



Site de Nancy - LABORATOIRE D'HYDROLOGIE DE NANCY

MODELISATION DE LA FORMATION DES THM DANS LES RESEAUX DE DISTRIBUTION D'EAU DESTINEE A LA CONSOMMATION HUMAINE EN FRANCE

Note de synthèse
Juin 2013

ANSES - Laboratoire d'Hydrologie de Nancy
40, rue Lionnois
54000 NANCY

Téléphone : + 33 (0)3 83 38 87 20 - Télécopie : + 33 (0)3 83 38 87 21 - www.anses.fr

InVS

12 rue du Val d'Osne
94 415 SAINT-MAURICE Cedex France

Téléphone : + 33 (0)1 41 79 67 00 - Télécopie : + 33 (0)1 41 79 67 67 - www.invs.sante.fr

Modélisation de la formation des THM dans les réseaux de distribution d'eau destinés à la consommation humaine en France.

Note de synthèse du travail effectué par Otmane BOUDOUCHE, dans le cadre d'un travail post-doctoral, co-encadré par l'Anses et l'InVS (C. Rosin¹, E. Joulin¹, C. Galey², A. Zeghnoun²).

1 : Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSeS) Direction de l'Evaluation des Risques, Unité évaluation des risques liés à l'eau, 27-31 Avenue du Général Leclerc BP19, 94701 Maisons-Alfort Cedex

2 : Institut de veille sanitaire (InVS), 12 rue du Val d'Osne, 94415 Saint Maurice Cedex

Résumé :

Dans le contexte de la sur-chloration des réseaux de distribution d'eau en lien avec d'éventuelles pollutions intentionnelles, il apparaît primordial d'avoir une bonne connaissance de l'exposition de la population aux sous-produits de chloration (SPC) qui sont susceptibles d'être à l'origine de certains cancers chez l'homme. En particulier, les trihalométhanes (THM), réglementés par le Code de la santé publique, font l'objet de contrôles réguliers dans l'eau traitée. Les teneurs en THM en réseau de distribution dépendent de nombreux facteurs parmi lesquels la teneur initiale en THM en sortie de station de traitement, la dose de chlore appliquée, la présence de précurseurs organiques, la température de l'eau, la présence de stations de rechloration et le temps de séjour dans le réseau. Afin de prédire l'évolution des teneurs en THM en réseau de distribution, des premiers travaux de modélisation ont été mis en œuvre en 2009 sur trois sites, par l'Institut de veille sanitaire (InVS) et l'Afssa (maintenant Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail - Anses). La comparaison des prédictions fournies par ce modèle aux données mesurées en cinq autres sites n'a toutefois pas permis d'établir la validité du modèle au-delà des trois sites considérés pour sa conception, bien que l'ensemble des huit sites soit alimenté par des eaux de surface.

Par conséquent, il est apparu nécessaire de faire évoluer ce premier modèle afin de disposer d'un modèle applicable à un domaine plus étendu et plus représentatif de la diversité de qualité des eaux de surface distribuées en France. Deux modèles ont été ainsi construits sur l'ensemble des données (données issues des trois sites ayant servi à l'établissement du précédent modèle et des cinq sites ayant servi à sa validation externe) : un modèle simplifié fondé sur un minimum de variables, en privilégiant celles qui sont facilement accessibles en routine (présentes dans la base nationale du contrôle sanitaire de l'eau destinée à la consommation humaine - Sise-eaux) et un modèle complet incorporant deux variables supplémentaires qui apportent un complément d'information sur la réactivité de la matière organique avec le chlore. Les modèles ont été conçus (choix des variables et relations entre elles) sur la base d'un jeu de données constitué par tirage aléatoire de 3/4 des données contenues dans la base de données complète. L'autre quart des données contenues dans la base a été utilisé pour la validation des modèles.

Les deux modèles, simplifiés et complet, expliquent respectivement 87 % et 88 % de la variance des données utilisées pour les tests de validation, et tous deux sont robustes. La capacité de généralisation a été estimée par la mesure de la fréquence des prévisions présentant moins de 25 % d'erreur relative (fréquence proche de 68 % pour les 2 modèles), et moins de 5 % d'erreur relative en prenant en compte les intervalles de confiance des données utilisées (respectivement 81 et 86 % pour les modèles simplifié et complet).

Au final, ces deux modèles apparaissent utilisables pour prédire des niveaux de THM en un point d'un réseau alimenté par une eau de surface, sous réserve de la disponibilité des données concernant les variables incluses dans les modèles pour le réseau considéré. Ces modèles peuvent notamment servir à déterminer les niveaux de concentrations en THM en différents points d'un même réseau, et ainsi contribuer à identifier les zones les plus critiques, proches du seuil réglementaire par exemple.

1. Contexte :

La chloration de l'eau potable est largement utilisée dans le monde pour prévenir et limiter le risque infectieux véhiculé par l'eau du robinet. En France, son utilisation date de plus d'un siècle dans plusieurs grandes villes. Depuis 2003, les autorités françaises ont recommandé d'étendre son utilisation à l'ensemble des réseaux d'eau quelle que soit la taille de la population desservie. En 2007, plus de 99 % des débits produits sont associés à une désinfection [1].

De par ses propriétés oxydantes, le chlore réagit avec la matière organique de l'eau, les ions bromure et iodure, pour former des sous-produits de chloration (SPC). Près de 600 SPC sont identifiés à ce jour [2]. Parmi les familles majoritaires, les trihalométhanes (THM) et les acides haloacétiques (HAA) représentent à elles deux entre 20 % et 30 % de la masse totale des SPC [3]. En France, les THM sont réglementés par le Code de la Santé Publique [4] et font l'objet de contrôles réguliers dans l'eau distribuée. Les prélèvements d'eau s'effectuent en sortie des stations de traitement disposant d'une étape de chloration, et en réseau si la teneur en chlore dans le réseau dépasse 0,5mg/L ou s'il y a une rechloration. La formation des SPC dépend de la nature de l'eau brute, des traitements mis en place pour éliminer la matière organique et de la stratégie de désinfection (points d'injection, doses appliquées, temps de contact).

La présence de SPC pose un problème de santé publique en raison des risques sanitaires associés et de la taille de la population exposée. Les études épidémiologiques indiquent une association entre l'exposition aux SPC, évaluée généralement par les mesures de THM effectuées dans le cadre des contrôles réglementaires, et l'apparition de cancers de la vessie, chez les hommes uniquement [5]. Une association entre THM et cancer colorectal est par ailleurs évoquée, sans consensus sur le sujet à ce jour [6]. Les effets suspectés sur la reproduction et le développement, bien que largement étudiés, sont toujours controversés [7-11]. L'estimation de l'exposition reste de façon générale le point faible des études épidémiologiques.

La formation des THM évolue dans le réseau de distribution de l'eau. Plusieurs études ont montré une augmentation des concentrations des THM d'un facteur 2 à 6 entre la sortie de l'usine et la périphérie du réseau de distribution d'eau potable [12].

2. Objectifs :

Un premier modèle de régression a été construit sur trois sites de production et de distribution d'eau destinée à la consommation humaine en 2009 par l'Anses et l'InVS [12,13] pour prédire les concentrations des THM dans les réseaux d'eau à partir de données mesurées à la sortie des usines de traitement, dans l'objectif de mieux estimer l'exposition de la population française aux THM. Les données provenant de 5 autres sites de production et de distribution ont été utilisées à des fins de validation externe. La comparaison des prédictions fournies par ce modèle (dit « 2009 ») aux données mesurées sur ces 5 sites n'a toutefois pas permis d'établir la validité du modèle au-delà des 3 sites considérés pour sa conception.

L'objectif de l'étude est donc de proposer deux variantes d'un nouveau modèle de régression fondé sur l'analyse de l'ensemble des données (données issus des 3 sites ayant servi à l'établissement du modèle « 2009 » et des 5 sites ayant servi à sa validation externe), afin de disposer d'un modèle applicable à un domaine plus étendu que le précédent. Un modèle complet utilisant toutes les variables fournies par les exploitants des différents sites a été construit, ainsi qu'un modèle simplifié retenant un sous-ensemble minimal de variables, réduit à celles qui sont indispensables, ou facilement accessibles et produites en routine.

3. Matériel et méthode :

3.1. La base de données du contrôle sanitaire de l'eau :

Comme pour l'ensemble des paramètres du contrôle sanitaire de l'eau, les résultats concernant les niveaux de THM sont enregistrés dans la base de données nationale SISE-Eaux gérée par le ministère de la Santé. Cette base, qui rassemble à ce jour plus de 70 millions de résultats d'analyses depuis les années 1995, est régulièrement utilisée pour estimer l'exposition de la population et contribue à la surveillance des risques d'origine hydrique en France. Néanmoins, dans le cas des THM, les données contenues dans cette base ne permettent pas d'estimer précisément les niveaux d'exposition au robinet du consommateur, car la variabilité spatiale des concentrations en THM n'est généralement pas prise en compte dans le programme de contrôle. En revanche, les informations contenues dans la base SISE-Eaux permettent d'avoir une bonne connaissance des concentrations à la sortie de toutes les usines de traitement d'eau disposant d'une étape de chloration.

3.2. Sites d’étude :

Huit sites ont été utilisés pour la construction et la validation des modèles. Tous ces sites sont alimentés par une eau de surface ou de retenue, et comportent une filière de traitement complète avec une étape d’affinage par filtration sur charbon actif en grains ou filtration bi-couche, et une étape d’ozonation. Il n’y pas d’étape de préchloration. La désinfection finale par du chlore est effectuée avant la mise en distribution de l’eau dans le réseau.

Les données proviennent de différentes campagnes de prélèvements d’échantillons et d’analyses, effectuées à des saisons différentes. Lors de chaque campagne, un prélèvement d’eau a été systématiquement effectué en sortie d’usine de traitement, à l’aval du point de chloration en usine (point de référence 0), et un à plusieurs prélèvements ont été effectués en différents points du réseau, avant ou après une éventuelle étape de rechloration dans ce dernier.

En conséquence, les données complètes utilisées se répartissent de la manière suivante entre les différents sites :

site	Nombre de campagnes	Nombre total de valeurs de THM en réseau	Temps de séjour hydraulique (min-max, en heures)
Sites 2009 -1	3	48	(11-27)
Sites 2009 -2	3	62	(26-160)
Sites 2009 -3	3	55	(30-210)
Site 4	4	16	(64-160)
Site 5	7	48	(5-280)
Site 6	7	14	(19-57)
Site 7	7	16	(5-53)
Site 8	2	3	(15-53)
total		262	

Tableau1 : Synthèse sur les échantillonnages des sites d’étude.

3.3. Domaine de variation des paramètres étudiés :

Le tableau 2 présente la description des variables de qualité d'eau et des variables d'exploitation qui peuvent influencer la formation des THM. L'incorporation de ces variables dans les modèles « simplifié » et « complet » est stipulée.

Variable	intitulé	unité	min	max	Modèle simplifié	Modèle complet
Variabiles du contrôle sanitaire, disponibles dans Sise-eaux						
THM ₀	[THM] au point de mise en distribution (sortie usine)	µg/L	1,3	68	-	-
		µmol/L	0,01	0,5	oui	oui
Cl ₂₀	Résiduel de chlore sortie usine	mg/L	0,05	1,3	oui	oui
Temp ₀	Température de l'eau	° C	7	23	oui	oui
COT ₀	Carbone organique total sortie usine	mg/L	1,1	4,0	oui	oui
pH ₀	pH sortie usine		7,2	8,5	oui	oui
Variabiles d'exploitation, dépendant du réseau, non réglementées						
Cl _{2inj}	Dose de chlore injectée dans le réservoir de chloration de la filière de traitement	mg/L	1,2	6,0	non	oui
TCtp	Temps de contact dans le réservoir de chloration de la filière de traitement	heures	0,5	6,9	non	oui
TSH _i	Temps de séjour hydraulique entre le point i du réseau et la sortie usine	heures	4,5	280	oui	oui
RCP _i	Présence d'une rechloration en amont du point i considéré.		Variable qualitative (Oui / Non)		oui	oui
Br ⁻	Ions bromures sortie usine	mg/L	0.003	0.97	non	non
Absuv ₀	Absorbance UV à 254nm, sortie usine	m ⁻¹	0,003	0,08	non	oui

Tableau.2 : Covariables des modèles « simplifié » et « complet » de formation des THM dans les réseaux de distribution et domaines de variation observés – Données provenant des 8 sites d'étude. Le point i est le point en réseau pour lequel la teneur en THM est estimée.

3.4. Modélisation :

La méthode utilisée pour ajuster ces modèles repose sur le découpage aléatoire des données en deux sous-échantillons. Le premier, appelé échantillon d'apprentissage ou d'entraînement, est constitué de 75 % des données disponibles et est utilisé pour construire le modèle. Le second, appelé échantillon test ou de validation, est constitué des 25 % des données restantes et est utilisé pour mesurer la capacité de généralisation du modèle en comparant ses prédictions aux valeurs observées.

Les co-variables sont introduites sous forme de fonctions polynomiales de degré 1 à 3, pour prendre en compte la possible non linéarité de la relation entre les niveaux de THM présents dans le réseau et les variables explicatives. Différents modèles de régression ont ensuite été testés avec les variables ainsi retenues, en introduisant d'éventuelles interactions. Ces modèles ont été appréciés en considérant :

1. le coefficient de détermination R^2 qui permet de connaître la contribution des variables testées dans l'explication de la variabilité de la réponse, et l'erreur standard résiduelle (RMSE) qui correspond à l'erreur faite sur la prédiction ;
2. l'analyse de la qualité d'ajustement du modèle aux données en regardant graphiquement la distribution des résidus ;
3. une validation sur les données non utilisées pour la construction du modèle (l'échantillon de test), appréciée sur la base de :
 - a. la racine carrée d'erreur quadratique moyenne (RMSE),
 - b. l'erreur relative (N25%), qui représente le pourcentage des prédictions ayant une erreur relative inférieure à 25 % en valeur absolue, et
 - c. l'erreur relative liée à l'incertitude (N5inc), qui représente le pourcentage des prédictions dont la valeur de l'erreur relative en valeur absolue calculée sur les intervalles de confiance est inférieure à 5 %¹.

Plus ces deux dernières erreurs sont élevées, plus la capacité de prévision ou de généralisation du modèle est importante.

La stabilité des deux modèles retenus a été vérifiée par validation croisée sur 8 sous-échantillons constitués aléatoirement à partir de l'échantillon de départ.

Le travail a été effectué avec le logiciel R (V2.14.2).

¹La valeur de l'erreur relative liée à l'incertitude représente la plus petite différence entre les extrémités des deux intervalles de confiance à 95%.

4. Résultats

4.1. Modèle simplifié :

Après exploration de la relation entre les THM_i et les variables explicatives disponibles, la forme du modèle simplifié est une forme polynomiale, de degré 1 à 3 selon les variables, présentant un terme d'interaction entre une rechloration en réseau et la température de l'eau (tableau 3).

Modèle simplifié- coefficients utilisables sous Excel	coefficient	erreur standard	valeur de t	Pr(> t)
Constante	1,45E+02	4,01E+01	3,6	0,0004
THM_0	1,25E+00	1,19E-01	10,508	< 0,0001
THM_0^2	-1,24E+00	2,67E-01	-4,638	< 0,0001
Cl_2_0	8,10E-02	1,69E-02	4,807	< 0,0001
TSH_i	1,20E-03	3,94E-04	3,057	0,0025
TSH_i^2	-9,49E-06	3,38E-06	-2,804	0,0055
TSH_i^3	2,61E-08	7,91E-09	3,297	0,0011
pH_0	-5,50E+01	1,54E+01	-3,575	0,0004
pH_0^2	6,97E+00	1,96E+00	3,546	0,0005
pH_0^3	-2,94E-01	8,35E-02	-3,518	0,0005
COT_0	1,11E-01	3,11E-02	3,559	0,0004
COT_0^2	-2,02E-02	5,87E-03	-3,444	0,0007
rcpi[1]	3,33E-01	7,54E-02	4,421	< 0,0001
Temp ₀ (sans rechloration)	-8,61E-03	8,07E-03	-1,067	0,2870
Temp ₀ ² (sans rechloration)	4,90E-04	2,75E-04	1,783	0,0758
Temp ₀ (avec rechloration)	-4,84E-02	8,45E-03	-5,73	< 0,0001
Temp ₀ ² (avec rechloration)	1,75E-03	2,90E-04	6,05	< 0,0001

Tableau 3. Résultats du modèle simplifié obtenus en utilisant 75% des données observées

Les performances de ce modèle sont les suivantes :

Construction sur l’échantillon d’apprentissage (75% des données) (N=197)		
R² = 87.15 %	RMSE = 0.0484	p < 2,2e-16
Validation sur l’échantillon test (25% des données)(N=65)		
RMSE = 0,0625	N ₂₅ = 67.7 %	N _{5inc} = 81.5 %

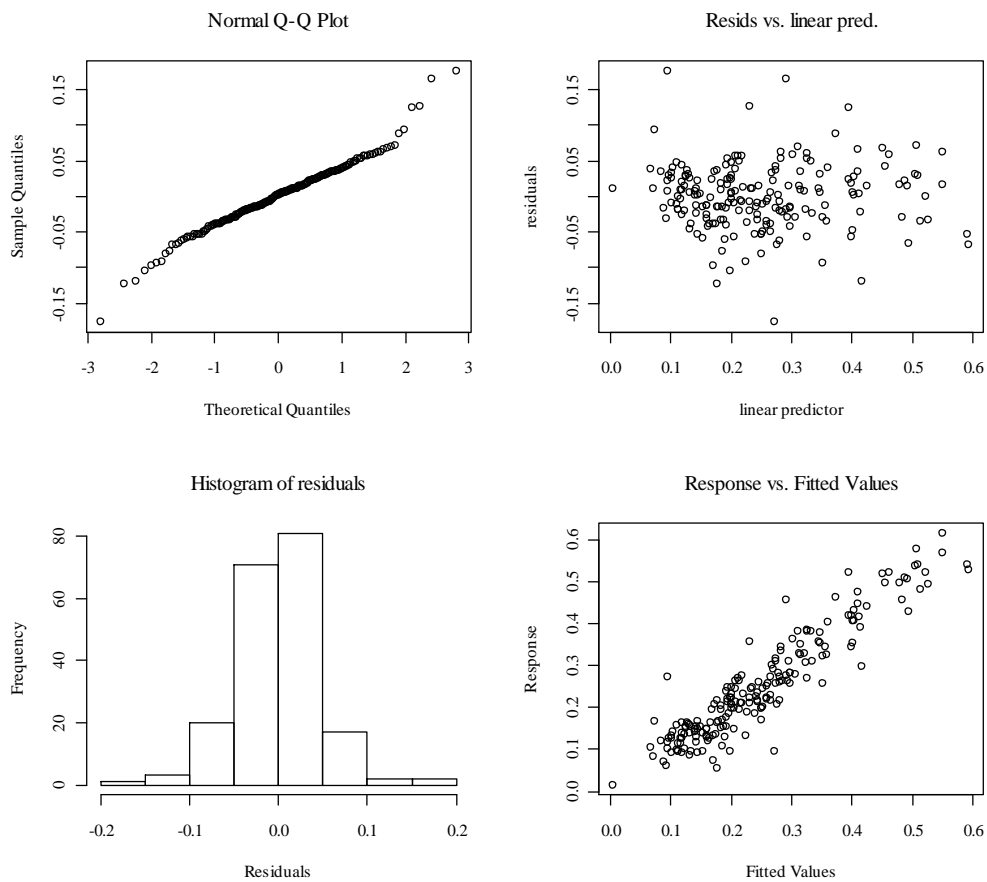


Figure 1 : Qualité d’ajustement du modèle simplifié

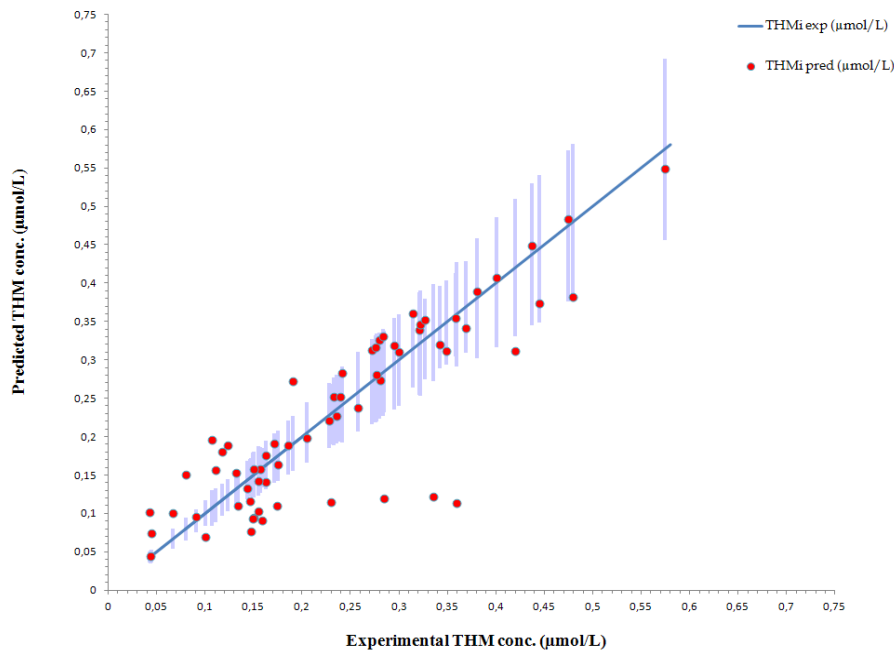


Figure 2. Validation du modèle simplifié sur l'échantillon test. Graphique des concentrations prédites par le modèle (THM_i en réseau) selon les concentrations expérimentales.

Les quatre valeurs expérimentales comprises entre 0,23 et 0,37 µmole/L, pour lesquels le modèle simplifié prédit une valeur autour de 0,1 µmole/L, appartiennent au même site (le site 7). Elles ont toutes été mesurées au printemps, 3 durant la même campagne.

Les 4 points de prélèvements sont différents, mais présentent une double chloration en réseau (pour 3 sur 4) ou un temps de contact mal connu et qui a pu être l'objet d'une forte sous-estimation (information >40 heures ou >48 heures, pour 3 sur 4. Les valeurs des temps de contacts entrés dans la base ont respectivement été 45 et 53 heures. Les temps réels pourraient dépasser les 200 heures).

La forme des relations entre les niveaux de THM présents dans le réseau et les variables explicatives retenues dans le modèle simplifié est présentée dans la figure 3.

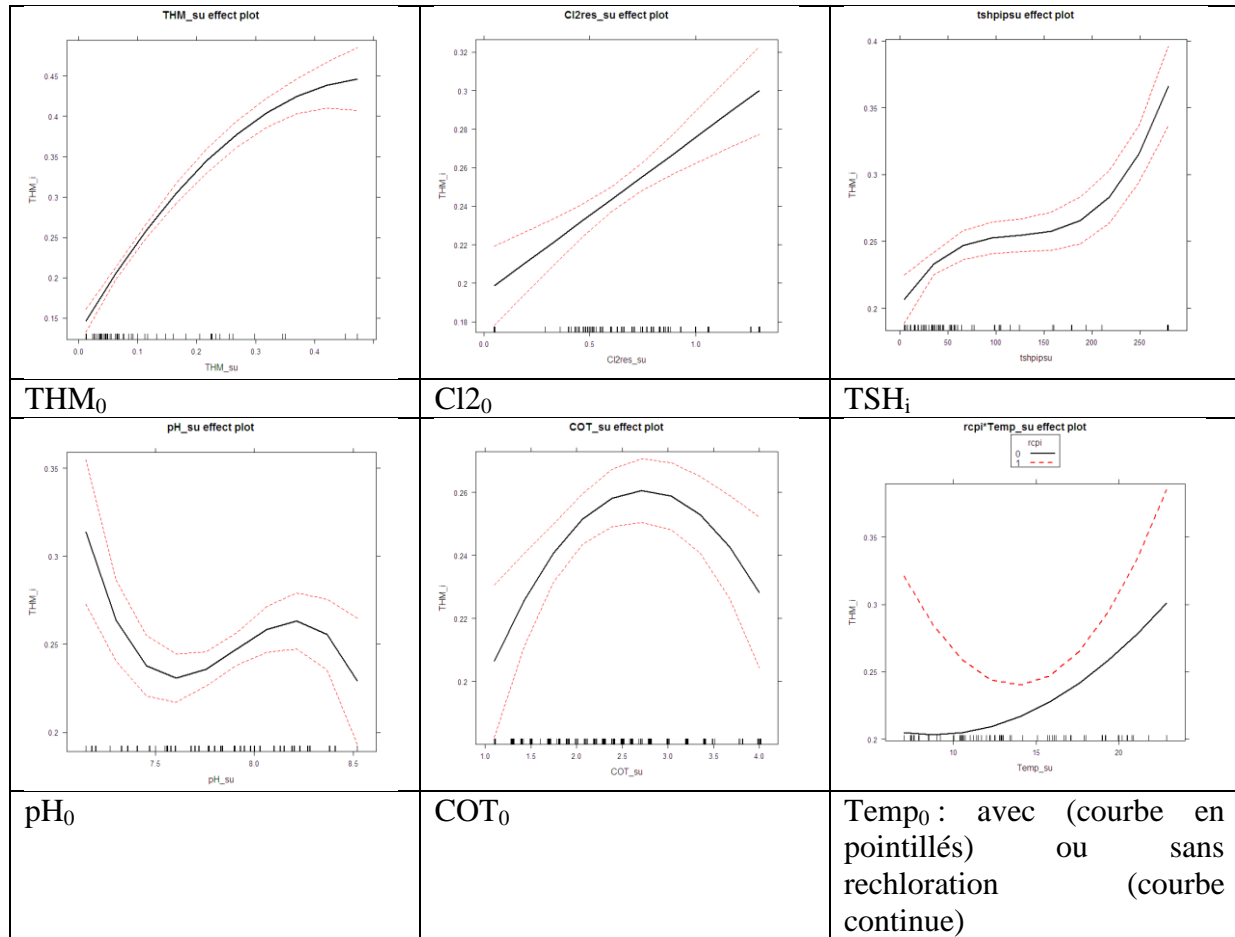


Figure 3. Relations entre les niveaux de THM présents dans le réseau et les variables explicatives retenues dans le modèle simplifié.

Une relation croissante est observée entre la formation des THM_i en réseau et les variables THM₀ (THM en sortie d'usine), Cl₂₀ (résiduel de chlore en sortie d'usine), TSH_i (temps de séjour de l'eau au point i de prélèvement), Temp₀ (température de l'eau) en absence de chloration.

Les relations observées pour les forts COT₀ (carbone organique de l'eau distribuée supérieure à 3,5 mg/L), fort pH (pH>8,3) et faibles températures (<10°C) en présence d'une chloration échappent à une explication simple. Les campagnes associées aux forts COT₀ et pH, et aux faibles températures sont en nombre limité et concernent uniquement certains sites : pour le COT₀, 2 sites et 3 campagnes (sur un total de 8 sites et 36 campagnes) ; pour le pH, 2 sites et 4 campagnes et pour la température avec chloration 4 sites et 7 campagnes. Ces 3 situations sont rencontrées simultanément dans une seule campagne.

4.2. Modèle complet :

Après exploration de la relation entre les THM_i et les variables explicatives disponibles, la forme du modèle complet est une forme polynomiale, de degré 1 à 3 selon les variables, présentant un terme d'interaction entre une rechloration en réseau et la température de l'eau (tableau 4). Le modèle complet utilise la variable R définie ci-dessous (proxi de la réactivité de la matière organique de l'eau au chlore dans l'usine. C'est une vitesse de consommation du chlore, dans l'usine), ainsi que l'absorbance UV (254nm) de l'eau.

$$R = (Cl_{2inj} - Cl_{20}) / TCtp$$

Modèle complet- coefficients utilisables sous Excel	coefficient s	erreur standard	valeur de t	Pr(> t)
Constante	1,39E+02	4,01E+01	3,458	0,0006
THM ₀	1,30E+00	1,32E-01	9,907	< 0,0001
THM ₀ ²	-1,55E+00	3,12E-01	-4,966	< 0,0001
Cl ₂₀	6,22E-02	1,80E-02	3,457	0,0006
TSH	1,03E-03	3,82E-04	2,682	0,0078
TSH ²	-7,36E-06	3,25E-06	-2,263	0,0245
TSH ³	2,13E-08	7,57E-09	2,817	0,0053
pH ₀	-5,32E+01	1,54E+01	-3,461	0,0006
pH ₀ ²	6,79E+00	1,96E+00	3,464	0,0006
pH ₀ ³	-2,89E-01	8,33E-02	-3,465	0,0006
COT ₀	1,37E-01	4,27E-02	3,213	0,0015
COT ₀ ²	-2,76E-02	8,41E-03	-3,279	0,0012
RCP _i	2,58E-01	7,21E-02	3,581	0,0004
R	-9,95E-02	2,95E-02	-3,371	0,0009
R ²	2,75E-02	1,22E-02	2,252	0,0252
R ³	-1,93E-03	1,40E-03	-1,383	0,1680
Absuv ₀	4,23E+00	2,43E+00	1,739	0,0834
Absuv ₀ ²	-1,07E+02	5,92E+01	-1,81	0,0716
Absuv ₀ ³	8,25E+02	4,32E+02	1,911	0,0572
Temp ₀ (sans rechloration)	-1,56E-02	7,68E-03	-2,034	0,0430
Temp ₀ ² (sans rechloration)	6,91E-04	2,61E-04	2,646	0,0087
Temp ₀ (avec rechloration)	-4,59E-02	8,10E-03	-5,67	< 0,0001
Temp ₀ ² (avec rechloration)	1,65E-03	2,78E-04	5,95	< 0,0001

Tableau 4. Résultats du modèle complet obtenus en utilisant 75% des données observées

Les performances de ce modèle sont les suivantes :

Construction sur l'échantillon d'apprentissage (75% des données) (N=197)		
$R^2 = 88.45 \%$	RMSE = 0.0467	$p < 2,2e-16$
Validation sur l'échantillon test (25% des données)(N=65)		
RMSE = 0,0563	$N_{25} = 67.7 \%$	$N_{5inc} = 86.1 \%$

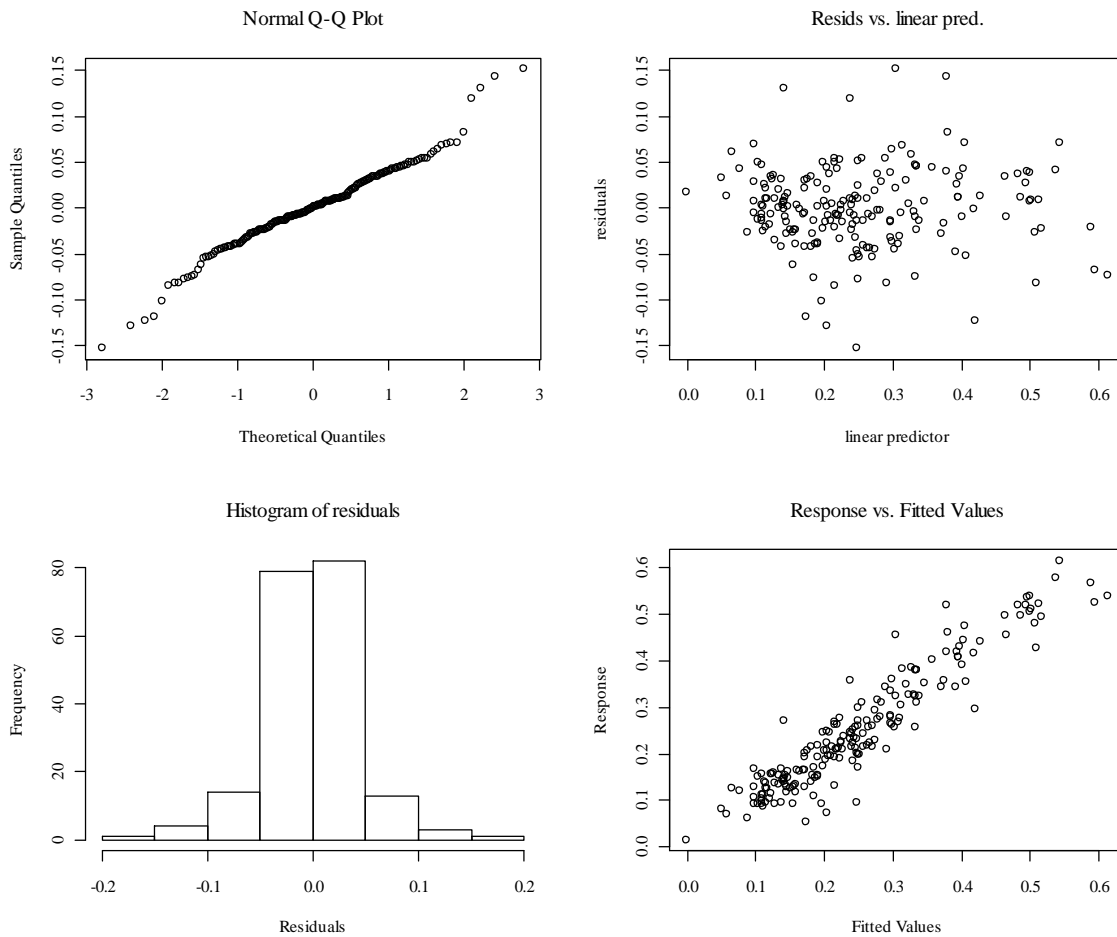


Figure 4. Qualité d'ajustement du modèle complet

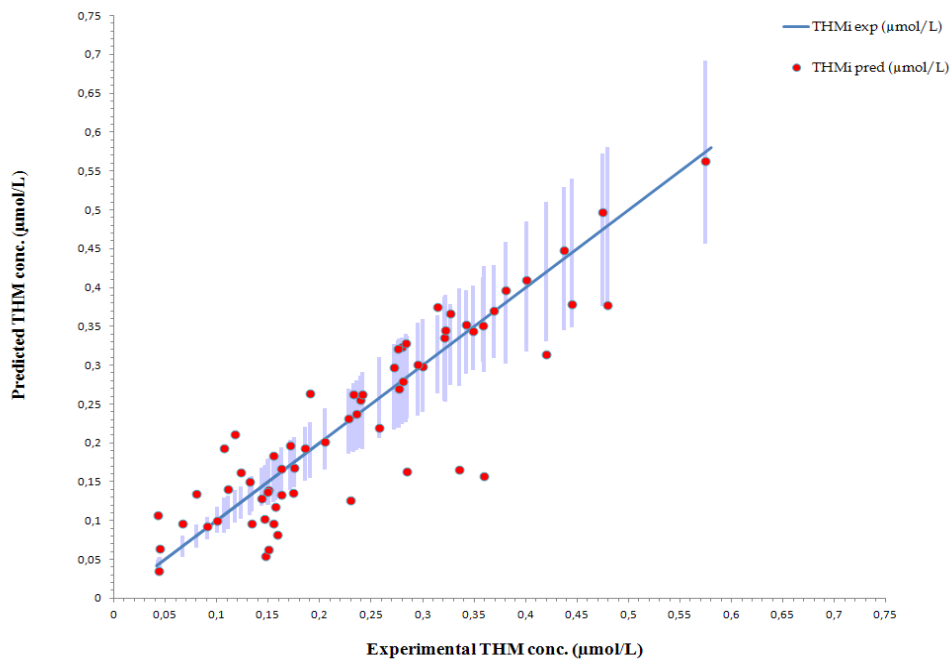


Figure 5. Validation du modèle complet sur l'échantillon test. Graphique des concentrations prédites par le modèle (THM_i en réseau) selon les concentrations expérimentales.

Les mêmes 4 points sont les moins bien validés.

La forme des relations entre les niveaux de THM présents dans le réseau et les variables explicatives retenues dans le modèle complet est similaire à celle observée pour le modèle simplifié (résultat non présenté).

5. Discussion et conclusion :

Le modèle de formation des THM en réseau construit en 2009 [13] à partir de données issues de 3 sites de production et de distribution d'eau n'avait pas été validé sur les nouvelles données recueillies à partir d'autres sites. La qualité de l'eau produite par les 3 sites initiaux était assez semblable, avec des teneurs en THM moyenne à forte (de 10 à près de 90 µg/L).

Une nouvelle modélisation a alors été entreprise, à partir de données provenant de 8 sites différents (les 3 sites utilisés pour la construction du modèle 2009, et les 5 sites utilisés pour sa validation). Tous ces sites sont alimentés par une eau de surface (de rivière ou de retenue), et comportent une filière de traitement complète avec une étape d'ozonation et une étape d'affinage par filtration sur charbon actif en grains ou sur filtres bicouches.

Deux modèles ont été construits. Le 1^{ier} est dénommé « simplifié ». Il est construit à partir des données disponibles dans le cadre du contrôle sanitaire (base nationale Sise-eaux) et des variables indispensables pour modéliser la formation des THM dans les réseaux de distribution (temps de séjour hydraulique et présence d'une rechloration dans le réseau). L'autre modèle est dénommé « complet ». Il est construit à partir de toutes les variables disponibles (issues du contrôle sanitaire, et de l'exploitation). Par rapport au modèle « simplifié », il comprend en plus des variables caractérisant mieux la réactivité de la matière organique de l'eau au chlore que le paramètre « COT » du modèle simplifié : absorbance UV et vitesse de consommation du chlore dans le bassin de chloration de l'usine.

Les performances de ces deux modèles sont très similaires, avec une légère amélioration quand on passe du modèle simplifié au modèle complet (augmentation du R^2 de 87,15 % à 88,45 %, même N_{25} à 67,7 % et augmentation du N_{5inc} de 81,5 à 86,1 %). Le domaine d'application de ces modèles semble couvrir des eaux de surface et les conditions françaises de traitement de l'eau pour la rendre potable, pour un large domaine de contamination dès la mise en distribution de l'eau en sortie de station de traitement (entre 1,3 à 68 µg/L de THM). Le domaine de validité des modèles (tableau 2) est sans doute à restreindre pour certaines co-variables ($COT_0 > 3,5$ mg/L ; $pH_0 > 8,3$; température de l'eau $< 10^\circ C$ et rechloration). L'utilisation de la validation croisée a permis de vérifier la stabilité des modèles proposés. En moyenne, les coefficients et les performances des modèles (en termes de prédictions) obtenues sont proches de ceux obtenus avec le modèle construit sur 75 % des données.

Au final, la validité globale des modèles « simplifié » et « complet » amène à proposer leur utilisation pour calculer des valeurs de THM_i dans un réseau.

Les conditions d'exploitation du réseau contribuent à la poursuite de la formation des THM. Toutefois les données permettant de modéliser finement leur influence sont pauvres. Les modèles prennent ainsi simplement en compte une estimation « grossière » du temps de séjour de l'eau dans le réseau entre la sortie de l'usine et le point de prélèvement en réseau, et la présence ou non de rechloration avant le point de prélèvement (sans connaissance de la dose injectée en rechloration, ni du nombre des rechlorations). L'incertitude très élevée sur le

paramètre temps de séjour, et un codage qualitatif (oui/non) de la rechloration en réseau (présence-absence) impactent probablement la qualité des prévisions fournies par les deux modèles.

Ces travaux ont mis en évidence les nombreuses difficultés existant pour disposer de données d'entrée fiables, notamment les temps de séjour en réseau, et exhaustives au sein des sites. Plusieurs sites initialement présagés pour contribuer à la modélisation n'ont ainsi pas été retenus faute de données concernant les variables pertinentes, et en particulier les temps de séjour hydraulique dans le réseau.

Ces modèles considèrent les THM dans leur globalité, sans individualiser les 4 espèces, ni s'intéresser à l'évolution dans le réseau de distribution de chaque espèce individuellement. Les résultats sont exprimés en $\mu\text{mol/L}$ de THM. La conversion en $\mu\text{g/L}$ peut s'effectuer en exprimant le résultat en équivalent chloroforme, en faisant l'hypothèse d'une conservation de la spéciation des THM entre la sortie d'usine et tout point du réseau, ou en utilisant la connaissance du site et les quelques mesures effectuées dans le réseau (auto-surveillance de l'exploitant, contrôle sanitaire en réseau).

L'utilisation de ces deux modèles pour prédire un niveau de THM en un point d'un réseau est possible très simplement, sous Excel®, sous réserve de la disponibilité des données concernant les variables incluses dans les modèles. Ces modèles peuvent notamment servir à déterminer les niveaux de concentrations en THM en différents points d'un même réseau, et ainsi contribuer à identifier les zones les plus critiques, proches du seuil réglementaire par exemple.

Les modèles n'ont pas été validés sur des eaux et des filières de traitements autres que celles utilisées pour leur construction. Si le domaine d'application de ces deux nouveaux modèles prend mieux en compte la diversité de qualité des eaux de consommation humaine de surface distribuées en France, il serait néanmoins intéressant de disposer d'autres jeux de données, sur des sites nouveaux, notamment avec des eaux d'origine souterraine, afin de vérifier leur capacité de généralisation.

Remerciements :

Ce travail n'aurait pas pu être réalisé sans la fourniture de données complètes par les exploitants des 8 sites d'études, que nous remercions pour leur participation. Nous remercions également pour leurs conseils avisés ou relecture du document Damien Mouly (InVS, CIRE Auvergne), Manuel J. Rodriguez (Université de Laval – Québec), Patrick Peigner (ARS – DT49), Pascal Beaudeau et Agnès Lefranc (InVS).

Bibliographie :

- [1] Davezac H, Grandguillot G, Robin A, Saout C. L'eau potable en France 2005-2006. 2008. 63 p.
- [2] Richardson SD, Plewa MJ, Wagner ED, Schoeny R, Demarini DM. Occurrence, genotoxicity, and carcinogenicity of regulated and emerging disinfection by-products in drinking water: a review and roadmap for research. *Mutat Res* 2007;636(1-3):178-242.
- [3] Weisel CP, Kim H, Haltmeier P, Klotz JB. Exposure estimates to disinfection by-products of chlorinated drinking water. *Environ Health Perspect* 1999;107(2):103-110.
- [4] Arrêté du 21 janvier 2010 relatif au programme de prélèvements et d'analyses du contrôle sanitaire pour les eaux fournies par un réseau de distribution, pris en application des articles R. 1321-10, R. 1321-15 et R. 1321-16 du code de la santé publique. *JO* 2010;(7 mars 2010)
- [5] Villanueva CM, Cantor KP, Grimalt JO, Malats N, Silverman D, Tardon A, *et al.* Bladder Cancer and Exposure to Water Disinfection By-Products through Ingestion, Bathing, Showering, and Swimming in Pools. *Am J Epidemiol* 2007;165(2):148-156.
- [6] Rahman MB, Driscoll T, Cowie C, Armstrong BK. Disinfection by-products in drinking water and colorectal cancer: a meta-analysis. *Int J Epidemiol* 2010;
- [7] Nieuwenhuijsen MJ, Martinez D, Grellier J, Bennett J, Best N, Iszatt N, *et al.* Chlorination disinfection by-products in drinking water and congenital anomalies: review and meta-analyses. *Environ Health Perspect* 2009;117(10):1486-1493.
- [8] Grellier J, Bennett J, Patelarou E, Smith RB, Toledano MB, Rushton L, *et al.* Exposure to disinfection by-products, fetal growth, and prematurity: A systematic review and meta-analysis. *Epidemiology* 2010;21(3):300-313.
- [9] Lewis C, Hoggatt KJ, Ritz B. The impact of different causal models on estimated effects of disinfection by-products on preterm birth. *Environ Res* 2011;111(3):371-376.
- [10] Hwang BF, Jaakkola JJK. Risk of stillbirth in the relation to water disinfection by-products: A population-based case-control study in Taiwan. *PLoS One* 2012;7(3)

- [11] Levallois P, Gingras S, Marcoux S, Legay C, Catto C, Rodriguez M, *et al.* Maternal exposure to drinking-water chlorination by-products and small-for-gestational-age neonates. *Epidemiology* 2012;23(2):267-276.
- [12] Mouly D, Joulin E, Rosin C, Beaudeau P, Zeghnoun A, Olszewski-Ortar A, *et al.* Variations in trihalomethane levels in three French water distribution systems and the development of a predictive model. *Water Res* 2010;44(18):5168-5179.
- [13] Mouly D, Joulin E, Rosin C, Beaudeau P, Zeghnoun A, Olszewski OA, *et al.* Les sous-produits de chloration dans l'eau destinée à la consommation humaine en France. Campagnes d'analyses dans quatre systèmes de distribution d'eau et modélisation de l'évolution des trihalométhanes. Saint-Maurice: Institut de veille sanitaire; 2009. 73 p.