A.

Les données des stations météorologiques d'Environnement et Changement Climatique Canada disponibles dans l'estuaire et Golfe du Saint-Laurent ont été téléchargées à l'aide du logiciel Cygwin.

Les données des indices climatiques étaient disponibles en ligne (Earth System Research Laboratory of the National Oceanic and Atmospheric Administration <u>ftp://ftp.cpc.ncep.noaa.gov/cwlinks)</u>.

Les données des marégraphes étaient disponibles sur le site de Pêches et Océans Canada. Certains marégraphes comprennent une courbe sinusoïdal lisse, ce qui nous a permis de détecter la pleine et la basse mer pour pouvoir calculer le marnage journalier. Par exemple, la figure 5-1 représente la courbe de marée au niveau d'IPE. Cependant pour d'autres marégraphes, ce n'était pas si simple. Par exemple, pour Cap-aux-meules, située au niveau des Iles de la Madeleine, la courbe est bruitée, alors le calcul du marnage devient plus difficile. Donc, on a utilisé un lissage de type polynomiale, par l'intermédiaire du filtre de Savitzky Golay, pour retirer ce bruit et pouvoir ensuite calculer le manage journalier. La figure 5-2 montre la courbe de marée avant et après le lissage à cette station.



Figure 5-1 : Courbe de marée au niveau du marégraphe Charletown (IPE)



Figure 5-2: Lissage de courbe de marée au niveau du marégraphe Cap aux meules (Iles de la Madeleine)

Afin de tester l'influence du débit fluvial du fleuve Saint Laurent sur la température de l'eau, on a utilisé les débits provenant de la station hydrométrique Lasalle. Les données étaient fournies en ligne par Environnement Canada.

La carte 5-3 ci-dessous représente les sites des thermographes côtiers, des bouées et des thermographes situés au niveau des bancs coquillers (zones de récoltes).



Figure 5-3: Localisation des stations de mesure de l'estuaire maritime et Golfe du Saint-Laurent

Dans ce rapport, nous allons présenter à titre d'exemple les résultats à la station Grande Rivière (thermographe testé sur la carte). Cette station est sélectionnée à cause de la longueur relativement grande de sa période d'enregistrement de données de la température de l'eau (1995-2016). Pour le reste des stations, des tableaux récapitulatifs des résultats sont présentés dans l'Annexe A.

# 5.3. La modélisation de la température de l'eau par l'intermédiaire des modèles statistiques

Les processus physiques, expliquant la relation entre la température de l'eau et les variables explicatives (météorologiques, océanographiques, indices climatiques, hydrologiques), sont assez complexes à modéliser à travers un modèle déterministe. Les approches statistiques paramétriques (régression linéaire multiple) et non paramétriques (les Réseaux des Neurones Artificiels, les Forêts Aléatoires) sont des alternatives souvent plus simples pour élaborer des modèles avec un nombre restreint de prédicteurs.

Les modèles d'apprentissage automatiques utilisés sont complètement différents. En effet, le modèle des Réseaux des Neurones Artificiels (Artificial Neural Network-ANN, en anglais) est inspiré par l'architecture du système nerveux formé par des couches des neurones. Par contre, le modèle des Forêts Aléatoires (Random Forest-RF, en anglais) est basé sur le principe de la méthode CART (Classification And Regression Tree), qui représente des arbres de décision binaire.

#### 5.3.1. Régression linéaire multiple

La régression linéaire multiple (Multiple Linear Regression-MLR, en anglais) est un modèle statistique paramétrique exprimé par l'équation (1) :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_{p-1} X_{p-1} + \varepsilon$$
(1)

Avec :

Y est la variable dépendante

X sont les variables explicatives

 $\beta_0$  est la constante (l'ordonnée à l'origine)

 $\beta_1 \dots \beta_{p-1}$  sont les paramètres de modèle (à estimer)

ε est l'erreur aléatoire de modèle

*p* est le nombre de paramètres (le nombre de variables explicatives plus1)

*n* est le nombre d'individus

Le modèle MLR est efficace pour modéliser seulement les relations linéaires entre les variables explicatives et la variable d'intérêt. En plus, malgré le fait que la relation entre la température de l'eau et la température de l'air est souvent considérée linéaire, pour des températures élevées cette relation devient non linéaire <sup>43</sup>. Pour cette raison, les modèles non-linéaires sont fréquemment privilégiés.

#### 5.3.2. Les Réseaux des Neurones Artificiels

Le modèle ANN est inspiré par l'architecture de système nerveux dont son élément de base est le neurone qui représente une fonction algébrique non linéaire.

Plusieurs algorithmes de modèle ANN ont été développés prenant l'exemple de Hopfield<sup>44</sup>, Hamming<sup>45</sup>, ART (Adaptative Resonance Theory)<sup>46</sup> et le perceptron à une couche<sup>47</sup> et multicouche<sup>48</sup>. L'algorithme le plus utilisé est le perceptron multicouche (MLP : Multilayerpercepton)<sup>49</sup>. Il est formé par une couche d'entrée puis une ou plusieurs couches cachées, suivant la complexité de la relation entre les variables d'entrée et de la sortie, et une couche de sortie. En effet, la couche d'entrée contient les variables explicatives et celle de sortie comprend la variable dépendante simulée. Le principe de modèle ANN consiste d'abord à calculer la somme pondérée des entrées plus un biais. Ces poids, associés pour chaque entrée, représentent la force de liaison entre les variables explicatives et la variable d'intérêt. Puis, la sortie de signal (somme pondérée) passe par une fonction d'activation (transfert) pour générer la sortie de modèle désirée. Cette relation est expliquée par l'équation (2):

$$Z_i = \sum_{i=1}^n w_i X_i + b \tag{2}$$

La fonction de transfert est une fonction non linéaire permettant de sélectionner les neurones qui peuvent entrer dans le calcul de la sortie du modèle et le plus important d'y introduire la non linéarité. En fait, l'ANN n'est qu'un modèle de régression linaire simple en absence des fonctions d'activation non linéaires. Ces transformations non linéaires permettent au modèle de comprendre les relations les plus complexes entre les variables explicatives ( $X_i$ ) et la variable d'intérêt. La fonction d'activation souvent utilisée est de forme sigmoïde représentée par l'équation (3) :

$$f(Z_i) = \frac{1}{1 + e^{-Z_i}}$$
(3)

L'architecture des réseaux des neurones de type MLP est expliquée par la figure 5-4 suivante:



Figure 5-4: Illustration d'architecture de ANN de type MLP

L'ANN du type MLP utilise la rétropropagation du gradient de l'erreur comme algorithme d'apprentissage supervisé. Ce processus permet de calculer l'erreur en comparant la sortie de modèle et les données observées de la variable souhaitée. Puis, un signal d'erreur se produit et se propage vers la couche d'entrée pour ajuster les poids, associés aux variables explicatives, jusqu'à ce qu'il minimise cet erreur et améliore l'ajustement des sorties simulées aux données réelles.

Dans la présente étude, on a utilisé le modèle ANN du type MLP avec une seule couche cachée en utilisant la rétropropagation du gradient de l'erreur. En effet, ce modèle est utilisé couramment dans la modélisation hydrologique <sup>50</sup> et dans la modélisation de température de l'eau des rivières <sup>49-51</sup>. Concernant la fonction d'activation, on a pris la fonction sigmoïde pour les neurones de la couche cachée et la fonction linéaire pour la couche de sortie.

Pour réduire le problème du surapprentisage, l'une des solutions proposées est l'utilisation de la validation croisée (k-fold-cross validation). En effet, l'ensemble de données est divisé en deux parties : les données d'entrainement et les données du test qui constitue la partie de données inconnue pour les modèles. La validation croisée consiste à diviser la partie apprentissage en k sous-ensembles. À chaque fois, l'un des sous-ensembles est utilisé pour la validation et le reste (k-1) forme un ensemble d'apprentissage. Ensuite, l'erreur est moyennée sur l'ensemble des k sous-ensembles. Donc, le modèle dans la validation croisée est ajusté k fois et testé k fois sur l'ensemble de données. On peut, en plus d'augmenter le nombre de sous-ensembles, répéter la validation croisée n fois, et donc l'erreur sera moyennée sur les sous-ensembles k\*n.

En plus de la validation croisée, on a procédé à utiliser une régularisation des paramètres d'hyperparamétrisation de chaque modèle à l'intérieur de la validation croisée afin de trouver les paramètres optimaux.

Dans notre étude, on a choisi d'utiliser 10 times-10 fold cross validation recommandée par Witten, et al. <sup>52</sup> pour réduire le surapprentissage des modèles.

La technique de la validation croisée est expliquée par la figure 5-5 ci-dessous :



Figure 5-5 : Principe de base de k-fold-cross validation

En fait, plusieurs paramètres peuvent être régularisés selon le type d'algorithme utilisé. Dans notre cas, on a choisi de contrôler le nombre des neurones cachés et le paramètre de la dégradation des pondérations ou « weight decay ». En effet, l'ajout de plusieurs couches ou même des plusieurs neurones au niveau d'une seule couche cachée pose le problème d'avoir un modèle surentrainé. Le « weight decay » exerce une pénalité sur la fonction d'erreur à minimiser, afin de réduire la magnitude des poids entre les neurones et donc minimise le risque de surapprentissage. En effet, Geman, et al. <sup>53</sup>Krogh and Hertz <sup>54</sup> ont démontré mathématiquement qu'un modèle avec une forte pondération des entrées est inefficace et ne peut pas être généralisé. Ces paramètres sont optimisés en utilisant une validation croisée.

#### 5.3.3. Les Forêts Aléatoires

Les forêts aléatoires (ou forêts d'arbres décisionnels) constituent un algorithme d'apprentissage automatique récent des années 2000 proposé par Breiman <sup>55</sup>.

Cet algorithme est une amélioration de l'algorithme des arbres décisionnels CART (Classification And Regression Tree), proposé pour la première fois par Breiman, et al. <sup>56</sup>, en combinant au CART l'algorithme de Bagging (Bootstrap aggregating) de Random Subspace.

En effet, le CART est formé par des arbres de décision binaire utilisés pour la prédiction de la réponse d'une variable continue ou catégorielle <sup>57</sup>. L'arbre contient initialement les données au niveau du nœud racine. Puis, ce nœud est divisé en deux nœuds filles qui sont eux-mêmes divisées ainsi de suite <sup>57</sup>. Cette technique est appelée le partitionnement récursif (Recursive Partitioning - RPART). La division se poursuit jusqu'à ce que les données au niveau des nœuds terminaux soient homogènes.

Le Bagging ,introduit par Breiman <sup>58</sup>, consiste d'abord à tirer des échantillons aléatoires à partir de l'échantillon initial. On appelle cette étape Bootstrap. Ensuite, on entraine le modèle de

CART, c'est-à-dire on construit un arbre de décision, pour chaque échantillon bootstrap. Puis, on effectue une agrégation en moyennant les prédictions obtenues par chaque modèle du CART en cas de régression, ou en votant pour la classe la plus fréquente en cas de la classification. Cet algorithme est une procédure de réduction de la variance du modèle en combinant les prédictions produites par différents modèles entrainées sur différents échantillons aléatoires. Il permet ainsi d'obtenir des prédictions plus précises qu'un modèle individuel. La figure 5-6 ci-dessous illustre le principe de Bagging.



Figure 5-6: Principe de Bagging

Par contre, le Random Subspace, introduit par Ho<sup>59</sup>, consiste à tirer aléatoirement k variables parmi les n variables prédicteurs à chaque division du nœud de l'arbre de telle sorte à minimiser l'erreur du modèle. Donc l'algorithme final du modèle des Forêts aléatoires comprend les mêmes étapes que l'algorithme du Bagging, mais en ajoutant l'algorithme de Random Subspace dans la construction de modèle CART (arbre) pour chaque échantillon.

La combinaison de ces deux algorithmes permet de construire des arbres de décision indépendants (moins corrélés), puisqu'ils sont entrainés aléatoirement sur différents échantillons et attributs (variables explicatives)<sup>55</sup>.

Dans notre étude, on va utiliser le modèle des forêts aléatoires de régression expliqué dans la figure 5-7 ci-dessous :



Figure 5-7: Structure des forêts aléatoires de régression <sup>60</sup>

Les paramètres de régularisation de ce modèle sont le nombre d'arbres arbitraires à construire, le nombre de variables sélectionnées aléatoirement pour chaque division et le nombre minimal des observations au niveau des nœuds terminaux.

Au contraire des Réseaux des Neurones Artificiels, Breiman <sup>55</sup> a prouvé que son modèle n'est pas sensible au problème du surapprentissage. En effet, chaque arbre de la forêt est construit sur une fraction de données tirée aléatoirement de la base de données d'apprentissage. En fait, cette fraction est divisée en deux fractions, une pour l'apprentissage (63,2%) qu'on appelle « in bag » et le reste (36,8%) pour la validation « out of bag ». Donc, pour chaque arbre développé le modèle est entrainé et validé en minimisant l'erreur qu'on appelle « out of bag error ». Cette technique permet de fournir une validation intégrée sans perdre une fraction de la base de données pour la validation <sup>55</sup>.

#### 5.3.4. Facteur d'inflation de la variance

L'occurrence de la multicolinéarité entre les prédicteurs pourrait poser problème pour les modèles linéaires. En effet, elle augmente la variance des coefficients estimés et affaiblit la puissance statistique d'un modèle linéaire pour l'identification des prédicteurs significatifs<sup>61</sup>. Parmi les indicateurs utilisés pour la vérification de l'existence de la multicollinéarité<sup>62</sup>, le facteur d'inflation de la variance (VIF) <sup>63,64</sup> est utilisé dans cette étude. Il est exprimé par la formule suivante :

$$VIF_k = \frac{1}{1 - {R_k}^2} \tag{3}$$

VIF<sub>k</sub> représente le facteur de l'inflation de la variance du prédicteur k et  $R_k^2$  représente le coefficient de détermination pour une régression linéaire dont le prédicteur k représente la variable réponse et le reste de prédicteurs représentent les variables explicatives. Une valeur de VIF dépassant une valeur critique (10) indique la présence d'une multicolinéarité significative. Cette dernière réduit la fiabilité du modèle de régression linéaire.

#### **5.3.5.** Sélection de variables (attributs)

La sélection de variables est un processus de réduction de dimension. Elle permet de choisir un sous ensemble optimal des prédicteurs pertinents et non redondants à partir de l'ensemble des variables explicatives utilisées en optimisant un critère de performance.

Dans cette étude, l'algorithme RFE (Recursive Feature Elimination) est utilisé dans la sélection d'attributs les plus pertinents et non redondants. Le RFE est une technique de type Wrapper basée sur l'approche descendante. Durant le processus d'apprentissage, le RFE permet de sélectionner le meilleur ensemble de prédicteurs qui contribue le plus à une meilleure performance prédictive du modèle.

Tout d'abord, il entraîne le modèle en utilisant tous les prédicteurs, puis il élimine séquentiellement la variable la moins importante en fonction de sa contribution à la précision prédictive du modèle, également appelée Variable Importance. Ensuite, il entraine de nouveau le modèle par l'intermédiaire de l'ensemble réduit de prédicteurs.

Le RFE sélectionne automatiquement le meilleur sous-ensemble de prédicteurs qui optimise un critère de performance sélectionné. Dans cette étude, le critère considéré est le RMSE. Cependant, l'ajout de prédicteurs n'assure pas une réduction significative du RMSE. Ainsi, afin de sélectionner la taille de sous-ensemble optimale, nous utilisons une tolérance de 10% comme étant un critère d'arrêt. Il représente le pourcentage de perte de performance. Par conséquent, si l'élimination d'un prédicteur n'entraîne pas une perte de performance supérieure à 10%, on peut l'éliminer.

La perte de la performance (loss) est exprimée par la formule suivante :

$$loss = \frac{(X-x)}{x} \times 100$$
(5)

Avec X qui est l'ensemble de performances des différents ensembles d'attributs et x qui est la valeur la plus basse du critère de performance sélectionnée.

Le RFE utilisé dans cette étude est incorporé dans un ré-échantillonnage de type 10-fold crossvalidation (validation croisée) afin d'obtenir la variation de la précision des performances due à la sélection des prédicteurs. En même temps, une optimisation des hyperparamètres de chaque modèle est effectuée à l'intérieur d'une validation croisée (10 times-10 fold cross validation) afin de trouver les paramètres optimaux permettant d'éviter le surapprentissage du modèle.

#### 5.3.6. Comparaison des modèles statistiques

#### 5.3.6.1. Critères de performance

Pour pouvoir comparer les deux modèles d'apprentissage automatiques utilisés, on a divisé la base de données en deux fractions, une pour l'apprentissage (80%) et l'autre pour le test (20%). Puis, on a utilisé une validation croisée «10 times-10 fold cross validation » accompagnée d'une régularisation des paramètres afin de trouver le modèle optimal qui minimise l'erreur et réduit le surapprentisage. Dans le cas du modèle RF, il se peut que le modèle ne surentraine pas, mais un nombre élevé d'arbres (500 et 1000) peut poser un problème au niveau de la vitesse de l'algorithme. Donc, la validation croisée nous aide à trouver un compromis entre la rapidité computationnelle des arbres de décision et les bonnes performances de prédiction. Ensuite, on compare la performance de ces modèles en calculant ces critères :

• Le RMSE (Root Mean Square Error)

$$RMSE = \sqrt{\sum_{1}^{n} \frac{(y_{i,obs} - y_{i,sim})^2}{n}}$$
(6)

• Le rRMSE (relative Root Mean Square Error)

$$rRMSE = 100 \times \sqrt{\sum_{1}^{n} \frac{(y_{i,obs} - y_{i,sim})^2}{n}}$$
 (7)

• Le Bias

$$Biais = \frac{\sum_{i=1}^{n} (y_{i,obs} - y_{i,sim})}{n}$$
(8)

• Le rBias (relative Bias)

$$rBias = \frac{100}{n} \times \sum_{i=1}^{n} \frac{y_{i,obs} - y_{i,p}}{y_{i,obs}}$$
(9)

• Le coefficient de Nash-Sutcliffe

$$Nash = 1 - \frac{\sum_{i=1}^{n} (y_{i,obs} - y_{i,sim})^2}{\sum_{i=1}^{n} (y_{i,obs} - \bar{y}_{obs})^2}$$
(10)

• Le coefficient de détermination R<sup>2</sup>

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (y_{i,obs} - y_{i,sim})^{2}}{\sum_{i=1}^{n} (y_{i,obs} - \bar{y}_{obs})^{2}}$$
(11)

Avec:

y<sub>i,sim</sub>: représente le vecteur des températures simulées

- y<sub>i,obs</sub>: représente le vecteur des températures observées
- yobs: représente moyenne du vecteur des températures observées
- n : longueur du vecteur des températures observées/simulées

#### 5.3.6.2. Test d'analyse de variance (ANOVA)

La comparaison des critères de performance des modèles testés est complétée en utilisant le test d'ANOVA (ANalysis Of Variance) simple (à un facteur) en sélectionnant l'un de ses critères.

Le test d'ANOVA est utilisé pour comparer les moyennes d'un ou plusieurs groupes. Les hypothèses préalables à ce test sont vérifiées en utilisant les graphiques quantiles-quantiles (QQ-plot) et le test de Shapiro pour l'hypothèse de la normalité des résidus, de même que le test de

Leven pour l'hypothèse de l'homogénéité de la variance. Les hypothèses de base de l'ANOVA sont :

- H<sub>0</sub> : les groupes proviennent de la même population, leurs moyennes ne sont pas significativement différentes.
- H<sub>1</sub> : Il y'a une différence significative entre les moyennes, c'est-à-dire qu'il existe au moins une moyenne différente des autres

Lorsque le test d'ANOVA est significatif, on passe à chercher le ou les groupe(s) différent(s) des autres en les comparant deux à deux. Cette inter-comparaison est exécutée au moyen du test Tukey Honestly Significant Difference (Tukey HSD).

# 5.4. Acquisition des scénarios climatiques et modélisation des températures de l'eau futures

Les modèles climatiques sont des outils numériques complexes qui nous permettent de modéliser l'évolution future plausible du climat en intégrant des équations des lois physiques et chimiques décrivant certains processus météorologiques, océanographiques, hydrologiques, etc. Il existe deux types de modèles climatiques qui diffèrent suivant la résolution de la grille varie de 200 à 300 km. Par contre, le modèle climatique régional couvre une grille de résolution 50 km ou moins, mais doit être piloté à ses frontières par un modèle global <sup>65</sup>.

Les scénarios possibles d'émission, qui font partie des entrées de modèle climatique, représentent l'évolution de la concentration des gaz à effet de serre et des aérosols dans l'atmosphère. Ces scénarios sont catégorisés selon le niveau et la tendance des émissions considérées représentés par les quatre niveaux de l'approche RCP (Representative Concentration Pathways), classés selon leurs forçages radiatifs <sup>65</sup>.

Le tableau ci-dessous représente ces quatre catégories de scénarios d'émission :

Nom	Forçage	Évolution	Équivalent	Réchauffement	Équivalent SRES
	radiatif vers		CO <sub>2</sub> (ppm)	moyen global	approximatif (en termes
	$2100 (W/m^2)$			vers 2100 p/r	du changement de
				1850 (°C)	température globale)
RCP8.5	8.5	Émissions	1370	4.9	A1FI
		fortes et			
		continues			
RCP6.0	6.0	Stabilisation	850	3.0	B2
		sans			
		dépassement			
RCP4.5	4.5	Stabilisation	650	2.4	B1
		sans			
		dépassement			
RCP2.6	2.6	Pic avant	490	1.5	Aucun
		2010 et			
		réduction			

Tableau 5-1: Caractéristiques des scénarios RCP 66

À la sortie de modèles climatiques, on obtient les projections climatiques qui représentent l'évolution future des variables décrivant le climat variant en fonction des scénarios d'émission à l'entrée de modèle.

Dans notre étude, nous avons choisi deux scénarios RCP : un scénario moyen (RCP4.5) et un scénario pessimiste (RCP8.5) appelé également « Business as usual ». Ces scénarios conduisent à un réchauffement des températures moyennes de l'air de 2.5°C à 5°C vers 2100.

Ouranos nous a fourni les projections climatiques de la température de l'air et la vitesse du vent de toutes les stations météorologiques sélectionnées dans le Golfe et l'estuaire du Saint-Laurent. Étant donné que seulement les projections de deux variables d'entrées étaient disponibles, nous avons ré-entrainé les trois modèles de nouveau et utilisé RFE pour le choix du meilleur sous ensemble des prédicteurs. Dans ce cas, seulement deux combinaisons des prédicteurs sont probables soit la température de l'eau comme variable d'entrée soit la température de l'eau et la vitesse du vent. On compare après les trois modèles selon un critère de performance (RMSE) en utilisant le test d'ANOVA. On choisit le modèle le plus performant dans la prédiction de la
température de l'eau pour l'ensemble des stations et on génère les projections de la température de l'eau journalière pour la période 2040-2100.

#### 5.5. Cartographie des zones à risque d'infection

Les métriques thermiques pertinentes à l'évaluation du risque potentiel de prolifération de *V*. *parahaemolyticus* et *V. vulnificus* ont été calculées à partir de ces scénarios. La principale métrique est le nombre de jours au-dessus du seuil de prolifération.

Afin d'évaluer le risque potentiel de prolifération de *Vibrios*, nous avons choisi comme métrique thermique le nombre de jours au-dessus de la température minimale de croissance (15°C). Puis, on calcule la moyenne mensuelle de cet indicateur du risque durant la période d'étude (juin-octobre). Ensuite, on moyenne les valeurs mensuelles obtenues sur une vingtaine d'années durant la période d'étude (2040-2100).

Par la suite, on utilise la méthode d'interpolation basée sur une pondération inverse à la distance, appelée en anglais Inverse Distance Weighting (IDW).

La méthode d'interpolation IDW est une approche déterministe. Elle consiste à attribuer à chaque site de mesure un poids  $(w_i)$  inversement proportionnel à la distance séparant ce site et le site de la valeur à estimer. Cette approche est exprimée par les formules suivantes :

$$\mathbf{z}_{j}^{*} = \frac{\sum_{i=1}^{N} w_{ij} x_{i}}{\sum_{i=1}^{N} w_{ij}}$$
(4)

Avec :

$$w_{ij} = \frac{1}{d_{ij}^p} \tag{5}$$

#### Avec :

 $z^*$ : La variable estimée au niveau de site j non échantillonné

 $x_i$ : La valeur de variable au niveau des sites mesurés

 $w_i$ : Le poids accordé pour chaque site de mesure i

d<sub>ij</sub>: La distance entre le site mesuré i et le site non jaugé j

p: La puissance de la pondération de la distance séparant le site de mesure et le site non échantillonné. La puissance généralement utilisée est 2.

Nous avons choisi cette méthode après avoir comparé l'erreur d'interpolation avec celle de l'interpolation par krigeage. Puisque l'interpolation linéaire est la plus simple et a donné des erreurs similaires au krigeage, nous l'avons sélectionnée. Les cartes d'interpolations pour deux horizons futurs (2040-2060) et (2080-2100) et pour les deux scénarios de changements climatiques (optimiste et pessimiste) ont été produites afin de pouvoir comparer le degré d'expansion possible du risque de prolifération dans la zone d'étude. La cartographie a été réalisée avec le logiciel ARCGIS.

Avant de produire les cartes d'interpolation, on calcule l'erreur relative de l'interpolation. Cette erreur correspond à l'erreur quadratique moyenne (RMSE) de l'interpolation produite pour un mois i divisé par la moyenne de l'indicateur du risque du même mois. Parfois l'erreur relative est assez forte, ce qui peut être dû à la grande variation spatiale de la variable étudiée. Pour les cas trop incertains, l'interpolation est inutile.

Ce projet a porté sur les températures de surface (profondeur moyenne de 1.5m). À cette profondeur, on peut trouver des mollusques sauvages (e.g. près des îles), de même que des cultures sur les estrans et en eau peu profonde. Cependant, pour d'autres cultures en eau qui se font parfois en eau plus profonde (comme les huîtres), la cartographie des zones du risque peut être biaisée (surestimation) vu qu'il peut y avoir une stratification thermique importante dans

32

certains sites et que la modélisation de la température de surface est moins représentative des conditions en profondeur.

### 6.0. Résultats

#### 6.1. Sélection des prédicteurs de la variation de la température de l'eau

Les résultats de l'analyse de la corrélation présentés dans le tableau A-2 de l'annexe A, indiquent que la température de la surface d'eau (SSTP) est significativement plus corrélée avec les MM-3jours la température de l'air (Tair), MM-30 jours de la vitesse du vent (WSPD), MM-30 jours du marnage maximal journalier (Tidal\_range), MM-60 jours de NAO et MM-120 jours du débit fluvial du Saint-Laurent (Flow).

SSTP est fortement corrélée avec Tair qui explique plus de 60% de sa variation pour la majorité des stations. La vitesse du vent est modérément corrélée avec Tair pour 90% des stations, elle explique entre 10 et 50% de la variation de SSTP.

La dépendance entre SSTP et le débit fluvial est faible. Le débit fluvial influence surtout la SSTP au niveau de l'estuaire maritime en expliquant 10% de la variation. Cette dépendance diminue entre 5 et 7 % en s'éloignant vers l'embouchure de l'estuaire et les stations côtières proches.

Concernant les indices climatiques, on a trouvé que seulement l'indice NAO est significativement corrélé avec SSTP. Son effet majeur est exercé au niveau des stations de la Côte Nord en expliquant entre 6 et 10% de la variation de SSTP.

Suite à la sélection des meilleures moyennes mobiles (filtres passe-bas) pour chaque prédicteur, une corrélation des paires entre la variable d'intérêt et les variables explicatives est appliquée.

33

La corrélation deux à deux, appelée également la corrélation des paires, permet de présenter à la fois les corrélations entre la variable d'intérêt et les prédicteurs, ainsi que les inter-corrélations qui peuvent exister entre les prédicteurs eux-mêmes.

La figure 6-1 ci-dessous présente la corrélation des paires au niveau de la station Grande Rivière.



Figure 6-1: Matrice de corrélations au niveau de la station Grande -Rivière

Chaque niveau de signification de la valeur de p (0.001, 0.01, 0.05, 0.1, 1) est associé aux symboles suivant ("\*\*\*", "\*\*", "\*", ".", "")

#### Avec :

SSTP : La température moyenne journalière de la surface d'eau Tair : la MM-3jours de la température moyenne journalière de l'air WSPD : La MM-30 jours de la vitesse moyenne journalière du vent Tidal\_range : La MM-30 jours du marnage maximal journalier NAO : La MM-60 jours de l'indice North Atlantic Oscilation (NAO) Flow : La MM-120 jours du débit fluvial du Saint-Laurent La figure 6-1 indique que la température de l'eau (SSTP) est fortement corrélée avec Tair et WSPD, alors qu'elle est faiblement associée au reste des prédicteurs. En ce qui concerne la corrélation entre les prédicteurs, WSPD est fortement corrélée à Tair, tandis que Tidal\_range est modérément corrélé au Flow.

Lorsque les corrélations entre les prédicteurs sont relativement fortes, le modèle de régression linéaire devient instable. On parle d'une multicolinéarité. L'occurrence de la multicolinéarité présente un problème pour les modèles de régression linéaire, car elle augmente la variance des coefficients à estimer et les rend instables, ce qui affaiblit la puissance du modèle dans l'identification des coefficients significatifs. Dans le but de tester l'occurrence de la multicolinéarité, on a calculé le facteur d'inflation de la variance. À la station Grande Rivière ainsi que pour le reste des stations étudiées, les valeurs de VIF pour tous les prédicteurs utilisés sont inférieures aux valeurs critiques signalées dans la littérature.

#### 6.2. Performance des modèles testés

La sélection du meilleur sous ensemble des prédicteurs les plus pertinents, assurant à la fois de simplifier le modèle et d'améliorer sa performance de prédiction, est effectuée par la méthode de RFE incorporée avec une validation croisée (10-fold). Pour chaque modèle, la sélection d'attributs pertinents est exécutée pendant la phase d'entrainement sur 80% de données. Les 20 % de données restantes (les données de test) sont inconnues pour le modèle. Ces données du test sont utilisées pour déterminer la performance de la prédiction de chaque modèle en comparant plusieurs critères tels que l'erreur quadratique moyenne (RMSE), le coefficient de détermination (R<sup>2</sup>), l'erreur relative quadratique moyenne (rRMSE), le critère de Nash-Sutcliff (Nash) et le biais moyen relatif (rBiais), comme indiqué dans le tableau 6-1 suivant.

Modèles	Attributs sélectionnés	RMSE (°C)	$\mathbb{R}^2$	rRMSE (%)	Nash	rBias (%)
MLR	Tair, Tidal range	1.60	0.74	16,68	0,74	-1,469
ANN	Tair, WSPD, Tidal range, Flow, NAO	1,46	0.79	14.67	0.78	-0.987
RF	Tair, WSPD, Tidal range, Flow	1.17	0.86	11.80	0.86	-0.988

Tableau 6-1: Résultats des critères de performance des modèles sur les données du test

On remarque que les modèles d'apprentissage automatique (ANN et RF) performent mieux que le modèle linéaire (MLR). La performance de l'ANN a été légèrement meilleure que celle de MLR. En comparaison avec le modèle linéaire, l'ANN présente les valeurs les plus faibles des critères RMSE, rRMSE, rBias et les valeurs de Nash et R<sup>2</sup> les plus élevées.

Par contre, le RF est plus performant que les modèles ANN et MLR grâce aux valeurs les plus faibles des critères RMSE, rRMSE et les valeurs de Nash et R<sup>2</sup> les plus élevées.

L'ajout de prédicteurs dans le modèle linéaire n'améliore pas significativement la performance étant donné les relations non linéaires qui peuvent exister entre les prédicteurs et la variable cible. En revanche, les modèles d'apprentissage automatique ont la capacité de capturer et de modéliser des relations non linéaires complexes.

La figure 6-2 ci-dessous montre que le modèle RF possède un meilleur ajustement et une meilleure précision de prédiction que l'ANN. Cette approche offre des performances plutôt robustes, grâce à la conception de son algorithme qui permet un compromis faible variance – faible biais.



Figure 6-2 : Comparaison de performance des modèles des Réseaux des Neurones Artificiels (ANN) et des Forêts Aléatoires (RF) au niveau de la station Grande Rivière

Comme démontré dans la figure 6-2 ci-dessus, le modèle RF était le meilleur modèle de prédiction de la température de l'eau à la station de Grande Rivière. Cependant, ce résultat peut être spécifique à cette station. Par conséquent, la même procédure est répétée pour le reste des stations. Le tableau A-3 de l'annexe A présente les résultats des performances des modèles testés pour les stations d'étude. Puis, à travers le test d'ANOVA, les modèles testés sont comparés en vérifiant que si les moyennes des valeurs de leurs RMSE sont significativement différentes ou pas.

Le test d'ANOVA (degré de liberté = 2, statistique-F=5.806) indique une différence significative entre les modèles au seuil de signification de 5% (valeur de p= 0,006). Ensuite, un test de Tukey Honestly Significant Difference (Tukey HSD) est exécuté entre les différents modèles testés pour l'ensemble des stations. Le test montre qu'il n'y a pas de différence significative entre ANN et MLR (valeur de p = 0,89). En revanche, il existe une différence

significative entre le modèle RF et les deux modèles MLR (valeur de p = 0,008) et ANN (valeur de p = 0,026) au seuil de signification de 5%. Le meilleur sous-ensemble d'entités pour le modèle RF diffère d'une station à l'autre mais il inclut pratiquement tous les prédicteurs testés. Concernant la performance de RF, la valeur moyenne de RMSE calculée pour l'ensemble de stations est 1,3 ° C.

#### 6.3. Modélisation des futures températures de l'eau

Vu la disponibilité des projections de deux prédicteurs seulement (température de l'air et la vitesse du vent), les modèles testés ont été reconstruits avec ces deux seuls intrants. Puis, nous avons utilisé RFE pour la sélection du meilleur sous ensemble d'entrées. Nous avons comparé par la suite les modèles de nouveau. Les résultats de test ANOVA (degré de liberté = 2, statistique-F=0.881) indique qu'il n'existe pas une différence significative entre les moyennes des RMSE des trois modèles pour l'ensemble des stations (valeur de p=0.881). Le Tableau A-4 de l'annexe A collige les résultats des critères des performances des trois modèles pour l'ensemble des stations avec l'ensemble des prédicteurs les plus pertinents. On remarque que le meilleur sous ensemble d'attributs pour le modèle RF comprend la température de l'air et la vitesse du vent alors que pour les modèles MLR et ANN, il comprend seulement la température de l'air comme variable d'entrée. Le modèle le plus parcimonieux est recherché, d'autant plus qu'il n'existe pas une différence significative entre les RMSE moyennes des trois modèles. On a choisi l'ANN pour la suite des travaux. Des scénarios futures de températures de l'eau ont été produits à l'aide du modèle ANN pour toutes les stations incluant les stations de Merinov en entrant seulement les projections de la Tair pour l'échéance (2040-2100) comme variable d'entrée. Puis, nous avons utilisé le test de Mann-Kendall Modifié, qui tient en compte des autocorrélations dans la série temporelle, pour tester la tendance des futures températures de

l'eau. Le tableau A-5 de l'annexe A collige les résultats de ce test pour toutes les stations. Nous avons calculé la pente de la tendance des futures températures de l'eau par décennie et vers 2100. En comparant les résultats du test de la tendance pour toutes les stations, on remarque que la pente décennale des futures températures de l'eau varie entre 0.2°C et 0.7°C et que les futures températures de l'eau vont augmenter entre 1.25°C et 4.27°C vers 2100.

En moyennant les pentes des tendances pour toutes les stations, on trouve que la température de l'eau va augmenter de 0.4 °C par décennie, pour un total de 2.5°C durant la période (2040-2100).

#### 6.4. Cartographie des futures zones à risque d'infections

Les cartes produites à partir de l'interpolation du nombre de jours dépassant 15 °C pour les mois d'août, juillet et septembre ont été produites car l'erreur relative n'a pas dépassé 50 %. Par contre, pour les mois de juin et d'octobre, seul un tableau récapitulatif par station est présenté puisque l'erreur relative calculée est très élevée (>50%).

Les figures 6-3 à 6-5 ci-dessous présentent les cartes d'interpolation du nombre de jours dépassant le seuil de 15°C au mois d'août. On a essayé de comparer d'une part l'expansion spatiale de l'indicateur de risque pour le même scénario (RPC 4.5 ou 8.5), mais entre deux périodes différentes (2040-2060; 2080-2010; Figures 6-3 et 6-4), ainsi que pour la même période (2080-2010), mais pour deux scénarios différents (RPC 4.5; 8.5; Figure 6-5).



Figure 6-3: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois d'Août



Figure 6-4: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois d'Août



Figure 6-5: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois d'Août

Le tableau 6-2 ci-dessous présente le nombre de jours au-dessus du seuil 15°C pour le scénario pessimiste des mois de juin et d'octobre pour la période (2080-2100). On a choisi de présenter les résultats de cette période, car les valeurs du nombre de jours dépassant 15°C sont plus élevées que pour la période précédente.

Localisation	stations	Type de Station	Juin	Octobre
	Montlouis		11	0
Estuaire	Gyre Anticosti	Bouée	15	0
	Courant de Gaspé		6	0
	Sept-Îles		13	0
Côte Nord	Natashquan		15	0
	Baie des Sept Iles		24	0
Le Daie des Chaleurs	Baie de Cascapedia		28	2
La Bale des Chaleurs	Baie de Tracadigache		26	1
	Grande Rivière		14	0
Gaspésie	Havre de Gaspé	Thermographes	25	2
	Haldimand		9	0
Ile-Prince Édouard	Borden		30	19
Iles de la Madeleine	Baie de Plaisance		22	6
	Bassin Havre Aubert		30	20
	Lagune du Havre aux			
	Maisons		29	16
	Lagune de la Grande			
	Entrée		28	14
	Île-Shag		20	4

Tableau 6-2 : Le nombre de jours au-dessus du seuil 15°C pour le scénario RCP8.5 des mois de juin et octobre durant la période (2080-2100)

Le nombre de jours au-dessus de 20°C a aussi été calculé pour localiser les zones à risque d'infection très élevée. L'erreur relative à l'interpolation de la variable du nombre de jours audessus du 20°C était très élevée. Ceci peut être expliqué par la haute variation de la température de l'eau entre le nord (plusieurs stations ont une valeur nulle pour cet indice) et le sud du Golfe du Saint Laurent. Le tableau 6-3 ci-dessous présente le nombre de jours dépassant le seuil 20°C pour le scénario 8.5 durant la période (2080-2100).

Localisation	stations	Juin	Juillet	Août	Septembre
Côte Nord	Natashquan	0	14	17	0
Ile-de la	Bassin Havre Aubert	8	31	31	22
	Lagune du Havre aux				
Widdeleffie	Maisons	5	30	31	16
	Lagune de la Grande Entrée	5	31	31	16
Gaspésie	Havre de Gaspé	1	25	30	4
Ile Prince- Édouard	Borden	0	15	16	0

Tableau 6-3 : Le nombre de jours au-dessus du seuil 20°C pour le scénario RCP8.5 durant la période (2080-2100

Le tableau 6-3 présente les stations avec des zones à risque d'infection très élevée, selon les scénarios étudiés. Pour le reste des stations, le nombre de jours dépassant 20°C est nul et n'ont donc pas été incluses au tableau.

### 7.0. Discussion

Durant cette étude, on a démontré que des variables explicatives autres que la température de l'air expliquent un pourcentage important de la variation de la température de l'eau dans la zone d'étude. Ces prédicteurs incluent une autre variable météorologique (vitesse du vent), une variable océanographique (marnage), hydrologique (débit) et une variable liée aux téléconnexions (indice climatique).

La plupart des travaux, qui portent sur l'indice climatique NAO, montrent l'occurrence de son effet sur la variation de la pression au niveau de la mer durant l'hiver<sup>67</sup>. Très peu études ont montré

que l'effet d'indice climatique peut se produire durant l'été<sup>68,69</sup>. Selon nos résultats, il semble pouvoir influencer la variation de la température de l'eau au niveau de la côte Nord.

Les modèles d'apprentissage automatique performent mieux que le modèle linéaire dans la modélisation de la température de l'eau dans l'estuaire et le Golfe du Saint-Laurent. L'ajout d'autres variables explicatives a amélioré la précision de la prédiction de la température de l'eau pour les modèles d'apprentissage automatique. Le modèle RF est le meilleur des modèles testés pour la prédiction de la température de l'eau. Cependant, il exige un nombre plus important de variables explicatives. Dans un contexte où les prédicteurs les plus facilement disponibles dans les sorties de modèle incluent la température de l'air et la vitesse du vent, le modèle ANN était le meilleur choix étant donné sa relative simplicité et le fait qu'il requiert seulement la température de l'air comme variable d'entrée pour la plupart des stations.

Durant la période (2040-2060) dans un cadre de scénario optimiste (Figure 6-3), on remarque que les eaux des bancs coquillers aux Iles-de-la-Madeleine, à l'Ile-Prince-Édouard, et pour une partie de la Côte Nord (Natashquan) ainsi qu'en Gaspésie et la Baie des Chaleurs seront à risque élevé d'infection, car le nombre de jours au-dessus du seuil de 15 °C excède 25 jours. Le nombre de jours dépassant le seuil, calculé aux bancs coquillers de la Côte Nord (Baie de Sept-Îles), varie quant à lui entre 20 et 25 jours. Pour la même période, mais dans un cadre de scénario pessimiste (Figure 6-4), la plupart des bancs coquillers seront à risque élevé d'infection par *vibrios* avec un nombre de jours au-dessus du seuil dépassant 25 occupant environ 72% de la surface totale de la zone d'étude.

En comparant les cartes d'interpolations du mois d'août (Figure 6-5) durant la période (2080-2100) pour les deux scénarios (optimiste et pessimiste), on remarque que le nombre de jours audessus du seuil (15°C), dépassant 25 jours, occupe entre 71.8% (scénario4.5) et 98.2% (scénario8.5) de la surface totale de l'estuaire et Golfe du Saint Laurent. Donc, vers 2100, la plupart des stations localisées dans les bancs coquillers, où se localisent des zones de récolte des mollusques bivalves, seront probablement sous un risque élevé d'infection par *vibrios* pour n'importe quel des deux scénarios climatiques envisagés.

L'expansion d'indicateur du risque au niveau de la zone d'étude durant le mois du juillet (Figures B-1, B-2 et B-3 de l'annexe B) est semblable au mois d'août. Les bancs coquillers seraient exposés à un risque similaire au mois d'août.

Durant le mois du septembre (Figures B-4, B-5 et B-6 de l'annexe B), le nombre de jours audessus du seuil, au niveau des bancs coquillers des îles de la Madeleine et Ile-Prince-Édouard excède 25 jours pour les deux scénarios climatiques envisagés, et ce, pour les deux horizons. Donc, elles seront sous un risque élevé d'infection par *vibrios*. Durant la période (2080-2100) dans un cadre de scénario pessimiste (Figure B-6), on remarque qu'en plus des îles de la Madeleine et l'Ile-Prince-Édouard, les bancs coquillers au niveau de la Gaspésie et la Baie des Chaleurs seront à risque élevé d'infection, car le nombre de jours au-dessus du seuil de 15 °C excède 20 jours.

Les résultats d'indicateur du risque pour les mois du juin et octobre (Tableau 6-2) montrent que le risque d'infection va se prolonger jusqu'au mois d'octobre sur les côtes de l'IPE et des Iles-dela-Madeleine. Donc, l'expansion du risque d'infection se prolonge spatialement et temporellement. Durant le mois du juin, tous les bancs coquillers du Golfe du Saint Laurent seront à risque élevé d'infection.

Les résultats du Tableau 6-3 montrent que durant la période (2080-2100) et dans un cadre de scénario pessimiste, un seuil de 20°C dans les zones de culture des Iles de la Madeleine serait dépassé durant environ 31jours. En septembre, le nombre de jours moyen dépassant le seuil 20°C

est d'environ 18 jours. Les bancs coquillers de la Gaspésie seront sous risque élevé durant le mois du juillet et d'août avec un nombre de jours au-dessus du seuil de 20°C qui dépasse 25 jours.

Les bancs coquillers de l'IPE et de côte Nord (station Natashquan) seront sous risque moins élevé que ceux des Iles de la Madeleine et la Gaspésie durant le mois du juillet et d'août avec un nombre de jours au-dessus du seuil de 20°C d'environ 16 jours. Durant le mois du juin, l'indicateur du risque ne dépasse pas 8 jours pour toutes les stations.

### **8.0.** Conclusions et recommandations

Les variables explicatives, autres que la température de l'eau, contribuent à expliquer la variation de la température de l'eau. Ces prédicteurs varient selon les stations et les modèles mais peuvent inclure des prédicteurs météorologique (vitesse du vent), océanographique (marnage), hydrologique (débit) et un indice de téléconnexion (Indice NAO). Les modèles d'apprentissage (Réseaux des Neurones Artificiels et Forêts Aléatoires) automatique permettent d'expliquer les relations non linéaires complexes qui peuvent exister entre la température de l'eau et ces variables. On a montré que ces modèles performent mieux que le modèle linéaire simple dans la prédiction de la température de l'eau. Le modèle des Forêts Aléatoires était le modèle le plus performant, mais on ne l'a pas utilisé dans la prédiction des futures températures de l'eau parce qu'il requiert plusieurs prédicteurs plus difficilement disponibles. Suite à la comparaison des modèles ANN et RF, en utilisant seulement comme variables d'entrées la température de l'air et la vitesse du vent et en appliquant la méthode RFE pour la sélection des meilleurs attributs, on a trouvé que le modèle ANN, en entrant seulement la température de l'air, est le modèle le plus simple et performant.

En utilisant les projections futures de la température de l'air en intrant au modèle sélectionné, on a pu produire les futures températures de l'eau pour l'ensemble des stations, puis calculer le nombre de jours au-dessus de la température minimale de la prolifération des *vibrios* pathogènes (15°C), comme indicateur du risque potentiel d'infection par *vibrios*.

En produisant les cartes du futur risque potentiel par *vibrios* suite à une interpolation spatiale de l'indicateur du risque, on a pu conclure que le risque d'infection par *vibrios* pourrait être élevé au niveau du Golfe du Saint-Laurent, notamment aux Iles-de-la-Madeleine, de la Gaspésie, de la Baie de la Chaleur ,de l'IPE et d'une partie de la Côte Nord selon les deux scénarios climatiques utilisés.

Le risque d'infection pourrait prendre de l'expansion en allant vers 2100 pour les deux scénarios. En effet, vers 2100 la couverture spatiale du nombre de jours au-dessus du seuil (15°C), dépassant 25 jours, passe de 72% (le scénario optimiste) à 98% (le scénario pessimiste) de la surface totale de l'estuaire et Golfe du Saint Laurent. En plus, cette expansion se prolonge temporellement en couvrant même le mois d'octobre pour certaines stations.

La température n'est pas la seule variable qui influence la prolifération des *vibrios*. La salinité est une autre variable importante. Une phase ultérieure de ce projet devrait prendre en compte cette variable. Pour ce faire, un modèle déterministe de circulation des masses d'eau du Golfe et de l'estuaire du St-Laurent devrait être mis à profit.

De plus, le présent projet a porté sur les températures de surface (profondeur moyenne de 1.5m). Cependant, il peut y avoir une importante stratification thermique dans certains sites côtiers. Cette possible stratification devrait aussi être prise en compte dans une éventuelle seconde phase du projet.

## 9.0. Références

- 1 Vezzulli, L., Colwell, R. R. & Pruzzo, C. Ocean warming and spread of pathogenic vibrios in the aquatic environment. *Microbial ecology* **65**, 817-825, doi:10.1007/s00248-012-0163-2 (2013).
- 2 Vezzulli, L. *et al.* Climate influence on Vibrio and associated human diseases during the past half-century in the coastal North Atlantic. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E5062-5071, doi:10.1073/pnas.1609157113 (2016).
- 3 McLaughlin. climate anomalies and the increasing risk of Vibrio. (2005).
- 4 Martinez-Urtaza, J., Bowers, J. C., Trinanes, J. & DePaola, A. Climate anomalies and the increasing risk of Vibrio parahaemolyticus and Vibrio vulnificus illnesses. *Food Research International* **43**, 1780-1790, doi:10.1016/j.foodres.2010.04.001 (2010).
- 5 Baker-Austin, C. *et al.* Emerging Vibrio risk at high latitudes in response to ocean warming. *Nature Climate Change* **3**, 73-77, doi:10.1038/nclimate1628 (2012).
- 6 Deeb, R., Tufford, D., Scott, G. I., Moore, J. G. & Dow, K. Impact of Climate Change on Vibrio vulnificus Abundance and Exposure Risk. *Estuaries and Coasts* **41**, 2289-2303, doi:10.1007/s12237-018-0424-5 (2018).
- 7 Dechet, A. M., Yu, P. A., Koram, N. & Painter, J. Nonfoodborne Vibrio infections: an important cause of morbidity and mortality in the United States, 1997-2006. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **46**, 970-976, doi:10.1086/529148 (2008).
- 8 Oliver, J. D. Wound infections caused by Vibrio vulnificus and other marine bacteria. *Epidemiology and Infection* **133**, 383-391, doi:10.1017/s0950268805003894 (2005).
- 9 Feldhusen, F. The role of seafood in bacterialfoodborne diseases. *Microbes and infection* **2**, 1651-1660 (2000).
- 10 Oliver, J. D. Vibrio vulnificus: Death on the Half Shell. A Personal Journey with the Pathogen and its Ecology. *Microbial ecology* **65**, 793-799, doi:10.1007/s00248-012-0140-9 (2013).
- 11 Banerjee, S. K., Rutley, R. & Bussey, J. Diversity and dynamics of the Canadian coastal Vibrio community: an emerging trend detected in the temperate regions. *Journal of bacteriology* **200**, e00787-00717 (2018).
- 12 Wormser, G. P. & Colebunders, R. L. Control of Communicable Diseases Manual, 19th Edition Edited by David L. Heymann Washington, DC: American Public Health Association, 2008. 746 pp. \$45.00 (hardcover). *Clinical Infectious Diseases* **49**, 1292-1293, doi:10.1086/605668 (2009).
- 13 Froelich, B. A. & Noble, R. T. Vibrio bacteria in raw oysters: managing risks to human health. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **371**, doi:10.1098/rstb.2015.0209 (2016).
- 14 Baker-Austin, C., Trinanes, J., Gonzalez-Escalona, N. & Martinez-Urtaza, J. Non-Cholera Vibrios: The Microbial Barometer of Climate Change. *Trends in microbiology* **25**, 76-84, doi:10.1016/j.tim.2016.09.008 (2017).
- 15 Gilliss, D. *et al.* Incidence and Trends of Infection with Pathogens Transmitted Commonly Through Food — Foodborne Diseases Active Surveillance Network, 10

U.S. Sites, 1996–2012. *MMWR. Morbidity and Mortality Weekly Report* **62**, 283-287 (2013).

- 16 Burge, C. A. *et al.* Climate change influences on marine infectious diseases: implications for management and society. *Annual review of marine science* **6**, 249-277, doi:10.1146/annurev-marine-010213-135029 (2014).
- 17 Baker-Austin, C., Stockley, L., Rangdale, R. & Martinez-Urtaza, J. Environmental occurrence and clinical impact of Vibrio vulnificus and Vibrio parahaemolyticus: a European perspective. *Environmental microbiology reports* **2**, 7-18, doi:10.1111/j.1758-2229.2009.00096.x (2010).
- Kaspar, C. W. & Tamplin, M. L. Effects of temperature and salinity on the survival of Vibrio vulnificus in seawater and shellfish. *Applied and environmental microbiology* 59, 2425-2429 (1993).
- 19 Motes, M. L. *et al.* Influence of Water Temperature and Salinity on Vibrio vulnificus in Northern Gulf and Atlantic Coast Oysters (Crassostrea virginica). *Applied and environmental microbiology* **64**, 1459-1465 (1998).
- 20 Heng, S.-P. *et al.* Vibrio vulnificus: An Environmental and Clinical Burden. *Frontiers in microbiology* **8**, 997-997, doi:10.3389/fmicb.2017.00997 (2017).
- 21 Miles, D. W., Ross, T., Olley, J. & McMeekin, T. A. Development and evaluation of a predictive model for the effect of temperature and water activity on the growth rate of Vibrio parahaemolyticus. *International Journal of Food Microbiology* **38**, 133-142, doi:<u>https://doi.org/10.1016/S0168-1605(97)00100-1</u> (1997).
- 22 Liu, B., Liu, H., Pan, Y., Xie, J. & Zhao, Y. Comparison of the Effects of Environmental Parameters on the Growth Variability of Vibrio parahaemolyticus Coupled with Strain Sources and Genotypes Analyses. *Frontiers in microbiology* **7**, 994, doi:10.3389/fmicb.2016.00994 (2016).
- 23 WHO, F. Risk assessment of Vibrio vulnificus in raw oysters: interpretative summary and technical report, microbiological risk assessment series 8: Geneva: World Health Organization. *Rome: Food And Agriculture Organization of the United Nations* (2005).
- 24 Nishibuchi, M. & DePaola, A. Vibrio species. *Foodborne pathogens: Microbiology and molecular biology*, 251-227 (2005).
- 25 McLaughlin, J. B. *et al.* Outbreak of Vibrio parahaemolyticus gastroenteritis associated with Alaskan oysters. *New England Journal of Medicine* **353**, 1463-1470 (2005).
- Kelly, M. T. Effect of temperature and salinity on Vibrio (Beneckea) vulnificus occurrence in a Gulf Coast environment. *Applied and environmental microbiology* 44, 820-824 (1982).
- 27 Leonard, B. Fish and Fishery Products: Hazards and Controls Guidance (4th Ed.). (DIANE Publishing Company, 2011).
- 28 Motes, M. L. & DePaola, A. Offshore suspension relaying to reduce levels of Vibrio vulnificus in oysters (Crassostrea virginica). *Applied and environmental microbiology* **62**, 3875-3877 (1996).
- 29 IPCC. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (Cambridge University Press, 2013).
- 30 Baker-Austin, C. *et al.* Heat Wave–Associated Vibriosis, Sweden and Finland, 2014. *Emerging Infectious Diseases* **22**, 1216-1220, doi:10.3201/eid2207.151996 (2016).

- 31 Cazorla, C., Guigon, A., Noel, M., Quilici, M.-L. & Lacassin, F. Fatal Vibrio vulnificus infection associated with eating raw oysters, New Caledonia. *Emerging infectious diseases* **17**, 136 (2011).
- 32 Harvell, C. D. *et al.* Climate Warming and Disease Risks for Terrestrial and Marine Biota. *Science* **296**, 2158-2162, doi:10.1126/science.1063699 (2002).
- 33 Martinez-Urtaza, J. *et al.* Emergence of Asiatic Vibrio Diseases in South America in Phase With El Niño. *Epidemiology* **19**, 829-837, doi:10.1097/EDE.0b013e3181883d43 (2008).
- 34 Lau, N.-C., Leetmaa, A. & Nath, M. J. Interactions between the Responses of North American Climate to El Niño–La Niña and to the Secular Warming Trend in the Indian–Western Pacific Oceans. *Journal of Climate* **21**, 476-494, doi:10.1175/2007jcli1899.1 (2008).
- 35 Higgins, R. W., Leetmaa, A. & Kousky, V. E. Relationships between Climate Variability and Winter Temperature Extremes in the United States. *Journal of Climate* **15**, 1555-1572, doi:10.1175/1520-0442(2002)015<1555:RBCVAW>2.0.C0;2 (2002).
- 36 Bell, G. D. *et al.* Climate Assessment for 1998. *Bulletin of the American Meteorological Society* **80**, 1040-1040, doi:10.1175/1520-0477(1999)080<1040:CAF>2.0.CO;2 (1999).
- 37 El-Sabh, M. I. & Murty, T. S. in *Oceanography of a Large-Scale Estuarine System* 10-50 (Springer, 1990).
- 38 Galbraith, P. S. *et al. Physical oceanographic conditions in the Gulf of St. Lawrence in* 2016. (Canadian Science Advisory Secretariat, 2017).
- 39 Bonsal, B. & Shabbar, A. *Large-scale climate oscillations influencing Canada, 1900-2008.* (Canadian Councils of Resource Ministers, 2011).
- 40 Shabbar, A. The impact of El Niño-Southern Oscillation on the Canadian climate. *Advances in Geosciences* **6**, 149-153 (2006).
- 41 Thiombiano, A. N., El Adlouni, S., St-Hilaire, A., Ouarda, T. B. M. J. & El-Jabi, N. Nonstationary frequency analysis of extreme daily precipitation amounts in Southeastern Canada using a peaks-over-threshold approach. *Theoretical and Applied Climatology* **129**, 413-426, doi:10.1007/s00704-016-1789-7 (2017).
- 42 C.S.A.S. *Oceanographic Conditions in the Atlantic Zone in 2016*. (Fisheries and Oceans Canada, Centre for Science Advice, 2017).
- 43 Benyahya, L., Caissie, D., St-Hilaire, A., Ouarda, T. B. M. J. & Bobée, B. A Review of Statistical Water Temperature Models. *Canadian Water Resources Journal* **32**, 179-192, doi:10.4296/cwrj3203179 (2007).
- 44 Hopfield, J. J. Artificial neural networks. *IEEE Circuits and Devices Magazine* **4**, 3-10 (1988).
- 45 Lippmann, R. P., Gold, B. & Malpass, M. L. A comparison of Hamming and Hopfield neural nets for pattern classification. (MASSACHUSETTS INST OF TECH LEXINGTON LINCOLN LAB, 1987).
- 46 Carpenter, G. A. & Grossberg, S. The ART of adaptive pattern recognition by a selforganizing neural network. *Computer* **21**, 77-88 (1988).
- 47 Rosenblatt, F. Principles of neurodynamics. perceptrons and the theory of brain mechanisms. (CORNELL AERONAUTICAL LAB INC BUFFALO NY, 1961).
- 48 Rumelhart, D. E., Hinton, G. E. & Williams, R. J. Learning representations by backpropagating errors. *Nature* **323**, 533 (1986).

- 49 Bélanger, M., El-Jabi, N., Caissie, D., Ashkar, F. & Ribi, J. M. Estimation de la température de l'eau de rivière en utilisant les réseaux de neurones et la régression linéaire multiple. *Revue des sciences de l'eau* **18**, doi:10.7202/705565ar (2005).
- 50 Chenard, J.-F. & Caissie, D. Stream temperature modelling using artificial neural networks: application on Catamaran Brook, New Brunswick, Canada. *Hydrological Processes* **22**, 3361-3372, doi:10.1002/hyp.6928 (2008).
- 51 Jeong, D. I., Daigle, A. & St-Hilaire, A. Development of a Stochastic Water Temperature Model and Projection of Future Water Temperature and Extreme Events in the Ouelle River Basin in QuÉbec, Canada. *River Research and Applications* **29**, 805-821, doi:10.1002/rra.2574 (2013).
- 52 Witten, I. H., Frank, E., Hall, M. A. & Pal, C. J. *Data Mining: Practical machine learning tools and techniques.* (Morgan Kaufmann, 2016).
- 53 Geman, S., Bienenstock, E. & Doursat, R. Neural networks and the bias/variance dilemma. *Neural computation* **4**, 1-58 (1992).
- 54 Krogh, A. & Hertz, J. A. in *Advances in neural information processing systems.* 950-957.
- 55 Breiman, L. Random forests. *Machine learning* **45**, 5-32 (2001).
- 56 Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. Classification and regression trees. (1984).
- 57 Quach, A. "Extensions and Improvements to Random Forests for Classification". *All Graduate Theses and Dissertations.* 6755 (2017).
- 58 Breiman, L. Bagging predictors. *Machine learning* **24**, 123-140 (1996).
- 59 Ho, T. K. in *Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on.* 545-549 (IEEE).
- 60 Pang, B., Yue, J., Zhao, G. & Xu, Z. Statistical Downscaling of Temperature with the Random Forest Model. *Advances in Meteorology* **2017**, 1-11, doi:10.1155/2017/7265178 (2017).
- 61 Zuur, A. F., Ieno, E. N. & Elphick, C. S. A protocol for data exploration to avoid common statistical problems. *Methods in ecology and evolution* **1**, 3-14 (2010).
- 62 Dormann, C. F. *et al.* Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27-46 (2013).
- 63 Belsley, D. A. Conditioning Diagnostics Collinearity and Weak Data in Regression. *JOURNAL-OPERATIONAL RESEARCH SOCIETY* **44**, 88-88 (1993).
- 64 Hair, J. F., Anderson, R. E., Tatham, R. L. & William, C. (Upper Saddle River, NJ: Prentice Hall, 1998).
- 65 CHARRON, I. Guide sur les scénarios climatiques : Utilisation de l'information climatique pour guider la recherche et la prise de décision en matière d'adaptation.
   94 p (Edition 2016.Ouranos, 2016).
- 66 Rogelj, J., Meinshausen, M. & Knutti, R. Global warming under old and new scenarios using IPCC climate sensitivity range estimates. *Nature climate change* **2**, 248 (2012).
- 67 Hurrell, J. W. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science* **269**, 676-679, doi:10.1126/science.269.5224.676 (1995).
- 68 Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of lowfrequency atmospheric circulation patterns. *Monthly weather review* **115**, 1083-1126 (1987).

69 Portis, D. H., Walsh, J. E., El Hamly, M. & Lamb, P. J. Seasonality of the North Atlantic oscillation. *Journal of Climate* **14**, 2069-2078 (2001).

## Annexe A

Sources	Type de station	Stations	
		Sept-Îles	
		Rivière aux Tonnerre	
		Havre St Pierre	
		Natashquan	
	Thermographes côtiers	Romaine	
	Thermographes couers	Tabatière	
l'Institut de Maurice- Lamontagne et Pêches et Océans Canada.		Blanc Sablon	
		Grande-Rivière	
		Iles Shag	
		Borden	
		Rimouski	
		Courant Gaspé	
	Bouées	Gyre Anticosti	
		Montlouis	
		Shediac Valley	
		Baie des Sept Iles	
		Baie de Cascapedia	
		Baie de Trascapedia	
	Thermographes côtiers aux	Baie de plaisance	
Merinov	bancs coquillers	Lagune du Havre aux	
	balles coquillers	Maisons	
		Bassin Havre Aubert	
		Lagune de la Grande	
		Entrée	
		Baie des Belles Amours	

Tableau A-1: Liste des stations dans le Golfe et estuaire du Saint Laurent

Stations	Tair		Vent		Mar	nage	Ru	noff	NAO		
Stations	R	$\mathbb{R}^2$	R	$\mathbb{R}^2$	R	$\mathbb{R}^2$	R	$\mathbb{R}^2$	R	$\mathbb{R}^2$	
Rimouski	0,892	80%	-0,328	11%	0,179	3%	0,318	10%	-0,305	9%	
Courant de Gaspé	0,913	83%	-0,712	51%	0,284	8%	0,23	5%	-0,318	10%	
Gyre Anticosti	0,571	33%	-0,425	18%	0,238	6%	0,267	7%	-0,285	8%	
Sept Iles	0,836	70%	-0,13	2%	0,312	10%	0,234	5%	-0,298	9%	
Rivière aux Tonnerre	0,765	59%	-0,283	8%	0,19	4%	0,272	7%	-0,315	10%	
Havre St Pierre	0,589	35%	-0,423	18%	Non_Sig	Non_Sig	0,096	1%	-0,255	7%	
Natashquan	0,86	74%	-0,309	10%	0,25	6%	0,205	4%	-0,171	3%	
Romaine	0,893	80%	-0,49	24%	-0,15	2%	0,04	0%	-0,287	8%	
Blanc Sablon	0,66	44%	-0,58	34%	-0,26	7%	Non_Sig	Non_Sig	-0,198	4%	
Montlouis	0,908	82%	-0,425	18%	0,299	9%	0,2345	5%	-0,193	4%	
Grande Rivière	0,84	71%	-0,696	48%	0,271	7%	0,099	1%	-0,159	3%	
Shediac Valley	0,94	88%	-0,469	22%	Non_Sig	Non_Sig	Non_Sig	Non_Sig	-0,097	1%	
Iles Shag	0,747	56%	-0,443	20%	0,304	9%	-0,104	1%	-0,223	5%	
Borden IPE	0,81	66%	-0,496	25%	0,458	21%	Non_Sig	Non_Sig	-0,159	3%	

 Tableau A-2: Tableau récapitulatif des corrélations entre la température de l'eau et les prédicteurs sélectionnées pour l'ensemble des stations étudiées

Non\_Sig : La corrélation entre la température de l'eau et le prédicteur n'est pas significative à un seuil de 5%

Stations	Attributs sélectionnés	Model	RMSE	$\mathbb{R}^2$	rRMSE	Nash	rBiais
	Tair	MLR	1,18	0,81	15,36	0,81	-0,62
Rimouski	Tair	ANN	1,14	0,83	14,51	0,83	-0,78
	Tair+Flow+WSPD	RF	0,99	0,87	12,91	0,87	-1,32
	Tair	MLR	1,49	0,81	19,34	0,81	-0,91
Courant de	Tair+WSPD+Tidal_range+NAO	ANN	1,37	0,84	18,19	0,84	-1,10
Uaspe	Tair+WSPD+flow	RF	1,37	0,84	16,94	0,84	-1,58
	Tair	MLR	1,29	0,88	13,73	0,85	-2,82
Gyre Anticosti	Tair+WSPD+NAO	ANN	1,26	0,86	13,60	0,86	-2,09
	Tair+NAO+WSPD+Flow	RF	0,77	0,95	7,82	0,86	-0,54
	Tair	MLR	1,36	0,85	17,38	0,85	-1,46
Montlouis	Tair+WSPD	ANN	1,32	0,85	16,87	0,85	-1,65
	Tair+WSPD+flow	RF	1,10	0,90	13,85	0,85	-1,56
	Tair	MLR	1,36	0,81	21,23	0,81	-2,06
Romaine	Tair	ANN	1,35	0,82	23,52	0,82	-2,93
	Tair+WSPD+Flow+Tidal_range	RF	1,06	0,89	17,40	0,82	-2,25
Sept-Îles	Tair	MLR	2,01	0,69	21,80	0,69	-2,54
	Tair	ANN	1,96	0,70	21,26	0,70	-3,17
	Tair+NAO+Tidal_range+Flow	RF	1,18	0,90	13,89	0,70	-2,11
	Tair	MLR	2,08	0,61	32,87	0,60	-5,60
Kiviere aux	Tair	ANN	2,01	0,63	31,74	0,63	-7,81
Tonnene	Tair+Flow+WSPD+Tidal_range	RF	$\begin{array}{c c c c c c c c c c c c c c c c c c c $				
	Tair	MLR	2,06	0,70	27,42	0,70	-5,45
Natashquan	$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	0,71	-6,41				
	Tair+Flow+WSPD	AO+WSPD+FlowRF $0,77$ $0,95$ $7,82$ $0,$ TairMLR $1,36$ $0,85$ $17,38$ $0,$ 'air+WSPDANN $1,32$ $0,85$ $16,87$ $0,$ +WSPD+flowRF $1,10$ $0,90$ $13,85$ $0,$ TairMLR $1,36$ $0,81$ $21,23$ $0,$ TairANN $1,35$ $0,82$ $23,52$ $0,$ D+Flow+Tidal_rangeRF $1,06$ $0,89$ $17,40$ $0,$ TairMLR $2,01$ $0,69$ $21,80$ $0,$ TairMLR $2,01$ $0,69$ $21,80$ $0,$ TairANN $1,96$ $0,70$ $21,26$ $0,$ $TairMLR2,010,6331,740,TairMLR2,010,6331,740,TairMLR2,030,7127,260,TairMLR2,030,7127,260,TairANN2,030,7127,260,TairMLR2,090,5241,820,'+Flow+WSPDRF1,810,7722,520,'+Flidal_rangeMLR2,070,5341,820,'+Tidal_rangeMLR2,470,2939,570,'+Tidal_rangeANN2,310,3836,300,'>D++Flow+NAORF1,800,6226,470,'air$	0,71	-4,30			
	Tair+Tidal_range	MLR	2,09	0,52	41,82	0,52	-7,83
Tabatière	Tair+Tidal_range	ANN	2,07	0,53	41,82	0,52	-8,11
	Tair+WSPD+Tidal_range+Flow+NAO	RF	1,51	0,75	29,49	0,52	-5,08
	Tair	MLR	2,47	0,29	39,57	0,27	-14,25
Havre St Pierre	Tair+WSPD	ANN	2,31	0,38	36,30	0,36	-13,30
	Tair+WSPD++Flow+NAO	RF	1,80	0,62	26,47	0,36	-7,61
	Tair+WSPD	MLR	2,12	0,65	18,16	0,64	-4,12
Ile-Shag	Tair+WSPD+Tidal_range	ANN	2,02	0,69	16,89	0,68	-4,51
	Tair+Flow+WSPD+NAO	RF	1,21	0,89	9,78	0,68	-1,92
	Tair	MLR	2,46	0,32	47,53	0,32	-13,76
Blanc Sablon	Tair+Tidal_range	ANN	2,26	0,43	38,81	0,42	-11,28
	Tair+Tidal_range+Flow+WSPD	RF	1,72	0,67	26,40	0,32	-7,87
Borden	Tair+WSPD	MLR	1,73	0,69	15,53	0,69	-1,21

## Tableau A-3: Résultats des critères des performances des modèles testés suite à la sélection des meilleurs attributs parmi tous prédicteurs sélectionnées

			-	-		-	
	Tair+WSPD	ANN	1,73	0,69	15,64	0,69	-1,29
	Tair+WSPD+Flow+Tidal_range	RF	1,42	0,80	15,27	0,69	-1,41
Shediac Valley	Tair	MLR	1,10	0,88	9,68	0,88	-0,42
	Tair+WSPD	ANN	1,06	0,89	8,88	0,89	-0,41
	Tair+WSPD+Flow	RF	0,77	0,95	6,41	0,89	-0,33
	Tair, Tidal_range	MLR	1,6	0,74	16,68	0,74	-1,469
Grande Rivière	Tair,WSPD,Tidal_range,Flow,NAO	ANN	1,46	0,79	14,67	0,78	-0,987
	Tair,WSPD,Tidal_range,Flow	RF	1,17	0,86	11,8	0,86	-0,988

Tableau A-4: Résultats des critères des performances des modèles testés en utilisant suite à la sélection des meilleurs attributs parmi la Tair et la vitesse du vent

		Attributs					
Stations	Model	sélectionnés	RMSE	$\mathbb{R}^2$	rRMSE	Nash	rBiais
Grande Rivière		Tair	1,751	0,690	18,466	0,690	-2,138
Borden		Tair+WSPD	1,734	0,686	15,533	0,686	-1,210
Blanc Sablon		Tair	2,461	0,324	47,531	0,317	-13,758
Courant de							
Gaspé		Tair	1,490	0,811	19,342	0,810	-0,914
Havre St							
Pierre	МІР	Tair	2,475	0,291	39,573	0,267	-14,246
Ile Shag	MLK (Multiple	Tair	2,221	0,615	19,103	0,609	-4,549
Montlouis	(Multiple	Tair	1,360	0,845	17,378	0,845	-1,461
Natashquan	Regression)	Tair	2,058	0,702	27,425	0,700	-5,454
Rimouski	Regression)	Tair	1,184	0,814	15,357	0,813	-0,616
Rivière aux							
Tonnerre		Tair	2,080	0,607	32,872	0,603	-5,604
Romaine		Tair	1,361	0,812	21,229	0,811	-2,058
Tabatière		Tair	2,389	0,367	51,540	0,367	-9,434
Sept-Îles		Tair	2,010	0,688	21,800	0,685	-2,539
Shediac Valley		Tair	1,102	0,884	9,677	0,884	-0,416
Moyenne			1,834	0,653	25,488	0,649	-4,600
Grande Rivière		Tair	1,700	0,708	17,506	0,708	-2,050
Borden		Tair	1,991	0,586	16,824	0,585	-1,529
Blanc Sablon		Tair	2,437	0,338	46,410	0,331	-13,774
Courant de	ANN						
Gaspé	(Artificial	Tair+WSPD	1,368	0,841	18,172	0,840	-1,235
Havre St	Neural						
Pierre	Network)	Tair	2,435	0,306	37,999	0,290	-12,670
Ile Shag		Tair+WSPD	2,097	0,656	18,127	0,651	-4,008
Montlouis	]	Tair+WSPD	1,323	0,854	16,875	0,854	-1,647
Natashquan		Tair	2,028	0,711	27,275	0,709	-6,409

Rimouski		Tair	1,139	0,828	14,512	0,827	-0,779
Rivière aux							
Tonnerre		Tair	2,015	0,633	31,737	0,627	-7,809
Romaine		Tair	1,345	0,816	23,504	0,816	-2,929
Tabatière		Tair	2,342	0,391	50,043	0,391	-9,577
Sept-Îles		Tair	1,957	0,704	21,259	0,702	-3,175
Shediac Valley		Tair	1,056	0,893	9,005	0,893	-0,568
Moyenne			1,802	0,662	24,946	0,659	-4,868
Grande Rivière		Tair+WSPD	1,654	0,725	16,409	0,708	-1,877
Borden		Tair+WSPD	1,837	0,650	16,947	0,585	-0,660
Blanc Sablon		Tair+WSPD	2,290	0,418	38,967	0,331	-10,614
Courant de							
Gaspé		Tair+WSPD	1,370	0,840	18,214	0,840	-1,018
Havre St							
Pierre		Tair+WSPD	2,519	0,307	36,931	0,290	-10,844
Ile Shag	RF	Tair+WSPD	1,958	0,702	16,515	0,685	-3,478
Montlouis	(Rondom	Tair+WSPD	1,418	0,804	15,273	0,687	-1,407
Natashquan	Forest)	Tair+WSPD	1,418	0,804	15,273	0,687	-1,407
Rimouski		Tair+WSPD	1,157	0,823	14,752	0,827	-0,794
Rivière aux							
Tonnerre		Tair+WSPD	2,080	0,613	30,344	0,627	-6,377
Romaine		Tair+WSPD	1,404	0,800	22,190	0,816	-2,496
Tabatière		Tair+WSPD	2,323	0,412	50,656	0,391	-7,193
Sept-Îles		Tair+WSPD	1,982	0,696	21,045	0,702	-3,676
Shediac Valley		Tair+WSPD	1,020	0,901	7,952	0,893	-0,430
Moyenne			1,745	0,678	22,962	0,648	-3,734

Stations	Corrected Zc	new P- value	N/N*	Original Z	old P- value	Tau	Sen's slope	old variance	new variance	Slope/10 years	Slope/60years
Natashquan	39,36	0,00E+00	0,48	27,22	3,37E-163	0,19	4,59E-04	8,60E+10	4,11E+10	0,711	4,266
Baie de Cascapedia	23,08	6,65E-118	1,41	27,38	5,14E-165	0,19	1,69E-04	8,60E+10	1,21E+11	0,262	1,570
Baie de Trascapedia	22,28	6,46E-110	1,51	27,38	5,13E-165	0,19	1,73E-04	8,60E+10	1,30E+11	0,269	1,613
Sept-Îles	26,67	9,23E-157	0,86	24,77	1,70E-135	0,17	2,42E-04	8,60E+10	7,42E+10	0,376	2,255
Baie de Plaisance	13,82	1,91E-43	4,22	28,40	2,00E-177	0,20	1,34E-04	8,60E+10	3,63E+11	0,208	1,245
Lagune Havre aux Maisons	24,89	8,87E-137	1,30	28,40	2,00E-177	0,20	3,31E-04	8,60E+10	1,12E+11	0,513	3,079
Bassin Havre Aubert	26,94	7,28E-160	1,11	28,40	2,00E-177	0,20	3,71E-04	8,60E+10	9,55E+10	0,575	3,447
Lagune de la Grande Entrée	21,21	7,92E-100	1,79	28,40	2,00E-177	0,20	3,16E-04	8,60E+10	1,54E+11	0,489	2,935
Baie des Belles Amours	40,00	0,00E+00	0,51	28,50	1,19E-178	0,20	2,81E-04	8,60E+10	4,36E+10	0,436	2,618
Blanc Sablon	35,17	4,99E-271	0,66	28,50	1,19E-178	0,20	2,64E-04	8,60E+10	5,64E+10	0,410	2,458
Borden (IPE)	20,14	3,02E-90	1,85	27,37	5,77E-165	0,19	1,49E-04	8,60E+10	1,59E+11	0,232	1,390
Grande Rivière	18,35	3,05E-75	2,67	30,01	7,90E-198	0,21	1,66E-04	8,60E+10	2,30E+11	0,257	1,540
Iles Shag	14,53	8,37E-48	3,82	28,40	2,00E-177	0,20	1,50E-04	8,60E+10	3,29E+11	0,233	1,400
Havre St Pierre	17,37	1,34E-67	2,35	26,66	1,54E-156	0,19	1,75E-04	8,60E+10	2,02E+11	0,271	1,624
Rivière aux Tonnerre	47,36	0,00E+00	0,40	29,92	9,34E-197	0,21	4,10E-04	8,60E+10	3,43E+10	0,635	3,813
Romaine	44,80	0,00E+00	0,56	33,49	7,14E-246	0,23	4,49E-04	8,60E+10	4,80E+10	0,696	4,176

Tableau A-5: Résultats de test de la tendance sur les futures températures de l'eau en utilisant le test du Mann Kendall Modifié

Tabatière	41,16	0,00E+00	0,50	29,15	7,44E-187	0,20	3,30E-04	8,60E+10	4,31E+10	0,512	3,072
Rimouski	50,82	0,00E+00	0,57	38,38	0,00E+00	0,27	3,77E-04	8,60E+10	4,90E+10	0,584	3,506
Courant de Gaspé	24,76	2,44E-135	1,42	29,49	4,26E-191	0,21	2,57E-04	8,60E+10	1,22E+11	0,399	2,391
Gyre Anticosti	22,58	7,12E-113	1,20	24,77	1,70E-135	0,17	2,02E-04	8,60E+10	1,04E+11	0,312	1,875
Montlouis	29,35	2,42E-189	1,01	29,49	4,26E-191	0,21	2,63E-04	8,60E+10	8,68E+10	0,407	2,443
Shediac Valley	17,18	3,65E-66	3,05	30,01	7,90E-198	0,21	2,03E-04	8,60E+10	2,62E+11	0,315	1,889

# Annexe B



Figure B-1 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois du Juille



Figure B-2 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois du Juillet



Figure B-3: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois du Juille



Figure B-4: Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario optimiste RCP4.5 durant le mois de Septembre


Figure B-5 : Cartes d'interpolation d'indicateur du risque d'infection par la méthode IDW dans l'estuaire et Golfe Saint Laurent dans un scénario pessimiste RCP8.5 durant le mois de Septembr



Figure B-6: Cartes d'interpolation d'indicateur du risque d'infection, par la méthode IDW, dans l'estuaire et Golfe Saint Laurent pour l'échéance (2080-2100) dans les scénarios (RCP4.5 et RCP8.5) durant le mois de Septembre